

УДК 519.24

The Numerical Probabilistic Approach to the Processing and Presentation of Remote Monitoring Data

Boris S. Dobronets and Olga A. Popova*

*Siberian Federal University
79 Svobodny, Krasnoyarsk, 660041, Russia*

Received 17.02.2016, received in revised form 18.03.2016, accepted 29.05.2016

The paper deals with a numerical probabilistic analysis as a method of processing and presentation of remote monitoring for the aggregation large amounts of data. For the aggregation we used histograms, polygons and splines. On the basis of aggregated data to identify relationships between the input and output characteristics are studied the theoretical and the practical aspects of regression modeling. On the basis of numerical examples we demonstrated the efficiency and reliability of the proposed methods.

Keywords: numerical probabilistic analysis, remote monitoring, data processing, histogram aggregation, frequency polygon, spline, regression.

Citation: Dobronets B.S., Popova O.A. The numerical probabilistic approach to the processing and presentation of remote monitoring data, J. Sib. Fed. Univ. Eng. technol., 2016, 9(7), 960-971. DOI: 10.17516/1999-494X-2016-9-7-960-971.

© Siberian Federal University. All rights reserved

* Corresponding author E-mail address: BDobronets@yandex.ru

Численный вероятностный подход к обработке и представлению данных дистанционного мониторинга

Б.С. Добронет, О.А. Попова
Сибирский федеральный университет
Россия, 660041, Красноярск, пр. Свободный, 79

В статье рассматривается численный вероятностный анализ как метод обработки и представления дистанционного мониторинга для агрегации больших объемов данных. Для агрегации мы использовали гистограммы, частотные полигоны и сплайны. На основе агрегированных данных для выявления взаимосвязей между входными и выходными характеристиками изучаются теоретические и практические аспекты регрессионного моделирования. На основе численных примеров демонстрируются эффективность и надежность предлагаемых методов.

Ключевые слова: численный вероятностный анализ, дистанционный мониторинг, обработка данных, гистограммная агрегация, частотный полигон, сплайн, регрессия.

Introduction

Control of complex systems, such as the distributed natural-technical objects, are classified as optimal control problems of a strategic nature. Typically, distributed natural-technical systems are characterized the interdisciplinary and interregional interaction of elements distributed over a large area. A distinctive feature of the management of such systems is to ensure their safety in the conditions of dangerous natural and man-made processes. An important problem to ensure the security of these systems is the preventively control of emergency and crisis states of the managed object. To increase the effectiveness of preventive control of distributed systems is necessary to calculate the assessment of potential risks based on remote monitoring data of the systems and the processes occurring in them, taking into account the impact of different types of uncertainty. For example, such methods include observation as land remote sensing. Remote sensing provides information about the object based on measurements taken at a distance from the object, that is, without direct contact with the object. An important aspect of remote monitoring is the ability to control large areas. It provides information generalizing the processes at the regional or all planet level. The information obtained on the basis of remote monitoring data are usually characterized by large volumes, varied methods of registration and the transfer forms. This information must be presented and properly processing to obtain useful knowledge. Preparation and data processing for research system processes includes a number of computational procedures. They must meet the requirements, such as reducing the uncertainty level, the accuracy and clarity of the results. By building a database for forecasting the state of complex distributed systems must be approached systematically, without losing a single opportunity to improve predictive accuracy of conclusions. Consider the features of the representation and data processing remote monitoring. Remote monitoring data are diverse in their qualitative and quantitative composition, the specificity of the methods of measurement, registration, submission. With access to a huge volume of remote monitoring data, researchers often need effective data processing technology and analysis.

This contributes to the development of new numerical methods and approaches to representation, numerical modeling and data analysis of remote monitoring in terms of different types of uncertainty. The variety of ways of presenting data determines the choice of methods of processing, modeling and analysis. The classical approach to data processing and analysis usually assumes that each value is the only point in n -dimensional space R^n [1]. In addition to such data in many practical problems of storage, processing, analysis, have to deal with “multi-value” data, examples of which include: a list of values, interval date and other types. A more complex type is modal – valued data containing the probability of weight, or any other related matter. A good example of a modal-valued data is a histogram. Such data gets to a data class, called “symbol” data [1]. An important difference between the symbol data is the presence of internal structures. It is does not apply the classical statistical theory and methods of data processing and analysis for symbolic data. Therefore, development of new methods for the analysis of such data and construct mathematical fundamentals are more important. Note that the large amount of information on the one hand, provides a more precise description of the object of research, on the other hand makes the difficulties for the search for solutions to the challenge. One way to solve this problem is to use a variety of data preprocessing procedures. For example, the aggregation method is a method of compressing large data. Procedures for the conversion of input data arrays into arrays of smaller dimension while maintaining useful information are the basis of aggregation. This article discusses the methods of presenting the data as a histogram, polygons, splines, histogram time series. On the basis of data representation methods we are solving the problem of data aggregation. Applying the numerical probabilistic analysis we are studying the existing relationships and trends in the data and are considering the questions of construction histogram regression models. For this purpose, the aggregated data we are presenting in the form of histograms and polygons.

Data representation

We will use a numerical probabilistic analysis for the processing and analysis of remote monitoring data, using such representation as histograms, polygons and splines. Histograms are one of the best known methods for presenting data.

Histograms. The histogram is called a random variable density which is represented piecewise constant function. Histogram P is defined grid $\{x_i | i = 0, \dots, n\}$, on each interval $[x_{i-1}, x_i]$, $i = 1, \dots, n$ histogram takes constant value of p_i . Histograms are widely used for the processing and analysis of remote sensing data. For example, in [2] is considered the problem of the study of natural processes on the basis of space and ground monitoring data. In addition to histograms we discuss the piecewise linear functions (polygons) and splines.

The piecewise linear function (Frequency Polygons). Piecewise linear functions can considered as a tool of approximation of the density function random variable. A piecewise linear function is a function composed of straight-line sections. The frequency polygon (FP) is a continuous density estimator based on the histogram. In one dimension, the frequency polygon is the linear interpolant of the midpoints of an equally spaced histogram.

Spline. A spline is a sufficiently smooth polynomial function that is piecewise-defined, and possesses a high degree of smoothness at the places where the polynomial pieces connect (which are known as knots). We will consider the probability density of the random variables are approximated spline.

Data aggregation

Aggregation can be considered as a data conversion process with a high degree of detail to a more generalized representation, by computing so-called aggregates or values obtained as a result of this conversion to a certain set of facts related to a specific dimension. An example of such procedures is a simple summation, calculation of the average, median, mode and range of maximum or minimum values. Application of aggregation procedures has its own advantages and disadvantages. Detailed data are often very volatile due to the impact of different random factors. It is making difficulties for detecting the general trends and patterns. It is important to bear in mind that the use of such procedures as averaging procedure, the exclusion of extreme values (emission), the smoothing procedure can lead to loss of important and considerable part of the useful information.

There are two typical situations where this happens [3]:

- If a variable is measured through time for each individual of a group, and the interest does not lie in the individuals but in the group as a whole. In this case, a time series of the sample mean of the observed variable over time would be a weak representation.
- When a variable is observed at a given frequency (say minutes), but has to be analyzed at a lower frequency (say days). In this case, if the variable is just sampled at the lower frequency, or if only the highest and lowest values between consecutive lower-frequency instants are represented (as is done when summarizing the intra-daily behavior of the shares in the stock markets), a lot of information is neglected.

These two situations describe contemporaneous and temporal aggregation, respectively.

Consider the histogram approach to data aggregation. This approach is useful for the following reasons. The histogram can be regarded as a mathematical object that is easy to describe and calculate the mathematical procedures and operations, while maintaining the essence of the frequency distribution of the data.

There is histogram arithmetic [4, 5], which allows you to perform various arithmetic operations on histogram variables, including the operation of calculating the maximum and minimum, raising to a power, the comparison procedure. Currently, histograms are used extensively in various fields. For example, histograms are widely used in databases and in the problems of approximating data, pattern recognition. Histograms are widely used in image processing problems, since the histogram characterizes the statistical distribution the number of pixels in the image depending on their values.

We note an important property of the histogram, which is associated with the aggregation of information. Note that the histogram adequately represents the distribution of random variables. Despite its simplicity, the histogram covers all possible ranges of probability density function estimation. Simple and flexible structure of histograms greatly simplifies their use in numerical calculations and has a clear visual image, which is useful for analytical conclusions.

Consider one example of a histogram approach to aggregate large amounts of data. Assume that the characteristics is monitored some of the observed object. For clarity, let it be temperature. Monitoring is carried out over the area Ω , which is represented as a union of subareas Ω_i . It is assumed that the temperature measurement points distributed uniformly in each of the subareas. Temperature as input parameter of standard model is described to average values x_i in the subarea Ω_i .

Remote sensing allows you to submit N_i temperature values for each subarea Ω_i . Where i is number of subarea, j is number of measurement. Thus, each x_i is represented as an average value

$$\bar{x}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{i,j}.$$

It is proposed to reduce the level of uncertainty in the data and instead of the average values of input parameters to use histogram representation for each subarea.

Consider the histogram P_x for the value x . Suppose that for the value x is known sample (x_1, x_2, \dots, x_N) . Denote by n_j the number members x_i of the re-sampling (x_1, x_2, \dots, x_N) got into interval $[z_{i-1}, z_i]$, then

$$p_j = \frac{n_j}{N(z_j - z_{j-1})}.$$

Thus, for each input variable x_i may be known, not only the average value \bar{x}_i but also the histogram P_i .

Fig. 1 shows a model example histogram P_i by values in a certain area Ω_i .

Assume that the area Ω_i is rectangle region. It is composed of 100×100 pixels. For each pixel value is put in appropriate value $P_{i,j}$. For clarity in Fig. 1 values $x_{i,j}$ are shades of gray. Lighter colors correspond to a higher temperature. Thus, the histogram describes the frequency distribution of the temperature into Ω_i

Considered approach demonstrates an effective way to aggregate data. So instead of 10^4 values represented in the subarea Ω_i for histogram is used only about 10^2 values. Moreover, histograms include a lot more information as interval of value change, the mean value, frequency and etc.

The frequency polygon (FP) is a continuous density estimator based on the histogram, with some form of linear interpolation. Transformation histogram in the frequency polygon can be represented graphically Fig. 2. [6].

In one dimension, the frequency polygon is the linear interpolant of the midpoints of an equally spaced histogram.

Theoretical basis of numerical probabilistic analysis

The subject of the NPA is to solve the various problems with data stochastic uncertainties using numerical operations on the probability densities of the random variables and functions with random

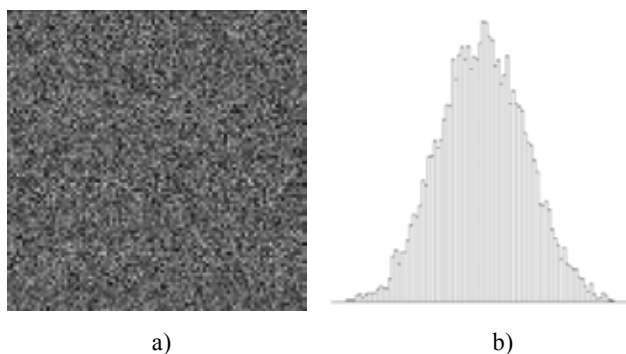


Fig. 1. a) – subarea Ω_i , b) – histogram P_i

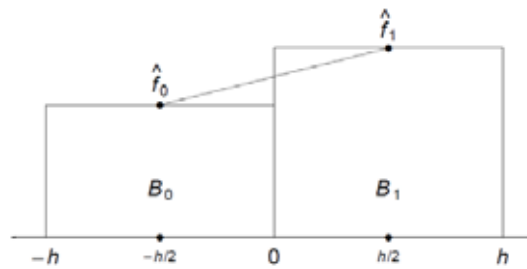


Fig. 2. The frequency polygon in a typical bin, $(-h/2; h/2)$, which is derived from two adjacent histogram bins

arguments. To this end, were developed varied tools, including concepts such as arithmetic, histogram, probability, histogram, probabilistic extension, second order histogram, histogram time series.

It is a non-parametric approach can be successfully applied for the probabilistic description of systems as part of an interactive visual simulation, thereby increasing the quality of research systems [4]. In the test cases and some practical problems were proved the advantage of this approach to the Monte Carlo method [5].

Consider the basic theoretical aspects of the NPA. Arithmetic for different data types is one of the important components of numerical probabilistic analysis. NPA-arithmetic reduces the uncertainty level in the data and helps to obtain the additional information on the distribution of random variables. These are operations “+”, “-”, “·”, “/”, “↑”, “max”, “min”, as well as binary relations “≤”, “≥” and some others. The numerical operations of the histogram arithmetic constitute the major component of NPA.

In [4, 5] developed numerical approaches to the various operations of the histogram variables, including the arithmetic operations, comparison operations, the procedure of calculating the maximum and minimum.

It is noted that the number of arithmetic operations to compute the procedure $x \cdot y$ by the histogram method over the variables x and y is near the order $O(n^2)$, where n is dimension of the histogram.

A histogram is a piecewise constant function approximates the probability density with an accuracy $O(1/n)$. However, even midpoints of histogram are approximated the probability density function with precision $O(1/n^2)$. Consequently, the frequency polygon approximates the function with the accuracy $O(1/n^2)$. Note that the arithmetic complexity does not increase.

Let us consider histogram operations. Let $p(x, y)$ be a joint probability density function of two random variables x and y . Let p_z be a histogram approximating the probability density of the operations between two random variables $x * y$, where $*$ \in $\{+, -, \cdot, /, \uparrow\}$. Then the probability to find the value z within the interval $[z_i, z_{i+1}]$ is determined by the formula [4, 5]

$$P(z_k < z < z_{k+1}) = \int_{\Omega_k} p(x, y) dx dy,$$

where $\Omega_k = \{(x, y) | z_k \leq x * y \leq z_{k+1}\}$ and the value P_k of the histogram on the interval $[z_k, z_{k+1}]$ is defined as

$$P_k = \int_{\Omega_k} p(x, y) dx dy / (z_{k+1} - z_k).$$

Then we extend the order relation $\succ \in \{<, \leq, \geq, >\}$ to random variables [7]:

$$x \succ y \text{ if and only if } x \succ y \text{ for all } x \in x, y \in y.$$

If the support of x, y are intersected, then we can talk about the probability of $x \succ y$

$$P(x \succ y) = \int_{\Omega} p(x, y) dx dy,$$

where $\Omega = \{(x, y) \mid x \succ y\}$ is the set of points $(x, y) \in R^2$ such that $x \succ y$, $p(x, y)$ is the joint probability density of x, y .

For example, consider the operation $\max(x, y)$. The probability $P(\max(x, y) < z)$ is determined by the formula

$$P(z) = \int_{\Omega_z} p(x, y) dx dy,$$

where $\Omega_z = \{(x, y) \mid (x < z) \text{ and } (y < z)\}$ and the value P_i of the histogram on the interval $[z_i, z_{i+1}]$ is defined as

$$P_i = (P(z_{i+1}) - P(z_i)) / (z_{i+1} - z_i).$$

Probabilistic extension. One of the most important problems that NPA deals with is to construct probability density functions of random variables. Let us start with the general case where (x_1, \dots, x_n) is a system of continuous random variables with the joint probability density function $p(x_1, \dots, x_n)$ and the random variable z is a function $f(x_1, \dots, x_n)$

$$z = f(x_1, \dots, x_n).$$

By *probabilistic extension* of the function f , we mean the probability density function of the random variable z .

Let us construct the histogram F approximating the probability density function of the variable z . Suppose the histogram F is defined on a grid $\{z_i \mid i = 0, \dots, n\}$. The domain is defined as $\Omega_i = \{(x_1, \dots, x_n) \mid z_i < f(x_1, \dots, x_n) < z_{i+1}\}$. Then the value F_i of the histogram on the interval $[z_i, z_{i+1}]$ is defined as

$$F_i = \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n / (z_{i+1} - z_i). \tag{1}$$

By *histogram probabilistic extension* of the function f , we mean the histogram F constructed according to (1).

Let $f(x_1, \dots, x_n)$ be a rational function. To construct the histogram F , we replaced the arithmetic operation by the histogram operation, while the variables x_1, x_2, \dots, x_n are replaced by the histogram of their possible values. It makes sense to call the resulting histogram of F as *natural histogram extension* (similar to “natural interval extension”).

Case 1. Let x_1, \dots, x_n be independent random variables. If $f(x_1, \dots, x_n)$ is a rational expression where each variable x_i occurs no more than once, then the natural histogram extension approximates a probabilistic extension.

Case 2. Let the function $f(x_1, \dots, x_n)$ admit a change of variables, so that $f(z_1, \dots, z_k)$ is a rational function of the variables z_1, \dots, z_k satisfying the conditions of Case 1. The variable z_i is a function of $x_i, i \in Ind_i$ and the Ind_i are mutually disjoint. Suppose for each z_i it is possible to construct the

probabilistic extension. Then the natural extension of $f(z_1, \dots, z_k)$ is approximated by the probabilistic extension of $f(x_1, \dots, x_n)$.

Case 3. We have to find the probabilistic extension for the function $f(x_1, x_2, \dots, x_n)$, but the conditions of Case 2 are not fulfilled. Suppose for definiteness that only x_1 occurs a few times.

If, instead of the random variable x_1 , we substitute a determinate value t , then it is possible to construct the natural probabilistic extension for the function $f(t, x_2, \dots, x_n)$.

Suppose that t is discrete random value approximating x_1 as follows. Let t take the value t_i with probability P_i and for each function $f(t_i, x_2, \dots, x_n)$ it is possible to construct the natural probabilistic extension.

Then the probabilistic extension f of the function $f(x_1, \dots, x_n)$ can be approximated by the probability density φ as follows [4]

$$\varphi(\xi) = \sum_{i=1}^n P_i \varphi_i(\xi).$$

Metrics for the probability density function

Due to the nature of histograms for numerical simulation of histogram regressions will use special approaches to quantify the distances between the probability density function. To this end, consider a few metrics that should be used in the histogram regression models

For this purpose we consider the metric for the probability density function of the random variables [3]

Let $f(x)$ and $g(x)$ are two probability density functions. Then distances between $f(x)$ and $g(x)$ are defined as

$$\rho_W(f, g) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt, \tag{2}$$

$$\rho_M(f, g) = \left(\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt \right)^{1/2}, \tag{3}$$

respectively, where $F^{-1}(t)$, $G^{-1}(t)$ – are the inverse cumulative distribution functions $f(x)$ and $g(x)$, respectively.

Let $h(x)$ be the histogram. Then distribution functions H corresponding to this histogram can be represented as

$$H(x) = \int_{-\infty}^x h(\xi) d\xi.$$

Due to the fact that the histogram is the piecewise constant function the calculation of the integral of it is not difficult. As a result, the distribution function is a piecewise linear function. Thus, the metrics (2) or (3) can be interpreted as the area between the distribution functions.

Numerical probabilistic approach to the representation of histogram time series

Histogram time series (HTS) describe the situations where for each time point are known histogram approximation of the probability density function of the studied parameter. It is important to

note that the forecast on the base of histogram time series presents the approximation of the probability density function of the studied process.

In order to construct the numerical procedures for data aggregation we define the histogram time series as a sequence of probability densities represented in the form of histograms. To construct and study the time histogram series will use tools of numerical probabilistic analysis.

Further based on the NPA and the concept of time histogram series are studied data aggregation tasks and construction the information analytical models.

Decomposition method of histogram time series. Consider the forecasting model construction based on the decomposition method of the histogram time series in the histogram subseries [8].

Denote by $h_i, i = 1, \dots, N$ some histogram time series. Let for each histogram is defined uniform grid $\omega^i = \{z_i^j, j = 0, \dots, k\}$, and values $p_i^j, j = 1, \dots, k$. Define by

$$z_i^0, z_i^k, p_i^j, j = 1, \dots, k, i = 1, \dots, N. \tag{4}$$

auxiliary histogram time series. Next, consider the approach of building a predictive model based on the decomposition of the time series in the histogram contracts.

To construct the forecast \hat{h}_{N+1} will be used the last d-th values of histogram time series (4).

Thus, using one of the methods for constructing the forecast, we construct values $\hat{z}_{N+1}^0, \hat{z}_{N+1}^k, \hat{p}_{N+1}^j$. Для гистограммы \hat{h}_{N+1} определим равномерную сетку $\omega^{N+1} = \{\hat{z}_{N+1}^j, j = 0, \dots, k\}$

Then, let's define an uniform grid $\omega^{N+1} = \{\hat{z}_{N+1}^j, j = 0, \dots, k\}$ for histograms \hat{h}_{N+1} with values \hat{p}_{N+1}^j . The final step in the construction of the forecast will be the normalization of the histogram \hat{h}_{N+1}

$$\hat{h}_{N+1}^{nor} := \frac{1}{\beta} \hat{h}_{N+1},$$

where

$$\beta = \int_{z_{N+1}^0}^{z_{N+1}^k} \hat{h}_{N+1}(\xi) d\xi.$$

$$\int_{z_{N+1}^0}^{z_{N+1}^k} \hat{h}_{N+1}^{nor}(\xi) d\xi = 1.$$

Numerical example. Consider the historical data of the maximum temperature in the city of Krasnoyarsk in the last hundred years. Сгруппируем данные по дням и построим соответствующие гистограммы.

To test the decomposition method we take histogram for 21–25 in August. According to the decomposition method we construct twelve auxiliary temporary series. Due to the properties of stationarity to construct forecast for the series $z_i^0 = 6^o, z_i^{10} = 35^o, i = 1, \dots, 5$, is trivial. Using the least squares method to ten subseries $p_i^j, j = 1, \dots, 10, i = 1, \dots, 5$, we can build a forecast for August 29th. Fig. 3 is shown the comparison of the true value and the forecast for August 29th. As can be seen from the comparison, the predicted value is well matched with the true value.

Regression Modeling

Regression modeling is research method to identify the existence of the various correlations between input and output data.

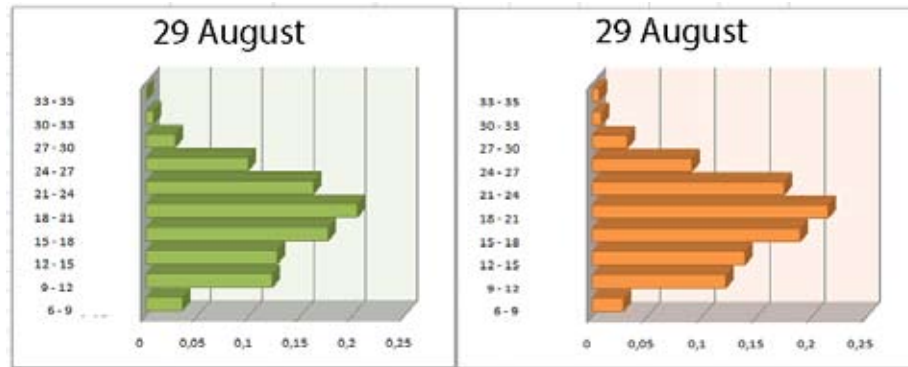


Fig. 3. True and predicted values

Regression analysis includes the following assumptions: the number of observations is sufficient to develop the statistical regularities concerning the factors and their interactions; processed data contains some errors (noise) due to measurement errors, and the influence of random factors; matrix of observations is the only information about the object being studied, the available before the start of the study.

To search for interconnections into remote monitoring data are proposed to use the data aggregation procedure and then to construct the regression model on aggregated data. Regression model to construct on the basis of the histogram data representation we call histogram regression.

Let $X = (x_1, \dots, x_n)$ be input data, and Y be output variable. Both, is represented as histogram variables. For $X = (x_1, \dots, x_n)$ we now joint probability density function $p(x_1, \dots, x_n)$.

Similarly classical nonparametric regression can be written for each pair (X_i, Y_i) by

$$Y_i = f(X_i, a) + \varepsilon_i, i = 1, \dots, N,$$

In the linear case the model can be defined as

$$Y_i = a_0 + \sum_{j=1}^n a_j x_{ij} + \varepsilon_i, i = 1, \dots, N.$$

Thus, for finding the unknown parameters the optimization problem can be represented as

$$\Phi(a) = \sum_{i=1}^N \rho(Y_i, a_0 + \sum_{j=1}^n a_j x_{ji})^2 \rightarrow \min, \quad (5)$$

where ρ be a metric in the space of histograms. Note that due to the nonlinearity of the addition operations under histogram variables to solve the problem (5) can be use the steepest descent method [9]. To calculate the gradient Φ' will use difference derivatives

$$\Phi'_i = \frac{\Phi(a_0, \dots, a_i + h, \dots, a_n) - \Phi(a_0, \dots, a_i, \dots, a_n)}{h},$$

where h be parameter. The initial approximation for the parameters vector can be obtained by solving the regression problem for mathematical expectations ($M[X_i]$, $M[Y_i]$).

Note that the histogram regression (5) can be regarded as a symbolic regression, since the input and output variables are histogram mathematical objects [1].

Numerical example. Consider the regression model based on the histogram time series aggregation. In this case, we know the measurements of the target variable Y_i . Measurements of the input variable will be treated as real.

Consider the temperature data for the last hundred years in the Krasnoyarsk city. For each day from April to October, the data is aggregated in the form of histograms. In this case, the regression model can be represented in the form

$$Y = A\varphi_1(t) + B\varphi_2(t) + C\varphi_3(t),$$

where A, B, C are the probability density functions, $\varphi_1, \varphi_2, \varphi_3$ — quadratic functions. Functions A, B, C represented in the form of Hermite cubic splines. Splines are defined by mesh $\{x_1, x_2, x_3\}$. Boundary conditions are $s(x_1) = 0, s'(x_1) = 0, s(x_3) = 0, s'(x_3) = 0$. In addition $s'(x_2) = 0$ and the value of $s(x_2)$ chosen from the conditions

$$\int_{x_1}^{x_3} s(\xi) d\xi = 1.$$

For variable x_1, x_3 is chosen by regression curves of minimum and maximum temperatures. For x_2 is chosen by regression curve of average temperatures

In Fig. 4 shows the regression of the probability density functions of the temperature data for the last hundred years in the Krasnoyarsk city, from April to October. Shades of gray define the values of the probability density function. The top and bottom line represents maximum and minimum temperature on each day over the past hundred years respectively. Midline denotes the mean of daily temperature over the last hundred years. Each vertical section is the probability density function of the temperature corresponding to a certain day of the year, according to the observations of the day in the last hundred years. At the first stage, the data presented for each day in a histogram. The regression data are presented in the form of Hermite cubic splines.

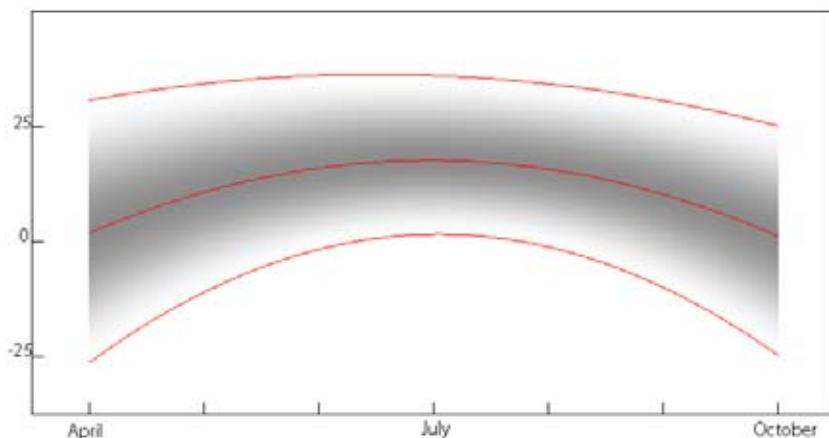


Fig. 4. NPA regression

Conclusion

Using NPA for preprocessing, processing and modeling of remote monitoring data based on the histogram aggregation contributes to the reliability of the study of natural systems and processes. The spatial and time aggregation procedures help to reduce the amount of computation in data processing and are an important basis for the extraction of useful knowledge from large volumes of data. Developed on the basis of the NPA submission methods, processing and numerical regression modeling, reduce the level of uncertainty in the information flow, significantly reduce the processing time and the implementation of numerical procedures.

This approach allows the mode of interactive visual modeling to provide the necessary data for operational decision making under remote surveillance techniques and distributed object systems.

References

- [1] Billard L., Diday E. *Symbolic Data Analysis: Conceptual Statistics and Data Mining* John Wiley & Sons, Ltd. 2006. 321 p.
- [2] Добронетц Б.С., Попова О.А. Гистограммный подход к представлению и обработке данных космического и наземного мониторинга. *Известия Южного федерального университета. Технические науки*, 2014, 6(155), 14-22. [Dobronets B.S., Popova O.A. Histogram approach to representation and processing of data space and data ground monitoring. *Izvestiya SFedU. Engineering science*, 2014, 6(155), 14-22. (In Russian)]
- [3] Arroyoa J., Mate C. Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting*, 2009, 25, 192–207
- [4] Dobronets B.S., Popova O.A. Numerical probabilistic analysis under aleatory and epistemic uncertainty. *Reliable Computing*. 2014, 19(3), 274-289.
- [5] Dobronets B.S., Krantsevich A.M., Krantsevich N.M. Software implementation of numerical operations on random variables. *J. Sib. Fed. Univ. Mathematics & Physics*, 2013, 6(2), 168–173.
- [6] Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, John Wiley, 2015, 380 p.
- [7] Popova O.A. Optimization problems with random data. *J. Sib. Fed. Univ. Mathematics & Physics*, 2013, 6(4), 506-515.
- [8] Попова О.А. Гистограммный информационно-аналитический подход к представлению и прогнозированию временных рядов. *Информатизация и связь*. 2014, 2, 43-47. [Popova O. Histogram and information — analytical approach to representation and for time series forecasting. *Informatization and Communications*, 2014, 2, 43-47. (In Russian)]
- [9] Васильев Ф.П. *Численные методы решения экстремальных задач*. М.: Наука, 1988. 552 с. [Vasil'ev, F. P. *Numerical methods for solving extremal problems*, Moscow, Nauka, 1988, 552 p. (In Russian)]