International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:5, No:6, 2011

Machine Learning Methods for Environmental Monitoring and Flood Protection

Alexander L. Pyayt, Ilya I. Mokhov, Bernhard Lang, Valeria V. Krzhizhanovskaya, Robert J. Meijer

***Abstract*—**More and more natural disasters are happening every year: floods, earthquakes, volcanic eruptions, etc. In order to reduce the risk of possible damages, governments all around the world are investing into development of Early Warning Systems (EWS) for environmental applications. The most important task of the EWS is identification of the onset of critical situations affecting environment and population, early enough to inform the authorities and general public. This paper describes an approach for monitoring of flood protections systems based on machine learning methods. An Artificial Intelligence (AI) component has been developed for detection of abnormal dike behaviour. The AI module has been integrated into an EWS platform of the *UrbanFlood* project (EU Seventh Framework Programme) and validated on real-time measurements from the sensors installed in a dike.

***Keywords*—**Early Warning System, intelligent environmental monitoring, machine learning, flood protection.

I. Introduction

More than two thirds of European cities are regularly confronted with natural disasters like floods, earthquakes, volcanic eruptions, etc. Early Warning Systems (EWS) play a crucial role in mitigating the effects of such disasters by detecting conditions which forecast the onset of a catastrophe and by evaluating the impact. The EWS provide decision support and information services to governments, companies and general public.

The goal of the *UrbanFlood* FP7 [1] project is the development of an Internet-based service platform for early warning systems validated for flooding. Most of the current EWS forecast the water heights in rivers and canals, but not the condition of dikes and the danger of dike breaches. The Ikdijk experiments [2] showed that information from sensor networks can help to predict dike failures. This approach shall be extended by in-time forecasts of such failures.

The role of the artificial intelligence (AI) component is to detect the abnormal behaviour of the object and to provide early indicators for the decision support system. The EWS infrastructure includes cloud and grid resources of the *UrbanFlood* project, distributed virtualization servers in the Netherlands, Russia and Poland. The advantage of a proposed EWS architecture is that a number of AI components can run on distributed resources, therefore the relatively complex computational tasks could be carried out effectively and efficiently, to provide a near real-time response in processing the sensor data.

Alexander L. Pyayt is with Siemens LLC, Russia (e-mail: alexander.pyayt@siemens.com).
Ilya I. Mokhov is with Siemens LLC, Russia (e-mail: ilya.mokhov@siemens.com).
Bernhard Lang is with Siemens AG, Germany (e-mail: bernhard.lang@siemens.com).
Valeria V. Krzhizhanovskaya is with Universiteit van Amsterdam (University of Amsterdam), the Netherlands (e-mail: V.Krzhizhanovskaya@uva.nl)
Robert J. Meijer is with Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO), the Netherlands (e-mail: robert.meijer@tno.nl).

II. The State of the Art

Environmental conditions like heavy rain, snowmelt, storms, etc. together with failures of water defence systems (e.g. caused by piping) are the main root causes of floods. Most of the state-of-the-art approaches mentioned in this section consider only the environmental data like the rain fall and winds. One of the advances of the *UrbanFlood* project is that a complete data set is used, including structural health parameters of the flood-protection dikes. The complexity of these data sets is very high and moreover it is not always possible to determine the inherent analytical dependencies within the data. Therefore some sophisticated methods are required in order to distinguish normal and abnormal situations and to estimate the level of emergency for a given state of the dike.

One of the tasks of the *UrbanFlood* EWS is to bring together the dike structural health data and the environmental data and use this combined data for further monitoring and abnormal behaviour detection. The dike behaviour monitoring model should use input data from the dike (such as water pressure inside the dike, core temperatures, displacements) and external parameters including environmental parameters, such as the weather conditions, dike repair works, traffic load, etc. as input data. The output of this model is the dike behaviour characteristic, e.g. confidence value of dike stability.

Let us consider two main classes of such models suggested in [3]: descriptive and cause-response models. Descriptive models are restricted to the pure geometrical description of the deformation process. They do not take into account influencing parameters like temperature, wind or traffic loads, whereas parametric cause-response models, for example finite element model, evaluate the reaction of the construction to external influence. In comparison to the parametric models, the structure of nonparametric cause-response model is not specified *a priori*. This type of models is directly derived from the measurements [3], including machine learning methods, such as neural networks and clustering methods. The high

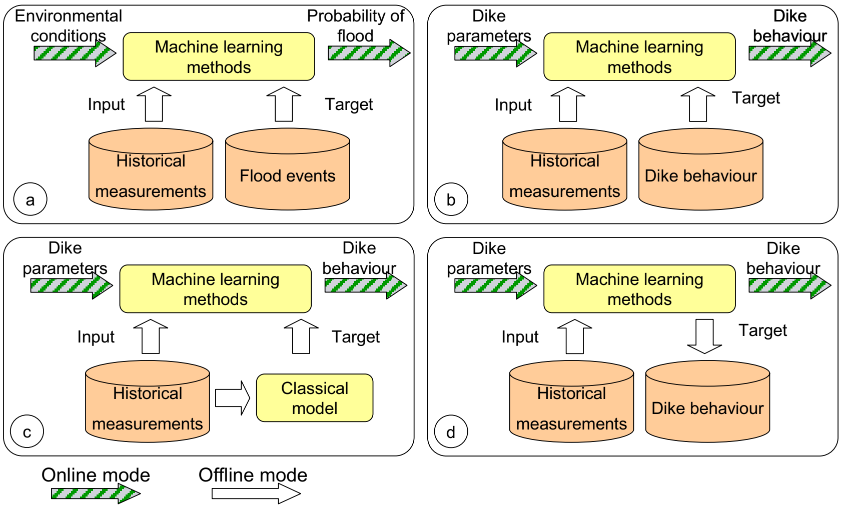

Fig. 1 Four types of machine learning methods for flood monitoring

potential of nonparametric methods for monitoring and analysis of dike deformation signals is described in [3].

In the *UrbanFlood* project various types of models are used as part of the EWS: the BREACH model developed by HR Wallingford [4] that combines features of descriptive and parametric cause-response models; and parametric cause-response Virtual Dike model developed by the University of Amsterdam [5]. Here and in the following sections these models will be referred as “classical”. The first version of the AI component was developed on the basis of nonparametric cause-response methods.

Dike abnormal behaviour can be detected by on-site or remote visual inspection; however it usually means that failure development is detected too late for proper maintenance actions. Analysis of the measurements collected at the object is required for early detection of the failure development. For dike monitoring, core temperature could be used as an indicator of leakage [6]. Characteristics of dikes stability include movements (horizontal, vertical, rotational, lateral), pore pressure and ground water flow through the dike [7]. Moreover, dike geometry (angle, type of ground, height, width) and environmental parameters (atmospheric pressure, air humidity, rainfalls, river flow, water height and others) are usually considered.

In many publications floods are forecasted using environmental parameters only (see Fig. 1a). For example, in [8] flood forecast was based on a river flow. It was the only output of prediction model, which was built by a linear regression model. Flow rate, air temperature and rainfall were used as input parameters. Flood forecast method based on rainfall data was suggested in [9].

Fig. 1b shows classical application of neural networks for dike behaviour analysis in case of available training set, for example, in [10] one of dike parameters was used as dike behaviour characteristic. Although there are problems with availability of such data sets, since usually there are not so many measurements related to development of failures. In this case some classical model is used to convert geometrical parameters of dike into slope stability, e.g. in [11] dike behaviour was calculated by the Bishop model and used as a target for neural network. It is illustrated in Fig. 1c.

In addition to the abovementioned examples, it would be beneficial to have an algorithm that can distinguish normal behaviour from abnormal based on only historical measurements of a *normal* dike behaviour. This is schematically shown in Fig. 1d, where the "Target" arrow is directed downwards. This approach is considered in this paper as a part of the abnormal behaviour detection strategy. Processing the sensor data is described in Section 3.

III. The Approach for Abnormal Behaviour Detection

A. Approach Description

Dike condition monitoring shall be carried out continuously. Numerous sensor data have to be processed in order to monitor dike condition. Obviously the amount of raw data is too large for manual inspections.

Automatic sensor data validation and data aggregation have to be carried out. Machine learning methods are selected to support the extraction of fault indicators for further analysis. Fig. 2 depicts a generic scheme of an automatic detection of abnormalities. Environmental conditions and dike parameters are input data, which are analyzed after pre-processing by the set of the classification agents – Neural Clouds (NC). Confidence values calculated by NC can be used after that for further analysis.

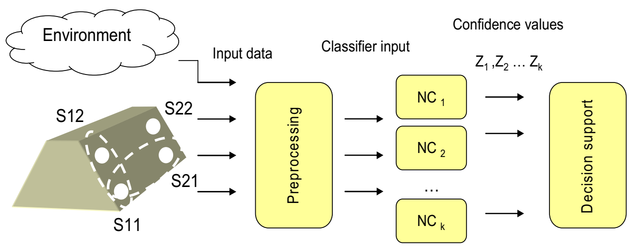

Fig. 2 Generic scheme of abnormality detection scheme

It is very important to distinguish sensor faults from dike failures. For example, one sensor indicates abnormal behaviour, whereas others show usual behaviour at the same time. It could be either a leakage or a sensor defect. In this case some redundant information is used. Clustering methods and statistical correlation are used for detection of analytical groups of installed sensors. This is a so called analytical redundancy. Physical redundancy (Fig. 2) means that sensors are grouped according to rules like: all sensors that measure the same characteristic – e.g. S11 and S21) or sensors from the same cross-section, e.g. S11 and S12). Moreover, the relations between environmental parameters and dike characteristics should be analyzed.

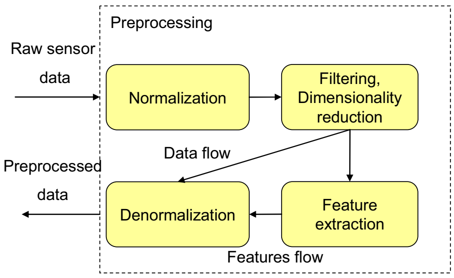

Fig. 3 Generic scheme of the pre-processing stage

Data pre-processing scheme, as the first part of data analysis flow, is presented in Fig. 3. It includes data normalization, filtering, feature extraction and denormalization blocks. There are two main data flows inside each pre-processing block: sensor data flow and calculated features flow. Further analysis of the sensor data flow involves the application of approach for abnormal behaviour detection. One of the aims of data pre-processing is to build a set of redundant features representing possible dependencies within the data set and the critical data properties. Features calculated by different methods in frequency and time domain for sliding windows, will be used both as an inputs for the abnormality detection block and as independent abnormality indicators. Detected outliers and data gaps caused by sensor network or communication problems will be filled in using interpolation methods.

A dike is an environmental object, therefore, methods commonly used for environmental time series analysis can be applied for dike behaviour analysis, e.g. singular spectrum analysis [12], Hilbert-Huang transform analysis [13], time series decomposition in seasonal, cyclical, trend and irregular component. Econometric analysis, fractal analysis and nonlinear dynamics methods will be used as independent indicators of time series changing properties.

Pre-processed data and calculated features will be used for the system behaviour classification. In *UrbanFlood* project, the authors suggest to use a one-side classification concept for abnormality detection. The basic idea behind the one-side classification in the field of environmental monitoring is that the majority of the measured data usually corresponds to the normal conditions of the environmental object (e.g. dike). Data collection of abnormal or critical conditions is hardly possible, and fault modelling gives mainly the information that can support experts in understanding the failure mechanisms, but decision support could be still rather complicated. Taking into account the abovementioned issues we suggest the application of so-called Neural Clouds [14] for the dike abnormal behaviour detection.

B. Neural Clouds

The NC classification algorithm receives the pre-processed data and a set of extracted features as an input. The core of the NC classification agent (single classification algorithm) is a combination of an Advanced K-Means (AKM) clustering algorithm and an extended Radial Basis Functions (RBF) network approach (Fig. 4) [14].

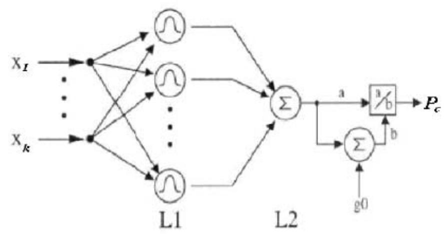

Fig. 4 Radial Basis Function network representation [14]
L1 part - Gaussian bells,
L2 part – superposition of Gaussian bells,
g_0 – normalizing parameter,
P_c - confidence level

AKM is a modification of the well-known k-means algorithm with an adaptive calculation of "optimal" number of clusters. Output of the AKM algorithm are centres of clusters representing historical data related to "normal" behaviour (training period).

After all the centres of clusters have been extracted from the input data, the data is encapsulated with the hyper surface. [14] For this purpose Gaussian distributions or the so called "Gaussian bells" are used.

$$R_i(x) = e^{-\frac{|x-m_i|^2}{2\sigma^2}} \qquad (1)$$

where m_i is the centre of the Gaussian bell, σ is the width of the Gaussian bell, x is the input data. The centres of the AKM clusters become the centres of corresponding Gaussian bells (Fig. 4 – L1 part).

The sum of all Gaussian bells is calculated, in order to obtain the encapsulating surface [14]. The sum of the Gaussian bells can be more than unity in case these bells overlap. Normalization factor g_0 (Fig. 4) is applied to make the confidence values P_c calculated by Neural Clouds in boundaries from 0 to 1 only. Confidence values close to 1 are reflecting normal behaviour while values close to 0 are reflecting anomalies.

The NC encapsulates all previously known configurations of selected parameters for a given training period. After training, the NC calculates a confidence value for every new state of the dike, describing the confidence value of abnormal behaviour.

The basic idea of NC for complex system monitoring is based on implementation of a single classification instance aimed to detect the deviations of the whole system behaviour from the normal state. In this work, we extend this approach by introducing a set of classification agents (Fig. 5), which can be trained and used independently to perform the fault or abnormality localization. This way, we also efficiently utilize the cloud computing infrastructure of the EWS.

The benefits from using smaller scale classification algorithms running on distributed computational resources are the reduced computational load and faster response times. It is proposed to group the data sets by data type, sensor location and possible dependencies within the data sets. Further, we train the committee of NC encapsulators separately for each set. Local classification agents are responsible for detection of abnormalities in the data from given sensor group, of correlations between the data from selected sensors, and of other extracted features (Fig. 2, 5). Such an extension could significantly improve the accuracy of the abnormality detection and make the follow up decision support processes more effective.

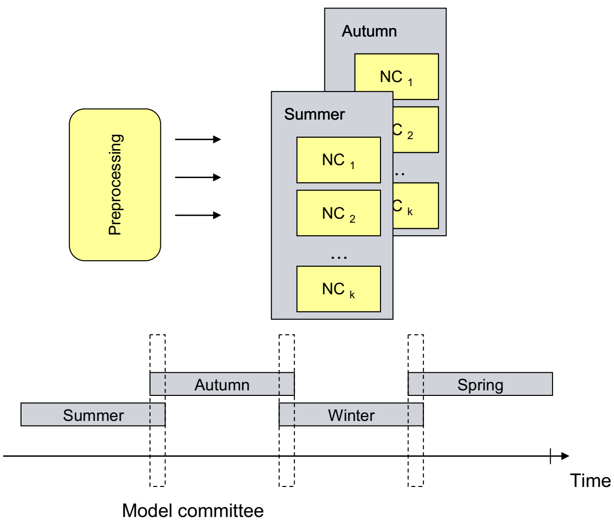

Fig. 5 Classification models adopted for seasonal changes

Another step towards the development of the NC application strategies comes from the nature of the environmental systems. Dike parameters depend on the external conditions: time of the day, high or low tide, atmospheric pressure, wind, seasonal temperature fluctuations, etc.

We suggest to create several models of classification agents (Fig. 5) and to train them according to the selected rules: for example, models with the same internal structure like the number of classification agents, inputs for the agents, etc. can be used for different seasons. Training is performed according to the predefined seasons; and the models are activated based on the selected schedule. Overlapping of the models would give a possibility to consider the transition intervals using the confidence values from their committees (Fig. 5).

The approach described in this section is used as a basis for implementation of the AI component in the *UrbanFlood* early warning system.

IV. Artificial Intelligence Component

The first version of the AI component was developed in C++ and Java programming languages and integrated into the EWS infrastructure (Fig. 6). Real-time dike measurements in XML format are published into Java Message Service (JMS) bus, and consumed by the AI component. The results of data analysis are sent back to the JMS topic to be used by decision support system and other components. The output message of the AI component also writes in XML format and contains calculated confidence value, timestamps of measurement and analysis. Data analysis block is based on the abnormal behaviour detection scheme. Input measurements and calculated confidence values are visualised by the WebDashBoard block. Only a web browser is required for visualization of dike measurements and results of data processing. Self-monitoring block is responsible for providing all the other EWS components with information about the AI component state. The main goal of this block is to increase robustness of the applications and of the EWS infrastructure. More details can be found in official website of the *UrbanFlood* project [1].

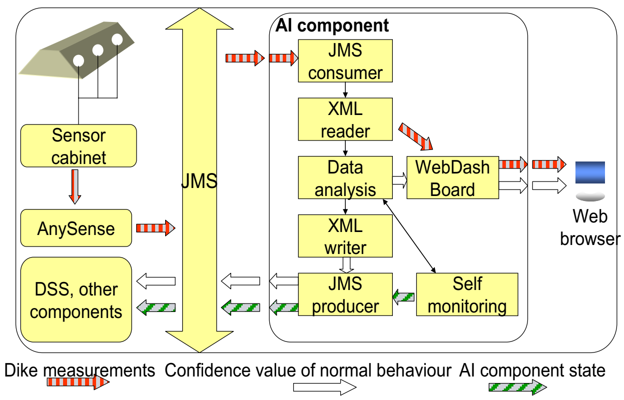

Fig. 6 Artificial Intelligence component as part of the *UrbanFlood* early warning system

V. Results of Stammerdijk Data Analysis

One of the dike failure mechanisms is the slope macro-instability. This mechanism has been extensively studied during the Ijkdijk experiments [15], when a full-scale levee crashed as a result of high water content inside the dike and a heavy load placed on the top of it.

The “Stammerdijk” dike in Amsterdam, the Netherlands was equipped with a detailed network of GeoBeads sensors continuously measuring pore water pressure, temperature and inclination (Fig. 8). Sensor modules have been installed in two cross-sections up to 10 meters below sea level in various ground layers (sand, clay and peat) [16].

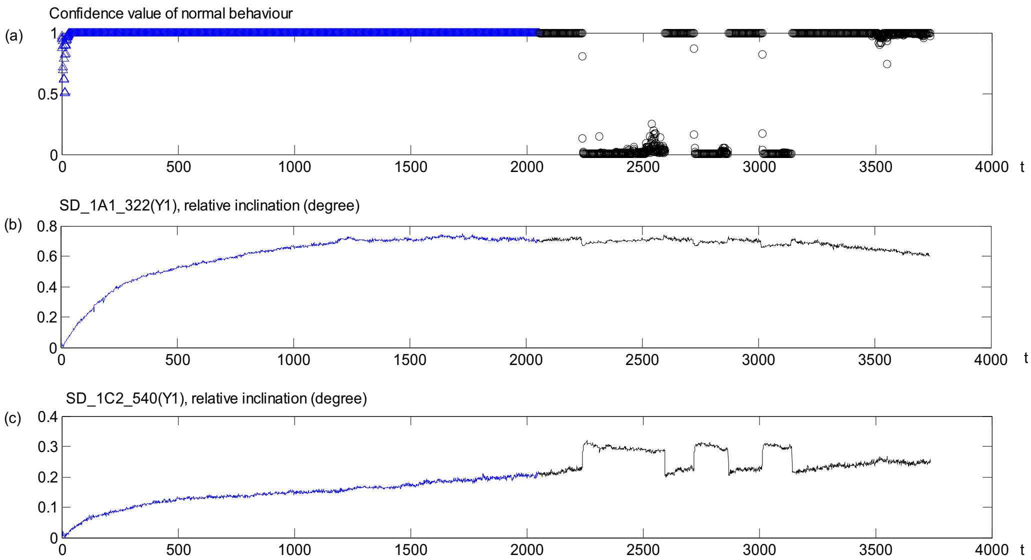

Fig. 7 Confidence values calculated by Neural Clouds for 2 parameters. (a): calculated confidence value of dike normal behaviour, (b) and (c) – relative inclination – in degrees (relative to reference time, averaged over 1 day). Training period of NC is from 1 till 2054 timestamp, testing period is from 2055 till 3735. X axis for (a), (b), (c) represents discrete time step number, with the time step of 10 minutes between the measurements.

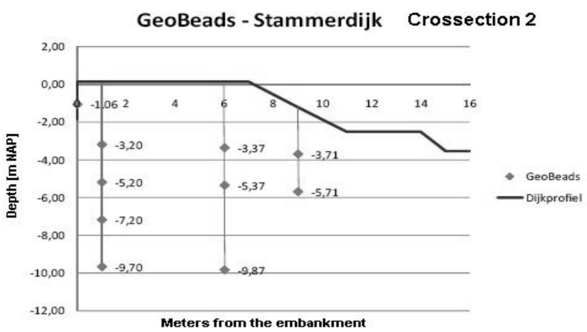

Fig. 8 One of the Stammerdijk cross-sections [17] with marked sensor locations

During the "macro-stability" tests done at the Stammerdijk, heavy load was placed on dike. Main aim of this test was analysis of ability of installed sensor network to translate the influence of external circumstances on levee stability [16]. This is suitable situation for testing the proposed approach for abnormal behaviour detection.

Some of sensors indicated influence of additional load on dike, for example, inclinometers SD_1A1_322(Y1) and SD_1C2_540(Y1), which measure relative inclination in degrees (relative to reference time, averaged over 1 day) of GeoBeads. Here "SD" stands for StammerDijk, "1" is the cross-section number, "C2" is the sensor location, "540" is the depth of the sensor in centimetres. Fig. 7b and Fig. 7c show that additional load placed three times at dike was successfully detected by selected sensors – three rapid changes in graphics are related to the events of placing (time steps with numbers 2238, 2711, 3012) and removal (steps with numbers 2607, 2866, 3140) of the load

These anomalies were detected by the Artificial Intelligence component (Fig. 7a). As training set the data related to normal mode were used. Test set includes data with anomalies as well as "normal" data. Fig.7a shows that suggested approach is able to detect such kind of anomalies: in case of "normal" data confidence values are close to 1, in case of anomalies close to 0.

Fig. 9 is a 2D projection of NC trained on Stammerdijk

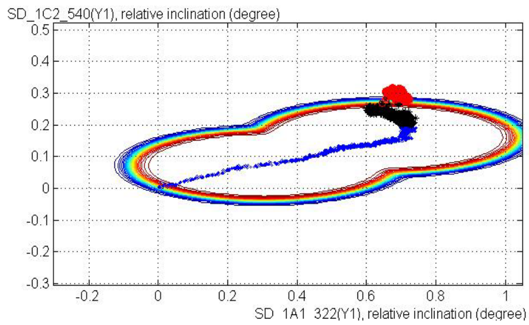

Fig. 9 2D projection of Neural Clouds. Confidence levels are represented by the contour lines

Input data from two selected sensors were pre-processed: measurements were normalised. Pre-processed dike parameters were used as input for NC training. Training set (Fig. 7a) was encapsulated by NC (Fig. 9, 10).

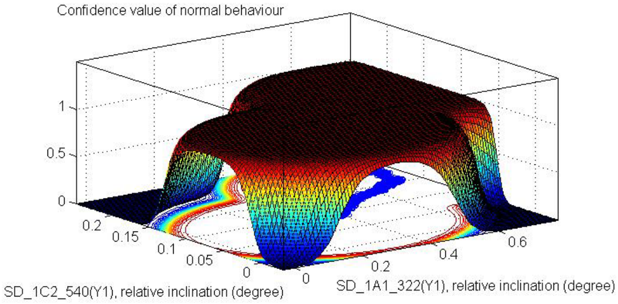

Fig. 10 Constructed Neural Clouds. Confidence levels are represented by the 3D surface

VI. Conclusions and Outlook

In this paper we presented the concept of abnormal behaviour detection for Early Warning Systems (EWS) for environmental applications. The Neural Clouds (NC) approach allows an early detection of the system state which was not presented in the training set, and could be considered as part of the EWS responsible for early detection of the abnormal behaviour. According to the review of the state-of-the-art, this is the first time such an approach is used for environmental applications.

Preliminary results of NC application to real-world Stammerdijk data have proved the efficiency of this approach for abnormal behaviour detection. This strategy has been implemented in the AI component and integrated into the *UrbanFlood* early warning system prototype. Most of the classical models are resource-intensive, whereas application of an AI component is an efficient method for online data processing.

The next step is to apply the presented approach to the analysis of the data from experimental dike IJkdijk [2] and real dike Livedijk [18]. It is important to extend the results of the Stammerdijk data analysis to the Ijkdijk macro-stability experiments, to be able to compare the results calculated for different dikes.

For further development of the methodology, it is planned to introduce a forecasting component allowing classification of the future states of the system.

We should mention that presented abnormal behaviour detection approach shouldn't be considered as a stand-alone application for dike health assessment. Low confidence values (detected anomalies) should be used as indicator for initiation of further data analysis by computational models, which can calculate dike stability.

As the next step, machine learning methods will be combined with classical models to make the system more robust. For example, the Virtual Dike [19] advanced computational model can be used for generating training sets for the AI component.

Dike is a complex object that is affected by many different factors, including environmental conditions. To improve the quality of classification, further in the project Neural Clouds should be adapted to the changing environmental conditions.

Acknowledgment

This work is supported by the EC FP7 project *UrbanFlood*, grant N 248767. The authors would like to thank WaterNet, particularly Rob van Putten, for providing Stammerdijk measurements and support.

References

[1] Official site of *UrbanFlood* project, http://UrbanFlood.eu
[2] Official site of IJkdijk project, http://www.ijkdijk.eu/IJkdijk
[3] Eichhorn, A.: Tasks and newest trends in geodetic deformation analysis: a tutorial. In: 15th European Signal Processing Conference (EUSIPCO 2007). Pp. 1156-1160, 2007
[4] HR Wallingford Breach model site, http://www.floodsite.net/html/HR_Breach_Model.htm
[5] Krzhizhanovskaya, V.V. et al. Flood early warning system: design, implementation and computational modules. Proceedings of the International Conference on Computational Science, ICCS 2011. Procedia Computer Science 2011 (to be published)
[6] Khan, A., Vrabie, V., Mars, J., Girard, A., D'Urso, G.: Automatic monitoring system for singularities detection in dikes by DTS data measurement. IEEE Transactions on Instrumentation and Measurement. Pp. 1-17, 2009
[7] Dam safety: an owner's guidance manual. Federal emergency management agency. FEMA 145 / August 1987.
[8] Basha, E., Ravela, S., Rus, D.: Model-Based Monitoring for Early Warning Flood Detection. In: proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys). Pp. 1-14, 2008
[9] Brath, A., Montanari, A., Toth, E.: Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. Hydrology and Earth System Sciences, 6(4). Pp. 627–640, 2002
[10] Kaunda, R., Chase, R., Kehew, A., Kaugars, K., Selegean, J.: Neural network modeling applications in active slope stability problems. Environmental Earth Sciences, 60. Pp. 1545–1558, 2010
[11] Choobbasti, A., Farrokhzad, F., Barari, A.: Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran). Arabian Geoscience Journal, 2. Pp. 311–319, 2009
[12] Ghil, M., Allen, M.R.: Advanced spectral methods for climatic time series. Reviews of geophysics, 40, 1/Month, pp. 1-41, 2002
[13] Rao, A.R., Hsu, E.-C.: Hilbert-Huang transform analysis of hydrological and environmental time series. Water Science and Technology Library. Vol. 60. Springer Netherlands, 2008
[14] Lang, B., Poppe, T., Minin, A., Mokhov, I., Kuperin, Y., Mekler, A., Liapakina, I.: Neural Clouds for Monitoring of Complex Systems. Optical Memory and Neural Networks (Information Optics). 2008. Vol. 17. No. 3. Pp. 183–192, 2008
[15] Ijkdijk macrostabilty experiment description, http://www.stowa.nl/Uploads/onderzoek_projecten/IJkdijk%20Macro%20stability%20English.pdf
[16] Geobeads and *UrbanFlood*, http://www.alertsolutions.nl/projectdocs/Flyer_GeoBeads_*UrbanFlood*_nov2011.pdf
[17] Peters, E.: GeoBeads, multiparameter network for levee monitoring. Joint *UrbanFlood* & SSG4Env Workshop, Thursday 11 & Friday 12 November 2010 – Amsterdam, the Netherlands, http://*UrbanFlood*.eu/Documents/p5_*UrbanFlood*_2010.pdf
[18] Official site of Livedijk, http://www.inspectiewaterkeringen.nl/content.asp?page=289
[19] Melnikova, N.B., Shirshov, G.S., Krzhizhanovskaya, V.V. Virtual Dike: multiscale simulation of dike stability. Proceedings of the International Conference on Computational Science, ICCS 2011. Procedia Computer Science 2011 (to be published)