7 Data-driven operation

7.1 Introduction

The introduction of informatics in the field of hydrodynamic modelling has led to many applications of data-driven modelling (DDM) and machine learning, studied in the cross-field of ‘hydroinformatics’. An overview of DDM in river basin management is given by Solomatine and Ostfeld (2008). Quoting the proceedings of the first international conference on the topic, “Hydroinformatics, understood as the application of advanced information technology to the problems of the aquatic environment, evolved through the 1980s, received its proper name in 1989 and was given a first published expression in 1991.” (Verwey et al., 1994). More recent examples are Pengel et al. 2012, Pyayt et al. 2011a-b, Krzhizhanovskaya et al. (2011) and Melnikova et al. (2011), who combine physics-based and data-driven model types into hybrid models in order to develop integral early warning systems for detecting dike instabilities and other flood safety issues. What sparked this data-driven paradigm is not only the steady increase in computational capacity and power, but also the growing availability and sophistication of sensors. In fact, it is not the hardware but the development of data processing algorithms that may be considered the bottleneck in most endeavours of putting the collected data to intelligent use.

In this chapter we work with the premise that the barrier gate contains certain sub-optimal properties with respect to dynamic loads. Reasons why this not unlikely were given in Section 1.1. Moreover, it is assumed that the FIV characteristics of the gate are not fully known. The aim of this chapter is to introduce a data-driven system for monitoring and preventing flow-induced gate vibrations based on data from sensors on the gates. That is, a model is built through data-intensive system identification and the model use is for prediction and control (see Section 1.2). Another way of saying essentially the same is that measurements are used to feed a self-learning system that facilitates (automated) decisions for gate operation. The trend of increasingly using remote operation of gates underlines the need to check that no long-running vibrations occur; an automated detection system would cover this need.

The past decade has seen several research groups working on the design and implementation of decision support systems for the control of flooding risks. The EU project Urban Flood (2012) made meaningful progress by proposing an artificial intelligence environment as a central component in an early-warning system that checks the states of flood barriers and detects anomalies. A five-year project with a similar scope, Flood Control 2015 (2012) consisted of a consortium of Dutch water-oriented companies, but did not look

---

into the dynamics of hydraulic structures with gates. The Leading Scientist Programme (LSP) comprises the development of an Advanced Computing Laboratory (2012), which has developed a computational infrastructure for studying complex systems and providing decision support, among others for flood defences. A literature survey revealed only one comparable study of a proposal to use sensors for the monitoring of gate dynamics, reported by Han et al. (2011). This proposal does not include artificial intelligence or control, however.

First a few remarks about measurements are made in Section 7.2. Then in Section 7.3 the overall set-up of the system is given. Sections 7.4 reflects on the applicability of the system and discusses possible extensions. Section 7.5 draws conclusions.

### 7.2 What quantities to measure?

The first step is to decide which physical quantities should be measured. In a situation where vibrations are observed or operational difficulties arise that are possibly the result of FIV, it is common practice to carry out field measurements. A good approach is to install acceleration sensors on the gates and monitor the gate behaviour for a certain period of time during which a representative range of conditions occurs. Figure 7.1 shows a temporary installation of an accelerometer that was used during several sessions of in situ measurements spread out over a few months.

*Figure 7.1. Temporary installations of acceleration sensors on the gate of a hydraulic structure. Both types measure acceleration in one direction, as shown by the black arrows. Pictures by Deltares.*

For the response, acceleration is generally easier to measure than displacement or velocity. For a harmonic oscillation these three quantities are naturally linked by differentiation (or integration) with respect to time. In the complex notation, differentiating with respect to time corresponds to multiplication with a factor $i\omega$. Support forces and pressures on the submerged gate are not easily measured in prototype, therefore the acceleration $\ddot{y}(t)$ is taken as the main structural response signal. Moreover, the gate lifting height or equivalently the gate opening in time $a(t)$ should be measured. This quantity is usually part of the gate lifting system anyway, so it just has to be registered.
For the hydraulic conditions, it is necessary to measure upstream and downstream water levels $h_1(t)$ and $h_2(t)$. These quantities are relatively easily measured for long periods of time. Main point of attention is to choose suitable measurement locations, so as not to be influenced by the surface roller or other effects directly related to the structure, including local bathymetry changes. Stationary measurements of point velocities near a structure don’t make much sense. This would have to be done at a great many locations in order to give a useful picture of the flow field. There exist different techniques to measure the discharge through a confined cross-section. Monitoring the flow discharge past a gate gives very useful information, for one because it would allow a data-driven model equivalent of the flow impact model cascade discussed in Chapter 4. However, accurate discharge measurements can be costly; some equipment is sensitive to sedimentation and debris, and a multi-gated barrier would need discharge meters in every gate. Moreover, one has to be aware that a single $Q-H$ relation does not describe all states of an unsteady flow due to hysteresis effects (e.g. Travaš et al. 2012). In this chapter discharge signals are not used.

The step from incidental prototype measurements to permanent gate monitoring is not expected to be complicated hardware-wise – it mostly relies on the installation of durable acceleration sensors. The use of wireless sensors would be a great improvement.

### 7.3 System set-up

#### 7.3.1 Overview of the components

The proposed system consists of a control loop with a numerical model fed by data from sensors attached to the gate. For now a single gate with submerged discharge is considered. Other than using the vertically measured gate opening as a measure for gate position, there is no principal reason to restrict the discussion to underflow gates. Figure 7.2 shows the functional block diagram of the system. The shaded grey rectangle marks the components that are discussed in this chapter.

![Figure 7.2. Schematic overview of the control system for gate dynamics. The components inside the grey rectangle are discussed in this chapter. The variables with hats are predictions.](image)
The starting point is the gate itself, which holds a certain position that is assumed to be described by its gate opening \( a \). The “data acquisition” module collects data from sensors installed on the gate, see previous section. The acquired signals are input for the “primary data analysis” module. It consists of appropriate signal processing methods (see Section 5.3) to derive the dominant frequency \( f_{\text{dom}} \) and mean displacement amplitudes \( A_i \) from the acceleration signals of the most recent time window. Subscripts \( i \) refer to the orientation and location of the sensor that produced the signal – for the one d.o.f. process discussed in previous chapters this simply corresponds to the vertical direction. The hydraulic head difference \( \Delta h = h_1 - h_2 \) and the mean gate opening \( a \) during this time are also registered.

The central module is the “machine learning” (ML) module which consists of a database of prior states plus a set of artificial intelligence algorithms that classify and interpret data. The ML module receives data from several sources. It receives input about the most recent situation (achieved hydraulic conditions and gate status) from the “primary data analysis” module. Furthermore it receives the latest predictions of water levels on both sides of the gate \( h_1 \) and \( h_2 \) from a separate, external numerical model of the flow system at a larger scale that produces water level predictions. The computed water levels of this system are used to propose the next gate opening \( a_{\text{new}} \) needed to accomplish certain far-field targets (e.g. minimum water depths for navigation). The ML module has the goal of determining the expected gate behaviour of the proposed setting \( a_{\text{new}} \) from the recently achieved state and the predicted water levels. If the recently measured situation closely resembles past states captured by data already present in the database, it will be relatively easy to draw a conclusion about the stability of proposed new gate settings. If necessary though, the ML unit will suggest an alternative, safer gate setting or opening/closing scenario.

Then this information is used in a decision step that chooses the most suitable scenario and prescribes (designs) the required operation. The decision step is shown in Figure 7.2 as a separate block, because it may not always be automated. The designed measure is carried out by the operation mechanism (controller) and the gate will attain a new position. The new situation will again be monitored by the sensors, thus completing the control loop. There is an optional extension to this system: the ML module can receive complementary data from another numerical model that performs physics-based simulations and produces a response prediction. The purpose of this is to fill in regions of missing data.

The system adopts a straightforward way of vibration control: avoidance of critical parameter range. This is done by adjusting settings of gate opening, by selection of starting time of the operation, or by adjusting the speed of the gate movement (this is not always possible technically).

### 7.3.2 Machine learning module

From a ML perspective, the sensory data (and possibly the simulation model) provide the database module with labeled data points. The ML operations on this data are therefore in the realm of supervised learning, more specifically, they are classification operations.

The reduced flow velocity \( V_r \) helps to classify flow-induced vibrations in different physical regimes (as seen in Chapter 5). Computing it as \( V_r = \sqrt{2g\Delta h/f_{\text{dom}}D} \) is convenient and
adheres to past studies. The acquired data (measured or calculated) occupies the three-dimensional space shown in Figure 7.3 (left).

For different gate openings $a_1$ and $a_2$, the gate response is quantified by displacement amplitude as function of the $V_r$ number. In this example, the strongest vibration at gate opening $a_2$ is found at about $V_r = 2.6$. Now, the problem that the control system should tackle is shown in the other plot of Figure 7.3. Suppose the gate has a constant position, so that the $a$-dimension drops out, and the head difference increases such that the dynamic state moves from $V_{r,1}$ to $V_{r,2}$, the question is then what response is expected if $V_r$ continues to grow to $V_{r,3}$. The empirical data points of the recent states between $V_{r,1}$ and $V_{r,2}$ and the known states for $V_r > V_{r,2}$ may be wrong indicators of the unknown state at $V_{r,3}$. Obviously, interpolation between the nearest known states in the $(V_r, a)$-plane will not predict the state at $V_{r,3}$ correctly.

To enable the use of efficient artificial intelligence, we introduce a critical mean amplitude threshold $A_{\text{crit}}$ (see Figure 7.3), below which the gate is said to be in a safe condition and above which vibrations are harmful. This allows a transformation of the problem into a binary classification problem: all data points belong either to a ‘critical’ or ‘unsafe’ class ($A > A_{\text{crit}}$) or to a ‘safe state’ class ($A \leq A_{\text{crit}}$). This is in essence a projection onto the $(V_r, a)$-plane.

Figure 7.4 shows the binary data points in the new plane. In this form, closure or opening under constant head difference is a vertical line. A changing head difference in time at fixed gate opening is shown as a horizontal line in this plane. The projected binary data points constitute a training set. The training objects are characterized by two attributes ($V_r$ and $a$) and one label (“safe” or “unsafe”). Most entries will be safe situations and in certain regions islands of critical vibration states will be found.
Figure 7.4. Fictitious example of measured data points in the \((V_r, a)\)-plane. Vibrations above the critical limit are indicated with triangles and vibrations below the critical limit with squares. This data is used as training data for classification of new states (+).

A classification algorithm is required that is capable of making accurate predictions to which class a new point is most likely to belong to. There exist many such algorithms. A basic choice is the K-nearest neighbours (KNN) algorithm (e.g., Rogers and Girolami, 2012 and Bishop, 2006). This non-probabilistic classification method associates a new point with a label based on the majority of the label values of the \(K\) data points nearest to it. In this algorithm there is in fact no specific training stage other than the construction of a matrix with all distances between the points. Parameter \(K\) must be chosen carefully as it determines the amount of smoothing; if \(K\) is too low there is a risk of overfitting and if it is too high underfitting can occur. Here, it is proposed to determine the value of \(K\) by a \(N\)-fold cross-validation, i.e. by repeatedly using different parts of the training set as validation data (also called test data). The data set is divided into \(N\) subsets of equal size. One subset plays the role of validation data and the other \(N-1\) subsets are the training points. This is done \(N\) times, so that all subsets play the role of validation data once, producing one error (defined as the proportion of incorrect classifications) each time. Below a concise pseudo-code for cross-validation applied to the KNN algorithm is given.

\[
\begin{align*}
\text{For } K &= 1 \text{ to } k \quad \text{% loop for } K \text{ values (using only odd values)} \\
\text{For } \text{fold=1 to } N \quad \text{% loop for N-fold cross-validation} \\
&\quad \text{randomly sudivide data in folds} \\
D &= [\text{empty}] \\
&\quad \text{construct matrix } D \text{ of Euclidean distances of training data} \\
D &= \text{sort}(D) \\
&\quad \text{For } j=1 \text{ to } K \quad \text{% loop of actual KNN} \\
&\quad \quad \text{find distances of } j \text{ closest points} \\
&\quad \quad \text{check binary label value of closest points} \\
&\quad \quad \text{determine majority of votes of closest points, this is the predicted label} \\
&\quad \quad \text{compute error of this fold} \\
\end{align*}
\]
A typical question the system should provide an answer to is whether it is safe to open a gate under certain conditions. First, $V_r$ should be determined using measured $\Delta h$ and response frequency $f$ estimated from nearby available data or using the gate's natural frequency. Figure 7.4 shows two options of opening scenarios that the system needs to evaluate, indicated by +’s, these are new points that require classification. In the case of opening in the lower of the two $V_r$ conditions, the system may give advice to wait with gate operation until the expected $V_r$ is reduced, in order to stay clear from the unsafe zone. Alternatively, the system may decide to open the gate at a faster rate, so that the critical zone is visited only briefly.

7.3.3 Physics-based model
The goal of employing a physics-based numerical model in this control system would be to provide the database with training data points in areas where no measured data is available. Next to water level data for boundary conditions, past response data would also be available to this model, making it essentially a prediction model. However, it was found in Chapter 6 that setting up, validating and applying a computational model for simulating fluid-structure interaction requires a large effort. It was concluded that it definitely has value for analysis of the (excitation) process, but that it not necessarily gives a sound representation of damping and amplitudes. Therefore, employing a physics-based model for providing $y_{dom}$ and $f_{dom}$ as shown in Figure 7.2 will be challenging. The best approach would be to carefully select a limited number of unknown states and attempt to use the analysis based on the outcomes of this physics-based model to evaluate the gate stability. The fact that running and interpreting this model is time-consuming and hard to automate, makes it as yet unfeasible to include it in the operational control chain.

7.4 Results of experimental data classification
The procedure of Section 7.3.2 is applied to the experimental data from Chapter 5. The data set used here for the classification consists of 145 observed vibrations. Figure 7.5 shows this data set in the same way as depicted by Figure 3. The dimensionless displacement amplitude was found by $A/D = (\hat{F}/k)/D$. The data represents measurements at various gate openings. During each test (corresponding to one data point in Figures 7.5 and 7.6), the discharge, gate opening and hence water levels were kept constant – the same can be assumed for field sensor data over short time windows.
Figure 7.5. Experimental vibration data of the original gate type from the physical scale model of Chapter 5. The plot shows non-dimensional displacement amplitude as a function of reduced velocity for steady-state conditions. The data set is divided into two classes using a threshold at $A/D = 0.004$. One high amplitude outlier at $V_r = 10$ is not plotted.

As known from Chapter 5, there are two regions of $V_r$ at which the vibrations grow in size considerably: $2 < V_r < 3.5$ and $V_r > 7$. A threshold at $A = 0.004D$ divides the data into two classes named "vibrations" and "no vibrations". The signals below this threshold contain irregular vibrations of small amplitude, above it more regular vibrations are found. Using this binary division, the experimental data is plotted as functions of $V_r$ and dimensionless gate opening $a/D$ in Figure 7.6 (left).

Figure 7.6. Left: binary identification of vibration regions: reduced velocity versus dimensionless gate opening. Right: result of cross-validation of classification showing that $K = 3$ gives the best classification.

It can be seen from Figure 7.6 (left) that vibrations are actually found for specific regions of $V_r$ in combination with certain gate opening ranges. While obviously the limited amount of
data does not give the full picture, it is for example shown that vibrations in the region \(2 < V_r < 3.5\) vanish when the gate is opened to 1.4 times the gate thickness and higher. The same cannot be concluded from the available data for vibrations detected at \(V_r > 7\).

To classify this data, it was first nominalized to scales from 0 to 1. A ten-fold cross-validation was then done on the value of \(K\) according to the algorithm of Section 7.3.2. This showed that \(K = 3\) is the choice that gives the minimal error in this case, as plotted in Figure 7.6 (right). The code that was used is an adaptation from Matlab scripts by Rogers and Girolami (2012). This KNN algorithm with \(K = 3\) is now ready to be used as an automated tool for tracking critical gate conditions. The contours of vibration regions are represented by classification decision boundaries, up to a certain accuracy. The purpose of this technique has overlap with that of physical model tests, but it automates and hence accelerates the identification of significant dynamic response regions.

### 7.5 Application challenges

#### 7.5.1 General

Some remarks on the proposed system:

- The system outline so far assumed a unique response for one gate opening and \(V_r\) value. In other words, the classification in Sections 7.3 and 7.4 was based on the assumption that the function \(f: (a/D, V_r) \rightarrow A/D\) is well-defined. But many dimensionless parameters from Section 2.4.2 are missing here. A more easily measured alternative to \(Fr\) is the submergence \(Cs\) defined in Section 5.5. Including this parameter as a classification label means including the effect of the varying free surface. Also, the amplitude is intimately linked to damping. Including \(Sc\) or \(\zeta\) does not make sense, however. It will be practically impossible to monitor these in prototype and besides, a single structure will experience very similar hydrodynamic damping for one hydraulic condition each time. Small changes in time of structural damping will be captured automatically by changes in the self-learning model. The same holds for all structural parameters that gradually change with time as the structure is aging. In conclusion, the ML module works best if the mapping \(A/D = f (a/D, V_r, Cs)\) is indeed well-defined. So we end up making a classification in three-dimensional space, which is computationally quite the same as what was done in Section 7.4. The experimental data is not suitable for testing 3D classification because it contains very little \(Cs\) variation.

- A challenge for the classification algorithm is that the training set is very asymmetric, because there will always be many more data points for safe situations than for situations with regular high-level vibrations. The experimental data set is in this respect not representative of real-life monitoring data, as the experiment focused on finding and recording vibration regions.

- A second aspect that is rather likely to be different in prototype, is the occurrence of more than one vibration region. There were two regions (fixed by values of reduced velocity and gate opening) in which significant vibrations are found in the experiment, by actively changing stiffness. This does not mean that one region cannot have multiple physical causes (excitation mechanisms) simultaneously.
The system is built with 'online ML', a self-learning process that gradually improves by taking up new sensor data as time progresses. The rate at which the database is filled with data points of critical situations depends on chosen critical limits (threshold) and occurrence of extreme events. Collected vibration data of real barrier gates will at first not be distributed in a way that makes determining the regime boundaries straightforward. In particular, water level variations and gate openings are usually not sufficiently varied during everyday operation to get a complete view on vibration states. Dedicated measurement events in which gate openings are prescribed when specific hydraulic heads occur are useful, if not necessary, to collect specific data to fill in blank spots.

- Every time a gate condition is visited that was already measured before, the database is updated. If the newly visited condition was an unknown point previously simulated by the physics-based model, the new measurements also act as calibration data for this model.

- The strategy parameters of the classification operation are linked with a safety factor that follows from the question of what the consequences are of making an incorrect classification. Optimisation of the threshold value of the binary split and the parameters of the classification algorithm (for the KNN-algorithm this is $K$) should take this into account.

### 7.5.2 Multiple degrees of freedom

The application of one sensor on a gate matches the single d.o.f. (SDOF) schematization, provided the vibration mode is known. The most common gate vibration modes are vertical translation, horizontal translation and rotational motion (for tainter valves), which are all described aptly by the single mass-spring analogy. This was discussed in Section 2.5. Ultimately, the hydro-elastic conditions determine if a certain vibration mode actually initiates. Any prototype investigation should first focus on identifying the primary vibration mode. The use of sensors can facilitate this process and confirm analytical studies. Gates with considerable longitudinal spans can suffer from bending vibration modes (Ishii, 1992), which are found only when additional accelerometers are placed strategically and their signals are analysed jointly. It is remarked by Kolkman and Jongeling (1996) that the rising demand for structures with larger dimensions results in relatively weaker, i.e. less stiff structural elements, thus increasing the risk for vibrations.

There could be more than one significant mode of vibration. Systems of multiple d.o.f. (MDOF) have $n > 1$ distinct eigenmodes and eigenfrequencies (for $n$ d.o.f.s) and are analytically described by a matrix-version of the motion equation:

$$ M\ddot{w} + C\dot{w} + Sw = f, $$

where the capitals are $n \times n$-matrices of mass, damping and stiffness, respectively, $w(t)$ is the response vector and $f(t)$ is the forcing vector. The added mass matrix already mentioned in Section 2.4 has to be included in $M$. An arbitrary damping matrix will cause displacements in one d.o.f. influencing the forces on another: coupling. The solution of (7.1) has the form

$$ w(t) = \sum_{i=1}^{n} \alpha_i(t) \cdot e_i, $$

(7.2)
with \( \mathbf{e} \), the eigenvectors and \( \alpha_i(t) \) 'participation functions' giving the contributions of the force vector to each d.o.f. (Kolkman and Jongeling, 1996).

The relative ease of analytical generalisation from the SDOF mechanical vibrations (Section 2.3) to multiple modes of vibration conceals the fact that analytical descriptions of fluid-solid interactions (Section 2.4) for multiple coupled modes are always complex and the physical understanding of MDOF FIV systems is incomplete. The study of gate vibrations of simultaneous cross-flow plus in-flow modes by Billeter and Staubli (2000) illustrates the extra complexity of experimentation. Moreover, they claim that the addition of only one d.o.f. increased the region of conditions in which non-linear response can be expected. This implies that hopes for tackling forward MDOF problems are slim; physics-based numerical modelling for real-life MDOF conditions is practically out of reach. This is another argument in favour of the data-driven approach.

Let us suppose it is not clear what the most stringent modes are and plenty of sensors are installed in an attempt to catch these. Then frequency domain decomposition (FDD) is one way of finding the modes from the collective matrix of all measured signals. Brincker et al. (2001) shows how to apply this for a modal identification using only output signals. This is a relatively easy computation that distinguishes separate vibration modes from a set of signals and can be added to the 'primary data analysis' block.

7.5.3 Applying machine learning in engineering

Even though computational hydraulics research has made use of artificial intelligence from relatively early dates (Verwey et al., 1994), state-of-the-art ML applications are not commonplace for real-life (consultancy) projects in the conservative field of hydraulic engineering. Approaches that bypass process analysis and expert judgement are often met with scepticism. Before a control system such as the one presented in this chapter will be trusted to replace manual operation, it has to be thoroughly tested in pilot projects. Seen from the opposite side, ML is sadly abused frequently as a quick way of producing scientific papers – this has resulted in an uncontrolled growth of superfluous ML ‘tricks’ that will never be picked up by others, let alone see any form of application. Attempting to reverse this trend, Wagstaff (2012) presents two necessary conditions for producing ‘ML that matters’: (i) instead of testing new algorithms on isolated benchmarks, choose meaningful and relevant domain tests; and (ii) communicate the results back to the problem's research field.

A numerical model in ML is trained on a dataset by considering errors between predicted and actual data points. To prevent the model from becoming a 'one-trick pony' that performs poor on other data of a similar problem, it is custom to split the dataset into a training set and a test (or validation) set. The trained model is then applied to the test dataset to produce a test error (or validation error), which is a better measure for how the model will actually perform on new problems (in the same domain). However, as Wagstaff (2012) points out, the fact that acceptable levels of test errors greatly vary over application domains is often overlooked.
7.6 Conclusions

This chapter has outlined and partly tested a data-driven approach to vibration characterisation for implementation in an operational system that controls the gate position such that dynamic excitations are avoided. Gate response data from sensors undergo a classification with the goal of predicting whether a future gate state is safe or not. The presented ML application fulfills Wagstaff’s (2012) conditions and included cross-validation. The test result suggests that in the KNN algorithm only a small number of neighbours should be used to determine the assigned class. These straightforward computations reflect a proof of concept rather than an optimised algorithm choice. There are many refined alternative classification methods available.

The proposed binary classification of the vibration regions in the \((V_r, a)\)-plane or \((V_r, a, C_s)\)-plane is an automated way to recognise transition characteristics between stable and unstable dynamic behaviour. It is reasonable to assume that the contours between safe/unsafe classes correspond to transitions from irregular to regular oscillations (on a signal level) and to a transition from net energy dissipation to net energy transfer from the flow to the gate motion (on a physics level). Ideally, the sensors are sensitive enough to record these transitions and thus produce training points of the unsafe state without actually entering states that are potentially dangerous for the structure.

Despite the fact that physics-based simulations lead to better insights into the gate response in unmeasured future conditions than the status quo of complete uncertainty faced otherwise, the incorporation of these simulations in an operational control loop seems inappropriate. They are too time-consuming and require careful interpretation that is hard to capture in an algorithm. Their value lies in analysis and possibly in offline generation of complementary input data for the classification database.

Finally, this system can also be used for stand-alone field monitoring, namely when the water level prediction, decision and operation blocks are disabled. As such it is a useful aid in analyses by tracking the conditions under which vibrations occur. This knowledge is useful for improving a flawed gate design.