International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

A semi-distributed modelling approach to support wastewater system management in large urban areas

Juan Pablo Rodríguez[1,2], Neil McIntyre[1], Mario Díaz-Granados[2], Juan Pablo Quijano[2] and Čedo Maksimović[1]
[1]Imperial College London, London, UK; [2]Universidad de los Andes, Bogotá, Colombia (jrodrigu@imperial.ac.uk; pabl-rod@uniandes.edu.co)

Abstract: Models that represent the integrated urban drainage system are needed to holistically assess and manage its performance. Furthermore such models should ideally be spatially distributed to support evaluation of individual components and their role within the system. However, due to data constraints and computational costs, high resolution models of the sewer system are unlikely to be applicable when dealing with large scale and complex systems. This work assesses the value of a relatively simple semi-distributed grey-box modelling approach, which, in comparison with a fully distributed approach, decreases supporting data requirements but still aims to provide a reasonable spatial representation of the sewer system and its temporal variability, thus enabling integrated assessment. New methods for characterising uncertainty in flow and pollution load inputs during dry weather conditions are also used. A large urban catchment in Bogotá (Colombia) is used as a case study. Modelled and observed flow and pollutant concentration data demonstrated high apparently random variability of dry weather flow profiles within the sewer system. Against this variability, the effects of in-sewer processes were not identifiable except where backwaters caused particularly high retention times.

***Keywords**:*
Bogotá's urban drainage; Grey-box modelling; Semi-distributed modelling; Wastewater quality modelling.

1 INTRODUCTION

Successful application of integrated urban drainage models has been recently reported for different case studies [e.g. Vanrolleghem et al., 2005; Devesa et al., 2009; Freni et al., 2009]. However, most of previous research has been applied to relatively small urban areas. Due to lack of application of integrated urban drainage models to large urban centres, there is no total clarity regarding appropriate sewer system model complexity in terms of spatial and temporal resolution, and complexity of process representation. Decisions about which properties of the system and which processes to explicitly represent in the model, which to neglect and which to represent stochastically can be greatly assisted by considering the sensitivity of the required model outputs, in the context of this work the relevant WWTP influent data, to these properties and processes. Previous research has demonstrated that, depending on WWTP configuration and characteristics, short-term fluctuations of a particular determinant may not be important [Langeveld et al., 2003]. This, for example, can reduce the requirements of the quality modelling, since exact calculation of short-term dynamics is unnecessary. Also, if data constraints and/or system variability mean that uncertainty in pollution loads dominates the uncertainty in modelled WWTP influent, then there may be no real value in representing in-sewer processes [e.g. Flamink et al., 2005]. To contribute to better understanding, this paper tests the hypothesis that a simple semi-distributed hydrological grey-box modelling scheme provides an appropriate characterisation of dry weather pollutant dynamics for large and complex sewer systems for the purpose of providing inputs to WWTP design and operation. A tailored version of the model of Achleitner et al. [2007] is applied, together with new methods for characterising uncertainty in flow and pollution load inputs. Such a characterisation, of both gauged and ungauged areas, is supported by means of a large water quality database available from recent monitoring in Bogotá (Colombia). The focus of the sewer system monitoring programme was on flow, biochemical oxygen demand (BOD), chemical oxygen demand (COD) and total suspended solids (TSS) as local environmental regulation focused on them when assessing wastewater discharges and WWTP performance.

2 THE BOGOTÁ CASE STUDY

In general, Colombia has made significant progress towards improving sanitation coverage. However, WWTPs still cover only a fraction of the country, with about 22% of municipalities having partial treatment facilities and only 10% of them performing adequately. In Bogotá and most large urban centres, only 20% to 25% of the wastewater is treated to some extent. For the Bogotá river basin, the National Planning Department recently issued a strategic planning document which called for upgrading and extending - from 4 m^3 s^{-1} to 8 m^3 s^{-1} - the only WWTP in Bogotá (the Salitre WWTP), the construction of a new and larger plant downstream of Bogotá and the expansion of wastewater treatment installations in the upper basin. The focus of this work is on the sewer system connected to the Salitre WWTP, which corresponds to a 150 km^2 urban catchment (also named Salitre) serving nearly 2.5 million inhabitants (about 1/3 of Bogotá's population). The first sewer system developments in this area were combined systems, while the newer systems (since 1965) are separated. The main problems related with the combined system are direct discharges of wastewater without treatment as a consequence of the absence or lack of adequate infrastructure, and combined sewer overflows (CSO) discharges even during dry weather periods. Regarding the separated system, there are many wrong connections that lead to sewage flowing into the storm drainage system and vice versa. As described by Díaz-Granados et al. [2009], Bogotá is composed of around 470 sub-catchments (where nearly 20% are rural or highly pervious, 25% combined sewer and 55% separate sewer sub-catchments). The Salitre catchment is composed of around 180 sub-catchments. The catchment is predominantly residential (residential water consumption corresponds to 84% of the total, with 16% standard deviation in this value between the 180 sub-catchments) with moderate industrial and commercial activities (on average 3% and 11% of the water consumption respectively). Nevertheless, industrial and commercial uses in particular areas reach up to 45% and 54% of the sub-catchment water consumption respectively.

3 METHODS

3.1 Monitoring programme

Díaz-Granados et al. [2009] described recent monitoring efforts covering different components of the Bogotá's urban drainage (i.e. sewer system, WWTP and receiving water courses). As part of this, a database which now covers quantity and quality from about 150 monitoring sites in Bogotá's sewer system was developed. Monitoring activities were carried out at one site at a time during dry weather week days. Field data were collected by the Environmental Engineering Research Centre at the Universidad de los Andes with support from the local water utility (Empresa de Acueducto y Alcantarillado de Bogotá - EAAB) in four monitoring phases: (*i*) 12 sites in 2006, (*ii*) 17 sites in 2007, (*iii*) 82 sites in 2009 and (*iv*) 36 sites in 2011. Over a period of 24 hours, phase (*i*) comprises 24 samples per site while phases (*ii*) to (*iv*) 8 samples per site. 29 monitoring sites characterised single sub-catchment outlets (covering the entire Bogotá system) and are used to estimate dry weather flow (DWF) inputs from both gauged and ungauged sub-catchments in the Salitre catchment, while data from gauged sub-catchments and downstream locations in the analysed sewer system are used for calibration and validation of the model. Observed daily mean BOD, COD and TSS concentrations at the 29 sub-catchment gauged outlets in Bogotá evidenced high spatial variability. 13 gauged sites were selected for validation of the model, they include combined and separate (both wastewater and stormwater systems) sub-catchment outlets, CSOs and main sewer trunks (see Table 1). The 13 sites allow quantification of the model's ability to represent flow and quality at structures such as CSOs and to handle the highly heterogeneous issue of wrong connections.

Table 1. Characteristics of 13 monitoring sites selected for model validation

Monitoring Site	Description	Hydraulic Retention Time (hours)
A	Combined sub-catchment outlet	0.6
B	Main sewer trunk	0.7
C	Separate sub-catchment outlet (Wastewater System)	0.7
D	Separate sub-catchment outlet (Stormwater System)	0.7
E	Main sewer trunk	0.8
F	Separate sub-catchment outlet (Wastewater System)	1.0
G	Separate sub-catchment outlet (Wastewater System)	1.0
H	Combined sewer overflow (CSO)	1.1
I	Combined sewer overflow (CSO)	1.3
J	Main sewer trunk (outlet flowing into WWTP)	2.3
K	Main sewer trunk (outlet flowing into WWTP)	4.1
L	Main sewer trunk	5.6
M	Main sewer trunk (outlet flowing into WWTP)	6.7

Apart from the previously described database, the EAAB collected wastewater samples providing continuous 4 hour time resolution flow, BOD, COD and TSS measurements at different locations in the Salitre catchment. Campaigns carried out in 2009 (2 August to 15 September) and 2010 (4 to 24 February) corresponds primarily to DWF conditions and are used here. Some of the locations coincide with sites reported in Table 1. In this work, 3 of these locations were selected for further model assessment: (*a*) a site downstream E (with a maximum retention time of about 1.8 hours), (*b*) a site downstream L (with nearly the same maximum retention time as site L but collecting additional wastewater contributions) and (*c*) site M. Only data of dry weather week days are used for analysis: 23, 20 and 13 monitored weekdays at (a), (b) and (c) respectively.

3.2 Sub-catchment pollution load model

In this work it is proposed a grey-box model to characterise pollutant loads. Such a model is composed of four sub-models: (*a*) one that estimates daily mean wastewater flows, BOD, COD and TSS concentrations, (*b*) one that characterises pollutants variability around mean values, (*c*) one that quantify pollutant loads that are transferred from the sanitary sewer system into the stormwater system via wrong connections and (*d*) one that represents infiltration flows into the sewer system. Due to a large observed spatial and temporal variability regarding sub-catchment wastewater generation, sub-models (*a*) and (*b*) are composed of two modules, one deterministic and one stochastic.

(a) Daily mean wastewater flows and pollutant concentrations: For the Bogotá case a detailed water consumption database is available. Daily mean wastewater flows for the urban sub-catchments were estimated based on those aggregated consumption rates multiplied by a return factor of 0.84, which is the reported value for Bogotá and the studied area. A step wise regression analysis is used to establish a linear model that relates wastewater average flows and the number and type of users at the sub-catchment level. In order to identify explanatory models of the observed daily mean concentrations, different data analysis techniques were used in this work. The final aim is to identify a relationship between sub-catchment properties (e.g. number and/or percentage of residential, industrial, commercial, office and multi users-related units, time lag, area, mean slope, hydraulic length and population) and the observed BOD, COD and TSS daily mean concentrations over the 29 gauged sub-catchment outflows. These techniques are: stepwise regression analysis and EPR MOGA [Giustolisi and Savic, 2006; 2009].

(b) Daily variability: Once the sub-catchment flow and pollutant concentration mean value and its variability are characterised, probabilistic prediction of the expected diurnal dynamics is needed to be used as inputs for the larger semi-distributed sewer system model. Different methodologies and models have been previously developed for generating DWFs and associated pollutographs [e.g. Butler and Graham, 1995; Langergraber et al., 2008; Gernaey et al., 2011]. However, these methodologies pose some limitations if used for generating input time series for large scale modelling purposes [Rodríguez et al., 2012]. Recently, De Keyser et al. [2010] developed a tool able to model DWFs where detailed information about sources is not available. De Keyser's tool uses normalised diurnal profiles coupled with local population-equivalent data. In general, the diurnal profiles have been commonly represented using polynomials or Fourier series. In this work the diurnal dynamics are represented as a combination of sine and cosine curves (i.e. 3rd order Fourier series) that allows two peaks of concentration and flows per day, with the timing and magnitude of peaks identified from the available measurements. The Fourier series ($\mathrm{X_{i,j}(t)}$, where subscripts i and j refer to a particular site and a particular determinant respectively) may be called a global time series as it aggregates the contributions from all water use components (i.e. the component of $\mathrm{X_{i,j}(t)}$ arising from each unit of residential, industrial, commercial, office-use and multi-user water use denoted as $\mathrm{R_j(t)}$, $\mathrm{I_j(t)}$, $\mathrm{C_j(t)}$, $\mathrm{O_j(t)}$ and $\mathrm{M_j(t)}$ respectively). A method for disaggregating the global series is implemented because the model must be used to estimate the determinants in sub-catchments which have mixtures of water use components that are different from the gauged sites,

$$\mathrm{X_{i,j}(t) = re_i \cdot R_j(t) + in_i \cdot I_j(t) + co_i \cdot C_j(t) + of_i \cdot O_j(t) + mu_i \cdot M_j(t)} \quad (1)$$

where for site i, $\mathrm{re_i}$, $\mathrm{in_i}$, $\mathrm{co_i}$, $\mathrm{of_i}$ and $\mathrm{mu_i}$ are the known number of residential, industrial, commercial, office-use and multi-user units respectively. These five time-series are assumed to be common across all sites and therefore applicable also to ungauged sites. The coefficients are optimised by minimising the sum of the squared differences between $\mathrm{X_{i,j}(t)}$ and those from (1). The optimisation was implemented using the *lsqnonneg* function in Matlab, which constrains the solution to non-negative coefficient estimates.

After applying this disaggregation procedure, one pattern per water use per flow/quality determinant is produced and thus the expected DWF patterns at both partially gauged and ungauged sub-catchments can be estimated. The limitations in both data and the model mean there is significant uncertainty in these expected DWF patterns, the quantification of which requires an error model, identified from the residuals between modelled and measured flows and concentrations. The error model, based on the methods described by Matalas [1967], focuses on representing autocorrelation and inter-determinant cross-correlations. The normalised residuals of flows, BOD, COD and TSS concentrations are found to be approximately described by a multivariate normal distribution and the serial correlation of each determinant can reasonably be described by a first-order linear autoregressive model. Applying this error model for a particular ungauged sub-catchment produces a time series of errors for each of the four determinants which are then superimposed on the corresponding deterministic model output. In order to disaggregate the historical error term into use-related error fractions, necessary to enable the characterisation of errors in ungauged areas, it is assumed that each of the water component units contributes proportionally to the total observed error series. It is then assumed that for each determinant the variance of the error associated with a unit of each water component is uniform over all sites, and that the cross-correlation and autocorrelation terms are also uniform, allowing the error distribution at any ungauged site to be defined. If multiple realizations are simulated, they represent the range of possible time series according to the model. Full details of the estimation of diurnal signals are in Rodríguez et al. [2012].

(c) Wrong connections: Different monitoring campaigns were carried out by EAAB on order to establish percentages of wrong connections from the sanitary into the stormwater system (i.e. number of properties that wrongly connected the wastewater outflow). For the Salitre catchment a global value of 22% was established. Mestra [2009] developed a GIS-based computational tool to identify the likelihood of cross-connections at the property level. Such a tool is used in this work in order to distribute the global value of 22% among system sub-catchments. The main factors which have a relevant effect on the presence of cross-connections are urban densification processes, sewer system age, the physical gap between the storm water and wastewater systems, socioeconomic level, land use, pipe depth and distance between property and the wastewater and storm water systems, pipe material, road type and property type. Mestra's tool uses 8 variables which take into account all these factors. The sum of the variables values is used to qualify the existence of wrong connections in three different ranges as low, medium, and high likelihood. The tool was tested and validated using a sub-catchment named Jaboque located in the Salitre sub-catchment in Bogotá.

(d) Infiltration flows: Regarding the infiltration flow, several studies have shown that the hydrology of urbanized areas can be very complex as the urban environment is highly heterogeneous in terms of land use, subsoil characteristics among other factors, and hence infiltration flows can be difficult to estimate. In this work a pragmatic approach to estimate those flows in dry weather periods is used. Sewer infiltration, which is assumed uniformly distributed over all urban areas ($6x10^{-6}$ $m^3s^{-1}km^{-2}$), was calculated as the difference between (*a*) water supply distribution system leakages (which are about 2.1 m^3s^{-1}) and (*b*) the percolation flowing to underground layers (0.2 m^3s^{-1}) plus the annual evapotranspiration (0.03 m^3s^{-1} = 25% of the real evapotranspiration).

3.3 Sewer system model

Achleitner et al. [2007] developed the City Drain toolbox for urban drainage system modelling. The sewer system model is built within the toolbox using conceptual blocks representing the source, the sub-catchment (both combined and separate systems), the sewer, the wastewater treatment plant and the river. The sewer blocks allow for flow and pollutant routing, although one of the main advantages City Drain is the possibility of modifying the code according to specific needs. A version of City Drain is used for the Bogotá case study to model the main sewer system, from sub-catchment outlets to the inlet of the Salitre WWTP. The resulting model of Salitre includes 180 sub-catchments drained by a number of main sewer pipes, 31 CSOs and 4 wastewater pumping stations. Physical properties and operational data for the CSSs and pumping stations were provided by the EAAB. The sub-catchment is modelled using an hourly time step which considered sufficient to capture the DWF dynamics.

4 RESULTS AND DISCUSSION

This section looks at two different spatial scales: sub-catchment and full-catchment scales. The first sub-section intends to assess if linear and/or non-linear models are able to characterise daily mean flows and concentrations of BOD, COD and TSS at the sub-catchment outlets. The second sub-section aims to quantify the performance of a simple hydrological semi-distributed model when replicating observed flows and pollutant concentrations within a large and complex sewer system (full catchment scale).

4.1 Sub-catchment wastewater outflows characterisation

Applying a stepwise regression analysis, as proposed in methods section, it was found (at the 95% significance level) that the sub-catchment outlet daily mean flow rate (in m^3s^{-1}) is linearly related to sub-catchment population, the number of residential, industrial, commercial, office-use and multi-use users with an R2 value of 0.89 (2). However, linear models were incapable of usefully characterising mean concentrations, with optimised R2 values of 0.14, 0.11 and 0.35 for BOD, COD and TSS respectively. Therefore the EPR MOGA-XL software [Giustolisi and Savic, 2009] was used to explore if nonlinear models perform better. The expressions in (3) and (4) were identified for sub-catchment outflow BOD and TSS daily mean concentrations (in mg l^{-1}). (3) characterises observed BOD daily mean concentrations with an R2 of 0.58 while (4) characterises observed TSS with an R2 of 0.61. COD is modelled using the linear relationship between BOD and COD as presented in (5) (no strong linear relationship exists for the couples BOD-TSS or COD-TSS). These models could be improved if additional data become available. However, they are considered the best model that can be identified from the currently available datasets.

$$\mathrm{Q} = 4.38\mathrm{x}10^{-6} \cdot \mathrm{R} + 5.71\mathrm{x}10^{-5} \cdot \mathrm{I} + 8.60\mathrm{x}10^{-6} \cdot \mathrm{C} + 0.00027 \cdot \mathrm{O} + 6.95\mathrm{x}10^{-6} \cdot \mathrm{M} \quad (2)$$

$$\mathrm{BOD} = 0.0048 \cdot \mathrm{r}^2 \cdot \mathrm{i}^{0.5} \cdot \mathrm{m}^{0.5} + 23992\,\frac{\mathrm{R}^{0.5} \cdot \mathrm{m}^{0.5} \cdot \mathrm{L}^{0.5}}{\mathrm{H}^{1.5}} + 1.51\,\frac{\mathrm{R}^{0.5} \cdot \mathrm{r}}{\mathrm{K}^{0.5}} \quad (3)$$

$$\mathrm{TSS} = 6.33 \cdot \frac{\mathrm{i}'^{0.5} \cdot (\mathrm{K})^{1.5} \cdot \mathrm{L}}{\mathrm{C}^2} + 0.00074 \cdot \mathrm{I} \cdot \mathrm{C} + 2.33 \cdot \mathrm{R}^{0.5} \quad (4)$$

$$\mathrm{CDO} = 2.05 \cdot \mathrm{BOD} \quad (5)$$

where r, i and m are the percentage of residential, industrial and multi-users units; L is the sub-catchment hydraulic length (in m); i′ is the percentage of industrial water use consumption; R, I, C, O and M are the number of residential, industrial, commercial, office-use and multi users respectively; K is the sub-catchment travel time (in seconds) and H is the number of inhabitants. The performances of (2) to (5) for other case studies are unknown and further research is needed to identify if the significant explanatory variables vary from one sewer system to the next. For this, averaging EPR-based models over multiple sewer systems has proven to be useful. In this work it is proposed to use (2) to (5) as the deterministic component of the models to estimate daily mean values, this coupled with a stochastic term based on observed models residuals.

4.2 Catchment scale model assessment

Assessment analysis firstly compared observed and modelled daily variability of pollutant concentrations at two different sites, downstream L and M, with relatively large hydraulic retention times of 5.6 and 6.7 hours respectively and where the effect of potential in-sewer process are more likely to be identifiable. Figure 1 presents results from 1000 model simulations (i.e. 1000 realisations of input flow and concentrations from the stochastic generator described in section 3.2), and their comparison to the observed values for COD (similar results were obtained for BOD and TSS). Figure 1 shows high variability of both the observed and modelled profiles. If the stochastic generator and the model are correct, and the observations do not suffer from significant bias, then the observations should appear as samples from the simulated distributions. This is impossible to evaluate with confidence because of the small number of observations at each time-step, however the plots can be used to subjectively identify where there is a major discrepancy, for example where observations consistently lie above or below the simulated confidence intervals. The latter is the case of site M (see Figure 1b). It is relevant to mention that there is a 600 m open channel from M to the WWTP that provides additional retention capacity depending on the plant pumping operation. Recall that the Salitre WWTP has a capacity of 4 m^3s^{-1}, consequently from approximately 10 am produced wastewater exceeds WWTP treatment capacity thus triggering backwater phenomena during DWF conditions. This has as a consequence an increased retention time (over the expected 6.7 hours) during about 60% of the time. Figure 1b clearly shows that when treatment capacity is exceeded, transformation and sedimentation processes become relevant; while when capacity is not a limiting factor, high loads are observed likely due to re-suspension of deposited material. The proposed model does not perform well at this site as these complex processes are not represented.

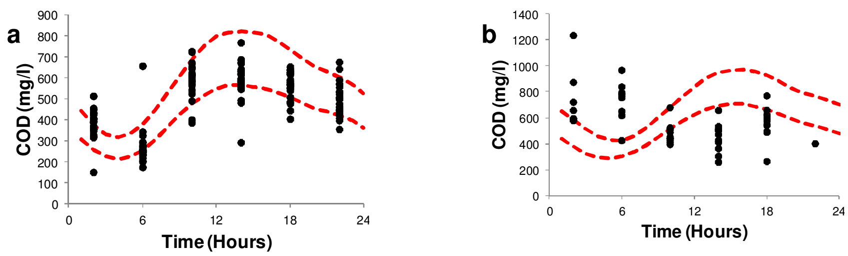

Figure 1. Comparison of observed and modelled (dashed lines corresponding to 95% confidence bounds) COD daily profiles. (a) downstream site L and (b) site M

Additional analyses included comparison of observed and modelled standard deviation, maximum and mean of pollutant concentrations and loads at sites summarized in Table 1. Loads are presented only for those sites where reliable flow and pollutant concentrations are simultaneously available. Figure 2 presents a summary of results from the same 1000 model simulations used above, and their comparison to the observed COD values (similar results were obtained for BOD and TSS). It can be concluded that model signals show high variability but most of the sites, including combined, sanitary and storm sewer sub-catchment outlets, CSO and main sewer trunks, were in general well represented by the model. It is again noticed the relatively poor characterisation of site M due to increased hydraulic retention time.

5 CONCLUSIONS

Urban water quantity and quality models are useful to support: performance assessment and definition of required upgrades to the existing urban drainage system and optimal planning and design of new systems. The main aim of these models is to quantify the efficiency of different measures in reducing the amount of pollutants discharged into receiving water bodies and minimise the consequent negative impacts on water quality. However, identifying a parsimonious modelling framework to represent large and complex sewer systems is a challenging task that has to take into account the expected uses of model results and limited knowledge and observations of the system. The inclusion of processes that do not significantly contribute to the important outputs results in a model that is unnecessarily difficult to use and has unnecessary data demands. Implementation of detailed distributed sewer models at large spatial scales is currently limited due to data availability requirements. Alternatively, it is possible to use semi-distributed models to represent the sewer system. This work demonstrated that a relatively simple semi-distributed hydrological modelling approach is, in general, able to support evaluation of individual components (i.e. sanitary and storm sewer sub-catchments, CSO, main sewer trunks) and their role within the system under dry weather conditions. This is supported through the following main conclusions: (*a*) there is high apparently random variability of dry weather flow profiles within the sewer system; (*b*) against this variability, the effects of in-sewer processes were not identifiable expect where backwaters caused particularly high retention times; (c) the reasonable success with which the model represents such variability indicates the potential of a simple water quality model approach when quantifiable uncertainty is accounted for.

REFERENCES

Achleitner, S., M. Möderl and W. Rauch, CITY DRAIN © - An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling & Software* 22(8), 1184-1195, 2007.

Butler, D. and N. J. D. Graham (1995). Modeling dry weather flow in sewer networks. *Journal of Environmental Engineering* 121(2), 161-173.

De Keyser, W., V. Gevaert, F. Verdonck, B. De Baets and L. Benedetti (2010). An emission time series generator for pollutant release modelling in urban areas. *Environmental Modelling & Software* 25(4), 554-561.

Devesa, F., J. Comas, C. Turon, A. Freixó, F. Carrasco and M. Poch, Scenario analysis for the role of sanitation infrastructures in integrated urban wastewater management. *Environmental Modelling & Software* 24(3), 371-380, 2009.

Díaz-Granados, M. A., J. P. Rodríguez, M. S. Rodríguez, J. C. Penagos, L. A. Camacho, S. Achleitner, Č. Maksimović and N. McIntyre, Towards a paradigm shift in urban drainage management and modelling in developing countries. *Revista de Ingeniería – Universidad de los Andes* 30, 133-150, 2009.

Flamink, C., J. Langeveld and F. Clemens, Aerobic transformations in sewer systems: are they relevant?. *Water Science and Technology* 52(3), 163-170, 2005.

Freni, G., G. Mannina and G. Viviani, Assessment of data availability influence on integrated urban drainage modelling uncertainty. *Environmental Modelling & Software* 24 (10), 1171-1181, 2009.

Gernaey, K. V., X. Flores-Alsina, C. Rosen, L. Benedetti and U. Jeppsson (2011). Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. *Environmental Modelling & Software* 26 (11), 1255-1267.

Giustolisi, O. and D. A. Savic, A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics* 8(3), 207-222, 2006.

Langergraber, G., J. Alex, N. Weissenbacher, D. Woerner, M. Ahnert, T. Frehmann, N. Halft, I. Hobus, M. Plattes, V. Spering and S. Winkler (2008). Generation of diurnal variation for influent data for dynamic simulation. *Water science and technology* 57(9), 1483-1486.

Langeveld, J. G., F. H. L. R. Clemens and J. H. J. M. van der Graaf, Interactions within the wastewater system: requirements for sewer processes modelling. *Water Science and Technology* 47(4), 101-108, 2003.

Matalas, N. C., Mathematical assessment of synthetic hydrology. *Water Resources Research* 3(4), 937-945, 1967.

Mestra, G. L., Identificación de conexiones erradas en alcantarillados separados mediante la aplicación de herramientas computacionales. M.Sc Thesis, Universidad de los Andes(Colombia), 2008.

Rodríguez, J. P., N. McIntyre, M. A. Díaz-Granados, S. Achleitner, M. Hochedlinger and Č. Maksimović, Characterising sub-catchment wastewater outflows using a stochastic time series generator. *Environmental Modelling & Software* (In review), Re-submitted 2012.

Vanrolleghem, P. A., L. Benedetti, J. Meirlaen, Modelling and real-time control of the integrated urban wastewater system. *Environmental Modelling & Software* 20 (4), 427-442, 2005.

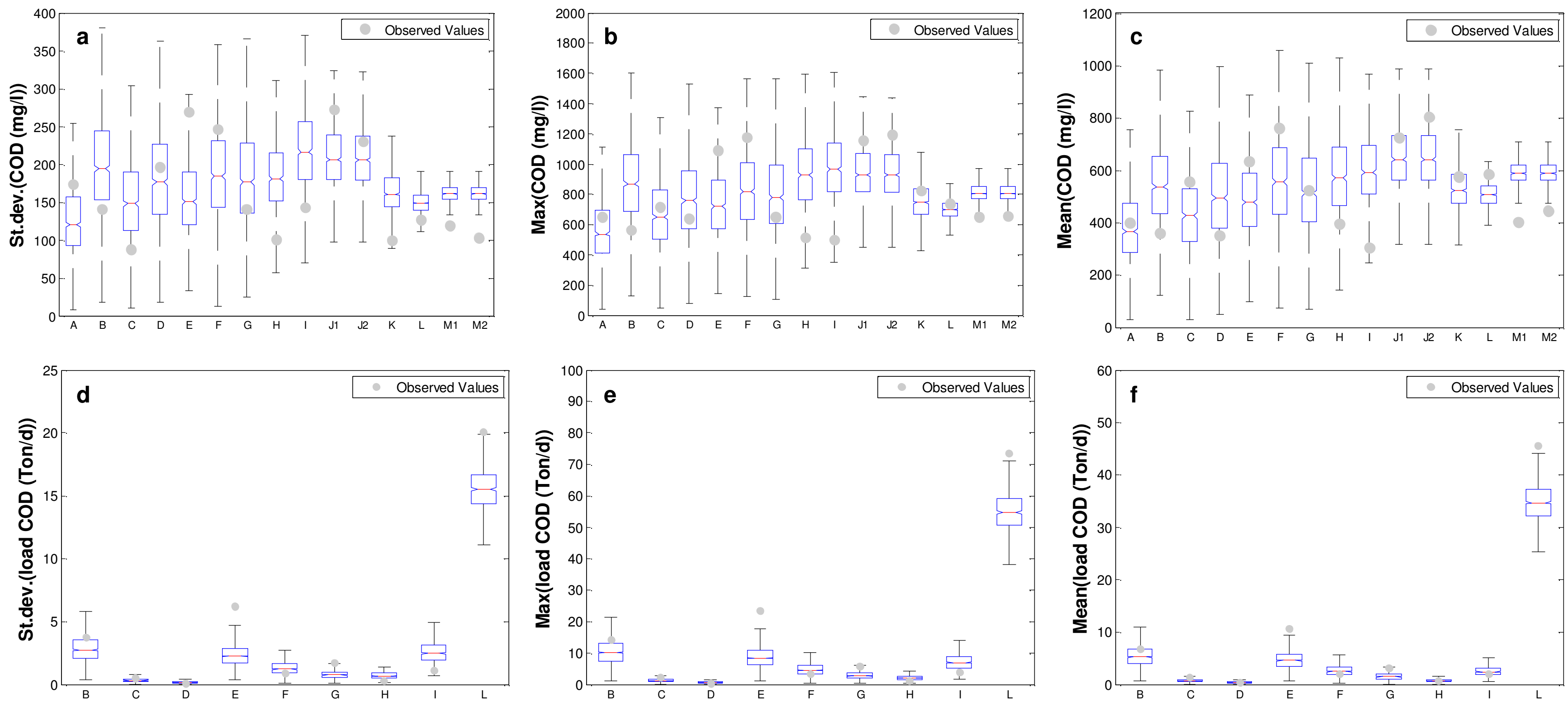

Figure 2. Comparison of observed and modelled COD concentration standard deviation, maximum, and mean values, and observed and modelled COD load standard deviation, maximum and mean values (box plots represent 95% intervals from 1000 model realisations