Surrogate modelling for simplification of a complex urban drainage model

Authors:

Mahmood Mahmoodian, Juan Pablo Carbajal, Vasilis Bellos,
Ulrich Leopold, Georges Schutz, Francois Clemens

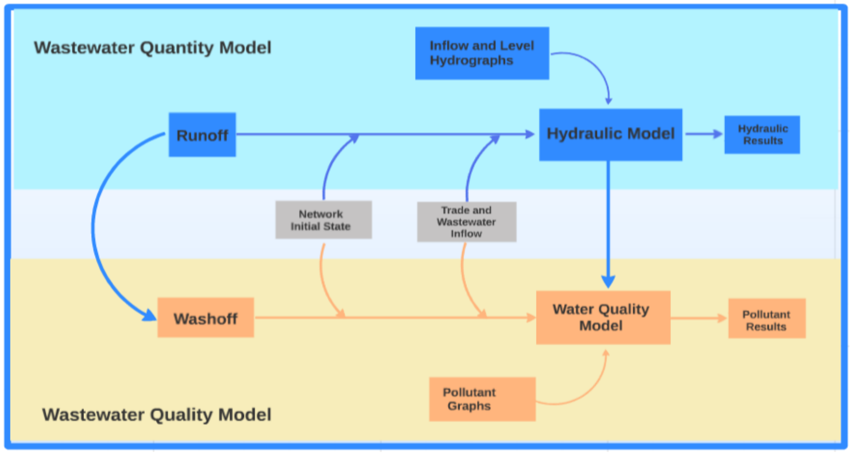
10th World Congress on water resources and environment
5-9 July 2017
Athens, Greece

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.

www.quics.eu

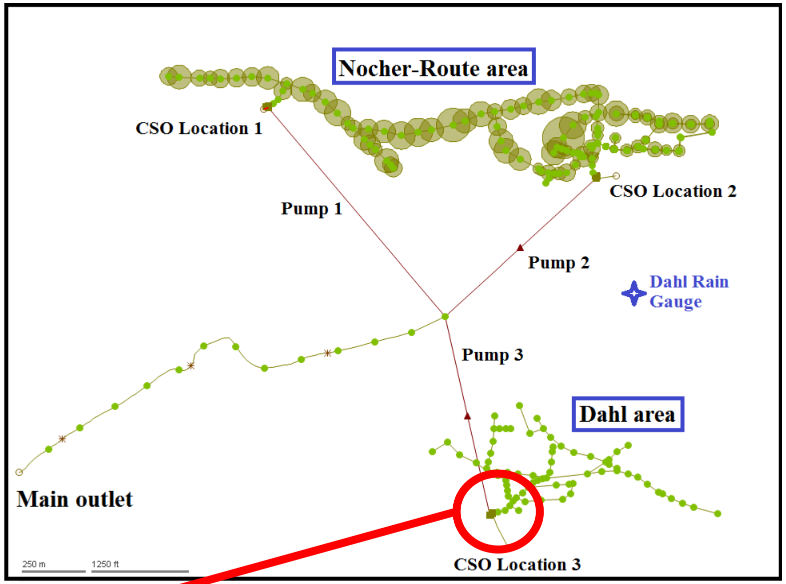
Complex Urban Drainage Models

Detailed Simulator Representing Physical Processes

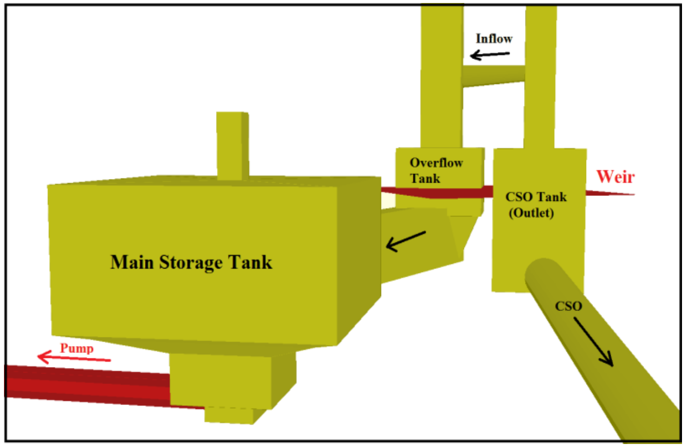


Adapted from **InfoWorks ICM®** help

Detailed Network Structure



Detailed CSO Structure



Zoom in

What is wrong with complex/detailed models?

- Long simulation **run-time**
- Non-linear **optimisation** required, solution might not converge

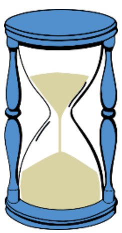
  (e.g. Real-time Control (RTC), Calibration, Structural Optimisation,…)
- **Uncertainty propagation** computationally expensive (for RTC even not feasible!)

Potential solution: surrogate modelling

Note:
Surrogate model = emulator
Complex/detailed model = simulator

Why surrogate modelling?

A solution for model-based real-time control (RTC)



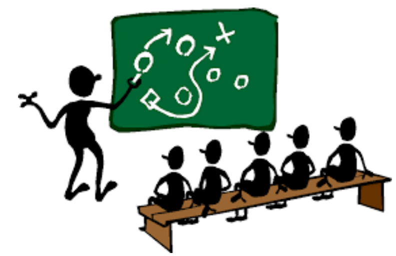
Optimisation in RTC ⟹ Fast (simple) model

General approaches for surrogate modelling:

1) To develop a simple, conceptual model **tailored to RTC.** (e.g. Mahmoodian et al. 2016);
2) To simplify/reduce the **already existing** computationally expensive models to construct the so-called **surrogate models or emulator**. (e.g. Carbajal et al. 2016; van Daal-Rombouts et al. 2016).
3) **Hybrid** method (this study).

Methodology:

The **strategy** for developing the surrogate model or emulator:



a) Identification of the variables to be emulated;
b) Development of a simplified model in which every component contributing to the variables identified in step (a), is replaced by a function;
c) Definition of these functions, which can be ad hoc or based on training data obtained with the detailed simulator; and
d) Validation of the results achieved by the emulator, by comparison with the simulator's results.

- **Step (a):** (In this case study)

Input:
Rainfall (intensity, duration)

→ **Emulator** →

Outputs:
Storage tank volume & CSO volume

Methodology:

Step (b): an intuitive simple **model**, based on **mass balance equation**

***Storage* tank volume**

$$\frac{dV}{dt} = D(t, d_c) + R(t, \alpha, \tau) - P(t, p_c) - C(t, V_{max}, \alpha, \tau)$$

Dry weather flow

Inflow generated by rainfall

Outflow by a pump (or a controllable valve,...)

CSO volume (over the weir)

Step (c): functions

$$D(t, d_c) = d_c d(t)$$

d (*t*): daily pattern unit waveform of wastewater flow;
d_c : a scaling constant (equal to 0. 66L/s in the specific case study).

$$\mathrm{P}(t, P_c) = \begin{cases} 0 & pump\ off \\ 6\ L/s & pump\ on \end{cases}$$

Methodology:

Functions for inflow due to rainfall event and outflow due to CSO event:

- ✓ Learnt from **data** provided by the **simulator** (virtual reality)
- ✓ Synthetic Rainfall **scenarios**: various constant intensities with different durations

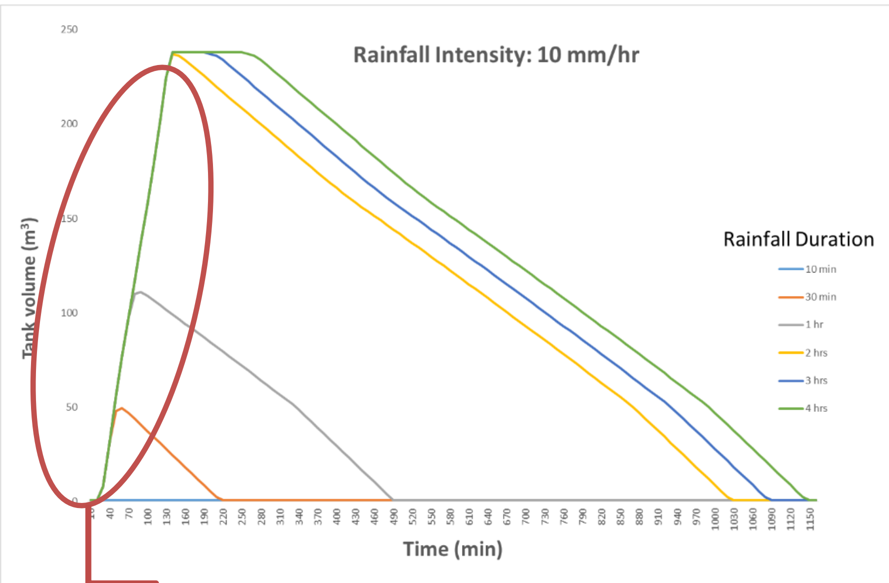


Figure 4. Tank volume change with various rainfall duration and same intensity

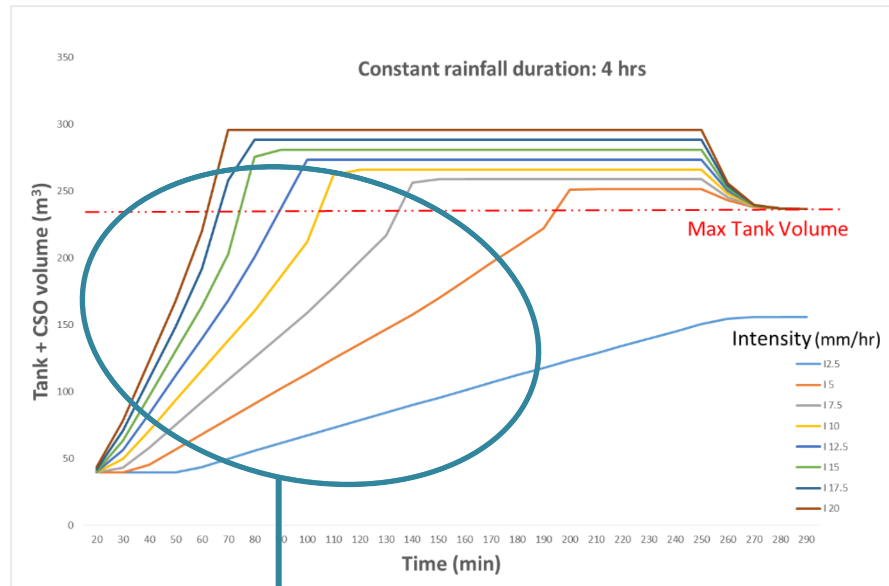


Figure 5. cumulative sum of storage tank volume and CSO volume for various rainfall scenarios with different intensities and constant duration of 4 hours (pump is off)

- ✓ **Tank filling function independent from the rainfall duration.**
- ✓ **R and C, only depend on rainfall intensity and a time lag.**

Methodology:

Inflow by rainfall:

$$R(t,\alpha,\tau) = \alpha r(t-\tau)$$

R: inflow due to rainfall (m^3)
r: rainfall intensity (mm/hr)
α *slope* obtained from the training data
(0.294 for this case)

τ: **lag time**, defined by **cross-correlation** of real rainfall time series and output (τ=30 min in this case).

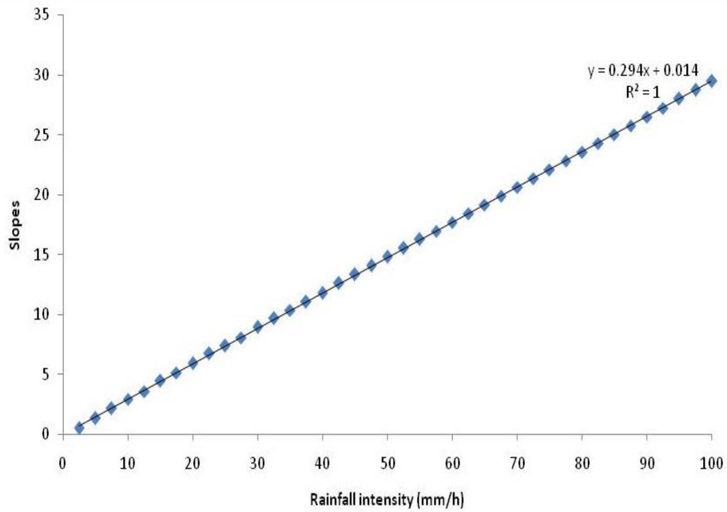


Figure 6. Tank filling curve slope versus different rainfall intensities

Outflow by CSO:

$$C(t,V_{max},\alpha,\tau) = \begin{cases} \alpha r(t-\tau) & if V \geq V_{max} \\ 0 & otherwise \end{cases}$$

C: CSO volume (m^3),
when the storage tank volume reaches the maximum capacity V_{max}.
Calculated from the training data as well.

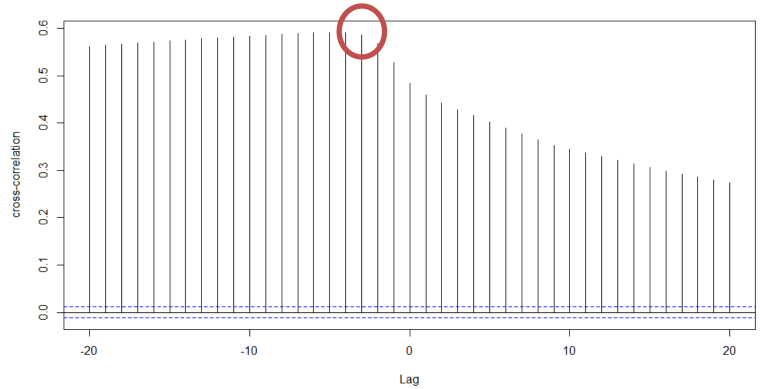
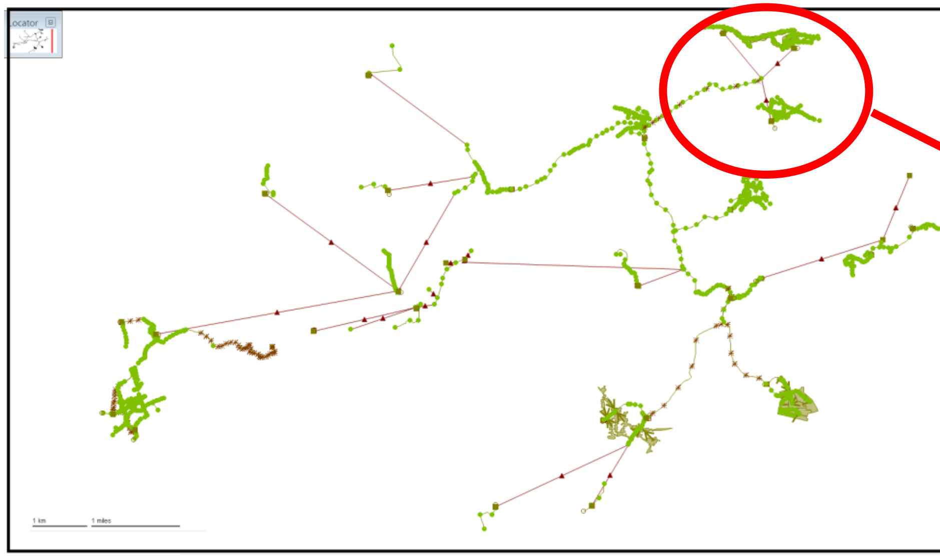


Figure 7. Cross-correlation between real rainfall time series and tank volume

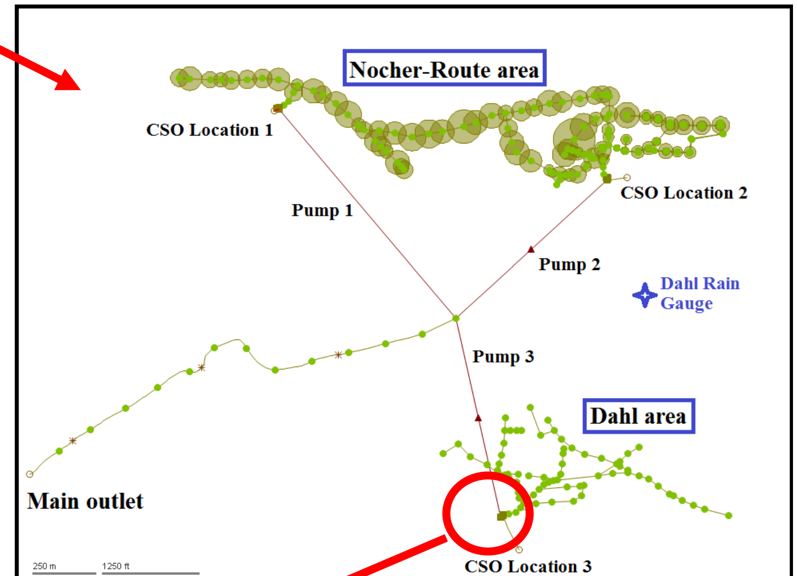
Maximum cross-correlation: 3 lags (30 min)

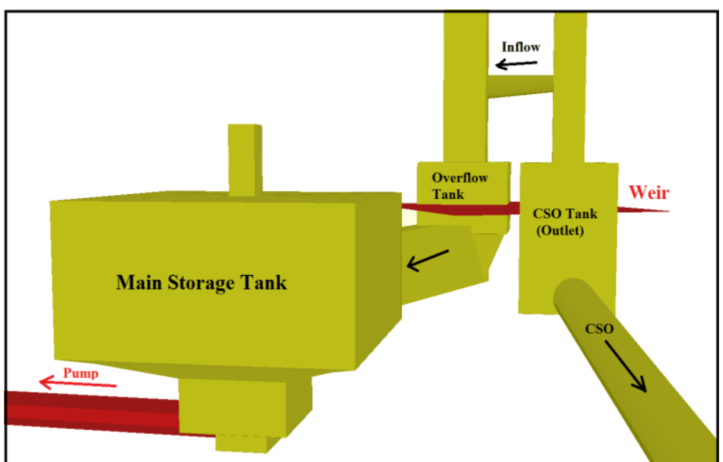
Case Study:

Haute-Sure Catchment, Luxembourg



Nocher-Route-Dahl Sub-catchment





Focus: CSO location

Results and Discussion

Validation (step d):

Real observed (unseen) rainfall events
(October 2007 - December 2009)

Storage tank volume:

- Ascending part (filling)
- Peaks

RTC application

- Descending part (emptying) !

P component simplification?
Artificial base flow in the simulator?

CSO volume:

RTC application

Larger events estimated better
Time period of event occurrence

Uncertainty quantification:

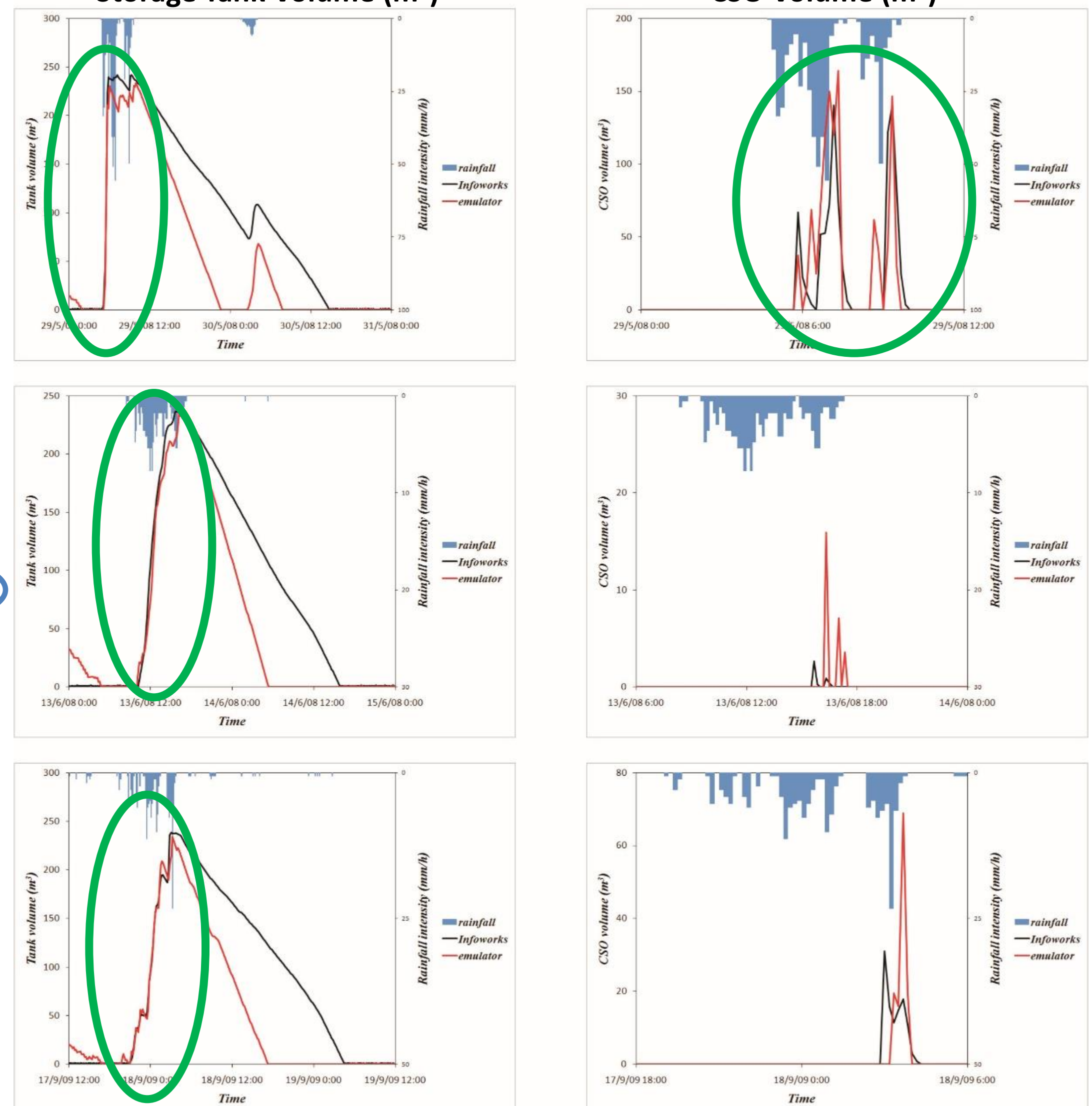
NRMSE = RMSE / (Ymax - Ymin)

- Storage tank volume: 0.072
- CSO volume: 0.002

Speed up:

Emulator about **1300 times** faster than the simulator.

a Storage Tank Volume (m^3) | ***b*** CSO Volume (m^3)



Conclusion: Take Away Messages

1. Although the introduced method was **simple (linear)** for a simple case study, it could turn into a **non-linear problem** in case of more complex networks. Hence, **more advanced** methods are required to solve such problems.

2. Surrogate modelling may **reduce the run-time** significantly, but, in return, it can **decrease the accuracy** of the simulated results as well. Finding a **balance** between the acceptable uncertainty and achieved run-time by surrogate modelling, is inevitable.

3. First, define clearly what the **purpose of your modelling** is; then choose your simulator. For instance, in model-based RTC we do not necessarily need fully-detailed dynamics of the system under study.

Future steps:

- Improvement of the method with more **advanced data-driven** surrogate modelling **techniques** (e.g. Gaussian Process Emulators)
- Quantification, propagation and reduction of **the uncertainty** induced by surrogate modelling.
- Consider waste water **quality modelling** in addition to its quantity modelling to be applied in **RTC application.**

Thank you for your attention

Any questions?

Keep It as Simple as Possible! (Law of parsimony, Ockham's Razor)

References:

- Carbajal, J. P., Leitão, J. P., Albert, C., 2016. Appraisal of data-driven and mechanistic emulators of nonlinear hydrodynamic urban drainage simulators. Environmental Modelling and Software, 92: 17-27.
- Van Daal-Rombouts, P., Sun S., Langeveld J., Bertrand-Krajewski J., Clemens F., 2016. Design and performance evaluation of a simplified dynamic model for combined sewer overflows in pumped sewer systems. Journal of Hydrology, 538: 609–624.
- Innovyze, 2017. InfoWorks ICM. Available at: http://www.innovyze.com/products/infoworks_icm/.
- Mahmoodian, M., Delmont, O., Schutz, G., 2017. Pollution-based model predictive control of combined sewer networks, considering uncertainty propagation. International Journal of Sustainable Development and Planning, 12(1): 98–111.
- Vanrolleghem, P. A., Benedetti, L., Meirlaen, J., 2005. Modelling and real-time control of the integrated urban wastewater system. Environmental Modelling and Software, 20(4): 427-442.

Partners and Acknowledgements

QUICS

The University Of Sheffield.
University of BRISTOL
TUDelft Delft University of Technology
Aquafin
WATERWAYS
RTC4Water Real Time Control of Water Infrastructure
ch2m
WAGENINGEN UNIVERSITY WAGENINGEN UR
RUHR UNIVERSITÄT BOCHUM RUB
Witteveen+Bos
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
Universidade de Coimbra
UNIVERSITÉ LAVAL
CTGA Centro Tecnológico de Gestão Ambiental
TECHNISCHE UNIVERSITÄT DRESDEN
LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST
eawag aquatic research
TUM Technische Universität München
British Geological Survey NATURAL ENVIRONMENT RESEARCH COUNCIL

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.

www.quics.eu