



# Parametric emulation and inference in computationally expensive integrated urban water quality simulators

Antonio M. Moreno-Rodenas<sup>1,2</sup> · Jeroen G. Langeveld<sup>1,3</sup> · Francois H. L. R. Clemens<sup>1,2</sup>

Received: 27 November 2018 / Accepted: 28 May 2019 / Published online: 4 July 2019  
© The Author(s) 2019

## Abstract

Water quality environmental assessment often requires the joint simulation of several subsystems (e.g. wastewater treatment processes, urban drainage and receiving water bodies). The complexity of these integrated catchment models grows fast, leading to potentially over-parameterised and computationally expensive models. The receiving water body physical and biochemical parameters are often a dominant source of uncertainty when simulating dissolved oxygen depletion processes. Thus, the use of system observations to refine prior knowledge (from experts or literature) is usually required. Unfortunately, simulating real-world scale water quality processes results in a significant computational burden, for which the use of sampling intensive applications (e.g. parametric inference) is severely hampered. Data-driven emulation aims at creating an interpolation map between the parametric and output multidimensional spaces of a dynamic simulator, thus providing a fast approximation of the model response. In this study a large-scale integrated urban water quality model is used to simulate dissolved oxygen depletion processes in a sensitive river. A polynomial expansion emulator was proposed to approximate the link between four and eight river physical and biochemical river parameters and the dynamics of river flow and dissolved oxygen concentration during one year (at hourly frequency). The emulator scheme was used to perform a sensitivity analysis and a formal parametric inference using local system observations. The effect of different likelihood assumptions (e.g. heteroscedasticity, normality and autocorrelation) during the inference of dissolved oxygen processes is also discussed. This study shows how the use of data-driven emulators can facilitate the integration of formal uncertainty analysis schemes in the hydrological and water quality modelling community.

**Keywords** Bayesian inference · Integrated catchment modelling · Dissolved oxygen simulation · Model emulation

## Introduction

Integrated urban water quality modelling focuses on the joint simulation of processes driving pollution dynamics through the urban-river system (Muschalla et al. 2009; Rauch et al.

2002). These models jointly evaluate wastewater treatment processes, urban drainage and river dynamics, which usually generate a rapid escalation of complexity (Benedetti et al. 2013). The representation of all subsystems involved produces highly parameterised conceptualisations, requiring a large amount of data in the calibration process (Langeveld et al. 2013a). Additionally, the dynamics of interest often occur at very different time-space scales. For instance, urban combined sewer overflow (CSO) discharges have a characteristic timescale of minutes-hours whereas river dissolved oxygen dynamics exhibit hourly to monthly scales. Quantifying and analysing uncertainties in these platforms is hence required to avoid over-confidence in modelling results and to guide further model improvement (Deletic et al. 2012; Tscheikner-Gratl et al. 2017). However, the computational effort required is a severe limitation for the applicability of uncertainty analysis techniques for most real-scale integrated catchment modelling studies (Tscheikner-Gratl et al. 2019).

---

Responsible editor: Marcus Schulz

---

✉ Antonio M. Moreno-Rodenas  
a.m.morenorodenas@tudelft.nl

<sup>1</sup> Section Sanitary Engineering, Water Management Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands

<sup>2</sup> Department of Hydraulic Engineering, Deltares, 2600 MH Delft, The Netherlands

<sup>3</sup> Partners4UrbanWater, Javastraat 104A, 6524 MJ Nijmegen, The Netherlands

Many uncertainty quantification strategies rely on intensive model sampling applications (Dotto et al. 2012). For instance, parametric inference schemes often require a large number of model evaluations (on the order of  $10^4$ – $10^5$ ) to reach convergence. This hampers the use of formal uncertainty inversion methods in real-scale integrated urban water systems.

One approach to speed up convergence time is the use of optimised sampling schemes. For instance exploiting information from parallel model evaluations (Goodman and Weare 2010; Laloy and Vrugt 2012) or by using informed adaptive Markov chain Monte Carlo schemes (Hoffman and Gelman 2014). However, this still requires a prohibitive number of model samples, which often fall beyond the computational budget of most model users. A key strategy for the acceleration of model sampling is the use of data-driven or mechanistic model emulation, where a mathematical representation is used to approximate an interpolation map between a vector of parameters-inputs and the dynamic response of the simulator. Laloy et al. (2013) proposed a two-stage sampling scheme, first generating a rough estimate from a model surrogate and later from the simulator itself to perform parametric inference in a groundwater model. Carbajal et al. (2017) compared the performance of mechanistic vs. data-driven emulation for urban drainage simulators, concluding that in general a fully data-driven approach is to be preferred unless confronted with highly sparse training datasets. Yang et al. (2018) used a Gaussian process data-driven emulator to study parameter uncertainty in a semi-distributed hydrological model.

Data-driven emulators are constructed by drawing samples of the computationally expensive model at a selected number of input-parameter combinations. These samples are then used to build a database of input-output relationships. The emulator creates an interpolation between the two multi-dimensional spaces, thus allowing the fast estimation of the response when using a new parameter-input combination. Unfortunately, most mathematical structures used to emulate dynamic models (e.g. polynomial expansions, Gaussian processes) are sensitive to the dimensionality of the problem. The number of required samples to train the emulator increases non-linearly with the dimension of the input space (Xiu and Karniadakis 2002), thus reaching a point in which the construction of the emulator has an equivalent computational burden as using the simulator directly. Consequently, emulators often deal with a low number of static global parameters and a fixed time-window model output. The discretisation of input time series (as a parameter vector) can allow for the emulation of short time series (Mahmoodian et al. 2018) yet the length of the time series is limited to a few discrete steps, thus hampering its use in most cases. Hybrid strategies can be used to encode system knowledge in the data-driven emulator thus representing input dynamics. For instance, Moreno-Rodenas et al. (2018) presented a methodology to emulate hydrodynamic simulators (2D shallow water equations) under variations of parameters and time-

dynamic rainfall inputs by encoding unitary response nonlinearities in a polynomial expansion scheme. However, the generalisation of such input-parametric response emulation schemes to other variables (e.g. non-conservative water quality pollutants) still remains unaddressed.

Nevertheless, formal inference and intensive sampling techniques for uncertainty analysis are not being generally applied in integrated urban water quality modelling studies (Tscheikner-Gratl et al. 2019). This is primarily due to the high computational cost involved with such applications. The use of emulators can facilitate dealing with such large-scale modelling schemes, and thus further stimulate the consideration of modelling uncertainties in environmental studies. Moreno-Rodenas et al. (2019) presented an uncertainty analysis for a large-scale integrated catchment system for the assessment of water quality dynamics in the Dommel River (the Netherlands). The contribution of different uncertainty sources in dissolved oxygen depletion simulations in a highly urbanised river system was quantified. Forward uncertainty propagation showed that the use of prior knowledge (extracted from literature, measurements and expert elicitation) of the river physical and biochemical parameters captured roughly 70% of the statistical uncertainty in the simulation of dissolved oxygen dynamics. Performing inference directly on the original model structure is however prohibitive due to its high computational cost. An emulator structure was used to accelerate the model evaluation and thus updates system knowledge based on local observations. The development of this emulator and the inference of the model parameters are here presented.

This study discusses the application of a fully data-driven emulation scheme to accelerate the estimation of the dynamics of dissolved oxygen and river flow when varying a set of global river parameters. An emulator platform (polynomial orthogonal expansion) is created to represent an interpolation map between a set of river parameters (four water quantity and eight water quality process parameters) and the dynamic time series of river flow and dissolved oxygen concentrations at a location of interest. The training is performed by generating a database of model parameter to output relationships during the full year of 2012 (hourly frequency). The emulator is then used to implement a global sensitivity analysis and an inference scheme under various likelihood function conceptualisations. Consequently, this work shows that the use of a dynamic emulator scheme can facilitate the use of sampling intensive applications in large-scale simulators for water quality studies.

## Materials and methods

### The integrated catchment model

This modelling study targets the simulation of dissolved oxygen dynamics in the Dommel River. This is a sensitive stream located in the south of the Netherlands (Fig. 1). The river has a discharge

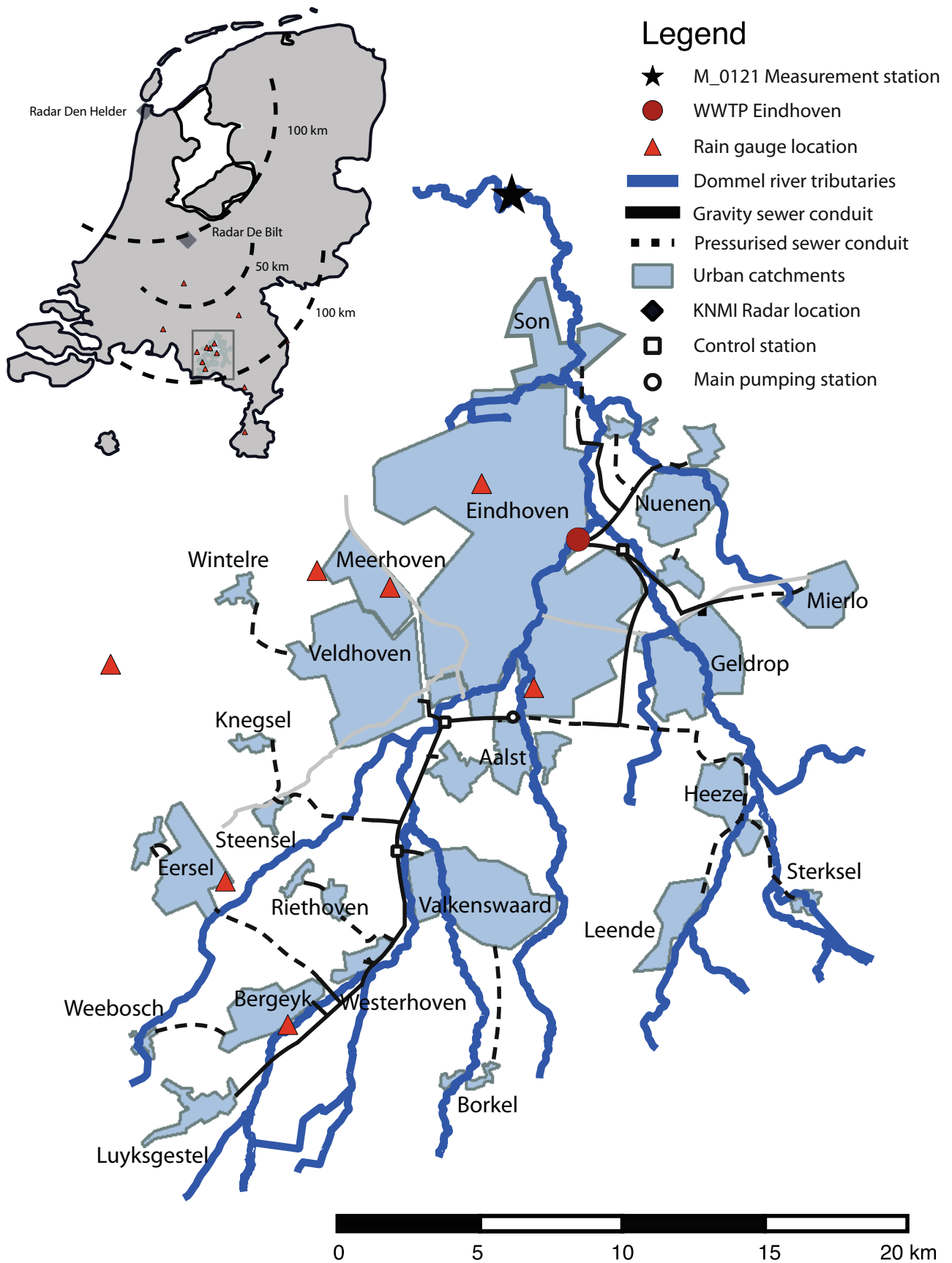


Fig. 1 Urban water system of the Dommel

between 2 and 20 m<sup>3</sup>/s, flowing through a mild-sloped lowland area. The river receives the discharge of 192 combined sewer overflow structures (CSOs) from several municipal urban drainage systems (connected urban area of ~4400 ha) and of a wastewater treatment plant (WWTP) of ~750,000 p.e. (population equivalent). High pollution loads from connected urban areas result in acute and chronic oxygen depletion events at the receiving water body. The integrated model accounts for the interaction between these three subsystems (river, urban drainage and WWTP). The river model is conceptualised as a tank-in-series hydrological scheme consisting 65 sections where the pollutant fluxes and transformation rates are computed (a conceptual scheme is presented in Annex A). The main river water quality processes are shown in Table 6 (Annex A). The WWTP was modelled through an a ASM2d (Gernaey et al. 2004) scheme representing three biological lines with primary clarifiers, activated sludge tanks and secondary clarifiers with a total capacity of 26,000 m<sup>3</sup>/h and a controlled bypass storm settling tank with a capacity of 9000 m<sup>3</sup>/h. Urban drainage flow was represented by 29 lumped rainfall-runoff and sewer transport schemes (Solvi 2006). Sewer water quality was represented by an influent generator at the WWTP (Langeveld et al. 2017) and by an event mean concentration multiplier at the CSO-receiving water links (Moreno-Rodenas et al. 2017b). The fully integrated model was implemented in the platform WEST (DHI). Further detail in the model and system characteristics can be found at Langeveld et al. (2013b) and Moreno-Rodenas et al. (2017a). Figure 14 (Annex A) depicts the model structure scheme. The river discharge and dissolved oxygen concentration were measured at a local station (M\_0121, Fig. 1) with hourly frequency. Table 1 depicts some of the main characteristics of the observed data,

**Dynamic emulator**

A polynomial chaos expansion (PCE) structure (Xiu (2010)) was used to create an interpolation map between a set of model parameters and the model outputs of river discharge and dissolved oxygen concentration (at the location M\_0121, Fig. 1) at hourly frequency during 1 year (01-01-2012–31-12-2012). PCE relies on fitting a series of orthogonal polynomials to the parametric-output multidimensional spaces. The model (*M*) consists of a large system of coupled differential equations, which represents the internal processes of the integrated water system to be emulated. An arbitrary time-dependent output state variable ( $Y_{sim} \in \mathbb{R}^D$ ) can be computed by solving the model:

$$Y_{sim} = M(\mathbf{x}_0, \mathbf{x}, \boldsymbol{\theta}_M, \boldsymbol{\theta}_I) \tag{1}$$

given a set of *m* initial conditions ( $\mathbf{x}_0 \in \mathbb{R}^m$ ), a set of *r* dynamic inputs of length *F* ( $\mathbf{x} \in \mathbb{R}^{xF}$ ) and a group of global model parameters [ $\boldsymbol{\theta}_M, \boldsymbol{\theta}_I$ ], which in this case is decomposed as two parameter subsets; a group of global model parameters ( $\boldsymbol{\theta}_M \in \mathbb{R}^S$ ), which value is fixed by the modeller during the emulation and a subset of *P* global model parameters ( $\boldsymbol{\theta}_I \in \mathbb{R}^P$ ) for which the modeller seeks to emulate.

The emulator is composed of a series of *N* orthogonal polynomials ( $\phi(\boldsymbol{\theta}_I) \in \mathbb{R}^{Nx1}$ ) such that the value of the dynamic variable of interest (at a certain combination of emulated parameters  $\boldsymbol{\theta}_I$ ) can be approximated:

$$Y_{sim}(t, \boldsymbol{\theta}_I) \approx \phi(\boldsymbol{\theta}_I)^T \cdot \mathbf{c} \tag{2}$$

where  $\mathbf{c} \in \mathbb{R}^{NxD}$  is a matrix of coefficients which is calibrated based on samples drawn from the simulator, thus creating a mapping between the parameter and output spaces ( $\mathbb{R}^P \rightarrow \mathbb{R}^D$ ). The training dataset is pre-computed by evaluating the model response at a number of *K* parameter combinations ( $\boldsymbol{\theta}_I = \mathbf{q}_i$  for  $i = 1 \dots K$ ). The training dataset is then used to calibrate the matrix of coefficients (*c*) such that:

$$\begin{bmatrix} Y_{sim}(t, \boldsymbol{\theta}_I = \mathbf{q}_1) \\ \vdots \\ Y_{sim}(t, \boldsymbol{\theta}_I = \mathbf{q}_K) \end{bmatrix} = \begin{bmatrix} \phi_1(\mathbf{q}_1) & \dots & \phi_N(\mathbf{q}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\mathbf{q}_K) & \dots & \phi_N(\mathbf{q}_K) \end{bmatrix} \cdot \begin{bmatrix} c_1(t) \\ \vdots \\ c_N(t) \end{bmatrix} \tag{3}$$

from which the polynomial values at each parameter sample,  $\phi_j(\mathbf{q}_i)$  and the model output  $Y_{sim}(t, \boldsymbol{\theta}_I = \mathbf{q}_i)$  are known. A least squares approach was used to calibrate the set of coefficients *c* for each variable of interest. Then, Eq. (2) can be used to approximate the output variable at a new combination of emulated model parameters. Further information about the fit of polynomial expansions and the selection of orthogonal series can be found at Hadigol and Doostan (2018), Feinberg (2015) and Xiu and Karniadakis (2002).

Two polynomial expansions were used to emulate the modelled flow and DO dynamics at the receiving water body (1-year, hourly frequency series at the location M\_0121, Fig. 1). The model response was emulated for variations of four (flow) and eight (DO water quality) model parameters respectively. The expansion was created using an orthogonal Legendre polynomial series (Gautschi 1994), truncated at 3rd order, and 200 training samples were drawn for both parameter spaces. Table 2 depicts the four river parameters for the emulation of flow and Table 3 shows the parameters for the emulation of DO along with their distribution (PCE training). Training samples were drawn using a Latin hypercube sampling (LHS) scheme.

**Table 1** Data characteristics at the observed period (2012)

Variable	Units	Location	Frequency	Mean	Std	Min	Max
River discharge	m <sup>3</sup> · s <sup>-1</sup>	M_0121	Hourly	6.5	3.3	2.2	23
River DO	gO <sub>2</sub> · m <sup>-3</sup>	M_0121	Hourly	6.9	2.1	0.3	11.6

**Table 2** River hydrology parameter PCE training ranges (emulation) and prior distributions (inference)

Name	Units	Description	PCE training	Prior distribution
$n$	$s \cdot m^{-1/3}$	Manning roughness	$\sim U(0.02, 0.15)^*$	$\sim U(0.025, 0.12)$
$k_z$	–	Embankment slope multiplier	$\sim U(0.3, 2)$	$\sim U(0.7, 1.3)$
$k_w$	–	River bed width multiplier	$\sim U(0.3, 2)$	$\sim U(0.5, 1.5)$
$k_{surface}$	–	Rural flow input multiplier	$\sim U(0.3, 2)$	$\sim U(0.7, 1.3)$

\* $\sim U(a, b)$  refers to uniformly distributed probability density function between  $a$  and  $b$

The integrated urban water quality model depends on a large number of dynamic inputs and submodel parameters (e.g. urban drainage in-sewer parameters, WWTP parameters, rainfall inputs) some of which were of a stochastic nature. All model inputs and parameters were fixed to a deterministic realisation and only the emulated parameters were varied during the training database sampling.

Additionally, 100 and 50 random parameter samples were drawn independently for the flow and DO-simulated outputs respectively. This independent set was used to test the emulator’s performance by comparing the Nash-Sutcliffe efficiency (NSE) between the emulator vs. simulator output series at random parameter realisations.

The fitted coefficients of the polynomial expansion have a direct interpretability in terms of the sensitivity of the different parameters (Xiu 2010). Also, the emulator can be used to cheaply evaluate any combination of the parameters within the training parameter range, hence facilitating the application of sampling intensive system analysis tools. In order to describe the sensitivity of the studied model parameters, the Sobol global sensitivity analysis (Sobol 1993) was applied to the flow and DO river dynamics.

**Parametric inference**

Prior knowledge of the river model parameters was encoded by means of independent uniform probability density functions. The parameter distribution ranges were defined based on literature values and expert criteria (non-formal elicitation). Tables 2 and 3 show the prior probability density function selected for each parameter.

**Table 3** River dissolved oxygen parameter PCE training ranges (emulation) and prior distributions (inference)

Name	Units	Description	PCE training	Prior distribution
$Kd1$	$d^{-1}$	Decay rate for BOD fast	$\sim U(0.3, 1)$	$\sim U(0.3, 0.8)$
$Kd2$	$d^{-1}$	Decay rate for BOD slow	$\sim U(0.2, 1)$	$\sim U(0.2, 0.4)$
$Vs1$	$m \cdot d^{-1}$	Sedimentation rate for BOD fast	$\sim U(0.2, 40)$	$\sim U(0.5, 20)$
$Vs2$	$m \cdot d^{-1}$	Sedimentation rate for BOD slow	$\sim U(5, 100)$	$\sim U(10, 60)$
$TKd$	–	Temperature coefficient for BOD oxidation	$\sim U(1, 1.1)$	$\sim U(1, 1.1)$
$TKL$	–	Temperature coefficient for reaeration	$\sim U(1, 1.1)$	$\sim U(1, 1.03)$
$TSOD$	–	Temperature coefficient for SOD	$\sim U(1, 1.1)$	$\sim U(1, 1.1)$
$VKL$	–	Velocity reaeration coefficient	$\sim U(2, 8)$	$\sim U(2, 5)$

This prior knowledge was updated by using an observation layout  $Y_{obs} \in \mathbb{R}^{1 \times L}$  for hourly measured flow and dissolved oxygen concentration at the outlet of the Dommel catchment during a period of approximately 7 months (15-Jan-2012–05-Aug-2012). The basic model observation layout was defined as:

$$Y_{obs} = M(x_0, x, \theta_M, \theta_I) + Z \tag{4}$$

where  $M$  refers to the integrated catchment model. Initial conditions  $x_0$  were computed using a warming up simulation period of 1 year for the WWTP initial conditions and a dedicated initialisation of the previous 15 days between 01-Jan-2012 until 15-Jan-2012 for the rest of the variables. The term  $Z$  refers to the residual structure between the simulated and measured series. This error term lumps measurement and model errors together. During the inference process, a probabilistic description of the model-measurement residuals  $Z$  is assumed a priori, and is later validated based on the posterior computed residuals. A common initial guess is to assume that residuals are independent, identically and Gaussianly distributed. This assumption leads to the following log-likelihood structure:

$$\ell(Y_{obs}|\theta_I) \propto -\frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (Y_{obs} - M(\theta_I))^T \cdot \Sigma^{-1} \cdot (Y_{obs} - M(\theta_I)), \tag{5}$$

where  $\Sigma$  represents the residual covariance function, which in this case:

$$\Sigma_Z = \sigma_1^2 \cdot \mathbf{I} \tag{6}$$

being  $\mathbf{I} \in \mathbb{R}^{L \times L}$  the identity matrix and  $\sigma_1^2$  the constant variance of the residuals.



Flow dynamics are also known to render heteroscedastic error structures. This implies that residuals trend to be systematically larger when the discharge is larger. This is often encoded by assuming that the residual standard deviation follows a linear relationship with the simulated variable. This results in a log-likelihood function with the form described in Eq. (5) with the following covariance matrix:

$$\Sigma_{Z\_het} = (\sigma_1 + \mathbf{Q}_F \cdot \sigma_2)^2 \cdot \mathbf{I} \quad (7)$$

where  $\mathbf{I} \in \mathbb{R}^{L \times L}$  is the identity matrix,  $\mathbf{Q}_F \in \mathbb{R}^{L \times L}$  is the computed output time series; meanwhile,  $\sigma_1$  and  $\sigma_2$  are the hyperparameters of the error generating process.

Additionally, the inference of dynamic models often leads to autocorrelated residual structures. Previous studies in the hydrological literature have often taken this into account by the use of a discrete autoregressive model of order  $p$  (Bates and Campbell 2001), or as formulated by Schoups and Vrugt (2010):

$$\Phi_p(B) \cdot z_t \sim N(0, \sigma_1) \quad (8)$$

being  $\Phi_p(B) = 1 - \sum_{i=1}^p p_i \cdot z_{t-i}$  an autoregressive polynomial of order  $p$  for the residual  $z_t$ , with Gaussian updates.

An equivalent formulation to account for a correlation structure was discussed by Honti et al. (2013) with the use of a bias description stochastic process  $B$  along with the error generating model ( $Z$ ). If assuming a stationary continuous constant bias and heteroscedastic residuals, Eq. (5) defines the log-likelihood function, with a covariance matrix defined as:

$$\Sigma_{(Z\_het+B)} = \left( \sigma_1 + \mathbf{Q}_{F(t_i)} \cdot \sigma_2 \right)^2 \cdot \delta_{ij} + \sigma_3^2 \cdot e^{-|d_{i,j}| \cdot \tau^{-1}} \quad (9)$$

which are the  $i$  and  $j$  elements of the covariance matrix  $\Sigma_{(Z\_het+B)} \in \mathbb{R}^{L \times L}$ , with  $\delta_{ij}$ , the Kronecker's delta,  $\mathbf{Q}_{F(t_i)}$  the expected flow (at time  $t_i$ ),  $d_{i,j}$  the distance in hours between  $i$  and  $j$  elements,  $\sigma_3$  a parameter of the stationary bias and  $\tau$  an extra hyperparameter which drives a correlation exponential decay. Del Giudice et al. (2013) discuss that in practice the model effect and the bias descriptors can have a poor identifiability, thus inferring both, model parameters and bias hyperparameters, which often require assigning strong priors to the latter. Table 4 presents the prior distributions for the hyperparameters of the different likelihood distribution structures.

Posterior samples were created using a Metropolis-Hasting algorithm (Hastings 1970; Metropolis et al. 1953). The joint prior probability distributions for the flow and dissolved oxygen river parameters were updated by drawing 50,000 samples from their posterior distribution by means of a Markov chain Monte Carlo sampling scheme (25,000 burn-in, 5 thinning). The Bayesian inference implementation was performed using the python probabilistic programming package PyMC version 2.3.6 (Patil et al. 2010).

Evaluating the likelihood distribution when using the full bias-description term involves inverting a covariance matrix of size  $n$ . In this case, this was prohibitively expensive when using the original measurement layout ( $n = 4892$ ). Thus, a shorter period was used to test the inference of the bias description (26-Jul-2012–14-Sep-2012). In this case, only 4000 accepted samples were used (2000 burn-in, 2 thinning).

## Results and discussion

### Dynamic emulation of flow and dissolved oxygen concentrations

The performance of the trained emulator to represent the integrated catchment model outputs (at new parameter combinations) was tested using an independent dataset. Figure 2 shows the Nash-Sutcliffe efficiency (NSE) between the emulated and simulated flow time series (1-year, hourly frequency) at 100 random parameter combinations for the flow emulator. Figure 3 shows the same test performed at 50 random samples drawn from the dissolved oxygen concentration emulation scheme. The performance of both emulator implementations is consistent across the parameter ranges and varies between 0.99–1 NSE. The observed performance during validation was considered sufficient for the substitution of the simulator by the emulator during the inference sampling.

Figures 17 and 18 (Annex C) show also a graphical comparison between the time series outputs from the emulator and the simulator at a series of random combinations of parameters independent from the samples drawn at the training dataset. Table 5 presents the computational effort required to sample from the original simulator, training and operation of the emulator. In this case, the computed average timings refer to a 2.2-GHz Intel Core i7 from mid 2014.

### Global sensitivity analysis of process parameters

The emulator structure was used to estimate the sensitivity of the integrated catchment model outputs to variations of the river physical and biochemical parameters. Figures 4 and 5 depict the first-order Sobol sensitivity indexes from the prior distribution of parameters at the river flow and DO dynamics. Figure 4 shows the simulated flow level is highly sensitive on the parameter  $k_{surface}$  (which drives the river base-flow input) during dry-weather periods, whereas the manning's roughness ( $n$ ) becomes more sensitive during the rising limb of the hydrographs.  $k_w$  (multiplier for the river bed width) shows a reduced influence.  $k_z$  (a multiplier for the slope of the embankment) has a similar, yet less pronounced effect when compared with hydraulic roughness.

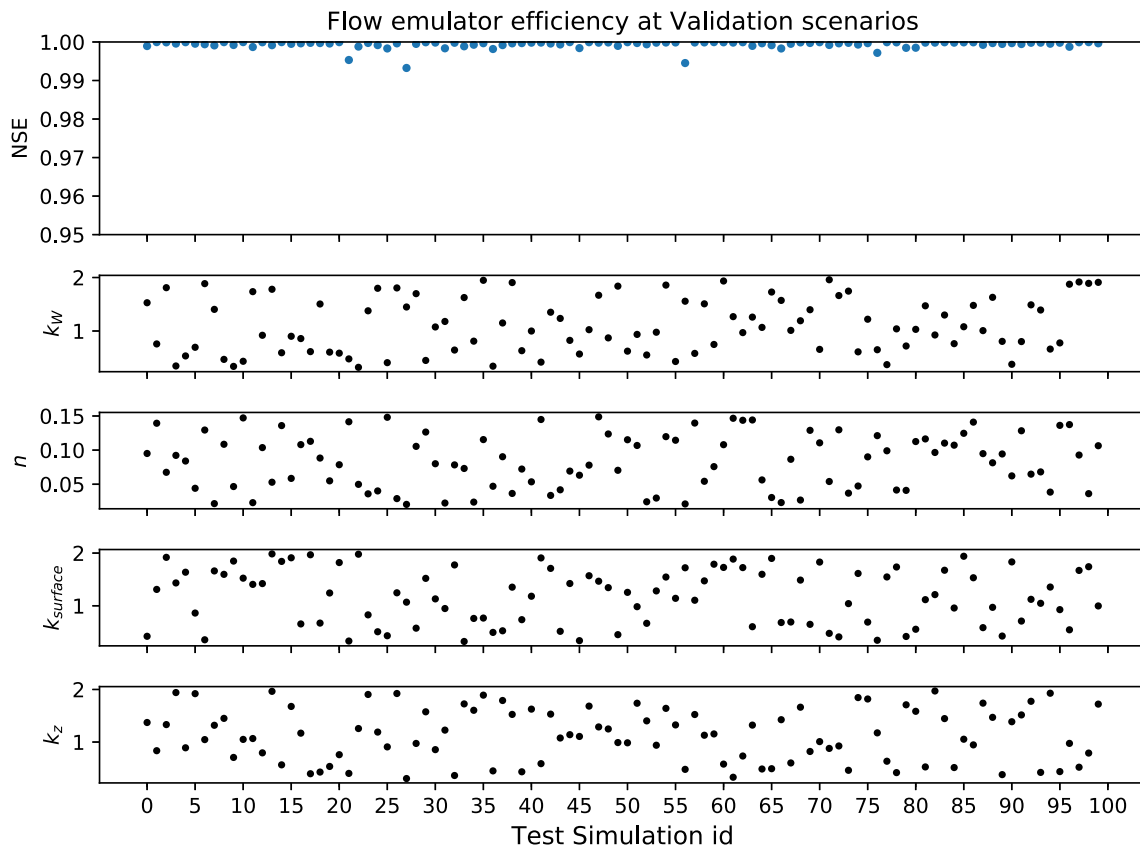
The results of the study of the sensitivity for the DO concentration simulation are shown in Fig 5. The parameter

**Table 4** Error model hyperparameters for the different hypotheses

Hyperparameter	Units	Description	Prior distribution
Flow i.i.d Gaussian			
$\sigma_1$	$m^3/s$	$\sigma_1$ , stationary standard deviation error	$\sim U(0, 10)$
Flow independent heteroscedastic Gaussian			
$\sigma_1$	$m^3/s$	$\sigma_1$ , stationary standard deviation error	$\sim U(0, 10)$
$\sigma_2$	$m^3/s$	$\sigma_2$ , stationary standard deviation error	$\sim U(0, 10)$
Flow AR(3) Gaussian updating			
$\sigma_1$	$m^3/s$	$\sigma_1$ , stationary standard deviation error	$\sim U(0, 10)$
$p_{1, 2, 3}$	–	Autocorrelation coefficients p1, p2, p3	$\sim U(0, 1)$
Flow heteroscedastic normal error and exponentially correlated bias			
$\sigma_1$	$m^3/s$	$\sigma_1$ , linear intercept standard deviation error	$\sim U(0, 10)$
$\sigma_2$	$m^3/s$	$\sigma_2$ , linear slope standard deviation error	$\sim U(0, 10)$
$\sigma_3$	$m^3/s$	$\sigma_3$ , bias standard deviation	$\sim U(0, 10)$
$\tau$	h	Tau, bias correlation exponential decay	$\sim U(10, 80)$
Dissolved oxygen i.i.d Gaussian			
$\sigma_1$	$mgO_2/l$	$\sigma_1$ , stationary error standard deviation	$\sim U(0, 10)$

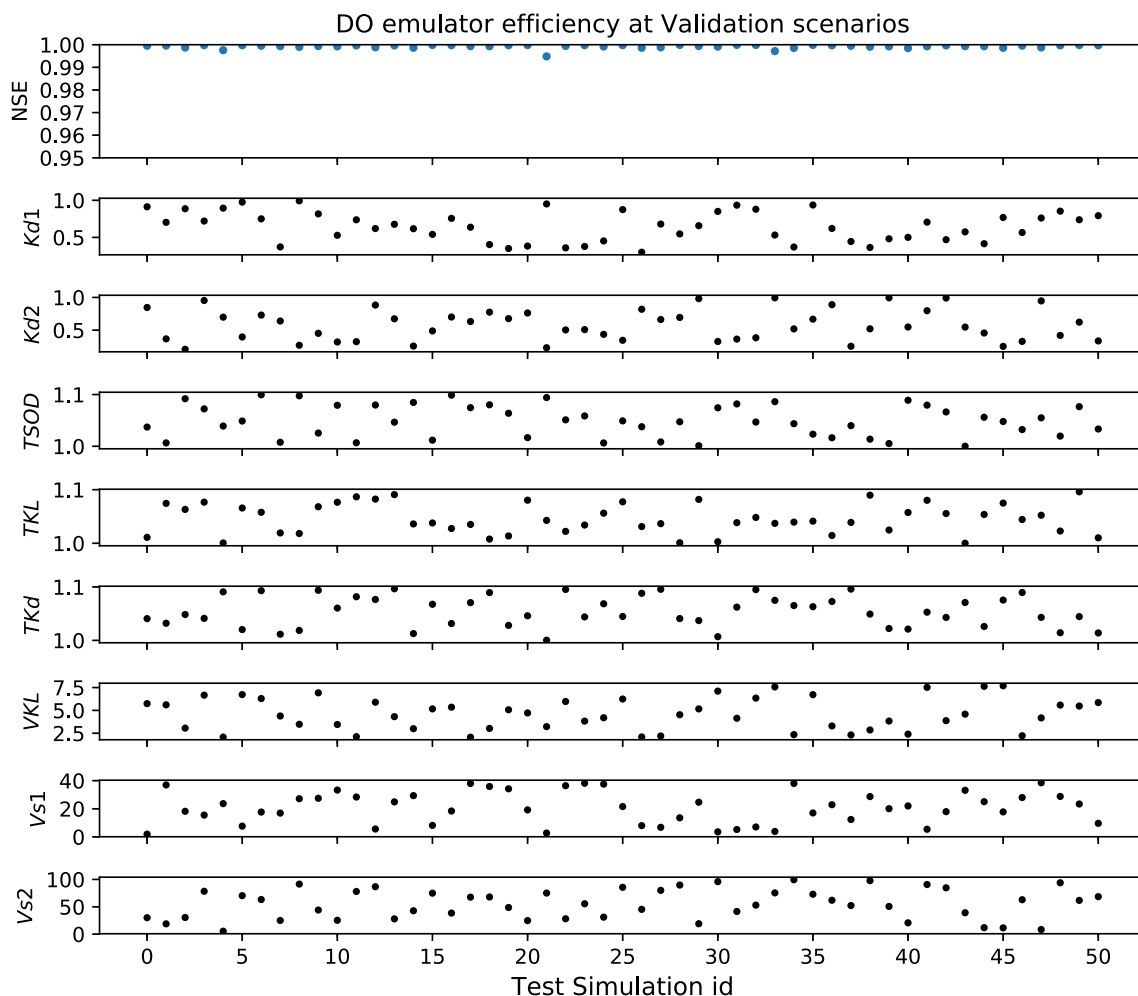
controlling the reaeration rate (*VKL*) dominates the dry-weather DO variability during summer times. This influence decreases during winter, where the temperature coefficient for the sediment oxygen demand (*TSOD*) becomes increasingly relevant. This has to do with the temperature inhibition model

structure, which influences the oxidation rate of organic matter for temperatures differing from 20 °C (Annex A, Table 6). During oxygen recovery patterns, *TSOD* is also relatively relevant, since the sediment layer becomes the main oxygen sink (days after a large storm event). Sensitivity indexes do not



**Fig. 2** Nash-Sutcliffe efficiency (NSE) at the flow emulator vs simulation for a four-dimensional parameter space under validation conditions. The x-axis shows the 100 combinations of the parameter values (simulation id

0 to 99). Above, the NSE between 1-year hourly frequency time series simulated by the model and the emulator for each parameter combination



**Fig. 3** Nash-Sutcliffe efficiency of dissolved oxygen emulator vs simulator for an eight-dimensional parameter space under validation conditions. The x-axis shows the 50 combinations of the parameter values

(simulation id 0 to 49). Above, the NSE between 1-year hourly frequency time series simulated by the model and the emulator for each parameter combination

show a consistent behaviour during acute oxygen depletion processes. Some depletion events, as the three occurring during July and September (also seen at Fig. 5b), present as dominant parameters *kd1* and *kd2* which are the oxidation rates for the two fractions of suspended BOD in the system. However, the events occurred in June and the three in October showed to be more sensitive to a different parameter combination as *TKL* or *TSOD*, which are related to temperature-driven reaeration or oxygen consumption. This is a good example of the complexity of the underlying process, in which interactions are

highly dependent on the dynamic state of the system. For instance, if a storm event activates predominantly northern CSOs (Fig. 1), which are closer to the outlet of the catchment, there is less time for the degradation of suspended matter to occur than a more upstream storm process. Also, events in which the WWTP is the main source of discharge (and not CSOs), the settling facilities of the WWTP might lead to a lower sediment build-up in the river and thus increasing the relevance of suspended organic matter degradation.

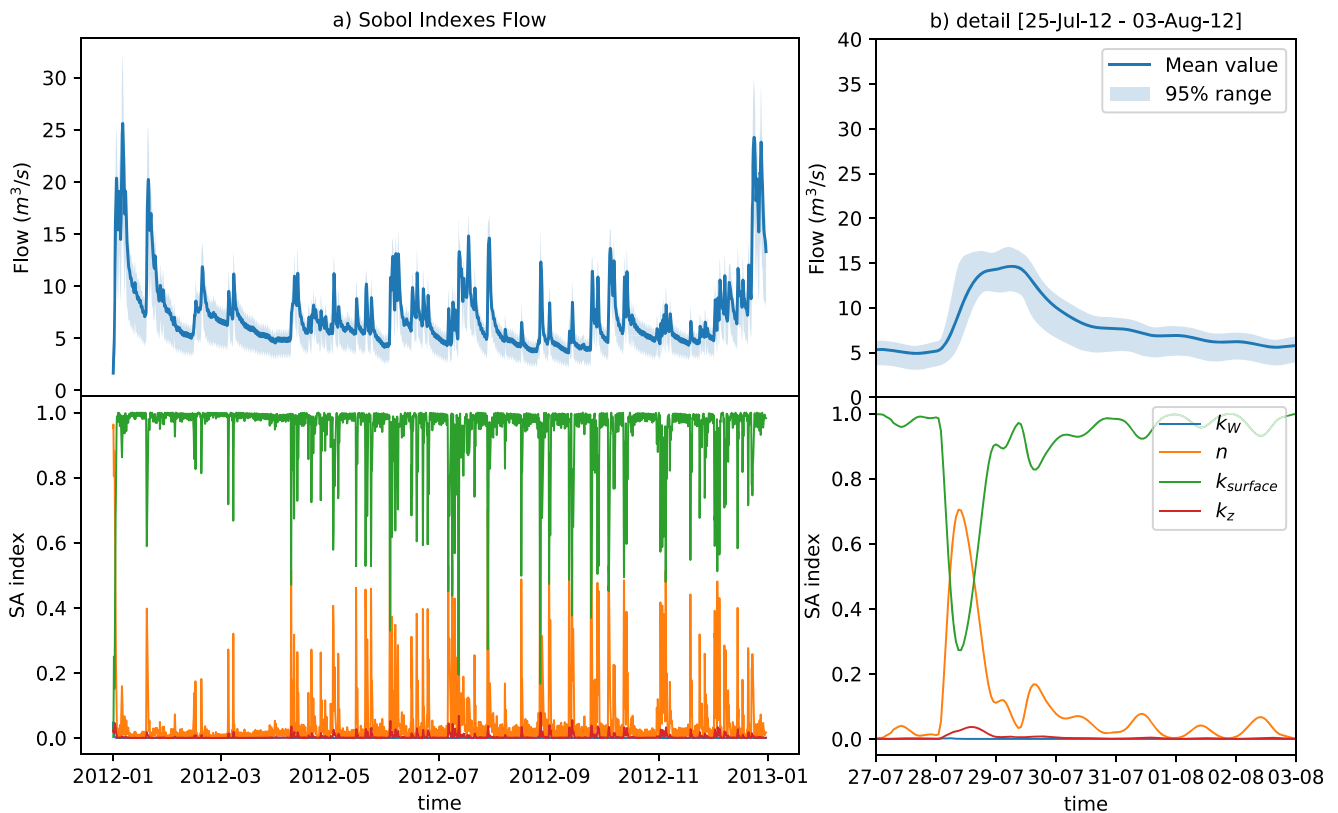
**Table 5** Emulation vs. model computational effort for 1-year hourly frequency series (in seconds)

Sample	Flow	DO
Simulator sample	3300 s	3300 s
Training database ( $\times 200$ simulator samples)	$660 \times 10^3$ s	$660 \times 10^3$ s
Emulator training	14 s	61 s
Emulator sample	0.06 s	0.07 s

### Parametric inference

A local dataset was used to update the river parameters prior knowledge. The emulator allowed drawing fast samples (Table 5) from the posterior distribution of the parameters given several hypotheses for the error generating process (Gaussian, independent and homoscedastic for DO dynamics and Gaussian, independent and heteroscedastic for the river flow dynamics). Figure 6 displays the comparison of





**Fig. 4** Sobol sensitivity indexes (first order) for the flow dynamics. Above, mean flow simulation and the 95% interval for the propagation of the parametric ranges. Below, sensitivity indexes for the four parameters. In the right, detail of the sensitivity during a medium-high intensity storm event (b)

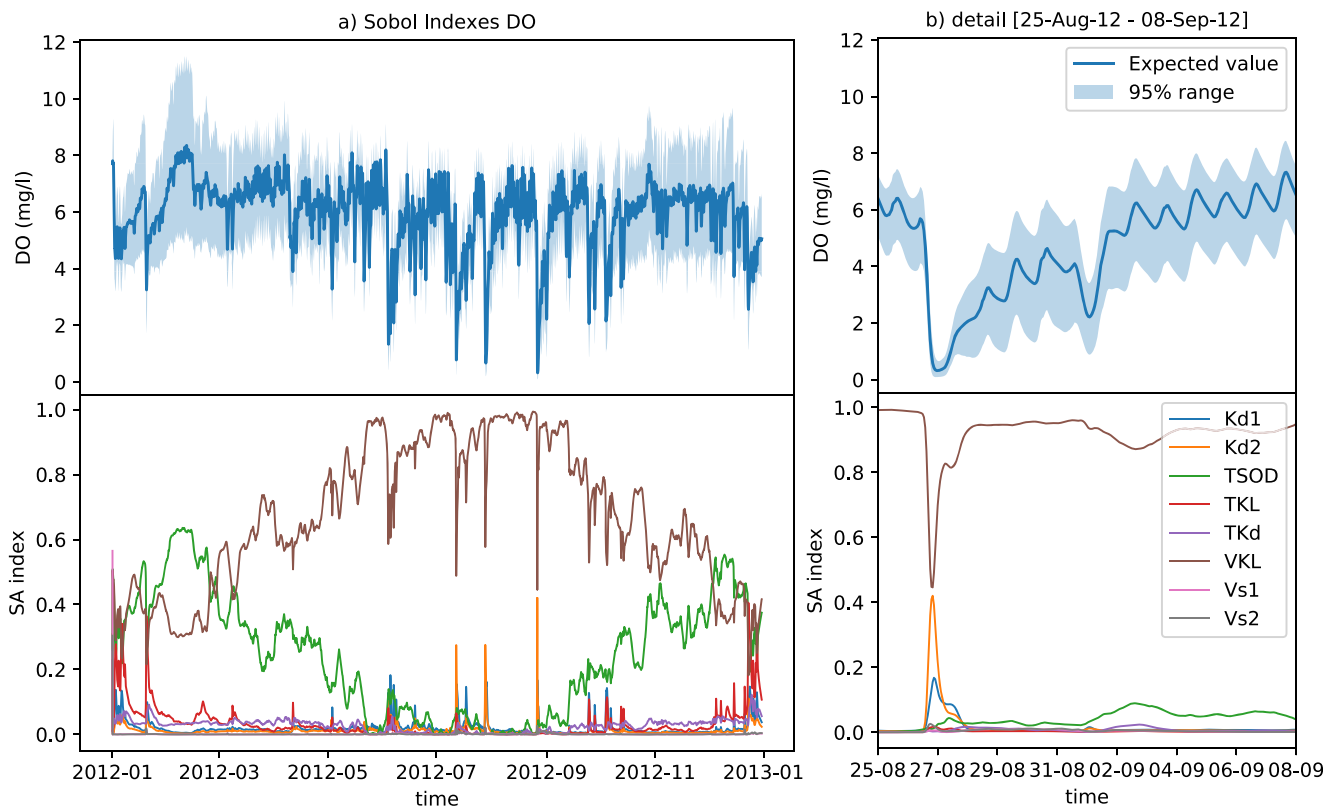
measured flow and the inferred mean model response. Also, a validation period (05-Aug-2012 until 31-Dec-2012) is shown. Intense dry weather periods induce a systematic overestimation of the flow as seen in July and in the beginning of September, this however is expected to have a limited influence in the water quality dynamics. The same comparison (measurement vs inferred and validation series) can be found in Fig. 7 for the simulation of dissolved oxygen. The general dynamics of DO are captured, especially the depletion processes, daily and seasonal variation.

The posterior probability density functions of the parameters for the water quantity and quality of the river section can be found in Figs. 8 and 9 respectively. The river variable  $k_W$  is poorly identified, which is denoted by the wide range of the posterior distribution (diagonal  $k_W$  at Fig. 8). This is also supported by the very low sensitivity of this parameter to the overall flow dynamics (Fig. 4). The rest of the parameters appear to be identifiable and are mostly mutually independent with the exception of a strong negative correlation between  $k_z$  and  $n$  (Spearman’s correlation coefficient,  $\rho_s = -0.78$ ). Therefore, the joint inference/calibration of both elements is not recommended, since the provided observations lack sufficient information to identify these parameters independently. Further use of this model should therefore prioritize fitting  $n$ , since it exhibits a larger sensitivity than  $k_z$ .

Water quality variables show a mostly independent joint posterior distribution with the exception of  $kd1$  and  $kd2$ , which show a mild negative correlation ( $\rho_s = -0.29$ ). This is explained by the fact that both parameters influence the same process (oxidation of organic matter) at two fractions of BOD for which DO measurements are probably insufficient to discriminate.

### Error-generating process and likelihood description

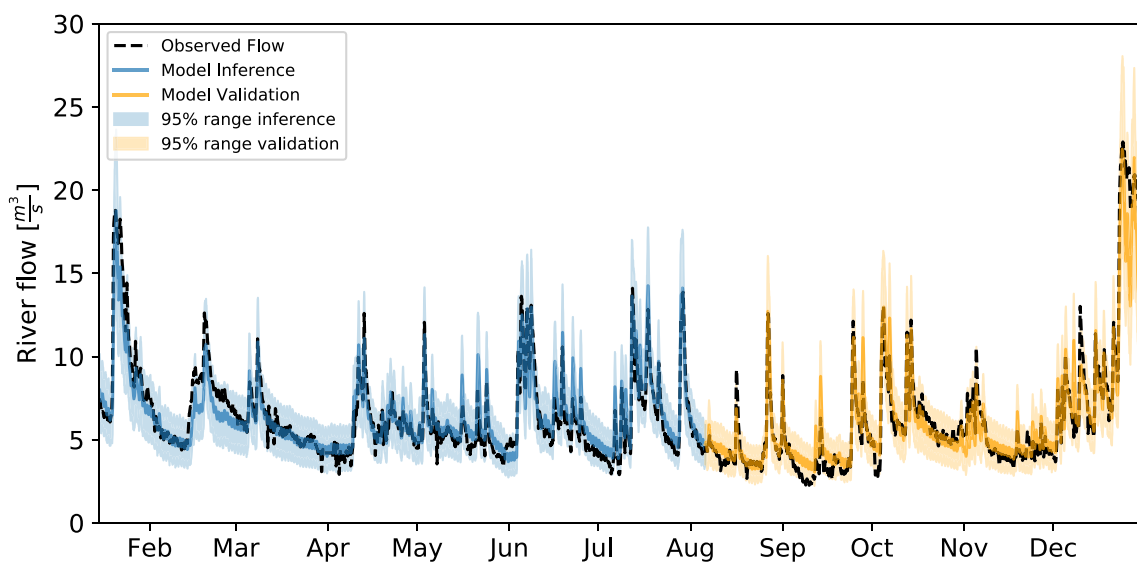
Bayesian inference relies on the a priori definition of an error-generating process (Eq. 4), which constitutes the likelihood structure used during the inference scheme. The error-generating process is selected based on a series of hypotheses, which can be encoded by expert guesses on the behaviour of the system. Yet those assumptions are still a subjective exercise and its validity should be checked once sampled from the posterior distribution. In this case, the initial error generation process for both flow and DO series was conceptualized as an independent, identically distributed Gaussian distribution. The posterior distribution for the flow process revealed a dependency of the residuals and the inferred flow dynamics. Such phenomena is well described in the hydrological literature (Sorooshian and Dracup 1980) and was corrected by the use of a linear dependent standard deviation structure in the river flow error-generating process (Eq. 7). Figure 10 represents three relevant characteristics of the residual structure at the posterior samples of river flow,



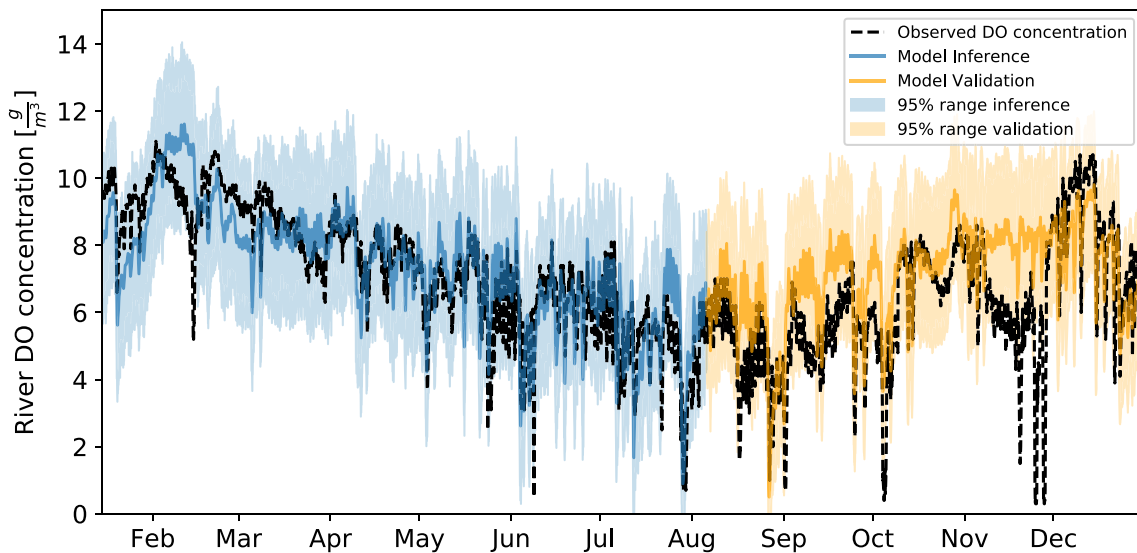
**Fig. 5** Sobol sensitivity indexes (first order) for the dissolved oxygen dynamics. Above, mean DO simulation and the 95% interval for the propagation of the parametric ranges. Below, sensitivity indexes for the eight parameters. In the right (b) detail of the sensitivity during a high intensity storm event

comparing the assumed error-generating process (in black) and the observed one (in blue). Figure 10a shows the heteroscedastic structure of flow residuals, Fig. 10b shows the comparison of the residual histograms and Fig. 10c shows the time-autocorrelation structure. It is apparent that the residual independency assumption is violated, since computed residuals present a strong time autocorrelation structure.

Fig. 11 shows the comparison of the computed residuals and the assumed error-generating process (independent, homoscedastic and Gaussian) for the dissolved oxygen in the river. The variance of the residuals is largely independent from the DO value. Also, residuals present a clear autocorrelation structure, albeit shorter than that of the flow inference.



**Fig. 6** Posterior sample for the inferred flow dynamics between 15-Jan-2012 and 05-Aug-2012. In orange, the posterior distribution under validation conditions 05-Aug-2012 until 31-Dec-2012, in black observed flow at the station M0121



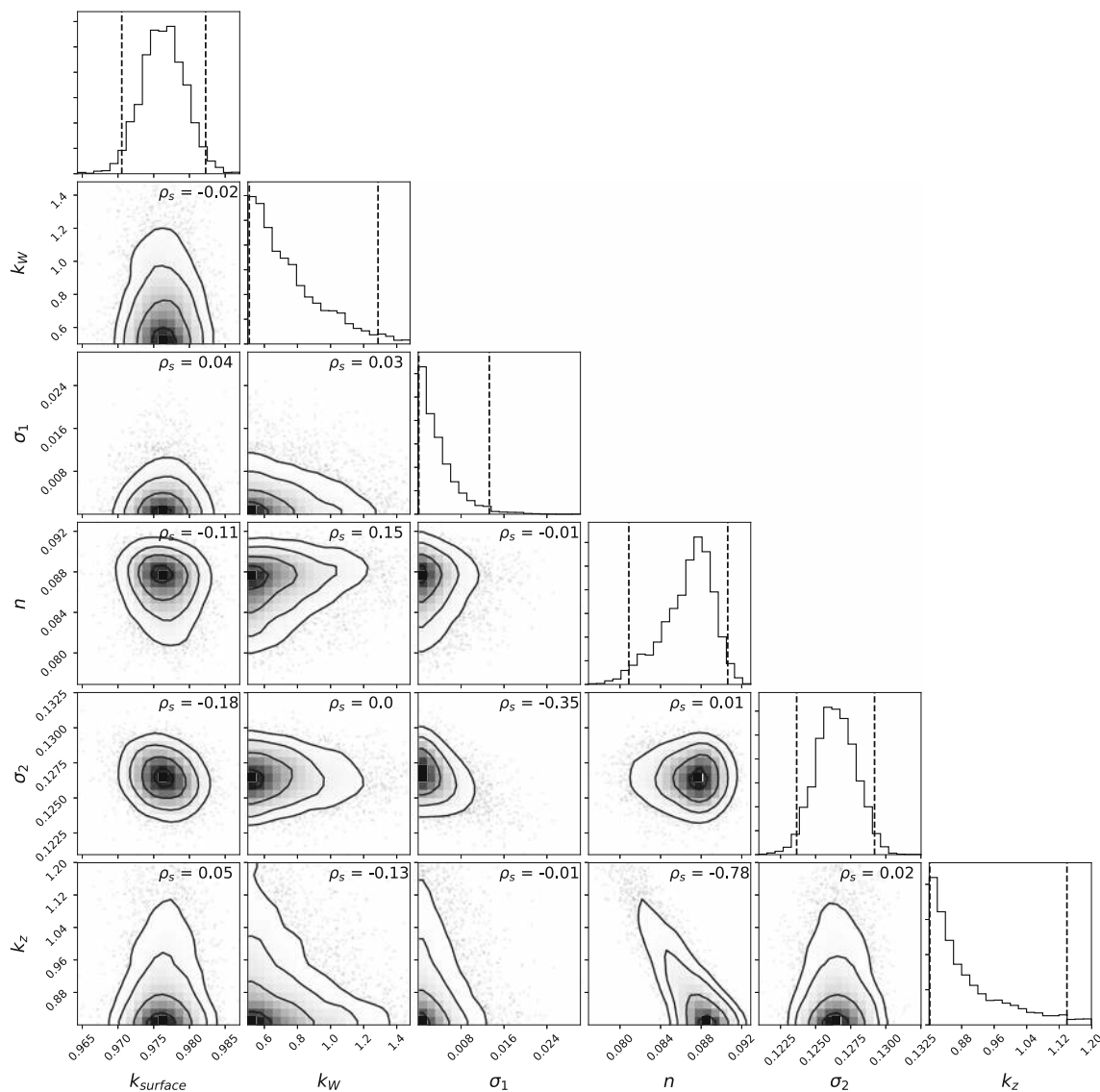
**Fig. 7** Posterior sample for the inferred dissolved oxygen dynamics between 15-Jan-2012 and 05-Aug-2012. In orange, the posterior distribution under validation conditions 05-Aug-2012 until 31-Dec-2012, in black observed flow at the station M0121

The time autocorrelation structure in hydrological inference has been discussed in several studies. For instance, Kuczera (1983) applied an ARMA (autoregressive moving-average) model to represent an autocorrelated likelihood structure in a hydrological model. Bates and Campbell (2001) argued that ARMA structures lead to local minima, and AR (autoregressive) models of order  $p$  are to be preferred. Schoups and Vrugt (2010) presented the use of an alternative likelihood structure, which addresses several common issues; like the non-normality of residuals, variance non-stationarity and the temporal correlation of the residuals (captured by an AR( $p$ ) model). Yet all these three studies simulated catchment hydrological flows at daily scales. In this work, the measurement layout has an hourly time-step, since relevant processes occur at those scales. It is expected that the autocorrelation structure becomes stronger when dealing with shorter timescales. As seen in Figs. 10 and 11 the correlation is still around 0.5 at 50–100 h lag time. Honti et al. (2013) and Del Giudice et al. (2013) presented the direct encoding of a bias description process within the error model. This was applied to urban drainage hydrodynamic simulation with time-steps of 1–2 min, which present strong autocorrelation structures. The bias description can be conceptualised as function of different variables or inputs, yet in its basic form, it constitutes a Gaussian multivariate distribution with an exponential covariance structure as in Eq. 9.

The use of an AR(3) model, in this case, rendered an almost negligible effect of the autoregressive parameters of higher order than one ( $< 0.01$ ), thus generating an equivalent AR(1) model. Measured series vs. inferred comparison and the residual structure can be seen in Fig. 15 (Annex B). Although the autocorrelation of residuals is better represented, the fit of the mean sample did not improve, rather was degraded through accounting for the autocorrelation term. This was also discussed by Evin et al.

(2013), who showed that using AR(1) models for hydrological inference on the raw residuals can lead to strong interactions with the inferred parameters and degraded outcomes.

On the other hand, the use of a bias description as in Del Giudice et al. (2013) becomes prohibitive for long time series. This implementation requires the inversion of a covariance matrix  $\Sigma \in \mathbb{R}^{L \times L}$  being  $L$  the size of the measurement layout. In this case, considering an hourly sampling layout during 15-Jan-2012 until 05-Aug-2012 leads to  $L = 4892$  elements. Expected values of the decay parameter  $\tau$  are likely to produce a highly sparse covariance matrix, thus sparse inversion optimisation could be applied (Betancourt and Alvarado 1986) yet intensive sampling for populating the posterior is still cumbersome. This large covariance matrix inversion renders the evaluation of the likelihood function computationally expensive, thus eliminating the benefits of the use of the dynamic emulator. A possible solution is to create a database of likelihood samples from the dynamic emulator which is used to build a second emulator linking the parameter space and the response of the likelihood function (Dietzel and Reichert 2014). An illustrative example of the residuals when using a bias description term is shown in Fig. 16 (Annex B) in which the inference was performed in a shorter time series (26-Jul-2012–14-Sep-2012). The description of the autocorrelation structure in the residuals did not allow for a better description of the process, or a better understanding of the parametric uncertainty. Ammann et al. (2018) recently studied the representation of autocorrelated likelihood structures with the conventional error models for hydrological applications. They discussed that the use of stationary autocorrelation models deteriorates the performance of the inferred model significantly (degrading even further when increasing the measurement layout frequency). They propose that the use of non-stationary autocorrelation schemes may overcome this problem,

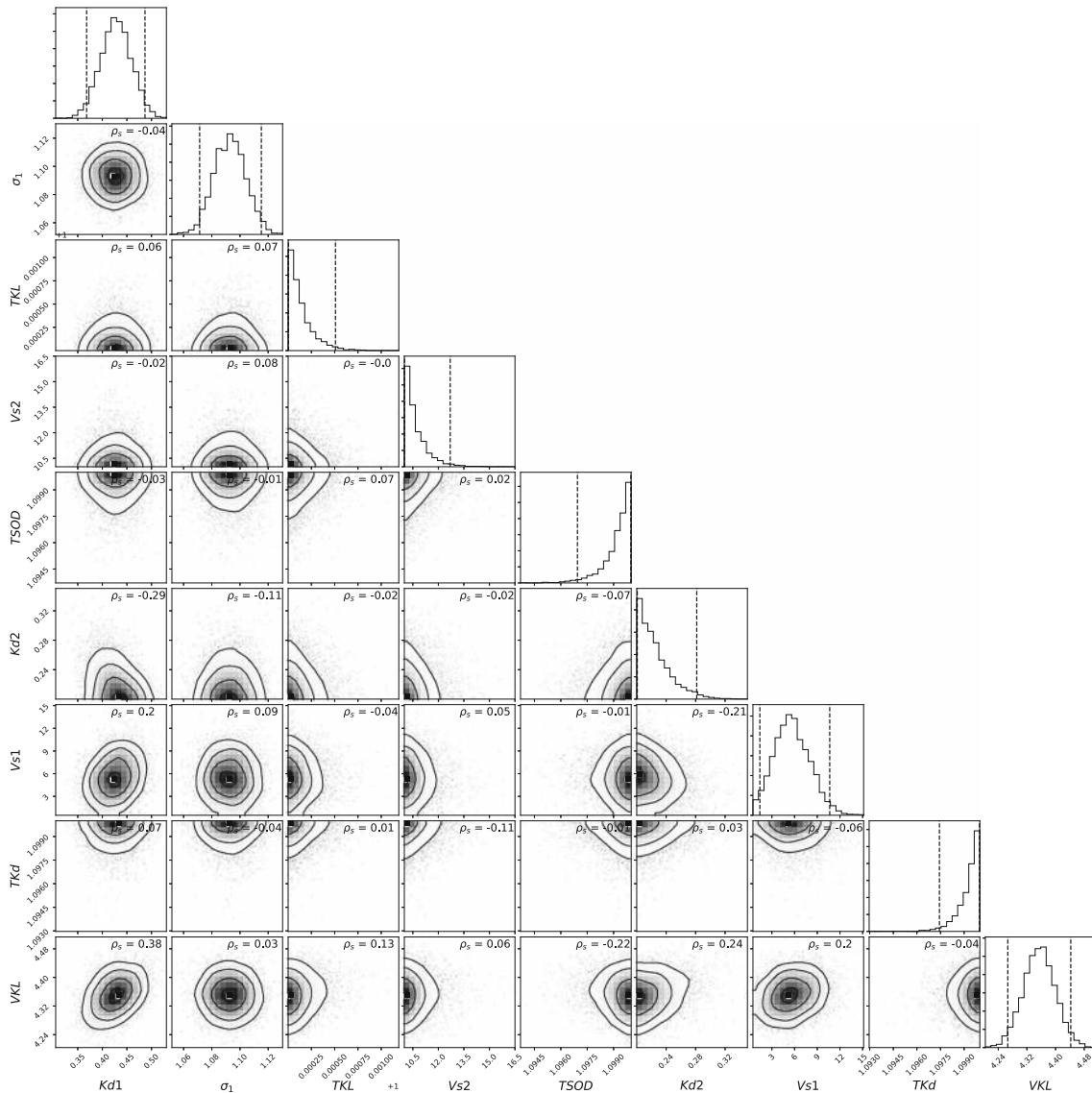


**Fig. 8** Posterior joint parametric distribution for the inference of the flow model parameters.  $\sigma_1$  and  $\sigma_2$  are hyperparameters of the selected error generation process (heteroscedastic, independent Gaussian). The spearman's correlation coefficient ( $\rho_s$ ) is shown at each parameter couple

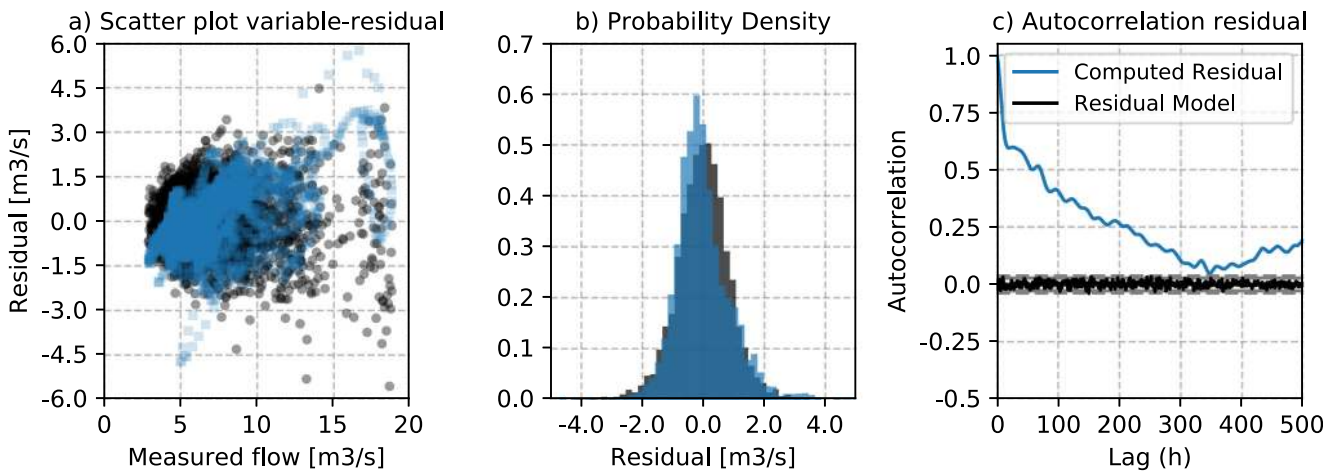
since hydrological models are expected to lose memory under storm events (thus dry-weather and wet-weather present different residual correlation patterns). This non-stationarity could not be found in this case, being the correlation structure in dry weather and wet weather for short scales (0–80 h lag) equivalent and the large-lag structures only slightly different (Fig. 12). Also, this is not expected to be applicable for DO series, where residuals are even less structured as in hydrological flow.

A strong autocorrelation structure is expected due to the nature of the process and the measurement layout. Both flow and dissolved oxygen concentrations present several dynamic modes, induced by storm events, daily fluctuation in the WWTP effluent and variation between dry-wet periods and temperature seasons. Small temporal shifts are expected due to model structural misfit (e.g. incorrect CSO timing in the urban drainage scheme or misrepresentation errors in rainfall data).

The temporal shift will likely render strongly correlated residuals in time. Yet these time-shifts are of limited influence for the model application. The objective of the model is to represent dynamics of oxygen in a receiving water body for environmental policy assessment studies. These studies use metrics which lumps the time-dynamics, as frequency-duration-concentration tables (FWR 2012); thus, the exact timing of the oxygen depletion is not highly relevant, but rather the correct representation of the magnitude and duration of each event. Consequently, the stiff likelihood conditions required to construct formal inference schemes (as shown in this study) might not render the most adequate approach when dealing with the with long-term dissolved oxygen dynamic series (in which the system exhibit multitude of complex dynamic states). Approximated Bayesian computation (Toni et al. 2009) could be of interest by allowing defining metrics which attend to the

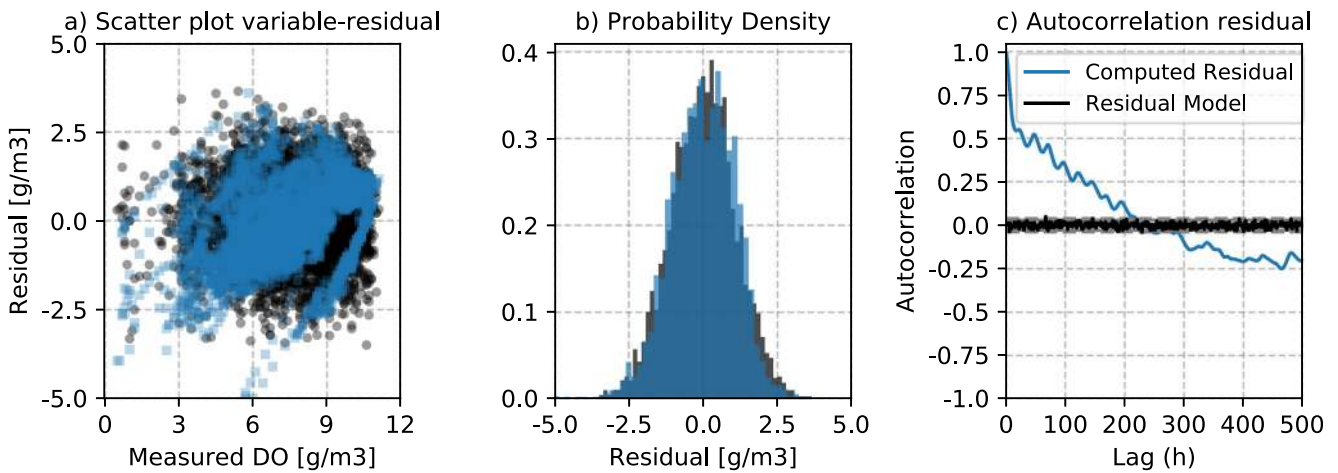


**Fig. 9** Posterior joint-parametric distribution for the inference of the water quality model parameters.  $\sigma_1$  is the hyperparameter of the selected error generation process (independent, identically distributed Gaussian). The spearman's correlation coefficient ( $\rho_s$ ) is shown at each parameter couple



**Fig. 10** Residual structure at the flow posterior mean sample. **a** Scatter plot variable-residual showing the dependency of the variance. **b** The residual probability density. **c** The autocorrelation plot at different time-lags





**Fig. 11** Residual structure at the dissolved oxygen posterior mean sample. **a** Scatter plot variable residual showing the dependency of the variance. **b** The residual probability density. **c** The autocorrelation plot at different time-lags

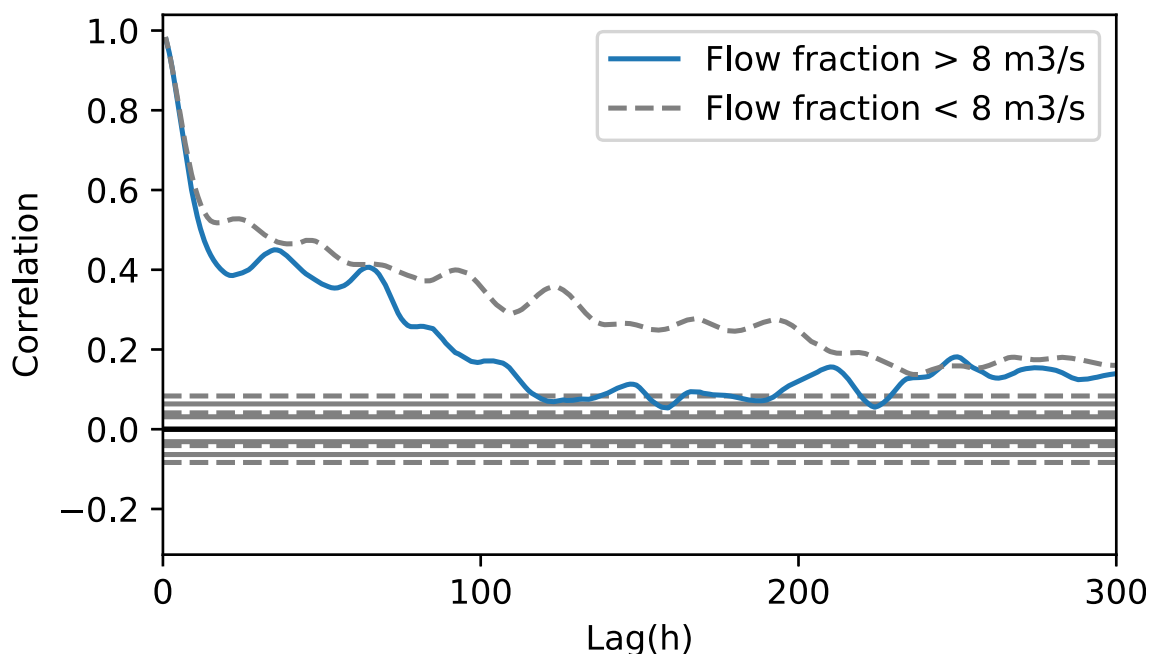
relevant features of the dissolved oxygen output space (e.g. duration and magnitude of events, slopes of the depletion/recovery patterns, etc.). The selection of signature-based metrics are being increasingly used for the diagnosis of hydrological modelling studies (Kavetski et al. 2018; Vrugt and Sadegh 2013) and could also facilitate the identification and calibration of urban water quality dynamics.

## Conclusions

This study presents the emulation and inference of river flow and dissolved oxygen dynamics in an integrated urban water quality system. The simulator jointly evaluates wastewater treatment

processes, urban drainage and receiving water quality processes. The use of an emulation scheme allowed accelerating significantly the approximation of the response of the simulator to variation of a set of global parameters. This facilitated the implementation of sampling intensive applications (e.g. sensitivity analysis and formal Bayesian inference schemes).

A polynomial orthogonal expansion emulator was fitted to represent flow and dissolved oxygen depletion for a 1-year-long time series (hourly frequency) under four and eight global parameters respectively. Two hundred model realizations sufficed to generate an acceptable interpolation in both cases. The emulator was validated using independent data, rendering a high-quality mapping between the parametric space and the dynamic response. This technique still exhibits severe limitations, like the impossibility to



**Fig. 12** Autocorrelation structure for flow residuals by magnitude

include large parametric spaces, dynamic inputs or non-smooth parametric to output relationships (e.g. bifurcation solution points).

The use of the emulator facilitated the computation of the sensibility of flow and DO dynamics to river parameters. Oxygen depletion processes exhibit a non-stationary dependency across storm events. In general, the reaeration rate showed to be the most relevant parameter during dry weather flow. Depletion of fast-slow biodegradable matter is often the responsible for the magnitude of the oxygen depletion event (attending only to the river process parameters). Meanwhile, oxygen recovery after a large depletion event is highly influenced by the sediment oxygen demand and reaeration processes, with a strong dependence in seasonality (temperature driven).

A set of observations of river flow and DO was used to update prior knowledge about the receiving water model parameters. This was achieved by performing an inference scheme on the emulator as a substitute of the simulator. Several hypotheses were used to define the likelihood structure. A homoscedastic, independent Gaussian distributed error was applied to the dissolved oxygen error process. Meanwhile, the flow residuals showed a heteroscedastic structure. Both inferred residual series rendered a highly temporally correlated structure, which violates the assumption of independence. The residual autocorrelation is related to the measurement layout frequency (hourly) and the nature of the simulated processes. Both flow and DO residuals are influenced by a strong memory effect, model structure-induced time shifts and input errors. Various formulations to deal with the residual autocorrelation or structural bias were tested. However, the inferred dynamics either deteriorated or did not improve. Detailed investigation on the effects of neglecting the correlation structure in the dissolved oxygen residual structure is still missing. Also, the use of alternative metrics for the inference of dissolved oxygen dynamics should be further studied.

The use of a dynamic emulation scheme allowed gaining insights on the underlying mechanistic relationships of the integrated urban water quality system. This can be easily extended to similar environmental modelling studies thus facilitating the application of sensitivity analysis, inference or calibration under long time series and low-dimensionality parametric spaces.

**Acknowledgements** This research was done as part of the Marie Curie ITN—Quantifying Uncertainty in Integrated Catchment Studies (QUICS) project. The authors would like to thank MIKE DHI for a free-of-charge academic licence agreement for the use of their software WEST. The authors would like to thank the financial contribution of the Dutch ‘ Kennisprogramma Urban Drainage ’ (Knowledge Programme Urban Drainage). The involved parties are ARCADIS, Deltares, Evides, Gemeente Almere, Gemeente Arnhem, Gemeente Breda, Gemeente’s-Gravenhage, Gemeentewerken Rotterdam, Gemeente Utrecht, GMB Rioleringsstechniek, Grontmij, KWR Watercycle Research Institute, Royal HaskoningDHV, Stichting RIONED, STOWA, Tauw, vanderValk + deGroot, Waterboard De Dommel, Waternet and Witteveen+Bos.

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

### Annex A: River model water quality processes

Table 6 presents the process matrix for the dissolved oxygen process routine at the river section. The model accounts for the transport and conversion rates of seven concentration state variables; BOD1 (biological oxygen demand of fast biodegradable suspended fraction), BOD2 (biological oxygen demand of slow biodegradable suspended fraction), BOD1p (biological oxygen demand of fast biodegradable particulate fraction), BOD2p (biological oxygen demand of slow biodegradable particulate fraction), BODs (biological oxygen demand at the sediment section), NH4 (ammonium) and DO (dissolved oxygen concentration).

**Table 6** Process matrix for the river water quality model structure

State variable	BOD1	BOD1p	BOD2	BOD2p	BODs	NH4	DO	Rate
	gO <sub>2</sub> /m <sup>3</sup>	gO <sub>2</sub> /m <sup>3</sup>	gO <sub>2</sub> /m <sup>3</sup>	gO <sub>2</sub> /m <sup>3</sup>	gO <sub>2</sub> /m <sup>3</sup>	gN/m <sup>3</sup>	gO <sub>2</sub> /m <sup>3</sup>	
Process								
1a. Oxidation of fast-suspended fraction (BOD1)	-1						-1	$TKd^T_{wat-20} \cdot Kd1 \cdot BOD1 \cdot \frac{DO}{KO2+DO}$
1b. Oxidation of fast-particulate fraction (BOD1p)		-1					-1	$TKd^T_{wat-20} \cdot Kd1 \cdot BOD1p \cdot \frac{DO}{KO2+DO}$
2a. Oxidation of slow-suspended fraction (BOD2)			-1				-1	$TKd^T_{wat-20} \cdot Kd2 \cdot BOD2 \cdot \frac{DO}{KO2+DO}$
2b. Oxidation of slow-particulate fraction (BOD2p)				-1			-1	$TKd^T_{wat-20} \cdot Kd2 \cdot BOD2p \cdot \frac{DO}{KO2+DO}$
3a. Sedimentation of BOD1p		-1			+1			$Vs1 \cdot BOD1p$
3b. Sedimentation of BOD2p				-1	+1			$Vs2 \cdot BOD2p$
4. Oxidation of organic matter in the sediment					-1		-1	$TSOD^T_{wat-20} \cdot K_{BOD} \cdot \frac{BODs}{d} \cdot \frac{DO}{KSO+DO}$
5. Constant sediment oxygen demand							-1	$TSOD^T_{wat-20} \cdot \frac{SOD}{d} \cdot \frac{DO}{KSO+DO}$
6. Nitrification						-1	-4.57	$TKnit^T_{wat-20} \cdot Knit \cdot NH4 \cdot \frac{DO}{KNO2+DO}$
7. Photosynthesis macrophyte							+1	$TKp^T_{wat-20} \cdot k_{prodM} \cdot Io \cdot \frac{MB}{d}$
8. Macrophyte oxygen consumption							-1	$TKp^T_{wat-20} \cdot k_{pcons} \cdot \frac{MB}{d}$
9. Reaeration							+1	$TKL^T_{wat-20} \cdot VKL \cdot (CS-DO)$

The river flow was conceptualized as a tank in series lumped scheme, in which each tank is a well-stirred reactor in which the biochemical processes take place (Fig. 13). The mass balance equation is computed at each node:

$$\frac{dV_i}{dt} = Q_{in_i} - Q_{out_i}, \quad (1)$$

where  $V_i$  is the volume at tank  $i$  and  $Q_{in_i}$  the sum of inflows from CSOs, upstream sections or rural hydrology.  $Q_{out_i}$  is the outflow from the tank at each instance

and is computed as a free-surface flow following the Gauckler-Manning equation:

$$Q_{out_i} = A \cdot R_H^{\frac{2}{3}} \cdot slp^{\frac{1}{2}} \cdot n^{-1}, \quad (2)$$

being  $A$  and  $R_H$  the area and hydraulic radius for a trapezoidal section with a given slope ( $slp$ ) and bed Manning's roughness ( $n$ ).

Figure 14 depicts the whole integrated system scheme which comprises the urban drainage networks connected to the area of interest in the Dommel, the WWTP of Eindhoven and the river stretch considered in the water quality model along with the parameter and input model spaces.

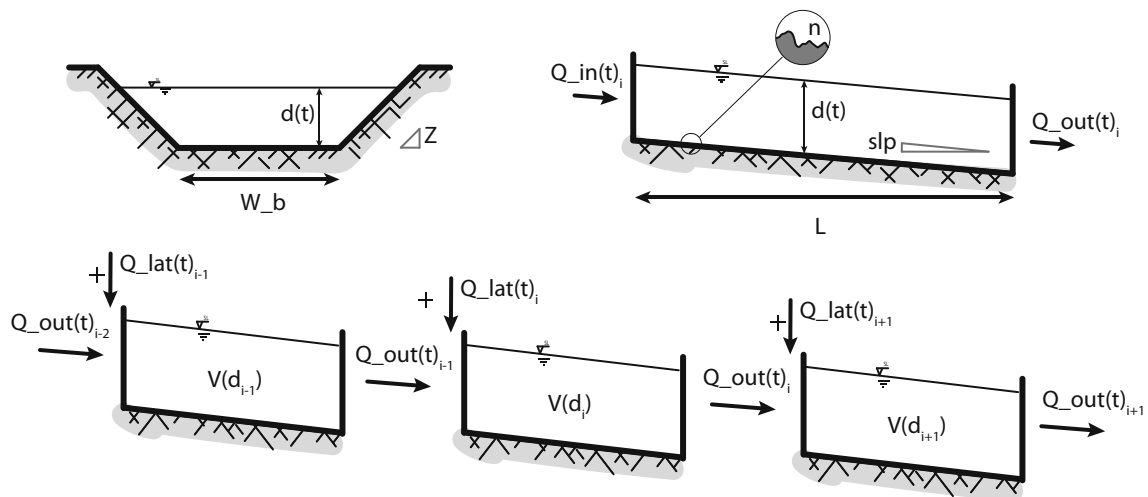


Fig. 13 Tank in series river flow scheme

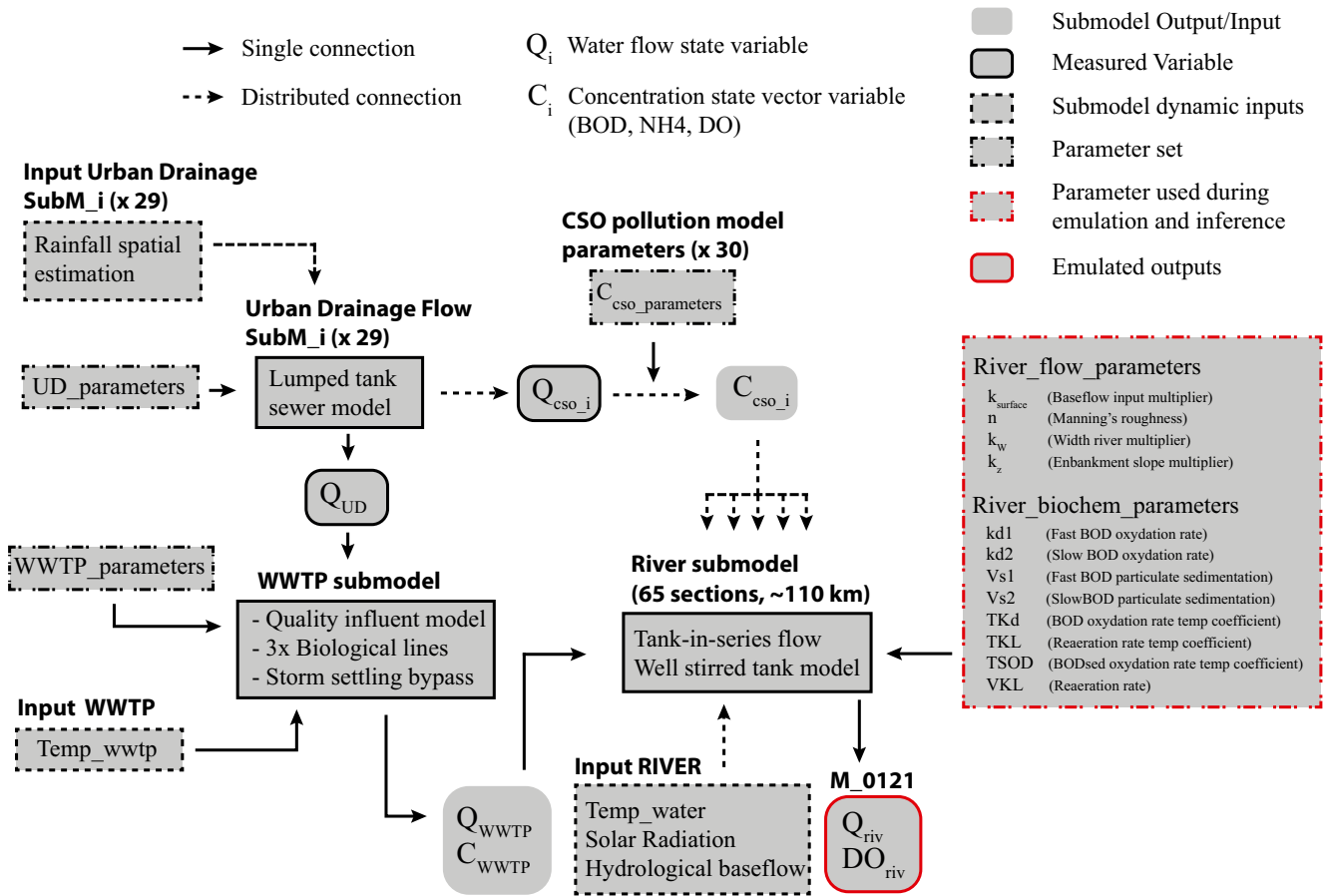
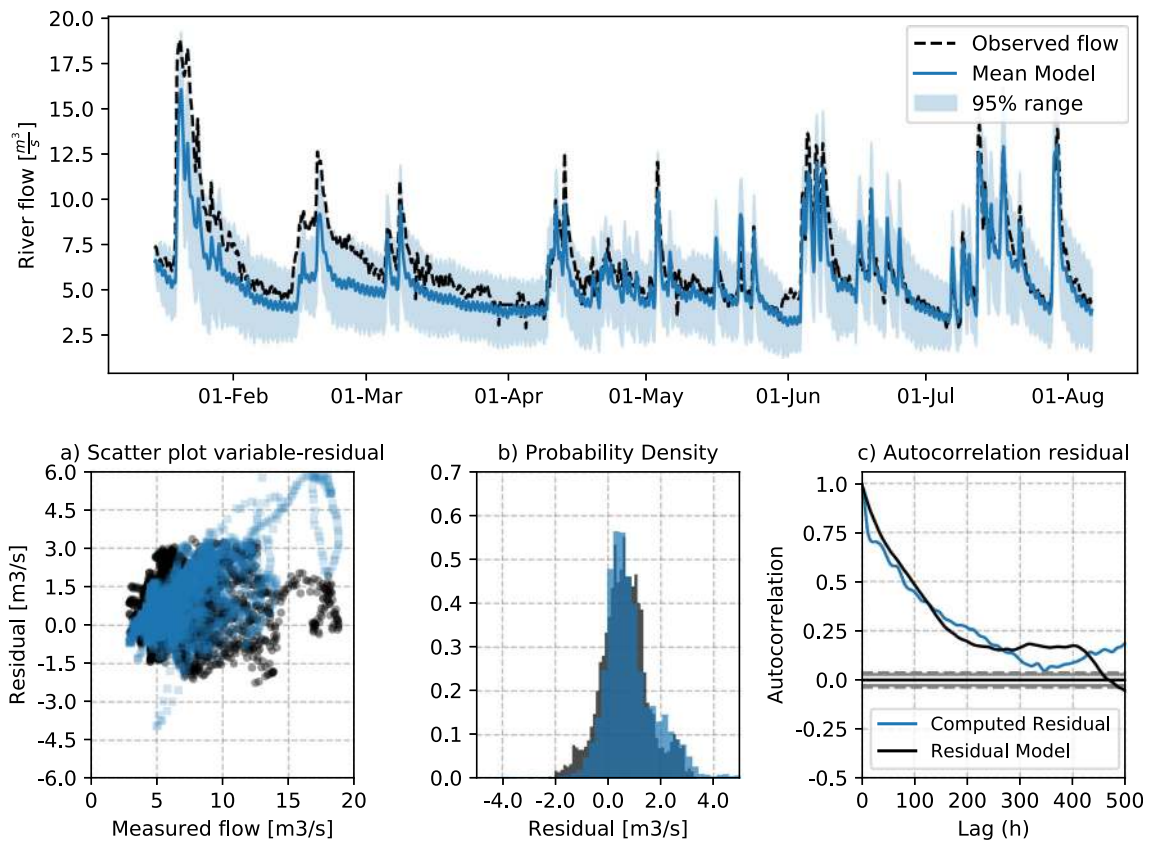


Fig. 14 Model scheme

## Annex B: Accounting for the autocorrelation structure



**Fig. 15** Autoregressive model order 3, comparison of measured and inferred dynamics and residual structure



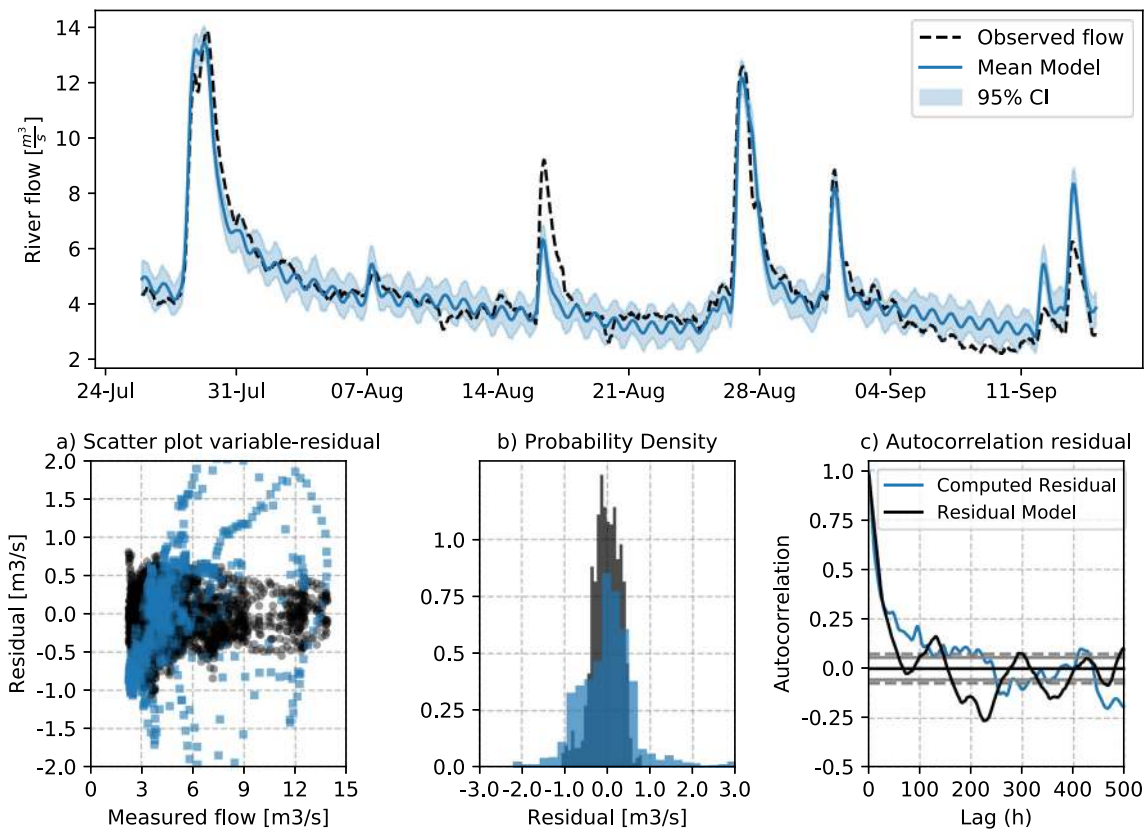


Fig. 16 Bias description, comparison of measured and inferred dynamics and residual structure

## Annex C: Test output series for the emulator-simulator comparison

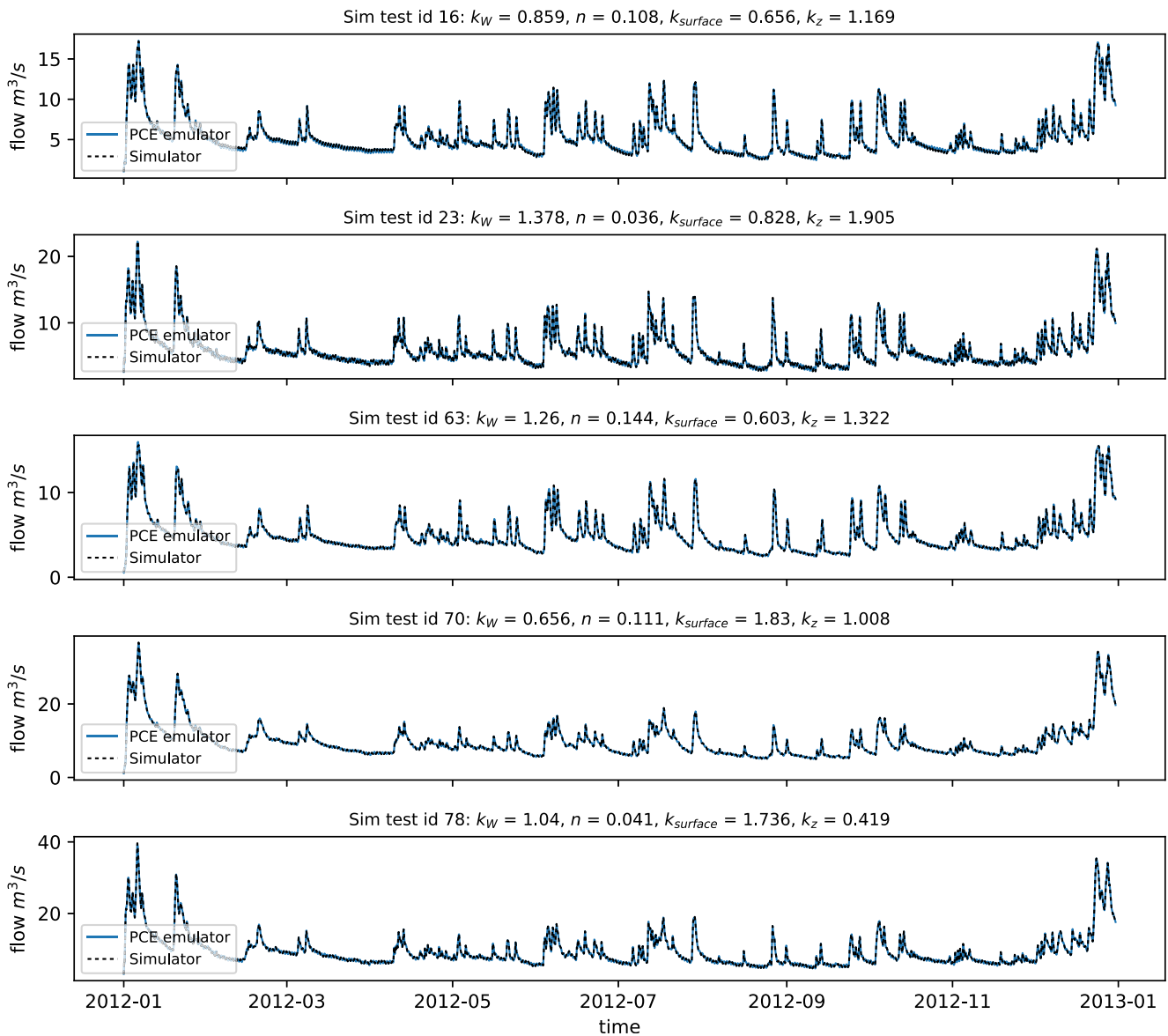
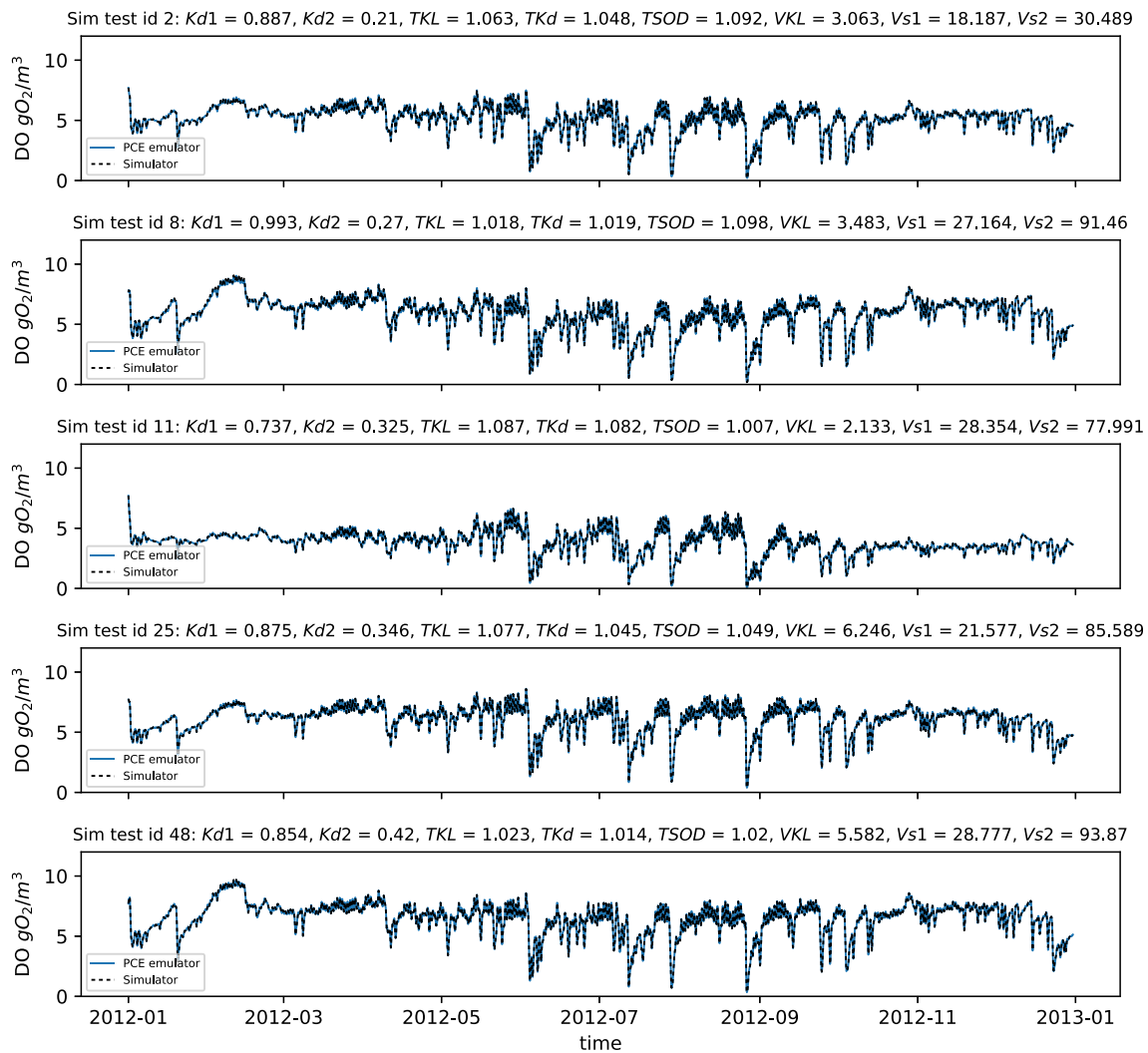


Fig. 17 Emulator vs. simulator flow time series graphical comparison for different test parameter combinations



**Fig. 18** Emulator vs. simulator dissolved oxygen time series graphical comparison for different test parameter combinations

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ammann L, Reichert P, Fenicia F (2018) A framework for likelihood functions of deterministic hydrological models. *Hydrol Earth Syst Sci Discuss* 2018:1–39
- Bates BC, Campbell EP (2001) A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resour Res* 37(4):937–947
- Benedetti L, Langeveld J, Comeau A, Corominas L, Daigger G, Martin C, Mikkelsen PS, Vezzaro L, Weijers S, Vanrolleghem PA (2013) Modelling and monitoring of integrated urban wastewater systems: review on status and perspectives. *Water Sci Technol* 68(6):1203–1215
- Betancourt R, Alvarado FL (1986) Parallel inversion of sparse matrices. *IEEE Trans Power Syst* 1(1):74–81
- Carbajal JP, Leitão JP, Albert C, Rieckermann J (2017) Appraisal of data-driven and mechanistic emulators of nonlinear simulators: the case of hydrodynamic urban drainage models. *Environ Model Softw* 92: 17–27
- Del Giudice D, Honti M, Scheidegger A, Albert C, Reichert P, Rieckermann J (2013) Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrol Earth Syst Sci* 17(10):4209–4225
- Deletic A, Dotto CBS, McCarthy DT, Kleidorfer M, Freni G, Mannina G, Uhl M, Henrichs M, Fletcher TD, Rauch W, Bertrand-Krajewski JL, Tait S (2012) Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C* 42:44:3–10
- Dietzel A, Reichert P (2014) Bayesian inference of a lake water quality model by emulating its posterior density. *Water Resour Res* 50(10): 7626–7647
- Dotto CB, Mannina G, Kleidorfer M, Vezzaro L, Henrichs M, McCarthy DT, Freni G, Rauch W, Deletic A (2012) Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Res* 46(8):2545–2558
- Evin G, Kavetski D, Thyer M, Kuczera G (2013) Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resour Res* 49(7):4518–4524

- Feinberg J (2015) Some improvements and applications of non-intrusive polynomial chaos expansions. PhD Thesis, University of Oslo
- FWR (2012) Urban pollution management manual. Foundation for Water Research, Marlow
- Gautschi W (1994) Algorithm 726: ORTHOPOL- a package of routines for generating orthogonal polynomials and gauss-type quadrature rules. *ACM Trans Math Softw* 20(1):21–62
- Gernaey KV, van Loosdrecht MC, Henze M, Lind M, Jørgensen SB (2004) Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environ Model Softw* 19(9):763–783
- Goodman J, Weare J (2010) Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* 5(1):65–80
- Hadigol M, Doostan A (2018) Least squares polynomial chaos expansion: a review of sampling strategies. *Comput Methods Appl Mech Eng* 332:382–407
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109
- Hoffman MD, Gelman A (2014) The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15(1):1593–1623
- Honti M, Stamm C, Reichert P (2013) Integrated uncertainty assessment of discharge predictions with a statistical error model. *Water Resour Res* 49(8):4866–4884
- Kavetski D, Fenicia F, Reichert P, Albert C (2018) Signature-domain calibration of hydrological models using approximate Bayesian computation: theory and comparison to existing applications. *Water Resour Res* 54(6):4059–4083
- Kuczera G (1983) Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resour Res* 19(5):1151–1162
- Laloy E, Vrugt JA (2012) High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. *Water Resour Res* 48(1). <https://doi.org/10.1029/2011WR010608>
- Laloy E, Rogiers B, Vrugt JA, Mallants D, Jacques D (2013) Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resour Res* 49(5):2664–2682
- Langeveld J, Nopens I, Schilperoord R, Benedetti L, de Klein J, Amerlinck Y, Weijers S (2013a) On data requirements for calibration of integrated models for urban water systems. *Water Sci Technol* 68(3):728–736
- Langeveld JG, Benedetti L, de Klein JJM, Nopens I, Amerlinck Y, van Nieuwenhuijzen A, Flameling T, van Zanten O, Weijers S (2013b) Impact-based integrated real-time control for improvement of the Dommel River water quality. *Urban Water J* 10(5):312–329
- Langeveld J, Van Daal P, Schilperoord R, Nopens I, Flameling T, Weijers S (2017) Empirical sewer water quality model for generating influent data for WWTP modelling. *Water* 9(7):491
- Mahmoodian M, Carbajal JP, Bellos V, Leopold U, Schutz G, Clemens F (2018) A hybrid surrogate modelling strategy for simplification of detailed urban drainage simulators. *Water Resour Manag* 32(15):5241–5256
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092
- Moreno-Rodenas AM, Cecinati F, Langeveld J, Clemens FHLR (2017a) Impact of spatiotemporal characteristics of rainfall inputs on integrated catchment dissolved oxygen simulations. *Water* 9(12):926
- Moreno-Rodenas AM, Langeveld J, Clemens FHLR (2017b) Accounting for correlation in uncertainty propagation. a copula approach for water quality modelling. Prage, Czech Republic
- Moreno-Rodenas AM, Bellos V, Langeveld JG, Clemens F (2018) A dynamic emulator for physically based flow simulators under varying rainfall and parametric conditions. *Water Res* 142:512–527
- Moreno-Rodenas AM, Tscheikner-Gratl F, Langeveld JG, Clemens FHLR (2019) Uncertainty analysis in a large-scale water quality integrated catchment modelling study. *Water Res* 158:46–60
- Muschalla D, Schutze M, Schroeder K, Bach M, Blumensaat F, Gruber G, Klepizewski K, Pabst M, Pressl A, Schindler N, Solvi AM, Wiese J (2009) The HSG procedure for modelling integrated urban wastewater systems. *Water Sci Technol* 60(8):2065–2075
- Patil A, Huard D, Fonnesebeck CJ (2010) PyMC: Bayesian stochastic modelling in Python. *J Stat Softw* 35(4):1–81
- Rauch W, Bertrand-Krajewski J, Krebs P, Mark O, Schilling W, Schutze M, Vanrolleghem P (2002) Deterministic modelling of integrated urban drainage systems. *Water Sci Technol* 45(3):81–94
- Schoups G, Vrugt JA (2010) A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour Res* 46(10). <https://doi.org/10.1029/2009WR008933>
- Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments* 1(4):407–414
- Solvi A-M (2006) Modelling the sewer-treatment-urban river system in view of the EU Water Framework Directive. Doctoral Thesis, Ghent University, Belgium
- Sorooshian S, Dracup JA (1980) Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resour Res* 16(2):430–442
- Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6(31):187–202
- Tscheikner-Gratl F, Lepot M, Moreno-Rodenas A, Schellart A (2017) A frame-work for the application of uncertainty analysis. <https://doi.org/10.5281/zenodo.1240926>
- Tscheikner-Gratl F, Bellos V, Schellart A, Moreno-Rodenas A, Muthusamy M, Langeveld J, Clemens F, Benedetti L, Rico-Ramirez MA, de Carvalho RF, Breuer L, Shucksmith J, Heuvelink GBM, Tait S (2019) Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Res* 150:368–379
- Vrugt JA, Sadeh M (2013) Toward diagnostic model calibration and evaluation: approximate Bayesian computation. *Water Resour Res* 49(7):4335–4345
- Xiu D (2010) Numerical methods for stochastic computations: a spectral method approach. Princeton university press
- Xiu D, Karniadakis GE (2002) The wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24(2):619–644
- Yang J, Jakeman A, Fang G, Chen X (2018) Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian process emulator. *Environ Model Softw* 101:289–300

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.