UDC 004.942
https://orcid.org/0000-0002-7620-1613

D.V. STEFANYSHYN

AN APPROACH TO PREDICTING BASED ON MONITORING DATA BY MEANS OF COMBINED SITUATIONAL-INDUCTIVE MODELING (THE MAIN IDEA AND EXPECTED RESULTS)

Abstract. *This approach is proposed to use to predict the behavior of complicated technogenous, ecological and economic systems based on regular monitoring data forming time series by means of simplified models of regression type. The approach uses the idea of decomposition of complex modeling and prediction tasks by means of regression models based on monitoring data to overcome the excessive structural and parametric uncertainty of real dynamic systems. According to the approach proposed, the prediction based on monitoring data by means of combined situational-inductive modeling consists of establishing relevant situational models of regression type being adequate in the future within certain limited time intervals.*
Keywords: *combined situational-inductive modeling, inductive and situational models, monitoring data, prediction, predictive background, time series.*

DOI: 10.35350/2409-8876-2019-17-4-75-87

Introduction

The main predictive models based on monitoring data and used for predicting the behavior of technogenous, ecological and economic systems are the regression ones. There are many modern methods based on regression models that are quite capable of modeling complex system relationships [1-9].

The use of regression models enables to simplify significantly problems of modeling and predicting the behavior of complicated systems based on experimental or observational data giving the possibility of making decisions in various challenging situations avoiding the use of much more complex system models – both deterministic and stochastic.

However, regression models can be the most successful in the case of data interpretation. In prediction, various regression models, including quite complex regressions, can get lost their accuracy and attractiveness. Often, regression models easily overemphasize patterns that are not reproducible and demonstrate the instability of extrapolation in the prediction zone (See an illustrative example in Fig. 1 below). So a researcher cannot know about the fatal prediction faults until the next set of samples appears [1].

Nevertheless, admittedly, the regression models are the most convenient models to solve practical prediction problems in various fields [1, 2]. They can easily be formalized and adapted to experimental data, including monitoring data presented in the form of time series [3, 4]. Also, if necessary, regression models may be modified in the appropriate way depending on the peculiarities and complexities of prediction problems [1, 2].

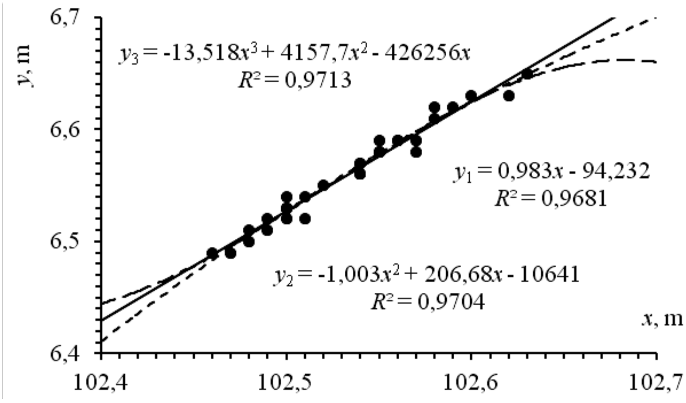

Figure 1 – An example of the instability of extrapolation by means of various regression models (y is the reaction of piezometer; x is upstream water level; the models were built according to data given in [5, 6])

This article presents some basic generalizations concerning the proposed approach to predicting based on monitoring data by means of combined situational-inductive modeling. They emerge from the results of our previous research that we performed when solving the problems of predicting based on monitoring data of the behavior of complex systems and processes having different natures and conditions to carry out monitoring [5-9].

1. Some common remarks concerning to situational and inductive modeling

1.1. Fundamentals of situational modeling

Situational modeling is usually undertaken to understanding and restoration of certain situations (the coincidence of the conditions and circumstances of the functionality) relating to the behavior of complex systems [5-10]. Nowadays, situational modeling has become especially popular in economics, medicine, military affairs, forensics, politics, and other similar spheres, as well as in artificial intelligence, where the development of a logical approach to modeling the behavior of complex systems and processes gave impetus to the creation of situational calculus theory [10-12].

The basic idea of situational modeling is that a complete description of the infinite set of all possible situations of the functioning of a real system is replaced by a certain finite set of generalized situations that reproduce to a certain degree the system's possible states [5-10]. The evolution of the dynamic system is modeled in the context of its "movement" along a series of situations that are the result of various actions. These different model situations (by R. Reiter [11]) do not determine literally appropriate states of the system; they show only the history of certain events as completed sequences of actions in certain periods of time.

Since situations cannot be described totally, and it is possible to speak only about some of their aspects, the non-monotonic output rule is used to describe the evolution of the system. In modeling, it is assumed (by J. McCarthy [12]) that on the basis of past facts, by which past situations are described, and on using some general rules or assumptions, according to which execution of actions and occurrence of events within situations take place, it is possible to predict some similar situations that will appear in the future.

1.2. Inductive modeling

Admittedly, inductive modeling is known as an original scientific approach founded by the Ukrainian scientist O.G. Ivakhnenko, boosted by his numerous doctoral students and followers. This approach found their theoretical and practical reflection in the widely famous method to be named the Group Method of Data Handling (GMDH) [13, 14].

It is now used to solve different problems the pattern recognition, the structural-parametric identification of models, the simulation and forecasting of complex processes and systems based on experimental data. According to this approach, on the basis of available observational data, a hypothesis about a possible class of models is put forward, the procedure of automatic generation of a set (it may consist of thousands and tens of thousands) of alternative models belonging to this class is formed, and the criterion of choosing the best model is established. It is assumed, since most routine work is transferred to a computer, the impact of human mistakes on the final result of modeling is minimized.

Nowadays, the GMDH method is also considered as one of the most appropriate and advanced information technologies to obtain knowledge from experimental data, or as one of the most effective methods of intellectual (or intelligent) data analysis [15]. However, the main theoretical and practical result of this approach to modeling based on experimental data is that the complexity of optimal predictive model depends on the level of uncertainty in the data: the higher this level is (e.g. due to noise or their abundance), the simpler must be the optimal model (with less number of predictors). Otherwise, the quite successful model for data interpolating can get lost its accuracy and attractiveness in predicting.

2. More about the challenges of prediction based on regression models

Practice shows [1, 2, 5-9], the construction of adequate regression models for predictive purposes, especially in case of extrapolation, can be an informal and challenging problem even in simple cases, as it can be seen for the example shown in Fig 1. So, the first challenge of prediction based on regression models is that, often, better interpretation models may have a tendency to overpredict (or underpredict) low values and underpredict (or overpredict) high ones [1].

However, problems related to the construction of reliable regression models can arise in the case of the prediction within the so-called "normal" values of observed data too [5-9].

Perhaps, the main predictive problem connected with regression modeling is that regression models are traditionally built as models that should suit the best way to all collected data following the principle of optimization. As a result, the increase in the number of observation data may complicate the execution of important limit restrictions of regression modeling [5-9].

Admittedly, complicating the structures of regression models by taking into account additional factors, parameters, and nonlinear effects, etc., the quality of regression models as interpretative models for observed data can be improved. But their quality as predictive models may be deteriorated. If there is more than one predictor, a researcher will have to further understand the characteristics of different predictors and the relationships among them. For example, between-predictor correlations (multicollinearity) may become a quite serious problem. It is well known, the simplest way to solve the multicollinearity problem is to remove the predictors that have the most correlated relationships. Usually, if predictors are highly intercorrelated, this implies that they are measuring the same underlying information. Removing some of them might lead to a more parsimonious and interpretable model without compromising the performance of the model [1, 2]. Undoubtedly, in some cases, the regression model simplification may be regarded as a practical way to improve it.

There are a lot of potential advantages of removing some predictors in regression modeling. Firstly, fewer predictors will decrease computational time and complexity. Secondly, some regression models can largely be improved in performance and (or) their stability will significantly increase without the problematic variables. But all these simplifications should be justified in terms of predicting. It should also be mentioned that the main idea of regression modeling, as it is well known, is that regression occurs when a dependent, endogenous variable depends not only on some independent, exogenous, explanatory variables (predictors) but also on some uncontrolled unknown factors. So, the choice of the structure and appropriate variables of regression models, both dependent and independent may be an informal task that requires prior research [5-9].

Essentially, the uncontrollable factors affect the structure and parameters of a regression model. So, they can be considered as certain circumstances and conditions in which a system under modeling exists in a certain period of time. In [5-9], we determined these circumstances and conditions as a peculiar predictive background. In other words, a predictive background can be considered as a set of external and (or) internal conditions including unknown predictors stipulating the structure and parameters of an accepted regression model.

The concept of predictive backgrounds is the best fit for the determination of situations in which a system is at particular points in time. In fact, the constancy of a predictive background relates to a limited period of time and can identify the only specific situation for which a relevant regression model can be built to be considered adequate (See an illustrative example in Fig. 2 below). Accordingly, various regression models corresponding to various predictive backgrounds or, in other words, to various situations describing states of a complex system in time may be adopted as situational models.

As can be seen in Fig. 2, situational models may be presented by the simplest single-factor regression models. But such significant simplification is not the only advantage of situational modeling. In general, changes of predictive backgrounds can be both regular and irregular, both stochastic (probabilistic) and principally unpredictable, uncertain or ambiguous. Some of these changes are actually impossible to forecast and estimate. Eventually, regular changes can be taken into account successfully when introducing some additional variables into regression models, whereas for irregular changes it is difficult or practically impossible to do this. Situational modeling not only enables avoiding these complexities but also to identify some peculiar situations being similar to “black swans” by Nassim Taleb [16].

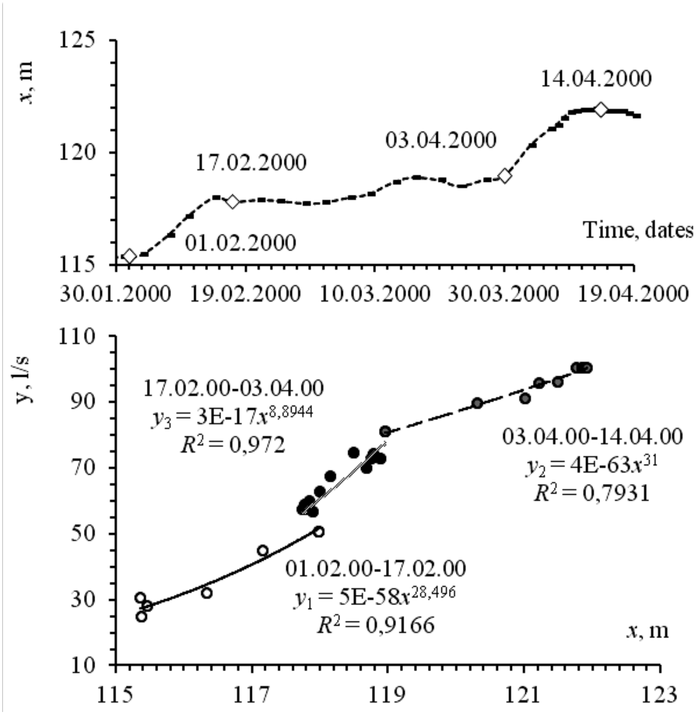

Figure 2 – Situational regression models corresponding to different predictive backgrounds (y is seepage discharges; x is upstream water levels; the models were built according to data given in [6, 8])

It is reasonable to assume that the smaller the duration of a model situation in time is, the more likely the constancy of the relevant predictive background can be expected. In Fig. 3 (See below, as illustrations), shown are four different models built for various durations and for various time intervals among which there are two “black swans” (situational models “1” and “3”).

From Fig. 3 there can be easily convinced that the regression model “0” based on all available data is the least reliable for forecasting purposes. Narrowing down the interval of retrospective modeling to the period of 1973-1998 years allowed obtaining a more reliable regression model (the model “2”) to predict. Finally, more detailed data processing revealed the possibility of peculiar “emissions” in the form of the situational regression models “1”, “3”, which took place in the past. So, it may be assumed the similar risky situations of the kind “black swans” events should be expected in the future too.

To sum up, let us emphasize the main definitions and assumptions. The definition “situational model” relates to a model adapted to a separate situation. Adequate situational models can be presented as single-factor regression models. The circumstances and conditions of a situational model adaptation to an appropriate situation are a predictive background. Situational models can reflect various phase states of a dynamic system at different time intervals.

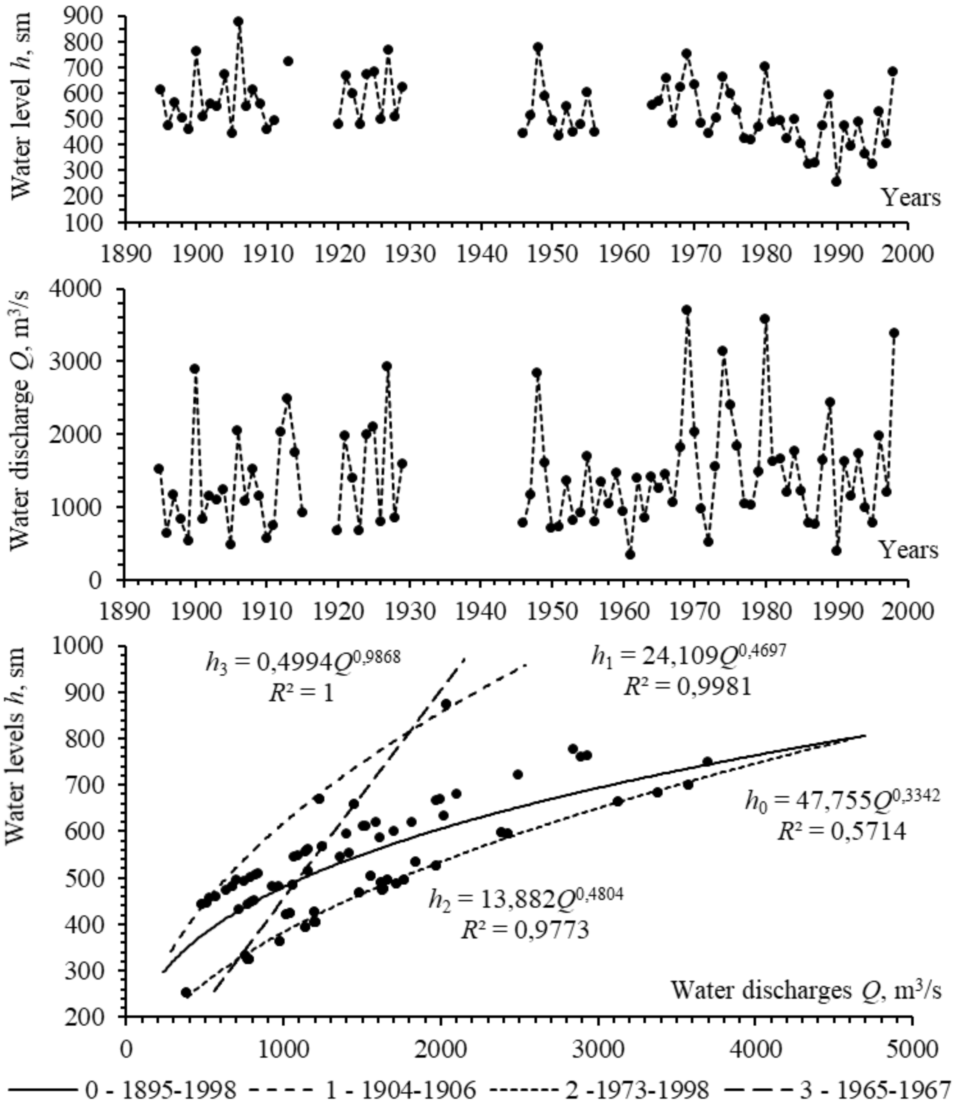

Figure 3 – Several regression models depending on data relating to different time intervals (the models were built according to data given in [9])

3. Presentation of the approach to combined situational-inductive modeling

The common block-scheme presenting the main stages and results of modeling and explaining the proposed approach to prediction based on monitoring data by means of combined situational-inductive modeling is shown in Fig. 4 below.

Preliminary modeling (pre-modeling) aims at selecting the type and variables of a regression model that is used further as a situational model, and decomposing the data obtained over the entire monitoring period into individual situations, resulting in a set of separate time series [5-9]. Different methods of time series data analysis can be used for this purpose [1-4, 15]. Correlation and regression analysis of monitoring data, lags analysis, etc. are performed. For example, when constructing the situational models shown in Fig. 2, the lag equal to one day between y and x was taken into account.

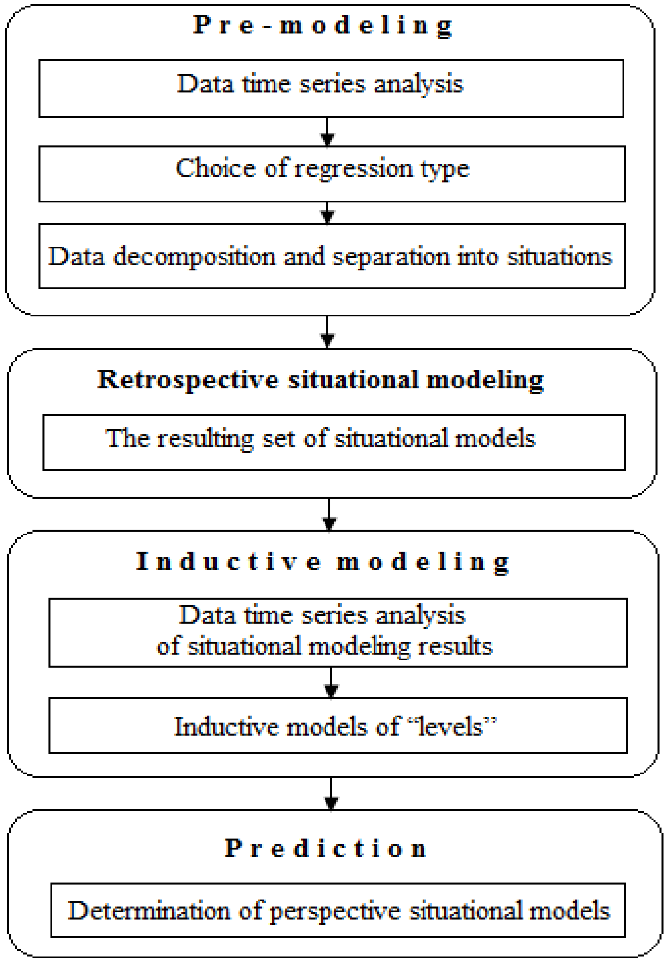

Figure 4 – The common block-scheme of the proposed approach to prediction based on monitoring data by means of combined situational-inductive modeling

To perform retrospective situational modeling, sample time series describing the behavior of dependent and independent model variables within separate time intervals are prepared. These separate samples (or clusters) have to meet certain established criteria of situational modeling adequacy that are formulated taking into account the behavior of variables of situational models relating to corresponding situations. In particular, the following common criteria of situational modeling adequacy concerning the behavior of independent variables may be considered.

– Non-stationary fluctuations of values of time series with monotonically increasing trends. Some separate time intervals characterized by relatively slow or relatively rapid monotonous trend growths of values may also be allocated.

– Non-stationary oscillations of values with monotonically increasing trends.

– Non-stationary fluctuations of values of time series with monotonically decreasing trends. Some separate time intervals characterized by relatively slow or relatively rapid monotonous trend declines of values may also be allocated.

– Non-stationary oscillations of values with monotonically decreasing trends.

– There are no trends; random stationary variations of values take place.

The result of retrospective situational modeling is a set of situational models (See, for example, a set of simple linear regressions shown in Fig. 5). It forms the basis for subsequent inductive modeling. As can be seen in the examples (Fig. 2, 3, 5), the transition from one situational model to other proceeds non-monotonically. Therefore, we assume that relevant adjacent phase states of the dynamic systems change non-monotonically too.

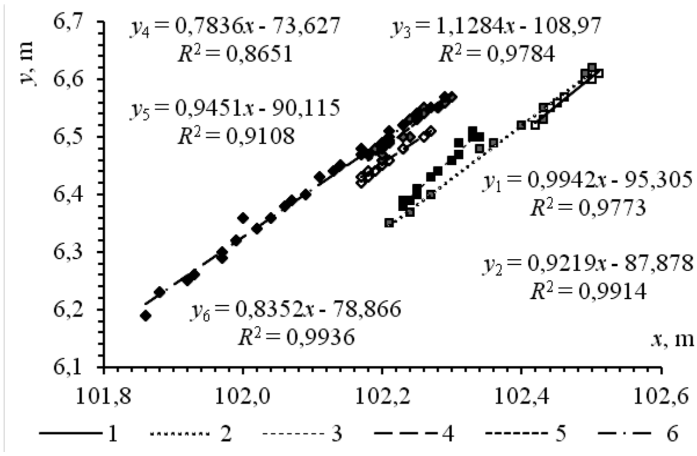

Figure 5 – Results of retrospective situational modeling
(y is the reaction of piezometer; x is upstream water level;
the situational models were built according to data given in [5, 6, 8])

In our study, the definition "inductive model" is related to a model obtained from the generalization based on a certain set situational (or, in more complex cases, inductive) models. The inductive models, in our interpretation, are models of "levels", which determine the behavior of dependent variables for some fixed values of predictors according to situations.

Overall, inductive models may have various structures (compositions). They can be presented in the form of a set of trends (Fig. 6) if the time factor is essential or taken into account, or in the form of a set of regression models (Fig. 7), if the time factor is not taken into account or it is unessential. More general inductive models may consist of trends and random "balances" after the extraction of these trends (See an illustrative example of extraction of trends while inductive modeling in Fig. 8 below), trends and regression models for random "balances" [6, 8] and so on. Inductive models may be modified in an appropriate way if new data and tendencies appear. As well as, if necessary, time or transportation lags between model variables can be taken into account. Additionally, some techniques of adaptive modeling can be applied during the construction of both inductive and situational models [1, 2, 6-9, 15].

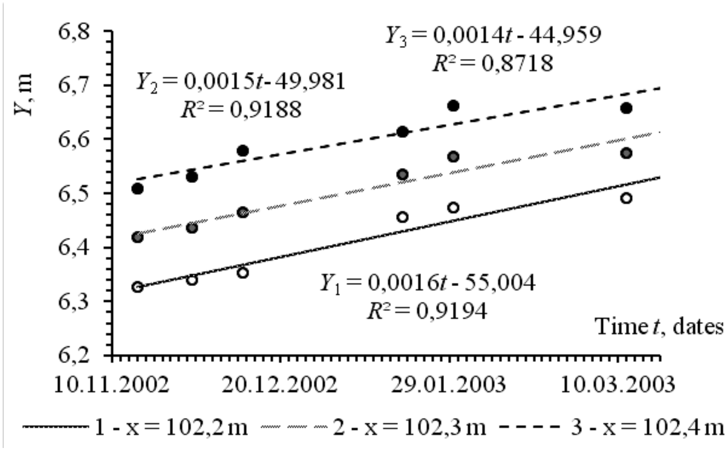

Figure 6 – Results of inductive modeling with prediction in the form of a set of trends (according to results of situational modeling, shown in Fig. 5)

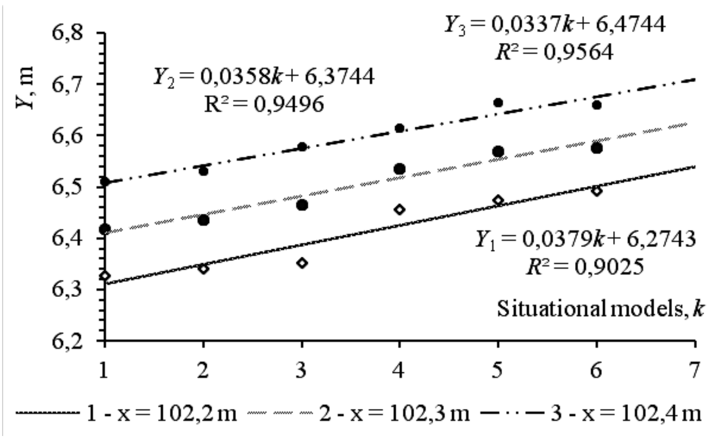

Figure 7 – Results of inductive modeling with prediction in the form of a set of regression models (according to results of situational modeling, shown in Fig. 5)

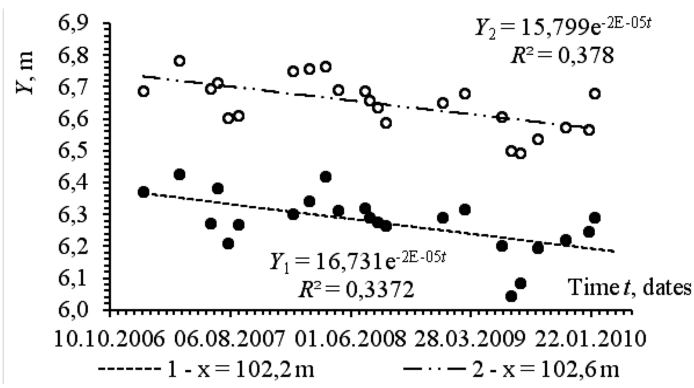

Figure 8 – An example of the extraction of trends while inductive modeling [6, 8]

Inductive models create the basis for predicting perspective situational models. In fig. 9, 10 are presented two examples of such predicting with verification obtained results on the base of observational data.

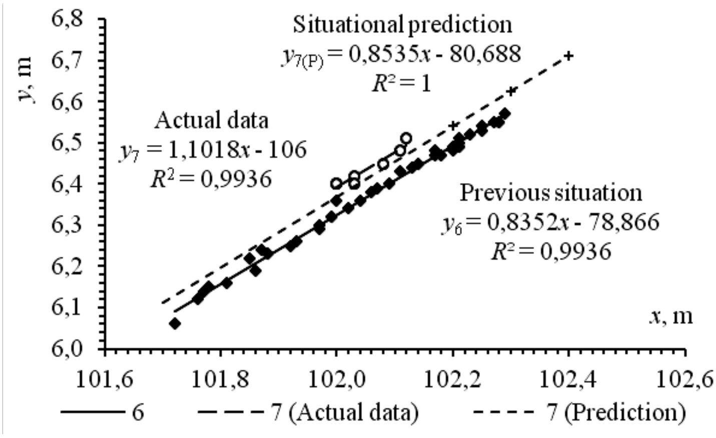

Figure 9 – Predicting of the perspective situational model 7 (data from Fig. 5, 6)

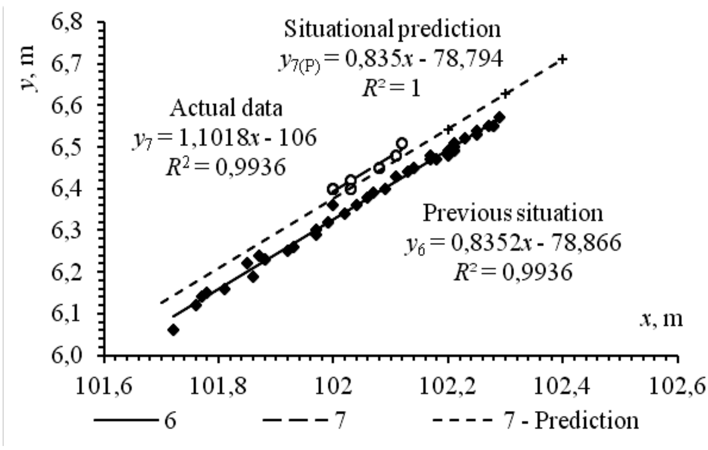

Figure 10 – Predicting of the perspective situational model 7 (data from Fig. 5, 7)

4. Some pre-conclusive remarks

Actually, the task of situational predicting, in our case, is an extrapolation task defined as the identification of the most probable situational model that will meet some expected situation in the future depending on situations that appeared in the past. The results of our studies (See also [6-8]) show that depending on the characteristics of the system under study, its interaction with the environment, the completeness of the monitoring data, and other factors that may determine the behavior of the dependent and independent variables of situational and inductive models, a predicted situational model may be unambiguous or the result of situational predicting will be a certain set of perspective situational models corresponding to various expected situations in the future.

It should be also noted the next. If inductive models are built on the basis of situational models of past periods that cover data of similar clusters of actual data (for example, taking into account the seasonal factor, etc.), the accuracy of the situational prediction on the basis of such inductive models may increase essentially [6, 8].

As well, in order to implement the proposed approach to predicting successfully, it is important to ensure to perform three basic monitoring principles formulated by R.A. Collacott [17]: 1) Consistency and regularity (continuity) of measurements for parameters and characteristics selected for the control; 2) Detection of changes in the behavior of these parameters and characteristics over time; 3) Prediction of future situations taking into account these changes.

Conclusions

Proposed is an approach to predict the behavior of complicated systems of the different origins, based on regular monitoring data forming time series, by means of simplified models of regression type. The approach uses the idea of decomposition of complex modeling and prediction tasks by means of regression models based on monitoring data to overcome excessive structural and parametric uncertainty of dynamic systems.

According to the approach proposed, the prediction based on monitoring data by means of combined situational-inductive modeling consists of establishing relevant situational models of regression type being adequate in the future within certain limited time intervals. The approach allows using simultaneously both the principle of optimization in modeling and the principle of adaptation to situational changes occurring in dynamic systems. Practical results indicate the opportunity to carry out such situational predicting with acceptable accuracy.

REFERENCES

1. Kuhn M., Johnson K. Applied Predictive Modeling. Springer Science+Business Media: New York, 2013. 600 p.
2. Geisser S. Predictive Inference: An Introduction. New York: Chapman & Hall, 2016. 264 p.
3. Brockwell P.J., Davis R.A. Time series: Theory and methods, second ed., Springer, 1991. 557 p.
4. Hamilton J. D. Time series analysis. Princeton University Press: Princeton, New Jersey, 1994. 782 p.
5. Stefanyshyn D.V. A method of forecasting of indexes of dynamic system that evolves slowly, based on time series analysis. ICIM 2013, Proc. of 4th Int. Conf. on Inductive Modelling, Kyiv, Ukraine, September 16-20, 2013. P. 221-224.
6. Stefanyshyn D.V. Situational and inductive modelling in problems of forecasting behavior of dams equipped with modern automated monitoring systems. Zeszyty Naukowe Inżynieria Lądowa i Wodna w Kształtowaniu Środowiska, Nr 14, Kalisz, 2016. P. 17-27.
7. Стефанишин Д.В. Ситуаційно-індуктивне моделювання в задачах екстраполяційного прогнозування за даними моніторингу. Системні дослідження та інформаційні технології, 2016, №4. С. 35-45.
8. Stefanyshyn D.V. Improving diagnostic models for forecasting the behavior of dams equipped with automated monitoring systems. Mathematical modeling in economy, 2017, № 3-4. P.P. 50-61.
9. Stefanyshyn D.V., Korbutiak V.M., Stefanyshyna-Gavryliuk Y.D. Situational predictive modelling of the flood hazard in the Dniester river valley near the town of Halych. Environmental safety and natural resources, 2019, Is. 1 (29). P. 16-27.

10. Russell S.J., Norvig P. Artificial Intelligence: A Modern Approach. 3rd ed. Pearson Education, Inc.: Upper Saddle River, New Jersey, 2010. 1132 p.
11. Reiter R. Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems. MIT Press, 2001. 424 p.
12. McCarthy J. Situations, actions, and causal laws. Memo 2: Stanford University Artificial Intelligence Project, 1963. 11 p. URL: http://www.dtic.mil/dtic/tr/fulltext/u2/785031.pdf.
13. Ivakhnenko A.G. Polinomial theory of complex systems. IEEE Transactions on Systems Man and Cybernetics, 4, 1971. P. 364-378.
14. Madala H.R., Ivakhnenko A.G. Inductive Learning Algorithms for Complex System Modeling. CRC Press, 1994. 368 p.
15. Berthold M., Borgelt Ch., Höppner F., Klawonn F. Guide to intelligent data analysis: how to intelligently make sense of real data. London : Springer-Verlag. 2010.
16. Taleb N.N. The Black Swan. The Impact of the Highly Improbable. Second enlarged edition. Random House. New York, 2010. 480 p.
17. Collacott R.A. Structural Integrity Monitoring. Chapman and Hall: London, New York, 1985. 455 p.

REFERENCES

1. Kuhn, M., Johnson, K. (2013). Applied Predictive Modeling. Springer Science+Business Media: New York, 600 p.
2. Geisser, S. (2016). Predictive Inference: An Introduction. New York: Chapman & Hall, 264 p.
3. Brockwell, P.J., Davis, R.A. (1991). Time series: Theory and methods, second ed., Springer, 557 p.
4. Hamilton, J. D. (1994). Time series analysis. Princeton University Press: Princeton, New Jersey, 782 p.
5. Stefanyshyn, D.V. (2013). A method of forecasting of indexes of dynamic system that evolves slowly, based on time series analysis. ICIM 2013, Proc. of 4th Int. Conf. on Inductive Modelling, Kyiv, Ukraine, September 16-20, 221-224.
6. Stefanyshyn, D.V. (2016). Situational and inductive modelling in problems of forecasting behavior of dams equipped with modern automated monitoring systems. Zeszyty Naukowe Inżynieria Lądowa i Wodna w Kształtowaniu Środowiska, Nr 14, Kalisz, 17-27.
7. Stefanyshyn, D.V. (2016). Situational and inductive modelling in problems of extrapolation forecasting on base of monitoring data. System Research and Information Technologies, №4, 35-45. (in Ukrainian).
8. Stefanyshyn, D.V. (2017). Improving diagnostic models for forecasting the behavior of dams equipped with automated monitoring systems. Mathematical modeling in economy, № 3-4, 50-61.
9. Stefanyshyn, D.V., Korbutiak, V.M., Stefanyshyna-Gavryliuk, Y.D. (2019). Situational predictive modelling of the flood hazard in the Dniester river valley near the town of Halych. Environmental safety and natural resources, Is. 1 (29), 16-27.
10. Russell, S.J., Norvig, P. (2010). Artificial Intelligence: A Modern Approach. 3rd ed. Pearson Education, Inc.: Upper Saddle River, New Jersey, 1132 p.
11. Reiter, R. (2001). Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems. MIT Press, 424 p.
12. McCarthy, J. (1963). Situations, actions, and causal laws. Memo 2: Stanford University Artificial Intelligence Project, 11 p. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/785031.pdf.
13. Ivakhnenko, A.G. (1971). Polinomial theory of complex systems. IEEE Transactions on Systems Man and Cybernetics, 4, 364-378.

14. Madala, H.R., Ivakhnenko, A.G. (1994). Inductive Learning Algorithms for Complex System Modeling. CRC Press, 368 p.
15. Berthold, M., Borgelt, Ch., Höppner, F., Klawonn, F. (2010). Guide to intelligent data analysis: how to intelligently make sense of real data. London, Springer-Verlag.
16. Taleb, N.N. (2010). The Black Swan. The Impact of the Highly Improbable. Second enlarged edition. Random House. New York, 480 p.
17. Collacott, R.A. (1985). Structural Integrity Monitoring. Chapman and Hall: London, New York, 455 p.

Стаття надійшла до редакції 18.08.2019.