ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL – 2012, Vol. 01, No. 2, 35–38

The algorithm of bifurcation points forecasting in the analitical researches of complex agro-ecological systems

V. OSYPENKO

Associate Professor of Automation and Robotics Systems Department of the National University of Life and Environmental Sciences of Ukraine, 03041, Kyiv, st. Geroiv Oborony 15, Ukraine, tel.: +038 050 411 7277, E-mail: vvo7@ukr.net

Received March 18.2012: accepted April 25.2012

Abstract. An original development is described in this work of the forecasting of so-called rare events concerning the agro-ecological systems. Under definition "rare" it is necessary to understand events which take place during some observed process, time intervals between them are so great that it is possible to consider that they practically do not influence each other. The beginning of the rare events can be called as a "bifurcation point" of the observed process. Both the multistage procedure of the forecasting based on principles of inductive modeling and the selection criterion of the best forecasting models are described.

Keywords: inductive modeling, algorithm of modeling, criterion of models selection, agro-ecological system, forecasting, bifurcation points, rare events.

INTRODUCTION

In the projects of the complex agro-ecological systems researches it is necessary to decide tasks when some phenomena have a substantial influence on the result, but they take place very rarely and are hardly predictable. Time intervals between such events are so great that it is possible to consider that they practically do not influence each other, i.e. it is possible to consider that cross-correlation dependence between them is absent. Such phenomena have got the name of «rare» events and the beginning of such rare events can be called as a "bifurcation point" of the observed process. The classic example of rare events can be a date (a top) of freezing of the river or water storage in a natural environment. Although in the middle zone of Europe freezing takes place practically annually, time intervals between its occurrences are so great, that it is possible to assume that each instance of it has no influence on any other. It is also possible to include natural cataclysms, for example typhoons, hurricanes, earthquakes etc. to the category of rare events which can have the substantial negative affect on the agrarian sector of economy and on the mankind habitat in general.

Another example of a rare event can be an application of some complex technical unit (engine, computer etc.) which is characterized by a natural similarity and almost identical operating conditions. It is possible to continue this list of rare events. Therefore, undoubtedly, the successful forecasting of rare events is essential for the decision of many tasks in the ecology, economics and in research of reliability while complex technical systems are being tested.

In [1] the decision of the forecasting task of the water basin freezing date determined with the harmonic algorithm of Group Method Data Handling (GMDH) [2]. The rare event was the moment (on forecast interval) in which the forecast value of the process passed its critical point.

In the present article [3] the original approach for the forecasting of rare events is offered. It uses a technique that is effectively used by experienced specialists in diagnosing and predicting failures of technical devices. By 'rare' it is suggested to understand an event, coming in the object under analysis, for which there have been no precedents in recent past (not exceeding the maximum time of delay τ_{max}).

THE PROBLEM OF RARE EVENTS FORECASTING IN AGRO-ECOLOGICAL SYSTEMS

Under the object of modeling design in the agro-ecological systems we will understand the same physical object investigated on a time interval long enough for the multiple (n times) observations of an interesting event, or few (n) observations of the same type ecological objects, which is observed at the same conditions in a time interval including only one event. Obviously, these cases can take place both in ecological and in technical applications. However, usually the first case goes to the ecological systems (where the same water basin is explored during the sufficient amount of years from the point of view of modern modeling methods of the freezing moment). The second one is closer to the technical systems. An example is when simultaneously the monitoring of a group of the same type engines or other technical devices takes place, from the beginning of exploitation to the moment of their failure.

Obviously, in case of our agro-ecological (ecological) modeling systems both cases can take place and both procedures of initial information reception are acceptable.

Traditionally, it is possible to formulate the task of forecasting as follows. There is the prehistory of an object's behavior on the observation interval $\grave{O}_{obs} = [t_0, t_k]$, fixed in a corresponding datasheet (in the database). It is required to synthesize a model describing the behavior of this object on the forecast interval $\grave{O}_f = [t_{k+1}, t_{\acute{o}}]$.

We will formulate the task of forecasting of the rare event ("bifurcation point" of the process) as follows: let (1) – the result of monitoring of some system on the interval $\grave{O}_{obs} = [t_0, t_k]$ n times of interesting us event ξ_i, $i = 1, 2, \ldots, n$. Or (2) - monitoring of n same type agro-ecological systems in which the event once took place. Further, we will follow the scheme (1).

Obviously, is possible to split up all the interval $\grave{O}_{obs}$ into n intervals: $\grave{O}_{obs} = [t_0, t_{S_1}, \ldots, t_{S_{i-1}}, t_{S_i}, \ldots, t_{S_{n-1}}, t_{S_n}]$, where t_{S_i} is the moment of i-th event. Such splitting follows the indicated assumption that only one event ξ_i took place on interval $[t_{S_{i-1}}, t_{S_i}]$. In addition, every interval $[t_{S_{i-1}}, t_{S_i}]$ is broken into l of narrower intervals $\Delta t' = [t'_{j-1}, t'_j] = const$, $j = 1, 2, \ldots l$ and in its knots the control of parameter tests of object are produced. Thus, there is the set of n moments of events ξ_i, $i = 1, 2, \ldots, n$ in our task. The forecast of $(n + 1)$-th moment of time is the subject of our researches.

Such a problem can be described through the following regressive equation of a model:

$$y_f = f\{x_{1(0)}, x_{1(-1)}, \ldots, x_{1(-\tau_1)},\\ x_{2(0)}, x_{2(-1)}, \ldots, x_{2(-\tau_2)},\\ \ldots\\ x_{m(0)}, x_{m(-1)}, \ldots, x_{m(-\tau_m)}, \theta_f\}, \tag{1}$$

where: y is the output (forecast) value, x_i, $i = 1, 2, ..., m$ - arguments, $\tau_1, \ldots, \tau_m$ are the delays of each argument taken into account, θ is the vector of the estimated parameters.

More laconically, model (1) can be presented as

$$y_f = f(X, \theta_f). \tag{2}$$

The difference of such approach from traditional forecasting procedures: (1) among the arguments of function $f(\cdot)$ the delay arguments of output value y are absent and (2) - output value is the time between the last supervision (control measuring) and beginning of rare event (bifurcation point of process). Thus, on the interval $\Delta t' = [t'_{j-1}, t'_j]$ of rare event, the occurrence $y \leq \Delta t'$, and on the interval of «non-occurrence» (precedence) – $y > \Delta t'$.

In [3], for the solution of the same problem, the original and effective method of initial informative base forming is also presented. This procedure got the name «floating scale» to indexation of delay arguments. «Floating» indexation means that index «0» appropriated to the control moment of event has occurred. In this case we have a situation $y \leq (t'_{j-1} - t'_j)$. «Floating» indexation must be used for the creation of datasheet. Values $x_{i(-\tau_i)}$ must correspond to the delays of i-th interval, $i = 1, 2, ..., n$.

PROCEDURE FOR PREDICTION OF RARE EVENTS IN AGRO-ECOLOGICAL SYSTEMS

The feature of the agro-ecological systems is that in complex processes in such systems the rare events of interest can take place in very limited times. If to consider the second variant of rare events research, i.e. the similar agro-ecological systems, then the amount of rare events in them can be rather small. In both cases the statistical data of initial supervisions are very limited. Thus, to the algorithms which could be possible to use for the modeling and decision of prediction task of rare event, strict requirements are demanded:

1) algorithms must save operability at limited low times of supervisions (n);
2) algorithms must save operability at high ratio signal/noise to be antijamming;
3) algorithms must have high speed and be able to process large datasheets for the modeling of optimal results in the form (1).

For today, the techniques of inductive self-organization of complex systems correspond to such strong conditions [4, 5]. Our multistage procedure of rare event forecasting can be presented as a next kind.

The 1-st stage: the designing of primary informative base (datasheets). Informative delay arguments (on τ) are needed to determine the extremum of autocorrelation function or rank correlation. It allows to take into account only those delays which mostly influence the investigated process in the agro-ecological system and to eliminate a large bust of all possible combinations of delay, which greatly simplifies the output expression (1).

The 2-nd stage: a synthesis of equation (1) - is teaching of a model. This stage is not less responsible than the stage of basic data preparation. Here, it is appropriate to apply the so-called combinatory (at small m) or multistage (at sufficient m) algorithms of Group Method of Data Handling (GMDH). Simplifying denotations, it is possible to present for example the chart of combinatory algorithm in another way.

Step 1. The equation of the model contains only 1 member (if necessary it is possible to include a free member of a_o) in the equation:

$$y = f(a_0, x_i),\ i = 1, \ldots, m \tag{3}$$

Step 2. Equation (1) contains 2 (with free member - three) elements:

$$y = f(a_0, x_i, x_j),\ i, j = 1, \ldots, m,\ i \neq j. \tag{4}$$

Step *s*. Equation (1) contains *s* (with free member – *s*+1) elements:

$$y = f(a_0, x_1, \ldots, x_2, \ldots, x_s).$$

The 3-rd stage: the prediction of $\xi_{(n+1)}$. Let on *(n*+1*)*-th interval $[t_{sn}, t_{s_{n+1}}]$ the regular observations under the above conditions be made. While reaching the length of the interval equal to the maximum delay in equation (5), *y* is calculated. If *ó* > $\Delta t'$, the next control measuring is produced. Such procedure is performed until the output value becomes smaller than interval $\Delta t'$. In this case, the value of *y* will be the time after expiration of which an event will occur after the last moment of observation $\xi_{(n+1)}$.

The increase of model complication takes place until reaching the minimum of selection criterion or until the moment of its stabilization takes place. In the last case the model corresponding to the minimum complexity is selected. Under "complication" in an inductive modeling, the amount of members in the right part of equation (1) is understood.

Thus, the teaching of model (1) is the task of identification which, from the positions of inductive modeling, is exhaustively formulated in [6], as follows. The task of identification consists of forming, from the observation data $W = (X \vdots y)$ of the same set of $\Im$ models having different structures of the kind $\hat{y}_f = f(X, \hat{\theta}_f)$, where θ is the vector of the estimated parameters and selecting the optimal model under the minimum criterion $CR(\cdot)$:

$$f^* = \arg\min_{f \in \Im} CR(y, f(X, \hat{\theta}_f)) \tag{5}$$

where: estimations of parameters $\hat{\theta}_f$ for each $f \in \Im$ are the decision of task

$$\hat{\theta}_f = \arg\min_{f \in R^{s_f}} Q(y, X, \hat{\theta}_f), \tag{6}$$

where: $Q(\cdot) \neq CR(\cdot)$ is the criterion of decision quality in the parametric identification task of private model of complexity s_f generated in the task of structural identification (1).

Most the often applied criterion of the models selection in the indicated algorithms is the criterion of regularity:

$$AR(s) = \left\| y_B - \hat{y}_{B_s} \right\|^2 = \left\| y_B - \hat{y}_{B_s} \hat{\theta}_{A_s} \right\|^2. \tag{7}$$

This criterion, as well as all the criteria in the inductive modeling of the complex systems, has the properties of external addition [2] which assume the breaking up of the set $W = (X \vdots y)$ into two non-overlapping subsets: teaching *A* (for the evaluation of models parameters) and verification (for the calculation of model errors, $A \cap B = \varnothing$). The form of record and the denotations are stored by the theoretical work [6].

Among the often applied criteria of selection $CR(\cdot)$ it is necessary, first of all, also to name the minimum of deviation criterion and the balance of forecasts criterion. Information about these criteria, conditions and ways of their applying can be found in [2, 4, 5, 7].

We will mark in the conclusion, that the described approach in one or another way was successfully used both in the tasks investigations of complex ecological processes (forecasting of large reservoirs of freezing, forecasting of processes of territories contamination and other ones) and at solving complex technical problems (forecasting the occurrence of anode effect, etc.).

CONCLUSIONS

In this article the original approach for forecasting of the so-called rare events taking place in the agro-ecological systems is described. Under term "rare" in the agro-ecological (ecological) systems it is necessary to understand the events which take place during some observed process, and time intervals between them are so great, that it is possible to consider that they practically do not influence each other. Two possible approaches to the forming of initial informative base (of data tables) for identification of such processes and phenomena with possibility of forecasting of a rare event beginning moment are described. A multistage procedure of forecasting based on inductive modeling principles as well as a criterion of selection of the best forecasting models is described.

The described approach has a wide range of application in agro-ecological, in technical, medical, biological and many other fields, where it is necessary to have a forecast of not only output value of a process (temperature, for example) but also the knowledge of the peak of rare event in the investigated process.

REFERENCES

1. **Ivakhnenko A.G. and Osypenko V.V. 1984.** Prediction of Rare Events on the Basis of the GMDH Algorithm // Scripta Technica, Inc., U.S.A., Soviet Automatic Control, № 5, vol. 17, pp. 6-10.
2. **Ivakhnenko A.G. 1982.** Induktiwnyj metod samoorganizatsii modelej stozhnykh sistem. Kijev: Nauk. dumka. pp. 296.
3. **Osypenko W. W. 1980.** Prognozirowanije daty zamierzanija oz. Bajkał po garminicheskomu algorytmu MGUA. W kn: Camoorganizatsja kiberneticheskikh sistem. Kiev an Ukrainy. 34-39.
4. **Ivakhnenko A. G. and Stepashkow S. 1985.** Pomekhoustoichiwost modelirowanija – Kiev: Nauka dumka – pp. 216.

5. **Madala H. R. and Ivakhnenko A.G. 1994.** Inductive learning algorithms for complex systems modeling. – New York: Boca Raton, CRC Press – 384.
6. **Stepashko W. S. 2003.** Teoretychni aspekty MGUA jak metody induktiwnogo modelewannia. Uprawlajushekije sistiemy i maszyny. No 2. 31-44.
7. **Battuk V., Basov M. and Dorundyak L.** Matematic model of the process of dust catching inan apparatus with a movable separator ECOTECHMOD an International Quarterly Journal on Economics of Technology and Modelling Processes. Vol. 1. No 1. 2012. 13-16.