QUICS: Quantifying Uncertainty in Integrated Catchment Studies

Deliverable 6.6 Guidance on trade-offs between model complexity and accuracy

Lead Partner: LIST

Revision: 31/05/18

Report Details

Title: Guidance on trade-offs between model complexity and accuracy

Deliverable Number: D6.6.

Author(s): J. Arturo Torres-Matallana (ESR3); Mahmood Mahmoodian (ESR4); Ulrich Leopold; Vivian Camacho Suarez (ESR9); James Shucksmith

Dissemination Level: Public

Document History

Version	Date	Status	Submitted by	Checked by	Comment
1.0	23/05/2018	First draft	Arturo Torres; Ulrich Leopold	Ulrich Leopold	Initial version
1.1	24/05/2018	Second draft	Mahmood Mahmoodian	Ulrich Leopold	ESR4 contribution
1.2	25/05/2018	Third draft	Arturo Torres; Ulrich Leopold	Ulrich Leopold	Revision
1.3	29/05/2018	Fourth Draft	James Shucksmith		Added material on river modelling. Made comments and suggestions.
1.4	31/05/2018	Final	Arturo Torres; Ulrich Leopold	Ulrich Leopold	Final version

Acronyms and Abbreviations

ADE	Advection Dispersion Equation
BCOD	Chemical oxygen demand load in the overflow
BNH4	Ammonium load in the overflow
COD	Chemical oxygen demand
CCOD	Chemical oxygen demand concentration in the overflow
CODr	Chemical oxygen demand rainwater pollution
CODs	Sewage chemical oxygen demand pollution
CSO	Combined sewer overflow
CNH4	Ammonium concentration in the overflow
MPC	Model predictive control
NH4	Ammonium
NH4s	Sewage ammonium pollution
RTC	Real-time Control

Vchamber	Volume in the combined sewer overflow chamber
Vsv	Spill overflow volume

Acknowledgements

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

Executive Summary

The partners involved in the deliverable D6.6 are the Luxembourg Institute of Science and Technology (LIST), and the University of Sheffield (UoS). This deliverable is a result of the research projects developed by the fellows ESR3 and ESR4 of the QUICS project.

The main objective of this deliverable is to contribute into the Work Package 6 (WP6), which is aim for external dissemination and outreach by presenting a guidance on trade-offs between model complexity and model accuracy.

Understanding model complexity and model accuracy in urban drainage modelling (UDM) is important because decision making for environmental protection requires that the accuracy of model outputs is known and meets pre-defined standards under a specific level of complexity. As a guidance to identify the trade-offs between model complexity and model accuracy accounting for uncertainty propagation (UP) in UDM we propose a **three-step procedure** as follows.

The **first step** is to set-up, within the same case study, at least two definitions of model complexity a so called simplified setting (Level 1) and a complex setting (Level 2). The aim of developing the Level 1 model setting is due mainly to the fact that full hydrodynamic urban drainage models are complex and require highly intense computational budget, which constitutes a constrain when long term simulation or UP analysis by Monte Carlo simulation is required. The Level 2 model can be implemented for short term simulation and UP analysis.

The **second step** can be developed according to two options. The first option aims to develop UP analysis. Statistical uncertainty analysis in UDM is a relatively new subject that largely needs to be developed while very few solid applications have been conducted. Here we establish a specific procedure to perform UP analysis in UDM in the temporal domain for the Level 1 and Level 2 model setups.

The second option consists of developing UP analysis in UDM in the spatio-temporal domain, allowing the possibility of different model resolution to agree with the resolution required by the model itself. Also, to support the change of scale in model output to reach the required by the user. All this done accounting for the associated uncertainties.

The **third step** aims to evaluate how each level of complexity interact according to the UP procedure and draw conclusions about model complexity and accuracy.

We present the three-step procedure to evaluate the trade-offs between model complexity and model accuracy, and apply it to a case study developed in the sewer system of the Haute-Sûre cathment in North-West Luxembourg.

CONTENTS

1 Introduction

1.1 Partners involved in the Deliverable

The partners involved in this deliverable are the Luxembourg Institute of Science and Technology (LIST) and the University of Sheffield (UoS). This deliverable is a result of the research projects developed by the fellows ESR3 and ESR4 of the QUICS project.

1.2 Deliverable objectives

1) Propose a procedure to evaluate the trade-offs between model complexity and model accuracy in UDM.
2) Illustrate the proposed procedure to evaluate the trade-offs between model complexity and model accuracy in UDM.

2 Proposed procedure for evaluating model complexity and accuracy in UDM.

To contribute into the WP6, which is aim for external dissemination and outreach, we present in this deliverable a guidance on trade-offs between model complexity and model accuracy, composed by a three step procedure.

2.1 Step 1: Model complexity definition

2.1.1 Level 1: Simplified or surrogate urban drainage system models.

Simplified models are commonly used in UDM for speed and convenience. Physically based simplified models may neglect a number of secondary processes, or operate at reduced dimensionality. For example, 1D pollutant transport and mixing models are commonly used to represent dispersion processes which arise from cross sectional averaging turbulent diffusion processes (Camacho et al 2017). Alternately “surrogate models” or “emulators” can be developed which aim to reproduce the behaviour of complex deterministic models using simpler mathematical or statistical functions. In general, there are four main strategies to develop so-called “surrogate models” or “emulators” (Asher, Croke, Jakeman, & Peeters, 2015):

1) Data-driven approach, in which the complex model is approximated through a statistical model which captures the input-output mapping of the original model.

2) Projection-based approach, in which the dimensionality of the parameter space is reduced by projecting the governing equations onto a basis of orthonormal vectors.

3) Hierarchical or multi-fidelity approach, where the surrogate is developed, for example, by ignoring some of the processes which are less relevant or by reducing the numerical resolution.

4) Hybrid approach, with combination of above methods.

In this project, we have investigated three approaches from categories 1, 3 and 4. A summary of each approach is presented in the following paragraphs.

Multi-fidelity mechanistic model. Given the current limitations that common complex software used in UDM face regarding the high computational budget required to perform long term simulation of water quality variables and to perform uncertainty propagation analysis by Monte Carlo simulation, we develop and implemented a simplified mechanistic sewer system model to simulate combined sewer overflow (CSO) called EmistatR (Emissions and Statistics in R for Wastewater and Pollutants in Combined Sewer Systems) (Torres-Matallana et al., 2018). This simplified mechanistic model is based in mass balance equations that describe the hydraulic dynamics in the CSO Chamber (CSOC) and the CSO spill volume. For water quality characterisation of the sewage are used two main variables: the chemical oxygen demand (COD) and ammonium (NH4). Load and concentration of these variables in the CSO are computed in EmiStatR. The model is implemented as an R-package (https://CRAN.R-project.org/package=EmiStatR), which allows a seamless integration with existing packages for e.g. time series analysis and sensitive analysis, and new routines for e.g. Monte Carlo simulation and temporal and spatio-temporal uncertainty propagation of model input.

Data-driven surrogate model. The challenge of this approach is to apply a data-driven Gaussian Process Emulator (GPE) technique to develop a surrogate model for a computationally expensive and detailed urban drainage simulator. The novelty is the consideration of (short) time series for the simulation inputs and outputs. Such simulation setup is interesting in applications such as Model Predictive Control (MPC) in which numerous, fast and frequent simulation results are required. Here, an emulator is developed to predict a storage tank's volume in a small case study in Luxembourg. Three main inputs are considered as the GPE's parameters: initial volume in the tank, the level in which the outlet pump of the tank must start to work, and the time series of expected rainfall in the upcoming 2 hours. The output of interest is the total volume of the storage tank for the next 24 hours. A dataset of 2000 input-output scenarios were produced using different possible combinations of the inputs and running the detailed simulator (InfoWorks® ICM). 80% of the dataset were applied to train the emulator and 20% to validate the results. Distributions of Nash-Sutcliffe efficiency (NSE) and Volumetric Efficiency (VE) were produced as indicators for quantification of the emulation error. Based on the results, it can be concluded that the introduced technique is able to reduce the simulations runtime significantly (300 times faster in this specific case), while imposing some inevitable accuracy cost. However, more investigation is required to validate the more generic applicability of this technique for multiple outputs and interactions between different urban drainage components. Figure 1 illustrates three random example validations using the emulator in comparison with the detailed simulator (InfoWorks ICM®).

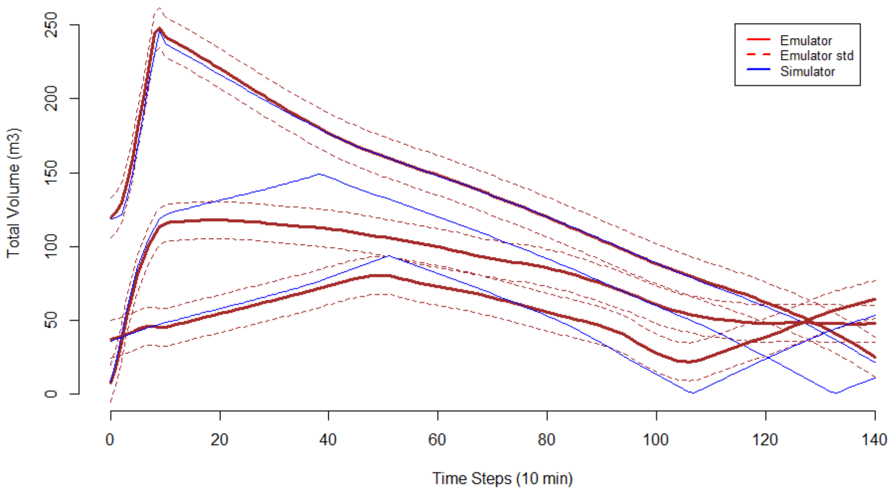

Figure 1. Comparison of emulator vs. simulator results for three random sample scenarios from validation dataset.

Note: this research will be presented at the UDM 2018 conference (September 2018).

Hybrid surrogate model. The focus of this approach is to present a rather simple surrogate modelling or emulation strategy to simplify and accelerate a detailed simulator, and make it available for RTC in our future studies. Hence, only the inputs and outputs of the simulator which are relevant for RTC are considered here. The proposed surrogate modelling strategy includes: a) identification of the variables to be emulated; b) development of a simplified conceptual 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 data-driven or ad-hoc (model-driven); and finally, d) validation of the results produced by the surrogate model in comparison with the original detailed simulator. Herein, a detailed InfoWorks ICM® simulator was selected for surrogate modelling. The case study area is a small urban drainage network in Luxembourg. A simple emulator was developed to map the rainfall time series, as input, to a storage tank volume and combined sewer overflow (CSO) in the case study network. The preliminary results show that the introduced strategy provides a reliable method to simplify the simulator and reduce its runtime significantly. For this specific case study, the emulator was approximately 1300 times faster than the original detailed simulator. For quantification of emulation error, an ensemble of 500 rainfall scenarios with one month duration is generated and the results produced by the emulator is compared with the ones produced by the simulator. Finally, distribution of Nash-Sutcliffe efficiency (NSE) between the emulator and simulator results for storage tank volume and CSO flow predictions was presented as an indicator of the emulation error. Figure 2 shows an example validation of the emulator vs. detailed simulator for a one-year-long simulation.

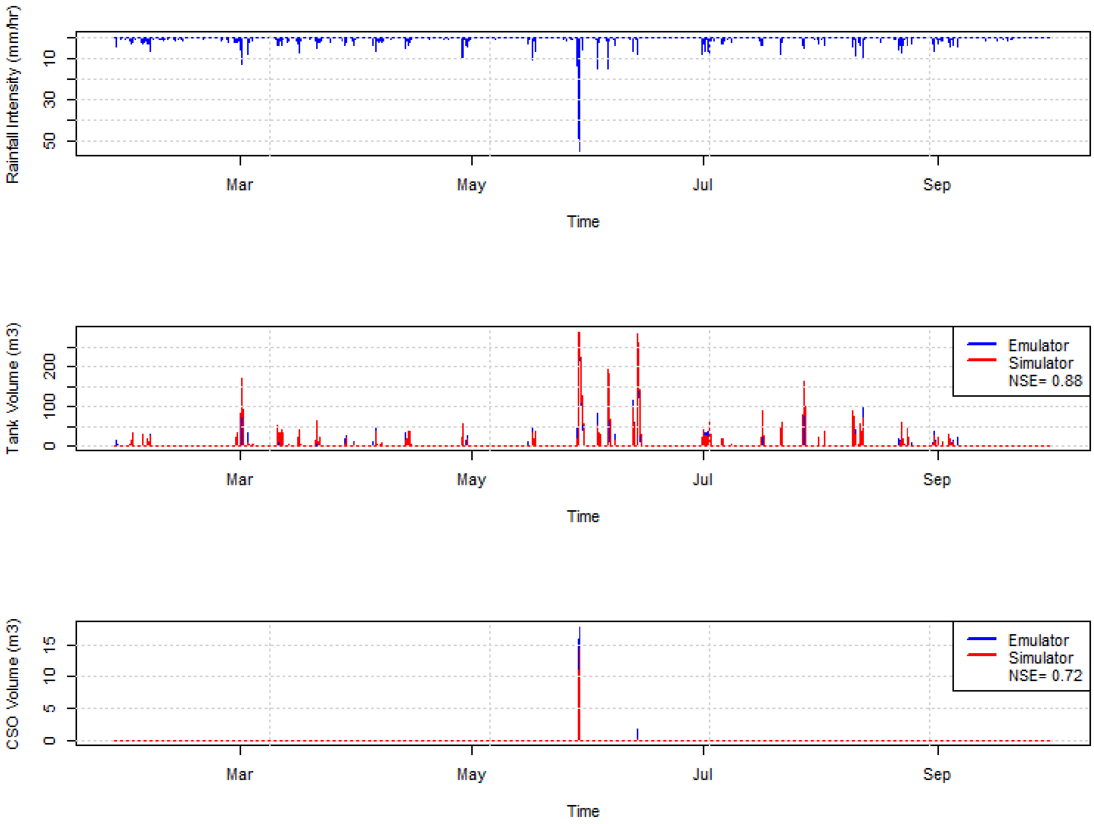

Figure 2. Validation of emulator vs. detailed simulator for simulation of total tank volume and CSO volume for one-year duration.

Note: An earlier version of this research has been presented at EWRA 2017 Conference. The abstract can be found at: http://www.ewra.net/ew/pdf/EW_2017_57_41.pdf. The latest version of the research was submitted to the Water Resources Management Journal in January 2018 (under review).

2.1.2 Level 2: Complex urban drainage system model.

InfoWorks ICM® is a typical example of highly detailed software which are commonly used for modelling urban drainage systems and receiving waters. 198 different parameters and numerous processes are involved in this software which makes it computationally too expensive to be applied applications such as UP or RTC. Figure 3 shows only the main elements of InfoWorks ICM® and the involved modules.

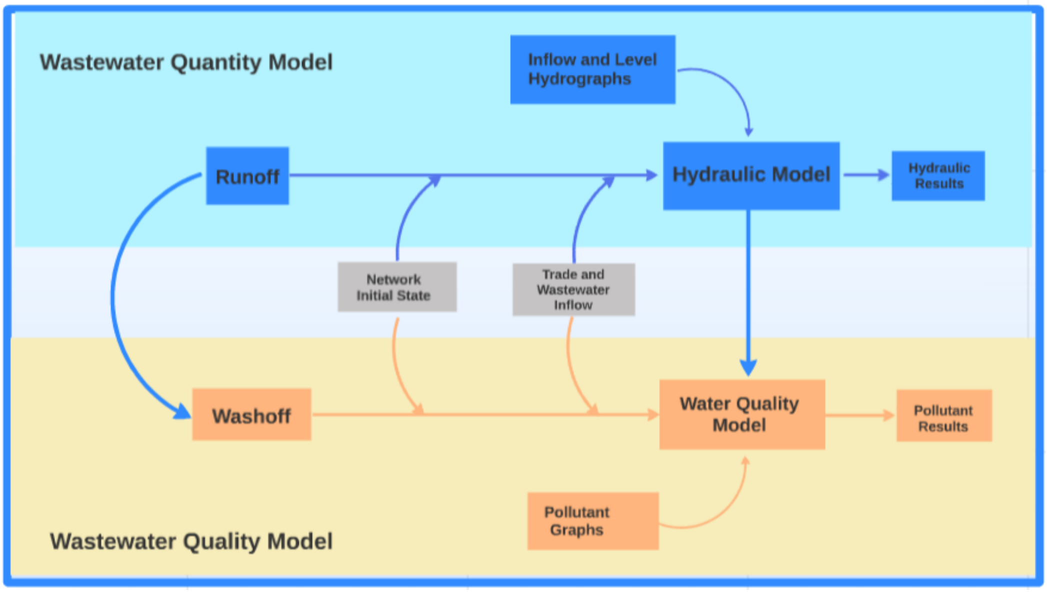

Figure 3. InfoWorks ICM® Model Structure (adapted from InfoWorks ICM® help)

The upper part of the Figure 3 (blue) is for wastewater quantity modelling and the lower part (orange) is showing the corresponding elements in wastewater quality modelling. The results of the quantity model (e.g. runoff and hydraulic model) are used as the input of the quality model, but the other way round is not true. The elements in the middle part (white) are common for both quantity and quality modelling.

For the runoff modelling in InfoWorks ICM® it is possible to select among 15 types of runoff volume models and 13 types of runoff routing models. Each of these models would require their own specific parameters and inputs. The hydraulic model is based on Saint-Venant equations for conservation of mass and momentum. The rainfall (the input of this sub-model) can be in forms of observed (recorded) or design rainfall.

Wastewater quality modelling is more detailed in InfoWorks ICM®. Four different sources of determinants inflow into the model are considered including: 1) wastewater event from domestic areas; 2) Trade waste event from industrial areas; 3) pollutant graph for specific inflows; and more importantly 4) Surface pollutant modelling. For the latter case, we have for example: 1) Wash-off model for sediments and attached pollutants (build-up and wash-off); and 2) Gully pot model for dissolved pollutants (build-up and wash-off).

The results of our research show that with a similar level of accuracy, the Level 1 models represents adequately the hydraulic dynamics in the CSOC and the CSO spill volume, and the load and concentration of COD and NH4 released to the environment along CSO spill events for a case study in the North-West of Luxembourg, the Haute-Sûre catchment.

2.2 Step 2: Input uncertainty propagation and model accuracy evaluation

This step comprises two options for developing a model input uncertainty propagation (UP) analysis and the quantification of the model accuracy. These options are described as follows:

2.2.1 Option 1: Uncertainty propagation in the temporal domain.

Monte Carlo technique is used to perform model input uncertainty propagation in the temporal domain. Three steps are followed to perform Monte Carlo model input uncertainty propagation analysis in the temporal domain through the model (Level 1 or Level 2). The first step is to define and evaluate model input uncertainty, the second step is the Monte Carlo simulation to propagate model input uncertainty to model output, and the third step is the computation of the contributions of each model input variables to the total uncertainty.

1. Evaluation of model input uncertainty.

Following Nol et al. (2010), not all model inputs can be taken into account in the Monte Carlo uncertainty propagation analysis because of the large computational budget required. Only those inputs that have a large uncertainty and to which the model is sensitive should be included, which reduces the number of model inputs analysed. The selection of model input for uncertainty quantification is based in the identification of model input and their level of uncertainty (low or high) and the level of model sensitivity (low or high). The level of uncertainty of the inputs can be defined by expert judgement, literature research, measurements of different model inputs in the sewer system, and interviews with experts. The level of model sensitivity can be derived by interpreting the model structure and components, interviews with experts, and model runs. The main task to quantify input uncertainty is to define the probability distribution function (pdf) that represents the uncertainty of the variable chosen. The uncertainties of selected model inputs can be characterized with pdfs following Heuvelink et al. (2007).

The results of sensitivity of the model output to model input in the Haute-Sûre catchment case study, show that the variables precipitation (P), impervious area (Aimp), pass-forward flow (Qd) to the wastewater treatment plant (WwTP), and volume (V) of the CSO chamber (CSOC) are the most sensitive variables for the output water quantity variables (water volume in the chamber, Vchamber; CSO spill volume, Vsv; and CSO spill flow QSv). Regarding water quality in terms of COD the input variables COD load in the sewage (CODs), COD concentration in the runoff (CODr), Aimp, Qd, V, and P have the greatest impact on output CSO COD load and concentration. The input variables water consumption (qs), NH4 load in the sewage (NH4s), infiltration flow (qf) in the sewer system, NH4 concentration in the runoff (NH4r), Aimp, population equivalents (pe), Qd, V and P have the greatest impact on output CSO NH4 load and concentration. After evaluation of the model output sensitivity to input variables, and taking into account the degree of uncertainties of each input, we selected four input variables to be included in the uncertainty analysis: P, CODs, NH4s, and CODr.

The field measurements were the basis to characterise input uncertainty of CODs and NH4s. Samples of COD and NH4 in milligram per litre were analysed in the dry weather flow produced in the villages of Goesdorf, Kaundorf and Nocher-Route. Regarding CODr, no field measurements were available. Thus, expert judgement and values from the literature were used to characterise input uncertainty in CODr. For all three input variables, CODs, NH4s, CODr, we proposed a normal distribution to characterise input uncertainty.

In order to avoid negative values, the variables were transformed by taking their natural logarithm. In the case of CODs or NH4s, for uncertainty propagation it is possible to simulate these variables by an autorregresive order one AR(1) model when no cross-correlation is considered. However, a more realistic model can be proposed by implementing a multivariate or vector autoregressive order one model VAR(1), which takes into account cross-correlation different to zero between variables.

Regarding the characterisation of precipitation uncertainty, due that P time series are highly skewed due to many zeros, it is required to apply a different approach for characterising uncertainty. We propose a multivariate autoregressive modelling and conditional simulation of precipitation time series (Torres-Matallana et al., 2017). This method, is suitable to simulate precipitation time series in a target catchment given a known precipitation time series in a second nearby location outside the catchment, while accounting for the uncertainty that is introduced due to spatial variation in precipitation.

2. Monte Carlo simulation to propagate model input uncertainty to model output.

The Monte Carlo method runs the model repeatedly, each time using different model input values, sampled from their probability distribution. The method thus consists of the following steps:

1. Repeat N times:

(a) Generate a set of realisations of the uncertain model inputs

(b) For this set of realisations, run the model and store the output

2. Compute and store sample statistics from the N model outputs.

Here, N is the number of Monte Carlo runs, i.e. the Monte Carlo sample size. Common sample statistics that measure the uncertainty are the standard deviation and the width of prediction intervals, which can be easily calculated from the N Monte Carlo outputs.

We made a deterministic run of the model. Additionally, we performed 1,500 Monte Carlo simulations allowing P, CODs, NH4s, and CODr as stochastic input variables with characteristics as defined in the previous section. In this way the total uncertainty of output variables due to input uncertainty was calculated. In this way, we defined the model output and the total uncertainty band of 5 and 95 percentile for CSO load and concentration of COD, and CSO concentration and load of NH4. In case of a rain event that produces a CSO, the uncertainty in the model output is quite large. Also, there is a systematic difference between the deterministic and the median run. The latter is always slightly above the deterministic run.

3. Contributions of each model input variables to the total uncertainty

The contribution of one input variable is calculated as the difference between the total uncertainty and the uncertainty obtained in the stochastic simulation of the other three variables. For instance, the uncertainty contribution of CODs was calculated as the total uncertainty minus the uncertainty of the simulations running only NH4s and CODr in stochastic mode. Therefore, 6,000 additional Monte Carlo simulations were conducted to calculate the uncertainty contribution of the four input variables. We concluded that the most important contribution in the total uncertainty corresponds to precipitation.

2.2.2 Option 2: Uncertainty propagation in the spatio-temporal domain.

Recent practice in urban drainage modelling incorporates characterisation of model input uncertainty in the temporal domain. Previous studies show that rainfall is one important source of uncertainty when uncertainty propagation is performed in the simulation of water volume in the combined sewer overflow tank and the emissions of pollutants to the receiving water body. However, studies often ignore the spatial dimension treating input rainfall as a non-spatially distributed time series, typically originated from rain gauge measurements. Neglecting spatial and space-time distribution of rainfall entering urban drainage systems may result in inaccurate quantification of rainfall and, hence, in substantial uncertainties associated to water quantity and quality predictions. This chapter hast the aim of developing a more realistic characterisation of rainfall as an input to urban sewer models in order to better evaluate its impacts on these predictions.

We developed a space-time model for predict rainfall fields at 10-minute temporal resolution and 500 meters as spatial resolution. We use ordinary global Kriging for prediction of the mean value and variance of rainfall over the entire country of Luxembourg by using 25 rain gauge stations. The region of study is the Haute-Sûre catchment in North-West Luxembourg. Given the mean and the variance maps, we compute rainfall maps for the lower and upper boundary of the 90% confidence interval. The mean, lower and upper boundaries are the main inputs for propagating uncertainties trough an integrated rainfall-runoff and sewer system modelling approach. We compare the deterministic temporal simulations made with a simplified sewer model and a complex mechanistic model with the space-time approach considering model input uncertainty.

To model precipitation fields in the spatio-temporal domain we use the concept of spatio-temporal variogram for ordinary global Kriging (Gräler et al., 2016). We implemented routines using the R package gstat (Pebesma, 2004) for defining spatio-temporal covariance models. Five models are available in gstat: Separable; Product-sum; Metric; Sum-metric; Simplified sum-metric. The method for Kriging prediction in space and time, considers the definition of one covariance model (or variogram) for the space domain and one covariance model (or variogram) for the time domain. We used the Sum-metric model for representing the covariance.

In order to calculate the theoretical spatio-temporal variogram, we selected a one-day period where the cumulative precipitation of the time series is maximum, retrieving a precipitation event in all stations. Upon the definition of the sum-metric model for predicting in space and time the rainfall fields, we proceed to compute the mean and variance maps for the Haute-Sûre catchment. We computed the lower boundary of the 90 percent confidence interval using the kriging mean minus twice the root squared kriging variance, and the upper boundary was computed as the kriging mean plus twice the the root squared of the variance. The prediction temporal interval is 10 minutes and the spatial resolution corresponds to a squared grid of 500 m per 500 m.

The rainfall fields were used as model input for an integrated rainfall-runoff and urban drainage model. Figure 4 illustrates the work flow for the integrated rainfall-runoff and sewer system modelling.

1 *Rainfall-runoff model*.

The rainfall-runoff model used is itzï (Courty et al., 2016). Itzï is a numerical model written in Python, and can be used to simulate surface flows induced by intense rainfall in the urban domain. The model is integrated into the open source GIS software GRASS, which allows a seamless integration of geospatial data as input for the model and model output in the native GRASS format for spatio-temporal raster datasets. Itzï uses an explicit finite-difference scheme to solve the simplified partial inertia shallow-water equations described by De Almeida et al. (2012) and De Almeida and Bates (2013). Besides rainfall maps stored in GRASS GIS as spatio-temporal raster datasets (strds) in [mm/h], itzi requires of maps for the coefficient of roughness of Manning [–] and the infiltration rate [mm/h]. Maps for defining the boundary conditions in the computational domain are also required.

The mean precipitation maps together with the lower and upper boundaries of the 90 percent confidence interval were fed into the rainfall-runoff model. The roughness coefficient and infiltration maps were also taken into account. We have chosen the Goesdorf sub-catchment to illustrate the results of the rainfall-runoff model, which corresponded to runoff depth over the land. Then the routing flow through the CSOC outlet was computed.

2 *Sewer system model*.

The sewer system models are used to compute the deterministic CSO spill volume, and loads and concentrations of COD and NH4, based on the point precipitation measured at Dahl station. The Level 1, simplified sewer system model, can be fed with point data of precipitation or runoff volume. The results of the deterministic simulations are compared with the second model, a complex mechanistic model (CMM). The computation from the runoff volume is done only with the simplified model.

- *Deterministic temporal simulation*

In order to compare the space-time approach for computation of the CSO spill volume, and loads and concentrations of COD and NH4, we computed the deterministic simulation only in the temporal domain, i.e. taking into account the point rainfall as measured in the rain gauge Dahl. Figure 5 shows the deterministic temporal simulation with the simplified model and the CMM.

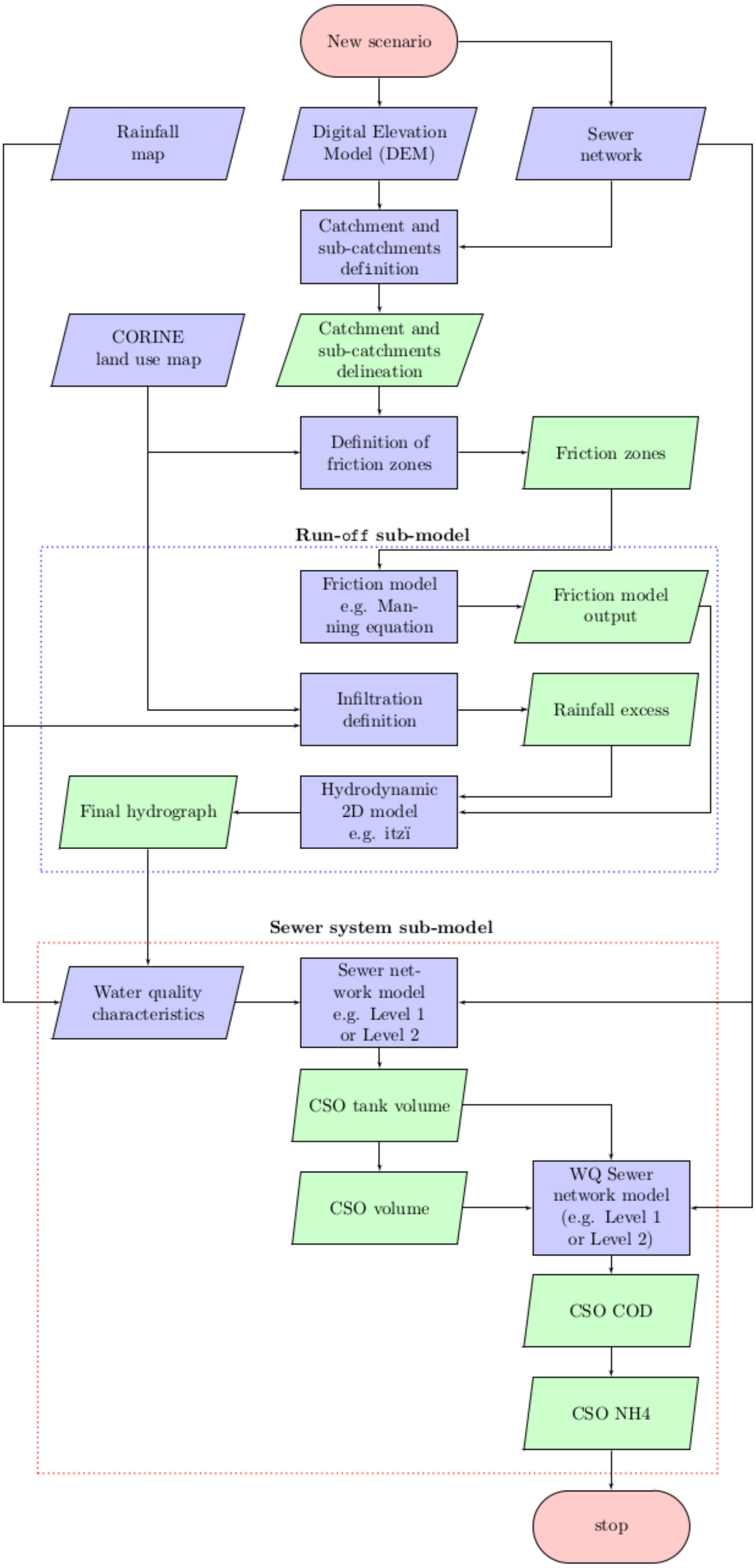

Figure 4. Work flow for the integrated rainfall-runoff and sewer system modelling.

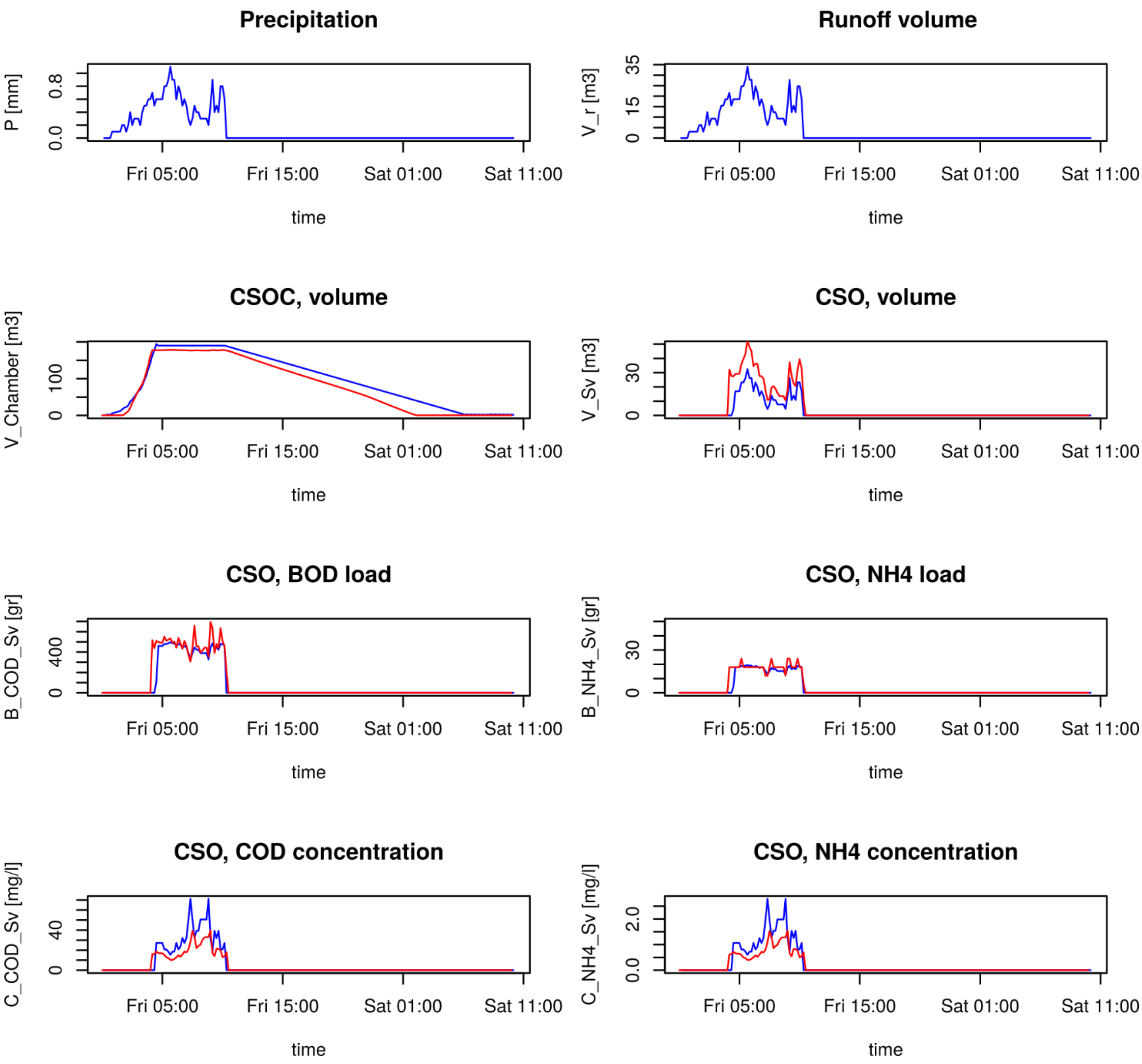

Figure 5. Comparison deterministic simulations for the integrated rainfall-runoff and sewer system models. EmiStatR (blue line) and CMM (red line). Goesdorf sub-catchment.

- *Spatio-temporal simulation*

We computed the CSO spill volume, and loads and concentrations of COD and NH4, based on the spatio-temporal rainfall fields predicted with the boundaries of the 90 percent confidence interval. Comparing these results with the deterministic simulation, we can infer that the deterministic model over estimates the runoff volume and, therefore, an over estimation of the CSOC volume and CSO spill volume. Consequently, the water quality variables, (COD and NH4) are over estimated as well. Indeed, the mean value of the predictions does not reflect any CSO spill volume and therefore no load and no concentration of pollutants. The deterministic simulations are comparable more to the upper boundary of the 90 percent confidence interval of the space-time simulation. Figure 6 shows the output from model Level 1 for the Goesdorf sub-catchment.

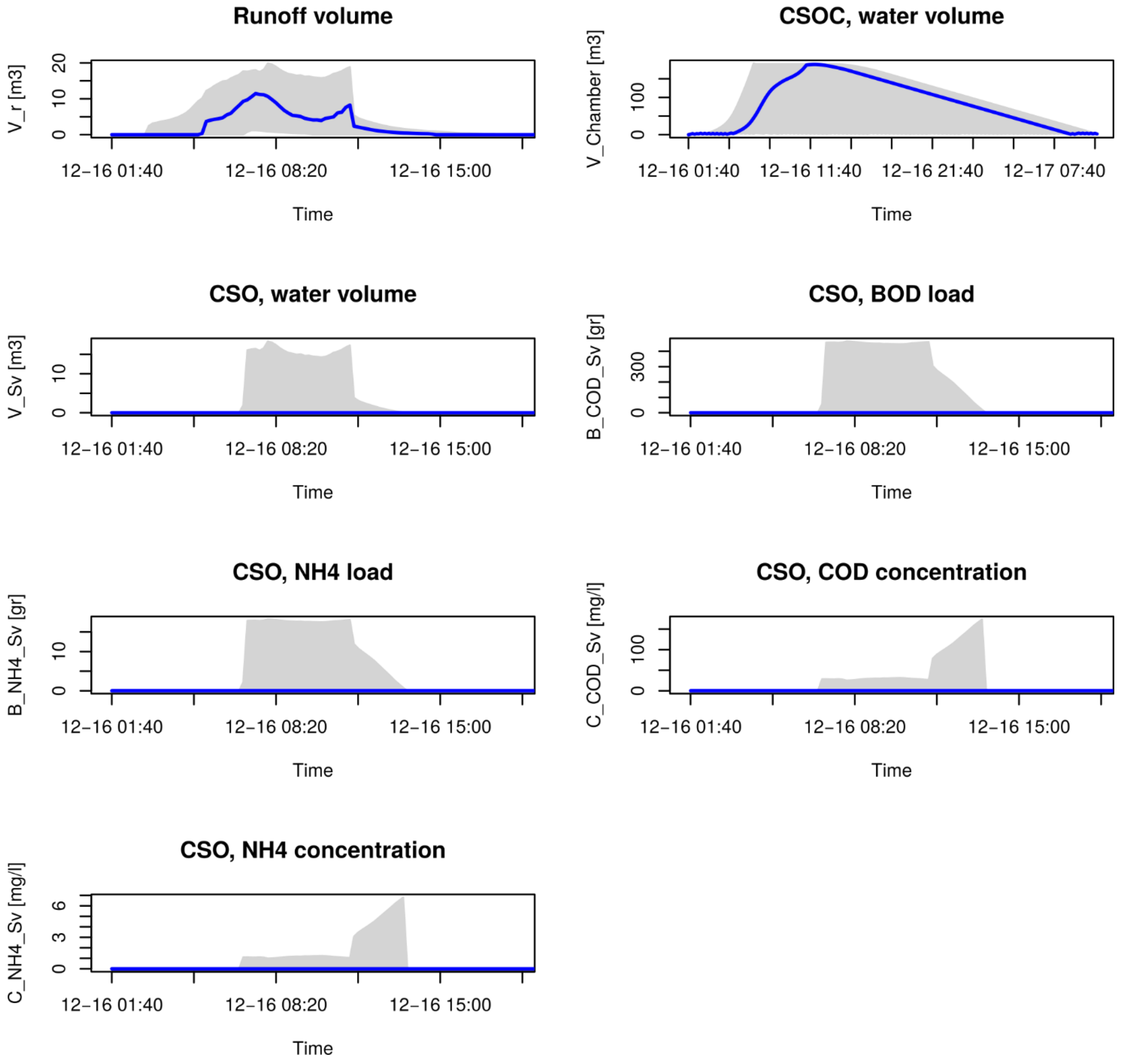

Figure 6. Output from model Level 1. Prediction value (blue line) and boundary of the 90 percent confidence interval (gray band). Goesdorf sub-catchment.

3 *River system modelling.*

Within receiving waters CSO impacts can be modelled with approaches that range from complex, detailed 3D hydrodynamic flow and diffusion models (e.g. Delft 3D) to highly simplified steady sate advection, or time invariant methods (e.g. SIMCAT). Previous work has shown that the impacts of uncertainty arising from river modelling tools are highly site specific (Camacho et al. submitted). In order to study the relationship between model structure (complexity) and uncertainty QUICS deliverable 4.7. (Camacho et al 2017) presented results of a study quantifying the variation between model outputs when utilising transport of mixing models with different complexities on the same system. The study focuses on the modelling of a CSO spill into a receiving water and the relative outputs of a 2D ADE mixing mode, a 1D ADE mixing model, a aggregated dead zone model and an advection only model. The variation in model outputs was found to be a function of distance downstream of the CSO release. Significant errors are introduced within simpler models close to the CSO source due to cross sectional averaging, however these effects become less significant with distance.

2.3 Step 3: Evaluation of the trade-offs between model complexity and model accuracy

Upon definition of the model complexity, and input uncertainty propagation and model accuracy evaluation, we proceed to the evaluation of the trade-offs between model complexity and model accuracy by means of an evaluation matrix as is shown in Figure 7.

		Model complexity	
		Simple (Level 1)	**Complex (Level 2)**
Model input	**Simple (temporal domain)**	**Simple model and simple model input.** This approach can lead to a parsimonious modelling technique, suitable for feasibility studies and long term simulation scenarios. Low computational burden required.	**Complex model and simple model input.**
	Complex (spatio-temporal domain)	**Simple model and complex model input.**	**Complex model and complex model input.** This approach can lead to a complete description of the processes modelled and accounting for complete description of model input uncertainty. This approach may be suitable for final design studies and not suitable for long term simulation scenarios because the high computational burden required.

Figure 7. Evaluation matrix for the identification of trade-offs between model complexity and model accuracy.

3 Conclusions

1) We presented in this deliverable a procedure as guidance on trade-offs between model complexity and model accuracy, composed by a three step procedure. In step 1 we illustrated different types of simplified or surrogate models as well as complex models. In step 2 we use model uncertainty analysis as a measure of model accuracy to analyse models of different complexity. A characterisation of the input uncertainty of the main input variables that control output uncertainty in water quantity and quality variables was done. We found that the uncertainty in loads, such as COD per capita per day in the sewage (CODs) and the concentration of COD in runoff (CODr) contribute to the uncertainty of the output variables: overflow COD load and overflow COD average concentration. CODr has the most important uncertainty contribution in the load and concentration of COD. The load of NH4 per capita per day in the sewage (NH4s) contributes totally in the uncertainty of overflow NH4 load and overflow COD average concentration. Rainfall is one of the most important drivers in the definition of uncertainty of output variables as load and concentration of COD and NH4.

2) To address rainfall uncertainty, we developed a space-time interpolation model for rainfall, based on space-time Kriging, using point rainfall measurements as the primary variable. This constitutes a useful technique for model input uncertainty characterisation and uncertainty propagation in the space-time domain. We interpolated rainfall over space and time and built a 90% confidence interval with the mean, lower and upper boundary for the Haute-Sûre urban drainage system catchment in North-West Luxembourg. The resulting space-time rainfall maps for mean, lower and upper bounds of 90% confidence interval were fed into a rainfall-runoff model simulating the routing of the runoff across the catchment to finally enter the urban drainage system model to predict water quantity and water quality in the combined sewer overflows (CSOs). The predicted space-time rainfall uncertainty propagation demonstrated that an over estimation of CSO spill volume and consequently pollutants (COD and NH4) is done when we consider only the deterministic simulation without taking into account the space-time model for rainfall. Also, we demonstrated that we can achieve a more realistic range of the physical processes for runoff generation and urban drainage hydraulics. Furthermore, the presented methodology is generic and can be applied to a wider range of integrated environmental assessment models.

3) This proposed procedure contributes to the evaluation of trade-offs between model complexity and model accuracy through input uncertainty propagation and model accuracy in urban drainage modelling. Also, we presented the feasibility and implementation to use the above approaches for both the scientific and the practitioners' communities.

4 References

Asher, M.J. et al., 2015. A review of surrogate models and their application to groundwatermodeling. In Water Resources Research. pp. 5957–5973.

Camacho, V., Schellart, A., Shucksmith, J., (2017) ‘Tool to advise on using appropriate river pollutant transport model’. QUICS Deliverable 4.7., Accessible from *https://www.sheffield.ac.uk/quics/dissemination/reports*

Camacho ,V., Schellart, A., Brevis, W., Shucksmith, J., Quantifying the impact of uncertainty in the dispersion coefficient on water quality modelling in rivers (Submitted to Water Resources Reseach)

Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-Temporal Interpolation using {gstat}. The R Journal, 8 (1):204–218.

Heuvelink G. B. M., Brown J. D., van Loon E. E. (2007). A probabilistic framework for representing and simulating uncertain environmental variables. International Journal of Geographical Information Science 21(5), 497–513.

Nol, L., Heuvelink, G. B. M., Veldkamp, A., de Vries W., Kros J. (2010). Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. Geoderma 159(1-2), 9–23.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the "gstat" package. Computers & Geosciences, 30:683–691.

Torres-Matallana, J.A., Leopold, U., & Heuvelink, G. (2017). Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. European Water, XX, 323–330.

Torres-Matallana, J.A.; Klepiszewski, K.; Leopold, U.; Heuvelink, G. EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows. (2018) Water. (Accepted with revision).