Bayesian Model Averaging Approach for Urban Drainage Water Quality Modelling

Gabriele Freni, Mariacrocetta Sambito and Stefania Piazza

November 2, 2023

Bayesian model averaging approach for urban drainage water quality modelling

Gabriele Freni*, Mariacrocetta Sambito and Stefania Piazza
* Facoltà di Ingegneria ed Archittettura, Università di Enna Kore, Cittadella Universitaria 94100 Enna
(E-mail: *gabriele.freni@unikore.it)*

Abstract The uncertainty in urban drainage water quality modelling is highly relevance in any practical application. Several models are available in the literature for such tasks, and one of the most problematic choices is the selection of the most appropriate approach for the specific application. The Bayesian Model Averaging approach attempts to support the modeller in such choices by providing a method to identify and select the best performing models and average their output response to reduce the related uncertainty. In the current report, the Bayesian Model Averaging is applied to a real catchment and is compared with several single water quality models. The analysis showed that the Bayesian Model Averaging approach outperformed all single model applications.

Keywords water quality modelling, uncertainty analysis, Bayesian approach, urban drainage, modelling average technique

INTRODUCTION

In urban drainage modelling, uncertainty analysis is necessary. However, uncertainty analysis in urban water-quality modelling is still in its infancy, and only a few studies have been conducted. Therefore, several methodological aspects still need to be studied and clarified, especially regarding water quality modelling (Dotto et al., 2012; Freni et al., 2010b). Uncertainty analysis is a powerful component in practical analyses because it is able to highlight the reliability of the simulation result with respect to a specific problem (Deletic et al., 2012). Model simulations/predictions are subject to various uncertainties and sources of forecast errors (Freni et al., 2010a; Candela et al., 2009). One of the most important sources of uncertainty is due to the model structure, especially in those cases where different competing model structures may be selected. Reliance on a single model typically overestimates the confidence and increases the statistical bias of the forecast (Parrish et al., 2012). The model-averaging techniques attempt to overcome the limitations of a single model by linearly combining a number of competing models into a single new model forecast. This method dates back to studies conducted by Bates and Granger (1969), whose analysis used model-averaging techniques in economic forecasting and determined that a pooled forecast of competing models outperformed any single model's forecast. Techniques such as equal weight, Granger-Ramanathan averaging, and Bates-Granger averaging (Granger and Ramanathan, 1984, Diks and Vrugt, 2010) linearly combine the deterministic model outputs into another single-point deterministic forecast. The linear combination of models can be easily criticised because the averaging process is not connected to model performance. As an alternative approach to overcome these shortcomings, Hoetting et al. (1999) considered the idea of Bayesian model averaging (BMA), which is a technique that weights a model by its performance and likelihood of predicting an observation and results in a probabilistic forecast. The model weights are all nonnegative with the total sum equalling the unity, thus acting as a probability measure of a model's likelihood of success. Rafterty et al. (2005) incorporated these techniques on an ensemble of meteorological models, while other authors (Duan et al., 2007) used these techniques on hydrological models. Duan et al. (2007) showed that the BMA techniques out-perform, or are comparable to, an individual model forecast using a variety of point-wise performance measures on a number of conceptual rainfall-runoff (CRR) models. These results were obtained using a single training period to determine the weights of each model. Other researchers (Rafterty et al., 2005; Vrugt and Robinson, 2007) used a sliding window of training periods to estimate the BMA weights. Recently, the sequential Bayesian combination approach was introduced by Hsu et al. (2009) as an alternative to BMA by using the sequential Bayes' law in recursively updating the posterior probability of a model likelihood function given new observations. The posterior probability acts as weights for the multi-model averaging. The BMA application requires the definition of some measure of uncertainty. The uncertainty of a model result can be stated by providing a range (or a band) of values that are likely to enclose the true value of a specific simulated variable. In this study, a Bayesian approach coupled with a Monte Carlo analysis has been used (Bayes, 1763). In the current study, the BMA approach was applied to several pollution wash-off and build-up models to identify the best combination of models to analyse urban drainage water quality. The static BMA strategy was used to verify its effectiveness in urban drainage in comparison with single models by using the experimental catchment in Fossolo (Bologna, Italy).

MATERIALS AND METHODS

The mathematical model

In this study, 4 build-up algorithms and 2 wash-off algorithms were considered. The algorithms were generated by combining 8 models with different complexity levels. These models were integrated in the widely used Storm Water Management Model (SWMM) model (Huber, 1986). To simulate the build-up on the catchment surfaces, 4 approaches, which have been described in the literature, are considered (Huber, 1986; Bertrand-Krajewski et al., 1993).

BP_1 (Build-up Model 1): a simple linear accumulation function (eq. 1) in which the accumulated mass, Ma (kg), is a linear function of the antecedent dry weather period ADWP (d). The value of Ma is also dependent on the catchment area, A (ha), imperviousness ratio, IMP (-), and residual mass, Mr (kg), present on the catchment surface at the beginning of the dry weather period. The only calibration parameter in this equation is the unit accumulation rate, Accu (kg/ha/d), which represents the mass accumulated on the impervious unit area in the time unit.

$$\mathrm{Ma} = \mathrm{Accu} \cdot \mathrm{A} \cdot \mathrm{IMP} \cdot \mathrm{ADWP} + \mathrm{Mr} \quad (1)$$

BP_2: a power accumulation function (eq. 2) in which Ma is a power function of the ADWP. The symbols have the same definitions as described above. This approach has two calibration parameters: Accu and the shape parameter of the power law, Kpower (-). The variable C is a dimensional constant equal to 1 day that is introduced herein to preserve the unit consistency.

$$\mathrm{Ma} = \mathrm{C} \cdot \mathrm{Accu} \cdot \mathrm{A} \cdot \mathrm{IMP} \cdot \left(\frac{\mathrm{ADWP}}{\mathrm{C}}\right)^{\mathrm{Kpower}} + \mathrm{Mr} \quad (2)$$

BP_3: a saturation function (eq. 3) in which the Ma functional dependency is asymptotic with the ADWP. The symbols have the same definitions as described above. This approach has two calibration parameters: Accu and the saturation parameter, Ksat (d).

$$Ma = \left(\frac{Accu}{Ksat + ADWP}\right) \cdot A \cdot IMP \cdot ADWP^2 + Mr \quad (3)$$

BP_4: the widely applied Alley-Smith (1981) model (eq. 4), in which Ma is an inverse exponential function of the ADWP. The model introduces the effect of external causes (such as wind or traffic) that can modify pollutant accumulation by moving parts of the accumulated pollutants outside of the catchment area. The symbols have the same definitions as described above. This approach has two calibration parameters: Accu and the dispersion parameter, Disp (d^{-1}), which take into account the causes that reduce the pollutant mass availability on the catchment.

$$Ma = Mr \cdot e^{(-Disp \cdot ADWP)} + \left(\frac{Accu}{Disp}\right) \cdot A \cdot IMP \cdot \left(1 - e^{(-Disp \cdot ADWP)}\right) \quad (4)$$

The first two approaches are not superiorly limited; therefore, they can produce indefinite build-up as the ADWP grows. The other two models introduce asymptotic behaviour, and BP_4 attempts to consider external causes that can remove part of the pollutants.

The wash-off of solids by overland flow during a storm event was simulated using the method derived by Jewell and Adrian (1978). Two models have been considered for the wash-off process, both of which are widely existent in literature (Huber, 1986):

WH_1 (Wash-off Model 1): the original formulation by Jewel and Adrian (1978), in which Me (kg) is the mass entering the network between t and t+Δt (h), M_a is the mass on the catchment at time t, P_n is the net rainfall intensity (mm/h), and Δt is the time step. The formulation requires the calibration of two parameters: the wash-off factor, Wh (-) and the wash-off coefficient, Arra ($mm^{-Wh} \cdot h^{(Wh-1)}$).

$$Me = Ma \cdot (1 - e^{-(Arra \cdot P_n^{Wh} \cdot \Delta t)}) \quad (5)$$

WH_2: a simplified version of WASHOFF_1 in which Wh is assumed to be equal to 1, thus reducing the number of parameters that require calibration.

$$Me = Ma \cdot (1 - e^{-(Arra \cdot P_n \cdot \Delta t)}) \quad (6)$$

The primary difference between the two wash-off approaches is the shape parameter, Wh, with the latter assuming that the exponent in eq. 6 is linearly dependent on the net rainfall intensity.

In this study, the build-up and wash-off models have been combined to obtain different modelling approaches with progressively decreasing levels of complexity. Table 1 shows the eight models that were considered in the current study and the number of parameters that each requires for calibration.

The Bayesian Model Averaging strategies

Consider a quantity, y, to be forecasted, such as the sewer flow at a particular location and time. Assume there are n models [m_1, m_2, …, m_k], each providing an independent model forecast, Ym_i, with an output time series with i=1, 2, …, k. In general, the BMA procedure seeks to compute a new forecast probability density of the modelling output as a weighted average of the competing models forecasts with weights that correspond to the comparative performance of the models over a certain training period of observations with duration T, Y= [y_1, y_2, …, y_T]. Usually, the BMA methodology assumes that the model errors are unbiased; that is, the expected value of the differences between the model prediction and the observations E [Ym_i – Y] = 0 for each model i. If a certain systematic bias in error is present, different correction strategies are available (such as Box-Cox transformation or Normal Quantile Transformations); in the current study, a linear regression is used as follows:

$$Y = a_i E[Ym_i] + b_i \quad (7)$$

Unique coefficients a_i and b_i for each model can be determined using a least squares approximation, with the observations in the training period as the dependent variable and the forecast as the explanatory variable. Then, these coefficients are applied to all future model forecasts. After the application of the bias correction strategy, all future references to model forecasts are assumed to be unbiased. The forecast density for y conditioned on the models forecast, m_i, and training period of observations, Y, can be expressed according to the law of total probability as follows:

$$P(Ym|m_1, m_2, ..., m_k, Y) = \sum_{i=1}^{k} P(Ym|m_i, Y)P(m_i|Y) \quad (8)$$

where $P(Ym|m_i, Y)$ is the posterior distribution of y based only on model m_i and the training data Y. $P(m_i|Y)$ is defined as the posterior probability, or the relative likelihood, of model m_i being correct given the training data Y. The classical BMA methodology assumes that the posterior probability of model m_i conditioned over the observations Y is associated with the uncertainty within an individual model (Duan et al., 2007). This approach seeks to obtain static parameters (weights and model posterior probabilities) from a fixed training period with a set of observations Y. The conditional posterior distribution of y, based only on model m_i, can be computed by the Bayesian approach, which leads to the hypothesis that modelling errors are normally distributed and unbiased. A possible systematic bias can be considered by equation 7. The Bayesian posterior probability distribution of the forecast can be obtained by the equation as follows:

$$P(Ym \mid m_i, Y) = \frac{P(Y \mid m_i, Ym)P(Ym)}{\int P(Y \mid m_i, Ym)P(Ym)\,dYm} \quad (9)$$

where the posterior parameter distribution is computed as a function of the prior knowledge $P(Ym)$ and the conditional probability for the measured data given the model output $P(Y \mid m_i, Ym)$, which is often referred to as the *likelihood function* of the model. Assuming that the residuals between the model and observations are distributed based on a normal distribution with a null average and variance, σ_e^2, $P(Y \mid m_i, Ym)$ can be written in the multiplicative form on the training period T as follows:

$$P(Y \mid m_i, Ym) = \prod_{k=1}^{T} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(\frac{(y_k - Ym(k))^2}{-2\sigma_e^2}\right) \quad (10)$$

where Ym (k) is the modelling response at the time k that corresponds to the available measures y_k of a specific variable (i.e., discharges, concentrations, loads, etc.) in the training dataset. The posterior distributions are populated by running Monte Carlo Simulations (1000 in the current study) using parameter values randomly drawn from prior distributions.

The BMA method needs to define $P(m_i|Y)$; a static weight w_i can be computed based on the average uncertainty band width as follows:

$$w_i = 1 - \frac{\max(\overline{l_{mi}}) - \overline{l_{mi}}}{\sum_{i=1}^{k}(\max(\overline{l_{mi}}) - \overline{l_{mi}})} \quad (11)$$

where $\overline{l_{mi}}$ is the average uncertainty bandwidth provided by the model m_i on the training set of observations Y using the Bayesian approach.

The case study

The Fossolo catchment is located in Italy in a residential area near Bologna (Figure 1). The catchment is characterised by an independent combined sewer network that does not receive any contributions from the surrounding catchment. Fossolo encompasses a total area of 40.71 ha, 74.80% of which is impermeable (approximately 30.45 ha). The buildings in the area are primarily residential, and service sector businesses are also present. The catchment has approximately 10,000 equivalent inhabitants. The catchment contains roads with a high vehicle flux (approximately 40,000 vehicles/day) and streets with a low vehicle flux (approximately 1,000 vehicles/day). The final collector section of the sewer network is polycentric with dimensions of 144 cm in height and 180 cm in width. The discharge is estimated from the water depth, which is measured by an ultrasonic probe placed in the main channel. A refrigerated automatic sampler with 24 l bottles is used. The BOD, COD and TSS were assessed for each sample according to Standard Methods (APHA, 1995). In the catchment, 12 events have been measured for both quantity and quality aspects (Table 1). The field campaign was conducted by Bologna University (Artina et al., 1997).

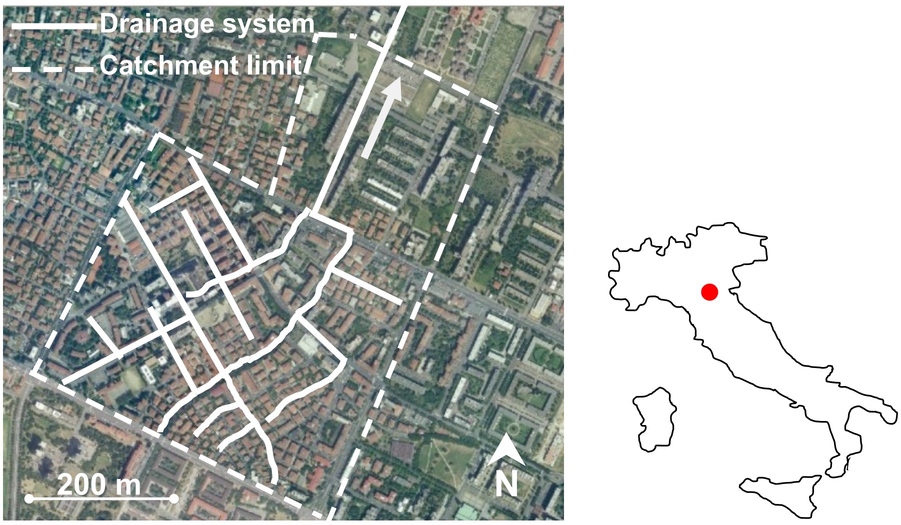

Figure 1 The Fossolo catchment (Bologna, Italy)

ANALYSIS OF RESULTS

The available data set was divided into two blocks of six events in order to have a training set and validation set. Each of the eight possible methods was used separately to run a classical

Bayesian analysis using the training set to update the parameter posterior distributions and uncertainty bands. The BMA was concurrently used to identify the most probable combination of the eight models to reduce the uncertainty of the forecast. The classical Bayesian analysis was performed using 1000 random Monte Carlo Simulations (MCS).

Table 1 Characteristics of 12 rain events used in the application

	Total rainfall (mm)	**Max intensity (mm/h)**	**Duration (min)**	**ADWP (hours)**
Minimum	2.8	12	45	50
Maximum	72.2	147	921	402
Mean	18.9	45	258	167

The same simulations were used to estimate the weights w_i and the posterior forecast probability needed to run the BMA. Assuming no prior knowledge about the distribution of parameters, the parameters were considered to be uniformly distributed in the ranges proposed in Table 2.

Table 2 Models and parameter variation ranges adopted for the BMA

Model →			**M_01**	**M_02**	**M_03**	**M_04**
Description ↓	**Symbol**	**Unit**	**BP_1 WH_1**	**BP_2 WH_1**	**BP_3 WH_1**	**BP_4 WH_1**
daily accumulation rate	Accu	[kg/ha/d]	0.16-5	0.25-5.4	0.3-5.2	0.2-4.8
decay rate in Alley-Smith model	Disp	[d^{-1}]	0.01-0.5	-	-	-
wash-off coefficient	Arra	[$mm^{-Wh} \cdot h^{(Wh-1)}$]	0.015-0.4	0.02-0.55	0.015-0.45	0.03-0.5
wash-off factor	Wh	-	0.5-1.5	0.5-2	0.5-1.5	0.5-2
half-saturation constant	Ksat	[d]			-	40-120
power law coefficient	Kpower	[-]			0.01-1	-
Model →			**M_05**	**M_06**	**M_07**	**M_08**
Description ↓	**Symbol**	**Unit**	**BP_1 WH_2**	**BP_2 WH_2**	**BP_3 WH_2**	**BP_4 WH_2**
daily accumulation rate	Accu	[kg/ha/d]	0.1-5.5	0.25-6.5	0.14-8.5	0.35-6
decay rate in Alley-Smith model	Disp	[d^{-1}]	-	-	-	0.07-0.8
wash-off coefficient	Arra	[$mm^{-Wh} \cdot h^{(Wh-1)}$]	0.03-0.55	0.01-0.6	0.012-0.8	0.01-0.5
wash-off factor	Wh	-	-	-	-	-
half-saturation constant	Ksat	[d]	18-150	-	-	-
power law coefficient	Kpower	[-]	-	0.01-1	-	-

After evaluating the weights, three models are clearly more likely to represent the analysed system in the training dataset. Specifically, models M_6, M_7, and M_8 are significantly more relevant for the analysed system. All of the models are characterised by the wash-off model WH_2, and the build-up models are not the most parameterised in the analysis. Models M_1, M_2, M_4 and M_5 reported weights lower than 0.1, and they can be neglected in the BMA to simplify the mixed approach. The weights connected with the neglected model can be proportionally distributed among the three models being used. Then, the final set of weights is $w_6 = 0.372$, $w_7 = 0.323$, and $w_8 = 0.305$. The validation set was then used to estimate the uncertainty on the output TSS concentration. Each model was considered separately using the classical Bayesian approach, which provided the uncertainty bands reported in figure 2. Different models provide different uncertainty bandwidth, thus indicating that they are essentially able to learn from the training dataset and are differently adaptable to the analysed case study. Figure 3 shows the same uncertainty bands provided by the BMA approach by averaging the three best performing models (M_6, M_7 and M_8) and neglecting all others. The uncertainty bandwidth is lower, indicating that the BMA approach outperforms the application of single models, even if it requires larger computational efforts because of the need to assess uncertainty bands for all possible candidate models. Table 3 shows the best behaving model for all 12 events and the comparison with BMA: the BMA approach always outperforms the single model applications, and the uncertainty bands are reduced up to approximately 30%.

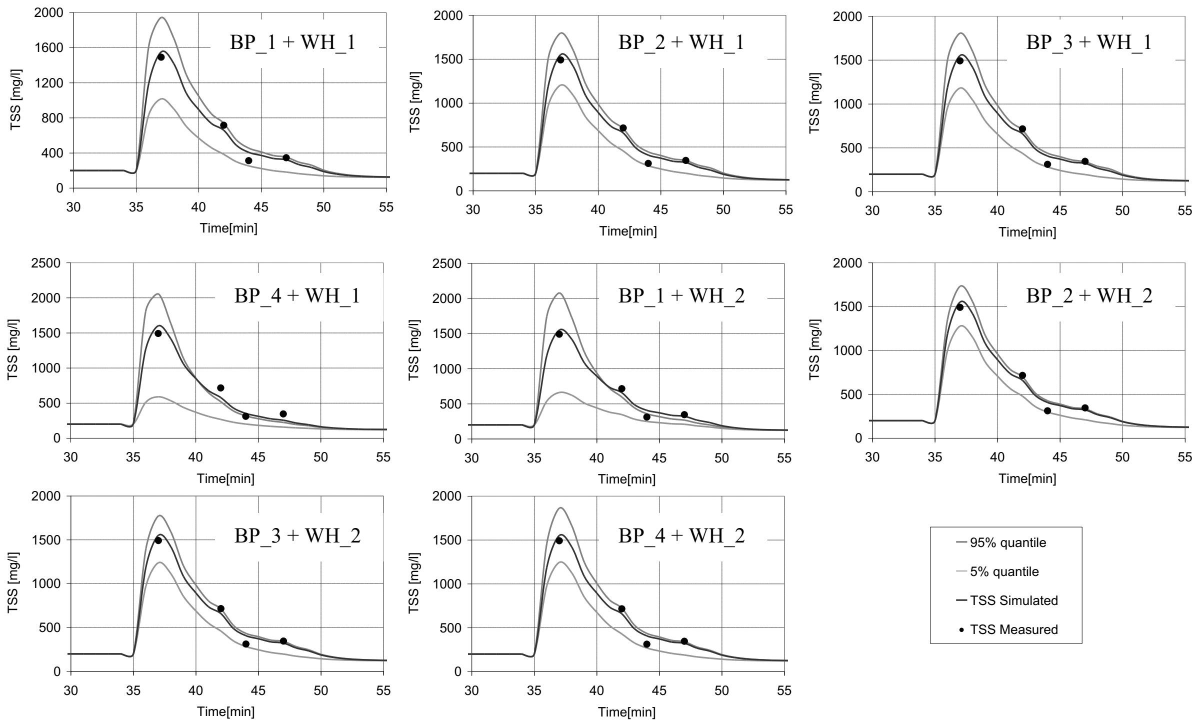

Figure 2 Uncertainty bands near TSS concentrations for the event of 21st August 1997

Table 3 Models and parameter variation ranges used for the BMA

Model	Average uncertainty band width for TSS concentration [mg/l]											
	1	2	3	4	5	6	7	8	9	10	11	12
M_01	543.4	673.5	543.4	652.4	834.2	542.2	632.5	765.4	543.5	657.3	643.2	413.3
M_02	488.2	601.2	511.2	406.5	405.4	354.5	313.4	698.6	531.2	332.3	411.2	365.4
M_03	437.3	566.3	401.2	322.7	311.2	300.2	214.5	665.6	489.7	320.2	326.9	846.1
M_04	645.6	768.7	896.5	554.3	324.3	767.4	876.3	564.3	675.3	875.3	923.1	513.2
M_05	603.4	834.2	753.2	611.2	743.2	456.3	543.4	534.2	769.3	564.3	1002.4	420.5
M_06	223.5	324.5	356.4	453.8	275.2	109.2	265.4	278.6	365.5	311.2	217.3	163.2
M_07	214.8	254.3	403.4	539.4	304.5	543.2	453.9	322	304.3	344.2	301.2	245.6
M_08	238.3	278.5	421.9	512.2	224.5	435.7	510.2	453.9	455.2	403.3	244.7	257.7
BMA	175.68	219.47	238.5	244.89	174.47	90.67	206.86	261.56	241.31	239.09	212.32	128.97

CONCLUSIONS

The paper proposed the application of Bayesian Model Averaging to solve the problem of selecting the most appropriate conceptual model for sewer water quality modelling. Such processes are characterised by a large uncertainty that propagates through the model to the modelling output, and the BMA was demonstrated to be a suitable approach to reduce the uncertainty and provide proper support for the modeller to select the most appropriate approach for a specific case study. The analysis demonstrated that the BMA outperforms the single model application, and the uncertainty is reduced by approximately 1/3 with respect to the best performing single model. Moreover, the BMA solves the problem of identifying the most appropriate model to be used in a specific application, considering that 4 different models could be selected as the best performing out of 12 events. On the other hand, BMA has higher computational requirements because all the candidate models have to be run instead of a single one selected by the modeller but this aspect is highly compensated by the fact that a wrong model selection can take up to 84% larger uncertainty bands than the BMA. Further analysis should be conducted to investigate different BMA strategies to reduce computational costs.

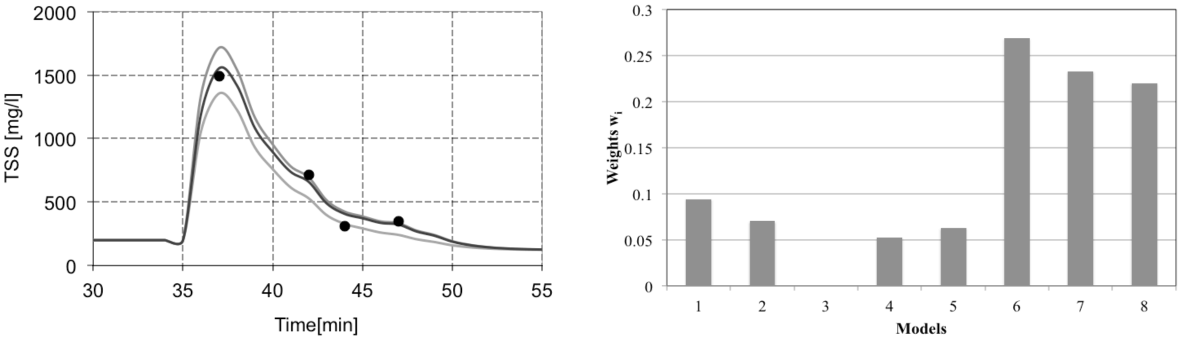

Figure 3 Uncertainty bands after BMA application for the event of 21st August 1997 and the weights of each modelling output in the BMA analysis

REFERENCES

APHA. (1995) Standard Methods for Examination of Water and Wastewater. APHA, AWWA and WPCF, Washington DC, USA

Artina, S., Maglionico, M., Marinelli, A., Raffaelli, G., Anzalone, C., Lanzarini, S. and Guzzinati, E. (1997). "Le misure di qualità nel bacino urbano Fossolo." L'acqua 2, 17-25. (in italian)

Bates, J. M., and C. W. J. Granger (1969), The combination of forecast, Operations Res., 20, 451–468.

Bayes, T. 1763 An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society London 53(A), 370–418.

Bertrand-Krajewski J.-L., Briat P. and Scrivener O. (1993). Sewer sediment production and transport modelling: a literature review. Journal of Hydraulic Research, 31(4), 435-460.

Candela, A., Freni, G., Mannina, G., Viviani, G. (2009) Quantification of diffuse and concentrated pollutant loads at the watershed-scale: An Italian case study. Water Science and Technology 59 (11), pp. 2125-2135

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, Parts A/B/C, 2012 vol. 42-44 pp. 3-10.

Diks, C. G. H., and J. A. Vrugt (2010), Comparison of point forecast accu- racy of model averaging methods in hydrologic applications, Stoch. En- viron. Res. Risk Assess., 24, 809–820, doi:10.1007/s00477-010-0378-z.

Dotto C B S, Mannina G., Kleidorfer M, Vezzaro L., Henrichs M., McCarthy D T, Freni G., Rauch W., and Deletic A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Research, 46(8), 2545–2558. doi:10.1016/j.watres.2012.02.009

Duan, Q., N. K. Ajami., X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, Adv. Water Resour., 30(5), 1371–1386.

Freni G., Mannina G., & Viviani G. (2010a) Urban water quality modelling: A parsimonious holistic approach for a complex real case study. Water Science and Technology, 61 (2) , pp. 521-536
Freni, G., La Loggia G., & Notaro V. (2010b) Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation. Water Science and Tech., 61 (12), 2979-2993
Granger, C. W. J., and R. Ramanathan (1984), Improved methods of combining forecast accuracy, J. Forecasting, 3, 197–204.
Hoetting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian modeling averaging: A tutorial, Stat. Sci., 14(4), 382–417.
Hsu, K., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, Water Resour. Res., 45, W00B12, doi:10.1029/2008WR006824.
Huber W.C. (1986) "Deterministic modeling of urban runoff quality," NATO ASI Series,Vol. G10, Urban Runoff Pollution, Edited by Torno H.C., Marsalek J., Desbordes M., Springler-Verlag Berlin Heidelberg.
Jewell, T.K. and Adrian, D.D. (1978). "SWMM storm water pollutant washoff function." J. Environ. Eng., 104(5), 1036-1040.
Parrish, M. A., Moradkhani, H., & Dechant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. Water Resources Research, 48(3), W03519. doi:10.1029/2011WR011116
Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, Mon. Weather Rev., 113, 1155–1174.
Vrugt, J., and B. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, Water Resour. Res., 43, W01411, doi:10.1029/2005 WR004838.