Leonardo Bacelar Lima Santos
Rogério Galante Negri
Tiago José de Carvalho *Editors*

Towards Mathematics, Computers and Environment: A Disasters Perspective

Springer

Towards Mathematics, Computers and Environment: A Disasters Perspective

Leonardo Bacelar Lima Santos
Rogério Galante Negri • Tiago José de Carvalho
Editors

Towards Mathematics, Computers and Environment: A Disasters Perspective

Springer

Editors
Leonardo Bacelar Lima Santos
National Centre for Monitoring and Early Warnings of Natural Disasters (CEMADEN)
São José dos Campos, São Paulo, Brazil

Rogério Galante Negri
Institute of Science and Technology
São Paulo State University (UNESP)
São José dos Campos, São Paulo, Brazil

Tiago José de Carvalho
Department of Informatics
Federal Institute of São Paulo (IFSP)
Campinas, São Paulo, Brazil

ISBN 978-3-030-21204-9 ISBN 978-3-030-21205-6 (eBook)
https://doi.org/10.1007/978-3-030-21205-6

Mathematics Subject Classification (2010): 86A10, 86A05, 86A60, 86A17, 86A22

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword by Dan Crisan

Climate change and the related challenges facing humanity have emerged as major societal priorities in recent times. The scientific community has responded strongly to these challenges, and, in particular, mathematicians are heavily involved in leading the research agenda particularly through the new area of Mathematics: Mathematics of Planet Earth. Similar to Mathematical Biology and Mathematical Finance, Mathematics of Planet Earth is defined not through its subject matter, but through its area of application. Its applications are directed toward the planetary issues that we face today: climate change, quantification of uncertainty, moving to an economy of sustainability, preservation of biodiversity, natural hazards, financial and social systems, adaptation to change, and many others. Such issues give rise to an abundance of challenging scientist problems with a strong multidisciplinary characteristics.

Recent initiatives within the international mathematics community are answering the need for interdisciplinary quantitative scientists to tackle diverse Planet Earth challenges. The Society for Industrial and Applied Mathematics (SIAM) has a Mathematics of Planet Earth activity group (www.siam.org/activity/mpe); the second SIAM MPE conference took place in September 2018 (www.siam.org/Conferences/CM/Main/mpe18). The Mathematics of Climate Network (mcrn.hubzero.org) coordinates with the researchers across the USA, while a research network in MPE has been established in the Netherlands (mathplanetearth.nl). The European Geophysical Union (EGU) has established an annual MPE theme in their General Assembly. Many other activities in this area take place all around the globe.

The book "*Towards Mathematics, Computers and Environment: A Disasters Perspective*" is a great addition to the scientific effort and the literature on Mathematics of Planet Earth. The topics covered are of exceptional importance. Examples include pollutant dispersion modeling, real-time monitoring and early warning of convective weather, modeling and simulation of surface water flooding, time series modeling for landslides and floods, and Bayesian analysis of disaster damage. The book will

interest a wide audience, including students, young researchers, and established scientists as well as emergency response and civil defense personnel, policymakers, and other stakeholders.

London, UK Dan Crisan
October 2018

Foreword by Carlos Nobre

Earth System Science is a relatively new field of science. It encompasses the complex interactions of natural and social systems with a view of unicity and integrality of the Planet and all living species, including Homo sapiens. It emerged more distinctly in the 1990s. Although one might say that the naturalists of the eighteenth and nineteenth centuries really saw the Earth and its biotic and abiotic environment as an indivisible whole, modern science of the twentieth century sliced knowledge on a number of building blocks, the scientific disciplines, which grew continuously apart. The 1987 UN's Brundtland Report on Sustainable Development created the political environment for seeking seeing the Earth and all its inhabitants and a whole, that is, the full integration of natural and social systems, given continuously more clout to inter- and transdisciplinarity in science.

One critical area of Earth System Science is "natural disasters." More and more, we have been able to develop quantitative understanding of the mechanisms underlying disasters, particularly the triggering mechanisms associated to hydrometeorological and climatic hazards. Progress is slower in developing quantitative models of social systems. That is particularly relevant for developing early warning systems for natural disasters based on uncovering predictability potential by utilizing disaster-specific environmental models. For instance, that is the case of quantitatively modeling human vulnerability to disasters. This area is receiving worldwide attention, and a few "predictors" are emerging. That is the case of years of schooling of women: the higher the schooling, the smaller the number of fatalities.

This book is an original contribution to demonstrate the potential of quantitative models to improve our understanding of natural hazards and improve predictability of natural disasters. It equally highlights important scientific development of disaster and Earth System Science in Latin America.

São José dos Campos, Brazil — Carlos Nobre
October 2018

Preface

This book covers relevant, timely topics that offer a glimpse of the current state of the art of disaster prevention research, with an emphasis on challenges within Latin America.

Professors, researchers, and students from both mathematical and environmental sciences, civil defense coordinators, policymakers, and stakeholders could all benefit from this book.

This book brings carefully selected, fully peer-reviewed scientific works presented at three multidisciplinary scientific events:

1. Brazilian Colloquium on Mathematics, Thematic Session "Natural disasters: the presence of mathematics—from understanding to prevention" that took place from July 26 to 31, 2015, at the Brazilian Institute of Pure and Applied Mathematics (IMPA), Rio de Janeiro, Brazil
2. "International Workshop on Mathematics of Climate Change and Natural Disasters" that took place from August 29 to September 2, 2017, at the Brazilian Institute of Space Research, São José dos Campos, Brazil
3. Brazilian National Congress of Applied and Computational Mathematics, Thematic Session "Mathematics for Disaster Risk Reduction", that took place from September 19 to 23, 2017, at the Federal University of São Paulo, organized by the Brazilian Society of Applied and Computational Mathematics (SBMAC)

The 12 book chapters that comprise this volume were selected among the works submitted for presentation in those events and some invited authors. Each chapter is being reviewed by two referees at two rounds.

All these selected chapters highlight the increasing importance of both physical and empirical modeling approaches. In this book, the tools from mathematics, statistics, and computing are presented in a cross-disciplinary way to give the researchers and policymakers a better understanding of how natural phenomena occur and what could be done to prepare communities and societies for the impacts of disasters.

It is worth observing that the thematic of this book is motivated by the actions behind the Biennium of Mathematics in Brazil (2017–2018), which focus on incentivizing the study and popularization of Mathematics in Brazil. To face this

context, dealing with issues related to environmental disaster risks in a mathematical point of view was one of the main characteristics of the contributions in this work.

São José dos Campos, Brazil — Leonardo Bacelar Lima Santos
São José dos Campos, Brazil — Rogério Galante Negri
Campinas, Brazil — Tiago José de Carvalho
November 2018

Acknowledgments

This book was possible due to dedication of more than 40 authors (organized in 12 chapters) and almost 30 reviewers from different countries, especially Argentina, Brazil, Colombia, Mexico, and Uruguay. We do thank you all for this hard and great work.

It is fundamental to highlight that the Brazilian Society of Applied and Computational Mathematics (SBMAC) develops a very important work for the Brazilian scientific community and keeps an enforcing relationship with other Mathematics societies in Latin America.

We would like to thank all the Springer team, especially the editor for Mathematics in Brazil, Robinson Nelson dos Santos, for all the assistance.

Finally, we register our institutional acknowledgments to the Brazilian National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), the Federal Institute of São Paulo (IFSP), and the São Paulo State University "Júlio de Mesquita Filho" (UNESP).

Introduction

We live in uncertain times. The global climate is undergoing a scenario of change and extreme weather, and climatic events are expected to increase in frequency and intensity. Precipitation patterns are changing; temperatures are rising. The need to understand the causes and mitigate against the potential effects is urgent.

Global environmental changes have caused an unprecedented rise in extreme climate events in Latin America where extreme droughts are perhaps the most costly natural extreme faced. Since 2005 until 2015, this region has experienced a huge number of record-breaking droughts unparalleled in severity with an estimated cost in excess of $13 billion (USD) from agricultural and livestock losses, according to the Food and Agriculture Organization of the United Nations. Further catastrophic effects of this include annual wildfires which become much more severe, and an effect of this is a huge increase in carbon emissions, further fuelling the cycle. Infrastructure is also affected as a huge percentage of electricity generation comes from hydroelectric sources; hence, the extreme droughts can cause power shortages. Landslides on populated slopes are particularly common in Latin America and have devastating consequences. In 2010 landslides and floods near Rio de Janerio, which resulted in over 500 deaths, 14,000 people were left homeless and had a loss of income of $408 million (USD). Significant research is required to improve predictive capabilities and provide advanced warnings which allow for mitigation through evacuation, saving lives and substantial capital through the reduction of material damage to infrastructure and helping safeguard the interests and development of the most vulnerable members of society.

This volume "*Towards Mathematics, Computers and Environment: A Disasters Perspective*" covers relevant timely topics that offer a glimpse of the current state of art of disaster prevention research, with an emphasis on challenges within Latin America. This text is aimed at interdisciplinary researchers from Mathematical and Environmental Sciences, civil defense coordinators, policymakers, and stakeholders. This book brings carefully selected fully peer-reviewed scientific works. The chapters highlight the increasing importance of both physical and data-driven modeling approaches.

Predictability and forecasts are of the utmost importance for climate models, and the current state of the art applies various well-studied techniques for this task. The systems of partial differential equations (PDEs) are posed that account for atmospheric movement and other advected species which can include water, trace chemicals, and aerosols. A numerical discretization of the underlying PDE model is itself formed of various components, advection schemes, linear algebra packages, spectral transformations, etc.

It is especially important to develop a rigorous understanding of these extremely complicated processes, not least to allow for reproducibility of results. Because of the huge undertaking that goes into building these models, many different choices are made in terms of a numerical model, various parameterizations, and different "physics" that enter into the models. In this light, it has been observed that different comprehensive climate models give vastly different answers when solving the same problem with the same forcing. Some of the nuances that must be studied are included in the first chapters of this book, dealing with ocean-atmosphere coupling, Rossby waves, and conservation of physical quantities and multiscale methods for advected quantities.

Given the huge variances in length and time scales and, in many cases, a fundamental lack of understanding in the physical causes of these events one turns to data-driven approaches. With the surge in activity in machine learning and the huge advances that have been made in data acquisition, these modeling approaches are becoming very successful. It is even possible to combine the two philosophies, which seem at a first glance to be so incompatible, through smart data assimilation strategies, making use of machine learning algorithms allowing us to reach toward the holy grail of disaster monitoring, a complete physics driven, data-informed disaster prediction model.

Reading, UK Tristan Pryer

Contents

Contributors

Paulo Henrique Barchi CAP-INPE, São José dos Campos, SP, Brazil

Guillermo Berri Facultad de Ciencias Astronómicas y Geofísicas – UNLP, La Plata, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) Buenos Aires, Argentina

Diego N. Brandão Centro Federal de Educação Tecnologica Celso Suckow da Fonseca, Rio de Janeiro, Brazil

Julián Bravo-Castillero Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Alan James Peixoto Calheiros National Institute for Space Research (INPE), São José dos Campos, Brazil

Rosio Camayo Center for Weather Forecasting and Climate Studies, National Institute for Space Research, São Paulo, Brazil

Tiago José de Carvalho Department of Informatics, Federal Institute of São Paulo (IFSP), Campinas, São Paulo, Brazil

Carolina Locatelli Colla Federal University of Sao Paulo, Institute of Science and Technology, São José dos Campos, SP, Brazil

Alessandra C. Corsi Centro de Tecnologias Geoambientais - CTGeo, Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, São Paulo, Brazil

Camila P. da Costa Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, RS, Brazil

Álvaro de Oliveira D'Antona School of Applied Sciences, UNICAMP, Limeira, Brazil

Glauston Roberto Teixeira de Lima National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos, Brazil

Eduardo S. de Macedo Centro de Tecnologias Geoambientais - CTGeo, Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, São Paulo, Brazil

Pier Francesco De Maria School of Applied Sciences, UNICAMP, Limeira, Brazil

Leonardo Tomazeli Duarte School of Applied Sciences, UNICAMP, Limeira, Brazil

Filipe A. M. Falcetta Centro de Tecnologias Geoambientais - CTGeo, Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, São Paulo, Brazil

Nelson Ferreira Fernandes Federal University of Rio de Janeiro, Department of Geography, Rio de Janeiro, RJ, Brazil

João Victor Cal Garcia National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos, SP, Brazil

Letícia de Carvalho Giannella Brazilian Institute of Geography and Statistics, National School of Statistical Sciences, Rio de Janeiro, RJ, Brazil

André Yoshizumi Gomes Serasa Experian, Decision Analytics, São Paulo, SP, Brazil

Marilia de Abreu Gregorio Centro de Estudios de Variabilidad y Cambio Climático (CEVARCAM) – FICH – UNL, Santa Fe, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

Marcela P. P. Guimarães Centro de Tecnologias Geoambientais - CTGeo, Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, São Paulo, Brazil

Lidiane dos Santos Lima Protection and Civil Defense Municipal Office, Maricá, RJ, Brazil

Luciana R. Londe Cemaden, São José dos Campos, SP, Brazil

Jader Lugon Jr. Instituto Federal Fluminense, Macaé, Brazil

Camila Bertini Martins Federal University of São Paulo, Paulista School of Medicine, Department of Preventive Medicine, São Paulo, SP, Brazil

Daniel S. Menasché UFRJ, Rio de Janeiro, RJ, Brazil

Gabriela V. Müller Centro de Estudios de Variabilidad y Cambio Climático (CEVARCAM) – FICH – UNL, Santa Fe, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

Viviana Aguilar Muñoz National Centre for Monitoring and Early Warnings of Natural Disasters (CEMADEN), São José dos Campos, SP, Brazil

Joshi Neelakshi CAP-INPE, São José dos Campos, SP, Brazil

Leslie D. Pérez-Fernández Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, RS, Brazil

Alex Sandro Aguiar Pessoa Climatempo Meteorologia, São José dos Campos, Brazil

Gabriel Augusto L. L. Pinheiro CAP-INPE São José dos Campos, SP, Brazil

Tristan Pryer Department of Mathematics and Statistics, University of Reading, Reading, UK

Enver Ramírez Center for Weather Forecasting and Climate Studies, National Institute for Space Research, São Paulo, Brazil

Pedro P. G. W. Rodrigues Instituto Politécnico, Universidade do Estado do Rio de Janeiro UERJ, Nova Friburgo, Brazil

Reinaldo Roberto Rosa Lab for Computing and Applied Mathematics-INPE, São José dos Campos, SP, Brazil

Paulo Cesar Colonna Rosman Federal University of Rio de Janeiro, Ocean Engineering Program, Rio de Janeiro, RJ, Brazil

Leonardo Bacelar Lima Santos National Centre for Monitoring and Early Warnings of Natural Disasters (CEMADEN), São José dos Campos, São Paulo, Brazil

Ricardo Manhães Savii Federal University of São Paulo, Institute of Science and Technology, São José dos Campos, SP, Brazil

Elcio Hideiti Shiguemori Geo-intelligence Division, Institute of Advanced Studies (IEAV), Rov. Tamoios, São José dos Campos, SP, Brazil

Antônio J. Silva Neto Instituto Politécnico, Universidade do Estado do Rio de Janeiro UERJ, Nova Friburgo, Brazil

Léo Siqueira University of Miami, Rosenstiel School for Marine and Atmospheric Science, Miami, FL, USA

Stephan Stephany National Institute for Space Research (INPE), São José dos Campos, Brazil

Julia Celia Mercedes Strauch Brazilian Institute of Geography and Statistics, National School of Statistical Sciences, Rio de Janeiro, RJ, Brazil

Cesar Strauss National Institute for Space Research (INPE), São José dos Campos, Brazil

Wagner R. Telles Universidade Federal Fluminense, Santo Antônio de Pádua, Rio de Janeiro, Brazil

Cristiano Torezzan School of Applied Sciences, UNICAMP, Limeira, Brazil

Didier A. Vega-Oliveros DCM-FFCLRP-USP, Ribeirão Preto, SP, Brazil

An Overview of the El Niño, La Niña, and the Southern Oscillation Phenomena: Theory, Observations, and Modeling Links

Léo Siqueira, Enver Ramírez, and Rosio Camayo

1 Overview

The term El Niño, La Niña, and the Southern Oscillation, collectively ENSO, refers to the large-scale ocean–atmosphere phenomenon linked to an occasional warming (cooling) during El Niño (La Niña) in sea-surface temperatures (SST) across the central and east-central equatorial Pacific (Fig. 1). It directly affects the equatorial central and eastern Pacific region with signatures in and over the western Pacific and Indian Ocean [31]. However, impacts of ENSO are experienced on a global scale [4, 45] through remote connections (teleconnections) with prominent implications over diverse human activities, like energy generation, agriculture, infrastructure, public health, transport, among others [31, 33, 35]. A typical pattern for El Niño (La Niña) consists in an abnormal warming (cooling) of the SST along the coast of Peru which is extended westward and forms an anomalous warm (cold) tongue shown in Fig. 1.

During an El Niño (La Niña) episode, changes in the oceanic heat content of upper layers is also observed with a shallowing (deepening) to the west and a deepening (shallowing) to the east. Regarding its temporal characteristics, ENSO is a naturally occurring phenomenon and displays irregular interannual variations (Fig. 3), with peak power between 3 and 7 years in SST, as highlighted in Fig. 2. However, its variability can also be found in other variables, e.g., sea level height, convective rainfall, surface air pressure, and atmospheric circulation. In the atmosphere, El Niño (La Niña) episodes are associated with the anomalous

L. Siqueira
University of Miami, Rosenstiel School for Marine and Atmospheric Science, Miami, FL, USA
e-mail: lsiqueira@miami.edu

E. Ramírez (✉) · R. Camayo
Center for Weather Forecasting and Climate Studies, National Institute for Space Research, São Paulo, Brazil
e-mail: enver.ramirez@inpe.br; rosio.camayo@inpe.br

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_1

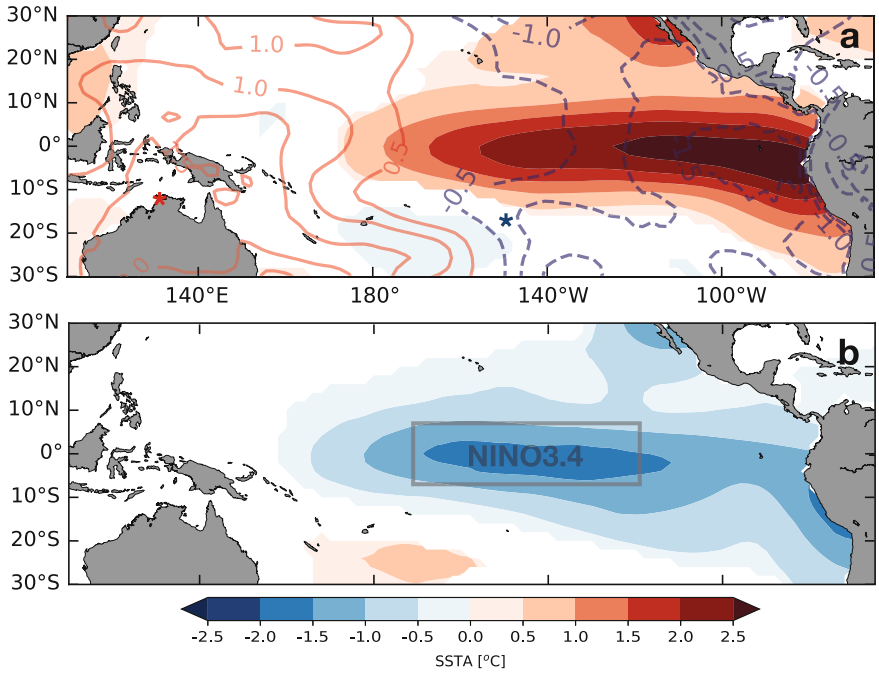


Fig. 1 Composite tropical Pacific SST anomalies (°C) displaying the typical patterns associated with the anomalous peak warming for strong (≥1.5) El Niño (**a**) and La Niña (**b**); Colored areas denote regions significant at the 95% level and contours show pressure departures in millibars. Colored stars show the approximate location of Darwin (red) and Tahiti (blue) from which pressure differences are used to calculate the Southern Oscillation Index in Fig. 3. Box in (**b**) shows the region (NINO3.4) where the largest SST variability occurs on ENSO timescales, typically used for ENSO monitoring. Asymmetry between the location of the maximum anomaly with El Niño peaking farther east and La Niña peaking more toward central Pacific is also noted

weakening (strengthening) of the climatological large-scale pressure gradients over the tropical region, typically measured as the normalized observed sea level pressure differences between Tahiti (eastern tropical Pacific) and Darwin, Australia (Fig. 1a). In general, the state of the Southern Oscillation is represented by a smoothed time series of the Southern Oscillation Index (SOI) which corresponds very well with changes in SST across the east-central tropical Pacific (NINO3.4 index). The negative phase of the SOI depicts below-normal air pressure at Tahiti and above-normal air pressure at Darwin (Fig. 1a). Extended periods of negative (positive) SOI values coincide with anomalous warm (cold) ocean waters across the eastern tropical Pacific indicative of El Niño (La Niña) episodes as shown in Figs. 1a and 3. The weakening (strengthening) of the large-scale pressure gradients enables a zonal shift of the strong atmospheric convection, which is frequently located over the western Pacific and maritime continent, to the east (west) of this region during El Niño (La Niña). This in turn affects the position of the heating sources and the

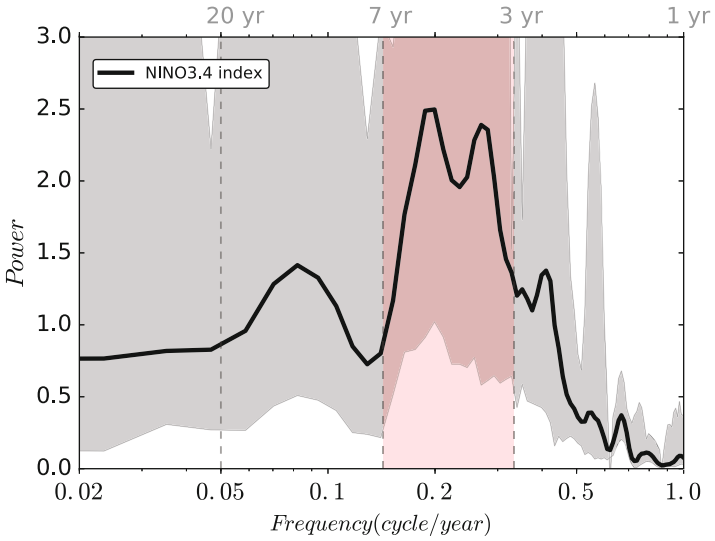


Fig. 2 Multitaper power spectral density of the observed NINO3.4 index, area averaged SST over [170°W:120°W;7°S:7°N]. Shaded gray area shows an estimate of the confidence interval of the spectrum generated using a jack-knifing procedure of [46]

energy partition between the tropical internal modes and the teleconnections related barotropic (external) modes [9, 36].

Concerning the irregularity of ENSO (Fig. 3), both east-central equatorial Pacific SST (NINO3.4) and SOI undergo oscillatory behavior as the system transits between above-normal sea level pressure differences (La Niña), below-normal sea level pressure differences (El Niño), and no significant pressure differences anomalies (neutral years). Major El Niño events recorded since the beginning of the last century are 1904/1905, 1913/1915, 1925/1926, 1940/1941, 1957/1958, 1972/1973, 1982/1983, 1986/1988, 1994/1995, 1997/1998 (strongest in last century), 2002/2003, 2004/2005, 2006/2007, 2009/2010, 2014/2016. Since the beginning of the observational record, ENSO frequency has changed from around 12 years to more frequently occurrences during the last part of the record. However, it must be acknowledged that the record length is far too small to be conclusive. In addition, extended El Niño episodes like 1913/1915, 1986/1988, 1991/1995, and 2014/2016 are cases that deserve more attention from the scientific community as well as the decadal like occurrences, types of variability that are also manifested in the ENSO spectrum (Fig. 2).

Although ENSO is a prominent variability that dominates the interannual timescale, it is able to interact with relatively higher [13] and lower [41] frequency variability. This in turn leads to local and regional changes that span from clouds and precipitation records to tropical–extratropical teleconnections of planetary scale. ENSO atmospheric teleconnections typically involve both the so-called tropical and tropical–extratropical "atmospheric bridges." The former occurs via modulation of

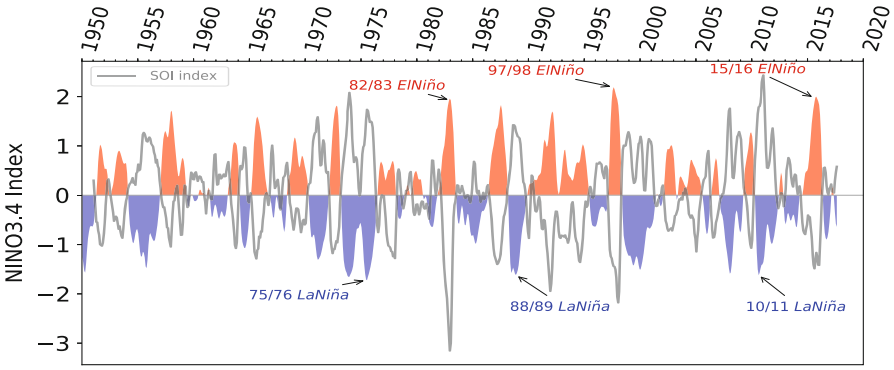


Fig. 3 Southern Oscillation Index (SOI; solid gray line), a measure of the atmospheric seesaw, derived from the normalized pressure difference between Tahiti (eastern Pacific) and Darwin (Australia) depicted by the blue and red stars in Fig. 1a, respectively. NINO3.4 index—area averaged SST over [170°W:120°W;7°S:7°N]—depicted by gray box in Fig. 1b. Positive (red) values show El Niño events, i.e., anomalous warming, while negative (blue) values show La Niña events, i.e., anomalous cooling. Strongest events recorded through both SOI and NINO3.4 index are indicated by arrows

the Walker circulation, with a time lag ranging from 3 to 6 months, producing changes in atmospheric subsidence, cloud cover, and evaporation over remote ocean basins such as the tropical North Atlantic [44], Indian Ocean [51], and South China Sea [50]. The tropical–extratropical bridge, on the other hand, is associated with ENSO-related upper-level atmospheric vorticity anomalies forcing large-scale atmospheric Rossby waves that propagate into the extratropics. The changes in deep convection due to ENSO lead to changes in atmospheric heating, low-level convergence, and upper-level divergence in the equatorial Pacific that are very effective at exciting the "Pacific-North American" pattern (PNA; [14]) and "Pacific-South American" pattern (PSA; [19]) in the Northern and Southern Hemispheres, respectively. These waves propagate into the extratropics establishing teleconnection patterns within 2–6 weeks and follow great circle routes that initially extend poleward and eastward, being eventually refracted away from the poles and return to the tropics [15, 48].

These teleconnections have a broad impact on the intensity and occurrence of extreme events over other parts of the globe. For example, temperature extremes are marked by an increase during El Niño over India, Southeast Asia, Australia, and Southern Africa, while temperatures become cooler over southeastern North America, and vice versa during La Niña [48]. Despite the fact that the remote influence of ENSO on precipitation is much less coherent than in the case of temperatures, outstanding heavy rains occurred in California, Ecuador, and Northern Peru during the strong 1982/1983 and 1997/1998 El Niño events. Furthermore, ENSO has been linked to the modulation of the Antarctic dipole pattern producing changes in sea ice concentration around West Antarctica through Rossby wave trains emanating from the tropical Pacific [32]. Although the above examples illustrate

some of the extent of ENSO impacts, these mostly are not stationary and of low predictability, except perhaps over Indonesia, eastern Australia, and Eastern Africa. The recent 2015/2016 El Niño was one of the strongest El Niño in history (Fig. 3); however, the precipitation anomalies considerably differed from the expectation associated with previous strong El Niño events not bringing heavy rains to Peru or Ecuador, and actually being opposite in sign compared to typical El Niño conditions in California [23, 37]. On the other hand, the unusual extent and duration of the melting in West Antarctica was likely favored by the strong 2014/2016 El Niño unlike the strong 1982/1983 and 1997/1998 episodes.

The theory for ENSO has been widely explored along the past 40 years by a large body of observational and theoretical work, with modeling attempts approaching a mature stage during the late 1980s, to the point where predictions were being made on a regular basis in the 1990s. Despite improvements in observations and coupled ocean–atmosphere models in recent decades, ENSO predictability and longer lead-time forecast skill remains limited. There have been different views and theories formulated about ENSO and some were eventually discarded, but despite that there is still lack of consensus between modelers and observers about what are the essential mechanisms for ENSO. Recent studies also suggest that there exists El Niño (La Niña) diversity regarding the parameters related to its amplitude, trigger mechanisms, spatial patterns, and life cycle [7] as well as its impacts on the globe. More recently, flavors of El Niño/La Niña characterizing two modes through which the ENSO is manifested have been obtained [7, 45], namely the Canonical El Niño/La Niña (central and eastern Pacific anomaly) and the El Niño/La Niña Modoki (central Pacific anomaly). A new, but old flavor, El Niño Costero (far eastern Pacific anomaly) with scientific registers back to the nineteenth century is being revisited and also claiming for its place into this classification [43]. The extended El Niño that happened earlier in the twenty-first century might suggest that our indices to measure when an extreme phase of the ENSO cycle is occurring appear to need reexamination. Perhaps, aware of this problem it has been suggested that a practical definition of El Niño is, therefore, more a matter of convenience for its users than strictly a scientific result [47].

2 Simplified but Realistic ENSO Models

A wide range of models of different complexity have been used in ENSO research and prediction. They range from very simple models (known as “toy models”) that attempt to recreate the very basic physics of the phenomena to sophisticated coupled ocean–atmosphere general circulation models (GCMs) that integrate our knowledge about global scale ocean and atmosphere dynamics. Statistical models based only on data have also shown that ENSO dynamics can be inferred to a surprisingly good degree using linear inverse models (LIMs).

From a historical perspective, a set of intermediate complexity models (ICMs) that combine simple physical models of the ocean and atmosphere into a coupled model for the equatorial Pacific were the first to capture the essential features of ENSO [1, 3, 5, 18, 53].

The paradigm of the ICMs consists of a shallow-water $1\frac{1}{2}$ layer reduced-gravity model of mean depth $H_{1.5}$, where the shallow (mixed) layer of relatively warm (and less dense) water overlies a much deeper layer of cold water in an ocean basin of length L. The two layers are separated by a sharp vertical temperature gradient (thermocline), and it is assumed that there is no motion in the deep layer. The idea is to approximate the thermal (and density) structure of the ocean in the simplest form possible. This simple ocean model is coupled to a simplified Gill-type [11] steady-state linear atmospheric model. In this type of model, the atmosphere is assumed to be in instantaneous equilibrium with the oceanic evolution since the timescale for atmospheric adjustment is fast (a few days) compared with the ocean timescales (months) and the ENSO cycle. One of the key distinctions between ICMs models and more elaborate models such as GCMs is that the former do not attempt to model the background flows of the system. Rather, perturbation equations are constructed such that all quantities computed are anomalies, and if necessary, observed background climatologies are used for the mean flows.

Dimensional analysis based on scales of variability of the system (e.g., zonal length scaled by the basin width L, meridional width scaled by the oceanic radius of deformation, neglecting meridional damping and meridional wind stress) indicates that the following "long wave" approximation (i.e., no $\partial_t v = 0$ term) can be used, which in dimensionless variables becomes

$$(\delta\partial_t + \epsilon_m)u_m - yv_m + \partial_x h = \tau^x, \tag{1a}$$

$$yu_m + \partial_y h = 0, \tag{1b}$$

$$(\delta\partial_t + \epsilon_m)h + \partial_x u_m + \partial_y v_m = 0, \tag{1c}$$

$$\epsilon_s u_s - yv_s = \delta_s \tau^x, \tag{2a}$$

$$\epsilon_s v_s + yu_s = 0, \tag{2b}$$

where u_m and v_m represent horizontal velocities (vertical-mean currents), h is the thickness perturbation, and ϵ_m is a fairly long oceanic damping (combining Rayleigh friction and thermal damping) coefficient. The vertical-mean component is governed by linearized shallow-water dynamics (1). Equation (2) represents the frictional balance of the Ekman layer and (u_s, v_s) are the contribution of vertical-shear currents associated with the fixed-depth surface layer. Furthermore, τ^x is the zonal wind stress (meridional wind stress is neglected) and the parameter δ measures the ratio of the timescale of oceanic adjustment (in the zonal direction) by wave dynamics to the timescale of SST change by coupled feedback and damping processes.

In these Eq. (1), the fast-wave (or fast-dynamics) limit is obtained as $\delta \to 0$, for which the wave dynamics timescales are fast compared to the coupled timescales

affecting SST, and the fast-SST limit as $\delta \to \infty$. The modes that result from the fast-wave limit are not very realistic, whereas those of the fast-SST limit, for which the adjustment of surface temperatures is much faster than the dynamic adjustment, leads to more realistic unstable and damped coupled modes. The latter is therefore the basis of the delayed oscillator model discussed in Sect. 3. In fact, in order to somehow explain the irregularity of ENSO some previous studies [28, 29] suggest that ENSO is actually a mixed mode instability located within both limits in parameter space, containing features of the fast-SST and fast-wave limits with continuous transition between the two.

In this class of model, the Ekman layer is shallower than the layer represented by (1) and the difference in flow between the mixed layer and the lower layer is controlled by a steady state Eq. (2) representing a heavily damped shear (Ekman) flow. In Eq. (2), δ_s is the surface layer coefficient, i.e., the ratio of the timescale of SST change by vertical-mean current and thermocline feedbacks to the timescale of SST change by coupled current perturbations associated with the active surface layer. This parameter governs the strength of feedbacks due to vertical-shear currents and upwelling created by viscous transfer between the surface layer and the rest of the thermocline.

In addition, a fully nonlinear SST Eq. (3) is also included, since SST is a key interfacial variable, in which temperature variations are determined by vertical mixing, horizontal advection, and surface heat fluxes

$$\partial_t T + u\partial_x T + v\partial_y T + \mathcal{H}(w)w(T - T_{\text{sub}})/H_{1.5} + \epsilon_T(T - T_0) = 0, \tag{3}$$

where u and v are horizontal velocities in the mixed layer, w the vertical velocity just below the mixed layer, T is the temperature of the equatorial surface, T_{sub} is the below thermocline ocean temperature approximated as a tanh function, T_0 is the radiation equilibrium temperature toward which the model tends in the absence of motion, ϵ_T is a constant representing Newtonian cooling by surface fluxes, and $\mathcal{H}(w) = 1/2(\tanh(w/\Delta w) + 1)$ is a continuous approximation to the Heaviside function used to switch between upwelling and downwelling. It is worth noting that SST variations, in this formulation (3), affect neither the pressure gradients nor the dynamics of the system.

From the atmospheric model (Gill's model) the following non-local relation between the zonal wind stress and equatorial SST can be derived [17, 18]:

$$\tau^x = \mu\mathcal{A}(T; x, y). \tag{4}$$

In Eq. (4), $\mathcal{A}(T; x, y)$ is a non-local function that relates the equatorial SST (T) to the zonal wind stress τ^x, and μ is a measure of the coupling strength coefficient. A formal derivation of (1)–(4) can be obtained from basic principles, but this would fall beyond the scope of this overview. It is worth noting that in this model, the nonlinearity enters through the thermodynamic Eq. (3) for the ocean. Furthermore, this coupled model does not contain high frequency internal variability in the atmosphere, in contrast with the real coupled ocean–atmosphere system, since it

considers a steady-state Gill-type atmosphere. Therefore, intraseasonal variability is explicitly neglected, although the response of the Gill-type atmosphere is an atmospheric Kelvin and long Rossby waves.

Equations (1)–(4) are, essentially, a stripped-down version of the [53] intermediate model (ZC). The major breakthrough in ENSO research introduced by this anomaly model was to produce sustained variability that is reasonably similar to ENSO. In this model, the recurrent warm and cold events of irregularly varying amplitude show realistic spatial structure and seasonal cycle as well as a period of about 3–4 years.

The ZC model was designed specifically to capture the interannual variability of large spatial scales of motion in the ocean. In this anomaly model, the major physical processes at work include a positive feedback between the large-scale ocean (downwelling) and atmosphere (convergence and latent heat release) leading to an ocean heating anomaly and further downwelling and deepening of the thermocline. The disequilibrium is carried by the ocean wave dynamics as eastward propagating equatorial Kelvin waves and westward propagating slightly off-equatorial Rossby waves. In the central and eastern parts of the basin the Kelvin waves grow unstable if they have large wavelengths and are slow enough to stay within the coupling region so that the coupling process feedbacks on the perturbation. The perturbations of the thermocline also produce negative off-equator depth anomalies (that is, a shallowing signal of the thermocline) in the central part of the basin that excite westward propagating Rossby waves. These Rossby waves eventually are reflected off the western boundary (no-net flow $\int udy = 0$ required) as large-scale equatorial Kelvin waves (long wave approximation) that move back to the east and eventually arrive to the eastern part of the basin months later, shallow the thermocline there, and cause cooling of the SST. This in turn implies in a break of the warm conditions as a result of the positive feedback of the thermocline deepening and the SST. Then the coupled feedback starts to operate with a different sign to amplify the negative temperature anomaly leading to a La Niña state. The equilibration of temperature anomalies is usually modeled through nonlinear effects, mainly in the model's SST Eq. (3), which limits the amplitude of the resulting oscillation. The ZC model then shows the recurrence of warm and cold events, deriving solely from self-interactions of the coupled system since there is no external forcing. The period of the oscillation is basically determined by a systematic time delay between dynamical changes in the eastern part of the basin and associated large-scale fluctuations in the equatorial wind stress. It is worth noting that by making a long wave approximation in the ocean dynamics, typically one ensures the dominance of Kelvin and long Rossby waves; however, other modes can be asymptotically found [34, 39]. By ensuring the dominance of Kelvin and long Rossby waves, one is assuming that all the energy associated with reflected Rossby waves goes into the largest Kelvin modes. Therefore, reflection becomes too efficient as it happens in the ZC model and is further enhanced by the use of unrealistic solid boundaries in this model.

3 Low Order Models: Delay Equations, Discharge–Recharge Theory, and Unification by the Two-Strip Model

After the ZC model and other concurrent efforts showed success in simulating ENSO, the understanding of equatorial wave dynamics was further refined providing conceptual models that account for important aspects of the ICMs results [2, 3, 17, 42]. Jin [17] formulated the two-strip model, a reduced model for ENSO capable of capturing the essential dynamics, based on the shallow-water response. For a matter of convenience, we start with the linearized shallow-water equations including dissipation and using the long wave approximation (i.e., $\partial_t v = 0$). Furthermore, we only consider the zonal wind stress $\mathbf{F} = \tau^x$ since it is the anomaly of this parameter that starts the destabilization process

$$\partial_t u - yv + \partial_x h + r_o u = \mathbf{F}, \tag{5a}$$

$$yu + \partial_y h = 0, \tag{5b}$$

$$\partial_t h + \partial_x u + \partial_y v + r_o h = 0. \tag{5c}$$

Then a single equation for the thermocline h is attained by taking the y derivative of (5a), multiplying the result by y, and eventually subtracting (5a) from the resulting equation

$$(yu_y - u)_t - y^2 v_y + y h_{xy} + r_o y u_y - h_x - r_o u = y\mathbf{F}_y - \mathbf{F}. \tag{6}$$

Using (5b) and (5c) we can fully re-write (6) as a function of h

$$y^2(\partial_t h + r_o h) + \left(\frac{2}{y}\partial_y - \partial_y^2\right)(\partial_t h + r_o h) - \partial_x h = y\partial_y \mathbf{F} - \mathbf{F}, \tag{7}$$

with boundary conditions

$$x = 0 : \int_{-\infty}^{\infty} \frac{1}{y}\frac{\partial h}{\partial y} dy = 0, \tag{8a}$$

$$x = 1 : \frac{\partial h}{\partial y} = 0. \tag{8b}$$

By looking at the thermocline structure associated with equatorial Rossby waves [17], it is reasonable to assume that h has a parabolic structure near to the equator $h(x, y, t) = h_e(x, t) + y^2 \Delta h(x, t)$. Taking $h(x, 1, t) = (h_n + h_e)/2$, where h_n is the departure of the thermocline at the northern domain, and thinking in terms of Taylor expansion for y near to the equator, we have $\partial_y h \approx y\Delta h(x, t) = y(h_n - h_e)/2$. Furthermore, using (5b) we obtain $u = -(h_n - h_e)$. Finally, substituting h in (7) and evaluating it at $y = 0$ and $y = y_n$ gives the two-strip model

$$(\partial_t + r_o)(h_e - h_n) + \partial_x h_e = \mathbf{F}\Big|_{y=0}, \tag{9a}$$

$$(\partial_t + r_o)(h_n) - \frac{1}{y_n^2}\partial_x h_n = \partial_y \left(\frac{\mathbf{F}}{y}\right)\Big|_{y=y_n}. \tag{9b}$$

At $y = 0$ the free wave solution for the zonally unbounded domain is a Kelvin wave with wavenumber k. Similarly, at $y = y_n$ the free wave solution for h_n is a Rossby wave with wavenumber $-k/h_n^2$. Thus, [17] obtained an elegant simplified model where the two modes of the ENSO are found. However, it is not only the wave dispersion that enables the system destabilization, but the wave reflection at the longitudinal borders. This condition is represented through

$$h_n(1, t) = r_E h_e(1, t), \tag{10a}$$

$$h_e(0, t) = r_W h_n(0, t). \tag{10b}$$

The basin width was normalized to 1, so that $h_n(1, t)$, $h_e(1, t)$ are evaluated at the eastern boundary and $h_n(0, t)$, $h_e(0, t)$ at the western boundary; r_E, r_W are a measure of the allowed mass flux in each border and accounts for the reflection of the waves. In fact, at the eastern boundary, the zonal velocity is given by $u_E = h_e(1, t) - h_n(1, t) = (1 - r_E)h_e(1, t)$, where $r_E \in [0, 1]$. In the extreme case $r_E = 1$ it results that $u_E = 0$. The western boundary, however, is not fully reflective so that energy leaks due to the open local topography, consequently $r_W < 1$ and the conservative choice of $r_W = 3/5$ is an appropriated value under the two-strip approximation with $h = 0$ for $y \geq 2y_n$. Thus, r_E and r_W monitor the mass exchange between the tropical Pacific and the adjacent regions.

The dispersion relation of the oceanic adjustment modes (eigenvectors of the forced problem) is

$$\sigma_j = -r_0 + \ln \frac{r_E r_W (1 + y_n^2) - r_E}{(1 + y_n^2) - r_E} + \mathrm{i}\frac{2\pi j}{1 + y_n^2}, \tag{11a}$$

$$h_n = H_n \mathrm{e}^{(\sigma_j t + (\sigma_j + r_0) x y_n^2)}, \tag{11b}$$

$$h_e = H_n \mathrm{e}^{(\sigma_j t - (\sigma_j + r_0) x)} + \frac{h_n}{1 + y_n^2}, \tag{11c}$$

where $j = \{\cdots, -2, -1, 0, +1, +2, \cdots\}$. The logarithmic term in (11a) is always negative, representing the damping associated with leakage to the west of the basin. For $j = 0$ the stationary mode of the ocean adjustment is obtained and the mean value of h over the whole basin is a solution of (9) for $r_0 = 0$. This mode can be thought as an approximation of the main dispersive mode of the full spectrum of the oceanic adjustment. For higher j's, the oscillatory modes are equivalent to the ocean basin modes.

Within this context, the long wave approximation forced by the wind stress, i.e., a model appropriate for the oceanic component, can be conveniently transformed to the two-strip model. Rossby and Kelvin waves are contained into (9) and both modes are milestone for the ENSO theory. Inclusion of the boundary conditions allows to obtain a higher degree of realism. In addition, the slow thermodynamic adjustment governing the equatorial temperature perturbations can be described by

$$\partial_t T + \vec{v} \cdot \nabla T - \alpha h + u_a^3/h = 0. \tag{12}$$

Neglecting the effect of the wind ($u_a^3/h = 0$) and using the advective part as a local damping $C_T(x)$, the temperature can be written as

$$\partial_t T_e + C_T(x) T_e - \alpha(x) T_e = 0, \tag{13}$$

where $\alpha(x)$ represents the effect of the thermocline variations (through background upwelling) on the temperature (thermocline feedback). Additional simplifications are employed in which the area average SST over the region with the highest variation during the ENSO development is considered and following the Gill atmosphere model [11], $\mathbf{F} = \mu \mathcal{A}(T_e) \mathrm{e}^{-\frac{\epsilon^2 y^2}{2}}$, where

$$\mathcal{A}(T_e) = A_0 T_e f(x), \tag{14}$$

with $f(x)$ being a fixed pattern and A_0 its amplitude without damping. The two-strip model can be integrated following the characteristic curves corresponding to the Kelvin and Rossby waves are given by

$$x - x_0 = t - t_0, \tag{15a}$$

$$x - x_0 = -\frac{t - t_0}{y_n^2}, \tag{15b}$$

where (x_0, y_0) represents an arbitrary point inside the domain. Thus, by integrating from the eastern to the western border (9b) along (15b) and then integrating along the reverse path along (15a) and neglecting the damping, the following delay ordinary differential equations are obtained:

$$\begin{aligned} h_{eW}(t) = {} & r_W r_E h_{eW}(t - 1 - y_n^2) \\ & + \mu A_0 r_W (r_E T_{eE}(t - 1 - x_P) - \theta T_{eE}(t - y_n^2 x_P)), \end{aligned} \tag{16a}$$

$$\begin{aligned} h_{eE}(t) = {} & r_W r_E h_{eE}(t - 1 - y_n^2) \\ & - \mu A_0 (\theta r_W T_{eE}(t - 1 - y_n^2 x_P) - T_{eE}(t - 1 + x_P)), \end{aligned} \tag{16b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} T_{eE} = -C_{TE} T_{eE} + \alpha_E h_{eE}, \tag{16c}$$

where x_P is a fixed point over the area of the wind response, T_{eE}, C_{TE}, and α_E are mean over the eastern half of the basin (region with largest thermal amplitude during ENSO). When the effect of the eastern boundary reflection is neglected ($r_E = 0$), then (16b) and (16c) give

$$\frac{\mathrm{d}}{\mathrm{d}t}T_{eE} = -C_{TE}T_{eE} + \mu A_0 \alpha_E (T_{eE}(t-1+x_P) - \theta r_W T_{eE}(t-1-y_n^2 x_P)), \quad (17)$$

showing the influence of local damping and remote signal due to propagation of Kelvin and Rossby waves on the average eastern basin temperature T_{eE}. The delay time $1 - x_P$ is the relatively fast effect due to the Kelvin wave; thus, it can be neglected on long timescales. However, it provides a local amplification of temperature perturbations by the thermocline feedback through a forced Kelvin wave response. The delay feedback $1 + y_n^2 x_P$ is the time taken for the Rossby wave to travel from the center of the wind response near x_P to the western boundary plus the time it takes the reflected Kelvin wave to cross the basin. In arriving to the eastern part of the basin, it provides a delayed negative feedback for the temperature perturbation (since $r_W > 0$).

Based on other ad hoc approximations the delayed-action oscillator was derived independently by [42] and [3]. In these studies, delay differential equations with local feedback in a basin with closed boundaries provided a convenient, and successful, paradigm for explaining interannual ENSO variability

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t) = aT(t) - bT(t-d) - cT^3(t). \quad (18)$$

Here T is the temperature disturbance and a represents the growth rate of T in the eastern Pacific and corresponds to $\mu A_0 \alpha_E - C_{TE}$ in (17), b is a measure of the delayed $(t - d)$ negative feedback corresponding to $(1 + y_n^2 x_P)$ in (17). The third term mimics the negative feedback mechanisms (nonlinear damping) that limit the growth of perturbations by letting small anomalies to grow fast in a linear damping regime, while large anomalies are strongly damped. A dimensionless equation can be obtained by scaling time by $1/a$ and the temperature by $\sqrt{a/c}$ so that

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{t}}\tilde{T}(\tilde{t}) = -\alpha \tilde{T}(\tilde{t}-\tau) + \tilde{T}(\tilde{t}) - \tilde{T}^3(\tilde{t}), \quad (19)$$

where $\tilde{T}$ represents the eastern equatorial SST anomalies, $\alpha = b/a$, $\tau = ad$ is the dimensionless delay time, $0 < \alpha < 1$, and $\tau > 0$. The first term on the right-hand side of (19) mimics the negative feedback by ocean adjustment processes, while the second term reflects the positive feedback mechanisms from air–sea coupling. The delay timescale τ represents the time taken by forced upwelling Rossby waves to reach the western boundary, reflect as a Kelvin wave, and reach the eastern Pacific. The above delayed oscillator equation has three equilibria $\bar{T} = 0, \pm\sqrt{1-\alpha}$, the

inner solution and one warm and one cold outer solutions. Considering infinitesimal perturbations about these equilibria by setting $\tilde{T} = \bar{T} + T'$ and linearizing

$$\frac{\mathrm{d}}{\mathrm{d}t}T'(t) = T'(t)(1 - 3\bar{T}^2) - \alpha T'(t - \tau). \tag{20}$$

A normal mode analysis of (20) with $T' = \hat{T}e^{\sigma t}$, where the complex frequency $\sigma = \sigma_r + i\sigma_i$, gives a transcendental algebraic equation

$$\sigma = 1 - 3\bar{T}^2 - \alpha \mathrm{e}^{-\sigma\tau}, \tag{21}$$

which can be solved for the frequency σ as function of the two nondimensional parameters α and τ. It turns out that the inner solution is always unstable with a non-oscillatory exponential growth, while the two outer solutions (warm and cold) may become oscillatory unstable depending on the parameters (α, τ). The system exhibits bounded oscillatory dynamics [42] if an outer steady state is unstable which appear for larger values of the negative feedback α, and for larger values of the delay time τ. The typical period of the unstable modes is about 2–3 times the Rossby delay time τ. The period, which is substantially shorter than the observed ENSO period, is not a robust outcome of this model and is one of the two main conceptual difficulties in the acceptance of the delayed oscillator theory. Another key element that fostered the debate regarding the central role of ocean wave dynamics is that western boundary reflections are considerably more intricate than formulated in the delayed oscillator theory for ENSO. Regarding the first issue [6] argued that multiple Kelvin waves coming off the western boundary are required to eliminate the perturbation growth in the east, while [20, 38, 49] argue, in contrast to [6], that off-equatorial Rossby waves (higher modes) forced by off-equatorial wind anomalies play a considerable role and explain the slower observed timescale due to their slower propagation speeds. The delayed oscillator engendered many other conceptual models based on one or another type of delayed-action equation, namely the western Pacific, advection, and unified oscillators each highlighting certain physical mechanisms involved in ENSO.

Concerning the second issue, [16, 17] proposed a different approach to derive the so-called recharge/discharge oscillator for ENSO. In this view, the details of western boundary reflections are of secondary importance and the slow timescale is associated with near equatorial balance between the wind stress curl and the depth-integrated meridional transport in the ocean (Sverdrup balance). The recharge oscillator assumes that differences in thermocline depth tilt between the equatorial eastern and western Pacific are largely in a Sverdrup balance with the equatorial wind stress force

$$h_E = h_W + [\tau], \tag{22}$$

where h_W denotes the thermocline depth anomaly in the western Pacific, h_E is the thermocline depth anomaly in the equatorial eastern Pacific, and $[\tau]$ is the

zonally integrated wind stress across the basin along the equator. In this view, the equatorial wave propagation process is relatively fast for establishing this thermocline slope that extends to the off-equatorial region as a result of the broadness of the atmospheric wind system. The Coriolis force becomes important off the equatorial band, and therefore there will be Sverdrup transport, which causes mass or, equivalently, heat content to converge toward the equator depending on wind forcing. Despite that the thermocline tilt along the equator is set up quickly to balance the equatorial wind stress as expressed by (22), the thermocline depth in the west takes time to adjust to the zonal integrated meridional transport, which is related to both the wind stress and its curl off the equatorial band

$$\frac{\mathrm{d}}{\mathrm{d}t}h_W = -rh_W - \alpha[\tau], \tag{23}$$

where the first term of the right-hand side represents ocean adjustment processes that are assumed to act at a constant rate r and collectively represents the damping of the upper ocean system through mixing, and the equatorial energy loss to the boundary layer currents at the east and west sides of the ocean basin. The second term represents the Sverdrup transport across the basin and the minus sign in front of α comes from the fact that a large positive wind stress is expected to depress the western thermocline depth anomaly [52]. As mentioned before, the variation of SST during ENSO is largely confined within the central to eastern equatorial Pacific and is described by

$$\frac{\mathrm{d}}{\mathrm{d}t}T_E = -cT_E + \gamma h_E + \delta_s \tau_E. \tag{24}$$

The first term on the right describes the local damping by surface fluxes, the second term reflects the thermocline downwelling, and the third term represents the advective feedback (Ekman pumping), possibly leading to additional downwelling. The set of equations can be closed by assuming that there is an overall westerly (easterly) wind anomaly for a positive (negative) SST anomaly averaged over the entire basin of the equatorial band, but a much weaker westerly (easterly) wind anomaly averaged over the eastern half of the basin

$$[\tau] = bT_E; \quad \tau_E = b_E T_E, \tag{25}$$

where b and b_E are coupling coefficients, and the feedback from Ekman pumping can be ignored, so $\delta_s b_E = 0$. Finally, the model reduces to a coupled set of ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}T_E = -cT_E + \gamma(h_W + bT_E), \tag{26a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}h_W = -rh_w - \alpha b T_E. \tag{26b}$$

The recharge oscillator and the delayed-action oscillator share the same ocean dynamics (low-frequency forced modes) and depend on two similar parameters: the damping r in (26b) and the Rossby wave propagation time in (19); and a free parameter describing how strongly the conditions in the west influence the SST in the east. Therefore, it is not possible to distinguish the two theories based on observations or model experiments. A major advantage of the recharge paradigm is that its essential process, interior Sverdrup flow, is easier to deal with measurements compared to western boundary reflections, so that monitoring of the low-frequency convergence of heat content can be tackled [26]. However, the challenges for both paradigms are similar. The delayed oscillator is criticized because Rossby waves may propagate through the Indonesian archipelago, while in the recharge oscillator the interior mass convergence may be returned poleward in western boundary currents or may pass through Indonesia just as the low-frequency Rossby waves. Put simply, there is little to be gained from separating these two views.

4 Summary and Discussion

Conceptual models of ENSO have played an important role in providing a basic dynamical framework for its cycle. Within this framework, the main simplified model abstractions involve a subtle trait where the internal uncoupled and fast intrinsic atmospheric variability is neglected—the atmosphere is rapidly adjustable to the ocean. Furthermore, in the earlier models (discussed in previous sections) the parameters weighting the proposed physical processes for growth and decay remain constant in time, placing the system in an unstable dynamical regime. However, analysis of observations later revealed that the relative importance of different physical processes may vary between events and no two ENSO episodes are completely alike.

The diversity of ENSO events and the observed asymmetries in its onset gave rise to the hypothesis that the system may not necessarily be in the unstable regime but could be a stable system driven by noise—the emergence of the stochastic forcing paradigm [22, 27, 30]. This stochastic view naturally makes use of scales-separation to divide the tropical dynamical system into fast and slow timescales. In this case, the slow timescale derives from the relaxation time of the equatorial ocean basin (modified by the atmospheric coupling), while the fast timescale is originated from the life cycle of (random) atmospheric convective disturbances. ENSO is therefore considered to behave like a stochastic oscillator, and linear stability analysis of coupled models (stochastic optimals) reveals that they are particularly susceptible to external (uncorrelated) forcing with specific large-scale atmospheric patterns. If variability from the fast (random) component "projects" significantly onto the pattern of susceptibility, then it can perturb the coupled system and convert the original regular oscillation to an irregular one. There are uncertainties about these patterns since they somewhat vary between different coupled models. However, once perturbed the system may rapidly (1–2 weeks) develop a very characteristic response which in some coupled models strongly resembles a westerly wind burst

(WWB; [21] and the references therein). In fact, this is supported by previous studies that show that every El Niño event during the past 50 years was accompanied by WWB activity [8, 10]. WWBs are linked to either paired or individual tropical cyclones, cold surges from midlatitudes, convective phase of the Madden–Julian oscillation [24, 25], or even a combination of them. The type of noisy WWB forcing of ENSO may fall into two categories: (a) an effectively white noise, in time domain, projected on the stochastic optimals or (b) stationary in time (seasonally dependent), in which case stochasticity is somehow correlated to a slow scale variability and thus produce multiplicative stochastic noise. Yet, caution must be taken since there are different definitions in use for WWBs between atmospheric scientists and oceanographers as reported in [54]. In considering ENSO as a stochastically forced system, the lack of atmospheric variability is alleviated in the earlier simplified models for ENSO. However, perhaps the most fundamental controversy regarding the role of atmospheric stochastic perturbations is in what dynamical regime the coupled system lies since there may be many indistinguishable behaviors between a system that is marginally stable and one that is slightly unstable. Nevertheless, this debate of stochastic forcing versus unstable dynamics has important implications for its predictability. The predictability of ENSO is more limited, in the order of a few months, if ENSO is a stable mode triggered by stochastic forcing than if ENSO is a self-sustained mode, for which useful ENSO predictions beyond a couple of years may be possible [40]. Furthermore, the dominance of one type of stochastic forcing also has implications for predictability [12]. This uncertainty renders its prediction a continuously challenging task since the real ENSO seems to operate near the critical boundary in "parameter space" (class of model) so that the dynamical regime may shift from one to the other. The lack of a long and reliable observed data set is one of the reasons for not being able to arrive at a statistically robust answer to this controversy so far.

References

1. Anderson, D.L.T., McCreary, J.P.: Slowly propagating disturbances in a coupled ocean-atmosphere model. J. Atmos. Sci. **42**(6), 615–629 (1985)
2. Battisti, D.: Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere-ocean model. J. Atmos. Sci. **45**(20), 2889–2919 (1988)
3. Battisti, D., Hirst, A.: Interannual variability in a tropical atmosphere-ocean model: influence of the basic state, ocean geometry and nonlinearity. J. Atmos. Sci. **46**(12), 1687–1712 (1989)
4. Camayo, R., Campos, E.: Application of wavelet transform in the study of coastal trapped waves off the west coast of South America. Geophys. Res. Lett. **33**, L22,601 (2006). https://doi.org/10.1029/2006GL026395
5. Cane, M.A., Zebiak, S.E.: A theory for El Niño and the Southern Oscillation. Science **228**(4703), 1085–1087 (1985)
6. Cane, M.A., Munnich, M., Zebiak, S.E.: A study of self-excited oscillations of the tropical ocean-atmosphere system. Part I: linear analysis. J. Atmos. Sci. **47**, 1562–1577 (1990)
7. Capotondi, A., Wittenberg, A.T., Newman, M., Lorenzo, E.D., Yu, J.Y., Braconnot, P., Cole, J., Dewitte, B., Giese, B., Guilyardi, E., Jin, F.F., Karnauskas, K., Kirtman, B., Lee, T., Schneider, N., Xue, Y., Yeh, S.W.: Understanding ENSO diversity. Bull. Am. Meteorol. Soc. **96**(6), 921–938 (2015). https://doi.org/10.1175/BAMS-D-13-00117.1

8. Eisenman, I., Yu, L., Tziperman, E.: Westerly wind bursts: ENSO's tail rather than the dog. J. Climate **18**, 5224–5238 (2005)
9. Garcia, R.R., Salby, M.L.: Transient response to localized episodic heating in the tropics. Part II: far-field behavior. J. Atmos. Sci. **44**, 499–532 (1987). https://doi.org/10.1175/1520-0469(1987)044<0499:TRTLEH>2.0.CO;2
10. Gebbie, G., Eisenman, I., Wittenberg, A., Tziperman, E.: Modulation of westerly wind bursts by sea surface temperature: a semistochastic feedback for ENSO. J. Atmos. Sci. **64**, 3281–3295 (2007)
11. Gill, A.: Some simple solutions for heat-induced tropical circulation. Q. J. R. Meteorol. Soc. **106**, 447–462 (1980)
12. Harrrison, D., Vecchi, G.: Westerly wind events in the tropical Pacific 1986–1995. J. Climate **10**, 3131–3156 (1997)
13. Hirst, A., Lau, K.M.: Intraseasonal and interannual oscillations in coupled ocean-atmosphere models. J. Climate **3**, 713–725 (1990)
14. Horel, J., Wallace, J.: Planetary-scale atmospheric phenomena associated with the southern oscillation. Mon. Weather Rev. **109**, 813–829 (1981)
15. Hoskins, B.J., Karoly, D.J.: The steady linear response of a spherical atmosphere to thermal and orographic forcing. J. Atmos. Sci. **38**, 1179–1196 (1981)
16. Jin, F.F.: An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. J. Atmos. Sci. **54**(7), 811–829 (1997)
17. Jin, F.F.: An equatorial ocean recharge paradigm for ENSO. Part II: a stripped-down coupled model. J. Atmos. Sci. **54**(7), 830–847 (1997)
18. Jin, F., Neelin, J.: Modes of interannual tropical ocean–atmosphere interaction – a unified view. Part I: numerical results. J. Atmos. Sci. **50**, 3477–3503 (1993)
19. Karoly, D.J.: Southern Hemisphere circulation features associated with El Niño–Southern Oscillation. J. Climate **2**, 1239–1252 (1989)
20. Kirtman, B.P.: Oceanic Rossby wave dynamics and the ENSO period in a coupled model. J. Climate **10**, 381–400 (1997)
21. Kleeman, R.: Stochastic theories for the irregularity of ENSO. Philos. Transact. A Math. Phys. Eng. Sci. **366**, 2511–26 (2008). https://doi.org/10.1098/rsta.2008.0048
22. Kleeman, R., Moore, A.M.: A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. J. Atmos. Sci. **54**, 753–767 (1997). https://doi.org/10.1175/1520-0469(1997)054!0753:ATFTLOO2.0.CO;2
23. Kumar, A., Chen, M.: What is the variability in US west coast winter precipitation during strong El Niño events? Clim. Dyn. **48**(7–8), 2789–2802 (2017)
24. Madden, R., Julian, P.: Description of a 40–50 day oscillation in the zonal wind in the tropical Pacific. J. Atmos. Sci. **28**, 702–708 (1971)
25. Madden, R., Julian, P.: Description of global-scale circulation cells in the tropics with a 40–50 day period. J. Atmos. Sci. **29**, 1109–1123 (1972)
26. McPhaden, M.J., Zhang, D.: Pacific ocean circulation rebounds. Geophys. Res. Lett. **31**, L18,301 (2004). https://doi.org/10.1029/2004GL020727.
27. Moore, A.M., Kleeman, R.: Stochastic forcing of ENSO by the Intraseasonal Oscillation. J. Climate **12**, 1199–1220 (1999). https://doi.org/10.1175/1520-0442(1999)012!1199:SFOEBTO2.0.CO;2
28. Neelin, J.D., Jin, F.F.: Modes of interannual tropical ocean-atmosphere interaction – a unified view. Part II: analytical results in the weak-coupling limit. J. Atmos. Sci. **50**(21), 3504–3522 (1993)
29. Neelin, J.D., Latif, M., Jin, F.F.: Dynamics of coupled ocean-atmosphere models: the tropical problem. Annu. Rev. Fluid Mech. **26**(1), 617–659 (1994)
30. Penland, C., Sardeshmukh, P.D.: The optimal growth of tropical sea surface temperature anomalies. J. Climate **8**, 1999–2024 (1995)
31. Philander, S.G.: El Niño, La Niña and the Southern Oscillation, International Geophysical Series. Academic Press, Cambridge (1999)

32. Pope, J.O., Holland, P.R., Orr, A., Marshall, G.J., Phillips, T.: The impacts of El Niño on the observed sea ice budget of West Antarctica. Geophys. Res. Lett. **44**, 6200–6208 (2017)
33. Ramírez, E., Silva Dias, P.L., Veiga, J.A., Camayo, R., Santos, A.: Multivariate analysis of the energy cycle of the South American rainy season. Int. J. Climatol. **29**, 2256–2269 (2009). https://doi.org/10.1002/joc.1858
34. Ramírez, E., Silva Dias, P.L., Raupp, C., Bonatti, J.P.: The family of anisotropically scaled equatorial waves. J. Adv. Model. Earth Syst. **3**, M12,002 (2011)
35. Ramirez, E., da Silva Dias, P.L., Raupp, C.F.: Multiscale atmosphere–ocean interactions and the low-frequency variability in the equatorial region. J. Atmos. Sci. **74**, 2503–2523 (2017). https://doi.org/10.1175/JAS-D-15-0325.1
36. Salby, M.L., Garcia, R.R.: Transient response to localized episodic heating in the tropics. Part I: excitation and short-time near-field behavior. J. Atmos. Sci. **44**(2), 458–498 (1987). https://doi.org/10.1175/1520-0469(1987)044<0458:TRTLEH>2.0.CO;2
37. Sanabria, J., Bourrel, L., Dewitte, B., Frappart, F., Rau, P., Solis, O., Labat, D.: Rainfall along the coast of Peru during strong El Niño events. Int. J. Climatol. **38**, 1737–1747 (2018)
38. Schneider, E.K., Huang, B., Shukla, J.: Ocean wave dynamics and El Niño. J. Climate **8**(10), 2415–2439 (1995)
39. Schubert, W., Silvers, L., Masarik, M., Gonzalez, A.: A filtered model of tropical wave motions. J. Adv. Model. Earth Syst. **1**(3), 1–11 (2009)
40. Siqueira, L., Kirtman, B.: Nonlinear dynamics approach to the predictability of the Cane-Zebiak coupled ocean-atmosphere model. Nonlin. Processes Geophys. **21**, 155–163 (2014)
41. Stefanova, L., Krishnamurti, T.: Kinetic energy exchange between the time scales of ENSO and the Pacific Decadal Oscillation. Meteorol. Atmos. Phys. **114**(95) (2011). https://doi.org/10.107/s00703-011-0162-8
42. Suarez, M., Schopf, P.: A delayed action oscillator for ENSO. J. Atmos. Sci. **45**, 3283–3287 (1988)
43. Takahashi, K., Martínez, A.G.: The very strong coastal El Niño in 1925 in the far-eastern Pacific. Clim. Dyn., 1–27 (2017). http://dx.doi.org/10.1007/s00382-017-3702-1
44. Taschetto, A.S., Rodrigues, R.R., Meehl, G.A., McGregor, S., England, M.H.: How sensitive are the Pacific–tropical North Atlantic teleconnections to the position and intensity of El Niño-related warming? Clim. Dyn. **46**(1841) (2016)
45. Tedeschi, R.G., Cavalcanti, I.F.A., Grimm, A.M.: Influences of two types of ENSO on South American precipitation. Int. J. Climatol. **33**(6), 1382–1400 (2013). http://dx.doi.org/10.1002/joc.3519
46. Thomson, D.: Jackknifing multitaper spectrum estimates. IEEE Signal Process. Mag. **24**, 20–30 (2007). http://dx.doi.org/10.1109/MSP.2007.4286561
47. Trenberth, K.E.: The definition of El Niño. Bull. Am. Meteorol. Soc. **78**(12), 2771–2777 (1997). http://dx.doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2
48. Trenberth, K.E., Branstator, G.W., Karoly, D., Kumar, A., Ropelewski, N.L.C.: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. J. Geophys. Res. **103**(C7), 14,291–14,324 (1998)
49. Wang, B., An, S.I.: Why the properties of El Niño changed during the late 1970s. Geophys. Res. Lett. **28**(19), 3709–3712 (2001)
50. Wang, C., Wang, W., Wang, D., Wang, Q.: Interannual variability of the South China Sea associated with El Niño. J. Geophys. Res. **111**(C03023) (2006)
51. Webster, P.J., Magaña, V.O., Palmer, T.N., Shukla, J., Tomas, R.A., Yanai, M., Yasunari, T.: Monsoons: processes, predictability, and the prospects for prediction. J. Geophys. Res. **103**(C7), 14,451–14,510 (1998)
52. Wyrtki, K.: El Niño – the dynamic response of the equatorial pacific ocean to atmospheric forcing. J. Phys. Oceanogr. **5**, 572–584 (1975). http://dx.doi.org/10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2
53. Zebiak, S., Cane, M.: A model El Niño southern oscillation. Mon. Weather Rev. **115**, 2262–2278 (1987)
54. Zhang, C.: Madden-Julian oscillation. Rev. Geophys. **43**, RG2003 (2005). http://dx.doi.org/10.1029/2004RG000158

Observation, Theory, and Numerical Modeling: Atmospheric Teleconnections Leading to Generalized Frosts over Southeast South America

Gabriela V. Müller, Guillermo Berri, and Marilia de Abreu Gregorio

1 Introduction

Natural phenomena, such flooding, landslides, droughts, frosts, among others, are responsible for significant social, economic, and environmental damages and losses, which are more recurrent and have increasingly intense impacts. In particular, frosts are extreme events that directly and significantly affect human and economic activities, causing impacts and damages in various sectors. Frosts in extratropical latitudes of South America are frequent since its climate regime is characterized by incursions of air masses of polar origin, especially during the winter. They are very important in the center-northeast of Argentina, a region called Pampa Húmeda (Wet Pampa), an extended plain of more than 750,000 km^2. It is considered one of the world regions where the impacts are greater due to the type of production and the number of inhabitants that live there [23, 29]. In order to give proper representation to the frequent frost phenomenon that in occasions cover extended regions, [28] defined a criterion for identifying generalized frosts (GF) events, by considering the days on which a surface temperature below 0 °C is recorded at more than 75% of the meteorological stations in the Wet Pampa.

G. V. Müller (✉) · M. de A. Gregorio
Centro de Estudios de Variabilidad y Cambio Climático (CEVARCAM) – FICH – UNL, Santa Fe, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
e-mail: gvmuller@fich.unl.edu.ar; mgregorio@fich.unl.edu.ar

G. Berri
Facultad de Ciencias Astronómicas y Geofísicas – UNLP, La Plata, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
e-mail: gberri@fcaglp.unlp.edu.ar

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_2

The processes that involve the development of this extremely cold condition are associated with complex interaction mechanisms in different atmospheric scales [32, 33]. Remote forcing, such as anomalous convection in the tropical region, can act as a trigger mechanism for Rossby waves generation that propagate to the extratropics, giving rise to the tropic–extratropic teleconnections [2, 12]. However, the wave propagation pattern will depend on the structure of the basic state, as well as the remote forcing [1]. The term teleconnection means remote connection and its study is a good example of the linkage between observations, theory, and numerical model results. In this work we present the linkage by means of the study of GF events in the Wet Pampa.

2 Theory of Rossby Waves

The identification of teleconnections and the analysis of their effect on the horizontal structure of the atmospheric circulation can be useful to understand the occurrence of anomalous events in various regions of the planet. It provides an overview of the atmospheric circulation in which a local forcing acts in such a way that influences remote regions. Such local forcing may be a heat source ([2, 10–12, 18, 35, 42], among others), or an orographic obstacle ([5, 7, 9, 12, 39], among others), which acts as wave generators that propagate in the atmosphere. One way to analyze the propagation of waves in the terrestrial sphere is by means of the theory of Rossby waves in a barotropic atmosphere.

2.1 Barotropic Equations

The simplest context for the study of Rossby waves is by considering a shallow layer of incompressible fluid on a rotating earth, with a simple geometry. By ignoring the effects of the earth sphericity and allowing the change with latitude of the vertical component of the earth rotation, the dynamics can be described by the two-dimensional rotating Euler equations:

$$\frac{d\mathbf{v}_H}{dt} + f\hat{k} \times \mathbf{v}_H = -\frac{1}{\rho_0}\nabla p \tag{1a}$$

$$\nabla \cdot \mathbf{v}_H = 0 \tag{1b}$$

where $\frac{d}{dt}$ is the total time derivative operator that applied to each scalar component A of the vector v_H is $\frac{d}{dt}A = \frac{d}{dt}A + (\mathbf{v} \cdot \nabla)A$, f is the Coriolis parameter, $\hat{k}$ is the unit vertical vector, ∇p is the pressure gradient, and $\mathbf{v}_H$ the horizontal component of the fluid flow.

Equation (1a) represents the fluid motion set by an external pressure gradient, while Eq. (1b) represents the incompressibility of the fluid. Based on the theory developed by Rossby [36, 37], the variation of the Coriolis parameter with latitude can be approximated by expanding the latitudinal dependence of f in a Taylor series about a reference latitude ϕ_0 and retaining only the first two terms to yield

$$f = f_0 + \beta y \tag{2}$$

where $\beta \equiv (df/dy)_{\phi_0} = 2\Omega\cos\phi_0/a$ and $y = 0\ at\ \phi_0$. This approximation is usually referred to as the *midlatitude β-plane* approximation.

2.2 Barotropic Vorticity Equation

By imposing incompressibility, the family of sound waves is removed from the set of possible solutions, yet the Coriolis parameter provides a restoring mechanism leading to the nontrivial, the so-called Rossby wave solution. To promptly set equations for describing the Rossby waves, the non-divergence condition $\nabla \cdot \mathbf{v}_H = 0$ allows us to describe the problem in stream-function formulation ψ, with $\mathbf{v}_H = \hat{k} \times \nabla\psi$.

Thus, taking the curl of Eq. (1a), the non-linear, the non-divergent barotropic Rossby waves are described by Charney [6], Pedlosky [34] as:

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = 0 \tag{3}$$

where $\zeta = \nabla^2\psi = \hat{k} \cdot (\nabla \times \mathbf{v}_H)$, and β is the meridional derivative of the Coriolis parameter.

The non-linear contribution in Eq. (3) is clearly due to the Jacobian

$$J(\psi, \zeta) = [\frac{\partial \psi}{\partial x}\frac{\partial \zeta}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \zeta}{\partial x}] \tag{4}$$

which represents the advection of the relative vorticity ζ since $\mathbf{v}_H = \hat{k} \times \nabla\psi$.

When the advective contribution is linearized by approximating $J(\psi, \zeta) \approx \bar{U}(y)\frac{\partial}{\partial x}\zeta$ with a time invariable westerly flow with meridional shear $\bar{U}(y)$, we obtain the following wave equation:

$$\frac{\partial \zeta}{\partial t} + \bar{U}(y)\frac{\partial \zeta}{\partial x} + \beta \frac{\partial \psi}{\partial x} = 0 \tag{5}$$

Equation (5) is in terms of the stream function ψ that can be expressed as:

$$\psi' = \psi_0 e^{(i(kx+ly-\sigma t))} \tag{6}$$

so that the relative vorticity results in terms of the perturbation stream function as $\zeta'=\nabla^2\psi'$(since ζ_0=0). Thus the solution allows to obtain the following frequency:

$$\sigma = k\bar{U} - \frac{\beta^+ k}{k^2 + l^2} \tag{7}$$

where the generalized β^+ parameter, the meridional gradient of absolute vorticity, is given by:

$$\beta^+ = \beta - \frac{\partial^2 \bar{U}}{\partial y^2} \tag{8}$$

On the other hand, the wave activity in the plane propagates with the group velocity whose components are obtained by deriving the frequency with respect to the wave numbers k and l:

$$c_g = (u_g, v_g) = (\partial\sigma/\partial k, \partial\sigma/\partial l) = (c, 0) + \left(\frac{2\beta^+}{K^2}\right)\cos(\alpha)\hat{K} \tag{9}$$

where $c = \sigma/k$ is the eastward phase velocity, $\hat{K}$ is the unit vector normal to the trough and ridge axis with an eastward positive component, and α is the angle the vector forms with the eastward direction. In addition, the total stationary Rossby wave number K_s is obtained by making $\sigma = 0$ in Eq. (7):

$$K_s = \sqrt{\frac{\beta^+}{\bar{U}}} \tag{10}$$

Therefore, the development of stationary Rossby waves is possible only in a westerly flow ($\bar{U} > 0$) as long as β^+ is always positive. From Eqs. (9) and (10), the group velocity for stationarity condition:

$$\mathbf{c}_{\mathrm{gs}} = 2\bar{U}\cos\alpha\,\hat{K} \tag{11}$$

Thus for the stationary Rossby waves, the energy propagates perpendicular to the axis of the ridges and troughs with velocity equal to $2\bar{U}\cos\alpha$. In sum, this theory describes how an initial vorticity source in a barotropic atmosphere with a stationary zonal flow is able to generate a series of ridges and troughs whose energy is dispersed with a group velocity given by Eq. (11). The resulting wave trains are linked to the observed teleconnection patterns ([12, 16], among others). In particular, the Southern Hemisphere offers a unique environment to test the propagation of these waves in observed atmospheric circulation patterns, due to the equivalent

barotropic structure of the atmosphere and the prevalent zonal structure of the time mean flow [19]. According to [17] a useful diagnostic tool for representing the mean background state in which the transients are embedded is the calculation of K_s given by Eq. (10) which is the total wave number at which a barotropic Rossby wave is stationary at a particular location and in a given background zonal flow ($\bar{U}$).

3 Local Forcings of Rossby Waves

Rossby waves are a type of atmospheric motion due to the conservation of absolute vorticity and the variation of the Coriolis parameter with latitude, the so-called β-effect. This synoptic scale wave motion in the horizontal plane is characterized by the alternation of regions of cyclonic and anticyclonic vorticity as the wave propagates, and the restoring force of this wave is the Coriolis force. One of the factors favoring Rossby waves generation is tropical diabatic heating, which is balanced by vertical upward motion and upper-level divergence resulting in a disturbance of the relative vorticity field. Hoskins et al. [12] used a barotropic model to study the atmospheric response to a thermal forcing and found the influence of tropical heating and upper-level divergence on the generation of Rossby waves that propagate along extratropical curved pathways in the form of wave trains associated with teleconnections. The forcing for the Rossby waves in a divergent field can be inferred by means of the barotropic vorticity equation:

$$\frac{\partial \xi}{\partial t} + \mathbf{v} \cdot \nabla \xi = -\xi D \tag{12}$$

where $\xi = \mathrm{f} + \zeta$ is the absolute vorticity, f the planetary vorticity, ζ the relative vorticity, $\mathbf{v}$ the wind vector, and $D = \partial u/\partial x + \partial v/\partial y$ is the horizontal divergence. The left-hand side of the equation describes the Rossby wave propagation, while the right-hand side represents the forcing. However, as ξ is relatively small in the tropics, the heating anomalies in the tropics would be ineffective to produce Rossby waves, which does not match the observed teleconnection patterns. In order to analyze in more detail the barotropic vorticity equation, the wind vector can be expressed as the sum of a rotational (v_ψ) and a divergent (v_χ) component, in terms of a stream function ψ and a velocity potential χ, respectively, as follows:

$$v_\psi = k \times \nabla \psi \tag{13}$$

$$v_\chi = \nabla \chi \tag{14}$$

Therefore, the relative vorticity can be written as $\zeta = \nabla^2 \psi$ and the divergence as $D = \nabla^2 \chi$. By replacing $\mathbf{v}$ as $v_\psi + v_\chi$, the barotropic vorticity equation becomes:

$$\frac{\partial \xi}{\partial t} + v_\psi \cdot \nabla \xi = -\xi D - v_\chi \cdot \nabla \xi \tag{15}$$

In this way, the left-hand side of the equation involves the rotational part and the right-hand side the divergent part of the wind. Sardeshmukh and Hoskins[38] discussed this equation and identified the right-hand side terms as Rossby waves sources by the divergent wind, S, as follows:

$$S = -\xi D - v_\chi \cdot \nabla \xi \tag{16}$$

Thus, the forcing for Rossby waves will be greater in areas where the divergence, the divergent wind, the absolute vorticity, or its gradient is larger. This may take place in the Southern Hemisphere regions southward of the equator where the upper-level divergence associated with deep convection is larger and there are large vorticity gradients associated with the subtropical jet. The previous equation indicates that Rossby waves can be excited by tropical diabatic heating even though ξ may be small in regions where the heating takes place, since the divergent flow will be larger there. Towards the subtropics the gradient of ξ is larger, contributing to higher values of S. We can rewrite the previous equation as:

$$S = S1 + S2 \tag{17}$$

where $S1 = -\xi D$ is the term of vortex stretching that represents the generation of vorticity by divergence and $S2 = -v_\chi \cdot \nabla \xi$ is the advection of absolute vorticity by the divergent flow. Lu and Kim[20] used these expressions to evaluate the role of tropical and extratropical heating in the generation of Rossby waves in the Northern Hemisphere. Sardeshmukh and Hoskins[38] highlighted the importance of the advection term of absolute vorticity by the divergent flow in their study of the atmospheric response to tropical divergence over the western Pacific Ocean. The local forcings are applied in the experiments described in Sect. 6 using the model presented in the following section.

4 Model IGCM

The baroclinic model IGCM (Intermediate Global Circulation Model—Center for Global Atmospheric Modeling, University of Reading) was used in several studies, like [2, 5, 15, 21]. Although the model is dry and does not include any physical process linked to the phase change of water vapor, many aspects of the atmospheric dynamics can be reproduced, from synoptic scale systems [13] to large-scale perturbations [43]. IGCM is a baroclinic model that runs on a global domain, with a T42 spectral truncation of the zonal wave number, and includes horizontal and vertical diffusion and Newtonian cooling [15]. The large-scale dynamics of the atmosphere can be interpreted in qualitative terms by means of the linear theory of waves because there is a good agreement between IGCM results and those obtained with the barotropic [2].

The model is global and is formulated on a sphere with a multi-level vertical coordinate. In the horizontal, it uses a spectral representation and in the vertical a finite difference scheme with a semi-implicit method in which the terms of linearized gravity waves are time averages so that the faster movements are smoothed for the time integration. Blackburn [3] included an alternative vertical scheme based on that used by the ECMWF [41] that preserves the angular momentum since the original finite difference scheme did not.

The horizontal equations of motion are expressed in the form of vorticity and divergence, which is convenient when using spectral techniques and semi-implicit time schemes. The equations of motion are formulated in dimensionless form, on a rotating sphere, assuming a hydrostatic and adiabatic perfect gas, without viscosity:

$$\frac{\partial \xi}{\partial t} = \frac{1}{1-\mu^2}\frac{\partial \Im_v}{\partial \lambda} - \frac{\partial \Im_u}{\partial \mu} \tag{18}$$

$$\frac{\partial D}{\partial t} = \frac{1}{1-\mu^2}\frac{\partial \Im_v}{\partial \lambda} - \frac{\partial \Im_u}{\partial \mu} - \nabla^2\left(\frac{U^2+V^2}{2(1-\mu^2)} + \phi + \overline{T} ln p_*\right) \tag{19}$$

$$\frac{\partial T'}{\partial t} = \frac{-1}{1-\mu^2}\frac{\partial}{\partial \lambda}(UT') - \frac{\partial}{\partial \mu}(VT') + DT' - \acute{\sigma}\frac{\partial T}{\partial \sigma} + k\frac{T\omega}{p} \tag{20}$$

$$\frac{\partial ln p_*}{\partial t} = -V \cdot \nabla ln p_* - D - \frac{\partial \acute{\sigma}}{\partial \sigma} \tag{21}$$

$$\frac{\partial \phi}{\partial ln \sigma} = -T \tag{22}$$

With $\Im_u$ and $\Im_v$ defined as follows:

$$\Im_u = V\xi - \acute{\sigma}\frac{\partial U}{\partial \sigma} - T'\frac{\partial ln p_*}{\partial \lambda} \tag{23}$$

$$\Im_v = U\xi - \acute{\sigma}\frac{\partial V}{\partial \sigma} - T'(1-\mu^2)\frac{\partial ln p_*}{\partial \lambda} \tag{24}$$

where ξ is the absolute vorticity, D the divergence, $T = T(\sigma) + T'$ the temperature, p_* the surface pressure, ϕ the geopotential, σ is the vertical coordinate pressure, ω is the vertical velocity, λ is the longitude, and $\mu = \sin\theta$, where θ is the latitude. The horizontal advection operator is

$$V \cdot \nabla = \frac{U}{1-\mu^2}\left(\frac{\partial}{\partial \lambda}\right) + V\left(\frac{\partial}{\partial \mu}\right) \tag{25}$$

The velocity potential χ and the stream function ψ are introduced in this form:

$$U = -\left(1-\mu^2\right)\frac{\partial \psi}{\partial \mu} + \frac{\partial \chi}{\partial \lambda} \tag{26}$$

$$V = \frac{\partial \psi}{\partial \lambda} + \left(1 - \mu^2\right) \frac{\partial \chi}{\partial \mu} \tag{27}$$

$$\xi = 2\mu + \frac{1}{1-\mu^2} \frac{\partial V}{\partial \lambda} - \frac{\partial U}{\partial \mu} = 2\mu + \nabla^2 \psi \tag{28}$$

$$D = \frac{1}{1-\mu^2} \frac{\partial U}{\partial \lambda} + \frac{\partial V}{\partial \mu} = \nabla^2 \chi \tag{29}$$

The vorticity ξ, the divergence D, and the temperature T at each level, as well as the logarithm of the surface pressure p_*, are expressed as a series of the form:

$$A = \sum_{m=-M}^{M} \sum_{n=|m|}^{N} A_n^m Y_n^m(\mu, \lambda) \tag{30}$$

where A represents any of those variables and λ is the longitude, μ the sine of the latitude, A_n^m is a complex coefficient, $Y_n^m = P_n^m(\mu) e^{im\lambda}$ is a spherical harmonic, where P_n^m a standard Legendre function, m is the zonal wave number, $n - m$ is the number of zeros between the north and south poles, n is the total wave number, and N,M are the truncation wave numbers defined as the largest number of wave retained in the spectral series.

The model employs jagged triangular truncation of the total wave number 42, which allows an independent representation of the position in the sphere. It can be executed in a wide range of horizontal and vertical resolutions. It has 12 vertical equally spaced levels which are chosen so that the resolution everywhere is lower than 105 hPa, with higher resolution near the tropopause. The model equations are integrated using a spectral transformation method as described in detail by Bourke [4].

Equation (18) is expressed in spherical coordinates, but the conservation of absolute vorticity $\xi = \zeta + f$ can be written in Cartesian coordinates as:

$$\frac{d\xi}{dt} = -V \cdot \nabla \xi \tag{31}$$

So that:

$$\frac{\partial \zeta}{\partial t} = u \frac{\partial}{\partial x}(\zeta + f) + v \frac{\partial}{\partial y}(\zeta + f) \tag{32}$$

$$= u \frac{\partial}{\partial x}(\zeta + f) + v \frac{\partial \zeta}{\partial y} + v \frac{\partial f}{\partial y} \tag{33}$$

Replacing in Eq. (31), $\frac{\partial f}{\partial y} = \beta$, $u = -\frac{\partial \psi}{\partial y}$, $v = \frac{\partial \psi}{\partial x}$, we obtain Eq. (3).

5 Preferential Paths of Rossby Waves Propagation

The wave theory provides a simple and useful interpretation of global Rossby wave propagation. This theory describes how an initial vorticity source in a barotropic atmosphere with a stationary zonal flow (Eq. (5)) is able to generate a series of ridges and troughs whose energy is dispersed with a group velocity given by (Eq. (11)). The resulting wave trains are linked to the observed teleconnection patterns ([12, 16], among others). In particular, the Southern Hemisphere offers a unique environment to test the propagation of these waves in observed atmospheric patterns, due to the equivalent barotropic structure of the atmosphere and the prevalent zonal structure of the time mean flow [19]. The following questions arise, which are the preferential paths of wave propagation, and where are the waves originated?

A useful diagnostic tool for representing the mean background state in which the transients are embedded is the calculation of K_s given by Eq. (10) which is the total wave number at which a barotropic Rossby wave is stationary at a particular location and in a given background zonal flow ($\bar{U}$) [17]. According to this theory, the characteristics of the Rossby wave propagation in a given basic state may be determined by the analysis of K_s. Hoskins and Ambrizzi [11], among others, showed that the distribution of K_s can be used to infer the location of critical lines and waveguides for stationary Rossby waves. For example, zonally oriented regions of relative high values of K_s bounded by lower values to the north and south (usually associated with strong jet streams) indicate favorable locations for guiding Rossby waves.

Based on the large-scale physical mechanisms associated with GF, manifested through teleconnection patterns given by stationary Rossby waves, [24] used K_s to give a qualitative picture of the effects of the basic state flow within the westerly duct on the subtropical and polar wave propagation. This was a first assessment of the impact of the GF basic flow on the dynamical structure relevant for the Rossby wave dispersion. Using this theoretical concept, the distribution of the stationary wave number (K_s) in the Southern Hemisphere is obtained for the two opposite basic states given by the composition of winters with maximum and minimum frequency of GF occurrence in the Wet Pampa. To understand the large-scale atmospheric circulation associated with these cold episodes, [30] identified years of extreme frequency of frost occurrence, during the period 1961–1990. A maximum (minimum) frequency of GF occurrence is considered when the number of frosts is one standard deviation (σ) above (below) the mean value during the austral winter (JJA), i.e., GF $+\sigma$ and GF $-\sigma$, respectively.

From these fields it is possible to determine preferential wave propagation paths that reach South America. The distribution of K_s emphasizes the importance of the jets as efficient waveguides, with a good agreement between their positions and the bands of local K_s maxima, with zonal orientation. Nevertheless, there are differences between both analyzed basic states, which explain in part why there are winters with GF $+\sigma$ and GF $-\sigma$, as shown by Müller and Ambrizzi [24]. The geographic distribution of K_s is shown in Fig. 1a, b with the zero wind contours,

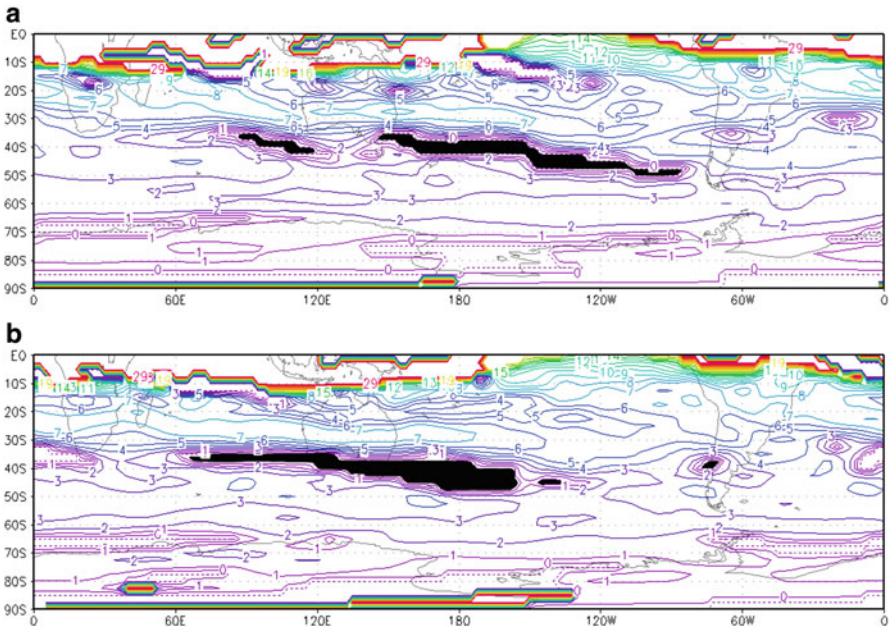


Fig. 1 Stationary wave number K_s at 250 hPa, corresponding to (**a**) GF $+\sigma$ and (**b**) GF $-\sigma$ (from [24])

$\bar{U} = 0$, in thick color lines, while those corresponding to zero K_s (i.e., $\beta^+ = 0$) in dotted lines. The equatorial belt is dominated by a sector where the propagation is inhibited due to the presence of a transition from westerly to easterly flow, which forms a "critical line" for the propagation of stationary Rossby waves. This region is marked with thicker lines in Fig. 1a, b indicating high wave number values ($K_s \geq 20$).

A local maximum of K_s in the region of maximum zonal winds, i.e., along the subtropical jet (see Fig. 2a, b), shows wave numbers 6 and 7 in both winter groups. This region also contains uniform values of K_s meridionally limited by smaller ones. In the polar side of the subtropical jet the presence of $K_s = 0$ (the black shaded region) inhibits the propagation of Rossby waves through it because, according to the wave theory, they must deviate before reaching this region, so that the $K_s = 0$ line acts as a waveguide boundary. An interesting feature of this last case is the difference in the longitudinal extension of K_s for GF $+\sigma$ with respect to GF $-\sigma$. In Fig. 1a this region extends very near to the South American continent. However, the K_s pattern for the GF $-\sigma$ winters (Fig. 1b) shows a shorter extension of the waveguide duct, besides of a region of zero K_s between $35°S$ and $40°S$ in the west of Argentina, which coincides with the main entrance region of weather systems over South America ([8, 40], among others). This feature is not seen in the GF $+\sigma$ winters and it may have some influence on the Rossby waves propagation paths and, therefore, in the frost events.

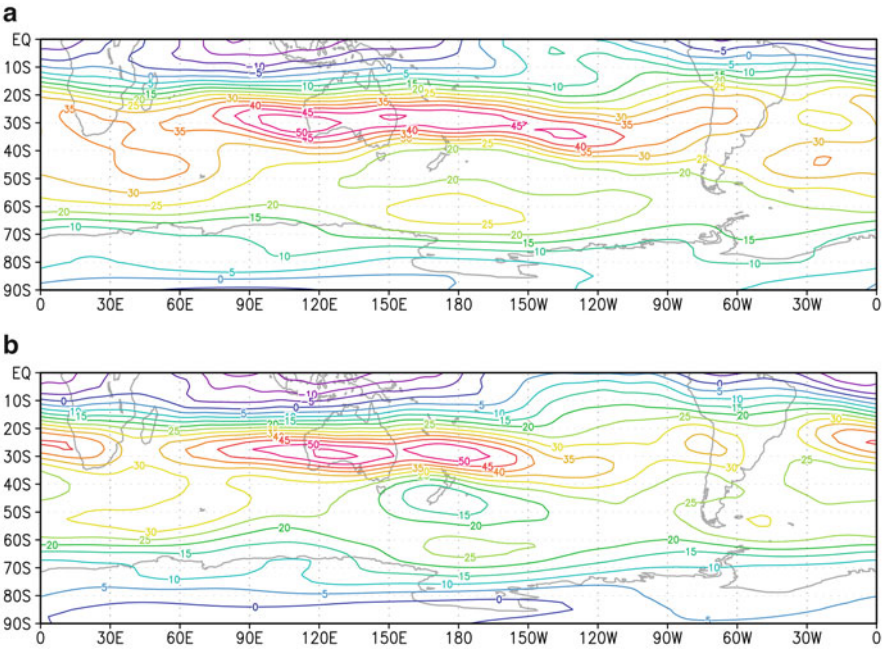


Fig. 2 Mean zonal wind component (ms^{-1}) at 250 hPa for the austral winters, corresponding to (**a**) GF $+\sigma$ and (**b**) GF $-\sigma$ (from [24])

Another characteristic of both basic states can be seen in Fig. 2a, b that shows regions of maximum mean zonal wind of winters with GF $+\sigma$ and GF $-\sigma$. The main basic state feature for GF $+\sigma$ winters (Fig. 2a) shows maximum values of mean zonal 250 hPa wind around 30° S extending from the Indian Ocean to the Pacific Ocean. The core maximum is around 50 ms^{-1} and is located over the eastern Indian Ocean, being displaced 60° to the west of the climatological reference mean (winter JJA 1961–1990). The 35 ms^{-1} contour reaches the South American continent with a second nucleus extending southeastward from southern South Africa. An outstanding characteristic is the magnitude of zonal wind gradient over the southwestern Indian Ocean. The zonal wind maxima at high latitudes, between 20 and 25 ms^{-1}, are located along the 60° S and the polar jet position is also different from that of the climatology with a "tongue" towards South America with a longitudinal extension that brings it closer to the continent [24]. In the case of GF $-\sigma$ winters (Fig. 2b) the subtropical jet is divided in two zones, one from the Atlantic Ocean up to South Africa and other from the Indian Ocean up to the central Pacific Ocean. In this last case, the subtropical jet has two maxima around 30° S, one to the southwest of Australia and the other one at approximately the dateline with a secondary maximum to the south of it.

However, the propagation of Rossby waves depends not only on the basic state in which they are, but also on the generating source of these waves [1]. In this

sense, [23, 24, 31] performed numerical experiments with the multi-level baroclinic general circulation model IGCM, which is an appropriate tool to study the physical mechanisms of wave propagation since the response can be directly attributed to the generating source.

6 Source Regions of Rossby Waves

In order to test the generation of extratropical wave patterns which favor the development of extreme cold events over the southeast of South America, numerical simulations using IGCM were carried out. Müller [23, 24, 31] simulated with the IGCM the Rossby wave pattern responsible for the teleconnection, with two basic states given by the composition of winters with GF $+\sigma$ and GF $-\sigma$ in the Wet Pampa.

The model IGCM is run in two steps according to [23, 24, 31]. The first one consists of the buildup of the initial basic state, with which the model reaches the steady state, using geopotential heights, zonal and meridional wind, and temperature at 12 vertical sigma levels. The second step consists of the perturbation of the model by introducing a thermal forcing. The heating source has an elliptical horizontal structure and a fixed position in latitude and longitude. The vertical heating profile follows a cosine function with the maximum amplitude at 400 hPa corresponding to a 5 °C/day and decaying to zero at the surface and the top of the model. This value is equivalent to a latent heat release associated with 10 mm of precipitation. Although one may question the realism of this heating source, its role here is purely of acting as a Rossby wavemaker [24, 31]. The combination of observed tropical convection and a specific basic state can create the appropriate environment to guide Rossby waves triggered by the tropical forcing towards South America [31]. In particular, stationary Rossby waves may be one important mechanism linking anomalous tropical convection with the extreme cold events in the Wet Pampas [31]. The theory of Rossby wave propagation supports the obtained result by means of ray tracing computation (derived from Eq. (10)), by placing the forcing in the position of the anomalous tropical convection [25], so that they act as a thermal forcing proxy for the model. The areas of anomalous convection and, consequently, the position of the heat source, are identified with the observed outgoing longwave radiation (OLR) compositions of the extreme GF $+\sigma$ and GF $-\sigma$ winters.[31] found, during GF $+\sigma$ winters, two main regions of anomalous tropical convection in the Southern Hemisphere located over the Indian and the western Pacific Oceans, respectively, with significant values. On the other hand, during GF $-\sigma$ winters there were no significant OLR regions that could act as sources of Rossby waves [31].

Numerical simulations with the IGCM model were performed by Müller et al. [31], considering the position of the forcing according to the observations in both basic states. The result of the experiment with the GF $+\sigma$ basic state and the heat source located over the western Pacific Ocean shows a pattern of wind anomalies with an arc-shaped trajectory of Rossby waves, which initially goes to the South

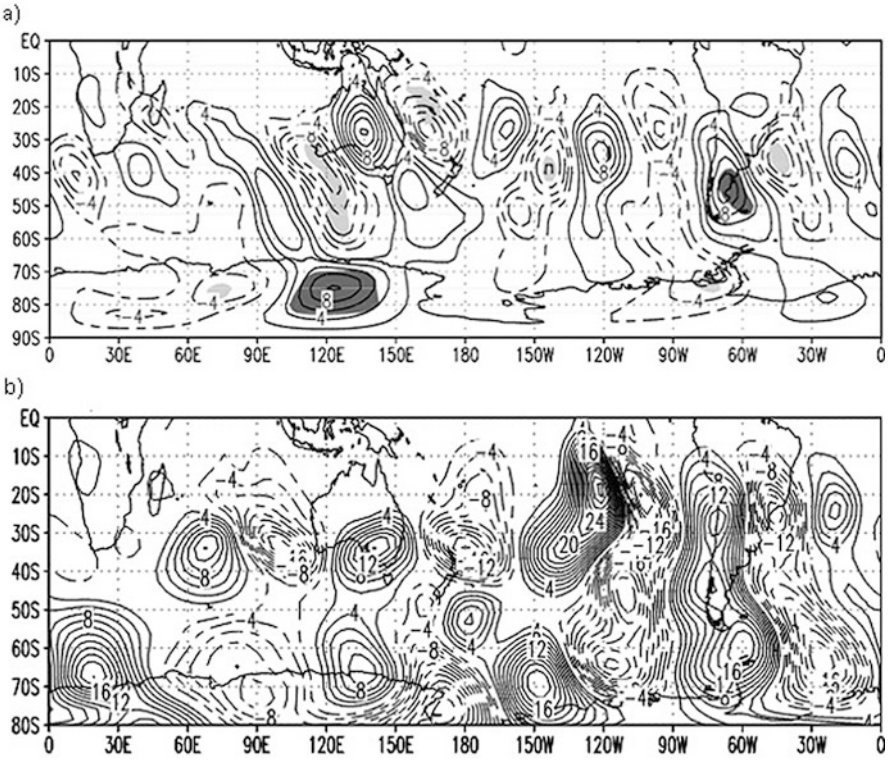


Fig. 3 (**a**) Composites of observed meridional wind field anomalies at 250 hPa for GF $+\sigma$. Positive (negative) contours are in solid (dotted) lines at 1.5 ms^{-1} intervals (from [30]). (**b**) Meridional wind component anomalies (ms^{-1}) at 250 hPa for day 14 of integration, corresponding to the simulation of GF $+\sigma$ basic state (from [24])

Pole and then turns towards the equator. This wave train reaches South America in its central part and creates southerly wind anomalies across the southern tip of the continent. However, when analyzing the experiment with the basic state GF $-\sigma$, the obtained pattern does not correspond to the preferred paths that the waves follow during the winter in the Southern Hemisphere. In this case it is possible to observe two waves trains that bifurcate from their region of origin [31].

Depending on the phase with which the waves enter the South American continent, they can favor the southerly cold and dry air advection by anomalous winds at low levels, over the whole southern extreme of the continent, leading to frosts occurrence in the Wet Pampa region [31]. However, the pattern obtained with the numerical experiments and the forcings located over the Indian Ocean and the western Pacific Oceans, respectively, are not able to explain the observed anomalies in the mass and wind fields as reported by Müller et al. [30]. These authors have shown that for GF $+\sigma$ winters, the synoptic scale waves spread over the South Pacific Ocean along the subpolar and the subtropical jets (Fig. 3a).

This behavior is clearly observed during the days previous to the event. When the waves are closer to the Andes Mountains they extend meridionally and form an NW–SE oriented arc, shifting progressively towards the NE leeward to the Andes. Once the waves have crossed the Andes Mountains, they spread towards the North in such a way that is consistent with the concept of Rossby wave dispersion on a sphere [14].

In order to obtain the results observed in [30], new numerical IGCM experiments were designed that allowed further understanding of the role played by the remote forcing as source of Rossby waves finally responsible for the GF events. Müller and Ambrizzi [24] performed new simulations with the same two basic states of GF $+\sigma$ and GF $-\sigma$, considering the regions where the propagation of these waves is favored by K_s and where the zonal wind maxima are, i.e., the jets location. In a particular numerical simulations performed for GF $+\sigma$ (Fig. 3b), the patterns are consistent with the observed ones (Fig. 3a). From the results obtained by Müller and Ambrizzi [24], a conceptual model is proposed which arises from the observations, the theory, and the numerical modeling.

7 Conceptual Model

In a basic state that favors the occurrence of GF $+\sigma$, Rossby waves are generated over the tropical Indian Ocean in a mean flow characterized by a meridional gradient of zonal wind (indicated with arrows in Fig. 4). The upper-level divergent motion (dashed arrow), probably generated by a tropical source (red circle), generates perturbations in the subtropics by interfering with the positive meridional gradient of absolute vorticity region that will finally become the source of Rossby waves. This source region is upstream of the Australian subtropical jet and upstream the region where the meridional gradient of absolute vorticity is negative (stripped area). According to the linear wave theory, this region will reflect the waves, not allowing their propagation through it, so that waveguide ducts are created to the north and the south of the region. In this way, the generated Rossby waves (represented by the sequence of positive—full lines—and negative—dashed lines—meridional wind anomaly regions) propagate along the subtropical and polar waveguides (thick arrows). Near the South American continent the phases of both wave trains coincide over the region where the propagation inhibition disappears, merging into one single pattern. The deep ridge over the western part of the continent produces strong cold air advection by southerly winds (thick arrows), extending through the whole troposphere, before the occurrence of a GF event. Hence, the cold air starts crossing the Andes mountain range, initially where the heights are lower, i.e., in southern South America, generating anticyclonic vorticity. The advection of anticyclonic vorticity feeds the continental anticyclone and the maritime cyclone situated downstream over the Atlantic Ocean. The increase of the pressure gradient over the area creates a strong southerly wind component between the high and

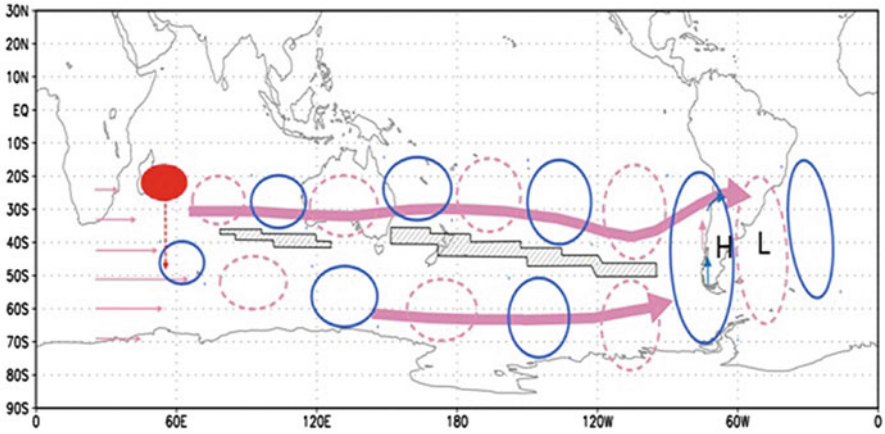


Fig. 4 Conceptual model of the physical mechanisms that favor a maximum frequency of GF occurrence during austral winters and the more persistent GF events. The Rossby wave train of meridional wind anomalies in solid (dashed) lines represents positive (negative) values. The combination of colors means that the results were obtained by: observations (pink), theory (black), and model (blue). The H and L symbols indicate the position of the low level anticyclone and cyclone, respectively. See text for details (from [24])

low pressure systems, bringing cold polar air to the region which finally leads to a significant drop of surface temperature over the eastern part of the continent.

The above described mechanism is also valid for explaining the conditions of the GF persistence as shown by Müller and Berri [26, 27] who grouped them based on the number of consecutive days during which the frost conditions persisted in the region. The persistence of the GF phenomenon depends on whether there is one or two wave trains approaching the South American continent and, in the last case, on how they merge. However, the former is not the only element that will define the GF persistence. It is at this point that linearity is diluted in the chaotic nature of any physical system and its mathematical treatment must take into account non-linear solutions such as the advection of cold air coming from the south in a sustained way, for a certain period of time, and at different levels of the atmosphere. In addition, there are other intervening factors in different scales of the atmospheric motion that should be taken into account such as the influence of the mountain range, the subsidence of the northerly flow resulting from a transverse circulation to the jet, and the confluence (diffluence) of the flow in the entry (exit) region of the subtropical jet ([26, 27], among others).

8 Conclusions

Extreme cold events have a strong impact in central-southern South America due to their spatial extension, and they are especially important when they become persistent and frequent. The impact of generalized frosts (GF) in midlatitude regions such as the Wet Pampa is more related to frost frequency and persistence with which they occur than to its intensity, in particular when they take place during cold and prolonged winters [27]. The atmospheric circulation features associated with the frequency of frosts occurrence in the Wet Pampa region have been explored [22–25, 30, 31] and, on the other hand, the dynamic conditions that favor the persistence of these events were studied by Müller and Berri [26, 27]. The results of these studies were obtained by relating observations, linear theory, and modeling, so it was possible to formulate a conceptual scheme (Fig. 4) to explain the teleconnection mechanisms that act in particular during the austral winters of maximum frequency of GF occurrence, in coincidence with those that explain the more persistent events. In this case, other scales of atmospheric motion intervene with the large scale that finally define the persistence of the GF event. Since the interaction between different scales of motion is mainly responsible for the energy exchanges in the atmosphere, new analyses were performed [32, 33], although it is necessary to profound the studies in order to further the understanding of non-linear GF phenomena.

References

1. Ambrizzi, T., Hoskins, B.J., Hsu, H.H.: Rossby wave propagation and teleconnection patterns in the austral winter. J. Atmos. Sci. **52**(21), 3661–3672 (1995)
2. Ambrizzi, T., Hoskins, B.J.: Stationary Rossby-wave propagation in a baroclinic atmosphere. Q. J. R. Meteorol. Soc. **123**, 919–928 (1997)
3. Blackburn M.: Program description for the multi-level global spectral model. Dept. of Meteorology, University of Reading (1985)
4. Bourke W.: A multi-level spectral model I: formulation and hemispheric integrations. Mon. Weather Rev. **102**, 678–701 (1974)
5. Cavalcanti, I.F.: Teleconnection patterns orographically induced in model results and from observational data in the austral winter of the Southern Hemisphere. Int. J. Climatol. **20**(10), 1191–1206 (2000)
6. Charney, J.G.: On the scale of atmospheric motions. Geofys. Publ. Oslo 17, 1–17 (1948)
7. Gan, M.A., Rao, V.B.: The influence of the Andes Cordillera on transient disturbances. Mon. Weather Rev. **122**(6), 1141–1157 (1994)
8. Garreaud, R.: Cold air incursions over subtropical South America: mean structure and dynamics. Mon. Weather Rev. **128**(7), 2544–2559 (2000)
9. Grose, W.L., Hoskins, B.J.: On the influence of orography on large-scale atmospheric flow. J. Atmos. Sci. **36**(2), 223–234 (1979)
10. Horel, J.D., Wallace, J.M.: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. Mon. Weather Rev. **109**(4), 813–829 (1981)
11. Hoskins, B.J., Ambrizzi, T.: Rossby wave propagation on a realistic longitudinally varying flow. J. Atmos. Sci. **50**(12), 1661–1671 (1993)

12. Hoskins, B.J., Karoly, D.J.: The steady linear response of a spherical atmosphere to thermal and orographic forcing. J. Atmos. Sci. **38**(6), 1179–1196 (1981)
13. Hoskins, B.J., Simmons, A.J.: A multi-layer spectral model and the semi-implicit method. Q. J. R. Meteorol. Soc. **101**(429), 637–655 (1975)
14. Hsu, H. H.: Propagation of low-level circulation features in the vicinity of mountain ranges. Mon. Weather Rev. **115**(9), 1864–1893 (1987)
15. Jin, F., Hoskins, B.J.: The direct response to tropical heating in a baroclinic atmosphere. J. Atmos. Sci. **52**(3), 307–319 (1995)
16. Karoly, D.J.: Rossby wave propagation in a barotropic atmosphere. Dyn. Atmos. Oceans **7**(2), 111–125 (1983)
17. Kiladis, G.N.: Observations of Rossby waves linked to convection over the eastern tropical Pacific. J. Atmos. Sci. **55**(3), 321–339 (1998)
18. Li, L., Nathan, T.R.: The global atmospheric response to low-frequency tropical forcing: zonally averaged basic states. J. Atmos. Sci. **51**(23), 3412–3426 (1994)
19. Loon, H., Jenne, R.L.: The zonal harmonic standing waves in the Southern Hemisphere. J. Geophys. Res. **77**(6), 992–1003 (1972)
20. Lu, R., Kim, B.J.: The climatological Rossby waves source over the STCZs in the Summer Northern Hemisphere. J. Meteorol. Soc. Jpn. **82**(2), 657–669 (2004)
21. Marengo, J.A., Ambrizzi, T., Kiladis, G., Liebmann, B.: Upper-air wave trains over the Pacific Ocean and wintertime cold surges in tropical-subtropical South America leading to Freezes in Southern and Southeastern Brazil. Theor. Appl. Climatol. **73**(3–4), 223–242 (2002)
22. Müller, G.V.: Patterns leading to extreme events in Argentina: partial and generalized frosts. Int. J. Climatol. **27**(10), 1373–1387 (2007)
23. Müller, G.V.: Temperature decrease in the extratropics of South America in response to a tropical forcing during the austral winter. Ann. Geophys. **28**(1), 1 (2010)
24. Müller, G.V., Ambrizzi, T.: Teleconnection patterns and Rossby wave propagation associated to generalized frosts over southern South America. Clim. Dyn. **29**(6), 633–645 (2007)
25. Müller, G.V., Ambrizzi, T.: Rossby wave propagation tracks in southern hemisphere mean basic flows associated to generalized frosts over southern South America. Atmósfera **23**(1), 25–35 (2010)
26. Müller, G.V., Berri, G.J.: Atmospheric circulation associated with persistent generalized frosts in central-southern South America. Mon. Weather Rev. **135**(4), 1268–1289 (2007)
27. Müller, G.V., Berri, G.J.: Atmospheric circulation associated with extreme generalized frosts persistence in central-southern South America. Clim. Dyn. **38**(5–6), 837–857 (2012)
28. Müller, G.V., Nuñez, M.N., Seluchi, M.E.: Relationship between ENSO cycles and frost events within the Pampa Humeda region. Int. J. Climatol. **20**(13), 1619–1637 (2000)
29. Müller, G.V., Compagnucci, R., Nuñez, M., Salles, A.: Surface circulation associated with frosts in the wet Pampas. Int. J. Climatol. **23**(8), 943–961 (2003)
30. Müller, G.V., Ambrizzi, T., Nuñez, M.N.: Mean atmospheric circulation leading to generalized frosts in central southern South America. Theor. Appl. Climatol. **82**(1–2), 95–112 (2005)
31. Müller, G.V., Ambrizzi, T., Ferraz, S.E.: The role of the observed tropical convection in the generation of frost events in the southern cone of South America. Ann. Geophys. **26**(6), 1379–1390 (2008)
32. Müller, G.V., Gan, M.A., Dal Piva, E., Silveira, V.P.: Energetics of wave propagation leading to cold event in tropical latitudes of South America. Clim. Dyn. **45**(1–2), 1–20 (2015)
33. Müller, G.V., Gan, M.A., Dal Piva, E. Energetics of wave propagation leading to frost events in South America: extratropical latitudes. Atmos. Sci. Lett. **18**(8), 342–348 (2017)
34. Pedlosky, J.: Geophysical Fluid Dynamics. Springer, Berlin (1987)
35. Rasmusson, E.M., Mo, K.: Linkages between 200-mb tropical and extratropical circulation anomalies during the 1986–1989 ENSO cycle. J. Clim. **6**(4), 595–616 (1993)
36. Rossby, C.G.: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. J. Mar. Res. **2**, 38–55 (1939)

37. Rossby, C.G.: On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. J. Meteorol. **2**(4), 187–204 (1945)
38. Sardeshmukh, P.D., Hoskins, B.J. The generation of global rotational flow by steady idealized tropical divergence. J. Atmos. Sci. **45**(7), 1228–1251 (1988)
39. Satyamurty, P., Dos Santos, R.P., Lems, M.A.M.: On the stationary trough generated by the Andes. Mon. Weather Rev. **108**(4), 510–520 (1980)
40. Seluchi M.E., Marengo J.A.: Tropical-midlatitude exchange of air masses during summer and winter in South America: climatic aspects and examples of intense events. Int. J. Climatol. **20**,1167–1190 (2000)
41. Simmons A.J., Burridge D.M.: An energy and angular-momentum conserving vertical finite-difference scheme and hybride vertical coordinates. Mon. Weather Rev. **109**, 758–766 (1981)
42. Tyrrell, G.C., Karoly, D.J., McBride, J.L.: Links between tropical convection and variations of the extratropical circulation during TOGA COARE. J. Atmos. Sci. **53**(18), 2735–2748 (1996)
43. Valdes, P.J., Hoskins, B.J.: Linear stationary wave simulations of the time-mean climatological flow. J. Atmos. Sci. **46**(16), 2509–2527 (1989)

Balances in the Atmosphere and Ocean: Implications for Forecasting and Reliability

Enver Ramírez, Léo Siqueira, and Rosio Camayo

1 Introduction

The spectrum of wavy motions, at the hydrostatic limit, falls in two classes: the vortical modes and the inertia-gravity waves. Vortical modes have mass and velocity fields close to geostrophy (quasi-geostrophic) and are mostly rotational with a slight divergence, whereas inertia-gravity waves (IGW) are largely divergent. Although IGWs are locally significant, in general they are radiated out from the origin region at speeds of hundreds of meters per second. Rotational modes are slower than IGWs and obey potential vorticity (PV) conservation principles. The application of rotational modes to the large-scale atmosphere and ocean was first explored by Rossby [42], from where they are also known as Rossby waves (RWs). Rossby waves are vortical modes that include geophysical effects (Earth's rotation).

Despite that for large-scale atmospheric flow, when both Rossby and Froude numbers are small, RWs and IGWs are clearly separated [12, 24, 25], they do not evolve completely independent from one another (rotational modes may have a slight divergence). Observational evidence of the dominance of quasi-geostrophic RWs has been subject of great scientific interest, as they were first identified in early studies of atmospheric planetary wave dynamics [42, 43]. It has inspired the development of efficient methods for numerical prediction and weather forecasting.

E. Ramírez (✉) · R. Camayo
Center for Weather Forecasting and Climate Studies, National Institute for Space Research, São Paulo, Brazil
e-mail: enver.ramirez@inpe.br; rosio.camayo@inpe.br

L. Siqueira
University of Miami, Rosenstiel School for Marine and Atmospheric Science, Miami, FL, USA
e-mail: lsiqueira@miami.edu

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_3

The propagation of RWs due to localized forcing is responsible for a major portion of the remotely forced variability around the planet, or teleconnections, associated with upper-air wave train-like patterns. Variations of tropical convection on intraseasonal to interannual time scales are a primary source of extratropical seasonal climate variability via the excitation of stationary RWs trains (e.g., [17]). The most well-known large-scale climate patterns associated with tropical convective anomalies, forced by interannual variability of the El Niño-Southern Oscillation in the equatorial central Pacific, are linked to the so-called Pacific-North American [15] and the Pacific-South American [19] patterns, in the Northern and Southern Hemispheres, respectively. On intraseasonal time scales, the Madden–Julian oscillation (MJO) and its induced convective heating has been recognized to significantly perturb geopotential height in the extratropics through RW teleconnections to the tropics in both modeling and observational studies [11, 18, 29]. The extratropical impacts induced by these RW trains can affect the local climate through a diverse set of phenomena including synoptic weather disturbances, jet stream meanderings, blocking events, and temperature extremes in the Euro-Atlantic region and in South America [6, 8, 41], the North Atlantic oscillation (e.g., [3, 9, 23]), U.S. West Coast precipitation events [14], extreme rainfall events in southeastern Brazil [2], among other impacts.

In short, previous studies suggest the existence of a variety of RWs, and interactions, with an abundance of impacts around the planet. The behavior of RWs can be studied from the theoretical viewpoint and by employing simple models and analytic techniques which make use of vorticity equation models or ray tracing to isolate the effects of RWs to the exclusion of all other processes.

2 Theoretical Analysis of Rossby Waves

2.1 Barotropic Equations

If we consider a shallow layer of incompressible and inviscid fluid flow on the rotating planet, then the dynamics can be described by the two-dimensional rotating Euler equations [25]

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v_H} + f\hat{k} \times \mathbf{v_H} = -\frac{1}{\rho_0}\nabla p, \tag{1a}$$

$$\nabla \cdot \mathbf{v_H} = 0, \tag{1b}$$

where $\frac{\mathrm{d}}{\mathrm{d}t}$ is the total time derivative operator and when applied to each scalar component A of the vector $\mathbf{v_H}$ is given by $\frac{\mathrm{d}}{\mathrm{d}t}A = \frac{\partial}{\partial t}A + (\mathbf{v} \cdot \nabla)A$, f is the Coriolis parameter, $\hat{k}$ is the unit vertical vector, ∇p is the pressure gradient, and $\mathbf{v_H}$ the horizontal component of the fluid flow. Equation (1a) represents the fluid motion set by an external pressure gradient, while (1b) represents the incompressibility of the

flow. The geophysical extension is given by the Coriolis parameter, which represents the effects introduced by the rotation and the curvature of the reference system. Furthermore, under the beta-plane approximation for the Coriolis parameter:

$$f = f_0 + \beta y, \tag{2}$$

a linear tangent plane is considered for f and a rectangular coordinate system can be used.

2.2 Barotropic Vorticity Equation

RW can be inferred by considering the two-dimensional Euler flow (1). By requesting the incompressibility condition, the family of sound waves is removed from the set of possible solutions, yet the Coriolis parameter provides a restoring mechanism leading to the nontrivial, the so-called RW solution. To promptly set the equations describing RWs, the non-divergence condition $\nabla \cdot \mathbf{v_H} = 0$ allows us to describe the problem in stream-function formulation ψ related to the winds by $\mathbf{v_H} = \hat{k} \times \nabla\psi$. Thus, taking the curl of (1), the nonlinear, non-divergent Barotropic Rossby waves are described by

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = 0, \tag{3}$$

where ψ is a stream function for the flow, $\zeta = \nabla^2\psi = \hat{k} \cdot (\nabla \times \mathbf{v_H})$ is the relative vorticity, β the meridional derivative of the Coriolis parameter f, and $\zeta + f$ is the absolute vorticity. The nonlinear contribution in (3) is clearly due to the Jacobian

$$J(\psi, \zeta) = \left[\frac{\partial \psi}{\partial x} \frac{\partial \zeta}{\partial y} - \frac{\partial \psi}{\partial y} \frac{\partial \zeta}{\partial x} \right] \tag{4}$$

which represents the advection of the relative vorticity ζ.

2.3 Rossby Waves

Assuming a basic zonal flow satisfying (3), a plane wave solution was proposed by Rossby [42]

$$\psi = \mathcal{R}_e \left[\mathcal{A}_0 \mathrm{e}^{\mathrm{i}(kx+ly-\sigma t+\phi)} \right] = \mathcal{A}_0 \cos\theta, \tag{5}$$

where $K = (k, l)$ is the wavenumber vector, σ the frequency of the oscillation, $\theta = kx + ly - \sigma t + \phi$ the wave phase, and $\mathcal{A}_0$ its amplitude. Thus, the following relations can be easily obtained:

$$\nabla^2 \psi = \left(k^2 + l^2\right) \mathcal{R}_e \left[\mathcal{A}_0 \mathrm{e}^{\mathrm{i}(kx+ly-\sigma t+\phi)}\right] \tag{6a}$$

$$\partial_t \psi = -\sigma \, \mathcal{R}_e \left[\mathcal{A}_0 \mathrm{e}^{\mathrm{i}(kx+ly-\sigma t+\phi)}\right] \tag{6b}$$

$$\partial_x \psi = k \, \mathcal{R}_e \left[\mathcal{A}_0 \mathrm{e}^{\mathrm{i}(kx+ly-\sigma t+\phi)}\right]. \tag{6c}$$

From the linearization of (3) and considering both the relations in (6) and that the flow consists of a time-independent component (basic state) plus a perturbation, the simplified Rossby wave dispersion is obtained

$$\sigma = -\frac{\beta k}{k^2 + l^2}. \tag{7}$$

It becomes clear from (7) that the physical restoring mechanism for the Rossby wave is the Coriolis parameter variation β. In the simplest case of barotropic flow on the β-plane, the potential vorticity is the absolute vorticity, $\beta y + \zeta$. Suppose that some initial perturbation displaces a material line η along a line of constant latitude to a perturbed position at $\eta(t = 0)$, as in Fig. 1. The conservation of potential vorticity, in either hemispheres, then leads to the production of positive (negative) relative vorticity for a southward (northward) displacement. The perturbation vorticity field and its associated velocity field then advects the chain of fluid parcels southward west of the vorticity maximum and northward west of the vorticity minimum. Thus, the fluid parcels oscillate back and forth about their equilibrium latitude, and the phase of the wave propagates westward. The key ingredient is therefore a basic state potential vorticity gradient, such as that provided by the change of f with latitude.

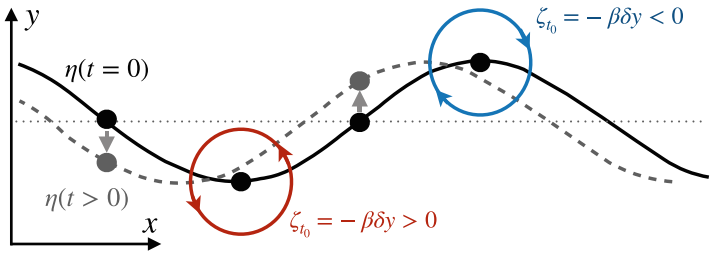


Fig. 1 Schematics of a two-dimensional Rossby wave mechanism. A chain of fluid parcels is initially displaced to the solid line $\eta(t = 0)$, leading to the production of relative vorticity. Colored circles show the velocity field and the material line evolves into the dashed gray line $\eta(t > 0)$. The fluid parcels oscillate back and forth about their equilibrium latitude (dotted straight line), and the phase of the wave propagates westward

Similar effects can be achieved by topography, where the topographic slope assumes the role of the β parameter and the resulting pattern is a topographic Rossby wave.

2.4 Rossby Wave as an Exact Solution to the Barotropic Vorticity Equation

For a geostrophically balanced flow, horizontally non-divergent, it results that as a consequence of $\mathbf{v_g} = \hat{k} \times \nabla\psi$, the horizontal velocity is parallel to the stream-function (i.e., perpendicular to the stream-function gradient). Therefore, a solution made of only one Rossby wave is an exact solution of the nonlinear equation, as a single Rossby wave cannot advect itself. Thus, the Jacobian contribution is null, $J(\psi, \zeta) = 0$, and the nonlinear equation reduces to the linear equation.

Physically, the Jacobian represents interaction between modes of the system, this means that a unique Rossby wave cannot interact with itself. However if more than one Rossby wave coexists, the velocities generated by one of the participating Rossby wave may advect the stream-function lines of the others, the field is modified and new waves can be created, the younger waves may interact with the older modes creating further new waves and eventually the field is completely distorted.

As any Rossby mode is a possible solution of the nonlinear Eq. (3), what ultimately determines how many modes are initially in the system is their initial energy distribution. For the general nonlinear case, the evolution of the system from this initial state is determined from both the initial mode distribution and the available nonlinear terms. These effects shall be discussed in the following sections.

2.5 Potential Vorticity Equation

A more complete set of equations for the potential vorticity conservation must consider the contribution of the planetary vorticity f as well as a contribution due to the oscillations of the free surface height of the fluid flow. These effects, for instance, are represented by $\mathrm{d}f/\mathrm{d}t = \beta\ (\partial\psi/\partial x)$ and $F\psi$, where $F = (L/L_R)^2$ and L_R the Rossby radius of deformation radius, defined in terms of the mean thickness $\bar{H}_0$, the gravity g, and the Coriolis parameter f

$$L_R = \frac{\sqrt{g\bar{H}_0}}{f}. \tag{8}$$

In practice, the re-inclusion of the free surface height oscillations restores the system gravitational adjustment and in the case of stratification, the buoyancy adjustment. In both cases, gravity waves are allowed in the system. In addition,

the free surface oscillation on a barotropic fluid flow allows for external waves, in this case the Lamb waves. This also implies that higher order divergence is present in the system.

Thus, with the aforementioned considerations, the potential vorticity $q = \zeta - F\psi + f$ shall be conserved, following

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right) q = 0. \tag{9}$$

3 Nonlinear Resonant Interactions in the Barotropic Case

3.1 Archetypal Model and Space-Time Scales Involved

In the previous sections, the nonlinear terms, due to the Jacobian in (3), allow the coupling between the wave modes. Here, the wind field of one of the waves can advect the stream-function of the others and the distortion of the stream-function is conductive to modification in the wind field. In turn, the modified wind field may interact with the stream-function of the advecting wave and, consequently a nonlinear coupling is established.

To better understand this process, consider either (3) or (9). As $J(\psi, \zeta - F\psi) = J(\psi, \zeta)$, $\frac{\partial f}{\partial t} = 0$, and $\frac{\partial f}{\partial y} = \beta$, then the archetypal equation for dispersive Rossby waves results in

$$\frac{\partial}{\partial t}(\zeta - F\psi) + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = 0. \tag{10}$$

In order to perform a dimensional analysis in (10), one must identify the scales involved. Thus, for a spatially isotropic phenomenon $L_x = L_y = L$ resulting that only L defines the typical spatial scale (c.f. [35, 36]), now if U is the characteristic velocity, then L/U is the characteristic advective time. In addition, as long as β is essential for the existence of Rossby waves, $1/(\beta L)$ defines the characteristic Rossby wave period. The dimensional timescale t can be defined using either of the above time scales $t = (\beta L)^{-1}\tilde{t} = (L/U)t_*$, from where the relationship between the dimensionless timescale parameters is given by

$$\tilde{t} = (\beta L^2/U)t_* = \beta_* t_*. \tag{11}$$

Thus, using $(\tilde{t}, t_*)$ a multi-time scale treatment can also be used. The convention used in this overview is that the frequency associated to $\tilde{t}$ is $\mathcal{O}[1]$ and consequently, the frequency associated with t_* is $\mathcal{O}[1/\beta_*]$. The fast time variable is therefore represented by $\tilde{t}$, whereas t_* represents the slow time variable. The reader can find further references on multi-time scale in [21, 27, 33, 36]. However, at this point and under the single-time scale consideration, the dimensionless archetypal model for barotropic dispersive Rossby wave is

$$\frac{\partial}{\partial \tilde{t}}(\zeta - F\psi) + \frac{1}{\beta_*} J(\psi, \zeta) + \frac{\partial \psi}{\partial x} = 0. \tag{12}$$

3.2 Weakly Nonlinearity: Searching for Approximate Solutions

For a rapidly rotating Rossby wave, $(\beta L)^{-1}$ must be small and in the limit $\beta_* = (\beta L^2/U) \gg 1$, the characteristic Rossby wave period $(\beta L)^{-1}$ is shorter than the advective time period (L/U). Consequently, the nonlinear term is smaller than the other terms, and to a good first approximation, the linear solutions are dominant. However, nonlinear interactions will still occur, as $1/\beta_*$ is arbitrarily small but finite. Physically, $1/\beta_*$ can be thought as a parameter relating the Rossby period and the advective period, but also, it must be said that it relates the relative vorticity gradient to the planetary vorticity gradient, such that:

$$\frac{1}{\beta_*} = \frac{\text{Rossby period}}{\text{Advective period}} = \frac{\text{Relative vort. gradient}}{\text{Planetary vort. gradient}}. \tag{13}$$

Now, since (12) is expressed in terms of the small parameter $1/\beta_*$, the answer may, in principle, be expanded in the same way (this is known as an asymptotic expansion). For the present case this is given by

$$\psi(\vec{x}, \tilde{t}, \beta_*) = \psi_0(\vec{x}, \tilde{t}) + \frac{1}{\beta_*}\psi_1(\vec{x}, \tilde{t}) + \frac{1}{\beta_*^2}\psi_2(\vec{x}, \tilde{t}) + \cdots + \mathcal{O}\left[\frac{1}{\beta_*^n}\right]. \tag{14}$$

The first term of this asymptotic expansion is the linear eigenmodes, characterizing the dominant role of them. Furthermore, the expansion (14) may be thought as successive and finer corrections to the streamlines (ψ) considering the interpretation of the right-hand side in (13). This expansion also serves for the other variables of the system since both $\zeta = \nabla^2\psi$ (vorticity) and $\mathbf{v} = \hat{k} \times \nabla\psi$ (winds).

3.3 Sequence of Easier to Solve Problems

The contribution of each new term in (14) is parameterized by the small number $1/\beta_*$. Thus, by plugging (14) into (12) and associating equal powers of $1/\beta_*$, a sequence of balanced problems is obtained. In general, one goes from a hard to solve mathematical problem to a sequence of relatively simpler problems and in order to obtain the solution, each term of the sequence $\{\psi_j(\vec{x}, \tilde{t})\}$ for $j = \{0, 1, 2, \ldots, n\}$ must be found. The lowest order $\mathcal{O}[1]$ balanced problem is simply the linear equation for Rossby waves

$$\frac{\partial}{\partial \tilde{t}}\left(\nabla^2\psi_0 - F\psi_0\right) + \frac{\partial \psi_0}{\partial x} = 0 \tag{15}$$

with solutions similar to that described by Eq. (5) in Sect. 2.3. The fact that ψ_0 satisfies (15) confirms that the leading order terms of the sequence are the linear modes of the barotropic vorticity equation.

The next problem to be solved is the $\mathcal{O}[1/\beta*]$ and it can be shown that

$$\frac{\partial}{\partial \tilde{t}}\left(\nabla^2\psi_1 - F\psi_1\right) + \frac{\partial \psi_1}{\partial x} = J\left(\nabla^2\psi_0, \psi_0\right). \tag{16}$$

A very important point to note here is that (16) is a linear, forced problem. The left-hand side of the problem is precisely the linear Rossby wave equation, similar to (15), whereas the right-hand side is the Jacobian formed by the solutions of the $\mathcal{O}[1]$ problem. Consequently, the nonlinear terms computed from the $\mathcal{O}[1]$ problem are the forcing for the small $\mathcal{O}[1/\beta_*]$ contributions. Here we stress that the analytic solutions for ψ_0 are already know from (15), then the forcing term for the $\mathcal{O}[1/\beta_*]$ is in principle computable. Instead of solving a hard problem, we solve a sequence of less complex to solve problems.

Employing the plane wave solution in (5) for two different waves labeled m and n, and considering that a Rossby wave cannot advect itself, the two waves have a non-null contribution from the Jacobian to the forcing of the problem (16)

$$\begin{aligned}\frac{\partial}{\partial \tilde{t}}&\left(\nabla^2\psi_1 - F\psi_1\right) + \frac{\partial \psi_1}{\partial x} \\ &= \sum_m \sum_n \frac{\mathcal{A}_m \mathcal{A}_n}{2}\left(K_m^2 - K_n^2\right)(k_n l_m - k_m l_n)\left(\frac{\cos(\theta_m + \theta_n) - \cos(\theta_m - \theta_n)}{2}\right),\end{aligned} \tag{17}$$

where trigonometric identities to transform the product of sine and cosine into a sum were used to replace the $\sin\theta$ functions resulting from the derivatives of the Jacobian operator to the $\cos\theta$ functions displayed in (17).

It shall also be noted that in addition to the constraint that the waves m and n must be different for the interaction to occur, two more conditions are evident for the forcing to be different from zero: $K_m^2 \neq K_n^2$ and $k_n l_m - k_m l_n \neq 0$. The second condition means that (k_m, l_m) must not be parallel to (k_n, l_n). By avoiding the above described situations, the forcing of ψ_1 then oscillates with a phase equal to the sum and differences of the modes m and n. Furthermore, as (17) is a linear forced problem, each forcing term can be computed separately and then linear superposition of the resulting ψ_1 can be used to form the $\mathcal{O}[1/\beta_*]$ solution.

For the sake of consistency, we define the interaction coefficient

$$N_{mn} = \frac{\left(K_m^2 - K_n^2\right)(k_n l_m - k_m l_n)}{2} \tag{18}$$

and each forcing can be considered separately since the problem is linear. Rather than studying the $\sum_m \sum_n$, we restrict our attention to only a set of two waves. Since the wave phase $\theta = kx + ly - \sigma t + \phi$, we have

$$\theta_m \pm \theta_n = (k_m \pm k_n)x + (l_m \pm l_n)y - (\sigma_m \pm \sigma_n)\tilde{t} + (\phi_m \pm \phi_n). \tag{19}$$

Therefore, the horizontal structure of the forcing is described by $K_{mn} = K_m \pm K_n$ and the frequency by $\omega_{mn} = \sigma_m \pm \sigma_n$. The forcing can then be Fourier expanded using $\mathrm{e}^{\mathrm{i}(k_{mn}x + l_{mn}y - \omega_{mn}\tilde{t} + \phi_{mn})}$, ψ_1 is also Fourier expanded $\psi_1 = \mathcal{A}_p \mathrm{e}^{\mathrm{i}(k_p x + l_p y - \sigma_p \tilde{t} + \theta_p)}$ to yield

$$\frac{d\mathcal{A}_p}{d\tilde{t}} = \frac{1}{2}\mathcal{A}_m \mathcal{A}_n \left(\frac{N_{mn}}{K_p^2 + F}\right) \mathrm{e}^{\mathrm{i}(\sigma_p - \omega_{mn})\tilde{t}} = F_s \mathrm{e}^{\mathrm{i}(\sigma_p - \omega_{mn})\tilde{t}}, \tag{20}$$

where $F_s = \frac{1}{2}\mathcal{A}_m \mathcal{A}_n \left(\frac{N_{mn}}{K_p^2 + F}\right)$ is the forcing amplitude. Solutions can be written as

$$\mathcal{A}_p = \left\{ \begin{array}{ll} F_s \frac{\mathrm{e}^{\mathrm{i}(\sigma_p - \omega_{mn})\tilde{t}}}{(\sigma_p - \omega_{mn})}, & \text{for } (\sigma_p - \omega_{mn}) \neq 0 \\ F_s \tilde{t}, & \text{for } (\sigma_p - \omega_{mn}) = 0 \end{array} \right\}. \tag{21}$$

Thus, whenever the forcing frequency does not coincide with the frequency of the mode ψ_1 (i.e., $\sigma_p \neq \omega_{mn}$), the amplitude $\mathcal{A}_p$ oscillates with time. However, if the forcing frequency coincides with that of the ψ_1 (i.e., $\sigma_p = \omega_{mn}$), resonance occurs and the amplitude of ψ_1 grows with time and in the limit $\tilde{t} \to t_*$, the amplitude $\mathcal{A}_p$ grows and eventually ψ_1 becomes comparable to ψ_0. In such a case, the expansion (14) should be no longer valid.

The condition for resonance to occur without invalidating (14) is known as the solvability condition (see [21]), so that if any growth occurs, it should be warranted to be sub-linear

$$\lim_{\tilde{t} \to +\infty} \left(\frac{\psi_1(\vec{x}, \tilde{t})}{|\tilde{t} + 1|}\right) = 0. \tag{22}$$

3.4 *Topographic Resonance in the Barotropic Flow*

The notion of topographic instability of a flow otherwise stable was developed by work as earlier as [4, 5, 13, 32]. Let's consider a purely steady, meridionally sheared, zonal flow $\bar{U}(y)$ driven by an external source, which in terms of vorticity

is given by $Q(y)$. Instead of having a free surface $F\psi$, the flow is subject to a bottom topography η_B. Furthermore, because of the externally forced nature of the flow a Rayleigh type dissipation is included $-r\nabla^2\psi$. Under such a condition, the barotropic vorticity (9) is given by

$$\Big(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\Big)(\nabla^2\psi + f + \eta_B) = -r\nabla^2\psi + Q(y), \tag{23}$$

where $\mathbf{v} = \hat{k}\times\nabla\psi$, $\partial_t f = \partial_x f = 0$, and $\partial_y f = \beta$. In the absence of topography, the purely steady zonal flow $\bar{U}(y)$ driven by $Q(y)$ is warranted if

$$-r\frac{\mathrm{d}\bar{U}}{\mathrm{d}y} = Q(y). \tag{24}$$

Thus, the stream function of the problem including both, the steady and transient contributions, is given by

$$\psi = -\int \bar{U}(y)\mathrm{d}y + \epsilon\phi, \tag{25}$$

and the vorticity may be written accordingly as

$$\nabla^2\psi = -\frac{\mathrm{d}}{\mathrm{d}y}\bar{U}(y) + \epsilon\nabla^2\phi = \frac{Q(y)}{r} + \epsilon\nabla^2\phi. \tag{26}$$

Inserting (26) in (23) results in

$$\begin{aligned}\Big(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\Big)\nabla^2\phi + \frac{\partial\phi}{\partial x}\Big(\beta - \frac{\mathrm{d}^2\bar{U}}{\mathrm{d}y^2}\Big) + \epsilon J(\phi,\nabla^2\phi) + J(\phi,\eta_B)\\ = -r\nabla^2\phi - \frac{\bar{U}}{\epsilon}\frac{\partial\eta_B}{\partial x}.\end{aligned} \tag{27}$$

Although nothing was said for the topography up to this point, it is the interaction of the zonal gradient of the topography and the steady mean flow that introduces a strong forcing $\mathcal{O}[\epsilon^{-1}]$ for the transient wavy motion (27). In addition, to ensure dominance of the normal-mode solutions for the leading order homogeneous problem, the stationary condition might be asked which results immediately in the topographic dependent stationary Rossby wave number K_{ST}

$$K_{\mathrm{ST}} = \sqrt{\frac{\beta^{\dagger\dagger}}{\bar{U}}}, \tag{28}$$

where

$$\beta^{\dagger\dagger} = \beta + \frac{\partial\eta_B}{\partial y} - \frac{\mathrm{d}^2\bar{U}}{\mathrm{d}y^2}. \tag{29}$$

In simple arguments, the bottom slope modifies β, which is a way that topography induces Rossby waves. Although certain heavy simplifications are required to proceed with the analytical treatment, the physical mechanisms for topographic resonance are likely to transcend these limitations [32]. The first of such heavy simplifications toward an analytic treatment is the consideration of the topography's shape. When studying the same flow but contained in a channel, [32] used a topography shape given by

$$\eta_B = h_0 \frac{\mathrm{e}^{ikx}}{2} \sin ly + C.C., \tag{30}$$

where $C.C.$ stands for the complex conjugated of the preceding term. Resonance might occur for values of k, l, $\bar{U}$, and β that enables stationary normal-mode solutions. In their study, [32] considered the simplest case for $\bar{U}$ being independent of y and the criteria for stationarity leading to near resonance conditions is achieved when

$$\bar{U} = \frac{\beta}{k^2 + l^2} + \Delta, \text{ with } \Delta \ll 1. \tag{31}$$

Although the reference time scale used is the advective $L/\bar{U}$, its development is a product of the nonlinear interactions occurring on the longer time scale $L/(\bar{U}\Delta^{1/2}\epsilon)$, which defines the slow time scale as estimated by Charney and Devore [4]

$$\tau = (\Delta^{1/2}\epsilon)\, t. \tag{32}$$

Taking into consideration the near resonance condition, the independence of $\bar{U}$ with respect to y, and the slow evolution scale τ, it is possible to obtain a representative equation

$$\begin{aligned} &\left[\Delta^{1/2}\epsilon \frac{\partial}{\partial \tau} + \left(\frac{\beta}{\left(k^2 + l^2\right)} + \Delta\right) \frac{\partial}{\partial x}\right] \nabla^2 \phi + \beta \frac{\partial \phi}{\partial x} + \epsilon J\left(\phi, \nabla^2 \phi\right) \\ &\quad + J(\phi, \eta_B) + r\nabla^2 \phi = -\epsilon^{-1} \left(\frac{\beta}{\left(k^2 + l^2\right)} + \Delta\right) \frac{\partial \eta_B}{\partial x}. \end{aligned} \tag{33}$$

Considering $r = \mathcal{O}[\epsilon^2]$ and with the aid of the asymptotic expansion for ϕ

$$\phi = \phi_0 + \epsilon\phi_1 + \epsilon^2 \phi_2 + \cdots. \tag{34}$$

The leading order problem $\mathcal{O}[1]$ of the expansion is subjected to

$$\frac{\beta}{(k^2 + l^2)} \frac{\partial}{\partial x} \left(\nabla^2 \phi_0 + (k^2 + l^2)\phi_0\right) = 0 \tag{35}$$

whose solution is given by

$$\phi_0 = \mathcal{A}(\tau)\,\frac{e^{ikx}}{2}\,\sin ly. \tag{36}$$

The topography imprinting over the leading order transient stream function ϕ_0 is clearly noted. We note that $\mathcal{A}(\tau)$ represents the slow amplitude evolution that has to be determined.

4 Increasing Complexity Through Vertical Interactions

4.1 The Barotropic–Baroclinic Interactions

In the tropics, the near field response to a localized heating anomaly is preferentially in terms of baroclinic modes, and the peak response to the tropical heating is associated with the first baroclinic mode. This mode is characterized by a single phase change between lower and upper levels in the flow fields; the second baroclinic structure is associated with two phase changes in the vertical, while the third baroclinic mode shows three phase changes and so on [7, 37, 44]. Despite that tropical baroclinic structures tend to stay trapped in latitude, they are efficient in radiating energy upward having an important role for the upper layers budget. Tropical baroclinic modes are usually described by the equatorial β-plane dynamics [28]. Highly damped shallow-water models [10, 28, 45] can actually give a reasonable first approximation to the low-level near wind field response.

The external or barotropic structures, on the other hand, are associated with the far field response to the tropical heating. The barotropic signal dominates the teleconnections far from the source in mid- and high-latitudes. The barotropic modes are less meridionally trapped, but are vertically confined [7, 44]. Thus, purely barotropic models have been extensively used to study the teleconnection response at midlatitudes [1, 17]. Earlier studies have emphasized the study of barotropic–baroclinic structures as a way to understand the tropical–extratropical interactions. However, the localized heating anomaly does not directly force a barotropic signal as aforementioned. Moreover, it has been suggested that tropical heating is inefficient to promote the vertical mode interactions. Then, vertical and horizontal mean shear flows have been introduced as a mechanism to promote the energy exchange.

Figure 2 displays observational evidence of an equatorial RW shown by pairs of twin cyclonic vortices forming an equatorial (baroclinic) wave train in the 850 hPa tropical winds for 7 October 2002. It exemplifies a case of long equatorial RW for which the zonal spatial scale (about 100° of longitude) is much larger than the meridional scale, with the former typically ranging from about 4000 to 10,000 km. The meridional distance scale of this wave is much shorter, about 20° of latitude, as the centers are located roughly 10° either side of the equator. The far field response

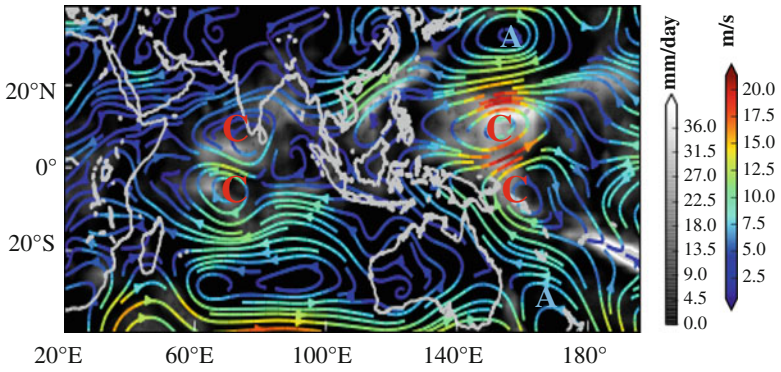


Fig. 2 Observational evidence of equatorial Rossby waves shown by pairs of twin cyclonic vortices (C) forming an equatorial wave train (baroclinic) in the 850 hPa tropical winds (m/s) for 7 October 2002 from the NCEP-NCAR Reanalysis Project—Climate Data Assimilation System I. Shading shows the rainfall field in mm/day. The figure shows one of the longest wavelengths possible, for which the zonal spatial scale (about 100° of longitude) is much larger than the meridional scale. The meridional distance scale of this wave is much shorter, about 20° of latitude, as the centers are located roughly 10° either side of the equator. The far field response is comprised by midlatitude (barotropic) wave trains

is comprised by midlatitude (barotropic) wave trains. In the following sections, we introduce the baroclinic vorticity equation, focusing on the new mechanisms that distinguish it from the barotropic case and immediately move to discuss basic models for the barotropic–baroclinic interactions.

4.2 General Baroclinic Vorticity

Similarly to the barotropic case, an equation for the baroclinic vorticity can be obtained. Thus, following [22], the baroclinic perturbation vorticity equation is given by

$$\frac{\partial \zeta'}{\partial t} = -\mathbf{v}' \cdot \nabla(\bar{\zeta} + f) - \bar{\mathbf{v}} \cdot \nabla \zeta' - (\mathbf{v}' \cdot \nabla \zeta')' + \left[(f + \zeta)\frac{\partial \omega}{\partial p}\right]' - \left(\hat{k} \cdot \nabla \omega \times \frac{\partial \vec{v}}{\partial p}\right)' + \nu, \tag{37}$$

where $\omega = \mathrm{d}p/\mathrm{d}t$ is the vertical velocity in pressure coordinates. The first three terms on the right-hand side represent the transport of vorticity, similar to those obtained for the barotropic case, and include the physical restoring mechanism for the RWs, evident if the β-plane approximation is used.

The fourth term represents the effects of the convergence or divergence in the baroclinic vortex and corresponds to a rate of change of the vorticity related to the stretching/shrinking of the vortex filaments. The fifth term represents the generation of vorticity related to the tilting of the vertical structures by a vertically non-uniform wind field, whereas the last term represents other source/sink terms not considered in the equation. In this term, production of vorticity by the horizontal tilt of isobaric (equal pressure) with respect to isentropic (equal density) surfaces is included, also dissipation of vorticity by interactions with the lower atmospheric layers (e.g., the atmospheric boundary layer) is represented, or even the effects of the latent heat release or radiative forcing in producing vorticity can be considered. Here, we note that in Eq. (37) new mechanisms that are not present in the barotropic case appear and in the following sections we will discuss in a simplified framework the interactions between the barotropic and baroclinic structures.

4.3 Two-Layer Equatorial β-Plane and Barotropic Vorticity

Let us consider, under the equatorial β-plane approximation, a basic two-layer model for the barotropic–baroclinic interactions. The barotropic or external mode represented by variables without phase changes along the vertical is given by $(\mathbf{v_0}, p_0)$, whereas the first baroclinic or internal mode (one phase change along the vertical) is represented by $(\mathbf{v_1}, p_1)$. Thus, following [26] the two-layer equatorial β-plane equations are given by

$$\left(\frac{\partial}{\partial t} + \mathbf{v_0}\cdot\nabla\right)\mathbf{v_0} + y\hat{k}\times\mathbf{v_0} + \mathbf{v_1}\nabla\cdot\mathbf{v_1} + (\mathbf{v_1}\cdot\nabla)\mathbf{v_1} = -\nabla p_0, \tag{38a}$$

$$\nabla\cdot\mathbf{v_0} = 0, \tag{38b}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v_0}\cdot\nabla\right)\mathbf{v_1} + y\hat{k}\times\mathbf{v_1} + (\mathbf{v_1}\cdot\nabla)\mathbf{v_0} = -\nabla p_1, \tag{38c}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v_0}\cdot\nabla\right)p_1 + \nabla\cdot\mathbf{v_1} = 0, \tag{38d}$$

where all variables are two-dimensional (x, y) and all vector differentiation is carried out in the horizontal. We note that the common transport operator in (38) advects the flow using the barotropic wind $\mathbf{v_0}$. Here, the role of tropical heating, radiative damping, and boundary layer drag is omitted in order to solely highlight the interactions through the dynamics.

The link with the barotropic dynamics in (3) is achieved by the use of the stream-function formulation, where $\zeta = \nabla^2\psi = \hat{k}\cdot(\nabla\times\mathbf{v_0})$. Then, the barotropic part of system (i.e., 38a–b) is written as

$$\frac{\partial\zeta}{\partial t} + J(\psi,\zeta) + \frac{\partial\psi}{\partial x} + \nabla\cdot\left(-\frac{\partial\mathbf{v_1}u_1}{\partial y} + \frac{\partial\mathbf{v_1}v_1}{\partial x}\right) = 0. \tag{39}$$

It must be noted that due to the coupling between external–internal structures, a new term $\nabla \cdot (-\partial_y(\mathbf{v_1}u_1) + \partial_x(\mathbf{v_1}v_1))$ appears in (39) compared to (3). Furthermore, since here there is a single baroclinic structure ($\mathbf{v_1}$, p_1), the equatorial β-plane wave theory of [28] can be applied and consequently, a family of waves (Kelvin, Rossby, mixed Rossby-gravity, and inertia-gravity) is being represented.

The baroclinic waves contained in (38c–d) must have an evanescent character for y values tending to the equatorial Rossby deformation radius $\sqrt{C/\beta}$, a necessary condition for the equatorial wave trapping. Moreover, as discussed in [35, 38–40] nonlinear interactions are allowed for equatorial waves. Therefore, it is necessary to verify whether these interactions might also be relevant for the barotropic–baroclinic interactions.

4.4 The Barotropic–Baroclinic Interactions in the Two-Layers Model

In order to gain insight into the interaction between midlatitude barotropic RWs and equatorial baroclinic RWs, [26] developed a set of simplified dynamic equations in the presence of mean shear. In their work, the zonal spatial scales of the equatorial baroclinic RWs are much larger than the meridional scales (i.e., $L_x \gg L_y$), therefore constituting the use of anisotropic scalings to (38). Furthermore, [26] gave a special attention to the selection of barotropic modes with significant midlatitude projection, i.e., those with smaller meridional index. After a suitable weakly nonlinear asymptotic expansion, a reduced model was obtained

$$\partial_\tau A - D\partial_x^3 A + B\partial_x A + A\partial_x B = 0, \tag{40a}$$

$$\partial_\tau B - \partial_x^3 B + A\partial_x A = 0, \tag{40b}$$

with A describing the baroclinic flow, B the barotropic flow, $\tau = \epsilon t$ the long time variable, and $\epsilon = 0.1$. The reduced model (40) comprises a set of coupled Korteweg–de Vries (KdV)-like equations occurring through the interaction of resonant wave trains at long wavelengths, considered in many aspects as novel equations in the applied mathematics literature.

A subsequent separation of zonal mean ($\bar{B}$, $\bar{A}$) and transient fluctuation (B', A') of both barotropic and baroclinic flows allowed to obtain equations for the transient barotropic–baroclinic interactions

$$\partial_\tau\Big(\int \frac{B'^2}{2}dx\Big) = -\partial_\tau\Big(\int \frac{A'^2}{2}dx\Big) \tag{41a}$$

$$= \frac{1}{2}\int (A')^2\partial_x B'dx + \frac{1}{2}\int \bar{A}(A'\partial_x B' - B'\partial_x A')dx. \tag{41b}$$

From (41), a non-trivial energy flow between transient waves is noted. The first term on the right-hand side $(A')^2\partial_x B'$ represents the interaction of baroclinic–barotropic transient waves, while the role of mean baroclinic shear ($\bar{A}$) to stimulate the energy flow between the barotropic–baroclinic transient modes is highlighted in the second term $\bar{A}(A'\partial_x B' - B'\partial_x A')$. Accordingly, the role of (vertical and general horizontal) shears for the barotropic–baroclinic interactions has been pointed by several researchers [16, 20, 45–47]. Majda and Biello [26] interpret the transient interaction term as a type of westerly wind burst (WWB) mechanism, as the gradients of the barotropic wave locally remove energy from itself and deposit into the baroclinic wave. An interesting discussion on the vertical structure of the WWB being either barotropic (related to synoptic transients) or baroclinic (related to the MJO) is presented in [48].

In addition, [35] studied possible interactions of a general family of ultra-long equatorial waves (i.e., ultra-long version of the [28] family) with the general family of classical equatorial waves. They found that resonant interactions between anisotropic equatorial waves and classical (isotropic) equatorial waves are possible, which extend the range of possible interactions. Thus, the study of [26] might be extended, involving other wave interactions even within their ultra-long framework. In the tropics, general lateral forcing by extratropical waves is believed to be a mechanism for the excitation of variability including low frequency variability at the intraseasonal [48] or even at the interannual [30] time scales. [31] studied the important role of midlatitude waves entering the equatorial region for the moisture budget.

Raupp [37] developed a similar study, but using as the starting point the dry (adiabatic) primitive equations and applying isotropic scalings. Primitive equations allow the coupling between the momentum equations and the thermodynamics, whereas the isotropic scalings retain the family of horizontal modes for each fixed vertical structure. They obtained an asymptotic reduced model that governs the weakly nonlinear amplitude of the waves in a given resonant triad (a closed system formed by the interactions of three waves). The reduced equations describe the evolution of the coupled amplitude (energy) of the triad members. Consequently, it determines the ways by which the triad members interact. Thus, in addition to the interaction of modes of a given vertical structure, interactions between vertical modes are also possible, and by using a more complex set of equations, interactions with the thermodynamic and even with water phase changes are also possible.

5 Scale Interactions

The rotating shallow water equations give us a simplified model for the large-scale atmospheric circulation. In the appropriate regime and far from the equatorial region, it gives a convenient scale separation between the vortical and the IGW modes. Traditionally, the balance problem is related to the dynamics of the slow modes that evolve and emerge even in the presence of unbalanced initial data

[12, 24, 25]. Balanced modes contribute to the predictability and the fast modes are generally filtered or controlled. However, geophysical vortical motions are also able to generate IGWs and although the source of IGWs by RWs is weak, it can be accumulated to a significant amount over sufficiently long time scales [12]. So, in this section a brief discussion about a few mechanisms that allow scale interactions are presented.

5.1 Fast Wave Average Framework

Using the multi-time scale procedure on a family of weakly coupled nonlinear oscillators, [25] derived a fast wave average framework that was used to explicitly show that the slow scale variability (amplitude modulation) of the modes is dependent on the nonlinear interactions. The inviscid rotating shallow water equations (42) in the quasi-geostrophic limit (i.e., for $\epsilon \to 0$) are written as

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \epsilon^{-1} f\hat{k} \times \mathbf{v} = -\epsilon^{-1}\nabla h, \tag{42a}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} + \epsilon^{-1}(1 + \epsilon h)\nabla \cdot \mathbf{v} = 0. \tag{42b}$$

Furthermore, because of the parameter dependency in ϵ, it is suggested that the solution of (42) may, in principle, be also formulated as a function of ϵ, i.e., $\vec{\phi}^{\epsilon}$. With $\vec{\phi}^{\epsilon} = (\mathbf{v}^{\epsilon}, rh^{\epsilon})^T$ given by

$$\vec{\phi}^{\epsilon} = \vec{\phi}^0(\tilde{t}, t) + \epsilon\vec{\phi}^1(\tilde{t}, t) + \mathcal{O}(\epsilon^2) + \cdots, \tag{43}$$

where $\tilde{t} = t/\epsilon$ is the fast time scale and t the slow time scale. The initial condition is given by $\vec{\phi}^{\epsilon}_{|_{t=0}} = \vec{\phi}$. Moreover, we make use of $\mathcal{L}$ to represent the linear operator and $\mathcal{N}$ the nonlinear operator. We note that far from the equatorial region, it is possible to apply Fourier transform to both zonal and meridional directions, leading to a skew-symmetric linear operator ($\mathcal{L} + \mathcal{L}^{T^*} = 0$) and the eigenvalues are purely imaginary

$$\left(\mathcal{L} - i\omega I\right)\begin{pmatrix}\vec{v}\\ rh\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \tag{44}$$

from which follows:

$$i\omega\left(\omega^2 - f^2 - k^2 - l^2\right) = 0. \tag{45}$$

The slow geostrophic mode ($\omega_{\mathcal{S}}$) corresponds to the solution with $\omega = 0$ in (45), whereas the eastward/westward inertia-gravity waves when $\omega = \pm(f^2 + k^2 + l^2)^{1/2}$

represent the fast inertia-gravity waves ($\omega_{\mathcal{F}}$). Here is important to note that in the tropical region, the trapping behavior of the equatorial wave guide prevents the utilization of the Fourier decomposition in the meridional direction. This leads to a different problem, in which the fast and slow equatorial modes are connected [34–36].

Going back to our problem, we note that from the sequence of easier to solve problems we get that

$$\frac{\partial \vec{\phi}^1}{\partial \tilde{t}} + \mathcal{L}(\vec{\phi}^1) = -\left(\frac{\partial \vec{\phi}^0}{\partial t} + \mathcal{N}\left(\vec{\phi}^0, \vec{\phi}^0\right)\right). \tag{46}$$

To solve (46), the Duhamel formula can be used and to ensure that (43) must be valid for a long time (solvability condition) it is requested that $\vec{\phi}^1 \mathrm{e}^{\tilde{t}\mathcal{L}}$ must be also bounded for the skew-symmetric, norm preserving $\mathcal{L}$ operator. Thus, it is possible to obtain an equation for the slow amplitude evolution $\mathcal{A}$ of the leading order modes, with $\vec{\phi}^0(\tilde{t}, t) = \mathrm{e}^{-\mathcal{L}\tilde{t}}\mathcal{A}(t)$. Consequently, the fast wave average framework results in

$$\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}t} + \lim_{\tilde{t}\to\infty} \frac{1}{\tilde{t}} \int_0^{\tilde{t}} \mathrm{e}^{s\mathcal{L}}\, \mathcal{N}(\mathrm{e}^{-s\mathcal{L}}\mathcal{A}, \mathrm{e}^{-s\mathcal{L}}\mathcal{A})\, \mathrm{d}s = 0. \tag{47}$$

As discussed, the leading order modes for the linear operator $\mathcal{L}$ are made of RWs and IGWs. Thus, the averaged equations are valid for both modes. To understand the dynamics of the framework, it is necessary to explore the nonlinear operator, which can also be written as

$$\mathcal{N} \equiv \mathcal{N}(\vec{\phi}^0, \vec{\phi}^0). \tag{48}$$

Furthermore, as $\mathbf{v^0}$ must also include both the fast IGW and slow RWs, i.e., $\vec{\phi}^0 = \vec{\phi}^0_{\mathcal{F}} + \vec{\phi}^0_{\mathcal{S}} = \mathcal{A}_{\mathcal{F}}\, \mathrm{e}^{-\mathcal{L}_{\mathcal{F}}\tilde{t}} + \mathcal{A}_{\mathcal{S}}\mathrm{e}^{-\mathcal{L}_{\mathcal{S}}\tilde{t}} = \mathcal{A}_{\mathcal{F}}\, \mathrm{e}^{-\mathcal{L}_{\mathcal{F}}\tilde{t}} + \mathcal{A}_{\mathcal{S}}$; $\mathcal{L}_{\mathcal{S}}\tilde{t} = i\omega_{\mathcal{S}}\tilde{t} = 0$. Then,

$$\mathcal{N} = \mathcal{N}(\vec{\phi}^0_{\mathcal{F}}, \vec{\phi}^0_{\mathcal{F}}) + \mathcal{N}(\vec{\phi}^0_{\mathcal{F}}, \vec{\phi}^0_{\mathcal{S}}) + \mathcal{N}(\vec{\phi}^0_{\mathcal{S}}, \vec{\phi}^0_{\mathcal{F}}) + \mathcal{N}(\vec{\phi}^0_{\mathcal{S}}, \vec{\phi}^0_{\mathcal{S}}), \tag{49}$$

corresponding to the interactions between fast waves, fast and slow waves, slow and fast waves, and between slow waves, respectively. The framework (47) can be written in the form of a coupled ordinary differential equations, one representing the slow modes and the other representing the fast modes. In addition, the resonance conditions prevent some of the interactions

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}_{\mathcal{F}} + \mathcal{C}^{\mathcal{F};\mathcal{F}}_{\mathcal{F}}\, \mathcal{N}(\mathcal{A}_{\mathcal{F}}, \mathcal{A}_{\mathcal{F}}) + \mathcal{C}^{\mathcal{F};\mathcal{S}}_{\mathcal{F}}\, \mathcal{N}(\mathcal{A}_{\mathcal{F}}, \mathcal{A}_{\mathcal{S}}) = 0, \tag{50a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}_{\mathcal{S}} + \mathcal{C}^{\mathcal{F};\mathcal{F}}_{\mathcal{S}}\, \mathcal{N}(\mathcal{A}_{\mathcal{F}}, \mathcal{A}_{\mathcal{F}}) + \mathcal{C}^{\mathcal{S};\mathcal{S}}_{\mathcal{S}}\, \mathcal{N}(\mathcal{A}_{\mathcal{S}}, \mathcal{A}_{\mathcal{S}}) = 0, \tag{50b}$$

where the set of interaction coefficients $\{\mathcal{C}_{\mathcal{F}}^{\mathcal{F};\mathcal{F}}, \mathcal{C}_{\mathcal{F}}^{\mathcal{F};\mathcal{S}}, \mathcal{C}_{\mathcal{S}}^{\mathcal{F};\mathcal{F}}, \mathcal{C}_{\mathcal{S}}^{\mathcal{S};\mathcal{S}}\}$ represents different combinations that have a non-null contribution for the mode interactions involving geostrophic and inertia-gravity waves. In the interaction coefficients, the superscript represents a combination of two of the interacting modes, while the subscript represents the mode that is being affected by the nonlinear interaction. In (50), it is possible to see that a necessary condition to have the evolution of geostrophic mode totally independent of the inertia-gravity wave is that $\mathcal{C}_{\mathcal{S}}^{\mathcal{F};\mathcal{F}} = 0$. This criteria is related to the conservation of vorticity.

6 Summary

In the tropics, the internal or near field response to localized heat sources occurs in terms of baroclinic modes, while the far field response is related to barotropic modes. The far field response has a strong projection on mid- and high-latitudes. In this overview, we start the study of Rossby waves by analyzing the properties of the barotropic equations. By using the incompressible bidimensional Euler equations, the non-divergence condition allows to describe the problem using the stream-function formulation and a nonlinear equation for the stream function is obtained.

The linearized version of the barotropic vorticity equation has RWs as solutions. However, it was shown that a single RW is an exact solution of the nonlinear barotropic vorticity equation. This peculiarity occurs because a single RW is not able to advect itself since the wind is parallel to the stream-function isolines. However, if more than one wave participates, the wind field of one wave may advect the streamlines of the other and the initial field is distorted, then newer RWs can be created. Approximate solutions for the nonlinear barotropic vorticity equation for the weakly nonlinear case can also be obtained. In this limit, the advective timescale (L/U) is larger than the timescale related to the RW $((\beta L)^{-1})$. This is used to build a parameter that measures the strength of the nonlinearity $1/\beta^* = U/\beta L^2$. In the weakly nonlinear case $1/\beta^* \ll 1$, the solutions for the stream function can be viewed as successive corrections for the relative vorticity gradient. In this case, a sequence of easier to solve problems is obtained, allowing the distinction between oscillatory and resonant conditions.

Resonant conditions are also important for the case of interaction with the topography in the presence of an externally forced shear. The zonal gradient of the topography introduces a strong forcing of $\mathcal{O}[\epsilon^{-1}]$ to the transient wavy motion. The simplified treatment of the topography shape allows mathematical tractability and clearly shows that the topography is able to generate a downstream RW following its shape.

Although the nonlinear barotropic vorticity description provides interesting insights, it is necessary to consider vertical mode interactions in order to further understand the tropical–extratropical interactions. Thus, an increasing complexity can be achieved by using a two-layer equatorial β-plane model. The two-layer

model is a simplified composition that allows to include both the barotropic and baroclinic descriptions, which can thus be used as a model for tropical–extratropical interactions. Baroclinic modes are meridionally trapped but radiates upward, whereas barotropic modes are vertical confined but can leave the tropics, dominating the tropical–extratropical teleconnections.

The barotropic description in a multilayer model includes new nonlinear terms that represent the effects of the baroclinic structures. The archetypal model used in the nonlinear interactions for the barotropic case in the two-layer model suggests that for $J(\psi, \zeta) = J(\psi, \zeta - F\psi)$, $F = (L/L_R)^2$ cannot vary in space; one condition that may violate this constraint is the existence of mechanisms that modify the phase speed of IGWs.

The baroclinic description of the flow is also modified by the presence of barotropic waves as well as by the presence of a basic state. A reduced model can be obtained in terms of a set of coupled Korteweg–de Vries (KdV)-like equations, in which non-trivial terms allow the energy exchange between the barotropic and baroclinic flows. In this description, the role of the mean baroclinic shear $\bar{A}$ in controlling the energy flow between the transient barotropic B' and baroclinic A' disturbances was verified. Finally, tropical heating is believed to be inefficient to promote vertical interactions; therefore, shears are to be important for the exchange.

In the midlatitudes, the basic balance relationship is geostrophy, where RWs are of primary concern at the expense of fast oscillations, although a balance between stringent and permissive control of the fast modes can allow improvements to weather forecasting.

In the tropics, a substantial fraction of the large-scale variability can also be explained by the equatorially trapped waves. However, even at the quasi-geostrophic limit, in the equatorial area the fast and slow modes are connected by the mixed Rossby-gravity wave and there is not a clear scale separation. Furthermore, as it was discussed much of the skill of seasonal forecast for different regions of the globe originates in the tropics. Therefore, there can be little doubt that a better representation of the fast modes and its importance to the evolution of slow modes, in the light of scale interactions, is key to improve reliability of seasonal forecast.

References

1. Ambrizzi, T., Hoskins, B.J., Hsu, H.H.: Rossby wave propagation and teleconnection patterns in the austral winter. J. Atmos. Sci. **52**(21), 3661–3672 (1995)
2. Carvalho, L.M.V., Jones, C., Liebmann, B.: The south Atlantic convergence zone: intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. J. Clim. **17**(1), 88–108 (2004) https://doi.org/10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2
3. Cassou, C.: Intraseasonal interaction between the Madden–Julian oscillation and the north Atlantic oscillation. Nature **455**,523–527 (2008). https://doi.org/10.1038/nature07286
4. Charney, J., Devore, J.: Multiple flow equilibria in the atmosphere and blocking. J. Atmos. Sci. **36**, 1205–1216 (1979)

5. Charney, J., Strauss, D.M.: Form drag instability, multiple equilibria and propagating planetary waves in baroclinic, orographically forced, planetary wave systems. J. Atmos. Sci. **37**, 1157–1176 (1980)
6. Coelho, C.A.S., Oliveira, C.P.O., Ambrizzi, T., Reboita, M.S., Carpenedo, C.B., Campos, J.L.P., Tomaziello, A.C.N., Pampuch, L.A., Custódio, M.S., Dutra, L.M.M., Da Rocha, R.P., Rehbein, A.: The 2014 southeast brazil austral summer drought: regional scale mechanisms and teleconnections. Clim. Dyn. **46**(11–12), 3737–3752 (2016). https://doi.org/10.1007/s00382-015-2800-1
7. Dias, P.L.D.S., Bonatti, J.P.: A preliminary study of the observed vertical mode structure of the summer circulation over tropical South America. Tellus Ser. A Dyn. Meteorol. Oceanogr. **37**(2), 185–195 (1985). https://doi.org/10.3402/tellusa.v37i2.11665
8. Dole, R., Hoerling, M., Kumar, A., Eischeid, J., Perlwitz, J., Quan, X.W., Kiladis, G., Webb, R., Murray, D., Chen, M., Wolter, K., Zhang, T.: The making of an extreme event: Putting the pieces together. Bull. Am. Meteorol. Soc. **95**(3), 427–440 (2014). https://doi.org/10.1175/BAMS-D-12-00069.1
9. Garfinkel, C., Benedict, J., Maloney, E.: Impact of the MJO on the boreal winter extratropical circulation. Geophys. Res. Lett. **41**(16), 6055–6062 (2014). https://doi.org/10.1002/2014GL061094
10. Gill, A.: Some simple solutions for heat-induced tropical circulation. Q. J. R. Meteorol. Soc. **106**, 447–462 (1980)
11. Gloeckler, L., Roundy, P.E.: Modulation of the extratropical circulation by combined activity of the Madden–Julian oscillation and equatorial Rossby wave during boreal winter. Mon. Weather Rev. **141**(4), 1347–1357 (2013)
12. Gurarie, D.: Long-range dynamics of a shallow water triad: renormalization, modulation, and cyclogenesis. J. Atmos. Sci. **60**(5), 693–710 (2003)
13. Hart, J.: Barotropic quasi-geostrophic flow over anisotropic mountains. J. Atmos. Sci. **36**, 1736–1746 (1979)
14. Higgins, R., Mo, K.: Persistent north pacific circulation anomalies and the tropical intraseasonal oscillation. J. Clim. **10**(2), 223–244 (1997)
15. Horel, J.D., Wallace, J.M.: Planetary-scale atmospheric phenomena associated with the southern oscillation. Mon. Weather Rev. **109**(4), 813–829 (1981)
16. Hoskins, B., Jin, F.F.: The initial value problem for tropical perturbations to a baroclinic atmosphere. Q. J. R. Meteorol. Soc. **117**, 299–317 (1991)
17. Hoskins, B.J., Karoly, D.J.: The steady linear response of a spherical atmosphere to thermal and orographic forcing. J. Atmos. Sci. **38**(6), 1179–1196 (1981)
18. Hoskins, B., Sardeshmukh, P.: A diagnostic study of the dynamics of the northern hemisphere winter of 1985–1986. Q. J. R. Meteorol. Soc. **113**, 759–778 (1987)
19. Karoly, D.J.: Southern hemisphere circulation features associated with El Niño-southern oscillation events. J. Clim. **2**, 1239–1252 (1989)
20. Kasahara, A., Silva Dias, P.: Response of equatorial planetary waves to stationary tropical heating in the global atmosphere with meridional and vertical shear. J. Atmos. Sci. **43**, 1893–1911 (1986)
21. Kevorkian, J., Cole, J.: Multiple Scale and Singular Perturbation Methods. Springer, New York (1986)
22. Lau, K., Lau, N.C.: The energetics and propagation dynamics of tropical summertime synoptic scale disturbances. Mon. Weather Rev. **120**, 2523–2539 (1992)
23. Lin, H., Brunet, G., Derome, J.: An observed connection between the north Atlantic oscillation and the Madden–Julian oscillation. J. Clim. (2009). https://doi.org/10.1175/2008JCLI2515.1
24. Lynch, P.: The Emergence of Numerical Weather Prediction Richardson's Dream. Cambridge University Press, Cambridge (2006)
25. Majda, A.: Introduction to PDEs and Waves for the Atmosphere and Ocean, vol. 9. American Mathematical Society, Providence (2003)
26. Majda, A.J., Biello, J.: The nonlinear interaction of barotropic and equatorial baroclinic Rossby waves. J. Atmos. Sci. **60**, 1809–1821 (2003)

27. Majda, A.J., Klein, R.: Systematic multiscale models for the tropics. J. Atmos. Sci. **60**(15), 393–408 (2003)
28. Matsuno, T.: Quasi-geostrophic motions in the equatorial area. J. Meteorol. Soc. Jpn. **44**, 25–42 (1966)
29. Matthews, A.J., Hoskins, B.J., Masutani, M.: The global response to tropical heating in the Madden-Julian oscillation during the northern winter. Q. J. R. Meteorol. Soc. **130**, 1991–2011 (2004)
30. McPhaden, M.: Genesis and evolution of the 1997–98 El Niño. Science **12**, 950–954 (1999)
31. Parsons, D., Yoneyama, K., J.L., R.: The evolution of the tropical western Pacific atmosphere-ocean system following the arrival of a dry intrusion. Q. J. R. Meteorol. Soc. **126**(563), 517–548 (2000)
32. Pedlosky, J.: Resonant topographic waves in barotropic and baroclinic flows. J. Atmos. Sci. **38**(12), 2626–2641 (1981)
33. Pedlosky, J.: Geophysical Fluid Dynamics. Springer, New York (1987)
34. Ramirez, E., Silva Dias, P.L., Raupp, C.: Asymptotic approach for the nonlinear equatorial long wave interactions. In: Journal of Physics: Conference Series, vol. 285, p. 012020. IOP Publishing, Bristol (2011)
35. Ramírez Gutiérrez, E., da Silva Dias, P.L., Raupp, C., Bonatti, J.P.: The family of anisotropically scaled equatorial waves. J. Adv. Model. Earth Syst. **3**(4), M12002 (2011)
36. Ramirez, E., da Silva Dias, P.L., Raupp, C.F.M.: Multiscale atmosphere–ocean interactions and the low-frequency variability in the equatorial region. J. Atmo. Sci. **74**(8), 2503–2523 (2017). https://doi.org/10.1175/JAS-D-15-0325.1
37. Raupp, C.F.: Interação não-linear entre ondas atmosféricas:um possível mecanismo para a conexão trópicos-extratrópicos em baixa-frequência. Tese (Doutorado em Meteorologia), Universidade de São Paulo – Instituto de Astronomia, Geofísica e Ciências Atmosféricas, São Paulo (2006)
38. Raupp, C.F.M., Silva Dias, P.L.: Dynamics of resonantly interacting equatorial waves. Tellus Ser. A Dyn. Meteorol. Oceanogr. **58**, 263–279 (2006). https://doi.org/10.1111/j.1600-0870.2006.00151.x
39. Ripa, P.: Weak interactions of equatorial waves in a one-layer model. Part I: general properties. J. Phys. Oceanogr. **13**, 1208–1226 (1983)
40. Ripa, P.: Weak interactions of equatorial waves in a one-layer model. Part II: applications. J. Phys. Oceanogr. **13**, 1227–1240 (1983)
41. Rodrigues, R.R., Woollings, T.: Impact of atmospheric blocking on South America in austral summer. J. Clim. **30**(5), 1821–1837 (2017). https://doi.org/10.1175/JCLI-D-16-0493.1
42. Rossby, C.G.: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. J. Mar. Res. **2**, 38–55 (1939)
43. Rossby, C.G.: On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. J. Meteorol. **2**(4), 187–204 (1945)
44. Salby, M.L., Garcia, R.R.: Transient response to localized episodic heating in the tropics. Part I: excitation and short-time near-field behavior. J. Atmos. Sci. **44**(2), 458–498 (1987). https://doi.org/10.1175/1520-0469(1987)044<0458:TRTLEH>2.0.CO;2
45. Webster, P.: Response of tropical atmosphere to local steady forcing. Mon. Weather Rev. **100**, 518–540 (1972)
46. Webster, P.: Mechanisms determining the atmospheric response to sea surface temperature anomalies. J. Atmos. Sci. **38**, 554–571 (1981)
47. Webster, P.: Seasonality in the local and remote atmospheric response to the sea surface temperature anomalies. J. Atmos. Sci. **39**, 41–52 (1982)
48. Zhang, C.: Madden-Julian oscillation. Rev. Geophys. **43** (2005). https://doi.org/10.1029/2004RG000158

Pollutant Dispersion Modeling via Mathematical Homogenization and Integral Transform-Based Multilayer Methods

Camila P. da Costa, Leslie D. Pérez-Fernández, and Julián Bravo-Castillero

1 Introduction

Dispersion of pollutants in the atmosphere is modeled via boundary/initial-value problems for advection–diffusion equations with variable coefficients. Various integral transform-based methods were developed to provide (semi-)analytical solutions to such models, for instance, the advection–diffusion multilayer method (ADMM—see [10, 25]), the generalized integral advection–diffusion multilayer technique (GIADMT—see [11, 12]), the generalized integral transform technique (GITT—see [14, 15]), the generalized integral Laplace transform technique (GILTT—see [24, 26]), and the classical integral transform technique (CITT—see [33]).

The ADMM and the GIADMT are based on the piecewise-constant approximation of the continuous wind velocity profile and eddy diffusivity coefficients in the vertical direction $z \in [0, h]$, here with uniform stepsize Δz, where h is the height of the planetary boundary layer (PLB), and the application of the Laplace transform. Such an approximation is the local average of the variable coefficients over each sublayer of thickness Δz. Then, the original problem with continuous coefficients is approximated by a problem with stepwise coefficients. Note that the finer such a stepwise approximation is, the more precise the results are, but the more the computational effort is required. In fact, the ADMM was developed to solve two-dimensional problems, whereas the GIADMT is its generalization to three-dimensional problems. Specifically, the GIADMT is the combination of the GITT

C. P. da Costa · L. D. Pérez-Fernández (✉)
Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, RS, Brazil
e-mail: camila.costa@ufpel.edu.br; leslie.fernandez@ufpel.edu.br

J. Bravo-Castillero
Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
e-mail: julian@mym.iimas.unam.mx

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_4

and the ADMM in which the ADMM is employed to obtain the coefficients of the Fourier series expansion of the pollutant concentration in the crosswind longitudinal direction arising from the application of the GITT (also, see [13, 40]).

In addition, it was reported in [27] that the ADMM (and, consequently, the GIADMT) produces estimations of pollutant concentrations which are as accurate as other integral transform-based methods such as the GILTT but with remarkably less computational cost. This last feature is essential in real-life situations such as industrial/natural disasters which require swift and accurate estimations of the ground-level distribution and concentration of the pollutants escaped to the atmosphere. Therefore, the availability of fast and precise estimations in operative conditions is required to prevent, or at least diminish, the impact of pollution emergencies on both health and economy.

In this contribution, in order to accelerate the availability of results with minimum loss of accuracy, both the purely applied ADMM and its realization as part of the GIADMT are combined with mathematical homogenization [3, 4, 35, 38]. Roughly speaking, homogenization of heterogeneous media is based on both the hypothesis of separation of structural scales and the continuum hypothesis, that is, the structure of the medium (here, the PBL) is characterized by a small geometrical parameter ε defined as the ratio of the characteristic lengths of both the continuous microscale and macroscale (here, $\varepsilon = \Delta z/h$) so macroscopic or effective properties can be provided for the heterogeneous medium (as long as the boundary conditions are uniform). With such assumptions, the hypothesis of equivalent homogeneity is valid, that is, there is an ideal homogeneous medium which is equivalent to the heterogeneous one in the sense that the constant physical properties of the former are the effective properties of the latter. From the mathematical point of view, the hypothesis of equivalent homogeneity is valid if the difference of the solutions of the original boundary-value problem with variable coefficients and the so-called homogenized problem with constant coefficients, which model the heterogeneous and related homogeneous media, respectively, is of order of a power of the ε with respect to the norm of the function space in which such solutions are sought. In other words, the solution of the original problem converges to the solution of the homogenized problem as $\varepsilon \to 0^+$, which, in this case, corresponds also to the coincidence of the original and the ADMM problems. Here, homogenization estimates of the effective wind velocity profile and eddy diffusivity coefficients are provided from the ADMM problem in the Laplace space for unstable atmospheric conditions.

Here, we focus on two- and three-dimensional steady-state models and consider various parameterizations of the wind velocity profile and eddy diffusivity coefficients for unstable atmospheric conditions. Then, homogenization estimates of the effective wind velocity profile and eddy diffusivity coefficients are provided from the multilayer problem in the Laplace space. Several computational experiments are performed to compare, for both computational cost and precision, the direct application of the multilayer methods with that of their combination with mathematical homogenization and to experimental data. In all the computational experiments, it is observed that the runtimes of the proposed approach combining mathematical

homogenization with multilayer methods are several orders of magnitude smaller that of the direct application of the multilayer methods with little loss of accuracy.

This work is organized as follows: Sect. 2 contains the derivation of the original three-dimensional problem, the formulation of its crosswind longitudinally integrated two-dimensional approximated version and other related two-dimensional problems, and some comments on the parameterizations of the coefficients for the mean wind vertical profile and the crosswind eddy diffusivity; Sect. 3 is devoted to describe the solution process via the ADMM of the generic two-dimensional problem formulated in Sect. 2, whereas the approximate solution provided by its combination with mathematical homogenization is presented and mathematically justified in Sect. 4; results from various simulations are presented and discussed in Sect. 5; and some concluding remarks are given in Sect. 6.

2 Formulation of Various Relevant Problems

2.1 Derivation of the Original Three-Dimensional Problem

Let $Oxyz$ be a fixed Cartesian coordinate system such that the pollutant source is located at coordinates $(0, 0, H_s)$, $H_s \in \mathbb{R}_+^*$. In this setting, air pollution modeling is based on the principle of mass conservation for one pollutant species with concentration $c(x, y, z, t)$ [36, 41]:

$$\frac{\partial c}{\partial t} + \mathbf{u} \cdot \nabla c - D\Delta c = S, \tag{1}$$

where $(x, y, z) \in \mathbb{R}_+ \times [-L_y, L_y] \times [0, h]$ are the space variables with $L_y, h \in \mathbb{R}_+^*$ and $H_s \in (0, h)$, $t \in \mathbb{R}_+^*$ is the time variable, $\mathbf{u}$ is the wind velocity vector field, D is the molecular diffusivity, and S is the pollutant source.

An initial simplification of conservation law (1) follows by assuming Reynolds decompositions for both the wind velocity field and the pollutant concentration, that is, $\mathbf{u} = \overline{\mathbf{u}} + \delta\mathbf{u}$ and $c = \overline{c} + \delta c$, where $\overline{(\cdot)}$ and $\delta(\cdot)$ represent the mean and fluctuating (turbulent) parts, respectively. Such decompositions are justified by the existence of the so-called spectral gap, which is the lack of variation at temporal or spatial mesoscales and separates macroscale mean motions from microscale turbulent ones (see Sections 2.2 and 2.3 of [36] for further details). Also, it is assumed [36, 41] that turbulence satisfies an ergodic hypothesis [5], that is, it is homogeneous and stationary, both statistically, so $\langle\overline{(\cdot)}\rangle = \overline{(\cdot)}$ and $\langle\delta(\cdot)\rangle = 0$, where $\langle\cdot\rangle$ stands for Reynolds average operator [36]. Also, turbulent motions smaller that the mesoscale, as the ones considered here, generally satisfy the conditions [7] for the so-called incompressibility approximation [22, 36], which produces the so-called continuity equation for turbulent fluctuations $\nabla \cdot \delta\mathbf{u} = 0$, implying that $\delta\mathbf{u} \cdot \nabla\delta c = \nabla \cdot (\delta c \delta\mathbf{u})$. With such considerations, and by averaging conservation law (1), it follows that

$$\frac{\partial \overline{c}}{\partial t} + \overline{\mathbf{u}} \cdot \nabla \overline{c} + \nabla \cdot \langle \delta c \delta \mathbf{u} \rangle - D \Delta \overline{c} = \langle S \rangle, \tag{2}$$

where $\langle \delta c \delta \mathbf{u} \rangle$ introduces three new unknowns representing the turbulent atmospheric diffusion eddies, so closure is needed. Usually, closure is achieved via the so-called K-theory, or gradient-transport theory [36, 41] which relies on the so-called Fickian (or first-order local) constitutive law $\langle \delta c \delta \mathbf{u} \rangle = -\mathbf{K} \nabla \overline{c}$, where $\mathbf{K}$ is the second-rank tensor field of turbulent diffusion. Other types of local and non-local closures can be considered (see, for instance, Chapter 6 of [36], and [13, 42], respectively) but this falls out of the scope of this work and will be considered in future studies.

In addition, it is possible to neglect term $D \Delta \overline{c}$ in the averaged law (2) as the dispersion effects of molecular diffusion are several orders of magnitude smaller than the ones corresponding to the turbulent diffusion eddies [36, 41]. With such considerations, and by assuming that the pollutant is nonreactive, so $\langle S \rangle = S$, conservation law (2) becomes

$$\frac{\partial \overline{c}}{\partial t} + \overline{\mathbf{u}} \cdot \nabla \overline{c} - \nabla \cdot (\mathbf{K} \nabla \overline{c}) = S. \tag{3}$$

Further simplifications of law (3) are considered. For instance, $\mathbf{K}$ is assumed to be a diagonal tensor with nonzero components K_x, K_y, and K_z [41] as cross-diagonal diffusion terms are usually insignificant [22]. In addition, it is considered that the x-axis is aligned with the wind direction, so that $\overline{\mathbf{u}} = (\overline{u}, 0, 0)$, and, in consequence, the turbulent diffusion along the x-axis is negligible in comparison to the corresponding advective transport:

$$\left| \overline{u} \frac{\partial \overline{c}}{\partial x} \right| \gg \left| \frac{\partial}{\partial x} \left[K_x \frac{\partial \overline{c}}{\partial x} \right] \right|. \tag{4}$$

Also, in the presence of a single pollutant source in steady-state emission regime and atmospheric conditions we have that $\partial \overline{c} / \partial t = 0$, and the source term S can be treated as a boundary condition. With such considerations, conservation law (3) simplifies to the following steady-state advection–diffusion equation:

$$\overline{u} \frac{\partial \overline{c}}{\partial x} - \frac{\partial}{\partial y} \left[K_y \frac{\partial \overline{c}}{\partial y} \right] - \frac{\partial}{\partial z} \left[K_z \frac{\partial \overline{c}}{\partial z} \right] = 0, \tag{5}$$

which is complemented with the boundary condition accounting for location and emission rate of the pollutant source:

$$\overline{uc}|_{x=0} = Q \delta(z - H_s) \delta(y), \tag{6}$$

where $Q \in \mathbb{R}_+^*$ is the emission rate of the pollutant source and $\delta(\cdot)$ is Dirac's delta function, and total reflexion conditions [41]:

$$K_y \frac{\partial \overline{c}}{\partial y}\bigg|_{y=\pm L_y} = K_z \frac{\partial \overline{c}}{\partial z}\bigg|_{z=0,h} = 0. \tag{7}$$

A final assumption is that the dominant atmospheric heterogeneity occurs in the vertical direction, so the mean wind velocity $\overline{u}$ and the crosswind eddy diffusivities K_y and K_z depend only on the vertical coordinate $z \in [0, h]$. Thus, the three-dimensional steady-state pollutant dispersion problem is formally stated as follows:

Given the data $L_y, h, Q \in \mathbb{R}^*_+$, $H_s \in (0, h)$, and $\overline{u}, K_y, K_z \in \mathscr{C}([0, h])$, find $\overline{c} \in \mathscr{H}^1(\mathbb{R}_+ \times [-L_y, L_y] \times [0, h])$ such that

$$\overline{u}(z)\frac{\partial \overline{c}}{\partial x} - K_y(z)\frac{\partial^2 \overline{c}}{\partial y^2} - \frac{\partial}{\partial z}\left[K_z(z)\frac{\partial \overline{c}}{\partial z}\right] = 0, \ (x, y, z) \in \mathbb{R}^*_+ \times (-L_y, L_y) \times (0, h), \tag{8}$$

$$\frac{\partial \overline{c}}{\partial y}\bigg|_{y=-L_y} = \frac{\partial \overline{c}}{\partial y}\bigg|_{y=L_y} = 0, \ (x, z) \in \mathbb{R}^*_+ \times (0, h), \tag{9}$$

$$K_z(0)\frac{\partial \overline{c}}{\partial z}\bigg|_{z=0} = K_z(h)\frac{\partial \overline{c}}{\partial z}\bigg|_{z=h} = 0, \ (x, y) \in \mathbb{R}^*_+ \times (-L_y, L_y), \tag{10}$$

$$\overline{u}(z)\overline{c}(0, y, z) - Q\delta(z - H_s)\delta(y) = 0, \ (y, z) \in (-L_y, L_y) \times (0, h), \tag{11}$$

where $\mathscr{C}(\cdot)$ is the space of continuous functions, and $\mathscr{H}^1(\cdot)$ is the space of square-integrable functions with square-integrable first-order derivatives.

2.2 Related Two-Dimensional Problems

A widely used approach to approximate the solution of problem (8)–(11) is to solve the related two-dimensional problem obtained by performing crosswind horizontal integration. Such a problem is stated as follows:

First, denote the crosswind horizontally averaged pollutant concentration by $\overline{c}^y(x, z) = \int_{-L_y}^{L_y} \overline{c}(x, y, z)dy$. Then, given $h, Q \in \mathbb{R}^*_+$, $H_s \in (0, h)$, and $\overline{u}, K_z \in \mathscr{C}([0, h])$, find $\overline{c}^y \in \mathscr{H}^1(\mathbb{R}_+ \times [0, h])$ such that

$$\overline{u}(z)\frac{\partial \overline{c}^y}{\partial x} - \frac{\partial}{\partial z}\left[K_z(z)\frac{\partial \overline{c}^y}{\partial z}\right] = 0, \ (x, z) \in \mathbb{R}^*_+ \times (0, h), \tag{12}$$

$$K_z(0)\frac{\partial \overline{c}^y}{\partial z}\bigg|_{z=0} = K_z(h)\frac{\partial \overline{c}^y}{\partial z}\bigg|_{z=h} = 0, \ x \in \mathbb{R}^*_+, \tag{13}$$

$$\overline{u}(z)\overline{c}^y(0, z) - Q\delta(z - H_s) = 0, \ z \in (0, h). \tag{14}$$

In this contribution, the solution $\overline{c}^y(x,z)$ of the two-dimensional approximated problem (12)–(14) is sought (in fact, approximated) via the ADMM. An alternative approach to estimate the solution of the original three-dimensional problem (8)–(11) is given by the GIADMT, which employees separation of variables with a Fourier series in the crosswind horizontal direction (instead of averaging) and applies the ADMM to a sequence of two-dimensional problems in order to obtain the coefficients of the series. Explicitly, the solution of the original problem (8)–(11) is sought as

$$\overline{c}(x,y,z) = \sum_{j=0}^{\infty} \frac{C_j(x,z)}{\sqrt{L_y}} \cos\frac{j\pi y}{L_y}. \tag{15}$$

The orthonormal basis $\left\{\cos(j\pi y/L_y)/\sqrt{L_y}\right\}_{j\in\{0\}\bigcup\mathbb{N}}$ of the Fourier series (15) was obtained by solving the related Sturm–Liouville problems (see, for instance, [29]), and the corresponding coefficient $C_j(x,z)$, $j \in \{0\}\bigcup\mathbb{N}$, is the solution of the following two-dimensional problem:

Given the data $L_y, h, Q \in \mathbb{R}_+^*$, $H_s \in (0,h)$, and $\overline{u}, K_y, K_z \in \mathscr{C}([0,h])$, find $C_j \in \mathscr{H}^1(\mathbb{R}_+ \times [0,h])$, $j \in \{0\}\bigcup\mathbb{N}$, such that

$$\overline{u}(z)\frac{\partial C_j}{\partial x} + \frac{j^2\pi^2}{L_y^2}K_y(z)C_j(x,z) - \frac{\partial}{\partial z}\left[K_z(z)\frac{\partial C_j}{\partial z}\right] = 0,\ (x,z) \in \mathbb{R}_+^* \times (0,h), \tag{16}$$

$$K_z(0)\left.\frac{\partial C_j}{\partial z}\right|_{z=0} = K_z(h)\left.\frac{\partial C_j}{\partial z}\right|_{z=h} = 0,\ x \in \mathbb{R}_+^*, \tag{17}$$

$$\overline{u}(z)C_j(0,z) - \frac{Q}{\sqrt{L_y}}\delta(z-H_s) = 0,\ z \in (0,h). \tag{18}$$

Problem (16)–(18) is also solved approximately via the ADMM, which completes the GIADMT approach.

In summary, we consider two direct approaches to estimate the solution of the original three-dimensional problem (8)–(11), namely the ADMM solution of the crosswind horizontally averaged two-dimensional problem (12)–(14), and its GIADMT solution (15) which in turn depends on the ADMM solution of a sequence of two-dimensional problems. The ADMM approach is detailed in Sect. 3.

In order to apply the ADMM in a unified way to both two-dimensional problems (12)–(14) and (16)–(18), notation is simplified by dropping symbol $\overline{(\cdot)}$, so $u \equiv \overline{u}$, $c \equiv \overline{c}$, and $c^y \equiv \overline{c}^y$, and also diffusivities are denoted as $k \equiv j^2\pi^2K_y/L_y^2$ and $K \equiv K_z$, respectively. Then, consider the following generic problem which contains problems (12)–(14) and (16)–(18) as particular cases:

Let $a \in \{0,1\}$ be a parameter representing the ADMM or the GIADMT approach. Then, given the data $L_y, h, Q \in \mathbb{R}_+^*$, $H_s \in (0,h)$, and $u, k, K \in \mathscr{C}([0,h])$, find $C \in \mathscr{H}^1(\mathbb{R}_+ \times [0,h])$, such that

$$u(z)\frac{\partial C}{\partial x} + ak(z)C(x,z) - \frac{\partial}{\partial z}\left[K(z)\frac{\partial C}{\partial z}\right] = 0,\ (x,z) \in \mathbb{R}_+^* \times (0,h), \tag{19}$$

$$K(0)\left.\frac{\partial C}{\partial z}\right|_{z=0} = K(h)\left.\frac{\partial C}{\partial z}\right|_{z=h} = 0,\ x \in \mathbb{R}_+^*, \tag{20}$$

$$\overline{u}(z)C(0,z) - \frac{Q}{L_y^{a/2}}\delta(z - H_s) = 0,\ z \in (0,h), \tag{21}$$

where problems (12)–(14) and (16)–(18) are obtained for $a = 0$ and $a = 1$, respectively, that is, $C = c^y$ for $a = 0$ and $C = C_j$ for $a = 1$. Next, we comment the specificities of the relations between the boundary conditions, the parameterizations of the coefficients, and the ADMM.

2.3 Some Comments on the Parameterizations of the Coefficients

Note that boundary conditions (20) are satisfied if $K(0) = K(h) = 0$, which indeed is the case of many parameterizations such as the ones in [16, 34, 39]. However, for vertical diffusion to occur, $K(z)$ must be nonzero at ground level, that is, $K(0) \neq 0$, as also noted in [18]. This feature is also desirable from the mathematical point of view in order to obtain regular Sturm–Liouville problems arising from the direct application of the Laplace transform to problem (19)–(21). This irregularity is overcome in a natural way by the ADMM by taking the local average of $K(z)$ in each sublayer of the PBL. Therefore, the ADMM allows using the following parameterizations of $K(z)$ for unstable atmospheric conditions:

- Parameterization of [30]:

$$\frac{K(z)}{w_* h} = \left(\frac{\kappa|L|}{h}\right)^{4/3} \frac{z}{|L|}\left(1 + 16\frac{z}{|L|}\right)^{-1/2} \tag{22}$$

- Parameterization of [34]:

$$\frac{K(z)}{w_* h} = \kappa \frac{z}{h}\left(1 - \frac{z}{h}\right) \tag{23}$$

- Parameterization of [16]:

$$\frac{K(z)}{w_* h} = 0.22\left(1 - e^{-4z/h} - 0.0003e^{8z/h}\right)\left(\frac{z}{h}\right)^{1/3}\left(1 - \frac{z}{h}\right)^{1/3} \tag{24}$$

- Parameterization of [39]:

$$\frac{K(z)}{w_* h} = \left(\frac{\kappa|L|}{h}\right)^{4/3} \frac{z}{|L|}\left(1 - \frac{z}{h}\right)\left(1 + 22\frac{z}{|L|}\right)^{1/4} \tag{25}$$

- Parameterization of [17]:

$$\frac{K(z)}{w_* h} = \frac{\kappa^{5/2}}{10}\left(\frac{\kappa|L|}{h}\right)^{-1/2}\left(1 - e^{-4z/h} - 0.0003e^{8z/h}\right)^{4/3}, \tag{26}$$

where $\kappa \sim 0.4$ is von Kármán's constant, $L < 0$ is Monin–Obukhov's length for unstable atmospheric conditions, and w_* is the convective velocity scale.

Interestingly, the related issues do not affect the crosswind horizontal eddy diffusivity $k(z)$, so the following adaptation of its parameterization from [16], which takes into account the notation introduced before problem (19)–(21), can be readily used:

$$\frac{k(z)}{w_* h} = \frac{j^2\pi^2}{L_y^2} 0.12\left(0.75 + \left(\frac{z}{|L|}\right)^{-2/3}\left(1 + \frac{z}{|L|}\right)^2\right)^{1/2}. \tag{27}$$

On the other hand, usual power-law [30] and logarithmic [6, 30] mean wind velocity profiles become null and negatively unbounded at $z = 0$, respectively. For the power-law case, boundary condition (21) is satisfied allowing nonzero pollutant concentration at the base of the pollutant source, that is, $c(0, 0) \neq 0$, which is a physical nonsense. Whereas, for the logarithmic case, $u(z) \to -\infty$ as $z \to 0^+$, which leads to a "$\infty = 0$" contradiction if $c(0, 0) \neq 0$; so, boundary condition (14) is satisfied only if either $c(0, 0) = 0$ (physically possible) or $u(z)c(0, z) \to 0^+$ as $z \to 0^+$, but this implies that $c(0, z) \to 0^-$, that is, $c(0, z) < 0$ (physical nonsense). On the other hand, if boundary condition (14) is written as $c(0, z) = Q\delta(z - H_s)/u(z)$, it leads to a "0/0" indeterminate form for the power-law case at $z = 0$, which in fact converges to $c(0, 0) = 0$ as $z \to 0^+$ because $\delta(-H_s) = 0$ and $u(z)$ is nonzero in the neighborhood of $z = 0$, whereas, for the logarithmic case, $c(0, z) \to 0^-$ as $z \to 0^+$, which is still nonsensical. Again, these situations are overcome naturally by the ADMM by approximating $u(z)$ by its local average in each sublayer of the PLB, which allows using the following parameterizations of $u(z)$ for unstable conditions [30]:

- Power-law profile:

$$u(z) = u_1\left(\frac{z}{z_1}\right)^p, \tag{28}$$

with wind velocity u_1 at height $z = z_1$, and turbulence intensity p [23].

- Logarithmic profile:

$$u(z) = \begin{cases} \dfrac{u_*}{\kappa}\left[\ln\left(\dfrac{z}{z_0}\right) - \Psi_m\left(\dfrac{z}{|L|}\right)\right], & z \le z_b \\ u(z_b), & z > z_b \end{cases}, \tag{29}$$

with friction velocity u_*, roughness length $z_0 > 0$, $z_b = \min\{|L|, 0.1h\}$, and stability function $\Psi_m(z/|L|)$

$$\Psi_m\left(\frac{z}{|L|}\right) = \ln P\left(\frac{z}{|L|}\right) - 2\arctan A\left(\frac{z}{|L|}\right), \tag{30}$$

where $A(z/|L|) = (1 + 16z/|L|)^{1/4}$ [32] and

$$P\left(\frac{z}{|L|}\right) = \frac{e^{\pi/2}}{8}\left(1 + A\left(\frac{z}{|L|}\right) + A^2\left(\frac{z}{|L|}\right) + A^3\left(\frac{z}{|L|}\right)\right). \tag{31}$$

Finally, note that all these issues can be avoided if the domain of original problem is defined above the roughness length, that is, for $z \in [z_0, h]$. The physical and mathematical implications of such a restriction will be addressed elsewhere. Next, the ADMM is applied to problem (19)–(21) in detail.

3 Application of the ADMM

Let $\{z_n\} \subset [0, h]$, with $z_n = n\Delta z$ and $n = \overline{0, N}$, be a uniform partition of $[0, h]$ with stepsize Δz, which represents the vertical dimension of the PBL. In each sublayer (z_{n-1}, z_n), $n = \overline{1, N}$, of thickness Δz of the PBL, consider the stepwise approximations of $u(z)$, $k(z)$, and $K(z)$ given by

$$\alpha_n = \frac{1}{\Delta z}\int_{z_{n-1}}^{z_n} \alpha(z)dz, \tag{32}$$

where $\alpha = u, k, K$. Then, denoting $C_n(x, z) = C(x, z)$ for $z \in (z_{n-1}, z_n)$, $n = \overline{1, N}$, the ADMM problem related to the generic problem (19)–(21) is

$$u_n\frac{\partial C_n}{\partial x} + ak_nC_n(x, z) - K_n\frac{\partial^2 C_n}{\partial z^2} = 0, \ (x, z) \in \mathbb{R}_+^* \times (z_{n-1}, z_n), \ n = \overline{1, N} \tag{33}$$

$$C_n(x, z_n) = C_{n+1}(x, z_n), \ x \in \mathbb{R}_+^*, \ n = \overline{1, N-1} \tag{34}$$

$$K_n\frac{\partial C_n}{\partial z}\bigg|_{z=z_n} = K_{n+1}\frac{\partial C_{n+1}}{\partial z}\bigg|_{z=z_n}, \ x \in \mathbb{R}_+^*, \ n = \overline{1, N-1} \tag{35}$$

$$\left.\frac{\partial C_1}{\partial z}\right|_{z=0} = \left.\frac{\partial C_N}{\partial z}\right|_{z=h} = 0,\ x \in \mathbb{R}_+^* \tag{36}$$

$$C_n(0,z) = \frac{Q}{u_n L_y^{a/2}}\delta(z-H_s)\delta_{n\bar{n}},\ z \in (z_{n-1},z_n),\ n=\overline{1,N}, \tag{37}$$

where continuity conditions (34) and (35) guarantee sufficient smoothness to the generic pollutant concentration $C(x,z)$ by imposing continuity at the interior partition points $z=z_n$, $n=\overline{1,N-1}$, to $C(x,z)$ and the related vertical turbulent flux $K(z)\partial C/\partial z$, respectively. Also, $\delta_{n\bar{n}}$ in boundary condition (37) is Kronecker's delta, so $n=\bar{n}$ indicates that $H_s \in (z_{\bar{n}-1}, z_{\bar{n}})$, that is, the layer which the pollutant source is located at.

By applying the Laplace transform $\mathscr{L}[\cdot]$ with respect to variable x to the ADMM problem (33)–(37), it follows that, for each $s \in \mathbb{C}$,

$$\frac{d^2\zeta_n}{dz^2} - R_n^2(s)\zeta_n(s,z) = -\frac{Q}{K_n L_y^{a/2}}\delta(z-H_s)\delta_{n\bar{n}},\ z \in (z_{n-1},z_n),\ n=\overline{1,N} \tag{38}$$

$$\zeta_n(s,z_n) = \zeta_{n+1}(s,z_n),\ n=\overline{1,N-1} \tag{39}$$

$$K_n\left.\frac{d\zeta_n}{dz}\right|_{z=z_n} = K_{n+1}\left.\frac{d\zeta_{n+1}}{dz}\right|_{z=z_n},\ n=\overline{1,N-1} \tag{40}$$

$$\left.\frac{d\zeta_1}{dz}\right|_{z=0} = \left.\frac{d\zeta_N}{dz}\right|_{z=h} = 0, \tag{41}$$

where $\zeta_n(s,z) = \mathscr{L}[C_n(x,z)]$ and $R_n^2(s) = (u_n s + ak_n)/K_n$.

The solution $\zeta_n(s,z)$ of the ADMM problem in the Laplace space (38)–(41) for $z \in (z_{n-1},z_n)$, $n=\overline{1,N}$, is

$$\zeta_n(s,z) = A_n e^{R_n(s)z} + B_n e^{-R_n(s)z} - \frac{Q\sinh\{R_n(s)(z-H_s)\}}{K_n R_n(s) L_y^{a/2}}H(z-H_s)\delta_{n\bar{n}}, \tag{42}$$

where $H(\cdot)$ is Heaviside's unit step function. In solution (42), constant coefficients A_n and B_n, $n=\overline{1,N}$, are obtained by solving the system of linear algebraic equations resulting by substituting (42) into conditions (39)–(41). Then, the solution of the ADMM problem (33)–(37) follows by applying the inverse Laplace transform $\mathscr{L}^{-1}[\cdot]$ to solution (42). It is worth noting that the complexity of solution (42) requires the numerical inversion of the Laplace transform, so the final solution is regarded as semi-analytical. Typically, the inversion algorithm of choice is the

Gaussian quadrature scheme, but here the more robust fixed-Talbot algorithm [1] is employed (also, see [28]). Thus, for $z \in (z_{n-1}, z_n)$, $n = \overline{1, N}$, the solution of the ADMM problem (33)–(37) is

$$C_n(x, z) = \frac{r}{M}\left[\frac{1}{2}\zeta_n(r, z)e^{rx} + \sum_{k=1}^{M-1} Re\{e^{xS(\theta_k)}\zeta_n\left(S(\theta_k), z\right)\left(1 + i\omega(\theta_k)\right)\}\right], \tag{43}$$

where parameter $r \in \mathbb{R}_+^*$ is fixed (so, naming this version of Talbot algorithm [37]), $i = \sqrt{-1}$, and $S(\theta_k) = r\theta_k(\cot(\theta_k) + i)$, $\omega(\theta_k) = \theta_k + (\theta_k \cot(\theta_k) - 1)\cot(\theta_k)$, and $\theta_k = k\pi/M \in (-\pi, \pi)$ (for further details, see [1]).

Finally, note that, for more general situations, other inversion algorithms of the Laplace transform [2] can be considered in order to improve the computational efficiency. For instance, a Fourier series-based inversion algorithm proved to be more efficient computationally when linear counter-gradient non-local closure is considered in the averaged conservation law (2) [13].

4 The Homogenization Estimate

The so-called homogenized problem related to the ADMM problem in the Laplace space (38)–(41) is

$$\frac{d^2\overline{\zeta}}{dz^2} - \hat{R}^2(s)\overline{\zeta}(s, z) = -\frac{Q}{\hat{K}L_y^{a/2}}\delta(z - H_s), \ z \in (0, h) \tag{44}$$

$$\left.\frac{d\overline{\zeta}}{dz}\right|_{z=0} = \left.\frac{d\overline{\zeta}}{dz}\right|_{z=h} = 0, \tag{45}$$

where $\hat{R}^2(s) = \left(\hat{u}s + a\hat{k}\right)/\hat{K}$, $\overline{\zeta}(s, z)$ is both the volume average of $\zeta(s, z)$ and the Laplace transform of the homogenization estimate $\overline{C}(x, z)$ of the solution of the ADMM problem (33)–(37), and the effective coefficients $\hat{u}$, $\hat{k}$, and $\hat{K}$ are given by the arithmetic averages of $u(z)$ and $k(z)$, and the harmonic average and $K(z)$, respectively, that is, with $\alpha = u, k$,

$$\hat{\alpha} = \frac{1}{h}\int_0^h \alpha(z)dz \approx \frac{\Delta z}{h}\sum_{n=1}^{N}\alpha_n, \quad \hat{K}^{-1} = \frac{1}{h}\int_0^h \frac{dz}{K(z)} \approx \frac{\Delta z}{h}\sum_{n=1}^{N}K_n^{-1}. \tag{46}$$

The mathematical justification of the homogenized problem in the Laplace space (44)–(46) is as follows. First, note that, for every $s \in \mathbb{C}$ and every $N \in \mathbb{N}$, the ADMM problem in the Laplace space (38)–(41) can be written as

$$-\frac{d}{dz}\left[\tilde{K}(z)\frac{d\zeta}{dz}\right] + (\tilde{u}(z)s + a\tilde{k}(z))\zeta(s,z) = \frac{Q}{L_y^{a/2}}\delta(z - H_s),\ z \in (0,h)\backslash\Gamma \tag{47}$$

$$[\![\zeta(s,z)]\!]_{z=z_n} = 0,\quad \left[\!\!\left[\tilde{K}(z)\frac{d\zeta}{dz}\right]\!\!\right]_{z=z_n} = 0,\ z_n \in \Gamma \tag{48}$$

$$\left.\frac{d\zeta}{dz}\right|_{z=0} = \left.\frac{d\zeta}{dz}\right|_{z=h} = 0, \tag{49}$$

where $\Gamma = \{z_n\}_{n=\overline{1,N-1}}$, $H_s \neq z_n$, $n = \overline{0,N}$, and $\tilde{K}(z) = K_n$, $\tilde{u}(z) = u_n$, $\tilde{k}(z) = k_n$ for $z \in (z_{n-1}, z_n)$, $n = \overline{1,N}$. Also, $[\![\cdot]\!]$ denotes the jump operator around the discontinuity points $z = z_n$ of $\tilde{K}(z)$ and $\tilde{u}(z)$, and the derivatives are understood in the generalized sense (see, for instance, Chapter 1 of [3]). Existence of a generalized solution $\zeta(z)$ of problem (47)–(49) can be proved via a maximum principle (for instance, Lemma 1.1, page 95, of [31]). Note that, as $h = N\Delta z$, the geometric parameter describing the separation of scales is $\varepsilon = 1/N$, so $N \to \infty$ is equivalent to $\varepsilon \to 0^+$, that is, the limit which mathematical homogenization is based on. By Lemma 4.1, page 63, of [38], it follows that if there exist $\hat{K}, \hat{k}, \hat{u} \in \mathbb{R}_+^*$ and $\overline{\zeta}(z) \in \mathscr{H}^1([0,h])$ such that $\tilde{K}^{-1}(z) \to \hat{K}^{-1}$, $\tilde{k}(z) \to \hat{k}$, $\tilde{u}(z) \to \hat{u}$, and $\zeta(z) \to \overline{\zeta}(z)$ as $N \to \infty$, then $\overline{\zeta}(z)$ solves Eq. (44) and $\tilde{K}(z)\partial\zeta/\partial z \to \hat{K}\partial\overline{\zeta}/\partial z$ as $N \to \infty$, which is applied here to obtain the boundary conditions (45). On the other hand, for the advection-free case, in [19] a formula is proved, via a Green's function approach and averaging theorems, whose one-dimensional version is $\hat{K} = \left\langle \tilde{K}^{-1}(z)\right\rangle^{-1}$, with $\langle\cdot\rangle = h^{-1}\int_0^h (\cdot)dz$, which is the second formula in (46). Moreover, if the domain is ε-periodic, Theorem 2.6, page 33, of [9] guarantees that $\tilde{K}^{-1}(\xi) \to \left\langle \tilde{K}^{-1}(\xi)\right\rangle_\Omega$, $\tilde{k}(\xi) \to \langle\tilde{k}(\xi)\rangle_\Omega$, and $\tilde{u}(\xi) \to \langle\tilde{u}(\xi)\rangle_\Omega$ as $N \to \infty$, with $\langle\cdot\rangle_\Omega = |\Omega|^{-1}\int_\Omega (\cdot)d\xi$ defined locally over the period $\Omega \ni \xi = z/\varepsilon$. As the local average coincides with the average over the whole domain in this case, it is possible to identify $\langle\cdot\rangle$ with $\langle\cdot\rangle_\Omega$, which leads to take $\hat{K} = \left\langle \tilde{K}^{-1}(z)\right\rangle^{-1}$, $\hat{k} = \langle\tilde{k}(z)\rangle$, and $\hat{u} = \langle\tilde{u}(z)\rangle$, which in the one-dimensional case become formulas (46) for the effective coefficients.

Then, the solution of the homogenized problem in the Laplace space (44)–(46) is

$$\overline{\zeta}(s,z) = \frac{Q}{\hat{K}\hat{R}(s)L_y^{a/2}}\left[\frac{\cosh\left\{\hat{R}(s)z\right\}\cosh\left\{\hat{R}(s)(h-H_s)\right\}}{\sinh\left\{\hat{R}(s)h\right\}} - \sinh\left\{\hat{R}(s)(z-H_s)\right\}H(z-H_s)\right], \tag{50}$$

and the homogenization estimate of the solution of the ADMM problem (33)–(37) follows by applying the inverse Laplace transform to (51). Again, due to the complexity of solution (51), the numerical inversion of the Laplace transform is required, so the final estimate is regarded as semi-analytical. As above, the fixed-Talbot inversion algorithm of [1] is employed. Thus, the homogenization estimate of the solution of the ADMM problem (33)–(37) is

$$\overline{C}(x,z) = \frac{r}{M}\left[\frac{1}{2}\overline{\zeta}(r,z)e^{rx} + \sum_{k=1}^{M-1} Re\{e^{xS(\theta_k)}\overline{\zeta}\left(S(\theta_k),z\right)\left(1+i\omega(\theta_k)\right)\}\right]. \tag{51}$$

5 Results and Discussion

5.1 Two-Dimensional Case

In order to evaluate the performance in both accuracy and computational effort of the approach combining the ADMM and mathematical homogenization (Eq. (51)), simulations involving the five parameterizations of the mean vertical eddy diffusivity $K(z)$ (Eqs. (22)–(26)) and the two parameterizations of the mean wind velocity profile $u(z)$ (Eqs. (28) and (29)) were carried out in comparison with the direct application of the ADMM (Eq. (43)) and to the observations of the Copenhagen experiment [20]. The experiment consisted of the release without buoyancy of tracer sulfur hexafluoride in northern Copenhagen, Denmark, from a tower of height $H_s = 115$ m with emission rate $Q = 100$ g/s. The tracer was collected at ground-level sampling units located at three crosswind arcs positioned 2–6 km from the source location. The relevant meteorological parameters considered in the computational simulations are summarized in Table 1. For the simulations, Fortran codes (Visual Fortran 6.1) for Eqs. (43) and (51) were implemented and run in a Dell Inspiron 1440 computer (Intel Core 2 Duo P8700 @ 2.53 GHz, 4 GB RAM, 32-bits OS). The parameters used for the numerical inversion algorithm of the Laplace transform are $M = 100$ and $r = 2M/21x$. Table 2 shows the simulations runtimes. Remarkably, the direct application of the ADMM (Eq. (43)) took about 8–9 min, whereas the combination of the ADMM with mathematical homogenization (Eq. (51)—labeled A + H) took no more than 0.3 s.

Table 1 Copenhagen experiment dataset [20]

Exp.	$u_1(m/s)$	$u_*(m/s)$	$L(m)$	$w_*(m/s)$	$h(m)$
1	2.1	0.37	−46	1.8	1980
2	4.9	0.74	−384	1.8	1920
3	2.4	0.39	−108	1.3	1120
4	2.5	0.39	−173	0.7	390
5	3.1	0.46	−577	0.7	820
6	7.2	1.07	−569	2.0	1300
7	4.1	0.65	−136	2.2	1850
8	4.2	0.70	−72	2.2	810
9	5.1	0.77	−382	1.9	2090

Table 2 Simulations runtimes (s)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	480	0.20	488	0.23
Eq. (23)	531	0.28	507	0.27
Eq. (24)	482	0.27	489	0.25
Eq. (25)	483	0.25	479	0.20
Eq. (26)	533	0.25	481	0.23

Simulations results are presented in Fig. 1, where red and black curves and markers correspond to the ADMM solution (43) and the homogenization estimate (51), respectively. The two columns on the left correspond to the power-law wind velocity profile (Eq. (28)) with turbulence exponent $p = 0.07$ and $z_1 = 10\,\mathrm{m}$, whereas the two columns on the right correspond to the logarithmic wind velocity profile (Eq. (29)) with roughness length $z_0 = 0.6\,\mathrm{m}$, and the lines correspond to the five parameterizations of the eddy diffusivity (Eqs. (22)–(26)), respectively. The horizontal profiles of the source-normalized near ground-level pollutant concentration (C/Q) in the first and third columns were obtained for $z = 10\,\mathrm{m}$ for the eighth Copenhagen experiment, whereas scatter diagrams depicting source-normalized predicted (c_p/Q) versus observed (c_o/Q) concentration values are in the second and fourth columns, respectively, with the factor of two indicated by the gray lines.

An initial inspection of Fig. 1 indicates overall good agreement between the two semi-analytical approaches. The greatest discrepancy is observed in the first and last lines, which correspond to parameterizations (22) and (26) of the mean vertical eddy diffusivity $K(z)$, respectively. In these cases, the comparison to the experimental data in the scatter diagrams, which exhibit several points outside the lines representing the factor of two, suggests that such a behavior is a consequence of the particular parameterizations, as it is not observed for the other three parameterizations of $K(z)$ (Eqs. (23)–(25)).

In order to evaluate the accuracy of the results of the proposed approach in comparison with those of the ADMM and to the experimental data, the performance of both approaches is quantified via the statistical indexes proposed by [21]. Tables 3, 4, 5, 6 and 7 present the statistical indexes employed to evaluate

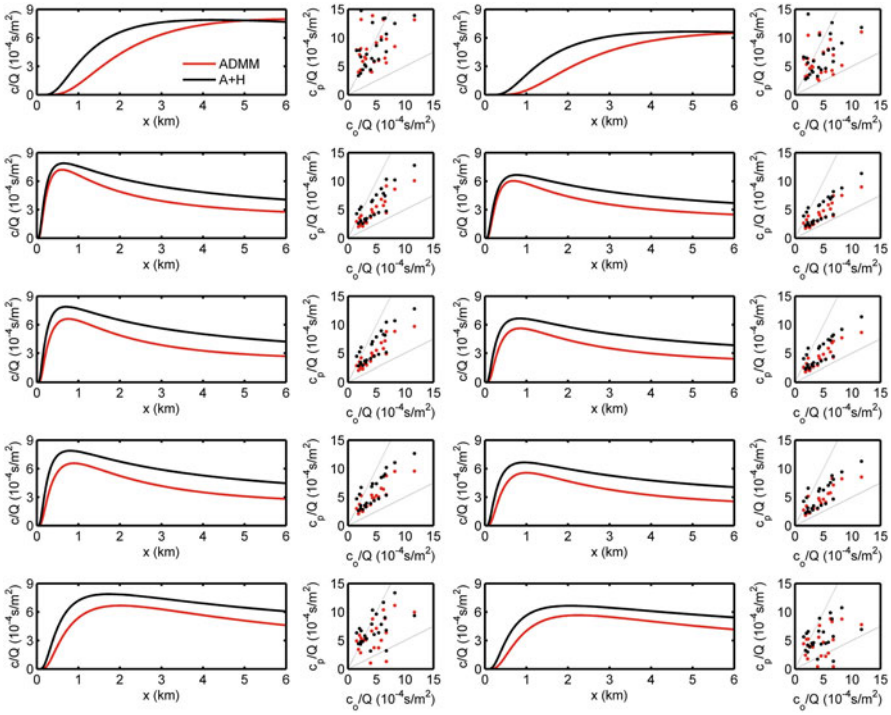


Fig. 1 Comparison of results of the ADMM solution (Eq. (43)) and the homogenization estimate (Eq. (51)—labeled A + H) from computational simulations in comparison to the data of the Copenhagen experiment for five eddy diffusivity parameterizations (Eqs. (22)–(26)) and two wind velocity profile parameterizations (Eqs. (28) and (29))

Table 3 Normalized mean square error (NMSE)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	0.59	0.61	0.39	0.46
Eq. (23)	0.05	0.14	0.07	0.09
Eq. (24)	0.05	0.16	0.06	0.10
Eq. (25)	0.07	0.18	0.07	0.10
Eq. (26)	0.29	0.38	0.33	0.32

Table 4 Correlation coefficient (COR)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	0.39	0.59	0.33	0.49
Eq. (23)	0.90	0.85	0.90	0.84
Eq. (24)	0.92	0.86	0.91	0.84
Eq. (25)	0.89	0.84	0.87	0.85
Eq. (26)	0.53	0.54	0.42	0.40

Table 5 Factor of two (FA2)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	0.52	0.65	0.78	0.78
Eq. (23)	1.00	0.87	1.00	0.87
Eq. (24)	1.00	0.87	1.00	0.87
Eq. (25)	0.96	0.87	0.96	0.96
Eq. (26)	0.70	0.70	0.65	0.74

Table 6 Fractional bias (FB)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	−0.55	−0.60	−0.28	−0.42
Eq. (23)	−0.03	−0.24	0.01	−0.10
Eq. (24)	−0.06	−0.29	0.07	−0.14
Eq. (25)	−0.12	−0.30	0.01	0.10
Eq. (26)	−0.24	−0.43	−0.01	−0.21

Table 7 Fractional standard deviation (FS)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	ADMM	A + H	ADMM	A + H
Eq. (22)	−0.24	−0.38	−0.06	−0.29
Eq. (23)	0.07	−0.14	0.21	0.01
Eq. (24)	0.01	−0.14	0.24	0.15
Eq. (25)	0.06	−0.15	0.19	0.14
Eq. (26)	−0.09	−0.07	−0.01	0.01

the accuracy of the proposed approach, namely: the normalized mean square error (NMSE—Table 3), the correlation coefficient (COR—Table 4), the factor of two (FA2—Table 5), the fractional bias (FB—Table 6), and the fractional standard deviation (FS—Table 7), which are calculated via the following formulas, where the overline indicates the arithmetic average and σ denotes the standard deviation:

$$\text{NMSE} = \frac{\overline{(c_p - c_o)^2}}{\overline{c_p c_o}}, \tag{52}$$

$$\text{COR} = \frac{\overline{(c_o - \bar{c}_o)(c_p - \bar{c}_p)}}{\sigma_o \sigma_p}, \tag{53}$$

$$\text{FA2} = (\text{percent of}) \ \frac{c_p}{c_o} \in [0.5, 2], \tag{54}$$

$$\text{FB} = \frac{\bar{c}_o - \bar{c}_p}{0.5(\bar{c}_o + \bar{c}_p)}, \tag{55}$$

$$\text{FS} = \frac{\sigma_o - \sigma_p}{0.5(\sigma_o + \sigma_p)}. \tag{56}$$

Table 8 Accuracy: this work versus results of [8]

Method	NMSE	COR	FA2	FB	FS
Double GITT	0.04	0.91	1.00	0.06	0.19
GILTT	0.09	0.85	1.00	0.11	0.13
ADMM	0.06	0.91	1.00	0.07	0.24
A + H	0.10	0.84	0.87	−0.14	0.15

Analysis of the data in Tables 3, 4, 5, 6 and 7 leads to the conclusion that the proposed approach yields results with little loss of accuracy. Remarkably, in the case corresponding to the parameterization (22) of $K(z)$, the proposed approach yields results that are better than those of the ADMM, which suggests that the combination of the ADMM with mathematical homogenization leads to improved estimations of the pollutant concentration when parameterizations are not accurate.

In addition, the accuracy of the proposed approach is compared to that of other methods by means of the statistical indexes (52)–(56) for the Copenhagen experiment data. Table 8 shows the comparison of the proposed approach to the ADMM and to two other integral transform-based methods, namely: the GITT (generalized integral transform technique) and the GILTT, with data taken from [8]. The situation corresponds to parameterizations (24) of $K(z)$ and (29) of $u(z)$. Note that the accuracy of the ADMM is similar to that of the double GITT, whereas the GILTT and the proposed approach exhibit similar accuracies. Again, the results provided by the proposed approach are more than acceptable, that is, the loss of accuracy is negligible, also considering the remarkably small computational cost.

5.2 Three-Dimensional Case

Now, in order to evaluate the performance in both accuracy and computational effort of the combination of the GIADMT and mathematical homogenization (Eqs. (15) and (51)), simulations involving the five parameterizations of the mean vertical eddy diffusivity $K(z)$ (Eqs. (22)–(26)), the mean crosswind longitudinal eddy diffusivity $k(z)$ (27), and the two parameterizations of the mean wind velocity profile $u(z)$ (Eqs. (28) and (29)) were carried out in comparison with the direct application of the GIADMT (Eq. (43)) and to the observations of the Copenhagen experiment [20]. All the relevant computational parameters are the same as in the previous section. Also, accuracy is evaluated via the statistical indexes (52)–(56), which are presented in Tables 9, 10, 11, 12 and 13. In addition, simulations runtimes of both approaches are presented in Table 14. In Tables 9, 10, 11, 12, 13 and 14, the proposed approach combining the GIADMT with mathematical homogenization is labeled G + H. Simulations results are illustrated in Figs. 2 and 3.

Table 9 Normalized mean square error (NMSE)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	0.35	0.21	0.44	0.20
Eq. (23)	0.23	0.21	0.25	0.20
Eq. (24)	0.21	0.18	0.24	0.17
Eq. (25)	0.17	0.16	0.20	0.16
Eq. (26)	0.36	0.19	0.57	0.31

Table 10 Correlation coefficient (COR)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	0.62	0.81	0.53	0.78
Eq. (23)	0.84	0.82	0.85	0.83
Eq. (24)	0.86	0.83	0.86	0.84
Eq. (25)	0.88	0.84	0.88	0.85
Eq. (26)	0.68	0.79	0.60	0.70

Table 11 Factor of two (FA2)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	0.74	0.78	0.52	0.78
Eq. (23)	0.83	0.83	0.83	0.83
Eq. (24)	0.83	0.96	0.83	0.87
Eq. (25)	0.87	0.91	0.83	0.87
Eq. (26)	0.70	0.83	0.61	0.78

Table 12 Fractional bias (FB)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	−0.15	−0.24	0.05	−0.12
Eq. (23)	0.21	0.06	0.26	0.12
Eq. (24)	0.22	0.02	0.27	0.08
Eq. (25)	0.17	−0.01	0.23	0.07
Eq. (26)	0.16	−0.05	0.33	0.11

Table 13 Fractional standard deviation (FS)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	0.02	−0.04	0.16	0.11
Eq. (23)	0.06	−0.04	0.14	0.05
Eq. (24)	0.11	−0.04	0.20	0.05
Eq. (25)	0.11	−0.03	0.19	0.06
Eq. (26)	0.17	0.25	0.28	0.37

Table 14 Simulations runtimes (s)

$u(z)$	Eq. (28)		Eq. (29)	
$K(z)$	GIADMT	G + H	GIADMT	G + H
Eq. (22)	3041	3	3243	4
Eq. (23)	3249	3	3555	3
Eq. (24)	3147	3	3403	3
Eq. (25)	3167	3	3397	3
Eq. (26)	3006	3	3284	4

Analysis of the data in Tables 9, 10, 11, 12 and 13 also leads to the conclusion that the proposed approach yields results with little loss of accuracy. In several cases, the proposed approach yields results that are better than those of the GIADMT, which suggests that its combination with mathematical homogenization leads to improved estimations of the pollutant concentration when parameterizations are not accurate. Finally, Table 14 shows the remarkable fact that the direct application of the GIADMT takes more than 50 min, whereas its combination with mathematical homogenization takes only about 3–4 s.

Figures 2 and 3 show contour plots of the proposed approach and scatter diagrams comparing it with the GIADMT approach for the power-law and logarithmic wind profiles (28) and (29), respectively, considering all the parameterizations of the vertical eddy diffusivity (22)–(26). Again, the greatest discrepancy is observed in the first and last lines, which correspond to parameterizations (22) and (26) of the mean vertical eddy diffusivity $K(z)$, respectively. In these cases, the comparison to the experimental data in the scatter diagrams, which exhibit several points outside the lines representing the factor of two.

6 Concluding Remarks

In this contribution, a mathematical homogenization approach was employed in combination with multilayer Laplace transform-based methods in order to accelerate the availability of results with minimum loss of accuracy. To the best of our knowledge, this approach using mathematical homogenization represents a novelty in air pollution modeling. Several computational simulations and statistical tests were performed for various parameterizations of mean wind velocity profile and vertical eddy diffusivity for unstable atmospheric conditions, in order to compare the new approach to the direct application of the multilayer methods for accuracy and computational time. The runtimes of the computational simulations of the proposed approach took less than one-third of a second and 3–4 s in the two- and three-dimensional cases, respectively, whereas the direct application of the ADMM and the GIADMT, which are the most computationally efficient among the integral transform-based methods, took about 8–9 min and 50–60 min, respectively. Moreover, the results of the proposed approach exhibited little loss of accuracy

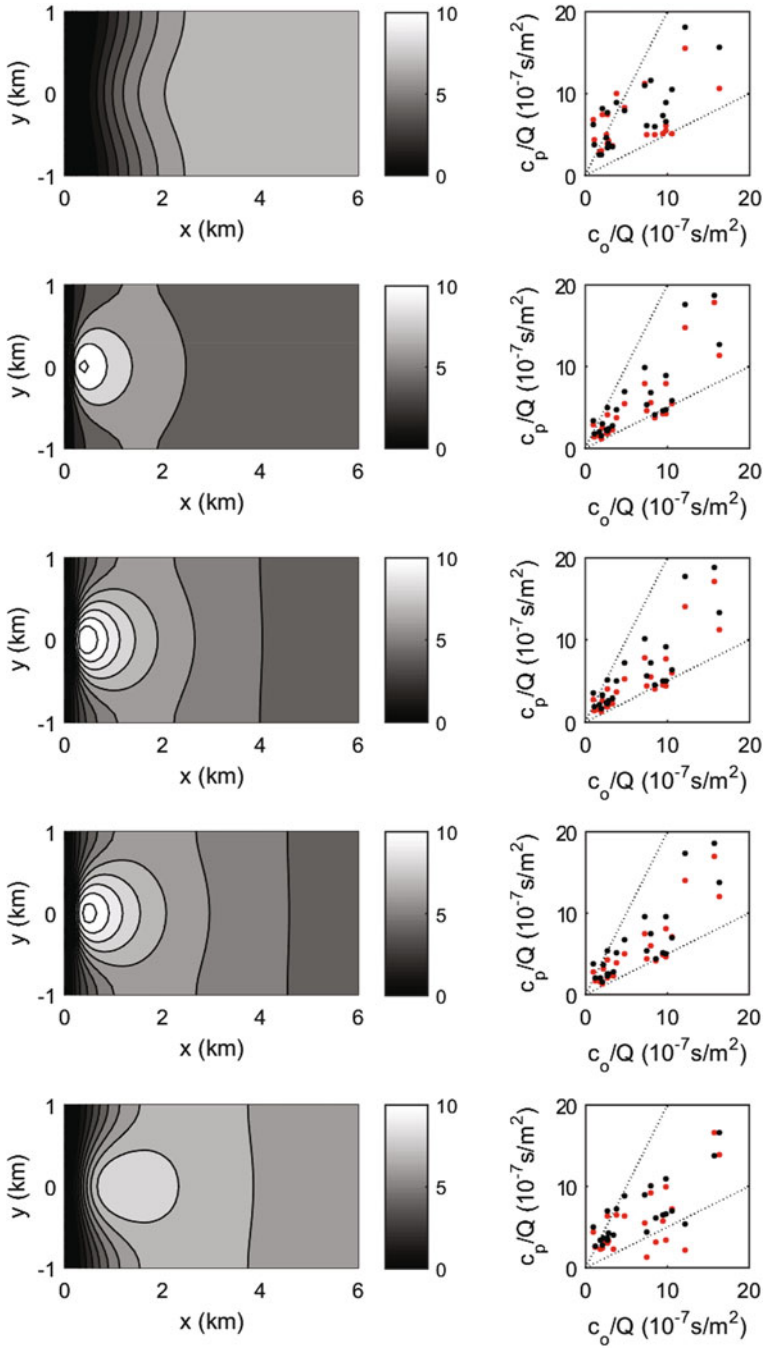


Fig. 2 Combination of the GIADMT solution (43) and the homogenization estimate (51) (left) and comparison of both approaches to the data of the Copenhagen experiment (right) for the eddy diffusivity parameterizations (22)–(26) and wind velocity profile power-law parameterization (28)

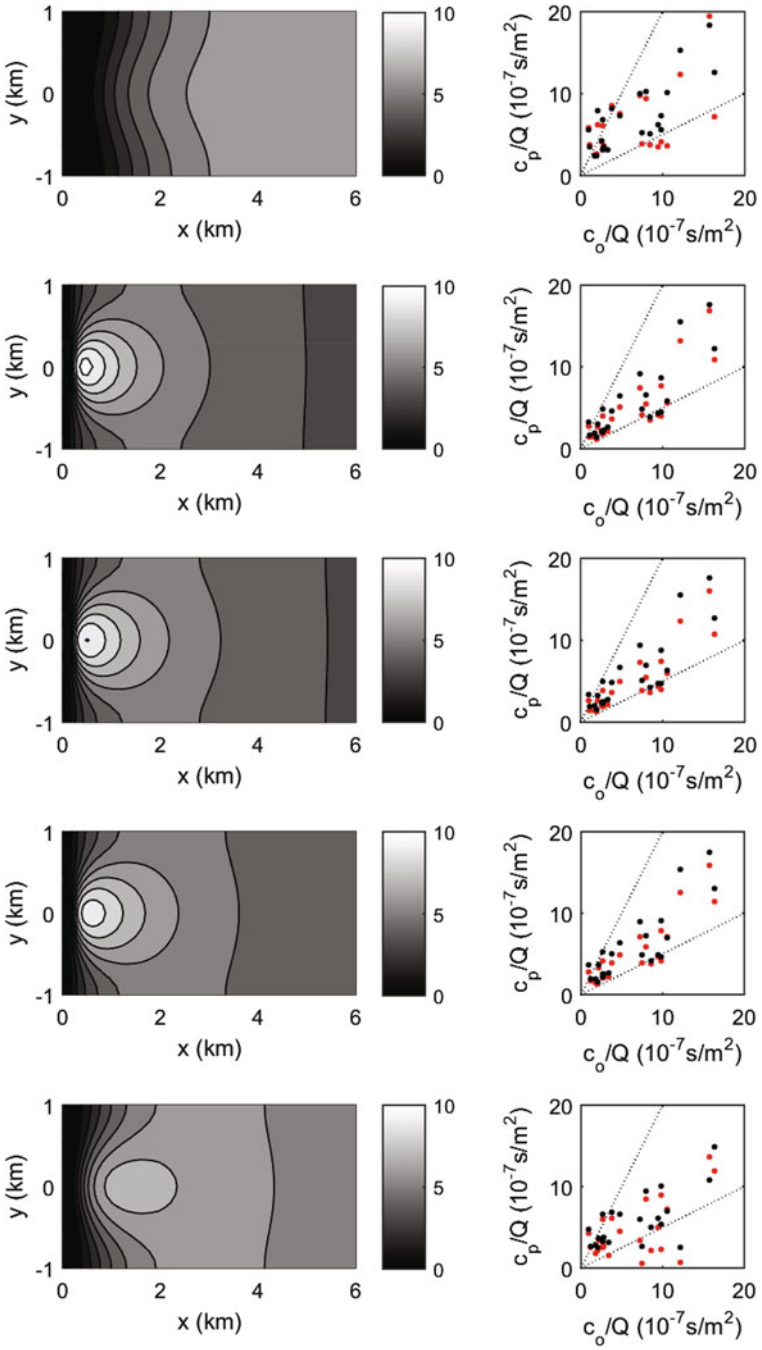


Fig. 3 Combination of the GIADMT solution (43) and the homogenization estimate (51) (left) and comparison of both approaches to the data of the Copenhagen experiment (right) for the eddy diffusivity parameterizations (22)–(26) and wind velocity profile logarithmic parameterization (29)

when compared with the results of the multilayer methods and to the Copenhagen experiment data. The results presented here are encouraging and suggest that mathematical homogenization can become a useful and powerful tool for solving pollutant dispersion models analytically.

Acknowledgements This study was partially financed by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*—Brazil (CAPES)—Finance Code 001, whose support via project No. 88881.030424/2013-01 is gratefully acknowledged.

References

1. Abate, J., Valkó, P.P.: Multi-precision Laplace transform inversion. Int. J. Numer. Meth. Eng. **60**, 979–993 (2004)
2. Abate, J., Whitt, W.: A unified framework for numerically inverting Laplace transforms. INFORMS J. Comput. **18**, 408–421 (2006)
3. Bakhvalov, N.S., Panasenko, G.P.: Homogenisation: Averaging Processes in Periodic Media. Kluwer, Dordrecht (1989)
4. Bensoussan, A., Lions, J.L., Papanicolau, G.: Asymptotic Analysis for Periodic Structures. North-Holland, Amsterdam (1978)
5. Beran, M.: Statistical Continuum Theories. John Wiley & Sons, New York (1968)
6. Berkowicz, R.R., Olesen, H.R., Torp, U.: The Danish Gaussian air pollution model (OML) – description, test and sensitivity analysis in view or regulatory applications. In: De Wispelaere, C., Schiermeier, F.A., Gillani, N.V. (eds.) Air Pollution Modeling and Its Application V. NATO Challenges of Modern Society, vol. 10, pp. 453–480. Springer, Boston (1986)
7. Businger, J.A.: Equations and concepts. In: Nieuwstadt, F.T.M., van Dop, H. (eds.) Atmospheric Turbulence and Air Pollution Modelling, pp. 1–36, Reidel, Dordrecht (1982)
8. Cassol, M., Wortmann, S., Rizza, U.: Analytical modeling of two-dimensional transient atmospheric pollutant dispersion by double GITT and Laplace Transform techniques. Environ. Modell. Softw. **24**, 144–151 (2009)
9. Cioranescu, D., Donato, P.: An Introduction to Homogenization. Oxford University Press, Oxford (1999)
10. Costa, C.P., Vilhena, M.T., Moreira, D.M., Tirabassi, T.: Semi-analytical solution of the steady three-dimensional advection-diffusion equation in the planetary boundary layer. Atmos. Environ. **40**, 5659–5669 (2006)
11. Costa, C.P., Vilhena, M.T., Moreira, D.M., Tirabassi, T.: The new GIADMT approach to simulate the pollutant dispersion in the planetary boundary layer. Develop. Environ. Sci. **6**, 805–807 (2007)
12. Costa, C.P., Tirabassi, T., Vilhena, M.T., Moreira, D.M.: A general formulation for pollutant dispersion in the atmosphere. J. Eng. Math. **74**, 159–173 (2012)
13. Costa, C.P., Rui, K., Pérez-Fernández, L.D.: Different numerical inversion algorithms of the Laplace transform for the solution of the advection-diffusion equation with non-local closure in air pollution modeling. Trends Appl. Comput. Math. **19**, 43–58 (2018)
14. Cotta, R.M.: Integral Transforms in Computational Heat and Fluid Flow. CRC Press, Boca Raton (1993)
15. Cotta, R., Mikhaylov, M.: Heat Conduction: Lumped Analysis, Integral Transforms, Symbolic Computation. John Wiley & Sons, New York (1997)

16. Degrazia, G., Rizza, U., Mangia, C., Tirabassi, T.: Validation of a new turbulent parameterization for dispersion models in convective conditions. Bound.-Lay. Meteorol. **85**, 243–254 (1997)
17. Degrazia, G., Anfossi, D., Carvalho, J., Mangia, C., Tirabassi, T., Campos Velho, H.F.: Turbulence parameterisation for PBL dispersion models in all stability conditions. Atmos. Environ. **34**, 3575–3583 (2000)
18. Essa, K.S.M., Etman, S.M., Embaby, M.: New analytical solution of the dispersion equation. Atmos. Res. **84**, 337–344 (2007)
19. Gorbachev, V.I., Pobedrya, B.E.: The effective characteristics of inhomogeneous media. J. Appl. Math. Mech. **61**, 145–151 (1997)
20. Gryning, S.E., Holtslag, A.A.M., Irwin, J.S., Sivertsen, B.: Applied dispersion modelling based on meteorological scaling parameters. Atmos. Environ. **21**, 79–89 (1987)
21. Hanna, S.R.: Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. Atmos. Environ. **23**, 1385–1398 (1989)
22. Hanna, S.R., Briggs, G.A., Hosker Jr., R.P.: Handbook on Atmospheric Diffusion. U. S. Department of Energy (1982)
23. Irwin, J.S.: A theoretical variation of the wind profile power-law exponent as a function of surface roughness and stability. Atmos. Environ. **13**, 191–194 (1979)
24. Moreira, D.M., Vilhena, M.T., Tirabassi, T., Buske, D., Cotta, R.M.: Near source atmospheric pollutant dispersion using the new GILTT method. Atmos. Environ. **39**, 6290–6295 (2005)
25. Moreira, D.M., Vilhena, M.T., Tirabassi, T., Costa, C., Bodmann, B.: Simulation of pollutant dispersion in atmosphere by the Laplace transform: the ADMM approach. Water Air Soil Pollut. **177**, 411–439 (2006)
26. Moreira, D.M., Vilhena, M.T., Buske, D., Tirabassi, T.: The state-of-art of the GILTT method to simulate pollutant dispersion in the atmosphere. Atmos. Res. **92**, 1–17 (2009)
27. Moreira, D.M., Vilhena, M.T., Tirabassi, T., Buske, D., Costa, C.P.: Comparison between analytical models to simulate pollutant dispersion in the atmosphere. Int. J. Environ. Waste Manag. **6**, 327–344 (2010)
28. Moreira, D.M., Moraes, A.C., Goulart, A.G., Albuquerque, T.T.A.: A contribution to solve the atmospheric diffusion equation with eddy diffusivity depending on source distance. Atmos. Environ. **83**, 254–259 (2014)
29. Özisik, M.N.: Heat Conduction. John Wiley & Sons, New York (1993)
30. Panofsky, H.A., Dutton, J.A.: Atmospheric Turbulence: Models and Methods for Engineering Applications. John Wiley & Sons, New York (1984)
31. Pao, C.V.: Nonlinear Parabolic and Elliptic Equations. Plenum, New York (1992)
32. Paulson, C.A.: The mathematical representation of wind and temperature profiles in the unstable atmospheric surface layer. J. Appl. Meteorol. **9**, 857–861 (1970)
33. Pérez-Guerrero, J.S., Pimentel, L.C.G., Oliveira-Júnior, J.F., Heilbron-Filho, P.F.L., Ulke, A.G.: A unified analytical solution of the steady-state atmospheric diffusion equation. Atmos. Environ. **55**, 201–212 (2012)
34. Pleim, J.E., Chang, J.S.: A non-local closure model for vertical mixing in the convective boundary layer. Atmos. Environ. **26A**, 965–981 (1992)
35. Pobedrya, B.E.: Mechanics of Composite Materials. Moscow State University Press, Moscow (1984) (in Russian)
36. Stull, R.B.: An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers, Dordrecht (1988)
37. Talbot, A.: The accurate inversion of Laplace transforms. J. Inst. Math. Appl. **23**, 97–120 (1979)
38. Tartar, L.: The General Theory of Homogenization: A Personalized Introduction. Springer-Verlag, Berlin (2009)
39. Ulke, A.G.: New turbulent parameterization for a dispersion model in the atmospheric boundary layer. Atmos. Environ. **34**, 1029–1042 (2000)

40. Vilhena, M.T., Costa, C.P., Moreira, D.M., Tirabassi, T.: A semi-analytical solution for the three-dimensional advection–diffusion equation considering non-local turbulence closure. Atmos. Res. **90**, 63–69 (2008)
41. Zannetti, P.: Air Pollution Modeling: Theories, Computational Methods and Available Software. Springer, New York (1990)
42. Zilitinkevich, S.: Non-local turbulent transport: pollution dispersion aspects of coherent structure of connective flows. Trans. Ecol. Environ. **6**, 53–60 (1995)

Data Mining Approaches to the Real-Time Monitoring and Early Warning of Convective Weather Using Lightning Data

Stephan Stephany, Cesar Strauss, Alan James Peixoto Calheiros, Glauston Roberto Teixeira de Lima, João Victor Cal Garcia, and Alex Sandro Aguiar Pessoa

1 Introduction

Predicting severe weather events plays an important role in the socio-economic development of a country. The ability of issuing short and medium-term warnings is relevant, in order to mitigate the hazards of heavy precipitation, strong winds, lightning, floods, and landslides. As a consequence, agriculture, telecommunications, transportation, and energy industries suffer frequent economic losses. Above everything, loss of human lives is the most important concern. Long-term predictions belong to climatology and are out of the scope of this work. Throughout the text, predictions generated by a numerical weather prediction (NWP) model are usually referred as forecasts, following its standard name.

Short-term approaches for severe weather prediction (up to 6 h) may provide immediate warnings, as shown, for instance, by real-time thunderstorm nowcasting [3]. An ensemble of techniques is necessary in order to give a fast feedback to the civil defense about the storms occurrence and severity. Such techniques are based on different datasets, mostly related to weather radar and satellite measurements, and high-resolution NWP models. Detailing the environmental condition, storm initiation, and severity are basic requirements to improve the quality and efficiency

S. Stephany (✉) · C. Strauss · A. J. P. Calheiros
National Institute for Space Research (INPE), São José dos Campos, Brazil
e-mail: stephan.stephany@inpe.br; cesar.strauss@inpe.br; alan.calheiros@inpe.br

G. R. T. de Lima · J. V. C. Garcia
National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos, Brazil
e-mail: glauston.lima@cemaden.gov.br; joao.garcia@cemaden.gov.br

A. S. A. Pessoa
Climatempo Meteorologia, São José dos Campos, Brazil
e-mail: asapessoa@gmail.com

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_5

of the convective nowcasting. Forecasting spatio-temporal patterns of meteorological variables helps to define the convective activity over a specific area where the thunderstorm will develop, since thunderstorms develop stronger and longer in high shear and instability conditions [28].

Once the thunderstorm is already acting over a specific area, weather satellite and radar images are more efficient to characterize the storm evolution for a very short-term forecasting (up to 2 h). Studies using geostationary satellite data defined important parameters to predict the intensity of thunderstorms [1, 24, 35, 36, 38]. Recent studies were applied to estimate the lightning initiation using both satellite and radar data [12, 25]. Karagiannidis et al. [15], for instance, presented a lightning activity nowcasting tool based on satellite and lightning network data with a good accuracy (80%), but with a significant false alarm rate (40%).

In [2], the tracking of convective cells was proposed by means of lightning data from the German-developed LINET European lightning detection network. This kind of data can be used for tracking flash rates along time, being important to identify the lightning jump (rapid increase in total lightning activity) which is correlated with the start of severe weather at the ground [39]. Such information is useful for nowcasting lightning activity and severe weather events occurrence over a risk area [5]. Nonetheless, a good performance of these extrapolation techniques was only observed over a very short-term prediction.

Furthermore, weather radar can provide volumetric information about the precipitation and wind field, which is useful to perform the tracking and forecasting of rainy systems with high electrical activity. A review of different nowcasting methods using radar is described by Pierce et al. [34]. Besides that, the coupling of radar and satellite information with lightning data provides additional parameters that improves the tracking and prediction of lightning activity [20, 21, 27, 45, 57] as well. As observed before, a synergy between different sensors provides data that enrich the analysis helping to improve the prediction of lightning strikes [23].

Medium-term approaches for weather prediction (typically from 6 to 72 h) depend primarily on NWP model forecasts. Nevertheless, the convective process is still a challenge for NWP models, mainly because of the lack of observational data or computing power (high spatio-temporal resolution) necessary to describe the process. Rainfall or strong winds associated with convective events are difficult to predict if such conditions were not attended. The numerical model itself may inaccurately simulate some meteorological variables, what is called model biases. Therefore, each NWP model presents its own biases that depends on forecast time. A great effort has been done to improve the NWP models skills using data assimilation. Sun [50] revised some techniques applied to convective-scale data assimilation using weather radar. Wang et al. [56] showed a significant improvement on the convection forecasting using the assimilation of lightning and radar reflectivity. Vendrasco et al. [54] used three-dimensional variational data assimilation (3DVAR) technique to assimilate radar information on the weather research and forecasting (WRF) model [44]. They found an improvement on the precipitation forecasting using radar data assimilation with large-scale analysis constraint. However, considering current NWP models employed in Brazil, the skill

for predicting rainfall is still poor, less than 30% for weak and less than 10% for heavy rainfall associated with convective events for 36 h forecasts.

Accurate weather monitoring and prediction requires the prompt analysis of huge amounts of data, generated by different sensors and by NWP models, making the application of automatic tools a necessary and current trend in meteorological centers [34]. Considering this scenario, this work proposes some alternative methods to enhance such monitoring and prediction using lightning data. Such methods are intended to complement standard approaches based on satellite and ground-based weather radar images, besides NWP models. They were developed by the National Institute for Space Research (INPE), and proposed to be used in Brazil, but may apply to other developing countries that lack a complete weather radar coverage, but have lightning detection networks, which demand lower acquisition and maintenance costs than the weather radar ones. These methods follow a data mining approach, which can be thought of as obtaining useful information from available data in the considered scope. It embeds a variety of statistical, pattern-recognition, and machine learning algorithms. In the case of weather monitoring, the methods shown here allow better visualization and analysis of the spatio-temporal evolution of lightning occurrences, identification of lightning clusters, and estimation of the associated rainfall. Weather prediction is tackled with machine learning algorithms applied to NWP model data.

The first method presented here for weather monitoring was shown in Strauss et al. [48]. It is a method based on kernel density estimation to process lightning data (EDDA, atmospheric discharge density estimator, in Portuguese). Assuming that cloud-to-ground (CG) lightning can be correlated to convective activity, Gaussian kernel density estimation is applied to generate a field of occurrence of CG lightning for consecutive periods of time. Instead of visualizing individual discharges, which are sparse in space and time, a smooth field is estimated making easy to visualize clusters of CG lightning occurrences. This is a well-known technique [45, 53] and was implemented by the EDDA software. Considering that lightning data acquired from a detection network is available in minutes, the lightning field of occurrence yielded by EDDA allows to depict in near real-time where convective activity is developing, providing an ancillary tool to the meteorologists.

Another weather monitoring method employs the same kernel density estimation approach to perform a spatio-temporal clustering using a temporal sliding-window [47, 49]. It was implemented by the EDDA-G software (G stands for clustering), providing a list of CG lightning clusters that is periodically updated. This list includes the coordinates of the cluster centroid, the number of CG lightning occurrences, the area covered by these CG discharges, among other attributes. It can be used to trigger a high-resolution numerical model in order to perform nowcasting for a longer period. A third method based on kernel density estimation of lightning data performs the estimation of convective rainfall, by means of a temporal sliding-window and a fitting function, as implemented by the EDDA-CHUVA software ("chuva" means rain in Portuguese). This estimation of the precipitated mass allows to evaluate possible impacts related to flooding or landslides.

In addition to the three weather monitoring methods, two weather prediction methods are also proposed. Data mining algorithms can be trained using NWP model data and lightning data yielding specific data mining models. These models can then be used to predict the occurrence of convective activity from NWP model forecasts. Such predictions may complement the forecasts themselves, helping meteorologists to improve the accuracy of early warnings and forecastings. Both weather prediction methods employ machine learning algorithms, the first is based in the rough set theory, while the second is an artificial neural network. They sweep the forecasts generated by a NWP model and classify the values of the meteorological variables for each pixel of the considered area as presenting or not convective activity. The output of such classifiers is a map depicting pixels that may present convective activity for that given NWP model forecast.

It is important to stress that the current chapter constitutes a short survey on former works of the authors at INPE using data mining methods applied to meteorological data [8, 9, 18, 19, 32, 33, 47–49]. Therefore, the methods shown here are a brief summary of these former works, being intended to constitute ancillary tools for weather monitoring and forecasting. As already mentioned, they are intended to be used mainly in developing countries exploiting low cost lightning detection networks, but may be useful in any country. The research initiated from these former works is still underway and will continue.

The following sections are briefly described as follows: Section 2 covers some methods formerly proposed by the authors for monitoring convective activity by means of lightning data, including the generation of fields of density of occurrence of lightning, the spatio-temporal clustering of lightning, and the estimation of the convective rainfall mass from such data. Section 3 shows two machine learning methods to predict convective activity from numerical model forecasts, one employs a rough set based classifier, while the other is a neural network. Final remarks appear in Sect. 4, followed by the references.

2 Methods for Monitoring Convective Activity

This section describes the methods proposed in the scope of near real-time monitoring of convective activity by means of the lightning occurrences: (1) visualization of these events using Gaussian kernel density estimation, which is performed by the EDDA software, (2) the spatio-temporal clustering of such occurrences, implemented by the EDDA-G software, and (3) the mapping of the number of occurrences to the rainfall mass by a fitting function, as implemented by the EDDA-CHUVA software.

2.1 Gaussian Kernel Density Estimation

The kernel density estimation is used to generate a field of density of occurrence of CG lightning from a set of lightning data composed of individual strokes for the defined time interval and area. The visualization of this field allows to identify electrically active cells that are associated with convective events. The kernel estimator yields the density of lightning events at a grid point $\mathbf{x}$ that results from a set of events at points $\mathbf{x_i}$ [41, 42]. The probability density $\hat{f}(\mathbf{x})$ is then given by

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{d(\mathbf{x}, \mathbf{x_i})}{h}\right) \tag{1}$$

where

$\mathbf{x}$ is the grid point coordinates (latitude, longitude),
$\mathbf{x_i}$ is the i-th lightning stroke coordinate,
n is the number of lightning occurrences,
$d(\mathbf{x}, \mathbf{x_i})$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x_i}$,
h is the defined window width,
$K(r)$ is the kernel function, and r the normalized distance $d(\mathbf{x}, \mathbf{x_i})/h$.

Here, the kernel function is given by the 2D Gaussian with zero mean and unity standard deviation

$$K(r) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} r^2\right) \tag{2}$$

The window width h can be calculated automatically in order to minimize the estimation error [42], but is empirically set in consonance of the scale of the convective event associated with the lightning [48]. The resulting field of density of occurrence is convenient since it deals better with lightning data, which is very sparse in space and time. However, electrically active cells may not persist from one time interval to the next. The choice of a convenient time interval is another issue concerning the adjustment of the density field to the scale of the observed event. The EDDA software allows to choose the parameters that define the resulting field like the time interval, window width, etc.

An integrated lightning detector network is composed of sensors that detect the spherical electromagnetic wave emitted by each lightning stroke. The position of the stroke is obtained by triangulation from the times of arrival of the wave at each sensor, recorded using a GPS time reference. Some type of sensors also detect the azimuth of the stroke, allowing to calculate the position more accurately. Lightning strokes are classified as being cloud-to-ground, ground-to-cloud, intracloud, cloud-to-cloud, and cloud-to-air. Current Brazilian lightning detection networks detect only ground-to-cloud and cloud-to-cloud lightning. In particular, this work employs lightning data acquired by the RINDAT network that stands for (in Portuguese)

national integrated network for detection of atmospheric discharges [29]. The ELAT (Atmospheric Electricity Group) of INPE (Brazilian Institute for Space Research) supplies lightning data to CPTEC/INPE (Center for Weather Prediction and Climatic Studies). Only cloud-to-ground lightning was considered in the monitoring and prediction of convective activity, based on a statistical correlation between this type of lightning and convective activity [17]. However, the proposed approaches would be enhanced with the use of intracloud lightning data, since this type of lightning precedes storms and its detection may be used as storm early warning. In addition, the ratio between intracloud to cloud-to-ground lightning may be used to classify the convective storm type [16]. RINDAT provides data recording in text files each lightning stroke in the UALF format (Universal ASCII Lightning Format). Each record corresponds to an individual stroke and shows its latitude, longitude, time of occurrence, polarity, multiplicity (in case of flashes composed of multiple strokes), and type (only cloud-to-cloud or cloud-to-ground).

2.2 Spatio-Temporal Clustering Based on Sliding-Window

Electrically active cells can be detected, identified, and tracked by the spatio-temporal clustering of lightning strokes, which allows the monitoring of their evolution along time. It is also possible to determine all lightning stroke occurrences of a given cluster and to obtain cluster parameters as the number of strokes, position of the cluster centroid, estimated area, or stroke rate. Clustering is defined as the process of grouping data in classes or clusters [11]. Similar objects are grouped together according to some criteria. In this case, lightning strokes are grouped using their positions and times of occurrence. The proposed clustering method is based on a temporal sliding-window, similarly to sliding-windows employed for data flow control in computer networks. A fixed-width temporal window slides in discrete time intervals with constant rate screening incoming lightning data. The resulting clusters correspond to the electrically active cells, which are associated with intense lightning. The clustering process is performed at every timestep of the sliding-window and must identify and keep track of all clusters for the considered area. Eventually, from one timestep to the next, clusters may merge, split into new ones, or disappear. The advance of the sliding-window was selected as half timestep, thus refreshing all strokes after two advances of the sliding-window. This behavior appears in an example shown in Fig. 1, which depicts four different clusters evolving from one timestep to the next.

A spatio-temporal clustering of lightning strokes was implemented in the EDDA-G software, which employs as clustering algorithm the 2D kernel-based DENCLUE 2.0 [14]. This algorithm finds local maxima of the density field by using an optimized gradient climbing algorithm, where the starting seed of the search is the position of a single lightning stroke. The location of the local maximum found by this search is then recorded as an attribute of this lightning stroke. The search is repeated several times, each time starting the search at the position of a different

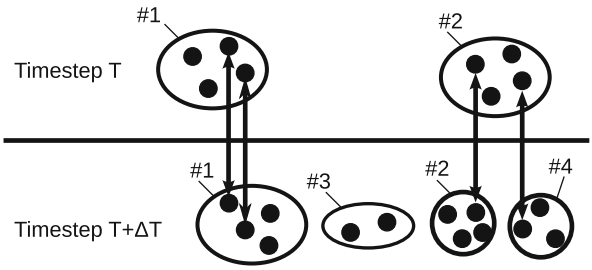


Fig. 1 Lightning stroke clusters evolving from one timestep to the next: cluster #1 persists, while cluster #2 is split into two giving raise to cluster #4, while cluster #3 is new (extracted from [47])

lightning stroke. After all lightning strokes are used as seeds, the clustering can be done. If two local maxima associated with two lightning strokes are close together, then the lightning strokes are assigned to the same electrically active cell. Multi-scale results can be obtained by varying the window width that governs the smoothing of the density field. EDDA-G outputs a text list of clusters and corresponding parameters. The EDDA software of the preceding section allows to visualize the field of density of occurrence of lightning strokes, while the EDDA-G software allows to extract parameters that help to evaluate the magnitude of the convective activity associated with the electrically active cells.

An example of the use of the EDDA and EDDA-G softwares appears in Fig. 2 for a thunderstorm occurred in January 16th 2010 20:10 UTC in the State of São Paulo. This figure shows a map with the convective and stratiform structures derived from radar reflectivity according to the Steiner criteria. The density of occurrence of CG lightning was generated by the EDDA software and a threshold of 1 stroke/degree^2/minute was applied to obtain contour lines corresponding to the electrically active cells. Clustering was performed by the EDDA-G software allowing to obtain different contour lines, corresponding to smaller cells that express more precisely their area.

2.3 Estimation of Convective Rainfall Mass from Lightning Data

As already mentioned, it is possible to estimate convective rainfall mass from the number of CG lightning strokes in places without weather radar coverage, as proposed in [8, 9], and implemented by the EDDA-CHUVA software, which estimates the mass of convective rainfall from the number of occurrences of CG lightning given by the RINDAT network. A former standard approach is the Tapia model [51], based on a constant rainfall–lightning ratio (RLR) to map both quantities, but the choice of a suitable RLR value may be difficult due to its high variability.

The approach formerly proposed by the authors is based on a temporal sliding-window that screens the occurrences of CG lightning strokes for the given area.

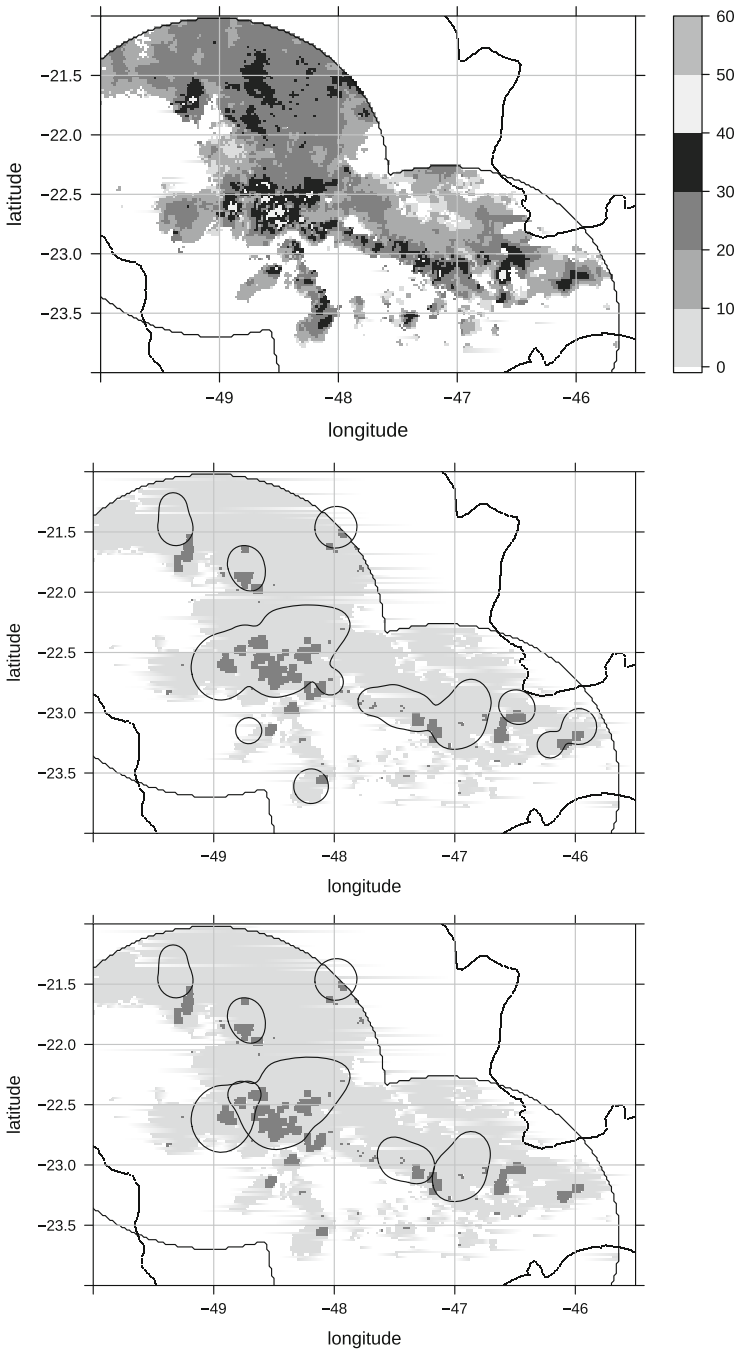


Fig. 2 Radar reflectivity (dBZ) of a thunderstorm occurred in January 16th 2010 20:10 UTC in the State of São Paulo (top), the corresponding electrically active cells detected solely by the EDDA software, using kernel density estimation (middle), and by the EDDA-G software, that performs clustering (bottom). Gray shades correspond to stratiform precipitation and black shades, to convective precipitation. Contour lines show electrically active cells for a given density threshold, while circles refer to the 150 km range of the Bauru and São Roque weather radars (extracted from [47])

The assumption that convective activity is correlated to electrically active cells is made. A WRLR (windowed-RLR) function is derived in the training phase by fitting a set of data points. Each point is given by the rainfall mass and the number of CG strokes for the time interval corresponding to the window width. Rainfall is estimated from weather radar data using a suitable Z-R relationship, which maps the reflectivity Z (i.e., the energy backscattered by hydrometeors) to the rainfall rate (in mm/h). Outliers corresponding to data points with very high precipitated mass were removed using the Tukey–Kramer method [52]. Obviously, since the WRLR function is derived in an area with radar coverage, it would be worth to estimate rainfall from lightning strokes in areas with a similar climate. It can happen to have data points with the same number of strokes, but with different values of rainfall mass. This is dealt with by considering the average rainfall mass for these data points.

Obviously, the correlation between lightning and convective precipitation is more significant than it is to stratiform precipitation [43]. Stratiform rainfall was filtered out using the criterion proposed in [46]. In the considered weather radar grid, this criterion marks a grid point as presenting convective rainfall if (1) it has reflectivity of at least 40 dBZ or (2) if it presents a significant gradient of reflectivity, above a certain threshold for a circle around it, and also all grid points inside the circle. Values of the threshold and the radius of the circle depend on the background intensity.

An example of such fitting function follows. Data points were derived for the entire year of 2009 in an area under the coverage of two Brazilian S-band weather radars located in the State of Sao Paulo (Brazil) at the cities of Bauru and Presidente Prudente. This area was divided into 32 squares with 50 km sides that are within the useful range of 150 km of both radars. The set of data points used to derive the fitting function is composed of the data points obtained for the 32 squares by the temporal sliding-window. Figure 3 shows the set of data points and the particular fitting function, which is described in Eq. 3.

$$\mathrm{WRLR}(N) = 941.3 \times N^{0.3878} - 182.1 \tag{3}$$

Fitting curves were derived for a set of weather radars for each season of the year considering 50 km × 50 km squares. These curves were then mapped to a grid of squares of the same size covering Brazil by specific constants (K's). Each K is given by the ratio of the amount of convective rainfall in the considered square and the one in the square over the weather radar. Outside weather radar coverage, rainfall masses were estimated from a network of pluviometers, with the addition of satellite rainfall estimates in areas of low density of pluviometers. The amount of convective rainfall was inferred from an assumed ratio between convective and stratiform rainfall masses. The EDDA-CHUVA software also supplies an estimate of the spatial distribution of convective rainfall that is based on the distribution of CG lightning itself, but using a coarser spatial resolution (5 km), since lightning and convective rainfall are not exactly coincident in space and time. As an example, a particular thunderstorm occurred in 22th April 2010 over the Presidente Prudente

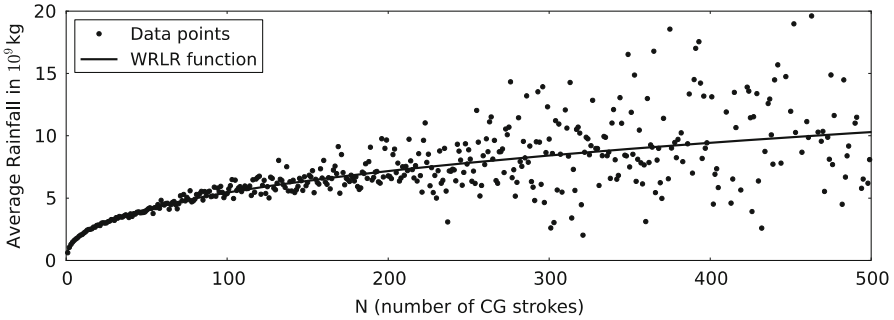


Fig. 3 Scatterplot of the data points employed in the training phase and the corresponding WRLR fitting function (extracted from [8])

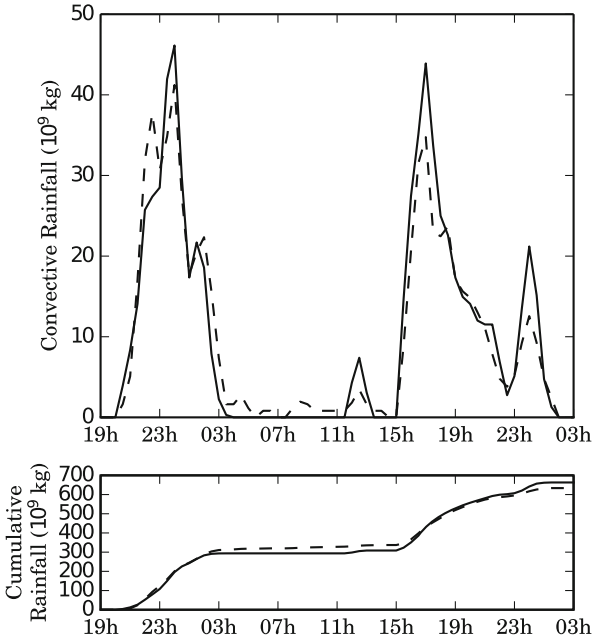


Fig. 4 Estimates of convective rainfall mass given by weather radar (solid line) and by the EDDA-CHUVA software (dashed line) for the thunderstorm occurred in 22nd April 2010 in Presidente Prudente. Top image shows the curve for 30 min accumulated rainfall mass, and bottom image, the cumulative rainfall mass, both in UTC time (extracted from [8])

weather radar was considered. Figure 4 compares the estimated rainfall by that weather radar and the EDDA-CHUVA software for a 50 km × 50 km square over the radar along the 32 h of such storm. The software yielded estimates of rainfall every 30 min that are closer to those obtained from the radar.

3 Machine Learning Methods for Predicting Convective Activity

Numerical models are able to perform accurate forecasts, but the ability of predicting convective activity still needs to be improved. The methods shown here were formerly proposed by the authors. They employ numerical model forecasts, but predicts the occurrence of convective activity using data mining models. Data mining algorithms are trained with pre-processed meteorological data, yielding data mining models, which in turn are used to predict such occurrences. In this scope, two classifiers were proposed for the prediction of convective activity: one based on the rough set theory (RST) and an artificial neural network. These classifiers are machine learning algorithms presented in former works of the authors.

It is assumed that the numerical model simulates adequately most of the atmospheric variables and that some patterns can be derived from the values of these variables as being associated with the presence of convective activity. In the same way, another set of patterns can be derived, associated with the absence of convective activity. The training phase employs analysis and forecast model data. Analysis model data is obtained by a data assimilation process that updates model variables using observational data, which is available for a much lower number of points than those of the model. Forecast model data is obtained by executing the model from the most recent analysis, i.e., using it as the set of initial conditions. The resulting database contains model data for several months or years, typically for every 6 h of simulated time (6, 12, 18, and 24 UTC). Model data for each output comprehend the values of hundreds of meteorological variables for each pixel of the model grid, covering the selected area. Many variables are calculated for several levels of atmospheric pressure. The resulting database is a collection of instances/samples. Each one is given by the values of the meteorological variables at a defined pixel and defined time. In the jargon of machine learning, these are the information attributes. However, in order to perform the training of the classifier, each instance must have a decision attribute, which characterizes that instance as presenting convective activity or not. The information attributes are numerical, and the decision attribute was taken as categorical, as being SCA (severe convective activity) or NSCA (non-SCA). The latter encompasses moderate, weak, and absent convective activity. Depending on the chosen classifier, information attributes can be taken as categorical, or the decision attribute, as numerical.

A key point is how to assign the decision attribute to each instance, since such task would require observational data that is not available outside the coverage of weather radars. In this approach, an instance is labeled as SCA or NSCA according to the value of the density of occurrence of CG lightning strokes in the corresponding pixel and for the strokes computed for that pixel during an interval centered at the time of the instance. A convenient threshold of density is chosen to separate both classes (SCA and NSCA). The EDDA software presented in Sect. 2.1 generates the density of occurrence of lightning using RINDAT data. Another option is to employ weather radar data if available, which would enhance

the accuracy of the method since the weather radar measures the reflectivity of the convective rainfall. Rainfall expresses better the time and location of a convective cell than the associated lightning strokes, since there is a discrepancy in space and time between the rainfall and the lightning of a cell. This method is intended for regional atmospheric models, more specifically the ETA model [26] that is employed operationally at CPTEC/INPE. However, it could be applied to any other regional model. A set of 26 or 56 model variables was selected by a group of meteorologist, but according to the chosen classifier, attribute reduction was performed, in order to use less variables and to reduce the algorithmic complexity of the problem. The coordinates of the model grid point corresponding to the pixel and the time were not taken as information attributes.

In the training phase, the chosen algorithm "learns" the patterns associated with both SCA and NSCA events, while in the test phase, it will classify each instance by matching it against these two sets of patterns. Therefore, lightning data is required only during the training phase, and model forecasts will be screened for matching patterns. Another issue in the training phase is to ensure a balance between the number of SCA and NSCA instances, since the former are seldom in comparison to the latter. All SCA instances can be considered, while the double of NSCA instances is randomly selected. This avoids a training that results in a classifier with a bias to classify every instance as being NSCA. The performance of the classifier is given by its ability of correctly classify SCA instances as being so (true positives or TP), and the same for the NSCA ones (true negatives or TN). False positives (FP) correspond to NSCA instances incorrectly classified as being SCA, and the opposite, give raise to the false negatives (FN). The latter are more harmful, since they represent a SCA event that goes undetected. The training process can be repeated as new data is made available in order to enhance the accuracy of the classifier, since increasing the number of instances of the database tends to improve the training performance. Sazonal or regional aspects can be tackled with specific training sets.

The results can be evaluated by means of a confusion matrix, which contains in its diagonal the number of instances correctly classified for each class, while out of diagonal numbers are misclassified instances. Each instance refers to a specific grid point of a weather forecast for the considered area. There are also some standard metrics for the evaluation of classifiers. A very common pair of metrics is given by the probability of detection (POD) and FAR (false alarm ratio). Also known as hit ratio, POD is defined as $TP/(TP + FN)$ and FAR as $FP/(TP + FP)$.

The following sections describe the two classifiers that were proposed for the prediction of convective activity using model data, the first one based on the rough set theory (RST) and the second one, an artificial neural network.

3.1 Rough Set Theory Classifier

The first classifier presented here is based on the rough sets theory (RST), which emerged in the 1980s to deal with uncertain, incomplete, or vague information

[30, 31]. It has a good mathematical formalism, and it is easy to use, since it does not require additional information such as the probability distribution, a priori probability, or pertinence degree. However, free RST software is not available. RST is an extension of the set theory and has the implicit feature of compressing the dataset. Such compression is due to the definition of equivalence classes based on indiscernibility relations and to the elimination of redundant or meaningless attributes. A central concept in RST is attribute reduction, which generates reducts. A reduct is any subset of attributes that preserves the indiscernibility of the elements for the considered classes, allowing to perform the same classification that would be obtained with the full set of attributes. RST can preprocess data obtaining reducts for other machine learning techniques. The typical high dimensionality of current databases precludes the use of greedy methods embedded in RST to find optimal or suboptimal reducts in the search space, requiring the use of stochastic methods.

In [32, 33], three alternative metaheuristics were proposed to calculate the reductions, with good results. Two are well-known metaheuristics: variable neighborhood search (VNS) and variable neighborhood descent (VND), and the third is new one called decrescent cardinality search (DCS). VNS generates a random candidate solution (or reduct) with any cardinality and at each iteration explores its neighborhood searching for new solutions that are then evaluated. VND is an extension of VNS, where the search for solutions is performed following a deterministic approach. DCS is a modified version of VNS that explores the search space accepting only lower cardinality reducts. It enforces randomly a cardinality reduction at each iteration before performing a local search. These metaheuristics were applied to benchmark classification problems, but were later applied to the prediction of convective activity [32], jointly with the random partition of the training set. Such partition allows to split the set of training data into smaller training sets, expecting that different patterns associated with SCA and NSCA events will be identified in each partition. Tests were performed using data from the ETA regional model, with a 20 km resolution. A pre-selection of attributes guided by meteorological expertise resulted in the choice of 58 ETA variables, 9 at surface level and 7 others at multiple levels (7 selected levels). Depending on the choice of the metaheuristic, the dimensionality of the resulting reducts was between 12 and 17, corresponding to the number of attributes employed for the training of the RST classifier.

In particular, as shown in [32], instead of considering only classes SCA and NSCA, convective activity events were divided into three classes, SCA (severe), MCA (moderate), and NSCA (weak/null). Table 1 shows classification results for the best and worst executions (out of 10 executions), expressed by three confusion matrices resulting of tests with 24, 48, and 72 h forecasts of the ETA model for a square with one-degree side in the State of São Paulo, embedding the cities of Bauru and Presidente Prudente. Each column shows the predicted number of instances as being of the column-class, while each line shows the actual number of instances of the line-class. In short, in the best execution, around 2/3 of the MCA and SCA instances were correctly predicted, and 99% of the NSCA ones. POD/FAR values for the tests with 24, 48, and 72 h forecasts were, respectively, 0.680/0.162,

Table 1 Confusion matrices for the best (3 first lines) and worst (3 last lines) executions obtained by the RST-based classifier for tests using 24, 48, and 72 h forecasts of the ETA 20 km model (extracted from [32])

	Predicted (24 h)			Predicted (48 h)			Predicted (72 h)		
	NSCA	MCA	SCA	NSCA	MCA	SCA	NSCA	MCA	SCA
Actual									
NSCA	7218	77	3	7080	91	2	6951	68	3
MCA	233	561	27	244	514	26	248	545	23
SCA	27	46	155	49	39	158	65	28	142
Actual									
NSCA	7201	85	12	7080	90	3	6926	92	4
MCA	303	488	30	332	425	27	326	465	25
SCA	53	36	139	71	39	136	83	24	128

0.642/0.151, and 0.604/0.155. Concerning the worst execution, POD/FAR values for the 24, 48, and 72 h forecast were, respectively, 0.772/0.390, 0.819/0.447, and 0.815/0.455.

3.2 Neural Network Classifier

The second classifier presented here is a neural network. Neural networks have a widespread employ in diverse areas of research [13, 37]. The classification performance of any neural network depends strongly on its architecture. After attempting to use an *ad hoc* algorithm based on the frequency of occurrence of values of pairs of variables [18], a neural network was conceived in [19] exploiting a specific architecture. Differently from the case of the RST classifier, a second preselection of attributes was performed using meteorological expertise, which resulted in the choice of only 26 ETA variables. The proposed network architecture connects each of the 325 possible pairs that result in the combination of the 26 inputs to a particular neuron in the sole hidden layer. The net input for each hidden neuron is then given by the values of its two input values multiplied by the respective weights plus its bias. This architecture that partially connects input nodes and hidden neurons was chosen in order to seek the information contained in pairs of variables. The nonlinear mapping is provided by the bipolar sigmoid activation function in the hidden layer. The resulting 325 hidden layer activations are then multiplied by the respective weights and summed up to provide the network output using a linear activation function. However, weights are adjusted differently for each class, resulting in two outputs, one that expresses the pertinence to class SCA and the other to the class NSCA, as shown in Fig. 5.

Tests were also performed using data from the ETA regional model, with a 20 km resolution, as presented in [19]. Some partial results of these tests are shown in Table 2, with the classification performance expressed by two confusion matrices (one for the training and another for the test) for a square with one-degree side in

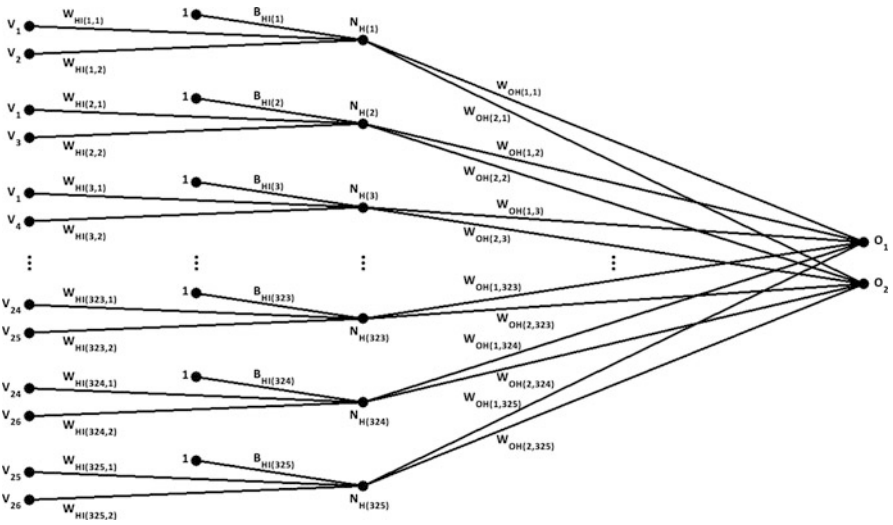


Fig. 5 Training architecture of the proposed neural network with the 650 input nodes given by the pairs of the 26 ETA variables V_i's, the 325 hidden layer nodes $N_{H(i)}$'s, and 2 outputs O_1 and O_2. The W_{HI}'s and W_{OH}'s denote the weights and the B_{HI}'s, the biases with unity input

Table 2 Training and test confusion matrices obtained by the neural network classifier for the ETA 20 km model for a 48 h forecast

	Predicted NSCA	Predicted SCA	Predicted NSCA	Predicted SCA
Actual				
NSCA training	506	44	498/520	30/52
SCA training	25	250	20/39	236/255
NSCA test	44	2	42/45	1/4
SCA test	2	21	0/4	19/23

Left matrix shows average values for 20 executions, while right matrix shows minimum/maximum values for these executions

the State of São Paulo, embedding the cities of Bauru and Presidente Prudente. As in the preceding table, each column shows the predicted number of instances as being of the column-class, while each line shows the actual number of instances of the line-class. That table shows that most instances were correctly classified in this two-class scheme. A 2:1 sampling between NSCA and SCA instances was applied, in order to eliminate the bias towards NSCA classifications in the training, since these are more numerous. It can be seen in the test results that 44 of the 46 NSCA instances were correctly predicted, as well as 21 of the 23 SCA ones, considering the average of 20 executions of the classifier. This result corresponds to a POD/FAR of 0.909/0.149 in the training and 0.913/0.087 in the test. The total number of instances employed in the training and test for the results shown in Table 2 was 41,415 of the NSCA class and 1,029 of the SCA class.

4 Final Remarks

As already mentioned before, this book chapter constitutes a short survey on former works of the authors using data mining techniques applied to meteorological data. The methods proposed for monitoring convective activity were passed to two Brazilian governmental institutes, CPTEC/INPE and CEMADEN, the National Center for Monitoring and Early Warning of Natural Disasters. The methods proposed for predicting convective activity are currently undergoing tests using another regional model used by CPTEC, the BRAMS model with a 5 km resolution [7]. The results that appear in the references [9, 19, 33, 49] show that such methods are potentially good in the proposed tasks, but only the operational implementations at these institutes will confirm their usefulness.

Besides, some of those methods shown before are able to be implemented to support other techniques, improving the nowcasting system operating on CPTEC. The EDDA-CHUVA can provide rainfall field estimation to the Hydroestimator Tracking and Nowcaster (Hydrotrack), a nowcasting method based on precipitation input [4]. In addition, the ForTraCC (Forecasting and Tracking of Cloud Cluster) [55] can assimilate the EDDA-G outputs to nowcasting the lightning activity, including the monitoring of lightning jump. With the recent launch of GOES-16 (geostationary operational environmental satellites) Lightning Mapper (GLM, [10]), optical total lightning measurements will be provided continually. The coupling of EDDA-G to ForTraCC can be used to track and extrapolate the lightning filed from GLM data, as proposed in a similar approach using proxy data [40].

Furthermore, fields campaigns over Brazil, like the CHUVA (Cloud processes of tHe main precipitation systems in Brazil: A contribUtion to cloud resolVing modeling) experiment [22] (or another experiment, the SOS-CHUVA) provides a richness of data allowing the application of new methods based on machine learning, as well as to carry out detailed lightning structure studies [6], in order to improve the thunderstorm nowcasting.

Finally, the use of multi-sensorial data seems to be a current trend in meteorology, as well as the correspondent information extraction by means of data mining techniques like the ones presented or cited in this work. The standard aim of data mining is to extract implicit, previously unknown, and potentially useful patterns from data, which in this case are associated with the occurrence of severe weather. The task of issuing weather forecasts and early warnings of severe events requires a complex and timely analysis of data from weather satellites and radars, in addition to NWP models and lightning data. Such task can be alleviated by the implementation of these automatic techniques in the operational environment of weather forecast.

Acknowledgements Author Stephan Stephany thanks CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for grant 307460/2015-0.

References

1. Adler, R.F., Fenn, D.D.: Thunderstorm intensity as determined from satellite data. J. Appl. Meteorol. **18**(4), 502–517 (1979)
2. Betz, H.D., Schmidt, K., Oettinger, W.P., Montag, B.: Cell-tracking with lightning data from LINET. Adv. Geosci. **17**, 55–61 (2008)
3. Browning, K.A.: Review lecture: local weather forecasting. In: Proceedings of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences, vol. 371, pp. 179–211. The Royal Society, London (1980)
4. Calheiros, A.J.P., Machado, L.A.T.: The HydroTrack: a nowcasting application using GOES data. In: Current Problems in Atmospheric Radiation (IRS 2008): Proceedings of the International Radiation Symposium (IRC/IAMAS), vol. 1100, pp. 361–364. AIP Conference Proceedings (2009)
5. Chronis, T., Carey, L.D., Schultz, C.J., Schultz, E.V., Calhoun, K.M., Goodman, S.J.: Exploring lightning jump characteristics. Weather Forecast. **30**(1), 23–37 (2015)
6. Chronis, T., Lang, T., Koshak, W., Blakeslee, R., Christian, H., McCaul, E., Bailey, J.: Diurnal characteristics of lightning flashes detected over the São Paulo lightning mapping array. J. Geophys. Res. Atmos. **120**(23) (2015)
7. Freitas, S., et al.: The Brazilian developments on the regional atmospheric modeling system (BRAMS 5.2): an integrated environmental model tuned for tropical areas. Geosci. Model Dev. **10**(5), 189–222 (2017)
8. Garcia, J.V.C.: Monitoring and prediction of convective events using data mining approaches. Ph.D. thesis, Applied Computing Post-graduate Program, INPE, Brazil (2014)
9. Garcia, J.V.C., Stephany, S., D'Oliveira, A.B.: Estimation of convective precipitation mass from lightning data using a temporal sliding-window for a series of thunderstorms in Southeastern Brazil. Atmos. Sci. Lett. **14**, 281–286 (2013)
10. Goodman, S.J., Blakeslee, R.J., Koshak, W.J., Mach, D., Bailey, J., Buechler, D., Carey, L., Schultz, C., Bateman, M., McCaul, E., et al.: The GOES-R geostationary lightning mapper (GLM). Atmos. Res. **125**, 34–49 (2013)
11. Han, J., Kambler, M.: Data Mining – Concepts and Techniques, 3 edn. Elsevier, New York (2011)
12. Harris, R.J., Mecikalski, J.R., MacKenzie Jr, W.M., Durkee, P.A., Nielsen, K.E.: The definition of GOES infrared lightning initiation interest fields. J. Appl. Meteorol. Climatol. **49**(12), 2527–2543 (2010)
13. Haykin, S.O.: Neural Networks and Learning Machines, 3 edn. Pearson/Prentice-Hall, Inc., Upper Saddle River (2008)
14. Hinneburg, A., Gabriel, H.H.: DENCLUE 2.0: fast clustering based on kernel density estimation. In: Berthold, M.R., Shawe-Taylor, J., Lavraĉ, N. (eds.) Advances in Intelligent Data Analysis VII. Lecture Notes in Computer Science, vol. 4723, pp. 70–80. Springer, Berlin (2007)
15. Karagiannidis, A., Lagouvardos, K., Kotroni, V.: The use of lightning data and Meteosat infrared imagery for the nowcasting of lightning activity. Atmos. Res. **168**, 57–69 (2016)
16. Lang, T.J., Rutledge, S.A.: Relationships between convective storm kinematics, precipitation, and lightning. Mon. Weather Rev. **130**, 2492–2506 (2002)
17. Lang, T.J., Rutledge, S.A.: A framework for the statistical analysis of large radar and lightning datasets: results from STEPS 2000. Mon. Weather Rev. **139**(8), 2536–2551 (2011)
18. Lima, G.R.T., Stephany, S.: A new classification approach for detecting severe weather patterns. Comput. Geosci. **57**, 158–165 (2013)
19. Lima, G.R.T., Stephany, S.: Training a neural network to detect patterns associated with severe weather. Learn Nonlinear Models **11**, 123–152 (2013)
20. López, R.E., Aubagnac, J.P.: The lightning activity of a hailstorm as a function of changes in its microphysical characteristics inferred from polarimetric radar observations. J. Geophys. Res. Atmos. **102**(D14), 16,799–16,813 (1997)

21. Lund, N.R., MacGorman, D.R., Schuur, T.J., Biggerstaff, M.I., Rust, W.D.: Relationships between lightning location and polarimetric radar signatures in a small mesoscale convective system. Mon. Weather Rev. **137**(12), 4151–4170 (2009)
22. Machado, L.A.T., Silva Dias, M.A.F., Morales, C., Fisch, G., Vila, D., Albrecht, R., Goodman, S.J., Calheiros, A.J.P., Biscaro, T., Kummerow, C., et al.: The CHUVA project: How does convection vary across Brazil? Bull. Am. Meteorol. Soc. **95**(9), 1365–1380 (2014)
23. Matthee, R., Mecikalski, J.R., Carey, L.D., Bitzer, P.M.: Quantitative differences between lightning and nonlightning convective rainfall events as observed with polarimetric radar and MSG satellite data. Mon. Weather Rev. **142**(10), 3651–3665 (2014)
24. Mecikalski, J.R., Bedka, K.M., Paech, S.J., Litten, L.A.: A statistical evaluation of GOES cloud-top properties for nowcasting convective initiation. Mon. Weather Rev. **136**(12), 4899–4914 (2008)
25. Mecikalski, J.R., Li, X., Carey, L.D., McCaul Jr, E.W., Coleman, T.A.: Regional comparison of GOES cloud-top properties and radar characteristics in advance of first-flash lightning initiation. Mon. Weather Rev. **141**(1), 55–74 (2013)
26. Mesinger, F., et al.: The step-mountain coordinate: model description and performance for cases of Alpine Lee cyclogenesis and for a case of an Appalachian redevelopment. Mon. Weather Rev. **116**(7), 1493–1518 (1988)
27. Meyer, V.K., Höller, H., Betz, H.D.: Automated thunderstorm tracking: utilization of three-dimensional lightning and radar data. Atmos. Chem. Phys. **13**(10), 5137–5150 (2013)
28. Moller, A.R.: Severe local storms forecasting. Meteorol. Monogr. **28**(50), 433–480 (2001)
29. Naccarato, K., Pinto Jr, O.: Improvements in the detection efficiency model for the Brazilian lightning detection network (BrasilDAT). Atmos. Res. **91**(2), 546–563 (2009)
30. Pawlak, Z.: Rough sets. Int. J. Comput. Inf. Sci. **11**(5), 341–356 (1982)
31. Pawlak, Z.: Rough Sets – Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
32. Pessoa, A.S.A.: Prediction of severe events from meteorological model outputs employing the Rough Sets Theory and metaheuristics for attribute reduction. Ph.D. thesis, Applied Computing Post-graduate Program, INPE, Brazil (2014)
33. Pessoa, A.S.A., Stephany, S.: An innovative approach for attribute reduction in Rough Set Theory. Intell. Inf. Manag. **06**, 223–239 (2014)
34. Pierce, C., Seed, A., Ballard, S., Simonin, D., Li, Z.: Nowcasting. In: Bech, J. (ed.) Doppler Radar Observations – Weather Radar, Wind Profiler, Ionospheric Radar, and Other Advanced Applications. InTech, London (2012). https://doi.org/10.5772/39054. https://www.intechopen.com/books/doppler-radar-observations-weather-radar-wind-profiler-ionospheric-radar-and-other-advanced-applications/nowcasting
35. Reynolds, D.W.: Observations of damaging hailstorms from geosynchronous satellite digital data. Mon. Weather Rev. **108**(3), 337–348 (1980)
36. Roberts, R.D., Burgess, D., Meister, M.: Developing tools for nowcasting storm severity. Weather Forecast. **21**(4), 540–558 (2006)
37. Samarasinghe, S.: Neural Networks for Applied Sciences and Engineering: from Fundamentals to Complex Pattern Recognition. Auerbach Publications, New York (2006)
38. Schmetz, J., Tjemkes, S.A., Gube, M., Van de Berg, L.: Monitoring deep convection and convective overshooting with METEOSAT. Adv. Space Res. **19**(3), 433–441 (1997)
39. Schultz, C.J., Petersen, W.A., Carey, L.D.: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. J. Appl. Meteorol. Climatol. **48**(12), 2543–2563 (2009)

40. Schultz, E., Schultz, C.J., Carey, L.D., Cecil, D.J., Bateman, M.: Automated storm tracking and the lightning jump algorithm using GOES-R Geostationary Lightning Mapper (GLM) proxy data. J. Oper. Meteorol. **3**(1), 1–7 (2016)
41. Scott, D.W.: Multivariate Density Estimation – Theory, Practice and Visualization. John Wiley & Sons, Inc., New York (1992)
42. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
43. Sist, M., Zauli, F., Biron, D., Melfi, D.: A study about the correlation link between lightning data and meteorological data. In: 2010 EUMETSAT Meteorological Satellite Conference, Córdoba, vol. 1 (2010)
44. Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.Y., Wang, W., Powers, J.G.: A description of the advanced research WRF version 3. Tech. Rep. NCAR/TN-4751STR, NCAR (2008)
45. Steinacker, R., Dorninger, M., Wölfelmaier, F., Krennert, T.: Automatic tracking of convective cells and cell complexes from lightning and radar data. Meteorol. Atmos. Phys. **72**(2), 101–110 (2000)
46. Steiner, M., Houze Jr., R.A., Yuter, S.E.: Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. J. Appl. Meteorol. **34**(9), 1978–2007 (1995)
47. Strauss, C.: Monitoring and prediction of convective events using data mining approaches. Ph.D. thesis, Applied Computing Post-graduate Program, INPE, Brazil (2013)
48. Strauss, C., Stephany, S., Caetano, M.: A ferramenta EDDA de geração de campos de densidade de descargas atmosféricas para mineração de dados meteorológicos. In: Anais. . . , vol. 3, pp. 269–275. Congr. Nac. de Mat. Apl. e Comput., SBMAC, São Carlos (2010)
49. Strauss, C., Rosa, M.B., Stephany, S.: Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events. Atmos. Res. **134**, 87–99 (2013)
50. Sun, J.: Convective-scale assimilation of radar data: progress and challenges. Q. J. Royal Meteorol. Soc. **131**(613), 3439–3463 (2005)
51. Tapia, A., Smith, J.A., Dixon, M.: Estimation of convective rainfall from lightning observations. J. Appl. Meteorol. **37**, 1497–1509 (1998)
52. Tukey, J.W.: Exploratory data analysis. Addison-Wesley, Boston (1977)
53. Tuomi, T.J., Larjavaara, M.: Identification and analysis of flash cells in thunderstorms. Q. J. Royal Meteorol. Soc. **131**(607), 1191–1214 (2005)
54. Vendrasco, E.P., Sun, J., Herdies, D.L., de Angelis, C.F.: Constraining a 3DVAR radar data assimilation system with large-scale analysis to improve short-range precipitation forecasts. J. Appl. Meteorol. Climatol. **55**(3), 673–690 (2016)
55. Vila, D.A., Machado, L.A.T., Laurent, H., Velasco, I.: Forecast and tracking the evolution of cloud clusters (ForTraCC) using satellite infrared imagery: methodology and validation. Weather Forecast. **23**(2), 233–245 (2008)
56. Wang, Y., Yang, Y., Wang, C.: Improving forecasting of strong convection by assimilating cloud-to-ground lightning data using the physical initialization method. Atmos. Res. **150**, 31–41 (2014)
57. Wang, C., Zheng, D., Zhang, Y., Liu, L.: Relationship between lightning activity and vertical airflow characteristics in thunderstorms. Atmos. Res. **191**, 12–19 (2017)

Methodological Proposal for the Prediction of Hydrological Responses to Land-Uses and Land-Cover Changes in a Brazilian Watershed

Lidiane dos Santos Lima, Paulo Cesar Colonna Rosman, Julia Celia Mercedes Strauch, Nelson Ferreira Fernandes, and Letícia de Carvalho Giannella

1 Introduction

Over the years, the land-use/land-cover (LULC) is undergoing significant changes due to the fast growth of population density. Thus, continuous urbanization together with the cited changes is a major challenge for the management of water resources [17]. In this context, the hydrology and hydrological cycle of a watershed can be altered considerably due to land-uses and land-cover changes (LULCC), agricultural practices, economics, cultural alternations, among other factors [6]. The hydrological changes can be noticed in the evaporation, evapotranspiration of the ecosystem, soil infiltration capacity, surface and subsurface flux regime, peak flow and water quantity and quality, etc. [10]. Therefore, to plan and manage water resources in the medium and long term, it is fundamental to understand in a broader and more detailed way the effects of LULCC on the flow in watersheds and subwatersheds [12].

Many studies show that LULCC directly affect the availability and flow of water in the hydrological cycle in a watershed [16, 20, 25, 30, 33, 34]. The researches cited are based on assessments of impacts on water resources. Increased urbanization is

L. dos Santos Lima (✉)
Protection and Civil Defense Municipal Office, Maricá, RJ, Brazil
e-mail: lslima@oceanica.ufrj.br

P. C. Colonna Rosman
Federal University of Rio de Janeiro, Ocean Engineering Program, Rio de Janeiro, RJ, Brazil
e-mail: pccrosman@ufrj.br

J. C. Mercedes Strauch · L. de Carvalho Giannella
Brazilian Institute of Geography and Statistics, National School of Statistical Sciences, Rio de Janeiro, RJ, Brazil

N. Ferreira Fernandes
Federal University of Rio de Janeiro, Department of Geography, Rio de Janeiro, RJ, Brazil

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_6

often associated with increases in maximum flows and decreases in minimum flows, and this impact of LULCC varies according to climatic conditions [20].

According to [30], LULCC in the Songkhram Watershed, Thailand, were responsible for the increase of 5.30–6.35% in the flow of the scenario considered. In the study by Kim et al. [20] in the Hoeya Watershed, Korea, flow increased in spring and winter, but decreased in summer and fall LULCC increased maximum flows in the rainy season, but decreased minimum flows in the dry periods. Thus, the authors concluded that LULCC may aggravate the problems of increasing seasonal variability in the flow due to climate change. At the Loess Plateau in China, urban sprawl accounted for 5.46% of all factors influencing LULCC in the region [33]. In a coastal watershed, Eight mile Creek, Alabama, local residential growth will continue into the near future [25]. The mathematical models showed that, due to urbanization, the maximum flows increased from 1966 to 2011 and from 2011 to 2022. In the Olifants Watershed in South Africa, [16] showed that LULCC in the region resulted in a 46.97% increase in surface runoff, with urbanization as the predominant factor for this. For the Tekeze Dam Watershed in Ethiopia, LULCC were also significant. The increase of exposed soil and agricultural areas resulted in an increase in the average annual flow rate of 6.02%.

As shown by all the authors described above, urbanization and other forms of LULCC may result in detrimental effects on the hydrology of a watershed. These effects may increase the number of hydrological incidents and/or disasters, e.g., floods. When precipitation rates are higher than infiltration capacity, the soil behaves like an impermeable surface [2]. As asserted by Noori et al. [25], the impacts of LULCC at maximum outflows may be associated mainly to low intensity but frequent precipitation events; thus, small flood events are more sensitive to urbanization than large events. To study and evaluate how LULCC occur in a region, as well as how these changes affect the flow of surface runoff, it is necessary to conduct research through field observations and/or mathematical modeling. The first method is onerous and usually has local and non-transferable results. Thus, modeling is a more viable and commonly used method to perform environmental diagnostics and prognoses of the region of interest [25]. However, it should be noted that the fieldwork is very valuable, because as such, it is possible to carry out calibration and validation of the mathematical models.

In the studies described above it is possible to observe some limitations, e.g., the prognosis of LULC because of models based on simple regressions. These types of constraints directly impact the outcome of the hydrological response because these limitations can introduce uncertainties in the assessments. To minimize this problem, the present research proposes that the prediction of LULC should be made with a dynamic model based on neural networks. Meteorological scenarios with complete hydrological cycles were chosen to represent excess and scarcity of precipitation in the region. Through all the above, the general objective of this research is to map and predict the integrated impacts of climate and LULCC on the flow of a watershed of interest for scenarios of precipitation excess and scarcity. The specific objectives are: (1) to map the pattern of LULCC in the past and present of a watershed; (2) to predict LULCC in the future of the watershed through the LULC

model; (3) to estimate the impacts on watershed discharges for each pluviometric scenario in the past, present, and future through the hydrological model; and (4) to analyze the estimated hydrological impacts for each subwatershed in the past, present, and future.

2 Methodology

2.1 Application Area

The Araruama Lagoon Watershed is at the coastal lowlands region in the state of Rio de Janeiro and covers the administrative limits of Saquarema, Rio Bonito, Araruama, Iguaba Grande, São Pedro da Aldeia, Cabo Frio, and Arraial do Cabo (Fig. 1). Its meteorological characteristics are peculiar presenting the lowest rainfall volume of the state and registering two types of microclimates, tropical in the west and semi-arid in the east, in a small extent. In the meteorological data recorded in the region, evaporation presents higher values than precipitation resulting in a water deficit, that is, a negative water balance for the most part of the watershed. In the climatology of the region, the average annual rainfall is in the range of 750 and 900 mm and average annual evaporation between 890 and 1370 mm. In the watershed in question is located the Araruama Lagoon, which is a permanent hypersaline lagoon with an average salinity of 52, this characteristic is due to the semi-arid climate previously mentioned [5, 7]. The connection of this lagoon with the sea is carried out through the Itajuru Channel, which during the years has been suffering constant shoaling, which makes it difficult for the internal waters of the lagoon to circulate and renew.

The total drainage area of the Araruama Lagoon Watershed is 430 km^2, with fresh water contributing to 2.3 m^3 of average volume. Its hydrographic system is composed of small subwatersheds and almost all its rivers show intermittent behavior. The most significant rivers that drain into the lagoon are, from west to east: Moças' River, Mataruna River, Salgado River, and Ubá River (Fig. 1) [7].

2.2 Land-Use and Land-Cover (LULC) Modeling

Currently the most commonly used instruments for remote sensing and combined image analysis are obtained through satellites orbiting the Earth. The data obtained by remote sensing tend to be more economical due to the possibility of phenomena monitoring in extensive and remote areas of the planet [15]. In conjunction with the use of remote sensing data, geographic information systems (GIS) are used to relate the environment and human activities. The products that result from the satellite data processed in the GIS can be LULC maps, as well as information

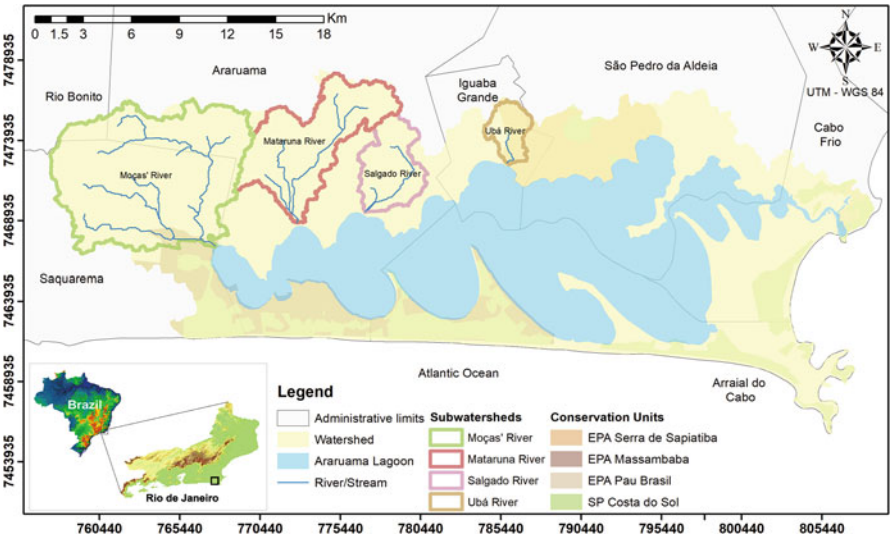


Fig. 1 Location of the Araruama Lagoon Watershed, application area considered in modeling, and its main rivers, from the lagoon system of the microregion of the lakes in state of Rio de Janeiro, Brazil (adapted from [21])

submitted to statistical and analytical procedures. Another relevant point to highlight is the quality and accessibility of data, including the availability of these images at reasonable resolution.

Through the data from the LULC's dynamics diagnosis, it is possible to predict future scenarios. To develop these prognoses, the modeling tool has been widely used for both management and environmental research [31]. The modeling of the LULC's dynamics must begin with a theoretical understanding of human behavior, through its interaction with different forms of land-use, as well as understanding how the land is distributed in the territory. These models began their development in the 1960s and since then their popularity and research have been growing significantly [18]. This growth is because changes in LULC are one of the most important ways man can directly and indirectly affect the environment in which he is inserted.

2.2.1 Land Change Modeler (LCM)

The LCM module of the TerrSet program aims to analyze and predict LULCC, as well as to improve the understanding of the system in its phenomena and effects, including influencing factors such as roads, slope, altitude, anthropogenic interferences, among others [14]. With this amplified understanding there is the possibility of achieving better support in regard to the planning and elaboration of public and private policies. In the context of LULC future projection, the model

is an effective tool to predict these changes in different scenarios, based on the trend pattern of the changes previously occurred. The performance of LCM will be through neural network and/or logistic regression, its accuracy is dependent on the variables that influence LULCC [3, 14, 28].

The input data for the model is composed of two LULC maps, one less recent and one more recent with the same LULC classes. These maps are analyzed presenting a quantitative evaluation of the changes that occurred between the chosen period. The analysis is explained through a comparative graph and maps of the losses and gains by LULC's classes and change of the contribution by class in several units as: hectare (ha), square kilometer (km^2), % of area, among others. The transition potential modeling creates maps with an acceptable degree of accuracy to perform the effective modeling by grouping a set of submodels and exploring the influence of explanatory variables, which may be either dynamic or static in time [14].

The transition structure of the submodels presents a relation of all the transitions, from the smallest to the largest, that occur between time t_1 and t_2. This modeling can be performed by both logistic regression and multi-layer perceptron neural network (MLP). The MLP neural network can model nonlinear relationships and the most robust LULCC models. The LULC's restrictions or incentives factors limit or expand the transformations to a region of interest giving it a degree of adequacy to change [28]. In this way it is possible to forecast trends for LULC, and finally to generate maps of possible future changes.

2.3 Hydrological Modeling

Over the years, many computational models have been developed to simulate the hydrology of watersheds. Thus, hydrological models are considered efficient tools of great importance for researching the complex hydrological processes that affect surface and subsurface hydrology of watershed. With the understanding of these processes, it is possible to evaluate the LULCC's impacts, agricultural activities, water resources management, among others [19].

2.3.1 Soil and Water Assessment Tool (SWAT)

For the trend simulations the SWAT model was chosen. This option was because the model has great acceptance in the academy, is widely used in urban regions and watersheds, and has several publications available. SWAT is a mathematical model created in 1996 and is being continuously developed since then by the Agricultural Research Service and Texas A&M University. The main function of the model is to analyze the impacts to the practice of agriculture in different types of soils and the patterns of the use of it on the surface and underground water flow, sediment production, and water quality in long periods. The SWAT model has been widely used worldwide to predict the nutrient flow discharge and loading

from watersheds of various sizes, for the development of maximum daily total load, simulating hydrology, sediments, nutrients, and pollutant load for basin scale studies dealing with the quantity and quality of water. Simulation scenarios can be useful for assessing the ecological status of the environment, considering climatic factors, soil, and water use. In general, the SWAT application to assess diffuse pollution has presented scientifically interesting results [24].

SWAT requires specific information on water, soil properties, topography, vegetation, and soil management practices in the watershed. Physically the process is associated with water movement, sediment movement, plant growth, nutrient cycling, which are directly modeled by SWAT, using this input data. The SWAT is a continuous model in time and usually works in daily time interval. One of its advantages is the possibility of making simulations for short or long time series. It considers the watershed divided into sub-basins based on relief, soils, LULC and thereby preserves the spatially distributed parameters of the entire basin and the homogeneous characteristics within the basin. The input data for each sub-basin can be grouped or organized according to the following categories: climatic, hydrologic response units (HRUs), ponds/wetlands, groundwater, main watercourse, and sub-basin drainage. To generate the HRUs in each sub-basin, a sensitivity level is adopted, eliminating the LULC's classes with an area smaller than the established value. SWAT controls the creation of HRUs based on the distribution of the different soil types and on the selected soil uses [13].

2.4 Model Calibration and Validation

SWAT calibration can be done automatically and/or manually. Manual calibration is performed by trial and error, while automatic can be performed by the SWAT-CUP[1] software that provides calibration, validation, and analysis of results' uncertainties generated by SWAT simulations [1]. The SWAT simulated data were initially calibrated in the SWAT-CUP, but they were not fully satisfactory and were therefore supplemented by manual calibration. After the complete calibration phase, the model was validated. In the validation a time series other than calibration was used, showing that the model can reproduce the series. After the simulation, calibration, and validation procedure, the results in the three stages were submitted to statistical analyses as described by Moriasi et al. [23]: Nash and Sutcliffe coefficient (NSE), normalized standard error (RSR), correlation coefficient (R^2), and percent bias (PBIAS) in addition to the visual analysis of the graphs generated by the time series.

[1] https://swat.tamu.edu/software/swat-cup/.

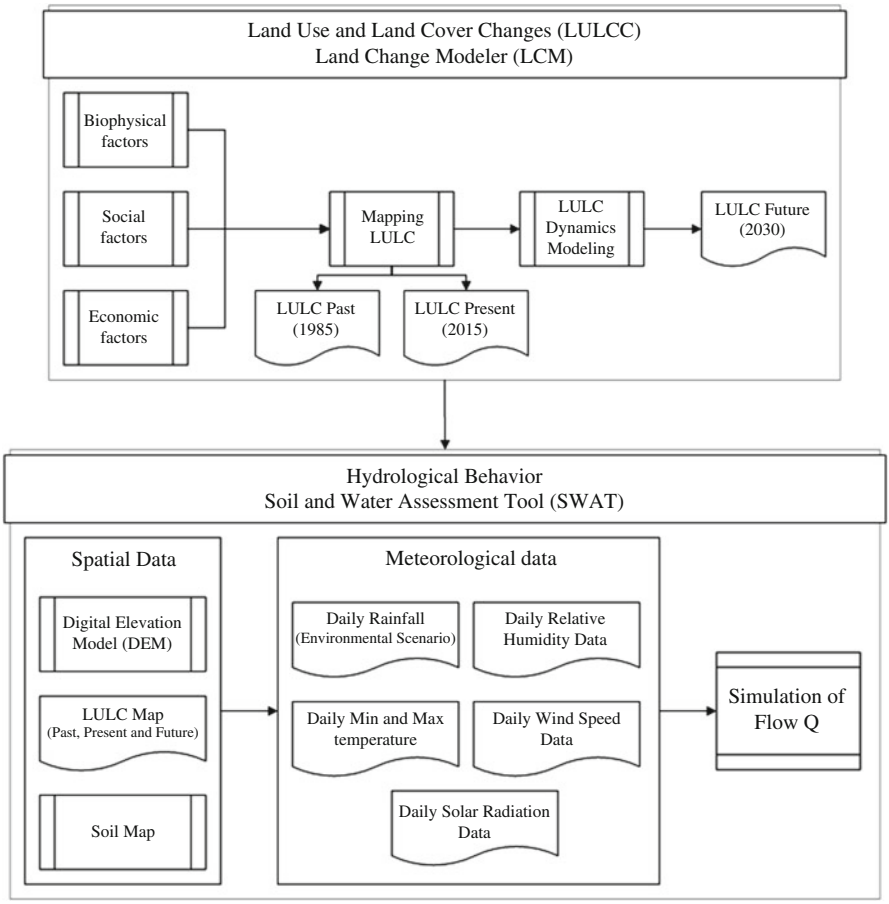


Fig. 2 Methodology framework used in this research

2.5 Research Methodology

The potential hydrological impacts of a watershed in response to LULCC were analyzed for the past (1985), present (2015), and future (2030) pluviometric scarcity and excess scenarios. The methodological framework used in the present research is illustrated in Fig. 2.

2.5.1 Data Preparation

The data used in the setup of SWAT hydrological model are described in Table 1. The input parameters are necessary to represent the watersheds' characteristics

Table 1 Data used in the SWAT model

Data	Source	Spatial/temporal resolution	Number/period	Description
Elevation (DEM)	ASTER GDEM version 2	30 m	–	USGS earth-explorer[a]
Soil map	[9]	1:250.000	2003	GeoPortal EMBRAPA[b]
LULC map	[22]	1:150.000	1985; 2015; 2030	–
Meteorological data	Climate forecast system reanalysis (CFSR)	–	6 stations/ 1990; 2005	Global weather data for SWAT[c]

[a]https://earthexplorer.usgs.gov
[b]http://mapoteca.cnps.embrapa.br
[c]https://globalweather.tamu.edu/

with quality and, consequently, to generate satisfactory results in the hydrological modeling. Secondary data from previous studies and primary data provided by Lagos de São Jão Intermunicipal Consortium (CILSJ) and the Fluminense Federal University (UFF) were used in the present research. The parameters adopted were divided into topographic data, LULC maps, soil map, and meteorological data.

2.5.2 Historical, Current, and Future (LULC)

The LULC maps used for the years 1985, 2015, and 2030 were obtained through the study of [22] carried out in the Araruama Lagoon Watershed as shown in Fig. 3a, b, and c. The distribution and summary of the LULCC statistics are illustrated and described in Fig. 3d and Table 2. Observing the changes from 1985 to 2015 the classes that showed the greatest gains in their areas were: medium density urban area, bare soil and disabled salines, and high-density urban area with an increase of 117.1%, 95.5%, and 44.2%, respectively. On the other hand, the classes that presented the highest losses in this period were: salines, low density urban area, and sand and dunes with decrease of 81.2%, 45.2%, and 20.0% in their respective areas. In Fig. 3 it is possible to verify that water is the class of greater area in the analyzed 45 years. The LULC simulated map for the future (2030) was modeled considering the past patterns of 1985 and 2015 in addition to the legal constraints imposed by the environmental conservation units currently in effect in the Araruama Lagoon Watershed: Environmental Protection Area (EPA) Serra de Sapiatiba, EPA Massambaba, EPA Pau Brasil, and Costa do Sol State Park (SP) (Fig. 3).

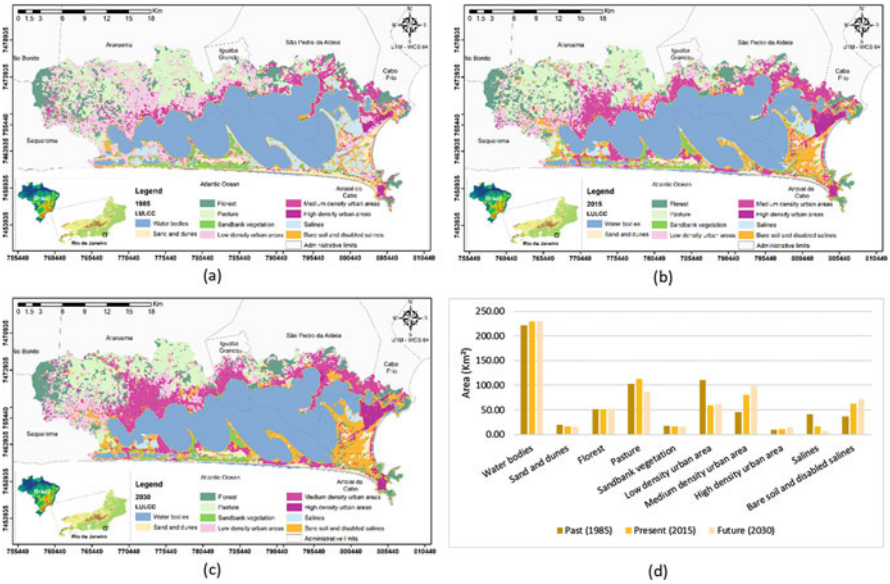


Fig. 3 LULCC for 1985 (**a**), 2015 (**b**), and 2030 (**c**) and distribution of areas in km^2 by classes LULCC (**d**) in Araruama Lagoon Watershed (adapted from [22])

2.6 Setting Scenarios

In this research, six different scenarios were considered (Table 3) to evaluate the impacts on the flow of the main tributaries of each subwatershed to the past, present, and future LULC in years with excess and scarcity of precipitation. The choice of representative years of excess (2005) and scarcity (1990) was determined by the analysis of the meteorological time series (Fig. 4).

3 Results and Discussion

The results concerning the influence, anthropic, and environmental in the subwatersheds flows were simulated by SWAT for the past (1985), present (2015), and future (2030). The subwatersheds considered in the simulation were: Moças' River, Mataruna River, Salgado River, and Ubá River, illustrated in Fig. 1. The physical and hydrological characteristics of each subwatershed, according to the hydrological model, are described in Table 4 for scenarios of scarcity and excess of precipitation.

Table 2 Statistics summary area concerning LULCC Araruama Lagoon Watershed for the periods of 1985, 2015, and 2030 (adapted from [22])

LULCC Araruama Lagoon Watershed	1985		2015		2030		Changes from 1985 to 2030		Changes from 2015 to 2030	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Water bodies	221.68	33.7	229.75	34.9	230.54	35.2	8.86	4.0	0.79	0.3
Sand and dunes	19.45	3.0	16.86	2.6	15.57	2.4	−3.88	−20.0	−1.29	−7.7
Forest	51.58	7.8	51.53	7.8	52.60	8.0	1.02	2.0	1.07	2.1
Pasture	102.58	15.6	113.21	17.2	87.32	13.3	−15.26	−14.9	−25.89	−22.9
Sandbank vegetation	17.85	2.7	16.56	2.5	15.53	2.4	−2.32	−13.0	−1.03	−6.2
Low density urban area	111.00	16.9	58.87	9.0	60.87	9.3	−50.13	−45.2	2.00	3.4
Medium density urban area	45.41	6.9	80.90	12.3	98.59	15.0	53.18	117.1	17.69	21.9
High-density urban area	10.25	1.6	10.80	1.6	14.78	2.3	4.53	44.2	3.98	36.9
Salines	40.98	6.2	16.81	2.6	7.69	1.2	−33.29	−81.2	−9.12	−54.3
Bare soil and disabled salines	36.67	5.6	62.20	9.5	71.68	10.9	35.01	95.5	9.48	15.2

Table 3 Environmental scenarios considered in hydrological modeling

Environmental scenario	Description
1. Precipitation scarcity in past	LULC 1985s and hydrological simulation considering precipitation scarcity 1990s
2. Precipitation scarcity in present	LULC 2015s and hydrological simulation considering precipitation scarcity 1990s
3. Precipitation scarcity in future	LULC 2030s and hydrological simulation considering precipitation scarcity 1990s
4. Precipitation excess in past	LULC 1985s and hydrological simulation considering precipitation excess 2005s
5. Precipitation excess in present	LULC 2015s and hydrological simulation considering precipitation excess 2005s
6. Precipitation excess in future	LULC 2030s and hydrological simulation considering precipitation excess 2005s

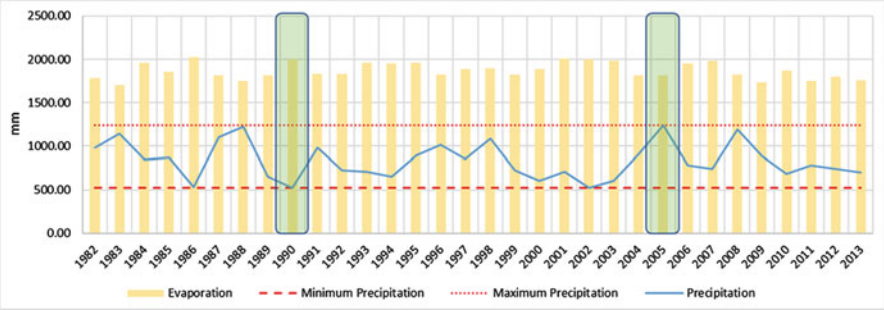


Fig. 4 Annual accumulated values of precipitation, evaporation, minimum and maximum precipitation for the Araruama Lagoon watershed, RJ, from 1982 to 2013. The years chosen to represent the pluviometric scarcity (1990) and the pluviometric excess (2005)

3.1 Model Calibration and Validation

For SWAT model calibration was necessary that the initial calibration was performed in a region adjacent to the area of application of Araruama Lagoon Watershed, which was the Silva Jardim Watershed. The region in question was chosen because it is the region closest to the application area with time series data available in [4]. The model was calibrated for the flows from 1983 to 1997 and validated from 1998 to 2010. After calibration and validation, the statistical analyses showed that the magnitude of the data (NSE) calibrated and validated in relation to the observed data presented a result classified as "good," as well as the normalized standard error (RSR). Percent bias (PBIAS), the simulated data were classified as "very good" (Table 5). Finally, the correlation coefficient, R^2, presented an uncalibrated value of 0.45, calibrated of 0.81, and validated of 0.91. The validated value approached 1, indicating that the data approached the trend line, and consequently with little dispersion.

Table 4 Physical and hydrological characteristics of the subwatersheds considered in the hydrological model

Characteristics/subwatershed	Moças' River			Mataruna River			Salgado River			Una River		
Area (km^2)	89.55			40.41			17.67			7.49		
Perimeter (km)	65.85			54.96			27.78			16.59		
Minimum relief (m)	4			5			4			5		
Maximum relief (m)	575			128			94			73		
Major land-use and land-cover (LULC)	Pasture			Pasture			Pasture			Pasture		
	Hydrological parameters											
	Scenario of precipitation scarcity											
	Past	Present	Future	Past	Present	Future	Past	Present	Future	Past	Present	Future
Minimum annual flow (m^3/s)	0.63	0.61	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum annual flow (m^3/s)	6.38	6.10	5.78	0.74	1.25	1.63	0.30	0.27	0.63	0.16	0.14	0.34
Accumulated annual precipitation (mm)	522.70											
Accumulated annual evaporation (mm)	2001.57											
	Scenario of precipitation excess											
	Past	Present	Future	Past	Present	Future	Past	Present	Future	Past	Present	Future
Minimum annual flow (m^3/s)	0.71	0.70	0.65	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00
Maximum annual flow (m^3/s)	129.20	131.70	129.90	45.94	46.41	46.52	19.43	19.88	20.02	9.72	9.20	9.97
Accumulated annual precipitation (mm)	1241.70											
Accumulated annual Evaporation (mm)	1818.14											

Table 5 Reference and classification according to criteria established by Moriasi et al. [23] of the statistics used to analyze flows data after model calibration and validation

	NSE				RSR				PBIAS (%)			
Category	Reference	(a)	(b)	(c)	Reference	(a)	(b)	(c)	Reference	(a)	(b)	(c)
Very good	$0.75 \leq NSE \leq 1.00$	0.06	0.55	0.65	$0.0 \leq RSR \leq 0.5$	0.97	0.67	0.59	$PBIAS \leq \pm 25$	35	6	−1
Good	$0.65 \leq NSE \leq 0.75$				$0.5 \leq RSR \leq 0.6$				$\pm 25 \leq PBIAS \leq \pm 40$			
Satisfactory	$0.50 \leq NSE \leq 0.65$				$0.6 \leq RSR \leq 0.7$				$\pm 40 \leq PBIAS \leq \pm 70$			
Unsatisfactory	$NSE \leq 0.50$				$RSR \geq 0.7$				$PBIAS \geq \pm 70$			

(a) Uncalibrated; (b) Calibrated; e (c) Validated

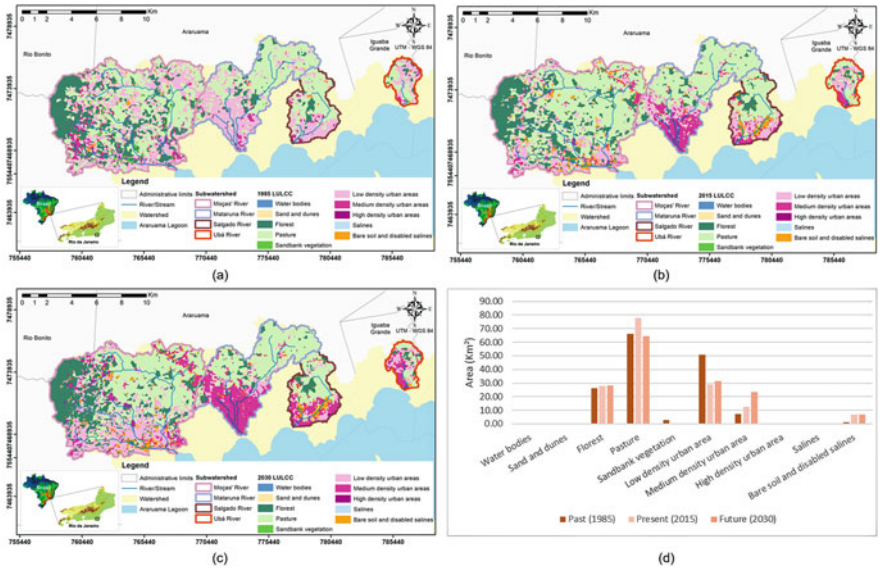


Fig. 5 LULCC for past (1985) (**a**), present (2015) (**b**), and future (2030) (**c**) and distribution of areas in km^2 by classes LULCC (**d**) in the simulated subwatersheds (adapted from [22])

3.2 LULCC

Figure 5a, b and c shows in detail the subwatersheds' LULC maps considered in hydrological modeling for the past, present, and future represented by the years 1985, 2015, and 2030, respectively. In Fig. 5d it is possible to observe the total distribution of subwatersheds areas (km^2) per LULCC's class in the 45 years analyzed. As described by Teixeira [32] and confirmed in the mapping carried out in this research, during the development of the lake's region, in the more interior areas, where the subwatersheds are located, pasture was the dominant constituent of LULCC's class in this region. Another important point to highlight is the increase in the population density represented by the decrease of low density urban areas class and the consequent and gradual increase of medium density urban areas class. This behavior is verified by Nunes [26], who emphasizes that since the 1990s the oil activity has played a significant role in the regional economy. This resulted not only in the injection of direct resources, but also indirect and secondary, increasing the population density in the region.

As shown in Fig. 3, the classifications and projections considered for the past (1985), present (2015), and future (2030) were obtained by Lima [22]. Figure 6 shows in detail each subwatershed considered in the modeling and their respective LULCC over the analyzed years. The Moças' River subwatershed presents the greater part of its area belonging to the administrative limits of Saquarema; Mataruna and Salgado entirely in the administrative limits of Araruama; and Ubá in

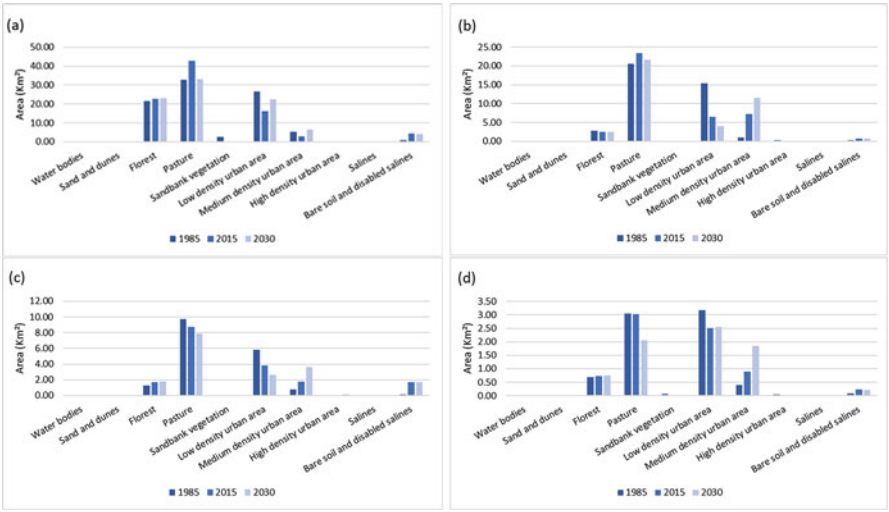


Fig. 6 Distribution of areas in km^2 by classes LULCC (d) in the simulated subwatersheds: Moças' River (**a**), Mataruna River (**b**), Salgado River (**c**), e Ubá River (**d**) for the periods of past (1985), present (2015), and future (2030)

Iguaba Grande (Fig. 1). As observed in Fig. 5d, the patterns of each subwatershed tend to have the following classes as dominants in the analyzed 45 years: pasture, low and medium density urban areas, and forest. This trend was also described by previous bibliographical references [7, 11]. Bidegain and Bizerril [7] in their study identified that the most significant classes are pasture, urban areas (population density), and saline. It should be noted that in the present study the modeling areas do not contain salines that are very significant in Araruama Lagoon watershed scale, as shown in Fig. 3 due to their regional economic importance, mainly before and during the 1990s.

Another LULC class that deserves attention is the forest. From 1900, scientists began to identify degradation of said class. The spaces of forest began to give way to sugarcane, coffee, orange groves, and finally livestock. When comparing maps of the past (1985), present (2015), and future (2030) it is possible to observe that even mathematically the class did not have significant changes, their presence was observed in different places. A plausible justification for this fact is the reforestation with *Eucalyptus* spp that cf. described by de Athayde Bohrer et al. [11] is present in the region along the RJ-106 highway, São Pedro da Aldeia and in smaller non-mappable areas.

Urban areas, both in the diagnosis (past and present) and prognosis period (future), were a class of significant changes in the four modeled subwatersheds. As found by [11], the largest urban concentrations are spatially distributed near the municipal headquarters in the surroundings of Araruama Lagoon, along the highways RJ-106, RJ-102, and RJ-140 and finally in the coastal zones. The rapid

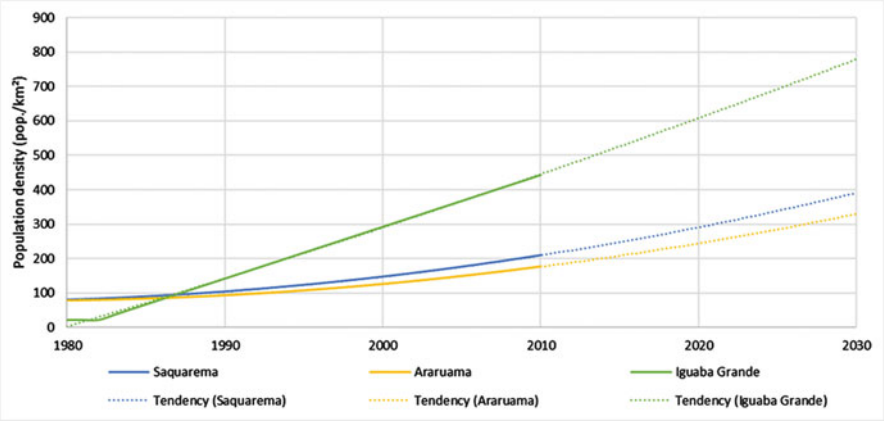


Fig. 7 Diagnosis (1980–2010) and population density (pop./km^2) prognosis (2011–2030) by administrative limits, Saquarema, Araruama, and Iguaba Grande—RJ (Source: CENSO 1980, 1991, 2000, and 2010)

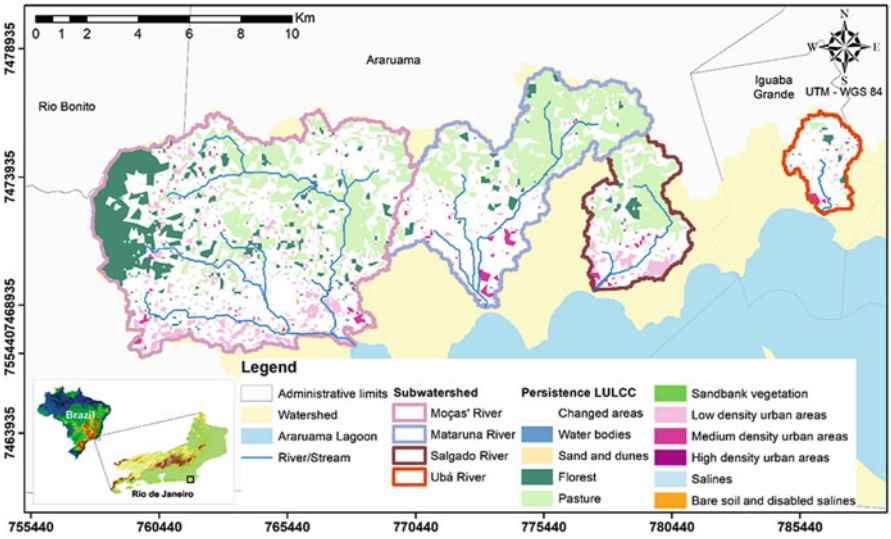


Fig. 8 LULCC's persistence map between the years of 1985 and 2030 to subwatersheds modeled

urban expansion is clearly seen in the evolution of the LULC mapping and in the population data provided by IBGE, as well as in the projections for the municipalities of Saquarema, Araruama, and Iguaba Grande in 2030 (Fig. 7).

During the 45 years analyzed, 59.0% of the total area of the subwatersheds underwent alteration in its LULC according to Fig. 8. The pasture and forest were the classes that persisted most during the years analyzed with 25% and 9% of their

areas, respectively. The changes helped to characterize and diagnose the LULC's transformations from 1985 to 2015 and to propose the future scenario for the year of 2030, considering that the conditions of infrastructural development and spatial restrictions remained the same for the current ones.

3.3 Impacts of Climate and LULCC on Streamflow

Quantifying the effect of LULCC on the hydrological behavior of a watershed is still a problem today. These changes can directly affect the hydrological processes of the region of interest through changes in evapotranspiration, rainfall interception, streamflow generation, water recharge, and soil moisture dynamics [35]. An alternative methodology that has been widely used is the temporal trend analysis based on the relation between the streamflow and precipitation for large watersheds. One of the great advantages of this method is that it can be applied in watersheds that have undergone LULCC without the necessity of a control watershed [8]. Another widely applied methodology for this purpose is the use of hydrological models, thus estimating the effects of changes local on different temporal scales. However, it should be noted that there are significant uncertainties associated with this method [27].

As described by Sanchez et al. [29] hydrological disasters classified by the Brazilian Classification and Codification of Disasters (Cobrade) in Rio de Janeiro state presented 497 records between the years of 1991 and 2012. Of the state registries cited, 13 were attested, in the Araruama Lagoon watershed, most frequently in the months of November to February of the analyzed period. The excessive population increase in the region of interest promoted a consequent increase in the waterproofing of the earth, which caused a greater volume of surface drainage; thus, municipalities with greater population density are more susceptible to flooding. It should be noted that due to the inexpressiveness of the municipality of Rio Bonito for the drainage of the hydrographic basin of interest, the data referring to it, with respect to hydrological disasters, were disregarded.

In the present research the hydrological responses to different conditions and changes of LULC in the past, present, and future were studied for maximum streamflows in the four hydrographic subwatersheds modeled in scenarios of pluviometric excess and scarcity. The results of the hydrological modeling, through the exposure of precipitation data, mean flows distributed in the months of the year in the scenarios of pluviometric scarcity and excess for each subwatershed are shown in Fig. 9. As described in item 3.2, the subwatersheds modeled presented significant changes in LULC during the analyzed 45 years and these alterations had effects on the maximum surface runoff flows of each subwatershed.

The impacts of the LULCC in the Moças River subwatershed showed that for the scenario of past precipitation scarcity to the present, it presented a mean decrease in its flow of 4% and for the present to the future of 5%. For the pluviometric excess scenario from 1985 to 2015, the reduction was cushioned by 2% and

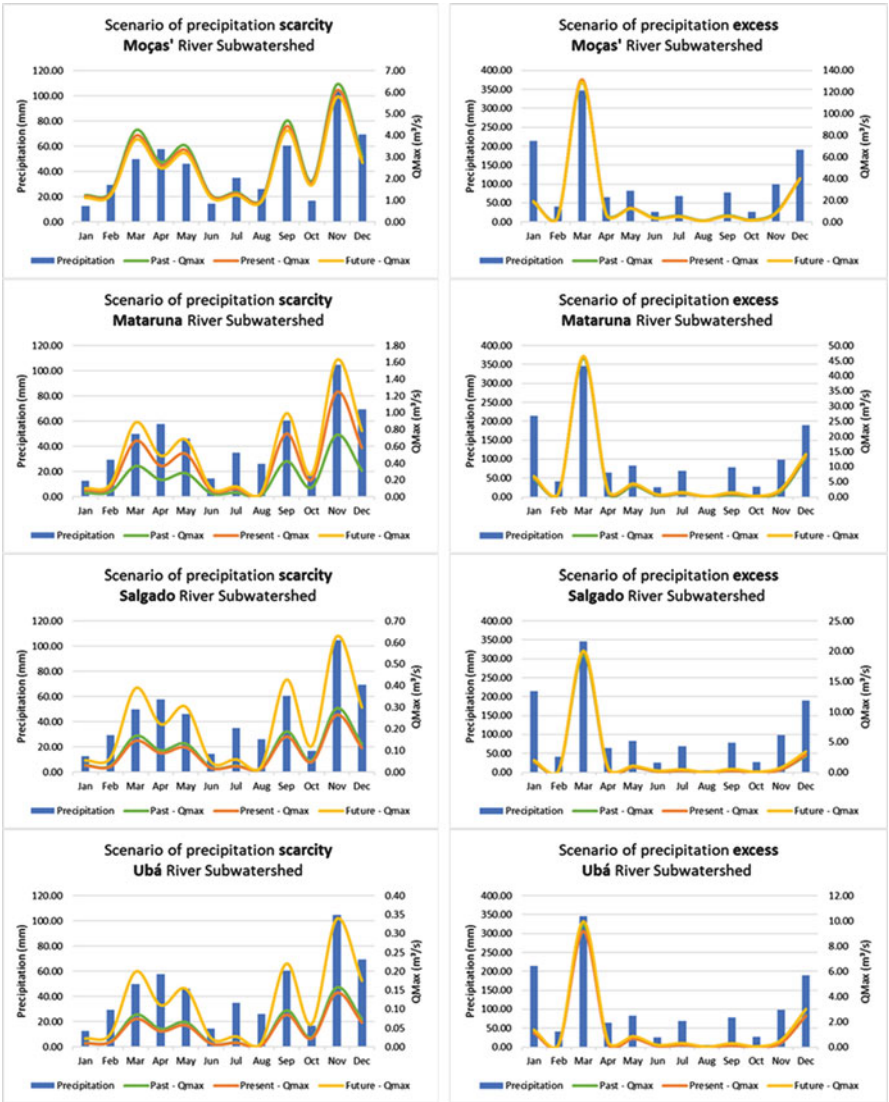


Fig. 9 Simulated streamflow to the scenario of scarcity and excess precipitation to LULCC of the past, present, and future for the Moças, Mataruna, Salgado e Ubá River subwatersheds

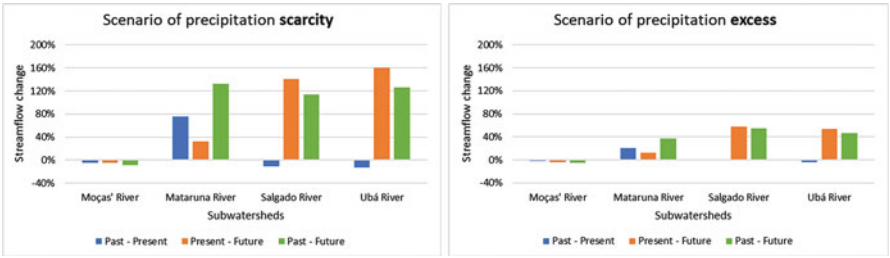


Fig. 10 Percentage variation of the past to present and future streamflow in the scenarios of precipitation scarcity and excess

from 2015 to 2030 by 4%. For Mataruna River subwatershed, the flow changes were more significant both during the period of scarcity and during the period of excessive precipitation. In the first scenario from the past to the present there was an increase of 75% and from the present to the future 32%. In the second scenario the percentages were 21% and 12%, respectively. The Salgado River subwatershed for the first 30 years analyzed presented an average annual reduction in its flow of 11%, in the prognosis the flow had an increase of 141% for the year of pluviometric scarcity. The Ubá River subwatershed presented the same hydrological behavior as the Salgado with reduction from the past to the present and increase of flow from the present to the future with percentage values of 13% and 160%, respectively (Fig. 10).

Figure 10 shows that, in general, the hydrological behavior of the subwatersheds is similar, emphasizing that for the pluviometric scarcity scenario, the percentage changes in flow are more significant. The reduction in the flows of subwatersheds, in some situations, mainly in Moças River subwatershed, was similarly found by Shrestha et al. [30] in Thailand. One possible way to reduce the flow rate is to increase evapotranspiration. On the other hand, the significant increase in flow rates, mainly from the present to the future, can be explained by the population density indicated by both the CENSO projections and the LULC prognosis. The conversion of non-urbanized areas into urban areas tends to increase the maximum flow which can cause flooding in areas near the margin of the rivers analyzed. As found in previous studies, e.g. [35], LULCC can exert influences on the flow of a system, resulting in an increase or decrease in its value depending on the nature of the change.

As found by Noori et al. [25] even if there is the same degree of urbanization in different areas of the watershed, its contribution to the streamflow depends on its location within the watershed. It should be noted that, in addition, other characteristics that need to be considered are topography and soil type which play an important role in the floods of a region. In this way, mapping and understanding the effects of urbanization on hydrological processes is essential for urban planning.

4 Conclusions

Sustainable planning and management of water resources can be more consistently accomplished when the capacity to predict LULC impacts is developed by reflecting the pressures of population densities and their local changes. This establishes a scientific challenge to understand the processes involved in these changes, as well as to formulate relationships and conceptualizations at different scales, making the predicted model of hydrological responses of a watershed to LULCC and climate more and more reliable. Thus, the present research fulfilled its objective of mapping and predicting the integrated impacts of climate and LULCC in the streamflow of a watershed of interest for scenarios of precipitation excess and scarcity.

The subwatersheds analyzed showed significant changes in their LULC over the simulated 45 years. As shown, 59.0% of the total area of the model domain suffered some type of change, which directly impacted the simulated results of the maximum streamflows of all subwatersheds. An LULCC that deserves to be highlighted is the gradual increase in medium population density area for the modeled regions, justifying some of the changes found in the modeled flows. It should be emphasized that the entire watershed also presented the same trend of population density increase. As the region has already registered hydrological disasters, knowledge of this hydrological tendency is important for urban planning and water resource management in order to mitigate such disasters.

It is important to point out, however, that during the analysis of the simulated data, some limitations of the research were identified, even though they did not compromise the research, these limitations show that there are gaps in the existence and availability of an environmental and socioeconomic basis, which would make more feasible and accurate the calibration and validation of hydrological models and LULCC models that will more actively assist in decision making.

From the above, some recommendations can be suggested, such as:

- Continuous monitoring of hydrometeorological, geological, and geotechnical conditions that allow to anticipate the occurrences of natural disasters, especially, hydrological. In this way, managers and decision makers will be able to mitigate the human and material damages resulting from such disasters [29].
- Another point to be addressed is the relationship between the hydrological disaster and the sustainability of multi-scalar urban drainage. The understanding of the complexity of the relationships between natural ecosystems, urban, and society is fundamental so that the control of floods in urban areas is reconceptualized technically and managerially. Thus, sustainable urban drainage can contribute to the control of runoff through structural and non-structural methodologies that reduce the exposure of the population vulnerable to local hydrological disasters [36].

The results of the present research can be used by several sectors and agents, e.g., LULC planners, decision makers, public administrators, among others with the purpose of promoting actions that allow the natural hydrological disaster mitigation

and/or control, considering environmental and social aspects. The identification and mapping of the main drivers of LULC and climatological changes contributes methodologically to quantify the hydrological effects in a watershed as response to these changes, considering the seasonality and the recurrence of these events. Thus, the present study is a contribution to the production of the knowledge needed to manage natural hydrological disasters in a watershed, providing subsidies so that communities can become resilient and sustainable.

Acknowledgements This research was supported in part by: Municipal Service of Emergency Response and Civil Protection of Maricá (SEPDEC), Federal University of Rio de Janeiro (UFRJ), International Virtual Institute of Global Changes (IVIG), and National School of Statistical Sciences (ENCE).

References

1. Abbaspour, K.C.: SWAT-CUP: SWAT calibration and uncertainty programs—a user manual (2015)
2. Abdulkareem, J.H., Sulaiman, W.N.A., Pradhan, B., Jamil, N.R. : Relationship between design floods and land use land cover (LULC) changes in a tropical complex catchment. Arab. J. Geosci. **11**(376), 1–17 (2018)
3. Ahmed, B., Ahmed, R.: Modeling urban land cover growth dynamics using multi-temporal satellite images: a case study of Dhaka, Bangladesh. ISPRS Int. J. Geo-Inf. **1**(1), 3 (2012)
4. ANA: Sistema de informações hidrológicas, Novembro de 2005 (2015)
5. Barbiéri, E.B.: Ritmo climático e extração do sal em cabo frio. Rev. Bras. Geogr. **7**(4), 23–109 (1975)
6. Berka, C., Schreier, H., Hall, K.: Linking water quality with agricultural intensification in a rural watershed. Water Air Soil Pollut. **127**(1), 389–401 (2001)
7. Bidegain, P., Bizerril, C.: Lagoa de Araruama: perfil ambiental do maior ecossistema lagunar hipersalino do mundo, vol. 12. Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável, Rio de Janeiro (2002)
8. Bosch, J.M., Hewlett, J.D.: A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. J. Hydrol. **55**(1), 3–23 (1982)
9. Carvalho Filho, A.D., Lumbreras, J.F. Wittern, K.P., Lemos, A.L., de Santos, R.D., Calderano Filho, B., Calderano, S.B., Oliveira, R.P., Aglio, M.L.D., Souza, J.S.D., Chaffin, C.E.: Mapa de reconhecimento de baixa intensidade dos solos do estado do Rio de Janeiro (2003)
10. Cuo, L., Zhang, Y., Gao, Y., Hao, Z., Cairang, L.: The impacts of climate change and land cover/use transition on the hydrology in the Upper Yellow River Basin, China. J. Hydrol. **502**, 37–52 (2013)
11. de Athayde Bohrer, C.B., Dantas, H.G.R., Cronemberger, F.M., Vicens, R.S., de Andrade, S.F.: Mapeamento da vegetação e do uso do solo no centro de diversidade vegetal de cabo frio, Rio de Janeiro, Brazil. Rodriguésia **60**(1), 1–23 (2009)
12. Dixon, B., Earls, J.: Effects of urbanization on streamflow using SWAT with real and simulated meteorological data. Appl. Geogr. **35**(1), 174–190 (2012)
13. Durão, A., Morais, M.M., Brito, D., Leitão, P.C., Fernandes, R.M., Neves, R.: Estimation of pollutant loads in Ardila watershed using the SWAT model. J. Environ. Sci. Eng. **1**(10B), 1179–1191 (2012)
14. Eastman, J.R.: TerrSet tutorial (2015)
15. Guidolini, J.F., de Almeida, A.M., Júnior, R.F.D.V., Schuler, A.E., Araújo, M.V.N.: Modelagem da dinâmica do uso e ocupação do solo (1975 a 2010) na bacia do rio uberaba, município de veríssimo—mg (2013)

16. Gyamfi, C., Ndambuki, J., Salim, R.: Hydrological responses to land use/cover changes in the Olifants Basin, South Africa. Water **8**(12), 588 (2016)
17. Hunter, P.R.: Climate change and waterborne and vector-borne disease. J. Appl. Microbiol. **94**(Suppl:37s–46s) (2003)
18. Jamal, J.A.: Dynamic Land Use/Cover Change Modelling, 1st edn. Springer, Berlin (2012)
19. Jaswinder, S., Knapp, H.V., Arnold, J.G., Demissie, M.: Hydrological modeling of the Iroquois River watershed using HSPF and SWAT1. J. Am. Water Resour. Assoc. **41**(2), 343–360 (2005)
20. Kim, J., Choi, J., Choi, C., Park, S.: Impacts of changes in climate and land use/land cover under IPCC RCP scenarios on streamflow in the Hoeya River Basin, Korea. Sci. Total Environ. **452–453**, 181–195 (2013)
21. Lima, L.D.S.: Acoplamento de modelos hidrológicos e hidrodinâmicos para prognóstico ambiental detalhado de longo prazo. Tese (2016)
22. Lima, L.D.S.: Alterações do uso e ocupação da terra associadas a transformações territoriais: passado, presente e futuro da bacia hidrográfica da Lagoa de Araruama—RJ. Monografia (2017)
23. Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Am. Soc. Agric. Biol. Engineers **50**(3), 885–900 (2007)
24. Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R.: Soil and water assessment tool: theoretical documentation (2011)
25. Noori, N., Kalin, L., Sen, S., Srivastava, P., Lebleu, C.: Identifying areas sensitive to land use/land cover change for downstream flooding in a coastal Alabama watershed. Reg. Environ. Chang. **16**(6), 1833–1845 (2016)
26. Nunes, N.D.S.: A influência do recebimento de royalties do petóleo nas fragmentaçes territoriais nas baixadas litorâneas, RJ (2014)
27. Post, D.A., Jakeman, A.J., Littlewood, I.G., Whitehead, P.G. and Jayasuriya, M.D.A.: Modelling land-cover-induced variations in hydrologic response: Picaninny Creek, Victoria. Ecol. Model. **86**(2), 177–182 (1996)
28. Rai, P.K., Mishra, V.N., Mohan, K.: Prediction of land use changes based on land change modeler (LCM) using remote sensing: a case study of Muzaffarpur (Bihar), India, vol. 60 (2014)
29. Sanchez, G.M., Machado, M.A., Martins, M.M., Oliveira, O.M.D., Higashi, R.A.D.R., Sbroglia, R.M., Dutra, R.D.C., Goerl, R.F., Bim, R.: Atlas brasileiro de desastres naturais: 1991 a 2012 (2013)
30. Shrestha, S., Bhatta, B., Shrestha, M., Shrestha, P.K.: Integrated assessment of the climate and landuse change impact on hydrology and water quality in the Songkhram River Basin, Thailand. Sci. Total Environ. **643**, 1610–1622 (2018)
31. Snowling, S.D., Kramer, J.R.: Evaluating modelling uncertainty for model selection. Ecolo. Model. **138**(1), 17–30 (2001)
32. Teixeira, V.M.D.L.: Dispersão e extensão urbana no município de araruama: uma análise da implantação de loteamentos residenciais voltados à atividade turística. Cadernos do Desenvolvimento Fluminense **23**, 36 (2015)
33. Wang, T., Yang, M.H.: Land use and land cover change in China's Loess Plateau: the impacts of climate change, urban expansion and grain for green project implementation. Appl. Ecol. Environ. Res. **16**(4), 4145–4163 2018.
34. Welde, K., Gebremariam, B.: Effect of land use land cover dynamics on hydrological response of watershed: case study of Tekeze dam watershed, Northern Ethiopia. Int. Soil Water Conserv. Res. **5**(1), 1–16 (2017)
35. Zhang, L., Cheng, L., Chiew, F., Fu, B.: Understanding the impacts of climate and landuse change on water yield. Curr. Opin. Environ. Sustain. **33**, 167–174 (2018)
36. Zhou, Q.: A review of sustainable urban drainage systems considering the climate change and urbanization impacts. Water **6**(4), 976 (2014)

Computational Modeling and Simulation of Surface Waterflood in Mountainous Urban Watersheds with the MOHID Platform: Case Study Nova Friburgo, Brazil

Wagner R. Telles, Diego N. Brandão, Jader Lugon Jr., Pedro P. G. W. Rodrigues, and Antônio J. Silva Neto

1 Introduction

Historically, civilizations have always sought to establish themselves in regions of floodplain because they allow for the best development of agriculture with its fertile soil, in addition to the abundant presence of water, indispensable for the maintenance of life. A classical example is the Egyptian civilization that prospered along the Nile River, where they found a fertile ground for growing crops. Other benefits of such location consist in allowing navigation, transportation and, later in human history, the industrial process [1].

These areas suffer a natural flooding process, though, with disorderly urban growth, this process may lead to disasters of vast proportions. Cities have some aggravating factors in this process, such as the increasing presence of buildings, and the large percentage of impervious areas, which prevent the flow and infiltration of water to the deeper soil layers, this meaning the surface runoff surpasses the drainage capacity.

W. R. Telles
Universidade Federal Fluminense, Santo Antônio de Pádua, Rio de Janeiro, Brazil
e-mail: wtelles@id.uff.br

D. N. Brandão (✉)
Centro Federal de Educação Tecnologica Celso Suckow da Fonseca, Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: diego.brandao@cefet-rj.br; diego.brandao@eic.cefet-rj.br

J. Lugon Jr.
Instituto Federal Fluminense, Macaé, Brazil
e-mail: jader@iff.edu.br

P. P. G. W. Rodrigues · A. J. Silva Neto
Instituto Politécnico, Universidade do Estado do Rio de Janeiro UERJ, Nova Friburgo, Brazil
e-mail: pwatts@iprj.uerj.br; ajsneto@iprj.uerj.br

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_7

According to Wang et al. [2], floodings have become one of the most significant natural hazards in urban areas. Data from Emergency Events Database (EM-DAT) created by the World Health Organization (WHO) show that only in Latin America floods made about 5036 fatal victims in the period from 2008 to June 2018 [3].

In mountainous cities, these events can be even more dangerous, due to the high water velocity, besides the vast amount of slopes sediments transported. An example was the disaster that occurred in the mountainous region of the State of Rio de Janeiro in Brazil in 2011, with about 900 fatal cases [3]. This number represents almost 18% of all victims in Latin America, and such event can be considered the most significant natural disaster caused by the urban flood in the region in the aforementioned period.

Figure 1 presents some Latin America disasters that took place in recent years. Another consideration is about the economic losses resulting from such disasters, exceeding billions of dollars [4]. Recent research shows that training local population to use systems capable of predicting this phenomenon reduces disaster risk [5].

In this context, mathematical models can help in the development of computational approaches that are capable of simulating the hydrological disasters caused by waterflood.

Due to the limit in the number of pages of the present work, we do not intend to carry out a systematic review on the research area, but we want merely to

Fig. 1 Examples of Latin America floods

demonstrate its research potential. For more details, there is a vast literature about flood [2, 6–15].

This chapter aims to present an introduction on the mathematical modeling involved in the problem of the urban flood, for it presents a classical model with differential equations, that can describe the behavior of water flow in a water body, as well as its computational representation. The use of the MOHID (MOdelagem HIDrodinâmica-Hydrological Modeling) platform allows simulating of a real case study obtaining high qualitative results.

2 Math Modeling and Proposed Problem Solution

Mathematical models have the capacity to simulate events that cause the elevation of water levels in the drainage channels of a river basin [16], which has as consequence, the production of flood waves that affect and provide risks to the population.

More specifically, studies involving flow in rivers and channels can be modeled by equations ranging from this one-dimensional formations with the Saint Venant equations to three-dimensional formulations with the Navier–Stokes equations, where the latter are approximate for incompressible fluids and with free surface [17]. In this chapter, emphasis is given to the modeling of flood waves using the equations of Saint Venant.

According to Liggett[18], considering the high number of variables that are involved in the characterization of flows in rivers, as well as the complex geometry of the channels, the application of the equations of Saint Venant is conditioned to the adoption of some simplifications based on the conservation principles, which are [19]:

- One-dimensional and incompressible flow;
- Speed is uniform in each cross section, varies only in the longitudinal direction;
- Vertical accelerations are not considered;
- Average depth of the bottom is small enough;
- Friction losses not present significant differences in relation to the permanent flows, allowing the use of the Manning equation;
- Erosion and sediment deposition in the bottom are neglected;
- Longitudinal axis of the canal can be represented by linear sections with low slope.

Based on the considerations described above, the equations of Saint Venant are formally described, based on the equation of continuity and motion, as [20]:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\beta . Q^2}{A}\right) + g \cdot A \frac{\partial h}{\partial x} + \frac{g \cdot n^2}{R^{4/3}} \frac{|Q| Q}{A} = 0 \tag{2}$$

where x is a spatial variable (m); t the time variable (s); Q is a discharge channel (m^3/s); A is a cross-sectional area (m^2); q is a side contribution (m^3/s); β is the *Boussinesq* number $(-)$; g is an acceleration of gravity (m/s^2); h is an elevation of the water surface above the assumed datum (m); n is the coefficient that relates all the elements that oppose channel resistance to flow, also known as Manning's roughness coefficient $(s/m^{1/3})$.

Considering as forces acting on a small control volume, in Eq. (2), called the dynamic equation, the first term is related to the local acceleration of the water without control volume due to the variation of velocity over time, while the second concerns the advection of momentum into a control volume due to the change of velocity along the channel; the third term represents a force associated with the pressure variation between the control and volume control levels due to the channel depth variation and, finally, the last term represents frictional force with the bottom and the river banks, which tends to slow the flow [21].

Other classical equations found in the literature can also be deduced from the Saint Venant equations, taking into account the preponderant terms related to the problem to be treated, namely: inertia, pressure, gravity, and friction. Taking only the gravitational and frictional forces, and taking into account that $h = Z + H$, where Z is the channel bottom elevation and H is the channel water column depth, Eq. (2), it can be written as:

$$Q = \frac{1}{n} R^{1/2} \cdot s^{1/2} \cdot A \tag{3}$$

where s is the longitudinal bed slope, defined as:

$$s = \left(-\frac{\partial Z}{\partial x} \right) \tag{4}$$

Equation (3), together with Eq. (1), is called the kinematic wave model. On the other hand, if in addition to friction and gravitational forces, we also take the hydrostatic force, then along with Eq. (1) we have the diffusive wave model, given by:

$$\frac{\partial H}{\partial x} + S - s = 0 \tag{5}$$

where S is the slope friction, defined as:

$$S = \frac{n^2}{R^{4/3}} \frac{|Q|Q}{A^2} \tag{6}$$

It should be noted that one of the most relevant parameters in the Saint Venant equations is the Manning coefficient, which, according to [22], is related to the physical characteristics of the channel, such as surface roughness and irregularities, background vegetation, sinuosity, erosion, and sedimentation leading to changes in

cross-sectional features, obstructions such as bridge or garbage piles, suspended material, and bottom loading.

According to Porto [23], the equations of Saint Venant described by Eqs. (1) and (2) applied to nonpermanent flows require analytical or numerical techniques developed for their solution, in addition to a significant amount of hydraulic data channel, especially when applied to natural waterways.

In this chapter, the Saint Venant equations are solved numerically using the MOHID platform, which has the numerical algorithms based on a finite volumes approach, constituting a flow oriented strategy that facilitates the coupling of different processes and allows the conservation of mass and movement [24].

For more details on the numerical solution methodology of the Saint Venant equations, the interested reader should consult [24].

3 MOHID (Modelagem Hidrodinâmica: Hydrological Modeling) Platform

MOHID water modeling system is a platform that simulates hydraulic and hydrological processes occurring in water bodies and soils, including river basins, estuaries, seas, and oceans. Its development began in 1985, by a team of collaborating technicians from the Marine and Environmental Technology Research Center (MARETEC), belonging to Instituto Superior Técnico (IST) and to the Engineering School of the Technical University of Lisbon, with the company Hidromod Ltda [25].

The basis of the programming used for the construction of the initial versions of the MOHID platform is FORTRAN 77 programming language, being developed based on a two-dimensional flow forced by the tide used in the study of estuaries and coastal zones, being solved through the classic method of finite differences. Over the years, this platform has been updated and improved due to its application in many researches and engineering projects [26].

In its current stage, this platform is based on the ANSI FORTRAN 95 programming language, which allows for an independence from the operating system (Windows, Linux, Unix, etc.) with which you want to run the model as well as an easy implementation of the code in any environment [27]. In addition, this programming is object oriented, allowing the modeling of different physical and biogeochemical processes that occur in different systems [28]. The latest update of this MOHID platform, called MOHID Studio 2016, version 3.0.0.1860, was released in June 2016 and runs on 64 bits in the Windows Vista, 7, 8, 10, or server version and is available for download for free on the company Action Modulers Consulting & Technology: http://www.actionmodulers.com/products/mstudio/products-mohidstudio2015.shtml.

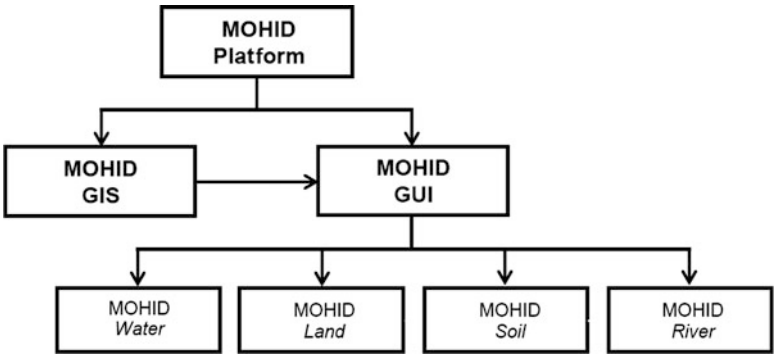


Fig. 2 Block diagram of MOHID platform

Basically, the MOHID platform is divided into two large blocks: MOHID GIS environment and MOHID GUI environment, which communicate with each other. Figure 2 shows the block diagram of MOHID platform.

The GIS environment is a geographic information system that allows the user to manipulate and manage the data of the time and space variables required or produced by the numerical programs of the MOHID platform [29]. It is in this environment that the computational meshes are created in which the simulations are applied [30], based on a file of topographic data of the region of interest, in the XYZ Points format.

With these files containing the computational meshes, the construction of the Digital Hydrologically Correct Terrain Model is started, where the first step is the construction of a mesh with regular or variable spacing, which is used to make an interpolation process with the topography data. In this process, for each cell in the space grid is assigned an altitude value according to the information contained in the XYZ Points file. This process of interpolation gives rise to a digital terrain file, which must be submitted to a process of removal of the possible depressions that may contain. This process allows the digital terrain to resemble the real physical environment, thus avoiding the emergence/disappearance of channels in the drainage network or, also, inappropriate storage places [31]. This process is performed in an iterative manner, where each cell of the file containing the digital model is analyzed, taking into consideration neighboring cells, in order to avoid that depressions that do not exist in the physical medium can appear in the digital file [32].

With the digital model of the terrain prepared, it is necessary a process of construction of the files referring to the slope of the cells, the direction of the flow, drainage area, drainage network, and delimitation of the hydrographic basins of interest. The delineation of the basin can be performed by determining the capacity to conduct the drainage channel conduction network [33]. Subsequently, the profile of the transverse networks of drainage networks is defined, when it is possible to construct the tracing of two distinct routes, relating it to the number of tributaries or

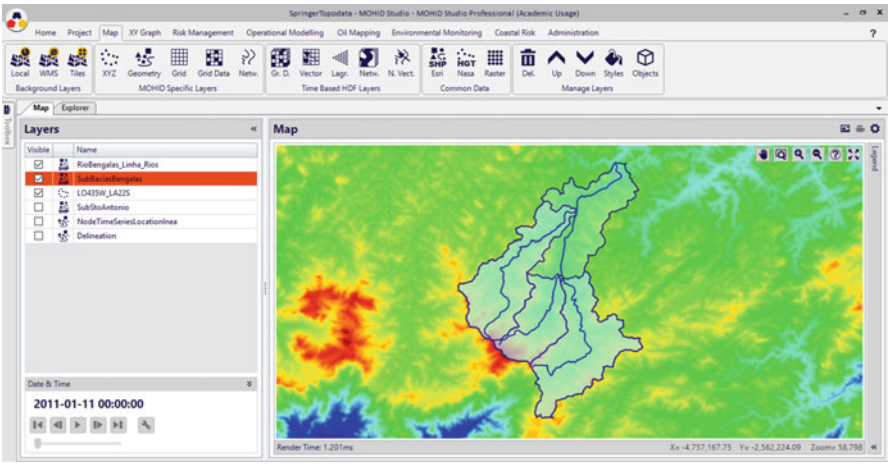


Fig. 3 MOHID GIS environment

to a drained area. The delimitation of the basin can be accomplished by determining an exit point (cell) in the drainage network—considered to be the exudation of the basin—or by the minimum area from which the process of formation of the drainage channels begins [34]. Subsequently, the user can construct the profile of the transversal sections of the drainage networks in two different ways, that is, relating them to the number of tributaries or the drained area, containing different geometries. Figure 3 shows the MOHID GIS environment.

Subsequently, after the entire process of preparing the files made in the MOHID GIS environment, it is possible to use the MOHID GUI environment to perform simulations linked to the events of interest. In this environment, the following numerical tools are present: MOHID Water (modeling of hydrodynamic processes, simulation of dispersion phenomena, wave propagation, sediment transport, water quality/biogeochemical processes in the water column and exchanges with the bottom), MOHID Land (hydrographic basin model), MOHID River Network (hydrographic network simulation) and MOHID Soil (water flow through porous media), which are available in the MOHID GUI [29, 30, 35].

For the simulation of the events of flood waves, the numerical tool MOHID Land is used, being the surface runoff and drainage network modeled by the Saint Venant equations, which were described in Sect. 2. This tool also necessarily requires a file of the digital terrain model, its drainage network, and precipitation data of the region of interest. In addition, for a simulation to be performed, the user must construct a working environment (Solution), a domain (Domain) and choose the phenomena that he wants to include in the simulation. In the middle range of the screen, there is a list with several files, in which the user will configure the parameters of interest and intervals of time. Figure 4 shows the MOHID GUI.

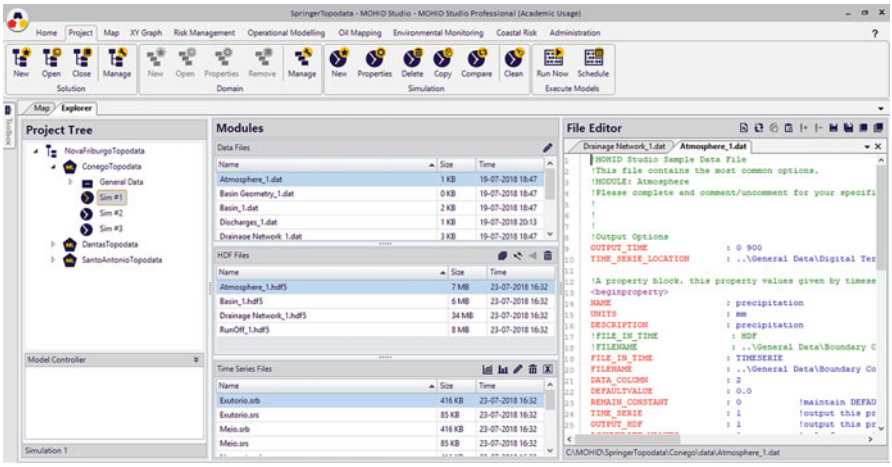


Fig. 4 MOHID GUI environment

4 Case Study: Nova Friburgo, Rio de Janeiro, Brazil

The city of Nova Friburgo is located in the Mountainous Region of the State of Rio de Janeiro, Brazil, and covers an area of approximately 933 km^2, under the geographical coordinates of the south parallels 22°11′and 22°24′ and the meridians of 42°37′ and 42°27′ (as described in [36]).

Its territorial area contains three main hydrographic basins: the Rio Grande watershed, the Bengalas river basin, and the Macaé river basin [37]. Particular emphasis is given to the catchment area of the Bengalas river, which covers the urban area of the city and is divided into four sub-basins: the D'Antas stream basin, the Cônego river basin, the Santo Antônio river basin, and the Bengalas river basin [37].

Among these sub-basins, the basins of the river Cônego and the river Santo Antonio stand out as the most important, since the confluence of these rivers occurs near the center of the city, giving rise to the Bengalas river. Therefore, an investigative analysis of these two basins becomes substantial importance with regard to the monitoring of eventual floods [38]. Figure 5 shows the geographical location of the city of Nova Friburgo, as well as the watersheds of the river Cônego and the river Santo Antônio.

According to the Nova Friburgo Rainforest Plan (PAPNF), finalized in 2007 and made available by the Municipality, the Cônego river is formed by the Caledônia river, it has its source located in the Serra da Boa Vista in Pico do Caledônia, near to the border with the municipality of Cachoeiras de Macacu, containing an extension of 10.6 km to the junction with the Santo Antônio river and drained area of approximately 29.1 km^2. In addition, there is intense urbanization of the banks

Fig. 5 Nova Friburgo city location containing the Bengalas, Cônego, and Santo Antônio rivers hydrographic basins. Source: IBGE

and the reduction of the width of its bed near the confluence with the Santo Antônio river. The canal is still uncoated, occupying the banks in some stretches [37].

On the other hand, the river Santo Antônio is born in the district of Mury, also near the border with the municipality of Cachoeiras de Macacu and follows to the confluence with the river Cônego, already in the center of the city of Nova Friburgo, totaling an extension of 16.2 km and its basin presents an area of 57.2 km^2. Part of

its banks is uncoated—but with good maintenance conditions—and part covered in stone mortar and concrete in the stretches where there is occupation [37].

5 Application and Results

This section presents the processes for the acquisition of experimental precipitation data and water levels in the watersheds of the Cônego and Santo Antônio rivers, for the month of January 2011, as well as the process of construction of the digital terrain model of the region of interest and computational simulations related to the drainage networks of each basin and consequent flood waves.

5.1 Experimental Data Acquisition

In order to analyze the applicability of the mathematical modeling, as well as the feasibility of using the MOHID platform in the simulation of flood events, it was considered the precipitation event that occurred in the municipality of Nova Friburgo in January 2011, and its respective flood waves caused by the elevations of water levels in the drainage networks of the Cônego and Santo Antônio river basins. This month was marked by a major natural disaster, leading to death a substantial number of inhabitants, approximately one thousand, as well as serious economic and social consequences.

Telles et al. [31, 38] also carried out the simulation involving the event of January 2011. However, in the referred work, the authors considered only the period from January 11 to 13, the moment of the tragedy's peak, unlike the simulations carried out in this chapter, in which considered every day of the month in question.

The precipitation measurements (mm) and water levels (m) of the rivers during the month of interest were obtained directly from the INEA website (http://inea.infoper.net/inea/), based on the telemetric stations of Olaria and Ypu, located in the Cônego and Santo Antônio rivers, respectively, whose coordinates are described in Table 1, and with 15-min intervals between measurements. As there is no information on the magnitude of the errors present in these measurements, for this work, it was considered that the instruments and the form of measurement are reliable to the point that the errors do not significantly influence the experimental data collected [33].

Table 1 Coordinates of telemetric stations at Cônego and Santo Antônio rivers basins

Station	Latitude	Longitude	Monitored River
Olaria	$22°18'31,83"S$	$42°32'31,96"W$	Cônego
Ypu	$22°17'45,09"S$	$42°31'35,41"W$	Santo Antônio

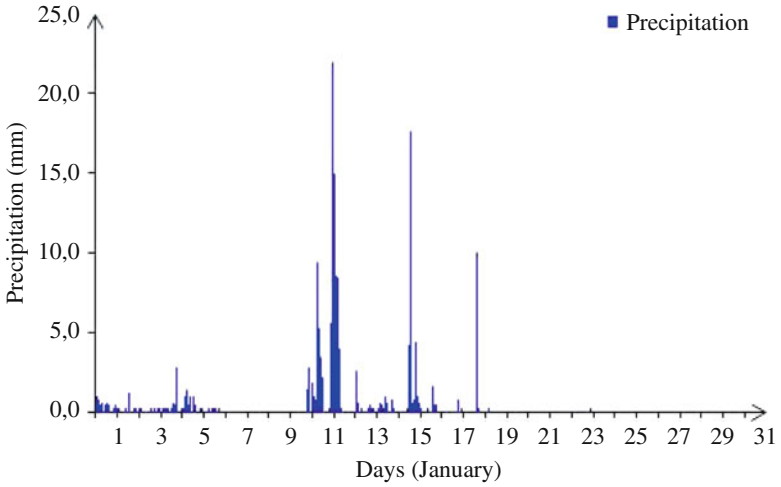


Fig. 6 Precipitation intensity measurements in the period from January, 1st 2011 at 0:00 to February, 1st of the same year at Olaria station

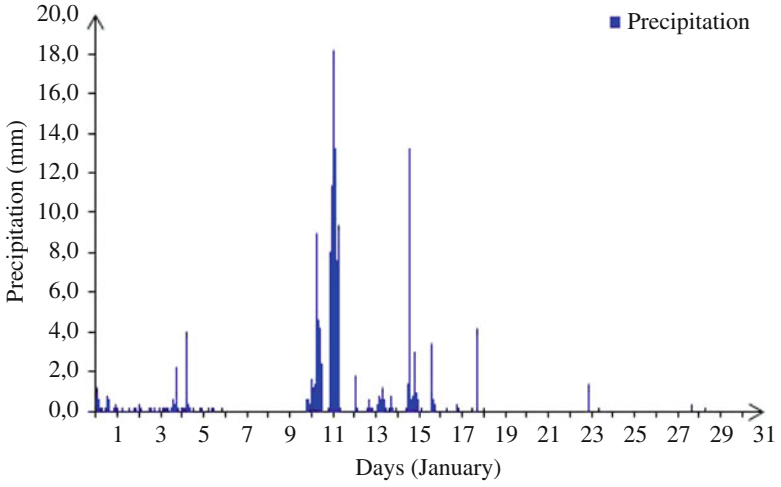


Fig. 7 Precipitation intensity measurements in the period from January, 1st 2011 at 0:00 to February, 1st of the same year at Ypu station

Figures 6 and 7 show the amounts of precipitation in the basins of the Cônego and Santo Antônio rivers from 0:00 h on January 1st, 2011 to 00:00 h on February 1st of the same year, at the Olaria and Ypu stations, which are measured in millimeters at 15-min intervals.

Based on these figures, the intensity of precipitation occurred in the period from 0:00 h on January 11 to 0:00 h on January 13, when the tragedy occurred in 2011.

5.2 The Construction of the Digital Model Terrain

For the simulation of the flood events that occurred in the period that includes the tragedy of 2011 in the Mountain Region of the State of Rio de Janeiro on the MOHID platform, it is necessary to construct a digital model of the terrain so that, based on the topographic profile of the region of interest, it is possible to reproduce the elevation of the water table in the drainage network, due to the rainfall occurring in the basin.

While Telles et al. [31, 34] used topographic data from 1: 50,000 Planialtimetric Letters, provided by the Brazilian Institute of Geography and Statistics (IBGE) and provided by the city hall of Nova Friburgo, the results presented in this section are based on the use of data from topography obtained from the site of the Geomorphometric Database of Brazil, referring to the TOPODATA project (http://www.dsr.inpe.br/topodata/index.php).

Launched for the first time in August 2008, the TOPODATA project allows the user to access the digital elevation model and basic local derivations of any region located within the Brazilian territory, based on Shuttle Radar Topography Mission (SRTM) data provided by United States Geological Survey (USGS) on the World Wide Web [39]. The data adopted in this chapter, downloaded free of charge in the ASCII format (.txt) directly from the TOPODATA project site, use WGS 84 datum and are the result of a re-sampling of the SRTM data covering the Brazilian territory, provided by the National Research Institute (INPE), generating a digital model of terrain with 30 m resolution [40]. In the MOHID GIS environment, the import and conversion of the ASCII (.txt) file into the XYZ Points format was performed, which is the basis for the environment to perform the construction process of the Hydrologically Correct Digital Terrain Model to be used, in later, by the MOHID Land tool in the process of simulating full wave events of interest.

As the MOHID Land tool is a spatially distributed and variable regime model, its application involves the conversion of all input data into a computational grid or in an evolutionary way in time [29]. Thus, a computational mesh was constructed for the Cônego river basin, containing 5760 cells, and could be interpreted as an array with 80 rows and 72 columns. As for the Santo Antônio river basin, the computational grid consisted of 13,650 cells, distributed in 130 rows by 105 columns. In both meshes, a regular spacing of 0.001° decimal was adopted for the cells.

Afterwards, based on the computational meshes and topography data obtained from the TOPODATA project site, the digital models of the hydrologically correct terrains were constructed, giving rise to the files referring to the slope of the cells, flow direction, drainage area, drainage network, and delimitation of the watersheds through the construction of the delimiting polygon of the basins of the Cônego and Santo Antônio rivers. Finally, the cross sections were constructed for each of the drainage channels, which have main water courses, Cônego and Santo Antonio rivers, as well as their monitoring stations, whose coordinates were described in Table 1. The profiles of the transversal sections constructed are based on the Nova

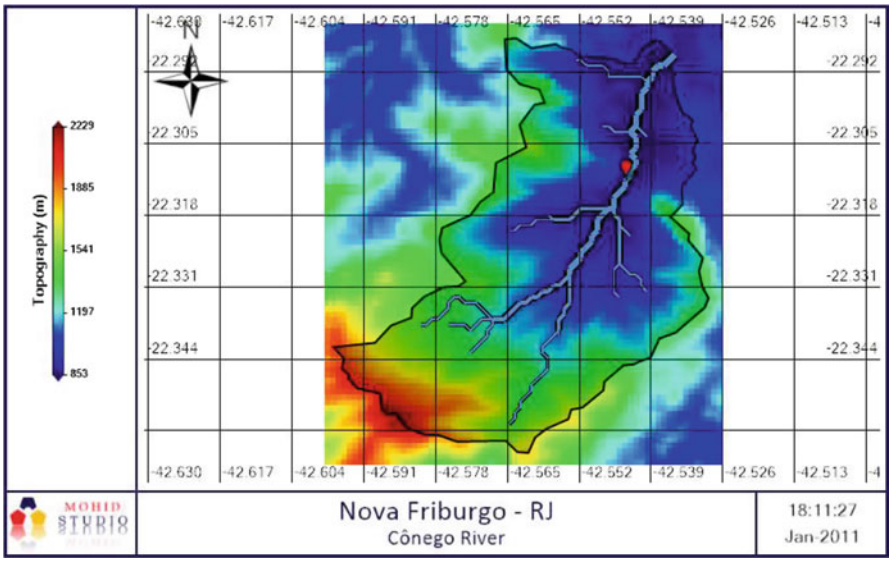


Fig. 8 Digital Model Terrain and delimiter polygon containing the drainage network for the Cônego river basin built by the MOHID GIS environment

Friburgo Pluvial Water Plan, finalized in 2007, and made available by the City Hall of that city [37]. Figures 8 and 9 show the digital terrain models, the delimiting polygon, the drainage network, and the monitoring stations related to the basins of interest. Such processes are described in detail in [31, 34].

5.3 Results Obtained for the Events of Interest

In this section the results of the hydraulic behavior model of the Cônego and Santo Antônio rivers, which compose the Bengalas river basin, located in the municipality of Nova Friburgo—RJ, are presented through the analysis of precipitation events and consequent elevation of the water table of the drainage network occurred in this municipality in January 2011, including the heavy rains that caused the tragedy in the mountainous region.

The results were obtained on an Acer computer, with an Intel Core i5-7200U of 2.50 GHz and 8 GB of RAM. As far as the execution time of the direct problem is concerned, it took approximately 5 min to carry out the simulations in the basins of interest.

Basically, to perform a simulation using the MOHID Land tool, it is necessary to file the hydrologically correct terrain digital model and the drainage network of the region of interest whose constructions are described in Sect. 5.1. The experimental

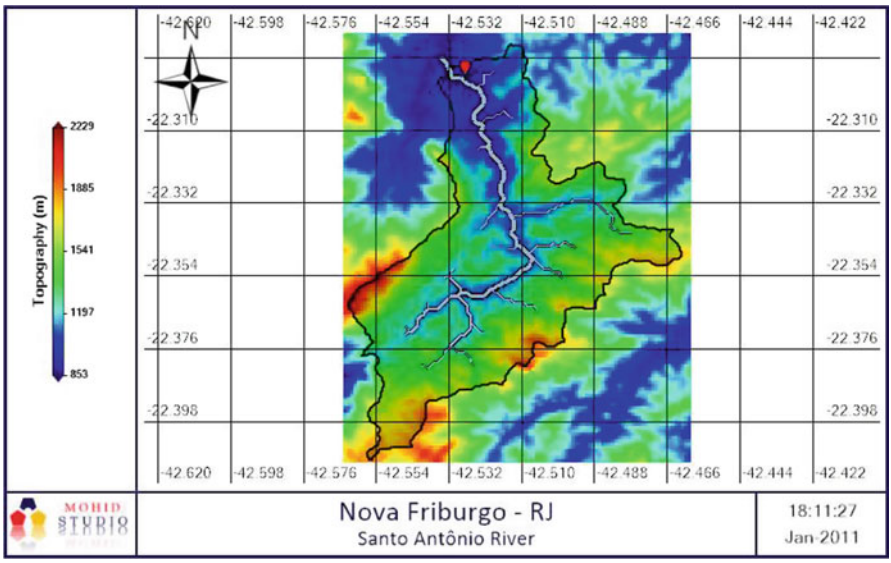


Fig. 9 Digital Model Terrain and delimiter polygon containing the drainage network for the Santo Antônio river basin built by the MOHID GIS environment

data of precipitation used in these simulations, measured in millimeters, are characterized in Sect. 5.1.

The configurations used in the MOHID platform have as reference the parameters adopted by Telles et al. [38] only for the period from January 11 to January 13 of that year, with particular emphasis given to basin and channel roughness coefficients, which were assumed 0.035 s/m$^{1/3}$ and 0.030 s/m$^{1/3}$, respectively. For the other parameters contemplated in the MOHID platform, such as minimum channel slope, hydraulic conductivity, numerical schemes used, among others, an impermeable basin was consecrated, except for the initial water level of the drainage network when starting the simulation, which was considered constant throughout the drainage canal bed, adopting a value of 0.30 m for the Cônego river basin and 0.80 m for the Santo Antônio river basin. Finally, for the time discretization, a time interval of $t = 10$ s was used. In Figures 10, 11, 12, and 13 the profiles of the water table and flood maps in the respective drainage networks of the basins of interest are shown. Based on these figures, it can be verified that the behavior of water levels during the month of January 2011 can be divided into three distinct periods, namely:

- First period (from January 1 to January 10)—low water levels in the drainage network;
- Second period (from January 11 to January 13)—interval of time in which the tragedy occurred in the mountainous region, with high levels of the water table in the drainage network;

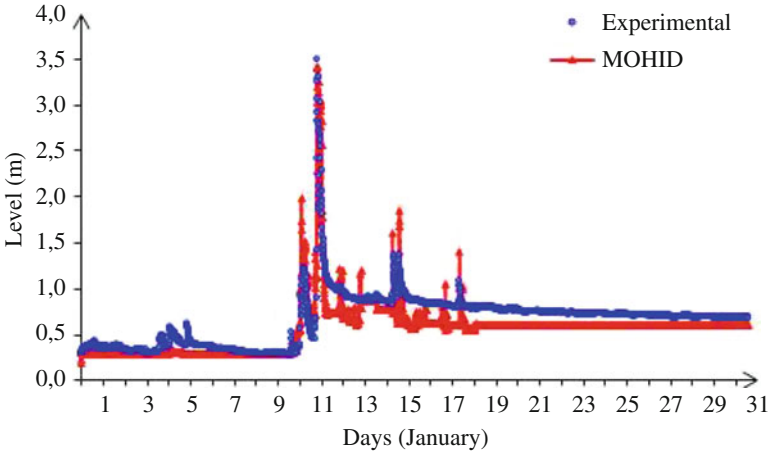


Fig. 10 Water depth profile during the 2011 January event at the drainage network at the Santo Antônio river basin obtained using the MOHID platform and real experimental data

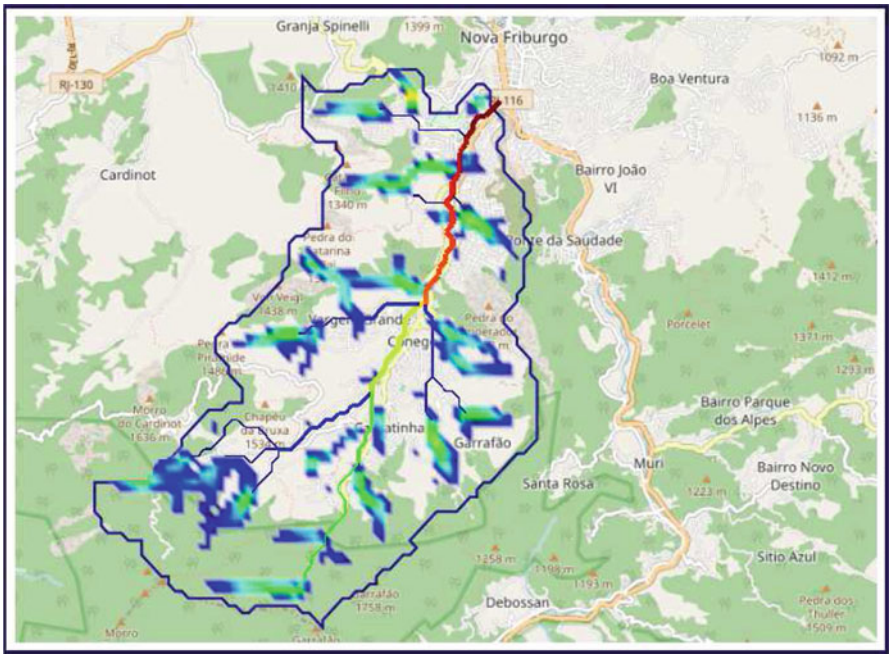


Fig. 11 Flood map during the 2011 January event at the Santo Antônio river basin obtained using the MOHID platform

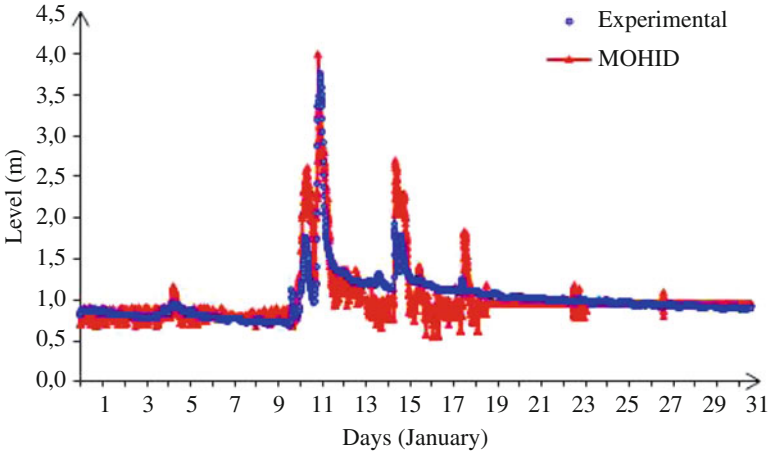


Fig. 12 Water depth profile during the 2011 January event at the drainage network at the Cônego river basin obtained using the MOHID platform and real experimental data

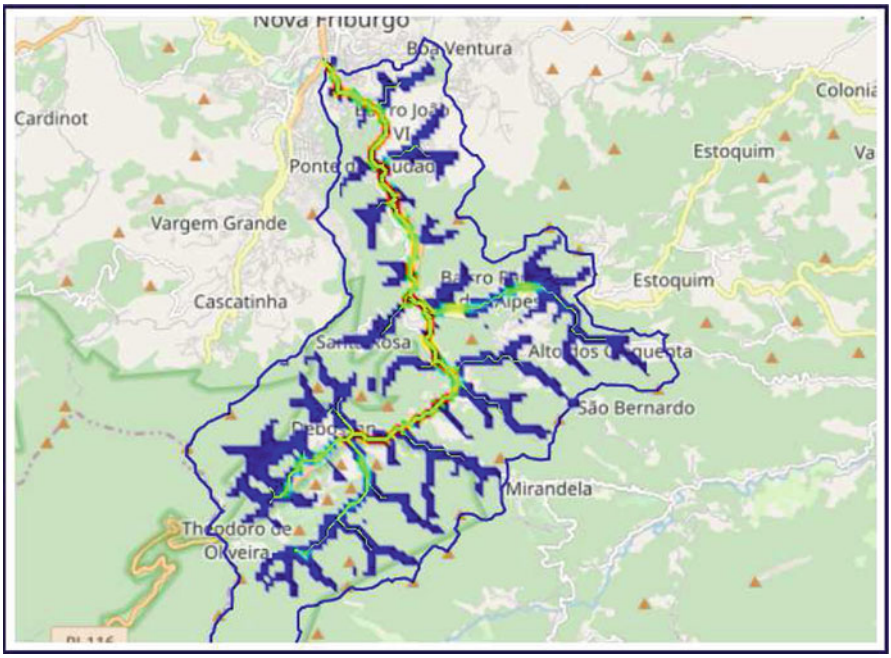


Fig. 13 Flood map during the 2011 January event at the Cônego river basin obtained using the MOHID platform

- Third period (from January 14 to January 31)—post-tragedy period, showing a median elevation of water levels in the drainage network.

Despite the distinct profiles in these three periods, there was a good fit between the experimental data and the numerical results obtained by the MOHID platform for the Olaria (Cônego) and Ypu (Santo Antônio) stations, regarding the time at which the flood waves occurred, where the sum of the squared residuals between the experimental data and those obtained by the MOHID platform was 28.8 and 66.23, respectively.

6 Conclusions

The objective of this chapter was to describe the computational modeling and consequent simulation of drainage flows in watersheds, which allows the understanding of the reasons that lead to the elevation of the water depth in the drainage network of these basins using the MOHID platform and the MOHID Land numerical tool. As a case study, the modeling, as well as the analysis of precipitation events occurred in the watersheds of the Cônego and Santo Antônio rivers located in the city of Nova Friburgo, State of Rio de Janeiro, for the period from January 1, 2011 to 31 of January of the same year, the period in which the greatest natural tragedy occurred in Brazil, more precisely in the Mountainous Region of the State of Rio de Janeiro.

The results obtained here were compared with experimental data from the website of the State Environmental Institute (INEA), where it was concluded that the approach based on the MOHID platform allowed a good fit between the experimental data and the calculated from the model. This result shows MOHID is a promising tool both from the environmental and social point of view since it allows its use in assisting the decision-making process in flood events and forecasts on the impacts and damages that the water level elevation may cause.

Acknowledgements The authors acknowledge the financial support provided by FAPERJ—Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (Rio de Janeiro Research Support Foundation, in Brazil), CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico (National Council for Scientific and Technological Development, in Brazil), and CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Coordination for Enhancement of Higher Education Personnel, Brazil).

References

1. Tingsanchali, T.: Urban flood disaster management. Proc. Eng. **32**, 25–37 (2012)
2. Wang, Y., Chen, A.S., Fu, G., Djordjević, S., Zhang, C., Savić, D.A.: An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features. Environ. Model. Softw. **107**, 85–95 (2018)
3. CRED, Centre for Research on the Epidemiology of Disasters: The International Disaster Database EM-DAT (2018). http://www.emdat.be/

4. Zhou, Q., Leng, G., Huang, M.: Impacts of future climate change on urban flood volumes in Hohhot in northern China: benefits of climate change mitigation and adaptations. Hydrol. Earth Syst. Sci. **22**, 305–316 (2018)
5. Sanders, B.F.: Hydrodynamic modeling of urban flood flows and disaster risk reduction. In: Oxford Research Encyclopedia of Natural Hazard Science (2017) http://naturalhazardscience.oxfordre.com/view/10.1093/acrefore/9780199389407.001.0001/acrefore-9780199389407-e-127?print=pdf
6. Ashley, R.M., Balmforth, D.J., Saul, A.J., Blanskby, J.D.: Flooding in the future predicting climate change, risks and responses in urban areas. Water Sci. Technol. **52**, 265–273 (2005)
7. Larsen, A.N., Gregersen, I.B., Christensen, O.B., Linde, J.J., Mikkelsen, P.S.: Potential future increase in extreme one-hour precipitation events over Europe due to climate change. Water Sci. Technol., **60**, 2205–2216 (2009)
8. Willems, P.: Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle. Belgium. J. Hydrol. **496**, 166–177 (2013)
9. Alfieri, L., Feyen, L., Di Baldassarre, G.: Increasing flood risk under climate change: a pan-European assessment of the benefits of four adaptation strategies. Clim. Chang. **136**, 507–521 (2016)
10. Arnbjerg-Nielsen, K., Leonardsen, L., Madsen, H.: Evaluating adaptation options for urban flooding based on new highend emission scenario regional climate model simulations. Clim. Res. **64**, 73–84 (2015)
11. Moore, T.L., Gulliver, J.S., Stack, L., Simpson, M.H.: Stormwater management and climate change: vulnerability and capacity for adaptation in urban and suburban contexts. Clim. Change **138**, 491–504 (2016)
12. Poussin, J.K., Bubeck, P., Aerts, J.C.J.H., Ward, P.J.: Potential of semi-structural and non-structural adaptation strategies to reduce future flood risk: case study for the Meuse. Nat. Hazards Earth Syst. Sci. **12**, 3455–3471 (2012)
13. Olsson, J., Berggren, K., Olofsson, M., Viklander, M.: Applying climate model precipitation scenarios for urban hydrological assessment: a case study in Kalmar City, Sweden. Atmos. Res. **92**, 364–375 (2009)
14. Willems, P., Arnbjerg-Nielsen, K., Olsson, J., Nguyen, V.T.V.: Climate change impact assessment on urban rainfall extremes and urban drainage: methods and shortcomings. Atmos. Res. **103**, 106–118 (2012)
15. Zahmatkesh, Z., Karamouz, M., Goharian, E., Burian, S.J.: Analysis of the effects of climate change on urban storm water runoff using statistically downscaled precipitation data and a change factor approach. J. Hydrol. Eng. **20**, 1 (2015)
16. Santos, L.L.: Hydrological models: concepts and apply (in Portuguese). Braz. J. Phys. Geogr. **2**, 1–19 (2009)
17. Steinstrasser, C.E.: Lax Diffusive Method applied to Saint Venant Equations (in Portuguese). Dissertation. Paraná Federal University, Curitiba (2005)
18. Liggett, J.A.: Basic equations of unsteady flow. In: Mahmood, K., Yevjevich, V. (eds.) Unsteady Flow in Open Channels. Water Resources Publications, Fort Collins (1975)
19. Ferreira, D.M., Fernandes, C.V.S., Gomes, J.: Verification of Saint-Venant equations solution based on the lax diffusive method for flow routing in natural channels. RBRH **22** (2017). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2318-03312017000100401&nrm=iso
20. Szymkiewicz, R.: Numerical Modeling in Open Channel Hydraulics. Springer, New York (2010)
21. Fan, F.M., Pontes, P.R.M., Paiva, R.C.D., Collischonn, W. Avaliação de um método de propagação de cheias em rios com aproximação inercial das equações de Saint-Venant. Rev. Bras. Recur. Hidr. **19**, 137–147 (2014)
22. Henderson, F.M.: Open Channel Flow. MacMillan, New York (1966)
23. Porto, R.M.: Basic of Hydrology (in Portuguese). EESC-USP, São Paulo (1999)
24. Trancoso A.R., Braunschweig, F., Leitão, P.C., Obermann, M, Neves, R.: An advanced modelling tool for simulating complex river systems. Sci. Total Environ. **407**, 3004–3016 (2009)

25. Maretec: MOHID: DescrDescrição do MOHID. Essentia Editora, Campos dos Goytacazes (2012)
26. Neves, R.J.J.: Étude Experimentale et Modélisation des Circulations Trasitoire et Résiduelle Dans Lestuairedu Sado. Universidade de Liège, Liège (1985)
27. Precioso, C.H.O., Kalas, F.A., Rodrigues, P.P.G.W., Lugon Jr., J.: Evaluation of environmental parameters variability in an urban lagoon (Campos dos Goytacazes/RJ) using the MOHID System (in Portuguese). Boletim do Observatório Ambiental Alberto Ribeiro Lamego **4**, 185–204 (2010)
28. Souza, M.P.G.: Simulação da Dispersão de Óleo na Baía do Espírito Santo Usando o Modelo Numérico MOHID-2D. Universidade Federal do Espírito Santo, Vitória (2010)
29. Braunschweig, F., Fernandez, L.: MOHID: User Manual (in Portuguese). Essentia Editora, Campos dos Goytacazes (2010)
30. Pessanha, C.M.D., Lugon Jr., J., Ferreira, M.I.P., Souza, P.R.N., Hora, H.M.C.: Uso de modelagem computacional aplicada a gestão sanitário-ambiental: uma proposta de adaptação da plataforma MOHID water para corpos lênticos aplicada a lagoa Imboacica, Macaé-RJ. Boletim do Observatório Ambiental Alberto Ribeiro Lamego **5**, 45–70 (2011)
31. Telles, W.R.: Prediction of Hydraulic Behaviour of a river based on the estimated coefficients that controls its flow. Case Study: Bengalas River, Nova Friburgo-RJ (in Portuguese). Ph.D. thesis. Universidade do Estado do Rio Janeiro, Nova Friburgo (2014)
32. Telles, W.R., Silva Neto, A.J., Rodrigues, P.P.G.W.: Avaliação do sistema de modelagem MOHID na delimitação de bacias hidrográficas. Anais do Congresso de Matemática Aplicada e Computacional. Natal 589–592 (2012). http://www.sbmac.org.br/cmacs/cmac-ne/2012/trabalhos/PDF/81.pdf
33. Telles, W.R., Rodrigues, P.P.G.W., Silva Neto, A.J.: Calibração automática de um simulador aplicado a um rio de montanha empregando dados experimentais de precipitação e nível—estudo de caso: Córrego D-Antas, RJ. Rev. Bras. Recur. Hidr. **21**, 143–151 (2016)
34. Telles, W.R., Rodrigues, P.P.G.W., Silva Neto, A.J.: Calibração automática da plataforma MOHID empregando um método estocástico de otimização e dados reais de um evento climático extremo em Nova Friburgo-RJ: Parte 1—preparação do modelo digital do terreno e aquisição dos dados experimentais de precipitação e nível. Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería **33**, 164–170 (2017)
35. Fernandes, R.: Modelação Operacional no Estuário do Tejo. Dissertation. Instituto Superior Técnico de Lisboa, Lisboa (2005)
36. Correia, E.F.G.: Hydrologic Modeling of Bengalas Basin, Nova Friburgo—RJ, using the Geotechnology Potential for the Definition of Areas with Flood Risk (in Portuguese). Dissertation. Universidade do Estado do Rio Janeiro, Rio de Janeiro (2011)
37. FCTH, Fundação Centro Tecnológico de Hidráulica: Plano de Águas Pluviais de Nova Friburgo—PAPNF. FCTH, São Paulo (2007)
38. Telles, W.R., Rodrigues, P.P.G.W., Silva Neto, A.J.: Calibração automática da plataforma MOHID empregando um método estocástico de otimização e dados reais de um evento climático extremo em Nova Friburgo—RJ: Parte 2—análise de sensibilidadee estimativa de parâmetros hidrológicos. Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería **33**, 204–211 (2017)
39. Valeriano, M.M., Albuquerque, P.C.G.: Topodata: processamento dos dados SRTM. INPE, São José dos Campos (2010)
40. Valeriano, M.M.: Dados topográficos. In: Florenzo, T.G. (eds.) Geomorfologia: Conceitos e Tecnologias Atuais, pp. 72–103. Oficina de Textos, São Paulo (2008)

Applied Time Series—Natural Disasters Perspective of Use: Landslide and Flood

Alessandra C. Corsi, Filipe A. M. Falcetta, Marcela P. P. Guimarães, and Eduardo S. de Macedo

1 Introduction

Societies, around the world, are continually living under the impact of the most varied of natural disasters. Natural disasters have affected more and more countries in the world, and Brazil is not immune to these events. To improve the forecasting and the prediction we use time series analyses for landslide and flooding.

Time series are collections of data recorded at a specified time [7, 10, 12]. According to the Pham [29], time series are ordered sequences of values of a variable at equally spaced time intervals.

Time series can be applied to understand the observed data and to fit a model to forecasting and monitoring.

Historically, time series methods were applied to problems in the physical and environmental sciences.

The Emergency Events Database (EM-DAT) from the Centre for Research on the Epidemiology of Disasters (CRED) data show that flooding caused the majority of disasters between 1994 and 2013, accounting for 43% of all recorded events and affecting nearly 2.5 billion people. Storms were the second most frequent type of disaster, killing more than 259,000 people and costing nearly 1 trillion dollars in recorded damages. This makes rainfalls the most expensive type of disaster during the past two decades, and the second most costly in terms of lives lost. Earthquakes (including tsunamis) killed more people than all other types of disaster put together, claiming nearly 750,000 lives between 1994 and 2013. Tsunamis were the most deadly sub-type of earthquake, with an average of 79 deaths for every 1000 people affected, compared to four deaths per 1000 for ground movements [19].

A. C. Corsi (✉) · F. A. M. Falcetta · M. P. P. Guimarães · E. S. de Macedo
Centro de Tecnologias Geoambientais - CTGeo, Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, São Paulo, Brazil
e-mail: accorsi@ipt.br; falcetta@ipt.br; mppg@ipt.br; esmacedo@ipt.br

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_8

Nunes [28] has studied the impacts of 863 natural disasters which occurred in the last five decades in South America, phenomena such as earthquakes and volcanisms resulted in twice as many deaths as hydro meteorological events. But events such as flooding, droughts and landslides affected a major number of people.

In the last decades were registered several events related to mass movement and floods in the country as in Santa Catarina in the year 2008, Alagoas and Pernambuco in 2010, and mountainous region of Rio de Janeiro in 2011. Such events caused losses and damages in the sectors of infrastructure, social, and productive, with a cost of around 15 billion reais, estimated by World Brank [39]. From 1988 to 2018, 3509 people were killed in Brazil, according to the Landslide Deaths Database developed by the Institute of Technological Research of the State of São Paulo (IPT).

In this chapter, we will present a review of the time series analysis applied to geological and hydrological problems, such as landslide and flooding. We illustrated the use of time series in two study cases applied to landslide and flooding.

2 Study Cases

Now, we presented the use of time series to forecasting landslide and the other the use of rainfall time series applied to flooding.

2.1 Time Series Applied to Landslides

Time series applied to landslide study can use several data types, like rainfall, landslide inventory, GPS, digital photogrammetry, water level, displacement data, and others.

Landslides are a worldwide natural hazard causing thousands of fatalities and severe monetary losses every year. They often occur as cascading effects in case of natural disasters such as earthquakes or hydrological extreme events (e.g., typhoons).

Landslide occurrence is affected by many factors, including geological conditions, rainfall, and others.

Landslides are recognized as dynamic and significantly hazardous phenomena. Time series observations could be used to improve the understanding of a landslide complex behavior and aid assessment of its geometry and kinematics [30].

Several works were developed in the world and in Brazil from the 1970s, correlating rainfall and landslide. Figure 1 shows the distribution of some these papers in the world.

The pioneer in this area was [26] when investigating the correlations between rainfalls and landslides in Hong Kong.

Lumb [26] analyzed the recurrence of slope failures in residual soils of Hong Kong (period 1950–1973), and he described various factors which contribute to the

Fig. 1 Some countries where rainfall and landslide correlation studies were performed

instability. It is postulated that the prime cause of the failures is direct infiltration of rainwater into the superficial zones of the slopes, producing a loss of effective cohesion following the saturation of the soil.

Lumb [26] established some useful correlations between the number and severity of rainfall events and the daily, and 15-day antecedent rainfalls. Prevention of slips implies protection against excessive infiltration. Several authors have continued the Lumb studies in Hong Kong, such as Brand et al. [9], Brand [8], Kay and Chen [25], Zhou et al. [41], Dai and Lee [14], among others.

Correlation studies between rainfall and landslides were also developed in Europe, such as Corominas and Moya [13] in Spain, Flageollet et al. [22] in France, Quinta Ferreira et al. [31] in Portugal, and Canuti et al. [11] in Italy are some examples.

Ávila-Parra and Martín Vide [6], Filun [21], and Angulo [2] were some studies developed in Chile.

Ávila Parra and Martín Vide [6] studied the spatial distribution of extreme precipitation risk between the fifth and eleventh regions of the Chilean mainland. To achieve that, probable maximum values of daily precipitation for different return periods are calculated, analyzing the maximum amounts of rainfall in 24 h in 46 rainfall stations with records between 1980 and 2010. The methodological approach has been the adjustment of the data employing different probability laws, especially Gumbel Max and Weibull, and the calculation of maximum values and returns periods. The results show a significant spatial variation of the maximum precipitations, the highest ones being located in the Andean foothills, decreasing gradually to pass the intermediate depression and increasing again in the Costa range. And the highest estimated daily precipitation occurs at Valparaíso, Maule, and Bío-Bío regions, with values equal to or greater than 200 mm in 50 years.

Filun [21] studied historical precipitation records associated with landslide processes on urbanized slopes of the city of Talcahuano in central-southern Chile. He found that the precipitation concentrated in a shorter period than 4 h represents the critical threshold of destabilization, with intensities on the order of 26.5 mm.

Angulo [2] studied the Copiapó river basin, in the southern portion of Chile, where he analyzed the maximum precipitation ratios and debris flows. He examined historical records of rainfall from 1940 to 2015. For the basin, the debris flows develop on steep slopes between 15 and 25° and maximum precipitation of 22 mm in 24, 48, and 72 h.

In Colombia, several studies were developed by Aristizábal and Gómez [3], Moreno et al. [27], Aristizábal et al. [4, 5] among others.

Moreno et al. [27] studied the relationship between rain and reported landslides in the department of Antioquia from 1929 to 1999. Their analysis indicated that the cumulative precipitation (15 days or more) influences the amount of subsequent 3-day rainfall that is needed to initiate ground movement.

Aristizábal and Gómez [3] compare the disaster inventory and rainfall data in the period 1880–2007, identifying its close relationship, with a bimodal tendency and peaks in May and October. Of every ten events that occur in the valley, 8 are of hydrometeorological origin, essentially mass movements, and floods.

Aristizábal et al. [4] propose the existence of four representative regions of antecedent rainfall for mass movements, called regions A, B, C, and D, for the Aburrá valley, based on the thresholds defined by Moreno et al. [27]. In region A, characterized by low levels of rainfall, 25% of events occur, which allows us to suppose that these events are associated with conditions different from the previous rain, such as anthropic intervention. In region B there are few movements in mass and correspond to conditions of low humidity in the ground and isolated downpours, which possibly do not reach to unleash landslide. Region C corresponds to a cumulative rainfall of 15 days between 70 and 150 mm, where 20% of the events occur, and finally region D, which corresponds to a cumulative 15-day rainfall of 150 mm, where the 55% of occasions.

Aristizábal et al. [5] analyzed critical rainfall thresholds for landslides forecasting in the Aburrá Valley. Their results show that the primary determinant for the occurrence of mass movements in the Aburrá Valley is the antecedent rainfall. In Aburrá Valley mass movements used in the analysis occurred for antecedent rainfall over 60 mm for 30 days, 160 mm for 60 days, and 200 mm for 90 days.

In Brazil, the correlation studies between rainfall and landslides had begun in the 1970s, with the efforts of [1, 15–18, 20, 23, 24, 33–38, 40].

Among the works developed in Brazil, it is worth highlighting the pioneering work of [23]. Guidicini and Iwasa [23] in this paper established a precipitation threshold necessary for landslides deflagration. For this purpose, they used time series analysis of heavy rainfall events between 1928 and 1976. They selected nine areas in different regions of Brazil to study the correlation between rainfall and landslide (Caraguatatuba/SP, Baixada Santista/SP, Highway Imigrantes/SP, Highway Anchieta/SP, Serra de Maranguape/CE, Rio de Janeiro/RJ, Serra das Araras/RJ, South State of Minas Gerais/MG, Vale do Tubarão/SC). In their first

analysis, they considered the precipitation levels of the events together with the standards of the days immediately before the events, that is, the sum of continued rains. When they analyzed together landslide occurrences and the rainfall data, the landslides are likely to occur when precipitation records are between 8 and 17% of the mean annual precipitation. The remaining events, those do not cause slides, have precipitation records up to 12% of the mean annual rainfall.

The work developed by Tatizana [35, 36] will be presented in detail in the case study, since until now the correlation established in these works is used by the Civil Defense of the municipality of Cubatão, located in the state of São Paulo, Brazil.

This case study addresses the use of rainfall and landslide time series to obtain rainfall thresholds that could result in landslides or debris flow. The rainfall thresholds are important to monitor and prevent accidents in urban areas and highways. This type of data is very useful to Civil Defense. They used rainfall thresholds during the operation of the Preventive Plan for Landslide in the State of São Paulo.

Tatizana et al. [35] performed a correlation analysis between rainfall and landslides in Serra do Mar, located in the Cubatão municipality (Brazil) (Fig. 2). The methodology used consisted of an analysis of the high rainfall events and the landslides by using retro-analysis.

Tatizana et al. [35] obtained the rainfall time series at the Curva da Onça station. The selected rainfall events were those with values greater than 100 mm in 1 day, 150 mm in 2 days, or 200 mm in 3 days. The total numbers of events found were 35

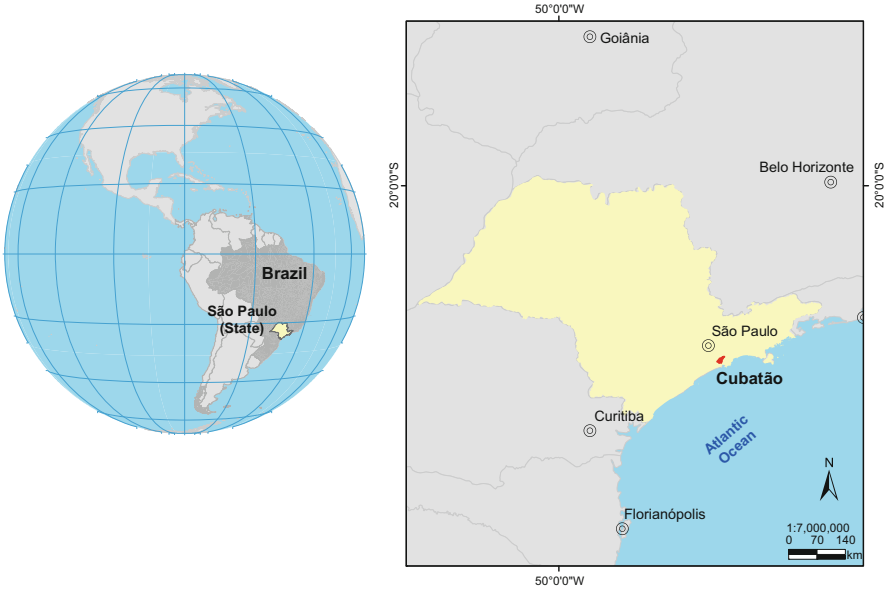


Fig. 2 Serra do Mar, Cubatão municipality, São Paulo State, Brazil

since 1956. They collected the hourly data for the previous 7 days and 1 day after for those events, obtained landslide data from several sources, such as newspaper archives, City Hall, Dersa—Desenvolvimento Rodoviário S/A, industries, technical works, field mapping, and verbal information.

The authors plotted the data in dispersion diagrams with accumulations of 8, 4, 3, and 2 days. The 4-day accumulated diagram showed the best dispersion. However, those simple correlations were not satisfactory to explain all the cases. They plotted the events again in diagrams of accumulated rainfall and the hourly intensity over time. They associated all of those events with heavy rainfall and the highest rainfall accumulated in 4 days (Fig. 3).

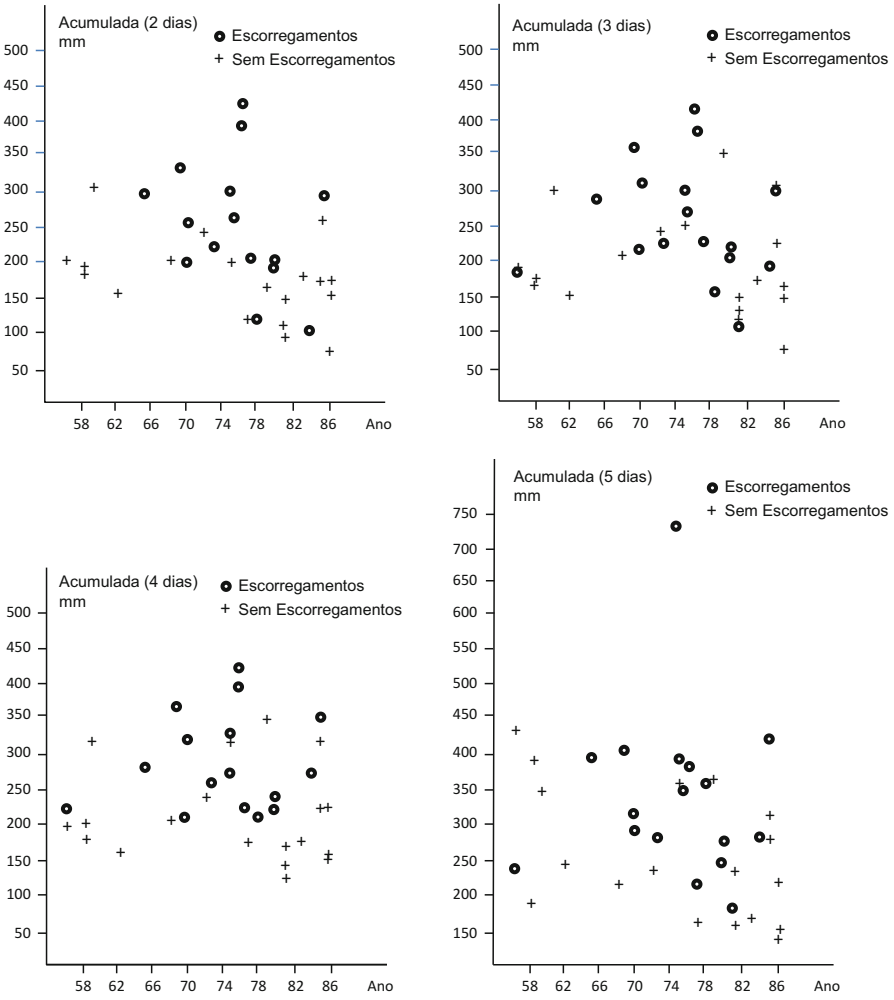


Fig. 3 Dispersion graphics

Tatizana et al. [35] plotted the events in a graphic where the axes x and y represent the accumulated in 4 days and the hourly intensity. They drew a curve that separated events with landslides from those without landslides. A geometric tracing curve was evident (Eq. (1)).

$$I(Ac) = k \times Ac^{-b} \tag{1}$$

I = hourly intensity (mm/h) Ac = accumulated rainfall in 4 previous days (mm) K and b = constants of the geometric relation (they vary with the geotechnical characteristics of the slopes and the climatic conditions).

Tatizana et al. [35, 36] determined the parameters k and b by the least squares numerical method (Eq. (2)):

$$I(Ac) = k \times Ac^{-0.933} \tag{2}$$

Equation (2) was called induced landslide envelopment because the major parts of those records were related to the occurrence of landslides in highways or landslides in slums. Heavy rainfall values beyond the induced landslides envelopment can generate few landslides, generalized landslides, or debris flow.

Tatizana et al. [35] traced the envelopment curves for sparse landslides, generalized landslides, and debris flow (Table 1).

Tatizana et al. [35] established a dimensionless index denominated CPC (critical precipitation coefficient) which measures the landslide susceptibility according to the precipitation evolution.

$$\text{CPC} = li/lci \tag{3}$$

li = hourly intensity (mm/h) recorded at hour i; Ici = critical hourly intensity for the occurrence of induced landslides; As: Aci = accumulated rainfall of the previous 4 days (mm).

So,

$$\text{CPC} = (li/2603) \times Aci^{-0.933} \tag{4}$$

Table 2 shows the relationship between CPC and landslide envelopment.

Tatizana et al. [36] analyzed the event occurred on January 22nd/23rd 1985 for five sectors of Serra do Mar (Fig. 4), in the Cubatão municipality. For those sectors,

Table 1 Landslide envelopment

Landslide envelopment	Equation
Induced landslide	$I(Ac) = 2603 \times Aci^{-0.933}$
Sparse landslide	$I(Ac) = 3579 \times Aci^{-0.933}$
Generalized landslide	$I(Ac) = 5466 \times Aci^{-0.933}$
Debris flow	$I(Ac) = 10646 \times Aci^{-0.933}$

Table 2 Relation between CPC and landslide envelopment

CPC	Landslide envelopment
1.0	Induced landslide
1.4	Sparse landslide
2.1	Generalized landslide
4.1	Debris flow

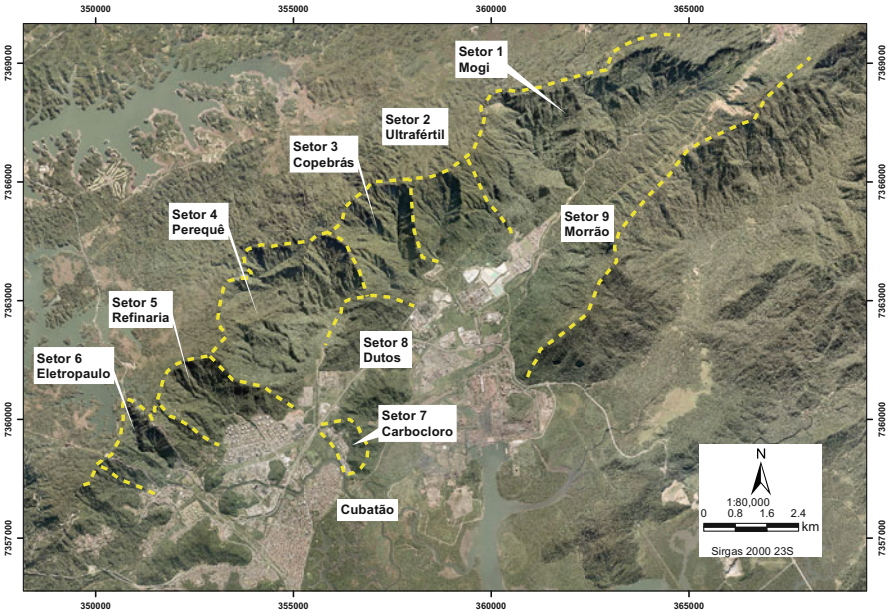


Fig. 4 Industrial sectors

Table 3 CPC sectors equations

Sector	Equation
Eletropaulo+bairros cota	$\text{CPC} = li/3467 \times Aci^{-0.933}$
Refinaria+sub-setor onça	$\text{CPC} = li/2603 \times Aci^{-0.933}$
Perequê	$\text{CPC} = li/2357 \times Aci^{-0.933}$
Copebrás+ultrafértil	$\text{CPC} = li/2033 \times Aci^{-0.933}$
Sub-setor paranapiacaba+ultrafértil	$\text{CPC} = li/3945 \times Aci^{-0.933}$

they determined the equation used to monitor those areas (Table 3). Cubatão Civil Defense used those equations to monitor the Industrial Areas.

According to the Resolução CMIL 17–610 [32], the Contingency Plan for the Serra do Mar in the Cubaão Industrial region aims to provide the participating entities with instruments of action for the prevention and minimization of impacts caused by possible floods and/or landslides on the Serra do Mar slopes, according to the area covered in Fig. 1, which is divided into sectors for the operationalization of the Plan.

The monitoring criteria were defined based on rainfall parameters, meteorological, and field surveys.

For the Cubatão industrial region, located near the Serra do Mar, we expected localized and generalized landslides and debris flow (Figs. 5 and 6). As previously discussed from the works of Tatizana et al. (1987), Civil Defense used to monitor the area the critical precipitation coefficient (CPC). The analysis of this parameter (CPC), associated with the information of the meteorological forecast, and the accumulated values of rainfall in 84 h, and information about landslide occurrences obtained in field surveys, allows the triggering of preventive or corrective actions associated with different states of contingency plan operation. The plan operates with four states: I—State Observation, II—State Attention, III—Critical State, and IV—State Emergency (Resolução CMIL 17–610) [32].

Civil Defense used a network of rain gauges to monitor the rain which is part of the monitoring network of the Department of Water and Electric Power of the São Paulo State (DAEE).

Fig. 5 Generalized landslide in the Serra do Mar, Cubatão, Brazil, by IPT

Fig. 6 Debris flow events occurred in 1994, basin Pedras river, by IPT

2.2 Time Series Applied to Flood

Hydrology is the Geoscience that studies phenomena allowing to determine temporal and spatial distribution of water resources, under attributes of quantity, quality, and their interaction with society. The intensities of these are space-temporally variable, as a result of climatic, geomorphological, and land use variability; therefore, they can be considered functions of time, space or both, at different geographic scales.

Hydrological processes, such as precipitation, evapotranspiration, and surface runoff, are considered stochastic—governed by probabilistic laws—since they are formed by arbitrary components that overlap their regularities. For instance, flood characteristics of a certain river basin can be considered random, since not all the causes of the phenomenon can be determined.

As the simultaneous usage of deterministic and stochastic approaches becomes increasingly more common, we may observe a trend of reducing the aleatoriety of hydrological processes; the study of time series is relevant to this trend, where the variability of the occurrence is recorded from sequential time/space measurements of the corresponding hydrological variable.

The approach onto time series in hydrology considers that all series are representative, stationary (i.e., it doesn't have any trend), and homogenous. This is important because the hydrological studies must allow us to extract from the sample data the probability that such a variable will equal or exceed a certain reference value which has not yet been observed.

The estimation of the recurrence probability of the studied phenomena becomes more challenging when it is possible to observe a seasonality of measurements and, eventually, trends resulting from regional variations induced by anthropic action.

Hydrological data are considered random variables with seasonal variations that could be irregular, enabling the occurrence of extremes.

Thus, the hydrological variables are always associated with occurrence probability, and the recurrence interval of hydrological phenomena could be evaluated using theoretical statistics. The recurrence interval or return period is defined as the inverse of the probability that the event will be exceeded in any 1 year. This parameter is essential for drainage works evaluation, and it is determined by statistical analysis of hydrometric time series of, for instance, precipitation, water level, or discharges.

If an extreme hydrologic event, as an extreme rainfall, for instance, is equalled or exceeded on average once in every 100 years, it can be said that the recurrence interval will be 100 years. It does not mean that the event will occur only once in a period of 100 years, but it is possible to infer that this event has a 1% probability of being equalled or exceeded every subsequent year.

The empirical probability distribution could be used in order to determine the probability of a maximum annual rainfall event. For this, the daily maximum rainfall for every hydrological year is sorted in the descending order and the exceedance empirical probability of each event could be obtained using the Weibull distribution formula.

$$P = \frac{m}{N+1} \tag{5}$$

N is the sample size (number of analyzed annual events) and m is the rainfall order ($m = 1$ for the higher rainfall of the time series and $m = N$ for the lower one).

In order to illustrate the concept, Table 4 below shows the empirical probability of maximum annual daily rainfall for the pluviometric data in the city of Lençóis Paulista, state of São Paulo (southeast of Brazil—Fig. 7), between the years 1972 and 2016.

The mainly disadvantage the using the empirical probability is the impossibility of extrapolate the estimation for recurrence intervals higher than the sample size. For doing this, data must be fitted in continuous probability distributions which assume the extreme value theory as a preliminary hypothesis on their premises. Thus, in extreme hydrological event studies, the most used probability distributions are Log-Normal, Log-Pearson type III, and Gumbel.

Log-Normal distribution is an asymmetric generalization of Normal distribution, considering a modification of the random variable $Y = \ln X$. The density function of this distribution is defined as below:

$$f(x) = \frac{1}{\sigma_y \sqrt{2\phi}} e^{-\frac{(Y-\mu_y)^2}{2\sigma_y^2}} \tag{6}$$

for x > 0.

Table 4 Maximum annual daily rainfall sorted in the descending order and empirical probability associated

Year	Maximum daily rainfall (mm)	Order	Empirical probability	Recurrence interval =1/p(years)
2001	143.8	1	0.02	41.0
2003	118.1	2	0.05	20.5
1975	117.5	3	0.07	13.7
2013	113.3	4	0.10	10.3
1989	110.2	5	0.12	8.2
1996	103.3	6	0.15	6.8
2002	103.0	7	0.17	5.9
1987	102.9	8	0.20	5.1
2015	102.5	9	0.22	4.6
1982	95.7	10	0.24	4.1
1983	90.7	11	0.27	3.7
1977	90.6	12	0.29	3.4
1998	89.2	13	0.32	3.2
1991	88.8	14	0.34	2.9
1985	85.7	15	0.37	2.7
1994	84.3	16	0.39	2.6
1980	83.0	17	0.41	2.4
2014	82.9	18	0.44	2.3
1995	82.8	19	0.46	2.2
1999	81.2	20	0.49	2.1
1976	80.7	21	0.51	2.0
1990	78.0	22	0.54	1.9
1973	75.0	23	0.56	1.8
2011	74.2	24	0.59	1.7
2005	73.3	25	0.61	1.6
1986	70.8	26	0.63	1.6
1997	70.3	27	0.66	1.5
1993	69.1	28	0.68	1.5
1988	66.0	29	0.71	1.4
1981	65.7	30	0.73	1.4
1984	64.9	31	0.76	1.3
2007	64.6	32	0.78	1.3
2006	64.2	33	0.80	1.2
2000	63.0	34	0.83	1.2
1978	61.6	35	0.85	1.2
2008	60.5	36	0.88	1.1
1974	58.0	37	0.90	1.1
1992	57.0	38	0.93	1.1
2004	53.2	39	0.95	1.1
1979	47.9740	0.98	1.0	

Fig. 7 Localization of Lençóis Paulista

The Log-Normal parameters are μy and σy, respectively, the average and the standard deviation of the variable $Y = \ln X$, determined using the expressions below:

$$\mu_y = \frac{\sum_{i=1}^{n} Y_i}{n} \tag{7}$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} \times (Y_i - \sigma_y)^2}{n-1}} \tag{8}$$

The Log-Pearson type III distribution is an asymmetric generalization of Pearson type III distribution, with a variable change similar to Log-Normal, $Y = \ln X$. The density function of this distribution is, then:

$$f(x) = \frac{\lambda^{\beta}(Y-\epsilon)^{\beta-1} e^{-\lambda(Y-\epsilon)}}{(\beta-1)!} \tag{9}$$

when $\ln x \geq \varepsilon$.

The parameters β, λ, and ε could be determined using the expressions below:

$$\beta = \left(\frac{2}{\gamma_y}\right)^2 \tag{10}$$

γ_y is the asymmetry coefficient

$$\lambda = \frac{\sigma_y}{\sqrt{\beta}} \tag{11}$$

$$\epsilon = \mu_y - \sigma_y\sqrt{\beta} \tag{12}$$

Lastly, the Gumbel distribution, also called the generalized extreme value distribution type-I, has its probability density function defined as below:

$$f(x) = \frac{1}{\alpha}e^{\left[-\frac{x-u}{\alpha}-e^{\left(\frac{x-u}{\alpha}\right)}\right]} \tag{13}$$

with $-\infty < x < +\infty$

The probability distribution function assumes the formula below:

$$F(x) = e^{-e^{\left(\frac{x-u}{a}\right)}} \tag{14}$$

Parameters α and u could be determined using the expressions below:

$$\alpha = \frac{\sqrt{6\sigma_x}}{\phi} \tag{15}$$

$$u = \mu_x - 0.5772\alpha \tag{16}$$

In order to use the probabilities distribution to estimate the recurrence interval of daily rainfall, the time series length must be at least 30-year long.

As an example, Fig. 8 shows the annual maximum daily rainfall of Lençóis Paulista time series fitting using the Gumbel distribution.

Using the rainfall extremes statistical analysis, it is possible to plot a graph correlating the recurrence interval (in years) with the probable maximum daily rainfall (in mm). This graph is presented in Fig. 9.

It could be noted that, for the selected five extreme daily rainfall events that occurred in Lençóis Paulista in 2011 and 2016, the recurrence intervals estimated by the statistical extrapolation using Gumbel distribution vary between 6 years and over 10,000 years, for the 260 mm daily rainfall that occurred on January 13th, 2016.

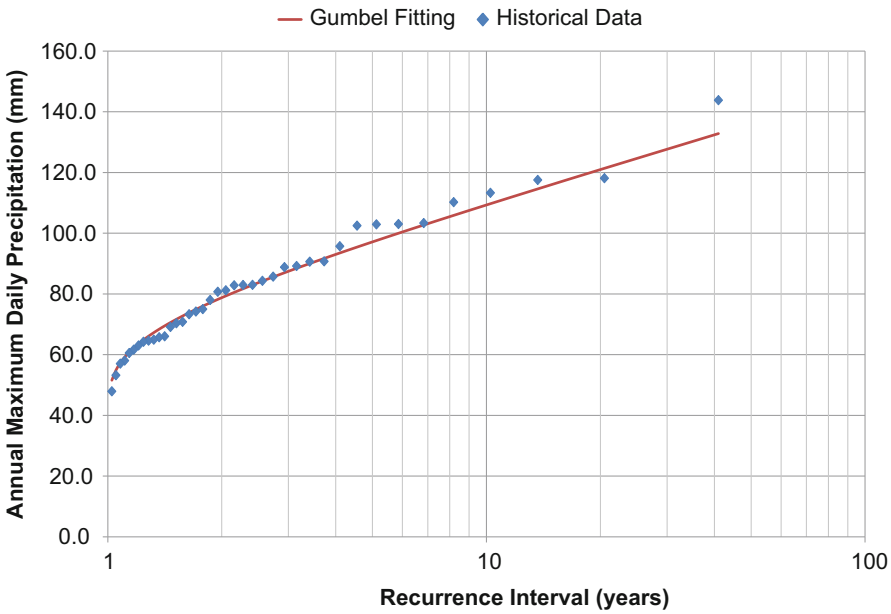


Fig. 8 Time series fitting using Gumbel distribution

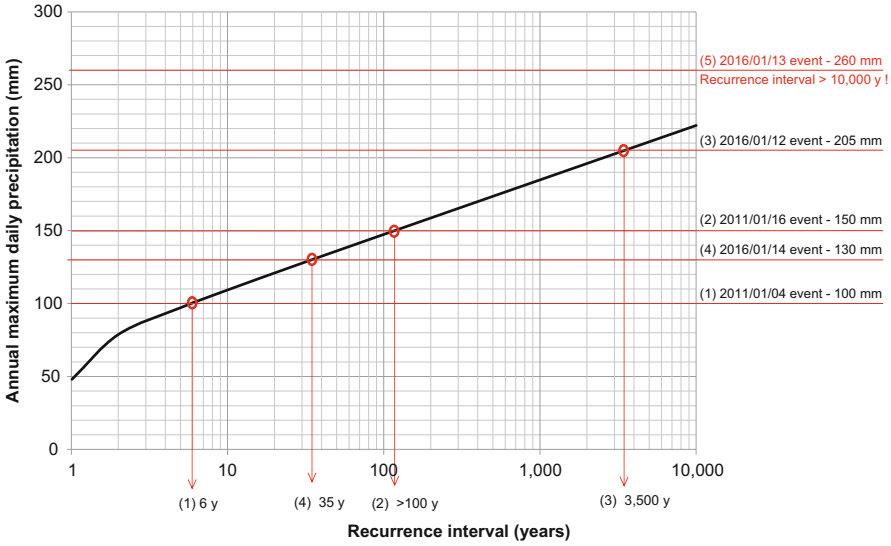


Fig. 9 Estimated recurrence intervals for the extreme daily rainfall in Lençois Paulista

3 Conclusion

The study of time series is important because, from past events observation, it could be possible to predict future behavior of climatic variables, reducing the probability of occurrence of disasters, by directing the recurrence of extreme events, and could be used to evaluate costs and, upon that, define priorities of action in municipal policies and/or urban planning.

In addition, data observations over such long data interval allowed the Civil Defense to monitor landslide and flood risk areas and the evolution of the process during an extreme event.

Obtaining time series of past landslides is an excellent opportunity to predict landslides and reduce its impact on society.

Therefore, in scenarios of climatic uncertainties, knowing the climatic time series allows a strategic gain in the capacity of resilience in extreme events (like floods or droughts) and in the planning of water resources.

References

1. Alheiros, M.M., Bitoun, J., Medeiros, S.M.G.M., Amorin Júnior, W.M.: manual de ocupação de morros da Região Metropolitana de Recife. Recife: Fundação de Desenvolvimento Minicipal (2003). Available via DIALOG. Accessed 30 Jan 2018. http://pt.scribd.com.br/doc/114373442/Manual-de-Ocupacao-de-Morros
2. Angulo, M.E.E.: Amenaza debido a eventos de precipitación máxima entre los 21° y 34° latitud Sur de Chile Continental: análisis de umbrales hidrometeorológicos, gatillantes de remoción em masa por flujos en la Cuenca del Rio Copiapó. Dissertación de Maestria. Facultad de Arquitectura y Urbanismo. Universidad de Chile, 188pp. (2015)
3. Aristizábal, E., Gómez, J.: Inventario de emergencias y desastres en el Valle de Aburrá: originados por fenómenos naturales y antrópicos en el periodo 1880–2007. Revista Gestión y Ambiente **10**(2), 17–30 (2007)
4. Aristizábal, E., Gamboa, M., Leoz, F.: Sistema de alerta temprana por movimientos em masa inducidos por lluvia para el valle de Aburrá, Colombia. Rev. EIA Esc. Ing. Antioq. **13**, 155–169 (2010)
5. Aristizábal, E., gonzález, T., Montoya, J.D., Vélez, J.I., Martínez, H., Guerra, A.: Análisis de umbrales empíricos de lluvia para el pronóstico de movimientos en masa en el valle de Aburrá, Colombia. Rev. EIA Esc. Ing. Antioq. **15**, 95–111 (2011)
6. Ávila-Parra, K., Martín-Vide, J.M.: Análisis estadístico de los eventos extremos de precipitación en la zona centro y sur de Chile Continental. Cuadernos Geográficos **52**(1), 69–83 (2013)
7. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. Wiley, New York (2015)
8. Brand, E.W.: Correlation between rainfall and landslides. In: Proceedings of 12th International Conference on Soil Mechanics and Foundation Engineering, Rio de Janeiro, vol. 1, pp. 70–72 (1989)
9. Brand, E.W., Premchitt, J., Phillipson, H.B.: Relationship between rainfall and landslides in Hong kong. In: Proceedings of 4th International Symposium on Landslides. Toronto, vol. 1, pp. 377–384 (1984)
10. Brockwell, P.J., Davis, R.A.: Introduction Time Series and Forecasting. Springer, Berlin (2002)

11. Canuti, P., Focardi, P., Garzonio, C.A.: Correlation between rainfall and landslides. Bull. Int. Assoc. Eng. Geol. **32**, 49–54 (1985). https://doi.org/10.1007/BF02594765
12. Chatfield, C.: The Analysis of Time Series: An Introduction. Chapman & Hall/CRC, London (2003)
13. Corominas, J., Moya, J.: Reconstructing recent landslide activity in relation to rainfall in the Llobregat River basin, Eastern Pyrenees, Spain. Geomorphology **30**, 79–93 (1999)
14. Dai, F.C., Lee, C.F.: Frequency-volume relation and prediction of rainfall-induced landslides. Eng. Geol. **59**, 253–266 (2001). https://doi.org/10.1016/S0013-7952(00)00077-6
15. D'Orsi, R.N.: Correlação entre pluviometria e escorregamentos no trecho da Serra dos Órgãos da Rodovia Federal BR-116 RJ (Rio-Tersópolis). In: Tese de Doutorado. Universidade Federal do Rio de Janeiro - UFRJ/COPPE. Rio de Janeiro, 287pp. (2011)
16. D'Orsi, R.N., D'Ávila, C., Ortigão, J.A.R., Dias, A., Moraes, L., Santos, M.D.: Rio-watch: the Rio de Janeiro Landslide Watch System. In: Conferência Brasileira sobre Estabilidade de Encostas, Anais. Rio de Janeiro, vol. 1, pp. 21–30 (1997)
17. D'Orsi, R.N., Feijó, R.L., Paes, N.M.: Rainfall and mass movements in Rio de Janeiro. In: Proceedings of 31st International Geological Congress, Rio de Janeiro (2000)
18. Elbachá, A.T., Campos, L.E.P., Bahia, R.F.C.: Tentativa de correlação entre precipitação e deslizamentos na cidade de Salvador. In: 1ł Conferência Brasileira sobre estabilidade de Encostas. Anais, Rio de Janeiro, vol. 2, pp. 647–656 (1992)
19. EM-DAT: The human cost of natural disasters - a global perspectives (2015). https://reliefweb.int/sites/reliefweb.int/files/resources/PAND_report.pdf
20. Feijó, R.L., Paes, N.M., DÒrsi, R.N.: Chuvas e movimentos de massa no município do Rio de Janeiro. In: 3 Conferência Brasileira sobre Estabilidade de Encostas. Anais, Rio de Janeiro, pp. 223–230 (2001)
21. Filun, P.A.L.: Análisis de umbrales de precipitación de procesos de remoción en masa en laderas urbanizadas de la costa de Chile centrosur. Ncuadernos de Geografía: Revista Colombiana de Geografía (2015). https://doi.org/10.15446/rcdg.v4n2.50212
22. Flageollet, J.C., Maquaire, O., Martin, B., Weber, D.: Landslides and climatic conditions in the Barcelonnette and Vars basins (Southern French Alps, France). Geomorphology **30**, 65–78 (1999)
23. Guidicini, G., Iwasa, O.Y.: Tentative correlation between rainfall and landslides in a humid tropical environment. Bull Int. Assoc. Eng. Geol. **16**, 13–20 (1977)
24. Ide, F.S.: Escorregamento, meteorologia e precipitação: uma proposta de método de investigação para a prevenção e monitoramento de riscos, aplicado em Campinas/SP. Dissertação de Mestrado em Tecnologia Ambiental, Instituto de Pesquisas Tecnológicas do Estado de São Paulo. São Paulo, 153pp. (2005)
25. Kay, J.N., Chen, T.: Rainfall-landslide relationship for Hong Kong. In: Proceedings of the Institution of Civil Engineers Geotechnical Engineering, Bangkok, vol. 113, pp. 117–118 (1995)
26. Lumb, P.: Slope failure in Hong Kong. Quar. J. Eng. Geol. **8**, 31–65 (1975)
27. Moreno, H.A., Velez, M,V., Montoya, J.D., Rhenals, R.L.: La lluvia y los deslizamientos de tierra en Antioquia: análisis de su ocurrencia en las escalas interanual, intraanual y diaria. Rev. EIA Esc. Ing. Antioq. **5**, 59–65 (2006)
28. Nunes, L.H.: Urbanização e desastres naturais na América do Sul: abangência América do Sul. Oficina de Textos, São Paulo (2015)
29. Pham, H.: Handbook of Engineering Statistics. Springer, Berlin (2006)
30. Peppa, M.V., Moore, P., Miller, P., Chambers, J.: 3D landslide motion from a UAV-derived time-series of morphological attributes. Nat. Hazards Earth Syst. Sci. Discuss. **17**, 2143–2150 (2017). https://doi.org/10.5194/nhess-2017-201
31. Quinta Ferreira, M., Lemos, L.J.L., Pereira, L.F.M.: Influência da precipitação na ocorrência de deslizamentos em Coimbra, nos últimos 139 anos. Revista Portuguesa de Geotecnia **104**, 17–30 (2005)
32. Resolução CMIL 17-610, de 27-11-2017. Available via DIALOG. Accessed 20 May 2018. http://www.defesacivil.sp.gov.br/?page_id=322

33. Salaroli, I.S.: Movimentos de massa no município de Vitória, Inventário, caracterização e indicativos de um modelo comportamental. Dissertação de Mestrado em Engenharia Ambiental, Universidade Federal do Espírito Santo, 148pp. (2003)
34. Silva, N.L.: Correlação entre pluviosidade e movimentos gravitacionais de massa no Alto Ribeirão do Carmo/MG. Dissertação de Mestrado em Geotecnia, Universidade Federal de Ouro Preto, 114pp. (2014)
35. Tatizana, C., Ogura, A.T., Cerri, L.E. da S., Rocha, M.C.M. da: Análise de correlação entre chuvas e escorregamentos - Serra do Mar, município de Cubatão. Congresso Brasileiro de Geologia de Engenharia **5**, 225–236 (1987)
36. Tatizana, C., Ogura, A.T., Cerri, L.E. da S., Rocha, M.C.M. da: Modelamento mumérico da análise de correlação entre chuvas e escorregamento aplicado as encostas da Serra do Mar no munucípio de Cubatão. Congresso Brasileiro de Geologia de Engenharia **5**, 237–248 (1987)
37. Tavares, R., Armani, G., Pressinotti, M.M.N., Santoro, J., Galina, M.H.: Análise da variabilidade temporal e espacial das chuvas associada aos movimentos gravitacionais de massa na Baixada Santista - SP. Congresso Brasileiro de Geologia de Engenharia e Ambinetal, Florianópolis, SC, vol. 11, 17pp. (2005)
38. Vieira, R.: Um olhar sobre a paisagem e o lugar como expressão do comportamento frente ao risco de deslizamento. Tese de Doutorado. Departamento de Geociências, Universidade Federal de Santa Catarina, Florianópolis, 197pp. (2004)
39. World Brank: Com gerenciamento de desastres, Brasil, poderia economizar bilhões de reais (2012). Available via DIALOG. Accessed 17 April 2018. http://www.worldbank.org/pt/news/feature/2012/11/19/Brazil-natural-disaster-management-costs-development
40. Xavier, H.: Percepção geográfica dos deslizamentos em áreas de risco no município de Belo Horizonte, MG. Tese de Doutorado. Universidade Estadual Paulista, Rio Claro, 222pp. (1996)
41. Zhou, C.H., Lee, C.F., Xu, Z.W.: On the spatial relationship between landslides and causative factors on Lantau Island, Hong Kong. Geomorphology **43**, 197–207 (2002)

Bayesian Analysis of the Disaster Damage in Brazil

Camila Bertini Martins, Viviana Aguilar Muñoz, André Yoshizumi Gomes, Ricardo Manhães Savii, and Carolina Locatelli Colla

1 Introduction

Disaster risk reduction (DRR) is a priority action for sustainable development, due to the disaster impacts on all societies; its study and understanding are a matter to several knowledge areas. Among the exact sciences, the statistical sciences play an important role in that sense. One of the implementation means of the sustainable development goals (SDG) [22] is to intensify our efforts to strengthen statistical capacities in developing countries. In fact, those goals and their targets must be followed up and reviewed by using a set of global indicators approved by a Statistical Commission and complemented by regional and national indicators [23].

Data is another sensible aspect of DRR and SDG strategies. The 2030 Agenda refers to data requirements needed to help with the measurement of SDGs progress, which will be paramount to decision-making: top quality, accessible, timely, and reliable, disjointed in relevant characteristics of national contexts. In the same way, the Sendai Framework [21] promotes collection, analysis, management and use of relevant, reliable and real-time access data to help global, regional, and national

C. B. Martins (✉)
Federal University of São Paulo, Paulista School of Medicine, Department of Preventive Medicine, São Paulo, SP, Brazil
e-mail: cb.martins@unifesp.br

V. A. Muñoz
National Centre for Monitoring and Early Warnings of Natural Disasters (CEMADEN), São José dos Campos, SP, Brazil

A. Y. Gomes
Serasa Experian, Decision Analytics, São Paulo, SP, Brazil

R. M. Savii · C. L. Colla
Federal University of São Paulo, Institute of Science and Technology, São José dos Campos, SP, Brazil

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_9

DRR programs. The disaster loss databases on detailed scale and relevant non-aggregated data are some useful supports to research disaster risk patterns, causes, and effects, as well as to strengthen disaster risk modeling, assessment, mapping, monitoring, and multi-hazard early warning systems [21]. The Sendai Framework also highlighted the importance of development and dissemination of science-based methodologies and tools to record and share related databases.

Statistical methodologies and tools play a relevant role for the organization and analysis of those data. The international, regional, and bilateral support to strengthen countries statistical capacity for DRR is one of the eight indicators of the Sendai Global Target F: it substantially enhances international cooperation to developing countries through adequate and sustainable support to complement their national actions for implementation of this framework by 2030. The development of all standards and metadata, methodologies, training and technical support, and technical material guidance for follow-up and operationalization of the Sendai indicators are relevant for DRR [24].

The exploratory statistical analysis of these data helps, first and foremost, in identifying variables and patterns that may be related to risk attributes, both qualitative and quantitative, and in simulating possible scenarios of risk. An example of spatial pattern could be the recurrent distribution of floods in certain territories; a time pattern may be their recurrence in the same month or season along many decades. On the other hand, disaster frequency may indicate levels of territory susceptibility and levels of population vulnerability.

This chapter presents a Bayesian inference statistical methodology applied to a disaster database to identify some impact patterns of natural phenomena in Brazil, in terms of their typology, frequency, and regional distribution (South, Southeast, Central West, North, and Northeast region). It is a diagnostic study that intends to help researches on socioeconomic component and disaster risk modeling. The systematic registry and organization of disaster occurrences, as well as the statistical processing of long-term data, allow the identification of disaster background causes in the country, as they may directly support monitoring and warning systems and, indirectly, the development of public policies to adapt to climate change and DRR.

This work is not intended to be an exhaustive study; it is also not the first work that presents a diagnosis of the distribution of risk and disaster patterns in Brazil. The focus of this work is to highlight the importance of investing in the exploration of statistical tools that are suitable for complex analysis of occurrence and disaster impact data, particularly in the Bayesian framework. It is important to draw attention to the importance of developing methodologies for statistical modeling capable of achieving results with the minimum of uncertainties in more detailed studies on risk and disasters.

2 Material and Methods

2.1 Field of Study

Brazil is one of the largest countries in the world, with more than eight million km^2 of land extension. In this territory, there is a great diversity of natural phenomena (i.e., hydrological, geological, meteorological, etc.) that can cause harm. Furthermore, according to the most up-to-date estimate projections by the national census authority [15], today, Brazil is the sixth most populous country worldwide, with more than 208 million of people. This geographical complexity is a prominent condition to understand the root causes of risks and disasters in this country, because its territory is especially susceptible to climate changes [6] and its population is notably vulnerable to the impact of climate-associated phenomena [12]. It is important to keep in mind that risk equation involves integrating natural and social forces; therefore, risk and disaster analysis must consider both natural hazards and social vulnerabilities [24].

To ease risk and disasters research at national level, it is advisable to consider some territorial zoning in sets of homogeneous regions, either physical or social. As a prior approach to achieve the objectives of this work, it was used, as a geographical reference, the officially recognized regional division of Brazil [15]: North, Northeast, Central West, South, and Southeast regions; note that each region is a group of several states (Fig. 1). Despite being a merely academic division (the regions do not have political autonomy), because it considers geographic, social, and economic factors, it is useful for statistical evaluations, public management, and economic and socio-environmental research.

2.1.1 Disaster Database

Disaster records for the 2003–2016 period were used as input data for this study. These data were obtained from the S2ID (Integrated Disaster Information System) database of the Secretaria Nacional de Proteção e Defesa Civil - SEDEC [4]. Those data are available to download from the SEDEC website; the records are arranged in a spreadsheet by date of occurrence, cities/state affected, and type of event or natural phenomena according to the Cobrade: Brazilian Code of Disasters [8]. Cobrade comprises geological, hydrological, meteorological, climatological, biological, and technological disasters. A spreadsheet with a total of 28,011 records was downloaded from the S2ID database; 51 of those were excluded from the studies because they did not belong to Cobrade. Then, 27,960 records were considered for this work, and they relate to the same number of documents about disasters caused by natural phenomena in Brazil between 2003 and 2016.

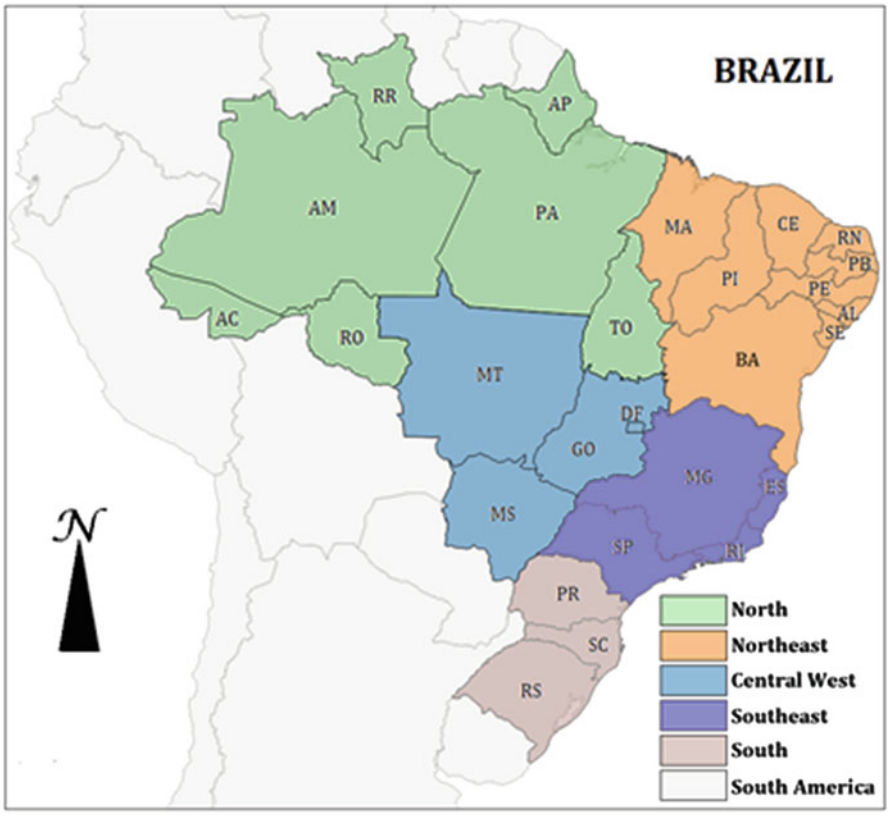


Fig. 1 Regional division of Brazil. North States: Acre (AC), Amapá (AP), Amazonas (AM), Pará (PA), Rondônia (RO), Roraima (RR); Northeast States: Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN), Sergipe (SE); Central West: Distrito Federal (DF), Goiás (GO), Mato Grosso (MT), Mato Grosso do Sul (MS); Southeast States: Espírito Santo (ES), Minas Gerais (MG), Rio de Janeiro (RJ), São Paulo (SP); South: Paraná (PR), Rio Grande do Sul (RS), Santa Catarina (SC)

The S2ID database was established in 2012 by the national government [7] and nowadays, it is the national system officially recognized for supplying federal resources to cities and states affected by disasters [5]. The files accessed in the Historic Series queries, S2ID database, contain the main information on the Federal Emergency Situations and State of Calamity carried out by SEDEC. It is important to realize that the information registered on S2ID documents refers to Brazilian cities; however, as this is a country diagnostic study, those documents were grouped by states and regions.

2.2 Statistical Inference

Whenever it comes to quantitative analysis, many fields of science use Statistics for assertive and data-driven solutions. A most common interest in real-life problems is to properly assess some unknown quantities, as well as any "uncertainty" concerning this assessment. One could, for example, study the probability of raining in a particular forest, or the mean number of wildfires in some dry regions. We collected a sample of the experiment in study and, after observing it, derived conclusions about these unknown quantities. Since we have only one realization (sample) from many other that are theoretically possible, we must also account for this "randomness" sample in measuring our unknown quantity—or *parameter*—of interest. This is a very well-known area of statistics called **inference**.

Statistical inference is all about studying some parameters of interest based on a random sample from an allegedly infinite population. The probability of each sample drawn from all possible population outcomes is very likely to depend on our parameter of interest and can be expressed by analytical devices called *probability distributions*. An unobserved sample point will always have some probability distribution behind it, and it is also called as a *random variable*—let us depict it by an upper-case X. An observed sample point is no longer random (now we know its value) and it is denoted by a lower-case x. A variable can be classified into *discrete*, when its range of possible values is enumerable (generally integers), or *continuous* when it is not (every real number in range).

Let X be a random variable with probability distribution depending on θ, where θ is our parameter of interest. A probability distribution depends on one or more parameters and is well defined by any of the two following formulations: a *probability density function* (or a *probability mass function* when our variable is discrete) represented by $p(x|\theta)$, which yields probabilities for continuous ranges of possible values of X (or even for point values if X is discrete), and a *cumulative distribution function*, the probability of a value lower than or equal to x, denoted by $F(x|\theta)$.

As an example, let X denote the number of disasters in some regions. For one to properly investigate X, it makes sense to know more about how it is probabilistically distributed in that specific region. Since we are interested in a counting variable (thus, discrete), a very common probability distribution to explain such class of variables is the *Poisson distribution*, which depends on a parameter θ—the *mean number* of disasters in that region (which is unknown)—and has a probability mass function given by

$$p(x|\theta) = P[X = x|\theta] = \frac{\theta^x e^{-\theta}}{x!} \tag{1}$$

and cumulative distribution function denoted by

$$F(x|\theta) = P[X \leq x|\theta] = \sum_{k=0}^{x} \frac{\theta^k e^{-\theta}}{k!}. \tag{2}$$

The Poisson probability distribution is defined by only one parameter—the population mean (several distributions are indexed by two or more parameters). A random variable X with Poisson distribution is indexed by $X \sim Poisson(\theta)$, where θ is the population mean.

Now, let $X_1, X_2, \ldots, X_n$ be a *random sample* of size n drawn from X. In our example, $X_1, X_2, \ldots, X_n$ stand for the yearly number of disasters in a given region collected for n years—X_i is the number of disasters at year i ($i = 1, 2, \ldots, n$). If we can properly use the sample data to "guess" the value of θ in a reliable way, then we will be able to make probabilistic relevant statements about how disaster frequency behaves in that region—a common procedure called *estimation*.

The next step is to convert the sample data into relevant information to estimate θ; and two basic assumptions about our sample are essential to carry this on:

1. the data points must be *independent* from each other, or equivalently; the occurrence of some value for X_i can *not* impact the probabilities of possible values of X_j ($i \neq j$). Usually, this is not the case for time-dependent sampling such as our example—temporal dependence matters, after all—but for the moment, let us assume independence for the sake of simplicity (see Section 4 for citations of time-dependent inference approaches);
2. all sample points (which are random variables) came from the same population; therefore, having *the same probability distribution*.

If $X_1, X_2, \ldots, X_n$ are independent random variables, it is possible to derive a joint probability distribution for all the sample data by multiplying their probability density (or mass) functions, in an analogous way of calculating joint probabilities of independent events in probability theory. Then, the *joint probability density (mass) function* is defined by $p(\mathbf{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$ and can be used to derive probabilities of possible samples of size n.

This quantity can also be seen as a function of θ, in which its interpretation changes drastically. Now, this function shows how much a value of θ is "likely" to be the true θ under the sample information. In other words, it shows which values of θ would give our sample a high probability of occurrence (since we managed to collect it anyway). When interpreted in this way, it is denoted by $L(\theta|\mathbf{x})$ and called as a *likelihood function* of θ. If $X \sim Poisson(\theta)$, the likelihood function of θ for a given sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\theta) = \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}. \tag{3}$$

A very common estimation procedure is to find the value of θ that maximizes the likelihood function or, in other words, the one θ that optimizes our sample probability of occurrence. This estimator of θ is known as its *maximum likelihood estimator* and composes an integral part of the classical inference context. One main issue of it, though, is that we cannot include any subjective statement about θ in its estimation—it is solely carried out based on the sampled data. In fact, the θ

parameter is a fixed, immutable point determined by Nature and no prior beliefs about it are allowed to support its estimation in the classic inferential procedure. On the other hand, Bayesian inference [2, 16, 18] presents an entirely different approach for the problem of inference about θ, by interpreting it as some measurable quantity via probability distributions. This context allows us to combine some prior belief about θ into the defined-by-data likelihood to assist the inferential process.

2.3 Bayesian Inference

2.3.1 Basic Concepts

The Bayesian methodology consists in specifying some probability distributions for the observed variables conditioned to one (or more) unknown parameter(s), denoted by θ for simplification purposes and constructing its likelihood function $L(\theta|\mathbf{x})$ based on sample data. It is assumed that θ is also a random variable with some *prior probability function* $p(\theta)$ defined before sample collection. The inference about the parameter is based on the *posterior probability distribution* obtained by the application of the well-known Bayes' theorem [18],

$$\begin{aligned} p(\theta|\mathbf{x}) &= \frac{L(\theta|\mathbf{x})p(\theta)}{\int_{\Theta} L(\theta|\mathbf{x})p(\theta)} \\ &= \frac{L(\theta|\mathbf{x})p(\theta)}{p(\mathbf{x})}, \end{aligned} \tag{4}$$

where Θ denotes the parametric space (the set of all possible values) of θ and $p(\mathbf{x})$ is the normalizing constant of $p(\theta|\mathbf{x})$, (since $p(\theta|\mathbf{x})$ is a probability density and must be integrable to one) called the marginal distribution or *predictive distribution* of $\mathbf{X}$ [14]. Note that $L(\theta|\mathbf{x})$ is actually a function of the sample, which explains why the Bayes' theorem applies. From (4), $p(\theta|\mathbf{x})$ is proportional to the multiplication of the likelihood and the prior function,

$$p(\theta|\mathbf{x}) \propto L(\theta|\mathbf{x})p(\theta). \tag{5}$$

Bayes' theorem is one of the few results of Mathematics that proposes to characterize learning with experience. In Bayesian inference, each problem is unique and θ is a quantity of interest taken with different levels of knowledge that depends on the problem in hand and on who analyzes it. Then, for Bayesian, the probability distribution that captures this variability is based on a prior information and is subjective by nature [18].

One of the most important questions in Bayesian inference concerns prior distributions, which represents the knowledge about an uncertain parameter θ before observing the results of a new experiment. Being aware of which information is

going into the prior distribution and the properties of the resulting posterior are crucial when setting up a prior distribution for a specific problem. An important note is that the prior distribution does not need to carry any information about θ at all, representing in this fashion our ignorance about its behavior. Examples of non-informative priors involve flat distributions for θ (such as Normal with $\sigma_\theta^2 > 10^6$) and functional forms designed to intentionally maximize our ignorance (or *entropy*) about θ for a given dataset.

These priors are included in the class of **objective priors**, in a sense that their impact on the posterior distribution does not depend on individual and subjective beliefs, thus leading to "objective" results. On the other hand, there are situations in which may be desirable to add some prior knowledge onto the elicitation of the prior distribution. Objective informative priors are given when knowledge about θ comes from a quantifiable source, like historical data. When such information, however, comes from an expert observation, we are dealing with subjective information, since it is not quantifiable in any unique and well-defined way and brings to surface a class of prior distributions called **subjective priors**.

One problem in the implementation of Bayesian methodologies is the analytical intractability. The class of **conjugate priors** aims to get over this issue by formulating a prior distribution which has the same functional form of the posterior distribution, when combined with the data information expressed through the likelihood function. It deserves special attention when we want to sequentially update our inference about θ as a new data that becomes available over time.

If $X \sim Poisson(\theta)$, we could consider a $Gamma(\alpha, \beta)$ prior distribution for θ, which depends on parameters α and β and has the following form:

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \tag{6}$$

where $\Gamma(\alpha)$ is the Gamma function applied on α. When combined with the Poisson likelihood given in (3), the posterior probability density of θ is given by

$$\begin{aligned} p(\theta|\mathbf{x}) &\propto L(\theta|\mathbf{x})p(\theta|\alpha, \beta) \\ &= \frac{\theta^{\sum_{i=1}^n x_i}e^{-n\theta}}{\prod_{i=1}^n x_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \\ &\propto \theta^{\sum_{i=1}^n x_i+\alpha-1}e^{-(n+\beta)\theta}, \end{aligned}$$

where constant quantities in θ were omitted (a useful practice when calculating posterior distributions). It turns out that the remaining expression is proportional to a Gamma distribution with parameters $\sum_{i=1}^n x_i + \alpha$ and $n + \beta$. This is an example of a conjugate prior: if $X \sim Poisson(\theta)$ and $\theta \sim Gamma\ (\alpha, \beta)$, then $\theta|\mathbf{X} \sim Gamma\ (\sum_{i=1}^n x_i + \alpha, n + \beta)$. Further details about conjugate priors are given in [16] and [18].

Posterior distributions can be used to provide point estimates or interval estimates (such as the highest posterior density ones [16]) of the parameter of interest, hypothesis testing, or to predict the value of future observations. It provides a unified set of Bayesian solutions to the conventional problems of scientific inference. The choice of the Bayesian point estimates of θ depends on the form of the posterior distribution, as well as the objectives of its use. The most used estimates are the posterior mode, posterior mean, and posterior median:

- Posterior mode: $\hat{\theta}$ such that $\max_{\theta \in \Theta} p(\theta|\mathbf{x})$;
- Posterior mean: $\hat{\theta} = E[\theta|\mathbf{x}]$;
- Posterior median: $\hat{\theta}$ such that $P[\theta \geq \hat{\theta}] \geq 1/2$ and $P[\theta \leq \hat{\theta}] \geq 1/2$.

Bayesian approach can be used to sequentially update the information about the parameter as new data become available. Suppose we formulate a prior for the parameter θ and observe a random sample $\mathbf{x}_1$. Then the posterior is

$$p(\theta|\mathbf{x}_1) \propto L(\theta|\mathbf{x}_1)p(\theta).$$

If a new sample $\mathbf{x}_2$ is observed, we can use the previous posterior as the new prior and derive a new posterior,

$$p(\theta|\mathbf{x}_2) \propto L(\theta|\mathbf{x}_2)p(\theta|\mathbf{x}_1).$$

This "sequential updating" process can continue indefinitely in the Bayesian setup.

The Bayesian operation is often difficult to execute and requires the use of numerical methods and the approximate Monte Carlo simulation method via Markov Chains. All results presented from this point on were produced with the statistical software R, version 3.3.3 [10].

2.3.2 Bayesian Inference for Rate of Disasters Occurrence

Then, let X be a random variable related to the number of disasters occurring in a given period. We have that X given θ has a Poisson distribution with parameter θ, the disaster occurrence rate in the selected period. Consider $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ an observed random sample of the random variable X. With likelihood function given by (3) and prior function for θ given by (6), that is, a $Gamma(\alpha, \beta)$ conjugate prior, the posterior distribution of θ is $\theta|\mathbf{X} \sim Gamma(\sum_{i=1}^{n} x_i + \alpha, n + \beta)$, with posterior mean given by

$$E[\theta|X] = \frac{\sum_{i=1}^{n} x_i + \alpha}{n + \beta}. \tag{7}$$

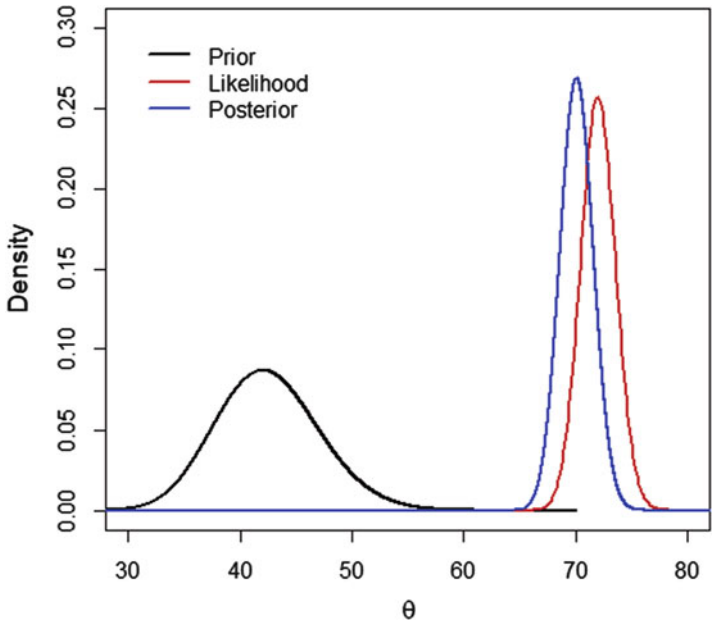


Fig. 2 Illustration of prior distribution, likelihood function, and posterior distribution, considering simulated data

and posterior variance

$$Var[\theta|X] = \frac{\sum_{i=1}^{n} x_i + \alpha}{(n+\beta)^2}. \tag{8}$$

Figure 2 represents the described Bayesian framework for some simulated data. It shows the comparison of the prior distribution with the posterior distribution to assess to what extent the experimental information can alter the initial belief.

2.3.3 Bayesian Prediction

Professionals from a multitude of fields often need to make some realistic statements about the likely outcome of a future "experiment of interest" based on the distribution of previously conducted related experiments. In the Bayesian framework, given an observed quantity X related to an unobserved parameter $\boldsymbol{\theta}$ through a posterior distribution $p_1(\theta|\mathbf{x})$, our interest relies on making inference about another random quantity Y related to X and $\boldsymbol{\theta}$ through $p_2(Y|\theta, X)$.

After observing a random sample $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, why not just predict some new observation x_{n+1} by plugging into the distribution of X some posterior point estimate $\hat{\theta}$? It turns out that we must account for the uncertainty about θ in

our prediction, otherwise, new observations would have underestimated variability. Since this uncertainty about θ is reflected on its posterior, averaging the distribution of x_{n+1} over it would yield

$$p(x_{n+1}|\mathbf{x}) = \int p(x_{n+1}|\theta)p(\theta|\mathbf{x})d\theta. \tag{9}$$

If X follows a Poisson distribution with parameter θ having a Gamma(α, β) conjugate prior, it was shown that $\theta|\mathbf{X} \sim Gamma\left(\sum_{i=1}^{n} x_i + \alpha; n + \beta\right)$. If we are interested in making inference about some unobserved value x_{n+1} from the same population of $\{x_1, x_2, \ldots, x_n\}$ and θ-conditionally independent from them, the predictive distribution is given by

$$\begin{aligned}
p(x_{n+1}|\mathbf{x}) &= \int_{-\infty}^{\infty} p(x_{n+1}|\theta)p(\theta|\mathbf{x})d\theta \\
&= \int_{-\infty}^{\infty} \frac{\theta^{x_{n+1}}e^{-\theta}}{x_{n+1}!} \frac{(n+\beta)^{\sum_{i=1}^{n} x_i+\alpha}}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)} \theta^{\sum_{i=1}^{n} x_i+\alpha-1} e^{-(n+\beta)\theta} d\theta \\
&= \frac{(n+\beta)^{\sum_{i=1}^{n} x_i+\alpha}}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right) x_{n+1}!} \int_{-\infty}^{\infty} \theta^{\sum_{i=1}^{n+1} x_i+\alpha-1} e^{-(n+\beta+1)\theta} d\theta \\
&= \frac{(n+\beta)^{\sum_{i=1}^{n} x_i+\alpha}}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right) x_{n+1}!} \frac{\Gamma\left(\sum_{i=1}^{n+1} x_i + \alpha\right)}{(n+\beta+1)^{\sum_{i=1}^{n+1} x_i+\alpha}},
\end{aligned}$$

since $\theta^{\sum_{i=1}^{n+1} x_i+\alpha-1} e^{-(n+\beta+1)\theta}$ is proportional to $Gamma\left(\sum_{i=1}^{n+1} x_i + \alpha; n+\beta+1\right)$.

Note that (9) is completely general, in that p can be any density or even an extremely complex model, and θ can represent one, two, or even thousands of unknown parameters. In practice, this can be a difficult calculation. So, most of the time, we simulate values from the predictive density, rather than getting the exact analytical solution:

1. Draw a sample from the posterior density of θ, $p(\theta|x)$.
2. Then plug it into $p(x_{n+1}|\theta)$ and draw an x_{n+1} observation from it.
3. Repeat steps 1 and 2 many times (at least 1000). Since a different value of θ is used in each time, we automatically restore the uncertainty missing when we plug in just a single value of θ.

2.3.4 Mixture of Distributions

An alternative way to evaluate the distribution of the number of disasters in each Brazilian state can be made possible after the idea of *mixture of distributions*, which is described by [17]. If $p_0(x), p_1(x), \ldots, p_k(x)$ is a sequence of either all discrete probability mass functions or all probability density functions, and $\omega_0, \omega_1, \ldots, \omega_k$

are a sequence of weights satisfying $\omega_i > 0$ and $\sum_{i=1}^{k} \omega_i = 1$, then $\sum_{i=1}^{k} \omega_i p_i(x)$ is also a probability mass/density function, which is called a *mixture of distributions*.

As an example, consider just one state of Brazil for some period of J years, and that we want to study its mean number of disasters θ, $\theta \in \Theta$. For the j-th year ($j = 1, \ldots, J$), let X_j be a random variable representing the number of occurred disasters at year j, with probability density function given by $p(x_j|\theta)$. The likelihood function for the j-th year is $L_j(\theta|\boldsymbol{x}_j) = \prod_{i=1}^{n_j} p(x_{ij}|\theta) = p(x_j|\theta)$, since we have only one observation for each year in this case ($n_j = 1$).

Under the Bayesian context, we derive the prior distribution of θ, $p(\theta)$, and the posterior distribution of θ for each year given their respective data, $p_j(\theta|x_j)$. Therefore, for the chosen state, the posterior distribution of θ is given by the mixture of J posterior distributions of θ. Then, the mixture of posterior distributions is given by

$$p_M(\theta|\boldsymbol{x}) = \sum_{j=1}^{J} \omega_j p_j(\theta|\boldsymbol{x}_j) \tag{10}$$

where ω_j is the weight of the j-th year.

3 Results

The Brazilian country has 5570 municipalities, from which 75% of them (4168) have declared an emergency situation at least once during the study period; for the remaining 25%, data is not available. Among the first 75%, some cities stand out for the number of times where an emergency situation was declared in Ceara State (CE): Irauçuba (28 occurrences), Caridade (27), Tauá, (27), and Pedra Branca (26); in Pernambuco State (PE): Lagoa Grande (26), and Santa Cruz (26). The most affected states of the country in the same period were Paraíba (PB) and Rio Grande do Sul (RS), with 3286 and 3215 emergency situations reported, respectively.

The map of Fig. 3 shows the spatial distribution of disasters in Brazil between 2003 and 2016. The map revels that six states (RR, AP, AC, RO, TO, GO) and Distrito Federal (DF), each one, had less than 200 emergency situations declared, a very low number if compared to the others. PB and RS are the most affected states, registering more than 3000 occurrences, and six other states (BA, CE, MG, SC, PE, PI) had between 2000 and 3000 occurrences in the same period. There are many factors that can explain these numbers, like frequency, magnitude, or intensity of hazardous events or the exposure level of municipalities to them, but this type of analysis is out of scope from this study.

It is advisable to approach risk and disaster research by classifying the territory in homogeneous regions of environmental determinants such as topography, climate, geological substrate, soils, hydrology, or vegetation because of their correlation with natural threats. In spite of that, another useful approach is by identifying the impacts

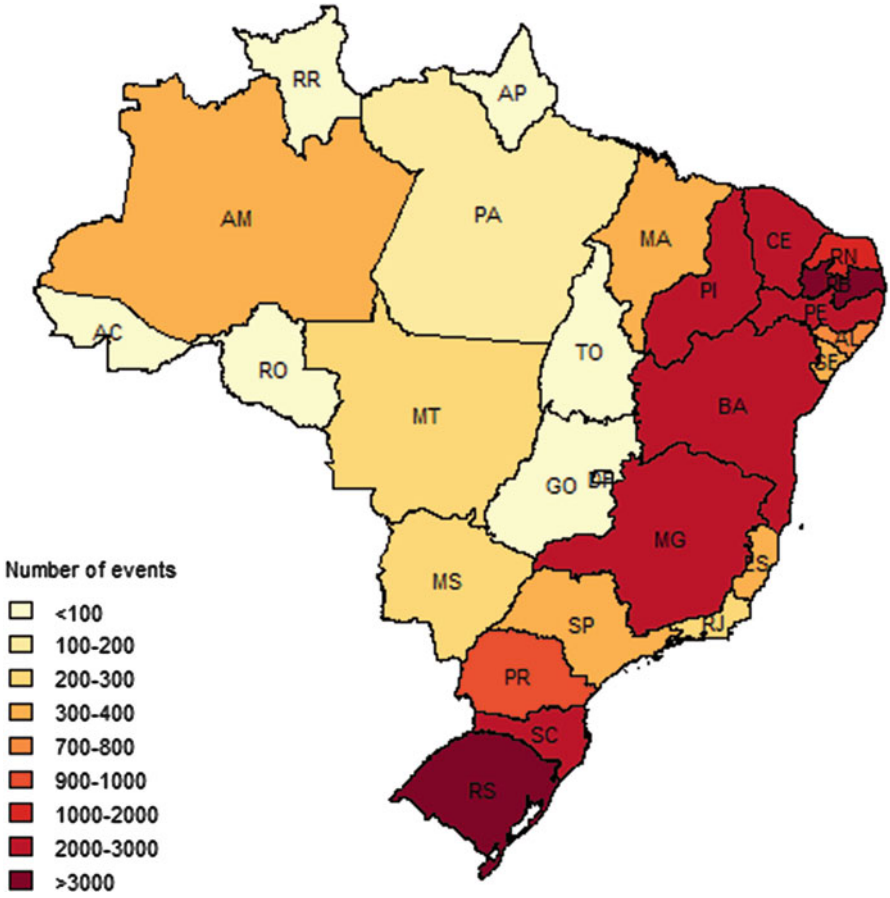


Fig. 3 Spatial distribution of disasters in Brazil. It refers to S2ID data from 2003 to 2016

and the level of hazards damage on the population; this information is related to the spatial distribution of natural phenomena that can cause damage, but also can reflect social asymmetries that are predominant in each social-physical region. For example, Fig. 3 shows that the greater frequencies of disaster occurrences are in the set of the most densely populated states, instead of those exposed to multi-hazards. However, this study does not consider if there is or not a civil defense in the municipalities responsible for generating S2ID documents.

Figure 4 shows a predominance of the flood event in much of the territory in the Northern Region: (AC, AM, AP, and PA states); dryness is spread to all Brazilian Regions: North (RO, RR, TO), Northeast (BA, CE, MA, PB, PE, PI), Central West (MS), Southeast (MG and RJ), and South (PR and RS); drought stands out in AL, SE, and RN states of Northeast; higher frequency of flash floods predominates in GO (Central West-CO), SP and ES (Southeast), and SC (South); MT and CO stand

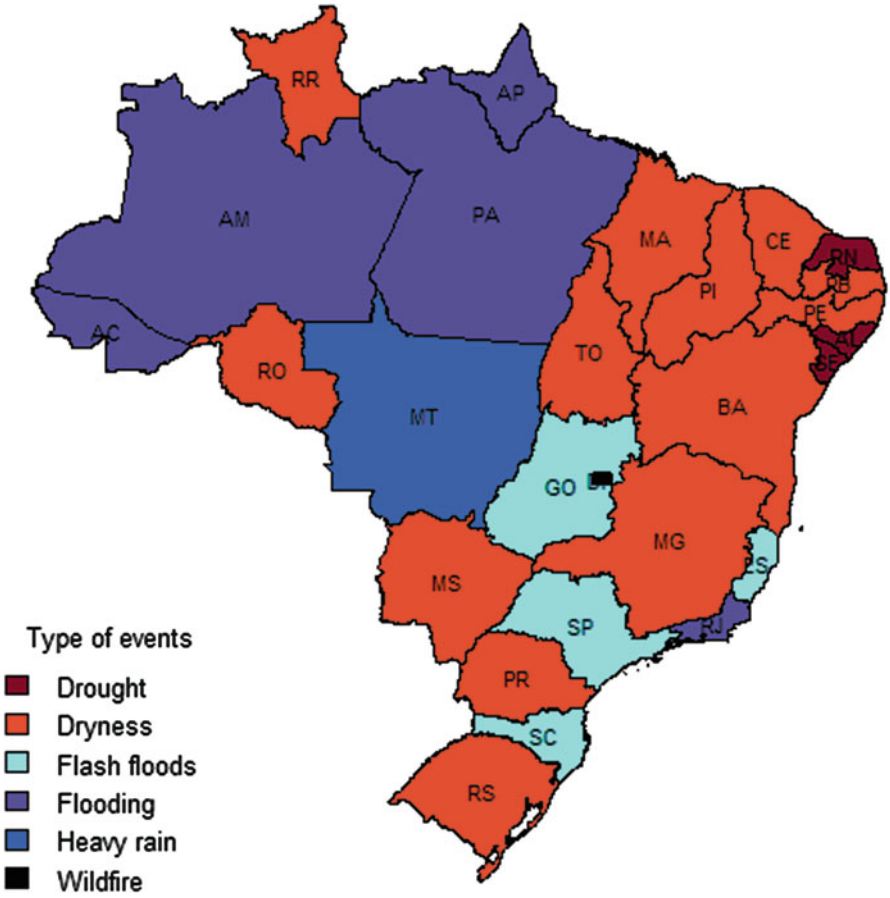


Fig. 4 Spatial distribution of disaster occurrences by cause typology in Brazil. It refers to S2ID data from 2003 to 2016

out due to intense frequency of rainfalls. A small portion of Federal District (DF) records is related to wildfires.

Although the spatial distribution of the number of occurrences is not related with the regional division of Brazil (Fig. 1), the typology has some correlation: the Northern region is more susceptible to floods; Central West region is more susceptible to intense rainfall and flash floods; Northeast is more susceptible to dryness and drought; Southeast and South regions cope with multiple hazards, standing out among those flooding, dryness, and flash floods.

Dryness and drought were the two more frequent types of threat that generate impacts in the country (2003 to 2012), encompassing 56 and 14% (respectively) over all records used for the study. On the other hand, 2013 was the year with the highest number of disaster occurrences reported on Brazilian S2ID database (Fig. 5),

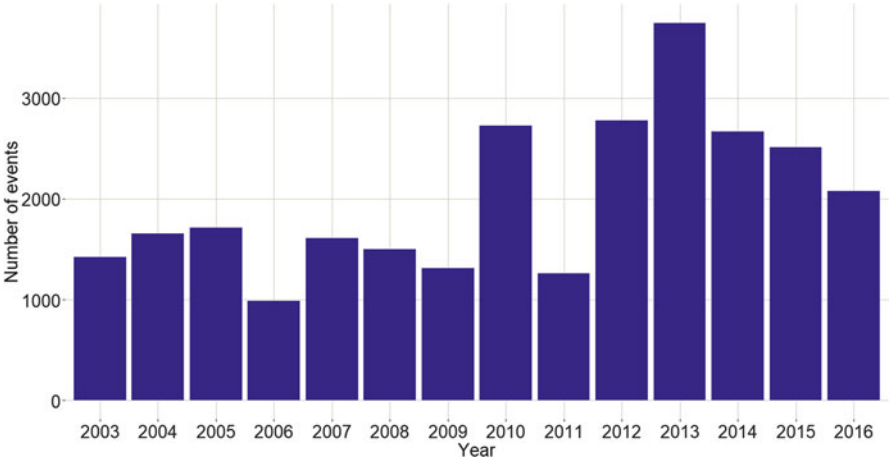


Fig. 5 Temporal distribution of the number of disasters per year, Brazil, 2003–2016

Table 1 Series of the number of disasters (dryness and droughts) for the State of Alagoas, Brazil (2003–2016)

State	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
AL	123	0	26	14	33	65	51	34	0	36	97	43	74	81

3743 in total; among those, the most frequent event was dryness (64%), followed by drought (23%). Those results are in agreement with the previous Brazilian studies: dryness and droughts, which are directly related to the reduction of rainfall and to the water deficit, respectively, were the most frequent types of hazard in Brazil period 1991–2012, and are considered the major national problems related to natural threats in this country, like documented on the Atlas Nacional de Desastres Naturais.

To perform our analysis in the Bayesian framework, we only considered the dryness or drought data of Alagoas, Northeast Region, in the study period (Table 1), as the procedure is analogous to the other states. Figure 6 illustrates the sequential updating procedure to derive Bayesian inference and shows how the posterior distribution evolves across the years (it "walks" over the θ axis) as more data is available to update it. In this context, the prior for θ in 2003 is a conjugate, non-informative one (Gamma with $\alpha = 0.5$ and $\beta = 0.0001$, resulting in a large variance) and the data from 2003 (Poisson with parameter θ) combines to form the posterior distribution of θ in 2003, also a Gamma distribution. For the subsequent years, the posterior from the previous year is used as the prior distribution and combined with the current year data to form the updated posterior distribution of θ (note that these posterior distributions are always Gamma).

It is noticeable how little the posterior evolves at the last years; in fact, it has already absorbed so much past information, especially the 2004–2012 period, that makes new data struggle to change our belief about θ, regardless of how impactful they are. The posterior distribution of θ in 2015 tells us that having 60 or more

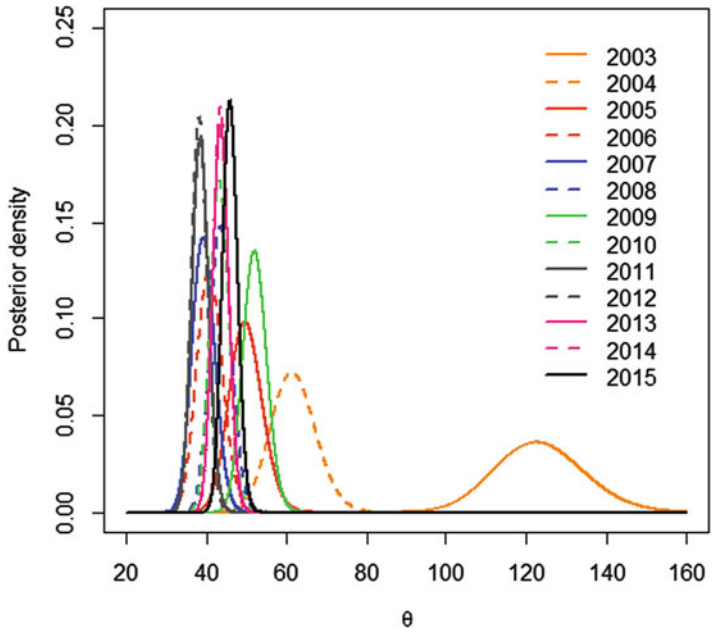


Fig. 6 Updating procedure per year, Alagoas, Brazil, 2003–2015

disasters in Alagoas is highly unlikely (almost zero probability), even though 97 disasters were reported in 2013, just 2 years before.

For comparison purposes, Fig. 7 shows the mixed posterior distribution of θ considering the period from 2003 to 2015 (2016 was left out to compare predictions from both approaches). Instead of sequential updating, here all the year samples are being observed at the same time. The weight of each sample is proportional to how fresh is the data: newer data has more weight (as its information is more up-to-date), which declines proportionally for older data. Since we have a 13-year time series for analysis (2003–2015), the most recent year of 2015 has a weight of $13/(1 + 2 + \ldots + 13) = 13/91$. The year of 2014 has a weight of $12/91$ and so on until 2003, with a weight of $1/91$. In this way, information loses value as it grows too old and no longer reflects the current reality.

The likelihood of each year is combined with a non-informative prior distribution that is the same for all years (again, Gamma(0.5, 0.0001)). Then, the calculated posteriors for each year are combined with the described weights, leading to a mixed posterior distribution that considers all the different information gathered at the period of 2003–2015, giving more weight to recent events. As 2013 showed 97 disasters, this is why low values (<50) are still predominant, but higher values (>80) are somehow feasible as well. Table 2 compares both posterior distributions and leads to the same conclusion: even though location measures are similar for both, the 95% HPD credibility interval includes a much wider range of values.

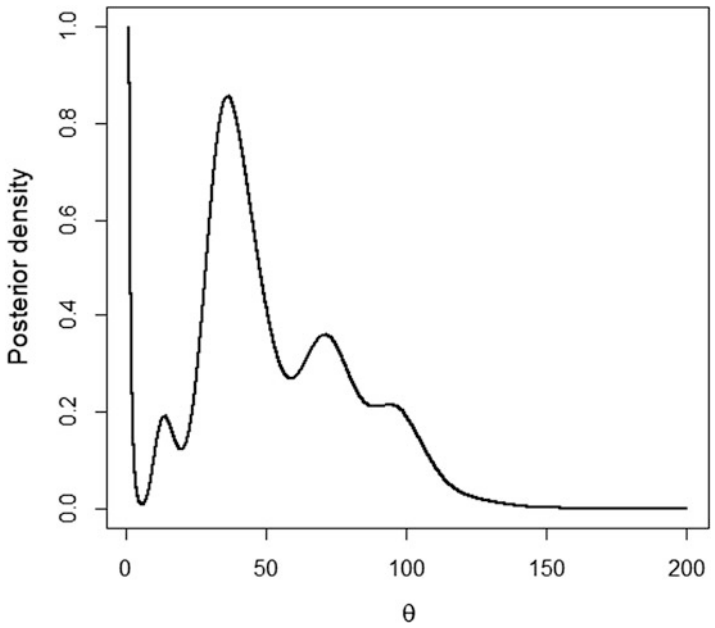


Fig. 7 Mixture of posterior distributions of θ, Alagoas, Brazil, 2003–2015

Table 2 Summary of posterior distributions of θ, Alagoas, Brazil, 2003–2015

Posterior	Median	Mean	95% HPD Interval
Sequential updating	45.9	45.9	[42.2, 49.6]
Mixture	43.0	48.1	[4.7 10^{-8}; 101.7]

After calculating the posterior for the year 2015 in both ways (sequential updating and mixing), Fig. 8 shows the predictive distributions attempting to predict the number of disasters in 2016, in the state of Alagoas. It is clear to see that the predictive distribution obtained by the sequentially updated posterior fails to capture the true number of disasters in that state in 2016 (81), as it is located in a region of lower counts and lacks the needed variability to, at least, consider the true 2016 number as possible—a very informative distribution indeed as it has low variance, but misses its target entirely ($P\left[X_{2016} \geq 81 | \theta\right] = 0$). The predictive given by the mixed posterior, in contrast, is much less informative as it has greater variance, but that is exactly a trait of Alagoas series: high variability. The "mixed" predictive distribution carried this trait on and, as a result, was able to capture the true value of 2016 with some degree of certainty ($P\left[X_{2016} \geq 81 | \theta\right] = 0.17$).

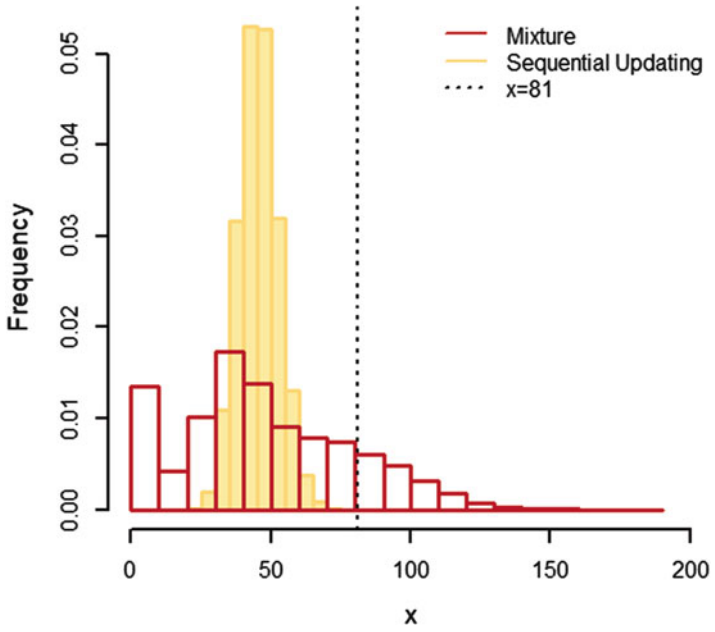


Fig. 8 Predictive distribution of x (number of disasters—dryness and droughts) for 2016, Alagoas, Brazil

4 Discussion and Conclusion

This work aimed to summarize some spatio-temporal data of disasters in Brazil while proposing a comprehensive, didactic methodology to study their behavior over time. Among the main results that we observed, in agreement with previous studies [9], it is important to highlight dryness and drought among the natural hazards that cause the most disasters in Brazil. These phenomena corresponded to 51.3% of the documents reviewed in [9] and to 70% of the documents analyzed in this book chapter.

According to [9], the number of disasters in Brazil increases over time, but these numbers alone are not strong enough to assure that disasters are really happening more often. The increase in the number of occurrences recorded in S2ID can be related to several variables; it is not necessarily due to the increase of disasters or intensification of climate change effects, but it may also be related to the new risk management policies established by the Brazilian government as a result of the national system restructuring [7], in which competent official institutions are required to record the damage to the municipality for declaration of emergency or state of public calamity. It is possible to spot a rise in the numbers for the last 7 years of the study period (2003–2016) compared to the earlier ones, but as much as it can represent an actual rise, an improvement of data collection and upkeep is a feasible hypothesis as well.

We presented a Bayesian approach for inference about the number of disasters in Brazil, picking the 2003–2015 period to construct our posterior model and using it to make predictions for 2016. The mixture of distributions is a very simple and intuitive procedure, as it managed to derive some fairly assertive conclusions regarding the prediction year when the traditional sequential update could not. The resulting mixed posterior distribution of θ revealed itself as a much better summary of θ—being non-unimodal and asymmetrical, it is safe to say that it succeeded to capture the variability trait from the series of Alagoas.

There are more sophisticated, suitable ways to extract information from the time series. Some elementary readings in time series model construction are given by [3] and [20], while [1] presents some ways to deal with time series modeling in a Bayesian framework; and a very broad paperwork on Bayesian dynamic models can be found in [11]. To sum it up, some disaster related works include [19], who proposes a generalized Pareto distribution to model rare event occurrences for small samples, and [13] developed a social vulnerability index adaptation for Brazilian reality.

To present a didactic way of data analysis, we limited our study only to the number of disasters, region, and year. Weather and social/geographic exposure-related features would be very interesting to enrich our study, though we are still in the development of a more appropriate national database for making it possible. This is a very important matter, since it would allow us to understand, among other things, some causes of disaster occurrences and their impact on the population. Such a source of information would enable statistical studies to plan and foresight against disasters, followed by risk management to lower population exposure and would improve their response time in face of an occurrence.

Acknowledgements The authors would like to thank to Federal University of São Paulo—UNIFESP, to CNPq—Brazil, to National Early Warning and Monitoring Center of Natural Disasters (CEMADEN) and to the Thematic Committee Mathematics & Disasters of the Brazilian Society of Applied (SBMAC), by its research support.

References

1. Barber, D., Cemgil, A.T., Chiappa, S.: Bayesian Time Series Models. Cambridge University Press, Cambridge Books Online (2011)
2. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. Springer Series in Statistics. Springer, Berlin (1985)
3. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting And Control. Holden-Day Series in Time Series Analysis and Digital Processing. Holden-Day, San Francisco (1976)
4. Brasil: Sistema Integrado de Informações sobre Desastres, S2ID. Available via DIALOG. https://s2id.mi.gov.br/. Accessed 15 May 2017
5. Brasil: Ministério da Integração Nacional. Gabinete do Ministro, Portaria N. 215, de 4 de Abril de 2017. Diário Oficial da União, n. 67. Available via DIALOG. http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=06/04/2017&jornal=1&pagina=40&totalArquivos=72. Accessed 17 Apr 2018

6. Brasil: Ministério da Ciência, Tecnologia e Inovação. Modelagem climática e vulnerabilidades Setoriais á mudança do clima no Brasil. Secretaria de Políticas e Programas de Pesquisa e Desenvolvimento. Coordenação-Geral de Mudanças Globais de Clima. Ministério da Ciência, Tecnologia e Inovação, Brasília (2016). Available via DIALOG. https://www.researchgate.net/profile/Wanderson_Silva7/publication/305084250_Modelagem_Climatica_e_Vulnerabilidades_Setoriais_a_Mudanca_do_Clima_no_Brasil/links/57818a7a08ae5f367d393b12/Modelagem-Climatica-e-Vulnerabilidades-Setoriais-a-Mudanca-do-Clima-no-Brasil.pdf. Accessed 17 Apr 2018
7. Brasil: Ministério da Integração Nacional. Instrução Normativa N. 01, de 24 de Agosto de 2012. Available via DIALOG. http://www.mi.gov.br/c/document_library/get_file?uuid=822a4d42-970b-4e80-93f8-daee395a52d1&groupId=301094. Accessed 17 Apr 2018
8. Brasil: Ministério da Integração Nacional. Instrução Normativa N. 01, de 24 de Agosto de 2012. Anexo 1: Classificação e Codificação Brasileira de Desastres (Cobrade). Available via DIALOG. http://www.integracao.gov.br/documents/3958478/0/Anexo+V+-+Cobrade_com+simbologia.pdf/d7d8bb0b-07f3-4572-a6ca-738daa95feb0. Accessed 17 Apr 2018
9. Centro Universitário de Estudos e Pesquisas Sobre Desastres, CEPED. Atlas Brasileiro de Desastres Naturais 1991 a 2012. 2. ed. rev. ampl. Florianópolis: CEPED UFSC (2013). Available via DIALOG. https://s2id.mi.gov.br/paginas/atlas/. Accessed 17 Apr 2018
10. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2016). Available via DIALOG. https://www.R-project.org/. Accessed 17 Apr 2018
11. Damien, P., Dellaportas, P., Polson, N.G., Stephens, D.A.: Bayesian Theory and Applications. Oxford University Press, Oxford (2015)
12. Debortoli, N.S., Camarinha, P.I.M., Rodrigues, R.R., Marengo, J.A.: Índice de vulnerabilidade aos desastres naturais no Brasil, no contexto de mudanças climáticas. Cap. 7 em: MCTI Modelagem climática e vulnerabilidades setoriais á mudança do clima no Brasil. MCTI, Brasília, pp. 321–385 (2016)
13. de Loyola Hummell, B.M., Cutter, S.L., Emrich, C.T.: Social vulnerability to natural hazards in Brazil. Int. J. Disaster Risk Sci. **7**, 111–122 (2016)
14. Ibrahim, J.G., Chen M.-H., Sinha, D.: Bayesian Survival Analysis. Springer Series in Statistics, 2nd edn. Springer, Berlin (2005)
15. Instituto Brasileiro de Geografia e Estatística – IBGE. Sinopse do Censo Demográfico 2010. Ministério do Planejamento, Orçamento e Gestão. Rio de Janeiro (2011). Available via DIALOG. https://biblioteca.ibge.gov.br/visualizacao/livros/liv49230.pdf. Accessed 17 Apr 2018
16. Migon, H.S., Gamerman, D., Louzada, F.: Statistical Inference: An Integrated Approach, 2nd edn. Taylor & Francis, Boca Raton (2014)
17. Mood, A.M., Graybill, F.A., Boes, D.C.: III.4.3 Contagious distributions and truncated distributions. In: Introduction to the Theory of Statistics, 3rd edn. McGraw-Hill, New York (1974)
18. Paulino, C.D.M., Turkman, M.A.A., Murteira, B.: Estatística Bayesiana, 1st edn. Fundação Calouste Gulbenkian, Lisboa (2003)
19. Pisarenko, V.F., Rodkin, M.V., Rukavishnikova, T.A.: Probability estimation of rare extreme events in the case of small samples: technique and examples of analysis of earthquake catalogs. Izv. Phys. Solid Earth **53**, 805–818 (2017)
20. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications: With R Examples. Springer Texts in Statistics. Springer, Berlin (2017)
21. United Nations Office for Disaster Risk Reduction, UNISDR: Sendai Framework for Disaster Risk Reduction 2015–2030 (2015). Available via DIALOG. http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf. Accessed 15 July 2017

22. United Nations, UN: Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. Seventieth session, October 2015. Available via DIALOG. http://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1E.pdf. Accessed Cited 12 Apr 2018
23. United Nations, UN.: Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators. Economic and Social Council. E/CN.3/2016/2/Rev.1, Statistical Commission, Forty-seventh session. March 2016. Available via DIALOG. https://unstats.un.org/unsd/statcom/49th-session/documents/2018-2-SDG-IAEG-E.pdf. Accessed 12 Apr 2018
24. United Nations, UN.: Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction. General Assembly. A/71/644. Seventy session. December 2016. Agenda item 19 (c) sustainable development: disaster risk reduction. Available via DIALOG. https://www.preventionweb.net/files/50683_oiewgreportenglish.pdf. Accessed 17 Apr 2018

About Interfaces Between Machine Learning, Complex Networks, Survivability Analysis, and Disaster Risk Reduction

Leonardo Bacelar Lima Santos, Luciana R. Londe, Tiago José de Carvalho, Daniel S. Menasché, and Didier A. Vega-Oliveros

1 Introduction

1.1 Background and Motivation

Modern society heavily relies on critical infrastructure systems. They are essential to support public interests and to keep health, security, and economic/social wellness. Their disruption or depletion would cause expressive consequences to society. Local and instantaneous failures or impairments could result in long-term effects on people's lives and severe damages to buildings and goods—characterizing human, social, and economic impacts [6], over various time and social scales. As these infrastructure systems are exposed to several kinds of risks, most of the preventive/protection plans are related to the recovery from a disaster and continuity of business.

L. Bacelar Lima Santos (✉)
National Centre for Monitoring and Early Warnings of Natural Disasters (CEMADEN), São José dos Campos, São Paulo, Brazil
e-mail: leonardo.santos@cemaden.gov.br

L. R Londe
Cemaden, São José dos Campos, SP, Brazil
e-mail: luciana.londe@cemaden.gov.br

T. José de Carvalho
Department of Informatics, Federal Institute of São Paulo (IFSP), Campinas, São Paulo, Brazil
e-mail: tiagojc@ifsp.edu.br

D. S. Menasché
UFRJ, Rio de Janeiro, RJ, Brazil
e-mail: sadoc@dcc.ufrj.br

D. A. Vega-Oliveros
DCM-FFCLRP-USP, Ribeirão Preto, SP, Brazil
e-mail: davo@icmc.usp.br

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_10

1.2 Disaster Risk Reduction (DRR) Management

To avoid losses, to increase profits, and to increase efficiency, planning and decision making should work well. To be able to manage, a deep and specific knowledge is necessary. For disaster risk management, we need to know the possible hazards, how vulnerable the communities are, and what has been done to improve resilience.

The UNISDR [75] defines disaster risk management as the application of disaster risk reduction policies and strategies to prevent new disaster risk, reduce existing disaster risk, and manage residual risk, contributing to the strengthening of resilience and reduction of disaster losses.

The disaster risk management can be performed by local and global governments, non-governmental organizations, the private sector, and communities. Also, the management can be applied both to either extensive (low intensity and high frequency) and intensive (high intensity and low frequency) disasters.

1.3 Defining Risk

The understanding of risk is a key component for the management of critical infrastructure systems. The word *risk* is widely discussed among authors who study disasters. Historically, the words *risk*, *fortune* (chance), and *hazards* appeared in related contexts. One of the first occurrences of such words appeared in manuscripts on sailors that went to the sea, with no warranties that they would either arrive at their destination or be able to go back to their origin. There was a need to characterize the *risk* of going to the sea [61, 79]. Later, the word *risk* was used for goods insurance and then risk and insurance became associated words.

The International Strategy for Disaster Reduction—ISDR, an UN office with focus on the disaster risk reduction, defines risk as "the combination of the probability of an event and its negative consequences" [75]. This concept is captured through the following equation:

$$R = HV \tag{1}$$

where R is disaster risk, H is hazard, and V is vulnerability.

Equation (1) explicitly indicates the role of both vulnerability (V) and hazard (H) on risk. This is in contrast to hazard-focused research and policy, which address hazards but not vulnerabilities [20]. Disaster risk is a function of the magnitude, potential occurrence, frequency, speed of onset, and spatial extent of a potentially harmful natural event or process—the "hazard"—and also a function of people's susceptibility to loss, injury, or death [20]. Taken together, susceptibility to harm and the process that creates and maintains that susceptibility to harm is referred to as "vulnerability." Vulnerability, in turn, may be counteracted either by individual and local capacity for protective action (C) or by protective actions carried out by

Table 1 Table of notation

Variable	Description
R	Risk
H	Hazard (impact, i.e., negative consequences of an event)
V	Vulnerability (exposure, i.e., probability of an event)
C	Protective capacity
M	Mitigation capacity

larger entities such as government (M, which stands for mitigation and prevention) [20]. To make such factors explicit, Eq. (1) is rewritten as follows [83]:

$$R = H((V/C) - M) \tag{2}$$

where R is disaster risk, V stands for vulnerability, C represents capacity for personal protection, and M denotes larger-scale risk mitigation by preventive action and social protection (see Table 1). It is important to note that mitigation is a widely emergency management community term. In transportation, a different sector, the mitigation word is actually synonym to adaptation.

1.4 Managing Risk

Next, we focus on strategies for risk management. One technical normative recommendation for risk management [59] defines risk management as "coordinated activities to direct and control an organization concerning risks." This is generic for different organizations and accounts for different consequences, both positive and negative. This normative recommendation states that the risk management must be specific (each risk scenario would imply different management decisions), including human and cultural factors, being transparent and inclusive, and being dynamic and able to react to changes.

The United Nations International Strategy for Disaster Reduction, UNISDR [75], defines disaster risk management as:

> The systematic process of using administrative directives, organizations, and operational skills and capacities to implement strategies, policies and improved coping capacities in order to lessen the adverse impacts of hazards and the possibility of disaster.

The people, society, or physical infrastructure at risk may be exposed to a same hazard under different ways. The hazard is not exclusively related to economic

conditions, but also to cultural, social, and environmental conditions, characteristics of these elements. Also, for each exposed element, there are approaches that better fit the study and management for risk reduction. When infrastructure systems are specifically the object of study, mathematical modeling and optimization techniques are useful to understand the problem and make projections.

1.5 Mathematical Risk Modeling

There are several approaches for mathematical analysis on critical infrastructures. According to Jenelius [42], there is a special category of studies that use mathematical modeling and optimization techniques to identify best responses or worst-case scenarios in critical infrastructures. Matisziw and Murray [51], for example, use an integer programming formulation to identify the most severe disruptions in a transport network of Ohio, USA, and Bell et al. [11] integrate a traffic model in a game-theoretic framework to determine routing strategies in London, UK, under the risk of antagonistic attacks [42]. In Brazil, Santos et al. [64, 65] presented methodologies based on historical datasets to quantify spatial and temporal potential impacts of disasters on urban mobility and transportation.

1.6 Chapter Goals

One of the goals of this chapter is to present mathematical models of risk in light of the impact of countermeasures on risk scores. Note that the definitions of risk presented in Eqs. (1) and (2) are typically coupled with a model that has either predictive or explanatory power. Without such models, the risk score may have limited applicability [28]. In fact, Eqs. (1) and (2), alone, do not capture the fundamental relationships between risks and opportunities, and do not tell the probability of a disaster given the implementation of specific countermeasures. Different models have been proposed in the literature to capture the impact of decisions on risk scores. Logistic [72] or multiple [22] regression, Bayesian networks [28], and Markovian phased-recovery models [7] are examples of such models.

Several concepts originally from areas such as Computer Science, Physics and Mathematics, especially Machine Learning, Complex Systems, and Dynamic Systems have been incorporated into the study of disaster risk reduction. *Another goal of this chapter is to identify interdisciplinary trends and opportunities for collaboration on risk management among different communities.*

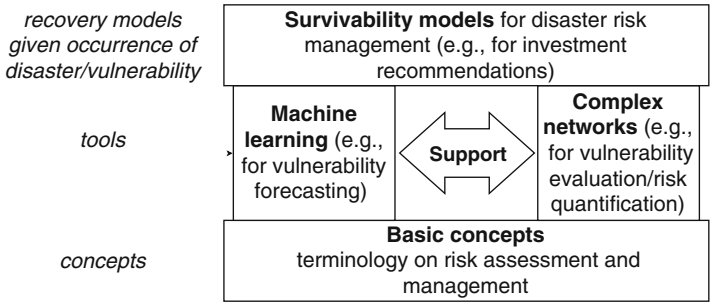


Fig. 1 Summarized structure of this chapter

1.7 Chapter Organization

The remainder of this chapter is organized as follows:

- in Sect. 2 we introduce basic terminology, and discuss the different interpretations of terms exploring the international literature and the definitions adopted by the Brazilian Civil Defense authority;
- in Sect. 3 we present basic machine learning concepts, and an example of a case study involving risk analysis;
- in Sect. 4 we discuss a specific type of vulnerabilities, namely topological vulnerabilities, which are key for studies on critical infrastructures;
- finally, in Sect. 5 we introduce survivability models as a way to cope with some of the challenges related to the modeling of risk.

The structure of this chapter is summarized in Fig. 1. The foundations of the chapter are laid in Sect. 2, where basic concepts are introduced. The pillars built on top of such foundations are tools such as machine learning and complex networks, used for vulnerability forecasting and risk quantification and presented in Sects. 3 and 4, respectively. Finally, survivability models that can leverage such tools are presented in Sect. 5.

2 Terminology and Background on Disasters Risk Reduction and Management—and Some Mathematical Interpretations

When critical infrastructures are affected by either natural or technological hazards, disasters may take place, taking municipalities out of their normal functioning status. There are many words in common for the disaster's terminology and the complex systems jargoons. It is common in science to borrow terms that make sense from one area to another. Some interactions between different knowledge areas have

been done with success. As an example, some concepts in this chapter, such as *adaptation*, *risk*, and *vulnerability* date back to Darwin's theory of evolution [23, 24, 29, 34, 41, 67].

The widespread word brought from physics to disaster's risk management is "resilience." Currently there are campaigns and prizes for "resilient cities" all over the world, aiming to promote preparedness actions and planning. In this case the concept changed not only when it was adapted to the disaster's field of knowledge, but also inside this sector.

The list of terms below brings together both the terminology adopted by the Brazilian Civil Defense authority, which is largely used in Brazil, and the definitions from the United Nations International Strategy for Disaster Reduction (UNISDR [75]).

- Emergency
  - Brazilian Civil Defense:
    1. Critical situation; dangerous event; accident
    2. Case of urgency.
  - UNISDR: *Emergency management:* The organization and management of resources and responsibilities for addressing all aspects of emergencies, in particular preparedness, response and initial recovery steps. Comment: A crisis or emergency is a threatening condition that requires urgent action. Effective emergency action can avoid the escalation of an event into a disaster. Emergency management involves plans and institutional arrangements to engage and guide the efforts of government, non-government, voluntary, and private agencies in comprehensive and coordinated ways to respond to the entire 14 spectrum of emergency needs. The expression "disaster management" is sometimes used instead of emergency management. *Emergency services:*

    > The set of specialized agencies that have specific responsibilities and objectives in serving and protecting people and property in emergency situations.

    Emergency services include agencies such as civil protection authorities, police, fire, ambulance, paramedic and emergency medicine services, Red Cross and Red Crescent societies, and specialized emergency units of electricity, transportation, communications, and other related services organizations.
- Susceptibility
  - The UNISDR terminology does not present a definition for this term. The Brazilian Civil Defense mentions the *susceptibility to erosion*:

    > Greater or lesser tendency of the soil to undergo erosion. The susceptibility to erosion depends on the slope and the characteristics of the terrain profile, the soil granulometry and anthropic factors related to: removal of vegetation; concentration of rainwater; exposure of susceptible land; inadequate execution of cuts and landfills.

In general, the susceptibility is associated with a determinated space. Saito [62] defines susceptibility as an inherent characteristic of the environment, representing the fragility of the environment.

- Vulnerability
  - Brazilian Civil Defense:
    1. Condition intrinsic to the body or receiver system that, when interacting with the magnitude of the event or accident, characterizes adverse effects, which are measured according to the intensity of the potential damage.
    2. Existing relationship between the magnitude of the hazard, if it happens, and the intensity of the resulting damage.
    3. Likelihood of a particular community or geographical area being affected by a hazard or potential risk of disaster, established from technical studies.
    4. It corresponds to the level of intrinsic insecurity from a disaster scenario to an adverse event. Vulnerability is the opposite of security.
  - UNISDR:

    The conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards.

- Risk
  - Brazilian Civil Defense:
    1. Measurement of potential damages or economic losses in terms of statistical probability of occurrence, intensity, or magnitude of the foreseeable consequences.
    2. Likelihood of occurrence of an accident or an adverse event related to the intensity of the resulting damage/losses.
    3. Probability of potential damage within a specified period of time and / or operational cycles.
    4. Factors established by systematic studies, which involve a significant probability of occurrence of an accident or disaster.
    5. Relationship between the likelihood of occurrence of an adverse event and the level of vulnerability of the affected system.
  - UNISDR: "The combination of the probability of an event and its negative consequences." Comment: This definition closely follows the definition of the ISO/IEC Guide 73. The word *risk* has two distinctive connotations: in popular usage the emphasis is usually placed on the concept of chance or possibility, such as in "the risk of an accident," whereas in technical settings the emphasis is usually placed on the consequences, in terms of "potential losses" for some particular cause, place, and period. It can be noted that people do not necessarily share the same perceptions of the significance and underlying causes of different risks.

- Acceptable risk
  - Brazilian Civil Defense: Very small risk, with limited consequences, associated with real or expressive benefits in a way that social groups would be willing to accept it. The acceptability of the risk is based on scientific, social, economic, and political information, including the benefits resulting from this condition.
  - UNISDR:

    > The level of potential losses that a society or community considers acceptable given existing social, economic, political, cultural, technical and environmental conditions. Comment: In engineering terms, acceptable risk is also used to assess and define the structural and non-structural measures that are needed in order to reduce possible harm to people, property, services and systems to a chosen tolerated level, according to codes or "accepted practice"İ which are based on known probabilities of hazards and other factors.

- Adaptation
  - Brazilian Civil Defense: *there is no mention to this term*
  - UNISDR:

    > The adjustment in natural or human systems in response to actual or expected climatic stimuli or their effects, which moderates harm or exploits beneficial opportunities. Comment: This definition addresses the concerns of climate change and is sourced from the secretariat of the United Nations Framework Convention on Climate Change (UNFCCC). The broader concept of adaptation also applies to non-climatic factors such as soil erosion or surface subsidence. Adaptation can occur in autonomous fashion, for example, through market changes, or as a result of intentional adaptation policies and plans. Many disaster risk reduction measures can directly contribute to better adaptation.

- Resilience
  - Brazilian Civil Defense: Ability of an individual to deal with problems, overcome obstacles, and resist to the pressure of adverse situations without going into psychological outbreak. It also refers to decision-making when someone faces a dilemma concerning the environment stress and the willingness to win.
  - UNISDR:

    > The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions.

  Resilience means the ability to "resile from" or "spring back from" a shock. The resilience of a community in respect to potential hazard events is determined by the degree to which the community has the necessary resources and is capable of organizing itself both prior to and during times of need.

3 Machine Learning in a Nutshell: Concepts and a Case Study in Disaster Risk Reduction Scenarios

Recent improvements in computational power and the availability of massive volumes of data produced in the big data era give rise to the increasing adoption of machine learning (ML) solutions. ML can play an important role in monitoring hazards. In particular, ML techniques have been employed in the analysis and detection of earthquakes [8, 44, 63], floods [2, 9, 25, 55], forest fires [19, 22, 53, 72], as well as forecasting future occurrences of such hazards.

According to Mitchell [54], a machine (or computer program) is said to learn from experience $\mathscr{X}$ with concerning to some task $\mathscr{T}$ measured by a performance measure $\mathscr{P}$, if its performance in the task according to $\mathscr{P}$ improves with experiences. The experience $\mathscr{X}$ is the training set that represents instances collected independent and identically distributed (*i.i.d.*) from the domain, in order to maximize the computer program performance. The task $\mathscr{T}$ is the operation or problem expected to be solved by the computer program, such as classifying new samples, finding clusters or patterns, predicting values, among others. The performance measurement $\mathscr{P}$, like accuracy, precision, error, etc., is a quantitative metric which evaluates the algorithm performance and success in the resolution of the task. Thus, the computer program is an algorithm that seeks for a mathematical function capable of solving a task or problem considering previous data.

Formally, we have the set of *i.i.d.* samples $\mathscr{X}$ and according to the task to be performed, the first l samples ($0 \leq l \leq n$) can contain labels $\{(x_1, y_1), \ldots, (x_l, y_l)\}$ and the remaining u samples are unlabeled, $U = \{x_{l+1}, \ldots, x_u\}$. Labeled samples are associated with the set $\mathscr{Y} = \{y_1, \ldots, y_l\}$, where $y_i \in \{1, \ldots, c\}, i = \{1, 2, \ldots, l\}$, with c representing the number of classes, and $(l+u = n)$ representing the total number of samples. A sample can be seen as a point in the dimensional space that is a collection of features. Each sample x_i is represented by a *features vector* $x_i \in R^d$ where d is the vector dimension or number of features. Each feature of x_i quantitatively measured a characteristic of the sample. For example, in a problem involving rain prediction, we can assume that each sample is represented by a vector with the location coordinates latitude, longitude, and the rain measure.

Learning tasks ($\mathscr{T}$) are associated with some experience or learning paradigm. Among many learning paradigms reported in the literature [17, 18, 21, 56, 74], the most commonly addressed are presented below:

- **Supervised learning:** in this paradigm the algorithms learn from the input data $\mathscr{X}$ and their associated labels $\mathscr{Y}$. These algorithms acquire experience from the dataset, where each sample is associated with a label, i.e., with $l = n$. The objective is to construct an inductive or transductive model that describes the dataset and predicts the class or a feature value of unseen samples;
- **Semi-supervised learning:** the algorithms employ only a small number of labeled samples and a large amount of non-labeled data for the training, thus seeking the construction of a function or model that requires less human effort

to label the samples. The objective is to infer the missing labels $\{y_{l+1}, \ldots, y_u\}$ corresponding to unlabeled samples $\{x_{l+1}, \ldots, x_u\}$, where $l \ll n$;

- **Unsupervised learning:** in this paradigm the samples have no label, with $l = 0$, i.e., the algorithms acquire experience from datasets where samples have no classes associated. The objective is to find patterns in the input datasets. Clustering is the usual task, where input samples are divided into groups based on their similarity, density, proximity, etc., since there is no prior knowledge about the classes.

3.1 Usual Tasks of Machine Learning for Disaster Risk

Tasks are the formal description of ML paradigms about how to process the samples. A particular task can be employed to solve many kinds of problems of disaster risk monitoring and management. In this section, we make the connection between ML and risk management. Below, we briefly describe three of these ML tasks:

- **Classification:** this is a supervised and semi-supervised learning task where, given a number of c possible classes or labels, the ML algorithm seeks to generalize and predict the label of samples not yet observed. One hypothetical example of ML classification tasks applied to environmental problems is the classification of deforestation areas in satellite images. The task defines one of two possible classes (deforestation or not deforestation) for each pixel from a satellite image. Based on the number of samples used for training, the algorithm must be able to predict for a satellite image not previously observed, whose pixels are deforested areas. The *accuracy* is the usual metric applied to evaluate the correctness of the algorithm. It is important to realize that we are interested in measure performance of proposed algorithms in a set of data not seen before. So, to quantify performance measure, we need to use a test set of data which is not contained in the training data applied into ML algorithm training step.
- **Regression:** it is a supervised task that tries to predict a future numerical value given a set of samples. Usually, the samples are captured over time and the values are related to some measurement. A good example of regression in the environmental scenario is the rain prediction, where given a time series of past rain measurements in a specific location, the algorithm will predict the expected amount of rain in the future. The performance measures in this task try to minimize the error among the predicted value and the real result. *The mean absolute deviation* (MAD) and *root mean square error* (RMSE) are instances of evaluation metrics, where lower values result in better predictive models.
- **Clustering:** this kind of ML task focuses on learning a probability density function (in cases where samples domain is continuous) or a probability mass function (otherwise) on the samples space in a way to learn the structure of the data. Clustering problems, where the objective splits a set of samples in k different groups, without any prior knowledge, is a good example of this

kind of problems. In the environmental scenario, for example, we can consider an approach of spatial clustering for detecting forest fire spots in satellite images [40]. In this task, there are several metrics for evaluating the quality of the separation, like *Silhouette* or *Jaccard* index, statistics like *GAP* or *adjusted Rand* indexes, *variation of information*, among many others.

The interest in ML approaches arouses due to the advances in computational technology and the vast amount of available information in the big data era. Traditional methods are studies dependent on a specific area, with inflexible models that cannot be easily generalized to other lands [22]. Another approach is the proposal of numerical indexes that incorporate meteorological data in the calculation. These indexes are employed for improving proactive capacity (e.g., public warnings) and mitigation capacity (e.g., defining guidelines for preventive actions, prioritizing regions, etc.). Nevertheless, these indexes highly correlated with the environmental characteristics and activity in specific regions where they were proposed.

On the other hand, ML approaches have the advantage of flexibility: models can be adapted to different lands and conditions. All kind of data sources, see Table 2, are valuable information to be employed for identifying trends and patterns (like Twitter, meteorological data, images, etc.), which can improve decision making. In this way, the classical statistical analysis can fail when such vast and complex data are presented [2, 22, 72]. Since there is an interest in developing automatic solutions, a fast detection is one of the key elements for a successful hazard prediction [25]. Hence, ML tools are a suitable tool to analyze different data and extract high-level information in a short time, for hazards monitoring.

However, traditional or ML methods face some drawbacks depending on the problem domain, like when there are important limitations on the quality and amount of available data, e.g., weather observations, disasters data, satellite images, etc. Fortunately, some techniques and ML methods can help in bypassing these problems. For instance, data augmentation [84] is the process of synthetically creating samples based on existing data. Other alternatives are algorithms like *k*NN or random forest, which have robust results when missing values or a small amount of data. They work well on non-linear and categorical data and adapt to the data structure taking into consideration the variance or the bias of samples. Also, according to the domain knowledge, the data analyst can employ common techniques for increasing the quality and instances. For example in time series data, a linear interpolation works well for data with some trend but without seasonality; for seasonal series is necessary a seasonal adjustment and the interpolation method; techniques like sample or multiple imputations, last observation carried forward (LOCF) or next observation carried backward (NOCB), and many others, can also be useful.

The point is that each ML method or specific algorithm is better adapted to solve some problems than others, which makes impossible to find a universally optimal ML algorithm. It can be said that for each problem, there is a unique ML algorithm. Therefore, the goal is to discover the most appropriate algorithm for a given domain

Table 2 Machine learning tasks applied to disaster risk and management

Classification	Fire	Vega-Garcia et al. [76] presented a supervised and regression method for predicting human-caused wildfire occurrence from geographic information and historical human-caused fire occurrences
	Fire	Mazzoni et al. [53] presented a method for detecting smoke in satellite images at 1.1 km pixel level
	Flood	Abrahart and See [2] presented a method for short-term forecasting of rainfall-runoff for a flood forecasting systems from hydrological data
	Flood	Mojaddadi et al. [55] presented a method for identification and prediction of the susceptible fluvial flood areas using multi-sensor remote-sensing data and GIS
Regression	Fire	Stojanova et al. [72] reported a method for detecting fire occurrences in the Slovenian forests using both satellite-based and meteorological data
	Fire	Cortez and Morais [22] proposed a supervised and semi-supervised approach for predicting the burned area of small forest fires using meteorological data
	Flood	Banihabib [9] presented a method for determining flash flood warning lead-time in a steep urbanized watershed using flood hydrographs data
	Flood	de Lima et al. [25] reported a method for predicting river levels and preventing hydrological disasters using pluviometric and hydrological data
Clustering	Fire	Hsu et al. [40] reported an unsupervised approach for detecting forest fire spots in satellite images
	Fire	Cheng and Wang [19] proposed a spatio-temporal clustering approach to discover the cells (hot spots) with highest probability of starting a fire, by using records of large forest fires in Canada
	Earthquake	AYuen et al. [8] presented a global warning system for the analysis and visualization of seismic data. The proposal is a semi-supervised method for detecting earthquakes and alerting the probability of tsunami according to live data and previous events
	Earthquake	Sakaki et al. [63] presented an earthquake reporting system that promptly detects the events and alerts to registered users. They employed an unsupervised clustering approach and construct a spatio-temporal model for detecting the center and trajectory of event location from Twitter messages. The system has high accuracy results, with a fast detection of earthquakes and alerts to the population

and problem. For this reason, we have a large number and variety of ML methods reported in the literature for solving disaster risk problems. In the next section, we discuss about some of the main adopted ML algorithms.

3.2 Machine Learning Based on Features Vector

Each of the problems described in Sect. 3.1 can be tackled using different algorithms. The range of these type of algorithms goes from simple linear regression to a robust ensemble of classifiers. Recalling that vector-based algorithms have as input the set of vectors $x_i \in R^d$, where each one represents a sample and d is the number of features. In this section, we briefly describe four examples of vector-based ML algorithms related to the previously mentioned problems.

3.2.1 Artificial Neural Networks

This is one of the most famous classes of machine learning algorithms. Artificial neural networks (ANN) algorithms can be associated with different kinds of tasks as classification, regression, and others. ANNs are inspired in biological neural networks that constitute animal brains. They are based in a set of connected units called artificial neurons [5, 16]. ANNs architectures frequently contain perceptron or sigmoid neurons, organized in layers and use the standard learning algorithm known as stochastic gradient descent for learning processes.

As defined by de Lima et al. [25], "an ANN is trained to learn the input-output relationships through an iterative process in which the weight assigned to each input attribute is adjusted to minimize the error between the ANN outputs and the true outputs (the output observed values). Once the learning process has been completed, it is expected that the ANN has good generalization ability and can properly respond to new data." ANN and derivations have been studied in different disaster-type problems. For instance, Vega-Garcia et al. [76] trained an ANN for predicting human-caused wildfire occurrence. Abrahart and See [2] did the same but for classifying short-term forecasting of rainfall-runoff. On the other hand, Banihabib [9] and de Lima et al. [25] trained an ANN in a regression task with the purpose of obtaining a risk score for flood warnings.

3.2.2 Regression

The most intuitive regression algorithm is associated with a linear regression task, which consists of estimating values based on the continuous variable(s) [26]. Regression tasks seek a relationship between independent and dependent variables by fitting an unknown function that can be represented by some kind of equation (linear, polynomial, logistic). Given a training dataset, the regression function can be derived by minimizing the sum of squared difference of distance between the data points and the predicted values [32]. Examples or regression methods are the Gaussian linear or additive models, multivariate adaptive splines, M5 regression tree models, logistic regression, and algorithms like ANNs, random forest, SVM, and many others [15, 55, 71, 87].

Works in disaster risk employ regression functions for understanding and approximating the magnitude of some phenomenon. Stojanova et al. [72] used a logistic regression function for detecting fire occurrences in the Slovenian forests. Biffis and Chavez [15] employed a random forest regression in satellite images identifying pixel-level of precipitation risk for maize culture in Mozambique. Sakaki et al. [63] proposed a mixture method of regression and clustering for a very fast detection of earthquakes, with a high precision of the epicenter and the trajectory. The authors employed the real-time message interaction from Twitter, considering each Twitter user as a sensor. In a comparative approach, Shortridge et al. [71] studied six regression models for empirical rainfall-runoff predictions, evaluating the accuracy, interpretability, among other performance measures. They highlighted some of the strengths and limitations of the approaches and showed important issues to be considered for regression models comparisons. Similarly, Cortez and Morais [22] test the multiple regression method against ANN and SVM for approximating the burned area of small forest fires. In this scenario, the regression method did not achieve the best results. However, they did not run a statistical analysis to show a significant difference among the several methods evaluated.

3.2.3 Support Vector Machine

Support vector machines (SVM) are an optimization learning technique that is very important in the literature because it obtains results equivalent or superior to other algorithms in different domains [22, 53, 55, 87]. The SVMs employ the statistical learning theory (SLT) [38], which establishes a set of principles that must be followed to obtain classifiers with good generalization capacity.

Let h be a classification function and H be the set of all classifiers (or functions) that a ML technique can generate. The SLT establishes the mathematical conditions to better choose a classifier $\hat{h}$. The purpose is to select a mathematical function that most minimize the expected error for the classifier $\hat{h}$ by separating the classes according to the hyperplane equation,

$$h(X) = W.X + b\,, \quad \text{with} \quad W \in X, \tag{3}$$

where $W.X$ is the scalar product between the vectors W and X, W is the normal vector of the hyperplane, and $b/|W|$ corresponds to the distance of the hyperplane in relation to the origin, with $b \in \Re$.

In relation to the type of ML paradigm, SVMs can be used for classification and regression, employing linear or non-linear kernels. Linear SVMs employ polynomial kernels and are good for linearly separable samples, i.e., the classes of the samples can be separated into hyperplanes. On the other hand, when it is not possible to divide the training dataset into hyperplanes, the non-linear strategy is adopted. For achieving this purpose, the algorithm maps the training set from their original space to a new space of larger dimension called feature space [38]. In this way, the problem becomes again a linear hyperplane separation.

3.2.4 *K*-Means Clustering

For density estimation, k-means is one of the most popular unsupervised learning algorithms for clustering. This algorithm divides training set $S = \{s_1, s_2, \ldots, s_{n-1}, s_n\}$ into k different clusters of samples where each cluster contains the closest samples [85]. To accomplish this task, the algorithm is initialized with k different centroids $\{c_1, c_2, \ldots, c_k\}$ and continuously executes the following steps until it converges:

1. Associate each sample s_i to one centroid c_j based on minimum distance (which can be measured using different distance metrics);
2. Following, each centroid c_j is updated to the mean of all training samples assigned to cluster j.

In environmental problems, Cheng and Wang [19] employed the K-Means method in spatio-temporal clustering for discovering the spot regions with highest probability of starting a fire. Hsu et al. [40] presented a similar detection approach but clustering forest fire spots in satellite images. Also, AYuen et al. [8] employed clustering for detecting earthquakes according to previous historical events.

3.2.5 Application in the Prevention of Disasters Related to Hydrological Processes

Recently, de Lima et al. [25] proposed a methodology to highlight the power of machine learning applied to environmental problems, especially when a quick answer is necessary. The authors used an ANN model and applied it to an operational scenario for rainfall prediction in advance of up to 2 h. The ANN used at this approach is a standard multilayer perceptron (MLP) with 15 nodes as input layer, which are fully connected to a single hidden layer. Finally, the hidden layer is connected to the single output neuron and employs sigmoid as hidden activations functions. The output neuron employs a linear function. The proposed ANN takes as input rainfall and river level for eight specific positions along a watershed (including river spring, where there is no level information), as depicted in Fig. 2, resulting in 15 input values. Once trained, the model is able to provide forecast for river levels, in a window between 15 and 120 min.

The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models. Such performance metric, denoted by P, was proposed by Nash and Sutcliffe [57], and is one of the most used performance metrics in hydrological problems. It is defined as one minus the mean squared error between the predicted and observed values normalized by the variance of the observed values during the period under investigation [46]. The index is given by

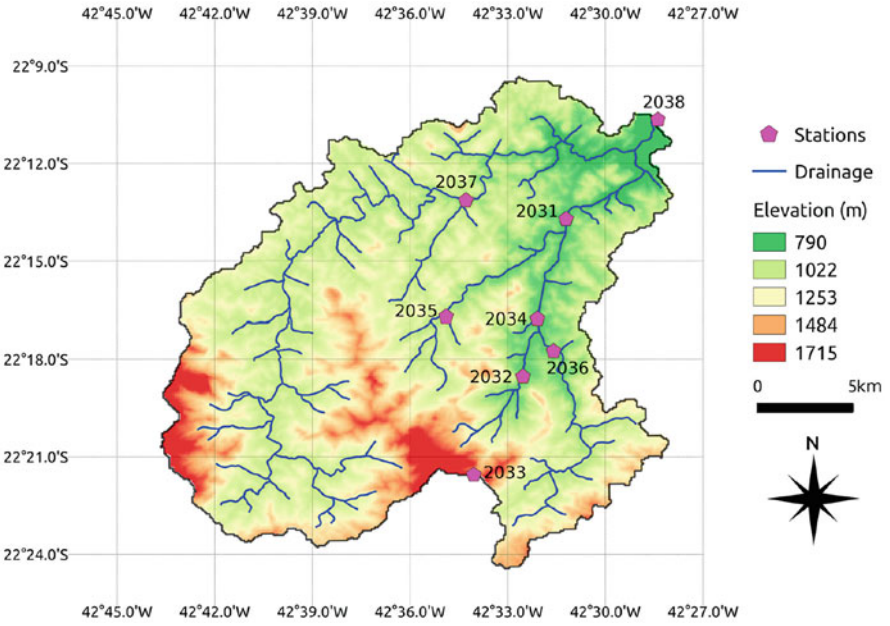


Fig. 2 Case study site considered in de Lima et al. [25] work. Source: de Lima et al. [25]

$$P = 1 - \frac{\sum_{i=1}^{n} \left(O_i - \tilde{O}_i\right)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \tag{4}$$

where n is the number of instances in the test set, O_i and $\tilde{O}_i$ are, respectively, the ith observed and predicted output values, and $\overline{O}$ is the average of the n observed values.

The forecasting of river levels 2 h in advance is an important challenge. In a monitoring scenario, forecasts in such a time window allow for preventive actions during a flood episode. For predictions 15 min ahead, de Lima et al. [25] report a Nash–Sutcliffe index of 0.9816 while for 120 min predictions the authors reported an index equal to 0.8688, which indicates that the predictive power of the proposed models is encouraging. A Nash–Sutcliffe index above 0.7 is already considered good.

3.3 Machine Learning Based on Complex Networks: The Cutting-Edge Approach

Methods based on complex networks are currently the more active and cutting-edge approaches in ML [21]. The data representation in complex networks inherently presents some interesting properties. It has the ability to deal with local and global characteristics of available data, identifying classes or groups independent of the data distribution, and to represent sub-manifold in dimensional space [14, 21], which translates into heterogeneous and nontrivial connections between nodes. Moreover, dynamical or evolutionary process can be integrated into the network structure due to the inclusion and removal of nodes and connections, the large-scale network states, distribution of functions or topological properties, and diversity of connections: the edges can have various physical meanings.

Datasets of different problem domains are not already represented in a network format. In conventional ML methods (Sect. 3.2), the sample data $\mathscr{X}$ are described such as an attribute-value table [54]. On the other hand in ML based on complex networks, the input dataset must be a network, or graph, G.

In order to apply network-based ML for data analysis, it is necessary to transform the original data into a network. The construction of a weighted sparse graph G from $\mathscr{X}$ is an important step for network-based ML methods [14]. In the network G, each sample $x_i \in \mathscr{X}$ is associated with the node i and each weighted edge, which connects the pair of nodes i and j, represents the similarity level between the nodes.

The information of the labeled nodes and G are the input for the ML methods. In supervised and semi-supervised tasks [13, 14], the objective is to predict the class of the unlabeled nodes. Among different approaches reported in the literature [21], we have the probabilistic methods, like the *weighted vote relational neighbor* or *network-only Bayes* classifiers; harmonic functions, like the *local and global consistency* method; minimum spanning tree and cuts, like the *MinCut* approach; label propagation, random walk, particle competition, etc. In the case of unsupervised learning, the clustering task can be considered as the problem of detecting communities on networks. The communities follow the manifold assumption [14] or the homophily principle of social networks [45, 77, 78], in which it is assumed that nodes with greater similarity to each other belong to the same class and tend to be part of the same group. The community detection corresponds to a network partition problem. Many works have reported algorithms to address this task [31], such as spectral methods, techniques based on modularity division, spectral decomposition, and random walk.

In Sect. 3.2, we present the current state-of-the-art approaches for inductive learning, i.e., methods that seek for deducting or inducing a mathematical function that describes the analyzed data [54]. On the other hand, here we present transductive methods, which consider the topological structure of the data and are a recent approach in ML [21]. In the transductive approach, the data is represented as nodes and edges in a graph that can have several properties, like symmetry, weighting,

irreflexivity, acyclicity, heterogeneity, multilayer, among others. Thus, graph-based ML methods are in the knowledge border as a novel research area to be explored for disaster risk and reduction problems.

4 Vulnerability on Complex Networks

Extreme weather and climatic events are expected to increase in frequency and intensity and cause more social and economic impacts on several sectors, including critical infrastructures like gas, oil and water networks, power grid, telecommunications, and transportation system.

Many countries, including Brazil, are signatories of the Sendai Framework for Disaster Risk Reduction 2015–2030. One out of the seven global targets of this document is addressed to "substantially reduce disaster damage to critical infrastructure and disruption of basic services."

Vulnerability is a key concept for disaster risk reduction (DRR). According to Cheung [20], vulnerability means "the characteristics of a person or group and their situation that influence their capacity to anticipate, cope with, resist and recover from the impact of a natural hazard (an extreme natural event or process)" (see Sect. 2). Measurement and mapping of vulnerability constitutes a subject of global interest.

Critical infrastructures, such as a network-type structure, can be represented (modeled) using a complex network approach. The complex networks approach may offer a valuable perspective in the context of DRR on critical infrastructures. There are a lot of topological measurements for complex networks [50], for example, degree—number of connections of a node—and betweenness—number of shortest paths those pass through a node. One measurement is particularly interesting in the context of critical infrastructures: the topological vulnerability.

4.1 Timeline of Vulnerability on Complex Networks

Albert et al. [3] and Holme et al. [39] studied the response of complex networks subject to attacks. They used the term "vulnerability," but they did not account for a pointwise *index* to quantify vulnerability, just a collective measurement [3] based on network's diameter and size of the largest cluster, and [39] based on the average inverse geodesic length.

The first paper that considered the pointwise vulnerability index was Goldshtein et al. [35]. In that paper, the authors cited two relevant previous works: Latora and Marchiori [47, 48].

Table 3 summaries the origins of vulnerability studies on complex networks.

Table 3 Timeline of the related works

Watts and Strogatz [82] and Barabasi and Albert [10]	Seminal works on complex networks
Albert et al. [3]	The authors discussed error and attack tolerance on complex networks based on measures as diameter and size of the largest cluster
Latora and Marchiori [47]	The authors proposed the efficiency index on complex networks
Holme et al. [39]	The authors studied network vulnerability index measured by the average inverse geodesic length
Latora and Marchiori [48][a]	The authors generalized the vulnerability index for any performance measurement, and discussed that idea under the context of a critical infrastructure as a complex network
Goldshtein et al. [35]	It was the first paper that defined a pointwise vulnerability V(i)—based on the ideas from Latora and Marchiori [47, 48]

[a]An arxiv version, later published in Latora and Marchiori [49]

4.2 Using Complex Networks for (Topological) Vulnerability Assessment

In the area of disaster risk management, the quantification of vulnerability is a leading research topic [70]. Contributions towards a quantitative understanding of vulnerabilities aim to (1) guide public policies on DRR, (2) clarify where the most vulnerable elements are, and (3) understand what are the impacts related to the vulnerabilities. Next, we introduce some basic terminologies and concepts related to the assessment of vulnerabilities using a complex network approach.

The shortest path length d_{ij} between two nodes i and j is the smallest number of links from i to j, across all the possible paths between i and j.

Definition 1 (Edge Efficiency) The efficiency e_{ij} in the communication between nodes i and j is inversely proportional to their shortest path length, i.e., $e_{ij} \sim 1/d_{ij}$.

Let's consider a graph $G = (V, L)$, where V is the set of $|V| = N$ nodes and L is the set of $|L| = M$ edges or links. Let E be the average efficiency of G.

Definition 2 (Graph Efficiency) E is given by

$$E = \frac{\sum_{i,j \in V,\ i \neq j} e_{ij}}{N(N-1)} \tag{5}$$

Let V_k be the vulnerability associated with a node (or edge) k of a graph G.

Definition 3 (Element Vulnerability) The vulnerability associated with element k of graph G, V_k, is given by

$$V_k = \frac{E - E_k^\star}{E} \tag{6}$$

where $E_k^\star$ is the efficiency of the graph considering the removal of element k.

Definition 4 (Global Vulnerability) The global vulnerability of a graph is the greatest vulnerability of its elements.

The vulnerability of (associated with) an element on a graph should therefore not be confused with the susceptibility of (associated with) the element to a threat. According to Sect. 2, the susceptibility represents an inherent fragility of the environment. The vulnerability of (associated with) an element on a graph can be understood as the way a system reacts under a concrete threat. Although it is a measurement associated with the element, assuming a possible value for each element (a pointwise measurement), the vulnerability on complex networks brings information about the dynamics throughout the whole network.

Figure 3 shows some complex network's indexes: degree, betweenness, and topological vulnerability. It is important to highlight that by using the vulnerability index we can identify the most relevant elements easier than by using the other indexes. Also, from the disaster risk reduction point of view, that index can be a good proxy of potential impacts, by showing the elements that, if removed, could cause the biggest impacts on the efficiency of the network.

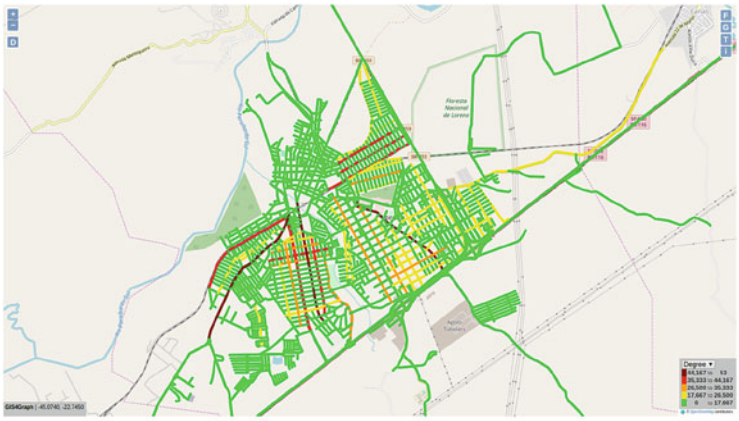
Those maps were produced using the tool GIS4Graph, presented in Santos et al. [66].

4.3 Recent Studies of Vulnerability on Complex Networks

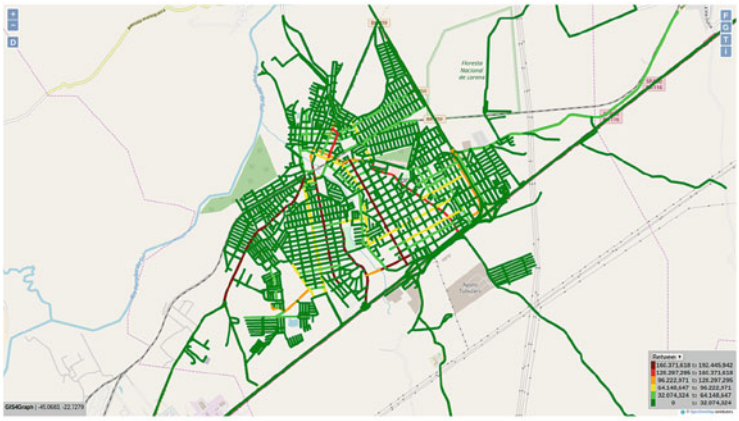
Recently, Wang et al. [81] analyzed vulnerability on complex networks using a performance based on the travel time between all pairs of nodes and the frequency of the services evaluated on the shortest paths—combining the physical structure, topologic character, and the social function of transport networks. Sun et al. [73] proposed a weighted approach, using the number of affected travelers when a station is attacked. Mattsson and Jenelius [52] presented a discussion of recent studies about vulnerability and resilience of transport systems, including works employing complex networks.

Other family of works analyzed vulnerability on complex networks using different performance measurements, as Gleyze and Rousseaux [33], who used the edge betweenness index, and, who used the accessibility index.

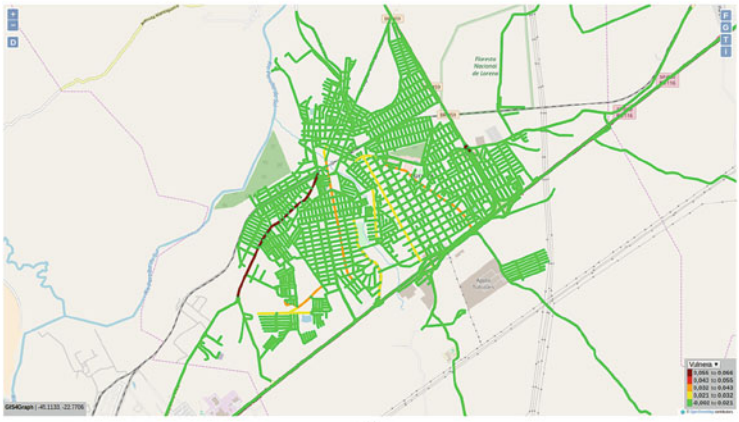
Finally, in Pregnolato et al. [58], a framework for assessing the disruption from flood events to transport systems was presented, combining a high-resolution urban



(a)



(b)



(c)

Fig. 3 Some complex network's indexes: (**a**) degree, (**b**) betweenness centrality, and (**c**) topological vulnerability. Maps produced using the open web tool GIS4Graph

flood model with transport modeling and network analytics to assess the impacts of extreme rainfall events.

None of those papers, however, analyzed the vulnerability index under the DRR point of view.

Several papers were published following similar ideas, even without the complex network approach, from the transportation literature, as Berdica [12], to the DRR literature, as Eleuterio et al. [27] for floods, and Khademi et al. [44] for earthquakes.

Wang and Taylor [80] studied other key concept in disaster risk reduction—resilience—under the complex networks approach. They discussed patterns and limitations of resilience of urban human mobility to disasters.

There are several works applying complex networks approaches to risk assessment [68], for many classes of services as illustrated in Fig. 4, like water supply [86], gas [37], energy [6] supply, and telecommunications [60].

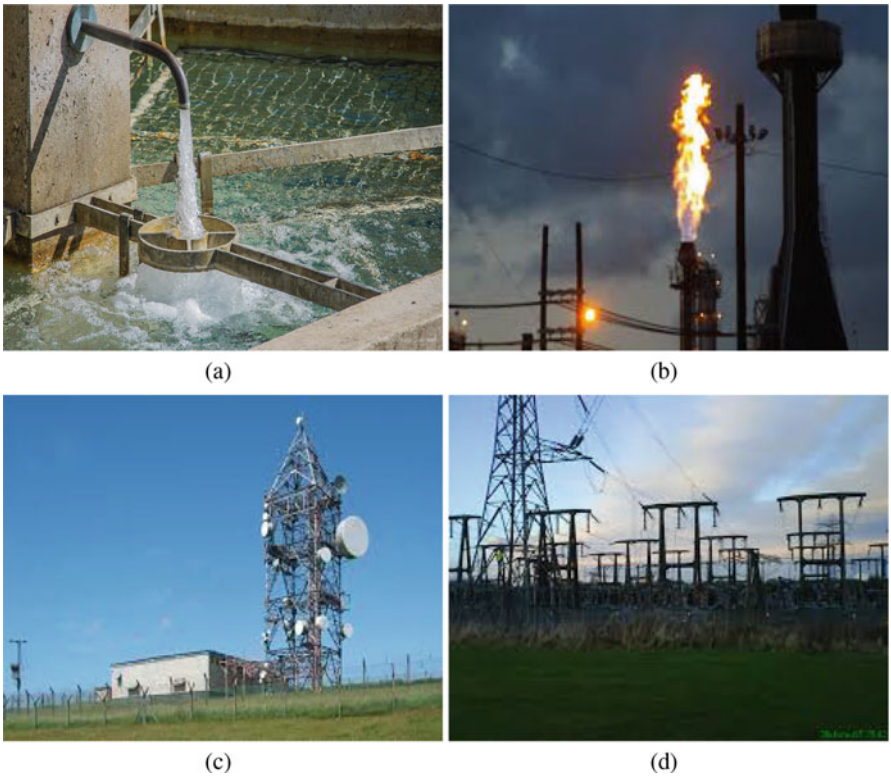


Fig. 4 Illustrative examples of networks supplies: (**a**) Water, (**b**) gas, (**c**) energy, and (**d**) telecommunication. Source: Google Images

In the following section, we introduce survivability models that can leverage the aforementioned complex network concepts to guide recommendations on investments.

5 Survivability of Critical Infrastructures

Background As our society evolves, more aspects of our daily lives depend on large-scale infrastructures such as rail, road, gas, water, power, and telecommunication networks, including the Internet, wired and wireless telephony. Critical infrastructures are widespread and are increasingly more interconnected and interdependent. The more the networks become smarter, the more they rely on information and communication technologies, also known as ICT. For this reason, a failure in the ICT network or in the power network can cause disruptions in many critical infrastructures.

Although smarter systems are, in concept, more robust than their traditional counterparts, failures in the ICT infrastructure can cause disasters. A failure in a signaling system of a railroad can cause human losses, problems in the dimensioning of dams can cause floods, and power outages are another example of disasters that can occur either due to failures in the physical system or due to cyberattacks.

Prior Art Traditional approaches to study the impact of failures and to understand how the system behaves when faced with disruptions involve detailed simulations. The result of these detailed simulations of systems is an accurate assessment of how the system will behave under the considered configurations. In addition, traditional solutions usually involve the full characterization of failure types and failure rates. They combine together failure models and recovery models to yield the so-called availability models.

Steady-state availability assessment consists in studying the system in steady state, and assessing the fraction of time in which the system is functional. As an example, metrics such as the system average interruption index, also known as SAIDI, are largely used by the power systems community.

Limitations of Prior Art and Challenges Survivability is the ability to remain alive or continue to exist. Traditional approaches for survivability assessment face challenges related to (1) scalability, (2) stiff problems due to events that occur in multiple time scales, and (3) sensitivity of the metrics of interest with respect to system parameters.

Although system assessment through simulations is very precise, it falls short on scalability. The high computational costs preclude the analysis of a large number of configurations, and practitioners have to focus on the most likely or promising setups. During the first stages of an exploratory analysis, it is worthwhile avoiding the use of detailed simulations. Instead, a big-picture, bird's-eye view analysis of the problem may be required. The outcome of such analysis can always be subject to more detailed investigation, for instance, through simulations at a second stage.

In addition, failures are rare. The recovery of the system usually occurs at a much faster time scale when compared against the time between failures. Therefore, decoupling failure rates from repair rates is beneficial. Finally, steady-state metrics are usually poorly sensitive to small changes in the system configuration. This is due, in part, because failures are very rare.

Goal Our aim is to present metrics, models, and heuristics to explore the state space in a principled, scalable, and effective way that can be used, for instance, to guide equipment upgrades and investments. To achieve this goal, we rely on survivability-related models which consider the behavior of the system from failure up to recovery, i.e., the initial state of the system is assumed to be a failure state and the focus is on the contingency plan. A survivability model is a reliability model, conditioned to the fact that the initial state is a failure state, which is a result of a failure event.

As an alternative to traditional simulation-based approaches, we focus on the use of model-based solution to perform an approximate analysis of the system. Rather than steady-state metrics, we consider transient metrics to assess how the system behaves from failure up to recovery. We refer to such approach for the assessment of transient metrics from failure up to repair as a *model-based survivability approach*.

Survivability is the ability of the system to recover service levels in a timely manner after the occurrence of disasters. Survivability models consider how the system behaves given that a disaster occurred. We refer to such models as GOOD models, which are models that study the system **g**iven **o**ccurrence **o**f **d**isasters. ROOD models, in contrast, consider **r**andom **o**ccurrences **o**f **d**isasters. Below we specifically analyze water, gas, and power systems, considering as a common theme the use of the model-based survivability approach for the analysis of such networks.

5.1 Water

In treatment systems, water moves through tanks which add different chemicals to it before it is either disposed or consumed [43]. The main goal of water supply companies is to reliably offer high quality water, whereas sewage facilities must ensure that a predefined maximum amount of water can be taken from the community sewage system and be cleaned and released with acceptable quality.

Our focus here is on wastewater management systems. Before being released, water is cleaned through multiple physical and chemical steps. The supervisory control and data analysis (SCADA) systems are used to remotely manage treatment and distribution facilities in all phases of operation [69].

One of the challenges of such systems consists of the fact that it involves both continuous and discrete state variables. The amount of water in each tank is continuous, as water is a fluid, while the setting of valves and the occurrence of failures is typically captured through discrete variables.

Hybrid fluid/discrete systems have been subject of study for a few decades. For the study of water systems, hybrid Petri networks have been used to model the state of the water level at the tanks, as well as the state of different valves. In a typical hybrid Petri net to characterize sewage systems, each tank is associated with a continuous random variable, and tanks are connected via fluid transitions, which represent how water flows between them. These transitions are governed by discrete controllers, which can be used to model the ICT part of the system. A detailed description of the model framework can be found in [7, 36].

Using a hybrid Petri net model, Avritzer et al. [7] concluded, as expected, that as the intake rate increases, the probability of no flood decreases. At the same time, the earlier a failure occurs, the larger the probability that no flood occurs. This happens because earlier failures occur while the tank level is still not very high. Therefore, chances are that a repair will take place before a flood. However, late failures favor floods, as water levels will rise over time and it is more likely that there will be no time for a repair before a flood occurs. If water levels tend to raise over time, late failures are more dangerous than earlier ones.

5.2 Gas

The survivability evaluation of gas distribution is receiving increased attention due to competitive challenges raised by novel features, such as demand-response control applications, smart monitoring, and actuation devices, as well as novel industrial organization of utilities [1, 30]. All such novel ICT features require additional attention when assessing systems survivability. Demand-response applications shift demand to reduce load during peak hours. Smart monitoring allows to quickly find failures and assists users to adjust consumption. Finally, gas distribution was traditionally executed by a centralized company, but is now being distributed among multiple competing utilities. For such reasons, the complexity of gas distribution is increasing, and its survivability requires additional attention.

Similar to the wastewater survivability model, the gas distribution survivability model is also a hybrid model, as gas is treated as a fluid whereas failures and repairs are analyzed using discrete variables. To model such systems, Avritzer et al. [7] consider stochastic time Petri networks. A Petri network is a language which allows us to capture system dynamics. In this case, it is used to model how the system behaves from failure up to recovery. Recovery can go through stages, such as failure detection and localization, pipe sectioning, pressure regulation, and repair and network restoration. The detailed analysis of such a network is out of the scope of this work, and is presented in [7].

A survivability-based model can be used to quantify the probability that gas distribution users are not served after failures. Immediately after a failure, the probability that users are not served increases. Afterwards, it decreases, and eventually reaches zero after a repair takes place.

5.3 Power

Power distribution grids are going through a paradigm shift. With the increased use of ICT, some of the same features already mentioned when discussing gas networks are also being used in power networks: demand-response applications, smart monitoring, and automated recovery from failures. Next, we introduce a model for power grid survivability.

As in the previous models, a power grid survivability model accounts for hybrid nature of the quantities involved. A Markov model with rewards rates can be used to capture how the system behaves from failure up to recovery. Different states represent different stages of recovery in the phased-recovery process, and reward rates capture the rate at which power is offered at each of the states.

Next, we consider a simple instance of a Markov model with rewards, where immediately after failure we have three possibilities: either ICT is functioning and there is enough backup energy to supply non-failed sections, which occurs with probability pq, or ICT is functioning but there is no energy to supply non-failed sections, which occurs with probability $p(1-q)$, or there is no communication, which occurs with probability $1-p$. In the first case, the system is amenable to automated recovery, whereas in the last case the system needs to go through intermediary adjustments before being recovered. Additional states can be considered to account for partially recovered systems. The final absorbing state corresponds to a fully recovered system. Varying the value of the probability that ICT and backup power are available immediately after a failure, which are given by p and q, respectively, the model can be used to assess the impact of different investments.

The Markov model can be used to numerically investigate system behavior under different sets of parameters. For instance, it can be used to compute the expected accumulated energy not supplied, from failure up to repair. The accumulated energy not supplied initially increases and then converges to its asymptotic value, after repairs take place. The effect of demand-response can be taken into account, for instance, by changing the rates between states, the state reward rates, or the state space itself. Then, the proposed model allows to quantify the impact of demand-response benefits. For further details, we refer the reader to [4, 7].

5.4 In a Few Words...

In this section, we presented tools to model the survivability of critical infrastructures such as water, gas, and power. We addressed challenges such as the fact systems are hybrid, having continuous and discrete features, and are deployed at large scale. The presented models (hybrid Petri nets and Markov models with rewards) can be used to bridge practitioners and analysts through a clear and simple language and an agreed set of parameters. Finally, we coped with the problem of

dealing with different time scales by analyzing the systems from failures up to repair.

6 Concluding Remarks

Why managing the risk? Instead of managing the impacts of a disaster, the management of the risks fosters the efficiency of the processes, avoiding losses and damages and increasing profits, in a way that eventually the decision making actions will also be more efficient.

We are only able to manage the processes we know. It is important, therefore, to understand the risks we are dealing with, their main causes, and main implications. Knowing the risks is a complex task, which demands the understanding of root causes, the large theoretical and multi-disciplinary background, and the net of interactions between actors and events.

Considering the equations of risk, there are several publications on machine learning for the "hazard" component, others on complex networks for the "vulnerability" component and also publications on survivability analysis which address applications to risk analysis.

The complex systems theory is a branch among the approaches that lead to risk understanding and efficient management. Some tools, as the machine learning, can easy the organization of ideas and information. Varying infrastructure systems can benefit from the use of these tools and from the understanding of complex systems applied to disaster's theory and practice.

Acknowledgements This research is partially supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, grant 2015/50122-0) and DFG-GRTK (grant 1740/2). DAVO acknowledges FAPESP (grant 2016/23698-1) and LBLS acknowledges FAPESP (grant 2018/06205-7) for the financial support.

References

1. Abbatantuono, G., Lamonaca, S., La Scala, M., Stecchi, U.: Monitoring and emergency control of natural gas distribution urban networks. In: IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS), 2016, pp. 1–6. IEEE, Piscataway (2016)
2. Abrahart, R., See, L.: Neural network modelling of non-linear hydrological relationships. Hydrol. Earth Syst. Sci. Discuss. **11**(5), 1563–1579 (2007)
3. Albert, R., Jeong, H., Barabasi, A.L.: Error and attack tolerance of complex networks (2000). arXiv:cond-mat/0008064v1
4. Alobaidi, I.A.: Dependability Analysis and Recovery Support for Smart Grids. Missouri University of Science and Technology, Rolla (2015)
5. Anderson, J.: An Introduction to Neural Networks. A Bradford Book. MIT Press, Cambridge (1995)

6. Arianos, S., Bompard, E., Carbone, A., Xue, F.: Power grids vulnerability: a complex network approach (2009). arXiv:08105278 [physicssoc-ph]
7. Avritzer, A., Carnevali, L., Ghasemieh, H., Happe, L., Haverkort, B.R., Koziolek, A., Menasche, D., Remke, A., Sarvestani, S.S., Vicario, E.: Survivability evaluation of gas, water and electricity infrastructures. Electron. Notes Theor. Comput. Sci. **310**, 5–25 (2015)
8. AYuen, D., Kadlec, B.J., Bollig, E.F., Dzwinel, W., Garbow, Z.A., da Silva, C.R.S.: Clustering and visualization of earthquake data in a grid environment. Vis. Geosci. **10**(1), 1–12 (2005)
9. Banihabib, M.E.: Performance of conceptual and black-box models in flood warning systems. Cogent Eng. **3**(1), 1127, 798 (2016)
10. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
11. Bell, M.G.H., Kanturska, U., Schmocker, J.D., Fonzone, A.: Attacker defender models and road network vulnerability. Philos. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci. **366**, 1893–1906 (2008). https://dx.doi.org/10.1098/rsta.2008.0019
12. Berdica, K.: An introduction to road vulnerability: what has been done, is done and should be done. Transp. Policy **9**, 117–127 (2002)
13. Berton, L., Vega-Oliveros, D.A., Valverde-Rebaza, J.C., da Silva, A.T., de Andrade Lopes, A.: The impact of network sampling on relational classification. In: 3rd Annual International Symposium on Information Management and Big Data - SIMBig, pp. 62–72 (2016)
14. Berton, L., de Andrade Lopes, A., Vega-Oliveros, D.A.: A comparison of graph construction methods for semi-supervised learning. In: 2018 International Joint Conference on Neural Networks (IJCNN), IJCNN'18, pp 1–8. IEEE, Piscataway (2018). https://dx.doi.org/10.1109/IJCNN.2018.8489524
15. Biffis, E., Chavez, E.: Satellite data and machine learning for weather risk management and food security. Risk Anal. **37**(8), 1508–1521 (2017). https://dx.doi.org/10.1111/risa.12847
16. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York (1995)
17. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, New York (2006)
18. Chapelle, O., Schölkopf, B., Zien, A.: Semi-Supervised Learning. MIT Press, Cambridge (2006)
19. Cheng, T., Wang, J.: Applications of spatio-temporal data mining and knowledge for forest fire. In: ISPRS Technical Commission VII Mid Term Symposium, pp. 148–153 (2006)
20. Cheung, N.K.W.: At risk: natural hazards, people's vulnerability and disasters. Geogr. J. **173**, 189–190 (2007)
21. Christiano Silva, T., Zhao, L.: Machine Learning in Complex Networks. Springer, Cham (2016)
22. Cortez, P., Morais, AdJR: A data mining approach to predict forest fires using meteorological data. In: Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, Associação Portuguesa para a Inteligência Artificial (APPIA), pp. 512–523 (2007)
23. Davis, L. (ed.): Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
24. De Jong, K.A.: An analysis of the behavior of a class of genetic adaptive systems, Ph.D. thesis, Ann Arbor, MI, USA, aAI7609381 (1975)
25. de Lima, G.R.T., Santos, L.B.L., de Carvalho, T.J., Carvalho, A.R., Cortivo, F.D., Scofield, G.B., Negri, R.G.: An operational dynamical neuro-forecasting model for hydrological disasters. Model. Earth Syst. Environ. **2**(2), 94 (2016)
26. Draper, N., Smith, H.: Applied regression analysis. No. v. 1. In: Wiley Series in Probability and Statistics: Texts and References Section. Wiley, Hoboken (1998)
27. Eleuterio, J., Hattemer, C., Rozan, A.: A systemic method for evaluating the potential impacts of floods on network infrastructures. Nat. Hazards Earth Syst. Sci. **13**, 983–998 (2013)
28. Fenton, N., Neil, M.: Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Boca Raton (2012)
29. Fogel, L.J., Owens, A.J., Walsh, M.J.: Artificial Intelligence Through Simulated Evolution. Wiley, Chichester (1966)

30. Folga, S., Talaber, L., McLamore, M., Kraucunas, I., McPherson, T., Parrott, L., Manzanares, T.: Literature review and synthesis for the natural gas infrastructure. Technical Report: ANL/GSS-15/5119262 Argonne National Lab. (ANL), Argonne, IL (United States) (2015). https://dx.doi.org/10.2172/1350046
31. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)
32. Gelman, A., Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models. Volume of Analytical Methods for Social Research. Cambridge University Press, New York (2007)
33. Gleyze, J.F., Rousseaux, F.: Impact of relief accuracy on flood simulations and road network vulnerability analysis. In: ECQTG (2003)
34. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
35. Goldshtein, V., Koganov, G.A., Surdutovich, G.I.: Vulnerability and hierarchy of complex networks (2004). arXiv:cond-mat/0409298v1
36. Gribaudo, M., Remke, A.: Hybrid Petri nets with general one-shot transitions. Perform. Eval. **105**, 22–50 (2016)
37. Gunes, E.F.: Optimal design of a gas transmission network: a case study of the Turkish natural gas pipeline network system. Graduate Theses and Dissertations, p. 13294 (2013)
38. Hearst, M.A.: Support vector machines. IEEE Intell. Syst. **13**(4), 18–28 (1998)
39. Holme, P., Kim, B.L., Yoon, C.N., Han, S.K.: Attack vulnerability of complex networks. Phys. Rev. E **65**, 056, 109 (2002)
40. Hsu, W., Lee, M.L., Zhang, J.: Image mining: trends and developments. J. Intell. Inf. Syst. **19**(1), 7–23 (2002)
41. Huning, A.: Evolutionsstrategie: optimierung technischer systeme nach prinzipien der biologischen evolution. 170 S. mit 36 Abb. Frommann Holzboog Verlag. Stuttgart 1973. Broschiert (1976)
42. Jenelius, E.: Large-scale road network vulnerability analysis. Doctoral thesis. KTH, School of Architecture and the Built Environment (ABE), Transport and Economics, Transport and Location Analaysis. ISBN: 978-91-85539-63-5 (2010)
43. Judi, D.R., Mcpherson, T.N.: Development of extended period pressure-dependent demand water distribution models. Technical Report: LA-UR-15–22068 Los Alamos National Lab. (LANL), Los Alamos, NM (United States) (2015). https://dx.doi.org/10.2172/1209272
44. Khademi, N., Balaei, B., Shahri, M., Mirzaei, M., Sarrafi, B., Zahabiun, M., Mohaymany, A.S.: Transportation network vulnerability analysis for the case of a catastrophic earthquake. Int. J. Disaster Risk Reduct. **12**, 234–254 (2015)
45. Kim, J., Hastak, M.: Social network analysis: characteristics of online social networks after a disaster. Int. J. Inf. Manag. **38**(1), 86–96 (2018). https://dx.doi.org/10.1016/j.ijinfomgt.2017.08.003
46. Krause, P., Boyle, D.P., Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. **5**, 89–97 (2005)
47. Latora, V., Marchiori, M.: Efficient behavior of small-world networks. Phys. Rev. Lett. **87**, 198, 701 (2001)
48. Latora, V., Marchiori, M.: Vulnerability and protection of critical infrastructures (2004). arXiv:cond-mat/0407491
49. Latora, V., Marchiori, M.: Vulnerability and protection of critical infrastructures. Phys. Rev. E **71**, 015, 103R (2005)
50. Lü, L., Chen, D., Ren, X.L., Zhang, Q.M., Zhang, Y.C., Zhou, T.: Vital nodes identification in complex networks. Phys. Rep. **650**, 1–63 (2016)
51. Matisziw, T.C., Murray, A.T.: Modeling s-t path availability to support disaster vulnerability assessment of network infrastructure. Comput. Oper. Res. **36**, 16–26 (2009)
52. Mattsson, G., Jenelius, E.: Vulnerability and resilience of transport systems - a discussion of recent research. Transp. Res. A **81**, 16–34 (2015)
53. Mazzoni, D., Tong, L., Diner, D., Li, Q., Logan, J.: Using misr and modis data for detection and analysis of smoke plume injection heights over north American during summer 2004, pp. B853+ (2005)

54. Mitchell, T.M.: Machine Learning. WCB. McGraw-Hill, Boston (1997)
55. Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., Bin Ghazali, A.H.: Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and gis. Geomat. Nat. Haz. Risk **8**(2), 1080–1102 (2017)
56. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge (2012)
57. Nash, J., Sutcliffe, J.: River flow forecasting through conceptual models: part I: a discussion of principles. J. Hydrol. **10**, 282–290 (1970)
58. Pregnolato, M., Ford, A., Robson, C., Glenis, V., Barr, S., Dawson, R.: Assessing urban strategies for reducing the impacts of extreme weather on infrastructure networks. R. Soc. Open. Sci. **3**, 160, 023 (2016)
59. Purdy, G.: Iso 31000: 2009—setting a new standard for risk management. Risk Anal. **30**(6), 881–886 (2010)
60. Ramaswami, R., Sivarajan, K.N.: Optical Networks: A Practical Perspective. Morgan Kaufmann, Burlington (2010)
61. Rebelo, F.: Geografia física e riscos naturais. Imprensa da Universidade de Coimbra/Coimbra University Press, Coimbra (2010)
62. Saito, S.: Estudo analítico da suscetibilidade a escorregamentos e quedas de blocos no maciço central de florianópolis - sc. PhD thesis, Dissertação (Mestrado de Geografia). Departamento de Geociências da Universidade Federal de Santa Catarina, Florianópolis-SC (2004)
63. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW '10, pp. 851–860. ACM, New York (2010)
64. Santos, L.B.L., Londe, L.R., Soriano, E., Souza, A., Coelho, A.F.: Potential flood-related daily urban mobility problems in Rio de Janeiro (Brazil). Revista do Departamento de Geografia **29**, 175–190 (2015)
65. Santos, L.B.L., Carvalho, T., Anderson, L., Rudordd, C.M., Marchezini, V., Londe, L.R., Saito, S.M.: A rs-gis-based comprehensive impact assessment of floods - a case study in madeira river, western Brazilian Amazon. IEEE Geosci. Remote Sens. Lett. **14**, 1614–1617 (2017)
66. Santos, L.B.L., Jorge, A.A.S., Rossato, M., Santos, J.D., Candido, O.A., Seron, W., de Santana, C.N.: (geo)graphs - complex networks as a shapefile of nodes and a shapefile of edges for different applications. CoRR abs/1711.05879 (2017). http://arxiv.org/abs/1711.05879
67. Schwefel, H.P.: Evolutionsstrategie und numerische optimierung. Ph.D. thesis, Technische Universität Berlin (1975)
68. Setola, R., De Porcellinis, S.: Complex networks and critical infrastructures. In: Chiuso, A., Fortuna, L., Frasca, M., Rizzo, A., Schenato, L., Zampieri, S. (eds.) Modelling, Estimation and Control of Networked Complex Systems. Understanding Complex Systems, pp. 91–106. Springer, Berlin (2009)
69. Sheikholeslami, R., Kaveh, A.: Vulnerability assessment of water distribution networks: Graph theory method. Int. J. Optim. Civil. Eng. **5**(3), 283–299 (2015)
70. Shi, P., Kasperson, R.: World Atlas of Natural Disaster Risk, vol. 366, pp. 1893–1906. Springer, Berlin (2015)
71. Shortridge, J.E., Guikema, S.D., Zaitchik, B.F.: Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrol. Earth Syst. Sci. **20**(7), 2611–2628 (2016). https://dx.doi.org/10.5194/hess-20-2611-2016
72. Stojanova, D., Panov, P., Kobler, A., Džeroski, S., Taškova, K.: Learning to predict forest fires with different data mining techniques. In: Conference on Data Mining and Data Warehouses (SiKDD 2006), Ljubljana, Slovenia, pp. 255–258 (2006)
73. Sun, D., Zhao, Y., Lu, Q.: Vulnerability analysis of urban rail transit networks: a case study of Shanghai, China. Sustainability **7**, 6919–6936 (2015)
74. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)

75. UNISDR: Terminology on disaster risk reduction. United Nations Office for Disaster Risk Reduction (UNISDR), p. 24 (2009). https://www.unisdr.org/we/inform/publications/
76. Vega-Garcia, C., Lee, B., Woodard, P., Titus, S.: Applying neural network technology to human-caused wildfire occurrence prediction. AI Appl. **10**(3), 9–18 (1996)
77. Vega-Oliveros, D., Berton, L., Lopes, A., Rodrigues, F.: Influence maximization based on the least influential spreaders. In: SocInf 2015, co-located with IJCAI 2015, vol. 1398, pp. 3–8 (2015)
78. Vega-Oliveros, D.A., Berton, L., Vazquez, F., Rodrigues, F.A.: The impact of social curiosity on information spreading on networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, pp. 459–466. ACM, New York (2017). http://doi.acm.org/10.1145/3110025.3110039
79. Villain-Gandossi, C.: Origines du concept de risque en occident. les risques martimes ou fortune de mer et leur compensation: les débuts de l'assurance martime. Annexe: Attestation d'emplois au Moyen Age du terme Risque Malta: Foundation for International Studies (1990)
80. Wang, Q., Taylor, J.E.: Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. PLoS One **11**, e0147,299 (2016)
81. Wang, Z., Chan, A., Li, Q.: A critical review of vulnerability of transport networks: From the perspective of complex network. In: Proceedings of the 17th International Symposium on Advancement of Construction Management and Real Estate Chapter, vol. 92, pp. 897–905 (2014)
82. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440 (1998)
83. Wisner, B., Gaillard, J., Kelman, I.: Framing disaster: theories and stories seeking to understand hazards, vulnerability and risk. In: Handbook of Hazards and Disaster Risk Reduction. Routledge, London (2011)
84. Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D.: Understanding data augmentation for classification: When to warp? In: International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–6 (2016)
85. Wu, J.: Advances in K-Means Clustering: A Data Mining Thinking, Springer Theses. Springer, Berlin (2012)
86. Yazdani, A., Jeffrey, P.: Complex network analysis of water distribution systems (2011). arXiv:11040121 [physicssoc-ph]
87. Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., Pourghasemi, H.R.: Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the three Gorges Reservoir area, China. Comput. Geosci. **112**, 23–37 (2018). https://dx.doi.org/10.1016/j.cageo.2017.11.019

Digital Humanities and Big Microdata: New Approaches for Demographic Research

Pier Francesco De Maria, Leonardo Tomazeli Duarte, Álvaro de Oliveira D'Antona, and Cristiano Torezzan

1 Introduction

Since the mid-twentieth century, quantitative research in Demography has benefited itself from the increasing availability of unidentified microdata [21]. In contrast to what we have available for the so-called *big data*, those data are constructed and organized from periodic population surveys [11] and they carry a considerable amount of information about the population at a given moment [37]. These characteristics have led to these population microdata (whether they are administrative records or traditional censuses) to be known as *big microdata* [28]. These data, available for broad temporal and spatial spans [18, 19], may allow one to reassess old demographic questions—in the light of new quantitative methods capable of extracting previously unobservable knowledge.

Using this richness (of big data or microdata) and the greater availability of techniques, new perspectives for studies in all fields of demographic knowledge can be opened [4]. Especially in the case of census data, these are new methods to analyze the so-considered most reliable data source of all [11]. It is important to note that this volume of data is not easily manageable [4, 17], since it requires significant computational capacity, as well as adequate treatment and preservation [16]. As alternatives in order to use this data, it has been common to resort to new quantitative techniques, which come from areas such as data mining and machine learning, capable of interpreting the data and obtaining new results. The combination of advanced data-processing and analysis techniques with big (micro)data can be addressed to the field of *Digital Humanities* [17].

P. F. De Maria (✉) · L. T. Duarte · Á. de O. D'Antona · C. Torezzan
School of Applied Sciences, UNICAMP, Limeira, Brazil
e-mail: dpierf@gmail.com; leonardo.duarte@fca.unicamp.br; alvaro.dantona@fca.unicamp.br; cristiano.torezzan@fca.unicamp.br

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_11

Demography can use this approach to analyze its vast datasets, since this combination allows not only to study new sources, but also to return to old datasets with the aim of understanding them better [1]. This area has grown since at least the 1980s, when surveys began to develop—more assiduously—seeking answers in structured data [3]. With the expansion of technological and computational capacities and the massive access to big data and big microdata, more complex researches have gained space and feasibility, as well as been motivated by the potential for an interdisciplinary dialog between the different areas of knowledge—in our case, in both Humanities and Social Sciences. This development has begun a fourth scientific paradigm, exploratory, data-intensive and data-driven, which has modified the way we do science [17].

In the case of researches involving themes and data about population issues, it is also possible to observe and delineate interfaces between Digital Humanities and big microdata. Considering the relevance of the Digital Humanities approach—especially in the field of *Computing Humanities* [22] and the recent deployment of *Digital Social Science* [29]—for studies of large databases from population censuses or sources such as mobile phone data [7, 25] or social networks [2, 38], it becomes essential to consider these approaches to study demographic and interdisciplinary issues.

Considering the potential of these intersections, we resort to a quantitative approach (including machine learning and spatial analysis) in order to analyze big microdata for demographic research. In this chapter, we give special attention to two fields of Demography that can benefit from this range of data and the blend of techniques. First, we show how population and environment (P&E) studies can benefit from Digital Humanities and big microdata in order to discover risk areas for disasters. Second, we discuss how the combination of internal migration and municipal data can reappraise the analysis of migratory dynamic.

This chapter starts with a presentation about what are big microdata and Digital Humanities, followed by a discussion concerning the relations between the formers and demographic research. Thereafter, we explore how this intersection works in two examples, one for each field mentioned above. From the analyses of those cases, we evaluate the potentiality of combining Digital Humanities and big microdata in order to improve demographic research with new features. Finally, we conclude this chapter remarking some challenges about the current uses of Digital Humanities with available big microdata.

2 Big Microdata, Digital Humanities, and Demography

During the decade of 2010, the concept of big data was widely used, with a lot of visibility in the scientific environment and outside it [35], with countless and varied definitions [17], without consensus on what, after all, should be “big data” [5]. A very broad definition states big data as “wide range of large data sets almost

impossible to manage and process by using traditional data management tools—due to their size, but also their complexity" [15]. Part of this popularity is due to the fact that these data, which run away from the conventional format of administrative records and demographic censuses [18], have been increasingly available due to the existence of computational methods capable of not only collecting them, but also storing and processing them [11].

Big data has been an object of study not restricted to the academy, but which, within research institutions and universities, has attracted the interest of several areas, such as Engineering, Mathematics, Economics, and Social Sciences [15]. Concerning the Social Sciences, big data can yield relevant and valuable impacts, since it helps us to: (1) understand issues related to society; (2) stimulate the formulation of public policies more well-directed; and (3) review our ways of observing, measuring, and interpreting human behaviors [8]. However, it is not since the beginning of big data era that Social Sciences could use these rich sources, the area lacked methods for the adequate treatment of these data [8].

This problem has been softened with the development and subsequent popularization of statistical software capable of handling large databases, such as SAS®, Python®, and R® [8]. Additionally, methods capable of handling big data have been developed in areas such as network analysis, textual processing [8], machine learning, and artificial intelligence [34]. However, there are not only these technical and methodological improvements that have promoted the use of big data in Humanities and Social Sciences. The information's level of detail, on the one hand, and the spatial granularity and the time interval of data collection, on the other hand, are reasons behind the interest of social scientists in big data [33]. Finally, the most important interface between big data and the Humanities and Social Sciences is established when we understand that "Social Science provides important context and theoretical insight to explain and understand big data" [33].

Although these elements evince the relevance of using big data in the Social Sciences, it is important to emphasize that the state of the art in this area recognizes that this is not a new issue, at least for population studies. Since before Demography was a scientific field [31], national states already collect data on their populations and organize them into administrative records; these can be considered the very first Big Data–especially considering the period in which they were produced [1]. While the world have witnessed a revolution in the late twentieth century, usually associated with the expansion of data mining [9], demographers point out that in their area there has also been a microdata revolution [21]. In terms of the wide availability of microdata for public use, it began approximately in the 1990s [20]; already in terms of microdata surveys, substantial growth had occurred in the late 1970s [13].

Since the census surveys have been systematized, both production and dissemination of unidentified data have grown—gaining the name of *big microdata* [28]. These big microdata are "consistent large-scale microdata that extend over many decades and span national boundaries with fine geographic detail [that] provide a unique laboratory for studying demographic processes and for testing social and economic models" [28, p. 293]. In addition, they meet the basic prerequisites to be

considered as part of the data revolution, since this can be defined as "an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there are data (...)" [30]. In this way, we can observe that the increasing spread of so-called big microdata is not only part of the data revolution, but also a key part of it and, in our understanding, it is part of big data.

In general, such microdata have the following characteristics: they are individually collected for the entire population, covering the main demographic themes, with low levels of missing values and high level of structuring [28]. However, in order to truly exploit the potential of these microdata, creative and innovative methods must be used for their analysis [28], which involve the use of data mining, machine learning, network analysis, and econometric techniques. At this intersection, the link between Digital Humanities (the quantitative instrumental) and the big microdata (the data source) begins to be better delineated. For this link to be clearer and more visible, it is necessary to understand how to access, analyze, and preserve this data [16], but also how to combine instances and variables in order to obtain meaningful results [12].

This involves, in the first case, to apply methods for organization, standardization, and data processing, and second, to use (among others) data mining techniques, network analysis, machine learning, multicriteria decision support methods, and geoprocessing. Demographic studies have greatly benefited from Digital Humanities, which also means seizing techniques to deal with the big data, on the one hand, and to broaden the applications of this data in Demography, on the other hand. Finally, the understanding and correct selection of data-processing techniques will also have the potential to benefit studies with big microdata (which are traditionally used by demographers), since the Digital Humanities approach allows us to look at the same data from another prism.

In this age of information, data have been increasingly important in relation to models, and both data-driven decisions (DDD) and data-analytic thinking have been more successful than model-based analysis [27]. This paradigm shift [17] is not only reflected in the data and methods available, but also in how we collect new data, such as the demographic census [11]. With these ongoing transformations, demographic researches may benefit from new, data-driven approaches to further explore the big set of available microdata, not necessarily relying on theories to analyze the data.

Thus, what and how much can we gain from this "data avalanche"? A new data source that needs creative strategies and techniques in order to properly be used [28]. If a first-tier demographer as Steven Ruggles invokes that we should be creative and innovative in order to properly explore high-quality big microdata, we can extend the techniques for big data to this domain. Therefore, there is an intersection between big microdata and Digital Humanities that can be exploited.

So far, we showed the relevance of big microdata in Social Sciences and for demographic research. Additionally, we argued that there is a link between big microdata and Digital Humanities. However, how can we draw this connexion? In order to answer this question, we should evaluate how to access, analyze, and preserve these data [16]. In order to answer this kind of question, we must be willing

to work on multidisciplinary teams and customize traditional big data methods to respond to the specific interests of the Social Sciences. Some ways to fulfill this task include (supervised and unsupervised) machine learning, econometric, spatial and multiple-criteria decision (MCDA) analyses.

Two relevant issues are related to what Digital Humanities stand for and if demographic researches fit in it. First, there is not a consensus about what *Digital Humanities* stands for and more fuzziness exists around the *Digital Social Sciences* concept [29]. Though the former considers the latter a different field, we propose that both live together, once Social Sciences and Humanities are, most of the time, integrated. By analyzing different techniques related to Digital Social Sciences, Spiro [29] observes that a relevant intersection exists between them and Digital Humanities. So, in this chapter, we are going to explore techniques that, elsewhere, can be defined as Digital Social Sciences, but that, hereafter, we propose to be part of the Digital Humanities.

Second, we should define how Demography can benefit itself by Digital Humanities and, consequently, how big microdata can be treated by these techniques. As a first approach, we should remember how we did (and still do) traditional data analysis: we prefer supervised models in order to process data. In the era of big (micro)data, datasets are more relevant than models: it is more important to know what is hidden in our data (structure, patterns, etc.) than how we discover that hidden information. This shift promotes advances in terms of machine learning techniques, with the aim of discovering unexpected relations and hidden patterns that traditional techniques do not observe.

Thus, demographic researches can benefit themselves from the new, data-driven approaches in order to deeper explore census-based big microdata. Hereafter, we present two examples related to the use of techniques that can discover unexpected patterns behind the data. First, we combine big microdata and spatial analysis in order to associate environmental risk areas to specific sociodemographic patterns. Second, we use data about internal migration in order to evaluate how evolved the structure and the characteristics of the migratory network. For both cases, we explore the sample census microdata for the Metropolitan Region of São Paulo (MRSP). Data treatment and analysis will be conducted using SAS® 9.4 (for the construction of indexes), Weka® 3.8 (for data mining), and Gephi® 0.92 (for spatial analysis).

In order to analyze population issues using new approaches, we choose the Brazilian demographic census. The choice is supported by the variety of data collected by the Brazilian Institute of Geography and Statistics (IBGE), and also because the demographic census is still one of (if not the most) important sources of data for developing countries [14]. In Brazil, the demographic census offers "the widest possibilities for population studies—their quantification, composition, structure and administrative political distribution (...)" ([26, p. 292], authors' translation). Moreover, as stated by the United Nations [32], the collection of census data is carried out for the whole population, for an uncountable number of demographic, economic, and social themes. Thus, using these highly detailed data is really important in order to implement complex analyses.

3 Revisiting Demographic Issues: Two Examples

Our first example of combining Digital Humanities with big microdata is about the identification of sociodemographic and economic vulnerabilities and their relations to environmental risks. A possible research question would be: "Is there an association between sociodemographic/economic vulnerability and environmental risks?". As a first approach, we have to discover which are the vulnerable area and where they are located. After this, we can overlap these data with information about areas subject to environmental risks. Lastly, we can evaluate if there is a spatial pattern relating vulnerability to environmental risk. In order to solve this problem, we can adopt a vulnerability index.

For this example, we opt for the "Family Vulnerability Index" (FVI), developed and implemented using census data by Furtado [10]. The author develops an index that approaches the idea of vulnerability as the "inability to respond adequately, in a timely manner, to unexpected social or environmental events" ([10, p. 8], authors' translation). This index has six dimensions: (1) social vulnerability (SV); (2) knowledge access (KA); (3) work access (WA); (4) resource scarcity (RS); (5) youth development (YD); and (6) housing conditions (HC). Each dimension has a set of indicators that, properly computed, give a final value between 0 (absence of vulnerability) and 100 (extreme vulnerability). We can obtain the FVI computing the average of the six dimensions. This final process is shown in Eq. (1).

$$FVI = 100 \times \frac{SV + KA + WA + RS + YD + HC}{6}, \quad 0 \leq FVI \leq 100 \qquad (1)$$

For each one of the 633 weighting areas (WA) of MRSP, we compute the FVI, using data from the census sample questionnaire. Additionally, we can explore data about households environment, available from census for the whole population, concerning the spatial presence of local afforestation and sewers, as well as the occurrence of garbage on the street and/or open sewage. With these variables (six from FVI and four from household environment), we can prepare a cluster analysis using EM (expectation-maximization) algorithm[1]—first proposed by Dempster et al. [6]—in order to discover: (1) the ideal number of clusters; (2) the centroid for these groups; and (3) which weighting areas belong to each cluster.

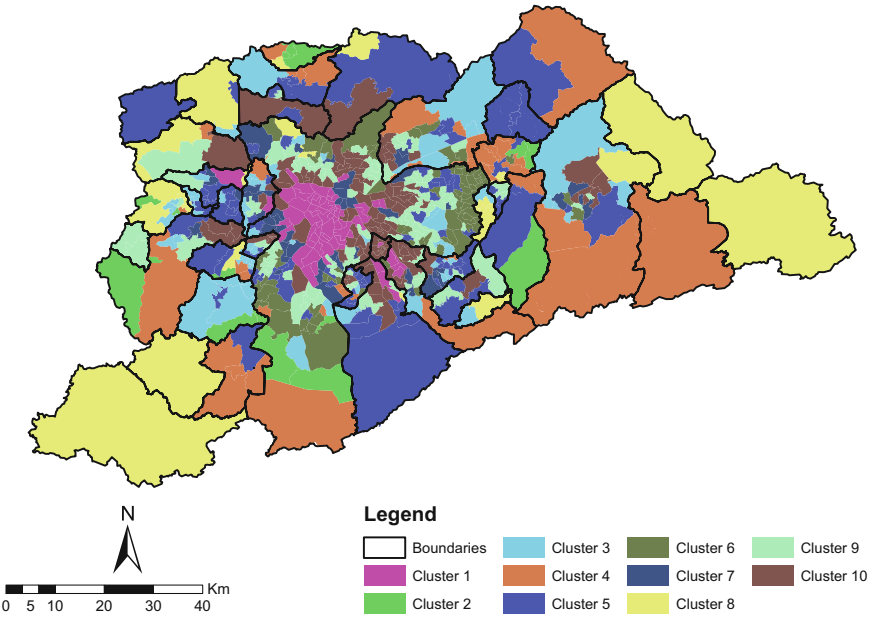
The results of the analysis are shown both in Table 1 and Fig. 1, where one observes that the 10 clusters are quite different from each other. As shown in Table 1, approximately 10% of weighting area belongs to Cluster 1, with a very high level of afforestation and sewers availability; moreover, the indexes of open sewage and garbage are very low. These results are associated with lower degrees of vulnerability in all the six dimensions of FVI. On the opposite side, less than

[1]We chose the EM algorithm, once we do not know neither how many clusters do we have in our data nor the possible characteristics of each cluster. These issues weaken the use of a priori methods like k-means, once they need a predetermined number of clusters.

Table 1 Clusters' centroids obtained by EM algorithm, MRSP (2010)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Have sewers	66.3	21.1	41.6	26.9	50.2	41.1	55.3	37.8	45.6	49.7
Afforestation	94.3	40.2	65.8	67.8	79.3	64.6	78.7	34.2	72.4	81.5
Sewage	0.7	45.1	17.8	9.6	4.9	4.8	7.5	9.0	1.9	0.9
Garbage	1.1	3.2	18.2	3.8	5.7	5.5	7.9	4.8	3.2	1.9
SV	15.1	20.2	19.3	20.0	19.0	19.7	17.4	19.4	18.6	17.3
KA	22.4	63.7	57.5	62.3	54.2	58.7	42.4	58.7	48.9	37.8
WA	8.5	22.1	18.8	21.9	17.7	19.2	13.3	19.2	15.2	11.8
RS	6.5	17.5	13.6	16.4	12.1	15.4	10.1	14.0	10.6	7.5
YD	3.6	6.8	6.9	8.3	6.1	7.6	5.7	6.3	6.2	4.6
HC	6.2	17.5	14.1	16.3	12.5	9.5	9.2	15.0	9.2	9.0
Instances	67	15	63	38	100	73	46	29	99	103
Percentage	10.6%	2.4%	10.0%	6.0%	15.8%	11.5%	7.3%	4.6%	15.6%	16.3%

Source: IBGE—Brazilian Demographic Census, 2010. Data processed and tabulated by the authors



Source: IBGE - Brazilian Demographic Census, 2010

Fig. 1 Spatial distribution of clusters by weighting areas, MRSP (2010)

3% belongs to Cluster 2, which has the worst results for the vulnerability index dimension combined to lower degrees of afforestation and sewers availability. In a similar way, we can interpret all the ten clusters, in order to observe differences concerning those areas.

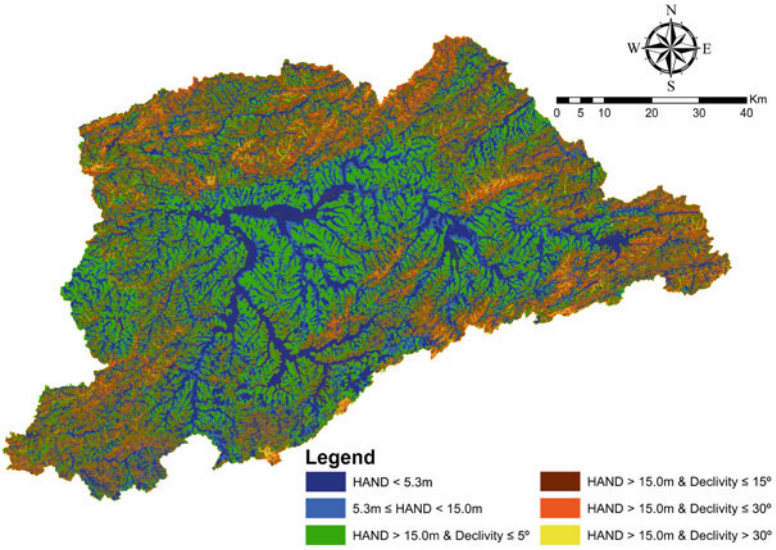
However, we still do not know where each cluster is located. Moreover, these results do not give us a broad and comprehensive analysis about the relations between: (1) the membership of each weighting area to a cluster and the general cluster characteristics; and (2) the spatial location of each cluster and its relation to environmental risk areas. In other words, the simple interpretation of a cluster analysis output is not sufficient in order to understand the relation between sociodemographic characteristics and spatial location, as well as their nexus with environmental risks. Thus, it is clear the importance of overlapping sociodemographic data—analyzed through a data-driven approach (as expectation-maximization)—to data collected from other sources.

First of all, Fig. 1 helps us to distinguish each weighting area in terms of cluster association. Cluster 1 is WA in the center of São Paulo, corresponding to a more developed area, both social and economical, while Clusters 2, 4, and 8 are located in peripheral areas. In this sense, we observe that more vulnerable (in social terms) is distributed far from the core of the MRSP. Main differences among those three clusters are related to the environment: Clusters 2 and 8 have the worst degree of afforestation, as Cluster 4 has a less developed sewers' system with a bit more of green areas. So, even if we observe some mathematical similarities in terms of clusters' centroids, the spatial distribution is an important issue to be assessed.

The last step consists in overlapping the results of cluster analysis with available data about environmental risks, with the aim to compare census-based big microdata with results from other sources. An example of data comes from the project "Brazilian megacities vulnerability to climatic changes: Metropolitan Region of São Paulo," developed by the National Institute for Space Research jointly to three state university: UNICAMP, USP, and UNESP (see [24]). This project gathered data about two kinds of environmental risks: flooding and landslides. Figure 2 is one of the maps produced by the project, where we can observe the spatial distribution of risk areas, considering proximity to water courses and declivity.

The combination of Figs. 1 and 2 suggests relations between socioeconomic vulnerability and environmental risks. For example, weighting areas in Cluster 8 (low afforestation and sewers availability) tend to have a higher declivity (between 15° and 30°), while areas in Cluster 2 (higher vulnerability associated with a very low occurrence of sewers and a high rate of open-air sewage) are located in areas with HAND values below 5.3m (indicating a higher risk of inundation). Curiously, less vulnerable areas are also subject to environmental risks: Clusters 1 (São Paulo core, with very good environment and socioeconomic indexes) and 10 (around this core, also with good indexes) are nearby flooding risk areas.

With this example, we observed that, starting with a considerable amount of unidentified big microdata (in the MRSP, a total of 1,216,611 cases), we can analyze

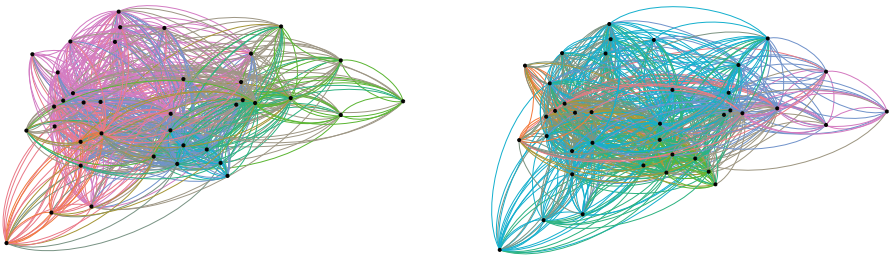


Source: HAND Project - INPE & IPT. Project "*Vulnerabilidade das megacidades brasileiras às mudanças climáticas: Região Metropolitana de São Paulo*" - INPE & UNICAMP
Note: Data for HAND model and other variables are free available for download at the project site: www.ccst.inpe.br/projeto/megacidades/sao_paulo/banco_dados/lista_dados.php.

Fig. 2 HAND (Height Above Nearest Drainage) model for the MRSP

patterns of vulnerability and their relations to environmental risks. This type of analysis can be extended to the whole set of Brazilian big microdata (around 22 million cases for the 2010 census). A possible deployment may combine these results in order to redefine clusters, using geo-referenced data for demographic variables and/or density-based model for cluster analysis. Moreover, it can be implemented through a multi-scalar model, combining data here presented with other sources, to answer questions like: "Why more vulnerable people, from a sociodemographic point-of-view, tend also to be more exposed to environmental risks?".

Our second example to enhance the use of big microdata by applying Digital Humanities concerns the discovery of spatial patterns in internal migration. Considering the same research area of the first example, a possible question would be: "What are the migration patterns within a metropolitan region?". Limiting our analyses to migration flows occurred in the MRSP, our approach can consider two dimensions: (1) the structure of the migration network; and (2) the relevance of each node of the structure. In this sense, we can combine a spatial analysis (more visual) with network and nodes measures (more technical), both using Gephi® 0.92. Doing this for two years of census data (e.g., 2000 and 2010), we can compare the



(a) Flows 1995-2000 (841 connections) (b) Flows 2005-2010 (816 connections)

Source: IBGE - Brazilian Demographic Censuses, 2000-2010
Note: Colors' graphs are independents

Fig. 3 Evolution of intrametropolitan migration network in MRSP (2000–2010)

dynamics of intrametropolitan migration, and draw more complex, wider/broader questions.

In order to explore this issue, we can analyze the individual dataset of census data (using the sample questionnaire), selecting only those who migrated within the 5 years prior to census date. Given that the MRSP has 39 municipalities, and excluding cases of intra-municipal displacement, we have a set of $n \times (n - 1)$ possible connections (i.e., 1.482 combinations), where each municipality can have 38 ingoing and 38 outgoing transactions. The migration network can be observed in Fig. 3, where each node is geographically referenced using city hall coordinates. Visually, both years have a dense network, representing an intense intrametropolitan migration dynamic. We can observe that peripheral nodes have less linkages, while the core is more dense and dynamic.

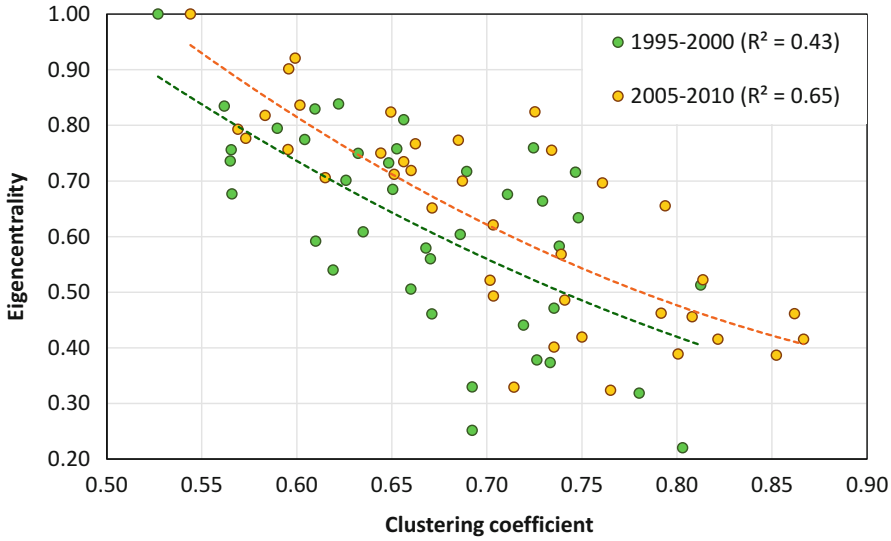
This is a first (visual and simple) approach in order to know more about the intrametropolitan dynamic of migration in MRSP. The colors of each graph show us that there are no more than 4 communities, so we have a clue about how the municipalities are divided into groups. In order to deeply understand the existent relations between the 39 municipalities, we can compute some general metrics about the network that are presented in Table 2. Between 1995–2000 and 2005–2010, though the average number of flows was stable, migratory volumes reduced in approximately three thousand people.

In the last decade, occurred a slight reduction in the density of this network and an increment in its modularity (i.e., it became a bit more sophisticated). Moreover, analyzing the levels rather than the trends, we can perceive that the graph density (i.e., how many combinations are made, out of the maximum possible) is stable, but not so high, and the modularity is very low. Combining these measures, we can deduct that the network is not so sophisticated, with a high number of flows with few migrants and few expressive linkages.

Table 2 Evolution of main network metrics for intrametropolitan migration in MRSP

Measures	1995–2000	2005–2010
Average degree	21.56	20.92
Average weighted degree	15,023	12,067
Network diameter	2	2
Graph density	0.57	0.55
Modularity	0.12	0.15
Number of communities	4	4
Average path length	1.43	1.45

Source: IBGE—Brazilian Demographic Censuses, 2000–2010
Note: For measures, see Wasserman and Faust [36] and Newman [23]



Source: IBGE - Brazilian Demographic Censuses, 2000-2010

Fig. 4 Relations between eigencentrality and clustering coefficient for the migration flows within MRSP (1995–2000, 2005–2010)

We can complement our analysis with metrics related to the nodes of the network; specifically, we can analyze the clustering and the eigencentrality coefficients. The first one evaluates node A, that is, (for example) linked to node B, in terms of the number of connections established by node B with all its possible neighbors. The second one measures the importance of a node, in terms of network construction: the higher the eigencentrality, the more relevant is the node for the whole network. With these two measures—jointly evaluated in Fig. 4—we can evaluate the network internal structure and its evolution.

Figure 4 shows that there is an inverse relation between these measures: municipalities with higher eigencentrality are less prone to have a clustered network. However, it appears that, from the 1990s to the 2000s, the average eigencentrality and clustering coefficients increased and reduced their variances. Combining this knowledge with the previous information about density and degree reductions, we infer that: (1) people are migrating or outside the capital of MRSP (São Paulo), or they are moving outward the boundaries of the MRSP; (2) we should include the whole Brazilian territory in order to understand migration; and (3) the network is becoming (slightly and slowly) more complex, even with a loss of migrants.

With this second example, we observed that, even combining cases that can explode the volume of data (e.g., with only 39 municipalities, we could have had 1482 flows), network analysis is capable to treat these data. Like the first example, we can extend the analysis to the whole range of available census-based microdata: in this case, we could have more than 30 millions of combinations (even if, discarding cases with no occurrences, we had more or less 300,000 unique flows). The expansion of the data range can also bring up questions about the evolution of internal migration, in a continental country like Brazil, that can be answered with Digital Humanities too, employing techniques such as network analysis and machine learning.

A possible deployment of this analysis may be done using the weighting area instead of the municipalities. This kind of analysis would be very useful in order to observe intra-urban flows, once it would make possible analyses about local migration sources and targets. Moreover, if one adds sociodemographic characteristics (or, e.g., a vulnerability index like in Fig. 1) to the analyses, a more complex study about vulnerability and migration could be accomplished. However, Brazilian census data, as they are collected and available nowadays, do not allow us to analyze migration at an intra-municipal level at both origin and destination. A possible path which can be followed is to analyze only destinations by weighting area, once we know where each person that answered the sample questionnaire currently lives, but we do not know, for the same spatial unit, where the person lived before.

4 Potentialities and New Horizons for Demography

The main feature of the approaches presented in this chapter is the ability to operate with very large census-based datasets (currently known as “big microdata”), which makes general (national) and disaggregated (by municipalities) studies feasible, in a potentially large time span. The treatment of uncountable variables (in order to evaluate issues like environment, vulnerability, and migration) by municipalities and, why not, for smaller areas (like weighting areas and censitary sectors) enables demographers to reassess old questions and investigate new themes. Moreover, big microdata, treated within the scope of Digital Humanities, can be used as a complementary source for researches based on other surveys.

On the one hand, there is a huge potential of reviewing consolidated theories using data-driven analyses (like machine learning, multivariate, and network analysis). Demographers of all specialties can come back to old censuses and investigate the same questions of years ago, starting from a new perspective; this can bring up reevaluations concerning what we know about population. On the other hand, actual computing power allows us to expand boundaries of knowledge: researches no longer need to be limited to spatial clippings that misidentify idiosyncrasies. In this sense, demographic researches can be more comprehensive, including more variables, spaces, and times, in order to deeper and broader understand sociodemographic and related phenomena.

Census-based data are usually presented on a political-administrative basis, while environmental data are disclosed in accordance with physical-geographical unities. The use of these divisions does not adequately fit to the representation of phenomena analyzed in Demography. Political-administrative and operational units are subject to vary in size and shape from one region and/or year to another. Thus, it is almost obligatory to spatially abstract in order to obtain more accurate analyses. Even though the examples here presented, as way of illustration, analyze the MRSP (one of the 10 world's biggest metropolitan regions), we suggest that, in order to adequately employ Digital Humanities on big microdata, researchers should not: (1) start with an excessively specific spatial clipping; and (2) define too wide spatial units of analysis.

Another front of deployment of demographic researches with big (micro)data and Digital Humanities—already in progress in some countries—is the development of administrative records in a more robust basis. Different from census data, those one are collected all the time, for all the individuals that live in a specific region, including a wider range of themes (not only demographic ones). Development in the collection of those data, besides the richness of the data itself and their time span, will not only increase sociodemographic, environmental, and economic researches (as well as cross-thematic analyses). These data will be a complementary source to understand census data and, in the long run, they may be a potential substitute for demographic censuses as we know them, allowing researches with constantly up-to-date information concerning population issues.

The solution to deal with so much (big micro)data is to adopt the tools provided by Digital Humanities in order to ensure new (but not only) spatial analyses, as well as panel studies (that may help in the development of causality researches). The purpose of this combination is to promote data-driven analyses that reveal spatial and social relations that, otherwise, would be hidden by data structure itself. Moreover, this combination may allow, in the near future, multi-scalar, multidimensional, and multilevel researches, unbiased by administrative and/or political boundaries, using large volume of relevant data, coming from both censuses and administrative records. Approaching demographic issues by the perspective of Digital Humanities will better "prepare" researchers to the data avalanche of big (micro)data, reshaping our comprehension about an uncountable number of social themes and reassessing relations in areas like P&E.

References

1. Billari, F., Zagheni, E.: Big data and population processes: a revolution? In: Petrucci, A., Verde, R. (eds.) Statistics and Data Science: New Challenges, New Generations, pp. 167–178. Firenze University Press, Firenze (2017)
2. Billari, F., D'Amuri, F., Marcucci, J.: Forecasting births using Google. In: *Proceedings of the 78th Annual Meeting of the Population Association of America*, New Orleans, USA (2013)
3. Burdick, A., Drucker, J., Lunenfeld, P.: et al.: Digital Humanities. Cambridge, MA: MIT Press (2012)
4. Cesare, N., Lee, H., McCormick, T., et al.: Promises and pitfalls of using digital traces for demographic research (2016). https://ssrn.com/abstract=2839585
5. De Mauro, A., Greco, M., Grimaldi, M.: What is big data? A consensual definition and a review of key research topics. In: *AIP Conference Proceedings*, vol. 1644, pp. 97–104. AIP Publishing, College Park (2015)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Brown J. World Aff. **39**(1), 1–38 (1977)
7. Deville, P., Linard, C., Martin, S., et al.: Dynamic population mapping using mobile phone data. Proc. Natl. Acad. Sci. **111**(45), 15888–15893 (2014)
8. Foster, I., Ghani, R., Jarim, R.S., et al.: Big Data and Social Science: A Practical Guide to Methods and Tools. Chapman & Hall, London (2016)
9. Friedman, J.H.: The role of Statistics in the data revolution? Int. Stat. Rev. **69**(1), 5–10 (2001)
10. Furtado, B.: Índice de Vulnerabilidade das Famílias (2000–2010): Resultados. Texto para Discussão 1835. Instituto de Pesquisa Econômica Aplicada, Brasília (2013)
11. Gołata, E.: Shift in methodology and population census quality. Stat. Transition **17**(4), 631–658 (2016)
12. González-Bailón, S.: Social science in the era of big data. Policy Internet **5**(2), 147–160 (2013)
13. Greenwood, M.J.: Migration and labor market opportunities. In: Fischer, M.M., Nijkamp, P. (eds.) Handbook of Regional Science, pp. 3–16. Springer, Berlin (2014)
14. Hakkert, R.: Fontes de Dados Demográficos. ABEP, Belo Horizonte (1996)
15. Halevi, G., Moed, H.: The evolution of big data as a research and scientific topic: Overview of the literature. Res. Trends **30**, 3–6 (2012)
16. King, G.: Ensuring the data-rich future of the social sciences. Science **331**, 718–721 (2011)
17. Kitchin, R.: Big data, new epistemologies and paradigm shifts. Big Data Soc. **1**(1), 1–12 (2014)
18. Letouzé, E., Jütting, J.: Official statistics, big data and human development. In: Data-Pop Alliance White Paper Series. Data-Pop Alliance, New York (2015)
19. Lin, J.P.: Data science as a foundation of integrating and enriching administrative data. In: Proceedings of the 9th International Conference on Social Science Methodology. Leicester, UK (2016)
20. McCaa, R.: The big census data revolution: IPUMS-international. Trans-border access to decades of census samples for three-fourths of the world and more. Rev. Demogr. Hist. **30**(1), 69–88 (2012)
21. McCaa, R., Ruggles, S.: The census in global perspective and the coming microdata revolution. Scand. Popul. Stud. **13**, 7–30 (2002)
22. McPherson, T.: Introduction: media studies and the digital humanities. Cinema J. **48**(2), 119–123 (2009)
23. Newman, M.E.: Network: An Introduction. Oxford University Press, Oxford (2010)
24. Nobre, C.A., Young, A.F.: Vulnerabilidades das megacidades brasileiras às mudanças climáticas: Região Metropolitana de São Paulo. Final Report. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (2011)
25. Okajima, I., Tanaka, S., Terada, M., et al.: Supporting grown in society and industry using statistical data from mobile terminal networks? Overview of mobile spatial statistics. NTT DOCOMO Tech. J. **14**(3) (2013)

26. Oliveira, L.A.P., Simões, C.C.S.: O IBGE e as pesquisas populacionais. Rev. Bras. Estud. Popul. **22**(2), 291–302 (2005)
27. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. Big Data **1**(1), 51–59 (2013)
28. Ruggles, S.: Big microdata for population research. Demogr. **51**(1), 287–297 (2014)
29. Spiro, L.: Defining digital social sciences (2014). http://acrl.ala.org/dh/
30. Stuart, E., Samman, E., Avis, W., Berliner, T.: The data revolution: finding the missing millions. Research report 03. Overseas Development Institute, London (2015)
31. Szmrecsányi, T.: Da aritmética política à demografia como ciência. Rev. Bras. Estud. Popul. **16**(1/2), 3–17 (1999)
32. United Nations: Principles and Recommendations for Population and Housing Censuses. United Nations Statistical Office, Department of International Social and Economic Affairs, New York (1980)
33. University of Cambridge: Big data Methods for Social Science and Policy. Cambridge Public Policy SRI, Cambridge, UK (2016). www.publicpolicy.cam.ac.uk/pdf/ppsri-big-data-report
34. Varian, H.: Big Data: new tricks for Econometrics. J. Econ. Perspect. **28**(2), 3–28 (2014)
35. Ward, J.S., Barker, A.: Undefined by data: a survey of big data definitions (2013). https://arxiv.org/pdf/1309.5821v1
36. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge Univeristy Press, Cambridge (1994)
37. Weber, I., State, B.: Digital demography. In: Proceedings of the 26th International World Wide Web Conference. Perth, Australia (2017)
38. Zagheni, E., Weber, I.: You are where you e-mail: using e-mail data to estimate international migration rates. In: Proceedings of the 4th Annual ACM Web Science Conference. Evanston, USA (2012)

Modeling Social and Geopolitical Disasters as Extreme Events: A Case Study Considering the Complex Dynamics of International Armed Conflicts

Reinaldo Roberto Rosa, Joshi Neelakshi, Gabriel Augusto L. L. Pinheiro, Paulo Henrique Barchi, and Elcio Hideiti Shiguemori

1 Introduction

In this chapter we address one of the great causes of a man-made disaster that can reach planetary proportions. These are the *armed conflicts*, which throughout the history of humanity have been the main cause of the great wars. A great war in general involves a great destruction of social, cultural, and artistic heritages and, in addition, a huge number of victims (e.g., the 2nd World War resulted in over 60 million deaths) [36]. In this way, great wars are *extreme events* because they induce all those disasters that are inherent to armed conflicts which can reach geopolitical proportions [3, 24] (See Appendix 1).

In a more general context than the one identified above, careful study involving causes and effects of extreme events that result in disasters (natural or man-made) is of paramount importance not only for society in general, but specifically for many areas of science and technology. This is due to the fact that it is mainly through science and technology that humanity will 1 day be able to predict, mitigate, and even control or avoid the causes and effects of major disasters. Therefore, a pragmatic study of armed conflicts, considering the great wars, must necessarily involve a multidisciplinary approach with emphasis on at least the following fundamental sciences: geopolitics, sociology, psychology, semiotics, mathematics,

R. R. Rosa (✉)
Lab for Computing and Applied Mathematics-INPE, São José dos Campos, SP, Brazil
e-mail: reinaldo.rosa@inpe.br

J. Neelakshi · G. A. L. L. Pinheiro · P. H. Barchi
CAP-INPE, São José dos Campos, SP, Brazil
e-mail: gabriel.pinheiro@inpe.br; paulo.barchi@inpe.br

E. H. Shiguemori
Geo-intelligence Division, Institute of Advanced Studies (IEAV), Rov. Tamoios, São José dos Campos, SP, Brazil

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6_12

physics, and chemistry. However, for the modeling of extreme events, especially within the context of social sciences, it is necessary to bring to the detailed study of the phenomenon the modern approaches of computational mathematics, statistical physics, and complex systems theory [27].

The concept of a complex system has evolved significantly over the past two decades. There are many approaches in physics, biology, engineering, mathematics, computer science, and in some interdisciplinary areas that mention, consider, and define a *complex system* [9, 38], however without rigorously presenting an approach that explicitly includes *extreme events* as a property of *complexity* [61, 64], and, even more, addressing its causes in terms of endogenous, exogenous, or hybrid nature [55]. That is, a concept that allows, at least partially, to identify and even measure the degree of complexity based on processes that manifest crises, extreme events (as shocks and disasters), and systemic phase transitions [65].

As will be seen throughout this chapter, complex systems far from the thermodynamic equilibrium [46, 60] can be studied effectively through the non-homogeneous energy transfer cascade model which is the main process describing turbulent-like behavior [33, 47]. Its adequacy to the problem of armed conflicts in a complex network allows to characterize and measure complexity in the so-called *sociosphere* [40], a systemic concept already addressed by the *general theory of systems* (GTS) [14, 60]. From the GTS the concept of sociosphere (a modern concept of biosphere) is presented from the complex interaction that occurs between the human being and the planet, giving rise to a region where conditions are such that our planet is theoretically capable of sustaining life.

The Earth ecosphere, where the interaction between the living and nonliving components takes place, is shaped and transformed by human activity. Thus, the human being merges nature into society. More precisely, the sociosphere can be considered as an inter-subjective ecosphere around our planet, defining the society-nature system [22]. The term comprises the way of social mediation among *agents* (human or social groups) takes place as economical and geopolitical processes.

The fact that such agents determine interactions that may be cooperative or conflicting, balanced or out of equilibrium, permanent, or fleeting in relation to the environment are some of the aspects that make the sociosphere a complex system with the following properties: expressive number and diversity of agents, direct and indirect connections, linear and nonlinear interactions, structurality, hierarchy, and collective phenomena that can strengthen or weaken the permanence of the system as a whole. In this scope a widely used concept for complex systems is as follows: complex systems are those consisting of many different agents interacting nonlinearly and it cannot be split into simpler subsystems without neglecting its collective properties [9, 38]. Although complex systems as a whole are difficult to model, the characterization of spatiotemporal complexity by scaling laws is highly attractive and has many applications in the study of complex dynamics from many systemic interacting elements [35, 61, 64].

In the next section we will describe spatiotemporal complexity based on the concept of scaling laws when associated with a complex dynamics that can be characterized by a multiplicative energy transfer cascade. In the rest of the text we

will use this description as a basis for modeling the dynamics of endogenous and exogenous armed conflicts.

2 Modeling Social Spatiotemporal Complexity

Formal complex systems are entities composed of well-defined components. The integrability of the components acts together as to form a functioning whole with spatiotemporal collective dynamics and responses to the environment. Some properties that distinct complex systems have in common are the following: (i) *high density*, (ii) *diversity*, (iii) *nonlinear connectivity*, (iv) *collective behavior*, (v) *structurality*, (vi) *thermodynamic permanence*, (vii) *far from equilibrium states*, and (viii) *scaling hierarchy* [53, 60]. As a practical example of complex systems exhibiting such properties, in the spatiotemporal domain, are: (a) the bird flocks [15], (b) the fish schools phenomena [50], and (c) the social crowds [23, 42]. In these high density complex systems, even when the diversity is low (but never null), their nonlinear interconnected nature leads to emergent behavior. From the property of **collective behavior** emerges **structurality** whose **permanence** is susceptible both to endogenous and exogenous stimuli. The structure is then dissipative and its stability can be characterized from the variation of its thermodynamic characteristics. This means that complex systems can trip across thresholds into sudden transitions and they can react disproportionately to seemingly small triggers, or transform as a result of influences from within the **scaling hierarchy** structure itself (e.g. hysteresis and self-organization).

In general, the complex exchange of non-homogeneous information across group agents may result in a spectrum of spatiotemporal patterns characterizing collective behaviors. While complex collective patterns in animal groups are readily identifiable by trained visual inspection, a mathematical method from raw stochastic data is not yet well established. However, considering the previous definition of complex systems, the possibility of measurement stochastic data brings information on the scaling hierarchy and its autocorrelation function. This reduction of the effective dimensionality of a complex system makes it possible to establish a mathematical approach known as Kramers–Moyal (KM) treatment [10, 23].

The KM approach is based on the hypothesis that there are several large Markov–Einstein time scales, which can be traced back to hydrodynamic-like memory effects. Above this time scale spectra the process can be mapped to an Ornstein–Uhlenbeck process [21] via the application of the developed extended structure function method [23]. This is in agreement with the classical theory of overdamped $1/f^{\beta}$ noise which characterizes, for example, Brownian motion, reaction–diffusion pattern formation, and turbulence, from its power spectrum density (PSD). In this context, complex dynamics such as non-homogeneous turbulence can be described from multifractal singularity spectra [12, 33, 43, 63].

Between the two example of systems, uniform [15, 50] and highly non-uniform [42], we identify the systems of social conflict that can be approximated by elements

containing only two distinct types: agents and reagents that, in practice, exchange their roles collectively. This allows us to initiate a first modeling that can be effective in the study of armed conflicts. Therefore, understanding the transfer energy cascade following a scaling hierarchy from an agent–reagent system and the correspondent non-homogeneous turbulent-like behavior becomes also essential for many social complex systems [6, 7, 28, 42].

2.1 Multiplicative Cascade for Extreme Conflicts Events

When an effective observable, expressing the energy of the system, exceeds a certain critical value, a sort of viscous character originating from agent–reagent sociological frictions may appear. Moreover in such a scaling hierarchy the exchange of information is limited only within a local region so that the transfer of the energy in a group may become incoherent. Therefore, in a systemic approach, the complex regimes of a social dynamical system may be similar to the turbulent motion of a complex viscous fluid. Along this line it is possible to propose an analytic method to characterize complexity from a systemic stochastic fluctuation which should be sensitive, in the time domain, to the occurrence of extreme events. A such time series can then be analyzed from their respective power spectrum density (PSD) [10, 21, 33] (see Appendix 2).

To simplify the notation, the elements of our agent–reagent multiplicative cascade are called simply *agents*. The diversity of agents is binomial. We use the colors black and white. When the colors are opposite the agents are in conflict (see Fig. 1a). We start the cascade from a non-homogeneous criterion containing a conflicting pair ($A_1 \times A_2$). The energy (or information) that characterizes the conflict is distributed between the agents and is interpreted as a function of a wave number $K = 1/scale$. In the cascade, as wave number falls over time, the amount of agents increases in a binary proportion. Agents of lower levels are called $pro - agents$ (PA) (they allies when they have the same color). The energy of the conflict will dissipate until a *disaster state* in which the maximum energy dissipation will occur. This scenario is compatible with a non-homogeneous cascade type model (see Fig. 1b). Hence, our approach is based on the so-called *p-model* mechanism [37, 56] that allows us to create a non-homogeneous cascade in a way which is entirely compatible with the fluctuations observed in stochastic time series.

2.2 P-Model for Social Conflicts

The P-model approach for non-homogenous turbulent-like cascade was proposed by Meneveau and Sreenivasan [37]. It gives a new insight into the kinetic energy dissipation in cascading process of eddies in the inertial range of a fully developed turbulence, and it is based on the special case of weighted transfer.

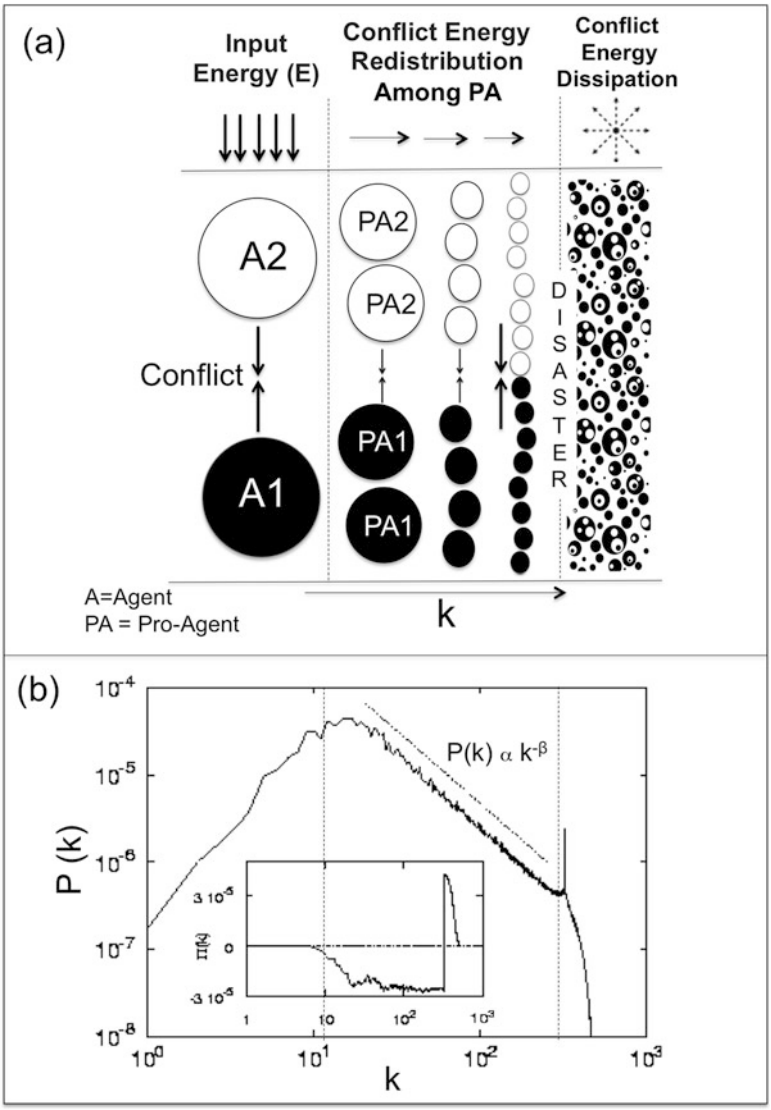


Fig. 1 The multiplicative cascade for non-homogeneous binomial turbulent-like process. (**a**) The scaling hierarchy for social conflicts; (**b**) the respective expected power spectrum density pattern showing the transition from the inertial range to the extreme event which is the response to the high dissipative regime. The embedded picture shows the enstrophy as a function of the wave number

From a theoretical point of view, P-model is a generalized form of two-scale cantor set with balanced distribution of length which shows multifractal properties of one-dimensional sections of the dissipation field. The generalized form starts from the classical view of eddy cascade before the inertial range of fully developed

turbulence, where flux of energy (E_K) actually dissipates at Kolmogorov length scale β from eddies of size L_K. Then, each eddy of size L_K is divided into two equal parts, $L_K/2$ expressed as L_{K_1} and L_{K_1} each; however, in each cascade step, the flux of energy is distributed, as a probability, unequally in fraction of p_1 and $p_2 = 1 - p_1$ where $p_1 + p_2 = 1$. This process is iterated over fixed p_1 until eddy reaches to Kolmogorov scale β [5, 31, 41].

The *p-model* multiplicative cascade is given by

$$\alpha = \frac{log_2\, p_1 \; + \; (\omega - 1)\, log_2\, p_2}{log_2\, l_1 \; + \; (\omega - 1)\, log_2\, l_2} \tag{1}$$

and

$$f(\alpha) = \frac{(\omega - 1)\, log_2\, (\omega - 1) \; - \; \omega\, log_2\, \omega}{log_2\, l_1 \; + \; (\omega - 1)\, log_2\, l_2} \tag{2}$$

Starting with a non-homogeneous energy distribution, one transfers a fraction $f(\alpha)$ of the *multifractal mass* from one half to the other in a randomly chosen direction [58]. This is equivalent to multiplying the originally uniform density field on either side by factors. The same procedure is repeated M times, recursively at ever smaller scales using fractions varying α on segments of length $L/2^n$, where the multiplicative weight ω is parameterized as $1 - (1 - 2p)$, resulting the discrete array $C(m)$ where m counts as time steps. This p-model algorithm procedure given by Venema [1, 16] can produce time series where the variance is finite if you would extrapolate its power spectrum to infinite large scales [49].

Time series, $C(m = t)$, with $M = 2^{11}$, representing non-homogeneous turbulent conflict is generated using the Venema algorithm [1], where the inputs are: the size of the time series in number of points (M), the PSD power spectrum (β_{PSD}), and the value of p, which is the fractional distribution of energy in non-homogeneous turbulent-like cascade [4, 18, 20, 26, 34]. The homogeneous dissipative process near the thermodynamic equilibrium is recovered when $(\beta_{PSD},\, p) = (-5/3, 0.5)$.

Deviations from the homogeneous cascade are compounded by abrupt changes in the frequency and magnitude of social conflict. Such changes are called *eXtreme Events* (XE) and their cause may be due to factors more internal than external. When the level of conflict increases significantly due to internal interactions, the extreme event is called endogenous (XE_{endo}). When the external energy transfer or abrupt dissipation is the main cause of XE, then it is called an extreme exogenous event (XE_{exo}). In the power law domain, events of the type XE_{endo} and XE_{exo} belong to different classes of universality. In order to generate time series XE_{endo} and XE_{exo} we will adopt the SDGA formalism [55] (see Appendix 2). Typical endogenous and exogenous processes, combining p-model and SDGA, are obtained for $(\beta_{PSD} \approx -0.4,\, p = 0.25)$ and $(\beta_{PSD} \approx -0.7,\, p = 0.25)$, respectively. Figure 2 shows XE time series for different combinations of p and β_{PSD}.

The cumulative energy of the process in the time domain is defined as normalized average $< C(\tau) >$ where τ is a chosen window time interval along the signal. The

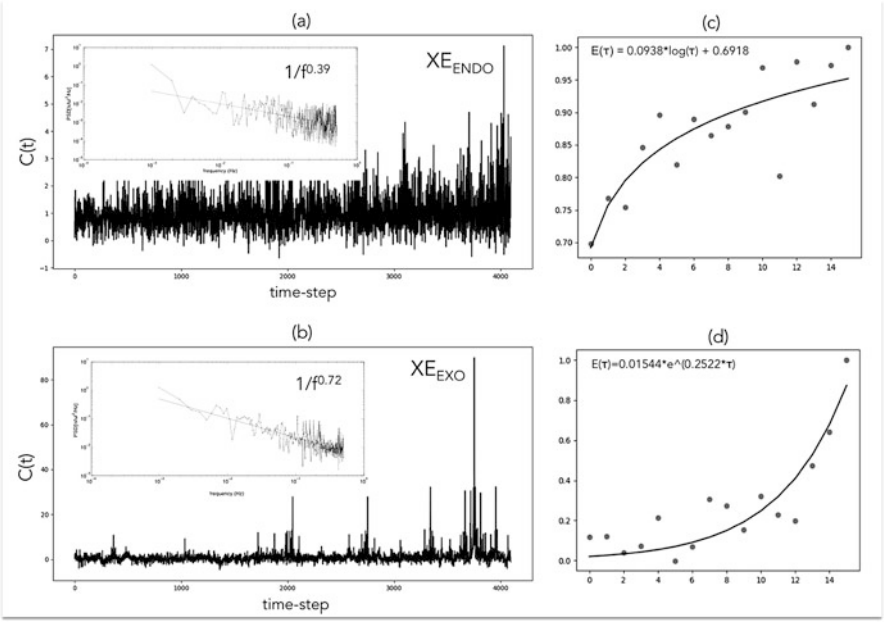


Fig. 2 Two typical time series simulated from p-model fixing $p = 0.25$ and varying the value of β. (**a**) A XE_{endo} pattern from $\beta = -0.39$. (**b**) A XE_{exo} pattern from $\beta = -0.72$. The respective PSDs are shown using embedded smaller pictures. Figure (**c**) and (**d**) shows the respective cumulative energy expressing the typical functional patterns for endogeny (log) and exogeny (exponential)

trend of the cumulative energy over each $\tau = 250$ time steps is shown in Fig. 5c and d. It is noteworthy that the typical trends are nonlinear being logarithmic for XE_{endo} and exponential for XE_{exo}.

3 Modeling Geopolitical Conflicts

Armed conflicts are, at some scale, associated with geopolitical factors (in their broadest meaning, which involves aspects of rights and security directly related to citizenship, culture, economics, and politics). In this way, as shown in Appendix 1, armed conflicts are those involving threats, defenses, and attacks with the active or passive use of weapons in a geopolitical domain.

In the geopolitical domain, the terrorist attacks are today one of the central problems in our social environment [62]. It is notable in this context that the rate per decade of international armed conflicts has increased more than 85% since the severe event on September 11, 2011 [24, 36]. Since then terrorist attacks have also been associated with other types of armed conflict involving geopolitical and

economic interests where small and large nations stand out as strategic parts of the complex and global armed sociological conflict (for simplicity, only international armed conflicts). Therefore, in order to formalize the study of the dynamics of international armed conflicts, the so-called Uppsala Conflict Data Program *(UCDP)* has been prepared since 2002 in close collaboration with researchers at the Department of Peace and Conflict Research at Uppsala University and the Departments of Sociology and Political Science and Geomatics at the Norwegian University of Science and Technology *(NTNU)* [24]. We will call here Armed Conflicts *(AC)* only to those defined in the *UCDP/PRIO* Dataset Codebook [57].

The number of AC as terrorist attacks worldwide by year is shown in Fig. 3 [57, 59]. The data reveal an endogeny in the frequency of attacks from 1970 until 1992. A period of decline is evident beginning in the wake of the Cold War's end and lasting roughly a dozen years [54]. For the past decade, however, there has been a exogenous rise in the number of terrorist attacks from just over a thousand in 2004 to almost 17,000 in 2014. The trends indicate some meaningful distinctions before and after the 9/11 Al-Qaeda attack. Since 9/11, these countries have experienced significantly more armed conflicts than they had previously [54]. It is also noteworthy, from the inserted picture, the hyper exogenous character of the September 11 attack.

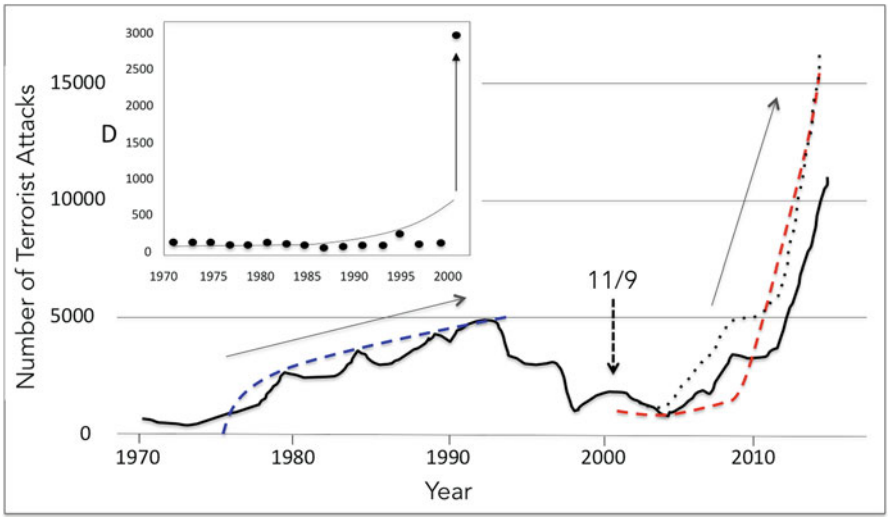
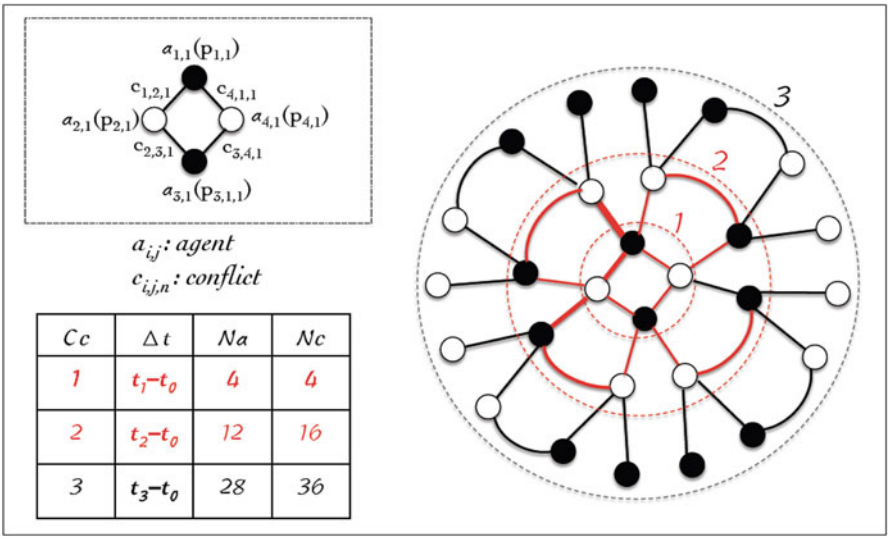


Fig. 3 Endogeny (blue) and exogeny (red) patterns in the total number of conflicts per year from terrorist attacks since 1970 until 2012. The exogenous pattern holds even when removing attacks in Iraq and Afghanistan (dashed curve). The smaller picture shows the GTD total number of deaths (D) per year from terrorist attacks against the USA state from 1970 until the Sep 11th 2001. This includes all victims and attackers who died as a direct result of the incident [59]



Cc	Δt	Na	Nc
1	t_1-t_0	4	4
2	t_2-t_0	12	16
3	t_3-t_0	28	36

Fig. 4 The Agent-Conflict Network, $ACN(SP_3)$ with $n = 3$ containing 28 agents and 36 conflicts. The partial number of agents and conflicts for sub-cells are shown in the inserted table. The embedded small picture on the left shows the details of the SP_3 core configuration

3.1 A Prototype for Agents-Conflict Network (ACN)

In order to illustrate how, in practice, the non-homogeneous p-model is designed for describing Geopolitical Conflicts, we introduce the simplest Agents-Conflict Network (ACN) shown in Fig. 4. We assume that the network has a core with four agents: $a_{1,1}$, $a_{2,1}$, $a_{3,1}$, and $a_{4,1}$. This set of agents is called the first *conflict cell*: C_1. In C_1 the agents $a_{1,1}$ and $a_{2,1}$ are in conflict. Following the cascade model described in Fig. 3, $a_{3,1}$ is pro-agent $a_{1,1}$ while $a_{4,1}$ is pro-agent $a_{2,1}$. Then, $a_{3,1}$ and $a_{4,1}$ are in conflict too. In the simplest configuration each agent will be in conflict with two new agents giving rise to cell C_2 which will contain $N_{C2} = N_{C1} + 2N_{C1}$ agents, where N_{C1} is the number of agents in C_1.

In a war of great proportions not all the agents involved (in general states) are in conflict with all their enemies. Geographical and diplomatic limits impose this restriction. Note that in the ACN prototype this happens naturally due to the connection adopted in the circular topology.

This prototype is referred to here as the *Conflict Core SP3* (SP_3). In this prototype, agent $a_{1,1}$ is always a *superpower state*[1] in conflict with some nation (or faction) with status of state (but not a superpower) that is agent $a_{2,1}$. A convenient initial condition to non-homogeneous cascade is that pro-agents $a_{3,1}$ and $a_{4,1}$ are

[1]The concept is based on the global hegemony introduced by Lyman-Muller [19].

Table 1 Tree pragmatic examples of SP_3 conflict core

Event	Date	$a_{1,1}$	$a_{2,1}$	$a_{3,1}$	$a_{4,1}$	$C(\tau_1 = 15)$	Deaths
NKB	10/09/17	USA	NK	Japan	China	16	0
SMS	4/07/17	USA	Syria	EU	Russia	08	25
WTC	9/11/01	USA	Al-Qaeda	EU	ISIS	06	2996

also necessarily superpowers [39]. Three contemporary examples involving the USA as the agent $a_{1,1}$ are set forth in Table 1: North Korea Border (NKB), Shayrat Missile Strike (SMS), and World trade Center (WTC). In this table, the acronyms EU and ISIS are, respectively, for European Union and Islamic State of Iraq and Syria, and τ_1 is given in days.

The conflict between two agents is denoted as $c_{i,j,n}$, where n denotes the cell level C_n. Each agent $a_{i,j}$ is assigned a probability $p_{i,j}$ which measures its chance of being active in the conflict (that is, $c_{i,j,n} \neq 0$).

The energy injected into the core SP_3 is the source that materializes the conflicts: $c_{1,2,1}$, $c_{2,3,1}$, $c_{3,4,1}$, and $c_{4,1,1}$ for a time interval $\tau_1 = t_1 - t_0$, where t_0 is defined as the initial conflict time. For a given τ_n it is possible to estimate the amplitude $C(\tau_n)$ by counting the number of active and significant conflicts such that $p_{i,j} \geq p_c$, where p_c is a certain critical probability as input to the model.

For the SP_3 an extreme event occurs when $C(\tau_n) \geq N_{C_n} + 4$, where N_{C_n} is the number of **active** agents contained in cell C_n. This criterion imposes on the model the form of distribution of the injected energy. Let's say for SP_3 with $n = 3$ we have all layer conflicts $n = 3$ disabled and all others activated which results then in 16 conflicts. In this case $C(\tau_3) = 12 + 4 = 16$. It is an extreme event where the energy balance did not reach the outermost layer and there is no any extra external energy source for this. Therefore characterizing in this case a typical XE_{endo}. If there were an complementary external energy injection, in the avalanche structure, this enclave would be transferred to the N_3 agents more quickly thus characterizing a more abrupt and more intense event, then a typical case of XE_{exo}.

3.2 The Prototype Simulation

In this section, we introduce a bidimensional cellular automata (CA) to simulate the ACN prototype [2]. The 2DCA-ACN is a two-dimensional cellular space with discrete time step where each cell holds one state from a finite state set that changes following a local update rule which depends on the neighborhood state around each cell at the previous time step [30]. At the core, the two black cells represent, according to the notation used in Fig. 4, the agents $a_{1,1}$ and $a_{2,1}$ and the white ones the respective pro-agents $a_{3,1}$ and $a_{4,1}$. The neighborhoods strategy is a Von Neumann configuration [30] that is composed of the cell itself and the four cells situated at north, south, east, and west. Moreover, the pressure in this conflict space

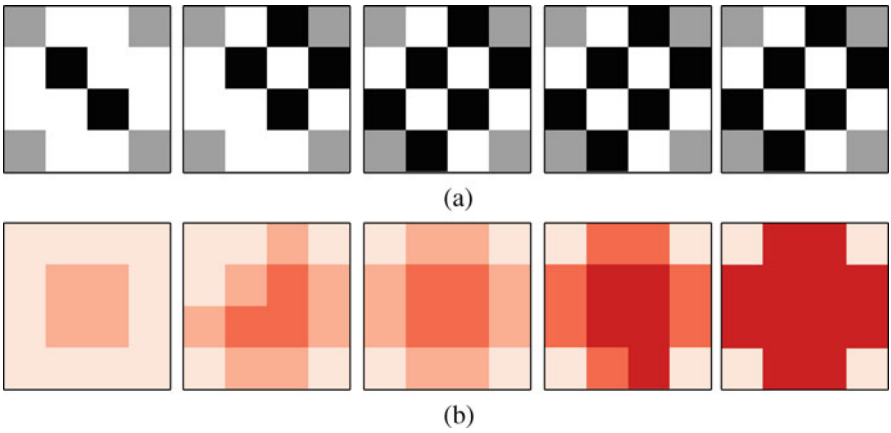


Fig. 5 An example of 2DCA-ACN simulation setting the parameter $P(A) = 0.5$ as the initial value to each activated cells. (**a**) Snapshots of the activated cells. (**b**) Evolution of the space conflict state represented by four red levels at each time step. The corresponding state with the intensity of red, in the ascending order, is given by: threats, strike, and war. The lighter color of the red levels does not draw any state, hence shown an inactive cell and is defined as level 0

only tends to grow. For the simulations to be considered in this article, the AC prototype is formed of only two-stage, that is $C_c = 2$: the initial state which is described as the catalyst iteration of the conflict space, while the second is characterized as a complete system, showing all allies of the two side of the conflict space, with a total agents $N_a = 12$ (the black and white cross-shaped area in Fig. 5) and a maximum number of conflicts of $N_c = 36$. Not that, to be compatible with the ACN topology, as a rule, there is no activation in the gray squares of the cellular space conflict. The time step follows the Von Neumann rule and is thus recurrent until the system converges to the maximum configuration of the second stage.

In this 2DCA-ACN, every cell holds a probability $P(A) \in [0, 1]$ of conflict with each neighborhood cell, a three-up-let state which is defined as (i) threats, (ii) strike, (iii) war and a counter to save the number of conflicts (i, ii, iii) that occurred. Through each state, we set as well an energy value $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ to threats, strike, and war, respectively such that $\mathcal{E}_3 \gg \mathcal{E}_2 > \mathcal{E}_1$. It is assumed that only cells with threats or superior state can influence in the environment, we call this an *active cell*. Otherwise, the cell is inactive and not assumes any value. At the initial time, the configuration of the system is given by activating all cells located in the first stage, setting their state as a threat, and keeping the other cells inactive (the gray cells of the space conflict are always inactive). Furthermore, a cell cannot bring down your state and the geopolitical conflicts only expand if a cell in the first stage turns up the state to the second level.

The chance of happens a conflict (η) in a cell by the joint probability of conflict between the cell (i, j) and your neighbors $(i + \alpha, j + \beta)$ is given by

$$P(\eta) = \prod P(A_{i+\alpha, j+\beta}) \tag{3}$$

where α is the x-axis and β is the y-axis coordinates, which assume the following indices: $\{(0, 0), (0, -1), (-1, 0), (0, 1), (1, 0), (1, 1)\}$. In this case, each cell is independent so the state of the other cells does not influence in your own state. Thereafter, we use a Bernoulli distribution to sample the probability $P(\eta)$ and decide if a conflict (threats, strike, war) get success or failure, in other words, if the conflict will happen or not.

Moreover, if some conflict happened, the counter must be incremented to register this action. In this way, the probability of $P(A)$ must be up to date according to the number of conflict in a cell, which is defined by:

$$P(A) = \frac{P(A) \times F}{z} \tag{4}$$

where F counts for the amount of conflict in a cell and $z = P(A) \times F + P(\overline{A})$ is the normalized probability.

The model was developed with the programming language Lua, a simple pseudo-code of the conflicts procedure is detailed in Algorithm 1. In the procedure, the parameter p represents all activated cells in the system at a specific time t. After receiving the cells, a loop is used to iterate over each neighbor of the current cell p for computing the joint probability (line 4). Worth to emphasize that the update of parameters occurs at the same time in every activated cell and to calculate the new value of counter and $P(A)$ it is only necessary the previous values. Therefore, if x is succeeded, the next step updates both the counter and the probability of conflict (line 10, 11).

Algorithm 1 Cellular automaton to geopolitical conflicts

1: Set parameters;
2: **Input:** $P(A)_p$: initial probability of conflict
3: **procedure** CONFLICT(p)
4: $\eta \leftarrow P(A)_p^{t-1}$;
5: **for** $\forall q \in N(p)$ **do** ▷ Neighbors of the current cell
6: **if** $State_q^{t-1}$ is active **then**
7: $\eta \leftarrow \eta * P(A)_q^{t-1}$;
8: **end if**
9: **end for**
10: **Compute** x $\sim Bernoulli(\eta)$;
11: **Update** $Counter_p^t$ according x;
12: **Update** $P(A)_p^t$ according x;
13: **end procedure**

Figure 5 shows the output for a 2DCA-ACN. The space conflict is shown at time = 10, 60, 70, 80, 160. The initial time is defined as the first stage for the agents: {(1, 1), (1, 2), (2, 1), (2, 2)} and every agent assumes the parameter $P(A) = 0.5$. While Fig. 5a shows the activated cells, Fig. 5b represents the state of each cell by means of four red color levels, such that, the level 1 illustrates the initial conflict state as a *threat* until the level 3 which is the conflict state *war*. Along the evolution of the conflict space a mean energy of the *state* in each cell at time t is registered, where the t is equivalent 15 days passed according to the UCDP data (see Sect. 4 and Appendix 1).

We find that the model is able to simulate both endogenous (XE_{endo}) and exogenous behavior (XE_{exo}) according to a specific initial value in the parameter probability of conflict $P(A)$. Figure 6 (Up) illustrates an endogenous case by initializing $P(A)$ with random probability for every cell activated at some time t. In general, the simulations showed that one can obtain similar results as well when $P(A)$ is defining, in average, with low probability. On the other hand, if $P(A)$ admit high probabilities ≈ 1, then an exogenous pattern XE_{exo} occurs in the state conflict space as shown in Fig. 6 (Bottom).

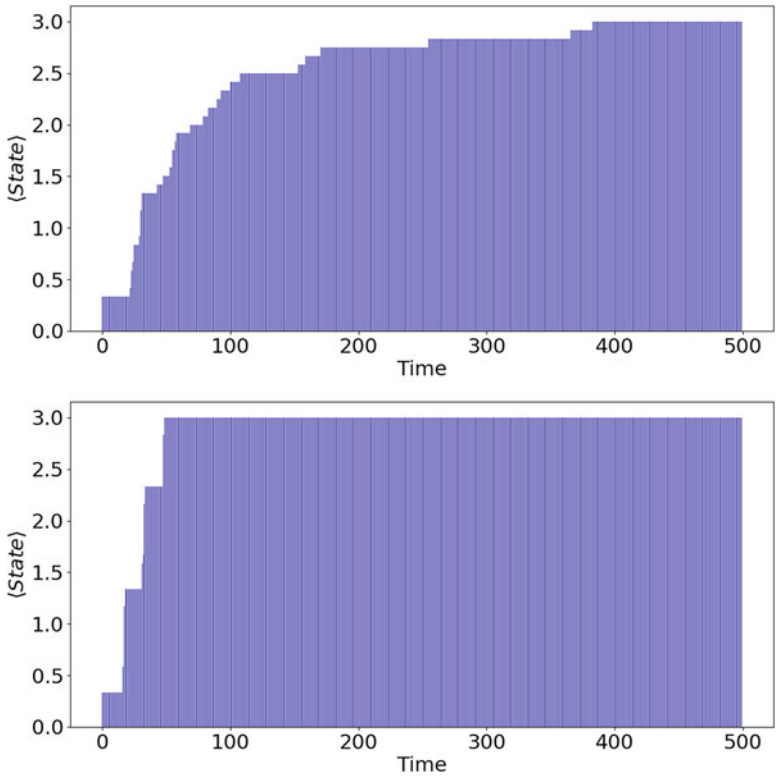


Fig. 6 **Up**: The average state of activated cells designing an endogenous behavior by simulating the model with $P(A)$ as a random probability. **Bottom**: The average state of activated cells illustrating an exogenous case when $P(A) = 0.95$

4 Building a World War Time Series

Looking through current international affairs and conflicts in the IACM dataset, we have designed new scheme to understand conflict intensity based on causalities [25]. The counts of armed conflicts are performed each 15 days. International Armed conflict data when plotted on $log - C(\tau) \times \tau$ scale distinct levels can be categorized as:

(i) *Level W* (War): $C_W(\tau) \gg 3\sigma$ (σ is the standard deviation of $C(\tau)$). Using the notation $C(15) = D_{15} = N_D$: number of deaths in the interval $\tau = 15$ days.

- *stability*: $N_D < 10^2$ on log scale. It has almost no fluctuation, continuous smooth unit vector.
- *conflicts*: $10^2 \leq N_D < 10^3$ on log scale. This is interpreted as one of the parties having conflicts with the other involved in the battle. But no use or threats of any missiles or nuclear weapons.
- *cold war*: $10^3 \leq N_D < 10^4$ on log scale. High production and large number of nuclear weapons tests. There are high counts of threats (conflicts).
- *warm war*: $N_D \geq 10^4$ on log scale. High production and large number of nuclear weapons tests. There are high counts of threats (conflicts). One of the parties uses the missiles. Threats of use of nuclear weapons are high (conflicts).
- *war or hot war*: $N_D \geq 10^5$ on log scale. Both parties use the missiles and nuclear weapons.

(ii) *Level S* (Strike): $C_S(\tau) \geq 3\sigma$ using $C(30) = N_S$ for UCDP monthly counting of armed attacks (strikes) (deaths are not computed in this index) [17]. The time series is shown in Fig. 7.

(iii) *Level T* (Threats): $C_T(\tau) \simeq \sigma$ using $C(\tau)_{p-model} = N_T$. This level can be also called *Level N* from *noise*.

To inspect an international conflict as an extreme event, we are interested in finding whether data is endogenous or requires some thrust and be categorized as exogenous. To analyze endogenous and exogenous patterns in *IACM*, a complex systems model has been built by inducing endogenous and exogenous *P-model* time series as a noise. This is equivalent to the complex daily threats among the agents and pro-agents in conflict involving international affairs conveyed by the official media (see Fig. 8, Appendix 1). The main cause of this component of the signal is political instability due to adverse regime. Actually from our understanding it works as a non-homogeneous turbulent-like information cascade, the underlying process which drives the energy until the state of extreme event [49–52]. War is a matter of making decisions in a very complex system from where the underlying turbulent-like fluctuations follow the patterns as shown in Fig. 3.

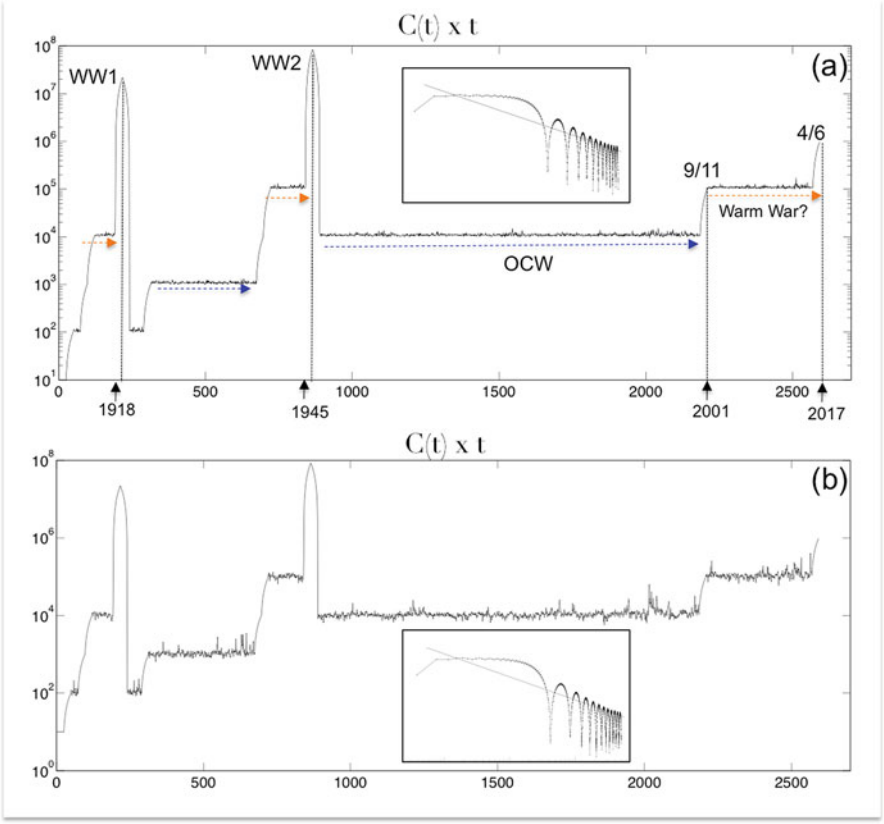


Fig. 7 The $C(t)$ times series for the interstate armed conflict model starting from the first World War conflict and ending at the USA missiles launched on Syria on 6 April 2017. The time series resolution is composed of 2592 points due to a 15 days sampling rate. (**a**) this model is based on the exogenous noise component σ_{XXe} and (**b**) based on the endogenous noise component σ_{DXe}. The respective PSDs are shown using embedded smaller pictures

Considering the three IAC counting components the model for the overall signal is defined as follows:

$$C(\tau) = \omega_D log N_D + \omega_S log N_S + \omega_T log N_T \tag{5}$$

Figure 7 shows the global (endogenous and exogenous) time series generated by Eq. (5). By consistency of conditional standard deviation, the set of weights are: $\omega_D = 10^{-1}$, $\omega_S = 10^{-2}$, and $\omega_T = 10^{-3}$.

The importance of this time series model, considering the results obtained with the prototype that describes a possible underlying process discussed in the

previous section, is to allow the application of forecasting tools for the study of a possible armed conflict that can reach global proportions. In this direction, we are already working on supervised learning methods in artificial intelligence, such as the artificial neural networks (ANNs) and the support vector machine (SVM) which are suitable for nonlinear time series prediction [8, 13, 29, 32].

5 Concluding Remarks

Data mining in the *UCPD/PRIO* dataset has inspired a more careful study on the fluctuation patterns of interstate armed conflict (IAC). Our approach leads to a multifractal p-model originally designed to simulate the highly intermittent spatial fluctuations of the kinetic energy dissipation in non-homogeneous turbulent-like dynamics. There exists a great potential for this approach through the exploration of the features of complex systems that make them distinct (diversity, memory, cross-scale interactions, sensitivity to environmental variability). The great challenge, however, remains to find generalities in the model dynamics to improve understanding and prediction. For example, the meaning of the evolution of the *state conflict space* defined here.

Based on this challenge, in this work, a mathematical and computational treatment has been proposed whose results show some consistency with the real data. The cascade interpretation of energy for non-homogeneous turbulence besides being consistent with the dynamics of complex social systems allowed to develop a prototype of multi-agents for which a simulation with cellular automata was performed. The simulation is validated from the main characteristics studied in this approach to AIC that are the occurrence of extreme endogenous and exogenous events from threats, attacks, and wars. An important question into this context, considering the results, is: what are the real causes of exogenous extreme events? That is, how should the high values of conflict probabilities of the model be interpreted in real life?

The occurrence of a XE_{exo} means that some important factor is not being considered in the dynamics of the system. Therefore, possible causes of exogeny are the following: existence and action of powerful and invisible terrorist cells; infiltrators and heavily camouflaged agents, and, finally, the existence of clandestine nuclear programs for high-destruction attacks. On the other hand, the most perverse aspect of endogeny is the possibility of protocol conflicts, that is, those devised and commissioned by undemocratic agreements between fake enemies.

Finally, from a practical point of view, we can say that geopolitical conflicts, in special the IAC, have become more frequent and intense maybe due to the increasing of exogeny in the conflict system. From the theoretical point of view, this probable increasing exogenous dynamics could gradually suppress smaller armed conflict and simultaneously forced the system to release tensions through

increasingly severe and frequent systemic armed conflict. Is it a self-organized criticality process that should be discussed within the geopolitical scope? Is it possible that the superpowers, for perverse reasons, are interested in carrying out a great protocol war? Exogenous behaviors are less likely since they can result in extreme events that can lead to an armed conflict of world proportions (world war) whose practical sociological results are of no interest to any of the agents involved. In this case, the result would be a global disaster with low resilience thus compromising the permanence of humanity on this planet. Anyway, no kind of war is welcome for most. Thus, this type of study challenges the geopolitical systems to create mechanisms of transparency that allow the population and the press to follow closely the great decisions taken by the most powerful leaders. Only then we will have the great power to prevent major social disasters.

Acknowledgements The authors are grateful for the financial support of the following agencies: CNPq, CAPES, and FAPESP.

Appendix 1: The UCDP Data Base

The Uppsala Conflict Data Program (UCDP) database [57] provides one of the most accurate and extensive information on armed conflicts including attributes like conflict intensity based on total number of battle-related deaths; number of conflicts (see Fig. 8); conflict type; details of warring party including geopolitical information; period with specific start and end date, etc. This database is updated annually and considered well-used data-sources on global armed conflicts. Its definition of armed conflict[2] is becoming a standard in how conflicts are systematically defined and studied. Conflict with minimum of D **battle-related deaths** per period τ and in which one of the warring party is government of a state is recorded as an *Interstate Armed Conflict* (IAC) [17]. UCDP DataBase (UDB) categorizes *IAC* in different intensity levels based on the total battle-related causalities:

- *Not active*: $D < 25$ per year.
- *Minor*: $D \geq 25$ per year but fewer than in the extreme event period.
- *Intermediate*: $25 \ll D < 1000$ total accumulated of at least 1000 deaths, but fewer than 1000 in any given year.
- *War*: $D \geq 1000$ per year.

[2]Based on the UDB the IAC is here defined as: contested incompatibility that concerns government and/or territory where there is a probability of using armed force between two parties, of which at least one is the government of a state.

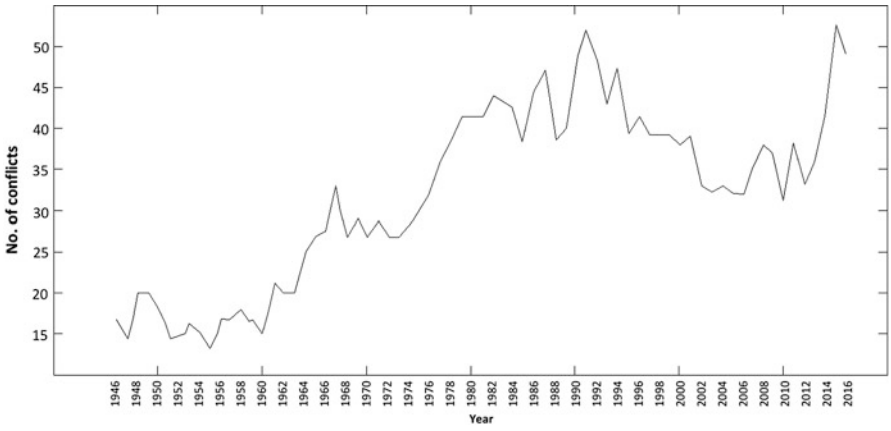


Fig. 8 UCDP monthly counting of armed attacks which is used as the component N_A in our time series model

Appendix 2: Power Laws for Endogenous and Exogenous Time Series

The systemic difference between endogeny and exogeny in abrupt events has been interpreted as a scaling process by Sornette–Deschtres–Gilbert–Ageon (SDGA) [55], where internal perturbations give rise to endogenous extreme events (XE_{endo}) which is characterized, as shown in Fig. 9a, by smoother average continuous fluctuations that increases slowly and after reaching its highest peak and gradually reduces by itself. Differently, an exogenous extreme event (XE_{exo}) results from a preponderant external perturbation and can be characterized by a sudden peak followed by unexpected rapid drop in the fluctuations (Fig. 9b).

The SDGA model is based on the book sales rank. While the book's selling rate, which has a XE_{endo} pattern, only relies on the advertising provided by the common sales system (basically, the publisher's advertising and, especially, the cascade of information between the readers and likely readers), the sales rate of the book with XE_{exo} pattern counted on an unusual systemic outsider high cost advertisement via a famous newspaper or TV broadcast interview.

According to the SDGA a time series can be modeled based on *social epidemic process* where in the beginning, first (*mother*) agent notices the book in advertisement or news or by chance and initiates buy at time t_i. Subsequent (*daughter*) generations of agents are build at different time t resulting in an epidemic that can be modeled by a memory kernel $\phi(t - t_i)$. The net sale is the sum of $1/f$ noise processes following a power law distribution that accounts for XE_{endo}, and impulsive distribution associated with XE_{exo}. The time series can be described by a conditional Poisson branching process given by

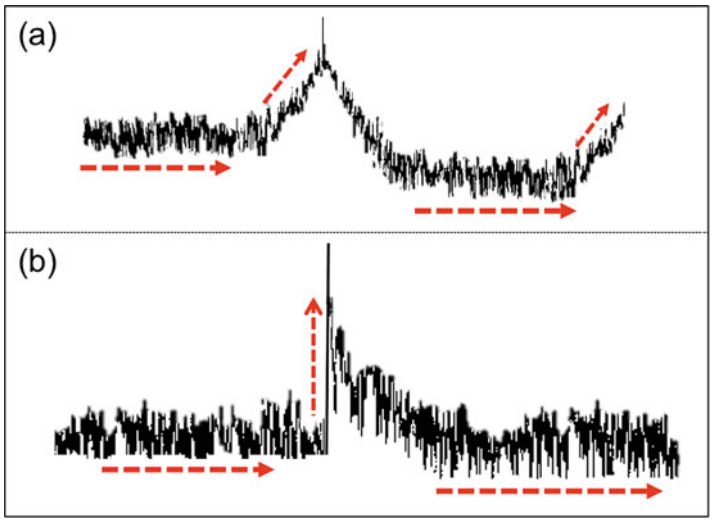


Fig. 9 Typical time series for (**a**) XE_{endo} and (**b**) XE_{exo}, according to the SDGA scaling approach

$$\lambda(t) = R(t) + \sum_{i/_1 \leq t} \mu_i \phi(t - t_i) \tag{6}$$

where μ_i is number of potential agents influenced by the agent i who bought earlier at time t_i. $R(t)$ is the rate of sales initiated spontaneously without influence from other previous agents.

For our generic complex systems scenario, the key idea in the SDGA approach is the invariance of the epidemic model but as a non-homogeneous network of potential daughter generations which can be considered through different values of branching ratio. The ensemble average yields a *branching ratio*, n, that signifies the average number of conflicts triggered by any *mother Agent* within her contact network and rely upon the network topology and impact of the systemic dissipative behavior. Authors considered the sub-critical regime $n < 1$ in order to ensure stationarity which accounts for efficient coarse-grained nature of the complex nonlinear dynamics. The exogenous response function is obtained from Laplace transform of the Green function $K(t)$ of the ensemble average.

According to Sornette [55] a *bare propagator* as $\phi(t - t_i) \sim 1/t^{(1+\theta)}$ with $0 < \theta < 1$ corresponds to long-range memory process which provides information on the conflicts propagation:

$$C_{exo}(t) \equiv K(t) \sim 1/(t - t_c)^{1-\theta}. \tag{7}$$

It provides information about the average number of agents influenced by one agent through any possible direct descent or ancestry. And thus average number of conflicts triggered by one agent can be given as:

$$\int_0^\infty K(t)dt = n/(1-n). \tag{8}$$

Continuous stochastic time series with spontaneous peaks indicates the lack of exogenous shock. Such series can be interpreted as an interaction between external factors over small-scale and enlarged effect of widespread cascade of social influences. This mechanism can explain peak in endogenous time series. Considering results for stochastic processes with finite variance and covariance for average growth of processes prior and later to the peak and applying to $\lambda(t)$ defined in Eq. (5) one get:

$$C_{endo}(t) \sim 1/|t - t_c|^{1-2\theta}. \tag{9}$$

Equations (7) and (9) agree with the prediction that XE_{exo} should occur faster with exponent $1-\theta$ compared to XE_{endo} with exponent $1-2\theta$. Therefore, after characterizing the power laws $1/(t-t_c)^{\beta}$ with highest correlation coefficient [55], the scaling interpretation presents two different universality classes characterizing XE_{endo} with β as $1-2\theta \approx 0.4$ and XE_{exo} with β as $1-\theta \approx 0.7$. These are compatible values with Eqs. (7) and (9) with the choice of $\theta = 0.3 \pm 0.1$.

References

1. http://www2.meteo.uni-bonn.de/staff/venema/themes/surrogates/pmodel/pmodel.m
2. https://github.com/galinslp/Geopolitical-conflicts
3. Albeverio, S., Jentsch, V., Kantz, H. (eds.): Extreme Events in Nature and Society. The Frontiers Collection, Springer (2006). https://doi.org/10.1007/3-540-28611-X
4. Arneodo, A., Bacry, E., Muzy, J.F.: Phys. A **213**(1–2), 232–275 (1995)
5. Arneodo, A.E., Baudet, C., Belin, F., Benzi, R., Castaing, B., Chabaud, B., Dubrulle, B., et al.: Structure functions in turbulence, in various flow configurations, at Reynolds number between 30 and 5000, using extended self-similarity. EPL (Europhys. Lett.) **34**(6), 411 (1996)
6. Bailey, Kenneth D.: Social Entropy Theory (term: "Prigogine entropy"), p. 72. State University of New York Press, New York (1990)
7. Bak, P.: How Nature Works. Springer, New York (1996)
8. Ben Taieb, S., Sorjamaa, A., Bontempi. G.: Multiple-output modeling for multi-step-ahead time series forecasting. Neurocomputing **73**(10), 1950–1957 (2010)
9. Boccara, N.: Modeling Complex Systems. Springer, New York (2010)
10. Bohr, T., Jensen, M.H., Paladin, G., Vulpiani, A.: Dynamical Systems Approach to Turbulence. Cambridge University Press, Cambridge (1998)
11. Bolzan, M.J.A., Ramos, F.M., Sa, L.D.A., Neto, C.R., Rosa, R.R.: Analysis of fine-scale canopy turbulence within and above an Amazon forest using Tsallis generalized thermostatistics. JGR **107-D20**, 8063 (2002)

12. Bolzan, J.M., Rosa, R.R., Sahay,Y.: Multifractal analysis of low-latitude geomagnetic fluctuations. Ann. Geophys. **27**(2) (Feb 2009)
13. Brownlee, J.: Time series prediction with LSTM recurrent neural networks with Keras. In: Deep learning with python, MLM (2016)
14. Buckley, W.: Sociology and the Modern Systems Theory. Prentice-Hall, Upper Saddle River (1967)
15. Couzin, I.D., Krause, J.: The social organization of fish schools. Adv. Ethology **36**, 64 (2001)
16. Davis, A., Marshak, A., Cahalan, R., Wiscombe, W.: The landsat scale break in stratocumulus as a three-dimensional radiative transfer effect: implications for cloud remote sensing. J. Atmos. Sci. **54**(2) (1997)
17. Dupuy, K., Gates, S., Nygard, H.M., Rudolfsen, I., Rustad, S.A., Strand, H., Urda, H.: Trends in Armed Conflict, 1946–2016. PRIO Conflict Trends (June 2017)
18. Enescu, B., Ito, K., Struzik, Z.: Geophys. J. Int. **164**(1), 63–74 (2006)
19. Epstein, J.M., Axtell, R.: Growing artificial societies: social science from the bottom up. The Brookings Institution/MIT Press, Cambridge (1996)
20. Farge, M.: Annu. Rev. Fluid Mech. **24**, 395–457 (1992)
21. Frisch, U.: Cambridge University Press, New York (1995)
22. Fuchs, C.: Internet and Society: Social Theory in the Information Age. Routledge, New York (2008)
23. Gardiner, C.W.: Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences, 3rd edn. Springer Series in Synergetics, Berlin (2004)
24. Gleditsch, N.P., Wallensteen, P., Eriksson, M., Sollenberg, M., Strand, H.: Armed conflict 1946–2001: a new dataset. J. Peace Res. **39**(5), 615–637 (2002)
25. Global Terrorism Index: Institute for Economics & Peace, pp. 94–95 (November 2016). ISBN 978-0-9942456-4-9
26. Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I.: Phys. Rev. A **33**, 1141 (1986)
27. Hersh, M.: Mathematical Modelling for Sustainable Development. Springer, New York (2006)
28. Hughes, R.L.: The flow of human crowds. Annu. Rev. Fluid Mech. **35**, 169–182 (2003)
29. Jiang, P., Chen, C., Liu, X.: Time series prediction for evolutions of complex systems: a deep learning approach. In: Proceedings of 2016 IEEE International Conference on Control and Robotics Engineering (ICCRE). https://doi.org/10.1109/ICCRE.2016.7476150
30. Kari, J.: Theory of cellular automata: a survey. Theor. Comput. Sci. **334**(1–3), 3–33 (2005)
31. Keylock, C.J.: Multifractal surrogate-data generation algorithm that preserves pointwise Hölder regularity structure, with initial applications to turbulence. Phys. Rev. E **95**(3), 032123 (2017)
32. Konar, A., Bhattacharya, D.: Time-Series Prediction and Applications: A Machine Intelligence Approach. Springer, Berlin (2017)
33. Majda, A.: Introduction to Turbulent Dynamical Systems in Complex Systems. Springer, New York (2016)
34. Mallat, S.: IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)
35. Margalef, R., Gutiérrez, E.: How to introduce connectance in the frame of an expression for diversity. Am. Nat. **5**, 601–607 (1983)
36. Marwick, A.: War and Social Change in the Twentieth Century: A Comparative Study of Britain. Macmillan Student Editions. Macmillan, London (1974)
37. Meneveau, C., Sreenivasan, K.R.: Simple Multifractal Cascade Model for Fully Developed Turbulence. Phys. Rev. Lett. **59**, 1424–1427 (1987)
38. Meyers, R.A. (Ed.): Encyclopedia of Complexity and Systems Science. Springer, New York (2009)
39. Miller, A.L.: China an emerging superpower? Stanf. J. Int. Rel. **6**(1) (2005)
40. Milsum, J.H.: The technosphere, the biosphere, the sociosphere: their systems modeling and optimization. IEEE Spectr. **5**(6) (1968). https://doi.org/10.1109/MSPEC.1968.5214690.
41. Muzy, J.F., Bacry, E., Arneodo, A.: Phys. Rev. Lett. **67**(25), 3515–3518 (1991)

42. Ohnishi, T.J.: A mathematical method for the turbulent behavior of crowds using agent particles. J. Phys. Conf. Ser. **738**, 012091 (2016)
43. Oswiecimka, P., Kwapien, J., Drozdz, S.: Phys. Rev. E **74**, 016103 (2006)
44. Page, S.E.: Diversity and Complexity. Princeton University Press, Princeton (2011)
45. Pei, S., Morone, F., Makse, H.A.: Theories for influencer in complex networks. In: Spreading Dynamics in Social Systems, Lehmann, S., Ahn, Y.-Y. (Eds.). Springer, New York (2017)
46. Prigogine, I.; Kondepudi, D., Modern Thermodynamics. Wiley, New York (1998)
47. Ramos, F.M., Rosa, R.R., Neto, C.R., Bolzan, M.J.A., Sá, L.D.A.: Nonextensive thermostatistics description of intermittency in turbulence and financial markets. Nonlinear Anal. Theory Methods Appl. **47**(5), 3521–3530 (2001)
48. Ramos, F.M., Bolzan, M.J.A., Sá, L.D.A., Rosa, R.R.: Atmospheric turbulence within and above an Amazon forest. Physica D Nonlinear Phenom. **193**(1–4), 278–291 (15 June 2004)
49. Ramos, F.M., Lima, I.B.T., Rosa, R.R., Mazzi, E.A., Carvalho, J.C., Rasera, M.F.F.L., Ometto, J.P.H.B., Assireu, A.T., Stech. J.L.: Extreme event dynamics in methane ebullition fluxes from tropical reservoirs. Geophys. Res. Lett. **33**(21) (2006)
50. Rieucau, G., Holmin, A.J., Castillo, J.C., Couzin, I.D., Handegard, N.-O.: School-level structural and dynamic adjustments to perceived risk promote efficient information transfer and collective evasion in herring. Anim. Behav. **117**, 69–78 (2016)
51. Rodrigues Neto, C., Zanandrea, A., Ramos, F.M., Rosa, R.R., Bolzan, M.J.A., Sá, L.D.A.: Phys. A **295**(1–2), 215–218 (2001)
52. Schertzer, D., Lovejoy, S.: Multifractal Generation of Self-Organized Criticality. In: Novak, M.M. (ed.) Fractals in the Natural and Applied Sciences, pp. 325–339. North-Holland, Elsevier (1994)
53. Sethna, J.P.: Statistical Mechanics: Entropy, Order Parameters, and Complexity. Oxford-Clarendom Press, Oxford (2017)
54. Smith, M., Zeigler, M.S.: Terrorism before and after 9/11—a more dangerous world? Res. Polit. **4**(4), 1–8 (2017)
55. Sornette, D., Deschâtres, F., Gilbert, T., Ageon, Y.: Endogenous versus exogenous shocks in complex networks: an empirical test using book sale rankings. Phys. Rev. Lett. **93**, 228701 (2004)
56. Struzik, Z.R.: Fractals **8**, 163–179 (2000)
57. Themnér, L.: The UCDP/PRIO Armed Conflict Dataset Codebook, Version 4- 2016 (2016)
58. Turiel, A., Perez-Vicente, C.J., Grazzini, J.: J. Comput. Phys. **216**, 362–390 (2006)
59. University of Maryland's Global Terrorism Database (GTD). https://www.start.umd.edu/gtd/
60. von Bertalanffy, K.L.: General Theory of Systems. Penguin University Books, Penguin (1978)
61. Weisbuch, G.: Complex Systems Dynamics. Santa Fé Institute. Westview Press, Boulder (1994)
62. WSJ Graphics: Wall Street J. **Nov. 14** (2015)
63. Xiong, G., Zhang, S., Yang, X.: Phys. A **391**, 6347–6361 (2012)
64. Yam, Y.B.: Dynamics of Complex Systems. Addison-Wesley, Boston (1992)
65. Zivieri, R., Pacini, N., Finocchio, G., Carpentieri, M.: Rate of entropy model for irreversible processes in living systems. Sci. Rep. **7**. Article number: 9134 (2017). https://doi.org/10.1038/s41598-017-09530-5

Index

L. Bacelar Lima Santos et al. (eds.), *Towards Mathematics, Computers and Environment: A Disasters Perspective*, https://doi.org/10.1007/978-3-030-21205-6