Data-driven methods in application to flood defence systems monitoring and analysis
Pyayt, A.

Citation for published version (APA):
Chapter 1. Introduction*

1.1. Levee health monitoring problem

According to the European Union (EU) Floods Directive [4] “flood” means “the temporary covering by water of land not normally covered by water”. There are about four times more floods registered nowadays in comparison to the 1980s [6]. A detailed classification of the floods can be found in [44].

Usually flood is a result of high water level due to different reasons, e.g., heavy rains or melting snow. There are several well-known projects dedicated to flood monitoring that consider water level as the main factor describing the onset of failure. The Flood Early Warning System (FEWS) described in [20] monitors rainfall, water levels and low water crossings in Austin. Information about Bandon flood early warning system can be found in [1]. This system is designed to give up to 5 hours prior warning of a flood event. Information about Local Flood Early Warning System (LFEWS) can be found in [10]. Most of the mentioned above projects consider water level forecasting as the only method of generation of alarm: if the forecasted water level is above some predefined threshold the experts are notified.

Some of the floods occur even in case of not critical water level, for example, flood due to dike failure in Wilnis (the Netherlands) happened as a result of drought [30]. Condition monitoring of flood defence structures (e.g., dams) becomes an important task nowadays.

A “dam” is as an “artificial barrier constructed across a watercourse for the purpose of storage, control, or diversion of water” [18]. Dike (or levee) is usually an earth-filled dam.

There are different reasons of levee failures. Most of them are caused by overtopping (see Figure 1.1(a)) (that often leads to external erosion of earth-fill dams), foundation defects, piping and seepage [125] that occur due to problems in the structural design or because of the lack of understanding of the relevant geological properties beneath and around the dam structures. Comprehensive analysis of various flood defence structure failure mechanisms can be found in [38] and [24].

According to a study of dam failures in the USA [16] overtopping is the reason of 34% of the observed floods. Foundation defects due to differential settlement, slides, slope instability, uplift pressures, and foundation seepage lead to 30% of all dam failures. Failure

---

* Parts of this chapter were published in:
due to piping and seepage accounts for 20% of all failures. The remaining 16% of failures are caused by the problems with conduits and valves, and other miscellaneous problems.

1735 dikes failed in the Netherlands between 1134 and 2006 [29]. 67% of the events were caused by erosion of inner slope protection, 11% by ice drift, 6% by erosion (Figure 1.1(c)) or instability of outer slope protection, 5% by sliding inner slope (Figure 1.1(e)), 4% by external reason (human and animal), 3% by sliding outer slope (Figure 1.1(f)), 2% by liquefaction of shore line, 1% by piping, 1% by micro instability (Figure 1.1(b)), horizontal shear (Figure 1.1(d)) and other related mechanisms.

Figure 1.1. Classification of dam failures: (a) wave overtopping; (b) micro-instability; (c) erosion of outer slope; (d) horizontal shear; (e) sliding of inner slope; (f) sliding of outer slope (based on classification presented in [112]).

Recent research projects that considered levee stability problem are FLOODsite [13], FloodControl 2015 [11] and UrbanFlood [12].

1.2. Overview of flood defence monitoring technologies

1.2.1. Sensor technologies

The mechanism of a possible failure is unknown beforehand and is difficult to be predicted. Visual inspection cannot guarantee detection of an onset of a levee failure early enough to prevent its collapse. Therefore a continuous levee health monitoring is required.

There are two general approaches of continuous dam monitoring: remote sensing, e.g., by LiDAR [47] or by satellite [53], and in situ sensing. The latter technology assumes installation of sensors inside the levee.

LiDAR technology measures distance by illuminating a target with a laser and analyzing the reflected light. This technology provides construction of high resolution maps. Comparison of the maps constructed at different moments in time provides detection of dam deformation.

GPS is also used for surface monitoring of concrete dam [117]; real-time GPS deformation monitoring system is presented in [2]. Thermo-graphic camera providing
information about temperature of the monitored object has been used during the IJkdijk experiments [58].

As we see, list of the monitored parameters is limited: deformation and temperature. This means that remote monitoring is useful for earth-fill and concrete dam monitoring in case of foundation defects detection (presented as cracks, for example). External erosion of an earth-fill dam can be detected; onset of a structural failure can be detected using remote sensing for monitoring of the concrete dams; dam leakage or seepage can be also detected (depending on location of seepage and sensitivity of the camera).

If the more detailed analysis of object behaviour is required sensors should be installed inside the object. Figure 1.2 illustrates installation of a distributed fibre optic cable and installation of individual sensors in cross-sections inside an earth-fill dam.

The use of fibre optic cables for deformation analysis for earth-fill dams is described in [119]. The pore water pressure sensors proved to be useful in levee stability analysis (e.g., detection of internal erosion) [3]. Inclinometers are generally used to measure tilt and to monitor lateral movements for embankments and dams (in other words – this type of sensor measures levee instability) [43]. Leakage can be detected by distributed fibre optic sensors measuring the temperature inside the levee [122]. A detailed overview and comparison of existing sensor technologies for levee monitoring can be found in [85].

The advantage of remote sensing techniques is that they are non-intrusive and they provide monitoring of large areas. One of the disadvantages of the sensors usage is that procedure of their installation can be dangerous for the monitored object. Selection of sensor placement and number of sensors is another tough task. If more accurate and reliable results are required, sensors should be installed inside the monitored object.

Availability of the monitoring technology is not enough for efficient levee condition monitoring. Development of the models supporting expert in decision making is required.

### 1.2.2. Modelling approach for flood monitoring

Usually data-driven (model-free methods) and model-based (numerical or physical modelling) approaches are applied in order to detect the failure development.

Data-driven approach requires availability of the monitoring system that is constantly measuring the most important object characteristics defined by the domain experts. This class includes different techniques: machine learning methods (e.g., the artificial neural networks (ANN)), statistical methods (e.g., central moments, linear correlation, clustering), soft computing, data mining and others. Importance of data-driven
techniques application for hydro-informatics is presented in [115]; perspectives of these methods for application in geodetic deformation analysis are presented in [62]. These methods can be applied for on-line/real-time computations.

Data-driven methods can be applied for analysis of data collected by both types of sensor technologies: remote sensing (e.g., pattern analysis of images) and in situ sensing (e.g., forecasting of sensor measurements).

Dike failure due to uplifting and piping is defined as the event in which the resistance (the critical head) drops below the stress (the outer water level, a combination of both sea level and river discharge, minus the inner water level) [86]. For instance, the application of singular value decomposition (SVD) to distributed temperature values is suggested for automatic leakage detection in [65]. Artificial neural networks (ANN) were applied for slope stability analysis in [63]. The ANN were trained in [126] to identify possible locations of weak zones given surface displacements.

Model-based approach (numerical modelling) requires information about the monitored object (e.g., initial geometry and soil properties); the constructed model will not depend on on-line measurements. Sensor values may be used for validation of the constructed model and as initial conditions for the modelling. Finite element method is widely used to simulate structural behaviour of dikes [116], [106], [35], [82]. Linear elastic perfectly plastic rheological soil model by Mohr-Coulomb [121] is employed for earthen structures. Flow through porous media is simulated by solving either Darcy’s equation for saturated flow or Richards’ equation with wetting and drying of the soil, for the cases when water table changes in time [34]. Usually physical modelling is time consuming and requires significant computational resources. Because of high complexity of the models, they cannot be used for real-time data processing.

Detection of the developing failure can be carried out in two different ways. A model of the monitored object can be constructed and detection of deviations of model’s output from real object behaviour can be interpreted as an alarm (model-based and data-driven (ANN as a black-box model, for example) approaches can be applied). The second approach implies identification of the specific failure mode development by detection of specific patterns of failures (both approaches can be applied to this problem solving).

Data-driven methods fully depend on availability of data. For example, in case of dike behaviour analysis the historical measurements should be available. If the data are not available due to some reasons, the model-based approach can be used for generation of the required data set.

For the task of failure pattern recognition the sensor data gathered during the dike collapse should be available, that is usually impossible. Physical model can produce patterns of failure development and these data instead of results of real dike collapses can be used for data-driven methods training: pattern recognition or classification task.

This means that computationally heavy model-based methods can be used in two different ways in combination with the “fast” data-driven methods: validation of alarm produced by one of the data-driven methods (one-point result validation) or preparation of the predefined tables with slope stability coefficients or “virtual” sensors measurements (related to normal or abnormal dam conditions). Example of combination of both
approaches for levee health monitoring can be found in [104]. The use of both types of models is required for robust monitoring of flood defence structures.

### 1.3 Motivation and scientific challenges

There are different ways of dam monitoring: remote sensing or installation of sensors inside the monitored object. The first approach provides monitoring of wide areas, but only limited number of parameters is available (deformation and temperature). Installation of sensors provides more reliable results but process of installation can be sometimes dangerous for the monitored object. The collected measurements should be stored in a database (Figure 1.3).

There are a lot monitoring techniques that can provide sufficient information about levee behaviour. The only problem is analysis of the collected data. Expert is usually not able to analyse in on-line all the measurements collected from the sensors. Delay in decision making can be very expensive. Intellectual monitoring system is required for analysis of the collected stream of measurements and for support in decision making. An expert can use two different classes of the models for further decision making: data-driven or model-based approach.

![Figure 1.3. Required scheme of data analysis.](image)

All the mentioned in the Sub-Section 1.2.2 examples of data-driven methods application for levee monitoring had a goal to solve specific (local) tasks, e.g., detection of leakage (only one of the classes of the levee failures). There is no example of the general-purpose solution based on the data-driven methods that is able to detect onset of instability due to any failure mechanism. Model-based approach can definitely give very precise assessment of levee behaviour but it is not suitable for the on-line analysis because it usually requires high computational efforts.

That is why development of the data-driven-based general purpose system is required in order to provide both properties: on-line alerting of any (even unknown or combined) class of levee failure. These requirements do not exclude possibility of combination with physical modelling approach.

We suggest to use the following classification of levees states (Figure 1.4): anomaly or abnormal behaviour that can be a sign of developing failure. Levee can come back to the normal state after the anomaly, or levee can start collapsing. The failure states are related to different levee collapse scenarios: piping, macrostability, erosion and others.
Anomaly can be typical for an unstable dike; stable dike under tough external condition can be close to unstable conditions. This means that anomaly can be registered as a result of a non-destructive experiment or it can be simulated by a physical model.

![Classification of the dike states](image)

Figure 1.4. Classification of the dike states.

The developed levee monitoring approach should be able to aggregate collected from multiple sensors information, provide data analysis.

We can reformulate all the above mentioned items as one research hypothesis: one generic approach can indicate anomalies (instabilities) in flood defence structure using data-driven methods on-line.

The scientific objectives can be formulated as follows:

1) Investigation of raw sensor data properties. Selection of the required pre-processing procedures.
2) Investigation of different variants of construction of an anomaly detection approach.
3) Validation of the investigated anomaly detection approach on real-world objects.
4) Development of the anomaly detection approach as a component of the early warning system.
5) Development of the approach for combination of the data-driven anomaly detection approach with physical modelling and validation of the developed approach.

1.4. Outline of the thesis

Chapter 2 contains short description of the projects where the anomaly detection approach has been developed and applied: UrbanFlood and IJkdijk experiment (All-in-One Sensor Validation Test). This chapter describes sensor technologies installed into the monitored objects. The monitored levees are shortly presented in this chapter.

Chapter 3 describes state-of-the-art in data processing and anomaly detection domains. This chapter presents problems of sensor data analysis. Stages of sensor measurements analysis and the developed anomaly detection concept are introduced in this chapter. The developed concept includes two different anomaly detection approaches: one-side classification and transfer function (scientific objectives 1 and 2).

Chapter 4 presents detection of anomalies in measurements collected from the stable levees that were results of non-destructive experiments, sensor faults or real accidents (scientific objective 3).
Chapter 5 presents short description of implementation of the selected anomaly detection approach as the software component of the early warning system: artificial intelligence (AI) component (scientific objective 4). This component provides on-line analysis of the collected measurements.

Chapter 6 describes approach for combination of data-driven methods with physical modelling. Results of combination of data-driven methods with physical modelling are presented in this chapter: for the UrbanFlood project on the example of combination with the Virtual Dike, for the IJkdijk experiment – combination with finite element model developed by the experts especially for the experiment (scientific objective 5).

Chapter 7 brings a summary of results and conclusions.