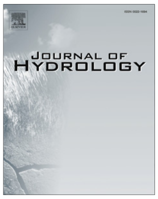
Journal of Hydrology 517 (2014) 997–1007

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data

Guido Petrucci [a,b,*], Céline Bonhomme [b]

[a] Vrije Universiteit Brussel (VUB), Earth System Sciences (ESSc), Brussels, Belgium
[b] Université Paris-Est, Laboratoire Environnement Eau Systèmes Urbains (UMR MA102), UPEC, UPEMLV, ENPC, AgroParisTech, 6-8 Avenue Blaise Pascal, 77455 Champs sur Marne cedex 2, France

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form 11 June 2014
Accepted 13 June 2014
Available online 27 June 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Purna Chandra Nayak, Associate Editor

Keywords:
Semi-distributed models
Urban drainage
Spatial distribution
Geographical data
Suspended solids

SUMMARY

Semi-distributed models are widely used in urban hydrology, supported by the abundance and detail of geographical data. The inclusion of these data into hydrological models requires however an increasing complexity of the model structure with spatially distributed parameters, potentially driving to over-parameterisation issues. In this paper, different configurations and model structures, including an increasing quantity of geographical information, are tested for both water quantity and water quality on the widely used SWMM5 model for a 2.3 km^2 catchment. The Nash criterion is used to calibrate the model and compare alternative configurations. Results for water quantity modelling show that the inclusion of some basic geographical information, particularly on land uses, clearly improves performances, but further refinements are less effective. Uncalibrated models with sufficient land use information reach performances comparable with those of calibrated models. For water quality modelling (suspended solids concentration), the best modelling performance is obtained by a compromise solution with moderate spatial distribution of parameters: no spatial distribution drives to limited performances, while an excessive one to severe over-parameterisation. A comparison to suspended solids measurements realized on a single road of the catchment shows that parameters providing good performances at the catchment scale are a realistic, although non optimal, representation of local scale processes.

1. Introduction

Nowadays many researchers and practitioners studying and managing stormwater sewer systems in urban catchments use semi-distributed models like SWMM, CANOE or MOUSE (Zoppou, 2001; Elliott and Trowsdale, 2007). These models are based on a description of the catchment as a set of subcatchments linked by a drainage network. The runoff and pollution generation processes are simulated for each subcatchment, and the network is used to simulate the routing of water and pollutants to the catchment outlet. Subcatchments are represented by conceptual models like non-linear reservoirs. Water quality is described by build-up and wash-off equations, sometimes adjusted for each subcatchment. Flow routing is usually based on hydraulic models. The term "semi-distributed" used for these models refers to the hybrid approach to the catchment description: on the one hand, the catchment is considered in its geographical characteristics and spatial distribution through the division in subcatchments and the physical description of the drainage network; on the other hand, spatial distribution stops at the level of subcatchments, described by conceptual lumped models. The main reason for using semi-distributed models is the necessity to predict the spatial distribution of hydrological variables within a catchment (Reggiani and Schellekens, 2005). In urban hydrology, this necessity is traditionally linked to the needs of managing artificial drainage systems: planning and management of drainage and sewer networks require knowing not only the total quantity of water and contaminants flowing through the system, but also the location of the inflows.

The use of semi-distributed models, however common, raises some questions and critics.

Recently, different works focusing on the good representation of surface flow paths have been published (Gironas et al., 2010; Jankowfsky et al., 2011). In fact, surface flows are crucial to model water quantity and quality and, today, topographic GIS data are

---

[*] Corresponding author at: Vrije Universiteit Brussel (VUB), Earth System Sciences (ESSc), Brussels, Belgium.
E-mail address: guido.petrucci@vub.ac.be (G. Petrucci).

http://dx.doi.org/10.1016/j.jhydrol.2014.06.019
0022-1694/© 2014 Elsevier B.V. All rights reserved.

easily accessible to a broad public (European Union, 2007; Rodriguez et al., 2003). Researchers and practitioners are, thus, encouraged to implement as much information as possible in their hydrological model to improve their outputs. But the reality remains more complex. The way these data are introduced in the model and the structure of the model itself may be crucial for determining the advantages provided by more detailed data. Because of the conceptual modelling of the subcatchments, several parameters in semi-distributed models are "effective" parameters representing a global hydrological behaviour, and cannot be directly measured or linked to geographical data. In addition, some parameters theoretically measurable are not measurable in practice with the necessary level of detail, because of costs and various technical difficulties (Siriwardene and Perera, 2006). It is the case, for instance, of infiltration parameters, initial losses or Manning roughness. Thus, it is not clear how a detailed representation of flow paths can improve simulations based on partially conceptual models.

The use of geographical data is also strictly linked to the issue of calibration. By a pragmatic point of view, the modeller must find a balance between two opposite needs: on the one hand, he has to increase the complexity of the model to get the most of available spatially distributed data; on the other hand, an increased complexity means a higher number of parameters in the model, with the risk of over-parameterisation (a similar dilemma was described by Kuczera and Mroczkowski, 1998).

In fact, the presence in the model of non-measurable parameters makes necessary a calibration process to obtain a good fit between simulations and observed hydrological data. If the number of parameters is large, the problem of "over-parameterisation" can emerge: when too many parameters are calibrated on little information, they remain largely undefined, and the resulting calibrated model shows poor predictive capabilities. Previous studies showed that the "information content" of rainfall/runoff time series allows the calibration of up to four parameters (Jakeman and Hornberger, 1993) and that a similar limitation exists for water quality measurements (Gaume et al., 1998). Semi-distributed models, where parameters are repeated for each subcatchment, easily involve tens to hundreds parameters (Kirchner, 2006; Muleta and Nicklow, 2005; van Griensven et al., 2006). On these bases, it seems inevitable that semi-distributed models lead to over-parameterisation. However, three arguments suggest that the question of over-parameterisation should be further analyzed.

The first argument is that, nowadays, rainfall, flow-rate, turbidity and other water quality measurements are increasingly available as long time series, with durations of months to years, and are measured at high frequency, typically of 1–10 min. The "information content" of these series appears to be much higher than that of the daily (Jakeman and Hornberger, 1993; van Griensven et al., 2006) or event-based (Gaume et al., 1998) series available in the past (Obropta and Kardos, 2007).

The second argument is that semi-distributed models, as discussed above, make an intensive use of topographical, physical and other data (e.g. land uses), constituting supplementary sources of information. These various data can be addressed, in this context, as "non-hydrological" data, to distinguish them from measurements of hydrological variables like rainfall, flow-rate or turbidity that can be grouped under the class of "hydrological data". All the cited studies about the number of parameters allowed by hydrological series were realized on lumped models using only hydrological data. Therefore, they did not take into account that, prior to calibration, a model can already contain a huge quantity of information on the catchment. This is recognized also by Jakeman and Hornberger (1993), considering that the inclusion of data on "physical catchment descriptors" (i.e. non-hydrological data) can increase the number of parameters allowed in rainfall/runoff models and reduce the "over-parameterisation threshold". In many cases, often coming from operational applications, semi-distributed models achieve satisfactory performances through a large use of geographical data and a simple manual calibration on rainfall-runoff series. This empirical evidence suggests that non-hydrological data can provide, at least in some cases, most of the information content required by the model, the calibration on hydrological data providing only a fine tuning. These arguments support the relevance of non-hydrological data but, to the authors' knowledge, no satisfactory attempt to go further on this issue has yet been realized. In comparison with the "simple" case of hydrologic data alone, the combined use of hydrological and non-hydrological data makes extremely difficult to understand the relationship between the information content and the maximum number of parameters allowed.

The third argument is that, in the hydrological community, two different points of view subsist about over-parameterisation: on the one hand, over-parameterisation is regarded as a flaw that, causing uncertainty in parameter determination, casts doubts on the model reliability and robustness (e.g. Perrin et al., 2001). On the other hand, over-parameterisation is considered as a consequence of the fact that a single natural process can have several different acceptable descriptions. This point of view (Beven, 2006) drove to the development of methodologies like the Generalised likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992) that combine the predictions of several different models and parameter sets, under the condition that their predictive capabilities are acceptable. It is clearly difficult to adopt the first point of view to investigate semi-distributed models, as it logically leads to dismiss these models in favour of more parsimonious ones. The second point of view, on the contrary, suggests exploring which modelling choices are more likely to produce satisfactory predictions, disregarding complexity as long as it does not limit performances.

These three arguments show how the question of the complexity allowed by rainfall/runoff models is still open. This work examines this question for semi-distributed models (i) on the basis of continuous high-frequency time-series, (ii) studying the effect of an increasing use of non-hydrological data and (iii) using model performance as the only criterion to evaluate models, disregarding parsimony.

Urban water quality models use as input variables outputs of rainfall/runoff models and rely for pollutants transport on their flow routing modules. Because of this strong dependency, water quality models share the questions mentioned above. Moreover, they present specific issues about the predictive capability of the fundamental modelling schemes used. Almost all models are based on the classical build-up and wash-off equations (Sartor et al., 1974). This approach, developed at the small-scale for simple urban surfaces (single roads or roofs) is highly debated (Bertrand-Krajewski, 2007).

Recent small-scale studies, although suggesting minor changes to the historical formulations, generally confirm the overall validity of classical build-up/wash-off models (Egodawatta et al., 2007, 2009; Wicke et al., 2012). On the contrary, catchment scale studies are extremely critic: model parameters are correlated, troubling calibration, and the predictive capability of lumped models is low (Kanso et al., 2005; Vezzaro and Mikkelsen, 2012). Most researches on the build-up/wash-off model at the catchment scale obtained, actually, poor results: Dotto et al. (2010), studying the uncertainty of this model on Suspended Solids (SS), obtain Nash values between 0.06 and 0.46 in calibration, and consider superfluous to proceed to a validation that would be *a priori* unsatisfactory; Kleidorfer et al. (2009) obtained Nash values in calibration up to 0.45; Dotto et al. (2012), up to 0.04. A common suggestion by these authors, coherent with the small-scale validity of the model, is that