Reply to interactive comment from Referee 1 on "Multivariate autoregressive modelling and conditional simulation for temporal uncertainty propagation in urban water systems" by Jairo Arturo Torres-Matallana et al.

Jairo Arturo Torres-Matallana, Ulrich Leopold, Gerard B.M. Heuvelink

Received and published: 7 October 2020

The manuscript "Multivariate autoregressive modelling and conditional simulation for temporal uncertainty propagation in urban water systems" aims to select and characterize the main sources of input uncertainty in urban water systems and quantifying the contributions of each uncertainty source to model output uncertainty over time. It provides a good and well-structured example of an uncertainty analysis for a quite simplified model. Especially the results on CSO water quantity and quality are interesting and useful for further studies. In general, I think the manuscript is a valuable addition to the field with some minor discussion points I would like to bring up:

Reply: Thank you for your kind words and valuable comments that helped us to improve the manuscript. We considered each comment and please find our replies below after each comment from Referee 1.

[1] Page 2 – Line 40: Is the minimization of CSO volume alone a goal in itself? There is the question if many events with a bad water quality (e.g. first flush) are better than fewer events with higher volume and better water quality? That may be a point for elaboration.

Reply: Good point. Minimization of CSO volume is not the only goal. In the revision, we have changed the sentence to: (Revision, Page 2 L38-39) "To reduce pollution in receiving waters it is important to minimise CSO load and concentration".

[2] Page 12 – Line 261: As your model is quite simple and requires "little" computational time the chosen method is feasible, but that is not the case in most of those integrated studies. Is that not a limitation worth mentioning and discussing in 4.4? How could the approach look like in a more complex model?

Reply: We agree that this point should be addressed in the Discussion and have made appropriate adaptations in the revised manuscript in Section 4.4 (Revision, Page 30, L638-647):

> **6. Uncertainty analysis with complex models.** In this research we were able to conduct a comprehensive Monte Carlo uncertainty propagation analysis, which required a large number of Monte Carlo runs. This was possible because we used a strongly simplified urban water system model, EmistatR. For more complex models that take much more computing time, application of a Monte Carlo uncertainty propagation analysis is more challenging. However, given sufficient resources it is possible, because each model run can be run independently and hence the analysis is extremely suitable for parallisation and cloud computing. In particular, the use of graphics processing units (GPU) for heavy computation is promising. Some recent examples that demonstrate the potential of GPU for this purpose are Eranen et al. (2014), Sten et al. (2016) and Sandric et al. (2019). Sriwastava et al. (2018) applied uncertainty propagation to a complex hydrodynamic model, by selecting a small subset of dominant input/model parameters that explain most of the model output variance.

The methodology used in our study may be replicated for a model of higher complexity because of the scalable approach that was followed. The main limitation of application to a higher model complexity case is not the method implementation itself but the hardware setup that is required to make the uncertainty propagation feasible. It is necessary to speed up the computations of a single model run, which is not always an easy task.

[3] Page 19 – Table 4: I quite like this very accessible and clear table for the decision- making of which input variables you select. Still, I think that the variables that are awarded ++ and + for uncertainty and sensitivity respectively must be discussed more. Especially I think that on the infiltration, NH4 in Rainwater, and C pervious where I don't necessarily agree with omitting them, at least not on the argues in the text of 3.1. On the other hand, I am surprised on the uncertainty of the total area. So, the distinction where the authors draw the deciding line in what to include into their analysis must be clearer. It could be maybe better explained by using graphical panels (e.g. in QUICS (Tscheikner-Gratl et al., 2017)) for illustrating that decision.

Reply: We agree that the decision between ++ and + for uncertainty and sensitivity needs more justification and have made appropriate adaptations in the revised manuscript for inflow of infiltration water, NH4 in rainwater, and C pervious:

(Revision, Page 14, L326-328) "To better support our decisions we also include a graphical assessment of the degree of uncertainty and sensitivity of each

input, as in Tscheikner-Gratl et al. (2017), see Figure S1 in the Supplementary Material.

Adaptation Section 3.1.2 (Revision, Page 15, L341-343): Regarding the inflow of infiltration water (4), "Although this is a very uncertain input, the quick-scan analysis showed that model output sensitivity is not very high as is indicated in Table 3. For this reason we did not include this variable in the uncertainty propagation analysis."

Adaptation Section 3.1.3 (Revision, Page 15, L353-354): Regarding NH_4 in rainwater (9), "Although model output is very sensitive to this model input variable, model input uncertainty is not very high as is indicated in Table 3. For this reason it was not included in the uncertainty propagation analysis."

Adaptation Section 3.1.4 (Revision, Page 15, L359-361): "Although model output is very sensitive to the input variable C_{per} (13), the uncertainty about this variable is not very high, as indicated in Table 3. The reason behind this is that C_{per} can be derived fairly accurately from GIS products, such as land use and soil type maps. Therefore, we did not include this variable in the uncertainty propagation analysis."

[4] Page 29 – Line 560: You don't start with the accuracy of Monte Carlo Analysis (which is then 4.2) but with Uncertainty and water quality impact (4.1).

Reply: Thank you for noting this mistake. We have corrected this in the revised manuscript, by changing the text to: (Revision, Page 25, L515-518) "In the following discussion, we start with the uncertainty and water quality impact of the model outputs to the environment, in relation to the uncertainty analysis. Next, we discuss the accuracy of Monte Carlo analysis, followed by a discussion of other sources of uncertainty. Finally, we highlight some limitations and possible solutions of the approach used in this work."

[5] Page 31 – Line 588: I agree that that is one of the very valuable contribution of this paper. Still I would like to see some comparisons to other attempts on quantity (e.g. Sriwastava et al., 2018) and quality (especially measurements taken at CSOs the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs - e.g. Brombach et al.(2005); Diaz-Fierros T et al. (2002))

Reply: Thank you for your kind words and suggestion. We have made appropriate adaptations in the revised manuscript by expanding the text and including comparisons with other quantity and quality studies as follows:

(Revision, Page 27, L544-566)"We also recognise other attempts on quantity (e.g. Sriwastava et al., 2018) and quality, especially measurements taken at CSOs, which demonstrate that the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs (e.g. Brombach et al.(2005); Diaz-Fierros T et al. (2002)). We present some comparisons with these studies in the following lines.

Sriwastava et al. (2018) apply uncertainty propagation to a complex hy-

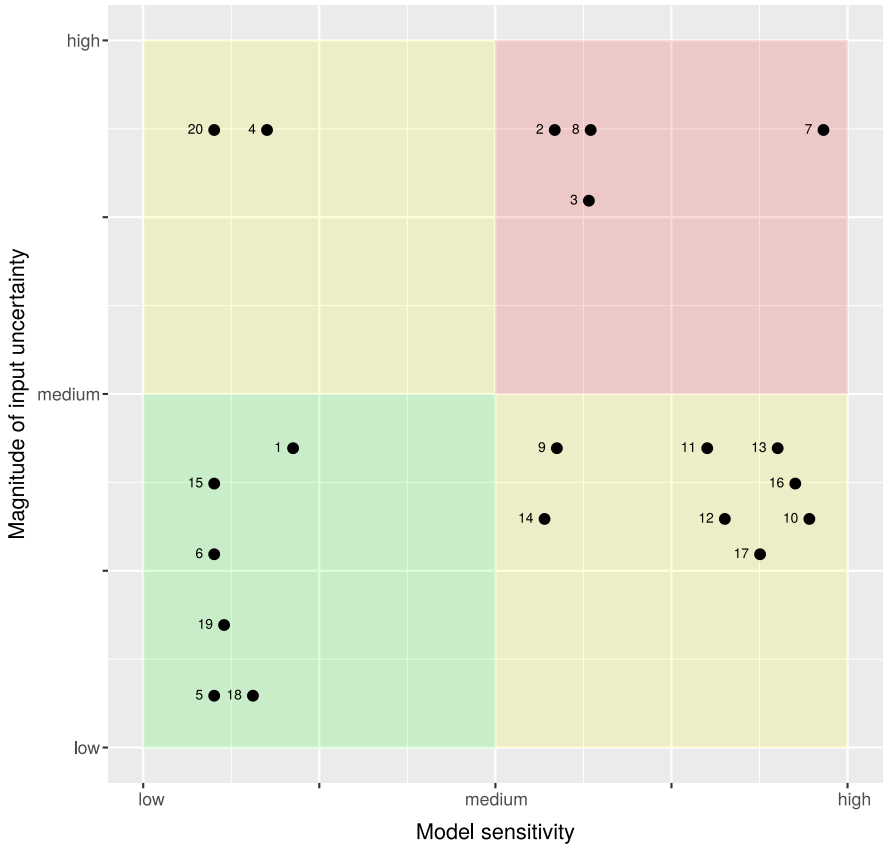

Figure 1: Graphical assessment of the contribution of input uncertainty to model output uncertainty. Numbers near each dot refer to the input variable number as defined in Table 4 of the manuscript. Panel layout after Tscheikner-Gratl et al. (2017).

drodynamic model for quantifying uncertainty in sewer overflow volume. They used MC for uncertainty propagation and Latin hypercube sampling (LHS) as an efficient sampling scheme. Although LHS ensures a full coverage of the sample space and provides a faster convergence than simple random sampling, the LHS application in the case of dynamic model inputs (e.g. precipitation, COD and NH4 inputs) is not trivial and its implementation is more complex than in the case of sampling from static variables (i.e., uncertain constants). In our study, we sampled time series of dynamic inputs using an implementation in stUPscales (Torres-Matallana et al., 2019; Torres-Matallana et al., 2018).

Diaz-Fierros et al. (2002), in a study in the city of Santiago de Compostela (North-West Spain, population about 100,000 inhabitants), where a combined sewer system feeds to a grossly under-sized wastewater treatment plant, reported an event mean concentration (Diaz-Fierros et al. (2002), Table 4) for the output variables $C_{COD,Sv,av}$ and $C_{NH4,Sv,av}$ of 329.1 mg·l^{-1} and 8.7 mg·l^{-1}, respectively. These values are larger than those found by Brombach et al. (2005), and more in agreement with our findings, especially for the case of $C_{NH4,Sv,av}$. Diaz-Fierros et al. (2002) reported values of $C_{COD,Sv,av}$ as high as 1073 mg·l^{-1}, which agrees with the right-hand tail of the distribution obtained in our study (i.e. a 0.995 quantile of 909.7 mg·l^{-1}). Similarly, for the case of $C_{NH4,Sv,av}$, Diaz-Fierros et al. (2002) reported values as high as 32.5 mg·l^{-1}, comparable with the 0.995 quantile (29.20 mg·l^{-1}) found in our study.

It is worth noting that regarding measurements taken at CSOs, the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs as reported by Diaz-Fierros et al. (2002) and Brombach et al. (2005). Thus, when comparing model outputs with independent measurements, one should bear in mind that discrepancies between measured and predicted are not only caused by errors in model inputs, model parameters and model structure, but are also the result of errors in the water quality measurements."

[6] Page 32 – Line 62: The point about linkage is an important one, but I don't see the big input from this paper on the topic. Can you elaborate on this, why is the quantification at sub-module level advisable? Only due to the computational budget limitations?

Reply: We agree that we did not address this aspect in our paper but in the Discussion we did want to point to the possibility of obtaining uncertainties at sub-model level. Some users may be interested at uncertainty levels of sub-modules of the model. For example, sub-module outputs are of particular interest in Bach et al. (2015), Burger et al. (2016) and Rauch et al. (2017).

Perhaps the text on lines 632-636 of the original manuscript was not very clear. We will reformulate it to: (Revision, Page 29, L610-615) "Tscheikner-Gratl et al. (2019) addressed the question as to whether there is an increase in uncertainty by linking integrated models or whether a compensation effect could take place by which overall uncertainty in key water quality parameters decreases. Some further insight into this topic could be obtained by quanti-

fying uncertainties at sub-model level, and analysing whether uncertainty at sub-model level is greater or smaller than at the overall level. With our implementation this is not a difficult task because EmistatR has a stringent modular design in which it is easy to analyse outputs and their uncertainties at sub-model level."

[7] Page 34 - Line 701: Your abstract starts with "Uncertainty is often ignored in urban water systems modelling." I would have therefore expected and would like to read how this can be improved and how studies like yours can provide guidance for the decision makers.

Reply: We believe that we have made a contribution towards making uncertainty propagation analysis in urban water systems modelling more routine. Clearly a single journal publication is not enough but we provide guidance, a simplified model that is very suited for Monte Carlo uncertainty propagation, and we shared the code scripts as well as the datasets to reproduce Figures 3 to 6, so that interested parties could more easily run an uncertainty analysis themselves. Please also note that our study was part of the larger 'QUICS' EU project (https://www.sheffield.ac.uk/quics), which aimed to stimulate the use of uncertainty analysis in integrated catchment modelling, and which involved partners from industry, water management authorities and consultancy firms.

Literature:

Brombach, H., Weiss, G., Fuchs, S., 2005. A new database on urban runoff pollution: comparison of separate and combined sewer systems. Water Sci Technol 51, 119–128. https://doi.org/10.2166/wst.2005.0039

Diaz-Fierros T, F., Puerta, J., Suarez, J., Diaz-Fierros V, F., 2002. Contaminant loads of CSOs at the wastewater treatment plant of a city in NW Spain. Urban Water 4, 291–299. https://doi.org/10.1016/S1462-0758(02)00020-1

Sriwastava, A.K., Tait, S., Schellart, A., Kroll, S., Van Dorpe, M., Van Assel, J., Shucksmith, J., 2018. Quantifying Uncertainty in Simulation of Sewer Overflow Volume. Journal of Environmental Engineering 144, 04018050. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001392

Torres-Matallana, J., Leopold, U., and Heuvelink, G.: stUPscales: an R-package for spatio-temporal Uncertainty Propagation across multiplescales with examples in urban water modelling, Water, 10(7), 1–30, https://doi.org/10.3390/w10070837, 2018.

Torres-Matallana, J., Leopold, U., and Heuvelink, G.: stUPscales: Spatio-Temporal Uncertainty Propagation Across Multiple Scales, https://CRAN.R-project.org/package=stUPscales, r package version 1.0.5.0, 2019.

Tscheikner-Gratl, F., Lepot, M., Moreno-Rodenas, A., Schellart, A., 2017. A Framework for the application of Uncertainty Analysis (Deliverable No. 6.7), QUICS. Zenodo, https://zenodo.org/record/1240926

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-342, 2020

Reply to interactive comment from Referee 2 on "Multivariate autoregressive modelling and conditional simulation for temporal uncertainty propagation in urban water systems" by Jairo Arturo Torres-Matallana et al.

Jairo Arturo Torres-Matallana, Ulrich Leopold, Gerard B.M. Heuvelink

Received and published: 7 October 2020

Reply: We welcome your comments, which were very helpful to improve our manuscript. In the text below we provide our replies (in blue) after each original comment.

* General

This manuscript presents a detailed case study on uncertainty propagation through a water quality model. The authors propose the use of auto-regressive models to describe the dynamic of the input time-series.

For the reader it is currently unclear if the focus is on the method ("this paper introduces an uncertainty analysis framework") or the application case. My suggestion is to focus on the case study, as the methodological contribution is rather limited. In any case, the focus should be set clearer in the introduction. In general, focusing more on the key points by moving some material to the supporting information would help.

Reply: We agree that the added value of the paper is mainly in the application, and we have made this clear in the revision, for example by no longer claiming that we introduce an uncertainty analysis framework but instead recognising that we build on existing methods (Revision, Page 1, L4-5). To emphasise that the application is most important we also changed the title of the manuscript to "Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system **in Luxembourg**".

Even though the focus is on the application we do think that our paper also makes a valid contribution on methods. For example, we use both univariate and multivariate autoregressive models to characterise uncertainty in dynamic variables, we use conditional simulation to sample from these models, and we use bootstrap computation to summarise the Monte Carlo outputs. None of

these methods are new but as far as we know they have not been used jointly in uncertainty analysis of urban water systems. We have moved some of the materials to the Supplementary Material, in particular the second part of Section 2.5.4 where we explain the calibration and conditional simulation of the precipitation time series (i.e., lines 223 to 255 in the original manuscript).

The authors use AR(1) processes to model dynamical inputs. However, it needs also to be shown that these models captures the characteristics of the inputs correctly, for example that the auto-correlation function and other statistics match.

Reply: We agree that we did not provide evidence of our statement on line 435 of the original manuscript that the simulated precipitation time series captured the main statistics of the observed time series well. Please find below evidence for this statement (Table 1 and Figure 1). The match is not perfect but we judge it close enough, given that the model is only an approximation of the real world (e.g., it assumes that the log-transformed precipitation has a normal distribution, constant a priori mean and variance, and a stationary near-exponential autocorrelation function that results from an AR(1) model formulation). We made a note in the text (Revision, Page 19, L411-412) and refer to the Supplementary Material, Table S1 and Figure S2.

Table 1: Mean and variance of the log-transformed observed precipitation time series at Esch-sur-Sure and Dahl rain gauges and the simulated precipitation time series at Goesdorf (random selection of simulation numbers 1, 750, 1500 and all).

	Esch-sur-Sure	Dahl	Sim 1	Sim 750	Sim 1500	Sims (All)
Mean	-6.6152	-6.5817	-6.3888	-6.3886	-6.3878	-6.3874
Variance	1.4188	1.5731	1.5636	1.5579	1.5594	1.5582

Table 2 shows comparisons of the means and variances for $C_{COD,S}$ and $C_{NH4,S}$ based on 91 measurements in the Haute-Sure catchment and simulations at Goesdorf (note that for COD_r a comparison cannot be made because we had no measurements of COD_r and a model for COD_r was based on expert judgement). The agreement between observed and simulated statistics is again quite close. We could not evaluate the autocorrelation functions of the observed $C_{COD,S}$ and $C_{NH4,S}$ because there were too few observations to be able to compute these (note that the 91 observations were from multiple locations within the catchment, see original manuscript lines 401-403).

'Uncertainty analysis' is an umbrella term. Therefore I would encourage to use always the more specific terms 'uncertainty propagation' and 'sensitivity analysis' (SA) when referring to these concepts. SA is a well established term for calculating the "contributions of input variables to total uncertainty".

Reply: We agree that uncertainty analysis is an umbrella term and made sure that we used it only in that way, while we use the terms 'uncertainty prop-

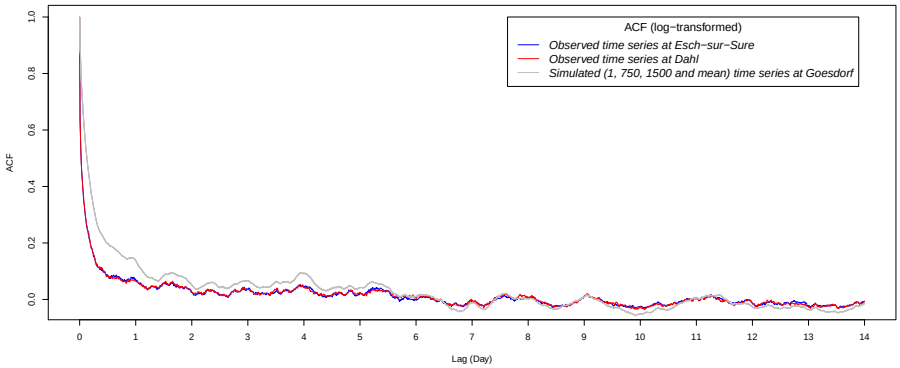

Figure 1: Autocorrelation function of the log-transformed observed precipitation time series at Esch-sur-Sûre and Dahl rain gauges and simulated precipitation at Goesdorf catchment.

Table 2: Mean and variance of log-transformed observed $C_{COD,S}$ and $C_{NH4,S}$ in the Haute-Sure catchment and of log-transformed simulated $C_{COD,S}$ and $C_{NH4,S}$ at Goesdorf (random selection of simulation numbers 1, 750, 1500 and all).

	Observations	Sim 1	Sim 750	Sim 1500	Sims (All)
Mean ($\log(C_{COD,S})$)	4.3783	4.3752	4.3737	4.4106	4.3780
Variance ($\log(C_{COD,S})$)	0.5637	0.5261	0.5257	0.5394	0.5640
Mean ($\log(C_{NH4,S})$)	1.4733	1.4656	1.4639	1.4865	1.4730
Variance ($\log(C_{NH4,S})$)	0.1679	0.1704	0.1684	0.1615	0.1681

agation' and 'stochastic sensitivity analysis' when we refer to a specific method. We prefer to use 'stochastic sensitivity analysis' instead of just 'sensitivity analysis' because in fact there is a lot of confusion about this term in the literature. Many deterministic modellers interpret sensitivity analysis as an approach that analyses how the model responds to (small) changes in its inputs and parameters, irrespective of how uncertain the inputs or parameters are. In fact that is also how we interpreted 'sensitivity' in Sections 2.4 and 3.1 (when we did the quick-scan). We quote from Smith and Smith (2007, page 71): "Whereas a sensitivity analysis defines how the model responds to changes in its components, an uncertainty analysis determines how much uncertainty is introduced into the model by each component of the model".

We checked the manuscript carefully to make the necessary changes, also in the title (replaced 'uncertainty propagation' by 'uncertainty analysis') and some section headings (i.e. Sections 2.6 (Revision, Page 11, L225) and 3.3 Revision, Page 19, L417) in the revised manuscript).

While hinted a several places, I think it would be beneficial to distinguish the kind of uncertainty that one tries to model more explicitly. Some inputs are intrinsically stochastic (e.g. precipitation), while others the uncertainty expresses our lack of knowledge about a parameter value.

Reply: In this paper uncertainty is always an expression of limited knowledge about a model input or model parameter. We are well aware that in the literature different types of uncertainty are distinguished (such as epistemic and stochastic, see Refsgaard et al. (2007, Section 3.3)), but we doubt that such distinction is useful and can in fact be made. For example, precipitation is the result of a physical-deterministic process and hence not intrinsically stochastic, no matter that it is practically impossible to know it without error everywhere, all the time. In fact, it is hard to imagine that there are truly stochastic processes in nature (note the fundamental difference with chaotic processes, which do exist but are deterministic). Some argue that there is stochastic uncertainty at the quantum-mechanistic level, but even there, there are doubts (we only need to quote Einstein who said "God does not play dice with the universe"). Our paper intentionally does not go this philosophical road, and that is why we refrained from distinguishing different kinds of uncertainty.

* Specific points

** Abstract

L4: the paper does not introduce a framework (which would be a theory about how to deal with uncertainties).

Reply: We acknowledge that the main contribution is the application and have removed all references to a 'new framework' from the paper. Please see also our reply to your first general comment.

** Introduction

L54: "determinism" is the absence of uncertainty, hence it cannot "represent uncertainty"

Reply: Here we merely cited the literature. But we see your point and have rephrased the sentence to: (Revision, Page 2, L53-54)"Five approaches to represent the presence or absence of uncertainty and how it is represented in the context of urban water systems are often distinguished. . . .".

** Material and methods

L115, eq2: Maybe I missed something, but is there no delay? The rain is transformed immediately to runoff?

That is correct. Indeed, we had not explained in Section 2.1 that a delay is included in EmiStatR (see also Torres-Matallana et al., 2018, Section 2.2.3) and have corrected this in the revision. We replaced the text in lines 116-120 with:

(Revision, Page 4, L116-120) The contribution of rainwater to the combined sewage volume, Q_r [$m^3 \cdot s^{-1}$], is derived from precipitation as follows:

$$Q_{r_t} = \frac{1}{6} \cdot P_{t-t_{fS}} \cdot [C_{imp} \cdot A_{imp} + C_{per} \cdot (A_{total} - A_{imp})] \quad (1)$$

where $\frac{1}{6}$ is a factor for units conversion, $P_{t-t_{fS}}$ precipitation at time $t - t_{fS}$ [mm$\cdot$ min^{-1}]; t_{fS} is a delay in time response related to flow time in the sewer system; A_{imp} is the impervious area of the catchment [ha]; A_{total} is the total area of the catchment [ha]; C_{imp} is the run-off coefficient for impervious areas [-]; and C_{per} is the run-off coefficient for pervious areas [-].

L160: "Some variable" - which one?

Reply: We assume that you refer to line 163: "Some of the variables were calibrated based on observations...". These variables are water consumption (qs), infiltration flow (q_f), time flow (t_{fS}), run-off coefficient for impervious area (C_{imp}), run-off coefficient for pervious area (C_{per}), orifice coefficient of discharge (C_d) and initial water level (lev_{ini}). We included the variable names into the text to be clearer which variables have been calibrated (Revision, Page 7, L165-167).

Table 1: I would remove all irrelevant "inputs", such as ID, name of structure, ...

Reply: (Revision, Page 8, Table 1) We deleted ID of the structure, Name of the structure, Name of the municipality, Name of the catchment and Number of the catchment from Table 1. We also changed the table caption to: "Most important general, CSO input and output variables of EmiStatR, with base values for the general input variables.".

Table 2: What are "flow time structure" and "curve level"? They do not show up in the model equations

Reply: Thank you for spotting this mistake. Flow time structure (t_{fS}) should be included in Eq. 2 in the original manuscript, this is the delay in time response related to the flow time in the sewer system. See also our reply to your comment above on missing the delay from rain to runoff. The curve level - volume ($lev2vol$) refers to a characteristic of a CSOT and translates the level of the water in the CSOT tank to the CSOT volume. To save space we did not include it in our paper (as mentioned in lines 101-102 of the original submission full details are provided in Torres-Matallana et al., 2018). To avoid confusion we removed this parameter from Tables 1 (Revision, Page 8) and 2 (Revision, Page 9).

L165: This is an example where "uncertainty propagation" should be used instead of "uncertainty analysis"

Reply: We could not find the term "uncertainty analysis" in line 165. Perhaps the reviewer refers to line 176? We have replaced "uncertainty analysis" by "uncertainty propagation analysis" in line 176 of the original manuscript (Revision: Page 9, L179).

L165ff: Why was the selection of inputs needed at all for the uncertainty propagation? The computational efforts do not change with the number of inputs (they do for the SA). Also there is not needed that the model is sensitive to an input. If it is not, than we will just get smaller uncertainties.

Reply: We agree that from a computational point of view there is no need to reduce the number of inputs. However, each uncertain input needs to be modelled by means of a (complex) probability distribution, and as we indicated in lines 176-177 of the original manuscript this is the most difficult and time-consuming step. From this perspective, it definitely pays off to reduce the number of uncertain inputs and focus on the most important ones.

L179: the equation is for the cdf, not the pdf. Also emphasize on 'marginal' implies that you want to model a joint distribution.

Reply: We had defined 'pdf' in line 175 as a 'probability distribution function'. This can refer to a cumulative probability distribution, a probability mass function (for discretely and categorical variables) and a probability density function (see statistical text books or https://en.wikipedia.org/wiki/Probability_distribution_function). Eq. 6 is a cumulative probability distribution function, hence in the notation that we introduced in line 175 it is a cumulative pdf.

L188ff: how did you determine the order of the AR process (seems always to be 1)?

Reply: We chose the order to be 1 in all cases to keep the model as simple as possible while still being able to handle autocorrelation.

L199: Not the uncertainties are correlated, but the values themselves

Reply: Thank you for this interesting perspective. However, we think that uncertainties can also be correlated. For instance, suppose we estimated $C_{COD,S}$ and $C_{NH4,S}$ using some model but are uncertain about their true value because of estimation errors. Then it is likely that if we overestimated $C_{COD,S}$ we will also have overestimated $C_{NH4,S}$, because the model may have missed an emission event. Similarly, underestimation of both $C_{COD,S}$ and $C_{NH4,S}$ will also tend to occur simultaneously. Thus, if we wish to model our uncertainty (i.e., the estimation errors) by random variables (one for $C_{COD,S}$ and one for $C_{NH4,S}$), then we would include a positive correlation between the two random variables.

L211: Precipitation lot of zeros. It is not clear how the presented model can describe the dry periods.

Reply: As mentioned in lines 229-231 of the original manuscript, we applied a Kernel smoothing to the precipitation time series prior to modelling. This removed many of the zeroes. In addition, as explained in line 233 of the original manuscript, we only used precipitation data for calibration of the AR(1) model if these were above the threshold of 0.01 mm.

L215-255: Remove the description as the model is already described in Torres-Matallana et al. (2017)

Reply: We agree that we provided too much detail. We have moved the text from line 223 onward to the Supplementary Material. We prefer to keep the text in lines 215-222 of the original manuscript in the main article to explain the reader what we did. We have rephrased this text to:

> (Revision, Page 11, L218-224) "Torres-Matallana et al. (2017) present a model to simulate precipitation inside a target catchment given a known precipitation time series in a nearby location outside the catchment, while accounting for the uncertainty that is introduced due to spatial variation in precipitation. The method used for input precipitation uncertainty characterisation is essentially the same as the application of a Kalman filter/smoother (Kalman, 1960; Webster and Heuvelink, 2006). Calibration of the model requires precipitation time series at two locations near the catchment of interest. Once the model is calibrated, it is used to simulate precipitation inside the target catchment from a single precipitation time series nearby the catchment. Details of the calibration and conditional simulation are presented in the Section S3 of the Supplementary Material."

L257-260: Remove or move to introduction.

Reply: We agree and have removed these lines.

L304: This is sensitive analysis and does not belong under the section "uncertainty propagation"

Reply: We changed the title of Section 2.6 to (Revision, Page 11, L225) "Uncertainty analysis".

L317: How did you aggregate for the whole year? Is it a problem that the individual indices are not independent due to the auto-correlation?

Reply: We aggregated to a yearly value by taking the arithmetic mean of all 10 minute indices within the year. We explain this in the revision. We agree that there likely is temporal structure in the 10 minute indices but that has no

effect on the usefulness of computing and interpreting a yearly aggregate. It still is an overall annual measure of the stochastic sensitivity.

L340: "may be computed" - did you do so?

Reply: Thank you for noting our sloppy formulation. We changed this sentence to: (Revision, Page 14, L306) "We used the German guideline ATV-A 128 (1992), which computes the throttle discharge at CSOs, $Q_{t,CSO}[l \cdot s^{-1}]$ as:"

** Results

L397: The wording is a bit confusing here: "evaluation of model output sensitivities" sounds like a SA, but you are referring to the "manual" analysis of the model.

Reply: As explained above the term 'sensitivity analysis' has multiple meanings. To many it simply refers to the sensitivity of the model output to changes in the model inputs. This is also the interpretation we used here. To avoid confusion we rephrased the sentence to (Revision Page 16, L371) "After ranking all inputs on level of uncertainty and model sensitivity, we selected..."

L415: was the COD modeled independently of the precipitation?

You are right, we did not include cross-correlation between COD_r and precipitation, as mentioned in lines 637-638. We agree that it would be more realistic to include a correlation and had mentioned this under 'possible improvements' in lines 637-641 of the original manuscript (Revision Page 29, L616-623).

Figure3: Please mention in the caption which density you used of the uncertainty propagation. Also, maybe move figure to supporting information (SI).

Reply: Thank you for this comment and suggestion. We have moved this figure to the Supplementary Material (Page 6, Figure S3) and extended the caption with: "Note that the blue densities were used in the uncertainty propagation".

L435: Please show some evidence for that.

Reply: Please see our response to your general comment. We provided the evidence in the Supplementary Material (Pages 1 and 3, Section S2).

L444-465: I propose to move this and figure 4 to the SI

We have moved Section 3.3.1 to the Supplementary Material (Page 5 and 7-8, Section S5) and refer to it in the main text (Revision, Page 19, L424-425).

** Discussion

L560: "we start with the accuracy..." - this topic comes second.

Reply: Thank you for noting this mistake, which was also pointed out by Referee 1. We have corrected this in the revised manuscript, by changing the

text to: (Revision, Page 25, L515-518) "In the following discussion, we start with the uncertainty and water quality impact of the model outputs to the environment, in relation to the uncertainty analysis. Next, we discuss the accuracy of Monte Carlo analysis, followed by a discussion of other sources of uncertainty. Finally, we highlight some limitations and possible solutions of the approach used in this work."

L637: This is an important suggestions. I'm a bit confused why you did not considered the correlation if the model is apparently able to do so ("We used the latest version of EmiStat R (version 1.2.2.0), which considers this kind of patterns.")

Reply: This seems to be a misunderstanding. EmiStatR version 1.2.2.0 can account for $C_{COD,S}$ and $C_{NH4,S}$ to be correlated with the German ATV-A 134 curve, i.e. daily consumption. However, it cannot account for a correlation between COD_r and precipitation.

** References

Kalman, R. E.: A new approach to linear filtering and prediction problems, Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering, 82D, 35–45, 1960.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process - A framework and guidance, Environmental Modelling and Software, 22, 1543–1556, https://doi.org/10.1016/j.envsoft.2007.02.004, 2007.

Smith, J. and Smith, P. Environmental Modelling: An Introduction. ISBN: 9780199272068. Oxford University Press. 2007.

Torres-Matallana, J. A., Leopold, U., and Heuvelink, G. B. M.: Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models, European Water, 57, 299–306, 2017.

Torres-Matallana, J. A., Klepiszewski, K., Leopold, U., and Heuvelink, G.: EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows, Water, 10(6), 1–24, https://doi.org/10.3390/w10060782, 2018.

Webster, R. and Heuvelink, G. B. M.: The Kalman filter for the pedologist's tool kit, European Journal of Soil Science, 57, 758 773, https://doi.org/10.1111/j.1365-2389.2006.00879.x, 2006.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-342, 2020

Multivariate autoregressive modelling and conditional simulation for temporal uncertainty ~~propagation in~~ analysis of an urban water ~~systems~~system in Luxembourg"

Jairo Arturo Torres-Matallana[1,2], Ulrich Leopold[2], and Gerard B.M. Heuvelink[1]

[1]Soil Geography and Landscape Group, Wageningen University
[2]Research Group for Sustainable Urban and Built Environment, Department for Environmental Research and Innovation, Luxembourg Institute of Science and Technology

Correspondence: Jairo Arturo Torres-Matallana (arturo.torres@list.lu)

Abstract. Uncertainty is often ignored in urban water systems modelling. Commercial software used in engineering practice often ignores uncertainties of input variables and their propagation because of a lack of user-friendly implementations. This can have serious consequences, such as the wrong dimensioning of urban drainage systems (UDS) and the inaccurate estimation of pollution released to the environment. This paper introduces an uncertainty analysis ~~framework~~ in urban drainage modelling~~and applies it~~, built on existing methods and applied to a case study in the Haute-Sûre catchment in Luxembourg. The ~~framework~~ case study makes use of the EmiStatR model which simulates the volume and substance flows in UDS using simplified representations of the drainage system and processes. A Monte Carlo uncertainty propagation analysis showed that uncertainties in chemical oxygen demand (COD) and ammonium (NH_4) loads and concentrations can be large and have a high temporal variability. Further, a stochastic sensitivity analysis that assesses the uncertainty contributions of input variables to the model output response showed that precipitation has the largest contribution to output uncertainty related with water quantity variables, such as volume in the chamber, overflow volume and flow. Regarding the water quality variables, the input variable related to COD in the wastewater has an important contribution to the uncertainty for COD load (66%) and COD concentration (62%). Similarly, the input variable related to NH_4 in the wastewater plays an important role in the contribution of total uncertainty for NH_4 load (34%) and NH_4 concentration (35%). The Monte Carlo simulation procedure used to propagate input uncertainty showed that among the water quantity output variables, the overflow flow is the most uncertain output variable with a coefficient of variation (cv) of 1.59. Among water quality variables, the annual average spill COD concentration and the average spill NH_4 concentration were the most uncertain model outputs (coefficients of variation of 0.99 and 0.82, respectively). Also, low standard errors for the coefficient of variation were obtained for all seven outputs. These were never greater than 0.05, which indicates that the selected MC replication size (1,500 simulations) was sufficient. We also evaluated how uncertainty propagation can explain more comprehensively the impact of water quality indicators for the receiving river. While the mean model water quality outputs for COD and NH_4 concentrations were slightly above the threshold, the 0.95 quantile was 2.7 times above the mean value for COD concentration, and 2.4 times above the mean value for NH_4. This implies that there is a considerable probability that these concentrations in the spilled CSO are substantially larger than the threshold. However,

COD and NH_4 concentration levels of the river water will likely stay below the water quality threshold, due to rapid dilution after CSO spill enters the river.

Keywords: Stochastic sensitivity analysis; uncertainty analysis; input uncertainty; temporal uncertainty; urban water modelling

1 Introduction

Combined sewer systems are important components of the urban water infrastructure. These systems are typically found in old and large cities (Baker, 2009; Litrico and Fromion, 2009) and are designed to transport the water generated and accumulated in an urban catchment to the receiving water body. During normal conditions all water is transported to the treatment facility before it is released to the environment. This is the so-called *throttled outflow* or pass-forward flow (Hager, 2010). However, during extreme conditions with heavy precipitation, the *combined sewer overflow* (CSO) discharges excess water directly to nearby streams, rivers, lakes or other water bodies (Baker, 2009). The CSO contains polluted water and solid matter (Hager, 2010), which, when released to the environment, can have a damaging impact on the water quality status of the receiving waters (Bachmann-Machnik et al., 2018; Gasperi et al., 2012). CSO pollutant load emissions are of similar or greater magnitude than the emissions from wastewater treatment plants (Gasperi et al., 2012; Bachmann-Machnik et al., 2018). CSO discharge impacts are mainly high peak flows, high organic loads from single events, which can lead to oxygen depletion, and ecotoxic concentrations of ammonia (NH_3) (Miskewitz and Uchrin, 2013; Bachmann-Machnik et al., 2018). To reduce pollution in receiving waters it is important to minimise CSO ~~volume~~load and concentration.

One of the main variables is chemical oxygen demand (COD), which is an indicator of organic compounds in water. It is used to measure the effluent quality (Viana da Silva et al., 2011). High levels of COD are correlated with a decrease of the amount of dissolved oxygen (DO) available for aquatic organisms. A depletion of DO concentration in the water column from near 9 mg/l (the maximum solubility of oxygen in estuarine water on an average summer day), to below 2 mg/l, is referred to as hypoxia. If hypoxic conditions are reached, the health of the ecosystem is affected, and cause physiological stress, and even death, to aquatic organisms (on Environmental and atural Resources - CENR, 2003). Ammonium (NH_4) is another important variable and is an indicator of nitrogen compounds in water. Concentrations of NH_4 in water and wastewater are relevant because high levels of nitrogen in receiving waters can cause eutrophication and, therefore, excessive growth of algae and other micro-organisms, resulting in oxygen dissolved depletion and fish toxicity (Huang et al., 2010).

To better assess environmental impacts, numerical models are applied in urban hydrology to simulate CSO emissions into the environment. It is recommended, however, that such modelling approaches consider the inherent uncertainty associated with the system representation and the approximation of the model to the reality (Hutton et al., 2011). Moreover, the model inputs are also not free of errors and associated uncertainties will also propagate to the model output (Heuvelink, 1998).

Five approaches to represent ~~uncertainty~~ the presence or absence of uncertainty and how it is represented in the context of urban water systems are often distinguished (Walker et al., 2003; Refsgaard et al., 2007; van der Keur et al., 2008; Bach et al., 2014): 1) determinism; 2) statistical uncertainty; 3) scenario uncertainty; 4) recognised ignorance; and 5) total (unrecognised) ignorance. Following van der Keur et al. (2008), *determinism* applies when we have knowledge with absolute certainty about the system under analysis. This is the "ideal world" case which is not realistic for urban hydrology systems. The *statistical* approach is useful when it is possible to describe uncertainty in statistical terms, i.e. when uncertainty can be characterised by probability distribution functions (pdfs). The *scenario* approach, in contrast, applies when quantitative probabilities cannot be determined, and instead qualitative measures of uncertainty are used. It is used when possible outcomes of uncertain inputs are known but not the probabilities of these outcomes (Brown, 2004). There is also no claim that the list of possible outcomes (scenarios) is exhaustive. *Recognised ignorance* occurs when there is awareness of lack of knowledge, but without any further possibility to process and address the recognised uncertainty. This is the case of very complex functional or inherently unidentifiable relationships, when e.g. predictions are infeasible due to chaotic behaviour of the system or when our understanding of the system behaviour is too limited (van der Keur et al., 2008). This is common in social systems where behaviour of humans and groups of humans may often unpredictable. Finally, *total ignorance* is the state of *"complete lack of awareness about imperfect knowledge"* (van der Keur et al., 2008). It is the opposite of determinism and reflects a state where we do not know that we do not know (Walker et al., 2003). Among the approaches described above, in this paper we will use the statistical approach to characterise and propagate uncertainties.

Three main sources of uncertainty in the context of performance evaluation analysis and design of urban water infrastructure and urban drainage modelling are identified (Walker et al., 2003; Neumann, 2007; Deletic et al., 2012). First, *model input* uncertainty is related to errors in input data, i.e. in driving forces such as precipitation. Second, *parameter uncertainty* is related to the uncertainty regarding the (calibrated) parameters of the model. Third, *model structural uncertainty* relates to uncertainty due to model conceptualisation and simplification. For instance, an urban drainage model might ignore certain sub-processes such as evaporation or chemical transformation or might simplify a non-linear relation between model variables to a linear relation. These type of uncertainties are not captured in model input and model parameter uncertainty and are represented by model structural uncertainty. The focus of this work is on the propagation of model input uncertainty.

Regarding methods for uncertainty propagation analysis, a distinction can be made between analytical methods, such as the Taylor series method (Heuvelink, 1998), and numerical techniques, such as Monte Carlo (MC) simulation. Numerical techniques are more flexible and hence more convenient to analyse uncertainty propagation with complex models (Zoppou, 2001). MC simulations are computationally demanding, especially in the case of complex models, but they can still be used if there are sufficient computational resources (Bastin et al., 2013), among others because it can greatly benefit from parallel computing.

Although uncertainty propagation analysis has been applied extensively in hydrologic modelling (e.g. Beven and Binley (1992); Kuczera and Parent (1998); Hutton et al. (2011); Vrugt et al. (2003b, a); Vrugt and Robinson (2007); Renard et al. (2010); Datta (2011)), the number of applications of long-term simulations in urban drainage modelling is limited and typically does not consider the influence of temporal and spatial correlation in the analysis of propagation of input uncertainty. Temporal

correlation occurs in uncertain dynamic variables such as precipitation and COD of household wastewater, because values of these variables over short time lags will be more similar than over large time lags. The same concept applies to variables that are spatially distributed (Webster and Oliver, 2007). It is important to take temporal (and spatial) correlation of uncertain inputs into account because this may have a major influence on the outcomes of an uncertainty analysis (Heuvelink, 1998). In this paper we perform a temporal uncertainty propagation analysis in urban water modelling, using MC simulation. As a case study we use the simplified model EmiStatR (Torres-Matallana et al., 2018b) to predict wastewater volume, COD and NH_4 concentrations in CSOs for three urban-rural sub-catchments of the Haute-Sûre catchment in the North-West of the Grand Duchy of Luxembourg.

The objectives of this study are to: 1) select and characterise the main sources of input uncertainty accounting for temporal auto- and cross-correlation within EmiStatR; 2) propagate input uncertainty through EmiStatR, taking into account temporal auto- and cross-correlation of uncertain dynamic inputs; 3) quantify and assess the contributions of each uncertainty source to model output uncertainty dynamically (over time) for the Luxembourg case study.

2 Materials and methods

2.1 The EmiStatR model

EmiStatR is used to simulate CSO flows and water quality concentrations. Details regarding the conceptual and mathematical model are provided in Torres-Matallana et al. (2018b). The main components of the EmiStatR model are: 1) Dry Weather Flow (DWF) including Infiltration Flow (IF); 2) Pollution of DWF; 3) Rain Weather Flow (RWF); 4) Pollution of RWF; 5) Combined Sewer Flow (CSF) and pollution; and 6) Combined Sewer Overflow (CSO) and pollution. Figure 1 illustrates the scheme of the sewer system analysed.

Basically, the total dry weather flow, Q_{DWF} $[\mathrm{l \cdot s^{-1}}]$ is calculated as:

$$Q_{DWF_t} = Q_{s_t} + Q_{f_t} \tag{1}$$

where Q_{DWF_t} $[\mathrm{l \cdot s^{-1}}]$ is the dry weather flow at time t and Q_{s_t} $[\mathrm{l \cdot s^{-1}}]$ is the dry weather flow of the residential sewage in the catchment at time t, calculated as $86,400^{-1} \cdot pe_t \cdot qs_t$ (where $86,400 = 24 \times 60 \times 60$ is a measurement unit conversion factor), with pe_t [PE] the population equivalents of the connected CSO structure at time t, and qs_t $[\mathrm{l \cdot PE^{-1} \cdot d^{-1}}]$ the individual water consumption of households at time t. Q_{f_t} $[\mathrm{l \cdot s^{-1}}]$ is the infiltration flow at time t that enters the pipes from groundwater flow through cracks and joints, calculated as $A_{imp} \cdot q_{f_t}$, where A_{imp} [ha] is the impervious area of the catchment, and q_{f_t} $[\mathrm{l \cdot s^{-1} \cdot ha^{-1}}]$ is the infiltration water inflow flux (specific infiltration discharge from groundwater flow) at time t. Variables qs_t and pe_t are dynamic and can be defined as time series with daily, weekly and seasonal patterns.

The contribution of ~~rain water~~ rainwater to the combined sewage flow, Q_r $[\mathrm{m^3 \cdot s^{-1}}]$, is derived from precipitation as follows:

$$Q_{r_t} = \frac{1}{6} \cdot P_t P_{t-t_{fS}} \cdot [C_{imp} \cdot A_{imp} + C_{per} \cdot (A_{total} - A_{imp})], \tag{2}$$

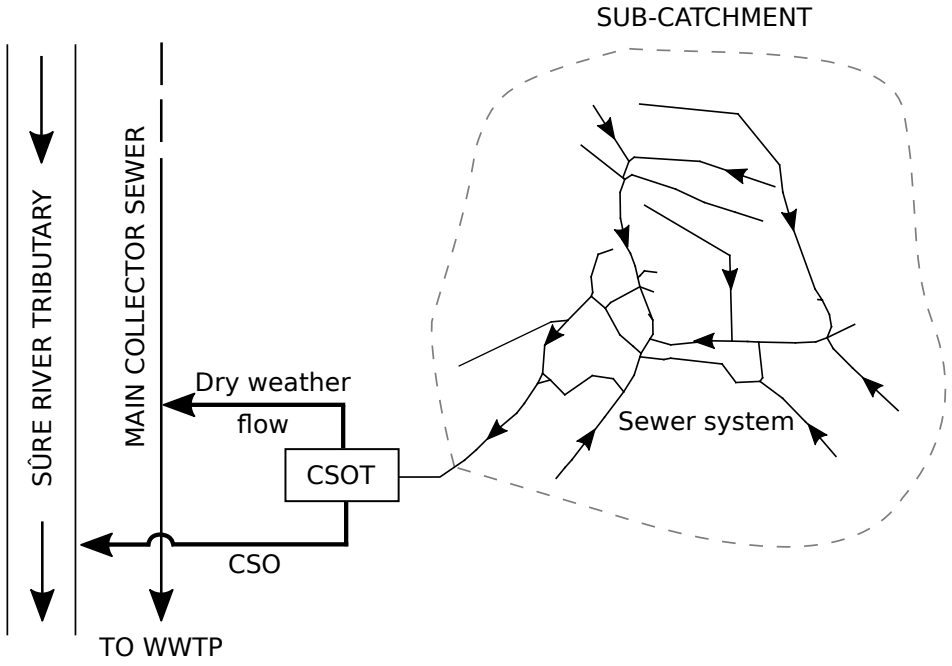

Figure 1. Scheme of the sewer system analysed. Adapted from: Andrés-Doménech et al. (2010)

where ~~1/6~~ $\frac{1}{6}$ is a factor for units conversion, ~~P_t is a time series of precipitation per unit time at time t~~ $P_{t-t_{fS}}$ precipitation at time $t - t_{fS}$ [mm· min^{-1}]; t_{fS} is a delay in time response related to flow time in the sewer system; A_{imp} is the impervious area of the catchment [ha]; A_{total} is the total area of the catchment [ha]; C_{imp} is the run-off coefficient for impervious areas [-]; and C_{per} is the run-off coefficient for pervious areas [-]. From Q_{r_t}, the CSO volume calculation is based on the exceeding volume stored in the Combined Sewer Overflow Chamber (CSOC). The CSO volume depends on four CSOC stages: (1) filling up; (2) CSO spill volume; (3) stagnation; and (4) emptying. The sum of the total dry weather flow, Q_{DWF_t}, and the rain water flow, Q_{r_t}, is called combined sewer flow at time t, Q_{CSF_t}.

The COD load, $B_{COD,Sv}$ [g], in the spill overflow volume is calculated as a function of the spill overflow volume at time t, V_{Sv_t} [m^3], a combined sewer mixing ratio at time t, cs_{mr_t} [-], the mean dry weather pollutant concentration at time t, C_{COD_t} [mg $\cdot$ l^{-1}], and the concentration due to rainwater pollution at time t, COD_{r_t} [mg $\cdot$ l^{-1}]:

$$B_{COD,Sv_t} = (cs_{mr_t} + 1)^{-1} V_{Sv_t} (cs_{mr_t} \cdot C_{COD_t} + COD_{r_t}) \tag{3}$$

The variable V_{Sv_t} depends directly on the water volume in the CSO chamber at time t, $V_{Chamber_t}$ [m^3]. It is computed as:

$$V_{Sv\,t} = \begin{cases} V_{r_t} + V_{dw_t} - V_{d_t}, & \text{if } V_{Chamber\,t} = V, \\ V_{Chamber\,t} - V & \text{if } V_{Chamber\,t} > V, \\ \epsilon & \text{if } V_{Chamber\,t} < V. \end{cases} \tag{4}$$

where V_{r_t} is the rain weather volume at time t accumulated during a time interval Δt [min], V_{dw_t} [m^3] is the total dry weather volume (amount of dry weather water in combined sewage flow) at time t, V_{d_t} is the volume of throttled outflow to the WWTP at time t [m^3], V [m^3] is the CSOC volume, and ϵ is a numerical precision term set equal to 10^{-5} [m^3]. While V_{Sv}, cs_{mr} and C_{COD} are dynamic, COD_{r_t} can either be dynamic or assumed constant if the pollution concentration is assumed constant in time. C_{COD_t} [$mg \cdot l^{-1}$] is calculated as (Torres-Matallana et al., 2018b):

$$C_{COD\,t} = \frac{10^3 \cdot pe_t \cdot C_{COD,S}}{qs_t \cdot pe_t + 86,400 \cdot A_{imp} \cdot q_{f_t}} \tag{5}$$

where $C_{COD,S}$ is the COD sewage pollution per capita [PE] load per day [$g \cdot PE^{-1} \cdot d^{-1}$]. Similar equations as above apply to the second water pollution indicator NH_4.

2.2 Sewer system in the Haute-Sûre catchment

The study area is composed of three sub-catchments of the Haute-Sûre catchment in the north-west of the Grand-Duchy of Luxembourg. The combined sewer system drains three villages: Goesdorf (GOE), Kaundorf (KAU), and Nocher-Route (NOR). The local sewer system downstream each village has a CSO tank to store pollutant peaks in the first flush of combined sewage flows. Table 2 shows the general characteristics of each CSO tank for each village. Figure 2 depicts the location of the CSO tanks and the delineation of the sub-catchments. The main land use types in the villages are residential, smaller industries and farms. Outside of the villages forest as well as agricultural arable and grassland are the dominating land uses. The receiving water bodies of the CSO structures are tributaries of the river Sûre (Sauer, in German).

2.3 Input data

The input variables of the EmiStatR model are shown in Table 1. Following Torres-Matallana et al. (2018b), seven input variables were calibrated: water consumption (qs_t), infiltration flow (q_{f_t}), flow time structure equivalent to the time of concentration to the combined sewer overflow tank (CSOT) structure (t_{fS}), run-off coefficient for impervious area (C_{imp}), run-off coefficient for pervious area (C_{per}), orifice coefficient of discharge (C_d), and the initial water level (Lev_{ini}). The main objective of the calibration process is to represent appropriately the water volume in the CSOT.

The observed precipitation (P_t) is a one year time series for 2010 at 10 minute time interval, measured at stations Esch-sur-Sûre and Dahl (Fig. 2). The variable water consumption (qs_t) is also dynamic and represented as a time series with a daily pattern according to factors proposed in the design German guideline ATV-A 134 (Evers et al., 2000).

The hydraulic variable measured is water level in the CSOT t, Lev [cm]. The temporal resolution of measurements of Lev is 30 seconds. Regarding wastewater quality (WWQ) characterisation, values of C_{COD_s} and C_{NH4_s} in the wastewater were derived from DWF measurements at Goesdorf, Kaundorf and Nocher-Route. A total of 91 two–hour composite samples were taken and measured in the laboratory for determination of concentrations of COD [$mg \cdot l^{-1}$] and NH_4 [$mg \cdot l^{-1}$]: 7 at Goesdorf on 4 May 2011, 48 between 19 June and 21 July 2010 at Kaundorf, and 36 between 9 March and 2 August 2011 at Nocher-Route. The variables COD_f and $NH4_f$ were set to zero because the pollution contribution of the infiltration water is negligible in the study area. The contribution of ammonium from rainwater $NH4_r$ was assumed constant and set to 2.00 [$mg \cdot l^{-1}$],

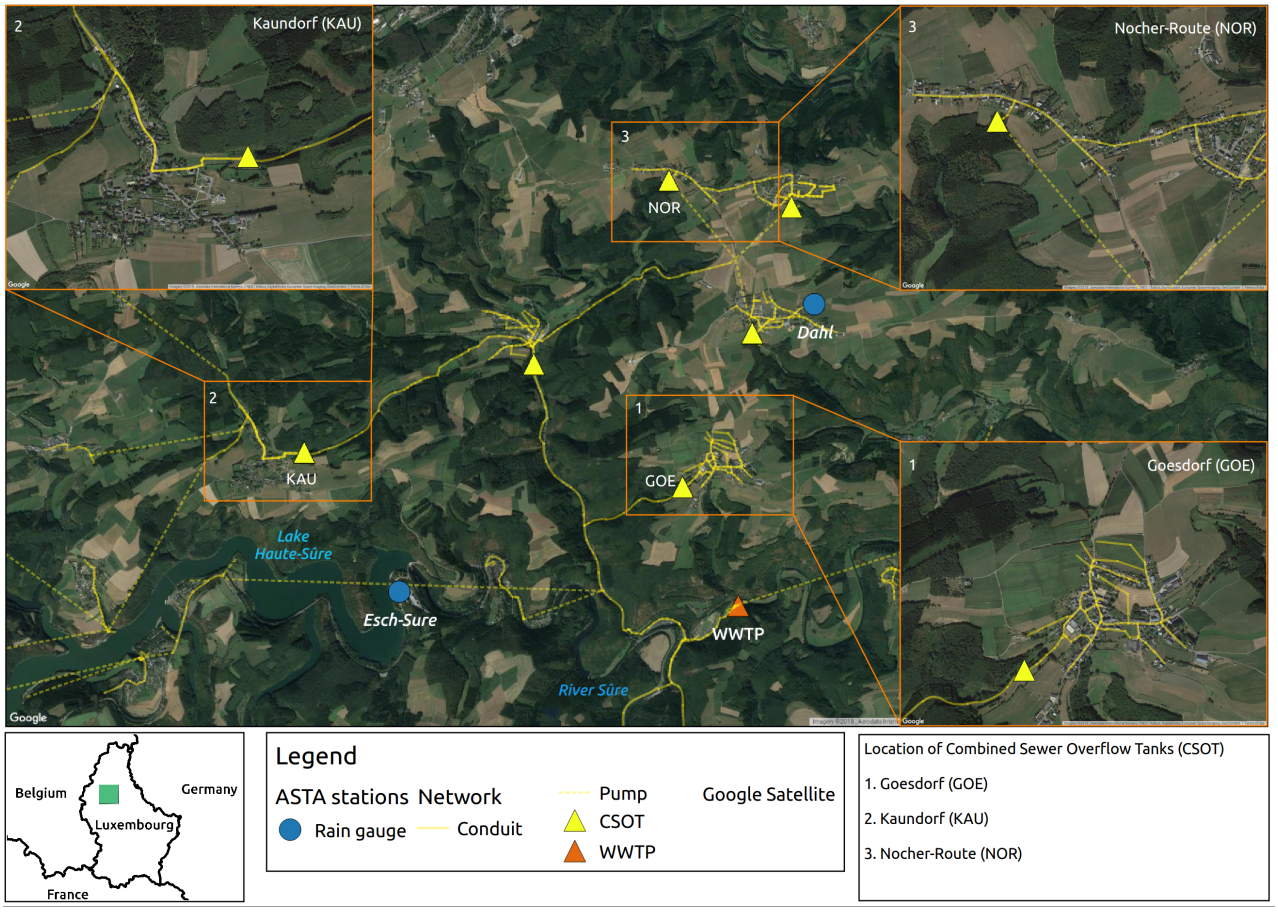

Figure 2. The three Haute-Sûre sub-catchments and locations of CSOT structures considered in this study. The background map is provided by © Google Maps.

while COD_r was equal to zero. Table 1 summarises the base values of the general input variables and Table 2 presents the base values of input variables for each individual CSO. These base values were used when running EmiStatR in deterministic mode (see Section 3.1). Some of the variables were calibrated based on observations in the CSOT to simulate water level and concentrations and loads of pollutants spilled in the CSO to the stream, river or lake. These variables are water consumption (q_S), infiltration flow (q_f), time flow (t_{fS}), run-off coefficient for impervious area (C_{imp}), run-off coefficient for pervious area (C_{per}), orifice coefficient of discharge (C_d) and initial water level (lev_{ini}).

2.4 Selection of model input for uncertainty quantification

Following recommendations from Nol et al. (2010), not all model inputs were taken into account in the uncertainty propagation analysis. Only inputs that are very uncertain and to which the model output is very sensitive were included because these are the ones that have the largest contribution to output uncertainty (Heuvelink (1998), Section 4.4). The level of uncertainty of the

Table 1. ~~General and~~ Most important general, CSO input ~~,~~ and output variables of ~~EmiStatR. Base~~ EmiStatR, with base values for the general input variables.

General input	Units	Base value	CSO input
1. Wastewater			*1. ~~Identification~~Catchment data*
Water consumption, qs	$[\mathrm{l(PE^a \cdot d)^{-1}}]$	152	~~ID of the structure~~ Total area, A_{total}
Pollution COD[b], $C_{COD,S}$	$[\mathrm{g(PE \cdot d)^{-1}}]$	104.2	~~Name of the structure~~ Impervious area, A_{imp}
Pollution NH_4[c], $C_{NH4,S}$	$[\mathrm{g(PE \cdot d)^{-1}}]$	4.7	~~2. Catchment data2. Infiltration waterName of the municipality - Inflow, q~~
~~Pollution COD, COD_r~~ *2. Infiltration water*	~~$[\mathrm{mg \cdot l^{-1}}]$~~	~~71.0~~	Run-off coeff. for pervious area, C_{per}
~~Pollution NH_4, $NH4_r$~~ Inflow, q_f	~~$[\mathrm{mg \cdot l^{-1}}]$~~ $[\mathrm{l(s \cdot ha)^{-1}}]$	~~2.0~~ 0.116	Flow time structure, t_{fS}
Pollution COD, COD_f	$[\mathrm{g(PE \cdot d)^{-1}}]$	0	Population equivalents, pe
Pollution NH_4, $NH4_f$	$[\mathrm{g(PE \cdot d)^{-1}}]$	0	~~3.~~ 2. *CSO structure data*
3. Rainwater			Volume, V
~~Curve level – volume, $lev2vol$~~ Rain time series, P	[~~m, m^3~~mm]		Initial water level, Lev_{ini}
Pollution COD, COD_r	$[\mathrm{mg \cdot l^{-1}}]$	71.0	Maximum throttled outflow, $Q_{d,max}$
Pollution NH_4, $NH4_r$	$[\mathrm{mg \cdot l^{-1}}]$	2.0	Orifice diameter, D_d
			Orifice coefficient of discharge, C_d
Output variables			
1. Quantity			
Volume in the CSO chamber, $V_{Chamber}$	[m^3]		
Overflow spill volume, V_{Sv}	[m^3]		
Overflow spill flow, Q_{Sv}	$[\mathrm{l \cdot s^{-1}}]$		
2. Quality			
Spill COD load, $B_{COD,Sv}$	[g]		
Average spill COD conc.[e], $C_{COD,Sv,av}$	$[\mathrm{mg \cdot l^{-1}}]$		
99.9th perc..[f] spill COD conc., $C_{COD,Sv,99.9}$	$[\mathrm{mg \cdot l^{-1}}]$		
Maximum overflow COD conc., $C_{COD,Sv,max}$	$[\mathrm{mg \cdot l^{-1}}]$		
Spill NH_4 load, $B_{NH4,Sv}$	[g]		
Average spill NH_4 conc., $C_{NH4,Sv,av}$	$[\mathrm{mg \cdot l^{-1}}]$		
99.9th perc. spill NH_4 conc., $C_{NH4,Sv,99.9}$	$[\mathrm{mg \cdot l^{-1}}]$		
Maximum spill NH_4 conc., $C_{NH4,Sv,max}$	$[\mathrm{mg \cdot l^{-1}}]$		

[a]PE = population equivalents; [b]COD = chemical oxygen demand; [c]NH_4 = ammonium;
[d]coef. = coefficient; [d]conc. = concentration; [f]perc. = percentile.

Table 2. The CSO structure input data for the EmiStatR model, after calibration. ~~For structures~~ Structures 2 and 3, only C_d was calibrated.

CSO input			
1. Identification			
ID of the structure	1	2	3
Name of the structure	FBH Goesdorf	FBN Kaundorf	FBH Nocher-Route
2. Catchment data			
Name of the municipality	Goesdorf	Kaundorf	Nocher-Route
Name of the catchment	Haute-Sûre	Haute-Sûre	Haute-Sûre
Number of the catchment	1	1	1
Land Use[a]	R/I	R/I	R/I
Total area, A_{total} [ha]	30.0	22	18.6
Impervious area, A_{imp} [ha]	5.0	11.0	4.3
Run-off coefficient for impervious area, C_{imp} [-]	0.28	0.3	0.3
Run-off coefficient for pervious area, C_{per} [-]	0.07	0.10	0.10
Flow time structure, t_{fS} [~~min~~time step]	1	2	2
Population equivalents, pe [PE]	611	358	326
3. CSO structure data			
Volume, V [m^3]	190	180	157
~~Curve level – volume, $lev2vol$ m, m^3 Goesdorf Kaundorf Nocher-Route~~ Initial water level, Lev_{ini} [m]	0.57	1.8	1.8
Maximum throttled outflow, $Q_{d,max}$ [$l \cdot s^{-1}$]	5.0	9	4
Orifice diameter, D_d, [m]	0.15	0.20	–
Orifice coefficient of discharge, C_d [-]	0.67	0.67	0.67

[a] R = residential, I = industrial.

inputs was defined by expert judgement and similar case studies in the literature. A quick-scan was used to determine the model sensitivity to each of the model inputs, by running EmistatR in deterministic mode with input base values given in Table 1. The level of model sensitivity was defined by analysing the mathematical model structure and components of the model, expert judgement and simulations with EmiStatR. Inputs that rank high on both the level of uncertainty and on model sensitivity were selected and included in the uncertainty propagation analysis.

2.5 Uncertainty quantification of selected model input

Because we used a statistical approach, probability distribution functions (pdfs) are the basis to represent uncertainties of the selected model inputs. This constitutes the most difficult step of an uncertainty propagation analysis and is done in different ways for constants and dynamic variables, as explained in the following sub-sections.

2.5.1 Uncertain constants

Following Heuvelink et al. (2007), an uncertain continuous numerical constant C can be characterised by its marginal (cumulative) pdf (mpdf):

$$F_C(c) = P(C \leq c) \tag{6}$$

Usually a parametric approach can be taken, meaning that a common shape for F_C is chosen (e.g., normal, lognormal, exponential, uniform) so that the mpdf is reduced to a number of parameters. In this study, the input variables that are in this category are: water consumption (qs), infiltration inflow (q_f), total area (A_{total}), impervious area (A_{imp}), the run-off coefficients for impervious area (C_{imp}) and pervious area (C_{per}), population equivalents (pe), flow time structure (t_{fS}), and initial water level (Lev_{ini}).

2.5.2 Univariate autoregressive modelling

Dynamic uncertain inputs may be temporally autocorrelated. This may dramatically influence the outcome of an uncertainty propagation analysis and must therefore be accounted for. One way of doing this is by assuming an autoregressive order one (AR(1)) model:

$$y_t = \mu + \phi\,(y_{t-1} - \mu) + w_t, \qquad t = 1, 2, ..., T, \quad y_0 \sim \mathcal{N}(\mu, \sigma^2) \tag{7}$$

where y_t is the uncertain input at time t, μ is its mean, ϕ is the autoregressive parameter ($0 \leq \phi < 1$), and w_t is a Gaussian white noise time series with mean zero and variance σ_w^2. The initial value y_0 is taken from a normal distribution with mean μ and variance σ^2. The parameters of the model can be estimated based on observations, or in absence of observations, suitable values are taken based on expert judgment or literature reference values. Note that the effect of the initial condition usually fades out quickly and hence is not of important concern.

The implementation of the AR(1) model in R was done via the R function arima.sim of the R base package stats (R-Core-Team and contributors worldwide, 2017), both for model calibration and simulation.

2.5.3 Multivariate autoregressive modelling

In case of multiple uncertain dynamic inputs, cross-correlation between these inputs may also need to be included. For example, $C_{COD,S}$ and $C_{NH4,S}$ and their uncertainties are likely correlated. This can be done using a multivariate AR(1) model (Luetkepohl, 2005), which is a natural extension of the univariate AR(1) model:

$$Y(t+1) = \boldsymbol{\mu} + \mathbf{A} \cdot [Y(t) - \boldsymbol{\mu}] + \varepsilon(t), \qquad t = 1, 2, ..., T, \quad Y_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \tag{8}$$

where $Y(t)$ is a vector of inputs at time t, $\mathbf{A}$ is a square matrix with parameters that define how the variables at time $t+1$ depend on those at time t, $\boldsymbol{\mu}$ is now a vector of means and $\varepsilon(t)$ a vector of zero-mean, normally distributed white noise processes. We further assume that the variance-covariance matrix $\mathbf{C}$ of $\varepsilon(t)$ is time-invariant. The initial value Y_0 is assumed

normally distributed and uncorrelated ($\mathbf{\Lambda}$ is a diagonal matrix). In order to estimate the vector $\boldsymbol{\mu}$ and matrices $\mathbf{A}$ and $\mathbf{C}$, a sample of the variables of interest is needed. Parameter estimation is done by means of the R-package mAr (Barbosa, 2015).

2.5.4 Input precipitation model

In case precipitation is selected as an uncertain input to be included in the uncertainty analysis, then it too must be characterised by a pdf. Since precipitation, however, is not normally distributed and has many zeros, we cannot make the Gaussian assumption and hence we cannot use the approach described in Section 2.5.2 to model its dynamic behaviour and uncertainty. In addition, we usually have precipitation measurements nearby so we need to condition the simulations to these measurements. Recall from Section 2.3 that in the case study precipitation data are recorded at stations Esch-sur-Sûre and Dahl.

Torres-Matallana et al. (2017) present a model to simulate precipitation inside a target catchment given a known precipitation time series in a nearby location outside the catchment, while accounting for the uncertainty that is introduced due to spatial variation in precipitation. The method used for input precipitation uncertainty characterisation is essentially the same as the application of a Kalman filter/smoother (Kalman, 1960; Webster and Heuvelink, 2006). Calibration of ~~that~~ the model requires precipitation time series at two locations near the catchment of interest. ~~We briefly summarise the method here. We denote the measured time series of precipitation at the first location as $P_1(t)$ and that at the second location as $P_2(t)$.~~ Once the model is calibrated, it is used to simulate precipitation inside the target catchment from a single precipitation time series nearby the catchment.

~~**Calibration**~~

~~We begin by relating the two precipitation time series as:~~

$$P_1(t) = P_2(t) \cdot \delta(t)$$

~~where $\delta(t)$ is a positive multiplicative factor that varies over time. We assume that $P_1(t)$, $P_2(t)$ and $\delta(t)$ are stationary and log-normally distributed stochastic processes. After log-transformation we get~~

$$\log(P_1(t)) = \log(P_2(t)) + \log(\delta(t))$$

~~We apply a Kernel (daniell) smoothing to the precipitation time series to avoid rapid fluctuation of the time series for precipitation depth values smaller than 0.1 mm. This also solves problems associated with taking logarithms of near-zero values. Next, in order to estimate the parameters of $\delta(t)$, we filter the time series allowing the computation of a ratio between the two measured time series. This ratio represents the difference in precipitation as registered in two nearby rain gauge stations. It is computed only for those cases where the precipitation depth of the two time series is greater than 0.01 mm.~~

To simplify notation we write $LP_1(t) = \log(P_1(t))$, $LP_2(t) = \log(P_2(t))$ and $L\delta(t) = \log(\delta(t))$. Since two out of three determine the third, we need only define two processes. We model the joint distribution of $LP_1(t)$ and $L\delta(t)$ by a bivariate AR(1) process, as introduced before:

$$\begin{bmatrix} LP_1(t+1) \\ \\ L\delta(t+1) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \\ \mu_\delta \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ \\ B_{21} & B_{22} \end{bmatrix} \left(\begin{bmatrix} LP_1(t) \\ \\ L\delta(t) \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \\ \mu_\delta \end{bmatrix} \right) + \begin{bmatrix} \varepsilon_1(t+1) \\ \\ \varepsilon_\delta(t+1) \end{bmatrix}$$

where ε_1 and ε_δ are zero-mean, cross-correlated and normally distributed white noise processes.

To calibrate this model, i.e. estimate its parameters μ_1, μ_δ, B_{11}, B_{12}, B_{21}, B_{22}, σ_1^2, σ_δ^2 and $\rho_{1\delta}$, where $\sigma_1^2 = \text{var}(\varepsilon_1)$, $\sigma_\delta^2 = \text{var}(\varepsilon_\delta)$ and $\rho_{1\delta}$ is the correlation between ε_1 and ε_δ, we used the Rpackage mAr (Barbosa, 2015). Calibration is based on two time series of LP_1 and $L\delta$ derived from observed time series P_1 and P_2.

Conditional simulation

To simulate a time series P for the target catchment from an observed time series P_o at a nearby location, we make use of the fact that the calibrated AR(1) model quantifies how precipitation at one location relates to that at a nearby location. We make use of Eq. **??**:

$$P(t) = P_o(t) \cdot \delta(t)$$

This requires simulations of $\delta(t)$. These are obtained using the calibrated model Eq. **??**, but now applied to the vector $[LP_o \; L\delta]^T$, which characterises the joint pdf of LP_o and $L\delta$. We use this model to simulate $L\delta$ conditional to the observed time series LP_o. Since the two processes are jointly normally distributed we can make use of a well known property of the multivariate normal distribution (**?**, page 47). Let U and V be two jointly normally distributed random vectors. The conditional distribution of U given $V = v$ is then also normal and given by:

$$(U|V) = v \sim N E[U] + cov(U,V) \cdot var(V)^{-1} \cdot (v - E[V]), \;\; var(U) - cov(U,V) \cdot var(V)^{-1} \cdot cov(V,U))$$

We make use of this equation to simulate δ by substituting:

$$U = L\delta(t+1) \qquad V = \begin{bmatrix} L\delta(t) \\ LP_o(t+1) \\ LP_o(t) \end{bmatrix}$$

for all $t = 1, ..., T$, while substituting the observed time series LP_o for v. For details we refer to Torres-Matallana et al. (2017)

Details of the calibration and conditional simulation are presented in section S3 of the Supplementary Material.

2.6 Uncertainty ~~propagation~~analysis

~~Various methods can be used to analyse uncertainty propagation. ? summarised the characteristics of the main methods, which range from deterministic methods such as minimum/maximum to hydrid methods as First- and Second-Order Reliability Methods (FORM/SORM), considering as well the Taylor series approximation and Monte Carlo (MC) simulation.~~ We used MC simulation (Hammersley and Handscomb, 1964; Kalos and Whitlock, 2008) to analyse how input uncertainty propagates through the EmiStatR model, because it is flexible and straightforward to implement. It is also feasible in our case study because EmistatR is a relatively simple model that does not involve a long computation time.

2.6.1 Monte Carlo simulation

The MC method runs the EmiStatR model repeatedly, each time using different model input values, sampled from their pdf. The method thus consists of the following steps:

1. Repeat n times:
   (a) Generate a set of realisations of the uncertain model inputs at 10 min resolution
   (b) For this set of realisations, run the model at 10 min resolution and store output. Later, in order to compute the summary statistics, a temporal aggregation of the model output to one hour intervals is done.
2. Compute and store sample statistics from the n model outputs.

Here, n is the number of MC runs, i.e. the MC sample size. Common sample statistics that measure the uncertainty are the standard deviation and quantiles of the distribution of MC outputs, such as the difference between the 0.95 and % 0.05 quantile, which can be easily calculated from the n Monte Carlo outputs.

Sampling from the pdf of uncertain inputs was done using simple random sampling.

2.6.2 Monte Carlo output summary

Proper presentation of MC outputs is important to get the most out of the experiment. Therefore, summary statistics are one important way to summarise the MC outputs. Commonly, a MC study yields n model outputs, which are stored in the MC result matrix $\mathbf{X}$ in Boos and Osborne (2015). From this matrix, various statistics can be computed. Basic summary statistics include the mean μ_{MC}, the standard deviation (σ_{MC}) and the variance σ^2_{MC}. From these we can compute the coefficient of variation CV_{MC} ($\sigma_{\text{MC}}/\mu_{\text{MC}}$), which is a dimensionless expression of relative uncertainty. The coefficient of variation is a standardised measure of the spread of a sampling distribution, being useful because it allows to directly compare variation in samples with different units, or with very different means (Marwick and Krishnamoorthy, 2019). We computed estimates and standard errors for these statistics and also for the interquartile range (IQR_{MC}), 0.005 ($\zeta_{0.005}$) and 0.995 ($\zeta_{0.995}$) quantiles, and the 99% width of the prediction band ($\zeta_{w,0.99}$).

2.6.3 Bootstrap computation for Monte Carlo summary

Following Boos and Osborne (2015) “Good statistical practice dictates that summaries in MC studies should always be accompanied by standard errors", we used the bootstrap method to compute standard errors of all MC statistics. These tools are particularly relevant in a case without analytic solutions (Boos, 2003). According to Boos and Osborne (2015, p. 228) standard errors for MC output statistics are often not computed, being an additional computational step on top of the overall analysis. Standard errors are straightforward to compute for simple statistics such as the sample mean over the replications of the MC output, but are more difficult to compute for more complex statistics, such as medians, sample variances and the classical Pearson measures of skewness and kurtosis. Therefore, to avoid burdensome computations we opted to compute the standard errors by the bootstrap method. We briefly explain the bootstrap method below. For a detailed explanation we refer to Efron (1979).

To compute the bootstrap variance of estimators we follow the logic given by Boos and Osborne (2015). From a MC sample $Y_1, ..., Y_n$, we draw a random sample of size n with replacement $Y_1^*, ..., Y_n^*$, and compute an estimator $\hat{\theta}$ of the MC statistic θ from this resample. We independently repeat this process B times, resulting in a sample of estimators $\hat{\theta}_1, ..., \hat{\theta}_B$. Then the bootstrap variance estimate, $\hat{V}_B$, is the sample variance of this sample of estimators:

$$\hat{V}_B = \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}_i - \bar{\hat{\theta}} \right)^2 \tag{9}$$

where $\bar{\hat{\theta}}$ is the mean of the sample of estimators. The MC standard error, se, is simply the square root of the bootstrap variance.

We implemented in stUPscales (Torres-Matallana et al., 2019) specific routines for computing, by means of the bootstrap method, the MC estimators and their standard error for all MC statistics, where the variance of the model output is the most important. We compared our results with the results obtained using the Monte.Carlo.se R-package (Boos et al., 2019).

2.6.4 Contributions of input variables to total uncertainty

A number of $m+1$ MC analysis are needed to compute the contributions of input variables to total uncertainty, where m is the number of model input variables selected for uncertainty quantification. The first MC analysis, MC_{tot}, is done to compute the total output uncertainty by varying stochastically all input variables. The uncertainty associated with the first variable x_1 is quantified by a second MC analysis MC_1, in which only x_1 is equal to its deterministic value, while the other input variables vary stochastically. Similarly, the other MC simulations MC_2, MC_3, ..., MC_m are used to quantify the uncertainty for the variables x_2, x_3, ..., x_m.

To quantify the contributions of individual input variables to the total uncertainty of the model inputs, the stochastic sensitivity S_i for each uncertain input x_i is computed. The first-order stochastic sensitivity index S_i is defined as (Saltelli et al., 2008, p. 160-161):

$$S_i = \frac{\text{Var}(\text{MC}_i)}{\text{Var}(\text{MC}_{tot})} \tag{10}$$

The first-order stochastic sensitivity index represents the main effect contribution of each input factor to the total variance of the output. The larger the index, the more important the input uncertainty. We computed stochastic sensitivity indices per time step and aggregated contributions for the whole year. For plotting purpose, we aggregated the outputs from 10 minutes time step to hourly time steps. The aggregation was done for each individual MC run before the contributions were computed.

2.7 Water quality impact

The results of the MC uncertainty propagation were also compared with the water quality standards. Standards are introduced to evaluate the impact of emissions of COD and NH_4 in CSOs into the receiving water. However, as Toffol (2006) recognises, although there are European emission standards for wastewater treatment plant effluent, standards for combined sewer overflow are not so clear. According to Steinel and Margane (2011), the European Water Framework Directive (WFD) is mainly concerned with the natural state of waters. Therefore, emission standards for effluent discharge are not set. The EU Directive 91/271/EEC (1991) sets standards for COD and total Nitrogen, hence similar values have been adopted in many European member states. For more details about guidelines and design procedures in Europe see Blumensaat et al. (2012). We assessed the emissions accordingly to the German guideline ATV-A 128 (1992), which is the standard for dimensioning and design of stormwater structures in combined sewers and commonly used in Luxembourg. The Austrian ÖWAV-RB 19 (2007) is also taken into account because it provides key reference guidelines for design of urban water infrastructure in central Europe. Three main indicators are taken into account: hydraulic impact, COD concentration, and acute ammonium toxicity.

2.7.1 Hydraulic impact

According to the Austrian guidelines and as summarised by Kleidorfer and Rauch (2011), the evaluation of the hydraulic impact is given by:

$$Q_1 \leq f_h \cdot Q_{r1} \tag{11}$$

where $0.1 \leq f_h \leq 0.5$, Q_1 [l·s^{-1}] is the maximum sewer overflow discharge with return period one year, and Q_{r1} [l·s^{-1}] is the maximum water discharge in the river with return period once per year. The factor f_h is taken as 0.1 in more sensitive streams, whereas it is 0.5 for streams with more stable bed and higher re-colonisation potential Toffol (2006). Time series of daily values recorded in 2006 to 2013 of the river Sûre at Heiderscheidergrund were used to compute the daily flow expected with return period once per year (1.01 years), Q_{r1}.

~~According to~~ We used the German guideline ATV-A 128 (1992), which computes the throttle discharge at CSOs, ~~$Q_{t,CSO}$ l·s^{-1}, may be computed using~~ $Q_{t,CSO}[l \cdot s^{-1}]$ as:

$$Q_{t,CSO} = f_t \cdot A_{imp} \tag{12}$$

where $7.5 \leq f_t \leq 15$ and A_{imp} [ha] is the impervious area connected to the combined sewer system. For the overflow flow MC output mean, 0.95 quantile and 0.995 quantile, we computed the exceedance percentage over the thresholds, calculated as the proportion of time steps exceeding the number of total time steps in the year (8,759 time steps at 1-hour).

2.7.2 COD concentration

Steinel and Margane (2011, Table 14) presents the effluent standards for discharging into freshwater adopted in selected European countries. A COD concentration of 125 [mg $\cdot$ l^{-1}] is reported for the European Union countries. Austria has stricter rules with a standard of 90 [mg $\cdot$ l^{-1}] for populations between 50 and 500 inhabitants, and 75 [mg $\cdot$ l^{-1}] for populations greater than 500 inhabitants. The Goesdorf population by 2001 was 1,025 inhabitants and by 2011 was 1,297 inhabitants (Statec, 2020). For NH_4 a similar approach was used.

2.7.3 Acute ammonium toxicity

Following Kleidorfer and Rauch (2011), "*the ammonia (NH_3) concentration depends on the ammonium (NH_4) concentration and on the dissociation equilibrium between NH_3 and NH_4 (which is influenced by temperature and pH-value)*". According to Kleidorfer and Rauch (2011), the Austrian guideline ÖWAV-RB 19 (2007) establishes a maximum value of 2.5 mg $\cdot$ l^{-1} for the ammonium (NH_4) concentration calculated for one hour duration for salmonid streams. For cyprinid streams a maximum value of 5.0 mg $\cdot$ l^{-1} is recommended.

3 Results

3.1 Selection of model inputs for uncertainty quantification

In this section we assess the degree of uncertainty and sensitivity for all input variables, following the procedure described in Section 2.4. We summarise the results in Tables 3 and 4. To better support our decisions we also include a graphical assessment of the degree of uncertainty and sensitivity of each input, as in Tscheikner-Gratl et al. (2017). See Figure S1 in the Supplementary Material.

3.1.1 Wastewater

Water consumption, qs, is a fairly uncertain input variable and the model output is sensitive to this variable. Volume and flow of CSO are sensitive to changes in qs. Regarding water quality output, total load of NH_4 is very sensitive to changes in qs. Pollution of sewage as COD load per capita per day, $C_{COD,S}$ is the first selected input variable for propagation of uncertainty, due to the fact that it is both a very uncertain input variable and the model output (average and 99.9 percentile overflow COD concentration) is very sensitive to it. Pollution of sewage as NH_4 load per capita per day, $C_{NH4,S}$, is also included in the uncertainty propagation analysis. It is a very uncertain input variable and the model output (overflow load and concentrations of NH_4) is very sensitive to it. The variables $C_{COD,S}$ and $C_{NH4,S}$ are very uncertain because these are correlated to the temporal and spatial pattern of water consumption, which has a daily, weekly and seasonal temporal variability.

3.1.2 Infiltration water

Inflow of infiltration water, q_f is a very uncertain input variable because this inflow depends of the number of anomalies in the pipes (cracks or wrong connections) that allow infiltrations flowing into and out of the system. The distribution of these anomalies has a strong random component and hence q_f is very uncertain, and model output is sensitive to it. Although this is a very uncertain input, the quick-scan analysis showed that model output sensitivity is not very high as is indicated in Table 3. For this reason we did not include this variable in the uncertainty propagation analysis.

Pollution of infiltration water as COD load per capita per day, COD_f and pollution of infiltration water as NH_4 load per capita per day, $NH4_f$ are not uncertain because in the Haute-Sûre study area the values of these variables are negligibly small.

3.1.3 Rainwater

Precipitation, P is the main driving force of the model and given the spatial variability of the rain fields, this input is considered very uncertain. The model output, additionally, is very sensitive to it. As a consequence, this input variable is treated as the third input variable in the uncertainty propagation analysis. Pollution of runoff as COD concentration, COD_r is the fourth input variable considered in the uncertainty propagation, given that it is a very uncertain and very sensitive input variable, particularly to load and concentration of COD in the overflow. ~~Pollution of runoff~~

Pollution in rainwater as NH_4 concentration, $NH4_r$ is considered fairly uncertain. The model output (overflow load and concentration of NH_4) is very sensitive to it. Although model output is very sensitive to this model input variable, model input uncertainty is not very high as is indicated in Table 3. For this reason it was not included in the uncertainty propagation analysis.

3.1.4 Sub-catchment

The model is very sensitive to the total area A_{total} and to the run-off coefficient for pervious area (C_{per}) and sensitive to the impervious area A_{imp} and to the run-off coefficient for impervious area (C_{imp}). However, we did not include A_{total} ~~and C_{per}~~ in the uncertainty analysis because ~~these~~ it can be fairly accurately derived from spatial databases and hence their uncertainty is not large. Although model output is very sensitive to the input variable C_{per}, the uncertainty about this variable is not very high, as indicated in Table 3. The reason behind this is that C_{per} can be derived fairly accurately from GIS products, such as land use and soil type maps. Therefore, we did not include this variable in the uncertainty propagation analysis. The population equivalents pe is a sensitive variable but not very uncertain. Hence this variable was not included in the uncertainty analysis. The theoretical largest flow time in the catchment t_{fS} is not uncertain and not sensitive.

3.1.5 CSO structure

Although model output is very sensitive to maximum throttled outflow $Q_{d,max}$ and volume V, these are not included in the uncertainty analysis because their values are accurately known. The same is true for the variables curve level - volume $lev2vol$, orifice diameter D_d and discharge coefficient C_d. These variables are accurately known and therefore not considered

Table 3. Results of deterministic sensitivity analysis. Average percentage of change of model output caused by ±10% change in model inputs (qs, $C_{COD,S}$, $C_{NH4,S}$, COD_r, pe, and P as time series, VAR(1) model for $C_{COD,S}$ and $C_{NH4,S}$, AR(1) model for COD_r and AR(1) conditioned for P. See Table 1 for nomenclature definition). Output change greater than 15% is considered very high. Variable C_d (not shown in the table) leads to a percentage of change less than 0.3%, while variables t_{fS} and Lev_{ini} (not shown in the table) lead to no change in the output. Values greater than 15 are shown in bold font.

	Input variable														
Output variable	qs	$C_{COD,S}$	$C_{NH4,S}$	q_f	COD_r	$NH4_r$	A_{total}	A_{imp}	C_{imp}	C_{per}	pe	$Q_{d,max}$	V	D_d	P
$V_{Chamber}$	4.2	0.0	0.0	3.0	0.0	0.0	8.6	7.1	5.9	7.2	4.2	**16.7**	7.5	0.8	13.4
V_{Sv}	2.9	0.0	0.0	1.7	0.0	0.0	**19.6**	11.6	13.5	**16.1**	2.9	13.1	**16.5**	0.2	**17.8**
Q_{Sv}	0.6	0.0	0.0	0.2	0.0	0.0	2.7	1.5	1.1	1.8	0.6	13.1	14.8	0.6	12.4
$B_{COD,Sv}$	2.9	2.4	0.0	1.7	7.7	0.0	**20.1**	11.8	13.7	**16.6**	5.3	**15.7**	14.5	0.2	**20.7**
$C_{COD,Sv,Av}$	0.7	4.6	0.0	0.5	5.4	0.0	0.5	0.6	0.6	0.7	4.0	3.2	4.0	0.2	1.1
$C_{COD,Sv,99.9}$	1.6	6.7	0.0	0.8	3.4	0.0	2.8	2.3	1.9	2.4	5.1	0.0	0.0	0.0	9.2
$C_{COD,Sv,Max}$	1.6	6.7	0.0	0.8	3.4	0.0	2.8	2.3	1.9	2.4	5.1	0.0	0.0	0.0	9.2
$B_{NH4,Sv}$	3.1	0.0	3.3	1.8	0.0	6.7	**20.4**	12.0	13.8	**16.8**	6.4	**17.0**	13.4	0.3	**22.1**
$C_{NH4,Sv,Av}$	0.9	0.0	5.8	0.6	0.0	4.2	0.6	0.8	0.9	0.9	5.3	4.3	5.4	0.2	1.5
$C_{NH4,Sv,99.9}$	1.6	0.0	7.6	0.8	0.0	2.4	3.5	2.6	2.3	2.9	6.1	0.0	0.0	0.0	11.3
$C_{NH4,Sv,Max}$	1.6	0.0	7.6	0.8	0.0	2.4	3.5	2.6	2.3	2.9	6.1	0.0	0.0	0.0	11.3

as uncertain variables. The initial water level in the chamber Lev_{ini} is very uncertain but the model output is not sensitive to this variable. Therefore, Lev_{ini} was not included in the uncertainty analysis.

3.2 Uncertainty quantification of selected model input

After ~~evaluation of the model output sensitivityand taking into account the degree of uncertainties of each input~~ranking all inputs on level of uncertainty and model sensitivity, we selected four input variables to be included in the uncertainty analysis. These are $C_{COD,S}$, $C_{NH4,S}$, COD_r and P (Table 4).

3.2.1 Sewage per capita COD and Ammonium

The fit of pdfs for the two uncertain inputs $C_{COD,S}$ and $C_{NH4,S}$ was based on measurements under dry weather flow conditions. Measurement campaigns were done in Goesdorf from 28th April to 24th June 2011, in Kaundorf from 22nd June to 18th August in 2010 and from 20th July to 5th August in 2011, and in Nocher-Route from 18th November 2010 to 27th April 2011. Samples of COD and NH_4 in mg·l^{-1} (91 in total for each variable) were analysed. An average wastewater amount was calculated for Goesdorf (153 l·PE^{-1}·d^{-1}), Kaundorf (112 l·PE^{-1}·d^{-1}) and Nocher-Route (94.3 l·PE^{-1}·d^{-1}). Table 5 presents summary statistics of the dry weather flow measurements of COD and NH4 and the corresponding value of $C_{COD,S}$ and

Table 4. Input variables of the EmiStatR model and selection of inputs for uncertainty analysis based on input uncertainty level and model sensitivity level (legend: from ++ very uncertain/sensitive to – – not uncertain/sensitive).

Input variable	Input uncertainty	Model sensitivity	Uncertainty analysis
Wastewater			
1. qs	+	+	no
2. $C_{COD,S}$	++	++	**yes**
3. $C_{NH4,S}$	++	++	**yes**
Infiltration water			
4. q_f	++	+	no
5. COD_f	– –	– –	no
6. $NH4_f$	– –	– –	no
Rainwater			
7. P	++	++	**yes**
8. COD_r	++	++	**yes**
9. $NH4_r$	+	++	no
Sub-catchment			
10. A_{total}	+	++	no
11. A_{imp}	+	+	no
12.C_{imp}	+	+	no
13. C_{per}	+	++	no
14. pe	+	+	no
15. t_{fS}	–	– –	no
CSO structure			
16. $Q_{d,max}$	–	++	no
17. V	–	++	no
18. D_d	– –	– –	no
19. C_d	– –	– –	no
20. Lev_{ini}	++	– –	no

$C_{NH4,S}$. COD is converted to $C_{COD,S}$ by means of a simple conversion from mg·l^{-1} to g·PE^{-1}·d^{-1}, by multiplying COD by the measured per capita flow (112 l·PE^{-1}·d^{-1}) and dividing by 1,000. NH_4 was converted to $C_{NH4,S}$ in a similar way.

Table 5. Summary statistics of dry weather flow measurements for $C_{COD,S}$ and $C_{NH4,S}$ characterisation.

	COD [mg·l^{-1}]	$C_{COD,S}$ [g·PE^{-1}·d^{-1}]	log($C_{COD,S}$) log(g·PE^{-1}·d^{-1})	NH4 [mg·l^{-1}]	$C_{NH4,S}$ [g·PE^{-1}·d^{-1}]	log($C_{NH4,S}$) log(g·PE^{-1}·d^{-1})
Min	61.9	6.9	1.936	16.10	1.745	0.556
P5	216.8	23.8	3.167	20.55	2.102	3.018
Mean	925.5	104.2	4.378	44.38	4.733	1.473
P95	2032.0	236.8	5.466	79.00	7.684	2.039
Max	3454.0	528.5	6.270	81.20	10.771	2.377
St. deviation	631.7	87.5	0.751	18.56	1.917	0.410

Closer inspection showed that $C_{COD,S}$ and $C_{NH4,S}$ observations are best characterised by a lognormal distribution (~~Fig. ??~~Supplementary Material, Section S4). Since $C_{COD,S}$ and $C_{NH4,S}$ are dynamic and cross-correlated, we calibrated a bivariate AR(1) model with state vector $Y = [\log(C_{COD,S}) \;\; \log(C_{NH4,S})]^T$. The estimated parameters of the model using the methodology described in Section 2.5.3 are:

$$\boldsymbol{\mu} = \begin{bmatrix} 4.40947 \\ 3.70411 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 0.99165 & -0.00319 \\ -0.00009 & 0.99455 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 0.00913 & 0.00224 \\ 0.00224 & 0.00185 \end{bmatrix} \tag{13}$$

The defined multivariate autoregressive model also capture the dynamic behaviour, temporal correlation and cross-correlation of the input variables, deriving the probability distributions of $C_{COD,S}$ and $C_{NH4,S}$ from measurements in the Haute-Sûre catchment, which agreed well with values reported in the literature (Katukiza et al., 2014; Heip et al., 1997).

3.2.2 Runoff COD concentration

Regarding COD_r, due to the fact that no field measurements were available, expert judgement and reference values from the literature were the basis to characterise the pdf of this input variable. The variable was assumed to be lognormally distributed with a mean value of 71 [mg·l^{-1}]. Although, House et al. (1993) and Welker (2008) reported a higher value, 107 [mg·l^{-1}] for COD_r, we selected a lower value due to the specific characteristics of the CSO system in the Haute-Sûre catchment. The value of 150 [mg·l^{-1}] as standard deviation of COD_r leads to a coefficient of variation (sd·mean^{-1}) equal to 2.11, which is greater than the coefficient of variation for $C_{COD,S}$ (0.84). We allow the standard deviation of COD_r to be greater than the standard deviation of $C_{COD,S}$, because COD measurements in rain water are very uncertain.

~~Histogram of observations, empirical density (red dashed line) and theoretical normal density (blue line) for (a) $log(C_{COD,S})$; (b) $log(C_{NH4,S})$; (c) $log(COD_r)$.~~

3.2.3 Input precipitation model

Precipitation and its associated uncertainty was modelled as an autoregressive model conditioned to the observed precipitation at a nearby measurement station. We assumed a multivariate lognormal distribution and included temporal correlation of the simulated time series. Calibration of the precipitation model is done with the mAr package as explained in Section 2.5.4 and using 10-minute precipitation time series of stations Esch-sur-Sûre and Dahl for 2010. Upon calibration of the multivariate autoregressive model, we proceeded with the conditional simulation of Y_c (Supplementary Material, Section S3.2, Eq. (??5)). For this, we computed the parameters of the model as shown in Eq. (14). The model parameters are given by (Torres-Matallana et al., 2017):

$$\mu_1 = 2.85501 \qquad \mathbf{B} = \begin{bmatrix} 0.95650 & 0.03980 \\ 0.02429 & 0.88304 \end{bmatrix} \qquad \sigma_1^2 = 0.07241$$
$$\mu_\delta = 0.10194 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \sigma_\delta^2 = 0.07951 \qquad (14)$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \rho_{1\delta} = -0.03876$$

Next we generated conditional simulations of the 10-minute precipitation for 2010 for each subcatchment using the approach described in Section 2.5.4. Note that this involves simulating log-transformed precipitation which can easily be transformed to precipitation data using the antilog. The simulation procedure was repeated as many times as simulated precipitation time series were required for the MC uncertainty propagation analysis.

The simulated precipitation time series captured the main statistics of the observed time series well. The reader can find evidence for this in the Supplementary Material (Table S1 and Figure S2). Despite the satisfactory performance of the proposed method, some cases showed an overestimation of the simulated precipitation, mainly due to high values of the ratio of the multiplicative factor $\delta(t)$. This behaviour was also recognised by McMillan et al. (2011), who stated that the multiplicative factor used in their study *"does not capture the distribution tails, especially during heavy precipitation where input errors would have important consequences for runoff prediction"*.

3.3 Uncertainty ~~propagation~~analysis

Model output sensitivity and the degree of uncertainties evaluation of each model input helped to define the four input variables included in the uncertainty analysis: $C_{COD,S}$, $C_{NH4,S}$, COD_r and P. In this section we present the results of the uncertainty propagation for these four selected input variables to the model output, both for water quantity (volume in the combined sewer overflow tank, CSOT, and overflow volume and flow) and for water quality (loads and concentrations of chemical oxygen demand, COD, and ammonium, NH_4).

3.3.1 Monte Carlo ~~simulation size~~output and uncertainty quantification

~~In order to perform the MC propagation analysis, we first did a convergence test to estimate the number of simulations required. Besides this test, we also computed the standard error of all MC outputs. These two methods have the same aim and are closely related. In the convergence test, the standard deviation of two different MC simulations with different random seeds were~~

computed and compared for the The seven output variables of from EmiStatR , three representing water quantity variables ($V_{Chamber}$, V_{Sv} and Q_{Sv}) and four for water quality ($B_{COD,Sv}$, $B_{NH4,Sv}$, $C_{COD,Sv,av}$, and $C_{NH4,Sv,av}$). The results of the test indicated that in most cases between 250 and 1,000 MC simulations are enough to reach stable results in terms of the Nash–Sutcliffe model efficiency coefficient (NSE), where a NSE of 1 means a perfect match between observations and model output. In this case we got a NSE $\approx$ 0.998 for overflow volume. Regarding the water quality variable $B_{COD,Sv}$, the test showed that a larger number of MC simulations is required. Between 1,000 and 2,000 simulations are required to reach stable results (NSE $\approx$ 0.880 for overflow COD load and 0.998 for overflow NH_4 load). Therefore, a number of 1,500 MC simulations was used to perform the uncertainty analysis of the water quantity and water quality outputs. Figure **??** illustrates results of the convergence test for the cases where the number of MC replications is 250, 1,000 and 1,500. In this figure the MC1 output is plotted on the x-axis and MC2 output on the y-axis. Although the model output corresponds to yearly time series at 10 minutes resolution, we only plotted those points where the overflow magnitude, and therefore COD and NH_4 load, is different from zero. As an indication, for a MC replication size of 1,500, the NSE values for overflow COD and NH_4 concentrations are 0.816 and 0.998, respectively.

Results of the MC convergence test for (a, b, c) volume in overflow; (d, e, f) overflow COD load; (g, h, i) overflow NH_4 load. Each open circle refers to a ten minute time instant in 2010. As an indication, for a MC replication size of 1,500, the NSE values for overflow COD and NH_4 concentrations are 0.816 and 0.998, respectively. Dotted line is the 1:1 line. SD = Standard Deviation.

The computing times per MC replication are presented in Table **??**. The computations were performed with two different Linux machines, a laptop with four cores for simulations between 50 and 500 replications, and a server with 80 cores for performing the simulations above 500 replications. Similar execution times were reached for MC1 were analysed by MC input uncertainty propagation. A detailed description of the Monte Carlo simulation size and MC2 for one-month time series at 10-min time steps (August 2010, 4,464 time steps), while substantial differences were obtained when the 80 cores server was used. We obtained similar timing for 1,500 replications with 50 cores as for 250 replications using three cores in the laptop. The timing reached demonstrates the feasibility to perform a solid MC uncertainty propagation analysis with EmiStatR.

Average running time in minutes for Monte Carlo (MC) replications and specific cores used with two different seeds for the pseudo-random number generator in R. The rainfall input used was a one year length time series with 10 minutes time steps from 1 to 31 August 2010 (4,464 time steps). Replications 250 500 1000 1500 2000 cores 3 3 50 50 50 MC1 7.12 14.23 4.84 7.33 9.40 MC2 7.09 14.63 4.96 7.26 9.53 Average 7.10 14.43 4.90 7.29 9.46 timing is presented in the Supplementary Material (Section S5).

3.3.2 Monte Carlo output and uncertainty quantification

The seven output variables from EmiStatRwere analysed by MC input uncertainty propagation. Figure 3 illustrates the uncertainty propagation outcomes for the first Monte Carlo simulation, where all input variables vary stochastically. The MC simulations were performed for the entire year 2010 at 10 minutes time step, which were aggregated to hourly time steps in the figure. The aggregation function used for precipitation, CSO chamber volume, CSO spill volume and loads was the sum,

whereas for CSO spill flow and concentrations the aggregation function it was the mean. The figure (top) shows input precipitation as main driving input. For illustration purposes, two events of two-day duration each are shown. The first event occurred in spring (May 2010), the second in both events, and shows that the uncertainty is high when there is a high precipitation event. The more intense the precipitation input, as seen in the figure inset at top-left (May event), the greater the uncertainty band width for overflow flow, as well as for COD and NH_4 loads (insets three and four) and concentrations (insets five and bottom). The MC estimated statistics and the standard errors (se) are presented in Table 6. The table shows the uncertainty quantification of outputs obtained from the MC uncertainty propagation for the first MC simulation (all selected input variables are uncertain).

Table 6. Monte Carlo estimated statistics and standard errors (se) by bootstrapping for the MC simulations, where all selected input variables are uncertain R (model run at 10 minute time steps and MC results aggregated to one hour averages over one year period). See Table 1 for output variable nomenclature and units. Interq. = Interquartile; quant. = quantile; pbw = prediction band width; sd = standard deviation; var = variance; cv = coefficient of variation.

	Mean	Interq. Range	0.005 quant.	0.995 quant.	99% pbw	0.05 quant.	0.95 quant.	90% pbw	sd	var	cv
	μ_{MC}	IQR	$\zeta_{0.005}$	$\zeta_{0.995}$	$\zeta_{w,0.99}$	$\zeta_{0.05}$	$\zeta_{0.95}$	$\zeta_{w,0.90}$	σ_{MC}	σ^2_{MC}	CV_{MC}
$V_{Chamber}$	92.51	13.65	77.55	109.16	31.60	81.02	104.28	23.25	8.27	2,984	0.100
se	2.53	1.17	2.18	3.17	1.48	2.27	2.95	1.36	0.56	482	0.001
V_{Sv}	3.18	3.69	0.85	6.60	5.76	0.94	5.76	4.82	1.98	1,100	0.070
se	0.51	0.73	0.24	0.95	0.95	0.26	0.91	0.83	0.35	259	0.012
Q_{Sv}	55.49	76.77	0.37	267.1	266.7	1.24	165.3	164.1	64.50	7,332	1.585
se	5.67	10.26	0.13	13.75	14.54	0.22	10.82	10.89	4.57	1,102	0.048
$B_{COD,Sv}$	1.18	1.69	0.04	6.11	6.06	0.07	3.49	3.41	1.27	394	0.087
se	0.20	0.31	0.02	1.02	1.00	0.03	0.59	0.59	0.21	81.71	0.013
$B_{NH4,Sv}$	0.052	0.077	0.004	0.174	0.170	0.006	0.125	0.120	0.045	0.546	0.075
se	0.009	0.014	0.002	0.029	0.028	0.002	0.022	0.021	0.008	0.115	0.013
$C_{COD,Sv,av}$	170.0	164.6	3.80	909.7	905.9	15.49	465.8	450.3	161.9	36,151	0.988
se	9.02	11.38	0.33	40.12	40.40	1.02	24.66	24.03	7.94	4,615	0.016
$C_{NH4,Sv,av}$	7.19	6.65	0.47	29.20	28.74	0.86	17.51	16.64	5.66	46.93	0.815
se	0.41	0.61	0.02	1.23	1.22	0.06	0.91	0.87	0.29	6.62	0.016

Table 6 shows the standard deviation (sd) and the coefficient of variation (cv) for the seven output variables considered in the uncertainty propagation. For the volume in the CSO chamber, $V_{Chamber}$, the annual mean standard deviation, σ_{MC}, (8.27 m^3) is lower than the mean, μ_{MC}, (92.51 m^3). This goes along with an annual mean coefficient of variation (CV_{MC}) of 0.100. A (CV_{MC}) greater than 1 means large uncertainty. The overflow spill volume, V_{Sv}, had a coefficient of variation of 0.070, while it was 1.585 for the overflow flow, Q_{Sv}. This shows that the relative uncertainty of the overflow flow is very large. Regarding the overflow COD load, the annual mean (1.18 kg) is similar as the annual mean standard deviation (1.27 kg). Similar behaviour was observed for the overflow COD concentration, which had an annual mean value of 170 mg/l and a standard deviation of 162 mg/l. For overflow NH_4 load and overflow NH_4 concentration the annual mean also had the same order of magnitude as the annual mean standard deviation. Overflow COD and NH_4 loads had a coefficient of variation of

0.087 and 0.075, respectively, whereas the coefficient of variation for concentrations were 0.988 and 0.815, respectively. This suggests that overflow concentrations are more uncertain.

Low standard errors (se) for the coefficient of variation were obtained for all seven outputs. These were never greater than 0.05, which indicates that the selected MC replication size (1,500 for $mc1$) is a suitable value. This holds for all output statistics, because in all cases the standard error is small to the estimated value.

3.3.2 Contributions of input variables to total uncertainty

The contributions of input variables to the total uncertainty of the model inputs were also computed using the procedure described in Section 2.6.4. A total of four MC simulations with a total of 6,000 runs were performed for estimating S_i (Eq. (10)). Afterwards, four contributions were evaluated per time step and aggregated for the whole year. Following Eq. (10), the per time step contributions of input variables to output variables in terms of percentage of variance, stochastic sensitivity S_i of the input variables $C_{COD,S}$, $C_{NH4,S}$, COD_r and P were calculated. An example of the contributions analysis per time step is presented in Fig. 4. Here we remark that a high uncertainty over time is shown mainly for the Spring event.

The aggregated over time contributions of input variables to output variables in terms of percentage of variance, stochastic sensitivity S_i of the input variables, were also calculated (Table 7). Note that P is the only source of uncertainty for $V_{Chamber}$ and V_{Sv}, while uncertainty in NH_4 inputs only propagates to NH_4 outputs, and similar for COD (Fig. 4).

Table 7. Aggregated over time contribution of input variables to output variables in terms of percentage of total variance

	Stochastic sensitivity, S_i, of input variable [%]				
Output variable	Total	$C_{COD,S}$	$C_{NH4,s}$	COD_r	P
$V_{Chamber}$	100.0	0.0	0.0	0.0	100.0
V_{Sv}	100.0	0.0	0.0	0.0	100.0
$B_{COD,Sv}$	100.0	65.7	0.0	2.9	31.4
$C_{COD,Sv,av}$	100.0	62.4	0.0	8.7	28.9
$B_{NH4,Sv}$	100.0	0.0	34.4	0.0	65.6
$C_{NH4,Sv,av}$	100.0	0.0	35.3	0.0	64.7

We found, as expected, that precipitation, P, is the only source of uncertainty from all uncertain input considered for water quantity output variables $V_{Chamber}$ and V_{Sv}. Regarding average values for the whole year, for the water quality output variables $B_{COD,Sv}$ and $C_{COD,Sv,av}$, $C_{COD,s}$ has the largest contribution to the output variance, about 66 percent for $B_{COD,Sv}$ and about 62 percent for $C_{COD,Sv,av}$. The second variable that contributes to uncertainty of these COD output variables is P, with about 3 percent for $B_{COD,Sv}$ and 9 percent for $C_{COD,Sv,av}$. Similarly, the input variable $C_{NH4,S}$ play an important role in the contribution of total uncertainty for $B_{NH4,Sv}$ (on average about 34 percent of the variance for the whole year) and $C_{NH4,Sv,av}$ (about 35 percent). Equally contributing to uncertainty of these NH_4 output variables is P with about 66 percent for $B_{NH4,Sv}$

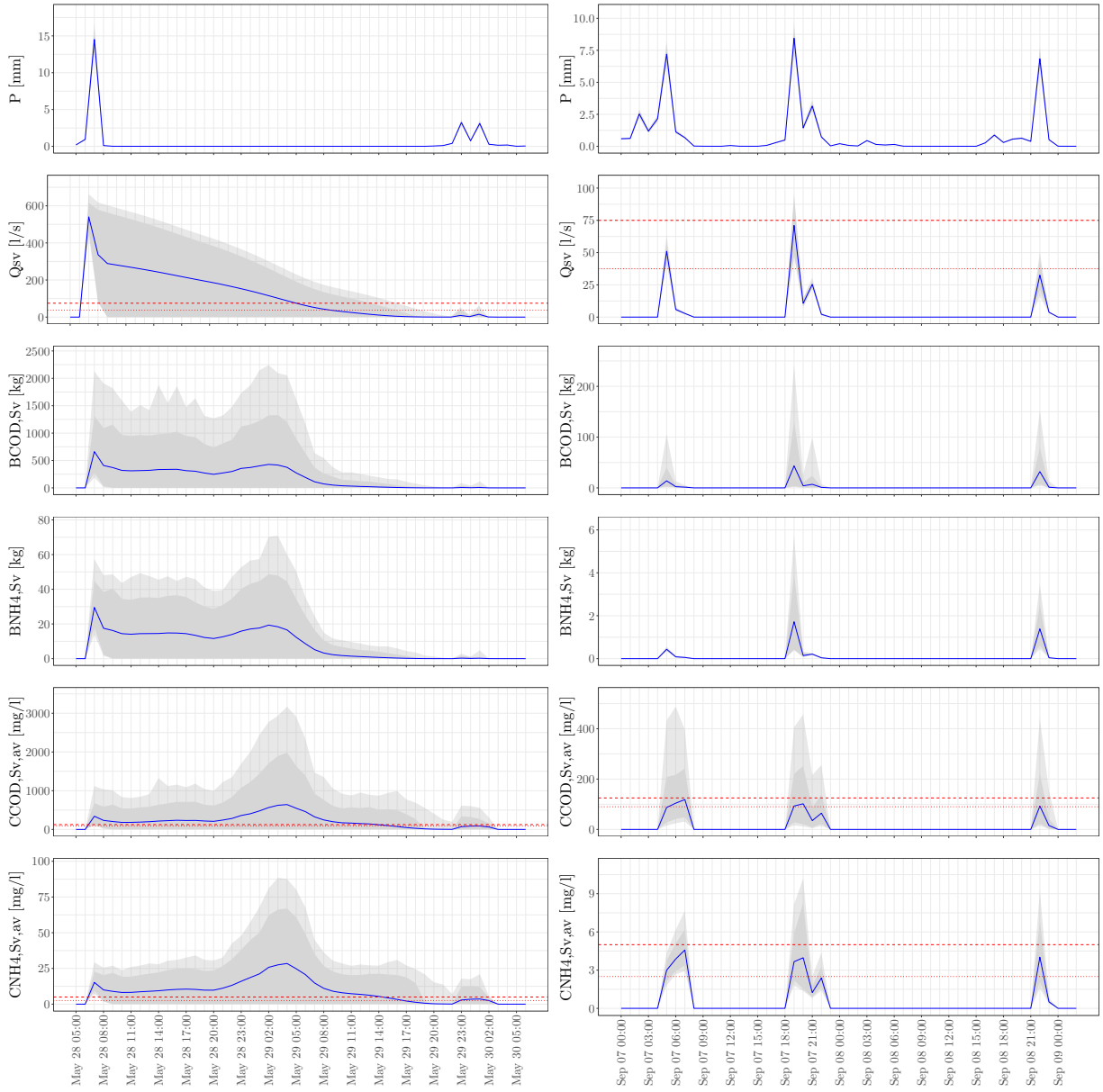

Figure 3. Uncertainty propagation outcomes for the first Monte Carlo simulation, where all input variables vary stochastically. The 99% prediction interval is shown as light grey shade, 90% prediction interval is shown as dark grey shade, mean value as blue line. The MC simulations were performed for the entire year 2010 at 10 minutes time step, aggregated to hourly time steps in the figure. Input precipitation (**top**). Overflow spill flow, the upper dashed red line indicates the 75 l/s threshold, lower dotted red line the 37.5 l/s threshold (**second**). Load of overflow COD (**third**). Load of overflow NH_4 (**fourth**). Average spill COD concentration. Upper dashed red line indicates the 125 mg/l threshold and the lower dotted red line indicates the 90 mg/l threshold (**fifth**). Average spill NH_4 concentration. Upper dashed red line indicates the 5.0 mg/l threshold, lower dotted red line the 2.5 mg/l threshold (**bottom**).

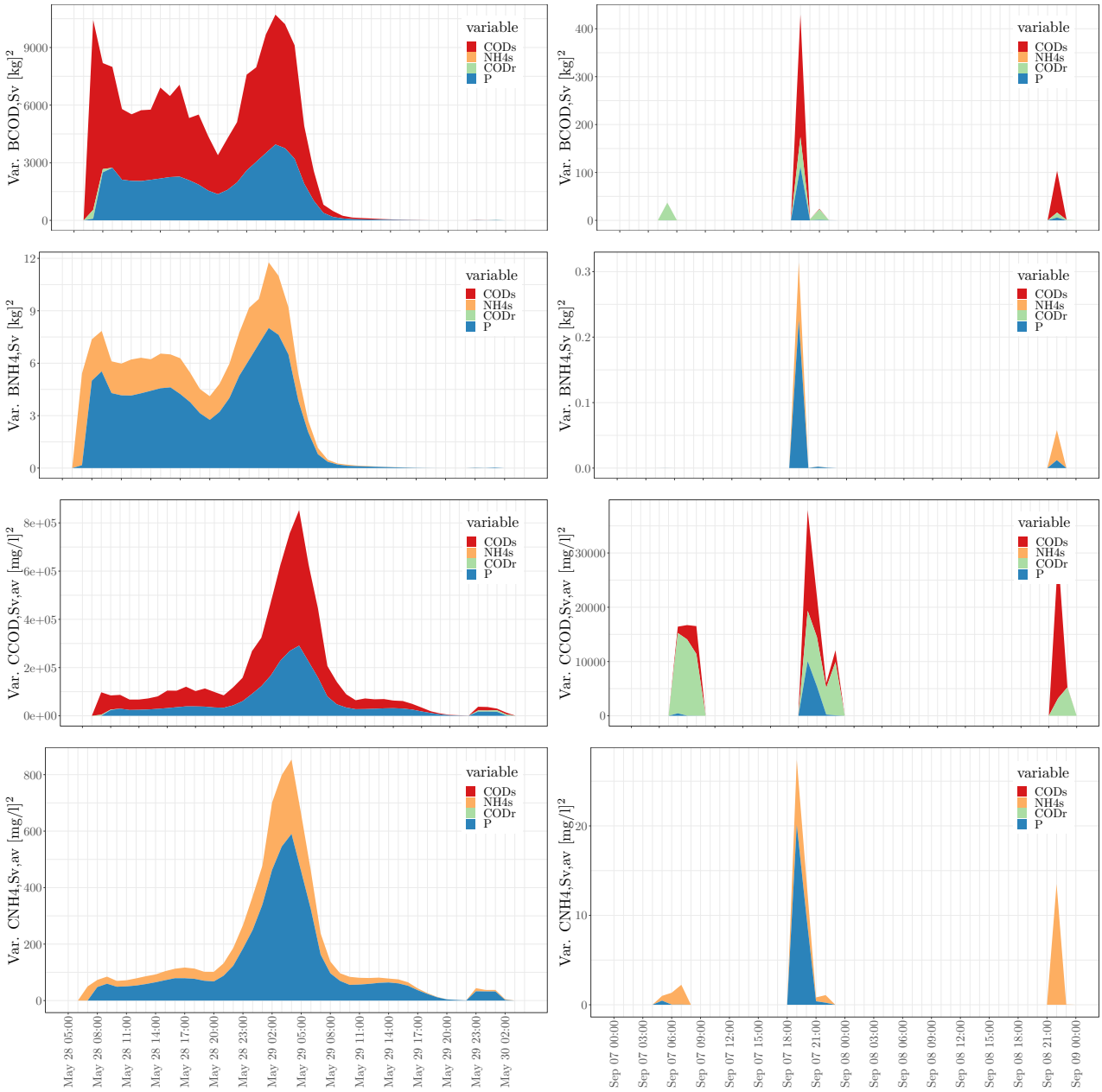

Figure 4. Temporal contributions of input variables to load of overflow COD (**top**) load of overflow NH_4 (**second**); concentration of overflow COD (**third**); concentration of overflow NH_4 (**bottom**) in terms of variance. The MC simulations were performed for the entire year 2010 at 10 minutes time step, which were aggregated to hourly time steps. For illustration two periods are shown from 28 to 30 May 2010 (**left**); and from 7 to 9 September 2010 (**right**).

and 65 percent for $C_{NH4,Sv,av}$. From these results we can infer that precipitation is a main source of uncertainty for all six outputs considered.

3.4 Uncertainty and water quality impact

Quantification and assessment of the water quality impact is an important step after the uncertainty propagation. As described in Section 2.7, the assessment of water quality standards was done taking into account the reference thresholds recommended in the European Union guidelines for COD, and the German and Austrian guidelines for hydraulic impact and acute ammonium toxicity.

3.4.1 Hydraulic impact

From the time series of daily values for 2006 to 2013 of the river Sûre, a daily flow expected with return period once per year (1.01 years), Q_{r1} of 16 $m^3 \cdot s^{-1}$ was computed at Heiderscheidergrund, which corresponds with the entire catchment area of the Haute-Sûre stormwater system (182.1 ha). Therefore, we estimated the river daily flow in the Goesdorf CSOT as a proportion to 30 ha, which is equal to 2.6 $m^3 \cdot s^{-1}$. Following Eq. (11), the maximum sewer overflow discharge with return period one year Q_1 can have a value between 0.26 $m^3 \cdot s^{-1}$ and 1.32 $m^3 \cdot s^{-1}$. Accordingly, with the German guideline ATV-A 128 (1992) (Eq. (12)), two additional thresholds are defined for the maximum sewer overflow discharge with return period one year for the Goesdorf catchment ($A_{imp} = 5.0$ ha). Q_1 is expected to vary between 37.5 $l \cdot s^{-1}$ and 75.0 $l \cdot s^{-1}$. We contrasted these values with those obtained from the uncertainty analysis. From Table 6, we obtained a one hour mean value for the overflow spill flow, Q_{Sv}, of 55.5 $l \cdot s^{-1}$, 90% prediction band width of 164.1 $l \cdot s^{-1}$, and standard deviation of 64.5 $l \cdot s^{-1}$. Figure 3 (second) presents the overflow spill flow for the two periods chosen for illustration. The upper dashed red line indicates the 75 $l \cdot s^{-1}$ threshold and the lower dotted red line indicates the 37.5 $l \cdot s^{-1}$ threshold. Table 8 (top) shows the exceedance percentage of overflow spill flow over the 37.5 and 75.0 $l \cdot s^{-1}$ thresholds for the mean, 0.95 quantile and 0.995 quantile. We found a 0.49% exceedance of the mean value over the 37.5 $l \cdot s^{-1}$ threshold and about 1.7% for the quantiles. As expected, slightly lower percentages were found for the 75.0 $l \cdot s^{-1}$ threshold.

3.4.2 COD concentration

A reference COD concentration emission in CSOs was presented in Section 2.7.2. For the European Union, a value of 125 $mg \cdot l^{-1}$ is used. We obtained a one hour average spill COD concentration with a mean of 170 $mg \cdot l^{-1}$, standard deviation of 162 $mg \cdot l^{-1}$, and a 90% prediction band width of 450 $mg \cdot l^{-1}$. Figure 3 (fifth) presents the average spill COD concentration. Upper dashed red line indicates the 125 mg/l threshold, lower dotted red line the 90 mg/l threshold. The mean COD concentration in the overflow volume was higher than the thresholds. However, note that when entering the river system it will quickly be diluted, suggesting that the negative impact on the environment will be dampened by the receiving water body.

Table 8 (centre) shows the exceedance percentage of overflow COD concentration over the 90 and 125 $mg \cdot l^{-1}$ thresholds for the mean, 0.95 quantile and 0.995 quantile. We found a 1.62% exceedance of the mean value over the 90 $mg \cdot l^{-1}$ threshold and

about 1.8% for the quantiles. Slightly lower percentages were found for the 125 mg·l^{-1} threshold for the mean value (1.03%). For the quantiles equal values were found as for the 90 mg·l^{-1} threshold.

3.4.3 Acute ammonium toxicity

We compared the acute ammonium toxicity reference values presented in Section 2.7.3 (2.5 mg $\cdot$ l^{-1} for the ammonium concentration calculated for one hour duration for salmonid streams, and for cyprinid streams a maximum value of 5.0 mg $\cdot$ l^{-1}), with the values we found for ammonium. An average spill NH_4 concentration with a mean of 7.19 mg $\cdot$ l^{-1}, standard deviation of 5.66 mg $\cdot$ l^{-1}, and 90% prediction band width of 16.64 mg $\cdot$ l^{-1} was obtained. Figure 3 (bottom) shows the average spill NH_4 concentration for the two periods chosen for illustration. The ammonium (NH_4) concentrations in the overflow flow are higher than the reference values, which are given for concentrations in the river.

Table 8 (bottom) shows the exceedance percentage of overflow NH_4 concentration over the 2.5 and 5.0 mg·l^{-1} thresholds for the mean, 0.95 quantile and 0.995 quantile. We found a 1.8% exceedance of the mean and quantile values over the 2.5 and 5.0 mg·l^{-1} thresholds. A slightly lower percentage (1.1%) was found for the 5.0 mg·l^{-1} threshold, regarding mean value.

Table 8. Frequency (percentage) over time that environmental thresholds are exceeded for different statistics of the overflow spill flow, COD and NH_4 concentration.

Output variable	Threshold	Statistic	Exceedance percentage
Q_{Sv} [l·s^{-1}]	37.5	Mean	0.49
	37.5	0.95 quantile	1.71
	37.5	0.995 quantile	1.74
	75.0	Mean	0.31
	75.0	0.95 quantile	1.51
	75.0	0.995 quantile	1.72
$C_{COD,Sv,av}$ [mg $\cdot$ l^{-1}]	90.0	Mean	1.62
	90.0	0.95 quantile	1.80
	90.0	0.995 quantile	1.82
	125.0	Mean	1.03
	125.0	0.95 quantile	1.80
	125.0	0.995 quantile	1.82
$C_{NH4,Sv,av}$ [mg $\cdot$ l^{-1}]	2.5	Mean	1.78
	2.5	0.95 quantile	1.80
	2.5	0.995 quantile	1.82
	5.0	Mean	1.05
	5.0	0.95 quantile	1.78
	5.0	0.995 quantile	1.82

4 Discussion

This study aimed to select and characterise the main sources of input uncertainty in urban water systems, while accounting for temporal auto- and cross-correlation of uncertain model inputs, by propagating input uncertainty through the EmiStatR model, and quantifying and assessing the contributions of each uncertainty source to model output uncertainty dynamically (over time). In the following discussion, we start with the ~~accuracy of Monte Carlo analysis. Then, we discuss the~~ uncertainty and water quality impact of the model outputs to the environment, in relation to the uncertainty analysis~~, and finally~~. Next, we discuss the accuracy of Monte Carlo analysis, followed by a discussion of other sources of uncertainty. Finally, we highlight some limitations and possible solutions of the approach used in this work.

4.1 Uncertainty and water quality impact

Next we discuss how the uncertainty propagation analysis done gives additional insight regarding hydraulics, COD concentration and acute ammonium toxicity impact on water quality over the river Sûre due to the CSO discharges under study. After doing the uncertainty propagation analysis we not only have predictions of model outputs but we also know how uncertain these are. An added value arises when we take into account the uncertainty information. For the case of the overflow spill flow, the expected model output (mean of 55.5 $l \cdot s^{-1}$) is below the environmental threshold of 75 $l \cdot s^{-1}$, but the 0.95 quantile (164.1 $l \cdot s^{-1}$) is much above the threshold. This indicates that there is a considerable chance of being above the threshold.

Regarding water quality outputs, although the mean model output for COD and NH_4 concentrations is fairly above of the thresholds, the 0.95 quantile is 2.7 times above the mean value for COD concentration, and 2.4 times above the mean value for NH_4. Also here we can conclude that we are not certain that we are below the threshold, because there is a considerable probability that the true values are above, even though the expected value is below the thresholds.

We were able to compute the water quantity and quality at CSO outlet to the river. We found that water quality (COD and NH_4) were sometimes above the environmental threshold. Even if the expected value was below the threshold there could still be a considerable probability that the quality was above the threshold because of the large uncertainty. Therefore, policy and decision makers and water managers need to be aware of this, because whenever concentrations are above the threshold this may harm the environment. Nevertheless it is worth noting that we computed concentration in the outlet of the CSO. When this spilled water enters the river it will quickly mix with the much cleaner river water and concentrations will drop quickly, so it is only a local problem. How local it is and how the river water quality is distributed in space and time is not an easy problem to solve and requires the use of hydrological and hydraulic river models e.g. SIMBA (IFAK, 2007) or MIKE 11 (DHI, 2017). Those models have been well-developed and for some of them uncertainty analyses have also been done (Beven and Binley, 1992; Refsgaard, 1997; Beven and Freer, 2001; Vrugt et al., 2003a, 2008; Beven et al., 2010; Andrés-Doménech et al., 2010; Beven, 2012; Jerves-Cobo et al., 2020; Yu et al., 2020), but obviously such uncertainty analyses can only be done if the inputs to these models are known as well as the uncertainty associated with these inputs. One of these inputs is inlet from CSO. That is where our paper makes a very valuable contribution, because our work has quantified water quantity and quality of CSO structures, including uncertainty, and that is exactly what these river models need to be able to do

an uncertainty propagation analysis. We also recognise other attempts on quantity (e.g. Sriwastava et al. (2018)) and quality, especially measurements taken at CSOs, which demonstrate that the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs (e.g. Brombach et al. (2005); Diaz-Fierros T et al. (2002)). We present some comparisons with these studies in the following lines.

Sriwastava et al. (2018) apply uncertainty propagation to a complex hydrodynamic model for quantifying uncertainty in sewer overflow volume. They used MC for uncertainty propagation and Latin hypercube sampling (LHS) as an efficient sampling scheme. Although LHS ensures a full coverage of the sample space and provides a faster convergence than simple random sampling, the LHS application in the case of dynamic model inputs (e.g. precipitation, COD and NH_4 inputs) is not trivial and its implementation is more complex than in the case of sampling from static variables (i.e., uncertain constants). In our study, we sampled time series of dynamic inputs using an implementation in stUPscales (Torres-Matallana et al., 2019, 2018a) .

Diaz-Fierros T et al. (2002), in a study in the city of Santiago de Compostela (North-West Spain, population about 100,000 inhabitants), where a combined sewer system feeds to a grossly under-sized wastewater treatment plant, reported an event mean concentration (Diaz-Fierros T et al. (2002), Table 4) for the output variables $C_{COD,Sv,av}$ and $C_{NH4,Sv,av}$ of 329.1 mg·l^{-1} and 8.7 mg·l^{-1}, respectively. These values are larger than those found by Brombach et al. (2005), and more in agreement with our findings, especially for the case of $C_{NH4,Sv,av}$. Diaz-Fierros T et al. (2002) reported values of $C_{COD,Sv,av}$ as high as 1073 mg·l^{-1}, which agrees with the right-hand tail of the distribution obtained in our study (i.e. a 0.995 quantile of 909.7 mg·l^{-1}). Similarly, for the case of $C_{NH4,Sv,av}$, Diaz-Fierros T et al. (2002) reported values as high as 32.5 mg·l^{-1}, comparable with the 0.995 quantile (29.20 mg·l^{-1}) found in our study.

It is worth noting that regarding measurements taken at CSOs, the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs as reported by Diaz-Fierros T et al. (2002) and Brombach et al. (2005) . Thus, when comparing model outputs with independent measurements, one should bear in mind that discrepancies between measured and predicted are not only caused by errors in model inputs, model parameters and model structure, but are also the result of errors in the water quality measurements.

4.2 Accuracy of Monte Carlo analysis

Regarding the Monte Carlo replication size for uncertainty propagation, we presented in Fig. ~~??~~ S4 (Supplementary Material) the results for three output variables and three replications size 250, 1,000 and the selected 1,500 (NSE closer to 1.0 for most of the output variables). We compute replications for 50, 100, and from 250 to 2,000 at steps of 250 replications the comparison of two equal MC runs (MC1 and MC2) with different seed for the pseudo-random number generator. the results suggest that the output variables related to COD (load and concentration) have a larger dispersion when we compare MC1 and MC2 for the same replications size. This is also reflected in the larger standard errors reported in Table 6 for e.g. the overflow COD load. Nevertheless, 1,500 runs are a feasible MC replication size for running a relative simple and fast model as EmiStatR (7.29 minutes in average execution time using parallel computing and 50 cores for a time series with 4,464 time steps). For a more complex full hydrodynamic model with a high computational burden, 1,500 replication four times to compute contributions

it may be not possible. Therefore, we suggest to check the intermediate results of the MC convergence test and we will find that e.g. for quantity variables as the spill overflow volume and quality variables as the overflow NH_4 load, 250 replications (7.10 minutes in average execution time using parallel computing and three cores for a time series with 4,464 time steps) per individual MC execution seem to be enough, which make more feasible the execution of this kind of uncertainty propagation.

Figures 3 and 4 shows that there is a large uncertainty for the early May event and smaller uncertainty for the September event. This is due mainly to the presence of a large dry period before the spill event in May, i.e. a shorter dry period preceding the spill flow leads to a lower uncertainty. This finding suggest also an interaction between the antecedent dry period and the concentration of pollutants.

4.3 Other sources of uncertainty

In this work we only looked at input uncertainty and not at parameter and model structural uncertainty. Further research can be done on those topics. Neumann (2007) address how are uncertainty ranges for parameters of full scale systems obtained and how does model structure uncertainty manifest itself and can be quantified for performance evaluation and design of urban water infrastructure. Moreno-Rodenas et al. (2019) also studied and depicted how model parameter is an important source of uncertainty. They emphasised that "*still, uncertainty analysis is seldom applied in practice and the relative contribution of the individual model elements is poorly understood.*". Also, they highlighted that after inferring the river process parameters with system measurements of flow and dissolved oxygen, combined sewer overflow pollution loads became the dominant uncertainty source along with rainfall variability. These findings agreed with our results.

Bachmann-Machnik et al. (2018) recognised that the most important parameters causing uncertainties in the sewer system model are connected area and the stormwater runoff quality. Our analysis confirms these findings, specifically regarding the stormwater runoff quality. In our study the input variable runoff COD was an important source of uncertainty with relation to the annual mean overflow COD released from the CSO.

4.4 Limitations and possible improvements

Despite the extensive temporal uncertainty propagation analysis the approach also has some limitations which we present hereafter addressing possible solutions in future work.

1. **Incorporation of the spatial distribution of model inputs.** Specifically for precipitation, Breinholt et al. (2012) stated that due to a poor representation of the spatial precipitation that is measured by point gauges and the complexity of the sewer systems, large output uncertainty can be expected. We also infer that we obtained a large output uncertainty due to neglect of the inherent spatial variability of precipitation. Therefore, we suggest that further research is needed to account for spatial variability of precipitation, that can bring light to understand how this variability impacts in the output uncertainty and quantify it properly. This issue should be related also to the problem of change of support. When modelling precipitation, we also ignored the support effect, i.e. we ignored that the sub-catchment area is much greater

than a point. Future research may address this issue of change-of-support. Studies that tackled this issue are found e.g. Leopold et al. (2006); Wadoux et al. (2017); Cecinati et al. (2018).

2. **Linkage of sub-models and uncertainty compensation effect.** Tscheikner-Gratl et al. (2019) addressed the question as to whether there is an increase in uncertainty by linking integrated models or ~~if~~ whether a compensation effect could take place ~~and that~~ by which overall uncertainty in key water quality parameters ~~actually decreases. We contribute in this discussion by advising to quantify~~ decreases. Some further insight into this topic could be obtained by quantifying uncertainties at sub-model level, ~~because as we demonstrated the computational budget can be reduced and make it feasible when dealing at the sub-module uncertainty propagation~~and analysing whether uncertainty at sub-model level is greater or smaller than at the overall level. With our implementation this is not a difficult task because EmistatR has a stringent modular design in which it is easy to analyse outputs and their uncertainties at sub-model level.

3. **Accounting for cross-correlation between the inputs precipitation and runoff COD concentration.** It is worth noting that we did not include correlation between COD_r and P. Including such correlation would yield a more realistic model of the uncertainty because these variables are known to have a strong correlation. It is highly recommendable to include correlations between COD_r and precipitation, because loads in chemical oxygen demand are correlated with the overland flow due to precipitation, which may transport distributed pollutants to the sewer system. Also the inputs $C_{COD,s}$ and $C_{NH4,s}$ can be related with a daily curve that reflect the pattern of consumption in the household like the German ATV-A 134 curve. We used the latest version of EmiStatR (version 1.2.2.0), which considers this kind of patterns.

4. **Absence of high frequency water quality observations to compare with model outputs and uncertainty prediction bands.** In order to gain understanding of the temporal dynamics of nutrients (nitrogen, N, and phosphorus, P), Yu et al. (2020) applied high frequency monitoring in a groundwater fed low-lying urban polder in Amsterdam (The Netherlands). They argued that although spatial and temporal concentration patterns from discrete sampling campaigns of water quality parameters, such as EColi, showed a clear dilution pattern, the temporal patterns of N and P were still poorly understood, given their reactive nature and more complex biogeochemistry. Therefore, high frequency measurement, is a key factor to understand these temporal dynamics and patterns.

5. **Absence of a joint spatio-temporal uncertainty analysis.** According to Zhou et al. (2020), the limitations in algorithms for classic uncertainty estimates is the cause that only the uncertainty in one dimension (either temporal variability or spatial heterogeneity) is considered, whereas the variation in the other dimension is dismissed, resulting in an incomplete assessment of the uncertainties. Zhou et al. (2020), also showed that classic metrics underestimate the uncertainty through averaging, which means a loss of information in the variation across spatio-temporal scales. To handle this limitation, suitable methods are the three-dimensional variance partitioning for a new uncertainty estimation in both spatio-temporal scales (Zhou et al., 2020), or spatio-temporal geostatistics (Gräler et al., 2016).

6. **Uncertainty analysis with complex models.** In this research we were able to conduct a comprehensive Monte Carlo uncertainty propagation analysis, which required a large number of Monte Carlo runs. This was possible because we used a strongly simplified urban water system model, EmistatR. For more complex models that take much more computing time, application of a Monte Carlo uncertainty propagation analysis is more challenging. However, given sufficient resources it is possible, because each model run can be run independently and hence the analysis is extremely suitable for parallisation and cloud computing. In particular, the use of graphics processing units (GPU) for heavy computation is promising. Some recent examples that demonstrate the potential of GPU for this purpose are Eränen et al. (2014), Sten et al. (2016), and Sandric et al. (2019). Sriwastava et al. (2018) applied uncertainty propagation to a complex hydrodynamic model, by selecting a small subset of dominant input/model parameters that explain most of the model output variance.

5 Conclusions

In this final section we conclude with highlighting the importance of temporal uncertainty propagation analysis and the selection and characterisation of uncertain model inputs impacting model sensitivity. We also point out that uncertainty propagation analysis helps to identify the most contributing sources and can provide better evidence for the impact assessment of pollutant release from sewer systems to the environment, in particular to the receiving waters.

1. **Uncertainty analysis is important because it quantifies the accuracy of model outputs and quantifies the uncertainty source contributions.** The latter provides essential information to take informed decisions about how to improve the accuracy of the model output. But MC uncertainty analysis is only possible if it is computationally feasible. We used a simplified urban water system model with capabilities to apply for *minimising transient pollution from urban wastewater systems* in parallel mode, which minimises model running time, allowing uncertainty propagation, long term simulations and evaluation of complex scenarios. These capabilities are crucial also for e.g. real time control applications, where simplified models of fast running times are desirable.

2. **Input variables that were very uncertain for which model output was very sensitive were selected to be included in the uncertainty propagation analysis.** We found four main input variables to be analysed: 1) Precipitation, P; 2) Chemical oxygen demand sewage pollution per capita load per day, $C_{COD,S}$; 3) Ammonium pollution per capita load per day, $C_{NH4,S}$; and 4) Chemical oxygen demand COD_r concentration.

3. **Selected input variables for uncertainty propagation can be characterised in terms of input uncertainty in four specific cases, depending on the type of input variable:** i) Uncertain constant inputs, characterised by their marginal (cumulative) pdf e.g. water consumption, infiltration flow, impervious area and run-off coefficients; ii) Temporally autocorrelated dynamic uncertain inputs, characterised by univariate time series autoregressive modelling e.g. COD_r; iii) Temporally cross-correlated multiple dynamic uncertain inputs, characterised by multivariate time series modelling, considering cross- and no-correlations among variables e.g. $C_{COD,S}$ and $C_{NH4,S}$; and iv) rain gauge input precipitation, characterised by autoregressive model conditioned to the observed precipitation (P).

4. **Model input uncertainty propagation through the simplified combined sewer overflow model (EmiStatR) helped to understand how does uncertainty propagate and how large is the uncertainty of EmiStatR outputs in a case study.** Three output variables were considered for water quantity and four variables for water quality. The Monte Carlo uncertainty propagation analysis showed that among the water quantity output variables, the overflow flow, Q_{Sv}, is the more uncertain output variable and has a large coefficient of variation (cv of 1.585). Among water quality variables, the annual average spill COD concentration, $C_{COD,Sv,av}$, and the average spill NH_4 concentration, $C_{NH4,Sv,av}$, were found to have large uncertainty (coefficients of variation of 0.988 and 0.815, respectively). Also, low standard errors (se) for the coefficient of variation were obtained for all seven outputs. They were never greater than 0.05, which indicated that the selected MC replication size (1,500 simulations) was a suitable value.

5. Regarding the **main sources of uncertainty model outputs**, for water quantity outputs, was precipitation, while for COD water quality outputs were P, $C_{COD,S}$ and COD_r, and for NH_4 outputs P and $C_{NH4,S}$.

6. Finally, we evaluated how **uncertainty propagation analysis can explain more comprehensively the impact of water quality indicators to the receiving river** for the Luxembourg case study. Although the mean model water quality outputs for COD and NH_4 concentrations is fairly above of the thresholds, the 0,95 quantile is 2.7 times above the mean value for COD concentration, and 2.4 times above the mean value for NH_4. We conclude that we are not certain that environmental thresholds are not exceeded, because there is a considerable probability that values are above, even though the expected value is below the thresholds. This is valid for concentrations in the spilled CSO, therefore, is important to highlight that the results confirmed our hypothesis that annual mean COD and NH_4 river concentrations are lower than the released CSO concentrations due to dilution and henceforth compliant with the water quality thresholds given by the guidelines consulted.

Code and data availability. The code scripts and datasets related to Figures 03 to 06 of this paper are available on Zenodo:
https://doi.org/10.5281/zenodo.3928079
and GitHub:
https://github.com/ArturoTorres/temporal_uncertainty_paper_reproducible.git

Author contributions. J. A. Torres-Matallana is the main author of the text in this article. Ulrich Leopold and Gerard Heuvelink contributed to this article with their statistical, geostatistical and programming knowledge, and reviewed and edited the text. J. A. Torres-Matallana developed all R-code scripts for the computations and Monte Carlo simulations, performed the simulations and analysis, with collaboration from Ulrich Leopold and Gerard Heuvelink.

Competing interests. The authors declare no competing interests.

Acknowledgements. The work presented was part of the QUICS (Quantifying Uncertainty in Integrated Catchment Studies) project. This project has received funding from the EU Marie Skłodowska-Curie research programme under the European Union's Seventh Framework Programme for research, technological development and demonstration with the grant agreement No. 607000, as well as the Luxembourg Institute of Science and Technology, LIST. We thank Dr. Kai Klepiszewski for his advice and involvement during early stages of this research and two anonymous reviewers for constructive and valuable comments that helped us to improve this paper.

References

Andrés-Doménech, I., Múnera, J. C., Francés, F., and Marco, J. B.: Coupling urban event-based and catchment continuous modelling for combined sewer overflow river impact assessment, Hydrology and Earth System Sciences, 14, 2057–2072, https://doi.org/10.5194/hess-14-2057-2010, 2010.

Bach, P. M., Rauch, W., Mikkelsen, P. S., McCarthy, D. T., and Deletic, A.: A critical review of integrated urban water modelling - Urban drainage and beyond, Environmental Modelling & Software, 54, 88 – 107, 2014.

Bachmann-Machnik, A., Meyer, D., Waldhoff, A., Fuchs, S., and Dittmer, U.: Integrating retention soil filters into urban hydrologic models – Relevant processes and important parameters, Journal of Hydrology, 559, 442 – 453, https://doi.org//10.1016/j.jhydrol.2018.02.046, 2018.

Baker, L. A., ed.: The Water Environment of Cities, Springer, 2009.

Barbosa, S. M.: Package "mAr": Multivariate AutoRegressive analysis, The Comprehensive R Archive Network, CRAN, 1.1-2 edn., 2015.

Bastin, L., Cornford, D., Jones, R., Heuvelink, G. B. M., Pebesma, E., Stasch, C., Nativi, S., Mazzetti, P., and Williams, M.: Managing uncertainty in integrated environmental modelling: The UncertWeb framework, Environmental Modelling & Software, 39, 116–134, 2013.

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction., Hydrological Processes, pp. 279–98, 1992.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11–29, https://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

Beven, K., Leedal, D., and Alcock, R.: Uncertainty and Good Practice in Hydrological Prediction, VATTEN, 66, 159–163, 2010.

Beven, K. J.: Rainfall-Runoff Modelling: The Primer, Wiley-Blackwell, second edn., Lancaster University, UK, 2012.

Blumensaat, F., Staufer, P., Heusch, S., Reussner, F., Schütze, M., Seiffert, S., Gruber, G., Zawilski, M., and Rieckermann, J.: Water quality-based assessment of urban drainage impacts in Europe where do we stand today?, Water Science & Technology, 66, 304, https://doi.org/10.2166/wst.2012.178, http://dx.doi.org/10.2166/wst.2012.178, iWA Publishing, 2012.

Boos, D., Matthew, K., and Osborne, J.: Monte.Carlo.se: Monte Carlo Standard Errors, https://CRAN.R-project.org/package=Monte.Carlo.se, r package version 0.1.0, 2019.

Boos, D. D.: Introduction to the Bootstrap and World, Statistical Science, 18, 168–174, institute of Mathematical Statistics, 2003.

Boos, D. D. and Osborne, J. A.: Assessing Variability of Complex Descriptive Statistics in Monte Carlo Studies Using Resampling Methods, International Statistical Review, 83, 228–238, https://doi.org/10.1111/insr.12087, 2015.

Breinholt, A., Moller, J. K., Madsen, H., and Mikkelsen, P. S.: "A formal statistical approach to representing uncertainty in rainfall–runoff modelling with focus on residual analysis and probabilistic output evaluation – Distinguishing simulation and prediction", Journal of Hydrology, pp. 36–52, 2012.

Brombach, H., Weiss, G., and Fuchs, S.: A new database on urban runoff pollution: comparison of separate and combined sewer systems, Water Science and Technology, 51, 119–128, https://doi.org/10.2166/wst.2005.0039, 2005.

Brown, J. D.: Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief, Transactions of the Institute of British Geographers, 29, 367–381, https://doi.org/10.1111/j.0020-2754.2004.00342.x, 2004.

Cecinati, F., Moreno-Ródenas, A. M., Rico-Ramirez, M. A., ten Veldhuis, M. C., and Langeveld, J. G.: Considering rain gauge uncertainty using kriging for uncertain data, Atmosphere, 9, 1–17, https://doi.org/10.3390/atmos9110446, 2018.

Datta, A. R.: Evaluation of Implicit and Explicit Methods of Uncertainty Analysis on a Hydrological Modeling, Ph.D. thesis, University of Windsor, Canada, 2011.

de Rocquigny, E., Devictor, N., and Tarantola, S., eds.: Uncertainty in Industrial Practice, John Wiley &Sons, Ltd, , 2008.

Deletic, A., Dotto, C., McCarthy, D., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T., Rauch, W., Bertrand-Krajewski, J., and Tait, S.: Assessing uncertainties in urban drainage models, Physics and Chemistry of the Earth, 42-44, 3–10, https://doi.org/10.1016/j.pce.2011.04.007, 2012.

DHI: MIKE11, A modeling system for rivers and channels, Reference Manual, DHI Water and Environment, 2017.

Diaz-Fierros T, F., Puerta, J., Suarez, J., and Diaz-Fierros V, F.: Contaminant loads of CSOs at the wastewater treatment plant of a city in NW Spain, Urban Water, 4, 291 – 299, https://doi.org/https://doi.org/10.1016/S1462-0758(02)00020-1, 2002.

Efron, B.: Bootstrap methods: Another look at the Jackknife, The Annals of Statistics, 7, 1–26, 1979.

Eränen, D., Oksanen, J., Westerholm, J., and Sarjakoski, T.: A full graphics processing unit implementation of uncertainty-aware drainage basin delineation, Computers & Geosciences, 73, 48 – 60, https://doi.org/https://doi.org/10.1016/j.cageo.2014.08.012, 2014.

Evers, P., Heinz, H., Hanitsch, P. H., Koch, G., Naupold, L., Tochtermann, W., Tornow, M., Zander, B., Mahret, H., and Warnow, D.: ATV-DVWK-A 134E: Planning and Construction of Wastewater Pumping Stations, Tech. rep., DWA, Germany, 2000.

Gasperi, J., Zgheib, S., Cladière, M., Rocher, V., Moilleron, R., and Chebbo, G.: Priority pollutants in urban stormwater: Part 2 – Case of combined sewers, Water Research, 46, 6693 – 6703, https://doi.org/10.1016/j.watres.2011.09.041, special Issue on Stormwater in urban areas, 2012.

Gräler, B., Pebesma, E., and Heuvelink, G.: Spatio-Temporal Interpolation using {gstat}, The R Journal, 8 (1), 204–218, http://journal.r-project.org/archive/2016-1/na-pebesma-heuvelink.pdf, 2016.

Hager, W. H.: Wastewater hydraulics, Springer, second edn., https://doi.org/10.1080/00221686.2011.614723, 2010.

Hammersley, J. and Handscomb, D.: Monte Carlo Methods, Methuen & Co Ltd, London, 1964.

Heip, L., Assel, J. V., and Swartentenbroekx, P.: Sewer flow quality modelling, Water Science & Technology, 36, 177–184, 1997.

Heuvelink, G. B. M.: Error Propagation in Environmental Modelling with GIS, Reserach Monographs in GIS, CRC Press Taylor & Francis Group, 1998.

Heuvelink, G. B. M., Brown, J. D., and van Loon, E. E.: A probabilistic framework for representing and simulating uncertain environmental variables, International Journal of Geographical Information Science, 21, 497–513, https://doi.org/10.1080/13658810601063951, 2007.

House, M. A., Ellis, J. B., Herricks, E. E., Hvitved-Jacobsen, T., Seager, J., Lijklema, L., Aalderink, H., and Clifforde, I. T.: Urban Drainage – Impacts on Receiving Water Quality, Water Science and Technology, 27, 117, https://doi.org/10.2166/wst.1993.0293, http://dx.doi.org/10.2166/wst.1993.0293, 1993.

Huang, H., Xiao, X., Yan, B., and Yang, L.: Ammonium removal from aqueous solutions by using natural Chinese (Chende) zeolite as adsorbent, Journal of Hazardous Materials, 175, 247–252, https://doi.org/10.1016/j.jhazmat.2009.09.156, 2010.

Hutton, C., Vamvakeridou-Lyroudia, L., Kapelan, Z., and Savic, D.: Uncertainty Quantification and Reduction in Urban Water Systems (UWS) Modelling: Evaluation Report, Tech. rep., The European Commission, 2011.

IFAK: SIMBA (Simulation of Biological Wastewater Systems): Manual and Reference, Tech. rep., Institut fÃ¼r Automation und Kommunikation e. V, Magdeburg, Germany, 2007.

Jerves-Cobo, R., Benedetti, L., Amerlinck, Y., Lock, K., De Mulder, C., Van Butsel, J., Cisneros, F., Goethals, P., and Nopens, I.: Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: Case study in the Cuenca River basin (Ecuador), Science of the Total Environment, 709, https://doi.org/10.1016/j.scitotenv.2019.136067, 2020.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering, 82D, 35–45, 1960.

Kalos, M. H. and Whitlock, P. A.: Monte Carlo Methods, Wiley-Blackwell, 2 edn., 2008.

Katukiza, A. Y., Ronteltap, M., Niwagaba, C. B., Kansiime, F., and Lens, P. N. L.: Grey water characterisation and pollutant loads in an urban slum, Int. J. Environ. Sci. Technol., 12, 423 436, https://doi.org/10.1007/s13762-013-0451-5, http://dx.doi.org/10.1007/s13762-013-0451-5, 2014.

Kleidorfer, M. and Rauch, W.: An application of Austrian legal requirements for CSO emissions, Water Science & Technology, 64, 1081, https://doi.org/10.2166/wst.2011.560, 2011.

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, Journal of Hydrology, 211, 69–85, https://doi.org/10.1016/S0022-1694(98)00198-X, 1998.

Leopold, U., Heuvelink, G. B. M., Tiktak, A., Finke, P. A., and Schoumans, O.: Accounting for change of support in spatial accuracy assessment of modelled soil mineral phosphorous concentration, Geoderma, 130, 368–386, 2006.

Litrico, X. and Fromion, V.: Modeling and Control of Hydroystems, Springer, https://doi.org/10.1017/CBO9781107415324.004, 2009.

Luetkepohl, H.: New Introduction to Multiple Time Series Analysis, Springer, 2005.

Marwick, B. and Krishnamoorthy, K.: cvequality: Tests for the Equality of Coefficients of Variation from Multiple Groups, https://github.com/benmarwick/cvequality, r package version 0.2.0, 2019.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, Journal of Hydrology, 400, 8394, https://doi.org/10.1016/j.jhydrol.2011.01.026, http://dx.doi.org/10.1016/j.jhydrol.2011.01.026, 2011.

Miskewitz, R. and Uchrin, C.: In-Stream Dissolved Oxygen Impacts and Sediment Oxygen Demand Resulting from Combined Sewer Overflow Discharges, Journal of Environmental Engineering, 139, 1307–1313, https://doi.org/10.1061/(ASCE)EE.1943-7870.0000739, 2013.

Moreno-Rodenas, A. M., Tscheikner-Gratl, F., Langeveld, J. G., and Clemens, F. H.: Uncertainty analysis in a large-scale water quality integrated catchment modelling study, Water Research, 158, 46 – 60, https://doi.org/10.1016/j.watres.2019.04.016, 2019.

Neumann, M. B.: Uncertainty Analysis for Performance Evaluation and Design of Urban Water Infrastructure, Ph.D. thesis, Swiss Federal Institute of Technology, ETH Zurich, 2007.

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

on Environmental, C. and atural Resources - CENR: An Assessment of Coastal Hypoxia and Eutrophication in U.S. Waters, Tech. rep., NSTC, National Science and Technology Council, 2003.

R-Core-Team and contributors worldwide: The R Stats Package, The R Project for Statistical Computing, 3.5.0 edn., https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html, 2017.

Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, J. Hydrol., 1997.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process - A framework and guidance, Environmental Modelling and Software, 22, 1543–1556, https://doi.org/10.1016/j.envsoft.2007.02.004, 2007.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, https://doi.org/10.1029/2009wr008328, 2010.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis : the primer, John Wiley & Sons, Ltd, Chichester, England, https://doi.org/10.1002/9780470725184, 2008.

Searle

Sandric, I., Ionita, C., Chitu, Z., Dardala, M., Irimia, R., and Furtuna, F. T.: Using CUDA to accelerate uncertainty propagation modelling for landslide susceptibility assessment, Environmental Modelling & Software, 115, 176 – 186, https://doi.org/https://doi.org/10.1016/j.envsoft.2019.02.016, 2019.

Sriwastava, A. K., Tait, S.: Linear Models, Wiley, 1997., Schellart, A., Kroll, S., Dorpe, M. V., Assel, J. V., and Shucksmith, J.: Quantifying Uncertainty in Simulation of Sewer Overflow Volume, Journal of Environmental Engineering, 144, 04018 050, https://doi.org/10.1061/(ASCE)EE.1943-7870.0001392, 2018.

Statec: Statistics Portal of the Grans-Duchy of Luxembourg, https://statistiques.public.lu, 2020.

Steinel, A. and Margane, A.: Best management practice guideline for wastewater facilities in karstic areas of Lebanon with special respect to the protection of ground- and surface waters, Tech. Rep. 2, Federal Ministry for Economic Cooperation and Development, Project: Protection of Jeita Spring, 2011.

Sten, J., Lilja, H., Hyväluoma, J., Westerholm, J., and Aspnäs, M.: Parallel flow accumulation algorithms for graphical processing units with application to RUSLE model, Computers & Geosciences, 89, 88 – 95, https://doi.org/https://doi.org/10.1016/j.cageo.2016.01.006, 2016.

Toffol, S. D.: Sewer system performance assessment - an indicators based methodology, Ph.D. thesis, Universität Innsbruck, Universität Innsbruck, Innsbruck, Austria, 2006.

Torres-Matallana, J., Leopold, U., and Heuvelink, G.: stUPscales: an R-package for spatio-temporal Uncertainty Propagation across multiple scales with examples in urban water modelling, Water, 10(7), 1–30, https://doi.org/https://doi.org/10.3390/w10070837, 2018a.

Torres-Matallana, J., Leopold, U., and Heuvelink, G.: stUPscales: Spatio-Temporal Uncertainty Propagation Across Multiple Scales, https://CRAN.R-project.org/package=stUPscales, r package version 1.0.5.0, 2019.

Torres-Matallana, J. A., Leopold, U., and Heuvelink, G. B. M.: Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models, European Water, 57, 299–306, 2017.

Torres-Matallana, J. A., Klepiszewski, K., Leopold, U., and Heuvelink, G.: EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows, Water, 10(6), 1–24, https://doi.org/10.3390/w10060782, 2018. 2018b.

Tscheikner-Gratl, F., Lepot, M., Moreno-Rodenas, A., and Schellart, A.: QUICS Deliverable 6.7: A Framework for the application of uncertainty analysis, Tech. rep., Delft University of Technology, University of Sheffield and CH2M, https://zenodo.org/record/1240926, 2017.

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M. A., de Carvalho, R. F., Breuer, L., Shucksmith, J., Heuvelink, G. B., and Tait, S.: Recent insights on uncertainties present in integrated catchment water quality modelling, Water Research, 150, 368 – 379, https://doi.org/10.1016/j.watres.2018.11.079, 2019.

van der Keur, P., Henriksen, H. J., Refsgaard, J. C., Brugnach, M., Pahl-Wostl, C., Dewulf, A., and Buiteveld, H.: Identification of major sources of uncertainty in current IWRM practice. Illustrated for the Rhine Basin, Water Resources Management, 22, 1677–1708, https://doi.org/10.1007/s11269-008-9248-6, 2008.

Viana da Silva, A. M. E., Bettencourt da Silva, R. J. N., and Camões, M. F. G. F. C.: Optimization of the determination of chemical oxygen demand in wastewaters, Analytica Chimica Acta, 699, 161–169, https://doi.org/10.1016/j.aca.2011.05.026, http://dx.doi.org/10.1016/j.aca.2011.05.026, 2011.

Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, Proceedings of the National Academy of Sciences of the United States of America, 2007.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resources Research, 39, https://doi.org/10.1029/2002wr001746, http://dx.doi.org/10.1029/2002WR001746, 2003a.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resources Research, 39, 2003b.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resources Research, 44, 1–15, https://doi.org/10.1029/2007WR006720, 2008.

Wadoux, A.-C., Brus, D., Rico-Ramirez, M., and Heuvelink, G.: Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model, Advances in Water Resources, 107, 126–138, https://doi.org/10.1016/j.advwatres.2017.06.005, cited By 13, 2017.

Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., and Krayer von Krauss, M.: Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, Integrated Assessment, 4, 5–17, https://doi.org/10.1076/iaij.4.1.5.16466, 2003.

Webster, R. and Heuvelink, G. B. M.: The Kalman filter for the pedologist's tool kit, European Journal of Soil Science, 57, 758 773, https://doi.org/10.1111/j.1365-2389.2006.00879.x, 2006.

Webster, R. and Oliver, M.: Geostatistics for Environmental Scientists, Wiley, 2nd edn., 2007.

Welker, A.: Emissions of pollutant loads from combined sewer systems and separate sewer systems – Which sewer system is better, in: 11th International Conference on Urban Drainage, edited by ICUD, Edinburgh, Scotland, UK, 2008.

Yu, L., Rozemeijer, J. C., Broers, H. P., van Breukelen, B. M., Middelburg, J. J., Ouboter, M., and van der Velde, Y.: Drivers of nitrogen and phosphorus dynamics in a groundwater-fed urban catchment revealed by high frequency monitoring, Hydrology and Earth System Sciences Discussions, 2020, 1–22, https://doi.org/10.5194/hess-2020-34, https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-34/, 2020.

Zhou, X., Polcher, J., Yang, T., and Huang, C.-S.: A new uncertainty estimation approach with multiple datasets and implementation for various precipitation products, Hydrology and Earth System Sciences, 24, 2061–2081, https://doi.org/10.5194/hess-24-2061-2020, https://www.hydrol-earth-syst-sci.net/24/2061/2020/, 2020.

Zoppou, C.: Review of urban storm water models, Environmental Modelling and Software, 16, 195–231, https://doi.org/10.1016/S1364-8152(00)00084-0, 2001.