20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Emulation modelling of salinity dynamics to inform real-time control of water quality in a tropical lake

S. Caietti-Marin [a], **S. Galelli** [b], **A. Castelletti** [a], **A. Goedbloed** [c]

[a] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy.*
[b] *Pillar of Engineering Systems & Design, Singapore University of Technology and Design, Singapore.*
[c] *Singapore-Delft Water Alliance, National University of Singapore, Singapore.*
Email: andrea.castelletti@polimi.it

Abstract: Emulation modelling has been successfully applied in many environmental applications to reduce large, computationally demanding, process-based models to low order surrogates to be used in place of the original model in problems involving hundreds or thousands of model simulations. Typical examples include optimal planning and management, data assimilation, and sensitivity analysis.

In this study, we describe the identification of a dynamic emulator of a 3D hydrodynamic reservoir model and its subsequent use within a real-time control framework for dam operation. In particular, we adopt a novel data-driven approach that combines the many advantages of data-driven modelling in representing complex, non-linear relationships, but preserves the state-space representation typical of process-based models, which is particularly effective in designing the controller.

The approach is demonstrated on Marina Reservoir, Singapore, which was recently reclaimed to the sea and transformed into a freshwater storage by constructing a barrage. A dynamic emulator of the salinity evolution in a control point near the dam was identified and then used in combination with Model Predictive Control to design the real-time operation of the barrage. Results show that the salinity levels, due to saline intrusion through groundwater seepage, can be dropped to drinking water standards by embedding the emulator in the real-time controller.

Keywords: *Dynamic emulation modelling, data-driven models, process-based models, water reservoirs opera-tion, water management*

1 INTRODUCTION

Process-based models are usually adopted to characterise time and space variability of hydrodynamic and ecological processes in lakes and reservoirs under the effect of hydro-climatic forcing and human activities. Unfortunately, these models are hardly usable in those problems for which hundreds or thousands of model evaluations are required, e.g. to inform decision-making and water resources system operation. Dynamic Emulation Modelling (DEMo, Castelletti et al., 2012) has been recently adopted in a number of applications to alleviate the computational burden associated to the conjunctive use of process-based models and dynamic optimisation algorithms (e.g., Castelletti et al., 2012; Xu et al., 2013). A dynamic emulator is a low-order, computationally efficient model identified from the original large process-based model and then used to replace it for computationally intensive applications. Literature shows a variety of emulation modelling approaches in water resources operation problems (Razavi et al., 2012), but, generally, they deal with relatively simple models (e.g., Chaves and Kojiri, 2007), very often implementing ad hoc solutions. Conversely, hydrodynamic and ecological processes in lakes and reservoir are generally described by complex and large 3D process-based models, able to accurately characterize processes that vary in time and space domain. Moreover, while for some applications a non-dynamic emulator is generally sufficient (e.g., Castelletti et al., 2010), for decision-making problem (e.g., optimal control) the emulator must be dynamic, that is it must reproduce the main system trajectories over any specified horizon.

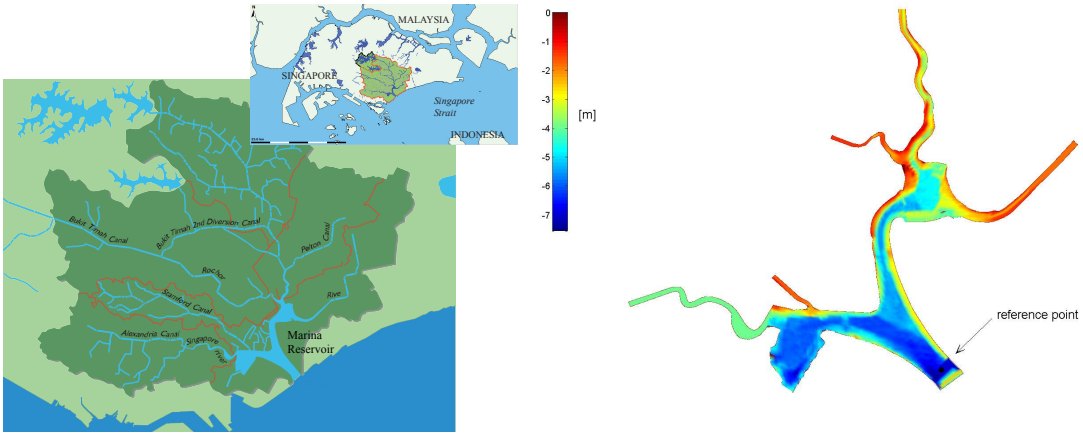
This work describes the application of an emulation modelling technique (Castelletti et al., 2012) to a 3D hydrodynamic reservoir model, in order to identify a data-driven low-order dynamic emulator to be used within a real-time control framework for dam operation. In particular, the proposed data-driven approach exploits the many advantages of data-driven modelling in representing complex, non-linear relationships, but preserves the state-space representation typical of process-based models, which is particularly effective in designing the controller. When applied to large 3D models, any DEMo exercise does require a pre-processing of the exogenous drivers and state variables in order to reduce, by spatial aggregation, the high number of candidate variables to a sub-set of lumped variables among which those appearing in the final emulator will be selected. In this work time series clustering (Magni et al., 2008) is adopted to identify spatial structures by objectively organizing data into homogeneous groups. The identified clusters are then processed with a recursive variable selection algorithm in order to single out the most relevant clusters in explaining the emulator's output, and the resulting emulator is then used to design a real-time controller.

The approach is demonstrated on a real-world case study concerning the reduction of a large process-based model (Delft3D) describing the hydrodynamic conditions of Marina Reservoir (Singapore): the identified emulation model is used to simulate salt intrusion dynamics within the reservoir, and subsequently coupled with a Model Predictive Control (MPC) scheme to design the real-time operation of the barrage.

2 CASE STUDY: MARINA RESERVOIR

2.1 System description

Marina Reservoir was created in late 2008 with the construction of a barrage that closed the homonymous marina from the sea (see Figure 1a). Five main tributaries discharge water into the reservoir draining a catchment of approximately 100 km^2 that produces a mean annual inflow of about 150 Mm^3. The catchment mainly consists of urbanised land and it is characterised by the presence of three further reservoirs, only managed for drinking water supply and whose discharge to Marina Reservoir is rare and negligible. The main reason for the construction of the reservoir is to increase Singapore drinking water supply through large-scale urban stormwater harvesting, and to ensure flood protection and lifestyle attraction in the surrounding lowland areas (Kristiana et al., 2011). Due to its peculiar location Marina Reservoir shows a number of water quality issues: the inflow is characterized by short bursts of high flow with sediment and nutrient rich water followed by dry periods with almost no flow, and because of its location in the tropics, temperature and light intensity are high. This typically leads to eutrophic in-reservoir water conditions (Antenucci et al., 2013). Moreover, the relatively recent formation of the impoundment from a former estuary makes salinity control another important objective: since the permanent closure of the barrage in April 2009, salinity have dropped to typical fresh water levels in many areas of the reservoir. However, some residual sources of salinity persist in the form of groundwater seepage driven by the water level gradient between the sea and the reservoir during high tide. This can result in high salinity in the deepest parts of the reservoir close to the barrage, and during prolonged



(a) (b)

Figure 1. Marina Reservoir water system (a) and bathymetry (b).

dry periods salinity increases also in the other parts of the reservoir through diffusion and mixing.

The barrage can be operated by actuating 9 surface gates and 7 pumps with a hourly decision time-step. Surface gates are used during low tide events, while pumps, with a total installed capacity of 280 m^3/s, can discharge water during high tide, when the sea water level is higher than the reservoir level. Moreover, a further pumping station is employed at the drinking water intake to pump water to the upstream treatment facilities. There are also 2 bottom pipes, used during low tide conditions to discharge saline water and to control temperature profiles within the reservoir. The reservoir regulation is aimed at meeting the drinking water demand, taking into proper consideration the issues of floods control around the lake shores and the energy costs due to the pumps usage during high tide events (Galelli et al., 2012).

2.2 Models available

For Marina Reservoir a set of models is available for simulation, control and forecasting. This modelling framework consists of three different modules (Twigt and Burger, 2010):

- Rainfall-runoff and 1D flow module (Sobek) describing the runoff generation from the different sub-catchments and the flow routing through the stormwater infrastructure;
- 3D Hydrodynamic model (Delft3D) that describes the hydrodynamic processes of Marina reservoir and transport phenomena due to the hydro-meteorological forcings (see Figure 1b for a representation of the reservoir bathymetry);
- Real-Time Control (RTC) module, which regulates the barrage hydraulic infrastructures as a function of the different operational objectives.

All the different modules are integrated into a 1D-3D coupled model. Its operation is as follows: the rainfall-runoff module runs independently first. This provides boundary conditions for the 1D and 3D flow modules, which are run together with the RTC one. The 1D and 3D flow coordinates discharge and water level at their combined boundary. The RTC module sets the states of controllable structures (gates, bottom pipes, and pumps) depending on the system state.

3 METHODOLOGY

As anticipated, the efficient management of Marina Reservoir should be capable of accounting not only for water quantity but also for quality targets. In order to incorporate salinity control as an operating objective, a prediction of salinity concentration is required. Unfortunately, Delft3D cannot be directly employed since its computational requirements does not allow for its direct integration within an optimisation framework. This problem can be solved by resorting to emulation techniques, namely by identifying a simple and computationally efficient surrogate of Delft3D.

Table 1. Selected number of clusters for the different sub-sets composing the vector $\mathbf{X}_t$. The symbols are explained in the text.

	sal	$temp$	u	v	w	h
# clusters	5	6	6	5	11	8

3.1 DEMo procedure

Given the process-based model Delft3D, with state vector $\mathbf{X}_t$, control (i.e. release decisions) $\mathbf{u}_t$ and exogenous driver $\mathbf{W}_t$, we applied the DEMo approach proposed in Castelletti et al. (2012) and identified the emulator over the data-set of tuples $\{\mathbf{X}_t,\ \mathbf{W}_t,\ \mathbf{u}_t,\ \mathbf{Y}_t,\ \mathbf{X}_{t+1}\ \}$ generated via simulation of Delft3D. In particular, the model used in this work has 6 state variables for each computational cell: salinity concentration (sal), temperature ($temp$), velocity in the three directions (u, v, w), and water level (h), stored for the 111 observation points (belonging to the model spatial domain) of the 12 computational layers (Delft3D real-to-run time ratio associated to this set-up is of about 100:1). The exogenous driver vector $\mathbf{W}_t$ includes 7 components accounting for the main hydro-meteorological processes, while the control vector $\mathbf{u}_t$ has 3 components, i.e. release from gates, pumps, and bottom pipes. The identification of a dynamic emulator includes the following 6 steps.

Step 1. Design of computer experiments and simulation runs. The purpose of the Design Of Experiments (DOE) is to set up a sequence of simulation runs for the model aimed at generating the data-set, which is subsequently utilized in the emulator identification. The data-set must be as much as possible informative, thus reproducing all possible model dynamic behaviours, forced by the widest spectrum of inputs ($\mathbf{W}_t$ and $\mathbf{u}_t$). Depending on the computational requirements of the utilised model, either statistical techniques (e.g. pseudo-random binary sequences in MacWilliams and Sloane, 1976) or expert-based design (e.g., Galelli et al., 2010) can be adopted.

In our application, as for the exogenous driver $\mathbf{W}_t$, the time-series of observational data over the period April 2009 - January 2011 is available, while, concerning $\mathbf{u}_t$, 10 different management scenarios are generated as pseudo-random sequences by perturbing the control trajectories obtained by operating the barrage while accounting for water quantity operating objectives only. The data are finally sampled with an hourly time-step, and finally stored in a data-set of $\sim 10^5$ tuples.

Step 2. Variable aggregation. As process-based models are spatially-distributed, the space discretization lead to a strong increase in the dimensionality of the state and exogenous driver vectors. In that case, the purpose of the variable aggregation is to transform $\mathbf{X}_t$ and $\mathbf{W}_t$ in two lower-dimension vectors $\tilde{\mathbf{X}}_t$ and $\tilde{\mathbf{W}}_t$ with a suitable aggregation scheme. The aggregation scheme can rely on expert-based skills or on fully automatic techniques, such as clustering algorithms (Liao (2005); Kavita and Punithavalli (2010), and references therein). Eventually, the data-set $\mathcal{F}$ is transformed into the lower-dimension data-set of tuples $\{\tilde{\mathbf{X}}_t, \tilde{\mathbf{W}}_t, \mathbf{u}_t, \tilde{\mathbf{X}}_{t+1}, \mathbf{Y}_t\}$.

In our work, the variable aggregation is performed using a hierarchical time-series clustering algorithm (Magni et al., 2008). To preserve a sort of physical interpretability of the aggregated time-series, the clustering algorithm is not directly applied to the complete data-set produced via simulation of Delft3D, but to six sub-sets, containing the temporal and spatial realizations of the state variables, in order to identify groups of homogeneous areas for each particular variable. The time series algorithm found 5 homogeneous areas for salinity concentration, while for temperature the identified number of clusters is 6. The final results obtained for the remaining sub-sets are reported in Table 1. Eventually, the selected output $\mathbf{Y}_t$ is the salinity concentration in the deepest point of the reservoir, located few hundred meters from the barrage.

Step 3. Variable selection. Based on the information content of $\tilde{\mathcal{F}}$, the process-based model is further simplified by selecting the components of $\tilde{\mathbf{X}}_t$ and $\tilde{\mathbf{W}}_t$ that will constitute the emulator's state $\mathbf{x}_t$ and exogenous driver $\mathbf{w}_t$ vectors. Generally, this operation relies on some automated technique, since $\tilde{\mathbf{X}}_t$ and $\tilde{\mathbf{W}}_t$ are often too large to be handled by a human operator.

In this paper this operation relies on the Recursive Variable Selection - Iterative Input Selection (RVS-IIS) algorithm (Castelletti et al., 2012), a data-driven input selection method that is able to identify the most relevant variables for building an emulator able to accurately reproduce the output values of the original process-based model, but with reduced dimensionality. The selection of the variables to appear in the emulator took two calls of the algorithm: the identified emulator is characterized by a state vector x_t with 1 component (i.e. the salinity concentration in Cluster 1, which is the deepest and the closest to the barrage) and a control vector u_t with 2 components (i.e. the discharges from gates and pipes).

Step 4. Structure identification In this step, the structure of the emulator is first identified and the relevant parameters estimated. As for the model class, the IIS algorithm is here combined with Extra-Trees, whose parameters are set according to Galelli and Castelletti (2013). The performance of the emulator being built is evaluated in terms of coefficient of determination R^2 and Root Mean Square Error (RMSE) (in k-fold cross-validation, with $k = 10$).

Step 5. Evaluation and physical interpretation. Once the emulator has been calibrated, its ability in reproducing the model input-output behaviour is cross-validated on the data-set $\tilde{\mathcal{F}}$. Eventually, this step aims at verifying the emulator credibility by the users' viewpoint. This property can be guaranteed by the emulator physical interpretability, which is based on the analysis of the inputs to the emulator being built.

Step 6. Emulation model usage. Once the emulator has been successfully validated against the data and the user/expert, it is ready to be employed in the solution of the control problem described next.

3.2 Control problem

The optimal operation of Marina barrage calls for the adoption of tools capable of accounting for water quantity and quality targets in a fast-varying hydro-meteorological system. To this purpose, a deterministic MPC scheme (Scattolini, 2009) is considered. At each step, MPC exploits a short-term prediction of the reservoir inflow and tidal conditions at the sea boundary based on the hydro-meteorological information available in real-time and computes the optimal sequence of the decisions for the barrage operation over a finite time horizon (open loop control). At the subsequent step the optimisation is relaunched and the decision updated based on new information (rolling out horizon). The operating objectives for the barrage are i) meeting the drinking water demand; ii) controlling floods on the lake shores; iii) minimizing the energy costs due to the pumps usage during high tide events; and iv) minimizing the salinity concentration in the barrage area. The decision vector $\mathbf{u}_t$ is composed of the release decisions from gates, pumps and bottom pipes, namely the volume to be released/pumped through each actuator.

The MPC control scheme is coupled with a dynamic emulator of salinity dynamics. Indeed, while the water quantity is described through simple mass balance equations, the accurate description of salinity dynamics requires the use of a large 3D hydrodynamic model, which must be substituted with a low order emulator to be combined with optimisation algorithms.

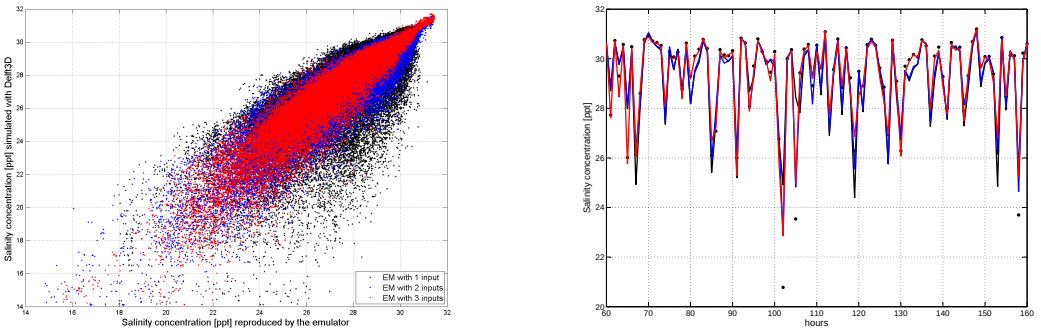
4 Numerical Results

The variables selected to appear in the emulator are the salinity concentration in a deep cluster close to the barrage (located in proximity of the reference point shown in Figure 1), and the discharges from gates and pipes. All these variables can be given a meaningful physical interpretation: the state variable provides information on the reservoir salinity concentration profile in an area close to the barrage. This information is complemented by the controls: the use of pipes is relevant because they are located at the bottom of the barrage and have an obvious key role in releasing water with higher salinity concentrations, while the gates discharge water from the upper layers. The performance of the final emulator is described by means of the coefficient of determination and of the RMSE, respectively equals to 0.91 and 0.6155. Figure 2 shows the comparison between salinity concentration simulated with Delft3D and reproduced by the emulator (3 hours ahead prediction) during the different stages of the variable selection process.

Control policy identification. According to the multi-objective nature of the problem, the resulting control policy must resort to a trade-off between the different objectives. In particular, the control of salinity and floods revealed to be strongly conflicting. Indeed, when limiting salinity concentration is prioritised over flood risks, a higher water level set-point is chosen (see Figure 3a). The rationale is that salinity concentrations are driven by the rate of groundwater intrusion, which depends on the water level difference between the reservoir and the sea: a set-point at the higher end of the range [-0.2; +0.3] m is thus the main way to influence salinity levels. However, a high water level set-point leads to increased flood risks. So, when flood risk prevention has priority, a lower set-point is required to create enough buffer for incoming events (see Figure 3b).

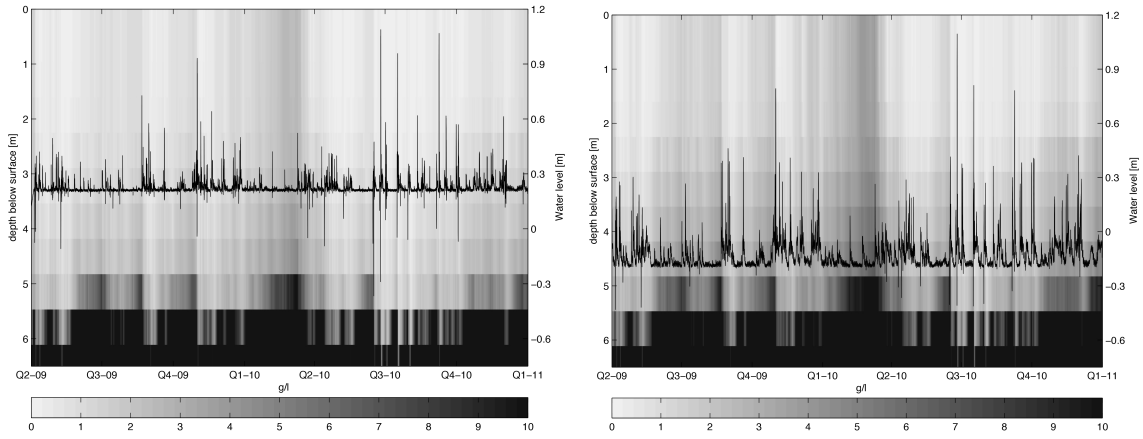
5 Conclusions

This work describes the identification of a dynamic emulator of a 3D hydrodynamic reservoir model and its subsequent use within a real-time control framework for dam operation. The proposed approach is demonstrated on the emulation of a large, 3D model used to simulate the salt intrusion dynamics in Marina Reservoir (Singapore), and then used in combination with a MPC scheme to design the reservoir operation. Results



(a) (b)

Figure 2. Scatter plot (a) and trajectories (b) of salinity concentration simulated with DelftD (black dots) and reproduced by the emulator with one, two and three selected inputs (black, blue and red line respectively).



(a) (b)

Figure 3. Comparison of salinity concentration at different depths in correspondence to a higher (a) and lower (b) water level set-point.

suggest that the reservoir operation can account for both water quantity and quality targets in a fast-varying hydro-meteorological system, and show the possibility of improving system performance by adopting a variable set-point for water level (i.e. raise the set-point when salinity levels are high to mitigate its impact and lower it again when salinity levels drop to prevent flood risks). The main advantage of this approach relies on the possibility to re-apply the procedure to any water quality problem, even though it is necessary to consider that the time scale characterising the water quality processes can strongly vary from case to case.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support and contributions of the Singapore-Delft Water Alliance. The research presented in this work was carried out as part of the Multi-objective Multiple Reservoir Management research programme (R-303-001-005-272).

REFERENCES

Antenucci, J., K. Tan, H. Eikaas, and J. Imberger (2013). The importance of transport processes and spatial gradients on in situ estimates of lake metabolism. *Hydrobiologia 700*(1), 9–21.

Castelletti, A., S. Galelli, M. Ratto, R. Soncini-Sessa, and P. Young (2012). A general framework for dynamic

emulation modelling in environmental problems. *Environmental Modelling & Software 34*, 5–18. 2012a.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2012). Data-driven dynamic emulation modelling for the optimal management of environmental systems. *Environmental Modelling & Software 34*, 30–43. 2012b.

Castelletti, A., F. Pianosi, R. Soncini-Sessa, and J. Antenucci (2010). A multi-objective response surface approach for improved water quality planning in lakes and reservoirs. *Water Resources Research 46*(W06502). doi: 10.1029/2009WR008389.

Chaves, P. and T. Kojiri (2007). Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. *Advances in Water Resources 30*(5), 1329–1341.

Galelli, S. and A. Castelletti (2013). Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. *Hydrology and Earth System Sciences 17*, 2669–2684.

Galelli, S., C. Gandolfi, R. Soncini-Sessa, and D. Agostani (2010). Building a metamodel of an irrigation district distributed-parameter model. *Agricultural Water Management 97*(2), 187–200.

Galelli, S., A. Goedbloed, D. Schwanenberg, and P. van Overloop (2012). Optimal real-time operation of multi-purpose urban reservoirs: A case study in singapore. *Journal of Water Resources Planning and Management*. doi: 10.1061/(ASCE)WR.1943-5452.0000342.

Kavita, V. and M. Punithavalli (2010). Clustering time series data stream - a literature survey. *International Journal of Computer Science and Information Security 8*(1), 289–294.

Kristiana, R., J. Antenucci, and J. Imberger (2011). Sustainability assessment of the impact of the Marina Bay development on Singapore: application of the index of sustainable functionality. *International Journal of Environment and Sustainable Development 10*(1), 1–35.

Liao, T. (2005). Clustering of time series data - a survey. *Pattern Recognition 38*(11), 1857–1874.

MacWilliams, F. and N. Sloane (1976). Pseudo-random sequences and arrays. *Proceedings of the IEEE 64*(12), 1715–1729.

Magni, P., F. Ferrazzi, L. Sacchi, and R. Bellazzi (2008). Timeclust: a clustering tool for gene expression time series. *Bioinformatics Application Note 24*(3), 430–432.

Razavi, S., B. A. Tolson, and D. Burn (2012). Review of surrogate modelling in water resources. *Water Resources and Research 48*(7). doi: 10.1029/2011WR011527.

Scattolini, R. (2009). Architectures for distributed and hierarchical model predictive control - a review. *Journal of Process Control 19*(5), 723–731.

Twigt, D. and D. Burger (2010). Water quality operational management system (WQ OMS), functional and technical design. Technical report, Deltares, Delft, The Netherlands.

Xu, M., P. J. van Overloop, and N. C. van de Giesen (2013). Model reduction in model predictive control of combined water quantity and quality in open channels. *Environmental Modelling & Software 42*, 72–87.