Performance assessment of a commonly used "accumulation and wash-off" model from long-term continuous road runoff turbidity measurements

Jérémie Sage, Céline Bonhomme, Saja Al Ali, Marie-Christine Gromaire

PERFORMANCE ASSESSMENT OF A COMMONLY USED "ACCUMULATION AND WASH-OFF" MODEL FROM LONG-TERM CONTINUOUS ROAD RUNOFF TURBIDITY MEASUREMENTS

SAGE Jérémie[1*], BONHOMME Céline[1], AL ALI Saja[1], GROMAIRE Marie-Christine[1]

[1]*Université Paris Est, LEESU, UMR MA 102 - Agro ParisTech, Ecole des Ponts ParisTech, Cité Descartes - 6-8 av. Blaise Pascal, 77455, Champs-sur-Marne Cedex 2, France*

[*]*Corresponding author: jeremie.sage@leesu.enpc.fr, + 00 33 1 64 15 39 13*

ABSTRACT: The suitability of a commonly used accumulation and wash-off model for continuous modelling of urban runoff contamination was evaluated based on 11-month turbidity and flow-rate records from an urban street. Calibration and uncertainty analysis were performed using a Markov Chain Monte-Carlo sampling method for both suspended solids loads (discharge rates) and concentration modelling. Selected models failed at replicating suspended solids concentration over the complete monitoring period. The studied dataset indeed suggests that the accumulation process is rather unpredictable and cannot be satisfactorily represented with usual accumulation models unless short periods are considered. Regarding suspended solid loads modelling, noticeably better performance was achieved, but similar results could as well be obtained with much simpler constant concentration models. Unless providing very accurate estimates of concentrations in runoff, accounting for their temporal variability during rain events may therefore not always be necessary for pollutant loads modelling, as loads are in fact mostly explained by runoff volumes.

KEYWORDS: Calibration, Concentrations, Loads, Pollutants, Suspended solids, Uncertainties

1. INTRODUCTION

Since the early eighties, several research programs have identified urban runoff as a major source of diffuse contamination and evidenced the need for better stormwater pollution control (Saget, 1994; US-EPA, 1983). Today, many local communities have already undertaken mitigation efforts to minimize the adverse impacts of stormwater discharge on the environment. More specifically, Low Impact Development (LID), advocating for on-site runoff and pollution control, has become increasingly popular (Ahiablame et al., 2012; Dietz, 2007). In

this context, simulation of temporal variations of pollutant concentrations in runoff originating from urban surfaces such as streets during rainfall (and from a storm to another) is of great interest to both researcher and practitioners for various applications related to the development of relevant stormwater management strategies for diffuse pollution control.

Various models have therefore been proposed in the past, among which, "accumulation and wash-off" models, implemented in most software solutions (Aryal et al., 2009; Obropta and Kardos, 2007) are often relied on to replicate time series of concentrations. While such models have been shown to perform relatively well (Crobeddu and Bennis, 2011; Piro and Carbone, 2010; Wang et al., 2011), investigation of the temporal variability of concentrations in runoff has, however, for a long time been restricted by experimental constraints related to sampling methods and thus relied on relatively scarce observations, with limited number of rain-events and very partial information on pollutant wash-off dynamics (Métadier and Bertrand-Krajewski, 2012). As a consequence, these models have generally not been verified against long-term continuous water quality measurements and several recent findings suggest that their ability to simulate temporal variability of pollutant concentrations in runoff might have been overestimated (Dotto et al., 2011; Freni et al., 2009; Kanso et al., 2005; Shaw et al., 2010).

The use of long-term continuous water quality measurements, which have only recently been made available, provides significant opportunities for in depth investigation of the processes associated with stormwater contamination (Deletic, 1998; Hannouche et al., 2014; Joannis et al., 2014; Métadier and Bertrand-Krajewski, 2012). The main purpose of this study is therefore to discuss and clarify the capacity of conventional water quality models to simulate sediment wash-off dynamics based on continuous monitoring of flow-rate and turbidity from an urban street over an 11-month period. While models' ability to replicate both suspended solids loads (e.g. discharge rates) and concentrations will be investigated, application of a "Markov Chain Monte-Carlo" (MCMC) method for calibration will furthermore enable quantification of uncertainties associated with parameters values so as to better identify potential limitations of usual accumulation and wash-off formulations.

2. MATERIAL AND METHODS

2.1. EXPERIMENTAL SETTINGS

The experimental site is located in "Sucy-en-Brie" municipality, a residential district within Paris conurbation. The studied catchment consists in a 800m² portion (½ roadway width + sidewalk) of an urban road carrying

moderate traffic loads (~8 000 vehicles per day), with a runoff length of 160m and an average slope of 0.8%. Runoff was collected by a storm drain where the monitoring equipment was installed. 11-litres tipping buckets were used for flow-rate measurement corresponding to a 0.014mm resolution in runoff height over drainage area.

Runoff quality was monitored with a YSI 6820VZ multi-parameter probe. In order to save storage capacity and reduce power consumption, the probe was driven by the flow-metre and measurements were performed at 1-min time step during runoff periods only (from the first bucket tipping of a rain event up to 30min after the last tipping). Turbidity measurements were here considered as a surrogate for runoff contamination. In order to facilitate the interpretation of results and comparison with other studies, turbidity data was however converted into Total Suspended Solids concentrations (TSS) from a linear TSS-turbidity relationship adjusted from event mean runoff samples performed for 7 rain events (R^2=0.96). Implication of the accuracy of this relation on the TSS values calculated or modelled will not be discussed here (further details on TSS-turbidity relationship may be found in Bertrand-Krajewski, 2004; Hannouche et al., 2011; the impact of input data uncertainty is discussed in Kleidorfer et al., 2009). A rain gage located nearby (500m from studied site), additionally provided rainfall measurements over the entire monitoring period.

Data were collected from September 2012 to December 2013. Technical maintenance was performed every two weeks to remove litter from the storm drain and to verify the measurement system (turbidity probe cleaning and tipping bucket system control). Despite regular inspection of the experimental system, several mechanical problems were encountered with the tipping bucket system during the monitoring period, resulting in absence of record over several weeks. Snow periods were as well excluded from the dataset (as selected water quality models do not apply for snowmelt). Overall, 175 rain events from January 2013 to November 2013 (considering a 30 minutes minimum inter-event time for their identification) were fully monitored.

As indicated in figure 1, a sudden increase in turbidity values, followed by a slower return to previous turbidity levels, was observed at the beginning of the monitoring period (after mid-January). This trend presumably does not result from a failure of the multi-parameter probe, for which calibration was verified three times during the experiment (from standard formazin solutions) and which never showed any drift in the measurements.

2.2. EXPONENTIAL ACCUMULATION AND WASH-OFF MODELS

The models selected for this study are based on SWMM “exponential” accumulation and wash-off formulations (Huber and Dickinson, 1988). Although widely adopted, these models have often been reported to fail to

replicate the variability of concentrations in runoff (Bai and Li, 2013; Shaw et al., 2010) and several studies cast doubt on their relevance for loads and concentrations modelling as compared to simpler formulations (Freni et al., 2009; Joannis et al., 2014; Kanso et al., 2006; Vezzaro, 2008). Simpler modelling approaches were thus also considered in this study so as to evaluate the benefits of accumulation and wash-off equations.

"Event mean concentration" (or EMC) models assume that concentrations in runoff remain invariant during a rain event. While such approaches do not allow for simulation of wash-off dynamics, recent studies on combined sewers suggest that it could be relevant for loads estimation (Joannis et al., 2014). Two "Event Mean Concentration" models were thus adopted for loads modelling, with, on the one hand, a constant concentration over the whole simulation period (referred to as constant EMC hypothesis, Eq. 1), and considering, on the other hand, an exponential EMC decrease from January to November in accordance with turbidity measurements (referred to as decreasing EMC hypothesis, Eq. 2). In both cases, concentrations are hence assumed to remain constant during a rain event i and can be computed from:

$$[TSS]_i = C_{CST} \quad (1)$$

$$[TSS]_i = (C_{INI} - C_{LIM}) \, Exp\,(-K \times T_i) + C_{LIM} \quad (2)$$

Where: $[TSS]_i$ = Event Mean Concentration (EMC) for suspended solids (mg/l); C_{CST}= constant EMC value (mg/l), C_{INI}= EMC at the beginning of simulation (mg/l); C_{LIM}= EMC at the end of simulation (mg/l); K = model parameter (d^{-1}); T_i = beginning date of rain event i (d).

Contrary to EMC models, the exponential wash-off equation allows for a description of both the inter- and intra-event variability of pollutant concentrations. Based on experimental measurements by Sartor and Boyd (1974), pollutant removal during rain event has traditionally been described as a "source-limited" process (Tsihrintzis and Hamid, 2001). Exponential wash-off functions thus consider that amount of sediment washed-off at time t directly depends on the mass available over road surface. While initial equations assumed that removal rate only depends on cumulative runoff volume, modified exponential wash-off models have been implemented in SWMM to account for non-linear dependency on runoff rate (Shaw et al., 2010). The general exponential equation can be written as follow:

$$\Phi(t) = M_{ACC}\,(t) \times C_1.q(t)^{\,C2} \quad (3)$$

$$M_{ACC}(t+dt) = M_{ACC}(t) - \Phi(t) \times dt \quad (4)$$

Where: $\Phi(t)$ = pollutant discharge rate washed off at t during time step dt ($g.m^{-2}.s^{-1}$); $M_{ACC}(t)$ = available mass at time t on road surface ($g.m^{-2}$); q(t) flow-rate over street surface ($mm.h^{-1}$); C1 = wash-off coefficient and C2 = wash-off exponent.

From previous equation, determination of the initial sediment storage $M_{ACC}(t_{0,i})$ ($g.m^{-2}$) at the beginning of a rain event i (e.g. $t = t_{0,i}$) is needed to compute suspended fluxes or concentrations. Pollutant accumulation is generally assumed to result from the equilibrium between pollutant deposition and removal due to traffic or wind erosion during dry periods (Alley and Smith, 1981). $M_{ACC}(t_{0,i})$ ($g.m^{-2}$) can therefore be computed as follow:

$$M_{ACC}(t_{0,i}) = D_{ACC}/D_{ERO} \times [1 - Exp(-D_{ERO} \times T_{DRY,i})] + M_{RES.} Exp(-D_{ERO} \times T_{DRY,i}) \quad (5)$$

Where: M_{RES} = residual pollutant loads remaining on street surface at the end of previous rain event ($g.m^{-2}$); $T_{DRY,i}$ = antecedent dry period duration associated with rain event i (in days); D_{ACC} = pollutant accumulation rate ($g.m^{-2}.d^{-1}$) and D_{ERO} = pollutant removal rate coefficient (d^{-1}). Pollutant load accumulated over road surface when equilibrium is reached (e.g. $T_{DRY,i} = +\infty$) is therefore D_{ACC}/D_{ERO} ($g.m^{-2}$). (Note that the initial sediment storage $M_{ACC}(t=0)$ at the beginning of simulation is also a parameter of the model)

Previous equation was however reported to be inappropriate because of the little explanatory values of antecedent dry period duration in loads variability (Kanso et al., 2005; Shaw et al., 2010). Alternative accumulation model were thus tested assuming either invariant initial storage $M_{ACC}(t_{0,i})$ from an event to another (Eq. 6) or an exponential decrease of the initial pollutant load $M_{ACC}(t_{0,i})$ available at the beginning of each rain event over the studied period (Eq. 7) so as to cope with turbidity measurements (cf. 2.1):

$$M_{ACC}(t_{0,i}) = M_{CST} \quad (6)$$

$$M_{ACC}(t_{0,i}) = (M_{INI} - M_{LIM}) \times Exp(-K \times t_{0,i}) + M_{LIM} \quad (7)$$

Where: M_{INI} = initial sediment load over road surface ($g.m^{-2}$); M_{LIM} = limit value for M_{ACC} ($g.m^{-2}$); $t_{0,i}$ = beginning date of rain event (d); K = model parameter (d^{-1}).

2.3. CALIBRATION AND PERFORMANCE ASSESSMENT

2.3.1. Metropolis-Hasting algorithm

While automatic calibration procedures are usually implemented to estimate parameter values that minimize discrepancies between model outputs and measurements, such approaches have often been reported to prevent meaningful assessment of models adequacy (Deletic et al., 2012; Gaume et al., 1998; Kavetski et al., 2006). A Markov Chain Monte-Carlo sampling method based on Metropolis-Hasting (1970) algorithm is hence adopted to perform model calibration from water quality measurements according to the approach proposed by Kanso et al (2004; 2005).

The Bayesian approach assumes that uncertainties in model parameters θ, considering a dataset D, can be represented by a probability distribution P(θ|D), corresponding to prior knowledge about model parameters P(θ) updated by observations D (Congdon, 2006). From Bayes' theorem, posterior probability distribution may be written as follow:

$$P(\theta|D) = P(\theta) \times P(D|\theta) / \int P(D|\theta).P(\theta).d\theta \quad (8)$$

Where $\int P(D|\theta).P(\theta).d\theta$ can be seen as a normalising constant and P(D|θ) is the probability of observed outcomes D given parameters values θ denoted as the likelihood L(θ|D) function of accumulation and wash-off model. In this study, P(θ) is considered as uniform since no prior knowledge about parameters is available (Kanso et al., 2005). Assuming that residuals ε between model's outputs and observations are independent and normally distributed ($\varepsilon \sim N(0,\sigma)$), likelihood function L(θ|D) becomes:

$$P(D \mid \theta) = L(\theta \mid D) = \prod_{i=0}^{N} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{\left[Y_i - f\left(X_i\right) \right]}{2\sigma^2} \right) \right] \quad (9)$$

Where $(Y_1,\ldots,Y_N)$ is a vector corresponding to measurements (e.g. suspended solids concentrations or loads records), $(X_1,\ldots,X_N)$ are input data associated with $(Y_1,\ldots,Y_N)$, and $f(X_i, \theta)$ represents model's response for X_i and a set of parameters θ. Both σ and θ are considered as parameters to be estimated through calibration.

While direct analytical calculation of posterior probability distribution P(θ|D) is generally impossible Metropolis-Hasting (1970) algorithm can be implemented to approximate P(θ|D). The principle of this method lies in its ability to generate a random walk through parameter space that converges to the posterior probability distribution (Chib and Greenberg, 1995). For each iteration, a set of parameters θ' is drawn from a normal candidate generating-density and likelihood associated with θ' is computed. Parameter acceptance depends upon a transition probability defined to ensure convergence towards posterior distribution P(θ|D) (for further details, see Chib and Greenberg, 1995)

2.3.2. Uncertainty analysis and bias description

Posterior distributions computed from Metropolis algorithm represent uncertainties associated with model parameters and therefore provide relevant information on the significance of calibration results. Likelihood as given in equation 9 is typically a least-square objective function and similar to the widely used Nash Sutcliffe (1970) model efficiency coefficient model (E). For convenience, model performance is here expressed in terms of E rather than $L(\theta|D)$.

It is however important to acknowledge that likelihood function assumes that residuals between model predictions and measurements are independent and normally distributed ($\varepsilon \sim N(0,\sigma)$). While such hypothesis is seldom questioned in urban hydrology, it is very likely to be violated (Del Giudice et al., 2013) and, as expected, model errors were here neither found to be independent or normally distributed (cf. figure 2). Data transformation to ensure normality and homoscedasticity of residuals has often been applied in hydrology (Li et al., 2011; Yang et al., 2007). While such approach was found to allow for "a more detailed" model examination, requiring the model to fit a wider portion of measured data (Dotto et al., 2013), it may not always be desirable as resulting in a change in the objective function (Dotto et al., 2011). Moreover, transformed residuals often remain auto-correlated (Del Giudice et al., 2013) and application of formal Bayesian Inference methods with correct bias description is yet a challenge (Evin et al., 2013).

In this study, posterior distributions were computed under the unverified assumption of normally distributed and independent residuals so as to preserve the least-square objective function. Although such a simplification probably introduces a bias in parameters uncertainty estimation (Dotto et al., 2013; Thyer et al., 2009), it was also found to produce similar results as non-formal techniques (such as Generalized Likelihood Uncertainty Estimation) and remains relevant for sensitivity analysis (Dotto et al., 2012, 2011; Vrugt et al., 2008). 1-99% confidence intervals associated with parameter uncertainties were thus computed from loads or concentrations simulated for each set of parameter of sample distribution (cf. Kanso et al., 2005). Conversely, violation of the Gaussian error assumption clearly prevents reasonable estimation of total predictive uncertainty (Dotto et al., 2011). As indicated in figure 2c, residuals between simulated and measured concentrations appear to depend on measured values (similar results were observed for loads) and total predictive uncertainty cannot therefore be plotted on the basis of σ value. In this paper, the 10-90% (total uncertainty) confidence intervals were estimated from the structure of the residuals, according to the method outlined by Dotto et al (2011), assuming a non-linear relationship between residuals and simulated values.

2.3.3. Application to dataset

The selected water quality model assumes that the amount of pollutant available at the beginning of a rain event depends on antecedent weather conditions and more generally results from a succession of wet and dry periods. Calibration and evaluation of these models hence requires continuous modelling of accumulation and wash-off processes over the studied period from runoff measurements. A simple hydrological model (runoff coefficient + non linear reservoir model) was thus calibrated (E=0.74, total runoff volume error = 0.8%) to generate flow-rate measurements during missing data periods from local rainfall records. Although higher model performance might have been expected, this inability to perfectly replicate flow-rate measurements is presumably related to the distance between studied site and the rain gage (Vezzaro et al., 2012). It should nonetheless be outlined that simulated flow-rates here only aim at modelling realistic evolution of available suspended solid loads on road surface when runoff measurements are missing.

The whole monitoring period (January 2013 to November 2013) was first considered to perform calibration for concentrations and loads modelling from flow-rate measurements completed by simulated data. 175 rain events were identified using a 30 min minimum inter event duration (period without runoff). Calibration was successively conducted with and without the first 12 events (cf. 3.1.1.). Model's ability to replicate sediment concentrations and fluxes was later evaluated for a shorter calibration period (03/06 to 23/07, 14 events). Whereas calibration only reflects how well the model can reproduce TSS measurements, validation allows for predictive power assessment and can therefore provide further information on model consistency for water quality modelling. Four contiguous periods (7 runoff events each), ranging from 30/05 to 13/09, were thus identified to perform validation. Each of them was successively considered for calibration and model predictions were hence each time evaluated for the three others. A summary of calibration and validation periods with corresponding water quality models is given in table 1. For each simulation, Metropolis algorithm was run for 500.000 iterations from optimal parameter values (e.g. maximum of likelihood) previously identified from an "initialization" run (500.000 iterations) to ensure sampling from a stationary parameter distribution.

3. RESULTS AND DISCUSSION

3.1. LONG-TERM SIMULATION OF ACCUMULATION AND WASH-OFF

3.1.1. Results for concentration modelling

Unsatisfactory results were first obtained from calibration over the entire monitoring period (Nash Sutcliffe efficiency coefficient E=0.33). This poor model performance was attributed to model's inability to cope with the

sudden increase in turbidity measurements that occurs after mid-January (cf. figure 1). Calibration period was thus redefined so as to eliminate the first 12 rain events occurring before 19/01.

A noticeably higher performance is apparently achieved after modification of the calibration period (E=0.61). However, while E value appears relatively high compared to other results from the literature (Dotto et al., 2011), the model seems mostly unable to properly simulate fluctuations in TSS concentration during a given rain event (cf. figure 3). Conversely, the accumulation and wash-off model succeeds in replicating overall decrease of concentrations in runoff over the calibration period (evidenced in figure 1). Optimal parameter values (cf. table 2) indeed indicate that the initial sediment storage M_{ACC} (t = 0) is much higher than the equilibrium load D_{ACC}/D_{ERO} ($T_{DRY,i} = +\infty$ in Eq. 5). Simulated concentrations are therefore, for the first events, driven by this high initial load M_{ACC} (t = 0), and progressively decrease as the storage M_{ACC} returns to its equilibrium value. The value calculated for E may therefore simply result from model's ability to adapt to this trend in TSS concentrations over the calibration period.

Dispersion of posterior probability distributions (cf. figure 4) is generally moderate, although more important for accumulation model parameters D_{ACC} and D_{ERO}, which indicates that an optimal set of parameters could easily be identified. As a consequence, uncertainties associated with calibration can be regarded as relatively low. Concerning simulated concentrations, narrow confidence intervals are obtained for parameter uncertainty, whereas the model fails to simulate wash-off dynamics at the event scale and displays a very high total predictive uncertainty (cf. figure 3). This first suggests that calibration results may be largely driven by the annual decrease in TSS concentrations; optimal accumulation parameters would therefore not necessarily reflect the ability of the model to simulate sediment accumulation between rain events. This hypothesis is supported by (1) the more important dispersion of D_{ACC} and D_{ERO} and (2) the cross-correlation observed between these two parameters (not shown here). An increase in D_{ERO} is indeed compensated by a decrease of D_{ACC}: the effect of these parameters as "erosion" or "accumulation rates" on the likelihood function is hence much lower than the effect of the equilibrium load D_{ACC}/D_{ERO} and the dispersion of D_{ACC}/D_{ERO} is thus lower than the one of D_{ACC} and D_{ERO} (cf. Figure 4). From figure 3 it is however quite clear that the poor fit between simulated concentrations and calibration data does not only result from erroneous prediction of sediment loads at the beginning of rain events. The performance of wash-off model itself will be discussed in 3.2, considering shorter calibration periods to attempt to avoid the influence of incorrect representation of long-term accumulation process.

So as to better evaluate the significance of previous calibration results, an alternative accumulation model (Eq. 7) is tested under the assumption that antecedent dry period has no effect on the initial pollutant load $M_{ACC}(t_{0,i})$ which simply decreases over the studied period.

Calibration results indicate that the omission of pollutant accumulation process modelling during dry periods does not significantly affect model performance (E=0.60 vs. E=0.61 previously). The little change in E value confirms that exponential accumulation and wash-off model performance period is here essentially related to its ability to replicate the annual decrease of TSS concentrations, with successions of wet and dry periods having only a limited effect on simulated concentrations. While antecedent dry period T_{DRY} has traditionally been considered as a reliable predictor of suspended solids concentrations or loads in runoff, representation of pollutant accumulation as a process solely dependant on this factor may therefore not be relevant. Kanso et al (2005) also found the initial sediment load $M_{ACC}(t_{0,i})$ to be independent from T_{DRY} and suggested that it might be considered as constant from an event to another (Eq. 6). Our results rather indicate that $M_{ACC}(t_{0,i})$ is highly variable and are thus consistent with those reported by Shaw et al (2010) who found pollutant accumulation to be mostly unpredictable and influenced by stochastic input of particulate matter.

In this study, the increase of TSS concentrations during the first half of the monitoring period could not be explained by direct on-site observations. It may however be hypothesized that various circumstances, such as construction work in the neighbourhood, or seasonal phenomena such as leaf fall or application of de-icing salts during winter periods (Deletic and Orr, 2005), can result in an acceleration of sediment deposition (independently from dry periods duration). The trend observed in TSS concentrations therefore indicates that such unpredictable occurrences can completely alter and drive the accumulation process which suggests that the traditional assumption of a pollutant accumulation only related to dry period duration may not be relevant for long term runoff quality modelling.

3.1.2. Results for load modelling

Calibration for suspended solids loads modelling was performed over the period ranging from January 19th to November 2013. Model's ability to replicate suspended solids discharge rates seems much better than for concentrations (E=0.79 vs. E=0.61) (cf. figure 5). This performance increase is in agreement with literature results which generally show that traditional water quality models produce more accurate estimates of loads than concentrations (Crobeddu and Bennis, 2011; Dotto et al., 2011; Shaw et al., 2010).

Optimal parameter values for loads however significantly differ from those calculated for concentrations (cf. table 2); TSS concentrations computed from these parameters hence do not really fit with measurements and result in a poor model performance (E=0.35 vs. E=0.61). Correspondingly, Nash efficiency coefficient is significantly affected when suspended solids loads are simulated from parameters associated with the best prediction of TSS concentrations (E=0.46 vs. E=0.79).

Despite the poor performance observed at the event scale for TSS concentration modelling (cf. 3.1.1), the exponential accumulation and wash-off model apparently provides relatively accurate load estimates. As suggested by Joannis et al (2014) in the case of combined sewers, suspended solid loads variability might thus be largely explained by runoff volumes (which can be measured or easily simulated) and satisfactory model performance for loads may not always be meaningful. Previous results indeed demonstrate that reasonable load prediction does not necessarily reflect the model's ability to replicate pollutant wash-off dynamics and TSS concentrations (and reciprocally). Model's relevance for loads estimation, as compared to simpler Event Mean Concentrations (EMC) formulations (cf. Eq. 1), should therefore probably be questioned.

Calibration was thus performed for the two EMC models (Eq. 1 and 2) assuming either a constant concentration over the simulation period or an exponential decrease of concentrations from mid-January to November. Nash model efficiency coefficients calculated for both "constant EMC" and "decreasing EMC" hypotheses are relatively high (E=0.61 and E=0.77 respectively vs. E=0.79). These results therefore support the idea that simple models may as well produce very acceptable load estimates as long as flow-rates are correctly measured or simulated. Eventually, "decreasing EMC" hypothesis performs almost as well as the initial exponential accumulation and wash-off model. Accounting for temporal variability in TSS concentrations therefore provides only slight improvement for loads prediction, which again indicates that this high E-value (associated with the decreasing EMC hypothesis) is mostly associated with the replication of TSS concentration decrease over the studied period.

3.2. APPLICATION OVER SHORTER PERIODS

Calibration over the 11 month period primarily suggests that the exponential accumulation and wash-off model fails at simulating concentration variations at the event scale. Random occurrences, resulting in unpredictable sediment inputs, are indeed likely to disrupt the process of sediment deposition and thus traditional formulations relating accumulation to antecedent dry period are probably inappropriate when applying these models over long periods.

The model investigated here was however reported to perform relatively well (at similar or wider scales) when applied to shorter periods or a limited number of rain events (Freni et al., 2009; Mannina and Viviani, 2010; Piro and Carbone, 2010). Several studies additionally evidenced that antecedent dry period was a relevant predictor for pollutant accumulation between rain events (Egodawatta and Goonetilleke, 2006; Vaze and Chiew, 2002), although, as suggested by Shaw et al (2010), accumulation rate may significantly vary over time. A shorter period, ranging from June 3rd to July 23rd (14 rain events over 51 days), far enough from mid-January concentration peak, was thus selected to evaluate model suitability for concentration and loads replication and prediction.

3.2.1. Results for TSS concentration modelling

Model calibration for the studied period results in a satisfactory fit with TSS concentration measurements (E=0.55, cf. Figure 7a). Posterior distributions computed from Metropolis Hasting algorithm (figure 6) indicate that uncertainties associated with most parameters are low. The dispersion obtained for M_{ACC} (t=0) simply demonstrates that the effect of initial sediment storage on model outputs quickly becomes negligible (since accumulation and wash-off are simulated for the whole monitoring period, although calibration is only performed over 51 days).

Values associated with D_{ERO} and D_{ACC} are however surprisingly high and similar to those obtained by Kanso et al (2004) who concluded that sediment desposition could be modelled as an instantaneous process (independent from antecedent dry period). Indeed, for D_{ERO} = 2.65 d^{-1}, 93% of the equilibrium load ($D_{ACC}/D_{ERO,}$ cf. Eq. 5) is reached within 24h. While uncertainties associated with these parameters appear as relatively moderate, the relevance of accumulation process representation may thus once again be questioned and current model was compared to a simpler formulation. Assuming constant initial sediment storage M_{ACC} ($t_{0,i}$) from an event to another results in a moderate decrease in E value (E=0.42 vs. 0.55 for the exponential accumulation model). The Alley and Smith (1981) model can hence be regarded as suitable for the studied period even though instantaneous accumulation assumption does not dramatically affect model performance. The high D_{ERO} value indeed indicates that initial sediment storage is mostly identical from an event to another unless very short antecedent dry periods are considered. The increase in E-value would therefore suggest that TSS concentrations during rainy periods, with successive rain events, are better represented when accumulation is simulated, although this conclusion is as well very dependant on rain event definition.

Even if satisfactory model performance is obtained for a studied period, calibration results may however differ from a period to another and do not reflect model's ability to replicate TSS concentrations for a wider range of events. So as to determine whether (1) calibration results are likely to differ from a period to another and (2) to assess the short-term predictive power of the model, calibration and validation was thus performed for 4 different periods. Studied period (from June 3rd to July 23rd) was therefore split in two 7-events intervals and two other intervals were additionally considered on each side of the June 3rd to July 23rd period.

As indicated in table 3, results remain decent as long as validation and calibration is performed within the June 3rd to July 23rd interval, which is not really surprising since a good model performance was observed for calibration over the whole period (E=0.55). Conversely, considering either first or last period for calibration or validation systematically produces very poor validation results (cf. figure 7b.), which reveals that the predictive power of the exponential accumulation and wash-off model should overall be regarded as unsatisfactory, unless considering very limited number of events immediately before or after calibration period.

The inability of the model to predict concentrations in runoff clearly reflects the fact that no single set of parameter can correctly simulate sediment wash-off dynamics for the 4 studied periods. As found by Métadier (2011) for larger catchments, optimal parameter values may indeed significantly vary from a calibration period to another (cf. table 4). These results are also consistent with those of Bai and Li (2012), who suggested that governing forces in suspended solids wash-off, resulting from both sediment supply and runoff ability to remove them, usually differ from an event to another. Interestingly, comparison of optimal parameters for studied periods evidences relatively small variations in wash-off parameter values, whereas accumulation parameters, which dictate the amount of pollutant available at the beginning of each event, significantly differ from a period to another. The lack of predictive power of the model is thus very likely to be explained by its inability to simulate sediment deposition between rain events.

It should nonetheless be underlined that this poor performance does not solely results from erroneous estimations of M_{RES} ($t_{0,i}$). As indicated in figure 7b, while the model indeed fails at predicting the order of magnitude of concentrations in runoff, replication of the temporal pattern of sediment concentrations is as well quite inaccurate. As a consequence, calibration results themselves remain quite unsatisfactory, although relatively short periods are considered. At this point it is however unclear whether this problem originates from the formulation of wash-off or data itself. As a matter of fact, while errors in turbidity or flow-rate records cannot be

completely excluded, the use of precise and high frequency data may also exacerbate discrepancies between measurements and simulated values (Del Giudice et al., 2013).

3.2.2. Results for suspended solid loads modelling

Optimal parameter values identified for loads over the June 3rd to July 23rd period are, once more, very different from those calculated for concentrations modelling (cf. table 5). While the model apparently produces outstanding results for loads estimation (E=0.96), very high uncertainties are nevertheless associated with accumulation model parameters (cf. figure 8).

As illustrated by scatter-plot in figure 8, the likelihood $L(\theta|D)$ in fact appears to be almost insensitive to D_{ERO} and D_{ACC}, which are again strongly cross-correlated (not shown here) and for which very high values, corresponding to an instantaneous pollutant accumulation, are identified through Metropolis algorithm. Assuming constant initial sediment storage $M_{ACC} = D_{ACC}/D_{ERO}$ (Eq. 6) for all rain events indeed results in similar model performance (E=0.96). Extreme simplification, considering constant TSS concentration (Eq. 1), however lead to a noticeable drop in Nash efficiency coefficient (E=0.81). These results therefore suggest that, although accounting for pollutant wash-off dynamics might still be relevant for loads modelling, attempts to simulate sediment deposition with the exponential model are clearly vain.

Interestingly, conclusions upon model suitability appear to depend on modelling objectives: whereas accumulation parameters have almost no effect on likelihood function when simulating sediment discharge rates, opposite results are indeed obtained for concentrations. This difference presumably lies in the characteristics of the objective function $L(\theta|D)$ which assigns more importance to higher observed values. As measured flow-rates generally exhibit a much wider range of variation than concentrations, calibration for loads basically depends on the correct replication of concentrations associated with the highest flow-rates, while calibration for concentrations basically requires model outputs to fit to the whole pollutograph.

Another consequence of the difference in the variances of loads and concentrations is that model calibration (with a least-square objective function) is generally easier for loads (cf. tables 2 & 5), as correct replication of the entire concentration record is not necessarily needed. For the same reason, validation results can be quite disappointing when considering short calibration periods (cf. table 6). Indeed, model fitting over a very limited number of event (and sediment discharge peaks) may lead to completely unrealistic short term (and long term) prediction of concentrations. It is thus presumably crucial that sufficiently long periods are used for calibration to expect any predictive power for loads modelling.

3.3. DISCUSSION AND PERSPECTIVES FOR FUTURE RESEARCH

From previous results, the traditional assumption that sediment accumulation can be described by a simple function of dry period duration is an obvious limitation of the model. The findings related to antecedent dry period duration (T_{DRY}) are consistent with several recent literature results which suggest that accumulation believably does not occur at a steady rate (Deletic and Orr, 2005; Shaw et al., 2010) and that the accumulation equation has thus little or no explanatory power (Dotto et al., 2011, 2010; Kanso et al., 2005). It should be however acknowledged that the dataset used in this study displays a particular temporal concentration pattern, presumably due to a large punctual sediment input over street surface. While such occurrences are likely to occur at any time and in any situation, the erratic behaviour of accumulation (and the inter-event variability of concentrations) might have been exacerbated by the fact that a relatively small catchment is considered here. The effect of isolated and unusual sediment inputs (such as construction works) is indeed likely to be smoothed for larger catchments, which might explain why satisfactory performance of conventional accumulation models are sometimes reported (Freni et al., 2009; Vezzaro, 2008). Nonetheless, seasonal phenomena such as leaf-fall or application of de-icing salts, which result in a more diffuse sediment supply, may still produce observable effects for larger catchments. Besides, further research is probably needed to better understand external processes that might affect pollutant accumulation and its long-term variability.

Regarding the wash-off function, this study indicates that both model performance and optimal parameter values may differ from a calibration period to another (cf. 3.2.1.). This variability suggests that the actual formulation, for which runoff-rate is the main explanatory variable, may not be able to capture all the processes involved in sediment wash-off. Previous results, based on laboratory and field experiment have indeed evidenced the importance of both raindrop energy (related to both raindrop diameter and intensity) or particle size in the wash-off process (Brodie and Rosewell, 2007; Egodawatta et al., 2007; Vaze and Chiew, 2003). Other model structures should therefore probably be tested to adequately replicate concentrations in runoff (although much simpler models may as well be acceptable for basic applications such as loads estimation).

More generally, in the perspective of stormwater management practices development, implementation of alternative modelling approaches, that incorporate the stochasticity of accumulation and wash-off, is presumably needed for simulation of realistic TSS concentrations over long periods.

Eventually, while MCMC method was shown to provide an interesting framework for sensitivity and uncertainty analysis and clearly evidenced the limitations of the exponential accumulation and wash-off model, this study

illustrates the issues arising from the selection of the likelihood function. Because the normality and independence of residuals assumption is usually unverified, rigorous implementation of the Bayesian framework generally requires data transformation or precise bias description which inevitably dictates the nature of the likelihood function (Del Giudice et al., 2013; Del Guidice et al., 2014). This likelihood function may however not systematically represent the modeller's perception of model performance, which is generally assessed from much simpler criterion (Dotto et al., 2011; McMillan and Clark, 2009). In this paper, model's ability to replicate loads and concentrations was for instance evaluated from the widely used Nash-Sutcliffe coefficient: the normality assumption was hence deliberately violated so as to preserve a least-square objective function for calibration. The choice of a performance criterion is always arguable and somewhat subjective (least-square objective functions for instance assigns more importance to highest measurement values). Nonetheless, implementation of informal likelihood functions (such as the Nash-Sutcliffe coefficient) that better represent modeller's judgement and allow for more complete exploration of the space of parameters (McMillan and Clark, 2009) is probably advisable as it provides more flexibility than formal Bayesian approaches.

4. CONCLUSION

The ability of a commonly used accumulation and wash-off' model to simulate total suspended solids (TSS) concentrations and loads in street runoff was investigated. Calibration and uncertainty analysis were conducted for both long and shorter periods from 11-month continuous turbidity and flow rate records, using a formal Bayesian approach (MCMC). From this analysis, it was found that:

- The accumulation and wash-off model did not manage to correctly replicate TSS concentrations for the 11-month period. Calibration was presumably hindered by an important increase of sediment deposition at the beginning of the monitoring period. The model, which assumes that accumulation solely depends on dry period between rain events, was thus unable to correctly simulate this process.

- When considering shorter periods for calibration, optimal accumulation parameter values were found to be significantly different from a period to another. Consequently, the predictive power of the model was mostly inexistent, unless considering very limited set of rain events.

- While assuming that sediment deposition and erosion occurs at steady rate might be acceptable for larger catchments (where the variability of sediment inputs may be smoothed), the results presented

here suggest that accumulation should rather be described as a stochastic process when small urban surfaces are considered.

- Eventually, suspended solid load predictions were relatively accurate despite the poor model performance for concentrations modelling. Accounting for fluctuations in TSS concentration during rain events may thus not systematically be relevant and respectable results can as well be obtained from very simple “event mean concentrations” models.

Confrontation of the “accumulation and wash-off” model with long-term continuous water quality records therefore clearly indicates that its relevance for both concentration and loads modelling should seriously be questioned. While simpler formulation can provide reasonable estimates of suspended solid loads, further research is believably needed to satisfactorily simulate concentrations. More specifically, analysis of longer water quality time-series could provide a better understanding of the accumulation process, which is probably essential to capture the temporal variability of concentrations in runoff.

5. ACKNOWLEDGEMENTS

This study has been conducted within the framework of the OPUR observatory. The authors gratefully acknowledge the Seine-Normandy Water Agency, Val-de-Marne Departmental Council, Seine-Saint-Denis Departmental Council, Hauts-de-Seine Departmental Council, City of Paris and the Interdepartmental Association for Sewerage Services in the Paris Metropolitan Area (SIAAP) for financial support.

6. REFERENCES

Ahiablame, L.M., Engel, B.A., Chaubey, I., 2012. Effectiveness of Low Impact Development Practices: Literature Review and Suggestions for Future Research. Water, Air, & Soil Pollution 223, 4253–4273. doi:10.1007/s11270-012-1189-2

Alley, W.M., Smith, P.E., 1981. Estimation of accumulation parameters for urban runoff quality modeling. Water Resources Research. doi:10.1029/WR017i006p01657

Aryal, R., Kanadasamy, J., Vigneswaran, S., Naidu, R., Lee, S.H., 2009. Review of Stormwater Quality, Quantity and Treatment Methds Part 2: Stormwater: Quality modelling. Korean Society of Environmental Engineers 14, 134–149.

Bai, S., Li, J., 2013. Sediment Wash-Off from an Impervious Urban Land Surface. Journal of Hydrologic Engineering 18, 488–498. doi:10.1061/(ASCE)HE.1943-5584.0000654

Bertrand-Krajewski, J.-L., 2004. TSS concentration in sewers estimated from turbidity measurements by means of linear regression accounting for uncertainties in both variables. Water Science and Technology 50, 81–8.

Brodie, I., Rosewell, C., 2007. Theoretical relationships between rainfall intensity and kinetic energy variants associated with stormwater particle washoff. Journal of Hydrology 340, 40–47. doi:10.1016/j.jhydrol.2007.03.019

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. Journal of the American Statistical Association 49, 327–335. doi:10.2307/2684568

Congdon, P., 2006. Bayesian Statistical Modelling. Wiley.

Crobeddu, E., Bennis, S., 2011. A New Model for Simulating TSS Washoff in Urban Areas. Applied and Environmental Soil Science 2011, 9. doi:doi:10.1155/2011/318765

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. Hydrology and Earth System Sciences 17, 4209–4225. doi:10.5194/hess-17-4209-2013

Del Guidice, D., Albert, C., Bares, V.B., Reichert, P., Jörg, R., 2014. The Effect of Model Complexity on Model Structure Uncertainty of Hydrodynamic Sewer Models, in: 13th International Conference on Urban Drainage (ICUD), 2014. Kuching, Malaysia.

Deletic, A., 1998. The first flush load of urban surface runoff. Water Research 32, 2462–2470. doi:10.1016/S0043-1354(97)00470-3

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

Deletic, A., Orr, D., 2005. Pollution Buildup on Road Surfaces. Journal of Environmental Engineering 131, 49–59. doi:10.1061/(ASCE)0733-9372(2005)131:1(49)

Dietz, M.E., 2007. Low impact development practices: A review of current research and recommendations for future directions. Water, Air, and Soil Pollution 186, 351–363. doi:10.1007/s11270-007-9484-z

Dotto, C.B.S., Deletic, A., McCarthy, D.T., 2013. Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals? Water Science and Technology 68, 1271–9. doi:10.2166/wst.2013.360.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T., Rauch, W., 2010. Stormwater quality models: performance and sensitivity analysis. Water Science and Technology 62, 837–843. doi:10.2166/wst.2010.325

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., Fletcher, T.D., 2011. Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. Environmental Modelling and Software 26, 1225–1239. doi:10.1016/j.envsoft.2011.03.013

Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W., Deletic, A., 2012. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Research 46, 2545–2558. doi:10.1016/j.watres.2012.02.009

Egodawatta, P., Goonetilleke, A., 2006. Characteristics of pollutants built-up on residential road surfaces, in: 7th International Conference on Hydroscience and Engineering, 10-13 September, 2006. doi:0977447405

Egodawatta, P., Thomas, E., Goonetilleke, A., 2007. Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. Water Research 41, 3025–3031. doi:10.1016/j.watres.2007.03.037

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 Resources Research 49, 4518–4524.

Freni, G., Mannina, G., Viviani, G., 2009. Urban runoff modelling uncertainty: Comparison among Bayesian and pseudo-Bayesian methods. Environmental Modelling and Software 24, 1100–1111. doi:10.1016/j.envsoft.2009.03.003

Gaume, E., Villeneuve, J.-P., Desbordes, M., 1998. Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. Journal of Hydrology. doi:10.1016/S0022-1694(98)00171-1

Hannouche, A., Chebbo, G., Joannis, C., 2014. Dynamics of pollutant discharge in combined sewer systems during rain events: chance or determinism? Water Science and Technology 69, 1751–8. doi:10.2166/wst.2014.088

Hannouche, A., Chebbo, G., Ruban, G., Tassin, B., Lemaire, B., Joannis, C., 2011. Relationship between turbidity and total suspended solids concentration within a combined sewer system. Water Science and Technology 64, 2445–52. doi:10.2166/wst.2011.779

Hasting, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109. doi:10.1093/biomet/57.1.97

Huber, W.C., Dickinson, R.E., 1988. Stormwater Management Model Version 4: Users Manual. Environmental Research Laboratory, EPA, Athens, GA, USA. doi:EPA 600/3-88/001a

Joannis, C., Hannouche, A., Chebbo, G., 2014. An assessment of the respective contributions of flow-rate and concentration variations to mass discharge variations at the outlets of two combined catchments during rain events. Urban Water Journal 1–7. doi:10.1080/1573062X.2014.939662

Kanso, A., 2004. Evaluation des modèles de calcul des flux polluants des rejets urbains par temps de pluie: Apport de l'approche bayésienne [Evaluation of urban stomwater quality models: benefits from bayesian approaches]. Ph.D. Thesis. Ecole des Ponts Paristech (France)

Kanso, A., Chebbo, G., Tassin, B., 2006. Application of MCMC-GSA model calibration method to urban runoff quality modeling. Reliability Engineering and System Safety 91, 1398–1405. doi:10.1016/j.ress.2005.11.051

Kanso, A., Tassin, B., Chebbo, G., 2005. A benchmark methodology for managing uncertainties in urban runoff quality models. Water Science and Technology 51, 163–170.

Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research 42. doi:10.1029/2005WR004368

Kleidorfer, M., Deletic, A., Fletcher, T.D., Rauch, W., 2009. Impact of input data uncertainties on urban stormwater model parameters. Water Science and Technology 60, 1545–1554. doi:10.2166/wst.2009.493

Li, L., Xu, C.-Y., Xia, J., Engeland, K., Reggiani, P., 2011. Uncertainty estimates by Bayesian method with likelihood of AR (1) plus Normal model and AR (1) plus Multi-Normal model in different time-scales hydrological models. Journal of hydrology 406, 54–65.

Mannina, G., Viviani, G., 2010. An urban drainage stormwater quality model: Model development and uncertainty quantification. Journal of Hydrology 381, 248–265. doi:10.1016/j.jhydrol.2009.11.047

McMillan, H., Clark, M., 2009. Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. Water Resources Research 45. doi:10.1029/2008WR007288

Métadier, M., 2011. Traitement et analyse de séries chronologiques continues de turbidité pour la formulation et le test de modèles des rejets urbains par temps de pluie [Treatment and analysis of continuous turbidities series for the test of urban stormwater quality models]. Ph.D. Thesis. INSA de Lyon, Villeurbanne (France).

Métadier, M., Bertrand-Krajewski, J.-L., 2012. The use of long-term on-line turbidity measurements for the calculation of urban stormwater pollutant concentrations, loads, pollutographs and intra-event fluxes. Water Research 46, 6836–6856. doi:http://dx.doi.org/10.1016/j.watres.2011.12.030

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology 10, 282–290. doi:10.1016/0022-1694(70)90255-6

Obropta, C.C., Kardos, J.S., 2007. Review of urban stormwater quality models: Deterministic, stochastic, and hybrid approaches. Journal of the American Water Resources Association. doi:10.1111/j.1752-1688.2007.00124.x

Piro, P., Carbone, M., 2010. Modelling approach to assess the variation of pollutant mass flow rates during storm events, in: GRAIE (Ed.), Novatech 2010, Lyon, France. Lyon, France.

Saget, A., 1994. Base de données sur la qualité des rejets urbains de temps de pluie: distribution de la pollution rejetée, dimensions des ouvrages d'interception [Stormwater quality database: distribution of pollutants and sizing of detention facilities]. Ph.D. Thesis. Ecole Nationale des Ponts et Chaussées (France)

Sartor, J.D., Boyd, G.B., Agardy, F.J., 1974. Water pollution aspects of street surface contaminants. Journal - Water Pollution Control Federation 46, 458–467.

Shaw, S.B., Stedinger, J.R., Walter, M.T., 2010. Evaluating Urban Pollutant Buildup/Wash-Off Models Using a Madison, Wisconsin Catchment. Journal of Environmental Engineering. doi:10.1061/(ASCE)EE.1943-7870.0000142

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research 45. doi:10.1029/2008WR006825

Tsihrintzis, V.A., Hamid, R., 2001. Modeling and Management of Urban Stormwater Runoff Quality: A Review. Water 11, 137–164. doi:10.1023/A:1007903817943

US-EPA, 1983. Results of the Nationwide Urban Runoff Program - Volume 1. Final Report.

Vaze, J., Chiew, F., 2003. Comparative evaluation of urban storm water quality models. Water Resources Research 39, 1280.

Vaze, J., Chiew, F.H.S., 2002. Experimental study of pollutant accumulation on an urban road surface. Urban Water 4, 379–389. doi:10.1016/S1462-0758(02)00027-4

Vezzaro, L., 2008. Sensitivity analysis and uncertainty evaluation of a conceptual stormwater quality model, in: 11the International Conference on Urban Drainage, Edinburgh, Scotland, 31st August-5th September 2008.

Vezzaro, L., Ledin, A., Mikkelsen, P.S., 2012. Integrated modelling of Priority Pollutants in stormwater systems. Physics and Chemistry of the Earth 42-44, 42–51. doi:10.1016/j.pce.2011.07.002

Vrugt, J.A., ter Braak, C.J.F., Gupta, H. V., Robinson, B.A., 2008. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment 23, 1011–1026. doi:10.1007/s00477-008-0274-y

Wang, L., Wei, J., Huang, Y., Wang, G., Maqsood, I., 2011. Urban nonpoint source pollution buildup and washoff models for simulating storm runoff quality in the Los Angeles County. Environmental Pollution 159, 1932–1940. doi:10.1016/j.envpol.2011.03.019

Yang, J., Reichert, P., Abbaspour, K.C., 2007. Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland). Water Resources Research. doi:10.1029/2006WR005497

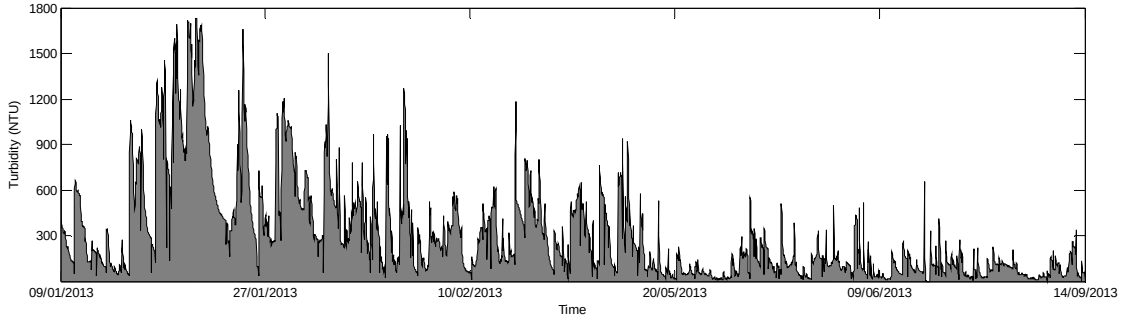

Figure 1 – Turbidity measurements from January to September 2013 (dry periods were excluded for better readability)

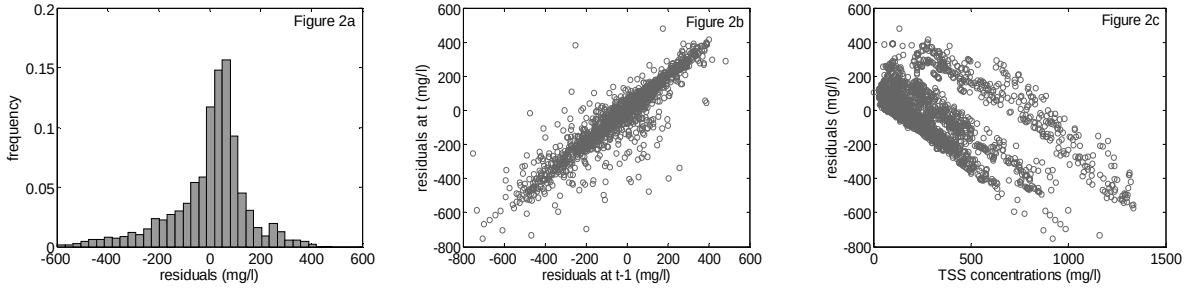

Figure 2 – Residuals analysis for concentration modelling: (a) Distribution, (b) first-order autocorrelation, (c) residuals against measured concentrations

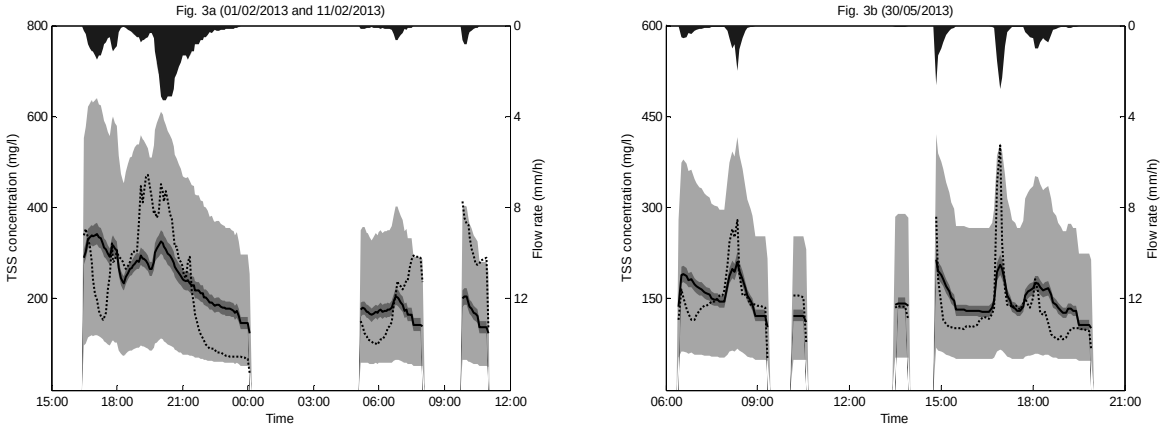

Figure 3 – Calibration results for 10/02 to 11/02 and 30/05 rain events. The black dashed line represents measured concentrations, the black bold line is simulated concentrations, dark shaded area is 1-99% parameter uncertainty, light shaded area is 10-90% total uncertainty and the black area is flow-rate over street surface.

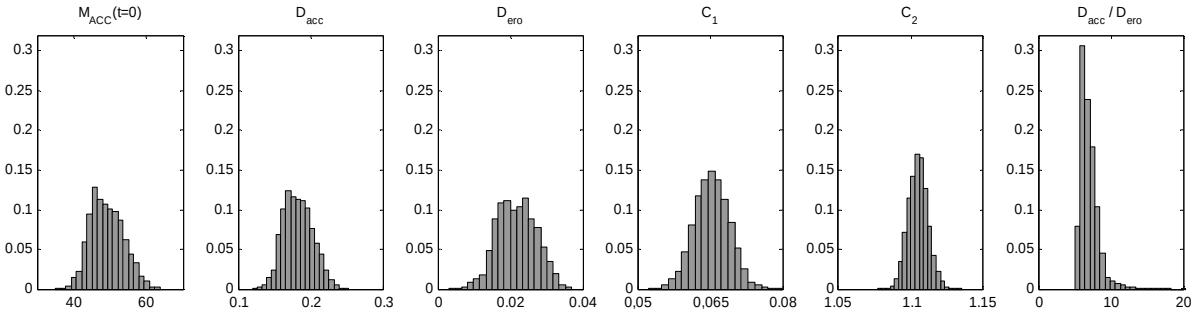

Figure 4 – Posterior probability distribution for model parameters

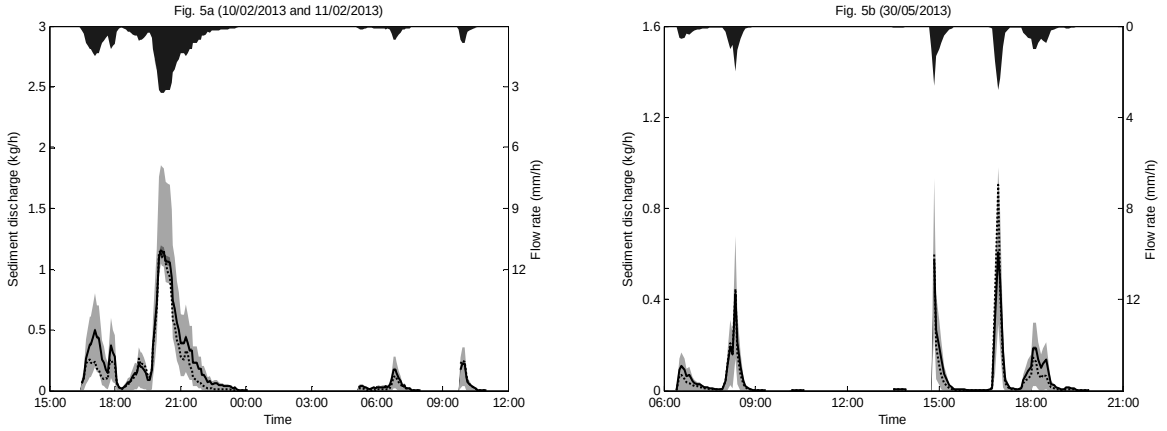

Figure 5 – Calibration results for 10/02 to 11/02 and 30/05 rain events - The black dashed line represents fluxes computed from flow-rate and turbidity measurements, the black bold line is simulated loads, light shaded area is total uncertainty and the black area is flow rate over street surface.

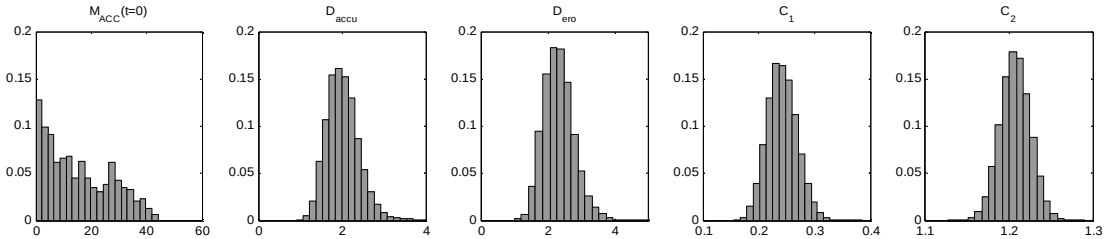

Figure 6 – Posterior probability distribution for model parameters (calibration on the June 3rd to July 23rd period)

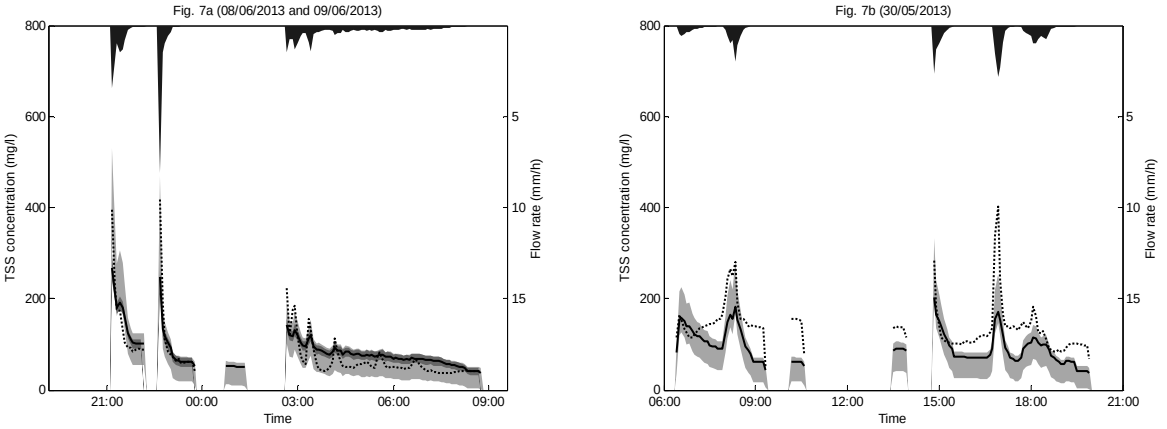

Figure 7 – Comparison between calibration (03/06 to 23/07) and validation results (30/05). The black dashed line represents measured concentrations, the black bold line is simulated concentrations, dark shaded area is 1-99% parameter uncertainty, light shaded area is 10-90% total uncertainty and the black area is flow rate over street surface.

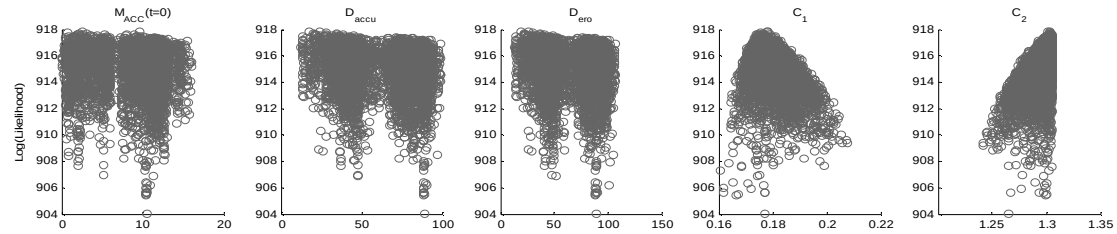

Figure 8 – Parameters value vs. likelihood (Calibration on loads for the 03/06 to 23/07 period).

Parameters	Calibration for concentrations	Calibration for loads
$M_{RES}(t=0)$ ($g.m^{-2}$)	47.3	31.9
D_{ACC} ($g.m^{-2}.d^{-1}$)	1.73×10^{-1}	8.8×10^{-3}
D_{ERO} (d^{-1})	1.90×10^{-2}	0.4×10^{-3}
C_1	0.065	0.021
C_2	1.10	1.18

Table 1 – Optimal parameter values for both concentration and load modelling

Calibration period	Validation period			
Dates	30/05 to 02/06	03/06 to 13/06	13/06 to 23/07	26/07 to 13/09
30/05 to 02/06	E_{CAL} = 0.40	-1.98	-1.67	-18.8
03/06 to 13/06	-0.17	E_{CAL} = 0.65	0.35	-1.51
13/06 to 23/07	-0.13	0.29	E_{CAL} = 0.63	-2.39
26/07 to 13/09	-0.79	0.16	-0.03	E_{CAL} = 0.26

Table 2 – Validation results (concentrations): Nash Sutcliffe efficiency criterion considering four 7-events periods for calibration (considering same period for calibration and validation gives model performance for calibration E_{CAL})

Calibration period	D_{ACC}	D_{ERO}	D_{ACC} / D_{ERO}	C_1	C_2
30/05 to 02/06	0.361	0.337	1.071	0.364	1.107
03/06 to 13/06	1.130	2.572	0.439	0.405	1.144
13/06 to 23/07	2.956	2.442	1.211	0.139	1.225
26/07 to 13/09	4.195	18.47	0.227	0.411	1.041

Table 3 – Optimal parameter values for 4 calibration periods

Parameters	Calibration for concentrations	Calibration for loads
$M_{RES}(t=0)$ ($g.m^{-2}$)	3.15	5.2
D_{ACC} ($g.m^{-2}.d^{-1}$)	1.79	16.5
D_{ERO} (d^{-1})	2.65	17.9
C_1	0.24	0.16
C_2	1.21	1.34

Table 4 – Optimal parameter values for both concentration and loads modelling (Calibration over the June 3rd to July 23rd period)

Calibration Period	Validation period			
Dates	30/05 to 02/06	03/06 to 13/06	13/06 to 23/07	26/07 to 13/09
30/05 to 02/06	E_{CAL} = 0.96	0.75	-2.87	-2.99
03/06 to 13/06	0.92	E_{CAL} = 0.99	0.44	-0.21
13/06 to 23/07	0.72	0.87	E_{CAL} = 0.96	0.78
26/07 to 13/09	0.54	0.72	0.90	E_{CAL} = 0.96

Table 5 – Validation results (loads): Nash Sutcliffe efficiency criterion considering four 7-events periods for calibration (considering same period for calibration and validation gives model performance for calibration E_{CAL})