ISSN 2349-4506
Impact Factor: 2.265

Global Journal of Engineering Science and Research Management

COMMON METHODOLOGY OF DATA ANALYSIS OF NATURAL DISASTERS

Mohammad Ayaz Ahmad *, Oleksandr Kuzomin, Vyacheslav Lyashenko, Bohdan Tkachenko
*Physics Department, Faculty of Science, P.O. Box 741, University of Tabuk-71491, Saudi Arabia

DOI: 10.5281/zenodo.51752

KEYWORDS: model, forecast, data analysis, natural disaster, statistic model.

ABSTRACT

Providing of safe human life is very connected with research and analysis of different disasters. Among of different disasters, there is a special place for natural disasters (ND), because ND have significant influence on human life. One of the methods to prevent uncontrolled ND influence on human life is wide data research which are connected with appearance and distribution of ND. Thus, in this work researched the whole methodology of data analysis of ND. Basis of this methodology is statistic model of ND data analysis and development of forecast ranking of ND development.

INTRODUCTION

In human life there are different situations which are connected with interruption of stable state of nature objects [1, 2, 3]. Such development of situations and events connected to them may occur under influence of a sum of external factors and in some cases it has chaotic distribution which develops and does not stop. It results to economical, social and human loss. In general, such situations can be characterized as natural disasters (for example: avalanches, mudflow, landslides, earthquake, tsunami, volcanic eruption, flood, hurricanes и tornadoes) prediction and planning of emergencies can minimize loss and as a result to minimize human loss. So we need to research different natural disasters to save humans life.

Traditionally in the core of building of such systems is continuous monitoring of such situations and data accumulating (about processes that are connected to ND). In this connection monitoring of ND is also actual – observing, control and foresight of dangerous processes and nature phenomenon (which are the source of ND) and the dynamic of development of ND, determining of their scale to alarm and to control results of ND. This is eventually the goal of this research - to get the basic concept when choosing the methodology of ND analysis.

MATERIAL AND METHODS

Model development to identify natural disasters

Model of source object is providing of object in some form different from its real form. Nowadays there are exists and wide used in science and engineering a lot of different model types and modeling methods. If to take as a basis the level of abstraction then it is possible to determine next types of models to analyze ND [4, 5, 6]:

Physical (natural) - they are restoring researched process with saving of its physical nature and they are tools of physical modeling;

Analog - they replace one object with another one with similar properties;
Mathematical - abstract model which are exists in the form of special meth constructions and they have meaning only for interpreter human of machine.

In the same time modeling of processes of appearance and distribution of ND can be categorized as identification task. Identification task can be formalized in the following way: by results of observing on input and output variables is being build optimal (in some meaning) system model of interaction of identification object with researched parameters.

This interaction is happening in input and output channels in ND in discrete time moments. Then we can assume that such observation is connected to some operator of ND observing object which we need to identify. So identification task can be decomposed to the following sub-tasks:
Selecting of input variables (factors or arguments);

ISSN 2349-4506
Impact Factor: 2.265
Global Journal of Engineering Science and Research Management

Selecting of output variables (parameters of optimization);
Selecting of model structure;
Research of algorithms of model usage to find extreme of output variables.

Basic regularities in solving of modeling task to identify ND
For modeling how to identify the problem ND, first of all, it is expedient to use the following basic laws:
$A-$ Associations;
$\Pi-$ Sequences;
$K-$ Classifications;
$KL-$ Clustering.
This is due to the fact that, for example, the association takes place in the case, if current, natural problem situation e_t, where $t-$ current time, and prior or early occurring situation S_i were linked with each other. In other words, it can be argued that there is link $A: e_t \to S_i$ [7].

At the same time, for Π is accurate statement $\Pi: S_i^{t-1} \to S_i^t$ as there is a chain of related situations in time [7].

With the help of classification K we can write the following expression $-S_i = K\{(S_i \in K_{kr}) \vee (S_i \in K_{\neg kr})\}$, where $K_{kr}-$ class of disasters, and $K_{\neg kr}-$ class of non-disasters, and identification signs of entered micro situations [7], that characterize a group of individual situations. Wherein, this is done by analyzing the situations that are already classified and formulation a set of rules.

Clustering $KL: S_i = KL\{(S_i \in KL_{kr}) \vee (S_i \in KL_{\neg kr})\}$ differs from the classification, that groups themselves are not set in advance. With the clusterization of tool Data Mining independently allocated different homogeneous groups of data [8].

RESULT AND DISCUSSION

The ideology of the data analysis for the origin and development of ND

Tasks of a common methodology of analysis of environmental data $\{Zps\}$ for the development of systems of information-analytical system (IAS) in preventing ND determine the functionality of the complex and meet the requirements of IAS structure $\{Str\}$ generally.

Optimal IAS development process $\{Pr\}$ (in the sense of satisfaction of relevant quality indicators $\{Pk\}$) should lead to the implementation of the most effective structure providing complex IAS $\{Str\}$.

At the same time, hardware components and subsystems Str_i $(i = \overline{1,K})$ of providing complex ISS should the best way implement the requirements and characteristics of the statistical model ND $\{Mod\}$, methods of development $\{Met\}$, algorithms for decision making search $\{Alg\}$ both for the design of ND and analysis of situations, ensure the decision of the best solutions in the ND. In this case we have to solve the Z problems $\{Zps\}$ developing of the structure of IAS $\{Str\}$ it should be presented in the following form:

$$\{Met, Alg, Prog, Mod, Zps, Pk\} \xrightarrow{\{Pr\}} \{Str\}, \quad (1)$$

$$\{Str_i \in Str\},\ i = \overline{1,K}, \quad (2)$$

$$Pk^e = \underset{Met^e \in Met}{extr} \underset{Alg_j^e \in Alg}{extr} \underset{Prog_\varphi^e \in Prog}{extr} \underset{Mod_t^e \in Mod}{extr} \underset{Zps_t^e Zps}{extr} Pk(Met^e, Alg_j^e, Prog_j^e, Mod_t^e, Zps_t^e), \quad (3)$$

ISSN 2349-4506
Impact Factor: 2.265
Global Journal of Engineering Science and Research Management

Where Met^e – e – th method of IAS development, $e = \overline{1,E}$;

Alg_j^e – j – th algorithm of algorithmic, IAS provides a complex method for IAS development, $j = \overline{1,J}$

.

$Prog_{\varphi_j}^e$ – φ – th software program that provides for the implementation of IAS complex of the algorithm in the method of the IAS development, $\varphi = \overline{1,\Pi}$;

Mod_t^e – t – th a model of the environment in e – th IAS development method, $t = \overline{1,T}$;

Zad_η^e – η – th the task of environmental monitoring and forecasting for e – th IAS development method, $\eta = \overline{1,Z}$;

Pk^e – Guarantee of the quality of the complex IAS for e – th method of IAS development.

Statistical model of ND analysis

Let us consider process of creating a statistical model ND {Mod} from system analysis positions. When developing a statistical model {Mod} you must, first of all, choose the most effective options for modeling, that deliver reliable estimates of ND given with a minimum of cost and time. Wherein, {Mod} ND representation should reflect the natural connection with the search for managerial situations, organizational decisions $\Im$, aimed at the prevention and elimination of consequences emergencies, which will provide a minimum of risk to human life R_e and require the necessary resources – Σ. Then the solution of the problem posed above for i situation of the environment in a given controlled area is need to develop a model of the ND as follows:

$$Mod_i = \underset{r_e^i \in R_e}{extr} Mod\left\{X^i, \Sigma^i, \Im^i, r_e^i\right\}, \quad (4)$$

where $i = \overline{1,N}$ – the number of possible situations of the environment.

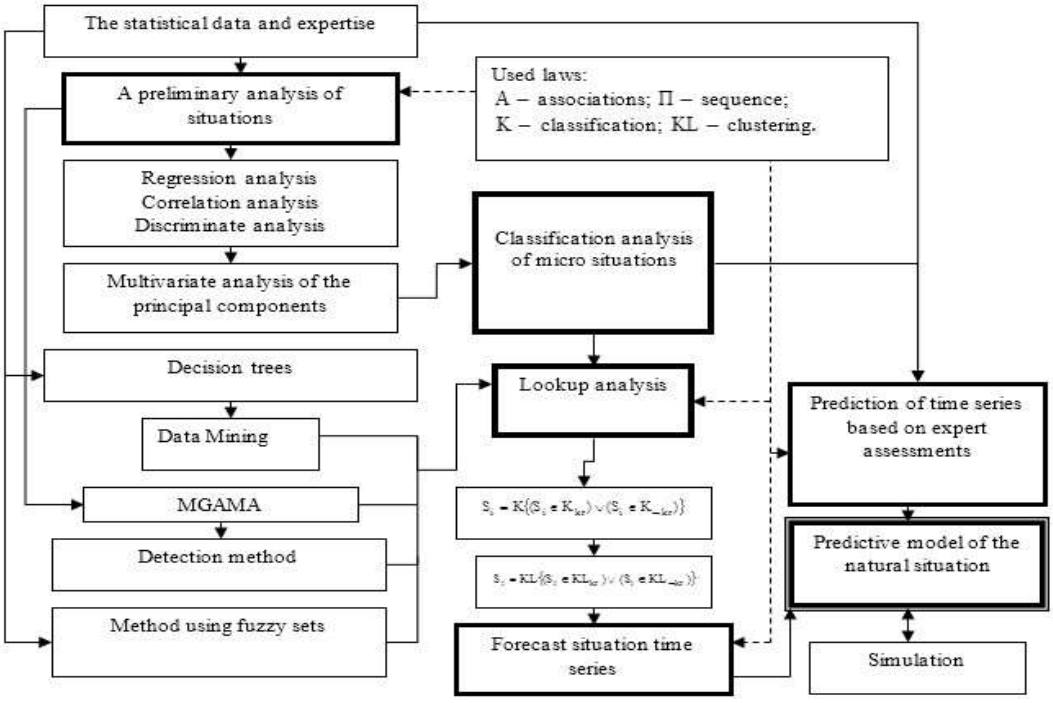

Figure. 1. ND data analysis as a set of different data aggregation techniques

In practice, statistical processing of data for any natural situations starts with the graphical data analysis. To do this, it is advisable to use the ideology of regression analysis, dispersion of multivariate analysis, building a decision tree (decision trees), Data Mining, method of group account of the main arguments (MGAMA), pattern recognition method, a method using fuzzy sets (Fig. 1).

Then the results of the preliminary analysis of a priori data is constructed situational model of the environment for j situations and i environment parameter:

$$\mathrm{Mod}_{\mathrm{nc}_i}^{j} = f\left(X_1^{1,f}, X_2^{2,f}, \ldots X_i^{j,f}, \ldots X_N^{M,f}\right), \tag{5}$$

Where $j = \overline{1,M}$ – the number of micro-situations;

$i = \overline{1,N}$ – The number of parameters; $\{X_i^j\}$ – set of environmental parameters. With condition $\bigcap_i W_i^j = \varnothing$.

At the same time in a controlled environment at time t we have generalized situational model at $L \in M$ (not all models of the same number of micro situations):

$$\mathrm{Sit}_t = \begin{cases} \mathrm{Sit}'_{1,1} = f\left(X_{1,1}^{M'_1}, X_{2,1}^{M'_1}, \ldots, X_{i,1}^{M'_1}, \ldots, X_{N,1}^{M'_1}\right) \\ \mathrm{Sit}'_{2,2} = f\left(X_{1,2}^{M'_2}, X_{2,2}^{M'_2}, \ldots, X_{i,2}^{M'_2}, \ldots, X_{N,2}^{M'_2}\right) \\ \ldots \\ \mathrm{Sit}'_{j,d} = f\left(X_{1,d}^{M'_f}, X_{2,d}^{M'_f}, \ldots, X_{i,d}^{M'_f}, \ldots, X_{N,d}^{M'_f}\right) \\ \ldots \\ \mathrm{Sit}'_{M,t} = f\left(X_{1,t}^{M'_F}, X_{2,t}^{M'_F}, \ldots, X_{i,t}^{M'_F}, \ldots, X_{N,t}^{M'_F}\right) \end{cases}. \tag{6}$$

Forecasting as an element of ND analysis

Based on a preliminary analysis of a priori information about ND (sampling) prediction problem can be solved (continuation) multivariate time series (random process with discrete time), as well as expert information provided by a set of uncoordinated probabilistic statements of several experts.

Decision function can be constructed to predict the time series, which will reflect the probabilistic properties of the studied stochastic process, and represents an estimate of the conditional distribution of variables predicted in the space at a given point in time prehistory.

Suppose that we have

$X = \{X_1, X_2, \ldots, X_j, \ldots, X_n\}$ – a set of data of natural situation;

D_j – the set of admissible values of the variable X_j (all D_j assumed to be bounded);

$D = \prod_{j=1}^{n} D_j$ – space of values,

$x = (x_1, \ldots, x_n) \in D$ – tuple of variables from the data set, which is available.

Lets record $N+1$ time samples $t = \overline{1, N+1}$.

Then, time series forecasting problem is to the known values for the variables N samples (history) to evaluate the values of the variables for $N+1$ – forecasted time. The indices corresponding to the countdown, we will place

ISSN 2349-4506
Impact Factor: 2.265
Global Journal of Engineering Science and Research Management

the left. In this way, ${}_t x$ denotes a tuple of variables at time t. For the convenience of recording history and introduce the notation $\overline{1,r}E^i$ and $\overline{1,m}_t X^{def} = ({}_{t-1}X,\ldots_{t-m}X)$ – composite tuple variables for history. Wherein $\overline{l,m}_t x \in D^{m-l+1}$, where D^k – k-th power of space D. Then D^{N+1} is a space of realizations of a random process under study.

Resulting Function of Nd Analysis Forecasting

Let's formulate the task of building of solving function. For this we will consider statistic game $\langle C, \tilde{C}, \Delta, V \rangle$, where C – set of nature strategies, $\tilde{C}$ – set of decisions (in our case $\tilde{C} \subseteq C$), $\Delta(C, \tilde{C})$ – loss function, V – observing space.

With nature strategy $c \in C$ we will link ${}_t P_c(D/\overline{1,r}_t x)$ – conditional distribution in space of variables of t countdown with known values of countdown variables $t-i, i = \overline{1,r}$, where $r = r(t,c)$ – length of essential history (affecting on the distribution for current countdown).

We assume that probabilistic properties of a process are not changed with time, i.e. ${}_t P_c(D/\overline{1,r}_t x) = P_c(D/\overline{1,r} x)$ – conditional distribution does not depend on t.

Then forecasting task can be solved by C strategy recovery. Wherein needed to specify an algorithm $Q: V \to C$, which by received empirical information $v \in V$ builds ${}^0E^{i,k} \in D$ – nature strategy evaluation. In empirical data can be at the same time appear realization of process history and probabilistic statements of the experts, where $v^1 = \overline{1,N}_{N+1} x$ and $v^2 = \{B_i \in B \mid i = \overline{1,M}\}$. Here B – subset of all kinds of statements like $B_i = \langle \overline{1,r}E^i, \{({}^0E^{i,k}, p^{i,k}) \mid k = \overline{1,K_i}\}, \gamma_i \rangle$, where $\overline{1,r}E^i \in D^r$, ${}^0E^{i,k} \in D$, $p^{i,k} = P({}^0x \in {}^0E^{i,k} / \overline{1,r} x)$ – estimation of probability of hitting ${}^0E^{i,k}$ with any $\overline{1,r}x \in \overline{1,r}E^i$; γ_i – estimation of statement credibility.

In this way, solving of forecasting task for ND analysis is to receive a posteriori (with given empirical information which sometimes is absent or its accumulation is not set up yet, so it is not possible to always use this forecasting method) probability measure on natural strategies using Bayes formula [9, 10]:

$$P(c/\nu) = \left(\int_C g_c(\nu) dP(C)\right)^{-1} g_c(\nu) P(C), \quad (7)$$

Where $g_c(\nu)$ – credibility function for empirical information (the definition of which requires additional costs for research),

$P(C)$ – a prior distribution on the class nature of the policies.

For the ND forecast correction by the MGAMA method [11] detected math description of natural situation as a functional transform «enter-escape» like:

$$W_t = F(\underline{x}(t)), \quad (8)$$

where W_t – result of natural situations (avalanche, mudflow, landslide as a result characterized in time of moving the mountain rapidly increasing masses of snow, rocks, etc. with increasing speed and volume);

$\underline{x_t}$ - Input vector of parameters that are being changed in time (meteorological influence) with limited capacity of sampling; $t_0 \le t < t_N$ - time range change of a controlled situation.

Then to determine of exit model measure matches W_t^{Mod} and of observed exit W_t serves quality criteria $Q(W_t, W_t^{Mod})$, which is the result of forecast analysis

$$Q = \begin{cases} r \le \Delta_{dop} - \text{forecast was confirmed} \\ r > \Delta_{dop} - \text{forecast has not been confirmed} \end{cases}, \quad (9)$$

Where r – proximity measure of real and model situation;
Δ_{dop} – Maximal allowed proximity value

We need to accumulate the statistical data to get rankings of short-term and long-term prediction. This data effects on the size of predicted parameter or correlated with it.

CONCLUSIONS

In this work we provided common methodology of data analysis of appearance and distribution of ND. The essence of such methodology is based on a building of aggregation of base models and procedures of data analysis which reveal the conditions of appearance and distribution of ND. First of all, it allows to take into account the range of appearance and distribution of ND influence factors changing and also to build adequate procedures of their prediction. Additionally, marked that the basis of ND prediction is a prior and historical information which is not always available for researchers. That's why the most acceptable option is comparisonal ND ranking like comparison of events on the basis of micro situations analysis. Such approach allows to increase reliability of a quantitative ND analysis data and to provide efficient of their prediction.

REFERENCES

1. Sargiacomo M. Earthquakes, exceptional government and extraordinary accounting. Accounting, Organizations and Society, 42: 67-89, 2015.
2. Blaikie P., Cannon T., Davis I., Wisner B. At risk: natural hazards, people's vulnerability and disasters. Routledge, 2014.
3. Kuzemin O., Lyashenko V. Microsituation Concept in GMES Decision Support Systems. In "Intelligent Data Processing in Global Monitoring for Environment and Security", Sofia: ITHEA, 2011, p. 217–238.
4. Lehning M., Bartelt P., Brown B., Fierz C. A physical SNOWPACK model for the Swiss avalanche warning: Part III: Meteorological forcing, thin layer formation and evaluation. Cold Regions Science and Technology, 35(3): 169-184, 2002.
5. Sovilla B., Burlando P., Bartelt P. Field experiments and numerical modeling of mass entrainment in snow avalanches. Journal of Geophysical Research: Earth Surface, 111(F3): 97-123, 2006.
6. Kuzemin A., Lyashenko V. Probabilistic and multivariate aspects of construction of the models and procedures for prediction of the avalanche-dangerous situations initiation. Fifth International Conference "Information Research and Applications", i.TECH 2007, 26-30 June, 2. p. 284–288.
7. Ahmad M. A., Kuzemin O., Lyashenko V., Ahmad N. A. Microsituations as part of the formalization of avalanche climate to avalanche-riskiness and avalanche-safety classes in the emergency situations separation. International Journal of Advanced Research, 3(4): 684-691, 2015.
8. Kuzemin A, Fastova D, Yanchevsky I. Methods of adaptive extraction and analysis of knowledge for knowledge-base construction and fast decision making. International Journal on Information Theories&Applicatios, 12 (1): 93-99, 2005.
9. Keller M., Pasanisi A., Marcilhac M., Yalamas T., Secanell R., Senfaute G. A Bayesian methodology applied to the estimation of earthquake recurrence parameters for seismic hazard assessment. Quality and Reliability Engineering International, 30(7): 921-933, 2014.

10. Eckert N., Keylock C. J., Bertrand D., Parent E., Faug T., Favier P., Naaim M. Quantitative risk and optimal design approaches in the snow avalanche field: Review and extensions. Cold Regions Science and Technology, 79: 1-19, 2012.
11. Ivakhnenko A. G., Muller I. A. Self-organization of predictive models. Tekhnika, Kiev, 1985.