Management and Industrial Engineering

Mangey Ram
J. Paulo Davim *Editors*

Diagnostic Techniques in Industrial Engineering

Springer

Management and Industrial Engineering

Series editor

J. Paulo Davim, Aveiro, Portugal

More information about this series at http://www.springer.com/series/11690

Mangey Ram · J. Paulo Davim
Editors

Diagnostic Techniques in Industrial Engineering

Springer

Editors
Mangey Ram
Department of Mathematics, Computer Science and Engineering
Graphic Era University
Dehradun, Uttarakhand
India

J. Paulo Davim
Department of Mechanical Engineering
University of Aveiro
Aveiro
Portugal

ISSN 2365-0532 ISSN 2365-0540 (electronic)
Management and Industrial Engineering
ISBN 978-3-319-65496-6 ISBN 978-3-319-65497-3 (eBook)
https://doi.org/10.1007/978-3-319-65497-3

Library of Congress Control Number: 2017949121

Printed on acid-free paper

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

Preface

Diagnostic techniques in industrial engineering are the most multi-dimensional study topic that can face an industrial engineering nowadays. Diagnostics techniques deals with the fault detection, isolation and identification. Fault detection is a task to indicate whether something is going wrong in the inspected engineering system; fault isolation is a task to locate the element that is faulty; and fault identification is a task to determine the nature of the fault when it is identified.

The main aim of the book is to give theoretical as well as practical information about diagnostic techniques in industrial engineering like methods of modern aircraft structures health monitoring; machine signature integrity and data trends monitoring; hazard and risk assessment and mitigation for objects of critical infrastructure; re-manufacturing strategy for economic benefits; computer interfaces; models for soil vapor extraction; vibration-based diagnostics of fatigued structures based on damage pattern recognition via minimization of misclassification probability; statistical techniques of reliability, availability, and maintainability; fuzzy logic in flexible manufacturing system.

This book "Diagnostic Techniques in Industrial Engineering" gripped on a comprehensive range of diagnostic techniques in industrial engineering areas for different tasks not only in a particular field. Several tools, techniques and strategies in industrial engineering applications will be discussed in the book.

Dehradun, India — Mangey Ram
Aveiro, Portugal — J. Paulo Davim

Acknowledgements

The Editors acknowledges Springer for this opportunity and professional support. Also, we would like to thank all the chapter authors for their availability for this work.

Mangey Ram
J. Paulo Davim

Contents

Editors and Contributors

About the Editors

Mangey Ram received the B.Sc. degree in science from Chaudhary Charan Singh University, Meerut, India, in 2000, the M.Sc. degree in mathematics from Hemwati Nandan Bahuguna Garhwal University, Srinagar (Garhwal), India, in 2004, and the Ph.D. degree major in mathematics and minor in computer science from G. B. Pant University of Agriculture and Technology, Pantnagar, India, in 2008. He has been a Faculty Member for around nine years and has taught several core courses in pure and applied mathematics at undergraduate, postgraduate, and doctorate levels. He is currently a Professor at Graphic Era University, Dehradun, India. Before joining the Graphic Era University, he was a Deputy Manager (Probationary Officer) with Syndicate Bank for a short period. He is Editor-in-Chief of *International Journal of Mathematical, Engineering and Management Sciences*; Executive Editor of Journal of Graphic Era University; Associate Executive Editor of Journal of *Reliability and Statistical Studies* and the Guest Editor & Member of the editorial board of many journals. He is a regular Reviewer for international journals, including IEEE, Elsevier, Springer, Emerald, John Wiley, Taylor & Francis and many other publishers. He has published 111 research publications in IEEE, Springer, Emerald, World Scientific and many other national and international journals of repute and also presented his works at national and international conferences. His fields of research are reliability theory and applied mathematics. Dr. Ram is a senior member of the IEEE, member of Operational Research Society of India, Society for Reliability Engineering, Quality and Operations Management in India, International Association of Engineers in Hong Kong, and Emerald Literati Network in the U.K. He has been a member of the organizing committee of a number of international and national conferences, seminars, and workshops. He has been conferred with “Young Scientist Award” by the Uttarakhand State Council for Science and Technology, Dehradun, in 2009. He has been awarded the “Best Faculty Award” in 2011 and recently Research Excellence Award in 2015 for his significant contribution in academics and research at Graphic Era University.

J. Paulo Davim received the Ph.D. degree in Mechanical Engineering in 1997, the M.Sc. degree in Mechanical Engineering (materials and manufacturing processes) in 1991, the Dipl.-Ing Engineer’s degree (5 years) in Mechanical Engineering in 1986, from the University of Porto (FEUP), the Aggregate title (Full Habilitation) from the University of Coimbra in 2005 and the D. Sc. degree from London Metropolitan University in 2013. He is Eur Ing by FEANI-Brussels and Senior Chartered Engineer by the Portuguese Institution of Engineers with a MBA and Specialist title in Engineering and Industrial Management. Currently, he is Professor at the Department of Mechanical Engineering of the University of Aveiro, Portugal. He has more than 30 years of teaching and research experience in Manufacturing, Materials and Mechanical Engineering with special emphasis in Machining & Tribology. He has also interest in Management & Industrial

Engineering and Higher Education for Sustainability & Engineering Education. He has received several scientific awards. He has guided large number of postdoc, Ph.D. and masters students. He has worked as evaluator of projects for international research agencies as well as examiner of Ph.D. thesis for many universities. He is the Editor in Chief of several international journals, Guest Editor of journals, books Editor, book Series Editor and Scientific Advisory for many international journals and conferences. Presently, he is an Editorial Board member of 25 international journals and acts as reviewer for more than 80 prestigious Web of Science journals. In addition, he has also published as editor (and co-editor) more than 100 books and as author (and co-author) more than 10 books, 60 book chapters and 400 articles in journals and conferences (more than 200 articles in journals indexed in Web of Science/h-index 41+ and SCOPUS/h-index 50+).

Contributors

Michael Kanisuru Adeyeri Department of Mechanical Engineering, The Federal University of Technology Akure, Akure, Nigeria; Institute for Mechatronics Engineering Systems, Leibniz University, Hannover, Germany

Mohini Agarwal Department of Operational Research, University of Delhi, Delhi, India

Deepti Aggrawal Delhi Technological University, East Delhi Campus, Delhi, India

Adarsh Anand Department of Operational Research, University of Delhi, Delhi, India

Gunjan Bansal Department of Operational Research, University of Delhi, Delhi, India

Andrey A. Bautin Central Aerohydrodynamic Institute (TsAGI), Zhukovsky, Russia

Dinesh C. S. Bisht Department of Mathematics, Jaypee Institute of Information Technology, Noida, India

Alexander V. Bochkov Risk Analysis Center, NIIgazeconomika LLC, Moscow, Russia

Irina V. Gadolina Mechanical Engineering Research Institute of Russian Academy of Science (IMASH RAN), Moscow, Russia

Mariem Kacem Univ Lyon, Ecole Nationale D'Ingénieurs de Saint-Etienne, Laboratoire de Tribologie et Dynamique Des Systèmes LTDS UMR 5513 CNRS, Saint-Etienne, France

Preeti Malik Department of Computer Science and Engineering, Graphic Era University, Dehradun, Uttarakhand, India

Lata Nautiyal Department of Computer Science and Engineering, Graphic Era University, Dehradun, Uttarakhand, India

Konstantin N. Nechval Department of Applied Mathematics, Transport and Telecommunication Institute, Riga, Latvia

Nicholas A. Nechval BVEF Research Institute, University of Latvia, Raina Blvd 19, Riga 1050, Latvia; Department of Mathematics, Baltic International Academy, Lomonosov Street 4, Riga 1019, Latvia

Ljubisa Papic Faculty of Technical Sciences Cacak, University of Kragujevac, Serbia and DQM Research Center, Prijevor, Serbia

Mangey Ram Department of Mathematics, Computer Science and Engineering, Graphic Era University, Dehradun, Uttarakhand, India

Pankaj Kumar Srivastava Department of Mathematics, Jaypee Institute of Information Technology, Noida, India

Yury A. Svirskiy Central Aerohydrodynamic Institute (TsAGI), Zhukovsky, Russia

Panagiotis Tsarouhas Department of Supply Chain Management & Logistics, Technological Educational Institute of Central Macedonia, Katerini, Hellas, Greece

Methods of Modern Aircraft Structural Health Monitoring and Diagnostics of Technical State

Yury A. Svirskiy, Andrey A. Bautin, Ljubisa Papic and Irina V. Gadolina

Abstract This chapter provides an overview of the most relevant methods for monitoring the integrity of structures made of isotropic (metals and alloys) and anisotropic (polymer composite) materials, which are widely used in the manufacture of parts and construction structures in the modern industry. Special attention is paid to ensuring the safe operation of aircraft structures, the use of onboard monitoring systems which will allow to reduce economic costs in the near future and to increase flying safety. The paper describes different variants of implementation of the onboard recording systems of data in the analysis of the actual state of the elements of the air frame. The advantages are shown, due to which the construction monitoring will be able to replace the existing system of provision and supporting of the airworthiness, implemented through periodic inspections.

Y.A. Svirskiy · A.A. Bautin (✉)
Central Aerohydrodynamic Institute (TsAGI), Zhukovsky, Russia
e-mail: andrey.bautin@tsagi.ru

Y.A. Svirskiy
e-mail: yury.svirsky@tsagi.ru

L. Papic
Faculty of Technical Sciences Cacak, University of Kragujevac, Serbia and DQM Research Center, Prijevor, Serbia
e-mail: dqmcenter@mts.rs

I.V. Gadolina
Mechanical Engineering Research Institute of Russian Academy of Science (IMASH RAN), Moscow, Russia
e-mail: gadolina@mail.ru

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*, Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_1

1 Introduction

In the last two decades, automated structural health monitoring (SHM) systems were developed and widely implemented in the most technically developed countries of the world. The SHM means the continuous and autonomous monitoring of damage, loading, the environmental parameters, and interaction of structural elements with the environment by means of permanently attached or embedded sensor systems ensuring the integrity of the structure [1]. After a successful approbation for land-based structures (bridges, power elements of high-rise buildings, etc.), SHM methods began to be introduced in aviation. In the 90s of the last century, the Airbus company with the purpose of increasing the strength, reliability, and durability of aircraft structures and reducing the downtime of aircraft and the cost of their maintenance started to develop common approaches on creation of the SHM system for aircraft plane (A/C) within the framework of the "philosophy of intelligent aircraft structures" [2]. The Aerospace Industry Steering Committee SHM-AISC was established in 2007. It is responsible for coordinating the development and implementation of automated onboard systems for monitoring the design integrity of aircraft and reusable orbital carrier with the use of networks of embedded sensors. The International Council of Management SHM-AISC includes such companies and organizations as Airbus, Boeing, BAE Systems, Embraer, Honeywell, Aviation Administration of the USA and Europe, as well as the scientific laboratory of the US armed forces, NASA, and some leading universities [3].

Operation of the structural health monitoring systems involves the installation of different types of sensors on the structures to determine the effects of physical (humidity and ambient temperature) and power (static, cyclic, impact, and other types of loads) actions on their strength and durability [1, 4, 5].

The ultimate goal of these developments is the creation of an artificial intelligence system, which is supposed not only to be able to detect defects and malfunctions, but also to respond adequately to defects' presence and to provide appropriate recommendations to the service personnel [2, 4].

The development of the health monitoring system is divided into several stages corresponding to the system generations. These stages are close to the stages that are considered by leading aviation firms like Airbus and Boeing up to the numbering of stages and minor details [2].

- At the first stage (zero and first generation), the SHM systems are used at the structures' testing, their maintenance, and restoration. These trials also serve to test different sensors and monitoring systems in terms of their applicability to real structures. The zero-generation SHM systems are widely used in ground tests of the aircraft in our days. As an example, the SHM has been used at the certification testing on the Airbus A380 [2].
- The next stage of the SHM systems development (the second generation) is characterized by the use of sensors, information from which is read after the flight or during maintenance. While ensuring an adequate level of sensors'

reliability, it is planned to pass to sensors operating in real time, i.e., recording and transmitting information during the flight.

- The full integration of the SHM systems with onboard A/C computing and control system means the transition to the third and final generation. It is expected that the SHM system will be used to develop new approaches to the aircraft design that will ensure weight reduction of metal and composite structures by 15% [2].

The general condition for obtaining and processing the diagnostic information about the presence of damage in structures is the use of built-in and external sensors, nondestructive testing (NDT) devices, systems for storage, and processing of information, algorithms, and programs for decision-making. As noted in paper [6], diagnostic information is usually limited in volume and indirect. Existing nondestructive inspections do not detect all damages and cracks, which can later become the reason of failure states. There is a high probability of defects missing due to the equipment imperfection, negligence of the operator or inaccessible location of defects and the ungrounded frequency control. For example, if the frequency of inspections is inconsistent with the timing of initiation and propagation of fatigue cracks, it can lead to the formation of defects of critical size and consequently to the destruction of the structure [7]. Thus, the data on loading regimes provide a valuable additional source of information on the technical condition of structures. On the basis of these data using various numerical schemes, it is possible to identify the load history of the test unit and the degree of damage accumulated during its service. While comparing the results of damage evaluation with the diagnostic data of the technical condition of constructions, the parameters are estimated which have not been identified with sufficient accuracy in the previous stages. Thus, two sources of information—diagnostic data on the state of the object and data on the loading history for metal structures—are closely related and mutually dependent [7]. For the construction of the polymer composite materials (PCM), the additional information of impact damage and its intensity is more important. The solution to the problem of obtaining such information, its processing and decision-making about the maintenance strategy of an individual sample of A/C and prediction of its residual life should be implemented in the framework of integrated intelligent systems, which monitor the A/C structural integrity [8]. It should be pointed out that the important part of the SHM system creation is to develop the standards, defining rules, procedures of use, and decision-making about receiving information from them.

2 A Brief Overview of Sensors Used

The sensors development is moving in the direction of the technologies through which it is possible to detect various types of practically important damages such as cracks, corrosion, the breach of adhesion, laminations, impact damages, etc. in metal and composite materials. The main requirements are as follows: small size and weight, ease of installation or adaptability, durability, and reliability. Currently, these requirements are satisfied by a number of sensors, which are based on different physical methods [9]:

- Acoustic emission (AE): Passive transducers listen to matrix cracking, lamination, and fiber breakage;
- Acousto-Ultrasonics (AU): A grid of piezoelectric sensors sends and receives ultrasonic pulses and analyzes changes in impulses patterns to identify and describe the damage;
- Comparative vacuum monitoring (CVM): The smallest air flow can be detected by leakage failure between atmospheric pressure and vacuum because of the crack in the matrix;
- Comparative vacuum monitoring through the thickness (CVM-TT): Using the drilled holes with a diameter of <1 mm, it is possible to reveal a breach of adhesion and lamination using the principle of similar systems in CVM;
- Crack wire sensors (CW): The break due to the occurrence of cracks or other damage serves as an alarm;
- Electromagnetic interference sensors (EMI): The system applies the built-in piezoelectric sensors and signal level analyzer. Increasing the level of signal relative to baseline values due to grease contamination or moisture indicates the presence of stratification;
- Eddy testing foil sensors (ETFS): Sensors generate a field of eddy currents in conducting materials which are violated by the cracks and corrosion damage;
- Fiber Bragg grating (FBG): The most used method of FBG is its use for measurement of temperature, strain, and vibration;
- Imaging ultrasonics (IU): Miniature, integrated network of sensors generate a signal through the material of the structural member. Changes in the reflected signal indicate the discontinuity or damage;
- Strain gages (SG): Using traditional strain gages to determine the strain.

Also, the new types of sensors are developed, in particular:

- Micro- and nanoelectromechanical systems MEMS and NEMS (microelectromechanical systems—MEMS, nanoelectromechanical systems—NEMS) (see Fig. 1);

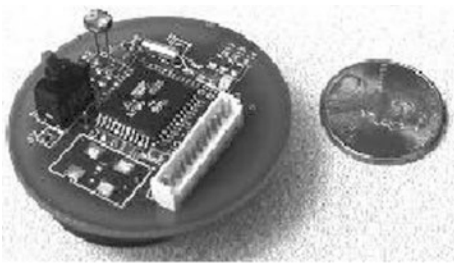


Fig. 1 Example of a MEMS sensor (≈16 cm^3), including a wireless telemetry system, microprocessor, sensors, and power supplies

- Sensors on the basis of printing conductive ink (printed conductive-ink sensors).
- Other types of fiber-optic sensors, including ones for registration acoustic emission signals [10] and for their use as sensors of cracks, use the effect of Rayleigh scattering (Rayleigh scattering). The FBG sensors are also used in conjunction with piezoelectric sensors [11]. In Ref. [11], it is noted that the measurement of deformation using FBG for the detection of damage caused by impacts is not effective enough. To illustrate, there is the following data. The impact energy of 40 J has caused significant damage but did not cause change of deformation at a distance of 40 mm from the point of impact. Reliably measured changes in the deformation caused by impact energy 30 J were localized at a distance not exceeding 20 mm.

Table 1 lists the applications of the main types of sensors (∨∨ indicates that the sensor suits the best for this type of damage).

We should also add the conventional sensors of displacement, acceleration (accelerometers), and temperature to this list. They are used in the following methods [12]:

- Stiffness monitoring: This enables the damage stages to be identified and any deflection limit to be maintained. For coupons, changes in the modulus can be determined from the stress–strain curve, either during fatigue testing or by static loading.
- Load–deflection curve (hysteresis) monitoring: For composites, the internal damping is much higher than in metals and is an indicator of damage in the material.

Table 1 Applications of the main types of sensors

Damage type	Sensor type										
	CVM	ETFS	AE	IU	CW	IDDS	AU	FOS	EMI	CVM-TT	SG
Crack detection and evaluation	∨∨	∨		∨							
Rupture detection and evaluation	∨∨			∨	∨						
Impact detection and evaluation			∨∨			∨		∨∨			
Delamination detection and evaluation			∨	∨			∨∨	∨∨		∨	
Adhesion quality assessment								∨∨	∨	∨∨	∨
Monitoring of sticky repair pads								∨∨		∨	∨
Disorders of adhesion detection and evaluation			∨					∨∨		∨∨	
Deformation/strain monitoring								∨∨			∨∨

- Resonant frequency and natural frequency monitoring: They provide information at a global level of the state of the material or structure at microlevel.
- Temperature monitoring: Damage induced in composites will generate heat, and as these materials have poor conductivity, it should be generally easy to detect. Infrared emissions can be detected with a suitable camera and will locate the areas of damage and how these will grow. Thermocouples are simple to install and have been placed in locations where high strains have been detected, possibly by a strain analysis.

In many cases, the sensors are combined in a distributed network, the use of which is performed in three ways [9]:

1 By the place of installation: The sensors are just the objects which are set on the construction in this case. The reading is collected manually using an external measuring system through the selected interval.
2 A network of sensors at the place of installation: In this case, the sensors are complemented by miniature measurement electronics. Stored data are periodically collected by the technical staff on the ground.
3 Wireless network sensors: Electronics provides wireless transmission of data as a development of the second method. Systems receiving this data diagnose construction in real time. The telemetry system gives the possibility of continuous wireless data transmission in the air or on the ground to a remote site. The Web site can be programmed to data scan, the determination of the threshold of structural damage, and informing staff about the need for repairs or other maintenance.

Currently, SHM systems have become a standard tool in the certification of static and fatigue full-scale tests of air constructions [9]. It reduces the switching-off of testing for control of the selected zones. For example, sensors of acoustic emission (Physical Acoustics Corp. (PAC, Princeton Junction, NJ, USA)) were used to monitor carbon horizontal stabilizer during field tests of the wide-bodied Boeing 777. Also, the AE systems were used for static tests on critical load and destruction for the identification and assessment of subcritical damage propagation and determination of a sequence of destruction of individual elements of the design.

Similarly, during the full-scale fatigue testing of the Airbus A380 in Toulouse (Toulouse, France) conducted by IABG (Dresden, Germany), 47500 flights were made during 26 months using a variety of health monitoring system (HMS) sensors, including the vacuum systems CVM, eddy current foil sensors (ETFS), AE sensors, and crack wire. Installed on the fuselage and wings, these HMS provided detailed information on crack generation in aluminum airframe, in the carbon fiber center section of the wing and parts of the fuselage made of the glass-reinforced aluminum (GLARE).

3 Methods of Location of Impact Damage Based on the Analysis of the Response of Polymer Composite Materials' Structures

The problem of damage detection and determination of their parameters in aircraft construction is a difficult task even for ground services. In this regard, for SHM systems the urgent task is not only to detect the damage but also to detect the fact of impact damage. It can be done monitoring the structural elements response to the impact. The fact of an impact serves as a signal to an extraordinary nondestructive testing of the integrity of the structure.

The problem of impact damage detecting on the response to the impact of structural elements of onboard monitoring system includes the following subproblems:

- Identification of the point of impact;
- Evaluation of damage parameters from impact;
- Determination of impact parameters.

Conventionally, the approaches to solving these subtasks can be divided into two types:

1. Creating the most accurate mathematical model which predicts the properties of the structure ("theoretical" approach) and
2. Conducting a large number of experiments and then predicting the properties of the structure via interpolation of data obtained in such experiments ("calculation-experimental" approach).

Due to the fact that the construction of sufficiently accurate models of complex aircraft constructions suitable for the SHM systems is currently unrealizable, the most popular is the "theory-experimental" approach.

Let e(i, t) be the deformation of the i-th sensor at time t, e_t is the vector composed of such deformations. Suppose that at time t = 0 the impact occurred, and at times t = 0, …, t = ∞ of the sensor readings were taken. In the general case, the dependence of the impact force at the moment t at point x would be as follows:

$$F_{xt} = f_t(e(t = 0), e(t = 1), \ldots, e(t = \infty)) \tag{3.1}$$

In the simplest cases (infinite perfectly rigid one-dimensional rod, two-dimensional infinite surface, the infinite speed of sound, etc.), the dependence f_t takes the simple enough forms. In such cases, we may build the appropriate mathematical model to determine the coordinates of the impact. So, in an environment with an infinite speed of sound and a very fast damping of the oscillations, members with t > 0 can be neglected, and dependence (3.1) takes the form:

$$F_{x0} = f_0(e(t = 0)) \tag{3.2}$$

However, in the case of the arbitrary configuration construction assembled from materials with random properties, the dependence of (3.1) is very difficult. In this case, the main factors preventing the creation of effective mathematical models are as follows:

- The complexity of mathematical modeling of composite materials and complex parts from them;
- The complexity of modeling processes in various joints of complex structures;
- Difficulties in modeling the construction made from various materials.

In this regard, in practice, there are various ways of overcoming the above difficulties.

3.1 Simplified Model Methods

In work [13], it was proposed to localize the point of impact according to a highly simplified mathematical model: The distance from sensor p_i to the impact point x is proportional to the maximum amplitude: $|x - p_i| \sim \max(e_{i,t=0}, \ldots, e_{i,t=\infty})$. In work [13], this model was checked experimentally, but the results were not satisfactory. In this regard, there is a reason to believe that the approach of using a deliberately simplified model is unsatisfactory.

Another common method of finding the point of impact is the method of time delays: It is supposed that the disturbance comes faster to the sensors located closer to the point of impact. This method is also unsatisfactory for complex anisotropic structures.

There are other approaches to finding the point of impact, based on simplified mathematical models: the measurement of power distribution [14], etc.

3.2 Linear Stationary System

One of the most promising approaches is the approach given in [14, 15]: a representation of the model in the form of a linear stationary system. In this case, a minimum number of initial assumptions are used: linearity (deformation of the i-th sensor is proportional to the impact force) and time invariance (a delay of indications of the i-th sensor relative to the exposure time is fixed). In such a system, the deformation depends on the strength in the following form:

$$e(i,t) = \sum_{k=0}^{t} h(i, k)F(x, k - t) \tag{3.3}$$

where F(x, t) is the force at point x at time t, h(i, t) is the transition function.

In matrix form, this equation can be written in the form:

$$e(i) = H(i)F_t(x) \tag{3.4}$$

where H is the lower triangular matrix composed of the transition functions of this linear system, F_t is a vector of power values at points in time t.

Thus, the problem is reduced to the following stages:

1 Definition of the matrix H: While doing this, it is necessary to use the experimentally obtained compliance of the strains e(i, t) on the history of the loading F (x, t). Using these data and using the least squares method, we can find the desired transition functions.

$$\sum (e(i) - H(i)F_t(x))^2 \rightarrow \min \tag{3.5}$$

2 The definition of the point of impact x was found by means of transition functions: Assuming the setting of force $F_0(t)$ is known, we solve the minimization problem (3.5) for all x;
3 Determination of the settings of the force F(t): Using the properties of linearity of our system, we find the inverse transition matrix H′, such that

$$F_t(x) = Y'(t) \cdot e(t) \tag{3.6}$$

Assuming that the impact location is known, we find the history of loading by the formula (3.6);
4 Alternately, assuming that the point of impact and the load history are known, we solve tasks p. 2 and p. 3 by mean of iterations. Eventually, the solution is converged.

3.3 The Dynamic System

The most promising approach for the recovery of impact force by sensor readings is a representation of the model as a dynamic system [13], because during the impact the condition of linearity and stationarity of the system typically fails. According to the theorem of Takens [16], it is possible to recover the properties of the random smooth dynamical system without knowing its exact mathematical model. It is done according to the following algorithm:

1 Selection of parameters τ (time, seconds) and D (constant): Parameters can be chosen arbitrarily, but in that case good results are not always obtained. In practice, some empirical assumptions are often used;
2 Let E(i, t) be matrix function of time t composed of sensor readings in some points of time.

$$E(i, t) = (e(i, t), e(i, t + \tau), \ldots, e(i, t + \tau(D - 1))) \tag{3.7}$$

3 In our dynamic system, the matrix will depend on the force of the impact at the point x at time t: E(i, t) = f(F(x, t)).

Next, we need to obtain the dependence of f.

$$\sum (E(i, t) - f(F(x, t)))^2 \rightarrow \min, \text{ (for all i, x, t)} \tag{3.8}$$

Usually, neural networks are used for the approximation of such high-dimensional functions. In work [14] for the approximation of f, the support vector machines were used.

3.4 Method Using Neural Network

The results of the method of determining the coordinates and power of the impact using a neural network developed by Rybakov are presented in work [17]. The method was applied to the specimen, which was a square 200 × 200 × 3 mm of the combined composite material. Layers of carbon and fiberglass are alternated in the specimen. Three (directed along axes *X*, *Y*, *Z*) piezoelectric IPC accelerometers were attached at the specimen at each points 1, 2, 3, 4 (see Fig. 2), a total of 12 sensors, for registering acceleration.

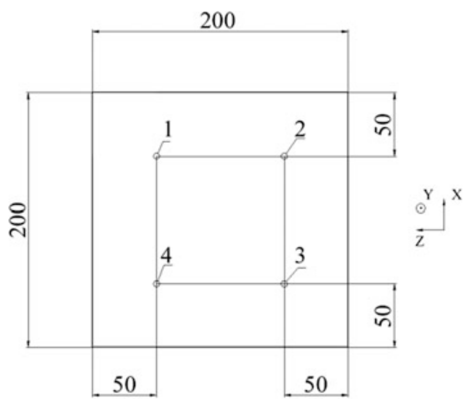


Fig. 2 Location and the directional axis of the sensor

The specimen was fixed on the edges. A hammer was used to strike and check force. The hammer was secured in a homemade impact machine for the replicability of frequency of occurrence of energies and angles of impacts. Impacts were made at random points with uniform distribution on the sample surface with different energies A, B, C at an angle of 90° to the surface. Also, the impacts were carried out with different heads on the hammer. The steel head made 10 impacts; a plastic head was made 230 impacts; and a rubber one made 160 impacts. Data logging was carried out using the LMS data acquisition system at a frequency of 51200 Hz.

3.4.1 A Dynamic System Construction

The approximation of properties of the specimen was conducted by means of restoration of properties of the corresponding dynamic system. For this purpose, the corresponding vectors of the phase space were built (1.7) $E(i, t) = (e(i, t), e(i, t + \tau),\ldots, e(i, t + \tau(D - 1)))$ at the moment of contact upon impact (3–10 ms).

The parameters τ, D are selected according to the maximum and minimum frequencies of the spectrum of sensor values at the moment of impact. A typical spectrum at the moment of contact upon impact of the tested plate is as follows (Fig. 3):

It is shown that significant signal frequencies are in the range from 600 to 6000 Hz. The time discretization was chosen with the focus on the theorem of Nyquist–Shannon [18]: To restore the dependency, it is necessary to take it equal to half of the minimum period of significant frequencies of the signal. Accordingly, $\tau = 0.5/6000\ \text{s} = 8.3\times10^{-5}\ \text{s}$.

Dimension of constant D was determined on the basis of the empirical assumption that the vector of the phase space should cover at least the minimum period of all significant frequencies of the signal.

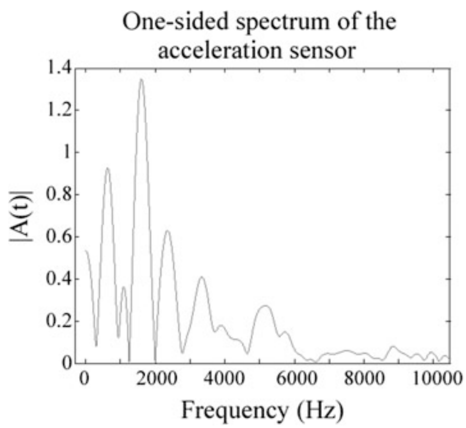


Fig. 3 Spectrum at the moment of contact upon impact, 3 ms

3.5 System Properties Restoration by Means of Neural Networks

The properties of the dynamic system were approximated by a back-propagation neural network [19]. The input of the network was supplied by vectors $E(i, t) = (e(i, t), e(i, t + \tau), \ldots, e(i, t + \tau(D - 1)))$. A total of $D \times N$ input neurons were used (D = 10 is the dimension of embedding, N = 12 is the dimension of the data vector at the particular moment of time, i.e., the number of sensors).

The output was supplied by the registered values of power F, the X, Y coordinates: a total of three output neurons.

Several different existing variants of back-propagation algorithm were tested. The best results (by far) were obtained during the training using the optimization of Levenberg–Marquardt [20]. The training was applied to 70% of the data of the entire set of data for inputs and outputs of the network, and testing was carried out on 30% of the data.

As a result of training, a neural network was obtained which is able to restore the history of loading at the moment of impact and the point of impact.

An example of a graph of restored power is shown at Fig. 4. The exact values of power cannot be restored; however, the general trend can be seen clearly.

Figure 5 shows the experimental and calculated coordinates of the impacts. The distance between the real and the estimated points is 14 ± 12 mm.

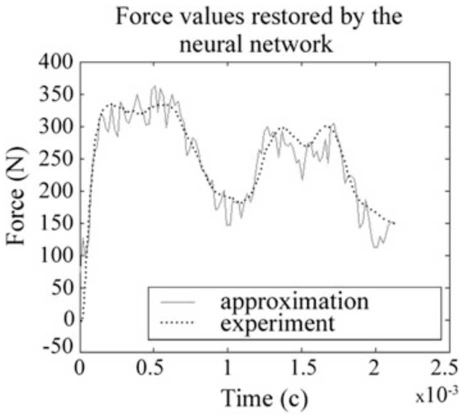


Fig. 4 Graph of power at the moment of impact restored by neural network

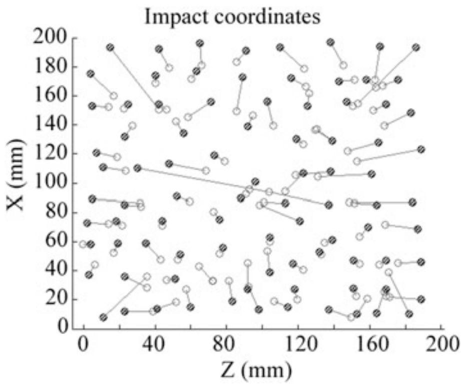


Fig. 5 Coordinates of the impacts (striped markers are experiment, white ones are calculation)

3.5.1 Restoration of Properties by the Support Vector Machine System

The support vector machine [21] is a promising algorithm for optimization and classification of the data that is rapidly developed recently. In this work, it was attempted to use the support vector machine for the purpose of approximation of properties of the achieved dynamic system. We used the library LibSVM in this work. The input of each machine was supplied by vectors $E(i, t) = (e(i, t), e(i, t + \tau), \ldots, e(i, t + \tau(D - 1)))$. The output of the first machine was supplied with power, the output of the second machine was supplied with X-coordinate values, and the output of the third machine was supplied with Y-coordinate values.

For the purpose of training, the supported vector machines with kernels were used:

- Radial basis function (RBF);
- Heterogeneous polynomial kernel.

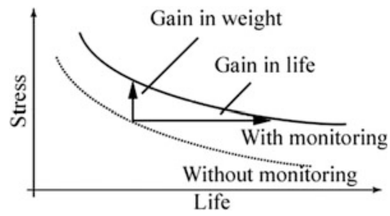
When using the machine with kernel-RBF, the convergence was very fast, but error at the end of training was much higher than in the case of a neural network (under fitting).

In case of polynomial kernel, the convergence rate was also very high. However, such machines have been overfitting (the error rate was too low on the learning set, and the high one was on the test set). Attempts to apply the cross-validation resulted in underfitting again.

Thus, it is necessary to conduct further researches of support vector machine, which has two advantages:

- High speed of convergence;
- The existence of local minima does not impede the determination of the optimal solution, which provides a higher reliability of the results.

Fig. 6 Benefits of HMS



4 Development of Methods of Damage Detection Based on Results of Measurements of Local Stress–Strain State Kinetics

4.1 Problem Formulation and Analysis of Existing Methods of Monitoring of Polymer Composite Materials Structures

Currently, the miniaturization and cost reduction of the measuring and computing technology has provided the possibility of application of embedded strain gage system that allows to determine the local stress–strain state accurately. Conventional strain gages as well as of becoming more widespread fiber-optic sensors of strain and temperature can be used. With their help, as the results of a number of studies show, it is possible not only to determine the service loading data of the structure, but also to discover the damages in the main load-bearing elements

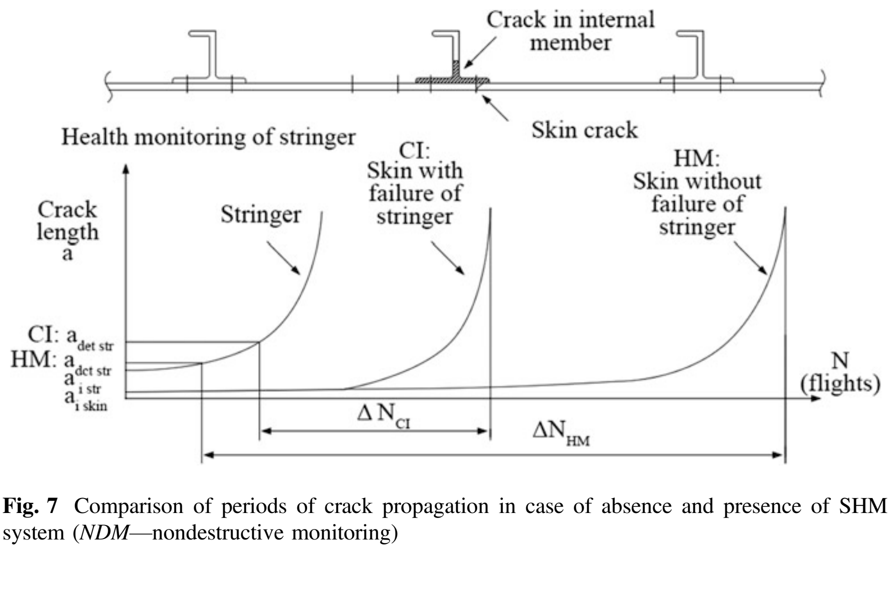


Fig. 7 Comparison of periods of crack propagation in case of absence and presence of SHM system (*NDM*—nondestructive monitoring)

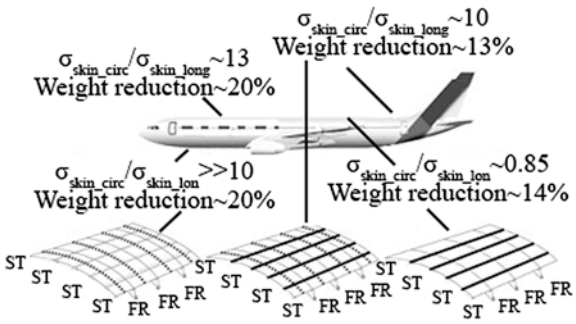


Fig. 8 Analysis of the possibilities of weight reduction for the various fuselage elements and the location of SHM sensors necessary for that [18]

of constructions. In this section, we examine works carried out in this direction and methods which can be used for their practical implementation.

The work [22] shows that information about the integrity of the main reinforcing element, for example, stringer, allows to increase the interval between inspections or reduce the weight (Fig. 6).

The increase of the interval between inspections is achieved due to the fact that the duration of crack propagation in case of non-failured stringer is substantially increased (Fig. 7) [23].

The advantage in weight is achieved due to the fact that if there is the SHM it is possible just to provide structural strength at much less lengths of the cracks.

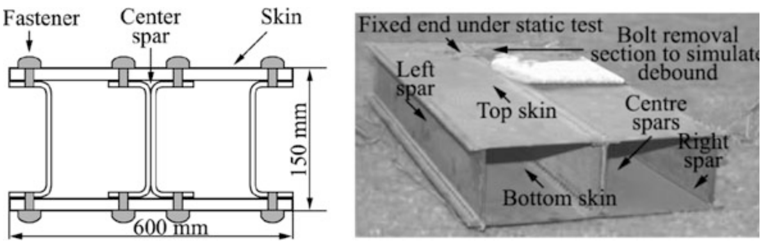


Fig. 9 Scheme (left) and photograph (right) of the examined fiberglass PCM construction with bolting connection

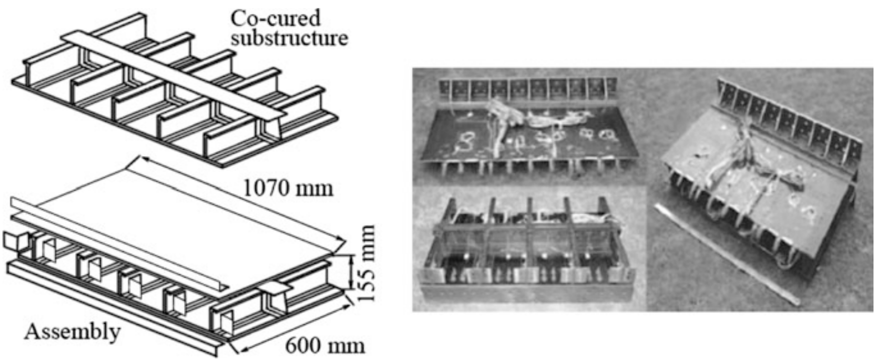


Fig. 10 Scheme (left) and photograph (right) of the examined carbon fiber PCM construction

Figure 8 shows the results of the analysis carried out for the wide-body passenger aircraft of a medium action radius which is made of aluminum alloy 2024. Sensors of monitoring of stringers' integrity are indicated by the solid lines, and sensors of monitoring of shell integrity are indicated by the dashed lines. Symbols: ST—stringer, FR—frame, $\sigma_{skin\_circ}/\sigma_{skin\_long}$—the ratio of circular stresses of the shell to longitudinal ones.

Along with monitoring of stringers' integrity and other load-bearing elements, it is relevant for PCM structures to monitor such specific for PCM types of destruction as the lamination. From this point of view, the subject of great interest is the result of work [24], which examines the ability to detect damage by monitoring local stress–strain state (Figs. 9 and 10) in the PCM construction.

Three types of specimens were examined: fiberglass PCM testboxes with bolted joint (Fig. 9) and adhesive connection, and also carbon fiber PCM testbox (Fig. 10). The loading type was the cantilever bend.

There were three options examined upon the fiberglass PCM testboxes with bolted joint:

- Integral structure: The testbox was tested with fully installed bolts to provide a basis for comparison with damaged constructions.
- Four bolts were removed. Four pairs of bolts were removed along the central spar near the fixed edge (Fig. 11). Since the distance between the bolts was equal to 20 mm, the length of lamination was 100 mm.
- Eight bolts were removed. Eight pairs of bolts were removed along the central spar near the fixed edge. The length of lamination was 180 mm.

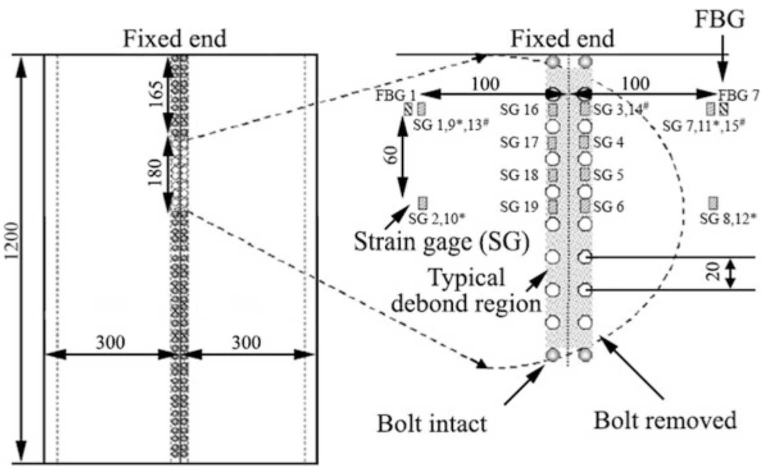


Fig. 11 Scheme of the lamination simulation of the fiberglass PCM structure with bolted joint (left). The location of the bolts and the scheme of sensors adhering (right)

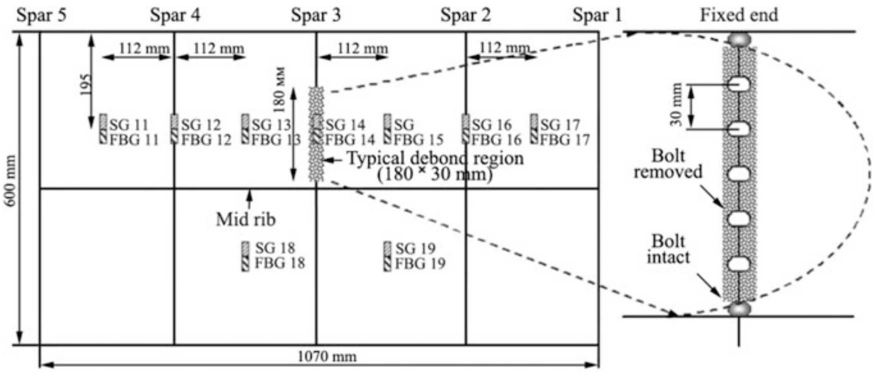


Fig. 12 Plan view of FBG and strain gage sensor location on testbox (left), bolt location on the spar for debond (right)

The width of lamination was equal to 30 mm—the width of the flange of the spar. In all three cases, deformation and displacement were measured. Figure 11 shows the schemes of the lamination simulation, the location of the bolts, and sensors' layout. The deformation was measured by both fiber Bragg grating (FBG) sensor and strain gages (SG).

One fiberglass PCM testbox with adhesive connection was made integral, and there was the simulation of 180 mm length lamination on the second one by means of the lack of adhesive connection in the same place where eight bolts were removed in case of the fiberglass PCM testboxes with bolted joint.

The carbon fiber PCM testbox was designed to test stiffness which is more typical for modern aviation structures and with the higher resistance to local buckling, particularly, due to the central rib. Lamination was simulated by

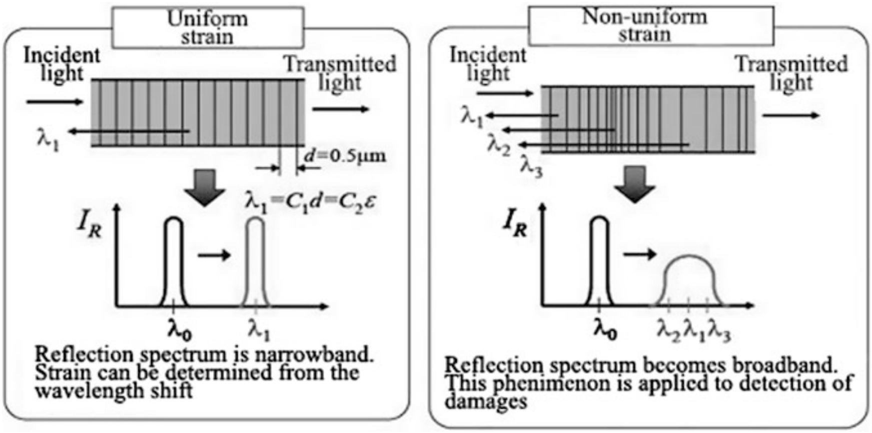


Fig. 13 Response of fiber-optic sensors with Bragg gratings under homogeneous and heterogeneous deformations

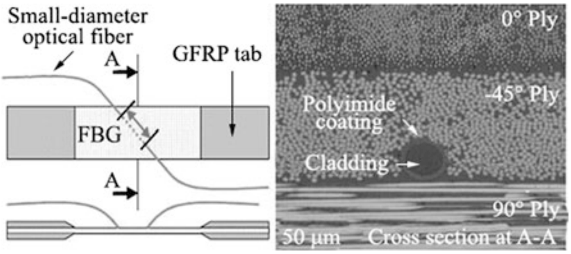


Fig. 14 Small diameter fiber-optic sensor, integrated into the layer −45° for the detection of transverse cracks in the adjacent layer 90° quasiisotropic laminate [45/0/-45/90]

removing five bolts consecutively (with a step of 30 mm), simulating 180 mm length lamination for different variants in three central spars (Fig. 12).

The main results that were obtained are as follows:

- In case of fiberglass PCM testboxes with bolted joint and adhesive connection, sensors located in the lamination zone allow to detect damage quite reliably, particularly, due to local buckling.
- In case of carbon fiber PCM testbox, sensors located in the lamination zone allow to detect damage quite reliably only if the detection is carried out by means of comparing sensor data in the same place of damaged and undamaged structure, i.e., not by absolute but by relative values of deformations.
- In addition to the deformation value, another damage indicator can be a violation of linearity between power applied and the deformation measured by strain gages as a result of local buckling (Fig. 13).

In addition, it was received:

- Removing the bolts on the fiberglass PCM testboxes allows to simulate delamination well enough; however, under the adhesive connection, local buckling occurs later and more abruptly.
- Measurement of deformation by fiber-optic sensors and strain gages gives alike results.

The subject of great interest is the work [25], which shows the possibility of detection of such defects of carbon fiber PCM as the lamination and transverse cracks due to the property of fiber-optic sensors with Bragg gratings to change the spectrum form of the output signal due to the appearance of heterogeneity caused by lamination and/or transverse cracks. However, it should be noted that the issues of practical use of this method require further study in order to provide monitoring not only of the area where the sensor is located, but also of the structure (Fig. 14).

4.2 *Monitoring of Multiple Site Damages of Structures Made of Aluminum Alloys*

Multiple site damages are the most dangerous type of damages in aircraft structure. Nowadays, designers of aircrafts are trying to prevent emergence of such damages at the stage of design [26]. Major aviation accidents connected with the emergence of multiple site damages forced air operators to set a lifetime of construction elements where the multiple site damages can occur, based on the principle of "safe life," which made the operation of such aircrafts much more expensive [27].

As a solution to this problem, there is an approach nowadays connected with such structure design when it is possible to confirm experimentally the impossibility of multiple site damages in it.

This approach contradicts to the concept which has been developed over many years of designing of aviation structures without significant differences between the durability in various locations, since such a design implies the absence of unreasonable reserves of the individual elements of airframe with high service life [28].

The use of monitoring of damages, which is able to provide multiple site damages detection in time, allows to set a lifetime of construction elements, where multiple site damages can occur based on the principle of "fail safe," which will make the maintenance of modern aircraft structure more cost saving without reduction in safety level during the operation.

The use of monitoring damage, which is able to provide timely detection of multiple site damages, provides the ability to assign resource for design elements with the multiple site damage, on the principle of "fail safe" that will make the maintenance of modern aircraft design more economical, not reducing the operation safety.

4.2.1 The Results of the Monitoring of Damages During Fatigue Testing of Longitudinal Joints

While determining the fatigue characteristics of elements of prospective passenger A/C, the authors of this work had carried out the testing of the damages monitoring by means of analysis of data of strain gages.

The prototypes of samples of the longitudinal joints of the fuselage, designed to obtain fatigue characteristics and determine the optimal structure of the connection, were used for the purposes of the experiment. All the samples were represented by two sheets of the same thickness, connected to stringer by three rows of rivets (Fig. 15).

During the testing, the strain was measured having different number of operating cycles, and the damages effect on the stress–strain state of the sheets was determined. The wire strain gages were used in the experiment, which were installed in two rows at the outer rivets on two sides of the inner and outer sheets. In this case, a large number of sensors used in the experiment were due to the examination of the size of the area where the changes of stress–strain state occurred as a result of emerging damages. In close proximity with the rows of rivets, due to the limitations connected with the available space for installation, small-base sensors with a base of 1 mm were used. In the regular part of the sheets, sensors with base size of 10 mm were used (Fig. 16, sensors with a base of 1 mm are shown in gray, and the ones with a base of 10 mm are shown in black).

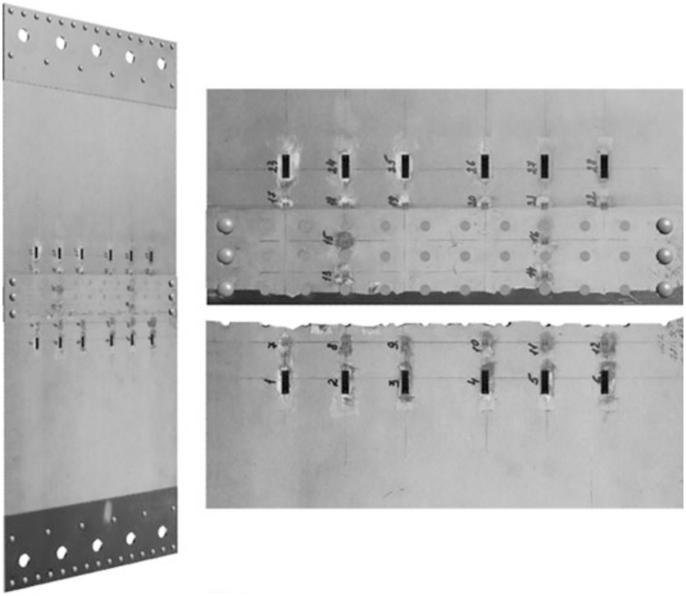


Fig. 15 Sample of the longitudinal joint of the fuselage with installed strain gages, and the specific destruction at an outer row of rivets

After the fatigue testing of the samples of longitudinal joints, the fracture surfaces of separated parts were carefully analyzed; as a result, the initial centers of cracks origin were identified, as well as the speed of cracks growth in different zones of the fracture and appearance of the fronts of cracks in different service time (Fig. 17). The received information was used to determine the zones of change of the stress–strain state by means of finite-element (FE) simulation.

The research of the longitudinal joint samples showed the following results:

- The destruction of samples in the longitudinal joint in most cases occurs in cross section that goes through the middle of the outer row of rivets on the outer connection sheet. Macropicture of fractures of the joint indicates that the emerging fatigue fractures related to the initially formed fatigue cracks are concentrated in the area of several holes. Some centers of small fatigue cracks observed in particular places of the damaged critical sections of the sample are secondary. Their initiation is triggered by increase of existing stresses due to the reduced area of the sample effective cross section during the propagation of the initially formed fatigue cracks. The initiation of all the fatigue cracks occurred on the surface of the outer sheet of the shell that contacts through the sealant layer with the inner sheet made of the same material. During almost all the period of propagation of the initial fatigue cracks, they evolved as blind cracks and could not be visually detected until the cracks became through and reached the length of reliably detectable size on the surface of the shell inverted outward

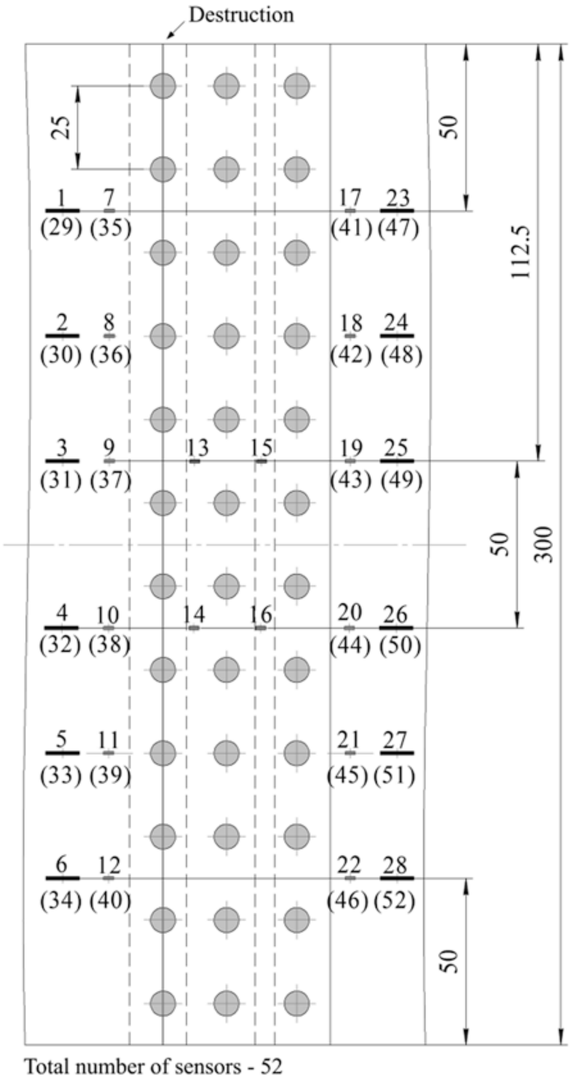


Fig. 16 Scheme of sensors' layout

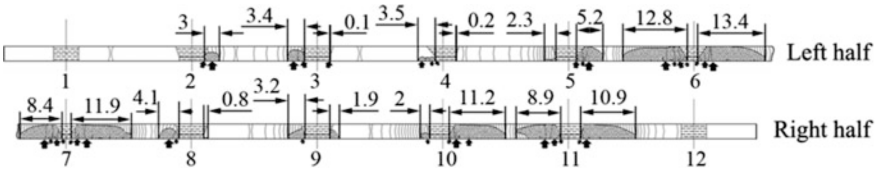


Fig. 17 Scheme of the fracture of the sample and size of the cracks (table of symbols is in the Table 2), received by means of fractography

Table 2 Table of symbols used in the diagrams of the fracture

	Arrow indicates the location of the fatigue fracture focus and the direction of propagation of closed fatigue cracks
	Fatigue zone where the crack rate is less than 5 mkm/cycle
	Fatigue zone where the crack rate is more than 5 mkm/cycle
	Rupture zone
	Hole

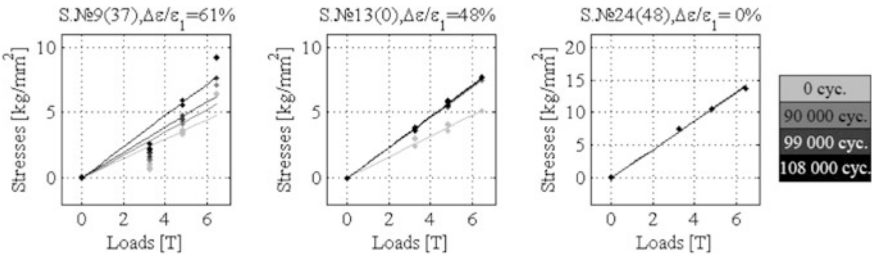


Fig. 18 Deformation change in different areas of the structure

of the fuselage. The overall unsatisfactory assessment, which characterizes possible maintenance of used longitudinal fuselage joints for damage tolerance, is caused by the poor inspection ability of the connection sheets by means of a visual control.

- Changes in the strain gage readings registered during the experiment can be divided into three groups. The first group includes the sensors located near the place of future destruction. The registered change in the strain gages readings in this group ranges from 10% to 50% and is found in all places across the width of the outer sheet used for installation of the sensors (Fig. 18, left graph). The change in the sensors' readings of this group is a result of the emergence of fatigue cracks in the structure. The second group includes sensors installed between the rows of stay bolts. There was a change in readings of this group of sensors as well, which is most likely caused by the changes in allocation of the stresses in the distances between the rows of rivets during the operating time of structure (Fig. 18, middle graph) A significant change in readings occurred in the interval between the first and second strain measurements, and further indications of the strain gages were on the same level. The third group of sensors is the strain gages which are maximally distant from the damage location. There was no change in the sensors readings of this group (Fig. 18, right graph).

The conducted experiment shows that in cases when the examination of the state of the structure by means of visual monitoring is impossible, damages in the structure can be detected by means of sensors measuring parameters of local stress–strain state.

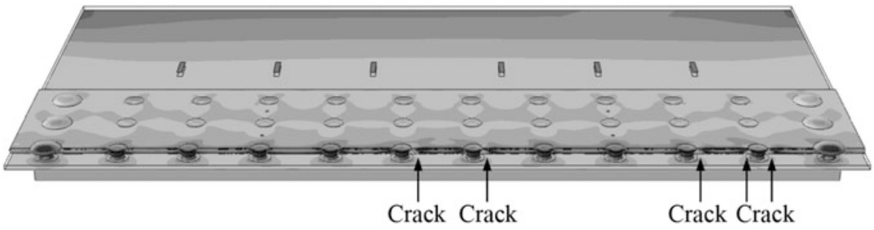


Fig. 19 Appearance of the destroyed section with a simulated crack

4.3 Analysis of the Kinetics of Stress–Strain State in the Area of Damage by Means of a Finite-element Model of the Structure

For analysis of the kinetics of the elastic deformations field, finite-element modeling was conducted by means of several FE models of three-row double-shear connection in the software package ABAQUS. The models consider the data about location of the fronts of fatigue cracks obtained during fractography of the samples during the fatigue life, when strain measurement was carried out. The model of the upper shell sheet was divided into three parts for the purpose of a correct subdivision of the simulated objects into the elements. In the geometry of the middle part,

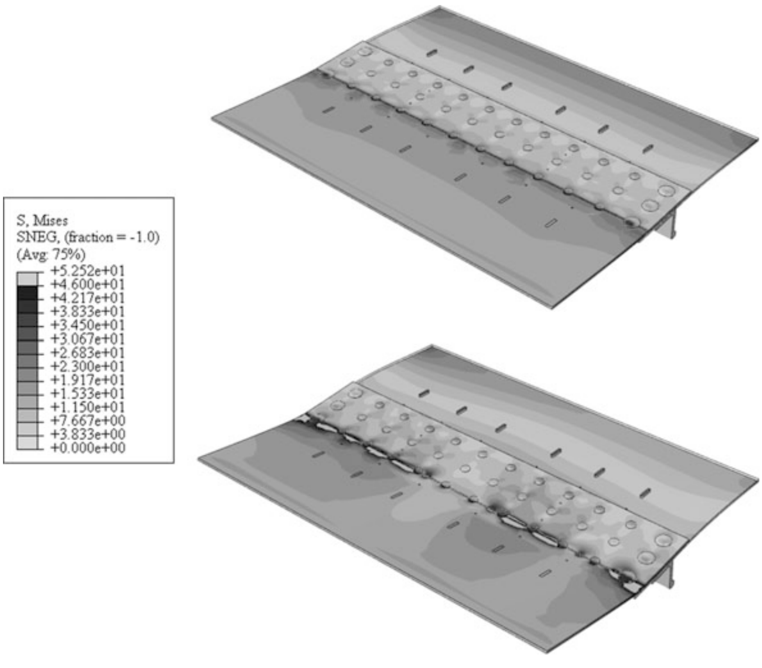
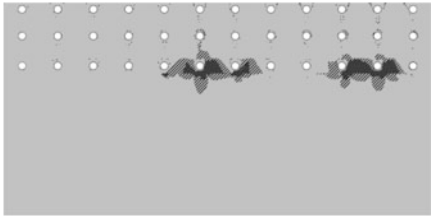
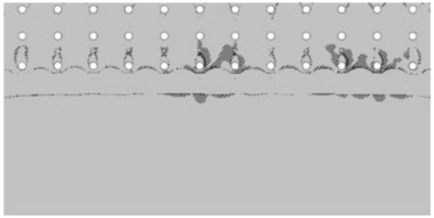
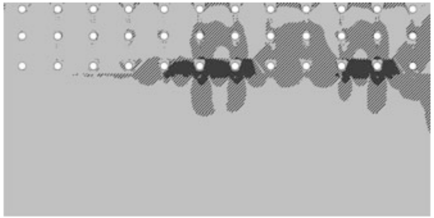
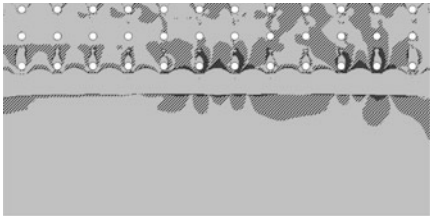
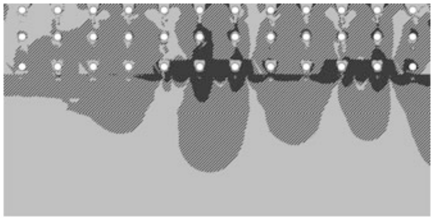
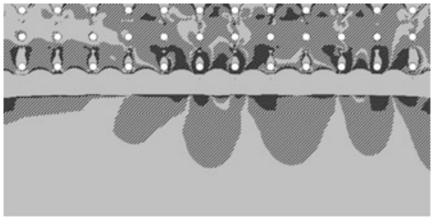


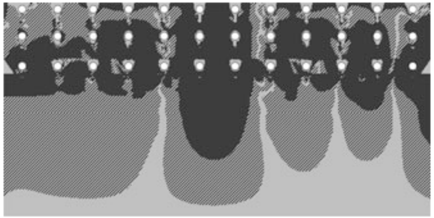
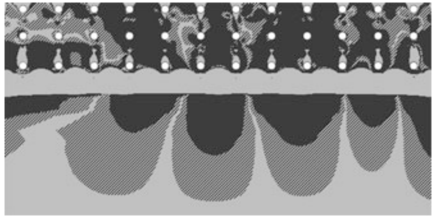
Fig. 20 FE model of the joint, calculated as for undamaged (top) and for fatigue damaged (bottom)

Table 3 Stress–strain state kinetics of the outer sheet connection

Lower surface of inner sheet	Upper surface of outer sheet
Joint fraction scheme by N = 90,000 cycles 	
	
Joint fraction scheme by N = 99,000 cycles 	
	
Joint fraction scheme by N = 108,000 cycles 	
	

(continued)

Table 3 (continued)

Lower surface of inner sheet	Upper surface of outer sheet
Joint fraction scheme by N = 113,500 cycles 	
	
Difference, $\Delta\varepsilon$ [%]	

the contours of the fronts of fatigue cracks were used which were received by means of analysis of the fractures of the tested samples (Fig. 19).

Figure 20 shows the change in the stress–strain state of the longitudinal fuselage joint which was reproduced in FE models corresponding to the initial (undamaged) and the final (before destruction) state of the joint.

The kinetics of the stress–strain state of the joint is shown in Table 3, which illustrates the change of the elastic stresses field in comparison with the initial state: Light gray is no change, shaded is a change of not less than 5%, dark grey is a change of not less than 30%. In the analysis of the kinetics, it was used the data about deformations obtained on the undamaged and damaged models taking into account the size of damages corresponding to the appropriate operating time.

According to the analysis of the kinetics of stress–strain state, you can make the following conclusions:

- Extensive zones, where changes of stress–strain state have occurred, are fixed similarly well on the upper and the lower surface of the destroyed sheet;
- The damage effect can be fixed in advance. There are several variants of sensors' layout in the area of stress–strain state changes;
- When the complete destruction of the joint would be not less than 15 thousand flight cycles, the region of maximum change of stress–strain state covers about two-thirds of the width of the top sheet and is located in the third row of the outer sheet;
- For the type of fatigue fracture with multiple sites of crack initiation investigated in the experiment, it can be defined and compiled the optimum sensors' layout scheme based on the analysis of strain gages and fractography of fractures by means of FE simulations, which do not require control of the stress–strain state of a large number of places. It will guarantee the early detection of fatigue damage.
- An estimated number of required sensors, which is necessary for the detection of damages of this type, is one monitoring location for every 100 mm of width of the sheet.

The examination of the kinetics of the stress–strain state showed that the monitoring of local stress–strain state can be used in the creation of a full-sized monitoring system. This measure is intended to expand the perspectives of the whole complex of methods providing the safety and maintenance of airworthiness of civil aircraft (Table 3).

5 Conclusion

This chapter gives a short review of the various methods of monitoring the condition of aircraft structure elements, of various sensors used in modern health monitoring systems and areas of sensor's application.

It is shown that localization of impact damage on composite structures by means of the analysis of the strain gages' readings is possible. The best result in detecting the location and force of impact is obtained through the use of neural networks.

The paper presents the results of the research and application of monitoring techniques based on monitoring local stress–strain state for the detection of fatigue damage of the elements of modern aircraft structures. The effective use of strain sensors was confirmed experimentally for monitoring damage of the structures from polymer composite materials and aluminum alloys with multiple sites of crack initiations. The paper shows that the inability of visual detection of fatigue damage of a longitudinal joint of the fuselage timely can be compensated by the analysis of local stress–strain state using a small number of strain gages installed on the structure.

To determine the parameters of sensors' layout scheme for monitoring the longitudinal joint, this work represents the analysis of the kinetics of stress–strain state calculated by FE model of the structure based on the detected fatigue damages. The analysis shows that the estimated number of sensors should be based on the assumption about the location of at least one sensor for every 100 mm along the longitudinal fuselage joint or alternatively one sensor should be placed for every five holes near the outer joint row.

References

1. Chang F-K (2005) Structural health monitoring 2005: Advancements and challenges for implementation. DEStech Publication Inc, Lancaster
2. Speckmann H, Roesner H (2006) Structural health monitoring: a contribution to the intelligent aircraft structure. In: Proceedings of 9th European NDT conference, Berlin, Germany, 25–29 Sept 2006
3. Coppinger R (2007) Airbus and Boeing back structural monitoring. http://www.flightglobal.com/articles/2007/02/20/212184/airbus-and-boeingback-structural-monitoring.html
4. Bartelds G (1997) Aircraft structural health monitoring, prospects for smart solutions from a European viewpoint. NLR TP 97489 A National Aerospace Laboratory NLR 13
5. Boller C, Buderath M (2007) Fatigue in aerostructures-where structural health monitoring can contribute to a complex subject. Philos Trans R Soc Lond A 365:561–587
6. Bolotin VV (1990) Longevity of machines and structures. Mashinostroenie, Moscow (in Russian)
7. Whittingham RB (2004) The blame machine: why human error causes accidents. Elsevier Butterworth-Heinema, Burlington
8. Ignatovich S et al (2011) The perspectives of using on-board systems for monitoring of residual life of airplane structures. Vestnik TNYU (mechanics and material science). Kiev Special issue 2, 136–143 (in Russian)
9. Gardiner G (2015) Structural health monitoring NDT-integrated aerostructures enter service. Composites World, 46–49
10. Fengming Yu et al (2014) Identification of damage types in carbon fiber reinforced plastic laminates by a novel optical fiber acoustic emission sensor. In: Proceedings of the European workshop on structural health monitoring (EWSHM)

11. Gëuemes A et al (2014) Damage detection in composite structures from fibre optic distributed strain measurements. In: Proceedings of the European workshop on structural health monitoring (EWSHM)
12. Mayer RM (1996) Design of composite structures against fatigue. Applications to wind turbine blades, ISBN 0 85298 957, 11–14
13. Coelhol CK, Hiche C, Chattopadhyay A (2010) Impact localization and force estimation on a composite wing using fiber bragg gratings sensors. In: 51th AIAA/ASME/ASCE/AHS/ASC structures. structural dynamics and materials conference, Orlando, 2010
14. Park J, Ha S, Chang F-K (2009) Monitoring impact events using a system-identification method. Stanford University California 94305. Stanford. doi: 10.2514/1.34895
15. Tajima M, Ning Hu, Fukunaga H (2010) Experimental impact force identification of CFRP Stiffened Panels. J Jpn Soc Comp Mater doi:/10.6089//jscm.35.106
16. Ricordeau J, Lacaille J (2010) Application of random forests to engine health monitoring. Paper 730 ICAS. In: Proceedings of 27th international congress of the aeronautical sciences, 2010
17. Pankov AV, Svirskiy Yu A, Sorina TG (2015) The main results of researches on creation of onboard systems monitoring (SHM), including the use of fiber-optic sensors. Report on the international round table of the SKOLKOVO Institute of Science and Technology, effective methods for forecasting resource. Moscow. 14–15 May 2015 (in Russian)
18. Nyquist H (1928) Certain topics in telegraph transmission theory. Trans AIEE
19. Bryson AE, Denham WF, Dreyfus SE (1963) Optimal programming problems with inequality constraints. Necessary conditions for extremal solutions. AIAA, 2544–2550
20. Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J App Math 11:431–441
21. Smola AJ, Scholkopf BA (2004) Tutorial on support vector regression. J Stat Comput 14 (3):199–222
22. Schmidt HJ, Schmidt-Brandecker B (2009) Design benefits in aeronautics resulting from SHM. In: Boller C, Chang F-K, Fujino Y (eds) Encyclopedia of structural health monitoring. Wiley, Chichester. ISBN: 978-0-470-05822-0, 1807-1813
23. Boller GA et al (2009) Commercial fixed-wing aircraft. In: Boller C, Chang F-K, Fujino Y (eds) Encyclopedia of structural health monitoring. Wiley. ISBN: 978-0-470-05822-0, 1917-1932
24. Sundaram R, Kamath GM, Gupta N, Subba RM (2010) Damage studies in composite structures for structural health monitoring using strain sensors. Advanced Composites Division, National Aerospace Laboratory, Bangalore
25. Takeda N (2008) Recent development of structural health monitoring technologies for aircraft composite structures. In: ICAS Proceedings of 26th international congress of the aeronautical sciences, Paper 706 (2008)
26. Nesterenko BG (2011) The requirements for the fatigue and durability of structures for civil aircraft. Scientific Bulletin of MSTU. Moscow, pp 49–59 (2011) (in Russian)
27. Nesterenko BG (2007) Fatigue and durability of longitudinal joints obshivka pressurized fuselages. In: Scientific Bulletin of MSTU. Moscow. pp 70–81 (2007) (in Russian)
28. Rajher VL, Selikhov AF, Khlebnikov IG (1984) Records of a plurality of structures critical indicators in the evaluation of durability and resource. In: Scientific notes of TSAGI XV. Publishing Department of TSAGI, Zhukovsky (in Russian)

Machine Signature Integrity and Data Trends Monitoring, a Diagnostic Approach to Fault Detection

Michael Kanisuru Adeyeri

Abstract This chapter proposes machine signature integrity and its violation detection as an approach to fault diagnostic. It uses machine history and its present (in situ) data through monitoring as potent tool in detecting incipient faults. This monitoring program is a function of set parameters that put in place as watch dog to provide signals or alert whenever there is fault initiation on the machine system. A robust flowchart on how fault could be detected, isolated, and identified along with its algorithm for the diagnostic program inherent on vibration-induced faults is presented, and if these algorithms are rightly appropriated, fault will not only be detected, but isolated and identified in any production system or manufacturing.

1 Introduction and Definition of Key Terms

Isermann [1] defined fault as "*an unpermitted deviation of at least one characteristic property of a variable from an acceptable behavior.*" Faults are disturbance to the machine systems which are not meant to be tolerated or endured. It is the outcome of system alteration and impairment. Whenever there is fault, be it minor or major, job shop or plant managers are not always happy because of many things that are at stake. Production processes, production hours, and inventory cost etcetera are being endangered. Thus, leading to the initiation of customer dissatisfaction on products and in the long run, the enterprise will start losing valuable customers, running at loss and will not be able to stand competitively with the world class leading manufacturing industries. To corroborate this, Yen [2] affirmed that fault assessment in machinery is characterized by robust control architecture.

M.K. Adeyeri (✉)
Department of Mechanical Engineering, The Federal University of Technology Akure, Akure, Nigeria
e-mail: adeyerimichaeltut@gmail.com; michael.adeyeri@imes.uni-hannover.de; mkadeyeri@futa.edu.ng

M.K. Adeyeri
Institute for Mechatronics Engineering Systems, Leibniz University, Hannover, Germany

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*,
Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_2

In view of this, there is need to have understanding about the science and mechanics of the industrial machinery which are inborn in the machine, hence its signature.

The machine signature is defined as the machine itself. It is likened to a scenario of someone who issued out a cheque to a third party, in accessing the amount of money written therein in the bank, at the cite and confirmation of the endorsed signature in line with issuer's name identity in the bank data base, even at his or her absence, the cheque would be honored. It is honored because the signature on it represents the presence of the account holder.

In the same vein, the signature of a machine defines the entirety of the machines. It defines its geometry, mechanism, operational behavior to vibration, thermal behavior, shock pulse attribute, acoustics, and sounds at no-load and on-loading characteristics. In a nutshell, these are inherent features exhibited in the machine for its identification. It is through this signature that assistance can be received in identifying various types of machines when used. Listening to the sound from various group of machines should be able to give distinct characteristics instinct and clue that this is airplane, that is helicopter and this is hammer mill.

Let us consider this scenario. In a plain description, in a group of ten (10) boys who are very close in relationship. If someone is picked, blindfolded, and saddled with the task of identifying the remaining nine (9) friends by name as the nine boys are engaged in discussion, without mincing word and mismatching of the voice character with individual names, the blindfolded boy will correctly identify the nine boys. Why? Because the voice signature produced by each of them over time in the course of their relationship has been registered in the memory database of the identifier (the blindfolded boy).

Another key word that needs to be explained is integrity. Integrity as a word discusses the undiluted, unpolluted behavior, or uncompromised signature of what something is known for. On a broader note, if "machine A" is known to process a 10 kg load of wheat in ten (10) minutes on the platform of ceteris paribus. Then four (4) days later, if such "machine A" is given another 10 kg of wheat to process, based on its integrity, it is expected that after ten (10) minutes the whole wheat would have been processed. To elucidate further, if the same "machine A" is now given 20 kg of wheat to process, drawing inference from its previous trend of activities and actions on wheat processes through the use of model formulation of its behavioral act, as expressed in model Eqs. (1) and (2).

$$t \propto m \tag{1}$$

where t is the process time and m is the mass of the wheat.

Introducing constant of proportionality k will transform Eq. (1) to:

$$t = km \tag{2}$$

With the scenario explained above, k is found to be 1.

Based on this model, it is expected that at the end of twenty (20) minutes, such "machine A" would have completed the given task of processing twenty (20) kg of wheat.

On the contrary, if such expectation of time from the "machine A" now extends beyond twenty (20) minutes to twenty-five (25) minutes, it is an indication that the machine integrity has been compromised due to one reason or another. Furthermore, it is possible that with the twenty (20) minutes spent, the expected grain size particle output may not be reached or attained.

These are the issues to be discussed in this chapter. The chapter is to identify why a given task meant for twenty (20) minutes has extended to twenty-five (25) minutes.

This is now called for the fourth key term of the topic that says "trends monitoring." Through the monitoring of the behavioral pattern(s) of the machines, it can give clue to how such faults could be detected, isolated, and identified.

There are facilities that are meant for these, forming basis for the ease of machines' condition monitoring. These facilities are meant to assist in machine data collection and its analysis. Data are needful because when processed, viable information is got and interpreted. The facilities and its setup shall be discussed in the following subsection.

Diagnostic is the process of discovering the unsatisfying status condition of an object through scientific analysis based on training of background historical data of the subject under consideration of investigation. However, Wang et al. [3] opined that diagnostics is carried out to know the reason behind the occurrence of the degradation in machines. In human medicine, the diagnostic approach in knowing or detecting what is wrong with human body and cells involves critical examination through various tests ranging from blood count test, body temperature measurement, blood pressure reading etcetera with the use of sophisticated medical laboratory diagnostic tools.

Likewise in industrial machinery systems, to be able to identify the causative agent behind system alteration from its initial signature state requires the use of many tested and trusted diagnostic tools which shall be considered briefly in the following subsection.

The underlining effect calling for this research is to address the catastrophic nature of damages and disturbance befalling the industries through incessant failures resulting from unidentifiable incipient faults. A stratified fault identification framework is proposed to meet the challenging needs of manufacturing based on training of machine integrity on past and in situ machinery data.

2 Literature Review

The basic strategies for fault identification in industrial machinery have been in place for over 3–4 decades ago. Prominent scholars had prevalently made meaningful contributions to this industrial menace. Some of the literatures consulted in the course of preparing this chapter are as stated by extracting the cogent contributions and examining areas for which the newly proposed approach could be of help in proffering solution to faults diagnosis in industrial engineering.

Mehra and Peschon [4] postulated an innovative approach to fault detection and diagnosis in dynamics systems by outlining the approach for establishing fault detection, its root cause, and prognosis with the introduction of mathematical models in affirming its postulation. With the duo's input, it is an indication that the struggle for fault identification has been on for a while.

Isermann [5] worked on fault diagnosis of machines through the use of parameter estimation, parametric physical coefficient, fault sign tree, and knowledge processing. Isermann [1] further posited that fault detection and its diagnosis is hinged on the formulation of robust dynamic model using the input and output signals of the systems. Isermann's work made use of dynamic process models on the systemic appropriation of the input and output signals generation from system being observed. The framework laid has its root on parametric estimation and equations' parity under the use of actuator, passage car, and a combustion engine.

Yen [2] researched on health monitoring of vibration signatures in rotor wings. His proposition was on the design and evaluation of onboard rotor craft machines using intelligent assessment tools. The aftermath of the proposition resulted in having a system that could house incipient faults, detect it, and as well identify it.

Frank and Koopen-Seliger [6] unveiled the usefulness and application of fuzzy logic and neural network to fault diagnosis. The work done entails the use of residual concept on generation and evaluation analysis all knitted on the fuzzy logic tools and "knowledge observer."

Benbouzid [7] discussed the review of induction motor signature analysis as a medium for fault detection. His work only gave an informative and educative resume of induction motors' signature analysis by laying foundation of the theory, implication, and outcome of using motor signature tools in determining faulty and degraded status of mechanical and electrical component wrapped on spectrum threshold of the stator current.

Venkatasubramanian et al. [8] explicitly reviewed the methods involved in fault diagnosis through the proposition of inherent characteristics for diagnostic system. Structural work for determining and comprehending diagnostic strategies based on qualitative model-based methods, process history-based record, and their limitations were as well documented.

Gang and Dahiya [9] worked on current signature analysis and its application in the condition monitoring of wind turbine for rotors. The work was hinged on the measurement of characterized frequency at excitation level of the wind turbine generator. The excitation produced was measured and analyzed at no load, full load, and different stages of wind conditions using the fast Fourier transform algorithm. Their submission was that the key approach to fruitful diagnosis is a function of putting up the right strategy in establishing frequency range and diagnostic index between normal and degraded condition.

Staroswiecki [10] researched on quantitative and qualitative models for fault detection and isolation. His research employed the use of differential informative knowledge of the system being observed by integrating both numerical and qualitative models in one entity. The explanation regarding the model principles for the fault detection and isolation was highlighted. But the work was limited by size as it cannot be applied on a large-scale system.

Kahraman et al. [11] gave a clear application of fuzzy sets in industrial engineering using topical classification approach. The work explicitly gave the survey and how the fuzzy set theory could be applied in industrial engineering in solving related faults challenges.

The formulation of mechanistic models for degradation and failed propagation arrest in rolling bearings was revealed in the work of Kurfess et al. [12]. Their work posited that the formulated models would be helpful in diagnostic and prognostic analysis of industrial system.

Rajakarunakaran et al. [13] used artificial neural network for the detection of faults in rotary system. The authors employed feed forward network algorithm and binary adaptive resonance of the artificial neural network phenomenon. The implementation of the proposition was carried out on centrifugal pumping machine encompassing the detection of seven (7) different faults ranging from shaft and bearing problem, leakage problem, impeller and mechanical problem, operability and vibration problem. The methodology on case-based reasoning system of industrial diagnosis was borne on Chebel-Morello et al. [14]. The methodology addressed diagnosis analysis based on “knowledge-oriented system” built on artificial reasoning.

Mendonca et al. [15] gave the architecture needed for fault detection and isolation using fuzzy methods hinged on tree search concept, residual computation analysis, and fuzzy decision factors. In another perspective, Ierace et al. [16] opined that electric signature is a potent tool if properly used will be a better and as well a cheap diagnostic and prognostics tool for fault detection and industrial machine upkeep. This proposition of Ierace was implemented on a vending machine to determine the economic importance of its quality, and it was submitted that the approach was quite relatively cheap.

Yin et al. [17] gave a comparative study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman Process. The work was data-driven meant to monitor and determine associated faults in the system. The outcome of the work has become industrial reference for monitoring processes. Also, Yin et al. [18] posited another theory that a well formulated data-driven design could lead to a robust fault detection system. This premise was used on wind turbines by harvesting the available residual data of generators. The robustness of the design was captioned on the performance index and optimization provision in the generic framework developed. Nonparametric dimension reduction algorithm was proposed by Chamay et al. [19] as another approach of neural networks application that could be used for diagnostic analysis. Their concept works on algorithm that reduces the dimension of the system-retrieved data to be trained for the neural application. This proposition was confirmed by its implementation on wind power machine. Their findings revealed that if and only if classification error rate is to be achieved, and only six input variables would be required for the artificial neural network.

Skelton [20] presented a synopsis on delivery reliability improvement and efficiency through emerging diagnostic techniques using Powercor as a base study. The work made use of historical details and the acquaintance interaction in the industrial field of Powercor. The findings established the probable cause of the fault associated with power poles and gave recommendation on how the pole structure should be.

Chamay et al. [21] developed a diagnostic system using linear predictive coding/cepstrum analysis in machine vibration purposely for diagnostics analysis. The development was on the use of linear predictive coding procedures in finding abnormality of failed components in engines' assembly line. Their findings revealed that the linear predictive coding cepstrum diagnostic tools are at advantage over the "Euclidean distance approach" being used in detecting faults.

Islam et al. [22] reviewed the condition monitoring techniques and it diagnostic tests for lifetime estimation of power transformers. Their work further gave the summary of what is obtainable in the power industry especially on transformer by establishing the strengths and weaknesses of all the various methods in vogue using the "health index" life estimation and probability "failure criteria."

Other documented related researches worthy of note in this chapter are on: overview of process fault diagnosis [23]; integrated diagnostic process for automotive systems [24]; failure diagnosis of complex systems [25]; data-mining approach for machine fault diagnosis based on associated frequency patterns [26]; review of condition monitoring parameters for fault diagnosis of fixed gearbox [27]; fault detection and classification in automated assembly machines using machine vision [28]; and state of the art and taxonomy of prognostics approaches, trends of prognostics applications, and open issues toward maturity at different technology readiness levels [29].

Furthermore, all other classical authors who had toiled greatly with their meaningful contributions in the niche area of fault diagnosis are as tabulated in Table 1.

With the aforementioned literatures on the enormous works that had been done in fault diagnosis, the approaches employed had been on fault detection and isolation using neural network, data-driven model, symptom-based model, qualitative model, quantitative model, fault-based model, normal-based model, process history, statistical knowledge, real-time analysis, vector support machine, auto regression, and just to mention a few. There is still a gap yet to be addressed in the area of machine signature integrity, and few literatures documented in this area do not give the holistic concept of it. Therefore, this chapter will proceed further in giving the algorithm of how machine signature integrity could be harnessed as a tool in fault diagnostic analysis.

Table 1 Summary of contributors who had worked on fault diagnosis-related research

S/n	Authors	Title of work done	Methods used/findings and contributions
1	Rao [30]	Condition monitoring and the integrity of industrial systems	It led to the introduction of condition monitoring and diagnostic engineering management (COMADEM)
2	Huang and Wang [31]	Related work on machine monitoring and diagnostics	Mainly a review work on signal processing tools, artificial intelligence, artificial network, and sensors
3	Stefanoiu and Ionescu [32]	Fuzzy statistical reasoning in fault diagnosis	Nonconventional approach was used. The approach initializes the automation of human reasoning integration into detection and classification of defects. And it was implemented on bearings
4	Dai et al. [33]	Machinery vibration signals analysis and monitoring for fault diagnosis and process control	Application of vibration for locating machine fault
5	Bongers and Gurgenci [34]	Fault detection and identification for longwall machinery using SCADA data	Application of data in detecting fault Use of semi-autonomous concept was adopted in addressing constraints in retrieving detailed equipment record on deterioration and failure
6	Sobhani-Tehrani and Khorasani [35]	Fault diagnosis of nonlinear systems	Encompassing fault diagnosis technique was introduced. Hybrid framework formulation for fault detection isolation
7	Peng et al. [36]	Current status of machine prognostics in condition-based maintenance: a review	Generic models on machinery prognostics analysis were formulated for the condition-based maintenance strategy identification (FDII) which was achieved
8	Cholewa et al. [37]	Diagnostic methods	It uses dynamic neural model hinged on bounded error Its parameters are based on knowledgeable information It is a belief-network-based diagnostic model
9	Qin [38]	Survey on data-driven process monitoring and diagnosis	Review work on state-of-the-art strategies used for fault detection
10	Hwang et al. [39]	A new machine condition monitoring method based on	Likelihood-based change of stochastic model hinged on live

(continued)

Table 1 (continued)

S/n	Authors	Title of work done	Methods used/findings and contributions
		likelihood change of a stochastic model	machine data to overcome previous defects likelihood was proposed It uses pretrained hidden Markov model The concept was implemented on machine and weld status condition
11	Wang et al. [40]	Health monitoring of engineering systems	Introduction of basic concepts in analysis and classification of faults and degradation prognosis and class function
12	Wang et al. [3]	Spectral kurtosis for fault detection, diagnosis, and prognostics of rotating machines: a review with applications, mechanical systems, and signal processing	Based on kurtosis concept
13	Tidriri et al. [41]	Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: a review of researches and future challenges	Model-driven fault diagnostic and health monitoring concepts for fault location and identification It enhances high accuracy of fault detection
14	Vog et al. [42]	A review of diagnostic and prognostic capabilities and best practices for manufacturing	Proposition of sustainable prognostic and health management system
15	Ogidi et al. [43]	Fault diagnosis and condition monitoring of axial flux permanent magnet wind generators	Axial flux fault was addressed stemming out from static eccentricities and circuits failure

3 Introduction to Machine Fault Diagnosis

Industrial machines and their components are bound to be faulty once they continually perform their day-to-day operations as the need arises. Fortuna et al. [23] opined that whenever a fault occurs, there are actions that needed to be taken in order to rectify such industrial operation interruption. The actions in diagnosing faults are strategically sequential and phased as found in Wang [40], Vachtsevanos et al. [44], Fortuna et al. [23] and Samantaray and Bouamama [45]. As shown in Fig. 1, the three phases identified with fault diagnosis are: fault detection state, fault-isolation state, and fault identification sate.

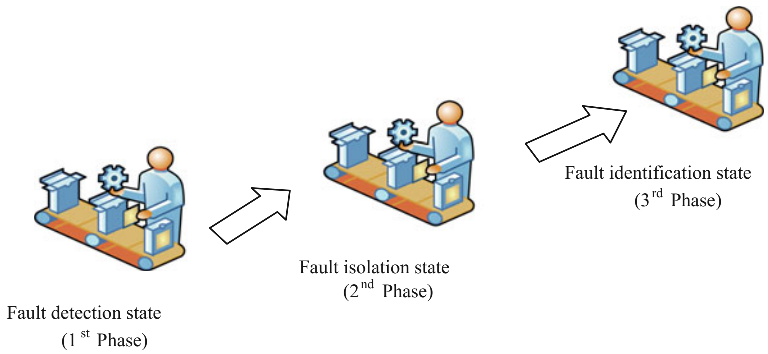


Fig. 1 Stages in fault diagnosis

The first phase which is fault detection state is to detect, and in other words, it connotes recognizing the initiating time that the fault begins to commence. As this is recognized, the next phase is to isolate (which is fault-isolation phase). Fault isolation implies establishing the nature, spot of occurrence, and period in which the fault occurs. After which, the fault will now be identified. Fault identification calls for knowing the degree, extent, and time transformation characteristics from which a normal system becomes being degraded or failed.

Faults are classified into three classes namely: abrupt fault; incipient fault; and intermittent fault. The detailed theory and characteristics of these classifications could be found in Vachtsevanos et al. [44], Fortuna et al. [23], Wang et al. [40], Yen [2] and Mendonca et al. [15].

There are several tools based on scientific proof which could be used for fault detection. Fortuna et al. [23] categorized these approaches as: parity check (checking the consistency of the parity through the model describing the industrial machine); observer-based scheme (processed harvested data from observation point of view are used in redesigning of the system); and parameter estimation (which works on the physical parameters as it is a reflection of the of the actual state of the system). Therefore, allowing residual inference to be drawn by using the comparative differences between the nominal and the estimated parameters. Thus, giving hint on the detection of the system fault.

While the fault diagnosis analysis techniques are inherent in the use of any or combination of these three tools, namely [44, 46]:

i. Data-driven-based tool: The approach is characterized with the use of sensor to harvest data read, and the data are trained for inferences with the help of decision logic. It finds its application on anticipated faults diagnosis;
ii. Model-based techniques: It makes use of sensors for data retrieving from machine to be diagnosed, data collected are modeled, and residuals or innovative knowledge is required which is built on the decision logic. This tool is good for addressing unanticipated faults; and

iii. Statistical regression and clustering techniques: This works on historical data. Signature patterning groupings and classification are done using historical data and algorithm of decision-making.

4 Machine Signature and Integrity

As earlier revealed in the introductory section of this chapter, machine signature exhibits the signals spectrum of the behavioral model of how the machine works. Ierace et al. [16] posited that the phenomenon lies on an action resulting into an effect.

In an approach to know what the machine signature holds, it is easily determined by machine learning and vector support machine principles. These two principles aid in determining the machine signature pattern.

Kavulya et al. [25] explained that machine learning is the science dealing with building up of procedural rules that enables computers to imitate characteristics of machine hinged on data training. In the same note, Behnia et al. [47] stated that machine learning and pattern recognition referred to *"the act of taking raw data and adopting an action based on the category of the pattern which is mostly a set of data linked to a set of classes."*

The machine learning is grouped into three, namely: supervised learning, unsupervised learning, and deep learning. While the widely used machine learning algorithms are support vector machines, linear regression, logistic regression, tree-based modeling, Bayesian modeling, k nearest neighbor, k-means, neural network, and perceptron ensemble modeling and anomaly detection [29, 53]. The explanation on machine learning, support vector machines, support vector regression can be found in Widodo and Yang [48], Wuest et al. [49], Nishani and Biba [50] and Behnia et al. [47] and Gui et al. [51]. But for the purpose of clarity on what the tools of the diagnostics analysis of this chapter are built on, it is on machine learning and support vector machine due to the following unique characteristics they possess:

i. Establishment of machine signature pattern [47];
ii. Provision of family classification grouping [47];
iii. Possibility of training of known data to learn from so as to be able to forecast on unforeseen situation [50];
iv. The learning algorithm implores the use of preliminary facts to kick start the process;
v. Good for extracting features and identification of defect [51]; and
vi. Has high accuracy and good generalization for a smaller group of samples [47].

Machine signature is made up of the individual components' signature it has, be it electric motor, actuator, gear, switch, couplings, belt, impeller, sprocket, sensors, rotor, etc.

Let the total component parts in any given “machine A” be n, such that the list of the components will flow from component $A1$, $A2$ ……….. $A(n-1)$, An.

At delivery of a newly produced “machine A,” the reliability is expected to be 100% functional provided all things are equal (ceteris paribus). As machine A commences operation, the efficiency and its reliability is expected to be approximately high as 99% for the first three to six months of its operation. With the passage of time, if proper maintenance practice is put in place, the efficiency will still be optimal. Figure 2 shows the transformation change with time when degradation starts to set in and the probable state of its reliability as well as the

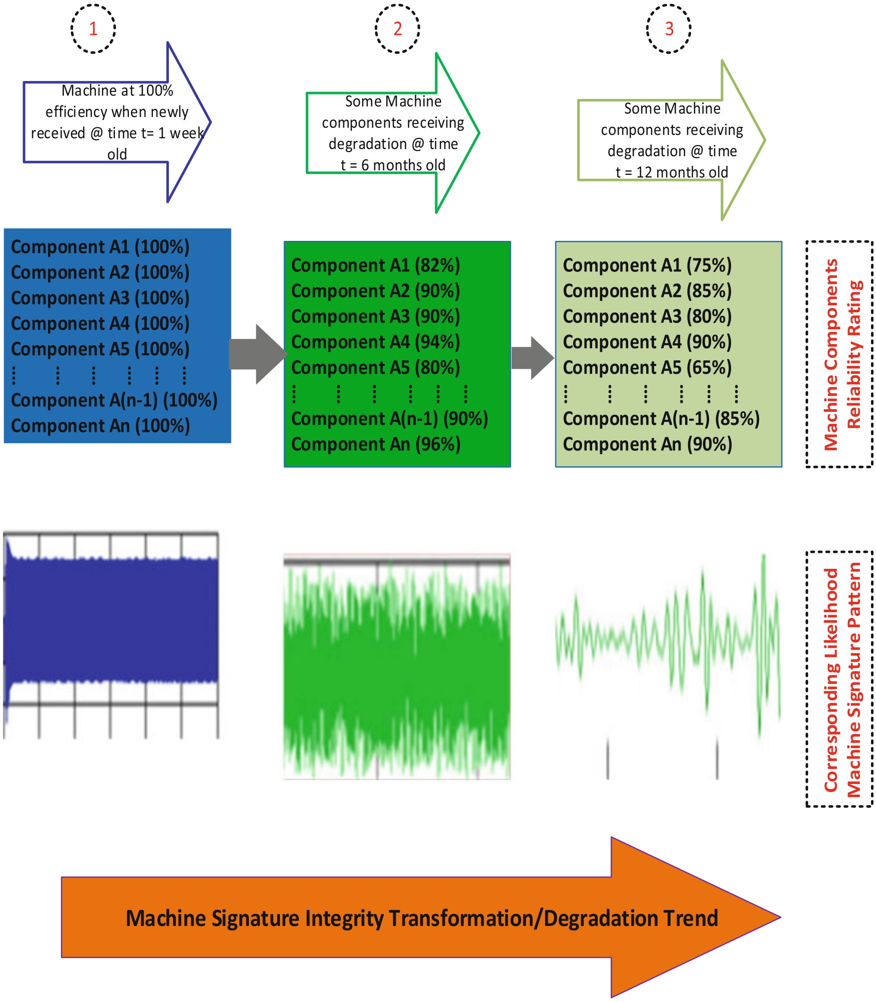


Fig. 2 Likelihood transformation change with time in machine signature according to degradation initiation in components’ parts

likelihood signatures the machine might display. As the machine system components are altered due to initiation of incipient fault(s), the signature pattern of the machine is bound to change.

In other words, as machine reliability changes, the machine signature changes accordingly. But for a well-maintained machine, its signature will probably remain constant. Mathematically, it could be inferred that machine signature should be directly proportional to its reliability rating as expressed in Eq. (3)

$$MS \propto R \tag{3}$$

where *MS* is the machine signature integrity pattern and R is the machine reliability or its efficiency. And *R* is the summation of all the individual components reliability divided by the total number of components it possesses as given in Eq. (4) or (5)

That is:

$$R_{A(\mathrm{total})} = \frac{R_{A1} + R_{A2} + R_{A3} + R_{A4} + \cdots R_{A(n-1)} + R_{An}}{n} (\%) \tag{4}$$

or

$$R_{A(\mathrm{total})} = \frac{1}{n}\left(\sum_{i=1}^{n} R_A\right) \tag{5}$$

where n is the total number of the components makeup of the machine being considered. R_{A1}, R_{A2}, R_{A3}… are the individual reliability rating of each component presents in machine A.

Machine Integrity

Machine integrity is a function of what the machine is known for and what it can do at any given time, t on any given operation, p when all the principles behind its operation are kept to the letter. Just as earlier noted in the introductory aspect of this chapter, when a manufacturer produces any industrial machine, before such is introduced to the market, series of test would have been carried out on it to make it certified, trademarked, and guaranteed for end users. For the machine to deny its integrity is an indication that something is wrong somewhere since it is a dummy object that awaits commands from its users or operators. For the course of clarity, let us consider Figs. 3 and 4 which are excerpt from the work of Garg and Dahiya [9]. Figure 3a–c excerpt presents the experimental result at which rotor signature wavelength of a wind turbine was being monitored under (no load) decomposition at normal condition, faulty condition at constant speed of 7 m/s, and faulty condition at varying step speed, respectively. While Fig. 4a–c excerpt presents rotor signature wavelength of the wind turbine at (full load) decomposition under normal condition, faulty condition at constant speed of 7 m/s, and faulty condition at varying step speed, respectively. From the excerpt, it could be noted that there are different wavelength patterns of the rotor as the frequency components increase with gravity of fault.

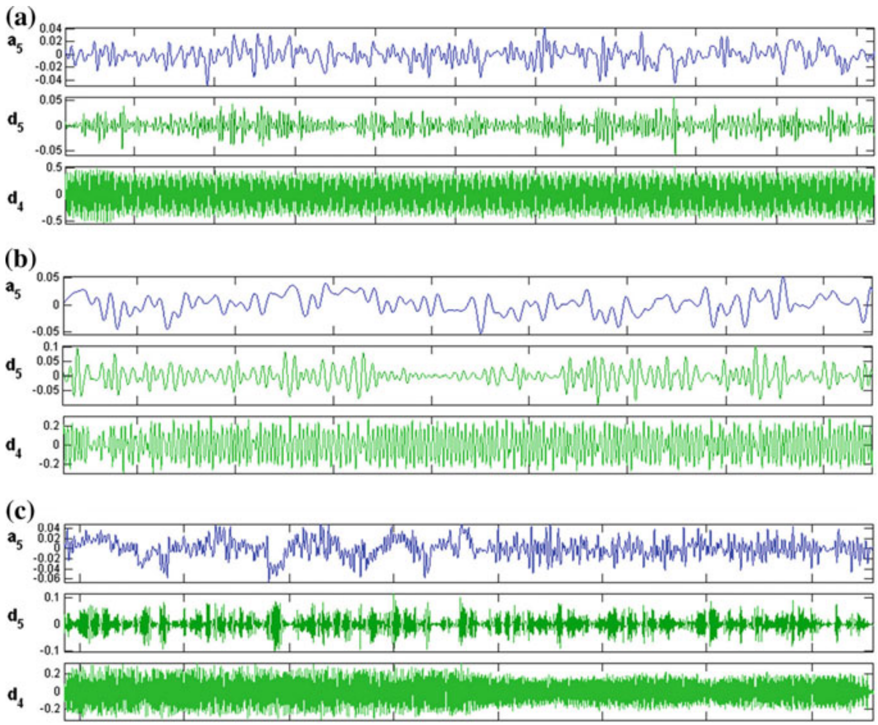


Fig. 3 Excerpt signature of the rotor modulating signal of *no-load* decomposition under: **a** normal healthy condition at constant wind speed of 7 m/s; **b** rotor eccentricity fault at constant wind speed of 7 m/s; and **c** rotor eccentricity fault at step variation in wind speed (from [9])

5 Machine Data Trending

Trend monitoring is an approach that is used in making known the acceptable and healthy signature patterns of machine. Trend analysis works on prediction of the future based on the existing facts and figures. It uses the present trend of operational event to predict what the likelihood of the machine would be in subsequent interval periods of time through the generated graphical pattern exhibiting or describing it. When extracted data from machine are to be trended, such data and the monitoring factors would be worked upon using regression model which could be linear, exponential, logarithmic, power, or moving average to establish the relationship between the generated data and the observed factors. The use of time series data for trending analysis would involve generating a model of $y = ax + b$, where "a" symbolizes the rate of change of the machine behavior and "b" the intercept on the y-axis. Once "a" is known, and "b" is established, relevant information could be derived in forecasting the behavior of the machine.

The approaches for data trending are:

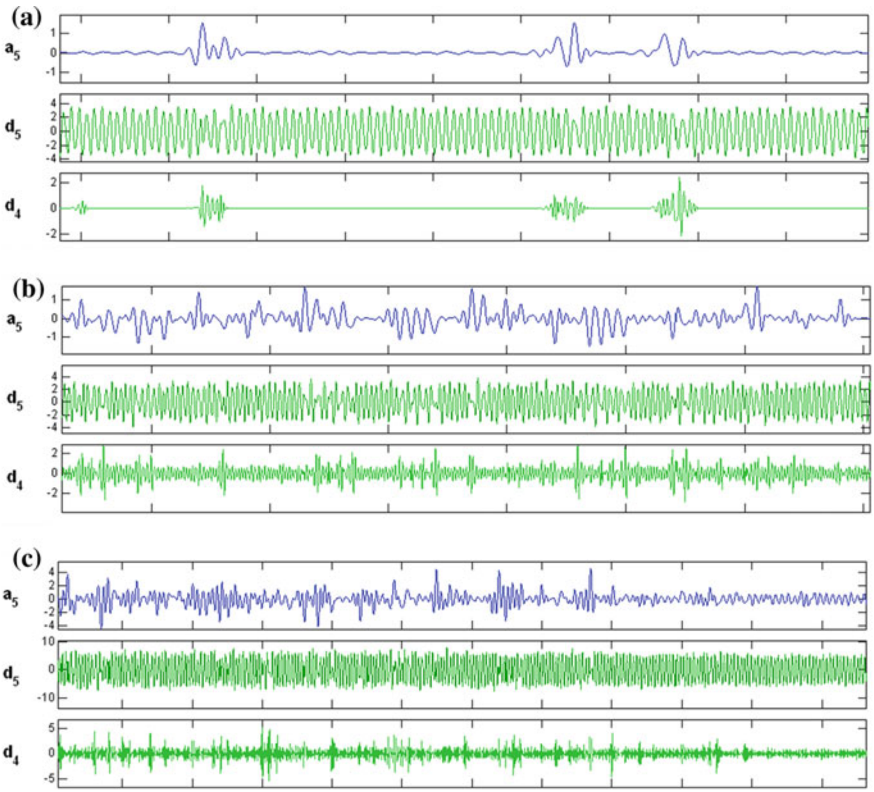


Fig. 4 Excerpt signature of the rotor modulating signal of *full load* decomposition under: **a** normal healthy condition at constant wind speed of 7 m/s; **b** rotor eccentricity fault at constant wind speed of 7 m/s; and **c** rotor eccentricity fault at step variation in wind speed (from [9])

i. Rule-based model;
ii. Data simulation;
iii. Statistical time series modeling; and
iv. Machine learning.

These approaches had been introduced in the previous subheadings.

Based on the knowledge of data trending, Okah-Avae [52] revealed that with experience of industrial machine users, the mechanical condition of machines is always proportional to its vibration level. And that a system degradation from its mechanical point of view would manifest itself in a corresponding change of its vibration response. Therefore, the use of historical data and in situ data taken continuously, intermittently, and periodically for establishing the trend pattern of the machinery behavior will give a reliable information of health status or condition of the machines. Based on this, the fault identification given in this chapter is built on vibration-related diagnosis trouble shooting tips.

6 The Proposed Concept for Machine Fault Diagnosis

The diagnostic analysis proposed in this chapter is as shown in Fig. 5.

Industrial machinery sends its behavioral data information into the data acquisition system either through wired or wireless communicating sensors on vibration-related data, temperature values, acoustic and sound data, humidity, wear rate, corrosion, lubrication data, etc. (However, this chapter will not discuss the necessary sensors needed for this). The data acquisition system retrieves the data and communicates it to the diagnostic analysis toolbox for various analyses. The

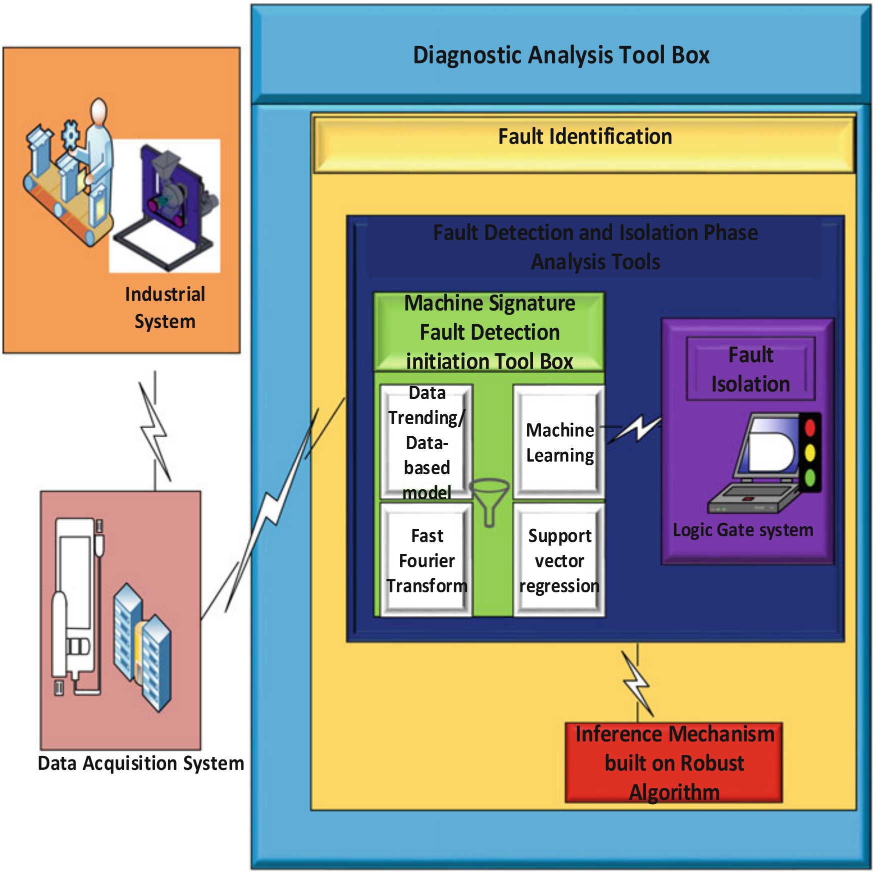


Fig. 5 Architectural framework of the proposed fault diagnostic system

machine signature toolbox receives the communicated data for analysis using data-based model, machine learning fast Fourier transform, and support vector machine knitted in one system for data training, filtering, signature patterning/features extraction, classification, and trending. Simultaneously, the data are communicated to the fault-isolation section where some decisions are being made based on "if, then else" principles. Also, the logic gate is integrated into this section to assist in robust filtering and decision-making, which is meant to aid the fault identification process.

For clarity on how this works, consider this scenario. The condition monitoring techniques are anchored on three different monitoring indices: vibration, temperature, and sound for detecting, isolating and identification of faults. The logic gate will function on the basis of truth table. Since there are three factors, the expected combination output will be 2^f. That is, the expected combination output Q will be 2^f as expressed in Eq. (6) where f is the number of factors being considered.

$$Q = 2^f \tag{6}$$

For this scenario under explanation, the truth table will have eight combinations as shown in Table 2 with the assumption that:

i. Logic 1 represents a state when an active factor is predominantly contributing to the machine fault;
ii. Logic 0 represents a state when a factor is not contributing to the machine fault;
iii. Vibration as a factor is denoted by "A";
iv. Temperature is represented with "B"; and
v. Sound be represented by "C."

The conversion of the truth table of Table 2 to Boolean equation is thus given in Eqs. (7) and (8)

$$F = \overline{A}BC + A\overline{B}C + AB\overline{C} + ABC \tag{7}$$

$$F = C(\overline{A}B + A\overline{B}) + A(B\overline{C} + BC) \tag{8}$$

Table 2 Truth table for fault-isolation assistance

S/N	*A*	*B*	*C*	*F*
1	0	0	0	0
2	0	0	1	0
3	0	1	0	0
4	0	1	1	1
5	1	0	0	0
6	1	0	1	1
7	1	1	0	1
8	1	1	1	1

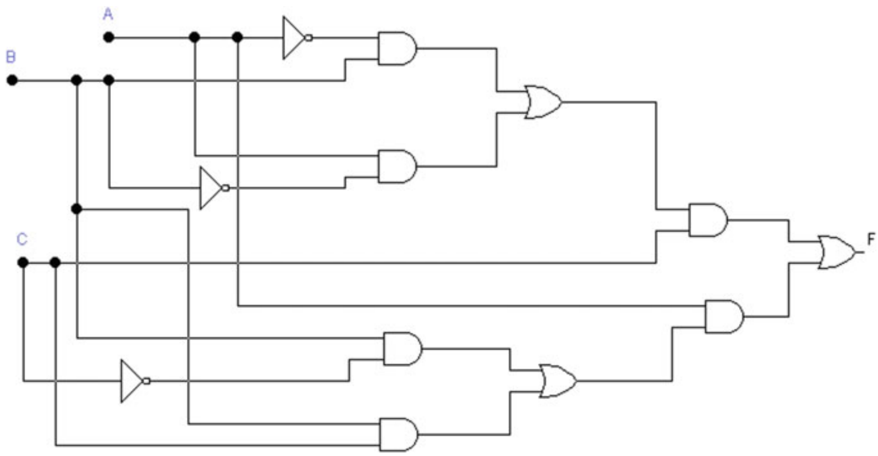


Fig. 6 Logic gate design for the assisted fault-isolation process

These equations are now used to build the logic gate which is as shown in Fig. 6.

With reference to the Truth Table of Table 2 and Fig. 6, at the fourth, sixth, seventh, and eighth combinations, there will be indicator signaling information into adjoining system to initiate the fault identification programs for the machines. How the fault will be identified would be discussed under the flowchart and algorithm of the proposed program.

The Supporting Flowchart and Algorithm for the Proposed Diagnostic Approach

Figure 7 refers. As serially numbered in red ink, serial number:

i. 1-3 gives the *pre-fault detection stage* of the machine signature integrity. This is needed because it will be a useless and wasteful act if the needed precautions that ought to have been adhered to by machine users are not been kept. These precautions check will ensure the reliance important in machine integrity. The needed precautions referred to include the way and manner on how the machine should be operated, be it on-load or no-load condition, recommended power rating and just to mention a few. The pre-fault detection algorithm gives the pre-check conformance test which requires the following tasks: input of manufacturer's data guidelines, setting of machine signature threshold spectrum pattern, and a decision control box. Once this is not met, the algorithm instructs the machine operator to task labeled "1" in the flowchart shown. If met, the program proceeds to the next stage which is signature patterning and data trending phase;

ii. 4-6 gives the **signature patterning and data trending phase**. In this phase, machine data are being extracted for signature patterning and data trending using the aforementioned analytical tools. Machine signature pattern is now compared with the set standard frequency pattern in task serial 1. After which, a decision based on the inference drawn would be made. If the machine

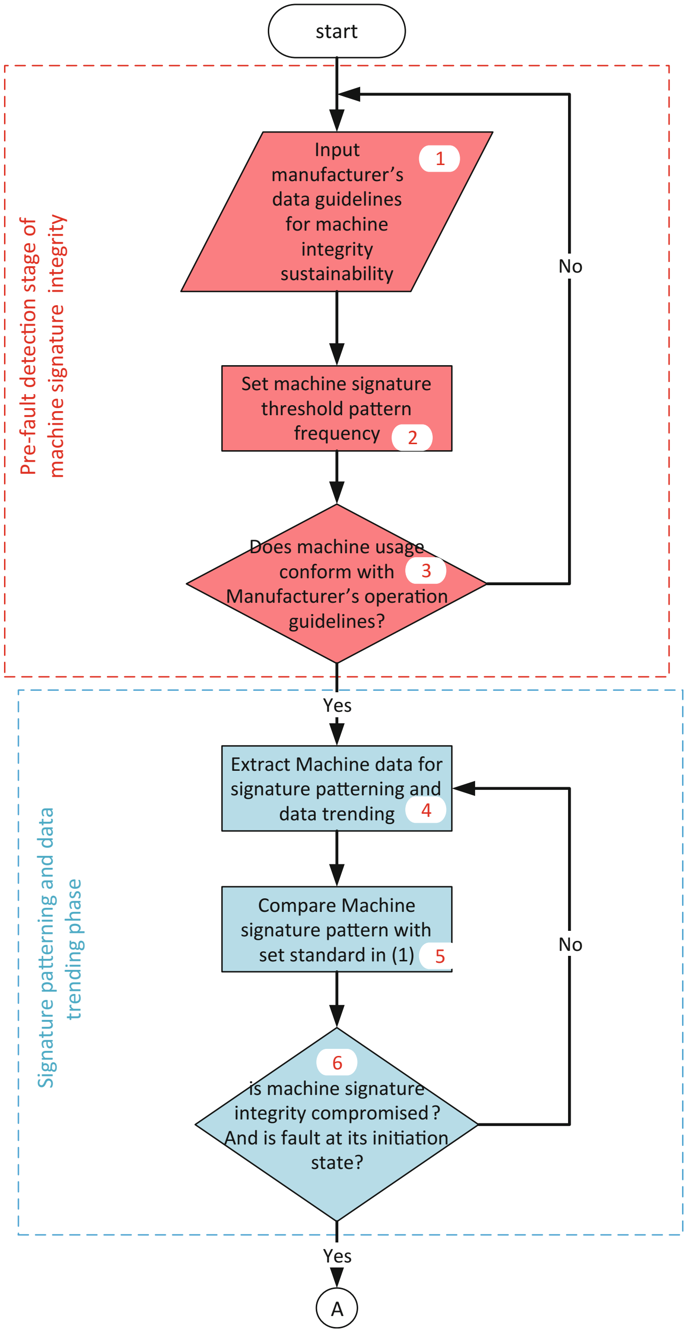


Fig. 7 Flowchart for the proposed fault diagnostic system

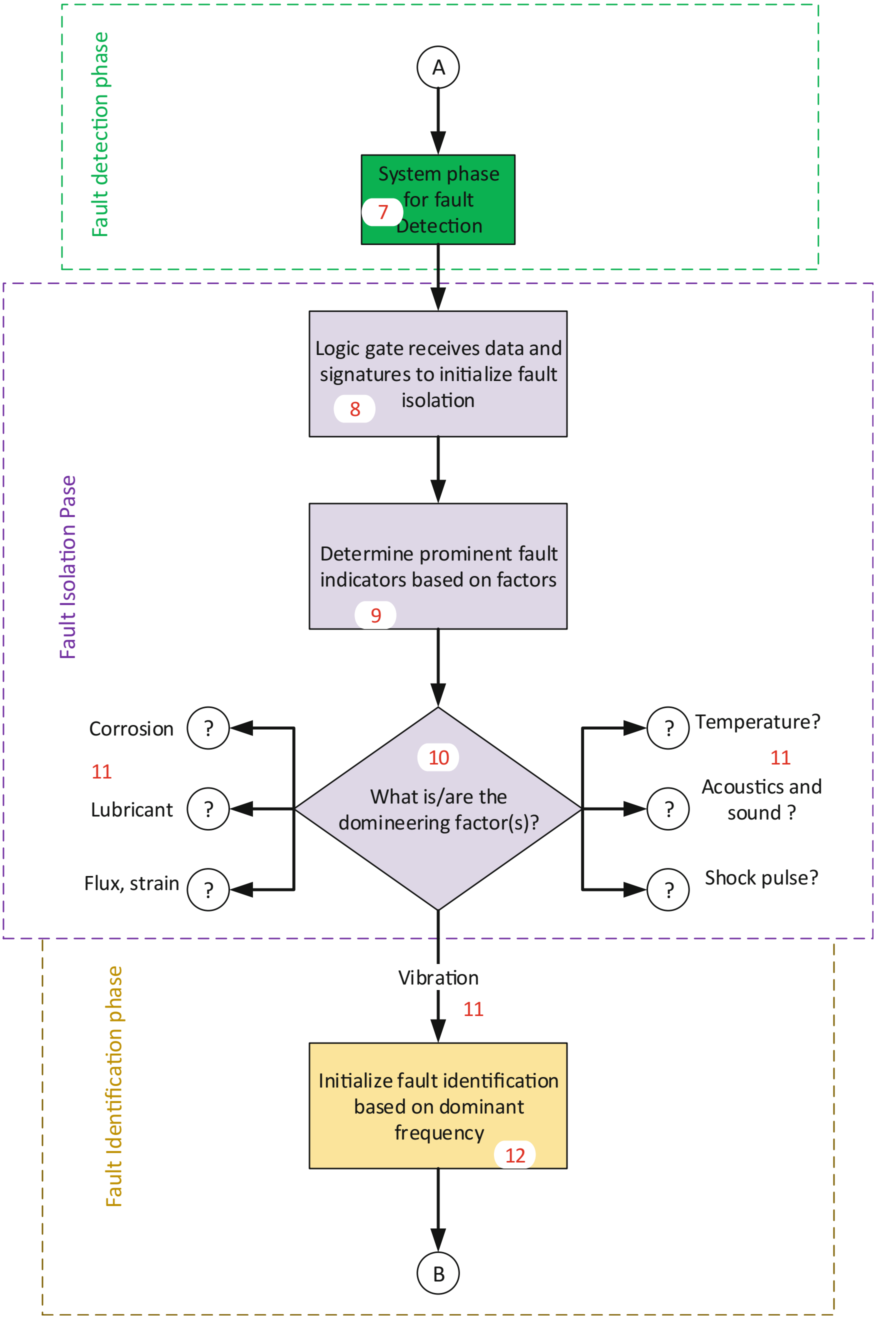


Fig. 7 (continued)

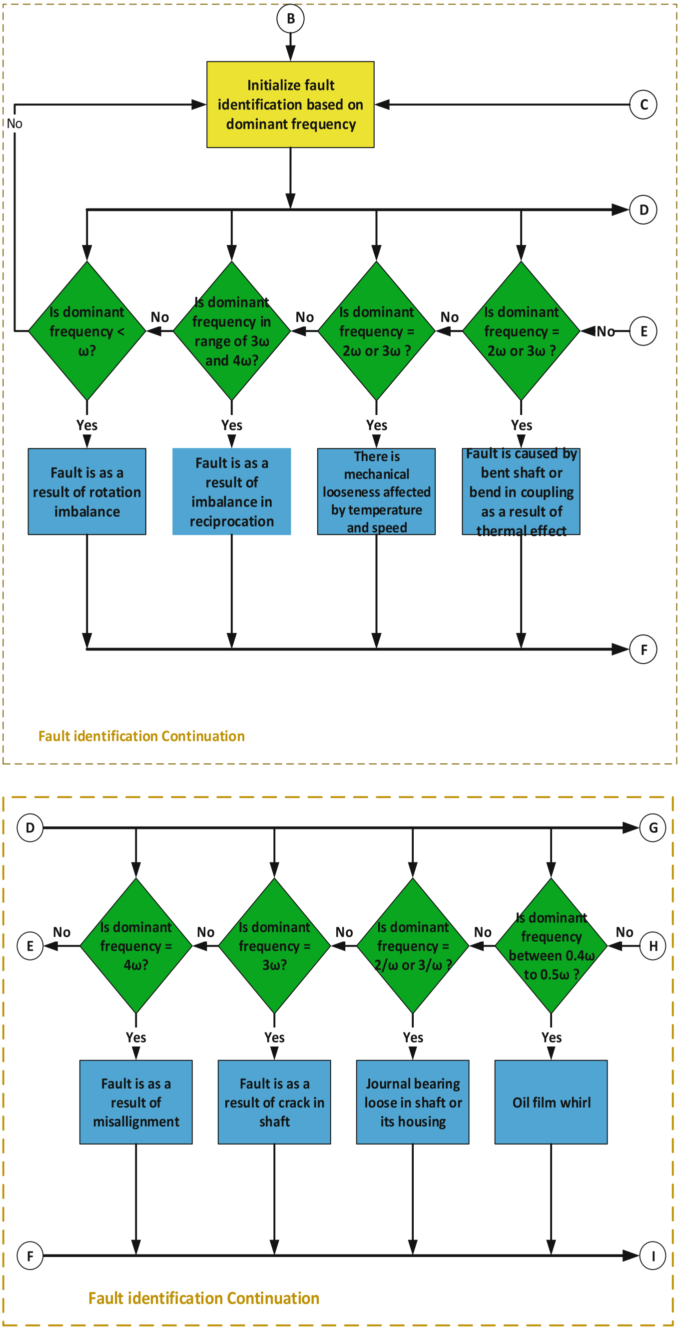


Fig. 7 (continued)

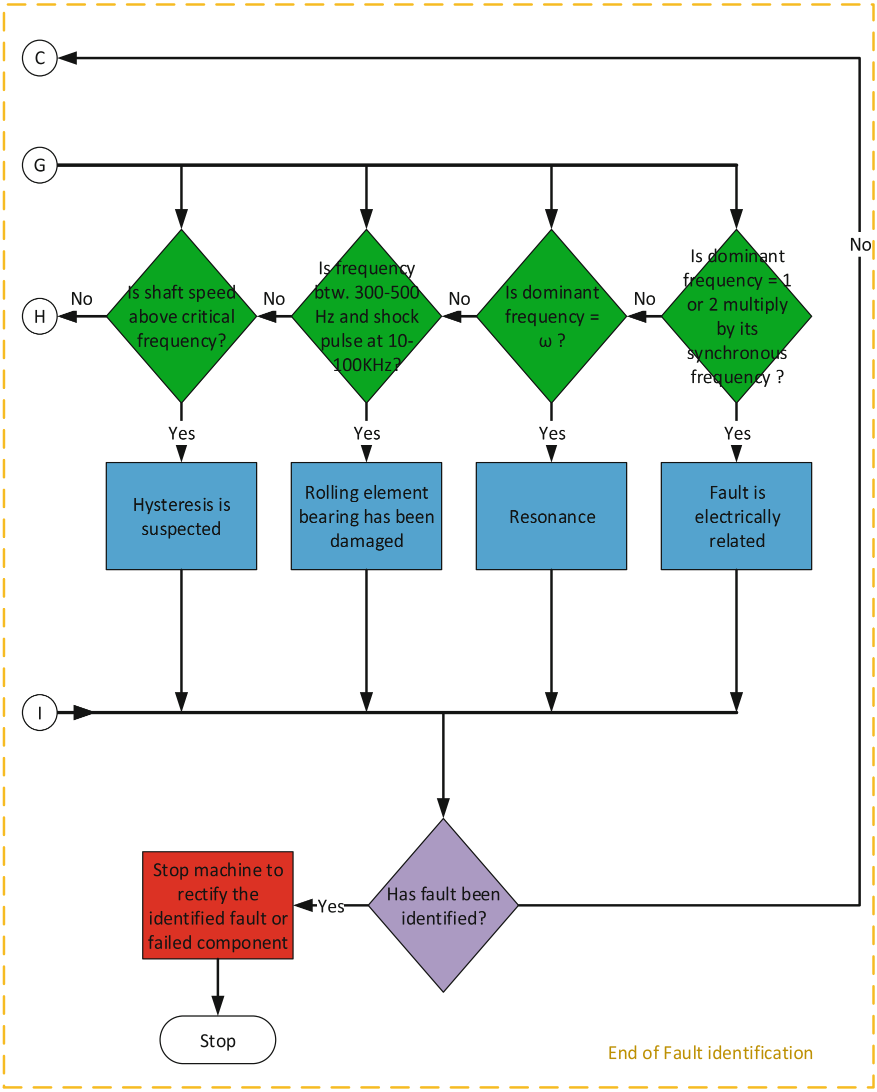


Fig. 7 (continued)

integrity is not compromised, the system will take it back to task "4," else, it will proceed to the next stage which is fault detection;

iii. 7 gives the task to be done under the **fault detection phase.**

iv. 8–10 discusses the operation task of the **fault-isolation state**. The logic gate in line with the machine learning and other data-driven model are used in the fault-isolation phase. Prominent fault indicators are determined as pointer to the knowledge of what to be identified; and

v. 11–12 highlights the tasks of **fault identification phase**. The stage hinges on the outcome from fault-isolation process. From the flowchart, seven factors were identified, all with question mark (?) except vibration. For the course of this chapter, only the fault identification based on vibration-related cases would be provided as the research work is just being kick started. The flowchart to identify the fault is contained in the remaining flowchart continuation shown in Fig. 7b–d

7 Diagnostic Algorithm

The set of programs that will implement the flowchart of Figs. 4, 5, and 6 depend solely on the algorithm describing the flow of activities sequentially and simultaneous actions that needed to be done for any point in time. Therefore, the algorithm of the proposed approach is thus as stated.

Algorithm I Algorithm I—Pre-setting/pre-fault detection and signature patterning stage

Step 1: *establish the manufacturers' instructions and requirement, input into the system*
Step 2: *set system connections (as rightly indicated above)*
Step 3: *initialize data retrieving process*
Step 4: *transform data into models*
Step 5: *train data, classify data, filter data by Support Vector Machines and Machine Learning tools*
Step 6: *generate machine signature pattern*
Step 6: *trend data and machine signature*
Step 7: *define frequency pattern, its boundary and associated classifiers*
Step 8: *commence diagnostic analysis*
Step 9: *is machine being used the way it should as preset and recommended by manufacturers?*
Step 10: *if yes, next step 11, else GOTO step1*
Step 11: *compare machine signature pattern with set reference standard of pattern*
Step 12: *is machine signature integrity compromised?*
Step 13: *if yes next step 14, else GOTO step3*

Algorithm II Algorithm II—Fault detection and Isolation Phase

Step 14: *initialize fault detection system*
Step 15: *logic gate receives communicated data to initialize isolation process*
Step 16: *system isolate fault using the designed logic gate*
Step 17: *determine prominent fault indicators*

Step 18: *is it vibration related, if yes, next step 19………. ** (the system can simultaneously work on all the factors that seem to be noticeable)*

Algorithm III Algorithm III—Fault Identification Phase

Step 19: *initialize fault identification*
Step 20: *system memory recall back set of signature pattern spectrum preset on it. (Here the vibration based signature patterns are activated)*
Step 21: *system makes decision based on the dominant frequency of the machine*
Step 22: *system compares and compares based on dominant frequency ****
Step 23: *is fault identified?*
Step 24: *if yes, next step, else GOTO 19*
Step 25: *stop machine and rectify fault.*
Step 26: *program paused*
Step 27: *Program restart*

Discussion of Algorithm Step 18 and Step 22

The sign "**" indicated in the algorithm means that the diagnostic program can simultaneously work on quite large factors identified. But for the sake of this chapter, only the vibration identification algorithm and flowchart are documented.

Similarly, the symbol "***" is shown so as to give a full explanation on how the fault is identified.

Therefore, with reference to Fig. 7b–d, the identification of vibration-related faults is integrated into the algorithm based on the indicated frequency range of documented vibration as found in Okah-Avae (1995). For instance, once the fault identification phase is initialized, the process starts to watch like a watch dog and using the inference mechanism built-in based on dominant frequency output translated from the various signature waveforms. The system will run through series of trouble shooting tips in his processor memory using generative decision questions based on the dominant frequency. Generative decision questions such as "is dominant frequency equals 3 multiply by the angular velocity, ω. If "NO," it asks another decision question that "is dominant frequency equals 4ω?" If "YES," the system quickly matches the "compromised signature integrity" or fault and drawn inference that the fault is as a result of "MISALIGNMENT." The system will proceed further to validate the authenticity of its submission by asking the question "HAS FAULT BEEN IDENTIFIED?" Then, the system pause for the rectification or repair of the identified fault or failed component(s) provided the answer to the question raised is "YES." But if "NO," the system will revert back to the "INITIALIZE FAULT IDENTIFICATION USING DOMINANT FREQUENCY" subroutine of the program as shown in Fig. 7b to continue to watch until fault is identified. The cycle of the diagnosis continues with the production cycle.

8 Conclusion

The concept presented is a preliminary work of a subpart outline program stemming out from a research work entitled smart predictive maintenance system for production system being supported under the "TWAS-DFG Cooperation Research Visit Program." The research is still at the framework level of which the explicit detail of result and finding could not be presented to readers, researchers, and industrialist at this instance, but the veracity of the program is of high importance. However, with the understanding of the proposed architectural frame work, flowchart and algorithm presented on machine signature integrity and data trending technique, if followed and fully implemented, stepwise/abrupt fault, drift/incipient fault, additive faults, and multiplicative faults which are based on time of appearance and addition would be detected, isolated, and identified. The anticipated further work to be done is to ensure the full implementation of the concept presented and as well move further by integrating a self-induced repairing module so that when a fault is detected, the system on its own should rectify the identified fault.

Acknowledgements The authors would like to acknowledge The World Academic of Sciences (TWAS) and German Research Foundation (DFG) for the award granted in supporting the research cooperation visit program. It is quite a research lift in crossbreeding ideas and sharing of scientific knowledge. Institute for Mechatronic system is quite appreciated for their hospitality, and the provision of conducive and enabling environment for the research activities.

References

1. Isermann R (2005) Model based fault detection and diagnosis: status and applications. Annu Rev Control 29:71–85
2. Yen GG (1996) Health monitoring of vibration signatures in rotorcraft wings. Neural Process Lett 4:127–137
3. Wang Y, Xiang J, Markert R, Liang M (2016) Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. Mech Syst Signal Process 66–67:679–698
4. Mehra RK, Peschon J (1971) An innovative approach to fault detection and diagnosis in dynamic systems. Automation 7:637–640
5. Isermann R (1993) Fault diagnosis of machines via parameter estimation and knowledge processing. Automatica 29(4):815–835
6. Frank PM, Koopen-Seliger B (1997) Fuzzy logic and neural network applications to fault diagnosis. Int J Approx Reason 16:67–88
7. Benbouzid MEH (2000) A review of induction motors signature analysis as a medium for faults detection. IEEE Trans Ind Electron 47(5):984–993
8. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN (2003) A review of process fault detection and diagnosis Part I: Quantitative model-based methods. Comput Chem Eng 27:293–311
9. Garg H, Dahiya R (2016) Current signature analysis and its application in the condition monitoring of wind turbine for rotor faults Energy System. doi:10.1007/s12667-016-0208-6
10. Staroswiecki M (2000) Quantitative and qualitative models for fault detection and isolation. Mech Syst Signal Process 14(3):301–325

11. Kahraman C, Gülbay M, Kabak Ö (2006) Applications of fuzzy sets in industrial engineering: a topical classification. In: Kahraman C (ed) Fuzzy application in Industrial Engineering. Springer, Berlin, Heidelberg, pp 1–55
12. Kurfess TR, Billington S, Liang SY (2006) Advanced diagnostic and prognostic techniques rolling element bearings. In: Wang L, Gao RX. (eds) Condition monitoring and control for intelligent manufacturing, pp 137–165
13. Rajakarunakaran S, Venkumar P, Devaraj D, Rao KSP (2008) Artificial neural network approach for fault detection in rotary system. Appl Soft Comput 8:740–748
14. Chebel-Morello B, Haouchine K, Zerhouni N (2009) A methodology to conceive a case based system of industrial diagnosis. In: Proceedings of the 4th World Congress on Engineering Asset Management, Athens, Greece, pp 474–486
15. Mendonca LF, Sousa JMC, Sa´ da Costa JMG (2009) An architecture for fault detection and isolation based on fuzzy methods. Expert Syst Appl 36:1092–1104
16. Ierace S, Garetti M, Cristaldi L (2009) Electric signature analysis as a cheap diagnostic and prognostic tool. In: Proceedings of the 4th World Congress on Engineering Asset Management Athens, Greece, pp 750–757
17. Yin S, Ding SX, Haghani A, Hao H, Zhang P (2012) A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. J Process Control 22:1567–1581
18. Yin S, Wang G, Karimi HR (2014) Data-driven design of robust fault detection system for wind turbines. Mechatronics 24:298–306
19. Chamay M, Oh S, Kim Y (2014) Non-parametric dimension reduction algorithm approach for neural networks applied to diagnostic systems. J Mech Sci Technol 28(9):3529–3536. doi:10.1007/s12206-014-0813-z
20. Skelton K (2006) Delivering reliability improvement & efficiency through emerging diagnostic techniques at Powercor. In: World Congress on Engineering Asset Management
21. Chamay M, Oh S, Kim Y (2013) Development of a diagnostic system using LPC/cepstrum analysis in machine vibration. J Mech Sci Technol 27(9):2629–2636
22. Islam MM, Lee G, Hettiwatte SN (2017) A review of condition monitoring techniques and diagnostic tests for lifetime estimation of power transformers. Electr Eng. doi:10.1007/s00202-017-0532-4
23. Fortuna L, Graziani S, Rizzo A, Xibilia MG (2007) Fault detection, sensor validation and diagnosis. In: Fortuna et al (eds) Soft sensors for monitoring and control of industrial processes. Springer, London, pp 183–226
24. Pattipati K, Kodali A, Luo J et al (2008) An integrated diagnostic process for automotive systems. Stud Comput Intell (SCI) 132:191–218
25. Kavulya SP, Joshi K, Giandomenico FD et al (2012). Failure diagnosis of complex systems. In: Wolter K et al (eds) Resilience assessment and evaluation of computing systems. Springer, Heidelberg, pp 239–260. doi:10.1007/978-3-642-29032-9_12
26. Rashid MM, Amar M, Gondal I, Kamruzzaman J (2016) A data mining approach for machine fault diagnosis based on associated frequency patterns. Appl Intell 45:638–651. doi:10.1007/s10489-016-0781-3
27. Goyal D, Vanraj Pabla BS, Dhami SS (2016) Condition monitoring parameters for fault diagnosis of fixed axis gearbox: a review. Arch Comput Methods Eng. doi:10.1007/s11831-016-9176-1
28. Chauhan V, Surgenor B (2017) Fault detection and classification in automated assembly machines using machine vision. Int J Adv Manuf Technol 90:2491–2512. doi:10.1007/s00170-016-9581-5
29. Javed K, Gouriveau R, Zerhouni N (2017) State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. Mech Syst Signal Process 94:214–236
30. Rao BKN (1998) Condition monitoring and the integrity of industrial systems. In: Davies A (ed) Handbook of condition monitoring. Springer, Netherland, pp 3–34

31. Huang HH, Wang B (1999) Related work on machine monitoring and diagnostics. In: Lee J, Wang B (eds) Computer aided maintenance Part one. Springer, New York, pp 41–58. doi: 10.1007/978-1-4615-5305-2_3
32. Stefanoiu D, Ionescu F (2006) Fuzzy-statistical reasoning in fault diagnosis. In: Palade V, Bocaniala CD, L Jain (eds) Computational intelligence in fault diagnosis. Springer, London, pp 125–177. doi: 10.1007/978-1-84628-631-5_5
33. Dai J, Chen CLP, Xu X et al (2008) Machinery vibration signals analysis and monitoring for fault diagnosis and process control. In: International conference on intelligent computing on advanced intelligent computing theories and applications with respect of theoretical and methodical issues. pp 696–703
34. Bongers DR, Gurgenci H (2008) Fault detection and identification for longwall machinery using SCADA data. In: Murthy K (ed) Complex system maintenance handbook Part V. Springer, London, pp 611–641
35. Sobhani-Tehrani E, Khorasani K (2009) Fault diagnosis of nonlinear systems using a hybrid approach. Lecture Notes Control Inform Sci 383:21–49. doi: 10.1007/978-0-387-92907-1
36. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50:297–313. doi:10.1007/s00170-009-2482-0
37. Cholewa W, Korbicz J, Ko´scielny JM et al (2011) Diagnostic method. In: Korbicz J, Ko´scielny JM (eds) Modelling, diagnostics and process control. Springer, Heidelberg, pp 153–231
38. Qin SJ (2012) Survey on data-driven industrial process monitoring and diagnosis. Annu Rev Control 36:220–234
39. Hwang KH, Lee JM, Hwang J (2013) A new machine condition monitoring method based on likelihood change of a stochastic model. Mech Syst Signal Process 41:357–365
40. Wang D, Yu W, Low CB, Arogeti S (2013) Health monitoring of engineering systems. In: Wang D et al (eds) Model-based health monitoring of hybrid systems. Springer, New York, pp 1–29. doi: 10.1007/978-1-4614-7369-5_1
41. Tidriri K, Chatti N, Verron S et al (2016) Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: a review of researches and future challenges. Annu Rev Control 42:63–81
42. Vog GW, Weiss BA, Helu M (2016) A review of diagnostic and prognostic capabilities and best practices for manufacturing. J Intell Manuf. doi:10.1007/s10845-016-1228-8
43. Ogidi OO, Barendse PS, Khan MA (2016) Fault diagnosis and condition monitoring of axial-flux permanent magnet wind generators. Electric Power Systems Res 136:1–7
44. Vachtsevanos G, Lewis F, Roemer M (2006). Fault diagnosis. In: Hess A, Wu B (eds) Intelligent fault diagnosis and prognosis for engineering systems. Wiley, New York, pp 172–279
45. Samantaray AK, Bouamama BO (2008) Model-based process supervision. A bond graph approach. Springer, London
46. Jayantha P. Liyanage, Jay Lee, Christos Emmanouilidis et al (2009) Integrated e-maintenance and intelligent maintenance systems. In: Ben-Daya et al (eds) Handbook of maintenance management and engineering Part V. Springer, London, pp 499–544
47. Behnia A, Ranjbar N, Chai HK, Masaeli M (2016) Failure prediction and reliability analysis of ferrocement composite structures by incorporating machine learning into acoustic emission monitoring technique. Constr Build Mater 122:823–832
48. Widodo A, Yang B-S (2007) Support vector in machine condition monitoring and fault diagnosis. Mech Syst Signal Process 21:2560–2574
49. Wuest T, Irgens C, Thoben KD (2014) An approach to monitoring quality in manufacturing using supervised machine learning on product state data. J Intell Manuf 25(5):1167–1180. doi:10.1007/s10845-013-0761-y
50. Nishani L, Biba M (2016) Machine learning for intrusion detection in MANET: a state-of-the-art survey. J Intell Inf Syst 46:391–407. doi:10.1007/s10844-015-0387-y

51. Gui G, Pan H, Lin Z et al (2017) Dat-driven support vector machine with optimization techniques for structural health monitoring and damage detection. KSCE J Civil Eng 21 (2):523–534
52. Okah-Avae BE (1995) The science of industrial machinery and systems maintenance. Spectrum Book Ltd, Ibadan Nigeria
53. Analytics Vidhya Content Team (2015) Ten machine learning algorithms explained to an "army soldier" https://www.analyticsvidhya.com/blog/2015/12/10-machine-learning-algorithms-explained-army-soldier/. Accessed 13 June 2017

Hazard and Risk Assessment and Mitigation for Objects of Critical Infrastructure

Alexander V. Bochkov

Abstract This chapter presents the basic definitions of hazard and risk assessment and mitigation for objects of critical infrastructure. The problem of investigation and assessment of risks, stability, vulnerability and survivability of large-scale systems (LSS) has been formulated. Approaches to the solution of a problem of assessing an abnormal situation in the a priori (pre-crisis, crisis) situation in the similar systems and construction of the system of its state indicators. The concept of rational distribution resources. The priority ranking that belongs to critical infrastructure (as exemplified by the fuel and energy complex facilities) has been discussed. The ranking algorithm as applied to similar facilities, which is based on their systemic value for the LSS, has been suggested. Besides, the algorithm for forming the ranked list of different types of facilities. In addition, the approaches to solving problems of analysis and risk management for critical infrastructure objects of LSS that allow making informed decisions on the rational allocation of funds for their protection are outlined. The noted features of situational management in the understanding of the management process as a process of transfer of information flows from one subject to another. It is shown that the selected persons can be described by including them in a broad, understandable category using special questionnaires. It is noted that such a questionnaire introduces systematicity in the process of apprehension in the analysis of assessed situations. It is shown that if there is sufficiently representative statistics on some criterion (for example, bad/good, regular/freelance, etc.), then using the support vector method, you can define rules in the automatic section of the stream. Consider an example of practical application of the method to the problem of risk prequalification of counterparties.

A.V. Bochkov (✉)
Risk Analysis Center, NIIgazeconomika LLC, 20/8 Staraya Basmannaya Str, Moscow 105066, Russia
e-mail: a.bochkov@gmail.com

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*, Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_3

1 Structural Complexity and Critical Infrastructure Objects Safety Problems Summary Experience of Risk Assessment

The so called and often discussed [1–5] problem of critical infrastructure is in the following: almost in all major sectors of economics there are systems which elements are so far distributed over a distance (sometimes these systems are classified as of territorial distributed) that it is almost impossible from economical point of view to protect completely all objects even of any one sector, without mentioning all sectors of the system. The main issues and a problem of the decision-maker (DM) in the field of safety of functioning of similar systems is the question of threats and risks assessment, significant for the system in general and for its elements, and definition of priority of elements and objects protection of critical infrastructure taking into account as a rule limited resources which are available at his disposal.

Besides huge sizes many sectors are so difficult that technologically and economically it is impossible to predict and estimate all unforeseen consequences of any incident regardless of whether an incident has been caused by consequences of people's malicious actions, or it was a consequence of natural disasters. As a rule, it is extremely difficult to predict consequences of small perturbation in one part of critical infrastructure for its other sectors. For example, all communications in the Internet in South Africa were completely stopped because of falling of the Twin Towers as a result of terrorist attack on the USA on September 9, 2001, and rather insignificant malfunctions in the electric useful power of First Energy in Ohio (USA) accelerated the blackout in August 2003 which has affected 50 million people for thousands of kilometers from the source of problem [6–8].

In fact, the existing infrastructure is vulnerable just because it contains so many very closely interconnected components that for most of technical consultants, analysts and DM defining its security policy it becomes an insoluble problem.

The concept of structural complexity, as well as the concept of the system in general still hasn't found unambiguous definition. At the same time, modern requirements to creation of systems of safety and efficiency of their functioning were and remain rather high. As a result, there are problems of the choice of priority objects of equipping from their general totality and optimum distribution at the disposal of the owner of the system (the owner, the state) of financial and material resources for their protection.

A person in his daily activities constantly faces the need of the choice using his logical apparatus, carrying out reasoning chain, addressing associations and analogies, remembering precedents in the past, stating forecasts, assumptions and guesses, applying to an intuition, and, at last, carrying out calculations. Analytical activity in general is always connected with the choice: anyway, its result is a decision made as a result of the analysis of some data set and some options. Naturally choosing the decision, DM seeks for receiving the best result for himself. Such way of action is usually called the optimum choice or the optimal solution.

With the expansion and complication of solvable problems also the cost of mistake grows during decision-making. Besides the more difficult the task is the more difficult it is to find not just optimal solution but also the correct decision in general. Proceeding from abovementioned further we will rather talk not about optimum but about rational decision. To possess methods of creation of such decision in the field of safety, planning and business of large energy companies, industries and the state in general is extremely important. The demand for adequate description and especially for modeling of reality grows constantly and along with it there is a need for synthesis and description of theoretical bases, ways of organization and maintaining, and also for creation of technological tools of information and analytical work.

Before going on it is necessary to define unambiguously an object of research and to agree about basic definitions of key concepts.

Hereinafter the term "system" is understood to be the set of operating interconnected elements and considered as uniform structural whole [9]. In general, any system means existence in its structure of a certain quantity of elements having a certain quality and characterized by relations among themselves. Some researchers pay attention that if the question is about system, in these relations "the principle of proportionality" must be observed [10, 11]. All this in total defines conditions for the solution of the main research task—the search of arguments of risk function.

The first exploratory phase is collection of data on the system, its surround, existing communications and interactions out of and in it, available resources, i.e., in other words—risk function synthesis after which it is possible to start the analysis of collected and systematized data.

There is quite large number of classifications of systems, each of which differs in the specifics. In this work, we consider only a class of so-called structural and difficult systems (the Latin word "structura" means interposition and communication of components of something).

Objects and the whole system in general are characterized by a set of properties.

One of basic—reliability—property of an object to function smoothly continuously with 100% efficiency level. In the analysis of reliability, the main criterion is failure criterion which divides everything into "yes" and "no".

Operational efficiency of functioning is a property of the system to function continuously though perhaps with the lowered level of "output parameters". Actually, it is the same reliability but without rigid failure criterion, and with several "levels" of quality/efficiency.

Stability is a property of the system to come back (for allowed time) to former 100% functioning level after disabling its separate components. At natural influences stability can be characterized by number of elements which casual "switching off" from the system leads to such state when the system stops satisfying to the concept 100% working capacity. At the same time elements are understood to be structural units ("blocks") of equal "scale". At hostile influences stability characterizes the number of elements arranged by importance, the removal of which leads the system to a state when it stops being 100% workable. Stability is an ability of the system to adapt and revert to the original state. In recent years during the analysis of stability of structural and difficult systems the terms resilience and "antifragility" [12] also started to be used.

Vitality is a property of an object to continue functioning in admissible limits after disabling of separate components. At natural influences vitality is characterized by number of elements which "switching off" from the system in a random way leads to its "dying". At hostile influences vitality is characterized "switching off" of the elements arranged by importance. Vitality is as margin of safety.

Safety is a property of an object to perform its functions without endamagement to operational personnel, environment and so forth.

One more concept about which it is already much written and told but without which it is impossible to consider approaches to the solution of the tasks designated in the introduction is danger. The best measure for the quantitative description of danger is risk today. We will dwell upon this concept.

The concept of risk is widely used in modern literature and often has absolutely various sense. So, for example, if they say that this car has high risk of accident then in this context the term "risk" can be simply replaced with the concept "accident risk".

Sometimes risk is used for the description of individual probability of death or any other consequence as a result of a certain activity, etc.

Today the concept of risk is generally characterized as: probability of an adverse effect uprising, probability that there is an adverse effect of just this type and probability that this type of influence causes a certain size of deviations of a condition of the subject of influence from its dynamic balance.

It follows that risk has to be vector value which can describe dangers of different type and where all its values given above enter as components. As main questions discussed below anyway are connected with safety of objects of critical infrastructure, therefore where it isn't stipulated especially, the term "risk" will be understood as risk of technogenic or more specifically industrial origin.

In questions connected with ensuring industrial safety most often was and sometimes is now the requirement of achievement of the negligible or "zero" risk connected with this or that production activity. Therefore, security systems which were created and used in the industry most often were engineering decisions directed on implementation of the requirement of absolute safety. The basic principle used for creation of these systems is so-called principle of ALAPA (As Low As Practicable Achievable). According to this principle it is necessary to increase industrial safety by all means and regardless of the reached level if it is technically effected.

In other words, according to the principle of ALAPA it was necessary to create technical security measures which would prevent emergencies, i.e. would nullify possibility and development of accident. And if an accident practically couldn't be, there was no need at the same time to consider actions for weakening of their consequences. The quality and efficiency of such security systems most often were judged according to statistical data of the results connected with their introduction.

However, despite all undertaken precautionary measures accidents at the industrial enterprises and sometimes very serious occurred and occur now. At the present stage of development when production itself and technological processes have become complicated and are connected with use of large numbers of energies, potential consequences of accidents became so serious that it became impossible to base the principles and criteria of safety on statistical approach any more. Besides complication of technologies has led to the fact that it is often just impossible to foresee all possible scenarios of development of accident and, respectively, to provide engineering and organizational decisions for their prevention that was convincingly confirmed once again by the Chernobyl accident and the accident in Fukushima.

All this has demanded essentially new approach in the solution of problems of industrial safety. It became necessary to develop new provisions and criteria of safety, to create a technique of the solution of these tasks. In the last two decades these questions are shown in a significant number of works which convincingly confirmed the statement which has already become axiomatic that achievement of absolute safety is impossible. But if it so, then there is a question: what the level of acceptable risk has to be, according to what criteria and what methods it has to be defined.

Thus, risk philosophy based on the concept of absolute safety with need has come to the concept of acceptable risk. The concept of acceptable risk demanded refusal of ALAPA principle and pass to a new ALARA principle (As Low As Reasonable Achievable). According to ALARA principle it is necessary to achieve a certain level of safety which has to be defined proceeding from social and economic conditions of social development. For accidents risk from which is higher than accepted it is necessary to use engineering decisions for their prevention and alleviation of the consequences, and for those accidents risk from which is less it is necessary to use only measures for alleviation of the consequences.

Realization of this principle for nuclear energetics found reflection in provisions on safety. The concept of safety of the NPP now they add not only design, but also beyond design basis accidents with possible damage of an active zone to its complete fusion. Consideration in the concept of safety of the NPP of beyond design basis accidents demanded a new level of protection provided with so-called accident management in the system of safety. Accident management means actions directed to prevention of development of design accidents into beyond design basis ones and alleviation of the consequences of beyond design basis accidents for what any serviceable technical means provided for reduction of consequences of beyond design basis accidents have to be used.

As development of design accident into beyond design basis one means the exit of the situation out of control, accident management at the first stage aims on restoration of this control of how active zone fusions will arise. If actions at this stage were unsuccessful, then there comes the second stage when accident management aims on weakening of its consequences.

Due to the fact that safety rules include the possibility of emergency mode on the NPP connected with beyond design basis accident there appeared the need for standard regulation of such actions as an assessment of possible doses of radiation of people in case of this type of accident, definition of zones of radioactive pollution and criteria for making decision on measures of protection of personnel and the population up to their evacuation. Therefore, an important feature is inclusion as one of criteria of safety the size of calculated value of probability of “limit emergency emission” which shouldn’t exceed 10^{-7} for the reactor in a year (in Russia).

It can be said that now the solution of problems of safety comes down to the following: on the basis of certain criteria to answer a question of what means and to what level it is necessary to reduce risk in a particular field of production activity in order safety both a person and the environment will be optimum.

Today one of possible and the most developed tool of the solution of this task is the tool united under the name “risk analysis”.

2 Risk Creation

We consider a problem of the characteristics of the functioning and sustainability of the structurally complex techno-economic systems (SCTES) in terms of different types of risk. The validity of the application to describe the behavior of this class of systems of semi-empirical mathematical models, which are based on a vector description of the system states, using the criteria approach for assessing the quality of its functioning, is demonstrated.

Under discussion is conceptual model for the interaction of the object and its environment, allowing to estimate the “optimal” allocation ratio between the productive system and its development potential.

The concept of non-formalizable threats for the sustainable functioning of this systems class was introduced. Expert procedure to account non-formalizable threats in case of risk assessment was proposed. For the construction of indicators for assessing the status the methods of quantitative analysis based on the theory and multi-criteria utility was used. Multi-criteria utility as an indicator of sustainability in the form of dimensionless complex hierarchy of indicators was proposed. Computing for through the convolution of the primary indicators.

The hierarchical model proposed to calculate the integral index of multi-criteria preference of one embodiment of the system over the other. Some results of case study are discussed.

Usually, the SCTES are characterized by distribution in space, big variety and interaction of objects types, non-uniform structure of processing chains, unique conditions of influence of risks of the various nature on objects of the subsystem and the system as a whole.

In the idea of situation management of SCTES principles of changeable (adaptive) behavior in terms of possible risks and uncertainties are initially put. Presence of such risks generated by different circumstances is capable to brake or change this or that way of movement, to force the system to live « under another scenario », different from all variety of plans shaped before.

If as sustainability of SCTES functioning to understand the plan performance of its development with admissible variation on volumes and terms of problems performance, then situation safety management in this system is reduced to minimization of hazardous losses at extraordinary situations and to carrying out of actions for their prediction. The success of such tactics depends substantially on intuition and talent of management of the company, on its ability to expect the possibility of weakly formalized threats outgrowth into notable risks and losses.

Under weakly formalized threats we understand here the threat for criticality estimation of which the development of original algorithm of the decision depending on a concrete situation is required and for which uncertainty and dynamism of the initial data and knowledge can be characteristic.

However, in the absence in enough of adequately estimated information necessary for decision-making, tactics of adaptive management quite often turns to a continuous chain of the "emergency" scenarios leading to disruption of controllability of all system. Hence, company management should be engaged not only the current work bringing quite notable results which utility is measured in economic factors, but also to care of creation the company condition monitoring system and the world surrounding it, watching dynamics of internal and external threats to its growth and development.

That is optimal control of SCTES aimed at reception of profit on its activity, consists in ability to find balance of redistribution of the available resources (material, human, information) by proprietor of the company between « productive activity » and « maintenance of development potential ».

The elementary model illustrating the abovementioned and allowing to estimate "optimal" proportions of resources distribution between "useful" (production) system and its development potential is the model of interaction of developing object and its environment [13, 14].

Let's present the activity of some SCTES, consisting of two subsystems (Fig. 1).

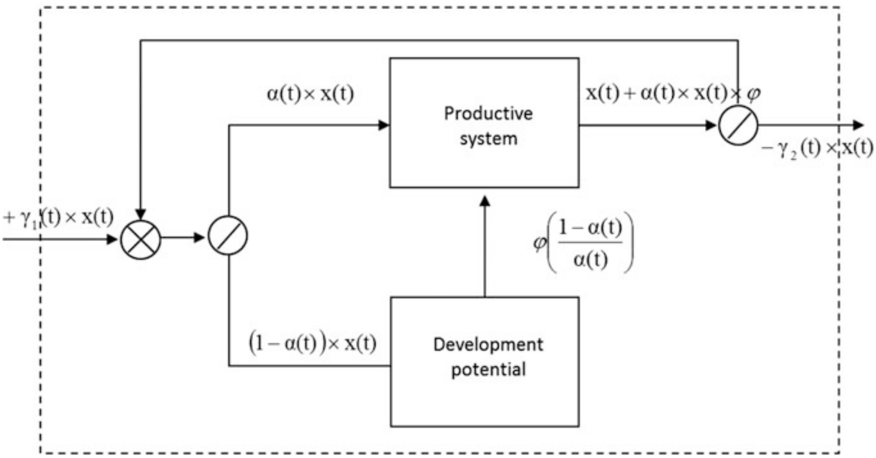


Fig. 1 Activity of productive system

The first, « productive subsystem » , brings profit, proportionally to quantity of received resources $(\alpha(t) \times x(t))$ with some positive resources increase speed coefficient

$$\varphi\left(\frac{1-\alpha(t)}{\alpha(t)}\right), \tag{2.1}$$

available in the system.

The second subsystem- « development potential » plays the role of the accelerator (retarder) of resources reproduction speed in the system.

Herein $\alpha(t)$ is proportion of resources distribution between productive subsystem and its development potential, $\gamma_1(t)$, $\gamma_2(t)$—coefficient of resources exchange intensity between investigated system and some external in relation to it system in the process of coexistence. Actually, the difference $\gamma(t) = (\gamma_1(t) - \gamma_2(t))$ is the share of resources deduced from the cycle of reproduction in the form of losses of one kind or another, for example, of final consumption, taxes, etc.

In the elementary representation "potential" influence is the value of function (2.1) and coefficient $\gamma(t)$ for large-scale systems we consider independent of times in an explicit form as constants. In this case system development is described by the homogeneous linear equation on a variable $x(t)$ at parameters α and γ

$$\frac{dx(t)}{dt} = \alpha(t) \times x(t) \times \varphi\left(\frac{1-\alpha(t)}{\alpha(t)}\right) + \gamma \times x(t) \tag{2.2}$$

The optimum proportion $\alpha^*(t)$ between productive system and its development potential is defined from the condition

$$\alpha^* \times \varphi\left(\frac{1-\alpha^*}{\alpha^*}\right) \to \max \tag{2.3}$$

At the natural assumption that $\varphi(\xi)$ is monotone function with saturation (Fig. 2) there is a simple way of its optimum definition, as

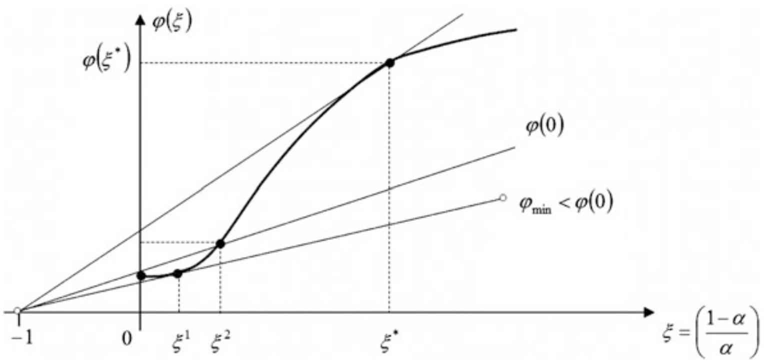


Fig. 2 Dependence of the "development potential" on the resources expended

$$\alpha = \left(\frac{1}{1+\xi}\right).$$

According to Fig. 2 it is clear that this optimum is reached in some point ξ^*, having quite certain sense. So, if resources for development potential are allocated « excessively much » $(\xi \succ \xi^*)$, then means $(\xi - \xi^*)$ are incorrectly withdrawn from the current reproduction and there is a situation when efforts to studying and counteraction to numerous risks which the developing system can never come across are spent.

The point $(\xi = 0)$ corresponds to the situation when all resources are spent exclusively for growth of productive system. The potential of similar system is low because of constant losses which it is possible to avoid if there is a potential for prediction of arising risks and struggle against them.

The segment $\xi = (0, \xi^1)$ shows that if the means allocated for studying and counteraction to threats and risks are small, then return from similar researches and done actions less than the resources allocated for them. Information gathering, research of internal and external threats on a low level doesn't allow to receive an adequate estimation for improvement of decision-making quality in most cases anyhow developing circumstances.

On the segment $\xi = (\xi^1, \xi^*)$ the contribution to development potential starts to give positive return, however only in the point ξ^2 the level of "self-support" of expenses for development of "potential" of system will be reached $(\varphi(\xi^2) = \varphi(0))$.

Therefore, it is expedient to consider this point as the point of "critical" position.

Decrease of potential $(\varphi(\xi))$ to the level $(\varphi(\xi^2))$ threatens that "in accordance with the circumstances" economically expediently there will be « strategy of survival »—strategy of full refusal of expenses for the decision of problems of prediction and anticipation of threats and risks and reproduction maintenance only at the expense of escalating inefficient capacities in productive subsystem $\xi \to 0$.

In spite of the fact that the conceptual model stated above is schematically, it, nevertheless, provides guidance that threats and risks as a matter of fact are "anti-potentials" of development, that is they are retarders of speed of all system reproduction. Since SCTES, as a rule, are non-uniform, they are subject to risks various by the nature and on influence levels. The received expert estimations of optimum proportions, certainly, need updating if to consider balanced development of the system consisting from many productively and territorially connected subsystems.

The logic of optimum proportional development in this case also remains. Received estimations should be considered only as "reference points" for the further researches, otherwise struggle for escalating of development potential will be carried out only in those territories and only in those productive-technological chains for which it by theoretical estimations is "economically expedient" that will lead to destruction of integrity of the system (connectivity loss), withdrawal from unified state and branch standards.

Let's suppose that the exit of investigated system runway from admissible corridor (a component of the vector of functioning efficiency indicators) can be caused to four reasons:

(a) owing to increase of importance of the problems put before the system to such level that default of these problems at occurrence of extraordinary situation (and furthermore in a normal mode of functioning) appears critical for system existence, up to necessity of its re-structuring as a whole;
(b) owing to system simplification or destruction at which locally arising extraordinary situations are really capable to outgrow in events with large-scale losses under scenarios of cascade type;
(c) owing to dramatic or long deterioration of operating conditions of objects and subsystems in one or several territorial formations formed, including, as a result of non-formalizable threats increase;
(d) owing to decrease in level of industrial and fire safety and (or) physical protection for technological blocks and objects of various type.

It is offered for an estimation of extraordinary situation threat level in SCTES to use the following hierarchical multi-criteria model [15].

Integrated risk of extraordinary situation $R(r_1, \ldots r_i, \ldots, r_n)$ represents function of risks of private extraordinary situations occurrence r_i $(i = 1, \ldots, n)$. The kind of dependence R on the arguments gets out proceeding from conditions:

$$0 \le R(r_1, \ldots, r_i, \ldots, r_n) \le 1 \tag{2.4}$$

$$R(0, \ldots, 0, \ldots, 0) = 0; \tag{2.5}$$

$$R(0, \ldots, r_i, \ldots, 0) = r_i; \tag{2.6}$$

$$0 \le R(r_1, \ldots, 1, \ldots, r_n) = 1 \tag{2.7}$$

for $\forall\ r_i = 1$ irrespective of values of other arguments.

Continuous function $R(r_1, \ldots r_i, \ldots, r_n)$, meeting (2.4)–(2.7), has the following general view

$$R(r_1, \ldots r_i, \ldots, r_n) = 1 - \left\{ \prod_{i=1}^{n} (1 - r_i) \right\} \times g(r_1, \ldots r_i, \ldots, r_n), \tag{2.8}$$

where $g(0, \ldots, r_i, \ldots, 0) = 1$.

If in special case $g(r_1, \ldots r_i, \ldots, r_n) \equiv 1$, then formula (2.8) is of the form

$$R(r_1, \ldots r_i, \ldots, r_n) = 1 - \left\{ \prod_{i=1}^{n} (1 - r_i) \right\} \tag{2.9}$$

states the underestimated estimation of integrated risk from calculation that the stream of extraordinary situations represents a mix of ordinary events taken from homogeneous, but differing with values r_i $(i = 1, \ldots, n)$ samples.

But for real systems risks, as a rule, are dependent.

Then we have

$$g(r_1, \ldots r_i, \ldots, r_n) = 1 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij} \times [r_i]^{\alpha_{ij}} \times [r_j]^{\beta_{ij}} \tag{2.10}$$

$$\begin{aligned} &\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij} \leq 1 \quad C_{ij} \geq 0 \\ &\alpha_{ij} > 0 \\ &\beta_{ij} > 0 \end{aligned} \tag{2.11}$$

where C_{ij}—coefficients of risks connection of i and j extraordinary situation; α_{ij} and β_{ij}—positive coefficients of elasticity of replacement of corresponding risks, allow to consider the facts of risks replacement, caused mainly by that simultaneously effective actions for decrease in all risks can't be done owing to limitation of time and resources.

The current risks values r_i $(i = 1, \ldots, n)$, entering integral risks factor R are values changed in time with various speeds (for example depending on the seasonal factor priorities of solved technological problems essentially change).

Thereof classical calculation of risks equation leads to problems of combinatory complexity on the initial data having objectively casual, uncertain, often qualitative (semi-quantitative) nature.

The decision of problems of risks analysis becomes complicated also because non-formalizable threats can play considerable role.

For the account of these factors it is offered to form values r_i as product of four components:

$$r_i = r_i^{(a)} r_i^{(b)} r_i^{(c)} r_i^{(d)}. \tag{2.12}$$

The component $r_i^{(a)}$ in (2.12) is estimated through categorizing of problems the performance of which is cancelled or delayed owing to the arisen extraordinary situation (for example, in gas supply systems categorizing can be defined through percentage distribution of categories of power users, affected in case of extraordinary situation because of the termination of gas supply).

The component $r_i^{(b)}$ is estimated through maximum permissible losses (MPL) in extraordinary situations at existing technological levels and materials (subjectively established) calculations of such losses.

Before achievement the level of MPL $r_i^{(b)}$ can be considered as linear function

$$r_i^{(b)} = \frac{\mathrm{L}_i}{\mathrm{MPL}}, \tag{2.13}$$

where L_i—current level of losses.

At exceeding of the level of MPL $r_i^{(b)}$ fixed with value 1.

The multiplier $r_i^{(c)}$ is estimated as dimensionless value, is calculated under empirically picked up statistical data about characteristics of objects in a binding to their territorial placing and has the meaning of indicator of absolute vulnerability of object on which the scenario of i extraordinary situation is initiated.

In general, $r_i^{(c)}$ can be considered as the indicator of environment aggression in which the object functions.

For each territory owing to geographical factors, features of productive structure, sociocultural, ethnic and other differences the construction of unique models of calculation significantly found on personal assessment of the experts familiar with this specificity is required.

And a last, valuation of $r_i^{(d)}$ is in terms of ranging of objects types. It reflects quality of relative "susceptibility" of objects of the set type on a wide range of external changes of the factors defining $r_i^{(c)}$. Values $r_i^{(d)}$ are used so that to result estimations of risks of extraordinary situations initiated by events on objects of various types to a uniform scale.

The offered scheme of calculation of integrated index R is mainly intended for the preliminary analysis of variants of system development on the basis of hierarchy of the indicators characterizing all aspects of extraordinary situations including both estimations of consequences $r_i^{(a)}$, $r_i^{(b)}$, and estimations of causes $r_i^{(c)}$, $r_i^{(d)}$.

Lines of levels of R values allow to build borders of reaction zones to changes of all spectrum of risks: $R \geq R_{\mathrm{red}}$ (a "red" zone demanding change of the existing mode of functioning or additional means for decrease of risks $r_i^{(c)}$ and $r_i^{(d)}$), $R_{\mathrm{red}} \geq R \geq R_{\mathrm{or}}$ (an « orange » zone demanding balancing of contract activity, carrying out of diagnostic and other actions), $R \leq R_{\mathrm{or}}$ (a "green" zone in which pertinently to speak about growth and the further development of the system, introduction of new capacities and new risks connected with their occurrence).

It is obvious that between developing object what is any of SCTES and its development potential, one of the components of which is the subsystem of safety (risk) management there should be a balance. High-yielding system with the underestimated risk is doomed to inefficient functioning owing to losses and on the contrary, the excessive safety concern leads to withdrawal of resources from a cycle of reproduction.

Thus, for SCTES, having diversified multiphasic production and difficult space-territorial topology of estimations having only economic character are unacceptable. Expediently complex development of development potential control system which can be realized, for example, within the limits of audit formly occurred crisis and pre-critical situations, estimation and generalization of experience, development of system of indicators of early detection of threats to steady functioning of objects (groups of objects).

Let's notice that during development of such indicators system for realization of situational approach to management of the company, it is necessary to accept a number of "reconciliatory agreements ».

The first main agreement consists that preservation of integrity of system and knowledge (information) about it is more significant, rather than economic success of separately taken productive element or productive-territorial formation.

The second agreement: threats and risks are considered as factors, braking development potential and, accordingly, use of the device of an estimation of the analysis and risks management without taking charge for risks corresponding to their competence by regulatory bodies is impossible. Concealment of risks or their revaluation'll become a subject of economic auction, inappropriate in the conditions of approaching threats.

The third agreement: pure « one-criteria » approach when to every discovered risk (social, economic, foreign policy) "cost" estimation of consequences of its realizations and (or) prevention of scenarios of their expansion is offered, isn't universal.

Activity of any person separately, groups of people, labor collective of the company as a whole many-sided and various, the use of multi-criteria approach with elements of "indistinct" logic, with use (as far as data permits) detection device of the latent laws in conjunction with and mutual strengthening of numerous factors therefore is the most pertinent.

The methodological approach taken as a principle offered method has advantage in comparison with the cost approaches, expressed that multi-criteria utility "absorbs" in itself in "share" participation all factors, but not just having cost expression ones (that, however, doesn't allow to remove all uncertainties).

Multi-criteria utility is formed on the indicators having in the basis different dimensions, units and scales of measurements which are easily arranged in the presence of computing resources to specific features of investigated objects and risks generated by numerous factors at different circumstances of place and time.

Classical schemes of multi-criteria analysis, based on linear and multiplicative convolutions are successfully enough used in design and predesign analysis, at the decision of some problems of situational management in marketing, at risks estimation of continuation or termination of research and developmental works but as SCTES is dynamic system, it is offered to use more developed model of dynamic multi-criteria analysis which will allow to include the situations generated so-called non-formalizable threats and risks-factors into consideration.

3 Risk Analysis

In many cases risk analysis is the only opportunity to investigate those safety issues on which the answer from statistics can't be received as, for example, accidents with small probability of realization, but with big potential consequences. Of course, risk analysis isn't the solution of all problems of safety but only using it, it is possible to compare risks from various sources of danger, to allocate the most essential of them, to choose the most effective and economic systems on increase in safety, to develop actions for decrease in consequences of accidents, etc.

In a foreign press together with the name "Risk Analysis" they sometimes use the name PRA (Probabilistic Risk Analysis), a probabilistic risk analysis approved by NRC (National Research Council). There is no basic distinction between them though it is considered that PRA is mainly aimed on the analysis of accidents with low probability, however through PRA events with a wide range of probability of emergence are often investigated.

At the present time the procedure of risk analysis can be divided conditionally into two main components and several intermediate each of which is characterized by its problems and uses its methods and models:

- risk assessment the main purpose of which is identification of dangers and quantitative description of potential damage (at this stage of analysis models are developed and measures which can be used for finding the risk size are defined, and also those factors which significantly influence this size are defined);
- risk management when priority areas and necessary means for increase in safety are defined (this stage uses results of risk assessment and intermediate phase of risk determination for security policy formulation).

Certainly risk analysis isn't completely finished and perfect tool for a research of problems of industrial safety yet and its improvements are possible and necessary, but even for modern level of industry development it became clear that the decisions made in this area on the basis of risk analysis already today will define the directions of industry and energetics development, production efficiency and wealth status of the population.

It is also an important circumstance that questions of risk analysis can't be considered separately from game statement. At the same time basic formulas in risk analysis are perverted, simplified, their belonging to game theory is forgotten. There are several reasons for it. The word "risk" became "fashionable", as a result experts "jumped on the term" without understanding from where it occurs, what axioms "are put" in this term. As a result, economists, insurers, ecologists, and others during many years produce false scientific results proceeding from their false thought-up definitions. Sometimes (multiplying "lie" by "lie" gives the "truth") they receive acceptable results. But, as a rule, it concerns only static and stationary cases (where there is the theory of "reliability"), but not dynamic cases. For a number of appendices, it was necessary to simplify a formula of risk calculation. As a result, risk as a dynamic characteristic, depends on time, means and information came down to two-dimensional "pictures of photos" where there are only probabilities and damage. In modern risk analysis "the theory of strength" and "the theory of reliability" have been left. But researches on "the theory of survivability", "theories of a gomeostazis", adaptive theories, including "the theory of decisions choosing", "the theory of perspective activity", "the theory of reflections", "the theory of the self-organized systems" have been curtailed.

It is necessary to remember basic difference between stochastic factors leading to decision-making in the conditions of risk and uncertain factors leading to decision-making in the conditions of uncertainty. Both result in dispersion of possible outcomes of management results. But stochastic factors are completely described by the known stochastic information, this information allows to choose the best decision on average.

In respect to uncertain factors similar information is absent. Generally, uncertainty can be caused either counteraction of reasonable opponent (more difficult case connected with the opponent's reflections (terrorist threat)), or insufficient awareness on conditions in which the choice of the decision is carried out.

Let us consider the principles of decisions choice with insufficient awareness concerning conditions in which the choice is carried out. Such situations are usually named "games with the nature".

In terms of "games with nature" the problem of decision-making can be formulated as follows. Let the person making the decision can choose one of m of possible options of his decisions: X_1, X_2, …, X_M and concerning conditions in which possible options will be realized it's possible to make n supposals: Y_1, Y_2, …, Y_N. Assessments of each option of the decision in each conditions (X_m, Y_n) are known and set in the form of gains matrix of the DM: $A = A\ (X_m, Y_n) = |A_{mn}|$.

Let us assume in the beginning that aprioristic information on probabilities of emergence of any situation of Y_n is absent.

The theory of statistical decisions offers several criteria of optimality of decisions choosing. We don't formalize the choice of this or that criterion, it is carried out by the DM, subjectively proceeding from his experience, intuition, etc. Let us consider these criteria.

Laplace criterion. As probabilities of occurring of this or that situation Y_n are unknown, let us consider all of them equiprobable. Then for every line of gain matrix the arithmetic average value of assessments is counted. To the optimal solution there will correspond such decision to which there corresponds the maximum value of this arithmetic average, i.e.

$$\overline{F} = F(\overline{X}, Y) = \max_{1 \le m \le M} \left(\frac{1}{N} \sum_{n=1}^{N} A_{mn} \right) \tag{3.1}$$

Wald criterion. In each line of matrix, we choose the minimum assessment. To the optimal solution there corresponds such decision to which there corresponds the maximum of this minimum, i.e.

$$\overline{F} = F(\overline{X}, Y) = \max_{1 \le m \le M} \left(\min_{1 \le n \le N} (A_{mn}) \right) \tag{3.2}$$

This criterion is very careful. It is focused on the worst conditions only among which you can find the best and now guaranteed result.

Savage criterion. In each column of matrix there is maximum assessment of $\overline{A}_n = \max_{1<m<M}(A_m)$ and a new matrix which elements are defined by a ratio of $R_{mn} = \overline{A}_n - A_{mn}$ is formed. It is the size of regrets that at the strategy of Y_n not optimum choice of X_m is made.

The value R_{mn} is called risk which means a difference between maximum win which would take place if it was well aware that there would come the most favorable situation $\overline{Y}_n$ for the DM, and a real gain at the choice of the decision of X_m in the conditions of Y_n.

This new matrix is called risk matrix. Further from risk matrix they choose such decision when the value of risk accepts the smallest value in the most adverse situation, i.e.

$$\overline{F} = F(\overline{X}, Y) = \min_{1 \le m \le M} \left(\max_{1 \le n \le N} (R_{mn}) \right). \tag{3.3}$$

The content of this criterion is in risk minimization. As well as Wald criterion, Savage criterion is very careful. They differ with different understanding of the worst situation: in the first case it is minimum gain, in the second it is maximum loss of gain in comparison with what could be reached in these conditions.

Hurwitz criterion. Some coefficient α, named "optimism coefficient", $0 < \alpha < 1$ is entered. In every line of gains matrix there is the biggest assessment of $\max_{1<n<N}(A_{mn})$ and the smallest one $\min_{1<n<N}(A_{mn})$.

They are multiplied respectively on α and $(1 - \alpha)$ and then their sum is calculated. To the optimal solution there will correspond such decision to which there corresponds the maximum of this sum, i.e.

$$\overline{F} = F(\overline{X}, Y) = \max_{1 \le m \le M} \left(\alpha \times \max_{1 \le n \le N} (A_{mn}) + (1 - \alpha) \times \min_{1 \le n \le N} (A_{mn}) \right) \tag{3.4}$$

At ($\alpha = 0$) Hurwitz criterion is transformed to Wald criterion. It is a case of extreme "pessimism". At ($\alpha = 1$) a case of extreme "optimism") the DM expects that he will be accompanied by the most favorable situation. "Optimism coefficient" α is appointed subjectively, on the basis of experience, intuitions, etc. The more the situation is dangerous, the more careful approach to the choice of the decision has to be and smaller value is appropriated to coefficient α.

But this criterion has no relation to risk analysis. Unless too subjective perception of "casual" and "voluntary" risks.

Proceeding from the above risk assessment is possible only in the presence of alternatives of the choice. If there is the only one option of the choice, then the risk is automatically equal to zero and the variation of payments is only the characteristic of uncontrollable environment. However, the alternative is always present as a refusal to make decision. In some cases, the refusal to make decision can give optimum in columns and then not zero risks in options due to the choice of wrong decision will appear.

Existence of assessments of probabilities $\sum_{n=1}^{N} p_n = 1$ for the description of environment condition $p_1 = p(Y_1)$, $p_2 = p(Y_2)$,…, $p_n = p(Y_n)$ allows to refuse the choice of the most negative case when using Savage criterion and to write down the required decision as follows:

$$\overline{F} = F(\overline{X}, Y) = \min_{1 \leq m \leq M} \left(\sum_{n=1}^{N} p_n \times \left(\max_{1 \leq n \leq N} (A_{mn}) - A_{mn} \right) \right) \tag{3.5}$$

It is more correct formula. Only in that case when for any couple (X_m, Y_n) the payment is defined only by the size of losses $A_{mn} = B - C_{mn}$

$$\bar{F} = F(\bar{X}, Y) = \min_{1 \leq m \leq M} \left(\sum_{n=1}^{M} p_n \times (B - C_{mn}) \right) = B + \min_{1 \leq m \leq M} \left(\sum_{n=1}^{M} p_n \times C_{mn} \right) \tag{3.6}$$

And only in that case when the level of losses at optimal variant for conditions of $Y_1, Y_2, ..., Y_N$ doesn't depend on n and is equal to $\bar{C}$, then:

$$\begin{aligned} \bar{F} = F(\bar{X}, Y) &= \min_{1 \leq m \leq M} \left(\sum_{n=1}^{M} p_n \times (B - C_{mn}) \right) \\ &= B - \bar{C} + \min_{1 \leq m \leq M} \left(\sum_{n=1}^{M} p_n \times C_{mn} \right) \end{aligned} \tag{3.7}$$

Only in this case the decision will be defined by the size of math expectation of losses. But adjusted for B and $\bar{C}$. Failure to take account of these amendments contains in many works. Usually put B and $\bar{C}$ equal to zero, and that's all. For example, in ecology to improve "air" goes for nothing (doesn't make profit) and if nobody is sick, then the optimum damage is accepted for 0.

Bayes Criterion leads to the same assessments:

$$\begin{aligned}\bar{F} = F(\bar{X}, Y) &= \max_{1 \le m \le M}\left(\sum_{n=1}^{M} p_n \times A_{mn}\right) = (B = 0; \bar{C} = 0) \\ &= \min_{1 \le m \le M}\left(\sum_{n=1}^{M} p_n \times C_{mn}\right)\end{aligned} \tag{3.8}$$

The present-day and prospective structure of the global fuel and energy complex substantially compounds solution of the problems of ensuring its steady functioning and reliable supply of consumers. The intersectoral nature of these problems generates a need for formulating the coordinated solutions aimed at ensuring reliability and safety of separate specialized subsystems included in fuel and energy complex with allowance for their interrelations beginning from the projected growth of the energy systems for 15–20 years and ending with working management of the systems during their operation. In a more comprehensive sense the intersectoral nature of the problem of ensuring the stable functioning of the fuel and energy complex is determined by interrelations of power industry with such other sectors of the national economy as power-plant industry, electrical engineering industry, metallurgy, instrument-making industry, transport, supply of materials and machinery, etc., as well as with the social and economic development plans of countries and regions.

The necessity of the intersectoral approach to the problem of ensuring the security and stable functioning of the fuel and energy complex requires the development of a uniform methodical approach to investigation of reliability, safety, and stability of the fuel and energy complex components represented by various energy systems. Such an approach should take into account existence of several common specific features of various energy systems making it possible to solve the above problem from the common standpoint both theoretically and methodically. These common specific features of the fuel and energy complex include: interrelation with other national economy systems (industries); territorial distribution and complexity; continuity and persistence of development; uninterrupted functioning and interrelation of the operation modes of the system components (parts); multipurpose nature and practical impossibility of the system's complete failure; irregularity of the products consumption process; exposure to large-scale external effects—both unintentional and intentional; possibility of cascaded development of breakdowns; dependence of capacities of the intercommunications on their locations, operation modes of the system and composition of the operating equipment; hierarchical pattern; diversity of reliability ensuring objects; active participation of a human being in the management process; uncertainty, incompleteness, insufficient authenticity of the information on the system parameters and modes. Inconsistent (adaptive) behavior principles under the conditions of possible risks and uncertainties are inherently laid as a foundation for the idea of management of the most of the fuel and energy complex subsystems. Existence of such risks originated by heterogeneous circumstances is able to block or vary this or that way of advancement and make the system exist "under another scenario" differing from multiplicity of previously shaped plans (Fig. 3).

The problem of ensuring security of the fuel and energy complex objects under the conditions of the varying scope and intensity of threats to stable development of the industry retains its relevance over extended periods [16–19].

The security requirements established for the higher- and medium-grade hazard objects are rather high and drastically increase the expenses of the objects' owners. It is practically impossible to elevate protection and security to the level required by the federal legislation at a single step. This brings up the question of ranking the objects within the preset grades for determining the order of priority for provision of the objects with the required protection means. To do this, it is necessary to specify the criterion, relative to which the importance (and, accordingly, the serial number) of this or that object in the ranked list will be determined.

Within the framework of countering the anthropogenic threats some corporations has developed and introduced the corporate standards, which are used to ensure uniform approaches to security organization for a certain category of objects, as well as establish the principles and rules for classification of protection sites by probable aftereffects (risks) of terrorist acts and are applicable for identification of vital facilities and the facilities first and foremost subject to be equipped with technical security equipment sets, formulation of the requirements to anti-terrorism security of the protection sites, and determination of sufficiency of time-sensitive and long-term solutions for their protection. For instance, some of the standards uses the class pertinent to the degree of risk of occurrence of the event, which is negative for the facility functioning, as a criterion for referral of the facilities to the category of the vital ones. This class takes into account the potential danger and terrorist vulnerability categories of the object to be determined by the respective

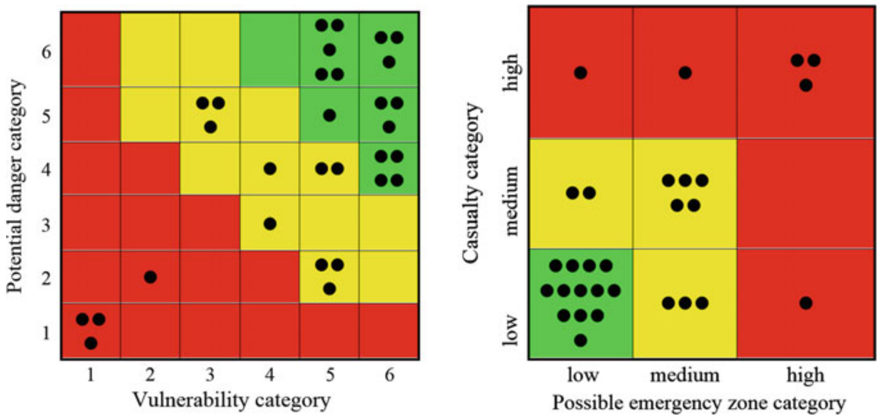


Fig. 3 Rating of objects by potential danger and vulnerability [43]; rating of objects by possible emergency zone and casualties [43]

procedures. While solving the problem of classification on the whole, these approaches do not allow unambiguous solution of the problem of rating the objects with allowance for their importance (significance) for the whole SCTES being considered. The differences in the classification (rating) and ranking problems are clearly shown in Figs. 4 and 5 [43].

The economic security aspects are always of immediate interest. It is a historically arisen concept of the need of both complex protections of everybody from the hazard actualization threats on the level of the "as low as reasonably practicable" principle (first-type problem) and the needs of specific individual protection, whose level is determined from the circumstances of place and time for the protected object (second-type problem). If the collective security mechanisms are ensured by the systems of the "armed forces", "common law-enforcement authorities", "emergency action services" types, specific individual security of high-security objects is provided by specialized bodies in compliance with the normative standards exceeding the normative standards for protection of average objects.

The increase of the protection level of the fuel and energy complex objects in response to the growth of the terrorist threat belongs to the second-type problems. If for protection against common criminal activities (thefts, vandalism, etc.) it is sufficient to satisfy the "average industry standards for all objects", protection against terrorism implies the understanding that the acts of terrorism are single and rare events.

The requirements, which are correct for the collective security mechanisms, in their pure form are not conceptually applicable for attaining the goal to be sought—ranking of the objects by their measure of significance for the system, where they are functioning. They are ineffective in the sphere of enhancement of stability of

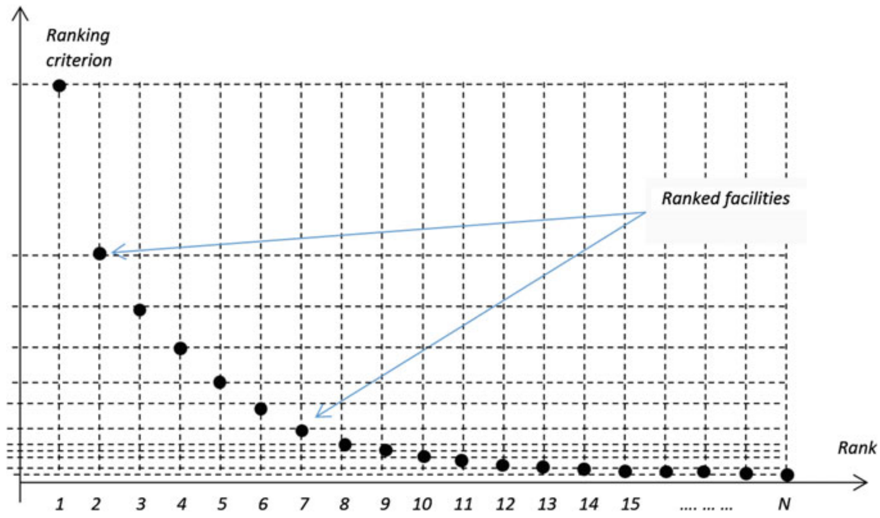


Fig. 4 Objects ranking results [43]

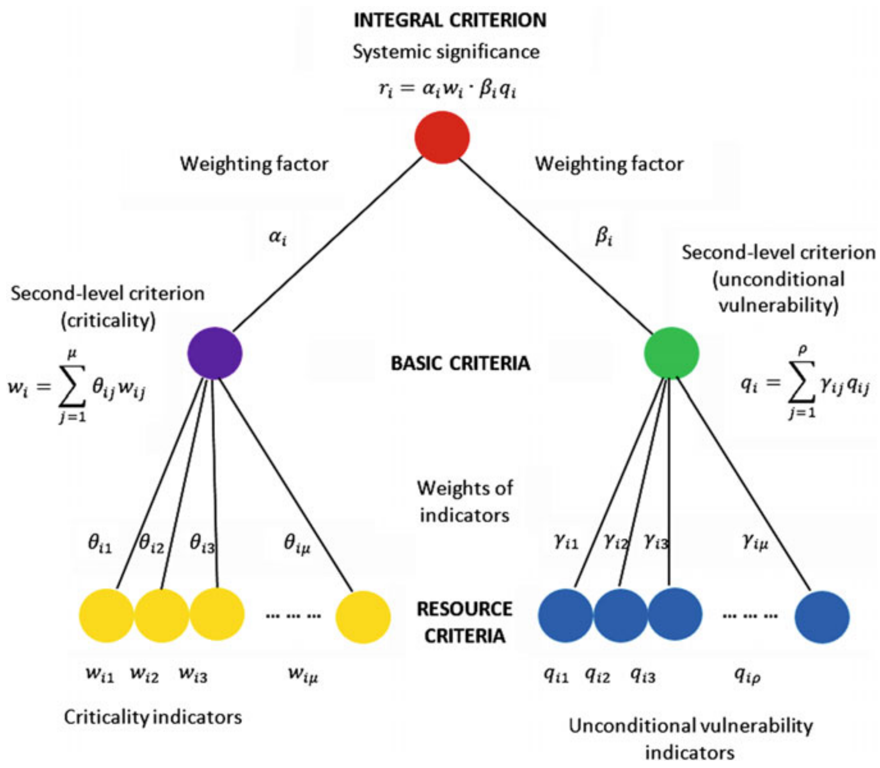


Fig. 5 Multilevel hierarchy of indicators and homogeneous objects assessment criteria [43]

these objects as construction of additional protective barriers on those objects, to which hardly anything threaten, is actually transformed into "underdelivery" of the protection means to the objects being "most attractive" for terrorists.

In a general case, meant by ranking is the procedure of seriation of any objects by the increase or decrease of a certain feature provided they really possess this feature. For instance, the fuel and energy complex objects can be ranked by their cost and volume of manufactured or pumped products, the research and development can be ranked by their importance for the Society development strategy, the investment projects can be ranked by NPV, etc.

The *system significance criterion* was selected as the ranking criterion for the SCTES objects. The system significance was defined as the object's feature characterizing the degree of its importance for the infrastructure and the life-support system of the fuel and energy complex, within which it functions. This is a complex feature that can comprise the criticality and unconditional vulnerability features, as well as a combination of these features.

The ranking problem proper is not a new one. The objects ranking methods are based on mathematical simulation, expert appraisals, decision-making theory, and interval estimation [20–22]. To any extent they take into account the interests of the organizations operating these objects, state supervisory agencies, and insurance companies. At the same time, the currently available ranking methods (for instance, ranking of the objects by the extent of protection in case of emergency situations on railroad transport [23], ranking of the hazardous production facilities of the gas distribution systems [24], etc.) do not take into account the characteristic properties of the structural connectivity of the ranking objects and importance of the specific object operation for interfacing systems and subsystems.

The objects ranking problem is a topical problem of the theory of measuring some complex synthetic features of the objects [25]. Formally solution of the problem is reduced to plotting some validity or usefulness function which links the measured feature with simpler resource indicators (factors) measured in actual values [26]. The validity function is used for both solution of the problems of selecting some best variant from the set of alternatives [27], and solution of the more compositional problems like the problem of business acquisition consisting of work orders with resource constraints (scopes of financing the objects' creation or modification) [28]. As the factors employed for arranging the ranks are often measured in qualitative scales rather than in quantitative ones, use should be made of the expert evaluation methods and expert system technologies for plotting the relationships between usefulness and primary resource factors [28, 29]. Owing to development of computer engineering there appeared a possibility of assessing the objects, whose description factors are set with an error. This fact necessitates development of a specific apparatus for statistical processing of the primary data [30] and use of the fuzzy logic tools [31–33].

The essential trait of solution of the ranking problems consists in the adaptive nature of the decision-making procedure for selection of optimal variants [32, 34], under which it is necessary to hold several experimental data and expert preferences correlation cycles for development of the final formula [33].

All other conditions being equal, the terrorist having some means for hitting a single target selects the object performing the greatest scope of the commodity-transport activity (W). The first basic criterion for assessment of systemic significance of the SCTES object as the fuel and energy complex component is obtained on the assumption that the effect from the functioning disruption is linear within wide variation range W. The given criterion assesses under deliveries of products as compared with the ideal functioning mode of the SCTES as the whole system. This is an estimate indicator closely related to the power rating of the SCTES object. It is calculated with aid of the object functioning models. In particular, use is made of gas flow models in the SCTES for the compressor and gas distribution stations and the underground gas storage facilities, the daily volumes of shipped products for the plants, etc.

However, the objects, owing to the circumstances of place and time of their functioning (seasonality), are distinguished by the principle difference in the effects from the under deliveries. One and the same gas is used in different process chains. Therefore, the terrorist finds it much more "profitable" (under otherwise equal conditions) to hit the object fulfilling the more important, more vital, and more "highly paid" work or the one that may entail great punitive sanctions (for instance, in case of violation of the export commitments).

So, the object acquires another indicator (let us call it α) showing the importance of fulfilling the unit of work unit of work. In other words, there can be cases when $W_1 > W_2$, but at the same time $\alpha_1 \cdot W_1 < \alpha_2 \cdot W_2$, i.e. when another object becomes more attractive for the terrorist due to the fact that he destroys a kind of a more "qualified object" or a "more critical resource".

The third indicator (Q) used by the suggested approach is the indicator of potential feasibility of the planned action. This feasibility is associated with a possibility of weapon delivery and presence of potential accomplices in the region of the target object. By analogy with the technical systems this indicator shows the aggression level of the external environment where the object is operating. With the advent of this indicator the objects hard-to-get for terrorists with a lower value of this indicator are not included in the list of potential targets by them as there are targets having the same effect, but located, for instance, closer to the terrorist base, to the RF border, etc.

The fourth indicator (β) reflects comparison of the objects of various types by their attractiveness for terrorists. This indicator rates average accessibility of the points of application of means of destruction depending on the "layout characteristics" of the objects in the SCTES. In this indicator the values of the killing effects are adjusted. Thus, in the case of the SCTES, for example, the gas-distributing station (GDS) is closer to the consumer and often has no backup, while the gas-compressor station (GCS), on the contrary, has, as a rule, both intrashop switchover of gas compressor units (GCU) and a ramified system of loop lines on the multi-line mains. Figuratively speaking, this indicator is the scale parameter of the object.

The basic criteria are developed with the use of the convolutions of resource indicators expressed in natural units, though the formulas of the convolutions are to reconstructed by expertise. Specific nature of the use of the convolution of indicators in terms of multipliers is associated with the fact that perception of expected losses (or rather perception of natural signals by human senses) has a logarithmic scale. To describe the relationships among the criteria, the oriented graph called the influence graph (Fig. 5) was introduced for consideration. The independent criteria called the resource ones are arranged in pairs on the lower level of this graph. Unlike the lower-level criteria, the systemic significance criterion is called the basic one.

Non-seclusion of the SCTES as an intricate system implies its interaction with the external environment and the effect of the latter on it. Generally speaking, this effect can be construed rather widely: there can be natural calamities (for instance, earthquakes leading to destruction of dams and other engineering structures), major accidents (for instance, explosion at the nuclear power plant, blackout of the whole region), as well as illegal actions characterized by the widest spectrum of effects. Precisely these external effects are characterized by great uncertainty of time, place and method of action, as well as selection of a specific object for commission of the action.

As importance of the object for the system under consideration and violator is the same, the required object protection level should be determined as a result of consideration of the nature of possible attacks. The following can be considered as such attacks:

First, the most commonly encountered “local criminal activities/practice” and the offences of the law hereto related, which do harm to the economic activity of the objects. As a rule, they are thefts. This local criminal activities (practice) also includes hooliganism (vandalism) and protest actions. Its level most probably correlates with the level of general criminal offences in the region of the object location. The latent (hidden) part of this kind of criminality can be rather adequately measured by such indicators as the unemployment level, share of migrants, and educational level of the population.

Second, migration of the intrastate criminal and terrorist activities. The zones of intensive terrorist activities tend to expand: criminal gangs “forced out” by the law-enforcement authorities migrate together with able-bodied population from the “trouble spots”. The most significant indicator reflecting this kind of offences is remoteness of the object from the zones of intensive terrorist activities.

Third, specially trained terrorist and guerrilla parties penetrating in full strength or partly as instructors from abroad. Their actions are characterized by good forethought, preparedness, and definiteness (planned nature of activity and weighted measurement of implement ability of this or that action by infliction of damage).

There can be mixed variants, as well. For instance, when the local criminal gangs and arriving “emissaries” merge on a common ideological or religious platform. This is particularly important under the conditions of current-day Russia, when the ideals and values of co-existence of different nationalities and population groups under protection of a strong paternalistic state.

For the purpose of the main problem—determination of the systemic significance of the SCTES objects it is suggested to consider the criminal underworld as the source of various external attacks against the objects, but the source with limited resources. We are more interested to a greater extent in the attacks with high and medium levels of preparedness and it is suggested to consider that the criminal underworld involves the whole spectrum of its capabilities. More specifically, it should be expect both launching large-scale attacks in order to make the owner of the fuel and energy complex objects bankrupt by forcing him to spend the finances for reinforcement of the objects’ physical protection and launching the attacks with medium-level preparation as excessive preparedness of the attacks is not expedient under the conditions when the company lack resources needed to protect all its objects according to the variant of the protection systems of the best industrial prototypes.

Thus, let us consider a certain (k-th) SCTES object. As a result of the anticipated attack by intruders having this or that training level, a certain damage (X) will be inflicted to this object through full (or partial) outage. Note that not every attack will be a priori successful for the attacker and therefore the protection profile of the k-th object can be described by interval representations through application of four matrixes of the following kind: $Q^{[k]}_{\min}(i,j)$, $Q^{[k]}_{\max}(i,j)$, $X^{[k]}_{\min}(i,j)$, $X^{[k]}_{\max}(i,j)$, where $i = 0, 1, \ldots, I^{[k]}$ is the protection level for the k-th object. The. zero level (i = 0) corresponds to the current protection state. Interpretation of the matrix elements is as follows: if the enemy attack with preparedness level j is launched against the above k-th object with protection level i, the damage ranging from $X^{[k]}_{\min}(i,j)$ to $X^{[k]}_{\max}(i,j)$ to will be inflicted the object owner with a probability from $Q^{[k]}_{\min}(i,j)$ to $Q^{[k]}_{\max}(i,j)$.

It is obvious that the $Q^{[k]}_{\min}(i,j)$, $Q^{[k]}_{\max}(i,j)$, $X^{[k]}_{\min}(i,j)$, $X^{[k]}_{\max}(i,j)$ values will be growing with preparedness level j and will decrease progressively as protection level i of the object grows. It is also obvious that protection at any level will require certain material expenses on the part of the company and state. let us denote the value of expenses for creation and maintenance of protection of object k at the i-th level by $Y^{[k]}\left(i^{[k]}\right)$. As the entire resource allocated for protection of all objects is limited, the following in equation should be implemented:

$$\sum\nolimits_k Y^{[k]}(i^k) \leq Y \tag{3.9}$$

Here Y is the sum of all expenses for protection of the objects provided protection system variant $i^{[k]}$ is selected for each object k.

If the criminal does not possess the advantage of choosing the target and variant of attack, i.e., criminal activities/practice is nonselective like the nature or technological failures, the optimal objects' protection profile could be attained by sequential execution of the following algorithm:

(a) assess probability $\lambda^{[k]}(j)$ of the attack against each k-th object by the enemy with the j-th preparedness level;
(b) calculate the median value of the risk from implementation of the attack against the k-th object by the enemy with the j-th preparedness level at the $i^{[k]}$ object protection system variant:

$$\begin{aligned} R[k, i^{[k]}] = \sum_{j=0}^{j} \Bigg\{ \lambda^{[k]}(j) \times \left(\frac{Q^{[k]}_{\min}(i^{[k]}, j) + Q^{[k]}_{\max}(i^{[k]}, j)}{2} \right) \\ \times \left(\frac{X^{[k]}_{\min}(i^{[k]}, j) + X^{[k]}_{\max}(i^{[k]}, j)}{2} \right) \end{aligned} \tag{3.10}$$

(c) determine the value of the prevented risk per unit of finances invested into protection:

$$\theta\left[k, i^{*[k]}\right] = \frac{R\left[k, i^{*[k]}\right]}{Y^{[k]}\left(i^{[k]}\right)} \tag{3.11}$$

(d) select the maximum prevented risk value for each k-th object:

$$\theta\left[k, i^{*[k]}\right] = \max_{i^{*[k]}}\left\{\theta\left[k, i^{*[k]}\right]\right\}, \tag{3.12}$$

i.e., the maximum risk reduction per unit of finances invested into protection for the k-th object is observed at selected variant $i^{*[k]}$.

(e) compile the ranked list of the objects and arrange them in the order of the value of indicator $\theta\left[k, i^{*[k]}\right]$ descending and then in the list count the first $\tilde{K}$ objects, the entire expenses for whose protection are covered by allocated finances Y, while the resources are insufficient for the $(\tilde{K}+1)$-th object. The essence of the above procedure is simple and clear: it is no use to raise money for additional protection of those objects, the threat to which is small to negligible (attack threat values $\lambda^{[k]}(j)$ are small). It is also inexpedient to additionally protect the objects if the temporary loss of their operating capability practically does not affect the value of the company's summary losses (the values of $X^{[k]}_{\max}(i^{[k]},j)$ are accordingly small). And, at last, additional protection is not expedient at the objects, which are so well protected that reduction of losses can be attained only by allocation of inadequately great funds (i.e., the values of $\theta\left[k, i^{*[k]}\right]$ are small).

The key point of the above algorithm consists in compiling the ranked list of the objects based on the minimization criterion for the losses' mathematical expectation per unit of finances invested into their protection (their stable functioning).

In the formula for $R\left[k, i^{[k]}\right]$ one can easily trace the necessity of the data collection and assessment by three components:

1. Values $X^{[k]}_{\min}(i^{[k]},j)$ and $X^{[k]}_{\max}(i^{[k]},j)$ of the losses caused by implementation of attacks.
2. Criminal environment aggressiveness indicator $\lambda^{[k]}(j)$.
3. Dependence of risks on the types of objects k.

As owing to the fact that the objects of SCTES are not stand alone enterprises the values of losses X should reflect the systemic effect (or socio-economic multi effect), which drastically increases depending on the users of the attacked object's products will suffer from the loss of the enterprise operating capability. So, it is needed not only to consider the upper boundaries of the damage indications, but also introduce additional 4th indicator, namely, the indicator showing importance of the continuous object functioning pertinent to a possibility of the cascading effect of increased consequences of the object outage for other national economy objects.

At last, it is expedient to additionally introduce the 5th component required for adequate ranking of the objects. The necessity of its introduction is caused by the fact that the enemy implements a task-oriented section of the attack. At the same time he has the validity factors and priorities displacing the values of $\lambda^{[k]}(j)$ from the "weighted average" not known either to the owner's security service, or to the competent state authorities. Sometimes these additional values have a specific nature: some criminals are prone to excessive blood-shedding and taking hostages, ritual killings, etc. Systemic significance of specific objects often grows for short time, for instance, during the stay of top public officials or ministers and particularly during the commissioning of the politically important production facilities of both international and regional (domestic) levels.

The above-mentioned fifth component is adjusting factor $\mu^{[k]}$ initially equal to unity for all objects can be (for instance, according to the security service, top management of the company, etc.) increased so that to increase the priority of inclusion of precisely the k-th object in the list of objects to be provided with additional protection measures due to the reasons not taken into account by the general rules.

Notwithstanding the fact that the considered problem theoretically has quite a large dimensionality and features great combinatorial complexity, it is perfectly well at hand due to monotonicity of the criteria employed and linearity of the systems of given constraints.

The key issues of this problem solution are of the informational and technological nature rather than the mathematical one:

- the assessments of the aftereffects of possible attacks launched by the enemy with preparedness level j should be available for each k-th object, which is not attainable for the time being;
- the whole SCTES needs appraisal of the risks, to which the objects are exposed within the complex of possible threats, including the weakly formalized ones (the more precise the assessment of the enemy's potential capabilities, which are both technologically and regionally non-homogeneous, the more effective the optimization of protection).

Within the framework of the considered setting allowing for the complex impact from the potential enemy activity the understanding of the protection systems' efficiency assessment cardinally changes. Thus, owing to limitedness of the resources immediately available for use by the criminal underworld one can obviously expect that the terrorists will shift the target setting from well protected objects (with low expected effectiveness of the attacks) to less protected objects (with greater effectiveness, but with lower one-time damages).

Another key element of the problem being considered consists in the fact that the search for effective solutions by both opposing forces is to a considerable extent carried out in the informational sphere:

- while preparing for the attack against the object, the criminal seeks for accomplices, which could help him choose the force application target attainable at his preparedness and equipment level;
- the protection system is capable of render a greater concentrated counteraction provided it is informed on the intentions of the crime.

Therefore, when describing the above procedure, it was repeatedly emphasized that we deal only with the appraisals made by both sides. Owing to unremovable uncertainty of appraisals in the quality of solving the problems of development of strategy and tactics for promoting the protection of the SCTES objects against possible deed of crime, including the acts of terrorism and attacks of subversive groups, it is expedient to "coarsen" the game formulation. In case of coarsening the "enemy capabilities should be idealized" and the characteristics of possible losses should be made more stringent, for instance, by way of changing over from the median risk assessments to the maximum ones.

So, the suggested approach operates on three key notions:

- systemic significance;
- criticality;
- unconditional vulnerability.

Systemic significance has been earlier defined.

Criticality is the feature of the object determining the extent of its influence on operating capability of the whole system with allowance for weighted consequences caused by its disconnection from various categories of consumers. Criticality cannot be determined by some element features, but should be determined within the framework of the whole system and its functional structure.

Unconditional vulnerability is the feature of the object characterizing the extent of reduction of its operating capability under the conditions of the environmental effects exceeding the boundaries of the normal operation conditions of the object. A specific class of external impacts associated with deliberate acts of human beings, for instance, sabotage can be singled out. Unconditional vulnerability characterizes a potential hazard for the object functioning at the present levels of the external factors, which are essentially the hazard characteristics of the territory of its location.

The methodological approach being the basic concept of the suggested method has an advantage over the cost approaches. This advantage consists in the fact that multicriterion usefulness "absorbs", on a shared basis, all factors, rather than only those expressed in monetary terms. Many present-day rating systems proceed only from the results of assessing one of the indicators describing the objects (for instance, activity of economic subjects, their criticality, etc.) [35, 36]. However, as practically both criticality and unconditional vulnerability of the objects (in the problem of objects ranking the by their systemic significance) arise from a great number of assessments, whose importance is not known beforehand, there appears a multicriterion ranking problem [37–40]. This problem belongs to the multicriterion decision making (selection) problems under the uncertainty conditions [31, 41] having a great importance for analysis of the systems of widely differing purposes [32, 42].

As in a general case the complicated system objects perform different functions and the results of their activity (or consequences of their outage) are differently assessed, it is important to know to what extent (how many times) one same-type object is more significant than another, as well as to be able to compare the assessments of the objects of various types. To do this, the necessity arose to introduce additional axioms specifying the classes of the functions of selection among heterogeneous objects (Table 1).

It should be understood that the general problem of selecting such axioms for the collections of objects comprising various types of objects has not been solved yet. There are several causes, of which the most substantial ones are:

1. A great dimensionality of the selection problem: the number of variants (objects, from which the choice should be made) and the number of indicators describing the state of each object are rather great. Data aggregation is required due to the fact that the time (number of operations) required for selection fundamentally grows with the dimensionality increase. Sorting and grouping of similar objects are most frequently used. In this case, simplifications of the real data (change-over from the quantitative indicators to the scoring and other ones) are implemented in the course of execution of the procedures, which permit deliberate accuracy decrease and information loss.
2. Diversity of the data types: Different attributes are measured against different scales and different objects are described by means of different sets of indicators.
3. Presence of "omitted values": "ovality" is often observed in statistics (due to various reasons). The authors of documents (especially text ones) omit the "implied" words and values "by default". It is often unjustified and explained by "lack of time".
4. Noisiness: Existence of indistinct and random indicators. The "measured" values used for selection are, as a rule, not equal to the true values, but just "close" to them. It is desirable to correct the systemic errors for the distorted values. The

Table 1 Axiomatics of arrangement of selection procedures

Basic axioms	Explanations
Inheritance axiom: if $O' \subseteq O$ then $\pi(o') \supseteq (\pi(o) \cap O')$	The axiom means that with the limited choice the "objects are the best of the best" variants belonging to $\pi(o) \cap O'$ and the objects, which are the best of those available in limited sampling $O' \subseteq O$, but would not be selected, if the choice is available in all alternatives O
Concordance axiom: $\prod_i \pi(O_i) \subseteq \pi\left(\bigcup_i O_i\right)$	If some object O was chosen as the best in each of sets O_i, it should be chosen from the whole aggregate of sets $\bigcup_i O_i$ as well
Omission axiom: $(\pi(o) \subseteq O' \subseteq O) \Rightarrow (\pi(o') = \pi(o))$	The axiom holds that if any part of the "rejected" objects is omitted, the selection result will not change

Note $\pi(o)$ is the true logic function, if the object is chosen to be the best; if O is the set of objects, $\pi(o)$ is the subset of the best of them features of additional distortions are different for the objects of various types and the selection variants should be agreed with the variants of processing these distortions.

5. Multicriteriality: It is practically impossible to indicate any single aim of functioning for any complex objects. The scales determining the target setting components are called the criterial scales and the respective variables, the criteria. As has been stated above, the practical selection problems are essentially the multicriterial ones.

Owing to the above causes it is expedient to solve the heterogeneous objects ranking problem in several steps. At the first step, particular models of systemic significance of the selected type of objects should be developed for each type of objects and then used for ranking. At the second step, it is necessary to carry out "sewing together" of the ranked lists of objects into a combined list. At the third step, the values of the assessments are adjusted if the necessity arises to take into account special functioning conditions of individual objects.

Uniformity of objects assumes that several describing variables (resource criteria) x_1, x_2, …, x_N can be suggested for them and the $Q(x_1, x_2, \ldots, x_N)$ scalar function can be set. This function for each object O takes on a value of $Q(x_1(O), x_2(O), \ldots, x_N(O))$.

By now several standardized approaches to description of the choice have been developed. The simplest variant is to imply that certain function Q called the criterion (quality criterion, objective function, preference function, utility function, etc.) can be set for all alternatives. This function possesses the feature establishing that if alternative x_2 is preferable to alternative x_1, then $Q(x_2) > q(x_1)$.

However, it is either difficult or practically impossible to construct such a function. At the same time the ideas of constructing, for instance, the utility functions for selection can be applicable at the primary variants selecting steps, when we try to interpolate a certain nonlinear utility scale on the basis of a limited amount of the data.

Moreover, to implement the procedure of selecting the most system-relevant significant object, in a "stepwise" manner we are interested only in the groups and their component objects, which are among the bidders for "the best of the best position" at each step. Selection of such objects in the simplest cases is reduced to the fact that the describing variables are considered to be a certain "test" (examination), while the attained values for the object are taken to be marks/credits under this test.

So, if there is certain limited amount M of describing variables (tests), the most system-relevant significant object is the one that will collect the greatest summary number of credits. The credits can be summed up with certain weighting factors reflecting relative significance of the m-th test $(m = 1, \ldots, M)$ with respect to the average test. The properly chosen tests should be generally capable of acting as adequate tools for exposition of criterial function $Q(x_1, x_2, \ldots, x_N)$ for assessment of systemic significance of the objects of the single type, but this is true only in the cases when the level of attaining a set of the object functioning purposes (as real objects are always multipurpose) can be adequately described by a linear combination of levels of attaining each specific purpose.

While turning back to solution of the initial problem of all lists merger let us note the known fact: if all pair-wise comparisons of different scales are sown together, there exists the procedure of arranging all objects in their common list, which will produce after renormalization into a certain basic type (for instance, comparison scale of gas-distributing stations, for example) universal assessments of systemic significance according to a unified scale of the objects of all types.

After arrangement of all objects in the common list, "objects — exceptions to the rules" "underestimated by the suggested method" will be found practically always in the course of the critical analysis of such an arrangement. The main cause of underestimation consists in the failure to take account of the factors (ecological, geological, ethnic, and other) and/or their combinations, which have not been included in the lists of resource indicators due to the fact that they do not feature the required generality for inclusion into a set of state variables of all objects.

In this case the experts should have a limited capability of shifting the object to the left in the list of the systemic significance objects by way of increasing the above-mentioned fifth factor from unity to a certain level. However, to prevent unscrupulous overestimation of this factor, the number of ranks by which the object significance can be elevated as compared with the initial rank determined according to the procedure. For instance, if the displacement limitation is taken to equal 20%, the 100th object can be advanced to the 80th rank, the 500th object, to the 400th one, and the 10th object by its significance, only to 8th rank and not higher.

After ranking the objects within the homogeneous groups (types) there occurs a sophisticated problem of developing a unified ranked list of the SCTES objects. To solve this problem, the authors have suggested the innovative algorithm for the merger of the heterogeneous ranked lists.

The first step for solution of this problem is the stage of measuring the quantitative and qualitative characteristics sufficient for consequent assessments. As the measurements can be performed only with certain accuracy, the list of indicators describing the state of the objects is always conveniently sufficient because it cannot embrace the whole possible multiplicity of internal and external circumstances of the objects' functioning. In this case, we can discuss only the measurements and assessments obtained with certain errors.

In fact, any assessment can be regarded as estimation of the dynamically varying values of the object indicator close to its equilibrium value. The equilibrium value has practically never been equal to the assessment value, but the assessment is always located in the field of the equilibrium point attraction.

The positions of objects O_1, O_2 and O_3 can be clarified on the significance scale only by way of comparing the selected objects with some other objects more than likely of another type—$\tilde{O}_1$, $\tilde{O}_2$ and $\tilde{O}_3$.

Three cases are possible:

- each of the objects of one type is recognized by the experts to be equal to the respective object of another type;

- all objects of the same type are recognized by the experts to be equal to the same object of another type;
- the first two objects of one type are recognized by the experts to be equal to the same object of another type, while the third object of the first type is equal to the less significant object of the second type.

Consideration of these cases has shown that indistinct assessment consisting in the use of indistinct equality as the comparison apparatus leads to the necessity of grouping (clusterization) of objects O_1, O_2 and O_3 of one type presented for assessment by way of their comparison with objects $\tilde{O}_1$, $\tilde{O}_2$ and $\tilde{O}_3$ of another type. In such a case, the assessments of the objects of the first type can be clarified through the assessments of the objects of the second type and alike the assessments of the objects of the second type can be clarified through the assessments of the objects of the first type.

Moreover, with such an approach nothing constrains the use of more than two compared scales. Sure enough, when changing over to more than two scales, some peculiarities associated, for instance, with "competition" between the second and third scales appear in the course of the process of forming the boundaries of the clusters in the scale of the first type objects, but the above idea served as the basis for elaboration of the algorithm for development of the integral estimation of systemic significance of the objects of different types.

Each type of the objects differing from the other ones is assessed in compliance with its own procedure. Naturally, different simply ordered lists of the objects measured with various errors and accordingly with different characteristic dimensions of the clusters are obtained.

The essence of the algorithm is well explained by the following analogy. An uphill road corresponds to each type of objects, i.e., each scale. The "mile stones" —assessments of their significance are placed near the objects located along this road and it be considered acceptable that the assessments are given with errors. These assessments increase incrementally and the locality sea level increases at the same time. Comparisons of the objects of different types satisfy the requirements of the same situation, when it is asserted that the i-th object of type t_1 is compared types can theoretically coincide, which in particular asserts that the respective "road" from the "smaller" object to the "larger" object is slightly sloping and runs over a flat terrain. In accordance with this analogy all objects participating in the comparisons, as well as those, which turn out to be nearly at the same sea level due to the aggregate of comparisons, are grouped into clusters. An arbitrary set of object of various types can exist in each cluster. The non-compared objects form, in their turn, the clusters consisting of a single object.

All conditions are satisfied by normalizing the arrangement of clusters and projecting the real data (own assessments of the objects included in the clusters) into the interval determined by the clusters' boundaries: the approximate parities of the compared objects are ensured by inclusion of these objects into one cluster, while the descending sorting of the objects' significance is ensured by the order of arrangement of clusters.

The general picture of the finite solution is given by the diagram represented in Fig. 6.

In the figure the ordinate axis contains the assessments of the general integral scale of significance of the objects of three types. Digital marks standing for the initial own assessments of the objects' significance are plotted on the individual graphs for the objects of different types. The clusters comprising the approximately equisignificant objects are shown by means of horizontal strips and Latin letters. The algorithm proper consists of two parts. The given data are prepared for calculations in the first part of the algorithm. In the second part of the algorithm, significance of the objects is assessed in all available types of the scales and the convolution of these assessments is finally formed to obtain the non-dimensional integral estimation of significance of the objects.

A detailed description of algorithm for different-type objects ranking problem is given in the paper [43].

The suggested methodology and adaptive algorithm for ranking the similar objects fuel and energy complex by their systemic significance takes into account the case of ambiguity of the given data on the resource and basic criteria or their

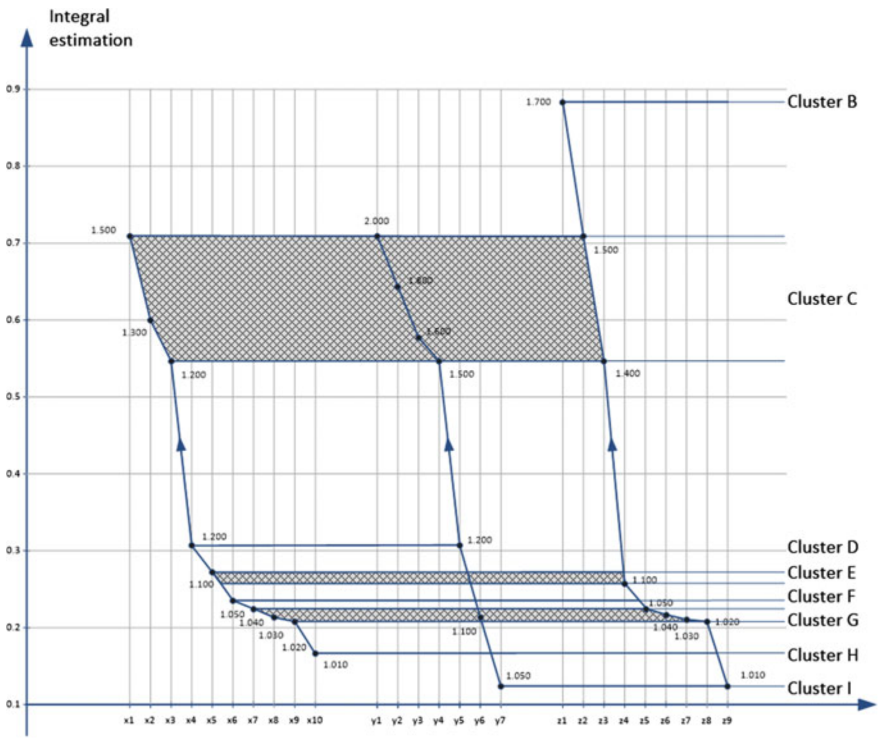


Fig. 6 Diagram showing solution of problem of developing integral estimation of systemic significance [43]

partial absence and substantiate a possibility of using the expert approach to the objects' ranking.

To merge the ranked lists of similar objects into a unified ranked list, use has been made of several novelties ensuring correctness of the procedure of comparing the objects of various types on the basis of partial expert findings of equisignificance of their separately selected representative pairs for solution of more general problems of enhancing protection of the fuel and energy complex objects.

The suggested approach can be used by the units in charge of security of the fuel and energy complex objects. A complex analysis of interrelated risks for separate industries and the fuel and energy complex wholes will make it possible to give substantiated recommendations for determination of required and sufficient security levels of hazardous production facilities proceeding from their significance for solution of a wide spectrum of management problems.

4 Risk Assessment

Risk assessment is the stage at which the adverse effects connected with this or that production activity is defined. And first of all it is necessary to identify danger sources for what it is necessary to define borders of the system under study. In other words, it is necessary to know what sources are needed to include in consideration and which aren't at risk assessment in the region or from the concrete system under study.

There are no hard rules here and they can't be. Sources and processes which are needed to consider and explore can be necessarily presented as one unified system or as a variety of independent subsystems. Degree of aggregation must depend on complexity of considered production, and also on degree of dependence of working processes.

However today there is a number of developed provisions which have to be considered during research of safety issues. So, for example, at risk assessment from energetic and other difficult technologies, by comparison with each other or when researches are conducted in regional or even nationwide, borders are defined so that to include all energetic or industrial system from fuel extraction or raw materials to final product production.

With reduction of studied geographical or technological scales the value of inclusion in consideration of full energetic or industrial cycles decreases. The most fully formulated provisions on delimitation of the studied regional or large industrial systems can be found in works [44, 45].

International organizations note the fact that at risk assessment even from one concrete technology in different countries in most cases receives various values. It is explained not only distinction in operating characteristics of fuel, its quality, in requirements for control of pollution level, but also not always adequate definition of the bodies of the system under study itself. Therefore, for simplification of collecting and data processing a unified set of terms and provisions for the description of energetic and industrial systems and their main components must be accepted. The formulation of basic concepts for such systems has been offered in work [45]. The main of them are:

- process is the smallest technological operation for which the main engineering characteristics and data on release and leak of dangerous substances are usually collected;
- action is the following level of aggregation (an action is one or several processes which answer one purpose, for example, coal transportation is an action since it can be executed by various processes: by rail, river, by car, etc.);
- trajectory is connected series of processes from production to use of the final product (the example of trajectory is stub oil production, its transportation to the coast, sublimation, transportation of the product to fuel station and fueling into the car);
- system is a set of corresponding trajectories (it can include all uses of this resource such as all coal trajectories from production to end use, or also can conclude all national energetic system, including all resources and final energy consumption).

The description within the offered definitions of supply of any energetic system will depend on the type of used fuel which has a number of related actions and within each action there can be one or several processes.

The description of use of the final product of any energetic system is subdivided at the certain level into concrete sectors (industrial, transport, municipal, etc.). Each sector consumes certain types of fuel, and each type of fuel is used by corresponding mechanism or device.

Use of such unified concepts and ways of the description of systems must help to facilitate use of information on systems for obtaining statistical data and their more adequate description.

To have an opportunity to define what danger this or that industrial system, production or an object constitutes it is necessary to have certain information on each of them. It is necessary to consider that since methods of determination of potential danger source parameters which must reflect the nature of their work and risk related to them are various depending on conditions of their functioning, then it is necessary to do distinction between normal (i.e. calculated, accident-free) work and emergencies which are usually characterized by low probability of implementation, but extensive possible damage.

Consideration of characteristics of the enterprise must be based on complete description of industrial complex, on inventory of all stored, processed or produced substances. For definition of potential danger, especially in the conditions of normal work of the enterprises it is necessary to have the full list of production wastes, what and in what quantities are emptied into water or air environment, what ways and methods of their processing, and main ways of burial.

It is also necessary to have information on types and intensity of transportations of dangerous substances to and from the enterprise under study. At the same time it is necessary to consider transportations on automobile, railway and water lines and their pipeline transportation.

Key aspects in risk assessment are detailed description of a source of danger and determination of the related possible damage.

There are various models of sources of danger which allow to define probability of this or that accident development and to determine the corresponding power of release of dangerous substances in the environment.

Depending on source type it is possible to select three categories of risk.

The first category is a usual risk connected with normal work of the enterprise. Into the conditions of normal work, we can also include accidents with insignificant damage which happen quite often. Thus, this category of risk is characterized by probability of realization equal or close to one. In most cases usual risk is either an integral part of the production, or easily controlled. As an example there are sulfur and ashes emissions from the combined heat and power plants using organic fuel, CO2emissions, etc. Sources of such risk are usually described by the power of emission or leak into the environment, connected with normal work or with some incident. The assessment of power of emission or leak for the working enterprises can be made on the basis of measurements or results of work experience of similar enterprises. For projected or when it is impossible to perform direct measurements, model calculations are carried out.

Other two categories of risk are connected with accidents on production, during transportation or storage of dangerous substances. By accident is meant an event with low probability of implementation (for example, less than one for all the time of existnece of the enterprise), but with considerable or even catastrophic consequences.

In the analysis of emergencies possible scenarios of development of accident are usually considered, i.e. the scenario constructed by the principle “what will be if…”. At the same time such factors as type of the initiating event, quantity of the available dangerous substance, efficiency of emergency security systems and many others have to be considered. Usually there is a large number of possible scenarios of development of accident and therefore in risk assessment it is necessary to define all range of possible scenarios and their probability. In which case sizes of probability can change from 10^{-6} to 10^{-8} events a year. It is so difficult to estimate rarer events that it is considered that they are almost improbable.

By consideration of accidents they can be divided into two classes and be characterized by two categories of risk respectively.

Periodic risk is connected with those accidents which quite often repeat, but cause limited damage even with loss of human life. It doesn’t mean that such accidents are planned. Of course they are undesirable, and for their prevention security systems are created and used. However, despite these measures according to statistics such accidents can happen and the related risk has quite wide range of values depending on type of production activity. The reason of such accidents is usually violations of technological process, incorrect use of the equipment and a human error.

For risk assessment of this category the frequency of accidents and other necessary parameters are evaluated by means of standard statistical techniques on the basis of the available data.

Hypothetical risk is connected with accidents which as it is considered, can happen to very small probability but lead to very big consequences. For such class of accidents absence or insufficient number of statistical data is characteristic. However, because of their huge potential damage it is impossible just to wait until sufficient practical experience is gained. Therefore, in these cases the analysis of hypothetical accidents is made for the purpose of determination of probability of realization of this accident and assessment of its possible consequences.

Usually the lack of statistical data belongs to behavior of a large industrial or energetic system in general. Therefore, such analysis is carried out either with the help of an expert assessment, or with method of "event trees" where the probability of hypothetical accident can be predicted on the basis of possible malfunctions or failures of separate knots or mechanisms which have relevant statistical data.

In many cases such method of predictions makes one of the main parts of the probabilistic risk analysis—PRA.

Usually it is almost impossible to take direct measurements which would show impact on the person by all possible ways. Therefore, for risk assessment it is necessary to use models of transfer of dangerous substances in the environment. Generally, models of transfer of dangerous substances in the environment are the simplified mathematical representations of the valid processes of dispersion.

Model type which can be used will depend on the required information, on characteristics of the considered dangerous substance and on the way using which it gets to the environment. The essential moment at the same time is the question of whether the time-dependent result is required to us, or it is possible to assume that process is quasistationary since in this case various models can be used [46–52].

When modeling distribution of chemical elements, it is often necessary to consider chemical reactions happening during accident or transfer in the environment. So, for example, when calculating risk from gas pipelines with high content of sulfur in gas mix for the case of nonflammable emission of gas in case of a rupture of the gas pipeline a toxic component, necessary for consideration, is H_2S. If gas flames up, then H_2S will be transformed to SO_2 which in this case is considered as toxic component.

However, it is necessary to remember that for risk assessment there is no need to use excessively complicated transfer models because of big uncertainty and averaging arising during calculating of risk sizes. By the way, finding the size of uncertainty and range of possible risk values is one more compound characteristic of risk in general.

So, according to various experts' points of view uncertainty in risk assessment from accidents at industrial enterprises can make one and even reach two orders of value. It is connected with a lack of knowledge base on a wide range of technical, ecological and social factors which it is necessary to consider in risk analysis.

Therefore, there is no need for strong complication of transfer models. There are even also conclusions based on the analysis of accuracy and uncertain when determining risk that transfer models allowing to receive value of concentration of dangerous substance in the explored place with an accuracy of 10 or maximum 20% are quite accepted [53].

In order to connect result of transfer models (usually it is concentration of dangerous substances in air, water, on the ground, etc.) with possible damage influence models are used.

In this situation three main groups of consequences which, respectively, will define three main risk classes can conditionally be understood as damage:

- the risk connected with transition to an irrevocable state (death, destruction of animal or plant species, total loss of working capacity, etc.),
- the risk connected with transition in steady, worse state in comparison with former (chronic disease, material losses, deterioration of life), and
- the risk connected with transition to the worst, but returnable state (temporary disability, restored material resources, environmental pollution, etc.).

There are developed methods and models of calculation of doses radioactive and chemical materials on various ways of influence which use the results received from transfer models. An ultimate goal of these calculations usually is definition of either collective, or individual dose (for radioactive materials the effective equivalent dose is defined since this value is proportional to radiation risk).

At assessment of injury for health of the person caused by a dose of this or that substance there is a series of problems. If for influence in high doses it is often possible to connect a damage to the reasons caused it, then with influences of small doses the situation is much more difficult. Chronic influences of small doses of some substances can cause symptoms of certain diseases while effects of other substances don't cause the diagnosed symptoms of illness, such as cancer and which can also be caused also by other reasons.

Direct researches of dependence a dose effect (or a dose damage) are complicated by many reasons. So, for example, in many cases other factors which aren't connected with existence of considered influence can also effect and therefore separation of its "pure" influence can be very difficult or just impossible. Development of disease can have long latent period equal to years or decades that also bring difficulties in diagnostics.

One of ways of finding of dependence a dose effect consists in that, using mathematical models to extrapolate data for high doses to the area of small doses at which directly it is impossible or it is very difficult to observe effects. The ratio influence-effect can be at the same time linear or some other function.

Thus steady functioning and development of any structural and difficult system depends on influence of a large number of external and internal factors, including factors of negative impact. For monitoring and assessment of these factors and decision-making directed to decrease in negative consequences of their appearing the so-called systems of balanced indicators (Balanced Scorecard), key performance indicators1 (KPI) (which are quantitatively characterizing factors of risks to which the system is subject) from among which strategic target indicators (STI) which quantitatively reflect strategic purposes of system functioning are chosen and represent the basic economic and operational performance characterizing efficiency of its development (indirectly, they're not achievement characterizes the level of the existing threats and degree of their realization in the considered period).

On the basis of these indicators the monitoring corporate systems of threats and risks are formed and they allow collecting data on changes and to carry out the analysis of efficiency of system functioning according to several hundred indicators in organizational, food, geographical and other sections in daily, quarter and annual planning horizon. Results of the analysis allow to exercise management by exception, focusing attention on problem areas of each management object by means of "go-no-go" indication. However, in process of accumulation of data there is a problem of interpretation of signals of these hundreds of " go-no-go indicators". It isn't obvious what to consider a "good" or "bad" signal in general for system if for example a half of indicators "lights" in green color, and a half with "red"? How to qualify a situation if it is a little more "green" indicators, than "red", etc. Also is unevident connection of analyzed indicators with indicators of the high level (STI) and degree of their influence on achievement of the STI target values approved by the management of the company. There is a so-called effect of "big data" when analysts don't manage to process the collecting information, and standard statistical methods just cease to work.

Besides threats and risks monitoring system formed on the basis of the analysis of trends of change of indicators isn't capable to predict crises and situations with negative dynamics. Such events are rare and proceed as a rule at various expected background, and in case of the analysis of historical dataset of rare events discrete dynamic probabilistic processes take place. Therefore the purpose of the analysis of structural and difficult system as object of forecasting is creation of such predictive model of situations dynamics arising at its functioning which will allow to reduce degree of events dates uncertainty and their scale by means of computing experiments and selection of acceptable parameters, that is to obtain expected information on object of forecasting due to detection of hidden regularities which indicate either changes of object condition, or regularities of changes of external environment parameters significantly influencing its functioning (so-called laws of variability of "forecast background").

Because of the discrete nature of crisis situations using of data analysis device based on classical laws of large numbers, is incorrectly. Actually convergence on probability in reality practically is never observed, except for the statistics which is saved up in systems of mass service (control of assembly operations on conveyors, casino, insurance of small incidents (road accidents, home and employment injuries, medicine, etc.). Compliance of reality and theory in these fields of activity is reached at the expense of very large number of independent realization. Thus indicators panel realized in the form of "traffic light" constructed on the basis of dispersion use as the main indicator can indicate during the whole year the normal state when actually the system passes into the area of pre-crisis values. Besides at officially declared hierarchical system of indicators as a rule there is no unambiguous functional connection and mutual influence of indicators of lower and upper level.

It is necessary to have correct primary analysis of long-term statistics, and on the basis of this analysis it is possible to draw the conclusion if it is possible to develop of adequate to the studied problem forecasting instrument and what share of accident of dates of arising of adverse situations and their scales can be eliminated with its help. It is also obvious that as true laws of distribution of the analyzed casual processes and mainly factors defining them will be continuously corrected (any hi-tech system, changes quicker than adequate statistics collect), it is necessary to use criteria "free from distributions". In particular, for example, as criteria of achievement of the predictive purpose it is necessary to take not sizes of rejections of model and real data, but the criteria used in methods of classification and recognition of images. For example, as measurement of accuracy of the forecast it is possible to use sizes of mistakes of prediction of the first and second kinds for various classes and types of situations, and if it is possible depending on classes of a physical object and depending on value of parameters of forecast background. The second circumstance is very important as, for example, it is incorrect to put statistics of accident rate of various seasons as during various seasons technological processes proceed differently.

Reliable performance by system of its functions is characterized by saving some set characteristics (reflected in the corresponding STI and KPI values) in set limits. In practice it is impossible completely to avoid deviations, however it is necessary to seek for minimization of deviations of current state from some set ideal—the purpose set for example in the form of STI values of the first level.

The measure of threat of not achievement of STI set values of the first level, is considered in this case as the variable representing function of rather current system statement: it increases in case of approximation of the assessed situation to some admissible boundary after achievement of which the system can't fulfill the obligations and reach appropriate set STI target values of the first level.

Mathematical statement. Let the set of signs of the current situation of X (for example, the current KPI values), the set of admissible realization of situations of Y give (for example, the current STI value of the first level is more (or less) than previous, etc.), and there is criterion function $y^*\colon X \to Y$ which values $y_i = y^*(x_i)$ are known only on a final subset of objects $\{x_1, \ldots, x_l\} \subset X$ (for example, the KPE values corresponding to the current STI value of the first level). Let us call pairs "object–answer" (x_i, y_i) precedents. Set of pairs $X_l = (x_i, y_i)_{i=1}^{l} = 1$ will make training selection.

It is required to restore on selection X_1 dependence y^*, that is to construct decision function $A\colon X \to Y$ which would bring closer criterion function $y^*(x)$, and not only on objects of training selection, but also on all set X. As at the same time decision function A has to allow effective computer realization let us call it also an algorithm.

In such statement KPI (in fact–risks factors) act as signs of the current situation characterized by STI value of the first level (risks indicators). Signs can be binary (1/0, red/green), nominal (a set of values), serial (a set of ordered values) or quantitative type. In case the current STI value of the first level in comparison with the previous value has improved, the current situation is assessed positively (for example, it is coded as "1"), otherwise—negatively (for example, it is coded as "0"). In the presence of three "measurements" of the situation (current, previous and pre-previous), it is possible more difficult (more exact assessment of current state through dynamic characteristics of speed and acceleration of change of STI value of the first level for this period of observations).

Let us consider work of an algorithm of classification of observed situations according to degree of their danger for the considered system. In general, the tasks connected with need of classification of situations arise rather often at diagnosis of diseases on the basis of a set of certain, characteristic factors, at technical diagnostics of cars and mechanisms, at assessment of financial stability of the contractors participating in competitions on performance of these or those services, borrowers of banks at an assessment of quality of management or at the analysis of risks factors and accidents. As a rule, the description of such situations is made out by means of questionnaires. Each subject of classification is represented as a vector (point) in p-dimensional space (the ordered set of p numbers). Though decisions on classification of the analyzed situations are made not only and not just on the basis of answers to questions of questionnaires (signs), however assessed questionnaires contain the hidden regularities allowing, in the presence of rather representative statistics, "to repeat" earlier made decision for new data. The method of basic vectors considered as a special case of regularization according to Tikhonov allows to construct the border dividing a set of similar objects into isolated classes. The main idea consists in construction on the basis of the analysis of signs, internal connections between unevident sets of "correct" parameters of the description and classifying signs of ensemble of trees of logical formulas. It allows to classify a situation unambiguously or to exclude it from further consideration.

Nominally it is possible to mark two classes of objects which experts have to face in the field of automation of management: "simple" and "complicated".

"Simple" ones are objects exact mathematical models of which, for example, in the form of algebraic equations system or linear programming model, at taking into the account of all necessary quantitative factors influencing behavior of an object are suitable for realization on the computer of the chosen class are and are quite adequate to an object.

"Complicated" objects of management have the following main distinctive features [54]: not all purposes of the choice of operating decisions and conditions influencing this choice can be expressed as quantitative ratios; formalized description of an object of management is absent or is unacceptably difficult; the considerable part of information necessary for mathematical description of an object exists in the form of ideas and wishes of specialist's experts having experience with this object. Creation of exact mathematical models of "complicated" objects suitable for realization and operation on modern computers is either difficult or often in general is impossible.

But it doesn't mean that the task has no decision. Generally, there can be two possible directions of search. The first is to try to apply a nontraditional mathematical apparatus for creation of the model considering all features of an object and suitable for realization. The second is to build not the model of an object, but model of object management (i.e., not an object, but the person operator in process of object management is modelled).

The main tool of similar developments is creation of software products within the uniform procedures which are carried out by the person—the Automated Control Systems (ACS). And the question is not about automatic systems, not about robotic systems making decisions independently, namely the question is about automated systems in which the main, defining role is assigned to the person. To emphasize this distinction, during further narration, we will call similar systems automated advising systems (AAS).

Now there are three basic platforms of development of AAS.

The first is the classical numerical mathematics focused on von Neumann structures of computing systems [55] based on the concept of algorithm which it is the final sequence of instructions implementation of which leads to the desirable (or acceptable) result. In case of small and non-uniform selections often there is no opportunity for a long time to create steady expert rules in the form of instructions and, therefore, in the studied case, use of artificial intelligence systems (AI) it is a situation when costs for creation of system won't be paid off with efficiency of its work. A way of creation of von Neumann systems is a way of difficult coordination with experts of algorithms of use of "general rules" which in the course of coordination become known to many and the control system are "unworthy", that is development of bad AAS a priori takes place.

The second is neuromathematics [56] where the desirable acceptable result is achieved due to training of some computing environment using precedents from the past. The main difference from the classical scheme are parallel connected nonlinear calculations. Owing to a trifle of selections and heterogeneity of initial massives of given neural systems (NS), as a rule, draw negative and positive conclusions for a set of additional "false images" [57]. Specific nature of neural network approach is that there are so many entrance signals (features of description of situation) that already at the first level of connected calculations when selection of communications is made for formation of necessary combinations of features, obvious ambiguity of the choice takes place. In these cases, experts accompanying process of training of NS have to nullify unclear for what reasons the majority of communications of neurons in the first layer of NS. Otherwise the weights of features will be averaged. There is no semantic "mixing" of entrance signals in packages of neuromathematics. Algorithms of training of NS are effective when entrance information is subject to preprocessing and is uniform, so neural network approach, in its classical understanding isn't suitable for the solution of considered tasks.

The third is parallel processors and implementation of a method of support vectors [58] in them in which qualities of speed of operation of trained NS are combined with severity of instructions of classical algorithmization. Preliminary formation of start statuses in training activity (additional training) of NS actually leads the neural network approach to support vectors offered by a great number of authors. The most known among them is the academician of the USSR Tikhonov who in the 40s of the last century used this method for the decision of incorrect tasks of zoning of oil fields according to the signals of reflected waves received in case of electrophysical and seismological investigation [59]. Also, as well as the NS which in case of correct approach must construct the same separating broken surfaces if it optimizes distances, but not amount of neurons, a method of regularization of the task of division of sets got the name "method of support vectors" uses information from precedents and makes the decision, based only on them. But in method of support vectors (with coordinates of the middle of the segments separating the closest neighbors) it is rigidly set the regulating principle—the closest neighbor. It makes related a method of support vectors with methods of cluster analysis [60, 61].

The idea of a method of support vectors is also close to methods of local approximation [62]: not the "universal world identical to all", but the world as "the world of the countries in which there are own laws, and these laws smoothly flow from one country into another, having, of course, some general part thanks to which such overflowing is possible" is under construction.

Essential difference method of support vectors from NS is that NS studying unreasonable "will displace the decision" to the area of rejected objects only because it is taught according to regression statistical signs—according to the method of the smallest squares. Considering that (for example, at the auction) only one wins, and several lose rejected objects must be more so and total contribution to the total amount of mistakes is more. The method of support vectors is more reliable: the place of support vector responsible for "its site" won't change until it won't be repealed (again locally) by data on arrival of a new closest to support vector neighbor. It won't be required to recalculate the decision anew as the influence of new data is "local" and won't influence other constructed rules.

General conclusion: the most preferable is the solution of considered tasks with "locally optimum methods", use of the principles of regularization of decisions underlain in the method of support vectors with adaptation under flows of new data.

The idea of creation of piecewise linear decisive rule of a general view is that the hyperplane is selected allowing to separate optimum in sense of some criterion (for example, maximum) a share of objects of any one from classes presented in the training selection then for the rest of objects they look for a new similar plane and so till full division of classes. Locally optimum planes used at each step of division of classes and creation of decisive rule of this kind in the procedure adopted by us are considered as orthogonal subspaces at first to the initial individually considered features, and then to new features synthesized from these initial ones.

Synthesis of the feature defining each following element of the decisive rule begins with the choice of "germ"—one of questions in a projection to which in the sense of the chosen criterion it is possible to receive the best almost "pure" separation of objects of one class from mix of remaining objects of this and other classes. After "germ" is found, an attempt to improve it is made, forming of it a new feature by a serial combination with each of all other features, also including earlier synthesized. And such combination can be made differently. In the elementary case trial improvements of a germ are carried out by summing up or subtraction. The best of pair combinations can be considered as "germ" for the following cycle of synthesis, etc., till synthesis stop criterion won't be satisfied. Then on the received feature it will be possible to design objects which aren't divided to this step of creation of the decisive rule, to separate a new part of objects of any class and if division of objects of different classes in selection hasn't come to the end yet, to pass to synthesis of the following feature and definition of the next element of the decisive rule.

For clarity let us consider the work of offered algorithm on a numerical example. Set of entering data is training selection—the so-called *table of features* (Fig. 1) which includes: *keys* identifying an object (or a situation); numbers of *features* of an object; *assessments of features* (1/0 – "successful"/"unsuccessful"); *indicator* " ±" defining the class of an object (is accepted/is rejected). In the example (Fig. 7) discussed further the following basic data are accepted: quantity of *objects*: 97; quantity of *features*: 125; quantity of *accepted* objects: 59; quantity of *rejected* objects: 38; *serial numbers of features* which participate in all subsequent operations are assigned to keys.

The main stages of algorithm and their summary are presented in Table 2. The algorithm has two options of realization: classical (index K) and simplified (index У). The part of the algorithm realizing the main purpose consists of points ^{К,У}[0.1] - ^{К,У}[0.7]. In processing of bivalent information in connection with development of the theory of information as the tool describing transfer of strings of signals consisting of zero and one Hemming's metrics is accepted:

$$\rho\left(O_i, O_j\right) = \sum_{k=1}^{K} \left[O_{ik}\left(1 - O_{jk}\right) + \left(1 - O_{ik}\right)O_{jk}\right] \tag{4.1}$$

The size of this metrics is distance between one-dimensional same objects (lines, columns, screens, size of symbols) is measured by the number of couples, incoincident among them. When performing a condition "excluding OR" value of a metrics increases by 1.

Ключи	+/-	1	2	3	4	5	6	7	8	9	10	11	12	114	115	116	117	118	119	120	121	122	123	124	125
Num1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num3	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num4	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num5	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num6	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num7	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num8	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num90	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Num91	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Num92	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Num93	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Num94	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num95	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num96	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Num97	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fig. 7 Table of features (training selection, basic data)

Table 2 The content of stages of algorithm

Index	The content of stage
$^{К,У}[0.1]$	Calculation of matrix “semi-hemming”
$^{К,У}[0.2]$	Elimination of logically contradictory basic data
$^{К,У}[0.3]$	Extraction of copies and “not working” samples of rejected objects
$^{К,У}[0.4]$	Extraction of copies of samples of accepted objects
$^{К}[0.5]$	Extraction of copies and “not working” samples of accepted objects
$^{К}[0.6]$	Calculation of ranks and weights of stiffness of rejected objects
$^{У}[0.6]$	Calculation of ranks and weights of stiffness of rejected objects
$^{К}[0.7]$	Calculation of ranks and weights of accepted objects in classical set
$^{У}[0.7]$	Calculation of ranks and weights of accepted objects in expanded set
$^{К}[0.8]$	Rule building. Classical optimized option
$^{У}[0.8]$	Rule building. Simplified option
$^{У}[0.9]$	Dropping of features copies and also features overly doubling the result of assessment choice

Discrepancy is interpreted as a mistake, and the proximity of objects, thereby, is assessed by the minimum quantity of “corrections” which it is necessary to bring either in one, or in another, or in both objects in order objects to become identical, indiscernible. Naturally equality is carried out:

$$\rho\left(O_i, O_j\right) = \rho\left(O_j, O_i\right) \tag{4.2}$$

As in this algorithm operations are performed over a natural class of monotonous functions, not all discrepancies in couple and only “ordered” are of interest (further —*semi-hemming measure*)

$$\mu\left(O_i, O_j\right) = \sum_{k=1}^{K} O_{ik}\left(1 - O_{jk}\right), \tag{4.3}$$

reflecting only quantity of “successful” features μ the first object of which has value equal 1, and the second –0.

On combinations of “successful” features accepted objects from rejected ones will “be separated”, and on a metrics μ the size of “boundary layer” will be measured. *The first step* of the algorithm is filling of a matrix of Semi-Hemming for all couples of objects. The result of such processing is a square matrix with zero diagonal elements (Fig. 8) whose lines and columns with smaller numbers have accepted objects.

The matrix is divided into 4 sectors according to participants of relations. So, for example, the left top sector represents the relation of accepted objects among themselves, the right lower—rejected among themselves. Two other sectors show the relation of accepted objects to rejected and vice versa. The right top sector of matrix is required. The ranges of cells of the table of Semi-Hemming having zero

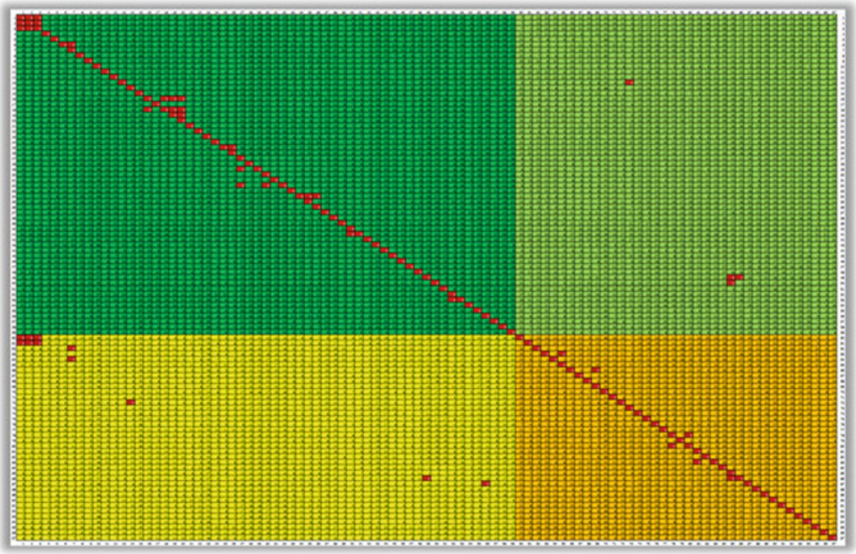


Fig. 8 Matrix of semi-hemming

values are coloured in Fig. 8 in more intense color that allows to observe the zero diagonal of matrix distinctly.

When the initial matrix of Semi-Hemming is constructed, it is possible to start main calculations. The first that Semi-Hemming's measure allows is to *eliminate contradictory couples of objects* which obviously don't meet conditions of strict inequalities setting a partial order. Technically it is necessary to overassess objects if their contents obviously indicate that in some compared couple of objects the object which is certainly the worst by data is assessed better than the opponent.

There can naturally be couples of objects $\langle O_j, O_i \rangle, (i < j)$ such that all units of the accepted object O_i contain in set of units of features of rejected object O_j. In this case we have the obvious contradiction which is reflected in Semi-Hemming matrix in such manner that the following equality is carried out:

$$\left(\mu(O_j, O_i) \geq 0\right) \wedge \left(\mu(O_i, O_j) = 0\right). \tag{4.4}$$

In the concerned example these are *objects 13, 49 and 50*. The contradictory (zero) relations are marked with color (Fig. 9).

To decide that the object O_j best by data has mistakenly been carried to rejected, or on the contrary, the object O_i worst by data has mistakenly been carried to accepted, i.e. to make reassessment of experimental material only the person forming the decision—PFD can (even not the expert). But assessment objects, as a rule, not the last stage of inspection, in all similar cases (there are 3 of them in reviewed example) the worst object in couple moves to the category of rejected

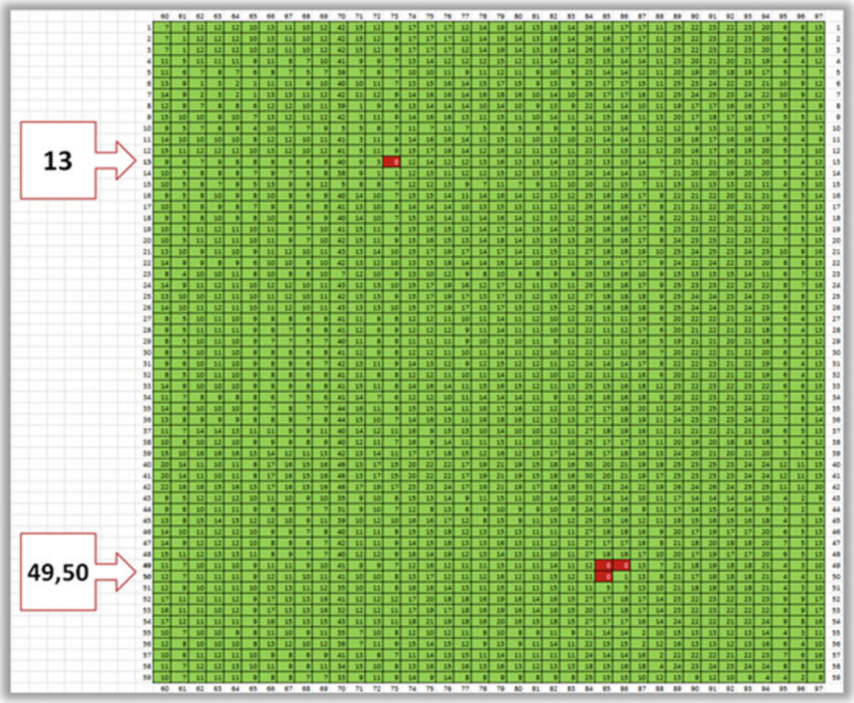


Fig. 9 Base purposeful matrix of semi-hemming

(Fig. 10). As a result, in Semi-Hemming matrix the quantity of lines and columns of accepted objects decreases, and the quantity of lines and columns of rejected objects increases by 3.

But in sum the number of objects hasn’t changed. By move of lines and columns of features of the overassessed objects in Semi-Hemming matrix of dimension 97 × 97 the submatrix of dimension 56 × 41 (Fig. 11) which is settling down in the right top corner of the matrix is formed, and it doesn’t contain zero any more. It means that there are all reasons to confirm that all strict inequalities explaining such choice of objects will be satisfied. In a full matrix 97 × 97 zero remain, but they aren’t critical.

So among accepted couples $\langle O_{i1}, O_{i2}\rangle, (i1 < i2)$ and ejected couples $\langle O_{j1}, O_{j2}\rangle, (j1 < j2)$ of objects there are objects which are coincident, there are also dominations (enclosure of features of one objects in data of other objects).

At the following stage of the algorithm there is a removal of copies and “idle examples” of objects in the corresponding submatrices. Actions over both sectors are made in three steps: removal of copies of objects, their Pareto sorting and streamlining of the relations in the submatrix with cyclic removal of minimum

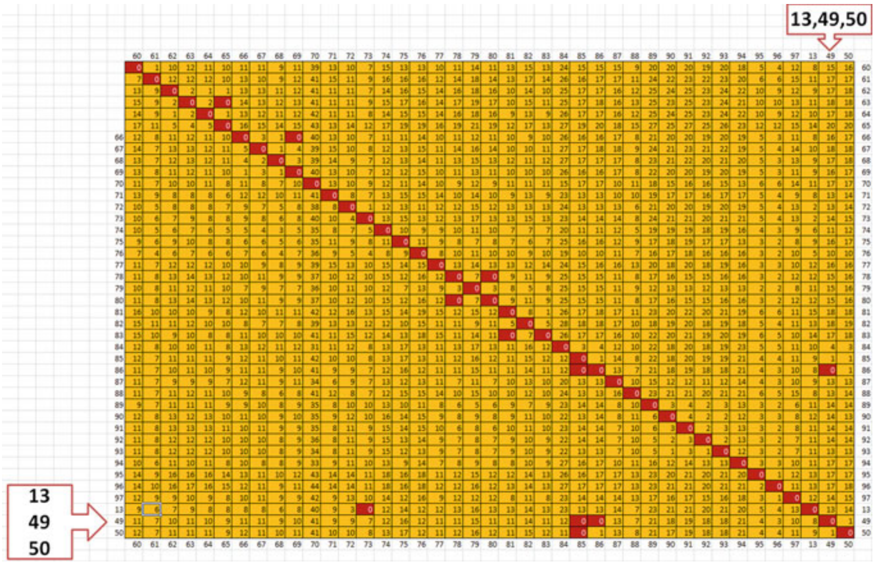


Fig. 10 Submatrix of rejected objects after removal of conflicting data

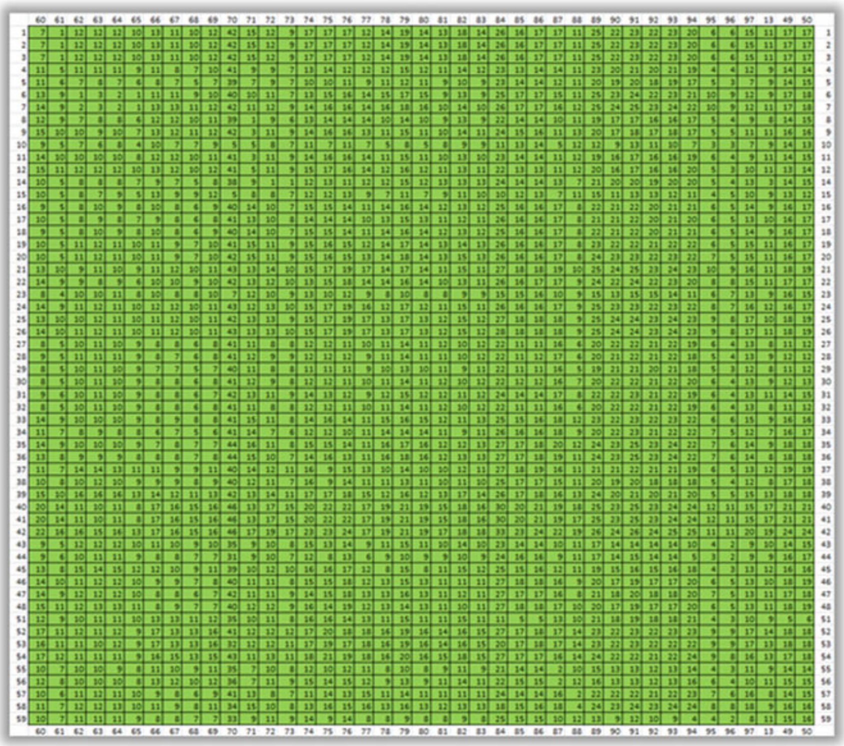


Fig. 11 Purposeful matrix after removal of conflicting data

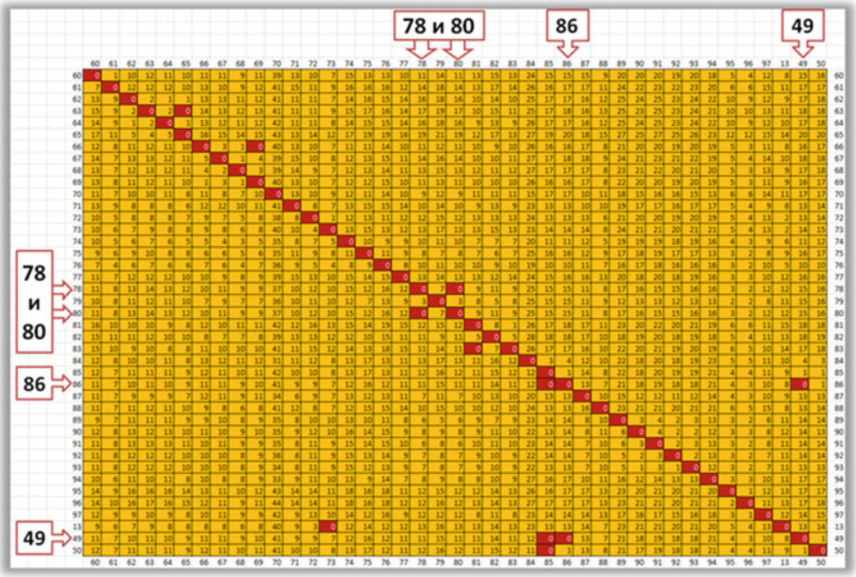


Fig. 12 Submatrix of semi-hemming before removal of copies of objects

(maximum) according to Pareto objects. Algorithms of removal of copies and “idle” examples for accepted and rejected objects are similar.

For removal of copies of objects, it is necessary to define symmetric off-diagonal couples of zero. In the reviewed example the are couples of objects under numbers 78 and 80, and also 86 and 49 (Fig. 12).

The corresponding objects (80 and 49) move (Fig. 13) to the end of the submatrix (lines and columns of the submatrix move to its end).

The quantity of rejected objects becomes equal 39. For the simplified option the quantity of accepted objects returned by this function is a border of purposeful submatrix which is used on the step У[0.6] of the algorithm.

At Pareto sorting of objects all “crosses” with the centers in diagonal elements which don’t contain zero values move to the left top corner of the submatrix (Fig. 14). The quantity of rejected objects accepts value 28. Streamlining of the relations between rejected (accepted) objects in the submatrix is carried out after shift of lines and columns so that all off-diagonal zero become higher (lower) than zero diagonal. As there are no cycles there has to be a triangular form. Further (until the triangle still contains zero), the rule is carried out: if in a line/column of any element there are zero, and there aren’t zero in a column (line) — it is an object not of maximum (minimum) layer according to Pareto. An object is displaced to the end (beginning) of the submatrix.

The lines which have remained after all cycles of an exception join number of maximum (minimum) objects according to Pareto (Fig. 15).

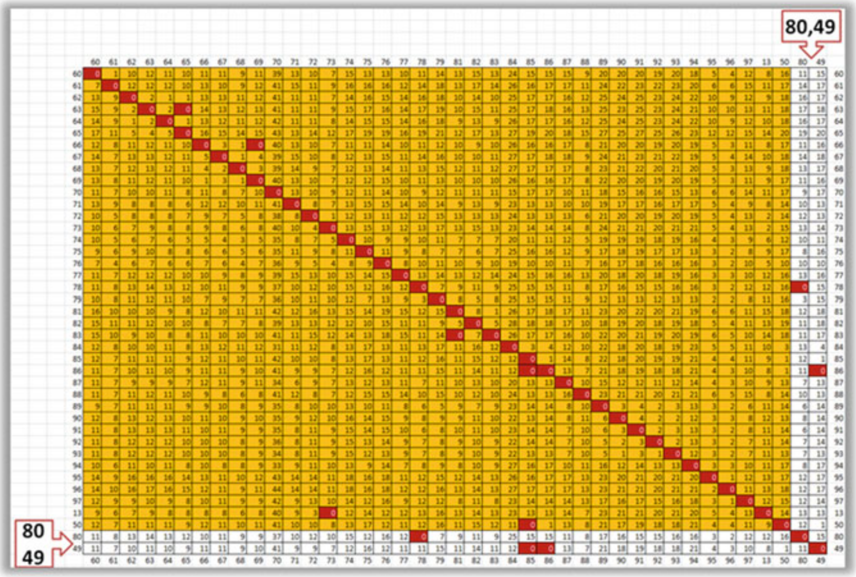


Fig. 13 Matrix of semi-hemming before Pareto sorting objects

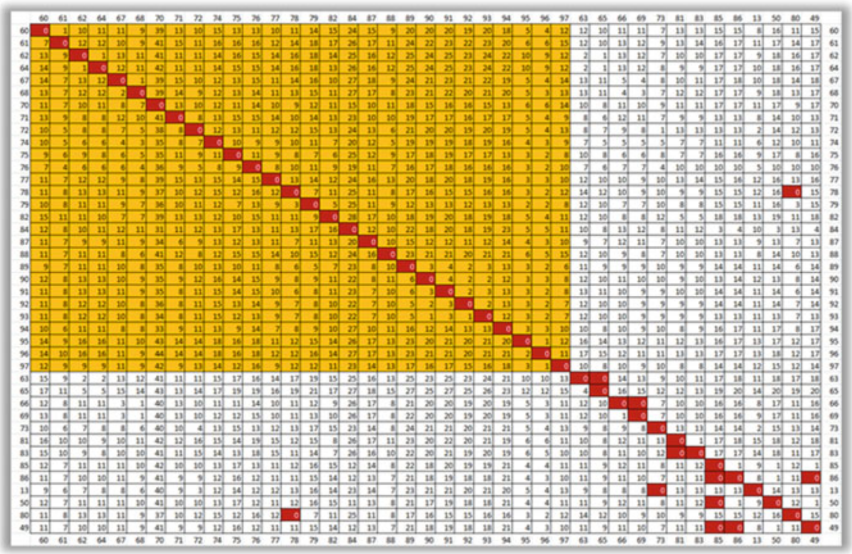


Fig. 14 Submatrix of semi-hemming before removal of Pareto minimum objects

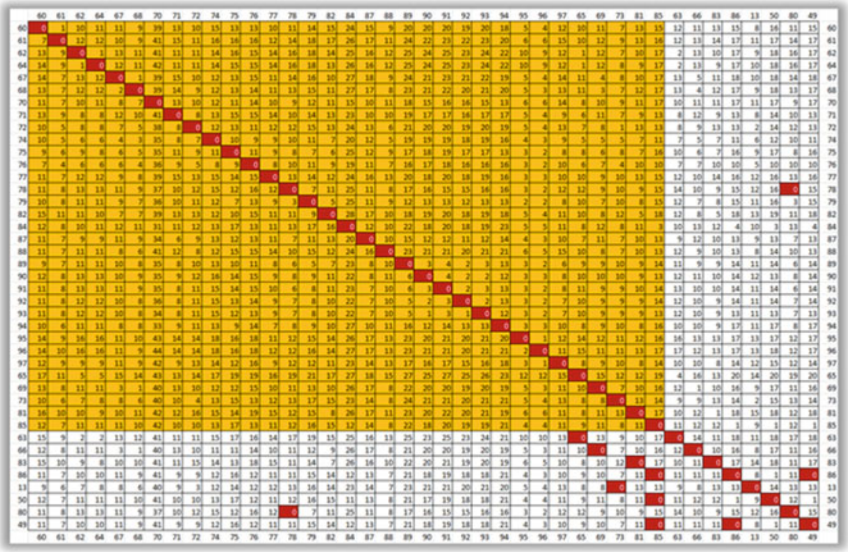


Fig. 15 Submatrix of semi-hemming after removal Pareto minimum objects

The algorithm returns total quantity of rejected or accepted objects which are borders of a required purposeful submatrix. The example of such submatrix for classical option is given in Fig. 16. The quantity of *rejected objects* for both options of reviewed example is identical and equal to 33.

It is caused by the fact that in both cases stages $^{\text{К,У}}[0.3]$ and $^{\text{К,У}}[0.4]$ are carried out. However the stage $^{\text{К}}[0.5]$ is intended only for classical option in this connection the quantity of *accepted objects* for it will decrease to 43 at 53 (for simplified). Calculation of ranks and weights of objects is made in three stages: selection of distances for three next objects, calculation of ranks of stiffness and, at last, calculation of weights. For every (accepted and rejected) object of purposeful submatrix (Figs. 14 and 15) defined at the previous stages three smaller values all of their cells are calculated.

They are used directly for calculation of ranks of stiffness: the objects having the smallest distances in relation to others will differ in the greatest stiffness.

After performance of the above-stated operations there is an opportunity to order lines and columns according to their ranks.

Weight are calculated by formulas:

$$\left(C_1^{\text{bad}}\right) \times \left(C_2^{\text{bad}}\right)^{RB[j]-1}, \tag{4.5}$$

$$\left(C_1^{\text{good}}\right) \times \left(C_2^{\text{good}}\right)^{RG[i]-1}, \tag{4.6}$$

	60	61	62	64	67	68	70	71	72	74	75	76	77	78	79	82	84	87	88	89	90	91	92	93	94	95	96	97	65	69	73	81	85	
1	7	1	12	12	11	10	42	15	12	17	17	17	12	14	19	18	26	17	11	25	22	23	22	23	20	6	6	15	10	12	9	13	16	1
4	11	5	11	11	8	7	41	9	9	13	14	12	12	12	15	14	23	14	11	23	20	21	20	21	19	4	4	12	9	10	7	11	13	4
5	11	6	7	7	7	5	39	7	9	10	10	11	9	11	12	10	23	12	11	20	19	20	18	19	17	5	3	7	6	7	7	9	14	5
8	12	9	7	8	12	10	39	1	9	13	14	14	14	10	14	13	22	10	11	19	17	17	16	16	17	5	4	9	6	11	6	9	14	8
9	15	10	10	10	12	11	42	3	11	14	16	16	13	11	15	14	24	11	13	20	17	18	17	18	17	5	5	11	7	12	9	10	15	9
10	9	5	7	8	7	7	5	5	8	11	7	11	7	5	8	9	11	5	12	12	9	13	11	10	7	3	3	7	4	9	7	8	13	10
11	14	10	10	10	12	10	41	3	11	14	16	16	14	11	15	13	23	11	12	19	16	17	16	16	19	6	4	9	8	11	9	10	14	11
12	15	11	12	12	12	10	41	5	11	15	17	16	14	12	16	13	22	11	12	20	16	17	16	16	20	5	3	10	10	12	9	11	13	12
14	10	5	8	8	7	5	38	9	1	12	13	11	12	12	15	13	24	13	7	21	20	20	19	20	20	5	4	13	7	8	1	13	14	14
15	10	5	8	9	9	9	5	8	8	12	12	13	9	7	11	11	10	7	11	15	11	13	13	12	11	4	5	10	5	12	7	9	12	15
17	10	5	8	8	8	6	41	13	10	14	14	14	10	13	15	12	26	17	8	21	21	22	20	21	20	6	5	13	7	8	8	11	16	17
21	13	10	9	10	12	10	43	13	14	15	17	19	17	14	17	15	27	19	10	25	24	25	23	24	23	10	9	16	9	11	10	11	18	21
22	14	9	9	9	10	9	42	13	12	13	15	18	14	14	16	13	26	17	9	24	22	24	22	23	20	8	8	15	6	10	10	10	16	22
23	8	4	10	11	8	8	7	12	10	13	10	12	9	8	10	9	15	10	9	15	13	15	15	14	11	6	7	13	8	10	9	8	15	23
24	14	9	11	11	12	10	43	12	13	15	17	19	16	12	17	15	26	17	9	25	23	23	22	23	22	8	7	16	10	11	10	11	16	24
28	9	5	11	11	7	6	41	12	9	12	12	12	9	11	14	10	22	17	6	20	21	22	21	22	18	5	4	13	9	8	9	11	11	28
31	9	6	10	10	8	6	42	13	11	14	13	12	9	12	15	11	24	17	8	22	22	23	21	22	19	6	4	13	9	7	9	12	14	31
33	14	9	10	10	9	8	41	15	11	14	16	14	11	15	16	11	25	18	12	23	22	23	22	23	22	6	6	15	9	8	8	12	15	33
37	11	7	14	13	9	9	40	14	12	16	9	15	13	10	14	12	27	16	11	21	21	22	21	21	19	6	5	13	11	11	11	10	18	37
38	10	8	10	10	9	8	40	12	11	16	9	14	11	11	13	11	25	15	11	20	19	20	18	18	18	5	4	12	9	9	7	10	17	38
39	15	10	16	16	12	11	42	13	14	17	17	18	15	12	16	17	26	16	13	24	20	21	20	21	20	5	5	15	13	13	11	13	17	39
42	22	16	16	16	16	15	46	17	19	23	23	24	17	19	21	18	33	22	19	26	24	26	24	25	25	11	11	20	13	16	17	17	23	42
43	9	5	12	12	10	9	35	9	10	15	13	14	9	11	15	14	23	10	11	17	14	14	14	14	10	4	2	9	10	10	8	10	14	43
44	9	6	10	11	8	7	31	9	10	12	8	13	6	9	10	10	24	9	11	17	14	15	14	14	5	3	2	9	9	7	7	9	16	44
45	13	8	15	15	10	9	39	10	12	16	16	17	12	8	15	15	25	12	11	19	15	16	15	16	18	4	3	13	12	11	10	11	15	45
46	14	10	11	12	9	7	40	11	11	15	15	18	12	13	15	11	27	16	9	20	17	19	17	17	20	6	5	13	10	8	8	11	18	46
47	14	9	12	12	8	6	42	11	11	14	15	18	13	13	16	12	27	16	8	21	18	20	18	18	20	7	5	13	10	7	9	11	17	47
48	15	11	12	13	9	7	40	12	12	16	14	19	12	13	14	10	27	17	10	20	17	19	17	17	20	6	5	13	11	7	9	11	18	48
51	12	9	10	11	13	11	35	10	11	16	16	14	13	11	15	15	11	13	10	21	18	19	18	18	21	4	3	10	10	12	8	11	5	51
54	17	12	11	11	15	13	43	11	13	18	21	19	18	16	20	18	27	16	14	24	22	22	21	22	24	9	8	16	9	15	11	15	17	54
55	10	7	10	9	10	9	35	7	10	12	10	12	11	8	10	11	21	2	10	15	13	13	12	13	14	4	3	11	8	11	8	9	14	55
56	12	8	10	10	12	10	36	7	11	15	14	15	12	9	13	14	22	2	12	16	13	13	12	13	16	6	4	10	8	12	9	11	15	56
57	10	6	11	11	8	6	41	13	8	11	14	13	15	11	14	11	24	16	2	22	22	22	21	22	23	7	6	16	10	9	7	11	14	57
58	11	7	12	13	9	8	34	15	10	13	16	15	16	13	16	13	18	18	4	24	23	24	23	24	24	8	8	18	10	11	8	12	15	58
59	10	7	11	11	8	7	33	9	11	14	9	14	8	8	9	9	25	10	12	13	9	12	10	9	4	4	2	8	9	7	9	8	15	59
6	13	9	1	2	11	9	40	10	11	13	15	16	14	15	17	13	25	15	11	25	23	24	22	23	21	10	9	12	1	10	7	9	17	6
16	9	5	8	9	8	6	40	14	10	15	15	14	11	14	16	13	25	17	8	22	21	22	20	21	21	6	5	14	8	9	7	12	16	16
25	13	10	10	11	12	10	42	13	13	15	17	19	17	13	17	15	27	18	9	25	24	24	23	24	23	9	8	17	10	11	9	12	18	25
29	8	5	10	10	7	5	40	11	8	11	11	11	9	10	13	9	22	16	5	19	21	21	20	21	18	5	3	12	9	7	8	11	11	29
32	8	5	10	10	8	6	41	11	8	12	12	11	10	11	14	10	22	16	6	20	22	22	21	22	19	6	4	13	9	8	8	12	11	32
34	11	7	8	8	7	5	41	14	7	12	12	10	11	14	14	9	26	18	9	20	22	23	21	22	22	7	5	12	8	6	6	11	16	34
41	20	14	11	11	16	15	46	13	17	20	22	22	17	19	21	18	30	19	17	25	23	25	23	24	24	12	11	15	8	16	15	15	20	41
53	16	11	11	12	13	13	32	12	12	17	19	17	18	16	19	16	20	17	14	23	22	23	22	22	22	9	9	17	9	16	11	14	17	53
	60	61	62	64	67	68	70	71	72	74	75	76	77	78	79	82	84	87	88	89	90	91	92	93	94	95	96	97	65	69	73	81	85	

Fig. 16 Purposeful submatrix for classical option

where $C^{\mathrm{bad}}\left(C^{\mathrm{good}}\right)$ is the weight of the most "stiffness" rejected (accepted) object with the rank equal 1; i, j-respectively, the number of the line and the column of the table.

The sequence and the summary of steps of the stage of building of decisive rules are given in Table 3. The algorithm of the stage ${}^{\mathrm{Y}}[0.8]$ differs from ${}^{\mathrm{K}}[0.8]$ in the fact that the leader (flagship) of each group isn't looked for from remained, and is appointed a new—features of g groups of tests are built so that there is always g object in it.

For the rest everything is made according to the same scheme: the optimum structure of objects group possessing the same key features as well as g object is looked for. Results of calculations are in Fig. 17.

Optimization of structure of the tests received according to the simplified scheme (optimization on the basis of the analysis of crossings of groups participants lists) is carried out in 2 steps: removal of enclosed tests and optimization of "overlappings" of tests.

Removal of enclosed tests is a procedure of sorting according to the size of groups—to the number of participants, "passed a test". Filling in the table on

Table 3 Structure of stage $^{K}[0.8]$

$^{K}[0.8.1]$	Unload and put in order of subsets of lines of work table;
$^{K}[0.8.2]$	Step-by-step complete objects groups passed tests;
$^{K}[0.8.2.1]$	From remained accepted objects not passed previous tests find “flagship” with the biggest weight of stiffness and against him and remained “expose” all rejected examples;
$^{K}[0.8.2.2.1]$	Find the most stiffness of remained opponents—”flagship” of rejected objects. Define group of questions using which accepted “ flagship” “wins” rejected one;
$^{K}[0.8.2.2.2]$	Choose the best from the selected group for continuation of creation of question groups—there has to be more “own” and less “strangers”;
$^{K}[0.8.2.2.3]$	Remember the chosen question—eliminate “ flagship” of opponents with its colleagues as who haven’t coped, and also a part of own which have also not answered this question. If not all opponents are eliminated pass to the point $^{K}[0.8.2.2]$;
$^{K}[0.8.2.3]$	«Flagship » and its “colleagues”, which answered all questions, are formed.
As questions were asked for “fight” - the final structure of group was unknown—the questions “for elimination” can describe properties of group not in the best way.	
$^{K}[0.8.2.4]$	To create the optimum choice of questions of the test to which every element from group satisfy, and none of opponents. Record of questions set in the questionnaire.
If not all groups are created (not all accepted objects have passed the previous tests), transition to the point $^{K}[0.8.2.1]$. The last groups can be small, it doesn’t mean that there are no accepted objects which have entered into earlier built groups, capable to fill up the last groups.	
$^{K}[0.8.3]$	Balancing of group members (optionally). Calculation of objects rating.

decrease, from the second line, we make verification with the lines which are earlier placed in the table. If a new “candidate” is the copy or a subset of the test which is earlier included in the table, this “candidate” is excluded from the table as obviously the worst test from recorded earlier.

As a result, 51 test continued participation in calculations (2 have been removed). On the basis of the data obtained from the previous step the binary matrix of relations “objects-tests” is created.

The sum of units in the lines ξ_i is the vector-column containing information on the number of tests to which an object satisfies (i — numbers of object). Its minimum value is minimum quantity of tests to which each object satisfies—a *threshold* ($\varepsilon_{\lim}$).

If $\varepsilon_{\lim} > 1$ then it can be reduced to value ($\varepsilon''_{\lim} > 0$):

$$\varepsilon''_{\lim} = \min\left(\varepsilon''_{\lim}, \varepsilon_{\lim}\right). \tag{4.7}$$

Each accepted object has to pass at least one test. In each column but only in the lines containing unit the minimum is looked for from ξ_i — it is *degree of freedom of the test*—a vector line φ_j. If $\varphi_j = \varepsilon''_{\lim}$ among subjects of the test there is an object which has reached $\varepsilon''_{\lim}$ — this test can’t be removed.

Num6	Num57	Num32	Num31	Num58	Num17	Num16	Num24	Num25	Num21
22	96	33	59	19					
96	22	59	33	19					
Num14	Num45	Num4	Num34	Num39	Num37	Num33	Num54	Num41	Num42
22	85	49	67	70	79				
22	85	49	79	67	70				
Num8	Num11	Num9	Num12	Num46	Num47	Num48	Num53		
70	96	75							
96	75	70							
Num1	Num22								
77	68	32	19						
77	32	19	68						
Num55	Num23								
82	34	6	19						
34	19	6							
Num44									
40									
40									
Num59	Num56	Num29	Num28						
50	98	51							
50	51	98							
Num43	Num5								
50	39	41							
39	50	41							
Num10	Num15	Num38							
102	75	21							
102	75	21							
Num51									
24	83	19							
24	83	19							
43									
35									
34									

Fig. 17 Result of the stage of creation of rules for classical option

In other tests search of a column with the smallest quantity of units among columns with minimum difference is made. The test which possesses this column is removed. Recalculation of ξ_i and φ_j is made. Process comes to the end if for all tests the following condition is satisfied:

$$\varphi_j = \varepsilon''_{\lim}. \tag{4.8}$$

In the reviewed example minimum quantity of tests appeared equally to 8. At this stage operation of an algorithm is complete. The algorithm models the principles of data structure, caused by solved by DM task of classification and its considerably not verbalized representations, feelings, semantic criteria and other entities intuitively clear but escaping formalization.

Essentially, the described algorithm is connected with creation of the field of structure of data and the analysis of its effects, including also clarification of the structure. At any data at the same time there is both an order, and a disorder. As it is difficult "to catch" the excluding OR—the idea of creation of decisive rules on monotonous functions setting a network order is realized. There is a network of subjects (questionnaires, situations, etc.) of classification refered to one class, there is a network of the objects refered to other class, and the algorithm looks for the line separating one network from another. The problem solver sense in geometry is rather simple: it is necessary to match features in such manner keeping properties of a private order that objects on a subset of features will be divided: "good"—have risen, "bad"—have fallen.

It is the classical task of the discrete mathematics about finding of the logic function accepting TRUE value (1) on "good" examples and FALSE value (0) on "bad" examples [63, 64] and it is solved in tens of different methods based on the method of expansion of any logic function in superposition of simpler functions (a formula of von Neumann) [55, 63]. Fine ideas of use of local superpositions of functions through relations of variables in model graphs conferred a prestigious award of von Neumann are offered [64].

There are robotic systems making any chips with given properties for the solution of specialized functions (so-called, chips). As well there are implementations of super-large-scale integral circuits [65] if there is a need of processing of arrays of very big dimensionality. But most often they are technically optimized decisions. Decision methods with optimization in case of all achievements of heuristic mathematics as a rule lead to big search of options, not guaranteeing that found solutions are the most optimum. Methods of creation optimum (containing less variables, or with not crossed multiplicands in logical amounts) formulas for partially given logic functions have algorithms of combinatorial complexity with the exponential growth of expenses of computing resources from the sizes of solvable tables (both by quantity of variables, and by quantity of training objects) [63, 64, 66]. In the reviewed example excessive optimization isn't required: from 125 features of an object 32 features automatically don't participate in creation of formulas of the solver.

Thus the task has actually come down to creation of model of "grey box" [67] testing entrances and exits of some simulator of real system. The model of "grey box" is partially "transparent" model in which elements of structure of decisions are visible and there is an opportunity to interfere with process of control of the solver. The question of extreme quantity of possible mistakes of the solver at "grey" approach can automatically be solved, on condition of introduction naturally of restriction for the type of logical functions adequate to a problem to solve.

In this article, we were restricted to rather representative class—a class of logic functions known as monotonic functions that is dictated by logic of axioms of choice. The monotonic logic function is a logic function which disjunctive normal form doesn't contain any variable with use of logical negation—the unitary operation NOT is absent.

That is, the solver has the following view:

$$y = \sum \left(x_{v,1} \cdot \ldots \cdot x_{v,D_v}\right) \tag{4.9}$$

where y is an object assessment ($y = 1$ for accepted objects, $y = 0$ for rejected objects); v is number of variables group; D_v is dimensionality (quantity of features in group); $x_{v,1}$ is the first feature in group; x_{v,D_v} is the last feature in group.

Value $y = 1$ (TRUE) is reached in that and only in that case when features at least of one group are reflected in the description of an object. By analogy the solver is similar to the examiner who asks examinees different groups of questions (different structure and quantity) and such options of groups only V, and at the answer to all questions gives pass $y = 1$ or not pass $y = 0$, giving an opportunity to be re-examined V times, every time with new fixed group of questions. The first who takes pass are those examinees who have answered all questions $x_{1,1} \cdot \ldots \cdot x_{1,D_1}$. Those who haven't passed in the first group for the second time will be examined again, but on group of questions $x_{2,1} \cdot \ldots \cdot x_{2,D_2}$ and so on until all list of groups of questions—"tests" V is exhausted.

The problem of creation of the solver-examiner consists in such way to match groups of questions that worthy, even not in the first group, but have got "pass", but not worthy have taken all V tests, but "pass" isn't got. Let us note that affirmative answer on questions of only one group is sufficient for getting "pass" (at creation of the solver without duplication increasing reliability) therefore in case of use such algorithm, for example, when tendering or an assessment of a creditworthiness of clients, the structure of groups of questions has to remain in secret—the examinee will get "pass" if he answers all questions in group irrespective of how well he would answer other questions.

There are many ways to strengthen the solver of view (4.9). For example, by means of a requirement task that the classified object has to be selected at least on 2 groups, 3 groups, on groups with large (excess) number of questions. It is possible "to cut off" the number of groups V, that is to increase reliability of selection having left those who have passed the first $\overline{V}$ tests. It is a way of mistakes management of the first sort.

It is possible to operate structure of features—whether to consider at once the main features or group of specific ones. Then dynamics of passing will either form groups with average specifics, the number passed through each test will be more uniform.

Interesting situation analysis turns out: as if automatically poorly crossed groups of similar descriptions of situations are under construction—the typology of "applicants" on a combination of features which enter the solver is created. Such typology can be useful at prequalification (reception or refusal of new applicants) at the auction: there is no need to take firm of that profile (type) which examples (among already chosen) are enough. Here it is possible to go to mistakes of the second sort, having taken in prequalification, for example, those who haven't answered all questions D_v, and have answered only $D_v - 1$ or $D_v - 2$ question.

If the group is constructed, then set of groups with the “sparing” admission is under construction automatically easily: instead of summand

$$X_{v,1} \times \ldots \times X_{v,D_v} \tag{4.10}$$

it is necessary to write the sum of summands at which one of factors is missed if we forgive one “unsuccessful” feature from group $X_{v,1} \times \ldots \times X_{v,D_v}$ or the sum of combinations $C^S_{D_v}$ summands in which from product in (4.10) about S features are omitted.

For example, if in group of 5 features “we forgive” 2, then instead of product of five factors we have to write down in the found formula of the solver 10 products on 3 factors with different three features from five:

$$\begin{aligned} &X_1X_2X_3 \cup X_1X_2X_4 \cup X_1X_2X_5 \cup X_1X_3X_4 \cup X_1X_3X_5 \\ &\cup X_1X_4X_5 \cup X_2X_3X_4 \cup X_2X_3X_5 \cup X_2X_4X_5 \cup X_3X_4X_5 \end{aligned} \tag{4.11}$$

Or simpler: the quantity of questions which have received the correct answer at this test has to be more or equally $(D_v - S_v)$. If mistakes are not forgiven, then $S_v = 0(C^0_{D_v} = 1)$.

Toughening of rules of selection is technically simple: if there is a test $\{X_{v,1}, \ldots, X_{v,D_v}\}$ to which a large number of objects satisfies, then it is necessary to fill its up with additional questions. It is desirable to have variable answers among “participants” of this group, for example on two questions $\overline{X}_{v,D_v+1}$ and $\overline{X}_{v,D_v+2}$ of “increased complexity”. It is ideal if a half of objects which have passed an “old” test contains only the 1st most informative feature, and the second half—the 2nd. Both will remain in prequalification, but will already begin to belong to two groups: $\{X_{v1,1}, \ldots, X_{v1,D_{v1}}; \overline{X}_{v1,D_{v1}+1}\}$ and $\{X_{v2,1}, \ldots, X_{v2,D_{v2}}; \overline{X}_{v2,D_{v2}+1}\}$.

In each of groups there will be an additional question and it will become more difficult “to get” into such groups. Thereby we carry out “seeing” reengineering of groups of questions without intricate algorithms of adaptation [68]. Groups become smaller, but the classification is more exact.

There is a mechanism of effective adaptive control. It will especially be shown when digitization not of logical, but quantitative variables where to a feature there corresponds some range of admissible values of key indices is used. Thereby “sparing”, we expand tolerance ranges of changes of values of these indices, and, on the contrary, “hardening”, we narrow these ranges. Last it is reached when as a feature of toughening the feature describing closeness to boundaries of admissible zones of values of key indices is selected.

Sharp growth of mistakes of the second sort demands retraining of the solver for “cultivation” of features belonging to old groups on new. And if this turns out to be ineffective (the solver safety margin won’t increase—quantity of mistakes of the first and second sort will approach threshold values), then introduction of new features will be required.

There is an important question: whether it is possible to consider that the rules synthesized by an algorithm really express certain “properties” of compared objects, and are not the result of game of casual fluctuations of selection in many-dimensional space of their description. In [69] it is shown that at a distribution of classes A and B presented by sets of objects of M_A and M_B in a K-dimensional subspace of H-dimensional space total number of subspaces in which incidentally separating planes will be looked for makes in a typical case $\frac{H^K}{K!}$. If the acceptable size of mathematical expectation makes no more $\frac{1}{H}$ of this size (i.e. approximately $\frac{H^{K-1}}{K!}$), then from inequality $\frac{H^{K-1}}{K!} > \frac{2^{K-M_A-M_B-1}H^K}{K!}$ follows that $2^{M_A+M_B+1-K} > H$. Setting $H = 2^{10}$ (that corresponds to usual number of features), it is shown that $M_A + M_B + 1 > 11\ K$. In more realistic situation when creation of the decisive rule in the form of the only K-dimensional plane is impossible, but the piecewise linear rules consisting of two planes are real it is shown that the assessment for mathematical expectation has a view [69]:

$$MO_{AB} > \frac{2^{K-M_B-1}H^K}{K!} \tag{4.12}$$

From this it follows that AAS will generate nonrandom two-element piecewise linear rules if each of divided classes is presented by not less than 23 objects at $K = 2$ and not less than 56 at $K = 5$.

5 Risk Management

Risk management is the main, final stage in risk analysis on which priority areas and optimum security aids are defined. Main questions considered at this stage are establishment of importance of the considered risk, comparison of costs for its reduction with those benefits which can be received as a result of it and support in carrying out policy of risk reduction. At this stage it is possible to receive the answer to a question of what risk level is excessively big and therefore demands creation for its technical systems decrease and what security systems from the point of view of material inputs are the most effective.

At first sight it can seem that the task of risk management is extremely clear and is in reducing risks from each technological unit to unified minimum or even just “zero” level. However, all researches in this area show that it is unreasonable to want to reduce risk from various technological units, for example, such as coal industry or nuclear power plants, or even in each concrete unit to one level.

The task of risk management can be brought to that on the basis of some certain criteria to answer a question of what means and to what level it is necessary to reduce risk in this or that field of production activity. One of possible and the most widely developed tool of the solution of this task is the assessment of efficiency of costs for risk reduction.

The analysis of efficiency of costs for risk reduction is directed to the solution of tasks connected with application of these or those measures and security systems. Therefore, it has to be applied preferably at a design stage of development and provide conditions for the choice of the most optimum actions for increase in safety both during normal work of the enterprise, and in emergencies.

As it has been already told earlier one of the main tasks of the stage of risk determination is in this fact that those criteria and corresponding methods which would allow to answer the question what is to consider dangerous and what is—safe. Especially sharply these questions rise for existing and again created difficult technical systems connected with use and production of large amounts of energy and dangerous substances where accidents can lead to environmental disasters. The similar situation is observed when a natural control system of safety is provided solution of problems which it never solved and, perhaps, isn't capable to solve, for example, on utilization of outlying to nature chemical compounds, large quantity of fusible substances, etc.

Considering this, it is necessary to use existing current regulations and to develop new ones for safety or more likely the rules of risk management which have to be based on determined criteria and norms.

Every day all people are affected by different types of danger. When it occurs from the person himself voluntarily, usually it doesn't cause worry. So, for example, by force of habit, social conditions or financial interest many people smoke, use cars or go to work as porters though the risk of death in these cases is rather high. The question of safety raises special worry when risks exist irrespective of people's desire. In this case the risk means something very bad that by all means needs to be avoided. However, all technical progress as it was shown earlier, along with those benefits which it brings at the same time leads to increase in technogenic risk. Therefore, regardless of desire of society it is forced to exist at a certain risk level.

There is even a statement that "without risk there can be no safety". Really in some cases society takes a certain risk because of receiving an opportunity to increase the safety finally. So according to the American data, every year 3.5 million children in the USA receive whooping cough, tetanus and diphtheria inoculations, 25 thousand from them feel after bacterination very hard they have big temperature and fever, 9 thousand are ill seriously, 50 people receive serious complications on brain, and about 20 people die. And still, despite such risk, health services every year hold an inoculation since otherwise the quantity of diseases and deaths will be many times as large.

So, for example, according to data for underdeveloped countries from 80 million children annually born and not receiving any inoculations 5 million dies from measles, whooping cough and similar illnesses on which prevention the program of immunization is referred. It is simple to count that the mortality for this reason is 6.25%. In a case with vaccines about 0.002% of children will die and will seriously suffer, i.e. more than 3000 times less.

However, it should be noted that now the risk is measured not only probability of a lethal outcome or a disease. From this point of view one lethal outcome can be equated to loss of 6000 working days, and, for example, one occupational health problem can be equated to loss of 338 working days. Risk concept also includes damage caused as a result of this influence in ecological, social and economic areas. Types of consequences which can be considered include:

1. untimely death;
2. latent cancerogenic effects;
3. early diseases;
4. lost man-expenses;
5. property damage;
6. financial losses, etc.

It can be effects of short or long impact on the local or regional industry, impacts on a social status of the population, effects of local, regional or even global character. Some of these effects always occur, others have only probabilistic character.

Because of a wide variety of possible effects one of the main problems is determination of sizes and units of measure of risk. Some types of risks can be measured by number of people subjected to influence, probability of death or disease or to be expressed in monetary units. One of possible units of measure of risk types from 1 to 4 can be lost resource days. Other risks, in particular in social area, are very difficult to present in numerical units so far and therefore, even participating in risk analysis in risk assessment they aren't considered yet.

From there, existing criteria of safety (or risk criteria) can be divided into two main groups—qualitative and quantitative.

Qualitative criteria unlike quantitative ones often depend on opinion of experts and therefore are more subjective and can have wide dispersion depending on knowledge of experts, their qualification, etc. Quantitative criteria are the basis for determination of those norms which are used for elaboration of strategy and tactics of risk management for the purpose of increase in safety.

From the definition of risk, it follows that it is characterized by three components:

- danger source;
- danger distribution area;
- subject of influence.

Therefore, criteria of safety and existing standards often belong to these components.

For danger sources which are technical systems, criteria of safety can be standards on probability of beyond design basis accident (for example, for the NPP this is melting of an active zone which probability shouldn't exceed 10^{-7} for the reactor in a year), or sizes of maximum permissible emissions (MPE) of dangerous substances in the atmosphere or discharges (MPD) into water in the conditions of normal operation of the enterprise. Also one of criteria of safety can be restriction for quantity of stored, processed or transported dangerous substance.

Criteria of safety refered to extension area (atmosphere, surface or ground waters, soil, etc.) traditionally are maximum permissible concentrations of dangerous substances (MPC) in this area. On the basis of maximum permissible concentrations values calculated most often on "the worst case" traditionally there will be such protective measures for increase in safety of the population as sanitary protection zones, restriction for emission of pollutant in the atmosphere, etc.

The subject of influence most often is understood as either a certain person, or group of the people united by the place of residence (for example, the population of the explored region), by the place of work (service personnel of the enterprise), etc. Criteria of safety for them most often were either individual, or collective doses in case of a possible adverse effect of dangerous substance, connected either with accident, or with normal operation—technical system. Usually the dose is the concentration of dangerous substance integrated on time and is connected with this or that type of consequences for person's health. At the same time the dose criterion of safety for the serving personnel most often has much bigger value in dose size, than for the population shivering near the enterprise.

However, the size of a dose isn't connected with a possibility of receiving it by person or group of people, especially in case of accident at the enterprise. We will easily eliminate this defect when using a concept of risk for danger assessment.

Today the main criteria of safety at risk assessment are either the individual, or collective risk of death or a certain disease. The individual risk is measured by probability of death (disease) of an individual in a year, collective—by the number of affected people from a certain group. Quantitative norms for this risk are being established. So the individual risk of death of the person in one year caused by industrial activity more than 10^{-6} is considered unacceptable, and risk less than 10^{-8} is considered negligible. The acceptable risk level is chosen in the range from 10^{-6} to 10^{-8} a year, proceeding from the economic and social reasons. It doesn't mean that the individual risk of death from any industrial activity in any society has to be not more, than 10^{-6} a year. So, owing to historically developed conditions or proceeding from social preferences society or its certain members can accept much higher sizes of risk, receiving at the same time material, economic or some other advantages.

As the purpose of safety is not only protection of the population, but also the environment, at consideration of safety issues it is necessary to have indicators which quantitatively would define its state. At today's level of knowledge, it is possible to refer to such characteristics proximity of ecosystem condition to its stability limit where predictability of changes of ecosystem in response to external influence will be lost. In Holland such sketches already use for introduction to the regulating activities for safety of values for maximum permissible environmental pressures. So possible criterion of safety of the environment as a result of industrial activity can be the level of impact on an ecosystem which shouldn't exceed values at which 5% of species in an ecosystem can suffer.

The economic damage connected with accidents at the enterprises even under conditions of their normal functioning existed practically always, and at the modern level of development of the industry it takes such scales of impact on the local and regional industry that need for development of economic criteria of safety appears. The example of such criterion can be the size of possible damage as a result of accident which has to be less than those benefits which are available or will be received as a result of functioning of technical system.

All above mentioned often leads to great difficulties in association of sizes of various risks and especially this task becomes complicated when attributing the scales depending on social preference to this or that type of risk at development of criteria of safety and standards at risk sizes. At the same time, it is necessary to consider that society accepts risks with high probability more often, but small, in its opinion, damage and even such as loss of human life, in comparison with the risks connected with accidents with low probability of implementation, but big potential consequences. Most often society has rejection on the relation even to single accident with a large number of loss of human life. Therefore, at decision-making often risks with small probability but with extensive possible damage can be represented by unfairly high values.

Nevertheless, concept of risk and tool for its assessment even more often is used in various countries for carrying out rational policy of industrial safety. So in Holland when planning industrial activity along with geographical and economic cards they use risk cards for all country. Some developed countries have already conducted special researches on risk assessment in large industrial regions. In these countries quantitative risk assessment and especially risk management at the level of industrial regions became important aspect of decision-making on regulation and protection of human health and the environment [54].

Determination of acceptable level of various types of risk in society, their consideration from the point of view of social, economic and political positions and preferences, unifying of various types of risks for determination of size of total risk, all these are one of the most important components of the procedure of risk analysis.

Especially sharply safety issues rise in regions with high concentration of industrial and energetic productions where at risk assessment from this or that enterprise it is also necessary to consider risk from other enterprises if they are located nearby and can influence each other, creating at the same time additional risk. So, for example, the neighborhood of the NPP which is the reason of the increased humidity of air at considerable distance (that doesn't constitute essential danger) and thermal power plant which represents a constant source of emissions of SO_2 actually means that in the region there is a source of emissions in the atmosphere of sulfuric acid representing big danger both to the population, and to the environment. Besides accident at one object or the enterprise can initiate accident at another one creating so-called "domino" effect that also has to be considered at danger assessment.

Therefore, according to modern understandings, the sketch and especially risk management for the purpose of ensuring necessary level of safety in society have to be carried out not for separate production installations but for all industrial complex. Because of this there was a need for development of unified methodology which would allow to define and describe main sources of danger in the region, to estimate related risks both for conditions of normal functioning of the enterprises, and in emergencies, to have an opportunity to compare various types of risks, to define the most effective measures and technical systems intended for increase in safety of the population and protection of the environment, to develop necessary measures for reduction of consequences of accidents, etc.

Besides this methodology in principle has to be directed to determination of level of safety which is optimum for the explored industrial region and therefore has to serve as the purpose to which achievement all actions for decrease in risk have to be directed.

Now basic provisions of such methodology directed to creation of the unified approach to studying of sources of danger and risk management, connected with them in regions with advanced industry are integrated in the procedure of the analysis of regional risk.

The analysis of regional risk as well as the analysis of risk in general includes three main interconnected sections: identification of dangers in the region, assessment of the risk connected with these dangers and its management for the purpose of increase in safety of the population and environment protection.

Possible realization of the questions connected with identification and ranging of sources of the danger for the purpose of allocation of the most essential intended for practical use in the analysis of regional risk can be the way of identification of danger in the region based on the actual material [53].

Considering the scale and complexity of practical realization of such methodology, researches are offered to be conducted from up to down that allows roughly at the first approximation to capture various power and industrial productions using simple technology of selection of kinds of risk under study.

According to this technique at the initial stage it is necessary to carry out qualitative description allowing to capture a set of productions of various type, to order and select for the subsequent analysis of production and separate installations constituting the greatest danger for the population and the environment in the region. At the subsequent stages according to the general methodology the risk assessment from the most potentially dangerous objects chosen at the first stage in the region has to be carried out, and then on this basis the strategy of regional risk management for the purpose of increase in safety can be developed.

Any assessment of regional risk must begin with the choice and write-off of the respective region. The basis for the choice of the region and definition of its borders in each case are defined by the specific features and therefore they can't be formalized or specified any tough rules for this purpose. However, there is a number of provisions which needs to be taken into account at the choice of the explored region:

- the region is chosen proceeding from its industrial and economic characteristics but not from administrative borders;
- at an initial stage of risk assessment, it is impossible to designate rigid borders of the region since areas which can have impact aren't defined yet;

- in consideration it may be necessary to include separate enterprises which are in borders of the same water or air basin as the explored region;
- in consideration of full production or power systems it is necessary to consider information on components of these systems beginning from fuel extraction before production of the final product though these parts can be far out of borders of the explored region;
- it is necessary to consider transport systems (automobile and railroads, waterways and pipe ducts) on which import and export from the region of dangerous substances is carried out and their characteristic;
- some sources of danger can make essential impact far outside the studied region; in such cases, except consideration of local influences it may be necessary to consider global.

After choosing the region under study it is necessary to describe it. For this purpose, it is necessary to have a certain database on state of environment, geographical and meteorological information on the region, data on placement of the enterprises, etc.

After the main information on the region is collected it is necessary to characterize the main sources of potential danger. At the same time the main attention has to be paid first for the enterprises at which potentially dangerous substances are used, stored or produced, for objects of storage of such substances and their transportation.

To have an opportunity to define what danger this or that industrial system, production or an object constitute it is necessary to have certain information on each of them. It is necessary to consider that since methods of determination of parameters of a source of potential danger which have to reflect the nature of their work and the related risk are various depending on conditions of their functioning, it is necessary to do distinction between normal (i.e. calculated, accident-free) work and emergencies which are usually characterized by low probability of implementation, but extensive possible damage.

Consideration of characteristics of the enterprise has to be based on a complete description of industrial complex, on inventory of all substances stored, processed or produced. For definition of potential danger, especially in the conditions of normal work of the enterprises, it is necessary to have full list of production wastes what of them and in what quantities are discharged into water or air environment, what ways and methods of their processing are available and main ways of burial.

It is also necessary to have information on types and intensity of transportations of dangerous substances in the explored region. At the same time, it is necessary to consider transportations by automobile, iron and water highways and their transportation by pipe ducts.

After information on the explored region stated above is obtained, primary identification of danger can be carried out. The main purpose of such identification is ranging of sources according to the level of danger in order to mark the most essential which have to become subjects of the further risk analysis first of all. It is based, mainly, on an expert assessment of possible consequences and for this reason is partly subjective and depends on the scope experts' knowledge and qualification. Therefore, at this stage of risk analysis it is necessary to involve as much as possible experts of various profile.

Depending on source type at an initial stage of an assessment it is offered to mark out two categories of risk: risk at accident and other extreme situations, and risk at normal operation.

According to risk category division into more several types can be carried out: risk of death, risk of chronic diseases, risk of economic and financial losses and risk of ecological damage which generally are considered when determining danger degree.

Those objects which as a result of an expert assessment will receive the greatest values of such numerical criterion can be considered as the most dangerous, i.e. mainly defining regional risk and therefore have to be marked for consideration first of all. Though absolute value of this criterion from the point of view of the size of risk doesn't bear any meaning, it allows to distinguish, and it is very important point in regional risk assessment procedure, the most essential sources of danger.

To make primary identification of danger in the region more objective in the offered technique they use databases on so-called threshold values of most potentially dangerous substances used in the industry. At the same time for creation of the unified approach the considered dangerous substances are subdivided on toxic, combustible and explosive.

Toxic substances, in turn, are subdivided on strongly toxic, toxic and substances which can be potentially dangerous only as a result of accident.

Gases which with a normal pressure and when mixing with air become fuel and substances which boiling point is equal or less than 20 °C belong to combustible substances. Also here there are flammable liquids which point of ignition is below 21 °C, and the boiling point is above 20 °C, and flammable liquids which point of ignition is below 55 °C and which remain in liquid state under pressure, but at a high temperature and pressure can become emergency and dangerous.

Those substances, which can blow up at ignition or those ones which are more sensitive to shock impacts, than dinitrogasoline, belong to explosives.

When considering emergencies the full number of everyone dangerous substances at the enterprise, in storage or in one transport unit is analyzed. At the same time criteria of danger with the corresponding consequences are threshold amounts of substance.

For normal work of the enterprise the main dangers both to the person, and to the environment present leaks and emissions of harmful substances in the atmosphere and storage of potentially dangerous wastes. At an initial stage of danger identification to a settlement operating mode it is offered to consider risk of chronic diseases, the second can be caused by these reasons, and possible economic and ecological damages.

Risk of chronic diseases may be caused by emissions in the atmosphere of gases as SO_x, NO_x, CO, O_3, H_2S, HCl and some others, and also radioactive materials, hydrocarbonate, dust and smoke. The reason of the diseases connected with industrial activity in the region may be ingress in water heavy metals, biocides, acids, nitrates, fertilizer, cancerogenic and radioactive materials.

When considering a question of economic damage, they consider possible losses of a harvest of forest lands, the cattle, fish, acid impact on buildings, etc. Possible ecological damages are assessed as well as in emergency cases.

Everything above mentioned doesn't mean that the productions which do not meet numerical criterion, i.e. having small value of danger index shouldn't be a research object in the analysis of regional risk. The described approach of primary identification of dangerous productions allows establishing only priorities for this analysis.

After primary identification and ranging of productions according to degree of their danger it is possible to pass to the following stage of the procedure of risk analysis—a risk assessment for determination directly of danger levels from industrial activity in the region. Principle points in risk assessment are the detailed description of danger source and determination of the related possible damage.

There are various models of danger sources which allow to define probability of this or that development of accident and to determine the corresponding power of emission of dangerous substances in the environment. For a case of normal work of the enterprise it can be measured sizes. There is also a number of models allowing to calculate distribution of various dangerous substances in air, water and the soil, to define doses of impact on the person and to assess possible damages.

However unfortunately there is no unified technique of detailed calculation of possible types of risk from various types of production which could be applied to regional risk assessment. Therefore, for practical activities approximate approaches of realization of methodology of the analysis of regional risk are developed and used now.

One of approximate approaches of regional risk analysis can serve the technique developed by the Dutch firm TNO [53]. This technique is intended for carrying out the unified integrated sketches of the regional risk connected with use, storage and transportation of large amounts of flammable, explosive and toxic chemicals. It doesn't consider risk for conditions of normal functioning of the enterprises, but can be used for the top assessments of emergencies risk.

The risk level connected with dangerous objects in the region is defined by the size of possible damage as a result of accidents on these objects and probability of emergence of these accidents. The size in this technique is meant as potential human losses which can be two types: in a zone of irrevocable losses people receive deadly injuries, and for those who get into a zone of sanitary infection hospitalization is necessary.

According to the offered technique, procedure risk assessment from accidents contains identification and classification of potentially dangerous objects and kinds of activity in the region, and also determination of parameters characterizing danger degree. At the same time it is considered that each object or kind of activity can be connected with use of only one potentially dangerous substance. These objects can be sources of accidents of various types which are defined by the class of used substance. Substances of one class lead to the same types of accidents and allow to use a unified algorithm for calculation of defeat zones.

In total in the technique data on properties of II classes of substances are offered. However, for substances of one class the size of zones of defeat is defined by a class of substance danger which is characterized by an integer number from 1 to 6 that corresponds to danger degrees from very low to extremely high.

Contours of each zone are represented either circle, or rectangle and the center of a defeat zone has surely to coincide with a point of an arrangement of production object under study. The size of these zones is characterized by radius for circle or length and width for rectangle, and shift concerning an object arrangement point. If there are several possible options of calculation of zones defeat, then the technique advises to choose the worst case, i.e. option with the largest area of a zone of irrevocable defeat. If real distribution of population density in the explored region is known, then the potential damage is defined by calculation of number of the people getting to limits of defeat zones.

For calculation of the risk connected with dangerous production it is necessary to define also probability of realization of accident with a certain type of damage.

The data provided in the technique allow receiving an assessment of the maximum defeat. Possibilities of other accident development are considered through corrections. The total assessment of probability is the product of frequency of accidents, on the objects of a certain type and corrections considering a factor of existence and non-uniformity of people distribution in the researched territory.

In technique applications there is the database on different types of accidents with different dangerous substances and minimum of reference information which is necessary for calculation of damage.

Thus the offered technique is directed to define the most essential sources of danger in the region and risk assessment connected with accidents in them. As a matter of fact, it is the first approach in the solution of tasks of the regional risk analysis where, unfortunately such important part as risk management isn't considered.

Nevertheless, this technique allows receiving on a unified basis rough, calculated on the worst case, risks assessments of each type from accidents on objects with various types of dangerous substances and therefore is the necessary tool for an assessment of regional risk now. It can also serve as a system basis on providing and supporting of decision-making in emergencies at industrial facilities.

According to the interest shown to this technique on its global reach of problems of regional risk analysis and on advantages of the uniform unified approach, most likely further it will change into more real models of danger sources, transfer of dangerous substances and impacts on the person and the environment. Inclusion in this technique of questions of risk management will make it the powerful tool for problems solution of industrial safety in the region.

With development and complication of productions and technological processes the security systems used on them to which creation and operation it is necessary to direct more and more and big material resources were integrated and became complicated. For some technologies which are especially connected with power, now the cost of security systems already makes considerable part from complexity of the technological installation itself. For example, in nuclear power cost can reach a half of the NPP cost. Therefore, safety level is always defined by a compromise between the aspiration to use limited resources most effectively and at the same time to reach the smallest risk level.

As recently it is realized that achievement of absolute danger is impossible and society is forced to exist at the given risk level, there is a question of what has to be the minimum degree of securities of the person and the environment from harmful effects of industrial production. In more widely used formulation of this question it sounds as a question of what has to be the level of the acceptable risk from industrial activity of the person.

Now the problem of optimum safety comes down to that on the basis of some certain criteria to answer a question of what means and to what level it is necessary to reduce risk in this or that field of production activity in order the safety both the person, and the environment will be maximum.

Today one of possible and the most widely developed tool of the solution of this task is the assessment of efficiency of costs for risk reduction [56].

The expenses efficiency analysis in a general sense is directed to assess economic efficiency of possible actions. In questions of industrial safety determination of acceptable levels of impact on the person and the environment can become an ultimate goal of such analysis. At the same time the all-round price of such influence has to be defined both by the cost of protection means from it, and the cost of damage caused by it.

Usually more difficulties arise in attempt to present in monetary terms the cost of such damage and, especially, cost of human life in connection with possible deaths or diseases. Due to the lack of unified methodology of human life assessment alternative approach has been developed for definition of social consequences from potential influence where the cost of human life isn't presented explicitly. Such approach is based on direct comparison of cost of measures for reduction of adverse effects and damage caused to health of the person which can be expressed by quantity of death or diseases connected with this level of danger.

At the same time the level of influence can be characterized by such indexes as concentration of harmful substances in this or that environment, or the dose power, power of emission of substances from a source or others.

Therefore, one of criteria of security measures efficiency can be their cost. Not only engineering decisions, but also organizational actions at the same time can be understood as security measures.

Often there is a question of what security system it is necessary to use in order that expenses will be minimum, and effect of introduction - maximum. The method allowing assessing and choosing among alternative an optimum security system on decrease in risk level is the analysis of efficiency of costs for its decrease.

The analysis of efficiency of expenses is the tool of the solution of tasks of risk management. Risk management in its turn is the main final stage of risk analysis at which priority areas and optimum security facilities are defined. At this stage of risk analysis, it is possible to receive the answer to a question of what risk level is excessively big and therefore demands creation of technical and organizational measures for its decrease and what from the point of view of material inputs security systems are the most effective.

At first sight it can seem that the task of risk management is extremely clear and is in reducing risks from each technological unit to the uniform minimum level. However, all researches in this area show that it is simply unreasonable to try to reduce risk from various enterprises or even in one enterprise for each concrete production installation to uniform level. The analysis of efficiency of expenses once again convincingly confirms it.

The procedure of the analysis of costs efficiency for decrease of risk begins with the fact that for sources of potential danger it is necessary to define all possible security systems and to estimate residual risks after their introduction.

Choosing a security system for possible use it is necessary to consider such factor as its cost. Assessment of security systems cost is connected with need to connect cost with specific actions of risk reduction. It is clear that it is logical to use only such systems which at identical expenses allow reducing risk most effectively. Therefore, the task of risk management can be lead to a problem of optimum distribution of limited funds for security systems.

It is known that expenses on creation and operation of such systems follow the law of return reduction. According to this law it is possible to lower each level of risk up to some size ΔR, having spent certain material resources of society ΔC for creation and operation of security systems necessary for this purpose.

New value of risk after introduction of these systems will determine the appropriate reached safety level. However further risk reduction and therefore increases in level of safety become more and more expensive. Value of the ratio of ΔR to ΔC can serve as a measure of efficiency of costs for risk reduction. This efficiency can be measured in the number of the prevented impacts on health of the person, on a unit of value of risk reduction.

Sometimes for the characteristic of expenses efficiency they use the inverse value equal to $\Delta C/\Delta R$. There is no fundamental difference here. Just in the latter case they consider that this ratio defines limit costs for risk reduction.

In both cases expenses efficiency is that criterion which is necessary for the choice of systems on risk reduction. Optimum security systems are those for which the ratio $\Delta R/\Delta C$ is maximum, or $\Delta C/\Delta R$ – minimum.

From above mentioned follows that limit costs for reduction increase with the level of reached safety. For each concrete risk level its further reduction is possible, however it is impossible to reduce him to zero.

It results in need of use of a concept not of the “zero”, but *acceptable* risk level. This level has to be defined both by extent of economic development of society, and social conditions of this society. Therefore, for each country the level of acceptable risk must have its value and a measure of efficiency of expenses is one of the criteria defining its value.

At estimation of cost of security systems there are several problems connected with need to determine what parameters will be read out as indicators of this cost. It is necessary to distinguish conditions of normal work and accidents where these indicators will be significantly various. The situation can also be complicated by the fact that the measures directed generally to decrease in one type of risk can bring to increase in risk levels of other type. There are no concrete instructions and recommendations about this occasion. Apparently for each type of production this issue has to be solved separately.

The analysis of efficiency of risk reduction costs is directed to the solution of the tasks connected with application of these or those measures and security systems. Therefore, it has to be applied preferably at a design stage of development and provide conditions for the choice of the most optimum actions for increase in safety both during the normal work of the enterprise, and in emergencies.

6 Risk Forecasting

In general, the level of reliability of any structurally difficult system is influenced by a large number of various factors to which it is possible to add both system structure and its controlling facilities, and reliability of the used equipment, level of the organization of system operation and other. The crucial role for prevention of negative influence of possible indignations on normal functioning of big systems, i.e. in ensuring system reliability, is played by stocks and reserves of all types.

Division of means of ensuring of system reliability into inventories and reserves is explained by the fact that in determination of an economic entity of their categories there is no complete clarity so far. They have different economic sense and perform different functions. Inventories provide a continuity and uninterrupted operation of a cycle of reproduction in the conditions of objectively originating gaps between time and the place of production of goods, and its consuming. Therefore, inventories are connected to sure events. Other economic nature is in reserves. Existence and need of reserves are connected to events which have (unlike the factors predetermining need of inventories) the probable characteristic. Value of inventories and reserves, and therefore costs for their creation is defined by different factors. The size of inventories is caused by terms of delivery of raw materials and materials, and also shipments of finished goods, the nature of technological process and a level of synchronism of production and consuming. The size of reserves will be caused by a type and value of deviations from normal activities which can arise at all stages of a reproduction cycle.

It is necessary to mark that despite distinction of an economic entity and assignment of inventories and reserves these categories are inseparably linked as two sides of a single system of support of dynamic stability of functioning of the considered structural and difficult system.

System approach to research of these tasks naturally results in need of modeling of various objects of the nature and society as developing systems which concept gains the increasing value in the most various branches of science [70–76]. Any object of a research added if it is necessary with certain communications with other developing objects, in particular, with subjects of a research, it is possible to interpret in the form of developing [77].

In prognostics the problem of the analysis of a subject of forecasting is reflection of more common problem of systems analysis in general. In each case the way and results of the analysis are defined by research purposes and character of the analyzed object [78].

At a research of ranks of historical data of rare events we deal with discrete dynamic probabilistic processes. Therefore, the purpose of the analysis of forecasting subject is development of predictive model of dynamics of emergence in it non-staff situations experiments with which will help to reduce uncertainty degree, for example, dates of events and their scale due to detection of hidden regularities. Hidden regularities indicate either changes of a condition of the modeled object, or regularities of changes of parameters of the external environment significantly influencing its functioning.

The main prediction phases are flashback, diagnosis and a prospection (prediction). At the same time upon transition from one stage to another the forecast is continuously specified: specification of structure of the studied object and optimization of structure of the description of an expected background is carried out (i.e. search of the most significant characteristics of the external environment is conducted).

Initially time (date of emergence of events) and derivative indicators defined through it—year, month, number of day in a month, day in working week act as primary characteristic of an expected background of any dynamic process. Further, along with identification of that how equally or variously these characteristics, derivative of time, influence dynamics of indicators of accident rate, the decision has to be made: on what periods to aggregate accident rate indicators, and if to aggregate, then in what way.

This stage is extremely important as aggregation in fact is “substitution” of difficult dynamic information data arrays by their two-three generalized characteristics:

- average value (mathematical expectation);
- dispersion which big emissions of data strongly influence (in the case under consideration—indicators of accidents with big damages and/or losses of gas) and, the main thing:
- type of distribution of the aggregated sum.

Mathematical expectation is one of the worst indicators for use in forecasts. Indicators like an average (weight-average [79]) value when in the studied statistics there are data with big emissions [80] are especially bad.

More steady and reliable indicators are considered:

- median of distribution of random variable [81], or
- a median of the set constructed of values of half-sums of all possible couples from statistical selection (so-called, Hodges-Lehmann statistics [82]).

One more indicator applied to an assessment of the sizes of a corridor of change of indicators of the predicted non-stationary process is the dispersion representing the sum of squares of a deviation from an average. Application of this indicator, when composed, forming its value, differ in several orders is also questionable. In case of such dispersion in data, the size of the size of a corridor is determined only seldom observable composed with great values. It is remarkable that the data forming the corridor size don't get to this corridor. And it, for example, in case of the analysis of accident rate of Unified Gas Supply System is the most significant non-staff situations with essential damages and losses of gas.

In case of the analysis of statistics of non-staff situations in Unified Gas Supply System it is incorrect to speak about an indicator of the general dispersion as the indicator of the size of ellipses of dispersion of the measured values in which (according to the theory) experimental data concentrate. This statement doesn't demand the proof as for set of events with small consequences, by definition, dispersion is obviously small, and for major accidents is obviously big and takes place obvious violation of uniformity of selection.

Major accidents are unique. As a rule, scenarios of cascade accidents are implemented in them (i.e. almost determined scenarios), the realization effect of which can't be described by a large sum of averages and small negative deviations. Therefore, in long-term statistics of severe accidents there are no accidents with a scale of losses of 70–90% from accident of the similar, for example, Chernobyl accident.

So, dispersion as the indicator of "temperature of the system" specifying predisposition of a status of the researched system to more vigorous chance fluctuations of rather "average" equilibrium parameters is also inadequate to the researched object statistics rate. As a result, rash use of the principles of statistical molecular physics for the dynamic description of scales of damages, loss of gas and frequencies of the accidents (which are quantized by the size), as a rule, is led to the conclusion that analyzable processes are processes of the accidental nature and any regularities are absent in them. At the same time the conclusion about a mismatch of model and reality is made on the amounts of a difference of squares between model and real data, i.e. besides on dispersion of discrepancies. There is an explanation of one mistake with using another.

The above defines difficulties in questions of studying and modeling of behavior of structural and difficult systems and a question of types of distribution of indicators of accidents, especially accidents with serious consequences. Because of the quantum nature of accidents use of the device of the analysis of data based on classical laws of large numbers is incorrect. Actually, convergence on probability in reality practically is never observed, except for the statistics which is saved up in systems of mass service. Compliance of reality of the theory in these fields of activity is reached at the expense of very large number of the same realization.

The qualitative and productive criticism of the statistical analysis of the data containing big emissions of values is available in many monographs, for example, a row "robust" (steady against similar emissions) methods of statistical processing contains, for example, in [79].

For example, the analysis of dynamics of dates of accidents of such large territorial distributed structural and difficult system as the Unified gas supply system (Russia), visually and convincingly shows that dates of emergence of accidents aren't realization of Bernoulli random walk value at all.

Mathematical expectation of number of accidents in day calculated as the relation of the actual number of emergency days in a year to duration of the analyzed period turns out below practically all points from the training selection that confirms the assumption that it is impossible to use population mean as the indicator describing reality for this system.

The dispersion indicator σ^2 also obviously doesn't reflect the size of dispersion of accident of dates of accidents. So, from "the theory of random walks" [82] in the area between the average value increased and reduced by value σ about 34,5% of accidents, i.e. up to 14–15 accidents have to be observed, and above top and below lower–up to 7–8. But the analysis clearly demonstrates discrepancy of the theory and reality. Thus, the indicator of "go-no-go" constructed with use of dispersion in the majority of observed cases would signal about finding system in a "green" zone while the system is in "yellow". Similar conclusions follow also from the analysis of damages and losses of gas for the mentioned system.

Reduction of the planning horizon also gives nothing: before the tendency on increase will be shown, we are doomed "to miss" in the forecast in estimates a considerable part of events at the beginning of realization of every "group of accidents". Decrease in the planning horizon to quarter leads to that before the termination of the third quarter there will "be expected" almost for 50% of emergency days a year more, than are observed in reality.

Thus, it is necessary to have correct primary analysis of long-term statistics on the basis of which it is possible to speak about what share of accident of dates of emergence of non-staff situations and their scales can be eliminated. Besides as any hi-tech system changes quicker than the adequate statistics [83] is kept, and true laws of distribution of the analyzed casual processes (and, the main thing, the factors defining them) are continuously corrected, it is necessary to use criteria "free from distributions".

As criteria of achievement of goals of the forecast it is offered to take not sizes of rejections of model and real data, but the criteria used in methods of classification and recognition of images (about this it will be told further in more detail). For example, as measurement of accuracy of the forecast it is possible to use sizes of mistakes of prediction of the first and second kinds for various classes and types of emergencies, and if it is possible, depending on classes of a physical object and depending on value of parameters of any forecast background. The second circumstance is very important as, for example, it is incorrect to put statistics of accident rate of various seasons as during various seasons various technological processes proceed.

The forecast should be built on the principles well-known in theories of modeling and similarity. Steady model designs for forecasting are the found constants (also they are called cycles with the infinite period) and many self-similar structures.

First, it is “cylindrical” self-similar structures—cycles (resonant coincidence of internal natural frequencies and stimulated external frequencies), different behavior of systems on different phases of a cycle (seasonal distinctions).

Secondly, it is “conical” self-similar structures (they are fractals, Fibonacci numbers, “golden sections” of Leonardo da Vinci) [84].

Thirdly, it is self-organized dissipative structures [85] modelled by solitons [86, 87] (lonely progressing wave of accident), wavelet-functions [88] and others steady space-time exciting and braking “wave” structures (quantum as structures “steadily work” as the radio receiver, only on assigned frequencies).

7 Conclusion

The basis of modern systems for monitoring hazard and risks should be the concept of risk management, the essence of which is the formation of mechanisms, methods and tools occurrence and development in the future. At the same time, the implementation of effective prevention measures to reduce accidents and prevent non-emergency situations is the cornerstone, which can be greatly facilitated by the introduction of a risk-based approach. The development and implementation of a subsystem for the forecasting of objects of critical infrastructure safety, based on the calculation of quantitative and qualitative indicators of risks and hazard, should be carried out using the methodology of so-called early warning systems. Attention should be paid to the influence of risk factors on the system of balanced indicators of safety and risks.

In general, the risk-oriented approach encompasses both probabilistic methods of modeling, emergency processes and events, and deterministic methods. Used probabilistic and deterministic assessments. However, the experience of using in the nuclear industry is a purely probabilistic safety analysis (in fact, a one-source instrument). Therefore, the security risk of critical infrastructure objects should be viewed as a multicomponent vector. A real assessment of the level of security based on risk factors. The risk assessment is always aimed at determining its quantitative indicators, which makes it possible to use it, not only to assess the state of safety, but also to justify the economic efficiency of measures, economic calculations and reimbursements, or compensate for the lost health of the workers and the environment. The resonance of the observation and analysis of the “traffic-light type”. At the stage of solving the problem, it is necessary to establish the means for measuring the level (for example, strategic targets) and the extent to which they affect the achievement of the targets for these indicators.

Because of the discrete nature of incidents and accidents, their relatively small number, the use of data analysis machines based on the classical laws of large numbers (control of assembly operations on conveyors, casinos, insurance of minor incidents, domestic and industrial injuries, medicine, etc.) is not correct for solving this problem. It is also obvious that when the true distribution laws of analyzed random processes and, most importantly, their determining factors, will be continuously adjusted (any high-tech system, change faster than adequate statistics accumulates), it is necessary to use the criteria “free from distributions”. For example, as criteria for achieving a predictive goal, one should take the deviations of the model and the results of the data, as well as in the methods of classification and pattern recognition. For example, depending on the forecast level, which can be used in the following cases: 1st and 2nd classes for different classes and types of situations, and, if possible, depending on the class of the physical object and depending on the values of the forecasting background parameters.

That is, the introduction of a decision criterion is necessary. As such a criterion, one can use the statistical characteristic of the state being evaluated, reflecting the “distance” normalized per unit. Only based on a correct primary analysis of multi-year statistics, it will be possible to provide an opinion on the most preferable methods for an integrated assessment of the safety of critical infrastructure facilities.

The ultimate goal of monitoring threats and security risks of critical infrastructure objects is to construct a predictive model of the dynamics of situations that will allow using numerical experiments and selection of acceptable parameters to reduce the degree of uncertainty in the probabilities of events and their scales, that is, to obtain predictive information by identifying hidden patterns that indicate or a change in state of an object, or a pattern of changes of environment parameters that significantly affect s on its functioning (the so-called laws of variability “forward-looking background”).

The control monitoring facility should be organized in such a way so that you can spend time management decisions, if state of the object is approaching the danger zone. This task is divided into many subtasks, as a vertically integrated company has several decision-making centers at various levels of management. Promising in solving this problem may be methods for assessing the reliability of achieving target indicators and methods of group analysis (the latter are preferable, because they allow you to build trajectories of changes in indicators without considering the “diffusion” components and, therefore, can serve as a basic element for building up aggregates, indicators and in the future an extensive system of monitoring indicators).

References

1. Critical Infrastructure Security (2012) Assessment, prevention, detection, response. In: Flammini (ed) WIT Transactions on State-of-the-art in Science and Engineering, vol 54, 325 p. ISBN 978-1-84564-562-5
2. National Infrastructure Protection Plan (2009) U.S. Department of Homeland Security, 100 p. Available at www.DHS.gov
3. National Strategy for the Physical Protection of Critical Infrastructure and Key Assets (2003) U.S. Department of Homeland Security Available at www.DHS.gov
4. Dudenhoeffer DD, Permann MR, Manic M (2006) CIMS: a framework for infra-structure interdependency modeling and analysis. In: Perronc LF, Wieland FP, Liu J, Lawson BG,

Nicol DM, Fujimoto RM (eds) Proceedings of the 2006 winter simulation conference. Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp 478–485
5. Perrow C (1999) Normal accidents. Princeton University Press, Princeton, NJ, 450 p
6. Ramo JC (2009) The age of the unthinkable. Little, Brown & Company, New York, 280 p
7. National Research Council (2003) The internet under crisis conditions: learning from September II National Academics Press, Washington, DC. Available at www.nap.edu
8. Final Report on the August 14. 2003 Blackout in the United States and Canada: causes and recommendations, April 2004, U.S. Canada Power System Outage Task Force, p 12 (January 12, 2004). Available at https://rcports.energy.gov
9. Riabinin IA (2007) Reliability and safety of structural-complex systems. St. Petersburg, St. Petersburg University, 276 p (Russ.)
10. Tokalin VN (1965) Comments to translating book systems analysis for business and industrial problem solving. Front Cover. Stanford L. Optner. Prentice-Hall. Business & Economics pp 116, 120 (Rus.) 120 p. Available at http://www.vixri.com
11. National Infrastructure Protection Plan (NIPP) (2013) Partnering for critical infrastructure security and resilience. 57 p. Available at https://www.dhs.gov/sites/default/files/publications/national-infrastructure-protection-plan-2013-508.pdf
12. Taleb NN (2014) Antifragile: things that gain from disorder (Incerto). Series: Incerto (Book 3), Publisher: Random House Trade Paperbacks; Reprint edition (January 28, 2014), 544 p. ISBN-13:978-0812979688
13. Klykov YI (1974) Big systems situational management. « Energiya » , Moscow, 130 p
14. Zhigirev NN, Vorobiev EI, Kuzmin II, Reznichenko VYu (1983) Model of interaction of developing object with environment. Preprint Institute of Atomic Energy 3799/16, 69 p
15. Russman IB, Kaplinskiy AI, Umyvakin VM (1991) Modeling and algorithmization non-formalizable problems of the best variants of system choice. VSU, Voronezh, 168 p
16. Gheorghe AV, Masera M, Weijnen M, De Vries LJ (2006) Critical infrastructures at risk. Securing the european electric power system. Topics in safety, risk, reliability and quality, vol 9, XXV, Springer, 371 p
17. Biringer B, Vugrin E, Warren D (2013) Critical infrastructure system security and resiliency. ISBN-10: 1466557508 CRC Press (April 12, 2013), 229 p
18. Flammini F (2012) Critical infrastructure security: assessment, prevention, detection, response. ISBN-10: 1845645626 WIT Press, Computational mechanics, 1st edn (February 13, 2012), 326 p
19. Lewis TG (2006) Critical Infrastructure Protection in homeland security: defending a networked nation SBN-10: 0471786284 Wiley-Interscience (April 21, 2006), 486 p
20. Kostrov AV (1996) Interval ranking of objects by many indicators. Problemy Bezopacnosti i Chrezvychainykh Situatsiy, no. 1, pp 46–68
21. Makhutov NA, Kryshevich OV (2002) Possibility of "man-machine-environment" systems of machine-building profile based on fuzzy sets, no. 2, pp 94–103
22. Gokhman OG (1991) Expert assessment. Voronezh, Voronezh University Publishers, 152 p
23. Zinevich SV, Tarasenko VA, Usolov EB (2006) Criteria of objects ranking by emergency hazard levels. In: Proceedings of the 11th Research and Practical Conference. East-Siberian Institute of RF Ministry of Interior, Irkutsk, pp 220–223
24. Buyko KV, Pantyukhova YuV (2010) Approaches to assessment of industrial safety in organizations operating hazardous industrial facilities. No.10
25. Bruck VM, Nikolayev VI (1977) Fundamentals of general systems theory. L.: SZPI Publishers
26. Von Neumann J, Morgenstern O (2007) Theory of games and economic behavior (60th Anniversary Commemorative Edition). One of Princeton University Press's Notable Centenary Titles, 776 p
27. Larichev OI (2002) Features of decision making methods in multicriterion individual choice problems. Avtomatika i Telemekhanika. No. 2, pp 146–158
28. Gokhman OG (1991) Expert assessment. VGU Publishers, Voronezh
29. Litvak BG (2004) Expert technologies in management. M.: DELO Publishers, 400 p

30. Cox DR, Hinkley DV (1979) Theoretical statistics. Chapman and Hall/CRC, 1st edn, 528 p
31. Kuvshinov BM, Shaposhnik II, Shiryayev VI, Shiryayev OV (2002) Use of committees in pattern-classification problems with inexact expert estimations. J Comp Sys Sci Int (Izvestia Rossiiskoi Akademii Nauk. Teoriya i Upravleniya) No. 5, pp 81–88
32. Zhukovsky VI, Zhukovskaya LV (2004) Risk in multicriterion and conflicting systems in uncertainty. M.: Yeditorial URSS Publishers, 267 p
33. Melikhov AN, Berstein LS, Korovin SY (1990) Situational advising systems with fuzzy logic. M.: Nauka Publishers
34. Saaty TL (2009) Mathematical principles of decision making. Principia Mathematica Decernendi. RWS Publications, 531 p
35. Karminskiy AM, Peresetskiy AA, Petrov AY (2005) Ratings in economics: methodology and practice. M.: *Finansy i Statistika* Publishers, 240 p
36. Baranov S, Skuf'ina T (2005) Analysis of interregional differentiation and development ratings of russian federation regions. Voprosy Ekonomiki. No. 8, pp 54–75
37. Keeney RL, Raiffa H (1993) Decisions with multiple objectives—preferences and value tradeoffs. Cambridge University Press, Cambridge, 569 p
38. Nogin VD (2005) Decision making in multicriterion environment: quantitative approach. M.: Phyzmatlit Publishers, 176 p
39. Podinovskiy VV, Nogin VD (1982) Pareto optimal solutions of multicriterion problems. M.: Nauka Publishers, 256 p
40. Podinovskiy VV (2004) Analysis of solutions with multiple estimations of criteria significance factors and probabilities of indefinite factor values and objective function. Avtomatika i Telemekhanika. No 11, pp 141–158
41. Mazurov, VD (1990) Committees method in optimization and classification. – M.: Nauka Publishers, 248 p
42. Analysis for Military Decisions (1964) Edited by Quade ES. RAND Corporation, 389 p
43. Bochkov AV, Lesnykh VV, Zhigirev NN, Lavrukhin YuN (2015) Some methodical aspects of critical infrastructure protection. Saf Sci 79:229–242. doi:10.1016/j.ssci.2015.06.008
44. Moskowitz PD, Hamilton LD (1988) Energy and complex industrial system environmental emissions data reporting and acquisition; Environmental Transport and Bose-Response Models: Cuddliness for case studies, WBO/PEP/88.13, New York
45. Blocker EJ, Clifton JJ, Dryden PS (1988) Hazard identification scheme for area risk management, IAEA, TIC, Consultant's meeting
46. Dickerson MH, Gudiksen PH, Sullivan TJ, Greenly GD (1985) ARAC status report: 1985, Lawrence Livermore National Laboratory, Livermore, CA, Report UCRL-53641
47. Clark MJ, Clarke RH, Grimwood PD, Hallam J, Jones JA, Kelly GN, Linsley GS, Shaw KB, Bouville A, Despres A, Fabre H, Guezengar JM, Madelmont C, Vilquin A, Le Grand J (1979) Methodology for evaluating the radiological consequences of radioactive effluents released in normal operations. July 1979, Joint Report by the National Radiological Protection Board and the Commissariat a L'Energie Atomique
48. Atmospheric Dispersion Models for Application in Relation to Radionuclide Releases (1986) IAEA-TECDOC-379, Vienna
49. Berliand MB (1975) Modern problems of atmospheric diffusion and air pollution [Sovremennye problemy atmosfernoi diffuzii i zagriaznenia atmosfery], L., Gidmeteoizdat, 448 p (Russ.)
50. Teverovsky EN, Artemova NE et al (1985) Permissible emissions of radioactive and chemical substances into the atmosphere [Dopustimye vybrosy radioaktivnyh i himicheskih veshchestv v atmosferu], Moscow, Energoatomizdat. 240 p (Russ.)
51. Accounting for dispersion parameters of the atmosphere in the selection of sites for the production of nuclear power plants [Uchet dispersionnyh parametrov atmosfery pri vybore ploshchadok doya atomnyh ehlektrostancij], Safety Guide No. 60-SG-S3, IAEA, Vienna (1982) (Russ.)
52. James A (1981) Mathematical methods of water pollution control [Matematicheskie metody kontrolya zagryazneniya vody], Moscow, MIR, 472 p (Russ.)

53. Whittaker JD (1980) Public safety aspects of sour gas facilities. Prepared for Alberta environment pollution control division, Canada, 55 p
54. Pospelov DA (1981) Logical-linguistic models in control systems [Logiko-lingvisticheskiye modeli v sistemkh upravleniya]. Moscow: Energoizdat, 231 p (Russ.)
55. Von Neumann J (1966) Theory of self-reproducing automata. In: Burks AW (ed) University of Illinois Press, Urbana, 408 p
56. Wasserman F (1992) Neurocomputing technique: theory and practice. Translated from the English. Moscow, Mir, 240 p (Russ.)
57. Leonov AA, Leonova MA, Fedoseyev YuN (2000) Neural network synthesis aimed at identification of an object state [Sintez neyronnoy seti dly resheniya zadachi identifi catsii sostoyaniya obyekta]. In: Proceedings of the Neuroinformatics-99 All-Russia conference Moscow, MIPhI, 2000, pt. 1, pp 100–109 (Russ.)
58. Joachims T (2002) Learning to classify text using support vector machines: methods, theory, and algorithms. Kluwer
59. Tihonov AN, Tyurkisher RI (1942) Influence of the intermediate layer on vertical electrical sounding [Vliyanie promezhutochnogo sloya pri vertikal'nom ehlektricheskom zondirovanii] Izv. AN SSSR. Ser. geogr. i geofiz., №5, pp 219–227 (Russ.)
60. Jambu M, Lebeaux M-O (1978) Classification automatique pour l'analyse des données: Logiciels ISSN 0181-5938. Dunod, 399 p
61. Duran, BS, Odell PL (1974) Cluster analysis. A survey Lecture Notes in Economics and Mathematical Systems. Econometrics # 100. Managing Editors: Beckmann M, Künzi HP, Springer, Berlin, Heidelberg, New York, 138 p
62. Katkovnik VY (1985) Nonparametric identification and data smoothing. Method of local approximation (Nepapametricheskaya identifikaciya i sglajivanie dannuh. Metod lokalnoi approksimatsii), Moscow, Nauka, 336 p (Russ.)
63. Yablonskiy SV (2003) Principals of discrete mathematics: educational guidance for universities [Vvedeniye v diskretnuyu matematiky: uchebnouye posobiye dlya vuzov]. Moscow, Vysshaya shkola, 384 p (Russ.)
64. Gorbatov VA (2006) Discrete mathematics: textbook for students of universities [Diskretnaya matematika: uchebnik dlya studentov vuzov]. Moscow, Astrel (Russ.)
65. Kung SY (1988) VLSI array processors. Prentice Hall information and system sciences series. Prentice Hall International, 667 p
66. Zhuravlev YI (2002) On recognition algorithms with representative sets (about logical algorithms) [Ob algoritmakh raspoznavaniya s predstavitelnymi naborami (o logicheskikh algoritmakh)]. Zhurnal vychislitelnoy matematiki i matematicheskoy fi ziki. vol 42, no 9, pp 1425–1435. ISSN 0044-4669. (Russ.)
67. Myers GJ, Sandler C, Badgett T (2012) The art of software testing, 3rd edn. Wiley, ISBN: 978-1-118-03196-4. 256 p
68. Ryazanov VV (1991) On the optimization of class of recognition models. J Pattern Recognit Image Anal 1(1):108–118
69. Pereverzev-Orlov VS (1990) Expert Advisor. Experience in developing a partner system [Sovetchik specialista. Opyt razrabotki partnyorskoj sistemy]. M.: Nauka, 133 p (Russ.)
70. Glushkov VV, Ivanov VV, Yanenko VM et al (1980) About modeling tools of developing systems [O sredstvah modelirovaniya razvivayushchihsya system] Kiev: Izd. IK AN USSR. 58 p (Preprint/ AN USSR. In-t kibernetiki; 80–37) (Russ.)
71. Zhabotinskij AM (1974) Concentration self-oscillations [Koncentracionnye avtokolebaniya] M.: Nauka, 178 p (Russ.)
72. Nicolis G, Prigogine I (1977) Self-organization in nonequilibrium systems. From dissipative structures to order through fluctuations. Wiley, New York, 491 p
73. Reutov AP, Savchenko RG, Suslov RM (1979) The system model as a relation of generalized qualities—orderliness, reliability and efficiency [Sistemnaya model' kak otnoshenie obobshchennyh kachestv – uporyadochennosti, nadyozhnosti i ehffektivnosti] –Voprosy kibernetiki (upravlenie razvitiem sistem) / Pod red. A.P. Reutova, R.M. Suslova. M., pp 5–34 (Russ.)

74. Romanovskij YM (1981) Problems of mathematical biophysics [Problemy matematicheskoj biofiziki] – M.: Izd-vo MGU. 36 p (Russ.)
75. Romanovskij YM (1981) The processes of self-organization in physics, chemistry and biology [Processy samoorganizacii v fizike, himii i biologii] M.: Znanie, 48 p (Russ.)
76. Ganti T (1979) A theory of Biomedical supersystems and its application to problems of natural and artificial biogenesis. Budapest; Akademiai, 136 p
77. Glushkov VV et al (1977) Modeling of developing systems [Text] [Modelirovanie razvivayushchihsya sistem]: / V.M. Glushko, V.V. Ivanov, V.M. YAnenko. M.: Nauka. Glavnaya redakciya fiziko-matematicheskoj literatury, 350 p (Russ.)
78. Sarkisyan SA. (ed) (1977) The theory of forecasting and decision making [Teoriya prognozirovaniya i prinyatiya reshenij]. Uch. Posobie. Vysshaya shkola, 351 p (Russ.)
79. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics—the approach based on influence functions, Wiley, 502 p. ISBN: 978-0-471-73577-9
80. Hettmansperger TP (1984) Statistical inference based on ranks. Wiley, New York, Wiley series in probability and mathematical statistics. Probability and mathematical statistics, 323 p
81. Hodges JL Jr, Lehmann EL Estimates of location based on rank tests. Ann Math Stat 34:598–611
82. Fristedt B, Gray L (1997) A modern approach to probability theory. Birkhäuser, Boston/Basel/Berlin, ISBN 978-0-8176-3807-8, S. 165
83. Sakhal D (1985) Tekhnicheskii progress: kontseptsii, modeli, otsenki [Technical progress: concepts, models, evaluations]. Moscow, Finansy i statistika Publ. 368 p
84. LaRouche LH Jr (1984) So, you wish to learn all about economics? a text on elementary mathematical economics. New Benjamin Franklin House, New York, 192 p. ISBN 0-943235-13-8.
85. Romanovski YM, Stepanova, NV, Chernavski DS (1984) Mathematical biophysics (Matematicheskaya biofisika), Moscow, Nauka, (Russ.)
86. Shimodaira H (1996) Метод выбора аналогичных данных обучения при предсказании временных рядов с помощью нейронных сетей. A method for selecting similar learning data in the prediction of time series using neural networks. Expert Syst Appl. 10(3–4): 429–434
87. Zhigirev NN, Korzh VV, Onykiy BN (1999) Using frequency asymmetry of informational indicators for construction of automated systems for text documents' classifi cation [Ispolzovaniye asimmetrii chastotnykh svoystv informatsionnykh priznakov dlya postroye-niya avtomatizirovannykh system klassifi katsii tekstovykh dokumentov]. In: Proceedings of the Neuroinformatics-99 All-Russia conference Moscow, MIPhI, pt. 3, pp 83–91 (Russ.)
88. Ho Daniel WC (2001) Zhang Ping-An, Xu JinHua. Fuzzy wavelet networks for learning functions. IEEE Trans Fuzzy Syst 9(1):200–211

Remanufacturing: An Industrial Strategy for Economic Benefits

Adarsh Anand, Gunjan Bansal, Mohini Agarwal and Deepti Aggrawal

Abstract A successful new product development (NPD) always requires a well-defined, systematically planned, coherently organized, and cost effectively executed strategy along with the objective to satisfy customers. Supply chain management (SCM) is one conquering approach that integrates all resources for these activities into an endeavor that maximizes the positive impact on customers. Therefore, a competent supply chain management (SCM) is well-adequate tangible way to influence customers through their performance regarding time and place. Moreover, reverse logistic supply chain (RLSC) is a closed-loop chain that works as 'Double Dividend' for the company in terms of generating economic and environmental benefits. With this study, we have demonstrated that the concept of remanufacturing of products leads to more economical and sustainable benefits for the manufacturer fraternity. To explain the concept, we have proposed a methodological cost modeling framework that is based upon the diffusion of new product wherein; we have considered two different scenarios pertaining to remanufacturing phenomenon. The proposed framework has been validated on real-life sales data set pertaining to automobile industry.

Keywords Remanufacturing · Reverse logistic supply chain · Diffusion model

A. Anand · G. Bansal (✉) · M. Agarwal
Department of Operational Research, University of Delhi, Delhi 110007, India
e-mail: gunjan.1512@gmail.com

A. Anand
e-mail: adarsh.anand86@gmail.com

M. Agarwal
e-mail: mohini15oct@gmail.com

D. Aggrawal
Delhi Technological University, East Delhi Campus, Delhi 110095, India
e-mail: deepti.aggrawal@gmail.com

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*,
Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_4

1 Introduction

Globalization of markets and rapid advancements in technologies always derives companies to create innovative products and services to generate competitive advantages. The whole process and required efforts of creating a new product is called new product development (NPD). A basic process of new product development requires five generic phases [1] such as:-

I. *Opportunity Identification and Selection*: In this phase, major objective of companies is to identify current needs of the customers that they may fulfill to satisfy their customers. Among all the opportunities (new desires and needs of the customers), companies discover most significant, imperative opportunity and move to the next step.
II. *Concept Generation*: After identifying the high potential opportunity, the next step is to collect available concepts of new product, tools that fit to opportunity and start generating new form.
III. *Concept/Project Evaluation*: Here, companies evaluate new product project in terms of technical, financial, and marketing criteria. Once the project has been finalized, a blueprint of product like product definition, working team, total budget, skeleton of development plan, and final product innovation charter (PIC) is designed.
IV. *Development*: This phase consists of two subtasks: (a) technical tasks, (b) marketing tasks.
    - Technical Tasks: Here, emphasis is given to the physical design of the product including its features and benefits and also validating its quality, reliability, and other required test.
    - Marketing Tasks: Prepare strategies, tactics, and launch details for marketing plans and also work on product augmentation (service, packaging, branding, etc.)
V. *Launch:* Manage the launch program to achieve the defined goals and objective set in the final business plans.

All specified steps require proper availability of raw materials, skilled human resources, advance tools and technologies, a strong integrated structure, and the system that can produce required products. The major factors that play significant role in the success of any NPD process are better, cheaper, and faster. The conjoint inclusion of these factors implies that NPD group has to create a qualitative product at lower cost and speeding time to deliver products [2]. An effective marketing mix strategy is one that integrates with factors (better, cheaper, and faster) along with the objective to fulfill customers' demands that can occur any time. Supply chain management is a very intricate yet a very simple approach that provides a wider outlook to convert an opportunity into profits at best be approached.

Supply chain management can be defined as 'an integrated, synchronized, and a closely knitted chain which link all the supply interacting organization in a two-way communication system in order to maintain a high quality of inventory in the most effective manner' [3–7]. In 1950s and 1960s, NPD was slow in terms of slow production (due to delay in receiving raw materials, lack of resources like workers, machineries) and delivering of final products to customers (due to lack of efficient transports and services). But, in today's competitive environment to speed up the NPD process, supply chain management (SCM) has become a broader marketing mix strategy for the companies [8, 9]. In the literature, positive effects have been analyzed on the relationship of NPD outcomes after incorporating the suppliers' involvement that reduces uncertainties like from producing goods and services to the delivery of products on time [10, 11]. Different companies follow different network designs of supply chains as per their convenience and requirements. These different supply chains can be explained as follows:-

(1) **Forward Supply Chain**: Traditional supply chain system that can be defined as 'system whose constituent parts include material suppliers, production facilities, distribution services, and customers linked together by the feed forward flow of materials and feedback flow of information' [12].
(2) **Reverse Supply Chain**: This system includes 'a reverse chain of collecting returned goods or refused goods which then passes through sorters for reprocessing (reuse, recycle, recondition, remanufacturing, refurbishing, and asset recovery) or to disposal' [3–5].
(3) **Green Supply Chain**: The system that integrates environmental thinking in terms of reducing health and environment impacts into supply chain management including product design, material sourcing and selection, manufacturing processes, delivery of the final product, and end-of-life management of the product after its useful life [13].
(4) **Sustainable Supply Chain**: Sustainable supply chain is responsible for considering three major aspects such as environmental, social, and financial into supply chain management. Or in other words, sustainable supply chain is a concept of embedded economic, environmental and societal decision in supply chains at design time [14, 15].

Figure 1, demonstrates the basic flow of a closed-loop supply chain management that includes forward and reverse logistic together, and incorporation of other network can also be utilized.

Over the years, companies have realized the strategic importance of planning, controlling, and designing as a whole. Therefore, supply chain management can be said as an upcoming evaluation stage that captures the synergy of inter-functional and interorganizational integration and coordination among suppliers and end users and to optimize the system as whole [16, 17]. Here, in this study we have emphasized on the concept of reverse logistic supply chain (RLSC), where the objective of using this strategy by manufactures is to reutilize the used products that have been stopped using by customers. Rest of the chapter has been structured as follows:-In next section (i.e., Sect. 2), we have explained how remanufacturing of products acts as an interface between generating profit for the companies and

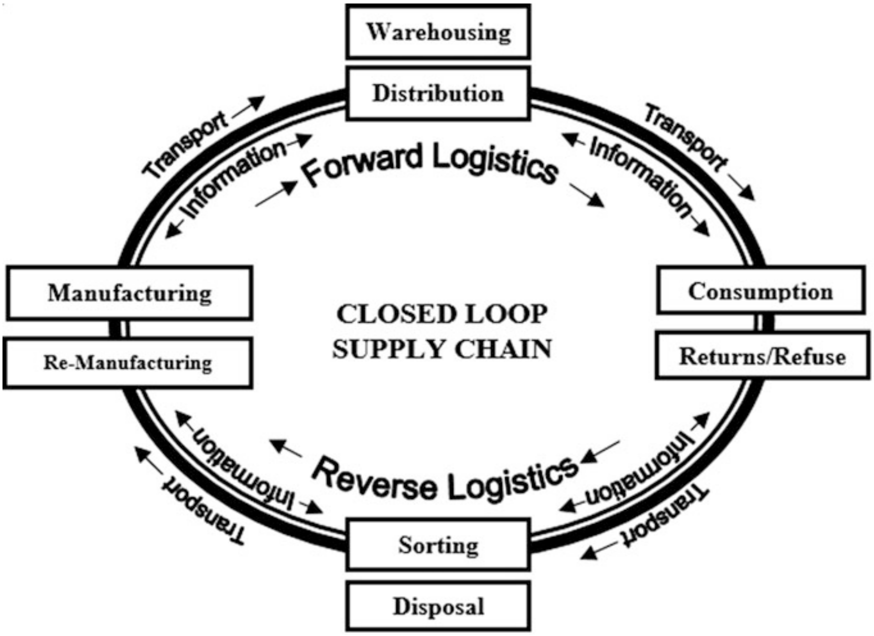


Fig. 1 Closed-loop supply chain. *Source* [3–5]

satisfying environmental issues. Further in Sect. 3, we have proposed our cost modeling by using the remanufacturing components into consideration. Section 4 demonstrates the validation of our proposition by a numerical illustration. Section 5 explains how this study can be useful for the manufactures, managers in adopting effective strategies for their companies. Further, this chapter has been concluded in the Sect. 6.

2 Remanufacturing as a Marketing Strategy

"A circular economy is one that is restorative and regenerative by design, and which aims to keep products, components, and materials at their highest utility and value at all times, distinguishing between technical and biological cycles" defined by Ellen MacArthur Foundation. Hence, it implies that circular economy provides efforts to minimize energy consumption, waste disposal, and preserving natural resources by employing a systematic approach of activities like reuse, sharing, repair, pay-per-use, refurbishment, remanufacturing, and recycling of products. Product remanufacturing plays a very significant role in circular economy. This term sometimes confused with other terms like refurbishing, reuse, repairing, reconditioning but they have different meanings. According to Center for

Remanufacturing and Reuse (CRR), remanufacturing can be defined as 'a process of returning a used product to at least its original performance with a warranty that is equivalent to that of the newly manufactured product' [18]. By using the strategy of remanufacturing of products, it helps companies to pursue their sustainable development goals. In other words, it can act as a 'Double Dividend' policy for the companies that not only generate more profits but also encouraging in achieving environmental benefits. Profits can be visualized in terms of creating new jobs, higher profit margins, breeding ground for problem-solving skills, embrace new manufacturing techniques, and better building relationships with their customers. Similarly, remanufacturing can be advantageous for environment that can be envision as first of all it reduces consumption of critical raw material specially when supply risk is present, it saves energy by consuming less resources as compared to new ones, reduces harmful gaseous emissions, and moreover sends lesser waste to landfill by consuming useful working materials. [19] have considered three different ways of collecting used products from the end users such as (i) manufactures can directly collect from the customers, (ii) can induce used products through retailers (or distributors), and (iii) by outsourcing to the third party, where they found that alternative (2) is most cost-effective option as compared to rest alternatives. A reverse logistic supply chain can design product in three different quality levels, i.e., (i) when all new components have been utilized, (ii) when mixture of new and recovered components have been considered, and (iii) when only used components/modules have been taken into consideration [20]. Different networks have been designed using the concept of reverse logistic supply chain such as reusability assessment of components has been analyzed using reliability measures for consumer products [21], a practical relevance model has been proposed that considered multi-period, components' facilities and its expansion facilities with profit-oriented objective function [22], optimal production inventory approach has been modeled [23]. Also, RLSC model has been designed using carbon footprint taken into consideration [24, 25].

New product development (NPD) and product life cycle (PLC) are the key marketing concepts that are directly dependent on the performance of the designed network of the supply chain. At every stage of NPD and PLC such as introduction, growth, maturity, and decline [26], logistic works as an encouraging hand that increases the acceptance of the product in the marketplace. Hence, integration of marketing and supply chain framework enhances the competitive advantages not only in terms of product and price but also supply chain services improved satisfaction level of customers. And, when the involvement of remanufacturing has also been added with the NPD and PLC, then it becomes a very cost-effective production strategy [9, 27]. Hence, reverse logistic acts as a production stratagem that recovers the residual value of used products [28]. Here, this chapter has utilized the concept of reverse logistic supply chain as a marketing strategy that helps in diffusing the product into the market in cost-effective manner. Further, mathematical cost modeling has been discussed where we have defined different modeling in different circumstances.

3 Mathematical Cost Modeling

The core objective behind the efficient supply chain is to satisfy customers' demand on time. Or in other words, maximization of customers' satisfaction at effectual cost is the main aspiration for any organization. Demand of the customers can be satisfied by either new product or commercial returns (after light repairs). And, it is worth to observe that the quantity of the manufactured product not only depends upon the demand of the product but also reliant on the commercial returns because they can also be considered as new product after light repairs [27]. In the presence of product diffusion, an optimization modeling framework for maximizing profits using remanufacturing concept has been proposed [29] and also, when to start collection of used product from the customers has been identified [30].

Here, this study quantifies the cost-saving strategy that defines a closed-loop system, i.e., conjointly employing forward and reverse logistic which is based upon the diffusion of the new product and its returns. The notations and assumptions for proposed modeling framework have been defined as follows:

Notations:

C_{fixed}	Fixed setup cost
$C_{\text{production}}$	Per unit production cost
$C_{\text{inventory}}$	Per unit inventory carrying cost
$C_{\text{collection}}$	Per unit cost of colleting used product
C_{remanuf}	Per unit cost of remanufacturing a new product ($C_{\text{remanf}} < C_{\text{production}}$)
C_{disposal}	Per unit cost of disposing waste
$C_{\text{component}}$	Per unit cost of holding components
Q	Quantity produced at time 't'
$D(t)$	Demand of the customers at time 't'
$N(t)$	Cumulative numbers of adoptions by time 't'
$n(t)$	Numbers of adopters at time 't'
r	Fraction of used products that can be remanufactured
c	Collection rate of the used products ($0 < c < 1$)
Pr	Price of the product
p	Coefficient of innovation
q	Coefficient of imitation
m	Potential initial market size
a	Minimum numbers of quantity of products produced
b	Fraction of units that have to be produced as per demand

Assumptions:

(a) The cost of remanufacturing a new product is less than manufacturing a totally new product.
(b) Only one major component of the product that pertains most valuable and considered one of the major driver is to be used for remanufacturing a new product.

(c) Remanufactured product is regarded as a new product and can be sold in primary/secondary market.
(d) No stock outs are allowed.
(e) Produced quantity is always greater than demand, or in other words demand is always fulfilled.

3.1 Demand Function

Companies frequently change their strategies, products, markets as per changing time, and demands of the customers. Therefore, the life cycle of the product has a certain life that consists of different stages such as introduction, growth, maturity, and decline [26]. Different products follow different patterns of the product life cycle (PLC); for example, *growth-slump-maturity* pattern follows by kitchen appliances, *cycle-recycle* pattern often describes by the sales of the new drugs, *scalloped* pattern follows by those products that can have many new uses identified such as Nylon's sales. Here in this study, we have considered the classical pattern of PLC that follows a hump-shaped curve that can be demonstrated in Fig. 2.

Demand of the customers signifies how rapidly product is being diffused in the marketplace. The potential adopters of new product can be classified on the basis of the mode from which they have been informed about the product such as: *innovators* and *imitators*. The individuals who are influenced by external sources like mass media are referred as innovators. And, those potential adopters who are informed by word-of-mouth are known as imitators. The classical model called Bass model [31] is the best representative of this phenomenon. The phenomenon of diffusion of new product has been widely studied by many researchers, to obtain the number of adoptions or the sales of new product [32–35], which can be utilized to determine the demand or the number of units that can be manufactured. Therefore, in this study, demand function follows the Bass model that has been defined below:-

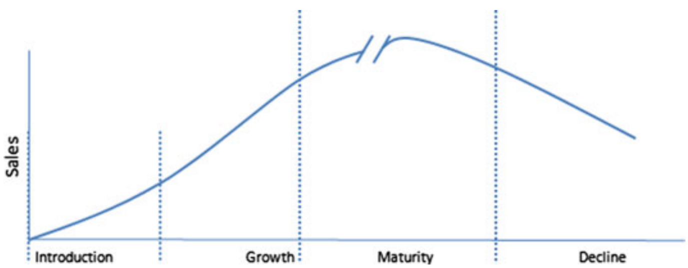


Fig. 2 Classic product life cycle

The diffusion process can be represented by using differential equation as:-

$$\frac{dN(t)}{dt} = \left[p + q\frac{N(t)}{m}\right][m - N(t)]. \tag{1}$$

Under the condition that at $t = 0, N(t) = 0$, the cumulative and noncumulative numbers of adopters can be shown as:-

$$N(t) = m\left(\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}\right) \tag{2}$$

$$n(t) = \frac{m(p+q)^2}{p}\frac{e^{-(p+q)t}}{\left[(q/p)e^{-(p+q)t} + 1\right]^2}. \tag{3}$$

Here, demand is considered equivalent to $n(t)$, i.e., $D(t) \simeq n(t)$ as we are focusing the quantity produced and deliver at time 't.'

3.2 Cost Function

This economic model of remanufacturing of products considers four different groups of activities, namely [36]:

I. **Manufacturing of the Products**:

This stage includes the processing of sourcing and processing of raw materials, producing goods, ensuring its quality checks, and marketing the products to end-customers and fulfilling the demands of them. This phase includes the costs that occur, while manufacturing products are like fixed cost (based on manpower, machinery, rent, etc.), production cost, inventory cost.

II. **Collection of Used Products**:

In this phase, different modes such as curb side collection, drop-off centers, and deposit or refund programs are being used to collect used products from the customers. The collection of returned product would be c% of the all sold products to the users even after products reach to their end of useful life because of the following reasons collection networks might not cover all required area, user might nor return the product or may be sold to third party, etc.

III. **Remanufacturing of the Product that Acted as a New Product**:

Once, the some fraction of used products have been collected by the companies then all used products have to pass through quality checks which identify the components of used products that can be used in remanufacturing process and can act as a new product. Therefore, among the 'c' proportion of used products,

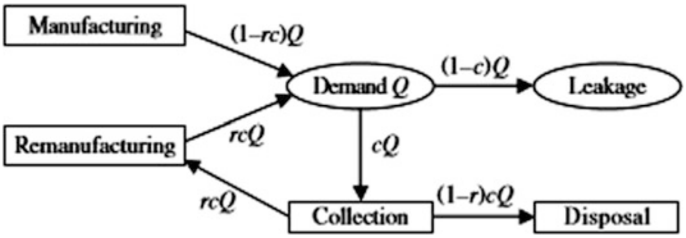


Fig. 3 Flow of product in a closed-loop supply chain. *Source* [36]

some 'r' fraction of collected products can be taken under the process of remanufacturing. Since, only one component is being used for creating a new product, therefore one component that has passed all required quality check is able to become a one new product. Further as stated by [29], the flow of returned goods might follow the Bass model pattern with different parameters in comparison with the new product adoption. Moreover, in this study we have not highlighted on this pattern of returned goods and solely focused on the profit obtained inculcating the concept of remanufacturing.

IV. **Disposal of the Waste**:

The used products that could not qualify the quality checks for remanufacturing process or declared as not useful for any processing are being disposed off. Disposing of waste may also incur some cost for the manufactures.

Further, we have defined two cost structures where in the case (1), it demonstrates the general scenario of product system, i.e., basic cost components have been used such as manufacturing cost and the cost incur in applying reverse logistic for producing products. Moreover, in case (2), a special cost component has also been taken into account that represents the holding cost of one main component that plays significant role in the product and can be used for remanufacturing process because of its long life. The basic objective of company is to maximize profits using various strategies. Here, we are comparing between two cost structures in each scenario, i.e., when only new products are being manufactured using all new components and; when the concept of remanufacturing has been used and one component of used product is required for making new product. The flow of products using RLSC can be represented diagrammatically (Fig. 3) as:-

Case 3.2.1: Cost Function: General Scenario (When Component Holding Cost Is Relaxed)

This scenario is the general scenario used by many production systems. The quantity produced by companies is the function of $n(t)$ that is $Q = a + b.n(t)$ or $Q = a + b.D(t)$, where a is the minimum numbers of products that a manufacturer has to produce for its customers and b is the coefficient of $n(t)$ or $D(t)$.

Therefore, it can be said that produced quantity is dependent on the demand of the customers over time. Cost structures of manufacturing using only new components and mixed manufacturing using used component are explained as:

(1) *When manufacturing is done using all new components*

$$\text{Cost} = C_{\text{fixed}} + C_{\text{production}} \cdot Q + C_{\text{inventory}} \cdot (Q - D(t)). \tag{4}$$

Here, first component, i.e., C_{fixed} is the total fixed cost associated in the production; second component is the total cost of producing Q units of products; third component is representing the inventory cost of holding finished goods when the produced quantity is higher than the demand, i.e., $Q > D(t)$. So the remaining units $(Q - D(t))$ would become inventory that can be utilized whenever demand is high.

(2) *When remanufacturing has taken into consideration*

$$\begin{aligned}\text{Cost} = {} & C_{\text{fixed}} + C_{\text{production}} \cdot Q \cdot (1 - r.c) + C_{\text{collection}} \cdot c.Q + C_{\text{remanf}} \cdot r.c.Q + C_{\text{disposal}} \cdot c.Q \cdot (1 - r) \\ & + C_{\text{inventory}}(Q - D(t)).\end{aligned} \tag{5}$$

Under this scenario, remanufacturing of used product has also been incorporated; therefore, production system may occur cost but in different aspects. This equation can be understood as the first and last component are the fixed cost and inventory cost as discussed above, but the production cost of producing Q units would be the mix of manufactured and remanufactured products, i.e., $C_{\text{production}} \cdot Q \cdot (1 - r.c) + C_{\text{remanf}} \cdot r.c.Q$, where $C_{\text{production}} > C_{\text{remanf}}$ and $r.c.Q$ are the quantity of units produced using used components. Third and fifth components are representing the cost of $c\%$ goods collected from the end users and disposing off the waste which could not qualify the quality checks, i.e., $(c.Q.(1 - r))$.

Case 3.2.2: Cost Function: Special Case (When Component Holding Cost Is Considered)

In order to minimize the impact on environment, manufacturers need to consider the importance of recovery of used objects so as to minimize the industrial waste. Objects can be product itself or the components used in a product. Component is an internal part or an element of a product that associated with value. Components itself have a vital role in improving the value and quality of the products. Therefore, original equipment manufacturers (OEMs) are stepping into remanufacturing activities via recovery of the components [37]. Here, this case represents a state when some manufactures hold stock of component that has vital role in producing products. And, the same component can be used for remanufacturing new products. The cost elements of the defined cost structure are similar as discussed (in case 1) except the factor $C_{\text{component}}.Q$. This factor implies that production system has to keep at least Q units of components so that Q units of products can be used. Therefore, the total cost of holding Q units of components is equal to $C_{\text{component}}.Q$, where $C_{\text{component}}$ is the cost of holding one component. But under mixed manufacturing and remanufacturing strategy, since $r.c.Q$ units of components can be reused for remanufacturing, therefore manufactures just need to keep $Q.(1-r.c)$ units of components in stock. Equations have been discussed as follows:-

(1) *When manufacturing is done using all new components*

$$\text{Cost} = C_{\text{fixed}} + C_{\text{production}}.Q + C_{\text{inventory}}.(Q - D(t)) + C_{\text{component}}.Q. \tag{6}$$

(2) *When remanufacturing has taken into consideration*

$$\begin{aligned}\text{Cost} = {} & C_{\text{fixed}} + C_{\text{production}}.Q.(1-r.c) + C_{\text{collection}}.c.Q + C_{\text{remanf}}.r.c.Q \\ & + C_{\text{disposal}}.c.Q.(1-r) + C_{\text{inventory}}(Q - D(t)) + C_{\text{component}}.Q.(1-r.c).\end{aligned} \tag{7}$$

3.3 *Profit Function*

In general, term ‘profit’ can be defined as the difference between the amount that has been earned and the amount that has been spent in buying/producing/operation [38, 39]. Here, in this study we have considered about ‘economic profit’ which implies the difference between total revenue generated by the firm and all cost incurred in manufacturing/remanufacturing the products. This can be written as:-

$$\text{Profit} = \varphi(\text{D(t)}) = \text{Revenue} - \text{Cost} \tag{8}$$

where
Revenue = $\text{Pr} * D(t)$.

Here, we are comparing two different profit functions (for both defined cases) such as: when only new components are utilized for manufacturing a product and when remanufacturing concept has been utilized, i.e., mixed (new and used) component has taken into consideration while manufacturing products. Therefore, using Eqs. 4 and 5 (for case 1) and Eqs. 6 and 7 (for case 2), we can formulate the profit maximization problem that considers all cost factors like fixed cost, manufacturing (or remanufacturing) cost, inventory cost, and component holding cost and optimizes the maximum profits and its corresponding time point.

Case 3.3.1: Problem Formulation: General Scenario (When Component Holding Cost Is Relaxed)

1. *When manufacturing is done using all new components* **(P.1.1)**

$$\text{Max } = \varphi(D(t)) = \Pr * D(t) - C_{\text{fixed}} + C_{\text{production}} \cdot Q + C_{\text{inventory}} \cdot (Q - D(t)),$$

subject to,

$$Q \geq D(t),$$

$$Q \geq 0,$$

$$D(t) \geq 0.$$

2. *When remanufacturing has taken into consideration* **(P.1.2)**

$$\begin{aligned}\text{Max} = \varphi(D(t)) = \Pr * D(t) - C_{\text{fixed}} + C_{\text{production}} \cdot Q.(1 - r.c) + C_{\text{collection}} \cdot c.Q \\ + C_{\text{remanf}} \cdot r.c.Q + C_{\text{disposal}} \cdot c.Q.(1 - r) + C_{\text{inventory}} (Q - D(t))\end{aligned}$$

subject to,

$$Q \geq D(t),$$

$$Q \geq 0,$$

$$D(t) \geq 0.$$

Case 3.3.2: Problem Formulation: Special Case (When Component Holding Cost Is Considered)

1. *When manufacturing is done using all new components* **(P.2.1)**

$$\begin{aligned}\text{Max} &= \varphi(D(t)) \\ &= \Pr * D(t) - C_{\text{fixed}} + C_{\text{production}} \cdot Q + C_{\text{inventory}} \cdot (Q - D(t)) + C_{\text{component}} \cdot Q,\end{aligned}$$

subject to,

$$Q \geq D(t),$$

$$Q \geq 0,$$

$$D(t) \geq 0.$$

2. *When remanufacturing has taken into consideration* **(P.2.2)**

$$\begin{aligned}\text{Max} = \varphi(D(t)) = \Pr * D(t) - C_{\text{fixed}} + C_{\text{production}}.Q.(1 - r.c) \\ + C_{\text{collection}}.c.Q + C_{\text{remanf}}.r.c.Q + C_{\text{disposal}}.c.Q.(1 - r) \\ + C_{\text{inventory}}(Q - D(t)) + C_{\text{component}}.Q.(1 - r.c),\end{aligned}$$

subject to,

$$Q \geq D(t),$$

$$Q \geq 0,$$

$$D(t) \geq 0.$$

4 Numerical Illustration

For validating our proposition, we have used Hyundai car sales data [40, 41]. The objective behind considering data related to cars is as follows: This product has many components (such as engine oil, oil filters, auto glasses, engines and transmissions, scrap metal, batteries, tires, water pumps, starters and alternators, belts, plastic, rubber hoses, and mat and carpets) that can be used in manufacturing a new car after repairing and refurbishing process [42]. The validation has been done in two steps. Initially using car sales data, we have fitted classical Bass model [31] by using nonlinear least square (NLLS) [43] and solved it in SAS software package [44]. The accuracy of fitting Bass model, values of actual car sales data, and predicted values can be seen graphically in Fig. 4. Also, the values of goodness of fit criteria: R^2 = 0.9938 and $AdjR^2$ = 0.9935 shows the accuracy of the prediction captured by the Bass Model. Therefore, we have taken coefficient of innovator ($p = 0.0125$) and imitator ($q = 0.201$) and expected market size ($m = 256667.2$) of adopters as initial inputs for the purpose of solving optimization problems (such that P.1.1, P.1.2, P.2.1, and P.2.2).

In the second step, we have identified the optimal time point of all defined optimization problems in previous section and found out the maximum profit for each case. We have made use of a mathematical calculator; MAPLE, that is being widely used for solving mathematical and optimization problems. Table 1 shows the assumed values which have been used for calculating the profits for the optimization problems (such that P.1.1, P.1.2, P.2.1, and P.2.2).

Using the parameter values as obtained for p, q and m and the cost parameter values, we find out the maximum profit for all four profit maximization problems. It is to observe that the maximum profit has been obtained at point ($t = 13$) but with

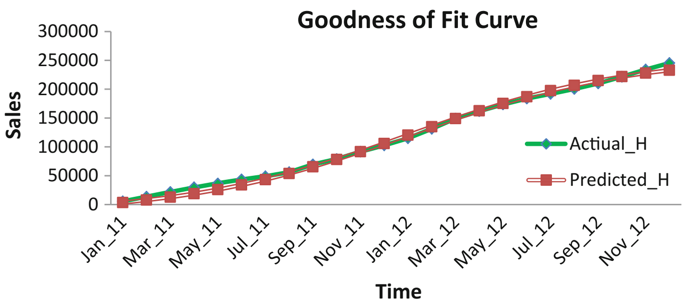


Fig. 4 Goodness of fit curve for car sales data

Table 1 Parameter values used in profit function

	Case 1: General scenario		Case 2: Special case	
	Manufacturing	Remanufacturing	Manufacturing	Remanufacturing
a	3000	3000	3000	3000
b	0.6	0.6	0.6	0.6
C_{fixed}	\$5	\$5	\$5	\$5
$C_{\text{production}}$	\$2	\$2	\$2	\$2
$C_{\text{inventory}}$	\$0.3	\$0.3	\$0.3	\$0.3
$C_{\text{collection}}$		\$0.1		\$0.1
C_{remanuf}		\$0.2		\$0.2
C_{disposal}		\$0.05		\$0.05
$C_{\text{component}}$				\$0.03
r		0.5		0.5
c		0.6		0.6
Pr	\$8	\$8	\$8	\$8

different values (due to variation in the cost components). Consider the time point ($t = 13$), when only new components have been used for manufacturing purpose; the maximum profit values are \$93,819.26 and \$93,467.26 in both the cases (i.e., P.1.1 and P.2.1), whereas when concept of remanufacturing has been employed as a economical strategy and a mixture of both; old and new components have been used for manufacturing purpose; the maximum profit obtained is \$99,275.25 and \$99,028.85, respectively, for case P.1.2 and P.2.2 which clearly shows an increment of \$5455.99 and \$5561.59, respectively. Thereby, it can be clearly seen that employing the concept of remanufacturing for goods production is beneficial for the firm.

Further, it can be seen that based on our assumptions and proposed modeling framework, the concept of remanufacturing talks about two different scenarios for understanding the cost structure wherein; the profit so obtained in special case (when basic cost factors like fixed cost, production cost, and inventory carrying cost of finished goods are accompanied by cost of holding components' inventory) is quite significant in comparison with the general case when the component holding cost is not considered in structuring the overall cost.

Profit values for all the proposed optimization problems at different time points have been presented in Table 2. And, it can be notified that for case 3.3.1 and case 3.3.2, utilization of remanufacturing concept in supply chain management (or closed-loop supply chain) provides more profits as compared to only using forward supply chain strategy.

Table 2 Net profit values

Time	Case 1: General scenario		Case 2: Special case	
	Manufacturing	Remanufacturing	Manufacturing	Remanufacturing
1	19,891.59154	22,366.97435	19,731.889	22,255.182
2	25,167.01473	27,855.09163	24,993.59	27,733.694
3	31,202.82369	34,134.25184	31,013.699	34,001.864
4	37,985.04934	41,189.92273	37,778.283	41,045.186
5	45,439.13048	48,944.53689	45,212.975	48,786.228
6	53,411.53949	57,238.37685	53,164.646	57,065.551
7	61,655.46133	65,814.67647	61,387.124	65,626.84
8	69,826.27573	74,314.9211	69,536.685	74,112.208
9	77,493.13845	82,290.89577	77,183.605	82,074.222
10	84,171.31537	89,238.32288	83,844.411	89,009.49
11	89,375.368	94,652.1921	89,034.927	94,413.883
12	92,686.35939	98,096.6758	92,337.306	97,852.338
13	93,819.2655	99,275.2583	93,467.26	99,028.858
14	92,673.2667	98,083.0552	92,324.248	97,838.742
15	89,350.33804	94,626.153	89,009.962	94,387.89
16	84,136.44803	89,202.04976	83,809.635	88,973.28
17	77,451.12091	82,247.18416	77,141.697	82,030.587
18	69,779.97596	74,266.75461	69,490.506	74,064.125
19	61,607.57037	65,764.85464	61,339.358	65,577.106
20	53,364.321	57,189.25461	53,117.551	57,016.515
21	45,394.29707	48,897.89589	45,168.258	48,739.668
22	37,943.74771	41,146.95589	37,737.089	41,002.294
23	31,165.69608	34,095.62732	30,976.668	33,963.307
24	25,134.29806	27,821.05588	24,960.958	27,699.718

5 Managerial Implication

In such high competitive and globalized environment, it becomes imperative for managers to infuse all activities into such a well defined and integrated system that would be benefited for the companies along with the society. Reverse logistic supply chain strategy has gained higher acceptance over the past years as compared to forward supply chain approach, where forward supply chain only focuses and works for delivering the right product to the right customers on the right time and ends the process at this stage. Closed-loop supply chain (or RLSC) not only includes these activities of traditional supply chain but also intends to capture additional values for manufactures. Therefore, RLSC (or remanufacturing scheme) is considered as an effective value-added service for both the parties (manufactures and customers). Moreover, the greatest contribution of utilizing remanufacturing process is to keep nonrenewable resources in circulation for multiple lifetimes. In other words, the end-of-life components of returned goods can be converted into same-as-new condition for manufacturing new products by using this approach.

Therefore, the objective of this study is to reinforce making use of this opportunity called remanufacturing or a closed-loop supply chain (or reverse logistics) as a fortify service that can benefit the whole system. We have considered that new products that have limited life cycle can be renewed and work as totally new product. The proposed cost structures have exhibited the significance of remanufacturing in terms of gaining profits and reducing wastes over using simple forwards supply chain management (in both the cases, i.e., general scenario and special case). Also in the special case, we have analyzed that keeping major components as stock in advance may incur some additional cost but if manufactures are using remanufacturing scheme then this would lead them in gaining higher profits. Hence, it can be said that new management requires minimizing the waste by converting old components (or used products) that failed or badly damaged or worn out into new and reusable form so that net profits can be maximized and product may retain their identity; so that satisfaction level of the customers and social gains can be obtained by companies.

6 Conclusion

This study is an endeavor to represent the significance of employing remanufacturing and reverse logistic supply chain (RLSC) under the constraint of finite product life cycles as an essential strategy to make supply chain operations more efficient of new product development. Here, we have proposed cost modeling of two different scenarios, i.e., (1) general case: when basic cost factors have been considered and; (2) special case: when component holding cost has also been measured with other basic cost factors. Under both scenarios, profit-based comparison on car sales data has been made between when only new components have been used for manufacturing new products and when mixed (new as well as used) components have been utilized for producing new products. And, the result has demonstrated that remanufacturing process drives more sustainability and organizational benefits to the product units. Furthermore, it is being observed that though keeping the component holding cost incur some additional cost but the remanufacturing policy can convert this loss into profit. Hence, it can be stated that reverse logistic supply chain (RLSC) not only alleviate depletion of recourses but also helpful in reducing environmental causes in terms of global warming and toxic materials.

Acknowledgements The research work presented in this chapter is supported by grants to the first author from Department of Science and Technology (DST) via DST PURSE phase II grant, India.

References

1. Crawford M, Benedetto AD (2006) New product management, 8th edn. Boston, McGraw Hill
2. Ragatz GL, Handfield RB, Scannell TV (1997) Success factors for integrating suppliers into new product development. J Prod Innov Manage 14(3):190–202
3. Bloomberg DJ, Lemay S, Hanna JB (2002) Logistics. Prentice Hall, Upper Saddle River
4. Sahay BS (ed) (2000) Supply chain management in the twenty-first century. Macmillan, New Delhi
5. Simchi-Levi D, Kaminsky P, Simchi-Levi E (2000) Designing and managing the supply chain: concepts, strategies, and case studies. Irwin McGraw-Hill, Singapore
6. Beamon BM (1998) Supply chain design and analysis: Models and methods. Int J Prod Econ 55(3):281–294
7. Calantone R, Dröge C (1999) Supply chain flexibility: an empirical study. J Supply Chain Manage 35(2):16–24
8. Tan KC (2001) A framework of supply chain management literature. Eur J Purch Supply Manage 7(1):39–48
9. Petersen KJ, Handfield RB, Ragatz GL (2005) Supplier integration into new product development: coordinating product, process and supply chain design. J Operat Manage 23 (3):371–388
10. Primo MA, Amundson SD (2002) An exploratory study of the effects of supplier relationships on new product development outcomes. J Operat Manage 20(1):33–52
11. Ragatz GL, Handfield RB, Petersen KJ (2002) Benefits associated with supplier integration into new product development under conditions of technology uncertainty. J Bus Res 55 (5):389–400
12. Stevens GC (1989) Integrating the supply chain. Int J Phys Distrib Mater Manage 19(8):3–8
13. Al-Zu'bi H (2014) What is difference between green supply chain management and sustainable supply chain management? https://www.researchgate.net/post/What_is_difference_between_green_supply_chain_management_and_sustainable_supply_chain_management, Accessed 25 March 2017
14. Neto JQF, Bloemhof-Ruwaard JM, Van Nunen JA, van Heck E (2008) Designing and evaluating sustainable logistics networks. Int J Prod Econ 111(2):195–208
15. Chaabane A, Ramudhin A, Paquet M (2012) Design of sustainable supply chains under the emission trading scheme. Int J Prod Econ 135(1):37–49
16. Min H, Zhou G (2002) Supply chain modeling: past, present and future. Comput Ind Eng 43 (1):231–249

17. Spekman RE, Spear J, Kamauff J (2002) Supply chain competency: learning as a key component. Supply Chain Manage: An Int J 7(1):41–55
18. LeBlanc R (2016) Remanufacturing in the circular economy. https://www.thebalance.com/remanufacturing-in-the-circular-economy-2878056, Accessed 25 March 2017
19. Savaskan RC, Bhattacharya S, Van Wassenhove LN (2004) Closed-loop supply chain models with product remanufacturing. Manage Sci 50(2):239–252
20. Das K, Chowdhury AH (2012) Designing a reverse logistics network for optimal collection, recovery and quality-based product-mix planning. Int J Prod Econ 135(1):209–221
21. Mazhar MI, Kara S, Kaebernick H (2005, February) Reusability assessment of components in consumer products—a statistical and condition monitoring data analysis strategy. In 4th Australian LCA conference, Sydney, pp 1–8
22. Alumur SA, Nickel S, Saldanha-da-Gama F, Verter V (2012) Multi-period reverse logistics network design. Eur J Oper Res 220(1):67–78
23. Bazan E, Jaber MY, Zanoni S (2016) A review of mathematical inventory models for reverse logistics and the future of its modeling: an environmental perspective. Appl Math Model 40 (5):4151–4178
24. Kannan D, Diabat A, Alrefaei M, Govindan K, Yong G (2012) A carbon footprint based reverse logistics network design model. Resour Conserv Recycl 67:75–79
25. Diabat A, Abdallah T, Al-Refaie A, Svetinovic D, Govindan K (2013) Strategic closed-loop facility location problem with carbon market trading. IEEE Trans Eng Manage 60(2):398–408
26. Rogers EM (2010) Diffusion of innovations. Simon and Schuster
27. Amin SH, Zhang G (2012) A proposed mathematical model for closed-loop network configuration based on product life cycle. Int J Adv Manuf Technol 58(5–8):791–801
28. Atasu A, Sarvary M, Van Wassenhove LN (2008) Remanufacturing as a marketing strategy. Manage Sci 54(10):1731–1746
29. Kapur PK, Sachdeva N, Singh O (2017) Optimal profit for manufacturers in product remanufacturing diffusion dynamics. J Ind Prod Eng, Taylor & Francis, doi:10.1080/21681015.2017.1295405
30. Sachdeva N, Kapur PK, Singh O (2017) When to start remanufacturing using adopter categorization. Accepted for publication in Quality, IT and Business Operations: Modeling and Optimization, Springer
31. Bass FM (1969) A new product growth for model consumer durables. Manage Sci 15(5):215–227
32. Singh O, Anand A, Kapur PK, Aggrawal D (2012) Consumer behaviour-based innovation diffusion modelling using stochastic differential equation incorporating change in adoption rate. Int J Technol Mark 7(4):346–360
33. Anand A, Singh O, Aggarwal R, Aggrawal D (2016) Diffusion modeling based on customer's review and product satisfaction. Int J Technol Diffus (IJTD) 7(1):20–31
34. Anand A, Agarwal M, Aggrawal D, Singh O (2016) Unified approach for modeling innovation adoption and optimal model selection for the diffusion process. J Adv Manage Res 13(2):154–178
35. Anand A, Bansal G (2016) Predicting customer's satisfaction (dissatisfaction) using logistic regression. Int J Math Eng Manage Sci 1(2):75–86
36. Geyer R, Van Wassenhove LN, Atasu A (2007) The economics of remanufacturing under limited component durability and finite product life cycles. Manage Sci 53(1):88–100
37. Han SH, Dong MY, Lu SX, Leung SC, Lim MK (2013) Production planning for hybrid remanufacturing and manufacturing system with component recovery. J Oper Res Soc 64 (10):1447–1460
38. Aggrawal D, Anand A, Singh O, Singh J (2014) Profit maximization by virtue of price & warranty length optimization. J High Technol Manage Res 25(1):1–8
39. Kapur PK, Anand A, Sachdeva N (2014) Profit estimation for a product under warranty: an evaluation based on customer's expectation and satisfaction. Int J Reliab Qual Saf Eng 21 (06):1450029
40. Team-bhp, http://www.team-bhp.com/. Accessed 03 July 2015

41. Motorbeam, http://www.motorbeam.com/. Accessed 03 July 2015
42. MNN (2017) 13 surprising car parts that can be recycled and reused. http://www.mnn.com/lifestyle/eco-tourism/sponsorstory/13-surprising-car-parts-that-can-be-recycled-and-reused. Accessed 11 April 2017
43. Srinivasan V, Mason CH (1986) Technical note-nonlinear least squares estimation of new product diffusion models. Mark Sci 5(2):169–178
44. SAS Institute Inc (2004) SAS/ETS User's Guide Version 9.1. SAS Institute Inc., Cary, NC

Computer Interfaces in Diagnostic Process of Industrial Engineering

Lata Nautiyal, Preeti Malik and Mangey Ram

Abstract As the technologies are getting advanced, development of technologies focused on requirements to consider interaction between human and computers. Harmony between human being and technology is the ultimate objective. Industrial engineering works on synthesis and uses of scientific beliefs to analyze, design, and enhancement of the systems. In this concern, this chapter presents some human–computer interface devices that are used in diagnosis in industrial engineering or some other fields. These devices work as an interface between human and machine. Some of these devices are intelligent and adaptive also, and some are not intelligent. This chapter presents an overview of these interfaces.

Keywords Industrial engineering · Human–computer interface · Intelligent interface · Adaptive interface · Unimodal HCI · Multimodal HCI · Human factors · User interface · Pointing devices · Functionality · Usability

1 Introduction

Industrial engineering is concerned with the synthesis, design, and regulation of productive systems [1]. A productive system is a system that manufactures a product. Industrial engineering deals in designing a system that manufactures required number of products with less cost and good quality. It associates values of person behavior with concepts of engineering processes and study [2].

L. Nautiyal (✉) · P. Malik · M. Ram
Department of Computer Science and Engineering,
Graphic Era University, Dehradun, Uttarakhand, India
e-mail: lata.nautiyal1903@gmail.com

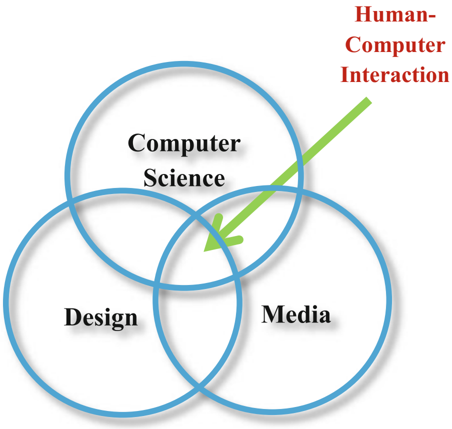
M. Ram
Department of Mathematics, Dehradun, Uttarakhand, India

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*,
Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_5

Industrial engineering merges the capabilities of a manager and engineer. For an industrial engineer, it is essential to understand the methodological details of a production processes and combine all the components of the system such as personnel, resources, equipment, and report so that a featured product can be developed on time and at an appropriate cost [2]. This process is industrial engineering. There are devices that are used in analysis process of industrial engineering, and these devices are called human–computer interfaces. These devices work as an interface between human and machine. Some of these devices are intelligent and adaptive also, and some are not intelligent. This chapter presents an overview of these interfaces. Interfacing was always major question in computer utilization. There are various approaches that are used for interacting with machines, and the voyage still endures and advance development methods emerge each day.

Enhancing the quality of interaction is not only a field of HCI research. There are also various branches for research which focuses on multimodality, smart adaptive, and active interfaces. Research of human–computer interface/interaction (HCI) field considers the development and usage of computing tools. It mainly focuses on the interaction between users and computer. HCI is positioned at the juncture of computer science, design, and media studies (Fig. 1). This term was promoted by Stuart K. Card, A. Newell, and T.P. Moran in 1983. It is also said that the ultimate interactions between human beings and computer systems which focuses on the user interface and the fundamental developments which yield the interactions are called as HCI. The contributing research areas include computer science, cognitive science, human factors, software engineering, management science, psychology, sociology, and anthropology (Fig. 2).

Fig. 1 Field of HCI research



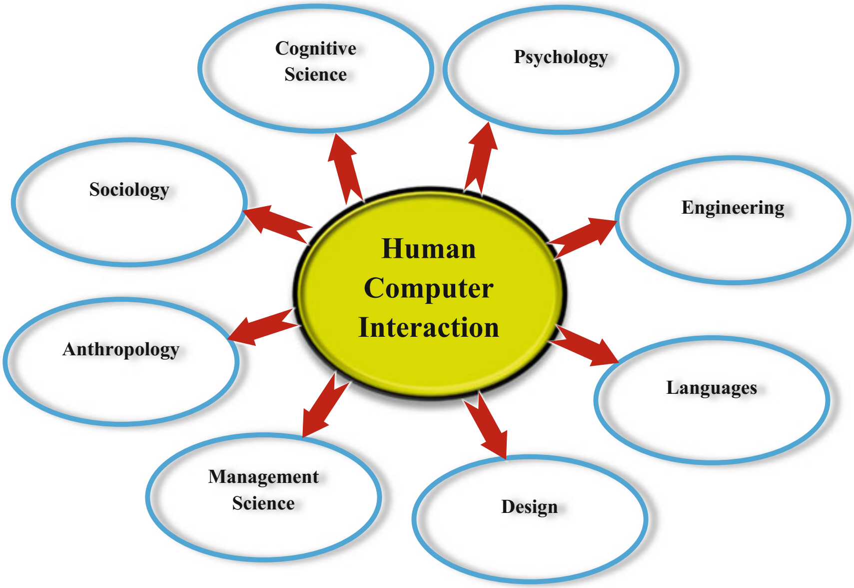


Fig. 2 Various disciplines of human–computer interaction research

1.1 *Objectives of HCI*

The ultimate objective of HCI is to produce operational or usable and secure systems and well-designed systems. The term operational or usable means making systems easy to learn and easy to use. A developer is required to do effort for following things while developing a usable HCI system:

- Realize the factors that conclude how a user uses technology.
- Design tools and techniques to support development of suitable systems.
- The interaction should be efficient, effective, and safe.
- User is priority.

In summary, the ultimate goal of designing a HCI system is the user satisfaction. People who are going to use the system should be the priority. Their requirements, abilities, and likings for accompanying various tasks should guide designers in the methodology that they design systems. User is not required to change themselves to work with the system; instead, the system should be developed to harmonize with user's requirements.

2 Basic Terminology

It can also be called man–machine interfacing. The notion of HCI was inevitably signified with the developing of computer itself because machines are not worthy unless we use them properly. It means there are two major terms in HCI design: functionality and usability [3].

A system can eventually be defined by its ***functionality***. It means how the very system helps to achieve its goals or functions. These functions are the collection of services or actions that are provided to the users. Though, the functionality can only be measured when the system is used by the user [4]. The second term is ***Usability***. Usability can be defined as a specific functionality as the span (range) and quantity (degree) by which the system can be utilized well to achieve specific objectives for specific public or user. The system is actually effective when there is a balance in both of these terms [5, 6].

HCI is about developing a system or machine or computer, so the clients can compile their tasks productively and securely. It is not only about usability; a wider definition of HCI can be:

HCI is a regulation concerned with development, synthesis, and implementation of interactive systems for human usage and also analysis of key facts surrounding them.

As the name infers, HCI has three components:

User: The term *user* means an individual user or a group of people who work together. Diverse people from diverse origin or mental models have diverse ways of understanding and storing the information. Social and nationwide dissimilarities also play an essential role.

Computer: The term computer means any technology ranging from laptop computer, desktop computer, or a large-scale computer. A website is itself a computer, and devices like mobile phones, and video cassette recorder are also referred as computer.

Interaction (the way user and computer communicate): Human and machine are two different entities with different features and characteristics, regardless of this HCI efforts for making their interaction easy. To develop a usable system, one requires to apply what he/she knows about human and machine and discuss with the possible users throughout the development process.

3 Overview of HCI

In the past era, the progresses built in HCI have caused the new tools to become accessible to everybody in no time although all the methods are not handy and/or reasonable for general user. Next two sections discusses these methods: First section gives a brief overview of approaches that are available to general user, and the second section presents a viewpoint of the track to which HCI research is going.

3.1 Existing HCI Technologies

Any HCI design must reflect several features of person behavior. Simplicity of the interaction is measured in terms of degree of participation of person interaction with the computer, and at times, it is more less when competed to the ease of interaction technique itself. The reason of difference in interfaces is difference in degree of functionality and usability. For example, an electrical kettle need not to be high level in terms of interface because its functionality is to increase the temperature of the substance and it would not be economical to have an interface more than a thermostatic on and off switch, whereas a simple webpage that may be restricted in functionality should be complicated enough in usability to draw the attention of customers [3].

Hence, while designing a HCI, one should be concerned about the degree of interface between the user and the machine. A user may be in three levels while interacting with a machine:

(1) **Physical**: This aspect of interaction governs the mechanism of interface between person and machine [7].
(2) **Cognitive**: This aspect is concerned about the way a user can interact and understand a machine [8].
(3) **Affective**: This aspect deals with the way that make user to use the computer [9]. It changes the attitude and emotion toward the user [3].

Vision-based input devices are the mostly used kind of devices. These devices are pointing devices or switch-based devices [10, 11]. Keyboard is an example of switch-based input device [12]. Mouse, touch screen panel, trackball, and pen-based input, etc. are the pointing input devices [13]. Joysticks have both the abilities of switch-based and pointing ability. Opposite to input devices, output devices can be of visual based or printing based [5]. Some advanced output devices depend on audition and they require speech recognition [14]. The main goal of designing these devices is that they should be more interactive, and this is a complex task [15]. On the contrary, it is easier to design auditory devices. Haptic devices are complex and expensive to develop [16]. These devices are usually developed for virtual realism [17] and disability assistance [18].

Nowadays, the traditional methods are combined with the modern techniques such as animation, networking, and intelligence. The modern devices can be classified into three categories: wearable devices [19], wireless devices [20], and virtual devices [21]. There are no clear boundaries among these three types of devices. GPS system [22], environmental monitoring, personal digital assistant (PDA), radio frequency identification (RFID) products, and virtual tour of real estate business [23] are some examples of these devices. Canesta keyboard is a virtual keyboard developed by Compaq's iPAQ (Fig. 3). This is designed by projecting a QWERTY-like model on a hard surface using a red light. Then, device traces finger movement of the user while typing on the surface with a motion sensor and sends the keystrokes back to the machine [24].

Fig. 3 Canesta virtual keyboard

3.2 Current Improvements in HCI

Recent improvements in HCI include networking, intelligent and adaptive interfaces and pervasive computing. Following section describes them:

3.2.1 Intelligent and Adaptive HCI

The devices used by the bulk of people are of command and action type. These devices are not using very classy physical tools and the research if flowing toward designing intelligent and adaptive devices. We can define the concepts of intelligence and adaptions as the enhancement in functionality and usability of newly developed devices. Development of such a device that is easier to use and satisfy the user is complex and not cost-effective. To achieve this goal, the interfaces of the devices are becoming easier to use. Note-taking machine is an excellent example of evolution, and another example is touch screen tablet PCs. Typewriters are replaced by keyboards, and keyboards are then replaced by touch screen table PCs. These PCs recognize your handwriting and convert it to text [25].

Attractive feature of modern interfaces are their intelligence [26] and adaption [27]. Intelligent HCI interfaces are those interfaces that include any kind of intelligence in the view from response to users. Some devices are empowered for verbal communications that interact to user in natural language [28–30]. It is not necessary to use intelligence in adaptive HCI interfaces, but they may use when interacting with the user [31]. A website to product selling is an example of adaptive HCI. These websites save the search and purchase history of users and suggest products according to their search history, so they are adaptive to some extent. These adaptations deal with affective and cognitive level of activity. PDAs and table PCs are the example of devices that uses intelligence and adaptation both the concepts. They can recognize handwriting and adapt the handwriting of the user that is logged in. These devices save the mistakes done by the user and also remember the correction made by the user.

Some non-intelligent HCI devices are passive in nature. They only react when the user invokes them, whereas intelligent and adaptive devices provide active interfaces [32, 33].

3.2.2 Pervasive Computing and Ambient Intelligence

Pervasive computing is recent field of research in HCI. These methods of HCI refer elimination of desktop and embed the computer in environment. Mark Weiser (1998) first gave the concept of pervasive (ubiquitous) computing. He wants to insert computer universally in the atmosphere so that a human can use more than one computer at a time, and these computers are wirelessly connected to each other [25]. This computing is said to be the third trend of computing after mainframe era (first wave) and PC era (second wave). In Mainframe era, there is one computer and many people, and in PC era, there is single computer and single person and pervasive computing deals with many machines and one human.

4 Architecture of HCI Systems' User Interface

Configuration is an essential issue of a HCI system. Even, a prearranged interface is normally demarcated by the large number of different inputs and outputs it delivers. HCI systems architecture demonstrates the working of these diversified inputs and outputs. Following segments of this chapter describe various design and configurations.

4.1 Unimodal User Interface

As stated before, an interface essentially depends upon large number of its inputs and outputs that are means of interaction. These interactions permit a user to communicate with machine. Every diverse autonomous single channel is named as modality [34]. Unimodal systems are the systems that depend upon a single modality. Depending upon the features of different modalities, systems are classified into three classes:

1. Visual-based System
2. Audio-based System
3. Sensor-based System.

4.1.1 Visual-Based Interface

The most well-known field in the research of HCI is visual-based HCI. Bearing in mind the range of uses, researchers attempted to handle many features of human reactions which can be acknowledged as a visual signal. Following are some major research fields:

- Facial Expression Analysis
- Body Movement Tracking (large scale)
- Gesture Recognition
- Gaze Detection (Eye Movement Tracking).

The aim of these fields varies because of their uses; a general commencement of all fields can be determined. The first one usually copes with identification of emotions visually [35–37]. The second [29, 38] and third [39–41] are typically the key center of this field and can have various drives; however, they are commonly intended for straight communication of person and computer in a command and action environment. The last one [30] is typically is not a direct practice of communication between human and computer which is normally used to well understand the consideration, aim of user in context-sensitive conditions [42].

4.1.2 Audio-Based Interface

Another significant field of HCI systems is interaction that is audio based. This field is concerned about the data obtained by various audio signals. These signals are not varying as visual signals, but the information collected from these signal is more reliable and useful. Following are the research fields of audio-based HCI:

- Speech Recognition
- Speaker Recognition
- Auditory Emotion Analysis
- Human-made Noise/Sign Detections (gasp, sigh, laugh, cry, etc.)
- Musical Interaction.

In starting days, first two fields were the major fields of research [14, 43]. Modern attempt to incorporate human emotions in intelligent HCI introduced the study of emotions in audio signals [44, 45]. Tone, pitch, sigh, and gasp, etc. are monitored to design a more intelligent HCI system [46]. Music production and interaction is a novel area of research. It can be applied in art industry [47].

4.1.3 Sensor-Based Interface

To provide interaction between user and machine, sensors are used in this type of HCIs. Following are some examples of sensor-based HCIs:

(i) Pen-based Interaction
(ii) Mouse and Keyboard
(iii) Joysticks
(iv) Motion Tracking Sensors and Digitizers
(v) Haptic Sensors
(vi) Pressure Sensors
(vii) Taste/Smell Sensors.

Pen-based sensors are used in mobile devices, and these are related to pen gesture [28] and handwriting identification fields. Motion tracking sensors are used in animation, art, and gaming. Wearable clothes are example of this type of sensors (refer Fig. 4). Pressure and haptic sensors are used in robotics and virtual realism

Fig. 4 Wearable motion capture cloth (taken from Operation Sports)

[16, 17, 19]. Hundreds of haptic sensors make the robots more aware and sensitive to touch [48, 49]. In Medical field, we also use haptic sensors [50].

4.2 Multimodal Interface

This type of HCI systems is combination of multiple modalities, and these modalities are the modes that the system reacts to the inputs [34] and these modes are communication channels. The basic idea behind these modes is the way humans interact to each other, and these communication channels are human senses: taste, smell, touch, hearing, and sight. Multimodal interfaces integrate diverse mixtures of speech, gesture, gaze, facial expressions, and other non-traditional modes of input [51]. A remarkable feature of multimodality is the association of various modalities to assist the identification.

4.3 Other User Interfaces

There are also some advanced paradigms of HCI such as World Wide Web. The web does not offer any technological innovation, because the functionalities already existed like protocols for transmission, file systems, etc. The revolution came with the start of browsers and hypertext markup language. Due to these functionalities, it got easier to access files on internet. Pervasive computing is also an advanced and emerging paradigm in the field of computing. As already discussed, the eventual goal of pervasive computing is to make the computer accessible everywhere it means in the environment itself. It is possible with the help of wireless networks, voice and vision systems, molecular electronics, and nanotechnology. The interfaces for these systems must use a rich set of interactive features.

5 Applications

Multimodal interfaces are being used in various applications counting map-based simulation like information center, biometric authentication systems, etc. The positive side of multimodal interfaces over conventional interfaces is that multimodal interface offers a more acceptable and user-friendly experience. For example, in a real estate system called REAL HUNTE, user can choose a house with a finger and get information about that particular house. Hence, it is a type of natural experience to point a house with finger and express doubts about that house. Additionally, multimodal interfaces offer redundancy to adapt unlike user and unlike circumstances. For example, MATCH kiosk permits user to use speech or handwriting to identify the kind of business to search for on a plot. Thus, if there is lot of noise, user can give his/her input query in written format. Some other applications of multimodal interfaces are as follows:

- Smart Video Conferencing
- Intelligent Homes/Offices
- Driver Monitoring
- Intelligent Games
- E-Commerce
- Helping People with Disabilities.

6 Future Direction of HCI

In the decade of 90s, HCI concern was shifted to communication between people enabled by machines. And network field was also growing in parallel to connect the computers. HCI concerned to identify how users interact effectively with a machine, but now HCI researchers concerned to interaction between human via a computer. HCI is being used by the researchers in various disciplines like Anthropology and Sociology. These sciences not only highlighted the consequences of computing on set of user but also how the user interpreted the machine or computer. Finally, the methodologies of these disciplines were combined and enhanced by more compound social modeling views and methods. Following are the recommendations for performing research and design in HCI.

6.1 Discover Novel Methods of Understanding Users

To discover novel methods of understanding users, one needs the articulation of various practices. In last decade, new techniques added human factors and cognitive sciences concepts. There are also other ways to improve existing techniques which includes views from more assorted disciplines and social traditions. The first phase of a new HCI is the analysis or synthesis, and this originates from analytic philosophy and involves clarifying the systems of meaning and value a particular set of actions involve.

6.2 Discover Novel Methods of Designing and Making

The development of models of new devices will need to be assumed in ways that are focused on specific types of users' value. The new development will be an extension of existing methods' development which was designed with the goal of usability, functionality, and balance between prototyped and engineered solutions.

In near future, less complex, rapid prototypes and design methods will be needed. Novel prototyping devices and methodologies will be particularly significant, letting the fast and less complex assembly of novel hardware and software to analyze together with and within normal objects and living spaces.

7 Conclusions

Analysis, design, and enhancement of the production system are the three main objectives of industrial engineering. A production system is a system that produces a good-quality product in time and within budget. The analysis or diagnosis process of industrial engineering requires a number of devices that can work as an interface between human and computer. This chapter discusses these device. HCI is a significant part of systems design. A good-quality system is easy to use and performs its functions correctly. Hence, development of HCI systems is the key area of research. Intelligent and adaptive devices are used in current era of industrial engineering. A device can be unimodal or multimodal. Pervasive computing aimed at embedding computer in the environment so that a person can use it everywhere. Virtual reality is the most advanced field of research in HCI. This chapter tried to describe these concerns and deliver a survey of open research.

References

1. Rahman CML, Uddin SM, Iqbal M (2014, December) Importance of human factors in industrial engineering and design. SEU J Sci Eng 8(1–2):7–13
2. Hicks PH (1994) Industrial engineering and management—a new perspective, International editions. McGraw-Hill, New York
3. Te'eni D, Carey J, Zhang P (2007) Human computer interaction: developing effective organizational information systems. Wiley, Hoboken
4. Shneiderman B, Plaisant C (2004) Designing the user interface: strategies for effective human-computer interaction, 4th edn. Pearson/Addison-Wesley, Boston
5. Nielsen J (1994) Usability engineering. Morgan Kaufman, San Francisco
6. Te'eni D (2016) Designs that fit: an overview of fit conceptualization in HCI. In Zhang P, Galletta D (eds) Human-computer interaction and management information systems: foundations. M.E. Sharpe, Armonk
7. Chapanis A (1965) Man machine engineering. Wadsworth, Belmont
8. Norman D (1986) Cognitive engineering. In: Norman D, Draper S (eds) User centered design: new perspective on human-computer interaction. Lawrence Erlbaum, Hillsdale
9. Picard RW (1997) Affective computing. MIT Press, Cambridge
10. Greenstein JS (1997) Pointing devices. In: Helander MG, Landauer TK, Prabhu P (eds) Handbook of human-computer interaction. Elsevier Science, Amsterdam
11. Myers BA (1998) A brief history of human-computer interaction technology. ACM interactions 5(2):44–54
12. Shneiderman B (1998) Designing the user interface: strategies for effective human-computer interaction (3rd edn). Addison Wesley Longman, Reading

13. Murata A (1991) An experimental evaluation of mouse, joystick, joycard, lightpen, trackball and touchscreen for pointing - basic study on human interface design. In: Proceedings of the fourth international conference on human-computer interaction 1991, pp 123–127
14. Rabiner LR (1993) Fundamentals of speech recognition. Prentice Hall, Englewood Cliffs
15. Karat CM, Vergo J, Nahamoo D (2003) Conversational interface technologies. In: Jacko JA, Sears A (eds) The human-computer interaction handbook: fundamentals, evolving technologies, and emerging application. Lawrence Erlbaum Associates, Mahwah
16. Robles-De-La-Torre G (2006) The Importance of the sense of touch in virtual and real environments. In: IEEE multimedia 13(3), Special issue on haptic user interfaces for multimedia systems, pp 24–30
17. Vince J (2004) Introduction to virtual reality. Springer, London
18. Iwata H (2003) Haptic interfaces. In: Jacko JA, Sears A (eds) The human-computer interaction handbook: fundamentals, evolving technologies, and emerging application. Lawrence Erlbaum Associates, Mahwah
19. Barfield W, Caudell T (2001) Fundamentals of wearable computers and augmented reality. Lawrence Erlbaum Associates, Mahwah
20. Yacoub MD (2002) Wireless technology: protocols, standards, and techniques. CRC Press, London
21. McMenemy K, Ferguson S (2007) A Hitchhiker's guide to virtual reality. A K Peters, Wellesley
22. Global Positioning System, "Home page", http://www.gps.gov/, visited on 10/10/2007
23. Chai JY, Hong P, Zhou MX (2004) A probabilistic approach to reference resolution in multimodal user interfaces. In: Proceedings of the 9th international conference on intelligent user interfaces, Funchal, Madeira, Portugal, pp 70–77
24. Bretz EA (2002) When work is fun and games. IEEE Spectrum 39(12):50–50
25. Riva G, Vatalaro F, Davide F, Alaniz M (2005) Ambient intelligence: the evolution of technology, communication and cognition towards the future of HCI, IOS Press, Fairfax
26. Maybury MT, Wahlster W (1998) Readings in intelligent user interfaces. Morgan Kaufmann Press, San Francisco
27. Kirlik A (2006) Adaptive perspectives on human-technology interaction. Oxford University Press, Oxford
28. Oviatt SL, Cohen P, Wu L, Vergo J, Duncan L, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J, Ferro D (2000) Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Human-Computer Interact 15:263–322
29. Gavrila DM (1999) The visual analysis of human movement: a survey. Comput Vis Image Underst 73(1):82–98
30. Sibert LE, Jacob RJK (2000) Evaluation of eye gaze interaction. Conf Human-Factors Comput Sys 281–288 (2000)
31. Various Authors (2007) Adaptive, intelligent and emotional user interfaces. In: Part II of HCI Intelligent Multimodal Interaction Environments, 12th International Conference, HCI International 2007 (Proceedings Part III), Springer, Heidelberg
32. Huhns MN, Singh MP (eds) (1998) Readings in agents. Morgan Kaufmann, San Francisco
33. Wasson CS (2006) System analysis, design, and development: concepts, principles, and practices. Wiley, Hoboken
34. Jaimes A, Sebe N (2007) Multimodal human computer interaction: a survey. Comput Vis Image Underst 108(1–2):116–134
35. Cohen I, Sebe N, Garg A, Chen L, Huang TS (2003) Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Underst 91(1–2):160–187
36. Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. Pattern Recogn 36:259–275
37. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans PAMI 22(12):1424–1445

38. Aggarwal JK, Cai Q (1999) Human motion analysis: a review. Comput Vis Image Underst 73 (3):428–440
39. Kettebekov S, Sharma R (2000) Understanding gestures in multimodal human computer interaction. Int J Artif Intell Tools 9(2):205–223
40. Wu Y, Huang T (1999) Vision-based gesture recognition: a review. In: Braffort A, Gherbi R, Gibet S, Richardson J, Teil D (eds) Gesture-based communication in human-computer interaction (Lecture notes in artificial intelligence), vol 1739. Springer, Berlin/Heidelberg
41. Kirishima T, Sato K, Chihara K (2005) Real-time gesture recognition by learning and selective control of visual interest points. IEEE Trans PAMI 27(3):351–364
42. Ruddaraju R, Haro A, Nagel K, Tran Q, Essa I, Abowd G, Mynatt E (2003) Perceptual user interfaces using vision-based eye tracking. In: Proceedings of the 5th international conference on multimodal interfaces, Vancouver, pp 227–233
43. Campbell JP Jr (1997) Speaker recognition: a tutorial. Proc IEEE 85(9):1437–1462
44. Oudeyer PY (2003) The production and recognition of emotions in speech: features and algorithms. Int J Hum Comput Stud 59(1–2):157–183
45. Chen LS (2000) Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction, PhD thesis, UIUC
46. Schröder M, Heylen D, Poggi I (2006) Perception of non-verbal emotional listener feedback. In: Proceedings of Speech Prosody 2006, Dresden, Germany, pp 43–46
47. Lyons MJ, Haehnel M, Tetsutani N (2003) Designing, playing, and performing, with a vision-based mouth interface. In: Proceedings of the 2003 conference on new interfaces for musical expression, Montreal, pp 116–121
48. Göger D, Weiss K, Burghart C, Wörn H (2006) Sensitive skin for a humanoid robot. In: Human-Centered Robotic Systems (HCRS'06), Munich
49. Khatib O, Brock O, Chang KS, Ruspini D, Sentis L, Viji S (2004) Human-centered robotics and interactive haptic simulation. Int J Robot Res 23(2):167–178
50. Legin A, Rudnitskaya A, Seleznev B, Vlasov Yu (2005) Electronic tongue for quality assessment of ethanol, vodka and eau-de-vie. Anal Chim Acta 534:129–135
51. Oviatt S (2003) Multimodal interfaces. In: Jacko JA, Sears A (eds) The human-computer interaction handbook: fundamentals, evolving technologies, and emerging application. Lawrence Erlbaum Associates, Mahwah

Models for Soil Vapor Extraction and Multiphase Extraction Design and Monitoring

Mariem Kacem

Abstract This chapter presents the tools needed to create a model of the soil vapor extraction (SVE) and multiphase extraction (MPE). It is based on the multiphase transport that can simulate the contaminant partition on the subsurface and perturbation caused by SVE and MPE methods. The model can predict the contaminant concentration on soil through time under variable soil saturation. The importance of a good definition of the non-aqueous phase liquid (NAPL)–gas phase mass transfer coefficient is discussed through SVE experimental and model result. The benefit of the model to design the MPE method is shown with regard to the number of wells and use of the aquifer drawdown principle.

Keywords SVE · MPE · Extraction · Soil · COV · NAPL · Model
Multiphase · Dispersion · Mass transfer coefficient · Transport
Stripping · Volatilization · Dissolution · Sorption · Saturation
Relative permeability · Simulation

1 Soil Vapor Extraction and Multiphase Extraction

The process of soil vapor extraction (SVE) has a role regarding the depollution of the unsaturated zone in the soil by the evaporation of the non-aqueous phase liquid (NAPL) trapped at soil residual saturation. Pollutant evaporation occurs by the pressure gradient in the pores that induces the circulation of air in contact with NAPL (Fig. 1a). In the above zone, pollutants are present in several states: dissolved in residual saturation water, adsorbed on organic matter and/or clay fraction of the soil, evaporated in gas phase, and water phase.

Multiphase extraction (MPE) involves simultaneous extraction of soil vapor and groundwater to remediate both types of contaminated media utilizing one pump and

M. Kacem (✉)
Univ Lyon, Ecole Nationale D'Ingénieurs de Saint-Etienne, Laboratoire de Tribologie et Dynamique Des Systèmes LTDS UMR 5513 CNRS, 42023 Saint-Etienne, France
e-mail: mariem.kacem@enise.fr

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*, Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_6

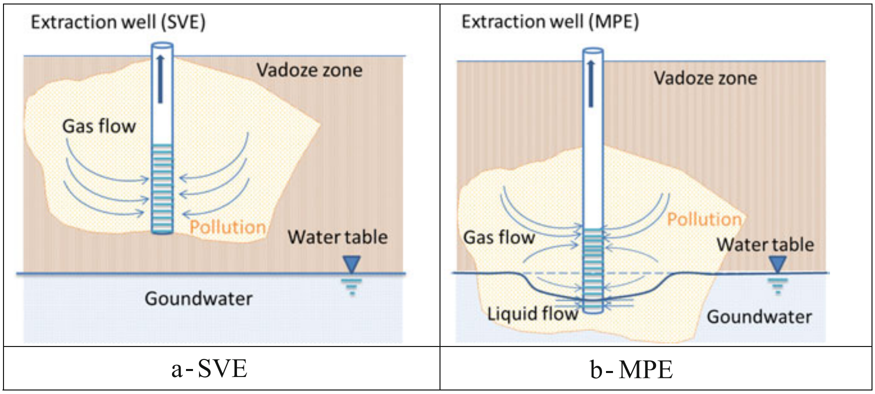


Fig. 1 Principal of SVE and MPE applications

one conduit (Fig. 1b). The extraction in this case is concerning multiple phases of fluids: NAPL, aqueous, and gas.

Soil vapor extraction and multiphase extraction are most efficient in porous media with moderate permeability. For MPE and in the case of highly permeable media, the dominant phase removed will be water, and there will be less influence on the soil gas. Near the aquifer, mostly groundwater will be removed with very little soil gas. In case of low permeability media, the dominant phase removed will be gas and the water will have much less influence. Near the aquifer, there will be much more drawdown, with little influence on the vadose zone. For the two methods SVE and MPE, the volatile compounds that tend not to adhere to the soil are responsive to the extraction.

Although developed in the industrial scale, these techniques need better control of the time of application and efficiency, to make energy and time savings. The existing works offer tools and relevant theoretical approaches, but give low sensibilities to certain parameter modifications, such as the humidity. In particular, there is a significant number of commercial or open-source software available for the flow and/or the transport in the porous media. Jennings and Patil [1] enumerated approximately thirty developed models only in the 90s. Few of these models are capable of simulating the multiphase flow and the multicompounds transport in conditions of real soils. Alternate methods are then run for the design of the process. As an example, those used for the MPE are based on rough solutions of radial unidimensional flow of the MPE [2]. This approach supposes a zone of homogeneous and isotropic treatment, not taking into account the spatial variation of the saturation in the zone. This kind of method gives approximate results and does not take into account the complexity of the porous media. On the academic plane, several models are developed in direct application of these techniques [3–6] or as theoretical developments of multiphase transport and transfer models [7, 8].

Both methods have common parameters of design. It is worth knowing the applied vacuum, the flows of the extracted phases, the location of wells in the capture field, and the location of the strainer of the well. To design those methods, several points must be controlled such as the characteristics:

- of the flow in the soil of the gas, the water (for the MPE) and of the NAPL (permeability, porosity, and soil retention curve);
- physical properties the NAPL (quantity, composition, physical parameters of transfer, and its distribution in the soil).

The purpose being to build a complete model in two components:

- a model of flow to predict the gas flow (and some liquid for the MPE) in the zone of treatment and in the well;
- a conceptual model for the mass transfer which determines transfer mechanisms and processes by the gaseous phase, the NAPL, (the dissolved phase for the MPE) and by the biodegradation.

2 Fundamentals of the Model

The conceptual model must be based on the understanding of its individual processes and mutual relations with the design of the SVE and MPE system. This model describes the various physio-chemical processes during the extraction with strong mathematical formulations.

The first point in the model construction is to identify the phase's state. In the vadose zone, the NAPL and the aqueous phase are considered to be non-mobile at residual saturation, in contrast to the gas phase experiencing a pressure gradient. The SVE model must describe the fluid transport and the mass transfer between phases according to the Fig. 2. In the saturated zone, all phases are mobile: gas,

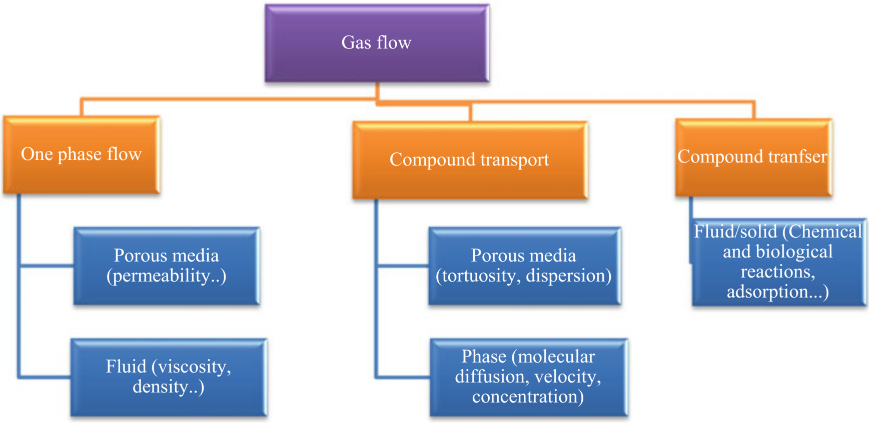


Fig. 2 SVE model used with neglecting the NAPL and water flow

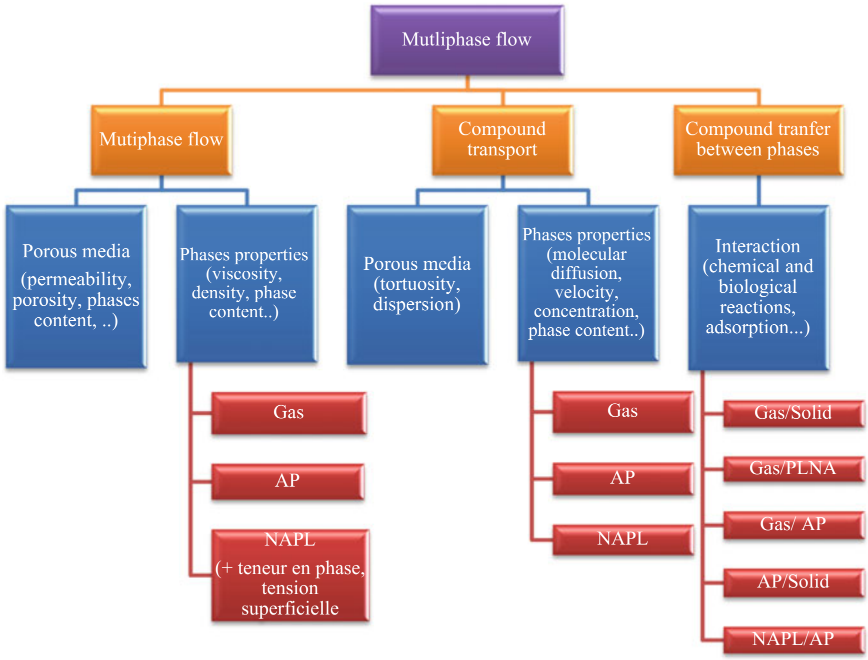


Fig. 3 MPE model (*AP* aqueous phase; *NAPL* non-aqueous phase liquid)

aqueous, and NAPL. The dominant processes of the MPE are the multiphase flow, the multicompound transport, and the transfer non-balanced between phases on soil, Fig. 3. The complete MPE model is composed of a model described in Fig. 2 applied on vadoze zone and a model described in Fig. 3 applied on saturated zone.

The model has to predict these processes and the distribution of the NAPL in the soil. It has to have the capacity of:

– identification of the optimal configuration of the extraction to implement (MPE: dual-phase extraction “DPE,” three-phase extraction “TPE,” bioSlurping);
– estimation of the airflow, the vacuum value, and the water extraction flow applied for a better design;
– determination of the wells and strainers locations to have an optimal efficiency;
– estimation of the mass and concentration of contaminants to eliminate from the soil in different phases (aqueous, gas, and solid) and then the estimation of the treatment time and extraction efficiency.

The modeling of the extraction consists of putting in equations of the processes by mathematical formulations and by solving the system by differential equations. The equations are governed numerically with a calculation code for a contamination treated by the MPE. Usually, the system of the mathematical equations for the multiphase flow and the isothermal transport of contaminants in the porous media is a result from the combination of several principles and laws of nature. Examples include the global mass conservation, mass conservation by compound, Darcy's law and its application in multiphase flow, the state equations, the constative relationships, and the mass transfer between phases.

3 Hypothesis

The processes involved in the SVE and MPE operations on soil are interdependent but happen simultaneously. The relevant understanding and the adoption of the simplifying hypotheses for the mathematical simulation are always necessary to obtain precise solutions on the efficiency of the operation at that time. Generally, the modeling of the processes of multiphase flow and pollutant transport is based on the following simplifying physical hypotheses:

- Water is the preferential wetting fluid, NAPL is the intermediate wetting fluid, and the gas phase is the no-wetting fluid [9];
- The sorption on solid phase is produced only by the aqueous phase [10], the gas phase and NAPL sorption are neglected;
- The Darcy's law is available for the multiphase flow system;
- The system is isotherm at T[K] temperature;
- The migration of the gas phase due to the water evaporation is neglected;
- Hysteresis retention curve is neglected, and the drainage process is always considered;
- The solid phase, aqueous phase, and NAPL are not compressible and the porosity is constant;
- The NAPL, solid, and aqueous phases densities are constant;
- The biodegradation is neglected.

4 SVE and MPE Models

4.1 Continuity Equations

The continuity differential equation for flow is seen in Eq. (1).

$$\frac{\partial(\varphi S_\alpha \rho_\alpha)}{\partial t} + \nabla[\rho_\alpha \vec{q}_\alpha] = Q^s_\alpha + \sum_\beta E_{\alpha,\beta} + \sum_\beta R_{\alpha,\beta} \tag{1}$$

where S_α is the saturation of α phase [−], φ is the porous media porosity, ρ_α is the bulk density of the α phase, q_α is the flow velocity (Darcy's velocity) [L.T^{-1}], Q^s_α is the α phase supply term [M.L^{-3} T^{-1}], $\sum_\beta E_{\alpha,\beta}$ is the sum representing the transfer of all compounds β to the α phase [M.L^{-3}·T^{-1}], and $\sum_{\beta} R_{\alpha,\beta}$ is the biochemical reaction sum in α phase for all compounds [M.L^{-3}·T^{-1}]. α phase can be gaseous, NAPL, or aqueous, written, respectively, g, o, and w.

The Darcy's velocity in multiphase system is calculated according to Darcy's law Eq. (2):

$$\vec{q}_\alpha = \frac{\overline{\overline{K}}\, k_{r\alpha}}{\mu_\alpha} (\nabla P_\alpha - \rho_\alpha\, g\, \nabla z) \tag{2}$$

where $\overline{\overline{K}}$ is the intrinsic permeability [L^2], $k_{r\alpha}$ is the relative permeability of α phase [−], μ_α is the dynamic viscosity of α phase [M.L^{-1}·T^{-1}], P_α is the pressure of α phase [M.L^{-1}·T^{-2}], g is the gravity acceleration [L.T^{-2}], and z is the Cartesian coordinate [−].

The mass conservation of compound β in α phase at the representative element volume scale (REV) is described by the transport equation (Eq. 3).

$$\frac{\partial}{\partial t}\left[\varphi\, S_\alpha\, C_{\alpha,\beta}\right] + \nabla\left[\vec{q}_\alpha C_{\alpha,\beta}\right] - \nabla\left[\varphi\, S_\alpha\, \overline{\overline{D}}_{\alpha,\beta} \nabla C_{\alpha,\beta}\right] = Q_{\alpha,\beta} + \sum_{\alpha} E_{\alpha,\beta} + \sum_{\beta} R_{\alpha,\beta} \tag{3}$$

where $C_{\alpha,\beta}$ is the concentration of compound β in phase α, and $\overline{\overline{D}}_{\alpha,\beta}$ is the hydrodynamic dispersion coefficient of the compound β in phase α defined by the following Eq. (4) [L^2.T^{-1}].

$$\overline{\overline{D}}_{\alpha\beta.} = \tau_\alpha\, D^o_{\alpha\beta}\, \delta_{ij} + \left[\alpha_T \left|\vec{U}_\alpha\right| \delta_{ij} + (\alpha_L - \alpha_T) \frac{U_{\alpha i}\, U_{\alpha j}}{\left|\vec{U}_\alpha\right|}\right] \tag{4}$$

in which $D^o_{\alpha,\beta}$ is the effective molecular diffusion coefficient of compound β in phase α [L^2.T^{-1}], δ_{ij} is the Kroncker symbol (equaling 1 if $i = j$, zero otherwise), α_L and α_T are the longitudinal and the transversal dispersivity, respectively, U_α is the pore velocity of phase α [L.T^{-1}], U_{ai}, U_{aj} are the pore velocity of α phase in i and j direction, respectively, [L.T^{-1}], and τ_α is the tortuosity factor.

The hydrodynamic dispersion coefficient was not usually expressed with all parts. A number of models have been presented to describe transport in both saturated and vadose zones that neglect molecular diffusion [11–13]. This simplification allows to analytically solving the transport equation. Recent works gives analytical solutions for model that incorporates molecular diffusion [14, 15].

The SVE model is composed of five partial differential equations (PDEs). The MPE model is composed of seven PDEs. For SVE, the continuity equation simulates the gas flow, while for MPE three global mass equations are used for mobile phases (gas, aqueous, and NAPL). The remaining four equations simulate the mass

Table 1 Summary of the PDEs used in SVE and MPE models

Equations	Application	
	Vadose zone	Saturated zone
Continuity for mobile phases $\frac{\partial(\varphi S_\alpha \rho_\alpha)}{\partial t} + \nabla[\rho_\alpha \vec{q}_\alpha] = Q_\alpha^s + \sum_\beta E_{\alpha,\beta} + \sum_\beta R_{\alpha,\beta}$	One equation for • Gas phase	Three equations for: • Gas phase • NAPL • Aqueous phase
Mass conservation of compound β in phase α at the REV scale $\frac{\partial}{\partial t}\left[\varphi S_\alpha C_{\alpha,\beta}\right] + \nabla\left[\vec{q}_\alpha C_{\alpha,\beta}\right] - \nabla\left[\varphi S_\alpha \overline{\overline{D}}_{\alpha,\beta} \nabla C_{\alpha,\beta}\right] = \sum_\alpha E_{\alpha,\beta} + \sum_\beta R_{\alpha,\beta} + Q_{\alpha,\beta}$	In the case of one compound β, four equations for: • Gas phase • NAPL • Aqueous phase • Solid phase	

conservation of: gaseous, NAPL, aqueous, and immobile solid phases according to Eq. (3). Table 1 summarizes the EDPs that were used for both methods.

It is assumed that the NAPL, aqueous, and solid phases are not compressible and have constant densities. Independent of whether or not the gas phase is compressible, their density is linked to the pressure and pollutants concentrations expressed by the perfect gas law (Eq. 5)

$$\rho_g(P_g, C_{g,\beta}) = \frac{P_g}{RT}\left[\frac{\rho_{air} + \sum_{\beta=1}^{N} C_{g,\beta}}{\frac{\rho_{air}}{M_{air}} + \sum_{\beta=1}^{N} \frac{C_{g,\beta}}{M_\beta}}\right] \tag{5}$$

where R is the perfect gas constant [M.L^2·mol^{-1}·T^{-3}], T is the temperature [T], M_β and M_{air} are the molecular mass, respectively, of compound β and of the air.

4.2 *Mass Transfer Between Phases*

4.2.1 Introduction

Mass transfer between phases can be described by Fig. 4. In saturated zone, contamination may exist in three physical states: sorbed to soil materials, dissolved in water, or immiscible liquid (NAPL). In the vadose zone, contaminants may exist in gas phase in addition. NAPL constituents may partition or move from one phase to

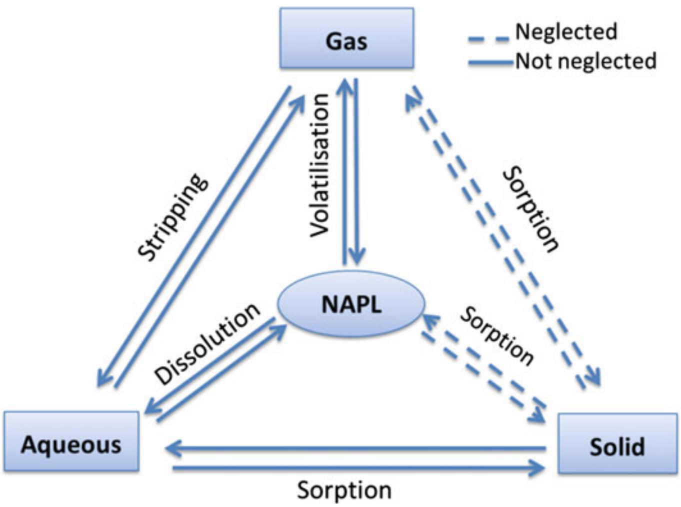


Fig. 4 Partitioning of NAPL among phases potentially found in the vadose zone

another, depending on environmental condition. In the hypothesis section, gas and NAPL adsorptions on solid phase are neglected. If they are not neglected, adsorption laws similar to the aqueous phase–solid phase mass transfer can be used. In the saturated zone, the mass transfers caused by gas phase presence are not considered.

4.2.2 NAPL–Gas Phase Transfer

The first mass transfer is the volatilization or mass transfer between NAPL and gaseous phases, simulated using a first-order kinetic model (Eq. (6)).

$$E_{o-g\,\beta} = \varphi\, S_g\, \lambda_{o-g,\beta} \left(\overline{C}^{sat}_{g,\beta} - C_{g,\beta} \right) \tag{6}$$

where $E_{o-g,\beta}$ is the term describing NAPL–gas phase mass transfer for compound β [M.L^{-3}·T^{-1}], $\lambda_{o-g,\beta}$ is the corresponding coefficient of mass transfer [T^{-1}], and $C_{g,\beta}$ and $\bar{C}^{Sat}_{g,\beta}$ are the equilibrium and saturation equilibrium concentrations of compound β in a mixture of gas phase [M.L^{-3}].

The ($\overline{C}^{Sat}_{g,\beta}$) is defined by Eq. (7).

$$\overline{C}^{Sat}_{g,\beta} = \omega_{g,\beta}\, C^{Sat}_{g,\beta} \tag{7}$$

where $C^{Sat}_{g,\beta}$ is the concentration of compound β in the gas phase at equilibrium and saturation [M.L^{-3}] and $\omega_{g,\beta}$ is the molar fraction of compound β in the gas phase [−].

4.2.3 NAPL–Aqueous Phase Transfer

The mass transfer between NAPL and aqueous phase is described by Eq. (8).

$$E_{o-w,\beta} = \varphi\, S_w\, \lambda_{o-w,\beta} \left(S^{eff}_{w,\beta} - C_{w,\beta} \right) \tag{8}$$

where $S^{eff}_{w,\beta}$ is the effective solubility of compound β in the aqueous phase [M.L^{-3}].

The NAPL–aqueous phase mass transfer coefficient can be calculated using the empirical correlation of modified Sherwood number (Eq. (9) [16]):

$$Sh^{m}_{o-w,\beta} = \frac{\lambda_{o-w,\beta}\, d^2_{50}}{D^o_{w,\beta}} = 12\,(\varphi - \theta_o)\, \mathrm{Re}^{0.75}_{w,\beta}\, \theta^{0.6}_o\, Sc^{0.5}_{w,\beta} \tag{9}$$

where $\lambda_{o-w,\beta}$ is NAPL–aqueous phase mass transfer coefficient of compound β [T^{-1}], d_{50} is the mean diameter of soil grains [L], $D^o_{w,\beta}$ is the molecular diffusion coefficient of compound β in the aqueous phase [$L^2 \cdot T^{-1}$], θ_o is the volumetric content of NAPL phase [$L^3 \cdot L^{-3}$], $Re_{w,\beta}$ is the Reynolds number of compound β in the aqueous phase, and $Sc_{w,\beta}$ is the Schmidt number of compound β in the aqueous phase.

4.2.4 Gas Phase–Aqueous Phase Transfer

To simulate mass transfer between aqueous and gas phases (stripping), the mass transfer coefficient can be estimated from the empirical model of Chao et al. [17] (Eq. (10)).

$$\begin{aligned} E_{w-g,\beta} &= \varphi\, S_w \lambda_{w-g,\beta} \left(C_{w\beta} - \frac{C_{g\beta}}{H_\beta} \right) \\ &= \varphi\, S_w\ 10^{-2.49} D^{o\,0.16}_{g,\beta} U_g^{0.84}\, d_{50}^{0,55} H_\beta^{-0.61} \left(C_{w,\beta} - \frac{C_{g,\beta}}{H_\beta} \right) \end{aligned} \tag{10}$$

in which $\lambda_{w-g,\beta}$ is the aqueous phase–gas phase mass transfer coefficient of compound β [T^{-1}], $C_{w,\beta}$ is the concentration of compound β in the aqueous phase [$M.L^{-3}$], U_g is the pore velocity of the gas phase [$L.T^{-1}$], and H_β is the Henry constant of compound β [-].

4.2.5 Solid Phase–Aqueous Phase Transfer

To simulate the mass phase transfer between aqueous and solid phases (sorption), the corresponding mass transfer coefficient can be estimated from the empirical model of Brusseau et al. [18] (Eq. (11)).

$$\begin{aligned} E_{w-s,\beta} &= \varphi\, S_w\ \lambda_{w-g,\beta} \left(C_{w,\beta} - \frac{C_{s,\beta}}{K_{d,\beta}} \right) \\ &= \varphi\, S_w\ Exp\ (0.301 - 0.668\ \log\ (K_{d,\beta})) \left(C_{w\beta} - \frac{C_{s,\beta}}{K_{d,\beta}} \right) \end{aligned} \tag{11}$$

where $C_{s,\beta}$ is the concentration of compound β in the solid phase [$M.L^{-3}$] and $K_{d,\beta}$ is the sorption coefficient of compound β in the soil [$L^3.M^{-1}$].

4.2.6 NAPL-gas Mass Transfer Coefficient Determination

The NAPL–gas mass transfer coefficient has an important effect on the model result [19]. Indeed, the most difficult part in the simulation process at real scale is to find the mathematical relationship that can estimate the coefficient based on the flow, porous media, pollutant characteristics, and pollutant content.

It is possible to estimate this coefficient based on the analytic solutions of the pollutant mass conservation equation. This equation expresses the gas phase and NAPL–gas phase transfers in 1D geometry (Eq. (3)) as proposed by van Genuchten et al. [20] for a stationary state (Eq. (12)).

$$\lambda_{o-g,\beta} = \frac{1}{4\,\overline{\overline{D}}_g}\left[\left(U_g - \frac{2\,\overline{\overline{D}}_g}{L_c}\ln(1 - \frac{C_{g,\beta}}{C_{g,\beta}^{sat}})\right)^2 - U_g^2\right] \tag{12}$$

in which L_c represents the column height [L].

Another method is the analytic solution developed by Wilkins et al. [21], which neglects the hydrodynamic dispersion $\overline{\overline{D}}_{g,\beta}$ (Eq. (13))

$$\lambda^*_{o-g,\beta} = -\frac{q_g}{L_c}\,\ln\,(1 - \frac{C_{g,\beta}}{C_{g,\beta}^{sat}}) \tag{13}$$

where $\lambda^*_{o-g,\beta} = \varphi S_g\,\lambda_{o-g,\beta}$.

Wilkins et al. [21] proposed the following empirical relationship:

$$\lambda^*_{o-g,\beta} = 10^{-0,4221}\,D^{o\,0,38}_{g,\beta}\,U_g^{0,62}\,\,d_{50}^{0,44} \tag{14}$$

The above equation relates the mass transfer coefficient to the interstitial pore velocity and the mean diameter of soil particles. The empirical model of Wilkins et al. [21] introduced a gas phase content in the estimation of the mass transfer coefficient and neglected the hydrodynamic dispersion.

Many authors have observed a reduction of the mass transfer coefficient with the decrease of NAPL saturation by evaporation. Van der Ham and Brouwers [22] offered an improvement of the model to link mass transfer coefficient to NAPL saturation and extraction time in transient state (Eq. (15)). Van der Ham and Brouwers [22] illustrated that this empirical parameter is linked to the saturation at the aqueous phase and to the soil homogeneity.

$$\lambda_{o-g,\beta} = \lambda^{init}_{o-g,\beta}\left(\frac{S_o}{S_o^{init}}\right)^{\varepsilon} \tag{15}$$

where $\lambda^{init}_{o-g,\beta}$ is the initial NAPL–gas mass transfer coefficient deduced from the expression of $\lambda^*_{o-g,\beta}$ and ε is an empirical parameter. The empirical models of those studies [21, 22] neglected dispersion coefficient which can be used for pore's velocity more than 0.1 cm·s^{-1} [23].

In order to generalize the model regarding the saturated and the unsaturated zone, Esrael et al. [24] proposed the use of the reference NAPL saturation and implemented the mass coefficient estimation with conditions mentioned in Eq (15). The obtained equation (Eq. (16)) can be used in case of variable saturation media such as the case of the Multiphase Extraction process. In this situation, the pollutant content and the water content are not limited. MPE is used in the capillary fringe and the saturated zone drained by dewatering in SVE. Then, the saturation in NAPL and water is varied in time and space.

$$\lambda_{o-g,\beta} = \begin{bmatrix} \lambda_{o-g,\beta}^{init} & ; S_o > S_o^{ref} \\ \lambda_{o-g,\beta}^{init} \left(\frac{S_o}{S_o^{ref}} \right)^{\varepsilon} & ; S_o \leq S_o^{ref} \end{bmatrix} \tag{16}$$

with S_o^{ref} is the reference NAPL saturation.

The NAPL–gas phase estimation use is depending also on the extraction application. It can be considered constant for one-dimensional approach when pore velocity is homogeneous, whereas this velocity varied with distance from extraction wells at the radial pilot or field scale [25, 26].

4.3 Constitute Equations

To resolve the continuity equations, a system of equations allows connecting phase relative permeability to the saturations and then to the employed pressure. It depends on the number of phases in the REV. Some works used an empirical model for relative permeability determination [27–29]. Other models are developed based on the phase saturation according to the mobile phase presence: the two-phase and three-phase systems.

4.3.1 Two-Phase System (Aqueous and Gas Phases)

The model uses the van Genuchten–Mualem model [30] for each phase in the two-phase system. The wetting phase is the aqueous one, the capillary pressure is defined by Eq. (17), and the saturation on aqueous phase is defined by Eq. (18).

$$P_{cgw} = P_g - P_w \tag{17}$$

$$\bar{S}_w(P_{cgw}) = \frac{S_w - S_{wr}}{1 - S_{wr}} = \begin{bmatrix} \left[1 + (\alpha_{vg} P_{cgw})^{n_{vg}} \right]^{-m_{vg}} & ; P_{cgw} > 0 \\ 1 & ; P_{cgw} \leq 0 \end{bmatrix} \tag{18}$$

$$S_g = 1 - S_w \tag{19}$$

where P_{cgw} is the capillary pressures for gas–aqueous phase, P_g and P_w are pressures of gas and aqueous phase, respectively, $\bar{S}_w$ is the effective residual saturation in the aqueous phase, α_{vg}, n_{vg}, and m_{vg} are the van Genuchten parameters.

The relative permeability for the two phases is expressed by Eq. (20) for aqueous phase and Eq. (21) for gas phase.

$$k_{rw}(\bar{S}_w) = \bar{S}_w^{0,5}\left[1-(1-\bar{S}_w^{1/m_{vg}})^{m_{vg}}\right]^2 \tag{20}$$

$$k_{rg}(\bar{S}_w) = (1-\bar{S}_w)^{0,5}\left[1-\bar{S}_w^{1/m_{vg}}\right]^{2m_{vg}} \tag{21}$$

4.3.2 Three-Phase System (NAPL, Aqueous, and Gas Phases)

The Parker et al. model [9] can be used. The saturation of each phase can be calculated by the Eqs. (22)–(24).

$$\bar{S}_l(P_{cgo}) = \frac{S_w + S_o - S_{wr} - S_{or}}{1 - S_{wr}} = \begin{bmatrix} \left(1+\left(\gamma_{go}\,\alpha_{vg}\,P_{cgo}\right)^{n_{vg}}\right)^{-m_{vg}} & ; P_{cgo} > 0 \\ 1 & ; P_{cgo} \le 0 \end{bmatrix} \tag{22}$$

$$\bar{S}_w(P_{cow}) = \frac{S_w - S_{wr}}{1 - S_{wr}} = \begin{bmatrix} \left(1+\left(\gamma_{go}\,\alpha_{vg}\,P_{cow}\right)^{n_{vg}}\right)^{-m_{vg}} & ; P_{cow} > 0 \\ 1 & ; P_{cow} \le 0 \end{bmatrix} \tag{23}$$

$$S_g = 1 - S_o - S_w \tag{24}$$

where $P_{cgo} = P_g - P_o$ and $P_{cow} = P_o - P_w$ are the capillary pressure for gas-NAPL and NAPL-aqueous phases, respectively, P_o is the pressure of NAPL, γ_{go} and γ_{ow} are scaling coefficients of two-phase system to three-phase to one, which can be estimated approximately by Eqs. (25) and (26) [31].

$$\gamma_{go} = \frac{\sigma_{go} + \sigma_{ow}}{\sigma_{go}} \tag{25}$$

$$\gamma_{ow} = \frac{\sigma_{go} + \sigma_{ow}}{\sigma_{ow}} \tag{26}$$

In which $\sigma_{gw}.\sigma_{go}.\sigma_{ow}$ are the interfacial tensions between phases gas–aqueous, gas–NAPL, and NAPL–aqueous, respectively.

Lenhard et al. [32] applied the above model to experimental results to determine the effective residual NAPL saturation, expressed by Eq. (27).

$$\overline{S}_{or} = \overline{S}_{or}^{\max}(\overline{S}_t^{\max} - \overline{\overline{S}}_w)^{0,5}(1 - \overline{\overline{S}}_w)^{1,5} \tag{27}$$

where $\overline{S}_{or}^{\max}$ is the maximal effective residual saturation in NAPL [-], $\overline{S}_t^{\max}$ is the maximal effective saturation in the total liquid phase [-], and $\overline{\overline{S}}_w$ is the apparent water saturation which is the sum of effective water saturation and the effective entrapped air saturation.

The van Genuchten–Mualem model [30] was used for all phases, the relative permeabilities of which were calculated using Eqs. (28)–(30).

$$k_{rw}(\overline{S}_w) = \overline{S}_w^{0,5}\left[1 - \left(1 - \overline{S}_w^{1/m_{vg}}\right)^{m_{vg}}\right]^2 \tag{28}$$

$$k_{ro}(\overline{S}_w, \overline{S}_t) = \left(\overline{S}_t - \overline{S}_w\right)^{0,5}\left[\left(1 - (\overline{S}_w + \overline{S}_{or})^{1/m_{vg}}\right)^{1/m_{vg}} - \left(1 - \overline{S}_t^{1/m_{vg}}\right)^{m_{vg}}\right]^2 \tag{29}$$

$$k_{rg}(\overline{S}_t) = \left(1 - \overline{S}_t\right)^{0,5}\left[1 - \left(1 - \overline{S}_t^{1/m_{vg}}\right)^{m_{vg}}\right]^2 \tag{30}$$

in which k_{ro} is the relative permeability of NAPL.

5 Application of SVE Model on 1D Column Experiment

The model was applied to simulate results of SVE applied on one-dimensional column. Two tests were running with dry sand and low water saturation. The used pollutant was the toluene. Table 2 summarizes the operating conditions of tests.

The SVE model was used introducing three models of NAPL-gas phase mass coefficient of the convection diffusion equation, the Wilkins et al. Model [21], and Esrael et al. model [24]. Several parameters values are tested, and best results were kept (Fig. 5).

The extracted gas concentration curves decrease continuously with similar shapes for all tests. It can be observed the three parts describing the typical extracted gas concentration: the flushing phase, evaporation phase, and diffusion phase separated by two inflexion points. At the end of the first part, about of 90% of

Table 2 Operating conditions of tests

Conditions	Unity	Test 1	Test 2
Porosity φ	[–]	0.4145	0.411
Toluene saturation	[–]	0.0294	0.03
Water saturation	[–]	–	0.076
Gas flow	[mL $\cdot$ min^{-1}]	3183	3203
Concentration at saturation	[mg $\cdot$L^{-1}]	108.9	106.2

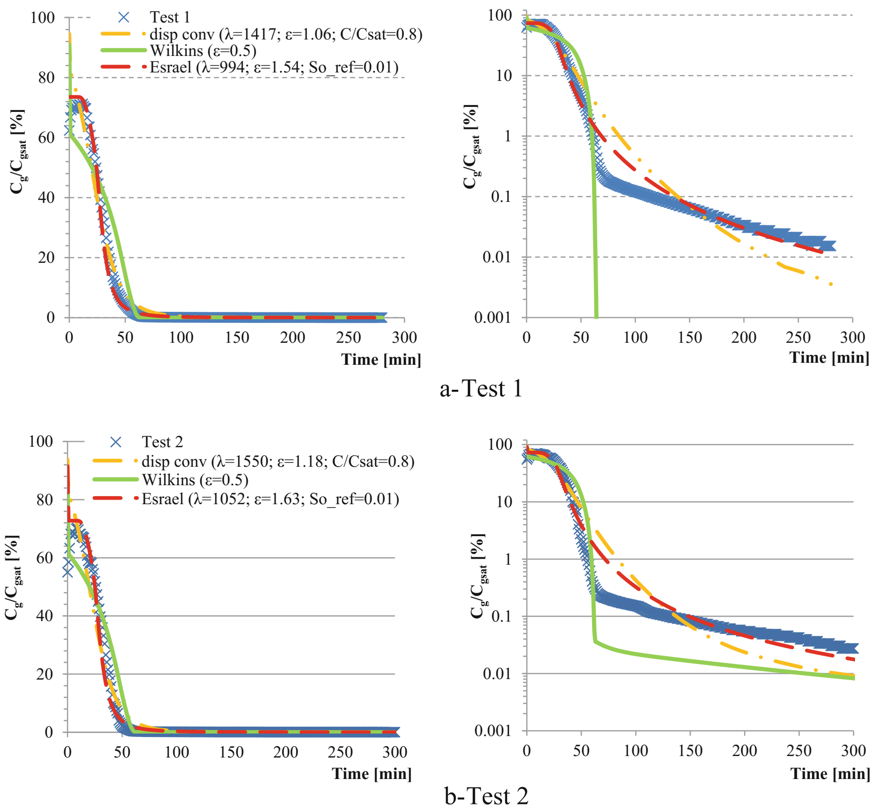


Fig. 5 Experiments' and models' results at semilogarithmic scale

the pollutant has been extracted, with the residual amount becoming progressively less accessible to the airflow due to being trapped in smalls boreholes (Test 1), or dissolved in water/adsorbed on soil grains for Test 2. The second inflexion point marks the disappearance of free NAPL, with the gas concentration after this point being determined by the diffusion step, where a new equilibrium between phases is established: adsorbed NAPL, dissolved NAPL, no accessible NAPL, and airflow. The logarithmic scale allows showing difference between models after the second inflexion point (Fig. 5).

For dry soil, Eq. (1.10) (dissolution) and Eq. (1.11) (adsorption) were not activated. Models assuming an initially low λ_{o-g} could not describe the new equilibrium state: The mass transfer coefficient decreased to values close or equal to zero. Subsequently, the concentration decreased to zero, not being able to model this point. Conversely, the above point and the curve beyond it could be estimated by Esrael model and the dispersion convection model (with overestimated λ_{o-g}). The presence of water in soil (tests 2) activated Eqs. (1.10) and (1.11), and a new equilibrium was established for all models, which could simulate this second inflexion point but not the subsequent curve evolution.

The values of the initial mass transfer coefficient for Test 2 are similar to those obtained for Test 1. The observed difference was due to the small flow rate difference and the saturation effect at 0.0755. The above saturation decreased the gas phase content and subsequently increased pore velocity. Tortuosity and the hydrodynamic dispersion were also reduced, and the concentration did not drop to zero, as in test 1. In this situation, the presence of water activated interphase mass transfer (aqueous–gas, NAPL–aqueous, and aqueous–solid), allowing exchange between phases and increasing the extracted gas concentration in the diffusion step.

6 Application of MPE Model

This example of application is excerpt from the PhD work of Esrael [33]. The MPE model was applied in the case assuming pollution. The proposed scenario constitutes pouring 8 m^3 of toluene on a circular surface of 1.5 m radius. During 12 h, pollutant infiltrates in the sandy soil with 0.157 $cm \cdot min^{-1}$ Darcy's velocity. The NAPL enriches the capillary zone and the water table situated at 3 m of depth. The NAPL can create a lens on the water surface of 8 m radius and 5 cm of thickness. Pollutant propagates in the soil in the three phases: aqueous (dissolution), gas (evaporation), and solid (sorption). We tested an intervention on site using the MPE after two months.

Pollutant relative concentration in gas phase $\left(\frac{C_{g,\beta}}{C_{g,\beta}^{sat}}\right)$ quickly drops in saturated zone and is still present in capillary fringe, where gas flow is low (Fig. 6a). Pollutant relative concentration in aqueous phase $\left(\frac{C_{w,\beta}}{S_{w,\beta}^{eff}}\right)$ remains present after 1 month (Fig. 6b) and decreases after 50 days to reach zero on unsaturated zone, where gas flow evaporates the pollutant dissolved in the aqueous phase. The pollutant remains present in the capillary fringe.

After 3 months, pollutant concentration in the solid phase in the unsaturated zone increases to 8 $mg \cdot kg^{-1}$. However, in the capillary fringe the concentration remains high, even after one year of treatment (Fig. 7).

After 20 days of extraction, NAPL disappears from the unsaturated zone under the effect of the convective gas flow. The aqueous flow dissolves the NAPL on the water table, and the pollutant disappears after 3 months. In the capillary zone, pollutant is present after this time were aqueous flow and gas flow are low (Fig. 8). It is possible to improve the method by using the aquifer drawdown principle which permits oil extraction from the entire capillary zone or by using numerous extraction wells. The simulation shows that we need a minimum of 3 extraction wells to extract pollutant from capillary fringe in 1 year.

The model gives the results of the MPE operation on this site; pollutant flow on extracted phase and pollutant extracted volumes in each phase show that on the first

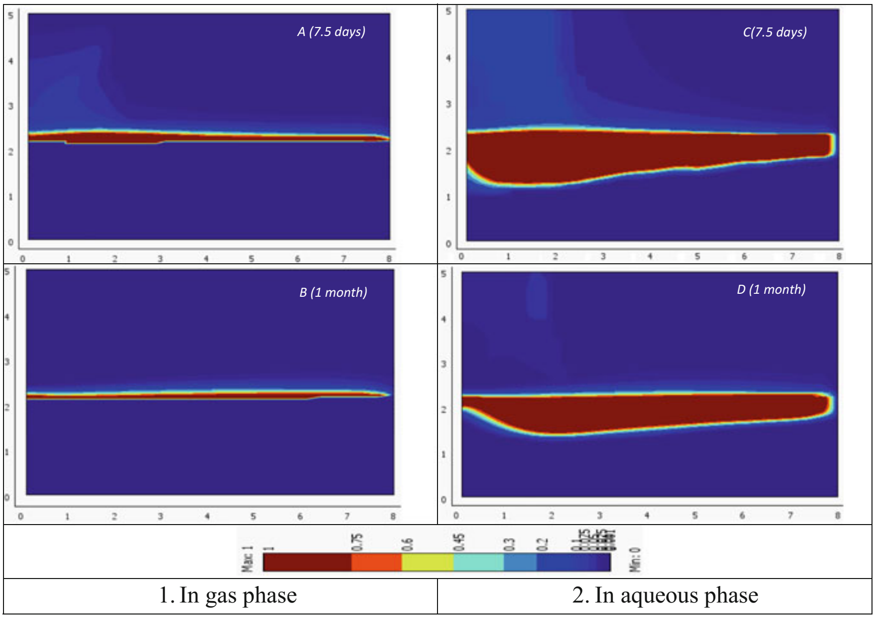


Fig. 6 Pollutant relative concentration

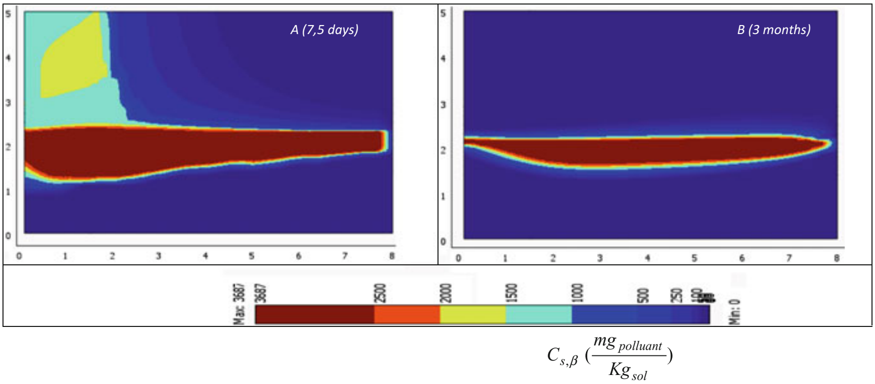


Fig. 7 Pollutant concentration on solid phase $C_{s,\beta}\ (\frac{mg_{polluant}}{Kg_{sol}})$

days, more than 180 L.day^{-1} of NAPL were extracted. When the free phase decreases at 100 days, NAPL flow decreases. At 50 days of extraction, the most mobile NAPL was recovered, 2385 L from an initial 2870 L, which is about 83% of efficiency. The gas flow has a similar behavior. During the four first days, the flow was more than 50 L.day^{-1} and decreases thickly with the NAPL disappearance in the no saturated zone. In contrast, the aqueous phase flows still in an order of

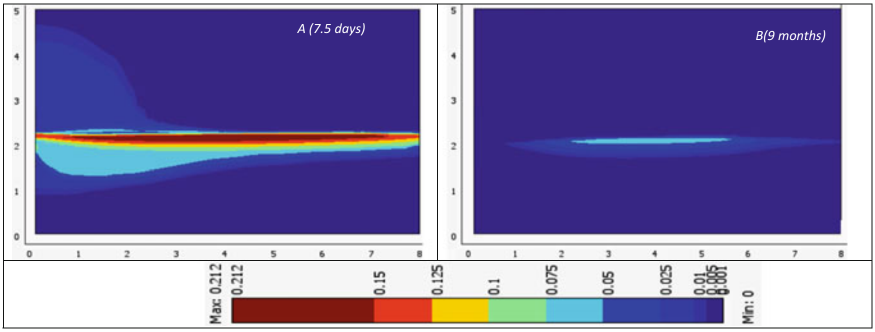


Fig. 8 NAPL saturation through the time

10 L.day^{-1} because the major pollutant quantity is trapped on the capillary fringe and there is no available space in the aqueous phase.

7 Conclusion

The multiphase transport equations can be used to create a model of the Soil Vapor Extraction (SVE) and multiphase extraction (MPE). The SVE model is composed of five partial differential equations (PDEs). The MPE model is composed of seven PDEs. For SVE, the continuity equation simulates the gas flow, while for MPE three global mass equations are used for mobile phases (gas, aqueous, and NAPL). The remaining four equations simulate the mass conservation of: gaseous, NAPL, aqueous, and immobile solid phases. To resolve the continuity equations, a system of equations allows connecting phase relative permeability to the saturations and then to the employed pressure.

A good simulation is depending on the definition of the mass transfers between phases and models used to describe the saturation state.

The first example of application shows the importance of a good definition of the non-aqueous phase liquid (NAPL)–gas phase mass transfer coefficient. In order to generalize the model regarding the saturated and the unsaturated zone, reference NAPL saturation is used. MPE is used in the capillary fringe and the saturated zone drained by dewatering in SVE. Then, the saturation in NAPL and water is varied in time and space. To have a global model for all range of velocity, a mass coefficient is not neglected.

The second example of application shows the benefit of the model to design the MPE method with regard to the number of wells and use of the aquifer drawdown principle.

References

1. Jennings AA, Patil P (2002) Feasibility modeling of passive soil vapor extraction. J Environ Eng Sci 1(2):157–172. Available on : http://dx.doi.org/10.1139/s02-009
2. USEPA (1999) Multi-phase extraction: state-of-the-practice. Ohaio, USA. 71 p. Available on http://www.epa.gov/tio/download/remed/mpe2.pdf, ISBN : EPA 542-R-99-004
3. Reddy KR, Adams J (2008) Conceptual modelling of air sparging for groundwater remediation. In: Proceedings of the 9th international symposium on environmental geotechnology and global sustainable development, Hong Kong, 18 p
4. Lu Y, Fan W, Yang YS, Du XQ, Zhang GX (2013) Mathematical modeling of differentiation processes in porous media during soil vapor extraction (SVE) remediation of contaminated soil/water. Water Air Soil Pollut 224:1491–1499
5. Hartsock J (2014) Evaluating the effectiveness of pre-remedial LNAPL recovery modeling using the 'LNAPL Distribution and Recovery Model' at two active remediation sites. Emporia State University, Thèse de doctorat, p 190
6. Souza MM, Oostrom M, White MD, da Silva GC, Brabosa MC (2016) Simulation of subsurface multiphase contaminant extraction using a Bioslurping well model. Transp Porous Med 114(3):649–673
7. Helmig R, Flemisch B, Wolff M, Ebigbo A, Class H (2013) Model coupling for multiphase flow in porous media. Adv Water Resour 51:52–66
8. Hwang HT, Parck YJ, Sudicky EA, Unger AJA, Illman WA, Frape SK, Shouakar-Stash O (2013) A multiphase flow and multispecies reactive transport model for DNAPL- involved compound specific isotope analysis. Adv Water Resour 59:111–122
9. Parker JC, Lenhard RJ, Kuppusamy T (1987) A parametric model for constitutive properties governing multiphase flow in porous media. Water Resour Res 23(4):618–624
10. Kaluarachchi JJ, Parker JC (1992) Multiphase flow with a simplified model for oil entrapment. Transp Porous Media 7(1):1–14. doi:10.1007/BF00617314
11. Chen JS, Liu CW, Liao CM (2002) A novel analytical power series solution for solute transport in a radially convergent flow field. J Hydrol 266:120–138. doi:10.1016/S0022-1694(02)00119-1
12. Wang Q, Zhan H (2013) Radial reactive solute transport in an aquifer-aquitard system. Adv Water Resour 61:51–61
13. Lai KH, Liu CW, Liang CP, Chen JS, Sie BR (2016) A novel method for analytically solving a radial advection-dispersion equation. J Hydrol 542:532–540. doi:10.1016/j.jhydrol.2016.09.027
14. Haddad AS, Hassanzadeh H, Abedi J, Chen Z, Ware A (2015) Characterization of scale-dependent dispersivity in fractured formations through a divergent flow tracer test. Groundwater 53:149–155. doi:10.1111/gwat.12187
15. Huang J, Goltz MN (2017) Analytical solutions for a soil vapor extraction model that incorporates gas phase dispersion and molecular diffusion. J Hydrol 549:452–460
16. Miller CT, Poirier-McNeil MM, Mayer AS (1990) Dissolution of trapped nonaqueous phase liquids: mass transfer characteristics. Water Resour Res 26(11):2783–2796. doi:10.1029/WR026i011p02783
17. Chao K-P, Say Kee O, Angelos P (1998) Water-to-air mass transfer of VOCs: laboratory-scale air sparging system. J Environ Eng 124(11):1054–1060. doi:10.1061/(ASCE)0733-9372(1998)124:11(1054)
18. Brusseau ML, Rao PSC (1989) Sorption nonideality during organic contaminant transport in porous media. Crit Rev Environ Control 19(1):33–99. doi:10.1080/10643388909388358
19. Zhao L (2007) Three-dimensional soil vapour extraction modeling. University of Guelph, Guelph, Canada, p 296
20. van Genuchten MT, Alves WJ (1982) Analytical solutions of the one-dimensional convective-dispersive solute transport equation. Washington, DC. 149 p

21. Wilkins MD, Abriola LM, Pennell KD (1995) An experimental investigation of rate-limited nonaqueous phase liquid volatilization in unsaturated porous media: steady state mass transfer. Water Resour Res 31(9):2159–2172. doi:10.1029/95WR01677
22. van der Ham AJG, Brouwers HJH (1998) Modelling and experimental investigation of transient, nonequilibrium mass transfer during steam of a nonaqueous phase liquid in unsaturated porous media. Water Resour Res 34(1):47–54
23. Kacem M, Esrael D, Benadda B (2017) Effect of dispersion coefficient on PLNA/gas mass transfer coefficient determination. EQA-Int J Environ Qual 22:33–42
24. Esrael D, Kacem M, Benadda B (2017) Modelling mass transfer during venting/soil vapour extraction: non-aqueous phase liquid/gas mass transfer coefficient estimation. J Contam Hydrol (In press). doi:10.1016/j.jconhyd.2017.05.003
25. Duggal A, Zytner RG (2009) Comparison of one- and three-dimensional soil vapour extraction experiments. Environ Technol 30(4):407–426
26. Boudouch O, Esrael D, Kacem M, Benadda B, Gourdon R (2016) Validity of the use of the mass transfer parameters obtained with 1D column in 3D systems during soil vapor extraction. J Environ Eng 142(6):04016018(1-9)
27. Stylianou C, deVantier BA (1995) Relative air permeability as function of saturation in soil venting. J Environ Eng 121(4):337–347
28. Falta RW, Javandel I, Pruess K, Witherspoon PA (1989) Density-driven flow of gas in the unsaturated zone due to the evaporation of volatile organic compounds. Water Resour Res 25:2159–2169
29. Boudouch O, Esrael D, Kacem M, Benadda B (2012) Influence of soil air permeability change on soil vapour extraction systems design. CLEAN–Soil, Air Water 40(5):461–471
30. van Genuchten MT (1980) A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44(5):892–898
31. Lenhard RJ (1992) Measurement and modeling of three-phase saturation-pressure hysteresis. J Contam Hydrol 9(3):243–269. Available on http://dx.doi.org/10.1016/0169-7722(92)90007-2
32. Lenhard RJ, Oostrom M, Dane JH (2004). A constitutive model for air-NAPL-water flow in the vadose zone accounting for immobile, non-occluded (residual) NAPL in strongly water-wet porous media. J Contam Hydrol 71(1–4):283-304. Available on http://dx.doi.org/10.1016/j.jconhyd.2004.07.005
33. Esrael D., (2015). Dépollution des Sols par l'Extraction Multiphasique MPE : Développement d'un Modèle Numérique. PhD thesis, University of Lyon INSA Lyon. 266p

A New Technique for Vibration-Based Diagnostics of Fatigued Structures Based on Damage Pattern Recognition via Minimization of Misclassification Probability

Nicholas A. Nechval and Konstantin N. Nechval

Abstract Vibration-based diagnostics provide various methods to detect, locate, and characterize damage in structural and mechanical systems by examining changes in measured vibration response. Research in vibration-based damage recognition has been rapidly expanding over the last few years. The basic idea behind this technology is that modal parameters (notably frequencies, mode shapes, and modal damping) are functions of the physical properties of the structure (mass, damping, and stiffness). Therefore, changes in the physical properties will cause detectable changes in the modal properties. In investigations, many techniques were applied to recognize damage in structural and mechanical systems, but recommendations on selecting the best technique for real monitoring systems are still insufficient and often contradictory. In the chapter presented, a novel technique of vibration-based diagnostics of fatigued structures based on damage pattern recognition through minimization of misclassification probability is proposed. This technique does not require the arbitrary selection of priors as in the Bayesian classifier and allows one to take into account the cases which are not adequate for Fisher's linear discriminant analysis (FLDA). The results obtained in this chapter agree with the simulation results, which confirm the validity of the theoretical predictions of performance of the suggested technique. A numerical example is given.

N.A. Nechval (✉)
BVEF Research Institute, University of Latvia,
Raina Blvd 19, Riga 1050, Latvia
e-mail: nechval@junik.lv

N.A. Nechval
Department of Mathematics, Baltic International Academy,
Lomonosov Street 4, Riga 1019, Latvia

K.N. Nechval
Department of Applied Mathematics, Transport and Telecommunication Institute,
Lomonosov Street 1, Riga 1019, Latvia
e-mail: konstan@tsi.lv

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*,
Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_7

Keywords Fatigued structure · Damage · Crack · Vibration-based diagnostics Pattern recognition · Misclassification probability · Minimization New technique · Gas turbine · Engine rotor · Disk · Turbine blades Damage detection · Measurement data · Modal parameters Frequency · Amplitude · Reference signature · Classification Separation threshold · Expected cost of misclassification

1 Introduction

Because of practical importance of fatigued structures such as bridges, highways, buildings, automobiles, air and space crafts, damage (crack) detection in structures has been the subject of intensive investigations in the last three decades. Many studies have been carried out in an attempt to find methods for nondestructive crack detection in structural members. Vibration-based techniques have been proved a fast and inexpensive means for crack detection (recognition) [1]. A crack in a structure induces a local flexibility which affects the dynamic behavior of the whole structure to a considerable degree. It results in reduction of measured frequencies and changes in mode shapes. An analysis of these changes makes it possible to detect (recognize) fatigue damages (cracks).

Dynamics of damaged rotors has been a subject of great interest for the last three decades, and detection and monitoring have gained increasing importance, recently. Failures of any high-speed rotating components (jet engine rotors, centrifuges, high-speed fans, etc.) can be very dangerous to surrounding equipment and personnel (see Fig. 1) and must always be avoided.

The ability to recognize a damaged component in aerospace, mechanical, and civil systems is becoming increasingly important [2–4]. Although extensive efforts have been devoted in the diagnostics and prognostics community to develop effective methods for damage diagnostics and recognition over the past few

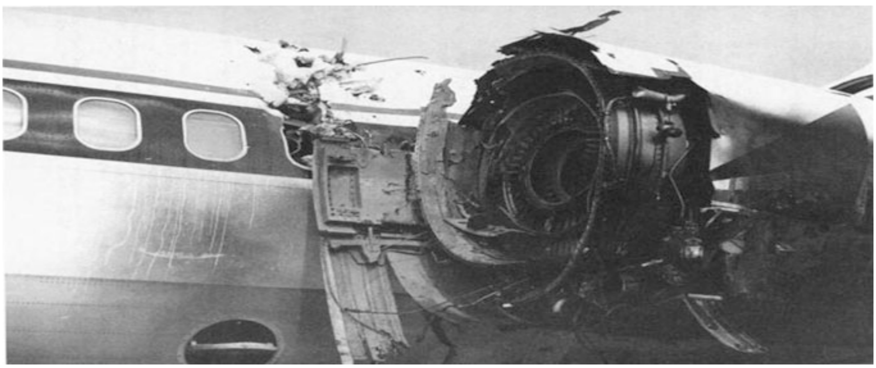


Fig. 1 Engine of Delta Air Lines Flight 1288 after it experienced catastrophic uncontained compressor rotor failure in 1996 (NTSB Report of Delta 1288, January 13, 1998)

decades, structural damage recognition is still regarded as a practical challenge for safety assurance of engineering structures [5].

The objective of this study is to develop an efficient probabilistic damage pattern recognition method for fatigued structure components with a minimal set of measurement data. The proposed method uses dynamical response measurement data. The damage pattern recognition problem is recasted to a probabilistic problem of minimization of misclassification probability, which is then solved using respective algorithms.

2 Problem of Vibration-Based Diagnostics of Gas Turbine Engine Rotor Disks

In this section, we look at a problem where vibration characteristics are used for gas turbine diagnostics. The present chapter focuses on turbine blade damage. Turbine blades undergo cyclic loading causing structural deterioration, which can lead to failure. Some recent studies have considered dynamical system approaches to model damage growth based on differential equations [6], while others have used physics-based models [7]. Selected studies have looked at modeling turbine blades as rotating Timoshenko beams with twist and taper [8–10]. Some studies have addressed damage in such beams using vibrational characteristics [11, 12].

Most rotor disks of gas turbine engine are limited by low-cycle fatigue (LCF) life, generally expressed in terms of mission equivalency cycles. The LCF life usually is defined as the number of cycles necessary to initiate a crack approximately 0.03 in. (0.76 mm) long. The distribution of LCF life is obtained for a given set of loading conditions (stress–strain, time, temperature). Traditionally, the LCF life prediction for gas turbine rotor disks is made for crack initiation life at an occurrence rate of l in 1000 disks. It is at this life that all LCF-limited disks are removed from service. This procedure has been very successful in preventing the occurrence of catastrophic failure of disks in the field. However, in retiring 1000 disks because 1 may fail, the remaining life of the 999 unfailed disks is not utilized.

An analysis methodology is developed in this chapter to recognize incurred quantifiable damage under which an engine disk is retired from service. The present methodology is based on the damage pattern classification technique minimizing misclassification probability.

3 Measurement Data Types from Vibration-Based Diagnostics

There are three main types of data obtained from analysis of vibration signals: time-domain analysis data, frequency-domain analysis data, and time-frequency analysis data. In time-domain analysis, features are extracted directly from the

Fig. 2 Measurement data of acceleration responses

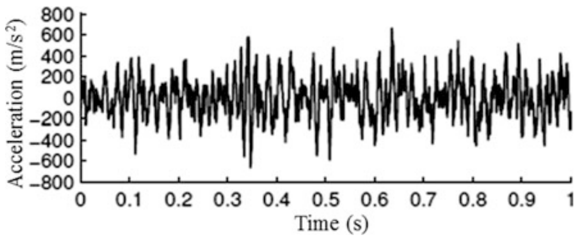


Fig. 3 Frequency components identified from acceleration measurement

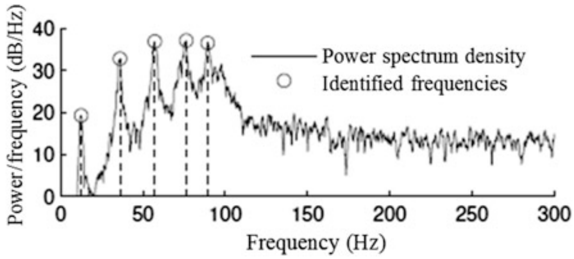
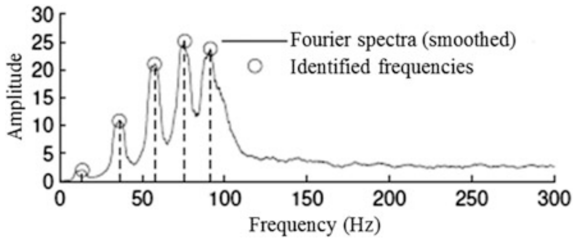


Fig. 4 Frequency components identified from the Fourier spectra (smoothed)



waveform time series [13]. In frequency-domain analysis, features are extracted through signal processing in the frequency domain [14]. Time-frequency analysis allows feature extraction from waveform type data in both the time and frequency domain [15, 16].

For illustration, measurement data of acceleration responses which can be obtained from vibration-based diagnostics are shown in Fig. 2.

Frequency components identified from acceleration measurement are shown in Fig. 3. The identification of frequency components can be made using peak-picking methods.

Frequency components identified from the corresponding Fourier spectra (smoothed) are shown in Fig. 4.

4 Damage Pattern Recognition via Vibration-Based Diagnostics

Damage pattern recognition (classification) via vibration-based diagnostics consists in the following. There are m classes (populations) of vibration signal, the elements (vibrational characteristics) of which are characterized by p measurements (features). Next, suppose that we are investigating a target vibration signal on the basis of the corresponding p measurements. We postulate that this signal can be regarded as a "random drawing" from one of the m populations (classes), but we do not know from which one. We suppose that m samples are available, each sample being drawn from a different class. The elements of these samples are realizations of p-dimensional random variables. After a sample of p-dimensional vectors of measurements of vibrational characteristics of the signal is drawn from a class known a priori to be one of the above set of m classes, the problem is to infer from which class the sample has been drawn. The decision rule should be in the form of associating the sample of observations on the target vibration signal with one of the m samples and declaring that this signal has come from the same class as the sample with which it is associated.

5 New Pattern Classification Technique for Vibration-Based Diagnostics Based on Minimization of Misclassification Probability

In this chapter, an efficient approach to pattern recognition (classification) is suggested. It is based on minimization of misclassification probability.

5.1 Pattern Classification into One of Two Classes

Let $(\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1})$ and $(\mathbf{y}_{21}, \ldots, \mathbf{y}_{2n_2})$ be samples of observed vectors of attributes of objects from two different classes C_1 and C_2, respectively. In this case, the proposed approach to pattern recognition (classification) into one of two classes C_1 and C_2 is given as follows.

Step 1 (*Transition from high-dimensional problem of pattern classification* (*dimension p* $\geq$ *2*) *to one-dimensional problem* (*dimension p* = *1*)). At this step, projection of high-dimensional data $(\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1})$ and $(\mathbf{y}_{21}, \ldots, \mathbf{y}_{2n_2})$ onto a line is carried out through a direction vector [17]

$$\mathbf{w} = \mathbf{S}_{12}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \tag{1}$$

where

$$\bar{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_1} \mathbf{y}_{1j}}{n_1}, \quad \bar{\mathbf{y}}_2 = \frac{\sum_{i=1}^{n_1} \mathbf{y}_{2j}}{n_2}, \tag{2}$$

$$\mathbf{S}_{12} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}, \tag{3}$$

$$\mathbf{S}_1 = \frac{\sum_{j=1}^{n_1} (\mathbf{y}_{1j} - \bar{\mathbf{y}}_1)(\mathbf{y}_{1j} - \bar{\mathbf{y}}_1)'}{n_1 - 1}, \quad \mathbf{S}_2 = \frac{\sum_{j=1}^{n_2} (\mathbf{y}_{2j} - \bar{\mathbf{y}}_2)(\mathbf{y}_{2j} - \bar{\mathbf{y}}_2)'}{n_2 - 1}. \tag{4}$$

We have:

$$(x_{11}, \ldots, x_{1n_1}) = (\mathbf{w}'\mathbf{y}_{11}, \ldots, \mathbf{w}'\mathbf{y}_{1n_1}) \tag{5}$$

and

$$(x_{21}, \ldots, x_{2n_2}) = (\mathbf{w}'\mathbf{y}_{21}, \ldots, \mathbf{w}'\mathbf{y}_{2n_2}). \tag{6}$$

The projection maximizes the distance between the means of the two classes while minimizing the variance within each class.

Step 2 (*Goodness-of-fit testing for statistical distributions*). At this step, the Anderson–Darling goodness-of-fit test is used to confirm goodness-of-fit of the two prespecified statistical distributions $f_{C_1}(x)$ and $f_{C_2}(x)$ with the data of (5) and (6), respectively.

Step 3 (*Determination of a separation threshold h*. At this step, the separation threshold h (see Fig. 5), which minimizes the probability of misclassification of a multivariate input observation $\mathbf{y}$ (target vibration signal) into two different classes C_1 and C_2,

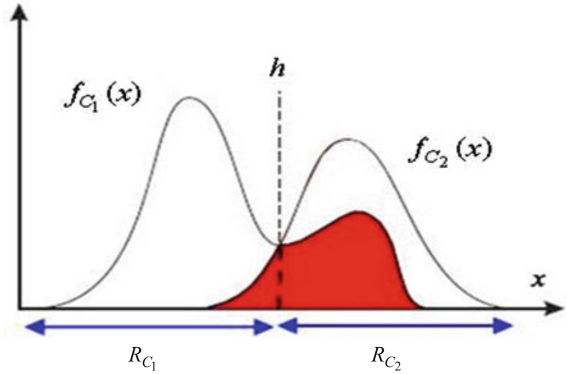
$$P(h) = \int_{R_{C_1}} f_{C_2}(x)dx + \int_{R_{C_2}} f_{C_1}(x)dx = \int_{\infty}^{h} f_{C_2}(x)dx + \int_{h}^{\infty} f_{C_1}(x)dx, \tag{7}$$

is determined via (7) as follows:

$$h = \arg\min_h P(h), \tag{8}$$

where $f_{C_j}(x)$ represents the probability density function (pdf) of an univariate input observation X from class C_j, $j \in \{1, 2\}$.

Fig. 5 Misclassification probability $P(h)$



Thus, h is used to obtain allocation of a multivariate input observation $\mathbf{y}$, which should be "optimal" in the sense of minimizing, on average, the number of incorrect assignments into two different classes C_1 and C_2.

Step 4 (*Pattern recognition* (*classification*) via *separation threshold h*). At this step, pattern recognition (classification) of the multivariate input observation $\mathbf{y}$ (target vibration signal) into one of two different classes C_1 and C_2 is carried out through $X = \mathbf{w}'\mathbf{y}$ as follows:

$$\mathbf{y} \in \begin{cases} C_1 & \text{if } X = \mathbf{w}'\mathbf{y} \leq h, \\ C_2 & \text{if } X = \mathbf{w}'\mathbf{y} > h. \end{cases} \tag{9}$$

Remark 1 The recognition (classification) rule (9) can be rewritten as follows:

$$\text{Assign } \mathbf{y} \text{ to class } C_j \text{ for which } f_{C_j}(x), j \in \{1, 2\}, \text{ is largest.} \tag{10}$$

5.2 Pattern Classification into One of Several Classes

Let $(\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}), \ldots, (\mathbf{y}_{m1}, \ldots, \mathbf{y}_{m\,n_m})$ be samples of observed vectors of attributes of objects from several different classes C_1, C_2, …, C_m, respectively. In this case, the proposed approach to pattern classification into one of several different classes C_1, C_2, …, C_m is given as follows.

Step 1 (*Transition from high-dimensional problem of pattern classification* (*dimension* $p \geq 2$) *to one-dimensional problem* (*dimension* $p = 1$)). At this step, projection of high-dimensional data $(\mathbf{y}_{k1}, \ldots, \mathbf{y}_{kn_k})$ and $(\mathbf{y}_{l1}, \ldots, \mathbf{y}_{ln_l})$, k, $l \in \{1, \ldots, m\}$, $k \neq l$, onto a line is carried out through a direction vector

$$\mathbf{w}_{kl} = \mathbf{S}_{kl}^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{y}}_l), \tag{11}$$

where

$$\bar{\mathbf{y}}_k = \frac{\sum_{i=1}^{n_k} \mathbf{y}_{ki}}{n_k}, \quad \bar{\mathbf{y}}_l = \frac{\sum_{i=1}^{n_l} \mathbf{y}_{li}}{n_l}, \tag{12}$$

$$\mathbf{S}_{kl} = \frac{(n_k - 1)\mathbf{S}_k + (n_l - 1)\mathbf{S}_l}{n_k + n_l - 2}, \tag{13}$$

$$\mathbf{S}_k = \frac{\sum_{i=1}^{n_k} (\mathbf{y}_{ki} - \bar{\mathbf{y}}_k)(\mathbf{y}_{ki} - \bar{\mathbf{y}}_k)'}{n_k - 1}, \quad \mathbf{S}_l = \frac{\sum_{i=1}^{n_l} (\mathbf{y}_{li} - \bar{\mathbf{y}}_l)(\mathbf{y}_{li} - \bar{\mathbf{y}}_l)'}{n_l - 1}. \tag{14}$$

We have:

$$(x_{k1}, \ldots, x_{kn_k}) = (\mathbf{w}_{kl}'\mathbf{y}_{k1}, \ldots, \mathbf{w}_{kl}'\mathbf{y}_{kn_k}) \tag{15}$$

and

$$(x_{l1}, \ldots, x_{ln_l}) = (\mathbf{w}_{kl}'\mathbf{y}_{l1}, \ldots, \mathbf{w}_{kl}'\mathbf{y}_{ln_l}). \tag{16}$$

The projection maximizes the distance between the means of the two classes while minimizing the variance within each class.

Step 2 (*Goodness-of-fit testing for statistical distributions*). At this step, the Anderson–Darling goodness-of-fit test is used to confirm goodness-of-fit of the prespecified statistical distributions $f_{C_k}^{kl}(x)$ and $f_{C_l}^{kl}(x)$, $k, l \in \{1, \ldots, m\}$, $k \neq l$, with the data of (15) and (16), respectively.

Step 3 (*Pattern recognition (classification)* via *statistical distributions*). At this step, pattern recognition (classification) of the multivariate input observation $\mathbf{y}$ (target vibration signal) into one of several different classes C_1, C_2, …, C_m is carried out through $X = \mathbf{w}_{kl}'\mathbf{y}$ as follows:

$$\mathbf{y} \in C_k \ \text{ if } \ f_{C_k}^{kl}(x) \geq f_{C_l}^{kl}(x), \quad \forall l \neq k. \tag{17}$$

6 Numerical Example

In order to show effectiveness of the proposed new technique for vibration-based diagnostics of fatigued structures based on damage pattern recognition via minimization of misclassification probability, a gas turbine rotor disk is considered (see Fig. 6).

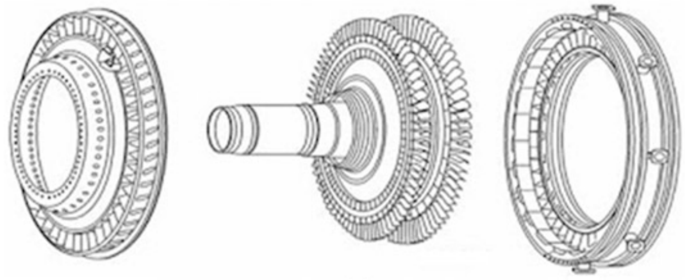


Fig. 6 High-pressure turbine assembly

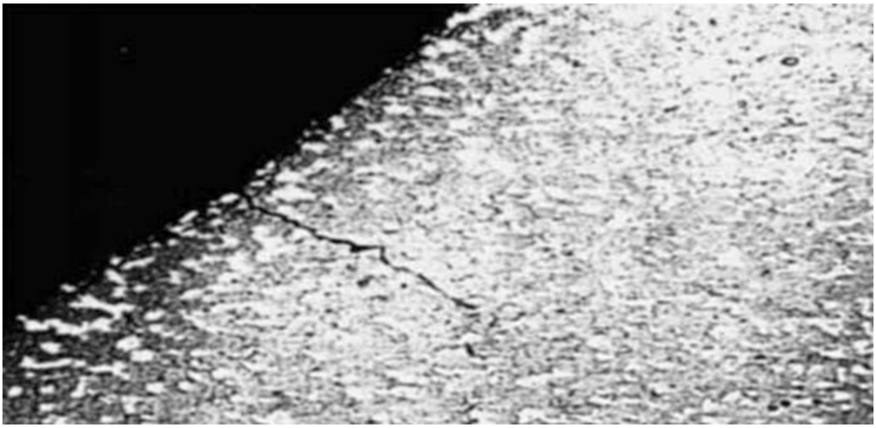


Fig. 7 Micrograph showing one of the cracks detected in the high-pressure turbine (bladed disk assembly)

In Fig. 7, micrograph showing one of the cracks detected in the bladed disk assembly of the high-pressure turbine for over 17,000 cycles is presented.

Schematic showing the orientation of the columnar grains at the leading and trailing edges in the blade root region of the failed turbine rotor blisk is presented in Fig. 8.

Table 1 shows, for illustration, two groups of analysis data from vibration-based diagnostics (frequency components identified from the corresponding Fourier spectra (smoothed)).

Random samples of n_1 = 7 disks from class C_1 and n_2 = 7 disks from class C_2 give the values of frequency components identified from the corresponding Fourier spectra (smoothed).

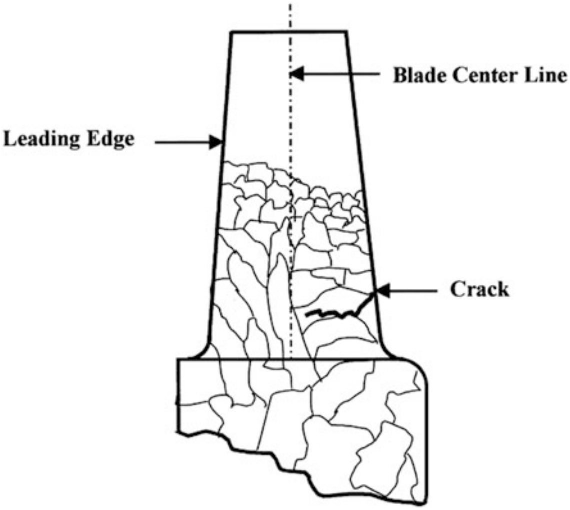


Fig. 8 Schematic showing the orientation of the columnar grains at the leading and trailing edges in the blade root region of the failed turbine rotor blisk

Table 1 Analysis data from vibration-based diagnostics (frequency components identified from the corresponding Fourier spectra (smoothed))

Class C_1: Disks have incurred detectable damage (crack)		Class C_2: Disks do not have incurred detectable damage (crack)	
Frequency components			
y_{11} (Frequency in Hz)	y_{12} (Amplitude)	y_{21} (Frequency in Hz)	y_{22} (Amplitude)
45.5	36.5	72.7	37.6
54.6	36.6	74.6	37.5
65.6	36.7	77.3	39.3
67.6	36.8	77.1	39.4
68.8	36.8	78	39.8
69.9	36.9	74.8	41
69	37.1	76.8	41.1

6.1 Constructing a Procedure for Detecting a Potential Damaged Disk

Using the analysis data, which are given in Table 1, the damage detection procedure can be constructed as follows.

Step 1 (*Transition from high-dimensional problem of detection of damaged disk* (*dimension* $p = 2$) *to one-dimensional problem* (*dimension* $p = 1$).

$$\mathbf{w} = \mathbf{S}_{12}^{-1}(\bar{\mathbf{y}}_2 - \bar{\mathbf{y}}_1) = \begin{bmatrix} 0.023138 & -0.03253 \\ -0.03253 & 0.990408 \end{bmatrix} \begin{bmatrix} 12.9 \\ 2.614286 \end{bmatrix} = \begin{bmatrix} 0.213432486 \\ 2.169553201 \end{bmatrix}, \tag{18}$$

where

$$\bar{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_1} \mathbf{y}_{1i}}{n_1} = \begin{bmatrix} 63 \\ 36.77143 \end{bmatrix}, \quad \bar{\mathbf{y}}_2 = \frac{\sum_{i=1}^{n_1} \mathbf{y}_{2i}}{n_2} = \begin{bmatrix} 75.9 \\ 39.38571 \end{bmatrix}, \tag{19}$$

$$\mathbf{S}_{12} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} = \begin{bmatrix} 45.31167 & 1.488333 \\ 1.488333 & 1.058571 \end{bmatrix}, \tag{20}$$

$$\mathbf{S}_1 = \frac{\sum_{i=1}^{n_1} (\mathbf{y}_{1i} - \bar{\mathbf{y}}_1)(\mathbf{y}_{1i} - \bar{\mathbf{y}}_1)'}{n_1 - 1} = \begin{bmatrix} 86.99667 & 1.526667 \\ 1.526667 & 0.039048 \end{bmatrix}, \tag{21}$$

$$\mathbf{S}_2 = \frac{\sum_{i=1}^{n_2} (\mathbf{y}_{2i} - \bar{\mathbf{y}}_2)(\mathbf{y}_{2i} - \bar{\mathbf{y}}_2)'}{n_2 - 1} = \begin{bmatrix} 3.626667 & 1.45 \\ 1.45 & 2.078095 \end{bmatrix}. \tag{22}$$

We have:

$$(x_{11}, \ldots, x_{1n_1}) = (\mathbf{w}'\mathbf{y}_{11}, \ldots, \mathbf{w}'\mathbf{y}_{1n_1}) \tag{23}$$

and

$$(x_{21}, \ldots, x_{2n_2}) = (\mathbf{w}'\mathbf{y}_{21}, \ldots, \mathbf{w}'\mathbf{y}_{2n_2}). \tag{24}$$

The data of (23) and (24) are given in Table 2.

Table 2 Projection of the data (p = 2) of Table 1 onto a line (p = 1)

Class C_1: Disks have incurred detectable damage (crack)	Class C_2: Disks do not have incurred detectable damage (crack)
x_1	x_2
88.89987	97.09174
91.05906	97.28031
93.62377	101.7618
94.26759	101.936
94.52371	102.996
94.97544	104.9164
95.21727	105.5603

Step 2 (*Goodness-of-fit testing for statistical distributions*).

Using the Anderson–Darling goodness-of-fit test for normality (significance level $\alpha = 0.05$), it was found that

$$f_{C_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right), \tag{25}$$

$$\mu_1 = \widehat{\mu}_1^{\mathrm{ML}} = \sum_{i=1}^{n_1} x_{2i} \Big/ n_1 = 93.22382, \tag{26}$$

$$\sigma_1 = \widehat{\sigma}_1^{\mathrm{ML}} = \left(\sum_{i=1}^{n_1} (x_{1i} - \widehat{\mu}_1^{\mathrm{ML}})^2 \Big/ n_1\right)^{1/2} = 2.358103, \tag{27}$$

$$f_{C_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right), \tag{28}$$

$$\mu_2 = \widehat{\mu}_2^{\mathrm{ML}} = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2} = 101.6489, \tag{29}$$

$$\sigma_2 = \widehat{\sigma}_2^{\mathrm{ML}} = \left(\frac{\sum_{i=1}^{n_2} (x_{2i} - \widehat{\mu}_2^{\mathrm{ML}})^2}{n_2}\right)^{1/2} = 3.359996, \tag{30}$$

Step 3 (*Determination of the separation threshold h*).

Theorem 1 If

$$f_{C_j}(x), \quad j = 1, 2, \tag{31}$$

are the probability density functions of the normal distribution with the parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, where $\mu_1 < \mu_2$, then the necessary and sufficient conditions for h to be a minimum point of the misclassification probability,

$$P(h) = \int_{-\infty}^{h} f_{C_2}(x)dx + \int_{h}^{\infty} f_{C_1}(x)dx, \tag{32}$$

are given by

(i) the necessary condition:

$$f_{C_1}(h) = f_{C_2}(h), \tag{33}$$

(ii) the sufficient condition:

$$\mu_1 < h < \mu_2. \tag{34}$$

Proof. The proof being straightforward is omitted here.

Corollary 1.1 If $\sigma_1 = \sigma_2$, then the separation threshold h is determined as

$$h = (\mu_1 + \mu_2)/2, \tag{35}$$

For example, in this case we deal with Fisher's separation threshold [17].
It follows from (33) that

$$h = 97.02663, \quad P(h) = 0.13786, \tag{36}$$

It follows from (35) that

$$h_{\text{Fisher}} = (\mu_1 + \mu_2)/2 = 97.43637, \quad P(h_{\text{Fisher}}) = 0.14198. \tag{37}$$

Indexes: In this case, the index of relative efficiency of Fisher's approach as compared with the proposed approach is

$$I_{\text{rel.eff.}}(h_{\text{Fisher}}, h) = P(h)/P(h_{\text{Fisher}}) = 0.13786/0.14198 = 0.97. \tag{38}$$

The index of reduction percentage in the probability of misclassification for the proposed approach as compared with Fisher's approach is given by

$$I_{\text{red.per.}}(h, h_{\text{Fisher}}) = (1 - I_{\text{rel.eff.}}(h_{\text{Fisher}}, h))100\% = 3\%. \tag{39}$$

Step 4 (*Pattern recognition* (*classification*) via *separation threshold h*).

For instance, consider the following data from Table 2:

$$x_{21} = 97.09174, \quad x_{22} = 97.28031. \tag{40}$$

Should these data be classified as C_1 (disks have incurred detectable damage (crack)) or C_2 (disks do not have incurred detectable damage (crack))?
Using Fisher's classification rule, we obtain:

$$x_{21} = 97.09174 < h_{\text{Fisher}} = 97.43637, \quad x_{22} = 97.28031 < h_{\text{Fisher}} = 97.43637. \tag{41}$$

Thus, x_{21} and x_{22} are incorrectly classified as C_1, i.e., in this case, Fisher's classification rule gives misclassification.

Using the proposed approach based on minimization of misclassification probability, we obtain:

$$x_{21} = 97.09174 > h = 97.02663, \quad x_{22} = 97.28031 > h = 97.02663. \tag{42}$$

Thus, x_{21} and x_{22} are correctly classified as C_2.

7 Conclusion

In this chapter, classification schemes are evaluated in terms of their misclassification probabilities, but this ignores misclassification cost. For example, even a seemingly small probability such as, say, 0.05 may be too large if the cost of making an incorrect assignment to class C_1 or C_2 is extremely high. A rule that ignores costs may cause problems.

The costs of misclassification can be defined as follows. The costs are: (1) zero for correct classification, (2) $c_{(1|2)}$ when an observation from C_2 is incorrectly classified as C_1, and (3) $c_{(2|1)}$ when a C_1 observation is incorrectly classified as C_2.

For any classification rule, the average or *expected cost of misclassification* (ECM) is given by

$$\mathrm{ECM} = c_{(1|2)} \int\limits_{R_{C_1}} f_{C_2}(x)dx + c_{(2|1)} \int\limits_{R_{C_2}} f_{C_1}(x)dx. \tag{43}$$

A reasonable classification rule should have an ECM as small, or nearly as small, as possible.

Theorem 2 The regions R_{C_1} and R_{C_2} that minimize the ECM are defined by the values x for which the following inequalities hold:

$$R_{C_1}: \frac{f_{C_1}(x)}{f_{C_2}(x)} \geq \frac{c_{(1|2)}}{c_{(2|1)}} \tag{44}$$

$$\begin{pmatrix} \text{density} \\ \text{ratio} \end{pmatrix} \geq \begin{pmatrix} \text{cost} \\ \text{ratio} \end{pmatrix}$$

$$R_{C_2}: \frac{f_{C_1}(x)}{f_{C_2}(x)} < \frac{c_{(1|2)}}{c_{(2|1)}} \tag{45}$$

$$\begin{pmatrix} \text{density} \\ \text{ratio} \end{pmatrix} < \begin{pmatrix} \text{cost} \\ \text{ratio} \end{pmatrix}$$

Proof. Noting that

$$\Omega = R_{C_1} \cup R_{C_2}, \tag{46}$$

so that the total probability

$$1 = \int_{\Omega} f_{C_2}(x)dx = \int_{R_{C_1}} f_{C_2}(x)dx + \int_{R_{C_2}} f_{C_2}(x)dx, \tag{47}$$

we can write

$$\mathrm{ECM} = c_{(1|2)}\left(1 - \int_{R_{C_2}} f_{C_2}(x)dx\right) + c_{(2|1)} \int_{R_{C_2}} f_{C_1}(x)dx. \tag{48}$$

By the additive property of integrals,

$$\mathrm{ECM} = \int_{R_{C_2}} [c_{(2|1)} f_{C_1}(x) - c_{(1|2)} f_{C_2}(x)]dx + c_{(1|2)}. \tag{49}$$

Now, $c_{(1|2)}$ and $c_{(2|1)}$ are nonnegative. In addition, $f_{C_1}(x)$ and $f_{C_2}(x)$ are nonnegative for all values of x and are the only quantities in ECM that depend on x. Thus, ECM is minimized if R_{C_2} includes those values of x for which the integrand

$$[c_{(2|1)} f_{C_1}(x) - c_{(1|2)} f_{C_2}(x)] < 0 \tag{50}$$

and excludes those values of x for which this quantity is nonnegative. This completes the proof.

Corollary 2.1 If

$$c_{(1|2)}/c_{(2|1)} = 1 \text{ (Equal misclassification costs)} \tag{51}$$

then

$$R_{C_1} : \frac{f_{C_1}(x)}{f_{C_2}(x)} \geq 1, \quad R_{C_2} : \frac{f_{C_1}(x)}{f_{C_2}(x)} < 1. \tag{52}$$

References

1. Dimarogonas AD (1996) Vibration of cracked structures: a state of the art review. Eng Fract Mech 55:831–857
2. Fan W, Qiao P (2011) Vibration-based damage identification methods: a review and comparative study. Struct Health Monit 10:83–111

3. Nechval NA, Nechval KN, Berzins G, Purgailis M (2007) CFAR vibration signal change test and its applications to real-time recognition of developing cracks in jet engine rotors. J VibroEng 9:49–52
4. Nechval NA, Nechval KN, Bausova I (2014) New procedures of pattern classification for vibration-based diagnostics via neural network. In: Golovko V, Imada A (eds) Neural networks and artificial intelligence. ICNNAI 2014. Communications in computer and information science, vol 440. Springer, International Publishing Switzerland, pp 63–75
5. Jardine A, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510
6. Adams DE, Nataraju M (2002) A nonlinear dynamics system for structural diagnosis and prognosis. Int J Eng Sci 40:1919–1941
7. Roy N, Ganguli R (2005) Helicopter rotor blade frequency evolution with damage growth and signal processing. J Sound Vib 283:821–851
8. Krupka R, Baumanis AM (1965) Bending-bending mode of rotating tapered twisted turbo machine blades including rotary inertia and shear deflection. J Eng Ind 91:10–17
9. Thomas J, Abbas BAH (1975) Finite element model for dynamic analysis of Timoshenko beam. J Sound Vib 41:291–299
10. Rao SS, Gupta RS (2001) Finite element vibration analysis of rotating Timoshenko beams. J Sound Vib 242:103–124
11. Takahashi I (1999) Vibration and stability of non-uniform cracked Timoshenko beam subjected to follower force. Comput Struct 71:585–591
12. Hou J, Wicks BJ, Antoniou RA (2000) An investigation of fatigue failures of turbine blades in a gas turbine engine by mechanical analysis. Eng Fail Anal 9:201–211
13. Zhan Y, Makis V, Jardine AKS (2003) Adaptive model for vibration monitoring of rotating machinery subject to random deterioration. J Qual Maint Eng 9:351–375
14. Goldman S (1999) Vibration spectrum analysis: a practical approach. Industrial Press, New York
15. Peng ZK, Chu FL (2004) Application of the wavelet transform in machine condition monitoring and fault diagnosis: a review with bibliography. Mech Syst Signal Process 18:199–221
16. Aretakis N, Mathioudakis K (1997) Wavelet analysis for gas turbine fault diagnostics. ASME J Eng Gas Turbine Power 119:870–876
17. Fisher R (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7:178–188

Statistical Techniques of Reliability, Availability, and Maintainability (RAM) Analysis in Industrial Engineering

Panagiotis Tsarouhas

Abstract This chapter introduces the basic statistical concepts of reliability, availability, and maintainability (RAM) analysis that used in industrial engineering. RAM are three system attributes that are of tremendous interest to systems engineers, and have significant impacts on the sustainment or total life cycle costs of a developed system. To analyze and measure RAM characteristics of a system, there must be a mathematical model of the system that shows the functional relationships among all the components, the subsystems, and the overall system. The analysis is a tool to evaluate the current operations management of the industrial system. Thus, engineers and managers will be able to make the right decisions about the system, optimizing efficiency, and productivity while reducing production costs.

1 Introduction

The good quality of a product requires the good quality of raw material that is used for the production of the product from one side and the appropriate operation of the machines that is required for its production by the other. In engineering and statistics, the reliability has a distinct significance that can be calculated, estimated, measured, and be drawn. Thus, the reliability is a concrete concept that is classified as the efficiency, the productivity, and the quality of the equipment or the system.

The reliability field presents the early 1930s when the probability concepts were applied to problems associated with electric power generators. However, the real application of reliability engineering may be traced back to World War II, when the Germans are reported to have first introduced the reliability concept to improve the reliability of their V1 and V2 rockets [1–3]. From the beginning of the industrial season, the problem of reliability was important; a classic example are the bearings (ball and roller bearings), where extensive studies have been carried out on the

P. Tsarouhas (✉)
Department of Supply Chain Management & Logistics, Technological Educational Institute of Central Macedonia, Kanellopoulou 2, 60100 Katerini, Hellas, Greece
e-mail: ptsarouhas@teicm.gr

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*, Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_8

characteristics of their lifetime from the first days of transport within the railway. Also, together with the growth and spread of airplanes began the reflections and the need of resolving the problems of reliability that are related not only to the equipment that is transported, but also to the persons that use the planes as means of transport.

The reliability entered a new season with the spread of computers, airplanes that fly in sound and over sound speeds, rockets, and spatial vehicles. Certain reliable products are produced by R&D teams and growth that daily applies the traditional beginnings of attachment in the experience of the observation for high quality. Today, however, a lot of pressures exist that dispute the effectiveness of traditional beginnings, i.e., competition, short time of delivery, new materials, reduction of cost, requirements of safety, that tend to increase the danger of the growth of a new product or system. The reliability engineering has been developed as an answer to the need for reduction or for restriction of these pressures with a scope to the improvement of the quality of the final products.

The beginning of the maintainability field may be traced back to 1901, to the United States Army Signal Corps contract for the development of the Wright brothers' airplane. In this document, it was clearly stated that the aircraft should be "simple to operate and maintain" [4]. In the modern context, the beginning of the maintainability discipline may be traced back to the period between World War II and the 1950s, when various efforts directly or indirectly concerned with maintainability were initiated [5]. In 1960, the US Air Force (USAF) initiated a program for developing an effective systems approach to maintainability that ultimately resulted in the development of maintainability specification MIL-M-26512 [6]. Many other military documents concerning maintainability appeared in the latter part of the 1960s [7, 8].

2 Comparing the Reliability with Quality Control

The reliability engineering ensures that the components, the equipment, and the entire system will operate reliably for a desirable period and will be maintained at the time interval of its lifetime that is from its development until its withdrawal. The quality control is the "*vital contact*" for the total circle of reliability. In other words, it initially ensures the equitable planning of components, equipment, and systems and then the equitable way of assembly for these during the production process. The quality control ensures that the components are produced inside the specifications and that the number of defective items is found in their lower level of acceptance. Also, the final products (components) are checked for the conformation with their specifications and then they are dispatched to use. The aim of control checked is to realize the actual quality level which should be bigger than determining. Also, no product is considered reliable if it has not passed through the process of the quality control [9].

Despite the relationship that presents the reliability with quality control, there are also differences that characterize:

1. The reliability deals with the behavior of the failure rate for a sufficiently long interval of operation, while the quality control deals with the percentage of defective items for a concrete moment of time.
2. The reliability deals in detail with the operation time of components and products, focusing basically on the phase of planning, while the quality control deals mainly with the phase of production for a product.
3. The reliability deals with the concepts of planning and methodology that ensure the biggest reliability of items, components, subsystems, or systems, while the quality control deals with the equitable and successful way of transformation of specifications and planning in good items, products, components, or systems.

The reliability ensures a high level of planning, the optimal reliability in a product and records all the phases of its lifetime, while the quality control ensures that the reliability is not degraded during the phase of processing [10].

The quality control guarantees that during the production process, the products will be uniform and will be medium in the forecasted specifications. Also, it will be supposedly checked each stage of the productive process, the variability of raw material, etc., so as to have the best quality of products with the most optimal cost for the consumer. Consequently, the process of quality control during the production process is considered the capable and necessary treaty so that are produced reliable products.

3 The Bathtub Curve

Each component, equipment, or system that is found in operational state, in a certain moment of time will fail. Failure is considered to be the condition to which it devolves a component, equipment, or system when it cannot achieve its mission. Characteristic types of failures are usually presented, excluding the damage that was caused by careless handling, storage, or inadequate operation by the users.

The failure behavior of various engineering components is called hazard rate function shown in Diagram 1 and represents the hazard rate of such components as a function of time. Because of its shape, it is known as the "*bathtub curve*" and is divided into three regions, i.e., I, II, and III.

Region I is known as *burn-in* that has decreasing failure rate because failures are caused by manufacturing defects, poor workmanship, and poor quality control.

Region II is characterized as the *useful life*, during which the component failure rate remains constant due to random loads, human error, natural failures, and abuse. Region III is called *wear-out* that has increasing failure rate because failures are caused by fatigue, aging, corrosion, and friction.

From the bathtub curve, it is obvious that the failure rate of the components or the systems is different according to the failure mode. Thus, in the burn-in region the failure rate is decreased, and in the useful life region, the failure rate is constant, whereas in the wear-out region, the failure rate is increased. Therefore, each failure mode follows a different failure distribution and requires different mathematic methods for its minimization [11].

4 Uptime, Downtime, and Repair Time

Uptime is the period during which the system or the equipment is in the operational state, whereas the downtime is the period during which the system or the equipment is in the failed state. The total operating time for the system is the sum of uptime and downtime.

Repair usually is referred to checkout or alignment which may extend beyond the downtime. The process of replacement/repair and reintroduction to the initial functional can be analyzed in several activities and delay times, as they appear in Fig. 1[12, 13]:

(i) *Identification time*: it is the time from the moment when the failure occurs until it is realized by the special technical personnel.

(ii) *Finding time*: it is the time that is required in order to find in which point of equipment exists a failure. It includes the time of unassembly of protective covers and safety protection for equipment from the technical personnel.

(iii) *Diagnosis time*: it is reported in the time that is needed in order to diagnose the failure type of equipment. The failure type can be mechanic, electric, electronic, etc.

(iv) *Spare part time*: it is the time that is needed for the finding and transport of necessary spare part(s) for the reparation of the failure.

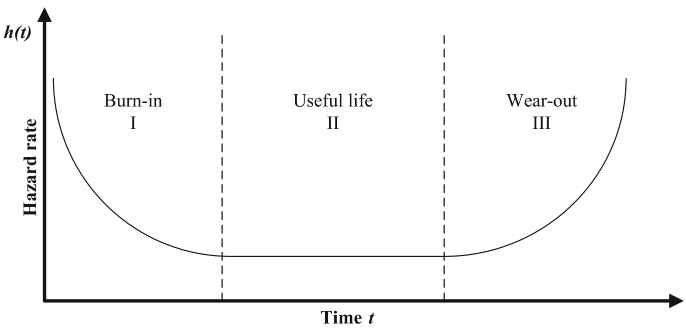


Fig. 1 Activities for downtime and repair time

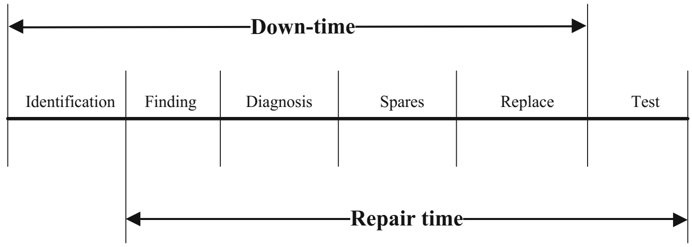


Fig. 2 Block diagram of an n-components series system

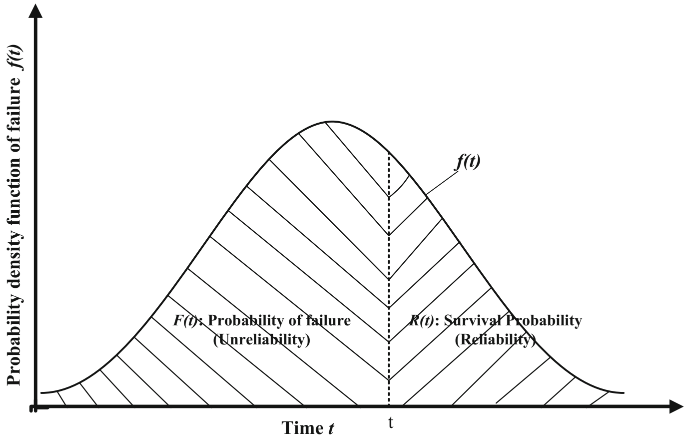


Fig. 3 Block diagram of an n-components parallel system

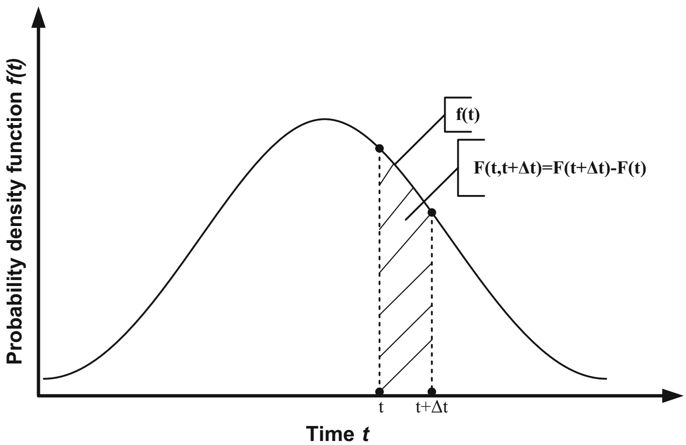


Fig. 4 Illustration of point, average, and asymptotic availability as a function of operating time

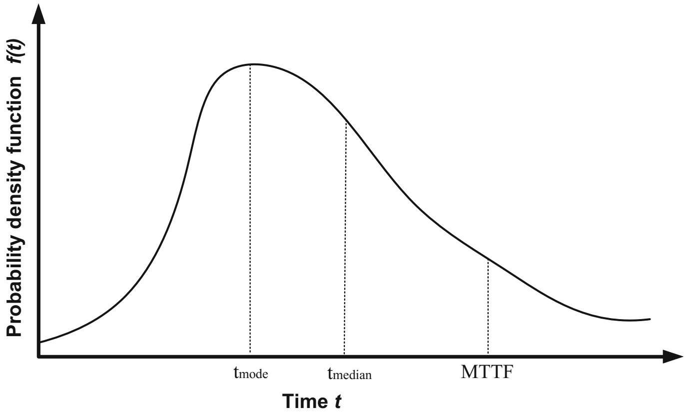


Diagram 1 Bathtub hazard rate curve

(v) *Replacement or repair time*: it is the required time for the repair or replacement of the failed component of the equipment.
(vi) *Test time*: it is the time that is required for adjustment and alignment of the repaired equipment before the mass production. It also contains the time of assembly for the protective covers and the safety protection of the equipment.

Those activities are called *active repair activities* that are included in time to repair.

In addition, the *passive repair activities* exit which are depended on for the maintenance strategy of the company:

(a) *Logistics time:* it is the necessary time waiting for the spare parts, additional technological equipment, and technicians to repair the failure of the system.
(b) *Administration time:* it is the time that is depended on for the operations of the system by the company, i.e., maintenance staff, spare parts level.

5 Reliability Function

Reliability engineering deals with the minimization of the premature failures observing their distributions. It also deals with the prediction of failure deterioration observing the statistical distribution of deterioration and determining the periods of preventive maintenance and replacement of components. Finally, it focuses on random failures and their forecast, aiming at their reduction or still better at their total obliteration. The reliability is defined as follows: *reliability is the probability a component, equipment, or a system realizes its predetermining mission without failures, for a certain time interval, when it functions correctly in a determined environment.*

The reliability is a probability expressed via the relative failures' frequency of a component or system. When it represents the population, it is replaced by the term of probability density function (*PDF*) which is the more basic function in theory of reliability and as the function of time is characterized as $f(t)$ Diagram 2.

Reliability is defined in terms of probability such as random variable, probability density functions, and cumulative distribution functions. The reliability $R(t)$ can be expressed as [14, 15]:

$$R(t) = P(T \geq t) = \int_{t}^{\infty} f(t)dt, \tag{1}$$

where $R(t) \geq 0, R(0) = 1$ and is defined as the probability that the system or component will function over some time period t, where T is a random variable that represents the *time to failure* of the component or system.

The cumulative distribution function (*CDF*) or failure probability$F(t)$ is the "unreliability function" and is defined as the probability that the *time to failure* is no greater than t, or

$$F(t) = 1 - R(t) = P(T \ < t) = \int_{0}^{t} f(t)dt, \tag{2}$$

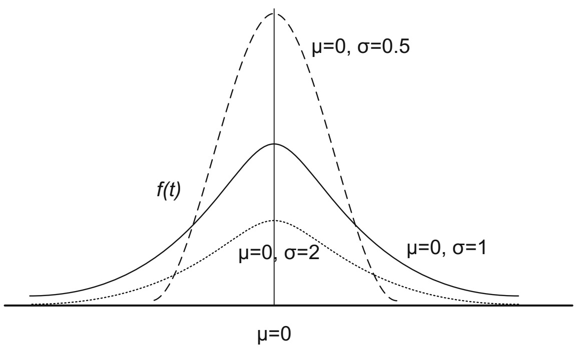
where $F(t) \geq 0, F(0) = 0$.

Therefore, the $F(t)$ is the probability that the failure occurs prior to some time t.

The PDF, $f(t)$, has the following properties:

$$f(t) \geq 0 \, \kappa\alpha\iota \int_{0}^{\infty} f(t)dt = 1.$$

Diagram 2 Survival probability *R(t)* as complement of probability failure *F(t)*



6 Failure and Hazard Rate Function

The *PDF*, $f(t)$, describes the shape of the failure distribution and is given by Ref. [16]:

$$f(t) = \frac{dF(t)}{dt} = \frac{d[1 - R(t)]}{dt} = -\frac{dR(t)}{dt}. \tag{3}$$

The failure probability in time interval $[t_1, t_2]$ can be expressed in terms of reliability as

$$R(t_1) - R(t_2) = \int_{t_1}^{\infty} f(t)dt - \int_{t_2}^{\infty} f(t)dt. \tag{4}$$

The conditional probability of a failure in the time interval $[t_1, t_2]$, given that the system has survived to time t_1, is

$$P(\frac{t_1 \le T \le t_2}{T \ge t_1}) = \frac{R(t_1) - R(t_2)}{R(t_1)}. \tag{5}$$

The condition probability of failure per unit time is defined as failure rate, $\lambda(t)$ and is expressed as follows:

$$\lambda(t) = \frac{R(t_1) - R(t_2)}{(t_2 - t_1) \times R(t_1)}. \tag{6}$$

If $t = t_1$ and $t_2 = t + \Delta t$, then Eq. (6) becomes

$$\lambda(t) = \frac{R(t) - R(t + \Delta t)}{\Delta t \times R(t)} = \frac{F(t + \Delta t) - F(t)}{\Delta t \times R(t)}. \tag{7}$$

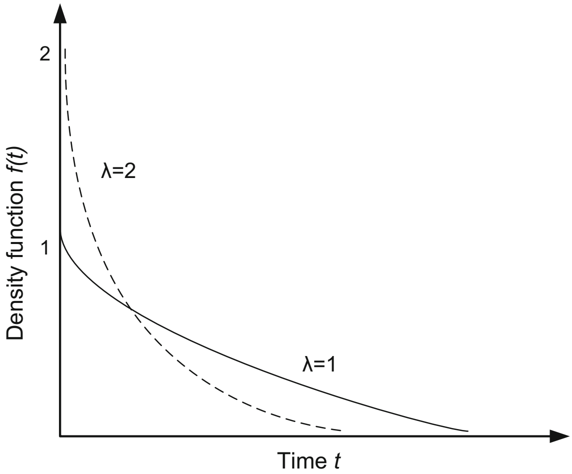
The condition probability is equal to the outline area under the probability function $f(t)$, and between t and $t + \Delta t$ (see Diagram 3) given that it has survived up to the time t.

The instantaneous failure rate or hazard rate function,$h(t)$, is defined as $\Delta t \to 0$ as

$$h(t) = \lim_{\Delta t \to 0} [\frac{R(t) - R(t + \Delta t)}{\Delta t \times R(t)}] = \lim_{\Delta t \to 0} \frac{-[R(t + \Delta t) - R(t)]}{\Delta t} \times \frac{1}{R(t)} = \frac{-dR(t)}{dt} \times \frac{1}{R(t)} = \frac{f(t)}{R(t)}, \tag{8}$$

which is the relationship between the density function of the time to failure and the reliability function with the hazard rate function at any time t. Equation (8) is very important because it is independent of the statistical distribution under consideration.

Diagram 3 Probability density function $f(t)$ (*PDF*)



Therefore the failure rate represents the average rate in a time interval whereas the hazard rate represents any point of time in an interval.

From Eq. (8), it is also easy to express the reliability function as

$$R(t) = \exp[-\int_0^t h(t)dt]. \tag{9}$$

7 Main Characteristics of Reliability

In addition, several other basic concepts of reliability may require development, i.e., the mean time to failure, mean time between failure, median, mode, and the 100Pth percentile.

Mean Time To Failure

The mean time to failure (*MTTF*) is the expected value of time without failure which refers to the lifetime of a non-repairable system and is defined by

$$MTTF = E(T) = \int_0^\infty t \times f(t)dt = \int_0^\infty R(t)dt. \tag{10}$$

From Eq. (3),

$$MTTF = \int_{0}^{\infty} \frac{-dR(t)}{dt} t dt = \int_{0}^{\infty} R(t) dt. \tag{11}$$

Mean Time to the First Failure and Mean Time Between Failure

The mean time to the first failure (*MTTFF*) describes the mean lifetime of a reparable component until its first failure. Therefore, *MTTFF* corresponds to the *MTTF* for non-repairable components.

The mean time between failure (*MTBF*) describes the lifetime after the first failure of the repairable component which is the mean lifetime until its next failure.

The median of time to failure

The median time to failure, t_{median}, divides the failure distribution into two halves where 50 percent of all the failures occurring before and the remaining 50 percent of the failures occurring after the media and is expressed as

$$R(t_{\text{median}}) = 0,5 = P(T \geq t_{\text{median}}) \tag{12}$$

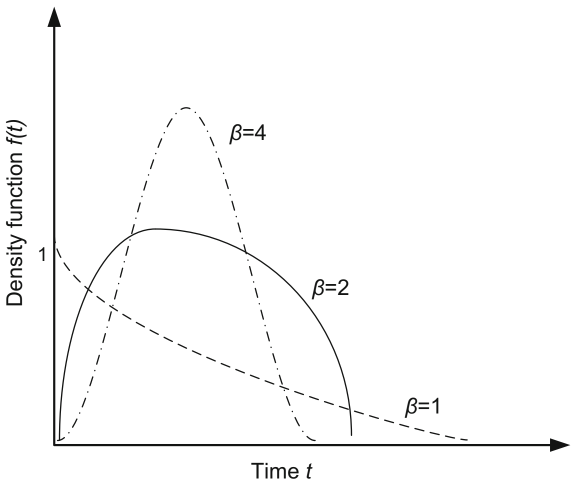
The mode of time to failure

The mode, t_{mode}, of the failure distribution is the maximum observed value of failure and is defined by

$$f(t_{\text{mode}}) = \max f(t), \quad \text{for } 0 \leq t \leq \infty. \tag{13}$$

Diagram 4 is the graphical presentation of the MTTF, t_{median}, and t_{mode} for a left symmetrical distribution.

Diagram 4 Presentation of basic concepts of reliability



The 100Pth percentile

The 100Pth percentile of a cumulative distribution functions $F(t)$ is the point in time t_p by which a proportion p of the population fails, mathematically it is expressed as

$$p = F(t_p). \tag{14}$$

8 Lifetime Distributions for Reliability

There are many probability distributions that have been developed to perform various types of statistical analysis. This section presents some useful theoretical continuous distributions to describe the lifetime of reliability and maintainability. Each probability distribution is characterized by the parameters, which in their own way control the distribution. These parameters can usually be classified into three categories:

1. *Shape parameter* that influences the shape of the distribution.
2. *Scale parameter* that influences both the mean and the spread of the distribution.
3. *Location or Source parameter* that locates the distribution along the abscissa.

Not all the probability distributions have three parameters; there are many distributions that have one or two parameters.

Normal distribution

The normal distribution which is also known as Gaussian distribution is the most extensively covered theoretical distribution in literature. It has the PDF,$f(t)$, perfectly symmetrical about the mean μ, which means that the mean, median, and the mode have the same numerical value or $\mu = t_{\text{median}} = t_{\text{mode}}$.

The location parameter μ in the normal distribution locates the distribution on the horizontal axis and represents the mean. The σ is the scale parameter and influences the range of the distribution and represents the standard deviation of the distribution. Diagram 5 shows the PDFs for different values of σ. The location parameter μ and the scale parameter σ have the same measurement units as lifetime, i.e., minutes, hours, cycles.

The equations that describe the normal distribution are:

(i) PDF:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]. \tag{15}$$

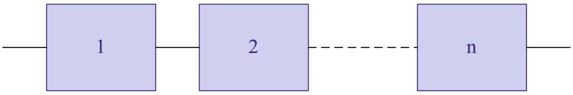


Diagram 5 Probability density function of normal distribution for different values of σ

(ii) CDF:

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\tau} \exp\left[-\frac{(\tau-\mu)^2}{2\sigma^2}\right] d\tau. \tag{16}$$

(iii) Survival probability:

$$R(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\tau}^{\infty} \exp\left[-\frac{(\tau-\mu)^2}{2\sigma^2}\right] d\tau. \tag{17}$$

(iv) Failure rate:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]}{\int_{\tau}^{\infty} \exp\left[-\frac{(\tau-\mu)^2}{2\sigma^2}\right] d\tau}. \tag{18}$$

The failure rate of the normal distribution is increasing; it may describe products within wear-out failure.

(v) The 100*P*th normal percentile is $t_p = \mu + z_p\sigma$, where z_p is the 100*P*th standard normal distribution. The standard normal distribution is the distribution with $\mu = 0$ and $\sigma = 1$.

Exponential distribution

The exponential distribution is probably the most widely used statistical distribution in reliability studies because it is easy to handle in performing reliability analysis and many engineering items exhibit constant failure rate during their useful life [17]. This distribution is defined by a single-scale parameter, λ, and is positively skewed. Diagram 6 presents some graphs of exponential density function for different values of λ.

The equations that describe the exponential distribution are:

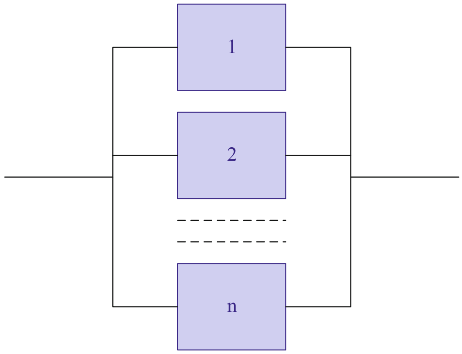
(i) PDF:

$$f(t) = \lambda \exp(-\lambda t). \tag{19}$$

(ii) CDF:

$$F(t) = 1 - \exp(-\lambda t). \tag{20}$$

Diagram 6 Probability density function of exponential distribution for different values of λ



(iii) Survival probability:

$$R(t) = \exp(-\lambda t). \tag{21}$$

(iv) Failure rate:

$$\lambda(t) = \frac{1}{t_{\text{mean}}} = \text{constant}, \tag{22}$$

where t_{mean} is the mean of lifetime.

(v) The 100Pth exponential percentile is $t_p = -t_{\text{mean}} \ln(1-p)$, where $p = F(t_p)$.

If the time t is equal to the mean, t_{mean}, then the reliability from Eq. (20) is $R(t_{\text{mean}}) = 0.368$ or 36.8% and the failure probability from Eq. (21) is $F(t_{\text{mean}}) = 0.632$ or 63.2%.

Lognormal distribution

The lognormal distribution is based on normal distribution; therefore, in some respect, it can be considered as a special case of the normal distribution because of the derivation of its probability density function. The equations that describe the lognormal distribution could be obtained if the random variable t is substituted with the logarithmized form log t in Eqs. (15)–(18):

(i) PDF:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log t - \mu)^2}{2\sigma^2}\right]. \tag{23}$$

(ii) CDF:

$$F(t) = \int_0^{\tau} \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log \tau - \mu)^2}{2\sigma^2}\right] d\tau. \tag{24}$$

(iii) Survival probability:

$$R(t) = \int_{\tau}^{\infty} \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log \tau - \mu)^2}{2\sigma^2}\right] d\tau. \quad (25)$$

(iv) Failure rate:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log t - \mu)^2}{2\sigma^2}\right]}{\int_{\tau}^{\infty} \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log \tau - \mu)^2}{2\sigma^2}\right] d\tau}. \quad (26)$$

(v) The 100*P*th lognormal percentile is $t_p = anti\ \log[\mu + z_p\sigma]$, where z_p is the 100*P*th standard normal percentile.

The scale parameter μ is called the *log mean* and is the mean of the *log* lifetime, whereas the shape parameter σ is called the *log standard deviation* and is the standard deviation of the *log* lifetime. Thus, these parameters are unitless pure numbers.

Weibull distribution

The Weibull distribution is widely used in reliability because this distribution has no characteristic shape such as the others distributions (see Diagram 7). The shape of the Weibull distribution depends on its parameters. The parameters of the Weibull distribution are its *shape* (β) and *scale* (ϑ). The shape of the Weibull distribution provides an insight into the behavior of the failure process and is unitless. Furthermore, it effects the distribution for several different values: (a) for $\beta < 1$; we have a decreasing failure rate function, (b) for $\beta > 1$; we have an increasing failure rate function, and (c) for $\beta = 1$; we have a constant failure rate

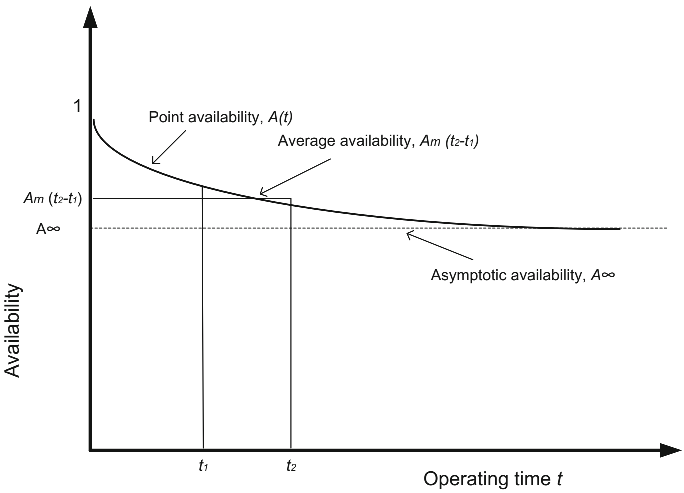


Diagram 7 Probability density function of Weibull distribution for different values of β

function and the distribution is identical to the exponential, with $\lambda = \frac{1}{\vartheta}$. When $1 < \beta < 2$, then the increasing failure rate is concave. When $\beta = 2$, then the increasing failure rate is linear. When $\beta > 2$, then the increasing failure rate is convex, and when $3 \leq \beta \leq 4$, then the increasing failure rate approaches the normal distribution [18].

The scale parameter, ϑ, of the Weibull distribution influences both the mean and spread of the distribution. Moreover, it is called the "*characteristic life*" and its units are identical to those of lifetime, i.e., minutes, hours, cycles. The parameter ϑ is equal to the duration of lifetime t at which 63.2% of the population will have failed.

The equations that describe the Weibull distribution are:

(i) PDF:

$$f(t) = \frac{\beta}{\vartheta}\left(\frac{t}{\vartheta}\right)^{\beta-1} \exp\left[(-t/\vartheta)\right]^{\beta}. \tag{27}$$

(ii) CDF:

$$F(t) = 1 - \exp\left[(-t/\vartheta)\right]^{\beta}. \tag{28}$$

(iii) Survival probability:

$$R(t) = \exp\left[(-t/\vartheta)\right]^{\beta}. \tag{29}$$

(iv) Failure rate:

$$\lambda(t) = \frac{\beta}{\vartheta}\left(\frac{t}{\vartheta}\right)^{\beta-1}. \tag{30}$$

for $\beta = 1$, then $\lambda(t) = \frac{1}{\vartheta}$ = constant. For $\beta > 1$, the $\lambda(t)$ is increased, whereas for $\beta < 1$, the $\lambda(t)$ is decreased.

(v) The $100P$th Weibull percentile is $t_p = \vartheta[-\ln(1-p)]^{1/\beta}$.

Depending upon the value of β, the Weibull distribution function can take the form of the following distribution as:(i) for $\beta < 1$, the Gamma distribution; (ii) for $\beta = 1$, the exponential distribution; (iii) for $\beta = 2$, the lognormal distribution; and (iv) for $\beta = 3.5$, approximate the normal distribution [8].

9 Reliability Network

A system may have various configurations or networks in performing reliability analysis. Components within a system may be related in two primary ways: series or parallel network.

Series Network

The simple case is considered of serial configuration that is the most commonly encountered reliability block diagram in engineering practice. In a serial configuration, all the consisting components of the system should be operating to maintain the required operation of the system. Thus, failure of any one component of the system will cause failure of the whole system. The series n-components are represented by the reliability block diagram of Fig. 2. Let $R_i(t)$ denote the reliability function of the component i, and it is presumed that the system consists of n-components. Then, the reliability of the system for t hours of operation is given by [19, 20]:

$$R_s(t) = R_1(t) * R_2(t) * \ldots * R_n(t) = \prod_{i=1}^{n} R_i(t),\ i = 1, 2, 3\ldots n. \tag{31}$$

In relation (7), it is assumed that all the n-components are independent; in other words, the failure or no failure of one component does not change the reliability of the other component. Therefore, the system operates if all the n mutually independent components in series operate, or:

$$R_s(t) = R_1(t) * R_2(t) * \ldots * R_n(t) \leq \min\{R_1(t), R_2(t), \ldots, R_n(t)\}. \tag{32}$$

The reliability of the system $R_s(t)$ can be no greater than the smallest component reliability.

In case each component has a constant failure rate λ_i, then the system reliability for Eq. (31) is given by

$$R(t) = \exp\left(-\sum_{i=1}^{n} \lambda_i t\right) = \exp\left(-\lambda_s t\right), \tag{33}$$

where $\lambda_s = \sum_{i=1}^{n} \lambda_i$.

Parallel Network

The case of parallel configuration is considered when two or more components are in parallel. The system operates if one or more components operate, and the system fails if all components fail. The parallel n-components are represented by the reliability block diagram of Fig. 3. The reliability of the system for n parallel and independent components is the probability that at least one component does not fail, or:

$$R_s(t) = 1 - \prod_{i=1}^{n} [1 - R_i(t)]. \tag{34}$$

The $R_s(t)$ can be greater than the most reliable component:

$$R_s(t) \geq \max\{R_1(t), R_2(t), \ldots, R_n(t)\}. \tag{35}$$

In case all components have a constant failure rate, then the system reliability for Eq. (10) is given by

$$R_s(t) = 1 - \prod_{i=1}^{n} [1 - \exp(-\lambda_i t)]. \tag{36}$$

where λ_i is the failure rate of ith component.

Therefore, if a system has two-components identical and independent in parallel that have the same constant failure rate λ each, the reliability of this system from Eq. (36) becomes:

$$R_s(t) = 1 - [1 - \exp(-\lambda t)][1 - \exp(-\lambda t)] = 2\exp(-\lambda t) - \exp(-2\lambda t).$$

10 Maintainability

Maintainability is defined as the probability that a failed component or system will be restored or repaired to a specified condition within a period of time when maintenance action is performed in perspective procedures. In general, there are two types of maintenance strategy: the corrective and the preventive maintenance. Corrective maintenance comprises of actions taken to restore a failed component or system to the operational state. The actions involve repair or replacement of all failed components necessary for successful operation of the equipment. Corrective maintenance actions are unscheduled actions intended to restore the equipment from a failed state into a working state. On the other hand, the preventive maintenance is scheduled and is performed usually periodical to a component that is near to failure. Preventive maintenance is widely considered an effective strategy for reducing the number of system failures, thus lowering the overall maintenance cost [21, 22].

Time to Repair

The time to repair (*TTR*) can be considered as a random variable, due to the time to repair a failure depending on from the particular failure mode, the maintenance staff, i.e., experience, skill level, training. To quantify the maintainability, the repair time is represented with T_r, which is the continuous random variable that represents the time to repair a failed component, with probability density function $g(t)$. The cumulative distribution function, $M(t)$, or repair probability is defined as [23–25]:

$$P(T \leq t) = M(t) = \int_0^t g(t)dt. \tag{37}$$

The repair rate $r(t)$ is defined as follows:

$$r(t) = \frac{g(t)}{1 - M(t)}. \tag{38}$$

Thus, from Eqs. (37) and (38), the probability density function, $g(t)$, is

$$g(t) = r(t) \times (1 - M(t)) = r(t) \times \exp[-\int_0^t r(t)dt]. \tag{39}$$

The mean time to repair (*MTTR*) could be calculated by

$$MTTR = \int_0^\infty t \times g(t)dt = \int_0^\infty (1 - \mathrm{M}(t))dt. \tag{40}$$

Normal repair times

In maintainability, the normal repair time distribution applies to maintenance actions where a fixed amount of time to complete is required, i.e., replacement of a particular component or equipment. The probability density function of the repair time is given by

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(t-\mu)^2}{2\sigma^2}] \tag{41}$$

with parameters μ which is the mean repair time, and σ which is the standard deviation of repair time for the normal distribution.

The maintainability function for the normal distribution is

$$M(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]dt. \tag{42}$$

Also, the mean time to repair and the median time to repair are equal; thus,

$$MTTR = t_{\text{median}} = \mu = \frac{\sum_{i=1}^{N} t_i}{N}, \tag{43}$$

and the variance is

$$Var(T_r) = \sigma^2 = \frac{\sum_{i=1}^{N} (t_i - \mu)^2}{N - 1}, \tag{44}$$

where N is the number of repair times and t_i is the repair time i for $i = 1, 2, \ldots N$.

Exponential repair times

In case the repair times are exponentially distributed, then the repair time probability density function is defined by

$$g(t) = r \times \exp(-rt). \tag{45}$$

The maintainability function for the exponential distribution is defined as

$$M(t) = \int_0^t \frac{\exp(-\frac{t}{MTTR})}{MTTR} dt = 1 - \exp(-\frac{t}{MTTR}), \tag{46}$$

where the repair rate is

$$r(t) = \frac{1}{MTTR} = r = \text{constant}. \tag{47}$$

Lognormal repair times

The lognormal repair law is usually applicable to represent the repair distribution. If the continuous random variable T_r that represents the time to repair is lognormally distributed, the Q = ln (T_r) is normally distributed. If μ_r and ${s_r}^2$ are the mean and variance of Q, respectively, then the corresponding probability density function with respect to repair time is defined by:

$$g(t) = \frac{1}{s_r t\sqrt{2\pi}} \exp\left[-\frac{(\ln t - \mu_r)^2}{2s_r^2}\right], t \geq 0. \tag{48}$$

The probability distribution of a repair being completed in time t is the cumulative function for repair, as follows:

$$M(t) = \mathrm{P}(T_r \leq t) = \Phi\left(\frac{1}{s_r} \ln\frac{t}{\exp(\mu_r)}\right), \tag{49}$$

where Φ (.) is the distribution function of the standard normal distribution.

The *MTTR* is the mean time required to complete a maintenance action; thus,

$$MTTR = \exp\left(\mu_r + s_r^2/2\right). \tag{50}$$

The variance for the repaired time is defined as

$$\sigma_r^2 = \exp\left(2\mu_r + s_r^2\right)\left[\exp\left(s_r^2\right) - 1\right]. \tag{51}$$

The mode is the maximum time required to completed a maintenance action, mathematically it is

$$t_{\text{mod}e,r} = \exp\left(\mu_r - s_r^2\right). \tag{52}$$

The median is the downtime within 50% of all maintenance actions which are completed; thus,

$$t_{\text{median},r} = \exp\left(\mu_r\right). \tag{53}$$

The repair rate function is given by

$$r(t) = \frac{g(t)}{1 - M(t)} = \frac{\frac{1}{s_r t\sqrt{2\pi}} \exp\left[-\frac{(\ln t_r - \mu_r)^2}{2s_r^2}\right]}{1 - \left[\Phi\left(\frac{1}{s_r}\ln\frac{t}{\exp\left(\mu_r\right)}\right)\right]}, \quad t \geq 0, \tag{54}$$

where μ_r is the mean of the lognormal distribution of the repair times and s_r is the standard deviation with which the lognormal distribution of the repair times is spread around the mean.

Weibull repair times

The Weibull distribution usually presents corrective maintenance repair times. The repair time probability density function is defined by

$$g(t) = \frac{2}{\vartheta^\beta} t^{\beta-1} \exp\left[(-t/\vartheta)\right]^\beta, \tag{55}$$

where β is the shape parameter and ϑ is the scale parameter of the distribution.

The repair probability for the Weibull repair time distribution is

$$M(t) = 1 - \exp\left[-\left(\frac{t}{\vartheta}\right)^\beta\right]. \tag{56}$$

System repair time

In practice, it is necessary to express system repair time as a function of the repair time of each component that is comprised. The mean time to repair the system ($MTTR_{\text{system}}$) is evaluated as a weighted mean of the individual component of the system based on the relative number of failures that are presented. Thus, the $MTTR_{\text{system}}$ is [26]:

$$MTTR_{\text{system}} = \frac{\sum_{i=1}^{n} q_i f_i MTTR_i}{\sum_{i=1}^{n} q_i f}, \tag{57}$$

where $MTTR_i$ is the mean time to repair the ith component, f_i is the expected number of failures of the ith component, and q_i is the number of identical components of type i.

11 Availability

Availability is defined as the ability of an item (under combined aspects of its reliability, maintainability, and maintenance support) to perform its required function at a stated instant of time or over a stated period of time [27, 28]. The availability of a system can never be less than system reliability because the availability is the probability that the component is currently in a non-failure state, even though it may have previously failed and been restored to its operational state [18]. Consider a reparable system that is in operation at time $t = 0$, when the system fails, a repair action takes place to restore the system in operational state. The state of the system can be given by a binary variable:

$$X(t) = \begin{cases} 1 & \text{if the system is operating at time } t \\ 0 & \text{otherwise} \end{cases}.$$

Availability could be expressed mathematically in several following ways [8, 29, 30]:

1. *Instantaneous or point availability*, $A(t)$: is the probability that the reparable system or equipment will be available to operate at any random time t:

$$A(t) = \Pr[X(t) = 1]. \tag{58}$$

When the system is not repaired, then $A(t) = R(t)$. The unavailability $\overline{A}(t)$ at time t of a reparable system is the probability that the system is in a non-operating state at time t, then

$$\overline{A}(t) = 1 - A(t) = \Pr[X(t) = 0]. \tag{59}$$

In case a system at any time t is either in operating state or in failure state, with constant failure rate λ and constant repair rate r (i.e., the system follows the exponential distribution for the failure data), then the point availability is given by

$$A(t) = \frac{r}{r+\lambda} + \frac{\lambda}{r+\lambda} \times \exp\left[-(r+\lambda) \times t\right]. \tag{60}$$

Substituting $\lambda = \frac{1}{MTTF}$ and $r = \frac{1}{MTTR}$ from Eq. (60) becomes

$$A(t) = \frac{MTTF}{MTTF+MTTR} + \frac{MTTR}{MTTF+MTTR} \times \exp\left[-\left(\frac{1}{MTTF} + \frac{1}{MTTR}\right) \times t\right]. \tag{61}$$

2. *Mission, interval, or average availability*, $A_m(t_2 - t_1)$: is the proportion of time over the interval $(t_2 - t_1)$ which the system or the equipment is available to operate:

$$A_m(t_2 - t_1) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} A(t)dt. \tag{62}$$

It is interesting in some applications that the interval availability from start-up time interval $(0, T)$ which is defined as:

$$A_m(T) = \frac{1}{T} \int_{0}^{T} A(t)dt. \tag{63}$$

In case a system has constant failure rate λ and constant repair rate r, the average availability is

$$A(t) = \frac{r}{\lambda+r} + \frac{\lambda}{(\lambda+r)^2 \times t} \times \left[1 - \exp\left(-(\lambda+r) \times t\right)\right]. \tag{64}$$

3. *Asymptotic availability*, *steady-state*, *or long-run availability*, A_∞: is the probability a system will be available to operate at any point of time t, as

$$A_\infty = \lim_{t \to \infty} A(t). \tag{65}$$

In Fig. 4, the relationship between point, average, and asymptotic availability as a function of operating time is shown.

4. *Inherent availability*, A_{inh}: is defined as the steady-state probability that a system or equipment will be in operational state, assuming that this probability depends merely on the failure and repair distribution. Thus, it can be expressed mathematically in terms of reliability and maintainability trade-offs:

$$A_{\text{inh}} = \lim_{T \to \infty} A(t) = \frac{MTTF}{MTTF + MTTR}. \tag{66}$$

5. *Achieved availability*, A_{ach}: is the probability that equipment or a system will be in operational state when used as specified, taking into account the scheduled and unscheduled maintenance, and is defined as follows:

$$A_{\text{ach}} = \frac{MTBM}{MTBM + MD}, \tag{67}$$

where *MTMB* is the mean time between maintenance and *MD* is the mean downtime of the system. The *MTMB* is given by

$$MTBM = \frac{T_D}{M(T_D) + \frac{T_D}{T_{PM}}}, \tag{68}$$

where $M(T_D)$ is the expected number of failures over the design life T_D, T_{PM} is the mean time of preventive maintenance.

The *MD* of the system is given

$$MD = \frac{M(T_D) \times MTTR + \frac{T_D}{T_{PM}} \times MPMT}{M(T_D) + \frac{T_D}{T_{PM}}}, \tag{69}$$

where the *MPMT* is the mean preventive maintenance time.

6. *Operational availability*, A_{oper}: is the probability that equipment or a system, when used under stated conditions in an actual operational environment, will operate satisfactorily when called upon [31]. The operational availability is defined by

$$A_{\text{oper}} = \frac{MTBM}{MTBM + MDT}, \tag{70}$$

where MDT is the mean maintenance down time and includes maintenance time, logistics delay time, and administrative delay time.

System Availability

A system is considered with n components in series, each component having point availability $A_i(t)$, and then, the system availability is given by

$$A_s(t) = \prod_{i=1}^{n} A_i(t). \tag{71}$$

The inherent availability for the system is

$$A_{\text{inh}} = \prod_{i=1}^{n} \frac{MTTF_i}{MTTF_i + MTTR_i}. \tag{72}$$

In case the series system has all the components with constant failure and repair rate, the inherent availability is given by

$$A_{\text{inh,sys}} = \frac{MTTF_{\text{sys}}}{MTTF_{\text{sys}} + MTTR_{\text{sys}}} \text{ otherwise,} \tag{73}$$

where $MTTF_{\text{sys}}$ and $MTTR_{\text{sys}}$ are the mean time to failure and repair of the system, respectively, and are defined as follows:

$$MTTF_{\text{sys}} = \frac{1}{\sum_{i=1}^{n} \lambda_i} = \frac{1}{\lambda_{\text{sys}}} \tag{74}$$

and

$$MTTR_{\text{sys}} = \frac{\sum_{i=1}^{n} \lambda_i \times MTTR_i}{\sum_{i=1}^{n} \lambda_i} = \frac{1}{\lambda_{\text{sys}}} \sum_{i=1}^{n} \lambda_i \times MTTR_i. \tag{75}$$

When the system consists of n components in parallel, the system availability is given by

$$A_{\text{sys}}(t) = 1 - \prod_{i=1}^{n} (1 - A_i(t)). \tag{76}$$

References

1. Advisory Group on Reliability of Electronic Equipment (AGREE) (1957) Reliability of military electronic equipment, Office of the Assistant Secretary of Defense (Research and Engineering). Department of Defense, Washington, DC
2. Coppola A (1984) Reliability engineering of electronic equipment: a historical perspective. IEEE Trans Reliab 33:29–35
3. Ding SH, Kamaruddin S (2015) Maintenance policy optimization—literature review and directions. Int J Adv Manuf Technol 76(5–8):1263–1283
4. AMCP706-133 (1976) Engineering design handbook: maintainability engineering theory and practice. US Department of Defense, Washington, DC
5. Retterer BL, Kowalski RA (1984) Maintainability: a historical perspective. IEEE Trans Reliab 33:56–61

6. Parida A, Kumar U, Galar D, Stenström C (2015) Performance measurement and management for maintenance: a literature review. J Qual Maint Eng 21(1):2–33
7. MIL-HDBK-472 (1966) Maintainability prediction, Department of Defense, Washington, DC, USA
8. MIL-HDBK-338B (1998) Electronic reliability design handbook, Department of Defence, USA
9. Berolini A (2007) Reliability engineering: theory and practice, 5th edn. Springer, Firenze, Italy
10. Blischke WR, Murthy DNP (2003) Case studies in reliability and maintenance. Wiley, Hooboken, pp 351–445
11. Dhillon BS (2006) Maintainability, maintenance and reliability for engineers. Taylor & Francis, Boca Raton
12. Smith DJ (2011) Reliability, maintainability, and risk: practical methods for engineers, 8th edn. Elsevier, Oxford, UK
13. Tsarouhas P (2014) Application of statistical approaches for analyzing the reliability and maintainability of food production lines: a case study of mozzarella cheese, In: Granato D, Ares G (eds) Chapter in Mathematical and statistical methods in food science and technology. Wiley, pp. 491–510
14. Wang H, Pham H (2006) Reliability and optimal maintenance. Springer, London
15. Garmabaki AHS, Ahmadi A, Block J, Pham H, Kumar U (2016) A reliability decision framework for multiple repairable units. Reliab Eng Syst Saf 150:78–88
16. Kumar U, Crocker J, Chitra T, Saranga H (2006) Reliability and six sigma. Springer, London
17. Dhillon BS (2008) Mining equipment reliability, maintainability and safety. Springer, London
18. Ebeling C (2008) An introduction to reliability and maintainability engineering, 10th edn. Tata McGraw-Hill, New Delhi
19. Rigdon SE, Basu AP (2000) Statistical methods for the reliability of repairable systems. Wiley, New York
20. Gustafson A, Schunnesson H, Kumar U (2015) Reliability analysis and comparison between automatic and manual load haul dump machines. Qual Reliab Eng Int 31(3):523–531
21. Okogbaa OG, Peng X (1996) Methodology for preventive maintenance analysis under transient response. In: Proceedings of the annual reliability and maintainability symposium, USA
22. Tsarouhas P, Makrygianni M (2016) A framework for maintenance and combat readiness management of a jet fighter aircraft. Int J Syst Assur Eng Manage. doi:10.1007/s13198-016-0481-8
23. Pascual DG, Kumar U (2016) Maintenance audits handbook: a performance measurement framework. CRC Press, Taylor & Francis Group, LLC, Boca Raton
24. Ben-Daya M, Duffua OS, Raouf A, Knezevic J, Ait-Kadi D (2009) Handbook of maintenance management and engineering. Springer, New York
25. Block J, Ahmadi A, Kumar U (2014) Fleet-level reliability of multiple repairable units: a parametric approach using the power law process. Int J Perform Eng 10(3):50–239
26. Kumar U, Crocker J, Knezevic J, El-Haram M (2000) Reliability, maintenance and logistic support-A life cycle approach. Kluwer Academic Press, Boston
27. De Sanctis I, Paciarotti C, Di Giovine O (2016) Integration between RCM and RAM: a case study. Int J Qual Reliab Manage 33(6):852–880
28. Rausand M, Hoyland A (2004) System reliability theory: models, statistical methods, and applications, 2nd edn. Wiley, New Jersey
29. Pecht M (2009) Product reliability, maintainability, and supportability handbook, 2nd edn. Taylor & Francis, Boca Raton
30. Kumar U, Ahmadi A, Verma AK, Varde P (2016) Current trends in reliability, availability, maintainability and safety: an industry perspective (Lecture notes in mechanical engineering). Springer International Publishing, Switzerland
31. Blanchard BS, Fabrycky WJ (1998) System engineering and analysis. Prentice Hall, Englewood Cliffs

Role of Fuzzy Logic in Flexible Manufacturing System

Dinesh C. S. Bisht, Pankaj Kumar Srivastava and Mangey Ram

Abstract In manufacturing systems there are two types of flexibilities, one in machines and other in routing. To get maximum output, manufacturer utilizes its best recourses even under uncertain environment. Fuzzy logic is a tool which easily handles uncertainties. This article describes basics of fuzzy set, fuzzy membership, and fuzzy rule base system and defuzzification. It also covers different aspects of fuzzy manufacturing system (FMS) and some standard fuzzy logic applications in FMS. Limitations of fuzzy modeling in flexible manufacturing system are also discussed. All discussed method can be further modified for individual problems. Future researchers can consider different aspects of fuzzy logic in flexible manufacturing system to handle more complex problems.

Keywords Fuzzy logic · Membership function · Membership grade
Fuzzy rule base system · Defuzzification · Centroid method · Flexible manufacturing system · Slack time · Optimization · Sequencing
Routing factor · Time factor · Scheduling · Mean machine utilization
Fuzzy multiple scheduling · Fuzzy logic controller · Flexible manufacturing system
Managerial objectives

D.C.S. Bisht · P.K. Srivastava (✉)
Department of Mathematics, Jaypee Institute of Information Technology,
Noida, India
e-mail: pankaj.srivastava@jiit.ac.in

D.C.S. Bisht
e-mail: drbisht.math@gmail.com

M. Ram
Department of Mathematics, Computer Science and Engineering,
Graphic Era University, Dehradun, Uttarakhand, India
e-mail: mangeyram@gmail.com

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*,
Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3_9

1 Fuzzy Logic

Life is full of uncertainties; it can be vague or imprecise. To deal with such uncertainties probability theory used to be a tool for mathematician, which is based on classical set theory. Prof. Zadeh [1], in 1965, challenged that there are some uncertainties which are out of the scope of probability theory. For example a company owner need honest person for his company. Now there are available choices can be extremely honest, very honest, honest some time, dishonest; which cannot be defined using classical logic because in this logic there are only two choices honest and dishonest. Prof. Zadeh named this new concept as fuzzy set theory based on membership functions. Classical set theory is about yes or no concepts whereas fuzzy set theory includes grey part also. Fuzzy set theory deals with appropriate reasoning in linguistic terms.

1.1 Definition of Fuzzy Set

Fuzzy set $F\,(m)$ is represented by pair of two components, first is the member m and the second is its membership grade $\mu_F(m)$ which maps any element m of universe of discourse M to the membership space [0,1], as given below

$$F(m) = \{(m, \mu_F(m)),\ m \in M \tag{1.1}$$

1.2 Equal Fuzzy Sets

Two fuzzy sets F_1 and F_2 are said to be equal if all the members of F_1 belong to F_2 with same membership grade as in F_1.

1.3 Membership Function

A function which describes the membership grades of elements in fuzzy set is said to be membership function. A membership function can be discrete or continuous. It needs a uniform membership function representation for efficiency. Some well known membership functions are as follows

1.3.1 Z-Shaped Membership Function

Z-Shaped membership function is given by-

$$Z(x;p,q) = \begin{cases} 1 - 2\left(\frac{x-p}{q-p}\right)^2; & \textit{if } p < x \geq (p+q)/2 \\ 2\left(\frac{x-p}{q-p}\right)^2; & \textit{if } (p+q)/2 < x \leq q \\ 1; & \textit{if } x \leq p \\ 0; & \text{otherwise} \end{cases} \tag{1.2}$$

1.3.2 Triangular Membership Function

Triangular membership function is given by-

$$T(x;p,q,r) = \begin{cases} \frac{x-p}{q-p}; & \textit{if } p < x \geq q \\ \frac{r-x}{r-q}; & \textit{if } q < x \leq r \\ 0; & \text{otherwise} \end{cases} \tag{1.3}$$

1.3.3 Trapezoidal Membership Function

Trapezoidal membership function is represented by-

$$T(x;p,q,r,s) = \begin{cases} \frac{x-p}{q-p}; & \textit{if } p < x \geq q \\ 1; & \textit{if } q < x \leq r \\ \frac{s-x}{s-r}; & \textit{if } r < x \leq s \\ 0; & \text{otherwise} \end{cases} \tag{1.4}$$

1.3.4 Gaussian Membership Function

Representation of Gaussian membership function is as follows-

$$G(x;\sigma,m) = \frac{1}{e^{\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}} \tag{1.5}$$

1.4 Fuzzy Rule Base System

To understand fuzzy rule base system let us take an statement “If there is *high* traffic jam and *heavy* rain then I may get *little* late”, Here high, heavy and little are fuzzy sets related to variables traffic jam, rainfall and late respectively. Mathematically it can be represented as

IF (x is F_1 and y is F_2) THEN (z is F_3); where F_1, F_2 and F_3 are fuzzy sets and x, y and z are variables. A collection of all such rule for a particular system is known as rule base [2].

1.5 Fuzzy Defuzzification

It is quite difficult to take decision on the bases of fuzzy output; in that case this fuzzy output is converted into crisp value. This process of converting fuzzy output into crisp output is known as defuzzification [3]. Different methods are available in literature; some widely used methods are discussed below

1.5.1 Center of Area (CoA) Method

This method is also known by names Centroid method and Center of Gravity method. CoA is most popular method of defuzzification among researchers. This method is based on center of area taken by the fuzzy set. Its defuzzified value is calculated after considering entire possibility distribution and is given by-

$$m* = \begin{cases} \dfrac{\int (\mu(m) \times m)dm}{\int \mu(m)\, dm}; & \text{for continuous membership value of } m \\ \dfrac{\sum \mu(m) \times m}{\sum \mu(m)}; & \text{for discrete membership value of } m \end{cases} \tag{1.6}$$

1.5.2 Max-Membership Function

This method is also called height method and applicable to peaked output functions. Expression for this method is given by-

$$\mu(m*) \geq \mu(m)\ \forall m \in M(\text{Universe of discourse}) \tag{1.7}$$

1.5.3 Weighted Average Method

This method is applied when output is symmetrical and give by-

$$m* = \frac{\sum \mu(\bar{m}) \times \bar{m}}{\sum \mu(\bar{m})} \tag{1.8}$$

where $\bar{m}$ is the Centroid of each symmetric membership function.

This method is computationally efficient but less popular.

1.5.4 Mean-Max Method

This method is similar to Max-membership method, only difference is the locations of maximum membership can be more than one. Expression is given by-

$$m* = \frac{m_1 + m_2}{2} \tag{1.9}$$

where $m_1 + m_2$ are the mean of maximum interval.

1.5.5 Center of Sums

This method is based on the algebraic sum of fuzzy subsets. This method is very fast in terms of calculations. The defuzzified value is give by-

$$m* = \frac{\sum_{i=1}^{N} m_i \sum_{k=1}^{n} \mu(m_i)}{\sum_{i=1}^{N} \sum_{k=1}^{n} \mu(m_i)} \tag{1.10}$$

2 Flexible Manufacturing System (FMS)

The market competition was at its best in the sixth decade of twentieth century as David Williamson, a British engineer introduced the concept of FMS. In next decade cost became the primary concern and later on it was followed by quality of products. Customers demand for speedy delivery, increased the complexity in the market. To gain competitive advantage FMS came into the picture more strongly.

FMS is a manufacturing technology which contains some amount of flexibility. Basically FMS is a manufacturing system which is not only reprogrammable but also capable of automatically producing a variety of products [4].

2.1 Features of FMS

On the basis of type of products FMS has different features. The financial supply is a pivot element of it and so we may describe the following two distinct features of FMS

1. Production of different variety of product types at a high cost.
2. Production of large volumes of product types at a lower cost.

2.2 Need of FMS

A few external changes such as change in the design of product and so the change in production system requires FMS. It will help in optimizing the manufacturing cycle time and so it will reduce the production cost. This cost management is converted into the necessity of FMS.

2.3 Symptoms of Being Flexible

There are some capabilities which make a manufacturing system flexible, such as-

1. Instant change in physical setup.
2. Instant change in operating instructions.
3. The quality of differentiating among the different part styles processed by the system.

2.4 Classification of FMS Related Problems

As per the objective associated with FMS it has following categories-

1. Facility design to accomplish long term managerial objectives.
2. Intermediate range planning to decide master production scheduling and deals.
3. Strategic analysis and economic justification to provide long range strategic business plan.

2.5 Short-Term FMS Control

An FMS system has short control in form of scheduling and on its basis it is governed by following four factors

(i) *Time factor*: decision of insertion timing of order in system;
(ii) *Sequence factor*: decision of ordering of order in system;
(iii) *Routing factor*: decision to choose a job for an operation out of multiple choices;
(iv) *Priority factor*: decision to choose priority out of jobs, workstations and resources.

2.6 Industrial Implications of FMS

Scheduling is the process of organizing, choosing and timing resource usage to carry out all the activities necessary to produce the desired outputs of activities and resources. In an FMS, the objective of scheduling is to optimize the use of resources. Various attempts are made to use an industrial implementation for experimentation [5–11].

2.7 Challenges with FMS

The objective of each enterprise involved in manufacturing is to achieve three goals: Reduce production cost, produce quality product and capability to response rapidly. The third goal responsiveness is most important to achieve and it becomes most challenging aspect for the enterprise. This also includes adjustable rates of productivity and flexibility to raise production at the lowest possible cost.

3 Fuzzy Logic in Flexible Manufacturing System

Decisions are precise when all available resources are used in well defined manner. Humans have the power to use their practical knowledge to get desired output. This motivate researcher to mimic human behavior in computational world. Fuzzy logic is a strong tool to deal with uncertainty and can deal with multiple criteria in one go. Even linguistic variables can be handled easily in fuzzy environment. Fuzzy logic can be applied to simulate FMS. Moreover the fuzzy based simulations are applied to sequencing of jobs. Out of various techniques applied to FMS, technique based on fuzzy logic has its own importance and most of the researchers use fuzzy logic and expert system theory to optimize resources scheduling. With the rapid development of fuzzy technologies, different fuzzy control strategies have been developed [12–19].

4 Different Fuzzy-Logic Based Approaches for Job Sequencing and Routing in FMS

4.1 Changing of Antecedents

Hintz and Zimmermann [20] proposed a new algorithm by changing the antecedents. The coming job is taken as consequent. The elements that are important to make a decision become the antecedents of the fuzzy rules in the scheduling procedure and described as

- slack time;
- waiting time;
- utilization uniformity across machines;
- unguarded utilization of machines;
- external priority.

And all such rules are stored in fuzzy system with their weight. With different antecedents similar approach is applied in problem of setting priority. This fuzzy system performed well while compared with other discrete methods as per following aspects:

1. Mean waiting time
2. Number of in-time
3. Mean machine utilization.

4.2 Fuzzy Set Theory Along with Possibility Theory

Fuzzy sets have a massive application to improve human planning and to overcome data uncertainty [21]. Every attribute is considered for possibility distribution of routing. With the help of weighted average possibility distributions are combined into one aggregate possibility distribution. Weights are fuzzy in nature. Center of gravity is taken as defuzzification method. That's why an additional constraint with minimum time routing is obtained. The added constraint contains different quality of every machine. If any part has no route then it will be in top priority. For other parts appropriateness index is calculated and rerouting is done on the bases of this index.

4.3 Fuzzy Multiple Scheduling

Hatono et al. [22] gave this idea of fuzzy multiple scheduling. This type of scheduling consists of evaluation, policy and dispatch module. All modules having its own fuzzy system with specific target. First module evaluates any sequence with the help of fuzzy rules, where resources, waiting time for resources and parts, earliness of batch are taken as antecedents for fuzzy rules. Consequent of such fuzzy rules are evaluation of schedule. Next module calculates the influence of different parts on schedule and generates weight for rules. The last module utilizes fuzzy rules in dispatch. Individual module has a unique fuzzy system. This architecture allows user the freedom of tuning the membership function which leads to modification of the target. This is a revolutionary idea of adjusting scheduling for better results.

4.4 Fuzzy Scheduling Mechanism for Job Shops

As per demand of clients Watanabe et al. [23] describe three categories of orders

(i) Normal,
(ii) Express
(iii) Just in time (JIT)

Each type of order possesses a different profit curve and two parameters, first measurement of manufacturing speed and second measurement of profit through order.

To determine parts priorities two rules based on fuzzy are introduced. First rule confirms that if product manufacturing is low and profit measurement is high then priority is high while second rule predicts that if product manufacturing is high and profit measurement is low then priority is low. This fuzzy scheduler is applicable on max-min inference, non-singleton fuzzifier and center of gravity defuzzifier.

4.5 Fuzzy Logic Controller (FLC) for Part Routing

To decide the electability of every feasible optional part route Ben-Arieh and Lee [24] proposed s fuzzy rule based approach and the route having highest selectibility factor is adopted. Each rule shows one consequent with the selectibility factor. It is based on max-min mechanism determining the output of each rule. Observation or aid of human experts creates all these rules.

The effectiveness of the fuzzy logic controller (FLC) was then examined through simulations on comparing with common routing heuristics. The FLC is always superior to those heuristics as per average time in system; number of parts completed and fraction of late parts. The fuzzy logic controller behaves better than heuristic rules especially in critical situations like a congested system.

5 Limitations

All discussed fuzzy approaches for FMS are all innovative and opened new doors for future researchers [25]. At the same time these methods have their limitations, some of these are:

1. Scheduling rules are made using human expert knowledge without any standard approach.
2. Comparison with other standard methods needs to be discussed.
3. Hard to decide correct weight of rules.
4. It is not an easy task to adjust weights of rules in dispatching module in complex problems.

6 Future Aspects

Future researchers can consider following aspects of fuzzy logic in flexible manufacturing system so that one can handle even a more complex problem.

1. Different decision points and more fuzzy rules can be set for a fuzzy system.
2. To evaluate actual performance realistic examples with constraints should be discussed.
3. There are many combinations for fuzzy rules, so more combination can be taken.
4. Adaptive fuzzy rules can be implemented with the help of other soft computing techniques like genetic algorithms and particle, swarm optimization etc.

References

1. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
2. Rajasekaran S, Vijayalakshmi Pai GA (2003) Neural networks, fuzzy logic and genetic algorithm: synthesis and applications. PHI Learning Pvt. Limited, New Delhi
3. Hellendoorn H, Thomas C (1993) Defuzzification in fuzzy controllers. Intelligent Fuzzy Syst 1:109–123
4. Chan FTS, Jiang B (2001) The applications of flexible manufacturing technologies in business process reengineering. Int J Flex Manuf Syst 13(2):131–144
5. Srinoi P, Shayan E, Ghotb F (2002) Scheduling of flexible manufacturing systems using fuzzy logic. Int J Prod Res 44(11):1–21
6. Huang L, Chen H-S, Hu T-T (2013) Survey on resource allocation policy and job scheduling algorithms of cloud computing. J Softw 8(2):480–487
7. Wu Z, Liu X, Ni Z, Yuan D, Yang Y (2013) A market-oriented hierarchical scheduling strategy in cloud workflow systems. J Supercomput 63(1):256–293
8. Xu B, Zhao C, Hu E, Hu B (2011) Job scheduling algorithm based on Berger model in cloud environment. Adv Eng Softw 42(7):419–425
9. Ma YB, Jang SH, Lee JS (2011) Ontology-based resource management for cloud computing. In: Intelligent information and database systems: third international conference, proceedings, Part II, vol 6592 of Lecture notes in computer science. Springer, Berlin, pp 343–352
10. Sun DW, Chang GC, Li FY, Wang C, Wang XW (2011) Optimizing multi-dimensional QoS cloud resource scheduling by immune clonal with preference. Chin J Electron 39(8):1824–1831
11. Byun EK, Kee YS, Kim JS, Maeng S (2011) Cost optimized provisioning of elastic resources for application workflows. Future Gener Comput Syst 27(8):1011–1026
12. Rattanatamrong P (2011) Real-time scheduling of ensemble systems with limited resources. Ph.D. thesis, University of Florida, Gainesville, FL, USA
13. Luo M, Zhang K, Yao L, Zhou X (2012) Research on resources scheduling technology based on fuzzy clustering analysis. In: Proceedings of the 9th international conference on fuzzy systems and knowledge discovery (FSKD ’12), pp 152–155
14. Rattanatamrong P, Fortes JAB (2014) Fuzzy scheduling of real-time ensemble systems. In: Proceedings of the international conference on high performance computing and simulation (HPCS ’14), pp 146–153, IEEE, Bologna, Italy
15. Masmoudi M, Ha¨ıt A (2013) Project scheduling under uncertainty using fuzzy modelling and solving techniques. Eng Appl Artif Intell 26(1):135–149

16. Er MJ, Sun YL (2001) Hybrid fuzzy proportional integral plus conventional derivative control of linear and nonlinear systems. IEEE Trans Ind Electron 48(6):1109–1117
17. Wu JC, Liu TS (1996) A Sliding-Mode Approach to Fuzzy Control Design. IEEE Trans Control Syst Technol 4(2):141–151
18. Hsu FY, Fu LC (2000) Intelligent robot deburring using adaptive fuzzy hybrid position/force control. IEEE Trans Robots Autom 16(4):325–334
19. Li HX, Gatland HB (1995) A new methodology for designing a fuzzy logic controller. IEEE Trans Syst Man, Cybernetics 25(3):505–512
20. Hintz GW, Zimmermann HJ (1989) A method to control flexible manufacturing systems. Eur J Oper Res 41:321–334
21. Choobineh F, Shivani M (1992) Dynamic process planning and scheduling algorithm. In: Proceedings of the Japan/USA symposium on flexible automation, pp. 429–432, ASME
22. Hatono I, Suzuka T, Umano M, Tamura H (1992) Towards intelligent scheduling for flexible manufacturing: application of fuzzy inference to realizing high variety of objectives. In: Proceedings of the Japan/USA symposium on flexible automation, vol 1, pp 433–440, ASME
23. Watanabe T, Tokumaru H, Hashimoto Y (1993) Job-shop scheduling using neural networks. Control Eng Pract 1(6):957–961
24. Ben-Arieh D, Lee ES (1995) Fuzzy logic controller for part routing. In: Parsaei HR, Jamshidi M (eds) Design and implementation of intelligent manufacturing systems, PTR Prentice Hall, Englewood Cliffs, pp 81–106
25. Dadone P (1997) Fuzzy control of flexible manufacturing systems. Dissertation, Virginia Polytechnic Institute and State University

Index

M. Ram and J.P. Davim (eds.), *Diagnostic Techniques in Industrial Engineering*, Management and Industrial Engineering, https://doi.org/10.1007/978-3-319-65497-3