Modern software for the environmental modeling and statistical data analysis

Kachiashvili K.J. a *, Melikdzhanian D.I. b

a Professor of Abdus Salam School of Mathematical Sciences of GC University, 68-B, New Muslim Town Lahore, Pakistan
Main Researcher of I. Vekua Institute of Applied Mathematics of the Tbilisi State University, 2, st. University, Tbilisi, 380043, Georgia
b Researcher of I. Vekua Institute of Applied Mathematics of the Tbilisi State University, 2, st. University, Tbilisi, 380043, Georgia

Abstract
There are considered three original software packages developed by authors: the first is for statistical processing of the experimental information; the second is a software package of realization of mathematical models of pollutants transport in rivers and the third is for identification of river water excessive pollution sources located between two controlled cross-sections of the river. They are designated on identical methodological basis and are intended for the users who are not professionals in the field of applied mathematics and computer science. The packages are universal, simple and convenient for understanding and application.

© 2010 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: software package, mathematical model, pollutant transport, identification, pollution sources

1. Introduction

The wide application of the methods of applied mathematics to the solution of practical tasks was promoted by intensive development of computers in the sixties of the last century, because the practical application of these methods is connected with realization of calculations in large volumes, which is possible with the help of the computer. The next fundamental push for wide introduction of these methods in practice was the intensive development and spreading of personal computers, for which, there were developed the works for creation of the specialized software for experimental data processing. The development of these packages moved to a qualitatively new level. The creation of the specialized program packages with specialized interfaces of data input-output began. In these packages, there are realized the sets of various methods of applied mathematics united in subject sections. These packages have the advanced means of input and interpretation of processable data. The representation of the initial data and the obtained results in the most convenient and evident form is of major importance for data analysis, as they allow us to choose most successfully those methods which will enable us to solve a problem and to attain the goal to be sought or to interpret logically the obtained results and to make well-founded conclusions. Therefore
modern software packages include convenient programs for data input-output and their graphic representation. The most of them are capable of building the three-dimensional diagrams. Below are offered three original packages developed by authors as examples of modern software of experimental data processing: 1) package for statistical processing of the experimental information, 2) package of realization of mathematical models of pollutants transport in rivers and 3) package for identification of river water excessive pollution sources located between two controlled cross-sections of the river. They are designated on identical methodological basis and are intended for the users who are not professionals in the field of applied mathematics and computer science. The packages are universal, simple and convenient for understanding and application.

2. Description of the packages

2.1. Statistical processing of the experimental information

The package is intended for processing the results of various experiments and investigations by the methods of applied statistics [1]. Applications include metrology, instrument-making, ecology, agriculture, medicine, biology, physics, chemistry, economy, sociology, etc.

The package is designed for the user unfamiliar with special sections of mathematics, applied statistics and programming. It is realized on the IBM-compatible personal computer.

The characteristic features of the package are the following: statistical efficiency, clearness for the user, convenience in operation and significant cheapness in comparison with the packages of a similar class. The management is carried out with the help of the user’s menu. Working with this package, you can use the program instructions (Help) applied to each task. The results of calculations are interpreted in the most user-friendly form (tables, graphs, comments).

The programs included in the package realize the tasks which are incorporated into eight sections.

1. Calculation of statistical numerical characteristics includes: graphic representation of the initial data and analysis of "gross blunders"; calculation of the basic numerical characteristics of the phenomenon under study; calculation of confidential intervals of the basic numerical characteristics (the classical method, a new method); calculation and construction of the histogram; calculation of tolerant intervals; test of admissibility of dispersion of estimations of variances of several groups of observation; test of admissibility of dispersion of arithmetic means by Fisher’s method; test of admissibility of dispersion of arithmetic means of groups at different intra-group variances; identification of two empirical distributions with the help of Chi-square and Kolmogoroff-Smirnoff tests (the classical case and its generalization when the number and the sizes of intervals of giving the random variables do not coincide); calculation of quantiles of probability distribution laws.

2. Calculation of dynamic characteristics includes: identification of differential equations of the first and the second orders of transient processes at various initial conditions; calculation of impulsive, transmitted, transient and amplitude-phase-frequency characteristics.

3. Identification of functional dependences (with the opportunity of calculation of prognosis values) includes the restoration of auto-, inter- and multi-regression dependences of the following kinds:

- \[ y = a \cdot x^b \] (geometrical);
- \[ y = a \cdot e^{b \cdot x} \] (exponential);
- \[ y = a \cdot \ln(b \cdot x) \] (logarithmic);
- \[ y = a \cdot x^b \cdot e^{c \cdot x} \] (geometric-exponential);
- \[ y = a \cdot (1 - e^{-b \cdot x}) \] (inverse-exponential);
- \[ y = a + b \cdot x^c \] (geometrical with a free term);
- \[ y = a + b \cdot e^{c \cdot x} \] (exponential with a free term);
- \[ y = (a + b \cdot x) \cdot e^{c \cdot x} \] (linear-exponential);
• \( y = h + (a + b \cdot x) \cdot e^{cx} \) (linear-exponential with a free term);
• \( y = a \cdot x^c \cdot (1 - bx)^d \) (product of geometrical dependences);
• \( y = a \cdot x^c + b \cdot x^d \) (sum of geometrical dependences);
• \( y = a \cdot e^{cx} + b \cdot e^{dx} \) (sum of exponential dependences);
• \( y = h + a \cdot e^{cx} + b \cdot x^d \) (sum of geometrical dependences with a free term);
• \( y = e^{cx} \cdot (a \cdot \cos(\alpha x) + b \cdot \sin(\alpha x)) \) (exponential-sine);
• \( y = h + e^{cx} \cdot (a \cdot \cos(\alpha x) + b \cdot \sin(\alpha x)) \) (exponential-sine with a free term);
• \( y = e^{cx} \cdot \sum_{k=0}^{m} A_k \cdot x^k; \) (polynomial);
• \( y = e^{cx} \cdot \sum_{k=0}^{m} A_k \cdot x^k; \) (geometric-polynomial);
• \( y = e^{cx} \cdot \sum_{k=0}^{m} A_k \cdot x^k; \) (exponential-polynomial);
• \( y = c \cdot \ln \left( \sum_{k=0}^{m} A_k \cdot x^k \right); \) (logarithmic-polynomial);
• \( y = A_k + \sum_{k=0}^{m} \left( A_k \cdot \cos(k \omega t) + B_k \cdot \sin(k \omega t) \right); \) (periodic);
• \( y = e^{cx} \cdot \sum_{k=0}^{m} A_k \cdot x^k; \) (linear multiple regression).

Restoration of the functional dependence by polynomial or trigonometrical splines.

These algorithms have been developed in regard to stationary and non-stationary variances within the observation range with consideration for the non-linearity factor. There are realized the algorithms of regression analysis. There is the possibility calculating the values of restored functional dependences at any points in the area of their definition.

4. Identification of probability distribution densities includes: estimation of unknown parameters of probability distribution densities: normal, uniform, triangular, trapezoidal, antimodal I, antimodal II, truncated Raileigh’s, chi-square, Student’s, binomial and Poisson’s; identification of the above mentioned probability distribution densities by the chi-square test; identification of the above probability distribution functions by the Kolmogoroff-Smirnoff and the omega-square tests, except for the Poisson and the Bernoulli distribution functions; testing of distribution normality at \( N < 50 \) by \( V \) and \( D \) tests.

The chosen power of these tests is ensured at using the estimations of unknown parameters of probability distributions defined by the given sample at identifying the distribution function by the Kolmogoroff-Smirnoff and the omega-square tests.

5. Applied tasks of decision making include: computation of density functions of suspended particle sizes in a liquid medium during a dynamic sedimentation process at specified theoretical and empirical distributions; identification of two empirical distribution densities at different intervals of representation of random variables; identification of \( M \)-dimensional objects; detection of the intensity variation in the Poisson flow; detection of the intensity “ejection” of the Poisson flow; testing of many simple hypotheses by using the unconstrained, constrained and quasi-optimal Bayesian rules of decision-making.

6. Nonparametrical methods: hypotheses testing by the Mann-Whitney test; hypotheses testing by the Wilcoxon test; sign test for one sample; sign test for the analysis of paired observations; Wilcoxon signed-rank sum test; One-factor analysis: Kruskal-Wallis test; Jonckheere test; one-factor analysis of variance; Two-factor analysis: Friedman test; Page test; two-way analysis of variance; Relationship of indicators measured in the scale of orders: Spearman’s rank correlation; Kendell’s correlation coefficient.
7. Time series processing includes: separation of the trends of multidimensional time series and computation of centred random series; calculation of forecasting values of time series; calculation of one-dimensional distribution laws of momentary values of the truncated realization of standardized random series; calculation of auto- and inter-covariance (correlation) functions and fulfillment of total correlation analysis; the latter includes the following tasks: the correlation coefficient, the homogeneity of correlation coefficients, the correlation index, the partial correlation coefficient, the multiplicity correlation coefficient; calculation of auto- and inter-spectral power densities; testing of the stationary state of centered random series by the trend; testing of the stationary state of centered random series according to the second moments; testing of the stationary state of centered random series using the errors of computation of covariance matrix.

8. Generation of the pseudo-random numbers and processes includes: generation of pseudo-random numbers distributed by normal, uniform, triangular, trapezoidal, antimodal I, antimodal II, truncated Raileigh’s, chi-square, Student’s, binomial and Poisson’s distribution laws; generation of the Poisson flow; generation of normally distributed random vectors; generation of multidimensional normal Markoff processes with calculation of absolute errors.

2.2. Mathematical models of pollutants transport in rivers

The applied program package of mathematical models of pollutants transport in rivers is designed an up-to-date as convenient reliable tool for experts in different fields of knowledge (biology, ichthyology, ecology, hydrology, building, agriculture etc.), allowing them to calculate the polluting substance concentrations at any point of the river depending on the quantity and the conditions of discharging from several pollution sources [2]. As the mathematical models of water quality formation in the river under the influence of several pollution sources, spatially one-, two- and three-dimensional advective-diffusion models at different initial and boundary conditions (see below) are realized in the package: a) an advective-diffusion equation with non-local boundary condition at the end of the controlled section with account for the coefficient of natural self-purification of the river; b) an advective-diffusion equation with boundary condition of full mixing at the end of the controlled section; c) an advective-diffusion equation ignoring the vertical advection with non-local boundary condition at the end of the controlled section with account for the coefficient of natural self-purification of the river; d) an advective-diffusion equation ignoring the vertical advection with boundary condition of full mixing at the end of the controlled section; e) a diffusion equation with non-local boundary condition at the end of the controlled section with account for the coefficient of natural self-purification of the river; f) a diffusion equation with boundary condition of full mixing at the end of the river controlled section. For the abovementioned mathematical models, there were developed new computation schemes and advanced the known finite-difference ones [3, 4].

In the package there are the options of inputting and editing the initial data describing the geographical, geometrical and hydrological features of the simulated section of the river, the condition and the specific features of pollution of the river, the quantity and the name of the pollutant of interest, the types of used models and restrictions, the ways of assignment of these data etc. There are the options of choosing the language of dialogue with the package, the conditions of computation realization, the desirable accuracy of computation, the type and format of output of the results. At any stage of working with the package, there is an option of help concerning the methods which are realized in the package, the capabilities and specific features of the package, the parameters of tasks and obtained results.

2.3. Identification of river water excessive pollution sources

For investigation, analysis and quality control of the environment, it is necessary to have adequate information about its condition. This information comprises an extremely great number of physical, chemical and biological parameters. For their regular checking, it is necessary to make a great number of measurements. Therefore, the problem associated with the environmental checking and control can be solved only by means of automatic on-line analyzers of environmental pollution, and automated systems of environmental monitoring [5]. Such systems are designed for the effective reduction of environmental pollution, and they are based on continuous checking,
prediction and control of effluents under unfavorable conditions by means of effective organizational and technical measures, and rational redistribution of facilities (if their number is restricted) among the objects.

The monitoring and control of the water environment conditions includes the problem of identification of pollution sources in order to take measures aimed at their elimination. This problem is especially topical for urban conditions, when the number of pollution sources is rather great, and there is no possibility of controlling each of them separately. Identification of pollution sources has not only ecological and technological effects, but a significant economic effect as well. The economic effect is reached by minimizing the technical facilities, in particular, the measurement facilities needed for the stand-alone control of each pollution source.

The program package for identification of river water excessive pollution sources located between two controlled cross-sections of the river has been developed by the authors on the basis of mathematical models of pollutant transport in the rivers and statistical hypotheses testing methods [6]. The identification algorithms were elaborated with the supposition that the pollution sources discharge different compositions of pollutants or (at the identical composition) different proportions of pollutants into the rivers. Here was formalized the problem and were developed the general scheme and working algorithms of its solution. Optimal and quasi-optimal decision-making algorithms for identification of the excessive pollution sources were developed. One-, two- and three-dimensional diffusion mathematical models of pollutants transport in rivers at different initial and boundary conditions were constructed. The information support and the software of the package of identification of the excessive pollution sources, containing the developed algorithms implemented as a set of programs written on programming language Delphi in operating system WINDOWS, were designed. The service programs of information input/output and representation of the obtained results were developed. The package is supplied with the possibility of choosing one of the languages realized in it or adding a new language, if necessary. The comprehensive experimental testing of the developed program packages justified their high computational, operational and service qualities.

3. Conclusion

Three software packages for experimental data processing designated on identical methodological basis are offered. The packages comprise original models, methods and algorithms developed by the authors. The software packages are realized for IBM-compatible personal computers according to the generally accepted standard for similar production all over the world. The consumer can use them as a modern, convenient, simple and reliable tool for resolution the problems he deals with in the considered areas. Experimental investigations of the developed software packages and the algorithms realized in it have confirmed their high computing, operational and service qualities.

References