Hydrology and Earth System Sciences

Open Access

Discussions

HESSD

Interactive comment

Interactive comment 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 et al.

arturo.torres@list.lu

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 will make this clear in the revision, for example by no longer claiming that we introduce an uncertainty analysis framework but instead recognise that we build on existing methods. 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 will move some of the materials to the supporting information, 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 judged 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).

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".

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

Reply: We agree that uncertainty analysis is an umbrella term and will make sure that we use it only in that way, while we use the terms 'uncertainty propagation' 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 will check the manuscript carefully to make the necessary changes, also in the title (replace 'uncertainty propagation' by 'uncertainty analysis') and some section headings (i.e. Sections 2.6 and 3.3 of the original 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 want to 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 we will remove 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: "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 will correct this in the revision. We will replace the text in lines 115-120 with:

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

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

where $1/6$ 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 will include the variable names into the text to be clearer which variables have been calibrated.

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

Reply: We will delete 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 will also change 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, 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 will remove this parameter from Table 2.

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 will replace "uncertainty analysis" by "uncertainty propagation analysis" in line 176.

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, 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, 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 Supporting Information. We prefer to keep the text in lines 215-222 in the main article to explain the reader what we did. We will rephrase this text to:

C9

"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."

L257-260: Remove or move to introduction.

Reply: We agree and will remove these lines.

L304: This is sensitive analysis and does not belong under the section "uncertainty propagation" Reply: We will change the title of Section 2.6 to "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 will 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 will change this sentence to: "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 will rephrase the sentence to "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.

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 will move this figure to the supporting information and will extend 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 are happy to provide the evidence in the Supporting Information.

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

We will move Section 3.3.1 to the Supporting Information and refer to it in the main text.

** 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 will correct this in the revised manuscript, by changing the text to: "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.

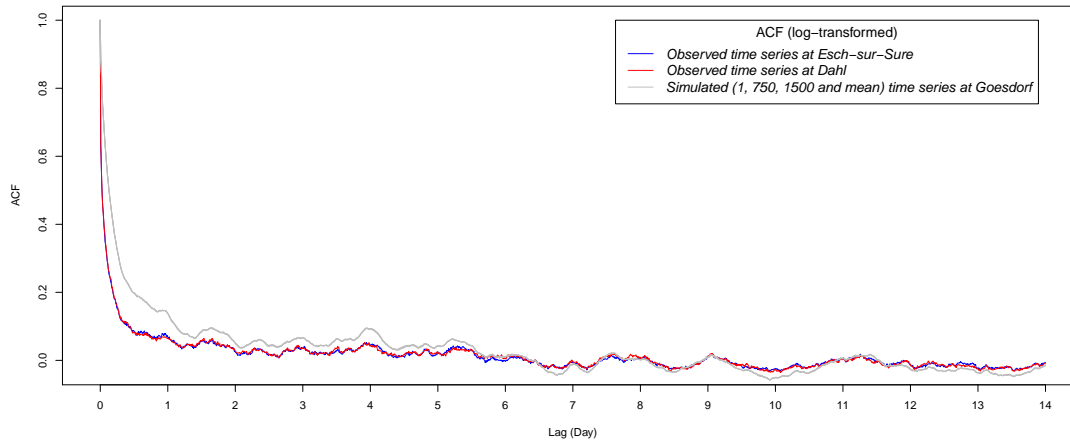Printer-friendly version

Discussion paper

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