ANALYSIS, EVALUATION AND FORECASTING IN ECONOMY

UDC 504.75 + 004.942; 519.25; 621.791: 626/627

D.V. STEFANYSHYN

IMPROVING DIAGNOSTIC MODELS FOR FORECASTING THE BEHAVIOR OF DAMS EQUIPPED WITH AUTOMATED MONITORING SYSTEMS

Abstract. *An approach to forecasting the behaviour of dams according to the data of instrumental observations with regard to capabilities of automated monitoring systems has been proposed. The approach is based on the use of situational and inductive models, where the situational models correspond to selective series of dynamics of observed data within limited time intervals and the inductive models which are constructed on model data derived from situational models simulate evolutions of diagnostic parameters.*
Keywords: *automated monitoring systems, dams, dependent and independent variables, instrumental observations, inductive and situational models, long-term and real-time forecasting, monotonic and non-monotonic data series of dynamics.*

Introduction

The long experience the construction of dams shows that accidents on these structures can lead to serious negative socioeconomic and environmental consequences including those of catastrophic proportions. Therefore, the problems of reliability and safety of dams are given considerable attention. At the international level the main work in this sphere is conducted by the International Commission on Large Dams (ICOLD). One of the most important challenges which are solved by engineers to maintain reliability and safety of dams is the creation of effective systems for monitoring of dams condition. The importance of such systems for ensuring reliability and safety of dams was repeatedly emphasized in the past. In particular, Bulletin 59 ICOLD ("Dam Safety – Guidelines" [1]) says that the majority of damaged dams had no monitoring systems or those systems were imperfect.

The problem with proper functioning of systems for monitoring of dams condition is a complex one and its solution does not only lie in sphere of the introduction of up-to-date equipment of automation and computerization. It should be noted the modern automated monitoring systems (AMS) which are installed at dams are not able to directly perform the functions of ensuring reliability and safety of the engineering facilities during operation. This is due to the fact that monitoring of dams can never be sufficient enough to include all possible influencing factors, important characteristics, parameters, elements and components, the condition of which may affect the overall condition of dams. The most modern types of instrumental control and samples of instrumentation installed at dams allow for doing monitoring of a relatively small number of factors and parameters. As a rule, monitoring is exercised to separate sections, cross-sections, etc. In addition the most advanced AMS is incapable of ensuring the proper modeling of dams' condition yet, which would allow predicting the future behavior of the waterworks. Adequately, they can only perform functions to storage of relevant data and control the state of instrumentation.

In this case, a new task arises which consists in ensuring the processes of modeling and forecasting of dams condition based on observational data under new circumstances when data may be collected in the great amount. Previously, when data were collected manually they were considered to be limited and insufficient to build adequate mathematical models. But without improving of approaches and methods for modeling and forecasting based on observational data, new capabilities of automated monitoring of dams condition are substantially minimized too. Practice shows that large amount of data does not always contribute to the quality of traditional regression models, whose accuracy can degrade. It was found that complex and well structured mathematical models based on observational data in conditions of large arrays of input data do not provide desired results [2]. In particular, the optimization principle, which lies at the basis of construction of traditional mathematical models based on observational data, requires for the systems under study to be in certain boundary limits. If there is a need to have taken into account large amount of data, this principle cannot be easy performed. As a result, challenges associated with the stability of solving optimization problems can even arise in the simplest of cases. Increased quantity of data in case of the traditional approach requires an increase in models dimension by taking into account additional factors and non-linear effects, etc. This leads to disruption the stability of complex models and they can not be used for forecasting purposes.

1. Basic principles of technical diagnostics on dams and principal challenges of regression modeling for forecasting purposes

The basic principles of technical diagnostics and monitoring of condition of technical systems, which were formulated by R.A. Collacott [3], are as follows:

1) Consistency and regularity (continuity) of measurements for characteristics which are selected as diagnostic parameters;

2) Detection of changes in behaviour of these parameters over time;

3) Predicting and forecasting of behaviour of the system which is under monitoring with taking into account these changes.

Automated monitoring systems allow maintaining regular and systematic measurement of diagnostic parameters and storing different data in sufficient quantities to form representative data samples for any situation and any time interval that can be considered in terms of monitoring changes in the environment and in behavior of dams. However, the experience of operating the system at the Kiev dam has showed that implementation of the two following above mentioned principles requires revision of traditional approaches to modeling based on observational data which are accumulating in large volume due to the increased capabilities of AMS. This is because the typical diagnostic models which are used for predicting and forecasting of condition of dams which are in operation are models of regression type.

Regression models determine the dependence of the mean value of some random variable y, which is accepted as a diagnostic parameter or as a dependent (endogenous or resulting) variable, from the other random variable x or the several such variables $x_1, x_2, ..., x_j, ..., x_m$ which are called independent (exogenous, explanatory) variables. The choice of an adequate regression model is based on the minimization of functional which is usually written as the sum of squared deviations $e_i = y_i - \bar{y}$ of the model values $\bar{y} = f(a_0, a_1, ..., x_j)$ of the diagnostic parameter y from observed values y_i, where x_j, $j = \overline{1, m}$, are independent parameters of the model with total number m:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = f\left(a_0, a_1, ...\right) \to \min , \qquad (1)$$

where $a_0, a_1, ...$ are required coefficients of the corresponding regression model. In this case the structure of the model is considered known.

The use of regression analysis in modeling according to empirical data can be right if certain requirements (boundary limits) are fulfilled, in particular [2]:

1) Data of observations for a dependent variable are random values which follow normal distribution;

2) Independent variables $x_1, x_2, ..., x_j, ..., x_m$ are measured with errors which can be neglected compared to the error of dependent variable y;

3) The factors $x_1, x_2, ..., x_j, ..., x_m$ are random variables that are not correlated to each other;

4) Random values $y_1, y_2, ..., y_i, ..., y_n$ of the resulting variable y should be obtained in the same conditions.

The modern regression analysis enables to simplify significantly the task of modeling with using of empirical data for on-line diagnostics of dams condition during their operation. This eliminates the need while modeling the causal relationships between different variables of solving more complex problems of structural and parametric identification of phenomenological models of processes and phenomena that determine behaviour of dams using systems of equations for the theory of elasticity, thermal conductivity, filtration theory, fluid mechanics, etc. with the relevant conditions of uniqueness [4]. This approach to technical diagnostics of dams on the basis of data of instrumental observations is the common one in the international practice. However, if the data of observations are heterogeneous, the construction of adequate regression models for forecasting purposes can be a serious challenge even in simple cases [5].

Searching for unknown coefficients of regression models, according to (1), is carried out so that the model in the statistical sense would better meet to empirical data. That is, to solve the problem (1) the principle of optimization is fulfilled with taking account the compliance with the above boundary limits [2]. But in practice, if the data are heterogeneous, these restrictions can not usually be performed. In this case the increase in the number of observations can disturb the execution of limit restrictions which modeling requires.

For example, Fig. 1 shows a dynamic series of daily observations of upstream water level (UWL) which was taken as an independent variable (a), and the scattering field of values of water level in a piezometer (PWL) (b) which was considered as a dependent variable on random values of UWL. As we can see there is a strong non-monotonicity of UWL values and there is a significant heterogeneity of PWL values depending on UWL.

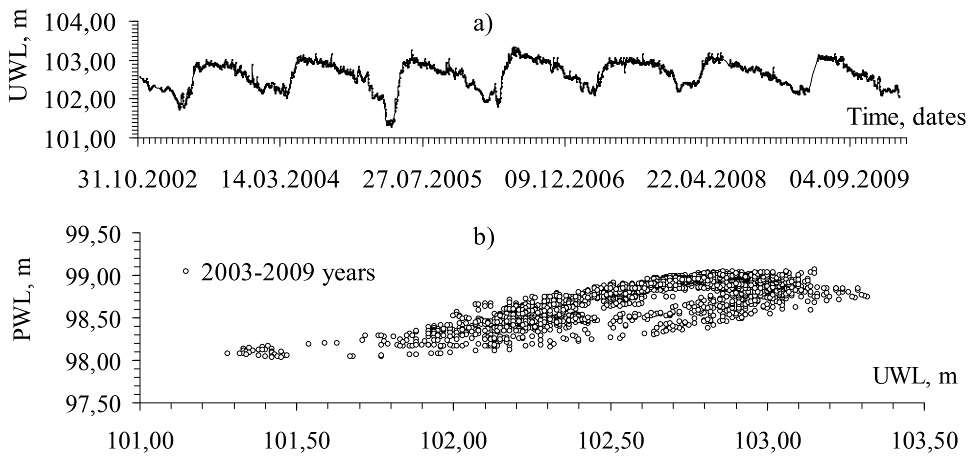

Fig. 1 – Illustration of non-monotonicity in observations of upstream water levels (UWL) (a) and heterogeneity of water levels in a piezometer (PWL) depending on UWL (b)

Increasing the structure dimension of a regression model by introducing into the model of additional independent variables cannot usually solve the problem of heterogeneity of variance. The presence of correlation between different independent variables in multivariate models (we know it as the multicollinearity problem) may become an additional challenge. And we know that under multicollinearity conditions the regression coefficients become highly unstable to small changes in the data, which violates the stability of solutions in the search for the unknown coefficients of regression models. Constructing such models like autoregressive models, distributed-lag models, etc., does not always bring success too.

2. A concept of situational regression models as the main idea of the new approach to regression modeling for forecasting purposes

In short time intervals compared to the total duration of instrumental observations it is easier to provide the monotony of observations series for variables of regression models and the homogeneity of samples of data and the independence of endogenous variables from the less significant factors [2, 4, 6]. It should be noted the main idea of regression analysis is that a suitable regression may take place if a dependent variable y depends not only on variables $x_1,...,\ x_j,...,x_m$, and the variable y may depend on uncontrolled, unknown factors which form something like a forecast background [7].

It can also be assumed, if in different periods of time these forecast backgrounds are relatively homogeneous and the corresponding series of dynamics of independent variables are monotonic (Fig. 2a), adequate regression models (Fig. 2b) may be constructed. Henceforth, we will call these suitable models as situational models. The situational models can be relatively simple. These can be single-factor models [7], which show how an endogenous variable y depends on one the most important exogenous variable x.

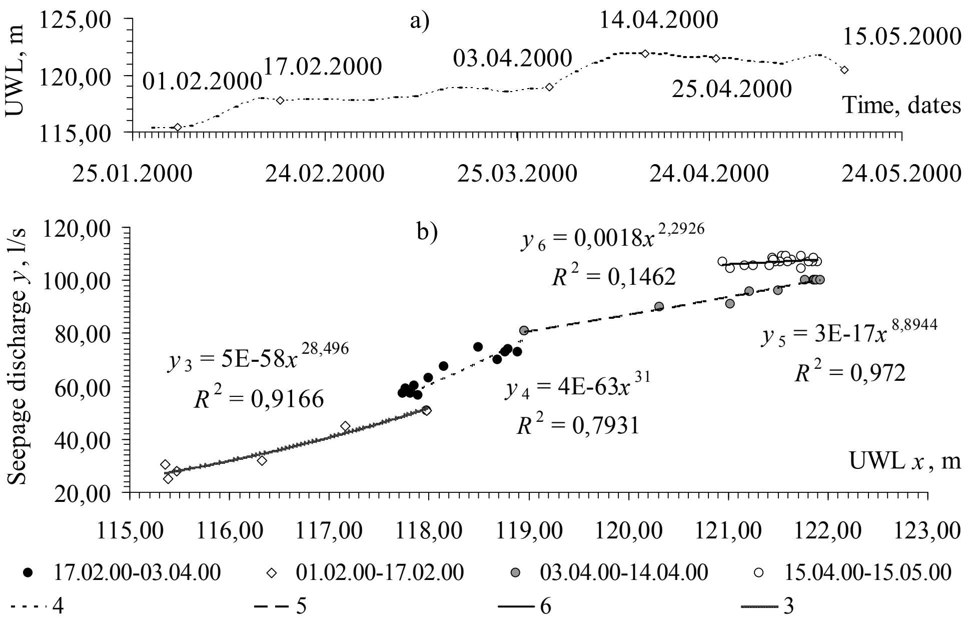

Fig. 2 – Data of observations of upstream water level (UWL) (a) and situational regression models for seepage discharge values through the dam foundation depending on UWL (b)

We suppose that upstream water level (UWL) for dams could be the principal independent variable x for situational modeling. This is the most suitable and convenient independent variable and the only independent variable which can be controlled if it is necessary.

We should emphasized that the main thing there is that situational models must adapt to particular situations (forecast backgrounds, etc.) that take place within limited time periods. It is very important they were the most adequate models to these situations.

In fact, the corresponding situational models, which are based on limited data, reflect different phase states of the dam as a dynamical system on respective time intervals. In the simulation we can get sets of adequate situational models that appropriately evolve over time (Fig. 2b). Although the transitions between the nearest situational models which define adjacent phase states of the dam as a dynamic system can be non-monotonic [4, 7], the prediction of future condition of the dam can be based on monitoring the evolutions of these situational models. This is the main idea of such simulation to obtain situational models.

3. Inductive models and forecasting future states of dams

Inductive diagnostic models are models obtained on base of generalization of results of construction of situational diagnostic models (Fig. 3a). In the most general case, inductive models are models of “levels” (Fig. 3b). These models, which are constructed with using of results of situational modeling at separate time intervals, can spread on the entire period of observations.

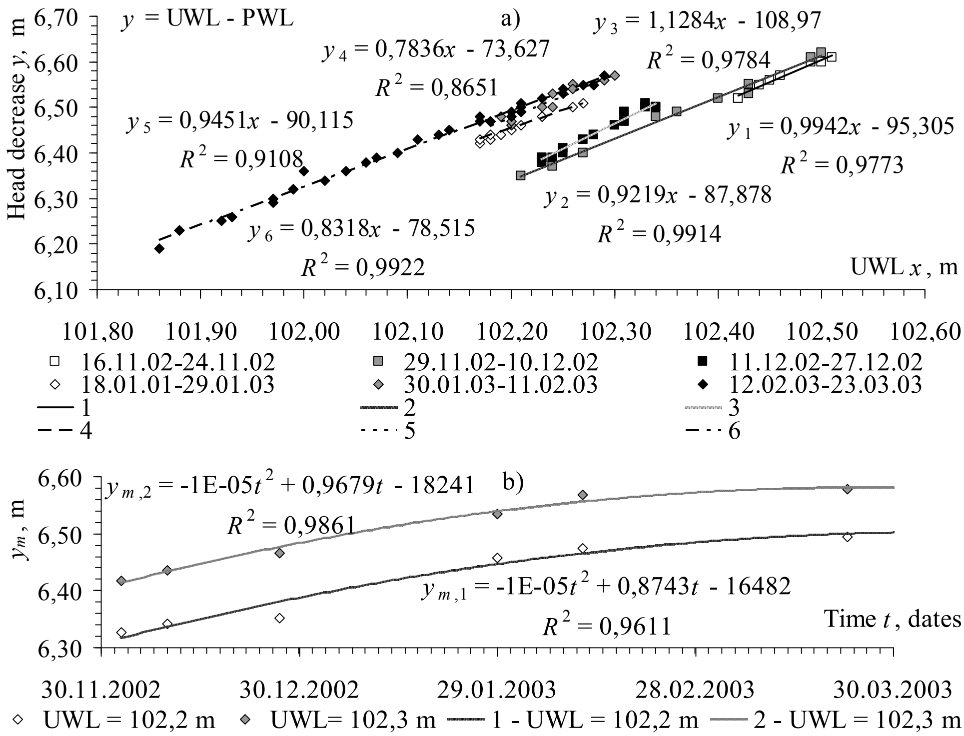

Fig. 3 – An example of a set of corresponding situational models (a) and two inductive models of “levels” (b) for dependencies of head decrease values on upstream water level (UWL) at a site of seepage between upstream and a piezometer

The structure of an inductive model is determined by particularities of behaviour of time series of simulated data obtained from corresponding situational models. In general, results of situational modeling may represent non-stationary (Fig. 3b) or stationary (quasi-steady) (See below Fig. 4a) time series of modeling data with presence or absence of trends respectively.

If trends have high coefficients of determination, inductive models can be described by these trends (Fig. 3b). Then general inductive models will consist of corresponding functions which show trends and random “balances” after the extraction of these trends. If results of situational modeling give stationary (quasi-steady) time series (there are no trends) (Fig. 4a), inductive models can be represented as regressions (Fig. 4b). In these cases general inductive models will consist of corresponding regressions and random “balances”. If trends in dynamic series of results of situational modeling of variable y have small coefficients of determination, an inductive model of i-level for y can be presented as composition of a selected trend $T\left(y_i(t)\right)$ (Fig. 5a) and a regression R for “balances” $\Delta y_i = y_i - T\left(y_i(t)\right)$ which are random values of the dependent variable (Fig. 5b):

$$y_i(t) = T\left(y_i(t)\right) + R\left(\Delta y_i(\hat{x}_L)\right). \tag{2}$$

If it is necessary, we can use a new explanatory variable $\hat{x}_L$ for modeling the regression of “balances” Δy_i. We may take into account a transport lag between the “balances” and the variable $\hat{x}_L$ too. In more complex cases if we need to take into account autocorrelation of the “balances”, in addition some cyclical components or deterministic components of corresponding series can be considered in inductive models.

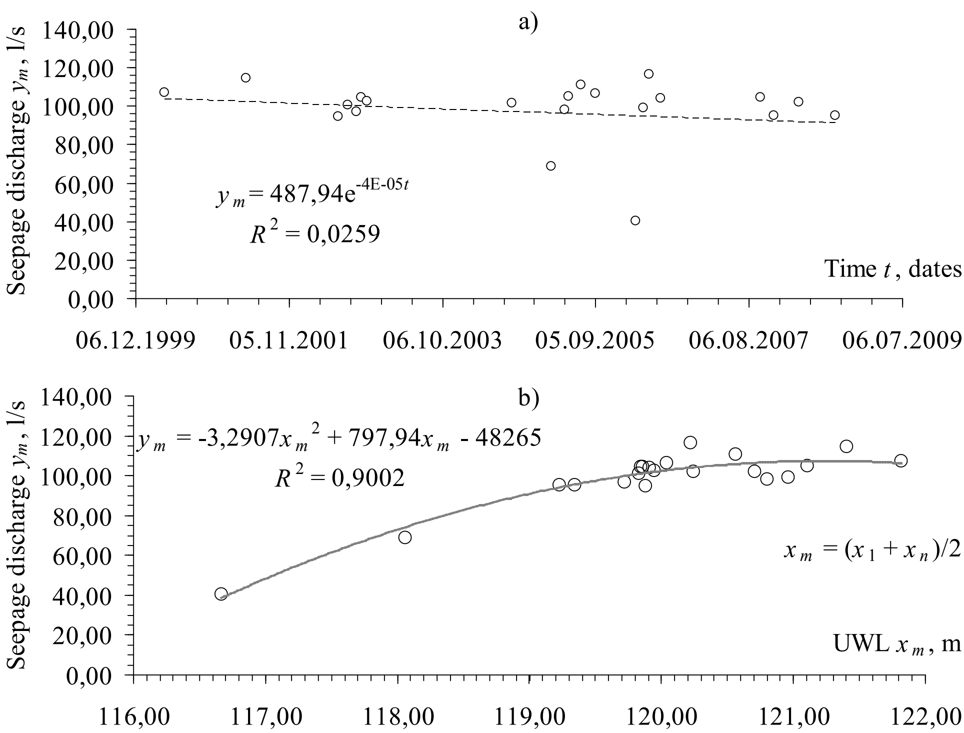

Fig. 4 – Stationary series of dynamics (a) of results situational modeling of seepage discharge values in conditions of stationary oscillations of upstream water level (UWL) and a corresponding inductive model in form of the regression (b)

Forecasting future states and behavior of dams is based on the method of extrapolation and is carried out in two main forms:

1) Real-time forecasts;

2) Long-term forecasts.

Real-time forecasts are made for the purpose of rapid assessment dams' condition in changed situations (See below Fig. 6) and performed by means of new situational models which require adjustments to previous situational models due to new data with extrapolation into region of expected values of independent variables. If new data comes, real-time forecasts may be constantly corrected. Observed values which differ significantly than situational models show can indicate changes of forecast backgrounds (Fig. 6).

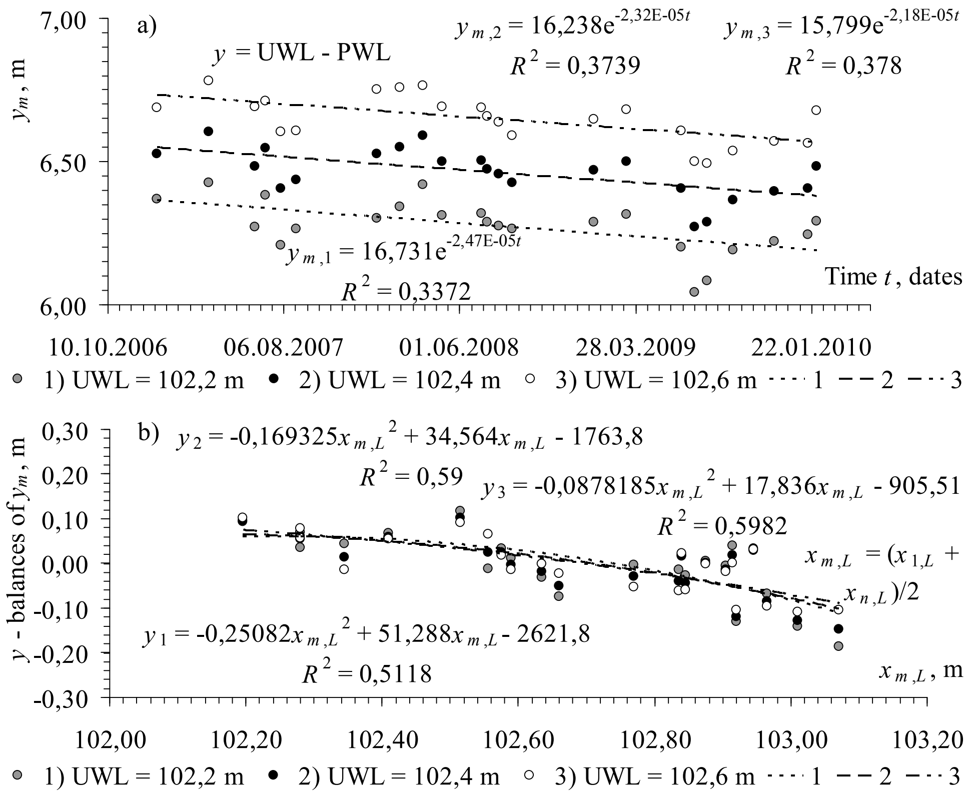

Fig. 5 – An example of non-stationary series of dynamics of results of situational modeling of head decrease values with relatively small coefficients of determination of trends

Long-term forecasts (Fig. 7) are usually based on inductive models which include trends and regressions for their balances (Fig. 5) but simple regression models (Fig. 4) may be used too. The expected results of the long-term forecasting are determination a new situational diagnostic model which corresponds to expected situation in the nearest future period (Fig. 7a) or series (variety of options) of situational diagnostic models which can correspond to several possible situations in the future period (Fig. 7b).

The accuracy of long-term forecasts which are made on the basis of inductive models can be greatly improved if the inductive models are based on results of situational modeling of past periods which are presented by homogeneous and interrelated clusters of the relevant data with taking into account behavior of independent variables and transport lags.

First, we should pay attention to behavior of upstream water level (UWL) (See below Fig. 8). In particularly, some following typical modes of behaviour of upstream water level (UWL) affecting dams can be allocated:

1) Slow rise of UWL;
2) Rapid rise of UWL;
3) Stationary fluctuations of UWL;
4) Slow lowering of UWL;
5) Rapid lowering of UWL.

If some transport lags exist between corresponding data of previous and next periods, forecasts can be unambiguous (Fig. 7a). Otherwise, we get several possible long-term forecasts concerning future periods (Fig. 7b).

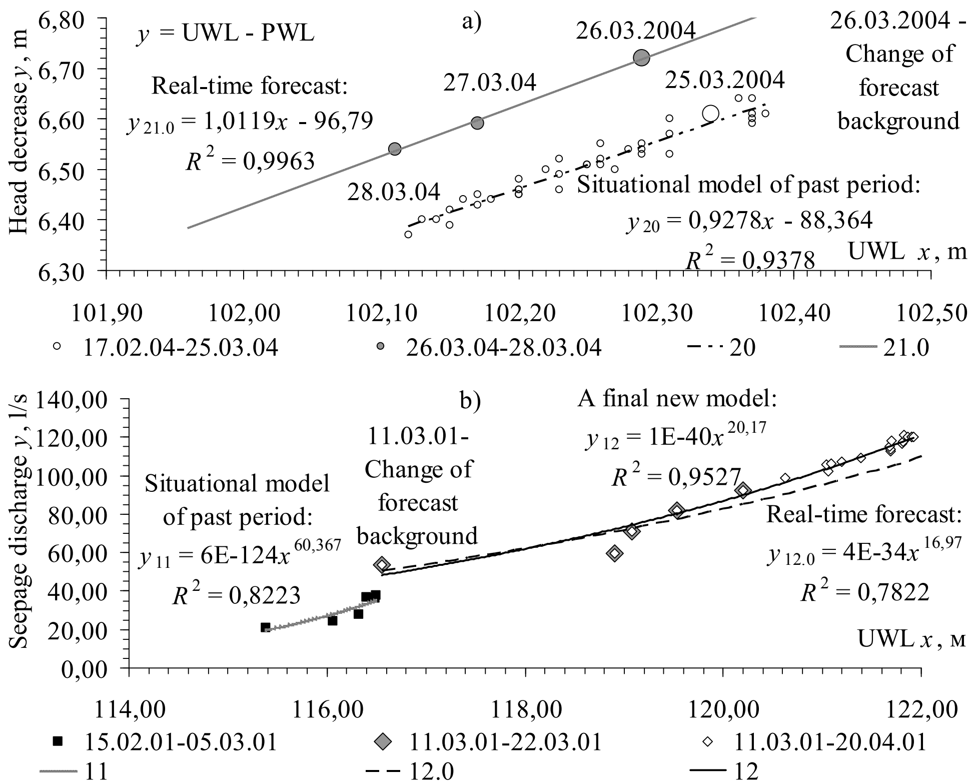

Fig. 6 – Examples of real-time forecasts: a) how head decrease values at a site of seepage between upstream and a piezometer can depend on upstream water level (UWL); b) how seepage discharge values can depend on upstream water level (UWL)

A general diagnostic model of an appropriate diagnostic parameter of a dam can be presented as a family of situational diagnostic models which are adapted to individual time intervals or as a family of inductive diagnostic models which allow producing situational diagnostic models for periods in the future. So, forecasts are performed on the basis of situational diagnostic models and by means of monitoring for evolutions of these models in time.

It should be noticed the simple mathematical models (trends and regressions) may be used as situational and inductive diagnostic models for dams where automated monitoring systems are installed. Such models and combinations thereof are the best to adapt to new data and changes of forecast backgrounds. When choosing the diagnostic models of a dam, it is also allowed using of modified diagnostic parameters and different data processing procedures which are aimed at the construction of adequate situational diagnostic models for forecasting of dams condition to assess their reliability according to the data of instrumental observations.

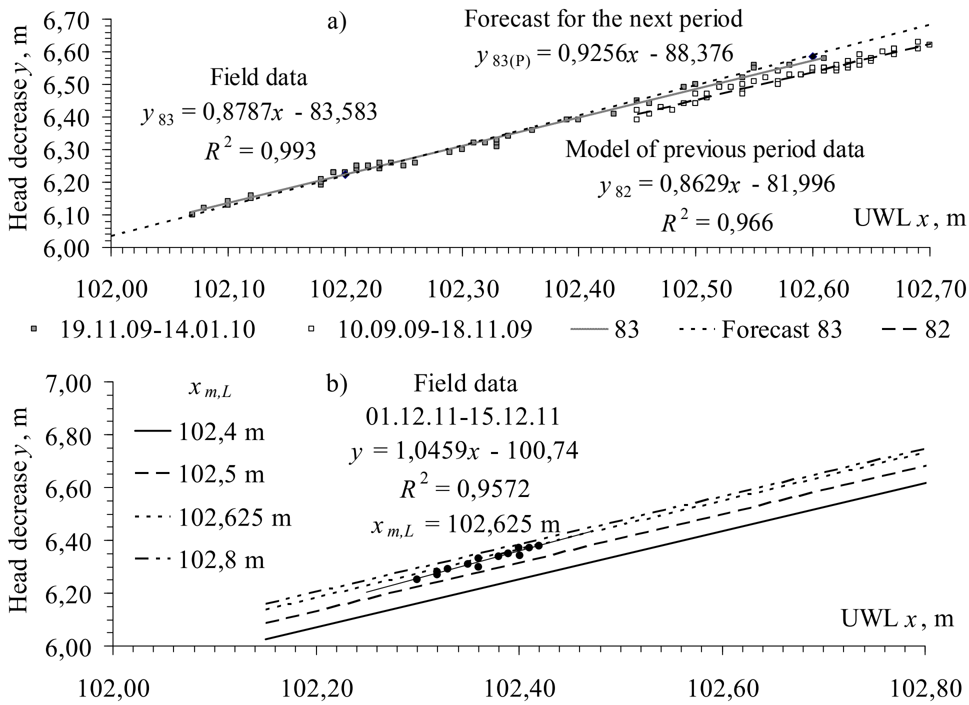

Fig. 7 – Examples of long-term forecasts how head decrease values at a site of seepage between upstream and a piezometer depend on upstream water level (UWL): a) an unambiguous forecast; b) some forecast options

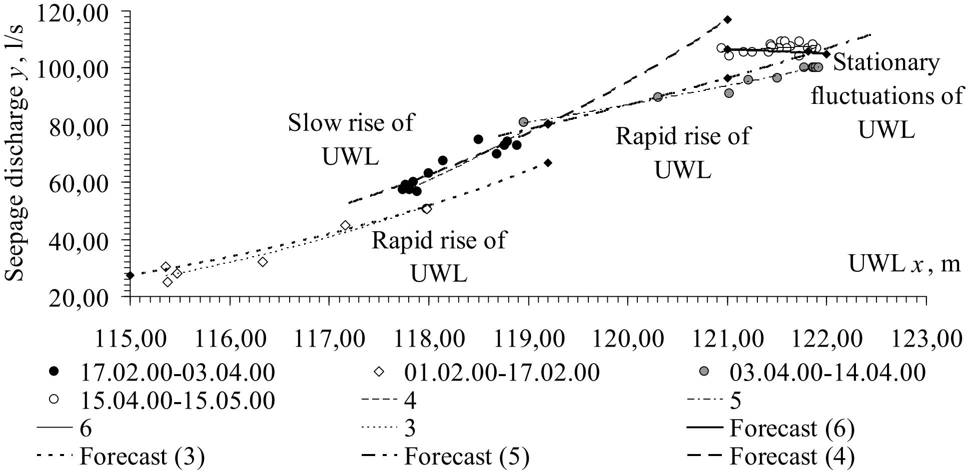

Fig. 8 – Examples of long-term retrospective forecasts how seepage discharge values depend on upstream water level (UWL)

Conclusions

1. A new approach to forecast condition and behavior of dams according to data of instrumental observations with regard to capabilities of automated monitoring systems installed on hydraulic structures has been proposed. The approach is based on the use of situational and inductive models of regression type where situational models correspond to selective series of dynamics of observed data within limited time intervals and inductive models which are constructed with model data derived from situational regression models enable to take into account evolutions of diagnostic parameters in time.

2. The simple mathematical models (trends and regressions) may be used as diagnostic models of dams if these models are easy adapted to new data and changes of corresponding forecast backgrounds. The accuracy of long-term forecasts which are made on the basis of inductive models can be greatly improved if the inductive models are based on results of situational modeling of past periods which are presented by homogeneous and interrelated clusters of the relevant data with taking into account behaviour of independent variables and transport lags.

REFERENCES

1. Dam safety – Guidelines. ICOLD, Bulletin No. 59, 1989.
2. Kuhn M. Applied Predictive Modeling / M. Kuhn, K. Johnson. – New York: Springer Science+Business Media, 2013. – 600 p.
3. Collacott R.A. Structural Integrity Monitoring / R.A. Collacott. London – New York; Chapman and Hall, 1985. 455 p.
4. Stefanyshyn D.V. Prediction of indexes of dynamic system with use of observational data as time series / D.V. Stefanyshyn // Problems of decision making under uncertainties (PDMU-2013). Abstracts of XXII International Conference. September 23-27. Foros-Yalta, Ukraine. Kyiv, 2013. – P. 31-32.
5. In-Soo Jung. Interpreting the Dynamics of Embankment Dams through a Time-Series Analysis of Piezometer Data Using a Non-Parametric Spectral Estimation Method / In-Soo

Jung, M. Berges, J.H. Garrett Jr, Ch.J. Kelly // Computing in Civil Engineering, 2013. – P.P. 25-32.
6. Hamilton J.D. Time series analysis / J.D. Hamilton. Princeton University Pr. Princeton. – N. J.: 1994. – 782 p.
7. Stefanyshyn D.V. A Method of Forecasting of Indices of Dynamic System that evolves slowly based on Time Series Analysis / D.V. Stefanyshyn // ICIM 2013. Proc. of 4th Int. Conf. on Inductive Modelling; Kyiv, Ukraine, September 16-20, 2013. – pp. 221-224.

Been received for revision 26.05.2017.