Springer Proceedings in Mathematics & Statistics

Priti Kumar Roy · Xianbing Cao ·
Xue-Zhi Li · Pratulananda Das ·
Satya Deo *Editors*

Mathematical Analysis and Applications in Modeling

ICMAAM 2018, Kolkata, India, January 9–12

Springer

Springer Proceedings in Mathematics & Statistics

Volume 302

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Priti Kumar Roy · Xianbing Cao · Xue-Zhi Li ·
Pratulananda Das · Satya Deo
Editors

Mathematical Analysis and Applications in Modeling

ICMAAM 2018, Kolkata, India, January 9–12

Springer

Editors

Priti Kumar Roy
Department of Mathematics
Jadavpur University
Kolkata, West Bengal, India

Xianbing Cao
College of Science
Beijing Technology and Business University
Beijing, China

Xue-Zhi Li
College of Mathematics
and Information Science
Henan Normal University
Xinxiang, Xinjiang, China

Pratulananda Das
Department of Mathematics
Jadavpur University
Kolkata, West Bengal, India

Satya Deo
Harish-Chandra Research Institute
Allahabad, Uttar Pradesh, India

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-981-15-0421-1 ISBN 978-981-15-0422-8 (eBook)
https://doi.org/10.1007/978-981-15-0422-8

Mathematics Subject Classification (2010): 35A24, 34D23, 92B05, 37N25, 03E75, 34A08, 03F20, 03G27, 11K06, 34D20

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

International Conference on "Mathematical Analysis and Applications in Modeling" was held on January 9–12, 2018 at the Department of Mathematics, Jadavpur University, Kolkata, India. The talks, both invited and contributory, during the conference period covered various branches of Pure and Applied Mathematics. The present volume of this book series, entitled *Mathematical Analysis and Applications in Modeling*, is based on the selected invited and contributory talks during the abovementioned international conference. The conference was inaugurated by Pro-Vice-Chancellor of Jadavpur University. World-renowned scientist Dr. Gaston N'Guerekata, currently holding the chair of Associate Dean for Undergraduate Studies and University Distinguished Professor of School of Computer, Mathematical and Natural Sciences at the Morgan State University, Baltimore, USA, chaired the inaugural session. Dr. Igor Schreiber, the renowned Professor at the Department of Chemical Engineering (UCHI) of the University of Chemistry and Technology, Prague, Czech Republic, was the honorable guest who delivered the keynote address. There were 15 plenary lectures, 17 invited lectures, and 168 eight contributory lectures given by the participants. There were more than 300 participants from different parts of India and abroad. NBHM, ISI Kolkata and DST-PURSE provided us financial support to organize this conference without any hurdle.

Both the invited and contributory talks touch various areas of pure and applied mathematics and illustrate the latest advances in the field of mathematics, medical sciences, oil exploration from environmentally friendly, renewable resources and production, dynamical systems, biological sciences, algebra, analysis, etc.

This book contains 37 chapters from different mathematical fields. These chapters include reaction network theory, periodic evolution equations, optimization models, topology, compositional square root functions, atherosclerotic plaque formation, effects of unequal diffusion coefficients, cellular neural network model, coordinate search method, two-echelon supply chain, quasi-isometric invariants, statistical outlook, gravitational waves, Banach spaces, large-scale production of biodiesel, bifurcation control, and many other branches of mathematics. Thus this book is useful to gain knowledge about the streams of mathematics covered by the chapters.

We expect this book will draw an immense impact on theoretical and application fields of Mathematics and hence on the whole world of Science. Our endeavor is to enlighten each and every unknown or less known branch of Mathematics, which was the one-point goal of our international conference.

It is not easy to sum up all the topics based on different fields of Mathematics and its application in 37 chapters. But we and our team tried our level best to cover all the branches to make the unknown known. We can call our initiative successful if this book will help the students in their research and contribute to societal benefits.

We thank the Almighty, our friends, scholars, well-wishers, and our family for their cordial support and assistance in finishing this job and publishing this book series for a better knowledge of different fields of Mathematics. With all our willpower, energy, constructive mind, and time, we have tried to be the editor of such a scientifically useful book series of various research articles with supreme concern. If there is any inaccuracy or drawbacks in this book, we are ready to take all the responsibilities and we wish to receive fruitful suggestions for improvement of our future work as editors.

Kolkata, India	Prof. Dr. Priti Kumar Roy
Beijing, China	Prof. Dr. Xianbing Cao
Xinxiang, China	Prof. Dr. Xue-Zhi Li
Kolkata, India	Prof. Dr. Pratulananda Das
Allahabad, India	Prof. Dr. Satya Deo

Acknowledgements

We would like to thank Shamim Ahamad, Editor for Mathematics & Statistics, Springer IN to write a book on the subject Springer Proceedings in Mathematics & Statistics, with book title *Mathematical Analysis and Applications in Modeling*, *ICMAAM 2018, Kolkata, India, and January 9–12*. We would like to thank our Ph.D. and post-doctoral research students of Jadavpur University, Kolkata, working under the guidance of Prof. Priti Kumar Roy, for giving us the motivation and incessant effort to write a book on the above subject. Without their active help and support, it is quite an impossible task for us to complete the book.

In recent years, the Department of Mathematics, Jadavpur University, in collaboration with renowned nationally and internationally acclaimed mathematicians, has published a significant number of research works in diverse topics of mathematics. Such collaborations and awards have given birth to an innovative exchange of ideas in the past decade, which have amounted to successful conferences. ICMAAM 2018 was hopefully, thus, a forerunner to more significant contributions that came into the light of publishing this book. We are grateful to all of the contributing authors for their wholehearted cooperation, constant perseverance, and excellent support during this painful time without whom we could not have finished this book.

We extend our heartfelt salute to the contributors to the volume for their articles and christening us the honor to do this work. We would like to make our gratitude to those who reviewed materials for the book and provided valuable comments.

We are also grateful to all of our previous and ongoing research scholars namely: Dr. Amar Nath Chatterjee, Dr. Jayanta Mondal, Dr. Nikhilesh Sil, Dr. Biplab Sinha Mahapatra, Dr. Abhirup Datta, Dr. Fahad al Basir, Dr. Mithun Kumar Ghosh, Dr. Dibyendu Biswas, Dr. Shubhankar Saha, Mr. Amit Kumar Roy and Mr. Salil Ghosh and post-doctoral researcher Dr. Arnab Banerjee, and all colleagues for their tremendous effort and their incessant encouragement and motivation. Without their hard-strained efforts and impulsive collaboration, it is impossible for us as editor to edit this book.

We are pleased to acknowledge the financial support by the Government of India, National Board of Higher Mathematics.

Finally, we would like to thank the production team of Springer IN for their proofreads and valuable corrections made during the production process.

Prof. Dr. Priti Kumar Roy
Prof. Dr. Xianbing Cao
Prof. Dr. Xue-Zhi Li
Prof. Dr. Pratulananda Das
Prof. Dr. Satya Deo

Contents

Optimal Strategies of the Psoriasis Treatment by Suppressing the Interaction Between T-Lymphocytes and Dendritic Cells 1
Ellina V. Grigorieva and Evgenii N. Khailov

The Use of Reaction Network Theory for Finding Network Motifs in Oscillatory Mechanisms and Kinetic Parameter Estimation 13
Igor Schreiber, Vuk Radojković, František Muzika, Radovan Jurašek and Lenka Schreiberová

An Existence Result for Some Fractional Evolution Equation with Nonlocal Conditions and Compact Resolvent Operator 25
G. M. N'Guérékata

Impact of Tobacco Smoking on the Prevalence of Tuberculosis Infection: A Mathematical Study 33
Mini Ghosh

Modelling the Effects of Stigma on Leprosy 53
Stephen G. Mosher, Christian Costris-Vas and Robert Smith?

One Simple Model—Various Complex Systems 69
Urszula Foryś

Reliability Analysis of Multi-state Two-Dimensional System by Universal Generating Function 85
K. Meenakshi and S. B. Singh

Quantum Symmetry of Classical Spaces 101
Debashish Goswami

Quasi-isometry and Rigidity 109
Parameswaran Sankaran

Existence of Compositional Square-Roots of Functions 117
V. Kannan

On Generalized Statistical Convergence and Strongly Summable Functions of Weight g 125
Ekrem Savaş and Rabia Savaş

On $\mathcal{I}$-Statistically Order Pre Cauchy Sequences in Riesz Spaces 133
Pratulananda Das and Ekrem Savas

On Ball Dentable Property in Banach Spaces 145
Sudeshna Basu

Some Observations Concerning Polynomial Convexity 151
Sushil Gorai

Solution of Large-Scale Multi-objective Optimization Models for Saltwater Intrusion Control in Coastal Aquifers Utilizing ANFIS Based Linked Meta-Models for Computational Feasibility and Efficiency 163
Dilip Kumar Roy and Bithin Datta

Cost Driven Optimization of Microgrid Under Environmental Uncertainties Using Different Improved PSO Models 173
Meenakshi De, G. Das and K. K. Mandal

Global Exponential Stability of Non-autonomous Cellular Neural Network Model with Time Varying Delays 187
M. Chowdhury and P. Das

Topological Configuration in the Heisenberg Spin Sequence and DNA 197
Subhamoy Singha Roy

Modelling Studies Focusing on Microphytobenthos and Its Role in Benthic-Pelagic Coupling 209
Swagata Sinha, Arnab Banerjee, Nabyendu Rakshit and Santanu Ray

Overview of Ecological Economics and Ecosystem Services Consequences from Shrimp Culture 225
Suvendu Das, Sagar Adhurya and Santanu Ray

Recent Trends of Ecosystem Health and Biodiversity Status of Pelagic-Benthic Coupled System in Indian Estuaries with Special Emphasis on Hooghly Estuary, India 237
Nabyendu Rakshit, Arnab Banerjee, Swagata Sinha and Santanu Ray

Guanotrophication by Waterbirds in Freshwater Lakes: A Review on Ecosystem Perspective 253
Sagar Adhurya, Suvendu Das and Santanu Ray

On the Volatility of High Frequency Stock Index Based on SV Model of MCMC . . . 271
Y. X. Zheng, Y. H. Zhang and X. H. Lu

The Effects of Unequal Diffusion Coefficients on Spatiotemporal Pattern Formation in Prey Harvested Reaction-Diffusion Systems . . . 279
Lakshmi Narayan Guin

Optimal Pricing Strategy in a Two-Echelon Supply Chain with Admissible Advanced and Delayed Payments . . . 293
B. C. Giri, R. Bhattacharjee and T. Maiti

Studies on Atherosclerotic Plaque Formation: A Mathematical Approach . . . 307
Debasmita Mukherjee, Lakshmi Narayan Guin and Santabrata Chakravarty

Threshold of a Stochastic Delayed SIR Epidemic Model with Saturation Incidence . . . 319
Yanli Zhou

Semi-stationary Equilibrium Strategies in Non-cooperative *N*-person Semi-Markov Games . . . 331
Prasenjit Mondal and Sagnik Sinha

Max Plus Algebra, Optimization and Game Theory . . . 345
Dipti Dubey, S. K. Neogy and Sagnik Sinha

Dynamics of Infectious Diseases with Periodic Awareness Campaigns . . . 363
Fahad Al Basir

A Vendor-Buyer Supply Chain Model for Deteriorating Item with Quadratic Time-Varying Demand and Pro-rata Warranty Policy . . . 371
B. Samanta, B. C. Giri and K. S. Chaudhuri

Pest Control of Jatropha curcas Plant for Different Response Functions . . . 385
Arunabha Sengupta, J. Chowdhury, Xianbing Cao and Priti Kumar Roy

Global Dynamics of a TB Model with Classes Age Structure and Environmental Transmission . . . 403
Yan-Xia Dang, Juan Wang, Xue-Zhi Li and Mini Ghosh

Gravitational Waves: The Mathematical Background . . . 449
Subenoy Chakraborty

Current Trends in the Bifurcation Methods of Solutions of Real World Dynamical Systems . . . 459
Jocirei D. Ferreira and V. Sree Hari Rao

A Modified Coordinate Search Method Based on Axes Rotation 499
Suvra Kanti Chakraborty and Geetanjali Panda

Dynamics of the Logistic Prey Predator Model in Crisp and Fuzzy Environment 511
S. Tudu, N. Mondal and S. Alam

About the Editors

Priti Kumar Roy is Professor in the Department of Mathematics, Jadavpur University, Kolkata, India. Earlier, he served at several government colleges in different parts of West Bengal, India. He is an eminent member of several national and international societies like Biomathematical Society of India, International Association of Engineers, European Society of Clinical Microbiology and Infectious Diseases and European Society for Mathematical and Theoretical Biology. Professor Roy has received the Best Paper Award at the World Congress on Engineering 2010 in London. He was awarded the "Siksha Ratan" Award in 2012. He is also recipient of Royal Society of Edinburg and Poland Academy of Science Fellowship through INSA. He has published a significant number of research papers on control therapeutic approaches and host–pathogen interactions on infectious as well as noninfectious diseases to enlighten new insights on the subjects. With over 120 peer-reviewed research papers, more than 50 invited talks (abroad), 30 invited talks (India) in different international and national institutes, Professor Roy is a dedicated researcher in mathematical modeling and has expertise in numerical and analytical solutions to complex problems on real-life system dynamics. He has published two books on "Mathematical Models for Therapeutic Approaches to Control HIV Disease Transmission" (Springer in 2015) and "Mathematical Models for Therapeutic Approaches to Control Psoriasis" (Springer in 2019). He also edited two books on "Insight and Control of Infectious Disease in Global Scenario (by Intech Publishers)" and "A collection of Writings Objective Articulation from Socio-Academic Spectrum". His research interests are also devoted to epidemiological issues on the chronic infectious disease such as HIV, Cutaneous Leishmaniasis and Leprosy, Biodiesel Production, Enzyme Kinetics, Methanol Toxicity, etc.

Professor Roy also works on the neglected tropical disease like Psoriasis and has formulated robust mathematical models on the dynamics of such disease.

Xianbing Cao is Professor and Director, Beijing Technology and Business University, China. Earlier he served as Dean, School of Science, BTBU. He received his Ph.D. in Stochastic Control Theory from the Chinese Academy of Sciences, in 2002. He is an eminent researcher in the field of Biostatistics. Internationally as a reputed Statistician as well as Mathematician, several research papers he has published in internationally reputed journals. He is the author or co-author of two books in the areas of stochastic and adaptive control systems and probability theory. His research interests include stochastic control systems, system identification, and financial time series analysis.

Xue-Zhi Li is Professor and Vice President, Henan Normal University, China. Before joining this administrative post, he served as a Vice President at Anyang Institute of Technology and also served as a Dean at the College of Mathematics and Information Science, Henan Normal University, Xinxiang, China.

He is an internationally reputed Mathematical Biologist. He has authored 2 books and more than 176 papers in national and international journals of repute.

Pratulananda Das is Professor in the Department of Mathematics, Jadavpur University, Kolkata, India. He received his Ph.D. from Kalyani University in 2000. He was awarded the "INSA Teachers Award" from the Indian National Science Academy, New Delhi, India, in 2017. He is a life member of several national and international societies like The National Academy of Sciences India, American Mathematical Society, Indian Science Congress Association, Ramanujan Mathematical Society, and Calcutta Mathematical Society. His research interests include set-theoretic and general topology, sequences and summability theory, and analysis.

Satya Deo is Senior Scientist at The National Academy of Sciences, India, and Honorary Professor at Harish-Chandra Research Institute, Allahabad, India. He was awarded the Distinguished Service Award by the Mathematical Association of India for outstanding contributions to the cause of teaching and research in 2006 and the Indian Science Congress Gold Medal by the then Prime Minister of India, in 2010. He has guided 17 Ph.D. students and published over 70 papers in reputed international journals and proceedings. His area of specialization includes algebraic and differential topology, topological and differentiable group actions, homology–cohomology theories, cohomological dimension theory, Burnside rings, Hopfian and co-Hopfian rings, spline modules, and topological combinatorics.

Optimal Strategies of the Psoriasis Treatment by Suppressing the Interaction Between T-Lymphocytes and Dendritic Cells

Ellina V. Grigorieva and Evgenii N. Khailov

Abstract This report contains the results devoted to the study of a mathematical model of psoriasis, proposed by P. K. Roy. This model is formulated as a Cauchy problem for the system of three nonlinear differential equations that describe the relationships between the concentrations of T-lymphocytes, keratinocytes and dendritic cells, the interactions of which cause the occurrence of psoriasis. Moreover, in this model we include a bounded scalar control responsible for a dose of a medication that suppresses the interaction of T-lymphocytes and dendritic cells. On the given time interval, for the control mathematical model, a problem of minimizing the concentration of keratinocytes at the end of time interval is considered. To analyze this problem, the Pontryagin maximum principle is applied. The adjoint system and the maximum condition for the optimal control are written. Using the corresponding system of differential equations, the switching function describing the behavior of the optimal control is studied. Such a system of equations allows us to determine the type of the optimal control: whether this control is only of a bang-bang type, or, in addition to the portions of a bang-bang type, it also contains a singular arc. When a singular arc occurs, the report discusses its order, the fulfillment of the necessary optimality condition for it, and its concatenation with portions of a bang-bang type. The obtained analytical results are illustrated by numerical calculations. The corresponding conclusions are made.

Keywords Psoriasis · Nonlinear system · Optimal control · Pontryagin maximum principle · Switching function · Singular arc

E. V. Grigorieva (✉)
Texas Woman's University, Denton, TX 76204, USA
e-mail: egrigorieva@mail.twu.edu

E. N. Khailov
Moscow State Lomonosov University, Moscow 119992, Russia
e-mail: khailov@cs.msu.su

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_1

1 Introduction

Psoriasis is an autoimmune disease with symptoms of chronic inflammation of the skin [6, 8]. In psoriasis, skin cells grow very rapidly, which leads to the appearance of red, dry and flake rashes. The main organ which is damaged is the skin, although other organs and systems of a person, in particular nails and joints, can be affected besides this. Psoriasis starts from the processes taking place in the epidermis. In the deep layer of the epidermis, immature skin cells called keratinocytes are formed. They produce keratin, a hard protein that is a building material for hair, nails and skin. Normally, keratinocytes grow and move from the lower layer to the surface of the skin almost imperceptibly. This process takes about a month. In people with psoriasis, keratinocytes proliferate very rapidly and move from the deep layer to the surface in about four days. The skin cannot get rid of these cells quickly enough, so that in a short time their amount increases dramatically, which leads to the formation of densified, dry patches on the skin or plaques. The lower layer of the dermis with blood, lymphatic vessels and nerves becomes inflamed and swollen.

Skin is an important immune organ. Specific skin cells, such as keratinocytes, promote the maturation of T-lymphocytes, which are the main element of the immune system of the skin. T-lymphocytes make up 90% of all lymphocytes of the skin and are located mainly in the upper and middle layers of the skin. The main function of dendritic cells is the presentation of antigens to T-lymphocytes. This means that they absorb antigens from the environment. Then they "process" them to the form that T-lymphocytes are able to recognize and develop an immune response. They also perform important immune-regulatory functions. Dendritic cells and activated T-lymphocytes play an important role in the development of psoriasis. The interactions between these two groups of cells trigger mechanisms leading ultimately to the development of the inflammatory process and the formation of psoriatic skin lesions.

Adequate treatment of psoriasis is challenging and drugs, leading to a complete cure, do not yet exist. Significant progress has been made, both in understanding the mechanisms of the disease and in finding new ways of treatment, and in standardizing the assessment of the severity of the disease. Mathematical models are effectively used to predict the behavior of skin cells, both in normal and in pathological states.

2 Optimal Control Problem

On a given time interval $[0, T]$ we consider the nonlinear control system of differential equations:

$$\begin{cases} l'(t) = \sigma - \delta v(t)l(t)m(t) - \gamma_1 l(t)k(t) - \mu l(t), \\ k'(t) = (\beta + \delta)v(t)l(t)m(t) + \gamma_2 l(t)k(t) - \lambda k(t), \\ m'(t) = \rho - \beta v(t)l(t)m(t) - \nu m(t), \\ l(0) = l_0, \ k(0) = k_0, \ m(0) = m_0; \ l_0, k_0, m_0 > 0. \end{cases} \tag{1}$$

It describes the interactions of various types of cells in a human body with drug therapy of psoriasis [3, 10, 11]. In system (1), $l(t)$, $k(t)$, and $m(t)$ are the concentrations of T-lymphocytes, keratinocytes and dendritic cells; l_0, k_0, m_0 are their initial conditions, respectively. The values σ, ρ, μ, λ, ν, γ_1, γ_2, δ, β are the given positive parameters of this system, which have the following meaning. The values σ and ρ are the appropriate inflow rates of T-lymphocytes and dendritic cells, μ and ν are the removal rates of these cells, respectively; λ is the decay rate of keratinocytes. In addition, the rate of activation of keratinocytes due to T-lymphocytes is indicated by γ_1 and the rate of keratinocytes growth is denoted by γ_2. The value δ is the activation rate of T-lymphocytes by dendritic cells, β is conversely the activation rate of dendritic cells due to T-lymphocytes. The interactions between T-lymphocytes and dendritic cells help to form keratinocytes through some cell biological procedures and thus the concentrations of both T-lymphocytes and dendritic cells are reduced by the terms δvlm and βvlm, respectively. On the other hand, under mixing homogeneity, the combined interaction of T-lymphocytes and dendritic cells contributes to the growth of concentration of epidermal keratinocytes by the term $(\beta + \delta)vlm$. Model (1) was kindly provided for analysis by Professor P. K. Roy (Centre for Mathematical Biology and Ecology, Department of Mathematics, Jadavpur University, Kolkata, India).

In system (1), $v(t)$ is a control function that satisfies the constraints:

$$0 < v_{\min} \leq v(t) \leq 1. \tag{2}$$

We note that the control $v(t)$ is an auxiliary. It is introduced into system (1) to simplify analytical analysis. The corresponding physical control $\widetilde{v}(t)$ in the same system is related to the control $v(t)$ by the formula $\widetilde{v}(t) = 1 - v(t)$. Therefore, where the auxiliary control $v(t)$ has a maximum value of 1, the appropriate physical control $\widetilde{v}(t)$ takes a minimum value of 0, and vice versa. The physical control $\widetilde{v}(t)$ is responsible for the dose of drug, which suppresses the interaction of T-lymphocytes and dendritic cells. Despite its importance, in the following arguments we focus on the analysis of auxiliary control $v(t)$. The set of admissible controls $\Omega(T)$ consists of all possible Lebesgue measurable functions $v(t)$ that satisfy constraints (2) for almost all $t \in [0, T]$.

Now, let us define the following positive constants:

$$\begin{aligned} &\eta = \min\{\mu; \lambda; \nu\}, \quad K = \rho\left(1 + \beta^{-1}\delta\right)\gamma_1 + \sigma\gamma_2, \\ &M = \gamma_2 l_0 + \gamma_1 k_0 + \left(1 + \beta^{-1}\delta\right)\gamma_1 m_0 + \eta^{-1}K, \end{aligned}$$

and introduce the set:

$$\Theta = \Big\{(l, m, k) : l > 0,\ m > 0,\ k > 0,\ \gamma_2 l + \gamma_1 k + \gamma_1\left(1 + \beta^{-1}\delta\right)m < M\Big\}.$$

Then, the boundedness, positivity, and extendability of solution of this system is established by the following lemma.

Lemma 1 *Let the inclusion* $(l_0, m_0, k_0) \in \Theta$ *be valid. For an arbitrary admissible control* $v(t)$*, the corresponding absolutely continuous solution* $(l(t), k(t), m(t))$ *to system (1) are defined on the entire interval* $[0, T]$ *and satisfy the inclusion:*

$$(l(t), m(t), k(t)) \in \Theta, \quad t \in (0, T]. \tag{3}$$

The proof of Lemma 1 is fairly straightforward and we omit it. Proofs of such statements are given, for examples, in [2, 4, 10].

For system (1) on the set of admissible controls $\Omega(T)$, we consider the problem of minimizing the functional

$$J(v) = k(T), \tag{4}$$

which means the concentration of keratinocytes at the final moment T of the psoriasis treatment. Justification for using such a functional for system (1) was previously discussed in [5].

In the minimization problem (1), (4) the restrictions (3) provide the existence of the optimal control $v_*(t)$ and the corresponding optimal solution $(l_*(t), k_*(t), m_*(t))$ (see [7]).

Finally, based on the results from [3, 11], we assume that the inequalities:

$$\gamma_1 \neq \gamma_2, \qquad (\beta + \delta)\gamma_1 > \delta\gamma_2, \qquad \lambda > \mu, \qquad \lambda > \nu$$

are valid.

3 Pontryagin Maximum Principle

In order to analyze the optimal control $v_*(t)$ and the corresponding optimal solution $(l_*(t), k_*(t), m_*(t))$, we apply the Pontryagin maximum principle [9]. Firstly, we write down the Hamiltonian

$$\begin{aligned} H(l, k, m, v, \psi_1, \psi_2, \psi_3) = {} & (\sigma - \delta vlm - \gamma_1 lk - \mu l)\psi_1 \\ & +((\beta + \delta)vlm + \gamma_2 lk - \lambda k)\psi_2 + (\rho - \beta vlm - \nu m)\psi_3, \end{aligned}$$

where ψ_1, ψ_2, ψ_3 are adjoint variables.

Secondly, we calculate the required partial derivatives:

$$\begin{aligned} H'_l(l, k, m, v, \psi_1, \psi_2, \psi_3) &= vm(-\delta\psi_1 + (\beta + \delta)\psi_2 - \beta\psi_3) \\ &\quad +k(\gamma_2\psi_2 - \gamma_1\psi_1) - \mu\psi_1, \\ H'_k(l, k, m, v, \psi_1, \psi_2, \psi_3) &= l(\gamma_2\psi_2 - \gamma_1\psi_1) - \lambda\psi_2, \\ H'_m(l, k, m, v, \psi_1, \psi_2, \psi_3) &= vl(-\delta\psi_1 + (\beta + \delta)\psi_2 - \beta\psi_3) - \nu\psi_3, \\ H'_v(l, k, m, v, \psi_1, \psi_2, \psi_3) &= lm(-\delta\psi_1 + (\beta + \delta)\psi_2 - \beta\psi_3). \end{aligned}$$

Then, in accordance with the Pontryagin maximum principle, for the optimal control $v_*(t)$ and the optimal solution $(l_*(t), k_*(t), m_*(t))$ there exists a vector-function $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$ such that:

- $\psi_*(t)$ is a nontrivial solution of the adjoint system:

$$\begin{cases} \psi_1^{*\prime}(t) = -v_*(t)m_*(t)(-\delta\psi_1^*(t) + (\beta+\delta)\psi_2^*(t) - \beta\psi_3^*(t)) \\ \qquad\quad -k_*(t)(\gamma_2\psi_2^*(t) - \gamma_1\psi_1^*(t)) + \mu\psi_1^*(t), \\ \psi_2^{*\prime}(t) = -l_*(t)(\gamma_2\psi_2^*(t) - \gamma_1\psi_1^*(t)) + \lambda\psi_2^*(t), \\ \psi_3^{*\prime}(t) = -v_*(t)l_*(t)(-\delta\psi_1^*(t) + (\beta+\delta)\psi_2^*(t) - \beta\psi_3^*(t)) + \nu\psi_3^*(t), \\ \psi_1^*(T) = 0,\ \psi_2^*(T) = -1,\ \psi_3^*(T) = 0; \end{cases} \tag{5}$$

- the control $v_*(t)$ maximizes the Hamiltonian

$$H(l_*(t), k_*(t), m_*(t), v, \psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$$

with respect to the variable $v \in [v_{\min}, 1]$ for almost all $t \in [0, T]$, and therefore it satisfies the relationship:

$$v_*(t) = \begin{cases} 1 & \text{, if } L(t) > 0, \\ \text{any } v \in [v_{\min}, 1] & \text{, if } L(t) = 0, \\ v_{\min} & \text{, if } L(t) < 0 \end{cases} \tag{6}$$

in which, by Lemma 1, the function

$$L(t) = -\delta\psi_1^*(t) + (\beta+\delta)\psi_2^*(t) - \beta\psi_3^*(t) \tag{7}$$

is the switching function describing the behavior of the control $v_*(t)$ according to formula (6).

4 Properties of the Switching Function

An analysis of the function $L(t)$ leads to the validity of the following lemma.

Lemma 2 *There is such a value $t_0 \in [0, T)$ that on the interval $(t_0, T]$ the switching function $L(t)$ is negative.*

Proof The functions $\psi_1^*(t)$, $\psi_2^*(t)$, $\psi_3^*(t)$, as the components of the absolutely continuous solution $\psi_*(t)$ to system (5), are absolutely continuous as well. Hence, by formula (7), the switching function $L(t)$ is also absolutely continuous, and therefore a continuous function. Due to formula (7) and the initial conditions of system (5), it takes the negative value at $t = T$:

$$L(T) = -(\beta+\delta) < 0.$$

Then, the stability of the sign of the continuous function $L(t)$ yields the required fact. This completes the proof.

Corollary 1 *From Lemma 2 and formula (6), it follows the relationship:*

$$v_*(t) = v_{\min}, \quad t \in (t_0, T].$$

Now, we introduce positive constants:

$$\alpha = \gamma_2^{-1}((\beta + \delta)\gamma_1 - \delta\gamma_2), \quad \varepsilon = \alpha(\lambda - \nu) + \delta(\lambda - \mu),$$

and then also the following functions:

$$\begin{aligned}
g_{11}(t) &= v_*(t)(\delta m_*(t) + \beta l_*(t)) + \nu, \\
g_{21}(t) &= v_*(t)m_*(t)(\gamma_1(\delta k_*(t) - (\beta + \delta)l_*(t)) + \delta(\mu - \nu)), \\
g_{22}(t) &= (\lambda - \mu)\varepsilon^{-1}\gamma_1(\delta k_*(t) - (\beta + \delta)l_*(t)) \\
&\quad + \varepsilon^{-1}(\alpha(\lambda - \nu)\lambda + \delta(\lambda - \mu)(\lambda - \nu + \mu)), \\
g_{31}(t) &= \gamma_1 v_*(t)m_*(t), \quad g_{32}(t) = (\lambda - \mu)\gamma_1\varepsilon^{-1}, \\
g_{33}(t) &= (\gamma_1 k_*(t) - \gamma_2 l_*(t)) - (\lambda - \mu)\varepsilon^{-1}\gamma_1(\delta k_*(t) - (\beta + \delta)l_*(t)) \\
&\quad + \varepsilon^{-1}(\alpha(\lambda - \nu)\mu + \delta(\lambda - \mu)\nu).
\end{aligned}$$

In addition, let us define the auxiliary functions:

$$\begin{aligned}
Q(t) &= \gamma_2\psi_2^*(t) - \gamma_1\psi_1^*(t), \\
G(t) &= (\delta k_*(t) - (\beta + \delta)l_*(t) + \gamma_2^{-1}(\beta + \delta)(\lambda - \nu))Q(t) + \varepsilon\psi_1^*(t),
\end{aligned}$$

and introduce the following function of two variables:

$$\begin{aligned}
\Phi(l, k) = &-\alpha\delta\gamma_1(\lambda - \nu)k^2 - \alpha(\beta + \delta)\gamma_2(\mu - \nu)l^2 - 2\delta(\beta + \delta)\gamma_1(\lambda - \mu)lk \\
&-(\beta + \delta)(\alpha(\lambda - \nu)\nu + \delta(\lambda - \mu)(2(\mu - \nu) + \nu))l \\
&+\delta(\alpha(\lambda - \nu)(2(\lambda - \mu) + \nu) + \delta(\lambda - \mu)(2(\lambda - \nu) + \nu))k \\
&+(\beta + \delta)(\sigma\varepsilon + \gamma_2^{-1}\delta(\lambda - \mu)(\lambda - \nu)(\mu - \nu)).
\end{aligned}$$

Then, using the equations of systems (1) and (5), we obtain the system of differential equations for the functions $L(t)$, $G(t)$ and $Q(t)$:

$$\begin{cases}
L'(t) = g_{11}(t)L(t) + G(t), \quad t \in [0, T], \\
G'(t) = g_{21}(t)L(t) + g_{22}(t)G(t) \\
\qquad + \left(2\delta(\beta + \delta)v_*(t)l_*(t)m_*(t) - \varepsilon^{-1}\Phi(l_*(t), k_*(t))\right) Q(t), \\
Q'(t) = g_{31}(t)L(t) + g_{32}(t)G(t) + g_{33}(t)Q(t), \\
L(T) = -(\beta + \delta), \quad Q(T) = -\gamma_2, \\
G(T) = -\left(\gamma_2(\delta k_*(T) - (\beta + \delta)l_*(T)) + (\beta + \delta)(\lambda - \nu)\right).
\end{cases} \tag{8}$$

Now, let us analyze formula (6). We have the following conclusions.

- If for some value $t_{(+)} \in [0, T]$ the switching function $L(t)$ is positive, then it is also positive in some neighborhood of this value. Then the corresponding optimal control $v_*(t)$ takes the value 1 in this neighborhood.
- Similarly, if for some value $t_{(-)} \in [0, T]$ the switching function $L(t)$ is negative, then it is also negative in some neighborhood of this value. Then the corresponding optimal control $v_*(t)$ takes the value $v_{\min}$ in this neighborhood.
- Since the function $L(t)$ is absolutely continuous, it can vanish either at separate points or on certain intervals. In the first case, the optimal control $v_*(t)$ is bang-bang, it takes only values $v_{\min}$ and 1. In this case, the value $t_0 \in (0, T)$, at which $L(t_0) = 0$ and such that passing this point the function $L(t)$ changes its sign, is a switching of this control. Then, naturally, the question of estimating the number of zeros of the switching function $L(t)$ arises. Such a question relates to the estimate of the number of switchings of the control $v_*(t)$, and here system (8) plays a significant role. In the second case, at such intervals the optimal control $v_*(t)$ has singular arcs [12, 13]. This phenomenon requires additional studies also using this system.

Now, let us study the existence of a singular arc of the optimal control $v_*(t)$, which means that the switching function $L(t)$ can vanish identically on some interval $\varDelta \subset [0, T]$. We use the first two equations of system (8) to find on this interval the first two derivatives of the function $L(t)$:

$$L'(t)\Big|_{L(t)=0} = 0, \qquad L''(t)\Big|_{L(t)=0,\ L'(t)=0} = 0.$$

As a result, the equality can be obtained:

$$\big(2\delta(\beta+\delta)v_*(t)l_*(t)m_*(t) - \varepsilon^{-1}\varPhi(l_*(t), k_*(t))\big)\, Q(t) = 0, \quad t \in \varDelta. \tag{9}$$

From the analysis of formula (9) the following conclusions are made.

- The second derivative of the function $L(t)$ contains the control $v_*(t)$. This means that the order q of the singular arc is equal to one.
- On the interval $\varDelta$ the optimal control $v_*(t)$ is singular and is given by the formula:

$$v^*_{\text{sing}}(t) = \frac{\varPhi(l^*_{\text{sing}}(t), k^*_{\text{sing}}(t))}{2\varepsilon\delta(\beta+\delta)l^*_{\text{sing}}(t)m^*_{\text{sing}}(t)}\,. \tag{10}$$

From this formula it follows that such a control has the form of feedback, that is, it depends only on the functions $l^*_{\text{sing}}(t)$, $k^*_{\text{sing}}(t)$, $m^*_{\text{sing}}(t)$, which are the corresponding components of the optimal solution $(l_*(t), k_*(t), m_*(t))$ on this interval. We assume that such a control is admissible, that is, the inclusion $v^*_{\text{sing}}(t) \in [v_{\min}, 1]$ is valid everywhere on the interval $\varDelta$.

- The necessary optimality condition of a singular arc (the Kelly condition [12, 13]) has the form:
$$2\delta(\beta+\delta)l^*_{\text{sing}}(t)m^*_{\text{sing}}(t)Q(t) \geq 0, \quad t \in \Delta. \tag{11}$$

The non-triviality of the vector-function $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$ implies the validity of the following lemma.

Lemma 3 *On the interval Δ the function $Q(t)$ is sign-definite, that is, it only takes either positive or negative values.*

Corollary 2 *From Lemma 3 and inequality (11) it follows that the Kelly condition either holds in the strengthened form:*

$$2\delta(\beta+\delta)l^*_{\text{sing}}(t)m^*_{\text{sing}}(t)Q(t) > 0, \quad t \in \Delta,$$

or is not satisfied at all, that is:

$$2\delta(\beta+\delta)l^*_{\text{sing}}(t)m^*_{\text{sing}}(t)Q(t) < 0, \quad t \in \Delta.$$

Finally, when the inclusion $v^*_{\text{sing}}(t) \in (v_{\min}, 1)$ is true for all $t \in \Delta$, Lemma 2, Corollary 2 and formula (10) provide the concatenation of the singular arc with the other bang-bang portions of the control $v_*(t)$.

5 Numerical Results

Further, only numerical investigation of optimal control $v_*(t)$ is possible. For the corresponding numerical calculations, the following values of the parameters and initial conditions of system (1) were used [3, 11], as well as the control constraints (2):

$\sigma = 15.0$	$\rho = 3.6$	$\beta = 0.4$	$\delta = 0.005$
$\mu = 0.01$	$\nu = 0.02$	$\gamma_1 = 0.8$	$\gamma_2 = 0.05$
$l_0 = 100.0$	$k_0 = 40.0$	$m_0 = 50.0$	
$v_{\min} = 0.3$	$T = 100.0$		

The numerical calculations were carried out using the software "BOCOP 2.0.5" (see [1]), and are shown in Figs. 1 and 2.

In Fig. 3 the surface $\Phi(l,k)$ is presented. It can be seen that positive values of the function $\Phi(l,k)$ in formula (10) in the region of variation of the variables l and k provide the admissibility of singular control $v^*_{\text{sing}}(t)$.

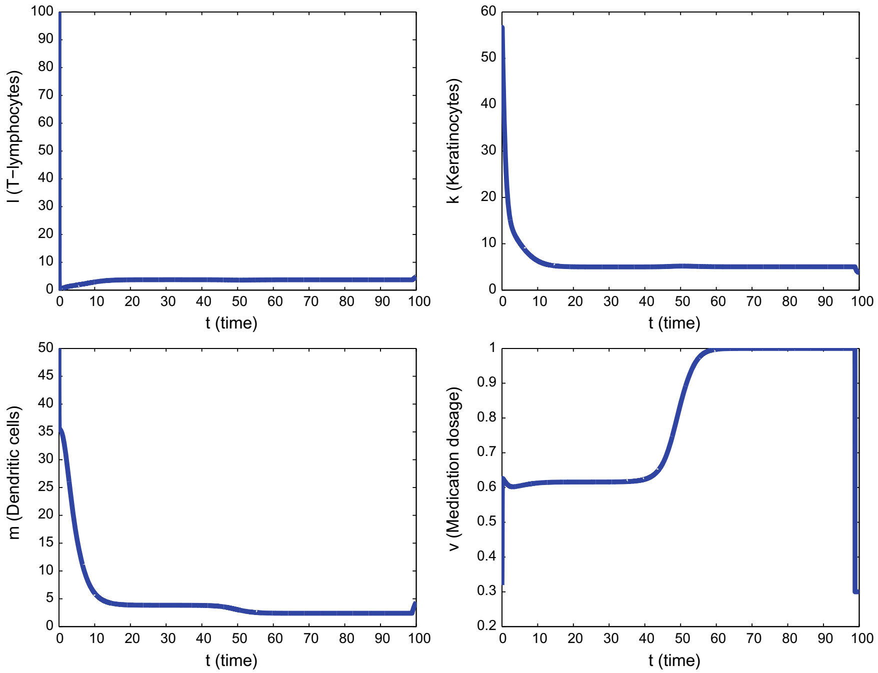


Fig. 1 Graphs of optimal solutions and optimal control for $\lambda = 0.9$: upper row: $l_*(t)$, $k_*(t)$; lower row: $m_*(t)$, $u_*(t)$

6 Conclusions

Physical optimal control $\widetilde{v}_*(t)$ according to Figs. 1 and 2 describes the situation when, first there is the period of the psoriasis treatment with greatest intensity. Next, it is followed by the period of the treatment with a smooth decrease in the dose of the used medication from the greatest intensity to the lower intensity. Then, there is a period of the psoriasis treatment with lower intensity, and finally the switching occurs to the period of the treatment with greatest intensity. Also, we emphasize that in all performed numerical calculations, the optimal concentration of keratinocytes $k_*(t)$ decreases to the end T to the level that is the minimal for the entire period $[0, T]$ of the psoriasis treatment (see Figs. 1 and 2).

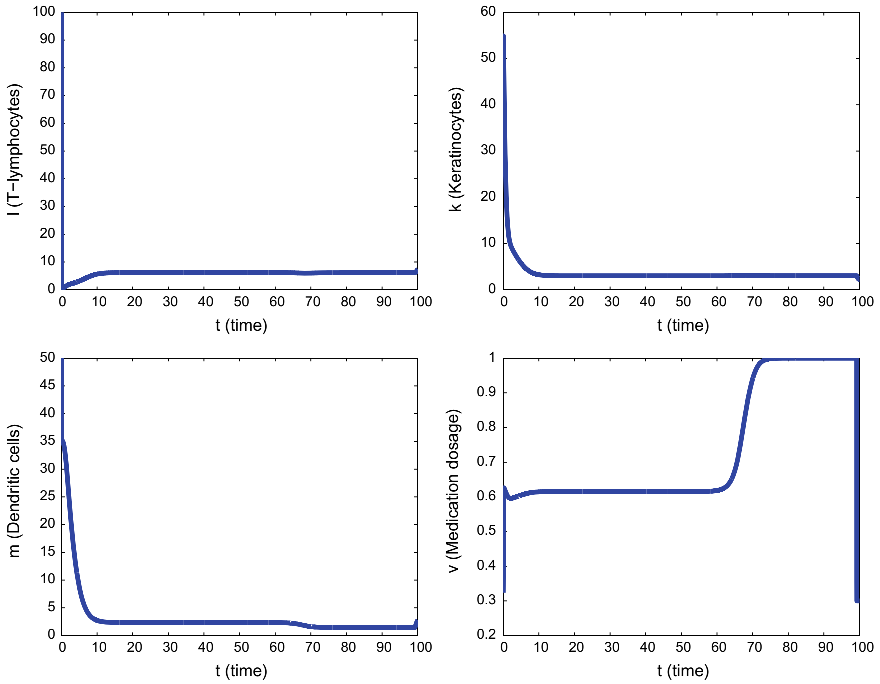


Fig. 2 Graphs of optimal solutions and optimal control for $\lambda = 1.5$: upper row: $l_*(t)$, $k_*(t)$; lower row: $m_*(t)$, $u_*(t)$

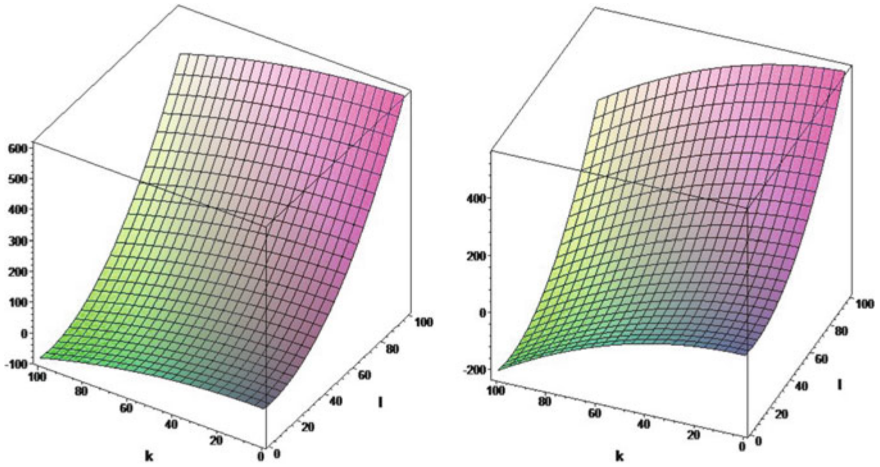


Fig. 3 Surface $\Phi(l, k)$ for $\lambda = 0.9$ and $\lambda = 1.5$

Acknowledgements E. N. Khailov was funded by Russian Foundation for Basic Research according to the research project 18-51-45003 IND_a.

References

1. Bonnans, F., Martinon, P., Giorgi, D., Grélard, V., Maindrault, S., Tissot, O., Liu, J.: BOCOP 2.0.5—User Guide. http://bocop.org. Accessed 8 Feb 2017
2. Datta, A., Roy, P.K.: T-cell proliferation on immunopathogenic mechanism of psoriasis: a control based theoretical approach. Control Cybern. **2013**, 23–42 (2013)
3. Datta, A., Li, X.-Z., Roy, P.K.: Drug therapy between T-cells and DCS reduces the excess production of keratinicytes: ausal effect of psoriasis. Math. Sci. Int. Res. J. **3**, 452–456 (2014)
4. Datta, A., Kesh, D.K., Roy, P.K.: Effect of $CD4^+$ T-cells and $CD8^+$ T-cells on psoriasis: a mathematical study. Imhotep Math. Proc. **3**, 1–11 (2016)
5. Grigorieva, E., Khailov, E., Deignan, P.: Optimal treatment strategies for control model of psoriasis. In: Proceedings of the SIAM Conference on Control and its Applications (CT17), Pittsburgh, Pennsylvania, USA, pp. 86–93, 10–12 July 2017
6. Gudjonsson, J.E., Johnston, A., Sigmundsdottir, H., Valdimarsson, H.: Immunopathogenic mechanisms in psoriasis. Rev. Clin. Exp. Immunol. **135**, 1–8 (2004)
7. Lee, E.B., Marcus, L.: Foundations of Optimal Control Theory. Wiley, New York (1967)
8. Lowes, M.A., Suárez-Fariñas, M., Krueger, J.G.: Immunology of psoriasis. Annu. Rev. Immunol. **32**, 227–255 (2014)
9. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)
10. Roy, P.K., Bradra, J., Chattopadhyay, B.: Mathematical modeling on immunopathogenesis in chronic plaque of psoriasis: a theoretical study. Lect. Notes Eng. Comput. Sci. **1**, 550–555 (2010)
11. Roy, P.K., Datta, A.: Impact of cytokine release in psoriasis: a control based mathematical approach. J. Nonlinear Evol. Equ. Appl. **2013**, 23–42 (2013)
12. Schättler, H., Ledzewicz, U.: Geometric Optimal Control. Theory, Methods and Examples. Springer, New York, Heidelberg, Dordrecht, London (2012)
13. Schättler, H., Ledzewicz, U.: Optimal Control for Mathematical Models of Cancer Therapies. An Application of Geometric Methods. Springer, New York, Heidelberg, Dordrecht, London (2015)

The Use of Reaction Network Theory for Finding Network Motifs in Oscillatory Mechanisms and Kinetic Parameter Estimation

Igor Schreiber, Vuk Radojković, František Muzika, Radovan Jurašek and Lenka Schreiberová

Abstract Stoichiometric network analysis (SNA) is a method of studying stability of steady states of reaction systems obeying mass action kinetics. Reaction rates are expressed as a linear combination of elementary subnetworks with nonnegative coefficients (convex parameters) as opposed to standard formulation using rate coefficients and input parameters (kinetic parameters). We present examples of core reaction subnetworks that provide for oscillatory instability. Frequently there is an autocatalytic cycle in the core subnetwork, but in biochemical reactions such cycle is often replaced by a pathway called competitive autocatalysis. Rate coefficients in complex networks are often only partly known. We present a method of estimating the unknown rate coefficients, in which known/measured kinetic parameters and steady state concentrations are used to determine convex parameters, which in turn allows for determination of unknown rate coefficients by solving a set of constraint equations.

Keywords Stoichiometric networks · Oscillating (bio) chemical reactions · Dynamical instabilities · Parameter estimation

1 Introduction

Reaction networks corresponding to biochemical processes occurring in living organisms, such as genome-scale metabolic networks [13] are typically large. To understand various operating modes embedded within such systems, the networks at steady states are decomposed into elementary subnetworks (elementary fluxes, extreme currents) by taking advantage of a pseudolinear form of the corresponding model equations. In terms of linear algebra, these modes are represented by vectors including a collection of reaction rates at steady state. The elementary fluxes describe

I. Schreiber (✉) · V. Radojković · F. Muzika · R. Jurašek · L. Schreiberová
Department of Chemical Engineering, University of Chemistry and Technology, Prague, Technická 5, 166 28 Praha 6, Czech Republic
e-mail: igor.schreiber@vscht.cz

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_2

simplest chemical processes/functions that are available in the network with only a limited number of reaction rates turned on. Subsequently, they can be linearly combined using arbitrarily chosen non-negative coupling coefficients, producing the full network with contributions from all reactions. Such decomposition relies on stoichiometry only and does not require specification of reaction kinetics. However, leaving out kinetics precludes determination of stability of the steady state and one must assume it to be stable. In many systems, dynamical instabilities are essential for performing appropriate function, such as periodic oscillations or bistable switches. In those cases, the framework for analysis is provided by stoichiometric network analysis [3], which assumes that power law kinetics are provided and examines stability of the network's steady states.

The elementary subnetworks, or small subnetworks that are formed by combining suitable elementary subnetworks, can be tested for potential instability. Such a test does not require knowledge of rate coefficients and steady state concentrations of participating species. When an unstable subnetwork is coupled with other subnetworks, its (potential) instability will dominate the entire network provided that the coupling of the unstable subnetwork is strong enough.

This estimate of stability allows for identification of a core subnetwork that gives rise to an oscillatory instability and thereby provides a natural explanation for observing chemical oscillations [3–6, 16, 18]. Moreover, the oscillatory core subnetworks may be arranged into groups sharing certain topological features, which allows for a categorization of chemical oscillators [5, 16], each category being represented by a prototype network or a motif. At the same time, species within each prototype can be classified based on the role they play in generating the oscillations.

As mentioned above, such a classification is still based on analysis that does not involve knowledge of rate coefficients and steady state concentrations. When describing a specific experimental system, some of these parameters are known while others are not. Below we outline a procedure, that uses the notion of potential instability as introduced by the SNA and attempts to estimate the set of unknown rate coefficients and/or steady state cocentrations. To that goal, we utilize the idea that the coupling coefficients of the elementary subnetworks and steady state concentrations (convex parameters) must be consistent with known rate coefficients and inflow/initial constraints (kinetic parameters).

In our previous work [11, 15] we initiated the outlined approach and applied it to an oscillatory enzyme reaction and the classical inorganic Belousov-Zhabotinsky reaction. In this work we extend the list of subnetworks having a distinct motif and provide a mathematical framework for employing these motifs as reference subnetworks in models with extensive and complex mechanisms.

2 Theoretical Part

All spatially homogeneous isothermal chemical oscillators are based on stoichiometry and kinetics and fall within the formal mathematical description given below.

Let us assume a system involving m reactions and a total number of species n^{tot},

$$\nu_{1j}^{L} A_1 + \cdots + \nu_{n^{tot}j}^{L} A_{n^{tot}} \rightarrow \nu_{1j}^{R} A_1 + \cdots + \nu_{n^{tot}j}^{R} A_{n^{tot}}\,,\ j = 1, \ldots, m, \tag{1}$$

where A$_i$ are the reacting species and ν_{ij}^{L}, ν_{ij}^{R} are left and right stoichiometric coefficients. Any reversible reaction is treated as a pair of forward and backward steps. In a spatially homogeneous system, such as a flow-through reactor, dynamics of $n \leq n^{tot}$ species that are not inert products or in a pool condition are governed by a set of coupled mass balance equations which have the following pseudolinear form:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{N}\,\mathbf{v}(\mathbf{x})\,, \tag{2}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is the vector of concentrations of the interacting dynamical species, $\mathbf{N} = \{\Delta\nu_{ij}\} = \{\nu_{ij}^{R} - \nu_{ij}^{L}\}$ is the $(n \times m)$ stoichiometric matrix and $\mathbf{v} = (v_1, \ldots, v_m)$ is the non-negative vector of reaction rates (fluxes) (All vectors are assumed being column vectors). The reaction rates are assumed to follow mass action kinetics,

$$v_j = k_j \prod_{i=1}^{n} x_i^{\kappa_{ij}} = k_j \bar{v}_j\,, \tag{3}$$

where $\kappa_{ij} = \partial \ln v_j / \partial \ln x_i \geq 0$ is the reaction order of species i in reaction j and k_j is the corresponding rate coefficient, which may include fixed concentration(s) of pooled species and $\bar{v}_j$ is the reduced reaction rate. In vector notation we have $\mathbf{k} = (k_1, \ldots, k_m)$ and $\bar{\mathbf{v}}(x) = (\bar{v}_1, \ldots, \bar{v}_m)$. For elementary reactions, $\kappa_{ij} = \nu_{ij}^{L}$. However, in general case power law terms may also be used for quasi-elementary steps with $\kappa_{ij} \neq \nu_{ij}^{L}$. The kinetic matrix $\{\kappa_{ij}\}$ is denoted as $\mathbf{K}$. In flow systems, the inflows and outflows are included as pseudoreactions of zeroth and first order, respectively; the rate coefficient corresponding to an inflow term is $k_j = k_0 x_{i0}$ and that for an outflow is $k_j = k_0$, where k_0 is the flow rate and x_{i0} is the feed concentration of any inflowing species i.

At steady state Eq. (2) reduces to

$$\mathbf{N}\mathbf{v} = 0\,. \tag{4}$$

Since the reaction rates are non-negative, the set of all $\mathbf{v}_s$ satisfying the steady state condition is a non-negative subset of the null space of $\mathbf{N}$ represented by an $(m - d)$-dimensional convex polyhedral cone delimited by faces of dimension $1, \ldots, (m - d) - 1$, where d is the rank of $\mathbf{N}$. One-dimensional faces (or edges) represent a set of minimal, irreducible, connected subnetworks called elementary subnetworks or extreme currents or elementary fluxes. There are f edges of the cone satisfying $f \geq m - d$. The edges should be properly normalized, a convenient way is to let the components of the rate vector corresponding an edge sum up to 1. Endpoints of the normalized edges are apexes of a convex polytope of dimension $(m - d - 1)$. If $f = m - d$, the edges form a basis of the cone which is then called simplicial and the corresponding polytope is a simplex. Extreme currents can be obtained by algorithms of linear programming [9] or other efficient algorithms [17].

Let $\mathbf{E}_k$ denote a normalized rate vector corresponding to an elementary subnetwork. The set of all such subnetworks can be put into a matrix

$$\mathbf{E} = [\mathbf{E}_1, \ldots, \mathbf{E}_f]\,. \tag{5}$$

Any feasible rate vector $\mathbf{v}_s$ satisfying the steady state condition can be conveniently expressed as a non-negative linear combination of the elementary subnetworks,

$$\mathbf{v}_s = \mathbf{E}\,\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_f)\,. \tag{6}$$

The rate vector $\mathbf{v}_s$ is determined by choosing linear combination vector $\boldsymbol{\alpha}$. Upon substituting for the rate vector from (3) and by choosing $\mathbf{x}_s$ the rate coefficients are

$$\mathbf{k} = (\mathbf{diag}\,\bar{\mathbf{v}}(\mathbf{x}_s))^{-1}(\mathbf{E}\,\boldsymbol{\alpha})\,. \tag{7}$$

Thus, for a given set of convex parameters $(\boldsymbol{\alpha}, \mathbf{x}_s)$, kinetic parameters are obtained via (7). However, unless the cone is a simplex, the convex parameters are redundant, which must be taken into account when constructing the network from its elementary subnetworks.

Using the convex parameters, the mass balances given by Eq. (2) are expressed as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{N}\,\mathbf{diag}\,(\mathbf{E}\,\boldsymbol{\alpha})\,(\mathbf{diag}\,\bar{\mathbf{v}}(\mathbf{x}_s))^{-1}\,\bar{\mathbf{v}}(\mathbf{x})\,. \tag{8}$$

Upon linearizing the r.h.s. at the steady state $\mathbf{x}_s$, the Jacobian matrix is obtained,

$$\mathbf{J} = \mathbf{N}\,\mathbf{diag}\,(\mathbf{E}\boldsymbol{\alpha})\,\mathbf{K}^T\mathbf{diag}\,\mathbf{h} = -\mathbf{V}\,\mathbf{diag}\,\mathbf{h}\,, \tag{9}$$

where $\mathbf{h} = (1/x_1^s, \ldots, 1/x_n^s)$ includes reciprocal values of the steady state concentrations and $\mathbf{K}^T = \{\kappa_{ji}\}$ is the transpose of kinetic matrix. Because of convenience suggested by the form of Eq. (9), the convex parameters are usually taken as $(\boldsymbol{\alpha}, \mathbf{h})$ rather than $(\boldsymbol{\alpha}, \mathbf{x}_s)$.

An instability of a steady state $\mathbf{x}_s$ can be determined by analyzing principal minors of the $(n \times n)$ matrix $\mathbf{V}$. If a principal minor of order ℓ involving a subset of indexes $i_1, \ldots, i_\ell$ of certain species is negative, then at least one eigenvalue of $\mathbf{J}$ is unstable, provided that the steady state concentrations of corresponding species $x_{i_1}^s, \ldots, x_{i_\ell}^s$ are sufficiently small [3]. It is sufficient to consider a leading negative minor with a minimal order ℓ since any higher order instability is derived from the minimal configuration. Edge is the simplest possible unstable subnetwork. Next in the hierarchy of unstable subnetworks is 2-face such that both edges constituting the 2-face are stable when analyzed separately. Such an instability is possible due to nonlinearity of kinetics. We call such a face primary unstable. Any primary unstable k-face,

$k = 1, 2 \ldots$ may be a suitable candidate for the core of oscillatory subnetwork. When there are multiple choices, the dominant subnetwork is selected as described below.

3 Identification of Dominant Subnetworks

As mentioned above, the instability induced by a negative principal minor reflects the susceptibility of the subnetwork to possessing an unstable steady state provided that the corresponding steady state concentrations are sufficiently small. Although there are special cases when more subtle criteria have to be applied to indicate oscillatory instability, [4, 6], generally the outlined features provide excellent guidelines in evaluating the potential of a reaction network to undergo a dynamical instability. A Hopf bifurcation represents the emergence of oscillations [10], which is of primary importance in this work.

When applying the SNA to oscillatory mechanisms of inorganic reactions that were discovered since the pioneering work of Belousov and Zhabotinsky [19], is has been found [5] that dominant subnetworks forming the core oscillator have only a few topological arrangements of their networks, which are called prototypes or motifs. They all possess an autocatalytic cycle, i.e. a cycle connecting species (denoted as type X) of which at least one has a stoichiometric overproduction. In addition, there is a negative feedback loop involving a noncyclic species (denoted as type Z) and a removal of a type X species either by decomposition or via reaction with an inhibitory species (denoted as type Y).

However, many biochemical oscillators do not possess an autocatalytic cycle. Instead, their core oscillator possesses two type X-like species competing for a type Y-like species. In addition, there is a negative feedback loop involving type Z species, but all cycles present in the network are "ordinary" or nonautocatalytic cycles that do not provide for stoichiometric overproduction. Yet the network admits an instability leading to oscillations. Such a feature is called competitive autocatalysis. As with the cyclic autocatalysis, only a few basic motifs are expected to constitute dominant subnetworks of many biochemical networks.

As yet the categorization of networks with noncyclic autocatalysis is unavailable, here we provide examples of four prototypes, see Fig. 1. Cases (a) and (b) are prevalent in inorganic chemical oscillators [5] and possess autocatalytic cycle. They differ in coupling of the type Z species into the network. In the case (a) Z is supplied with feed and consumed by the cycle thereby creating negative feedback, whereas in the case (b) Z is produced by the cycle and induces negative feedback by producing type Y species, which inhibits the autocatalytic growth. Cases (c) and (d) represent examples of competitive autocatalysis. In the case (c) there are two type X species connected in a nonautocatalytic cycle, with X_1 being supplied externally, which allows for an autocatalytic-like accumulation if the type Y species is low. On the other hand, if Y is high, the autocatalysis is inhibited. The type Z species controls occurrence of the phase of accumulation (when Z is low) and depletion of X_1 (when Z is high). Availability of Z is inflow-controlled as in the case (a). Case (d) is an exam-

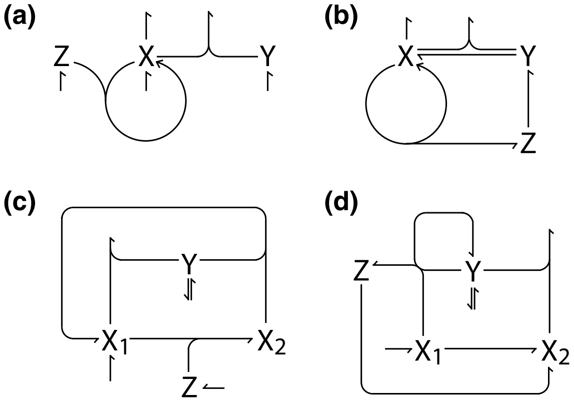


Fig. 1 Basic motifs of oscillatory reaction networks, **a** inflow-limited cyclic autocatalysis with external supply of Z, **b** inflow-unlimited (batch) cyclic autocatalysis with internal production of Z, **c** competitive autocatalysis with external supply of Z, **d** competitive autocatalysis with internal production of Z

ple od topological arrangement where there is no cycle connecting type X species (not even a nonautocatalytic cycle). Instead, type Y species is self-regenerating. The negative feedback may be arranged in several ways, either by external supply of Z as in the case (c) or by internal production of Z feeding back to X_2 as shown in the Figure or feeding back to X_1 (not shown). The internal production of Z is resembling the case (b), both X_1 and Y are either provided externally as shown in the Figure, or internally by embedding the motif into a larger network. The motif in case (a) represents numerous flow-limited inorganic oscillating reactions [5], case (b) represents a batch oscillator, in particular the Belousov-Zhabotinsky reaction [12]. Case c) is a generalization of an enzyme reaction with substrate inhibition, where X_1 and X_2 are two enzyme forms, Y is the inhibitory substrate and Z is another substrate that controls the oscillations. Case (d) represents a transcriptional network with X_1 and X_2 being the activator and inhibitor, respectively, Y is the protomer and Z represents the mRNA coding for the inhibitor [2]. Alternatively, case (d) with feedback to X_1 is found in phosphorylation cascades [8]. Essentially, all known inorganic oscillators possess either case (a) or case (b) motif [5]. Additional motifs with autocatalytic loop possess higher order autocatalysis as found in the glycolytic oscillator [5]. Although additional motifs involving competitive autocatalysis are quite likely, they are yet to be discovered.

4 Determination of Parameters Based on Stoichiometric Constraints

Our major goal is to use the SNA approach to networks to estimate unknown parameters in a system, for which we have or assume a detailed mechanism implying power law kinetics, and are able to perform experiments leading the emergence of oscillations via Hopf bifurcation. Typically, some of the rate coefficients are known from previous research, thus our aim is to determine a subset of rate coefficients. Parameter determination can be based on choosing an appropriate subset of Eq. (6) with $\mathbf{v}_s = \mathbf{v}(\mathbf{x}_s)$ consistent with our experiments and already known parameters. These equations can be used for finding various unknown quantities including rate coefficients and steady state concentrations, given that other quantities are available, such as rate coefficients known from independent experiments or taken from literature, experimentally measured steady state concentrations, known inflow rate and inflow concentrations at a distinct dynamical instability. The major instabilities are: (i) emergence of oscillations at a Hopf bifurcation or (ii) switch to another steady state at a saddle-node bifurcation. To preserve linearity, we choose a subset of rate equations such that, upon substituting the known quantities, the rate expressions are either linear in a particular unknown or fully determined. We call such equations constraint equations.

In order to have a compact form of the constraint equations, let us arrange the order of elementary subnetworks in $\mathbf{E}$, the order of species in $\mathbf{x}_s$ and the order of reactions in $\mathbf{v}_s$ and $\mathbf{k}$ as follows (below we drop the subscript s so that $\mathbf{x}$ and $\mathbf{v}$ now denote steady state quantities):

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{fv}, \boldsymbol{\alpha}^{uv}), \mathbf{x} = (\mathbf{x}^{fv}, \mathbf{x}^{uv}, \mathbf{x}^{iv}), \mathbf{k} = (\mathbf{k}^{fv}, \mathbf{k}^{uv}, \mathbf{k}^{iv}), \mathbf{v} = (\mathbf{v}^{fv}, \mathbf{v}^{uv}, \mathbf{v}^{iv}). \tag{10}$$

Each vector consists of subvectors, where the superscript fv means a fixed or given value of the relevant quantity, uv means an unknown value to be determined from the constraint equations and iv means an implied value determined from those equations in Eq. (6) that are not used as the constraint equations and can be obtained only after the unknown values are determined. Moreover, the reaction rates possessing some unknowns fall within two distinct subsets, $\mathbf{v}^{uv} = (\mathbf{v}^{uvx}, \mathbf{v}^{uvk})$; in $\mathbf{v}^{uvx}$ only the concentrations $\mathbf{x}^{uv}$ occur as (linear) unknowns, while in $\mathbf{v}^{uvk}$ the rate coefficients $\mathbf{k}^{uv}$ are the only unknowns. To maintain linearity of the constraint equations, concentration of a species i is an admissible unknown if i is of first order in any reaction j included in the constraint equations. By contrast, any α_k or k_j satisfy linearity by definition.

Choice of $\boldsymbol{\alpha}^{fv}$. The fixed values of α_k's can be split into two parts, $\boldsymbol{\alpha}^{fv} = (\boldsymbol{\alpha}^{uds}, \boldsymbol{\alpha}^{nds})$. The choice of $\boldsymbol{\alpha}^{uds}$ is based on the preceding stability analysis (Sect. 2) that identifies an unstable subnetwork having a dominant contribution in Eq. (6). The other subvector $\boldsymbol{\alpha}^{nds}$ specifies contributions from nondominant subnetworks that are not involved in the constraint equations due a limited set of known data.

Choice of $\mathbf{x}^{fv}$. Certain x_i's in $\mathbf{x}^{fv}$ are known from experimental measurements (such as by measuring pH). Others may be set according to requirements of the stability analysis as outlined above. In particular, the leading principal minor of order ℓ for the unstable dominant subnetwork indicates species with indexes $i_1, \ldots, i_\ell$ that must have small values relative to others if the instability is to be manifested.

Choice of $\mathbf{k}^{fv}$. Some of the rate coefficients can be often found in literature, presumably evaluated by standard experimental procedures while others can be assumed based on a number of clues, such as the upper bound for diffusion-limited rate processes, forward/backward rate coefficients related via equilibrium constants, etc.

Choice of unknown and implied variables. All values in $\boldsymbol{\alpha}$ that are not fixed are taken as unknowns. The choice is less straightforward for steady state concentrations. Those x_i's that are not fixed and appear with first order in the reaction rates where all other quantities are fixed, may be selected as unknowns. The remaining concentrations correspond either to species with order other than one in some rate expressions or the rates where they appear include other unknown quantities and thus fall within $\mathbf{x}^{iv}$. The set of unknown rate coefficients has two parts, $\mathbf{k}^{uv} = (\mathbf{k}^{irr}, \mathbf{k}^{fwr})$, where $\mathbf{k}^{irr}$ includes unknown rate coefficients of irreversible steps, which requires the corresponding subset of reduced reaction rates $\bar{\mathbf{v}}^{irr}$ to be given by fixed values of the relevant concentrations. In the case of reversible reactions, there are pairs of forward/backward reaction steps and $\mathbf{k}^{fwr}$ denotes the set of forward rate coefficients. The reduced rates in both directions, $\bar{\mathbf{v}}^{fwr}$, $\bar{\mathbf{v}}^{bwr}$, are again assumed to be determined by the fixed values of relevant concentrations. Provided that the values of the equilibrium constants are known, the corresponding backward rate coefficients are expressed in terms of the forward rate coefficients, $\mathbf{k}^{bwr} = (\mathbf{diag}\,\mathbf{K}_{eq})^{-1}\mathbf{k}^{fwr}$ where $\mathbf{K}_{eq}$ is the corresponding vector of equilibrium constants.

4.1 Formulation of the Constraint Equations

After identifying fixed, unknown and implied quantities we can finally set up the constraint equations by selecting certain equations from Eq. (6) and rearranging them to obtain linear equations in a standard matrix form. First, we express $\mathbf{E}$ in terms of three blocks containing edge(s) involved in the unstable dominant subnetwork, non-dominant subnetworks not to be used in the constraint equations and the remaining edges that will be part of the constraints, respectively:

$$\mathbf{E} = [\mathbf{E}^{uds}, \mathbf{E}^{nds}, \mathbf{E}^{uv}]. \tag{11}$$

Equation (6) can be then rewritten as

$$\mathbf{E}^{uds}\boldsymbol{\alpha}^{uds} + \mathbf{E}^{nds}\boldsymbol{\alpha}^{nds} + \mathbf{E}^{uv}\boldsymbol{\alpha}^{uv} = \mathbf{v}\,. \tag{12}$$

Depending on the available input data, we choose constraint equations based on those rates v_j which are either known entirely or expressed as a linear function of either k_j^{uv} or x_i^{uv}. As a result, the remaining equations are all included in the term $\mathbf{E}^{nds}\boldsymbol{\alpha}^{nds}$ representing edges that have no contribution to the selected v_j's and are removed. The constraint equations then read

$$\hat{\mathbf{E}}^{uv}\boldsymbol{\alpha}^{uv} = \hat{\mathbf{v}} - \hat{\mathbf{E}}^{uds}\boldsymbol{\alpha}^{uds}\,, \tag{13}$$

where $\hat{\mathbf{v}} = (\mathbf{v}^{fv}, \mathbf{v}^{uvk}, \mathbf{v}^{uvx})$ is the set of rates used as constraints and $\hat{\mathbf{E}}^{uv}$, $\hat{\mathbf{E}}^{uds}$ retain only rows corresponding to $\hat{\mathbf{v}}$. Since $\mathbf{v}^{uvk}$ depends linearly on $\mathbf{k}^{uv}$ and $\mathbf{v}^{uvx}$ on $\mathbf{x}^{uv}$, these terms can be moved to the l.h.s. of Eq. (13) to yield a system of linear constraint equations in a simple block form,

$$\begin{bmatrix} & 0 & 0 \\ \hat{\mathbf{E}}^{uv} & \mathbf{A} & 0 \\ & 0 & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}^{uv} \\ \mathbf{k}^{uv} \\ \mathbf{x}^{uv} \end{bmatrix} = \begin{bmatrix} \mathbf{v}^{fv} \\ 0 \\ 0 \end{bmatrix} - \hat{\mathbf{E}}^{uds} \boldsymbol{\alpha}^{uds} \,. \tag{14}$$

The matrix **A** possesses a banded structure,

$$\mathbf{A} = \begin{bmatrix} -\mathbf{diag}\,\bar{\mathbf{v}}^{irr} & 0 \\ 0 & -\mathbf{diag}\,\bar{\mathbf{v}}^{fwr} \\ 0 & -\mathbf{diag}\,\bar{\mathbf{v}}^{bwr} \end{bmatrix}, \tag{15}$$

whereas the structure of **B** is given by how many constrained reactions involve a given species as a linear unknown and does not possess any specific structure of nonzero elements, except that each row contains only one nonzero value which is negative.

4.2 Solving the Constraint Equations

The constraint equations Eq. (14), may be concisely written as

$$\mathbf{U}\,\mathbf{a} = \mathbf{b}\,. \tag{16}$$

As a rule, the system is underdetermined due to a large number of elementary subnetworks and limited data on reaction rates known from experiments or literature that can be used in formulating constraint equations. Consequently, Eq. (16) is not expected to provide a unique solution and a suitable solution needs to be selected by utilizing an optimization procedure with an appropriate objective function. A clue to defining an objective function is readily provided by employing the stability analysis of stoichiometric networks as outlined in Sect. 2. More specifically, for chemical oscillators the emergence of oscillations via Hopf bifurcation is implied by dominance of the chosen (leading) unstable subnetwork. Therefore we can postulate that the contributions of the elementary subnetworks other than the leading unstable subnetwork should be as small as possible at the oscillatory instability. Thus the objective function to be minimized may be taken as the sum of the contributions of all subnetworks involved in the constraint equations other than the unstable dominant one, whose contribution is used as a free bifurcation parameter, which is varied until a Hopf bifurcation is found. Since the constraint equations are constructed to be linear, a linear programming solver [14] was used for solving the constrained system Eq. (16) by minimizing

$$f(\mathbf{a}) = \sum_{k=1}^{p} \alpha_k^{uv}. \tag{17}$$

In general, the set of all admissible solutions of Eq. (16) with non-negative components of **a** is restricted to a set which may be a convex bounded polytope or a convex unbounded polytope, which arises by shifting the non-negative cone (if it exists) of the homogeneous subsystem of Eq. (16) in the space of **a** due to **b** and has a set of apexes in some directions but extends without bounds in other directions. The minimal solution sits in one of the apexes.

5 Discussion and Conclusions

The approach outlined above has been applied to the glucose oxidase–catalase reaction [11] and the Belousov-Zhabotinsky reaction [15]. However, main applications are expected in identifying kinetic parameters in models of biological oscillating systems, such as circadian clocks [7]. Also, when temperature dependence of the rate coefficients is of interest, the input experimental information at two (or more) different temperatures needs to be provided and results subsequently fitted to Arrhenius law.

There are certain caveats that must be taken care of to obtain the solution of Eq. (16). Some of the parameters $\mathbf{x}^{fv}$ and $\mathbf{k}^{fv}$ that are not available from measurements must be assigned fixed values chosen heuristically. It may happen that such a choice violates solvability of the system Eq. (16). In this case, an effective way of resolving the problem is to find incompatible constraint equations and remove them. Likewise, some constraint equations may be linear combinations of others, which causes the linear programming solver to fail. Both incompatible and linearly dependent equations can be removed by applying singular value decomposition [14]. Another limitation is the linearity of constraint equations. In future work a nonlinear constrained optimization [1] should be considered.

Acknowledgements This work has been supported by the grant 18-24397S from the Czech Science Foundation.

References

1. Albers, C.J., Critchley, F., Gower, J.C.: Quadratic minimisation problems in statistics. J. Multivar. Anal. **102**(3), 698–713 (2011)
2. Cerveny, J., Salagovic, J., Muzika, F., Safranek, D., Schreiber, I.: Influence of circadian clocks on optimal regime of central C-N metabolism of cyanobacteria. In: Mishra, A.K., Tiwari, D.N., Rai, A.N. (eds.) Cyanobacteria: From Basic Science to Applications, Chap. 9, pp. 193–206. Academic Press, London (2019)
3. Clarke, B.L.: Stability of complex reaction networks. Adv. Chem. Phys. **43**, 1–278 (1980)
4. Eiswirth, M., Bürger, J., Strasser, P., Ertl, G.: Oscillating Langmuir-Hinshelwood mechanisms. J. Phys. Chem. **100**(49), 19118–19123 (1996)
5. Eiswirth, M., Freund, A., Ross, J.: Mechanistic classification of chemical oscillators and the role of species. Adv. Chem. Phys. **80**, 127–199 (1991)
6. Errami, H., Eiswirth, M., Grigoriev, D., Seiler, W.M., Sturm, T., Weber, A.: Detection of Hopf bifurcations in chemical reaction networks using convex coordinates. J. Comput. Phys. **291**, 279–302 (2015)

7. Gonze, D.: Modeling circadian clocks: from equations to oscillations. Cent. Eur. J. Biol. **6**(5), 699–711 (2011)
8. Hadac, O., Muzika, F., Nevoral, V., Pribyl, M., Schreiber, I.: Minimal oscillating subnetwork in the Huang-Ferrell model of the MAPK cascade. Plos One **12**(6) (2017)
9. Hadley, G.: Linear Programming. Addison-Wesley Publishing Company (1962)
10. Marsden, J.E., McCracken, M.: The Hopf Bifurcation and Its Applications. Springer, New York (1976)
11. Muzika, F., Jurasek, R., Schreiberova, L., Radojkovic, V., Schreiber, I.: Identifying the oscillatory mechanism of the glucose oxidase-catalase coupled enzyme system. J. Phys. Chem. A **121**(40), 7518–7523 (2017)
12. Noyes, R.M., Field, R.J., Koros, E.: Oscillations in chemical systems. 1. Detailed mechanism in a system showing temporal oscillations. J. Am. Chem. Soc. **94**(4), 1394–1395 (1972)
13. Palsson, B.: Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, Cambridge, New York (2006)
14. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in Fortran. Cambridge University Press, Cambridge (1986, 1992)
15. Radojkovic, V., Schreiber, I.: Constrained stoichiometric network analysis. PCCP **20**, 9910–9921 (2018)
16. Ross, J., Schreiber, I., Vlad, M.O.: Determination of Complex Reaction Mechanisms. Oxford University Press Inc., New York (2006)
17. Schilling, C.H., Letscher, D., Palsson, B.Ø.: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. **203**(3), 229–248 (2000)
18. Schreiber, I., Ross, J.: Mechanisms of oscillatory reactions deduced from bifurcation diagrams. J. Phys. Chem. A **107**(46), 9846–9859 (2003)
19. Zhabotinskii, A.M.: Periodic course of the oxidation of malonic acid in a solution (studies on the kinetics of Belousov's reaction). Biofizika **9**, 306–11 (1964)

An Existence Result for Some Fractional Evolution Equation with Nonlocal Conditions and Compact Resolvent Operator

G. M. N'Guérékata

Abstract We are concerned with the existence of mild solutions to the fractional differential equation $D_t^\alpha u(t) = Au(t) + J_t^{1-\alpha} f(t, u(t)),\ \ 0 < t \leq T$, with nonlocal conditions $u_0 = u(0) + g(u)$, where $0 < \alpha < 1$, D_t^α is the Caputo derivative, $J_t^{1-\alpha} h(t) := \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{h(s)}{(t-s)^\alpha} ds$ is the fractional integral of order α of the function h, and $A : D(A) \subset X \to X$ is a linear operator which generates a compact analytic resolvent family $(R_\alpha(t))_{t \geq 0}$, X being a Banach space. We obtain our results using the Krasnoleskii's fixed point theorem.

Keywords S-asymptotically ω-periodic functions · Bloc-periodic functionc · Fractional differential equations

1 Introduction

Fractional differential equations are one of the most useful mathematical tools in both pure and applied analysis. We can find numerous applications of fractional differential and integral equations of fractional order in viscoelasticity, electrochemistry, control, porous media, electromagnetism, etc., There has been a significant development in ordinary and partial fractional integral equations in recent years; see the monographs of Abbas et al. [1, 2], Miller [10], Podlubny [13] and the references therein.

The problem of finding solutions of phenomena modeled by fractional differential equations has been extensively studied over the last decades (cf. instance [3, 4, 7, 11, 12, 14]). We are concerned with the following fractional differential equation with nonlocal initial conditions.

$$\begin{cases} D_t^\alpha u(t) = Au(t) + J_t^{1-\alpha} f(t, u(t)), \quad 0 < t \leq T \\ \\ u_0 = u(0) + g(u) \end{cases} \tag{1}$$

G. M. N'Guérékata (✉)
Department of Mathematics, Morgan State University, Baltimore, MD 21251, USA
e-mail: Gaston.N'Guerekata@morgan.edu

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_3

Our aim is to study the existence of mild solutions under appropriate conditions on the functions f and g assuming that the (generally unbounded) linear operator A generates an $(\alpha, 1)$ resolvent family. The following articles motivated our work.
In [12], we studied the existence (and uniqueness) of mild solution to the following problem in a finite dimensional space X

$$\begin{cases} D_t^\alpha u(t) = f(t, u(t)), \quad 0 < t \leq T \\ \\ u_0 = u(0) + g(u) \end{cases} \tag{2}$$

using the Banach's fixed point theorem and the Krasnoleskii's fixed point theorem. This paper has inspired several works in infinite dimensional spaces including the paper [5]. In [8], the authors studied the existence of mild solution to the problem

$$\begin{cases} D_t^\alpha u(t) = Au(t) + J_t^{1-\alpha} f(t, u(t)), \quad 0 < t \leq T \\ \\ u_0 = u(0) + g(u) \end{cases} \tag{3}$$

where D_t^α is the Caputo derivative and the operator A generates an $(\alpha, 1)$-resolvent family. They used the Leray-Schauder theorem to achieve their results.
In the present work, we consider the same problem with different (say weaker) assumptions on the functions f and g and use the Krasnoselskii theorem to obtain our result.

2 Preliminaries

Throughout this paper, X will be a Banach space, $\mathcal{C}$ will be the space of all continuous functions $[0, T] \to X$ equipped with the supnorm, $B(X)$ the Banach space of all bounded linear operators $X \to X$. Let's recall some definitions and facts (cf. for instance [8]).

Definition 2.1 Let $\alpha > 0$. A closed linear operator $A : D(A) \subset X \to X$ is said to be the generator of an $(\alpha, 1)$-resolvent family if there exists $\omega \geq 0$ and a strongly continuous function $R_\alpha : \mathbb{R}^+ \to B(X)$ such that $\{\lambda^\alpha : Re\lambda > \omega\} \subset \rho(A)$ and

$$\lambda^{\alpha-1}(\lambda^\alpha - A)^{-1}x = \int_0^\infty e^{-\lambda t} R_\alpha(t)x dt, \quad Re\lambda > \omega, \ x \in X.$$

In this case, the family $(R_\alpha(t))_{t\geq 0}$ is called an $(\alpha, 1)$-resolvent family generated by A.

Remark 2.2 If A is the generator of an $(\alpha, 1)$-resolvent family $(R_\alpha(t))_{t\geq 0}$, then we have the following properties:

(i) $R_\alpha(t)$ is strongly continuous for $t \geq 0$ and $R_\alpha(0) = I$, the identity operator on X
(ii) $R_\alpha(t)A \subset AR_\alpha(t)$ for $t \geq 0$
(iii) For $x \in D(A)$, the resolvent equation

$$R_\alpha(t)x = x + \int_0^t g_\alpha(t-s)R_\alpha(t)Axds$$

holds for all $t \geq 0$.

Definition 2.3 We say that a strongly continuous family $(T(t))_{t\geq 0}$ is exponentially bounded, or of type (M, ω) if there exists $M > 0$ and $\omega \in \mathbb{R}$ such that

$$\|T(t)\| \leq Me^{\omega t}, \quad \forall t \geq 0.$$

Theorem 2.4 ([8]) *Let* $0 < \alpha \leq 1$ *and* $(R_\alpha(t))_{t\geq 0}$ *be an* $(\alpha, 1)$*-resolvent family of type* (M, ω) *generated by the operator A. Suppose that* $R_\alpha(t)$ *is continuous in the uniform operator topology for all t* > 0*. Then the following assertions are equivalent*
(i) $R_\alpha(t))$ *is a compact operator for all t* > 0
(ii) $(\mu - A)^{-1}$ *is a compact operator for all* $\mu > \omega^{\frac{1}{\alpha}}$.

Theorem 2.5 (Ascoli-Arzela) *Let* $\Gamma := (\gamma(t))$ *be a family of continuous maps* $I \to X$ *where I is a bounded interval of* $\mathbb{R}$ *and X a Banach space. If* Γ *is uniformly bounded and equicontinuous and for any* $\bar{t} \in I$*, the set* $(\gamma_n(\bar{t})), n = 1, 2, \ldots$ *is relatively compact, then there exists a uniformly convergent function sequence* $(\gamma_n(t)), n = 1, 2, \ldots, t \in I$ *in* Γ.

Lemma 2.6 (Mazur) *If K is a compact subset of X, then its convex closure* $\overline{co}(K)$ *is compact.*

Theorem 2.7 (Krasnoselskii) *Let M be a convex closed nonempty subset of a Banach space X. Suppose that A, B are operators mapping M into X such that*
(i) $Au + Bv \in M$ *whenever* $u, v \in M$*;*
(ii) A is a contraction mapping;
(iii) B is a continuous compact mapping.
Then there exists $z \in M$ *such that* $z = Az + Bz$

3 Main Results

Let $J_t^{1-\alpha}h(t) := \frac{1}{\Gamma(1-\alpha)}\int_0^t \frac{h(s)}{(t-s)^\alpha}ds$ be the fractional integral of order $\alpha > 0$ of the function h with values in a Banach space X. The integrals here and in this paper are considered in the sense of Bochner that is those measurable functions $f : [0, T] \to X$ such that $\|f\|$ is Lebesgue integrable.
For a function $f \in C^m(\mathbb{R}^+, X)$, where $m \in \mathbb{N}$, its Caputo fractional derivative of order $\alpha \in (m-1, m)$ is defined by

$$D_t^\alpha f(t) := \frac{1}{\Gamma(m-\alpha)} \int_0^t (t-s)^{m-\alpha-1} f^{(m)}(s)ds.$$

It is easy to prove the following for $\alpha > 0$:

(i) The Caputo derivative of a constant is zero.
(ii) If $D_t^\alpha h(t) = 0$, then

$$h(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^{n-1}$$

where c_i are reals and $n = [\alpha] + 1$. For more details on fractional calculus, we refer for instance to [2, 10, 13].

Now we consider the following problem

$$\begin{cases} D_t^\alpha u(t) = Au(t) + J_t^{1-\alpha} f(t, u(t)), & 0 < t \le T \\ u_0 = u(0) + g(u) \end{cases} \tag{4}$$

where $0 < \alpha < 1$, D_t^α is the Caputo derivative. The linear operator $A : D(A) \subset X \to X$ generates an $(\alpha, 1)$-resolvent family $R_\alpha(t)$ of type (M, ω). $f : [0, T] \times X \to X$ and $g : \mathcal{C} := C([0, T], X) \to X$ with some assumptions to be defined below. As indicated in his pioneering paper [4], Deng proves that the nonlocal condition $u(0) = u_0 - g(u)$ can be applied in physics with better effect than the classical Cauchy Problem initial condition $u(0) = u_0$. We refer also the reader to [9] for more information about this concept.

Definition 3.1 [6] A function $u \in \mathcal{C} := C([0, T]; X)$ is said to be a mild solution to Eq. (4) if it satisfies the integral equation

$$u(t) = R_\alpha(t)[u_0 - g(u)] + \int_0^t R_\alpha(t-s) f(s, u(s))ds, \quad t \in [0, T].$$

We make the following assumptions

A1. The function f is of Carathéodory, i.e. $f(\cdot, u)$ is strongly measurable for each $u \in X$ and $f(t, \cdot)$ is continuous for each $t \in X$.

A2. There exist constant $c_f > 0$, $c_g > 0$ such that

$$\|f(t, u)\| \le c_f(1 + \|u\|_\mathcal{C}), \forall (t, u) \in [0, T] \times X,$$
$$\|g(u)\| \le c_g(1 + \|u\|_\mathcal{C}), \quad \forall u \in \mathcal{C}.$$

A3. There exists a constant $L_g > 0$ such that

$$\|g(u) - g(v)\| \le L_g \|u - v\|_\mathcal{C}, \quad \forall u, v \in \mathcal{C}.$$

Theorem 3.2 *Assume that the operator A generates an $(\alpha, 1)$-resolvent family $(R_\alpha(y))_{t\geq 0}$ of type (M, ω) where $\omega < 0$ and $(\lambda^\alpha - A)^{-1}$ is compact and $R_\alpha(t)$ is continuous in the usual operator topology for all $t > 0$. Under the above assumptions $A1 - A3$, Eq. (4) possesses a mild solution provided $\max\{L_g, (1 + c_g + \frac{c_f}{\omega})\} < \frac{1}{M}$.*

Proof Let r such that $r > \frac{M(c_f+c_g)}{1-M(1+c_g+\frac{c_f}{|\omega|})}$ and consider the closed convex subset $\Gamma_r := \{u \in \mathcal{C} : \|u\|_{\mathcal{C}} \leq r\}$.
Define the operators $P, Q : \Gamma_r \to \Gamma_r$ by

$$(Pu)(t) := R_\alpha(t)[u_0 - g(u)]$$

and

$$(Qu)(t) : \int_0^t R_\alpha(t-s) f(s, u(s)) ds.$$

We will use several steps.

Step 1.
Let $u, v \in \Gamma_r$. We will show that $Pu + Qv \in \Gamma_r$. Indeed we have

$$\begin{aligned}
\|Pu + Qu\| &\leq M[e^{\omega t}(\|u_0\| + \|g(u)\| + \int_0^t e^{\omega(t-s)} c_f (1 + \|v(s)\|) ds] \\
&\leq M[e^{\omega t}(\|u_o\| + c_g(1 + \|u\|_{\mathcal{C}}) + c_f(1 + \|v\|_{\mathcal{C}}) \int_0^t e^{\omega(t-s)} ds] \\
&\leq M[e^{\omega t}(\|u_o\| + c_g(1 + \|u\|_{\mathcal{C}}) + \frac{c_f}{\omega}(1 + \|v\|_{\mathcal{C}})] \\
&\leq M[\|u_0\| + c_g(1 + r) + \frac{c_f}{\omega}(1 + r)] \\
&\leq r.
\end{aligned}$$

This completes the proofs of Step 1.

Step 2. Let $u, v \in \mathcal{C}$. Then we have

$$\begin{aligned}
\|(Pu)(t) - (Qu)(t)\| &= \|R_\alpha(t)[u_0 - g(u)] - R_\alpha(t)[u_0 - g(v)]\| \\
&\leq \|R_\alpha(t)[g(u) - g(v)]\| \\
&\leq Me^{\omega t}\|g(u) - g(v)\| \\
&\leq ML_g e^{\omega t}\|g(u) - g(v)\|_{\mathcal{C}} \\
&\leq ML_g\|g(u) - g(v)\|_{\mathcal{C}},
\end{aligned}$$

which shows that P is a contraction since $ML_g < 1$.

Step 3. Q is a continuous operator. Indeed let $(u)_n \subset B_r$ such that $u_n \to u$ in B_r. Then for each $s \in [0, T]$, $f(s, u_n(s)) \to f(s, u(s))$ in view of A1. Now fix $t \in [0, T]$. We have

$$\begin{aligned}
(Qu_n)(t) - (Qu)(t)\| &= \int_0^t R_\alpha(t-s)[f(s,u_n(s)) - f(s,u(s))]ds\| \\
&\leq M \int_0^t e^{\omega(t-s)} \|f(s,u_n(s)) - f(s,u(s))\| \\
&\leq M \int_0^t e^{\omega(t-s)} (\|f(s,u_n(s))\| + \|f(s,u(s))\|)ds \\
&\leq Mc_f \int_0^t e^{\omega(t-s)} (2 + \|u_n(s)\| + \|u(s)\|)ds \\
&\leq 2Mc_f(1+r) \int_0^t e^{\omega(t-s)} ds \\
&< \infty.
\end{aligned}$$

Therefore, by the Lebesgue Dominated Convergence Theorem, $Qu_n \to Qu$.
Step 4. Q is a compact operator. First, consider an arbitrary sequence in B_r. Then we have

$$\begin{aligned}
\|(Qu_n)(t)\| &\leq M \int_0^t e^{\omega(t-s)} \|f(s,u_n(s))\|ds \\
&\leq Mc_f \int_0^t e^{\omega(t-s)} (1 + \|u_n(s)\|)ds \\
&\leq Mc_f(1+r) \int_0^t e^{\omega(t-s)} ds \\
&\leq Mc_f \frac{(1+r)}{|\omega|},
\end{aligned}$$

for every $n = 1, 2, \ldots$. This shows that (Qu_n) is uniformly bounded.

Now we will show that (Qu_n) is equicontinuous. Let's take t_1, t_2 such that $0 \leq t_1 < t_2 \leq T$. Then we get

$$\begin{aligned}
\|(Qu_n)(t_1) - (Qu_n)(t_2)\| &= \| \int_0^{t_1} R_\alpha(t_1 - s) f(s, un(s))ds - \int_0^{t_2} R_\alpha(t_2 - s) f(s, u_n(s))ds\| \\
&= \| \int_0^{t_1} R_\alpha(t_1 - s) f(s, u_n(s))ds \\
&- \int_0^{t_1} R_\alpha(t_2 - s) f(s, u_n(s))ds \\
&- \int_{t_1}^{t_2} R_\alpha(t_2 - s) f(s, u_n(s))ds\| \\
&\leq I_1 + I_2
\end{aligned}$$

where

$$I_1 = \| \int_0^{t_1} [R_\alpha(t_1 - s) - R_\alpha(t_2 - s)] f(s, u_n(s))ds\|, \quad I_2 = \| \int_{t_1}^{t_2} R_\alpha(t_2 - s) f(s, u_n(s))ds\|.$$

Now we note that

$$\begin{aligned} I_2 &\le M \int_{t_1}^{t_2} e^{\omega(t_2-s)} \|f(s,u_n(s))\| ds \\ &\le Mc_f \int_{t_1}^{t_2} e^{\omega(t_2-s)}(1+\|u_n(s)\|) ds \\ &\le Mc_f(1+r) \int_{t_1}^{t_2} e^{\omega(t_2-s)} ds \\ &\le Mc_f(1+r)(t_2-t_1), \end{aligned}$$

which shows that $\lim_{t_1\to t_2} I_2 = 0$. Also we have

$$\begin{aligned} I_1 &\le \| \int_0^{t_1} \|R_\alpha(t_1-s) - R_\alpha(t_2-s)\| \|f(s,u_n(s))\| ds \\ &\le c_f \| \int_0^{t_1} \|R_\alpha(t_1-s) - R_\alpha(t_2-s)\|(1+\|u_n(s)\|) ds \\ &\le c_f \| \int_0^{t_1} \|R_\alpha(t_1-s) - R_\alpha(t_2-s)\|(1+\|u_n(s)\|) ds \\ &\le c_f(1+r) \| \int_0^{t_1} \|R_\alpha(t_1-s) - R_\alpha(t_2-s)\| ds. \end{aligned}$$

Let's note that

$$\|R_\alpha(t_1-\cdot) - R_\alpha(t_2-\cdot)\| \le 2Me^{\omega(t-\cdot)}$$

and that $2Me^{\omega(t-\cdot)} \in L^1([0,T],\mathbb{R})$. Also $\lim_{t_1\to t_2}[R_\alpha(t_1-s)R_\alpha(t_2-s)] = 0$ in $BC(X)$.

Thus by the Lebesgue Dominated Convergence Theorem, $\lim_{t_1\to t_2} I_1 = 0$, which proves the equicontinuity of (Qu_n).

Step 5. In view of Theorem 2.4, $R_\alpha(t)$ is compact for all $t > 0$. Therefor the set

$$K := \{R_\alpha(t-s)f(s,u(s)) : u \in \mathcal{C}, s \in [0,T]\}$$

is relatively compact for each $t \in [0,T]$. It follows that $\overline{co}(K)$ is compact, using the Mazur Theorem. Now let $u \in B_r$. It is clear that for all $t \in [0,T]$ we have

$$(Qu)(t) \in \overline{co}(K).$$

Consequently, the set $\{(Qu)(t) : u \in B_r\}$ is relatively compact in X. Finally we conclude that Q is compact by the Arzela-Ascoli's theorem. □

Acknowledgements We would like to thank the referee for his/her careful reading and valuable suggestions.

References

1. Abbas, S., Benchohra, M., N’Guérékata, G.M.: Topics in Fractional Differential Equations. Springer, New York (2012)
2. Abbas, S., Benchohra, M., N’Guérékata, G.M.: Advanced Fractional Differential Equations. Nova Science Publishers, New York (2015)
3. Araya, D., Lizama, C.: Almost automorphic mild solutions to fractional differential equations. Nonlinear Anal. **69**, 3692–3705 (2008)
4. Deng, K.: Exponential decay of solutions of semilinear parabolic equations with nonlocal initial conditions. J. Math. Anal. Appl. **179**, 630–637 (1993)
5. Dong, X., Wang, J., Zhou, Y.: On local problems for fractional differential equations in Banach spaces. Opusc. Math. **31**(3) (2011). https://doi.org/10.7494/OpMath.2011.31.341
6. Fan, Z.: Characterization of compactness for resolvents and its applications. Appl. Math. Comput. **232**, 60–67 (2014)
7. Lizama, C., N’Guérékata, G.M.: Mild solutions for abstract fractional differential equations. Appl. Anal. **92**(8), 1731–1754 (2013)
8. Lizama, C., Pereira, A., Ponce, R.: On the compactness of fractional resolvent operator functions. Semigroup Forum (2016). https://doi.org/10.1007/s00233-016-9788-7
9. Lizama, C., Pozo, J.C.: Existence of mild solutions for semilinear integrodifferential equations with nonlocal conditions. Abstr. Appl. Anal. **2012**(2012) (Art. ID 647103), 15 (2012). https://doi.org/10.1155/2012/647103
10. Miller, K., Ross, B.: An Introduction to the Fractional Calulus and Fractional Differential Equations. Wiley, New York (1993)
11. Mophou, G., N’Guérékata, G.M.: Existence of mild solutions of some semilinear neutral fractional functional evolution equations with infinite delay. Appl. Math. Comput. **216**, 61–69 (2010)
12. N’Guérékata, G.M.: A Cauchy problem for some abstract fractional differential equations with nonlocal conditions. Nonlinear Anal. **70**, 11–14 (2009)
13. Podlubny, I.: Fractional Differential Equations. Academic Press, San Diego (1999)
14. Zhou, Y., Jiao, F.: Nonlocal Cauchy problem for fractional evolution equations. Nonlinear Anal. **11**, 4465–4475 (2010)

Impact of Tobacco Smoking on the Prevalence of Tuberculosis Infection: A Mathematical Study

Mini Ghosh

Abstract Tobacco smoking is a social problem. It is a well-established fact that the smoking of tobacco products can cause severe health problems. It is also observed that the smokers are having higher risk of getting tuberculosis infection than the non-smokers. In this article a non-linear mathematical model is proposed and analyzed to demonstrate the impact of tobacco smoking on transmission dynamics of the tuberculosis. The basic reproduction number R_0 of the model is computed and the stabilities of different equilibria of the model are studied in-detail. The model system exhibits backward bifurcation for some specific set of parameters which suggests that mere reducing the basic reproduction number corresponding to TB below one is not enough to establish the stability of TB-free equilibrium point. So efforts are needed to reduce this basic reproduction number much below one to have TB-free equilibrium to be stable. It is also observed that the smoking habits can influence the stability of co-existence equilibrium and it can lead to persistence of TB in the population even though the basic reproduction number corresponding to TB is much less than one. Furthermore, this model is extended to the optimal control problem and the optimal control model is analyzed using the Pontryagin's Maximum Principle and is solved numerically using MATLAB^{*TM}. Finally, numerical simulations are performed to analyze the effect of optimal control on the infected population. We observe that the optimal control model gives better result as compared to the model without optimal control as it reduces the number of infectives significantly within desired interval of time.

Keywords Epidemic model · Basic reproduction number · Stability · Optimal control

M. Ghosh (✉)
Division of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Chennai Campus, Chennai 600127, India
e-mail: minighosh@vit.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_4

1 Introduction

Tuberculosis (TB) is an infectious disease which is caused by bacterium *Mycobacterium tuberculosis* (MTB). TB mostly affects the lungs but it is not limited to it. TB bacteria can also attack to other parts of the body e.g. kidney, spine, brain etc. It is an established fact that individuals who smoke have increased risk of getting infected with TB. So in order to reduce the infection prevalence of TB in the population, it is desirable to control smoking habits of the individuals. Although, there are several research papers to study the transmission dynamics of TB, and smoking dynamics separately, there are very few mathematical models which incorporate the impact of smoking habits in the transmission dynamics of TB. In [1], authors formulated and analyzed a mathematical model to see the effect of smoking on transmission dynamics of TB by considering standard incidence type incidence. Later Choi et al. [2] extended this model to optimal control problem and compared the results by applying different types of optimal control strategies. In [3], authors have tried to model social, environmental and biological determinants of TB. Here too, authors have shown that TB incidence declines under a range of smoking reduction strategies. In this article, a mathematical model is formulated to study the transmission dynamics of TB in presence of smokers by considering simple mass-action type incidence which is more suitable for study of TB. Additionally, it is assumed that smokers infected with TB need to quit smoking to recover fully from TB. So individuals from the class of smokers infected with TB mainly move to the class of TB-infected or to the class of recovered individuals depending upon their immunity level when they quit smoking. This fact was not incorporated in the existing mathematical models. Further the proposed model is extended to optimal control problem by incorporating two types of optimal controls. Here the recovery rate constant for TB and the rate at which smokers infected with TB quit smoking are considered as the optimal control parameters. These parameters one can manipulate with suitable control efforts.

The remaining of this article is organized as follows: Section 2 describes the basic model and Sect. 3 exhibits the basic reproduction numbers and the existence of equilibria. Section 4 shows the stability analysis of different equilibria. Numerical simulation of the proposed model is demonstrated in Sect. 5. Section 6 deals with the optimal control problem and simulation results of optimal control problem is discussed in Sect. 7. Section 8 presents the results and discussion. Finally, Sect. 9 concludes the article and identifies some of the important future scope of the research.

2 Model Formulation

Here the whole population is divided into the following classes: susceptibles (S), latently infected (TB) individuals (E_T), TB-infected individuals (I_T), smokers (I_S), smokers latently infected with TB (E_{TS}), smokers infected with TB (I_{TS}), and recovered individuals (R_T). Hence the total human population $N(t)$ at time t is given by

$$N(t) = E_T + I_T + I_S + E_{TS} + I_{TS} + R_T.$$

Assuming simple mass-action type incidence, the force of infection for TB is denoted as λ_T and is given by

$$\lambda_T = \beta_T(I_T + \eta I_{TS}),$$

where β_T is the rate of transmission of TB to susceptible individuals and $\eta > 1$ is the modification parameter which reflects that the rate of transmission of TB will be more due to individuals who are smokers infected with TB compared to individuals infected with TB alone. Similarly, the strength of peer influence on smoking habits is denoted as λ_S and is given by

$$\lambda_S = \beta_S(I_S + \epsilon I_{TS}),$$

where β_S denotes the rate of transmission of smoking habits to susceptible individuals. Here the parameter $\epsilon < 1$ is the modification parameter which reflects the fact that the rate of transmission of smoking habits due to smokers infected with TB will be less compared to smokers without TB as they will be less socialized and will be under treatment.

As the risk of TB increases in smokers so the force of infection for TB in the class of smokers is denoted by λ_{TS} and is given by

$$\lambda_{TS} = \beta_{TS}(I_T + \eta I_{TS}) = \frac{\lambda_T \beta_{TS}}{\beta_T},$$

where β_{TS} denotes the rate of transmission of TB in the class of smokers. It is assumed that the susceptibles are getting infected with TB at a rate λ_T and are becoming smokers at a rate λ_S. A fraction p of TB infected susceptibles moves to latently infected compartment and the remaining $(1 - p)$ fraction moves to class of individuals infected with TB. This mechanism describes the slow and fast progression of TB disease. Individuals in latently infected compartment may move to class of TB-infected compartment due to exogenous reinfection at the rate $\omega\lambda_T$. Additionally, individuals in latently infected compartment move to TB-infected compartment at the disease progression rate ρ. Here it is assumed that smokers infected with TB may recover from TB when they quit smoking and depending upon their immunity level they may directly move to recover class at the rate ν_S or they may join TB-only class at rate σ_S. The individuals in latently infected compartment and recovered compartment may move to the class of smokers latently infected with TB at a rate λ_S if they come into contact with smokers with or without TB. With these assumption the mathematical model is formulated as follows:

$$
\begin{aligned}
\frac{dS}{dt} &= \Lambda - \mu S - \lambda_S S - \lambda_T S + \sigma I_S,\\
\frac{dE_T}{dt} &= p\lambda_T S - (\mu + \rho)E_T + \alpha E_{TS} + \zeta R_T - \lambda_S E_T - \omega\lambda_T E_T,\\
\frac{dI_T}{dt} &= (1-p)\lambda_T S - (\mu + \delta + \nu)I_T + \omega\lambda_T E_T + \rho E_T + \sigma_S I_{TS},\\
\frac{dI_s}{dt} &= \lambda_S S - (\mu + \sigma)I_S - \lambda_{TS} I_S,\\
\frac{dE_{TS}}{dt} &= \lambda_S(E_T + R_T) + p\lambda_{TS} I_S - \omega\lambda_T E_{TS} - (\alpha + \mu + \rho_S)E_{TS},\\
\frac{dI_{TS}}{dt} &= (1-p)\lambda_{TS} I_S + \omega\lambda_T E_{TS} + \rho_S E_{TS} - (\mu + \delta_S + \nu_S + \sigma_S)I_{TS},\\
\frac{dR_T}{dt} &= \nu I_T + \nu_S I_{TS} - (\mu + \zeta)R_T - \lambda_S R_T. \qquad (1)
\end{aligned}
$$

The schematic flow diagram of our model is shown in Fig. 1 and description of parameters is given in Table 1.

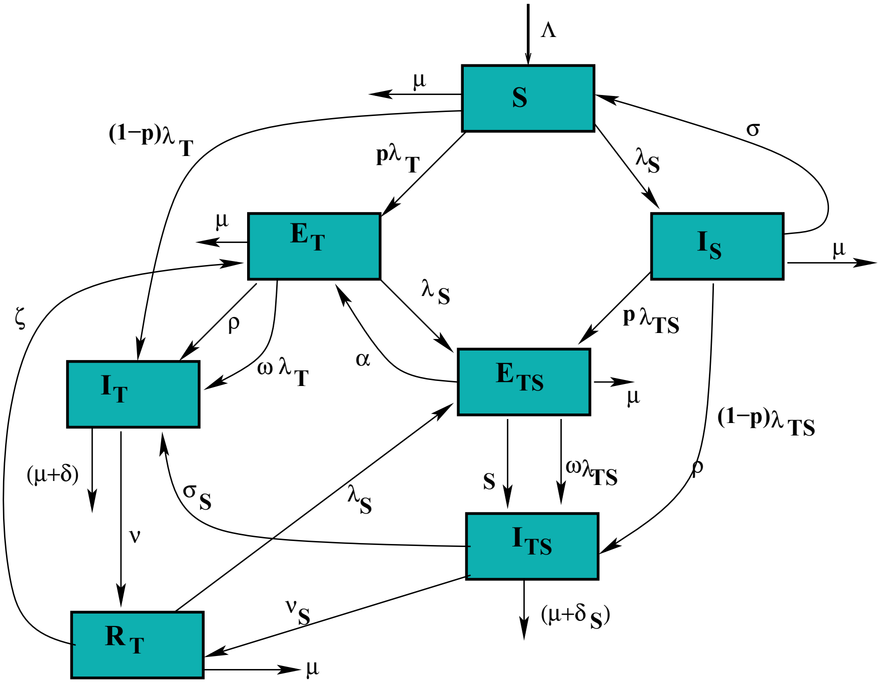


Fig. 1 Schematic diagram of transmission dynamics of TB in presence of smokers

Table 1 Description of parameters

Λ	Rate of recruitment
β_S	Rate of transmission of smoking habits
β_T	Rate of transmission of TB
β_{TS}	Rate of transmission of TB in the class of smokers
$\epsilon < 1$	Modification parameter
$\eta > 1$	Modification parameter
p, $(1 - p)$	Rate of slow and fast progression of TB
σ	Rate at which smokers quit smoking
μ	Natural death rate
ρ	Disease progression rate
ζ	Rate of movement of recovered individuals to class of latent TB
α	Rate at which smokers with latent TB quit smoking
ω	Parameter corresponding to exogenous re-infection
δ	TB related death rate
ν	Rate of recovery from TB in the class of TB infectives
σ_S	Rate at which smokers infected with TB quit smoking
ρ_S	Disease progression rate in the class of smokers with latent TB
δ_S	TB related death rate in the class of smokers infected with TB
ν_S	Rate of recovery from TB in the class of smokers infected with TB

Here all the parameters of the mathematical model (1) are assumed to be non-negative for all $t \geq 0$ and the region of attraction for all feasible solution is given by

$$\Omega = \left\{ (S, E_T, I_T, I_S, E_{TS}, I_{TS}, R_T) \in \mathcal{R}^7_+ : S + E_T + I_T + I_S + E_{TS} + I_{TS} + R_T \leq \frac{\Lambda}{\mu} \right\}.$$

It is easy to verify that the region Ω is positively invariant.

3 The Basic Reproduction Number and Existence of Equilibria

The model system (1) has four different types of equilibria, namely, (i) TB-free non-smokers equilibrium point $E_0(\frac{\Lambda}{\mu}, 0, 0, 0, 0, 0, 0)$, (ii) TB-free smokers only equilibrium point $E_1\left(\frac{\mu+\sigma}{\beta_S}, 0, 0, \frac{\beta_S\Lambda - \mu(\mu+\sigma)}{\beta_S\mu}, 0, 0, 0\right)$, (iii) Smoking-free TB-only equilibrium point $E_2(S^*, E_T^*, I_T^*, 0, 0, 0, R_T^*)$ and (iv) Co-existence equilibrium point $E_3(S^{**}, E_T^{**}, I_T^{**}, I_S^{**}, E_{TS}^{**}, I_{TS}^{**}, R_T^{**})$. Here it is noted that the equilibrium E_0 always exists, the equilibrium E_1 exists only when $\frac{\beta_S\Lambda}{\mu(\mu+\sigma)} = R_0^S$ (say) > 1. There are possibility of no, one or two smoking-free equilibrium E_2 depending upon existence of no, one or two positive root/roots S^* of the following quadratic:

$$\begin{aligned}&(\mu+\zeta)\beta_T\{(1-p)\mu+\rho-\omega\mu\}S^2+[\rho\zeta\nu-(\mu+\zeta)(\mu+\rho)(\mu+\delta+\nu)\\&\quad+\ (\mu+\zeta)\omega\{\beta_T\Lambda+\mu(\mu+\delta+\nu)\}-\omega\mu\nu\zeta]S\\&\quad-\ \omega\Lambda\{(\mu+\zeta)(\mu+\delta)+\mu\nu\}=0,\end{aligned}$$

and $I_T^*=\frac{\Lambda-\mu S^*}{\beta_T S^*}$, $R_T^*=\frac{\nu I_T^*}{\mu+\zeta}$, $E_T^*=\frac{p\beta_T S^* I_T^*+\zeta R_T^*}{(\mu+\rho+\omega\beta_T I_T^*)}$. The explicit expressions for the variables in the equilibrium point E_3 are not easy to obtain but one can show the existence of this equilibrium by graphical method. In the present work it is left and existence and stability of this equilibrium are discussed only in numerical simulation section.

The basic reproduction number $R_0=R_0^{ST}$ (say) is computed as follows using the next generation matrix method discussed in Driessche and Watmough (2002) [4].

$$R_0^{ST}=\max\left\{\frac{\beta_S\Lambda}{\mu(\mu+\sigma)},\ \frac{\rho(\mu+\zeta)p\beta_T\frac{\Lambda}{\mu}+(\mu+\rho)(\mu+\zeta)(1-p)\beta_T\frac{\Lambda}{\mu}}{(\mu+\rho)(\mu+\nu+\delta)(\mu+\zeta)-\zeta\rho\nu}\right\}=\max\{R_0^S,R_0^T\},$$

where $R_0^S=\frac{\beta_S\Lambda}{\mu(\mu+\sigma)}$ and $R_0^T=\frac{\rho(\mu+\zeta)p\beta_T\frac{\Lambda}{\mu}+(\mu+\rho)(\mu+\zeta)(1-p)\beta_T\frac{\Lambda}{\mu}}{(\mu+\rho)(\mu+\nu+\delta)(\mu+\zeta)-\zeta\rho\nu}$ denote the basic reproduction numbers corresponding to transmission of smoking-only and TB-only.

4 Stability Analysis

This section deals with the stability of different equilibria of the model. Here it is noted that the TB-free smoker-only equilibrium point $E_1\left(\frac{\mu+\sigma}{\beta_S},0,0,\frac{\beta_S\Lambda-\mu(\mu+\sigma)}{\beta_S\mu},0,0,0\right)$ is unrealistic and further analysis of this equilibrium is ignored. However numerical simulation shows that this equilibrium is unstable whenever R_0^S is greater than one.

Theorem 4.1 *The equilibrium point* $E_0=(\frac{\Lambda}{\mu},0,0,0,0,0,0)$ *is locally asymptotically stable whenever* $R_0^{ST}<1$ *and unstable otherwise.*

Proof The Jacobian matrix of the system (1) evaluated at $E_0=(\frac{\Lambda}{\mu},0,0,0,0,0,0)$ is given by the following matrix

$$\begin{pmatrix}-\mu & 0 & -\frac{\beta_T\Lambda}{\mu} & -\frac{\beta_S\Lambda}{\mu}+\sigma & 0 & -(\beta_S\epsilon+\beta_T\eta)\frac{\Lambda}{\mu} & 0\\ 0 & -(\mu+\rho) & \frac{p\beta_T\Lambda}{\mu} & 0 & \alpha & \frac{p\eta\beta_T\Lambda}{\mu} & \zeta\\ 0 & \rho & j_{33} & 0 & 0 & \frac{(1-p)\eta\beta_T\Lambda}{\mu}+\sigma_S & 0\\ 0 & 0 & 0 & \frac{\beta_S\Lambda}{\mu}-(\mu+\sigma) & 0 & \frac{\beta_S\epsilon\Lambda}{\mu} & 0\\ 0 & 0 & 0 & 0 & -(\alpha+\rho_S+\mu) & 0 & 0\\ 0 & 0 & 0 & 0 & \rho_S & -(\mu+\delta_S+\nu_S+\sigma_S) & 0\\ 0 & 0 & \nu & 0 & 0 & \nu_S & -(\mu+\zeta)\end{pmatrix},$$

where $j_{33} = \frac{(1-p)\beta_T\Lambda}{\mu} - (\mu + \delta + \nu)$. Clearly, four eigenvalues are $-\mu$, $\frac{\beta_S\Lambda}{\mu} - (\mu + \sigma)$, $-(\alpha + \rho_S + \mu)$ and $-(\mu + \delta_S + \nu_S + \sigma_S)$ and other eigenvalues are given by the roots of the following qubic equation:

$$\psi^3 + a_1\psi^2 + a_2\psi + a_3 = 0,$$

where

$$\begin{aligned}
a_1 &= 2\mu + \rho + \zeta - \left\{(1-p)\beta_T\frac{\Lambda}{\mu} - (\mu + \delta + \nu)\right\},\\
a_2 &= (\mu + \rho)(\mu + \zeta) + (2\mu + \rho + \zeta)(\mu + \delta + \nu) - (1-p)\beta_T\frac{\Lambda}{\mu}(2\mu + \rho + \zeta) - \rho p\beta_T\frac{\Lambda}{\mu},\\
a_3 &= -\left[(\mu + \rho)(\mu + \zeta)\left\{(1-p)\beta_T\frac{\Lambda}{\mu} - (\mu + \delta + \nu)\right\} + \rho p\beta_T\frac{\Lambda}{\mu}(\mu + \zeta) + \nu\zeta\right].
\end{aligned}$$

It is easy to verify that a_1, a_2, a_3 are positive and $a_1a_2 - a_3 > 0$ under the condition $R_0^{ST} < 1$. Hence from Routh-Hurwitch criteria the equilibrium point E_0 is locally asymptotically stable whenever the basic reproduction number R_0^{ST} is less than one.

Theorem 4.2 *The equilibrium point* $E_0 = (\frac{\Lambda}{\mu}, 0, 0, 0, 0, 0, 0)$ *may not be globally asymptotically stable in presence of smokers.*

Proof This theorem is proved by following the method in Castillo-Chavez et al. (2002) [5]. The system (1) can be written as follows:

$$\begin{aligned}
\frac{dX}{dt} &= F(X, Z),\\
\frac{dZ}{dt} &= G(X, Y),\ \ G(X, \mathbf{0}) = 0,
\end{aligned}$$

where $X = (S, R_T)$ and $Y = (E_T, I_T, I_S, E_{TS}, I_{TS})$ with $X \in \mathcal{R}_+^2$ denoting the classes of uninfected individuals and $Y \in \mathcal{R}_+^5$ denoting the infectious (i.e. smoking or TB or both) compartments. The disease-free equilibrium is now denoted by

$$E_0 = (X^*, \mathbf{0}),\ \text{where}\ X^* = \left(\frac{\Lambda}{\mu}, 0\right).$$

The global stability of the equilibrium point E_0 is guaranteed provided the following two conditions are met.
H1: For $\frac{dX}{dt} = F(X, \mathbf{0})$, X^* is globally asymptotically stable (g.a.s),
H2: $G(X, Y) = AY - G(\widehat{X}, Y),\ \widehat{G}(X, Y) \geq 0$ for $(X, Y) \in \Omega$,
where $A = D_Y G(X^*, \mathbf{0})$ is an M-matrix (i.e., all off-diagonal elements of A are non-negative) and Ω is the region where the model system is biologically feasible.

The model system (1) can be represented as follows:

$$F(X,\mathbf{0}) = \begin{pmatrix} \Lambda - \mu S \\ -(\mu+\zeta)R_T \end{pmatrix}$$

$G(X, Y) = AY - G(\widehat{X}, Y)$, where

$$A = \begin{pmatrix} -(\mu+\rho) & \frac{p\beta_T\Lambda}{\mu} & 0 & \alpha & \frac{p\eta\beta_T\Lambda}{\mu} \\ \rho & a_{22} & 0 & 0 & \frac{(1-p)\eta\beta_T\Lambda}{\mu} + \sigma_S \\ 0 & 0 & \frac{\beta_S\Lambda}{\mu} - (\mu+\sigma) & 0 & \frac{\beta_S\epsilon\Lambda}{\mu} \\ 0 & 0 & 0 & -(\alpha+\rho_S+\mu) & 0 \\ 0 & 0 & 0 & \rho_S & -(\mu+\delta_S+\nu_S+\sigma_S) \end{pmatrix},$$

where $a_{22} = \frac{(1-p)\beta_T\Lambda}{\mu} - (\mu+\delta+\nu)$ and

$$G(\widehat{X}, Y) = \begin{pmatrix} G_1(\widehat{X}, Y) \\ G_2(\widehat{X}, Y) \\ G_3(\widehat{X}, Y) \\ G_4(\widehat{X}, Y) \\ G_5(\widehat{X}, Y) \end{pmatrix} = \begin{pmatrix} p\lambda_T\left(\frac{\Lambda}{\mu} - S\right) + (\lambda_S + \omega\lambda_T)E_T - \zeta R_T \\ \lambda_T\left\{(1-p)\left(\frac{\Lambda}{\mu} - S\right) - \omega E_T\right\} \\ \lambda_S\left(\frac{\Lambda}{\mu} - S\right) + \lambda_{TS}I_S \\ -\lambda_S(R_T + E_T) - p\lambda_{TS}I_S + \omega\lambda_T E_{TS} \\ -(1-p)\lambda_T I_S - \omega\lambda_T E_{TS} \end{pmatrix}.$$

It is easy to notice that the matrix A is an M-matrix as all its off-diagonal elements are non-negative. To establish the global stability of E_0 one needs to show $G(\widehat{X}, Y) \geq 0$ but $G_5(\widehat{X}, Y) < 0$. Hence the global stability of the equilibrium point E_0 is not guaranteed. Here it can be noted that in the absence of smoking and exogenous re-infection the equilibrium point E_0 can be globally stable provided the parameter ζ which corresponds to latent TB in recovered individuals too is zero.

Theorem 4.3 *If* $\omega\frac{\mu}{\rho^2}\left\{\left\{(\mu+\delta+\nu) - (1-p)\beta^*\frac{\Lambda}{\mu}\right\} - \frac{\beta^{*2}\Lambda}{\mu^2}\left\{1 + \frac{\mu}{\rho}(1-p)\right\}\right\}$ > 0, *i.e.* $a > 0$, *then model system (1) undergoes a backward bifurcation at* $R_0^T = 1$, *otherwise* $a < 0$ *and a unique non-smoking TB-only equilibrium point* $E_2(S^*, E^*, I^*, 0, 0, 0, R_T^*)$ *is locally asymptotically stable for* $R_0^T > 1$, *but close to* 1.

Proof The stability of E_2 is proved using the Centre Manifold theory by Carr [6]. Using the Theorem 4.1 of [7], the local stability of the non-smoking TB-only equilibrium point is established. The original system (1) is rewritten by changing the variables as follows: $S = x_1, E_T = x_2, I_T = x_3, R_T = x_4$ and noting that $I_S = 0, E_{TS} = 0, I_{TS} = 0$. Let us consider the vector $X = (x_1, x_2, x_3, x_4)^T$ and write the model (1) in the form of $\frac{dX}{dt} = (f_1, f_2, f_3, f_4)^T$, where

$$\begin{aligned}
x_1' &= f_1 = \Lambda - \mu x_1 - \beta_T x_1 x_3 \\
x_2' &= f_2 = p\beta_T x_3 x_1 - (\mu+\rho)x_2 + \zeta x_4 - \omega\beta_T x_3 x_2 \\
x_3' &= f_3 = (1-p)\beta_T x_3 x_1 - (\mu+\delta+\nu)x_3 + \omega\beta_T x_3 x_2 + \rho x_2 \\
x_4' &= f_4 = \nu x_3 - (\mu+\zeta)x_4
\end{aligned} \tag{2}$$

The reproduction number of this model is same as R_0^T of the original model (1). Therefore β_T is chosen as bifurcation parameter. By solving $R_0^T = 1$ for β_T one can get $\beta_T = \beta^*$ as follows:

$$\beta^* = \frac{(\mu+\rho)(\mu+\nu+\delta)(\mu+\zeta) - \zeta\rho\nu}{\dfrac{\Lambda}{\mu}(\mu+\zeta)\{\mu(1-p)+\rho\}}.$$

The Jacobian matrix J_β^* of the system (2) evaluated at TB-free equilibrium point is given by

$$\begin{pmatrix}
-\mu & 0 & -\beta^*\dfrac{\Lambda}{\mu} & 0 \\
0 & -(\mu+\rho) & p\beta^*\dfrac{\Lambda}{\mu} & \zeta \\
0 & \rho & (1-p)\beta^*\dfrac{\Lambda}{\mu} - (\mu+\delta+\nu) & 0 \\
0 & 0 & \nu & -(\mu+\zeta)
\end{pmatrix}.$$

The above matrix has one simple eigenvalue as 0 and all other eigenvalues have negative real part. The right eigenvector associated with zero eigenvalue is given by $w = [w_1, w_2, w_3, w_4]^T$, where

$$\begin{aligned}
w_3 &= \frac{\mu+\zeta}{\nu} w_4,\ w_1 = -\frac{(\mu+\zeta)\beta^*\Lambda}{\nu\mu^2} w_4, \\
w_2 &= -\frac{(\mu+\zeta)\{(1-p)\beta^*\dfrac{\Lambda}{\mu} - (\mu+\delta+\nu)\}}{\rho\nu} w_4,\ w_4 > 0.
\end{aligned}$$

The left eigenvector associated with zero eigenvalue is given by $z = [z_1, z_2, z_3, z_4]^T$, where

$$z_1 = 0,\ z_2 = \frac{(\mu+\zeta)}{\zeta} z_4,\ z_3 = \frac{(\mu+\rho)(\mu+\zeta)}{\rho\zeta} z_4,\ z_4 > 0.$$

Following the same notation as discussed in [7], the bifurcation coefficients a and b are computed as follows:

$$a = \frac{(\mu+\zeta)^3}{\nu^2\zeta}\left[\beta^*\omega\frac{\mu}{\rho^2}\{(\mu+\delta+\nu) - (1-p)\beta^*\frac{\Lambda}{\mu}\} - \frac{\beta^{*2}\Lambda}{\mu^2}\{1 + \frac{\mu+\rho}{(1-p)}\}\right] z_4 w_4^2,$$

$$b = \frac{(\mu+\zeta)^2}{\zeta\nu}\frac{\Lambda}{\mu}\{1+\frac{\mu}{\rho}(1-p)\}z_4 w_4 > 0.$$

Here $b > 0$ and a can be positive or negative. So using the results discussed in [7] it is concluded that the model system (1) exhibits backward bifurcation when $a > 0$ at $R_0^T = 1$.

The existence and stability of non-trivial equilibrium point is not easy to prove analytically. However from the above discussion one can say that local stability of non-trivial equilibrium point E_3 should be guaranteed under some restriction on parameters. Also when R_0^T is less than one, one can have backward bifurcation for the original model. This fact is demonstrated in numerical simulation.

5 Numerical Simulation

Here first the model system (1) is simulated for the following set of parameters:

$$\Lambda = 500, \mu = 0.0166, \beta_T = 0.00001, \beta_S = 0.000014, \beta_{TS} = 0.000015, \sigma = 0.06,$$

$$p = 0.45, \rho = 0.000013, \alpha = 0.03, \zeta = 0.0945, \omega = 1.05, \delta = 0.04,$$

$$\nu = 0.08, \rho_S = 0.00002, \sigma_S = 0.12, \delta_S = 0.06, \nu_S = 0.05, \eta = 1.04, \epsilon = 0.8.$$

For this set of parameters both R_0^T and R_0^S are greater than one. Here $R_0^T = 1.214$ and $R_0^S = 5.505$. The equilibria $(30120, 0, 0, 0, 0, 0, 0)$ where both disease and smokers are not there. This equilibrium point is unstable. The smoking-free equilibrium point $(11033, 7227.4, 2872, 0, 0, 0, 2068)$ and TB-free equilibrium point $(5471.4, 0, 0, 24649, 0, 0, 0)$ are also unstable. The non-trivial equilibrium point $(7175.6, 3949.5, 2557.9, 2206.9, 3251.3, 697.13, 1598.7)$ is stable. Here Fig. 2 is demonstrating the stability of non-trivial equilibrium point. Next the parameter p is changed from 0.45 to 0.75 and all other parameters are kept as mentioned above. For this set of parameters, $R_0^T = 0.5527 < 1$ and $R_0^S = 5.505 > 1$. Hence $R_0^{ST} = 5.505 > 1$. In this case also a stable non-trivial equilibrium point as $(6996.8, 3420.8, 1525.8, 4422.7, 5869.1, 722.79, 873.5)$ exists. This suggest that backward bifurcation occurs. The stability of this equilibrium point is shown in Fig. 3.

Next, the model is simulated for different values of β_S to see the impact of this parameter on the equilibrium level of different population. The parameter p is taken as 0.65 and all other parameters are as mentioned above. Here it is observed that with the increase in this parameter the number of smokers and smokers infected with TB and TB-infectives increase (Figs. 4, 5 and 6). Here for all values of β_S, R_0^T is less than 1 and R_0^S is greater than one. But similar results appear for the set of parameters where both R_0^S and R_0^T are greater than one.

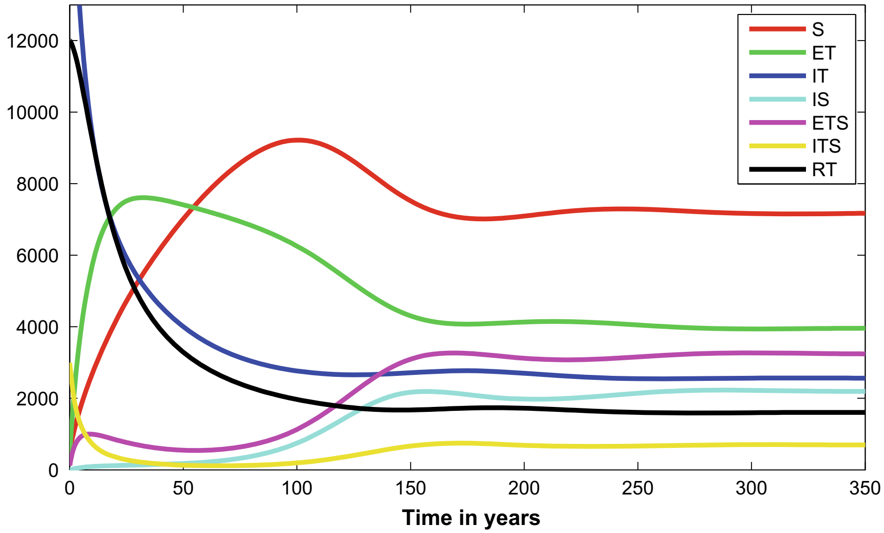


Fig. 2 Variation of different variables with time when both R_0^S and R_0^T are greater than one

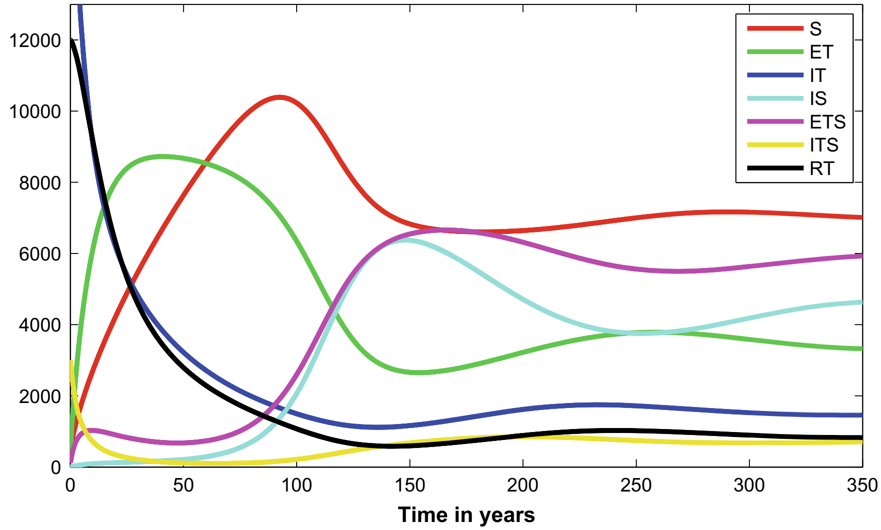


Fig. 3 Variation of different variables with time when $R_0^S > 1$ and $R_0^T < 1$

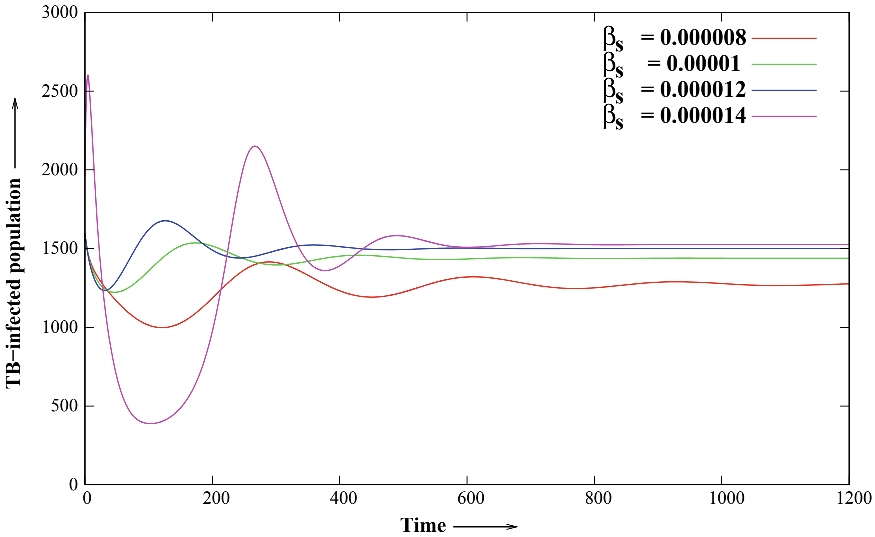


Fig. 4 Variation of TB-infected population with time for different values of β_S

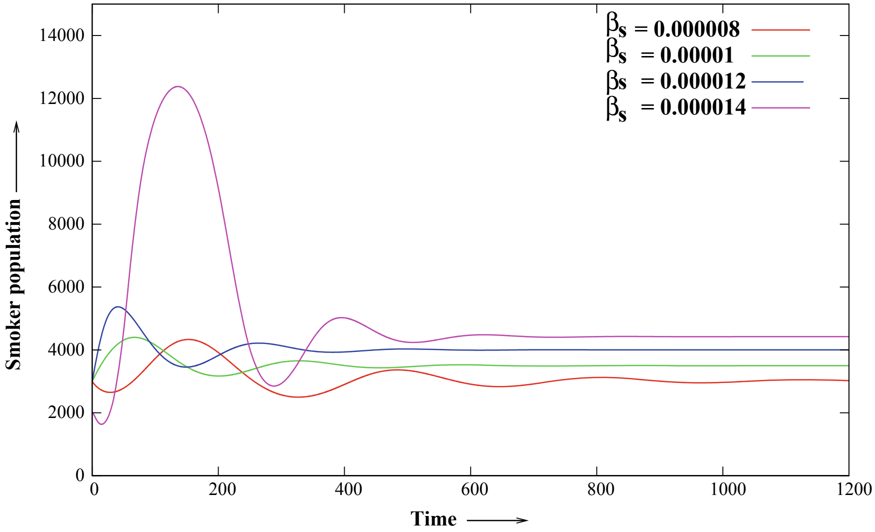


Fig. 5 Variation of smoker population with time for different values of β_S

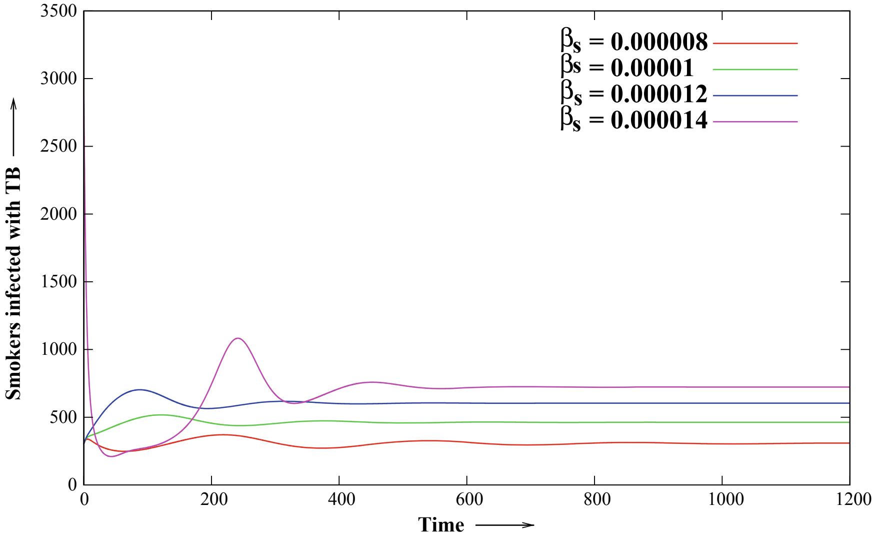


Fig. 6 Variation of smokers population infected with TB for different values of β_S

6 Optimal Control Problem

Here the mathematical model (1) is extended to optimal control problem. Optimal control is a method to determine time dependent control and state variables for a given dynamical system over a period of time to design suitable cost-effective strategies. Here the parameters ν (the recovery rate constant for TB-infectives) and σ_S (the rate at which smokers infected with TB leave smoking) are made time dependent. These two parameters are identified keeping in mind the fact that it is possible to regulate these parameters by increasing the number of TB-detection centres and counseling centres for smokers. If these are zero then there is no effort being placed in these controls at time *t* and if they are equal to one then maximum effort is applied. Keeping in view of the above assumptions, the optimal control model is formulated as follows:

$$
\begin{aligned}
\frac{dS}{dt} &= \Lambda - \mu S - \lambda_S S - \lambda_T S + \sigma I_S,\\
\frac{dE_T}{dt} &= p\lambda_T S - (\mu + \rho)E_T + \alpha E_{TS} + \zeta R_T - \lambda_S E_T - \omega\lambda_T E_T,\\
\frac{dI_T}{dt} &= (1-p)\lambda_T S - (\mu + \delta + \nu(t))I_T + \omega\lambda_T E_T + \rho E_T + \sigma_S(t) I_{TS},\\
\frac{dI_s}{dt} &= \lambda_S S - (\mu + \sigma)I_S - \lambda_{TS} I_S,\\
\frac{dE_{TS}}{dt} &= \lambda_S(E_T + R_T) + p\lambda_{TS} I_S - \omega\lambda_T E_{TS} - (\alpha + \mu + \rho_S)E_{TS},
\end{aligned}
$$

$$\begin{aligned}\frac{dI_{TS}}{dt} &= (1-p)\lambda_{TS}I_S + \omega\lambda_T E_{TS} + \rho_S E_{TS} - (\mu + \delta_S + \nu_S + \sigma_S(t))I_{TS},\\ \frac{dR_T}{dt} &= \nu(t)I_T + \nu_S I_{TS} - (\mu + \zeta)R_T - \lambda_S R_T,\end{aligned} \tag{3}$$

The objective functional for fixed time t_f is given below:

$$J = \int_0^{t_f} (C_1 I_T + C_2 I_S + C_3 I_{TS} + \frac{1}{2}C_4\nu^2 + \frac{1}{2}C_5{\sigma_S}^2). \tag{4}$$

Here the parameter $C_1 \geq 0$, $C_2 \geq 0$, $C_3 \geq 0$, $C_4 \geq 0$, $C_5 \geq 0$ and they represent the weight constants. Our objective is to find the control parameters ν^* and ${\sigma_S}^*$, such that

$$J(\nu^*, {\sigma_S}^*) = \min_{\nu,\sigma_S \in \Omega} J(\nu, \sigma_S), \tag{5}$$

where Ω is the control set and is defined as
$\Omega = \{\nu, \sigma_S : \text{measurable and } 0 \leq \nu, \sigma_S \leq 1\}$ and $t \in [0, t_f]$. The Lagrangian of this problem is defined as:

$$L(I_T, I_S, I_{TS}, \nu, \sigma_S) = C_1 I_T + C_2 I_S + C_3 I_{TS} + \frac{1}{2}C_4\nu^2 + \frac{1}{2}C_5{\sigma_S}^2.$$

The Hamiltonian $\mathcal{H}$ is given by

$$\begin{aligned}\mathcal{H} = L(I_T, I_S, I_{TS}, \nu, \sigma_S) &+ \lambda_1\frac{dS}{dt} + \lambda_2\frac{dE_T}{dt} + \lambda_3\frac{dI_T}{dt} + \lambda_4\frac{dI_S}{dt}\\ &+ \lambda_5\frac{dE_{TS}}{dt} + \lambda_6\frac{dI_{TS}}{dt} + \lambda_7\frac{dE_{R_T}}{dt},\end{aligned}$$

where λ_i are the adjoint variables and $i = 1$ to 7. Now adjoint variables in the form of differential equation can be written as follows:

$$\begin{aligned}\frac{d\lambda_1}{dt} &= -\frac{\partial H}{\partial S} = \lambda_1(\mu + \lambda_S + \lambda_T) - \lambda_2 p\lambda_T - \lambda_3(1-p)\lambda_T - \lambda_4\lambda_S\\ \frac{d\lambda_2}{dt} &= -\frac{\partial H}{\partial E_T} = \lambda_2(\mu + \rho + \Lambda_S + \omega\lambda_T) - \lambda_3(\omega\lambda_T + \rho) - \lambda_5\lambda_S\\ \frac{d\lambda_3}{dt} &= -\frac{\partial H}{\partial I_T} = -C_1 + \lambda_1 S\beta_T - \lambda_2(p\beta_T S - \omega\beta_T E_T)\\ &\quad -\lambda_3\{(1-p)\beta_T S - (\mu + \delta + \nu) + \omega\beta_T E_T\}\\ &\quad +\lambda_4\beta_{TS}I_S - \lambda_5(p\beta_{TS}I_S - \omega\beta_T E_{TS}) - \lambda_6\{(1-p)\beta_T I_S + \omega\beta_T E_{TS}\} + \lambda_7\nu(t)\\ \frac{d\lambda_4}{dt} &= -\frac{\partial H}{\partial I_S} = -C_2 + \lambda_1(\sigma - \beta_S S) + \lambda_2\beta_S E_T - \lambda_4\{\beta_S S - (\mu + \sigma) - \beta_{TS}(I_T + \eta I_{TS})\}\end{aligned}$$

$$
\begin{aligned}
&-\lambda_5\{\beta_S(E_T+R_T)+p\lambda_{TS}\}-\lambda_6(1-p)\lambda_T\\
\frac{d\lambda_5}{dt}&=-\frac{\partial H}{\partial E_{TS}}=-\lambda_2\alpha+\lambda_5(\omega\lambda_T+\alpha+\mu+\rho_S-\lambda_6(\omega\lambda_T+\rho_S)\\
\frac{d\lambda_6}{dt}&=-\frac{\partial H}{\partial I_{TS}}=-C_3+\lambda_1 S(\beta_T\eta+\epsilon\beta_S)-\lambda_2(p\eta\beta_T S-\epsilon\beta_S E_T)\\
&-\lambda_3\,\{(1-p)S\beta_T\eta+\omega E_T\beta_T\eta\\
&+\sigma_S(t)\}-\lambda_4(\epsilon\beta_S S-\beta_{TS}\eta I_S)-\lambda_5\{(E_T+R_T)\beta_S\epsilon+p\eta I_S\beta_{TS}-\omega\eta\beta_T E_{TS}\}\\
&-\lambda_6\{(1-p)I_S\eta\beta_T+\omega\eta\beta_T E_{TS}-(\mu+\delta_S+\nu_S+\sigma_S(t))\}-\lambda_7\nu_S\\
\frac{d\lambda_7}{dt}&=-\frac{\partial H}{\partial R_T}=-\lambda_2\zeta-\lambda_5\lambda_S+\lambda_7(\mu+\zeta)
\end{aligned}
\tag{6}
$$

Let $\widetilde{S}$, $\widetilde{E_T}$, $\widetilde{I_T}$, $\widetilde{I_S}$, $\widetilde{E_{TS}}$, $\widetilde{I_{TS}}$, $\widetilde{R_T}$ be the optimum values of $S, E_T, I_T, I_S, E_{TS}, I_{TS}, R_T$ respectively, and $\widetilde{\lambda}_1$, $\widetilde{\lambda}_2$, $\widetilde{\lambda}_3$, $\widetilde{\lambda}_4$, $\widetilde{\lambda}_5$, $\widetilde{\lambda}_6$, $\widetilde{\lambda}_7$ be the solution of the system (3).

By using [8, 9], the following theorem is stated below without proof:

Theorem 6.1 *There exist optimal controls $\nu^*, {\sigma_S}^* \in \Omega$ such that $J(\nu^*, {\sigma_S}^*) = \min J(\nu, \sigma_S)$ subject to system (3).*

With the help of Pontryagin's maximum principle [9] and Theorem (6.1), one can prove the following theorem:

Theorem 6.2 *The optimal controls $(\nu^*, {\sigma_S}^*)$ which minimizes J over the region Ω given by*

$$
\nu^* = min\{1, max(0, \widetilde{\nu})\}, \quad {\sigma_S}^* = min\{1, max(0, \widetilde{\sigma_S})\}, \quad where
$$

$$
\widetilde{\nu} = \frac{(\lambda_3 - \lambda_7)I_T}{c_4}, \quad \widetilde{\sigma_S} = \frac{(\lambda_6 - \lambda_3)I_{TS}}{c_5}.
$$

7 Simulation of Optimal Control Model

The optimal control model is simulated using MATLAB by considering the set of parameters which corresponds to the stability of endemic equilibrium point of the model (1). The weight constants are taken as follows:

$$
C_1 = 1, C_2 = 1, C_3 = 22, C_4 = 10, C_5 = 20.
$$

The time interval is considered as [0, 20]. The system (3) is solved by iterative method with the help of forward and backward difference approximation. Here Figs. 7 and 8 are showing the control profiles of the controls $\nu(t)$ and $\sigma_S(t)$ respectively.

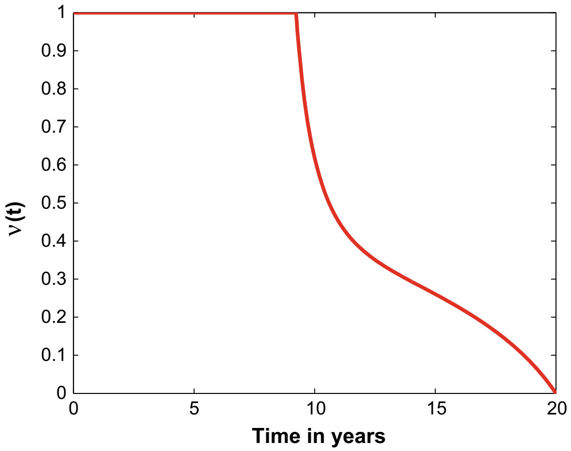


Fig. 7 Control profile of $\nu(t)$

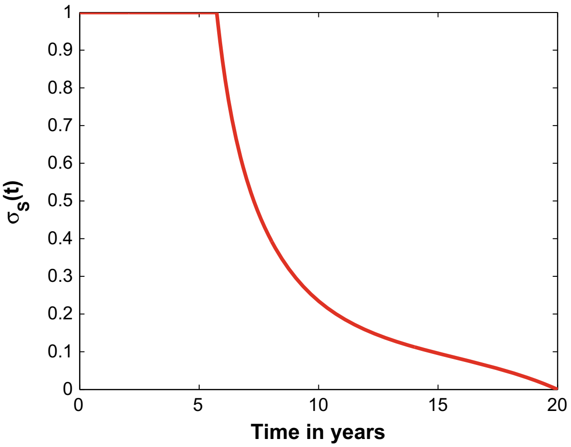


Fig. 8 Control profile of $\sigma_S(t)$

Finally, to see the effects of optimal controls, the TB-infected population and smokers infected with TB are plotted against time with and without optimal control in Figs. 9 and 10. It is easy to observe that optimal control is very much effective in reducing the number of infectives in the desired interval of time.

8 Discussion

From the simulation results discussed in Sect. 5, it is clear that even if the reproduction number corresponding to TB is less than one and if the reproduction number corresponding to smoking habits is greater than one, then the system tends to achieve a stable non-trivial equilibrium point, i.e. TB persists in the population under con-

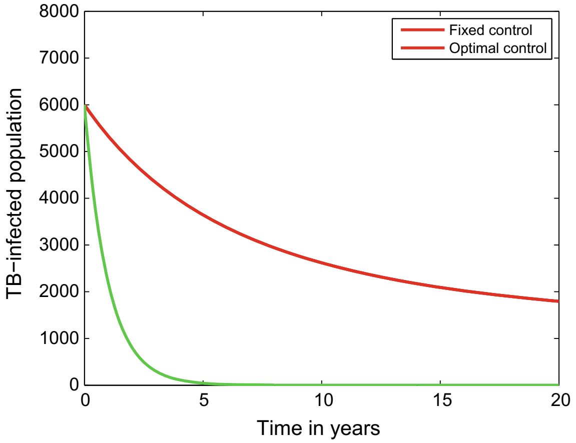


Fig. 9 Variation of TB-infected population against time with and without optimal control

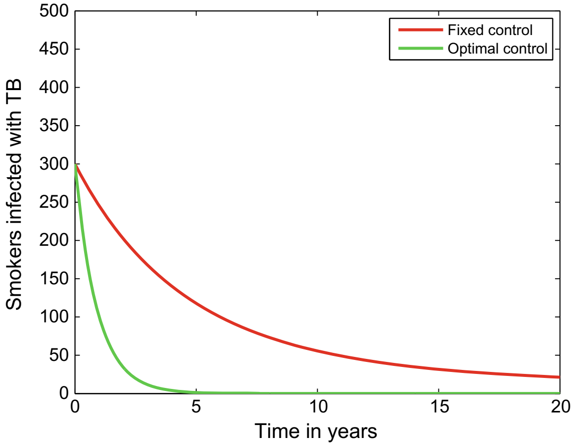


Fig. 10 Variation of smokers infected with TB population against time with and without optimal control

sideration. The increase in the rate of transmission β_S corresponding to the smoking habits causes an increase not only in the equilibrium level of smokers but also in the class of TB-infected population. The proposed model can help in quantifying the required decrease in the parameter β_S and β_T which can lead to the stability of TB-free non-smokers equilibrium point. Although there are several parameters which can be treated as optimal control parameters, in the present article the recovery rate constants corresponding to TB and the rate at which smokers infected with TB quit smoking are considered as optimal control parameters as these are easy to implement and also are easy to quantify.

From the simulation of optimal control model, it is observed that one needs to maintain maximum control up to 9 years for $\nu(t)$ and up to 6 years for $\sigma_S(t)$. Then, these controls can be decreased slowly to get the optimal cost effective results in the overall duration of 20 years. From Figs. 9 and 10, it is clear that both the TB-infected population and the population of smokers infected with TB decrease to zero within 5 years of time and remains at this level throughout the further period of 15 years. This suggests that to keep the population TB-free, continuous efforts are required and one should not abandon the control strategies by seeing the prevalence of TB to its minimum, otherwise the disease may reappear.

9 Conclusion

In this article a non-linear mathematical model for the transmission dynamics of TB in presence of smokers has been formulated and analyzed. The analysis of the proposed model emphasizes the need of reduction in smoking habits to have reduction in the prevalence of TB. This model exhibits backward bifurcation and there exists stable co-existence equilibrium point even when $R_0^T < 1$ and $R_0^S > 1$. Here, it can be noted that backward bifurcation occurs due to re-infection and slow to fast progressions of TB. So, for the stable co-existence equilibrium, R_0^S must be greater than one. Furthermore, this model has been extended to the optimal control problem. The numerical simulation of optimal control problem reflects the control profiles of the optimal control parameters which cause the significant decrease in the TB-infected individuals and the smokers infected with TB. Thus, by adjusting the recovery rate of TB and the rate at which smokers infected with TB quit the smoking, one can control both the TB and smoking within a given interval of time.

However, the present model does not incorporate the control in the transmission of TB through the social distancing programs whereas it is possible to have this strategy once individual is identified as infected with TB. We all know that smoking is not good for health and passive smoking also increases the risk of respiratory disease. So, the advertisement through media and other sources can increase the social distancing from the smokers or group of smokers. And this can reduce the smoking habits of smokers and can impact greatly in the infection prevalence of TB. So far, these facts have not been analyzed through mathematical model and this can be considered as one of the future endeavors in this area. Additionally, it is observed that heavy alcohol consumption adds to the risk of active TB. It would be interesting to study the synergistic effects of smoking and alcohol drinking on the infection prevalence of TB. Some of the planned future works will go in these directions and currently they are under investigation.

Acknowledgements The author thanks the handling editor and anonymous referees for their valuable comments and suggestions which have led to an overall improvement of the original article.

References

1. Bhunu, C.P., Mushayabasa, S., Tchuenche, J.M.: A theoretical assessment of the effects of smoking on the transmission dynamics of tuberculosis. Bull. Math. Biol. **73**, 1333–1357 (2011)
2. Choi, S., Jung, E., Lee, S.-M.: Optimal intervention strategy for prevention tuberculosis using a smoking-tuberculosis model. J. Theor. Biol. **380**, 256–270 (2015)
3. Murray, M., Oxlade, O., Lin, H.-H.: Modeling social, environmental and biological determinants of tuberculosis. Int. J. Tuberc. Lung Dis. **15**(6), S64S70 (2011)
4. Driessche, P.V., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
5. Castillo-Chavez, C., Feng, Z., Huang, W.: On the computation of R_0 and its role on global stability. In: Mathematical Approaches for Emerging and Reemerging Infectious Diseases. An Introduction, pp. 229–250. Springer, Berlin (2002)
6. Carr, J.: Applications Centre Manifold Theory. Springer, New York (1981)
7. Castillo-Chavez, C., Song, B.: Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. **1**(2), 361404 (2004)
8. Lenhart, S., Workman, J.T.: Optimal Control Applied to Biological Models. CRC Press (2007)
9. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Interscience Publishers (1962)

Modelling the Effects of Stigma on Leprosy

Stephen G. Mosher, Christian Costris-Vas and Robert Smith?

Abstract The World Health Organization's leprosy-elimination campaign has significantly reduced global leprosy prevalence, but approximately 214,000 new cases of leprosy are reported each year. An ancient and neglected affliction, leprosy is also one of the most heavily stigmatised diseases of all time. We developed a mathematical model to examine the effects of stigma on sustaining disease transmission, using low and high degrees of stigma, as well as in its absence. Our results show that stigma does indeed play a central role in the long-term sustainability of leprosy. We also examined sensitivity of the outcome to all parameters and showed that the effects of stigma could increase the number of infected individuals by a factor of 80. Therefore both targeted education and shifts in cultural attitudes towards leprosy will be necessary for the eventual eradication of the disease.

Keywords Leprosy · Stigma · Mathematical model · Latin Hypercube Sampling · Partial rank correlation coefficients

S. G. Mosher
Department of Earth and Environmental Sciences,
The University of Ottawa, Ottawa, Canada
e-mail: stephenmosher@gmail.com

C. Costris-Vas
Department of Mathematics, The University of Ottawa, Ottawa, Canada
e-mail: ccost069@uottawa.ca

R. Smith? (✉)
Department of Mathematics and Faculty of Medicine, The University of Ottawa,
150 Louis-Pasteur Pvt, Ottawa, ON K1N6N5, Canada
e-mail: rsmith43@uottawa.ca

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_5

1 Introduction

Leprosy is a disease that has affected human beings for millennia; however, its causative agent, the bacteria *Mycobacterium Leprae*, was not identified until 1872 by Armauer Hansen. In the 1940s, a cure was developed with the drug dapsone; however, dapsone requires treatment for life. In the 1960s, drug resistance evolved due to widespread use of dapsone until the 1970s and 80s, when, upon recommendation from the World Health Organization (WHO), multi-drug therapy (MDT) was developed. Combining dapsone with the drugs clofazimine and rifampicin [17] resulted in a cure rate of 98% [4]. In addition to solving the drug-resistance problem, the MDT cocktail for leprosy also lessened the drug treatment timescale from life to a maximum of 24 months, depending on the type and severity of infection [18]. Thus, following the success of MDT for leprosy, the WHO launched a leprosy-elimination campaign in 1991 [10]. The target date of the 1991 WHO elimination campaign was the year 2000, for which elimination was defined in terms of a global prevalence threshold of less than 1 case in 10,000. In 1995, it was resolved that the MDT cocktail for leprosy would be provided to all patients worldwide for free [18]; in 2000, the WHO claimed to have achieved their elimination goal, citing a global prevalence of less than 600,000 cases. Yet, while all but six countries reported leprosy prevalences of less than 1 case in 10,000 in 2005 [17], approximately 214,000 new cases of leprosy are currently reported each year, with a global case-detection rate of 3.78 per 100,000 [4]. About 80% of all new cases come from India, Brazil and Indonesia [1, 4]. The WHO's original use of the term "elimination" has been criticised [10, 17] as an inhibiter to the progress of further leprosy reduction after 2000. The WHO later rebranded its leprosy efforts as an "Enhanced global strategy for further reducing the disease burden due to leprosy 2011–2015" [1].

Leprosy is a bacterial infection of the skin and is a leading infectious cause of disability [13, 18]. Yet, even among the neglected tropical diseases, it is one of the most overlooked [1]. This is in part because an effective treatment exists for the disease and in part because it is hard to quantify the cumulative socio-economic impact of the disease, as the disease is not fatal. Rather, leprosy infections lead to a plethora of secondary problems, such as infection of untreated wounds, debilitating ulcers on palms and soles, nerve-function impairment and damage, chronic disability, blindness and severe disfigurement [20].

The only known reservoirs of the *M. Leprae* bacteria are humans and South American armadillos [18]. However, *M. Leprae* can also survive outside the body for up to 45 days [10]. While the exact transmission mechanism is still unknown, most scholars agree that it involves direct contact with nasal fluids from the infected. Furthermore, it is thought that most people infected with *M leprae* do not develop clinical infections [18]. Two types of clinical infections may develop, either paucibacillary (PB) or multibacillary (MB) leprosy, with MB leprosy being the more severe. Of the two types of infections, MB leprosy is thought to be the only infectious type or at least the main source of *M. Leprae* [1, 18]. The type of infection that develops in an individual is thought to be largely mediated by the response of their immune system [1]. In the case of MB leprosy, the bacteria spread systematically, and lesions tend to contain higher levels of bacilli. To simplify the diagnosis process in the field, an operational classification of leprosy has been developed by the WHO, in which patients are diagnosed simply by the number of skin lesions they have [4]. In cases where an infectee has five or more skin lesions, they are classified as having MB leprosy; otherwise, the classification is PB leprosy [18]. The incubation period from sub-clinical to clinical infections is extremely slow for leprosy, ranging from 2–12 years [18], and there is currently a lack of diagnostic tools for detecting early levels of *M. Leprae* [4]. While a vaccine specifically geared toward leprosy immunisation does not exist, the bacillus Calmette–Guérin (BCG) vaccine, originally developed for tuberculosis (TB), is known to provide variable protection against leprosy [13]. However, given that a new TB vaccine will likely supersede BCG in the future, the eventual consequences for leprosy control efforts remain unclear [13].

Stigma confers itself in several forms: exterior social forces, sometimes denoted "community stigma", and the emotional harm contained by an individual within themselves [23]. A further understanding has been reached concerning the layers of cognitive categories and the ways that they complement the predisposed beliefs of a particular disease. These include labelling, stereotyping, cognitive separation and emotional reactions [9]. Perception of stigma and experiences of discrimination cause people to feel ashamed and may cause them to isolate themselves from society [19], thus perpetuating the stereotype that leprosy is something shameful to be hidden away [23]. Alongside the emotional trauma is the added effect of prolonging individual instances of infection and increasing the chance of spread to others [7, 26]. The impact of knowing that one carries the disease and the anticipated stigma is in some instances as great or an even greater source of suffering than symptoms of the disease itself [22, 24], These factors play into the propensity to hide the ailment, which prolongs its affliction on the individuals involved and society as a whole [26]. Recently, there has been a substantial interest in understanding and diluting the overarching trend of stigma in many of today's diseases [24]. Several initiatives are being explored to address the prominence of stigma in sustaining the disease and the impact it has from the perspective of the individual [14]. These include alleviating health problems with improved social policies, unhinging the inclination to stigmatise on the part of perpetrators and better supporting those already affected by social neglect [24]. These all aspire to the same end goal, which is the transformation of stigma into social support rather than an increased burden [3].

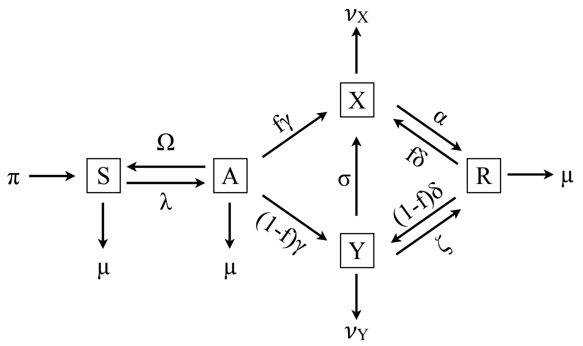
The effects of leprosy stigma are widespread, negatively affecting employment, marriages and social activities of those infected, recovering or recovered [16]. Such aspects of disease-related stigma further deteriorate both the psychological and physiological states of individuals, promoting a feedback loop of negative overall health [24]. Furthermore, there is much ignorance regarding leprosy in countries where it persists. For example, stigmatisation of infected individuals is promoted through local legislation in some countries, while many believe the disease to be hereditary when it is not [21]. The reason for this misconception is that family members of those infected are disproportionately at higher risk of infection due to frequent close contact with infected individuals [1]. Leprosy also disproportionately affects poorer citizens who lack access to care, which further promotes unfounded stigma against the poor [11]. Hence recovered infectees face substantial risks to their overall health and well-being even after being released from MDT [23]. It is also known that stigma can lead to significant delays with respect to case detection and self-diagnosis [24]. However, it is when individuals first develop symptoms that they derive the most benefit from MDT. This is because, while MDT is an effective treatment at all stages, severe nerve damage or further complications arising from the disease grow with time and remain permanent after infections are cleared [21].

To date, only a handful of mathematical models have addressed leprosy [6]. Models have used compartments [8], been simulation based [12] or individual based [5] and have considered treatment and relapse [15]. However, to the best of our knowledge, there have been no attempts to model the effects of stigma in the transmission dynamics of leprosy. Therefore, we propose a simple model for the transmission dynamics of leprosy that can account for the effects of stigma. We address the following research questions: 1. What role does stigma play in the transmission dynamics of leprosy? 2. How sensitive is the outcome to variation in disease parameters, including stigma? 3. Can leprosy be eradicated if stigma is removed?

2 The Model

Our model consists of susceptible, exposed, infected and recovered individuals, with the productively infected compartment split in two. The model allows for otherwise healthy individuals to contract leprosy, clear asymptomatic infections, progress from asymptomatic to symptomatic infection states, recover through MDT or relapse to symptomatic infection. Specifically, the five classes under consideration are as follows: the susceptible class 'S', those with sub-clinical or asymptomatic infections 'A', those with symptomatic infections that they choose to disclose 'X', those with symptomatic infections that they choose to conceal 'Y', and those who have recovered from the disease 'R'. By splitting the non-asymptomatic infection compartment into two discrete groups, the model mimics the choice, available to members of the population who develop symptomatic leprosy infections, to either conceal or disclose their infection. Likewise, the same choice is available to members of the population who relapse into symptomatic infections. Two further possibilities are accounted for by the model, in which members who originally concealed their symptomatic infection may later change their minds, disclosing their infection or are discovered. This is manifested as a path from 'Y' to 'X'. Incorporating this split in the infection compartment is a simple way to explicitly account for the effects of stigma on the transmission of leprosy.

Fig. 1 Model flow diagram



The differential equations governing our model are as follows:

$$
\begin{aligned}
S' &= \pi + \Omega A - (\mu + \lambda)S \\
A' &= \lambda S - (\mu + \gamma + \Omega)A \\
X' &= f\gamma A + \sigma Y + f\delta R - (\nu_X + \alpha)X \\
Y' &= (1 - f)\gamma A + (1 - f)\delta R - (\nu_Y + \sigma + \zeta)Y \\
R' &= \alpha X + \zeta Y - (\mu + \delta)R,
\end{aligned}
\tag{1}
$$

with the force of infection given by $\lambda = \beta_1(1 - \eta)A + \beta_1 Y + (\beta_1 - \frac{\beta_2 X}{m+X})X$.

A schematic of the model is shown in Fig. 1, and the model parameters are described in Table 1.

In deriving this model of leprosy, we make the following assumptions:

1. The chance of infection depends upon interactions between S and classes A, X and Y, although the most prominent course of infection remains the interaction between S and Y (the stigma class); the β_1 term is thus the largest of the transmission terms.
2. Because it remains difficult to detect asymptomatic infections, members of the A class act as usual, interact with susceptibles as usual and die at the natural death rate.
3. The asymptomatic class, which carries the *M. Leprae* bacteria, has a naturally reduced transmission compared to the stigma class.
4. The effect of transmission from the X class is modified by a dampening term that reduces infectivity. This dampening term takes the form of a Holling Type II function with the property that the effect saturates at a level β_2 when there are large numbers of individuals in the disclosing class. This reflects the fact that susceptible individuals will likely attempt to reduce contact with individuals who openly display symptoms, but the effect of such avoidance is limited when numbers are large. As such, $\beta_1 > \beta_2$.
5. A large fraction Ω of members with asymptomatic leprosy infections clear such infections [18], after which they return to the susceptible class.

Table 1 Variables and parameters

Symbol	Description	Range	Units	Reference
S	Susceptible individuals	–	people	–
A	Asymptomatically infected individuals	–	people	–
X	Infected individuals with disclosed, symptomatic infections	–	people	–
Y	Infected individuals with undisclosed, symptomatic infections	–	people	–
R	Recovered individuals	–	people	–
π	Birth rate	10–30	$\frac{\text{people}}{\text{year}}$	[25]
λ	Force of infection	–	$\frac{1}{\text{year}}$	–
β_1	Transmissibility between susceptibles and non-disclosing infectees	0.001–0.01	$\frac{1}{\text{people} \cdot \text{year}}$	[4]
β_2	Dampened transmissibility between susceptibles and disclosing infectees	$\frac{\beta_1}{2}$	$\frac{1}{\text{people} \cdot \text{year}}$	Estimate
η	Coefficient of reduced infection between for asymptomatic infectees	0–1	–	–
m	Half-saturation constant	0–5	people	Estimate
Ω	Clearance rate of asymptomatic infection	0.5–1	–	[18]
μ	Natural death rate	0.01–0.03	$\frac{1}{\text{year}}$	(Lifespan of 33–100 years)
γ	Rate of developing symptoms	0–0.2	$\frac{1}{\text{year}}$	[18]
f	Fraction who disclose infection	0–1	–	–
σ	Rate of eventual disclosure of infection	0.5–1	$\frac{1}{\text{year}}$	[18]
ν_X	Disease death rate (unstigmatised)	0.03–0.09	$\frac{1}{\text{year}}$	Estimate
ν_Y	Disease death rate (stigmatised)	0.09–1	$\frac{1}{\text{year}}$	Estimate
α	Recovery rate without stigma	0.1–1	$\frac{1}{\text{year}}$	[18]
ζ	Recovery rate with stigma	0–0.1	$\frac{1}{\text{year}}$	Estimate
δ	Relapse rate	0–0.1	$\frac{1}{\text{year}}$	[4, 18]

6. A fraction f of individuals who develop symptomatic infections will disclose their infection and seek MDT; we assume this fraction is the same at the onset of initial infection as it is when immunity lapses.
7. Those who initially conceal their symptomatic infection are identified at rate σ, either by changing their minds or due to discovery.
8. Individuals who do not disclose their symptoms consequently do not seek MDT and do not recover.

9. In the absence of stigma, all members who progress to symptomatic infection have the opportunity to seek MDT, which is highly successful and freely available.
10. There is a small chance for symptomatic infection relapse (approximately 2–3%) [18].
11. Some stigmatised individuals may disclose only to their doctors and receive MDT; we assume this factor will be much smaller than stigmatised individuals who seek treatment.

In addition to these primary assumptions, we further assume constant birth and death rates and ignore vaccination.

Remark We note that, when $f = 1$, then $Y = 0, \sigma$ is not relevant, $\lambda = \beta_1(1-\eta)A + (\beta_1 - \frac{\beta_2 X}{m+X})X$, and our model collapses to the special case of leprosy without stigma, corresponding to assumption (9).

3 Analysis

3.1 Equilibria

The disease-free equilibrium (DFE) is given by $(\overline{S}, \overline{A}, \overline{X}, \overline{R}) = (\pi/\mu, 0, 0, 0)$.

To find the endemic equilibrium, we start by setting the governing equations of our model to zero:

$$0 = \pi + \Omega A - \mu S - \beta_1(1-\eta)SA - \beta_1 SY - \left(\beta_1 - \frac{\beta_2 X}{m+X}\right)SX \tag{2}$$

$$0 = \beta_1(1-\eta)SA + \beta_1 SY + \left(\beta_1 - \frac{\beta_2 X}{m+X}\right)SX - (\mu+\gamma+\Omega)A \tag{3}$$

$$0 = f\gamma A + \sigma Y + f\delta R - (\nu_X + \alpha)X \tag{4}$$

$$0 = (1-f)\gamma A + (1-f)\delta R - (\nu_Y + \sigma + \zeta)Y \tag{5}$$

$$0 = \alpha X + \zeta Y - (\mu+\delta)R. \tag{6}$$

We rearrange Eq. (6) to obtain

$$R(X,Y) = \frac{\alpha}{\mu+\delta}X + \frac{\zeta}{\mu+\delta}Y. \tag{7}$$

Next, we substitute Eq. (7) into Eq. (5), giving

$$0 = (1-f)\gamma A + (1-f)\frac{\alpha\delta}{\mu+\delta}X + (1-f)\frac{\zeta\delta}{\mu+\delta}Y - (\nu_Y + \sigma + \zeta)Y, \tag{8}$$

which we then rearrange to get

$$A(X,Y)=\left((\nu_Y+\sigma+\zeta)Y-(1-f)\frac{\zeta\delta}{\mu+\delta}Y-\frac{(1-f)\alpha\delta}{\mu+\delta}X\right)\frac{1}{(1-f)\gamma}. \quad (9)$$

Next, substituting our expressions for $A(X,Y)$ and $R(X)$ into Eq. (4), we get

$$Y(X)=\frac{(\nu_X+\alpha)(1-f)}{f(\nu_Y+\sigma+\zeta)+\sigma(1-f)}X\equiv qX. \quad (10)$$

We re-express (9) as

$$A(X)=\left[(\nu_Y+\sigma+\zeta)q-\frac{(1-f)\zeta\delta}{\mu+\delta}q-\frac{(1-f)\alpha\delta}{\mu+\delta}\right]\frac{1}{(1-f)\gamma}X\equiv nX, \quad (11)$$

and it is easy to show that $n>0$. We now substitute $A(X)$ and $Y(X)$ into Eq. (3) and use the fact that $X\neq 0$ to get

$$S(X)=\frac{n(\mu+\gamma+\Omega)}{n(1-\eta)\beta_1+\beta_1 q+\left(\beta_1-\frac{\beta_2 X}{m+X}\right)}. \quad (12)$$

Note that

$$\beta_1-\frac{\beta_2 X}{m+X}>\beta_1-\beta_2>0,$$

and hence $S(X)>0$.

Finally, we substitute Eq. (12) into Eq. (2) along with the condensed forms of Eqs. (10) and (11) to find that

$$\begin{aligned}0=X^2&\Big[\beta_1(1-\eta)n^2(\mu+\gamma)+\beta_1 qn(\mu+\gamma)+\beta_1 n(\mu+\gamma)-\beta_2 n(\mu+\gamma)\Big]\\&+X\Big[\mu n(\mu+\gamma+\Omega)+\beta_1(1-\eta)n^2(\mu+\gamma)m+\beta_1 qn(\mu+\gamma)m+\beta_1 mn(\mu+\gamma)\\&\quad+\pi\beta_2-\pi n(1-\eta)\beta_1-\pi\beta_1 q-\pi\beta_1\Big]\\&+\mu n(\mu+\gamma+\Omega)m-\pi n(1-\eta)\beta_1 m-\pi\beta_1 qm-\pi\beta_1 m\\&=a_1X^2+b_1X+c_1.\end{aligned}$$

Since $\beta_1>\beta_2$, it follows that $a_1>0$, and hence this is an upward-facing parabola. Furthermore, since $S\leq\frac{\pi}{\mu}$ (the upper bound of the population), then, using Eq. (12), the constant term satisfies

$$\begin{aligned} c_1 &= \mu n(\mu + \gamma + \Omega)m - (\pi n(1-\eta)\beta_1 + \pi\beta_1 q + \pi\beta_1)m \\ &\leq \pi n(1-\eta)\beta_1 m + \pi\beta_1 q m + \pi\beta_1 m - \frac{\pi\beta_2 X m}{m+X} - (\pi n(1-\eta)\beta_1 + \pi\beta_1 q + \pi\beta_1)m \\ &= -\frac{\pi\beta_2 X m}{m+X} < 0 \end{aligned}$$

if $X > 0$. It follows that the parabola has a negative y-intercept. An upward-facing parabola with a negative y-intercept can only have a single positive x-intercept, regardless of the sign of b_1. It follows that there is a unique positive endemic equilibrium given by

$$\bar{X} = \frac{-b_1 + \sqrt{b_1^2 - 4a_1 c_1}}{2a_1}. \tag{13}$$

If $X = 0$, we have $c_1 = 0$. Hence the endemic equilibrium collides with the DFE at this point.

3.2 Basic Reproduction Number

Using the next-generation method, R_0 is defined to be the largest eigenvalue of the matrix FV^{-1}, with F representing newly arising infections in the system and V the balance of transfers of existing infections between the classes. Thus

$$F = \begin{bmatrix} \beta_1(1-\eta)\frac{\pi}{\mu} & \beta_1\frac{\pi}{\mu} & \beta_1\frac{\pi}{\mu} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } V = \begin{bmatrix} (\mu + \gamma + \Omega) & 0 & 0 \\ -f\gamma & (\nu_X + \alpha) & -\sigma \\ -(1-f)\gamma & 0 & (\nu_Y + \sigma + \zeta) \end{bmatrix}.$$

We find the reproduction number for leprosy with stigma to be:

$$\begin{aligned} R_0 = &\beta_1 \frac{\pi}{\mu} \frac{(1-\eta)}{(\mu + \gamma + \Omega)} + \beta_1 \frac{\pi}{\mu} \frac{\sigma(1-f)\gamma + f\gamma(\nu_Y + \sigma + \zeta)}{(\mu + \gamma + \Omega)(\nu_X + \alpha)(\nu_Y + \sigma + \zeta)} \\ &+ \beta_1 \frac{\pi}{\mu} \frac{(1-f)\gamma}{(\mu + \gamma + \Omega)(\nu_Y + \sigma + \zeta)}. \end{aligned}$$

These three terms represent the contributions from asymptomatic, unstigmatised and stigmatised individuals, respectively.

4 Numerical Simulations

To assess the effects of stigma, we considered three regimes: (a) moderate stigma, (b) high stigma, and (c) no stigma. We used sample values chosen within the ranges from the table and varied the effect of stigma using the disclosure rate σ to assess moderate versus high stigma regimes and turned off the proportion of individuals entering the stigma compartment for the no-stigma case.

Using recent leprosy-prevalence statistics [4], we conduct approximate order-of-magnitude estimates of our infection coefficients. Roughly 214,000 cases of leprosy were reported in 2014, with approximately 81% of those cases coming from Brazil, India and Indonesia. Approximating the populations of these nations by 200 million, 1.25 billion and 250 million, respectively, we estimate the coefficient of infection to be

$$\beta_1 = \frac{0.81 \times 214000}{1700000000} = 0.000101965 \approx 0.0001. \tag{14}$$

We considered a small village of 1000 individuals. Initial conditions were chosen so that there were 1000 susceptibles and a single infected non-stigmatised individual. Figure 2 illustrates the case of moderate stigma, showing an infection wave and a substantial number of uninfected individuals ($S + R = 636$ at the end of this simulation).

Next, we used the same parameters and initial conditions as in Fig. 2 except that we changed the rate of disclosure from $\sigma = 1$ to $\sigma = 0.1$. This reflects the

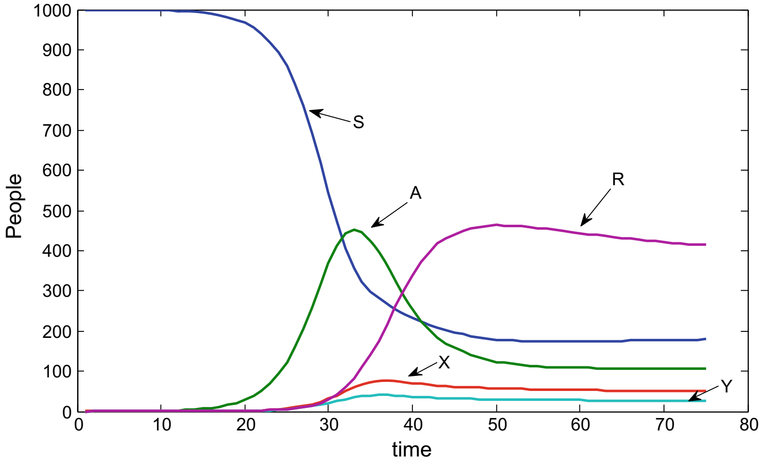


Fig. 2 Leprosy modelled in a moderate stigma regime. There is an infection wave, with an endemic disease outcome but a substantial number of recovered individuals. Parameters used were $\alpha = 1$; $\zeta = 0.1$; $\beta_1 = 0.01$; $\beta_2 = \beta_1/2$; $\gamma = 0.2$; $\delta = 0.1$; $\mu = 0.03$; $\nu = 0.09$; $\pi = 30$; $\sigma = 1$; $\Omega = 1$; $f = 0.5$; $m = 0.5$; $\eta = 0.8$. Note in particular that stigmatised individuals remain so for $\frac{1}{\sigma} = 1$ year

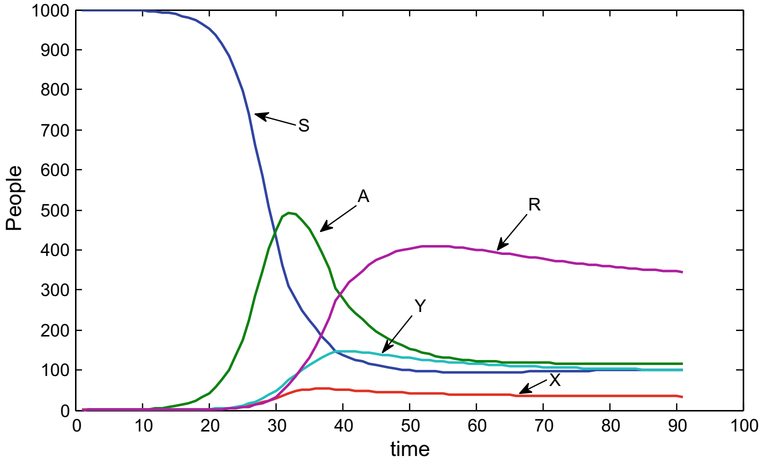


Fig. 3 Leprosy modelled in a high stigma regime. Parameters were as in Fig. 2 except that $\sigma = 0.1$ (i.e., stigmatised individuals remain so for an average of ten years). The number of recovered individuals is low, while stigmatised individuals persist. Note the re-ordering of the final outcome, compared to Fig. 2

case where individuals remain stigmatised for ten years (since the length of time remaining in a compartment is inversely proportional to the rate of leaving it). In this case, the number of stigmatised individuals exceeds the number of non-stigmatised or asymptomatic individuals, sustaining a high level of infected individuals. The number of uninfected individuals was significantly lower than in Fig. 2 ($S + R = 376$ in this simulation).

Finally, we modelled the case of no stigma. Parameters were as in Fig. 2 except that $f = 1$ to ensure that no individuals entered the stigma compartment. The outcome is similar to Fig. 2 except that there are slightly more uninfected individuals ($S + R = 729$ at the end of this simulation).

Additionally, we numerically explored the dependency of the results on the parameters β_2 and m. However, although these had an effect on the shape of the curves, they did not produce outcomes significantly different to Figs. 2, 3 and 4. (Results not shown.)

We also performed a sensitivity analysis on the reproduction number using Latin Hypercube Sampling (LHS) and Partial Rank Correlation Coefficients (PRCCs). Latin Hypercube Sampling is a statistical sampling method that evaluates sensitivity of an outcome variable to all input variables. PRCCs measure the relative degree of sensitivity to each parameter, regardless of whether the parameter has a positive or negative influence on the outcome variable [2].

PRCCs were calculated for the model and are displayed in Fig. 5 using the ranges from the table but with the range for β_1 extended to $0 \leq \beta_1 \leq 0.005$ to illustrate a wider outcome. This analysis provides a way to measure the sensitivity of a model to each parameter it contains. Figure 5 shows that β_1 and η are the most sensitive

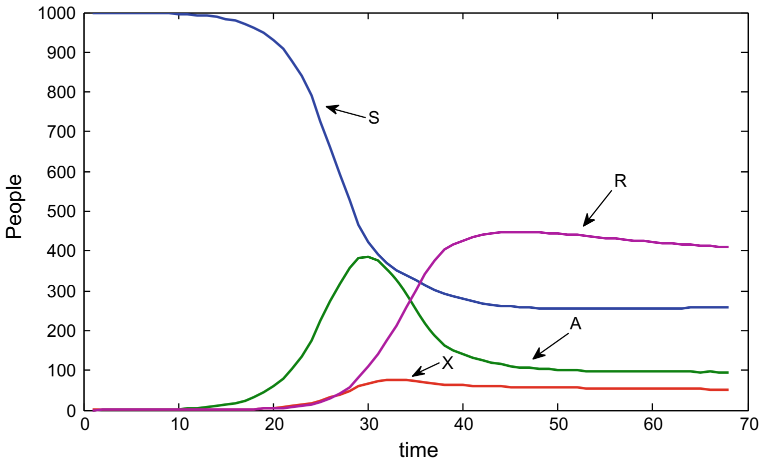


Fig. 4 Leprosy modelled in a stigma-free regime. Parameters were as in Fig. 2 except that $f = 1$ so that there were no stigmatised individuals. The number of infected cases is low, and the number of recovered individuals is high

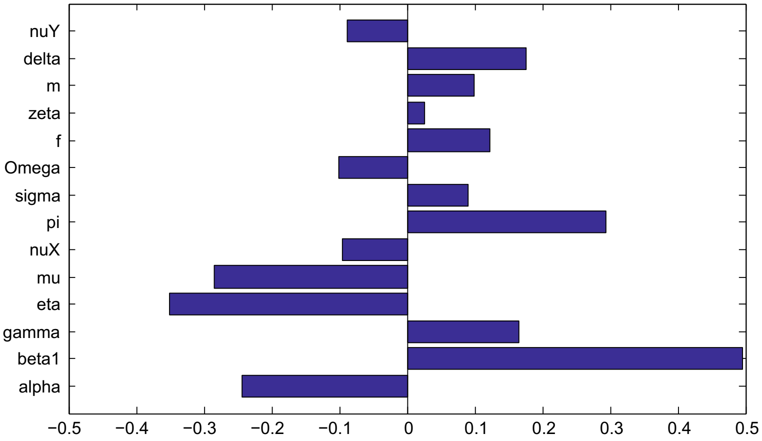


Fig. 5 Partial rank correlation coefficients for the model

parameters in the model. Note that the stigma parameters β_2 and m do not affect R_0, which is a measure of initial disease invasion.

In Fig. 6, we plot the individual Monte Carlo simulations of the two most sensitive model parameters. For $0 \leq \beta_1 \leq 0.5 \times 10^{-3}$, eradication will result, regardless of the values of the other parameters (i.e., if the disease transmission is extremely low). However, there is no value of η that can guarantee eradication; even if $\eta = 1$, fluctuations in the other parameters could still maintain the epidemic. Note that the threshold value of β_1 is larger than the value found in the literature using (14). This

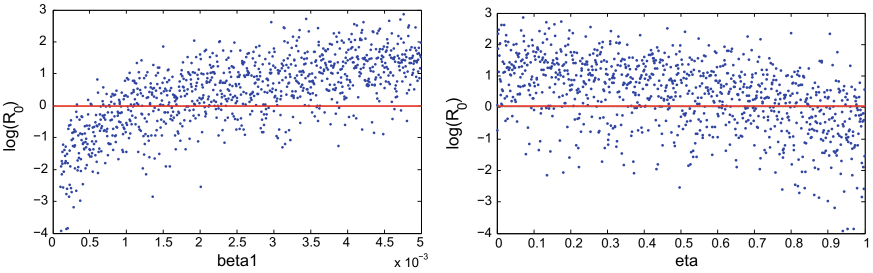


Fig. 6 Monte Carlo simulations for the two most sensitive model parameters, β_1 and η. The $R_0 = 1$ threshold is indicated by the horizontal line. Eradication can be achieved if β_1 is reduced below 0.0005. However, no value of η can guarantee eradication

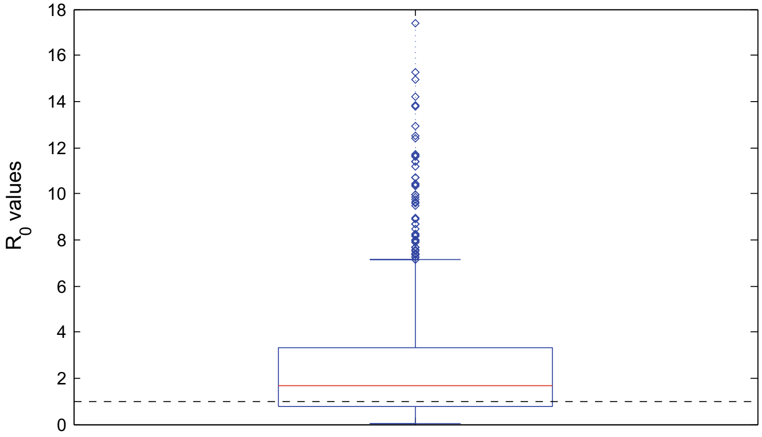


Fig. 7 Box plot of the range of R_0 values from Monte Carlo simulations

suggests that eradication is theoretically possible, but fluctuations in the parameters could have a significant effect on the outcome.

Figure 7 shows the complete range of R_0 values for all simulations. While the interquartile range crosses the $R_0 = 1$ threshold, suggesting that eradication is theoretically possible, the outlier values are quite extreme, suggesting that fluctuations in the parameter values could lead to significant epidemics. Note that this figure uses the extended range $0 \leq \beta_1 \leq 0.005$.

Since the endemic equilibrium can be uniquely determined from (13), we used the LHS method to determine the effects of stigma on the outcome. We ran Monte Carlo simulations on the endemic equilibrium $\bar{X}$ and then applied the ratio of stigma to non-stigma cases using (10). The outcome is illustrated in the ordered scatterplot in Fig. 8. The blue dots (lower half of graph) illustrate the number of simulations where there were fewer stigma cases, while the red circles (upper half of graph) illustrate the number of simulations where there were more stigma cases. In this example,

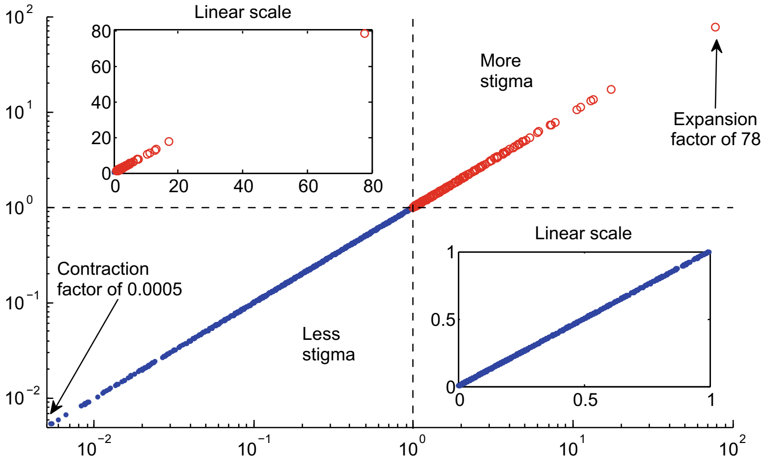


Fig. 8 Ordered scatterplot of the ratio of more-stigma versus less-stigma cases using LHS on the endemic equilibrium values on a log-log scale using ranges from the table. The red circles (upper right) are the Monte Carlo simulations where there are more stigma cases at equilibrium. The blue dots (lower left) are the simulations where there is less stigma at equilibrium. The inset graphs illustrate the two cases in a linear scale. Note that, at the extreme, there are 78 times as many stigma cases as non-stigma cases

there were 796/1000 cases where less stigma occurred and 204/1000 cases where more stigma occurred. However, although there were more cases where stigma was reduced, the degree of stigma expansion in the more-stigma case was extensive. At the extremes, the ratio of less-stigma to more-stigma cases ranged from a factor of 0.000545 to a factor of 78.

5 Discussion

We examined the effects of stigma on leprosy, both in the short term of initial disease outbreak (Figs. 5 and 6) and in the long term (Figs. 2, 3 and 4). Although stigma had no effect on initial disease invasion or eradication, differences in stigma levels had the potential to substantially alter the overall prognosis of leprosy in the population. To the best of our knowledge, this is the first mathematical model of leprosy to incorporate stigma.

Additionally, we used an ordered scatterplot to examine the ratio of increased stigma to decreased stigma across a range of parameter values. Although there were more cases where the amount of stigma was reduced, in the cases where it was increased, the result could be as much as an 80-fold increase in the number of stigmatised individuals. It follows that stigma is an important factor in the spread of leprosy.

Our model has some limitations, which should be acknowledged. We conflated "disclosure" and "stigma", whereas in practice the social phenomenon of leprosy-related stigma is more complex. We represent stigma by way of parameter choices, which is a simplification. Leprosy is an extremely heterogeneous disease in at least two major senses: it is heterogeneous with respect to those at risk of exposure to the *M. Leprae* bacteria and with respect to the spatial distribution of the disease. Therefore, given this heterogeneity, mass-action disease transmission may not be suitable for modelling this disease. We have also assumed constant birth and death rates, but, for any long-term disease, modelling time-varying birth and death rates may be more appropriate. Likewise, a further refinement would be to model the stigma parameters as randomly time-varying functions. Future work will examine the effect of TB vaccines against leprosy, which have been shown to be efficacious [18].

We thus see that stigma, whether moderate or high, plays a significant role in sustaining leprosy. Our sensitivity analysis showed that R_0 tends to range from 1 to 18 for typical model parameter ranges (Fig. 7). In practice, we may never have direct control over the transmission rate β_1, yet we may be able to influence σ, the rate at which non-disclosing symptomatic infectees either change their minds and disclose their infection or are discovered. Figure 3 shows that reducing this rate is critical.

In practice, leprosy-eradication strategies should focus on the reduction of leprosy-related stigma through a combination of targeted education about the disease and shifts in cultural attitudes towards leprosy. However, since the measure of stigma estimated in this model may conflate actual disease-related stigma with other important factors, such as knowledge or access to care, this model also highlights the importance of continuing to make leprosy MDT accessible while simultaneously educating at-risk populations about the possibility of such care. Finally, depending on how likely it is that asymptomatic infections can in turn generate new infections, such strategies may need to be supplemented by continued efforts to detect sub-clinical infections, which has been emphasised in the literature [17]. The persistence of leprosy is a complex problem; therefore a simple solution for its eradication is likely not possible; however, any approach incorporating the reduction of leprosy-related stigma will likely go a long way towards true leprosy eradication.

Acknowledgements The authors are grateful to Alison Kealey and Nyuk Sian Chong for invaluable assistance and are also grateful to an anonymous reviewer for insightful comments that greatly improved the manuscript. RS? is supported by an NSERC Discovery Grant. For citation purposes, please note that the question mark in "Smith?" is part of the author's name.

References

1. Blok, D., de Vlas, S., Fischer, E., Richardus, J.: Mathematical modelling of leprosy and its control. Adv. Parasitol. **87**, 33–51 (2014)
2. Blower, S., Dowlatabadi, H.: Sensitivity and unvertainty analysis of complex models. Int. Stat. Rev. **62**, 229–243 (1994)

3. Cross, H., Choudhary, R.: STEP: an intervention to address the issue of stigma related to leprosy in Southern Nepal. Lepr. Rev. **76**, 316–324 (2005)
4. Dara, S., Gadde, R.: Epidemiology, prognosis, and prevention of leprosy worldwide. R. Curr. Trop. Med. Rep. (2016)
5. Fischer, E., Vlas, D., Meima, A., Habbema, D., Richardus, J.: Different mechanisms for heterogeneity in leprosy susceptibility can explain disease clustering within households. PLoS ONE **5** (2012)
6. Kealey, A., Smith?, R.: Neglected tropical diseases: infection, modelling and control. J. Health Care Poor Underserved **21**, 53–69 (2010)
7. Kumaresan, J., Maganu, E.: Socio-cultural dimensions of leprosy in North-Western Botswana. Soc. Sci. Med. **39**, 537–541 (1994)
8. Lechat, M., Misson, C., Lambert, A.: Simulation of vaccination and resistance in leprosy using an epidemiometric model. Int. J. Lepr. **53**, 461–467 (1985)
9. Link, B.G., Phelan, J.C.: Conceptualizing stigma. Annu. Rev. Sociol. **27**, 363–385 (2001)
10. Lockwood, D., Suneetha, S.: Leprosy: too complex a disease for a simple elimination paradigm. Bull. World Health Organ. **83**, 230–235 (2005)
11. Lustosa, A., Nogueira, L., Pedrosa, J., Teles, J., Campelo, V.: The impact of leprosy on health-related quality of life. Revista da Sociedade Brasileira de Medicina Tropical **44**, 621–626 (2011)
12. Meima, A., Gupte, M., Oortmarssen, G.V., Habbema, J.: Simlep: a simulation model for leprosy transmission and control. Int. J. Lepr. Other Mycobact. Dis. **67**, 215–236 (1999)
13. Merle, C., Cunha, S., Rodrigues, L.: BCG vaccination and leprosy protection: review of current evidence and status of BCG in leprosy control. Expert. Rev. Vaccines **9**, 209–222 (2012)
14. Mg, W.: Stigma and the social burden of neglected tropical diseases. PLoS Negl. Trop. Dis. **2**, e237 (2008)
15. Mushuyabasa, S., Bhunu, C.: Modelling the effects of chemotheraphy and relapse on the transmission dynamics of leprosy. Math. Sci. **6**, 12 (2012)
16. Rafferty, J.: Curing the stigma of leprosy. Lepr. Rev. **76**(2), 119–126 (2005)
17. Richardus, J., Habbema, J.: The impact of leprosy control on the transmission of M. leprae: is elimination being attained? Lepr. Rev. **78**(4), 330–337 (2007)
18. Rodrigues, L., Lockwood, D.: Leprosy now: epidemiology, progress and research gaps. Lancet Infect. Dis. **11**(6), 464–470 (2011)
19. Sengupta, S., Banks, B., Jonas, D., Miles, M.S., Smith, G.C.: HIV interventions to reduce HIV/AIDS stigma: a systematic review. AIDS Behav. **15**, 1075–1087 (2011)
20. Smith, W., Anderson, A., Withington, S., Van Brakel, W., Croft, R., Nicholls, P., Richardus, J.: Steroid prophylaxis for prevention of nerve function impairment in leprosy: randomised placebo controlled trial (tripod 1). Br. Med. J., 328 (2004)
21. Suzuki, K., Akama, T., Kawashima, A., Yoshihara, A., Yotstu, R., Ishii, N.: Current status of leprosy: epidemiology, basic science and clinical perspectives. Adv. Exp. Med. Biol. **582**, 22–33 (2012)
22. Tsutsumi, A., Izutsu, T., Islam, M.D.A., Amed, J.U., Nakahara, S., Takagi, F., Wakai, S.: Depressive status of leprosy patients in Bangladesh: association with self-perception of stigma. Lepr. Rev., 57–66 (2004)
23. van Brakel, W.H., Sihombing, B., Djarir, H., Beise, K., Kusumawardhani, L., Yulihane, R., Kurniasari, I., Kasim, M., Kesumaningsih, K.I., Wilder-Smith, A.: Disability in people affected by leprosy: the role of impairment, activity, social participation, stigma and discrimination. Glob. Health Action **5**, 18394 (2012)
24. Weiss, M., Ramakrishna, J., Somma, D.: Health-related stigma: rethinking concepts and interventions. Pyschology Health Med. **11**(3), 277–287 (2006)
25. World Bank: Crude birth rate for developing countries in Middle East and North Africa (2016)
26. Yang, L.H., Kleinman, A., Link, B.G., Phelan, J.C., Lee, S., Good, B.: Culture and stigma: adding moral experience to stigma theory. Soc. Sci. Med. **64**, 1524–1535 (2007)

One Simple Model—Various Complex Systems

Urszula Foryś

Abstract In this chapter we consider a simple system of two ODEs that could be used to describe various phenomena. Examples of these phenomena are presented. In general, we focus on two interacting agents, like two animal populations, two people or groups of people, two neuronal populations and so on. The system have the following structure: the first part of an equation describes the inner dynamics, while the second part is responsible for interactions. We consider two actors/agents having similar inner dynamics, as well as interaction function is similar for both of them. In the simplest case, when the inner dynamics is linear, the behavior of the system depends on the interaction functions. We discuss similarities and differences between the models with different interaction terms.

Keywords Ordinary differential equations · Mathematical modeling of interacting agents · Stability analysis

1 Introduction

Mathematical modeling of interacting species has a long history, starting from classic Lotka-Volterra model. Almost 100 years ago Lotka [3] proposed a system of two differential equations to show that oscillatory dynamics is possible for chemical reactions. The same model was considered by Volterra [9, 10] in the context of prey-predator interactions. Since that time many mathematical models of interacting species have been proposed and many text-books on that topic have been published; cf. e.g. [4, 7].

In this paper we would like to consider a model of two interacting agents having the same inner dynamics and the same interaction function which describes impacts of one of the agents into the other one. Let $x_i(t)$ describes the state of the ith agent at

U. Foryś (✉)
Faculty of Mathematics, Informatics and Mechanics, Institute of Applied Mathematics
and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland
e-mail: urszula@mimuw.edu.pl

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_6

time t. In terms of interacting animal species, x_i denotes density of the ith species, and therefore, $x_i \geq 0$; in terms of two interacting persons, x_i reflects the intensity of emotions of the ith person and $x_i \in \mathbb{R}$ in such a case; in terms of interacting neurons, it reflects an intensity of signals. Hence, it is obvious that building a model one needs to be careful about the range of applicability.

The general form of the model we consider reads

$$\begin{aligned} \dot{x_1} &= f_1(x_1) + c_1 g_1(x_1, x_2) =: F_1(x_1, x_2), \\ \dot{x_2} &= f_2(x_2) + c_2 g_2(x_1, x_2) =: F_2(x_1, x_2), \end{aligned} \tag{1}$$

where:

- $f_i(\cdot)$ describes the inner dynamics of the ith agent;
- $c_1 g_1(\cdot, \cdot)$ describes the impact of the first agent into the second one;
- $c_2 g_2(\cdot, \cdot)$ describes the impact of the second agent into the first one.

Parameters c_i are introduced in order to keep as similar as possible forms of interaction functions g_i. These parameters reflect the strength and direction of interactions between agents. Positive values of c_1 mean positive influence of the second agent into the first one and vice versa.

For the inner dynamics we assume that:

A $f_i : \mathbb{R} \to \mathbb{R}$ (or we restrict the domain and values of f_i to the non-negative values $\mathbb{R}_+$, depending on the model);
B f_i is of class $\mathbf{C}^1$;
C the equation $\dot{x} = f_i(x)$ has exactly one globally stable steady state.

Notice that B is a technical assumption guaranteeing existence and uniqueness of solutions, while C says that in the absence of interactions (that is when an agent is in solitude) the agent remains in its steady state and came back to this state after any disturbance. Assumption C does not mean that f has a unique steady state, as it could also have zero steady state being unstable, and then the positive steady state is globally stable in positive quadrant. However, this is related to positive values of the variables x_i, so if the variables could have different signs, the globally stable steady state is unique.

The simplest function f fulfilling Assumptions A-C is just a linear function

$$f(\zeta) = a - b\zeta, \quad b > 0, \tag{2}$$

which has the steady state $\bar{\zeta} = \frac{a}{b}$ and $\bar{\zeta} > 0$ for $a > 0$. The parameter b is a time scale and this time scale reflects the time needed to came back to the inner steady state after a disturbance. The linear function could be used both when we need to restrict x_i to non-negative values and when it is not the case. Another commonly used function f is logistic one:

$$f(\zeta) = r\zeta\,(1 - b\zeta)\,, \quad r,\ b > 0. \tag{3}$$

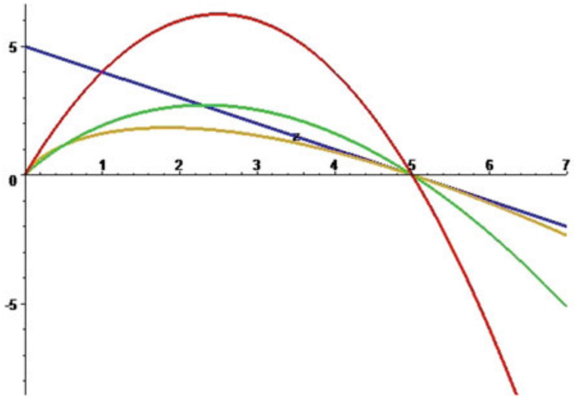


Fig. 1 Graphs of several possible functions f_i describing the inner dynamics of agents for which $\dot{x} = f_i(x)$ have the same globally stable steady state (here it is equal to 5): linear (blue curve), logistic (red curve), generalized logistic with $\alpha = 2/3$ (green curve), Gompertz (brown curve)

However, this function can be applied only for $x_i \geq 0$. Here, r is the growth rate, which again could be interpreted as a time scale, b is the competition rate, while $K = \frac{1}{b}$ is the steady state from Assumption C.

In this paper we mainly focus on the inner dynamics described by Eq. (2) or Eq. (3), but many other functions fulfilling A-C could be used, e.g. generalized logistic: $f(\zeta) = r\zeta\,(1 - (b\zeta)^{\alpha})$; Gompertz: $f(\zeta) = -r\zeta\,\ln(b\zeta)$ and others; cf. Fig. 1.

As regards the interaction functions g_i we also assume that they are of class $\mathbf{C}^1$, while more specific properties will be discussed in the context of specific models we would like to present here. However, in completely symmetric case we assume $g_1(x_1, x_2) = g_2(x_2, x_1)$, while in general these functions could have different coefficients.

2 Lotka-Volterra Models of Interacting Species

In this section we recall classic models describing interacting species apart the original Lotka-Volterra predator-prey model, that is we focus on the case with logistic f_i and bilinear $g_i(x_1, x_2)$, namely $g_i(x_1, x_2) = x_1 x_2$. Notice that we can distinguish two types of interactions:

1. competing species, when both $c_i < 0$;
2. mutualizm, when both $c_i > 0$.

It is also possible to consider different signs of c_i, that is $c_1 c_2 < 0$. However, this would rather correspond to predator-prey interactions, and it is worth to notice that for this type of interactions the inner dynamics of two agents could be different. Hence, this model is out of our scope.

2.1 Analysis of the Models of Interacting Species

First notice, that in both cases solutions are positive for positive initial data, while in the cases listed above:

1. for each of the species we have $\dot{x}_i \le r_i x_i(1 - b_i x_i)$, meaning that the set $\{(x_1, x_2) \in \mathbb{R}^2 : 0 \le x_i \le K_i\}$ ($K_i = 1/b_i$ is the carrying capacity for the ith species) is positively invariant, implying existence of solutions for all $t \ge 0$;
2. solutions can tend to ∞, depending on the parameters.

Notice also that we are able to scale both variables by b_i obtaining exactly the same model but with $b_i = K_i = 1$ and c_i scaled accordingly. Hence, we assume $b_i = 1$ for $i = 1, 2$.

Next, we focus on the existence and stability of steady states. There can be up to 4 steady states. Clearly, (0, 0), (0, 1) and (1, 0) always exist, while existence of a positive steady state depends on the values of the model parameters. This state satisfies the following system of equations:

$$r_i(1 - x_i) + c_i x_j = 0 \quad \text{for} \quad i, j = 1, 2, \quad i \neq j.$$

Clearly, linear function $x_2 = \frac{r_1}{c_1}(x_1 - 1)$ is the null-cline for the first variable and $x_2 = 1 + \frac{c_2}{r_2}x_1$ is the null-cline for the second one and they have a cross section in $\mathbb{R}^2_+$:

1. if $|c_i| > r_i$ or $|c_i| < r_i$ for both $i = 1, 2$;
2. if $r_1 r_2 > c_1 c_2$.

Using the Dulac-Bendixson Criterion it is easy to check that the models have no periodic orbits in positive quadrant. Clearly, let define $B(x_1, x_2) = \frac{1}{x_1 x_2}$ and calculate the divergence of the vector field (BF_1, BF_2). We obtain

$$\frac{\partial}{\partial x_1}\left(\frac{r_1(1 - x_1)}{x_2} + c_1\right) + \frac{\partial}{\partial x_2}\left(\frac{r_2(1 - x_2)}{x_1} + c_2\right) = -\frac{r_1}{x_2} - \frac{r_2}{x_1} < 0 \text{ for } x_1, x_2 > 0.$$

Thus, we can conclude that if solutions remain in the bounded region, then any solution tend to one of the steady states. This implies the following dynamics:

1. either there is exactly one stable steady state and it is globally stable, or there are two stable steady states and we observe bi-stability;
2. either there exists a positive steady state and it is globally stable, or there is no such state and solutions are unbounded.

2.2 *Illustration of the Interacting Species Models Dynamics*

In this subsection we present phase space portraits illustrating possible dynamics of the models of interacting species.

Let us first consider competing species. Figure 2 presents four exemplary portraits.

(I) In this case the inequality $r_2/|c_2| > 1 > r_1/|c_1|$ is satisfied. There is no positive steady state and the only stable steady state $(1, 0)$ attracts all solutions in the positive quadrant. Competition leads to the extinction of the second species, while the first species tends to its carrying capacity.

(II) This case is symmetric to the previous one: $r_2/|c_2| < 1 < r_1/|c_1|$ and all solutions are attracted by $(0, 1)$.

(III) In this case the inequalities $r_i/|c_i| < 1$, $i = 1, 2$, are satisfied. Competition is not so harmful for both species and they coexist in the environment.

(IV) This case is reverse to the previous one: $r_i/|c_i| > 1$, $i = 1, 2$, and competition is harmful for both species. The species with better initial condition wins the competition and tends to its carrying capacity, while the other goes to

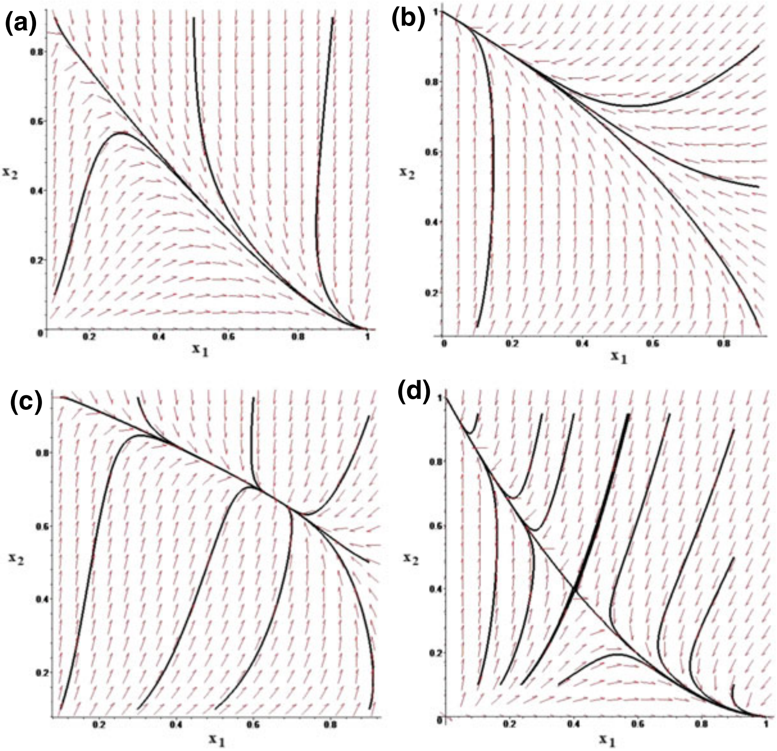


Fig. 2 Phase space portraits for competing species and various possibilities of the model dynamics: **a** $r_2/|c_2| > 1 > r_1/|c_1|$; **b** $r_2/|c_2| < 1 < r_1/|c_1|$; **c** $r_i/|c_i| < 1, i = 1, 2$; (III) $r_i/|c_i| > 1, i = 1, 2$

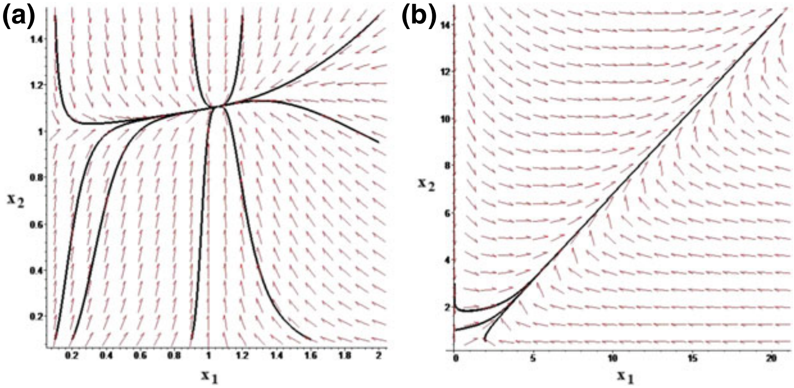


Fig. 3 Phase space portraits for mutualistic species and various possibilities of the model dynamics: **a** $r_1r_2 > c_1c_2$; **b** $r_1r_2 < c_1c_2$

extinction. The phase space is divided into the basins of attraction of the states $(1, 0)$ and $(0, 1)$, and the separatrix is formed from the stable manifold of the positive steady state which is a saddle.

Next, we turn to the case when the species interact in a mutualistic way. There are two possible phase space portraits presented in Fig. 3.

(I) In this case the inequality $r_1r_2 > c_1c_2$ is satisfied. There exists the positive steady state which attracts all solutions.

(II) In this case the inequality $r_1r_2 < c_1c_2$ is satisfied. There is no positive steady state and all solutions for positive initial data tend to infinity.

Notice that in all considered cases the semi-trivial steady states—$(1, 0)$ and $(0, 1)$—are either saddles or stable nodes. This is a simple consequence of the inner dynamics for which these states are stable. Moreover, both states are always saddles for mutualistic species, while for competing species all combinations (both saddles, both stable, one stable and one unstable) are possible. Moreover, dynamics of these two different systems is similar when there exists stable positive steady state. However, for all other cases competition leads to extinction of one of the species, which corresponds to the ecological rule of competitive exclusion.

3 Influence Function Depending on One Variable Only

In this section we would like to discuss System (1) of the following specific form:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1) + c_1 g_1(x_2), \\ \dot{x}_2 &= f_2(x_2) + c_2 g_2(x_1), \end{aligned} \tag{4}$$

which is related to the modeling of dyadic interactions. Modeling dyadic interactions using ODEs started with a short note by Strogatz [8] who noticed that fluctuations of emotions in the relationship of Romeo and Juliet could be described using a simple linear mathematical oscillator. However, it is obvious that humans' emotions are non-linear, and this was a reason to propose a model of the form (4) by Rinaldi and his coauthors; cf. [6] for details and the history of this model. Similar approach but in a little different context was used by Liebovitch et al. [2]. They used the notion of actors instead of partners to mark that they do not describe love relationships. Here we focus on the interpretation presented in our paper [5], which is similar to those by Liebovitch et al.

In System (4), the inner dynamics is described by linear functions $f_i(\zeta) = a_i - b_i\zeta$. In our interpretation (cf. [5]) coefficients a_i of the inner dynamics correspond to the natural optimism/pessimism of the ith actor. More precisely, when $a_i > 0$, then the ith actor is an optimist with the natural level of optimism equal to $\frac{a_i}{b_i}$ which is called uninfluenced steady state, while for $a_i < 0$ this steady state reflects the natural level of pessimism for the ith actor. The coefficient $b_i > 0$ is called a forgetting coefficient in this context.

Both functions g_i are the same, that is $g_1(\zeta) = g_2(\zeta) = g(\zeta)$ and g have the following properties:

(a) $g(0) = 0$;
(b) $g'(0) = 1$;
(c) g is increasing;
(d) $\zeta g''(\zeta) < 0$ for $\zeta \neq 0$.

Assumption (a) says that there is no influence if the second actor is absent. Assumption (b) is a normalization assumption for the derivative of g and together with (d) means that 1 is a maximal value. Assumption (c) reflects the fact that more intensive impact is related to greater values of the variable (that is greater intensity of emotions, greater density of the species, and so on). However each additional unit of the variable gives less profit, which is associated with natural saturation of emotions or other processes. Coefficients c_i reflect direction and strength of the impact of the other actor to the ith one. Clearly, if $c_1 > 0$, then the first actor likes the second one, and therefore the second actor has a positive impact to the first one and vice versa. In [5] we focused on the analysis who want to establish closer relationships with whom. We present some discussion on that topic below.

3.1 Analysis of System (4)

First notice that System (4) has at least one steady state. Clearly, steady states satisfy the following system of equations:

$$a_1 - b_1x_1 + c_1g(x_2) = 0, \quad a_2 - b_2x_2 + c_2g(x_1) = 0,$$

and the curve $x_2 = \frac{a_2+c_2g(x_1)}{b_2}$ is the null-cline for the second variable, which is monotonic and defined for every $x_1 \in \mathbb{R}$, while the curve $x_1 = \frac{a_1+c_1g(x_2)}{b_1}$ is the null-cline for the first variable which takes all values from $\mathbb{R}$. Due to Assumption (d) the curves $x_1 = g(x_2)$ and $x_2 = g(x_1)$ have a cross-section at zero, while the curves $x_1 = c_1g(x_2)$ and $x_2 = c_2g(x_1)$ could have two others, depending on the parameters c_i. This property is inherited by the null-clines, and this implies that System (4) has from one to three steady states.

It is worth to notice that there is no periodic orbits within solutions of System (4). Clearly, calculating the divergence of the right-hand side F we obtain

$$\frac{\partial}{\partial x_1}\left(a_1 - b_1x_1 + c_1g(x_2)\right) + \frac{\partial}{\partial x_2}\left(a_2 - b_2x_2 + c_2g(x_1)\right) = -\left(b_1 + b_2\right) < 0 \text{ in } \mathbb{R}^2.$$

Hence, the Dulac Criterion implies that periodic orbits do not exist, and therefore any solution remaining in a bounded region tends to a steady state. This implies that:

- if there is only one steady state then this state is globally stable in $\mathbb{R}^2$;
- if there are three steady states then we observe bi-stability, that is two of the three steady states are stable with a saddle between them and the stable manifold of the saddle divides the whole space into the basins of attraction of stable states;
- if there are two steady states then the saddle-node bifurcation is observed.

Notice that this behavior is similar to those observed for competitive species. However, there are situations in which we observe damping oscillations. It occurs that the more similar the agents are, the more probable the oscillatory dynamics is.

3.2 *Illustration of the Dynamics of System* (4)

Now, we turn to numerical examples of the model dynamics for various signs of coefficients a_i and c_i, $i = 1, 2$. Interpreting the results we follow the ideas of our paper [5] and focus on the relations between two actors depending on their attitude to life, that is their level of optimism/pessimism. In the numerical examples of the phase portraits presented in this subsection the interaction function g was chosen as arctan. Both forgetting coefficients were fixed, $b_i = 1$ and $|a_i| = 1$. Hence, the uninfluenced steady state is either 1 for an optimist or -1 for a pessimist. We want to check how these uninfluenced steady states change due to the interactions between actors. We say that the ith actor gains due to the interactions when the ith coordinate of the steady state of System (4) is greater than the uninfluenced steady state of this actor, while the ith actor loses in the opposite case.

We start our analysis from two actors being pessimists—three exemplary phase portraits are presented in Fig. 4.

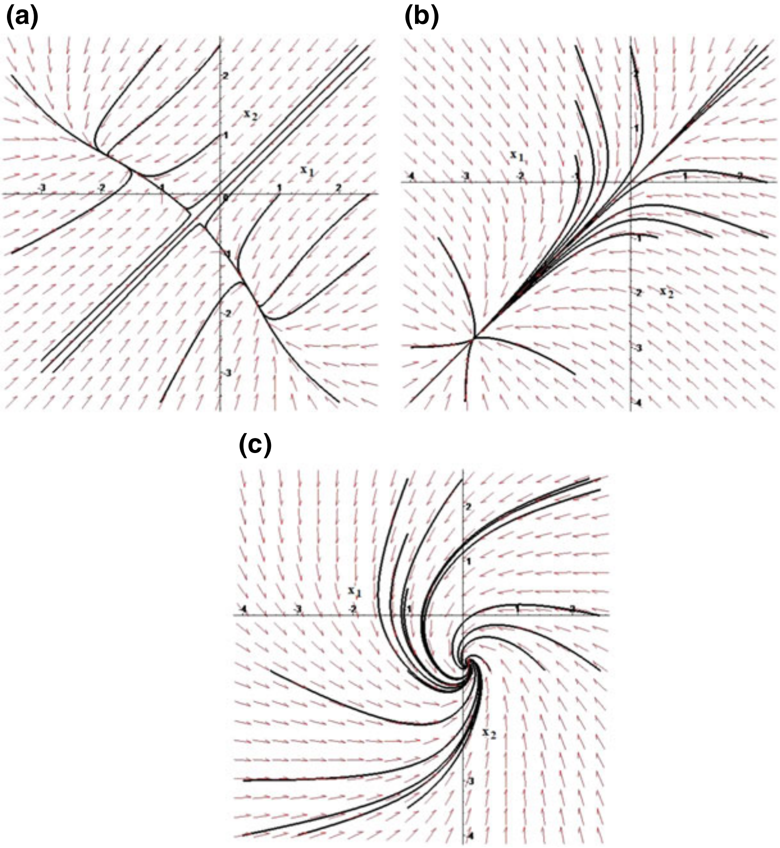


Fig. 4 Phase space portraits for System (4) and two pessimists. Various possibilities of the model dynamics: **a** $c_i < 0, i = 1, 2$; **b** $c_i > 0, i = 1, 2$; **c** $c_1 < 0, c_2 > 0$

(I) In this case the actors do not like each other ($c_i < 0$, $i = 1, 2$) and the nullclines cross three times yielding the existence of three steady states. We see that only in the stable manifold of a saddle both actors can gain due to the interactions, while for all other cases one of the actors loses and the other gains. If at the beginning the first actor is in a better mood than the second one, then this situation will be extended for all $t > 0$ and vice versa.

(II) In this case the actors like each other ($c_i > 0$, $i = 1, 2$) and there is one steady state. We see that both actors lose due to the interactions.

(III) In this case the first actor does not like the second one ($c_1 < 0$), while the second actor likes the first one ($c_2 > 0$). We observe oscillations resulting in the gaining of the first actor.

As we can see for a pessimist, it is difficult to be a friend of another pessimist, as being his friend we take over his emotions and get worse in his company.

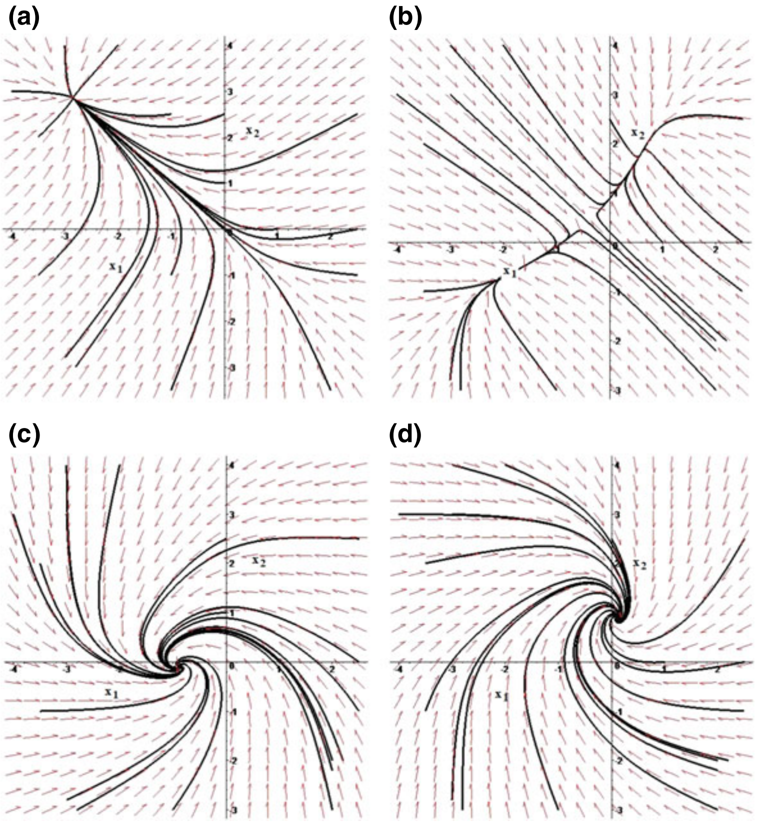


Fig. 5 Phase space portraits for System (4) and one pessimists and one optimist. Various possibilities of the model dynamics: **a** $c_i < 0, i = 1, 2$; **b** $c_i > 0, i = 1, 2$; **c** $c_1 < 0, c_2 > 0$; **d** $c_1 > 0, c_2 < 0$

Next, we consider relationships between a pessimist and an optimist. For the first actor we assume $a_1 = -1$, while for the second one $a_i = 1$. Various possibilities are presented in Fig. 5.

(I) In this case the first actor being a pessimist does not like ($c_1 < 0$) the second one who is an optimist and also does not like ($c_2 < 0$) the first one as well. We see that the first actor loses, while the second one gains.

(II) In this case both actors like each other ($c_i > 0, i = 1, 2$). If they are both in relatively good moods then they both gain. However, if their moods are relatively bad then both actors lose.

(III) In this case the first actor does not like ($c_1 < 0$) the second one, while the second actor likes ($c_2 > 0$) the first one. We see that the second actor loses.

(IV) In this case the first actor likes ($c_1 > 0$) the second one, while the second actor does not like ($c_2 < 0$) the first one. We see that the first actor gains.

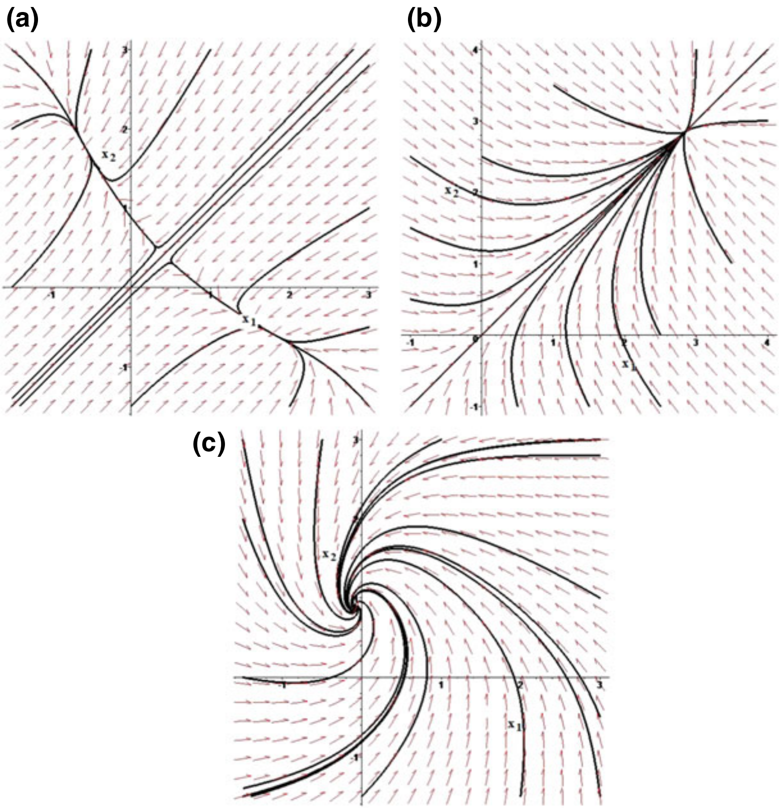


Fig. 6 Phase space portraits for System (4) and two optimist. Various possibilities of the model dynamics: **a** $c_i < 0, i = 1, 2$; **b** $c_i > 0, i = 1, 2$; **c** $c_1 < 0, c_2 > 0$

For relationships between a pessimist and an optimist it is not so obvious how the interactions will avoid the steady state. It is worth to notice that the above mentioned possibilities are only examples of the model dynamics, and the situation could change with changing parameter values; cf. [5] for details.

Lastly, we consider the relationships between two optimists. The exemplary phase space portraits illustrating the model dynamics are drawn in Fig. 6.

(I) In this case both actors do not like each other ($c_i < 0, i = 1, 2$). For the used parameters there are three steady states and depending on initial data one of the actors gains while the other loses. Only within stable curve for a saddle both of them lose.

(II) In this case both actors like each other ($c_i > 0, i = 1, 2$). This is the best case as both actors gain due to the interactions.

(III) In this case the first actor does not like ($c_1 > 0$) the second one, while the second actor likes ($C_2 > 0$) the first one. We see that the first actor lose.

Notice that for the cases when there are three steady states for System (4) other types of the model dynamics are also possible. Clearly, the number of steady states depends on the parameter values, so there can be also two or one steady state and except bi-stability (sadle-node bifurcation with two steady states can be also interpreted as bi-stability) global stability is also possible. Many interesting conclusions could be drawn on the basis of this simple model. For example, basing on changes of strategy (that is changing sign of the coefficient c_i) one can explain so called Stockholm syndrom; cf. [5] for details.

4 Influence Function of the Form $x_i g(x_i x_j)$

In this section we consider the model proposed in the context of perceptual decision making in [1]. In that paper we focused on modeling the most basic perceptual decision-making in neuronal networks in which the network needs to disambiguate between two sensory stimuli. We described changes in firing rates x_1, x_2 of two neuronal populations. In the simplest version without time delay this model reads

$$\begin{aligned} \dot{x}_1 &= \alpha_1\Big(a_1 - b_1 x_1 + c x_2 g(x_1 x_2)\Big), \\ \dot{x}_2 &= \alpha_2\Big(a_2 - b_2 x_2 + c x_1 g(x_1 x_2)\Big), \end{aligned} \tag{5}$$

where $\frac{1}{\alpha_i}$ are time scales for these neuronal populations, the ith population receives an external input a_i, our populations are self-inhibited with inhibition coefficients b_i, c is a coefficient describing the maximal capacity of a synapse, while g characterizes interactions between neuronal populations related to so-called synaptic plasticity for which we assumed a sigmoid shape. In this model we are interested in positive values of the variables x_i and this is the reason we define the function g only for non-negative values. However, in general the model could be extended for all real values of x_i and then the function g should be considered as a function on $\mathbb{R}$.

We assume that $g : \mathbb{R}_+ \to \mathbb{R}_+$ of class $\mathbf{C}^1$ satisfies the following assumptions:

- $g(0) = 0$;
- $g'(\zeta) > 0$ for $\zeta > 0$;
- there exists $\zeta_1 > 0$ such that in $(0, \zeta_1)$ the function g is convex and for $\zeta > \zeta_1$ it is concave;
- g is bounded.

4.1 Analysis and Illustration of the Model Dynamics for the Case Considered by Foryś et al. [1]

In [1] we assumed g in the form of Hill function with Hill coefficient equal to 2, that is $g(\zeta) = \frac{\zeta^2}{1+\zeta^2}$. Moreover, without additional input we considered the model being fully symmetric, meaning that all coefficients are the same with reference values $\alpha_i = 3$, $a_i = 0.4$, $b_1 = 1$, $c = 1$. The main aim of the analysis presented in [1] was to check if the model is able to correctly predict to which neuronal population an additional input was applied.

Looking at the model dynamics for the basic symmetric case it is easy to see that the phase space portrait is symmetric, as exchanging the variables $x_1 \leftrightarrow x_2$ does not change the model. Moreover, the straight line $x_2 = x_1$ is a specific trajectory on which the following equation is satisfied:

$$\dot{\zeta} = \alpha\Big(a - \zeta + \zeta g\left(\zeta^2\right)\Big) = 3\left(0.4 - \frac{\zeta}{1+\zeta^2}\right). \tag{6}$$

Due to the properties of the function $\frac{\zeta}{1+\zeta^2}$ (which is unimodal with one maximum equal to 0.5 at $\zeta = 1$ we see that there are two steady states of Eq. (6), $0 < \zeta_1 < 1 < \zeta_2$ and ζ_1 is stable while ζ_2 is unstable. Within the straight line $x_2 = x_1$ all solutions with initial value below ζ_2 tends to ζ_1, while those above ζ_2 tends to ∞ as $t \to \infty$. These corresponds to the dynamics in the whole phase space where (ζ_1, ζ_1) is a stable node while (ζ_2, ζ_2) is a saddle. The stable curve of a saddle divides the phase space into two regions: below this curve solutions are attracted by the first steady state, while above this curve solutions tend to ∞. It should be noticed that we are interested in the model dynamics for the values of variables near the first steady state, as large values are not biologically plausible. Hence, in Fig. 7 we present only a part of the phase space corresponding to such values.

In Fig. 7a we see our reference symmetric case with all solutions attracted by the steady state with coordinates near 0.4. Giving additional input to the first variable results in moving this state to the right (cf. Fig. 7b), while when the additional input is given to the second variable, then the steady state moves up. This shows that the system correctly recognizes to which population this additional input is applied.

In [1] we presented an extended analysis of System (5) also for parameter values different than reference values, especially focusing on the role of parameters a and c (denoted by I and ϵ in the original paper) for the number of steady states. It occurs that there can be from one to three steady states, similarly to other models considered in this chapter. The stability of these states is also similar to the cases studied above in previous sections.

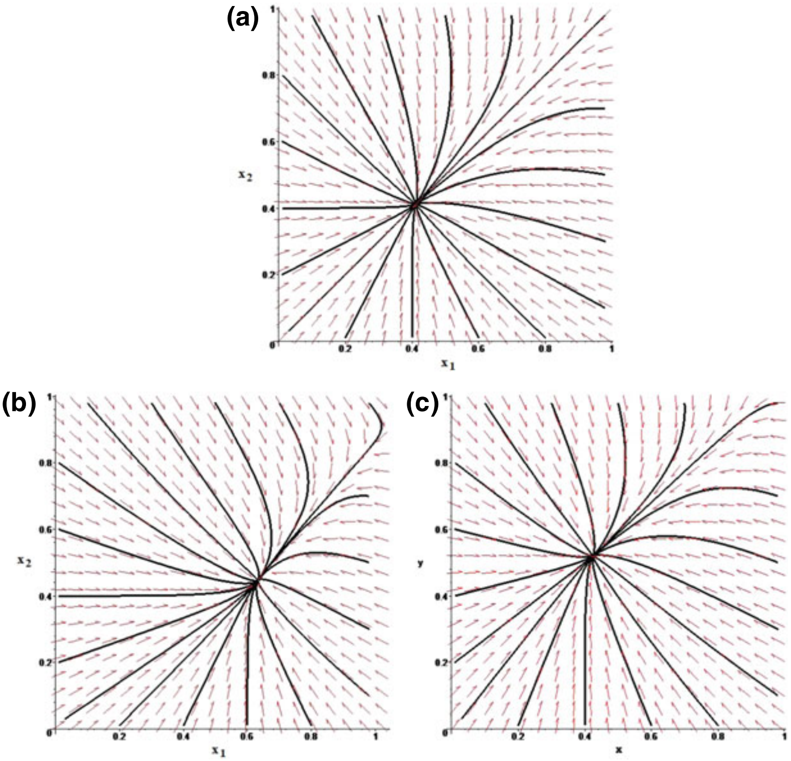


Fig. 7 Phase space portraits for System (5): **a** symmetric case; **b** for additional input given to the first population; **c** for additional input given to the second population

5 Conclusions

The main aim of the chapter was to show that simple systems of ODEs are capable to describe complex processes appearing in nature. We have discussed three examples of general System (1) reflecting interactions between two animal species, between two persons during a meeting and two neuronal populations acting in perceptual decision making. For all the examples from one to three steady states appear. If there is only one steady state, then it is typically globally stable, while for three steady states we observe bi-stability. These two type of the behavior are separated by saddle-node bifurcation. In the case of mutualistic species solutions can be unbounded when there is no positive steady state. Although the model is able to reflect some types of the dynamics observed in nature, but it should be noticed that periodic dynamics does not appear for all ODEs presented in this paper. Hence, to reflect such type of the behavior some modifications are necessary. One of the possibilities is to include time delay into the model. Clearly, in [1] we introduced time delay in self-inhibition and obtained periodic dynamics which is able to reflect ambiguity in perceptual decision making.

References

1. Foryś, U., Bielczyk, N., Piskała, K., Płomecka, N., Poleszczuk, J.: Impact of time delay in perceptual decision-making: neuronal population modeling approach. Complexity Article ID 4391587 (2017)
2. Liebovitch, L., Naudot, V., Vallacher, R., Nowak, A., Biu-Wrzosinska, L., Coleman, P.: Dynamics of two-actor cooperation-competition conflict models. Physica A **387**, 6360–6378 (2008)
3. Lotka, A.J.: Undamped oscillations derived from the law of mass action. J. Amer. Chem. Soc. **42**, 1595–1599 (1920)
4. Murray, J.D.: Mathematical Biology: I. An Introduction. Springer, Berlin (2002)
5. Piotrowska, M.J., Górecka, J., Foryś, U.: The role of optimism and pessimism in the dynamics of emotional states. Discret. Contin. Dyn. Syst. Ser. B **23**(1), 401–423 (2018)
6. Rinaldi, S., Rossa Della, F., Dercole, F., Gragnani, A., Landi P.: Modeling love dynamics. In: World Scientific Series on Nonlinear Science Series A, vol. 89. World Scientific Publishing Co. Pte. Ltd. (2015)
7. Smith, H.L.: Monotone dynamical systems; an introduction to the theory of competitive and cooperative systems. In: Mathematical Surveys and Monographs 41. American Mathematical Society, Providence (1993)
8. Strogatz, S.: Love affairs and differential equations. Math. Mag. **65**(1) (1988)
9. Volterra, V.: Variazionie fluttuazioni del numero d'individui in specie animali conviventi. Mem. Acad. Lincei. **2**, 31–113 (1926)
10. Volterra, V.: Variations and fluctuations of a number of individuals in animal species living together. Translation by R.N. Chapman. In: Animal Ecology, pp. 409–448. McGraw Hill, New York (1931)

Reliability Analysis of Multi-state Two-Dimensional System by Universal Generating Function

K. Meenakshi and S. B. Singh

Abstract In this paper, two dimensional multi-state non-repairable systems having *m* rows and *n* columns have been studied. Markov stochastic process has been applied for obtaining probabilities of the components. Reliability metrics such as reliability, mean time to failure and sensitivity analysis of the target system with the application of universal generating function are evaluated. Finally, the developed model is demonstrated with the help of a numerical example.

Keywords Multi-state system · Two-dimensional system · Reliability · Universal generating function · Mean time to failure · Sensitivity analysis

Notations

$R_{stm}(t)$	Reliability of system at time *t*
$\lambda^i_{s^*\omega}$	Failure rate of component *i* from state s^* to ω
$p_{ie}(t)$	Probability of component *i* in state *e*
D	Demands of system performance
$U_{stm}(z)$	UGF of system
$\varphi(U_{stm}(z), D)$	Function contain only those terms which have sum of the performances $\geq D$
E_i	*i*th components of system
I	Symbol for counting elements of $k \times l$ matrix.

K. Meenakshi (✉) · S. B. Singh
Department of Mathematics, Statistics and Computer Science, G.B. Pant University of Agriculture and Technology, Pantnagar 263145, India
e-mail: gariameenakshi86@gmail.com

S. B. Singh
e-mail: drsurajbsingh@yahoo.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_7

1 Introduction

A system is a set of mutually connected and dependent components which works together [16]. The systems are basically classified in two different categories according to their states viz. binary state systems (BSS) and multi-state systems (MSS). In BSS the system and its components can have two possible states: working and failed whereas in MSS, a system and its components can be in finite number of states. MSS reliability models allow both the system and its components to assume more than two or finite number of levels of performance. Further, if we inquire the engineering systems we see that they are represented more realistically and precisely through multi-state reliability models. Modern engineering systems are much more complex and present difficulties in system performance evaluation. Especially in today's real world problems with the development of science and technology and increasing amount of design of complex and interrelated system, the more number of system states needs to be considered. Moreover, increasing high demands for the accurate reliability evaluation, it is difficult to use traditional binary reliability techniques as it can't always insure the realistic requirements of the complex engineering systems. Therefore, the reliability evaluation of multi-state system is high in demand [13].

Now a days high definition engineering systems are emerging. The two dimensional (two-dim) systems are also increasingly significant complex systems used in engineering systems like temperature monitoring systems for chemical reactor, video monitoring systems etc. Before development of two-dim system, one dim consecutive k-out-of-n: F system is introduced [4] due to motivation of theory development and engineering application in reliability theory. Later, the consecutive k -out-of-n: F system firstly extended to two-dim situation in 1990s by Salvia and Lasher. The said, first two-dim system was k^2/n^2: F system which fails if k^2 $(2 \leq \mathrm{k} \leq n - 1)$ elements in sub-square of square containing n^2 elements are failed [19]. Similarly [2], propound linear connected (r, s)-out-of-(m, n): F system, which consists of components in rows and columns. This type of system experiences failure if at least r rows and s columns components are failed in the system.

There are many methods to analysis reliability of the multi-state systems, viz. stochastic process, Monte Carlo approach, Markov chain approach, universal generating function (UGF) etc. UGF approach is one of the best methods ever seen to analysis reliability of any system. Universal generating function was first emerged in 1986 [18], after that considerable research efforts have been presented by many authors [3, 8, 9, 11, 12, 15] not only in applying but also in extending this approach in their research works. As far as computation of two-dim system reliability is concerned various methods like Monte Carlo method, Markov chain embedding approach, YM algorithm etc. are proposed by different researchers. Although UGF approach is efficient, systematic and well defined for reliability analysis of any system but reliability analysis of two-dim system with help of Markov stochastic process [10, 14], and UGF is not yet done. Further, in their article, authors [6, 7, 15] investigated methods for analyzing sensitivity of the system corresponding to the component's states probabilities or failure rates. These investigated methods are based on Birnbaum importance measure [1], and also for measuring expected time to failure of the system which is known as mean time to failure (MTTF). It is worth mentioning here that although, it is a critical factor of system design but the said study is also not earlier conducted for two-dim repairable and non-repairable multi-state system with the application of UGF.

Keeping aforementioned facts in view, in the present study a two-dim non-repairable multi-state (*k*, *l*)/(*m*, *n*): G system has been taken. Markov stochastic process is applied for finding probabilities of multi-state components of the system corresponding to their performances. Then universal generating function approach has been applied to evaluate reliability indices such as reliability, MTTF of the system and sensitivity analysis of the system's components. Finally, a numerical example has been taken to demonstrate the proposed approach for the considered system.

2 Preliminaries

2.1 Definition

2.1.1 Multi-state Two-Dim (*k*, *l*)/(*m*, *n*): G System

A non-repairable multi-state two-dim (*k*, *l*)/(*m*, *n*): G system consists *m*, *n* components, which are arranged in matrix form ($m \times n$) where *m* and *n* are row and column respectively. The system works if sum of the performances of consecutive *kl* components of sub matrix ($k \times l$), where row *k* ($2 < k < m$) and column $l(1 < l < n)$ is greater than equal to demand performance *D*.

2.2 Universal Generating Function

UGF is a form of ordinary moment generating function which represents the probability mass function of discrete random variables. Let a discrete random variable X has M possible values $(x_1, x_2, \ldots, x_M)$ and $(p_1, p_2, \ldots, p_M)$ be the corresponding probabilities. Then UGF of discrete random variable $X = (x_1, x_2, \ldots, x_M)$ is represented by a polynomial

$$U(z) = \sum_{j=1}^{M} p_j z^{(X=x_j)}, \; j = 1, 2, 3, \ldots, M \tag{1}$$

Consider a system having $X_i (i = 1, 2, \ldots, n)$ components and every component has $v(v = 1, 2, \ldots, M_j)$ possible state denoted by x_{jv}. Let probability distribution of each variable x_{jv} (jth variable at vth state) is represented by p_{jv}. Then the UGF of random variable X_j is defined as

$$u_{X_j}(z) = \sum_{v=1}^{M_j} p_{jv} z^{x_{jv}} \tag{2}$$

Combination of m universal generating functions of m variables is defined by composition operator $\otimes_f$, where the properties of the composition operator $\otimes_f$ totally depend on properties of the function $f(X_1, X_2, \ldots, X_m)$. Therefore, the composition is given by

$$\begin{aligned} U(z) &= \otimes_f (u_{X_1}(z), u_{X_2}(z), \ldots, u_{X_m}(z)) \\ &= \otimes_f (\sum_{v_1=1}^{M_1} \sum_{v_2=1}^{M_2} \sum_{v_3=1}^{M_3} \cdots \sum_{v_m=1}^{M_m} (\prod_{i=1}^{m} p_{iv_i} z^{f(x_{iv_1}, \ldots, x_{iv_m})}) \end{aligned} \tag{3}$$

2.3 Probability Measure

Consider a system having $E_i (i = 1, 2, \ldots, l^*)$ component with corresponding probabilities $p_{ie}(e = 1, 2, \ldots, \omega)$. Every component ith has $\lambda^i_{s^*\omega}$ transition rate from state s^* to another lower state ω. If system's components are non-repairable then the probabilities corresponding to the system's components are evaluated as

$$\left.\begin{array}{l} \frac{dp_{i1}(t)}{dt} = -p_{i1}(t)\lambda^i_{12} \\ \frac{dp_{il^*}(t)}{dt} = \lambda^i_{l^*-1\,l^*} p_{i\,l^*-1}(t) - p_{il^*}(t)\lambda^i_{l^*l^*+1} \\ \vdots \\ \frac{dp_{i\omega}(t)}{dt} = \lambda^i_{\omega-1\,\omega} p_{i\,s^*}(t) \end{array}\right\} \tag{4}$$

2.4 Mean Time to Failure

Mean Time to Failure (MTTF) is an expected time to failure of a non-repairable system. If $R(t)$ is the reliability function of the random variable Y at time t, then MTTF is defined as

$$\text{MTTF} = \int_0^{\infty} R(t)dt \tag{5}$$

2.5 Sensitivity Analysis

Sensitivity analysis is a method used to identify most sensitive components of a system corresponding to overall system reliability. Let $p_{i\omega}(i = 1, 2, \ldots, m$ is number of components and $\omega = 1, 2, \ldots, n$ is possible state of the component) probability corresponding to the component's states of the system and $R(t)$ be reliability of the system. Then the sensitivity of the system component corresponding to its state is defined as

$$S_{i\omega} = \partial R(t)/\partial p_{i\omega} \tag{6}$$

If the system reliability depends upon component's failure rate then the sensitivity of system reliability is given by

$$S^{i}_{s^{*}\omega} = \partial R(t)/\partial \lambda^{i}_{s^{*}\omega} \tag{7}$$

where $\lambda^{i}_{s^{*}\omega}$ be failure rate of component i from state $s^{*}$to ω.

3 Reliability, MTTF and Component's Sensitivity Analysis of Two Dimensional System by UGF

Preposition 3.1 The reliability of the two-dim multi-state (*k*, *l*)/(*m*, *n*): G system having $m \times n$ components is given by $R_{stm}((k, l)/(m, n)\text{: G}) = \varphi(U_{stm}(z), D) = \sum\limits_{K_{mn}=1}^{\beta_{mn}} \cdots \sum\limits_{K_2=1}^{\beta_2} \sum\limits_{K_1=1}^{\beta_1} (\prod\limits_{j=1}^{mn} p_{jK_j})\eta(z)$

$$\begin{aligned} \text{where, } \eta(z) &= z^{\begin{bmatrix} g_{1K_1} & g_{2K_2} & \cdots & g_{nK_n} \\ g_{n+1K_{n+1}} & g_{n+2K_{n+2}} & \cdots & g_{2nK_{2n}} \\ \vdots & \vdots & & \vdots \\ g_{((m-1)n)+1K_{(m-1)n+2}} & g_{((m-1)n)+2K_{((m-1)n)+2}} & \cdots & g_{mnK_{mn}} \end{bmatrix}_{m\times n}} \\ &= \begin{cases} 1, \text{ if} \sum\limits_{I=1}^{kl} (g^{I}_{\alpha K_\alpha}) \geq D \\ 0, \text{ otherwise} \end{cases} \\ g^{I}_{\alpha K_\alpha} &= \text{element of any } k \times l \text{ submatrix of } m \times n \text{ matrix,} \\ \alpha &= 1, 2, \ldots mn, \quad K_\alpha = 1, 2, \ldots, N \end{aligned} \tag{8}$$

Proof Consider a two-dim multi-state (*k*, *l*)/(*m*, *n*): G system with $m \times n$ components. Let every component has more than two performance states ($g_{\alpha K_\alpha}$where $\alpha = 1, 2, \ldots mn, \ \ K_\alpha = 1, 2, \ldots, N$) and $P_{\alpha K_\alpha}$ be the probability corresponding to the state and D be the performance demand of the working components then the UGF of the considered system is obtained with the help of Eq. (3) as

$$U_{stm}(z) = \sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j}) z^{\begin{bmatrix} g_{1K_1} & g_{2K_2} & \cdots & g_{nK_n} \\ g_{n+1K_{n+1}} & g_{n+2K_{n+2}} & \cdots & g_{2nK_{2n}} \\ \vdots & \vdots & & \vdots \\ g_{((m-1)n)+1K_{(m-1)n+2}} & g_{((m-1)n)+2K_{((m-1)n)+2}} & \cdots & g_{mnK_{mn}} \end{bmatrix}_{m\times n}}$$

Let sum of all probabilities of the system component's states for which sum of the elements $\left(\sum_{I=1}^{kl} (g^I_{\alpha K_\alpha})\right)$ of any sub matrix $k \times l$ is greater than demand performance D be

$$= \sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j}) \eta(z)$$

Then the reliability of (k, l)/(m, n): G system is given by

$$R_{stm}((k, l)/(m, n)\text{: G}) = \varphi(U_{stm}(z), D) = \sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j}) \eta(z)$$

$$\text{where, } \eta(z) \;= z^{\begin{bmatrix} g_{1K_1} & g_{2K_2} & \cdots & g_{nK_n} \\ g_{n+1K_{n+1}} & g_{n+2K_{n+2}} & \cdots & g_{2nK_{2n}} \\ \vdots & \vdots & & \vdots \\ g_{((m-1)n)+1K_{(m-1)n+2}} & g_{((m-1)n)+2K_{((m-1)n)+2}} & \cdots & g_{mnK_{mn}} \end{bmatrix}_{m\times n}}$$

$$= \begin{cases} 1, \text{ if } \sum_{I=1}^{kl} (g^I_{\alpha K_\alpha}) \geq D \\ 0, \text{ otherwise} \end{cases}$$

Corollary 1 If $l = k$ and $m = n$ in (k, l)/(m, n): G, then

$$R_{stm}((k, k)/(n, n)\text{: G}) = \varphi(U_{stm}(z), D) = \sum_{K_{n^2}=1}^{\beta_{n^2}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{n^2} p_{jK_j}) \eta(z)$$

$$\text{where, } \eta(z) \;= z^{\begin{bmatrix} g_{1K_1} & g_{2K_2} & \cdots & g_{nK_n} \\ g_{n+1K_{n+1}} & g_{n+2K_{n+2}} & \cdots & g_{2nK_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ g_{((n-1)n)+1K_{(n-1)n+2}} & g_{((n-1)n)+2K_{((n-1)n)+2}} & \cdots & g_{n^2K_{n^2}} \end{bmatrix}_{n\times n}}$$

$$= \begin{cases} 1, \text{ if } \sum_{I=1}^{k^2} (g_{\alpha K_\alpha}^I) \geq D \\ 0, \text{ otherwise} \end{cases}$$

Preposition 3.2 If $R_{stm}((k,l)/(m,n))\colon \text{G})$ be the reliability of $(k,l)/(m,n))\colon$ G system and $MT_{stm}((k,l)/(m,n)\colon$ G is MTTF of the system, then MTTF of the considered system is obtained as

$$MT_{stm}((k,l)/(m,n)\colon \text{G} = \int_0^\infty (\sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j})\eta(z))(t)dt \tag{9}$$

Proof If system reliability of a two dim multi-state $(k,l)/(m,n)\colon$ G system is $R_{stm}((k,l)/(m,n))\colon \text{G})(t)$ then the mean time to failure of the system is $MT_{stm}((k,l)/(m,n)\colon \text{G} = \int_0^\infty R_{stm}((k,l)/(m,n))\colon \text{G})(t)dt$

Now putting value of the system reliability from Eq. (8), we get MTTF of considered system as

$$MT_{stm}((k,l)/(m,n)\colon \text{G} = \int_0^\infty (\sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j})\eta(z))(t)dt$$

Preposition 3.3 If $R_{stm}((k,l)/(m,n))\colon \text{G})$ and $S_{i\omega}((k,l)/(m,n))\colon \text{G})$ are reliability and components sensitivity of (*k*, *l*)/(*m*, *n*): G system respectively, then the system components sensitivity is evaluated as

$$S_{i\omega}((k,l)/(m,n)\colon \text{G}) = \frac{\partial}{\partial p_{i\omega}}(\sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j})\eta(z)) \tag{10}$$

Proof Sensitivity of $((k,l)/(m,n))\colon \text{G})$ system is expressed as $S_{i\omega}((k,l)/(m,n))\colon \text{G})$ which can be evaluated with the application of Eq. (5) as

$$S_{i\omega}((k,l)/(m,n)\colon \text{G}) = \frac{\partial}{\partial p_{i\omega}}(R_{stm}((k,l)/(m,n)\colon \text{G})$$

By substituting the value of the system reliability from Eq. (7), we have

$$S_{i\omega}((k,l)/(m,n)\text{: G}) = \frac{\partial}{\partial p_{i\omega}}(\sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j})\eta(z))$$

Case 1: Assume that $\lambda^i_{s^*\omega}$ be failure rate of system's component i from state s^* to ω. If failure rates are parameter of system reliability then sensitivity is evaluated as

$$S^i_{s^*\omega}((k,l)/(m,n)\text{: G}) = \frac{\partial}{\partial \lambda^i_{s^*\omega}}(\sum_{K_{mn}=1}^{\beta_{mn}} \cdots \sum_{K_2=1}^{\beta_2} \sum_{K_1=1}^{\beta_1} (\prod_{j=1}^{mn} p_{jK_j})\eta(z)) \tag{11}$$

4 Illustration

Consider a two-dim (2, 2)/(3, 3) system with components E_i, $i = 1, 2, \ldots, 9$ and let the components have $p_{i\omega}$ probabilities corresponding to the states $\omega = 1, 2, \ldots n$. Let the system component E_{i^*}, $i^* = 1, 2, \ldots, 6$ and E_{j^*}, $j^* = 7, 8, 9$ have three and two states of performance respectively and demand performance be $D = 21$ (Fig. 1).

Universal generating functions ($u_{E_n}, n = 1, 2, \ldots, 9$) of the system's every component E_i, $i = 1, 2, \ldots, 9$ can be obtained with the help of Eq. (1) as

$$\begin{aligned}
u_{E_1}(z) &= p_{11}z^5 + p_{12}z^3 + p_{13}z^0,\\
u_{E_2}(z) &= p_{21}z^6 + p_{22}z^4 + p_{23}z^0,\\
u_{E_3}(z) &= p_{31}z^4 + p_{32}z^2 + p_{33}z^0\\
u_{E_4}(z) &= p_{41}z^7 + p_{42}z^5 + p_{43}z^0,\\
u_{E_5}(z) &= p_{51}z^3 + p_{52}z^2 + p_{53}z^0,\\
u_{E_6}(z) &= p_{61}z^2 + p_{62}z^1 + p_{63}z^0\\
u_{E_7}(z) &= p_{71}z^6 + p_{72}z^0,\\
u_{E_8}(z) &= p_{81}z^7 + p_{81}z^0,\\
u_{E_9}(z) &= p_{91}z^8 + p_{92}z^0
\end{aligned}$$

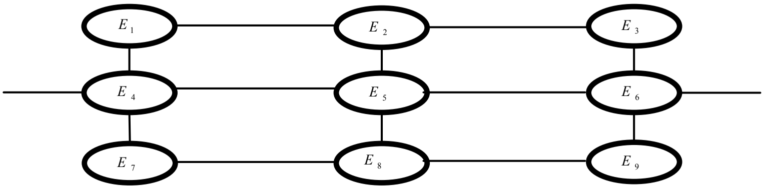


Fig. 1 Two dimension system structure

After composition (from Eq. (3)) of all of the component's UGFs we have UGF of the system.

$$U_{stm}(z) = \sum_{K_9=1}^{2}\sum_{K_8=1}^{2}\sum_{K_7=1}^{2}\sum_{K_6=1}^{3}\sum_{K_5=1}^{3}\sum_{K_4=1}^{3}\sum_{K_3=1}^{3}\sum_{K_2=1}^{3}\sum_{K_1=1}^{3}(\prod_{i=1}^{9} p_{iK_i})z^{\begin{bmatrix} g_{1K_1} & g_{2K_2} & g_{3K_3} \\ g_{4K_4} & g_{5K_5} & g_{6K_6} \\ g_{7K_7} & g_{8K_8} & g_{9K_9} \end{bmatrix}} \tag{12}$$

where p_{iK_i} be the probability, g_{iK_i} be performance corresponding to the component $i = 1, 2, 3, \ldots, 9$ and component's state K_i. Values of component's state performances $g_{11} = 5, g_{12} = 3,\ g_{13} = 0, g_{21} = 6, g_{22} = 4,\ g_{23} = 0, g_{31} = 4,$ $g_{32} = 2,\ g_{33} = 0, g_{41} = 7, g_{42} = 5,\ g_{43} = 0, g_{51} = 3, g_{52} = 2,\ g_{53} = 0, g_{61} = 2,$ $g_{62} = 1, g_{63} = 0, g_{71} = 6, g_{72} = 0,\ g_{81} = 7, g_{82} = 0, g_{91} = 8,\ g_{92} = 0.$

Then the reliability of the two dim multi-state (2, 2)/(3, 3): G system is evaluated from Eq. (12) with the application of Eq. (8) as

$$\begin{aligned} R(t) = &\sum_{K_9=1}^{2}\sum_{K_6=1}^{3}\sum_{K_4=1}^{2}\sum_{K_3=1}^{3}\sum_{K_2=1}^{3}\sum_{K_1=1}^{3}(p_{1K_1}p_{2K_2}p_{3K_3}p_{4K_4}p_{51}p_{6K_6}p_{71}p_{81}p_{9K_9}) \\ &+ \sum_{K_9=1}^{2}\sum_{K_6=1}^{3}\sum_{K_3=1}^{3}\sum_{K_2=1}^{3}\sum_{K_1=1}^{3}(p_{1K_1}p_{2K_2}p_{3K_3}p_{41}p_{52}p_{6K_6}p_{71}p_{81}p_{9K_9}) \\ &+ \sum_{K_9=1}^{2}\sum_{K_6=1}^{3}\sum_{K_3=1}^{3}(p_{11}p_{21}p_{3K_3}p_{41}p_{51}p_{6K_6}p_{72}p_{81}p_{9K_9}) \\ &+ \sum_{K_9=1}^{2}\sum_{K_7=1}^{2}\sum_{K_6=1}^{3}\sum_{K_3=1}^{3}(p_{11}p_{21}p_{3K_3}p_{41}p_{51}p_{6K_6}p_{7K_7}p_{82}p_{9K_9}) \end{aligned} \tag{13}$$

Probabilities $p_{iK_i},\ i = 1, 2, \ldots, 6,\ K_i = 1, 2, 3,$ and $p_{rK_r},\ r = 7, 8, 9,\ K_r = 1, 2$ of the considered system components can be evaluated with the application of Eq. (4) as listed in Table 1.

4.1 Reliability

Reliability of the two-dim multi-state (2, 2)/(3, 3) system with respect to time are shown in Fig. 2 with the help of Eq. (13) and probabilities listed in Table 1.

Table 1 Probabilities of system components w.r.t. time

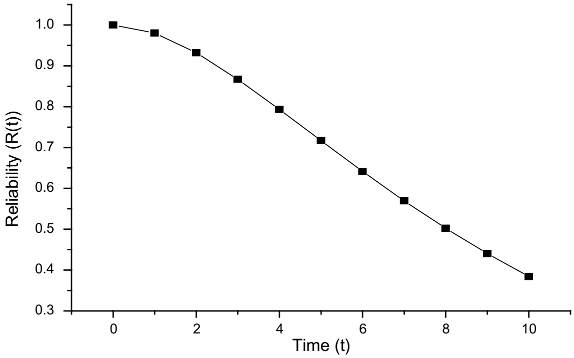
t	0	1	2	3	4	5	6	7	8	9	10
P_{11}	1	0.941765	0.88692	0.83527	0.786628	0.740818	0.697676	0.657047	0.618783	0.582748	0.548812
P_{12}	0	0.055391	0.102285	0.141678	0.174462	0.201432	0.223297	0.240692	0.254182	0.264268	0.271398
P_{13}	0	0.002845	0.010795	0.023052	0.038911	0.05775	0.079027	0.102261	0.127035	0.152984	0.17979
P_{21}	1	0.936632	0.923116	0.88692	0.852144	0.818731	0.786628	0.755784	0.726149	0.697676	0.67032
P_{22}	0	0.037673	0.070973	0.100293	0.125995	0.148411	0.167844	0.184575	0.198857	0.210924	0.220991
P_{23}	0	0.001537	0.005911	0.012787	0.021861	0.032859	0.045528	0.059642	0.074994	0.0914	0.108689
P_{31}	1	0.980199	0.960789	0.941765	0.923116	0.904837	0.88692	0.869358	0.852144	0.83527	0.818731
P_{32}	0	0.019217	0.036935	0.053247	0.068244	0.08201	0.094622	0.106156	0.11668	0.126261	0.384395
P_{33}	0	0.980391	0.931999	0.866865	0.793304	0.716963	0.641578	0.569524	0.502219	0.440413	0.13496
P_{41}	1	0.99005	0.980199	0.970446	0.960789	0.951229	0.941765	0.932394	0.923116	0.913931	0.904837
P_{42}	0	0.009705	0.01884	0.027434	0.035515	0.043107	0.050237	0.056926	0.063199	0.069076	0.074577
P_{43}	0	0.000245	0.000961	0.00212	0.003696	0.005663	0.007999	0.01068	0.013685	0.016993	0.020586
P_{51}	1	0.970446	0.941765	0.913931	0.88692	0.860708	0.83527	0.810584	0.786628	0.763379	0.740818
P_{52}	0	0.02844	0.053931	0.076718	0.097025	0.115058	0.131009	0.145055	0.157358	0.168068	0.177322
P_{53}	0	0.001115	0.004304	0.009351	0.016055	0.024234	0.033721	0. 044361	0.056014	0.068553	0.081859
P_{61}	1	0.951229	0.904837	0.860708	0.818731	0.778801	0.740818	0.704688	0.67032	0.637628	0.606531
P_{62}	0	0.046623	0.086959	0.121661	0.151318	0.176466	0.197587	0.21512	0.22946	0.240963	0.249951
P_{63}	0	0.002148	0.008204	0.017631	0.029951	0.044733	0.061594	0.080192	0.10022	0.121409	0.143518
P_{71}	1	0.932394	0.869358	0.810584	0.755784	0.704688	0.657047	0.612626	0.571209	0.532592	0.496585
P_{72}	0	0.067606	0.130642	0.189416	0.244216	0.295312	0.342953	0.387374	0.428791	0.467408	0.503415
P_{81}	1	0.923116	0.852144	0.786628	0.726149	0.67032	0.618783	0.571209	0.527292	0.486752	0.449329

(continued)

Table 1 (continued)

t	0	1	2	3	4	5	6	7	8	9	10
P_{82}	0	0.076884	0.147856	0.213372	0.273851	0.32968	0.381217	0.428791	0.472708	0.513248	0.550671
P_{91}	1	0.980199	0.960789	0.941765	0.923116	0.904837	0.88692	0.869358	0.852144	0.83527	0.818731
P_{92}	0	0.019801	0.039211	0.058235	0.076884	0.095163	0.11308	0.130642	0.147856	0.16473	0.181269

Fig. 2 Reliability of the (2, 2)/(3, 3): G and 4-within-(3, 2)-out-of-(3, 3): G system w.r.t. time



4.2 MTTF

Mean time to failure of the two-dim multi-state (2, 2)/(3, 3): G system can be obtained with help of Eqs. (9), (12). The effect over the MTTF with respect to failure rates is shown in Table 2 for the aforementioned target system.

Where $\lambda_{12}^1 = 0.06$, $\lambda_{12}^2 = 0.04$, $\lambda_{12}^3 = 0.02$, $\lambda_{12}^4 = 0.015$, $\lambda_{12}^5 = 0.037$, $\lambda_{12}^6 = 0.05$, $\lambda_{12}^7 = 0.07$, $\lambda_{12}^8 = 0.08$, $\lambda_{12}^9 = 0.02$, $\lambda_{23}^1 = 0.01$, $\lambda_{23}^2 = 0.08$, $\lambda_{23}^3 = 0.06$, $\lambda_{23}^4 = 0.055$, $\lambda_{23}^5 = 0.077$, $\lambda_{12}^6 = 0.09$.

4.3 Sensitivity Analysis

Sensitivity analysis of the (2, 2)/(3, 3): G system can measure when reliability depends on probability only is evaluated with the help of Eq. (10). Similarly, if failure rates are parameter of the reliability then it can be calculated with the help of Eq. (11) is presented by Fig. 3.

5 Conclusion

In this work, two-dim non-repairable multistate $(k, l)/(m, n)$: G system is taken for the study. Our main target in the study is to analysis reliability, MTTF and components sensitivity measure of the system with application of universal generating function and Markov process. Finally, numerical example has been taken to show the effectiveness and flexibility of the presented method. It is observed that UGF is a systematic and a good approach to calculate reliability, MTTF and importance of the target system. It is seen from the considered examples that the reliability of two-dim non-repairable multistate (2, 2)/(3, 3): G system is decreasing with increment of time. Further, observation reveals that the MTTF decreases with respect to the failure rates

Table 2 Mean time to failure of the (2, 2)/(3, 3): G system w.r.t. different failure rates

λ_{12}^{1}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	11.84991	11.12602	10.53449	10.0439	9.631834	9.281927	8.981949	8.722601	8.496695	8.298595
λ_{23}^{1}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192
λ_{12}^{2}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	10.53449	10.0439	9.631834	9.281927	8.981949	8.722601	8.496695	8.298595	8.123824	7.968788
λ_{23}^{2}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192
λ_{12}^{3}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192
λ_{23}^{3}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192
λ_{12}^{4}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.514096	9.065925	8.676472	8.335517	8.035026	7.768604	7.531102	7.31833	7.126846	6.953803
λ_{23}^{4}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.352019	9.33381	9.317278	9.302203	9.2884	9.275715	9.264017	9.253196	9.243155	9.233815
λ_{12}^{5}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	10.71272	10.11113	9.597986	9.155918	8.771748	8.435308	8.138636	7.87541	7.640552	7.42994
λ_{23}^{5}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.571709	9.515132	9.464358	9.418537	9.376979	9.339114	9.304473	9.272659	9.24334	9.216234
λ_{23}^{5}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192

(continued)

Table 2 (continued)

λ_{23}^{5}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192
λ_{12}^{7}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	12.33514	11.50674	10.86689	10.35703	9.941151	9.595687	9.304473	9.055998	8.841814	8.655566
λ_{12}^{8}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	13.17459	12.15388	11.38239	10.77758	10.29059	9.890295	9.55585	9.272659	9.03017	8.820549
λ_{12}^{9}	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192	9.28192

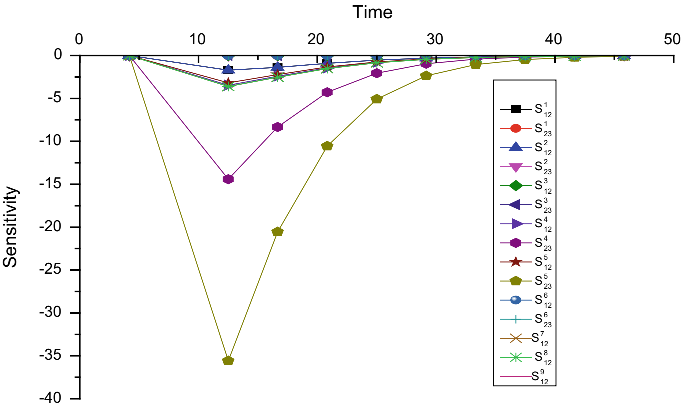


Fig. 3 Variation in sensitivity corresponding to every state failure rate of the (2, 2)/(3, 3): G system w.r.t. time

(λ^1_{12}, λ^2_{12}, λ^4_{12}, λ^4_{23}, λ^5_{12}, λ^5_{23}, λ^7_{12} and λ^8_{12}) whereas it is unchanged w.r.t. failure rates λ^1_{23}, λ^2_{23}, λ^3_{12}, λ^3_{23}, λ^6_{12}, λ^6_{23} and λ^9_{12}. Sensitivity analysis of the considered systems have also been done with respect probability and failure rate. Sensitivity of the system reliability corresponding to the parameter probability is found to be decreasing and tending to zero as time passes away. However sensitivity corresponding to those probability parameters which are not part of the reliability expression is found to be zero as expected. Further, study reveals that sensitivity of the proposed system corresponding to failure rates λ^4_{12}, λ^4_{23}, λ^5_{12}, λ^5_{23} and λ^7_{12} increasing first, then decreasing and finally its value tending to zero whereas for λ^1_{12}, λ^2_{12} and λ^8_{12} its value decreasing first, then increasing after some time values tend to zero with respect to time. For remaining failure rates its value is zero because their reliability is independent of the said failure rates.

References

1. Birnbaum, L.W.: On the Importance of Different Components in a Multi-component System, Multivariate Analysis 2. Academic Press, New York (1969)
2. Boehme, T.K., Kossow, A., Preuss, W.: A generalization of consecutive-*k*-out-of-*n*: F system. IEEE Trans. Reliab. **41**(3), 451–457 (1992)
3. Destercke, S., Sallak, M.: An extension of universal generating function in multi-state system considering epistemic uncertainties. IEEE Trans. Reliab. **62**(2), 504–514 (2013)
4. Kontoleon, J.M.: Reliability determination of a *r*-successive-*k*-out-of-*n*: F system. IEEE Trans. Reliab. **29**(5), 437 (1980)
5. Koutras, M.V., Papadopoulos, G.K., Papastavridis, S.G.: A reliability bound for 2-dimensional consecutive-*k*-out-of-*n*: F systems. Nonlinear Anal. **30**, 3345–3348 (1997)
6. Leemis, L.M.: Reliability Probabilistic Models and Statistical Methods. Prentice Hall, Inc. Englewood Clifs, N Jersey (1995)
7. Levitin, G., Lisnianski, A.: Importance and sensitivity analysis of multi-state systems using the universal generating function method. Reliab. Eng. Syst. Saf. **65**(3), 271–282 (1999)

8. Levitin, G.: Universal Generating Functions in Reliability Analysis and Optimization. Springer, London (2005)
9. Li, Y.F., Ding, Y., Zeo, E.: Random fuzzy extension of universal generating function approach for the reliability assessment of multi-state systems under aleatory and epistemic uncertainties. IEEE Trans. Reliab. **63**(1), 13–25 (2014)
10. Lisnianski, A., Frenkel, I., Ding, Y.: Multi-state System Reliability Analysis and Optimization for Engineers and Industrial Managers. Springer, London (2010)
11. Meenakshi, K., Singh, S.B.: Availability assessment of multi-state system by hybrid universal generating function and probability intervals. Int. J. Perform. Eng. **12**(4), 321–339 (2016)
12. Meenakshi, K., Singh, S.B.: Reliability analysis of multi-state complex system having two multi-state subsystems under uncertainty. J. Reliab. Stat. Stud. **10**(1), 161–177 (2017)
13. Natvig, B.: Multistate Systems Reliability Theory with Applications. Wiley (2010)
14. Natvig, B., Hjort, N., Funnemark, E.: The association in time of a markov process with application to multistate reliability theory. J. Appl. Probab. **22**, 473–479 (1985)
15. Negi, S., Singh, S.B.: Reliability analysis of non-repairable complex system with weighted subsystems connected in series. Appl. Math. Commun. **261**, 79–89 (2015)
16. Rausand, M.: System Reliability Theory. Wiley
17. Salvia, A.A., Lasher, W.C.: 2-dimensional consecutive-*k*-out-of-*n*: F models. IEEE Trans. Reliab. **39**(3), 382–385 (1990)
18. Ushakov, I.: Universal generating function. Sov. J. Comput. Syst. **24**, 118–129 (1986)
19. Yamamoto, H., Miyakawa, M.: Reliability of a linear connected-(r, s)-out-of-(m, n): F lattice system. IEEE Trans. Reliab. **44**, 333–336 (1995)
20. Wang, Z., Chen, W.: Time-variant reliability assessment through equivalent stochastic process transformation. Reliab. Eng. Syst. Saf. **152**, 166–175 (2016)

Quantum Symmetry of Classical Spaces

Debashish Goswami

Abstract We give a brief overview of generalized symmetry of classical spaces (manifolds/metric spaces/varieties etc.) in terms of (co)actions of Hopf algebras, both in the algebraic and the analytic set-up.

Keywords Compact quantum group · Quantum isometry group · Riemannian manifold · Smooth action

Subject classification: 81R50 · 81R60 · 20G42 · 58B34

1 Introduction

It is hard to over-emphasize the importance of groups in mathematics, physics and even in other physical and social sciences. They are used to describe symmetry of mathematical and physical systems. Quantum groups are generalization of groups and have their origin in different problems in mathematical physics as well as the theory of classical locally compact groups. The pioneering work by Drinfeld and Jimbo [9, 10, 17, 18] and others (e.g. [24, 25]) gave the formulation of quantum groups in the algebraic setting as Hopf algebras typically obtained by deformations of the universal enveloping algebras of semisimple Lie algebras. This led to a deep and successful theory having connections with physics, knot theory, number theory, representation theory etc. On the other hand, Woronowicz [28, 29] approached it from a point of view of harmonic analysis on locally compact groups and came up with a set of axioms for defining compact quantum groups (CQG for short) as the generalization of compact topological groups. A similar theory of locally compact quantum groups remained elusive for quite some time. Finally, a satisfactory theory

Partially supported by J C Bose Fellowship from D.S.T. (Govt. of India).

D. Goswami (✉)
Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, India
e-mail: goswamid@isical.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_8

of locally compact quantum groups was proposed in [19, 20, 23]. However, in this article we mostly confine ourselves in the framework of compact quantum groups.

Quantum groups correspond to some kind of 'generalized symmetry' of physical systems and mathematical structures. Indeed, the idea of a group acting on a space was extended to the idea of a quantum group 'co-acting' on a noncommutative space (that is, a possibly noncommutative C^* algebra). The question of defining and finding 'all quantum symmetries' arises naturally in this context. Such an approach was taken in the pioneering work by Manin [22], though in a purely algebraic framework. Indeed, Manin's quantum semigroups such as $M_q(2)$ [21, 22] were constructed as universal symmetry objects of suitable mathematical entities. It was Shuzhou Wang who began the study of the formulation of such universal symmetry objects in the analytic framework of CQG. In [27] he defined and proved the existence of quantum automorphism groups of finite dimensional C^* algebras. Since then, many interesting examples of such quantum groups, particularly the quantum permutation groups of finite sets and finite graphs, were extensively studied by a number of mathematicians (see for example [1–7] and references therein).

The underlying basic principle of defining a quantum automorphism group corresponding to some given mathematical structure (for example a finite set, a graph, a C^* or von Neumann algebra) consists of two steps: first, to identify (if possible) the group of automorphism of the structure as a universal object in a suitable category, and then, try to look for the universal object in a similar but bigger category, replacing groups by quantum groups of appropriate type.

We will briefly review the basics in the next section covering only those aspects of the quantum group theory which concern us. For a comprehensive idea about the vast and rich theory of quantum groups, the reader should consult any of the several standard textbooks on the subject, some of which are listed in the bibliography.

2 Basic Definitions and Terminologies

We denote by $\otimes_{\rm alg}$ the algebraic tensor product, whereas $\otimes$ will denote some kind of topological tensor product to be explained later. In particular, for two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, we denote by $\mathcal{H} \otimes \mathcal{K}$ the (unique) Hilbert space obtained by completing their algebraic tensor product.

Definition 2.1 A Hopf algebra $\mathcal{Q}$ over a filed F is a vector space over F equipped with linear maps $m : \mathcal{Q} \otimes_{\rm alg} \mathcal{Q} \to \mathcal{Q}$, $\Delta : \mathcal{Q} \to \mathcal{Q} \otimes_{\rm alg} \mathcal{Q}$, $\kappa : \mathcal{Q} \to \mathcal{Q}$ and $\epsilon : \mathcal{Q} \to F$ and an element 1 of $\mathcal{Q}$ such that
(i) $(\mathcal{Q}, m)$ is an associative F-algebra with the unit element 1;
(ii) Δ is a unital homomorphism satisfying the co-associativity condition $(\Delta \otimes {\rm id}) \circ \Delta = ({\rm id} \otimes \Delta) \circ \Delta$;
(iii) κ is an anti-homomorphism, ϵ is a unital homomorphism;
(iv) $m \circ (\kappa \otimes {\rm id}) \circ \Delta = m \circ ({\rm id} \otimes \kappa) \circ \Delta = \epsilon(\cdot)1$;
(v) $(\epsilon \otimes {\rm id}) \circ \Delta = ({\rm id} \otimes \epsilon) \circ \Delta = {\rm id}$.

The maps $m, \Delta, \kappa, \epsilon$ are called the multiplication, co-multiplication (or co-product), antipode and co-unit respectively. If $F = \mathbb{C}$ and the algebra has an involution $*$ such that Δ and κ are $*$-homomorphism and κ is invertible with $(* \circ \kappa)^2 = \text{id}$, we say that $\mathcal{Q}$ is a Hopf $*$-algebra.

For a finite group G, the algebra $\mathcal{F}(G)$ of all complex valued functions on G forms a Hopf $*$ algebra, by taking the point-wise operations as the algebra operations, involution given by the complex conjugation of functions and taking Δ, ϵ, κ as follows:

$$\Delta(f)(g,h) = f(gh), \ \ \kappa(f)(g) = f(g^{-1}), \ \ \epsilon(f) = f(e),$$

where e is the identity of G. Another Hopf algebra, dual to the above in a suitable sense, is given by the convolution algebra $\mathbb{C}G$. It is same as $\mathcal{F}(G)$ as a vector space but the algebra structure is given by $\chi_g \cdot \chi_h = \chi_{gh}$, where χ_g denotes the characteristic function of the point g. The co-product sends χ_g to $\chi_g \otimes \chi_g$ for every $g \in G$. The counit $\hat{\epsilon}$ and the antipode $\hat{\kappa}$ are given by, $\hat{\epsilon}(\chi_g) = 1$, $\hat{\kappa}(\chi_g) = \chi_{g^{-1}}$ for all g.

Definition 2.2 A unital homomorphism $\alpha : \mathcal{C} \to \mathcal{C} \otimes_{\text{alg}} \mathcal{Q}$, where $\mathcal{C}$ is a unital algebra over F and $\mathcal{Q}$ is a Hopf algebra over F, is called a co-action if $(\text{id} \otimes \Delta) \circ \alpha = (\alpha \otimes \text{id}) \circ \alpha$ and $(\text{id} \otimes \epsilon) \circ \alpha = \text{id}_{\mathcal{C}}$. The co-action is called inner faithful if it does not factor through a proper Hopf subalgebra of $\mathcal{Q}$.

Now, we come to the analytic counterparts of the above concepts.

Definition 2.3 A C^* algebra is a complex algebra $\mathcal{A}$ which is complete (i.e.Banach space) with respect to some norm $\| \cdot \|$ defined on it and also has an involution $*$ satisfying $\|xy\| \leq \|x\|\|y\|$, $\|x^*\| = \|x\|$ and $\|x^*x\| = \|x\|^2$.

The celebrated Gelfand's Theorem states that any commutative C^* algebra is isomorphic with $C_0(X)$, i.e. the algebra of continuous functions which vanish at infinity, for some locally compact Hausdorff space X. This justifies the viewpoint that a general (possibly noncommutative) C^* algebra may be thought of as a 'noncommutative space'. Moreover, unital C^* algebras are the analogues of compact spaces, because a commutative unital C^* algebra is (isomorphic with) $C(X)$ for some compact Hausdorff space X.

A general (possibly noncommutative) C^* algebra can be embedded as a closed, $*$-subalgebra of $\mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Given $\mathcal{A}_i \subseteq \mathcal{B}(\mathcal{H}_i)$, $i = 1, 2$, we can consider their algebraic tensor product as a subalgebra of $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ and the closure of this algebraic tensor product w.r.t. the norm inherited from that of $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ is called the minimal tensor product and will be denoted by $\mathcal{A}_1 \otimes \mathcal{A}_2$. It actually does not depend on the embedding $\mathcal{A}_i \subseteq \mathcal{B}(\mathcal{H}_i)$. If $\mathcal{A}_1 = C(X)$, one has $\mathcal{A}_1 \otimes \mathcal{A}_2 \cong C(X, \mathcal{A}_2)$, i.e. the algebra of continuous functions from X to $\mathcal{A}_2$.

Definition 2.4 A **c**ompact quantum group (CQG for short) a la Woronowicz is a pair $(\mathcal{A}, \Delta)$ where $\mathcal{A}$ is a unital C^*-algebra, Δ is a coassociative comultiplication, i.e. a unital C^*-homomorphism from $\mathcal{A}$ to $\mathcal{A} \otimes \mathcal{A}$ (minimal tensor product) satisfying $(\Delta \otimes \text{id}) \circ \Delta = (\text{id} \otimes \Delta) \circ \Delta$, and linear span of each of the sets $\{(b \otimes 1)\Delta(c) \colon b, c \in \mathcal{A}\}$ and $\{(1 \otimes b)\Delta(c) \colon b, c \in \mathcal{A}\}$ is dense in $\mathcal{A} \otimes \mathcal{A}$.

A CQG which is commutative as a C^* algebra is isomorphic with $C(G)$ for a compact group G, with $\Delta(f)(g,h) = f(gh)$. Every CQG $\mathcal{A}$ contains a unital dense $*$-subalgebra $\mathcal{A}_0$ and maps $\kappa : \mathcal{A}_0 \to \mathcal{A}_0$, $\epsilon : \mathcal{A}_0 \to \mathbb{C}$ such that $\mathcal{A}_0$ becomes a Hopf-$*$ algebra. This Hopf algebra will be called the Hopf algebra associated to the CQG $\mathcal{A}$.

There is an analogue of the Haar measure, namely a (unique) positive linear functional h, called the Haar state, on $\mathcal{A}$ such that $h(1) = 1$ and $(h \otimes \mathrm{id})(\Delta(a)) = (\mathrm{id} \otimes h)(\Delta(a)) = h(a)1$ for all $a \in \mathcal{A}$. Moreover, there is an exact analogue of unitary co-representation and Peter-Weyl theory. A unital $*$-homomorphism $\pi : \mathcal{Q}_1 \to \mathcal{Q}_2$ (where $\mathcal{Q}_1$, $\mathcal{Q}_2$ are CQG's with the coproduct Δ_i, $i = 1, 2$ respectively) is called a CQG morphism if $\Delta_2 \circ \pi = (\pi \otimes \pi) \circ \Delta_1$.

We next define a generalization of group action on spaces.

Definition 2.5 We say that a CQG $(\mathcal{A}, \Delta)$ (co)-acts on a (unital) C^*-algebra $\mathcal{C}$ if there is a unital $*$-homomorphism $\alpha : \mathcal{C} \to \mathcal{C} \otimes \mathcal{A}$ such that $(\alpha \otimes \mathrm{id}) \circ \alpha = (\mathrm{id} \otimes \Delta) \circ \alpha$, and the linear span of $\alpha(\mathcal{C})(1 \otimes \mathcal{A})$ is norm-dense in $\mathcal{C} \otimes \mathcal{A}$.

In Woronowicz theory, it is customary to drop 'co', and call the above co-action simply 'action' of the CQG on the C^* algebra. However, to avoid confusion, let's say that $\mathcal{A}$ acts on a space X, or, α is an action on X to mean α is a co-action on $C(X)$ in the sense of the above definition. A co-action α on $\mathcal{C}$ is called faithful if the $*$-subalgebra generated by $\{(\omega \otimes \mathrm{id})(\alpha(b))\}$, where $b \in \mathcal{C}$ and ω varying over the set of bounded linear functionals on $\mathcal{C}$, is dense in $\mathcal{A}$.

3 Quantum Symmetries of Classical Spaces

3.1 Examples of Genuine Quantum Symmetries of Classical Non-smooth Spaces

It is a natural question to ask: are there genuine symmetries (i.e. not given by a group action) of a classical space? Mathematically, this can be phrased either purely algebraically or analytically. For example, one may ask whether there are Hopf algebras which are not commutative as algebras having inner faithful co-actions on the coordinate ring of some variety? In the analytical framework, an analogous question will be: is there a CQG which is noncommutative as a C^* algebra and which co-acts on $C(X)$ where X is a topological space? In fact, if there are such 'genuine' quantum symmetries of classical spaces, one may hope to understand the space better in terms of them.

Let us first consider the simplest classical space, namely a finite set with cardinality n, say $X = \{1, 2, ..., n\}$. Let $G = S_n$ be the group of permutations of X. It can be identified as the universal object in the category of groups acting on X. For a similar (bigger) category of compact quantum groups co-acting on $C(X)$, Wang proved in [27] that this category has a universal (initial) object $\mathcal{S}_n^+$ which is the unital C^* algebra generated by n^2 elements q_{ij} satisfying

$$q_{ij} = q_{ij}^* = q_{ij}^2, \ \sum_i q_{ij} = 1 = \sum_j q_{ij}.$$

The coproduct is given by $\Delta(q_{ij}) = \sum_k q_{ik} \otimes q_{kj}$, and the co-action on $C(X)$ is given by $\alpha(\chi_i) = \sum_j \chi_j \otimes q_{ji}$.
This CQG is naturally called 'quantum permutation group' of n objects. It is genuine quantum group (i.e. noncommutative as C^* algebra) for $n \geq 4$. In other words, given any CQG $\mathcal{Q}$ with co-action $\beta : C(X) \to C(X) \otimes \mathcal{Q}$, there is a unique CQG morphism $\pi : \mathcal{S}_n^+ \to \mathcal{Q}$ such that $\beta = (\mathrm{id} \otimes \pi) \circ \alpha$.

One may interpret the above as an abundance of genuine quantum symmetries for a finite set of cardinality 4 or more, in fact, infinite dimensional spaces of such symmetries in contrast to the finitely many classical (permutation) symmetries. Using this, one can construct a genuine quantum symmetry of any disconnected space with 4 or more homoemorphic components. However, it is not so clear how to construct genuine quantum symmetries for compact connected spaces and the author of the present review made a conjecture that a compact connected metric space does not admit any genuine quantum symmetry. Adapting the above quantum permutation, Huichi Huang disproved this conjecture by constructing in [16] a faithful co-action of $\mathcal{S}_n^+$ on a compact connected topological space obtained by gluing n copies of a given space (e.g. the circle) at a common point.

On the other hand, there are quite a few interesting examples of inner faithful co-actions of genuine Hopf algebras on polynomial algebras as well as coordinate algebras of certain varieties. In [11], Etingof and Walton gave an example of inner faithful co-action of the finite dimensional genuine Hopf algebra $\mathbb{C}S_3$ (the convolution algebra of the group S_3) on the coordinate ring of the (non-smooth) variety $\{xy = 0\}$. The co-action (say α) is very easy to describe: on the generators x, y of the coordinate ring, we define $\alpha(x) = x \otimes \chi_{(12)}, \ \alpha(y) = y \otimes \chi_{(13)}$, where $\chi_{(12)}, \chi_{(13)}$ denote the elements of the group algebra given corresponding to the transpositions (12), (13) respectively. Let us also mention that in [26], the authors constructed an inner faithful co-action of a genuine infinite dimensional (but not associated to any CQG) Hopf algebra on the polynomial ring of n variables, which is the coordinate ring of the n-plane.

4 The No-go Results and Open Questions for Quantum Symmetries on Smooth Spaces

It is certainly very interesting to construct genuine quantum symmeries of smooth, connected manifolds and varieties. If we look carefully at the examples of connected spaces with genuine quantum symmetry given in the previous subsection, we see that in all the examples, either the space is non-smooth in some sense, or the Hopf algebra is not coming from a CQG. It thus makes sense to ask whether it is possible to get an example of faithful co-action of a genuine compact quantum group on a compact, smooth, connected manifold. An answer to this question turned out to be much more challenging than it would appear at a first glance.

If a compact group acts smoothly on a compact smooth manifold, one can use an averaging trick to get a Riemannian structure on the manifold for which the group action becomes isometric. Thus, the study of smooth action by compact groups on compact smooth manifolds is equivalent to the study of isometric actions. This motivated the present author and his collaborators to formulate and investigate a notion of quantum isometry in the context of classical as well as more general noncommutative manifolds a la Connes (see [8]). We do not go into the precise formulation of this theory but refer the reader to [14] and the references therein, some of which are listed in the bibliography as well. Let us just mention that given a compact Riemannian manifold M there is a universal compact quantum group in the category of all CQG $\mathcal{Q}$ with faithful co-action α of $\mathcal{Q}$ on $C(M)$ which commutes with the Hodge Laplacian on functions in a natural sense. This universal or largest CQG is called the quantum isometry group for M, denoted by $QISO(M)$. After the initial formulation in [12], there has been a flurry of computations of the quantum isometry groups of classical and noncommutative manifolds by several authors. It was rather interesting to notice that for the connected compact manifolds like the spheres and the tori, the quantum isometry groups turn out to be $C(ISO(M))$. That is, there is no genuine quantum isometry. It led the present author to conjecture that $QISO(M) = C(ISO(M))$ for an arbitrary compact connected Riemannian manifolds. In fact, he made an even stronger conjecture there is no faithful, smooth co-action of a CQG on a compact smooth connected manifold. Let us explain here that by a smooth co- action, we mean the following:

Definition 4.1 A co-action $\alpha : C(M) \to C(M) \otimes \mathcal{Q}$, where $\mathcal{Q}$ is a CQG and M is a compact smooth manifold, is called smooth if $\alpha(C^\infty(M)) \subseteq C^\infty(M, \mathcal{Q})$ and the linear span of $\alpha(C^\infty(M))(1 \otimes \mathcal{Q})$ is dense in $C^\infty(M, \mathcal{Q})$ in its natural Frechet topology coming from the smooth structure of M.

Now, we can precisely state the conjecture: **C**onjecture I Let a CQG $\mathcal{Q}$ have a faithful smooth co-action on $C(M)$, where M is a compact, smooth, connected manifold. Then $\mathcal{Q}$ must be a commutative as a C^* algebra, i.e. $\mathcal{Q} \cong C(G)$ for some compact group G acting smoothly on M.

After a long effort, the isometric case of the above conjecture, i.e. $QISO = ISO$ for compact connected Riemannian manifolds, was verified in [15] by the present author and Soumalya Joardar. This involved an intricate mixture of analytic, algebraic and geometric ideas. There had been a parallel, independent development in a somewhat similar direction by Etingof and Walton who proved the following:

Theorem 4.2 *Let $\mathcal{C}$ be a commutative domain and $\mathcal{Q}$ be a finite dimensional semisimple (equivalently, co-semisimple) Hopf algebra with an inner faithful co-action on $\mathcal{C}$. Then $\mathcal{Q}$ must be commutative as an algebra.*

These results gave a strong indication in favour of an affirmative resolution of the conjecture. However, Etingof-Walton's techniques could not be applied to prove

Conjecture I as their arguments crucially used the semi-simplicity and finite dimensionality of the Hopf algebra. Indeed, a resolution of Conjecture I seemed quite elusive and an initial announcement of a proof by the present author and two other collaborators posted on the archive had to be withdrawn due to a flaw. Finally, the present author could achieve a complete proof of the truth of Conjecture I using probabilistic tools in a very recent article [13] posted in the archive.

We end with the interesting and important open questions:
Question 1: Can the result of Etingof and Walton be extended to more general class of possibly infinite dimensional Hopf algebras, e.g. Hopf algebras associated to compact quantum groups? **Q**uestion 2: Can one get many more interesting examples of genuine quantum symmetries by Hopf algebras of non-compact type on smooth manifolds and algebraic varieties?

References

1. Banica, T.: Quantum automorphism groups of small metric spaces. Pac. J. Math. **219**(1), 27–51 (2005)
2. Banica, T.: Quantum automorphism groups of homogeneous graphs. J. Funct. Anal. **224**(2), 243–280 (2005)
3. Banica, T., Bichon, J.: Quantum groups acting on 4 points. J. Reine Angew. Math. **626**, 75–114 (2009)
4. Banica, T., Bichon, J., Collins, B.: Quantum permutation groups: a survey. noncommutative harmonic analysis with applications to probability. Polish Acad. Sci. Inst. Math. **78**, 13–34. Banach Center Publications, Warsaw (2007)
5. Banica, T., Moroianu, S.: On the structure of quantum permutation groups. Proc. Am. Math. Soc. **135**(1), 21–29 (2007)
6. Bichon, J.: Quantum automorphism groups of finite graphs. Proc. Am. Math. Soc. **131**(3), 665–673 (2003)
7. Bichon, J.: Algebraic quantum permutation groups. Asian-Eur. J. Math. **1**(1), 1–13 (2008)
8. Connes, A.: Noncommutative Geometry. Academic Press, London, New York (1994)
9. Drinfeld, V.G.: Quantum groups. In: Proceedings of the International Congress of Mathematicians, Berkeley (1986)
10. Drinfeld, V.G.: Quasi-Hopf algebras. Leningr. Math. J. **1**, 1419–1457 (1990)
11. Etingof, P., Walton, C.: Semisimple hopf actions on commutative domains. Adv. Math. **251**, 47–61 (2014)
12. Goswami, D.: Quantum group of isometries in classical and non commutative geometry. Commun. Math. Phys. **285**(1), 141–160 (2009)
13. Goswami, D.: Non-existence of genuine (compact) quantum symmetries of compact, connected smooth manifolds. preprint (2018)
14. Goswami, D., Bhowmick, J.: Quantum Isometry Groups. Springer, Infosys Series (2017)
15. Goswami, D., Joardar, S.: Non-existence of faithful isometric action of compact quantum groups on compact, connected Riemannian manifolds. Geom. Funct. Anal. **28**(1), 146–178 (2018)
16. Huang, H.: Faithful compact quantum group actions on connected compact metrizable spaces. J. Geom. Phys. **70**, 232–236 (2013)
17. Jimbo, M.: Quantum R-matrix for the generalized Toda system. Commun. Math. Phys. **10**, 63–69 (1985)
18. Jimbo, M.: A q-diff erence analogue of U(g) and the Yang-Baxter equation. Lett. Math. Phys. **10**(1), 63–69 (1985)

19. Kustermans, J., Vaes, S.: Locally compact quantum groups. Ann. Sci. Ecole Norm. super. **33**(6), 837–934 (2000)
20. Kustermans, J., Vaes, S.: Locally compact quantum groups in the von Neumann algebraic setting. Math. Scand. **92**(1), 68–92 (2003)
21. Manin, Y: Some remarks on Koszul algebras and quantum groups. Ann. Inst. Fourier Grenoble **37**(4), 191–205 (1987)
22. Manin, Y.: Quantum Groups and Non-commutative Geometry. CRM, Montreal (1988)
23. Masuda, T., Nakagami, Y., Woronowicz, S.L.: A C^*-algebraic framework for quantum groups. Int. J. Math. **14**(9), 903–1001 (2003)
24. Reshetikhin, N.Y., Takhtajan, L.A., Faddeev, L.D.: Quantization of Lie groups and Lie algebras. Algebra i analiz 1, **178** (1989) (Russian), English translation in Leningrad Math. J. 1
25. Soibelman, Y.S., Vaksman, L.L.: On some problems in the theory of quantum groups. representation theory and dynamical systems. Adv. Soviet Math. Am. Math. Soc. Providence, RI, **9**, 3–55 (1992)
26. Walton, C., Wang, X.: On quantum groups associated to non-noetherian regular algebras of dimension 2. arXiv:1503.0918
27. Wang, S.: Quantum symmetry groups of finite spaces. Commun. Math. Phys. **195**, 195–211 (1998)
28. Woronowicz, S.L.: Compact matrix pseudogroups. Commun. Math. Phys. **111**(4), 613–665 (1987)
29. Woronowicz, S.L.: Compact quantum groups. In: Connes, A. (ed.) et al. Symétries Quantiques (Quantum symmetries) (Les Houches), p. 1998. Elsevier, Amsterdam (1995)

Quasi-isometry and Rigidity

Parameswaran Sankaran

Abstract This is a brief exposition on the quasi-isometric rigidity of irreducible lattices in Lie groups. The basic notions in coarse geometry are recalled and illustrated. It is beyond the scope of these notes to go into the proofs of most of the results stated here. We shall be content with pointing the reader to standard references for detailed proofs. These notes are based on my talk in the *International Conference on Mathematics and its Analysis and Applications in Mathematical Modelling* held at Jadavpur University, Kolkata, in December 2017.

Keywords Quasi-isometry · Invariants · Rigidity

Mathematics Subject classification: 20F32 · 20E99

1 Introduction

Let (X, d_X), (Y, d_Y) be metric spaces and let $\lambda \geq 1$, $\epsilon \geq 0$ be real numbers. A set-map $f : X \to Y$ is called a (λ, ϵ)-*quasi-isometric embedding* if the following condition holds: $-\epsilon + \lambda^{-1} d_X(x_0, x_1) \leq d_Y(f(x_0), f(x_1)) \leq \lambda d_X(x_0, x_1) + \epsilon$ for all $x_0, x_1 \in X$. If there exists a $C \geq 0$ such that each $y \in Y$ is at a distance at most C from the image $f(X)$, we say that f is C-dense. A quasi-isometric embedding which is C-dense for some $C \geq 0$, is called a (λ, ϵ, C)-*quasi-isometric equivalence* or, more briefly, a *quasi-isometry*. When $f : X \to Y$ is a (λ, ϵ)-quasi-isometric embedding with $\epsilon = 0$, then f is necessarily continuous. In general, however, f need not be so. When f is

The author was partially supported by a XII Plan Project, Department of Atomic Energy, Government of India.

P. Sankaran (✉)
Chennai Mathematical Institute, H1 SIPCOT IT Park, Siruseri, Kelambakkam,
Chennai 603103, India
e-mail: sankaran@cmi.ac.in

Institute of Mathematical Sciences, (HBNI), CIT Campus Taramani,
Chennai 600113, India

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_9

a (λ, ϵ, C) quasi-isometry, there exist $\mu \geq 1, \delta \geq 0, D$ and a $g : Y \to X$ which is a (μ, δ, D)-quasi-isometry such that $g \circ f$ and $f \circ g$ are a bounded distance away from id_X and id_Y respectively, that is, $||g \circ f - id_X|| := \sup_{x \in X} d_X(g(f(x)), x) < \infty$ and $||f \circ g - id_Y|| = \sup_{y \in Y} d_Y(f(g(y)), y) < \infty$. We say that f, g are quasi-inverses of each other and that X, Y are of the same quasi-isometry type; we write $X \sim_{\text{qi}} Y$. Being of the same quasi-isometry type is an 'equivalence relation' on the class of all metric spaces. A $(\lambda, 0, 0)$-quasi-isometry $f : (X, d_X) \to (Y, d_Y)$ is nothing but a bi-Lipschitz homeomorphism. Quite often, the specific values of λ, ϵ, C are not so important and so we usually omit explicit mention of them. As a first example, $\mathbb{Z} \hookrightarrow \mathbb{R}$ is a $(1, 0, 1/2)$-isometry equivalence with $(1, 1, 0)$-quasi-inverse $\mathbb{R} \to \mathbb{Z}$ defined as $x \mapsto \lfloor x \rfloor$.

On the set of all quasi-isometry self-equivalences of (X, d_X), one has an equivalence relation where $f \sim g$ if $||f - g|| = \sup_{x \in X} d_X(f(x), g(x)) < \infty$ for quasi-isometries $f, g : X \to X$. The set of equivalence classes form a group $\mathcal{QI}(X)$, called the group of quasi-isometries of (X, d_X), where $[f] \cdot [g] = [f \circ g]$. The group of isometries will be denoted $\text{Isom}(X)$. One has a natural homomorphism $\text{Isom}(X) \to \mathcal{QI}(X)$ defined as $f \mapsto [f]$. In general, $\text{Isom}(X)$ and $\mathcal{QI}(X)$ are not closely related. For example, $\mathcal{QI}(\mathbb{S}^n)$ is trivial whereas $\text{Isom}(\mathbb{S}^n)$ is the orthogonal group $\text{O}(n + 1)$. On the other hand, when $X = \mathbb{Z} \cup \{3/4\} \subset \mathbb{R}$ the group $\text{Isom}(X)$ is trivial whereas it can be seen $\mathcal{QI}(X) \cong \mathcal{QI}(\mathbb{R})$ contains the group $\text{GL}(1, \mathbb{R})$; indeed $\mathcal{QI}(\mathbb{R})$ is a rather large group. See [15].

The notion of quasi-isometry captures the essential features of the *large scale* geometry of a metric space, that is, those features which are remain when "viewed from far away." For example, any two (non-empty) bounded metric spaces are quasi-isometrically equivalent to each other. Also if $B \subset X$ is a bounded subset of X, then X is quasi-isometrically equivalent to $X \setminus B$. More generally, if $Y \subset X$ is a C-dense subset of X, (i.e., if any $x \in X$ is at a distance at most C from a $y \in Y$) then X and Y are quasi-isometrically equivalent. Conversely, if the inclusion $Y \hookrightarrow X$ is a quasi-isometry, then Y is C-dense for some $C > 0$. *Coarse geometry* is the study of properties of metric spaces which remain invariant under quasi-isometric equivalence. An important problem in coarse geometry is the classification problem, which asks to classify metric spaces according to their quasi-isometry type. A part of this problem is to study invariants of quasi-isometry, which may be used to distinguish quasi-isometry types. Our aim here will be to give a brief exposition of the concept of quasi-isometric rigidity and to state the results concerning rigidity properties of lattices in semisimple Lie groups. I omit all the proofs and point the reader to relevant sources.

1.1 Groups as Geometric Objects

One of main objectives of coarse geometry is the study of (finitely generated) groups viewed as geometric objects via the word metric—a point of view that resulted in explosive growth of the subject since the seminal work of Gromov [8]. More precisely, suppose that Γ is finitely generated group and $S \subset \Gamma$ a finite generating set. One has the word metric defined as $d(\gamma_0, \gamma_1) = l_S(\gamma_0^{-1}\gamma_1)\ \forall \gamma_0, \gamma_1 \in S$; here $l_S(\gamma)$ denotes the length of γ as a word in $S \cup S^{-1}$, that is, $l_S(\gamma) = k$ where $k \geq 0$ is the smallest integer for which γ has an expression $\gamma = a_1 \cdots a_k, a_j \in S \cup S^{-1}, 1 \leq j \leq k$, with $l_S(1) := 0$. It turns out that, changing the generating set S to another finite generating set S' leads to another metric $d_{S'}$ but does not change the quasi-isometry type of (Γ, d_S). In fact, it is easily seen that d_S and $d_{S'}$ are bi-Lipschitz equivalent, i.e, the identity map $(\Gamma, d_S) \to (\Gamma, d_{S'})$ is bi-Lipschitz.

Recall that the Cayley graph $\mathcal{C} = \mathcal{C}(\Gamma, S)$ of (Γ, S) is a graph whose set of vertices is Γ and $(\gamma, \gamma') \in \Gamma \times \Gamma$ is an (oriented) edge whenever $\gamma^{-1}.\gamma'$ is in S. There is a natural metric on $\mathcal{C}$ in which each edge has length 1 and $d(\gamma, \gamma') = l_S(\gamma^{-1}\gamma')$. This metric is invariant under the natural Γ-action on the left of $\mathcal{C}$. The inclusion $\Gamma \hookrightarrow \mathcal{C}$ is 1-dense and hence is a quasi-isometry.

A metric space (X, d_X) is *proper* if closed all balls in X of finite radii are compact. (X, d_X) is a *length space* if $d_X(x_0, x_1) = \inf l(\sigma)$ where the infimum is taken over all (rectifiable) paths σ from x_0 to x_1 and $l(\sigma)$ denotes the length of σ. It is said to be a *geodesic metric space* if, for any $x_0, x_1 \in X$, there is a path $\sigma : [0, l] \to X$ from x_0, to x_1 such that $d_X(\sigma(t), \sigma(t')) = |t - t'|, \forall t, t' \in [0, l]$. Such a path is called a *geodesic*.

Suppose that Γ acts on a metric space (X, d_X). The action is *properly discontinuous* if, given any $x \in X$, there exists an open neighbourhood U of x such that $U \cap \gamma U \neq \emptyset$ for at most finitely many elements $\gamma \in \Gamma$. The action is said to be *cocompact* if $\Gamma \backslash X$ is compact; equivalently, there is a compact set $K \subset X$ such that $X = \cup_{\gamma \in \Gamma} \gamma K$.

The following theorem is often referred to as the fundamental theorem of coarse geometry. It was first proved by V.A. Efremovich in 1953 and by A. Švarc in 1955. It was later rediscovered by J. Milnor in 1968. It generalizes the above observation that a finitely generated group Γ (with word metric d_S) is quasi-isometric to its Cayley graph $\mathcal{C}(\Gamma, S)$. We refer the reader to [1, Chapter I.8, Prop. 8.19] for a proof.

Theorem 1.1 (Švarc-Milnor Lemma) *Suppose that a group Γ acts properly discontinuously and cocompactly by isometries on a proper length metric space (X, d_X). Then Γ is a finitely generated group. If $S \subset \Gamma$ is any finite generating set and if $x_0 \in X$ is arbitrary, then the map $\gamma \mapsto \gamma.x_0$ is a quasi-isometry* $(\Gamma, d_S) \to (X, d_X)$. □

Example 1.2 (i) Let Γ be a group generated by a finite set $S \subset \Gamma$. If $\Lambda \subset \Gamma$ is a finite index subgroup, then $\Gamma \sim_{\text{qi}} \Lambda$. This follows from Theorem 1.1 by restricting the action of Γ on $\mathcal{C}(\Gamma, S)$ to Λ. Also if N is a finite subgroup of Γ, then $\Gamma \sim_{\text{qi}} \Gamma/N =: \bar{\Gamma}$. To see this, we assume, as we may, that $N \setminus \{1\} \subset S$ and that no two distinct elements of $S \setminus N$ are in the same coset γN. Then it is readily seen that $l_{\bar{S}}(\bar{\gamma}) \leq l_S(\gamma) \leq l_{\bar{S}}(\bar{\gamma}) + 1$ where $\bar{\gamma}$ denotes $\gamma N \in \bar{\Gamma}$ and $\bar{S} := \{\bar{s} \mid s \in S \setminus N\} \subset \bar{\Gamma}$. It follows that the canonical quotient map $\eta : (\Gamma, d_S) \to (\bar{\Gamma}, d_{\bar{S}})$ is a (λ, ϵ, C)-quasi-isometry where $\lambda = 1, \epsilon = 1, C = 0$. Alternatively one may apply Švarc-Milnor lemma (Theorem 1.1) to the action of Γ (via the quotient map $\Gamma \to \bar{\Gamma}$) on the Cayley graph $\mathcal{C}(\bar{\Gamma}, \bar{S})$ where $\bar{S}$ is any finite generating set of $\bar{\Gamma}$. Note that the Γ-action is proper since N is finite.

(ii) Two groups Γ_0 and Γ_1 are said to be *commensurable* if there exists a group Γ such that Γ is isomorphic to a finite index subgroup of Γ_i for $i = 0, 1$. Since intersection of two finite index subgroups is again of finite index, commensurability is an equivalence relation. For example, any free group F_n of rank $n \geq 2$ may be realised as a finite index subgroup of F_2. Thus any two non-abelian free groups of finite rank are commensurable. As another example, let G_0 and G_1 be finite groups with $o(G_0) \geq 2$, $o(G_1) \geq 3$, then their free product $G := G_0 * G_1$ contains a non-abelian free group F of finite rank such that F has finite index in G. Thus G is commensurable with F_2.

We say that Γ_0 and Γ_1 are *weakly commensurable* if there exist finite normal subgroups $N_i \subset \Gamma_i, i = 0, 1$, such that Γ_0/N_0 and Γ_1/N_1 are commensurable. Weak commensurability is an equivalence relation. This follows from two observations: (a) the normal subgroup $NN' \subset \Gamma$ is finite whenever N, N' are finite normal subgroups a group Γ, and, (b) commensurability is an equivalence relation. From (i), we see that weakly commensurable groups are quasi-isometrically equivalent (with respect to any word metrics). As an application, recall that $\mathrm{PSL}(2, \mathbb{Z}) = \mathrm{SL}(2, \mathbb{Z})/\pm I$ is isomorphic to a free product $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z}$. It follows that SL(2) is weakly commensurable with F_2 and so $\mathrm{SL}(2, \mathbb{Z}) \sim_{\mathrm{qi}} F_n$ for any $n \geq 2$.

(iii) Suppose that S_g is a closed connected oriented surface of genus g. If $g \geq 2$, then one has a finite covering projection $S_g \to S_2$. It follows that the fundamental group $\Gamma_g := \pi_1(S_g)$ is a finite index subgroup of $\pi_1(S_2)$. So $\Gamma_g \sim_{\mathrm{qi}} \Gamma_2$ for $g \geq 2$. On the other hand, $S_g = \mathcal{H}/\Gamma_g$ when $g \geq 2$ where $\mathcal{H}$ is the Poincaré upper half space and Γ_g acts freely and properly discontinuously via isometries on $\mathcal{H}$. Applying Švarc-Milnor lemma, we see that $\Gamma_g \sim_{\mathrm{qi}} \mathcal{H}$ for ≥ 2. In the case of the torus $S_1 = \mathbb{S}^1 \times \mathbb{S}^1 = \mathbb{R}^2/\mathbb{Z}^2$, and, again by the Švarc-Milnor lemma, $\Gamma_1 \sim_{\mathrm{qi}} \mathbb{Z}^2 \sim_{\mathrm{qi}} \mathbb{R}^2$.

The group $\Gamma_g, g > 1$, is not quasi-isometric to $\mathbb{Z}^n$ for any n. The fact that Γ_g contains a non-abelian free group F implies that the number $b_k(\Gamma_g)$ of elements in a ball of radius k (with respect to a word metric) grows exponentially in k, whereas it grows at a polynomial rate in the case of $\mathbb{Z}^n$. It is known that the growth rate of the function $k \to b_k(\Gamma)$ of a finitely generated group (Γ, d_S) is a quasi-isometric invariant. This proves our assertion. See [1, Chapter I.8] for details.

It has been shown that if Γ is a finitely generated group which is virtually nilpotent, then Γ has polynomial growth. A major landmark result, whose proof due to Gromov [8] greatly influenced the development of geometric group theory, is the converse: *A finitely generated group* Γ *is virtually nilpotent if it has polynomial growth.*

2 Quasi-isometric Rigidity

When two finitely generated groups are quasi-isometrically equivalent, one would like to know how closely they are related as *algebraic* objects. Since weakly commensurable groups are quasi-isometric, one may ask whether it is possible to *recover* the group, up to weak commensurability, from its quasi-isomorphism type. This leads to the notions of quasi-isometric rigidity. We refer the reader to [1, Chapter I.8] and [7, §8.6] for detailed discussions on this topic.

There are several variants of rigidity, we consider only two: one version captures the idea that a rigid group is one with the property that any group quasi-isometric to it should be weakly commensurable to it. Another version is based on the idea that a rigid group is one which has a relatively ‘small’ quasi-isometry group.

Definition 2.1 (i) We say that a finitely generated group Γ with a word metric is *quasi-isometrically rigid.* if any finitely generated group quasi-isometric to Γ is weakly commensurable to Γ. (ii) Let $\mathcal{G}$ be a class of finitely generated groups which is closed under weak commensurability. We say that $\mathcal{G}$ is *quasi-isometrically rigid* if a finitely generated group Λ is quasi-isometric to a $\Gamma \in \mathcal{G}$, then $\Lambda \in \mathcal{G}$.

Obviously, the trivial group is quasi-isometrically rigid. For a non-trivial example, Bridson and Gersten [2] have shown that if a group is quasi-isometrically equivalent $\mathbb{Z}^n$, then it is virtually $\mathbb{Z}^n$; see [13] for a more general result. Thus $\mathbb{Z}^n$, $n \in \mathbb{N}$, is quasi-isometrically rigid. By applying a theorem of Stallings on the structure of groups with infinitely many ends, it can be shown that any group quasi-isometric to a finitely generated (non-abelian) free group is virtually free.

Example 2.2 (i) The class of all finitely presented groups is quasi-isometrically rigid; see [1, Chapter I.8, Prop. 8.24]. (ii) The class of all finitely generated virtually nilpotent groups is quasi-isometrically rigid, since, by Gromov’s polynomial growth theorem, any such group is virtually nilpotent. (iii) If Γ is rigid, then the class $\mathcal{G}(\Gamma)$ of all groups which are weakly commensurable with Γ is quasi-isometrically rigid.

Let Γ be a finitely generated group and let $\mathcal{A}(\Gamma)$ be the set of all isomorphisms $\phi : H_0 \to H_1$ where H_0, H_1 are arbitrary finite index subgroups of Γ. One has an equivalence relation on $\mathcal{A}(\Gamma)$ defined as $\phi \sim \psi$ where $\psi : H_0' \to H_1'$ if $\phi|_K = \psi|_K$ for some finite index subgroup $K \subset H_0 \cap H_1$. The equivalence classes are called virtual automorphisms of Γ and the set $\mathcal{A}(\Gamma)/\sim$ is denoted $\text{Vaut}(\Gamma)$. If Γ has no finite index subgroup, then $\text{Vaut}(\Gamma) = \text{Aut}(\Gamma)$. When Γ is residually finite, the group $\text{Vaut}(\Gamma)$ is particularly interesting. For example, it is not difficult to show that $\text{Vaut}(\mathbb{Z}^n) \cong \text{GL}(n; \mathbb{Q})$. Note that any isomorphism $\phi : H_0 \to H_1$ defines an element of $\mathcal{QI}(\Gamma)$ leading to a well-defined homomorphism $\text{Vaut}(\Gamma) \to \mathcal{QI}(\Gamma)$. In general this homomorphism is not surjective. For example, the linear action of $\text{GL}(n, \mathbb{R})$ on $\mathbb{R}^n$ yields an embedding of $\text{GL}(n, \mathbb{R})$ into $\mathcal{QI}(\mathbb{R}^n) \cong \mathcal{QI}(\mathbb{Z}^n)$ whereas $\text{Vaut}(\mathbb{Z}^n) = \text{GL}(n, \mathbb{Q})$.

Definition 2.3 (i) A finitely generated group Γ is said to be *strongly quasi-isometrically rigid* if the natural homomorphism $\text{Vaut}(\Gamma) \to \mathcal{QI}(\Gamma)$ is a surjection. (ii) A metric space X is quasi-isometrically rigid if $\text{Isom}(X) \to \mathcal{QI}(X)$ is an isomorphism of groups (Sec §3.4 [9]).

The infinite cyclic group is quasi-isometrically rigid but not strongly.

Before proceeding further, we recall some standard notions concerning lattices in semisimple Lie groups. The reader is referred to [14, 19] for detailed expositions.

2.1 Lattices in Semisimple Lie Group

Let G be a connected Lie group. One says that G is *simple* if it has no connected normal subgroup. G is called *semisimple* if G is an almost direct product $G_1 \cdots G_k$ where G_i are simple normal subgroups of G. Here almost direct product means that $G_i \cap G_j$ is a finite normal subgroup of G. When G is semisimple, $G_i \cap G_j$ is in fact contained in the centre of G. We will assume that G is a connected non-compact semisimple Lie group and that it has finite centre, denoted $Z(G)$. Let K be a maximal compact subgroup of G and then $X := G/K$ is connected and admits a G-invariant Riemannian metric. It is a globally symmetric space of non-compact type. Our assumption that G is non-compact implies that X has non-positive sectional curvature.

The real rank of a linear semisimple Lie group $G \subset \mathrm{GL}(N)$ may be defined as the maximum number d such that G has a diagonalizable subgroup isomorphic to $(\mathbb{R}^\times)^d$. When G is not linear, (example a non-trivial cover of $\mathrm{SL}(2, \mathbb{R})$) one defines its real rank to be that of $G/Z(G)$ (which is always linear). For example, the real rank of $\mathrm{SL}(n, \mathbb{R})$ equals $n - 1$ whereas the real rank of $\mathrm{SO}_0(p, q)$ is $\min p, q$. (Recall that $\mathrm{SO}_0(p, q)$ is the identity component of all linear transformations of $\mathbb{R}^{p+q}$ which preserve the quadratic form $\beta(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2$.)

A discrete subgroup Γ of G is called a *lattice* if the homogeneous space $\Gamma\backslash G$ carries a G-invariant measure with respect to which its volume is finite. For example, it is known that $\mathrm{SL}(n, \mathbb{Z}) \subset \mathrm{SL}(n, \mathbb{R})$ is a lattice. A lattice $\Gamma \subset G$ is *uniform* if $\Gamma\backslash G$ is compact; otherwise it is *nonuniform*. One says that a lattice $\Gamma \subset G$ is *reducible* if it contains infinite subgroups Γ_0, Γ_1 which generate a subgroup Λ isomorphic to an almost direct product $\Gamma_0 \cdot \Gamma_1$ such that Λ has finite index in Γ. We say that Γ is *irreducible* if it is not reducible. It can be shown that if Γ is a lattice in a non-compact semisimple Lie group G with finite centre and finitely many components, then it is weakly commensurable with a lattice in a connected Lie group with trivial centre.

Let Γ be an irreducible nonuniform lattice in a connected semisimple Lie group G. Then the centre $Z(\Gamma)$ of Γ is finite and $\Gamma/Z(\Gamma)$ is a lattice in $G/Z(G)$. Define the commensurator of Γ, $\mathrm{Comm}(\Gamma)$, to be the group $\{g \in G \mid \Gamma \text{ is commensurable with } g\Gamma g^{-1}\}$. It is clear that $\Gamma \subset \mathrm{Comm}(\Gamma)$.

If G has trivial centre and no compact factors and if Γ is an irreducible lattice, then: either Γ is non-arithmetic and the group $\mathrm{Comm}(\Gamma)$ is also a lattice in G, or, Γ is an 'arithmetic lattice' and $\mathrm{Comm}(\Gamma)$ is dense in G. This result due to Margulis. In the latter case, under a further hypothesis (namely that G *equals* the $\mathbb{R}$-points of a $\mathbb{Q}$-algebraic group), $\mathrm{Comm}(\Gamma)$ is the rational points $G(\mathbb{Q})$ of G. Thus, in this case, the group $\mathrm{Comm}(\Gamma)$ is a countable dense subgroup of G. See [19] for these results and for the definition of arithmetic lattices.

Restriction of the conjugation by an element $g \in \mathrm{Comm}(\Gamma)$ to $\Gamma \cap g\Gamma g^{-1}$ yields a well-defined element of $\mathcal{QI}(\Gamma)$ since $\Gamma \cap g\Gamma g^{-1} \hookrightarrow \Gamma$ is a quasi-isometry. This leads to a homomorphism $\mathrm{Comm}(\Gamma) \to \mathcal{QI}(\Gamma)$.

We are ready state the result concerning quasi-isometry group of lattices in semisimple Lie groups. The final result is the outcome work of several mathematicians. The beautiful survey article by Farb [5] outlines not only the proofs, but also the history of the problem including description of the contributions to this problem of the mathematicians whose work we have merely cited.

Theorem 2.4 (Schwartz *[16, 17],* Eskin *[3],* Farb and Schwartz *[6]*) *Suppose that* Γ *is an irreducible nonuniform lattice in a semisimple Lie group G with trivial centre and without compact factors. If G is not locally isomorphic to SL*$(2, \mathbb{R})$*, then* $\mathcal{QI}(\Gamma)$ *is isomorphic to Comm*(Γ).

When G is locally isomorphic to $\mathrm{SL}(2, \mathbb{R})$, then any nonuniform lattice Γ is virtually free. Thus $\mathcal{QI}(\Gamma) \cong \mathcal{QI}(F_2)$ and the conclusion of the above theorem fails. Denote by ∂F_2 the end space of the Cayley graph $\mathcal{C}(F_2)$. It is homeomorphic to the Cantor space. It is known that $\mathcal{QI}(F_2)$ is a certain group of homeomorphisms of the Cantor space known as the group of *quasisymmetric homeomorphisms*.

When Γ is uniform, it is quasi-isometric to $X = G/K$ by Švarc-Milnor lemma. It follows that $\mathcal{QI}(\Gamma) \cong \mathcal{QI}(X)$. It turns out that when G has real rank at least 2, $\mathrm{Isom}(X) \to \mathcal{QI}(X)$ is an isomorphism. This is also true when X is a quaternionic hyperbolic space $\mathrm{Sp}(n, 1)/K$ or the Cayley hyperbolic plane F_4/K.

Theorem 2.5 (Mostow *[12],* Tukia *[18],* Koryani and Reimann *[11],* Kleiner and Leeb *[10],* Eskin and Farb *[4],* Pansu *[13]*) *Suppose that* Γ *is a uniform lattice in a non-compact simple Lie group G such that either the real rank of G is at least* 2 *or* $X = G/K$ *is either a quaternionic hyperbolic the Cayley hyperbolic plane. Then* $\mathcal{QI}(\Gamma) \cong$ *Isom*(X).

We have left out irreducible uniform lattices in the rank 1 groups locally isomorphic to $\mathrm{SO}_0(n, 1)$, $n \geq 3$, or to $\mathrm{SU}(n, 1)$, $n \geq 2$. In these cases, the corresponding symmetric spaces X are real and complex hyperbolic spaces, denoted $\mathcal{H}^n$ and $\mathbb{C}\mathcal{H}^n$ respectively. Note that if Γ is any such lattice, then $\Gamma \sim_{\mathrm{qi}} X$ by the Švarc-Milnor lemma. Associated to X is its 'boundary' ∂X, which is homeomorphic to the sphere $\mathbb{S}^{n-1}$ or $\mathbb{S}^{2n-1}$ according as X is the real or complex hyperbolic n-space. It turns out that any quasi-isometry of X induces a homeomorphism of ∂X leading to a homomorphism $\mathcal{QI}(\Gamma) \cong \mathcal{QI}(X) \to \mathrm{Homeo}(\partial X)$. It is known that this is a monomorphism and the image is a certain group known as the group of *quasi-conformal group of homeomorphisms* of ∂X.

Finally we end this note with the following theorem:

Theorem 2.6 (Quasi-isometric rigidity for irreducible lattices) *Let G be a connected semisimple Lie group with trivial centre and without compact factors and let* Γ *be an irreducible lattice in G. If* Λ *is any finitely generated group that is quasi-isometric to* Γ*, then there exists a finite normal subgroup* $F \subset \Lambda$ *such that* Λ/F *is isomorphic to a lattice* Γ' *in G.*

References

1. Bridson, M.R., Haefliger, A.: Metric spaces of non-positive curvature. Grundlehren der Mathematischen Wissenschaften, vol. 319. Springer, Berlin (1999)
2. Bridson, M.R., Gersten, S.M.: The optimal isoperimetric inequality for torus bundles over the circle. Quart. J. Math. Oxford Ser. **47**(2)(185), 1–23 (1996)
3. Eskin, A.: Quasi-isometric rigidity of nonuniform lattices in higher rank symmetric spaces. J. Am. Math. Soc. **11**, 321–361 (1998)
4. Eskin, A., Farb, B.: Quasi-flats and rigidity in higher rank symmetric spaces. J. Am. Math. Soc. **10**, 653–692 (1997)
5. Farb, B.: The quasi-isometry classification of lattices in semisimple Lie groups. Math. Res. Lett. **4**, 705–717 (1997)
6. Farb, B., Schwartz, R.: The large-scale geometry of Hilbert modular groups. J. Diff. Geom. **44**, 435–478 (1996)
7. Druţu, C., Kapovich, M.: Geometric group theory. with an appendix by Bogdan Nica. In: American Mathematical Society Colloquium Publications, American Mathematical Society, Providence, RI, vol. 63 (2018)
8. Gromov, M.: Groups of polynomial growth and expanding maps. Inst. Hautes Études Sci. Publ. Math. No. **53**, 53–73 (1981)
9. Gromov, M., Pansu, P.: Rigidity of lattices: an introduction. In: Geometric Topology: recent Developments (Montecatini Terme, 1990), Lecture Notes in Math, vol. 1504, pp. 39–137. Springer, Berlin (1990)
10. Kleiner, B., Leeb, B.: Rigidity of quasi-isometries for symmetric spaces for symmetric spaces and Euclidean buildings. Inst. Hautes Études Sci. Publ. Math. **86**, 115–197 (1997)
11. Koranyi, A., Reimann, H.M.: Foundations for the theory of quasiconformal mappings on the Heisenberg group. Adv. Math. **111**, 1–87 (1995)
12. Mostow, G.D.: Quasiconformal mappings in n-space and rigidity of hyperbolic space forms. Publ. Inst. Hautes Études Sci. **34**, 53–104 (1968)
13. Pansu, P.: Metriques de Carnot-Caratheodory et quasiisometries des espaces symmetriques de rang un. Ann. Math. **129**, 1–60 (1989)
14. Raghunathan, M.S.: Discrete subgroups of Lie groups. Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge, vol. 68, Springer, Berlin (1972)
15. Sankaran, P.: On homeomorphisms and quasi-isometries of the real line. Proc. Am. Math. Soc. **134**(7), 1875–1880 (2006)
16. Schwartz, R.: The quasi-isometry classification of rank one lattices. Inst. Hautes Études Sci. Publ. Math. **82**(1995), 133–168 (1996)
17. Schwartz, R.: Quasi-isometric rigidity and diophantine approximation. Acta Math. **177**, 75–112 (1996)
18. Tukia, P.: Quasiconformal extension of quasisymmetric mappings compatible with a Möbius group. Acta Math. **154**, 153–193 (1985)
19. Zimmer, R.J.: Ergodic Theory and Semisimple Lie Groups. Birkhaüser, Boston (1984)

Existence of Compositional Square-Roots of Functions

V. Kannan

Abstract If f and g are two functions such that $g \circ g = f$, we say that g is a compositional square root of f. Here we discuss the question as to which maps admit a square root in this sense.

Keywords Compositional square root · Cycles · Piecewise monotonic maps · Piecewise linear maps · Critical point

Summary of Results

For interval maps and real maps, we have the following theorems:

(1) Every increasing map admits a square root.
(2) No decreasing map admits a square root.
(3) A bijection on any set admits a (possibly discontinuous) square root iff, for every even integer n or infinity, the number of cycles of that length is even or infinity.
(4) A quadratic polynomial $ax^2 + bx + c$ admits a square root iff $b^2 - 4ac \leq 2b$.
(5) A cubic polynomial $ax^3 + bx^2 + cx + d$ admits a square root iff it is strictly increasing, which is the case iff $b^2 \leq 3ac$.
(6) There is a surjective polynomial with exactly n critical points admitting a square root iff n is an even integer greater than or equal to 4.
(7) Every interval map f can be extended to another map g on a bigger interval such that $g \circ g = f$ on the smaller interval.
(8) Every interval map f can be extended to another map g on a bigger interval such that the extended map does not admit a square root.
(9) Every piecewise linear map from [0, 1] to [0, 1] that interchanges 0 and 1 and keeps the interior interval invariant, fails to admit a square root.

Of these nine results, (1), (2), (3) and (4) have been proved in [3]. While (3) is a folk result found easily on online sources, we provide a clearer proof here. The result (4)

A part of the work was done when the author was a Visiting Professor, C. R. Rao AIIMS, University of Hyderabad.

V. Kannan (✉)
SRM University Amaravati, Amaravati, Andhra Pradesh, India
e-mail: vksm.uoh@nic.in; kannan.v@srmap.edu.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_10

itself is not new as it follows as a special case of the results in [4]. Results (6), (8) and (9) are new and are proved in Sect. 3 here. While (7) is new, we state it without proof for the time being.

1 Preliminaries

A compositional square root of a map $f : X \to X$ is a map $g : X \to X$ such that $g \circ g = f$. In what follows, we sometimes omit the word compositional and call g simply a square root of f.

An interval map is a continuous self map of $I = [0, 1]$. Similarly, a real map is a continuous self map of the real line $\mathbb{R}$. The set of all fixed points of a function f is denoted by $Fix(f)$. For interval maps and real maps, this is always a closed subset of the domain.

If n is a positive integer, f^n always denotes the n-fold composite of f with itself (and not the pointwise product).
An element $x \in X$ is called a periodic point of f if there is a positive integer n such that $f^n(x) = x$.

For a map on I or $\mathbb{R}$, a critical point is a point c in every neighbourhood of which f fails to be one one.

A piecewise monotonic map, abbreviated henceforth as p.m. map, is one for which the domain can be divided into a finite number of subintervals on each of which it is monotonic. Equivalently, a map is piecewise monotonic iff it has only finitely many critical points.

Similarly, a piecewise linear map, abbreviated henceforth as p.l. map, is one for which the domain can be divided into a finite number of subintervals on each of which it is linear. Obviously all p.l. maps are p.m.

A subset E of the domain is said to be invariant with respect to f if $f(E) \subset E$. If f is a bijection, the full orbit of x is $\{f^n(x) | n \in \mathbb{Z}\}$.

2 Square Roots of Bijections

We first show that this question of the existence of square roots (or for that matter, of any nth roots) can be completely answered, at least for bijections, if we know the complete orbit structure of the function.

Theorem 1 *Let X be any set and $f : X \to X$ any bijection. Then f admits a square root on X iff the number of full orbits of any given non-odd size is non-odd (i.e., either even or infinity).*

The necessity part remains valid for any function $f : X \to X$:
If the map f admits a square root, then the following numbers cannot be odd, i.e. are either even or infinity:

1. for every even integer n, the number of cyclic orbits of that size
2. the number of full orbits that are infinite.

Proof

Necessity: Suppose f admits a compositional square root, that is, say there is a $g : X \to X$ such that $g \circ g = f$.

Consider the orbit of a point $x \in X$.

Case I: Suppose the orbit of x is finite, that is, there is $n \in \mathbb{N}$ such that $x, g(x), g(g(x)), ..., g^n(x) = x$ is an n-cycle.

(i). When n is odd, say $2k + 1$, the f-orbit of x becomes

$$x, f(x) = g^2(x), f^2(x) = g^4(x), \dots, f^k(x) = g^{2k}(x),$$

$$f^{k+1}(x) = g^{2k+2}(x) = g(x), f^{k+2}(x) = g^3(x), \dots, f^n(x) = x$$

Thus, the f-orbit of x, as a set, is the same as the g-orbit of x.

(ii). If n is even, say $2k$, the f-orbit of x is

$$x, f(x) = g^2(x), f^2(x) = g^4(x), \dots, f^k(x) = g^{2k}(x) = x$$

and the f-orbit of $g(x)$ is

$$f^{k+1}(x) = g^{2k+2}(x) = g(x), f^{k+2}(x) = g^3(x), \dots, f^{2k+1}(x) = g(x).$$

Thus, the g-orbit of x splits into two disjoint f-orbits, namely that of x and $g(x)$.

Coupled with the fact that every f-cycle has to be part of a g-cycle, this proves that the number of f-cycles of length $2k$ has to be equal to twice the number of g-cycles of length k.

Case II: Suppose the orbit of x is infinite.

The f-orbit of x is infinite iff the g-orbit of x is infinite as well.

Also, every infinite g-orbit $\{g^n(x)|n \in \mathbb{Z}\}$ splits into two f-orbits, namely $\{f^n(x)|n \in \mathbb{Z}\}$ and $\{f^n(f(x))|n \in \mathbb{Z}\}$.

This shows that the number of infinite orbits of f has to be even or infinity.

Note that this part of the proof is valid for all functions and do not use the bijectivity of f.

Sufficiency: Suppose it is true that, for every even integer n or infinity, the number of cyclic orbits of that size is either even or infinity.

We prove the theorem by constructing a $g : X \to X$ such that $g \circ g = f$.

Since f is bijective, the*full orbits* of x, defined by $\{f^n(x)|n \in \mathbb{Z}\}$, form a partition of X. To define g on the whole of X, it suffices to define g on each orbit constituting this partition. Our construction now proceeds in three steps:

Step I: Whenever there are two full orbits of the same size, say the orbits of x and y, we can define g such that g takes the orbit of x to the orbit of y and vice versa in the following manner:
$g(f^n(x)) = f^n(y)$ and $g(f^n(y)) = f^{n+1}(x)$ for all $n \in \mathbb{Z}$.

By thus shuttling between the orbits of x and y, we are able to produce a g, on the union of these two orbits, such that $g \circ g = f$, *whenever we have two orbits of the same size.*

The following figure illustrates this for two infinite cycles: (the dotted arrows are for g)

$$\begin{tikzcd}
\ldots \arrow[r] & f^{-1}(x) \arrow[r] & x \arrow[r] \arrow[d, dashed] & f(x) \arrow[r] \arrow[d, dashed] & f^2(x) \arrow[r] \arrow[d, dashed] & f^3(x) \arrow[r] & \ldots \\
\ldots \arrow[r] & f^{-1}(y) \arrow[r] \arrow[ur, dashed] & y \arrow[r] \arrow[ur, dashed] & f(y) \arrow[r] \arrow[ur, dashed] & f^2(y) \arrow[r] \arrow[ur, dashed] & f^3(y) \arrow[r] & \ldots
\end{tikzcd}$$

The following figure illustrates this for two finite cycles: (the dotted arrows are for g)

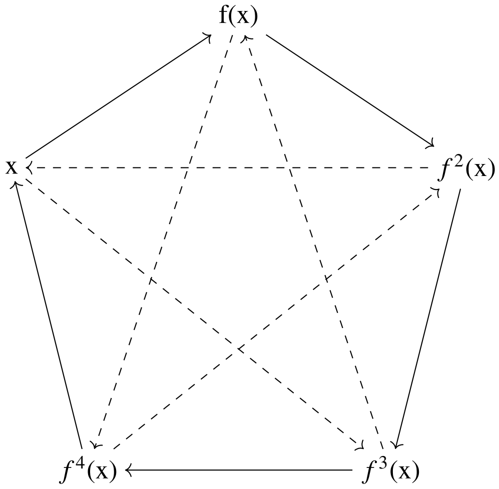
$$\begin{tikzcd}
x \arrow[r] \arrow[d, dashed] & f(x) \arrow[r] \arrow[d, dashed] & f^2(x) \arrow[r] \arrow[d, dashed] & f^3(x) \arrow[r] \arrow[d, dashed] & .....f^{q-1}(x) \arrow[r] \arrow[d, dashed] & f^p(x) = x \arrow[d, dashed] \\
y \arrow[r] \arrow[ur, dashed] & f(y) \arrow[r] \arrow[ur, dashed] & f^2(y) \arrow[r] \arrow[ur, dashed] & f^3(y) \arrow[r] & .....f^{p-1}(y) \arrow[r] \arrow[ur, dashed] & f^p(y) = y
\end{tikzcd}$$

Step II: Let x have a cyclic f-orbit of odd size, say $2n+1$. Then define g by $g^{2k}(x) = f^k(x)$ and $g^{2k+1}(x) = f^{n+k+1}(x)$ for all $k \leq n$.

Then, as can be easily verified, $g \circ g = f$.

Note that here we crucially make use of the oddness of the size of the orbit to construct g; had n been even, the two parts of the formula above would have been inconsistent and g could not have been defined.

The following figure illustrates this when $n = 2$: (the dotted arrows are for g)



Step III: If $f : X \to X$ is any bijection satisfying the hypothesis of our theorem, on each odd cycle of f, define g as in Step II. For every even n, the set of all orbits of size n divides into two equal parts, enabling us to produce a pairing of them. For each pair of such n-cycles, define g as in Step I.

Thus g has been defined on the whole of X satisfying $g \circ g = f$.

Put in another way, let $\tilde{X}$ be the quotient space of X consisting of the f-orbit decompositions. Our hypothesis ensures that there is a map $\sigma : \tilde{X} \to \tilde{X}$ such that

(1) σ preserves the orbit size.
(2) σ fixes all odd orbits and no even or infinite orbits, i.e., $\sigma(\tilde{x}) = \tilde{x}$ iff $\tilde{x}$ is an orbit of an $x \in X$ of odd size.
(3) $\sigma \circ \sigma =$ Identity on $\tilde{X}$.

Then σ gives the pairing $(\tilde{x}, \sigma(\tilde{x}))$ among f-orbits of even size. □

A similar result can be proved more generally for injective maps along similar lines. However, no such characterisation seems to be easily available for non-injective maps.

3 Its All About the Domain

In this section, we prove the results listed (6), (8) and (9) in the opening section.

We quote an interesting result from a recent Ph.D. thesis [1], also published as [2]:

Theorem 2 ([1], [2]) *Let f be an interval map. Let J be any interval containing $Fix(f \circ f)$. Then all periodic points of f lie in the interval $f(f(J))$.*

This result has an immediate consequence regarding the matter we have at hand:

Theorem 3 *Let f be an interval map such that it has a periodic point outside $f(J)$ where J is the smallest interval containing $Fix(f)$. Then f does not admit a square root.*

Proof Let, if possible, g be an interval map such that $g \circ g = f$. Then J contains $Fix(g \circ g)$, and therefore, by the above theorem, all periodic points of g should lie in $g(g(J))$. But f and g have the same set of periodic points as $g \circ g = f$. It follows that all periodic points of f should lie within J, contrary to the hypothesis.

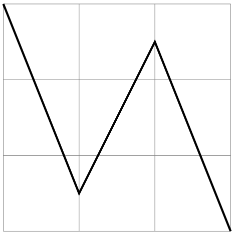
The above theorem can be used to give a lot of examples of interval maps that do not admit a square root. Here is one:

Example 1 *Let f be the piecewise linear map defined by $f(0) = 1$, $f\left(\frac{1}{3}\right) = \frac{1}{6}$, $f\left(\frac{2}{3}\right) = \frac{5}{6}$ and $f(1) = 0$. (i.e., f is linear on the intervals $\left[0, \frac{1}{3}\right]$, $\left[\frac{1}{3}, \frac{2}{3}\right]$, $\left[\frac{2}{3}, 1\right]$).*

In fact we actually have

$$f(x) = \begin{cases} 1 - \frac{5x}{2}, 0 \leq x < \frac{1}{3} \\ 2x - \frac{1}{2}, \frac{1}{3} \leq x \leq \frac{2}{3} \\ \frac{5-5x}{2}, \frac{2}{3} \leq x \leq 1 \end{cases} \tag{1}$$

The fixed points of this function are $\frac{2}{7}, \frac{1}{2}, \frac{5}{7}$, while it is easily observed that 0 and 1 are periodic points (they form what is called a 2-cycle). Here, the smallest interval containing all the fixed points is $J = \left[\frac{2}{7}, \frac{5}{7}\right]$, and $f(J) = \left[\frac{1}{6}, \frac{5}{6}\right]$. The periodic points 0 and 1 are outside $f(J)$. Therefore, by the above theorem, f does not admit a square root.



Graph of f

The idea we used can be crystallised in the following more general form:

Corollary 1 *No interval map f such that $f^{-1}(0) = \{1\}$ and $f^{-1}(1) = \{0\}$ admits a square root.*

A polynomial example can be obtained by translating and scaling $x - x^3$ on $[-\sqrt{2}, \sqrt{2}]$.

Our next theorem exploits this idea to establish that every interval map can be extended to a map on a bigger interval such that it does not admit a square root.

Theorem 4 *Let $c < a < b < d$. Then any map from $[a, b]$ to itself can be extended to a map from $[c, d]$ to itself such that this extension does not admit a square root.*

Proof Let $f : [a, b] \to [a, b]$ be the given map. Define its extension g on $[c, d]$ as

$$\begin{cases} g(x) = f(x) \text{ for all } x \in [a, b] \\ g(c) = d \\ g(d) = c \\ g \text{ is linear on } [c, a] \text{ and } [b, d]. \end{cases} \tag{2}$$

On the one hand, c and d, being mapped to each other, are periodic points (what is called a 2-cycle). On the other, because the diagonal line does not meet this graph outside the inner square $[a, b] \times [a, b]$, it is clear that g has no fixed points outside $[a, b]$. Also the smallest interval containing $Fix(g)$ is contained in $[a, b]$. It follows that $g(g(J)) = f(J) \subset [a, b]$, whereas the periodic points c and d are outside this interval. Therefore, by our previous theorem, g does not admit a square root. □

Hence the adage that there are plenty of maps without a square root.

We now present a result in the opposite direction: roughly speaking, every interval map admits a square root, provided some minimal margin of space for manipulation.

Theorem 5 *Let $c \leq a < b < d$. Let $f : [a, b] \to [a, b]$ be continuous. Then there is a continuous extension $g : [c, d] \to [c, d]$ such that $g \circ g = f$ on $[a, b]$.*

Proof First choose an element r such that $b < r < d$. Next, choose p and q such that the map $px + q$ takes $[a, b]$ onto $[r, d]$ (taking a to r and b to d). More specifically, $p = \frac{d-r}{b-a}$ and $q = \frac{br-ad}{b-a}$.

Take g to be any map from $[c, d]$ to $[c, d]$ satisfying the following:
$g(t) = pf(t) + q$, if $a \leq t \leq b$, and
$g(t) = \frac{t-q}{p}$, if $r \leq t \leq d$.

To be more specific, we can take g to be linear on $[b, r]$ and constant on $[c, a]$. We particularly note that

$$g(r) = \frac{r-q}{p} = \frac{r - \frac{br-ad}{b-a}}{\frac{d-r}{b-a}} = a \text{ and}$$

$$g(d) = \frac{d-q}{p} = \frac{d - \frac{br-ad}{b-a}}{\frac{d-r}{b-a}} = b$$

and that the linear map $\frac{t-q}{p}$ is the inverse of the map $pt + q$. This observation is crucial to complete the proof as follows:
Here g takes $[a, b]$ to $[r, d]$ and brings back $[r, d]$ to $[a, b]$, and while doing so,

$$g(t) = pf(t) + q \text{ and}$$

$$g(g(t)) = \frac{pf(t) + q - q}{p} = f(t)$$

holds for all $t \in [a, b]$. This g satisfies $g \circ g = f$ on $[a, b]$ as promised. □

Note that f may not admit a square root without enlarging the domain; think of any of the examples in the negative results above.

References

1. Archana, M.: Forcing relation and conjugacy classification for a class of interval maps. Ph.D. Thesis, University of Hyderabad (2018)
2. Archana, M., Kannan, V.: Intervals containing all periodic points. R. Anal. Exch. 263–270 (2017)
3. Kannan, V., Unnikrishnan, N.: Compositional square roots. Pre-print
4. Rice, R.E., Schweizer, B., Sklar, A.: When is $az^2 + bz + c$ of the form $f(f(z))$? American Mathematical Monthly. **87**, 252–263 (1980)

On Generalized Statistical Convergence and Strongly Summable Functions of Weight *g*

Ekrem Savaş and Rabia Savaş

Abstract In this paper, by using a nonnegative real-valued Lebesque measurable function in the interval $(1, \infty)$ we introduce the concepts of strong (V, λ, p)-summability and λ-statistical convergence of weight $g : [0, \infty) \to [0, \infty)$ where $g(x_n) \to \infty$ for any sequence (x_n) in $[0, \infty)$ with $x_n \to \infty$. We also examine some relations between λ- statistical convergence of weight g and strong (V, λ, p)-summability of weight g.

Keywords Lebesque measurable function · Weight g · Statistical convergence · Strong summability

2000 Mathematics Subject Classification Primary 42B15 · Secondary 40C05

The idea of statistical convergence is found in Zygmund [1], the first edition of his published monograph in Warsaw in 1935. The concept of statistical convergence was introduced by Fast [2] and then reintroduced by Schoenberg [3] independently and later on studied by many authors. The active researches on this area were interesting after the paper of Fridy [4]. It has been investigated in a number of papers ([4–10]). Most of the existing work on statistical convergence appears to have been restricted to real or complex sequences except the works of Kolk, Maddox, and Cakalli. This notion was used by Kolk in [8] to extend statistical convergence to normed spaces; by Maddox [9] to extend to locally convex Hausdorff topological linear spaces; and used by Cakalli [11] to extend to topological Hausdorff groups. Moreover, statistical convergence is closely related to the concept of convergence in probability.

E. Savaş (✉)
Usak University, Usak, Turkey
e-mail: ekremsavas@yahoo.com

R. Savaş
Sakarya University, sakarya, Turkey
e-mail: rabiasavas@hotmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_11

1 Definition and Preliminaries

Spaces of strongly summable sequences were discussed by Kuttner [12], Maddox [13] and others. The definitions of statistical convergence and strong p-Cesàro convergence of a sequence of real numbers were introduced in the literature independently of one another and followed different lines of development since their first appearance. It turns out, however, that the two definitions can be simply related to one another in general and are equivalent for bounded sequences. The idea of statistical convergence depends on the density of subsets of the set $\mathbb{N}$ of natural numbers. The density of a subset E of $\mathbb{N}$ is defined by

$$\delta(E) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \chi_E(k) \text{ provided the limit exists,}$$

where χ_E is the characteristic function of E. It is clear that any finite subset of $\mathbb{N}$ has zero natural density and $\delta(E^c) = 1 - \delta(E)$.

A sequence $x = (x_k)$ is said to be statistically convergent to L if for every $\varepsilon > 0$, $\delta(\{k \in \mathbb{N} : |x_k - L| \geq \varepsilon\}) = 0$. In this case we write $x_k \overset{stat}{\to} L$ or $S - \lim x_k = L$. The set of all statistically convergent sequences will be denoted by S. The more general idea of λ-statistical convergence was introduced by Mursaleen in [14]. Subsequently a lot of interesting investigations have been done by various authors on several related notions of statistical convergence(see for example [15–20].

The order of statistical convergence of a sequence of numbers was given by Gadjiev and Orhan in [21]. Later, in [15] and [16], the notions of statistical convergence of order α and λ-statistical convergence of order α were introduced and studied, respectively. The notions of λ-statistical convergence and strongly λ-summable function of order α were introduced and studied by Et et al. ([22]).

Let $\lambda = (\lambda_m)$ be a non-decreasing sequence of positive numbers tending to ∞ such that $\lambda_{m+1} \leq \lambda_m + 1,\ \lambda_1 = 1$. The collection of such a sequence λ will be denoted by Δ.

The generalized de Valée-Pousin mean is defined by

$$t_m\left(x\right) = \frac{1}{\lambda_m} \sum_{k\in I_m} x_k,$$

where $I_m = [m - \lambda_m + 1, m]$, for $m = 1, 2, \ldots$.

A sequence $x = (x_k)$ is said to be (V, λ)-summable to a number L if $t_m(x) \to L$ as $m \to \infty$. If $\lambda_m = m$, then (V, λ)-summability is reduced to Cesàro summability.

Mursaleen defined λ-statistical convergence as follows:

A sequence $x = (x_k)$ is said to be λ-statistically convergent or S_λ-convergent to L if for every $\varepsilon > 0$

$$lim_m \frac{1}{\lambda_m} |\{k \in I_m : |x_k - L| \geq \epsilon\}| = 0,$$

In this case we write $S_\lambda - limx = L$ or $x_k \to L(S_\lambda)$.

Strongly summable functions are introduced and studied by Borwein [23] . A nonnegative real-valued Lebesque measurable function $x(t)$ in the interval $(1, \infty)$ is said to be strongly summable to L if

$$\lim_{m\to\infty} \frac{1}{m} \int_1^m |x(t) - L|^p dt = 0, \quad 1 \le p < \infty.$$

$[W_p]$ will denote the space of real-valued Lebesque measurable function in the interval $(1, \infty)$. The space $[W_p]$ is a normed space with the norm

$$\|x\| = \sup_{m\ge1} \left(\frac{1}{m} \int_1^m |x(t)|^p dt \right)^{\frac{1}{p}}.$$

In this paper, using the notion of (V, λ)-summability and λ -statistical convergence, we introduce and study the concepts of strong (V, λ, p)-summability and λ-statistical convergence of weight g of real-valued Lebesque measurable functions $x(t)$ in the interval $(1, \infty)$.

Throughout by function $x(t)$ we shall mean a nonnegative real-valued Lebesque measurable function in the interval $(1, \infty)$ and we will consider functions $g : [0, \infty) \to [0, \infty)$ such that $g(x_n) \to \infty$ if $x_n \to \infty$. The class of all such functions will be denoted by G.

2 Main Results

In this section, we give the main results of this paper. Before establishing the main theorems we have some definitions.

Definition 2.1 Let $\lambda = (\lambda_m) \in \Delta$ and $g \in \mathbf{G}$. A function $x(t)$ is said to be strongly (V, λ, p)-summable of weight g (or $[W^g_{\lambda p}]$-summable) if there is a number L such that

$$\lim_{m\to\infty} \frac{1}{g(\lambda_m)} \int_{m-\lambda_m+1}^m | x(t) - L|^p dt = 0, \quad 1 \le p < \infty,$$

where $I_m = [m - \lambda_m + 1, m]$. In this case we write $[W^g_{\lambda p}] - \lim x(t) = L$. The set of all strongly (V, λ, p)-summable functions of weight g will be denoted by $[W^g_{\lambda p}]$. For $\lambda_m = m$ for all $m \in \mathbb{N}$, we shall write $[W^g_p]$ instead of $[W^g_{\lambda p}]$.

Definition 2.2 Let $\lambda = (\lambda_m) \in \Delta$ and $g \in \mathbf{G}$. A function $x(t)$ is said to be λ-statistically convergent of weight g(or $[S^g_\lambda]$ -statistical convergence) to a number L for every $\varepsilon > 0$,

$$\lim_m \frac{1}{g(\lambda_m)} |\{t \in I_m : |x(t) - L| \ge \varepsilon\}| = 0.$$

The set of all λ-statistically convergent functions of weight g will be denoted by $[S_\lambda^g]$. In this case we write $[S_\lambda^g] - \lim x(t) = L$. For $\lambda_m = m$, for all $m \in \mathbb{N}$, we shall write $[S^g]$ instead of $[S_\lambda^g]$.

The following theorem gives the algebraic characterization of nonnegative real-valued Lebesque measurable functions.

Theorem 2.3 *Let $g \in \mathbf{G}$ and $x(t)$ and $y(t)$ be nonnegative real-valued Lebesque measurable functions in the interval $(1, \infty)$, then*

(1) If $[S_\lambda^g] - \lim x(t) = L$ and $c \in \mathbb{R}$, then $[S_\lambda^g] - \lim(cx(t)) = cL$
(2) If $[S_\lambda^g] - \lim x(t) = L_1$ and $[S_\lambda^g] - \lim y(t) = L_2$, then $[S_\lambda^g] - \lim(x(t) + y)t))$ $= L_1 + L_2$

Proof (i) For $c = 0$, the result holds easily. Let $c \neq 0$. Now we write that

$$\frac{1}{g(\lambda_m)} |\{k \in I_m : |cx(t) - cL| \geq \varepsilon\}| = \frac{1}{g(\lambda_m)} \left| \left\{ k \in I_m : |x(t) - L| \geq \frac{\varepsilon}{|c|} \right\} \right|$$

and the result follows.

(ii) The result follows from the fact that

$$\frac{1}{g(\lambda_m)} |\{k \in I_m : |x(t) + y(t) - (L_1 + L_2)| \geq \varepsilon\}|$$
$$\leq \frac{1}{g(\lambda_m)} \left|\{k \in I_m : |x(t) - L_1| \geq \frac{\varepsilon}{2}\}\right| + \frac{1}{g(\lambda_m)} \left|\{k \in I_m : |y(t) - L_2| \geq \frac{\varepsilon}{2}\}\right|.$$

□

Theorem 2.4 *Let $\lambda = (\lambda_m) \in \Delta$ and $g_1, g_2 \in G$ be such that there exist $M > 0$ and $r \in \mathbb{N}$ such that $\frac{g_1(\lambda_m)}{g_2(\lambda_m)} \leq M$ for all $m \geq r$. Then $[S_\lambda^{g_1}] \subset [S_\lambda^{g_2}]$.*

Proof Observe that,

$$\frac{1}{g_2(\lambda_m)} |\{k \in I_m : |x(t) - L| \geq \varepsilon\}| = \frac{g_1(\lambda_m)}{g_2(\lambda_m)} \cdot \frac{1}{g_1(\lambda_m)} |\{k \in I_m : |x(t) - L| \geq \varepsilon\}|$$
$$\leq M \cdot \frac{1}{g_1(\lambda_m)} |\{k \in I_m : |x(t) - L| \geq \varepsilon\}|$$

for all $m \geq r$. If $x(t)) \in \left[S_\lambda^{g_1}\right]$ then the right hand side tends to zero for every $\varepsilon > 0$ and consequently

$$\frac{1}{g_2(\lambda_m)} |\{k \in I_m : |x(t) - L| \geq \varepsilon\}| = 0$$

and so $x(t) \in \left[S_\lambda^{g_2}\right]$. Therefore $\left[S_\lambda^{g_1}\right] \subseteq \left[S_\lambda^{g_2}\right]$. □

We have the following result from theorem 3.4

Corollary 2.5 *Let $\lambda = (\lambda_m) \in \Delta$ and $x(t)$ be a nonnegative real-valued Lebesque measurable function in the interval $(1, \infty)$. In particular if $g \in \mathbf{G}$ be such that there exist $M > 0$ and a $r \in \mathbb{N}$ such that $m/g(\lambda_m) \leq M$ for all $m \geq r$ then $[S_\lambda^g] \subseteq [S_\lambda]$.*

Theorem 2.6 *Let $\lambda = (\lambda_m) \in \Delta$ and $g \in \mathbf{G}$. $[S] \subseteq [S_\lambda^g]$ if $\liminf_{m\to\infty} \dfrac{g(\lambda_m)}{m} > 0$.*

Proof For any $\varepsilon > 0$, we have

$$\{k \leq m : |x(t) - L| \geq \varepsilon\} \supseteq \{k \in I_m : |x(t) - L| \geq \varepsilon\}.$$

Hence it follows that for $m \in \mathbb{N}$

$$\begin{aligned}\frac{1}{m}\left|\{k \leq m : |x(t) - L| \geq \varepsilon\}\right| &\geq \frac{1}{m}\left|\{k \in I_m : |x(t) - L| \geq \varepsilon\}\right| \\ &\geq \frac{g(\lambda_m)}{m}\cdot\frac{1}{g(\lambda_m)}\left|\{k \in I_m : |x(t) - L| \geq \varepsilon\}\right|.\end{aligned}$$

If $x \to L(S)$, then $\frac{1}{m}\left|\{k \leq m : |x(t) - L| \geq \varepsilon\}\right| \to 0$ as $m \to \infty$, and so

$$\frac{1}{g(\lambda_m)}\left|\{k \in I_m : |x(t) - L| \geq \varepsilon\}\right| \to 0$$

as $m \to \infty$. This gives that $x(t) \to L\left[S_\lambda^g\right]$. □

In view of Theorem 3.4, we state the following result without proof

Theorem 2.7 *Let $\lambda = (\lambda_m) \in \Delta$. If $g_1, g_2 \in \mathbf{G}$ be such that there exist $M > 0$ and a $r \in \mathbb{N}$ such that $g_1(\lambda_m)/g_2(\lambda_m) \leq M$ for all $m \geq r$ then $\left[V_{\lambda p}^{g_1}\right] \subseteq \left[V_{\lambda p}^{g_2}\right]$.*

We have the following result from theorem 3.7.

Corollary 2.8 *Let $\lambda = (\lambda_m) \in \Delta$ and p be a positive real number, then $[W_{\lambda p}^{g_1}] \subseteq [W_{\lambda p}^{g_2}]$.*

Theorem 2.9 *Let $\lambda = (\lambda_m) \in \Delta$ and p be a positive real number. Let $g_1, g_2 \in \mathbf{G}$ be such that there exist $M > 0$ and a $r \in \mathbb{N}$ such that $g_1(\lambda_m)/g_2(\lambda_m) \leq M$ for all $m \geq r$ and let $0 < p < \infty$. If a function $x(t)$ is strongly $(V, \lambda, p)-$ summable of weight g_1 to L then it is $\lambda-$ statistically convergent of weight g_2 to L i.e $\left[V_{\lambda p}^{g_1}\right] \subset S_\lambda^{g_2}$.*

Proof Let function $(x(t)) \in \left[V_{\lambda p}^{g_1}\right]$ and let $\varepsilon > 0$ be given. We observe that

$$\begin{aligned}\int\limits_{k\in I_m} |x(t)-L|^p\,dt &= \int\limits_{\substack{k\in I_m\\ |x(t)-L|\geq\varepsilon}} |x(t)-L|^p\,dt + \int\limits_{\substack{k\in I_m\\ |x(t)-L|<\varepsilon}} |x(t)-L|^p\,dt\\ &\geq \int\limits_{\substack{k\in I_m\\ |x(t)-L|\geq\varepsilon}} |x(t)-L|^p\,dt\\ &\geq |\{k\in I_m : |x(t)-L|\geq\varepsilon|\,.\varepsilon^p.\end{aligned}$$

Now, it follows that

$$\begin{aligned}\frac{1}{g_1(\lambda_m)}\int\limits_{k\in I_m} |x(t)-L|^p &\geq \frac{1}{g_1(\lambda_m)}\,|\{k\in I_m : |x(t)-L|\geq\varepsilon|\,.\varepsilon^p\\ &= \frac{g_2(\lambda_m)}{g_1(\lambda_m)}\cdot\frac{1}{g_2(\lambda_m)}\,|\{k\in I_m : |x(t)-L|\geq\varepsilon|\,.\varepsilon^p\\ &\geq \frac{1}{M}\cdot\frac{1}{g_2(\lambda_m)}\,|\{k\in I_m : |x(t)-L|\geq\varepsilon|\,.\varepsilon^p\end{aligned}$$

for all $m\geq r$. If $x(t)\to L(\left[V_{\lambda p}^{g_1}\right])$, then the left hand side tends to zero and consequently the right hand side also tends to zero . Hence $x(t)\to L\left[S_\lambda^{g_2}\right]$. □

Corollary 2.10 *If $g\in\mathbf{G}$ be such that there exist $M>0$ and a $r\in N$ such that $\frac{m}{g(\lambda_m)}\leq M$ for all $m\geq r$ and $0<p<\infty$, then $\left[V_{\lambda p}^{g}\right]\subseteq S_\lambda$.*

Theorem 2.11 *Let $\lambda=(\lambda_m)$ and $\mu=(\mu_m)$ be two sequences in Δ such that $\lambda_m\leq\mu_m$ for all $m\in\mathbb{N}$, and $g_1,g_2\in\mathbf{G}$. If*

$$\lim_{m\to\infty}\inf\frac{g_1(\lambda_n)}{g_2(\mu_m)}>0, \tag{2.1}$$

then $[S_\mu^{g_2}]\subseteq[S_\lambda^{g_1}]$.

Proof Suppose that $\lambda_m\leq\mu_m$ for all $m\in\mathbb{N}$ and let (3.1) be satisfied. Then $I_m\subset J_n$ and so that $\varepsilon>0$ we may write

$$\{t\in J_m : |x(t)-L|\geq\varepsilon\}\supset\{t\in I_m : |x(t)-L|\geq\varepsilon\}$$

and so

$$\frac{1}{g_2(\mu_m)}\,|\{t\in J_m : |x(t)-L|\geq\varepsilon\}|\geq\frac{g_1(\lambda_m)}{g_2(\mu_m)}\frac{1}{g_1(\lambda_m)}\,|\{t\in I_m : |x(t)-L|\geq\varepsilon\}|$$

for all $m\in\mathbb{N}$, where $J_m=[m-\mu_m+1,m]$. Now, taking the limit as $n\to\infty$ in the last inequality and using (3.1), we get $[S_\mu^{g_2}]\subseteq[S_\lambda^{g_1}]$. □

From Theorem 3.11, we have the following results.

Corollary 2.12 *Let $\lambda = (\lambda_m)$ and $\mu = (\mu_m)$ be two sequences in Δ such that $\lambda_m \leq \mu_m$ for all $n \in \mathbb{N}$. If (2.1) holds, then*

(1) $[S_{\mu}^{g_1}] \subseteq [S_{\lambda}^{g_1}]$,
(2) $[S_{\mu}] \subseteq [S_{\lambda}^{g_1}]$,
(3) $[S_{\mu}] \subseteq [S_{\lambda}]$.

Theorem 2.13 *Let $\lambda = (\lambda_m)$ and $\mu = (\mu_m)$ be two sequences in Δ such that $\lambda_m \leq \mu_m$ for all $m \in \mathbb{N}$ and $g_1, g_2 \in \mathbf{G}$. If (2.1) holds, then $[W_{\mu p}^{g_2}] \subseteq [W_{\lambda p}^{g_1}]$*

Proof The proof of the theorem is straightforward, thus omitted. □

Theorem 2.14 *Let $\lambda = (\lambda_m)$ and $\mu = (\mu_m)$ be two sequences in Δ such that $\lambda_m \leq \mu_m$ for all $m \in \mathbb{N}$. Let $g_1, g_2 \in \mathbf{G}$ and (2.1) holds. If a real-valued function $x(t)$ is strongly (V, μ, p)-summable of weight g_2 to L, then it is λ-statistically convergent of of weight g_1 to L.*

Proof Let $x(t)$ be a real-valued function such that $x(t)$ is strongly (V, μ, p)-summable of weight $g_2$to L and $\varepsilon > 0$. Then we have

$$\begin{aligned}\int_{t \in J_m} |x(t) - L|^p dt &= \int_{\substack{t \in J_m \\ |x(t)-L| \geq \varepsilon}} |x(t) - L|^p dt + \int_{\substack{t \in J_m \\ |x(t)-L| < \varepsilon}} |x(t) - L|^p dt \\ &\geq \int_{\substack{t \in I_m \\ |x(t)-L| < \varepsilon}} |x(t) - L|^p dt \\ &\geq |\{t \in I_m : |x(t) - L| \geq \varepsilon\}| \cdot \varepsilon^p\end{aligned}$$

and so that

$$\frac{1}{g_2(\mu_n)} \int_{t \in J_m} |x(t) - L|^p dt \geq \frac{g_1(\lambda_m)}{g_2(\mu_m)} \frac{1}{g_1(\lambda_m} |\{t \in I_m : |x(t) - L| \geq \varepsilon\}| \cdot \varepsilon^p.$$

Since (2.1) holds, it follows that if $x(t)$ is strongly (V, μ, p)-summable of weight g_2 to L, then it is λ-statistically convergent of weight g_1 to L. □

We have the following corollary.

Corollary 2.15 *Let $\lambda = (\lambda_m)$ and $\mu = (\mu_m)$ be two sequences in Δ such that $\lambda_m \leq \mu_m$ for all $n \in \mathbb{N}$. Let $g_1, g_2 \in \mathbf{G}$. If (2.1) holds, then*

(1) A real-valued function $x(t)$ is strongly (V, μ, p)-summable of weight g_1 to L, then it is λ-statistically convergent of weight g_1 to L.
(2) A real-valued function $x(t)$ is strongly (V, μ, p)-summable to L, then it is λ-statistically convergent of weight g_1 to L.
(3) A real-valued function $x(t)$ is strongly (V, μ, p)-summable to L , then it is λ-statistically convergent to L.

References

1. Zygmund, A.: Trigonometric Series. Cambridge University Press, Cambridge (1979)
2. Fast, H.: Sur la convergence statistique. Colloq. Math. **2**, 41–44 (1951)
3. Schoenberg, I.J.: The integrability of certain functions and related summability methods. Amer. Math. Monthly **66**, 361–375 (1959)
4. Fridy, J.A.: On statistical convergence. Analysis **5**, 301–313 (1985)
5. Fridy, J.A., Miller, H.I.: A matrix characterization of statistical convergence. Analysis **11**, 59–66 (1991)
6. Fridy, J.A., Orhan, C.: Lacunary statistical convergence. Pacific J. Math. **160**(1), 43–51 (1993)
7. Salat, T.: On statistically convergent sequences. Math. Slovaca **30**, 139–150 (1980)
8. Kolk, E.: The statistical convergence in Banach spaces. Tartu Ulikooli Toimetised, Acta et Commentationes Universitatis Tartueneis **928**, 4152 (1991)
9. Maddox, I.J.: Statistical convergence in locally convex spaces. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 104, pp. 141–145 (1988)
10. Nuray, F. λ- strongly summable and λ statistically convergent functions, Iranian J. Sci. Tech. Trans. A, vol. 34, no. A4, pp. 335–339 (2010)
11. Cakalli, H.: On statistical convergence in topological groups. Pure Appl. Math. Sci. **43**(1–2), 27-31 (1996)
12. Kuttner, B.: Note on strong summability. J. London Math. Soc. **21**, 118–122 (1946)
13. Maddox, I.J.: On strong almost convergence. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 85, no. 2, pp. 345–350 (1979)
14. Mursaleen, M.: λ-statistical convergence. Math. Slovaca **50**(1), 111–115 (2000)
15. R. Çolak, Statistical convergence of order α, Modern methods in Analysis and its Applications, pp. 121-129. Anamaya Pub., New Delhi, India (2010)
16. Çolak, R., Bektaş, C.A.: λ-statistical convergence of order α. Acta Math. Scientia **31B**(3), 953–959 (2011)
17. Malkowsky, E., Savaş, E.: Some λ-sequence spaces defined by a modulus. Arch. Math. (Brno) **36**, 219–228 (2000)
18. Savaş, E.: Generalized summability methods of functions using ideals. In: AIP Conference Proceedings, vol. 1676, 020009 (2015). https://doi.org/10.1063/1.4930435
19. Savaş, E.: Strong almost convergence and almost λ-statistical convergence. Hokkaido Math. J. **29**(3), 531–536 (2000)
20. Savaş, E., Savaş, R.: Some λ- sequence spaces defined by Orlicz functions. Indian J. Pure. Appl. Math. **34**(12), 1673–1680 (2003)
21. Gadjiev, A.D., Orhan, C.: Some approximation theorems via statistical convergence. Rocky Mountain J. Math. **32**(1), 508–520 (2002)
22. Et, M., Mohiuddine, S.A., Alotaibi, A.: On λ-statistical convergence and strongly λ-summable functions of order α, J. Ineq. Appl. 2–8 (2013)
23. Borwein, D.: Linear functionals with strong Cesaro Summability. J. Lond. Math. Soc. **40**, 628–634 (1965)

On $\mathcal{I}$-Statistically Order Pre Cauchy Sequences in Riesz Spaces

Pratulananda Das and Ekrem Savas

Abstract In this paper. we continue to investigate in line of the recent work of Sencimen and Pehlivan, Das and Savas and consider the notion of $\mathcal{I}$-statistical order pre Cauchy condition related to a new type of order convergence, namely $\mathcal{I}$-statistical order convergence in Riesz spaces and establish some of its basic properties. We mainly investigate their inter-relationship.

Keywords Ideal · Filter · Riesz spaces · $\mathcal{I}$-statistical order convergence · $\mathcal{I}$-statistical order pre Cauchy condition

Mathematics Subject Classification (2010) 40A05 · 46B42

1 Introduction

The concept of Riesz spaces or vector lattices was first introduced by Riesz in [1]. After the significant advancements by Freudenthal [2] and Kantorovich [3] in this field, several mathematicians have over the years developed the subject. Riesz spaces occur very naturally during many studies in analysis as also they play important role in other branches of Mathematics like optimization, problems of Banach spaces, measure theory and operator theory as also have applications in economics (see [4]).

One of the fundamental concepts in the study of Riesz spaces is "order convergence" from which the idea of order continuity comes. The significance of order convergence is that it is not generally a "topological concept" per say i.e. it does not

P. Das (✉)
Department of Mathematics, Jadavpur University,
Kolkata 700032, West Bengal, India
e-mail: pratulananda@yahoo.co.in

E. Savas
Department of Mathematics, Istanbul Commerce University,
Sütlüce-Istanbul, Turkey
e-mail: ekremsavas@yahoo.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_12

necessarily correspond to the notion of convergence with respect to a topology. One can consult the book by Zaanen [5] where a detailed investigation of this matter is given. Very recently the notion of order convergence has been extended to statistical order convergence by Sencimen and Pehlivan [6].

Fast [7] and Schoenberg [8] independently introduced the notion of statistical convergence, which is an extension of the idea of usual convergence and consequently its topological implications were studied first by Fridy [9] and Šalát [10] (also later by Maddox [11]). Connor [12] established some interesting properties of this convergence. Di Maio and Kočinac [13] investigated the same concept in topological spaces and statistical Cauchy condition in uniform spaces. Pehlivan and Albayrak [14, 15] studied statistical convergence of sequences in locally solid Riesz spaces and more importantly the idea of order statistical convergence was studied in Riesz spaces by Sencimen and Pehlivan [6] in which we are interested here. One can see [16, 17] for more recent developments in this direction. Also more investigations on generalized convergences have been done using filters in related structures like (l)-groups in [18–21].

In [22] Connor et al. introduced a very interesting Cauchy condition called pre-statistical Cauchy condition associated with the notion of statistical convergence and it was observed that statistically convergent sequences of real numbers are always statistically pre-Cauchy. On the other hand, under certain general conditions, the converse is also true whose importance lies in the fact that one do not has to guess the limit of the statistically convergent sequence beforehand.

In [23] the notion of ideals was used to extend the idea of statistical convergence to $\mathcal{I}$-convergence with many interesting consequences and most importantly it turned out to be one of the most general type of convergence. More investigations in this direction and more applications of ideals can be found in [16, 17, 24–31] where many important references can be found.

Very recently in [27] we used ideals to unify the above two approaches and introduced the concepts of $\mathcal{I}$-statistical convergence and $\mathcal{I}$-lacunary statistical convergence for sequences of real numbers and investigated their basic properties. Subsequently the notion of $\mathcal{I}$-statistically pre-Cauchy sequences of real numbers have been investigated in [28].

Finally it should be mentioned that over the years a lot of work has been done to study these general summability methods in different structures in order to enhance their applicability. In a step forward in this direction we extended the ideas of order and monotone convergence in Riesz spaces and introduced the ideas of $\mathcal{I}$-lacunary statistical order and monotone convergence in a Riesz space and established some of its properties by using the mathematical tools of the theory of Riesz spaces and basic properties of order sequences in [17]. As a natural consequence in this paper we study the notions of $\mathcal{I}$-statistical order convergence and $\mathcal{I}$-statistically order pre-Cauchy sequences and establish some properties. It should be noted that the methods of proofs are not exactly analogous to those of [27] or [28] and are more complicated.

2 Preliminaries

We first recall some basic notions of Riesz spaces from [4] (see also [6]).

Definition 2.1 A real vector space E (with elements $x, y, \ldots$) with a partial order $\leq$ is called an ordered vector space if E is partially ordered in such a way that the vector space structure and the order structure are compatible, i.e.

(i) $x \leq y$ implies $x + z \leq y + z \; \forall z \in E$,
(ii) $x \geq \theta$ implies $\alpha x \geq \theta \; \forall \alpha \geq 0$ in $\mathbb{R}$, where θ is the null element with respect to addition.

If, in addition, E is a lattice with respect to partial ordering, then E is called a Riesz space or also a vector lattice.

Let E be a Riesz space. For any $x \in E$, we write $x^{+} = x \vee \theta$, $x^{-} = (-x) \vee \theta$ and $|x| = x \vee (-x)$. If $|x| \wedge |y| = \theta$, then we write $x \perp y$ and x, y are called disjoint.

A sequence (x_n) in E is said to be increasing if $x_1 \leq x_2 \leq \ldots$ and decreasing if $x_1 \geq x_2 \geq \ldots$. We denote this by $x_n \uparrow$ and $x_n \downarrow$ respectively. An increasing or decreasing sequence is just called monotonic. If $x_n \uparrow$ and $\sup x_n = x$ exists in E then we write $x_n \uparrow x$. If $x_n \downarrow$ and $\inf x_n = x$ exists in E then we write $x_n \downarrow x$.

A sequence (x_n) in E is said to converge in order to x if there exists a sequence $p_n \downarrow \theta$ such that $|x_n - x| \leq p_n$ holds $\forall n$. In this case we write $x_n \overset{ord}{\rightarrow} x$. Note that if $x_n \uparrow x$ or $x_n \downarrow x$ then $x_n \overset{ord}{\rightarrow} x$. A monotonically decreasing sequence which is convergent to θ is called an (O)-sequence.

A sequence (x_n) in E is said to be order bounded if there exists an order interval $[y, z]$ such that $y \leq x_n \leq z \; \forall n \in \mathbb{N}$.

We now recall the following basic facts from [28].

A family $\mathcal{I}$ of subsets of a non-empty set X is said to be an ideal if (i) $A, B \in \mathcal{I}$ imply $A \cup B \in \mathcal{I}$, (ii) $A \in \mathcal{I}$, $B \subset A$ imply $B \in \mathcal{I}$. $\mathcal{I}$ is called nontrivial if $\mathcal{I} \neq \{\varnothing\}$ and $X \notin \mathcal{I}$. $\mathcal{I}$ is admissible if it contains all singletons. If $\mathcal{I}$ is a proper nontrivial ideal then the family of sets $\mathcal{F}(\mathcal{I}) = \{M \subset X : M^{c} \in \mathcal{I}\}$ is a filter on X (where c stands for the complement). It is called the filter associated with the ideal $\mathcal{I}$.

Throughout the paper $\mathcal{I}$ will stand for a proper nontrivial admissible ideal of $\mathbb{N}$.

Definition 2.2 ([23], see also [29]) A sequence (x_n) of elements of $\mathbb{R}$ is said to be $\mathcal{I}$-convergent to $x \in \mathbb{R}$ if for each $\epsilon > 0$ the set $A(\epsilon) = \{n \in \mathbb{N} : |x_n - x| \geq \epsilon\} \in \mathcal{I}$. In this case we write $\mathcal{I} - \lim x_n = x$.

3 Main Definitions and Results

We first introduce the following definition.

Definition 3.1 A sequence (x_n) is said to be $\mathcal{I} - statistically$ order convergent to x_0 if there exists an (O)-sequence (σ_l) such that for any $\delta > 0$

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

for every $l \in \mathbb{N}$. In this case we will write $x_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x_0$.

Theorem 3.1 *In a Riesz space R we have the following,*

(i) The $\mathcal{I}$-statistical order limit is unique.
(ii) The $\mathcal{I}$-statistical order limit is linear.
(iii) The Lattice operations $\vee$ and $\wedge$ are $\mathcal{I}$-statistically order continuous.
(iv) The mappings defined on R by $x \to x^+$, $x \to x^-$ and $x \to |x|$ are $\mathcal{I}$-statistically order continuous.
(v) If (x_n) is a sequence in R such that $x_n \to z$, $\mathcal{I}$-statistically orderly convergent and $x_n \geq y$ for a.a.n then $x \geq y$.

Proof (i) Let (x_n) be a sequence in R such that $x_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x$ and $y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} y$. Then there are two (O) $-$sequences (σ_l) and (η_l) such that for a given δ, $\frac{3}{4} < \delta < 1$

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

and

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |y_k - y| \nleq \eta_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

for any $l \in \mathbb{N}$. Fix $l \in \mathbb{N}$. Now

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x| \nleq \sigma_l\right\}\right| < 1 - \delta\} \in \mathcal{F}(\mathcal{I})$$

and

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |y_k - y| \nleq \eta_l\right\}\right| < 1 - \delta\} \in \mathcal{F}(\mathcal{I}).$$

Since $\mathcal{F}(\mathcal{I})$ is closed under finite intersection and $\phi \notin \mathcal{F}(\mathcal{I})$ so we can choose a $n_0 \in \mathbb{N}$ for which

$$\frac{1}{n_0} \left|\left\{k \leq n_0 : |x_k - x| \nleq \sigma_l\right\}\right| < 1 - \delta$$

as well as

$$\frac{1}{n_0} \left|\left\{k \leq n_0 : |y_k - y| \nleq \eta_l\right\}\right| < 1 - \delta.$$

Writing $B = \left\{k \leq n_0 : |x_k - x| \nleq \sigma_l\right\}$ and $C = \left\{k \leq n_0 : |y_k - y| \nleq \eta_l\right\}$ we observe that

$$\frac{|B \cup C|}{n_0} < \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Hence $B \cup C \neq [1, n_0]$ and so we can choose a $k \leq n_0$ such that $k \notin (B \cup C)$. Consequently,

$$|x_k - x| \leq \sigma_l \text{ and } |x_k - y| \leq \eta_l.$$

Therefore $|x - y| \leq |x_k - x| + |y_k - y| \leq \sigma_l + \eta_l$. Since this is true for any $l \in \mathbb{N}$ and $(\sigma_l + \eta_l)$ is also an (O)-sequence so we must have $x = y$.

(ii) Let (x_n) and (y_n) be two sequences in a Riesz space R such that $x_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x$ and $y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} y$. Then there are two (O)-sequences (σ_l) and (η_l) such that for any $\delta > 0$,

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

and

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |y_k - y| \nleq \eta_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

for any $l \in \mathbb{N}$. Since

$$|(x_k + y_k) - (x + y)| \leq |x_k - x| + |y_k - y|$$

so

$$\{k \leq n : |(x_k + y_k) - (x + y)| \nleq \sigma_l + \eta_l\}$$

$$\subset \left\{k \leq n : |x_k - x| \nleq \sigma_l\right\} \cup \{k \leq n : |y_k - y| \nleq \eta_l\}$$

which implies that

$$\frac{1}{n} \left|\{k \leq n : |(x_k + y_k) - (x + y)| \nleq \sigma_l + \eta_l\}\right|$$

$$\leq \frac{1}{n} | \left\{k \leq n : |x_k - x| \nleq \sigma_l\right\} | + \frac{1}{n} \left|\{k \leq n : |y_k - y| \nleq \eta_l\}\right|.$$

Thus for $\delta_1 > 0$,

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\{k \leq n : |(x_k + y_k) - (x + y)| \nleq \sigma_l + \eta_l\}\right| \geq \delta_1\}$$

$$\subset \{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x| \nleq \sigma_l\right\}\right| \geq \frac{\delta_1}{2}\} \cup \{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |y_k - y| \nleq \eta_l\right\}\right| \geq \frac{\delta_1}{2}\} \in \mathcal{I}.$$

Since $(\sigma_l + \eta_l)$ is also an (O) $-$sequence this shows that $x_n + y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x + y$.

(iii) Let $x_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x$ and $y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} y$. we will have to show that $x_n \vee y_n$ $(x_n \wedge y_n) \overset{\mathcal{I}-stat-ord}{\longrightarrow} x \vee y$ $(x \wedge y)$. The first assertion follows from the same arguments as presented in the proof of (ii) and the relation that for any $n \in \mathbb{N}$

$$\left\{k \leq n : |(x_k \vee y_k) - (x \vee y)| \nleq \sigma_l + \eta_l\right\} \subset \left\{k \leq n : |x_k - x| \nleq \sigma_l\right\} \cup \left\{k \leq n : |y_k - y| \nleq \eta_l\right\}.$$

Similarly we can prove that $x_n \wedge y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} (x \wedge y)$ are omitted. ■

Theorem 3.2 *Let R be a Riesz space and (x_n) , (y_n) , (z_n) are three sequences in R such that $x_n \leq z_n \leq y_n$, $\forall n \in K$ where $K \subset \mathbb{N}$ and $K \in \mathcal{F}(\mathcal{I})$, $x_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x_0$ and $y_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x_0$ then $z_n \overset{\mathcal{I}-stat-ord}{\longrightarrow} x_0$.*

Proof From our hypothesis there exist two (O)-sequences (σ_l) and (η_l) such that for any $\delta > 0$

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

and

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |y_k - y_0| \nleq \eta_l\right\}\right| \geq \delta\} \in \mathcal{I}.$$

Since for any $k \in K$, $|z_k - x_0| \leq |x_k - x_0| + |y_k - x_0|$, so we get

$$\{k \leq n : |z_k - x_0| \nleq \sigma_l + \eta_l\} \cap K \subset (\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\} \cup \{k \leq n : |y_k - x_0| \nleq \eta_l\}) \cap K.$$

Note that for $n \in \mathbb{N}$, if we consider only those k which belong to K then we have

$$\frac{1}{n} \left|\{k \leq n : |z_k - x_0| \nleq \sigma_l + \eta_l\}\right| \leq \frac{1}{n} \left|\{k \leq n : |x_k - x_0| \nleq \sigma_l\}\right| + \frac{1}{n} \left|\{k \leq n : |y_k - x_0| \nleq \eta_l\}\right|.$$

Therefore

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |z_k - x_0| \nleq \sigma_l + \eta_l\right\}\right| \geq \delta\}$$

$$\subset \{n \in \mathbb{N} : \frac{1}{n} \left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\} \geq \frac{\delta}{2}\} \cup \{n \in \mathbb{N} : \frac{1}{n} \left\{k \leq n : |y_k - x_0| \nleq \eta_l\right\} \geq \frac{\delta}{2}\} \cup K^c.$$

Since all the the three sets on the right hand side belong to the ideal $\mathcal{I}$ so the set on the left hand side also belong to $\mathcal{I}$. Finally in view of the fact that $(\sigma_l + \eta_l)$ is also an (O)-sequence we get the desired result. ■

Finally we end the discussion on $\mathcal{I}$-statistical order convergence with the observation that every $\mathcal{I}$-statistically order convergent sequence must have an order convergent subsequence in the usual sense. As the observation is not at all trivial as in the statistical case, we present its proof below.

Let (x_n) be $\mathcal{I}$-statistically order convergent to x_0. Then there is an (O)-sequence (σ_l) satisfying the property that for any $\delta > 0$,

$$\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

for every $l \in \mathbb{N}$. Without any loss of generality assume that (σ_l) is strictly decreasing. Now

$$C = \{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_1\right\}\right| \geq 1\} \in \mathcal{I}.$$

Clearly $C \neq \mathbb{N}$ (as $\mathcal{I}$ is non-trivial). Choose $n_1 \in \mathbb{N}\backslash C$. Then

$$\frac{1}{n_1} \left|\left\{k \leq n_1 : |x_k - x_0| \nleq \sigma_1\right\}\right| < 1$$

$$\text{i.e.} \frac{1}{n_1} |\{k \leq n_1 : |x_k - x_0| \leq \sigma_1\}| > 0.$$

Thus there exists a $k_1 \leq n_1$ for which $\left|x_{k_1} - x_0\right| \leq \sigma_1$. Again taking $\delta = \frac{1}{2}$ we have

$$D = \{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_2\right\}\right| \geq \frac{1}{2}\} \in \mathcal{I}.$$

Since $\mathcal{I}$ is admissible so $D \cup \{1, 2, \ldots, 3n_1\} \in \mathcal{I}$. Hence we can choose $n_2 \in \mathbb{N}$ such that $n_2 \notin D$, $n_2 > 3n_1$ and

$$\frac{1}{n_2} \left|\left\{k \leq n_2 : |x_k - x_0| \nleq \sigma_2\right\}\right| < \frac{1}{2}$$

$$1.\text{e.}\ \frac{1}{n_2} |\{k \leq n_2 : |x_k - x_0| < \sigma_2\}| > \frac{1}{2}.$$

Note that if $|x_k - x_0| \geq \sigma_2\ \forall k,\ n_1 < k \leq n_2$ then

$$\frac{1}{n_2} |\{k \leq n_2 : |x_k - x_0| < \sigma_2\}| \leq \frac{n_1}{n_2} < \frac{1}{3}.$$

Consequently there must exist a k_2, $n_1 < k_2 < n_2$ such that $\left|x_{k_2} - x_0\right| < \sigma_2$. Proceeding in this way we obtain an increasing sequence of indices $\{k_1 < k_2 < k_3 < \ldots\}$ satisfying $\left|x_{k_j} - x_0\right| < \sigma_j$. Evidently $\left(x_{k_j}\right)$ is order convergent to $x_{0.}$

Definition 3.2 A sequence (x_n) is said to be $\mathcal{I}$-*statistically* order pre-Cauchy if there exists an (O)-sequence (σ_l) such that

$$\{n \in \mathbb{N} : \frac{1}{n^2} \left|\left\{(j, k) : \left|x_k - x_j\right| \nleq \sigma_L,\ j, k \leq n\right\}\right| \geq \delta\} \in \mathcal{I}$$

for every $l \in \mathbb{N}$.

Theorem 3.3 *An $\mathcal{I}$-statistically order convergent sequence is $\mathcal{I}$-statistically order pre-Cauchy.*

Proof Let $x = (x_k)$ be $\mathcal{I}$-statistically order convergent to x_0. Let $\delta > 0$ be given. Now there is an (O)-sequence (σ_l) such that

$$C_l = \{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| \geq \delta\} \in \mathcal{I}$$

for every $l \in \mathbb{N}$. Note that for any $n \in C_l^c$ where c stands for the complement,

$$\frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| < \delta$$

$$\text{i.e.} \frac{1}{n} \left|\{k \leq n : |x_k - x_0| \leq \sigma_l\}\right| > 1 - \delta.$$

Now writing $B_n = \{k \leq n : |x_k - x_0| \leq \sigma_l\}$ we see that for $j, k \in B_n$

$$\left|x_k - x_j\right| \leq |x_k - x_0| + \left|x_j - x_0\right| \leq 2\sigma_l.$$

Therefore

$$B_n \times B_n \subset \left\{(j, k) : \left|x_j - x_i\right| \leq 2\sigma_l,\ j, k \leq n\right\}$$

which implies

$$\left[\frac{|B_n|}{n}\right]^2 \leq \frac{1}{n^2} \left|\left\{(j, k) : \left|x_k - x_j\right| \leq 2\sigma_l,\ j, k \leq n\right\}\right|.$$

Note that $(2\sigma_l)$ is also an (O)-sequence and for all $n \in C_l^c$,

$$\frac{1}{n^2} \left|\{(j, k) : \left|x_k - x_j\right| \leq 2\sigma_l,\ j, k \leq n\}\right| \geq \left[\frac{|B_n|}{n}\right]^2 > (1 - \delta)^2$$

$$\text{i.e.} \frac{1}{n^2} \left|\{(j, k) : \left|x_k - x_j\right| \nleq 2\sigma_l,\ j, k \leq n\}\right| < 1 - (1 - \delta)^2.$$

Let $\delta_1 > 0$ be given. Choosing $\delta > 0$ so that $1 - (1 - \delta)^2 < \delta_1$ we see that $\forall n \in C^c$

$$\frac{1}{n^2} \left|\{(j, k) : \left|x_k - x_j\right| \nleq 2\sigma_l,\ j, k \leq n\}\right| < \delta_1$$

and hence

$$\{n \in \mathbb{N} : \frac{1}{n^2} \left|\{(j, k) : \left|x_k - x_j\right| \nleq 2\sigma_l, j, k \leq n\}\right| \geq \delta_1\} \subset C_l.$$

Since $C_l \in \mathcal{I}$, hence

$$\{n \in \mathbb{N} : \frac{1}{n^2} \left|\{(j,k) : \left|x_k - x_j\right| \nleq 2\sigma_l,\, j,k \leq n\}\right| \geq \delta_1\} \in \mathcal{I}$$

which shows that (x_k) is $\mathcal{I}$-statistically order pre-Cauchy. ■

We now present a necessary and sufficient condition for a sequence (x_k) to be $\mathcal{I}$-statistically order pre-Cauchy. We recall the following definition from [28] where the original references can be seen.

Definition 3.3 Let $\mathcal{I}$ be an admissible ideal of $\mathbb{N}$ and Let $x = (x_n)$ be a real sequence. Let

$$A_x = \{a \in R : \{k : x_k < a\} \notin \mathcal{I}\}$$

Then the $\mathcal{I}$-Limit inferior of x is given by

$$\mathcal{I} - \liminf x = \begin{cases} \inf A_x \text{ if } A_x \neq \phi \\ \infty, \text{ if } A_x = \phi \end{cases}$$

It is known (see [28]) that $\mathcal{I} - \liminf x = \alpha$ (finite) if and only if for arbitrary $\varepsilon > 0$,

$$\{k : x_k < \alpha + \varepsilon\} \notin \mathcal{I} \text{ and } \{k : x_k < \alpha - \varepsilon\} \in \mathcal{I}.$$

Theorem 3.4 *Suppose $x = (x_k)$ is $\mathcal{I}$-statistically order pre-Cauchy. If x has a subsequence $\left(x_{p_k}\right)$ which converges to x_0 and*

$$0 < \mathcal{I} - \liminf_n \frac{1}{n} \left|\{p_k \leq n : k \in \mathbb{N}\}\right| < \infty$$

then x is $\mathcal{I}$-statistically order convergent to x_0.

Proof Since $\lim_k x_{p_k} = L$, there is an (O)-sequences (σ_l) such that for every $l \in \mathbb{N}$ there is a $M \in \mathbb{N}$ such that

$$\left|x_j - x_0\right| < \sigma_l$$

whenever $j > M$ and $j = p_k$ for some k. Let $A = \{p_k : p_k > M, k \in \mathbb{N}\}$ and $A(\varepsilon) = \left\{k : |x_k - x_0| \nleq \sigma_l\right\}$. Now observe that

$$\frac{1}{n^2} \left|\{(j,k) : \left|x_k - x_j\right| \nleq \frac{\sigma_l}{2},\; j,k \leq n\}\right| \geq$$

$$\frac{1}{n^2} \sum_{j,k \leq n} \chi_{A(\varepsilon)\times A}(j,k) = \frac{1}{n} \left|\{p_k \leq n : p_k \in A\}\right| . \frac{1}{n} \left|\{k \leq n : |x_k - x_0| \nleq \sigma_l\}\right| .$$

Since x is $\mathcal{I}$-statistically order pre-Cauchy, for $\delta > 0$,

$$C_l = \{n \in \mathbb{N} : \frac{1}{n^2} \left|\{(j,k) : \left|x_k - x_j\right| \nleq \frac{\sigma_l}{2}, j,k \leq n\}\right| \geq \delta\} \in \mathcal{I}.$$

Then for all $n \in C_l^c$,

$$\frac{1}{n^2} \left|\{(j,k) : \left|x_k - x_j\right| \nleq \frac{\sigma_l}{2},\ j,k \leq n\}\right| < \delta..........(1)$$

Again as $\mathcal{I} - \liminf\limits_{n} \frac{1}{n} |\{p_k \leq n : k \in \mathbb{N}\}| = b$ (say) > 0, hence

$$\left\{n \in \mathbb{N} : \frac{1}{n} |\{p_k \leq n : k \in \mathbb{N}\}| < \frac{b}{2}\right\} = D(say) \in \mathcal{I}.$$

Consequently for $n \in D^c$,

$$\frac{1}{n} |\{p_k \leq n : k \in \mathbb{N}\}| \geq \frac{b}{2}..........(2)$$

Now from (1) and (2) we have for all $n \in C_l^c \cap D^c = (C_l \cup D)^c$,

$$\frac{1}{n} \left|\{k \leq n : |x_k - x_0| \nleq \sigma_l\right| < \frac{2\delta}{b}.$$

Let $\delta_1 > 0$ be given. Then choosing $\delta > 0$ such that $\frac{2\delta}{b} < \delta_1$ we see that $\forall n \in (C_l \cup D)^c$,

$$\frac{1}{n} \left|\{k \leq n : |x_k - x_0| \nleq \sigma_l\right| < \delta_1$$

$$\text{i.e.} \left\{n \in \mathbb{N} : \frac{1}{n} \left|\left\{k \leq n : |x_k - x_0| \nleq \sigma_l\right\}\right| \geq \delta_1\right\} \subset (C_l \cup D).$$

Since $C_l, D \in \mathcal{I}$, so $(C_l \cup D) \in \mathcal{I}$ and therefore the set on the left hand side also belongs to $\mathcal{I}$. This shows that x is $\mathcal{I}$-statistically order convergent to x_0. ■

Next we observe an interesting property of $\mathcal{I} - statistically$ order pre-Cauchy sequences in a Riesz space under certain conditions, in Line of Lemma 2.4 [28].

Theorem 3.5 *Let R be a Riesz space which is totally ordered with respect to the partial order "$\leq$" Let $x = (x_k)$ be a sequence and let $y_1, y_2 \in R$ be such that $y_1 < y_2$ and $x_k \notin (y_1, y_2)$ $\forall k \in \mathbb{N}$. Write $A = \{k : x_k \leq y_1\}$ and $B = \{k : x_k \leq y_2\}$ and further assume that*

$$\limsup D_n(A) - \liminf D_n(A) < v$$

for same $0 \leq v \leq 1$. Then if x is $\mathcal{I}$-statistically order pre-Cauchy then either

$$\mathcal{I}-\lim_n D_n(A)=0 \; or \; \mathcal{I}-\lim_n D_n(B)=0$$

where as usual $D_n(A)=\frac{1}{n}\left|\{k\leq n:k\in A\}\right|$.

Proof Clearly $B=\mathbb{N}\backslash A$ and hence $D_n(B)=1-D_n(A)\;\;\forall n\in\mathbb{N}$. We will have to show that $\mathcal{I}-\lim_n D_n(A)=0$ or 1. Note that

$$A\times B\subset\{(j,k):\left|x_k-x_j\right|\nleq y_2-y_1\}.$$

Since x is $\mathcal{I}$-statistically order pre-Cauchy so there is an (O) −sequence (σ_l) such that

$$\mathcal{I}-\lim_n\frac{1}{n^2}\left|\{(j,k):\left|x_k-x_j\right|\nleq\sigma_l,\;\;j,k\leq n\}\right|=0$$

for every $l\in\mathbb{N}$. Since $\sigma_l\downarrow\theta$, $y_2-y_1>\theta$, so choose $l\in\mathbb{N}$ such that $\sigma_l<y_2-y_1$. Then

$$\frac{1}{n^2}\left|\{(j,k):\left|x_k-x_j\right|\nleq y_2-y_1,\;\;j,k\leq n\}\right|\leq\frac{1}{n^2}\left|\{(j,k):\left|x_k-x_j\right|\nleq\sigma_l,\;\;j,k\leq n\}\right|$$

$\forall n\in\mathbb{N}$ and so we get

$$\begin{aligned}0&=\mathcal{I}-\lim_n\frac{1}{n^2}\left|\{(j,k):\left|x_k-x_j\right|\nleq y_2-y_1,\;\;j,k\leq n\}\right|\\&=\mathcal{I}-\lim_n D_n(A)\;\;[D_n(N\backslash A)]\\&=\mathcal{I}-\underset{n}{Lim}D_n(A)\;\;[1-D_n(A)\;].\end{aligned}$$

Now proceeding as in Theorem 2.4 [28] we can obtain the desired result. ■

Acknowledgements The first author is thankful to TUBA, Tukish Academy of Sciences for arranging a visit during which this work was done. The first author is also thankful to SERB, DST, New Delhi for granting a research project No. SR/S4/MS:813/13 during the tenure of which this work was done.

References

1. Riesz, F.: Sur la decomposition des operations functionelles lineaires. Alti del Congr. Internaz. del Mat., Bologna. Zanichelli . **3**, 143–148 (1930)
2. Freudenthal, H.: Teilweise geordnete Modulen. Proc. Acad. Amsterdam. **39**, 641–651 (1936)
3. Kantorovich, L.V.: Lineare halbgeordnete Raume. Receueil Math. **2**, 121–168 (1937)
4. Aliprantis, C. D., Burkinshaw, O.: Locally solid Riesz spaces with applications to economics, 2nd edn. AMS (2003)
5. Zaanen, A.C.: Introduction to operator theory in Riesz spaces. Springer Verlag, Berlin (1997)
6. Sencimen, C., Pehlivan, S.: Statistical order convergence in Riesz spaces. Math. Slovaca. **62**(2), 257–270 (2012)

7. Fast, H.: Sur la convergence statistique. Colloq. Math. **2**, 241–244 (1951)
8. Schoenberg, I.J.: The integrability of certain functions and related summability methods. Amer. Math. Monthly. **66**, 361–375 (1959)
9. Fridy, J.A.: On statistical convergence. Analysis. **5**, 301–313 (1985)
10. Šalát, T.: On Statistically convergent sequences of real numbers. Math. Slovaca. **30**, 139–150 (1980)
11. Maddox, I.J.: Statistical convergence in locally convex spaces. Math Proc. Camb. Phil. Soc. **104**, 141–145 (1988)
12. Connor, J.: R-type summability methods, Cauchy criteria, P-sets and statistical convergence. Proc. Amer. Math. Soc. **115**(2), 319–323 (1992)
13. Di Maio, G., Kočinac, L.D.R.: Statistical convergence in topology. Topology Appl. **156**, 28–45 (2008)
14. Albayrak, H., Pehlivan, S.: Statistical convergence and statistical continuity on locally solid Riesz spaces. Topology Appl. **159**, 1887–1893 (2012)
15. Albayrak, H., Pehlivan, S.: Erratum to "Statistical convergence and statistical continuity on locally solid Riesz spaces". Topology Appl. **160**(3), 443–444 (2013)
16. Das, P., Savas, E.: On I_λ-statistical convergence in locally solid Riesz spaces. Math. Slovaca
17. Das, P., Savas, E.: On $\mathcal{I}$-lacunary statistical order convergence in Riesz spaces. An. Stiint. Univ. Al. I. Cuza Iasi Math
18. Boccuto, A., Dimitriou, X., Papanastassiou, N.: Brooks-Jewett-type theorems for the pointwise ideal convergence of measures with values in (l)-groups. Tatra Mt. Math. Publ. **49**, 17–26 (2011)
19. Boccuto, A., Dimitriou, X., Papanastassiou, N.: Some versions of limit and Dieudonne-type theorems with respect to filter convergence for (l)-group-valued measures. Cent. Eur. J. Math. **9**(6), 1298–1311 (2011)
20. Boccuto, A., Dimitriou, X., Papanastassiou, N.: Basic matrix theorems for I-convergence in (l)-groups. Math. Slovaca. **62**(5), 885–908 (2012)
21. Boccuto, A., Dimitriou, X., Papanastassiou, N.: Schur lemma and limit theorems in lattice groups with respect to filters. Math. Slovaca. **62**(6), 1145–1166 (2012)
22. Connor, J., Fridy, J.A., Kline, J.: Statistically pre-Cauchy sequences. Analysis. **14**, 311–317 (1994)
23. Kostyrko, P., Šalát, T., Wilczyński, W.: I-convergence. Real Anal. Exchange. **26**(2), 669–685 (2000/2001)
24. Das, P., Kostyrko, P., Wilczyński, W., Malik, P.: On I and I* - convergence of double sequences. Math. Slovaca. **58**(5), 605–620 (2008)
25. Das, P., Ghosal, S.: Some further results on I-Cauchy sequences and condition (AP). Comp. Math. Appl. **59**, 2597–2600 (2010)
26. Das, P., Ghosal, S.K.: On I-Cauchy nets and completeness. Topology Appl. **157**(7), 1152–1156 (2010)
27. Das, P., Savas, E., Ghosal, S.K.: On generalizations of certain summability methods using ideals. Appl. Math. Lett. **24**, 1509–1514 (2011)
28. Das, P., Savas, E.: On $\mathcal{I}$-statistically pre-Cauchy sequences. Taiwanese J. Math. **18**(1), 115–126 (2014)
29. Lahiri, B.K., Das, P.: I and I^*-convergence in topological spaces. Math. Bohem. **130**, 153–160 (2005)
30. Savas, E., Das, P.: A generalized statistical convergence via ideals. Appl. Math. Lett. **24**, 826–830 (2011)
31. Savas, E., Das, P., Dutta, S.: A note on strong matrix summability via ideals. Appl. Math Lett. **25**(4), 733–738 (2012)
32. Caserta, A., Di Maio, G., Kočinac, L.D.R.: Statistical convergence in function spaces. Abstr. Appl. Anal. **Vol 2011** (2011), Article ID 420419, p. 11
33. Kostyrko, P., Macaj, M., Šalát, T., Sleziak, M.: I-convergence and extremal I-limit points. Math. Slovaca. **55**, 443–464 (2005)
34. Kuratowski, K.: Topology I. Warszawa, PWN (1961)
35. Luxemburg, W.A.J., Zaanen, A.C.: Riesz spaces - I. North Holland, Amsterdam (1971)

On Ball Dentable Property in Banach Spaces

Sudeshna Basu

Abstract In this work, we introduce the notion of Ball dentable property in Banach spaces. We study certain stability results for the w^*-Ball dentable property leading to a discussion on Ball rentability in the context of ideals of Banach spaces. We prove that the w^*-Ball-dentable property can be lifted from an M-ideal to the whole Banach Space. We also prove similar results for strict ideals of a Banach space. We note that the space $C(K, X)^*$ has w^*-Ball dentable property when K is dispersed and X^* has the w^*-Ball dentable property.

Keywords Slices · M-Ideals · Strict ideals

Classifications 46B20 · 46B28

1 Introduction

Let X be a *real* Banach space and X^* its dual. We will denote by B_X, S_X and $B_X(x, r)$ the closed unit ball, the unit sphere and the closed ball of radius $r > 0$ and center x. We refer to the monograph [3] for notions of convexity theory that we will be using here.

Definition 1 (i) We say $A \subseteq B_{X^*}$ is a norming set for X if $\|x\| = \sup\{|x^*(x)| : x^* \in A\}$, for all $x \in X$. A closed subspace $F \subseteq X^*$ is a norming subspace if B_F is a norming set for X.

(ii) Let $f \in X^*$, $\alpha > 0$ and $C \subseteq X$. Then the set $S(C, f, \alpha) = \{x \in C : f(x) > \sup f(C) - \alpha\}$ is called the open slice determined by f and α. We assume without loss of generality that $\|f\| = 1$. One can analogously define w^* slices in X^*.

S. Basu (✉)
Department of Mathematics, George Washington University,
Washington DC 20052, USA
e-mail: sbasu@gwu.edu; sudeshnamelody@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_13

(iii) A point $x \neq 0$ in a convex set $K \subseteq X$ is called a denting point point of K, if for every $\varepsilon > 0$, there exist slices S of K, such that $x \in S$ and $diam(S) < \varepsilon$. One can analogously define w^*-denting point in X^*.

The following result will be useful in our discussion.

Proposition 1.1 [7] *Suppose $x \in B_X$. Then x is denting point if and only if x is a PC (point of continuity, i.e. points for which the identity mapping on the unit ball, from weak topology to norm topology is continuous.) and an extreme point of B_X.*

Remark 2 The above result is also true for w^*-denting points.

Definition 3 A Banach Space is said to have Ball-Dentable property (BDP) if the unit ball has slices of arbitrarily small diameter. Analogously we can define w^*-Ball dentability in a dual space.

Remark 4 Clearly Radon Nikodym Property (RNP) implies BDP. (see [3].)

In this short note, we study certain stability results for w^*-BDP leading to a discussion on BDP in the context of ideals of Banach spaces, see [4, 10]. We use various techniques from the geometric theory of Banach spaces to achieve this. The spaces that we will be considering have been well studied in the literature. A large class of function spaces like the Bloch spaces, Lorentz and Orlicz spaces, spaces of vector-valued functions and spaces of compact operators are examples of the spaces we will be considering, for details, see [5]. We provide some descriptions of w^*-denting points in Banach spaces in different contexts. We need the following definition.

Definition 5 Let X be a Banach space.

(i) A linear projection P on X is called an M-*projection* if
$\|x\| = max\{\|Px\|, \|x - Px\|\}$, for all $x \in X$.
A linear projection P on X is called an L-*projection* if
$\|x\| = \|Px\| + \|x - Px\|$ for all $x \in X$. A linear projection P on X is called an L^p-*projection*$(1 < p < \infty)$ if
$\|x\|^p = \|Px\|^p + \|x - Px\|^p$ for all $x \in X$.

(ii) A subspace $M \subseteq X$ is called an M-summand if it is the range of an M-projection. A subspace $M \subseteq X$ is called an L-summand if it is the range of an L-projection. A subspace $M \subseteq X$ is called an L^p-summand if it is the range of an L^p-projection.

(iii) A closed subspace $M \subseteq X$ is called an M-ideal if $M^\perp$ is the kernel of an L-projection in X^*.

We recall from Chap. I of [5] that when $M \subset X$ is an M-ideal, elements of M^* have unique norm-preserving extension to X^* and one has the identification, $X^* = M^* \oplus_1 M^\perp$. Several examples from among function spaces and spaces of operators that satisfy these geometric properties can be found in the monograph [5], see also [8]. First, we prove for an M- ideal $M \subset X$, any w^*-denting point of B_{M^*}, is also a

w^*-denting point of B_{X^*}. We prove similar results for a strict ideal $Y \subset X$(see Sect. 2 for the definition) i.e., we prove that a w^*-denting point of a strict ideal continues to be so in the bigger space. We also prove corresponding results for the Ball dentable property. The techniques used in the proofs are adapted from [2].

2 Stability Results

We will use the standard notation of $\oplus_1$,$\oplus_p$, and $\oplus_\infty$ to denote the $\ell^1 \ell^p$, and ℓ^∞-direct sum of two or more Banach spaces. Let $M \subseteq X$ be an M-ideal. It follows from the results in Chap. I in [5] that any $x^* \in X^*$, if $\|m^*\| = \|x^*|M\| = \|x^*\|$, then x^* is the unique norm preserving extension of m^*. For notational convenience we denote both the functionals by m^*. Clearly any M-ideal is also an ideal.

Proposition 2.1 *Let $Z = X \oplus_1 Y$. Let x_0 be a denting point of B_X. Then x_0 is a denting point of B_Z.*

Proof Let $\{z_n\}$ be a sequence in B_Z such that $z_n \longrightarrow x_0$, weakly. If P denotes the projection mapping to X, it follows that $P(z_n) \longrightarrow x_0$. Since x_0 is a denting poing of B_X, it follows from Proposition 1.1, that x_0 is a PC. So $P(z_n) \longrightarrow x_0$ in norm. Since x_0 is denting, it is an extreme point as well, so $\|x_0\| = 1$, and it follows that $\underline{lim} P(x_n) = 1$ and $\underline{lim}(x_n) = 1$. This implies $\lim_n \|z_n - P(z_n)\| = 0$. Therefore $\|z_n - x_0\| \leq \|z_n - P(z_n)\| + \|z_n - x_0\| \longrightarrow 0$. Thus x_0 is a point of continuity of B_Z Also, since x_0 is an extreme point of B_X, it is a an extreme point of B_Z. Again using Proposition 1.1 we have, x_0 is a denting point of B_Z. □

Remark 6 Similar conclusion follow for w^*-denting points also.

The following corollary is immediate.

Corollary 2.2 *Suppose $M \subseteq X$ is an M-ideal. If $m^* \in B_{M^*}$ is a w^*-denting point of B_{M^*}, then $m^* \in B_{X^*}$ is a w^*-denting point of B_{X^*}.*

In the case of an M-ideal $M \subset X$, the following is true.

Proposition 7 *Let $M \subseteq X$ be an M-ideal, then if M^* has the w^*-BDP then X^* has the w^*-BDP.*

Proof Suppose M^* has the w^*-BDP, then for any for any $\varepsilon > 0$ there exists slices S_M and $S_M = \{m^* \in B_{M^*}/m^*(m_0) > 1 - \alpha\}$ and $(diaS_M) < \varepsilon$. Since M is an M-ideal, for any $x^* \in X^*$ we have the unique decomposition, $x^* = m^* + m^\perp$, where $m^* \in M^*$ and $m^\perp \in M^\perp$. Suppose we have $0 < \mu < \alpha$. Then

$$\begin{aligned} S_X &= \{x^* \in B_{X^*}/x^*(m_0) > 1 - \mu\} \\ &= \{x^* \in B_{X^*}/m^*(m_0) + m^\perp(m_0) > 1 - \mu\} \\ &\subseteq S_M \times \mu B_{M^\perp} \end{aligned}$$

Choose $\beta = min(\mu, \varepsilon)$. Then

$$S'_X = \{x^* \in B_{X^*}/x^*(m_0) > 1 - \beta\} \subseteq S_X \times \beta B_{M^\perp}$$

Thus $dia(S'_X) \leq dia(S_M) + 2\varepsilon < \varepsilon + 2\varepsilon = 3\varepsilon$.
Also, since $\|m_0\| = 1$, there exists $m_0^* \in B_{M^*}$ such that $m_0^*(m_0) > 1 - \beta$. Hence $m_0^* \in S'_X$. □

Arguing similarly it follows that

Corollary 8 *Suppose $X = \oplus_p X_i$, $1 \leq p \leq \infty$. If X_i^* has the w^*-BDP for some i, then X^* has the w^*-BDP.*

The above arguments extend easily to vector-valued continuous functions. We recall that for a compact Hausdorff space K, $C(K, X)$ denotes the space of continuous X-valued functions on K, equipped with the supremum norm. We recall from [6] that dispersed compact Hausdorff spaces have isolated points.

Corollary 9 *Suppose K is a compact Hausdorff space with an isolated point. If X^* has the w^*-BDP, then $C(K, X)^*$ has the w^*-BDP.*

Proof Suppose X^* has the w^*-BDP. For an isolated point $k_0 \in K$, the map $F \to \chi_{k_0} F$ is a M-projection in $C(K, X)$ whose range is isometric to X. Hence we see that $C(K, X)^*$ has the w^*-BDP. □

Definition 10 We recall that an ideal Y is said to be a strict ideal if for a projection $P : X^* \to X^*$ with $\|P\| = 1, ker(P) = Y^\perp$, one also has $B_{P(X^*)}$ is w^*-dense in B_{X^*} or in other words $B_{P(X^*)}$ is a norming set for X.

In the case of an ideal also one has that Y^* embeds (though there may not be uniqueness of norm-preserving extensions) as $P(X^*)$. Thus we continue to write $X^* = Y^* \oplus Y^\perp$. In what follows we use a result from [9], that identifies strict ideals as those for which $Y \subset X \subset Y^{**}$ under the canonical embedding of Y in Y^{**}.

Proposition 11 *Suppose Y is a strict ideal of X. If $y^* \in B_{Y^*}$ is a w^*-denting point of B_{Y^*}, then y^* is a w^*-denting point of B_{X^*}.*

Proof Since $y^* \in B_{Y^*}$ is a w^*-denting point of B_{Y^*}, for any $\varepsilon > 0$ there exists w^* slices S and $S = \{y^* \in B_{Y^*}/y^*(y_0) > 1 - \alpha\}$ and $dia(S_) < \varepsilon$. Since Y is a strict ideal in X , we have $B_{X^*} = \overline{B_{Y^*}}^{w^*}$, hence we have the following

$$\begin{aligned} S' &= \{x^* \in B_{X^*}/x^*(x_0) > 1 - \alpha\} \\ &= \{x^* \in \overline{B_{Y^*}}^{w^*}/x^*(x_0) > 1 - \alpha\} \\ \Longrightarrow dia(S') &= dia(S) < \varepsilon \end{aligned}$$

Hence y^* is a w^*-denting point of B_{X^*}. □

Arguing similarly it follows that:

Proposition 12 *Suppose Y is a strict ideal of X. If Y^* has w^*-BDP then X^* has w^*-BDP.*

Remark 13 A prime example of a strict ideal is a Banach space X under its canonical embedding in X^{**}. It is known that any w^*-denting point of $B_{X^{**}}$ is a point of X.

References

1. Acosta, A.D., Kaminska, A., Mastylo, M.: The Daugavet property in rearrangement invariant spaces. Trans. Am. Math. Soc. **367**, 4061–4078 (2015)
2. Basu, S., Rao, T.S.S.R.K.: On small combination of slices in banach spaces. Extr. Math. **31**, 1–10 (2016)
3. Bourgin, R.D.: Geometric aspects of convex sets with the Radon-Nikodm property. Lecture Notes in Mathematics, vol. 993, p. xii+ 474. Springer, Berlin (1983)
4. Godefroy, G., Kalton, N.J., Saphar, P.D.: Unconditional ideals in Banach spaces. Studia Math. **104**, 13–59 (1993)
5. Harmand, P., Werner, D., Werner, W.: M-Ideals in Banach Spaces and Banach Algebras. Lecture Notes in Mathematics, vol. 1547. Springer, Berlin (1993)
6. Lacey, H.E.: The Isometric Theory of Classical Banach Spaces. Die Grundlehren der mathematischen Wissenschaften, Band, vol. 208, pp. x+270 . Springer, New York (1974)
7. Lin, B.L., Lin, P.K., Troyanski, S.: Charecterisation of denting points. Proc. Am. Math. Soc. **102**, 526–528 (1988)
8. Oja, E.: On M-ideals of compact operators in Lorentz sequence spaces. J. Math. Anal. Appl. **259**, 439–452 (2001)
9. Rao, T.S.S.R.K. : On ideals in Banach spaces. Rocky Mt. J. Math. **31**, 595-609 (2001)
10. Rosenthal, H.P.: On the structure of non-dentable closed bounded convex sets. Adv. Math. **70**, 1–58 (1988)

Some Observations Concerning Polynomial Convexity

Sushil Gorai

Abstract In this paper we discuss a couple of observations related to polynomial convexity. More precisely,

(i) We observe that the union of finitely many disjoint closed balls with centres in $\bigcup_{\theta\in[0,\pi/2]} e^{i\theta}V$ is polynomially convex, where V is a Lagrangian subspace of $\mathbb{C}^n$.

(ii) We show that any compact subset K of $\{(z,w)\in\mathbb{C}^2 : q(w)=\overline{p(z)}\}$, where p and q are two non-constant holomorphic polynomials in one variable, is polynomially convex and $\mathcal{P}(K)=\mathcal{C}(K)$.

Keywords Polynomial convexity · Closed ball · Totally real · Lagrangians

2010 Mathematics Subject Classification Primary: 32E20

1 Introduction

For a compact set $K\subset\mathbb{C}^n$ the *polynomially convex hull* is defined by

$$\widehat{K} := \left\{ z\in\mathbb{C}^n : |p(z)| \leq \sup_K |p|,\ p\in\mathbb{C}[z_1,\dots,z_n] \right\}.$$

K is said to be *polynomially convex* if $\widehat{K}=K$. Similarly, we define *rationally convex hull* of a compact set $K\subset\mathbb{C}^n$ as

This work is partially supported by an INSPIRE Faculty Fellowship (IFA-11MA-02) funded by DST and is also supported by a research grant under MATRICS scheme (MTR/2017/000974).

S. Gorai (✉)
Department of Mathematics and Statistics,
Indian Institute of Science Education and Research Kolkata, Mohanpur,
Nadia 741246, West Bengal, India
e-mail: sushil.gorai@iiserkol.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_14

$$\widehat{K}_R := \left\{ z \in \mathbb{C}^n : |f(z)| \leq \sup_K |f|,\ f \text{ is a rational function} \right\}.$$

K is said to be *rationally convex* if $\widehat{K}_R = K$. We note that $K \subset \widehat{K}_R \subset \widehat{K}$. Any compact convex subset of $\mathbb{C}^n$, $n \geq 1$, is polynomially convex. Thanks to Runge's approximation theorem, any compact subset of $\mathbb{C}$ is rationally convex. A compact subset $K \subset \mathbb{C}$ is polynomially convex if and only if $\mathbb{C}\backslash K$ is connected. Hence, in $\mathbb{C}$, polynomial convexity becomes a purely topological property on the compact set; of course, the reason is the very deep interconnections between topology and complex analysis in one variable. In $\mathbb{C}^n$, $n \geq 2$, it is not a topological property. In fact, there exist two compact subsets in $\mathbb{C}^2$, which are homeomorphic, but one of them is polynomially convex and the other is not. For instance, consider the unit circle placed in $\mathbb{R}^2 \subset \mathbb{C}^2$ and in $\mathbb{C} \times \{0\} \subset \mathbb{C}^2$. The first circle is polynomially convex while the later is not. Polynomial convexity is very closely related with polynomial approximation. Below we mention a theorem that exhibit such a connection (see Stout's book [13] for more on these).

Theorem 1.1 (Oka-Weil) *Let $K \subset \mathbb{C}^n$ be a compact polynomially convex. Then any function that is holomorphic in a neighborhood of K can be approximated uniformly on K by polynomials in $z_1, \ldots, z_n$.*

Although the questions of polynomial convexity appear naturally in connections with questions in function theory, it is, however, very difficult to determine whether a given compact in $\mathbb{C}^n$, $n \geq 2$, is polynomially convex. For instance, no characterization of a finite union of pairwise disjoint polynomially convex sets is known. Characterization is not known even for convex compact sets. The union of two disjoint compact convex sets is polynomially convex, thanks to Hahn-Banach separation theorem. The union of three disjoint compact convex set is not necessarily polynomially convex (see Kallin [7]). This leads researchers to focus on certain families of compacts having with some geometrical properties in $\mathbb{C}^n$ to study the question of polynomial convexity. In these paper we present two families of compacts which are polynomially convex. The first one is finite union of disjoint closed balls with centres lying in some particular region in $\mathbb{C}^n$. Let us now make brief survey about works done about polynomial and rational convexity for finite union of pairwise disjoint closed balls. In the same paper Kallin [7] showed that the union of three disjoint closed balls is polynomially convex. It is an open problem whether the union of four disjoint closed balls in $\mathbb{C}^n$, $n \geq 2$, is polynomially convex. The most general result in this direction is given by Khudaiberganov [8].

Result 1.2 (Khudaiberganov) The union of any finite number of disjoint balls in $\mathbb{C}^n$ with centres lying in $\mathbb{R}^n \subset \mathbb{C}^n$ is polynomially convex.

The question of rational convexity of the union of finitely many disjoint closed balls in $\mathbb{C}^n$ is studied by Nemirovskiĭ [10]. He proved that any finite union of disjoint closed balls is rationally convex using a result of Duval-Sibony [2].

In this note we report an interesting (at least to the author) observation proceeding along the similar argument as Khudaiberganov [8] (see also the observation we need to recall few basic notions in symplectic geometry. We consider $(\mathbb{C}^n, \omega_0)$ as a symplectic manifold with the standard symplectic form

$$\omega_0 = \sum_{j=1}^{n} dx_j \wedge dy_j.$$

A linear subspace V of $\mathbb{C}^n$ is said to be a Lagrangian subspace of $\mathbb{C}^n$ if $V = \{u \in \mathbb{C}^n : \omega_0(u, v) = 0 \ \ \forall v \in V\}$. For a Lagrangian subspace V, it follows that for every $\theta \in \mathbb{R}$, $e^{i\theta}V := \{e^{i\theta}v \in \mathbb{C}^n : \ v \in V\}$ is also a Lagrangian subspace.

Remark 1.3 We note that if a subspace V of $\mathbb{C}^n$ is Lagrangian, then the image under a unitary transformation is also a Lagrangian subspace. Also there exists a unitary $T : \mathbb{C}^n \to \mathbb{C}^n$ such that

$$T(V) = \mathbb{R}^n \subset \mathbb{C}^n.$$

By Result 1.2 we know that the union of finitely many disjoint closed balls are polynomially convex if the centres lie in a Lagrangian subspace of $\mathbb{C}^n$.

Our first observation is:

Theorem 1.4 *Let V be a Lagrangian subspace of $\mathbb{C}^n$. The union of finitely many disjoint closed balls is polynomially convex if their centres lie in $\bigcup_{\theta\in[0,\pi/2]} e^{i\theta}V$.*

We now fix some notations: $B(a; r)$ denotes the open ball in $\mathbb{C}^n$ centred at $a = (a_1, \ldots, a_n)$ and with radius r, i.e., $B(a; r) = \{z \in \mathbb{C}^n : |z_1 - a_1|^2 + \cdots + |z_n - a_n|^2 < r^2\}$ and $\mathbb{B}$ denotes the open unit ball. Open unit disc in $\mathbb{C}$ is denoted by $\mathbb{D}$. For a compact $K \subset \mathbb{C}^n$, let $\mathcal{C}(K)$ denotes the algebra of all continuous function and $\mathcal{P}(K)$ denotes the closed subalgebra of $\mathcal{C}(K)$ generated by polynomials in $z_1, \ldots, z_n$.

The other class of compact subsets that we consider in this note are subsets lying in certain real analytic variety in $\mathbb{C}^2$ of the form $\left\{(z, w) \in \mathbb{C}^2 : q(w) = \overline{p(z)}\right\}$, where p and q are two non-constant holomorphic polynomials in one variable. Our next observation is:

Theorem 1.5 *Any compact subset K of $S := \{(z, w) \in \mathbb{C}^2 : q(w) = \overline{p(z)}\}$, where p and q are two non-constant holomorphic polynomial in one variable, is polynomially convex and $\mathcal{P}(K) = \mathcal{C}(K)$.*

If one of p and q is constant a compact patch $K = \left\{(z, w) \in \mathbb{C}^2 : q(w) = \overline{p(z)}\right\} \cap \overline{B(a; r)}$ is polynomially convex but $\mathcal{P}(K) \neq \mathcal{C}(K)$.

2 Technical Preliminaries

In this section we mention some results from the literature that will be useful in the proof. The first one is a lemma due to Kallin [6] (see [1] for a survey on the use of Kallin's lemma)

Lemma 2.1 (Kallin) *Let K_1 and K_2 be two compact polynomially convex subsets in $\mathbb{C}^n$. Suppose further that there exists a holomorphic polynomial P satisfying the following conditions:*

(*i*) $\widehat{P(K_1)} \cap \widehat{P(K_2)} \subset \{0\}$*; and*
(*ii*) $P^{-1}\{0\} \cap (K_1 \cup K_2)$ *is polynomially convex.*

Then $K_1 \cup K_2$ is polynomially convex.

Next, we mention a basic but nontrivial result from Hörmander's book [5].

Result 2.2 ([5], Theorem 4.3.4) Let K be a compact subset of a pseudoconvex domain Ω in $\mathbb{C}^n$. Then $\widehat{K}_\Omega = \widehat{K}_\Omega^P$, where $\widehat{K}_\Omega = \{z \in \Omega : |f(z)| \le \sup_{w\in K} |f(w)|\ \forall f \in \mathcal{O}(\Omega)\}$ and $\widehat{K}_\Omega^P = \{z \in \Omega : u(z) \le \sup_{w\in K} u(w)\ \forall u \in \mathsf{psh}(\Omega)\}$.

We note that, when $\Omega = \mathbb{C}^n$, Result 2.2 gives us that the polynomially convex hull $\widehat{K}$ is equal to the plurisubharmonically convex hull $\widehat{K}^P$. It plays a vital role in our proof of Theorem 1.5. The main idea behind our proof of approximation part of Theorem 1.5 is to look at the points where the set S is totally real. A real submanifold M of $\mathbb{C}^n$ is said to be *totally real* at $p \in M$ if $T_pM \cap iT_pM = \{0\}$, where T_pM denotes the tangent space of M at p viewed as a subspace in $\mathbb{C}^n$. A real submanifold M is said to be *totally real* if it is totally real at every point $p \in M$. Following result from [3] gives a characterization of a level set of certain map from $\mathbb{C}^n$ to $\mathbb{R}^n$ to be totally real.

Result 2.3 ([3], Lemma 2.5) Let $\rho_1, \dots, \rho_n$ be real valued functions so that $\rho := (\rho_1, \dots, \rho_n) : \mathbb{C}^n \to \mathbb{R}^n$ is a submersion. The level set $S := \{z \in \mathbb{C}^n : \rho(z) = 0\}$ is totally real at a point $p \in S$ if and only if $\det A_p \neq 0$, where

$$A_p = \begin{pmatrix} \frac{\partial \rho_1}{\partial \overline{z_1}}(p) & \dots & \frac{\partial \rho_1}{\partial \overline{z_n}}(p) \\ \frac{\partial \rho_2}{\partial \overline{z_1}}(p) & \dots & \frac{\partial \rho_2}{\partial \overline{z_n}}(p) \\ & \vdots & \\ \frac{\partial \rho_n}{\partial \overline{z_1}}(p) & \dots & \frac{\partial \rho_n}{\partial \overline{z_n}}(p) \end{pmatrix}$$

It is well-known that any totally-real submanifold in $\mathbb{C}^n$ is locally polynomially convex at every point (see [4, 14]) i.e., for each $p \in M$ there exists a ball $B(p; r)$ such that $M \cap \overline{B(p; r)}$ is polynomially convex. We now mention the following approximation result due to O'Farrell, Preskenis and Walsh [11] for compact sets that are locally contained in totally-real submanifolds of $\mathbb{C}^n$.

Result 2.4 (O'Farrell-Preskenis-Walsh) Let $K \subset \mathbb{C}^n$ be a compact polynomially convex subset of $\mathbb{C}^n$ and $E \subset K$ be such that $K \setminus E$ is locally contained in totally-real submanifolds of $\mathbb{C}^n$. Then

$$\mathcal{P}(K) = \{f \in \mathcal{C}(K) : f|_E \in \mathcal{P}(E)\}\,.$$

Next, we mention another approximation result that will be useful in our proof of Theorem 1.5.

Result 2.5 ([3], Lemma 2.3) Let K be a compact subset of $\mathbb{C}^n$ such that $\mathcal{P}(K) = \mathcal{C}(K)$. Then any closed subset L of K is polynomially convex and $\mathcal{P}(L) = \mathcal{C}(L)$.

3 Union of Balls

Our aim in this section is to prove Theorem 1.4. Before going into the proof we state and prove a lemma about the image of a ball centred at $\mathbb{R}^n \subset \mathbb{C}^n$ under the polynomial $p(z_1, \ldots, z_n) = \sum_{j=1}^n z_j^2$. This will play a very crucial role in our proof of Theorem 1.4.

Lemma 3.1 *Let $a \in \mathbb{R}^n$ and $0 \leq r \leq 1$ be such that $|a| - r > 1$. Then the image of the closed ball $\overline{B(a,r)}$ under the polynomial $p(z_1, \ldots, z_n) = \sum_{j=1}^n z_j^2$ lies in the affine half-space $\{w \in \mathbb{C} : \mathfrak{Re}\, w > 1\}$.*

Proof Let $z \in \overline{B(a,r)}$, where $a \in \mathbb{R}^n$. Writing $z = x + iy$, $x, y \in \mathbb{R}^n$, we get that

$$|x - a|^2 + |y|^2 \leq r^2. \tag{3.1}$$

For all $z \in \overline{B(a,r)}$ we obtain that

$$\begin{aligned}\mathfrak{Re}\, p(z) &= |x|^2 - |y|^2\\ &\geq |x|^2 - r^2 + |x-a|^2 \quad \text{(using Eq. (3.1))}\\ &= |x|^2 - r^2 + |x|^2 - 2\langle x, a\rangle + |a|^2\\ &\geq 2|x|^2 - 2|x||a| + |a|^2 - r^2.\end{aligned}$$

We now consider the function $\varphi(t) = 2t^2 - 2t|a| + |a|^2 - r^2$. The function $\varphi(t)$ has a minimum at $t = \dfrac{|a|}{2}$ and is increasing for $t > \dfrac{|a|}{2}$. Since, by assumption, $|a| - r > 1$ and $0 \leq r \leq 1$, we get that $r < |a|/2$. This implies that $\dfrac{|a|}{2} < |a| - r$. Therefore, for all $t \geq |a| - r$,

$$\begin{aligned}\varphi(t) &\geq \varphi(|a|-r)\\ &= 2(|a|-r)^2 - 2(|a|-r)|a| + |a|^2 - r^2\\ &= |a|^2 - 2r|a| + r^2\\ &= (|a|-r)^2. \end{aligned} \tag{3.2}$$

For $z \in \overline{B(a,r)}$, $z = x + iy$, we have $|x| \geq |a| - r$. Hence, in view of Eq. (3.2), we obtain that

$$\begin{aligned}\mathfrak{Re}\, p(z) &= \varphi(|x|)\\ &\geq \varphi(|a|-r) > 1 \ \ \forall z \in \overline{B(a,r)}.\end{aligned}$$

Hence,

$$p(\overline{B(a;r)}) \subset \{w \in \mathbb{C} : \mathfrak{Re}\, w > 1\}.$$

□

In this section we provide a proof of Theorem 1.4. The main idea behind the proof is due to Khudaiberganov [8] (see also [12]).

Proof of Theorem 1.4 Since V is a Lagrangian subspace of $\mathbb{C}^n$, there exists a unitary transformation $T : \mathbb{C}^n \to \mathbb{C}^n$ such that $T(V) = \mathbb{R}^n$. $\mathbb{C}$-linearity of T gives us $T(\lambda V) = \lambda\mathbb{R}^n$ for all $\lambda \in \mathbb{C}$; in particular,

$$T(e^{i\theta}V) = e^{i\theta}\mathbb{R}^n.$$

Since unitary transformations of $\mathbb{C}^2$ maps balls to balls, it is enough to consider the disjoint closed balls with centres lying in $\bigcup_{\theta\in[0,\pi/2]} e^{i\theta}\mathbb{R}^n$. Without loss of generality we assume that the closed disjoint balls are as follows: $\overline{\mathbb{B}}$, the closed unit ball, and $\overline{B(a_j;r_j)}$ such that $a_j \in \bigcup_{\theta\in[0,\pi/2]} e^{i\theta}\mathbb{R}^n$ and $0 \leq r_j \leq 1$, $j = 1, \ldots, N$. Since the closed balls are pairwise disjoint, we note that

$$|a_j| - r_j > 1 \ \ \forall j = 1, \ldots, N. \tag{3.3}$$

We show that $\overline{\mathbb{B}} \cup \left(\bigcup_{j=1}^{N} \overline{B(a_j;r_j)}\right)$ is polynomially convex. We will use the induction on N for that. For $N = 1$, clearly, $\overline{\mathbb{B}} \cup \overline{B(a_1;r_1)}$ is polynomially convex for any ball $B(a_1;r_1)$ with $a_1 \in \bigcup_{\theta\in[0,\pi/2]} e^{i\theta}\mathbb{R}^n$ and $\overline{\mathbb{B}} \cap \overline{B(a_1;r_1)} = \varnothing$. As the induction hypothesis we assume that the union $\overline{\mathbb{B}} \cup \left(\bigcup_{j=1}^{N-1} \overline{B(\alpha_j;r_j)}\right)$ of N pairwise disjoint closed balls, one of them being the closed unit ball and the others being any $(N-1)$ pairwise disjoint balls with centres $\alpha_j \in \bigcup_{\theta\in[0,\pi/2]} e^{i\theta}\mathbb{R}^n$ and radii $r_j \leq 1$, is polynomially convex.

Assume the compact sets $K_1 := \overline{\mathbb{B}}$ and $K_2 := \bigcup_{j=1}^{N} \overline{B(a_j;r_j)}$. Since K_2 is a union of $N - 1$ disjoint balls with centres in $\bigcup_{\theta\in[0,\pi/2]} e^{i\theta}\mathbb{R}^n$. Without loss of generality assume that $r_N \geq r_j$, $j = 1, \ldots, (N-1)$. There exists an invertible $\mathbb{C}$-affine transformation S on $\mathbb{C}^n$ of the form

$$S(z) = \mu(z + b),$$

where $\mu, b \in \mathbb{C}$, such that

$$S(\overline{B(a_N; r_N)}) = \overline{\mathbb{B}} \text{ and } S(\overline{B(a_j; r_j)}) = \overline{B(c_j; s_j)},$$

where $c_j \in \bigcup_{\theta \in [0,\pi/2]} e^{i\theta}\mathbb{R}^n$ and $0 \leq s_j \leq 1$ for all $j = 1, \ldots, (N-1)$. We also have $|c_j| - s_j > 1$ for all $j = 1, \ldots, N-1$. By induction hypothesis, $\overline{\mathbb{B}} \cup \left(\bigcup_{j=1}^{N-1} \overline{B(c_j; s_j)}\right)$ is polynomially convex. Hence, K_2 is polynomially convex.

We now use Kallin's lemma (Lemma 2.1) with the polynomial

$$p(z_1, \ldots, z_n) = z_1^2 + \cdots + z_n^2.$$

to show $K_1 \cup K_2$ is polynomially convex. Clearly,

$$|p(z)| \leq 1 \quad \forall z \in K_1. \tag{3.4}$$

Since $a_j \in \bigcup_{\theta \in [0,\pi/2]} e^{i\theta}\mathbb{R}^n$, we assume that $a_j = e^{i\theta_j} b_j$, where $b_j \in \mathbb{R}^n$ and $\theta_j \in [0, \pi/2]$ for all $j = 1, \ldots, N$. We first fix a $j_0 : 1 \leq j_0 \leq N$. Corresponding to j_0 we consider a unitary map $T_{j_0} : \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$T_{j_0}(z) = e^{i\theta_{j_0}} z.$$

Clearly, $T_{j_0}(b_{j_0}) = a_{j_0}$ and $T_{j_0}(\overline{B(b_{j_0}; r_{j_0})}) = \overline{B(a_{j_0}; r_{j_0})}$. In view of Lemma 3.1, we obtain that

$$\mathfrak{Re}\, p(z) > 1 \quad \forall z \in \overline{B(b_{j_0}; r_{j_0})}.$$

Since p is a homogeneous holomorphic polynomial of degree two, we get

$$p(T_{j_0}(z)) = e^{2i\theta_{j_0}} p(z) \quad \forall z \in \overline{B(b_{j_0}; r_{j_0})}.$$

Hence, we get that

$$\mathfrak{Re}\left(e^{-2i\theta_{j_0}} p(z)\right) > 1 \quad \forall z \in \overline{B(a_{j_0}; r_{j_0})}.$$

Therefore, the image of $\overline{B(a_{j_0}, r_{j_0})}$ under the polynomial p lies in the half plane

$$\left\{w \in \mathbb{C} : \mathfrak{Re}\left(e^{-2i\theta_{j_0}} w\right) > 1\right\}.$$

Since we have chosen j_0 arbitrarily, hence, for each $j = 1, \ldots, N$, we obtain that

$$p(\overline{B(a_j, r_j)}) \subset \left\{w \in \mathbb{C} : \mathfrak{Re}\left(e^{-2i\theta_j} w\right) > 1\right\} =: H_{\theta_j}.$$

Writing $w = u + iv$ in $\mathbb{C}$, we get the half space as

$$H_{\theta_j} = \left\{u + iv \in \mathbb{C} : u \cos 2\theta_j + v \sin 2\theta_j > 1\right\}.$$

Since the boundary line of H_{θ_j} is tangent to the unit circle, $H_{\theta_j} \cap \overline{\mathbb{D}} = \varnothing$.

We get the image of K_2 under the polynomial p

$$p(K_2) \subset \bigcup_{j=1}^{N} H_{\theta_j}.$$

We also obtain that

$$\left(\bigcup_{j=1}^{N} H_{\theta_j}\right) \cap \overline{\mathbb{D}} = \varnothing. \tag{3.5}$$

We note that

$$H_0 = \{u + iv \in \mathbb{C} : u > 1\} \ \text{ and } \ H_{\pi/2} = \{u + iv \in \mathbb{C} : u < -1\},$$

and $H_{\theta_j} \subset \{u + iv\mathbb{C} : v > 0, u^2 + v^2 > 1\} \cup H_0 \cup H_{\pi/2}$ for all $j = 1, \ldots, N$. Hence, the strip $\{u + iv \in \mathbb{C} : -1 \le u \le 1, v \le 0\}$ does not intersect the union of half spaces $\left(\bigcup_{j=1}^{N} H_{\theta_j}\right)$. Hence, we get that $\mathbb{C} \setminus \left(\bigcup_{j=1}^{N} H_{\theta_j}\right)$ is connected. Therefore, in view of Eqs. (3.4) and (3.5), we conclude

$$\widehat{p(K_1)} \cap \widehat{p(K_2)} = \varnothing.$$

All the conditions of Kallin's lemma are satisfied with the above polynomial p. Hence, $K_1 \cup K_2 = \bigcup_{j=0}^{N} B_j$ is polynomially convex. □

4 Compact Subsets of Certain Real Analytic Variety

In this section we provide a proof of Theorem 1.5. The idea is to construct a non-negative plurisubharmonic function on $\mathbb{C}^n$ such that the set S lies on the zero set of that function.

Proof of Theorem 1.5 Let B be a closed ball in $\mathbb{C}^2$. If $S \cap B = \varnothing$, then there is nothing to prove. Therefore, assume $S \cap B \neq \varnothing$. We divide the proof into two steps. First we show that $S \cap B$ is polynomially convex. In the second step we show that any compact subset K of S is polynomially convex and $\mathcal{P}(K) = \mathcal{C}(K)$.

Step I: To show $S \cap B$ is polynomially convex.
Consider the function $\Psi : \mathbb{C}^2 \to \mathbb{R}$ defined by

$$\Psi(z, w) = |\overline{p(z)} - q(w)|^2.$$

Clearly, $S = \Psi^{-1}\{0\}$.

A simple computation gives us

$$\begin{aligned}
\frac{\partial^2 \Psi}{\partial z \partial \overline{z}}(z, w) &= \left| \frac{\partial p}{\partial z}(z) \right|^2 \\
\frac{\partial^2 \Psi}{\partial z \partial \overline{w}}(z, w) &= 0 = \frac{\partial^2 \Psi}{\partial w \partial \overline{z}}(z, w) \\
\frac{\partial^2 \Psi}{\partial w \partial \overline{w}}(z, w) &= \left| \frac{\partial q}{\partial w}(w) \right|^2 .
\end{aligned}$$

The Levi-form of Ψ:

$$\begin{aligned}
\mathscr{L}\Psi((z, w); (u, v)) &= \left| \frac{\partial P}{\partial z}(z) \right|^2 |u|^2 + \left| \frac{\partial q}{\partial w}(w) \right|^2 |v|^2 \\
&\geq 0 \quad \forall (u, v) \in \mathbb{C}^2.
\end{aligned}$$

Therefore, Ψ is plurisubharmonic in $\mathbb{C}^2$. Hence, $S \cap B$ is plurisubharmonically convex. In view of Result 2.2, $S \cap B$ is polynomially convex.

Step II: To show any compact subset $K \subset S$ is polynomially convex and $\mathcal{P}(K) = \mathcal{C}(K)$. The main insight here is to show that off a very small set S is totally real. In this case we show that there is a finite set $E \subset S$ such that $S \setminus E$ is locally contained in totally real submanifold of $\mathbb{C}^2$. We will use Result 2.3 for that. In this case the defining function ρ is

$$\rho(z, w) = (\rho_1(z, w), \rho_2(z, w)),$$

where

$$\rho_1(z, w) := \mathfrak{Re}(p(z) - q(w)) \quad \text{and} \quad \rho_2(z, w) := \mathfrak{Im}(-p(z) - q(w))$$

Let $(z_0, w_0) \in S$. The matrix

$$A_{(z_0, w_0)} = \begin{pmatrix} \dfrac{\partial \rho_1}{\partial \overline{z}}(z_0, w_0) & \dfrac{\partial \rho_1}{\partial \overline{w}}(z_0, w_0) \\[2ex] \dfrac{\partial \rho_2}{\partial \overline{z}}(z_0, w_0) & \dfrac{\partial \rho_1}{\partial \overline{w}}(z_0, w_0). \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}\overline{\frac{\partial p}{\partial z}(z_0)} & -\frac{1}{2}\overline{\frac{\partial q}{\partial w}}(w_0) \\ -\frac{i}{2}\overline{\frac{\partial p}{\partial z}(z_0)} & -\frac{i}{2}\overline{\frac{\partial q}{\partial w}(w_0)} \end{pmatrix}$$

We obtain that $\det A_{(z_0,w_0)} = 0$ if and only if $\frac{\partial p}{\partial z}(z_0)\frac{\partial q}{\partial w}(w_0) = 0$. Consider the set

$$Z = \left\{(z_0, w_0) \in \mathbb{C}^2 : \ q(w_0) = \overline{p(z_0)},\ \frac{\partial p}{\partial z}(z_0)\frac{\partial q}{\partial w}(w_0) = 0\right\} =: Z_1 \cup Z_2,$$

where

$$Z_1 := \left\{(z_0, w_0) \in \mathbb{C}^2 : \ q(w_0) = \overline{p(z_0)},\ \frac{\partial p}{\partial z}(z_0) = 0\right\}$$
$$Z_2 := \left\{(z_0, w_0) \in \mathbb{C}^2 : \ q(w_0) = \overline{p(z_0)},\ \frac{\partial q}{\partial w}(w_0) = 0\right\}.$$

Since p and q are non-constant holomorphic polynomials, the holomorphic polynomials $\frac{\partial p}{\partial z}$ and $\frac{\partial q}{\partial w}$ are not identically zero. $\det A_{(z_0,w_0)} \neq 0$ gives us that ρ is locally a submersion at (z_0, w_0). Hence, both the sets Z_1 and Z_2 are finite sets. Hence, by Result 2.3, $S\setminus Z$ is locally contained in totally-real submanifold.

Let K be any compact subset of S. There exists a closed ball B in $\mathbb{C}^n$ such that

$$K \subset S \cap B.$$

Since S is totally-real except finitely many points, in view of Result 2.4, we obtain that

$$\mathcal{P}(S \cap B) = \mathcal{C}(S \cap B).$$

Hence, by Result 2.5, we get that K is polynomially convex and $\mathcal{P}(K) = \mathcal{C}(K)$. □

Corollary 4.1 *If* $K = \{(z^m, \overline{z}^n) \in \mathbb{C}^2 : z \in \overline{\mathbb{D}}\}$ *then* $\mathcal{P}(K) = \mathcal{C}(K)$.

Proof Consider the set

$$S := \{(z, w) \in \mathbb{C}^2 : w^m = \overline{z}^n\}.$$

Clearly, K is a compact subset of S. By using Theorem 1.5, we get that K is polynomially convex and $\mathcal{P}(K) = \mathcal{C}(K)$. □

Remark 4.2 A special case, when $\gcd(m, n) = 1$, of Corollary 4.1 gives us Minsker's theorem [9].

Acknowledgements I would like to thank the referee for valuable comments. I would also like to thank Professor Peter de Paepe for pointing out a mistake in the earlier version of Corollary 4.1.

References

1. de Paepe, P.J.: Eva Kallin's lemma on polynomial convexity. Bull. Lond. Math. Soc. **33**(1), 1–10 (2001)
2. Duval, J., Sibony, N.: polynomially convexity, rational convexity, and currents. Duke Math. J. **79**(2), 487–513 (1995)
3. Gorai, S.: On polynomially convexity of compact subsets of totally-real submanifolds in $\mathbb{C}^n$. J. Math. Anal. Appl. **448**, 1305–1317 (2017)
4. Hörmander, L., Wermer, J.: Uniform approximation on compact sets in $\mathbb{C}^n$. Math. Scand. **23**, 5–21 (1968)
5. Hörmander, L.: An Introduction to Complex Analysis in Several Variables. North-Holland Mathematical Library, vol. 7, 3rd edn. North-Holland Publishing Co., Amsterdam (1990)
6. Kallin, E.: Fat polynomially convex sets, function algebras. In: Proceedings of International Symposium on Function Algebras, pp. 149–152. Tulane University, 1965, Scott Foresman, Chicago, IL (1966)
7. Kallin, E.: Polynomial convexity: the three spheres problem. In: Proceedings of Conference Complex Analysis (Minneapolis, 1964), pp. 301-304. Springer, Berlin (1965)
8. Khudaiberganov, G.: Polynomial and rational convexity of the union of compacta in $\mathbb{C}^n$ Izv. Vyssh. Uchebn. Zaved., Mat. (2), 70–74 (1987)
9. Minsker, S.: Some applications of the Stone-Weierstrass theorem to planar rational approximation. Proc. Am. Math. Soc. **58**, 94–96 (1976)
10. Nemirovskiĭ, S.Y.: Finite unions of balls in $\mathbb{C}^n$ are rationally convex, Uspekhi Mat. Nauk **63**(2)(380), 157–158 (2008) (translation in Russian Math. Surv. **63**(2), 381–382 (2008))
11. O'Farrell, A.G., Preskenis, K.J., Walsh, D.: Holomorphic approximation in Lipschitz norms. In: Proceedings of the Conference on Banach Algebras and Several Complex Variables (New Haven, Conn., 1983), pp. 187–194. Contemp. Math., 32, Amer. Math. Soc., Providence, RI (1984)
12. Smirnov, M.M., Chirka, E.M.: Polynomial convexity of some sets in $\mathbb{C}^n$. Mat. Zametki **50**(5), 81–89 (1991); translation in Math. Notes **50**(5–6), 1151–1157 (1991)
13. Stout, E.L.: Polynomial Convexity. Birkhäuser, Boston (2007)
14. Wermer, J.: Approximations on a disk. Math. Ann. **155**, 331–333 (1964)

Solution of Large-Scale Multi-objective Optimization Models for Saltwater Intrusion Control in Coastal Aquifers Utilizing ANFIS Based Linked Meta-Models for Computational Feasibility and Efficiency

Dilip Kumar Roy and Bithin Datta

Abstract Saltwater intrusion in coastal aquifers poses significant challenges in the management of vulnerable coastal groundwater resources around the world. To develop a strategy for regional scale sustainable management of coastal aquifers, solution of large-scale multi-objective decision models is essential. The flow and solute transport equations are also density dependent, where the flow parameters are dependent on salt concentration; hence, the flow and solute transport equations need to be solved as coupled equations. In a linked optimization simulation model, the numerical simulation model as a predictor of the physical processes need to be solved enormous number of times to be able to identify an optimum solution as per the specified objectives and constraints. This problem becomes even more complicated when multiple objectives are included and the Pareto optimal solutions need to be determined. Therefore, to ensure the computational efficiency and feasibility of determining a regional scale strategy for control and sustainable use of a coastal aquifer, meta-models that are trained, tested and validated using randomized solutions of the numerical simulation models can be utilized. These meta-models once trained and tested serves the purpose of an approximate emulator of the complex numerical models rendering the solution of a complex and large scale linked optimization model computationally efficient and feasible. The optimal groundwater extraction patterns can be obtained through linked simulation-optimization (S/O) technique in which the simulation part is usually replaced by computationally efficient meta-models. This study proposes a computationally efficient meta-model to emulate density reliant integrated flow and solute transport scenarios of coastal aquifers. A meta-model, Adaptive Neuro Fuzzy Inference System (ANFIS) is trained and developed for an illustrative coastal aquifer study area. Prediction accuracy of the developed ANFIS based meta-model is evaluated for suitability. The meta-model is then integrated with

D. K. Roy (✉) · B. Datta
Discipline of Civil Engineering, College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia
e-mail: dilip.roy@my.jcu.edu.au

B. Datta
e-mail: bithin.datta@jcu.edu.au

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_15

a multiple objective coastal aquifer management model to demonstrate the potential application of this methodology. The optimization algorithm utilized for solution is the Controlled Elitist Multi-objective Genetic Algorithm. Performance evaluation results show acceptable accuracy in the obtained optimized management strategies. Therefore, use of trained and tested meta-models linked to an optimization model results in significant computational efficacy. It also ensures computational practicability of solving such large-scale integrated S/O approach for regional scale coastal groundwater management.

Keywords Saltwater intrusion · Coastal aquifer · Parallel processing · Fuzzy c-mean clustering · Fuzzy inference system · Genetic algorithm

1 Introduction

Coastal areas are vulnerable to saltwater intrusion due to unplanned exploitation of groundwater resources. One of the most important approaches to minimize saltwater intrusion in these areas is to adopt optimal use of groundwater resources in different sectors to prevent degradation of water quality as well as to warrant sufficient supply of potable water. Coupled simulation-optimization (S/O) approach can be utilized to derive optimum groundwater management schemes that provide different alternate extraction strategies while maintaining the saltwater concentration at specified monitoring locations to maximum allowable limits [3, 8]. This approach requires two important constituents: a numerical simulation model, which simulates density reliant integrated flow and solute transport processes, and an optimization algorithm that is utilized to obtain global Pareto optimal solution in terms of different alternative groundwater extraction strategies. As part of the regional scale management strategy, different management measures can be utilized. These may include creation of hydraulic barriers near the shoreline for controlling saltwater intrusion. This research adopts barrier extraction wells [5, 6, 18] in combination with beneficial production wells to develop multiple-objective groundwater extraction strategies by utilizing the coupled S/O approach. Water extracted from barrier wells are generally very saline and cannot be used for beneficial purposes; therefore, the aim is to reduce groundwater abstraction from these wells while minimizing saltwater intrusion. Another aim of the management problem is to make best use of advantageous withdrawal from the production bores.

However, a major challenge in implementing such a coupled S/O approach is to decrease the computational burden. In an integrated S/O approach, simulation models are called by the optimization algorithm numerous number of times in order to achieve global optimal solution. These repeated calls are associated with huge computational burden. For instance, if the numerical simulation model for a study needs only 10 min to converge, and the numerical simulation model is evaluated 1000 times by the optimization routine, then approximately 7 days of CPU time will be required to obtain an optimal solution for this problem. One way to get rid of the computational infeasibility issue is to use computationally efficient approximate meta-models to replace time consuming and memory intensive numerical models [4]. Previous studies of saltwater intrusion management modelling have utilized different meta-models as computationally efficient substitutes of complex simulation models in the optimization formulation. The most commonly used meta-models are based on Artificial Neural Network (ANN) [1, 2, 9], Multivariate Adaptive Regression Spline (MARS) [13, 16] and Genetic Programming (GP) [17, 19]. However, these meta-models have certain disadvantages. For example, the ANN has associated computational burden and model overfitting issues [21], GP tends to converge to local optima [12] etc. To get rid of few drawbacks of the commonly used meta-models for saltwater intrusion processes prediction, Roy, Datta [14] proposed a fuzzy logic based meta-model, which is more accurate, stable, and computationally efficient compared to existing meta-models in saltwater intrusion problems in coastal aquifers.

This study extends the use of fuzzy logic based meta-modelling approach by externally linking Adaptive Neuro Fuzzy Inference System (ANFIS) models within an integrated S/O methodology to obtain multiple-objective groundwater abstraction policies for controlling salinity intrusion. The applicability of the proposed approach is verified for a demonstrative multiple layered coastal aquifer study area. Each layer of the stratified aquifer represents different aquifer materials. However, the values of hydraulic conductivity within each soil layer are assumed constant.

2 Methodology

The proposed methodology includes a numerical model for simulating aquifer processes; the ANFIS meta-models trained and tested using the solution result of the numerical simulation model to emulate the physical processes of the stratified aquifer system; and an optimization algorithm to prescribe optimal groundwater extraction strategies as solutions. The optimization algorithm utilized in the present study is a Controlled Elitist Multiple Objective Genetic Algorithm (CEMOGA) [7].

2.1 Numerical Model

A three-dimensional (3D) density reliant coupled flow and salt transport numerical model, FEMWATER [10] is used to simulate the integrated flow and transport processes, and to produce input-output training pairs for the ANFIS based meta-models. The governing 3D density dependent integrated flow and transport equations are described in Lin et al. [10] and is not repeated here.

2.2 ANFIS

Physical processes of coastal aquifers are approximated by using ANFIS based meta-models to achieve computational efficiency, and to ensure computational achievability of the proposed integrated S/O methodology. Input-output datasets acquired from the solution results of the simulation model are used to train and test the meta-models.

The Sugeno FIS, also referred to as Takagi-Sugeno-Kang FIS [20] is perfect for developing a preliminary FIS structure for training of the desired ANFIS model. The computational framework of a Sugeno FIS follows the theory of fuzzy logic. FISs can be applied to nonlinear mapping of input and output spaces by utilizing a number of fuzzy rules.

Fuzzy if-then rule set for a first-order Sugeno type FIS can be written as:

$$\text{Rule 1 : If} \propto \text{is } P_1 \text{ and } \beta \text{ is } Q_1 \text{ then } f_1 = p_1\alpha + q_1\beta + r_1 \tag{1}$$

$$\text{Rule 1 : If} \propto \text{is } P_2 \text{ and } \beta \text{ is } Q_2 \text{ then } f_2 = p_2\alpha + q_2\beta + r_2 \tag{2}$$

This can be graphically represented in Fig. 1 that illustrates an ANFIS architecture based on a Sugeno type FIS.

2.3 Management Model

A regional scale coastal aquifer management model is developed through ANFIS meta-model based integrated S/O approach. Two contradictory aims of groundwater abstraction strategy are implemented. The first objective ensures the maximum withdrawal of groundwater for beneficial purposes. The second objective minimizes groundwater abstraction from barrier extraction wells to reduce the extent of saltwater intrusion by establishing a hydraulic head barrier near the coastal boundary. The proposed coastal aquifer management model is represented by the following

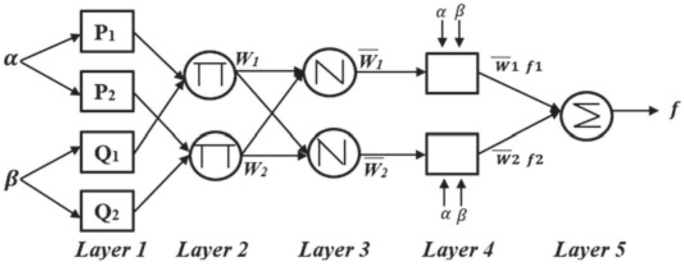


Fig. 1 ANFIS model structure derived from a Sugeno type FIS that has two inputs (after Roy, Datta [15])

equations

$$\text{Maximize}: f_1(Q_{PW}) = \sum_{r=1}^{R} \sum_{t=1}^{T} Q_{PW_r^t} \tag{3}$$

$$\text{Minimize}: f_2(Q_{BW}) = \sum_{k=1}^{K} \sum_{t=1}^{T} Q_{BW_k^t} \tag{4}$$

Subject to

$$C_i = \xi(Q_{PW}, Q_{BW}) \tag{5}$$

$$C_i \leq C_{max} \forall_i \tag{6}$$

$$Q_{PW\,min} \leq Q_{PW_r^t} \leq Q_{PW\,max} \tag{7}$$

$$Q_{BW\,min} \leq Q_{BW_k^t} \leq Q_{BW\,max} \tag{8}$$

where, PW = production wells, BW = barrier extraction wells, $Q_{PW_r^t}$ = groundwater abstraction from the rth production well at tth time step, $Q_{BW_k^t}$ = groundwater abstraction from the kth barrier extraction well at tth time step, C_i = salinity concentrations at ith monitoring locations at the termination of the management phase, ξ () = density reliant integrated flow and solute transport meta-model.

3 Application of the Developed Methodology

The applicability of the meta-model and integrated S/O based coastal aquifer management model is illustrated for an demonstrative coastal aquifer site as shown in Fig. 3 (after Roy, Datta [14]). The 4.35 km^2 coastal aquifer study area has a thickness of 80 m, which is divided into four different soil layers. Each soil layer is 20 m thick. The simulation of the aquifer processes is carried out for a period of 5 years, the groundwater extraction is assumed constant during each 1-year period. The transient simulation is accomplished using a period of 5 days.

A 5-year management phase is considered. The pumping variables are symbolized by X1–X80 in which variables X1–X55 denotes water abstraction from 11 production wells for 5 years of management period, and variables X56–X80 indicates extraction of water from 5 barrier wells during the period of 5 years. The range of pumping values of the eighty decision variables is 0–1300 m^3/day. Salinity concentrations at the completion of the management phase are observed at five designated monitoring locations. The monitoring locations are symbolized by OP1–OP5 (Fig. 2). Individual ANFIS meta-models are developed for each of these monitoring locations.

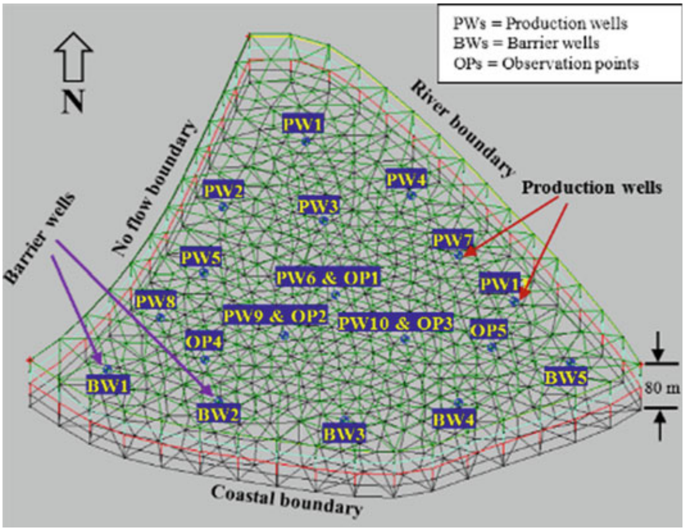


Fig. 2 Illustrative study area

4 Results and Discussion

4.1 Performance of ANFIS Meta-Models

The prediction capability of the developed ANFIS meta-models is evaluated by comparing the actual and ANFIS predicted salinity concentrations using different performance evaluation indices. In this context, actual concentration denotes saltwater concentrations obtained from numerical simulation model solution as a response to spatiotemporal groundwater extraction values from the aquifer, and predicted concentration values indicate the values predicted by the developed ANFIS meta-models at different locations. In general, at all monitoring locations, ANFIS predictions are in good covenant with the numerical model simulation results. Therefore, the prediction accuracy of all developed ANFIS meta-models is satisfactory.

Prediction precision of the developed ANFIS meta-models are assessed based on RMSE, MARE, R, NS, and TS criteria as depicted in Table 1. It is showed from Table 1 that results of all performance evaluation indices are satisfactory for all ANFIS meta-models predicting salinity concentrations at five monitoring locations. The proposed ANFIS meta-models result in higher values of R and NS, and lower values of RMSE and MARE showing good prediction capability.

Table 1 ANFIS performance for testing dataset

Monitoring locations	Evaluation indices						
	RMSE	MARE	R	NS	TS (RE threshold)		
					<1%	<2%	<5%
OP1	0.2940	0.5451	0.9981	0.9960	81.33	99.33	100
OP2	6.3862	0.3602	0.9993	0.9985	96	100	–
OP3	3.7025	0.2507	0.9997	0.9994	99	100	–
OP4	6.5414	0.0736	0.9998	0.9996	100	–	–
OP5	3.5376	0.0443	0.9999	0.9997	100	–	–

OPs = Monitoring locations, RMSE = Root mean square error, MARE = Mean absolute relative error, R = Correlation coefficient, NS = Nash-Sutcliffe efficiency coefficient

4.2 Management Model Performance

The optimization model provides optimal solution represented by a Pareto optimal front that shows the non-dominated tradeoff between the two contradictory objectives of the management problem. The Pareto front shown in Fig. 3 provides several sets of non-dominated groundwater abstraction values obtained while ensuring that the maximum permissible salinity levels at designated monitoring locations are not exceeded.

It is perceived from Fig. 3 that increasing the rate of extraction from production wells requires an additional amount of water withdrawal from the barrier wells to meet specified salinity standards. Abstracted groundwater from the barrier wells generally cannot be utilized for beneficial purposes due to high salinity. Therefore, managers can choose the rate of barrier well pumping based on the demand for beneficial water use, while keeping the pre-set maximum allowable saltwater concentrations at certain monitoring locations in mind. These Pareto optimal solutions show the contradictory

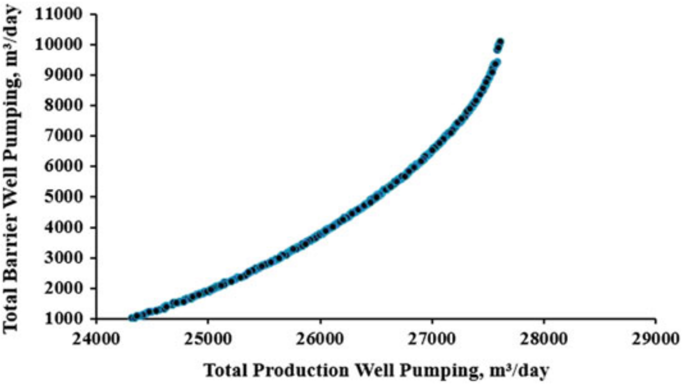


Fig. 3 Pareto front acquired as solution of the management policy

feature of the two objectives, a necessary condition for a multiple objective optimal management.

The optimization routine for this multiple objective coastal aquifer management model is performed in parallel by utilizing multiple MATLAB workers in a parallel pool [11]. The parallel computation is performed by allocating the objective function and the constraints to four workers (4 cores) using a seven core PC [Intel (R) Core (TM) i7-4790 CPU@3.60 GHz]. The optimization model with parallel processing facility takes around 5.674 h to obtain global optimal solution compared to 17.012 h of CPU time without the parallel processing platform.

5 Summary and Conclusions

This study presents and evaluates the utilization of a trained ANFIS meta-model as an effective soft computing tool to emulate salt transport scenarios of a coastal aquifer study area. Moreover, the feasibility of incorporating the proposed ANFIS meta-models with CEMOGA optimization algorithm in an integrated S/O approach to obtain the Pareto optimal solution of groundwater abstraction policies is also illustrated. Individual ANFIS meta-models are developed at five designated monitoring locations. Properly trained and validated ANFIS meta-models are then integrated externally with a CEMOGA based optimization algorithm. The proposed ANFIS-CEMOGA approach is implemented for solving the large-scale multiple objective optimization problem. Additional computational efficiency is achieved by implementing the management model in a parallel computation approach. An example problem consisting of a multiple layered coastal aquifer study area is utilized to assess the applicability of the proposed approach. Results demonstrated the suitability of ANFIS meta-models in developing an integrated S/O based large-scale coastal aquifer management policy. Therefore, ANFIS meta-models have the potential to make development of such management policies in a large-scale coastal aquifer system computationally efficient and feasible. Future study may attempt to extend the application of this methodology to heterogeneous coastal aquifer systems incorporating random fields of different aquifer parameters.

References

1. Bhattacharjya, R.K., Datta, B.: Optimal management of coastal aquifers using linked simulation optimization approach. Water Resour. Manage **19**(3), 295–320 (2005). https://doi.org/10.1007/s11269-005-3180-9
2. Bhattacharjya, R.K., Datta, B., Satish, M.G.: Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers. J. Hydrol. Eng. **12**(3), 273–282 (2007). https://doi.org/10.1061/(asce)1084-0699(2007)12:3(273)
3. Bhattacharjya, R.K., Datta, B.: ANN-GA-based model for multiple objective management of coastal aquifers. J. Water Resour. Plan Manage **135**(5), 314–322 (2009). https://doi.org/10.1061/(asce)0733-9496(2009)135:5(314)

4. Blanning, R.W.: The construction and implementation of metamodels. Simulation **24**, 177–184 (1975)
5. Das, A., Datta, B.: Development of management models for sustainable use of coastal aquifers. J. Irrig. Drain. Eng **125**(3), 112–121 (1999). https://doi.org/10.1061/(asce)0733-9437(1999)125:3(112)
6. Das, A., Datta, B.: Development of multiobjective management models for coastal aquifers. J. Water Resour. Plan Manage **125**(2), 76–87 (1999). https://doi.org/10.1061/(asce)0733-9496(1999)125:2(76)
7. Deb, K., Goel, T.: Controlled elitist non-dominated sorting genetic algorithms for better convergence. In: Zitzler, E., Thiele, L., Deb, K., Coello Coello, C.A., Corne, D. (eds.), Evolutionary Multi-criterion Optimization: First International Conference, EMO 2001 Zurich, Switzerland, March 7–9, 2001 proceedings, pp. 67–81. Springer, Berlin (2001). https://doi.org/10.1007/3-540-44719-9_5
8. Dhar, A., Datta, B.: Saltwater intrusion management of coastal aquifers. I: linked simulation-optimization. J. Hydrol. Eng. **14**(12), 1263–1272 (2009). https://doi.org/10.1061/(asce)he.1943-5584.0000097
9. Kourakos, G., Mantoglou, A.: Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. Adv. Water Resour. **32**(4), 507–521 (2009). https://doi.org/10.1016/j.advwatres.2009.01.001
10. Lin, H.-C.J., Rechards, D.R., Talbot, C.A., Yeh, G.-T., Cheng, J.-R., Cheng, H.-P., Jones, N.L.: FEMWATER: a three-dimensional finite element computer model for simulating density-dependent flow and transport in variable saturated media. Technical Rep No CHL-97–12 Vicksburg, MS: US Army Engineer Waterways Experiment Station Coastal and Hydraulics Laboratory (1997)
11. MATLAB (2017a) MATLAB Version R2017a. The Mathworks Inc, Mathworks, Natick
12. Pillay, N.: An investigation into the use of genetic programming for the induction of novice-procedural programming solution algorithms in intelligent programming tutors. Dissertation, University of KwaZulu-Natal, Durban (2004)
13. Roy, D.K., Datta, B.: Multivariate adaptive regression spline ensembles for management of multilayered coastal aquifers. J. Hydrol. Eng. **22**(9), 04017031 (2017)
14. Roy, D.K., Datta, B.: Fuzzy c-mean clustering based inference system for saltwater intrusion processes prediction in coastal aquifers. Water Resour. Manage **31**(1), 355–376 (2017). https://doi.org/10.1007/s11269-016-1531-3
15. Roy, D.K., Datta, B.: Genetic algorithm tuned fuzzy inference system to evolve optimal groundwater extraction strategies to control saltwater intrusion in multi-layered coastal aquifers under parameter uncertainty. Model Earth Syst. Environ. **3**(4), 1707–1725 (2017). https://doi.org/10.1007/s40808-017-0398-5
16. Roy, D.K., Datta, B.: A surrogate based multi-objective management model to control saltwater intrusion in multi-layered coastal aquifer systems. Civ. Eng. Environ. Syst. 1–26 (2018). https://doi.org/10.1080/10286608.2018.1431777
17. Sreekanth, J., Datta, B.: Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. J. Hydrol. **393**(3–4), 245–256 (2010). https://doi.org/10.1016/j.jhydrol.2010.08.023
18. Sreekanth, J., Datta, B.: Optimal combined operation of production and barrier wells for the control of saltwater intrusion in coastal groundwater well fields. Desalin Water Treat **32**(1–3), 72–78 (2011). https://doi.org/10.5004/dwt.2011.2680
19. Sreekanth, J., Datta, B.: Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. Water Resour. Manage **25**(13), 3201–3218 (2011). https://doi.org/10.1007/s11269-011-9852-8
20. Sugeno, M.: Industrial Applications of Fuzzy Control. Elsevier Science Inc. (1985)
21. Tu, J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J. Clin. Epidemiol. **49**(11), 1225–1231 (1996). https://doi.org/10.1016/S0895-4356(96)00002-9

Cost Driven Optimization of Microgrid Under Environmental Uncertainties Using Different Improved PSO Models

Meenakshi De, G. Das and K. K. Mandal

Abstract This paper presents a micro grid generation scheduling model using Non-linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) and Time Varying Acceleration Co-efficient Particle Swarm Optimization (TVAC-PSO) techniques. Here energy management in micro grid is done in presence of renewable energy sources such as wind and solar power. In this research work, implementation of Demand Response (DR) schedules are carried out as incentive based payment i.e., on offered price packages. In the typical microgrid, different power components including Wind Turbine (WT), Photovoltaic (PV) cell, Micro-Turbine (MT), Fuel Cell (FC), battery hybrid power source and responsive loads are used. Analytical approaches and case studies are conducted for obtaining minimum operating costs and comparative studies are carried out without demand response participation and with demand response participation respectively. The results obtained represent the superiority of the proposed approach for effective generation scheduling in micro grids.

Keywords Non-linear decreasing inertia weight particle swarm optimization · Generation scheduling · Demand response · Micro grids

1 Introduction

The micro grid concept of power production is relatively new aimed at reducing harmful emissions from fossil-fueled power plants at the same time enhancing the utilization of new technologies and concepts in the form of inclusion of renewable energy resources in the existing power networks. But, the disadvantage of renewable energy resources lies in their intermittent nature of power production. Thus, in order to maintain stable system operation, the power system operators provide certain system reserve to overcome uncertainties of power production. Some research stud-

M. De (✉) · G. Das · K. K. Mandal
Department of Power Engineering, Jadavpur University, Kolkata, India
e-mail: meenakshide.ju@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_16

ies conducted previously investigated the inclusion of distributed generations such as wind and solar energy [1–3]. But these solutions suffer from drawbacks such as increase in costs and problems in commitment of power units etc. Another solution to this problem includes increasing the total energy reserve [4] to maintain system security. So a demand side reserve or demand response (DR) can be included by network operators by way of decreasing energy consumption during energy shortage period [5]. So, demand response (DR) is the change in consumption pattern of electricity in customer's side in response to changing electricity prices over a specified duration of time. Thus it increases incentive based payment modes with respect to utilization of electricity. These incentive based payments encourage lower use of electricity in times of high market price and higher use when prices are within threshold value. The incentive based demand response management in micro grids provide diverse advantages such as it covers the uncertainty associated with solar and wind power production. At the same time the customers will have a wide variety of choices from the offered packages to suit their needs and their budget. Some studies focused on micro grid (MG) operation and management involving a stochastic model in order to achieve optimized set points of operation [6, 7] while others stressed on investigating MG operation by heuristic algorithm [8]. In this paper optimal operating cost of a micro grid system is found using two meta-heuristic techniques viz. Non-linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) and Time Varying Acceleration Co-efficient Particle Swarm Optimization (TVAC-PSO). The results obtained in this research are compared with those of results obtained in other published literature. The results prove the effectiveness of the proposed approach.

Thus, the main contribution of this research paper is as follows:

(1) Utilization of incentive based demand response program to cover the uncertainties resulted from power production by solar PV array and wind turbine.
(2) Novelty of this work lies in the development of microgrid generation scheduling model using Non-linear Decreasing Inertia Weight Particle Swarm Optimization algorithm (NDIW-PSO) and Time Varying Acceleration Co-efficient Particle Swarm Optimization algorithm (TVAC-PSO).
(3) Analytical approaches and case studies are conducted for obtaining minimum operating cost in microgrid. Comparative studies are also carried out without demand response participation and with demand response participation respectively to obtain the results which establish the superiority of this research work.

2 Problem Formulation

2.1 Specifications of Demand Response Participants

Various types of electricity consumers with varying consumption behaviors can be considered for case study. In this paper, consumers are of three types i.e., residential, commercial and industrial. The micro grid structure followed in this work has renewable energy resources such as wind and solar photovoltaic array where the incentive based demand response programs are utilized to cover up the environmental uncertainties associated with these types of sources. Energy is given to consumers within 24 h period consisting of resources such as solar PV, wind turbine and battery.

$$RP(r,t) = RC(r,t) * \pi_{r,t}\,;\, RC(r,t) <= RC_t^{max} \tag{1}$$

$$CP(c,t) = CC(c,t) * \pi_{c,t}\,;\, CC(c,t) <= CC_t^{max} \tag{2}$$

$$IP(i,t) = IC(i,t) * \pi_{i,t}\,;\, IC(i,t) <= IC_t^{max} \tag{3}$$

In above equations r, c and i present the number of residential, commercial and industrial consumers; $RC(r,t)$, $CC(c,t)$ and $IC(i,t)$ indicate the amount of load reduction planned by each residential, commercial and industrial consumer in period t; RC_t^{max}, CC_t^{max} and IC_t^{max} represents maximum load reduction proposed by each consumer respectively in time period t; $\pi_{r,t}$, $\pi_{c,t}$ and $\pi_{i,t}$ indicates amount of incentive payments towards each customer in time period t and $RP(r,t)$, $CP(c,t)$ and $IP(i,t)$ presents cost due to load reduction by residential, commercial and industrial customers in period t for proposed load reduction respectively.

2.2 Objective Function

The objective function consists of several components such as storage costs, start-up and shutdown costs of the generating units, costs for exchange with the utility and costs of demand response program participation which can be mathematically formulated as follows:

$$\begin{aligned} minf_1(X) = \sum_{t=1}^{T} Cost = \sum_{t=1}^{T}\sum_{i=1}^{Ng}[u_i(t)P_{Gi}(t)B_{Gi}(t) + S_{Gi}|u_i(t) - u_i(t-1)|] \\ + \sum_{j=1}^{Ns}[u_j(t)P_{sj}(t)B_{sj}(t) + S_{sj}|u_j(t) - u_j(t-1)|] \\ + P_{Grid}(t)B_{Grid}(t) + P_{DR}(t)B_{DR}(t) \end{aligned} \tag{4}$$

$$P_{DR}(t) = \sum_{r} RC(r,t) + \sum_{c} CC(c,t) + \sum_{i} IC(i,t) \tag{5}$$

where $P_{Gi}(t)$ and $P_{Sj}(t)$ are the real power output of ith generator and jth storage in period t respectively, $B_{Gi}(t)$ and $B_{Sj}(t)$ are bids of DG and storage unit at hour t respectively. $S_{Gi}(t)$ and $S_{Sj}(t)$ represent the start up and shut down costs for ith and jth storage respectively, $P_{Grid}(t)$ is the active power which is bought/sold from/to the utility at time t and $B_{Grid}(t)$ is the bid of utility at time t, u represents the state vector which indicates the ON/OFF states of all units in period t. $P_{DR}(t)$ and $B_{DR}(t)$ is the active power and bid price to participate in the Demand Response Programs in period t.

2.3 Problem Constraints

Power balance boundaries: The total power generated from the distributed units should be equal to the total demand of power in the microgrid for maintaining a steady supply of power. It can be mathematically described as follows:

$$\sum_{i=1}^{Ng} P_{Gi}(t) + \sum_{j=1}^{Ns} P_{sj}(t) + P_{Grid}(t) = \sum_{k=1}^{Nk} P_{lk}(t) \tag{6}$$

Real power generation capacity: For efficient and reliable operation of the microgrid system, the active power output of each DG unit is bounded by lower and upper limits:

$$\begin{gathered} P_{Gi,min}(t) <= P_{Gi}(t) <= P_{Gi,max}(t) \\ P_{sj,min}(t) <= P_{sj}(t) <= P_{sj,max}(t) \\ P_{Grid,min}(t) <= P_{Grid}(t) <= P_{Grid,max}(t) \end{gathered} \tag{7}$$

where $P_{G,min}(t)$, $P_{s,min}(t)$ and $P_{Grid,min}(t)$ are the minimum active powers of ith DG, jth storage and grid at period t respectively. In similar way, $P_{G,max}(t)$, $P_{s,max}(t)$ and $P_{Grid,max}(t)$ are the maximum power generations of corresponding units at hour t.

Battery charge and discharge boundaries: During each time period, there are limitations on charge and discharge of battery, the equations and constraints can be expressed as in the following equation:

$$\begin{gathered} \phi_j(t) = \phi_j(t-1) - 1/\lambda_{Dj} u_{Dj}(t) P_{Dsj}(t) + \lambda_{Cj}(t) u_{Cj}(t) P_{Csj}(t) \\ \phi_{min} <= \phi_j(t) <= \phi_{max} \end{gathered} \tag{8}$$

where $\phi_j(t)$ and $\phi_j(t-1)$ represent the reserved energy at present time and at the previous time count. $P_{Dsj}(t)$ and $P_{Csj}(t)$ is the allowed rate of discharge and charge during a definite time interval. $\lambda_{Dj}(t)$ and $\lambda_{Cj}(t)$ present the battery efficiency at times of discharge and charge respectively.

3 Particle Swarm Optimization

Particle swarm optimization is one of the meta-heuristic techniques which have drawn much attention in recent times. Its main concept is based on ideas of bird flocking and fish schooling in nature. Eberhart and Kennedy developed the particle swarm optimization algorithm in 1995. Since then, it has gained immense acceptance in its efficiency for solving complex optimization problems in various fields of engineering. In this work, two improved particle swarm optimization techniques are utilized to solve generation scheduling in micro grids. Here, the algorithm is initialized using a random number of particles. Each particle corresponds to the candidate solution to the problem to be considered. The implemented algorithm can be carried out using the following steps. Firstly, the input data are defined such as micro grid structure, operation characteristics of micro sources and grid, PV array and wind turbine predicted power output for each study period, real time price offer package for micro sources and grid, and the daily demand. Secondly, an initial population is considered based on following equation.

$$X^{o} = [X_1, X_2, \ldots X_N]^T \tag{9}$$

where X is the decision variable vector.

(i) Non-linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) Inertia weight, w(t): It is a parameter that is used to control the influence of previous velocities on the current obtained velocity. It reflects the balance between the global exploration and local exploration capabilities of particles in search space. Huang Chongpeng [9, 10] provided a new method of PSO with non-linearly decreasing inertia weight.

$$w_o = w_{end} + (w_{end} - w_{start}) * (1 - (t/t_{max})(k_1))(k_2) \tag{10}$$

where k_1, k_2 are two natural numbers, w_{start} is the initial inertia weight, w_{end} is the final value of weighting co-efficient, t_{max} is the maximum number of iteration and t is current iteration. The specified values of k_1, k_2 used are $k_1 > 1$ and $k_2 = 1$.

(ii) Time Varying Acceleration Co-efficients incorporated Particle Swarm Optimization Algorithm (TVAC PSO) The acceleration co-efficients are c_1 and c_2, kept constant for classical PSO. For PSO considering time varying acceleration coefficients (TVAC-PSO) the acceleration co-efficients c_1, c_2 are modified as follows:

$$c_1 = c_{1f} - c_{1i} * (iter/iter_{max}) + c_{1i} \tag{11}$$

$$c_2 = c_{2f} - c_{2i} * (iter/iter_{max}) + c_{2i} \tag{12}$$

where c_{1i}, c_{1f}, c_{2i}, c_{2f} are constants, *iter* is the current number of iteration, $iter_{max}$ is the maximum number of iterations [11]. Here, c_{1i} and c_{2i} represent initial values whereas c_{1f} and c_{2f} represent the final values of cognitive and social acceleration factors respectively. The range of values of c_{1i} and c_{1f} is chosen as 2.5 and 0.5, and the range of values of c_{2i} and c_{2f} is chosen as 0.5 and 2.5 respectively for the present work. Fitness is calculated using Eq. 4. The position vector and velocity vector in the above mentioned search space can be represented as $X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, v_{i3}, \ldots, v_{id})$ respectively. The best fitness value obtained by each particle at any given iteration is given by $P_i = (p_{i1}, p_{i2}, p_{i3}, \ldots, p_{id})$ which is termed *pbest* and the fittest particle found so far is termed as *gbest* given by $P_g = (p_{g1}, p_{g2}, p_{g3}, \ldots, p_{gd})$. Next, updated velocity and position of each particle is calculated using the following equations:

$$v_{(id,iter)} = w_{iter}v_{(id,iter-1)} + c_1r_1[p_{(id,iter-1)} - x_{(id,iter-1)}] + c_2r_2[p_{(gd,iter-1)} - x_{(id,iter-1)}] \tag{13}$$

$$x_{(id,iter)} = x_{(id,iter-1)} + v_{id,iter} \tag{14}$$

Individual best and global best: As the particle navigates through the search space, it compares its fitness value at present position to best fitness value it has attained at any time up to the present time. The best position which is associated with best fitness value encountered so far is termed individual best. Global best is the best position among total individual best positions achieved so far.

Stopping criteria: These are the conditions or criteria under which the search process will terminate. In this work the search process will terminate if the number of iterations reaches the maximum allowable limit of iterations.

Analytical approach for Proposed Algorithm Decision Variables: During implementation of the objective function, X is considered as decision variable vector. This comprises of generating units output power, quantity of power exchanged with utility, power in microgrid during demand response programs and on/off mode in a vision planned for day ahead, which are expressed as follows:

$$\begin{gathered} X = [P_g, U_g] \\ P_g = [P_{DG1}, P_{DG2}, \ldots, P_{DGNDg}, P_{s1}, P_{s2}, \ldots, P_{sNs}, P_{Grid}, P_{DR}] \\ U_g = [U_{DG1}, U_{DG2}, \ldots, U_{DGNDg}, U_{s1}, U_{s2}, \ldots., U_{sNs}, U_{Grid}, U_{DR}] \end{gathered} \tag{15}$$

In the above equation, N_{Dg} and N_s are total number of generation and storage units respectively, P_g is the active power vector including all distributed generation and storage units, utility power and active power which participate in the demand response programs. U_g represent the on or off states of all utility during each hour of day.

Case Study Analysis and constraint handling There are two different methods for constraint handling in optimization problems. One method is to take the total penalty violations into consideration with the objective function and create a fitness function. Here, the objective is to optimize the fitness function. Another technique for constraint handling is to start the optimization with a pre-specified feasible set of solutions and carry out the study with feasible set of solutions within the optimization process. The first approach is used in this paper for handling the constraints.

4 Results and Discussions

The proposed algorithm is implemented using MATLAB. It is applied on micro grid system comprising of micro turbine, fuel cell, battery, solar PV cell, wind turbine [5]. For the present problem effects of demand response has been analyzed. The number of iterations and population are chosen as 300 and 24 respectively.

Figure 1 shows the cost with NDIW PSO and TVAC PSO without demand response programs. Figure 2 shows the cost with NDIW PSO and TVAC PSO with demand response programs included in problem formulation. Table 5 compares the result obtained by PSO.

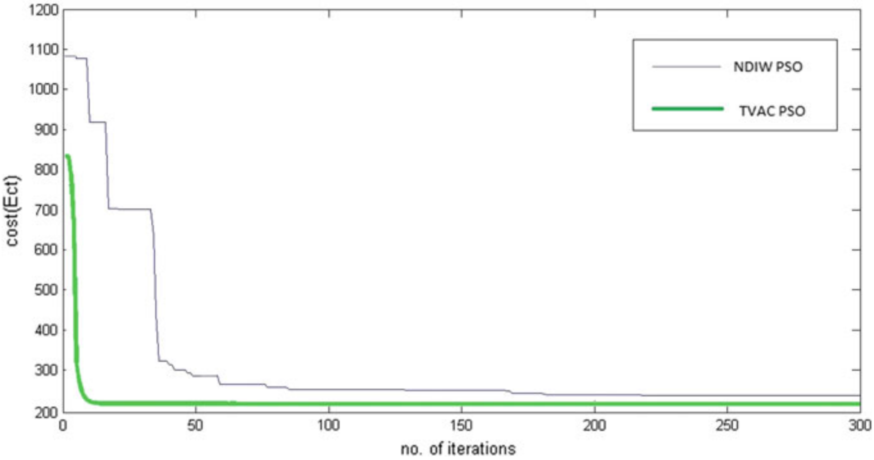


Fig. 1 Convergence characteristics for case 1A and case 1B (without DR)

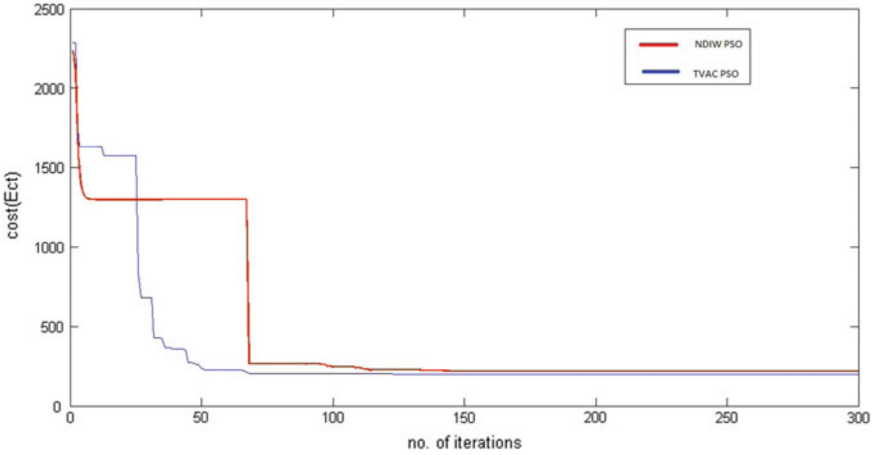


Fig. 2 Convergence characteristics for case 2A and case 2B (with DR)

Table 1 Economic power dispatch for case 1A

Hour	MT (kW)	FC (kW)	PV (kW)	WT (kW)	Battery (kW)	Utility (kW)
1	12.1	3.3595	0	7.2974	6.1026	24.486
2	11.3	4.0048	0	1.6640	−0.007	23.71
3	15.5	2.7146	0	4.6620	1.4944	8.0384
4	28.1	28.178	0	10.703	−26.60	14.877
5	11.5	5.9347	0	3.7475	−9.198	−28.72
6	27.0	20.735	0	8.3353	−25.26	−7.186
7	27.8	20.433	0	12.535	−24.84	−24.80
8	6.78	2.0636	1.0476	8.1519	8.4023	−8.155
9	5.90	22.643	20.036	9.9712	−13.75	−2.745
10	7.00	12.782	0.0782	7.8931	9.0488	9.4169
11	25.2	24.199	1.3504	7.0764	−14.33	−8.012
12	5.52	1.4050	10.596	4.7227	−30.0	10.697
13	22.2	10.528	10.358	13.712	−1.313	−30.0
14	9.40	27.436	7.4307	1.5970	14.085	−22.71
15	22.4	14.613	17.164	9.8536	−30.0	−29.78
16	9.80	9.6326	9.075	0.3533	9.8974	14.165
17	10.2	4.7275	6.206	4.4553	18.543	9.5755
18	18.1	24.066	0	3.3274	−28.81	27.306
19	8.10	18.333	0	4.4776	−4.799	−15.94
20	29.2	21.359	0	14.250	19.434	23.331
21	16.2	3.1121	0	7.3193	−28.76	−10.39
22	5.90	14.077	0	0.3442	−11.68	−11.32
23	19.9	28.451	0	7.5080	−14.67	−5.651
24	13.3	8.3979	0	7.0646	−16.60	−24.33

The micro grid considered for the present work has three different feeders namely residential, commercial and industrial consumers; the maximum electricity demand of residential load is 576.3 KWh, the maximum electricity demand of commercial load is 271.2 KWh, and the maximum electricity demand of industrial load is 847.5 KWh. It is considered here that the reactive power required by loads is compensated locally by means of capacitor placement in the relevant buses. The total load demand in 24 h without DR participation of responsive loads is 1703.0671 kW and with DR participation of responsive loads it is found 1702.0789 kW. Two case scenarios are considered where operating costs of all units are considered along with problem constraints in such manner that case 1 aims to obtain the optimum value of operating cost using Non-linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) and Time-Varying Acceleration Co-efficients incorporated Particle Swarm Optimization (TVAC-PSO) without demand response whereas case 2 minimizes operation costs using NDIW-PSO and TVAC-PSO via demand response participation.

Table 2 Economic power dispatch for case 1B

Hour	MT (kW)	FC (kW)	PV (kW)	WT (kW)	Battery (kW)	Utility (kW)
1	2.0017	21.437	0	7.549	13.9	20.27
2	27.894	15.763	0	10.16	−20.6	−2.73
3	3.8525	5.8831	0	6.436	−3.19	16.45
4	5.2013	21.722	0	9.667	−30.0	−18.03
5	18.823	20.357	0	2.219	−21.9	−8.94
6	9.8381	6.8176	0	13.30	−30.0	−24.59
7	18.145	28.571	0	8.346	19.2	−19.36
8	10.806	22.392	19.967	10.23	−12.2	17.493
9	8.5472	6.9221	7.0792	1.120	−18.6	6.2586
10	11.037	0.4209	14.340	4.202	2.47	11.714
11	24.213	4.6427	14.985	10.84	−30.0	−30.0
12	13.856	19.522	21.353	5.569	−24.3	13.551
13	4.5503	24.634	3.3261	3.121	−15.5	26.026
14	12.712	11.064	12.762	11.17	−12.8	−27.77
15	0.9197	15.490	17.624	1.513	−30.0	−30.0
16	20.732	18.127	8.2171	3.514	−1.08	−30.0
17	4.6522	6.8800	3.5641	8.490	−6.31	−18.72
18	8.3975	2.9198	0	11.23	−12.5	3.1837
19	23.164	10.351	0	11.30	−26.3	1.4144
20	16.153	15.925	0	10.20	4.19	9.2543
21	15.462	13.634	0	8.848	−7.08	−15.77
22	13.203	23.524	0	1.244	−1.29	−23.02
23	13.576	25.300	0	8.949	−8.46	0.6586
24	12.891	4.3340	0	12.32	−30.0	−5.371

4.1 Case 1A

For this case optimal operating cost without demand response is obtained using Nonlinear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO). Table 1 shows the optimal result for this case. In this case the total operating cost using Non linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) is found out to be 233.0964 Ect when demand response program is not considered.

4.2 Case 1B

To obtain the optimum value of operating cost using Time Varying Acceleration Co-efficients Particle Swarm Optimization (TVAC PSO) without demand response.

Table 3 Economic power dispatch for case 2A

Hour	MT (kW)	FC (kW)	PV (kW)	WT (kW)	Battery (kW)	Utility (kW)
1	24.1	27.779	0	1.7597	−0.100	−3.878
2	27.4	9.7459	0	12.466	12.451	19.407
3	28.1	3.3481	0	4.9371	3.9188	11.381
4	13.2	23.258	0	4.5196	9.3773	8.4341
5	13.0	25.412	0	9.6896	7.1226	−12.23
6	17.4	10.810	0	4.3200	8.3878	7.2181
7	24.4	21.189	0	9.7934	−0.144	−13.80
8	13.9	13.117	16.727	6.1376	−20.48	1.3509
9	18.6	18.172	9.4123	8.4919	23.274	−6.727
10	25.1	21.215	17.316	8.7513	5.2368	5.6983
11	22.2	2.8005	16.994	8.8651	1.7595	6.2197
12	28.8	25.909	2.9280	7.7693	−11.67	−30.0
13	9.10	4.2104	17.721	2.8031	−23.08	−30.0
14	14.5	11.582	1.9442	9.3002	−5.245	−18.61
15	24.3	2.8778	9.3524	8.2713	−4.126	−9.691
16	24.0	10.029	0.5615	3.2429	−17.20	−30.0
17	12.0	20.687	20.517	3.4287	−30.0	14.088
18	21.3	24.734	0	3.0930	−23.35	4.2453
19	6.66	19.833	0	10.257	−8.062	10.522
20	6.25	17.881	0	0.6772	−1.045	−5.994
21	17.8	17.036	0	11.113	28.470	−0.200
22	7.54	15.119	0	12.696	−30.0	22.990
23	16.3	24.040	0	3.2087	13.465	16.604
24	12.0	2.7116	0	6.1302	13.317	−8.674

Table 2 shows the optimal result for this case. In this case the total operating cost using Time Varying Acceleration Coefficients Particle Swarm Optimization (TVAC PSO) is found out to be 202.5812 Ect when demand response program is not considered.

4.3 *Case 2A*

To obtain the optimum value of operating cost using Non-linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) with demand response. Table 3 shows the optimal result for this case. In this case the total operating cost using Non linear Decreasing Inertia Weight Particle Swarm Optimization (NDIW-PSO) is found out to be 219.6025 Ect when demand response program is considered.

Table 4 Economic power dispatch for case 2B

Hour	MT (kW)	FC (kW)	PV (kW)	WT (kW)	Battery (kW)	Utility (kW)
1	7.7844	2.72	0	1.3545	28.9	20.43
2	9.6211	13.0	0	9.3101	27.4	18.06
3	13.861	6.11	0	1.4234	−5.62	−4.259
4	25.461	9.27	0	13.783	25.5	−7.201
5	26.829	25.3	0	0.8054	−23.9	18.942
6	8.7090	8.85	0	7.1667	−12.3	25.80
7	23.905	6.90	0	11.601	−27.6	−6.323
8	7.0480	15.7	5.3255	4.4398	14.73	−13.66
9	26.149	17.7	15.670	7.7713	−4.56	−30.0
10	28.526	8.45	2.4843	4.8186	−4.45	−7.671
11	22.310	20.5	1.5043	7.0173	−19.1	−18.42
12	18.160	4.61	6.8131	0.6172	9.813	−8.311
13	18.606	15.1	17.614	7.5087	−7.12	6.9401
14	20.835	13.2	12.945	7.8937	8.566	−29.15
15	21.033	7.34	8.5031	3.5817	21.98	−5.541
16	4.0334	24.7	24.162	3.3025	7.816	1.6828
17	12.106	14.5	14.945	9.4633	−30.0	−30.0
18	8.9138	23.1	0	6.0980	22.44	15.926
19	27.444	16.9	0	10.008	17.78	−25.98
20	26.992	24.4	0	12.983	−26.5	−0.963
21	28.319	20.4	0	6.0517	−21.5	11.613
22	24.353	13.0	0	0.5559	−22.4	5.0096
23	11.562	5.23	0	10.165	12.27	−30.0
24	9.8282	10.2	0	8.2742	17.86	−5.980

Table 5 Comparative analysis of results

Method	Parameter	NDIW-PSO	TVAC-PSO	PSO [5]
Without DR	Cost (Ect)	233.0964	202.5812	241.3
With DR	Cost (Ect)	219.6025	200.9721	231.3

4.4 Case 2B

To obtain the optimum value of operating cost using Time varying Acceleration Co-efficients incorporated Particle Swarm Optimization (TVAC PSO) with demand response. Table 4 shows the optimal result for this case. In case the total operating cost using Time Varying Acceleration Co efficients particle swarm optimization (TVAC PSO) is found out to be 200.9721 Ect when demand response program is considered (Table 5).

5 Conclusion

In this research work, uncertainties associated with fluctuating nature of wind and PV resources was considered, the objective function was analyzed for different cases. Today, with the emergence of renewable energies such as wind and solar and due to the uncertainties associated with their power production, there is a need to provide the necessary reserve and find precise solution for the further cover of these uncertainties. Recently, significant studies have been conducted on MG operation management. Since the main objective in operating a MG is to achieve operation at the minimum possible cost, therefore, most of the previous studies have investigated this area from different points of view. Some of the previous studies have been focused on the uncertainties caused by renewable resources of energy resulted from the prediction error of wind speed and solar radiation in a system. In this research work, the investigation on microgrid operation management is done to achieve minimum cost by using two different intelligent optimization algorithms which can be considered as a novel contribution to this area.

The results obtained show that operating costs were reduced by adaptive demand response programs which prove the superiority of proposed approach. A comparison of results for both NDIW-PSO and TVAC-PSO is carried out, also it is found that TVAC-PSO performs better than NDIW-PSO and PSO.

Acknowledgements The authors express gratitude towards the Department of Power Engineering, Jadavpur University for providing facilities for carrying this research work.

References

1. Jiayi, H., Chuanwen, J., Rong, X.: A review on distributed energy resources and micro grid. Int. J. Renew. Sustain. Energy Rev. **12**(9), 2472–2483 (2008)
2. Chowdhury, S., Crossley, P.: Microgrids and active distribution networks. The Institution of Engineering and Technology (2009)
3. Rouholamini, M., Mohammadian, M.: Energy management of a grid-tied residential-scale hybrid renewable generation system incorporating fuel cell and electrolyzer. J. Energy Build. **102**, 406–16 (2015)
4. Pandit, M., Srivastava, L., Sharma, M.: Environmental economic dispatch in multi area power system employing improved differential evolution with fuzzy selection. Appl. Soft Comput. **28**, 498–510 (2015)
5. Aghajani, R.G., Shayanfar, A.H., Shayeghi, H.: Presenting a multi-objective generation scheduling model for pricing demand response rate in micro-grid energy management. Energy Convers. Manag. **106**, 308–321 (2015)
6. Jabbari-Sabet, R., Moghaddas-Tafreshi, S.M., Mirhoseini, S.S.: Microgrid operation and management using probabilistic reconfiguration and unit commitment. Electr. Power Energy Syst. **75**, 328–336 (2016)
7. Abido, M.A.: Optimal power flow using particle swarm optimization. Electr. Power Energy Syst. **24**, 563–571 (2002)
8. Chen, C., Duan, S., Cai, T., Liu, B., Hu, G.: Smart energy management system for optimal micro grid economic operation. IET Renew. Power Gener. **5**(3), 258–267 (2011)

9. Chongpeng, H., Yuling, Z., Dingguo, J., Baoguo, X.: On some non-linear decreasing inertia weight strategies in particle swarm optimization. In: Proceedings of the 26th Chinese Control Conference, Hunan, China, Zhangjiajie, pp. 570–753 (2007)
10. Imran, M., Hashima, R., Khalidb, Noor Elaiza Abd: An overview of particle swarm optimization variants. Procedia Eng. **53**, 491–496 (2013)
11. Ratnaweera, A., Halgamuge, S.K., Watson, H.C.: Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Trans. Evol. Comput. **8**(3), 240–255 (2004)

Global Exponential Stability of Non-autonomous Cellular Neural Network Model with Time Varying Delays

M. Chowdhury and P. Das

Abstract We have considered a general form of non-autonomous cellular neural network with time varying delays in this paper. We have estimated the upper bound of solutions of the system by introducing different parameters and considered some conditions on it. We have derived the conditions of boundedness and global exponential stability of the model which is initially unstable for some parameter values using Young Inequality technique and Dini derivative. Several examples and their computer simulations are given to illustrate the effectiveness of obtained results.

Keywords Non-autonomous system · Time-varying delay · Boundedness · Global exponential stability

1 Introduction

In recent times, delayed cellular neural networks have attracted substantial attention due to their ability for functions of pattern recognition, image processing, optimization signal processing and associative memories. An Artificial Neural Network (ANN) is a conceptual model based on the structure and functions of biological neural networks which consist of numerous cells called neurons and their interconnections. The mathematical models in the light of neural dynamics represent the complex behavior of the human brain. Cellular Neural Network or CNN is a computing model of locally connected circuits (called cells) similar to neural networks, but communication occurs in neighboring units only.

In brain, neurons are connected to each other by axons. The phenomenon of polarization and depolarization or change in cell potential occurs during neural transmis-

M. Chowdhury · P. Das (✉)
Department of Mathematics, Indian Institute of Engineering Science
and Technology, Shibpur, Howrah 711103, India
e-mail: prithadas01@yahoo.com

M. Chowdhury
e-mail: madhusreechowdhury92@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_17

sions. Initially the neurons are negatively charged. Influx of sodium ions make the interior of the cell positive. Efflux of potassium ions begins towards the end of this stage, the interior of the cell tends to become electrically negative. This prevents further progress of depolarization beyond +35 mV which demands boundedness of the state of the neuron.

Previously, the dynamics of autonomous neural networks have been vastly studied and are applied in many fields such as control system, pattern identification, signal processing, associative memory [1–4]. It is seen that due to evolutionary processes of physical systems as well as internal and external disturbances, the strength of the neuron vary with time. So it is important to study the non-autonomous mathematical model of the neural system [3, 5–7].

Synaptic delay is the time taken by the nerve impulse to cross one synapse. Changes in the delay parameter causes inconvenient shifts in the phases of the neural signals leading to consequences explained as tremor in the fingers, Parkinson's disease, motion disorders in the case of burst of epilepsy, etc. [8–10]. Time delay leads to oscillation, divergence, instability and chaos in neural networks [11–13]. In this case, the stability of the system is particularly important to manufacture microelectronic neural networks. So it is important to consider time varying delay in the model to analyse different dynamical behavior.

Instability of the dynamical system initiates the phenomenon of neurodegeneration, which leads to the gradual loss of structure or function of neurons resulting in neurodegenerative diseases like amyotrophic lateral sclerosis, Parkinson's, Alzheimer's and Huntington's diseases [4, 7, 14]. So it is essential to attain global exponential stability to intensify the convergence rate of neural network approaching to equilibrium.

In this paper the main motivation is to investigate the boundedness and global exponential stability of solutions for the non-autonomous system of the cellular neural network model. In this model we have considered delay only in the interconnecting neurons.

The paper is organized in the following manner. In Sect. 2, we have described the model and have considered few conditions on the parameters. In Sect. 3 we have shown that the solutions of the model are bounded. In Sect. 4 we have found a criterion for the global exponential stability of system. In Sect. 5 computer simulations of various examples are discussed showing variety of dynamical scenario for different parameter values.

2 Model Description

Here, we have considered the following artificial non-autonomous n-neural network model with time varying delays:

$$\frac{dy_i(t)}{dt} = -s_i(t)y_i(t) + \sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n} d_{ijl}(t)g_{ijl}(y_j(t-\tau_{ijl}(t))) \\ + d_{iii}(t)g_{iii}(y_i(t)) + J_i(t), \qquad i = 1, 2, \ldots, n \tag{2.1}$$

Here $y_i(t)$ corresponds to the state variable associated with the ith neuron at time t; n is the number of state variables and m is the number of axons emitting signal from ith unit in m number of pathways to the jth unit; $g_{ijl}(y_i(t))$ represents the neuron activation function at time t; $d_{ijl}(t)$ represents the connection weights of the jth unit on the ith unit at time $t - \tau_{ijl}(t)$; $\tau_{ijl}(t)$ corresponds to the transmission delay signal of the ith unit along the lth axon of the jth unit at time t and is a nonnegative function; $J_i(t)$ denotes the external input signal on the ith unit at time t; $s_i(t)$ represents the rate with which the ith unit will reset its potential to the resting state when detached from network.

In the above system we consider the following conditions on the parameters of the model:

- **(A1)** $s_i(t), d_{ijl}(t) \quad (i, j = 1, 2, \ldots, n, l = 1, 2, \ldots, m)$ are bounded and continuous functions defined on t $\in R_+$.
- **(A2)** $g_{ijl}(v) \quad (i, j = 1, 2, \ldots, n, l = 1, 2, \ldots, m)$ satisfy the Lipschitz condition, i.e. there are constants $\varsigma_{ijl} > 0$ such that;
  $|g_{ijl}(v_1) - g_{ijl}(v_2)| \leq \varsigma_{ijl}|v_1 - v_2| \; \forall \; v_1, v_2 \in R.$
- **(A3)** $\tau_{ijl}(t) \; (i, j = 1, 2, \ldots, n, l = 1, 2, \ldots, m)$ are nonnegative, bounded and continuous functions defined on R_+.
- **(A4)** If there exist constants $a_{ijl}, c_{ijl} \in R, u_i > 0 \, (i, j = 1, 2, n, l = 1, 2, m), b > 1, \alpha > 0$ such that;
  $bu_i s_i(t) - (b-1)\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n} u_j|d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}} \varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} - bu_i|d_{iii}(t)|\varsigma_{iii}$
  $- \sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n} u_j \varsigma_{ijl}^{l_{ijl}} |d_{ijl}(t)|^{a_{ijl}} > \alpha \; ; \forall \; t \geq 0$ and i=1,2,…,n.

Let τ=sup$\left\{\tau_{ijl} : t \in [0, \infty), i, j = 1, 2, \ldots, n, l = 1, 2, \ldots, m\right\}$ and $b > 1$ be a known constant. Let $C[-\tau, 0]$ be the Banach space of continuous functions $\varphi(\theta) = (\varphi_1(\theta), \varphi_2(\theta), \ldots, \varphi_n(\theta)) : [-\tau, 0] \to R^n$. We define $\|\varphi\| = sup_{-\tau<\theta<0}|\varphi(\theta)|$ where $|\varphi(\theta)| = [\max_{1\leq i\leq n} |\varphi(\theta|^s]^{\frac{1}{s}}$.

We assume that all the solutions of system (2.1) satisfy the initial condition:

$$y_i(\theta) = \varphi_i(\theta), \quad \theta \in [-\tau, 0]; \quad i = 1, 2, \ldots, n \tag{2.2}$$

We know by the fundamental theory of functional differential equations, system (2.1) has a unique solution $(y_1(t), y_2(t), \ldots, y_n(t))$ which satisfies the initial condition (2.2).

In this paper, we use some other definitions of boundedness and global exponential stability and lemmas on Young inequality and Dini derivative which will be used to prove the results from [3].

3 Boundedness

Theorem 3.1 *Let A_1-A_4 hold. Then all the solutions of the system (2.1) are defined and are bounded on R_+*

Proof Let us assume that y(t)=$(y_1(t), y_2(t), \ldots, y_n(t))$ be any solution of the system (2.1) with initial function $\varphi \in \mathrm{C}[-\tau, 0]$ at $t = 0$. Let $y_i(t) = u_i v_i(t)$, $(i = 1, 2, \ldots, n)$ then (2.1) becomes:

$$\frac{dv_i(t)}{dt} = -s_i(t)v_i(t) + \sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{1}{u_i} d_{ijl}(t) g_{ijl}(u_j v_j(t - \tau_{ijl}(t))) + \frac{1}{u_i} d_{iii}(t) g_{iii}(u_i v_i(t)) + \frac{1}{u_i} J_i(t). \tag{3.1}$$

where $i = 1, 2, 3, \ldots, n$. By calculating the upper right derivative $D^+|v_i(t)|^s$ and get :

$$D^+|v_i(t)|^s \leq -bs_i(t)|v_i(t)|^b + b\sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)||v_i(t)|^{(b-1)} \varsigma_{ijl}|v_j(t - \tau_{ijl}(t))|$$
$$+ \frac{b}{u_i}|J_i(t)||v_i(t)|^{(b-1)} + b|d_{iii}(t)||v_i(t)|^b \varsigma_{iii} + b\sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{1}{u_i}|d_{ijl}(t)||g_{ijl}(0)||v_i(t)|^{(b-1)}$$
$$+ \frac{b}{u_i}|d_{iii}(t)||v_i(t)|^{(b-1)}|g_{iii}(0)| \tag{3.2}$$

By using Young's Inequality: $b \sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)||v_i(t)|^{(b-1)} \varsigma_{ijl}|v_j(t - \tau_{ijl}(t))|$

$$\leq (b-1)[\sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}} \varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} |v_i(t)|^b] + \sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)|^{a_{ijl}} \varsigma_{ijl}^{c_{ijl}} |v_j(t - \tau_{ijl}(t))|^b$$

Therefore from (3.2),

$$D^+|v_i(t)|^b \leq \frac{1}{u_i}[-bu_i s_i(t) + (b-1)\left[\sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} u_j|b_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}} \varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} |v_i(t)|^b\right]$$
$$+bu_i|d_{iii}(t)|\varsigma_{iii} + bR^* v_i(t)^{-1}]|v_i(t)|^b + \sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)|^{a_{ijl}} \varsigma_{ijl}^{c_{ijl}} \max_{t-\tau\leq b_1\leq t} |v_j(b_1)|^b$$
$$\forall t \geq 0; i = 1, 2, \ldots, n \tag{3.3}$$

where
$R^* = \sup_{t\geq 0; i=1,2,\ldots,n} \Big\{|J_i(t)| + \sum_{l=1}^{m} \sum_{j=1, j\neq l}^{n} |d_{ijl}(t)||g_{ijl}(0)| + |d_{iii}(t)||g_{iii}(0)|$: $t \geq 0, i = 1, 2, \ldots, n\}$

By choosing a large constant q from A_4 such that:

$$-bu_i s_i(t) + (b-1)\sum_{l=1}^{m}\sum_{j=1, j\neq l}^{n} u_j |d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}} \varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} + bu_i|d_{iii}(t)|\varsigma_{iii} + bR^* q^{-1} + \sum_{l=1}^{m}\sum_{j=1, j\neq l}^{n} u_j |d_{ijl}(t)|^{a_{ijl}} \varsigma_{ijl}^{c_{ijl}} < 0 \tag{3.4}$$

$\forall t \geq 0 \quad and \quad i = 1, 2, \ldots, n \quad and \quad ||\varphi|| \leq \min_{1\leq i\leq n}\{u_i\}_q$. Thus we get; $|y_i(t)| < u_i q$, $\forall\ t \geq 0$, $i = 1, 2, \ldots, n$. We prove this by contradiction. If this statement is false then $\exists$ some i and time $t_1 \geq 0$ such that $|y_i(t_1)| = u_i q$, $D^+|y_i(t_1)|^b \geq 0$ and $|y_j(t)| \leq u_j q$; $\forall - \tau \leq t \leq t_1$ and $j = 1, 2, \ldots, n$.

From this we further obtain $|v_i(t_1)| = q$, $D^+|v_i(t_1)|^b \geq 0$ and $|v_j(t)| \leq q$; $\forall - \tau \leq t \leq t_1$ and $j = 1, 2, \ldots, n$.

Then from (3.3) and (3.4) we get, $D^+|v_i(t_1)|^b < 0$. This is a contradiction.

So, $|v_i(t)| < q$; $\forall t \geq 0$. So $|y_i(t)| \leq u_i q$; $\forall t \geq 0$.

Hence this solution $y(t) = (y_1(t), y_2(t), \ldots, y_n(t))$ is defined and is bounded on R^+.

Hence proved.

4 Global Exponential Stability

Theorem 4.1 *Let $A_1 - A_4$ hold. Then (2.1) is globally exponentially stable.*

Proof For proving this at first we need to assume two solutions of the system (2.1). Let $y_1(t) = (y_{11}(t), y_{12}(t), \ldots, y_{1n}(t))$ and $y_2(t) = (y_{21}(t), y_{22}(t), \ldots, y_{2n}(t))$ be any two solutions with initial conditions: $y_1(\theta) = \varphi_1(\theta)$ and $y_2(\theta) = \varphi_2(\theta)$; $\forall\theta \in [-\tau, 0]$ where $\varphi_1(\theta) = (\psi_{11}(\theta), \varphi_{12}(\theta), \ldots, \varphi_{1n}(\theta))$ and $\varphi_2(\theta) = (\varphi_{21}(\theta), \varphi_{22}(\theta), \ldots, \varphi_{2n}(\theta))$ By Theorem 3.1 we obtain that $y_1(t)$ and $y_2(t)$ are defined and are bounded $\forall t \in R^+$

Let $z_i(t) = y_{1i}(t) - y_{2i}(t)$ and $z(t) = (z_1(t), z_2(t), \ldots, z_n(t))$.
From A_2 we get:

$$\frac{dz_i(t)}{dt} \leq -s_i(t)z_i(t) + \sum_{l=1}^{m}\sum_{j=1, j\neq l}^{n} |d_{ijl}(t)|\varsigma_{ijl}|z_j(t - \tau_{ijl}(t))| + |d_{ijl}(t)|\varsigma_{iii}|z_i(t)| \tag{4.1}$$

Let $z_i(t) = u_i x_i(t)$; $i = 1, 2, \ldots, n$.
Then (4.1) is transformed to:

$$\frac{dx_i(t)}{dt} \leq -s_i(t)x_i(t) + \sum_{l=1}^{m}\sum_{j=1, j\neq l}^{n} \frac{u_j}{u_i}|d_{ijl}(t)|\varsigma_{ijl}|x_j(t - \tau_{ijl}(t))| + |d_{iii}(t)|\varsigma_{iii}|x_i(t)| \tag{4.2}$$

From (A_4), we choose a constant $\varepsilon > 0$ such that:

$$bu_i s_i(t) - (b-1)\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}\frac{u_j}{u_i}|d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}}\varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} - bu_i|d_{iii}(t)|\varsigma_{iii}$$
$$-e^{b\varepsilon t}\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}\frac{u_j}{u_i}|d_{ijl}(t)|^{a_{ijl}} - b\varepsilon u_i > \frac{\alpha}{2}$$

$\forall t \geq 0$ and i=1,2,…,n.
Let $q_i(t) = e^{b\varepsilon t}|x_i(t)|^b$.

We calculate the upper right derivative $D^+ q_i(t)$ and by using Young's Inequality in (4.2) we get;

$$D^+(q_i(t)) \leq b(\varepsilon - s_i(t) + d_{iii}(t)\varsigma_{iii}(t))q_i(t) + (b-1)\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}\frac{u_j}{u_i}|d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}}\varsigma_{ijl}^{\frac{b-c_{ijl}}{b-1}} q_i(t)$$
$$+\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}\frac{u_j}{u_i}|d_{ijl}(t)|^{a_{ijl}}\varsigma_{ijl}^{c_{ijl}}e^{b\varepsilon t}\sup_{t-\tau\leq b_1\leq t} p_j(b_1) \tag{4.3}$$

We choose a constant R $\geq$ 1 such that $R^b \geq (\min_{1\leq i\leq n}\{u_i^b\})^{-1}$, then for each $i = 1, 2, \ldots, n$;

$$q_i(0) = |x_i(0)|^b = \frac{|\varphi_{1i}(0) - \varphi_{2i}(0)|^b}{u_i^b} \leq \frac{||\varphi_1 - \varphi_2||}{min\{u_i^b\}} < R^b||\varphi_1 - \varphi_2||^b \tag{4.4}$$

Hence we further get $q_i(t) < R^b||\varphi_1 - \varphi_2||^b \quad \forall t \geq 0$ and $i = 1, 2, \ldots, n$.

We prove this by contradiction. If it is not true, then $\exists$ some i and $t_1 > 0$ such that $q_i(t_1) < R^b||\varphi_1 - \varphi_2||^b$; $D^+(q_i(t_1) \geq 0$ and $q_j(t) < R^b||\varphi_1 - \varphi_2||^b \quad \forall -\tau \leq t \leq t_1$; $j = 1, 2, \ldots, n$. But from (4.3) we get;

$$D^+(q_i(t_1)) \leq \frac{1}{u_i}b(\varepsilon - s_i(t) + d_{iii}(t)\varphi_{iii})u_i + (r-1)\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}u_j|d_{ijl}(t)|^{\frac{b-a_{ijl}}{b-1}}\varphi_{ijl}^{\frac{b-c_{ijl}}{b-1}}$$
$$+\sum_{l=1}^{m}\sum_{j=1,j\neq l}^{n}u_j|d_{ijl}(t)|^{a_{ijl}}\varsigma_{ijl}^{c_{ijl}}e^{b\varepsilon t}]R^b||\varphi_1 - \varphi_2||^b$$
$$\leq -\frac{1}{2u_i}\alpha R^b||\varphi_1 - \varphi_2||^b < 0$$

This is a contradiction. Therefore, $q_i(t) < R^b||\varphi_1 - \varphi_2||^b \, \forall t \geq 0$ and $i = 1, 2, \ldots, n$ That is; $e^{b\varepsilon t}|x_i(t)|^b \leq R^b||\varphi_1 - \varphi_2||^b$; $\forall t \geq 0$ Hence $|x_i(t)| \leq R||\varphi_1 - \varphi_2||e^{-\varepsilon t}$; $\forall t \geq 0$ Further, we have $\sum_{j=1}^{n}|y_{1i}(t) - y_{2i}(t)| \leq nuR||\varphi_1 - \varphi_2||e^{-\varepsilon t}; \forall t \geq 0$ Therefore $u = max_{1\leq i\leq n}\{u_i\}$. This shows that from definition (2), system (2.1) is globally exponentially stable. Hence proved. □

5 Numerical Simulations and Results

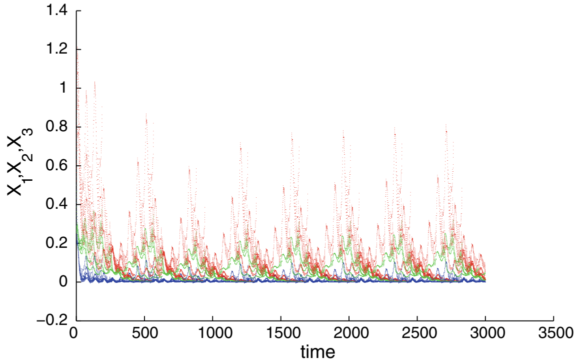
Here we have considered the following examples of our model for the verification of the analytical results obtained. The considered model is of the form:

$$\frac{dy_1(t)}{dt} = -s_1(t)y_1(t) + d_{11}(t)g_{11}(y_1(t)) + d_{12}(t)g_{12}(y_2(t-\tau_{12}(t))) + d_{13}(t)g_{13}(y_3(t-\tau_{13}(t)))$$
$$\frac{dy_2(t)}{dt} = -s_2(t)y_2(t) + d_{21}(t)g_{21}(y_1(t-\tau_{21}(t)) + d_{22}(t)g_{22}(y_2(t)) + d_{23}(t)g_{23}(y_3(t-\tau_{23}(t)))$$
$$\frac{dy_3(t)}{dt} = -s_3(t)y_3(t) + d_{31}(t)g_{31}(y_1(t-\tau_{31}(t))) + d_{32}(t)g_{32}(y_2(t-\tau_{32}(t))) + d_{33}(t)g_{33}(y_3(t))$$

Example 5.1 Here we consider the activation function $g_{ij}(x) = tanh(x)$ and the parameters are chosen as: $r_1(t) = r_2(t) = r_3(t) = 1 + \sin t \quad \tau_{ij} = -0.49 + 0.2 \sin t i, j = 1, 2, 3$
$b_{11}(t) = 1 + \cos t$; $b_{12}(t) = -(1 + \sin t)$; $b_{13}(t) = 5 + \cos t$
$b_{21}(t) = 1 + \cos t$; $b_{22}(t) = -(1 + \sin t)$; $b_{23}(t) = 1 + \sin t$
$b_{31}(t) = -(1 + \cos t)$; $b_{32}(t) = 1 + \sin t$; $b_{33}(t) = -(1 + \sin t)$
Here we choose $u_i = a_{ij} = c_{ij} = \eta_{ij} = 1$ & $s = 2$. We see that the system experiences unstable behavior. Figures 1 and 2 depicts the unstable nature.

Example 5.2 Again we have considered the activation function $g_{ij}(x) = \tanh(x)$. Obviously it is unbounded and satisfy Lipschitz condition. We take: $s_1(t) = s_2(t) = s_3(t) = 5 + \sin t$
$\tau_{ij} = 0.8 - 0.2 \sin t \quad i, j = 1, 2, 3$
$d_{11}(t) = 5 + 4 \cos t$; $d_{12}(t) = -(5 + 5 \sin t)$; $d_{13}(t) = 1 + \cos t$
$d_{21}(t) = 5 + \frac{1}{5} \cos t$; $d_{22}(t) = -(5 + 5 \sin t)$;
$d_{23}(t) = -(2 + \sin t)$
$d_{31}(t) = 1 + \cos t$; $d_{32}(t) = -(1 + \frac{1}{2} \sin t)$;
$d_{33}(t) = -(1 + \sin t)$

Fig. 1 Solution trajectories $x_1(red), x_2(green), x_3(blue)$ of the model showing unstable behavior when delay $\tau_{ij} = -0.49 + 0.2 \sin t$



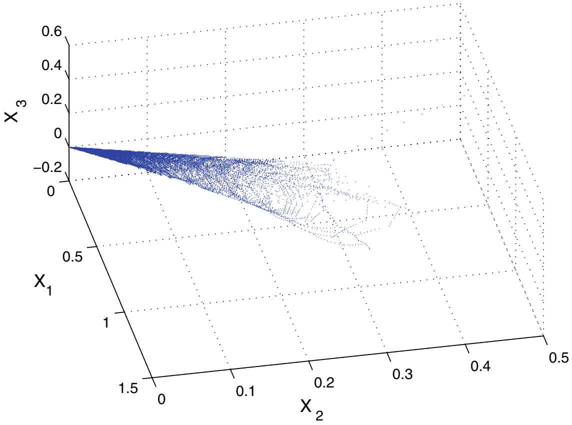


Fig. 2 3D plot showing unstable behavior when delay $\tau_{ij} = -0.49 + 0.2 \sin t$

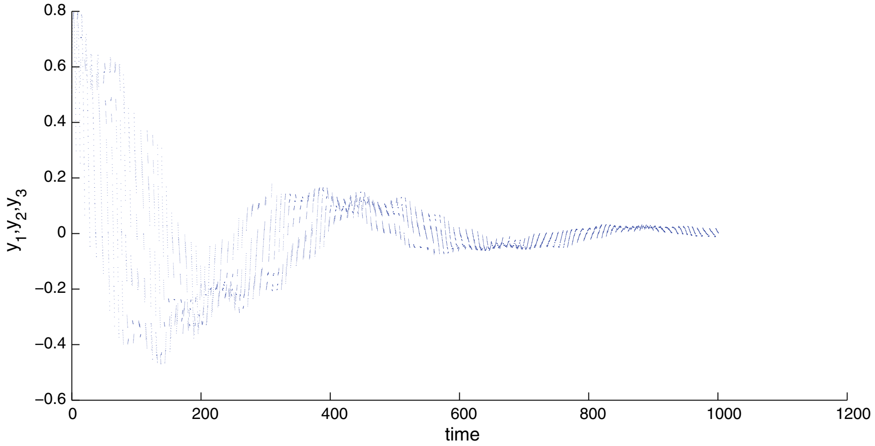


Fig. 3 Solution trajectories y_1, y_2, y_3 of the model showing Stable behavior satisfying A_4 when delay $\tau_{ij} = 0.8 - 0.2 \sin t$

We choose $u_i = a_{ij} = c_{ij} = \varsigma_{ij} = 1$ & $b = 2$. Clearly the parameters satisfy A_4 and thus by Theorem 3.1 and 4.1 the solutions of system (2.1) are bounded and globally exponentially stable. Figures 1 and 2 depict the dynamical behavior.

In Figs. 1 and 2 we see the unstable behavior of the system (2.1) which shows that the system is unstable in nature. But in the Figs. 3 and 4 we see the stable behavior of the system (2.1). As in Example 5.2 we have chosen the parameters in such a way such that (A4) is satisfied. So it clearly illustrates that Theorem 4.1 is satisfied. Here we have considered the delay functions $\tau_{ijl}(t)$ $(i, j = 1, 2, \ldots, n, l = 1, 2, \ldots, m)$ as nonnegative, bounded and continuous functions on R_+ which make the results simpler alike [4] where discrete and distributed delay is considered. Also here we have

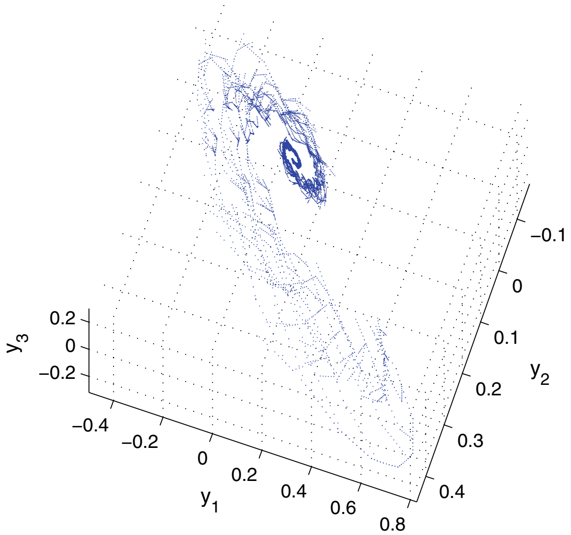


Fig. 4 3D plot showing stable behavior when delay $\tau_{ij} = 0.8 - 0.2 \sin t$

considered non-autonomous system alike the model in [4] which is an autonomous system. Our model is more realistic as it is seen that due to progressive processes of physical systems as well as internal and external disruptions, the strength of the neuron vary with time.

6 Conclusions

Here we have considered the non-autonomous cellular neural network model with time varying delays. In many prevailing studies of dynamics of neural networks constant delays have been taken into account. But in this paper we have considered time varying delay which is very crucial as it has been observed that the occurrence of the awkward shifts in the phases of neural signals substantially affects the performance of the system. Numerous diseases are the consequences of this. In this paper the delay functions $\tau_{ijl}(t)$ are non-negative, bounded and continuous in the set of positive real numbers. Our neural network model is more practical as we have omitted the delay in the self connecting neurons. A signal transmission from a neuron to itself is clearly negligible. This modification simplifies the existing results. Here we have proved the boundedness of the state of neuron and the global exponential stability of the system based on certain criteria which is feasible for analyzing and designing the neural network. We have estimated the upper bound of the solutions of the system by applying Young inequality technique and Dini derivative.

The analytical results involving global exponential stability are demonstrated by numerical results using Matlab 15. These results can be used to design stable neural networks with time varying parameters and delays in practice.

References

1. Chua, L.O., Yang, L.: Cellular neural networks: theory. IEEE Trans. Circuits Syst. **35**, 1257–1272 (1988)
2. Chua, L.O., Yang, L.: Cellular neural networks: applications. IEEE Trans. Circuits Syst. **35**, 1273–1290 (1988)
3. Jiang, H., Teng, Z.: Dynamics of neural networks with variable coefficients and time-varying delays. Neural Netw. **19**, 676–683 (2006)
4. Li, Q., Wang, S., Wu, Y.: Exponential stability of the neural networks with time-varying discrete and distributed delays. In: 2010 Chinese control and decision conference. IEEE (2010)
5. Cao, J., Yang, J.: Boundedness and stability for Cohen Grossberg neural network with time-varying delays. J. Math. Anal. Appl. **296**, 665–685 (2004)
6. Cao, J., Wang, J.: Global asymptotic stability of a general class of recurrent neural networks with time-varying delays. IEEE Trans. Circuits Syst.-I **50**, 3444 (2003)
7. Jiang, M., Shen, Y., Liu, M.: Global exponential stability of non-autonomous neural networks with variable delay. In: Advances in neural networks (2005). ISNN 108–113
8. Baldi, P., Atiya, A.F.: How delays effect Neural Dynamics and Learning. IEEE Trans. Neural Netw. **5**(4) (1994)
9. Das, P., Kundu, A.: Bifurcation and chaos in delayed cellular neural network model. J. Appl. Math. Phys. **2**, 219–224 (2014)
10. Kundu, A., Das, P.: Global stability, bifurcation, and chaos control in a delayed neural network model. In: Advances in artificial neural systems, vol. 2014, Article ID 369230, 8 p. (2014)
11. Cao,J., Song, Q.: Stability in Cohen Grossberg-type bidirectional associative memory neural networks with time-varying delays. Nonlinearity **19**, 1601–1617 (2006)
12. Jiang, H., Li, Z., Teng, Z.: Boundedness and stability for nonautonomous cellular neural networks with delay. Phys. Lett. A **306**, 313325 (2003)
13. Zhang, Q., Wei, X., Xu, J.: Delay-dependent exponential stability criteria for non-autonomous cellular neural networks with time-varying delays. Chaos Solitons Fractals (2008) (Elsevier)
14. Zhou, L., Zhang, Y.: Global exponential stability of cellular neural networks with multi-proportional delays. Int. J. Biomath. **8**(6), 1550071 (17 pages) (2015) (World Scientific Publishing Company)

Topological Configuration in the Heisenberg Spin Sequence and DNA

Subhamoy Singha Roy

Abstract We have considered here that the conformational properties of a DNA molecule can be mapped onto a Heisenberg spin system when spins are located on the axis forming an antiferromagnetic chain. This helps us to study the topological properties of a DNA molecule in terms of $SU(2)$ gauge fields. The entanglement entropy of the spin system in a supercoil has been determined and it is pointed out that this effectively corresponds to the thermodynamic entropy. The model reproduces salient features of the Rod-Like-Chain model avoiding the "RLC model crisis".

Keywords Heisenberg spin system · Antiferromagnetic spin chain · Lagrangian · Entanglement entropy · Gauss linking number · Chern-Simons topology · DNA supercoil · Nonquantized monopole

1 Introduction

It is now well known that DNA can be regarded as a physical elastic object in a viscous environment. Two strands of double helix are antiparallel and two polynucleotide chains are coiled about the same axis such that B-DNA (Z-DNA) has right-handed (left-handed) helical sense. In earlier studies it has been shown that elongation verses force characteristics of a DNA molecule [1] can be described [2] by the worm-like-chain (WLC), which essentially depicts a chain by an elastic continuous curve at thermal equilibrium with a single elastic constant, the persistence length A characterizing the bending energy [3]. It has been shown that the analytic solution of WLC can be achieved through the mapping to a quantum mechanical problem. In fact its partition function corresponds to Euclidean path integral for a quantum dumbbell when rotations and vibrations are ignored so that it represents a rigid rotator. The generalization of the WLC with twist rigidity was considered by Bouchiat and Merzard [4] and it was shown that a DNA molecule at small supercoiling can be described

S. Singha Roy (✉)
Department of Physics, JIS College of Engineering (Autonomous), West Bengal University of Technology, Kalyani, Nadia 741235, India
e-mail: ssroy.science@gmail.com; sub-hamoy.singharoy@jiscollege.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_18

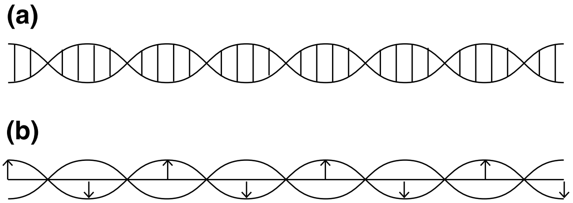


Fig. 1 DNA molecule as a spin chain. **a** DNA molecule and **b** the corresponding antiferromagnetic spin system

by a thin elastic rod involving the twist rigidity. It is called the rod-like-chain (RLC) model. The partition function can be mapped onto the path integral representation for a quantum charged particle in the field of a magnetic monopole with nonquantized charge [4]. The theory is singular in the continuum limit and needs to be regularized at an intermediate length scale. It is found that the model is in good agreement with the experimental data at small supercoiling region.

In this note we shall study the specific properties of DNA by taking into consideration that the conformational properties of a DNA molecule can be mapped onto a Heisenberg spin system. In fact as two polynucleotide chains are coiled about the same axis with a specific helical sense in a DNA molecule, we may visualize it such that a spin with a specific orientation is inserted on the axis indeed as the rotation of the two strands about the axis in the positive direction through an angle π is identical with that of rotation about the axis in the negative direction through an angle π, we can uplift the rotation group $SO(3)$ to $SU(2)$. A spin is represented as an $SU(2)$ gauge bundle [5]. In view of this we can consider that twisting of the two strands can be taken to be represented by a spin inserted on the axis such that two adjacent coils have opposite orientations of the spin. This essentially implies that two strands twisting about the axis in the opposite direction can be designated by two spins having orientations $+1/2$ and $-1/2$. When these spins are inserted on the axis such that two spins having opposite orientations are located in the two adjacent coils with lattice spacing of one period of helix this represents an antiferromagnetic spin chain. The nearest neighbour spin—spin interaction can be viewed as to represent the twisting energy (Fig. 1).

In an earlier paper [6] we have studied the denaturation of DNA by mapping its conformational properties onto a Heisenberg spin chain. It has been shown that denaturation occurs when the entanglement entropy of the spin chain vanishes and appears to be caused by quantum phase transition induced by a quench [7] when the temperature effect is incorporated in the quench time. Here torsion takes the role of the external field. In a recent paper [8] it has been pointed out that the sequence heterogeneity interplays with the entropy effects and we have the onset of denaturation bubbles. The melting profiles for different sequence-sequence DNA molecules have been studied and the results are found to be in excellent agreement with experiments. We shall study here the topological and elastic properties of DNA by mapping it onto Heisenberg spin chain by utilising the fact that a spin can be treated as an $SU(2)$ gauge bundle so that a gauge theoretic treatment of DNA can be formulated. We have shown in some earlier papers [5, 9] that a spin can be represented by an

$SU(2)$ gauge bundle and we can study a spin system in terms of these gauge fields. Here we shall show that when the conformational properties of DNA are mapped onto a Heisenberg spin chain we can represent the topological properties such as the linking number in terms of these gauge fields. The bending as well as the twisting energy can be formulated in terms of the Lagrangian involving these gauge fields. It is pointed out that the thermodynamic entropy of a DNA supercoil can be represented by the entanglement entropy of the spin chain. We have analytically estimated the free energy per unit length of a DNA supercoil, which corresponds to that of a strongly confined polymer in a tube and the result is found to be in good agreement with that obtained from computer simulation. It is shown that the antiferromagnetic spin chain manifests certain salient features of the rod-like chain (RLC) model when the partition function is mapped onto the path integral representation of a quantum charged particle in the field of a magnetic monopole with nonquantized monopole charge. It is argued that this formulation helps us to avoid the "RLC model crisis", which involves the violation of rotational invariance due to the nonquantized value of angular momentum.

In Sect. 2 we depict DNA as a spin chain. In Sect. 3 we study the topological properties of a DNA molecule and also consider the entropy of a DNA supercoil and we compare the present formalism with the RLC model from an analysis of the spin system.

2 DNA as a Spin Chain

We consider that two polynucleotide chains are coiled about the same axis with a specific helical sense in a DNA molecule. This can be viewed as if a spin with a specific orientation is inserted on the axis of the coil such that two adjacent coils have opposite orientations of the spin. In fact with each turn two strands move in the opposite side of the axis and so the spin orientation assigned for the two adjacent coils should be opposite to each other. It may be noted that twisting of the two strands in mutually opposite directions can be taken to imply that two strands can be designated by two spins having orientations $+1/2$ and $-1/2$. When these two spins having opposite orientations are inserted on the axis such that these are located in the two adjacent sites with lattice spacing of one period of helix this represents an antiferromagnetic spin chain. The nearest neighbour spin—spin interaction can be viewed as to represent the twisting energy.

As the two strands can be taken to be designated by two opposite spin orientations we can consider that a unit vector depicting the tangent $\mathbf{t} = \partial_s \mathbf{r}$, where $\mathbf{r}(s)$ is a space curve parametrized by the arc length s of the helix can be mapped onto the spin vector when the spin is located at the spatial point x on the molecular axis about, which the two strands move. It is noted that any arbitrary deformation of the interwound helices such as bending, twisting and stretching will deviate the axis from a simple straight line. A spin vector in the Lie algebra of $SU(2)$ representation can be constructed with bosonic or fermionic oscillators. We write the spin vector $\mathbf{S}(x)$ as

$$\mathbf{S}(x) = \psi_\alpha^\dagger(x)\boldsymbol{\sigma}_{\alpha\beta}\psi_\beta(x), \tag{1.1}$$

where $\psi^\dagger(\psi)$ is the fermionic oscillator function and $\boldsymbol{\sigma}_{\alpha\beta}$ is the vector of Pauli matrices. A unit vector $\mathbf{n}$ is constructed as

$$n = \left(\psi_1^*\psi_2^*\right)\boldsymbol{\sigma}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \tag{1.2}$$

with

$$\begin{aligned}\psi_1 &= \left(\cos\tfrac{\theta}{2}\right)e^{i\phi/2}\\ \psi_2 &= \left(\sin\tfrac{\theta}{2}\right)e^{-i\phi/2}\end{aligned} \tag{1.3}$$

This helps us to write the spin vector (1.1) in terms of the unit vector (1.2) as

$$\mathbf{S}(x) = (\sqrt{3}/2)\psi_\alpha^\dagger(x)\boldsymbol{\sigma}_{\alpha\beta}\psi_\beta \tag{1.4}$$

We can construct a unit vector n_μ with $\mu = 0, 1, 2, 3$ in $3 + 1$ dimensions incorporating the unit vector n given by Eq. (1.2)

$$n_\mu = (1/\sqrt{2})\left(\phi_1^*\phi_2^*\right)\sigma_\mu\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \tag{1.5}$$

with $\sigma_0 = I$, being the identity matrix and $\boldsymbol{\sigma}$ is the vector of Pauli matrices. We construct the topological current

$$\mathbf{J}_\mu = \left(1/12\pi^2\right)\varepsilon_{\mu\nu\lambda\sigma}\varepsilon_{abcd}n_a\partial_\nu n_b\partial_\lambda n_c\partial_\sigma n_d, \tag{1.6}$$

where (a, b, c, d) correspond to $(0, 1, 2, 3)$ and $(\mu, \nu, \lambda, \sigma)$ correspond to space-time indices. The current J_μ can be written in the form [10].

$$\mathbf{J}_\mu = \left(1/24\pi^2\right)\varepsilon_{\mu\lambda\sigma}Tr\left(g^{-1}\partial_\nu g\right)\left(g^{-1}\partial_\lambda g\right)\left(g^{-1}\partial_\sigma g\right) \tag{1.7}$$

with $g = n_0 I + i\mathbf{n}\cdot\boldsymbol{\sigma}$, which belongs to the group $SU(2)$. If we demand that in Euclidean four-dimensional spacetime the field strength $F_{\mu\nu}$ of a gauge potential A_μ vanishes at all points on the boundary S^3 of a certain volume V^4 inside, which $F_{\mu\nu} \neq 0$ the gauge potential tends to a pure gauge towards the boundary and we write

$$A_\mu = g^{-1}\partial_\mu g \tag{1.8}$$

with $g \in SU(2)$.

We can now write the topological current given by (1.7) as [5]

$$J_\mu = \left(1/16\pi^2\right)\varepsilon^{\mu\nu\lambda\sigma}Tr\{A_\nu F_{\lambda\sigma} + (2/3)A_\nu A_\lambda A_\sigma\} \tag{1.9}$$

with A_μ given by Eq. (1.8). It is noted that as the spin vector is constructed from the unit vector n given by (1.2), which is incorporated in the current J_μ as is evident from Eq. (1.6), we can associate spin with this current J_μ. In fact we can consider the topological Lagrangian in terms of the $SU(2)$ gauge fields in affine space

$$L = -(1/3)Tr\varepsilon^{\mu\nu\alpha\beta}F_{\mu\nu}F_{\alpha\beta} \tag{1.10}$$

This gives rise to the topological current [11]

$$\mathbf{J}_\mu = \varepsilon^{\mu\nu\lambda\sigma}\mathbf{a}_\nu \times \mathbf{f}_{\lambda\sigma} = \varepsilon^{\mu\nu\lambda\sigma}\partial_\nu\mathbf{f}_{\lambda\sigma}, \tag{1.11}$$

where we have taken the $SU(2)$ gauge field A_μ and the corresponding field strength $F_{\mu\nu}$ as

$$A_\mu = \mathbf{a}_\mu \cdot \boldsymbol{\sigma} \text{ and } F_{\mu\nu} = \mathbf{f}_{\mu\nu} \cdot \boldsymbol{\sigma} \tag{1.12}$$

$\boldsymbol{\sigma}$ being vector of Pauli matrices. From this it appears that the spin vector $\mathbf{S}(x)$ can be depicted as the topological current $\mathbf{J}_\mu$ given by Eq. (1.11). In terms of this current a spin system on a lattice can be viewed as if currents are located on the vertices when gauge fields lie on links [9].

3 Topological Properties of a DNA Molecule

In length scales of a large number of base pairs DNA in vivo is organized into topologically independent loops. The two strands of a circular DNA molecule possess as a topological invariant the number of times they wind around each other, which is known as the linking number. A B-DNA molecule has one right-handed twist per h = 3.4 nm along its length. When these are closed in a planar circle without twisting of the ends the resulting linking number is $Lk_0 = L/h = \omega_0 L/(2\pi)$, where L is the length and ω_0 is the spatial rotation rate of the base pairs about the central axis [12]. Deviations in the twisting rate from ω_0 is measured relative to Lk_0 through the parameter defining the excess linking $\sigma = (\Delta Lk/Lk_0)$, where $\Delta Lk = Lk - Lk_0$. The linking number Lk is expressed as $Lk = Tw + Wr$, where Tw represents the twist corresponding to the rotation of the internal degrees of freedom about the molecule axis and Wr represents the writhe [13–15]. The twist measures the winding of one curve about the other. It can be mathematically expressed as [12]

$$Tw = \int_0^L (ds/2\pi)[\omega_0 + \Omega(s)] = Lk_0 + \Delta Tw \tag{1.13}$$

where Ω is the twist strain measuring the excess or deficit rotation of the base pairs about the axis and s defines the arc length. The writhe characterizes the chiral deformation of a curve. One can assign an orientation to a curve and compute the sum of signed crossings in a planar projection along every direction. Wr is equal to the average of such sums over all projections [16].

We can understand twist and writhe more explicitly from the following considerations [17]. Let us take that there are two oriented non-intersecting closed curves γ_1 and γ_2 and imagine that γ_2 carries a unit current in the direction of its orientation. Evidently this gives rise to a magnetic field. Now, applying Ampere's law to deduce the number of times γ_1 encircles γ_2 and then using the Biot-Savert law, which helps us to estimate the magnetic field due to the current we can write

$$Lk(\gamma_1, \gamma_2) = \oint_{r_1} \mathbf{B}(\mathbf{r}_1)d\mathbf{r}_1 = \frac{1}{4\pi}\oint_{\gamma_1}\oint_{\gamma_2} \{(\mathbf{r}_1 - \mathbf{r}_2)(d\mathbf{r}_1 \times d\mathbf{r}_2)\}/|\mathbf{r}_1 - \mathbf{r}_2| \tag{1.14}$$

This is called the Gauss linking number, which is symmetric under the interchange of $\gamma_1 \leftrightarrow \gamma_2$. For a closed loop of DNA we identify the curves with the strands. Two such parallel curves (γ_1, γ_2) form the edges of a ribbon of width ε. We now consider (γ_1, γ_2) the curve $\mathbf{r}(t)$ running along the axis of the ribbon midway between and γ_1 and γ_2 denote it as γ. The unit tangent to γ at the point $\mathbf{r}(t)$ is given by

$$t(t) = \dot{\mathbf{r}}(t)/|\mathbf{r}(t)| \tag{1.15}$$

the dot denoting the differentiation with respect to t. A unit vector $\mathbf{u}(t)$ perpendicular to $\mathbf{t}(t)$ lies in the ribbon pointing from $\mathbf{r_1}(t)$ to $\mathbf{r_2}(t)$. We can now write

$$\begin{aligned}\mathbf{r}_1(t) &= \mathbf{r}(t) - (1/2)\varepsilon\mathbf{u}(t)\\ \mathbf{r}_2(t) &= \mathbf{r}(t) + (1/2)\varepsilon\mathbf{u}(t)\end{aligned} \tag{1.16}$$

Expressing

$$\dot{\mathbf{u}} = \boldsymbol{\omega} \times \mathbf{u}, \tag{1.17}$$

where $\boldsymbol{\omega}(t)$ is the angular velocity vector we can define twist as

$$Tw = \frac{1}{2\pi}\oint_{\gamma} (\boldsymbol{\omega} \cdot \mathbf{t})dt \tag{1.18}$$

If we let $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ equal to the single axis curve $\mathbf{r}(t)$ in the integrand of (1.14) the self-linking integral known as the writhe is given by

$$Wr = \frac{1}{4\pi}\oint_{\gamma}\oint_{\gamma} \big\{(\mathbf{r}(t_1) - \mathbf{r}(t_2))\big(\dot{\mathbf{r}}(t_1) \times \dot{\mathbf{r}}(t_2)\big)\big\}dt_1dt_2/|(\mathbf{r}(t_1) - \mathbf{r}(t_2))|^3 \tag{1.19}$$

The write is a property of the overall shape of the curve γ and is independent of the ribbon that contains it. We have the relation

$$Lk = Tw + Wr, \tag{1.20}$$

which is known as the Calugareanu-White-Fuller relation. It may be mentioned here that the expressions (1.19) and (1.20) were also derived by Frank-Kamenetskii and Vologodskii [18] in the context of their study on the physical properties of circular DNAs. Indeed these authors have presented the mathematical basis of the theory of knots and links as well as the theory of ribbons in their analysis on the topological aspects of polymers and their biophysical applications. From the point of view that a DNA molecule can be depicted as a spin system we can determine the linking number from the spin degrees of freedom. It is noted that the expression of the current J_μ associated with the spin given by Eq. (1.9) essentially corresponds to the Chern-Simons secondary characteristics class. The topological charge

$$q = \int J_0 d^3x \tag{1.21}$$

corresponds to the winding number associated with the homotopy $\pi_3\left(S^3\right) = Z$ and can be written as

$$q = 2\mu = \left(11/24\pi^2\right) \int_S^3 \varepsilon^{\mu\nu\lambda\sigma} Tr\left(g^{-1}\partial_\nu g\right)\left(g^{-1}\partial_\lambda g\right)\left(g^{-1}\partial_\sigma g\right) \tag{1.22}$$

This charge q essentially represents the Pontryagin index, which is an integer and the relation $q = 2\mu$ implies that μ corresponds to the magnetic monopole strength. This Pontryagin index can be written as the integral in the four-dimensional manifold M_4 as

$$q = \left(1/16\pi^2\right) \int_{M_4} Tr(F \wedge F) \tag{1.23}$$

where F is the two-form related to the field strength associated with the $SU(2)$ gauge field A_μ. Now, from the relation

$$\int_{M_4} Tr(F \wedge F) = \int_{M_3} Tr(A \wedge dA + (2/3)A \wedge A \wedge A) \tag{1.24}$$

where M_3 is a three-dimensional manifold and A is the one-form corresponding to the $SU(2)$ gauge field A_μ we note that the RHS of Eq. (1.24) represents the Chern-Simons invariant and is thus found to be associated with the Pontryagin index in a four-dimensional manifold. Noting that the Pontryagin index corresponding to the charge related to the gauge current J_μ given by Eq. (1.9), which is associated with the spin, we can consider spin in the framework of Chern-Simons topology. In fact from Eq. (1.11) we note that any component of the spin vector can be written as

$$J^a_\mu(a = 1, 2, 3) = \varepsilon^{\mu\nu\lambda\sigma} a_\nu \partial_\lambda a_\sigma \tag{1.25}$$

where a_ν corresponds to an Abelian gauge field. When we project it onto a three-dimensional manifold this corresponds to the Chern-Simons term $\varepsilon^{\nu\lambda\sigma} a_\nu \partial_\lambda a_\sigma$. In the Abelian theory we consider the one-form a associated with the gauge field a_ν and choose the action

$$S = (k/8\pi) \int_{M_3} \varepsilon^{ijk} a_i \partial_j a_k, \tag{1.26}$$

where k is an integer. We now pick up some circles C_a and some integers n_a corresponding to representations of the Abelian gauge group. It is assumed that two curve C_a and C_l do not intersect for $a \neq b$. As shown by Polyakov [19] the expectation value of the product

$$W = \prod_a \exp(in_a) \int_{C_a} a \tag{1.27}$$

with respect to the measure determined by $e^i S$ is given by

$$\langle W \rangle = \exp\left[(I/2k) \sum_{a,b} n_a n_b \int_{C_a} dx^i \int_{C_b} dy^j \varepsilon_{ijk} \left\{ (x - y)^k / |x - y|^3 \right\} \right] \tag{1.28}$$

For $a \neq b$ this integral is essentially the linking number

$$\phi(C_a, C_b) = \frac{1}{4\pi} \int_{C_a} dx^i \int_{C_b} dy^j \varepsilon_{ijk} \left\{ (x - y)^k / |x - y|^3 \right\} \tag{1.29}$$

As long as C_a and C_b do not intersect $\phi(C_a, C_b)$ is a well-defined integer. Thus, ignoring the term $a = b$, we have

$$\langle W \rangle = \exp\left[(2\pi i/k) \sum_{a,b} n_a n_b \phi(C_a, C_b) \right] \tag{1.30}$$

The appearance of linking number from the current representing the spin suggests that the linking number can be associated with a spin system.

From this analysis it appears that when a DNA molecule is represented as a spin system the linking number can be considered as a topological invariant. It should be mentioned that though the linking number is a topological invariant when it is split into twist (*Tw*) and writhe (*Wr*) these entities are not topological invariants. Since the linking number of a closed DNA molecule remains constant during any deformation of the molecule that preserves chemical bonding, it can only be changed by mechanisms in which chemical bonds are disrupted [16].

The entanglement entropy can be written in the form

$$S \approx \pi r^2/4r_0^2 \tag{1.31}$$

r being the radius of the tube and r_0 a fundamental area unit. We now identify $r \approx A$, A being the bending elastic constant and r_0 is taken to represent the radial displacement of a given point on the coil, which satisfies the condition $|r_0| < R$ [17]. Taking the mean value $|r_0| = R/2$, the entanglement entropy is given by

$$S \approx \pi A^2/R^2 \tag{1.32}$$

Similarly for the displacement of a given point on the coil along the supercoil axis, which is of the order of πP, P being the pitch

$$S \approx \pi A^2/(\pi P)^2 \tag{1.33}$$

So the total entropy is given by

$$S = \pi A^2/R^2 + \pi A^2/(\pi P)^2 \tag{1.34}$$

Now, we observe that this entanglement entropy effectively corresponds to the thermodynamic entropy. Indeed for a tube of narrow radius entanglement entropy cannot vanish, whereas in the limiting case of radius $r \to 0$ we can think of zero radius (straight line) when the total elastic energy vanishes at zero temperature. In this case the entanglement entropy also vanishes.

In fact the Berry phase acquired by a spin state in a spin 1/2 chain when the Hamiltonian is parallel transported along a closed circuit is given by [18–21]

$$\int_0^{2\pi} A_\phi d\phi = 2\pi(1 - \cos\theta) \tag{1.35}$$

$$\phi_B = \pi(1 - \cos\theta) \tag{1.36}$$

It is noted that the holonomy given by Eq. (1.35) is twice this phase factor. The effective monopole charge associated with a spin state in an entangled spin system is given by the relation

$$\mu_{\text{eff}} = (1/2)(1 - \cos\theta) \tag{1.37}$$

which follows from the relation $\phi_B = 2\pi\mu$ [22]. Evidently the monopole charge takes nonquantized value apart from the situation when the polar angle θ of the spin axis with the quantization axis is given by $\theta = 0, \pi/2$ and π. This nonquantized monopole charge takes values on the *RG* flow of the monopole charge in an entangled spin system. Thus, we observe that when we map a DNA molecule onto an antiferromagentic spin system the salient features of the RLC model are well reproduced avoiding the “RLC model crisis”.

4 Discussion

We have shown that a DNA molecule can be treated as a quantum spin system such that spins are located on the axis forming an antiferromagnetic chain. These spins can be associated with $SU(2)$ gauge field currents when gauge fields lie on the links. We have studied the topological properties of a DNA molecule as bending (curvature) and twisting (torsion) in terms of these gauge fields. A significant result of this formalism is that bending and twisting are not independent entities. In fact bending influences the propagation of twisting strain along the DNA, which has been supported by experiments.

The formulation of DNA supercoil in terms of an antiferromagnetic spin chain gives rise to the entanglement entropy, which induces the entropic potential associated with the free energy per unit length corresponding to the entropy cost of confining a stiff polymer inside a narrow tube. The entanglement entropy effectively represents the thermodynamic entropy and this repulsive entropic potential opposes the elastically driven collapse of a supercoil, which can occur at zero temperature.

Finally, we have observed that the spin chain model reproduces the salient feature of the RLC model when the partition function is mapped onto the path integral representation of a quantum charged particle in the field of a magnetic monopole with nonquantized charge. However, in this case the RLC model crisis is avoided as the nonquantized monopole charge appears to take values on the *RG* flow of the monopole charge in an entangled spin system.

Acknowledgements The authors are grateful to express my deep gratitude to my beloved Sir Prof. Pratul Bandyopadhyay for helpful discussion.

References

1. Smith, S.B., Finzi, L., Bustamante, C.: Science **258**, 1122 (1992)
2. Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S.B.: Science **265**, 1599 (1994)
3. Fixman, M., Kovac, J.: J. Chem. Phys. **58**, 1564 (1973)

4. Bouchiat, C., Mezard, M.: Phys. Rev. Lett. **80**, 1556 (1997); Euro. Phys. J. E **2**, 377 (2000)
5. Bandyopadhyay, P.: Proc. Roy. Soc (London) A. **466**, 2917 (2010)
6. Singha Roy, S., Bandyopadhyay, P.: Phys. Lett. A **337**, 2884 (2013)
7. Basu, B., Bandyopadhyay, P., Majumdar, P.: Phys. Rev. A **86**, 022303 (2012)
8. Singha Roy, S., Bondyopadhyay, P.: Euro. Phys. Lett. **109**, 48002 (2015)
9. Goswami, G., Bandyopadhyay, P.: J. Math. Phys. **34**, 749 (1995)
10. Abanov, A.I., Wiegmann, P.B.: Nucl. Phys. B **570**, 685 (2000)
11. Carmeli, M., Malin, S.: Ann. Phys. **103**, 208 (1977)
12. Marko, J.F., Siggia, E.D.: Phys. Rev. E **52**, 2912 (1995)
13. Calugareanu, I.: Crechoslovak. Math. J. **11**, 588 (1961)
14. White, J.H.: Am. J. Math. **91**, 693 (1969)
15. Fuller, F.B.: Proc. Natl. Acad. Sci. (USA) **75**, 3557 (1978)
16. Swigon, D., Benham, C.J., et al. (eds.): Mathematics of DNA Structure, Function and Interactions. Springer (2009)
17. Burkhardt, T.W.: J. Phys A Math. Gen. **28**, L 629 (1995)
18. Basu, B., Bandyopadhyay, P.: Int. J. Geo. Meth. Mod. Phys. **9**, 707 (2007)
19. Basu, B., Bandyopadhyay, P.: J. Phys. A **41**, 055301 (2008)
20. Singha Roy, S.: Theor. Phys. **2**(3), 141 (2017)
21. Singha Roy, S., Bandyopadhyay, P.: Phys. Lett. A **382**, 1973 (2018)
22. Bandyopadhyay, P.: Proc. Roy. Soc (London) A **467**, 427 (2011)

Modelling Studies Focusing on Microphytobenthos and Its Role in Benthic-Pelagic Coupling

Swagata Sinha, Arnab Banerjee, Nabyendu Rakshit and Santanu Ray

Abstract Microphytobenthos (MPB) are much less conspicuous than other groups of organisms occupying the benthic photic zone but are ecologically much important due to their roles in benthic-pelagic coupling, nutrient cycling, association with macrophytes and as food source for many benthic as well as pelagic predators. While there have been numerous studies on MPB, modelling studies on its role in benthic-pelagic couple (BPC) and BPC in general has been negligible. Most modelling studies on MPB have focused on its productivity which depends on many factors such as irradiance, tidal wave, sediment texture and so on. The role of MPB in benthicpelagic coupling helps in maintaining the resilience of the ecosystem, in increasing the productivity of the system and also regulating the dynamics of the entire system. The present work is a review article that highlights such modelling studies that have focused on MPB, its role in coupling and their effectiveness in depicting the real system and to suggest a future modelling approach that could be applied to study the aquatic ecosystem as whole including both benthic and pelagic food webs.

Keywords Microalgae · Light attenuation · Nutrient recycling · Vertical migration · Biofilm resuspension

1 Introduction

Benthos (Greek: '*bevoo*'—meaning bottom) refers to the community of organisms that live on, in, or near the seabed—the benthic zone. Benthos in euphotic zone is predominately characterised by photosynthetically active microorganisms though macroscopic vegetation is also seen. The term microphytobenthos (MPB) refers to microscopic, unicellular eukaryotic algae (Baccilariophyceae, Chlorophyceae and Dinophyceae) and prokaryotic Cyanobacteria that grow in a wide range of habitats

S. Sinha · A. Banerjee · N. Rakshit · S. Ray (✉)
Systems Ecology & Ecological Modelling Laboratory, Department of Zoology,
Visva-Bharati University, Santiniketan 731235, India
e-mail: sray@visva-bharati.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_19

such as intertidal sand and mud flats, salt marshes, submerged aquatic vegetation beds. Concentration of benthic microalgae in sediment-water interface microhabitat is very high and may exceed biomass of phytoplankton of overlying water column. Contribution of MPB to primary production in littoral zones is significant, though it is less conspicuous than macroalgae or vascular plants [1].

Modelling studies on aquatic ecosystem till date mainly focus on pelagic part of the food web. However, the distinction between benthic and pelagic life styles is not possible in the strictest sense, especially in shallow water systems since, the exchange of organisms and nutrients between the sediment and water column is common—the MPB being resuspended by water currents and wave action, dwell in the water column and contribute to the planktonic community [2]. Benthic-Pelagic coupling (BPC) connects the two food webs inticrately and effects the entire system. This review work focuses on the different studies done on MPB, BPC and MPB's role in BPC and in the end, suggests an approach, that can be used to study the entire aquatic food web highlighting the role of MPB and BPC.

1.1 Importance of Microphytobenthos

Diversity and functional role of microphytobenthic communities has been a major research topic since a long time [3]. MPB have very important ecological role as follows

1. Benthic microalgae and bacteria are the most productive within tropical and temperate intertidal sediments [4].
2. They also serve as important habitats for smaller invertebrates such as chironomids, amphipods and several smaller meiofauna [5].
3. Benthic microalgae attached to macrophytes trap nutrients before they reach water column and returns them to sediments when epiphytic algae settle [6].
4. They serve as chemical modulators in aquatic ecosystems [7].
5. Stabilising the substrata is another significant role of benthic microalgae [8]. Several species of benthic microalgae such as pennate diatoms and blue green algae, while growing on the surface layer of the bottom sediments, get attached to the sand and detritus by the mucilaginous substances called extracellular polymeric substances (EPS) secreted by themselves, thus stabilising the substrata.
6. MPB are excellent indicators of short term stress owing to their short life span, sensitivity to nutrient variation and, widespread habitat from marine to freshwater ecosystems.
7. Owing to their role in nutrient cycling and trophic interactions, MPB also acts as connecting or coupling link between benthic and pelagic food webs (discussed below).

2 Role of Microphytobenthos in Nutrient Cycling

MPB strongly impacts mineralization pathways and nutrient fluxes at the sediment interface by virtue of photosynthesis and nutrient assimilation [9]. Lake epipelic and epipsammic algae also produce extrapolymeric substances, an important source of fresh organic matter, that can fuel heterotrophic bacteria, thus indirectly influencing the benthic food web [10]. Several studies have shown that MPB modulate the flux of nutrients across the water-sediment interface [11].

Amount of particular nutrient (Nu) sequestered by sediments can be estimated as a function of nutrient sedimentation rate ($dSNu/dt$) and net nutrient efflux rate ($dENu/dt$) across sediment-water interface [12]. Particulate detrital algal fluff resuspension is an important process in most aquatic ecosystems but cannot be measured directly. To prevent an overestimation of particulate organic matter sedimentation, a resuspension factor, assuming resuspension of 50% of apparent nutrient storage, was also included. Estimated apparent Nu-storage was quantified, assuming steady state for MPB biomass using the following mass balance equation (1)

$$\text{Apparent } Nu_{\text{storage}} = 0.5 \times \frac{dsNu}{dt} - \frac{deNu}{dt} \tag{1}$$

2.1 Nitrogen Uptake by Microphytobenthos

Denitrification—the conversion of nitrate or nitrites to nitrogen and mineralization—the conversion of organic nitrogen to ammonium are two of the most important steps of nitrogen cycling, determining the fate of the nitrogen in the system; more denitrification results in more loss of nitrogen from the system.

1. Sundb*ä*ck et al. [13] showed that MPB incorporate between 40 and 100% of remineralized nitrogen and Hochard et al. [14] showed that competition for nitrate and ammonia between MPB and bacteria led to reduction of coupled nitrification-denitrification.
2. Studies have shown that N assimilation by MPB is greater by a factor of 70 than the loss of N by denitrification [14].
3. It plays an important role in regulating the timing and dynamics of nitrogen fluxes between sediment and overlying water column over time scales of days [15].
4. MPB-colonized-sediment is a major source of organic matter, directly sustaining both benthic and pelagic higher trophic levels [16] and contributes to water column via re-suspension of living cells and particulate organic nitrogen [17].
5. Nitrogen loss from the sediment is however, regulated by both MPB and physical processes and the extent of effect of both varies seasonally [18].

3 Factors Affecting Microphytobenthos Productivity

MPB show vertical migration along the sediment in response to different factors, thus only the proportion of MPB occupying the top most layer of sediment is effectively productive. Productivity of MPB is regulated by many factors, highlighted in Table 1.

Table 1 Factors affecting MPB productivity

Factors	Effects
Light	1. Light is only available on the mud surface during day time emersion periods and high turbidity prevents even that [19]
	2. Microalgae migrate actively to the top illuminated sediment layer [20]
	3. The algae actively migrate back into the deeper layers to prevent resuspension and predation before immersion period or when light availability is low
	4. Migratory behaviour of MPB helps MPB to maximise photosynthesis by reducing exposure to photoinhibitory levels of light [21]
Resuspension of MPB	1. Wind induced waves [22] or bioturbation [23] cause resuspension of MPB
	2. Changes the biomass of MPB on the sediment surface
	3. Increases the turbidity of the water column causing light attenuation
	4. More nutrients are available to the pelagic life forms resulting in their increased reproduction, which increases the turbidity of the water column
	5. Light attenuation results in decreased MPB productivity [22]
Tidal waves	MPB productivity at low tide (when sufficient light reaches the sediment) is almost twice the productivity at high tide [24]
Temperature	[25] showed that the mud surface temperature influences the photosynthetic capacity of the MPB especially during low tide
Grazing pressure	1. When grazing pressure reduces the total biomass of the MPB in the photic layer of the sediment [26], productivity is reduced
	2. Increased MPB diversity (under grazing pressure) affects productivity [27]
Nutrition	Changes in nutrient concentrations due to resuspension by disturbance or any sediment biotic factor can significantly alter community structure of MPB
Sediment texture	1. [28] showed that sediment texture type affect the vertical distribution of MPB
	2. Chlorophyll-a was distributed down to 10 cm in the sandy sites as compared to the distribution down to 4 cm in the muddy sites

4 Benthic-Pelagic Coupling

In freshwater lakes, benthic primary production was seen to account for up to 98% of total primary production [29]. Griffiths et al. [30] defined benthic-pelagic coupling (BPC) as those processes which connect the bottom substrate and the water column habitats through the exchange of mass, energy and nutrient. Essential ecosystem functions, such as production and energy transfer in food webs, biogeochemical cycling and provisioning of fish nursery areas [31] are supported by multiple and interacting benthic-pelagic coupling processes [32]. Contributions of 30–80% of the phytoplanktonic nitrogen requirement in shallow (5–50 m) coastal environments originate from the sediments [33]. Nutrients exported to sediment in the form of detritus, when re-introduced in water column as re-mineralised inorganic nutrients creates potential for high primary productivity in the water column even when terrestrial nutrient input is low [34].

4.1 Mechanisms of BPC

BPC mechanisms are essential for the ecological understanding of the structure and function of aquatic ecosystems [35]. Processes involved in benthic pelagic coupling are shown in Fig. 1.

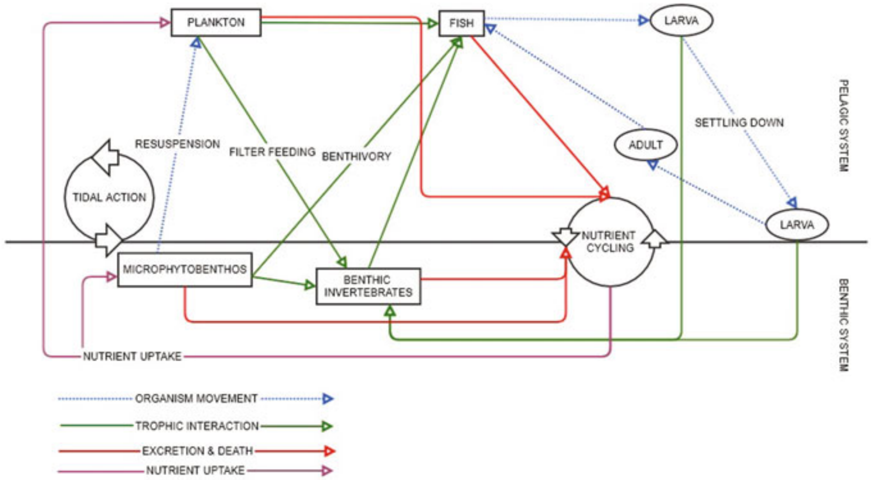


Fig. 1 Diagrammatic representation of different processes involved in benthic pelagic coupling (adapted from [36])

The major mechanisms for benthic-pelagic coupling can be grouped into the following according to Baustian et al. [36].

(1) **Organism movement**: Zooplankton and fish by virtue of their migratory behaviour couple benthic and pelagic systems [37]. Pelagic larval stages of marine benthic macrofauna and many freshwater and marine benthic macroinvertebrates during emergence serve as prey for pelagic fish [38]. Fishes experience ontogenetic shifts in behavior and feeding patterns between pelagic and benthic zones.

(2) **Trophic interactions**: Studies in the Gironde estuary showed that demersal fishes fed on supra, epibenthic, and pelagic preys [39]. Phytoplankton and microzooplankton face considerable grazing pressure from benthic suspension feeders such as bivalves [40]. Demersal fishes on the Bay of Biscay continental shelf consume benthic epifauna and are consumed by pelagic top predators [41].

(3) **Biogeochemical cycling**: Benthic production in shallow lakes is stimulated by nutrient sources such as decomposing carcasses of fishes [37]. Benthic zebra mussels in freshwater systems feed on pelagic phytoplankton and then excrete dissolved nutrients (P and N) back into the pelagic water column [42]. In shallow ecosystems where sufficient light penetrates below the surface mixed layer, macrophytes, macroalgae, and microphytobenthos dominate benthic primary production at the sediment-water interface [43] and gross primary production can exceed respiration [44]. Bioturbation links the benthic and pelagic systems by suspending sediment that influences phytoplankton and zooplankton recruitment [45] and also stimulates mineralization of organic matter and the release of nutrients [46], thereby affecting the growth of phytoplankton in the pelagic zone [47].

5 Microphytobenthos and Benthic Pelagic Coupling

Processes that connect benthic and pelagic habitats influence the ecology of both, particularly in ecosystems with large areas of benthic habitat relative to water volume [48].

MPB in Nutrient Cycling: MPB can potentially regulate the light and nutrient environment of an aquatic system. If MPB restricts resuspension of particulate material and efflux of dissolved nutrients, the overlying water column is clear and nutrient poor which allows more light to reach the benthos, thus facilitating more productivity in light limited MPB and does not facilitate bloom. Both benthic and pelagic systems are intimately linked and there are two stable states-one benthic dominated and one pelagic dominated as shown in Fig. 2. Role of MPB in nutrient cycling has already been discussed in details Sect. 2 of this article. **MPB in Trophic Interaction**: In the study of Zanden et al. [49] it was seen that prey fish population alone was not sufficient to support the piscivorous fish population and top-down effect was weak. However, in presence of zoobenthos (whose population was supported by MPB), piscivorous

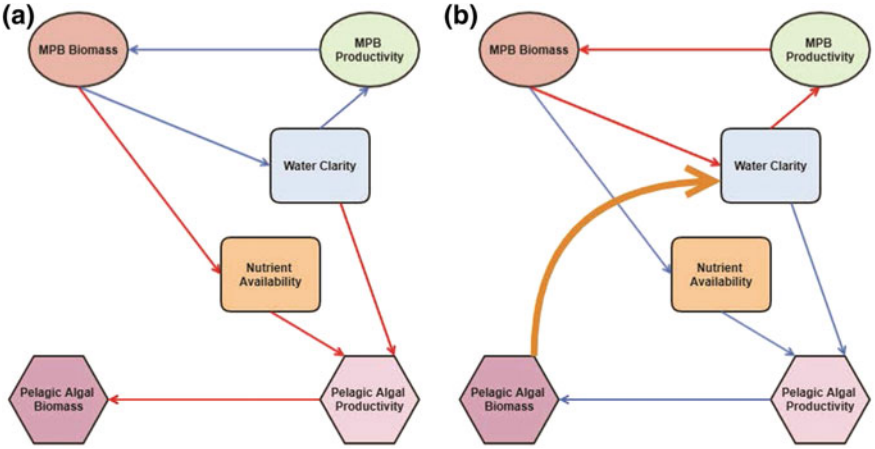


Fig. 2 **a** A benthic dominated system, which is characterized by high benthic biomass and productivity which increases water clarity and reduces nutrient availability in to the pelagic system; **b** A pelagic dominated system-resuspension decreases benthic biomass and productivity and water clarity and increases nutrient availability which in turn results in increased pelagic production. Red arrows indicate negative effect and blue arrows indicate positive effect (adapted from [17])

fish population thrived and a strong top-down effect was seen, thus establishing the importance of benthic-pelagic coupling in intensifying trophic interactions. **MPB in Organism Movement**: Resuspension to the water column allows MPB to be a part of the pelagic system and when it settles back it again participates in the benthic food web. Thus, it is quite evident that MPB plays a role in BPC through all the three mechanisms of BPC. However, modelling studies focusing on its role in BPC has been negligible.

5.1 Effect of Anthropogenic Stress on BPC and MPB

Table 2 summarises the effects of such stressors on different processes that comprises the benthic-pelagic coupling linkages as well as the probable effects on MPB dynamics.

6 Microphytobenthos Dynamics (Modelling Approach)

(1) Pinckney and Zingmark [61]: Microalgae migrate up and down in the sediment, moving in and out of the sediment photic zone which is responsible for the regular periodicities in production. This simulation model incorporates three important factors, viz. (1) in-situ irradiance at sediment surface (taking into account all

Table 2 Effects of anthropogenic stressors on mechanisms of BPC and MPB dynamics

Stressors	Benthic pelagic coupling mechanisms			MPB dynamics
	Organism movement	*Trophic interaction*	*Biogeochemical cycling*	
Nutrient overload	Increased sinking of pelagic organisms [50]; less resuspension of benthic biomass [30]	Shift from benthic production to pelagic production [51]	Increased flow of pelagic organic matter to sediments [52]	Productivity and biomass decreases due to reduced light availability because of high pelagic production
Invasive species	Increased bioturbation [53]; change in abundance of migratory organisms [36]	Increased benthic production results in increased pelagic production via trophic cascade [54]	Increased bioturbation results in release of nutrients from sediment [55] and decreased water clarity	Effect is species-specific; depending on the feeding habit of species introduced
Overfishing	Abundance and behaviour of migratory organisms altered [56, 57]; resuspension of benthic biomass depends on type of fishing gear used [30]	Predators removed [58]; if the predator targets benthic organisms then abundance of benthic organisms increase and if pelagic organisms are targets then abundance of pelagic organisms increase [30]	Filter feeding organisms decrease in case of bottom trawling [30]	In case the fish species removed feed on MPB then MPB biomass increases; otherwise it is reduced
Climate change	Ice cover loss may increase wave induced bottom shear [30]	Drastic climate change can result in changes in food web structure [59]	Vertical flux of carbon decreases [60]	May affect community structure of MPB resulting in changes in productivity

the factors that affect its quality and quantity), (2) biomass specific production and (3) vertical migration of the microalgae. The drawback of this model is that secondary factors affecting production rates such as desiccation at sediment water interface, high pH values, carbon limitation, photo-inhibition and so on, seasonal fluctuation of photosynthetic biomass, and effect of tidal resuspended benthic algae, grazing, self-shading are not taken into account.

(2) (a) Guarini et al. [19]: Vertical migratory behaviour of the microalgae causes the biomass of microalgae present at the surface of the sediment during day

time emersion period and immersion period to vary greatly which results in a difference in productivity. To represent this difference in the biomass in two different layers of the sediment at two different times (emersion and immersion), a deterministic two-compartment model representing the simultaneous evolution of S-chlorophyll-a concentration (indicating biomass) in the biofilm at the sediment surface and F-chlorophyll-a concentration in the underlying first centimetre of mud was formulated. The drawback of this model is that two of the major regulatory processes, loss and production are considered as linear and their spatial variability is not accounted for. Also, the effect of light and temperature is not considered.

(b) Guarini et al. [62]: This updated model takes the effect of light and temperature into consideration and compartments comprise of the illuminated photic layer S (which is equivalent to the biofilm of the previous compartment) and the aphotic layer B where sunlight does not penetrate. The equations are modified from the earlier model only in the term of productivity p which now includes the effects of light and temperature and is denoted by pB.

(c) Guarini et al. [63]: When the migration period (that is the total amount of time spent by microalgae at the surface of the sediment) exceeds the duration of the emersion period, the biofilm is resuspended at the beginning of the flooding tide. The main objective of this model was to include a critical time period and critical biomass, beyond which the biofilm is resuspended.

In this model, the daytime emersion periods are divided into productive period Ep and the period when the biofilm disappears En; night emersion period N and immersion period I were also separated. Potential duration of the biofilm T_p is a function of biomass and the critical biomass B_c above which the biofilm is exported to the water column is a function of TE, the duration of the day time emersion. When, T_p is more than TE, that is cell renewal is not completed and the biofilm is still at the surface during the emersion period, S is resuspended or exported to the water column.

The model only takes into account resuspension due to tidal effects and fails to account for resuspension occurring due to bioturbation and sloppy feeding by grazers. Also, similar to the very first model by these authors, the global mortality rates are considered to be linear.

(3) Ser$\hat{a}$dio and Catarino [64]: It is a model similar to the first model of Guarini et al. (2000b), the only difference being that dark level chlorophyll-a fluorescence is measured.

(4) Hochard et al. [14]: The earlier MPB models developed for estimating production in the intertidal zone [63] did not take into account the nitrogen cycle or the diagenetic processes occurring in the sediment which effect the MPB dynamics considerably [65]. This model is derived from the OMEXDIA model [66] and includes four extra variables depicting MPB carbon concentration, MPB nitrogen concentration, MPB chlorophyll a concentration and EPS. Another major

difference of this model with the OMEXDIA model is the assumption that all ammonium released during oxic mineralization is directly transformed to nitrate was not retained, to account for the competition between nitrifying bacteria and MPB.

This model takes into account a change in concentration with time as well as with depth which is advantageous because impact of MPB on diagenesis largely resulted from changes in the spatial distribution of the compounds-MPB accumulation near the surface increases oxygen levels due to photosynthesis which directly impacts mineralization. One major drawback of this model is that, one of the most important factors affecting MPB dynamics—vertical migration—is not included.

7 Modelling Studies Focusing on Both MPB and BPC

(1) Brito et al. [67]: Preliminary results of the model which couples benthic and pelagic compartments of the shallow coastal lagoon—Ria Formosa indicate presence of a large biomass of benthic microalgae, which is mainly dependent on the nutrients in pore water and by the phenomenon of resuspension strongly influences the pelagic chlorophyll concentration.

(2) Hochard et al. [68]: The benthic pelagic coupled model of southwest lagoon of New Caledonia clearly showed that sediment erosion and wave bottom currents driven pore water advection causes significant disturbances to the system that relaxes the control of MPB on nutrient fluxes and results in liberation of nutrients to the water column, along with increased inputs of suspended sediment organic matter (OM) and MPB.

(3) Lindegren et al. [69]: The occurrence of regime shifts and how multiple drivers such as climate change, eutrophication and so on affect the regime shifts in a eutrophied and heavily exploited marine system Kattegat were tested by the authors by studying the state and regulatory pathways of the system. When a large part of the primary production sinks to the bottom and benthic individuals are favoured more than pelagic consumers, a regime shift from pelagic to benthic system occured.

(4) Rakotomalala et al. [70]: The common cockle *Cerastoderma edule* regulates the MPB erosion flux in the water column in Baie des Veys (Lower Normandy, France) as it is a major bioturbator in terms of biomass. The authors of this study developed a numerical model that reproduces the export of MPB associated with the biogenic layer erosion considers two compartments—the water column and the sediment where cockles are present.

Recent modelling studies on BPC [71, 72] have focused on the nutrient dynamics and physical parameters and not on the biological aspects.

8 Difficulties of Conducting Studies on Benthic-Pelagic Coupling

– Coupled links integrate different depths within the water column and may also do the same on a smaller scale over depths within sediments. However, most of the surveillance techniques and experimental methodologies available usually focus on the sediment-water interface. Thus proper quantification of the linkages between two compartments is a huge undertaking [73].
– Moreover, because of the size constraints of the recording devices of most surveillance and experimental approaches, the materials are observed only when they fall within a very specific and restricted size range.
– Simple knowledge about the amount of material leaving or arriving at the sediment-water interface, without following the subsequent fate of the organisms as to their spatial distribution provides only a limited view of the benthic-pelagic coupling.
– Biological components of both benthic and pelagic systems have patchy distribution, along both vertical and horizontal axes which poses many difficulties in sampling.
– Relating emergence and settlement processes of macrofaunal organisms to sediment physico-chemical and microbiological features is also potentially complicated by the scale at which those features are routinely measured.

9 Future Prospects

Even though MPB performs many functions—from maintaining the resilience of the ecosystem to increasing its productivity and preventing nutrient loss from the system (through benthic pelagic coupling and its role in nutrient cycling respectively)—its importance in the studied ecosystem models are understated. There are many studies focusing on the functional role of MPB, however, studies of aquatic food web dynamics especially those in the marine systems have mostly focused on pelagic interactions of phytoplankton production and consumption by herbivorous grazers while completely ignoring the fate of detritus that settle to bottom and the role of MPB in recycling the nutrients obtained from these organic matters to the pelagic system.

Whole system can only be successfully studied if benthic and pelagic systems are considered in unison. As discussed in Sect. 3, BPC is what connects the two systems. In Sect. 4, the importance of MPB over other benthic components in BPC has been highlighted. The main aim of this work was to study the available modelling works focusing on MPB and its role in BPC. While there are many studies focusing on MPB, they mainly focus on MPB productivity and very few have been done on its role in BPC. Studies focusing on BPC mainly consider the biogeochemical aspect and have not been considered here. The few studies mentioned in Sect. 6, are the only ones considering some other mechanisms as well but still they consider only MPB

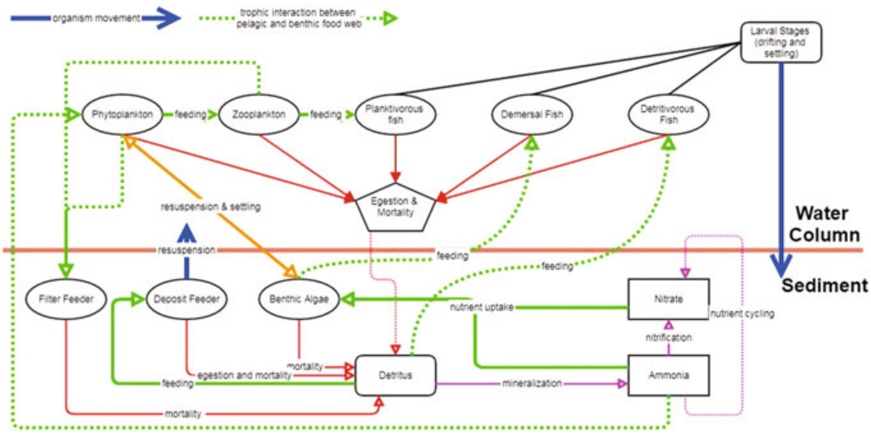


Fig. 3 A conceptual model developed to study the whole aquatic ecosystem of an estuary

and not other benthic compartments. Whole system studies achieved through static modelling approaches can be used for conducting such in-depth studies.

Static modelling approaches using ecological network analyses have been applied in various situations. For example, recent applications include study of the trophic structure and determination of robustness of reservoir ecosystem [74, 75]. A study by Rakshit et al. [76] using static models for temporal comparison of trophic structure of the Hooghly-Matla estuarine system considers a few benthic compartments as well. However, such an approach has not yet been applied to study the aquatic system as a whole with equal focus on both pelagic and benthic systems.

A conceptual model integrating both benthic and pelagic food webs has been developed as shown in Fig. 3.

Development of an MPB based food web model using the above conceptual diagram and studying it using ENA could yield valuable information about the significance and role of the MPB coupled links in different aquatic ecosystems.

Acknowledgements Author Swagata Sinha acknowledges Council of Scientific and Industrial Research, India (CSIR) F.No. 09/202(0080)/2018-EMR-I for financial assistance. Author Arnab Banerjee acknowledges the University Grants Commission, Govt. of India, for providing Dr. D. S. Kothari Post Doctoral fellowship, F No: BL/17–18/0490.

References

1. Pinckney, J., Zingmark, R.: Biomass and production of benthic microalgal communities in Estuarine habitats. Estuaries **16**(4), 887897 (1993)
2. De Jonge, V.N., Van Beusekom, J.E.E.: Wind-and tide-induced resuspension of sediment and microphytobenthos from tidal flats in the Ems Estuary. Oceanogr. Lit. Rev. **3**(43), 250 (1996)

3. Montagna, P.A., Blanchard, G.F., Dinet, A.: Effect of production and biomass of intertidal microphytobenthos on meiofaunal grazing rates. J. Exp. Mar. Bio. Ecol. **185**(2), 149165 (1995)
4. Gillespie, P.A., Maxwell, P.D., Rhodes, L.L.: Microphytobenthic communities of subtidal locations in New Zealand: taxonomy, biomass, production, and food-web implications. New Zeal. J. Mar. Freshw. Res. **34**(1), 4153 (2000)
5. Dodds, W.K., Gudder, D.A.: The ecology of Cladophora. J. Phycol. **28**(4), 415427 (1992)
6. Burkholder, J.M., Wetzel, R.G., Klomparens, K.L.: Direct comparison of phosphate uptake by adnate and loosely attached microalgae within an intact biofilm matrix. Appl. Environ. Microbiol. **56**(9), 28822890 (1990)
7. Lock, M.A., Wallace, R.R., Costerton, J.W., Ventullo, R.M., Charlton, S.E.: River epilithon: toward a structural-functional model, Oikos **42**, 10–22 (1984)
8. Holland, A.F., Zingmark, R.G., Dean, J.M.: Quantitative evidence concerning the stabilization of sediments by marine benthic diatoms. Mar. Biol. **27**(3), 191196 (1974)
9. Nils, R.P.: Coupled nitrification-denitrification in autotrophic and heterotrophic estuarine sediments: on the influence of benthic microalgae. Limnol. Oceanogr. **48**(1), 93105 (2003)
10. Middelburg, J.J., Barranguet, C., Boschker, H.T.S., Herman, P.M.J., Moens, T., Heip, C.H.R.: The fate of intertidal microphytobenthos carbon: an in situ 13C-labeling study. Limnol. Oceanogr. **45**(6), 12241234 (2000)
11. Christensen, P.B., Glud, R.N., Dalsgaard, T., Gillespie, P.: Impacts of longline mussel farming on oxygen and nitrogen dynamics and biological communities of coastal sediments. Aquaculture **218**(1), 567588 (2003)
12. Dale, A.W., Prego, R.: Physico-biogeochemical controls on benthic-pelagic coupling of nutrient fluxes and recycling in a coastal upwelling system. Mar. Ecol. Prog. Ser. **235**(2), 1528 (2002)
13. Sundb$\breve{a}$ck, K., Miles, A., Linares, F.: Nitrogen dynamics in nontidal littoral sediments: role of microphytobenthos and denitrification. Estuar. coasts **29**(6), 11961211 (2006)
14. Hochard, S., Pinazo, C., Grenz, C., Burton, J.L., Pringault, O., Evans, J.L.B., Pringault, O.: Impact of microphytobenthos on the sediment biogeochemical cycles: a modeling approach. Ecol. Modell. **221**(13), 16871701 (2010)
15. Tobias, C., Giblin, A., McClelland, J., Tucker, J., Peterson, B.: Sediment DIN fluxes and preferential recycling of benthic microalgal nitrogen in a shallow macrotidal estuary. Mar. Ecol. Prog. Ser. **257**, 2536 (2003)
16. Kang, C.K., Kim, J.B., Lee, K.S., Bin Kim, J., Lee, P.Y., Hong, J.S.: Trophic importance of benthic microalgae to macrozoobenthos in coastal bay systems in Korea: dual stable C and N isotope analyses. Mar. Ecol. Prog. Ser. **259**, 7992 (2003)
17. MacIntyre, H.L., Lomas, M.W., Cornwell, J., Suggett, D.J., Gobler, C.J., Koch, E.W., Kana, T.M.: Mediation of benthic pelagic coupling by microphytobenthos: an energy- and material-based model for initiation of blooms of Aureococcus anophagefferens. Harmful Algae **3**, 403437 (2004)
18. Nielsen, S.L., Risgaard-Petersen, N., Banta, G.T.: Nitrogen retention in coastal marine sediments—a field study of the relative importance of biological and physical removal in a Danish Estuary. Estuar. Coasts **40**, 12761287 (2017)
19. Guarini, J.M., Blanchard, G.F., Gros, P., Gouleau, D., Bacher, C.: Dynamic model of the short-term variability of microphytobenthic biomass on temperate intertidal mudflats **195**, 291303 (2000)
20. J. Ser$\hat{a}$dio, J. da Silva, and F. Catarino, Nondestructive tracing of migratory rhythms of intertidal benthic microalgae using in vivo chlorophyll a fluorescence. J. Phycol. **33**(3), 542–553 (1997)
21. Cartaxana, P., Cruz, S., Gameiro, C., Khl, M.: Regulation of intertidal microphytobenthos photosynthesis over a diel emersion period is strongly affected by diatom migration patterns. Front. Microbiol. **7**, 111 (2016)
22. Arfi, R., Guiral, D., Bouvy, M.: Wind Induced resuspension in a shallow tropical lagoon. Estuar. Coast. Shelf Sci. **36**, 587604 (1993)
23. Blanchard, G.F., Cariou-Le Gall, V.: Photosynthetic characteristics of microphytobenthos. J. Exp. Mar. Bio. Ecol. **182**, 114 (1994)

24. Pinckney, J.L., Zingmark, R.G.: Effects of tidal stage and sun angles on intertidal benthic microalgal productivity. Mar. Ecol. Prog. Ser. **76**, 8189 (1991)
25. Blanchard, G., Guarini, J.: Studying the role of mud temperature on the hourly variation of the photosynthetic capacity of microphytobenthos in intertidal areas. Oceanogr. Lit. Rev. **6**(44), 612 (1997)
26. Aberle-Malzahn, N., Wiltshire, K.H., Brendelberger, H.: The microphytobenthos and its role in aquatic food webs. Christian-Albrechts-Universitt Kiel (2004)
27. McCormick, P.V., Stevenson, R.J.: Grazer control of nutrient availability in the periphyton. Oecologia **86**(2), 287291 (1991)
28. Yin, K., Zetsche, E.M., Harrison, P.J.: Effects of sandy versus muddy sediments on the vertical distribution of microphytobenthos in intertidal flats of the Fraser River Estuary. Canada. Environ. Sci. Pollut. Res. **23**(14), 1419614209 (2016)
29. Vadeboncoeur, Y., Jeppesen, E., Zanden, M., Schierup, H.H., Christoffersen, K., Lodge, D.M.: From greenland to green lakes: cultural eutrophication and the loss of benthic pathways in lakes. Limnol. Oceanogr. **48**(4), 14081418 (2003)
30. Griffiths, J.R., Kadin, M., Nascimento, F.J.A., Tamelander, T., Trnroos, A., Bonaglia, S., Bonsdorff, E., Brchert, V., Grdmark, A., Jrnstrm, M., Kotta, J., Lindegren, M., Nordstrm, M.C., Norkko, A., Olsson, J., Weigel, B., ydelis, R., Blenckner, T., Niiranen, S., Winder M.: The importance of benthic-pelagic coupling for marine ecosystem functioning in a changing world. Glob. Chang. Biol., 21792196 (2017)
31. Seitz, R.D., Wennhage, H., Bergstrm, U., Lipcius, R.N., Ysebaert, T.: Ecological value of coastal habitats for commercially and ecologically important species. ICES J. Mar. Sci. **71**(3), 648665 (2013)
32. Chauvaud, L., Jean, F., Ragueneau, O., Thouzeau, G.: Long-term variation of the Bay of Brest ecosystem: benthic-pelagic coupling revisited. Mar. Ecol. Prog. Ser. **200**, 3548 (2000)
33. Boynton, W.R., Kemp, W.M.: Nutrient regeneration and oxygen consumption by sediments along an estuarine salinity gradient. Mar. Ecol. Prog. Ser. **23**, 4555 (1985)
34. Tilstone, G.H., Miguez, B.M., Figueiras, F.G., Fermin, E.G.: Diatom dynamics in a coastal ecosystem affected by upwelling: coupling between species succession, circulation and biogeochemical processes. Mar. Ecol. Prog. Ser. **205**, 2341 (2000)
35. Chatterjee, A., Klein, C., Naegelen, A., Claquin, P., Masson, A., Legoff, M., Amice, E., LHelguen, S., Chauvaud, L., Leynaert, A.: Comparative dynamics of pelagic and benthic micro-algae in a coastal ecosystem. Estuar. Coast. Shelf Sci. **133**, 6777 (2013)
36. Baustian, M.M., Hansen, G.J.A., de Kluijver, A., Robinson, K., Henry, E.N., Knoll, L.B., Rose, K.C., Carey, C.C., Sciences, W., Lane, F., Lansing, E.: Linking the bottom to the top in aquatic ecosystems: mechanisms and stressors of benthic-pelagic coupling. In: Eco-DAS X Symposium Proceedings, pp. 2547 (2014)
37. Vanni, M.J.: Nutrient cycling by animals in freshwater ecosystems. Annu. Rev. Ecol. Syst. **33**(1), 341370 (2013)
38. Pitt, K.A., Clement, A.L., Connolly, R.M., Thibault-Botha, D.: Predation by jellyfish on large and emergent zooplankton: implications for benthicpelagic coupling. Estuar. Coast. Shelf Sci. **76**(4), 827833 (2008)
39. Pasquaud, S., Pillet, M., David, V., Sautour, B., Elie, P.: Determination of fish trophic levels in an estuarine system. Estuar. Coast. Shelf Sci. **86**(2), 237246 (2010)
40. Lonsdale, D., Cerrato, R., Holland, R., Mass, A., Holt, L., Schaffner, R., Pan, J., Caron, D.: Influence of suspension-feeding bivalves on the pelagic food webs of shallow, coastal embayments. Aquat. Biol. **6**, 263279 (2009)
41. Lassalle, G., Lobry, J., Le Loch, F., Bustamante, P., Certain, G., Delmas, D., Dupuy, C., Hily, C., Labry, C., Le Pape, O., Marquis, E., Petitgas, P., Pusineri, C., Ridoux, V., Spitz, J., Niquil, N.: Lower trophic levels and detrital biomass control the Bay of Biscay continental shelf food web: implications for ecosystem management. Prog. Oceanogr. **91**(4), 561575 (2011)
42. Arnott, D.L., Vanni, M.J.: Nitrogen and phosphorus recycling by the zebra mussel (Dreissena polymorpha) in the western basin of Lake Erie. Can. J. Fish. Aquat. Sci. **53**(3), 646659 (1996)

43. Karlsson, J., Säwström, C.: Benthic algae support zooplankton growth during winter in a clear-water lake. Oikos **118**(4), 539544 (2009)
44. Staehr, P.A., Baastrup-Spohr, L., Sand-Jensen, K., Stedmon, C.: Lake metabolism scales with lake morphometry and catchment conditions. Aquat. Sci. **74**(1), 155169 (2012)
45. Gyllström, M., Lakowitz, T., Brönmark, C., Hansson, L.A.: Bioturbation as driver of zooplankton recruitment, biodiversity and community composition in aquatic ecosystems. Ecosystems **11**(7), 11201132 (2008)
46. DAndrea, A.F., DeWitt, T.H.: Geochemical ecosystem engineering by the mud shrimp Upogebia pugettensis (Crustacea: Thalassinidae) in Yaquina Bay, Oregon: density-dependent effects on organic matter remineralization and nutrient cycling. Limnol. Oceanogr. **54**(6), 19111932 (2009)
47. Welsh, D.T.: Its a dirty job but someone has to do it: The role of marine benthic macrofauna in organic matter turnover and nutrient recycling to the water column. Chem. Ecol. **19**(5), 321342 (2003)
48. Vadeboncoeur, Y., Vander Zanden, M.J., Lodge, D.M.: Putting the lake back together: reintegrating benthie pathways into lake food web models. Bioscience **52**(1), 4454 (2002)
49. Vander Zanden, M.J., Essington, T.E., Vadeboncoeur, Y.: Is pelagic top-down control in lakes augmented by benthic energy pathways?, Can. J. Fish. Aquat. Sci. **62**(6), 14221431 (2005)
50. Heip, C.H.R., Goosen, N.K., Herman, P.M.J., Kromkamp, J.C., Middelburg, J.J., Soetaert, K.E.R.: Production and consumption of biological particles in temperate tidal estuaries. Oceanogr. Mar. Biol. Annu. Rev. **33**, 1149 (1995)
51. Turner, R.E.: Some effects of eutrophication on pelagic and demersal marine food webs. Coast. Hypoxia Conseq. Living Resour. Ecosyst., 371398 (2001)
52. Chandra, S., Jake Vander Zanden, M., Heyvaert, A.C., Richards, B.C., Allen, B.C., Goldman, C.R.: The effects of cultural eutrophication on the coupling between pelagic primary producers and benthic consumers. Limnol. Oceanogr. **50**(5), 13681376 (2005)
53. Weber, M.J., Brown, M.L.: Effects of common carp on aquatic ecosystems 80 years after carp as a dominant: ecological insights for fisheries management. Rev. Fish. Sci. **17**(4), 524537 (2009)
54. Roohi, A., Yasin, Z., Kideys, A.E., Hwai, A.T.S., Khanari, A.G., Eker-Develi, E.: Impact of a new invasive ctenophore (Mnemiopsis leidyi) on the zooplankton community of the Southern Caspian Sea. Mar. Ecol. **29**(4), 421434 (2008)
55. Matsuzaki, S.S., Usio, N., Takamura, N., Washitani, I.: Effects of common carp on nutrient dynamics and littoral community composition: roles of excretion and bioturbation. Fundam. Appl. Limnol. **168**(1), 2738 (2007)
56. Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., others: Rebuilding global fisheries. Science **325**(5940), 578585 (2009)
57. Madin, E.M.P., Gaines, S.D., Warner, R.R.: Field evidence for pervasive indirect effects of fishing on prey foraging behavior. Ecology **91**(12), 35633571 (2010)
58. Estes, J.A., Duggins, D.O.: Sea otters and kelp forests in Alaska: generality and variation in a community ecological paradigm. Ecol. Monogr. **65**(1), 75100 (1995)
59. Edwards, M., Richardson, A.J.: Impact of climate change on marine pelagic phenology and trophic mismatch. Nature **430**(7002), 881884 (2004)
60. Nixon, S.W., Fulweiler, R.W., Buckley, B.A., Granger, S.L., Nowicki, B.L., Henry, K.M.: The impact of changing climate on phenology, productivity, and benthic-pelagic coupling in Narragansett Bay. Estuar. Coast. Shelf Sci. **82**(1), 118 (2009)
61. Pinckney, J.L., Zingmark, R.G.: Modeling the annual production of intertidal benthic microalgae in estuarine ecosystems. J. Phycol. **29**(4), 396407 (1993)
62. Guarini, J.M., Blanchard, G.F., Gros, P.: Quantification of the microphytobenthic primary production in european intertidal mudflats—a modelling approach. Cont. Shelf Res. **20**(1213), 17711788 (2000)
63. Guarini, J.M., Sari, N., Moritz, C.: Modelling the dynamics of the microalgal biomass in semi-enclosed shallow-water ecosystems. Ecol. Modell. **211**(34), 267278 (2008)

64. Serôdio, J., Catarino, F.: Modelling the primary productivity of intertidal microphytobenthos: time scales of variability and effects of migratory rhythms. Mar. Ecol. Prog. Ser. **192**, 1330 (2000)
65. Blackford, J.C.: The Influence of microphytobenthos on the Northern Adriatic ecosystem: a modelling study. Estuar. Coast. Shelf Sci. **55**(1), 109123 (2002)
66. Soetaert, K., Herman, P.M.J., Middelburg, J.J.: A model of early diagenetic processes from the shelf to abyssal depths. Geochim. Cosmochim. Acta **60**(6), 10191040 (1996)
67. Brito, A.C., Newton, A., Fernandes, T.F., Tett, P.: The role of microphytobenthos on shallow coastal lagoons: a modelling approach **106**, 207228 (2011)
68. Hochard, S., Pinazo, C., Rochelle-newall, E., Pringault, O.: Benthic pelagic coupling in a shallow oligotrophic ecosystem: importance of microphytobenthos and physical forcing. Ecol. Modell. **247**, 307318 (2012)
69. Lindegren, M., Blenckner, T., Stenseth, N.C.C., Jolla, L., Castle, C., Centre, S.R.: Nutrient reduction and climate change cause a potential shift from pelagic to benthic pathways in a eutrophic marine ecosystem. Glob. Chang. Biol. **18**(12), 34913503 (2012)
70. Rakotomalala, C., Granger, K., Ubertini, M., Fort, M., Orvain, F.: Modelling the effect of Cerastoderma edule bioturbation on microphytobenthos resuspension towards the planktonic food web of estuarine ecosystem. Ecol. Modell. **316**, 155167 (2015)
71. Provoost, P., Braeckman, U., Van Gansbeke, D., Moodley, L., Soetaert, K., Middelburg, J.J., Vanaverbeke, J.: Estuarine, coastal and shelf science modelling benthic oxygen consumption and benthic-pelagic coupling at a shallow station in the southern North Sea. Estuar. Coast. Shelf Sci. **120**, 111 (2013)
72. Seidel, M., Beck, M., Greskowiak, J., Riedel, T., Waska, H., Suryaputra, I.G.N.A., Schnetger, B., Niggemann, J., Simon, M., Dittmar, T.: Benthic-pelagic coupling of nutrients and dissolved organic matter composition in an intertidal sandy beach. Mar. Chem. **176**, 150163 (2015)
73. Raffaelli, D., Bell, E., Weithoff, G., Matsumoto, A., Cruz-Motta, J.J., Kershaw, P., Parker, R., Parry, D., Jones, M.: The ups and downs of benthic ecology: considerations of scale, heterogeneity and surveillance for benthicpelagic coupling. J. Exp. Mar. Bio. Ecol. **285286**, 191203 (2003)
74. Banerjee, A., Banerjee, M., Mukherjee, J., Rakshit, N., Ray, S.: Trophic relationships and ecosystem functioning of Bakreswar Reservoir. India. Ecol. Inform. **36**, 5060 (2016)
75. Banerjee, A., Chakrabarty, M., Rakshit, N., Mukherjee, J., Ray, S.: Indicators and assessment of ecosystem health of Bakreswar reservoir, India: an approach through network analysis. Ecol. Indic. **80**(May), 163173 (2017)
76. Rakshit, N., Banerjee, A., Mukherjee, J., Chakrabarty, M., Borrett, S.R., Ray, S.: Comparative study of food webs from two different time periods of Hooghly Matla estuarine system, India through network analysis. Ecol. Modell. **356**, 2537 (2017)

Overview of Ecological Economics and Ecosystem Services Consequences from Shrimp Culture

Suvendu Das, Sagar Adhurya and Santanu Ray

Abstract Recent trends of aquaculture, as well as mariculture, go towards more profits in minimum time investment. Every growing country is following this trend. Shrimp culture (both freshwater and saline water) is one of the best-suited cultures in relation to all of these demands. There is a direct relationship between the ecosystem and Shrimp culture. The particular ecosystem provides suitable factors which makes Shrimp culture successful. Every ecosystem has its own uniqueness which causes it to be different from others. Human society derives their benefits from this natural unique ecosystem and these benefits are considered as the part of ecosystem services. Ecosystem services can be calculated in terms of ecological economic analysis. Ecological economics and its application is the major pathway to understand the relationship between unique ecosystem components and ecosystem services. In Shrimp culture, especially in case of brackish water culture, the farmers add lots of external substances in water. These elements cause environmental damages and affect ecosystem health. Eco-economic study of ecosystem services can enlighten the conflicts between economic benefits and ecosystem loss. The decision makers can get an overall idea about how the optimization of ecosystem services could be done in case of Shrimp culture.

Keywords Aquaculture · Mari-culture · Ecosystem health · Environmental damage · Ecosystem components · Eco-economic study

S. Das · S. Adhurya · S. Ray (✉)
Systems Ecology & Ecological modeling Laboratory Department of Zoology,
Visva-Bharati University, Santiniketan 731235, India
e-mail: santanu.ray@visva-bharati.ac.in

S. Das
e-mail: suvendudas.rs@visva-bharati.ac.in

S. Adhurya
e-mail: sagaradhurya.rs@visva-bharati.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_20

1 Introduction

Aquaculture has emerged rapidly in last few decades. The aquaculture is transforming into future food supplier to human overgrowing population and it causes socioeconomic limitations, affecting ecosystem from local to global situations [1]. Form different purposes of human society are being benefited by marine and aquatic ecosystem. The ecosystem approach to aquaculture (EAA) turns primary concern of researchers and the concept EAA is not fully understood though EAA defines an interdisciplinary perspective toward interruption of the ecosystem by aquaculture [2]. Best management practice (BMP) for good aquaculture practice (GAP) is important to "reduce key impact" and BMP helps to increase producers efficiency, reduce waste and improve income [1]. Shrimp culture is "high value" product of mariculture. There are many species of shrimp but in worldwide Pacific White Shrimp (*Litopaneus monodon*) and Black Tiger Shrimp (*Penaeus monodon*) are major cultivation species. More or less same process is followed by the entire world for the shrimp culture [3]. Marine shrimp species generally are cultivated in water bodies of coastal region. Trend of marine shrimp culture is arising inland by creating an artificially ideal environment for shrimp culture. Most intensive practice for economic growth with increasing demand in world market is befalling in marine shrimp culture. Ecological and environmental manipulation is a maximum outcome of this culture. Service provided by ecosystem and ecosystem functioning is changing from the perspective of value, drastically with some magnitudes. Eco-economic valuation of natural capital is derelict. Time has come to make some changes to establish some sustainable development goals to diminish the conflicts between mariculture and sustainable ecosystem and good decision making [4].

2 Mari-Culture as New Form of Aquaculture

Mariculture is a branch of aquaculture where cultivation of fish and others marine organisms are done for human consumption. Major mariculture organisms are different marine fish, oysters, shrimps, marine algae etc. Environmental resources play main function in culture of marine organisms and the supply of the resources should be in accurate for each and every culture and cultured species [5]. From 1984, production of food by aquaculture is growing steadily near about 10% per year in comparison to live stock meat and capture fisheries and expansion of aquaculture happened in the year of 1980 [6]. There are lots of different techniques for the mariculture, evolving day by day. Among these techniques, cage culture and closed system culture are important ones. In some countries (i.e. France), closed system culture is superior over widely accepted cage farming. Closed system culture is more eco-friendly and follows more conventional hatchery operation [7]. Mariculture with offshore cage culture system first grounded its foot in the 1950s at Japan and from last few decades mostly salmon farming in North America and northern Europe is occurred by cage farming [7].

3 Shrimp Culture as a Trend of Recent Aquaculture as well as Mariculture

Every coastal country's recent trend of mariculture goes towards shrimp harvesting industries and a large part of their coastal economy is dependent on foreign export of shrimp. The boost up of shrimp production has been gradually raised from 1975 and before 1975; it was accounted that 2% world market occupied by shrimp [8].

Global Status of Shrimp Culture

R. R Stickney reported in Encyclopedia of Aquaculture, 2000 that "shrimp culture increased 300% from 1975 to 1985, 250% from 1985 to 1995, and if it increases 200% between 1995 and year 2005, would shrimp culture production will be at 2.1 million metric tons (hereafter abbreviated as MTD 1.1 standard tons D 2,204.6 lb, or 1,000 kg). Present world shrimp culture production is 737 thousand MT annually, or 24.5% of the 3 million MT would market for shrimp." From decades shrimp is one of the most-traded product and now it is second in value term among the other aquaculture products (Fig. 1). Different countries from Asia lead the World shrimp production. Shrimp export contributes in economic growth but now shrimp producing countries face the challenges of decline in production by disease outbreak and global price of shrimp is falling [9]. Over production of shrimp is the main cause of decreasing price of shrimp in Global market. From FAO report of The State of World Fisheries and Aquaculture 2016, the scenario of increasing production of crustacean (shrimp) has been cleared—"annual per capita availability of crustaceans grew substantially from 0.4 kg in 1961 to 1.8 kg in 2013". Shrimps and others aquaculture production contributes employment to rural poor people by labour inputs, seed and feed collection [10]. Shrimp, Prawn and lobster cultivation is followed by different countries which are environmentally equipped by proper ecological condition for cultivation of these crustaceans. Recently production of shrimp has shown alteration of ecosystem services and worldwide shrimp culture faces the challenges to sustainable development. Sustainability is necessary to maintain the balance between ecosystem health and economic growth.

4 Ecosystem Services and Ecological Economics

Ecosystem services are ecological features or functions that have direct impacts on human society. The benefits are driven by human from features and functioning of ecosystem is known as ecosystem services [11]. Different factors such as economic growth, politics, religious activities, culture and lots of biophysical factors regulate

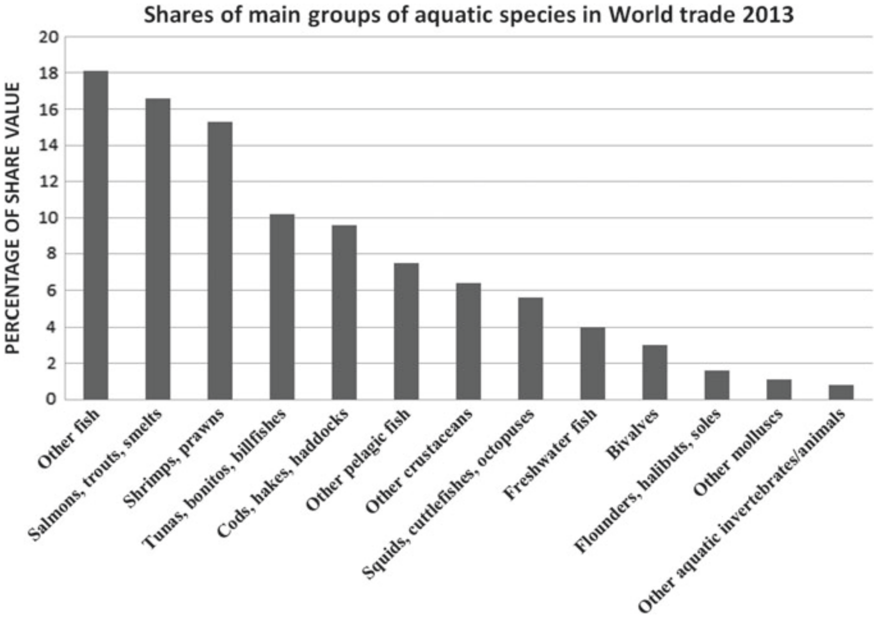


Fig. 1 Shares of the main group of species. Here it shows among the most trading products of aquaculture, Shrimps and prawns are in the second position according to their share value percentage neglecting the "other fish" group (which considers all other marine and freshwater species). Adapted from FAO [9]

the ecosystem services of a particular ecosystem [12]. The major factor for changing and losing native ecosystem services is anthropogenic stress. With relation to increasing human population size and economic growth over last few decades, near about 60% of ecosystem services had been disrupted or lost [13].

4.1 Forms of Ecosystem Services

Ecosystem which gives services to human society is also referred as 'natural capital' and the term 'capital' is connecting ecosystem services with human economy [14]. In the year 1997, Robert Costanza and other co-workers listed different ecosystem services as functions for a particular study. These are as follow: (1) Gas regulation (e.g.—CO_2/O_2 regulation, SOx level) 2. Climate regulation (e.g.—greenhouse gas regulation) 3. Disturbance regulation (e.g.—storm protection, flood control) 4. Water regulation (e.g.—providing water for agricultural) 5. Water supply (e.g.—provisioning water by watersheds) 6. Erosion control and sediment retention (e.g.—prevent loss of soil by wind) 7. Soil formation (e.g.—weathering of rocks and accumulation of organic materials) 8. Nutrient cycling (e.g.—nitrogen, phosphorus and other elements cycle) 9. Waste treatment (e.g.—waste treatment, pollution control) 10. Pollination (e.g.—provide pollinator for reproduction) 11. Biological control (e.g.—predators, parasitoid control herbivory) 12. Refugia (e.g.—habitat for migratory species) 13. Food production (e.g.—production of fish, crop) 14. Raw materials (e.g.—production of lumber) 15. Genetic resources (e.g.—medicines, gene for resistance against pathogen of organism) 16. Recreation (e.g.—eco-tourism) 17. Cultural (e.g.—spiritual, cultural support) [15].

4.2 Valuation of Services Provided by the Ecosystem

The crucial part of ecosystem service is to define value of a natural capital. The ecosystem service has a holistic approach toward evaluation of ecosystem structure, function and processes for proper valuation. The value includes whole marketed and non-marketed mechanism to detect the value of a particular ecosystem and its services [16]. There are lots of doubts. The basic point of views behind the doubt has completely contrast outcomes. One of them considers nonanthropogenic approach to derive the value and another one considers anthropocentric, economic approach to derive value of ecosystem services [17]. To detect the proper value of ecosystem services, the holistic approach has been taken which considers anthropocentric approach and all value of services are interpreted into economic valuation. Few misapprehensions are present in terms of economic valuation. There are lots of believes that mention economic valuation is just an assessment of the commercial value of anything. Actually economic valuation incorporates many components which have no commercial or market value [18]. In the book of "Valuing Ecosystem Services toward Better Environmental Decision-Making" reported to clear concept of economic valuation of any ecosystem service that is: "The instrumental value of an ecosystem service is a value derived from its role as a means toward an end other than itself. In other words, its value is derived from its usefulness in achieving a goal. In contrast, intrinsic value is the value that exists independently of any such contribution; it reflects the value of something for its own sake. For example, if a fish population provides a source of food for either humans or other species, it has instrumental value. This value stems from its contribution to the goal of sustaining the consuming population. If it continues to have value even then it were no longer "useful" to these populations (e.g., if an alternative, preferred food source were discovered), that remaining value would be its intrinsic value. For example, if the Grand Canyon and the Florida Everglades have intrinsic value, that component of value would be independent of whether humans directly or indirectly use them—either as sites for recreation, study, or even contemplation. Intrinsic value can also stem from heritage or cultural sources, such as the value of culturally important burial grounds. Because intrinsic value is the value of something unrelated to its instrumental use of any kind, it is often termed 'noninstrumental' value" [17].

4.3 Ecological Economics: Approach to Connect Natural Capitals with Ecosystem Services

Ecological economics is an interdisciplinary or transdisciplinary subject which relates social science with economics and economics with natural science. Ecological economics deals with quantification of ecosystem services. Human societal pattern now has been entrapped into economic growth. Economic growth creates different problems like unequal distribution of wealth, reduction of birth rate of human population, environmental pollution, ecosystem degradation and loss of natural capitals [16]. Measures of economic activity which is used by most nations like gross domestic product (GDP), mostly neglect depletion of natural capitals [19]. Natural capital is one of the main controlling factors of consumption of goods and individual utility. For a holistic overview of ecological economics natural capital is included along with others three capitals (built, social and human capitals) [20]. It has been cleared that in the absence of "human capital", "social capital" and "built capital" ecosystem never dispense any benefits [21]. The estimation of value of ecosystem services provided by any biome and its alteration by human manipulation can be estimated by quantifying benefits before and after human manifestation [22].

4.4 Ecological Footprint and Natural Capital

Ecological Footprint (EF) is one of the key indicators of natural capital. Ecological Footprint is defined as "area of productive land and water ecosystems required to produce the resources that the population consumes and assimilate the wastes that the population produces, wherever on Earth that land and water may be located" [23]. Robert Costanza denoted Ecological Footprint as, "the EF a useful provisional indicator of sustainability at the global scale, but it should be cast in these terms; as a technologically skeptical indicator" [24]. Though assessment of the theoretical concept of Ecological Footprint does not have any particular method but it is an important indicator of natural capitals.

5 Farmer's Manipulation in Shrimp Culture

In case of shrimp culture, particular ecosystem and particular environmental conditions are very important. The suitable ecosystem provides specific ecosystem services instead of that the increasing production of shrimp or any others product in a particular ecosystem could never be possible but the major question is alteration of ecosystem services.

- Different types of culture methods and harvesting procedures are followed worldwide for shrimp culture. Among these culture methods, semi-intensive culture is most popular aquaculture method. There are many types of semi-intensive cultural methods like continuous and batch culture [25].

- In continuous culture, regular stocking and selective harvesting are done and water is not drained out. It results that large remaining shrimp is predominant and lowering the growth rate of new stocks by competition.
- In batch culture, stocking is done in each pond once and after harvesting product in desire size entire water would be drained out. In contrast to the continuous culture, it avoids competition and predatory effect of remaining shrimps.
- In low, medium, high technology semi-intensive system the stocking density is maintained from low number to high number respectively. According to intensive culture methods, different parameters are set and water quality is maintained. The yearly cycle of production varies in tropical and subtropical countries and due to seasonal variation implementation of culture methods also differ.
- In culture methods, periodic or daily measurement of pH, DO_2, temperature, hardness, alkalinity, ammonia, nitrite, nitrate is important to maintain proper growth rate of shrimps. Weeds grows in the shallow water and shows r-selected growth [26]. These aquatic weeds act as shelter and food for shrimp. Farmers generally provide food two to three times daily.
- In open intensive and semi-intensive culture of brackish water shrimp, exchange of water, application of fertilizer and food is practiced [27].
- Graslund and Bengtsson (2001) reported in an enlisted manner all chemicals, fertilizers, insecticide and disinfectants are used in shrimp farming, shrimp hatcheries with the impact of these substances [28].

6 Alteration of Ecosystem Services by Shrimp Culture

For brackish water shrimp culture by intensive or semi-intensive methods, generally saline water is influent and effluent from source by proper channel. There is a huge difference in quantity of factors between influent and effluent water of shrimp farms. Effluent water of shrimp culture contains different chemical substances, organic material, excreta etc. Effluent water is generally drained into riverine or marine system. In marine water, nitrate is main limiting element of phytoplankton growth but in case of shrimp (most likely marine species: *Penaeus vannamei*, *Penaeus monodon*, *Penaeus stylirostris*) the nitrate concentration exceeds the standard value from more than 55% [29]. This causes major threat of eutrophication of sea and river. Apart from farming manipulation by different chemicals water gets contaminated by sediment and a direct relation occurs between sediment accumulations, pond water, food consumption rate, stocking density at different stages of shrimp [30].

Marine and riverine ecosystems provide lots of services: (a) water supply, (b) food, (c) salt production, (d) oxygen supply, (e) medium of transport, (f) maintenance of biogeochemical cycles, (g) cultural support, (h) spiritual support etc. Marine shrimp culture directly or indirectly alters these services. Popular practice of exotic shrimp culture decreases the value of native cultivated species. Discharged water from shrimp culture pollutes water and increases contamination of disease. For domestic water supply, the purification cost increases due to proper implementation of purification methods. Conversion of fertile land into shrimp farm is a recent trend in coastal countries of Asia, as results, soil fertility is permanently degraded and these changes are unaltered. This brief scenario of alteration of ecosystem services by shrimp culture is retention point of decision making.

7 Eco-Economic Analysis of Loss of Ecosystem Services and Economic Gain by Shrimp Culture

Shrimp culture as well as mariculture increase value of some services (e.g. food production value) consequently, decrease the value of some ecosystem services. Cost and benefit analysis is a useful economic method which can help to detect actual holistic benefits with relation to cost of ecosystem services [31].

$$PB_t = \frac{B_t}{(1+r)^t} \tag{1}$$

$$PC_t = \frac{C_t}{(1+r)^t} \tag{2}$$

Here, r is the discount rate; B_t is the benefit in *t*th year; C_t is the cost in *t*th year; PB_t and PC_t are discounted present value of cost and benefit respectively. Net present value (NPV) is another important factor of cost-benefit analysis. NPV represents net sum of benefits in every year.

$$NPV = \sum_{t=0}^{T} \frac{B_t - C_t}{(1+r)^t} \tag{3}$$

Benefit to cost ratio (BCR) is an important method of cost benefit analysis and formula of BCR is:

$$BCR = \sum_{t=0}^{T} \frac{B_t}{(1+r)^t} \Big/ \sum_{t=0}^{T} \frac{C_t}{(1+r)^t} \tag{4}$$

Higher value of BCR indicates efficient use of resource and environment. Among the economics, cost benefit analysis is popular method and now it shows its relevancy by its usefulness [32].

Emergy is a measure that determines transformation of energy into different forms. Each of the transformed energy contributes in different support system for human welfare and the flow of solar energy through different transformation has role in economy of human society. For production of one joule of any type of energy how much available energy is directly or indirectly required known as Emergy [33]. Solar energy is actual source of all energy. Collection of data, construction of energy flow diagram, "transformation of the flows into solar emergy", calculation emergy based indices are main analysing procedure [34]. Among these indices, environmental load ratio (ELR), emergy yield ratio (EYR) and environmental sustainability index (ESI) are important ones [34]. These indexes are important for accounting energy in social and economic sector for different services.

$$ELR = \frac{F + N}{F} \tag{5}$$

$$EYR = \frac{Y}{F} \tag{6}$$

$$ESI = \frac{EYR}{ELR} \tag{7}$$

Here, Y = the energy output, F = the energy input, measured by feedback from economy, N = non-renewable input.

Environmental load ratio (ELR) is useful to detect total nonrenewable energy released per unit of local renewable energy. Emergy yield ratio (EYR) expresses total energy output per total energy input. Environmental sustainability index (ESI) is the total emergy yield per environmental loading. So these indexes relate principles of thermodynamics with socioeconomic aspects as well as ecoeconomic consequences. In ecological economic studies the emergy is represented by emdollar which defines the ratio of the total yield from shrimp culture (expressed in currency) and total emergy in money. Elevated value of emdollar specifies more loss of ecosystem functioning and services [34, 35].

For the restoration of ecosystem services public support is mandatory [36]. Contingent Valuation Method (CVM) is used to detect the value of ecosystem services from people perspective by evaluating people willingness to pay (WTP) for restoration or maintaining specificity of ecosystem services [37]. Distinctness of demographic and other situation could be used as independent variables during formulation to forecast variation of WTP in relation to people's present condition [36].

FAO's dataset of Indian freshwater and marine shrimp production (in ton) from 2012 to 2016 reflected emerging production of shrimp but maximum production increases for the marine shrimp. Environmental and ecosystem damage is greater in case of marine shrimp culture. So, Fig. 2 indirectly indicates the increasing ecosystem alteration along with the shrimp production.

8 Ecosystem Service, Ecological Economics, Policy Making and Ecosystem Based Fisheries Management (EBFM)

Nonlinear change in ecosystem functioning and services always makes difficulties for policy makers. The concept of an externality of economics deals with consecutive result of ecosystem services [38]. For policy making, there is a need to proper assess-

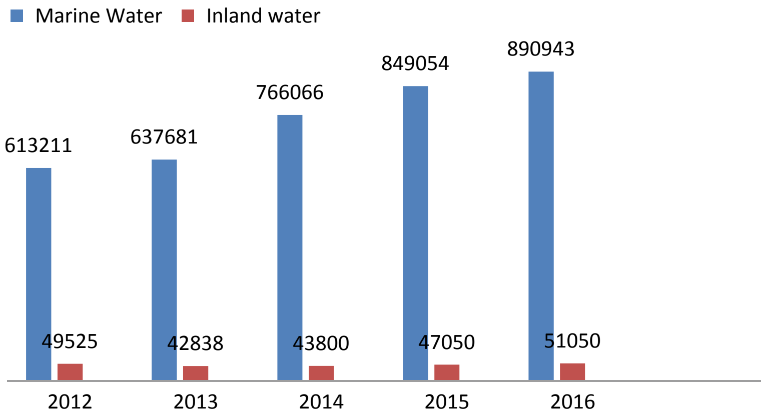


Fig. 2 Comparison productions of fresh water and marine shrimps in India (2012–2016)

ment of the linked effect of economy on environment and vice versa [39]. Decision makers or public policy makers always focus on some prime objective to overcome the conflicts between eco-economic and ecosystem services with human population's needs. In this aspect a good example has been reported by T. D. Crocker and J. Tschirhart: "the decision of a nation to gazette a new national park could require decision makers to (1) educate the public about the benefits of this decision, (2) attempt to ensure the equitable distribution of benefits (3) impute an economic valuation of the services provided by the park and (4) consider landscape management issues inside of and adjacent to the park" [39]. How can all ecological, sociocultural, economic cost and benefits act as a driving force of changes in ecosystem services and value of natural capital and how far the policy is viable with relation to the ecosystem dynamics are important questions before drawing any sustainable decision [40]. A holistic approach towards ecosystem and its all criteria are important aspects in policy making. Ecosystem based fisheries management (EBFM) depicts a holistic approach to maintaining ecosystem qualities and sustainability with relation to the benefits from ecosystem services [41]. The relevance of EBFM in policy making is useful as EBFM is accumulated overcome of 'ecological factor', 'human welfare' and 'management'.

9 Conclusion

Overall view of shrimp culture and its relationship with ecosystem services and ecological economics has been enlightened. In relation to growing human population size supply of protein diet is increasing day by day and that is enforcing growing aquaculture. In growing phase of aquaculture it modifies itself by innovative strategies and techniques. These transitions are cause of experimentation with environment as outcomes different alterations from ecosystem function to services occurs. The impact on human population is indirect and silent. Here, in this review focus is very specific. Shrimp is one of the major aquaculture products, emerging extensively in tropical and subtropical countries. Analysis of ecological economics, ecosystem services is very relevant for decision making and drawing a good policy to create balance between sustainable growth of shrimp culture and ecosystem individualities by its functions and features. The dynamic changes of ecosystem and shrimp as well as any cultivation is difficult to predict and it needs to establish a good model to understand the complexity in simple way.

References

1. Holmer, M., Black, K., Duarte, C.M., Marbà, N., Karakassis, I.: Aquaculture in the Ecosystem. Springer (2007)
2. Brugère, C., Aguilar-Manjarrez, J., Beveridge, M.C.M., Soto, D.: The ecosystem approach to aquaculture 10 years on—a critical review and consideration of its future role in blue growth. Rev. Aquac., 1–22 (2018)
3. Fast, A.W., Lester, L.J.: Marine Shrimp Culture: Principles and Practices, vol. 23. Elsevier (2013)
4. Ward, M., Possingham, H., Rhodes, J.R., Mumby, P.: Food, money and lobsters: valuing ecosystem services to align environmental management with sustainable development goals. Ecosyst. Serv. **29**, 56–69 (2018)
5. Black, K.D.: Mariculture, environmental, economic and social impacts. In: Encyclopedia of Ocean Sciences (2001)
6. Tacon, A.J.: Aquaculture Production Trends Analysis. Rome (1997)
7. Kataviæ, I.: Mariculture in the New Millennium Marikultura u novom tisuæljeæu. Agric. Conspec. Sci. **64**(3), 223–229 (1999)
8. Stickney, R.R.: Encyclopedia of Aquaculture. Wiley, New York (2000)
9. FAO: The State of World Fisheries and Aquaculture 2016. Rome (2016)
10. Tacon, A.G.J.: Increasing the contribution of aquaculture for food security and poverty alleviation. In: Third Millennium, Technical Proceedings of Conference on Aquaculture, pp. 63–72. Bangkok, Thailand, 20–25 Feb 2000 (2001)
11. Costanza, R., et al.: Twenty years of ecosystem services: how far have we come and how far do we still need to go? Ecosyst. Serv. **28**, 1–16 (2017)
12. Liu, Y., Li, J., Zhang, H.: An ecosystem service valuation of land use change in Taiyuan City, China. Ecol. Modell. **225**, 127–132 (2012)
13. Xu, Z., et al.: Energy modeling simulation of changes in ecosystem services before and after the implementation of a grain-for-green program on the Loess Plateau—a case study of the Zhifanggou valley in Ansai County, Shaanxi Province, China. Ecosyst. Serv. **31**, 32–43 (2018)
14. Costanza, R., Daly, H.E.: Natural capita and sustainable development. Conserv. Biol. **6**(1), 37–46 (1992)
15. Costanza, R., et al.: 'The value of the world's ecosystem services and natural capital. Nature **387**(6630), 253–260 (1997)
16. Van den Belt, V.: Ecological economics of Estuaries and Coasts. In: Van den Belt, V., Costanza, R. (eds.) Treatise on Estuarine and Coastal Science, pp. 1–13 (2011)
17. N. R. Council: Valuing Ecosystem Services: Toward Better Environmental Decision-making. National Academies Press (2005)
18. Freeman III, A.M., Herriges, J.A., Kling, C.L.: The Measurement of Environmental and Resource Values: Theory and Methods. RFF ORESS, New York (2014)

19. Hueting, R.: New Scarcity and Economic Growth: More Welfare Through Less Production? North-Holland Publishing, Amsterdam (1980)
20. Costanza, R., et al.: An Introduction to Ecological Economics. CRC Press (2014)
21. Costanza, R., et al.: Changes in the global value of ecosystem services. Glob. Environ. Chang. **26**(1), 152–158 (2014)
22. Balmford, A., Bruner, A., Cooper, P., Costanza, R., et al.: Economic reasons for conserving wild nature. Science **297**(5583), 950–953 (2002)
23. Rees, W.E.: Eco-footprint analysis: merits and brickbats. Ecol. Econ. **32**(3), 371–374 (2000)
24. Costanza, R.: The dynamics of the ecological footprint concept. Ecol. Econ. **32**, 341–345 (2000)
25. Valenti, W.C., New, M.B.: Grow-out systems—monoculture. In: New, M.B., Valenti, W.C. (eds.) Freshwater Prawn Culture: The Farming of *Macrobrachium -rosenbergii*, pp. 157–176 (2007)
26. New, M.B.: Farming Freshwater Prawns: A Manual for the Culture of the Giant River Prawn (*Macrobrachium rosenbergii*), no. 428. Food & Agriculture Org. (2002)
27. Dierberg, F.E., Kiattisimkul, W.: Issue, impacts, and implication of shrimp aquaculture in Thailand. Environ. Manage. **20**(5), 649–666 (1996)
28. Graslund, S., Bengtsson, B.: Chemicals and biological products used in south-east Asian shrimp farming, and their potential impact on the environment—a review. Sci. Total Environ. **280**(1–3), 93–131 (2001)
29. Ziemann, D.A., Walsh, W.A., Saphore, E.G., Fulton-Bennett, K.: A survey of water quality characteristics of effluent from hawaiian aquaculture facilities. J. World Aquac. Soc. **23**(3), 180–191 (1992)
30. Anh, P.T., Kroeze, C., Bush, S.R., Mol, A.P.J.: Water pollution by intensive brackish shrimp farming in south-east Vietnam: causes and options for control. Agric. Water Manag. **97**(6), 872–882 (2010)
31. Zheng, W., Shi, H., Chen, S., Zhu, M.: Benefit and cost analysis of mariculture based on ecosystem services. Ecol. Econ. **68**(6), 1626–1632 (2009)
32. Molinos-Senante, M., Hernàndez-Sancho, F., Sala-Garrido, R.: Economic feasibility study for wastewater treatment: a cost-benefit analysis. Sci. Total Environ. **408**(20), 4396–4402 (2010)
33. Odum, H.T.: Environmental Accounting: Emergy and Environmental Decision Making. Wiley (1996)
34. Shi, H., Zheng, W., Zhang, X., Zhu, M., Ding, D.: Ecological-economic assessment of monoculture and integrated multi-trophic aquaculture in Sanggou Bay of China. Aquaculture **410–411**, 172–178 (2013)
35. Odum, H.T.: Emergy evaluation of salmon pen culture (2001)
36. Nicosia, K., et al.: Determining the willingness to pay for ecosystem service restoration in a degraded coastal watershed: a ninth grade investigation. Ecol. Econ. **104**, 145–151 (2014)
37. Follain, J.R., Jimenez, E.: Estimating the demand for housing characteristics: a survey and critique. Reg. Sci. Urban Econ. **15**(1), 77–107 (1985)
38. Fisher, B., Turner, R.K., Morling, P.: Defining and classifying ecosystem services for decision making. Ecol. Econ. **68**(3), 643–653 (2009)
39. Crocker, T.D., Tschirhart, J.: Ecosystems, externalities, and economies. Environ. Resour. Econ. **2**(6), 551–567 (1992)
40. de Groot, R.S., Alkemade, R., Braat, L., Hein, L., Willemen, L.: Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. Ecol. Complex. **7**(3), 260–272 (2010)
41. Viet, T.: Ecosystem-based fishery management: a review of concepts and ecological economic models. **13**(2) (2012)
42. Valderrama, D., Anderson, J.L.: Shrimp Production Review. Guangzhou (2016)

Recent Trends of Ecosystem Health and Biodiversity Status of Pelagic-Benthic Coupled System in Indian Estuaries with Special Emphasis on Hooghly Estuary, India

Nabyendu Rakshit, Arnab Banerjee, Swagata Sinha and Santanu Ray

Abstract An Ecosystem can be defined as a community of interconnected elements comprising of living (biotic) and nonliving (abiotic) components of their surrounding environment interacting with each other. Among different types of ecosystems, estuarine system is of interest because the most sensitive land-water-atmosphere interactions are pronounced at these regions. It provides diverse habitat for wide variety of aquatic resources of ecological and economic significance including finfish, prawn, bivalve, gastropod, fiddler crab and plankton and so on. But recent years have seen gradual degradation of estuarine ecosystem, mainly in coastal landscape of India, owing to the different anthropogenic factors such as overfishing, development of agriculture and sewage from aquaculture farms, expansion of human settlements. These hamper the ecological balance affecting the food web of the concerned systems. It may lead to impacts including extinction of species, alternation of species diversity of different trophic levels, declination of mean trophic level within the system and significant habitat modification or destruction. Beside this, it also affects the social and economic wellbeing of the coastal communities. So it is important for us to know about recent trends of Indian estuarine ecosystems biodiversity and their proper sustainable management in future. For this purpose, understanding of how ecosystems are structured and how they function is much necessary and ecosystem health analysis is a more scientific and appropriate approach than any other ecological studies. So, our study emphasizes on three major aspects: (1) current scenario of ecosystem heath and biodiversity of Indian estuaries at both temporal and spatial scales; (2) Importance of Hooghly estuary and associated modelling studies (3) The necessary actions required for improvement of their ecosystem health status.

N. Rakshit · S. Sinha · S. Ray (✉)
Systems Ecology & Ecological Modelling Laboratory, Department of Zoology,
Visva-Bharati University, Santiniketan 731235, India
e-mail: sray@visva-bharati.ac.in

A. Banerjee
Centre for Mathematical Biology and Ecology, Department of Mathematics, Jadavpur University,
Kolkata 700032, India

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_21

1 Introduction

The term estuary is derived from the Latin ‘aestus’ meaning heat, boiling or tide. Specifically, the adjective ‘asetuarium’ means tidal. According to Oxford Dictionary estuary defines as “the tidal mouth of a great river that is where the tide meets the current”. However the most widely accepted definition of an estuary in the scientific literature is given by Pritchard [44]. An estuary is a semi-enclosed coastal body of water which has a free connection with the open sea and within which sea water is measurably diluted with fresh water derived from land drainage. The fresh water discharge from rivers varies with the seasonal and tidal changes during a year causes considerable fluctuations in salinity and other physico-chemical conditions resulting in significant changes in estuarine ecosystem. Besides this, estuaries have great ecological significance because the most sensitive land-water-atmosphere interactions are pronounced in these regions. Along with tropical rainforests and coral reefs, estuaries are considered as the world’s most productive ecosystems, more productive than the rivers and oceans that influence them from either side. This is due to the fact that nutrient rich river waters combine with warmer, light infused shallow coastal waters and the resulting upwelling process of nutrient-rich deep ocean waters support high primary productivity. The mixing of lighter fresh water and heavier salt water trap and circulate nutrients in such a manner that they are often retained and recycled by benthic (bottom dwelling) organisms to create a self-enriching system. Estuaries also act as energy subsidies through tidal transport of food and nutrients and removal of wastes. This vast primary productivity is the cornerstone of the estuarine food chain, providing food for large populations that inhabit the estuary. In some estuaries, where productivity exceeds the amount that can be used within the estuary, the action of regular tidal flushing moves nutrients and organic materials to adjacent coastal waters thereby increasing their productivity. So, estuaries play a key role in sustenance and maintenance of other ecosystems by enriching their productivity. An important characteristic of most estuaries is the presence of mangrove forest in the deltaic region of the river mouth. This unique vegetation has a significant role in promoting fish diversity. Thus, the estuaries and connected mangroves constitute important fishery resources of the country. As for example, the Hooghly Matla estuarine ecosystem with adjacent mangroves is one of the largest ecosystems of the world [41] and has great ecological importance in coastal landscape of India as it supports many essential fisheries of high economic value like Hilsa Population. It provides a nursery for the larval forms of some marine fish species and provides shelter and food for many young and adult fish and shellfish. These in turn provide food for other levels of the food chain including shore birds, waterfowl, larger fish and marine mammals.

Estuarine ecosystems are some of the most heavily used and threatened natural systems globally [12, 24, 68]. The degree of exploitation due to human activities is intense and increasing; 50% of salt marshes, 35% of mangroves, 30% of coral reefs, and 29% of seagrasses are either lost or degraded worldwide [40, 63, 64, 66, 67].

2 Studies on Indian Estuaries

India has a rich enormous bounding coastline and vast stretches of estuaries with an area of approximately 7500 km. India has 14 major, 44 medium and 162 minor rivers with a total catchment area of 3.12 × 106 km^2, discharging 1645 km^3 of freshwater every year to the seas around the country. The major rivers are Ganga, Mahanadi, Godavari, Krishna and Kaveri on the east coast and Narmada and Tapti on the west coast. These seven rivers have a catchment area of 1.83 × 106 km^2 and discharge 812 km^3 of freshwater transporting 1194 × 106 ton of silt to the marine waters every year [70]. Almost 50 estuaries are present along both east and west coast in India. These are summarized in the following Table 1.

Table 1 Summary of different Indian estuaries along both east and west coast in India

East Coast estuaries	West Coast estuaries
1. Hooghly estuary	25. Kadinamkulam estuary
2. Rushikulya estuary	26. Estuaries of Kochi
3. Bahuda estuary	27. Korapuzha estuary
4. Mahanadi estuary	28. Beypore estuary
5. Godavari and Krishna estuaries	29. Olipuram Kadavu backwaters
6. Gosthani estuary	30. Edava-Nadayara and Paravur backwaters
7. Kandaleru estuary	31. Poonthura estuary
8. Swarnamukhi and Konderu estuaries	32. Puthuponnani and Chandragiri
9. Araniar estuary	33. Shiriya, Thotapally and Pozhikara
10. Ennore estuary	34. Netravathi and Gurupur estuaries
11. Cooum estuary	35. Mulki estuary
12. Adyar estuary	36. Pavenje estuary
13. Muttukadu backwaters	37. Gangolli estuary
14. Edaiyur–Sadras estuarine complex	38. Kali estuary
15. Uppanar estuary	39. Mandovi–Zuari estuarine complex
16. Vellar estuary	40. Estuaries of Mumbai
17. Kollidam (Coleroon) estuary	41. Waghotana estuary
18. Kaveri (Cauvery) estuary	42. Vashishti estuary
19. Agniar estuary	43. Purna estuary
20. Kallar estuary	44. Mahi estuary
21. Pinnakayal and Pullavazhi estuaries	45. Damaganga - Kolak river estuaries
22. Athankarai and Kanjirangudi	46. Par river estuary
23. Kottakkarai, Uppar, Vaigai, Kottakkudy	47. Ambika-Kaveri-Kareira estuarine complex
24. Thengapattanam estuaries	48. Mindola river estuary
	49. Tapti and Narmada river estuaries
	50. Auranga estuary

Considerable work has been carried out on Indian estuaries such as Hooghly, Mahi, Narmada, Tapi, Puma, Par, Ambika, Auranga, Kolak, Damanganga, Ulhas, Mahim, Savitri, Kundalika, Vashisti, Ashatmudi and Erniore estuaries and Cochin Backwaters during last four decades. The Zoological Survey of India (ZSI), Kolkata and ENVIS centre, Annamalai University have published many reports on estuaries between 1985 and 1995. Many estuaries of the Indian subcontinent were thus studied and several experiments were performed on their hydrochemistry, geomorphology, and so on. These are summarized in following Tables 2 and 3.

3 Hooghly Estuary

Hooghly Matla estuarine ecosystem (HMES) ecosystem is one of the largest detritus-based ecosystems of the world [41]. This estuarine complex along with mangrove forest has achieved a notable place on the global map due to its greatest halophytic formation in the world [51]. It extends over two countries: India (West Bengal) and Bangladesh. The entire region is characterised by a dense network of rivers, canals and creeks. It is situated 21° 32′N and 22° 40′N; 88° 05′E and 89°E, at an altitude 0–10 m above sea level and just south of Kolkata. It houses the estuarine phase of the River Ganges and measures 9630 km^2, out of which 4262 km^2 is intertidal area. It also supports many essential fisheries of high economic value. Maximum catches are recorded (90%) from lower part of the estuary, including some fishing centres such as Diamond Harbour, Kakdwip, Namkhana, Bokkhali, Jambudwip, Frasergung and Sagar Island [26]. But in recent years this ecosystem is degrading gradually owing to the different anthropogenic factors such as overfishing, development of agriculture, sewage discharge from aquaculture farms expansion of human settlements (900 km^2; 2001 census), establishment of Farakka barrage in the upper stream of river, construction of Kolkata-Haldia ports and climatic factors such as rise in temperature, sea level, salinity and increasing frequency of severe cyclones such as Nargis, Bijli and Aila in recent years [14, 15]. These hamper the ecological balance affecting the food web of the concerned system. The increasing trend of fish yield over years (from 27014.5 tonnes in 1984–93 to 64204.3 tonnes in 1999–2003) has resulted in a drastic decline in the fish catch per unit effort (CPUE) from 189.7 kg in 1990–91 to 44 kg in 2002–03 [27, 38, 59]. Futhermore, it may lead to critical situations where fish population will no longer be able to sustain themselves and may become extinct. For example, species like Lisa tade, Lates calcarifer have declined and Tenualosa toli, Chanos chanos are totally absent in recent years [26, 59]. These situations may lead to impacts including alternation of species diversity of other trophic levels, declination of mean trophic level within the system and significant habitat modification or destruction. Besides this, it affects the social and economic wellbeing of the coastal communities that depend on fish for their way of life.

Table 2 Summary of work done in different Indian estuaries with respect to Hydro-chemical, Geo-chemical

Worker	Work done
Hydro-chemical properties	
Nair and Azis [37]	Studied hydrochemical and geochemical properties of water and sediment nutrients of Ashtamudi estuary, Kerala
Upadhyay [65]	Monitored seasonal pattern of temperature, salinity, pH, dissolved oxygen, phosphate, nitrate and silicate profiles of the Mahanadi estuary
Sastry and Chandramohan [58]	Comprehensive survey of the pollution status of Godavari estuary
Anilkumar et al. [2]	Monitored mixing characteristics and seasonal dynamics of Beypore estuary
Saha et al. [55]	Focused on physicochemical characteristics in relation to pollution and phytoplankton production potential of brackish water in Sundarbans of India
Balakrishna and Probst [4]	Studied the source of organic carbon and nitrogen in the Godavari river and its tributaries, the yield of organic carbon from the catchment, seasonal variability in their concentration and the ultimate flux of organic and inorganic carbon into the Bay of Bengal
Biswas et al. [6]	Studied seasonal and spatial variation of dissolved and atmospheric methane (CHU) in the estuaries of the Sundarban mangrove ecosystem.
Anil Kumar et al. [2]	Studied water quality of Adimalathura estuary, a small brackish water biotope of Kerala, exposed to pollution from the domestic wastes and coconut husk rotting. This study revealed the deleterious effects of waste disposal on the water quality and showed marked increase in the concentration level of nutrients and a decrease in dissolved oxygen
Prabu et al. [43]	Reported higher nutrient concentrations during the monsoon season and lower during summer while studying seasonal variations in Uppanar estuary
Krithika et al. [21]	Seasonal and tidal dynamics of dissolved nutrients, chlorophyll and primary production in the Pichavaram mangrove system, Southeast coast of India
Soundarapandian et al. [61]	Investigated physicochemical parameters such as rainfall, temperature, salinity, pH, dissolved oxygen and nutrients like nitrate, nitrite, inorganic phosphate and reactive silicate in Uppanar estuary, Cuddalore, southeast coast of India
Pradhan et al. [43]	Applied multivariate statistics and principal component analysis (PCA) to mark the influence of modernization on the quality of the Devi estuary
Geo-chemical properties	
Ghosh and Choudhury [9]	Focused on the sediments of Hoogly estuary and also distribution of organic carbon, total nitrogen, available nitrogen, total phosphorus and available phosphorus in relation to the texture of the sediment
Nasnolkar et al. [39]	Studied sediment organic carbon, total nitrogen, total phosphorous and hydrography of the overlying waters of the estuarine region in Mandovi estuary and a significant linear variation was indicated among nutrients and sediment characteristics
Rajasegar et al. [45]	Emphasized the effect of nutrient rich water from the shrimp farms on sediment composition, organic carbon, total phosphorus and total nitrogen content of sediments in Vellar estuary
Ram et al. [47]	Determined concentrations of Total Organic Carbon (TOC), Hg, Al, Fe in sediment of the Amba estuary in Mumbai harbor

Table 3 Summary of work done in different Indian estuaries with respect to biological aspects

Biological properties	
Goswami and Devassy [11]	Observed seasonal changes in the cladocera in Mandovi and Zuari estuaries
Godhantaraman [10]	Studied seasonal abundance and the relationship of microzooplankton with higher trophic levels in the tropical estuarine and mangrove waters, Parangipettai
Perumal et al. [41]	Investigated the hydrography, composition and community structure of phytoplankton and zooplankton, including chlorophyll-a content and primary productivity at the Kaduviyar estuary (Southeast coast of India)
Sanilkumar et al. [55]	Monitoring and surveillance of algal blooms along the southwest coast of India
Naik et al. [36]	Measured seasonal variations of phytoplankton and chlorophyll. A along with its environmental variations including nutrients in Mahanadi estuary
Ansari et al. and Ansari and Parulekar [3]	Revealed that the estuarine sediment was inhabited by a wide variety of benthic organisms whose population varies with season and location in Mandovi estuary
Harkantra and Rodrigues [13]	Investigated species diversity, biomass and population density of soft bottom macro fauna in relation to environmental influences of Mandovi and Zuari estuaries
Kailasam and Sivakami [16]	Monitored the effect of thermal effluent discharge on benthic fauna of Tuticorin bay
Murugesan et al. [31]	Explored the marine zone of Vellar estuary for temporal changes in community structure of benthos using advanced statistical tools
Rao et al. [50]	Explored the community structure of polychaete fauna of Godavari by using multivariate analysis

4 Modelling Studies on Ecosystem Health and Biodiversity Status of Hooghly Estuary, India

There have been several modeling studies regarding ecosystem health and biodiversity status of Hooghly estuary including static modelling and as well as processed dynamic modelling approaches. In static modelling approach, the system is assumed to be steady state condition and all inputs and outputs of each compartment are equal. But In process based dynamic modelling each compartment including both biotic and abiotic compartments are stimulated in respect to time variation. Both these modelling approaches are able to analyze qualitative as well as quantitative properties of the system such as dynamic behavior of the system, overall system exploitation status.

4.1 Process Based Dynamic Modelling Approaches

Two multi-dimensional dynamic simulation models were proposed by Mandal et al. [25] in order to quantify the contribution of dissolved inorganic nitrogen (DIN) to estuarine water from adjacent mangrove forest and also the impact of this DIN to plankton dynamics of the Hooghly-Matla estuarine system. In their first modelling approach, Nitrogen is considered as important nutrient which occurs in various organic and inorganic forms such as soil total nitrogen, soil organic nitrogen, soil inorganic nitrogen, total organic nitrogen of water, dissolved organic nitrogen of water, particulate organic nitrogen of water and dissolved inorganic nitrogen of water and it plays a crucial role in the regulation of productivity in this estuarine system. Modelling of nitrogen dynamics from mangrove litterfall and particularly the release of dissolved inorganic nitrogen in this estuarine system is important because of its role in primary productivity by means of growth of phytoplankton and other higher plants and all other biological components of grazing food chain. The most important aspect of model output is sensitivity analysis with respect to proper sustainable management of estuarine system and this study reveals that the leaching rate of soil organic nitrogen to total organic nitrogen of water and loss rate of soil organic nitrogen as humic acid and fulvic acids are very sensitive parameters in this system. So, these parameters should be considered for future management practices. Afterwards, the final study revealed that the availability of nutrient throughout the stretch of estuary depends on leaching rate of inorganic nitrogen and uptake of nutrient by phytoplankton which reinforces the previous study and helps to understand the grazing food chain more elaborately. Maximum nutrient load in estuary occurs in monsoon. Phytoplankton biomass solely depends upon the nutrient availability and grazing of zooplankton. These studies are focused on nitrogen biogeochemical process and gazing pathway in this system.

But in every system mainly mangrove-based detritus pathway is much more important than gazing pathway. Decomposition and subsequent remineralization of mangrove litter is main source of detritus and it also paly important role in nutrient dynamics both in the forest as well as estuarine system. A five-compartment dynamic model of detritus food web dynamics have been developed by the authors Roy et al. [54] to study the impact of detritivorous fish on the mangrove estuarine detritus food web. Almost 70% of the detritus formed in the soil was being washed to the estuarine water which acts as source or sink of nutrient for the primary producers of aquatic food chain. A significant amount of detritus in the estuarine water is readily consumed by a group of detritivorous fishes before it is rematerialized completely to its inorganic nutrient form. Model results depict the dependence of detritivorous fishes on detritus biomass which is further dependent upon several factors like mortality of phytoplankton, zooplankton, and detritivorous fishes; and chiefly on litter biomass and litter decomposition. So, this study revealed the importance of detritus pathway for maintaining mineral cycle and subsequently the ecosystem health of this estuary.

It is a well-known fact that mangrove forests serve as an important source of carbon and other nutrients to the adjacent lagoonal and coastal systems. To represent important role of carbon as nutrient, the authors Mukherjee et al. [30] proposed a seven compartment model focusing on the dynamics of carbon in this estuarine system. Different forms of carbon present in soil (soil organic carbon (SOC), soil inorganic carbon (SIC)) and in water (dissolved inorganic carbon (DIC), dissolved carbon dioxide (DCO2), dissolved bicarbonate (DBC), dissolved organic carbon (DOC) and particulate organic carbon (POC)) are taken into consideration. Model simulation results show seasonal variations of litterfall biomass—the main source of SOC pool that is ultimately transported to the estuary. Besides this, the death of organisms in soil and water increases the SOC and POC content respectively. pH of water is major controlling factor and depending on this, DIC is converted to DCO2 and DBC, which serve as raw material during photosynthesis of phytoplankton. Mineralization rate of SOC to SIC and uptake rate of DCO2 and DBC are considered as most sensitive parameters and need more attention in future.

All these studies regarding biogeochemical processes and nutrient cycling of major elements like nitrogen, phosphorus and carbon helps to understand quantitative and qualitative measurement of the concerned system.

4.2 Static Modelling Approaches Regarding Ecosystem Health

To evaluate ecosystem health and the anthropogenic effect on Sundarban mangrove estuarine ecosystem, best-known detritus-based ecosystems, two approaches like spatial as well as temporal scale are done by Prof. Ray and his research team. During his first study [51], two islands, one from virgin part (Prentice island) and the second from reclaimed part (Sagar island) of the estuary are selected for comparative study of benthic food webs of both by using network analyses. Prentice island is almost free from both direct and indirect human interferences and fully covered by different mangrove vegetations. The soil of this island is typically muddy, alluvial in nature and full of detritus coming from huge litterfall of the mangroves. On the other hand Sagar island, the largest in the deltaic system, is fully disturbed due to all sorts of direct and indirect anthropogenic stresses. Most part of the island is occupied by human habitation. The results demonstrate a remarkable difference between these two islands- where the virgin ecosystem is dominantly controlled by detritus supplied from the litterfall of mangroves, the bottom community of reclaimed island receives a large contribution from the phytoplankton populations. Using mineral cycling as an indicator, the Virgin island appeared to be in good health as the detritus pathway is more dominant there. The number of pathways of recycling is found to be much higher in undisturbed system in comparison to that of the reclaimed as indicated by the low Finn cycling index in the disturbed part. Litterfall comprises only 16 in reclaimed island where as in virgin Island it is about 70%. Pathway redundancy is rather high in disturbed system, indicating the system is highly resilient to further perturbation which is a desirable criterion for healthy ecosystem. However, in virgin forest the ascendancy value is much higher than the redundancy, showing that the system is healthy and almost free from any anthropogenic stress.

To understand exploitation status of ecosystem health Ecopath with Ecosim software as static modelling tool for temporal comparison of trophic structure of the Hooghly-Matla estuarine system have applied by Rakshit et al. [49] for both qualitative and quantitative assessment of due to anthropogenic effects. Two mass-balanced network models of Hooghly Matla estuarine system, from two different time periods (less exploited phase 1985–1990 and highly exploited phase 1998–2003 were constructed for this purpose. The models were used to estimate the important biological interactions and relationships among different ecologically important groups. Twenty functional groups based on species of different habitats from coastal areas in this ecosystem have been selected, including shrimps, squids, crabs, mackerel, small pelagics, demersal fishes, benthic feeders, predator fishes and trash fish (Fig. 1). The biomass values for all components are estimated from catch production and bottom trawling surveys.

The values of Ecotrophic Efficiency in the models are high (>0.5) for most groups of higher trophic levels. Most fish population approached high degree of exploitation with change in the overall trophic structure mainly due to top down effects. Several system statistics and network flow indices from the model outputs indicate that this estuary is facing degradation and stress resulting in some degrees of instability (see Table 2).

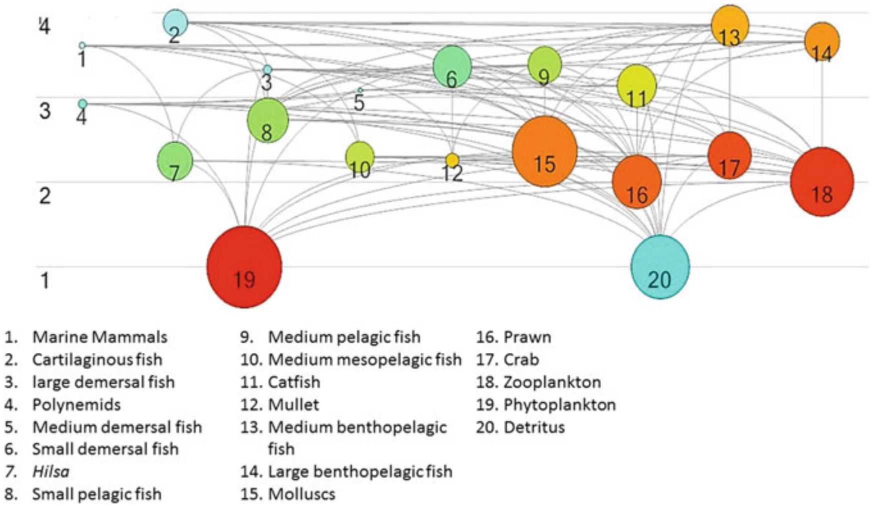


Fig. 1 Flow diagram of the HMES where the nodes represent the components, curved lines show food web connectivity and horizontal straight lines represent the trophic levels

5 Future Prospects

Even Existing ecosystem models of Hooghly estuary both dynamic and static models are briefly discussed and summarized. But coupling of physical and biological models are still unexplored in this system due to much more complexity. In future, it will be better predicting biodiversity status of whole system, its productivity and as well as dynamic behavior of benthic pelagic coupling. In this regard, it should be mentioned that pelagic-benthic coupling plays an important role in nutrient cycling and increasing productivity of system. Recent study by Rakshit et al. [49] using static models for temporal comparison of trophic structure of the Hooghly-Matla estuarine system considers only a few benthic compartments. However, modelling approaches including benthic-pelagic coupling have been negligible. So. In future more scientific and better predicting aquatic models including pelagic as well as benthic environment are expected. Now it should be concluded for development of better predicting models, extensive and more scientific empirical works should be required (Tables 4 and 5).

Table 4 Functional fish groups including species with references included in modelling of Hooghly estuary

Group	Included species	Biomass	P/B	Q/B
Cartilaginous fish	Skates or Guitarfish (*Rhinobatos* sp.), Sawfish (*Pristis microdon*), sting ray (*Himantura* sp.) and Shark (*Scoliodon laticaudus*)	Srinath et al. [62]	Mohamed et al. [28]	
Polynemids	*Polynemus paradiseus*, *Eleutheronema tetradactylum* (Gurjeoli)	Nath et al. [38], Sinha [58]	Khan et al. [34] and Nabi et al. [18]	
Demersal fish	Large (*Lates calcarifer*), Medium (*Sillaginopsis panijus*) or small (*Harpadon nehereus*)	Nath et al. [38], Sinha [58]	Balli et al. [5], Karmakar [17], Khan et al. [19, 20]	
Pelagic fish	Medium Pelagic fish (*Rastrelliger kanagurta*) and small pelagic fish (*Setipinna* sp., *Coilia* sp.)	Nath et al. [38], Sinha [58]	Khan et al. [19, 20], Nabi et al. [33]	
Benthopelagic fish	Medium (*Trichiurus gangeticus*–Ganges hairtail) and large (*Daysciaena albida, Sciaena biauritus, Otolithoides pama*)	Nath et al. [38], Sinha [58]	Chakraborty [7], Reuben et al. [52]	
Mesopelagic Fish	*Pampus argenteus* (Butter fish)	Nath et al. [38], Sinha [58]	Khan et al. [20]	
Mullets	*Liza parsia, Liza tade*	Nath et al. [38], Sinha [58]	Moorthy et al. [29], Rangaswamy [48]	
Catfish	*Tachysurus jella, Mystus gulio, Plotosus canius, Pangasius pangasius, Osteogeneiosus militaris*	Nath et al. [38], Sinha [58]	Raje et al. [4]	
Hilsa	*Tenualosa ilisha; Ilisha megaloptera; Ilisha elongate; Ilisha toil*	Nath et al. [38], Sinha [58]	Amin et al. [1], Reuben et al. [52]	

Table 5 System statistics, flows higher-order indices of Hooghly Matla estuary for the time periods 1985–1990 and 1998–2003

Parameter/ecosystem indices	Unit	Phase 1 (1985–1990)	Phase 2 (1998–2003)
Sum of all consumption	t km^{-}2 year^{-}1	5743.80	4009.51
Sum of all exports	t km^{-}2 year^{-}1	5262.76	2124.3
Sum of all respiratory flows	t km^{-}2 year^{-}1	3051.45	2313.11
Sum of all flows into detritus	t km^{-}2 year^{-}1	8332.25	4168.85
Total system throughput	t km^{-}2 year^{-}1	22,390.26	12615.76
Sum of all production	t km^{-}2 year^{-}1	11374.84	11276.31
Calulated total net primary production	t km^{-}2 year^{-}1	9831.25	10381.80
Total primary production/total respiration		3.22	4.49
Net system production	t km^{-}2 year^{-}1	6779.81	8068.7
Total Primary production/total biomass		40.34	41.57
Total biomass/total throughput	year^{-}1	0.02	0.01
Total biomass (excluding detritus)	t km^{-}2	236.53	257.35
System omnivory index		0.24	0.24
Ecopath pedigree index		0.54	0.54
Measure of fit, t*		2.65	2.65
Ascendancy	Flowbits	22287 (36.32%)	12993 (30%)
Throughput	t/km^2/year	22390.26	126165.760
Overhead	Flowbits	39077 (63.68%)	30324 (70%)
Development capacity		61365	43317
Throughput cycled (excluding detritus)	t/km^2/year	39.61	42.65
Predatory cycling index (excluding detritus)	% of throughput	0.69	0.99
Throughput cycled (including detritus)	t/km^2/year	1735	1059
Finn's cycling index	% of throughput	7.75	8.4
Finn's mean path length		2.69	2.84
Finn's straight through path length (excluding detritus)		1.86	1.83
Finn's straight through path length (excluding detritus)		2.48	2.61

6 Conclusion

Estuaries are also highly productive and play an important role in the food chain. Now a day the conservation importance of estuaries has already been noted due to the fact that most of the estuaries areas around the world are approaching towards degradation. At the same time, urbanizations that have developed alongside estuaries are often low lying and prone to flooding. The problem is made more complicated due to the complex nature of estuary systems. Within an estuary, form and process are extremely linked and there are no obvious dependent and independent variables or clear cause-effect hierarchies. This interdependence means that changes in one part of the system can cause responses elsewhere in the estuary. So, management issues for the estuaries should be dealt depending upon different aspects like habitat resoration, resource conservation, policies and planning along with examples and cases studies from various part of India. On the other hand, it is necessary that system analyses and management outlines should be focused on the policy makers and managers interested in understanding and protecting the precious estuarine systems of the country. By that way, a strategic approach to estuary management must be developed in future to overcome biodiversity loss, extinction of species and future endangerment as well.

Acknowledgements Author Nabyendu Rakshit acknowledges the University Grants Commission, Govt. of India, for NET-JRF fellowship, Ref. No. 20161430758. Author Arnab Banerjee acknowledges the University Grants Commission, Govt. of India, for providing Dr. D S Kothari Post Doctoral fellowship, F No: BL/17-18/0490.

References

1. Amin, S.M.N., Rahman, M.A., Haldar, G.C., Mazid, M.A., Milton, D.: Population dynamics and stock assessment of Hilsa shad, *Tenualosa ilisha* in Bangladesh. Asian Fish. Sci. **15**, 123–128 (2002)
2. Anilkumar, N., Sankaranarayanan, V.N., Josanto, V.: Studies on mixing of the waters of different salinity gradients using Richardsons number and the suspended sediment distribution in the Beypore estuary, south west coast of India (1999)
3. Ansari, Z.A., Parulekar, A.H.: Distribution, abundance and ecology of the meiofauna in a tropical estuary along the west coast of India. Hydrobiologia **262**(2), 115–126 (1993)
4. Balakrishna, K., Probst, J.L.: Organic carbon transport and C/N ratio variations in a large tropical river: Godavari as a case study. India. Biogeochem. **73**(3), 457–473 (2005)
5. Balli, J.J., Chakraborty, S.K., Jaiswar, A.K.: Population dynamics of Bombay duck *Harpadontidae nehereus* (Ham, 1822)(Teleostomi/Harpadontidae) from Mumbai waters. India. Indian J. Mar. Sci. **40**, 67 (2011)
6. Biswas, H., Mukhopadhyay, S.K., Sen, S., Jana, T.K.: Spatial and temporal patterns of methane dynamics in the tropical mangrove dominated estuary, NE coast of Bay of Bengal. India J. Marine Syst. **68**(1–2), 55–64 (2007)
7. Chakraborty, S.K.: Fishery, age, growth and mortality estimates of *Trichiurus lepturus* Linnaeus from Bombay waters. Indian J. Fish. **37**, 1–7 (1990)
8. Chakraborty, S.K., Devadoss, P., Manojkumar, P.P., Feroz Khan, M., Jayasankar, P., Sivakami, S., Gandhi, V., Appanasastry, Y., Raju, A., Livingston, P., Amcer Hamsa, K.M.S., Badruddin, M., Ramalingam, P., Dhareswar, V.M., Seshagirl Rao, C. V, Nandakumaran, K., Chavan, B.B.,

Seetha, P.K.: The fishery, biology and stock assessment of jew fish resources of India. Mar. Fish. Res. Manag. 604–616 (2000)

9. Ghosh, P.B., Choudhury, A.: Copper, zinc and lead in the sediment of Hooghly estuary. Environment and Ecology **7**, 427–430 (1989)
10. Godhantaraman, N.: Seasonal variations in taxonomic composition, abundance and food web relationship of microzooplankton in estuarine and mangrove waters, Parangipettai region, southeast coast of India (2001)
11. Goswami, S.C. and Devassy, V.P.: Seasonal fluctuations in the occurrence of Cladocera in the Mandovi-Zuari estuarine waters of Goa (1991)
12. Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E. and Fujita, R., : A global map of human impact on marine ecosystems. Science **319**(5865), 948–952 (2008)
13. Harkantra, S.N. and Rodrigues, N.R.: Environmental influences on the species diversity, biomass and population density of soft bottom macrofauna in the estuarine system of Goa, west coast of India (2004)
14. Hazra, S., Ghosh, T., DasGupta, R., Sen, G.: Sea level and associated changes in the Sundarbans. Sci. Cult. **68**, 309–321 (2002)
15. Hazra, S., Samanta, K., Mukhopadhyay, A., Akhand, A.: Temporal change detection (2001–2008) of the Sundarban. Unpubl, Report, WWF-India (2010)
16. Kailasam, M., Sivakami, S.: Effect of thermal effluent discharge on benthic fauna off Tuticorin bay, south east coast of India. Jndian Journal of Marine Sciences **33**(2), 194–201 (2004)
17. Karmakar, J.K.: Study on some aspect of biology with special reference to population dynamics of Bhetki; *Lates calcarifer* (Bolch, 1970). University of Chittagong, Bangladesh (2003)
18. Khan, M.G., Islam, M.S., Quayum, S.A., Sada, M.N.U., Chowdhury, Z.A.: Biology of the fish and shrimp population expoloited by estuarine set bagnet, in: BOBP Seminar. Coxs Bazar, Bangladesh, p. 20 (1992)
19. Khan, M.Z., Kumaran, M., Jayaprakash, A.A., Scariah, K.S., Deshmukh, V.M., Dhulkhed, M.H.: Stock assessment of pomfrets off west coast of India. Indian J. Fish. **39**, 249–259 (1992)
20. Khan, S., Banu, N., Isabella, B.: Studies on some aspects of the biology and fecundity of *Mystus tengra* (Hamilton-Buchanan). Bangladesh J. Zool **20**, 151–160 (1992)
21. Krithika, K., Purvaja, R., Ramesh, R.: Fluxes of methane and nitrous oxide from an Indian mangrove. Current Science 218–224 (2008)
22. Kumary, K.A., Azis, P.A., Natarajan, P.: Water quality of the Adimalathura Estuary, southwest coast of India. J. mar. biol. Ass. India **49**(1), 01–06 (2007)
23. Kumar, R.S.: Soil macro-invertebrates in a prawn culture field of a tropical estuary. Indian Journal of Fisheries **49**(4), 451–455 (2002)
24. Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., Jackson, J.B.: Depletion, degradation, and recovery potential of estuaries and coastal seas. Science **312**(5781), 1806–1809 (2006)
25. Mandal, S., Ray, S., Ghosh, P.B.: Modelling of the contribution of dissolved inorganic nitrogen (DIN) from litterfall of adjacent mangrove forest to Hooghly-Matla estuary. India. Ecological Modelling **220**(21), 2988–3000 (2009)
26. Mitra, P.M., Karmakar, H.C., Ghosh, A.K.: Fisheries of Hooghly-Matlah estuarine system: further appraisal 1994-95 to 1999-2000 (Bulletin no. 109) (2001)
27. Mitra, P.M., Karmakar, H.C., Sinha, M., Ghosh, A., Saigal, B.N.: Fisheries of the Hooghly-Matlah estuarine system-an appraisal (CIFRI Bulletin no. 67) (1997)
28. Mohamed, K.S., Zacharia, P.U., Muthiah, C., Abdurahiman, K.P., Nayek, T.H.: A Trophic Model of the Arabian Sea Ecosystem of Karnatakaand Simulation of Fishery Yields for its Multigear Marine Fisheries. Kerala, India (2005)
29. Moorthy, K.S.V., Reddy, H.R.V., Annappaswamy, T.S.: Age and growth of blue spot mullet, *Valamugil seheli* (Forskal) from Mangalore. Indian J. Fish. **50**, 73–79 (2003)
30. Mukherjee, J., Ray, S., Ghosh, P.B.: A system dynamic modeling of carbon cycle from mangrove litter to the adjacent Hooghly estuary, India. Ecological modelling **252**, 185–195 (2013)

31. Murugesan, P., Ajmal Khan, S., Ajithkumar, T.T.: Temporal changes in the benthic community structure of the marine zone of Vellar estuary southeast coast of India. Journal of the Marine Biological Association of India **49**(2), 154–158 (2007)
32. Nabi, M.R.: Management of Estuarine Set Bag Net Fishery of Bangladesh: Application of Traditional Scientific Methods, Sustainable Livelihoods Approach and Local Indigenous Knowledge. Ph. D. Dissertation, Borneo Marine Research Institute, University Malaysia Sabah, Malaysia (2007)
33. Nabi, M.R., Hoque, M.A., Rahman, R.A., Mustafa, S., Kader, M.A.: Population dynamics of *Polynemus paradiseus* from estuarine set bag net fishery of Bangladesh. Chiang Mai J. Sci. **34**, 355–365 (2007)
34. Nandy, A.C., Bagchi, M.M., Majumder, S.K.: Ecological changes in the Hooghly estuary due to water release from Farakka Barrage. Mahasagar **16**, 209–220 (1983)
35. Narasimham, K.A., Kripa, V., Balan, K.: Molluscan shellfish resources of India-an overview. Indian J. Fish. **40**, 112–124 (1993)
36. Naik, S., Acharya, B.C. and Mohapatra, A.: Seasonal variations of phytoplankton in Mahanadi estuary, east coast of India (2009)
37. Nair, N.B. and Azis, P.A.: Hydrobiology of the Ashtamudi estuary- a tropical backwater system in Kerala (1987)
38. Nath, D., Misra, R.N., Karmakar, H.C.: The Hooghly estuarine system-ecological flux, fishery resources and production potential, p. 130. Bull. Cent. Inl. Fish. Res, Inst (2004)
39. Nasnolkar, C.M., Shirodkar, P.V., Singbal, S.Y.S.: Studies on organic carbon, nitrogen and phosphorus in the sediments of Mandovi Estuary. Goa. Oceanographic Literature Review **12**(43), 1208 (1996)
40. Pillay, T.V.R.: Biology of the Hilsa, *Hilsa ilisha* (Hamilton) of the river Hooghly. Indian J. Fish. **5**, 201–257 (1958)
41. Perumal, N.V., Rajkumar, M., Perumal, P. and Rajasekar, K.T.: Seasonal variations of plankton diversity in the Kaduviyar estuary, Nagapattinam, southeast coast of India (2009)
42. Prabu, V.A., Rajkumar, M., Perumal, P.: Seasonal variations in physico-chemical characteristics of Pichavaram mangroves, southeast coast of India. J. Environ. Biol **29**(6), 945–950 (2008)
43. Pradhan, U.K., Shirodkar, P.V., Sahu, B.K.: Physico-chemical characteristics of the coastal water off Devi estuary, Orissa and evaluation of its seasonal changes using chemometric techniques. Current Science 1203–1209 (2009)
44. Pritchard, D.W.: Estuarine classification—a help or a hindrance. In Estuarine circulation (pp. 1-38). Humana Press (1989)
45. Rajasegar, M., Srinivasan, M. and Khan, S.A.: Distribution of sediment nutrients of Vellar estuary in relation to shrimp farming (2002)
46. Raje, S.G., Dineshbabu, A.P., DAS, T.: Biology and stock assessment of *Tachysurus jella* (Day). Indian J. Fish. **55**, 295–299 (2008)
47. Ram, A., Rokade, M.A. and Zingde, M.D.: Mercury enrichment in sediments of Amba estuary (2009)
48. Rangaswamy, C.P.: Maturity and spawning of Mugil cephalus of Lake Pulicat. Recent Res. Estuar. Biol. 47–60 (1975)
49. Rakshit, N., Banerjee, A., Mukherjee, J., Chakrabarty, M., Borrett, S.R., Ray, S.: Comparative study of food webs from two different time periods of Hooghly Matla estuarine system, India through network analysis. Ecological Modelling **356**, 25–37 (2017)
50. Rao, D.S., Rao, M.S., Annapurna, C.: Polychaete community structure of Vasishta Godavari estuary, east coast of India. Journal of the Marine Biological Association of India **51**(2), 137–144 (2009)
51. Ray, S.: Comparative study of virgin and reclaimed islands of Sundarban mangrove ecosystem through network analysis. ecological modelling, 215(1-3), pp.207-216 (2008)
52. Reuben, S., Dan, S.S., Somaraju, M.V., Philipose, V., Sathianandan, T.V.: The resources of hilsa shad, *Hilsa ilisha* (Hamilton), along the northeast coast of India. Indian J. Fish. **39**, 169–181 (1992)

53. Reuben, S., Vijayakumaran, K., Achayya, P., Prabhakar, R.V.D.: Biology and exploitation of *Trichiurus lepturus* Linnaeus from Visakhapatnam waters. Indian J. Fish. **44**, 101–110 (1997)
54. Roy, M., Ray, S., Ghosh, P.B.: Modelling of impact of detritus on detritivorous food chain of Sundarban mangrove ecosystem, India. Procedia Environmental Sciences **13**, 377–390 (2012)
55. Saha, S.B., Mitra, A., Bhattacharyya, S.B., Choudhury, A.: Status of sediment with special reference to heavy metal pollution of a brackishwater tidal ecosystem in northern Sundarbans of West Bengal. Tropical Ecology **42**(1), 127–132 (2001)
56. Sanilkumar, M.G., Joseph, K.J. and Saramma, A.V.: Microalgae in the southwest coast of India (Doctoral dissertation, Cochin University of Science & Technology) (2009)
57. Sarkar, S.K., Singh, B.N., Choudhury, A.: Composition and variations in the abundance of zooplankton in the Hooghly estuary, West Bengal. India. Proc. Anim. Sci. **95**, 125–134 (1986)
58. Sastry, A.G.R. and Chandramohan, P.: Diel and tidal fluctuations in the water quality of Vasishta Godavari estuary (1990)
59. Sinha, M.: Farakka barrage and its impact on the hydrology and fishery of Hooghly estuary, in: The Ganges Water Diversion: Environmental Effects and Implications. Springer, pp. 103–124 (2004)
60. Smith, B.D., Braulik, G., Strindberg, S., Ahmed, B., Mansur, R.: Abundance of Irrawaddy dolphins (*Orcaella brevirostris*) and Ganges river dolphins (*Platanista gangetica gangetica*) estimated using concurrent counts made by independent teams in waterways of the Sundarbans mangrove forest in Bangladesh. Mar. Mammal Sci. **22**, 527–547 (2006)
61. Soundarapandian, P., Premkumar, T., Dinakaran, G.K.: Studies on the physico-chemical characteristic and nutrients in the Uppanar estuary of Cuddalore, South east coast of India. Curr. Res. J. Biol. Sci **1**(3), 102–105 (2009)
62. Srinath, M., Kuriakose, S., Ammini, P.L., Prasad, C.J., Ramani, K., Beena, M.R.: Marine Fish Landings in India 1985–2004. C. Spec. Publ. **89**, 1–161 (2006)
63. UNEP World Conservation Monitoring Centre and Census of Marine Life on Seamounts (Programme). Data Analysis Working Group.: Seamounts, deep-sea corals and fisheries: vulnerability of deep-sea corals to fishing on seamounts beyond areas of national jurisdiction (No. 183). UNEP/Earthprint (2006)
64. UNFAO.: Statistical Year Book (2012)
65. Upadhyay, S.: Physico-chemical characteristics of the Mahanadi estuarine ecosystem, east coast of India. Indian J Mar Sci (1988)
66. Valiela, I., Bowen, J.L., York, J.K.: Mangrove Forests: One of the World's Threatened Major Tropical Environments: At least 35 of the area of mangrove forests has been lost in the past two decades, losses that exceed those for tropical rain forests and coral reefs, two other well-known threatened environments. Bioscience **51**(10), 807–815 (2001)
67. Waycott, M., Duarte, C.M., Carruthers, T.J., Orth, R.J., Dennison, W.C., Olyarnik, S., Calladine, A., Fourqurean, J.W., Heck, K.L., Hughes, A.R., Kendrick, G.A.: Accelerating loss of seagrasses across the globe threatens coastal ecosystems. Proceedings of the National Academy of Sciences **106**(30), 12377–12381 (2009)
68. Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B., Lotze, H.K., Micheli, F., Palumbi, S.R. and Sala, E.: Impacts of biodiversity loss on ocean ecosystem services. science, 314(5800), pp.787-790 (2006)
69. Zafar, M., Mustafa, M.G., Amin, S.M.N., Akhter, S.: Studies on population dynamics of *Acetes indicus* from Bangladesh coast. J. Nat. Ocea. Mar. Inst **14**, 1–15 (1997)
70. Zingde, M.D., Bhosle, N.B., Narvekar, P.V., Desai, B.N.: Hydrography and water quality of Bombay harbour. Society of Biosciences, Muzaffarnagar (1989)

Guanotrophication by Waterbirds in Freshwater Lakes: A Review on Ecosystem Perspective

Sagar Adhurya, Suvendu Das and Santanu Ray

Abstract Freshwater lakes throughout the world are nowadays threatened by climate change and eutrophication. As a result the crisis of drinking water and disease outbreak resulting from poor water quality has become a global issue for policymakers. Waterbirds are an inherent part of freshwater lake ecosystem and the effects of waterbird on lake ecosystem remain an important area of research in the last century. Many freshwater lakes throughout the globe support a huge number of waterbird congregations. Nutrients from waterbird droppings may serve as an important source of nutrients to those freshwater lakes and sometime may lead to eutrophication (by excess nitrogen and phosphorus loading). Waterbird can modulate the nutrient dynamics of the waterbody in two ways: (i) by changing the form of nutrient in the waterbody (nutrient cycling), (ii) by bringing nutrient from another area to waterbody (external nutrient loading) and the vice versa (nutrient input). Several attempts have been taken to estimate nutrient loading by waterbirds in waterbodies and mostly relies on extrapolation from bird count. Recent approaches incorporate avian bioenergetics for nutrient loading estimation. Effects of bird faeces on water quality, the faecal nutrient content of different species, nutrient loading estimation methods, factors affecting nutrient loading by waterbirds, ways to reduce the nutrient level in the waterbody and nutrient cycling pathways of waterbird faeces in the aquatic ecosystem are reviewed in this article with a brief future perspective. The study of guanotrophication is not only important with respect to water quality maintenance but also for conservation of waterbirds and their habitat.

Keywords Eutrophication · Guano · Nitrogen · Phosphorus · Food chain

S. Adhurya · S. Das · S. Ray (✉)
Systems Ecology & Ecological modeling Laboratory Department of Zoology,
Visva-Bharati University, Santiniketan 731235, India
e-mail: santanu.ray@visva-bharati.ac.in

S. Adhurya
e-mail: sagaradhurya.rs@visva-bharati.ac.in

S. Das
e-mail: suvendudas.rs@visva-bharati.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_22

1 Introduction

Freshwater lakes throughout the world are now threatened by climate change and eutrophication. As a result crisis of drinking water and disease outbreak resulting from poor water quality has become a global issue for policy makers. In addition to atmospheric input, anthropogenic (fertilizer, drainage etc.) and biotic input (excreta of different animals, dead materials etc.) are sources of nutrient to waterbodies. Being an inherent part of the aquatic ecosystem, the study about effect of waterbirds on the ecosystem remained an important area of research since middle of the last century. It is evident from previous studies that nutrients from bird guano may serve as a vital agent for eutrophication of the waterbodies with the high avian congregation (discussed in Sect. 3). Guanotrophication is a popular term for the nutrient enrichment of water bodies by bird droppings [1]. According to previous studies, congregation of huge number of waterbird in water bodies like lakes for long period may have undesirable effect on water quality through destruction of wetland vegetation [2], reduction in water quality [3, 4], increase of heavy metal concentration [5], increase of coliform bacteria [6] and increase of water hardness [7, 8]. Degraded water quality also may result in a disease outbreak in waterbirds [9, 10]. In addition, guanotrophication also may promote changes in zooplankton community [11, 12] (Fig. 1).

The term waterbird or aquatic avifauna refers to cormorants, ducks, grebes, moorhen, waterhen, jacana etc., which feeds and roosts on the waterbody. Another term used multiple times in this review is waterfowl, which indicates mainly the member of the Order Anseriformes (Ducks, Swans, Geese etc.). Waterbird can modulate the nutrient dynamics of the waterbody in three ways: (i) by changing the form of nutrient in the waterbody (nutrient cycling or internal loading), (ii) by bringing nutrient from another area to waterbody (external nutrient loading) and (iii) by exporting nutrient from aquatic body to terrestrial ecosystem [13].

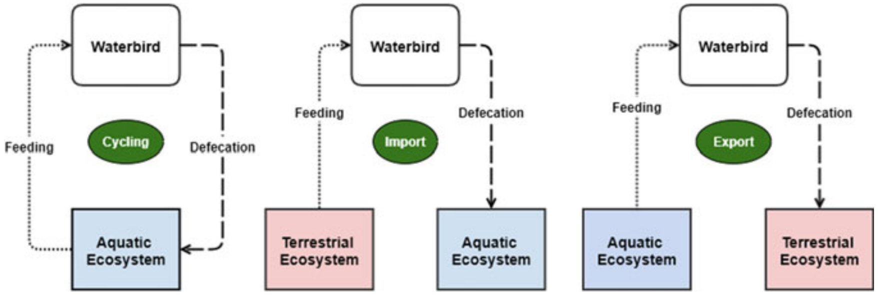


Fig. 1 The waterbirds as nutrient vectors. (i) Nutrient cycling: The waterbirds (like moorhen, coots, pygmy goose, jacana, grebe) which feeds and defecate in the same waterbody, converts one form of nutrient to another. (ii) The waterbirds (like most of the ducks, geese, swans) which feed in terrestrial habitats but roost at aquatic habitat, causes the net nutrient import to the aquatic ecosystem. (iii) The waterbirds (like herons, egrets, cormorants, kingfishers), which feed on aquatic ecosystem and roost on terrestrial ecosystem, leads to ultimate net nutrient export from aquatic to terrestrial ecosystem

Table 1 Summary of the literature reviewed

Subject	Literature
Estimation of nutrient loading	[3, 4, 14, 15]
Faecal nutrient analysis	[4, 16–24, 24–29]
Correlative study between waterbird and limnochemical variables	[7, 8, 30–32]
Bioassay study	[14, 29, 33, 34]

In this article, the faecal nutrient content of different waterbirds, the effect of faecal addition on water quality, different nutrient loading estimation method, factors affecting nutrient loading by waterbirds, ways to reduce the nutrient level in the waterbody and nutrient cycling pathways of waterbird faecal matter in the freshwater ecosystem are discussed with brief future perspective (Table 1).

2 Faecal Matter Analysis

The faecal composition of waterbird is highly variable because it depends on several factors (Table 2). Again, the measurement of nutrient concentration in waterbird droppings also varies considerably between studies due to different measurement approach by different workers [35]. The dry weight of single waterbird dropping lies between 0.2–0.4% of body weight [18]. The daily dropping production as dry mass for waterfowls was assumed between 2.25% [36] and 3.2% [18] of body weight for estimation of defecation rate of waterfowls, for which defecation rate is unknown. For domestic ducks, the defecation rate estimated at 3.8% of body weight [37]. The pH of waterfowl faecal matter lies between 5.0 to 8.5 [18]. The moisture content is between 66.69% to 92.5% [16, 18, 22]. The bird droppings consist of mainly two parts: labile and recalcitrant portion. The labile portion readily dissolved in water and increase the level of inorganic nutrient within 3 days of faecal addition [19]. The recalcitrant portion is resistant to decomposition and consist majority of the faecal matters including insoluble uric acid and undigested solid food materials [38]. As a result, most of this portion readily settles down to the sediment [39]. Inorganic nutrient from the recalcitrant portion started to appear in the water column after 15–30 days of addition as a result of decomposition [19]. As the guano decomposes, nitrogen (N) content is lost gradually as ammonia and guano become more phosphatic [40]. The faecal nutrient content of some waterbirds is given in Table 2.

Table 2 Body mass, feeding habit (FH), Total Nitrogen (TN) and Total Phosphorous (TP) of some waterbirds found in the literature. In the table, "H" indicates herbivorous feeding habit and "C" indicates carnivorous feeding habit

Scientific name	Body weight (kg) [41]	FH	TN (mg gDW^{-1})	TP (mg gDW^{-1})	Reference
Anas platyrhynchos	1.141 ± 0.07	H	52.3	17.4	[24]
			–	17.4	[22]
			53.1	8.5	[17]
			26.2 ± 0.8	13.2 ± 1.3	[21]
Anser albifrons	2.531 ± 0.219	H	32.6 ± 17.18	4.09 ± 3.60	[18]
A. anser	3.308 ± 0.298	H	23.633 ± 17.79	3.933 ± 2.85	[18]
A. brachyrhynchus	2.645 ± 0.184	H	19.563 ± 9.253	4.829 ± 2.646	[18]
			40 ± 2.6	3.6 ± 0.3	[19]
A. caerulescens	2.641 ± 0.211	H	26.48 ± 4.802	–	[16]
A. fabalis	2.771 ± 0.256	H	29.3	0.656	[18]
Branta canadensis	3.727 ± 0.314	H	5.72	14	[24]
			20.8	3.671	[18]
			5.7	14	[22]
			48 ± 22	15 ± 6	[4]
			11.9 ± 3.2	2 ± 0.7	[20]
B. leucopsis	1.687 ± 0.09	H	13.267 ± 1.963	6.031 ± 1.112	[18]
			11 ± 2.062	3.3 ± 0.225	[29]
Cygnus columbianus	2.19 ± 0.62	H	50.610 ± 21.462	8.885 ± 5.335	[27]
C. cygnus	9.35	H	11.05 ± 0.636	4.785 ± 4.172	[18]
C. olor	10.735 ± 0.775	H	15.8 ± 7.778	7.101 ± 0.278	[18]
Mareca penelope	0.772 ± 0.067	H	27	0.874	[18]
Dropping collected from heronry	–	C	88.160 ± 5.480	7.529 ± 1.572	[28]
Larus argentatus	1.085 ± 0.093	C	73.05	4.62	[42]
			8.4 ± 1.242	9.1 ± 1.329	[23]
L. canus	0.404 ± 0.04	C	70.25	4.24	[42]
L. fuscus	0.818 ± 0.06	C	68.58	4.33	[42]
L. marinus	1.659 ± 0.241	C	–	47.533 ± 10.486	[23]
L. michahellis	1.154 ± 0.095	C	44.2 ± 25.7	7.3 ± 4.1	[26]
L. ridibundus	0.284 ± 0.02	C	54.28	3.39	[42]
			72.4	78.6	[17]
Nordic Geese	–	H	–	8.85 ± 0.001	[25]
Gull	–	C	29.6 ± 3.7	16.2 ± 2.9	[21]
Heron	–	C	42.1 ± 6.7	114.7 ± 12.1	[21]
Cormorant	–	C	32.8 ± 0.2	143.2 ± 3.5	[21]

3 Effect of Nutrient Loading by Waterbird on Water Quality

Most of the study in this regard compared the effect of waterbird-born nutrient loading with other sources of nutrient entering the waterbody. The effect of nutrient loading on water quality varies between studies. Some studies found that nutrient contribution by waterbird may degrade water quality [4, 15, 25, 43], but other study found nutrient contribution by waterbird is negligible in comparison to other allochthonous sources [21, 37, 44] (Table 3).

The results of these studies are mostly incomparable, because of the variability of estimation method, varied anthropogenic pressure etc. Some studies only considered external input, while other studies include both internal and external input; and studies which considered both kinds of nutrient loading, actually overestimated the actual contribution by waterbirds. Again, instead of direct estimation as previously discussed, some study measured nutrient loading effect on water quality indirectly. Some study compared the water quality of inflow and outflow of a waterbird refuge [34, 38], some compared between water quality of high water-use pond and low water-use pond [31, 45] and some compared between changes in water quality indicators with changes in bird number [7, 8, 32]. Olson et al. found primary productivity at lake inlet was limited by both N and P, whereas output is limited by only N [34]. This indicates an important P subsidy by waterbird in a waterbody. Some studies found that waterbird can contribute a significant amount of nutrient and sustains lake productivity when other sources of nutrient were scarce [32, 46]. Nutrient loading by birds varies seasonally due to fluctuation in both waterbird number and nutrient content in food [15]. Tobiessen and Wheat found that nutrient added by waterbirds do not have any long-term and short-term effect on water quality [32]. Scherer et al. was also unable to find any short-term effect but hypothesized that bird dropping may have long-term effect in water quality degradation [37]. Some studies found a positive correlation of nutrient content in water and sediment with the bird number [7, 8, 31, 46], but other studies were unable to find this correlation for phosphorus [37, 39, 49]. Most studies were unable to find the change in chlorophyll-a level and Secchi depth due to nutrient loading by waterbird [29, 32, 37, 39, 49], except [33, 45]. In the bioassay experiment, it has been found that chl-a level increased after some days of faecal addition [22], similar observation was also found in the natural ecosystem [1]. Again, some study unable to find any correlation between fecal addition and chl-a, both in the natural environment [32] and also in mesocosm experiment [39]. Harris et al. found a significant difference ($p<0.05$) in chlorophyll-a and TP between high bird use pond and low bird use pond [45]. This lack of correlation may be the result of low arctic temperature or zooplankton grazing pressure on phytoplankton [29] or increased competition with macrophyte [24, 50].

Table 3 Summary of nutrient loading estimates found in literatures as Nitrogen (N) and Phosphospus (P). The relative contribution of guanotrophy in comparison to other nutrient sources indicated as %N and %P

Lake/Area of the study	Waterbird group concerned	Area (ha)	N (g m^{-2} yr^{-1})	P (g m^{-2} yr^{-1})	%N	%P	Reference
Bay Beach WLS, USA	Ducks and Geese	1.97	4.8	4.8	–	–	[45]
Baton, Rouge, USA	”	3.7	–	0.27	–	7.4	[44]
		1.4	–	0.43	–	7.9	[44]
Wintergreen lake, USA	”	15	1.87	0.59	27	70	[4]
Green Lake, USA	All	105	–	0.15	28.7	17.43	[37]
Grand Lieu, France	”	5150	0.11	0.04	0.7	2.4	[21]
			0.15	0.05	0.4	6.6	[21]
Gull Pond, USA	Gull	44	–	0.12	–	42	[23]
Middle Creek Reservoir, USA	Geese	162	3.67	0.52	–	–	[34]
Lake Waban, USA	”	360	–	0.008	–	0.4	[46]
Greenfield Lake, USA	All	37	0.14	0.23	–	–	[14]
Bosque del Apache NWR, USA	Geese	50.2	14.29	1.76	–	–	[15]
All inland lakes of Netherland	Herbivorous waterbirds	3.57×10^5	0.074–0.14	0.009–0.01	–	–	[3]
	Carnivorous waterbirds	”	0.026–0.065	0.012–0.016	–	–	[47]
Lake Arendsee, Germany	Geese	514	–	0.54	–	88	[25]
			–	0.336	–	92	[25]
Brown Moss, UK	Ducks and Geese	33	0.741	0.234	17	73	[43]
Swan Lake, Canada	Geese	5.48	–	0.26	–	74.6	[48]

4 Estimation of Nutrient Loading (NL) by Waterbird

As discussed earlier the estimation of NL is very important from a management perspective. This is impossible to measure NL directly from a field experiment. Commonly, nutrient loading is estimated by counting the number of birds and extrapolating it with the defecation rate and faecal nutrient content. To get an accurate measure of allochthonous nutrient loading by waterbirds, following factors must be considered: (i) duration of a day birds spent at the lake, (ii) the defecation rate, (iii) season and species-specific energy requirement of waterbirds, (iv) life history and behavioral pattern, and (v) species-specific feeding habit of waterbirds for the location of interest. The methods of estimating nutrient loading by waterbirds found in the different literature are given bellow chronologically.

4.1 Nutrient Loading Model by Harris et al. [45]

It was the pioneering approach towards estimating the NL by waterfowls. It was quantified by multiplying the number of bird-days (average monthly number of individuals counted per day multiplied by the number of days per month) with the nutrient content in faecal matter. The main drawback of this method is internal and external cycling of nutrient was not differentiated. Moreover, species variation was also neglected in this method.

4.2 Phosphorus Loading Model by Gremillion and Malone [44]

This nutrient loading model considered the allometric relationship between body mass, time spent by waterbirds in lake and defecation rate to estimate species-specific nutrient loading by waterfowls. The model is as follows:

$$NL = \sum_{S=1}^{n} N * DM * BM * RT * B \tag{1}$$

Here, N is a nutrient concentration in faecal matter, DM is dropping mass, BM is body mass, RT is residence time in the lake and B is a number of bird-days, S is number of species and n indicates the n-th species. The main drawback of this model is, the bioenergetics is neglected. Similar concept has been adopted in several studies with little modification for NL estimation [4, 14, 21, 23, 23, 34, 37, 46, 51, 52].

4.3 Bioenergetics Model by Post et al. [15]

This model includes the bioenergetics model [53] to estimate daily energy requirements of waterbird. The daily energy requirement (DER) includes energetic costs of existence metabolism, routine metabolism of free-living birds and flight. Existence metabolism, in turn, depends on body mass and average daily temperature. Nutrient uptake will depend on its energy requirement. If bird feeds on low energy containing food, they need more food and conversely, if they feed on high energy containing food item, they need less to eat [2]. So, eating high energy content feed results in lower faecal mass and consequently, lower nutrient loading. Again, they also included gut passage time (or retention time) (RT) to estimate in which site waterbirds were defecating. This is the first approach to differentiate between autochthonous and allochthonous input by waterbirds. This kind of model is more biologically sound, accurate and realistic but much complex than previous approaches.

4.4 Bioenergetics Model by Hahn et al. [3, 47]

Two different models were proposed for NL estimation for both carnivorous [47] and herbivorous waterbirds [3]: (i) for estimation from feeding data, (ii) estimation from defecation data. For carnivorous birds, the author further proposed three different models for chicks, non-breeding and breeding adults. In case of herbivorous waterbirds, the feeding model consists of RT, foraging time (T_f), species and season-specific proportion of energy obtained from terrestrial food (f_t), DER, the apparent metabolizable energy of food (AM) and elemental composition of food (X_{food}). The defecation model includes, f_t, RT, dropping rate (DrR), dropping mass (DrM) and elemental composition of dropping (X_{drop}). For non-breeding carnivorous waterbirds, the feeding model includes: fraction of total released nutrients contributed to waterbody (A), DER, AM, E and X_{food}. In defecation model, instead of X_{food}, X_{drop} is taken and an additional term (fixed relation between food intake and excrement considered. The model for breeding birds is same except an additional term nesting period (T_{nest}) is included. For chicks, the author considered for feeding model total energy requirement for entire chick-rearing period instead of DER, clutch size (CS), survival probability of chicks, nutrient needed for chick development (X_{syn}), X_{food}, AM and E. The excretion model differs similarly like non-breeding bird model from feeding model. Author used this model to estimate nutrient loading by waterbirds in different lakes in Netherland.

Nutrient loading outputs from both the feeding and excretion models were compared, but it was found that later model significantly underestimates nitrogen loading as compared with previous one. Two possible reasons were suggested by the author: (i) if waterfowls made more than one feeding-migration in a day and, (ii) excretory loss of a certain portion of nitrogen as volatile ammonia. This defecation model was further used for quantification of gull derived external nitrogen input [54]. The different nutrient loading studies as found in previous literature regarding guanotrophy is summarized in Table 3.

5 Variability in Nutrient Loading Estimation

As mentioned earlier, variability encountered in this review in nutrient loading estimates is mainly due to the different method followed in different studies. Some studies included internal loading which actually made overestimation of the nutrient loading [25, 45]. Estimation of nutrient loading in [45] relies on nutrient loading rate of domestic ducks. But, nutrient content in wild waterfowl faecal matter varies considerably from domestic waterfowls [4, 55]. Seasonal variation in N:P ratio in waterbird dropping was neglected in most of the studies [15]. Most of the studies neglected the nutrient demand during pre-migration fattening of bird except [3]. Moreover, the basis of the most of the nutrient loading study was bird count and it is the biggest problem in nutrient loading estimation. Increasing sample size by counting birds for multiple times during a month or by taking the averaged count data from different birders will provide a better estimation of number of birds [4]. In addition, the use of citizen science platform like eBird will be useful to obtain data about bird abundance for large lakes [56, 57].

6 Factors Affecting Allochthonous Nutrient Input by Waterbirds

In addition to food quality, allochthonous nutrient input by waterbirds is mainly dependent on the feeding and roosting behaviour of waterbird. These factors are described as follows: **i. Daily migration pattern:** Compared with more than one daily migration to foraging site, waterbirds making only one migration contributes less nutrient input to the roosting site [15]. **ii. Food quality:** Waterbirds fed with food of high and low energy value show similar elemental composition [58]. Again, low amount of high energy containing food can fulfil the daily energy need of waterbirds than low energy containing food [59]. As a result, low energy containing food produces more waste (similarly caused more nutrient loading) in compared to high energy containing food [15, 60]. **iii. Retention time:** Food retention time or gut passage time is proportional to nutrient loading in the lake [3, 15]. **iv. Temperature:** Energetic costs of foraging exceed gain during very low temperature, so waterbirds prefer to stay at the roosting site at very low temperature [61, 62]. Again existence metabolism increases with temperature [15]. So, bird needs more food with increasing temperature. **v. Wind speed:** At high wind speed waterbirds prefer single feeding migration in a day [15]. It was also reported that, high wind speed and rainfall results in lower feeding activity of some dabbling ducks [63, 64].

7 Ways to Reduce the Nutrient Level in the Lake

It is known that congregation of a huge number of waterbirds (feeds mostly outside the lake) in a small space may lead to reduced water quality [3, 4]. The nutrient loading in the lake can be reduced in several ways as discussed as follows: **1. Dredging:** Dredging is a common practice to improve water quality in wetlands by removing nutrient-rich sediments. But, it is not a sustainable way to improve water quality [45, 65]. **2. Shifting bird-use inside and between wetland:** Forcing waterbirds to shift between water bodies or spreading the waterbird in a lake will improve water quality [49, 61]. But this method will be unethical and disturbance to bird will reduce bird number in a waterbody. **3. Nutrient inactivation:** Another method to reduce the nutrient content in the water column is just to precipitate them chemically (with the application of alum, fly ash, PhoslockTM etc.) to sediment [20, 45]. But, an ecological consequence of this method needs more study. **4. Decreasing nutrient retention time in the lake:** Decreasing hydraulic retention time by increasing flushing rate is another way to prevent water quality degradation. But, it is always not possible in a wetland, especially in a dry area and closed lake. **5. Increasing ecological complexity:** The excess nutrient can be utilized in the wetland by increasing ecological complexity by introducing exotic species to wetland [45]. Eutrophic lakes have the potential to support a greater number of waterbirds (both species richness and abundance) [66, 67]. So, control of eutrophication should consider the trade-off between cleaner water and reduced biodiversity [31].

8 Fates of Guano Derived Nutrient in Aquatic Ecosystem

As discussed earlier, most of the bird dropping settles down quickly to bottom sediment and most of the feces not readily decomposed. After gradual decomposition, the organic nutrients converted into inorganic bioavailable forms. This decomposition process is enhanced by temperature and mixing of sediments. The flocculant sediment in lakes is disturbed by wind flow, mixing by benthic organism and waterbird etc., which further boost the mineralization process. The main element of bird droppings we discussed so far is the N and P, which affect the water chemistry. The nutrient cycling process of these elements is very different. Details of nitrogen and phosphorus cycle can be found in different literatures [68–70]. For the ease of discussion these processes are discussed below under the following two subsections.

8.1 Cycling of Nitrogen

As discussed earlier, the main nitrogenous waste in bird droppings is uric acid. In addition to this, bird feces contain some amount of ammonia and nitrogen contents

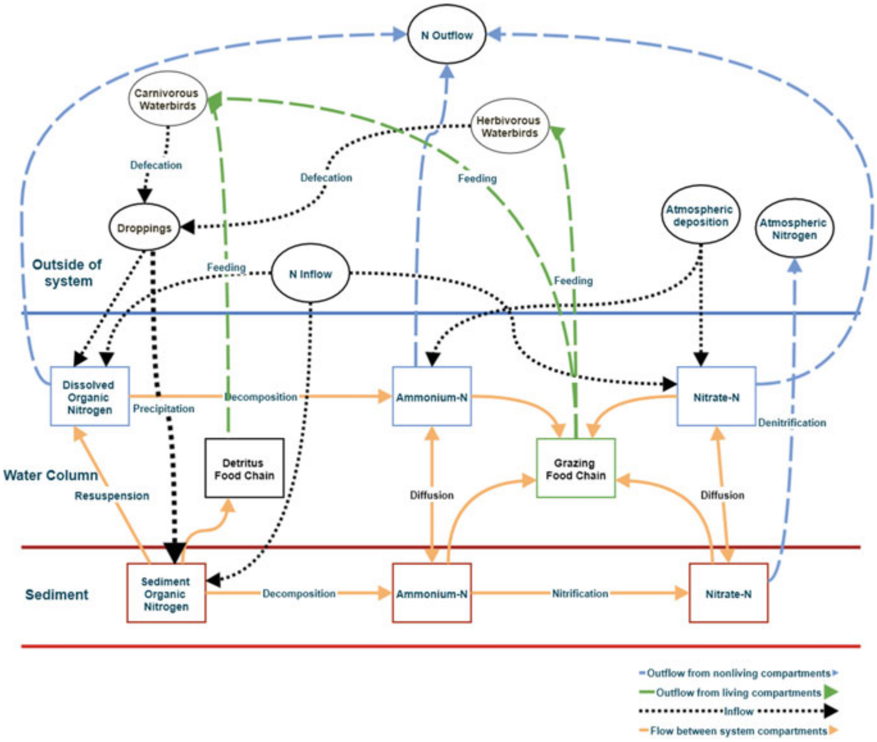


Fig. 2 Possible pathways of nitrogen cycling of waterbird-derived nutrient and nutrient from other sources in waterbody

of undigested food materials. Being mostly insoluble, waterbird dropping mostly settles down to sediment. Ammonia and a very little fraction of uric acid remain in the water column. Organic nitrogenous wastes in sediments decomposed to simpler form and ultimately converted to ammonium (NH_4^+) by the action of microbes in aerobic and anaerobic conditions. NH_4^+ diffused between water column and sediments as per concentration gradient. NH_4^+ in oxidized sediment layer converted to nitrite (NO_2^-) and ultimately to nitrate (NO_3^-) by nitrification. Some proportion of NH_4^+ is converted to NH_3 in high pH (more than 7.5). NH_3 is volatile and lost to atmosphere. Some amount of NH_4^+ and NO_2^- lost from the sediment by annamox. NO_2^- being very unstable readily converted to NO_3^- and present in very low amount in water column and sediments. A fraction of NO_3^- is further converted to atmospheric nitrogen and lost from wetland by denitrification. Plant and phytoplankton can utilize NH_4^+ and NO_3^- as food. Phytoplankton further utilized by zooplankton and which is further taken by fishes. Some portion of feces recycled through detritus pathway. In case of shallow lakes or ponds, most of the ponds are dominated by macrophytes. Macrophytes are able to uptake nitrogen inorganic nitrogen from the sediments (Fig. 2).

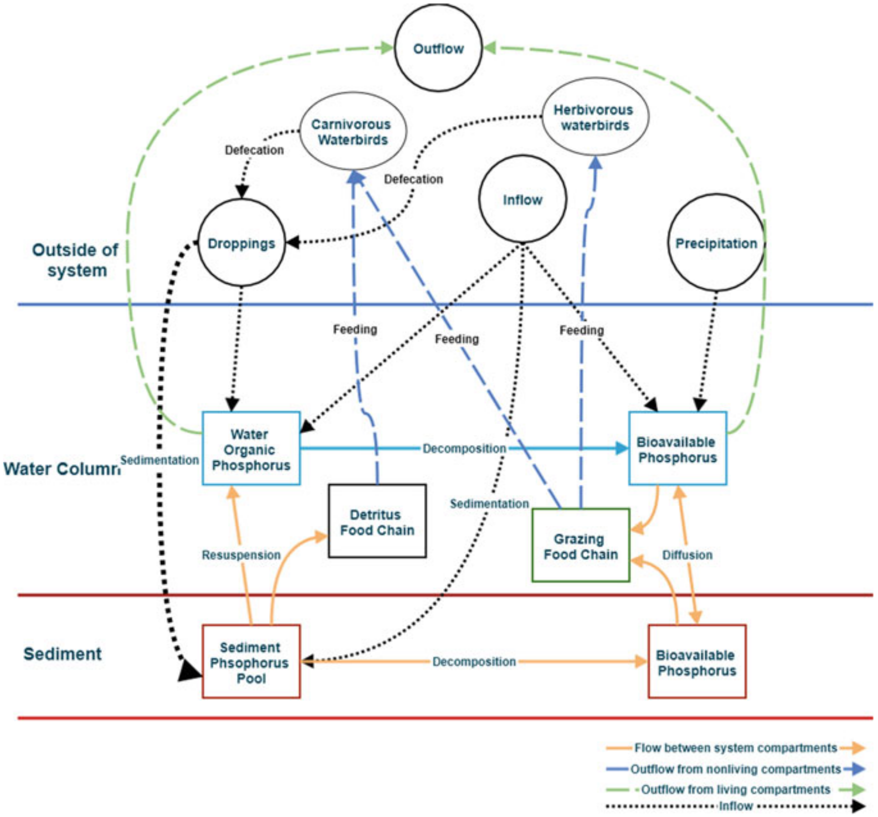


Fig. 3 Possible pathways of phosphorous cycling of waterbird-derived nutrient and nutrient from other sources in waterbody

8.2 *Cycling of Phosphorous*

Unlike nitrogen cycle, phosphorus cycle is not operated by redox potential in sediment. As stated above, most of the droppings settle down in sediments, most of phosphorus goes to sediment phosphorus pool. Sediment organic phosphorus also can resuspend to water. Organic phosphorus after degradation forms inorganic phosphorus. Inorganic phosphorus is not always bioavailable, only bioavailable form is orthophosphate. The inorganic phosphorus in lake remains in different form: bioavailable orthophosphate, phosphorus adsorbed in organic colloids, phosphorus adsorbed in metals in sediments and phosphorus bound to organic matter in sediments. Because of this nature of phosphorus, phosphorus is limiting in most of the waterbody and unlike nitrogen do not show atmospheric cycling instead shows sedimentary cycling. When temperature increases, the hypolimnion of lake become anoxic and this anoxic environment promotes release of orthophosphate from sediments. Again, release of bioavailable orthophosphate is enhanced by the disturbance in sediment. Autotroph can uptake bioavailable orthophosphate from sediment and phosphorus from autotroph then goes to grazing food chain. Again, sediment phosphorus can also be cycled through detritus food chain. As the guano decomposed, sediment lose the N as it is more soluble in water, results in phosphatic sediment (Fig. 3).

9 Conclusion

The relative importance of waterbird-derived nutrient depends on the land-use pattern. Lake which received a large amount of nutrient through drainage, non-point sources have little effect of waterbird input. But, in the arid area or close lake with no inlet and outlet or in oligotrophic lakes waterbirds contributes a significant amount of nutrients and in most cases, support the productivity of that lake. Estimates of nutrient loading by waterbirds have mostly been done with data from existing literature. But, development in technology, changes in fertilizer application etc. have changed the agricultural practice over the decades, which in turn have affected elemental concentration and energy quality of food of waterbird. So, using historical loading rates may lead to bias towards old studies and may give inaccurate results. Again, species composition of waterbird and nutrient content in food vary geographically. The retention time of food also varies with a different diet of waterbird and the diet compositions also show geographical variations. No study has so far been done in the tropics where the physical conditions are very much different from temperate regions. The hot tropical region may increase energy demand of bird by increasing energy of existence (a function of temperature). So, using nutrient loading or defecation rate values form the studies of temperate origin may give wrong results if applied in tropics. As mentioned earlier, ducks are neglected globally. But the large congregation of ducks may result in considerable amount of external load to the lake. Data unavailability of defecation rates and retention time of species like Whistling-Ducks, Pygmy-Goose have become a problem for estimating nutrient loading for this So, in future species, site and season-specific studies are required to understand more deeply and accurately the external nutrient inputs to the lake by waterbirds. Little study has been done on ecosystem approach of guanotrophy. A recent review concluded regarding the need of experimental ecosystem approach [35]. There are many studies exist regarding what effect waterbird put on the aquatic ecosystem. But, there is a little study performed about how question except for some isotopic study [33] and bioassay study [34]. Process-based dynamic model approach, as has been developed for semiaquatic waders at Okefenokee marshland [71], can be a best answer for the how question. This kind of model can be also useful in management perspective.

Acknowledgements The authors greatly acknowledge the financial support provided by University Grants Commission (UGC), New Delhi, India under the scheme of UGC-NET JRF fellowship. The author is thankful to Dr. Joyita Mukherjee, Dr. Fahad Al Basir, Dr. Arnab Banerjee, Nabyendu Rakshit and Swagata Sinha for their kind suggestions. The authors are also thankful to anonymous reviewer for important suggestions for the improvement of the article.

References

1. Leentvaar, P.: Observations in guanotrophic environments. Hydrobiologia **29**, 441489 (1967). https://doi.org/10.1007/BF00189906
2. Kerbes, R.H., Kotanen, P.M., Jefferies, R.L.: Destruction of wetland habitats by Lesser Snow Geese: a keystone species on the West Coast of Hudson Bay. J. Appl. Ecol. **27**, 242258 (1990). https://doi.org/10.2307/2403582
3. Hahn, S., Bauer, S., Klaassen, M.: Quantification of allochthonous nutrient input into freshwater bodies by herbivorous waterbirds. Freshw. Biol. **53**, 181193 (2008). https://doi.org/10.1111/j.1365-2427.2007.01881.x
4. Manny, B.A., Johnson, W.C., Wetzel, R.G.: Nutrient additions by waterfowl to lakes and reservoirs: predicting their effects on productivity and water quality. Hydrobiologia **279**, 121132 (1994). https://doi.org/10.1007/978-94-011-1128-7_12
5. Mathis, B.J., Kevern, N.R.: Distribution of mercury, cadmium, lead and thallium in a eutrophic lake. Hydrobiologia **46**, 207222 (1975). https://doi.org/10.1007/BF00043141
6. Alderisio, K.A., DeLuca, N.: Seasonal enumeration of fecal coliform bacteria from the feces of ring-billed gulls (*Larus delawarensis*) and Canada geese (*Branta canadensis*). Appl. Environ. Microbiol. **65**, 562830 (1999)
7. Chatterjee, A., Adhikari, S., Mukhopadhyay, S.K.: Effects of waterbird colonization on limnochemical features of a natural wetland on buxa tiger reserve, India. During Wintering Period. Wetlands **37**, 177190 (2017). https://doi.org/10.1007/s13157-016-0864-2
8. Roy, U.S., Roy Goswami, A., Aich, A., Mukhopadhyay, S.K.: Changes in densities of waterbird species in Santragachi Lake, India: Potential effects on limnochemical variables. Zool. Stud. **50**, 7684 (2011)
9. Friend, M.: Evolving changes in diseases of Waterbirds. In: Boere, G.C., Galbraith, C.A., Stroud, D.A. (eds.) Waterbirds Around the World: A Global Overview of the Conservation, Management and Research of the Worlds Waterbird Flyways, p. 412417. The Stationery Office Limited, Edinberg, UK (2006)
10. Wobeser, G.A.: Diseases of Wild Waterfowl, 2nd edn. Springer, US, Boston, MA (1997)
11. Krylov, A.V., Kulakov, D.V., Chalova, I.V., Tselmovich, O.L.: The effect of vital activity products of hydrophilic birds and the degree of overgrowth on zooplankton in experimental microcosms. Inl. Water Biol. **6**, 114123 (2013). https://doi.org/10.1134/S1995082913020065
12. Krylov, A.V., Kulakov, D.V., Tsvetkov, A.I., Papchenkov, V.G.: Effect of atmospheric precipitation and the abundance of semiaquatic bird colonies on zooplankton in the littoral of a small high-trophic lake. Biol. Bull. **41**, 862868 (2014). https://doi.org/10.1134/S1062359014100069
13. Polis, G.A., Anderson, W.B., Holt, R.D.: Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. Ann. Rev. Ecol. Syst. **28**, 289316 (1997). https://doi.org/10.1146/annurev.ecolsys.28.1.289
14. Mallin, M.A., McIver, M.R., Wambach, E.J., Robuck, A.R.: Algal blooms, circulators, waterfowl, and eutrophic Greenfield Lake. North Carolina. Lake Reserv. Manag. **32**, 168181 (2016). https://doi.org/10.1080/10402381.2016.1146374
15. Post, D.M., Taylor, J.P., Kitchell, J.F., et al.: The role of migratory waterfowl as nutrient vectors in a managed wetland. Conserv. Biol. **12**, 910920 (1998). https://doi.org/10.1111/j.1523-1739.1998.97112.x
16. Bazely, D.R., Jefferies, R.L.: Goose Faeces: A Source of Nitrogen for Plant Growth in a Grazed Salt Marsh. J. Appl. Ecol. **22**, 693703 (1985)

17. Gwiazda, R.: Contribution of water birds to nutrient loading to the ecosystem of mesotrophic reservoir. Ekol. Pol. **44**, 289297 (1996)
18. Kear, J.: The agricultural importance wild goose droppings. Wildfowl **14**, 7277 (1962)
19. Liu, Y., Hefting, M.M., Verhoeven, J.T.A., Klaassen, M.: Nutrient release characteristics from droppings of grass-foraging waterfowl (*Anser brachyrhynchus*) roosting in aquatic habitats. Ecohydrology **7**, 12161222 (2014). https://doi.org/10.1002/eco.1454
20. Lrling, M., van Oosterhout, F.: Case study on the efficacy of a lanthanum-enriched clay (Phoslock) in controlling eutrophication in Lake Het Groene Eiland (The Netherlands). Hydrobiologia **710**, 253263 (2013). https://doi.org/10.1007/s10750-012-1141-x
21. Marion, L., Clergeau, P., Brient, L., Bertru, G.: The importance of avian-contributed nitrogen (N) and phosphorus (P) to Lake Grand-Lieu. France. Hydrobiologia **279280**, 133147 (1994). https://doi.org/10.1007/BF00027848
22. Pettigrew, C.T., Hann, B.J., Goldsborough, L.G.: Waterfowl feces as a source of nutrients to a prairie wetland: Responses of microinvertebrates to experimental additions. Hydrobiologia **362**, 5566 (1997). https://doi.org/10.1023/A:1003167219199
23. Portnoy, J.W.: Gull contributions of phosphorus and nitrogen to a Cape Cod kettle pond. Hydrobiologia **202**, 6169 (1990). https://doi.org/10.1007/BF00027092
24. Purcell, S.L.: The Significance of Waterfowl Feces as a Source of Nutrients to Algae in a Prairie Wetland. University of Manitoba, Winnipeg, Thesis Department (1999)
25. Rnicke, H., Doerffer, R., Siewers, H., et al.: Phosphorus input by nordic geese to the eutrophic Lake Arendsee, Germany. Fundam. Appl. Limnol./Arch. fr Hydrobiol **172**, 111119 (2008). https://doi.org/10.1127/1863-9135/2008/0172-0111
26. Signa, G., Mazzola, A., Vizzini, S.: Effects of a small seagull colony on trophic status and primary production in a Mediterranean coastal system (Marinello ponds, Italy). Estuar. Coast. Shelf. Sci. **111**, 2734 (2012). https://doi.org/10.1016/J.ECSS.2012.06.008
27. Somura, H., Masunaga, T., Mori, Y., et al.: Estimation of nutrient input by a migratory bird, the Tundra Swan (*Cygnus columbianus*), to winter-flooded paddy fields. Agric. Ecosyst. Environ. **199**, 19 (2015). https://doi.org/10.1016/j.agee.2014.07.018
28. Telesford-Checkley, J.M., Mora, M.A., Grant, W.E., et al.: Estimating the contribution of nitrogen and phosphorus to waterbodies by colonial nesting waterbirds. Sci. Total Environ. **574**, 13351344 (2017). https://doi.org/10.1016/j.scitotenv.2016.08.043
29. van Geest, G.J., Hessen, D.O., Spierenburg, P., et al.: Goose-mediated nutrient enrichment and planktonic grazer control in arctic freshwater ponds. Oecologia **153**, 653662 (2007). https://doi.org/10.1007/s00442-007-0770-7
30. Gunaratne, A.M., Jayakody, S., Amarasinghe, U.S.: Ornithological eutrophication as a source of allochthonous nutrient enrichment in Anavilundawa reservoir. Sri Lanka. Int. Rev. Hydrobiol. **100**, 151157 (2015). https://doi.org/10.1002/iroh.201501804
31. Hoyer, M.V., Canfield, D.E.: Bird abundance and species richness on Florida lakes: influence of trophic status, lake morphology, and aquatic macrophytes. Hydrobiologia **279**, 107119 (1994). https://doi.org/10.1007/BF00027846
32. Tobiessen, P., Wheat, E.: Long and short term effects of waterfowl on Collins Lake, an urban lake in upstate New York. Lake Reserv. Manag. **16**, 340344 (2000). https://doi.org/10.1080/07438140009354241
33. Kitchell, J.F., Schindler, D.E., Herwig, B.R., et al.: Nutrient cycling at the landscape scale: The role of diel foraging migrations by geese at the Bosque del Apache National Wildlife Refuge. New Mexico. Limnol. Oceanogr. **44**, 828836 (1999). https://doi.org/10.4319/lo.1999.44.3_part_2.0828
34. Olson, M.H., Hage, M.M., Binkley, M.D., Binder, J.R.: Impact of migratory snow geese on nitrogen and phosphorus dynamics in a freshwater reservoir. Freshw. Biol. **50**, 882890 (2005). https://doi.org/10.1111/j.1365-2427.2005.01367.x
35. Dessborn, L., Hessel, R., Elmberg, J.: Geese as vectors of nitrogen and phosphorous to freshwater systems. Inl Waters **6**, 111122 (2016). https://doi.org/10.5268/IW-6.1.897
36. Sanderson, G.C., Anderson, W.L.: Waterfowl studies at Lake Sangchris, 1973–1977. Illinois Nat. Hist. Surv. Bull. **32**, 656690 (1978)

37. Scherer, N.M., Gibbons, H.L., Stoops, K.B., Muller, M.: Phosphorus Loading of an Urban Lake by Bird Droppings. Lake Reserv. Manag. **11**, 317327 (1995). https://doi.org/10.1080/07438149509354213
38. Brandvold, D.K., Popp, C.J., Brierley, J.A.: Waterfowl refuge effect on water quality: II. Chemical and physical parameters. Water Pollut. Control Fed. **48**, 680687 (1976)
39. Unckless, R.L., Makarewicz, J.C.: The impact of nutrient loading from Canada Geese (*Branta canadensis*) on water quality, a mesocosm approach. Hydrobiologia **586**, 393401 (2007). https://doi.org/10.1007/s10750-007-0712-8
40. Copeman, P.R.V.D.R., Dillman, F.J.: Changes in the composition of guano during storage. J. Agric. Sci. **27**, 178187 (1937)
41. Dunning, J.B.: Body masses of birds of the world. In: Dunning, J.B. (ed.) CRC Handbook of Avain Body Masses, vol. 672, 2nd edn. CRC Press, Taylor and Francis Group, Boca Raton (2008)
42. Gould, D.J., Fletcher, M.R.: Gull droppings and their effects on water quality. Water Res. **12**, 665672 (1978). https://doi.org/10.1016/0043-1354(78)90176-8
43. Chaichana, R., Leah, R., Moss, B.: Birds as eutrophicating agents: A nutrient budget for a small lake in a protected area. Hydrobiologia **646**, 111121 (2010). https://doi.org/10.1007/s10750-010-0166-2
44. Gremillion, P.T., Malone, R.F.: Waterfowl Waste As a Source of Nutrient Enrichment in Two Urban Hypereutrophic Lakes. Lake Reserv. Manag. **2**, 319322 (1986). https://doi.org/10.1080/07438148609354650
45. Harris, H., Ladowski, J., Donald, J.: Water-Quality Problems and Management of an Urban Waterfowl Sanctuary. J. Wildl. Manag. **45**, 501507 (1981)
46. Moore, M., Zakova, P., Shaeffer, K., Burton, R.: Potential effects of Canada Geese and climate change on phosphorus inputs to suburban lakes of the Northeastern USA. Lake Reserv. Manag. **14**, 5259 (1998)
47. Hahn, S., Bauer, S., Klaassen, M.: Estimating the contribution of carnivorous waterbirds to nutrient loading in freshwater habitats. Freshw. Biol. **52**, 24212433 (2007). https://doi.org/10.1111/j.1365-2427.2007.01838.x
48. Nrnberg, G.K., LaZerte, B.D.: Trophic state decrease after lanthanum-modified bentonite (Phoslock) application to a hyper-eutrophic polymictic urban lake frequented by Canada geese (*Branta canadensis*). Lake. Reserv. Manag. **32**, 7488 (2016). https://doi.org/10.1080/10402381.2015.1133739
49. Gwiazda, R., Wonica, A., ozowski B, et al.: Impact of waterbirds on chemical and biological features of water and sediments of a large, shallow dam reservoir. Oceanol. Hydrobiol. Stud. **43**, 418426 (2014). https://doi.org/10.2478/s13545-014-0160-9
50. Sndergaard, M., Moss, B.: Impact of submerged macrophytes on phytoplankton in shallow freshwater lakes. In: Jeppesen, E., Sndergaard, M., Sndergaard, M., Christoffersen, K. (eds.) The Structuring Role of Submerged Macrophytes in Lakes. Ecological Studies (Analysis and Synthesis), p. 115132. Springer, New York (1998)
51. Andersen, D.C., Sartoris, J.J., Thullen, J.S., Reusch, P.G.: The effects of bird use on nutrient removal in a constructed wastewater-treatment wetland. Wetlands **23**, 423435 (2003). https://doi.org/10.1672/17-20
52. Comber, S.D.W., Smith, R., Daldorph, P., et al.: Development of a chemical source apportionment decision support framework for lake catchment management. Sci. Total Environ. **622623**, 96105 (2018). https://doi.org/10.1016/J.SCITOTENV.2017.11.313
53. Kendeigh, S.C., Dolnik, V.R., Gavrilov, V.M.: Avian Energetics. In: Pinowski, J., Kendeigh, S.C. (eds.) Grainivorous Birds in Ecosystems, p. 127204. Cambridge University Press, London (1977)
54. Winton, R.S., River, M.: The biogeochemical implications of massive gull flocks at landfills. Water Res. **122**, 440446 (2017). https://doi.org/10.1016/j.watres.2017.05.076
55. Paloumpis, A.A., Starrett, W.C.: An Ecological Study of Benthic Organisms in Three Illinois River Flood Plain Lakes. Am. Midl. Nat. **64**, 406 (1960). https://doi.org/10.2307/2422672

56. Callaghan, C.T., Gawlik, D.E.: Efficacy of eBird data as an aid in conservation planning and monitoring. J. F. Ornithol. **86**, 298304 (2015). https://doi.org/10.1111/jofo.12121
57. Sullivan, B.L., Aycrigg, J.L., Barry, J.H., et al.: The eBird enterprise: an integrated approach to development and application of citizen science. Biol. Conserv. **169**, 3140 (2014). https://doi.org/10.1016/j.biocon.2013.11.003
58. Perry, M.C., Kuenzel, W.J., Williams, B.K., Serafin, J.A.: Influence of Nutrients on Feed Intake and Condition of Captive Canvasbacks in Winter. J. Wildl. Manag. **50**, 427434 (1986). https://doi.org/10.2307/3801099
59. Paulus, S.L.: Time-activity budgets of nonbreeding Anatidae: a review. In: Weller, M.W. (ed.) Waterfowl in Winter: Selected Papers From Symposium and Workshop Held in Galveston, p. 135152. 7–10 January 1985 [1988]. University of Minnensota Press, Texas (1988)
60. Bedard, J., Gauthier, G.: Comparative energy budgets of Greater Snow Geese Chen caerulescens atlantica staging in two habitats in spring. Ardea **77**, 320 (1989)
61. Frederick, R.B., Klaas, E.E.: Resource Use and Behavior of Migrating Snow Geese. J. Wildl. Manag. **46**, 601 (1982). https://doi.org/10.2307/3808550
62. Ravelling, D.G., Crews, W.E., Klimstra, W.E.: Activity Patterns of Canada Geese during Winter. Wilson Bull. **84**, 278295 (1972)
63. Miller, M.R.: Time budgets of northern pintails wintering in the Sacramento Valley. California. Wildfowl **36**, 5364 (1985)
64. Paulus, S.L.: Activity Budgets of Nonbreeding Gadwalls in Louisiana. J. Wildl. Manag. **48**, 371380 (1984). https://doi.org/10.2307/3801168
65. Peimin, P., Guoxiang, W., Chunhua, H., et al.: Can We Control Lake Eutrophication by Dredging? J. Lake Sci. **12**, 269279 (2000). https://doi.org/10.18307/2000.0312
66. Murphy, S.M., Kessel, B., Vining, L.J.: Waterfowl Populations and Limnologic Characteristics of Taiga Ponds. J Wildl. Manag. **48**, 11561163 (1984). https://doi.org/10.2307/3801776
67. Nilsson, S.G., Nilsson, I.N.: Breeding Bird Community Densities and Species Richness in Lakes. Oikos **31**, 214221 (1978). https://doi.org/10.2307/3543565
68. Mitsch, W.J., Gosselink, J.G.: Wetlands. Fourth. Wiley, New Jersey (2007)
69. Reddy, K.R., DeLaune, R.D.: Biogeochemistry of Wetlands: Science and Applications. CRC Press, Taylor and Francis Group, Boca Raton (2008)
70. Wetzel, R.G.: Limnology: Lake and Reservoir Ecosystem. Academic, California (2001)
71. Oliver, D., Legovi, T.: Okefenokee marshland before, during and after nutrient enrichment by a bird rookery. Ecol. Model. **43**, 195223 (1988). https://doi.org/10.1016/0304-3800(88)90004-X

On the Volatility of High Frequency Stock Index Based on SV Model of MCMC

Y. X. Zheng, Y. H. Zhang and X. H. Lu

Abstract By using of 5-min high-frequency data in CSI 300 index stock high-frequency data from 15th Jan. 2018 to 5th Mar. 2018, basing on Bayesian Analysis simulated by MCMC, this paper adopts the Stochastic Volatility model to do empirical researches on China's stock market and utilizes DIC criterion to do model fitting comparison. The result shows that China's stock market has higher volatility persistence, and the fitting effect for SV model to 5-min high-frequency data is better than the low-frequency data, and the standard stochastic volatility model (SV-N) is more suitable for high frequency-data of 5-min than the heavy-tail finance stochastic volatility model (SV-T).

Keywords SV model · Gibbs sampling · Bayesian analysis · Monte Carlo method

1 Introduction

Volatility is one of the most important characteristics of financial market, it is directly related to market uncertainty and risk, and is one of the most concise and effective indicators of financial market quality and efficiency [18]. We know that the degree of change between yields is called volatility, but it is not observable. Therefore, how to effectively describe the dynamic behavior of financial market fluctuations has become the focus research on financial market.

Y. X. Zheng
Technology Support Group, Migu Culture Technology Com., Ltd, Beijing 100032, China

Y. X. Zheng · Y. H. Zhang (✉)
Department of Mathematics, Beijing Technology and Business University,
Beijing 100048, China
e-mail: zhangyanhui@th.btbu.edu.cn

X. H. Lu
School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100026, China

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_23

With the development of computer and communication technology, high frequency data can better capture real-time information in financial market and reflect market changes more comprehensively. Andersen and Bollerslev (1998) proposed the measurement of volatility based on the "day sum squared rate of return" and created a precedent for the study of volatility using intraday high frequency data [3, 14]. Therefore, studying high-frequency data in the financial market helps us to understand the micro-structural characteristics of the financial market, and has important applications in risk management and finance engineering, such as derivatives pricing, and has become an advantageous tool for market surveillance agencies. As a fast-growing field, there are some typical characteristics for the empirical analysis of high-frequency data in the US stock market such as volatility day-type trend, calendar effect, kurtosis and first-order negative correlation of price series [4.5]. The research on the volatility of China's stock high-frequency data mainly focuses on the non-parametric estimation of integral volatility and modeling and forecasting of volatility [5, 8, 15, 18], but there are still many problems not to be solved.

In the research article [16, 17], which eliminates the day pattern and establishes the ACD Model by using of high frequency data of CSI 300 stock index. They also discusses the information about the transmission mechanism for the duration of trading volume. By using ACD model, studies of high-frequency data of one minute before and after, the implementation of the fuse mechanism analyzes the impact of the fuse mechanism on China's A-share market. Inspired by these, this paper will empirically analyze the 5-min high frequency data of the CSI 300 stock index from 15th January 2018 to 5th March 2018, and study the volatility characteristics of the micro market returns. We also analyze the effect on high frequency data of SV model. Bayesian inference for standard stochastic volatility (SV-N) model and heavy-tail stochastic volatility (SV-T) model will be deduced combing with R and WinBUGS, and the model will be fitted by using Markov Monte Carlo(MCMC) method. Then we analyze and compare the fit of the two models under different K-lines. This paper will fit the low-frequency daily rate data and the 5-min high-frequency rate data with SV model. The results show that the SV model has a better fitting effect on the 5-min high-frequency data.

2 Model Principle

Let P_t be the closing price of the stock index on the tth trading day, $y_t = \ln P_{t+1} - \ln P_t$ be the daily return sequence. The standard stochastic volatility (SV-N) model is:

$$\begin{cases} y_t = \exp(\theta_t/2)\varepsilon_t, & \varepsilon_t \sim i.i.d.N(0,1), t = 1, 2, \ldots, n, \\ \theta_t = \mu + \phi(\theta_{t-1} - \mu) + \eta_t, & \eta_t \sim i.i.d.N(0, \sigma^2), t = 1, 2, \ldots, n, \end{cases} \tag{1}$$

where ε_t is the white noise interference term that obeys the standard normal distribution; θ_t is the logarithmic volatility, and is a Gaussian AR(1) process that obeys the persistence parameter ϕ, reflecting the impact of current volatility on future volatility, the SV model is stable in covariance when $\mid \phi \mid < 1$; η_t is the volatility disturbance level of independent and identical distribution, obeying $(0, \sigma^2)$ normal distribution; moreover ε_t and η_t are white noise sequences mutually independent.

If we look y_t as t distribution with ω degrees of freedom in the SV-N model (2.1), we will get a heavy-tail stochastic volatility model (SV-T for short), here the perturbation ε_t obeys mean is 0, a normalized t-distribution with ω degree of freedom variance is 1, which has the ability to capture the high peak and heavy-tail of the actual financial gain sequence.

Heavy-tail finance stochastic volatility model (SV-T) is

$$\begin{cases} y_t = \exp(\theta_t/2)\varepsilon_t, & \varepsilon_t \sim i.i.d.t(0, 1, \omega), t = 1, 2, \ldots, n, \\ \theta_t = \mu + \phi(\theta_{t-1} - \mu) + \eta_t, & \eta_t \sim i.i.d.t(0, \sigma^2), t = 1, 2, \ldots, n. \end{cases} \tag{2}$$

We will apply Bayesian inference method to deal with the volatility of SV model. The posterior density distributions of the model parameters will be discussed separately.

Set $\phi = 2\phi - 1$. The prior distribution of the parameters μ, ϕ_1, σ and θ_0 in SV-N is [10]

$$\phi_1 \sim Be(20, 1.5); \sigma^2 \sim IGa(2.5, 0.025); \mu \sim N(0, 100); \theta_0 \sim N(\mu, \sigma^2), \tag{3}$$

The prior distribution of SV-T model with an extra ω

$$\omega \sim \chi(8), freedom : I(4, 40). \tag{4}$$

Because the calculation of posterior distribution of the SV model will use high-dimensional integrals, and the Markov chain can converge to a stable probability distribution by shifting the probability matrix, we will simulate the high-dimensional probability distribution by means of MCMC with the steady distribution characteristics of the Markov chain. When the Markov chain passes the preburning stage to eliminate the influence of the initial parameters and reaches the stationary state, each state transition can generate a sample of the distribution to be simulated. Gibbs sampling can decompose a high-dimensional estimation problem into several low dimensional problems by using the conditional distribution of all parameters.

The Gibbs sampling algorithm [11] is as follows. Given starting point $\theta(0) = (\theta_1^0, \theta_2^0, \ldots \theta_d^0)$, $i = 0, 1, \ldots t$, iterate the following steps

(1) Extract random number $\theta_1^{(i+1)}$ from $p(\theta_1 \mid \theta_2^{(i)}, \ldots, \theta_d^{(i)}, y)$;

(2) Extract random number $\theta_2^{(i+1)}$ from $p(\theta_2 \mid \theta_1^{(i)}, \ldots, \theta_d^{(i)}, y)$;

(3) Extract random number $\theta_3^{(i+1)}$ from $p(\theta_3 \mid \theta_1^{(i)}, \ldots, \theta_d^{(i)}, y)$;

$\ldots$

(4) Extract random number $\theta_d^{(i+1)}$ from $p(\theta_d \mid \theta_1^{(i)}, \ldots, \theta_d - 1^{(i)}, y)$.

Then we get a sequence of Gibbs with length t which is $\theta^{(t)} = (\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(t)})$, after m iterations of annealing, the Gibbs sequence can converge to a smooth distribution that independent with the initial value. After the pre-burning sample is discarded, the Markov chain's implementation value is obtained, that is the sequence $X^{(m+1)}, \ldots, X^{(t)}$.

As there are many unknown variables and observations in SV model, we will choose the DIC criterion to compare and analyze the degree of fitting of the two models. The DIC criterion [7] consists of two parts, one is $\overline{D}$ representing the good and bad of the model fitting data, dispersion of posterior mean. Another one is PD that to measure the complexity of the model.

$$DIC = \overline{D} + P_D; \tag{5}$$

$$\overline{D} = E_{\theta\mid y}[-\ln L(y|\theta)]; \tag{6}$$

$$P_D = E_{\theta|y}[-\ln L(y|\overline{\theta})]. \tag{7}$$

where $\overline{\theta}$ is the posterior mean of θ, $L(y \mid \overline{\theta})$ is the likelihood function under the condition of known parameters and fluctuation mean. $E_{\theta|y}[X]$ denotes the mean under the posterior distribution g$(\theta \mid y)$[12].

3 Empirical Analysis

We will simulate the volatility of China's stock market by using 5-min high-frequency data of the CSI 300 stock index from 15th Jan. 2018 to 5th Mar. 2018 in normal trading time. The data comes from the Great wisdom 365 software, and the sample size is 1488. The CSI 300 stock index has a wide sampling and strong representativeness, which can more accurately reflect the general trend of China's stock price. We analyze the statistical characteristics of the series of returns and study the volatility aggregation effect of China's stock market by R. we knew that the yield sequence is generally stable from its logarithmic yield from Fig. 1.

Table 1 shows the basic statistical characteristics of the sequence y_t. The yield variability of the sequence y_t is not significant and the trend is relatively stable. The left-handed for the distribution of yields shows that the stock has risen slowly and dropped sharply. The peak value of the sequence is much larger than that of the normal distribution, indicating that the sequence distribution has the proterty of Peak and heavy-tail compared to the normal distribution. The J-B statistic [4] shows that the sequence y_t does not obey the normal distribution. From the Box.test value,

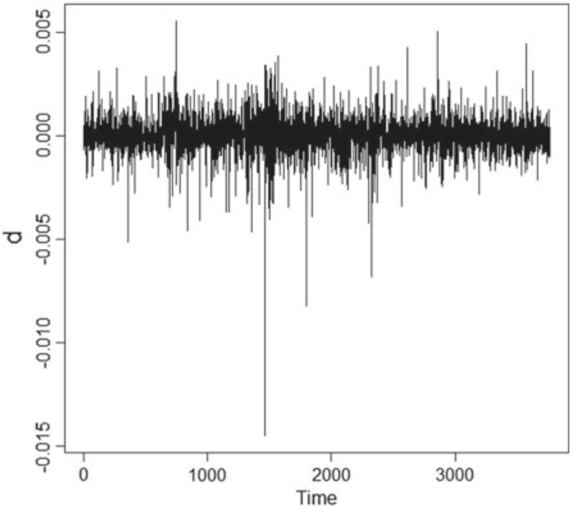


Fig. 1 CSI 300 logarithmic return rate chart

Table 1 Basic statistical characteristics of the sequence

Stock index	Mean	Standard deviation	Skewness	Kurtosis	J-B	Box.test
CSI 300	$2.61 * 10^{-5}$	$9.685 * 10^{-5}$	–1.21676	20.43131	$2.2 * 10^{-16}$	$2.103 * 10^{-3}$

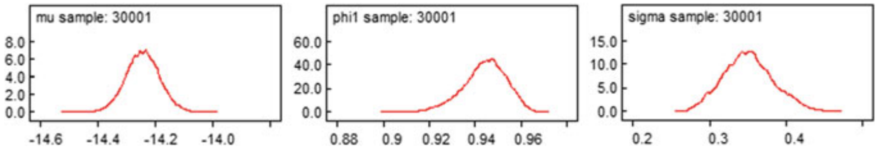


Fig. 2 Estimation of parameters posterior distribution density for the 5-min SV-N model of CSI 300

we can reject the original hypothesis sequence is irrelevant, so the sequence has autocorrelation.

In this paper, we estimate Bayesian parameters of two models with WinBUGS. It takes 10000 iterations for each parameter to be estimated as a preburning sample to ensure that Gibbs is close enough to the random sample from the joint distribution and the convergence of the parameters. After 30001 iterations, the model was simulated and estimated to obtain the following posterior distribution density function graph.

Parameter μ has symmetry in these two models from Figs. 2 to 3, ϕ is slightly left-biased. Parameter σ is symmetric in SV-N model. σ is slightly right-biased and ω is obviously right-biased in SV-T model. The posterior distribution curve of the

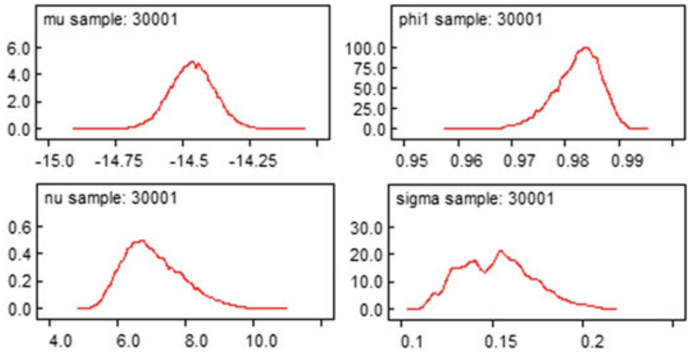


Fig. 3 Estimation of parameters posterior distribution density of the 5-min SV-T model in CSI 300

Table 2 Bayesian estimation results of CSI 300 SV-N and SV-T model parameters

	Node	Mean	sd	MC error	2.5%	Median	97.5%	Start	Sample
SV-N	mu	−14.24	0.05827	0.001647	−14.36	−14.24	−14.13	10000	30001
	phi1	0.9442	0.00943	5.911E-4	0.9234	0.9449	0.9604	10000	30001
	sigma	0.351	0.03266	0.002288	0.2907	0.3499	0.418	10000	30001
SV-T	mu	−14.47	0.08354	0.003295	−14.63	−14.47	−14.3	10000	30001
	nu	7.042	0.8692	0.05323	5.657	6.919	8.993	10000	30001
	phi1	0.9823	0.004324	2.727E-4	0.9726	0.9828	0.9894	10000	30001
	sigma	0.1526	0.01985	0.001467	0.1182	0.1526	0.1941	10000	30001

two model parameters is smooth, so the Markov chain has been basically stable. It shows that the parameters of the simulated estimation are effective.

The MC errors of the SV-N and SV-T models are far smaller than the standard deviation of the parameters from Table 2. The quantile interval estimation of each parameter is basically stable, and the fluctuation caused by the random model is small, so the estimation result is convergent and effective. SV-T model has stronger volatility and higher risk than SV-N model from the value of μ. There is continuity in the yield of the CSI 300 stock index from the value of ϕ. We can calculate the half-life of a fluctuating shock, the duration of the shock is half of its initial value decay by the formula $-\ln(2)/\ln(\phi)$[9]. The half-lives of SV-N and SV-T are 12.1 and 38.8 weeks respectively, indicating that the CSI 300 stock index has obvious fluctuation accumulation characteristics and the SV-T model has a stronger fluctuation persistence [6]; ε_t and η_t are independent error processes in the SV-N and SV-T models, and therefore are not taken into account leverage effects [13]; SV-N is noisy in the simulation process from the value of ω. The CSI 300 stock index yield does not obey the normal distribution, but it has the characteristics of high peak and heavy-tail (Fig. 3).

We analyze DIC fitting of the low-frequency daily yield and 5-min high-frequency return rate for SV-N and SV-T models by WinBUGS. SV-N model is better than SV-T

Table 3 Com parison of DIC of SV-N and SV-T

Time	Model	Dbar	Dhat	PD	DIC
5-min high -frequency	SV-N	−42967.600	−43355.000	387.389	−42580.200
	SV-T	−42663.100	−42829.000	165.850	−42497.300
Daily low-frequency	SV-N	−63.908	−63.956	0.048	−63.860
	SV-T	−51.615	−52.185	0.570	−51.045

model data, but it is more complex for high-frequency 5-min and low-frequency daily rate data from the DIC comparison (Table 3). The high-frequency of 5-min is better than the low-frequency daily data fitting from the perspective of DIC synthesis. The SV-N model is more suitable for simulating the CSI 300 stock index for the high-frequency 5-min yield data, which proves that the SV model is suitable for fitting the high frequency data.

4 Conclusion

We use high-frequency 5-min data of the CSI 300 index as sample data to study the volatility of the China's stock market with the high-frequency data, and analyze the statistical characteristics of yield data by using SV-N model and SV-T model. Then Bayes analysis is applied, and a priori distribution of parameters is set, the MCMC simulation process of Gibbs sampling is constructed. We compare the fitting degree under different models of the same data respectively by the DIC criterion.

The results indicates that:

China's CSI 300 index yield does not obey the normal distribution, but shows the characteristics of the heavy-tail,which further reflects the strong persistence of high-frequency income in China's stock market,shows that the current high volatility level will not return to a lower level of volatility quickly, it will continue to maintain high volatility in a couple of time periods.

By DIC analysis of the daily yield data and 5-min yield data, we conclude that the stochastic volatility model fits better to high-frequency 5-min yield data.

The SV-N model performs better than the SV-T model.

Acknowledgements The work was partly supported by National Natural Science Foundation of China (Grant No. 11971042) and National Science Foundation of Beijing Municipality (Grant No. 1182008).

References

1. Andersen, T., Bollerslev, T.: (Super) High frequency data analysis and modeling. Stat Study (11), 28–31 (2002)
2. Andersen, T., Bollerslev, T.: The distribution of exchange rate volatility. J. Am. Stat. Assoc. **96**(457), 42–55 (2000)
3. Andersen, T., Bollerslev, T.: Answering the skeptics: yes, Standard volatility models do provide accurate forecasts. Int. Econ. Rev. **39**(4), 885–905 (1998)
4. Gao, T.M.: Method and Modeling of Econometric Analysis, vol. 14. Tsinghua University Press (2009)
5. Li, S.G., Zhang, S.Y.: Financial volatility model based on high frequency data. Stat Decis (1), 7–8 (2008)
6. Li, G.H.: Study on fluctuation of coastal dry bulk freight rate based on stochastic volatility model. Res. Dev. Sci. Technol. World **38**(3) (2016)
7. Liu, F.Q.: Comparison of SV model based on DIC criterion. Stat. Decis. (9) (2004)
8. Lu, X.H., Zhang, Y.H., Zheng, Y.X.: Research and comparison of wavelet estimation methods for high frequency data volatility. Stat. Decis. (2018)
9. Madhu K., Raul S.: Regime-switching stochastic volatility and short-term interest rates. J. Empir Financ. **11**(3), 309–329 (2004)
10. Meyer, R., Yu, J.: BUGS for a Bayesian analysis of stochastic volatility models. Econ. J. (S1368-4221) **3**, 198–215 (2000)
11. Robert, C.P., Casella, G.: Monte Carlo Statistical Mothods. Springer, New York (2004)
12. Spiegelhalter, D.J., Best, N.G., Carlin, B.P.: Bayesian measures of model complexity and fit (with Discussion). J. R. Stat. Soc. Ser. B **64**(4), 583–616 (2002)
13. Sun, Z.H.: An analysis of the pulsating characteristics of the gold market in China. J. Changsha Univer. Sci. Technol. **30**(2) (2015)
14. Wang, T.Y., Huang, Z.: Research on modeling and application of volatility based on high frequency data. Econ. Perspect (3), 141–146 (2012)
15. Wang, C.F., Zhuang, H.G., Fang, Z.M., Lu, T.: Estimation and volatility prediction of long memory stochastic volatility models research on high frequency data of China's stock market. Syst. Eng. 7(26), 29–34 (2008)
16. Yang, J.Y., Zhang, Y.H.: Research on the market liquidity of stock index futures based on ACD model. Math. Prac. Theory **46**(9), 54–60 (2016)
17. Yang, J.Y., Zhang, Y.H.: Impact of circuit-breaker mechanism on chinas a share market. Stat. Decis. **13**, 153–155 (2017)
18. Zhang, B., Yu, C., Bi, T.: High Frequency Financial Data Modeling Theory, Method and Application, vol. 1, pp. 3–6. Tsinghua University Press (2015)

The Effects of Unequal Diffusion Coefficients on Spatiotemporal Pattern Formation in Prey Harvested Reaction-Diffusion Systems

Lakshmi Narayan Guin

Abstract In this investigation, we explore the spatiotemporal dynamics of reaction-diffusion predator-prey systems with Holling type II functional response. For partial differential equation, we consider the diffusion-driven instability of the coexistence equilibrium solution through spatiotemporal patterns. We find the conditions for Turing bifurcation of the system in a two-dimensional spatial domain by making use of the linear stability analysis and the bifurcation analysis. By choosing the ecological system parameter as the bifurcation parameter, we show that the system experiences a sequence of spatiotemporal patterns. The results of numerical simulations unveil that there are various spatial patterns including typical Turing patterns such as hot spots, spots-stripes mixture and stripes pattern through Turing instability. Our results show that the ecological system parameter plays a vital function in the proposed reaction-diffusion predator-prey models. Numerical design has been finally carried out through graphical representations of those outcomes towards the end in order to recognize the spatiotemporal behaviour of the system under study. All the outcomes are predictable to be of use in the study of the dynamic complexity of flora and fauna.

Keywords Reaction-diffusion equations · Prey refuge · Prey harvesting · Turing instability · Spatiotemporal pattern

AMS 35B36 · 35G31 · 35K55 · 37C75 · 37H20 · 70Kxx · 82B26

1 Introduction

The dynamical correlation between interacting species has been explored extensively in recent years due to its worldwide existence and significance in mathematical biology and ecology. The purpose of this research is to study systematically the dynamical properties of reaction-diffusion predator-prey model with Holling type II

L. Narayan Guin (✉)
Department of Mathematics, Visva-Bharati, Santiniketan 731235, West Bengal, India
e-mail: guin_ln@yahoo.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_24

functional response. A usual interaction of predator and prey is well known as the Holling type II model, which is broadly used in real life ecological relevance, and it has also been used to illustrate the spatiotemporal dynamics of reaction-diffusion predator-prey model systems [6, 7, 10, 11, 19–21]. In the present literature, reaction-diffusion systems typically take place as models for interactions between predator-prey species where the species can vary in size in space and move according to physical diffusion. The central focus of this paper is to investigate the predator-prey dynamics with both refuge and harvesting where, the interacting species moves in two-dimensional space consistent with self-diffusion.

Refuge exercise is any approach that reduces predation risk. The existence of refuges can evidently have significant effects on the co-existence of predator and prey species. Usually, prey species may escape being killed by predator species either by defending themselves or by evading. One approach to escape is to shift into a refuge where predation possibility is reduced [3, 16, 17, 22, 23]. By adding a large refuge to a model, Hassel [14] explained that the system revealed divergent oscillations in the absence of refuge, replaced the oscillatory behaviour by means of a stable equilibrium. Chen et al. [2] argued the instability and global stability properties of the equilibria and the existence and uniqueness of limit cycle of a Holling-Tanner type II predator-prey model incorporating a prey refuge. They establish that dynamic behaviour of the model system very much depends on the prey refuge parameter and increasing amount of refuge could increase prey densities and guide to predator eruptions. Even though there are many biological as well as ecological research articles available on the predator-prey temporal dynamics by incorporating a prey refuge, the significance of spatial prey refugia and diffusion on predator-prey model have received much less attention in the literature. The effects of refuges used by prey species on the dynamics of considered reaction-diffusion predator-prey model is examined at this point.

It is familiar that the harvesting has an important consequence on the dynamics of a model. In a predator-prey model with non-constant prey harvesting, the aim is to find out how much we can harvest without changing seriously the harvested species. The utilization of biological reserves and the harvesting of some ecological species are frequently observed in fishery, forestry, and wildlife management. Mathematical models have been exercised at length and effectively to gain insight into the scientific management of renewable resources like fishery and forestry [15, 18, 26]. Consequently, it is essential to revise the suitable interacting species model with non-constant prey harvesting. However, the effect of non-constant prey harvesting on the spatiotemporal dynamics of a reaction-diffusion predator-prey system has not been analyzed regularly. To the best of our information, little interest has been paid to the dynamics of a reaction-diffusion predator-prey model incorporating non-constant prey harvesting.

A number of research papers have been published over the last three decades on reaction-diffusion predator-prey models and different types of spatial patterns have been accounted for these reaction-diffusion systems [8, 9, 12, 13, 24]. A number of ecological models that explain reaction-diffusion predator-prey dynamics with harvesting have been enlarged, stemming from the significant research articles in the literature [4, 5, 13]. In these efforts the spatiotemporal dynamics of a predator-prey system with linear/nonlinear harvesting is modeled by the reaction-diffusion system. In [4, 5], it is demonstrated that the considered harvested model can generate spatiotemporal patterns through diffusion-driven instability. We summarize the result of interest from [4, 5] which measures the usefulness of the Turing pattern formation of a diffusive system when supplemented with mutual impacts of both prey refuge and harvesting. In this paper, a reaction-diffusion model describing the Holling-Tanner type II predator-prey interaction with intra-specific competition among predators and combined impacts with both prey refuge and harvesting is proposed. The aim of this manuscript is to study the influences of both prey refuge and harvesting on the spatiotemporal dynamics of a reaction-diffusion model with Holling type II prey-dependent functional response. Efforts have also been prepared to investigate the Turing pattern formation in a two-species interaction through diffusion-driven instability in the proposed predation model [25].

The rest of this paper is organized in the following manner: In Sect. 2, the predator-prey model in presence of diffusion is introduced with both prey refuge and harvesting. In Sect. 3, the model is examined in presence of spatial diffusion and the Turing bifurcation analysis around the unique positive spatially homogeneous steady state have been carried out. In Sect. 4, the Turing pattern formation via computational scheme is discussed briefly and numerical simulation results are analyzed. Finally, in Sect. 5, the ecological implications of analytical and numerical findings are talked about independently for the intention of sustaining ecological stability in nature.

2 The Mathematical Model with Diffusion and Its Analysis

In this investigation, we consider an extended model of the classical Bazykin's model [1] to explore the outcomes of both refuge and harvesting on the spatial pattern formation via diffusion-driven instability. In view of that, an effort is prepared to explore the following two-dimensional continuous pursuit-evasion predator-prey model with both linear prey harvesting and constant proportion of prey refuge as

$$\frac{\partial u}{\partial t} = ru\left(1 - \frac{u}{K}\right) - \frac{p_1(1-m)uv}{(1-m)u + p_2} - Hu + D_1\nabla^2 u, \tag{2.1a}$$

$$\frac{\partial v}{\partial t} = -p_3 v + \frac{p_4(1-m)uv}{(1-m)u + p_2} - p_5 v^2 + D_2\nabla^2 v, \tag{2.1b}$$

$$u(0, x, y) > 0, \quad v(0, x, y) > 0, \tag{2.1c}$$

where u, v denote prey and predator species respectively at (t, x, y) in a given spatial domain of $\mathbb{R}^2$, and $r, K, p_1, p_2, H, p_3, p_4, p_5, D_1, D_2$ are all positive ecological parameters. Here r represents intrinsic growth rate and K, the carrying capacity of the prey species; $m \in [0, 1)$ is the prey refuge parameter; p_1 is the maximum number of prey that can be eaten by each predator in unit time; p_2 is the interference coefficient of the predator; p_3 is the death rate of the predator; p_4 is the maximal predator growth rate; p_5 is the intra-specific competition among predators and $H \in [0, 1)$ stands for linear harvesting rate of prey species. The terms $\nabla^2 u$, $\nabla^2 v$ represent the diffusion of prey and predator species respectively, where ∇^2 indicates two-dimensional Laplacian operator; D_1 and D_2 are the self-diffusion coefficients for u and v, respectively.

System (2.1) takes place as a Holling-Tanner type II diffusive predator-prey model of interacting populations in the same spatial domain of $\mathbb{R}^2$. Assuming that all the utilized ecological parameters do not depend on space or time, that is, the environment is uniform. It is supposed that the refuge protecting mu of prey, where $m \in [0, 1)$, is constant and hence $(1 - m)u$ is only prey available to predator for predation. Throughout the analysis, we believe that the predator species in the model (2.1) is not of commercial/economical significance and accordingly the prey species is continuously being harvested with non-constant rate in time by a harvesting organization.

In order to minimize the number of parameters in model (2.1), we set $(u, v, t, x, y) = (K\tilde{u}, \frac{Kp_4\tilde{v}}{p_1}, \frac{\tilde{t}}{r}, \tilde{x}L, \tilde{y}L)$; L designates the characteristic length of the spatial domain Ω, and accordingly we arrive at the following equations containing dimensionless quantities of the system (2.1) (after dropping tildes):

$$\frac{\partial u}{\partial t} = F_1(u, v) + d_1\nabla^2 u, \quad (x, y) \in \Omega, \ t > 0, \tag{2.2a}$$

$$\frac{\partial v}{\partial t} = F_2(u, v) + d_2\nabla^2 v, \quad (x, y) \in \Omega, \ t > 0, \tag{2.2b}$$

$$u(0, x, y) > 0, \quad v(0, x, y) > 0, \quad (x, y) \in \Omega, \tag{2.2c}$$

where $F_1(u, v) = u(1 - u) - \frac{b(1-m)uv}{(1-m)u+a} - hu$, $F_2(u, v) = -cv + \frac{b(1-m)uv}{(1-m)u+a} - dv^2$ and the dimensionless system parameters are $a = \frac{p_2}{K}$, $b = \frac{p_4}{r}$, $c = \frac{p_3}{r}$, $d = \frac{Kp_4p_5}{p_1r}$, $h = \frac{H}{r}$, $d_1 = \frac{D_1}{rL^2}$ and $d_2 = \frac{D_2}{rL^2}$.

To solve the reaction-diffusion model (2.2), we employ Neumann boundary condition as $\frac{\partial}{\partial \xi}\begin{bmatrix} u \\ v \end{bmatrix}_{(x,y)} = 0, \ (x, y) \in \partial\Omega, \ t > 0$ where $\partial\Omega$ is the closed smooth boundary of the reaction-diffusion domain Ω and ξ is the unit outward normal to $\partial\Omega$.

With the purpose of spatial pattern formation through diffusion-driven instability of system (2.2), first we have to think over the non-spatial system of (2.2). Actually, the non-spatial system of (2.2) has three equilibria, which correspond to spatially homogeneous equilibria of system (2.2), in the positive quadrant as follows:

(i) $e_0(0, 0)$ (total extinct);

(ii) $e_1(1 - h, 0)$ (extinct of the predator);

(iii) a non-trivial feasible stationary state $e_2(u_2, v_2)$ (coexistence of both the species) where $v_2 = \frac{[a+(1-m)u_2](1-u_2-h)}{(1-m)b}$; $u_2 \in (0, 1 - h)$, $h \in [0, 1)$ and u_2 be the root of the cubic polynomial equation

$$\alpha u^3 + \beta u^2 + \gamma u + \delta = 0, \tag{2.3}$$

$$\alpha = (1-m)^2 d > 0,$$
$$\beta = -(1-h)(1-m)^2 d + 2(1-m)da,$$
$$\gamma = -(1-m)^2 cb - 2(1-h)(1-m)da + (1-m)^2 b^2 + a^2 d,$$
$$\delta = -(1-h)da^2 - (1-m)cab < 0.$$

Due to the ecological meaning, prime interest has been focused to investigate the dynamic behaviour about the positive and unique interior equilibrium point $e_2(u_2, v_2)$.

Now, we will discuss about the possible equilibrium solutions of the non-spatial system of (2.2). Figure 1 illustrates the nullcline of the model system, which is the intersection of prey nullcline (the blue one) and predator nullcline (the green one). The prey nullcline consists of the axis $u = 0$ and the curve $v = \frac{(1-u-a_4)[(1-a_2)u+a_3]}{a_1(1-a_2)}$ and the predator nullcline is given by $v = 0$ and the curve $v = \frac{1}{a_6}(-a_5 + \frac{a_1(1-a_2)u}{[(1-a_2)u+a_3]})$. One can observe the positive and unique interior equilibrium point $e_2(0.07767770319458, 0.33798353957001)$ corresponding to the parameter values $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.1$, $h = 0.1$ (cf. Fig. 1).

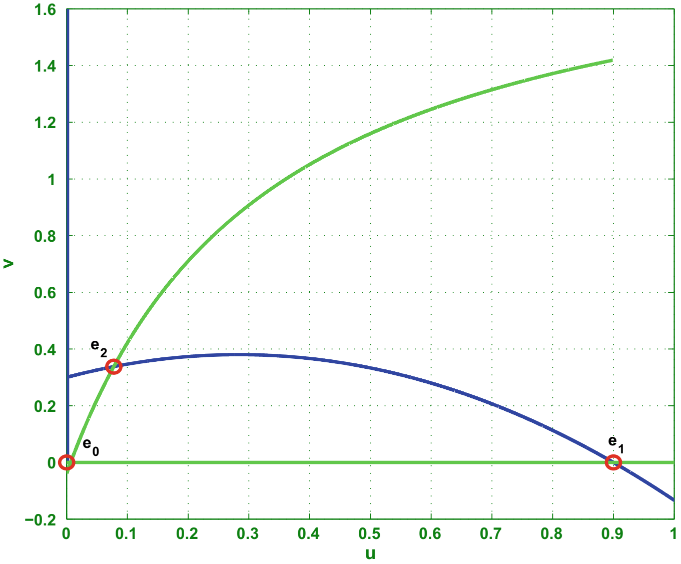


Fig. 1 Nullclines of the non-spatial system of (2.2) in uv-plane (blue colour curve for the prey nullcline, green colour curve for the predator nullcline) for $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.1$, $h = 0.1$.

3 Analysis in Presence of Diffusion of the System (2.2): Turing Bifurcation

To carry out a linear stability analysis for the non-trivial stationary state $e_2(u_2, v_2)$, we have to linearized the reaction-diffusion system (2.2) around the spatially homogeneous equilibrium point $e_2(u_2, v_2)$ for small space- and time-dependent perturbations and expand them in Fourier space. For this, let

$$\begin{aligned} u(\overrightarrow{x}, t) &= u_2 + \overline{u}(\overrightarrow{x}, t), \ |\overline{u}(\overrightarrow{x}, t)| << u_2, \\ v(\overrightarrow{x}, t) &= v_2 + \overline{v}(\overrightarrow{x}, t), \ |\overline{v}(\overrightarrow{x}, t)| << v_2, \\ \text{and } \begin{bmatrix} \overline{u}(\overrightarrow{x}, t) \\ \overline{v}(\overrightarrow{x}, t) \end{bmatrix} &= \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} e^{\lambda t} e^{i \overrightarrow{k} \overrightarrow{x}}, \ \overrightarrow{k} = (k_x, k_y), \end{aligned}$$

where $\overrightarrow{x} = (x, y)$ and $\overrightarrow{k} . \overrightarrow{k} = k^2$; k and λ are the wave number vector and fluctuated growth rate in time t, respectively; ζ_1, ζ_2 are the corresponding amplitudes. Then, we find the corresponding characteristic equation as follows:

$$|J - k^2 d - \lambda I_2| = 0, \tag{3.1}$$

$$J = \begin{bmatrix} 1 - 2u_2 - \frac{b(1-m)av_2}{[(1-m)u_2+a]^2} - h & -\frac{b(1-m)u_2}{[(1-m)u_2+a]} \\ \frac{b(1-m)av_2}{[(1-m)u_2+a]^2} & -c + \frac{b(1-m)u_2}{[(1-m)u_2+a]} - 2dv_2 \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}, \ d = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}, \ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The solutions of (3.1) can be achieved by the subsequent form:

$$\lambda_{2,1} = \frac{-B \pm \sqrt{B^2 - 4C}}{2}, \tag{3.2}$$

$$B(k^2) = k^2(d_1 + d_2) - tr(J), \tag{3.3}$$

$$C(k^2) = det \ J + k^4 d_1 d_2 - k^2(d_1 J_{22} + d_2 J_{11}). \tag{3.4}$$

The system (2.2) will be unstable if at least one of the eigenvalues λ_2 or λ_1 is positive. So, diffusion-driven instability can only be reached if $C(k^2) < 0$ and this condition guarantees that the coefficient of k^2 in (3.4) is positive i.e.

$$d_1 J_{22} + d_2 J_{11} > 0 \tag{3.5}$$

Here, $C(k^2)$ is a parabolic type quadratic function of k^2. Let, the minimum of $C(k^2) = 0$ is arrived at

$$k_{cr}^2 = \frac{d_1 J_{22} + d_2 J_{11}}{2d_1 d_2} > 0. \tag{3.6}$$

At $k^2 = k_{cr}^2$, $C(k^2) < 0$ transfers into

$$(d_1 J_{22} + d_2 J_{11})^2 > 4 d_1 d_2 (det\ J). \tag{3.7}$$

Specifically, Turing instability involves the following necessary and sufficient conditions:

$$\begin{aligned}
&(i)\ J_{11} + J_{22} < 0,\\
&(ii)\ J_{11} J_{22} - J_{12} J_{21} > 0,\\
&(iii)\ d_1 J_{22} + d_2 J_{11} > 0,\\
&(iv)\ \frac{(d_1 J_{22} + d_2 J_{11})^2}{4 d_1 d_2} > (J_{11} J_{22} - J_{12} J_{21}).
\end{aligned}$$

Mathematically, the Turing bifurcation of the system (2.2) happens when $\mathrm{Im}(\mu(k)) = 0$, $\mathrm{Re}(\mu(k)) = 0$ at $k = k_T \neq 0$ where the wave number k_T satisfies $k_T^2 = \sqrt{\frac{(J_{11} J_{22} - J_{12} J_{21})}{d_1 d_2}}$. The equilibria that can be found in the Turing space, are stable with regard to homogeneous perturbations but they mislay their stability with respect to perturbations of specific wave number k. The diffusion-driven instability cannot happen except the ratio of diffusion coefficient $\frac{d_2}{d_1}$ is suitably away from unity. From the conditions $J_{11} + J_{22} < 0$ and $d_1 J_{22} + d_2 J_{11} > 0$, one can able to find $d_2 > d_1$ easily, which points out that diffusivity of the predator is greater than that of the prey species.

4 Numerical Simulation Results: Spatiotemporal Pattern

In this section, numerical simulations are executed for the reaction-diffusion system (2.2) in two-dimensional square domain $\Omega = [0, 100] \times [0, 100]$ and the corresponding system parameters values chooses within the regime of Turing space. Extensive testing was presented through numerical simulation to understand the complex dynamical behaviour of the model (2.2), and the consequences are revealed here. In order to solve the reaction-diffusion system numerically, we use explicit finite difference scheme for the spatial derivatives and an explicit Euler method for the time integration with appropriate time step satisfying the CFL (Courant-Friedrichs-Lewy) stability criterion for two dimensional diffusion equations (i.e. $|d_i(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2})\Delta t| < \frac{1}{2}$, $i = 1, 2$) by means of space step size $\Delta x = \Delta y = 1.0$ and time step $\Delta t = 0.005$. We carried out extensive numerical simulations of the reaction-diffusion system (2.2) in Ω = [0, 100] × [0, 100] and the qualitative outcomes are revealed in the study. Our numerical simulations make use of the following initial conditions:

$$\begin{aligned}
&u(0, x, y) = u_2,\\
&v(0, x, y) = v_2 * ((x - 50)^2 + (y - 100)^2 >= 100) + 0.00 * ((x - 50)^2 + (y - 100)^2 < 100).
\end{aligned}$$

Throughout the numerical simulations, special types of dynamics are obtained and the allocations of prey and predator species are always of the similar type. As a result, we can confine our analysis of pattern formation to prey species.

4.1 The Effect of Unequal Diffusion Coefficients d_1 and d_2

Figure 2 illustrates the progress of the stationary Turing pattern of the prey species at (i) $d_2 = 5.0$, (ii) $d_2 = 10.0$, (iii) $d_2 = 20.0$, (iv) $d_2 = 40.0$ with small random perturbation of the stationary solution u_2 and v_2 of the spatially homogeneous system (2.2) through $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.0$, $h = 0.0$, $d_1 = 0.1$. In addition, we can examine that as d_2 increases from 5.0 to 40.0, the radius of hot spots are uniformly increases and significantly the concentration of the prey species increases also. The entire thing occurs over the spatial domain, starting from hot spots with small radius, and finally the dynamics of the system does not undergo any further changes.

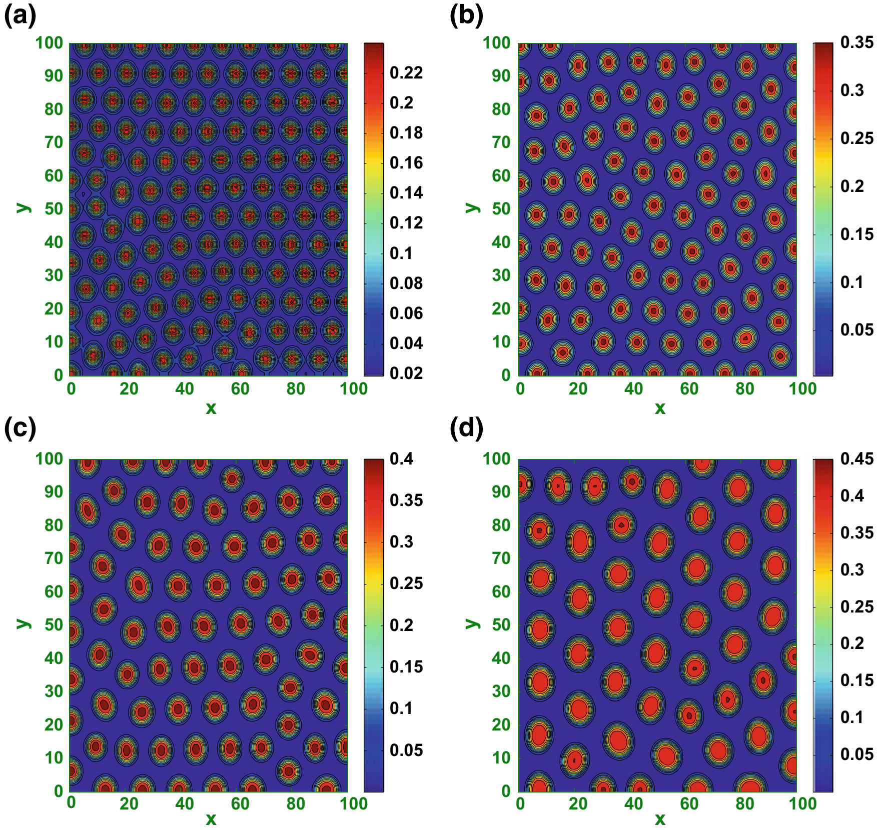


Fig. 2 Distinctive spatial pattern of the prey size in two-dimensional space for the system (2.2). Here the parameter values are $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.0$, $h = 0.0$, $d_1 = 0.1$. **a** $d_2 = 5.0$, **b** $d_2 = 10.0$, **c** $d_2 = 20.0$, **d** $d_2 = 40.0$. Predator size shows similar properties

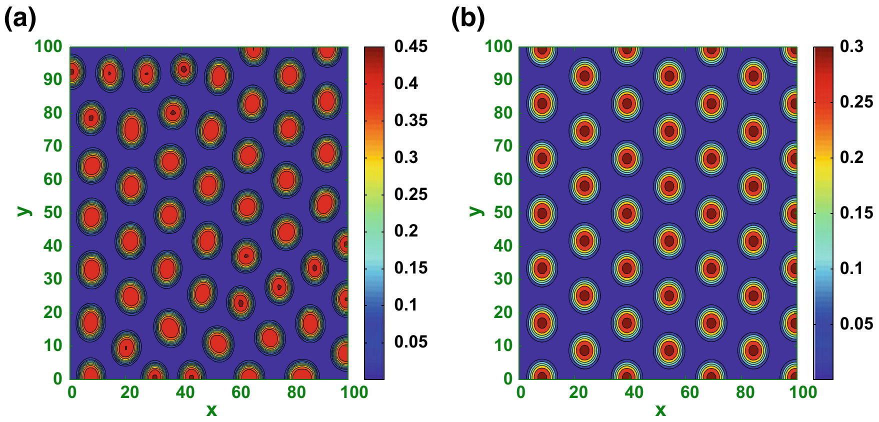


Fig. 3 Distinctive spatial pattern of the prey size in two-dimensional space for the system (2.2). Here the parameter values are $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.0$, $d_1 = 0.1$, $d_2 = 40.0$. **a** $h = 0.0$, **b** $h = 0.18$. Predator size shows similar properties

4.2 The Effect of h in Absence of Refuge

Figure 3 shows the evolution of the stationary Turing pattern of the prey species at (i) $h = 0.0$, (ii) $h = 0.18$ with small random perturbation of the stationary solution u_2 and v_2 corresponding to the other parameters $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $m = 0.0$, $d_1 = 0.1$, $d_2 = 40.0$. As can be seen from Fig. 3, as h increases, hot spots patterns are gradually formed and consequently the prey concentration decreases as well.

4.3 The Effect of m in Absence of Harvesting

Figure 4 demonstrates three typical spatiotemporal patterns of the prey species for special values of m. From the Fig. 4, one can notice that the steady-state pattern takes a long time to settle down, starting with a homogeneous state $e_2(u_2, v_2)$, and the random perturbation leads to the structure of hot spots (cf. Fig. 4a), and ending with the coexistent of hot spots and stripes patterns (cf. Fig. 3c).

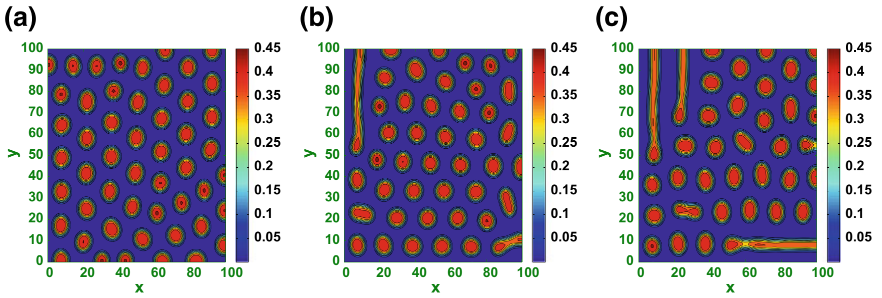


Fig. 4 Distinctive spatial pattern of the prey size in two-dimensional space for the system (2.2). Here the parameter values are $a = 0.3$, $b = 1.0$, $c = 0.02$, $d = 0.5$, $h = 0.0$, $d_1 = 0.1$, $d_2 = 40.0$. **a** $m = 0.0$, **b** $m = 0.2$, **c** $m = 0.3$. Predator size shows similar properties

4.4 The Effect of d in Presence of Both Refuge and Harvesting

Figure 5 explains four typical spatial patterns through diffusion-driven instability of u for the system (2.2) for different values of the parameter d. From the Fig. 5, we can detect that as the significant parameter d increases, the sequence "hot spots (cf. Fig. 5a, b) → hot spots and stripe mixture patterns (cf. Fig. 5c) → almost stripes patterns (cf. Fig. 5d)" is observed.

4.5 The Effect of c in Presence of Both Refuge and Harvesting

Similarly, from Fig. 6, we can observe that as the value of the parameter c increases, the sequence "hot spots (cf. Fig. 6a) → hot spots and short stripes (cf. Fig. 6b) → hot spots and stripe mixture patterns (cf. Fig. 6c) → stripes patterns (cf. Fig. 6d) " is found to emerge.

5 Conclusion and Discussion

In the current paper, we have reflected on a diffusive model for predator-prey interaction with Holling-Tanner type II functional response within two-dimensional space. A variety of spatiotemporal patterns generated by a reaction-diffusion predator-prey system with both prey refuge and harvesting are investigated analytically and numerically. A series of numerical simulations illustrated that pattern transition will emerge when both prey refuge and harvesting are added. We obtained the required conditions for diffusion-driven instability in terms of system parameters. We have also confirmed the idea that resulting spatial pattern lies in the interior of Turing

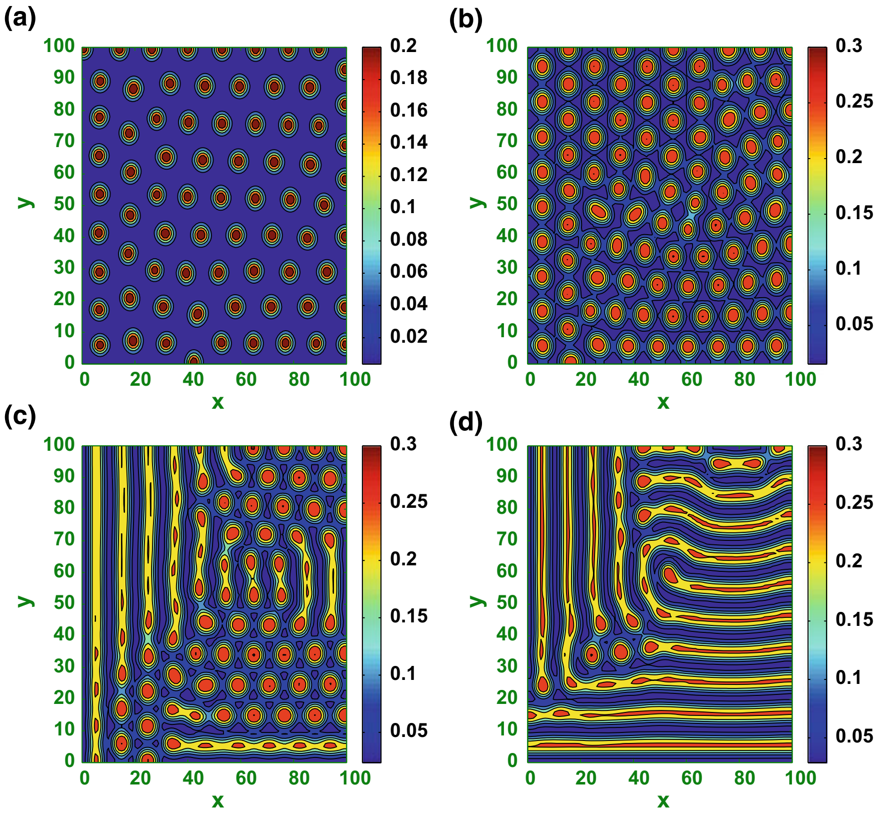


Fig. 5 Distinctive spatial pattern of the prey size in two-dimensional space for the system (2.2). Here the parameter values are $a = 0.3$, $b = 1.0$, $c = 0.02$, $m = 0.1$, $h = 0.1$, $d_1 = 0.1$, $d_2 = 10.0$. **a** $d = 0.35$, **b** $d = 0.75$, **c** $d = 0.82$, **d** $d = 0.85$. Predator size shows similar properties

space. Numerical simulation results indicate that Turing patterns (cf. Figs. 2, 3, 4, 5 and 6) can emerge through the interaction between the diffusion and both prey refuge and harvesting with other momentous system parameters in the system (2.2). Throughout the investigation, we have noticed that the variety of parameter values is important to study the effect of diffusion and diffusion-driven Turing instability cannot happen for $d_1 = d_2$.

Figure 3 demonstrates the evolution of the stationary spatial pattern of the prey species (u) at different h by keeping the value of prey refuge m fixed at $m = 0.0$ and as h increases, uniform hot spot patterns are gradually formed and consequently the prey concentration decreases. Different Turing patterns of the prey species emerge as we vary constant proportion of prey refuge m by keeping the value of linear prey harvesting h fixed at $h = 0.0$ (cf. Fig. 4a–c). The absence of prey refuge $m = 0.0$ results in spot patterns, that is, where the prey abundance is higher in isolated zones (cf. Fig. 4a). Now, as we increase the value of prey refuge, we get a mixture of stripe

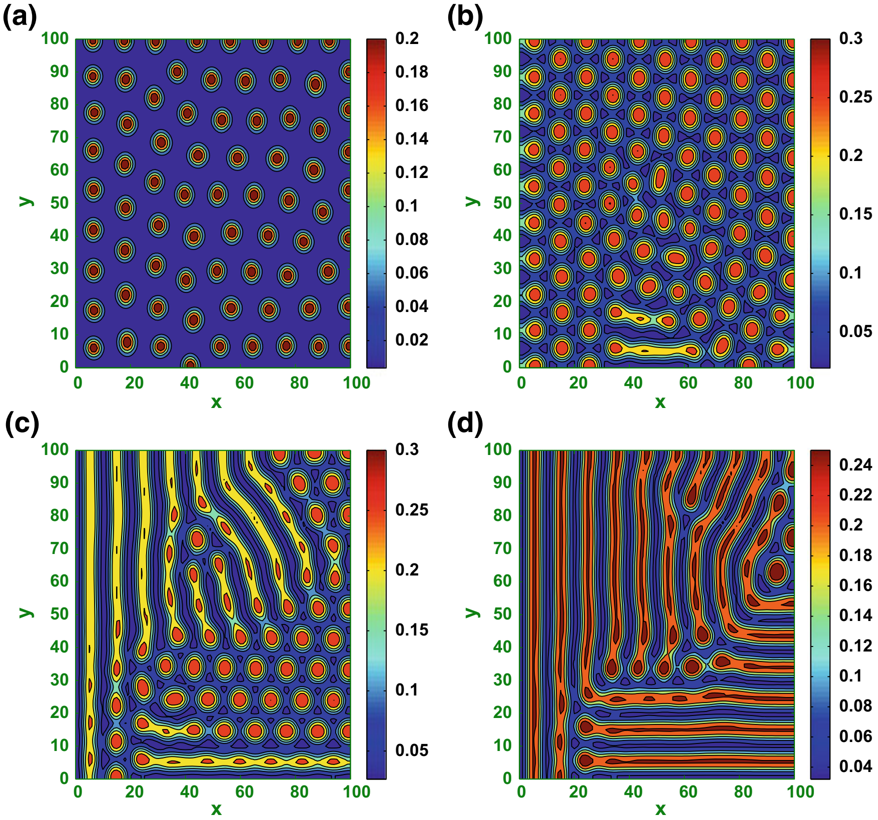


Fig. 6 Distinctive spatial pattern of the prey size in two-dimensional space for the system (2.2). Here the parameter values are $a = 0.3$, $b = 1.0$, $d = 0.5$, $m = 0.1$, $h = 0.1$, $d_1 = 0.1$, $d_2 = 10.0$. **a** $c = 0.02$, **b** $c = 0.12$, **c** $c = 0.13$, **d** $c = 0.145$. Predator size shows similar properties

and spot patterns. In the presence of prey refuge and harvesting, variations in intra-specific competition among predators (d) generate spots, mixture of spots and stripes (cf. Fig. 5). Therefore, by increasing the value of intra-specific competition among predators (d), the pattern sequence "hot spots (cf. Fig. 5a, b) → hot spots and stripe mixture patterns (cf. Fig. 5c) → almost stripes patterns (cf. Fig. 5d) " is detected. So, the consequences based on the current system explain that the effect of both prey refuge, harvesting and intra-specific competition among predators for spatial pattern formation is insightful on the dynamic complexity of ecosystems.

In this paper, we have studied the effect of internal competition response on the formation and characters of Turing patterns. In reality, the formation and characteristics of Turing patterns may be influenced by some other factors in ecological systems, including fluctuate environment feedback, circumstance noise and the periodic travelling wave feedback, and these will be investigated in the future.

Acknowledgements The author confers thank to the anonymous referee for the very helpful ideas and statements which led to developments of our original manuscript. Also, the author gratefully acknowledges the financial support in part from Special Assistance Programme (SAP-III) sponsored by the University Grants Commission (UGC), New Delhi, India (Grant No. F.510/3/DRS-III/2015 (SAP-I)).

References

1. Bazykin, A.D., Khibnik, A.I., Krauskopf, B.: Nonlinear Dynamics of Interacting Populations, vol. 11. World Scientific Publishing Company Incorporated. Singapore (1998)
2. Chen, L., Chen, F., Chen, L.: Qualitative analysis of a predator-prey model with Holling type II functional response incorporating a constant prey refuge. Nonlinear Anal.: Real World Appl. **11**, 246–252 (2010)
3. Collings, J.B.: Bifurcation and stability analysis of a temperature-dependent mite predator-prey interaction model incorporating a prey refuge. Bull Math Biol **57**, 63–76 (1995)
4. Duque, C., Lizana, M.: On the dynamics of a predator-prey model with nonconstant death rate and diffusion. Nonlinear Anal: Real World Appl **12**, 2198–2210 (2011)
5. Feng, P.: On a diffusive predator-prey model with nonlinear harvesting. Math Biosci Eng **11**, 807–821 (2014)
6. Guin, L.N., Baek, H.: Comparative analysis between prey-dependent and ratio-dependent predator-prey systems relating to patterning phenomenon. Math Comput Simul **146**, 100–117 (2018)
7. Guin, L.N., Mandal, P.K.: Effect of prey refuge on spatiotemporal dynamics of reaction-diffusion system. Comput. Math. Appl. **68**, 1325–1340 (2014a)
8. Guin, L.N., Mandal, P.K.: Spatial pattern in a diffusive predator-prey model with sigmoid ratio-dependent functional response. Int. J. Biomath. **7**, 1450047 (2014b)
9. Guin, L.N., Mandal, P.K.: Spatiotemporal dynamics of reaction-diffusion models of interacting populations. Appl. Math. Model. **38**, 4417–4427 (2014c)
10. Guin, L.N., Haque, M., Mandal, P.K.: The spatial patterns through diffusion-driven instability in a predator-prey model. Appl. Math. Model. **36**, 1825–1841 (2012)
11. Guin, L.N., Chakravarty, S., Mandal, P.K.: Existence of spatial patterns in reaction-diffusion systems incorporating a prey refuge. Nonlinear Anal.: Model. Control **20**, 509–527 (2015)
12. Guin, L.N., Mondal, B., Chakravarty, S.: Existence of spatiotemporal patterns in the reaction-diffusion predator-prey model incorporating prey refuge. Int. J. Biomath. **9**, 1650085 (2016)
13. Guin, L.N., Mondal, B., Chakravarty, S.: Spatiotemporal patterns of a pursuit-evasion generalist predator-prey model with prey harvesting. J. Appl. Nonlinear Dyn. **7**, 165–177 (2018)
14. Hassell, M.P.: The Dynamics of Arthropod Predator-Prey Systems. Princeton University Press (1978)
15. Huang, J., Gong, Y., Ruan, S.: Bifurcation analysis in a predator-prey model with constant-yield predator harvesting. Discr. Contin. Dyn. Syst.-Ser. B **18**, 2101–2121 (2013)
16. Ko, W., Ryu, K.: Qualitative analysis of a predator-prey model with Holling type II functional response incorporating a prey refuge. J. Differ. Equ. **231**, 534–550 (2006)
17. Ma, Z., Li, W., Zhao, Y., Wang, W., Zhang, H., Li, Z.: Effects of prey refuges on a predator-prey model with a class of functional responses: the role of refuges. Math. Biosci. **218**, 73–79 (2009)
18. Makinde, O.D.: Solving ratio-dependent predator-prey system with constant effort harvesting using adomian decomposition method. Appl. Math. Comput. **186**, 17–22 (2007)
19. Murray, J.D.: Mathematical Biology II: Spatial Models and Biomedical Applications (2002)
20. Okubo, A., Levin, S.A.: Diffusion and Ecological Problems: Modern Perspective (2001)
21. Okubo, A.: Diffusion and Ecological Problems: Mathematical Models (1980)
22. Pallini, A., Janssen, A., Sabelis, M.W., et al.: Predators induce interspecific herbivore competition for food in refuge space. Ecol. Lett. **1**, 171–177 (1998)

23. Sih, A.: Prey refuges and predator-prey stability. Theor. Popul. Biol. **31**, 1–12 (1987)
24. Sun, G.Q., Sarwardi, S., Pal, P.J., Rahaman, S.: The spatial patterns through diffusion-driven instability in modified Leslie-Gower and Holling-type II predator-prey model. J. Biol. Syst. **18**, 593–603 (2010)
25. Turing, A.M.: The chemical basis of morphogenesis, philosophical transactions of the royal society of London. Biol. Sci. **237**, 37–72 (1952)
26. Xiao, D., Jennings, L.S.: Bifurcations of a ratio-dependent predator-prey system with constant rate harvesting. SIAM J. Appl. Math. **65**, 737–753 (2005)

Optimal Pricing Strategy in a Two-Echelon Supply Chain with Admissible Advanced and Delayed Payments

B. C. Giri, R. Bhattacharjee and T. Maiti

Abstract In the traditional economic order quantity model, the purchasing cost of an order is paid at the time of its receipt. However, in reality, installment payment of purchasing cost is very common and many distributors pay the purchasing cost to the manufacturers in installments (upstream partial prepayment). In a similar manner, distributors allow retailers to pay the purchasing cost after the goods are received (downstream partial delay payment). This article develops a two-echelon supply chain model with a single-manufacturer and a single-retailer in which the manufacturer adopts a lot-for-lot policy for meeting the demand of the retailer under admissible advanced and delayed payment options. The market demand at the retailer is assumed to be linearly dependent on the selling price. The effect of advanced/delayed payment on the optimal payment time is analyzed. The optimal results of the developed model and sensitivity analysis are presented for a numerical example. It is revealed that the supply chain's average profit attains the maximum when the optimal cycle time is longer than the optimal payment time.

Keywords Supply chain · Lot-for-lot policy · Price-dependent demand · Advanced/delayed payment

1 Introduction

The competitive strategy of a supply chain largely depends on its payment procedure. The optimal payment is an important driver in the suitable balance in efficient-responsiveness spectrum of supply chain framework. In the traditional EOQ

B. C. Giri (✉) · R. Bhattacharjee · T. Maiti
Department of Mathematics, Jadavpur University, Kolkata 700032, India
e-mail: bcgiri.jumath@gmail.com

R. Bhattacharjee
e-mail: rupakb13@yahoo.co.in

R. Bhattacharjee
Department of Mathematics, JIS College of Engineering, Kalyani 741235, India

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_25

(Economic Order Quantity) model, the purchasing cost of an order is paid at the time of its receipt. In some cases, retailers request purchasers to pay all or a fraction of the purchasing cost in advance and the remaining fraction in several installments. In practice, retailers are often allowed delayed payments by their suppliers in order to increase sales volume. For retailers, the delayed payment not only lowers the opportunity cost of capital but also allows them to earn interest on the revenue of goods sold.

Timing of payment of purchasing cost plays a major role in deriving the optimal decisions of the distributors or retailers. There are three basic payment strategies, viz. (i) delivery time payment, (ii) delayed payment and (iii) advanced payment. Goyal [1] initiated a fixed delay payment period in the traditional EOQ model. Subsequently, many researchers proposed inventory models relating to the permissible delay in payments (Jamal et al. [2]; Teng et al. [3]; Chung [5] and others). Significant amount of work has so far been done in this area assuming that the length of the delay payment period is linked to the retailer's order quantity (Huang [6]), and two-level trade credit policy (Giri and Maiti [7]; Giri and Sharma [8]). Some researchers assumed that the supplier provides the delay payment period and cash discount simultaneously (Yang [9]).

In response to the supplier's trade credit offer, the retailer may pay before time or fail to pay in time. The supplier can charge additional cost on delay time or provide cash discount for prior payment. In this situation, the optimal payment time decision is more important to the retailer in doing business. However, the optimal payment time decision from the point of view of the retailer has been studied by some researchers. Jamal et al. [2] and Sarker et al. [10] determined optimal cycle time and payment time when the supplier allows a specified credit period to the retailer for payment without penalty. Liao and Chen [11] focused on a retailer's inventory control system in the optimal delay in payment time for initial stock-dependent consumption rate when a wholesaler permits a delayed payment. Huang [12] investigated the buyer's optimal cycle time and optimal payment time under the supplier's trade credit policy and cash discount policy. Chang et al. [13] developed a mathematical model to determine the optimal payment period and replenishment cycle when the supplier offers the retailer a permissible delay in payments. Maiti et al. [14] studied the effect of advanced payment for price-dependent demand in a stochastic environment. Recently Giri et al. [15] investigated optimal payment time with trade credit financing. Li et al. [16] focused on advance-cash-credit payments by a discounted cash flow analysis. Correlation between expiration date and advanced payment was interestingly discussed by Teng et al. [17]. Zhang et al. [18] exhibited ordering policy with advanced payment for stable supply capacity.

In this paper, we consider the optimal payment time of the retailer in a two-echelon supply chain system where the retailer has the full liberty to pay whenever he wants (before the distribution cycle starts, in the intermediate time with variations) under some mutually agreed conditions on the wholesale price with the manufacturer. The paper is structured as follows. In the next section, notations and assumptions for developing the model are given. The proposed model is formulated in Sect. 3. Section 4 deals with model analysis and some important properties of the proposed model. Section 5 demonstrates the developed model for a numerical example. Critical value of the controlling parameter for flipping the choice of the model is also presented in this section. Sensitivity analysis is presented in Sect. 6. Section 7 concludes the paper and gives some directions for future research.

2 Notations and Assumptions

The following notations are used to develop the proposed model:

p	retailer's unit selling price (decision variable)
$D(= D(p))$	demand rate at the retailer
$P(> D)$	production rate at the manufacturer
Q	retailer's order quantity
T	replenishment interval of the retailer, i.e. the cycle time (decision variable)
R	time period after/before which the actual payment is made by the retailer
t_m	time delay for the manufacturer to begin production
w	unit cost for the retailer
β	controlling factor for payment time variation
w_1	wholesale price for advanced payment
w_2	wholesale price for delayed payment
c	unit cost for the manufacturer
a_r	retailer's ordering cost per order
a_m	manufacturer's set up cost per set up
h_r	retailer's holding cost (excluding the opportunity cost) per unit per unit time
h_m	manufacturer's holding cost (excluding the opportunity cost) per unit per unit time
i_e	interest rate at which interest is earned from the bank
i_p	interest rate at which interest is paid to the bank
i_v	interest rate lost by the manufacturer (opportunity cost).

The following assumptions are made for developing the proposed model.

- The supply chain under consideration consists of a single-manufacturer and a single-retailer for trading a single product where the manufacturer follows a lot-for-lot production policy in response to the retailer's order quantity.
- The market demand D at the retailer is linearly dependent on the selling price p. We take $D = D(p) = a - bp$, where $a > 0, b > 0$ such that $p < a/b$.
- Shortages in the retailer's inventory are not allowed and all replenishments are made instantaneously.
- The manufacturer's production rate is sufficiently greater than the retailer's demand rate.
- The manufacturer offers an optional payment procedure to the retailer. This offer contains advanced payment or delayed payment. Advanced payment has to be made before the initiation of replenishment. The delayed payment is classified into two categories. The payment can be made during any time of replenishment i.e., within T or after the completion of a replenishment cycle. If the payment is made in advance then the wholesale price w_1 will be $(1 - \beta)w$ and, in case of delayed payment, the wholesale price w_2 will be $(1 + \beta)w$, where β is the controlling factor for payment time variation.

- If the payment time is adopted by the retailer as the advanced mode, then at the time of payment, the retailer has to take a bank loan to fulfill the payment at an interest rate i_p and repay it at the end of the cycle.
- If the payment time is adopted by the retailer as the delayed mode, then before the time of payment, the manufacturer incurred an opportunity loss at an interest rate i_v.
- During the whole process, the retailer deposits his earning to an interest giving organization (bank) with an interest rate i_e and according to the requirement of the retailer, it is accessible without any condition.
- The planning horizon is infinite.

3 Model Formulation

3.1 Retailer's Perspective

The instantaneous states of the retailer's inventory level $I_r(t)$ at any instant t in the time interval $[0, T]$ can be described as

$$\frac{dI_r(t)}{dt} = -D, \quad 0 \le t \le T; \quad I_r(T) = 0.$$

Solving, we get

$$I_r(t) = D(T - t), \quad 0 \le t \le T.$$

Since $I_r(0) = Q$, therefore, we have $Q = DT$. Then the holding cost per unit time of the retailer is

$$\frac{h_r}{T} \int_0^T I_r(t)dt = \frac{h_r DT}{2}.$$

We now consider the following two cases based on the relation of the payment time R and the replenishment interval T whose variation is basically termed as advanced payment and delayed payment.

Case I: Advanced payment

In this payment mode, it is assumed that the payment towards the manufacturer is done before the item is procured for next phase delivery. The total sales revenue in the period is pDT and the profit is $(p - w_1)DT$. Due to advanced payment, the retailer takes loan from bank and repays at the end of the cycle i.e., the interest is to be paid with a duration of $(T + R)$ time. Since the interest rate at which interest is paid to the bank is i_p, the total amount of interest is $i_p w_1 DT(T + R)$. Meanwhile, when the disbursement of the procured quantity is started from the retailer's end, the accumulated revenue is kept in an interest giving organization (bank) with interest rate i_e. The corresponding gain is

$$i_e p \int_0^T Dt\,dt = \frac{i_e p D T^2}{2}.$$

So the net profit of the retailer
$= sales\ profit - ordering\ cost - holding\ cost - interest\ paid\ to\ the\ bank + interest\ earned\ from\ the\ bank.$
Therefore, the retailer's net profit per unit time $TP_{rI}(T, p, R)$ is given by

$$TP_{rI}(T, p) = (p - w_1)D - \frac{a_r}{T} - \frac{h_r DT}{2} - i_p w_1 D(T + R) + \frac{i_e pDT}{2} \qquad (1)$$

Case II: Delayed payment
In this case, two subcases may arise depending on the payment time. Payment can be made any time within the cycle i.e., $0 < R < T$, or after the completion of a cycle i.e., $T < R$.

Subcase IIA: $0 < R < T$
In this scenario, the total sales revenue is pDT and the total wholesale price is $w_2 DT$. So the net profit becomes $(p - w_2)DT$. As the payment time is after the starting of the product sales inception, in this duration, the retailer keeps the earning into an interest giving concern (bank) with an interest rate i_e. So the accumulated interest in this time interval will be

$$i_e p \int_0^R Dt\,dt = \frac{i_e pDR^2}{2}.$$

After the payment made at time R, the sales revenue is also kept in the bank to get the interest during the residual time. The corresponding interest will be

$$i_e p \int_0^{T-R} Dt\,dt = \frac{i_e pD(T - R)^2}{2}.$$

We now have the following two important observations:

(i) At the time of payment, if the sum of total sales revenue and acquired interest is less than or equal to $w_2 DT$, then the deficit amount has to be taken from a bank with an interest rate i_p. So the interest to be paid to the bank will be

$$i_p \int_0^{T-R} \left[w_2 DT - p \int_0^R Dt\,dt - i_e p \int_0^R Dt\,dt \right] dt.$$

(ii) At the time of payment, if the sum of total sales revenue and acquired interest is greater than or equal to $w_2 DT$, then the residual amount and the collected sales revenue will be deposited into an interest giving organization (bank) for the duration of $(T - R)$ time at an interest rate i_e. The interest to be earned from the bank will be

$$i_e \int_0^{T-R} \Big[p \int_0^R Dt\, dt + i_e p \int_0^R Dt\, dt - w_2 DT \Big]\, dt.$$

In both the cases, the ordering cost and the holding cost remain same as A_r and $\frac{h_r DT^2}{2}$.

So the net profit of the retailer
$= sales\ profit + interest\ earned\ from\ the\ bank - ordering\ cost - holding\ cost - interest\ paid\ to\ the\ bank(or + interest\ earned)$.

Therefore, the retailer's net profit per unit time in the two situations are given by

$$\begin{aligned} TP_{rIIA_1}(T,p) = (p - w_2)D + i_e p\frac{DR^2}{2T} + \frac{i_e pD(T-R)^2}{2T} - \frac{a_r}{T} - \frac{h_r DT}{2} \\ - i_p \frac{T-R}{T}\Big(w_2 DT - i_e p\frac{DR^2}{2} - p\frac{DR^2}{2}\Big) \end{aligned} \tag{2}$$

and

$$\begin{aligned} TP_{rIIA_2}(T,p) = (p - w_2)D + i_e p\frac{DR^2}{2T} + \frac{i_e pD(T-R)^2}{2T} - \frac{a_r}{T} - \frac{h_r DT}{2} \\ + i_e \frac{T-R}{T}\Big(p\frac{DR^2}{2} + i_e p\frac{DR^2}{2} - w_2 DT\Big) \end{aligned} \tag{3}$$

Subcase IIB: $T < R$

In this case, payment will be made after completion of the cycle i.e., $T < R$. Sales revenue, ordering cost and holding cost components are all same as the subcase IIA. But as the payment is made after time T, the interest earned by the retailer consists of two parts. Firstly, the interest earned during the cycle $[0, T]$ is

$$i_e p \int_0^T Dt\, dt = \frac{i_e pDT^2}{2}$$

and the interest earned during $[T, R]$ is

$$i_e pDT \int_T^R dt = i_e pDT(R - T).$$

So the net profit of the retailer
$= sales\ profit + interest\ earned\ from\ the\ bank - ordering\ cost - holding\ cost$.

Therefore, the retailer's net profit per unit time $TP_{rIIB}(T, p)$ is given by

$$TP_{rIIB}(T,p) = (p - w_2)D + i_e p\Big(\frac{DT}{2} + D(R - T)\Big) - \frac{a_r}{T} - \frac{h_r DT}{2} \tag{4}$$

3.2 Manufacturer’s Perspective

The manufacturer’s inventory level $I_m(t)$ at any time t is given by

$$I_m(t) = P(t - t_m); \quad t_m \leq t \leq T.$$

Therefore, the holding cost of the manufacturer is

$$h_m \int_{t_m}^{T} I_m(t)dt = \frac{h_m P(T - t_m)^2}{2}.$$

Case I: Advanced payment
As the payment is made in advance, the acquired amount will be kept in an interest giving concern at a rate i_e. So then the accumulated interest will be $i_e w_1 DTR$. Therefore, the manufacturer’s net profit per unit time will be

$$TP_{mI}(T, p) = (w_1 - c)D + i_e w_1 DR - \frac{a_m}{T} - \frac{h_m P(T - t_m)^2}{2T} \tag{5}$$

Case II: Delayed payment
In this case, the wholesale price is considered as w_2 and then the total sales profit is $(w_2 - c)DT$.
Subcase IIA: $0 < R < T$
As the payment is made after the delivery, during this time, manufacturer may have the rolling money which is not achieved in reality. So the corresponding opportunity loss is $i_v w_2 DTR$, where i_v is the opportunity loss rate. Considering the ordering cost and the holding cost, the manufacturer’s net profit per unit time is given by

$$TP_{mIIA}(T, p) = (w_2 - c)D - i_v w_2 DR - \frac{a_m}{T} - \frac{h_m P(T - t_m)^2}{2T} \tag{6}$$

Subcase IIB: $T < R$
As the payment is made after the cycle time, the opportunity loss is $i_v w_2 DT(T + R)$. Other cost components remain same as case IIA. Hence the manufacturer’s net profit per unit time becomes

$$TP_{mIIB}(T, p) = (w_2 - c)D - i_v w_2 D(T + R) - \frac{a_m}{T} - \frac{h_m P(T - t_m)^2}{2T} \tag{7}$$

Since the manufacturer is a make-to-order producer, he follows a lot-for-lot production policy in response to the retailer’s demand. Therefore, if the lot demanded by the retailer is $Q(= DT)$ and the manufacturer produces that lot during the period $(T - t_m)$ at the rate P, then we have $Q = P(T - t_m)$ which gives $\frac{P(T-t_m)}{T} = \frac{Q}{T} = D$. Substituting $D = \frac{P(T-t_m)}{T}$, we obtain the manufacturer’s profit per unit time as follows:

$$TP_{mI}(T,p) = (w_1 - c)D + i_e w_1 DR - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \tag{8}$$

$$TP_{mIIA}(T,p) = (w_2 - c)D - i_v w_2 DR - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \tag{9}$$

$$TP_{mIIB}(T,p) = (w_2 - c)D - i_v w_2 D(T+R) - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \tag{10}$$

3.3 Integrated Supply Chain

Taking into account all the cases discussed in the previous subsections, the net profit per unit time of the integrated supply chain is given by

$$\begin{aligned} TP(T,p) &= TP_r(T,p) + TP_m(T,p) \\ &= \begin{cases} TP_1(T,p), & \text{if payment is made in advance;} \\ TP_2(T,p) & \text{if } 0 \le R \le T \text{ and } w_2 DT > i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}; \\ TP_3(T,p), & \text{if } 0 \le R \le T \text{ and } i_e p\frac{DR^2}{2} + p\frac{DR^2}{2} > w_2 DT; \\ TP_4(T,p), & \text{if } T \le R. \end{cases} \end{aligned} \tag{11}$$

where

$$\begin{aligned} TP_1(T,p) &= TP_{rI}(T,p) + TP_{mI}(T,p) \\ &= (p-c)D - \frac{a_r}{T} - \frac{h_r DT}{2} - i_p w_1 D(T+R) + \frac{i_e pDT}{2} + i_e w_1 DR \\ &\quad - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \\ TP_2(T,p) &= TP_{rIIA_1}(T,p) + TP_{mIIA}(T,p) \\ &= (p-c)D + i_e p\frac{DR^2}{2T} + \frac{i_e pD(T-R)^2}{2T} - \frac{a_r}{T} - \frac{h_r DT}{2} \\ &\quad - i_p\frac{T-R}{T}\Big(w_2 DT - i_e p\frac{DR^2}{2} - p\frac{DR^2}{2}\Big) - i_v w_2 DR - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \\ TP_3(T,p) &= TP_{rIIA_2}(T,p,R) + TP_{mIIA}(T,p) \\ &= (p-c)D + i_e p\frac{DR^2}{2T} + \frac{i_e pD(T-R)^2}{2T} - \frac{a_r}{T} - \frac{h_r DT}{2} \\ &\quad + i_e\frac{T-R}{T}\Big(p\frac{DR^2}{2} + i_e p\frac{DR^2}{2} - w_2 DT\Big) - i_v w_2 DR - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \\ TP_4(T,p) &= TP_{rIIB}(T,p) + TP_{mIIB}(T,p) \\ &= (p-c)D + i_e p\Big(\frac{DT}{2} + D(R-T)\Big) - \frac{a_r}{T} - \frac{h_r DT}{2} \\ &\quad - i_v w_2 D(T+R) - \frac{a_m}{T} - \frac{h_m D^2 T}{2P} \end{aligned}$$

4 Model Analysis

Theorem 1 *For given values of p and R, there exists a unique optimal cycle length* $T_i^*(p)$ *which maximizes* $TP_i(T|p),\;\; i = 1, 2, 3, 4$ *provided that* $i_p < \frac{pi_e - h_r}{2w(1-\beta)}$, $i_p > \frac{1}{R}(2 - \frac{1}{i_e})$, $i_e > \frac{h_r}{(p+2w(1+\beta))}$, *and* $i_e < \frac{h_r}{p}$.

Proof For given p and R, the profit function is well defined and continuous in T. Therefore, differentiating $TP_1(T|p)$ with respect to T, we get

$$TP_1^{'}(T|p) = \frac{a_r + a_m}{T^2} - \frac{(a - bp)(ah_m - pbh_m + P(h_r - pI_e - 2I_p w(-1+\beta)))}{2P}$$
$$TP_1^{''}(T|p) = -\frac{2(a_r + a_m)}{T^3} < 0, \quad \text{as } a_r, a_m > 0.$$

Then the optimal value of T, say $T_1^*(p)$, can be obtained by equating $TP_1^{'}(T|p) = 0$ which gives

$$T_1^*(p) = \sqrt{\frac{2P(a_r + a_m)}{(a - bp)(ah_m - pbh_m + P(h_r - pI_e - 2wI_p(-1+\beta)))}} \tag{12}$$

For the delayed payment, when $0 \leq R \leq T$ and $w_2 DT > i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}$ we have

$$TP_2^{'}(T|p) = \frac{1}{2PT^2}(2P(a_r + a_m) - (a - bp)((ah_m - pbh_m + Ph_r)T^2 -$$
$$pI_e P(R^2(-2 + I_p R) + T^2) + pI_p(-R^3 p + 2wT^2(1+\beta))))$$

which gives the optimal cycle length as

$$T_2^*(p) = \sqrt{\frac{(2(a_r + a_m) + R^2(-2I_e + (1 + I_e I_p R)p(a - bp)P))}{(a - bp)(ah_m - pbh_m + P(h_r - I_e p + 2wI_p(1+\beta)))}} \tag{13}$$

On the other hand, for the delayed payment when $0 \leq R \leq T$ and $w_2 DT < i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}$ we have

$$TP_3^{'}(T|p) = \frac{1}{2PT^2}(2P(a_r + a_m) - (a - bp)(pI_e^2 R^3 P + (ah_m - pbh_m + h_r P)T^2$$
$$+pIe(2R^2 p + R^3 p - T^2(p + 2w(1+\beta)))))$$

which gives the optimal cycle length as

$$T_3^*(p) = \sqrt{\frac{(2(a_r + a_m) + I_e R^2(2 + R + I_e R)p(-a + bp)P))}{(a - bp)(ah_m - pbh_m + P(h_r - I_e(p + 2w(1+\beta)))}} \tag{14}$$

For delayed payment, in case of $R > T$, we have

$$TP_4^{'}(T|p) = -\frac{(a-bp)(ah_m - pbh_m + (h_r + pI_e)P)}{2P} + \frac{a_r + a_m}{T^2}$$
$$TP_4^{''}(T|p) = -\frac{2(a_r + a_m)}{T^3} < 0.$$

Hence the optimal cycle length is

$$T_4^*(p) = \sqrt{\frac{2P(a_r + a_m)}{(a-bp)(ah_m - pbh_m + P(h_r - pI_e))}} \tag{15}$$

We now consider the optimal cycle length for the cases discussed above. We draw the following relations from the delayed payment cases and classifications:
$T_2^*(p) \geq R$ and $w_2DT > i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}$ give

$$\triangle_1(p) \equiv R^2\Big[(a-bp)(ah_m - pbh_m + P(h_r - I_e p + 2wI_p(1+\beta)))\Big]$$
$$-\Big(2(a_r + a_m) + R^2(-2I_e + (1 + I_e I_p R)p(a-bp)P)\Big) \leq 0.$$

Similarly, $T_3^*(p) \geq R$ and $i_e p\frac{DR^2}{2} + p\frac{DR^2}{2} > w_2DT$ give

$$\triangle_2(p) \equiv R^2\Big[(a-bp)(ah_m - pbh_m + P(h_r - I_e(p + 2w(1+\beta)))\Big]$$
$$-\Big(2(a_r + a_m) + I_e R^2(2 + R + I_e R)p(-a + bp)P)\Big) \leq 0,$$

and $T_4^*(p) \leq R$ gives

$$\triangle_3(p) \equiv R^2\Big[(a-bp)(ah_m - pbh_m + P(h_r - pI_e))\Big] - 2P(a_r + a_m) \geq 0.$$

Using the values of $\triangle_i(p)$, i = 1 to 3 , we have the following result:

Theorem 2 *For given values of p and R,*

(i) if $\triangle_1(p) \leq 0$, then the optimal cycle length $T^*(p) = T_2^*(p)$;
(ii) if $\triangle_2(p) \leq 0$, then the optimal cycle length $T^*(p) = T_3^*(p)$;
(iii) if $\triangle_3(p) \geq 0$, then the optimal cycle length $T^*(p) = T_4^*(p)$.

5 Numerical Results

To demonstrate the proposed model numerically, we consider the following numerical examples satisfying the conditions in each of the four cases:

Table 1 Optimal results in four different cases (examples)

Case	p^*	T^*	$\Pi_r(T, p)$	$\Pi_m(T, p)$	$\Pi(T, p)$
I	316.265	1.6165	2517.97	4269.39	6787.36
IIA_1	320.262	3.64405	7886.72	799.266	8685.98
IIA_2	304.812	2.82131	6524.22	832.307	7356.53
IIB	303.531	1.16745	7758.46	516.058	8274.51

Example 5.1 (***Case I***: Advanced payment)
$a = 100$; $b = 0.2$; $c = 60$; $a_m = 650$; $a_r = 620$; $h_m = 1.5$; $h_r = 1.8$; $\beta = 0.04$; $P = 600$; $w = 150$; $i_e = 0.15$; $i_p = 0.25$; $R = 2$.

Example 5.2 (***Case IIA***: Delayed payment: $0 \leq R \leq T$ and $w_2 DT > i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}$) $i_v = 0.22$; and all other parameter-values are same as given in Example 5.1.

Example 5.3 (***Case IIA***: Delayed payment: $0 \leq R \leq T$ and $i_e p\frac{DR^2}{2} + p\frac{DR^2}{2} > w_2 DT$) $i_p = 0$ and all other parameter-values are same as given in Example 5.2.

Example 5.4 (***Subcase IIB***: Delayed payment: $T \leq R$) All parameter-values are same as given in Example 5.2.

Table 1 shows the optimal results obtained in Examples 5.1–5.4. It is clear from Table 1 that the whole supply chain's profit attains the maximum in case IIA_1 where the optimal cycle time is longer than the optimal payment time. Also, the optimal payment time is longer than the delayed payment period but $w_2 DT > i_e p\frac{DR^2}{2} + p\frac{DR^2}{2}$ holds true. It is further noted that the variation of the controlling parameter β leads to a critical situation where the requirement of the model flips from delayed payment to advanced payment.

From Table 2 it is clearly observed that, with gradual increment of β, the total optimal profit for delayed payment decreases and the total optimal profit for advanced payment increases. For $\beta = 0.04$, the total profit for delayed payment is 8685.98 and for advanced payment is 6787.36. At $\beta = 0.278112$, the total profit for both the models are same and after that the corresponding total profit for advanced payment becomes higher than that of the delayed payment. At this particular moment, the supply chain will adopt the decision for flipping the model from delayed payment mode to advanced payment mode.

6 Sensitivity Analysis

In this section, we examine the effects of changes in the parameter-values on the optimal decisions of the proposed model. We change one parameter value at a time keeping the other parameter-values unchanged.

Table 2 Critical value for β

Value of β	Delayed payment			Advanced payment ($R < T$)		
	p^*	T^*	Π^*	p^*	T^*	Π^*
0.04	320.26	3.644	8685.98	316.26	1.616	6787.36
0.06	320.19	3.564	8595.35	315.99	1.662	6854.63
0.08	320.72	3.490	8506.91	315.73	1.711	6923.30
0.15	321.65	3.264	8212.76	315.08	1.933	7176.77
0.25	323.14	3.010	7827.57	315.59	2.525	7590.80
0.27	323.46	2.966	7754.74	316.20	2.739	7685.44
0.275	323.54	2.956	7736.73	316.41	2.803	7710.04
0.277	323.57	2.951	7729.55	316.50	2.830	7719.99
0.278	323.59	2.949	7725.96	316.54	2.844	7725.00
0.278112	323.59	2.949	7725.56	316.55	2.845	7725.56
0.279	323.60	2.947	7722.38	316.59	2.858	7730.02
0.28	323.62	2.945	7718.80	316.64	2.872	7735.06
0.29	323.78	2.924	7683.17	317.23	3.032	7786.55
0.31	324.10	2.885	7612.80	319.12	3.487	7896.79

Table 3 Sensitivity analysis: advanced payment

Parameter	Value	Delayed payment			Advanced payment ($R < T$)		
		p^*	T^*	Π^*	p^*	T^*	Π^*
a	92	296.35	1.624	5228.96	298.18	3.400	6796.48
	96	306.27	1.617	5985.73	309.18	3.517	7711.49
	100	316.26	1.616	6787.36	320.26	3.644	8685.98
	104	326.33	1.621	7633.83	331.44	3.783	9721.20
	108	336.46	1.632	8525.17	342.73	3.936	10818.50
b	0.19	330.09	1.672	7476.54	335.14	3.848	9541.68
	0.195	322.98	1.642	7122.28	324.49	3.740	9100.61
	0.20	316.26	1.616	6787.36	320.26	3.644	8685.98
	0.205	309.90	1.594	6470.27	313.42	3.556	8295.53
	0.21	303.85	1.574	6169.69	306.94	3.476	7937.24
R	1.90	315.54	1.610	6840.38	318.06	3.310	8545.44
	1.95	315.90	1.613	6813.84	319.18	3.475	8613.19
	2	316.26	1.616	6787.36	320.26	3.644	8685.98
	2.05	316.62	1.620	6760.93	321.32	3.817	8763.92
	2.10	316.98	1.623	6734.55	322.34	3.995	8847.11

The following observations are made from the results presented in Table 3 for advanced payment.

- The selling price increases when the basic market demand a increases. A higher basic market demand generates higher total demand. The higher selling price increases the profit as well as the cycle length for both the cases.

- The sensitivity coefficient of the selling price b has direct impact on the market demand. For higher value of b, the demand is low. Then the retailer must decrease his selling price in order to protect the market demand. As a result, the net profit and the cycle length decrease. These observations are similar in both the cases.
- When the optimal payment time increases, the optimal selling price and the cycle length increase whereas the whole system's profit decreases nominally in delayed payment mode. But in advanced payment mode, the whole system's profit increases.

7 Conclusions

In this study, we develop a two-level supply chain model with one retailer and one manufacturer for trading a single product, where the system offers advanced payment and delayed payment options. The model will provide an expert decision making machinery depending on the available input parameters.

Here one advanced payment option and three delayed payment modes are considered with the relationship amongst the cycle length T and payment time R. The study reveals that the whole supply chain's profit attains the maximum where the optimal cycle time is longer than the optimal payment time. Payment in advanced mode implies that the manufacturer's profit is higher than the retailer's profit and naturally in the delayed mode, the retailer's profit is higher than the manufacturer's profit.

Most importantly, it is identified that for a critical value of β the model should flip its character in view of the total profit of the supply chain.

For future research, the model can be extended in several ways. For instance, one can incorporate shortages, deterioration in inventory, cash discounts, non-linear demand pattern, etc. in this model. Moreover, one can extend the proposed model by considering credit period with variations. Upstream partial prepayment and downstream partial delay payment will also be useful considerations in extending the model.

References

1. Goyal, S.K.: Economic order quantity under conditions of permissible delay in payment. J. Oper. Res. Soc. **36**, 335–338 (1985)
2. Jamal, A.M.M., Sarker, B.R., Wang, S.: Optimal payment time for a retailer under permitted delay of payment by the wholesaler. Int. J. Prod. Econ. **66**, 59–66 (2000)
3. Teng, J.T., Chang, C.T., Goyal, S.K.: Optimal pricing and ordering policy under permissible delay in payments. Int. J. Prod. Econ. **97**, 121–129 (2005)
4. Cheng, M.C., Lou, K.R., Ouyang, L.Y., Chiang, Y.H.: The optimal ordering policy with trade credit under two different payment methods. TOP **18**, 413–428 (2010)
5. Chung, K.J.: The correct proofs for the optimal ordering policy with trade credit under two different payment methods in a supply chain system. TOP **20**, 768–776 (2012)

6. Huang, Y.F.: Economic order quantity under conditionally permissible delay in payments. Eur. J. Oper. Res. **176**, 911–924 (2007)
7. Giri, B.C., Maiti, T.: Supply chain model with price- and trade credit-sensitive demand under two-level permissible delay in payments. Int. J. Syst. Sci. **44**, 937–948 (2013)
8. Giri, B.C., Sharma, S.: Optimal ordering policy for an inventory system with linearly increasing demand and allowable shortages under two levels trade credit financing. Oper. Res. **16**, 25–50 (2016)
9. Yang, C.T.: The optimal order and payment policies for deteriorating items in discount cash flows analysis under the alternatives of conditionally permissible delay in payments and cash discount. TOP **18**, 429–443 (2010)
10. Sarker, B.R., Jamal, A.M.M., Wang, S.: Optimal payment time under permissible delay in payment for products with deterioration. Prod. Plan. Control **11**, 380–390 (2000)
11. Liao, H.C., Chen, Y.K.: Optimal payment time for retailer's inventory system. Int. J. Syst. Sci. **34**, 245–253 (2003)
12. Huang, Y.F.: Buyer's optimal ordering policy and payment policy under supplier credit. Int. J. Syst. Sci. **36**, 801–807 (2005)
13. Chang, C.T., Wu, S.J., Chen, L.C.: Optimal payment time with deteriorating items under inflation and permissible delay in payments. Int. J. Syst. Sci. **40**, 985–993 (2009)
14. Maiti, A.K., Maiti, M.K., Maiti, M.: Inventory model with stochastic lead-time and price dependent demand incorporating advance payment. Appl. Math. Model. **33**, 2433–2443 (2009)
15. Giri, B.C., Bhattacharjee, R., Maiti, T.: Optimal payment time in a two-echelon supply chain with price-dependent demand under trade credit financing. Int. J. Syst. Sci. Oper. Logist. **5**(4), 374–392 (2018)
16. Li, R., Chan, Y.L., Chang, C.T., Barron, L.E.C.: Pricing and lot-sizing policies for perishable products with advance-cash-credit payments by a discounted cash-flow analysis. Int. J. Prod. Econ. **193**, 578–589 (2017)
17. Teng, J.T., Barron, L.E.C., Chang, H.J., Wu, J., Hu, Y.: Inventory lot-size policies for deteriorating items with expiration dates and advance payments. Appl. Math. Model. **40**, 8605–8616 (2016)
18. Zhang, Q., Zhang, D., Tsao, Y.C., Luo, J.: Optimal ordering policy in a two-stage supply chain with advance payment for stable supply capacity. Int. J. Prod. Econ. **177**, 34–43 (2016)

Studies on Atherosclerotic Plaque Formation: A Mathematical Approach

Debasmita Mukherjee, Lakshmi Narayan Guin and Santabrata Chakravarty

Abstract Atherosclerosis is one of the main causes behind several cardiovascular diseases (CVDs). It occurs due to plaque accumulation in the innermost layer of artery, known as intima. The present study deals with a qualitative mathematical model consisting of a system of ten nonlinear ordinary differential equations (ODEs) involving ten important interacting cellular components coming into play namely, low density lipoprotein, free radicals, chemoattractants, monocytes, macrophages, T-cells, smooth muscle cells, foam cells and collagen in order to describe the evolution of atherosclerotic plaque. Numerical results obtained for this model are shown through several plots represented by phase portraits of some subsystems. As the present model is encountered with various important cellular components involving atherosclerosis, so it is believed to provide a platform for in silico treatment of this pernicious disease.

Keywords Atherosclerosis · Non-linear ODE system · Routh-Hurwitz criterion

1 Introduction

Atherosclerosis is a chronic vascular disease which is the main cause behind deaths from cardiovascular diseases (CVDs) in the form of heart attack (myocardial infarction) and stroke (cerebrovascular accident), worldwide. Previous studies recorded that 31% of deaths occur due to CVDs.

D. Mukherjee (✉) · L. N. Guin · S. Chakravarty
Department of Mathematics, Visva Bharati, Santiniketan, WB, India
e-mail: debasmita.mukherjee24@gmail.com

L. N. Guin
e-mail: guin_ln@yahoo.com

S. Chakravarty
e-mail: santabrata2004@yahoo.co.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_26

When cholesterol level increases in blood, mainly bad cholesterol (or LDLs) passes through endothelium lesions and comes into the intima. In intima, LDLs get oxidised with free radicals and thus formed new species, called oxidised LDLs. In presence of oxidised LDL, endothelium layer and SMCs secrete chemoattractants. Chemoattractants attracts monocytes and T-cells from lumen to intima. In intima, monocytes diferentiate into macrophages. These macrophages have an affinity towards ox-LDL. They absorb ox-LDL via scavenger receptor and form a fatty steak called foam cells. SMCs move from media to intima in the presence of chemoattractant in the intima and they form a fibrous cap surrounding the foam cells. The foam cell surrounded with fibrous cap together is called plaque, when this plaque accumulation increases, the shear stress of endothelium layer crosses its limit and then plaque bulges into lumen. Plaque is thrombogenic and it hinders the smooth blood flow in the lumen. The process is described in Fig. 1.

Quite a good number of previous model studies in terms of ODEs and PDEs are available in the recent past where this entire biochemical process is described in mathematical terms. To name a few, [2, 3, 5, 6] have described the formation of atherosclerotic plaques in terms of ODE models whereas [4] has provided both ODE and PDE models of early stage of this disease. References [1, 7, 8] have decribed the models of atherosclerosis in terms of PDE models. In [9] an extensive detailed list of previous models and their significance are provided.

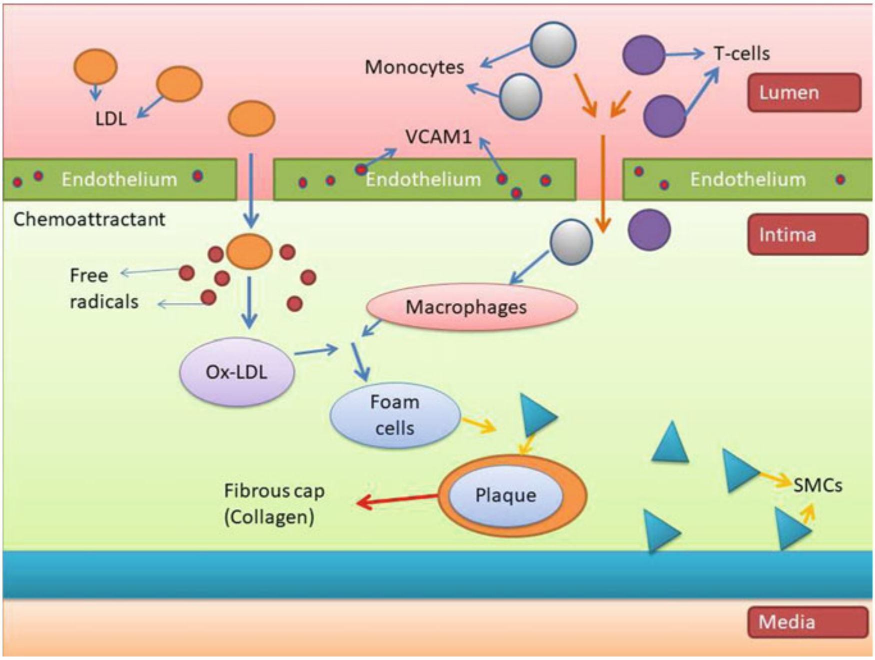


Fig. 1 Various stages of atherosclerotic plaque formation

In this paper, the entire biochemical process is described through ten dimensional nonlinear autonomous system of ODEs and it is a modified model described in [4]. Model equations for collagen and foam cells are significantly modified and there are fewer modification in the overall model. The present model is found to be locally stable from both theoretical and numerical points of view. As the system is ten dimensional, so several subsystems are considered to reveal the complex features of this model. Most of the previous models involved only foam cells as the output of atherosclerosis whereas this current model has taken care of foam cells and the fibrous cap surrounded the plaque made up from collagen as two separate equations. The aim of this present article is to provide only a stable mathematical model to contribute in the clinical investigation on this topic.

The paper is organised as follows, in Sect. 2, the model is formulated through a system of ten nonlinear ODEs, in Sect. 3 the model is nondimesionalised and in Sect. 4 stability analysis has been performed. In Sect. 5 numerical simulation is explained at length while in Sect. 6, concluding remarks is provided.

2 Formulation of the Model with Proper Justification

To understand the dynamics of biochemical process an autonomous system of ten non-linear ODEs is presented here. Among several important components involved in atherosclerotic plaque formation the components like LDLs, free radicals, modified/oxidised LDLs, chemoattractants, monocytes, macrophages, T-cells, smooth muscle cells, foam cells and collagens are chosen as dependent variables. The model is described as follows:

$$\frac{d\tilde{L}}{dt} = \underbrace{\sigma_L}_{\text{LDL resource}} - \underbrace{k_L \tilde{R}\tilde{L}}_{\text{LDL interaction with radical}} - \underbrace{d_L \tilde{L}}_{\text{diffusion of LDL}}, \tag{1}$$

$$\frac{d\tilde{R}}{dt} = \underbrace{\sigma_R}_{\text{radical resource}} - k_L \tilde{R}\tilde{L} - \underbrace{d_R \tilde{R}}_{\text{diffusion of radical}}, \tag{2}$$

$$\frac{d\tilde{X}}{dt} = k_L \tilde{R}\tilde{L} - \underbrace{\rho_{in} \tilde{M}\tilde{X}}_{\text{ingestion of oxLDL by macrophages}} - \underbrace{d_X \tilde{X}}_{\text{diffusion of oxLDL}}, \tag{3}$$

$$\frac{d\tilde{C}}{dt} = \underbrace{\rho_C \tilde{X}}_{\text{insertion of chemoattractant in intima}} - \underbrace{d_C \tilde{C}}_{\text{death of chemoattractant}}, \tag{4}$$

$$\begin{aligned}\frac{d\tilde{m}}{dt} = {} & \underbrace{\rho_m \tilde{C}}_{\text{insertion of monocytes into intima}} - \underbrace{\rho_M \tilde{m}}_{\text{differentiation of monocytes into macrophages}} \\ & - \underbrace{d_m \tilde{m}}_{\text{death of monocytes}}, \end{aligned} \tag{5}$$

$$\frac{d\tilde{M}}{dt} = \underbrace{\rho_M \tilde{m}}_{\text{differentiation of monocytes into macrophages}} - \rho_{in}\tilde{M}\tilde{X} - \underbrace{d_M \tilde{M}}_{\text{death of macrophages}}, \tag{6}$$

$$\frac{d\tilde{T}}{dt} = \underbrace{\rho_T \tilde{C}}_{\text{insertion of T cell into intima}} - \underbrace{d_T \tilde{T}}_{\text{death of T cells}}, \tag{7}$$

$$\frac{d\tilde{S}}{dt} = \underbrace{\rho_S \tilde{C}}_{\text{insertion of SMCs into intima}} - \underbrace{m_S \tilde{S}}_{\text{migration of SMCs from media to intima}} - \underbrace{d_S \tilde{S}}_{\text{death of SMCs}}, \tag{8}$$

$$\frac{d\tilde{F}}{dt} = \alpha\rho_{in}\tilde{M}\tilde{X} - \underbrace{d_F \tilde{F}}_{\text{death of foam cells}}, \tag{9}$$

$$\frac{d\tilde{G}}{dt} = \beta\rho_{in}\tilde{M}\tilde{X} + m_S\tilde{S} - \underbrace{d_G \tilde{G}}_{\text{degradation of collagen}}. \tag{10}$$

Here $\tilde{L}$, $\tilde{R}$, $\tilde{X}$, $\tilde{C}$, $\tilde{m}$, $\tilde{M}$, $\tilde{T}$, $\tilde{S}$, $\tilde{F}$ and $\tilde{G}$ represent the respective time-dependent concentration of LDLs, free radicals, oxidized LDLs, chemoattractants, monocytes, macrophages, T-cells, smooth muscle cells, foam cells and collagens and the remaining parameters of the model have their nomenclatures listed in Table 2. All the above Eqs. (1)–(10) represent the respective rate of concentration for all the cellular components arising from the interplay among the components in the entire biochemical process. The genesis of these equations are further explained sequentially so as to have complete understanding of the model formulation. Equation (1) depicts constant source of LDL, interaction of LDL with radicals and diffusion term for LDL. Equation (2) has similar explanations to that of (1) for radicals. In Eq. (3), first term is the formation term of oxidised LDL from interaction of LDL and radicals, second term is the loss term due to ingestion by macrophages and the third term is also the loss term due to its natural diffusion. Equation (4) refers insertion of chemoattractant in intima depending on the presence of ox-LDL in the intima followed by its loss due to death. Equation (5) represents the source term for monocyte concentration as proportional to that of chemoattractant in the intima, then its loss due to its differentiation into macrophages followed by its natural death. In the Eq. (6) first term is the source term for macrophages obtained from differentiation of monocytes, then loss of macrophages due to its involvement in the ingestion process followed by its natural death term. Equation (7) models the source term for T-cell concentration in intima as proportional to the presence of chemoattractant followed by its natural death term. Equation (8) represents concentration of SMCs in the intima as proportional to the presence of chemoattractant, then its loss due to its involvement in the collagen formation and the third term stands for its natural death. Equation (9) shows the concentration of foam cells, its source term is the ingestion term obtained from macrophage and ox-LDL followed by its natural death. The last equation of this current model equation (10) depicts the collagen concentration, first term is a fractional part of the ingested term, second term is because SMCs are involved in the process of collagen formation and finally its own natural degradation term. Here $\alpha\rho_{in}\tilde{M}\tilde{X}$ in (9) and $\beta\rho_{in}\tilde{M}\tilde{X}$ in (10) represent the fractional part of ox-LDL ingested macrophages contributing into atherosclerotic foam cell formation and fibrous cap formation, where $\alpha + \beta = 1$.

3 Rescaled Model

For the purpose of introduction of rescaling, we introduce two parameters r (sec^{-1}) and δ (concentration^{-1}). The time variable t is changed to $\tau = rt$, in which r has unit sec^{-1}. The cell concentrations are changed as $y_i = \frac{\tilde{y_i}}{\delta}$, where $y_i = L,\ R,\ X,\ C,\ m,\ M,\ T,\ S,\ F,\ G$ for $i = 1, 2, \ldots, 10$ respectively. The value of r is chosen to be equal to 6×10^{-7} sec^{-1} as in [8] which leads to estimate the value of non-dimensional time $\tau = 1$ when t is nearly 19 days approximately. The respective system of rescaled equations should now be read as follows:

$$\frac{dL}{d\tau} = u_{11} - u_{12}RL - u_{13}L, \tag{11}$$

$$\frac{dR}{d\tau} = u_{21} - u_{12}RL - u_{23}R, \tag{12}$$

$$\frac{dX}{d\tau} = u_{12}RL - u_{32}MX - u_{33}X, \tag{13}$$

$$\frac{dC}{d\tau} = u_{41}X - u_{42}C, \tag{14}$$

$$\frac{dm}{d\tau} = u_{51}C - a_{52}m - u_{53}m, \tag{15}$$

$$\frac{dM}{d\tau} = u_{52}m - u_{32}MX - u_{63}M, \tag{16}$$

$$\frac{dT}{d\tau} = u_{71}C - u_{72}T, \tag{17}$$

$$\frac{dS}{d\tau} = u_{81}C - u_{82}S - u_{83}S, \tag{18}$$

$$\frac{dF}{d\tau} = \alpha u_{32}MX - u_{92}F, \tag{19}$$

$$\text{and } \frac{dG}{d\tau} = \beta u_{32}MX + u_{82}S - u_{103}G \tag{20}$$

where,

$$u_{11} = \sigma_L/\delta r, u_{12} = k_L\delta/r, u_{13} = d_L/r, u_{21} = \sigma_R/\delta r, u_{23} = d_R/r$$
$$u_{32} = \rho_{in}\delta/r, u_{33} = d_X/r, u_{41} = \rho_C/r, u_{42} = d_C/r, u_{51} = \rho_m/r$$

$$u_{52} = \rho_M/r, u_{53} = d_m/r, u_{63} = d_M/r, u_{71} = \rho_T/r, u_{72} = d_T/r$$
$$u_{81} = \rho_S/r, u_{82} = m_S/r, u_{83} = d_S/r, u_{92} = d_F/r, u_{103} = d_G/r.$$

4 Stability Analysis

The following theorem shows that under certain suitable conditions, all the solutions of the system of equations (11)–(20) are non-negative. A set Γ in $\mathbb{R}^{10}_+$ may be ascertained such that all solutions starting from Γ remain bounded.

Theorem 1 *Let all the parameters of the system of equations* (11)–(20) *be positive and* Γ *be a region in* $\mathbb{R}^{10}_+$ *defined as,* $\Gamma = \{(L, R, X, C, m, M, T, S, F, G) \in \mathbb{R}^{10}_+ | 0 \leq L \leq \bar{L}, 0 \leq R \leq \bar{R}, 0 \leq X \leq \bar{X}, 0 \leq C \leq \bar{C}, 0 \leq m \leq \bar{m}, 0 \leq M \leq \bar{M}, 0 \leq T \leq \bar{T}, 0 \leq S \leq \bar{S}, 0 \leq F \leq \bar{F}, 0 \leq G \leq \bar{G}\}$. *Then* Γ *is positive invariant and all the solutions starting from* Γ *are uniformally bounded, the parameters over bar are being the respective upper bounds.*

Proof First one may make an attempt to prove the positive invariant part.
Let $(L(0), R(0), X(0), C(0), m(0), M(0), T(0), S(0), F(0), G(0)) \in \Gamma$.
If possible, suppose $L(\tau)$ be non positive. Then there exists $\tau_0 > 0$ such that $L(\tau_0) = 0$ and $L(\tau) > 0$ for any τ satisfying $0 \leq \tau \leq \tau_0$.Then necessarily $\frac{dL}{d\tau}|_{\tau=\tau_0} \leq 0$. This is a contradiction, because
$\frac{dL}{d\tau}|_{\tau=\tau_0} = u_{11} - u_{12}R(\tau_0)L(\tau_0) - u_{13}L(\tau_0) = u_{11} > 0.$
Hence, $L(\tau)$ is positive $\forall \tau \geq 0$.
In exactly similar way $R(\tau)$, $X(\tau)$, $C(\tau)$, $m(\tau)$, $M(\tau)$, $T(\tau)$, $S(\tau)$, $F(\tau)$ and $G(\tau)$ are positive $\forall \tau \geq 0$ can be shown, so these proves are omitted here.

Next the part of boundedness may be shown as follows.
From equation (11), one may have
$\frac{dL}{d\tau} = u_{11} - u_{12}RL - u_{13}L \leq u_{11} - u_{13}L.$
$\implies \frac{dL}{d\tau} + u_{13}L \leq u_{11}.$
$\implies L(\tau) \leq \frac{u_{11}}{u_{13}} + \kappa_2 e^{(-u_{13}\tau)}$, where κ_2 is a positive integrating constant. The term $\kappa_2 e^{(-u_{13}\tau)}$ vanishes as $\tau \to \infty$. So, $L(\tau) \leq \frac{u_{11}}{u_{13}}$ as, $\tau \to \infty$, i.e $L(\tau)$ remains bounded $\forall \tau \geq 0$.

Theorem 2 *The system represented by* (11)–(20) *is locally asymptotically stable.*

Proof To prove the system is locally stable, one needs to consider the corresponding linearised system. Let us suppose the Jacobian matrix corresponding to (11)–(20) is J. After evaluating this Jacobian matrix at the equlibrium positions of the system (11)–(20) one gets the characteristic polynomial at each of these equlibrium. As the calculation is quite complex, so here a general approach is provided to prove the stability of this system. Denote the jacobian matrix obtained

Table 1 Routh table of V^*

x^{10}	v_{101}	v_{102}	v_{103}	v_{104}	v_{105}	v_{106}
x^9	v_{91}	v_{92}	v_{93}	v_{94}	v_{95}	0
x^8	v_{81}	v_{82}	v_{83}	v_{84}	v_{85}	0
x^7	v_{71}	v_{72}	v_{73}	v_{74}	0	0
x^6	v_{61}	v_{62}	v_{63}	v_{64}	0	0
x^5	v_{51}	v_{52}	v_{53}	0	0	0
x^4	v_{41}	v_{42}	v_{43}	0	0	0
x^3	v_{31}	v_{32}	0	0	0	0
x^2	v_{21}	v_{22}	0	0	0	0
x	v_{11}	0	0	0	0	0
1	v_{01}	0	0	0	0	0

at equlibrium point by J^*. Suppose the characteristic polynomial of J^* may be expressed as, $x^{10} + c_9x^9 + c_8x^8 + c_7x^7 + c_6x^6 + c_5x^5 + c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0 = g(x)$, where $c_k = (-1)^{10-k} \sum_{|\mathcal{D}|=10-k} V^*[\mathcal{D}]$, i.e, c_k are the sum of all principal minors of order $(10-k)$ of the matrix V^* multiplied by $(-1)^{10-k}$. Here $\mathcal{D} \subseteq \{1, 2, \ldots, 10\}$ and $V^*[\mathcal{D}]$ denotes the principal minors of V^* formed by the columns and rows with indices from $\mathcal{D}$. Let us denote, $1 = v_{101}$, $c_8 = v_{102}$, $c_6 = v_{103}$, $c_4 = v_{104}$, $c_2 = v_{105}$, $c_0 = v_{106}$, $c_9 = v_{91}$, $c_7 = v_{92}$, $c_5 = v_{93}$, $c_3 = v_{94}$, $c_1 = v_{95}$. Then Routh table of the polynomial $g(x)$ may be formed as in Table 1; where $v_{ij} = \frac{v_{(i+1)j}v_{(i+2)(j+1)} - v_{(i+1)(j+1)}v_{(i+2)(j)}}{v_{(i+1)j}}$, for $i = 0, 1, 2, \ldots, 10$ and $j = 1, 2, \ldots, 6$. Routh-Hurwitz criterion states that a system is stable if and only if all terms in the first column are positive and if these conditions are not satified then there will be an unstable solution. Numerical justification of this theorem is shown in next section by considering several subsytems and showing the behaviour of phase portraits for different choice of initial conditions around the non-zero equilibrium point.

5 Numerical Simulation and Discussion of the Results

The model parameters involved in system (11)–(20) are not available in the existing literatures for human models, so few of them are collected from previous studies and rest of them are assumed for this particular model and as provided in Table 2. The model equations have been solved numerically using Runge-Kutta 4th order method. The non-zero equilibrium point is found to be, $v_0 = [L = 0.414714574672321,\ R = 2.28646012986491 \times 10^{-5},$

Table 2 List of parameter values used in the model

Parameters	Descriptions	Numeric values	Source
u_{11}	Constant source of LDL	64	[3]
u_{12}	Reaction rate of LDL and radicals	0.5×10^7	[3]
u_{13}	Death rate of LDL	40	[3]
u_{21}	Constant source of radicals	100	Present study
u_{23}	Death rate of radicals	2.3×10^6	[3]
u_{32}	Ingestion rate of macrophages and ox-LDL	10^4	[4]
u_{33}	Death rate of ox-LDL	40	Present study
u_{41}	Rate of production of chemoattractant	1000	Present study
u_{42}	Death rate of chemoattractant	10	Present study
u_{51}	Monocyte insertion rate into intima inversely varying with WSS	1000	Present study
u_{52}	Differentiation rate of monocyte into macrophages	2	[5]
u_{53}	Death rate of monocytes	10	Present study
u_{63}	Death rate of macrophages	0.1	[8]
u_{71}	Rate of production of T-cells	10	Present study
u_{72}	Death rate of T-cell	7	[1]
u_{81}	Rate of production of SMCs	10	Present study
u_{82}	Mmigration rate of SMCs from media to intima	0.1	Present study
u_{83}	Death rate of SMCs	17	[1]
u_{92}	Death rate of foam cells	0.6	[1]
u_{103}	Death rate of of collagen	0.05	Present study
α	Fractional part of lipid core contributing in plaque formation	0.75	Present study
β	Fractional part of lipid core contributing in fibrous cap formation	0.25	Present study

$X = 0.00284781527236790,\ C = 0.284781527236790,$
$m = 23.7317939363991,\ M = 1.66083470585807,\ T = 0.406830753195414,$
$S = 0.166538904816836,\ F = 59.1218805027656,\ G = 236.820599820696]$.
It is sufficient to take the time-series plots over a span of 400 dimensionless time-scale because it is equivalent to 20 years real time and an atherosclerotic plaque takes $14 - 15$ years to grow fully. So the figures presented here are taken over 400 dimensionless time-scale. Figure 2 represents time -series plots of all cellular components involved in the present model, (a) for larger time-scale $\tau = 400$ and (b) for smaller time-scale 20. Similarly, Fig. 3 represent time series plots of LDLs, radicals, ox-LDL, chemoattractants, T-cells and SMCs which are separately provided because they are not clearly visible in the Fig. 2.

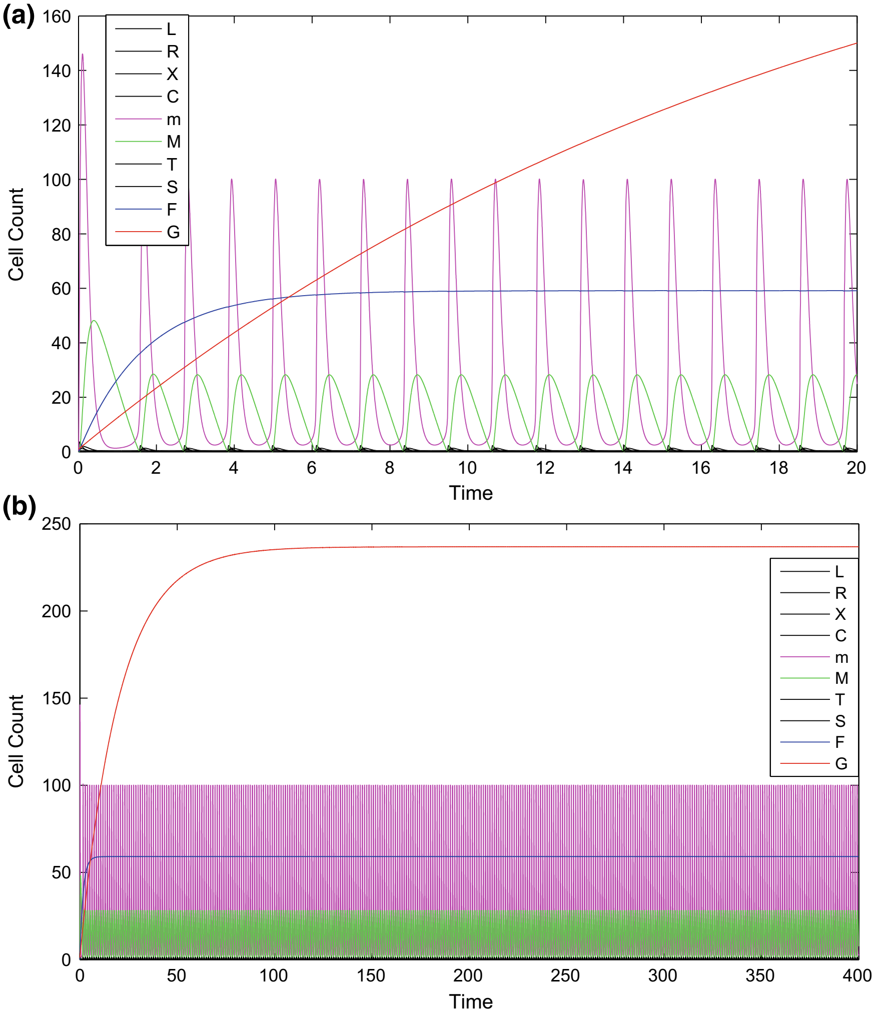


Fig. 2 Time series representations of all cellular species corresponding to **a** smaller ($\tau = 20$) and **b** larger ($\tau = 400$) span of time

The system (11)–(20) is a stable system which has been shown theoretically in Theorem 4.2. Also it has been found that the Jacobian matrix corresponding to the system (11)–(20) has either negative real eigenvalues or complex conjugates with negative real parts at ν_0, which suffices the stability nature of the formulated model. To show the stable nature of this particular model, several phase portraits of two-dimensional or three- dimensional corresponding to some subsytems considered in random order are shown. Figure 4 shows the two-dimensional phase

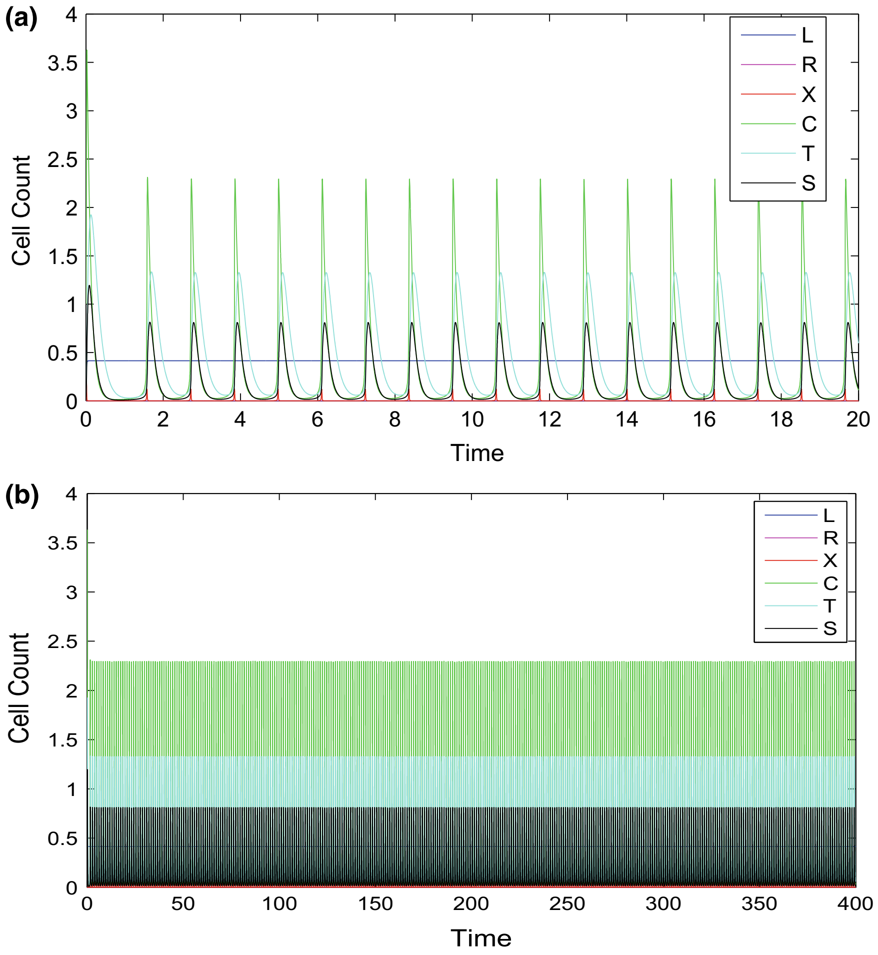


Fig. 3 Time series representation of the concentrations of rest of cellular species of the model not clearly visible in Fig. 2 for **a** $\tau = 20$ and **b** $\tau = 400$ using parameter value of Table 2

portraits of few subsytems converging around the non-zero equilibrium point for different initial conditions. Similarly, Fig. 5 shows the same for three-dimensional phase portraits.

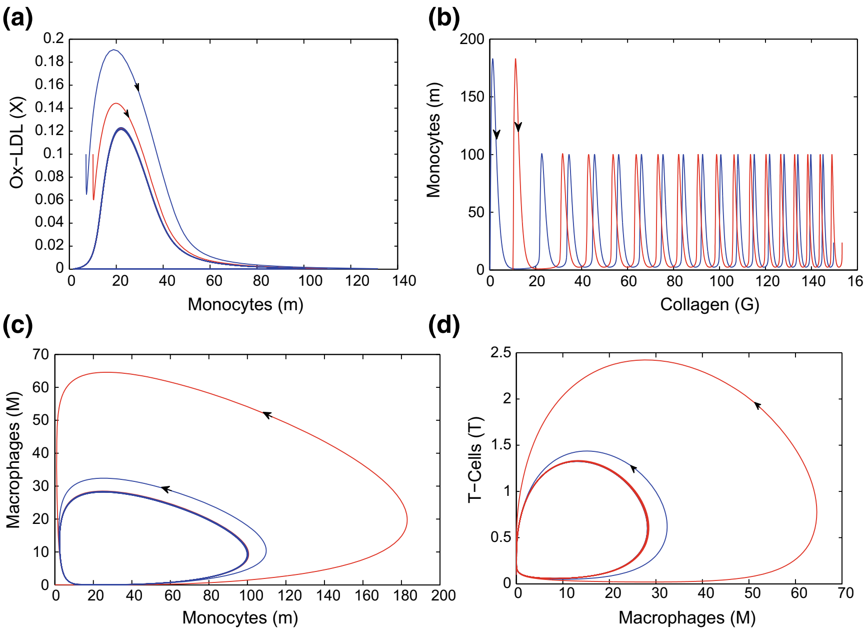


Fig. 4 Two dimensional projection of the phase portraits for several subsystems arising from the original model corresponding to **a** (m,X) **b** (G,m), **c** (m,M), **d** (M,T)-spaces by using parameter values of Table 2 for two different initial conditions

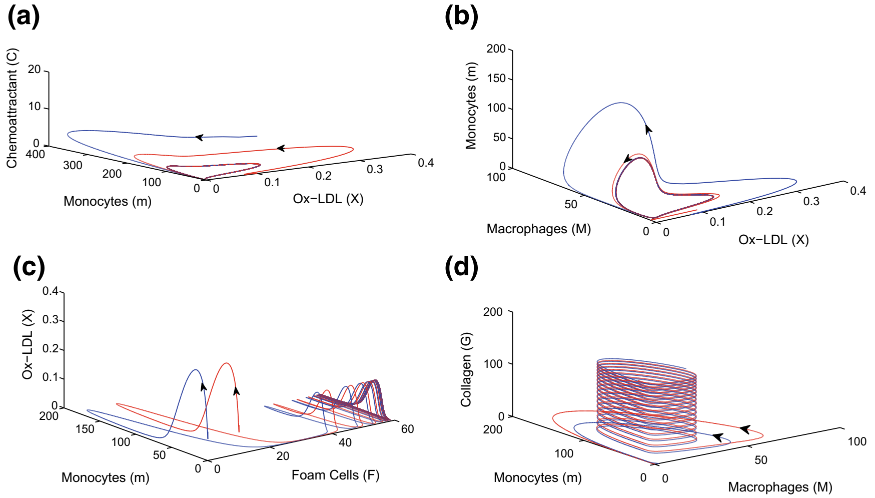


Fig. 5 Three dimensional projection of the phase portraits for several subsystems arising from the original model corresponding to **a** (X,m,C), **b** (X,M,m), **c** (F,m,X), **d** (M,m,G)-spaces by using parameter values of Table 2 for two different initial conditions

6 Concluding Remarks

In the presented article, a mathematical model comprising a system of ten nonlinear ODEs describing the early stages of atherosclerotic plaque formation is provided in details. The model is being proved to have all the cellular components positive, bounded and the system is locally stable. The purpose of this article is to provide a stable system to enhance better understanding on the dynamics of atherosclerotic plaque. The model formulated in the present study has been shown to be stable both theoretically and numerically. Normal diffusion has been applied in the spatial atherosclerotic dynamics by some investigators [1, 7, 8], which is a more general form for describing the diffusion phenomena in the nature. This issue needs further investigation and discussion and hence has a future scope of research.

Acknowledgements This paper was presented in ICMAAM-2018. The authors gratefully acknowledge the financial support by Special Assistance Programme (SAP-III) sponsored by the University Grants Commission (UGC), New Delhi, India [Grants no. F.510/3/DRS-III/2015(SAP-I)].

References

1. Hao, W., Friedman, A.: The LDL-HDL profile determines the risk of atherosclerosis: a mathematical model. PLoS One **9**(3), e90497 (2014)
2. Ougrinovskaia, A., Thompson, R.S., Myerscough M.R.: An ode model of early stages of atherosclerosis: mechanisms of the inflammatory response. Bull. Math. Biol. **72**(6), 1534–1561 (2010)
3. Cobbold, C.A., Sherratt, J.A., Maxwell, S.R.J.: Lipoprotein oxidation and its significance for atherosclerosis: a mathematical approach. Bull. Math. Biol. **64**(1), 65–95 (2002)
4. McKay, C., McKee, S., Mottram, N., Mulholland, T., Wilson, S., Kennedy, S., Wadsworth, R.: Towards a Model of Atherosclerosis, pp. 1–29. University of Strathclyde (2005)
5. Bulelzai, M.A.K., Dubbeldam, J.L.A.: Long time evolution of atherosclerotic plaques. J. Theor. Biol. **297**, 1–10 (2012)
6. Anlamlert, W., Lenbury, Y., Bell, J.: Modeling fibrous cap formation in atherosclerotic plaque development: stability and oscillatory behavior. Adv. Diff. Eq. **2017**(1), 195 (2017)
7. Friedman, A., Hao, W.: A mathematical model of atherosclerosis with reverse cholesterol transport and associated risk factors. Bull. Math. Biol. **77**(5), 758–781 (2015)
8. Chalmers, A.D., Cohen, A., Bursill, C.A., Myerscough, M.R.: Bifurcation and dynamics in a mathematical model of early atherosclerosis. J. Math. Biol. **71**(6–7), 1451–1480 (2015)
9. Parton, A., McGilligan, V., Okane, M., Baldrick, F.R., Watterson, S.: Computational modelling of atherosclerosis. Brief. Bioinform. **17**(4), 562–575 (2015)

Threshold of a Stochastic Delayed SIR Epidemic Model with Saturation Incidence

Yanli Zhou

Abstract A stochastic SIR epidemic model with time delay and saturation incidence is formulated in this paper. We show that the disease dynamics of the stochastic delayed SIR model can be governed by its related threshold R_0^S, whose value completely determines the disease to go extinct and prevail for any size of the white noise. The results are improved and the method is simpler than the previously-known literature. And the related results recover the known results in the earlier literature as special cases. The presented results are illustrated by numerical simulations.

Keywords The threshold · Stochastic SIR epidemic model · Time delay · Saturation incidence

1 Introduction

In recent years, stochastic epidemic system driven by Brownian motion has been received great attention and has been studied extensively, see [1–10]. Under some simple conditions, the authors showed that Brownian motion can suppress the explosion of the solution. These important results reveal that environmental variations may have significant impacts on the properties of epidemic systems.

For human disease, the nature of epidemic growth and spread is inherently random due to the unpredictability of connections between people and population is subject to a continuous spectrum of disturbances. Hence the variability and randomness of the environment are fed through to the state of the epidemic. Stochastic differential equation (SDE) models could be a more appropriate way of modeling epidemics in many circumstances.

There are different possible approaches to introduce random effects in the epidemic models affected by environmental white noise from biological significance and mathematical perspective. Some scholars demonstrated that one or more system

Y. Zhou (✉)
Shanghai University of Medicine and Health Sciences, Shanghai 201318, China
e-mail: zhouyanli_math@163.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_27

parameter(s) can be perturbed stochastically with white noise term to derive environmentally perturbed system. In [3], Jiang et al. investigated a stochastic epidemic SIR models by introducing randomness into the natural death rates, where they proved existence of the positive solution and discussed the asymptotic behavior. In [11], their method to include stochastic perturbation is similar to that of Jiang et al. [3]. They discussed the asymptotic properties of a stochastic delayed SIR epidemic model with temporary immunity. The studied model in [11] is as follows:

$$\begin{cases} \mathrm{d}S(t) = \big[\Lambda - \mu S(t) - \beta S(t)I(t) + \gamma e^{-\mu\tau} I(t-\tau)\big]\mathrm{d}t + \sigma_1 S(t)\mathrm{d}B_1(t), \\ \mathrm{d}I(t) = \big[\beta S(t)I(t) - (\mu+\gamma+\delta)I(t)\big]\mathrm{d}t + \sigma_2 I(t)\mathrm{d}B_2(t), \\ \mathrm{d}R(t) = \big\{[\gamma I(t) - \gamma e^{-\mu\tau} I(t-\tau)] - \mu R(t)\big\}\mathrm{d}t + \sigma_3 R(t)\mathrm{d}B_3(t), \end{cases} \tag{1.1}$$

where S, I, R denote the host population the susceptible, the infective and the recovered sub population, respectively. Λ is the birth rate, μ is the natural rates, δ is mortality rate induced by the disease, β denotes the transmission coefficient between compartments S and I, γ is the recovery rate of the infective individuals, τ is the length of immunity period. B_1, B_2 and B_3 are mutually independent Brownian motions on the suitable probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ with the intensity of environmental white noise σ_1, σ_2 and σ_3. The parameters are all supposed to be positive.

Using the same method as in [3], Yang et al. [12] considered the SIR model with saturated incidence and investigated their dynamics according to the basic reproduction number under additional conditions. The studied model in [12, 13] is as follows:

$$\begin{cases} \mathrm{d}S(t) = \big[\Lambda - \mu S(t) - \frac{\beta S(t)I(t)}{1+\alpha I(t)}\big]\mathrm{d}t + \sigma_1 S(t)\mathrm{d}B_1(t), \\ \mathrm{d}I(t) = \big[\frac{\beta S(t)I(t)}{1+\alpha I(t)} - (\mu+\gamma+\delta)I(t)\big]\mathrm{d}t + \sigma_2 I(t)\mathrm{d}B_2(t), \\ \mathrm{d}R(t) = \big[\gamma I(t) - \mu R(t)\big]\mathrm{d}t + \sigma_3 R(t)\mathrm{d}B_3(t). \end{cases} \tag{1.2}$$

Finding the threshold value is a very meaningful research topic. However, they [11, 12] did not accurately point out the threshold whose value can completely determine the dynamics of the considered models. In this paper, we will try to find the threshold conditions for the following delayed SIR epidemic model with saturation incidence:

$$\begin{cases} \mathrm{d}S(t) = [\Lambda - \mu S(t) - \frac{\beta S(t)I(t)}{1+\alpha I(t)} + \gamma e^{-\mu\tau} I(t-\tau)]\mathrm{d}t + \sigma_1 S(t)\mathrm{d}B_1(t), \\ \mathrm{d}I(t) = [\frac{\beta S(t)I(t)}{1+\alpha I(t)} - (\mu+\gamma+\delta)I(t)]\mathrm{d}t + \sigma_2 I(t)\mathrm{d}B_2(t), \\ \mathrm{d}R(t) = [\gamma I(t) - \gamma e^{-\mu\tau} I(t-\tau) - \mu R(t)]\mathrm{d}t + \sigma_3 R(t)\mathrm{d}B_3(t). \end{cases} \tag{1.3}$$

Since $R(t)$ does not appear explicitly in the first two equations of (1.3), so we omitted the third equation without loss of generality. Then, we only discuss the following system

$$\begin{cases} \mathrm{d}S(t) = \big[\Lambda - \mu S(t) - \dfrac{\beta S(t)I(t)}{1+\alpha I(t)} + \gamma e^{-\mu\tau} I(t-\tau)\big]\mathrm{d}t + \sigma_1 S(t)\mathrm{d}B_1(t), \\ \mathrm{d}I(t) = \big[\dfrac{\beta S(t)I(t)}{1+\alpha I(t)} - (\mu+\gamma+\delta)I(t)\big]\mathrm{d}t + \sigma_2 I(t)\mathrm{d}B_2(t), \end{cases} \tag{1.4}$$

instead of (1.3).

Remark 1 The coefficients of system (1.4) are locally Lipschitz continuous, using the Lyapunov analysis method (see [11]) by defining a C^2-function

$$V(S(t), I(t)) = \big(S(t) - 1 - \log S(t)) + (I(t) - 1 - \log I(t)\big) + \gamma e^{-\mu\tau} \int_{t-\tau}^{t} I(s)\mathrm{d}s$$

for $(S(t), I(t)) \in \mathbb{R}_+^2$, we can prove that the solution of system (1.4) is positive and global. The proof is similar to Liu et al. [11], and hence is omitted.

The outline of the paper is as follows. In Sect. 2, by using simpler method, the threshold of the stochastic delayed SIR epidemic model is obtained, whose value is below 1 or above 1 will completely determine the disease to go extinct or prevail for any size of the white noise. In Sect. 3, numerical simulations are carried out to illustrate the theoretical results.

2 The Main Results

For simplicity, we introduce the following notations and lemmas which will be used later.

$$\langle x(t)\rangle = \frac{1}{t}\int_0^t x(u)\mathrm{d}u.$$

Let

$$R_0^S = \frac{1}{\mu+\delta+\gamma}(\beta\frac{\Lambda}{\mu} - \frac{{\sigma_2}^2}{2}),$$

which is an important parameter.

Lemma 2.1 (see [14]) *Let $A(t)$ and $U(t)$ be two continuous adapted increasing process on $t \geq 0$ with $A(0) = U(0) = 0$ a.s. Let $M(t)$ be a real-valued continuous local martingale with $M(0) = 0$ a.s. Let X_0 be a nonnegative $\mathcal{F}_0$-measurable random variable such that $EX_0 < \infty$. Define $X(t) = X_0 + A(t) - U(t) + M(t)$ for all*

$t \geq 0$. *If* $X(t)$ *is nonnegative, then* $\lim_{t\to\infty} A(t) < \infty$ *implies* $\lim_{t\to\infty} U(t) < \infty$, $\lim_{t\to\infty} X(t) < \infty$ *and* $\lim_{t\to\infty} M(t) < \infty$ *hold with probability one.*

Lemma 2.2 (see [15]) *Let* $M(t)$, $t \geq 0$ *be a local martingale vanishing at time* 0 *and define*

$$\rho_M(t) := \int_0^t \frac{d\langle M, M\rangle(s)}{(1+s)^2}, t \geq 0$$

where $\langle M, M\rangle(t)$ *is Meyers angle bracket process. Then*

$$\lim_{t\to\infty} \frac{M(t)}{t} = 0 \quad a.s. \ \ provided\ that \ \ \lim_{t\to\infty} \rho_M(t) < \infty \quad a.s.$$

Lemma 2.3 *Let* $(S(t), I(t), R(t))$ *be the solution of system (1.3) with initial value* $S(0) > 0$, $I(\zeta) \geq 0$ *for all* $\zeta \in [-\tau, 0)$ *with* $I(0) > 0$ *and* $R(0) > 0$*, then*

$$\limsup_{t\to\infty}[S(t) + I(t) + R(t)] < \infty \quad a.s. \tag{2.1}$$

Moreover,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \sigma_1 S(s)\mathrm{d}B_1(s) = 0, \lim_{t\to\infty} \frac{1}{t}\int_0^t \sigma_2 I(s)\mathrm{d}B_2(s) = 0,$$
$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \sigma_3 R(s)\mathrm{d}B_3(s) = 0 \quad a.s. \tag{2.2}$$

Proof 1 By (1.3), we get

$$\begin{aligned}\mathrm{d}(S+I+R) = &[\Lambda - \mu(S+I+R) - \delta I]\mathrm{d}t + \sigma_1 S\mathrm{d}B_1(t) + \sigma_2 I\mathrm{d}B_2(t)\\ &+ \sigma_3 R\mathrm{d}B_3(t).\end{aligned}$$

This admits a unique solution which can be written as follows:

$$\begin{aligned}S+I+R = &\frac{\Lambda}{\mu} + [S(0)+I(0)+R(0) - \frac{\Lambda}{\mu}]e^{-\mu t} - \delta\int_0^t e^{-\mu(t-s)}I(s)\mathrm{d}s + M(t)\\ \leq &\frac{\Lambda}{\mu} + [S(0)+I(0)+R(0) - \frac{\Lambda}{\mu}]e^{-\mu t} + M(t),\end{aligned}$$

where $M(t) = \sigma_1 \int_0^t e^{-\mu(t-s)}S(s)\mathrm{d}B_1(s) + \sigma_2 \int_0^t e^{-\mu(t-s)}I(s)\mathrm{d}B_2(s) + \sigma_3 \int_0^t e^{-\mu(t-s)}$ $R(s)\mathrm{d}B_3(s)$ is a continuous local martingale with $M(0) = 0$ a.s. Define

$$X(t) = X(0) + A(t) - U(t) + M(t),$$

with $X(0) = S(0) + I(0) + R(0)$, $A(t) = \frac{\Lambda}{\mu}(1 - e^{-\mu t})$ and $U(t) = \big(S(0) + I(0) + R(0)\big)(1 - e^{-\mu t})$ for all $t \geq 0$. According to the stochastic comparison theorem, $S(t) + I(t) + R(t) \leq X(t)$ a.s. It is easy to check that $A(t)$ and $U(t)$ are continuous adapted increasing processes on $t \geq 0$ with $A(0) = U(0) = 0$. By using Lemma 2.1, we obtain that $\lim_{t\to\infty} X(t) < \infty$ a.s. Then, we complete the proof of (2.1).

For simplicity, we denote

$$M_1(t) = \int_0^t \sigma_1 S(s)\mathrm{d}B_1(s),\ M_2(t) = \int_0^t \sigma_2 I(s)\mathrm{d}B_2(s),\ M_3(t) = \int_0^t \sigma_3 I(s)\mathrm{d}B_3(s).$$

Compute that $\langle M_1, M_1\rangle(t) = \int_0^t \sigma_1^2 S^2(s)\mathrm{d}s$ and by (2.1), we get

$$\lim_{t\to\infty} \rho_1(t) = \lim_{t\to\infty} \int_0^t \frac{\sigma_1^2 S^2(s)\mathrm{d}s}{(1+s)^2} \leq \sigma_1^2 \sup_{t\geq 0}[S^2(t)] < \infty.$$

Thus, from lemma 2.2, $\lim_{t\to\infty} \frac{1}{t}\int_0^t \sigma_1 S(s)\mathrm{d}B_1(s) = 0$. Obviously, the left can be proved similarly. Then, the proof is completed. □

Lemma 2.4 (see [16]) *Let* $f \in C\big[(0,\infty)\times\Omega,(0,\infty)\big]$ *and* $F \in C\big([0,\infty]\times\Omega, R\big)$. *If there exist positive constants* λ_0, λ *and* T *such that*

$$logf(t) = \lambda t - \lambda_0 \int_0^t f(s)ds + F(t) \quad a.s.$$

for all $t \geq T$, *and* $\lim_{t\to\infty} \dfrac{F(t)}{t} = 0$ *a.s., then*

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t f(s)ds = \frac{\lambda}{\lambda_0} \quad a.s.$$

Lemma 2.5 *Let* $(S(t), I(t))$ *be the solution of model (1.4) with initial value* $S(0) > 0$ *and* $I(\zeta) \geq 0$ *for all* $\zeta \in [-\tau, 0)$ *with* $I(0) > 0$. *Then*

$$\lim_{t\to\infty} \frac{S(t) + I(t) + \gamma e^{-\mu t}\int_{t-\tau}^t e^{\mu s} I(s)\mathrm{d}s}{t} = 0 \ \ a.s.$$

Moreover,

$$\lim_{t\to\infty} \frac{S(t)}{t} = 0\,,\ \lim_{t\to\infty} \frac{I(t)}{t} = 0\,,\ \lim_{t\to\infty} \frac{e^{-\mu t}\int_{t-\tau}^t e^{\mu s} I(s)\mathrm{d}s}{t} = 0 \ \ a.s. \tag{2.3}$$

Remark 2 Lemma 2.5 can be easily proved by the same way as Lemma 2.1 in [7, 11], where they obtained the results by Burkholder-Davis-Gundy inequality, Bore-Cantelli lemma, Doob's martingale inequality and Strong Law of Large Numbers. So we omit theirs proofs. The details of the proof can be found in [7, 11].

Theorem 2.1 *Let $(S(t), I(t))$ be the solution of model (1.4) with initial value $S(0) > 0$ and $I(\zeta) \geq 0$ for all $\zeta \in [-\tau, 0)$ with $I(0) > 0$.*
(I) *If $R_0^S < 1$, then $\limsup_{t\to\infty} \log \frac{I(t)}{I(0)} \leq (\mu+\delta+\gamma)(R_0^S - 1) < 0$ a.s. namely, $I(t)$ tends to zero exponentially;*
(II) *If $R_0^S > 1$, then $\limsup_{t\to\infty} \langle I(t)\rangle = \frac{\mu}{\alpha\mu+\beta}(R_0^S - 1) > 0$ a.s. namely, the disease will be persistent in mean.*

Proof 2 Notice that

$$\begin{aligned}\mathrm{d}(S + I + \gamma e^{-\mu\tau}\int_{t-\tau}^{t} I(s)\mathrm{d}s) =& [\Lambda - \mu S - (\mu+\delta+\gamma(1-e^{-\mu\tau}))I]\mathrm{d}t\\ &+ \sigma_1 S(t)\mathrm{d}B_1(t) + \sigma_2 I(t)\mathrm{d}B_2(t).\end{aligned}$$

Integrating from 0 to t on both sides of the above, yields

$$\begin{aligned}&\frac{S(t)-S(0)}{t} + \frac{I(t)-I(0)}{t} + \frac{\gamma e^{-\mu\tau}\int_{t-\tau}^{t} I(s)\mathrm{d}s - \gamma e^{-\mu\tau}\int_{-\tau}^{0} I(s)\mathrm{d}s}{t}\\ &= \Lambda - \mu\langle S(t)\rangle - (\mu+\delta+\gamma(1-e^{-\mu\tau}))\langle I(t)\rangle + \frac{1}{t}M_1(t) + \frac{1}{t}M_2(t).\end{aligned}$$

Then, we have

$$\langle S(t)\rangle = \frac{\Lambda}{\mu} - \frac{\mu+\delta+\gamma(1-e^{-\mu\tau})}{\mu}\langle I(t)\rangle + \frac{1}{t}\varphi_1(t), \tag{2.4}$$

where

$$\begin{aligned}\varphi_1(t) = -\frac{1}{\mu}[S(t) - S(0) + I(t) - I(0) + \gamma e^{-\mu\tau}\int_{t-\tau}^{t} I(s)\mathrm{d}s\\ - \gamma e^{-\mu\tau}\int_{-\tau}^{0} I(s)\mathrm{d}s - M_1(t) - M_2(t)].\end{aligned}$$

Applying Lemma 2.3 and Lemma 2.5, we obtain that

$$\lim_{t\to\infty} \frac{\varphi_1(t)}{t} = 0 \quad a.s.$$

By integrating (1.4), we have that

$$\Lambda - (\mu + \frac{\beta}{\alpha})\langle S(t)\rangle + \frac{1}{\alpha}\langle \frac{\beta S(t)}{1+\alpha I(t)}\rangle + \gamma e^{-\mu\tau}\langle I(t)\rangle = \frac{1}{t}\varphi_2(t), \tag{2.5}$$

where $\varphi_2(t) = S(t) - S(0) + \gamma e^{-\mu\tau}\int_{t-\tau}^{t} I(s)\mathrm{d}s - \gamma e^{-\mu\tau}\int_{-\tau}^{0} I(s)\mathrm{d}s - M_1(t)$.

By Lemma 2.3 and Lemma 2.5, we can easily get

$$\lim_{t\to\infty}\frac{\varphi_2(t)}{t}=0 \quad a.s.$$

Define $V(I(t)) = \log I(t)$ and by Itô formula, we have

$$\mathrm{d}\log I(t) = [\frac{\beta S(t)}{1+\alpha I(t)} - (\mu+\delta+\gamma) - \frac{1}{2}{\sigma_2}^2]\mathrm{d}t + \sigma_2 \mathrm{d}B_2(t). \tag{2.6}$$

Integrating from 0 to t and dividing t on both sides of (2.6), we have

$$\frac{1}{t}\log\frac{I(t)}{I(0)} = \langle\frac{\beta S(t)}{1+\alpha I(t)}\rangle - (\mu+\delta+\gamma) - \frac{1}{2}{\sigma_2}^2 + \frac{\sigma_2 B_2(t)}{t}. \tag{2.7}$$

Substituting (2.4) and (2.5) into (2.7), we have

$$\begin{aligned}\frac{1}{t}\log\frac{I(t)}{I(0)} &= \frac{\beta}{\mu}\Lambda - (\mu+\delta+\gamma) - \frac{1}{2}{\sigma_2}^2\\ &\quad - \frac{\alpha\mu+\beta}{\mu}[\mu+\delta+\gamma(1-e^{-\mu\tau})]\langle I(t)\rangle + \frac{F(t)}{t},\end{aligned} \tag{2.8}$$

where $F(t) = -\dfrac{\alpha\mu+\beta}{\mu}\varphi_1(t) + \alpha\varphi_2(t) + \sigma_2 B_2(t)$. By the strong law of large numbers and Lemma 2.3, we can get $\lim_{t\to\infty}\dfrac{F(t)}{t} = 0$.

If $R_0^S < 1$, by the fact that $I(t) > 0$ a.s., we complete the proof of (I). If $R_0^S > 1$, by Lemma 2.4 and (2.8) yields the desired result (II). The proof is completed.

Remark 3 In Theorem 2.1, it is easy to see that the disease dynamics of SDE model with delay can be governed by its related parameter R_0^S: if $R_0^S < 1$, the disease will die out stochastically; if $R_0^S > 1$, the disease will break out. Then, the parameter R_0^S completely determines the disease to go extinct or prevail for any size of the white noise. It shows R_0^S is the threshold of (1.3). Comparing with [11, 12], the related results are improved, and the method is simplified due to the nonnegative semimartingale convergence theorem.

3 Numerical Simulations and Conclusion

In this section, computer simulation of the path of $(S(t), I(t))$ for model (1.4) is given by using the EM method with initial value $(S(0), I(0)) = (0.6, 0.2)$. At the same time, we shall assume that all parameters are given in appropriate units (see [13, 17, 18]).

Example 1 As numerical example, let us suppose that

$$\Lambda = 0.6, \mu = 0.2, \alpha = 0.2, \beta = 0.3, \gamma = 0.3, \delta = 0.1, \sigma_1 = 0.3, \sigma_2 = 0.8, \sigma_3 = 0.5.$$

We can therefore compute that $R_0^S = 0.967 < 1$. By (I) of Theorem 2.1, the disease will die out. The computer simulation in Fig. 1 supports this result clearly, illustrating extinction of the disease.

Example 2 Assume that the system parameters and the initial conditions are the same as in Example 1, except $\sigma_2 = 0.2$, such that the conditions of Theorem 2.1 (II) hold. Note that $R_0^S = 1.476 > 1$. That is to say, the disease will prevail. The computer simulation in Fig. 2 supports this result clearly, illustrating persistence of the disease.

Movement of infective individuals can be seen for different values of delay τ in Fig. 3. As the value of delay τ increasing, the amplitude of these oscillations decrease correspondingly. For large values of time delays τ, complex dynamics of the model is observed as oscillations occur at the early stages and the oscillations are damped. Effects of recovered individuals can be seen for different values of delay τ in Fig. 4, in the beginning, as the value of delay τ increasing, the amplitude of these oscillations increase correspondingly. For sufficiently large τ, initial oscillations are quickly damped.

Investigating epidemic models, we pay more attention to the extinction and persistence of the disease. In the deterministic models, the value of the basic reproductive number R_0 determines the prevalence or extinction of the disease. If $R_0 < 1$, the disease will be eliminated from the community, whereas an epidemic occurs if $R_0 > 1$.

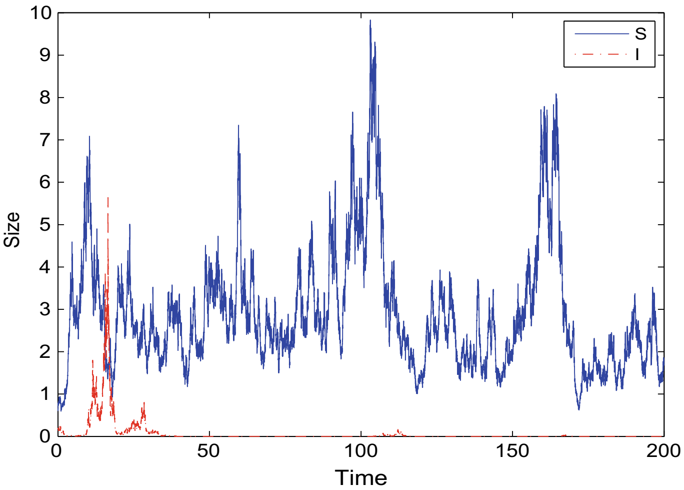


Fig. 1 Simulations of the path $S(t)$, $I(t)$ for the stochastic delayed (1.4) with noise parameter value $\sigma_2 = 0.8, \tau = 0.1$

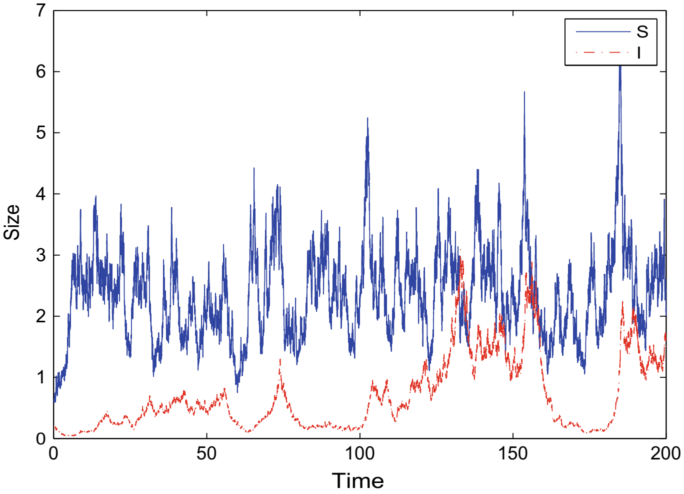


Fig. 2 Simulations of the path $S(t)$, $I(t)$ for the stochastic delayed (1.4) with noise parameter value $\sigma_2 = 0.2, \tau = 0.1$

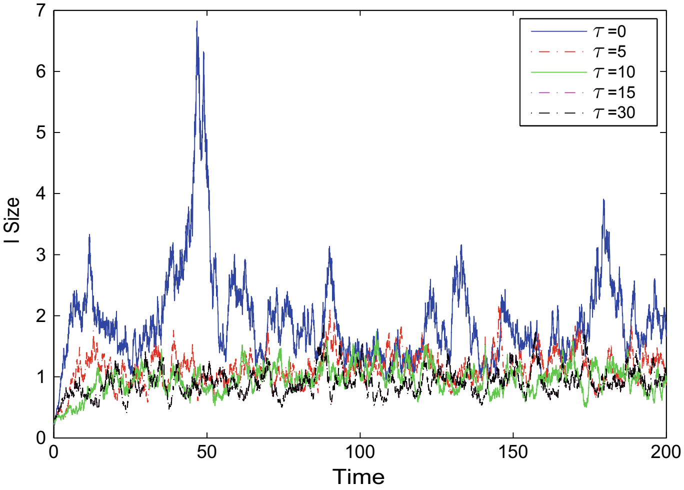


Fig. 3 Variation of infective population with time for different values of delay τ for $\sigma_2 = 0.2$, $\beta = 3$

For stochastic models, there may be not endemic equilibrium and there different types of persistence and extinction. In this paper, we provide a simple but effective method for estimating the threshold of a class of the stochastic epidemic models by use of the nonnegative semimartingale convergence theorem. The threshold R_0^S is obtained for the stochastic delayed SIR model, whose value will completely determine persistence or extinction of the disease. If $R_0^S < 1$, the disease is eliminated,

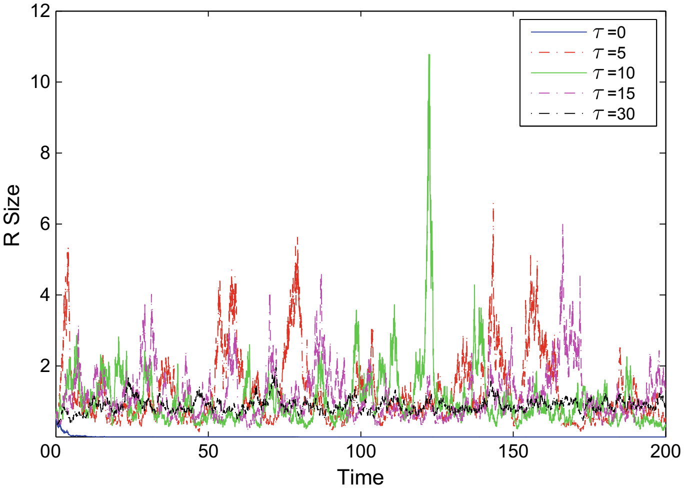


Fig. 4 Variation of recovered population with time for different values of delay τ for $\sigma_2 = 0.2$, $\beta = 3$

whereas if $R_0^S > 1$ the disease persists in the population. Besides, when $R_0^S > 1$, the system is proved to be convergent in time mean. Our results recover the known results in the earlier literature special cases. On the other hand, comparing with the previously-known literatures, the method is simpler than before and the results are improved.

References

1. Ji, C., Jiang, D., Shi, N.: The behavior of an SIR epidemic model with stochastic perturbation. Stoch. Anal. Appl. **30**, 755–773 (2012)
2. Gray, A., Greenhalgh, D., Hu, L., Mao, X., Pan, J.F.: A stochastic differential equation SIS epidemic model. SIAM J. Appl. Math. **71**, 876–902 (2011)
3. Jiang, D., Yu, J., Ji, C., Shi, N.: Asymptotic behavior of global positive solution to a stochastic SIR model. Math. Comput. Model. **54**, 221–232 (2011)
4. Artalejo, J., Economou, A., Lopez-Herrero, M.: On the number of recovered individuals in the SIS and SIR stochastic epidemic models. Math. Biosci. **228**, 45–55 (2010)
5. Lu, Q.: Stability of SIRS system with random perturbations. Phys. A **388**, 3677–3686 (2009)
6. Zhao, Y., Jiang, D., O'Regan, D.: The extinction and persistence of the stochastic SIS epidemic model with vaccination. Phys. A **392**, 4916–4927 (2009)
7. Zhao, Y., Jiang, D.: The threshold of a stochastic SIS epidemic model with vaccination. Appl. Math. Comput. **243**, 718–727 (2014)
8. Ji, C., Jiang, D.: Threshold behaviour of a stochastic SIR model. Appl. Math. Model. **38**, 5067–5079 (2014)
9. Lin, Y., Jiang, D., Wang, S.: Stationary distribution of a stochastic SIS epidemic model with vaccination. Phys. A **394**, 187–197 (2014)
10. Lahrouz, A., Settati, A.: Necessary and sufficient condition for extinction and persistence of SIRS system with random perturbation. Appl. Math. Comput. **233**, 10–19 (2014)

11. Lu, Q., Chen, Q., Jiang, D.: The threshold of a stochastic delayed SIR epidemic model with temporary immunity. Phys. A **450**, 115–125 (2016)
12. Yang, Q., Jiang, D., Shi, N.: The ergodicity and extinction of stochastically perturbed SIR and SEIR epidemic models with saturated incidence. J. Math. Anal. Appl. **388**, 248–271 (2012)
13. Zhao, D.: Study on the threshold of a stochastic SIR epidemic model and its extensions. Commun. Nolinerar Sci. Numer. Simulat. **38**, 172–177 (2016)
14. Mao, X.: Stochastic Differential Equations and Applications, 2nd edn. Horwood, Chichester, UK (2008)
15. Lipster, R.: A strong law of large numbers for local martingale. Stochastics **3**, 217–228 (1980)
16. Zhao, Y., Jiang, D.: The threshold of a stochastic SIRS epidemic model with saturated incidence. Appl. Math. Lett. **34**, 90–93 (2014)
17. Kyrychko, Y., Blyuss, K.: Global properties of a delayed SIR model with temporary immunity and nonlinear incidence rate. Nonlinear Anal. Real World Appl. **6**, 495–507 (2005)
18. Naresh, R., Tripathi, A., Tchuenche, J., Sharma, D.: Stability analysis of a time delayed SIR epidemic model with nonlinear incidence rate. Comput. Math. Appl. **58**, 348–359 (2009)

Semi-stationary Equilibrium Strategies in Non-cooperative N-person Semi-Markov Games

Prasenjit Mondal and Sagnik Sinha

Abstract For a limiting ratio average (undiscounted) non-cooperative N-person semi-Markov game with finite state and action spaces, we prove that the solutions in the game where all players are restricted to semi-stationary strategies (that depend only on the initial state and the current state) are solutions for the unrestricted game. Furthermore, we consider zero-sum two-person semi-Markov games with action independent transitions (where the transition probabilities are independent of the actions of the players in each state) and prove the existence of an optimal semi-stationary strategy for each player. An example is provided to show that the semi-stationary optimal strategies cannot be strengthened further for such class of games.

Keywords Semi-Markov games · Limiting ratio average payoff · Nash equilibrium · Action independent transitions · Optimal semi-stationary strategies

Mathemetics Subject Classification (2000) Primary: 91A15 · Secondary: 60G99

1 Introduction

A semi-Markov game (SMG) is a generalization of a stochastic game [17] using variable sojourn times which depend not only on the present state and the actions chosen but also on the state at the next decision epoch. In the literature of SMGs with limiting ratio average (undiscounted) payoff, to prove the existence of optimal/Nash equilibrium strategies, various recurrence like conditions are assumed at the outset. For example, [5, 8, 16] gave an ergodicity condition whereas, [6, 20]

P. Mondal (✉)
Mathematics Department, Government General Degree College, Ranibandh,
Bankura 722135, India
e-mail: prasenjit1044@yahoo.com

S. Sinha
Mathematics Department, Jadavpur University, Kolkata 700032, India
e-mail: sagnik62@yahoo.co.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_28

used some variants of a Lyapunov-like condition which yields the so-called weighted geometric ergodicity property. Limiting ratio average SMGs have also been studied using various special structures on the data (i.e. rewards and transitions)—a subclass of SeR-SIT (Separable Reward and State Independent Transition) SMGs [10], single controller unichain SMGs [11, 13].

In this paper, we study non-cooperative N-person finite state and action spaces SMGs with a general multichain structure. A form of limiting ratio average (undiscounted) payoff is the criterion for comparing different strategies. Jianyong and Xiaobo [7] showed in an example that an undiscounted semi-Markov decision process (SMDP, i.e. a one player SMG) may not have an optimal stationary policy. Mondal and Sinha [11] showed the existence of an optimal semi-stationary policy (that depends only on initial and current state) in that example. Sinha and Mondal [18] proved the existence of semi-stationary optimal policies for a finite state and action spaces SMDP. Mondal [15] studied the zero-sum two-person undiscounted SMGs with an absorbing set of states and using an existence result of SMDP [14], the author proved that a pair of strategies optimal in the class of semi-stationary strategies will also be optimal in the class of all strategies. The author also proved that such an absorbing SMG has an optimal pair of semi-stationary strategies when the transition probabilities are independent of the actions of the players (such games are called action independent transition SMGs). However, the results of [15] are applicable only for the class of zero-sum two-person SMGs where all states but one are absorbing for any pair of stationary strategies. Some natural questions arise as what happen if we consider the case of a general multichain non-cooperative nonzero-sum N-person SMG and the case of a zero-sum two-person multichain SMG with action independent transitions. We solve such problems in this paper.

This paper is organized as follows. In Sect. 2, we define a non-cooperative N-person undiscounted SMG with finite state and action spaces. We prove that if a Nash equilibrium exists in the class of semi-stationary strategies, then it will be an equilibrium point in the class of all strategies. In Sect. 3, a class of zero-sum two-person action independent transition SMGs has been studied. The existence of an optimal pair of semi-stationary strategies is proved for such games. An example is provided to show that this result cannot be strengthened further to obtain an optimal pair of stationary strategies.

2 Non-cooperative N-person Finite Semi-Markov Games

A non-cooperative N-person finite (state and action spaces) semi-Markov game (SMG) is defined by a collection of objects $< S, A_i = \{A_i(s) : s \in S\}, p, Q, r_i; i = 1, 2, \ldots, N >$. Here $S = \{1, 2, \ldots, z\}$ is the state space and for each $s \in S$, $i = 1, 2, \ldots, N$, $A_i(s)$ is the set of admissible actions of player i in state s. We assume that S, $A_i(s)$, $s \in S$, $i = 1, 2, \ldots, N$ are finite (nonempty) sets. Let $\mathcal{K} = \{(s, a_1, a_2, \ldots, a_N) \mid s \in S, a_i \in A_i(s), i = 1, 2, \ldots, N\} \subseteq S \times A_1 \times A_2 \times \cdots \times A_N$. For each $k \in \mathcal{K}$, $p(\cdot \mid k)$ is the stochastic kernel (probability distribution) on

S and it governs the transition from one state to another. For each $k = (s, a_1, a_2, \ldots, a_N)$, $s' \in S$, $Q_{ss'}(\cdot \mid a_1, a_2, \ldots, a_N)$ is a probability distribution function on $[0, \infty)$ given $\mathcal{K} \times S$, and is called the conditional transition time distribution. Finally, $r_i : \mathcal{K} \to \mathbf{R}, i = 1, 2, \ldots, N$ is called the payoff function and it represents the immediate (expected) reward for player i.

The game is played over the infinite future as follows: At the first decision epoch, the game starts at a state $s_1 \in S$ and the players choose their actions $a_i^1 \in A_i(s_1), i = 1, 2, \ldots, N$, simultaneously and hence independently of one another. Consequently, player i, $i = 1, 2, \ldots, N$ receives an immediate reward $r_i(s_1, a_1^1, a_2^1, \ldots, a_N^1)$; the system moves to a new state s_2 with probability $p(s_2|s_1, a_1^1, a_2^1, \ldots, a_N^1)$ and the time for this transition is determined by $Q_{s_1 s_2}(\cdot \mid a_1^1, a_2^1, \ldots, a_N^1)$. The process repeats itself from state s_2 and the game continues on indefinitely.

Non-cooperative nonzero-sum N-person SMGs were studied in [16] and [5]. Clearly, when the transition times are identical, the game reduces to a discrete-time stochastic game [17]. A semi-Markov decision process (SMDP) is defined as an SMG where all the players except one are restricted to one action in each state. Thus an SMDP is a one person SMG. For $n \in \mathbf{N}$ (the set of all natural numbers), let $\mathcal{H}_n$ be the space of all histories upto the nth decision epoch, i.e., $\mathcal{H}_n = \mathcal{H}_{n-1} \times A_1 \times \ldots \times A_N \times S$ with $\mathcal{H}_1 = S$. A generic element $h_n \in \mathcal{H}_n$ is defined as

$$h_n = (s_1, a_1^1, a_2^1, \ldots, a_N^1, s_2, a_1^2, a_2^2, \ldots, a_N^2, \ldots, s_n),$$

where $(s_l, a_1^l, a_2^l, \ldots, a_N^l) \in \mathcal{K}$ are respectively the state and actions of the players at the lth decision epoch.

A behavioral strategy π_i of player i is a sequence $\{\pi_{in}\}_{n=1}^{\infty}$ of stochastic kernel on A_i given $\mathcal{H}_n$ satisfying the constraint

$$\pi_{in}(A_i(s_n)|h_n) = 1, s_n \in S, h_n \in \mathcal{H}_n, n \in \mathbf{N}.$$

A strategy $\{\pi_{in}\}_{n=1}^{\infty}$ of player i is called semi-Markov if

$$\pi_{in}(\cdot|h_n) = \pi_{in}(\cdot|s_1, s_n) \text{ for all } h_n \in \mathcal{H}_n, n \in \mathbf{N},$$

i.e., if each π_{in} depends on the history h_n only through the initial state s_1, current state s_n and the decision epoch number n.

A strategy $\{\pi_{in}\}_{n=1}^{\infty}$ of player i is called semi-stationary if there exists a function g_i such that

$$\pi_{in}(\cdot|h_n, s_1 = s, s_n = s') = g_i(s, s') \text{ for all } s, s' \in S, h_n \in \mathcal{H}_n, n \in \mathbf{N}.$$

A semi-stationary strategy depends only on the initial state and the current state of the game. Thus, a semi-stationary strategy is a semi-Markov strategy which is independent of time count n.

A strategy $\{\pi_{in}\}_{n=1}^{\infty}$ of player i is called stationary if there exists a function f_i such that

$$\pi_{in}(\cdot|h_n, s_1 = s) = f_i(s) \text{ for all } s \in S, h_n \in \mathcal{H}_n, n \in \mathbf{N}.$$

A strategy is called pure if it is not randomized.

For simplifying the notation, we shall write g_i and f_i respectively as the semi-stationary and stationary strategy for player i $(i = 1, 2, \dots, N)$. Let Π_i, $\mathcal{G}_i$ and $\mathcal{F}_i$ be respectively the classes of all behavioral, semi-stationary and stationary strategies of player i $(i = 1, 2, \dots, N)$ in a finite N-person semi-Markov game. Similar classes of strategies (policies) in an SMDP will be denoted as Π, $\mathcal{G}$ and $\mathcal{F}$ respectively.

Let $(X_1, A_1^1, A_2^1, \dots, A_N^1, X_2, A_1^2, A_2^2, \dots, A_N^2, \dots)$ be a coordinate sequence in $\Omega = \mathcal{K}^\infty \times S$. Given an initial state $s \in S$ and an N-tuple of strategies $(\pi_1, \pi_2, \dots, \pi_N)$, Kolmogorov's theorem ([1]) guarantees the existence and uniqueness of a probability measure $P_{\pi_1,\dots,\pi_N}(\cdot \mid X_1 = s)$ on the product σ-field of Ω. Let $E_{\pi_1,\dots,\pi_N}(\cdot \mid X_1 = s)$ be the corresponding expectation operator.

Definition 1 For $(\pi_1, \dots, \pi_N) \in \Pi_1 \times \cdots \times \Pi_N$, the expected limiting ratio average payoff to player i $(i = 1, 2, \dots, N)$ in state $s \in S$ is defined as

$$\phi_i(s, \pi_1, \dots, \pi_N) = \liminf_{n\to\infty} \frac{E_{\pi_1,\dots,\pi_N}[\sum\limits_{m=1}^{n} r_i(X_m, A_1^m, \dots, A_N^m) \mid X_1 = s]}{E_{\pi_1,\dots,\pi_N}[\sum\limits_{m=1}^{n} \tau(X_m, A_1^m, \dots, A_N^m) \mid X_1 = s]},$$

where

$$\tau(s, a_1, \dots, a_N) = \sum_{s'\in S} p(s'|s, a_1, \dots, a_N) \int_0^\infty t\, dQ_{ss'}(t|a_1, \dots, a_N)$$

is the expected (mean) sojourn time in state s when the players choose their actions $a_i \in A_i(s)$, $i = 1, 2, \dots, N$.

Remark 1 For an initial state $s \in S$ and an N-tuple of strategies $(\pi_1, \pi_2, \dots, \pi_N)$, the payoff function $\phi_i(s, \pi_1, \dots, \pi_N)$ is a function of the probability measure $P_{\pi_1,\dots,\pi_N}(\cdot \mid X_1 = s)$. Note that the strategies considered here do not use the information about the previous and current sojourn times. [4, Theorem 4.3, p. 266] showed that the probability measure $P_{\pi_1,\dots,\pi_N}(\cdot \mid X_1 = s)$ (and hence the expectation operator $E_{\pi_1,\dots,\pi_N}(\cdot \mid X_1 = s)$) is independent of the sojourn times of the game.

Assumption 1 There exist $\epsilon > 0$, $M > 0$ such that

$$\epsilon \le \tau(s, a_1, \dots, a_N) \le M \;\forall (s, a_1, \dots, a_N) \in \mathcal{K}.$$

For simplicity of notation, we shall write

$$\phi_i(s,\pi_1,\ldots,\pi_N) = \liminf_{n\to\infty} \frac{\sum\limits_{m=1}^{n} r_i^m(s,\pi_1,\pi_2,\ldots,\pi_N)}{\sum\limits_{m=1}^{n} \tau^m(s,\pi_1,\pi_2,\ldots,\pi_N)},$$

where $r_i^m(s,\pi_1,\pi_2,\ldots,\pi_N)$ and $\tau^m(s,\pi_1,\pi_2,\ldots,\pi_N)$ are respectively the expected reward to player i, $i=1,2,\ldots,N$ and the expected sojourn time at the m-th decision epoch when the players use the strategies $\pi_1,\pi_2,\ldots,\pi_N$ respectively and when the initial state is s.

Definition 2 *(Nash equilibrium)* An N-tuple of strategies $(\pi_1^*,\ldots,\pi_N^*) \in \Pi_1 \times \cdots \times \Pi_N$ is said to be a Nash equilibrium if

$$\phi_i(s,\pi_1^*,\ldots,\pi_N^*) \geq \phi_i(s,\pi_1^*,\ldots,\pi_{i-1}^*,\pi_i,\pi_{i+1}^*,\ldots,\pi_N^*) \text{ for all } s\in S, \\ \text{for any } \pi_i \in \Pi_i, i=1,2,\ldots,N.$$

For a one player SMG (i.e., SMDP), we have the following result.

Theorem 1 ([18], Theorems 1 and 2, p. 866) *Let $< S, A, p, Q, r >$ be a semi-Markov decision process and ϕ be the limiting ratio average payoff function. Then there exist pure semi-stationary strategies $g^*, g^{**} \in \mathcal{G}$ such that*

$$\phi(s,g^*) \geq \phi(s,\pi) \textit{ and } \phi(s,g^{**}) \leq \phi(s,\pi) \textit{ for all } \pi \in \Pi \textit{ and all } s \in S.$$

We observe the following result for a non-cooperative N-person SMG:

Theorem 2 *Let $G =< S, A_i = \{A_i(s) : s \in S\}, p, \tau, r_i; i = 1,2,\ldots,N >$ be a non-cooperative N-person semi-Markov game and let $s \in S$ be a fixed and arbitrary initial state. If there exists an N-tuple of stationary strategies $(f_1^*, f_2^*,\ldots,f_N^*)$ such that*

$$\phi_i(s,f_1^*,f_2^*,\ldots,f_N^*) \geq \phi_i(s,f_1^*,\ldots,f_{i-1}^*,f_i,f_{i+1}^*,\ldots,f_N^*) \\ \textit{for all } f_i \in \mathcal{F}_i, i=1,2,\ldots,N$$

then $(f_1^*, f_2^*,\ldots,f_N^*)$ is a Nash equilibrium of the game for the initial state s.

Proof We follow the approach of [19] in our proof. Define a semi-Markov decision model $G_0 =< S_0, A_0, p_0, r_0, \tau_0 >$ with player 1 as the decision maker as follows: $S_0 = S$, $A_0(s) = A_1(s)$ for each $s \in S$ and

$$p_0(s'|s,a) = \sum_{a_2\in A_2(s)} \cdots \sum_{a_N\in A_N(s)} p(s'|s,a,a_2,\ldots,a_N) \prod_{i=2}^{N} f_i^*(s,a_i), \\ = p(s'|s,a,f_2^*,\ldots,f_N^*)$$

$$r_0(s,a) = \sum_{a_2\in A_2(s)} \cdots \sum_{a_N\in A_N(s)} r_1(s,a,a_2,\ldots,a_N) \prod_{i=2}^{N} f_i^*(s,a_i),$$

$$
\begin{aligned}
&= r_1(s, a, f_2^*, \dots, f_N^*) \\
\tau_0(s, a) &= \sum_{a_2 \in A_2(s)} \cdots \sum_{a_N \in A_N(s)} \tau(s, a, a_2, \dots, a_N) \prod_{i=2}^{N} f_i^*(s, a_i) \\
&= \tau(s, a, f_2^*, \dots, f_N^*)
\end{aligned}
$$

for all $s, s' \in S, a \in A_1(s)$. Let $\pi_1 = \{\pi_{1n}\}_{n=1}^{\infty} \in \Pi_1$ be any strategy for player 1 in the SMG. We apply this strategy in our decision model by creating pseudo-histories $h_n = (s_1, a_1^1, \dots, a_N^1, s_2, \dots, s_{n-1}, a_1^{n-1}, \dots, a_N^{n-1}, s_n)$ for the n-th decision epoch ($n \in \mathbf{N}$). Denote by $s_1 = s \in S$ the initial state of the model and let $h_1 = (s_1)$ for the first decision epoch. Then π_{11} is well defined and we can use π_{11} to choose the initial action $a_1^1 \in A_1(s_1)$. Let s_2 be the new state for the second decision epoch. Then choose $a_i^1 \in A_i(s_1)$, $i = 2, 3, \dots, N$ according to the joint probability distribution $P(a_2^1 \in A_2^1, \dots, a_N^1 \in A_N^1 | h_1) = \frac{\prod_{i=2}^{N} f_i^*(s_1, a_i^1) p(s_2 | s_1, a_1^1, \dots, a_N^1)}{p_0(s_2 | s_1, a_1^1)}$ and define $h_2 = (s_1, a_1^1, a_2^1, \dots, a_N^1, s_2)$. Now we assume that h_{n-1} has been created. Then choose $a_1^{n-1} \in A_1(s_{n-1})$ using $\pi_{1(n-1)}$ conditioning on the pseudo-history h_{n-1}. Let s_n be the new state at the n-th decision epoch. Then choose $a_i^{n-1} \in A_i(s_{n-1})$, $i = 2, 3, \dots, N$ according to the joint probability distribution

$$
\begin{aligned}
&P(a_2^{n-1} \in A_2^{n-1}, a_3^{n-1} \in A_3^{n-1}, \dots, a_N^{n-1} \in A_N^{n-1} | h_{n-1}) \\
&= \frac{\prod_{i=2}^{N} f_i^*(s_{n-1}, a_i^{n-1}) p(s_n | s_{n-1}, a_1^{n-1}, \dots, a_N^{n-1})}{p_0(s_n | s_{n-1}, a_1^{n-1})}.
\end{aligned} \tag{1}
$$

By this construction of pseudo-histories on each decision epoch, we can apply the strategy π_1 to the decision model. Now we show that

$$
\phi_0(s_1, \pi_1) = \phi_1(s_1, \pi_1, f_2^*, \dots, f_N^*), \tag{2}
$$

where ϕ_0 is the undiscounted payoff function of the decision model G_0.
For this, we first show by induction on the decision epoch number n that

$$
P_{\pi_1}(h_n) = P_{\pi_1, f_2^*, \dots, f_N^*}(h_n) \text{ for all } n \in \mathbf{N}, \tag{3}
$$

where $P_{\pi_1}(h_n)$ and $P_{\pi_1, f_2^*, \dots, f_N^*}(h_n)$ are respectively the probabilities of the pseudo-history h_n in the semi-Markov decision model and the history h_n in the original semi-Markov game when the initial state is $s_1 = s$.
The equality (3) is obvious for $n = 1$. By induction hypothesis, suppose that (3) holds for decision epoch number less than n. Then using (1), we have

$$
\begin{aligned}
P_{\pi_1}(h_n) &= P_{\pi_1}(h_{n-1}) P_{\pi_1}(a_1^{n-1} | h_{n-1}) P(a_2^{n-1}, \dots, a_N^{n-1} | h_{n-1}) p_0(s_n | s_{n-1}, a_1^{n-1}) \\
&= P_{\pi_1, f_2^*, \dots, f_N^*}(h_{n-1}) P_{\pi_1}(a_1^{n-1} | h_{n-1}) p_0(s_n | s_{n-1}, a_1^{n-1})
\end{aligned}
$$

$$\left[\frac{\prod_{i=2}^{N} f_i^*(s_{n-1}, a_i^{n-1}) p(s_n|s_{n-1}, a_1^{n-1}, \dots, a_N^{n-1})}{p_0(s_n|s_{n-1}, a_1^{n-1})}\right]$$
$$= P_{\pi_1, f_2^*, \dots, f_N^*}(h_n).$$

Hence (3) holds and it follows that

$$r_0^n(s_1, \pi_1) = r_1^n(s_1, \pi_1, f_2^*, \dots, f_N^*) \text{ and } \tau_0^n(s_1, \pi_1) = \tau^n(s_1, \pi_1, f_2^*, \dots, f_N^*) \quad \text{for all } n \in \mathbf{N},$$

which proves (2). Next we show that f_1^* is optimal for our decision model G_0 of maximization type with initial state s. The first part of Theorem 1 implies that there exists a stationary optimal strategy $f_0^s \in \mathcal{F}$ for our semi-Markov decision model G_0 with initial state s. Then we have

$$\phi_0(s, f_0^s) = \phi_1(s, f_0^s, f_2^*, \dots, f_N^*) \leq \phi_1(s, f_1^*, f_2^*, \dots, f_N^*) = \phi_0(s, f_1^*).$$

Thus, f_1^* is also optimal for G_0 with initial state s. Let $\pi_1 \in \Pi_1$ be any behavioral strategy for player 1 in the SMG. Then

$$\phi_1(s, \pi_1, f_2^*, \dots, f_N^*) = \phi_0(s, \pi_1) \leq \phi_0(s, f_1^*) = \phi_1(s, f_1^*, f_2^*, \dots, f_N^*).$$

Similarly, we can show that for any behavioral strategy $\pi_i \in \Pi_i$, $i = 2, 3, \dots, N$

$$\phi_i(s, f_1^*, f_2^*, \dots, f_N^*) \geq \phi_i(s, f_1^*, \dots, f_{i-1}^*, \pi_i, f_{i+1}^*, \dots, f_N^*).$$

Hence the theorem is proved. □

Corollary 1 *Suppose there exists an N-tuple of semi-stationary strategies $(g_1^*, g_2^*, \dots, g_N^*)$ such that for all $s \in S$,*

$$\phi_i(s, g_1^*, g_2^*, \dots, g_N^*) \geq \phi_i(s, g_1^*, \dots, g_{i-1}^*, g_i, g_{i+1}^*, \dots, g_N^*) \quad \textit{for all } g_i \in \mathcal{G}_i, i = 1, 2, \dots, N$$

then $(g_1^*, g_2^*, \dots, g_N^*)$ is a Nash equilibrium of the game.

Note: Corollary 1 states that if $(g_1^*, g_2^*, \dots, g_N^*)$ is Nash equilibrium in the class of semi-stationary strategies, then they will be so among all strategies.

Proof Given a semi-stationary strategy g, let g^s be the stationary strategy determined by g when the initial state is $s \in S$. Let $s \in S$ be fixed and arbitrary. By the given condition, there exists an N-tuple of stationary strategies $(g_1^{*s}, g_2^{*s}, \dots, g_N^{*s})$ such that

$$\phi_i(s, g_1^{*s}, g_2^{*s}, \dots, g_N^{*s}) \geq \phi_i(s, g_1^{*s}, \dots, g_{i-1}^{*s}, g_i^s, g_{i+1}^{*s}, \dots, g_N^{*s}) \quad \text{for all } g_i^s \in \mathcal{F}_i, i = 1, 2, \dots, N.$$

By Theorem 2, $(g_1^{*s}, g_2^{*s}, \ldots, g_N^{*s})$ is a Nash equilibrium of the game when the initial state is s. Since the above inequality holds for any $s \in S$, the N-tuple of semi-stationary strategies $(g_1^*, g_2^*, \ldots, g_N^*)$ is a Nash equilibrium of the N-person game. □

3 Action Independent Transition Semi-Markov Games

We begin with the following:

Definition 3 An action independent transition (AIT) semi-Markov game is a semi-Markov game where the transition probabilities are independent of the actions of the players in each state. This means

$$p(s'|s, a_1, a_2, \ldots, a_N) = p(s'|s) \text{ for all } (s, a_1, a_2, \ldots, a_N) \in \mathcal{K}, s' \in S.$$

We obtain some results on zero-sum two-person AIT semi-Markov games. An N-person SMG is called a zero-sum two-person SMG if $N = 2$ and $r_1(s, a_1, a_2) = -r_2(s, a_1, a_2)$ for all $(s, a_1, a_2) \in \mathcal{K}$. In this case, we simply write $r_1(s, a_1, a_2) = r(s, a_1, a_2)$ and the payoff function $\phi_1(s, \pi_1, \pi_2) = \phi(s, \pi_1, \pi_2)$ for all $s \in S$ and $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$. For the zero-sum case, we are concerned with value and optimal strategies.

Definition 4 A zero-sum two-person undiscounted SMG is said to have a value vector $\phi = \left[\phi(s)\right]_{z\times 1}$ if

$$\sup_{\pi_1\in\Pi_1} \inf_{\pi_2\in\Pi_2} \phi(s, \pi_1, \pi_2) = \inf_{\pi_2\in\Pi_2} \sup_{\pi_1\in\Pi_1} \phi(s, \pi_1, \pi_2) = \phi(s) \text{ for all } s \in S.$$

A pair of strategies (π_1^*, π_2^*) is average optimal for the two players if

$$\phi(s, \pi_1, \pi_2^*) \le \phi(s) \le \phi(s, \pi_1^*, \pi_2) \text{ for all } s \in S \text{ and all } (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2.$$

For a zero-sum two-person undiscounted SMG, Mondal [13, Theorem 3.2, p .8] proved that the existence of an optimal in the class of semi-Markov strategies will be optimal in the class of all strategies. The following result is more sharper than that in the sense that we need to consider only the class of semi-stationary strategies.

Theorem 3 *Let $G =< S, A_1, A_2, p, r, \tau >$ be a zero-sum two-person semi-Markov game with limiting ratio average payoff. If there exists a pair of semi-stationary strategies (g_1^*, g_2^*) such that for all $s \in S$,*

$$\phi(s, g_1, g_2^*) \le \phi(s, g_1^*, g_2^*) \le \phi(s, g_1^*, g_2) \textit{ for all } (g_1, g_2) \in \mathcal{G}_1 \times \mathcal{G}_2,$$

then (g_1^*, g_2^*) is a pair of optimal strategies of the game.

Proof The proof that $\phi(s, g_1^*, g_2^*) \geq \phi(s, \pi_1, g_2^*)$ for all $s \in S$, $\pi_1 \in \Pi_1$ follows exactly in similar lines as in Theorem 2 and Corollary 1. The other requirement $\phi(s, g_1^*, g_2^*) \leq \phi(s, g_1^*, \pi_2)$ for all $s \in S$, $\pi_2 \in \Pi_2$ can be proved in analogous way with the help of second part of Theorem 1. □

Remark 2 Theorem 3 is the zero-sum version of Corollary 1 for the case $N = 2$. The converse is not true, i.e., the existence of value of a limiting ratio average semi-Markov game does not necessarily imply the existence of value in semi-stationary strategies. See Big Match [2].

We associate with each $(f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2, r(f_1, f_2) = \Big[r(s, f_1, f_2)\Big]_{z\times 1}, \tau(f_1, f_2)$ $=\Big[\tau(s, f_1, f_2)\Big]_{z\times 1}$ and $P(f_1, f_2) = \Big[p(s'|s, f_1, f_2)\Big]_{z\times z}$ as the expected reward vector, mean sojourn time vector and the transition probability (Markov) matrix respectively. By [3, theorem 2.1, p. 175], there exists a Markov matrix $P^*(f_1, f_2)$ (called the Cesaro limiting matrix) such that

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} P^m(f_1, f_2) = P^*(f_1, f_2).$$

Now we have the following expression for the undiscounted payoff.

Lemma 1 *For each pair of strategies* $(f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$\phi(s, f_1, f_2) = \frac{\big[P^*(f_1, f_2)\, r(f_1, f_2)\big](s)}{\big[P^*(f_1, f_2)\, \tau(f_1, f_2)\big](s)} \textit{ for all } s \in S.$$

Action independent transition SMGs with undiscounted payoffs have been studied in [15] but only for the class of SMGs with absorbing states. Mondal and Sinha [12] studied zero-sum two-person AR-AIT-ATT (Additive Reward-Action Independent Transition and Additive Transition Time) SMGs with discounted payoff. We study AIT SMGs with a general multichain structure and obtain the following:

Theorem 4 *A zero-sum two-person action independent transition semi-Markov game with limiting ratio average payoff has value and an optimal pair of semi-stationary strategies.*

Proof Since the transition probabilities are independent of the actions of the players in each state, we have

$$P(f_1, f_2) = P = \Big[p(s'|s)\Big]_{z\times z} \text{ for each } (f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2.$$

Thus, $P^*(f_1, f_2) = P^* = \Big[p^*(s'|s)\Big]_{z\times z}$. Therefore, by Lemma 1, for each $(f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$\phi(s, f_1, f_2) = \frac{\sum\limits_{s' \in S} p^*(s'|s) r(s', f_1, f_2)}{\sum\limits_{s' \in S} p^*(s'|s) \tau(s', f_1, f_2)} \text{ for all } s \in S.$$

The above payoff function can be expressed in the form $\frac{x^t U y}{x^t V y}$, where $x = (x_1, x_2, \ldots, x_z)$ with $x_s = (f_1(s, a_1) : a_1 \in A_1(s)) \in \mathbf{R}^{|A_1(s)|}$,[1] $y = (y_1, y_2, \ldots, y_z)$ with $y_s = (f_2(s, a_2) : a_2 \in A_2(s)) \in \mathbf{R}^{|A_2(s)|}$ and

$$U = \begin{bmatrix} U_1 & & & \\ & U_2 & & \\ & & \ddots & \\ & & & U_z \end{bmatrix}, U_{s'} = \Big[p^*(s'|s) r(s', a_1, a_2) \Big]_{|A_1(s')| \times |A_2(s')|} \text{ for each } s' \in S,$$

$$V = \begin{bmatrix} V_1 & & & \\ & V_2 & & \\ & & \ddots & \\ & & & V_z \end{bmatrix}, V_{s'} = \Big[p^*(s'|s) \tau(s', a_1, a_2) \Big]_{|A_1(s')| \times |A_2(s')|} \text{ for each } s' \in S,$$

the other elements in U and V are zero. Note that

$$\begin{aligned} x^t U y &= x_1^t U_1 y_1 + \cdots + x_z^t U_z y_z \\ &= \sum_{a_1 \in A_1(1)} \sum_{a_2 \in A_2(1)} p^*(1|s) r(1, a_1, a_2) f_1(1, a_1) f_2(1, a_2) + \cdots \\ &\quad + \sum_{a_1 \in A_1(z)} \sum_{a_2 \in A_2(z)} p^*(z|s) r(z, a_1, a_2) f_1(z, a_1) f_2(z, a_2) \\ &= \sum_{s' \in S} \sum_{a_1 \in A_1(s')} \sum_{a_2 \in A_2(s')} p^*(s'|s) r(s', a_1, a_2) f_1(s', a_1) f_2(s', a_2) \\ &= \sum_{s' \in S} p^*(s'|s) r(s', f_1, f_2). \end{aligned}$$

Similarly, $x^t V y = \sum\limits_{s' \in S} p^*(s'|s) \tau(s', f_1, f_2)$.

The minimax theorem of rational form $\frac{x^t U y}{x^t V y}$ was established by Loomis [9]. Thus, we obtain a pair of stationary optimal strategies (f_1^{*s}, f_2^{*s}) for the initial state $s \in S$. Then the pair of semi-stationary strategies (g_1^*, g_2^*), where $g_1^* = (f_1^{*1}, f_1^{*2}, \ldots, f_1^{*z})$ and $g_2^* = (f_2^{*1}, f_2^{*2}, \ldots, f_2^{*z})$, is optimal for the semi-Markov game and the theorem is proved. □

Theorem 4 cannot be strengthened further in the sense that an action independent transition semi-Markov game may not have a stationary optimal strategy pair. This fact follows from the following example.

[1] We use the notation $|A|$ to denote the number of elements in the set A.

Example 1 Consider an SMG with state space $S = \{1, 2, 3\}$, action sets $A_1(1) = A_2(1) = A_1(3) = A_2(3) = \{1\}$, $A_1(2) = A_2(2) = \{1, 2\}$. Rewards, transition probabilities and mean sojourn times are given below.

State 1:

1 (1, 0, 0) 1

State 2:

3 (0, 1, 0) 4	3.76 (0, 1, 0) 5
2 (0, 1, 0) 5	1 (0, 1, 0) 3

State 3:

5 $(\frac{1}{4}, \frac{3}{4}, 0)$ 2

where a cell

r (q_1, q_2, q_3) τ

specifies that r is the immediate reward; q_1, q_2 and q_3 are the transition probabilities that the next states are states 1, 2 and 3 respectively and τ is the expected sojourn time if this cell is chosen. Note that states 1 and 2 are absorbing. Thus, there exists an optimal pair of stationary strategies when the initial states are states 1 and 2. We have

$$\phi(1) = val[1] = 1 \text{ and } \phi(2) = val\begin{bmatrix} \frac{3}{4} & \frac{3.76}{5} \\ \frac{2}{5} & \frac{1}{3} \end{bmatrix} = \frac{3}{4},$$

where val denotes the minimax value of the matrix game. The unique optimal stationary strategies for states 1 and 2 are given by $(f_1^*(1), f_2^*(1)) = ((1), (1))$ and $(f_1^*(2), f_2^*(2)) = ((1, 0), (1, 0))$. Let $f_1^* = \{f_1^*(1), f_1^*(2), f_1^*(3)\} = \{(1), (1, 0), (1)\}$ and $f_2^* = \{f_2^*(1), f_2^*(2), f_2^*(3)\} = \{(1), (1, 0), (1)\}$ which are extensions of f_1^* and f_2^* for state 3. It follows that

$$\phi(3, f_1^*, f_2^*) = \frac{\frac{1}{4} \times 1 + \frac{3}{4} \times 3}{\frac{1}{4} \times 1 + \frac{3}{4} \times 4} = 10/13.$$

If we choose $f_2 = \{f_2(1), f_2(2), f_2(3)\} = \{(1, 0), (0, 1), (1)\}$, then it is easy to see that

$$\phi(3, f_1^*, f_2) = 12.28/16 < 10/13 = \phi(3, f_1^*, f_2^*)$$
$$\text{and } \phi(2, f_1^*, f_2) = 3.76/5 > 3/4 = \phi(2, f_1^*, f_2^*).$$

Thus, player 2 has no optimal in the class of stationary strategies.

4 Conclusions

In this paper, we investigate the role of semi-stationary strategies in studying the general limiting ratio average (undiscounted) semi-Markov games. We obtain a structured class of multichain zero-sum two-person action independent transition semi-Markov games that possess semi-stationary optimal strategies. For the nonzero-sum case of such games, the existence of Nash equilibrium is still an open problem. It is also interesting to find other structured classes of semi-Markov games which possess semi-stationary optimal/Nash equilibrium strategies under undiscounted payoff criterion.

Acknowledgements The authors wish to thank the unknown referees who have patiently gone through this paper and whose suggestions have improved its presentation and readability considerably.

References

1. Bertsekas, D.P., Shreve, S.E.: Stochastic optimal control: The discrete time case. Academic Press NY, 139 (1978)
2. Blackwell, D., Ferguson, T.S.: The big match. Ann. Math. Statist **39**(1), 159–163 (1968)
3. Doob, J.L.: Stochastic Processes. Willey ISBN 0-471-52369 (1953)
4. Feinberg, E.A.: Constrained Semi-Markov decision processes with average rewards. Math. Methods Operat. Res. **39**(3), 257–288 (1994)
5. Ghosh, M.K., Sinha, S.: Non-cooperative N-person semi-Markov games on a denumerable state space. Comp. Appl. Math. **21**, 833–847 (2002)
6. Jaskiewicz, A.: Zero-sum semi-Markov games. SIAM journal on control and optimization **41**(3), 723–739 (2002)
7. Jianyong, L., Xiaobo, Z.: On average reward semi-Markov decision processes with a general multichain structure. Math. Operat. Res. **29**(2), 339–352 (2004)
8. Lal, A.K., Sinha, S.: Zero-sum two-person semi-Markov games. J. Appl. Prob. **29**(1), 56–72 (1992)
9. Loomis, L.H.: On a theorem of von Neumann. Proc. Nat. Acad. Sci. **32**, 213–215 (1946)
10. Mondal, P., Sinha, S.: Ordered field property in a subclass of finite SeR-SIT Semi-Markov Games. Int. Game Theor. Rev. **15**(4) (2013). 1340026-1-1340026-20
11. Mondal, P., Sinha, S.: Ordered Field Property for Semi-Markov Games when One Player Controls Transition Probabilities and Transition Times, International Game Theory Review, 17(2). (2015). 1540022-1-1540022-26
12. Mondal, P., Sinha, S., Neogy, S.K., Das, A.K.: On Discounted AR-AT Semi-Markov games and its complementarity formulations. Int. J. Game Theor. **45**(3), 567–583 (2016)
13. Mondal, P.: Linear Programming and Zero-Sum Two-Person Undiscounted Semi-Markov Games. Asia-Pacific Journal of Operational Research, 32(6), 1550043-1-1550043-20 (2015)
14. Mondal, P.: On undiscounted Semi-Markov decision processes with absorbing states. Math. Methods Oper. Res. **83**(2), 161–177 (2016)
15. Mondal, P.: On zero-sum two-person undiscounted Semi-Markov games with a multichain structure. Advanc. Appl. Probab. **49**(3), 826–849 (2017)
16. Polowczuk, W.: Nonzero-sum semi-Markov games with countable state spaces. Applicationes Mathematicae **27**(4), 395–402 (2000)
17. Shapley, L.: Stochastic Games. Proc. Natl. Acad. Sci. USA **39**(10), 1095–1100 (1953)

18. Sinha, S., Mondal, P.: Semi-Markov decision processes with limiting ratio average rewards. J. Math. Anal. Appl. **455**(1), 864–871 (2017)
19. Stern, M.A.: On Stochastic Games with Limiting Average Payoff. University of Illinois at Chicago Circle, Chicago, Illinois, PhD Thesis (1975)
20. Vega-Amaya, O.: Zero-sum average semi-Markov games: fixed-point solutions of the Shapley equation. SIAM J. Control Optimizat. **42**(5), 1876–1894 (2003)

Max Plus Algebra, Optimization and Game Theory

Dipti Dubey, S. K. Neogy and Sagnik Sinha

Abstract Max-plus algebra has been applied to several fields like matrix algebra, cryptography, transportation, manufacturing, information technology and study of discrete event systems like subway traffic networks, parallel processing systems, telecommunication networks for many years. In this paper, we discuss various optimization problem using methods based on max-plus algebra, which has maximization and addition as its basic arithmetic operations. We present some sub-classes of mathematical optimization problems like linear programming, convex quadratic programming problem, fractional programming problem, bimatrix game problem and some classes of stochastic game problem in max algebraic framework and discuss various connections between max-plus algebra and optimization.

Keywords Max-plus algebra · Linear programming · Convex quadratic programming problem · Fractional programming problem · Bimatrix game problem

1 Introduction

We begin by introducing max-plus algebra or max algebra (in short) defined by the algebraic structure $(\mathbb{R}_{max}, \oplus, \otimes) = (\mathbb{R}_{max}, \max, +)$ where $\mathbb{R}_{max} = \mathbb{R} \cup -\infty$. In max-plus algebra, the classical arithmetic operations of addition and multiplication are replaced by maximum and addition respectively. We denote $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$ where $a, b \in \mathbb{R}_{max}$. The max-plus algebra started developing in the late 1950s, just after the area of Operations Research started to develop [11].

D. Dubey · S. K. Neogy (✉)
Indian Statistical Institute, 7 S.J.S. Sansanwal Marg, New Delhi 110016, India
e-mail: skn@isid.ac.in

D. Dubey
e-mail: diptidubey@isid.ac.in

S. Sinha
Jadavpur University, Kolkata, India

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_29

Several systems which are not linear in the conventional algebra turn out to be linear in max algebra and hence become ammenable to mathematical treatment, similar to that we use for linear systems. Over the time max-algebra has drawn lot of attention to the researchers due to the fact that both its algebraic operations are commutative and associative, and also they satisfy the distributive law. Therefore many of the basic tools from classical linear algebra are also usable in max algebra.

The list of algebraic properties of max plus algebra [17] are listed below. We summarize below the well known key points about max-plus algebra or max algebra.

- The algebraic structure $(\mathbb{R}_{max}, \oplus, \otimes) = (\mathbb{R} \cup -\infty, \max, +)$ is called the max-plus algebra.
- We introduce the notation of $\varepsilon = -\infty$ and $e = 0$ following Baccelli et al. [2].
- Notations ε, e is used instead of $-\infty$ and 0 respectively for emphasis their special meanings and to avoid confusion with their roles in conventional algebra.
- Algebraic structure $\mathbb{R}_{max}$ is an example of a commutative, idempotent semiring (or dioid).

This structure satisfies **axioms: Operation** $\oplus$

- Associativity: $x \oplus (y \oplus z) = (x \oplus y) \oplus z \;\forall\, x, y, z \in \mathbb{R}_{max}$
- Commutativity: $x \oplus y = y \oplus x \;\forall\, x, y \in \mathbb{R}_{max}$
- Existence of a zero element(ε): $x \oplus \varepsilon = \varepsilon \oplus x = x, \;\forall\, x \in \mathbb{R}_{max}$
- idempotency of $\oplus$: $x \oplus x = x \forall\, x \in \mathbb{R}_{max}$,

axioms: Operation $\otimes$

- Associativity: $x \otimes (y \otimes z) = (x \otimes y) \otimes z \;\forall\, x, y, z \in \mathbb{R}_{max}$.
- Commutativity: $x \otimes y = y \otimes x \;\forall\, x, y \in \mathbb{R}_{max}$.
- Existence of a unit element(e): $x \otimes e = e \otimes x = x \;\forall\, x \in \mathbb{R}_{max}$.
- existence of absorbing element(ε): $x \otimes \varepsilon = \varepsilon \otimes x = \varepsilon \;\forall\, x \in \mathbb{R}_{max}$.
- Distributivity of $\otimes$ over $\oplus$: $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z) \;\forall\, x, y, z \in \mathbb{R}_{max}$.

The structure $(\mathbb{R}_{max}, \oplus, \otimes)$ satisfies all the semi-ring axioms, i.e.,

- $\oplus$ is associative, commutative, with zero element ε.
- $\otimes$ is associative, distributes over $\oplus$ and $\otimes$ has a unit element e.
- ε is absorbing for $\otimes$.

Such a structure satisfying above axioms is called semi-ring. This semi-ring is commutative if $x \otimes y = y \otimes x$, idempotent if $x \oplus x = x$. The term dioid is used some times for idempotent semi-ring. Therefore Max-plus algebra is an example of a commutative and idempotent semiring. Note that max algebra is not a ring, because the operation $\oplus$ does not have a inverse. For instance the equation $a \oplus x = b$ does not have a solution if $a > b$. See [10] and [9] for details.

In this paper, we observe that this algebraic structure enables us to formulate and solve certain non-linear problems in a linear-like way. For details of algebraic properties of max-algebra and its applications see [3–7].

2 Notations and Definitions

We use the notation $\oplus = \max$ and $\otimes =$ plus. We consider max-plus matrices and max-plus vectors with entries from $\mathbb{R}_{\max}$. All max-plus vectors are column vectors unless otherwise specified. Let $p \in \mathbb{R}$. The pth max-algebraic power of $x \in \mathbb{R}$ is denoted by $x^{\otimes^p}$ and corresponds to px in conventional algebra. Let $M = \{1, \ldots, m\}$ and $N = \{1, \ldots, n\}$. An $m \times n$ max-plus matrix $A = [a_{ij}] \in \mathbb{R}_{max}^{m\times n}$ is an $m \times n$ array of entries where $a_{ij} \in \mathbb{R}_{max}$. Max-plus matrix addition and multiplication is analogous to classical matrix addition and multiplication. For three max-plus matrix $A, B \in \mathbb{R}_{max}^{m\times n}, C \in \mathbb{R}_{max}^{n\times p}$, we define $A \oplus B = [a_{ij} \oplus b_{ij}], k \otimes A = [k \otimes a_{ij}]$ $= [k + a_{ij}]$ where $k \in \mathbb{R}_{max}$ and $B \otimes C = [\oplus_{j=1}^{n} b_{ij} \otimes c_{jk}] = [\oplus_{j=1}^{n}(b_{ij} + c_{jk})] =$ $[\max_j(b_{ij} + c_{jk})]$ is an $m \times p$ matrix. The matrix operations are illustrated below in the following example.

Example 1

$$\begin{bmatrix} 0 & -1 & 1 \\ -2 & 6 & -5 \\ 0 & -3 & 9 \end{bmatrix} \oplus \begin{bmatrix} 5 & -2 & 3 \\ -5 & 6 & 2 \\ 2 & -5 & 7 \end{bmatrix} = \begin{bmatrix} 5 & -1 & 3 \\ -2 & 6 & 2 \\ 2 & -3 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 5 & -2 & 3 \\ -5 & 6 & 2 \end{bmatrix} \otimes \begin{bmatrix} 1 & -2 \\ -1 & 7 \\ 4 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} (5+1) \oplus (-2+(-1)) \oplus (3+4) & (5+(-2)) \oplus (-2+7) \oplus (3+2) \\ (-5+1) \oplus (6+(-1)) \oplus (2+4) & (-5+(-2)) \oplus (6+7) \oplus (2+2) \end{bmatrix} = \begin{bmatrix} 7 & 5 \\ 6 & 13 \end{bmatrix}$$

$$7 \otimes \begin{bmatrix} 5 & -2 & 3 \\ -5 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 5 & 10 \\ 2 & 13 & 9 \end{bmatrix}$$

The system

$$\begin{bmatrix} 5 & -2 & 3 \\ -5 & 6 & 2 \\ 2 & -5 & 7 \end{bmatrix} \otimes \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

written in the conventional notation are as follows.

$$\max(5 + x_1, -2 + x_2, 3 + x_3) = 5,$$

$$\max(-5 + x_1, 6 + x_2, 2 + x_3) = -2,$$

$$\max(2 + x_1, -5 + x_2, 7 + x_3) = 9.$$

Example 2 We illustrate the usefulness of max-plus algebra in real life using the following example. Suppose three flights leave three different cities Mumbai, Chennai and Kolkata but arrive at Delhi. From Delhi there is a flight to New York connecting all three flights to Delhi. Let us denote the departure times of the flights as x_1, x_2, x_3 and x_4 respectively. The given duration of the three flights (including the times for changing the aircraft) are d_1, d_2 and d_3, respectively. Here, we need to determine the departure times of flights of the flights from Mumbai, Chennai and Kolkata. Therefore

$$x4 = \max(x_1 + d_1, x_2 + d_2, x_3 + d_3).$$

In max algebraic framework this can be written as

$$x4 = x_1 \otimes d_1 \oplus x_2 \otimes d_2 \oplus x_3 \otimes d_3.$$

So if d is the given departure time of the flight from Delhi to New York then in order to find out x_1, x_2 and x_3, we need to solve the following max-algebraic linear equation.

$$x_1 \otimes d_1 \oplus x_2 \otimes d_2 \oplus x_3 \otimes d_3 = d.$$

2.1 System of Multivariate Polynomial Equalities and Inequalities in the Max Algebra

De Schutter and De Moor [12] consider the following problem involving a system of multivariate polynomial equalities and inequalities as a generalized framework for many important max-algebraic problems such as matrix decompositions, transformation of state space models, state space realization of impulse responses, construction of matrices with a given characteristic polynomial which is stated as follows.

Suppose $\{1, \ldots, \tilde{m}_k\}$ be a given a set of integers and $\{\tilde{a}_{ki}\}$, $\{\tilde{b}_k\}$ and $\{\tilde{c}_{kij}\}$ are three sets of coefficients where $i \in \{1, \ldots, \tilde{m}_k\}$, $j \in \{1, \ldots, n\}$, $k \in \{\tilde{p}_1 + 1, \ldots, \tilde{p}_1 + \tilde{p}_2\}$. The problem is to find a vector $z \in \mathbb{R}^n$ that satisfies

$$\bigoplus_{i=1}^{\tilde{m}_k} \tilde{a}_{ki} \otimes \bigotimes_{j=1}^{n} z_j^{\otimes^{\tilde{c}_{kij}}} = \tilde{b}_k \text{ for } k = 1, \ldots, \tilde{p}_1, \tag{1}$$

$$\bigoplus_{i=1}^{\tilde{m}_k} \tilde{a}_{ki} \otimes \bigotimes_{j=1}^{n} z_j^{\otimes^{\tilde{c}_{kij}}} \leq \tilde{b}_k \text{ for } k = \tilde{p}_1 + 1, \ldots, \tilde{p}_1 + \tilde{p}_2. \tag{2}$$

or show that no such vector z exists.

In conventional algebra this can be written as

$$\max_{i=1,\ldots,\tilde{m}_k}(\tilde{a}_{ki}+\sum_{j=1}^{n}\tilde{c}_{kij}z_j)=\tilde{b}_k \text{ for } k=1,\ldots,\tilde{p}_1,$$

$$\max_{i=1,\ldots,\tilde{m}_k}(\tilde{a}_{ki}+\sum_{j=1}^{n}\tilde{c}_{kij}z_j)\leq\tilde{b}_k \text{ for } k=\tilde{p}_1+1,\ldots,\tilde{p}_1+\tilde{p}_2.$$

De Schutter and De Moor [12] also observe that the general solution set of a system of multivariate max-algebraic polynomial equalities and inequalities is the union of a set of bounded and unbounded polyhedra. Further a method was developed to find all solutions of a system of multivariate polynomial equalities and inequalities in the max algebra.

2.2 Max Algebraic Linear System and Set Covering Problem

In this section, we discuss about the equivalence of a max algebraic linear system and a set covering problem. See [8] and [24] for details.

Let $A\otimes x=b$ is a max-algebraic linear system where $A=[a_{ij}]\in\mathbb{R}^{m\times n}_{max}$ and $b\in\mathbb{R}^{m}_{max}$. This can be written as

$$\max_{j=1,\ldots,n}(a_{ij}+x_j)=b_i \ (i=1,\ldots,m).$$

By subtracting the right-hand side values we get a linear system whose all right hand-side constants are zero

$$\max_{j=1,\ldots,n}(a_{ij}-b_i+x_j)=0 \ (i=1,\ldots,m).$$

Using the max-plus matrix notation, this is written as

$$B\otimes A\otimes x=B\otimes b=0$$

where

$$B=\operatorname{diag}\left[b_1^{\otimes^{-1}}\cdots b_m^{\otimes^{-1}}\right].$$

This linear system with all right hand-side constants zero ($A\otimes x=0$) is called normalized linear system and the process obtaining it is called *normalization.*

Let $S=\{x\in\mathbb{R}^n : A\otimes x=b\}$ and $M_j=\{k\in M : a_{kj}-b_k=\max_{i\in M}(a_{ij}-b_i)\,\forall\, j\in N\}$ where $M=\{1,\ldots,m\}$ and $N=\{1,\ldots,n\}$. Let $x^*_j=-\max_{i\in M}(a_{ij}-b_i)\,\forall\, j\in N\}$.

In what follows we consider the following two problems.

1. [Unique] Solvability: Given $A \in \mathbb{R}_{max}^{m\times n}$ and $b \in \mathbb{R}_{max}^{m}$ does the system $A \otimes x = b$ have a [unique] solution?
2. [Minimal] Set Covering: Given a finite set $M = \{1, \ldots, m\}$ and subsets $M_1, \ldots, M_n$ of M is $\cup_{j=1}^{n} M_j = M$ (is $\cup_{j=1}^{n} M_j = M$ but $\cup_{j=1, j\neq k}^{n} M_j \neq M$ for some $k \in \{1, \ldots, n\}$)?

The following result is observed in [8] and [24].

Theorem 1 *1.* $S \neq \emptyset \Longleftrightarrow \cup_{j=1}^{n} M_j = M$
2. $|S| = 1 \Longleftrightarrow \cup_{j=1}^{n} M_j = M, \cup_{j\in N'} M_j \neq M$ *for any* $N' \subseteq N, N' \neq N$.

The above theorem shows that the solvability of max algebraic linear system is equivalent to set covering and unique solvability of max algebraic linear system is equivalent to minimal set covering.

Example 3 We make use of the example provided by Butkovic [8] for illustration of the procedure. Consider the max algebraic linear system $A \otimes x = b$ where

$$A = \begin{bmatrix} -2 & 2 & 2 \\ -5 & -3 & -2 \\ -5 & -3 & 3 \\ -3 & -3 & 2 \\ 1 & 4 & 6 \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \; b = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 5 \end{bmatrix}$$

For each $i \in M$, we compute $(a_{ij} - b_i) \, \forall \, j \in N$ we obtain the following matrix

$$\bar{A} = \begin{bmatrix} -5 & -1 & -1 \\ -3 & -1 & 0 \\ -6 & -4 & 2 \\ -3 & -3 & 2 \\ -4 & -1 & 1 \end{bmatrix}. \tag{3}$$

So, $x_1^* = -\max\{-5, -3, -6, -3, -4\} = 3$. Therefore $-x_1^*$ is maximum entry of column 1, $-x_2^*$ is maximum entry of column 2 and $-x_3^*$ is maximum entry of column 3 of the matrix $\bar{A}$. Hence $x_1^* = 3$, $x_2^* = 1$ and $x_3^* = -2$.

3 Max Version of Various Optimization Problem

In this section we consider a class of optimization problems which lead to systems of multivariate max-algebraic polynomial equalities and inequalities. First we look at linear programming problem.

3.1 Linear Programming

Consider the primal linear program (P): Minimize $\tilde{c}^T\tilde{x}$ subject to $A\tilde{x} \geq \tilde{b}$, $\tilde{x} \geq 0$ and its dual (D): Maximize $\tilde{b}^T\tilde{y}$ subject to $A^T\tilde{y} \leq \tilde{c}$, $\tilde{y} \geq 0$ where $A \in \mathbb{R}^{m\times n}$, $\tilde{b} \in \mathbb{R}^m$ and $\tilde{c} \in \mathbb{R}^n$.

If $\tilde{x}^*$ is feasible to P and $\tilde{y}^*$ is feasible to D then $\tilde{x}^*$, $\tilde{y}^*$ are optimal to the respective problem iff

$$\tilde{y}^*(A\tilde{x}^* - \tilde{b}) + \tilde{x}^*(\tilde{c} - A^T\tilde{y}^*) = 0.$$

This implies

$$\tilde{y}^*(A\tilde{x}^* - \tilde{b}) = 0,$$

$$\tilde{x}^*(\tilde{c} - A^T\tilde{y}^*) = 0.$$

In max algebraic notation solving an LP reduces to a multivariate max-algebraic polynomial equalities and inequalities system as follows:

Find $\tilde{x} \in \mathbb{R}^n$, $\tilde{y} \in \mathbb{R}^m$ such that

$$\tilde{b}_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} \tilde{x}_j^{\otimes^{a_{ij}}} \geq 0 \text{ for } i = 1, \ldots, m, \tag{4}$$

$$\tilde{c}_i \bigotimes_{j=1}^{m} \tilde{y}_j^{\otimes^{(-a_{ji})}} \geq 0 \text{ for } i = 1, \ldots, n, \tag{5}$$

$$\bigoplus_{i=1}^{n} \tilde{x}_i \oplus \tilde{c}_i \bigotimes_{j=1}^{m} \tilde{y}_j^{\otimes^{(-a_{ji})}} = 0. \tag{6}$$

$$\bigoplus_{i=1}^{m} \tilde{y}_i \oplus \tilde{b}_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} \tilde{x}_j^{\otimes^{a_{ij}}} = 0. \tag{7}$$

3.2 Convex Quadratic Programming

Quadratic programming problems arises in a number of applications in economics and engineering. Hence through quadratic and linear programming problem, complementary slackness principle is also highly useful in the economic theory and models and has been recognized as equilibrium condition.

Consider the convex Quadratic Programming problem.

Minimize $f(x) = \tilde{c}^T x + \frac{1}{2} x^T Qx$ subject to $Ax \geq \tilde{b}, x \geq 0$ where $Q \in \mathbb{R}^{n \times n}$ is symmetric, $A \in \mathbb{R}^{m \times n}$, $\tilde{b} \in \mathbb{R}^m$ and $\tilde{c} \in \mathbb{R}^n$.

The Karush-Kuhn-Tucker necessary and sufficient optimality conditions specialized to this problem yields the following equations and inequalities.

$$\tilde{c} + Q\tilde{x} - A^T \tilde{y} - \tilde{u} = 0$$
$$-A\tilde{x} + \tilde{v} = -\tilde{b}$$
$$\tilde{x}^T \tilde{u} = \tilde{y}^T \tilde{v} = 0$$
$$\tilde{x} \geq 0,\ \ \tilde{y} \geq 0,\ \ \tilde{u} \geq 0, \tilde{v} \geq 0$$

In max algebraic notation solving a convex quadratic programming problem reduces to a multivariate max-algebraic polynomial equalities and inequalities system as follows:

Find $\tilde{x} \in \mathbb{R}^n$, $\tilde{y} \in \mathbb{R}^m$ such that

$$\tilde{c}_i \bigotimes_{j=1}^{n} \tilde{x}_j^{\otimes^{(q_{ij})}} \bigotimes_{j=1}^{m} \tilde{y}_j^{\otimes^{(-a_{ji})}} \otimes \tilde{u}_i^{\otimes^{-1}} = 0 \text{ for } i = 1, \ldots, n, \tag{8}$$

$$\tilde{b}_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} \tilde{x}^{\otimes^{a_{ij}}} \geq 0 \text{ for } i = 1, \ldots, m, \tag{9}$$

$$\bigoplus_{i=1}^{m} \tilde{y}_i \oplus \tilde{b}_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} \tilde{x}_j^{\otimes^{a_{ij}}} = 0. \tag{10}$$

$$\bigoplus_{i=1}^{n} \tilde{x}_i \oplus \tilde{u}_i = 0 \tag{11}$$

3.3 Linear Fractional Programming Problem

The problem of minimizing a linear fractional function subject to linear inequality conditions also reduces to a multivariate max-algebraic polynomial equalities and inequalities system via the Karush-Kuhn-Tucker conditions.

Given a matrix $A \in \mathbb{R}^{m \times n}$, vectors $\tilde{b} \in \mathbb{R}^m, \tilde{c}, \tilde{d} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, the linear fractional programming problem is defined as follows:

$$\text{Minimize} f(\tilde{x}) = \frac{(\tilde{c}^T \tilde{x} + \alpha)}{(\tilde{d}^T \tilde{x} + \beta)} \tag{12}$$

subject to

$$A\tilde{x} \leq \tilde{b}, -\tilde{x} \leq 0 \tag{13}$$

Let $\mathcal{S} = \{\tilde{x} \mid A\tilde{x} \leq \tilde{b}, \tilde{x} \geq 0\}$. It is assumed that $\tilde{d}^T\tilde{x} + \beta \neq 0 \; \forall \; \tilde{x} \in \mathcal{S}$ or without loss of further generality, we assume that $\tilde{d}^T\tilde{x} + \beta > 0 \, \forall \, \tilde{x} \in \mathcal{S}$. With this assumption the function $f(\tilde{x})$ is both pseudo-convex and pseudo-concave. The Karush-Kuhn-Tucker optimality conditions are both necessary and sufficient for a point $\bar{x}$ to be a solution to (12) and (13). Thus $\bar{x}$ is a solution to (12) and (13) iff there exists a $\bar{y}, \bar{u}, \bar{v} \geq 0$ (where $\bar{y}, \bar{u} \in \mathbb{R}^m$, and $\bar{v} \in \mathbb{R}^n$) such that

$$\begin{aligned}
&\nabla f(\bar{x}) + A^T\bar{u} - \bar{v} = 0\\
&A\bar{x} + \bar{y} = \tilde{b}\\
&\bar{x}^T\bar{v} + \bar{y}^T\bar{u} = 0\\
&\bar{x} \geq 0, \bar{u} \geq 0\\
&\bar{v} \geq 0, \bar{y} \geq 0
\end{aligned}$$

For linear fractional programming problem, it is easy to compute $\nabla f(\bar{x})$. This is given by

$$\nabla f(\bar{x}) = (\tilde{d}^T\bar{x} + \beta)^{-2}[(\tilde{d}^T\bar{x} + \beta)\tilde{c} - (\tilde{c}^T\bar{x} + \alpha)\tilde{d}]$$

which reduces to $(\tilde{d}^T\bar{x} + \beta)^{-2}[\tilde{D}\bar{x} + \beta\tilde{c} - \alpha\tilde{d}]$.

Here $\tilde{D}$ is a $n \times n$ matrix whose ijth element $\tilde{d}_{ij}$ is given by $\tilde{d}_j\tilde{c}_j - \tilde{d}_i\tilde{c}_i$ for $1 \leq i \leq n,\; 1 \leq j \leq n$. We see that $\bar{x}$ is a solution to (12) and (13) iff there exists a $\bar{y} \in \mathbb{R}^m,\; \bar{u} \in \mathbb{R}^m$ and $\bar{v} \in \mathbb{R}^n$ such that

$$\begin{aligned}
&\tilde{D}\bar{x} + \beta\tilde{c} - \alpha\tilde{d} + A^T\bar{u} - \bar{v} = 0\\
&A\bar{x} + \bar{y} = \tilde{b}\\
&\bar{x}^T\bar{v} + \bar{y}^T\bar{u} = 0\\
&\bar{x} \geq 0, \bar{u} \geq= 0\\
&\bar{v} \geq 0, \bar{y} \geq 0
\end{aligned}$$

In max algebraic notation solving a fractional programming reduces to a multivariate max-algebraic polynomial equalities and inequalities system as follows:

Find $\bar{x} \in \mathbb{R}^n, \bar{y} \in \mathbb{R}^m$ such that

$$\tilde{c}_i^{\otimes^{\beta}} \otimes \tilde{d}_i^{\otimes^{-\alpha}} \bigotimes_{j=1}^{n} \bar{x}_j^{\otimes^{(\tilde{d}_{ij})}} \bigotimes_{j=1}^{m} \bar{u}_j^{\otimes^{a_{ji}}} \otimes \bar{v}_i^{\otimes^{-1}} = 0 \text{ for } i = 1, \ldots, n, \tag{14}$$

$$\tilde{b}_i \bigotimes_{j=1}^{n} \bar{x}_j^{\otimes^{-a_{ij}}} \geq 0 \text{ for } i = 1, \ldots, m, \tag{15}$$

$$\bigoplus_{i=1}^{m} \bar{u}_i \oplus \tilde{b}_i \bigotimes_{j=1}^{n} \bar{x}_j^{\otimes^{-a_{ij}}} = 0. \tag{16}$$

$$\bigoplus_{i=1}^{n} \bar{v}_i \oplus \bar{x}_i = 0 \tag{17}$$

3.4 Computation Nash Equilibrium Pair in Bimatrix Games Using Max-Plus Algebra

A bimatrix game is a noncooperative nonzero-sum two person game in which each player has a finite number of actions (known as pure strategies). Let player 1 have m pure strategies and player 2 have n pure strategies. In a game if player 1 chooses strategy i and player 2 chooses strategy j they incur the costs a_{ij} and b_{ij} respectively where $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ and $B = [b_{ij}] \in \mathbb{R}^{m \times n}$ are given cost matrices.

A mixed strategy for player 1 is a probability vector $x \in \mathbb{R}^m$. We denote mixed strategy for player 2 as a probability vector $y \in \mathbb{R}^n$. If player 1 adopts a mixed strategy x and player 2 adopts a mixed strategy y then their *expected costs* are given by $x^T Ay$ and $x^T By$ respectively.

A pair of mixed strategies (x^*, y^*) with $x^* \in \mathbb{R}^m$ and $y^* \in \mathbb{R}^n$ is said to be a *Nash equilibrium pair* if

$$(x^*)^T Ay^* \leq x^T Ay^* \; \forall \; x \in \mathbb{R}^m \;\text{ and}$$

$$(x^*)^T By^* \leq (x^*)^T By \; \forall \; y \in \mathbb{R}^n.$$

We assume that all entries of the matrices A and B are positive (since addition of a constant to all entries of A or B leaves the set of equilibrium points invariant). In max algebraic notation solving a bimatrix game reduces to a multivariate max-algebraic polynomial equalities and inequalities system as follows:

Find $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ such that

$$e_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} y_j^{\otimes^{a_{ij}}} \geq 0 \text{ for } i = 1, \ldots, m, \tag{18}$$

$$e_i^{\otimes^{-1}} \bigotimes_{j=1}^{m} x_j^{\otimes^{(b_{ji})}} \geq 0 \text{ for } i = 1, \ldots, n, \tag{19}$$

$$\bigoplus_{i=1}^{m} x_i \oplus e_i^{\otimes^{-1}} \bigotimes_{j=1}^{n} y_j^{\otimes^{(a_{ij})}} = 0. \tag{20}$$

$$\bigoplus_{i=1}^{n} y_i \oplus e_i^{\otimes^{-1}} \bigotimes_{j=1}^{m} x_j^{\otimes^{b_{ji}}} = 0. \tag{21}$$

where e_m and e_n are m vectors and n vectors whose components are all 1's.

Note that if (x^*, y^*) is a Nash equilibrium pair then $(\tilde{x}, \tilde{y})$ is a solution to (18)–(21) where

$$\tilde{x} = x^*/(x^*)^T By^* \;\text{ and }\; \tilde{y} = y^*/(x^*)^T Ay^*. \tag{22}$$

Conversely, if $(\tilde{x}, \tilde{y})$ is a solution of (18)–(21) then $\tilde{x} \neq 0$ and $\tilde{y} \neq 0$ in (22) is ensured from the positivity of the cost matrices A and B. Therefore (x^*, y^*) is a Nash equilibrium pair where

$$x^* = \tilde{x}/e_m^T\tilde{x} \text{ and } \ y^* = \tilde{y}/e_n^T\tilde{y}.$$

We can make use of the method proposed by Schutter and De Moor [12] to find all solutions of the system of multivariate polynomial equalities and inequalities in the max algebra that arises in the context of linear programming.

4 Application of Max Plus Algebra in Stochastic Games

In 1953, Shapley [21] introduced Stochastic games. The theory of stochastic games have been used to study many problems like search problems, military applications, advertising problems, the travelling inspector problem, and problems related to various economic applications [14]. In what follows, we first introduce stochastic games and its various subclasses with special structures. In this section, we consider the problem of solving a stochastic game using max algebraic approach.

A two-player zero-sum stochastic game with finite state/action space is defined by the following objects.

- A state space $\mathcal{S} = \{1, 2, \ldots, \mathcal{N}\}$.
- For each $s \in \mathcal{S}$, finite action sets $\mathcal{A}(s) = \{1, 2, \ldots, m_s\}$ for Player 1 and $\mathcal{B}(s) = \{1, 2, \ldots, n_s\}$ for Player 2.
- A reward law $R(s)$ for $s \in S$ where $R(s) = [r_{ij}(s)]$ is an $m_s \times n_s$ matrix whose (i, j)th entry denotes the payoff from Player 2 to Player 1 corresponding to the choices of action $i \in \mathcal{A}(s), \ j \in \mathcal{B}(s)$ by Player 1 and Player 2 respectively.
- A transition law $q = (q_{ij}(s, s') : \ (s, s') \in S \times S, \ i \in \mathcal{A}(s), \ j \in \mathcal{B}(s))$, where $q_{ij}(s, s')$ denotes the probability of a transition from state s to state s' given that Player 1 and Player 2 choose actions $i \in A(s)$, $j \in B(s)$ respectively.

The game is played in stages $t = 0, 1, 2, \ldots$ At some stage t, the players find themselves in a state $s \in \mathcal{S}$ and independently choose actions $i \in \mathcal{A}(s)$, $j \in \mathcal{B}(s)$. Player 2 pays Player 1 an amount $r_{ij}(s)$ and at stage $(t + 1)$, the new state is s' with probability $q_{ij}(s, s')$. Play continues at this new state.

The players guide the game via strategies. In general, strategies can depend on complete histories of the game until the current stage. We are however concerned with the simpler class of *stationary strategies* which depend only on the current state s and not on stages. So for Player 1, a stationary strategy

$$f \in F_{\mathcal{S}} = \{f_i(s) \ : \ s \in \mathcal{S}, i \in \mathcal{A}(s), f_i(s) \geq 0, \ \sum_{i \in \mathcal{A}(s)} f_i(s) = 1\}$$

indicates that the action $i \in \mathcal{A}(s)$ should be chosen by Player 1 with probability $f_i(s)$ when the game is in state s.

Similarly for Player 2, a stationary strategy

$$g \in G_{\mathcal{S}} = \{g_j(s) \ : \ s \in \mathcal{S}, \, j \in \mathcal{B}(s), \, g_j(s) \geq 0, \ \sum_{j \in \mathcal{B}(s)} g_j(s) = 1\}$$

indicates that the action $j \in \mathcal{B}(s)$ should be chosen with probability $g_j(s)$ by Player 2 when the game is in state s.

Here, $F_{\mathcal{S}}$ and $G_{\mathcal{S}}$ will denote the set of all stationary strategies for Player 1 and Player 2 respectively. Let $f(s)$ and $g(s)$ be the corresponding vectors of dimension m_s and n_s respectively.

Fixed stationary strategies f and g induce a Markov chain on S with transition matrix $P(f, g)$ whose (s, s')th entry is given by

$$P_{ss'}(f, g) = \sum_{i \in \mathcal{A}(s)} \sum_{j \in \mathcal{B}(s)} q_{ij}(s, s') f_i(s) g_j(s)$$

and the expected current reward vector $r(f, g)$ has entries defined by

$$r_s(f, g) = \sum_{i \in \mathcal{A}(s)} \sum_{j \in \mathcal{B}(s)} r_{ij}(s) f_i(s) g_j(s) = f^T(s) R(s) g(s).$$

With fixed general strategies f, g and an initial state s, the stream of expected payoff to Player 1 at stage t, denoted by $v_s^t(f, g)$, $t = 0, 1, 2, \ldots$ is well defined and the resulting discounted and undiscounted payoffs are

$$\phi_s^{\beta}(f, g) = \sum_{t=0}^{\infty} \beta^t v_s^t(f, g) \text{ for a } \beta \in (0, 1)$$

and

$$\phi_s(f, g) = \lim_{T \uparrow \infty} \inf \frac{1}{T+1} \sum_{t=0}^{T} v_s^t(f, g).$$

A pair of strategies (f^*, g^*) is optimal for Player 1 and Player 2 in the undiscounted game if for all $s \in \mathcal{S}$

$$\phi_s(f, g^*) \leq \phi_s(f^*, g^*) = v_s^* \leq \phi_s(f^*, g),$$

for any strategies f and g of Player 1 and Player 2 respectively. The number v_s^* is called the *value of the game* starting in state s and $v^* = (v_1^*, v_2^*, \ldots, v_{\mathcal{N}}^*)$ is called the *value vector.* The definition for discounted case is similar.

The games considered by Shapley [21] are now called two-person zero-sum discounted games with finite state and action space. In this fundamental paper, Shapley

[21] proved the existence of a value and optimal stationary strategies for discounted case and gave a method for iterative computation of the value of a stochastic game with discounted payoff. Subsequently, Gillette [16] studied undiscounted case or limiting average payoff case. Since then, most of the researchers in stochastic game have dealt with the problem of finding sufficient conditions for the existence of value and optimal or ϵ-optimal strategies. Shapley's stochastic games were also extended to non-zero-sum games by many researchers. See [15] and [22].

Researchers over the years have tried to identify those classes of zero-sum stochastic games for which there is a possibility of obtaining a finite step algorithm to compute a solution. Many of the results in this area are for zero-sum games with special structures. These zero-sum stochastic games with special structure are referred collectively as the class of *structured stochastic games*. For some known classes of structured stochastic games, it is known that optimal stationary strategies exist and the game satisfies the orderfield property (i.e., the solution to the game lies in the same ordered field as the data of the game (e.g., rational)).

Max algebraic approach may be used for some known classes of games which possess ordered field property. Note that in general, it is difficult to find a pair of equilibrium (optimal strategies) strategies. Of course one can approximate it in the discounted case as Shapley has done it in his seminal paper on stochastic games but it is not an efficient procedure. See also the excellent survey paper by Raghavan and Filar [20]. Also see Mohan et al. [19].

- **Single controller stochastic games**: In the case where player 2 is *single controller* this means $q_{ij}(s, s') = q_j(s, s')\ \forall\ i, j, s, s'$.
- **Switching controller stochastic games**: In a switching control stochastic game the law of motion is controlled by Player 1 alone when the game is played in a certain subset of states and Player 2 alone when the game is played in other states. In other words, a switching control game is a stochastic game in which the set $\mathcal{S}$ of states are partitioned into sets $\mathcal{S}_1$ and $\mathcal{S}_2$ where the transition function is given by

$$q_{ij}(s, s') = \begin{cases} q_i(s, s'), & \text{for } s' \in \mathcal{S}, s \in \mathcal{S}_1, i \in \mathcal{A}(s) \text{ and } \forall\ j\ \in \mathcal{B}(s), \\ q_j(s, s'), & \text{for } s' \in \mathcal{S}, s \in \mathcal{S}_2, j \in \mathcal{B}(s) \text{ and } \forall\ i\ \in \mathcal{A}(s). \end{cases}$$

Many of the researchers have attempted to formulate the problem of computing a value and optimal strategies as a complementarity model and obtain a finite step method.

Now we mention a few *open problems* from Raghavan and Filar [20].

Problem I: Characterize the class of stochastic games which possess the ordered field property.

Problem II: Is it possible to obtain a finite step algorithm for the class of stochastic games which possess the ordered field property?

This problem is known to have an affirmative answer for the class of games discussed above in this section.

Filar [13] introduced the class of switching controller (SC) stochastic games as a natural generalization of the single controller game. However from the algorithmic point of view this class of games appear to be more difficult. The special game structure was used to develop a finite step algorithm by Vrieze et al. [23], but that algorithm requires solving a large number of single control stochastic games.

5 Discounted Switching Controller Stochastic Games

An algorithm was proposed by Mohan and Raghavan [18] for discounted switching controller games which is based on two linear programs. Even though this procedure converges to the value vector, it need not terminate in finitely many steps since basis vectors vary continuously. It is possible to put in the max algebraic framework using the following result.

Theorem 2 *A β-discounted zero-sum stochastic game possesses values v_s^β for $s \in \mathcal{S}$ and optimal stationary strategies f and g for player 1 and 2 respectively, if and only if (v^β, f, g) solves the following nonlinear system. Find (v^β, f, g) where $v_s^\beta \in R^{|\mathcal{S}|}$ such that*

$$v_s^\beta \bigotimes_{s' \in S} q_i(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [R(s)g(s)]_i^{\otimes^{-1}} \geq 0, \;\; i \in \mathcal{A}(s), s \in \mathcal{S}_1 \tag{23}$$

$$v_s^{\beta^{\otimes^{-1}}} \bigotimes_{s' \in S} q_j(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [f(s)R(s)]_j \geq 0, \;\; j \in \mathcal{B}(s), s \in \mathcal{S}_2 \tag{24}$$

$$f \in F_s, g \in G_s \tag{25}$$

Remark 1 Note that if v_s^β, $f(s)$, $g(s)$ satisfy (23), (24) and (25) then

$$v_s^\beta = [\beta P(f,g)v^\beta]_s + r_s(f,g) \tag{26}$$

In (26), the quadratic form $f(s)R(s)g(s)$ has notation $r_s(f,g)$.

We show that solving a discounted switching control game is equivalent to solving the following multivariate max-algebraic polynomial equalities and inequalities system.

Theorem 3 *A β-discounted switching control stochastic game has values v_s^β for $s \in \mathcal{S}$ and optimal stationary strategies f and g, if and only if*

$$f \in F_{\mathcal{S}}, \;\; g \in G_{\mathcal{S}} \tag{27}$$

$$v_s^\beta \bigotimes_{s' \in S} q_i(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [R(s)g(s)]_i^{\otimes^{-1}} \geq 0, \;\; i \in \mathcal{A}(s), s \in \mathcal{S}_1 \tag{28}$$

$$v_s^\beta \otimes {\theta_s}^{\otimes^{-1}} \otimes [R(s)g(s)]_i{}^{\otimes^{-1}} \geq 0,\ \ i \in \mathcal{A}(s), s \in \mathcal{S}_2 \tag{29}$$

$$v_s^{\beta^{\otimes^{-1}}} \otimes \theta_s \otimes [f(s)R(s)]_j \geq 0,\ \ j \in \mathcal{B}(s), s \in \mathcal{S}_1 \tag{30}$$

$$v_s^{\beta^{\otimes^{-1}}} \bigotimes_{s' \in S} q_j(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [f(s)R(s)]_j \geq 0,\ \ j \in \mathcal{B}(s), s \in \mathcal{S}_2 \tag{31}$$

$$\bigoplus_{i \in \mathcal{A}(s)} f_i(s) \oplus v_s^\beta \bigotimes_{s' \in S} q_i(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [R(s)g(s)]_i{}^{\otimes^{-1}} = 0, s \in \mathcal{S}_1 \tag{32}$$

$$\bigoplus_{i \in A(s)} f_i(s) \oplus v_s^\beta \otimes {\theta_s}^{\otimes^{-1}} \otimes [R(s)g(s)]_i{}^{\otimes^{-1}} = 0, s \in \mathcal{S}_2 \tag{33}$$

$$\bigoplus_{j \in B(s)} g_j(s) \oplus v_s^{\beta^{\otimes^{-1}}} \otimes \theta_s \otimes [f(s)R(s)]_j = 0, s \in \mathcal{S}_1, s \in \mathcal{S}_2 \tag{34}$$

$$\bigoplus_{j \in \mathcal{B}(s)} g_j(s) \oplus v_s^{\beta^{\otimes^{-1}}} \bigotimes_{s' \in S} q_j(s,s')^{\otimes^{-\beta v_{s'}^\beta}} \otimes [f(s)R(s)]_j = 0, s \in \mathcal{S}_1, s \in \mathcal{S}_2 \tag{35}$$

Remark 2 We recommend to use the method proposed by Schutter and De Moor [12] for finding the solutions of the above system of multivariate polynomial equalities and inequalities in max algebraic frame work to obtain the value vector and optimal stationary strategies for a discounted switching control game. It is well known that finding a finite step alogorithm to compute value vector and optimal stationary strategies is an open problem in stochastic games [20] for last two decades. Max algebraic approach has a huge potential for application of other classes of structured stochastic games and semi-Markov games.

6 Conclusion

Many of the theorems and techniques available in classical linear algebra have max-plus analogues. Most of the initial studies carried out on this topic were limited to what are called path algebras in recent literature. Max-plus semi-ring has been successfully applied to discrete event systems and dynamic programming. See [1] for a nice survey on this topic. Application in max-Algebraic framework has been successfully extended to other areas like analytic hierarchy process, multi-objective optimization, scheduling problems, cryptography, combinatorial optimization, genetic algorithms and queueing theory.

So far we have discussed max-plus algebra in finite dimensional settings. However, applications of max-plus algebra in infinite dimensional settings is an emerging area of research.

Acknowledgements The authors would like to thank the anonymous referees for their constructive suggestions which considerably improve the overall presentation of the paper. The authors would like to thank Professor R. B. Bapat, Indian Statistical Institute, Delhi Centre for his valuable comments and suggestions. The first author wants to thank the Science and Engineering Research Board, Department of Science & Technology, Government of India, for financial support for this research.

References

1. Akian, M., Bapat, R.B., Gaubert, S.: Max-plus algebra. Hogben, L., Brualdi, R., Greenbaum, A., Mathias, R. (eds) Handbook of linear algebra, discrete mathematics and its applications, Vol. 39, 2nd edn, Chapman and Hall (2014)
2. Baccelli, F., Cohen, G., Olsder, G.J., Quadrat, J.-P.: Synchronization and linearity. Wiley, New York (1992)
3. Bapat, R.B., Stanford, D., van den Driessche, P.: The eigenproblem in max algebra, DMS-631-IR, University of Victoria, British Columbia (1993)
4. Bapat, R.B.: Max algebra and graph theory. In: Krishnamurthy, Ravichandran, N. (eds) Proceedings of the Advance Workshop and Tutorial in Operations Research, Allied Publisher Pvt Ltd. (2012)
5. Bapat, R.B.: Pattern properties and spectral inequalities in max algebra. SIAM J. Matrix Anal. Appl. **16**, 964–976 (1995)
6. Bouquard, J.-L., Lent, C., Billaut, J.-C.: Application of an optimization problem in max-plus algebra to scheduling problems. Discrete Appl. Math. **154**, 2064–2079 (2006)
7. Burkard, R.E., Butkovic, P.: Max algebra and the linear assignment problem. Math. Program. Ser. B **98**, 415–429 (2003)
8. Butkovic, P.: Max-algebra: the linear algebra of combinatorics? Linear Algebra Appl. **367**, 313–335 (2003)
9. Butkovic, P.: Max-linear systems: theory and algorithms. Springer, Berlin (2010)
10. Cuninghame-Green, R.A.: Minimax algebra, lecture notes in economics and mathematical systems (1979)
11. Cuninghame-Green, R.A.: Minimax algebra. Lecture notes in economics and mathematical systems, pp. 166. Springer, Berlin, New Yotk (1979)
12. De Schutter, B., De Moor, B.: The extended linear complementarity problem. Math. Program. **71**(3), 289–325 (1995)
13. Filar, J.A.: Orderfield property for stochastic games when the player who controls transitions changes from state to state. JOTA **34**, 503–515 (1981)
14. Filar, J.A., Vrieze, O.J.: Competitive markov decision processes. Springer, New York (1997)
15. Fink, A.M.: Equilibrium in a stochastic n-person game. J. Sci. Hiroshima Univ. Ser. A **28**, 89–93 (1964)
16. Gillette, D.: Stochastic game with zero step probabilities. In: Tucker, A.W., Dresher, M., Wolfe, P. (eds) Theory of games. Princeton University Press, Princeton, New Jersey (1957)
17. Heidergott, B., Jan Olsder, G., van der Woude, J.: Max plus at work modeling and analysis of synchronized systems: a course on max-plus algebra and its applications, Princeton University Press (2006)
18. Mohan, S.R., Raghavan, T.E.S.: An algorithm for discounted switching control games. OR Spektrum **9**, 41–45 (1987)
19. Mohan, S.R., Neogy, S.K., Parthasarathy, T.: Pivoting algorithms for some classes of stochastic games: a survey. Int. Game Theory Rev. **3**, 253–281 (2001)

20. Raghavan, T.E.S., Filar, J.A.: Algorithms for stochastic games, a survey. Zietch. Oper. Res. **35**, 437–472 (1991)
21. Shapley, L.S.: Stochastic games. Proc. Natl. Acad. Sci. **39**, 1095–1100 (1953)
22. Sobel, M.J.: Noncooperative stochastic games. Ann. Math. Stat. **42**, 1930–1935 (1971)
23. Vrieze, O.J., Tijs, S.H., Raghavan, T.E.S., Filar, J.A.: A finite algorithm for the switching controller stochastic game. OR Spektrum **5**, 15–24 (1983)
24. Zimmermann, U.: Linear and combinatorial optimization in ordered algebraic structures. North-Holland, Amsterdam (1981)

Dynamics of Infectious Diseases with Periodic Awareness Campaigns

Fahad Al Basir

Abstract In this article, we investigate the effect of an awareness campaign on the prevalence of infectious diseases, assuming the awareness campaign is organised periodically. On this basis, an SIS model is developed, considering susceptible and infected humans, using impulsive differential equations to describe the awareness campaign. To attain an effective control of the disease, the period and rate of awareness is determined using the mathematical model. Numerical simulations illustrate the analytical outcomes.

Keywords Mathematical model · Impulsive differential equation · Basic reproduction number · Stability analysis · Numerical simulations

1 Introduction

Infectious diseases that cause mortality, disability, and social and economic disruption are a major threat to mankind, which are responsible for a quarter of all deaths annually in the world [1]. Once an infectious disease appears and spreads in a region, the government or health organizations do their best to stop the propagation of the disease. One of the way is to make people aware of the appropriate preventive knowledge of diseases as soon as possible through media and education which make people take precautions to reduce their chances of being infected [2].

The prevalence of any epidemic is strongly dependent on the social behavior of individuals in a population which makes them susceptible to infection [3, 4]. Behavioral changes alone are capable of making a significant difference on the size of epidemic [5]. By conducting campaigns through media, human behavior can be modified towards the disease. These campaigns may create awareness among various high-risk groups which help in limiting the spread of infection. Such campaigns basically focus upon increasing people's knowledge about disease transmission and facilitate measures that can reduce chances of being infected [4]. Reducing daily

F. Al Basir (✉)
Department of Mathematics, Asansol Girls' College, Asansol 713304, West Bengal, India
e-mail: fahadbasir@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_30

contacts in response to information about the presence of disease is a rational action that can be adopted by many people in the community [6].

In developing and underdeveloped countries, the mass media plays an important role in changing behavior related to public health [7]. The media not only make the population acquainted with the disease but also suggest the necessary preventive practices such as social distancing, wearing protective masks or vaccination. In general the people who are aware adopt these practices so that their chances of becoming infected are minimized. As the awareness disseminates, people respond toward it and eventually will change their behavior to alter their susceptibility [8].

Recently, mathematical models have investigated the impact on disease spread and control [9]. In many studies, awareness campaigns are proportional to number of infected individuals reported by a health organization, which has been considered as a continuous process. Here, we assume that the awareness campaign is a discontinuous, periodic process. Mathematically, this awareness implementation feature can be appropriately modeled through impulsive differential equations, as it has the ability to capture periodic events [10–12].

2 Mathematical Model Derivation

Let $S(t)$ and $I(t)$ be the density of the susceptible and infected populations respectively at time t. The cumulative density of awareness programs driven by media in that region is $M(t)$. The total susceptible population is divided into two subclasses: the unaware susceptible population S_u and the aware susceptible population S_a. It is also assumed that infected individuals recover through treatment. After recovery, a fraction p of recovered people will join the aware susceptible class, whereas the remaining fraction q (p + q = 1) will join unaware susceptible class.

People immigrate at a rate π into the susceptible unaware population. Unaware susceptible individuals become infected after interaction with infected people at a rate β. Let d be the natural mortality rate of the population, e is the additional mortality rate of infected population due to disease and r is the recovery rate.

Here, it is assumed that the unaware susceptible population becomes aware at the rate α. Due to the effects of awareness programs that is repeatedly performed at a constant time instants, cumulative density of awareness campaign increases by a fixed amount ω. Awareness programs cut down at a rate θ, due to ineffectiveness.

The above assumptions lead to the following model with impulsive awqreness campaign,

$$\begin{aligned}
\frac{dS_u}{dt} &= \pi - \beta S_u I - \alpha S_u M - dS_u + qrI \\
\frac{dS_a}{dt} &= \alpha S_u M - dS_a + prI \\
\frac{dI}{dt} &= \beta S_u I - (d + e + r)I
\end{aligned}$$

$$\begin{aligned} \frac{dM}{dt} &= -\theta M, && t \neq t_k, \\ M(t^+) &= M(t) + \omega, && t = t_k \end{aligned} \tag{1}$$

with the initial conditions: $S_u(0) > 0$, $S_a(0) > 0$, $I(0) > 0$, $M(0) > 0$.

3 Dynamics of the System

Consider the following subsystem,

$$\begin{aligned} \frac{dM}{dt} &= -\theta M, && t \neq t_k \\ \Delta M &= \omega, && t = t_k \end{aligned} \tag{2}$$

Let $\tau = t_{k+1} - t_k$ be the period of the campaign. The solution of the system (2) is,

$$M(t) = M(t_k^+)e^{-\theta(t-t_k)}, \quad \text{for } t_k < t \leq t_{k+1}. \tag{3}$$

In presence of impulsive effect, we have a recursion relation at the moments of impulse, given by

$$M(t_k^+) = M(t_k^-) + \omega.$$

Thus the awareness campaign before and after the impulse is taken is,

$$M(t_k^+) = \frac{\omega(1 - e^{-k\theta\tau})}{1 - e^{-\theta\tau}}$$

and

$$M(t_{k+1}^-) = \frac{\omega(1 - e^{-k\theta\tau})e^{-\theta\tau}}{1 - e^{-\theta\tau}}.$$

Thus the limiting case of the awareness campaign before and after one cycle is as follows:

$$\lim_{k\to\infty} M(t_k^+) = \frac{\omega}{1 - e^{-\theta\tau}} \quad \text{and} \quad \lim_{k\to\infty} M(t_{k+1}^-) = \frac{\omega e^{-\theta\tau}}{1 - e^{-\theta\tau}}$$

and

$$M(t_{k+1}^+) = \frac{\omega e^{-\theta\tau}}{1 - e^{-\theta\tau}} + \omega = \frac{\omega}{1 - e^{-\theta\tau}}.$$

Here we have found out that

$$M(t_k^+) - \frac{\omega}{1 - e^{-\theta\tau}} = \omega\frac{1 - e^{-k\theta\tau}}{1 - e^{-\theta\tau}} - \frac{\omega}{1 - e^{-\theta\tau}} = -\omega\frac{e^{-k\theta\tau}}{1 - e^{-\theta\tau}} < 0.$$

There does not exist an equilibrium point for the impulsive system. Hence, we can evaluate equilibrium-like periodic orbits. There are two periodic orbits: disease-free periodic orbit and endemic periodic orbit. We only focus on the disease free orbit. We now recall the following results [13, 14].

Definition 1 Let $\Lambda \equiv (S_u, S_a, I, M)$, $B_0 = [B : R_+^4 \rightarrow R_+]$. Then B is said to belong to class B_0 if

(1) B is continuous on $(t_k, t_{k+1}] \times R_+^3$, $n \in N$ and for each $\Lambda \in R^4$, $\lim_{(t,\mu)\rightarrow(t_k^+,\Lambda)} B(t, \mu) = B(t_k^+, \Lambda)$ exists,

(2) B is locally Lipschitzian in Λ.

Lemma 1 *Let $\mathbf{Z}(t)$ be a solution of the system (1) with $\mathbf{Z}(0^+) \geq 0$. Then $\mathbf{Z}_i(t) \geq 0$, $i = 1, \ldots, 4$ for all $t \geq 0$. Moreover, $\mathbf{Z}_i(t) > 0$, $i = 1, \ldots, 4$ for all $t > 0$ if $\mathbf{Z}_i(0^+) > 0$, $i = 1, \ldots, 4$.*

Lemma 2 *There exists a constant G such that $S_u(t) \leq G$, $S_a(t) \leq G$, $I(t) \leq G$ and $M(t) \leq G$ for each and every solution $\mathbf{Z}(t)$ of system (1) for all sufficiently large t.*

Lemma 3 *Let us consider $B \in B_0$ and assume that*

$$\begin{aligned} D^+B(t, Z) &\leq j(t, B(t, \mathbf{Z}(t))), \quad t \neq t_k, \\ B(t, Z(t^+)) &\leq \Phi_n(B(t, \mathbf{Z}(t))), \quad t = t_k, \end{aligned}$$

where $j : \mathbf{R}_+ \times \mathbf{R}_+ \rightarrow \mathbf{R}$ is continuous in $(t_k, t_{k+1}]$ for $e \in \mathbf{R}_+^2$, $n \in N$, the limit $\lim_{(t,V)\rightarrow(t_k^+)} j(t, g) = j(t_k^+, x)$ exists and $\Phi_n^i (i = 1, 2) : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ is non-decreasing. Let $y(t)$ be a maximal solution of the following scalar impulsive differential equation

$$\begin{aligned} \frac{dx(t)}{dt} &= j(t, x(t)), \quad t \neq t_k, \\ x(t^+) &= \Phi_n(x(t)), \quad t = t_k, \quad x(0^+) = x_0, \end{aligned} \tag{4}$$

existing on $(0^+, \infty)$. Then $B(0^+, \mathbf{Z}_0) \leq x_0$ implies that $B(t, \mathbf{Z}(t)) \leq y(t)$, $t \geq 0$, for any solution $\mathbf{Z}(t)$ of system (1). If j satisfies additional smoothness conditions to ensure the existence and uniqueness of solutions for (4), then $y(t)$ is the unique solution of (4).

We now consider the following sub-system:

$$\frac{dM(t)}{dt} = -\theta M, \quad t \neq t_k, \quad M(t_k^+) = M(t_k) + \omega, \quad M(0^+) = M_0. \tag{5}$$

Thus from above Lemma, we obtain the following result,

Lemma 4 *System (5) has a unique positive periodic solution $\tilde{M}(t)$ with period τ and given by*

$$\tilde{M}(t) = \frac{\omega\, exp(-\theta(t-t_k))}{1-exp(-\theta\tau)}, \quad t_k \le t \le t_{k+1}, \quad \tilde{M}(0^+) = \frac{\theta}{1-exp(-\theta\tau)}.$$

On this basis, we have the following theorem.

Theorem 1 *The disease free periodic solution $(\tilde{S_u}, 0, 0, \tilde{M})$ of the system (1) is locally asymptotically stable if*

$$\tilde{R_0} < 1 \tag{6}$$

where,

$$\tilde{R_0} = \frac{\beta}{\tau(d+e+r)} \int_0^{\tau} \tilde{S_u} dt.$$

Proof Let the solution of the system (1) without infected people be denoted by $(\tilde{S_u}, 0, 0, \tilde{M})$, where

$$\tilde{M}(t) = \frac{\omega\ \exp(-\theta(t-t_k))}{1-\exp(-\theta\tau)}, \quad t_k \le t \le t_{k+1},$$

with initial condition $M(0^+)$ as in Lemma 4. We now test the stability of the equilibria. The variational matrix at $(\tilde{S_u}, 0, 0, \tilde{M})$ is given by

$$V(t) = \begin{pmatrix} -\alpha\tilde{M} - d & 0 & qr - \beta\tilde{S_u} & -\alpha\tilde{S_u} \\ \alpha\tilde{M} & -d & pr & -\alpha\tilde{S_u} \\ 0 & 0 & \beta\tilde{S_u} - (d+e+r) & 0 \\ 0 & 0 & 0 & -\theta \end{pmatrix}.$$

The monodromy matrix $\mathbb{M}$ of the variational matrix $V(t)$ is

$$\mathbb{M}(T) = I \exp\left(\int_0^{\tau} V(t)dt\right),$$

where I is the identity matrix. We thus have: $\mathbb{M}(T) = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Here, $\lambda_i, i = 1, 2, 3, 4$, are the Floquet multipliers and given by

$$\lambda_1 = \exp\left[-\alpha \int_0^\tau \tilde{M} dt - d\tau\right] < 1, \quad \lambda_2 = \exp\left[-d\tau\right] < 1,$$
$$\lambda_3 = \exp\left[\beta \int_0^\tau \tilde{S}_u dt - (d + e + r)\tau\right], \quad \lambda_4 = \exp(-\theta\tau) < 1.$$

Now, $\lambda_3 < 1$ holds if (6) is true. Thus, the periodic solution $(\tilde{S}_u(t), 0, 0, \tilde{M}(t))$ of the system (1) is locally asymptotically stable if the conditions given in (6) hold.

4 Numerical Simulation

In this section, numerical simulations are performed to investigate the dynamics of the system obtained in previous sections. Simulations results are plotted in figures with the parameters values given in Table 1.

Figure 1 displays the comparison between the behavior of the system whenever awareness campaign is carried out at an interval of 10 days and in continuous way. Here, the aware humans are initially increasing and attaining a stable state faster than in the continuous awareness program, after approximately 900 days. The infected human population becomes extinct within less than 500 days.

Time series solution of the system with impulses is plotted in Fig. 2 with three different choices of time interval τ. On a close inspection, we observe that the infection can be controlled much faster in 10 days interval compare to the result obtained for 15 and 30 days intervals. It is to be noted that the aware population drops to the half strength in comparison to the observed strength with a 10-days time interval among three consecutive campaigns. Also, we observed that the time interval of 10 days is more beneficial for overall performance of the system.

Table 1 List of parameters used in numerical simulations and their references [6]

Parameter	Definition	Value	Unit
π	Constant recruitment rate	30	Person day^{-1}
β	Disease transmission rate	0.00025	Day^{-1}
α	Maximum rate of awareness from unaware to aware human	0.02	Day^{-1}
d	Susceptible class natural death rate	0.005	Day^{-1}
e	Additional death rate due to infection	0.02	Day^{-1}
θ	Depletion rate of awareness program due to ineffectiveness	0.05	Day^{-1}

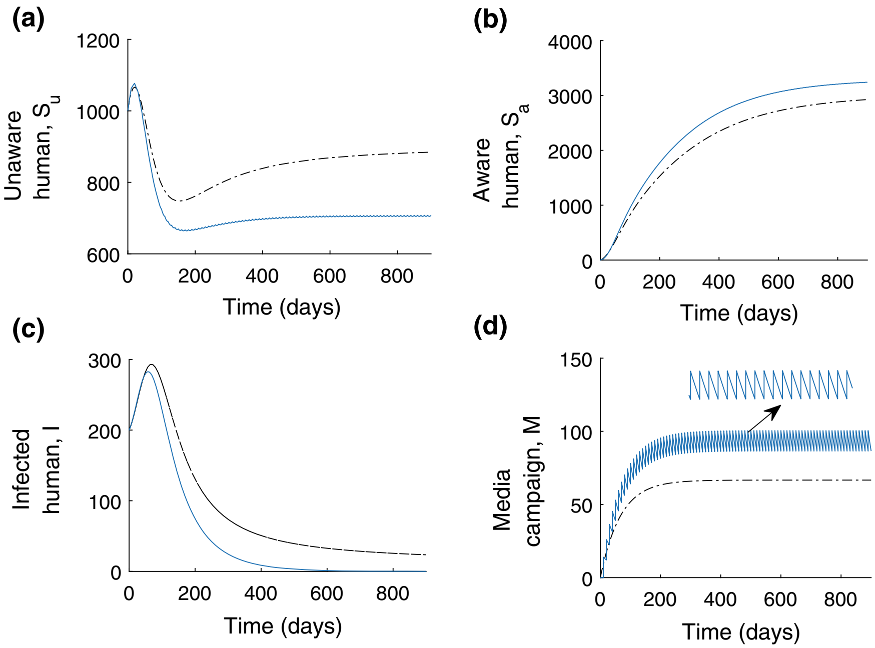


Fig. 1 Solution of the system for two cases, Case I: Continuous campaign (dashed line), Case II: with impulsive way taking $\tau = 10$ days (solid line). Parameters are taken from Table 1

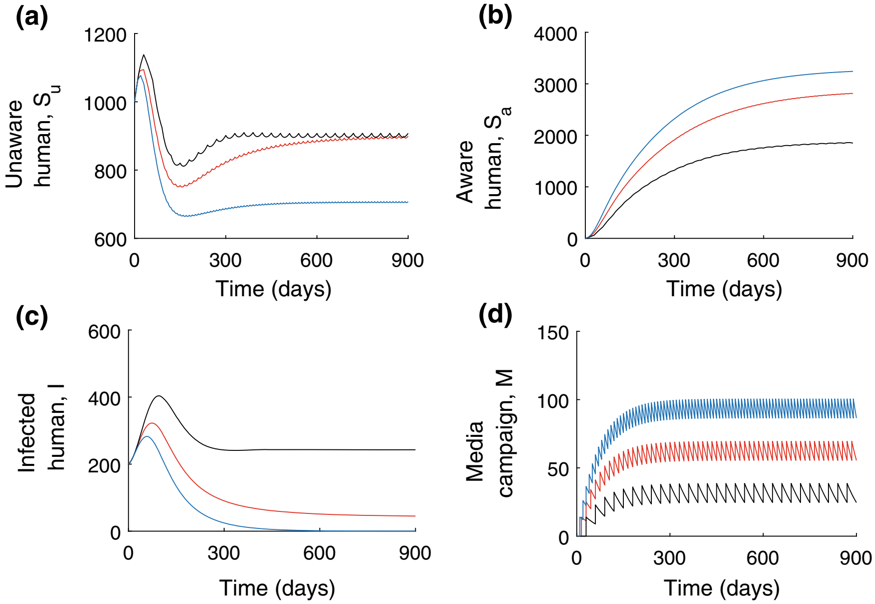


Fig. 2 System populations are plotted for different time interval τ. Blue line represents $\tau = 10$, red line is for $\tau = 15$, black line for $\tau = 30$ days, $\omega = 1$

5 Discission and Conclusion

In this article, we have proposed an epidemic models to estimate and analyze the effects of an impulsive awareness campaign on infectious disease control. We give the threshold $\tilde{R_0}$ of the infected human free impulsive system (1). When $\tilde{R_0} < 1$, periodic equilibrium $\tilde{E^*}$ is locally asymptotically stable.

We can control the rate of impulsive campaigning and the period of impulse. The results obtained here imply that we can control the period of impulsive campaign such that the infection-free periodic solution is locally stable. This means the infected people could be eradicated in cost effective way if we implement the impulsive control strategies with proper rate and interval.

Acknowledgements Author is grateful to the reviewers for their useful comments which enrich the content of the paper.

References

1. WHO: The world health report 2004: Changing history. http://www.who.int/whr/2004/en/ (2004)
2. Brodie, M., Hamel, E., Brady, L.A., et al.: AIDS at 21: Media coverage of the HIV epidemic 1981–2002. Columbia J. Rev. **42**(6), A1
3. Cui, J., Sun, Y., Zhu, H.: The impact of media on the control of infectious diseases. J. Dyn. Differ. Equ. **20**, 31–53 (2007)
4. Liu, Y., Cui, J.A., Zhu, H.: The impact of media coverage on the dynamics of infectious disease. Int. J. Biomath. **1**(1), 65–74 (2008)
5. Smith, R.J., Cloutier, P., Harrison, J., Desforges, A.: A mathematical model for the eradication of Guinea Worm Disease. In: Mushayabasa, S., Bhunu, C.P. (eds) Understanding the dynamics of emerging and re-emerging infectious diseases using mathematical models, pp. 133–156. (2012)
6. Al Basir, F., Ray, S., Venturino, E.: Role of media coverage and delay in controlling infectious diseases: a mathematical model. Appl. Math. Comput. **337**, 372–385 (2018)
7. Dupas, P.: Health behavior in developing countries. Annu. Rev. Econ. **3**(1), 425–449 (2011)
8. Agaba, G.O., Kyrychko, Y.N., Blyuss, K.B.: Time-delayed SIS epidemic model with population awareness. Ecol. Complex. **31**, 50–56 (2017)
9. Roy, P.K., Saha, S., Basir, A.F.: Effect of awareness programs in controlling the disease HIV/AIDS: an optimal control theoretic approach. Adv. Differ. Equ. **2015**(1), 217 (2015)
10. Basir, F.A., Venturino, E., Roy, P.K.: Effects of awareness program for controlling mosaic disease in Jatropha curcas plantations. Math. Methods Appl. Sci. **40**(7), 2441–2453 (2017)
11. Li, X., Bohner, M., Wang, C.K.: Impulsive differential equations: periodic solutions and applications. Automatica **52**, 173–8 (2015)
12. Liu, X.N., Chen, L.S.: Complex dynamics of Holling type II Lotka-Volterra predator-prey system with impulsive perturbations on the predator. Chaos Solitons Fractals **16**(2), 311–320 (2003)
13. Yu, H., Zhong, S., Agarwal, R.P.: Mathematics analysis and chaos in an ecological model with an impulsive control strategy. Commun. Nonlinear Sci. Numer. Simul. **16**(2), 776–786 (2011)
14. Lakshmikantham, V., Bainov, D.D., Simeonovm, P.S.: Theory of impulsive differential equations, series in modern applied mathematics. World Scientific Publishing, pp. 6 (1989)

A Vendor-Buyer Supply Chain Model for Deteriorating Item with Quadratic Time-Varying Demand and Pro-rata Warranty Policy

B. Samanta, B. C. Giri and K. S. Chaudhuri

Abstract The article considers a single-vendor a single-buyer supply chain in which the buyer sells a seasonal deteriorating product to customers over a single period of time. At the buyer's end, the demand is assumed to be quadratic in time. The vendor follows a lot-for-lot policy for replenishment made to the buyer and (s)he agrees to refund the customer's purchase price on a pro-rata basis, if any item fails during the warranty period offered by the buyer. We derive and optimize the average expected total cost of the supply chain in order to obtain the optimal decisions of the integrated system. An algorithm is developed for finding the optimal solution of the model numerically. A numerical example is taken to demonstrate the coordination policy between the vendor and the buyer. Sensitivity analysis is carried out to investigate the effects of key parameters on the optimal decisions.

Keywords Supply chain · Quadratic demand · Deterioration · Warranty

1 Introduction

Due to rapid technological progress, competition in various types of business gradually increases. Company managers are facing tough challenges from almost all circumstances in the real market.To overcome these challenges, they learn that the adjustment and cooperation among all the members of supply chain is more beneficial than those policies obtained separately from the buyer's and vendor's perspectives.

The concept of coordination in a supply chain was first introduced by Goyal [1] by considering a single-vendor single-buyer supply chain scenario. Later on, numerous works have been carried out to establish vendor-buyer coordination under different circumstances. It was realized that the supply chain coordination practices could reduce the total relevant cost or increase the total profit over the chosen planning horizon [2–6]. Glock [7] considered a buyer sourcing a product from het-

B. Samanta · B. C. Giri (✉) · K. S. Chaudhuri
Department of Mathematics, Jadavpur University, Kolkata 700032, India
e-mail: bcgiri.jumath@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_31

erogeneous suppliers and tackled both the supplier selection and lot size decision with the objective of minimising the total system cost. Li and Huang [8] considered a single-vendor and multiple buyers supply chain model and realised that under vendor-managed inventory, the production decision does not involve all operational information of the vendor and buyers. Jauhari [9] considered an integrated vendor-buyer model with defective items, inspection error and stochastic demand. Hariga et al. [10] studied a two-stage supply chain model under consignment stock policy. Hossain et al. [11] investigated vendor-buyer cooperative policy under generalized lead time distribution with penalty cost for delivery lateness.

In all the above-mentioned works, non-perishability characteristic of buyer's inventory is considered. However, in practical situation, perishability or deterioration is a natural phenomenon for many items. Agricultural products, food items, chemicals, drugs, fashion goods, electronic components, films are some examples of items in which sufficient deterioration may occur during normal storage periods of the units. The loss due to deterioration must be taken into account while analyzing the inventory system. There is a vast literature on deteriorating inventory system. Readers can go through the latest review article contributed by Goyal and Giri [12] to learn the summary of the works done on deteriorating inventory system until that time.

Supply chain researchers considered buyer's demand rate in different ways. Several models were developed by considering constant demand rate. However, in practice, the buyer's demand rate may not always remain constant; it may vary with time. Giri and Maiti [6] developed a supply chain model with linearly time-dependent demand. A linearly time varying demand means uniform change in demand rate which is rarely seen to occur in real market. On the other hand, the exponential pattern also seems to be unrealistic in the sense that there is hardly any product whose demand changes at an exponential rate. In this paper, we consider a quadratic time dependent demand for a vendor-buyer supply chain model over a short selling period. This type of demand is more realistic because it can represent both accelerated and retarded growth in demand, and it is quite appropriate for seasonal products like winter garments or cosmetics.

The analysis of warranty policy and cost model associated with short-term or fixed term policy have received a lot of attention among the researchers. Failure of a product may occur due to faulty design, bad workmanship, age, usage or the increase of operational and environmental stresses above the designed level. It is impossible to totally avoid all failures. The manufacturer (vendor)/service provider can prevent or minimize the effect of such failure by ensuing after sales service through warranty and service contract. Chun and Tang [13] developed an inventory model with the free-replacement and fixed period warranty policy within a constant warranty period. Yeh et al. [14] studied the imperfect production system where a repairable product is sold with free minimal repair warranty. Chen and Lo [15] developed an imperfect production system with allowable shortages for products extending Yeh et al.'s [14] model. All these models considered constant warranty cost. In this paper, pro-rata warranty policy is adopted in which warranty cost is taken as a function of warranty period and production cost of the product.

We organize the paper as follows: Assumptions and notations are given in the next section. Formulations and analysis of the model under perspectives of the buyer, the vendor and the entire supply chain are given in Sect. 3. An algorithm for finding the optimal solution of the supply chain is provided. Numerical illustration and sensitivity analysis are performed in Sect. 4. The paper is concluded with some remarks and practical implications in Sect. 5.

2 Assumptions and Notations

The following assumptions are made for developing the proposed model.

- The supply chain consists of a single-vendor and a single-buyer.
- Over a short selling period, the buyer sells a seasonal product which is deteriorating in nature.
- The demand rate at the buyer's end is quadratic in time.
- A constant fraction of the buyer's on-hand inventory deteriorates per unit time.
- The buyer receives order from the vendor in every T time interval.
- The production rate of the vendor varies with time and is greater than the demand rate of the buyer.
- The vendor follows the lot-for-lot policy for replenishment made to the buyer.
- The buyer sells each item under Pro-Rata Warranty (PRW) policy. Under this policy, the vendor agrees to refund a fraction of customer's purchase price, if any item fails during warranty period offered by the buyer.
- The warranty does not renew.
- Shortages are not allowed to occur in the buyer's inventory.
- As the vendor's production rate is greater than the buyer's demand rate, the vendor may start production with a time delay in every production cycle.

The following notations are used for developing the proposed model.

$d(t)$ buyer's demand rate; we take $d(t) = a + bt + ct^2$; $a \geq 0,\ b > 0,\ c > 0$.
$p(t)$ vendor's production rate; we take $p(t) = kd(T + t),\ k > 1$.
θ deterioration rate at the buyer; $0 \leq \theta < 1$.
t_m time delay for the vendor to begin production.
$a_r(a_m)$ order (set up) cost per order (set up) for the buyer (vendor).
$c_r(c_m)$ cost per single item for the buyer (vendor).
$h_r(h_m)$ holding cost per unit item per unit time for the buyer (vendor).
$I_{ri}(t)$ buyer's inventory level at any time t where $iT \leq t \leq (i + 1)t$.
$I_{mi}(t)$ vendor's inventory level at any time t where $iT \leq t \leq (i + 1)t$.
α defective item production rate, $0 \leq \alpha < 1$.
w_p warranty period from initial purchase given by the vendor to the buyer.
$f(\cdot)$ probability density function associated with time when machine shifts from an in-control state to an out-of-control state.

$E(N)$ expected number of defective items produced in a cycle.

λ failure intensity of a product.

$g(t)$ failure density function.

$G(t)$ cumulative failure distribution of a product associated with $g(t)$.

$R(t)$ failure rate per item at any time t.

$r(x)$ refund cost of a failure item failed at any time $x \in (t, t + w_p)$ from initial purchase.
We take $r(x) = c_r(1 - \frac{x-t}{w_p}), \;\; t \leq x \leq (t + w_p)$.

$w(t)$ warranty cost at time $t \in [\ iT,\ \ (i+1)T\ ]$

w_m total warranty cost in $[\ iT,\ \ (i+1)T\ ]$.

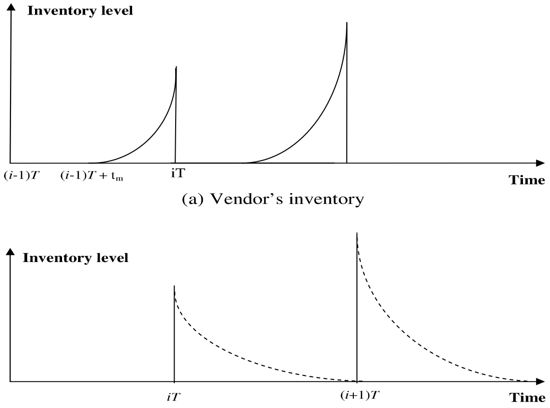
3 Model Development and Analysis

We consider a vendor-buyer integrated system where the buyer receives the first lot from the vendor at time T and sells these items with a warranty during next T time according to customer's demand. If any item fails within warranty period from customer's initial purchase time, the vendor agrees to refund a fraction of customer's purchase price. At time $2T$, the buyer's inventory level becomes zero and s(he) receives the next lot from the vendor. This process will continue. Since the vendor's production rate is greater than the buyer's demand rate, therefore, to produce the buyer's order quantity, the vendor may start production after some delay time (t_m) in every production cycle. Then the following questions arise:
(i) How much time can be delayed by the vendor to start production in each cycle?
(ii) What would be the cycle length for which cost is minimum?
To answer these questions, we focus our attention to the i-th inventory cycle of the buyer, that is, the time interval $[iT, (i+1)T];\ \ i = 1, 2, 3, \ldots$ The quantity replenished at the beginning of this cycle is the total conforming items produced by the vendor in the time interval $[(i-1)T, iT];\ \ i = 1, 2, 3, \ldots$ Since shortages are not allowed to occur, the expected number of conforming items produced by the vendor during the time interval $[(i-1)T + t_m, iT]$ is equal to the buyer's order quantity, which is again equal to the total demand and deterioration during the time interval $[iT, (i+1)T];\ \ i = 1, 2, 3, \ldots$ at the buyer's end. The fluctuations of the buyer's and the vendor's inventories are reflected in Fig. 1.

3.1 Buyer's Perspective

The buyer's inventory level changes due to the combined effect of demand and deterioration. Therefore, if $I_{ri}(t)$ denotes the buyer's inventory level at any time $t,\ iT \leq t \leq (i+1)T;\ i = 1, 2, 3, \ldots$ then the differential equation governing the instantaneous states of the inventory level at time t is given by

Fig. 1 Schematic diagram of vendor-buyer inventory



(a) Vendor's inventory

(b) Buyer's inventory

$$\frac{dI_{ri}(t)}{dt} = -\theta I_{ri}(t) - d(t); \quad iT \le t \le (i+1)T \tag{1}$$

with $I_{ri}((i+1)T) = 0$.
Solving (1), we get

$$I_{ri}(t) = Ae^{-\theta t} - \frac{1}{\theta^3}[(a+bt+ct^2)\theta^2 - (b+2ct)\theta + 2c] \tag{2}$$

$$\text{where } A = \frac{e^{(i+1)T}}{\theta^3}[c(i+1)^2(\theta T)^2 + (b\theta - 2c)(i+1)(\theta T) + a\theta^2 - b\theta + 2c]$$

Now, the sum of inventory holding cost and deteriorating cost during the time interval $[iT, (i+1)T]$ is given by

$$\begin{aligned}&(h_r + \theta c_r)\int_{iT}^{(i+1)T} I_{ri}(t)dt\\ &= \frac{(h_r+\theta c_r)}{\theta^3}\Big[\frac{e^{\theta T}-1}{\theta}\{c(i+1)^2(\theta T)^2 + (b\theta-2c)(i+1)\theta T + a\theta^2 - b\theta + 2c\}\\ &\quad - \frac{T}{6}\{2c(3i^2+3i+1)(\theta T)^2 + 3(b\theta-2c)(2i+1)\theta T + 6(a\theta^2 - b\theta + 2c)\}\Big].\end{aligned}$$

Hence, the buyer's average total cost for the i-th cycle is given by

$$\begin{aligned}Ac_{ri}(T) = \frac{a_r}{T} + \frac{(h_r+\theta c_r)}{\theta^3}\Big[&\frac{e^{\theta T}-1}{\theta T}\{c(i+1)^2(\theta T)^2 + (b\theta-2c)(i+1)\theta T\\ &+a\theta^2 - b\theta + 2c\} - \frac{1}{6}\{2c(3i^2+3i+1)(\theta T)^2 + 3(b\theta-2c)(2i+1)\theta T\\ &+6(a\theta^2 - b\theta + 2c)\}\Big]; \quad i = 1, 2, 3, ...\end{aligned} \tag{3}$$

Lemma 1 *The function $f(x) = e^x(x^2 - 2x + 2) - 2$ is positive and strictly increasing $\forall\, x > 0$.*

Lemma 2 *The function $f(x) = (x + 2)e^x - \frac{2}{3}$ is positive and strictly increasing for all $x > 0$. Also, $f(x)$ cannot be less than $\frac{4}{3}$ for all $x > 0$.*

Proof Clearly $f(x)$ is continuous for all $x \geq 0$. Also, $f'(x) = (x + 3)e^x > 0$ for all $x > 0$. Therefore, $f(x)$ is a strictly increasing function on $[0, \infty)$. Again, $f(0) = \frac{4}{3} > 0$. This proves the result.

Lemma 3 $f(n) = \frac{2}{3}\{2(n + 1)^2 - n^2\} > 0,\ \forall\ n \in N.$

Lemma 4 $f(x, n) = (n + 1)^2\{(x + 2)e^x - \frac{2}{3}\} - \frac{2}{3}n^2$ *is positive and strictly increasing for all $x > 0$ and $n \in N$.*

Proposition 1 *The average cost function $Ac_{ri}(T)$ is convex provided that $\frac{b}{4a} \leq \frac{2c}{b} \leq \theta < 1$, if $a > 0$ and $\frac{2c}{b} = \theta < 1$, if $a = 0$.*

Proof Differentiating (3) twice with respect to T, we obtain

$$\frac{d^2Ac_{ri}(T)}{dT^2} = \frac{2a_r}{T^3} + \frac{h_r + c_r\theta}{\theta}\Big[c(i + 1)^2\{(\theta T + 2)e^{\theta T} - \frac{2}{3}\} - \frac{2}{3}ci^2 + (b\theta - 2c)(i + 1)e^{\theta T} + \frac{a\theta^2 - b\theta + 2c}{(\theta T)^3}\{e^{\theta T}((\theta T)^2 - 2\theta T + 2) - 2\}\Big]$$

Now, $a\theta^2 - b\theta + 2c = a\{\theta^2 - 2\theta\frac{b}{2a} + (\frac{b}{2a})^2\} + 2c - \frac{b^2}{4a}$.
Case (i): $a > 0$
Clearly, $a\theta^2 - b\theta + 2c \geq 0$ if $2c - \frac{b^2}{4a} \geq 0$, i.e. if $\frac{b}{4a} \leq \frac{2c}{b}$.
Also, $b\theta - 2c \geq 0$ if $\theta \geq \frac{2c}{b}$. Thus, $\frac{b}{4a} \leq \frac{2c}{b} \leq \theta < 1$.
Case (ii): $a = 0$
The relations $b\theta - 2c \geq 0$ and $a\theta^2 - b\theta + 2c \geq 0$ give $b\theta - 2c \geq 0$ and $b\theta - 2c \leq 0$, respectively which then imply $\theta = \frac{2c}{b}$. Using the above two cases, Lemmas 1 and 4, we see that $\frac{d^2Ac_{ri}(T)}{dT^2} > 0$ if $\frac{2c}{b} \leq \theta < 1$ and $ac > b^2$ when $a > 0$. Hence, the proposition follows.

3.2 Vendor's Perspective

We assume that the time to machine shift from a in-control state to an out-control state follows a uniform distribution with probability density function

$$\begin{aligned} f(t) &= \frac{1}{T - t_m}; \qquad t_m \leq \ t \leq \ T \\ &= 0; \qquad elsewhere. \end{aligned}$$

Then the expected number of defective items produced in the time period $[(i-1)T + t_m, iT]$ is given by

$$\begin{aligned} E(N) &= \int_{(i-1)T+t_m}^{iT} \alpha p(t) f(t)(iT - t_m)dt \\ &= \frac{\alpha k}{12}\Big[6aT + 2b(3i+1)T^2 + c(6i^2+4i+1)T^3 - ((6i^2+4i+1)cT^2 \\ &\quad +2(3i-1)bT + 6a)t_m - (4b + (8i-1)cT)t_m^2 - 3ct_m^3\Big] \end{aligned} \tag{4}$$

Differentiating partially with respect to t_m, we obtain

$$\begin{aligned} \frac{\partial E(N)}{\partial t_m} &= -\frac{\alpha k}{12}\Big[(6i^2 - 4i - 1)cT^2 + 2(3i-1)bT + 6a + 2((8i-1)cT + 4b)t_m \\ &\quad +9ct_m^2\Big] \end{aligned} \tag{5}$$

Proposition 2 *For a prescribed scheduling period T, the expected number of defective items $(E(N))$ produced by the vendor is decreasing in the delay time t_m.*

Proof The result can be easily proved from (5) as $6i^2 - 4i - 1 > 0$ for i=1, 2, 3, ...

The instantaneous states of the vendor's inventory level can be described by the following differential equation:

$$\frac{dI_{mi}}{dt} = p(t) - \frac{E(N)}{T - t_m}, \quad (i-1)T + t_m \leq t \leq iT \tag{6}$$

with $I_{mi}((i-1)T + t_m) = 0$.
Solving (6), we get

$$\begin{aligned} I_{mi}(t) &= (ka - \frac{E(N)}{T - t_m})\{(T+t) - (iT + t_m)\} + \frac{kb}{2}\{(T+t)^2 - (iT + t_m)^2\} \\ &\quad + \frac{kc}{3}\{(T+t)^3 - (iT + t_m)^3\} \end{aligned} \tag{7}$$

If Q_{mi} denotes the production lot size of the vendor, then we have

$$Q_{mi} = I_{mi}(iT) = k(T - t_m)\Big[a + \frac{b}{2}((2i+1)T + t_m) + \frac{c}{3}((3i^2 + 3i + 1)T^2 + (3i+1)Tt_m + t_m^2)\Big] - E(N) \tag{8}$$

The vendor's holding cost for the period $[(i-1)T, iT]$ is given by

$$h_m \int_{(i-1)T+t_m}^{iT} I_{mi}(t)dt = h_m(2-\alpha)(T - t_m)E(N)/(2\alpha) \tag{9}$$

Suppose that the vendor sells an item with warranty period w_p at any time $t \in [iT, (i+1)T]$ in the i-th cycle. Let the product defect be found at time $x \in (t, t + w_p)$ from initial purchase. Then, according to PRW policy, the buyer has to refund money for remaining $(t + w_p - x)$ period for the product. Since c_r is the cost of the product, the refund from the vendor should be $r(x,t) = c_r \frac{(t+w_p-x)}{w_p} = c_r(1 - \frac{x-t}{w_p})$. Since the vendor rejects the defective items, all the items supplied to the buyer are good in condition. Still a few of those items may fail at the customer end. We assume that the failure of the product follows an exponential distribution with probability density function $g(x) = \lambda e^{-\lambda x}$, where, λ is the failure intensity of a product and $x \in (t, t + w_p)$. Then, the corresponding cumulative failure distribution function $G(x)$ is given by $G(x) = 1 - e^{-\lambda x}$, where $x \in (t, t + w_p)$. If $R(x)$ be the failure rate function of any product in time $x \in (t, t + w_p)$, then we can write $R(x) = \frac{g(x)}{1-G(x)} = \lambda$.
Then the warranty cost at any time $t \in [iT, (i+1)T]$ is given by

$$w_c(t) = \int_{t-iT}^{t+w_p-iT} r(x,t)d(t)R(x)dx = \lambda c_r(w_p/2 + iT)(a + bt + ct^2)$$

Hence, the total warranty cost during the period $[iT, (i+1)T]$ is

$$w_{TC} = \int_{iT}^{(i+1)T} w_c(t)dt = \lambda T c_r(\frac{w_p}{2} + iT)[a + b(i + \frac{1}{2})T + c(i^2 + i + \frac{1}{3})T^2]$$

So, the average expected total warranty cost in $[iT, (i+1)T]$ is

$$w_{ATC} = \frac{w_{TC}}{T} = \lambda c_r(\frac{w_p}{2} + iT)[a + b(i + \frac{1}{2})T + c(i^2 + i + \frac{1}{3})T^2] \tag{10}$$

As the vendor's total cost includes ordering cost, holding cost, defective item cost and warranty cost, therefore, the average expected total cost is given by

$$AC_{mi}(T, t_m) = \frac{a_m + c_m E(N)}{T} + \frac{h_m}{T}(T - t_m)E(N)(\frac{1}{\alpha} - \frac{1}{2}) + \lambda c_r(\frac{w_p}{2} + iT) \times [a + b(i + \frac{1}{2})T + c(i^2 + i + \frac{1}{3})T^2] \tag{11}$$

3.3 Integrated Approach

The average expected total cost of the integrated supply chain is $AC = AC_{ri} + AC_{mi}$. Our objective is to

$$\begin{aligned} \text{Minimize} \quad & AC \\ \text{subject to} \quad & Q_{ri} = Q_{mi} \end{aligned} \tag{12}$$

$$\begin{aligned} \text{where } Q_{ri} &= \int_{iT}^{(i+1)T} (d(t) + \theta I_{ri}(t))dt \\ &= T(a + b(i + \frac{1}{2})T + c(i^2 + i + \frac{1}{3})T^2) + \frac{1}{\theta^2}\Big[\frac{e^{\theta T - 1}}{\theta T}(a\theta^2 - b\theta + 2c \\ &\quad +(b\theta - 2c)(i+1)\theta T + c(i+1)^2(\theta T)^2) - (a\theta^2 - b\theta + 2c \\ &\quad +(b\theta - 2c)(i + \frac{1}{2})\theta T + c(i^2 + i + \frac{1}{3})(\theta T)^2)\Big] \end{aligned} \tag{13}$$

and Q_{mi} is given in (8). The constrained optimization problem (12) can be solved numerically following a simple algorithm given below:

Algorithm:

Step 1. Initialize $t_m = 0$ and $\varepsilon =$ a small positive number.
Step 2. Use any suitable one-dimensional search technique to find the optimal value T^* by minimizing $AC(T, t_m)$.
Step 3. Calculate $S = Q_{mi}(T^*, t_m) - Q_{ri}(T^*)$.
Step 4. If $|S| < \varepsilon$ then $t_m^* = t_m$ and go to Step 5. Otherwise, do increment or decrement of t_m according to $S > 0$ or $S < 0$, respectively and then go to Step 2.
Step 5. Calculate $AC(T^*, t_m^*)$. Stop.

4 Numerical Illustration

We consider the following numerical data to illustrate the developed model:
$a = 200, b = 2, c = 0.06, a_m = 80, h_m = 0.08, c_m = 8, a_r = 50, h_r = 0.2,$ $c_r = 10, \theta = 0.05, \alpha = 0.02, k = 1.4, \lambda = 0.4, w_p = 1$ in appropriate units. For this data set, the surface generated by the cost function $AC(T, t_m)$ for wide ranges of values of T and t_m is found convex. Hence, for the chosen data set, the proposed algorithm can be applied for finding the numerical solution. The optimal results obtained for the supply chain model are shown in Table 1.
We now consider the model from the buyer's and vendor's points of view.

Table 1 Optimal results of the supply chain model for successive ten cycles

ith cycle	T^*	t_m^*	$AC(T^*, t_m^*)$
1	0.38137	0.09943	1096.10
2	0.27561	0.07184	1356.46
3	0.22656	0.05905	1559.34
4	0.19680	0.05128	1731.56
5	0.17628	0.04593	1883.92
6	0.16105	0.04197	2022.07
7	0.14916	0.03887	2149.39
8	0.13954	0.03636	2268.11
9	0.13156	0.03428	2379.78
10	0.12479	0.03251	2485.54

Table 2 Optimal results from buyer's perspective and comparison with integrated model when $t_m = 0$

ith cycle	T^*	$AC_{ri}(T^*)$ (A)	$AC_{mi}(T^*)$ (B)	(A) + (B)	$AC(T^*)$
1	0.82183	120.00	1200.21	1320.21	224.11
2	0.81464	120.52	1870.54	1991.06	634.60
3	0.80735	121.06	2540.69	2661.75	1102.41
4	0.79999	121.61	3210.56	3332.17	1600.61
5	0.79260	122.17	3880.08	4002.25	2118.33
6	0.78521	122.74	4549.21	4671.95	2649.88
7	0.77784	123.32	5217.94	5341.26	3191.87
8	0.77052	123.91	5886.29	6010.20	3742.09
9	0.76325	124.50	6554.31	6678.81	4299.03
10	0.75607	125.10	7222.04	7347.140	4861.60

Let us first consider the model from the buyer's perspective. For fifth cycle ($i = 5$), the buyer's optimum cycle length is obtained as $T^* = 0.792604$ unit and the average total cost is $AC_{ri} = 122.172$ units. Substituting $T^* = 0.792604$ and $t_m = 0$ in (11), we get $AC_{mi} = 3880.08$ units. Hence the sum of the buyer's and vendor's average cost is $AC_{ri}^* + AC_{mi}^* = 4002.252$ units, which is costlier of 2118.332 units than the average total cost of the proposed supply chain model. Similar characteristics of the proposed model is observed for all values of $i = 1, 2, ..., 10$ as shown in Table 2.

We now consider the model from the vendor's perspective. For the fifth cycle ($i = 5$), the vendor's optimum cycle length is obtained as $T^* = 0.139701$ unit and the average expected total cost is $AC_{mi} = 1563.12$ units. Substituting $T^* = 0.139701$ in (5), we get $AC_{ri} = 367.789$ units. Then $AC_{ri}^* + AC_{mi}^* = 1930.909$ units,

Table 3 Optimal results from vendor's perspective and comparison with integrated model when $t_m = 0$

ith cycle	T^*	$AC_{mi}(T^*)$ (C)	$AC_{ri}(T^*)$ (D)	(C) + (D)	$AC(T^*)$
1	0.31134	935.062	182.62	1117.68	21.58
2	0.22085	1144.8	242.01	1386.81	30.35
3	0.18042	1306.33	289.89	1596.22	36.88
4	0.15624	1442.76	331.08	1773.84	42.28
5	0.13970	1563.12	367.79	1930.91	46.99
6	0.12748	1672.06	401.23	2073.29	51.22
7	0.11798	1772.32	432.16	2204.48	55.09
8	0.11031	1865.72	461.07	2326.79	58.68
9	0.10396	1953.51	488.32	2441.83	62.05
10	0.09858	2036.59	514.18	2550.77	65.23

which is costlier of 46.989 units than the average expected total cost of the proposed supply chain model. Similar characteristics of the proposed model is observed for other values of $i = 1, 2, ..., 10$ as shown in Table 3. From the above discussions we thus observe that the integrated supply chain model results in lower cost.
We further note that the buyer's optimum cycle length is always greater than the vendor's optimum cycle length in any cycle. Moreover, the average total cost of the integrated supply chain in any cycle is closer to the average total cost from vendor's perspective than that from buyer's perspective.

4.1 Sensitivity Analysis

In this sub-section, we examine the effects of changes in the parameter-values on the optimal solution of the proposed supply chain model. We change one parameter value at a time and keep other parameter-values interchanged. We conduct the analysis for the integrated model for $i = 3$. The results are shown in Table 4.

Table 4 indicates that the model is highly sensitive with respect to the parameters a, c_r and λ and moderately sensitive with respect to the parameters a_m, a_r and w_p. When the parameter a_m increases, the optimum cycle length of the integrated supply chain increases. If warranty period increases, then the number of defective items also increases. More items may fail during this extra warranty period. For these more failure items, the vendor has to return more warranty cost. That is why the average expected total cost of integrated system also increases. We also check that the model is insensitive with respect to the remaining parameters h_m, c_m, h_r, θ, c and k.

Table 4 Sensitivity analysis

Parameter	% change in parameter value	T^*	t_m^*	$AC(T^*, t_m^*)$	% change in $AC(T^*, t_m^*)$
a	+50	0.18586	0.04842	2020.48	29.57
	+20	0.20733	0.05402	1749.89	12.22
	−20	0.25230	0.06577	1357.78	−12.93
	-50	0.31486	0.08217	1024.66	−34.29
a_m	+50	0.25877	0.06744	1724.17	10.57
	+20	0.23998	0.06254	1627.93	4.40
	−20	0.21228	0.05532	1486.42	−4.68
	−50	0.18878	0.04919	1366.74	−12.35
a_r	+50	0.2472	0.06443	1664.88	6.77
	+20	0.23504	0.06126	1602.67	2.78
	−20	0.21775	0.05675	1514.33	−2.89
	−50	0.20379	0.05311	1443.16	−7.45
c_r	+50	0.18561	0.04837	2012.61	29.07
	+20	0.20716	0.05399	1746.84	12.02
	−20	0.25271	0.06587	1360.57	−12.75
	−50	0.31759	0.08279	1030.36	−33.92
λ	+50	0.18623	0.04853	2007.91	28.77
	+20	0.20751	0.05408	1744.75	11.89
	−20	0.25208	0.06570	1363.13	−12.58
	−50	0.31449	0.08198	1038.39	−33.41
w_p	+50	0.22624	0.05896	1760.97	12.93
	+20	0.22643	0.05901	1639.99	5.17
	−20	0.22669	0.05908	1478.69	−5.17
	−50	0.22689	0.05913	1357.72	−12.93

5 Conclusions

Reliability of products is becoming increasingly important due to rapid technological development and tough competition in the market. In the proposed model, we reject all the defective items that are produced. Manufacturing cost then increases but it provides customer/user higher peace of mind. It is impossible to totally avoid all failures of any product. Warranty protects the manufacturer against any damage/failure to the product due to misuse or abuse by the user during specified warranty period. Customers/buyers purchase the product and the decision to purchase is influenced by several factors such as quality brands and provisions for after sales service such as warranty. The numerical experiment shows that vendor-buyer coordination policy provides lower average expected total cost than that of the policy derived from the vendor's or the buyer's perspective only. The proposed model has been developed based on simplistic assumptions viz., quadratic demand at the buyer's end and PRW warranty policy offered by the vendor. The model, however, can be further extended to consider more practical situations such as multiple vendors, multiple buyers or items sold with FRW policy or combination of PRW and FRW policy.

References

1. Goyal, S.K.: An integrated inventory model for single supplier single-customer problem. Int. J. Product. Res. **15**, 107–111 (1976)
2. Viswanthan, S.: Optimal strategy for the integrated vendor-buyer inventory model. Eur. J. Operat. Res. **105**(1), 38–42 (1998)
3. Ben-Daya, M., Hariga, M.: Integrated single vendor single buyer model with stochastic demand and variable lead time. Int. J. Product. Econom. **92**, 75–80 (2004)
4. Hsiao, J., Lin, C.: A buyer-vendor EOQ model with changeable lead-time in supply chain. Int. J. Adv. Manuf. Technol. **26**, 917–921 (2005)
5. Sajadieh, M.S., Akbari-Jokar, M.R.: An integrated vendor buyer cooperative model under stochastic supply lead-time. Int. J. Adv. Manuf. Technol. **41**, 1043–1050 (2009)
6. Giri, B.C., Maiti, T.: Supply chain Model for a deteriorating product with time-varying demand and production rate. J. Oper. Res. Soc. **63**, 665–673 (2012)
7. Glock, C.H.: A multiple-vendor single-buyer integrated inventory model with a variable number of vendors. Comput. Ind. Eng. **60**, 173–182 (2011)
8. Li, Y., Huang, X.: A one-vendor multiple-buyer production-distribution system: The value of vendor managed inventory. INFOR **53**, 13–25 (2015)
9. Jauhari, W.A.: Integrated vendor-buyer model with defective items, inspection error and stochastic demand. Int. J. Math. Oper. Res. **8**(3), 342–359 (2016)
10. Hariga, M., Asad, R., Khan, Z.: Manufacturing-remanufacturing policies for a centralized two stage supply chain under consignment stock partnership. Int. J. Prod. Econ. **183**, 262–274 (2017)
11. Hossain, S.J.M., Ohaiba, M.M., Sarker, B.R.: An optimal vendor-buyer cooperative policy under generalized lead-time distribution with penalty cost for delivery lateness. Int. J. Prod. Econ. **188**, 50–62 (2017)
12. Goyal, S.K., Giri, B.C.: Recent trend in modeling of deteriorating inventory: an invited review. Eur. J. Oper. Res. **134**, 1–16 (2001)
13. Chun, Y.H., Tang, K.: Determining the optimal warranty price based on the producers and customers risk preferences. Eur. J. Oper. Res. **85**, 97–110 (1995)
14. Yeh, R.H., Ho, W.T., Tseng, S.T.: Optimal poduction run length for products sold with warranty. Eur. J. Oper. Res. **20**(3), 575–582 (2000)
15. Chen, C.K., Lo, C.C.: Optimal production run length for products sold with warranty in an imperfect production system with allowable shortages. Math. Comput. Model. **44**, 319–331 (2006)

Pest Control of Jatropha curcas Plant for Different Response Functions

Arunabha Sengupta, J. Chowdhury, Xianbing Cao and Priti Kumar Roy

Abstract Nowadays Jatropha curcas plants are being considered as a renewable energy feedstock for the production of biodiesel to overcome the crisis of natural fuel. The seeds of this plant contain oil which is one of the significant resources for alternative fuel production i.e. biodiesel. Though Jatropha curcas plant is proven to be toxic to many insects and animals, it is not pest and disease resistant. On this regard our research article presents formulation and analysis of a mathematical model for Jatropha plantation with a view to control its natural pests through application of biological pesticide. We assume linear and hyperbolic functional responses of predator for susceptible pest, where for infected pest the functional response is linear, as infected pests are weaker than susceptible pest and easy to catch. We study the dynamics of the system around each of the ecological feasible equilibrium. The reduction of disease eradication and predator-pest coexistence are observed around the predator free and disease free equilibrium respectively.

Keywords *Jatropha curcas* plant · Biodiesel · Pest control · Stability · Mathematical modeling · Functional response · Biological pesticide

1 Introduction

In current global scenario the demand of alternative energy sources are highly thriving to overcome the crisis of conventional energy sources. As a result production of renewable energy sources are truly required. Biodiesel is one of the most useful

A. Sengupta · J. Chowdhury · P. K. Roy (✉)
Centre for Mathematical Biology and Ecology, Department of Mathematics,
Jadavpur University, Kolkata 700032, India
e-mail: pritiju@gmail.com

X. Cao
College of Science, Beijing Technology and Business University,
Beijing, China
e-mail: xbcao3613@sina.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_32

alternative fuel which is renewable, clean-burning and cost effective [1, 2]. Jatropha curcus seed is more effective resource for biodiesel due to its higher oil content and non-edible nature [3].

There are many biodiesel producing resources like Soybean oil, Palm oil, Mustard oil, Sunflower oil, but Jatropha curcas oil is the most promising as it produces the purest quality biodiesel [4]. Jatropha plants are generally affected by pests. This affects the growth of the plant and consequently the oil production. Pest control is a worldwide problem in agricultural ecosystem management. So we need to control these pests for healthy growth of the plant and for oil productivity improvement. Many researchers formulated mathematical model for controlling pest to study the different aspects of the pest management strategies with probable results for elevated applications using analysis of the system within the mathematical illustration. Though broad spectrum chemical pesticides have been suggested [5, 6] to use, those chemical pesticides are harmful to our health and plant growth and it causes environmental pollution, health hazards and uneconomic crop production [7]. This leads to the interest in biological control methods for plant pests. Beside that, the most effective measures in pest management are determined by the ecology of a pest . Thus to resolve this type of problem, the concept of Integrated Pest Management (IPM) [8] is being generated and its application is being increased in the field by marginal farmers recently. IPM seeks to diminish dependence on pesticides by emphasizing the contribution of biological control. Role of microbial pesticides in the IPM has been recently examined for agriculture and other fields [9–11]. As components of an integrated approach, bio-pesticide can play significant role in pest control [12]. Potentially, the use of virus is one of the most effective biological methods for controlling pests. This advantage is subdued in the case where the transmission is carried out by an insect vector. In North America and European countries we can see the practical evidence of the use of virus against insect pests [13]. The experimental and field use of pathogenic viruses in Europe is listed by Falcon et al. [14].

As the fuel crop is economically very important, our main object is to protect the crop from its complications to enhance its oil production. Pests are the main barrier for natural growth of Jatropha curcus plant. Many pests are known to occur on Jatropha curcus but less than ten type occur quite frequently, e.g. the leaf miner Stomphastis thraustica, the leaf and stem miner Pempelia morosalis, and the shield-backed bug Calidea panaethiopica etc. [15].

The natural enemy (predator) in the system survives on the susceptible and infected pest. The predator consumes the infected pest in linear mass action due to the fact that the viral infection makes some behavioral changes and sub lethal effect on host. In this research article we want to compare the change in nature of the system due to the consumption of susceptible pest by predator with Holling type I, II functional responses in the view of mathematical and numerical analysis.

In the next section we formulate the mathematical model. After that we perform local stability analysis around different equilibria. In the section after stability analysis, numerical simulations are shown. Finally, we conclude our findings.

2 The Model

2.1 General Model Formulation with Suitable Assumptions

Here we have formulated a five dimensional mathematical model, containing biomass of Jatropha plant $j(t)$, susceptible pest $s(t)$, infected pest $i(t)$, predator $p(t)$ and virus population $v(t)$. Individual plant growth follows logistic fashion where r_j denotes the maximum growth rate and k_j denotes the carrying capacity of the plant. The pest population is partitioned into two classes, Susceptible and Infected pest. Pest consumes the plant resource at a rate α which is further converted into the susceptible pest with maximum growth rate r_s. The carrying capacity of the susceptible pest is assumed to be k_s. The virus population interact with the susceptible pests and turn them into infected pest class at a λ. Here we have considered the functional response $f(s)$ of the predator population on the susceptible pest population as linear, hyperbolic and sigmoid which helps the predators in their growth at a rate θ_1. The predators consumes the infected pests at a rate l. ξ is the natural mortality rate of the infected pests. d_p is the natural death rate of predators and ε_p is the intra-specific competition coefficient among predators present in the predatory guild of infected pest. θ_2 is the growth rate of the predator due to predation of the infected pests. π_v is the constant recruitment rate of the virus. κ is the virus replication rate. The mortality rate of the virus population is μ_v.

$$\begin{aligned}
\frac{dj}{dt} &= r_j j(1-\frac{j}{k_j}) - \alpha j s,\\
\frac{ds}{dt} &= r_s j s(1-\frac{s+i}{k_s}) - \lambda s v - f(s)p,\\
\frac{di}{dt} &= \lambda s v - \xi i - l i p,\\
\frac{dp}{dt} &= p(-d_p - \varepsilon_p p) + \theta_1 f(s)p + \theta_2 i p,\\
\frac{dv}{dt} &= \pi_v + \kappa \xi i - \mu_v v - \gamma s v.
\end{aligned} \tag{1}$$

Where $j(0) \geq 0, s(0) \geq 0, i(0) \geq 0, p(0) \geq 0, v(0) \geq 0$ and all the parameters are assumed to be non-negative. The function $f(s)$ is of two different following types:

(i) **Holling Type I** or **linear functional response** and (ii) **Holling Type II** or **hyperbolic functional response**.

2.2 Linear Functional Response

We first consider the Holling type-I functional response. We consider the equilibria of the above system and discuss their local stability properties. For linear functional response, system takes the following form:

$$
\begin{aligned}
\frac{dj}{dt} &= r_j j(1 - \frac{j}{k_j}) - \alpha js,\\
\frac{ds}{dt} &= r_s js(1 - \frac{s+i}{k_s}) - \lambda sv - \beta sp,\\
\frac{di}{dt} &= \lambda sv - \xi i - lip,\\
\frac{dp}{dt} &= p(-d_p - \varepsilon_p p) + \theta_1 \beta sp + \theta_2 ip,\\
\frac{dv}{dt} &= \pi_v + \kappa \xi i - \mu_v v - \gamma sv.
\end{aligned}
\tag{2}
$$

2.2.1 Existence of Equilibria and Stability

Theorem 1 *The axial equilibrium point* $E = (0, 0, 0, 0, \pi_v)$ *exists and the system (2) is unstable around* E *for all the parametric values.*

Theorem 2 *The pest free equilibrium point* $E_0 = (k_j, 0, 0, 0, \frac{\pi_v}{\mu_v})$ *exists and the system (2) is stable around* E_0 *if* $k_j < \frac{\lambda \pi_v}{r_s \mu_v}$.

Pest free equilibrium: $E_0 = (k_j, 0, 0, 0, \frac{\pi_v}{\mu_v})$,
The Jacobian matrix for pest free equilibrium point is given by,

$$
J = \begin{pmatrix}
-r_j & -\alpha k_j & 0 & 0 & 0\\
0 & r_s k_j - \lambda \frac{\pi_v}{\mu_v} & 0 & 0 & 0\\
0 & \lambda \frac{\pi_v}{\mu_v} & -\xi & 0 & 0\\
0 & 0 & 0 & -d_p & 0\\
0 & -\gamma \frac{\pi_v}{\mu_v} & \kappa \xi & 0 & -\mu_v
\end{pmatrix}
$$

At E_0 the above system is stable if $k_j < \frac{\lambda \pi_v}{r_s \mu_v}$.

Theorem 3 *The predator free equilibrium point* E_1 *exists if* $\kappa > \frac{\gamma}{\lambda}$ *and* $\frac{r_j}{\alpha} < \bar{s} < \frac{\mu_v}{k\lambda - \gamma}$ *and the system (2) is stable around* E_1 *for condition (4).*

The predator free equilibrium point $E_1 = (\bar{j}, \bar{s}, \bar{i}, 0, \bar{v})$,
where $\bar{s}$ is the positive root of the cubic equation $A\bar{s}^3 + B\bar{s}^2 + C\bar{s} + D = 0$ (A, B, C, D are given in Appendix A) and
$\bar{v} = \frac{\pi_v}{\mu_v - (k\lambda - \gamma)\bar{s}}$, $\bar{j} = k_j(1 - \frac{\alpha \bar{s}}{r_j})$, $\bar{i} = \frac{\pi_v \lambda \bar{s}}{\xi[\mu_v - (\kappa\lambda - \gamma)\bar{s}]}$.

The predator-free equilibrium exists when $\kappa > \frac{\gamma}{\lambda}$ and $\frac{r_j}{\alpha} < \bar{s} < \frac{\mu_v}{k\lambda - \gamma}$.
The characteristic equation corresponding to the variational matrix at predator free equilibrium point given in Appendix A is,

$$\lambda^5 + B_1\lambda^4 + B_2\lambda^3 + B_3\lambda^2 + B_4\lambda + B_5 = 0. \tag{3}$$

Where $B_i s (i = 1, 2, 3, 4)$ are given in Appendix A.
Then by Routh-Hurwitz criterion it follows that the predator free equilibrium point $E_1(\bar{j}, \bar{s}, \bar{i}, 0, \bar{v})$ is locally asymptotically stable if

$$\begin{aligned} &(i)\ B_i (i = 1, 2, 3, 4, 5) > 0 \\ &(ii)\ B_1 B_2 B_3 > B_3^2 + B_1^2 B_4 \\ &(iii)\ (B_1 B_4 - B_5)(B_1 B_2 B_3 - B_3^2 - B_1^2 B_4) > B_5 (B_1 B_2 - B_3)^2 + B_1 B_5^2. \end{aligned} \tag{4}$$

Theorem 4 *The interior equilibrium point E^* exists if $s^* < \frac{r_j}{\alpha}$ and the system (2) is stable around E^* for condition (7).*

The interior equilibrium: $E^*(j^*, s^*, i^*, p^*, v^*)$.
Here $j^* = k_j(1 - \frac{\alpha s^*}{r_j})$,
i^* is the positive root of the equation

$$C_1 i^{*2} + C_2 i^* + C_3 = 0. \tag{5}$$

where C_is are given in Appendix B and
$p^* = \frac{\theta_1 \beta s^* + \theta_2 i^* - d_p}{\varepsilon_p}$, $v^* = \frac{i^*[\xi \varepsilon_p + l(\theta_1 \beta s^* + \theta_2 i^* - d_p)]}{\varepsilon_p \lambda s^*}$.
In Eq. (4) we have a single variation of sign. Then by Descartes' Rule of Sign we should have unique positive root. Therefore $C_2 < 0$.
i.e. $\kappa\lambda\xi - \mu_v \xi - \frac{l\mu_v \theta_1 \beta}{\varepsilon_p} - \frac{l\mu_v d_p}{\varepsilon_p s^*} - \gamma\xi - \frac{\gamma l}{\varepsilon_p}(\theta_1 \beta s^* - d_p) < 0.$
$l\gamma\beta s^{*2} + (\varepsilon_p \mu_v \xi + l\mu_v \theta_1 \beta + \varepsilon_p \gamma \xi - \kappa \varepsilon_p \xi - l\gamma d_p)s^* + l\mu_v d_p > 0.$
This can be written as $(s - a)(s - b) > 0 \cdot (a < b)(a, b$ are given in Appendix B)
The interior equilibrium point E^* exists when $s^* < \frac{r_j}{\alpha}$.
The characteristic equation corresponding to the variational matrix at interior equilibrium point given in Appendix B is,

$$\lambda_1^5 + D_1\lambda_1^4 + D_2\lambda_1^3 + D_3\lambda_1^2 + D_4\lambda_1 + D_5 = 0. \tag{6}$$

Where $D_i (i = 1,\ 2,\ 3,\ 4)$s are given in Appendix B.
Then by Routh-Hurwitz criterion it follows that the interior equilibrium point $E^* = (j^*, s^*, i^*, p^*, v^*)$ is locally asymptotically stable if

$$\begin{aligned} &(i)\ D_i (i\ =\ 1,\ 2,\ 3,\ 4,\ 5) > 0 \\ &(ii)\ D_1 D_2 D_3 > D_3^2 + D_1^2 D_4 \\ &(iii)\ (D_1 D_4 - D_5)(D_1 D_2 D_3 - D_3^2 - D_1^2 D_4) > D_5 (D_1 D_2 - D_3)^2 + D_1 D_5^2. \end{aligned} \tag{7}$$

2.3 Hyperbolic Functional Response

To catch a susceptible pest predators need some time to search for its food. Hence the searching efficiency of the predators can play an important role in the system. Hence the objective of this subsection is to introduce the model with hyperbolic functional response to observe the dynamics of the system. For hyperbolic functional response, system takes the following form:

$$\begin{aligned}
\frac{dj}{dt} &= r_j j(1-\frac{j}{k_j}) - \alpha js,\\
\frac{ds}{dt} &= r_s js(1-\frac{s+i}{k_s}) - \lambda sv - \frac{\beta sp}{a+s},\\
\frac{di}{dt} &= \lambda sv - \xi i - lip,\\
\frac{dp}{dt} &= p(-d_p - \varepsilon_p p) + \theta_1 \frac{\beta sp}{a+s} + \theta_2 ip,\\
\frac{dv}{dt} &= \pi_v + \kappa\xi i - \mu_v v - \gamma sv.
\end{aligned} \tag{8}$$

Where a is the searching efficiency of the predator.

2.3.1 Existence of Equilibria and Stability

Theorem 5 *The axial equilibrium point $E^{11} = (0,0,0,0,\pi_v)$ exists and the system (8) is unstable around E^{11} for all the parametric values.*

Theorem 6 *The pest free equilibrium point $E_0^{11} = (k_j,0,0,0,\frac{\pi_v}{\mu_v})$ exists and the system (8) is stable around E_0^{11} if $k_j < \frac{\lambda\pi_v}{r_s\mu_v}$.*

The Jacobian matrix for pest free equilibrium point is given by,

$$J = \begin{pmatrix} -r_j & -\alpha k_j & 0 & 0 & 0 \\ 0 & r_s k_j - \lambda\frac{\pi_v}{\mu_v} & 0 & 0 & 0 \\ 0 & \lambda\frac{\pi_v}{\mu_v} & -\xi & 0 & 0 \\ 0 & 0 & 0 & -d_p & 0 \\ 0 & -\gamma\frac{\pi_v}{\mu_v} & \kappa\xi & 0 & -\mu_v \end{pmatrix}$$

At E_0^{11} the above system is stable if $k_j < \frac{\lambda\pi_v}{r_s\mu_v}$.

Theorem 7 *The predator free equilibrium point E_1^{11} exists if $\kappa > \frac{\gamma}{\lambda}$ and $\frac{r_j}{\alpha} < \hat{s} < \frac{\mu_v}{k\lambda-\gamma}$ and the system (8) is stable around E_1^{11} for condition (10) .*

Predator free equilibrium: $E_1^{11}(\hat{j}, \hat{s}, \hat{i}, 0, \hat{v})$.
Where $\hat{s}$ is the positive root of the cubic equation $\grave{A}\hat{s}^3 + \grave{B}\hat{s}^2 + \grave{C}\hat{s} + \grave{D} = 0$ ($\grave{A}$, $\grave{B}$, $\grave{C}$, $\grave{D}$ are given in Appendix C) and
$\hat{v} = \frac{\pi_v}{\mu_v - (\kappa\lambda - \gamma)\hat{s}}$, $\hat{j} = k_j(1 - \frac{\alpha\hat{s}}{r_j})$, $\hat{i} = \frac{\pi_v \lambda \hat{s}}{\xi[\mu_v - (\kappa\lambda - \gamma)\hat{s}]}$
The predator free equilibrium exists when $\kappa > \frac{\gamma}{\lambda}$ and $\frac{r_j}{\alpha} < \hat{s} < \frac{\mu_v}{k\lambda - \gamma}$.
The characteristic equation of the variational matrix given in Appendix C is,

$$\mu^5 + b_1\mu^4 + b_2\mu^3 + b_3\mu^2 + b_4\mu + b_5 = 0. \tag{9}$$

where $b_i s (i = 1, 2, 3, 4)$ are given in Appendix C.
Then by Routh-Hurwitz criterion it follows that the predator free equilibrium point $E_1 = (\hat{j}, \hat{s}, \hat{i}, 0, \hat{v})$ is locally asymptotically stable if

$$\begin{aligned} &(i)\, b_i (i = 1, 2, 3, 4, 5) > 0 \\ &(ii)\, b_1 b_2 b_3 > b_3^2 + b_1^2 b_4 \\ &(iii)\, (b_1 b_4 - b_5)(b_1 b_2 b_3 - b_3^2 - b_1^2 b_4) > b_5(b_1 b_2 - b_3)^2 + b_1 b_5^2. \end{aligned} \tag{10}$$

Theorem 8 *The equilibrium point E_* exists if $k_j > j_*$ and the system (8) is stable around E_* for condition (13).*

Interior equilibrium: $E_*^{11}(j_*, s_*, i_*, p_*, v_*)$.
Here $s_* = \frac{r_j}{\alpha}(1 - \frac{j_*}{k_j})$,
$p_* = -\frac{d_p}{\varepsilon_p} + \frac{\theta_1 \beta r_j (k_j - j_*)}{\varepsilon_p (a\alpha k_j + r_j(k_j - j_*))} + \frac{\theta_2}{\varepsilon_p} i_*$
$v_* = \frac{\alpha k_j}{\lambda r_j (k_j - j_*)}[\xi i_* + l i_* (-\frac{d_p}{\varepsilon_p} + \frac{\theta_1 \beta r_j (k_j - j_*)}{\varepsilon_p (a\alpha k_j + r_j(k_j - j_*))} + \frac{\theta_2}{\varepsilon_p} i_*)]$
j_* is the positive root of the equation

$$c_1 {j_*}^2 + c_2 j_* + c_3 = 0. \tag{11}$$

where c_is are given in Appendix D.
The interior equilibrium exists if $k_j > j_*$.
The characteristic equation of the variational matrix given in Appendix D is,

$$\mu_1^5 + \grave{B}_1\mu_1^4 + \grave{B}_2\mu_1^3 + \grave{B}_3\mu_1^2 + \grave{B}_4\mu_1 + \grave{B}_5 = 0. \tag{12}$$

where $B_i (i = 1, 2, 3, 4)$s are given in Appendix D.
Then by Routh-Hurwitz criterion it follows that the interior equilibrium point $E_*^{11} = (j_*, s_*, i_*, p_*, v_*)$ is locally asymptotically stable if

$$\begin{aligned} &(i) \grave{B}_i (i = 1, 2, 3, 4, 5) > 0 \\ &(ii) \grave{B}_1 \grave{B}_2 \grave{B}_3 > {\grave{B}_3}^2 + {\grave{B}_1}^2 \grave{B}_4 \\ &(iii) (\grave{B}_1 \grave{B}_4 - \grave{B}_5)(\grave{B}_1 \grave{B}_2 \grave{B}_3 - {\grave{B}_3}^2 - {\grave{B}_1}^2 \grave{B}_4) > \grave{B}_5(\grave{B}_1 \grave{B}_2 - \grave{B}_3)^2 + \grave{B}_1 {\grave{B}_5}^2. \end{aligned} \tag{13}$$

3 Numerical Simulation and Discussion

In this section, we will present some numerical simulation results using Matlab to validate our analytical findings. To reduce the pest, we introduce virus spraying. Here our main objective is to control the pest population by applying virus to get healthy Jatropha plant. This numerical experiment and simulation is done under parameters given as Table 1.

In Fig. 1, we plotted the model variables as function of time. It is clear that the system moves towards its stable region as time increases. In figure (a) the trajectories of plant biomass for Holling type I and II functional responses have been compared. Here we can observe that for hyperbolic functional response the plant biomass reaches its stability more faster than other functional responses. We can also observe from figure (b) and (c)that susceptible pest is transformed into infected pest for and exterminated by the virus interference. For hyperbolic functional response transformation of susceptible pest to infected pest is more rapid than other functional responses. The figure (d) shows that the predator population is initially decreased to a certain level and then increased gradually. Finally predator population assumes a steady state (Fig. 2).

Figure 3 depicts the effect of the virus replication parameter κ on different population. We vary the value of κ from 5 to 500. For less value of κ the system becomes unstable. Since κ is the virus replication parameter, with the increasing value of κ the virus population is increasing. Consequently the susceptible pest population is readily converted into infected pest and susceptible pest population size is decreased. Here the model system moves towards stable pest-free equilibrium point at $\kappa = 500$. As a result plant biomass is reaching its maximum value within 100 days (Tables 2 and 3).

Table 1 Existence and stability for Holling type I functional response

Equilibrium point	Existence	Stability
$E = (0,\ 0,\ 0,\ 0, \pi_v)$	Exists	Always unstable
$E_0 = (k_j,\ 0,\ 0,\ 0,\ \frac{\pi_v}{\mu_v})$	Exists	Stable if $k_j < \frac{\lambda\pi_v}{r_s\mu_v}$
$E_1 = (\bar{j}, \bar{s}, \bar{i}, 0, \bar{v})$	Exists if $\kappa > \frac{\gamma}{\lambda}$ and $\frac{r_j}{\alpha} < \bar{s} < \frac{\mu_v}{k\lambda - \gamma}$	Stable if $B_i(i = 1, 2, 3, 4, 5) > 0, B_1B_2B_3 > B_3^2 + B_1^2B_4, (B_1B_4 - B_5)(B_1B_2B_3 - B_3^2 - B_1^2B_4) > B_5(B_1B_2 - B_3)^2 + B_1B_5^2$
$E^* = (j^*,\ s^*,\ i^*,\ p^*,\ v^*)$	Exists if $s^* < \frac{r_j}{\alpha}$	Stable if $D_i(i = 1, 2, 3, 4, 5) > 0, D_1D_2D_3 > D_3^2 + D_1^2D_4, (D_1D_4 - D_5)(D_1D_2D_3 - D_3^2 - D_1^2D_4) > D_5(D_1D_2 - D_3)^2 + D_1D_5^2$

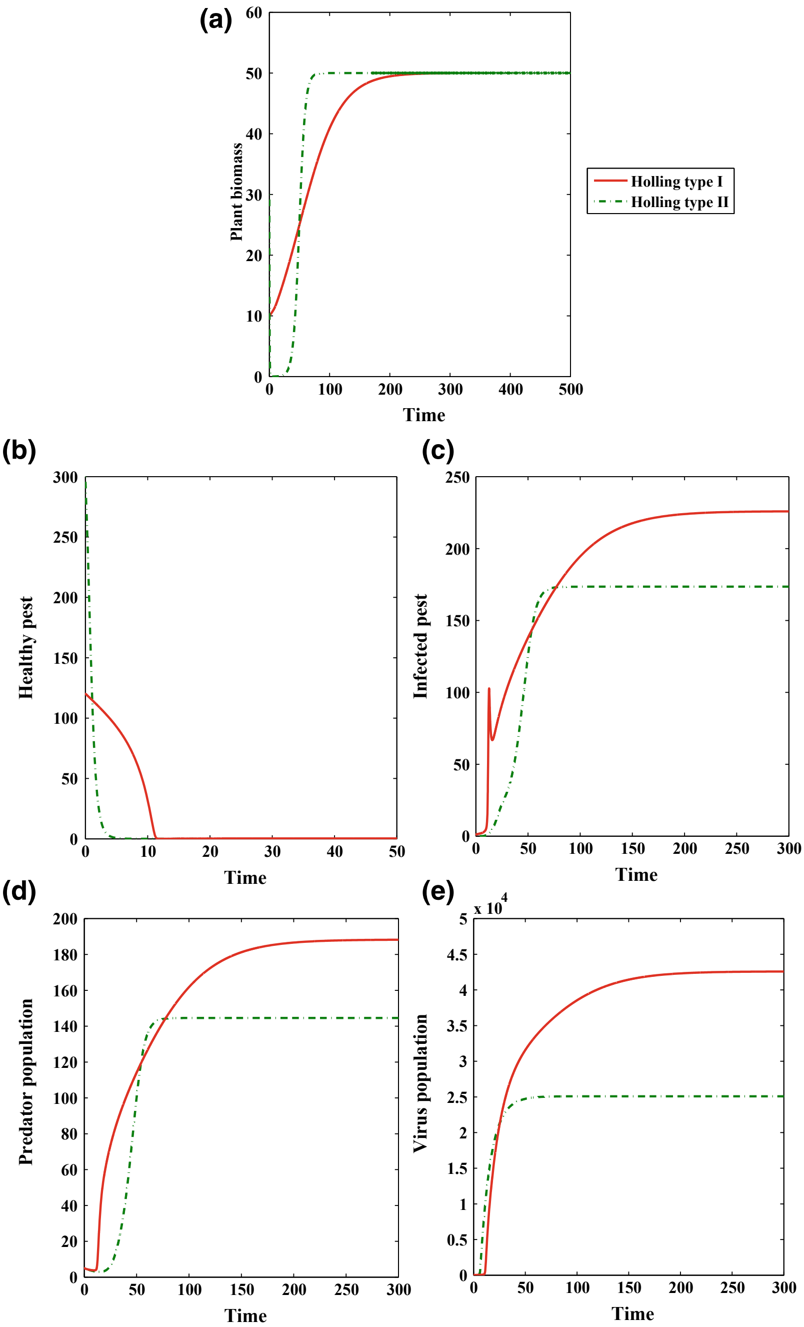


Fig. 1 Comparison of trajectories for different functional responses for: Plant biomass, Healthy pest, Infected pest, Predator population and Virus population at $\kappa = 500$, other parameter values are given in Table 1

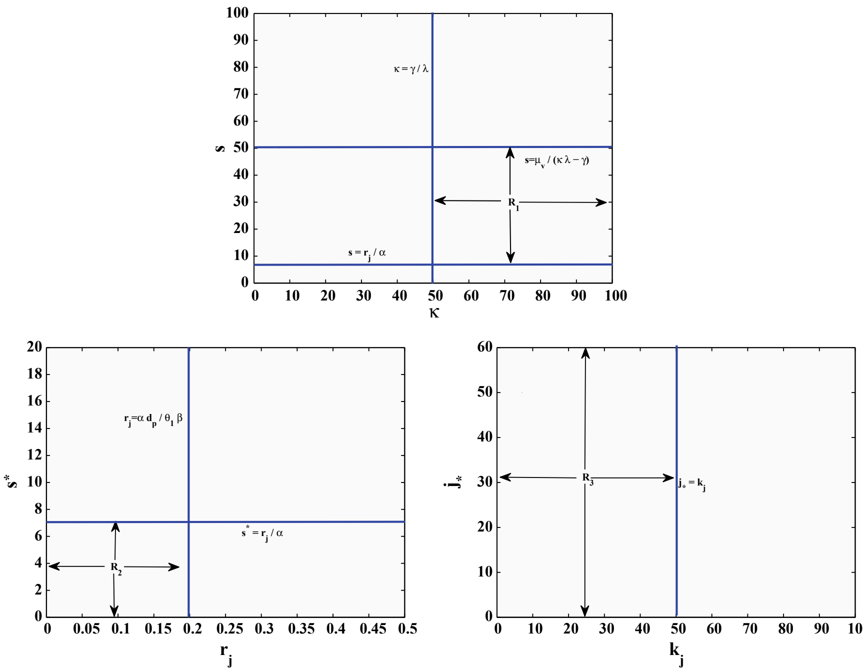


Fig. 2 Region of existence for different equilibria. Here R_1 is the existence region of predator free equilibrium points for the system with both the functional responses, R_2 and R_3 are the existence regions of interior equilibrium points for the system with functional response I and II respectively

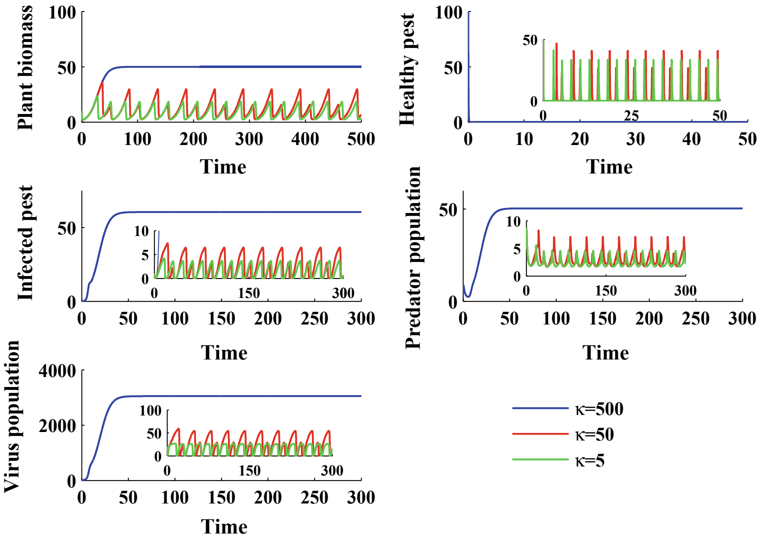


Fig. 3 Dynamics of the system with different values of κ for Holling type II functional responce, other parameter values are given in Table 1

Table 2 Existence and stability for Holling type II functional response

Equilibrium point	Existence	Stability
$E^{11} = (0, 0, 0, 0, \pi_v)$	Exists	Always unstable
$E_0^{11} = (k_j, 0, 0, 0, \frac{\pi_v}{\mu_v})$ $E_1^{11} = (\hat{j}, \hat{s}, \hat{i}, 0, \hat{v})$	Exists Exists if $\kappa > \frac{\gamma}{\lambda}$ and $\frac{r_j}{\alpha} < \hat{s} < \frac{\mu_v}{k\lambda - \gamma}$	Stable if $k_j < \frac{\lambda \pi_v}{r_s \mu_v}$ Stable if $b_i (i = 1, 2, 3, 4, 5) > 0, b_1 b_2 b_3 > b_3^2 + b_1^2 b_4, (b_1 b_4 - b_5)(b_1 b_2 b_3 - b_3^2 - b_1^2 b_4) > b_5 (b_1 b_2 - b_3)^2 + b_1 b_5^2$
$E_*^{11} = (j_*, s_*, i_*, p_*, v_*)$	Exists if $k_j > j_*,$	Stable if $\grave{B}_i (i = 1, 2, 3, 4, 5) > 0, \grave{B}_1 \grave{B}_2 \grave{B}_3 > {\grave{B}_3}^2 + {\grave{B}_1}^2 \grave{B}_4, (\grave{B}_1 \grave{B}_4 - \grave{B}_5)(\grave{B}_1 \grave{B}_2 \grave{B}_3 - {\grave{B}_3}^2 - {\grave{B}_1}^2 \grave{B}_4) > \grave{B}_5 (\grave{B}_1 \grave{B}_2 - \grave{B}_3)^2 + \grave{B}_1 {\grave{B}_5}^2$

Table 3 Parameters value used for numerical simulation [4, 6, 16]

Parameters	Definition	Values (Unit)
r_j	The growth rate of plant biomass	0.03 $kg\ day^{-1}$
k_j	The maximum density biomass of plant	50 $kg\ plant^{-1}$
r_s	Conversion factor of susceptible pest	0.05 day^{-1}
k_s	The pest carrying capacity	300 $plant^{-1}$
α	The interaction rate between pest and plant	0.0001 $plant^{-1} day^{-1}$
γ	reduction rate constant of virus	0.008 day^{-1}
λ	The infection rate of pest by virus	0.003 $pest^{-1} day^{-1}$
ξ	The mortality rate of infected pest	0.01 day^{-1}
d_p	The mortality rate of predator	0.006 day^{-1}
ε_p	lysis of predator due to competition	0.002 day^{-1}
θ_1	conversion factor for predator	0.05
θ_2	conversion factor for predator	0.01
μ_v	The decay rate of virus	0.1 $gm\ day^{-1}$
β	consumption rate of susceptible pest by predator	0.015 $pest^{-1} day^{-1}$
l	consumption rate of infected pest by predator	0.7 $pest^{-1} day^{-1}$
a	the half-saturation coefficient	0.5 (constant)

Figures 4 and 5 represent the time series solution of the model equation for hyperbolic functional response in virus-plant-healthy pest plane with initial values [50, 100, 50], [20, 200, 20], [5, 300,100] and [30, 300, 60], [20, 200, 20], [10,100, 40] respectively . Figure 3 describes that as time increases the system converges to pest-free equilibria for $\kappa = 500$. Figure 4 depicts that as time increases the system converges to endemic equilibria for $\kappa = 100$.

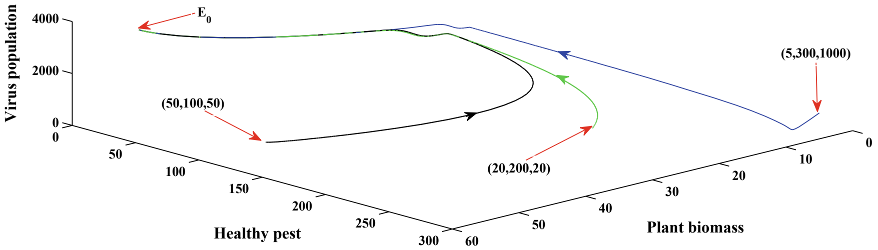


Fig. 4 Phase portrait in virus-plant-healthy pest plane showing that the system moves towards pest-free equilibrium point and the system becomes stable for Holling type II functional responce at $\kappa = 500$, other parameter values are given in Table 1

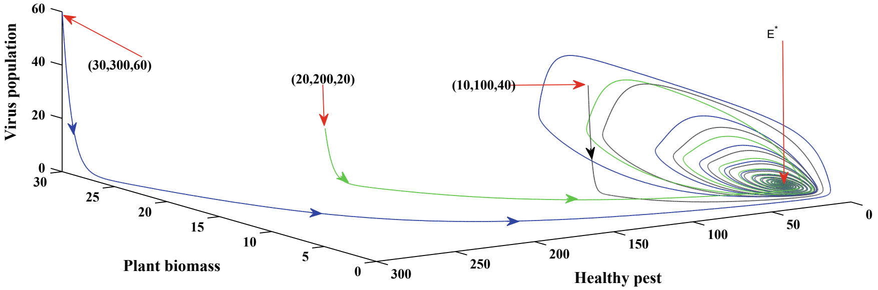


Fig. 5 Phase portrait in virus-plant-healthy pest plane showing that the system moves towards endemic equilibrium point and the system becomes stable for Holling type II functional responce at $\kappa = 100$, other parameter values are given in Table 1

In Figure 6 we have shown a mesh plotting in $\kappa - \lambda$-susceptible pest plane. Figure (a) shows that for Holling type I functional response with the increasing value of κ and λ, the pest population decreases but not gets eradicated totally. In figure (b) we have seen that for Holling type II functional response, the healthy pests are exterminated when κ lies between 85 to 200 and λ lies between 0.005 to 0.01.

4 Conclusion

In this research article, our main aim is to control pest of Jatropha curcas plant using virus as controlling agent. Analytically we inspected the system from the viewpoint of stability and existence. We have checked the local stability at pest free equilibrium, predator free equilibrium points and interior equilibrium point for different functional

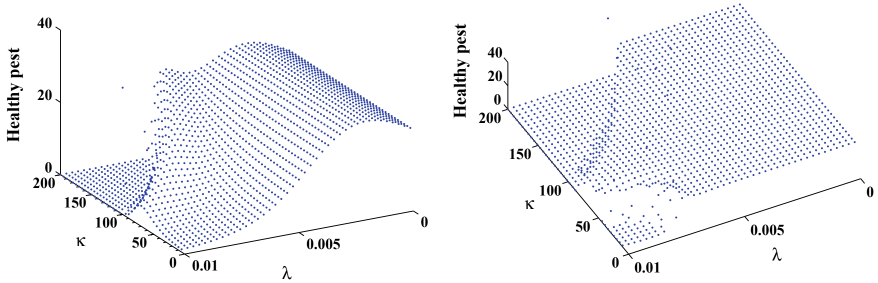


Fig. 6 Mesh plotting for **a** linear and **b** hyperbolic in κ — λ-susceptible pest plane

responses on predator. Numerically, we examine the effect of virus replication. Here, we observe the changes of dynamical behavior with respect to different time intervals. The reason for using different types of Holling functional responses is to observe which functional response would be a suitable candidate to represent pest eradication. If the pest becomes dominant, then Jatropha will be affected heavily with economic loss. Also, if the prey becomes extinct, then the natural predator will die out, that may affect the biological balance of the ecosystem. Thus, it is very important to maintain the biological balance of the ecosystem in such a way so that in one hand crop yield will be maximized and predators also survive. We can easily see that hyperbolic functional response is more effective than that of linear functional response as within 60 days higher number of infected pest will be saturated and the system becomes stable, whereas for linear functional response it will take 150 days to stabilize the system. Finally, our work reveals that an introduction of predators/ natural enemies with a hyperbolic functional response would be most effective to control pest and maximize healthy plant production.

Acknowledgements The research work is financially supported by DST-INSPIRE fellowship, Government of India, Department of Mathematics, Jadavpur University, Kolkata-700032, India.

Appendix A

Coefficients of $A\bar{s}^3 + B\bar{s}^2 + C\bar{s} + D = 0$ are as follows,
$A = \frac{A'}{r_j k_s \xi[\mu_v - (\kappa\lambda - \gamma)\bar{s}]}$, $B = \frac{B'}{r_j k_s \xi[\mu_v - (\kappa\lambda - \gamma)\bar{s}]}$,
$C = \frac{C'}{r_j k_s \xi[\mu_v - (\kappa\lambda - \gamma)\bar{s}]}$, $D = \frac{D' - \lambda\pi_v r_j k_s \xi}{r_j k_s \xi[\mu_v - (\kappa\lambda - \gamma)\bar{s}]}$; where
$A' = r_s k_j \alpha\xi(\gamma - \kappa\lambda)$, $B' = r_s k_j(\alpha\xi k_s \kappa\lambda - \alpha\xi\lambda k_s + \alpha\xi\mu_v + \xi r_j \kappa\lambda - r_j \xi\gamma + \alpha\mu\lambda)$,
$C' = r_s k_j(\xi\lambda r_j k_s - \alpha\xi\mu_v k_s - \xi k_s \kappa\lambda r_j - \xi\mu_v r_j - \lambda\mu_v r_j)$, $D' = \xi\mu_v r_s r_j k_j k_s$,

The Jacobian matrix for predator free equilibrium point for type I functional response is given by,

$$J = \begin{pmatrix} a^{11} & a^{12} & 0 & 0 & 0 \\ a^{21} & a^{22} & a^{23} & a^{24} & a^{25} \\ 0 & a^{32} & a^{33} & a^{34} & a^{35} \\ 0 & 0 & 0 & a^{44} & 0 \\ 0 & a^{52} & a^{53} & 0 & a^{55} \end{pmatrix}$$

where
$a^{11} = r_j - \alpha\bar{s} - 2\bar{j}\frac{r_j}{k_j}, a^{12} = -\alpha\bar{j}, a^{21} = r_s\bar{s}(1 - \frac{\bar{s}+\bar{i}}{k_s}),$
$a^{22} = r_s\bar{j}(1 - \frac{\bar{s}+\bar{i}}{k_s}) - \frac{r_s}{k_s}\bar{j}\bar{s} - \lambda\bar{v}, a^{23} = -\frac{r_s}{k_s}\bar{j}\bar{s},$
$a^{24} = -\beta\bar{s}, a^{25} = -\lambda\bar{s}, a^{32} = \lambda\bar{v}, a^{33} = -\xi,$
$a^{34} = -l\bar{i},\ \ a^{35} = \lambda\bar{s},\ \ a^{44} = -d_p + \theta_1\beta\bar{s} + \theta_2\bar{i},\ \ a^{52} = -\gamma\bar{v},\ \ a^{53} = k\xi,\ \ a^{55} = -(\mu_v + \gamma\bar{s}).$

The coefficients of Eq. 1 are as follows,
$B_1 = -\sum a^{ii},$
$B_2 = \sum a^{ii}a^{jj} - \sum a^{ij}a^{ji},$
$B_3 = -\sum a^{ii}a^{jj}a^{kk} + \sum a^{ij}a^{ji}a^{kk} - \sum a^{ij}a^{jk}a^{ki},$
$B_4 = \sum a^{ii}a^{jj}a^{kk}a^{ll} - \sum a^{ij}a^{ji}a^{kk}a^{ll} + \sum a^{ij}a^{jk}a^{ki}a^{ll} - \sum a^{ij}a^{ji}a^{kl}a^{lk},$
$B_5 = -a^{ii}a^{jj}a^{kk}a^{ll}a^{mm} + \sum a^{ij}a^{ji}a^{kk}a^{ll}a^{mm} - \sum a^{ij}a^{jk}a^{ki}a^{ll}a^{nn} + \sum a^{ij}a^{ji}a^{kl}a^{lk}a^{mm}.$
$(i, j, k, l, m = \{1, 2, 3, 4, 5\}$ and $i \neq j \neq k \neq l \neq m)$

Appendix B

The coefficients of Eq. 5 are given as follows,
$C_1 = \frac{l\mu_v\theta_2}{\varepsilon_p s^*} + \gamma l\theta_2,$
$C_2 = k\lambda\xi - \mu_v\xi - \frac{l\mu_v\theta_1\beta}{\varepsilon_p} - \frac{l\mu_v d_p}{\varepsilon_p s^*} - \gamma\xi - \frac{\gamma l}{\varepsilon_p}(\theta_1\beta s^* - d_p),$
$C_3 = \lambda\pi_v.$

The constants in $(s - a)(s - b) > 0(a < b)$ are
$a = \frac{-(\varepsilon_p\mu_v\xi + l\mu_v\theta_1\beta + \varepsilon_p\gamma\xi - \kappa\varepsilon_p\xi - l\gamma d_p) - \surd((\varepsilon_p\mu_v\xi + l\mu_v\theta_1\beta + \varepsilon_p\gamma\xi - \kappa\varepsilon_p\xi - l\gamma d_p)^2 - 4l^2\gamma\theta_1\beta\mu_v d_p)}{2l\gamma\theta_1\beta}$ and
$b = \frac{-(\varepsilon_p\mu_v\xi + l\mu_v\theta_1\beta + \varepsilon_p\gamma\xi - \kappa\varepsilon_p\xi - l\gamma d_p) + \surd((\varepsilon_p\mu_v\xi + l\mu_v\theta_1\beta + \varepsilon_p\gamma\xi - \kappa\varepsilon_p\xi - l\gamma d_p)^2 - 4l^2\gamma\theta_1\beta\mu_v d_p)}{2l\gamma\theta_1\beta}.$
The Jacobian matrix for the interior equilibrium point for type I functional response is given by,

$$J = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ 0 & a_{32} & a_{33} & a_{34} & a_{35} \\ 0 & a_{42} & a_{43} & a_{44} & 0 \\ 0 & a_{52} & a_{53} & 0 & a_{55} \end{pmatrix}$$

where,
where, $a_{11} = r_j - \alpha s^* - 2j^*\frac{r_j}{k_j}, a_{12} = -\alpha j^*, a_{21} = r_s s^*(1 - \frac{s^*+i^*}{k_s}),$

$a_{22} = r_s j^*(1 - \frac{s^*+i^*}{k_s}) - \frac{r_s}{k_s} j^* s^* - \lambda v^* - \beta p^*, a_{23} = -\frac{r_s}{k_s} j^* s^*,$
$a_{24} = -\beta s^*, a_{25} = -\lambda s^*, a_{32} = \lambda v^*, a_{33} = -(\xi + l p^*),$
$a_{34} = -l i^*, a_{35} = \lambda s^*, a_{42} = \theta_1 \beta p^*, a_{43} = \theta_2 p^*,$
$a_{44} = -d_p - 2\varepsilon_p p^* + \theta_1 \beta s^* + \theta_2 i^*, a_{52} = -\gamma v^* \ a_{53} = \kappa \xi \ a_{55} = -(\mu_v + \gamma s^*).$

The coefficients of Eq. 6 are as follows,
$D_1 = -\sum a_{ii},$
$D_2 = \sum a_{ii} a_{jj} - \sum a_{ij} a_{ji},$
$D_3 = -\sum a_{ii} a_{jj} a_{kk} + \sum a_{ij} a_{ji} a_{kk} - \sum a_{ij} a_{jk} a_{ki},$
$D_4 = \sum a_{ii} a_{jj} a_{kk} a_{ll} - \sum a_{ij} a_{ji} a_{kk} a_{ll} + \sum a_{ij} a_{jk} a_{ki} a_{ll} - \sum a_{ij} a_{ji} a_{kl} a_{lk},$
$D_5 = -a_{ii} a_{jj} a_{kk} a_{ll} a_{mm} + \sum a_{ij} a_{ji} a_{kk} a_{ll} a_{mm} - \sum a_{ij} a_{jk} a_{ki} a_{ll} a_{nn} +$
$\sum a_{ij} a_{ji} a_{kl} a_{lk} a_{mm} - \sum a_{ii} a_{jk} a_{kl} a_{lm} a_{mj}.$
$(i, j, k, l, m = \{1, 2, 3, 4, 5\}$ and $i \neq j \neq k \neq l \neq m)$

Appendix C

Coefficients of $\grave{A}\hat{s}^3 + \grave{B}\hat{s}^2 + \grave{C}\hat{s} + \grave{D} = 0$ are as follows,
$\grave{A} = \frac{\tilde{A}}{r_j k_s \xi [\mu_v - (\kappa\lambda - \gamma)\hat{s}]}, \grave{B} = \frac{\tilde{B}}{r_j k_s \xi [\mu_v - (\kappa\lambda - \gamma)\hat{s}]},$
$\grave{C} = \frac{\tilde{C}}{r_j k_s \xi [\mu_v - (\kappa\lambda - \gamma)\hat{s}]}, \grave{D} = \frac{\tilde{D} - \lambda \pi_v r_j k_s \xi}{r_j k_s \xi [\mu_v - (\kappa\lambda - \gamma)\hat{s}]};$
where
$\tilde{A} = r_s k_j \alpha \xi (\gamma - \kappa\lambda), \qquad \tilde{B} = r_s k_j (\alpha \xi k_s \kappa \lambda - \alpha \xi \lambda k_s + \alpha \xi \mu_v + \xi r_j \kappa \lambda - r_j \xi \gamma +$
$\alpha \mu \lambda),$
$\tilde{C} = r_s k_j (\xi \lambda r_j k_s - \alpha \xi \mu_v k_s - \xi k_s \kappa \lambda r_j - \xi \mu_v r_j - \lambda \mu_v r_j), \tilde{D} = \xi \mu_v r_s r_j k_j k_s,$

The Jacobian matrix for the predator-free equilibrium point for type II functional response is given by,

$$J = \begin{pmatrix} b^{11} & b^{12} & 0 & 0 & 0 \\ b^{21} & b^{22} & b^{23} & b^{24} & b^{25} \\ 0 & b^{32} & b^{33} & b^{34} & b^{35} \\ 0 & 0 & 0 & b^{44} & 0 \\ 0 & b^{52} & b^{53} & 0 & b^{55} \end{pmatrix}$$

where
$b^{11} = r_j - \alpha\hat{s} - 2\hat{j}\frac{r_j}{k_j}, b^{12} = -\alpha\hat{j}, b^{21} = r_s\hat{s}(1 - \frac{\hat{s}+\hat{i}}{k_s}),$
$b^{22} = r_s\hat{j}(1 - \frac{\hat{s}+\hat{i}}{k_s}) - \frac{r_s}{k_s}\hat{j}\hat{s} - \lambda\hat{v}, b^{23} = -\frac{r_s}{k_s}\hat{j}\hat{s},$
$b^{24} = -\beta\frac{\hat{s}}{\alpha+\hat{s}}, b^{25} = -\lambda\hat{s}, b^{32} = \lambda\hat{v}, b^{33} = -\xi,$
$b^{34} = -l\hat{i}, \ b^{35} = \lambda\hat{s}, \ b^{44} = -d_p + \theta_1\beta\frac{\hat{s}}{\alpha+\hat{s}} + \theta_2\hat{i}, \ b^{52} = -\gamma\hat{v}, \ b^{53} = \kappa\xi, \ b^{55} =$
$-(\mu_v + \gamma\hat{s}).$

The coefficients of Eq. 9 are as follows,
$b_1 = -\sum b^{ii},$
$b_2 = \sum b^{ii} b^{jj} - \sum b^{ij} b^{ji},$
$b_3 = -\sum b^{ii} b^{jj} b^{kk} + \sum b^{ij} b^{ji} b^{kk} - \sum b^{ij} b^{jk} b^{ki},$

$b_4 = \sum b^{ii}b^{jj}b^{kk}b^{ll} - \sum b^{ij}b^{ji}b^{kk}b^{ll} + \sum b^{ij}b^{jk}b^{ki}b^{ll} - \sum b^{ij}b^{ji}b^{kl}b^{lk}$,
$b_5 = -b^{ii}b^{jj}b^{kk}b^{ll}b^{mm} + \sum b^{ij}b^{ji}b^{kk}b^{ll}b^{mm} - \sum b^{ij}b^{jk}b^{ki}b^{ll}b^{nn} +$
$\sum b^{ij}b^{ji}b^{kl}b^{lk}b^{mm}$.
$(i, j, k, l, m = \{1, 2, 3, 4, 5\}$ and $i \neq j \neq k \neq l \neq m)$

Appendix D

The coefficients of Eq. 11 are as follows,
$c_1 = \pi_v\lambda\varepsilon_p r_j^2 + \kappa\xi\lambda\varepsilon_p r_j^2 - \gamma\xi\varepsilon_p r_j^2 i_* - \gamma\theta_1\beta r_j^2 + \gamma\theta_2\alpha k_j a r_j - \gamma\theta_2 r_j i_*$,
$c_2 = -\pi_v\lambda r_j\varepsilon_p a\alpha k_j - \kappa\xi\lambda r_j\varepsilon_p a\alpha k_j i_* - 2\pi_v\lambda r_j^2\varepsilon_p k_j - 2\kappa k_j\xi\lambda r_j^2\varepsilon_p +$
$\xi\mu_v\alpha k_j r_j\varepsilon_p i - \mu_v\alpha k_j l d_p r_j i_* + \theta_1\beta r_j + \theta_2 i_* + \gamma\xi\varepsilon_p a\alpha r_j k_j i_* + 2\gamma\xi\varepsilon_p r_j^2 k_j i_* -$
$\gamma r_j d_p a l\alpha r_j i_* - \gamma r_j^2 d_p k_j l i_* + \gamma r_j^2\theta_1\beta k_j + \gamma r_j k_j\theta_1\beta + \gamma r_j k_j^2\theta_2 a\alpha + \gamma r_j^2\theta_2 k_j i_* +$
$\gamma r_j k_j\theta_2 i_*$,
$c_3 = \pi_v\lambda r_j\varepsilon_p a\alpha k_j^2 + \kappa k_j^2\xi\lambda r_j^2\varepsilon_p + \kappa\xi\lambda r_j\varepsilon_p a\alpha k_j^2 i_* - \xi\mu_v\alpha^2 k_j^2\varepsilon_p - \xi\mu_v\alpha k_j^2 r_j\varepsilon_p i_*$
$+ \mu_v\alpha^2 k_j^2 l d_p a i_* + \mu_v\alpha k_j^2 l d_p r_j i_* - \theta_1\beta r_j k_j - \theta_2 a\alpha k_j i_* - \theta_2 k_j r_j i_* -$
$\lambda\xi\varepsilon_p a\alpha r_j k_j^2 i_* - \gamma\xi\varepsilon_p r_j^2 k_j^2 i_* + \gamma r_j k_j^2 d_p a l\alpha i_* + \gamma r_j^2 k_j^2 d_p l i_* - \gamma r_j^2 k_j^2\theta_1\beta -$
$\gamma r_j^2 k_j^2\theta_2 i_*$.

The Jacobian matrix for the interior equilibrium point for type II functional response is given by,

$$J = \begin{pmatrix} b_{11} & b_{12} & 0 & 0 & 0 \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ 0 & b_{32} & b_{33} & b_{34} & b_{35} \\ 0 & b_{42} & b_{43} & b_{44} & 0 \\ 0 & b_{52} & b_{53} & 0 & b_{55} \end{pmatrix}$$

where,
$b_{11} = r_j - \alpha s^* - 2j^*\frac{r_j}{k_j}$, $b_{12} = -\alpha j^*$, $b_{21} = r_s s^*(1 - \frac{s^*+i^*}{k_s})$,
$b_{22} = r_s j^*(1 - \frac{s^*+i^*}{k_s}) - \frac{r_s}{k_s}j^*s^* - \lambda v^* - \beta p^*$, $b_{23} = -\frac{r_s}{k_s}j^*s^*$,
$b_{24} = -\beta s^*$, $b_{25} = -\lambda s^*$, $b_{32} = \lambda v^*$, $b_{33} = -(\xi + lp^*)$,
$b_{34} = -li^*$, $b_{35} = \lambda s^*$, $b_{42} = \theta_1\beta p^*$, $b_{43} = \theta_2 p^*$,
$b_{44} = -d_p - 2\varepsilon_p p^* + \theta_1\beta s^* + \theta_2 i^*$, $b_{52} = -\gamma v^*$, $b_{53} = \kappa\xi$, $b_{55} = -(\mu_v + \gamma s^*)$.

The coefficients of Eq. 12 are as follows,
$\grave{B}_1 = -\sum b_{ii}$,
$\grave{B}_2 = \sum b_{ii}b_{jj} - \sum b_{ij}b_{ji}$,
$\grave{B}_3 = -\sum b_{ii}b_{jj}b_{kk} + \sum b_{ij}b_{ji}b_{kk} - \sum b_{ij}b_{jk}b_{ki}$,
$\grave{B}_4 = \sum b_{ii}b_{jj}b_{kk}b_{ll} - \sum b_{ij}b_{ji}b_{kk}b_{ll} + \sum b_{ij}b_{jk}b_{ki}b_{ll} - \sum b_{ij}b_{ji}b_{kl}b_{lk}$,
$\grave{B}_5 = -b_{ii}b_{jj}b_{kk}b_{ll}b_{mm} + \sum b_{ij}b_{ji}b_{kk}b_{ll}b_{mm} - \sum b_{ij}b_{jk}b_{ki}b_{ll}b_{nn} +$
$\sum b_{ij}b_{ji}b_{kl}b_{lk}b_{mm} - \sum b_{ii}b_{jk}b_{kl}b_{lm}b_{mj}$.
$(i, j, k, l, m = \{1, 2, 3, 4, 5\}$ and $i \neq j \neq k \neq l \neq m)$

References

1. Ju, L.P., Chen, B.: Embodied energy and energy evaluation of a typical biodiesel production chain in China. Ecol. Model. **222**, 2385–2392 (2011)
2. Chowdhury, J., Al Basir, F., Pal, J., Roy, P.K.: Studies on biodiesel production from Jatropha curcas oil using chemical and biochemical methods–a mathematical approach. Fuel **158**, 503–511 (2015)
3. Roy, P.K., Datta, S., Nandi, S., Basir, F.A.: Effect of mass transfer kinetics for maximum production of biodiesel from Jatropha Curcas oil: a mathematical approach. Fuel **134**, 39–44 (2014)
4. Chowdhury, J., Al Basir, F., Pal, J., Roy, P.K.: Pest control for Jatropha curcas plant through viral disease: a mathematical approach.Nonlin. Stud. **23**(4), 517–532 (2016)
5. Dias, W.O., Wanner, E.F., Cardoso, R.T.: A multiobjective optimization approach for combating Aedes aegypti using chemical and biological alternated step-size control. Math. Biosci. 269, 37–47 (2015)
6. E. Venturino, P. K. ROY, F. AL BASIR, A. DATTA, A model for the control of the mosaic virus disease in Jatropha Curcas plantations. Energy Ecology Environ. 10.1007-40974-016-0033-8 (2016)
7. Georghiou, G.P.: Overview of insecticide resistance, In: ACS Symposium Series-American Chemical Society (USA) (1990)
8. Thomas, M.B.: Ecological approaches and the development of truly integrated pest management. Proc. Natl. Acad. Sci **96**(11), 5944–5951 (1999)
9. Van Frankenhuyzen, K., Reardon, R.C., Dubois, N.R.: Forest defoliators, Field Manual of Techniques in Invertebrate Pathology, pp. 481–504. Springer, Netherlands (2007)
10. G. M. TATCHELL, Microbial insecticides and IPM: current and future opportunities for the use of biopesticides. In: BCPC Symposium Proceedings (1997)
11. Dent, D.R.: Integrated pest management and microbial insecticides in Microbial Insecticides: Novelty or Necessity? In: Proceedings of the British Crop Protection Council Symposium **127–138**, (1997)
12. Bhattacharya, D.K., Karan, S.: On bionomic model of integrated pest management of a single pest population. J. Differ. Equat. Dyn. Syst. **12**(4), 301–330 (2004)
13. Franz, J.M., Huber, J.: Feldversuche mit insektenpathogenen Viren in Europa. Entomophaga **24**(4), 333–343 (1979)
14. Van den Bosch, R., Messenger P.S., Gutierrez, A.P.: Microbial control of insects, weeds, and plant pathogens. In: An Introduction to Biological Control. Springer US, pp. 59–74 (1982)
15. Terren, M., Mignon, J., De Clerck, C., Jijakli, H., Savery, S.: Principal disease and insect pests of Jatropha curcas L. in the Lower Valley of the Senegal River. Tropicultura, **30**(4), 222–229 (2012)
16. Bhattacharyya, S., Bhattacharya, D.K.: Pest control through viral disease: mathematical modeling and analysis. J. Theor. Biol. **238**(1), 177–197 (2006)

Global Dynamics of a TB Model with Classes Age Structure and Environmental Transmission

Yan-Xia Dang, Juan Wang, Xue-Zhi Li and Mini Ghosh

Abstract In this article, an age structured SVEIR epidemic model for TB is formulated and analyzed by considering three types of ages e.g., latent age, infection age and vaccination age. The presented model also incorporates the environmental transmission of TB. The dynamics of the disease is governed by a system of differential-integral equations. We assume that the vaccines for TB are partially effective. Some vaccinated individuals get permanent immunity to this disease, but some vaccinated individuals lose its protective power over a time and become susceptible again. The dynamical property of the model is established by using LaSalle's invariance principle and constructing suitable Lyapunov functions. It has been shown that the dynamics of the model is governed by basic reproductive number $\mathcal{R}(\xi)$. The disease-free equilibrium is globally stable if the basic reproductive number $\mathcal{R}(\xi) < 1$. The endemic equilibrium is locally and globally stable if $\mathcal{R}(\xi) > 1$. As the basic reproduction number plays an important role in determining the stability of the system, reducing this number below one through vaccination can lead to decrease in the transmission of this disease. Additionally, contaminated environment also contributes to

Supported partially by the NNSF of China (11601465, 11771017).

Y.-X. Dang
Department of Public Education, Zhumadian Vocational and Technical College,
Zhumadian 463000, China
e-mail: dyx125@126.com

J. Wang
Department of Mathematics, Xinyang Normal University,
Xinyang 464000, China
e-mail: dotey99@126.com

X.-Z. Li (✉)
College of Mathematics and Information Sciences, Henan Normal University,
Xinxiang 453007, China
e-mail: xzli66@126.com

M. Ghosh
School of Advanced Sciences, Vellore Institute of Technology,
Chennai Campus, Chennai 600127, India
e-mail: minighosh@vit.ac.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_33

the increase in $\mathcal{R}(\xi)$, so we also need to keep the environment clean to decrease the basic reproduction number $\mathcal{R}(\xi)$ below one. These types of control measures are easy to implement in our society and certainly this will improve the well-being of the society.

Keywords TB model · Age-since-latency · Age-since-infection · Age-since-vaccination · Environmental transmission · Reproduction number · Global stability · Lyapunov function

AMS subject classifications 92D30

1 Introduction

Tuberculosis (TB) [1] is a bacterial disease caused by *Mycobacterium tuberculosis*. It is believed that at least one-third of the world human population is the reservoir of this disease [2, 3]. Usually, this disease is acquired through airborne infection from someone who has active TB (smear-positive TB).

One of the important features of TB is that a relatively small proportion of those infected go on to develop clinical disease [4]. Most people are assumed to mount an effective immune response to the initial infection that limits proliferation of the bacilli and may lead to long-lasting partial immunity both to further infection and to reactivation of latent bacilli existing from the original infection.

Active-TB (the clinical disease) can develop into pulmonary or extrapulmonary form. Extrapulmonary TB is common in children while pulmonary TB is prevalent in adults. Mycobacterium tuberculosis, the causal agent of the disease, is transmitted almost exclusively via pulmonary cases (exceptions could include laryngeal TB).

Individuals who have latent infection are not clinically ill or capable of transmitting TB [3]. Exposed individuals may remain in this latent stage for long and variable periods of time (in fact, may die without ever developing active TB). Apparently, the longer we carry this bacterium the less likely we are to develop active TB unless our immune system becomes seriously compromised by other disease. Consequently, the development of active TB after infection is highly dependent on the infection age. As a result, in our study, we consider both the latent age and the age of infection. Additionally, Center for Disease Control (CDC) trace data on TB infection cases had validated that TB can be transmitted while traveling by air plane [5]. These data suggests that TB infections may occur without having extensive and repeated exposure to TB-active individuals. It is quite possible for an individual to become infected while using crowded public transportation for several hours a day. This assumption was supported by an outbreak of tuberculosis which was originated in a neighborhood bar (see [6]). Thus an individual may acquire infection by using the mass-transportation or by involving himself/herself in a community meetings/get-togethers. And the urbanization process that exploded after the industrial revolution has played a fundamental role in the observed patterns of TB spread in industrialized nations. The process of industrialization caused damage to the ecological environment leading to urban environmental pollution. This suggests the necessity of modeling both direct and environmental transmission of this disease. Meanwhile, the persistence and the pandemic threat of avian influenza as well as the very publicized cholera outbreak in Haiti have increased the awareness of diseases which transmit both directly and environmentally. Therefore, the environmental transmission of bacteria is an important factor that affects the disease dynamics and needs to be considered in the prevention and control of such diseases.

Recently, the epidemiological models with vaccination have been analyzed by many researchers [7–15]. In this article, we assume that the vaccines for TB bacterium are partially effective, i.e., some vaccinated individuals get permanent immunity to this disease while some vaccinated individuals lose its protective power over time and get infected by TB bacteria.

Tuberculosis (TB) is an ancient disease that still supports huge levels of prevalence across the world. Global tuberculosis control is facing major challenges today. In general, much effort is still required to make quality care accessible without barriers of gender, age, type of disease, social setting, and ability to pay. Coinfection with Mycobacterium tuberculosis and HIV(TB/HIV), especially in Africa, and multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in all regions, make control activities more complex and demanding.

The formulation of suitable mathematical model depending upon geographic region, societal structure, culture tradition etc. is very important for better prediction of the future dynamics of this disease. Here we formulate an age structured SVEIR model of TB disease with environmental transmission by considering three ages, e.g., latent age, infection age and vaccination age. The purpose of this article is to analyze the system of differential-integral equations that describes the dynamics of disease transmission for tuberculosis (TB). We aim to look at how the vaccination and the contaminated environment influence the transmission dynamics of TB. The main interest in studying the proposed model is to understand the long-time behavior of the disease transmission dynamics i.e., whether the disease will die out eventually or will persist in the population. A clear answer to this question is practically important to design the suitable disease control strategies.

This article is organized as follows. In Sect. 2, we introduce the model. In Sect. 3, we compute the equilibria and analyze their local stability. In Sect. 4, we show that the disease is uniformly persistent if the reproduction number $\mathcal{R}(\xi) > 1$ by applying the persistence theory for infinite-dimensional systems. In Sect. 5, we prove the global stability of the disease-free equilibrium. In Sect. 6, the global stability of the endemic equilibrium is established. We conclude with a discussion in Sect. 7.

2 The Model

To formulate the model first we denote the number of susceptible human individuals by $S(t)$. We structure the infected individuals by age-since-infection τ. Let $E(\tau, t)$ be the density of individuals infected by TB bacteria at any time t who are infected with age-since-infection τ but not yet infectious. Let $i(\tau, t)$ be the density of individuals infected by TB bacteria at any time t with age-since-infection τ. Let $W(\theta, t)$ be the density of TB bacteria of age $\theta > 0$ at time t in the contaminated environment. Let $V(a, t)$ be the density of vaccinated individuals with respect to the age of vaccination $a > 0$ at any time t. Let $R(t)$ be the number of recovered or immune individuals at any time t. Keeping in view of the above notation, we formulate our age-structured SVEIR model with environmental transmission. Here three types of ages e.g. latent age, infection age and vaccination age are incorporated in the model.

$$\begin{cases} \dfrac{dS}{dt} = \Lambda - S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau - S\int_0^\infty \xi(\theta)W(\theta,t)d\theta - (\mu+\zeta)S + \int_0^\infty \alpha_2(a)V(a,t)da, \\ \dfrac{\partial E(\tau,t)}{\partial \tau} + \dfrac{\partial E(\tau,t)}{\partial t} = -(\mu + m(\tau))E(\tau,t), \\ E(0,t) = S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S\int_0^\infty \xi(\theta)W(\theta,t)d\theta, \\ \dfrac{\partial i(\tau,t)}{\partial \tau} + \dfrac{\partial i(\tau,t)}{\partial t} = -(\mu + \alpha_1(\tau) + r_1(\tau))i(\tau,t), \\ i(0,t) = \displaystyle\int_0^\infty m(\tau)E(\tau,t)d\tau, \\ \dfrac{\partial W(\theta,t)}{\partial \theta} + \dfrac{\partial W(\theta,t)}{\partial t} = -\delta(\theta)W(\theta,t), \\ W(0,t) = \displaystyle\int_0^\infty \eta(\tau)i(\tau,t)d\tau, \\ \dfrac{\partial V(a,t)}{\partial a} + \dfrac{\partial V(a,t)}{\partial t} = -(\mu + \alpha_2(a) + r_2(a))V(a,t), \\ V(0,t) = \zeta S, \\ \dfrac{dR}{dt} = \displaystyle\int_0^\infty r_1(\tau)i(\tau,t)d\tau + \int_0^\infty r_2(a)V(a,t)da - \mu R. \end{cases} \tag{2.1}$$

Here Λ is the birth/recruitment rate, $m(\tau)$ denotes the removal rate from the latent period, $\alpha_1(\tau)$ gives the additional host mortality due to the bacteria, $\beta(\tau)$ is the transmission coefficient and $\eta(\tau)$ is the shedding rate of an infected individual of infection age τ infected by TB bacteria. Furthermore, $\alpha_2(a)$ denotes the rate at which vaccine wanes and is a bounded general function of vaccination-age a, ζ is the rate of vaccination of the susceptible individuals, $r_1(\tau)$ denotes the recovery rate of infected individuals, and $r_2(a)$ is the rate of the vaccinated individuals acquiring vaccination throughout their lives. Finally, $\xi(\theta)$ is the transmission rate of TB bacteria from the contaminated environment, and $\delta(\theta)$ is the clearance rate of bacteria of age θ from the environment.

To understand the model, notice that susceptible individuals are recruited at a rate Λ. Susceptible individuals can become infected with TB bacteria either through a direct contact with an infected individual with TB bacteria or through coming into contact with TB bacteria that are in the environment. Infection through direct contact with infected individuals can happen through contact with individuals of any age-since-infection at a specific age-specific transmission rate. As a consequence, the force of infection of susceptible individuals through direct contact is given by the integral over all ages-since-infection. So is the force of infection of susceptible individuals through the contaminated environment. Upon infection through direct or indirect transmission, the newly infected individuals move to the latent class, then progress into the infectious class with the progression rate $m(\tau)$. The non-infectious and infectious individuals infected by TB bacteria with age-since-infection equal to zero constitute the boundary conditions. Individuals infected with TB shed the TB bacteria into the environment at a rate $\eta(\tau)$. All bacterial particles shed by individuals infected with TB of all age classes are given by the integral. That gives the number of bacterial particles in the environment with age-since-infection equal to zero at time t in the environment. Again, the newly vaccinated individuals coming into the vaccinated class with age-since-vaccination equal to zero constitutes the boundary condition. The vaccine lose its protective power with time and eventually some vaccinated individuals become susceptible again. The total number of the vaccinated individuals becoming susceptible is given by the integral over all ages-since-vaccination. But the vaccine is partially effective, some vaccinated individuals acquire immunity throughout their lives. The number of recovered individuals from the infected class is given by the integral over all ages-since-infection, and the number of individuals acquiring immunity is given by the integral over all ages-since-vaccination.

We notice that the equation for the recovered individuals is decoupled from the system and the analysis of system (2.1) is equivalent to the analysis of the following system

$$\begin{cases} \dfrac{dS}{dt} = \Lambda - S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau - S\int_0^\infty \xi(\theta)W(\theta,t)d\theta - (\mu+\zeta)S + \int_0^\infty \alpha_2(a)V(a,t)da, \\ \dfrac{\partial E(\tau,t)}{\partial \tau} + \dfrac{\partial E(\tau,t)}{\partial t} = -(\mu + m(\tau))E(\tau,t), \\ E(0,t) = S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S\int_0^\infty \xi(\theta)W(\theta,t)d\theta, \\ \dfrac{\partial i(\tau,t)}{\partial \tau} + \dfrac{\partial i(\tau,t)}{\partial t} = -(\mu + \alpha_1(\tau) + r_1(\tau))i(\tau,t), \\ i(0,t) = \displaystyle\int_0^\infty m(\tau)E(\tau,t)d\tau, \\ \dfrac{\partial W(\theta,t)}{\partial \theta} + \dfrac{\partial W(\theta,t)}{\partial t} = -\delta(\theta)W(\theta,t), \\ W(0,t) = \displaystyle\int_0^\infty \eta(\tau)i(\tau,t)d\tau, \\ \dfrac{\partial V(a,t)}{\partial a} + \dfrac{\partial V(a,t)}{\partial t} = -(\mu + \alpha_2(a) + r_2(a))V(a,t), \\ V(0,t) = \zeta S. \end{cases} \tag{2.2}$$

Model (2.2) is equipped with the following initial conditions:

$$S(0)=S_0,\quad E(\tau,0)=\varphi(\tau),\quad i(\tau,0)=\psi(\tau),\quad W(\theta,0)=\phi(\theta),\quad V(a,0)=V_0(a).$$

All parameters are nonnegative, i.e., $\Lambda>0,\ \zeta>0$, and $\mu>0$. We make the following assumptions on the parameter-functions.

Assumption 2.1 The parameter-functions satisfy the following.

1. The functions $\beta(\tau)$ and $\xi(\theta)$ are bounded and uniformly continuous. When $\beta(\tau)$ and $\xi(\theta)$ are of compact support, the support has non-zero Lebesgue measure;
2. The functions $m(\tau)$, $\alpha_1(\tau)$, $\eta(\tau)$, $r_1(\tau)$, $\delta(\theta)$, $\alpha_2(a)$, $r_2(a)$ belong to $L^\infty(0,\infty)$;
3. The functions $\varphi(\tau)$, $\psi(\tau)$, $\phi(\theta)$ and $V_0(a)$ are integrable.

Define the space of functions

$$X=\mathbb{R}\times(L^1(0,\infty))\times(L^1(0,\infty))\times(L^1(0,\infty))\times(L^1(0,\infty)).$$

It can be verified that solutions of (2.2) with nonnegative initial conditions belong to the positive cone for $t\geq 0$. Furthermore, adding all the equations we have

$$\begin{aligned}&\frac{d}{dt}\left(S(t)+\int_0^\infty E(\tau,t)d\tau+\int_0^\infty i(\tau,t)d\tau+\int_0^\infty V(a,t)da\right)\\ &\leq\Lambda-\mu\left(S(t)+\int_0^\infty E(\tau,t)d\tau+\int_0^\infty i(\tau,t)d\tau+\int_0^\infty V(a,t)da\right).\end{aligned}$$

Hence,

$$\limsup_t\left(S(t)+\int_0^\infty E(\tau,t)d\tau+\int_0^\infty i(\tau,t)d\tau+\int_0^\infty V(a,t)da\right)\leq\frac{\Lambda}{\mu}.$$

The free bacteria in the environment can be bounded as follows:

$$\begin{aligned}&\frac{d}{dt}\left(\int_0^\infty W(\theta,t)d\theta\right)\\ &\leq W(0,t)-\dot{\delta}\left(\int_0^\infty W(\theta,t)d\theta\right)\\ &=\int_0^\infty \eta(\tau)i(\tau,t)d\tau-\dot{\delta}\left(\int_0^\infty W(\theta,t)d\theta\right)\\ &\leq\bar{\eta}\int_0^\infty i(\tau,t)d\tau-\dot{\delta}\left(\int_0^\infty W(\theta,t)d\theta\right)\\ &\leq\frac{\Lambda}{\mu}\bar{\eta}-\dot{\delta}\left(\int_0^\infty W(\theta,t)d\theta\right),\end{aligned}$$

where $\bar{\eta} = \sup_{\tau}\{\eta(\tau)\}$ and $\dot{\delta} = \inf_{\theta}\{\delta(\theta)\}$. Hence,

$$\limsup_{t} \int_0^{\infty} W(\theta, t)d\theta \leq \frac{\bar{\eta}\frac{\Lambda}{\mu}}{\dot{\delta}} = \frac{\bar{\eta}\Lambda}{\dot{\delta}\mu}.$$

Therefore, the following set is positively invariant for system

$$\Omega = \left\{(S, E, i, W, V) \in X_+ \middle| \int_0^{\infty} W(\theta, t)d\theta \leq \frac{\bar{\eta}\Lambda}{\dot{\delta}\mu}, \right.$$
$$\left. \left(S(t) + \int_0^{\infty} E(\tau, t)d\tau + \int_0^{\infty} i(\tau, t)d\tau + \int_0^{\infty} V(a, t)da\right) \leq \frac{\Lambda}{\mu}\right\}.$$

Finally, since the exit rate from the latent compartment is given by $\mu + m(\tau)$, the probability of still being incubated after τ time units is given by

$$\pi_1(\tau) = e^{-\mu\tau}e^{-\int_0^{\tau} m(s)ds}, \tag{2.3}$$

the exit rate from the infective compartment is given by $\mu + \alpha_1(\tau) + r_1(\tau)$, so the probability of still being infectious after τ time units is given by

$$\pi_2(\tau) = e^{-\mu\tau}e^{-\int_0^{\tau}(\alpha_1(s)+r_1(s))ds}, \tag{2.4}$$

similarly the probability of TB bacteria of age θ still being in the environment is given by

$$\pi_3(\theta) = e^{-\int_0^{\theta} \delta(s)ds}. \tag{2.5}$$

And the probability of still being vaccinated after a time units is given by

$$\pi_4(a) = e^{-\mu a}e^{-\int_0^{a}(\alpha_2(s)+r_2(s))ds}. \tag{2.6}$$

The reproduction number corresponding to system (2.2) is given by the following expression

$$\begin{aligned}&\mathcal{R}(\zeta)\\ &= \frac{\Lambda\left(\int_0^{\infty} m(\tau)\pi_1(\tau)d\tau \int_0^{\infty} \beta(\tau)\pi_2(\tau)d\tau + \int_0^{\infty} m(\tau)\pi_1(\tau)d\tau \int_0^{\infty} \eta(\tau)\pi_2(\tau)d\tau \int_0^{\infty} \xi(\theta)\pi_3(\theta)d\theta\right)}{\mu + \zeta - \zeta\int_0^{\infty} \alpha_2(a)\pi_4(a)da}.\end{aligned} \tag{2.7}$$

Here the reproduction number gives the number of secondary infections that an individual infected with TB bacteria will produce in an entirely susceptible population. $\mathcal{R}$ gives the strength of TB bacteria to invade when rare and alone.

In the next section we compute explicit expressions for the equilibria and establish their local stability.

3 Equilibria and Their Local Stability

System (2.2) always has a unique disease-free equilibrium $\mathcal{E}_0$, which is given by

$$\mathcal{E}_0 = (S_0^*, 0, 0, 0, V_0^*(a)),$$

where

$$S_0^* = \frac{\Lambda}{\mu + \zeta - \zeta \int_0^\infty \alpha_2(a)\pi_4(a)da}, \qquad V_0^*(\theta) = \zeta S_0^* \pi_4(a).$$

In addition, there is endemic equilibrium $\mathcal{E}_1$ given by

$$\mathcal{E}_1 = \Big(S^*, E^*(\tau), i^*(\tau), W^*(\theta), V^*(a)\Big).$$

The endemic equilibrium $\mathcal{E}_1$ exists if and only if $\mathcal{R}(\zeta) > 1$. The non-zero components of the equilibrium $\mathcal{E}_1$ are given by

$$\begin{aligned} S^* &= \frac{1}{\int_0^\infty m(\tau)\pi_1(\tau)d\tau \int_0^\infty \beta(\tau)\pi_2(\tau)d\tau + \int_0^\infty m(\tau)\pi_1(\tau)d\tau \int_0^\infty \eta(\tau)\pi_2(\tau)d\tau \int_0^\infty \xi(\theta)\pi_3(\theta)d\theta}, \\ E^*(\tau) &= E^*(0)\pi_1(\tau), \quad i^*(\tau) = i^*(0)\pi_2(\tau), \quad W^*(\theta) = W^*(0)\pi_3(\theta), \quad V^*(a) = \zeta S^* \pi_4(a), \end{aligned} \tag{3.1}$$

where

$$E^*(0) = \Lambda\left(1 - \frac{1}{\mathcal{R}}\right), \quad i^*(0) = E^*(0)\int_0^\infty m(\tau)\pi_1(\tau)d\tau, \quad W^*(0) = i^*(0)\int_0^\infty \eta(\tau)\pi_2(\tau)d\tau.$$

Here it can be noticed that

$$\mathcal{R}(\zeta) = \frac{S_0^*}{S^*}. \tag{3.2}$$

Next, we turn to the linearized equations for the disease-free equilibrium. To introduce the linearization at the disease-free equilibrium $\mathcal{E}_0$, let $S(t) = S_0^* + x(t)$, $E(\tau, t) = y(\tau, t)$, $i(\tau, t) = z(\tau, t)$, $W(\theta, t) = e(\theta, t)$, $V(a, t) = V_0^*(a) + h(a, t)$. The linearized system becomes

$$\begin{cases} \dfrac{dx}{dt} = -S_0^* \displaystyle\int_0^\infty \beta(\tau)z(\tau,t)d\tau - S_0^* \int_0^\infty \xi(\theta)e(\theta,t)d\theta - (\mu+\zeta)x + \int_0^\infty \alpha_2(a)h(a,t)da, \\ \dfrac{\partial y(\tau,t)}{\partial \tau} + \dfrac{\partial y(\tau,t)}{\partial t} = -(\mu + m(\tau))y(\tau,t), \\ y(0,t) = S_0^* \displaystyle\int_0^\infty \beta(\tau)z(\tau,t)d\tau + S_0^* \int_0^\infty \xi(\theta)e(\theta,t)d\theta, \\ \dfrac{\partial z(\tau,t)}{\partial \tau} + \dfrac{\partial z(\tau,t)}{\partial t} = -(\mu + \alpha_1(\tau) + r_1(\tau))z(\tau,t), \\ z(0,t) = \displaystyle\int_0^\infty m(\tau)y(\tau,t)d\tau, \\ \dfrac{\partial e(\theta,t)}{\partial \theta} + \dfrac{\partial e(\theta,t)}{\partial t} = -\delta(\theta)e(\theta,t), \\ e(0,t) = \displaystyle\int_0^\infty \eta(\tau)z(\tau,t)d\tau, \\ \dfrac{\partial h(a,t)}{\partial a} + \dfrac{\partial h(a,t)}{\partial t} = -(\mu + \alpha_2(a) + r_2(a))h(a,t), \\ h(0,t) = \zeta x. \end{cases} \tag{3.3}$$

To study the system (2.2), we look for solutions of the form $x(t) = \bar{x}e^{\lambda t}$, $y(\tau,t) = \bar{y}(\tau)e^{\lambda t}$, $z(\tau,t) = \bar{z}(\tau)e^{\lambda t}$, $e(\theta,t) = \bar{e}(\theta)e^{\lambda t}$ and $h(a,t) = \bar{h}(a)e^{\lambda t}$. We obtain the following eigenvalue problem

$$\begin{cases} \lambda\bar{x} = -S_0^* \displaystyle\int_0^\infty \beta(\tau)\bar{z}(\tau)d\tau - S_0^* \int_0^\infty \xi(\theta)\bar{e}(\theta)d\theta - (\mu+\zeta)\bar{x} + \int_0^\infty \alpha_2(a)\bar{h}(a)da \\ \dfrac{d\bar{y}(\tau)}{d\tau} = -(\lambda + \mu + m(\tau))\bar{y}(\tau), \\ \bar{y}(0) = S_0^* \displaystyle\int_0^\infty \beta(\tau)\bar{z}(\tau)d\tau + S_0^* \int_0^\infty \xi(\theta)\bar{e}(\theta)d\theta, \\ \dfrac{d\bar{z}(\tau)}{d\tau} = -(\lambda + \mu + \alpha_1(\tau) + r_1(\tau))\bar{z}(\tau), \\ \bar{z}(0) = \displaystyle\int_0^\infty m(\tau)\bar{y}(\tau)d\tau, \\ \dfrac{d\bar{e}(\theta)}{d\theta} = -(\lambda + \delta(\theta))\bar{e}(\theta), \\ \bar{e}(0) = \displaystyle\int_0^\infty \eta(\tau)\bar{z}(\tau)d\tau, \\ \dfrac{d\bar{h}(a)}{da} = -(\lambda + \mu + \alpha_2(a) + r_2(a))\bar{h}(a), \\ \bar{h}(0) = \zeta\bar{x}. \end{cases} \tag{3.4}$$

Solving the differential equations, we obtain

$$\bar{y}(\tau) = \bar{y}(0)e^{-\lambda\tau}\pi_1(\tau), \quad \bar{z}(\tau) = \bar{z}(0)e^{-\lambda\tau}\pi_2(\tau), \quad \bar{e}(\theta) = \bar{e}(0)e^{-\lambda\theta}\pi_3(\theta),$$
$$\bar{h}(a) = \zeta\bar{x}e^{-\lambda a}\pi_4(a).$$

Substituting for $\bar{y}(\tau)$ in the equation for $\bar{z}(0)$, expressing $\bar{z}(0)$ in term of $\bar{y}(0)$, and replacing $\bar{z}(0)$ in the equation for $\bar{e}(\theta)$, we obtain

$$\bar{z}(\tau) = \bar{y}(0)e^{-\lambda\tau}\pi_2(\tau)\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau,$$
$$\bar{e}(\theta) = \bar{y}(0)e^{-\lambda\theta}\pi_3(\theta)\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau.$$

Similarly, substituting for $\bar{z}(\tau)$ and $\bar{e}(\theta)$ in the equation for $\bar{y}(0)$, then dividing $\bar{y}(0)$ from both side of the resulting equation, we get the following characteristic equation: $\mathcal{G}_1(\lambda) = 1$ where

$$\begin{aligned}\mathcal{G}_1(\lambda) = S_0^* \bigg(&\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau + \\ &\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)e^{-\lambda\theta}\pi_3(\theta)d\theta\bigg).\end{aligned} \tag{3.5}$$

Now we are ready to establish the following result.

Proposition 3.1 *If $\mathcal{R}(\zeta) < 1$, then the disease-free equilibrium is locally asymptotically stable. If the reproduction numbers is larger than one, it is unstable.*

Proof Consider λ with $\Re\lambda \geq 0$. We have

$$|\mathcal{G}_1(\lambda)| \leq |\mathcal{G}_1(\Re\lambda)| \leq |\mathcal{G}_1(0)| = \mathcal{R}(\zeta) < 1.$$

Hence, the equations $\mathcal{G}_1(\lambda) = 1$ do not have a solution with non-negative real part. Therefore, the stability of $\mathcal{E}_0$ depends on the eigenvalues of the following equation

$$\lambda\bar{x} = -(\mu+\zeta)\bar{x} + \zeta\bar{x}\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da.$$

Canceling $\bar{x}$ from both side, we obtain

$$\lambda = -(\mu+\zeta) + \zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da.$$

We rewrite the above characteristic equation in the form

$$\lambda + \mu + \zeta = \zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da. \tag{3.6}$$

We set

$$LHS \overset{def}{=} \lambda + \mu + \zeta, \qquad RHS \overset{def}{=} \zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da. \tag{3.7}$$

If λ is a root with $\Re\lambda \geq 0$, it follows from (3.7) that

$$
\begin{aligned}
\left|RHS\right| &= \left|\zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da\right| \le \left|\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da\right| \\
&\le \zeta\int_0^\infty \alpha_2(a)e^{-\int_0^a \alpha_2(s)ds}d\theta = \zeta\Big(- e^{-\int_0^a \alpha_2(s)ds}\Big|_0^\infty\Big) = \zeta, \\
\left|LHS\right| &= \left|\lambda+\mu+\zeta\right| > \zeta = \left|RHS\right|.
\end{aligned}
$$

It is a contradiction. Hence we conclude that the Eq. (3.6) cannot have any root with a non-negative real part. Therefore, the disease-free equilibrium is locally asymptotically stable.

Now assume $\mathcal{R}(\zeta) > 1$. For λ real we have $\mathcal{G}_1(0) = \mathcal{R}(\zeta) > 1$. Furthermore, $\lim_{\lambda\to\infty}\mathcal{G}_1(\lambda) = 0$. Hence, the equation $\mathcal{G}_1(\lambda) = 1$ has a real positive root. Therefore, the disease-free equilibrium is unstable. □

Now we turn to the local stability of the endemic equilibrium $\mathcal{E}_1$.

Proposition 3.2 *Assume* $\mathcal{R}(\zeta) > 1$. *Then single-strain equilibrium* $\mathcal{E}_1$ *is locally asymptotically stable.*

Proof We introduce the following notation for the perturbations: $S(t) = S^* + x(t)$, $E(\tau, t) = E^*(\tau) + y(\tau, t)$, $i(\tau, t) = i^*(\tau) + z(\tau, t)$, $W(\theta, t) = W^*(\theta) + e(\theta, t)$, $V(a, t) = V^*(a) + h(a, t)$. Then the perturbed system can be written as follows:

$$
\begin{cases}
\dfrac{dx(t)}{dt} = -S^*\displaystyle\int_0^\infty \beta(\tau)z(\tau,t)d\tau - x(t)\int_0^\infty \beta(\tau)i^*(\tau)d\tau - S^*\int_0^\infty \xi(\theta)e(\theta,t)d\theta \\
\qquad -x(t)\displaystyle\int_0^\infty \xi(\theta)W^*(\theta)d\theta - (\mu+\zeta)x(t) + \int_0^\infty \alpha_2(a)h(a,t)da, \\
\dfrac{\partial y(\tau,t)}{\partial\tau} + \dfrac{\partial y(\tau,t)}{\partial t} = -(\mu+m(\tau))y(\tau,t), \\
y(0,t) = S^*\displaystyle\int_0^\infty \beta(\tau)z(\tau,t)d\tau + x(t)\int_0^\infty \beta(\tau)i^*(\tau)d\tau + S^*\int_0^\infty \xi(\theta)e(\theta,t)d\theta \\
\qquad +x(t)\displaystyle\int_0^\infty \xi(\theta)W^*(\theta)d\theta, \\
\dfrac{\partial z(\tau,t)}{\partial\tau} + \dfrac{\partial z(\tau,t)}{\partial t} = -(\mu+\alpha_1(\tau)+r_1(\tau))z(\tau,t), \\
z(0,t) = \displaystyle\int_0^\infty m(\tau)y(\tau,t)d\tau, \\
\dfrac{\partial e(\theta,t)}{\partial\theta} + \dfrac{\partial e(\theta,t)}{\partial t} = -\delta(\theta)e(\theta,t), \\
e(0,t) = \displaystyle\int_0^\infty \eta(\tau)z(\tau,t)d\tau, \\
\dfrac{\partial h(a,t)}{\partial a} + \dfrac{\partial h(a,t)}{\partial t} = -(\mu+\alpha_2(a)+r_2(a))h(a,t), \\
h(0,t) = \zeta x.
\end{cases}
\tag{3.8}
$$

An approach similar to [16] (see Appendix B in [16]) can show that the linear stability of the system is in fact determined by the eigenvalues of the linearized system (3.8). To investigate the point spectrum, we look for exponential solutions (see the case of the disease-free equilibrium) and obtain the following linear eigenvalue problem.

$$\begin{cases} \lambda x = -S^* \int_0^\infty \beta(\tau)z(\tau)d\tau - x\int_0^\infty \beta(\tau)i^*(\tau)d\tau - S^*\int_0^\infty \xi(\theta)e(\theta)d\theta \\ \qquad -x\int_0^\infty \xi(\theta)W^*(\theta)d\theta - (\mu+\zeta)x + \int_0^\infty \alpha_2(a)h(a)da, \\ \dfrac{dy(\tau)}{d\tau} = -(\lambda+\mu+m(\tau))y(\tau), \\ y(0) = S^*\int_0^\infty \beta(\tau)z(\tau)d\tau + x\int_0^\infty \beta(\tau)i^*(\tau)d\tau + S^*\int_0^\infty \xi(\theta)e(\theta)d\theta \\ \qquad +x\int_0^\infty \xi(\theta)W^*(\theta)d\theta, \\ \dfrac{dz(\tau)}{d\tau} = -(\lambda+\mu+\alpha_1(\tau)+r_1(\tau))z(\tau), \\ z(0) = \int_0^\infty m(\tau)y(\tau)d\tau, \\ \dfrac{de(\theta)}{d\theta} = -(\lambda+\delta(\theta))e(\theta), \\ e(0) = \int_0^\infty \eta(\tau)z(\tau)d\tau, \\ \dfrac{dh(a)}{da} = -(\lambda+\mu+\alpha_2(a)+r_2(a))h(a), \\ h(0) = \zeta x. \end{cases} \tag{3.9}$$

Similar to the disease-free equilibrium, we obtain the following characteristic equation

$$\begin{cases} \Big(\lambda+\mu+\zeta+\int_0^\infty \beta(\tau)i^*(\tau)d\tau + \int_0^\infty \xi(\theta)W^*(\theta)d\theta - \zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da\Big)x+ \\ S^*\Big(\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau + \\ \quad \int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)e^{-\lambda\theta}\pi_3(\theta)d\theta\Big)y(0) = 0, \\ \\ -\Big(\int_0^\infty \beta(\tau)i^*(\tau)d\tau + \int_0^\infty \xi(\theta)W^*(\theta)d\theta\Big)x+ \\ \quad \Big(1 - S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau - \\ \quad S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)e^{-\lambda\theta}\pi_3(\theta)d\theta\Big)y(0) = 0. \end{cases} \tag{3.10}$$

Thus, from (3.10), we obtain the following characteristic equation for the endemic equilibrium $\mathcal{E}_1$.

$$\begin{aligned}&\left(\lambda+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da\right)\times\\ &\quad\left(1-S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau-\right.\\ &\qquad\left.S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)e^{-\lambda\theta}\pi_3(\theta)d\theta\right)+\\ &\qquad\qquad\int_0^\infty \beta(\tau)i^*(\tau)d\tau+\int_0^\infty \xi(\theta)W^*(\theta)d\theta=0.\end{aligned}\tag{3.11}$$

Therefore, the stability of $\mathcal{E}_1$ depends on the eigenvalues of the equation. We introduce the following notations

$$\begin{cases}F_1(\lambda)=\lambda+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)e^{-\lambda a}\pi_4(a)da,\\ F_2(\lambda)=1-S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau-\\ \qquad S^*\int_0^\infty m(\tau)e^{-\lambda\tau}\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)e^{-\lambda\tau}\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)e^{-\lambda\theta}\pi_3(\theta)d\theta,\\ F(\lambda)=F_1(\lambda)F_2(\lambda)+\int_0^\infty \beta(\tau)i^*(\tau)d\tau+\int_0^\infty \xi(\theta)W^*(\theta)d\theta.\end{cases}\tag{3.12}$$

Thus the Eq. (3.11) turns into the following equation

$$F(\lambda)=0.\tag{3.13}$$

Differentiating $\Re\lambda$ derivative of $F_1(\Re\lambda)$, $F_2(\Re\lambda)$ with respective to $\Re\lambda$, we obtain

$$F_1'(\Re\lambda)\geq 0,\qquad F_2'(\Re\lambda)\geq 0.$$

We know

$$\begin{cases}F_1(0)=\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da\geq 0,\\ F_2(0)=1-S^*\int_0^\infty m(\tau)\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau-\\ \qquad S^*\int_0^\infty m(\tau)\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta=0.\end{cases}$$

Consider λ with $\Re\lambda \geq 0$, we have

$$\begin{cases} F_1(\Re\lambda) \geq F_1(0) \geq 0, \\ F_2(\Re\lambda) \geq F_2(0) \geq 0. \end{cases}$$

Then we can get

$$\begin{aligned} |F(\lambda)| &= \left| F_1(\lambda)F_2(\lambda) + \int_0^\infty \beta(\tau)i^*(\tau)d\tau + \int_0^\infty \xi(\theta)W^*(\theta)d\theta \right| \\ &\geq \int_0^\infty \beta(\tau)i^*(\tau)d\tau + \int_0^\infty \xi(\theta)W^*(\theta)d\theta > 0. \end{aligned}$$

So for λ with $\Re\lambda \geq 0$, the characteristic equation (3.13) has solutions with negative real parts. Thus, the endemic equilibrium $\mathcal{E}_1$ is locally asymptotically stable. This completes the proof. □

4 The Uniform Strong Persistence of TB Bacteria

We need to integrate the differential equation along the characteristic lines. Denote the initial condition by $B(t)$: $B_1(t) = E(0, t),\ B_2(t) = i(0, t),\ B_3(t) = W(0, t),\ B_4(t) = V(0, t)$. Integrating along the characteristic lines, we obtain

$$E(\tau, t) = \begin{cases} B_1(t-\tau)\pi_1(\tau), & t > \tau, \\ \varphi(\tau - t)\dfrac{\pi_1(\tau)}{\pi_1(\tau - t)}, & t < \tau, \end{cases}$$

$$i(\tau, t) = \begin{cases} B_2(t-\tau)\pi_2(\tau), & t > \tau, \\ \psi(\tau - t)\dfrac{\pi_2(\tau)}{\pi_2(\tau - t)}, & t < \tau, \end{cases}$$

$$W(\theta, t) = \begin{cases} B_3(t-\theta)\pi_3(\theta), & t > \theta, \\ \phi(\theta - t)\dfrac{\pi_3(\theta)}{\pi_3(\theta - t)}, & t < \theta, \end{cases}$$

$$V(a, t) = \begin{cases} \zeta S(t-a)\pi_4(a), & t > a, \\ V_0(a - t)\dfrac{\pi_4(a)}{\pi_4(a - t)}, & t < a. \end{cases} \tag{4.1}$$

We define the following set

$$\hat{\Omega}_1 = \left\{ \varphi \in L^1_+(0,\infty) \middle| \exists s \geq 0 : \int_0^\infty m(\tau + s)\varphi(\tau)d\tau > 0 \right\},$$

$$\hat{\Omega}_2 = \left\{ \psi \in L^1_+(0,\infty) \middle| \exists s \geq 0 : \int_0^\infty \beta(\tau + s)\psi(\tau)d\tau > 0 \text{ or } \int_0^\infty \eta(\tau + s)\psi(\tau)d\tau > 0 \right\},$$

and

$$\hat{\Omega}_3 = \left\{ \phi \in L^1_+(0,\infty) \middle| \exists s \geq 0 : \int_0^\infty \xi(\theta + s)\phi(\theta)d\theta > 0 \right\}.$$

Define

$$\Omega_0 = \mathbb{R}_+ \times \hat{\Omega}_1 \times \hat{\Omega}_2 \times \hat{\Omega}_3 \times (L^1_+(0,\infty)).$$

Finally, define $X_0 = \Omega \times \Omega_0$. We notice that X_0 is forward invariant. It is not hard to see that Ω is a forward invariant. To see that $\hat{\Omega}_2$ is forward invariant, let us assume that the first inequality holds for the initial condition. The first inequality says that the condition is such that if the support of $\beta(\tau)$ is transferred s units to the right, it will intersect the support of the initial condition. But if that happens for the initial time, it will happen for any other time since the support of the initial condition only moves to the right. Similarly, $\hat{\Omega}_1$, $\hat{\Omega}_3$ are also forward invariant.

We will show that when $\mathcal{R}(\xi) > 1$ the disease persists. Consequently, we identify conditions which lead to the prevalence in individuals. There are many different types of persistence [17]. We identify here the two that we will be working on. We call TB bacteria uniformly weakly persistent if there exists some $\gamma > 0$ independent of the initial conditions such that

$$\limsup_{t\to\infty} \int_0^\infty E(\tau, t)d\tau > \gamma \quad \text{whenever} \quad \int_0^\infty \varphi(\tau)d\tau > 0,$$

$$\limsup_{t\to\infty} \int_0^\infty i(\tau, t)d\tau > \gamma \quad \text{whenever} \quad \int_0^\infty \psi(\tau)d\tau > 0,$$

$$\limsup_{t\to\infty} \int_0^\infty W(\theta, t)d\theta > \gamma \quad \text{whenever} \quad \int_0^\infty \phi(\theta)d\theta > 0.$$

and

$$\limsup_{t\to\infty} \int_0^\infty V(a, t)da > \gamma \quad \text{whenever} \quad \int_0^\infty V_0(a)da > 0,$$

for all solutions of model (2.2). One of the important implications of uniform weak persistence of the disease is that the disease-free equilibrium is unstable. We call TB bacteria uniformly strongly persistent if there exists some $\gamma > 0$ independent of the initial conditions such that

$$\liminf_{t\to\infty}\int_0^\infty E(\tau,t)d\tau > \gamma \quad \text{whenever} \quad \int_0^\infty \varphi(\tau)d\tau > 0,$$

$$\liminf_{t\to\infty}\int_0^\infty i(\tau,t)d\tau > \gamma \quad \text{whenever} \quad \int_0^\infty \psi(\tau)d\tau > 0,$$

$$\liminf_{t\to\infty}\int_0^\infty W(\theta,t)d\theta > \gamma \quad \text{whenever} \quad \int_0^\infty \phi(\theta)d\theta > 0.$$

and

$$\liminf_{t\to\infty}\int_0^\infty V(a,t)da > \gamma \quad \text{whenever} \quad \int_0^\infty V_0(a)da > 0,$$

for all solutions of model (2.2). It is evident from the definitions that, if the disease is uniformly strongly persistent, it is also uniformly weakly persistent. To show uniform strong persistence for TB bacteria, we need to show two components.

1. We have to show that TB bacteria is uniformly weakly persistent.
2. We need to show that the solution semiflow of system (2.2) has a global compact attractor $\mathfrak{T}$.

First, we show uniform weak persistence of TB bacteria. The following proposition states that result.

Proposition 4.1 *Assume $\mathcal{R}(\zeta) > 1$. Then, for all initial conditions that belong to X_0, TB bacteria is uniformly weakly persistent, i.e., there exists $\gamma > 0$ such that*

$$\limsup_{t}\left(\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta\right) \geq \gamma.$$

Proof We argue by contradiction. Assume that TB bacteria dies out. In particular assume that for every $\varepsilon > 0$ and an initial condition in X_0 we have

$$\limsup_{t}\left(\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta\right) < \varepsilon.$$

Hence, there exist $T > 0$ such that for all $t > T$, we have

$$\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta < \varepsilon.$$

By shifting the dynamical system we may assume that the above inequality holds for all $t \geq 0$. From the first equation in (2.2), and taking into account the above inequality, we have

$$S'(t) \geq \Lambda - (\varepsilon + \mu + \zeta)S + \zeta\int_0^t \alpha_2(a)S(t-a)\pi_4(a)da + \int_t^\infty \alpha_2(a)V_0(a-t)\frac{\pi_4(a)}{\pi_4(a-t)}da.$$

Therefore,

$$\limsup_{t\to\infty} S(t) \geq \liminf_{t\to\infty} S(t) \geq \frac{\Lambda}{\varepsilon+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da}.$$

Recall that we are using the following notation $B_1(t) = E(0,t)$, $B_2(t) = i(0,t)$, $B_3(t) = W(0,t)$. Using the inequality above we obtain

$$\begin{cases} B_1(t) \geq \dfrac{\Lambda}{\varepsilon+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da}\left(\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta\right), \\ B_2(t) = \displaystyle\int_0^\infty m(\tau)E(\tau,t)d\tau, \\ B_3(t) = \displaystyle\int_0^\infty \eta(\tau)i(\tau,t)d\tau. \end{cases} \tag{4.2}$$

Now, we apply expression (4.1) to obtain the following system of inequalities in $B_1(t)$, $B_2(t)$ and $B_3(t)$:

$$\begin{cases} B_1(t) \geq \dfrac{\Lambda}{\varepsilon+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da}\left(\displaystyle\int_0^t \beta(\tau)B_2(t-\tau)\pi_2(\tau)d\tau \right. \\ \qquad\qquad\qquad \left. + \displaystyle\int_0^t \xi(\theta)B_3(t-\theta)\pi_3(\theta)d\theta\right), \\ B_2(t) \geq \displaystyle\int_0^t m(\tau)B_1(t-\tau)\pi_1(\tau)d\tau, \\ B_3(t) \geq \displaystyle\int_0^t \eta(\tau)B_2(t-\tau)\pi_2(\tau)d\tau. \end{cases} \tag{4.3}$$

We will take the Laplace transform of both sides of inequalities (4.3). Since all functions above are bounded, their Laplace transform exists for $\lambda > 0$. We denote by $\hat{B}_1(\lambda)$ the Laplace transform of $B_1(t)$, by $\hat{B}_2(\lambda)$ the Laplace transform of $B_2(t)$, and by $\hat{B}_3(\lambda)$ the Laplace transform of $B_3(t)$. Furthermore,

$$\begin{cases} \hat{K}_1(\lambda) = \displaystyle\int_0^\infty m(\tau)\pi_1(\tau)e^{-\lambda\tau}d\tau, \\ \hat{K}_2(\lambda) = \displaystyle\int_0^\infty \beta(\tau)\pi_2(\tau)e^{-\lambda\tau}d\tau, \\ \hat{K}_3(\lambda) = \displaystyle\int_0^\infty \xi(\theta)\pi_3(\theta)e^{-\lambda\theta}d\theta, \\ \hat{K}_4(\lambda) = \displaystyle\int_0^\infty \eta(\tau)\pi_2(\tau)e^{-\lambda\tau}d\tau. \end{cases} \tag{4.4}$$

Taking the Laplace transform of inequalities (4.3) and using the convolution property of the Laplace transform, we obtain the following system of inequalities for $\hat{B}_1(\lambda),\ \hat{B}_2(\lambda),\ \hat{B}_3(\lambda)$.

$$\begin{cases} \hat{B}_1(\lambda) \geq \dfrac{\Lambda}{\varepsilon + \mu + \zeta - \zeta \int_0^\infty \alpha_2(a)\pi_4(a)da}\Big(\hat{K}_2(\lambda)\hat{B}_2(\lambda) + \hat{K}_3(\lambda)\hat{B}_3(\lambda)\Big), \\ \hat{B}_2(\lambda) \geq \hat{K}_1(\lambda)\hat{B}_1(\lambda), \\ \hat{B}_3(\lambda) \geq \hat{K}_4(\lambda)\hat{B}_2(\lambda). \end{cases} \tag{4.5}$$

Eliminating $\hat{B}_2(\lambda)$ and $\hat{B}_3(\lambda)$ from the system above, we obtain

$$\hat{B}_1(\lambda) \geq \frac{\Lambda}{\varepsilon + \mu + \zeta - \zeta \int_0^\infty \alpha_2(a)\pi_4(a)da}\Big(\hat{K}_1(\lambda)\hat{K}_2(\lambda) + \hat{K}_1(\lambda)\hat{K}_3(\lambda)\hat{K}_4(\lambda)\Big)\hat{B}_1(\lambda).$$

The inequality should hold for the given $\varepsilon \approx 0$ and for any $\lambda > 0$. But this is impossible since for $\varepsilon \approx 0$ and $\lambda \approx 0$, the coefficient of $\hat{B}(\lambda)$ on the right hand side is approximately $\mathcal{R}(\zeta) > 1$, i.e.,

$$\frac{\Lambda}{\varepsilon + \mu + \zeta - \zeta \int_0^\infty \alpha_2(a)\pi_4(a)da}\Big(\hat{K}_1(\lambda)\hat{K}_2(\lambda) + \hat{K}_1(\lambda)\hat{K}_3(\lambda)\hat{K}_4(\lambda)\Big) \approx \mathcal{R}(\zeta) > 1.$$

This contradicts our assumption that

$$\limsup_t \left(\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta\right) < \varepsilon.$$

Therefore, there exists $\varepsilon_0 > 0$ such that for any initial condition in X_0, we have

$$\limsup_t \left(\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta\right) > \varepsilon_0.$$

It remains to be seen that each of the components is bounded below. The inequality above implies that

$$\limsup_{t\to\infty} B_1(t) \geq \varepsilon_0 \frac{\Lambda}{\varepsilon + \mu + \zeta - \zeta \int_0^\infty \alpha_2(a)\pi_4(a)da},$$

and

$$\begin{aligned}\limsup_{t\to\infty} B_2(t) &\geq \limsup_{t\to\infty}\int_0^\infty m(\tau)E(\tau,t)d\tau\\ &\geq \frac{\Lambda\varepsilon_0}{\varepsilon+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da}\int_0^\infty m(\tau)\pi_1(\tau)d\tau \geq \gamma_1,\\ \limsup_{t\to\infty} B_3(t) &\geq \limsup_{t\to\infty}\int_0^\infty \eta(\tau)i(\tau,t)d\tau \geq \gamma_1\int_0^\infty \eta(\tau)\pi_2(\tau)d\tau \geq \gamma_2.\end{aligned}$$

Hence,

$$\limsup_{t\to\infty}\int_0^\infty \beta(\tau)i(\tau,t)d\tau \geq \gamma_1\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau \geq \gamma_3.$$

Similarly,

$$\limsup_{t\to\infty}\int_0^\infty \xi(\theta)W(\theta,t)d\theta > \gamma_2\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta \geq \gamma_4.$$

In addition,

$$\int_0^\infty V(\theta,t)d\theta = \zeta\int_0^a S(t-a)\pi_4(a)da + \int_t^\infty V_0(a-t)\frac{\pi_4(a)}{\pi_4(a-t)}da.$$

Thus we have

$$\begin{aligned}\limsup_{t\to\infty}\int_0^\infty V(a,t)da &\geq \liminf_{t\to\infty}\int_0^\infty V(a,t)da\\ &\geq \frac{\zeta\Lambda}{\varepsilon+\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da}\int_0^\infty \pi_4(a)d\theta \geq \gamma_5.\end{aligned}$$

This completes the proof. □

Our next goal is to prove that system (2.2) has a global compact attractor $\mathfrak{T}$. As a first step, we define the semiflow Ψ of the solutions of system (2.2)

$$\Psi\bigg(t: S_0, \varphi(\cdot), \psi(\cdot), \phi(\cdot),\ V_0(\cdot)\bigg) = \bigg(S(t), E(\tau,t), i(\tau,t), W(\theta,t), V(a,t)\bigg).$$

The semiflow is a mapping $\Psi : [0,\infty)\times X_0 \to X_0$. A set $\mathfrak{T}$ in X_0 is called a *global compact attractor* for Ψ, if $\mathfrak{T}$ is a maximal compact invariant set and if for all open sets $\mathfrak{U}$ containing $\mathfrak{T}$ and all bounded sets $\mathcal{B}$ of X_0 there exists some $T > 0$ such that $\Psi(t,\mathcal{B}) \subseteq \mathfrak{U}$, for all $t > T$. The following proposition establishes the presence of a global compact attractor.

Proposition 4.2 *Assume $\mathcal{R}(\zeta) > 1$. Then, there exists $\mathfrak{T}$, a compact subset of X_0, which is a global attractor for the solution semiflow Ψ of (2.2) in X_0. Moreover, $\mathfrak{T}$ is* *invariant under the solution semiflow, i.e.,*

$$\Psi(t, x_0) \subseteq \mathfrak{T} \quad \textit{for every} \quad x_0 \in \mathfrak{T},\ \forall t \geq 0.$$

Proof To establish this result, we will apply Lemma 3.1.3 and Theorem 3.4.6 in [18]. To show the assumptions of Lemma 3.1.3 and Theorem 3.4.6 in [18], we split the solution semiflow into two components. For an initial condition $x_0 \in X_0$ we have $\Psi(t, x_0) = \hat{\Psi}(t, x_0) + \tilde{\Psi}(t, x_0)$. The splitting is done in such a way that $\hat{\Psi}(t, x_0) \to 0$ as $t \to \infty$ for every $x_0 \in X_0$, and for a fixed t and any bounded set $\mathcal{B}$ in X_0, the set $\{\tilde{\Psi}(t, x_0) : x_0 \in \mathcal{B}\}$ is pre-compact. The two components of the semiflow are defined as follows:

$$\hat{\Psi}\Big(t : S_0, \varphi(\cdot), \psi(\cdot), \phi(\cdot),\ V_0(\cdot)\Big) = \Big(0, \hat{E}(\tau, t), \hat{i}(\tau, t), \hat{W}(\theta, t), \hat{V}(a, t)\Big),$$

$$\tilde{\Psi}\Big(t : S_0, \varphi(\cdot), \psi(\cdot), \phi(\cdot),\ V_0(\cdot)\Big) = \Big(S(t), \tilde{E}(\tau, t), \tilde{i}(\tau, t), \tilde{W}(\theta, t), \tilde{V}(a, t)\Big), \tag{4.6}$$

where

$$\begin{cases} E(t, X^0) = \hat{E}(t, X^0) + \tilde{E}(t, X^0), \\ i(t, X^0) = \hat{i}(t, X^0) + \tilde{i}(t, X^0), \\ W(t, X^0) = \hat{W}(t, X^0) + \tilde{W}(t, X^0), \\ V(t, X^0) = \hat{V}(t, X^0) + \tilde{V}(t, X^0), \end{cases}$$

and $\hat{E}(\tau, t), \hat{i}(\tau, t), \hat{W}(\theta, t), \hat{V}(a, t), \tilde{E}(\tau, t), \tilde{i}(\tau, t), \tilde{W}(\theta, t), \tilde{V}(a, t)$ are the solutions of the following equations (the remaining equations are as in system (2.1))

$$\begin{cases} \dfrac{\partial \hat{E}}{\partial t} + \dfrac{\partial \hat{E}}{\partial \tau} = -(\mu + m(\tau))\hat{E}(\tau, t), \\ \hat{E}(0, t) = 0, \\ \hat{E}(\tau, 0) = \varphi(\tau), \end{cases}$$

$$\begin{cases} \dfrac{\partial \hat{i}}{\partial t} + \dfrac{\partial \hat{i}}{\partial \tau} = -(\mu + \alpha_1(\tau) + r_1(\tau))\hat{i}(\tau, t), \\ \hat{i}(0, t) = 0, \\ \hat{i}(\tau, 0) = \psi(\tau), \end{cases}$$

$$\begin{cases} \dfrac{\partial \hat{W}}{\partial t} + \dfrac{\partial \hat{W}}{\partial \theta} = -\delta(\theta)\hat{W}(\theta, t), \\ \hat{W}(0, t) = 0, \\ \hat{W}(\theta, 0) = \phi(\theta), \end{cases}$$

$$\begin{cases} \dfrac{\partial \hat{V}}{\partial t} + \dfrac{\partial \hat{V}}{\partial a} = -(\mu + \alpha_2(a) + r_2(a))\hat{V}(a, t), \\ \hat{V}(0, t) = 0, \\ \hat{V}(a, 0) = \hat{V}_0(a). \end{cases} \tag{4.7}$$

And

$$\begin{cases} \dfrac{\partial \tilde{E}}{\partial t} + \dfrac{\partial \tilde{E}}{\partial \tau} = -(\mu + m(\tau))\tilde{E}(\tau, t), \\ \tilde{E}(0, t) = S\left(\displaystyle\int_0^\infty \beta(\tau)\tilde{i}(\tau, t)d\tau + \int_0^\infty \xi(\theta)\tilde{W}(\theta, t)d\theta\right), \\ \tilde{E}(\tau, 0) = 0, \end{cases}$$

$$\begin{cases} \dfrac{\partial \tilde{i}}{\partial t} + \dfrac{\partial \tilde{i}}{\partial \tau} = -(\mu + \alpha_1(\tau) + r_1(\tau))\hat{i}(\tau, t), \\ \tilde{i}(0, t) = \displaystyle\int_0^\infty m(\tau)\tilde{E}(\tau, t)d\tau, \\ \tilde{i}(\tau, 0) = 0, \end{cases}$$

$$\begin{cases} \dfrac{\partial \tilde{W}}{\partial t} + \dfrac{\partial \tilde{W}}{\partial \theta} = -\delta(\theta)\tilde{W}(\theta, t), \\ \tilde{W}(0, t) = \displaystyle\int_0^\infty \eta((\tau))\tilde{i}(\tau, t)d\tau, \\ \tilde{W}(\theta, 0) = 0, \end{cases}$$

$$\begin{cases} \dfrac{\partial \tilde{V}}{\partial t} + \dfrac{\partial \tilde{V}}{\partial a} = -(\mu + \alpha_2(a) + r_2(a))\tilde{V}(a, t), \\ \tilde{V}(0, t) = \zeta S, \\ \tilde{V}(a, 0) = 0. \end{cases} \tag{4.8}$$

System (4.7) is decoupled from the remaining equations. Using the formula (4.1) to integrate along the characteristic lines, we obtain

$$\hat{E}(\tau,t)=\begin{cases}0, & t>\tau,\\ \varphi(\tau-t)\dfrac{\pi_1(\tau)}{\pi_1(\tau-t)}, & t<\tau,\end{cases}$$
$$\hat{i}(\tau,t)=\begin{cases}0, & t>\tau,\\ \psi(\tau-t)\dfrac{\pi_2(\tau)}{\pi_2(\tau-t)}, & t<\tau,\end{cases}$$
$$\hat{W}(\theta,t)=\begin{cases}0, & t>\theta,\\ \phi(\theta-t)\dfrac{\pi_3(\theta)}{\pi_3(\theta-t)}, & t<\theta,\end{cases} \tag{4.9}$$
$$\hat{V}(a,t)=\begin{cases}0, & t>a,\\ \hat{V}_0(a-t)\dfrac{\pi_4(a)}{\pi_4(a-t)}, & t<a.\end{cases}$$

Integrating $\hat{E}$ with respect to τ, we obtain:

$$\int_t^\infty \varphi(\tau-t)\frac{\pi_1(\tau)}{\pi_1(\tau-t)}d\tau=\int_0^\infty \varphi(\tau)\frac{\pi_1(t+\tau)}{\pi_1(\tau)}d\tau\le e^{-\mu t}\int_0^\infty \varphi(\tau)d\tau\to 0,$$

as $t\to\infty$. Similarly,

$$\int_t^\infty \psi(\tau-t)\frac{\pi_2(\tau)}{\pi_2(\tau-t)}d\tau=\int_0^\infty \psi(\tau)\frac{\pi_2(t+\tau)}{\pi_2(\tau)}d\tau\le e^{-\mu t}\int_0^\infty \psi(\tau)d\tau\to 0,$$

$$\int_t^\infty \phi(\theta-t)\frac{\pi_3(\theta)}{\pi_3(\theta-t)}d\theta=\int_0^\infty \phi(\theta)\frac{\pi_3(t+\theta)}{\pi_3(\theta)}d\theta\le e^{-\mu t}\int_0^\infty \phi(\theta)d\theta\to 0,$$

$$\int_t^\infty \hat{V}_0(a-t)\frac{\pi_4(a)}{\pi_4(a-t)}da=\int_0^\infty \hat{V}_0(a)\frac{\pi_4(t+a)}{\pi_4(a)}da\le e^{-\mu t}\int_0^\infty \hat{V}_0(a)da\to 0.$$

This shows the first claim, i.e., $\hat{\Psi}(t,x_0)\to 0$ as $t\to\infty$ uniformly for every $x_0\in\mathcal{B}\subseteq X_0$, where $\mathcal{B}$ is a ball of a given radius.

To show the second claim, we need to show compactness. We fix t and let $x_0\in X_0$. Note that X_0 is bounded. We have to show that for that fixed t the family of functions defined by

$$\tilde{\Psi}(t,x_0)=\Big(S(t),\tilde{E}(\tau,t),\tilde{i}(\tau,t),\tilde{W}(\theta,t),\tilde{V}(a,t)\Big)$$

obtained by taking different initial conditions in X_0 is a compact family of functions. The family

$$\{\tilde{\Psi}(t, x_0)|x_0 \in X_0, t - \text{fixed}\} \subseteq X_0,$$

and, therefore, it is bounded. Thus, we have established the boundedness of the set. To show compactness we first see that the third condition in the Frechét-Kolmogorov Theorem [19] for compactness in L^1 is trivially satisfied since $\tilde{E}(\tau, t) = 0$, $\tilde{i}(\tau, t) = 0$, $\tilde{W}(\theta, t) = 0$, $\tilde{V}(a, t) = 0$ for $\min\{\theta, \tau, a\} > t$. To see the second condition of that Theorem, we have to find the bounds for the L^1-norm of $\frac{\partial E}{\partial \tau}$, $\frac{\partial i}{\partial \tau}$, $\frac{\partial W}{\partial \theta}$, $\frac{\partial V}{\partial a}$. To derive that bound, first notice that

$$\begin{aligned}
\tilde{E}(\tau, t) &= \begin{cases} \tilde{B}_1(t-\tau)\pi_1(\tau), & t > \tau, \\ 0, & t < \tau. \end{cases} \\
\tilde{i}(\tau, t) &= \begin{cases} \tilde{B}_2(t-\tau)\pi_2(\tau), & t > \tau, \\ 0, & t < \tau. \end{cases} \\
\tilde{W}(\theta, t) &= \begin{cases} \tilde{B}_3(t-\theta)\pi_3(\theta), & t > \theta, \\ 0, & t < \theta. \end{cases} \\
\tilde{V}(a, t) &= \begin{cases} \zeta S(t-a)\pi_4(a), & t > a, \\ 0, & t < a \end{cases}
\end{aligned} \tag{4.10}$$

where

$$\begin{aligned}
\tilde{B}_1(t) &= S(t)\int_0^\infty \beta(\tau)\tilde{i}(\tau, t)d\tau + S(t)\int_0^\infty \xi(\theta)\tilde{W}(\theta, t)d\theta \\
&= S(t)\int_0^t \beta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau + S(t)\int_0^t \xi(\theta)\tilde{B}_3(t-\theta)\pi_3(\theta)d\theta, \\
\tilde{B}_2(t) &= \int_0^\infty m(\tau)\tilde{E}(\tau, t)d\tau = \int_0^t m(\tau)\tilde{B}_1(t-\tau)\pi_1(\tau)d\tau, \\
\tilde{B}_3(t) &= \int_0^\infty \eta(\tau)\tilde{i}(\tau, t)d\tau = \int_0^t \eta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau.
\end{aligned} \tag{4.11}$$

First, we notice that for $x_0 \in X_0$, $\tilde{B}_1(t)$ is bounded. We can see that by recalling that S is bounded. Hence, the $\tilde{B}_1(t)$ satisfies the following inequality

$$\begin{aligned}
\tilde{B}_1(t) &= S(t)\int_0^t \beta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau + S(t)\int_0^t \xi(\theta)\tilde{B}_3(t-\theta)\pi_3(\theta)d\theta \\
&\leq k_1'\int_0^t \tilde{B}_2(t-\tau)d\tau + k_1''\int_0^t \tilde{B}_3(t-\theta)d\theta \\
&= k_1'\int_0^t \tilde{B}_2(\tau)d\tau + k_1''\int_0^t \tilde{B}_3(\theta)d\theta \\
&= k_1'\int_0^t\int_0^\tau m(\sigma)\tilde{B}_1(\tau-\sigma)\pi_1(\sigma)d\sigma d\tau \\
&\quad + k_1''\int_0^t\int_0^\theta \eta(\tau)\pi_2(\tau)\int_0^{\theta-\tau} m(\sigma)\tilde{B}_1(\theta-\tau-\sigma)\pi_1(\sigma)d\sigma d\tau d\theta \\
&= k_1'\int_0^t m(\sigma)\pi_1(\sigma)\int_\sigma^t \tilde{B}_1(\tau-\sigma)d\tau d\sigma \\
&\quad + k_1''\int_0^t m(\sigma)\pi_1(\sigma)\int_0^{t-\sigma}\eta(\tau)\pi_2(\tau)\int_{\tau+\sigma}^t \tilde{B}_1(\theta-\tau-\sigma)d\theta d\tau d\sigma \qquad (4.12) \\
&= k_1'\int_0^t m(\sigma)\pi_1(\sigma)\int_0^{t-\sigma}\tilde{B}_1(\tau)d\tau d\sigma \\
&\quad + k_1''\int_0^t m(\sigma)\pi_1(\sigma)\int_0^{t-\sigma}\eta(\tau)\pi_2(\tau)\int_0^{t-\tau-\sigma}\tilde{B}_1(\theta)d\theta d\tau d\sigma \\
&\leq k_1'\int_0^t m(\sigma)\pi_1(\sigma)\int_0^t \tilde{B}_1(\tau)d\tau d\sigma \\
&\quad + k_1''\int_0^t m(\sigma)\pi_1(\sigma)\int_0^{t-\sigma}\eta(\tau)\pi_2(\tau)\int_0^t \tilde{B}_1(\theta)d\theta d\tau d\sigma \\
&\leq k_1'''\int_0^t \tilde{B}_1(\tau)d\tau + k_1''''\int_0^t \tilde{B}_1(\theta)d\theta \\
&= k_1\int_0^t \tilde{B}_1(\tau)d\tau,
\end{aligned}$$

where k_1', k_1'', k_1''', k_1'''' and k_1 are constants that depend on the bounds of the parameters as well as the bounds of the solution. Gronwall's inequality implies that

$$\tilde{B}_1(t) \leq \tilde{B}_1(0)e^{k_1 t}.$$

In the following, we derive that for $x_0 \in X_0$ the sets $\tilde{B}_2(t)$, $\tilde{B}_3(t)$ are also bounded.

$$\begin{aligned}
\tilde{B}_2(t) &= \int_0^\infty m(\tau)\tilde{E}(\tau,t)d\tau = \int_0^t m(\tau)\tilde{B}_1(t-\tau)\pi_1(\tau)d\tau \\
&\le K_2' \int_0^t \tilde{B}_1(t-\tau)d\tau = K_2' \int_0^t \tilde{B}_1(\tau)d\tau \\
&\le K_2' \int_0^t \tilde{B}_1(0)e^{k_1\tau}d\tau = \frac{K_2'\tilde{B}_1(0)}{k_1}(e^{k_1 t}-1) \\
&\le k_2 e^{k_1 t}, \\
\tilde{B}_3(t) &= \int_0^\infty \eta(\tau)\tilde{i}(\tau,t)d\tau = \int_0^t \eta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau \\
&\le K_3' \int_0^t \tilde{B}_2(t-\tau)d\tau = K_3' \int_0^t \tilde{B}_2(\tau)d\tau \\
&\le K_3' \int_0^t k_2 e^{k_1\tau}d\tau = \frac{K_3' k_2}{k_1}(e^{k_1 t}-1) \\
&\le k_3 e^{k_1 t}.
\end{aligned}$$

Next, we differentiate (4.10) with respect to (τ, θ, a):

$$\left|\frac{\partial \tilde{E}(\tau,t)}{\partial \tau}\right| \le \begin{cases} |(\tilde{B}_1(t-\tau))'\pi_1(\tau) + \tilde{B}_1(t-\tau)(\pi_1(\tau))'|, & t > \tau, \\ 0, & t < \tau. \end{cases}$$

$$\left|\frac{\partial \tilde{i}(\tau,t)}{\partial \tau}\right| \le \begin{cases} |(\tilde{B}_2(t-\tau))'\pi_2(\tau) + \tilde{B}_2(t-\tau)(\pi_2(\tau))'|, & t > \tau, \\ 0, & t < \tau. \end{cases}$$

$$\left|\frac{\partial \tilde{W}(\theta,t)}{\partial \theta}\right| \le \begin{cases} |(\tilde{B}_3(t-\theta))'\pi_3(\theta) + \tilde{B}_3(t-\theta)(\pi_3(\theta))'|, & t > \theta, \\ 0, & t < \theta, \end{cases}$$

$$\left|\frac{\partial \tilde{V}(a,t)}{\partial a}\right| \le \begin{cases} \zeta|(S'(t-a))\pi_4(a) + S(t-a)(\pi_4(a))'|, & t > a, \\ 0, & t < a. \end{cases}$$

We have to see that $|(\tilde{B}_1(t-\tau))'|$, $|(\tilde{B}_2(t-\tau))'|$ and $|(\tilde{B}_3(t-\theta))'|$ are bounded. Differentiating (4.11), we obtain

$$\begin{cases} (\tilde{B}_1(t))' = S' \int_0^t \beta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau + S\beta(t)\tilde{B}_2(0)\pi_2(t) \\ \qquad +S \int_0^t \beta(\tau)(\tilde{B}_2(t-\tau))'\pi_2(\tau)d\tau + S'(t)\int_0^t \xi(\theta)\tilde{B}_3(t-\theta)\pi_3(\theta)d\theta \\ \qquad +S\xi(t)\tilde{B}_3(0)\pi_3(t) + S\int_0^t \xi(\theta)(\tilde{B}_3(t-\theta))'\pi_3(\theta)d\theta, \\ (\tilde{B}_2(t))' = m(t)\tilde{B}_1(0)\pi_1(t) + \int_0^t m(\tau)(\tilde{B}_1(t-\tau))'\pi_1(\tau)d\tau, \\ (\tilde{B}_3(t))' = \eta(t)\tilde{B}_2(0)\pi_2(t) + \int_0^t \eta(\tau)(\tilde{B}_2(t-\tau))'\pi_2(\tau)d\tau. \end{cases}$$

Taking an absolute value and bounding all terms, we can rewrite the above equality as the following inequality:

$$\begin{aligned}(\tilde{B}_1(t))' &\le k_4' \int_0^t \tilde{B}_2(t-\tau)d\tau + k_4'' \int_0^t \tilde{B}_3(t-\theta)d\theta + k_5 \\ &= k_4' \int_0^t \tilde{B}_2(\tau)d\tau + k_4'' \int_0^t \tilde{B}_3(\theta)d\theta + k_5 \\ &\le k_4 \int_0^t \tilde{B}_1(\tau)d\tau + k_5.\end{aligned}$$

The technique here is similar to the one used for $\tilde{B}(t)$ in (4.12). Gronwall's inequality then implies that,

$$|(\tilde{B}_1(t))'| \le k_5 e^{k_4 t}.$$

It is evident that

$$|(\tilde{B}_2(t))'| \le k_6, \qquad |(\tilde{B}_3(t))'| \le k_7.$$

Putting all these bounds together, we have

$$\begin{aligned}\parallel \partial_\tau \tilde{E} \parallel &\le k_5 e^{k_4 t} \int_0^\infty \pi_1(\tau)d\tau + \tilde{B}_1(0)e^{k_1 t}(\mu + \bar{m}) \int_0^\infty \pi_1(\tau)d\tau \\ &< k_8, \\ \parallel \partial_\tau \tilde{i} \parallel &\le k_6 \int_0^\infty \pi_2(\tau)d\tau + k_2 e^{k_1 t}(\mu + \bar{\alpha}_1 + \bar{r}_1) \int_0^\infty \pi_2(\tau)d\tau \\ &< k_9, \\ \parallel \partial_\theta \tilde{W} \parallel &\le k_7 \int_0^\infty \pi_3(\theta)d\theta + k_3 e^{k_1 t}\bar{\delta} \int_0^\infty \pi_3(\theta)d\theta \\ &< k_{10}, \\ \parallel \partial_a \tilde{V} \parallel &\le k_{11} \int_0^\infty \pi_4(a)da + k_{12}(\mu + \bar{\alpha}_2 + \bar{r}_2) \int_0^\infty \pi_4(a)da \\ &< k^0,\end{aligned}$$

where $\bar{r}_1 = \sup_\tau\{r_1(\tau)\}$, $\bar{r}_2 = \sup_a\{r_2(a)\}$, $\bar{\alpha}_1 = \sup_\tau\{\alpha_1(\tau)\}$, $\bar{\alpha}_2 = \sup_a\{\alpha_2(a)\}$.

To complete the proof, we notice that

$$\int_0^\infty |\tilde{E}(\tau+h,t)-\tilde{E}(\tau,t)|d\tau \leq \| \partial_\tau \tilde{E} \| \, |h| \leq k_8|h|,$$
$$\int_0^\infty |\tilde{i}(\tau+h,t)-\tilde{i}(\tau,t)|d\tau \leq \| \partial_\tau \tilde{i} \| \, |h| \leq k_9|h|,$$
$$\int_0^\infty |\tilde{W}(\theta+h,t)-\tilde{W}(\theta,t)|d\theta \leq \| \partial_\theta \tilde{W} \| \, |h| \leq k_{10}|h|,$$
$$\int_0^\infty |\tilde{V}(a+h,t)-\tilde{V}(a,t)|da \leq \| \partial_a \tilde{V} \| \, |h| \leq k^0|h|.$$

Thus, the integral can be made arbitrary small uniformly in the family of functions. That establishes the second requirement of the Frechét-Kolmogorov Theorem. We conclude that the family is compact. □

Now we have all components to establish the uniform strong persistence. The next proposition states the uniform strong persistence of E, i, W and V.

Proposition 4.3 *Assume $\mathcal{R}(\zeta) > 1$. Then, for all initial conditions that belong to X_0, TB bacteria persists, i.e., there exists $\gamma > 0$ such that*

$$\liminf_t \left(\int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta \right) \geq \gamma.$$

Proof We apply Theorem 2.6 introduced in [20]. We consider the solution semiflow Ψ on X_0. We define a functional $\rho : X_0 \rightarrow \mathrm{R}_+$ as follows:

$$\rho(\Psi(t,x_0)) = \int_0^\infty \beta(\tau)\tilde{i}(\tau,t)d\tau + \int_0^\infty \xi(\theta)\tilde{W}(\theta,t)d\theta.$$

Proposition 4.1 implies that the semiflow is uniformly weakly ρ-persistent. Proposition 4.2 shows that the solution semiflow has a global compact attractor $\mathfrak{T}$. Total orbits are solutions to the system (2.2) defined for all times $t \in \mathbb{R}$. Since the solution semiflow is nonnegative, we have that for any s and any $t > s$

$$\begin{aligned}
&\int_0^\infty \beta(\tau)\tilde{i}(\tau,t)d\tau + \int_0^\infty \xi(\theta)\tilde{W}(\theta,t)d\theta \\
&= \int_0^t \beta(\tau)\tilde{B}_2(t-\tau)\pi_2(\tau)d\tau + \int_0^t \xi(\theta)\tilde{B}_3(t-\theta)\pi_3(\theta)d\theta \\
&\geq \mathfrak{b}_1 \int_0^t \tilde{B}_2(t-\tau)d\tau + \mathfrak{b}_2 \int_0^t \tilde{B}_3(t-\theta)d\theta \\
&= \mathfrak{b}_1 \int_0^t \tilde{B}_2(\tau)d\tau + \mathfrak{b}_2 \int_0^t \tilde{B}_3(\theta)d\theta \\
&= \mathfrak{b}_1 \int_0^t \int_0^\tau m(\sigma)\tilde{B}_1(\tau-\sigma)\pi_1(\sigma)d\sigma d\tau \\
&\quad + \mathfrak{b}_2 \int_0^t \int_0^\theta \eta(\tau)\pi_2(\tau) \int_0^{\theta-\tau} m(\sigma)\tilde{B}_1(\theta-\tau-\sigma)\pi_1(\sigma)d\sigma d\tau d\theta \\
&= \mathfrak{b}_1 \int_0^t m(\sigma)\pi_1(\sigma) \int_\sigma^t \tilde{B}_1(\tau-\sigma)d\tau d\sigma \\
&\quad + \mathfrak{b}_2 \int_0^t m(\sigma)\pi_1(\sigma) \int_0^{t-\sigma} \eta(\tau)\pi_2(\tau) \int_{\tau+\sigma}^t \tilde{B}_1(\theta-\tau-\sigma)d\theta d\tau d\sigma \\
&= \mathfrak{b}_1 \int_0^t m(\sigma)\pi_1(\sigma) \int_0^{t-\sigma} \tilde{B}_1(\tau)d\tau d\sigma \\
&\quad + \mathfrak{b}_2 \int_0^t m(\sigma)\pi_1(\sigma) \int_0^{t-\sigma} \eta(\tau)\pi_2(\tau) \int_0^{t-\tau-\sigma} \tilde{B}_1(\theta)d\theta d\tau d\sigma \\
&\geq \mathfrak{b}_3 \int_0^t \int_0^{t-\sigma} \tilde{B}_1(\tau)d\tau d\sigma + \mathfrak{b}_4 \int_0^t \int_0^{t-\sigma} \int_0^{t-\tau-\sigma} \tilde{B}_1(\theta)d\theta d\tau d\sigma \\
&\geq \left(\frac{1}{2}\mathfrak{b}_3 t^2 + \frac{1}{6}\mathfrak{b}_4 t^3\right)\tilde{B}_1(s)
\end{aligned}$$

Therefore, $\int_0^\infty \beta(\tau)\tilde{i}(\tau,t)d\tau + \int_0^\infty \xi(\theta)\tilde{W}(\theta,t)d\theta > 0$ for all $t > s$, provided $\tilde{B}_1(s) > 0$. Theorem 2.6 in [20] now implies that the semiflow is uniformly strongly ρ-persistent. Hence, there exists ς such that

$$\liminf_{t\to\infty} \left(\int_0^\infty \beta(\tau)\tilde{i}(\tau,t)d\tau + \int_0^\infty \xi(\theta)\tilde{W}(\theta,t)d\theta \right) \geq \varsigma.$$

Corollary 4.1 *Assume $\mathcal{R}(\zeta) > 1$. There exists constants $\varsigma > 0$ and $M > 0$ such that for each orbit $(S(t), E(\tau,t), i(\tau,t), W(\theta,t), V(a,t)$ of Ψ in $\mathfrak{T}$, we have*

$$\varsigma \leq \int_0^\infty \beta(\tau)i(\tau,t)d\tau + \int_0^\infty \xi(\theta)W(\theta,t)d\theta \leq M,$$

and

$$\varsigma \leq \int_0^\infty m(\tau)E(\tau,t)d\tau \leq M, \quad \varsigma \leq \int_0^\infty \eta(\tau)i(\tau,t)d\tau \leq M, \quad \forall t \in \mathrm{R}.$$

Now, we will emphasize that through the change of variables given above, the general form of system (2.2) is equivalent to the special case. For $t \geq 0$, $r \in \mathbb{R}$, we have (see [7, 21])

$$\begin{aligned}&\int_0^\infty \alpha_2(a)V(t+r)(a)da\\ &= \zeta \int_0^t \alpha_2(a)S(t-a+r)\pi_4(a)da + \int_t^\infty \alpha_2(a)V_0(r)(a-t)\frac{\pi_4(a)}{\pi_4(a-t)}da.\end{aligned}$$

Setting $\hat{t} = t + r$, it follows that for $t \geq r$,

$$\begin{aligned}&\int_0^\infty \alpha_2(a)V(t)(a)d\alpha\\ &= \zeta \int_0^{t-r} \alpha_2(a)S(t-a)\pi_4(a)da + \int_{t-r}^\infty \alpha_2(a)V_0(r)(a-(t-r))\frac{\pi_4(a)}{\pi_4(a-(t-r))}da,\end{aligned}$$

and for $t \geq r$,

$$\left| \int_{t-r}^\infty \alpha_2(a)V_0(r)(a-(t-r))\frac{\pi_4(a)}{\pi_4(a-(t-r))}da \right| \leq e^{-\mu(t-r)} \parallel V_0(r) \parallel_{L^1(0,+\infty)} .$$

It follows that (as $r \to -\infty$) for $t \in \mathbb{R}$.

$$\int_0^\infty \alpha_2(a)V(a,t)d\theta = \int_0^\infty \alpha_2(a)V(t)(a)da = \zeta \int_0^\infty \alpha_2(a)S(t-a)\pi_4(a)da.$$

The system (2.2) becomes

$$\begin{cases} \dfrac{dS}{dt} = \Lambda - S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau - S\int_0^\infty \xi(\theta)W(\theta,t)d\theta - (\mu+\zeta)S \\ \qquad +\zeta \displaystyle\int_0^\infty \alpha_2(a)\pi_4(a)S(t-a)da, \\ \dfrac{\partial E(\tau,t)}{\partial \tau} + \dfrac{\partial E(\tau,t)}{\partial t} = -(\mu + m(\tau))E(\tau,t), \\ E(0,t) = S\displaystyle\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S\int_0^\infty \xi(\theta)W(\theta,t)d\theta, \\ \dfrac{\partial i(\tau,t)}{\partial \tau} + \dfrac{\partial i(\tau,t)}{\partial t} = -(\mu + \alpha_1(\tau) + r_1(\tau))i(\tau,t), \\ i(0,t) = \displaystyle\int_0^\infty m(\tau)E(\tau,t)d\tau, \\ \dfrac{\partial W(\theta,t)}{\partial \theta} + \dfrac{\partial W(\theta,t)}{\partial t} = -\delta(\theta)W(\theta,t), \\ W(0,t) = \displaystyle\int_0^\infty \eta(\tau)i(\tau,t)d\tau. \end{cases} \tag{4.13}$$

Once system (4.13) is solved, we can use (4.1) to obtain $V(a,t)$. In the sequel, it is system (4.13) to be considered. Once the stability of system (4.13) is obtained, it then gives the stability of system (2.2)

5 Global Stability of the Disease-Free Equilibrium

In the previous section, we have established that equilibria are locally stable, i.e, given the conditions on the parameters, if the initial conditions are close enough to the equilibrium, the solution will converge to that equilibrium. In this section our objective is to extend these results to global stability results. That is, given the conditions on the parameters, convergence to the equilibrium occurs independent of the initial conditions.

As a first step, we establish the global stability of the disease-free equilibrium. We will use a Lyapunov function to approach the problem.

Theorem 5.1 *Assume $\mathcal{R}(\zeta) \leq 1$. Then the disease-free equilibrium $\mathcal{E}_0$ is globally asymptotically stable.*

Proof We will use a Lyapunov function. We adopt the logistic function used in [22, 23]. Define

$$f(x) = x - 1 - \ln x.$$

We note that $f(x) \geq 0$ for all $x > 0$. $f(x)$ achieves its global minimum at one, with $f(1) = 0$. Let

$$\begin{cases} g(\theta) = \displaystyle\int_\theta^\infty \xi(s) e^{-\int_\theta^s \delta(\sigma)d\sigma} ds, \\ q(\tau) = \displaystyle\int_\tau^\infty \Big(\beta(s) + g(0)\eta(s)\Big) e^{-\int_\tau^s (\mu+\alpha_1(\sigma)+r_1(\sigma))d\sigma} ds, \\ p(\tau) = q(0) \displaystyle\int_\tau^\infty m(s) e^{-\int_\tau^s (\mu+m(\sigma))d\sigma} ds. \end{cases} \tag{5.1}$$

Noticing that

$$p(0) = \frac{\mathcal{R}(\zeta)}{S_0^*}.$$

Differentiating (5.1) first, we obtain

$$\begin{cases} g'(\theta) = -\xi(\theta) + \delta(\theta)g(\theta), \\ q'(\tau) = -\Big(\beta(\tau) + g(0)\eta(\tau)\Big) + \Big(\mu + \alpha_1(\tau) + r_1(\tau)\Big)q(\tau), \\ p_j'(\tau) = -q(0)m(\tau) + \Big(\mu + m(\tau)\Big)p(\tau). \end{cases} \tag{5.2}$$

We define the following Lyapunov function:

$$U_1(t) = U_s(t) + U_e(t) + U_{ii}(t) + U_w(t) + U_v(t)$$

where

$$\begin{cases} U_s(t) &= S_0^* f\left(\dfrac{S}{S_0^*}\right), \\ U_e(t) &= S_0^* \displaystyle\int_0^\infty p(\tau) E(\tau, t) d\tau, \\ U_{ii}(t) &= S_0^* \displaystyle\int_0^\infty q(\tau) i(\tau, t) d\tau, \\ U_w(t) &= S_0^* \displaystyle\int_0^\infty g(\theta) W(\theta, t) d\theta, \\ U_v(t) &= \zeta S_0^* \displaystyle\int_0^\infty \alpha_2(a) \pi_4(a) \int_0^a f\left(\dfrac{S(t-\sigma)}{S_0^*}\right) d\sigma da. \end{cases}$$

Because of the complexity of the expressions, we take the derivative of each component of the Lyapunov function separately

$$\begin{aligned} U_s'(t) &= S_0^* \left(1 - \frac{S_0^*}{S}\right) \frac{1}{S_0^*} S' \\ &= (1 - \frac{S_0^*}{S}) \left[\Lambda - E(0, t) - (\mu + \zeta) S + \zeta \int_0^\infty \alpha_2(a) S(t-a) \pi_4(a) da \right] \\ &= (1 - \frac{S_0^*}{S}) \left[(\mu + \zeta) S_0^* - \zeta S_0^* \int_0^\infty \alpha_2(a) \pi_4(a) da - E(0, t) \right. \\ &\quad \left. -(\mu + \zeta) S + \zeta \int_0^\infty \alpha_2(a) S(t-a) \pi_4(a) da \right] \\ &= -\frac{(\mu + \zeta)(S - S_0^*)^2}{S} - E(0, t) + S_0^* \int_0^\infty \beta(\tau) i(\tau, t) d\tau + S_0^* \int_0^\infty \xi(\theta) W(\theta, t) d\theta \\ &\quad + \zeta S_0^* \int_0^\infty \alpha_2(a) \pi_4(a) \left(\frac{S(t-a)}{S_0^*} - 1\right)\left(1 - \frac{S_0^*}{S}\right) da \\ &= -\frac{(\mu + \zeta)(S - S_0^*)^2}{S} - E(0, t) + S_0^* \int_0^\infty \beta(\tau) i(\tau, t) d\tau + S_0^* \int_0^\infty \xi(\theta) W(\theta, t) d\theta \\ &\quad + \zeta S_0^* \int_0^\infty \alpha_2(a) \pi_4(a) \left(\frac{S(t-a)}{S_0^*} - 1 - \frac{S(t-a)}{S} + \frac{S_0^*}{S}\right) da. \end{aligned} \tag{5.3}$$

$$\begin{aligned}
U_v'(t) &= \zeta S_0^* \int_0^\infty \alpha_2(a)\pi_4(a) \int_0^a f'\left(\frac{S(t-\sigma)}{S_0^*}\right)\frac{1}{S_0^*}\frac{dS(t-\sigma)}{dt}d\sigma da \\
&= -\zeta S_0^*\left[\int_0^\infty \alpha_2(a)\pi_4(a) \int_0^a f'\left(\frac{S(t-\sigma)}{S_0^*}\right)\frac{1}{S_0^*}\frac{dS(t-\sigma)}{d\sigma}d\sigma da\right] \\
&= -\zeta S_0^*\left[\int_0^\infty \alpha_2(a)\pi_4(a) f\left(\frac{S(t-\sigma)}{S_0^*}\right)\bigg|_0^a da\right] \\
&= \zeta S_0^* \int_0^\infty \alpha_2(a)\pi_4(a)\left[f\left(\frac{S(t)}{S_0^*}\right) - f\left(\frac{S(t-a)}{S_0^*}\right)\right]da \\
&= \zeta S_0^* \int_0^\infty \alpha_2(a)\pi_4(a)\left[\frac{S(t)}{S_0^*} - \frac{S(t-a)}{S_0^*} + \ln\frac{S(t-a)}{S(t)}\right]da.
\end{aligned} \tag{5.4}$$

And

$$\begin{aligned}
U_e'(t) &= S_0^* \int_0^\infty p(\tau)\left[-\frac{\partial E(\tau,t)}{\partial \tau} - \Big(\mu + m(\tau)\Big)E(\tau,t)\right]d\tau \\
&= -S_0^*\left[\int_0^\infty p(\tau)dE(\tau,t) + \int_0^\infty \Big(\mu + m(\tau)\Big)p(\tau)E(\tau,t)d\tau\right] \\
&= -S_0^*\left[\left(p(\tau)E(\tau,t)\bigg|_0^\infty - \int_0^\infty E(\tau,t)dp(\tau)\right) + \int_0^\infty \Big(\mu + m(\tau)\Big)p(\tau)E(\tau,t)d\tau\right] \\
&= -S_0^*\bigg\{-p(0)E(0,t) - \int_0^\infty \left[-q(0)m(\tau) + \Big(\mu + m(\tau)\Big)p(\tau)\right]E(\tau,t)d\tau \\
&\quad + \int_0^\infty \Big(\mu + m(\tau)\Big)p(\tau)E(\tau,t)d\tau\bigg\} \\
&= S_0^*\left[p(0)E(0,t) - \int_0^\infty q(0)m(\tau)E(\tau,t)d\tau\right] \\
&= S_0^*\Big[p(0)E(0,t) - q(0)i(0,t)\Big] \\
&= \mathcal{R}(\zeta)E(0,t) - S_0^*q(0)i(0,t).
\end{aligned} \tag{5.5}$$

Similarly, we have

$$\begin{aligned}
U_{ii}'(t) &= -S_0^*\left[\int_0^\infty q(\tau)di(\tau,t) + \int_0^\infty \Big(\mu + \alpha_1(\tau) + r_1(\tau)\Big)q(\tau)i(\tau,t)d\tau\right] \\
&= -S_0^*\bigg\{-q(0)i(0,t) - \int_0^\infty \left[-\Big(\beta(\tau) + \eta(\tau)g(0)\Big) + \Big(\mu + \alpha_1(\tau) + r_1(\tau)\Big)q(\tau)\right] \\
&\quad i(\tau,t)d\tau + \int_0^\infty \Big(\mu + \alpha_1(\tau) + r_1(\tau)\Big)q(\tau)i(\tau,t)d\tau\bigg\} \\
&= S_0^*q(0)i(0,t) - S_0^*\int_0^\infty \beta(\tau)i(\tau,t)d\tau - S_0^*g(0)W(0,t).
\end{aligned} \tag{5.6}$$

$$
\begin{aligned}
U_w'(t) &= -S_0^*\Big[\int_0^\infty g(\theta)dW(\theta,t) + \int_0^\infty \delta(\theta)g(\theta)W(\theta,t)d\theta\Big] \\
&= -S_0^*\Big\{-g(0)W(0,t) - \int_0^\infty \Big[-\xi(\theta)+\delta(\theta)g(\theta)\Big]W(\theta,t)d\theta \\
&\quad + \int_0^\infty \delta(\theta)g(\theta)W(\theta,t)d\theta\Big\} \\
&= S_0^* g(0)W(0,t) - S_0^*\int_0^\infty \xi(\theta)W(\theta,t)d\theta.
\end{aligned}
\tag{5.7}
$$

It can be noted here that

$$
i(0,t) = \int_0^\infty m(\tau)E(\tau,t)d\tau, \qquad W(0,t) = \int_0^\infty \eta(\tau)i(\tau,t)d\tau.
$$

Now adding all the derivatives, we get

$$
\begin{aligned}
U_1'(t) &= -\frac{(\mu+\zeta)(S-S_0^*)^2}{S} - E(0,t) + S_0^*\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S_0^*\int_0^\infty \xi(\theta)W(\theta,t)d\theta \\
&\quad +\zeta S_0^*\int_0^\infty \alpha_2(a)\pi_4(a)\Big(\frac{S(t-a)}{S_0^*} - 1 - \frac{S(t-a)}{S} + \frac{S_0^*}{S})\Big)da \\
&\quad +\zeta S_0^*\int_0^\infty \alpha_2(a)\pi_4(a)\Big[\frac{S}{S_0^*} - \frac{S(t-a)}{S_0^*} + \ln\frac{S(t-a)}{S(t)}\Big]da \\
&\quad +\mathcal{R}(\zeta)E(0,t) - S_0^* q(0)i(0,t) \\
&\quad +S_0^* q(0)i(0,t) - S_0^*\int_0^\infty \beta(\tau)i(\tau,t)d\tau - S_0^* g(0)W(0,t) \\
&\quad +S_0^* g(0)W(0,t) - S_0^*\int_0^\infty \xi(\theta)W(\theta,t)d\theta \\
&= -\frac{(\mu+\xi)(S-S_0^*)^2}{S} + \Big(R(\zeta)-1\Big)E(0,t) \\
&\quad +\zeta S_0^*\int_0^\infty \alpha_2(a)\pi_4(a)\Big[\Big(\frac{S_0^*}{S} + \frac{S(t)}{S_0^*} - 2\Big) - \Big(\frac{S(t-a)}{S} - 1 - \ln\frac{S(t-a)}{S(t)}\Big)\Big] \\
&= -\frac{\Big(\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da\Big)(S-S_0^*)^2}{S} + \Big(R(\zeta)-1\Big)E(0,t) \\
&\quad -\zeta S_0^*\int_0^\infty \alpha_2(a)\pi_4(a)f\Big(\frac{S(t-a)}{S(t)}\Big)d\theta \le 0.
\end{aligned}
\tag{5.8}
$$

The last inequality follows from the fact that $\mathcal{R}(\zeta) \le 1$. Notice that U_1' equals zero if and only if the first term equals zero, i.e., $S = S_0^*$, and if each of the terms in the above sum is equal to zero, i.e.,

$$E(0,t)=0, \qquad \frac{S(t-a)}{S(t)}=1.$$

We define a set

$$\Theta_1=\left\{(S,E,i,W,V)\in\Omega\middle|U_1'(t)=0\right\}.$$

LaSalle's Invariance Principle [24] implies that the bounded solutions of (2.2) converge to the largest compact invariant set of Θ_1. We will show that this largest compact invariant set is the singleton given by the disease-free equilibrium. First, we notice that equality in (5.8) occurs if and only if $S(t)=S_0^*,\ E(0,t)=0,\ S(t-a)/S(t)=1$. Thus, from the solutions for the equations along the characteristic lines (4.1), we have that $E(\tau,t)=0$ for all $t>\tau$. Hence, for $t>\tau$ we have

$$\begin{aligned}&\lim_{t\to\infty} i(0,t)=\lim_{t\to\infty}\int_0^\infty m(\tau)E(\tau,t)d\tau=0,\\ &i(\tau,t)=0, \quad \text{for all} \quad t>\tau.\end{aligned}$$

Similarly, we also have

$$\begin{aligned}&\lim_{t\to\infty} W(0,t)=\lim_{t\to\infty}\int_0^\infty \eta(\tau)i(\tau,t)d\tau=0,\\ &W(\theta,t)=0, \quad \text{for all} \quad t>\theta.\end{aligned}$$

We conclude that the disease-free equilibrium is globally stable. This completes the proof. □

In the next section we show that the endemic equilibrium $\mathcal{E}_1$ is globally stable.

6 Global Stability of the Endemic Equilibrium

From Proposition 3.2 we know that the endemic $\mathcal{E}_1$ is locally asymptotically stable. Now we are ready to establish the global stability of the equilibrium $\mathcal{E}_1$. To demonstrate that with Lyapunov function $U_2(t)$ (say) we have to establish that $U_2'(t)\le 0$ along the solution curves of system (2.2). With $f(x)=x-1-\ln x$, we define the following Lyapunov function

$$U_2(t)=U_S(t)+U_E(t)+U_i(t)+U_W(t)+U_V(t)$$

where

$$\begin{cases} U_S(t) &= S^* f(\frac{S}{S^*}), \\ U_E(t) &= S^* \int_0^\infty p(\tau)E^*(\tau) f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)d\tau, \\ U_i(t) &= S^* \int_0^\infty q(\tau)i^*(\tau) f\left(\frac{i(\tau,t)}{i^*(\tau)}\right)d\tau, \\ U_W(t) &= S^* \int_0^\infty g(\theta)W^*(\theta) f\left(\frac{W(\theta,t)}{W^*(\theta)}\right)d\theta, \\ U_V(t) &= \zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a) \int_0^a f\left(\frac{S(t-\sigma)}{S^*}\right)d\sigma da. \end{cases}$$

The following theorem summarizes the result.

Theorem 6.1 *Assume $\mathcal{R}(\zeta) > 1$. Then, equilibrium $\mathcal{E}_1$ is globally asymptotically stable, i.e., for any initial condition $x_0 \in X_0$ the solution semiflow converges to $\mathcal{E}_1$.*

Proof Following the same approach as in Theorem 5.1, we differentiate $U_2(t)$. Because of the complexity of the expressions, we take the derivative of each component of the Lyapunov function separately.

$$\begin{aligned} U_S'(t) &= \left(1 - \frac{S^*}{S}\right)\left[\Lambda - E(0,t) - (\mu+\zeta)S + \zeta \int_0^\infty \alpha_2(a)S(t-a)\pi_4(a)da\right] \\ &= \left(1 - \frac{S^*}{S}\right)\left[E^*(0) + (\mu+\zeta)S^* - \zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)da\right. \\ &\quad \left. -E(0,t) - (\mu+\zeta)S + \zeta \int_0^\infty \alpha_2(a)S(t-a)\pi_4(a)da\right] \\ &= -\frac{(\mu+\zeta)(S-S^*)^2}{S} + E^*(0) - \frac{S^*}{S}E^*(0) - E(0,t) \\ &\quad + S^* \int_0^\infty \beta(\tau)i(\tau,t)d\tau + S^* \int_0^\infty \xi(\theta)W(\theta,t)d\theta \\ &\quad + \zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)\left(\frac{S(t-a)}{S^*} - 1 - \frac{S(t-a)}{S} + \frac{S^*}{S}\right)da. \end{aligned} \tag{6.1}$$

$$U_V'(t) = -\zeta S^* \left[\int_0^\infty \alpha_2(a)\pi_4(a) f\left(\frac{S(t-\sigma)}{S^*}\right)\Big|_0^a da\right]$$

$$= \zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)\left[f\left(\frac{S(t)}{S^*}\right) - f\left(\frac{S(t-a)}{S^*}\right)\right]da \tag{6.2}$$

$$= \zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)\left[\frac{S(t)}{S^*} - \frac{S(t-a)}{S^*} + \ln\frac{S(t-a)}{S(t)}\right]da.$$

Next, we need to take the time derivative of U_E.

$$\begin{aligned} U_E'(t) &= S^*\int_0^\infty p(\tau)E^*(\tau)f'\left(\frac{E(\tau,t)}{E^*(\tau)}\right)\frac{1}{E^*(\tau)}\frac{\partial E(\tau,t)}{\partial t}d\tau \\ &= S^*\int_0^\infty p(\tau)E^*(\tau)f'\left(\frac{E(\tau,t)}{E^*(\tau)}\right)\frac{1}{E^*(\tau)}\left(-\frac{\partial E(\tau,t)}{\partial \tau}-(\mu+m(\tau))E(\tau,t)\right)d\tau \\ &= -S^*\int_0^\infty p(\tau)E^*(\tau)df\left(\frac{E(\tau,t)}{E^*(\tau)}\right) \\ &= -S^*\left[p(\tau)E^*(\tau)f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)\Big|_0^\infty-\int_0^\infty f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)d\left(p(\tau)E^*(\tau)\right)\right] \\ &= -S^*\left[-p(0)E^*(0)f\left(\frac{E(0,t)}{E^*(0)}\right)-\int_0^\infty f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)\left(-q(0)m(\tau)E^*(\tau)\right)d\tau\right] \\ &= S^*p(0)E^*(0)f\left(\frac{E(0,t)}{E^*(0)}\right)-S^*\int_0^\infty q(0)m(\tau)E^*(\tau)f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)d\tau. \end{aligned} \tag{6.3}$$

The above equality follows from (5.2) and the fact

$$\begin{aligned} &p'(\tau)E^*(\tau)+p(\tau)E'^*(\tau) \\ &= E^*(\tau)\Big[-q(0)m(\tau)+(\mu+m(\tau))p(\tau)\Big]-p(\tau)E^*(\tau)(\mu+m(\tau)) \\ &= -q(0)m(\tau)E^*(\tau). \end{aligned}$$

Similarly, we have

$$\begin{aligned} &q'(\tau)i^*(\tau)+q(\tau)i'^*(\tau) \\ &= i^*(\tau)\Big[-\Big(\beta(\tau)+g(0)\eta(\tau)\Big)+(\mu+\alpha_1(\tau)+r_1(\tau))q(\tau)\Big]-q(\tau)i^*(\tau)(\mu+\alpha_1(\tau)+r_1(\tau)) \\ &= -\Big(\beta(\tau)+g(0)\eta(\tau)\Big)i^*(\tau), \\ &g'(\theta)W^*(\theta)+g(\theta)W'^*(\theta) \\ &= W^*(\theta)\Big[-\xi(\theta)+\delta(\theta)g(\theta)\Big]-g(\theta)W^*(\theta)\delta(\theta) \\ &= -\xi(\theta)W^*(\theta). \end{aligned}$$

Differentiating $U_i(t)$, we have

$$\begin{aligned} U_i'(t) &= -S^* \int_0^\infty q(\tau)i^*(\tau)df\left(\frac{i(\tau,t)}{i^*(\tau)}\right) \\ &= S^*q(0)i^*(0)f\left(\frac{i(0,t)}{i^*(0)}\right) - S^* \int_0^\infty \beta(\tau)i^*(\tau)f\left(\frac{i(\tau,t)}{i^*(\tau)}\right)d\tau \\ &\quad -S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)f\left(\frac{i(\tau,t)}{i^*(\tau)}\right)d\tau. \end{aligned} \tag{6.4}$$

Now we turn to the derivative of the environmental component. Differentiating $U_W(t)$, we have

$$\begin{aligned} U_W'(t) &= -S^* \int_0^\infty g(\theta)W^*(\theta)df\left(\frac{W(\theta,t)}{W^*(\theta)}\right) \\ &= S^*g(0)W^*(0)f\left(\frac{W(0,t)}{W^*(0)}\right) - S^* \int_0^\infty \xi(\theta)W^*(\theta)f\left(\frac{W(\theta,t)}{W^*(\theta)}\right)d\theta. \end{aligned} \tag{6.5}$$

Adding derivatives of all the five components of the Lyapunov function, we have

$$\begin{aligned} U_2'(t) &= -\frac{(\mu+\zeta)(S-S^*)^2}{S} + E^*(0) - \frac{S^*}{S}E^*(0) - E(0,t) \\ &\quad +S_0^* \int_0^\infty \beta(\tau)i(\tau,t)d\tau + S_0^* \int_0^\infty \xi(\theta)W(\theta,t)d\theta \\ &\quad +\zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)\left(\frac{S(t-a)}{S^*} - 1 - \frac{S(t-a)}{S} + \frac{S^*}{S}\right)da \\ &\quad +\zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a)\left[\frac{S}{S^*} - \frac{S(t-a)}{S^*} + \ln\frac{S(t-a)}{S}\right]da \\ &\quad +S^*p(0)E^*(0)f\left(\frac{E(0,t)}{E^*(0)}\right) - S^*q(0)\int_0^\infty m(\tau)E^*(\tau)f\left(\frac{E(\tau,t)}{E^*(\tau)}\right)d\tau \\ &\quad +S^*q(0)i^*(0)f\left(\frac{i(0,t)}{i^*(0)}\right) - S^*\int_0^\infty \beta(\tau)i^*(\tau)f\left(\frac{i(\tau,t)}{i^*(\tau)}\right)d\tau \\ &\quad -S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)f\left(\frac{i(\tau,t)}{i^*(\tau)}\right)d\tau \\ &\quad +S^*g(0)W^*(0)f\left(\frac{W(0,t)}{W^*(0)}\right) - S^*\int_0^\infty \xi(\theta)W^*(\theta)f\left(\frac{W(\theta,t)}{W^*(\theta)}\right)d\theta. \end{aligned} \tag{6.6}$$

It can be noticed that

$$\begin{aligned}E^*(0) &= S^*\left(\int_0^\infty \beta(\tau)i^*(\tau)d\tau + \int_0^\infty \xi(\theta)W^*(\theta)d\theta\right)\\ &= S^*\left(i^*(0)\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau + W^*(0)\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta\right)\\ &= S^*\left(i^*(0)\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau + \int_0^\infty \eta(\tau)i^*(\tau)d\tau\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta\right)\\ &= S^*\left(\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau + \int_0^\infty \eta(\tau)\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta\right)i^*(0)\\ &= S^*i^*(0)\int_0^\infty (\beta(\tau) + g(0)\eta(\tau))\pi_2(\tau)d\tau\\ &= S^*i^*(0)q(0)\\ &= S^*q(0)\int_0^\infty m(\tau)E^*(\tau)d\tau\\ &= S^*q(0)E^*(0)\int_0^\infty m(\tau)\pi_1(\tau)d\tau\\ &= S^*p(0)E^*(0).\end{aligned} \tag{6.7}$$

From Equation (6.7), we obtain

$$S^*p(0) = 1, \quad S^*i^*(0)q(0) = E^*(0).$$

Using $f(x) = x - 1 - \ln x$, we have

$$
\begin{aligned}
U_2'(t) = &-\frac{(\mu+\zeta)(S-S^*)^2}{S} + E^*(0) - \frac{S^*}{S}E^*(0) - E(0,t)\\
&+S^*\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S^*\int_0^\infty \xi(\theta)W(\theta,t)d\theta\\
&+\zeta S^*\int_0^\infty \alpha_2(a)\pi_4(a)\bigg[\bigg(\frac{S^*}{S}+\frac{S}{S^*}-2\bigg)-\bigg(\frac{S(t-a)}{S}-1-\ln\frac{S(t-a)}{S}\bigg)\bigg]\\
&+E^*(0)\bigg(\frac{E(0,t)}{E^*(0)}-1-\ln\frac{E(0,t)}{E^*(0)}\bigg)\\
&-S^*q(0)\int_0^\infty m(\tau)E^*(\tau)\bigg(\frac{E(\tau,t)}{E^*(\tau)}-1-\ln\frac{E(\tau,t)}{E^*(\tau)}\bigg)d\tau\\
&+S^*q(0)i^*(0)\bigg(\frac{i(0,t)}{i^*(0)}-1-\ln\frac{i(0,t)}{i^*(0)}\bigg)\\
&-S^*\int_0^\infty \beta(\tau)i^*(\tau)\bigg(\frac{i(\tau,t)}{i^*(\tau)}-1-\ln\frac{i(\tau,t)}{i^*(\tau)}\bigg)d\tau\\
&-S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)\bigg(\frac{i(\tau,t)}{i^*(\tau)}-1-\ln\frac{i(\tau,t)}{i^*(\tau)}\bigg)d\tau\\
&+S^*g(0)W^*(0)\bigg(\frac{W(0,t)}{W^*(0)}-1-\ln\frac{W(0,t)}{W^*(0)}\bigg)\\
&-S^*\int_0^\infty \xi(\theta)W^*(\theta)\bigg(\frac{W(\theta,t)}{W^*(\theta)}-1-\ln\frac{W(\theta,t)}{W^*(\theta)}\bigg)d\theta.
\end{aligned}
$$

$$
\begin{aligned}
= &-\frac{(\mu+\zeta)(S-S^*)^2}{S} + E^*(0) - \frac{S^*}{S}E^*(0) - E(0,t)\\
&+S^*\int_0^\infty \beta(\tau)i(\tau,t)d\tau + S^*\int_0^\infty \xi(\theta)W(\theta,t)d\theta\\
&+\zeta S^*\int_0^\infty \alpha_2(a)\pi_4(a)\bigg[\bigg(\frac{S^*}{S}+\frac{S}{S^*}-2\bigg)-\bigg(\frac{S(t-a)}{S}-1-\ln\frac{S(t-a)}{S}\bigg)\bigg]\\
&+E(0,t)-E^*(0)-E^*(0)\ln\frac{E(0,t)}{E^*(0)} - S^*q(0)i(0,t) + S^*q(0)i^*(0)\\
&+S^*q(0)\int_0^\infty m(\tau)E^*(\tau)\ln\frac{E(\tau,t)}{E^*(\tau)}d\tau + S^*q(0)i(0,t) - S^*q(0)i^*(0)\\
&-S^*q(0)i^*(0)\ln\frac{i(0,t)}{i^*(0)} - S^*\int_0^\infty \beta(\tau)i(\tau,t))d\tau\\
&+S^*\int_0^\infty \beta(\tau)i^*(\tau)d\tau + S^*\int_0^\infty \beta(\tau)i^*(\tau)\ln\frac{i(\tau,t)}{i^*(\tau)}d\tau\\
&-S^*g(0)W(0,t) + S^*g(0)W^*(0) + S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)\ln\frac{i(\tau,t)}{i^*(\tau)}d\tau\\
&+S^*g(0)W(0,t) - S^*g(0)W^*(0) - S^*g(0)W^*(0)\ln\frac{W(0,t)}{W^*(0)}\\
&-S^*\int_0^\infty \xi(\theta)W(\theta,t))d\theta + S^*\int_0^\infty \xi(\theta)W^*(\theta)d\theta\\
&+S^*\int_0^\infty \xi(\theta)W^*(\theta)\ln\frac{W(\theta,t)}{W^*(\theta)}d\theta.
\end{aligned}
\tag{6.8}
$$

Simplifying (6.8), we obtain

$$\begin{aligned}U_2'(t) = &-\frac{(\mu + \zeta - \zeta\int_0^\infty \alpha_2(a)\pi_4(a)da)(S - S^*)^2}{S}\\ &-\zeta S^*\int_0^\infty \alpha_2(a)\pi_4(a)f\left(\frac{S(t-a)}{S}\right)da\\ &-\frac{S^*}{S}E^*(0) + E^*(0) - E^*(0)\ln\frac{E(0,t)}{E^*(0)}\\ &+S^*\int_0^\infty \beta(\tau)i^*(\tau)\ln\frac{i(\tau,t)}{i^*(\tau)}d\tau + S^*\int_0^\infty \xi(\theta)W^*(\theta)\ln\frac{W(\theta,t)}{W^*(\theta)}d\theta\\ &+S^*q(0)\int_0^\infty m(\tau)E^*(\tau)\ln\frac{E(\tau,t)}{E^*(\tau)}d\tau - S^*q(0)i^*(0)\ln\frac{i(0,t)}{i^*(0)}d\tau\\ &+S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)\ln\frac{i(\tau,t)}{i^*(\tau)}d\tau - S^*g(0)W^*(0)\ln\frac{W(0,t)}{W^*(0)}.\end{aligned} \tag{6.9}$$

It is easy to visualize that

$$\begin{cases} E^*(0) = S^*\int_0^\infty \beta(\tau)i^*(\tau)d\tau + S^*\int_0^\infty \xi(\theta)W^*(\theta)d\theta,\\ i^*(0) = \int_0^\infty m(\tau)E^*(\tau)d\tau,\\ W^*(0) = \int_0^\infty \eta(\tau)i^*(\tau)d\tau.\end{cases}$$

Thus we have

$$\begin{aligned}U_2'(t) = &-\frac{(\mu + \zeta - \zeta\int_0^\infty \alpha_2(a)\pi_4(a)da)(S - S^*)^2}{S}\\ &-\zeta S^*\int_0^\infty \alpha_2(a)\pi_4(a)f\left(\frac{S(t-a)}{S}\right)da\\ &+S^*\int_0^\infty \beta(\tau)i^*(\tau)\left(-\frac{S^*}{S} + 1 + \ln\frac{i(\tau,t)}{i^*(\tau)}\frac{E^*(0)}{E(0,t)}\right)d\tau\\ &+S^*\int_0^\infty \xi(\theta)W^*(\theta)\left(-\frac{S^*}{S} + 1 + \ln\frac{W(\theta,t)}{W^*(\theta)}\frac{E^*(0)}{E(0,t)}\right)d\theta\\ &+S^*q(0)\int_0^\infty m(\tau)E^*(\tau)\ln\frac{E(\tau,t)}{E^*(\tau)}\frac{i^*(0)}{i(0,t)}d\tau\\ &+S^*g(0)\int_0^\infty \eta(\tau)i^*(\tau)\ln\frac{i(\tau,t)}{i^*(\tau)}\frac{W^*(0)}{W(0,t)}d\tau.\end{aligned} \tag{6.10}$$

Next, we notice that

$$\begin{cases} S^*\displaystyle\int_0^\infty \beta(\tau)i^*(\tau)\left(1-\frac{S\ i(\tau,t)E^*(0)}{S^*i^*(\tau)E(0,t)}\right)d\tau \\ \qquad +S^*\displaystyle\int_0^\infty \xi(\theta)W^*(\theta)\left(1-\frac{S\ W(\theta,t)E^*(0)}{S^*W^*(\theta)E(0,t)}\right)d\theta=0, \\ \displaystyle\int_0^\infty m(\tau)E^*(\tau)\left(1-\frac{E(\tau,t)i^*(0)}{E^*(\tau)i(0,t)}\right)d\tau=0, \\ \displaystyle\int_0^\infty \eta(\tau)i^*(\tau)\left(1-\frac{i(\tau,t)W^*(0)}{i^*(\tau)W(0,t)}\right)d\tau=0. \end{cases} \tag{6.11}$$

Indeed

$$\begin{aligned} & S^*\int_0^\infty \beta(\tau)i^*(\tau)\left(1-\frac{S\ i(\tau,t)E^*(0)}{S^*i^*(\tau)E(0,t)}\right)d\tau \\ & \quad +S^*\int_0^\infty \xi(\theta)W^*(\theta)\left(1-\frac{S\ W(\theta,t)E^*(0)}{S^*W^*(\theta)E(0,t)}\right)d\theta \\ & = S^*\int_0^\infty \beta(\tau)i^*(\tau)d\tau - S\int_0^\infty \beta(\tau)i(\tau,t)d\tau\frac{E^*(0)}{E(0,t)} \\ & \quad +S^*\int_0^\infty \xi(\theta)W^*(\theta)d\theta - S\int_0^\infty \xi(\theta)W(\theta,t)d\theta\frac{E^*(0)}{E(0,t)} \\ & = E^*(0)-E(0,t)\frac{E^*(0)}{E(0,t)}=0. \end{aligned} \tag{6.12}$$

$$\begin{aligned} & \int_0^\infty m(\tau)E^*(\tau)\left(1-\frac{E(\tau,t)i^*(0)}{E^*(\tau)i(0,t)}\right)d\tau \\ & = \int_0^\infty m(\tau)E^*(\tau)d\tau - \int_0^\infty m(\tau)E(\tau,t)d\tau\frac{i^*(0)}{i(0,t)} \\ & = i^*(0)-i(0,t)\frac{i^*(0)}{i(0,t)}=0, \end{aligned} \tag{6.13}$$

and

$$\begin{aligned} & \int_0^\infty \eta(\tau)i^*(\tau)\left(1-\frac{i(\tau,t)W^*(0)}{i^*(\tau)W(0,t)}\right)d\tau \\ & = \int_0^\infty \eta(\tau)i^*(\tau)d\tau - \int_0^\infty \eta(\tau)i(\tau,t)d\tau\frac{W^*(0)}{W(0,t)} \\ & = W^*(0)-W(0,t)\frac{W^*(0)}{W(0,t)}=0. \end{aligned} \tag{6.14}$$

Using (6.11) to simplify (6.10), we obtain

$$
\begin{aligned}
U_2'(t) =& -\frac{(\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da)(S-S^*)^2}{S} \\
&-\zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a) f\left(\frac{S(t-a)}{S}\right)da \\
&+S^* \int_0^\infty \beta(\tau)i^*(\tau)\left(2-\frac{S^*}{S}-\frac{S}{S^*}\frac{i(\tau,t)}{i^*(\tau)}\frac{E^*(0)}{E(0,t)}+\ln\frac{i(\tau,t)}{i^*(\tau)}\frac{E^*(0)}{E(0,t)}\right)d\tau \\
&+S^* \int_0^\infty \xi(\theta)W^*(\theta)\left(2-\frac{S^*}{S}-\frac{S}{S^*}\frac{W(\theta,t)}{W^*(\theta)}\frac{E^*(0)}{E(0,t)}+\ln\frac{W(\theta,t)}{W^*(\theta)}\frac{E^*(0)}{E(0,t)}\right)d\tau \\
&+S^* q(0)\int_0^\infty m(\tau)E^*(\tau)\left(-\frac{E(\tau,t)}{E^*(\tau)}\frac{i^*(0)}{i(0,t)}+1+\ln\frac{E(\tau,t)}{E^*(\tau)}\frac{i^*(0)}{i(0,t)}\right)d\tau \\
&+S^* g(0)\int_0^\infty \eta(\tau)i^*(\tau)\left(-\frac{i(\tau,t)}{i^*(\tau)}\frac{W^*(0)}{W(0,t)}+1+\ln\frac{i(\tau,t)}{i^*(\tau)}\frac{W^*(0)}{W(0,t)}\right)d\tau.
\end{aligned}
\tag{6.15}
$$

Finally, we have

$$
\begin{aligned}
U_2'(t) =& -\frac{(\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da)(S-S^*)^2}{S} \\
&-\zeta S^* \int_0^\infty \alpha_2(a)\pi_4(a) f\left(\frac{S(t-a)}{S}\right)da \\
&-S^* \int_0^\infty \beta(\tau)i^*(\tau)\left[f\left(\frac{S^*}{S}\right)+f\left(\frac{S}{S^*}\frac{i(\tau,t)}{i^*(\tau)}\frac{E^*(0)}{E(0,t)}\right)\right]d\tau \\
&-S^* \int_0^\infty \xi(\theta)W^*(\theta)\left[f\left(\frac{S^*}{S}\right)+f\left(\frac{S}{S^*}\frac{W(\theta,t)}{W^*(\theta)}\frac{E^*(0)}{E(0,t)}\right)\right]d\tau \\
&-S^* q(0)\int_0^\infty m(\tau)E^*(\tau) f\left(\frac{E(\tau,t)}{E^*(\tau)}\frac{i^*(0)}{i(0,t)}\right)d\tau \\
&-S^* g(0)\int_0^\infty \eta(\tau)i^*(\tau) f\left(\frac{i(\tau,t)}{i^*(\tau)}\frac{W^*(0)}{W(0,t)}\right)d\tau \le 0.
\end{aligned}
\tag{6.16}
$$

Define,

$$
\Theta_2 = \left\{(S, E, i, W, V) \in X_0 \,\middle|\, U_2'(t) = 0\right\}.
$$

We want to show that the largest invariant set in Θ_2 is the singleton $\mathcal{E}_1$. First, we notice that equality in (6.17) occurs if and only if

$$
\begin{aligned}
&S = S^*, \quad \frac{S(t-a)}{S(t)} = 1, \quad \frac{E(\tau,t)}{E^*(\tau)}\frac{i^*(0)}{i(0,t)} = 1, \quad \frac{i(\tau,t)}{i^*(\tau)}\frac{W^*(0)}{W(0,t)} = 1, \\
&\frac{S}{S^*}\frac{i(\tau,t)}{i^*(\tau)}\frac{E^*(0)}{E(0,t)} = 1, \quad \frac{S}{S^*}\frac{W(\theta,t)}{W^*(\theta)}\frac{E^*(0)}{E(0,t)} = 1.
\end{aligned}
\tag{6.17}
$$

From the conditions (6.17) we have

$$\begin{aligned} &S(t-a)=S(t)=S^*, \qquad E(\tau,t)=\frac{i(0,t)}{i^*(0)}E^*(\tau),\\ &i(\tau,t)=\frac{E(0,t)}{E^*(0)}i^*(\tau)=\frac{W(0,t)}{W^*(0)}i^*(\tau), \quad W(\theta,t)=\frac{E(0,t)}{E^*(0)}W^*(\theta). \end{aligned} \tag{6.18}$$

Furthermore, $S=S^*,\ S(t-a)/S=1$ implies that $dS/dt=0$. Consequently, using (6.18), we have

$$\begin{aligned} 0&=\Lambda-S^*\int_0^\infty\beta(\tau)i(\tau,t)d\tau-S^*\int_0^\infty\xi(\theta)W(\theta,t)d\theta-(\mu+\zeta)S^*\\ &\quad+\zeta S^*\int_0^\infty\alpha_2(a)\pi_4(a)da\\ &=S^*\int_0^\infty\beta(\tau)i^*(\tau)d\tau-S^*\int_0^\infty\beta(\tau)\left(\frac{E(0,t)}{E^*(0)}i^*(\tau)\right)d\tau\\ &\quad+S^*\int_0^\infty\xi(\theta)W^*(\theta)d\theta-S^*\int_0^\infty\xi(\theta)\left(\frac{E(0,t)}{E^*(0)}W^*(\theta)\right)d\theta\\ &=\left(1-\frac{E(0,t)}{E^*(0)}\right)\left(S^*\int_0^\infty\beta(\tau)i^*(\tau)d\tau+S^*\int_0^\infty\xi(\theta)W^*(\theta)d\theta\right). \end{aligned} \tag{6.19}$$

Hence, we must have $E(0,t)=E^*(0)$. Thus, using condition (6.18), we have $i(\tau,t)=i^*(\tau)$, $W(\theta,t)=W^*(\theta)$. We conclude that the largest invariant set in Θ_2 is the singleton $\mathcal{E}_1$. Following the similar argument as in [22], one can conclude that the compact global attractor $\mathfrak{T}=\{\mathcal{E}_1\}$. □

7 Discussion

Here an SVEIR model is proposed to describe the transmission of TB bacteria by incorporating the age of latency, age of infection, age of vaccination and biological age of pathogen in the environment. Under our assumptions, the global dynamics of the proposed model is shown to be determined completely by the basic reproduction number $\mathcal{R}(\zeta)$. The disease dies out if $\mathcal{R}(\zeta)<1$. If $\mathcal{R}(\zeta)>1$, the disease persists. The global stability results for the disease-free and the endemic equilibrium are proved by constructing suitable Lyapunov functions as used in [22, 23, 25, 26].

The population level prevalence of TB bacteria is governed by the reproduction number $\mathcal{R}(\zeta)>1$. $\mathcal{R}(\zeta)$ is a decreasing function of the vaccination rate ζ, so increasing the rate of vaccination reduces the infection prevalence of TB bacteria in the population. This helps in controlling the transmission of TB bacteria. We also note that

$$-\int_0^\infty \Big(\mu+\alpha_2(a)+r_2(a)\Big)e^{-\int_0^a(\mu+\alpha_2(s)+r_2(s))ds}da = e^{-\int_0^a(\mu+\alpha_2(s)+r_2(s))ds}|_0^\infty = -1. \tag{7.1}$$

From the Eq. (7.1), we get

$$-\int_0^\infty \alpha_2(a)\pi_4(a)da = \int_0^\infty (\mu+r_2(a))\pi_4(a)d\theta - 1. \tag{7.2}$$

Thus we can obtain

$$\mu+\zeta-\zeta\int_0^\infty \alpha_2(a)\pi_4(a)da = \mu+\zeta\int_0^\infty (\mu+r_2(a))\pi_4(a)da. \tag{7.3}$$

Hence, $\mathcal{R}(\zeta)$ can be rewritten as

$$\mathcal{R}(\zeta) = \frac{\Lambda\Big(\int_0^\infty m(\tau)\pi_1(\tau)d\tau\int_0^\infty \beta(\tau)\pi_2(\tau)d\tau + \int_0^\infty m(\tau)\pi_1(\tau)d\tau\int_0^\infty \eta(\tau)\pi_2(\tau)d\tau\int_0^\infty \xi(\theta)\pi_3(\theta)d\theta\Big)}{\mu+\zeta\int_0^\infty(\mu+r_2(a))\pi_4(a)da}. \tag{7.4}$$

Here it can be noted that $\mathcal{R}(\zeta)$ is an increasing function of age of the vaccination a, and if the age of the vaccination a is large, the TB bacteria can persist. It may be because of this that the TB bacteria still exist with high level of prevalence across the world. From our presented results, one can observe that for a given set of parameters, the sensitivity of $\mathcal{R}(\zeta)$ to the vaccination rate and age of the vaccination can be used to guide disease control policies.

Indirect transmission affects the basic reproduction number and hence it leads into an additional contribution to the final size relation of the epidemic model. Here the basic reproduction number $\mathcal{R}$ is a decreasing function of the clearance rate $\delta(\tau)$, and it suggests that keeping the environment clean is an effective way to prevent the transmission of TB disease to some extent in the long-term perspective. There are some practical examples to support this observation. Also, the sensitivity of $\mathcal{R}$ to contaminated environment can be used to guide the disease control strategies.

Acknowledgements The authors thank the handling editor and anonymous referees for their valuable comments and suggestions which led to an improvement of our original manuscript.

References

1. Feng, Z.L., Chavez, C.C., Capurro, A.F.: A model for tuberculosis with exogenous reinfection. Theoret. Popul. Biol. **57**, 235–247 (2000)
2. Bloom, B.R.: Tuberculosis: Pathogenesis, Protection, and Control. ASM Press, Washington (1994)
3. Miller, B.: Preventive therapy for tuberculosis. Med. Clin. North Am. **77**, 1263–1275 (1993)
4. Smith, P.G., Moss, A.R.: Epidemiology of tuberculosis. In: Bloom, B.R. (eds.) Tuberculosis: Pathogenesis, Protection, and Control. ASM Press, Washington (1994)

5. Kolata, G.: First documented case of TB passed on airliner is reported by the U.S., New York Times, 3 March 1995
6. Kline, S.E., Hedemark, L.L., Davies, S.F.: Outbreak of tuberculosis among regular patrons of a neighborhood bar. N. Engl. J. Med. **333**, 222–227 (1995)
7. Duan, X.C., Yuan, S.L., Li, X.Z.: Global stability of an SVIR model with age of vaccination. Comput. Math. Appl. **226**, 528–540 (2014)
8. Kribs-Zaleta, C.M., Velasco-Hernandez, J.X.: A simple vaccination model with multiple endemic states. Math. Biosci. **164**, 183–201 (2000)
9. Li, J.Q., Ma, Z., Zhou, Y.: Global analysis of SIS epidemic model with a simple vaccination and multiple endemic equilibria. Acta Math. Sci. **26**, 83–93 (2006)
10. Liu, X., Takeuchi, Y., Iwami, S.: SVIR epidemic models with vaccination strategies. J. Theoret. Biol. **253**, 1–11 (2008)
11. Kribs-Zaleta, C.M., Martcheva, M.: Vaccination strategies and backward bifurcation in an age-since-infection structured model. Math. Biosci. **177&178**, 317–332 (2002)
12. Buonomo, B., d'Onofrio, A., Lacitignola, D.: Global stability of an SIR epidemic model with information dependent vaccination. Math. Biosci. **216**, 9–16 (2008)
13. Cao, B., Huo, H.-F., Xiang, H.: Global stability of an age-structure epidemic model with imperfect vaccination and relapse. Phys. A Stat. Mech. Appl. **486**, 638–655 (2017)
14. Wang, X., Zhang, Y., Song, X.: An age-structured epidemic model with waning immunity and general nonlinear incidence rate. Int. J. Biomath. **11**(5), 1850069 (2018)
15. Wang, C., Fan, D., Xia, L., Yi, X.: Global stability for a multi-group SVIR model with age of vaccination. Int. J. Biomath. **11**(5), 1850068 (2018)
16. Martcheva, M., Thieme, H.R.: Progression age enhanced backward bifurcation in an epidemic model with super-infection. J. Math. Biol **46**, 385–424 (2003)
17. Allen, L.J.S.: An Introduction to Mathematical Biology. Pearson, NJ (2007)
18. Hale, J.K.: Asymptotic Behavior of Dissipative Systems. AMS, Providence (1988)
19. Yosida, K.: Functional Analysis, 2nd edn. Springer, Berlin (1968)
20. Thieme, H.R.: Uniform persistence and permanence for non-autonomous semiflows in population biology. Math. Biosci. **166**, 173–201 (2000)
21. D'Agata, E., Magal, P., Ruan, S., Webb, G.F.: Asymptotic behavior in nosocomial epidemic models with antibiotic resistance. Differ. Integral Equ. **19**, 573–600 (2006)
22. Magal, P., McCluskey, C.C., Webb, G.F.: Lyapunov functional and global asymptotic stability for an infection-age model. Appl. Anal. **89**(7), 1109–1140 (2010)
23. Martcheva, M., Li, X.Z.: Competitive exclusion in an infection-age structured model with environmental transmission. J. Math. Anal. Appl. **408**, 225–246 (2013)
24. LaSalle, J.P.: Some extensions of Lyapunov's second method. IRE Trans. Circuit Theory CT **7**: 520–527 (1960)
25. Brauer, F., Shuai, Z.S., Driessche, P.V.D.: Dynamics of an age-of-infection cholera model. Math. Biosci. Eng. **10**, 1335–1349 (2013)
26. Huang, G., Liu, X., Takeuchi, Y.: Lyapunov functions and global stability for age-structured HIV infection model. SIAM J. Appl. Math. **72**, 25–38 (2012)

Gravitational Waves: The Mathematical Background

Subenoy Chakraborty

Abstract The geometrical construction of General Relativity which led Einstein to the prediction of gravitational waves is discussed and formation of Cauchy problem is shown. Due to complexity of the Cauchy problem, it is not possible to solve it by the methods of analysis and geometry—approximation methods and numerical algorithms are used. Finally, the consequences of detecting gravitational waves are discussed.

Keywords Lorentzian manifold · Initial value problem · Gravitational waves

1 Introduction

The theoretical prediction of Einstein about the wave nature of gravity came immediately [1] after the formulation of general relativity by himself. But 99 more years have passed before gravitational waves (GW) became in reality. The LIGO-VIRGO collaborators detected the GW for the first time in 2015. GW are considered as vibrations in space-time, propagating at the speed of light away from the source. Research on GW (radiation) is the common platform for interaction among astronomers, physicists and mathematicians. The mathematical construction of GW are based on the properties of the Einstein vacuum equations and the initial value problem (Cauchy problem). Based on the theory of non-linear partial differential equations (PDE) gravitational field dynamics is studied as a Cauchy problem of the Einstein field equations. The basic differential geometry needed to define the universe as a geometric object is presented below in Sect. 2. Also the mathematical properties of the Einstein vacuum equations are presented in this section. A detailed study about the initial value formulation i.e. the Cauchy problem is presented in Sect. 3. Also a discussion about gravitational radiation is included in this section. Both linearized and post-Newtonian approximation of Einstein equations is studied in Sect. 4 in the

S. Chakraborty (✉)
Department of Mathematics, Jadavpur University, Kolkata 700032, West Bengal, India

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_34

context of gravitational waves. Numerical techniques and possible computer simulation is discussed in Sect. 5. Finally, GW detection is summarized in Sect. 6.

2 Geometry of Space-Time

Our universe is considered as 4D space-time manifold (M) with a Lorentzian metric $g_{\mu\nu}$ and the curvature tensor characterizes the properties of the gravitational field. Usually, the manifold M is assumed to be oriented differentiable manifold and the metric tensor is a non-degenerate quadratic form

$$g = \sum_{\mu,\nu=0}^{3} g_{\mu\nu}\, dx^{\mu} \otimes dx^{\nu} \tag{1}$$

which is defined smoothly over the manifold (known as the tangent space) $T_q M$ at every point $q \in M$. In $T_q M$ any four vector $\overrightarrow{X}$ can be classified as (a) time-like if $g\left(\overrightarrow{X}, \overrightarrow{X}\right) < 0$, (b) space-like if $g\left(\overrightarrow{X}, \overrightarrow{X}\right) > 0$ and (c) null or light-like if $g\left(\overrightarrow{X}, \overrightarrow{X}\right) = 0$. In fact at each point q the collection of null vectors form a cone, termed the null cone. So one can introduce the notion of causal curve as a differentiable curve having time-like or null tangent vector at each point. Similarly, one can introduce the notion of space-like hypersurface if the normal vector is time-like. So the metric tensor restricted to the hypersurface should be positive definite (i.e. the hypersurface is a Riemannian manifold). A space-like hypersurface $(\mathscr{H})$ can be characterized as Cauchy hypersurface if every causal curve through any point $x \in M$ intersects $\mathscr{H}$ exactly at one point. In this context, a space-time manifold (M, g) is said to be globally hyperbolic if it has a Cauchy hypersurface. Hence from the very definition, a globally hyperbolic space-time manifold should be causal and it is possible to define a time function (t) at every point on the manifold so that its level surfaces (i.e. $t =$constant) are Cauchy surfaces.

In a globally hyperbolic space-time a Cauchy hypersurface $(\mathscr{H})$ is a Riemannian manifold equipped with a metric h_{ab} and there also exists a 2nd rank symmetric tensor K_{ab} (known as extrinsic curvature) which characterizes how the hypersurface $\mathscr{H}$ is embedded in the space-time manifold M. These two (symmetric) 2nd rank tensors are determined from the following first and second fundamental forms on the hypersurface:

$$h_{ab} = g_{\alpha\beta}\, e^{\alpha}_{a}\, e^{\beta}_{b} \tag{2}$$

$$K_{ab} = n_{\alpha;\beta}\, e^{\alpha}_{a}\, e^{\beta}_{b}. \tag{3}$$

Here $e^{\alpha}_{a} = \frac{\partial x^{\alpha}}{\partial y^{a}}$ are tangent to curves contained in Σ, hypersurface of the 4D space-time manifold and n^{α} is normal to the hypersurface.

The vacuum Einstein equations (i.e. $R_{\mu\nu} = 0$) can be written on the hypersurface as the evolution equations of 3-metric h_{ab} and the extrinsic curvature K_{ab} as

$$\frac{\partial h_{ab}}{\partial t} = -2NK_{ab} + \mathscr{L}_{\overrightarrow{\underset{\sim}{N}}} h_{ab} \tag{4}$$

$$\text{and} \quad \frac{\partial K_{ab}}{\partial t} = -\nabla_a \nabla_b N + \mathscr{L}_{\overrightarrow{\underset{\sim}{N}}} K_{ab} + \left(R_{ab} + K_{ab}\,\mathrm{Tr}\,K^{ab} - 2K_{ad}K^{d}_{b}\right) N. \tag{5}$$

Here N is the lapse function and $\overrightarrow{\underset{\sim}{N}}$ is the shift (three) vector which are essentially the g_{tt} and g_{ta} components of the metric. In fact this scalar and vector functions are related to the evolution vector field $\overrightarrow{T}$ $\left(i.e.\ \frac{\partial}{\partial t}\right)$ as

$$\overrightarrow{T} = N\overrightarrow{n} + \overrightarrow{N} \tag{6}$$

with $\overrightarrow{n}$, the unit normal to the $t =$ constant hypersurface. Also in the above ∇_a and $\mathscr{L}$ at the covariant and Lie derivative on the hypersurface. However, it should be noted h_{ab} and K_{ab} can not evolve independently rather they are related to the constraint equations:

$$\nabla^a K_{ab} - \nabla_b\,\mathrm{tr}\,K = 0 \tag{7}$$

$$\text{and} \quad \underset{\sim}{R} + (\mathrm{tr}\,K)^2 - |K|^2 = 0 \tag{8}$$

These constraint equations play an important role in the following initial value formulation.

3 Initial Value Formulation and Gravitational Radiation

In General Relativity (GR) an initial data set is a 3D manifold $\mathscr{H}$ equipped with a Riemannian metric h_{ab} and the extrinsic curvature k_{ab} i.e. $(\mathscr{H}, h_{ab}, k_{ab})$, which satisfy the constraint Eqs. (7) and (8). Usually, it is assumed that initial data set should be asymptotically flat i.e. outside a sufficiently large compact set $\mathbb{C}$, $\mathscr{H} \setminus \mathbb{C}$ is diffeomorphic to the complement of a closed ball in $\mathbb{R}^3$ and $h_{ab} \to \eta_{ab},\ k_{ab} \to 0$ at infinity.

Although many mathematicians took interest in the beauty and challenges of GR by the Cauchy problem was formulated properly much later because geometry and partial differential equation theory were not so rich at the time. The pioneering result was due to Choquet-Bruhat in 1952 [2] and it was local in nature. According to them: "for an initial data set $(\mathscr{H}, h, k)$ satisfying vacuum constraint equations there always exists a space-time (M, g) satisfying the Einstein vacuum equations and for which $\mathscr{H}$ is a spacelike hypersurface with induced metric 'h' and 2nd fundamental form 'K'." They used a typical co-ordinate system (known as wave coordinates) in which Einstein's vacuum equations can be expressed as a system of hyperbolic partial differential equations.

The above local Cauchy problem does not address the following questions: (i) Will the local solutions of the Einstein equations exist for all time ? (ii) Will the local solutions form singularities and what type will it be? In 1969 Choquet-Bruhat and Geroch [3] showed that it is possible to have a unique, globally hyperbolic maximal spacetime (M, g) (corresponding to the solutions of the vacuum Einstein equations) which is evolved from a Cauchy surface $\mathscr{H}$. This 4D manifold M is called the maximal future development of the initial data set $(\mathscr{H}, h_{ab}, k_{ab})$.

A further generalization of the above result has been done by Christodoulou and Klainerman [4] in 1993. They have proved that it is possible to construct a unique causally geodesically complete and globally hyperbolic and globally asymptotically flat 4D space-time manifold (M, g) which is evolved from a strong asymptotically flat initial data set.

Mathematically, based on L^2 curvature conjecture (by Klainerman et al. [5–10]) the time of existence of the solution depends only on the L^2-norms of the Riemann curvature tensor and on the gradient of the second fundamental form.

In the present talk, we are interested about the properties of radiation i.e. the behaviour of curvature at large distances. As gravitational radiation propagates at the speed of light so the asymptotic behaviour of gravitational waves near infinity will be important and it is characterized by radiative space-times with asymptotic structures. Due to propagation of gravitational radiation along null hypersurfaces in the space-time so it is expected that the wave will reach to an observer at future null infinity (I^+). I^+ is usually defined as the collection of all endpoints of future-directed null geodesics along which $r \to \infty$. Geometrically, I^+ has the topology $\mathbb{R} \times S^2$ and it intersects a null hypersurface in a 2-sphere (S^2). The Trautman-Bondi mass on a null hypersurface at I^+ measures the amount of mass that remains in an isolated gravitational system in that this mass measures the remaining mass after radiation through I^+.

From the geometry of the space-time at I^+ it can be shown that a test mass will go to rest after the gravitational wave has passed and consequently the geodesics will not deviate anymore. But the question is, will the test masses be at the same position as before the wave passed or will they be dislocated ? Mathematically, will the geometry of space-time change permanently? The affirmative answer is known as the memory effect of gravitational waves. This memory effect has been analyzed in the linearized theory by Zel'dovich and Polnarev [11] and in the full non-linear formulation by Christodoulou [12]. The linearized effect (also known as ordinary) was not detectable (at that time due to its very small nature) while the non-linear effect (also known as null) can in principle be measured. Later, Bieri and Garfinkle [13] showed that these two memories are two different effects namely the difference of a specific component of the Weyl tensor is the source for ordinary memory while the null memory effect is due to fields that do reach null infinity I^+. Mathematically, the permanent displacement (i.e. memory) is related to [14]

$$F = c\int_{-\infty}^{\infty} |\sigma(u)|^2\, du$$

where $\frac{F}{4\pi}$ measures the total energy radiated in a given direction per unit solid angle.

4 Linearized and Post-Newtonian Approximation of Einstein Equations: Gravitational Waves

Gravitational waves become weaker and weaker as they propagate away from the source. So it is reasonable to assume that the space-time metric far away from the source to be very close to Minkowskian geometry i.e.

$$g_{\mu\nu} = \eta_{\mu\nu} + q_{\mu\nu}$$

where $\eta_{\mu\nu}$ is the usual Minkowskian metric i.e. $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$ and $q_{\mu\nu}$ is considered to be the small deviation from the flat geometry. In the linearized Einstein theory the field equations contain only linear power of q and its derivatives. Due to general covariance of Einstein gravity there is gauge invariance in linearized gravity. As a result any function A written as $A = A_0 + \delta A$ where A_0 is its value at the background Minkowskian geometry while δA is its first order perturbation. Thus for an infinitesimal diffeomorphism along an vector field ξ, δA changes to $\delta A \to \delta A + \mathscr{L}_{\overrightarrow{\xi}} A_0$. Similar to harmonic co-ordinates, the vector field $\overrightarrow{\xi}$ is chosen to satisfy the Lorentz gauge condition: $\partial_\mu \overline{q}^{\mu\nu} = 0$ with $\overline{q}_{\mu\nu} = q_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} q$. Then the linearized Einstein field equations reduces to the wave equation

$$\Box \overline{q}_{\mu\nu} = -16\pi G T_{\mu\nu}$$

where $\Box$ is the usual wave operator in Minkowski space-time.

Further, in vacuum the remaining freedom of choosing $\overrightarrow{\xi}$ imposes the conditions that the deviation metric $q_{\mu\nu}$ has only spatial components and is tracefree but still it satisfies the Lorentz gauge. Then it is termed TT gauge in which the only two propagating degrees of freedom of the metric perturbation are transverse (i.e. $\partial^i h_{ij} = 0$) and spatially traceless (i.e. $\eta^{ij} q_{ij} = 0$). In this (TT) gauge the linearized Riemannian curvature tensor takes the form: $R_{it,jt} = -\frac{1}{2}\ddot{q}_{ij}^{TT}$ and it can be considered as the source for geodesic deviation. As a result, it shows how matter behaves in the presence of GW.

Now using the above equation in the geodesic deviation equation one may be able to calculate the change in distance between two free falling test particles

$$\Delta d^i(t) = \frac{1}{2} h_{ij}^{TT}(t) d_0^j$$

with d_0^j, the initial distance between the test masses.

Moreover, the gravitational wave perturbations have only two polarizations due to the above TT gauge. So if the wave travelling along the z-direction such that $q_{ij}^{TT}(t - z)$ is a solution of $\Box\, q_{ij}^{TT} = 0$, then there are only two independent propagating degrees of freedom:

$$\begin{aligned} q_{+}(t-z) &= q_{xx}^{TT} = -q_{yy}^{TT} \\ \text{and} \quad q_{\times}(t-x) &= q_{xy}^{TT} = q_{yx}^{TT} \end{aligned}$$

where one has to take into account that the metric perturbation vanishes for larger 'r' and its trace-free nature in addition to Lorentz condition. In the above q_+ gravitational wave stretches the x-direction in space while it squeezes the y-direction and vice-versa. Hence in aLIGO the interferometer used to detect GW has two long perpendicular arms which measure this distortion.

The above linearized theory can be extended to higher order in the metric perturbation, assuming the object producing gravitational field moves with non-relativistic speed. Such formulation of Einstein theory is known as post-Newtonian approximation. This method was developed by Einstein, Will, Schaefer and others [15, 16]. Using harmonic gauge (i.e. $\partial_\alpha\left(\sqrt{-g}g^{\alpha\beta}\right) = 0$) the Einstein field equations in PN theory take the form

$$\begin{aligned} \Box\, q^{\alpha\beta} &= -\frac{16\pi G}{c^4}\tau^{\alpha\beta} \\ \text{where} \quad \tau^{\alpha\beta} &= -(g)T^{\alpha\beta} + (16\pi)^{-1}N^{\alpha\beta} \end{aligned}$$

with $N^{\alpha\beta}$ containing quadratic forms of the metric perturbation.

The above inhomogeneous wave equation in the post-Newtonian approximation is normally solved order by order in the perturbation through Green function methods with integral over the past light cone of Minkowski space. At each order, the previously calculated information is used in the expression for $N^{\alpha\beta}$ and to find the motion of the matter sources. Normally, at higher post-Newtonian order, the integrals appear to be divergent. But these infinities can be overcome either by asymptotic matching methods [17–19] or through regularization techniques [20, 21]. Thus the PN iterative process provides an approximate perturbative solution to the Einstein equations upto a given order restricting gravitational interaction and the speed of the objects i.e. the speed of the body should not be comparable to the speed of light and the object should not be highly gravitating namely black hole (BH) or neutron stars. However, Damour [22] showed that even for significant self-gravitating object, PN approximation can be described upto a given order in the perturbation. Recently, it has been shown [23] that PN approximation can accurately describe the merger of binary BHs the and neutron stars. Finally, it is to be noted that the accuracy of the approximate solutions can be improved by using resummation techniques.

5 Numerical Techniques

In numerical relativity, Einstein field equations are solved by the method of simulation. In fact it is very useful when gravity is very strong and highly dynamical (for example, the case of two BH merger). Normally, for a differential equation, the most common and widely used technique for simulation is finite difference method [24]. According to this method, a function $f(x)$ can be approximated by its values on equally spaced points

$$f_i = f(i\delta)\ , \quad i \in \mathbb{N},$$

so that the derivatives can be approximated as

$$f'_{(i)} \approx \left(f_{i+1} - f_{i-1}\right)/(2\delta)$$

$$f''_{(i)} \approx \left(f_{i+1} + f_{i-1} - 2f_i\right)/\delta^2.$$

However, for a partial differential equation with an initial value formulation, one replaces the fields by their values on a space-time lattice and the field equations by finite difference equations which evaluate the fields at time step 'n'. Thus the vacuum Einstein equations can be written as difference equations with '0' step information as the initial data set. However, there are difficulties with the finite difference technique as follows:

(a) Normally, the solution of the finite difference equation should converge to a solution of the corresponding differential equation in the limit as $\delta \to 0$. But it is not always the case. Due to coordinate invariance of general relativity, it is possible to express the Einstein field equations in many different forms which are not all strongly hyperbolic. These forms of the Einstein field equations do not converge in computer simulations.
(b) Due to the constraint equations the initial data as well as its evolution should satisfy the constraint equations. But in computer simulation the initial data only satisfy the finite difference version of the constraints and hence they have a small amount of constraint violation. So it is possible that initial data with small constraint violation may give rise to rapidly growing constraint violation and hence the accuracy of the simulation is destroyed.
(c) The simulation using infinite difference method will not be applicable if the space-time contains BH (i.e. singularity). The simulation cannot be continued to the past if a time slice contains a space-time singularity. Even if the time slice advances slowly inside the BH and rapidly outside it, then stretching of the slices may lead to inaccuracies in the finite difference approximation.

In 2005, the above problems were resolved by Pretorius [25] showing fully successful binary BH simulation. Subsequently, in other works [26, 27], these issues were resolved for various BH masses and spins. The most efficient simulations are done by the $S \times S$ (i.e. simulating extreme space-times) collaboration using spectral methods instead of finite difference approaches [28]. In spectral methods the grid values f_i are chosen as approximation to the function $f(x)$ in the expansion with a particular basis of orthogonal functions. Then the expansion coefficients and the derivatives of the basis functions are used to compute the derivatives of $f(x)$. A given accuracy of the derivatives by spectral methods can be obtained with significantly fewer grid points compared to finite difference methods.

6 Summary of GW Detection

The search for GW was started long back in 1960s using resonance bar detectors [29]. Subsequently, in early 1970s Weiss et al. [30–32] used laser interferometers for the detection of GW. In the interferometer a laser beam is split into two sub-beams that travel down orthogonal arms, bounce off mirrors and then recombine on return. So if the light travel time is the same then there will be constructive recombination of light beams. But due to propagation of gravitational wave the light travel time will not be identical in each arm and as a result interference will occur. Interferometer measures this small time difference very accurately to get ideas about the GW and in turn about the properties of the source of GW.

Initially, the Laser Interferometer Gravitational wave Observatory (iLIGO) started operation in the early 2000s. In fact there are two LIGO facilities in operation (one at Hanford, Washington and the other at Livingstone and Louisiana). Recently, there is Italian collaboration (Virgo) and a Japanese extension (KAGRA) both of which will be in operation soon. India also joins in this group and LIGO-India will be online in 2020s. The use of multiple detectors to obtain redundancy and to increase the confidence of a detection by observing the signal through independent detectors with uncorrelated noise. In spite of being four order more sensitive over a wide frequency band than Weber's original instrument, unfortunately iLIGO was not able to detect GW.

In late 2000s iLIGO was upgraded as follows:

(a) an increase in the laser power to reduce quantum noise

(b) larger and heavier mirrors to reduce thermal and radiation pressure noise

(c) better suspension fibers for the mirrors to reduce suspension thermal noise.

And also there are other technological developments. This upgraded Laser Interferometer GW observatory is then termed the advanced LIGO (aLIGO) and it started detection in 2015 with $3-4$ times more sensitive then the previous one. Surprisingly, within a few days of first scientific operation (on 14th September, 2015), aLIGO detectors noted interference pattern related to GW produced due to merger of two BHs far away at a distance 1.3 billion light years [33]. Amazingly, the signal was so prominent (compared to the noise) that the probability that the recorded event was a GW is much larger than 5σ (i.e. probability of false alarm $\ll 10^{-7}$). GW was detected second time in the same year on 26th December [14].

Thus the GW detection by aLIGO shows first direct evidence of the existence of BH binaries and their coalescence. Also the observation shows that BHs truly merge in nature within an amount of time smaller than the age of the universe.

Finally, aLIGO observations demonstrate that GR is not only highly accurate in the solar system but also in extreme gravity scenarios when gravitational interaction is strong, highly non-linear, and highly dynamical. Consequently, any modified gravity theory is ruled out in the context of formulation of quantum gravity theory. Further, it is expected that future GW detection will help us to verify other predictions of GR namely (i) gravitational interaction is parity invariant, (ii) GW propagates at the speed of light and (iii) GW is characterized by two transverse polarizations. Therefore, GW detection is not only a remarkable confirmation of Einstein's gravity theory, it opens a new era in astrophysics.

References

1. Einstein, A.: Näherungsweise integration der Feldgleichungen der gravitation. SAW, pp. 688–696. Sitzungesber K. Preuss. Akad. Wiss., Berlin (1916)
2. Choquet-Bruhat, Y.: Théoréme d'existence pour certain systémes d'equations aux dérivées partielles nonlinéaires. Acta Math. **88**, 141 (1968)
3. Choquet-Bruhat, Y., Geroch, R.: Global aspects of the Cauchy problem in general relativity. Commun. Math. Phys. **14**, 329–335 (1969)
4. Christodoulou, D., Klainerman, S.: The Global Non-linear Stability of the Minkowski Space. Princeton Mathematical Series, vol. 41. Princeton University Press, Princeton (1993)
5. Klainerman, S., Rodnianski, I., Szeftel, J.: Invent. Math. **202**, 1 (2015)
6. Klainerman, S., Rodnianski, I., Szeftel, J.: arXiv:1204.1772 [math.AP]
7. Szeftel, J.: arXiv:1204.1768 [math.AP]
8. Szeftel, J.: arXiv:1204.1769 [math.AP]
9. Szeftel, J.: arXiv:1204.1770 [math.AP]
10. Szeftel, J.: arXiv:1204.1771 [math.AP]
11. Zel'dovich, Ya.B., Polnarev, A.G.: Astron. Zh. **51**, 30 (1974)
12. Christodoulou, D.: Phys. Rev. Lett. **67**, 1486 (1991)
13. Bieri, L., Garfinkle, D.: Phys. Rev. D **89**, 084039 (2014)
14. Bieri, L., Garfinkle, D., Yunes, N.: arXiv:1710.03272 [gr-qc]
15. Blanchet, L.: Living Rev. **9**, 4 (2006)
16. Poisson, E., Will, C.M.: Gravity. Cambridge University Press, Cambridge (2014)
17. Will, C.M., Wiseman, A.G.: Phys. Rev. D **54**, 4813 (1996)
18. Will, C.M.: Prog. Theoret. Phys. Suppl. **136**, 158 (1999)
19. Pati, M.E., Will, C.M.: Phys. Rev. D **62**, 124015 (2000)
20. Blanchet, L., Faye, G.: J. Math. Phys. **41**, 7675 (2000)
21. Damour, T., Jaranowski, P., Schaofer, G.: Phys. Lett. B **513**, 147 (2001)
22. Damour, T.: Fundam. Theor. Phys. **9**, 89 (1984)
23. Babak, S., Taracchini, A., Buonanno, A.: Phys. Rev. D **95**, 024010 (2017)
24. Press, W., Teukolsky, S., Velterling, W., Flannery, B.: Numerical Recipes in Fortran. Cambridge University Press, Cambridge (1992)
25. Pretorius, F.: Phys. Rev. Lett. **95** (2005)
26. Camponelli, M., Lousto, C.O., Marronetti, B., Zlochower, Y.: Phys. Rev. Lett. **96** (2006)
27. Baker, J.G., et al.: Phys. Rev. Lett. **96** (2006)
28. Mroue, A., et al.: Phys. Rev. Lett. **111**, 241104 (2013)
29. Weber, J.: Phys. Rev. Lett. **24**, 276 (1970)
30. Weiss, R.: MIT Report No. 105 (1972)
31. Press, W., Thorne, K.: Annu. Rev. Astron. Astrophys. **10**, 335 (1972)
32. Drever, R.W.P., et al.: Laser interferometer gravitational-wave observatory (LIGO). Technical report (1989)
33. Abbott, B.P., et al. [LIGO Scientific and Virgo collaborations]: Phys. Rev. Lett. **116**, 061102 (2016)

Current Trends in the Bifurcation Methods of Solutions of Real World Dynamical Systems

Jocirei D. Ferreira and V. Sree Hari Rao

Abstract In this survey paper we discuss a class of dynamical systems giving rise to three types of bifurcations: Hopf bifurcation, Turing bifurcation and Zip bifurcation. We present methods from dynamical systems theory that help to classify the studied bifurcations. As application of the proposed methods, a class of predator-prey systems are analyzed. Special attention is bestowed to a class of models describing the interactions of two predator species competing for one prey. Under certain natural assumptions, it has been observed that the models admit a one dimensional continuum of equilibria leading to, what is described as a zip bifurcation phenomenon. The models presented in this survey research for these species exhibit rich dynamics for varying values of the vital parameters involved. Conditions for the existence and stability of equilibria of the model equations are established. The effort in the article is to create an awareness among the researchers that certain real world dynamical systems often give rise to the existence of non-isolated equilibria also known as a continuous of equilibria. We propose to present methodologies that would help one to analyze systems with a continuous of equilibria.

Keywords Real world problems · Stability · Hopf bifurcation · Bogdanov takens bifurcation · Turing bifurcation · Zip bifurcation

The ordering of the authors is alphabetical since all authors contributed equally to the planning and execution of this research.

Research supported by CNPq Grants No. 445356/2014-6, Brazil.
Research supported by CNPq Grant No. 445356/2014-6 and by Foundation for Scientific Research and Technological Innovation—A constituent division of Sri Vadrevu Seshagiri Rao Memorial Charitable Trust, Hyderabad, 500102, India.

J. D. Ferreira (✉)
Institute of Exact and Earth Science, Federal University of Mato Grosso,
Barra do Garças, Brazil
e-mail: jocirei@ufmt.br

V. Sree Hari Rao
Foundation for Scientific Research and Technological Innovation,
Hyderabad 500 102, India
e-mail: vshrao@researchfoundation.in; vshrao@gmail.com

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_35

1 Introduction

Bifurcation theory is a subject with classical mathematical origin. The modern development of this subject dates back to the pioneering work of Poincaré, and over the past four decades this theory has as witnessed rapid developments with new ideas and methods from dynamical systems theory, singularity theory and many others. This subject is the mathematical study of possible changes in the qualitative or topological structure of parameter-dependent systems (families). Bifurcations occur in both continuous systems (described by ODEs, DDEs or PDEs), and discrete systems (described by maps).

It is useful to divide bifurcations into two principal classes, local and global. A local bifurcation occurs when a parameter variation causes the change in the stability of an equilibrium (or fixed point). The topological changes in the phase portrait of the system may be confined to arbitrarily small neighborhoods of the bifurcating fixed points by moving the bifurcation parameter close to the bifurcation point (hence "local"). Global bifurcations occur when "larger" invariant sets, such as periodic orbits collide with equilibria. This causes changes in the topology of the trajectories in the phase space which cannot be confined to a small neighborhood, as is the case with local bifurcations. In fact, the changes in topology extend out to an arbitrarily large distance (hence "global"). There are several important application areas of bifurcation theory. Bifurcation theory has also been applied to the study of several theoretical examples which are difficult to access experimentally.

In the recent years, many types of bifurcations have been studied and classified including saddle node bifurcations, Hopf bifurcations, Bogdanov-Takens bifurcations, umbilic bifurcations, cusp bifurcations, Turing bifurcations, zip bifurcations and others. We plan to organize a survey of thoroughly reviewed original articles that are of interest to the scientific community working in the area of bifurcation theory and applications.

A model that describes the competition of two predator species for a single regenerating prey species was introduced by Hsu et al. [1, 2] (see also Koch [3, 4]) and has been studied since then by several workers, e.g., Butler [5], Keener [6], Smith [7], and Wilken [8]. In this model of a three-dimensional system of ordinary differential equations the prey population is assumed to have a logistic growth rate in the absence of predators, and the predator populations are assumed to obey a Holling-type functional response (Michaelis-Menten kinetics). Butler [5] has shown that most of the results concerning the model of Hsu et al. can be achieved for a whole class of two-predator-one-prey models whose common feature is that the prey's growth rate and the predators' functional response are arbitrary functions satisfying certain natural conditions.

Finally, we would like to highlight that for the ODE system of three dimension studied in [1–8], the authors have considered conditions under the parameters to get and analyze isolated equilibria. Our purpose in this paper is to connect the other types of bifurcations (Hopf, Turing, etc.) with the non isolated equilibria.

2 Hopf Bifurcation

Some of the basic ideas that are used to study the existence and stability of periodic solutions will be introduced in this section. We shall give a brief notion of what is a Lyapunov coefficient and we will show its importance to conclude about the existence and stability of periodic solutions. For this first part we follow basicaly [9–11]. Furthermore, we present a summary of the projection method described in [12–14] for the calculation of the first Lyapunov coefficient associated with the Hopf bifurcation.

2.1 Poincaré Map and Displacement Function

A very strong notion in the study of periodic orbits is the Poincaré map. It is the crucial concept of the "geometric theory of differential equation". To define the Poincaré map, also called the return map, let ϕ_t be the flow of the differential equation

$$\dot{x} = f(x)$$

where $x \in \mathbb{R}^n$, f is of class C^∞ in $\mathbb{R}^n$, and suppose that $S \subset \mathbb{R}^n$ is a $(n-1)$-dimensional submanifold.

Definition 1 If $p \in S$ and $(p, f(p)) \notin T_pS$, we say that the vector $(p, f(p))$ is *transverse* to S at p. If $(p, f(p))$ is transverse to S at each $p \in S$, then S is a *section* for ϕ_t.

Remark 1 If $p \in S$, then the curve $t \mapsto \phi_t(p)$ "passes through" S as t passes through $t = 0$. It is possible to have a $T = T(p) > 0$ such that $\phi_T(p) \in S$. In this case, we say that p returns to S at time T.

Definition 2 If there exists an open set $\sum \subseteq S$ such that each point in $\sum$ returns to S, then $\sum$ is called a *Poincaré section*.

Definition 3 Define the map $P : \sum \to S$ by $P(p) = \phi_{T(p)}(p)$ where $T(p) > 0$ is the time of first return to S. The map P is called *Poincaré map* (also known as *return map or succession function*) on $\sum$ and $T : \sum \to \mathbb{R}$ is called the *return time map*. The functions P and T are smooth functions on $\sum$ since the solution of the differential equation is smoothly dependent on its initial value.

The crucial idea of Poincaré to determine periodic orbits is that fixed points of the return map belong to periodic orbits. That is, periodic points of the return map correspond to periodic solutions of the differential equation. If P denotes the return map, then p is a fixed point of P if $P(p) = p$.

Definition 4 A *periodic point* of P with *period k* is a fixed point of the kth iterate of P. It passes through the Poincaré section $k-1$ time before closing. Thus, $p \in \sum$ is a periodic point of period k if $P^k(p) = p$.

Instead of studying fixed points of P^k, it is more convenient to study the zeros of the corresponding displacement function, which we will be defined in the following.

Definition 5 The map $d : \sum \to \mathbb{R}^n$, defined by $d(p) = P^k(p) - p$ is called displacement function associated with P. The periodic points of period k for the Poincaré map correspond to the roots of the map $d(p)$.

Remark 2 An important stability theorem for periodic orbits, states that if $P^k(p) = p$ and $DP^k(p)$ has all its eigenvalues inside the unit circle, then the periodic orbit with initial point p is asymptotically stable.

2.2 Hopf Bifurcation and Lyapunov Coefficients

The main purpose of this section is to present the properties of the succession function and also define the concept of focal values and multiplicity of a multiple focus. Also, we introduce the concept of Lyapunov coefficients.

Consider the family of differential equations

$$\dot{u} = f(u, \zeta), \quad u \in \mathbb{R}^n, \quad \zeta \in \mathbb{R}^m, \tag{1}$$

where ζ is a vector of parameters. Assume that f is of class $\mathbb{C}^\infty$ in $\mathbb{R}^n \times \mathbb{R}^m$.

Definition 6 An equilibrium state (x_0, ζ_0) of the dynamical system (1) with pure imaginary characteristic roots is called a focus (either stable or unstable), a center, or a center-focus. An equilibrium state of a dynamical system wich has pure imaginary characteristic roots and is a focus will be called a multiple focus (for more details see [9] and [10] pp. 93).

Definition 7 An ordered pair $(x_0, \zeta_0) \in \mathbb{R}^n \times \mathbb{R}^m$ is called a *Hopf point* for the system (1) if there is a curve C in $\mathbb{R}^n \times \mathbb{R}^m$, that is given by $\epsilon \mapsto (c_1(\epsilon), c_2(\epsilon))$ and satisfies the properties:

(i) $C(0) = (x_0, \zeta_0)$ and $f(c_1(\epsilon), c_2(\epsilon)) = 0$.

(ii) The linear transformation given by the derivative $f_u(c_1(\epsilon), c_2(\epsilon)) : \mathbb{R}^n \to \mathbb{R}^n$ has a pair of nonzero complex conjugate eigenvalues $\alpha(\zeta) \pm i\beta(\zeta)$, each with algebraic (and geometric) multiplicity one. Also, $\alpha(0) = 0$, $\alpha'(0) \neq 0$ and $\beta(0) \neq 0$.

(iii) Except for the eigenvalues $\pm i\beta(0)$, all other eigenvalues of $f_u(x_0, \zeta_0)$ have nonzero real parts.

Remark 3 It is sufficient to consider the bifurcations for a planar family of differential equations associated with the family (1) since, as we will show in the next section, there exists a center manifold reduction method for the family (1) that produces a family of planar differential equations

$$\dot{u} = f(u, \zeta), \quad u \in \mathbb{R}^2, \quad \zeta \in \mathbb{R}^m, \tag{2}$$

with a corresponding Hopf point. Moreover, there is a neighborhood $U \times V \subset \mathbb{R}^n \times \mathbb{R}^m$ of the Hopf point (x_0, ζ_0) such that if $\zeta \in V$ and the corresponding member of the family (2) has a bounded orbit in U, then this same orbit is an invariant set for the corresponding member of the planar family (1).

Definition 8 The planar family (2) has a *supercritical Hopf bifurcation* at a Hopf point with associated curve $\epsilon \to (c_1(\epsilon), c_2(\epsilon))$ if there are positive numbers ϵ_0, K_1 and K_2 such that for each ϵ in the open interval $(0, \epsilon_0)$ the differential equation $\dot{u} = f(u, c_2(\epsilon))$ has a hyperbolic limit cycle with radii

$$(K_1\sqrt{\epsilon} + O(\epsilon), K_2\sqrt{\epsilon} + O(\epsilon))$$

relative to the rest point $u = c_1(\epsilon)$. The bifurcation is called *subcritical* if there is a similar limit cycle for the system with parameter values in the range $-\epsilon_0 < \epsilon < 0$. Also, we say that the family (1) has a supercritical (respectively, subcritical) Hopf bifurcation at a Hopf point if the corresponding (center manifold) reduced system (2) has a supercritical (respectively, subcritical) Hopf bifurcation.

Observe that, after the translation $v = u - c_1(\epsilon)$, the differential equation (2) becomes

$$\dot{v} = f(v + c_1(\epsilon), \zeta)$$

with $f(0 + c_1(\epsilon), c_2(\epsilon)) \equiv 0$. In particular, in the new coordenates, the associated rest points remain at the origin for all values of the parameter ϵ. Thus it suffices to consider the family (2) to be of the form

$$\dot{u} = f(u, \zeta), \quad u \in \mathbb{R}^2, \quad \zeta \in \mathbb{R}, \tag{3}$$

only now with a Hopf point at $(u, \zeta) = (0, 0) \in \mathbb{R}^2 \times \mathbb{R}$ and with associated curve c given by $\zeta \to (0, \zeta)$.

Remark 4 If $(u, \zeta) = (0, 0) \in \mathbb{R}^2 \times \mathbb{R}$ is a Hopf point for the family (3) with eigenvalues $\alpha(\zeta) \pm i\beta(\zeta)$, there exists a smooth parameter-dependent linear change of coordenates that transforms the system (3) into the form

$$\begin{cases} \dot{x} = \alpha(\zeta)x - \beta(\zeta)y + \varphi(x, y, \zeta) \\ \dot{y} = \beta(\zeta)x + \alpha(\zeta)y + \psi(x, y, \zeta) \end{cases} \tag{4}$$

where the functions φ and ψ together with their first partial derivatives with respect to the space variables vanishe at the origin, $\alpha(0) = 0$, $\alpha'(0) \neq 0$ and $\beta(0) \neq 0$ (for details see [11]).

Changing the family (4) to polar coordinates we obtain the family

$$\begin{cases} \dot{\rho} = F(\rho, \theta, \zeta) \\ \quad = \alpha(\zeta)\rho + \varphi(\rho\cos\,\theta, \rho\sin\,\theta, \zeta)\cos\,\theta + \psi(\rho\cos\,\theta, \rho\sin\,\theta, \zeta)\sin\,\theta \\ \dot{\theta} = \beta(\zeta) + \Phi(\rho, \theta, \zeta) \\ \quad = \beta(\zeta) + \dfrac{\psi(\rho\cos\,\theta, \rho\sin\,\theta, \zeta)}{\rho}\cos\,\theta - \dfrac{\varphi(\rho\cos\,\theta, \rho\sin\,\theta, \zeta)}{\rho}\sin\,\theta \end{cases} \tag{5}$$

and then the equation

$$\frac{d\rho}{d\theta} = \frac{F(\rho, \theta, \zeta)}{\beta(\zeta) + \Phi(\rho, \theta, \zeta)} \tag{6}$$

where it is assumed that $\Phi(0, \theta, \zeta) = 0$ for all θ which ensures the continuity of the function $\Phi(\rho, \theta, \zeta)$.

Definition 9 Let $\rho = \chi(\theta, \theta_0, \rho_0, \zeta)$ the solution of the differential equation (6) with initial conditions

$$\chi(\theta_0, \theta_0, \rho_0, \zeta) = \rho_0.$$

We define the succession function for system (6) as

$$\rho = \chi_{\theta_0}(\rho_0, \zeta) = \chi(\theta_0 + 2\pi, \theta_0, \rho_0, \zeta) \tag{7}$$

on the ray $\theta = \theta_0$. Also, we define the displacemente function for Eq. (6) as

$$d_{\theta_0}(\rho_0, \zeta) = \chi_{\theta_0}(\rho_0, \zeta) - \rho_0. \tag{8}$$

where for $\theta = 0$ we denote the function $\chi(\rho_0, \zeta)$ and $d(\rho_0)$ respectively, and

$$\rho = \chi(\rho_0, \zeta) = \chi(2\pi, 0, \rho_0, \zeta) \tag{9}$$

$$d(\rho_0, \zeta) = \chi(\rho_0, \zeta) - \rho_0. \tag{10}$$

Remark 5 The solution $\rho = \chi(\theta, \theta_0, \rho_0, \zeta)$ of Eq. (6) has continuous partial derivatives with respect to ρ_0 up to order N (for details see [9] and [10] pp. 243). Hence, the functions $\chi(\rho_0, \zeta)$ and $d(\rho_0, \zeta)$ are continuously differentiable N times.

Definition 10 Suppose that $(u, \lambda) = (0, 0) = O$ is a Hopf point for the family (3), that is, system (3) can be transformed in the form (4) and has pure imaginary characteristic roots at $(x, y, \zeta) = (0, 0, 0)$. The value of the i-th derivative of the function $d(\rho_0)$ at the point O, that is, $d^{(i)}(0, 0)$, is called i-th focal value of the focus O.

Theorem 1 *If there exists k such that*

$$d'(0,0) = d''(0,0) = \cdots = d^{(k-1)}(0,0) = 0 \text{ and } d^{(k)}(0,0) \neq 0, \qquad (11)$$

k is an odd number, where the derivative is with respect to ρ_0.

Proof See [10]. □

Definition 11 If conditions (11) are satisfied, and $k = 2m+1, m \geq 0$ then the focus O is a focus of multiplicity m.

Remark 6 It can be proved (see [10] pp. 93) that

$$d'(0,0) = e^{2\pi\frac{\alpha}{\beta}} - 1. \qquad (12)$$

Then, if $m = 0$ it implies that $k = 1, d'(0,0) \neq 0$ and $\alpha(0) \neq 0$. But then the focus O has complex, though not pure imaginary characteristic roots so it is a simple focus. Conversely, if $m > 0$, then $k \geq 3$, $d'(0,0) = 0$, $\alpha(0) = 0$ and the caracteristic roots are pure imaginary, that is the focus is multiple according with Definition 6. Note that a multiple focus does not always have a defined multiplicity since, if system (4) is of class N, but not of class $N+1$, and if $d'(0,0) = d''(0,0) = \cdots = d^{(N)}(0,0) = 0$, Definition 6 is not applicable.

Definition 12 The first nonzero focal value of a multiple focus is called the *first Lyapunov value* (or *Lyapunov coefficient*). In other words, the Lyapunov value is the number $d^k(0,0)$ provided that relations (11) holds and $k \geq 3$. If $k = 2m+1$, the Lyapunov value will also be called the m-th Lyapunov value ($m \geq 1$).

Remark 7 From now on, we denote the m-th Lyapunov value by $l_m(0)$, $m \geq 1$.

Remark 8 Other equivalent definitions and algorithmic procedures to write the expressions for the Lyapunov coefficients l_m, for two dimensional systems can be found in Gasull and Torregrosa [15] and Kuznetsov [14], among others. These procedures apply also to the n-dimensional systems, what is our study in the next section.

2.3 Reduction to the Center Manifold and the Hopf Bifurcation

In this section, we present a summary of the projection method described in [12–14] for the calculation of the first Lyapunov coefficient associated with the Hopf bifurcation. This helps to conclude that the equilibria of a studied dynamical system undergoes a *Hopf bifurcation*. Other equivalent definitions and algorithmic procedures to write the expressions of the Lyapunov coefficients for two-dimensional systems can be found in [10, 16], among others, and can be adapted to n-dimensional systems with $n \geq 3$, if it is restricted to the center manifold. We present the projection method for the case $n = 4$ since, as an application example, we analyze a four dimensional predator-prey model that undergoes a Hopf bifurcation. The procedure can be generalized easily for the cases $n > 4$.

Consider the differential equation

$$\mathbf{x}' = f(\mathbf{x}, \zeta), \tag{13}$$

where $\mathbf{x} \in \mathbb{R}^4$ is a vector representing phase variables and $\zeta \in \mathbb{R}$ is a parameter representing control parameter. Assume that f is of class $\mathbb{C}^\infty$ in $\mathbb{R}^4 \times \mathbb{R}$. Suppose that (13) has an equilibrium point $\mathbf{x} = \mathbf{x}_0$ at $\zeta = \zeta_0$ and, denoting the variable $\mathbf{x} - \mathbf{x}_0$ also by $\mathbf{x}$, we write

$$F(\mathbf{x}) = f(\mathbf{x}, \zeta_0) \tag{14}$$

in the form

$$F(\mathbf{x}) = \mathbf{J}\mathbf{x} + \frac{1}{2}B(\mathbf{x}, \mathbf{x}) + \frac{1}{6}C(\mathbf{x}, \mathbf{x}, \mathbf{x}) + O(\|x\|^4), \tag{15}$$

where $\mathbf{J} = f_{\mathbf{x}}(\mathbf{0}, \zeta_0)$ and $B(\mathbf{x}, \mathbf{y})$, $C(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are *multilinear functions*. In coordinates, we have

$$B_i(\mathbf{x}, \mathbf{y}) = \sum_{j,k=1}^{4} \left.\frac{\partial^2 F_i(\xi)}{\partial \xi_j \xi_k}\right|_{\xi=0} x_j y_k, \qquad C_i(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{j,k,l=1}^{4} \left.\frac{\partial^3 F_i(\xi)}{\partial \xi_j \xi_k \xi_l}\right|_{\xi=0} x_j y_k z_l,$$

where $i = 1, 2, 3, 4$.

Suppose that $(\mathbf{x}, \zeta_0)$ is an equilibrium point of (13) where the Jacobian matrix $\mathbf{J}$ has a pair of pure imaginary eigenvalues $\lambda_{2,3} = \pm i\omega_0$, $\omega_0 > 0$, and these eigenvalues are the only pure imaginary eigenvalues. Let $\mathbf{T}^c$ be the generalized eigenspace of $\mathbf{J}$ corresponding to $\lambda_{2,3}$. By this it is meant the largest subspace invariant by $\mathbf{J}$ on which the eigenvalues are $\lambda_{2,3}$.

Let $\mathbf{p}, \mathbf{q} \in \mathbb{C}^4$ be *complex* eigenvectors corresponding to $-i\omega_0$ and $+i\omega_0$ respectively, such that

$$\mathbf{J}\mathbf{q} = i\omega_0\mathbf{q}, \qquad \mathbf{J}^T\mathbf{p} = -i\omega_0\mathbf{p}, \qquad \langle \mathbf{p}, \mathbf{q}\rangle = 1, \tag{16}$$

where $\mathbf{J}^T$ is the transpose of the matrix $\mathbf{J}$ and $\langle \mathbf{p}, \mathbf{q}\rangle = \sum_{i=1}^4 \overline{p}_i q_i$ is the standard scalar product in $\mathbb{C}^4$ (linear with respect to the second argument). Thus, the critical real eigenspace $\mathbf{T}^c$ corresponding to $\pm i\omega_0$ is now two-dimensional, is spanned by $\{\mathbb{R}e\ \mathbf{q}, \mathbb{I}m\ \mathbf{q}\}$ and any vector $\mathbf{y} \in \mathbf{T}^c$ can be represented as $\mathbf{y} = w\mathbf{q} + \overline{w\mathbf{q}}$, where $w = \langle \mathbf{p}, \mathbf{y}\rangle \in \mathbb{C}$.

The two-dimensional center manifold associated with the eigenvalues $\lambda_{2,3} = \pm i\omega_0$ can be parameterized by the variables w and $\overline{w}$ by means of an immersion of the form $\mathbf{x} = H(w, \overline{w})$, where $H : \mathbb{C}^2 \to \mathbb{R}^4$ has a Taylor expansion of the form

$$H(w, \overline{w}) = w\mathbf{q} + \overline{w\mathbf{q}} + \sum_{2 \le j+k \le 3} \frac{h_{jk}}{j!k!} w^j \overline{w}^k + O(|w|^4), \tag{17}$$

with $\mathbf{h}_{jk} \in \mathbb{C}^4$ and $\mathbf{h}_{jk} = \overline{\mathbf{h}}_{kj}$. Substituting this expression into (13) we obtain the following differential equation

$$H_w w' + H_{\overline{w}} \overline{w}' = F(H(w, \overline{w})), \tag{18}$$

where F is given by (14). The complex vectors $\mathbf{h}_{ij}$ are obtained solving the system of linear equations defined by the coefficients of (18), taking into account the coefficients of F, so that system (18), on the chart w for a central manifold, can be written as follows

$$w' = i\omega_0\omega + \frac{1}{2}G_{21}w|w|^2 + O(|w|^4),$$

with $G_{21} \in \mathbb{C}$.

Definition 13 The first *Lyapunov coefficient* l_1 is defined by

$$l_1 = \frac{1}{2}\mathbb{R}e\ G_{21}, \tag{19}$$

where $G_{21} = \langle p, \mathcal{H}_{21}\rangle$, $\mathcal{H}_{21} = C(\mathbf{q}, \mathbf{q}, \overline{\mathbf{q}}) + B(\overline{\mathbf{q}}, \mathbf{h}_{20}) + 2B(\mathbf{q}, \mathbf{h}_{11})$, $\mathbf{h}_{11} = -\mathbf{J}^{-1} B(\mathbf{q}, \overline{\mathbf{q}})$, $\mathbf{h}_{20} = (2i\omega_0\mathbf{I}_4 - \mathbf{J})^{-1}B(\mathbf{q}, \mathbf{q})$ and $\mathbf{I}_4$ is the unit 4×4 matrix.

Proposition 1 *Consider system (13) where $\mathbf{x} \in \mathbb{R}^4$ is a vector representing phase variables and $\zeta \in \mathbb{R}$ is a parameter representing control parameter. Assume that f is of class $\mathbb{C}^\infty$ in $\mathbb{R}^4 \times \mathbb{R}$. Suppose that (13) has an equilibrium point $\mathbf{x} = \mathbf{x}_0$ at $\zeta = \zeta_0$. Then the following holds:*

(i) Suppose that at $(\mathbf{x}_0, \zeta_0)$ the Jacobian matrix $\mathbf{J}$ of (13) has a pair of pure imaginary eigenvalues $\lambda_{2,3}(\zeta_0) = \pm i\omega_0$, $\omega > 0$, and these eigenvalues are the only eigenvalues with null real part. Then $(\mathbf{x}_0, \zeta_0)$ is a Hopf point of (13).

(ii) At $(\mathbf{x}_0, \zeta_0)$ a two-dimensional center manifold is well-defined, it is invariant under the flow generated by (13) and can be continued with arbitrary high class of differentiability to nearby parameter values.

(iii) If $\frac{d}{d\zeta}\mathbb{R}e\ \lambda(\zeta)|_{\zeta=\zeta_0} \neq 0$ then $(\mathbf{x}_0, \zeta_0)$ is a transversal Hopf point. In this case, if $l_1 \neq 0$ then in a neighborhood of $(\mathbf{x}_0, \zeta_0)$ the dynamic behavior of the system (13), reduced to the family of parameter-dependent continuations of the center manifold, is orbitally topologically equivalent to the following complex normal form

$$w' = (\eta + i\omega)w + l_1 w|w|^2,$$

where $w \in \mathbb{C}$, η, ω and l_1 are real functions having derivatives of arbitrarily higher order, which are continuations of 0, ω_0 and the first Lyapunov coefficient at the Hopf point. When $l_1 < 0$ ($l_1 > 0$) a family of stable (unstable) periodic orbits can be found in these manifolds, shrinking to an equilibrium point at the Hopf point.

Proof See [14]. □

In the next example, we apply the projection method discussed in this section to show that a four dimensional predator-prey ODE model undergoes a Hopf bifurcation. For a detailed study of the model we refer the readers to [17].

Example 1 Consider the system

$$\begin{aligned}\dot{N} &= \tfrac{4\varepsilon}{K} N^2 \left(1 - \tfrac{N}{K}\right) - \alpha N P \\ \dot{P} &= -\gamma P + \beta P \textstyle\int_{-\infty}^{t} N(\tau)\, G(t-\tau)\, d\tau,\end{aligned} \tag{20}$$

that describes the dynamical interactions between a predator and a prey species with a weak Allee effect in the prey population $N(t)$ and a distributed delay $G(t-\tau)$ in the predator population $P(t)$, at time t. The parameters $K, \varepsilon, \alpha, \gamma$ and β are positive. K represents the carrying capacity of the environment with respect to the prey, ε is the intrinsic growth rate of the prey, α is the rate of predation, γ is the intrinsic mortality of the predator and β is the conversion rate.

In order to determine the equilibria for the system (20), we rewrite (20) in the following form

$$\begin{aligned}\dot{N}(t) &= \frac{4\varepsilon}{K} N^2(t) \left(1 - \frac{N(t)}{K}\right) - \alpha N(t)\, P(t) \\ \dot{P}(t) &= -\gamma P(t) + \beta P(t)\, Q(t) \\ \dot{Q}(t) &= a\,(N(t) - Q(t)) \\ \dot{R}(t) &= a\,(N(t) - R(t)),\end{aligned} \tag{21}$$

where

$$\begin{aligned}Q(t) &= a^2 \int_{-\infty}^{t} (t-\tau) N(\tau) \exp(-a(t-\tau))\, d\tau \\ R(t) &= a \int_{-\infty}^{t} N(\tau) \exp(-a(t-\tau))\, d\tau, \quad t \geq 0.\end{aligned}$$

The equivalence of (20) and (21) on $\mathbb{R}_+$ is to be understood in the following sense. Let $\overline{C}^0(-\infty, 0]$ denote the set of real valued functions that are continuous and bounded on $(-\infty, 0]$. Let $(N, P) : [0, \infty) \to \mathbb{R}^2$ be the solution of (20) that corresponds to the "*initial function*" $\widetilde{N} \in \overline{C}^0(-\infty, 0]$ and the initial value $P_0 = P(0)$ that is, on $(-\infty, 0)$, N is considered to be equal to $\widetilde{N}$; then $(N, P, Q, R) : [0, \infty) \to \mathbb{R}^4$ is the solution of (21) that satisfies the initial conditions $N_0 = \widetilde{N}$, $P_0 = P(0)$ and

$$\begin{aligned}Q(0) = Q_0 &= a^2 \int_{-\infty}^{0} (-\tau) \widetilde{N}(\tau) \exp(-a(-\tau))\, d\tau \\ R(0) = R_0 &= a \int_{-\infty}^{0} \widetilde{N}(\tau) \exp(-a(-\tau))\, d\tau.\end{aligned}$$

The change of variables,

$$N = Kn, \quad P = Kp, \quad Q = Kq \quad R = Kr \quad and \quad t = \frac{s}{\varepsilon},$$

transforms the system (21) into

$$\begin{aligned} \mathbf{x}' &= f(\mathbf{x}, K) \\ &= \left(4n^2(1-n) - \frac{\alpha K}{\varepsilon} np, -\frac{\gamma}{\varepsilon} p + \frac{\beta K}{\varepsilon} pq, \frac{a}{\varepsilon}(r-q), \frac{a}{\varepsilon}(n-r)\right), \end{aligned} \tag{22}$$

where the prime represents the derivative respect to s, $\mathbf{x} = (n, p, q, r) \in \mathbb{R}^4$ and $K \in (0, \infty)$. It is easy to see that system (22) has the following equilibria

$$\begin{aligned} &E_0 = (0, 0, 0, 0), \quad E_1 = (1, 0, 1, 1) \quad \text{and} \\ &E_2 = \left(\frac{\gamma}{\beta K}, \frac{4\varepsilon\gamma(\beta K - \gamma)}{\alpha\beta^2 K^3}, \frac{\gamma}{\beta K}, \frac{\gamma}{\beta K}\right). \end{aligned}$$

For the system (22) the equilibrium E_0 is unstable for all $K > 0$. On the other hand, the equilibrium E_1 is locally asymptotically stable if

$$K < \frac{\gamma}{\beta} \text{ and unstable if } K > \frac{\gamma}{\beta}. \tag{23}$$

In the following, we study the stability of the equilibrium E_2. In this case, the characteristic polynomial associated with the Jacobian matrix of system (22) evaluated at E_2 is given by

$$\begin{aligned} p(\lambda) = \lambda^4 &+ \left[\frac{2a}{\varepsilon} + \frac{4\gamma}{K\beta}\left(\frac{2\gamma}{K\beta} - 1\right)\right]\lambda^3 + \left[\frac{a^2}{\varepsilon^2} + \frac{8a\gamma}{K\beta\varepsilon}\left(\frac{2\gamma}{K\beta} - 1\right)\right]\lambda^2 \\ &+ \frac{4a^2\gamma}{K\beta\varepsilon^2}\left(\frac{2\gamma}{K\beta} - 1\right)\lambda + \frac{4a^2\gamma^2}{K\beta\varepsilon^3}\left(1 - \frac{\gamma}{K\beta}\right). \end{aligned}$$

The critical point E_2 is unstable for the system (22) if

$$a < \frac{1}{2}\frac{K\beta - \gamma}{2\gamma - K\beta}\gamma \tag{24}$$

and it is locally asymptotically stable if

$$a > 2\frac{K\beta - \gamma}{2\gamma - K\beta}\gamma. \tag{25}$$

Furthermore, if

$$\frac{1}{2}\frac{K\beta-\gamma}{2\gamma-K\beta}\gamma < a < 2\frac{K\beta-\gamma}{2\gamma-K\beta}\gamma, \tag{26}$$

then there exist an $\varepsilon_0 > 0$ such that E_2 is asymptotically stable for $\varepsilon > \varepsilon_0$ and unstable for $0 < \varepsilon < \varepsilon_0$, where

$$\varepsilon_0 = \frac{aK^2\beta^2}{2\gamma\,(2\gamma-K\beta)}\,\frac{\sqrt{2\gamma\,(K\beta-\gamma)}-\sqrt{a\,(2\gamma-K\beta)}}{2\sqrt{a\,(2\gamma-K\beta)}-\sqrt{2\gamma\,(K\beta-\gamma)}}. \tag{27}$$

To simplify the calculations we introduce the notation $b = \frac{\gamma}{K\beta}$ and rewrite conditions (26) and (27) respectively, as

$$\frac{\gamma}{2}\,\frac{1-b}{2b-1} < a < 2\gamma\,\frac{1-b}{2b-1}, \quad \varepsilon_0 = \frac{a}{2b\,(2b-1)}\,\frac{\sqrt{2\gamma\,(1-b)}-\sqrt{a\,(2b-1)}}{2\sqrt{a\,(2b-1)}-\sqrt{2\gamma\,(1-b)}\,.}$$

At $\varepsilon = \varepsilon_0$ the eigenvalues of system (22) are given by

$$\lambda_0 = \frac{-A_1 \pm \sqrt{A_1^2-4A_0}}{2} < 0, \ \ \lambda_{1,2}\,(\varepsilon_0) = \pm i\omega$$

in which

$$\begin{aligned}\omega^2 &= \frac{4\gamma^2\,(2\gamma-K\beta)^{\frac{3}{2}}\left[2\sqrt{a\,(2\gamma-K\beta)}-\sqrt{2\gamma\,(K\beta-\gamma)}\right]^2}{(K\beta)^4\sqrt{a}\left[\sqrt{2\gamma\,(K\beta-\gamma)}-\sqrt{a\,(2\gamma-K\beta)}\right]} > 0,\\ A_1 &= \frac{4\gamma\,(2\gamma-K\beta)\sqrt{a\,(2\gamma-K\beta)}}{K^2\beta^2\left[\sqrt{2\gamma\,(K\beta-\gamma)}-\sqrt{a\,(2\gamma-K\beta)}\right]} > 0\end{aligned} \tag{28}$$

$$A_0 = \frac{8\gamma^3\,(2\gamma-K\beta)^{\frac{3}{2}}\,(K\beta-\gamma)\left[2\sqrt{a\,(2\gamma-K\beta)}-\sqrt{2\gamma\,(K\beta-\gamma)}\right]}{(K\beta)^4\sqrt{a}\left[\sqrt{2\gamma\,(K\beta-\gamma)}-\sqrt{a\,(2\gamma-K\beta)}\right]^2} > 0. \tag{29}$$

The inequalities given in (28) and (29) follow assuming that $\beta K-\gamma > 0$ and $2\gamma-\beta K > 0$.

Now in accordance with (16) it is easy to check that the complex vectors

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} \dfrac{4b^2K\beta\left(1-b\right)}{\varepsilon_0\left[16b^2\left(2b-1\right)^2+\omega^2\right]} + \dfrac{16b^3K\beta\left(1-b\right)\left(2b-1\right)}{\varepsilon_0\omega\left[16b^2\left(2b-1\right)^2+\omega^2\right]}\,i \\ -\dfrac{4b\beta\left(1-b\right)}{\alpha\omega}\,i \\ 1 \\ 1+\dfrac{\varepsilon_0\omega}{a}\,i \end{bmatrix}$$

$$\widetilde{\mathbf{p}} = \begin{bmatrix} a\varepsilon_0\omega \\ -abK\alpha i \\ -\dfrac{4ab^2K\beta\varepsilon_0\left(1-b\right)\left[-\varepsilon_0\omega+ai\right]}{a^2+\varepsilon_0^2\omega^2} \\ \varepsilon_0^2\omega\left[4b\left(2b-1\right)-\omega i\right] \end{bmatrix}, \tag{30}$$

are proper eigenvectors of $\mathbf{J}$ and $\mathbf{J}^T$ respectively. To achieve the necessary normalization given in (16), we can take, for example,

$$\mathbf{p} = \frac{1}{\langle\widetilde{\mathbf{p}},\mathbf{q}\rangle}\widetilde{\mathbf{p}}. \tag{31}$$

Using the notation of the previous section (see expression (15)) the multilinear symmetric functions may be written as

$$B\left(\mathbf{x},\mathbf{y}\right) = \begin{bmatrix} 8\left(1-3b\right)x_1y_1 - \frac{K\alpha}{\varepsilon_0}x_1y_2 - \frac{K\alpha}{\varepsilon_0}x_2y_1 \\ \frac{K\beta}{\varepsilon_0}x_3y_2 + \frac{K\beta}{\varepsilon_0}x_2y_3 \\ 0 \\ 0 \end{bmatrix},$$

$$C\left(\mathbf{x},\mathbf{y},\mathbf{z}\right) = \begin{bmatrix} -24x_1y_1z_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{32}$$

The real part, $\eta = \eta(\varepsilon)$, of the pair of complex eigenvalues at the critical parameter $\varepsilon = \varepsilon_0$ verifies

$$\eta'\left(\varepsilon_0\right) = -\frac{1}{\varepsilon_0^2}\,\frac{ux+vy}{x^2+y^2} < 0 \tag{33}$$

where

$$\begin{aligned} u &= a\left[a\varepsilon_0\omega^2 - 6ab\gamma\left(1-b\right) + 4b\varepsilon_0^2\omega^2\left(2b-1\right)\right], \\ v &= a\varepsilon_0\omega\left[\varepsilon_0\omega^2 - 4ab\left(2b-1\right)\right] \\ x &= 2b\left(2b-1\right)\left(a^2-3\varepsilon_0^2\omega^2\right) - 3a\varepsilon_0\omega^2, \\ y &= \omega\left[a^2 - 2\varepsilon_0^2\omega^2 + 8ab\varepsilon_0\left(2b-1\right)\right]. \end{aligned}$$

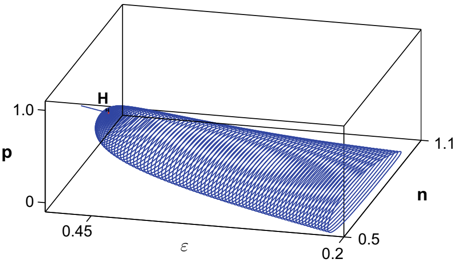


Fig. 1 Bifurcation diagram for $0.2 < \varepsilon < 0.5$ in ε, n, p space where **H** is the bifurcation point

Accordingly, we have the following theorem.

Theorem 2 *Consider the one-parameter family of differential equations (22) and suppose that (26) holds. The first Lyapunov coefficient associated with the equilibrium E_2 is given by*

$$l_1(\varepsilon_0) = \frac{1}{2\omega} Re\left\{ -24\bar{p}_1 q_1 q_1 \bar{q}_1 - \frac{2a\varepsilon_0}{bH}\left[\omega q_1 - \frac{bK\beta}{\varepsilon_0}\right] \times \left[8(1-3b)q_1\bar{q}_1 - \frac{K\alpha}{\varepsilon_0}(q_1\bar{q}_2 + \bar{q}_1 q_2)\right] + \frac{a\varepsilon_0\omega}{H}\left[8(1-3b)\bar{q}_1 r_1 - \frac{K\alpha}{\varepsilon_0}(\bar{q}_1 r_2 + \bar{q}_2 r_1)\right] + \frac{abK^2\alpha\beta}{H\varepsilon_0}[r_2 + \bar{q}_2 r_3] \right\}$$

where $\mathbf{r} = (2i\omega\mathbf{I}_4 - \mathbf{J})^{-1} B(\mathbf{q}, \mathbf{q}) = [r_1, r_2, r_3, r_4]^T$, $\mathbf{I}_4$ is the unit 4×4 matrix and $\mathbf{J} = \mathbf{J}(E_2)$ is the Jacobian matrix of system (22) evaluated at E_2. If $l_1(\varepsilon_0)$ is different from zero then the one-parameter family of differential equations (22) has a transversal Hopf point at E_2 for $\varepsilon = \varepsilon_0$, i.e a Hopf point of codimension 1 at E_2. If $l_1(\varepsilon_0) < 0$ then for $\varepsilon > \varepsilon_0$ the equilibrium point E_2 is asymptotically stable and for each $0 < \varepsilon < \varepsilon_0$, but close to ε_0, there exists a asymptotically orbitally stable periodic orbit around the unstable equilibrium point E_2. If $l_1(\varepsilon_0) > 0$ then for $0 < \varepsilon < \varepsilon_0$ the equilibrium point E_2 is unstable and for each $\varepsilon > \varepsilon_0$, but close to ε_0, there exists an unstable periodic orbit around the asymptotically stable equilibrium point E_2.

Using the software MATCONT we illustrate the occurrence of Hopf bifurcation for the equilibrium $E_2 = (\frac{\gamma}{\beta K}, \frac{4\varepsilon\gamma(\beta K-\gamma)}{\alpha\beta^2 K^3}, \frac{\gamma}{\beta K}, \frac{\gamma}{\beta K})$. If we consider system (22) with values of parameters: $K = 1.0$, $\alpha = 0.5$, $\gamma = 0.4$, $\beta = 0.6$, $a = 0.5$, we get $l_1 \approx -0.01735$ which shows that the periodic orbit generated by the Hopf bifurcation is supercritical. In this case, we obtain Fig. 1 (the projection on the ε, n, p-space of the bifurcation diagram) illustrating the occurrence of stable periodic orbit around the unstable equilibrium E_2 when ε changes on the interval $0.2 < \varepsilon < 0.5$. Note that we have the presence of stable periodic orbit up to the bifurcation point $\varepsilon_0 \approx 0.4054270614$ which is denoted by **H** in Fig. 1.

3 Turing Bifurcation

The basic reaction-diffusion theory of morphogenesis has been put forward in the classical article written by the English mathematician Turing [18] describing the way in which non-uniformity (stripes, spots, spirals, etc.) may arise naturally out of a homogeneous, uniform state. The theory (which can be called a reaction-diffusion theory of morphogenesis), has served as a basic model in theoretical biology, and is seen by some as the very beginning of chaos theory. Morphogenesis is a biological process that causes an organism to formulate its shape in the development of pattern and form. It is one of three fundamental aspects of developmental biology along with the control of cell growth and cellular differentiation.

Spatial ecology addresses the fundamental effects of space on the dynamics of individual species and on the structure, dynamics, diversity, and stability of multispecies communities. Essentially, this subject is designed to highlight the importance of space in the areas of stability, patterns of diversity, invasions, coexistence, and pattern generation. The mathematical formulation of the ideas dealing with the spacial aspect of species leads to reaction diffusion models. For more interesting account of various aspects and examples in population ecology we refer the readers to [19].

Reaction-diffusion systems have attracted much interest as a prototype model for pattern formation. The above-mentioned patterns (fronts, spirals, targets, hexagons, stripes and dissipative solitons) can be found in various types of reaction-diffusion systems in spite of large discrepancies e.g. in the local reaction terms. It has also been argued that reaction-diffusion processes are an essential basis for processes connected to animal coats and skin pigmentation [20, 21]. Another reason for the interest in reaction-diffusion systems is that although they represent nonlinear partial differential equation, there are often possibilities for an analytical treatment.

The science of *pattern formation* deals with the visible, (statistically) orderly outcomes of self-organization and the common principles behind similar patterns in nature. In developmental biology, pattern formation refers to the generation of complex organizations of cell fates in space and time. Pattern formation is controlled by genes. The role of genes in pattern formation is well seen in the anterior-posterior patterning of embryos from the model organism Drosophila melanogaster (a fruit fly). Animal markings, segmentation of animals, phyllotaxis, neuronal activation patterns like tonotopy, and predator-prey equations' trajectories are all examples of how natural patterns are formed. In developmental biology, pattern formation describes the mechanism by which initially equivalent cells in a developing tissue in an embryo assume complex forms and functions.

One of the best understood examples of pattern formation is the patterning along the future head to tail (antero-posterior) axis of the fruit fly Drosophila melanogaster. The development of this fly is particularly well studied, and it is representative of a major class of animals, the insects. Other multicellular organisms sometimes use similar mechanisms for axis formation, although signal transfer between the earliest cells of many developing organisms is often more important than in Drosophila.

Vegetation patterns such as fir waves [22] and tiger bush [23] form for different reasons. Fir waves occur in forests on mountain slopes after wind disturbance, during regeneration. When trees fall, the trees that they had sheltered become exposed and are in turn more likely to be damaged, so gaps tend to expand downwind. Meanwhile, on the windward side, young trees grow, protected by the wind shadow of the remaining tall trees. In contrast, Tiger bush consists of stripes of bushes on arid slopes in countries such as Niger where plant growth is limited by rainfall. Each roughly horizontal stripe of vegetation absorbs rainwater from the bare zone immediately above it [23].

Turing [18] suggested that under certain conditions systems of chemical substances, although it may originally be quite homogeneous, can react and diffuse in such a way as to produce spatial patterns or structure due to an instability of the homogeneous equilibrium. In this section we will be devoted to mechanisms which can generate spatial pattern and form. We introduce and analyze reaction diffusion pattern formation mechanisms mainly with developmental biology in mind.

The Turing instability [18] refers to "diffusion-driven instability", i.e., the stationary solution stays stable with respect to a *kinetic system* (a system without diffusion) but becomes unstable with respect to the system with diffusion. This phenomenon is interesting because the general experience is that diffusion is a uniform phenomenon that is, helps stability by evening out differences, and now the opposite happens, and it is also of interest because Turing instability may go together with the occurrence of a spatially non-constant stationary solution, which is called a *pattern*. To be more precise, suppose that the studied reaction diffusion system has a bifurcation parameter denoted by b and let E^* be an equilibrium point for the associated kinetic system, that is, a homogeneous solution for the reaction diffusion system. Let b_0 be a fixed parameter and suppose that for $b < b_0$ the equilibrium solution E^* is asymptotically stable with respect of both the kinetic and the reaction diffusion system; however, for $b > b_0$ E^* remains asymptotically stable with respect the kinetic while it is unstable for the diffusion system; Finally, suppose that in a neighborhood of b_0 the associated nonlinear reaction diffusion system has a spatially non-constant stationary solution, a pattern. Then, we say that at b_0 the constant solution E^* undergoes a *Turing bifurcation*.

It is quite natural in many biological processes to consider Cross-diffusion. Note, for example, that in a predator-prey iteraction we might expect the predators to develop migratory strategies to take advantage of the preys defense switching behavior. Such migratory behavior, which depends on the concentration of the predators, constitutes a cross-diffusion which is in addition to each species a natural tendency to diffuse to areas of smaller population concentration. As the predators cross diffuse, and the prey switches its defense, we might expect such an ecosystem to exhibit a rich dynamical interplay among the three species.

In the following, we present some results classifying the existence of Turing bifurcation in two dimensional reaction diffusion mathematical models represented by PDEs. We follow basically [24].

Consider the reaction diffusion system satisfying the Newman boundary conditions as follows

$$U_t = DU_{xx} + F(U) \tag{34}$$

$$U_x(t,0) = U_x(t,l) = 0. \tag{35}$$

Clearly, a spatially constant solution $U(t) = (N(t), P(t))$ of (34) satisfies the boundary conditions (35) and the kinetic system

$$U_t = F(U). \tag{36}$$

The equilibrium $U^* = (N^*, P^*)$ of system (36) is at the same time a constant solution of (34).

Definition 14 The equilibrium $U^* = (N^*, P^*)$ of (34) is *Turing (diffusionally) unstable* if it is an asymptotically stable equilibrium of the kinetic system (36) but is unstable with respect to solutions of (34)–(35).

Remark 9 The latter requirement in Definition 14 means that there are solutions of (34)–(35) that have initial values $U(0,x)$ arbitrarily close to U^* (in the supremum norm) but do not tend to U^* as t tends to infinity.

In the following, we perform some calculations to find a criterion for the Turing instability. Without loss of generality, we can assume that $U^* = (0,0)$ and the linearized system of (34) at U^* assumes the form

$$V_t = DV_{xx} + JV \tag{37}$$

$$V_x(t,0) = V_x(t,l) = 0 \tag{38}$$

where

$$J = F_U(U^*) \tag{39}$$

We solve the linear boundary value problem by Fourier's method. Let $0 = \mu_0 < \mu_1 < \mu_2 < \cdots \to \infty$ and $\{\phi_j\}_{j=0}^{\infty}$ be the eigenvalues and eigenfunctions of the Laplacian operator in $(0,l)$ with Neumann boundary i.e.,

$$\phi_j''(x) = \mu_j \phi_j(x) \text{ on } (0,l), \quad \phi_j'(x) = 0 \quad \text{at } x = 0, l. \tag{40}$$

We can assume that $\{\psi_j\}_{j=0}^{\infty}$ is an orthonormal basis of $L^2(\Omega)$.

So, the solution of (37) with initial condition $V(\cdot,0) = V_0$ is given by

$$V(x,t) = \sum_{k=0}^{\infty} e^{(J-\mu_j D)t} \langle V_0, \phi_k \rangle \phi_k(x), \tag{41}$$

where $\langle V_0, \phi_j \rangle = \int_0^l V_0(x)\phi_j(x)\,dx$.

It follows from the linearization principle that a 'non-trivial' homogeneous solution of (34–35) is asymptotically stable if the eigenvalues of all matrix $J - \mu_j D$ have negative real part; if there exists a $k \geq 1$ such that $J - \mu_j D$ has an eigenvalue with positive real part then the solution is unstable.

Definition 15 We say that U^* undergoes a Turing bifurcation at $\lambda_0 \in (0, \infty)$ if for $0 < \lambda < \lambda_0$ the solution U^* is asymptotically stable, for $\lambda_0 < \lambda$ it is unstable (or vice versa), and in some neighbourhood of λ_0 the problem (34)–(35) has non-constant stationary solutions (i.e. solutions which do not depend on time t but are not constant in space, are varying with x).

In the following example, we present a two dimensional reaction-diffusion predator-prey model that exhibits a Turing bifurcation. For a detailed study of this model, we refer the readers to [25].

Example 2 We consider the following predator-prey system

$$\begin{aligned} \frac{\partial N}{\partial t} &= D_N \nabla^2 N + rN\left(1 - \frac{N}{K}\right) - \frac{\beta(1-m)NP}{K_1 + \delta(1-m)N + \alpha\eta A + \theta P}, \\ \frac{\partial P}{\partial t} &= D_P \nabla^2 P + \frac{c[\beta(1-m)N + \eta A]P}{K_1 + \delta(1-m)N + \alpha\eta A + \theta P} - \gamma P, \end{aligned} \tag{42}$$

where $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the *Laplacian Operator* in the two dimensional space $\Omega = [0, R] \times [0, R]$; D_N, D_P are the self-diffusion coefficients for prey and predator, respectively. If h_1 and e_1 represents the handling time of the predator per prey item and the ability of the predator to detect the prey, respectively, then $\beta = 1/h_1$ and $K_1 = 1/e_1 h_1$ represent the maximum rate of predation and half saturation value of prey uptake by the predator, respectively. Let h_2 and e_2, respectively, be the handling time of the predator per unit quantity of additional food and ability for the predator to detect the additional food. Thus, introduce the parameters $\eta = e_2/e_1$ and $\alpha = h_2/h_1$. The term ηA designates effectual additional food level and the constant θ scales the impact of predator interference. The description of the parameters composing the model (42) is given as follows: r is the maximum growth rate of prey; K is the carrying capacity of prey; δ is the proportionality constant; m represents the proportion of refuge protecting the prey; α is the relative handling time for additional food to prey item; η represents the relative ability of the predator to detect additional food to prey; A is the amount of additional food to the predators; c represents the maximum growth rate of the predator; γ is the death rate of predator.

Model (42) is to be analyzed under the following non-zero initial condition and *zero-flux* (or Neumann) boundary conditions:

$$N(x, y, 0) > 0, \;\; P(x, y, 0) > 0, \;\; (x, y) \in \; \Omega = [0, R] \times [0, R], \tag{43}$$

$$\frac{\partial N}{\partial \nu} = \frac{\partial P}{\partial \nu} = 0, \ (x, y) \in \partial\Omega. \tag{44}$$

where $\nu = \vec{\nu}(x)$ denotes the outer unit normal to $\partial\Omega$.

The constant solutions of system (42)–(44) are the same as the constant solution of the kinetic system. Following the ideas in [25], the constant solutions of (42)–(44) are: $E_0 = (0, 0)$; $E_1 = (K, 0)$; $E_2 = (0, P_2)$; where $P_2 = \dfrac{c\eta A - \gamma(K_1 + \alpha\eta A)}{\theta\gamma}$, which is feasible if $c\eta A - \gamma(K_1 + \alpha\eta A) > 0$; and $E^*(N^*, P^*)$, where $P^* = \dfrac{(1 - m)(c\beta - \gamma\delta)N^* + c\eta A - \gamma(K_1 + \alpha\eta A)}{\gamma\theta}$ and N^* is the positive root of the following quadratic equation: $a_0N^2 + a_1N + a_2 = 0$, where $a_0 = rc\theta\beta(1 - m)$, $a_1 = rc\theta\eta A + \beta K(1 - m)^2(c\beta - \gamma\delta) - rc\theta K\beta(1 - m)$ and $a_2 = \beta K(1 - m)[c\eta A - \gamma(K_1 + \alpha\eta A)] - rc\theta\eta AK$. The previous quadratic equation in N, has exactly one positive root if $a_2 < 0$.

We will focus our attention on the effect of diffusion on the system (42)–(44) around the equilibrium point E^*. To this purpose, we consider the variational system of (42)–(44) corresponding to E^* which is given by

$$\begin{aligned}
\frac{\partial n}{\partial t} &= a_{11}n + a_{12}p + D_N\left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2}\right),\\
\frac{\partial p}{\partial t} &= a_{21}n + a_{22}p + D_P\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right),
\end{aligned} \tag{45}$$

where $N = N^* + n$, $P = P^* + p$. Here, (n, p) are small perturbation of (N, P) around the interior equilibrium point $E^*(N^*, P^*)$. We observe that the Jacobian matrix corresponding to the interior equilibrium E^* of the kinetic system associated with (42) can be written as

$$J(E^*) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

where

$$\begin{aligned}
a_{11} &= -\left(\frac{r}{K} - \frac{\beta\delta(1 - m)^2P^*}{[K_1 + \delta(1 - m)N^* + \alpha\eta A + \theta P^*]^2}\right)N^*,\\
a_{12} &= -\frac{\beta(1 - m)N^*[K_1 + \delta(1 - m)N^* + \alpha\eta A]}{[K_1 + \delta(1 - m)N^* + \alpha\eta A + \theta P^*]^2},
\end{aligned}$$

$$a_{21} = \frac{c(1-m)P^*[\beta(K_1 + \alpha\eta A + \theta P^*) - \delta\eta A]}{[K_1 + \delta(1-m)N^* + \alpha\eta A + \theta P^*]^2},$$
$$a_{22} = -\frac{c\theta P^*[\beta(1-m)N^* + \eta A]}{[K_1 + \delta(1-m)N^* + \alpha\eta A + \theta P^*]^2}.$$

The corresponding characteristic equation of J is

$$\lambda^2 + A\lambda + B = 0, \tag{46}$$

where $A = -(a_{11} + a_{22})$ and $B = a_{11}a_{22} - a_{12}a_{21}$.

Let us consider the solution of system (45) in the form

$$\begin{pmatrix} n \\ p \end{pmatrix} = \begin{pmatrix} n_k \\ p_k \end{pmatrix} e^{\xi t + i(\kappa_x x + \kappa_y y)}$$

where ξ is the growth rate of perturbation in time t, κ_x and κ_y represent the wave numbers of the solutions. The Jacobian matrix of the linearized system (45) can be written as

$$\tilde{J} = \begin{pmatrix} a_{11} - D_N(\kappa_x^2 + \kappa_y^2) & a_{12} \\ a_{21} & a_{22} - D_P(\kappa_x^2 + \kappa_y^2) \end{pmatrix}.$$

In the spatial model, the value of ξ depends on the sum of the square of wave numbers $\kappa_x^2 + \kappa_y^2$ [26]. As a result, both wave numbers affect the eigenvalues. This makes it clear that some Fourier modes will vanish in the long-term limit whereas others will amplify. For the sake of simplicity, we can make use of ξ being rotational symmetric function on the $\kappa_x\kappa_y$-plane and substitute $\kappa^2 = \kappa_x^2 + \kappa_y^2$ and obtain the results for the two-dimensional case from the one-dimensional formulation. The corresponding characteristic equation is given by

$$\xi^2 + \tilde{A}\xi + \tilde{B} = 0, \tag{47}$$

where

$$\tilde{A} = A + \kappa^2(D_N + D_P),$$
$$\tilde{B} = B - (a_{11}D_P + a_{22}D_N)\kappa^2 + D_N D_P \kappa^4.$$

By the Routh-Hurwitz criterion together with Eq. (47), we conclude that the equilibrium point E^* is locally asymptotically stable in the presence of diffusion if and only if $\tilde{A} > 0$ and $\tilde{B} > 0$. Clearly, $A > 0$ implies $\tilde{A} > 0$. Therefore, Turing instability occurs only in the case when $B > 0$, but $\tilde{B} < 0$. Hence, the condition for Turing instability is given by

$$D_N D_P \kappa^4 - (a_{11}D_P + a_{22}D_N)\kappa^2 + B < 0. \tag{48}$$

The minimum of $D_N D_P \kappa^4 - (a_{11} D_P + a_{22} D_N)\kappa^2 + B = 0$ is reached at $\kappa^2 = \kappa_c^2$, where κ_c^2 is given by

$$\kappa_c^2 = \frac{a_{11} D_P + a_{22} D_N}{2 D_N D_P} > 0. \tag{49}$$

As $a_{11} + a_{22} < 0$ and κ_c is real then we must have $a_{11} a_{22} < 0$. Thus, a sufficient condition for instability is that inequality (48) be satisfied. Therefore, with the above value of κ_c^2, the condition for Turing instability given by inequality (48) can be written as

$$(a_{11} D_P + a_{22} D_N)^2 > 4 D_N D_P B. \tag{50}$$

The critical wave number κ_c of the growing perturbation is given by Eq. (49). It is not difficult to find that a change of sign in inequality (48) occurs when κ^2 enters or leaves the interval (κ_-^2, κ_+^2) where

$$\kappa_\pm^2 = \frac{a_{11} D_P + a_{22} D_N \pm \sqrt{(a_{11} D_P + a_{22} D_N)^2 - 4 D_N D_P B}}{2 D_N D_P}.$$

In particular, inequality (48) is satisfied (i.e., instability is present) for $\kappa_-^2 < \kappa^2 < \kappa_+^2$.

Note that the Turing instability cannot occur unless the diffusivity ratio is sufficiently away from unity. Indeed, recall that $a_{11} + a_{22} < 0$ and therefore $a_{11} < -a_{22}$ (where $a_{22} < 0$). Then, from the condition (49), we obtain:

$$\frac{D_P}{D_N} > -\frac{a_{22}}{a_{11}} > 1. \tag{51}$$

Condition (51) is a general necessary condition of Turing instability applicable to any two-species system. In particular, it means that the diffusive instability cannot occur for $D_P = D_N$.

Looking at the above analytic conditions, it is not clear how the local asymptotic stability and the Turing instability depend on the prey refuge and the additional food parameters. Therefore, in the following we present some numerical simulation to verify the occurrence of Turing instability. We consider system (42) with values of parameters: $r = 0.5$, $K = 3$, $\beta = 0.6$, $\delta = 1$, $m = 0.1$, $\alpha = 0.01$, $\eta = 0.01$, $A = 0.04\, c = 1$, $\gamma = 0.25$, $\theta = 0.4$, $K_1 = 0.4$, $D_N = 0.01$, $D_P = 1$. We investigate different Turing patterns of system (42). To explore the spatio-temporal dynamics of the system (42) in two dimensional spatial domain, the system of partial differential equations is numerically solved using a finite difference method. During numerical simulation different types of dynamics have been observed and it is found that the distributions of predator and prey are always of the same type. Thus, we have restricted our analysis of pattern formation to one distribution and only shown the distribution of prey for instance. We look at different situations by varying one of the parameters between m and A and keeping the other one fixed.

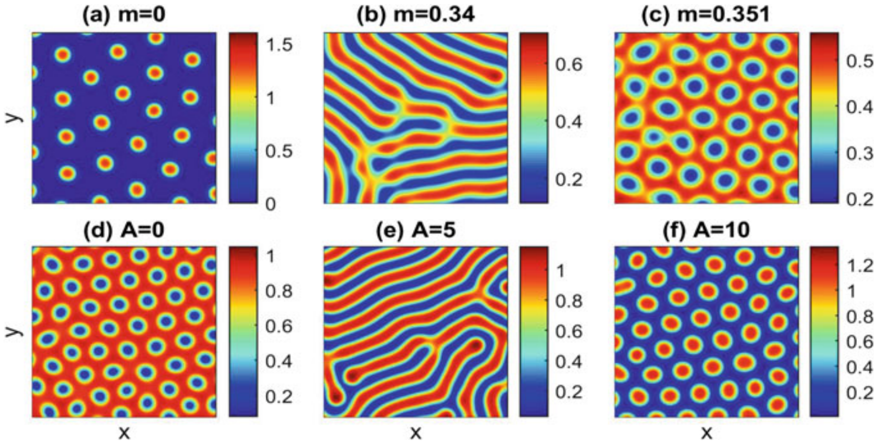


Fig. 2 Stationary Turing patterns developed by prey at **a**–**c** $A = 14$ and different values of m, **d**–**f** $m = 0.15$ and different values of A at time $t = 20{,}000$. Variations in m results in the pattern sequence spots to stripes to holes, whereas variations in A results in the opposite pattern sequence holes to stripes to spots

Different spatial patterns of the prey population emerge as we vary prey refuge (m) by keeping the value of additional food (A) fixed at $A = 14$ (Fig. 2a–c). The absence of prey refuge ($m = 0$) results in spot patterns, i.e., where the prey abundance is higher in isolated zones (Fig. 2a). In view of population dynamics, "spots" pattern means that the prey population is driven by the predator to a very low level in majority of the spatial region and the final result is the formation of patches of high prey density surrounded by areas of low prey densities. At $m = 0.34$, there is regular peaks and troughs of prey density, hence the stripe pattern emerges (Fig. 2b). However, for large prey refuge ($m = 0.351$), the model dynamics exhibits a transition from stripes-holes growth to holes replication, where the prey population is in the isolated zone with low density and the remaining region is of high density (Fig. 2c). Therefore, by increasing the value of m, the pattern sequence "spots → stripe-spot mixture → stripes → stripe-hole mixtures → holes" is observed.

Next, we look at different pattern formations by varying additional food A by keeping prey refuge as constant ($m = 0.15$) (Fig. 2d–f). In the absence of additional food supply ($A = 0$), we observe holes patterns (Fig. 2d). As we reach the value $A = 5$, the whole space is dominated by stripe patterns (Fig. 2e). For larger values of additional food ($A = 10$), spot pattern dominates the space (Fig. 2f). Therefore, in this case, an increase in additional food results in the pattern sequence "holes → stripe-hole mixture → stripes → stripe-spot mixtures → spots". Thus, it is inferred that by increasing the level of additional food to the predator, there is a transition from the prey predominant state to predator predominant state when prey refuge is present in the system.

4 Zip Bifurcation for ODE Systems Without Delay

In a recent study on a model describing the interactions of two predator species competing for one prey [27, 28], it is assumed that the threshold quantities of prey at which predator growth rate becomes positive are equal. As a consequence, it has been observed that the model system admits a one dimensional continuum of equilibria leading to, what is described as a *zip bifurcation* phenomenon. The zip bifurcation concept was introduced by Farkas [27] in connection with the study of the following ordinary differential system

$$\begin{cases} \dot{S} = \gamma S\left(1-\dfrac{S}{K}\right) - \dfrac{m_1 S}{a_1+S}u_1 - \dfrac{m_2 S}{a_2+S}u_2 \\ \dot{u_1} = \dfrac{m_1 S}{a_1+S}u_1 - d_1 u_1 \\ \dot{u_2} = \dfrac{m_2 S}{a_2+S}x_2 - d_2 u_2. \end{cases} \tag{52}$$

modeling the interaction between two predator species competing for a single prey with the Holling type-II functional response $\frac{mS}{a+S}$. In system (52), S denotes the density of the prey population and u_1, u_2 represent the densities of the predator populations. The constants γ and K are positive and respectively, denote the intrinsic growth rate and carrying capacity of the environment with respect to prey. The parameters m_i, d_i and a_i, for $i = 1, 2$, correspond to the maximal growth rates, mortality rates and saturation constants of the respective predator populations. Also the parameters α and β are positive and satisfy the inequality $0 < \alpha \leq \beta < 1$. The functional response of the predator i is given by $p(S, a_i) = \frac{m_i S}{a_i+S}$, $i = 1, 2$. One of the predators could be identified as a k-strategist and the other as an r-strategist. Roughly speaking, a species is said to be a k-strategist if it has a relatively low growth rate and may survive with low carrying capacity K, while the other predator, which exhibits high growth rate, is identified as a r-strategist (for details see [27–29]). It was establish that this model is not structurally stable, however, it illustrates the intuitively evident fact that at low values of the carrying capacity K, both predators might survive but as K grows, the k-strategist loses ground and only the r-strategist survives with the prey.

It is interesting to note that competition in predator-prey models differs from that of competition in microbial populations more popularly known as chemostat models. Evidently in the case of predator models, the predators have to put effort in securing prey, while in microbes the predator is not required to put any effort as they grow in cultured medium and the nutrients reach the microbes. For this reason, though competition occurs in microbes, the study will yield different dynamics. The basic response law in microbial populations is expressed in terms of Michaelis-Menten type formulation (For more details and relevant models we refer the readers to Nisbet and Gurney [30], Sree Hari Rao and Rajasekhara Rao [31–33], Smith and Waltman [34]). The mathematical representation of the Holling type-II type response function resembles the Holling type II response function used largely for predator-prey competition systems. This often leads to confusion among the readers who expect that the dynamics induced by the two respective competition systems be alike. We wish to state clearly that the present study restricts our considerations solely to predator-prey systems.

In the following we present a very brief summary of the results in [27] for system (52). The system (52) admits the following equilibria: $(S, x_1, x_2) = (0, 0, 0)$; $(S, x_1, x_2) = (K, 0, 0)$ and the points on the straight line

$$L_K = \Big\{ \big(S, u_1, u_2\big) \in R^3; \quad S = \lambda, \quad u_1 \geq 0, \; u_2 \geq 0 \; \text{ and } \\ m_1 \frac{S}{a_1 + S} u_1 + m_2 \frac{S}{a_2 + S} u_2 = \gamma(1 - \frac{\lambda}{K}) \Big\} \tag{53}$$

in the positive octant of the S, u_1, u_2-space, provided we assume that

$$m_i > d_i \; \text{ for } \; i = 1, 2, \; \text{ and } \; \lambda = \frac{a_1 d_1}{m_1 - d_1} = \frac{a_2 d_2}{m_2 - d_2}. \tag{54}$$

The equilibria on L_K will be denoted by $u^* = (\lambda, \xi_1, \xi_2)$. Hence $u^* \in L_K$. The end points of the line segment, i.e. the equilibria in the $u_1 = 0$ and $u_2 = 0$ planes, respectively are

$$P_2 = (\lambda, 0, \xi_2) = \Big(\lambda, 0, \frac{\gamma(a_2 + \lambda)(K - \lambda)}{m_2 K}\Big) \text{ and}$$
$$P_1 = (\lambda, 0, \xi_1) = \Big(\lambda, \frac{\gamma(a_1 + \lambda)(K - \lambda)}{m_1 K}, 0\Big).$$

It is easy to see that the trivial equilibria $(0, 0, 0)$ and $(K, 0, 0)$ are unstable provided $0 < \lambda < K$ (see [27, 29] for details). Also, all equilibria on L_K are stable for all K that satisfy the inequality $\lambda < K \leq a_2 + 2\lambda$. This means that if food is scarce both the r and k-strategists may live together in the long run in a steady state that depends on the initial values of the species. When $a_2 + 2\lambda < K < a_1 + 2\lambda$ (i.e, when $a_1 > a_2$) the family of equilibria on the line L_K undergoes a split and a part of L_K is unstable; that is, there exists a point $(\lambda, \xi_1(K), \xi_2(K))$ on L_K such that the equilibria on

$$L_U = \big\{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 < \xi_1(K)\big\}$$

are unstable for the flow of (52) and the equilibria on

$$L_S = \big\{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 > \xi_1(K)\big\}$$

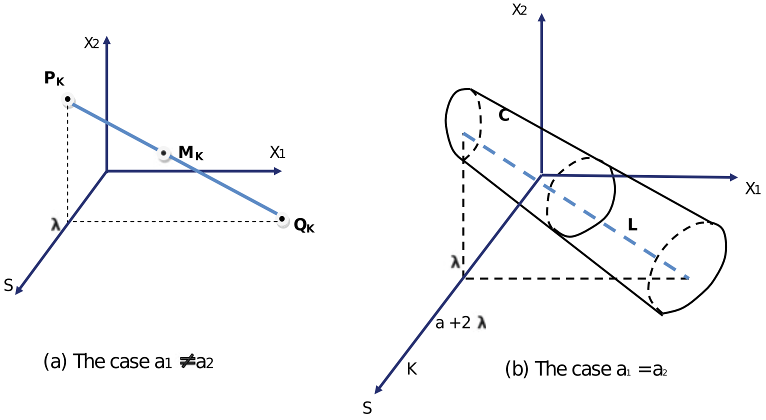
are stable. The point $(\xi_1(K), \xi_2(K))$ is obtained solving the system

$$\begin{cases} \dfrac{m_1\xi_1}{a_1+\lambda} + \dfrac{m_2\xi_2}{a_2+\lambda} & = \dfrac{\gamma(K-\lambda)}{K} \\ \dfrac{m_1\xi_1}{(a_1+\lambda)^2} + \dfrac{m_2\xi_2}{(a_2+\lambda)^2} & = \dfrac{\gamma}{K}. \end{cases} \tag{55}$$

As K takes on the value $a_1 + 2\lambda$ the equilibrium that exists in (S, x_1)-plane is stable while all other equilibria on L_K lose stability. When $K > a_1 + 2\lambda$ all equilibria on L_K become unstable.

Note that as K increases from $a_2 + 2\lambda$ to $a_1 + 2\lambda$ the point $(\lambda, \xi_1(K), \xi_2(K))$ moves along L_K continuously from $(\lambda, 0, \xi_2(K))$ to $(\lambda, \xi_1(K), 0)$ so that the points left behind become unstable. This phenomenon has been termed as *zip bifurcation* (see [27, 29]). From the point of view of the competition, as the quantity of available food increases the k-strategist loses ground and those equilibria where the relative growth of k-strategist is high compared to the growth of r-strategist, are the first to be destabilized. When K reaches the value $a_1 + 2\lambda$ all interior equilibria become destabilized and the only stable equilibrium remaining is the endpoint of L in the (S, x_1)-plane. This means that at this value of the carrying capacity the k-strategist dies out. One may prove that if K is increased further then even the equilibrium in the (S, x_1)-plane gets destabilized but the prey and the r-strategist continue to coexist in a periodic manner due to the occurrence of Andronov-Hopf bifurcation.

It is interesting to observe that the line L_K (blue line in Fig. 3) represents the line of nonisolated equilibria in the positive octant of R^3. This line connects points P_K in the (S, x_2)-plane and Q_K in the (S, x_1)-plane. Note that there exists a point M_K on L_K with the property that the equilibria on L_K between P_K and M_K are unstable, while those between M_K and Q_K are stable in the case where $a_1 > a_2$ or vice-versa



(a) The case a1 ≠a2

(b) The case a1 =a2

Fig. 3 **a** Zip bifurcation: the part of L_K between P_K and M_K is unstable and that between M_K and Q_K is stable if $a_1 > a_2$ or vice-versa (if $a_1 < a_2$); **b** degenerate zip bifurcation: in this case $a + 2\lambda < K < a + 2\lambda + \delta$ and all equilibria on L_K are unstable with an invariant topological cylinder around L_K which is a local attractor for system (52) in the case where $a_1 < a_2$ (see Fig. 3a). We present in the following the main result obtained in [27] related to the zip bifurcartion phenomenon.

Theorem 3 *For any k satisfying $a_2 + 2\lambda < K < a_1 + 2\lambda$ the point $(\lambda, \xi_1(K), \xi_2(K))$ divides L_K into two segments; the equilibria of system (52) in the set*

$$L_U = \{(\lambda, \xi_1(K), \xi_2(K)) \in L_K : \xi_1 < \xi_1(K)\}$$

are unstable, the equilibria in the set

$$L_S = \{(\lambda, \xi_1(K), \xi_2(K)) \in L_K : \xi_1 > \xi_1(K)\}$$

are stable in the Lyapunov sense (L_S stands for the "stable part of L_K").

Also, when $a_1 = a_2$ it is clear that all equilibria on L_K are stable for all K that satisfies the inequality $\lambda < K \leq a_2 + 2\lambda$. Furthermore, when K crosses $a + 2\lambda$ all equilibria on L_K lose stability and at $K = a + 2\lambda$ the segment L_K bifurcates into a cylinder, that is there exists a $\delta > 0$ such that for $a + 2\lambda < K < a + 2\lambda + \delta$ system (52) has an invariant topological cylinder C which is the union of closed paths and is an attractor of the system. Further it has a neighbourhood in which the trajectories with initial condition in this neighbourhood tend to C as t tends to infinity (See Fig. 3b). In other words, we observe that the zip bifurcation that shows up for all $a_1 \neq a_2$, vanishes when $a_1 = a_2$. This is an interesting scenario in the dynamics of the system (52) and henceforth, we term this phenomenon as a *degenerate zip bifurcation*.

We simulate the system (52) with the following values of the parameters $m_1 = 0.6$, $m_2 = 0.7$, $d_1 = 0.3$, $d_2 = 0.2$, $\beta_1 = 0.3$, $\beta_2 = 0.5$, $\gamma = 0.8$, $a_1 = 0.16$, $a_2 = 0.4$, $\lambda = 0.16$ and $K = 0.6$ (see Fig. 4). Clearly Fig. 4 explains the occurrence of zip bifurcation of the equilibria on the line L_K with unstable part initiating near the (S, x_1)-plane and also on the line L_K (the brown line) and directed towards the (S, x_2)-plane with the stable part lying near this plane. We note that the transition from instability to stability occurs at $(\lambda, \xi_1(K), \xi_2(K)) = (0.16, 0.1137777778, 0.2986666667)$ on L_K.

In the study of mathematical models of real world problems usually conditions are placed on the parameters of the model equations to yield isolated equilibrium points. Local stability analysis of an isolated equilibrium would explain the circumstances under which such an equilibrium would lose or gain stability when the parameters pass through certain critical values. Also it may be established under certain additional assumptions that the change in the stability behavior may lead to the presence of small amplitude periodic solutions when the parameters pass through these critical values. In such a situation we say that the equilibrium bifurcates into small amplitude periodic solutions. Now in the models presented in this survey, following the ideas developed in [27, 28], we observe that a continuum of equilibrium points or a line of

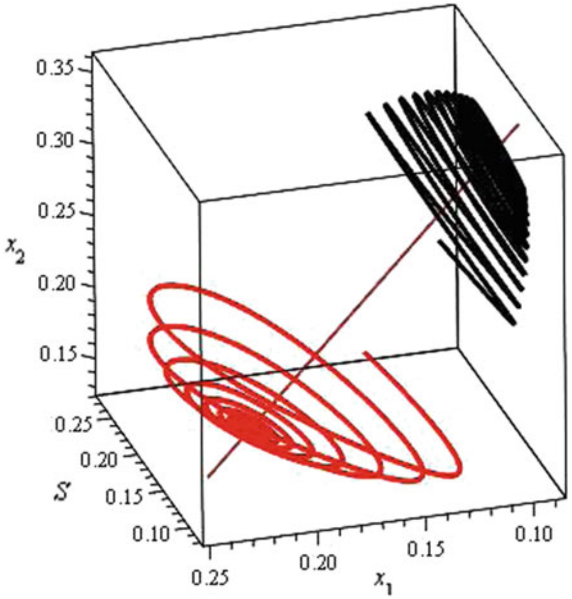


Fig. 4 Phase portrait: The red curve is an unstable orbit initiating near the (S, x_1)-plane and the line L_K, directed towards the (S, x_2)-plane with the stable part lying in it; the black curve is a stable orbit initiating near the (S, x_2)-plane and the line L_K, directed towards the (S, x_1)-plane with the unstable part lying in it

non isolated equilibrium points exists when the carrying capacity K is varied. The stability analysis of this continuum in many situations is very complicated.

There are many studies wherein the researchers have studied predator prey systems of the type described by system (52), by placing appropriate conditions that yield isolated equilibria. The feature of our work is to analyze systems that posses non isolated (a continuum of equilibria) for predator prey models described by system (52). Thus, the models presented in this survey is different from the work that appears in the literature. We have not provided the complete bibliography of the results related to the study of isolated equilibria. We refer the readers to the work in the following references in which the non isolated equilibria situation has been studied [35–41].

5 Zip Bifurcation for ODE Systems with Delay

Time delays are natural in any biological process therefore, in this section a delay term modeling the delayed logistic growth of the prey is included in the prey equation of system (52). Fixed points of the system are identified, and the main result concerning the linearized stability analysis of the nonisolated equilibria is presented. Accordingly, in this survey paper we consider the following system of ordinary differential equations involving a discrete time delay

$$\begin{cases} S'(t) = \gamma\big(1 - \frac{S(t-\tau)}{K}\big)S(t) - m_1\dfrac{S(t)}{a_1 + S(t)}u_1(t) - m_2\dfrac{S(t)}{a_2 + S(t)}u_2(t) \\[2ex] u_1'(t) = \dfrac{m_1 S(t)}{a_1 + S(t)}u_1(t) - d_1 u_1(t) \\[2ex] u_2'(t) = \dfrac{m_2 S(t)}{a_2 + S(t)}u_2(t) - d_2 u_2(t). \end{cases} \tag{56}$$

The meaning of the parameters composing model (56) are the same as in model (52).

The equilibria of (56) satisfy the equations

$$\begin{aligned} &\gamma\big(1 - \frac{S(t-\tau)}{K}\big)S(t) - m_1\frac{S(t)}{a_1 + S(t)}u_1(t) - m_2\frac{S(t)}{a_2 + S(t)}u_2(t) = 0 \\ &\frac{m_1 S(t)}{a_1 + S(t)}u_1(t) - d_1 u_1(t) \qquad\qquad\qquad\qquad\qquad\qquad\quad = 0 \\ &\frac{m_2 S(t)}{a_2 + S(t)}u_2(t) - d_2 u_2(t) \qquad\qquad\qquad\qquad\qquad\qquad\quad = 0. \end{aligned}$$

Following the arguments given in [27, 28, 37, 38] we conclude that the equilibria of (56) are

$$(S, u_1, u_2) = (0, 0, 0), \quad (S, u_1, u_2) = (K, 0, 0)$$

and the points on the straight line segment

$$L_K = \Big\{ \big(S, u_1, u_2\big) \in R^3; \quad S = \lambda, \quad u_1 \geq 0, \ u_2 \geq 0 \text{ and } \\ m_1\frac{S}{a_1 + S}u_1 + m_2\frac{S}{a_2 + S}u_2 = \gamma\big(1 - \frac{\lambda}{K}\big) \Big\}$$

in the positive octant of (S, u_1, u_2)-space, provided we assume that

$$m_i > d_i \text{ for } i = 1, 2, \text{ and } \lambda = \frac{a_1 d_1}{m_1 - d_1} = \frac{a_2 d_2}{m_2 - d_2}. \tag{57}$$

The characteristic equation associated with the variational system of system (52) at $(\lambda, \xi_1, \xi_2) \in L_K$ is given by

$$\mu\Big[\mu^2 - a\mu + c + be^{-\mu\tau}\mu\Big] = 0 \tag{58}$$

where $a = \lambda \sum_{i=1}^{2} \frac{m_i}{(a_i+\lambda)^2}\xi_i$; $b = \frac{\lambda\gamma}{K}$ and $c = \lambda \sum_{i=1}^{2} \frac{m_i \beta_i}{(a_i+\lambda)^2}\xi_i$.

We now discuss the stability of the equilibria (λ, ξ_1, ξ_2) on L_K for a fixed K in the interval $(a_2 + 2\lambda, a_2 + 2\lambda)$. In this case, the family of equilibria on L_K undergoes a split and there exists a point $(\lambda, \xi_1(K), \xi_2(K))$ on L_K such that the equilibria on L_U are unstable for the flow of (52) and the equilibria on L_S are stable in view of the case when $\tau = 0$. For details we refer the readers to [39].

Theorem 4 *The number of different imaginary roots with positive (negative) imaginary parts of (58) can be zero, one, or two only.*

(I) If there are no such roots, then the stability of each equilibrium $(\lambda, \xi_1, \xi_2) \in L_K$ does not switch for any $\tau \geq 0$.

(II) If there is one imaginary root of (58) with positive imaginary part, an unstable solution (λ, ξ_1, ξ_2) on L_K never becomes stable for any $\tau \geq 0$. If the solution is asymptotically stable for $\tau = 0$, then there exists a $\tau_0 > 0$ such that for all $\tau < \tau_0$ this solution is uniformly asymptotically stable, and is unstable for all $\tau > \tau_0$.

(III) If there are two imaginary roots with positive imaginary part, ω_+, ω_-, such that $0 < \omega_- < \omega_+$, then the stability of each equilibrium (λ, ξ_1, ξ_2) on L_K can change at most a finite number of times, as τ increases, and eventually becomes unstable.

Corollary 1 *Suppose that $E = (\lambda, \xi_1, \xi_2) \in L_K$ is an equilibrium point of system (56) or equivalently of system (52). Then the following statement holds.*

(i) If E is stable (or unstable) for system (56) when $\tau = 0$, then stability switches may not occur for all $\tau \geq 0$.

(ii) If E is stable for system (56) when $\tau = 0$, we may have a $\tau_0 > 0$ such that E loses its stability when τ passes through τ_0.

(iii) If E is stable (or unstable) for system (56) when $\tau = 0$, we may have a finite number of stability switches for E as τ increases.

Corollary 2 *For any fixed K satisfying $a_2 + 2\lambda \leq K \leq a_1 + 2\lambda$, there exists an equilibrium point $(\lambda, \xi_1(K), \xi_2(K))$ which splits L_K in two parts L_K^u and L_K^s (one of which may be empty); for $\tau = 0$, the equilibria of (56) on the set*

$$L_K^u = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 < \xi_1(K)\}$$

are unstable and the equilibria on the set

$$L_K^s = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 > \xi_1(K)\}$$

are stable. Furthermore, depending of the values of the parameters of the system (56) the zip bifurcation phenomenon may or may not be preserved. In other words the zip bifurcation is unsustainable.

Remark 10 The above result is surprising since it shows that the zip bifurcation phenomenon cannot be sustained in the presence of a discrete delay in system (52). In the section we consider a delay-free system and conclude taht the sustainability of the zip bifurcation is achieved even in the presence of diffusion.

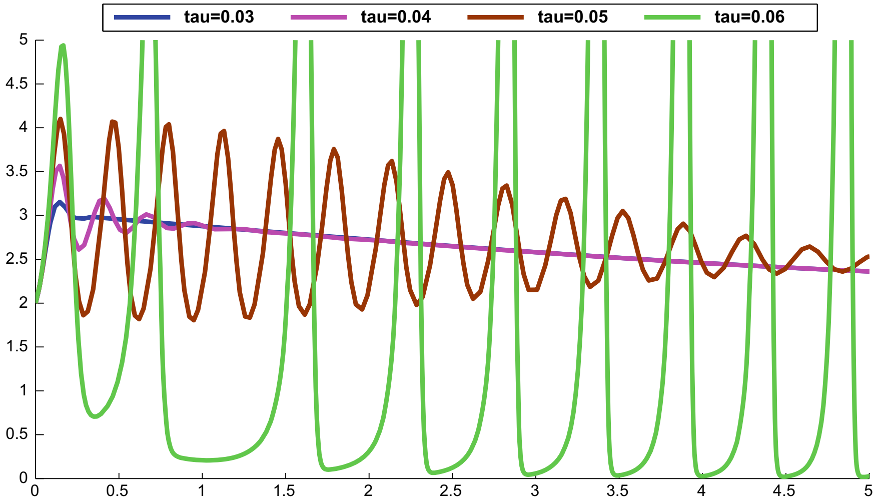


Fig. 5 Unsustainable zip bifurcation

In the Fig. 5, we present the simulation illustrating the dynamics of the time delay system (56). We choose the following values for the parameters in system (56) given by $m_1 = 0.6$, $m_2 = 0.3$, $d_1 = 0.3$, $d_2 = 0.2$, $\beta_1 = 0.3$, $\beta_2 = 0.1$, $\gamma = 46$, $a_2 = 1$, $a_1 = 2$, $\lambda = 2$ and $K = 5.5$. We find that the equilibria (λ, ξ_1, ξ_2) on the line L_K is stable for all values of τ satisfying $0 \leq \tau \leq \tau_1 = 0.04163939054$ and is unstable for $\tau > \tau_1$ (see Fig. 5). This confirms the unsustainability of the zip bifurcation when $\tau > \tau_1$. It is clear from Fig. 5 that $\tau = 0.03$ (blue curve) represents the stable behaviour of S and for $\tau = 0.04$ (pink curve) S loses stability since this value of τ is nearer to τ_1. Also, for $\tau = 0.05$ and 0.06, S exhibits more instability tendency (brickred and green curves respectively).

6 Zip Bifurcation for Reaction Diffusion Systems Without Delay

The model to be discussed is described by the partial differential equation system

$$\begin{cases} \dfrac{\partial S}{\partial t}(x,t) = \delta_0 \Delta S + \gamma(1 - \dfrac{S}{K})S - \displaystyle\sum_{i=1}^{2} m_i \dfrac{S}{a_i + S} u_i \quad \text{in} \quad \Omega \times (0, \infty) \\ \dfrac{\partial u_1}{\partial t}(x,t) = \delta_1 \Delta u_1 + (m_1 \dfrac{S}{a_1 + S} - d_1) u_1, \\ \dfrac{\partial u_2}{\partial t}(x,t) = \delta_2 \Delta u_2 + (m_2 \dfrac{S}{a_2 + S} - d_2) u_2 \end{cases} \tag{59}$$

where $\Omega \subset \mathbb{R}^N$ $(N = 1, 2, 3)$ is an open connected bounded domain with smooth boundary $\partial\Omega$. We will assume that the functions S and u_i satisfies the Neumann boundary conditions

$$\frac{\partial S}{\partial \nu}(x,t) = 0, \ \ \frac{\partial u_i}{\partial \nu}(x,t) = 0, \ \ i = 1, 2, \ \ \text{on } \partial\Omega \times (0, \infty), \tag{60}$$

where $\nu = \vec{\nu}(x)$ denotes the outer unit normal to $\partial\Omega$ and $\Delta = \sum_{j=1}^{N} \frac{\partial}{\partial x_j}$ is the Laplacian operator. The meaning of the parameters composing model (59) are the same as in model (52). Positivity and global existence of the solutions for system (59)–(60), have been studied in [38].

The homogeneous solutions of system (59)–(60) are the same as the constant solutions of system (52). To study the stability of these equilibria, let $E = (S^*, u_1^*, u_2^*)$ be a homogeneous solution of system (59)–(60). Thus the variational system of (59)–(60) corresponding to E is given by

$$\begin{cases} \dfrac{\partial S(x,t)}{\partial t} = \delta_0 \Delta S + \Big[\gamma\big(1 - \dfrac{S^*}{K}\big) - \displaystyle\sum_{i=1}^{n} m_i \frac{a_i}{(a_i + S^*)^2} u_i^*\Big] S(t,x) \\ \qquad\qquad - \displaystyle\sum_{i=1}^{2} \frac{m_i S^*}{a_i + S^*} u_i(t,x) - \frac{\gamma\lambda}{K} S(t-\tau, x) \\ \dfrac{\partial u_i(x,t)}{\partial t} = \delta_i \Delta u_i + \dfrac{\beta_i u_i^*}{a_i + S^*} u_i(t,x), \ \ i = 1, 2, \ \ \text{on } \Omega \times (0, \infty). \end{cases} \tag{61}$$

Let $0 = \mu_0 < \mu_1 < \mu_2 < \ldots \to \infty$ and $\{\psi_k\}_{k=0}^{\infty}$ be the eigenvalues and eigenfunctions of the Laplacian operator in Ω with Neumann boundary conditions on $\partial\Omega$:

$$\begin{cases} \Delta\psi_k = \lambda_k \psi_k \ \text{ in } \Omega \\ \dfrac{\partial \psi_k}{\partial \nu} = 0 \qquad \text{on } \partial\Omega. \end{cases} \tag{62}$$

Without loss of generality, we suppose that $\{\psi_k\}_{k=0}^{\infty}$ is an orthonormal basis for $L^2(\Omega)$.

Let $w = (S, u_1, u_2)$, $D = \text{diag}(\delta_0, \delta_1, \delta_2)$ and J be the matrix

$$J = \begin{pmatrix} \lambda \displaystyle\sum_{i=1}^{2} \frac{m_i}{(a_i + S^*)^2} u_i^* - \frac{\gamma S^*}{K} e^{-\tau\nu} & -\dfrac{m_1 S^*}{a_1 + S^*} & -\dfrac{m_2 S^*}{a_2 + S^*} \\ \dfrac{\beta_1 u_1^*}{a_1 + S^*} & 0 & 0 \\ \dfrac{\beta_2 u_2^*}{a_2 + S^*} & 0 & 0 \end{pmatrix}.$$

Then, (61) can be written as

$$\frac{\partial w}{\partial t} = D\Delta w + Jw,$$

hence, the solution of (60), (61) with initial condition $w(\cdot, 0) = w_0$ is given by

$$w(x,t) = \sum_{k=0}^{\infty} e^{(J-\mu_k D)t} \langle w_0, \psi_k \rangle \psi_k(x), \tag{63}$$

where $\langle w_0, \psi_k \rangle = \int_\Omega w_0(x)\psi_k(x)\, dx$.

It follows from the linearization principle that a 'non-trivial' homogeneous solution of (59)–(60) is asymptotically stable if all the eigenvalues of the matrix $J - \mu_k D$ have negative real parts and if there exists a $k \geq 0$ such that $J - \mu_k D$ has an eigenvalue with positive real part then the solution is unstable. In the following, we present the mains results obtained in [39] concerning the stability of the equilibria on L_K of the system (59)–(60).

Theorem 5 *Suppose that E is an equilibrium of (59)–(60) independent of x. If E is stable for the flow of (52), then E is stable for the flow of (59)–(60) independently of $D = \mathrm{diag}(\delta_0, \delta_1, \delta_2)$.*

Theorem 6 *For any K satisfying $a_2 + 2\lambda \leq K \leq a_1 + 2\lambda$, the point $(\lambda, \xi_1(K), \xi_2(K))$ splits L_K into two parts L_K^u and L_K^s; the equilibria of (59)–(60) in the set*

$$L_K^u = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 < \xi_1(K)\}$$

are unstable and those in the set

$$L_K^s = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 > \xi_1(K)\}$$

are stable, independently of the diffusion matrix $D = \mathrm{diag}(\delta_0, \delta_1, \delta_2)$.

So, the result given in Theorem 5 shows that the zip bifurcation phenomenon may be preserved by the introduction of a diagonal diffusion matrix D in the model (59), independent of the domain Ω.

7 Zip Bifurcation for Reaction Diffusion Systems with Delay

Together with the diffusion of the species, the reaction-diffusion model studied in this section is represented by the following PDE system

$$\begin{cases} \dfrac{\partial S}{\partial t}(x,t) = \delta_0 \Delta S(t,x) + \gamma\big(1 - \dfrac{S(t-\tau,x)}{K}\big)S(t,x) - \displaystyle\sum_{i=1}^{2} m_i \dfrac{S}{a_i + S} u_i(t,x) \\ \dfrac{\partial u_1}{\partial t}(x,t) = \delta_1 \Delta u_1(t,x) + \big(m_1 \dfrac{S}{a_1+S} u_1(t,x) - d_1\big) u_1(t,x) \\ \dfrac{\partial u_2}{\partial t}(x,t) = \delta_2 \Delta u_2(t,x) + \big(m_2 \dfrac{S}{a_2+S} u_2(t,x) - d_2\big) u_2(t,x), \quad \text{on} \quad \Omega \times (0,\infty) \end{cases} \tag{64}$$

where we assume that the functions S and u_i satisfy the Neumann boundary conditions

$$\frac{\partial S}{\partial \nu} = 0, \quad \frac{\partial u_i}{\partial \nu} = 0, \quad i = 1, 2, \quad \text{on} \quad \partial\Omega \times (0,\infty), \tag{65}$$

$\nu = \vec{\nu}(x)$ denotes the outer unit normal to $\partial\Omega$ and $\Delta = \sum_{j=1}^{N} \partial/\partial x_j$ is the Laplacian operator.

We are interested on the stability of the equilibria of the system (64)–(65) on L_K. If $E \in L_K$, following the results presented in [39] we have the following theorem.

Theorem 7 *(i) Assume that conditions* $(\overline{H}1)$ *given in [39] pp. 1231 hold. If the equilibrium E is stable (unstable) at* $\tau = 0$ *then E remains stable (unstable) for all* $\tau \geq 0$.

(ii) Assume that conditions $(\overline{H}2)$ *given in [39] pp. 1231 hold. Then E is unstable for all* $\tau \geq 0$.

(iii) Assume that either $(\overline{H}3)$ *or* $(\overline{H}4)$ *given in [39] pp. 1232 holds. As* τ *increases, stability switches may occur.*

Remark 11 Note that if part (i) of Theorem 7 holds, then the zip bifurcation phenomenon sustains. On the other hand, if (iii) holds the phenomenon may not sustain.

8 Zip Bifurcation for Reaction Diffusion Systems with Cross-Diffusion

We are analyzing an ecosystem consisting of two predators competing for one single prey. In such a system, we might expect the prey to develop two separate sets of defensive capabilities, one effective against each of the predators, and would switch from one set to the other depending on the relative abundance of the two predator species. Such defense switching behavior has been described, for example, for a fish species in the Lake Tanganyika against two phenotypes of the scale-eating cichlid P. microlepis [42]. On the other hand, we might also expect the predators to develop migratory strategies to take advantage of the defense switching behavior of the prey. In the present work this is indeed the case. The purpose of this paper is to investigate such a migration which involves cross-diffusion process. In particular, we will demonstrate the emergence of *Turing instability*. For a detailed discussion and examples on cross-diffusion models describing real world phenomena, we refer the readers to [18, 20, 21, 43–45].

Accordingly, noting that the relationship between the two predators in (52) is competitive, we shall introduce the cross-diffusion terms as follows:

$$\begin{cases} -div\ (\delta_0 \nabla S) & = \gamma(1-\frac{S}{K})S - m_1\frac{S}{a_1+S}u_1 - m_2\frac{S}{a_2+S}u_2 \\ -div\left[(\delta_1 + \frac{k_1}{a_1+u_2})\nabla u_1 - \frac{k_1 u_1}{(a_1+u_2)^2}\nabla u_2\right] & = (m_1\frac{S}{a_1+S} - d_1)u_1 \\ -div\left[(\delta_2 + \frac{k_2}{a_2+u_1})\nabla u_2 - \frac{k_2 u_2}{(a_2+u_1)^2}\nabla u_1\right] & = (m_2\frac{S}{a_2+S} - d_2)u_2 \quad \text{on} \quad \Omega \times [0,\infty) \end{cases} \tag{66}$$

with Neumann boundary conditions

$$\frac{\partial S}{\partial \nu}(x,t) = 0, \ \frac{\partial u_i}{\partial \nu}(x,t) = 0, \ \ i = 1, 2, \ \ \text{on} \ \ \partial\Omega \times [0,\infty). \tag{67}$$

As a starting point in this exposition, we consider the one dimensional form of the system (66)–(67) by letting $\Omega = [0, l]$, l a fixed constant.

Remark 12 In the system (66), k_1 and k_2 denote the *cross-diffusion coefficients*. The predators u_1, and u_2 diffuse with flux

$$J_1 = -\left(\delta_1 + \frac{k_1}{a_1+u_2}\right)\nabla u_1 + \frac{k_1 u_1}{(a_1+u_2)^2}\nabla u_2.$$

We observe that, as $\frac{k_1 u_1}{(a_1+u_2)^2} \geq 0$, the part $\frac{k_1 u_1}{(a_1+u_2)^2}\nabla u_2$ of the flux of u_1 is directed towards the increasing population density of the predator u_2. In this way, the first predator moves in anticipation of the predator u_2 and of the defense switching behavior of the prey. Similarly, we can observe that the flux of u_2 is directed towards the increasing population density of the predator u_1 so that the second predator also moves in anticipation of predator u_1 and of the defense switching behavior of the prey, which shows the competition between the two predators.

Keeping $\Omega = [0, l]$ and l fixed, we also consider a study of the one dimensional form of the system (52) given by

$$\begin{cases} S_t = \Delta\big(k_{11}S + k_{12}u_1 + k_{13}u_2\big) + \gamma(1-\frac{S}{K})S - m_1\frac{S}{a_1+S}u_1 - m_2\frac{S}{a_2+S}u_2 \\ u_{1t} = \Delta\big(k_{21}S + k_{22}u_1 + k_{23}u_2\big) + (m_1\frac{S}{a_1+S} - d_1)u_1 \\ u_{2t} = \Delta\big(k_{31}S + k_{32}u_1 + k_{33}u_2\big) + (m_2\frac{S}{a_2+S} - d_2)u_2, \\ S_x = u_{1x} = u_{2x} = 0, \ \ \text{on} \ \ [0,l] \times [0,\infty), \end{cases} \tag{68}$$

where $k_{ij},\ i, j = 1, 2, 3$ are constants.

We notice that the homogeneous solutions for systems (66) and (68) are $(S, u_1, u_2) = (0, 0, 0)$, $(S, u_1, u_2) = (K, 0, 0)$ and those described in equation of L_K given in (53).

8.1 Analysis of Equilibria for Model (66) on the Line L_K

In the following, we consider the cross-diffusion model (66) and present the results concerning the stability properties of the equilibria $u^* = (\lambda, \xi_1, \xi_2) \in L_K$. The results presented here were submitted for publication and shall appear in the future.

Proposition 2 *If $K > a_1 + 2\lambda$, then all equilibria on L_K are unstable.*

The following result presents the conditions under which the PDE system (66) and its corresponding ODE system (52) share the same stability behavior.

Theorem 8 *Suppose that u^* is an equilibrium of (66) independent of x. If u^* is stable for the flow of (52), then u^* will be stable for the flow of (66).*

Theorem 8 may be viewed as a result on the preservation of stability under self-diffusion for prey population and cross-diffusion for both predator populations for model (66). As a consequence, if $\lambda < K < a_2 + 2\lambda$, the line of equilibrium points L_K of (52) is locally asymptotically stable for the flow of (66). If $K > a_1 + 2\lambda$, then L_K is unstable. We have the following corollary.

Corollary 3 *For any K satisfying $a_2 + 2\lambda \leq K \leq a_1 + 2\lambda$, the point $(\lambda, \xi_1(K), \xi_2(K))$ splits L_K in two parts L_K^u and L_K^s; the equilibria of (66) on the set*

$$L_K^u = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 < \xi_1(K)\}$$

are unstable and those on the set

$$L_K^s = \{(\lambda, \xi_1, \xi_2) \in L_K : \xi_1 > \xi_1(K)\}$$

are stable, regardless the cross-diffusion coefficients given in the system (66).

Remark 13 The result in Corollary 3 shows that the zip bifurcation phenomenon is sustainable even by the introduction of cross-diffusion in the model (52).

8.2 Analysis of Equilibria for Model (68) on the Line L_K

The Turing instability [18] refers to “diffusion-driven instability, that is the stability of the constant equilibrium $u^* = (\lambda, \xi_1, \xi_2)$ changing from stability for the ODE system

(52), to instability for the PDE system (68). In what follows, we discuss the Turing instability for the model (68) which depends on the cross-diffusion coefficients.

Consider the linearized system of (68) at u^*

$$\begin{cases} \psi_t = K(u^*)\psi_{xx} + G_u(u^*)\psi, \quad 0 < x < l, \quad t \geq 0 \\ \psi_x = 0, \quad x = 0, l, \quad t \geq 0, \end{cases} \tag{69}$$

where $\psi = (\psi_1, \psi_2, \psi_3)^T$,

$$K(u^*) = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix}$$

and

$$F_u(u^*, 0) = G_u(u^*) = \begin{pmatrix} -\frac{\gamma\lambda}{K} + \lambda \sum_{i=1}^{2} \frac{m_i}{(a_i+\lambda)^2}\xi_i & -\frac{m_1\lambda}{a_1+\lambda} & -\frac{m_2\lambda}{a_2+\lambda} \\ \frac{\beta_1\xi_1}{a_1+\lambda} & 0 & 0 \\ \frac{\beta_2\xi_2}{a_2+\lambda} & 0 & 0 \end{pmatrix},$$

with $\beta_i = m_i - d_i, i = 1, 2$ and k_{ij} are constants with $k_{11}, k_{22}, k_{33}, k_{12}, k_{13} \geq 0$ and $k_{21}, k_{31}, k_{23}, k_{32} \leq 0$.

We denote by $P_k(\nu)$ the characteristic polynomial of $G_u(u^*) - \mu_j K(u^*)$. For each $k \geq 0$, the eigenvalues of $G_u - \mu_j K$ are the roots of the polynomial

$$P_j(\nu) = \nu^3 + A_j\nu^2 + B_j\nu + C_j, \tag{70}$$

We will omit the complete expressions for A_j, B_j and C_j since they are too long.

It is easy to see that the inequalities

$$A_j > 0, B_j > 0, C_j > 0 \text{ and } A_jB_j - C_j > 0 \tag{71}$$

ensure the stability of u^* while the equilibrium u^* is unstable if either of (72) or (73) holds for $j = 1, 2, \ldots$

$$A_j > 0, C_j > 0 \text{ and } B_j < 0 \tag{72}$$

$$A_j > 0, B_j < 0 \text{ and } C_j < 0. \tag{73}$$

The next theorem shows that if u^* is stable for the flow of (52) then u^* may not be stable for the flow of (68).

Theorem 9 *Suppose that u^* is an equilibrium of (68) independent of x. Then, if u^* is stable for the flow of (52), u^* will be stable for the flow of (68) provided (71) holds for all $j = 1, 2, \ldots$. Further, the equilibrium u^* is unstable, if any of the mentioned inequalities (72) or (73) holds for some $j = 1, 2, \ldots$*

9 Discussion

The dynamical systems modeling real world phenomena discussed in this survey, exhibit three types of bifurcations: Hopf bifurcation, Turing bifurcation and Zip bifurcation. Special attention is bestowed on model (52) and its generalizations. Under certain natural assumptions, it has been observed that model (52) and its generalizations admit a one dimensional continuum of equilibria leading to Zip bifurcation. The effort in this exposition is to create an awareness among the researchers that certain real world dynamical systems often give rise to the existence of nonisolated equilibria also known as a continuum of equilibria. In this article we have presented methodologies that would help one to analyze such systems.

It is our belief that in the years to come the bifurcation theory plays a more important role in almost all application domains of science and technology. Our thoughts and efforts along these lines have culminated in the organization of the current survey of thoroughly reviewed original articles that are of interest to the scientific community working in the area of bifurcation theory and applications. The models presented in this survey research exhibit rich dynamics for varying values of the vital parameters involved.

Acknowledgements The authors wish to thank the anonymous reviewers for their constructive suggestions, which have led to the improvement of our earlier presentation.

References

1. Hsu, S.H., Hubbell, S.P., Wallman, P.: Competing predators. SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **35**, 617–625 (1978)
2. Hsu, S.H., Hubbell, S.P., Wallman, P.: A contribution lo the theory of competing predators. Ecol. Monogr. **48**, 337–349 (1978)
3. Koch, A.L.: Coexistence resulting from an alteration of density dependent and density independent growth. J. Theor. Biol. **44**, 373–386 (1974)
4. Koch, A.L.: Competitive coexistence of two predators utilizing the same prey under constant environmental conditions. J. Theor. Biol. **44**, 387–395 (1974)
5. Butler, G.J.: Competitive predator-prey systems and coexistence. Population Biology Proceedings, Edmonton 1982. Lecture Notes in Biomathematics, vol. 52, pp. 210–217. Springer, Berlin (1983)
6. Keener, J.P.: Oscillatory coexistence in the chemostat: a codimension two unfolding. SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **43**, 1005–1018 (1983)
7. Smith, H.L.: The interaction of steady stale and Hopf bifurcations in a twopredator- one-prey competition model. SIAM (Soc. Ind. Appl. Math.) J. Appl Math. **42**, 27–43 (1982)

8. Wilken, D.R.: Some remarks on a competing predators problem. SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **42**, 895–902 (1982)
9. Andronov, A.A., Leontovich, E.A., Gordon, I.I., Maier, A.G.: Qualitative Theory of Dynamic Systems of Second Order. Nauka, Moskva (1966)
10. Andronov, A.A., Leontovich, E.A., Gordon, I.I., Maier, A.G.: Theory of Bifurcations of Dynamic Systems on a Plane. Nauka, Moskva (1971)
11. Chicone, C.: Ordinary Differential Equations with Applications, 2nd edn. Springer (2006)
12. Hassard, B.D., Kazarinoff, N.D., Wan, Y.H.: Theory and Application of Hopf Bifurcation. Cambridge University Press, Cambridge (1981)
13. Sotomayor, J., Mello, L.F., de Caravallo Braga, D.: Lyaponuv Coefficients for Degenerate Hopf Bifurcations. arXiv:0709.3949v1 [math.DS], 25 Sep 2007
14. Kuznetsov, Y.A.: Elements of Applied Bifurcation Theory. Springer-Verlag (1998)
15. Gasull, A., Torregrosa, J.: A new approach to the computation of the Lyapunov constants. Comp. Appl. Math. **20**, 149–177 (2001)
16. Marsden, J.E., McCracken, M.: The Hopf bifurcation and its applications. Springer-Verlag, New York (1976)
17. Ferreira, J.D., Salazar, C.A.T., Tabares, P.C.C.: Weak Allee effect in a predator-prey model involving memory with a hump. Nonlinear Anal. R. World Appl. **14**, 536–548 (2013)
18. Turing, A.: The chemical basis of morphogenesis. Philos. Trans. Soc. London Ser. B **237**, 37–72 (1952)
19. Tilman, D., Kareiva, P. (eds.): Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions. Monographs in Population Biology, vol. 30. Princeton University Press, NJ, USA (1997)
20. Murray, J.D.: Mathematical Biology I: An Introduction, 3rd edn. Springer, USA (2002)
21. Murray, J.D.: Mathematical Biology II: Spatial Models and Biomedical Applications, 3rd edn. Springer, USA (2003)
22. D'Avanzo, C.: Fir Waves: Regeneration in New England Conifer Forests. TIEE, (2004). Retrieved 26 May 2012
23. Tongway, D.J., Valentin, C., Seghieri, J.: Banded Vegetation Patterning in Arid and Semiarid Environments. Springer-Verlag, New York (2001)
24. Cavani, M., Farkas, M.: Bifurcations in a predator-prey model with memory and diffusion II: turing bifurcation. Acta Math. Hungar **63**(4), 375–393 (1994)
25. Chakraborty, S., Tiwari, P.K., Sasmal, S.K., Biswas, S., Bhattacharya, S., Chattopadhyay, J.: Interactive effects of prey refuge and additional food for predator in a diffusive predator-prey system. Appl. Math. Modell. **47**, 128–140 (2017)
26. Baurmann, M.: Turing instabilities and pattern formation in a benthic nutrient-microorganism system. Math. Biosci. Eng. **1**, 111–130 (2004)
27. Farkas, M.: Zip bifurcation in a competition model. Nonlinear Anal. Theory Methods Appl. **8**, 1295–1309 (1984)
28. Farkas, M.: Competitive exclusion by zip bifurcation in dynamical systems. In: IIASA Workshop 1985 Sopron. Lecture Notes in Economics and Mathematical Systems, vol. 287, pp. 165–178. Springer (1987)
29. Farkas, M.: Periodic Motions. Springer-Verlag, New York (1994)
30. Nisbet, R.M., Gurney, W.S.C.: Modelling Fluctuating Populations. Wiley, New York, NY (1982)
31. Sree Hari Rao, V., Raja Sekhara Rao, P.: Dynamic Models and Control of Biological Systems. Springer, New York (2009)
32. Sree Hari Rao, V., Raja Sekhara Rao, P.: Global stability of Chemostat models involving time delays. Differ. Equ. Dyn. Syst. **8**, 1–28 (2000)
33. Sree Hari Rao, V., Raja Sekhara Rao, P.: Basic chemostat model revisited. Differ. Equ. Dyn. Syst. **17**(1–2), 3–16 (2009)
34. Smith, H.L., Waltman, P.: The Theory of Chemostat. Cambridge University Press, London (1995)

35. Bocsó, A., Farkas, M.: Political and economic rationality leads to velcro bifurcation. Appl. Math. Comput. **140**, 381–389 (2003)
36. Sáez, E., Stange, E., Szántó, I.: Simultaneous zip bifurcation and limit cycles in three dimensional competition models. SIAM J. Appl. Dyn. Syst. **05**, 1–11 (2006)
37. Ferreira, J.D., Fernandes de Oliveira, L.A.: Hopf and zip bifurcation in a specific (n + 1)-dimensional competitive system. Matemáticas: Enseñanza Universitaria **15**, 33–50 (2007)
38. Ferreira, J.D., Fernandes de Oliveira, L.A.: Zip bifurcation in a competitive system with diffusion. Differ. Equ. Dyn. Syst. Int. J. Theory Appl. Comput. Simul. **17**, 37–53 (2009)
39. Ferreira, J.D., Sree Hari Rao, V.: Unsustainable zip-bifurcation in a predator-prey model involving discrete delay. Proc. R. Soc. Edinb. **143A**, 1209–1236 (2013)
40. Farkas, M., Ferreira, J.D.: Zip bifurcation in a reaction-diffusion system. Differ. Equ. Dyn. Syst. Int. J. Theory Appl. Comput. Simul. **15**, 169–183 (2007)
41. Farkas, M., Sáez, E., Szántó, I.: Velcro bifurcation in competition models with generalized Holling functional response. Miskolc Math. Notes **6**(2), 185–195 (2005)
42. Takahashi, S., Hori, M.: Unstable evolutionarily stable strategy and oscillation: a model of lateral asymmetry in scale-eating cichlids. Amer. Nat. **144**, 1001–1020 (1994)
43. Okubo, A., Levin, S.A.: Diusion and Ecological Problems: Modern Perspectives, 2nd edn. Springer-Verlag, Berlin (2000)
44. Wang, M.: Stationary patterns of strongly coupled prey-predator models. J. Math. Anal. Appl. **292**, 484–505 (2004)
45. Lou, Y., Ni, W.-M.: Diffusion, self-diffusion and cross-diffusion. J. Differ. Equ. **131**, 79–131 (1996)

A Modified Coordinate Search Method Based on Axes Rotation

Suvra Kanti Chakraborty and Geetanjali Panda

Abstract In this paper, a traditional coordinate search method is modified through rotation of axes and an expansion of square-stencil to capture the solution in a better and faster way. The scheme remains derivative free with global convergence property. The iterative process is explained for two-dimensional function in detail, which is followed by its extension to higher dimensions. Numerical illustrations and graphical representations for the sequential progress of the proposed scheme are provided. The comparison with the traditional coordinate search schemes through performance profiles are also provided to coin the advantages of the proposed scheme.

Keywords Derivative free optimization · Coordinate search · Forward-track line search · Stencil expansion

1 Introduction

In the context of solving optimization problems from real life, derivative of functions often turns out to be unavailable, unreliable or impossible to obtain. Such situation invokes several methods of derivative-free optimization, which involves computation of the functional values at some sample points. A study on these schemes can be found in [4, 10, 11, 14], among which, GPS (generalized pattern search) is an efficient class of methods that moves to an improving point following a specific pattern. One of its ancestors, the basic coordinate search method has two variants: non-opportunistic and opportunistic search. Non-opportunistic search, alternatively called complete search, finds the sample points in all possible directions of the coordinates, whereas opportunistic search tries to find an improving point by fixing an ordering in the set of coordinate directions.

S. K. Chakraborty (✉) · G. Panda
Department of Mathematics, Indian Institute of Technology, Kharagpur 721302,
West Bengal, India
e-mail: suvrakanti@maths.iitkgp.ernet.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_36

Several natural improvements to the coordinate search methods exist in the literature of derivative-free optimization. Adaptive search method [8] adapts the transformation of axes, Hookes-Jeeves method [7] creates a set of linearly independent directions and proposes an exploratory step. In the primitive stage, the convergence of coordinate search method was not established. Later on, exploiting the idea and structure of positive bases, the convergence of coordinate search method was proposed (see [12]). Under some assumptions, the theory of convergence to the stationary points was also proposed except for some exceptional cases (see [15]). Due to the simplicity of inherent framework, coordinate search method never failed to draw the attention of the researchers. The ease of its implementation has been exploited, and several hybrid structures are proposed by the researchers during the last few decades (see [6, 9]).

In this paper, the existing coordinate search method is modified. Initially, the points are selected as coordinate directions, each at equal distance. The convex combination of these points completes an n-dimensional structure (viz, CXn). For example, it provides a square in two-dimension and an octahedron in three-dimension, respectively. The initial set of coordinate directions eventually forms a maximal positive basis. This paper accepts some logic of existing GPS methods but differs in several contexts. Like the other GPS methods, the progress of its iterative points depends on the success of obtaining an improving point at some vertex of the above mentioned n-dimensional structure (CXn). In addition to the success step, we propose a line search that tracks forward in the improving direction and an expansion of CXn, which puts the current iteration point in its centre. The failure step contracts CXn, which halves all its diagonals. However, this hybridization of line-search technique along with its direct search structure, does not employ any change in the simplicity of its framework. Rather, this adds a new efficient path to the existing coordinate search methods in a significant way. The most interesting fact in the proposed scheme is that the geometrical concept of expansion of CXn automatically generates another new set of positive basis, which is used in the next search-step. The number of maximum possible bases, however, remains to be finite. This fact is used to prove the global convergence of the proposed scheme.

The proposed scheme, along with its necessary algorithms, are described in Sect. 2, which is followed by its execution in two-dimensions. Section 3 proposes the higher dimensional extension of the proposed scheme, whereas Sect. 4 provides the global convergence of the scheme. In Sect. 5, the numerical experiences on test functions for the proposed and existing schemes are summarized. Finally some concluding remarks are provided in Sect. 6.

2 Proposed Scheme in Two Dimensions

We first refer to the algorithm written in the book of [11]. It is known that the classical coordinate search method is designed for unconstrained optimization problem. It starts with computing the functional value at an initial point and 2n points in the stencil, which is the set $S(x, h) = \{z|z = x + he_i\}$ centred at x, where e_i is the unit vector in i-th coordinate direction. This set of points forms a maximal positive basis at the initial point. In non-opportunistic search for minimization problem, the entire stencil is sampled and the centre x is replaced with $x_{\min}$ if $f(x_{\min}) < f(x)$, where $x_{\min} = arg\, \min_{z \in S(x,h)} f(z)$. If $f(x_{\min}) \geq f(x)$, the stencil is shrunk by preferably a factor of 2. On the contrary to non-opportunistic search, opportunistic search maintains an order of the unit vectors and shifts the centre once it gets a better point. We refer to Algorithm 4.1 and 4.2 of the book [11] for the detailed algorithms of non-opportunistic and opportunistic coordinate search method. Both of these methods can be divided into two parts: poll step and search step. In the poll step, the functional value at the points on the stencils are computed and compared. In the search step, the shifting of the centre of the stencil or the contraction occurs.

In this paper, a forward track line search in the improving direction is introduced along with a possible rotation and expansion of the stencil square under specified condition. The scheme in 2-Dimension for $\min_{x \in \mathbb{R}^2} f(x)$ is explained first, which will be extended to n-dimension in the next sections. First, in the poll step, $2n$ stencil points are found keeping the initial point at its centre and the functional values at those points are computed. Next, the search step is divided into two parts. If the stencil-failure occurs, that is, if none of the stencil points provides improving functional value, we contract the stencil, otherwise stencil-success occurs, that is, if any of the points provides improving functional value, we choose that direction and perform a forward tracking line search until we fail to capture a better point. The centre of the stencil is shifted to this point, and the squared stencil is expanded with a factor $\sqrt{2}$ in the next step by its construction. In this process, the stencil-square gets rotated by 45° angle. The iterative scheme is stopped once the length of the square becomes less than a preassigned small positive real number.

The following color picture explains the idea of forward-track along improving direction, contraction, and expansion of square-stencil. The yellow vertices form the initial square-stencil with the red point as its centre. The smaller square with blue vertices forms the new square-stencil in case of stencil-failure. On the other hand, in case of stencil-success, an arrow is assigned to the yellow vertex to demonstrate the forward-tracking at the vertex of success. With the assumption that no other green vertex comes out to be an improving point, the current green point becomes the centre of new square-stencil, and the corresponding expanded square is formed by adding two sky-blue points to the exiting structure (Fig. 1).

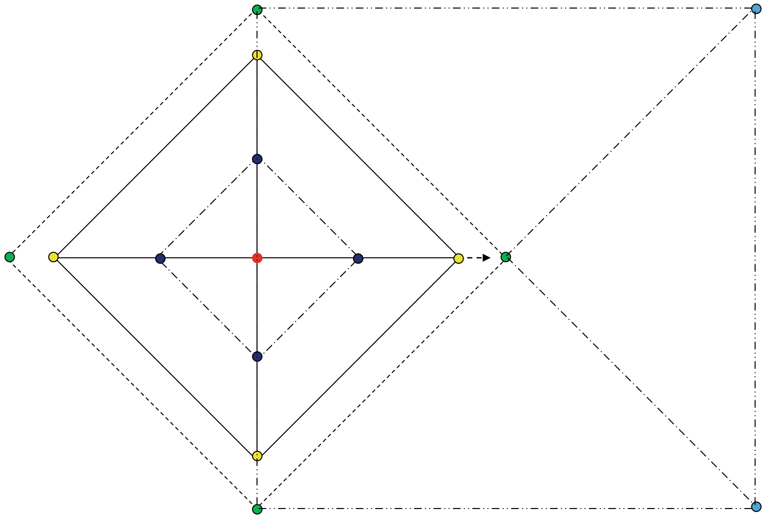


Fig. 1 Forward-tracking along improving direction, contraction and expansion of square-stencil

The scheme for 2-Dimension is described in the following algorithms. Algorithm 1 will be used to perform the main algorithm.

Algorithm 1: Forming a square-stencil with centre x and two vertices x_1 and x_2

```
1. Input $x, x_1, x_2$;
2. Initialize $Y$, a matrix of order 5-by-2;
3. $Y(1, :) = x$;
4. $Y(2, :) = x_1$;
5. $Y(3, :) = x_2$;
6. $Y(4, :) = 2x - x_1$;
7. $Y(5, :) = 2x - x_2$;
8. Output $Y$
```

2.1 Numerical Illustration in Two Dimension

In this subsection, the proposed scheme is justified with a well-known test problem, minimizing Rosenbrock's function with the standard initial guess at $(-1.2, 1)$. The code for the Algorithm 2 is written and executed in Matlab-R2015a. It is observed that the stencil square is sometimes contracted, expanded, or rotated as well. Figure 2 gives a clear view to this idea. Finally, the length of the square gets reduced, and when it shrinks to less than 0.00001, the execution of the code gets stopped. The scheme accepts the solution point as (0.99998, 0.99999), which is within the ϵ tolerance limit to the ideal solution of the function at (1, 1).

Algorithm 2: Scheme for two dimension

```
      1. Input $x^{(0)}$, $\epsilon > 0$;
      2. $\beta = .1, h = 1$;
      3. Initialize matrix $Y$ of order 5-by-2;
      4. $x_1 = x + he_1,\ x_2 = x + he_2, e_1, e_2$ are standard bases in 2D;
for $k = 0, 1, 2, \ldots$
      5. k=k+1;
      6. Use Algorithm 1 with input $x^{(k-1)}, x_1, x_2$ to obtain Y;
      7. Find functional values at the 2 dimensional points $Y(i, :),\ i = 1(1)5$;
      8. If stencil failure occurs, $h = \frac{h}{2}, x_1 = (x + Y(2, :))/2$,
$x_2 = (x + Y(3, :))/2$; else go to step 10;
      9. if $h > \epsilon$ : go to Step 5; else break ;
      10. $x_{min}$ is one of the vertices of the stencil square, $d = x_{min} - x^{(k)}, x_{old} = x^{(k)}$.
      11. Along $d$, perform forward-track line search with factor $\beta$.
          If $f(x_{min} + \beta d) > f(x_{min})$, go to Step 15.
      12. Continue Step 11 until the forward-track does not produce an improving point.
          The current iterating point is $x^{(new)}$.
      13. Treat $x_{old}$ as the centre of a square-stencil and $x^{(new)}$ as a vertex;
          Find the two adjacent vertices $V_1$ and $V_2$ of $x^{(new)}$;
      14. $x^{(k)} = x^{(new)}, x_1 = V_1,\ x_2 = V_2$,
          Go to step 5;
      15. Square Expansion: $x^{(k)} = x_{min}, x_1 = Y(2, :), x_2 = Y(4, :), h = \sqrt{2}h$, and go to
step 5.
      16. Output: $x^{(k)}$;
end;
```

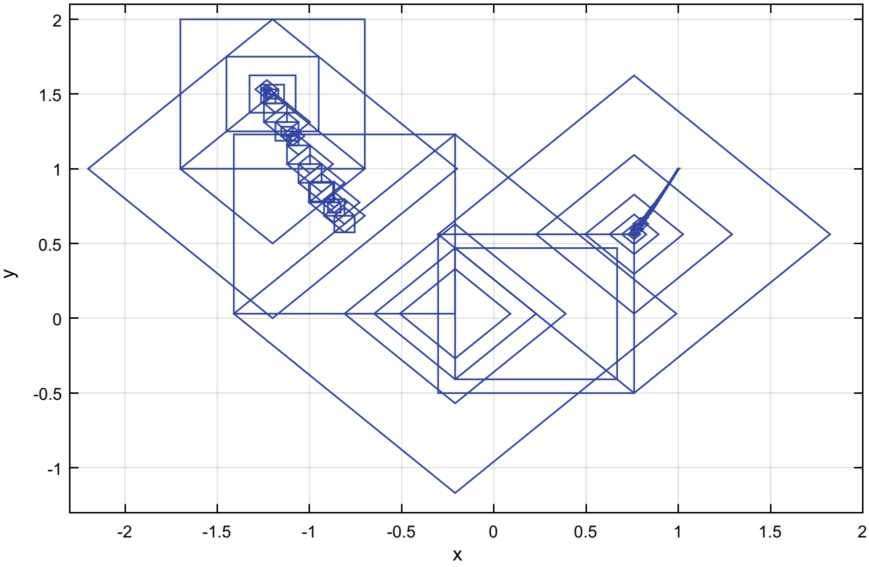


Fig. 2 The sequential progress of the proposed scheme of Rosenbrock function in two dimension

3 Extension to n-Dimension

In the previous section, the proposed scheme has been illustrated along with an algorithm and a numerical example in two-dimensions. In this section, the scheme is extended to higher dimensional problems.

Consider an initial point, say $x^{(0)}$ in $\mathbb{R}^n$ and n points along n basis vectors of the n-dimensional plane. Take the reflection of these chosen n points with respect to the initial point. Now, these $2n$ points form a maximal positive basis of n-dimensional plane at $x^{(0)}$. Figure 3 shows $2n$ number of basis vectors at the initial red point. On joining these points, the required n-dimensional stencil is obtained. At any of these points, $n-1$ number of square can be drawn. In Fig. 3, since the maximal positive basis is drawn for three dimensions, at any of its verticds, two squares pass by.

An ordering of $2n$ basis elements is maintained as $[e_1, e_2, e_3, \ldots, e_n, e_{n+1}(=-e_1), e_{n+2}(=-e_2), e_{n+2}(=-e_3), \ldots, e_{2n}(=-e_n)]$. Suppose in the first search step, we successfully found an improving point and the point is $x^{(0)} + e_m h$, $1 \leq m \leq 2n$. Then, we take the $(m-1)$-th or $(m+1)$-th basis element and consider the corresponding square formed by the m-th basis element and any of these two. Once we get the square, we apply Algorithm 2, for expansion on contraction on this particular square. If expanded, the distance of any vertex of the expanded square from its centre is computed, and remaining basis vectors are expanded to the same length. It helps the n-dimensional structure to keep its geometry as intact. If the search step leads to failure, the basis elements are halved as to contract the structure as a whole. This idea is briefed in Algorithm 3.

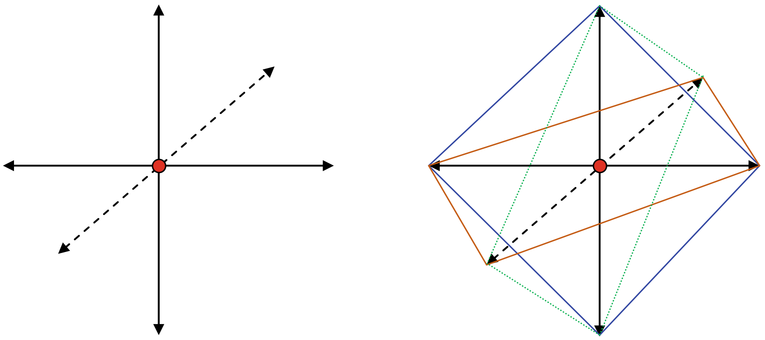


Fig. 3 Maximal positive basis and squares at the vertices

Algorithm 3: Scheme for n-dimension

```
        1. Input $x^{(0)}$, $\epsilon > 0$, dimension: $n$;
        2. Known data : $\beta = .1$, $h = 1$;
        3. Find $n$ dimensional $2n$ stencil points;
  for $k = 0, 1, 2, \ldots$
        4. k=k+1;
        5. Perform search-step to check if any vertex of the $n$-dimensional stencil
           comes out to be an improving point than $x^{(0)}$.
        6. If stencil failure occurs, $h = \frac{h}{2}$ ;
        7. if $h > \epsilon$ : go to Step 4; else break ;
        8. If stencil success occurs at some vertex, identify any of its adjacent
  vertex,
            and complete a square;
        9. Use Algorithm 2 on this square (2-D structure),
             and impose forward-track or expansion of square as applicable;
        10. Compute the length of the diagonal of the expanded square and treat its
  half as h;
        11. At the centre of the square, extend the other basis vectors with the
  length h;
        12. The current iterating point is $x^{(k)}$.
        13. Go to Step 4;
        14. Output: $x^{(k)}$;
  end;
```

Advantages of the proposed scheme: The primary advantages of the scheme are as follows.

1. The scheme retains the simplicity of the basic coordinate search scheme.
2. The construction in search-step invokes the rotation of the stencil-square, which helps to justify the nature of the function concerning another set of maximal positive bases.
3. This scheme provides a scope of progress further to get an improving point in the success-direction by a forward tracking line search.
4. The expansion of the stencil-square by its construction can capture the domain in a shorter time than the other methods. Even if the initial point is chosen far from the solution, this property helps to reach its vicinity very fast. This provides the global convergence property of the proposed scheme.

4 Convergence Analysis of the Proposed Scheme

The scheme starts with an initial point and a set of maximal positive bases. The process of finding a new iterate $x^{(k)}$ passes through two successive stages. One is search step that finds the functional values at some finite number of points. Next, depending on success or failure of the search step, the forward track and stencil expansion is introduced or it is determined whether the stencil structure would get contracted. So, the search step is optional for the proof of convergence. Rather, one must concentrate on failure, that introduces the poll step, which polls the points along with the positive basis with reduced size from current iterate.

Unless the current iterate point is a stationary one, the poll step gives a successful step that finds an improving point after a finite number of reduction in step size parameter. With this notion, we take some mild assumptions and state the convergence proof of the scheme following the results from (Chap. 7, [4]).

Since the iterates $f(x^{(k)})$ corresponding to $x^{(k)}$ is monotonically decreasing, a convenient way of imposing this assumption is as following

Assumption 1 The level set $L(x^{(0)}) = \{x \in \mathbb{R}^n : f(x) \leq f(x^{(0)})\}$ is compact.

Since we are interested in global convergence of the scheme, it relies on proving first that there exists a subsequence of step size parameters converging to 0. For this, we must guarantee the following assumption. Here h_k is the step length h at k-th iteration.

Assumption 2 If there exists $h_0 > 0$ such that $h_k > h_0$ for each iteration, then the algorithm visits only a finite number of points.

Based on Assumption 2, the following theorem and corollary can be proved [4].

Theorem 1 *If Assumption 2 holds, then the sequence of step size parameters satisfies* $lim_{k\to\infty} inf\ h_k = 0$.

Corollary 1 *Let the above assumptions hold. There exists a point x^* and a subsequence $\{k_i\}$ of unsuccessful iterates for which $lim_{i\to\infty} h_{k_i} = 0$ and $lim_{i\to\infty} x_{k_i} = x^*$.*

If the function is differentiable, this scheme stops at a stationary point. To prove this, we need to assume the following. This assumption is admissible since for n-dimensional problem, the proposed algorithm uses atmost n number of positive bases.

Assumption 3 The set of positive bases used by the algorithm is finite.

Another assumption of differentiability becomes necessary to take to show that the solution point is a stationary one, which is as follows.

Assumption 4 f is continuously differentiable in an open set containing $L(x^{(0)})$.

Theorem 2 *Let Assumptions 1–4 hold. Then the sequence of iterates $\{x^{(k)}\}$ has a limit point x^* for which $\nabla f(x^*) = 0$.*

The proof of the above theorem follows from (Chap. 7, [4]).

Regarding the convergence of the scheme, we must conclude that the assumptions stated in this section are admissible to our proposed scheme. Hence, without taking differentiability condition by successive stencil failure, a stencil gets contracted to a point. Moreover, under the continuously differentiability condition and finite positive basis set, it may be proved that a stationary point is reached, which eventually satisfies the first order necessary condition for optimality of unconstrained minimization problem.

5 Numerical Illustration

For numerical illustration, we run the proposed algorithm on a set of functions from [1] in MATLAB-R2015a platform with 8 GB RAM. This test set includes convex, non-convex as well as non-differentiable functions. The three schemes are abbreviated as follows.

CSO: Coordinate Search (Opportunistic);
CSNO: Coordinate Search (Non-opportunistic);
PM: Proposed Method.

The schemes mentioned above are all derivative free search schemes and ensures global convergence. So, the initial guesses for their respective algorithms are chosen usually very far from the solution point, except for some functions that are steep and valley shaped at the solution point. First, the number of iteration for each function from the test set is observed. In the context of computational experience, the number of times they are called for each scheme is counted. The algorithm for each scheme gets stopped once the n-dimensional structure becomes small as desired. The centre of this final structure is considered to be the solution of the minimization problem. However, one may also include the budgets for the stopping criteria. To be more specific, the barrier of function count and run-time may be included. As per the computational experience, it is seen that run-time is not a suitable measure since it is not platform independent. This fact motivates us to restrict on counting the number of function evaluation, which is platform independent and put a comparison of the proposed scheme with the traditional methods on this aspect.

With a fine introspection into the algorithm, it is observed that the proposed scheme identifies two basis elements at the current iteration point and with the points at distance 'h' on these two orthogonal basis element, completes a square. On that plane, it performs the algorithm that is mentioned for two-dimension. The algorithm for two-dimension clearly defines whether the square gets expanded or contracted. Next, according to that, the other basis elements at the centre of n-dimensional structure are expanded or contracted to keep the geometrical shape of the n-dimensional stencil as intact.

For these three solvers- CSO, CSNO, and PM, one can observe the computational results in Table 1. The comparative study of the results of different solvers through the study of performance profiles are an area of interest for current researchers [2, 3]. We also present and discuss the numerical findings in the same approach. The initial guesses for each test are taken as two n-dimensional vectors f_n and g_n, which are defined as mentioned below.

$$f_n = [1, -1, 1, \dots, -1], g_n = [-1.2, 1, -1.2, 1, \dots, -1.2, 1].$$

Besides putting the computational detail in Table 1, we further study the performance profiles on the test set for the three schemes. The performance ratio defined in [5, 13] is $\rho_{(p,s)} = \frac{r_{(p,s)}}{\min\{r_{(p,s)}:1 \leq r \leq n_s\}}$, where $r_{(p,s)}$ refers to the iteration (or, num-

Table 1 Numerical comparisons among the proposed method and traditional methods

S. no	Function	Dim	Initial guess	Iteration			Function count		
				OCS	NOCS	PM	OCS	NOCS	PM
1	Almost perturbed quadratic	10	$100\ f_{10}$	1017	1017	221	10840	20341	4226
2	Bartei	2	$100\ f_2$	217	217	60	568	869	256
3	BIGGSB1	10	$100\ f_{10}$	2879	1990	739	35290	39801	13930
4	Broyden Tridiagonal	10	$100\ f_{10}$	1542	1536	885	16651	30721	16716
5	CURLY 20	10	$100\ f_{10}$	1082	1082	232	11500	21641	4436
6	DIXON3DQ	10	$100\ f_{10}$	1017	1990	739	10840	39801	13930
7	Extended BD1	10	f_{10}	132	132	169	1530	2641	3231
8	Extended Beale	10	f_{10}	922	922	986	5075	18441	18511
9	Extended DENSCHNB	10	$50\ f_{10}$	512	502	178	5520	10041	3420
10	Extended Penalty	10	$100\ f_{10}$	1071	1071	2111	11335	214021	4046
11	Extended PSC1	10	$100\ f_{10}$	1097	1087	221	11625	21741	4227
12	Extended Rosenbrock	10	g_{10}	27247	27367	8878	171060	547347	165011
13	Extended TET	10	$100\ f_{10}$	1777	1067	274	24160	21341	5397
14	Extended Tridiagonal1	10	$100\ f_{10}$	1962	1022	1629	21165	20441	30708
15	Extended Trigonometric	10	f_{10}	109	109	153	1270	2181	2901
16	Full Hessian FH1	10	$100\ f_{10}$	3233	1640	1648	37876	32801	32801
17	HIMMELBG	10	$100\ f_{10}$	1087	1087	204	21741	11570	3922
18	LIARWHD	10	$10\ f_{10}$	443	456	262	3091	9121	4972
19	Price 1	2	$100\ f_2$	207	207	67	543	829	265
20	Quartic	10	$20\ f_{10}$	217	217	173	2395	4341	3314

ber of function evaluation) for solver s spent on problem p and n_s refers to the number of problems in the model test set. In order to obtain an overall assessment of a solver on the given model test set, the cumulative distribution function is $P_s(\tau) = \frac{1}{n_p} size\{p \in \mathscr{T} : \rho_{(p,s)} \leq \tau\}$. In other words, $P_s(\tau)$ is the probability that a performance ratio of $\rho_{(p,s)}$ is within a factor of τ of the best possible ratio. We present the performance profiles on the number of iteration, and the number of objective function evaluation on the test set $\mathscr{T}$ for the three schemes CSO, CSNO and PM in Figs. 4 and 5 respectively. The figures on performance profiles express the advantage of developing proposed scheme PM over the other traditional two methods.

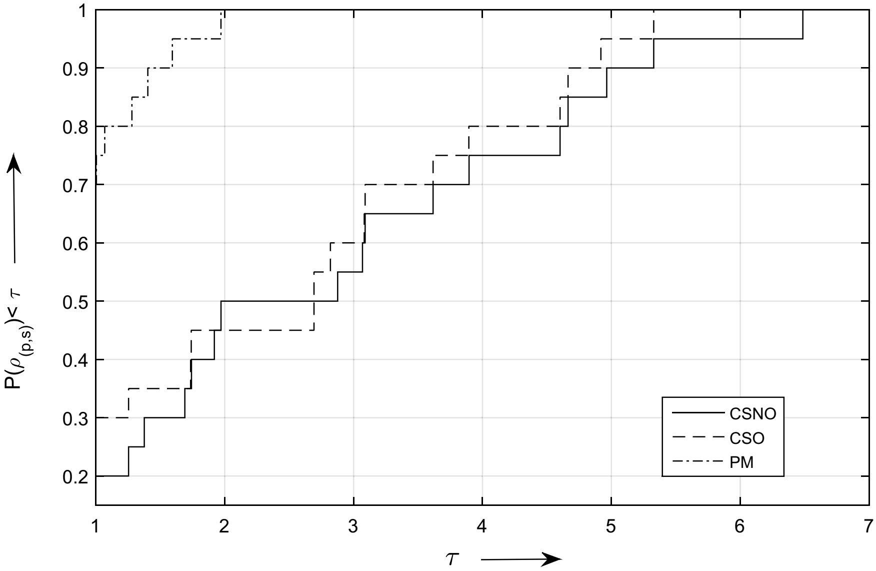


Fig. 4 Performance profile for iteration

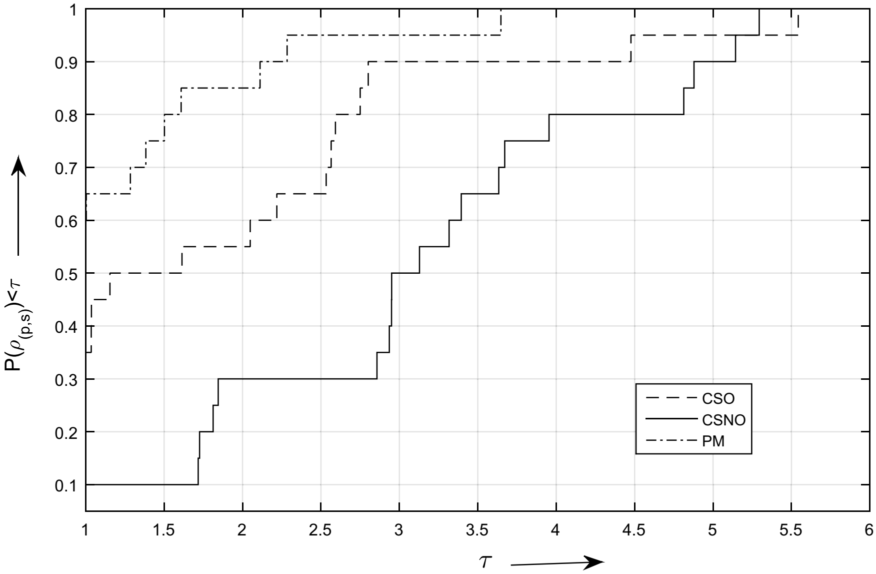


Fig. 5 Performance profile for objective function evaluation

6 Conclusion

The basic two coordinate search methods are very simple and easy to implement. In this paper, keeping their basic structure as intact, an idea of expansion of stencil and rotation of maximal positive basis vectors at the iteration point has been introduced. The global convergence of the proposed scheme is proved, and the advantage of its implementation is justified with numerical illustration on a test set of benchmark problems. The scope of future modification of the scheme is possible by imposing a further restriction on the conditions of expansion of the n-dimensional stencil.

References

1. Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim **10**(1), 147–161 (2008)
2. Barbosa, H.J.C., Bernardino, H.S., Barreto, A.M.S.: Using performance profiles to analyze the results of the 2006 CEC constrained optimization competition. In: 2010 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. IEEE (2010)
3. Chakraborty, S.K., Panda, G.: Descent line search scheme using Gersgorin circle theorem. Oper. Res. Lett. **45**(6), 565–569 (2017)
4. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. SIAM (2009)
5. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Progr. **91**(2), 201–213 (2002)
6. Frandi, E., Papini, A.: Coordinate search algorithms in multilevel optimization. Optim. Methods Softw. **29**(5), 1020–1041 (2014)
7. Hooke, R., Jeeves, T.A.: Direct search solution of numerical and statistical problems. J. ACM (JACM) **8**(2), 212–229 (1961)
8. Hu, J., Fu, M.C., Marcus, S.I.: A model reference adaptive search method for global optimization. Oper. Res. **55**(3), 549–568 (2007)
9. Huyer, W., Neumaier, A.: Global optimization by multilevel coordinate search. J. Glob. Optim. **14**(4), 331–355 (1999)
10. Kelley, C.T.: Iterative Methods for Optimization. SIAM (1999)
11. Kelley, C.T.: Implicit Filtering. SIAM (2011)
12. Lewis, R.M., Torczon, V.: Rank ordering and positive bases in pattern search algorithms. Technical report, Institute for Computer Applications in Science and Engineering Hampton VA, 1996
13. Moré, J.J., Wild, S.M.: Benchmarking derivative-free optimization algorithms. SIAM J. Optim. **20**(1), 172–191 (2009)
14. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer Series in Operations Research and Financial Engineering, Springer, NY (2006)
15. Torczon, V.: On the convergence of pattern search algorithms. SIAM J. Optim. **7**(1), 1–25 (1997)

Dynamics of the Logistic Prey Predator Model in Crisp and Fuzzy Environment

S. Tudu, N. Mondal and S. Alam

Abstract Recently the study of fuzzy dynamical system is growing rapidly in various field specially in biological system dynamics. In this article a dynamical model of two species population has been studied taking intrinsic growth rate, natural mortality rate and rate of conversion as triangular fuzzy number. Here the dynamics of the model system was discussed both in fuzzy and crisp environment. Also the analytical finding has been supported through numerical simulations.

Keywords Prey-predator · Stability · Graded mean value · Fuzzy set

1 Introduction

In the year 1838, a very realistic logistic model was developed by Belgian mathematician Verhulst [1] where logistic growth was projected as appropriate growth term for biological species. This Logistic model is applicable and more appropriate when overcrowding and many competition resources are in the population. In the area of theoretical biology two species prey predator model was firstly initiated by Lotka [2] and Volterra [3]. For the improvement of human society, it is very much essential how to utilise the biological resources. In the last decades, ordinary differential equation is used to modelling this type of problems and provided some mathematical answers and explanations [4–7] in this regard. It may be noted that biological parameters used in the differential equations are not always fixed. In ecosystem many parameters oscillates simultaneously with the periodically varying environments. They are also varying due to both human society and nature, such as earthquake, fire, financial crisis, climate warming etc. In real world periodical environmental cycle is vary common. Furthermore, there are other common influence to varies the model parameters, for example temperature variations highly influenced the reproduction rate of bacteria

S. Tudu · N. Mondal · S. Alam (✉)
Department of Mathematics, Indian Institute of Engineering Science and Technology,
P.O.: Botanic Garden, Howrah 711103, West Bengal, India
e-mail: salam50in@yahoo.co.in

P. K. Roy et al. (eds.), *Mathematical Analysis and Applications in Modeling*,
Springer Proceedings in Mathematics & Statistics 302,
https://doi.org/10.1007/978-981-15-0422-8_37

during the day time. Also, the interaction between the species and consequently the dynamics of the whole system are highly influenced by this environmental variations. Many researches made an attempt to explain the system dynamics incorporating the concept of stochastical fluctuation arguing the fact that the environmental change is more or less stochastical in nature. In real world problems, some parameters of model system can be roughly estimated incorporating the said stochastical concept. But there is a question arise the choice of suitable probability distribution function for all imprecise parameters and it makes the model very complicated, to overcome this difficulty there is need of fuzzy set theory [8]. In view of behaviour of population communities and varies environmental factors, it is natural to consider the biological parameters in emphasize environment. The fuzzy prey-predator model may be more meaningful than the deterministic prey-predator model. There are some literature present in the field of biomathematics under emphasise environment. Bassanezi et al. [9] considered the variables and parameters of fuzzy dynamical system as imprecise in nature and used fuzzy differential equation to study the stability of the system. Barros et al. [10] used environmental fuzziness of a life expectancy model by taking the parameters as fuzzy in nature. Guo et al. [11] studied logistic model and gompertz model under fuzziness. Mizukoshi et al. [12] considered the fuzzy initial value problem with parameters or initial conditions under fuzziness. Peixoto et al. [13] and Pal et al. [14] represented prey predator model under fuzziness.

In this article we consider a simple prey-predator model with one prey and predator species under imprecise environment where biological parameters are taken as fuzzy parameters. We analyse the model system both in crisp environment as well as fuzzy environment. In fuzzy model, for simplicity, we consider all the biological parameters as triangle fuzzy numbers. We also introduce a noble concept of Graded Mean Value technique to defuzzify the parameters and the system dynamics is interpreted in different optimistic level.

2 Preliminaries

For the development of the model in fuzzy environment we need some preliminary definition and concepts which will be used later on. Zadeh [8] first introduced fuzzy sets as a mathematical way of representing vagueness in everyday life.

Definition 1 Fuzzy set: Let X be a non empty set. A fuzzy set A in X is characterised by its membership function $\mu_A : X \to [0, 1]$ and $\mu_A(x)$ is interpreted as the degree of membership of element x in fuzzy set A for each $x \varepsilon X$. The fuzzy set A is completely determined by the set tuples $A = \{(x, \mu_A(x)) : x \varepsilon X\}$.

Definition 2 α-cut and strong α-cut of a fuzzy set: Let A be a given fuzzy set defined on a non empty set X and μ be its membership function. Then for any $\alpha \varepsilon [0, 1]$ the α-cut and strong α-cut are denoted by μ^{α} and $\mu^{\alpha+}$ and defined by $\mu^{\alpha} = \{x \mid \mu_A(x) \geq \alpha, \forall x \varepsilon X\}$ and $\mu^{\alpha+} = \{x \mid \mu_A(x)\ \alpha, \forall x \varepsilon X\}$.

Definition 3 Fuzzy Number: A fuzzy number u is a pair $(\underline{u}, \bar{u})$ of functions $\underline{u}(\alpha), \bar{u}(\alpha)$; $0 \leqslant \alpha \leqslant 1$ which satisfy the following requirements:

1. $\underline{u}(\alpha)$ is bounded monotonic increasing left continuous function.
2. $\bar{u}(\alpha)$ is bounded monotonic increasing left continuous function.
3. $\underline{u}(\alpha) \leqslant \bar{u}(\alpha), 0 \leqslant \alpha \leqslant 1$.

Definition 4 Triangular fuzzy number: A triangular fuzzy number (TFN) is represented by $\tilde{A} = (a, b, c)$ and its membership function is defined as bellow

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, \text{if } x \leqslant a \\ \frac{x-a}{b-a}, \text{if } a \leqslant x \leqslant b \\ \frac{c-x}{c-b}, \text{if } b \leqslant x \leqslant c \\ 0, \text{if } x \geqslant c \end{cases} \tag{1}$$

Definition 5 Graded mean value: Let $\lambda \varepsilon [0, 1]$ be a pre-assigned parameter called the degree of optimism. The Graded mean Value Chen (1985), Mahapatra and Roy (2003) or total λ integral value of $\tilde{a}$ is defined as convex combination of the right and left integral values through the degree of optimism and it is denoted by

$$I_\lambda(\tilde{a}) = \lambda I_r(\tilde{a}) + (1-\lambda) I_l(\tilde{a}) \tag{2}$$

where $I_l(a)$ and $I_r(a)$ are the left and right integral values of $\tilde{a}$ defined as

$$I_l(\tilde{a}) = \int_0^1 (\mu_{l\tilde{a}})^{-1} \alpha d\alpha \quad \text{and} \quad I_r(\tilde{a}) = \int_0^1 (\mu_{r\tilde{a}})^{-1} \alpha d\alpha \tag{3}$$

For a triangular fuzzy number $\tilde{a} = (a_1, a_2, a_3)$

$$(\mu_{l\tilde{a}})^{-1} \alpha = a_1 + \alpha(a_2 - a_1) \quad \text{and} \quad (\mu_{r\tilde{a}})^{-1} \alpha = a_3 - \alpha(a_3 - a_2) \tag{4}$$

Therefore the left and right integral values are

$$I_l(\tilde{a}) = \frac{a_1 + a_2}{2} \text{ and } I_r(\tilde{a}) = \frac{a_3 + a_2}{2} \tag{5}$$

Hence the total λ-integral values of $\tilde{a}$ is

$$I_\lambda(\tilde{a}) = \left[\lambda \left(\frac{a_3 + a_2}{2} \right) + (1-\lambda) \left(\frac{a_1 + a_2}{2} \right) \right] \tag{6}$$

The pessimistic viewpoint of a decision maker is reflected by left integral value and optimistic viewpoint is reflected by the right integral value. The highest degree of optimism is specified by a large value of λ. As for example when $\lambda = 1$, the total integral value $I_l(\tilde{a}) = (\frac{a_2+a_3}{2}) = I_r(\tilde{a})$ represents an optimistic standpoint on the contrary, when $\lambda = 0$ the total integral is $I_l(\tilde{a}) = (\frac{a_1+a_2}{2}) = I_r(\tilde{a})$ represents an optimistic viewpoint when $\lambda = 0.5$ the total integral value reflects a moderately optimistic decision maker's standpoint. Similarly $\lambda = 0.5$ and $\lambda = 0.7$ represents about optimistic and about pessimistic standpoints of the decision maker's respectively.

2.1 Crisp Logistic Prey-Predator Model

The mathematical expression of the logistic prey-predator model with two species precise biological parameters is given by

$$\begin{aligned} \frac{dx}{dt} &= rx\left(1 - \frac{x}{k}\right) - axy \\ \frac{dy}{dt} &= axy - dy \end{aligned} \tag{7}$$

where x, and y denotes the population density of the prey and predator species respectively at time t. In this model $r(> 0)$ denotes the intrinsic growth rate of the prey species in the absence of predator and $d(> 0)$ is natural death rate of the predator in the absence of the prey. Finally a denotes the conversion coefficient of prey biomass into predator biomass.

2.2 Fuzzy Logistic Prey-Predator Model

Due to the imprecise nature of the intrinsic growth rate (r) of the prey species in the absence of predator we consider r as a triangular fuzzy number. Similarly for the imprecise nature of the natural mortality rate (d) of the predator species in the absence of prey population and finally due to the imprecise nature of the conversion coefficient (a) of prey biomass into predator biomass we consider both a and d as triangular fuzzy number. Consequently the crisp logistic prey-predator model becomes the Fuzzy logistic prey-predator model and is given by

$$\begin{aligned} \frac{d\tilde{x}}{dt} &= \tilde{r}x\left(1 - \frac{x}{k}\right) - \tilde{a}xy \\ \frac{d\tilde{y}}{dt} &= \tilde{a}xy - \tilde{d}y \end{aligned} \tag{8}$$

3 Positive Invariance of the System

Let $Y = \begin{pmatrix} x \\ y \end{pmatrix}$, where $Y \in \mathbb{R}^2$ $Y = \begin{pmatrix} F_1(Y) \\ F_2(y) \end{pmatrix}$, i.e $Y = \begin{pmatrix} rx\left(1 - \frac{x}{k}\right) - axy \\ axy - dy \end{pmatrix}$ where $F : C_+ \to \mathbb{R}^2$ and $F \in C^\infty(\mathbb{R}^2)$ and C^∞ stands for continuously differentiable function. The above system becomes, $\dot{Y} = F(Y)$ with $Y(\theta) = (\phi_1(\theta), \phi_2(\theta)) \in C_+$ and $\phi_i(\theta) > 0$ for $i = 1, 2$. It is easy to check in the above equation that whenever choosing $Y(\theta) = 0$ such that $Y_i = 0$, then $F_i(Y)|y_i(t) \in C_+ \geq 0$ for $i = 1, 2$. Due to Lemma (Yang et al. [2, 4]) any solution of the above equation with $Y(\theta) \in C_+$, say $Y(t) = Y(t, Y(\theta))$ such that $Y(\theta) \in \mathbb{R}^2 \forall t > 0$.

4 Boundedness of the System

Theorem 1 *The solutions of the above system is bounded.*

Proof Let us define $W = x + y$. The time derivatives

$$\frac{dW}{dt} = \frac{dx}{dt} + \frac{dy}{dt} = rx\left(1 - \frac{x}{k}\right) - dy$$

Now

$$\frac{dW}{dt} + qW = rx\left(1 - \frac{x}{k}\right) - dy + q(x + y) \tag{9}$$

$$= \frac{k(r+q)^2}{4r} - \frac{r}{k}\left(x - \frac{k(r+q)}{2r}\right)^2 + (q - d)y \tag{10}$$

if $q \leq d$, Then $\frac{dW}{dt} + qW \leq M$ where $M = \frac{k(r+q)^2}{4r}$ (say). Therefore, $\frac{dW}{dt} + qW \leq M(constant)$, which is a linear differential equation in W. After solving we get, $W \leq \frac{M}{q} + Ce^{-qt}$, where C is an integrating constant. At $t = 0$, $W = 0$, we have $C \geq -\frac{M}{q}$. Therefore, $W \leq \frac{M}{q}(1 - e^{-qt})$ and $W \geq 0$, since $W = x + y$, x and y both are ≥ 0 So, $0 \leq W(x(t), y(t)) \leq \frac{M}{q}(1 - e^{-qt})$, which implies that $0 \leq W(x(t), y(t)) \leq \frac{M}{q}$ as $t \to \infty$. Hence all the solution of the above system are bounded and so we can now analyze the stability of the system.

5 Dimensionless of the Logistic Model

$$\begin{aligned} \frac{dx}{dt} &= rx\left(1 - \frac{x}{k}\right) - axy \\ \frac{dy}{dt} &= axy - dy \end{aligned} \tag{11}$$

Let $U = \frac{x}{k}$, $V = ay$. The Dimensionless system becomes

$$\begin{aligned}\frac{dU}{dt} &= rU\,(1-U) - UV\\ \frac{dV}{dt} &= (cU-d)V\end{aligned}\tag{12}$$

where $ak = c$ (say), r, d all are positive constants.

6 Logistic Prey-predator Model in Different Environment

In this section we will discuss the logistic prey-predator model in crisp and fuzzy environment.

Equilibrium points and Existence Criteria:
The equilibrium points of (12) are given by $E_0(0,0)$, $E_1(1,0)$, and $E_2\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$. $E_0(0,0)$ is called trivial equilibrium point while $E_1(1,0)$ are called axial equilibrium point and $E_2\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is called interior as well as planer equilibrium point. $E_0(0,0)$ and $E_1(1,0)$ exists with out any restriction. But $E_2\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ exists if $c > d$.

6.1 Stability Analysis of System in Crisp Environment

$$\begin{aligned}\frac{dU}{dt} &= rU\,(1-U) - UV && = F(U,V)\\ \frac{dV}{dt} &= (cU-d)V && = G(U,V)\end{aligned}\tag{13}$$

Theorem 2 *The equilibrium point $E_0(0,0)$ is unstable.*

Proof The proof is obvious.

Theorem 3 *The equilibrium point $E_1(1,0)$ is stable if $c < d$ otherwise it is unstable.*

Proof From the variational matrix the eigen values are $-r$, and $c-d$. Again, Since $-r$ is negative constant and if $c-d < 0$ then the point $E_1(1,0)$ is stable and otherwise it is unstable.

Theorem 4 *The Interior Equilibrium Point $E^*\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is always asyamptotically stable.*

Proof The characteristic equation at $E^*\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is given by $\Rightarrow \lambda^2 + \frac{rd}{c}\lambda + rd(1-\frac{d}{c}) = 0 \Rightarrow \lambda = -\frac{rd}{2c} \pm \frac{1}{2c}\sqrt{(rd)^2 - 4cdr(c-d)}$
There may arise three cases:
Case I: When $rd = 4c(c-d), Then\ \lambda = -\frac{rd}{2c}$ (of multiplicity 2), The two eigen value are negative. So the interior point $E^*\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is stable.
Case II: Clearly, from the Theorem 4 we can conclude that one eigen value is negative and other is negative if $\sqrt{(rd)^2 - 4cdr(c-d)} < rd$, so two eigen value is negative and hence the interior point $E^*\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is stable otherwise it is unstable.
Case III: Finally, when $rd < 4c(c-d)$, Then the eigen values reduces to in the form $\lambda = -\frac{rd}{2c} \pm i\left(\frac{1}{2c}\sqrt{4cdr(c-d) - (rd)^2}\right)$, which is complex number (eigen value) but real part of the complex eigen value is negative, so the interior point $E^*\left(\frac{d}{c}, r(1-\frac{d}{c})\right)$ is Asymptotically Stable.

6.2 Logistic Prey-Predator Model in Fuzzy Environment

In this section we will discuss the logistic prey-predator model in fuzzy environment.

Theorem 5 *The fuzzy differential equations*

$$\frac{d\tilde{U}}{dt} = \tilde{r_0}U(1-U) - UV \tag{14}$$

$$\frac{d\tilde{V}}{dt} = (\tilde{c_0}U - \tilde{d_0})V \tag{15}$$

where $\tilde{r_0} = (r_{01}, r_{02}, r_{03})$, $\tilde{c_0} = (c_{01}, c_{02}, c_{03})$ and $\tilde{d_0} = (d_{01}, d_{02}, d_{03})$ for positive real numbers r_{0i}, c_{0i} and d_{0i}, $i = 1, 2, 3$ is provided by the differential equation

$$\frac{dU}{dt} = I_\lambda(\tilde{r_0})U(1-U) - UV \tag{16}$$

$$\frac{dv}{dt} = (I_\lambda(\tilde{c_0})U - I_\lambda(\tilde{d_0}))V \tag{17}$$

where $I_\lambda(\tilde{r_0}) = (1-\lambda)\frac{r_{01}+r_{02}}{2} + \lambda\frac{r_{02}+r_{03}}{2}$, $I_\lambda(\tilde{c_0}) = (1-\lambda)\frac{c_{01}+c_{02}}{2} + \lambda\frac{c_{02}+c_{03}}{2}$, $I_\lambda(\tilde{d_0}) = (1-\lambda)\frac{d_{01}+d_{02}}{2} + \lambda\frac{d_{02}+d_{03}}{2}$, $\forall\lambda\varepsilon[0, 1]$.

Proof The given fuzzy differential equation can be written as

$$\frac{d\tilde{U}}{dt} = (r_{01}, r_{02}, r_{03})U(1-U) - UV \tag{18}$$

$$\frac{d\tilde{V}}{dt} = (c_{01}, c_{02}, c_{03})U - (d_{01}, d_{02}, d_{03})V \tag{19}$$

Taking the α-cut of the triangular fuzzy numbers $\tilde{r_0}$, $\tilde{c_0}$ and $\tilde{d_0}$ Eqs. (14) and (15) can be written as

$$[(\mu^L_{\frac{d\tilde{U}}{dt}})^{-1}\alpha, (\mu^R_{\frac{d\tilde{U}}{dt}})^{-1}\alpha] = [(\mu^L_{\tilde{r_0}})^{-1}\alpha, (\mu^R_{\tilde{r_0}})^{-1}\alpha]U(1-U) - UV \tag{20}$$

$$[(\mu^L_{\frac{d\tilde{V}}{dt}})^{-1}\alpha, (\mu^R_{\frac{d\tilde{V}}{dt}})^{-1}\alpha] = [(\mu^L_{\tilde{c_0}})^{-1}\alpha, (\mu^R_{\tilde{c_0}})^{-1}\alpha]U - [(\mu^L_{\tilde{d_0}})^{-1}\alpha, (\mu^R_{\tilde{d_0}})^{-1}\alpha]V \tag{21}$$

where $(\mu^L_{\tilde{r_0}})^{-1}\alpha = r_{01} + \alpha(r_{02} - r_{01})$, $(\mu^R_{\tilde{r_0}})^{-1}\alpha = r_{03} - \alpha(r_{03} - r_{02})$, $(\mu^L_{\tilde{c_0}})^{-1}\alpha = c_{01} + \alpha(c_{02} - c_{01})$, $(\mu^R_{\tilde{c_0}})^{-1}\alpha = c_{03} - \alpha(c_{03} - c_{02})$, $(\mu^L_{\tilde{d_0}})^{-1}\alpha = d_{01} + \alpha(d_{02} - d_{01})$ and $(\mu^R_{\tilde{d_0}})^{-1}\alpha = d_{03} + \alpha(d_{03} - d_{02})$.
Now following the properties of Graded Mean Value Technique the differential equations (16) and (17) can be written as

$$\frac{dU}{dt} = [(1-\lambda)I_L(\tilde{r_0}) + \lambda I_R(\tilde{r_0})]U(1-U) - UV$$
$$\frac{dV}{dt} = [(1-\lambda)I_L(\tilde{c_0}) + \lambda I_R(\tilde{c_0})]U - ([(1-\lambda)I_L(\tilde{d_0}) + \lambda I_R(\tilde{d_0})])V$$

where $I_L(\tilde{r_0}) = \int_0^1 (\mu^L_{\tilde{r_0}})^{-1}\alpha d\alpha = (\frac{r_{01}+r_{02}}{2})$
$I_R(\tilde{r_0}) = \int_0^1 (\mu^R_{\tilde{r_0}})^{-1}\alpha d\alpha = (\frac{r_{02}+r_{03}}{2})$, similarly $I_L(\tilde{c_0}) = (\frac{c_{01}+c_{02}}{2})$, $I_R(\tilde{c_0}) = (\frac{c_{02}+c_{03}}{2})$, $I_L(\tilde{d_0}) = (\frac{d_{01}+d_{02}}{2})$ and $I_R(\tilde{d_0}) = (\frac{d_{02}+d_{03}}{2})$.

Therefore the crisp form of the fuzzy differential equation is given by

$$\frac{dU}{dt} = I_\lambda(\tilde{r_0})U(1-U) - UV \tag{22}$$
$$\frac{dv}{dt} = (I_\lambda(\tilde{c_0})U - I_\lambda(\tilde{d_0}))V \tag{23}$$

and finally the theorem is proved.

Equilibrium points and Existence Criteria: The equilibrium points of the above system are given by $E^0(0,0)$, $E^1(1,0)$ and $E^*\left(\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}, I_\lambda(\tilde{r_0})(1 - \frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})})\right)$. $E^0(0,0)$ and $E^1(1,0)$ exists without any conditions but the interior point $E^*\left(\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}, I_\lambda(\tilde{r_0})(1 - \frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})})\right)$ exits if $I_\lambda(\tilde{c_0}) > I_\lambda(\tilde{d_0})$. Stability conditions of the equilibrium points of system (22) and (23) are stated in the following theorems.

Theorem 6 *The trivial equilibrium point $E^0(0,0)$ is always unstable.*

Proof The proof is obvious.

Theorem 7 *The axial equilibrium point $E^1(1,0)$ is stable if $I_\lambda(\tilde{c_0}) < I_\lambda(\tilde{d_0})$ and otherwise it is unstable.*

Proof Calculating the Jacobian matrix at $E_1(1,0)$ we have the eigenvalues $-I_\lambda(\tilde{r_0}) < 0$ and $I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})$. Now if $I_\lambda(\tilde{c_0}) < I_\lambda(\tilde{d_0})$ then the equilibrium point $E_1(1,0)$ is stable and otherwise it is unstable.

Theorem 8 *The interior equilibrium point* $E_2\left(\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}, I_\lambda(\tilde{r_0})(1-\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})})\right)$

(i) Stable if $I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0}) = 4I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right)$
(ii) Stable but not asymptotically stable if $I_\lambda(\tilde{c_0}) = I_\lambda(\tilde{d_0})$
(iii) Stable if $I_\lambda(\tilde{c_0}) > I_\lambda(\tilde{d_0})$
(iv) Asymptotically stable if $I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0}) < 4I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right)$.

Proof The characteristic equation at $E_2\left(\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}, I_\lambda(\tilde{r_0})(1-\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})})\right)$ corresponding to the jacobian matrix is given by

$$\lambda^2 + \frac{I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}\lambda + \frac{I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right) = 0 \tag{24}$$

(i) From the characteristic equation it is clear that when $I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0}) = 4I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right)$ then $\lambda = -\frac{I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})}{2I_\lambda(\tilde{c_0})}$ (two times), which is negative so the interior equilibrium point is stable.
(ii) Again, from the above characteristic equation when $I_\lambda(\tilde{c_0}) = I_\lambda(\tilde{d_0})$ then the two eigenvalues are 0 and $-\frac{I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}$ since one eigenvalue is zero (0) and other is negative, so the equilibrium point is stable but not Asymptotically stable.
(iii) To prove the stability condition $I_\lambda(\tilde{c_0}) > I_\lambda(\tilde{d_0}$, from the characteristic equation we have
$-I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0}) \pm \sqrt{\left(I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})\right)^2 - 4I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right)},$
this shows that one eigenvalue is always negative and other is negative if
$4I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0})I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right) > 0$, i.e. $I_\lambda(\tilde{c_0}) > I_\lambda(\tilde{d_0})$.
(iv) Finally, to prove the stability condition from Eq. (24) we see that the two eigenvalues are complex with negative real part if $I_\lambda(\tilde{r_0})I_\lambda(\tilde{d_0}) < 4I_\lambda(\tilde{c_0})\left(I_\lambda(\tilde{c_0}) - I_\lambda(\tilde{d_0})\right)$. Since the real part of the eigenvalues are negative so the interior equilibrium point is asymptotically stable and otherwise it is unstable.

7 Numerical Example

For a numerical example let us take the intrinsic growth rate, the conversion rate and the mortality rate as triangular fuzzy numbers given by $\tilde{r_0} = (0.9, 1.2, 1.5)$, $\tilde{c_0} = (0.3, 0.4, 0.5)$ and $\tilde{d_0} = (0.2, 0.3, 0.5)$ then their λ-integrals are

$$I_\lambda(\tilde{r_0}) = 1.05(1-\lambda) + 1.35\lambda \tag{25}$$

$$I_\lambda(\tilde{c_0}) = 0.35(1-\lambda) + 0.45\lambda \tag{26}$$

$$\text{and } I_\lambda(\tilde{d_0}) = 0.25(1-\lambda) + 0.4\lambda \tag{27}$$

For $\lambda = 0$, $I_\lambda(\tilde{r_0}) = 1.05$, $I_\lambda(\tilde{c_0}) = 0.35$ and $I_\lambda(\tilde{d_0}) = 0.25$.

Now from the Theorem 6 we see that the trivial equilibrium point $E_0(0, 0)$ is always unstable as one of the eigen values $I_\lambda(\tilde{r_0}) > 0$ and the another eigen value $I_\lambda(\tilde{d_0}) < 0$, Here $I_\lambda(\tilde{c_0}) > I_\lambda(\tilde{d_0})$ and hence from Theorem 7 we can say that the axial equilibrium point $E_1(1, 0)$ is unstable equilibrium point. Similarly using the condition (iii) of Theorem 8 we can conclude that the interior equilibrium point $E_2(\frac{5}{7}, 3)$ is stable equilibrium point. To represent the stability of the predator prey system here we give a graphical representation of the system taking the above mentioned integral values of $\tilde{r_0}$, $\tilde{c_0}$ and $\tilde{d_0}$ for $\lambda = 0$.

Taking the particular values for the parameter of the logistic prey predator model we have simulate predator-prey verses time solution graph in MATLAB. Here the values of the parameter are taken as $r = 0.5, k = 12, a = 0.3, d = 0.4, ts = [0, 200], z0 = [0.4, 0.8]$ for Fig. 1. It is clear from Fig. 1 the prey-predator populations becomes stable after a certain stage of time period.

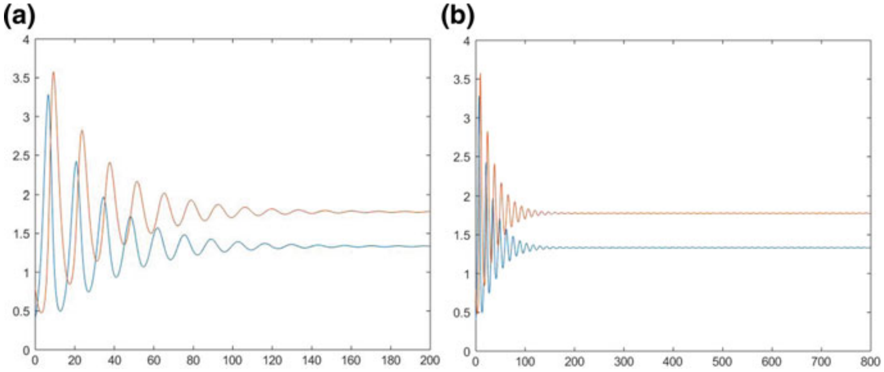


Fig. 1 **a** and **b** depicts the fact that the model system will reach to stable coexistence point E^* while **a** shows the system is stable for 0 to 200 and **b** is also stable for long time of period varies from 0 to 800

8 Conclusion

Prey-Predator model has different development in theoretical and practical applications in the field of biomathematics. Most of the researchers have developed the prey-predator model based on the assumption that the biological parameters are precisely known but the scenario is different in real life situation. In this article, we have developed a method to find the biological equilibrium points, when some biological parameters are imprecise in nature. In this article we have considered the biological parameters as triangular fuzzy number and have analysed the system dynamics of the predator prey model taking the defuzzification value of the triangular fuzzy numbers. Here for defuzzification we have used graded mean integral value technique. Again from Fig. 1b we can see that the interior equilibrium point $E_2\left(\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})}, I_\lambda(\tilde{r_0})(1-\frac{I_\lambda(\tilde{d_0})}{I_\lambda(\tilde{c_0})})\right)$ remains stable for long time period where $I_\lambda(\tilde{r_0}) = (1-\lambda)(\frac{r_{01}+r_{02}}{2}) + \lambda(\frac{r_{02}+r_{03}}{2})$, $I_\lambda(\tilde{c_0}) = (1-\lambda)(\frac{c_{01}+c_{02}}{2}) + \lambda(\frac{c_{02}+c_{03}}{2})$ and $I_\lambda(\tilde{d_0}) = (1-\lambda)(\frac{d_{01}+d_{02}}{2}) + \lambda(\frac{d_{02}+d_{03}}{2})$, $\forall\lambda\varepsilon[0, 1]$. $I_\lambda(\tilde{r_0})$, $I_\lambda(\tilde{c_0})$ and $I_\lambda(\tilde{d_0})$ all are the total λ integral of $\tilde{r_0}$, $\tilde{c_0}$ and $\tilde{d_0}$. The pessimistic viewpoint of a decision maker is reflected by left integral value and optimistic viewpoint is reflected by the right integral value. The highest degree of optimism is specified by a large value of λ. As for example when $\lambda = 1$, the total integral value $I_l(\tilde{r_0}) = (\frac{r_{02}+r_{03}}{2}) = I_r(\tilde{r_0})$ represents an optimistic standpoint on the contrary, when $\lambda = 0$ the total integral is $I_l(\tilde{r_0}) = (\frac{r_{01}+r_{02}}{2}) = I_r(\tilde{r_0})$represents an optimistic viewpoint when $\lambda = 0.5$ the total integral value reflects a moderately optimistic decision maker's standpoint. Similarly $\lambda = 0.5$ and $\lambda = 0.7$ represents about optimistic and about pessimistic standpoints of the decision maker's respectively. Similarly, this concept is applicable for $I_\lambda(\tilde{c_0})$ and $I_\lambda(\tilde{d_0})$. Form Fig. 2 shows that the prey versus predator population is stable and

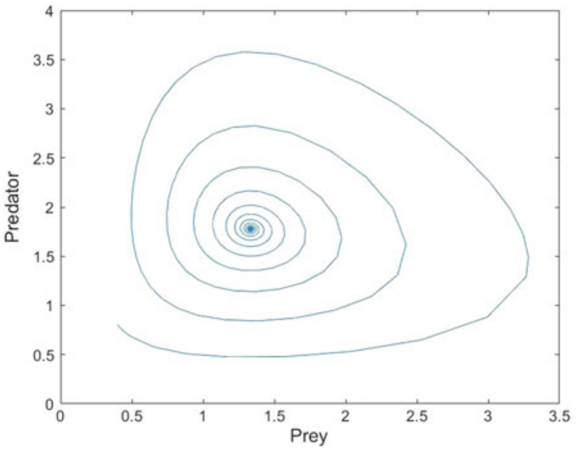


Fig. 2 Figure shows that the model system will goes to stable focus. Here, predator versus prey solution in crisp environment

depicts the fact that the population will reach to stable focus. From Fig. 3a–f we draw the graph for different value of λ, such as $\lambda = 0$, $\lambda = 0.3$, $\lambda = 0.5$, $\lambda = 0.7$, $\lambda = 0.85$ and $\lambda = 1$ and see that there is no change of stability for different value of λ that is the system is always stable. Here in this article we try to project how the dynamics is changing from crisp to Fuzzy environment taking various level of optimistic view.

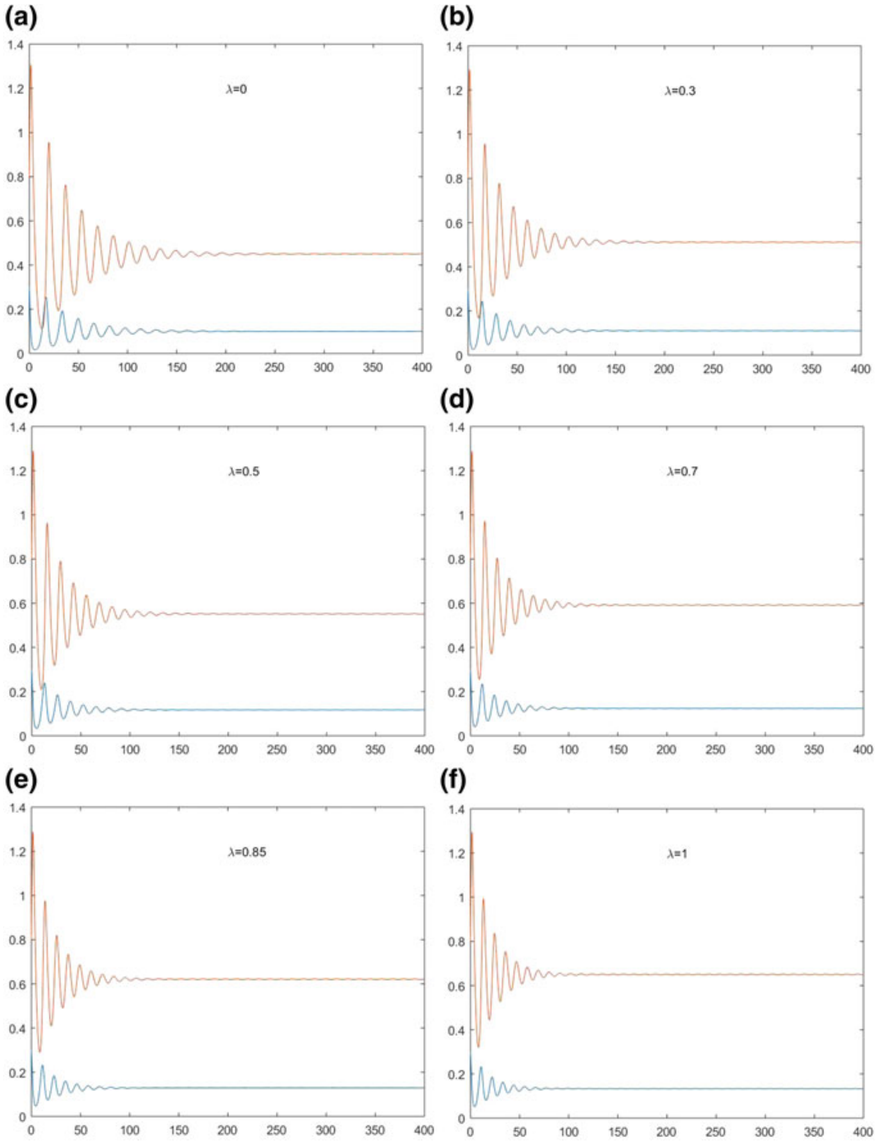


Fig. 3 Depicts the stability of the system for different values of λ

References

1. Verhulst, P.F.: Notice sur la loi que la population suit dans son accroissement. Corresp. Math. Phys. **10**, 113–121 (1838)
2. Lotka, A.J.: Elements of Physical Biology. Williams and Wilkins, Baltimore (1925)
3. Volterra, V.: Lecons sur la theorie mathematique de la lutte pour la vie. Gauthier-Villars, Paris (1931)
4. Shih, S.D., Chow, S.S.: Equivalence of n-point Gauss-Chebyshev rule and 4n-point midpoint rule in computing the period of a Lotka-Volterra system. Adv. Comput. Math. **28**, 63–79 (2008)
5. Liao, X., Chen, Y., Zhou, S.: Traveling wavefronts of a prey-predator diffusion system with stage-structure and harvesting. J. Comput. Appl. Math. **235**, 2560–2568 (2011)
6. Qu, Y., Wei, J.: Bifurcation analysis in a time-delay model for prey-predator growth with stage-structure. Nonlinear Dyn. **49**, 285–294 (2007)
7. Seo, G., DeAngelis, D.L.: A predator-prey model with a Holling type I functional response including a predator mutual interference. J. Nonlinear Sci. **21**, 811–833 (2011)
8. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
9. Bassanezi, R.C., Barros, L.C., Tonelli, A.: Attractors and asymptotic stability for fuzzy dynamical systems. Fuzzy Sets Syst. **113**, 473–483 (2000)
10. Barros, L.C., Bassanezi, R.C., Tonelli, P.A.: Fuzzy modelling in population dynamics. Ecol. Model. **128**, 27–33 (2000)
11. Guo, M., Xu, X., Li, R.: Impulsive functional differential inclusions and fuzzy population models. Fuzzy Sets Syst. **138**, 601–615 (2003)
12. Mizukoshi, M.T., Barros, L.C., Bassanezi, R.C.: Stability of fuzzy dynamic systems. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **17**, 69–84 (2009)
13. Peixoto, M., Barros, L.C., Bassanezi, R.C.: Predator-prey fuzzy model. Ecol. Model. **214**, 39–44 (2008)
14. Pal, D., Mahaptra, G.S., Samanta, G.P.: Quota harvesting model for a single species population under fuzziness. IJMS **12**(1–2), 33–46 (2013)