Value of distributed water level and soil moisture data in the evaluation of a distributed hydrological model: Application to the PUMMA model in the Mercier catchment (6.6 km 2) in France

M. Fuamba, F. Branger, Isabelle Braud, E. Batchabani, P. Sanzana, B. Sarrazin, S. Jankowfsky

Journal of Hydrology 569 (2019) 753–770

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Value of distributed water level and soil moisture data in the evaluation of a distributed hydrological model: Application to the PUMMA model in the Mercier catchment (6.6 km^2) in France

Musandji Fuamba[a,*], Flora Branger[b], Isabelle Braud[b], Essoyeke Batchabani[a], Pedro Sanzana[c], Benoit Sarrazin[d], Sonja Jankowfsky[e]

[a] POLYTECHNIQUE MONTRÉAL, Montréal, Québec, Canada
[b] IRSTEA, UR RiverLy, Centre de Lyon-Villeurbanne, 69625 Villeurbanne, France
[c] Departamento de Ingeniería Hidráulica y Ambiental, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE, Santiago, Chile
[d] ISARA-LYON, Lyon, France
[e] RMS, San Francisco Bay Area, USA

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Venkat Lakshmi, Associate Editor

Keywords:
Distributed hydrological modelling
Object-oriented modelling
Model diagnostic
Soil moisture
Water level sensor network
Hydrological signatures
Semi-rural catchment

ABSTRACT

This paper emphasizes the importance of integrating outlet discharge and observed internal variables in the evaluation of distributed hydrological models outputs. It proposes a general methodology for a diagnostic evaluation of a complex distributed hydrological model, based on discharge data at the outlet and additional distributed information such as water level and surface soil moisture data. The proposed methodology is illustrated using the PUMMA model in the Mercier sub-catchment (6.6 km^2). Model parameters are specified according to field data and a previous study performed in a neighbouring catchment (Jankowfsky et al., 2014), without calibration. The distributed water level and soil moisture network of sensors were useful in the model evaluation process. Thus, model parameters are specified either using in situ information or results from previous studies. A stepwise approach is used for model evaluation. It includes standard water balance assessment as well as comparison of observed and simulated outlet discharge, whether on annual or event timescales. Soil moisture sensors are used to assess the ability of the model to simulate seasonal water storage dynamics based on a normalized index. The water level sensors network is used on two timescales: on a seasonal timescale, sensors network is used to assess the model's ability to simulate intermittency; whereas on event timescales, sensors network is used in determining the model's ability to reproduce observed reaction as well as response times. Event timescales do also focus on the correlation between hydrological response and either rainfall event or antecedent soil moisture variables. Results show that the non-calibrated model is quite effective at capturing water flow and soil water-storage dynamics, but it fails to reproduce observed runoff volume during events. There is strong indication of a deficiency in the characterization of catchment storage and upstream flowpath description. The soil water content and a network of water level sensors provide interesting information about soil moisture and river flow dynamics. They however fail to provide quantitative information about catchment storage. This study opens interesting perspectives for the evaluation of distributed hydrological models using hydrological signatures. Furthermore, it highlights the requirement of quantitative as well as qualitative signatures for improving such models.

1. Introduction

Distributed hydrological models are valuable in our understanding of hydrological processes and watershed behaviour. They provide explicit representations of processes and components involved in hydrological balance, and are able to integrate the natural variability of soil and sub-soil, the spatial organisation of landscape and water management practices. As such, they are powerful hypothesis-testing tools, as they formalize our assumptions on the main drivers of catchment response, and allow for testing their relevance against observed data (Gupta et al., 2008; Fenicia et al, 2008; Clark et al., 2011).

In this hypothesis-testing framework, it is now well established that

* Corresponding author at: Polytechnique Montréal, C.P. 6079, succ. Centre-ville, Montréal, Québec H3C 3A7, Canada.
E-mail address: musandji.fuamba@polymtl.ca (M. Fuamba).

https://doi.org/10.1016/j.jhydrol.2018.12.035
Received 30 May 2017; Received in revised form 4 December 2018; Accepted 21 December 2018
Available online 04 January 2019

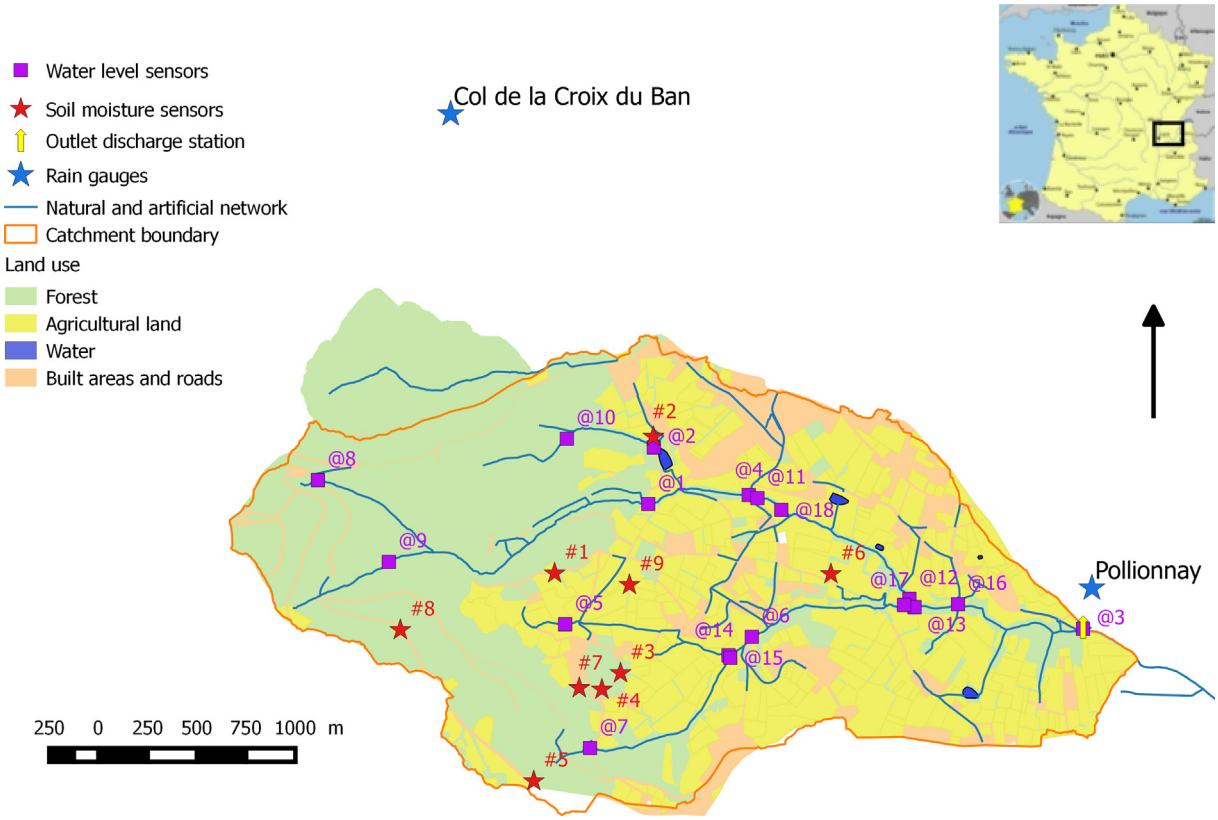

Fig. 1. Location of the study area and simplified land use map, drainage network (natural and artificial), rain, discharge, water level and soil moisture measurements stations in the Mercier catchment.

distributed models offer far better results when their outputs are compared not only to the outlet discharge, but also to observed internal variables (Hrachowitz et al., 2013, Smith et al., 2013). This ensures more robust representations of the catchment's internal process dynamics and improved internal models consistency, thus providing "the right answers for the right reasons" (Klemes, 1986; Grayson et al., 1992; Kirchner, 2006). Several studies show the importance of using internal discharge gauges for model calibration and evaluation (e.g.: Uhlenbrook and Sieber, 2005; Moussa et al., 2007). However, the use of additional data sources, although called for in many papers (Hrachowitz et al., 2013), is not so widespread.

A way of using such additional data is to incorporate them in model calibration process using data assimilation techniques, as illustrated by Pereira-Cardenal et al. (2011), Paiva et al. (2013) or Wanders et al. (2014), who used remote sensing data in large catchments. Another way is to incorporate such data in a diagnostic approach as defined by Gupta et al. (2008). It consists of comparing the model outputs with alternative indicators. Known as "hydrological signatures", these alternative indicators describe the behaviour of the catchment hydrological response. Hydrological signatures are derived from field data and provide physically interpretable metrics for functional behaviour of catchments (McMillan et al., 2011, 2014), as opposed to traditional performance criteria (Nash Sutcliffe efficiency, typically) that are too restrictive to fully describe how well a model represents the different aspects of hydrological response. The advantage of hydrological signatures is also that they can point toward the causes of (model) bad performance and provide elements to improve it (Gupta et al., 2008). In the recent years, some authors recommended to perform diagnostic evaluation on an uncalibrated model (e.g. McMillan et al., 2016). They argued that, although uncalibrated models have generally lower performance scores than calibrated models, they offer several advantages: explicit link between the parameter values and the physics; no compensations for biases in input forcing data (typically rainfall) or errors in model structure (Clark et al., 2011). Therefore, uncalibrated approach seems particularly appropriate for diagnostic evaluation of a model's strengths and weaknesses and to assess the value of qualitative distributed observations for such a diagnostic evaluation.

The main objective of this study is thus to build a diagnostic evaluation of a complex distributed model, based on additional distributed data. Besides outlet discharge, two types of additional data were used: surface soil moisture and water level in ephemeral streams. All these data had been produced by previous research projects and were reused opportunistically for the present study. This main general objective is specified into three sub-objectives that are:

1. To assess usability of additional data in a model evaluation process, and to assess strengths and weaknesses of corresponding hydrological signatures. These data cannot be compared directly to model outputs, either because their quality is not good enough (semi-qualitative data) or because they do not have adequate spatial/temporal resolution. So defining appropriate hydrological signatures from this data is a challenging task.
2. To evaluate all components of the used model (surface runoff, infiltration in the soil, groundwater flow, routing in the stream network), or specific catchment zones (specific sub-catchments, given soil type/land use).
3. To test the model's underlying, functioning hypotheses, as the last part of the model diagnostic. This would allow prioritizing any further developments of the model.

The case study used in this paper is the Mercier catchment (6.6 km^2), located near Lyon, France and the considered model is the distributed *peri*-urban model PUMMA, Peri-Urban Model for landscape Management, (Jankowfsky et al., 2014). PUMMA was used in this study because its level of detail is well adapted to the size of the Mercier catchment and to the resolution of the additional data. It also

incorporates urban hydrology processes that are of importance on the Mercier catchment. However, the methodology proposed in the paper is quite general and could be used for other case studies and/or distributed hydrological models. The paper is organised as follows. In Section 2, study site, available data, modelling tool and model setup are presented. Section 3 presents the model evaluation protocol that was built and the signatures derived from the distributed soil moisture and water level data. Results are presented in Section 4 and discussed in Section 5. Finally, conclusions and perspectives are given in Section 6.

2. Case study, model description and set up

2.1. Mercier catchment and available data

The Mercier catchment is part of the Yzeron basin, located in the south west of Lyon, France (Fig. 1). The Yzeron is representative of periurban areas with a highly heterogeneous land use (Jacqueminet et al., 2013), dominated upstream by forests and agriculture, and increasing urbanization downstream. The Mercier semi-rural sub-catchment has a surface of 6.6 km^2 with altitudes ranging from 300 m to 785 m. Its geology consists mainly of gneiss and granite. The soils are quite shallow, especially in upslope areas, leading overall to low water storage capacity. Land use is 46% agricultural, 40% forests and 14% urban and impervious areas (Braud et al., 2013). The climate is temperate with continental and Mediterranean influences. The average annual precipitation between 1997 and 2010 was 741 mm with a standard deviation of 156 mm. The average response time of the Mercier catchment varies between 1 and 3–4 h. The catchment is part of the *Observatoire de Terrain en Hydrologie Urbaine* (OTHU) and has been monitored since 1997.

2.1.1. Permanent data

Consists of rainfall measurements at two gauges (one starting in 1997 and the other in 2005) and discharge measurement at the catchment outlet (since 1997). These data are available with a variable timestep. Reference evapotranspiration at a hourly timestep was calculated from the national SAFRAN database (Quintana-Seguí et al., 2008; Vidal et al., 2010) by Braud et al. (2013), using the Penman-Monteith equation (FAO, 1998).

2.1.2. Additional data from the networks of sensors

These data come from previous research projects conducted independently from the present study and in limited timeframes. We opportunistically reused the existing data without possibility of changing the experimental setup since all sensors were already removed from the field.

The network of water level sensors was installed for a PhD thesis, which aimed at documenting the activation of ephemeral streams during events and its links with topography (Sarrazin, 2012). Distributed series of water levels are available with a 2–5 min timestep (17 measurement points located in permanent or ephemeral streams, Sarrazin, 2012, see location in Fig. 1) from 2007 to 2010. Sensors were located throughout the Mercier catchment, with upstream areas ranging from 0.06 km^2 to 6.6 km^2 and contrasted land uses. The main characteristics are presented in Table 1. Since sensor accuracy (± 1.6 cm obtained by comparing the sensors and manual field measurements, Sarrazin, 2012) is of the same order of magnitude as the measured water levels, continuous series cannot be used directly for model evaluation. However, simple flow/no-flow patterns (for continuous periods), as well as water level variations and response during rainfall-runoff events can all be extracted (Sarrazin, 2012).

Surface soil moisture (0–5 cm depth layer) data were installed for a project monitoring surface runoff (Dehotin et al., 2015) during one year. Data are available at 9 locations with a 2 min timestep from April 2010 to April 2011. This dataset is described in Dehotin et al. (2015). Location of the sensors is shown in Fig. 1.

Table 1
Main characteristics of the upstream catchments of the water level sensors. The reach length includes the natural and artificial channels (ditches, sewer pipes) connected to the stream network that is represented in the model. Upstream areas and land use characteristics are taken from Sarrazin (2012). Upstream stations are highlighted in bold.

Sensor	Upstream area (km^2)	Reach length (m)	Land use Agriculture (%)	Land use Forest (%)	Land use Residential (%)	Land use Roads (%)
@1	1,46	2000	8	91	0	4
@2	**0,65**	**950**	**21**	**68**	**8**	**2**
@3	6,6	4500	46	40	10	4
@4	2,45	2500	18	74	5	3
@5	**0,21**	**300**	**38**	**57**	**2**	**3**
@6	1,75	1450	53	36	8	4
@7	**0,20**	**200**	**25**	**55**	**16**	**6**
@8	**0,11**	**250**	**6**	**87**	**0**	**7**
@9	**0,34**	**500**	**2**	**94**	**0**	**4**
@10	**0,23**	**500**	**0**	**99**	**0**	**1**
@11	**0.06**	**700**	**30**	**0**	**57**	**13**
@12	3,29	3500	30	57	10	3
@13	5,62	3500	40	46	10	4
@14	0,88	1200	49	41	6	4
@15	0,65	1200	44	41	11	4
@16	**0,09**	**400**	**64**	**0**	**28**	**8**
@17	2.27	2250	56	31	10	3

2.1.3. GIS layers

They were mainly obtained or purchased through research projects (Braud et al., 2010): a Lidar Digitial Elevation Model with a 2 m resolution, geology map in the scale 1:50000 digitized by Gnouma (2006), and the pedological map of the French DONESOL program. A detailed land use map was obtained by manual digitisation from a 2008 BDOrtho IGN image (Jacqueminet et al., 2013).

Fig. 1 shows the land use map (Jacqueminet et al. 2013) simplified in four main classes, the natural and artificial drainage network and the location of the various sensors used in this study.

2.2. Model description

The PUMMA model that is used in this study was specifically designed for the hydrology of *peri*-urban areas. It is presented in detail by Jankowfsky et al. (2014). In particular, it takes into account the potential effect of each landscape object on hydrology, especially linear objects (river network, roads, ditches, hedgerows, thin stripes of alluvial forest etc.), and urban objects (sewer networks, roads, impervious surfaces), that are represented with the same degree of detail and in a unified framework. These characteristics make PUMMA quite unique among the existing distributed hydrological models – see Jankowfsky et al. (2014).

PUMMA has the specificity of representing explicitly the hydrological objects of the landscape that are interconnected thanks to the use of the LIQUID modelling framework (Branger et al., 2010). Each of these objects is represented by a specific module: HEDGE for agricultural and forest areas, URBS for urban cadastral units, RIVER1D for the natural river network and artificial sewer network, TDSO for storm water overflow devices, SISTBA for lakes and retention basins. Additional modules allow the flow transfer between modelling units (WTI, WTRI, OLAF) (Jankowfsky et al., 2014) (see Fig. 2). Model mesh consists of irregular modelling units that follow the land use patterns.

The HEDGE runoff generation module simulates rural areas such as forests and agricultural fields, as well as hedges. The soil is represented using two interconnected compartments: drainable porosity where the water table is computed and micro-porosity where infiltration and evapotranspiration take place. Overland flow can occur when the soil column is fully saturated. Hortonian runoff is not accounted for in this module. The RIVER1D runoff routing module, based on the kinematic

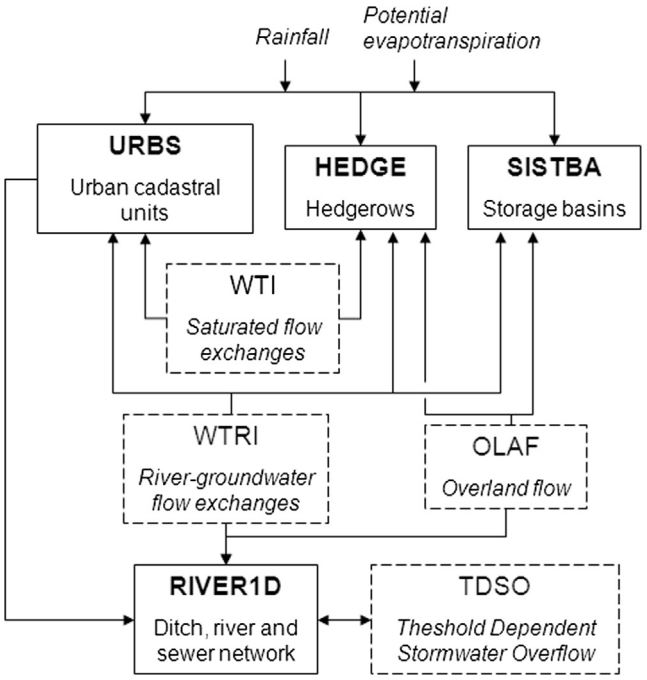

Fig. 2. Structure of the PUMMA model and coupling between the modules (from Jankowfsky et al., 2014). The main modules, representing hydrological processes within the *peri*-urban objects are shown with a continuous line. Interfaces, calculating lateral fluxes between the main modules are shown in dashed lines. Arrows indicate possible water fluxes.

wave equation and simplified channel geometry (trapezes), assures routing of water towards the outlet. The lateral subsurface flow is simulated by the Water Table Interface (WTI) module. It uses the Darcy law to calculate the subsurface flow between agricultural fields, hedge rows and urban cadastral units. The WTRI (Water Table River Interface) is similar to the WTI interface. It calculates the groundwater exchange between a model unit, and a river reach or a lake. Instead of the Darcy equation, it uses the Miles approach (Miles, 1985), which considers the water table deformation close to the river based on the Dupuit–Forchheimer assumption. For built areas, the URBS module is the LIQUID implementation of the Urban Runoff Branching Structure model by Rodriguez et al. (2008). Each cadastral unit is divided into natural, built and road areas, each one having its own water balance. The latter is computed using three reservoirs representing respectively the surface, the non-saturated and saturated zones. Interception by trees is also taken into account. The water table level is computed as the average of the three components and is used to compute rain water drainage by sewer pipes. The Threshold Dependent Stormwater Overflow module (TDSO) simulates the storm water overflow between sewer network and natural streams. The overflow is calculated with a weir or orifice equation based on a water-control depth threshold value. Overland flow from the surface reservoirs of the HEDGE module and the natural part of the URBS module is calculated with the OLAF (OverLAnd Flow) interface. The flux between two units is computed based on the Manning Strickler approach. Finally, the SISTBA (SImulation of STorage BAsins) module simulates retention basins and lakes. It represents retention basins or natural lakes with a simple linear reservoir.

According to the principles of the LIQUID framework (Branger et al., 2010), the way the PUMMA model handles timestep is as follows: each module manages its own variable timestep, according to inner numerical constraints (typically variation of state variables within a timestep). Moreover, each module has the ability to interrupt itself and to recalculate a new timestep if required (for example if a new value is received by another module). This is the way the modules exchange data. Timesteps for a model like PUMMA range typically from a few seconds to several hours, depending on the module and on the rainfall situation.

2.3. Model set up

The simulation was performed for five years from January 1st 2006 to December 31st 2010. In the continuous simulation, year 2006 was considered as the warm-up period and model results were only analysed and discussed for the 2007–2010 period.

2.3.1. Model mesh

The object-oriented model mesh of the PUMMA model is composed of a combination of Hydrological Response Units (HRUs) (Flügel, 1995) for rural land use units and urban hydrological elements (UHE) (Rodriguez et al., 2008) for urban cadastral units, including roads. In the Mercier catchment, agricultural and road ditches were mapped using field survey and were incorporated to the natural river network.

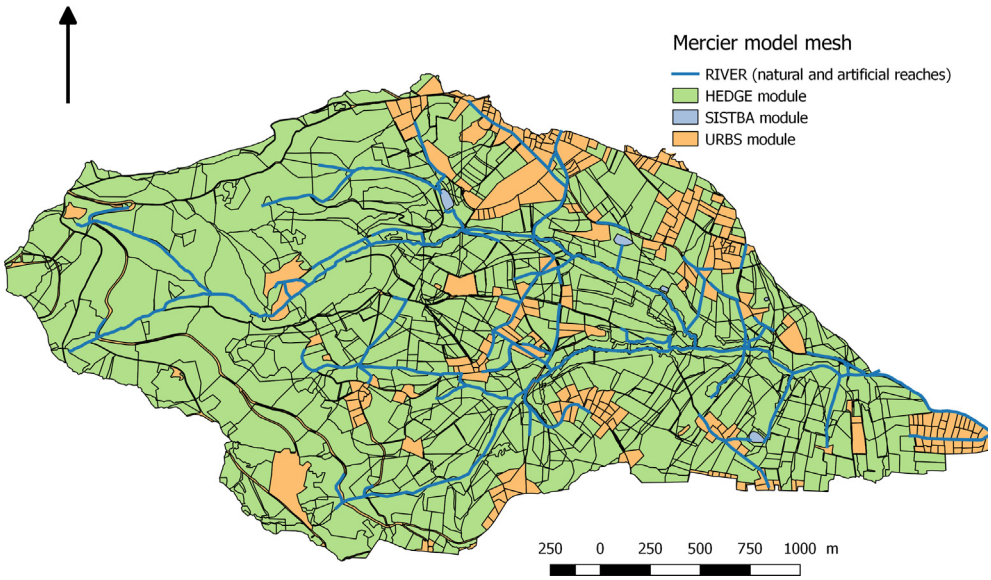

Fig. 3. Mercier catchment PUMMA model mesh. The different colours correspond to the various PUMMA modules: HEDGE representing natural areas, SISTBA representing lakes, URBS representing urban cadastral units. The river network (composed of natural, rain water and ditches network) is shown in blue.

In the small urban area located in the catchment, rain water is collected in a separate sewer system, which was also incorporated to the natural river network. The river network presented in Fig. 3 incorporates all categories of reaches (natural river, road and agricultural ditches, sewers) and forms the basis for the RIVER1D module that simulates flow routing. Note also that water flow in some reaches is intermittent. Furthermore, ephemeral streams were represented using the RIVER1D module. HRUs were obtained using GIS overlay of available land-use, geology, soil and sub-basin maps (Jankowfsky, 2011). Using the 2 m Lidar DTM, rural sub-basins were calculated using the flow direction forcing technique (Kenny and Matthews, 2005) in order to determine one sub-basin for each river and artificial ditch branch (Jankowfsky et al., 2013). The UHEs encompass one cadastral unit (typically house plus surrounding garden) and half of the adjoining street. UHEs and HRUs, as well as lakes and retention basins (for the SISTBA module) were merged in order to get a non-overlapping model mesh. The object oriented model mesh consists of irregular vector objects: polygons for HRUs/UHEs and lakes/retention basins, and lines for the drainage network. It is shown in Fig. 3 with the associated PUMMA process modules. Note that UHEs exclusively composed of roads represent 18% of the URBS modules surface. Each model elements is characterized by various geometric attributes (average altitude, slopes, etc…) which were computed using the DTM.

The WTI/WTRI interfaces are line objects defined by intersection between polygons. Pre-processing methods were also developed to extract the interface geometries and their connections automatically (Jankowfsky, 2011). The automatic preprocessing also provides the flow routing on the irregular model mesh for the OLAF module and determines the urban pipe connections. The model mesh considered in this study is based on the original land use map of Jacqueminet et al. (2013).

2.3.2. Model inputs

Due to its specific handling structure, PUMMA model is able to take into account input series (rain, potential evapotranspiration) at any timestep relevant for the simulation. In this case, we used the highest temporal resolution available for input series: variable timestep for rainfall at both permanent rain gauges, and hourly timestep for reference evapotranspiration. Rainfall was interpolated to the model mesh using weighting averages of the Thiessen polygons. Reference evapotranspiration was considered homogenous over the whole catchment. Crop coefficients are used in PUMMA to modulate *ET0*, according to the vegetation type. Values proposed by Jankowfsky were used (2011).

As stated in the Introduction, an uncalibrated approach was chosen for this study. Therefore, parameters were set according to available data, physical considerations of hydrological processes or the literature. Model parameters were set as in Jankowfsky et al. (2014) for the neighbouring Chaudanne catchment. Soil parameters were specified for each model unit using the DONESOL soil data base and in situ measurements (Gonzalez-Sosa et al., 2010). For the urban UHEs, built, road and natural areas were computed from the detailed land use map of Jacqueminet et al. (2013). Natural and road permeability were assigned values proposed by Jankowfsky et al. (2014), study which -based on a sensitivity study – shows that some parameters sets were more consistent with observations. Their recommendations were used in this study given the proximity between both catchments for interception reservoirs of built areas, roads and natural areas. The link coefficients, representing the fraction of runoff from both built and road areas directly connected to the network were similarly specified to 0.6 and 0.5 respectively. Jankowfsky et al. (2014) recommended the use of a variable soil depth, with larger values at lower altitudes. A uniform soil depth of 1 m was used in this simulation and the evaluation of the variable soil depth is left for further sensitivity studies. Table 2 summarizes the main parameters and input data and how they were obtained.

2.3.3. The evaluation presented in the remaining of the paper uses the simulation configuration described in this section. Model outputs

PUMMA is able to produce outputs at a variable timestep and for each modelling unit. In order to save computation time, most of the variables were set at a fixed timestep of 6 min. This is done by interpolating linearly between the actual calculated variable timesteps. Then comes temporal aggregation (daily means for example) if required for comparison with field data. Observed discharge data at the catchment outlet were interpolated linearly to a 6-min timestep for comparison with the model output.

3. Model evaluation protocol

Model evaluation protocol is designed following the three specific objectives presented in the introduction.

3.1. Assessing how additional data can be processed in the model evaluation

"Hydrological signatures" computed for model evaluation and the hydrological processes they document are presented here.

3.1.1. Hydrological signatures derived from soil moisture data

Soil moisture is quite commonly monitored in field experiments. The use of soil moisture data for model evaluation raises several issues, among which the issue of scaling (point measurements vs. integrated modelling units) and the issue of representativeness (how to compare model values using conceptual storage for soil compartment). In this study, soil moisture data were available only in natural areas, where hydrological processes were simulated using the HEDGE module. Although the soil is divided into layers in this module, soil moisture output of each layer was not available. Only total water storage (mm) of the whole modelling unit was available. The observed soil moisture, however, was a point surface measurement. Moreover, the model was run based on Thiessen polygons calculated from a mere 2 rain gauges – which means that very local variations of rainfall, to which surface soil moisture sensors are sensitive – are likely to be lost. Therefore it was unnecessary to compare observation and model at short, event-based timesteps. As a result, model outputs and data were processed at a daily timestep and normalized values were considered for both soil water storage and surface soil moisture, as shown in Eq. (1), where X is the variable value; min, max, and *norm* respectively for minimum, maximum, and normalized.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Such an indicator provides information about seasonal variations of soil moisture.

3.1.2. Hydrological signatures derived from the water level sensor networks

Distributed water level network consists of a dense network of sensors located in (sometimes ephemeral) streams (Maréchal, 2011). These sensors do not necessarily provide accurate numerical values ("soft" pieces of information according to Hrachowitz et al., 2013), but they can document how water in the streams is distributed spatio-temporally in response to rainfall events. Such piece of information can also be obtained from direct observation of ephemeral streams activation during events (Maréchal et al., 2013). Similar experiments (with temperature sensors) were conducted in ecohydrology for documenting the streamflow patterns of ephemeral streams (Constantz et al., 2001) and snowmelt spatio-temporal dynamics (Lyon et al., 2008). In our case, quantitative water levels cannot be used directly, as they are highly dependent on local stream conditions at the location of the sensor, which a model can never reproduce. It was also impossible to estimate stable rating curves and convert water levels into discharges. No volume/water balance indicators could thus be derived from the water level data. However, we could derive two types of information:

Table 2
Input Parameters for the PUMMA Modules.

Module Name	Parameter Type	Parameter Description (unit)	Source	Parameter range
HEDGE	Soil	Drainable porosity (−)	Field, Rawls and Brakensiek, 1985)	0.16–0.28
		Retention porosity (−)	Idem	0.08–0.25
		Soil depth (m)	Estimated	1
	Vegetation	Crop coefficient, depending on land use (−)	Viaud et al. (2005), Jankowfsky (2011)	0.35–2.0
RIVER1D	Network Reach	Manning roughness parameter (−)	Field observation, Chow (1973)	0.013–0.08
		Slope (−)	Derived from a 2 m DTM	0.001–0.71
		River width (m)	Field survey	0.15–3.4
WTI	Model Unit	Lateral hydraulic conductivities (m s^{-1})	Average of adjacent model units	4.3e-7–1.e-5
WTRI	Model Unit	Lateral hydraulic conductivities (m s^{-1})	Permeability of adjacent model unit	9.0e-7–1.e-5
URBS	Trees	Min. level of interception reservoir (m)	Rodriguez et al. (2008)	0.0003–0.001
		Scaling parameter for intercepted water draining to the surface reservoir (−)	Rodriguez et al. (2008)	0.0067
		Root depth (m)	Estimated	1.1
	Surface	Maximal capacity of street surface reservoirs (m)	Jankowfski et al. (2014)	0.008
		Maximal capacity of built surface reservoirs (m)	Jankowfski et al. (2014)	0.002
		Maximal capacity of natural surface reservoirs (m)	Jankowfski et al. (2014)	0.015
		Saturated hydraulic conductivity of street (m s^{-1})	10* Rodriguez et al. (2008)	7.5e-7
		Saturated hydraulic conductivity of built (m s^{-1})	Rodriguez et al. (2008)	
	Soil	Natural saturated hydraulic conductivity (m s^{-1})	100* Rodriguez et al. (2008)	4.3e-5–1.e-4
		Scaling parameter of the hydraulic conductivity (−)	Rodriguez et al. (2008)	0.2
		Representative position of the vadose zone (m)	Rodriguez et al. (2008)	0.5
		Water content at natural saturation (m^3 m^{-3})	Field, Cosby et al. (1984)	0.4–0.43
		Retention curve exponent (−)	Field, Cosby et al. (1984)	4.6–6.4
		Suction head at air entry (m)	Field, Cosby et al. (1984)	0.11–0.22
	Sewer	Depth of drainage pipes (m)	Estimated	1
		Groundwater drainage coefficient (−)	Jankowfski et al. (2014)	20
		Groundwater drainage exponent (−)	Jankowfski et al. (2014)	2
	Link coefficient	Percentage of water from saturated zone draining into the rain water network (−)	Jankowfski et al. (2014)	1
		Percentage of surface runoff from natural areas leading to overland flow (−)	Jankowfski et al. (2014)	1
		Percentage of surface runoff from streets directly connected to the drainage network (−)	Jankowfski et al. (2014)	0.5 (1 for roads)
		Percentage of surface runoff from roofs directly connected to the drainage network (−)	Jankowfski et al. (2014)	0.6
OLAF	Land	Manning roughness parameter (−)	Depending on land use (Chow, 1973)	0.035–0.24
SISTBA	Storage basin	Maximal level of reservoir (m)	From size	1–2
		Retention parameter (−)	No discharge at the bottom	–

flow/no flow patterns and characteristic times.

(a) Flow/no flow patterns: Sarrazin (2012) estimated that a threshold of 1 cm is representative of these patterns and distinguished three flow categories that he calculated for fortnight periods:

- If water level remains above the threshold for the whole period, it is a “flow” period
- If water level remains below the threshold, it is a “no flow” period
- The last case (water level varies above and below the threshold) corresponds to an “intermittent flow” period.

In terms of hydrological processes, these stream intermittency patterns can be linked with the base flow dynamics in the catchment. Independently from rainfall events, the seasonal patterns, with drying out of streams in summer and incrementally continuous flow in autumn are good indicators of how this base flow is generated and transferred through the catchment.

(b) Characteristic times. Characteristic times used in this study are reaction and response to rainfall events. Reaction time is the time between the beginning of rain and the beginning of significant increase in water level. Response time is defined in this paper as the time between the beginning of the rain and the peak flow. H_{deb} and H_{max} are respectively water depth at the reaction and response times (see Fig. 4).

These characteristic times can document fast flow generation processes (surface runoff, quick subsurface flow), especially for the upstream sensors for which the reach routing process can be neglected and appears in bold in Table 1.

A selection of 32 rainfall events made by Sarrazin (2012) was used for this purpose. It was found that below a cumulative rainfall of 10 mm and/or a rainfall intensity of 2 mm hr^{-1}, no significant water level response was recorded in the downstream stations, which were the most reactive. In addition, for being able to jointly analyse hydrological response at the various locations, it was necessary to spatially select homogeneous rainfalls, based on available rain gauges. Only events with rainfall accumulation differences and timing lower than 30% were kept. This led to the 32 events sample used subsequently. For these

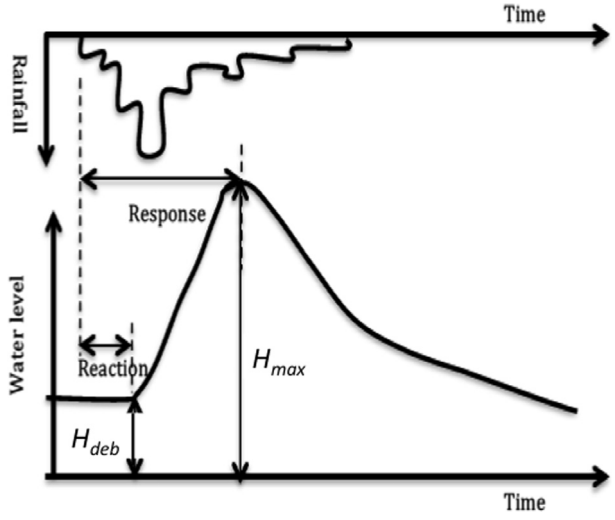

Fig. 4. Conceptual scheme of a flood hydrograph showing the reaction and response times.

events observed between 2007 and 2010, cumulative precipitation volumes range between 4.75 and 73.75 mm and their durations range between 0.2 and 11.5 h.

Sarrazin (2012) also tested several hypotheses in terms of control of the hydrological response by evaluating correlations between H_{max} and H_{max}-H_{deb} and various rainfall and antecedent soil moisture characteristics on the other side. Rainfall events were characterized using the total rainfall amount, event duration, average intensity, and maximum intensity. Antecedent soil moisture was characterized using an antecedent rainfall index and the cumulative rainfall amount in the last 5, 10 or 30 days. The correlation between water level values and these rainfall and soil moisture characteristics was computed using the Spearman rank correlations to identify possible cofluctuations.

3.2. Evaluation of the model's components

We followed a stepwise approach by using standard criteria such as water balance assessment and comparison of observed and simulated discharge at the outlet. Then, we used previously defined signatures derived from additional data to successively evaluate the soil water balance component, base flow component and fast flow component. This evaluation was done using both water level data and statistical analysis of the runoff response's main controls.

3.2.1. Water balance assessment

On an annual scale, model outputs were analysed in terms of water balance and contribution from various components. For each year in the simulation period (2007–2010), rainfall, runoff, actual evapotranspiration (*AET*) and change in soil storage (ΔS) were computed in order to assess interaction or correlation between those elements. System behaviour under dry or wet conditions was analysed through computation of water balance components – which are: contributions from surface runoff or sub-surface flow, and contributions from natural or artificialized areas. Consistency of these contributions was discussed in light of previous findings or observations.

3.2.2. Comparing outlet discharge data to simulation results

The model was evaluated based on the discharge calculated at the Mercier outlet compared to discharge data observed at that location. Two types of comparison were performed: (1) long-term continuous-based comparison; and (2) short-term event-based comparison.

For (1), the Nash-Sutcliffe efficiency (NSE) and the bias (PBIAS) were calculated according to Eqs. (A1) and (A2) in Appendix. Criteria were calculated for each year from 2007 to 2010 as well as over the entire simulation period (January 2007 to December 2010) at hourly and daily timesteps. NSE was calculated both for discharge and square root of discharge, which contrary to NSE on discharge gives less weight to high values.

For the event-based analysis (2), simulations were evaluated using differences in peak value, and peak time (Eqs. (A3) and (A4) in appendix), *NSE* and volume errors to analyze the capacity to simulate runoff events. For this comparison we also added the Pearson correlation coefficient R^2 and weighted correlation coefficient ωR^2 defined by Krause et al. (2006) (see Eqs. (A5) and (A5)). These criteria were calculated at a 6-min timestep, for the selection of 32 events. Eight of these events were not observed at the outlet (gaps in the data) and 4 had spurious observed data, resulting in 20 events retained for the statistical analysis.

3.2.3. Evaluation of the soil water balance component using soil moisture data

Soil water storage (also expressed as soil moisture) is the main state variable of the hydrological system. It is influenced by rainfall, evapotranspiration and soil infiltration and drainage properties. In PUMMA, the soil infiltration component is the core component of the model. Comparison with soil moisture data focuses directly on this component.

The comparison between observation and model was conducted using the Pearson and Kendall τ correlation coefficients between observed and simulated normalized soil moisture defined by Eq. (1) in order to assess the ability of the model to reproduce the seasonal dynamics of soil filling and emptying. Both absolute values of the normalized values and increments from one day to the next were compared.

In terms of model parameters, the soil-filling process is controlled mainly by the rainfall and the modelling-unit's soil depth and porosity (total soil water storage capacity). Soil emptying is driven by evapotranspiration, which is potential evapotranspiration modulated by the vegetation-related crop coefficient, soil wetness, and the lateral conductivity of WTI interfaces (groundwater lateral flow). Since in HEDGE only saturation excess runoff is represented, no additional parameter is involved for surface runoff. When considering only soil moisture increments (i.e. relative variations), soil drainage capacity and lateral conductivity are the main governing parameters.

3.2.4. Evaluation of the base flow component using water level data

The ability of the model to reproduce stream intermittency was tested using the same classification as observed data, for the continuous year 2009, which was divided in 24 (approximate) fortnight periods (see Section 3.1.2). That year was selected because a maximum of sensors were operating then. In addition, a few technical sensors issues were resolved so that the 1 cm threshold could be considered as reliable (Sarrazin, 2012). Because modelled stream reaches are quite simplified as compared to terrain reality (trapeze cross-sections, uniform width throughout the reach), the same 1 cm threshold was deemed irrelevant for the simulated water levels. Three different thresholds were tested: 0.5 cm, 0.7 cm and 1 cm. Assuming the model's actual evapotranspiration calculation performs well, model parameters corresponding to base flow generation in PUMMA are: soil depth in HEDGE (it influences the groundwater depth in each modelling unit), the lateral conductivity in WTI interfaces, and topology (spatial organisation of the modelling units, which control water pathways). For the two sensors corresponding to drainage areas with significant residential zone, the parameter controlling groundwater drainage in the URBS module by sewer pipes is also important.

3.2.5. Evaluation of fast flow components using water level data

In the PUMMA model, fast flow corresponds to surface runoff. On forest or agricultural areas, it is generated by the HEDGE module. In HEDGE, surface runoff is produced when the whole soil profile is filled up. This surface runoff is routed between modelling units by the OLAF module. The corresponding parameters are soil depth and porosity, and topology of the OLAF module. Whereas for residential areas and roads, URBS module is prefered. The main drivers for controlling surface runoff are the interception reservoir, the surface hydraulic conductivity and link coefficients (corresponding to the fraction of surface runoff – especially on impervious road and built surfaces – directly connected to the drainage network). As for the surface, runoff collected from road and built surfaces is directly connected to the stream network. No specific routing parameters are to be taken into account if the length of the stream reach is small enough.

The reaction and response times of the basin at different stations of Mercier were calculated. A comparison is done between the model outputs and the observed values for the 32 events using the Pearson correlation coefficient. Stations @2, @5, @7 to @11 and @16 are located in small head watersheds and the length of the stream reach in the upstream catchment is < 1000 m (Table 1). For these stations, we consider that only surface runoff processes are involved in the model's response. For the other stations, interpretation is more difficult as the stream routing component of the model cannot be neglected. Therefore, the comparison highlights these upstream stations, for which the dominant land uses are respectively forest (stations @8, @9, @10),

Table 3
Catchment averaged components of the simulated water balance for years 2007–2010. The values are calculated for the Mercier catchment. AET is actual evapotranspiration. Percentage is the proportion of rainfall for each component.

Year	Rainfall (mm)	Runoff (mm)	*AET* (mm)	ΔS (mm)	Observed runoff (mm)
2007	892	236 (26%)	631 (71%)	+25 (2.8%)	201
2008	882	259 (29%)	550 (62%)	+73 (+8.3%)	212
2009	607	141 (23%)	626 (103%)	−160 (−26%)	125
2010	885	133 (15%)	642 (72%)	+110 (+12%)	201

forest and agriculture (stations @2, @5, @7) and residential-agriculture (stations @11 and @16).

In addition, the same correlations as defined in 3.1.2 between water level values H_{max} or H_{max}-H_{deb}, and rainfall and antecedent soil moisture characteristics were also calculated using the model outputs to verify whether the model was able to reproduce the observed correlations. These correlations are interesting as Sarrazin (2012) showed. For some stations, amplitude of the response H_{max} or H_{max}-H_{deb} were more correlated to antecedent rainfall and rainfall amounts. This means that the hydrological response is more sensitive to catchment wetness. Other stations, on the other hand, were more related to rainfall intensity. Thus meaning that the hydrological response is more sensitive to infiltration capacity and routing towards the outlet.

4. Results

4.1. Water balance assessment

Table 3 shows components of the annual water balance for the four years and Table 4, discharge components. Table 3 shows that simulated runoff is generally lower when rainfall is lower (2009), but 2010 is an exception: the computed runoff is low and underestimated as compared to observation. Note also the significant simulated change in water storage from one year to the next. The catchment stores water during wet years, but releases it during dry years such as 2009 where decrease of soil water storage is used to fulfil *AET*. The different behaviour of year 2010 may be explained by a dry winter when *AET* is low and more rainfall in summer when vegetation is active.

In terms of runoff components, 2010 is also different from other years, with more sub-surface flow than surface runoff and a larger contribution of URBS (Table 4). This is mainly related to infiltration in sewer networks, which is proportionally higher than other years. Results from Table 4 show that contribution of direct runoff from the natural part of URBS is low (1%), but contribution from HEDGE is large (16–33% of rainfall). Maps of average ponding (see Fig. 10 in Sanzana et al. 2018) show that ponding is higher in downstream catchment area with agricultural land use than it is in the forested area in upstream catchment. This result is consistent with runoff observations from Dehotin et al. (2015) and with saturated hydraulic conductivity estimations provided by Gonzalez-Sosa et al. (2010) showing higher values in forest than in cultivated areas. But this result might also be due to the unique soil depth taken in the model set up, which is probably overestimated in the upstream catchment and underestimated downstream.

Table 4
Components of the river + lakes discharge (%) as computed by the model for years 2007–2010. The percentages are calculated with reference to the total discharge volume.

Components of the river discharge	2007	2008	2009	2010
Direct surface runoff from URBS elements	21	21	23	22
Built	3	3	3	6
Road	17	17	19	15
Natural	1	1	1	1
Direct surface overland flow from HEDGE	33	33	33	16
Sub-surface flow from HEDGE	18	22	22	21
Sub-surface flow from URBS	27	24	21	42
Total surface runoff	55	54	56	38
Total sub-surface flow	45	46	44	62
Total Runoff from HEDGE (natural areas)	52	55	55	37
Total Runoff from URBS (artificialized areas)	48	45	45	63

Table 5
Indicators for the model evaluation of outlet discharge for continuous long-term evaluation.

	Hourly time step			Daily time step	
Period	**NSE_Q**	**NSE_$\sqrt{Q}$**	**PBIAS (%)**	**NSE_Q**	**NSE_$\sqrt{Q}$**
2007	−0.77	0.20	1.9	−0.58	0.24
2008	0.41	0.39	9.3	0.53	0.44
2009	−0.33	0.45	1.3	0.01	0.47
2010	0.10	0.02	−43.0	0.08	−0.05
2007–2010	0.01	0.27	−8.0	0.15	0.29

4.2. Comparison between observed and simulated discharge at the outlet

4.2.1. Continuous long-term comparison

Table 5 shows statistical criteria for each year and the whole period. Table 5 shows contrasted results from one year to the next in terms of NSE with a value of 0.01 for the whole period, hourly values, and 0.15 for daily values. NSE is negative for 2007 and 2009, and positive for 2008 and 2010. The year 2008 is particularly well simulated but includes the highest discharge event ever recorded in the catchment (November 1–3), event which is well simulated by the model (see Fig. 5e), with a large positive impact on the NSE computed for year 2008. NSE on the square root of discharge is much larger than NSE on discharge, with positive values, except on daily values for 2010. This means that the model captures fairly well the general hydrological regime and the base flow, which has low values in summer and higher values in winter. PBIAS is satisfactory with values of less than 10% in absolute values, except for year 2010 where model runoff is underestimated by 43%. Note however that winter 2009–2010 was affected by snowfalls and frost that perturbed rainfall measurement that may have been underestimated during this period. In addition, snow processes are not included in the model, which may explain model underestimation during the 2009–2010 winter.

According to Moriasi et al. (2007) simulation results are satisfactory if NSE > 50% and PBIAS < 25%. Based on these thresholds, the model can be deemed satisfactory only for the 2008-period's daily timesteps. Model performance is therefore quite poor according to statistical criteria. Additionally, the model is uncalibrated and captures variations in seasonal discharge quite well. It is however generally too reactive to small and moderate events rainfall with a very noisy simulated discharge.

4.2.2. Short-term event-based comparison

Table 6 shows performance criteria statistics at the 6-min timestep for the 20 events mentionned in Section 3.2.2. Fig. 5 illustrates model results for six representative events. Results show that the model generally captures event dynamics well with relatively low peak-flow-lag-error values. This is quite remarkable as no roughness coefficients calibration (in the flow routing module) was done to better fit peak times.

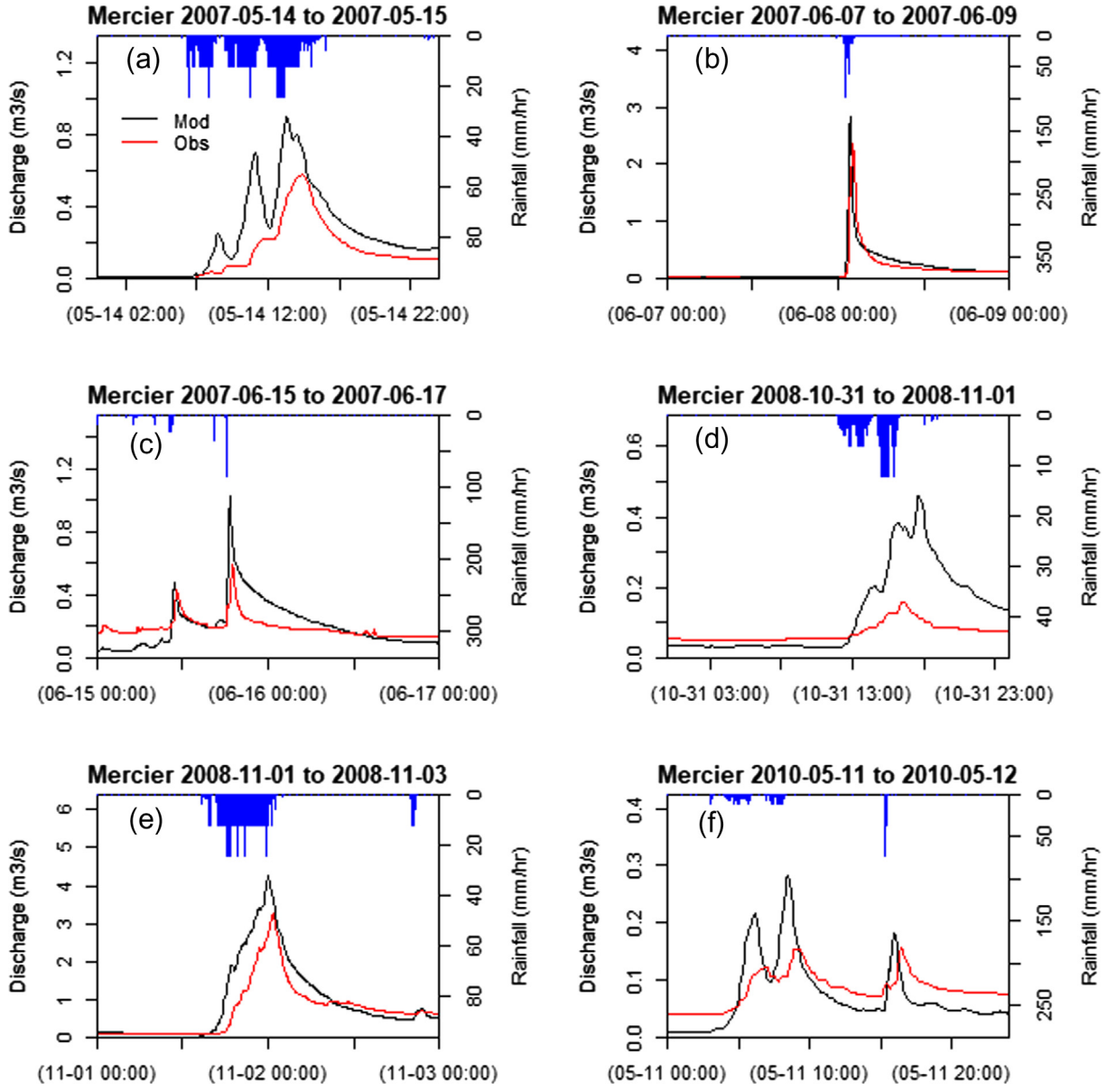

Fig. 5. Observed and calculated outlet hydrographs for 6 representative rainfall events. Note that the scale for discharge and rainfall is different from one figure to the other. Rainfall intensity is provided with the original variable time step, as used as input of the model and discharge values are 6-min time step average values.

Performance is not so good in terms of PBIAS, as the model generally overestimates the volume (median of about 63% for PBIAS). This consistently leads to very poor values of NSE (negative median). On the other hand, R^2 and ωR^2 show much higher values, thus confirming model's ability to reproduce the dynamics. Volume overestimation is greater under dry conditions (Table 6), consistently leading to better model performance in wet conditions. The shape of the simulated hydrographs shows that the model is generally too reactive, with a much peakier behavior than through observation. Such model over-reaction leads to general overestimation of peak discharge.

Table 6
Statistics of the performance indicators for the model evaluation of simulated discharge using 20 events. Statistics are provided for the whole sample and for events occurring in dry and wet conditions using data with 6 time steps.

	All events (20 events)			Dry conditions (8 events)			Wet conditions (12 events)		
	Median	Min	Max	Median	Min	Max	Median	Min	Max
NSE_Q_6min	−3.5	−47.8	0.7	−6.5	−47.8	0.7	−2.4	−17.5	0.5
R^2_6min	0.6	0.0	0.9	0.4	0.0	0.7	0.7	0.4	0.9
ωR^2_6min	0.4	0.0	0.7	0.2	0.0	0.7	0.4	0.2	0.7
PBIAS_6min (%)	62.7	−57.4	275.0	86.9	17.4	275.0	46.7	−57.4	120.7
Peak flow lag time (h)	−0.4	−13.9	5.3	−0.4	−6.3	0.6	−0.4	−13.9	5.3
Peak flow error (−)	1.4	−0.4	8.9	2.2	0.2	8.9	0.9	−0.4	2.0

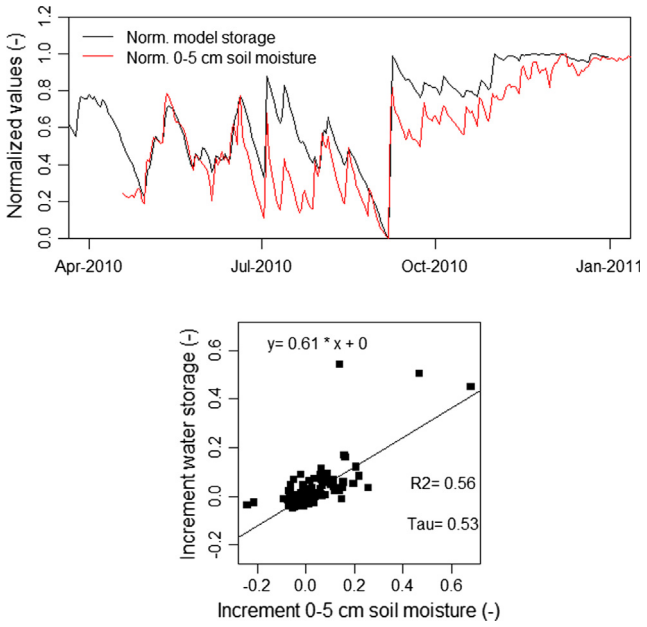

Fig. 6. (Top) Time series of normalized simulated soil water storage (black) and normalized observed surface soil moisture (red) at station #6. (Bottom) Scatterplot of normalized daily increments of modelled soil water storage versus normalized daily increments of surface soil moisture at station #6.

4.3. Evaluation of soil water component

Fig. 6 shows comparison between simulated normalized daily soil-water storage and observed normalized surface soil moisture (top), and correlation between daily increments (bottom). Table 7 summarizes coefficients of determination of the Pearson regression and the Kendall τ for rank correlation between both variables for all the measurement points.

Results show that the model captures seasonal variations of soil-water storage well, and more precisely when soil-water storage is permanently high. Results are also very good for increments. It shows that the model reproduces soil-water storage dynamics quite well. This is less true for station #7, where surface soil moisture does not show an increase in autumn; increase however seen with simulated soil water storage. Discrepancy may be due to particular features of the soil moisture sensor location. For this station, the $\tau_{Kendall}$ for the increments is nevertheless correct. Note also that the $\tau_{Kendall}$ is generally higher than the coefficient of determination. It means that both observed and modelled variables have similar variations, but that this relation is not necessarily linear, which is expected for variables that are different.

In terms of modelled processes and parameters, this means that at daily timestep, reconstitution of local rainfall using Thiessen polygons on the 2 rain gauges is relevant. The representation of evapotranspiration processes and emptying of the soil profile through lateral subsurface flow seem to be correct as well.

Table 7
Coefficients of linear regressions between normalized soil water storage and normalized soil moisture, and Kendall τ value for rank correlation $\tau_{Kendall}$. Results are provided for the absolute values (left) and daily increments (right).

	$S_{norm} = f(\theta_{norm})$		$\Delta S_{norm} = f(\theta_{norm})$	
Soil moisture	R^2	$\tau_{Kendall}$	R^2	$\tau_{Kendall}$
#1	0.36	0.51	0.26	0.41
#2	0.45	0.63	0.56	0.44
#3	0.25	0.33	0.17	0.34
#4	0.49	0.53	0.57	0.48
#5	0.57	0.65	0.28	0.61
#6	0.79	0.71	0.56	0.53
#7	0.04	0.01	0.20	0.46

4.4. Evaluation of base flow component using water level data

The results are presented in Fig. 7 for each station and each period. Summary statistics are presented in Table 8. Observed intermittency patterns are quite balanced, with a predominance of continuous flow (average 46% over the year) in winter, 36% of intermittent flow and 15% only of no-flow, which occurs predominantly in summer. The model does not succeed in reproducing these general patterns. Whatever the threshold, intermittent flow is strongly dominant (53% for 0.5 cm threshold to 72% for 1 cm threshold). Frequency of flow (41% for 0.5 cm and 14% for 1 cm) and no flow (6% for 0.5 cm and 14% for 1 cm) periods are both underestimated compared to observations. The model is more successful in reproducing winter predominance of continuous flow periods, although this remains underestimated. The more realistic picture seems to be obtained with the 0.7 cm threshold, which best reproduces the length of the continuous flow period from January to May, and presents a few continuous flow spots at the end of the year. The 1 cm threshold generates an overall dry picture, whereas the 0.5 cm threshold produces unrealistically long continuous flow periods. On the other hand, whatever the threshold – occurrence of dry periods during summer is missed by the model. But this is less problematic since the distinction between intermittent flow and no flow is quite subtle, and probably also very uncertain and dependent on local conditions at the sensor's location. Looking more closely, we notice that discrepancies come mainly from stations @2 and @7 to @10 which are all upstream stations dominated by forests. On these stations, the simulated behaviour is too dry, with absence of baseflow during the winter months – whatever the threshold. In the model, this can be mainly due to soil depth and lateral flow conductivity, or topology. Comparison with soil moisture data shows that lateral flow conductivity seemed adequate. But this was tested only for agriculture land use where soil moisture sensors were located, and not forests. In the wooded areas, it is possible that we underestimated soil depth and lateral conductivity. But the first-order factor could be topology. Indeed, PUMMA is based on irregular modelling units, and in the wooded upstream areas – units which are very large (see Fig. 3). This can create erroneous flow paths and inaccurate flow transfers between units, since the model takes into account only the mean altitude of each unit. On the other hand, the model behaves much better for stations located in agricultural or residential areas, and areas with small upstream and smaller modelling units (stations @11, @5, @16). The stations where the model overestimates baseflow are @14 @6 and @13 (only @14 for 0.7 cm threshold) which are located downstream each other in this order.

4.5. Evaluation of fast flow component using water level data

Fig. 8 shows comparison between the model output and observed values of reaction and response time, as well as of H_{max}-H_{deb} on the Mercier catchment. Points are colored according to the dominant land use: forest, forest-agriculture and residential-agriculture. Table 9 presents the values of the regression equations $Y_{mod} = b + a^*X_{obs}$ and square coefficient R^2 for different partitions of the data set according to land use, upstream/downstream location (see Table 1), event duration (threshold of 4 h), rainfall depth (threshold of 20 mm), maximum intensity (threshold of 10 mm hr^{-1}) and antecedent rainfall over the last 10 days (threshold of 20 mm).

A low overall correlation between the model and observations can be noticed for reaction time. The slope of the regression is 0.05 over the whole sample; far enough from the target value of 1 to conclude that the model generally reacts too quickly and underestimates reaction time. Table 9 shows that reaction is better simulated for long duration events ($R^2 = 0.33$ for $D > 4$ h) and low intensities ($R^2 = 0.15$ for $I_{max} < 10$ mm/h) as reaction can be determined more accurately under

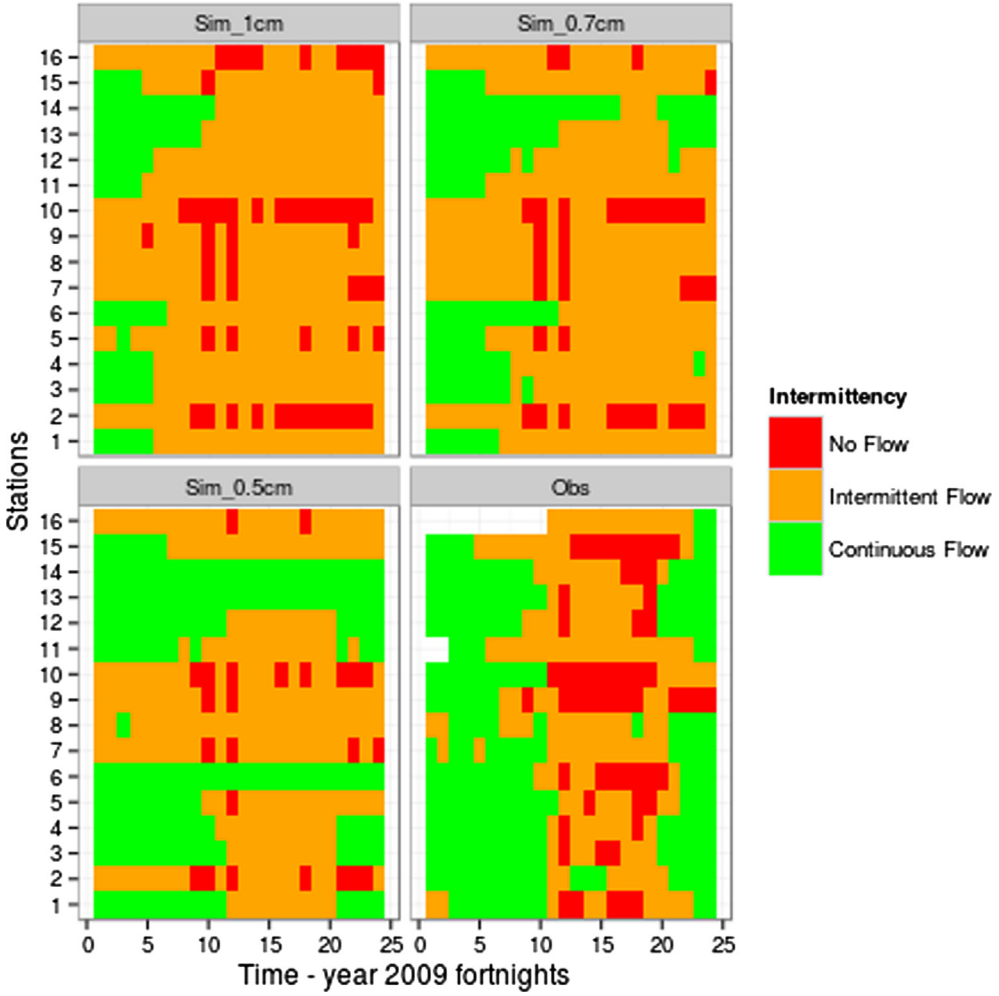

Fig. 7. Comparison of Simulated (Top and Bottom Left) and Observed (Bottom Right) stream intermittency patterns for year 2009 divided into 24 approximate fortnight periods for water level stations @1 to @16.

Table 8
Average frequency of each flow pattern for all the stations with comparison of observed values and simulated values for the three tested thresholds.

Frequency (%)	No flow	Intermittent flow	Continuous flow
Observation (NA values: 3%)	15	36	46
Simulation – 1 cm	14	72	14
Simulation – 0.7 cm	9	66	24
Simulation – 0.5 cm	6	53	41

these conditions. Indeed, observation shows that values are quite uncertain due to sensor accuracy and fluctuations, especially for low values (about a few cm), so that change in just a few centimetres must be detected to determine that there is a reaction. Model results are more stable and it is easier to identify the onset of the hydrological reaction. This may partly explain reaction times underestimation by the model. In terms of processes, reaction can provide information on the ability of the model to generate more or less quick surface runoff by infiltration excess or saturation excess. Reaction is therefore very sensitive to flow paths and model topology for surface runoff. The latter is supposed to be less accurate in the wooded part due to larger modelling units. This may explain the absence of significant correlation for forested catchments as compared to agricultural ones (see Table 9). For downstream stations, correlation in terms of response is not significant and is very low (0.03) for upstream stations (Table 9). This may be related to more complex transfer processes in downstream stations, not properly simulated by the model.

The model better simulates response time than reaction time with an overall $R^2 = 0.39$. The model showed a tendency to satisfactorily simulate reaction time for forested and forested-agricultural land use, with a general model's underestimation, as shown by values of the regression slope of less than 1. None of the factors is well simulated for the residential-agriculture land use, but the sample size is probably too small to be representative. Response is better simulated for the upstream stations, long duration events, low rainfall amount, low maximum intensity events, and high antecedent rainfall events. Response is the sum of reaction and of the rising limb on the hydrograph. It reflects the velocity with which runoff generated in the sub-catchment is transferred to the outlet. The modelled response depends on the way water pathways are represented into the model. Water pathways are shorter in upstream catchments than in downstream ones, thus simulation of the response is less prone to errors than downstream.

Amplitude of the response is also globally not well simulated by the model with an overall correlation of 0.16, reaching 0.32 for forest-agriculture catchments. High correlation is not expected as change in water level is very dependent on local configurations at the sites of observations, which are not represented in the model. The orders of magnitude of simulated and observed response amplitude are very different with slopes of the regressions – generally around 0.3. Results are not very sensitive to the type of rainfall or antecedent soil moisture as performance is very similar between the various classes.

Tables 10 and 11 present each station's Spearman rank correlation

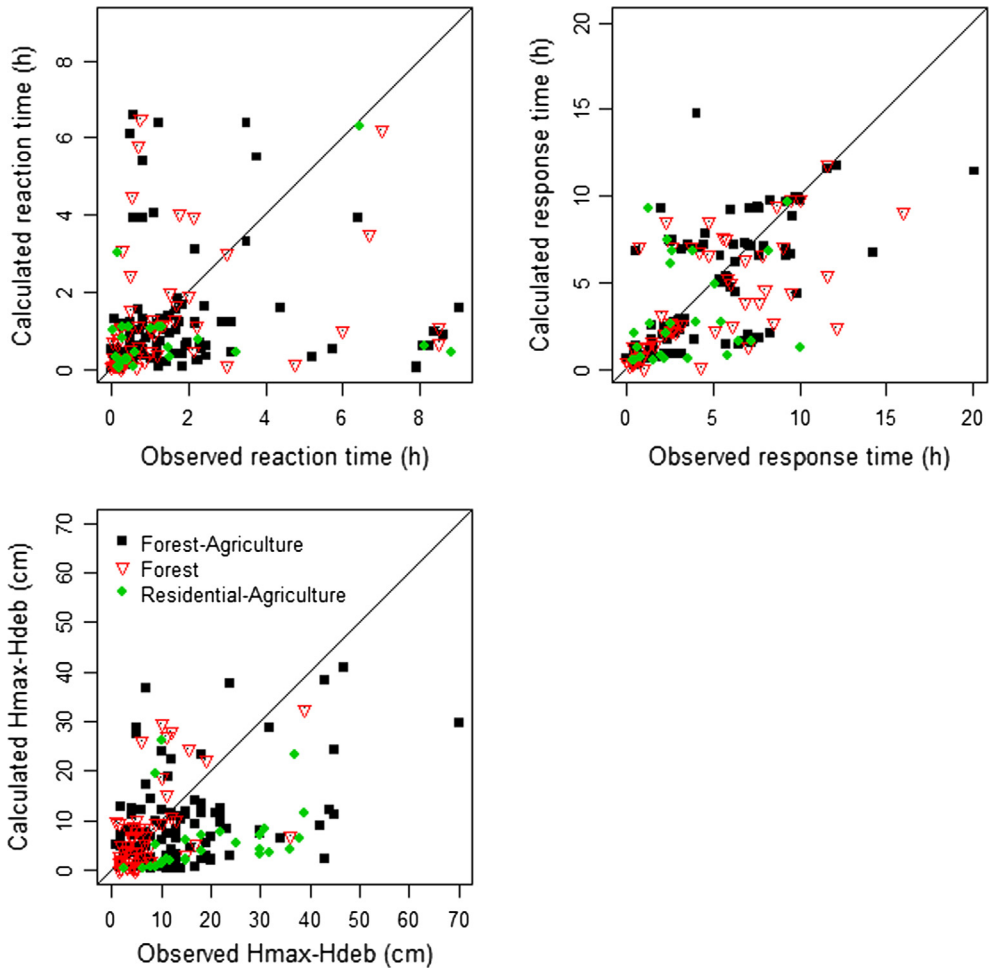

Fig. 8. Comparison between observed and calculated reaction and response times and response amplitude H_{max}-H_{deb}. Points are coloured according to their land use.

coefficients between H_{max} and H_{max}-H_{deb} and variables describing rainfall events and antecedent soil moisture. Only values with a significant correlation test ($p < 0.1$) are reported.

For both variables, observations show high correlation with rainfall volume, maximum rainfall intensity or both for half of the stations. Consistent with observation, the model simulates those relationships between amplitude of the hydrological response, rainfall amount and maximum intensity. But these are stronger and more systematic than observed. In addition, correlations are not as strong with observed H_{max}-H_{deb} than with H_{max} but the model does not detect this difference. This may be due to identification of H_{deb} which is more difficult in observations than in the model (see above), which could lower the observed correlation. For station @11, which corresponds to the most urbanized sub-catchment, the model does not at all reproduce the

Table 9
Coefficients of the regression $Y_{mod} = b + a^*X_{obs}$ and square coefficient R^2 between modelled and observed reaction times, response times and response amplitude H_{max}-H_{deb}. Results are given for various partitioning of the soil sample according to land use (F = forest, FA = forest-agriculture, RA = residential-agriculture; upstream/downstream stations (see their identification in Table 1); event duration D, rainfall event amount R, maximum intensity I_{max}; antecedent rainfall over the last 10 days (*Ant10*). N is the sample size. Only correlation significant at the 0.1 level are provided.

	Reaction time				Response time				H_{max}-H_{deb}			
	b	a	R^2	N	b	a	R^2	N	b	a	R^2	N
All points	0.13	0.05	0.1	196	1.38	0.59	0.39	196	4.28	0.3	0.16	203
F	–	–	–	111	1.26	0.65	0.44	111	4.42	0.32	0.2	114
FA	1	0.2	0.07	62	1.33	0.56	0.43	58	2.66	0.66	0.32	62
RA	–	–	–	27	–	–	–	27	–	–	–	27
Upstream	0.89	0.12	0.03	122	0.83	0.68	0.52	122	5.72	0.33	0.19	125
Downstream	–	–	–	74	2.15	0.45	0.23	74	2.86	0.14	0.08	78
$D < 4$ h	–	–	–	114	3.41	0.38	0.2	114	4.46	0.34	0.2	120
$D > 4$ h	0.27	0.45	0.33	82	0.83	0.29	0.29	82	3.78	0.26	0.13	83
$R < 20$ mm	0.99	0.16	0.06	105	1.88	0.58	0.39	105	5.96	0.31	0.18	109
$R > 20$ mm	–	–	–	91	1.28	0.46	0.24	91	2.68	0.24	0.13	94
$I_{max} < 10$ mm/h	0.56	0.31	0.15	153	1.08	0.46	0.31	153	4	0.29	0.16	157
$I_{max} > 10$ mm/h	–	–	–	43	6.78	0.21	0.13	43	5.24	0.3	0.18	46
Ant10 < 20 mm	0.93	0.27	0.07	137	1.56	0.5	0.33	137	3.51	0.27	0.14	139
Ant10 > 20 mm	–	–	–	59	0.92	0.8	0.56	59	6.33	0.32	0.19	64

Table 10
Spearman correlations between rainfall/soil moisture characteristics and H_{max} for a series of rainfall events. Stations are coloured according to land use (blue: forest; black: forest-agriculture, red: residential-agriculture). Upstream stations appear in bold.

Station	n		Rainfall	Duration	Mean intensity	Max intensity	API	Rain_5days	Rain_10days	Rain_30days
@1	15	Obs	0.72	-	-	0.57	-	-	-	-
		Mod	-	-	-	-	-	-	-	-
@2	**19**	**Obs**	**0.56**	**-**	**-**	**0.6**	**0.5**	**-**	**0.63**	**0.69**
		Mod	**0.7**	**-**	**-**	**0.48**	**0.44**	**-**	**0.38**	**0.43**
@3	14	Obs	0.73	0.55	-	-	-	-	-	0.55
		Mod	0.92	0.62	-	0.55	-	-	-	0.68
@4	18	Obs	0.71	-	-	0.45	-	-	0.4	0.6
		Mod	0.64	-	-	0.53	0.45	0.4	0.42	0.5
@5	**14**	**Obs**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**-**	**-**	**-**	**-**	**0.79**	**-**	**0.89**	**-**
@6	7	Obs	-	-	0.62	0.72	-	-	-	-
		Mod	-	-	-	-	0.9	0.77	0.63	-
@7	**12**	**Obs**	**-**	**-**	**-**	**0.58**	**-**	**-**	**-**	**-**
		Mod	**0.84**	**0.47**	**-**	**0.73**	**-**	**-**	**-**	**-**
@8	**12**	**Obs**	**0.67**	**-**	**0.57**	**0.8**	**-**	**0.54**	**-**	**-**
		Mod	**0.63**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
@9	**8**	**Obs**	**-**	**-**	**-**	**-**	**-**	**0.82**	**-**	**-**
		Mod	**-**	**-**	**-**	**0.96**	**-**	**-**	**-**	**-**
@10	**10**	**Obs**								
		Mod	**-**	**-**	**0.63**	**0.82**	**-**	**-**	**-**	**-**
@11	**21**	**Obs**	**-**	**-0.76**	**0.68**	**0.49**	**-**	**-**	**-**	**-**
		Mod	**0.77**	**-**	**-**	**0.63**	**0.64**	**0.64**	**0.54**	**0.53**
@12	10	Obs	-	-	-	-	-	-	-	-
		Mod	0.73	-	-	0.63	-	-	-	-
@13	15	Obs	0.6	-	-	-	-	-	-	-
		Mod	0.81	-	-	0.48	-	-	-	-
@14	11	Obs	-	-	-	0.74	-	-	-	-
		Mod	0.64	-	-	0.61	-	-	-	-
@15	11	Obs	0.87	0.78	-	-	-	-	-	-
		Mod	0.98	0.69	-	-	-	-	-	-
@16	**7**	**Obs**	**0.65**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**0.59**	**-**	**-**	**0.68**	**-**	**-**	**-**	**-**

Table 11
Spearman correlations between rainfall/soil moisture characteristics and H_{max}-H_{deb} for a series of rainfall events. Stations are coloured according to land use (blue: forest; black: forest-agriculture, red: residential-agriculture). Upstream stations appear in bold.

Station	n		Rainfall	Duration	Mean intensity	Max intensity	API	Rain_5days	Rain_10days	Rain_30days
@1	15	Obs	0.63	-	-	0.53	-	-	-	-
		Mod	-	-	-	-	-	-	-	-
@2	**19**	**Obs**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**0.73**	**-**	**-**	**0.49**	**0.37**	**-**	**-**	**0.43**
@3	14	Obs	0.82	0.55	-	0.5	-	-	-	0.64
		Mod	0.91	0.57	-	0.6	-	-	-	0.7
@4	18	Obs	0.48	-	0.45	0.64	-	-	-	-
		Mod	0.66	-	-	0.49	0.42	0.39	-	0.47
@5	**14**	**Obs**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**-**	**-**	**-**	**-**	**0.96**	**-**	**0.79**	**-**
@6	7	Obs	-	-	-	-	-0.62	-	-0.9	-0.61
		Mod	0.7	-	-	-	0.83	0.72	-	-
@7	**12**	**Obs**	**-**	**-**	**-**	**0.47**	**-**	**-**	**-**	**-**
		Mod	**0.77**	**-**	**-**	**0.75**	**-**	**-**	**-**	**-**
@8	**12**	**Obs**	**0.64**	**-**	**0.64**	**0.84**	**-**	**-**	**-**	**-**
		Mod	**0.7**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
@9	**8**	**Obs**								
		Mod	**-**	**-**	**-**	**1**	**-**	**-**	**-**	**-**
@10	**10**	**Obs**	**0.84**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**-**	**-**	**0.63**	**0.82**	**-**	**-**	**-**	**-**
@11	**21**	**Obs**	**-**	**-0.82**	**0.61**	**-**	**-**	**-**	**-**	**-**
		Mod	**0.76**	**-**	**-**	**0.65**	**0.69**	**0.67**	**0.58**	**0.57**
@12	10	Obs	-	-	-	-	-	-	-	-
		Mod	0.67	-	-	0.62	-	-	-	-
@13	15	Obs	-	-	-	-	-	-	-	0.5
		Mod	0.79	-	-	0.5	0.47	-	-	-
@14	11	Obs	-	-0.77	0.88	0.76	-	-	-	-
		Mod	-	-	-	0.61	-	-	-	-
@15	11	Obs								
		Mod	0.98	-	-	0.68	-	-	-	-
@16	**7**	**Obs**	**0.65**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
		Mod	**0.59**	**-**	**-**	**0.68**	**-**	**-**	**-**	**-**

significantly high anti-correlation with rainfall duration and mean intensity. However, it shows high correlation with rainfall amount and antecedent rainfall. This could point out to a misrepresentation of surface runoff by the URBS runoff, the response of which seems to be governed by soil saturation level in the model.

5. Discussion

To introduce the discussion, we would like to underline that PUMMA, as every hydrological model, was built on specific assumptions (functioning hypothesis). As mentioned in the Introduction Section, the third specific objective of the present paper is to assess whether these assumptions can be dismissed or strenghtened according to the simulation results. Structural assumptions are as follows:

- PUMMA considers only one surface-free aquifer created by filling up the soil profile from impervious bedrock (with no-flow boundary condition at the bottom). The aquifer is consequently quite shallow, and there is no possibility to simulate temporary water saturation above groundwater level (perched water tables). This also prevents the simulation of water storage in deep weathered bedrock if it is not considered as part of the soil.
- The only way to generate surface runoff is by soil saturation. Hortonian surface runoff is not implemented.

In terms of model parameters, other assumptions were also made for the application to the Mercier catchment:

- Spatial discretization (model mesh) was based on land use only, which leads to significantly larger modelling units in the upstream zones of the catchment. PUMMA is able to perform calculations in such a heterogeneous mesh, but one can wonder whether this does not generate distortions in simulated water pathways.
- For simplicity we used a uniform soil depth for the whole catchment.

In addition, PUMMA is uncalibrated, and increasing model performance is not this paper's objective. The model is run using parameter values of either the literature or in situ measurements. Model evaluation is performed on this reference simulation in order to focus on functioning hypotheses testing. Model evaluation also helps to relate discrepancies between model and observation in order to improve process representation and/or parameter specification. This discussion focuses on (1) the lessons that can be learned in terms of simulated hydrological processes from comparison of PUMMA with distributed data sets of soil moisture and water level; (2) on the relevance and limits of the computed hydrological signatures.

5.1. On the use of discharge data

In this study, discharge at the catchment outlet was used for model evaluation in a traditional manner, using statistical criteria to compare simulation and observation. Because our interest is in evaluating the model over the whole hydrological regime, computing several criteria to highlight model performance (for varied ranges) seemed compulsory. The present example shows that using the NSE criterion on discharge alone may be misleading. In our case, NSE values are very poor mainly due to systematic discharge overestimation by the model. Dynamics, however, at all temporal scales are captured quite well, as shown by the relatively good values of R^2 or ωR^2. In addition, NSE is very sensitive to the occurrence of high events, as shown by a much better performance for the year 2008 (see Fig. 5e), during which the highest flood event of the whole study period occurred. In terms of model diagnostic, it is difficult to conclude at this stage.

Even if the model has clearly major weaknesses, these cannot be directly related to any specific components. Nevertheless, the overall volume estimation is not so bad. However, values of the dynamics indicators show that the parameters controlling water transfer (lateral hydraulic conductivity and reach routing parameters) seem reasonable. Future improvements in the use of statistical performance criteria should probably focus on uncertainty assessment, as these criteria are very sensitive to data accuracy. In the case of the Mercier catchment, the stream channel at gauging station is wide and water level is less than 3 cm at low flow, leading to low discharge accuracy. This point was improved recently by adding a V Notch weir to increase sensitivity at low flow.

Uncertainty on the stage-discharge relationship was also assessed using the BaRatin Bayesian method (Le Coz et al., 2014), and work is in progress to propagate this uncertainty and the stage measurement uncertainty to the hydrological records (Horner et al., 2018). This will allow a more robust appraisal of model performance.

5.2. On the accuracy and information brought by soil water content data

Available soil moisture sensors were only monitoring the surface soil-water content at nine locations, all located in agricultural land use. The data are point measurements, whereas model outputs are provided at the scale of a whole modelling unit. There was indeed a scale mismatch between observed and simulated values. In addition, in the model, we were mainly interested in assessing soil-water storage simulation over the whole soil column depth, rather than only the soil moisture behaviour in top soil. It is interesting to see that, once normalised, surface soil moisture provides interesting information about seasonal soil water storage, and in particular about the time when soil water storage is sufficiently replenished to get permanent wet conditions. In terms of model diagnostic, the results give us good confidence in the formulations of input and output flows of the soil water storage in the model. However, the drawback of normalization is that it does not provide any information on the absolute value of soil water storage and this is not very helpful to assess the relevance of the homogeneous 1 m soil depth hypothesis. In order to better constrain soil depth specification, studies should be completed with approaches such as the one proposed by Vannier et al. (2014) in assessing catchment water storage.

5.3. On the accuracy and information brought by the water level network at continuous scale (intermittency)

In its present version, PUMMA is not successful at reproducing stream intermittency as defined in our test. Seasonal variation is approximately reproduced if the flow/no flow threshold is reduced to 0.7 cm, which is consistent with other results (seasonality of discharge at the outlet and soil moisture). It is difficult to interpret where discrepancies originate, as water level in the stream depends on hydrology of the upstream catchment, but also on local hydraulic conditions. Therefore potentially all components of PUMMA could be involved.

We must also underline that this particular test deals with high uncertainty, regarding both observations and model. The choice of a single threshold applied for all stations is a strong assumption, and the chosen value of 1 cm for observations is debatable given sensor accuracy. The existence of flow/no flow depends also on local conditions at the location of each sensor, which were not investigated in detail for simplicity of the application. Moreover the flow routing component of PUMMA is a routing model with simplified geometries: although calculated, the variable "water level" is not necessarily representative, particularly for very low values, where numerical considerations can also interfere with the results. Yet we still can derive a few elements of diagnostic. It seems that the model does not simulate enough baseflow in the upstream part of the catchment, and too much in the downstream part. For the upstream stations, it indicates that either soil water storage is not sufficient, or topology (water pathways) for groundwater flow is not correct. For downstream stations, it is more difficult to conclude. Model's baseflow overestimation is not observed at the outlet, where

statistical criteria indicate that the model is able to successfully reproduce seasonal flow variations. These contradictory results could be explained by compensation effects (overestimation being compensated by underestimation of baseflow further downstream), or by erroneous interpretation of sensor measurements due to flow conditions (there is no overestimation of baseflow at these locations). Additional work is thus required to draw more precise conclusions, in particular by having a closer look at the water level series and defining more specific signatures.

5.4. On the accuracy and information brought by the water level network at the event scale

5.4.1. Characteristic times and amplitude of hydrological response

Although there is quite a high level of uncertainty in the observed data for the determination of reaction and response times, we found that the model seems to react rather too quickly. Since quick surface runoff can only be produced in the model by soil saturation, this also clearly indicates that soil water storage is too small. This is consistent with results of the comparison of outlet hydrographs. This is also consistent with results when comparing observed amplitude of the hydrological response, H_{max}-H_{deb}, with the simulated values much larger than observations. A uniform value of soil depth, as used in this simulation, is probably not realistic given the geological context (gneiss). We could have expected overestimation of soil depth in the upstream catchment and underestimation downstream, as tested by Jankowfsky et al. (2014), but here it seems that soil depth is underestimated everywhere, although deeper soil can be probably found downstream. A geophysical survey (Goutaland, 2009) showed that the bedrock depth was very irregular with a large thickness (about 10 m) of a weathered layer in the talwegs. The fact that response is better simulated upstream than downstream in the catchment (as compared to reaction time, which is badly simulated everywhere) also indicates that misspecification of water pathways (i.e. model topology) might not be a first-order factor, at least for surface runoff.

Proper use of the water level sensors network also requires good knowledge of the rainfall spatial variability as it may induce much localized responses. To overcome this problem, the 32 events selected for the present analysis agree with some criteria about rainfall homogeneity on the available rain gauges. The use of Thiessen polygons using only two rain gauges may also be responsible for discrepancies between model and observation.

5.4.2. Correlations with rainfall characteristics and antecedent soil moisture

The main outcome of this analysis is that representation or parameter specification in the URBS module should be revisited in order to better reproduce the dynamics of the hydrological response in these areas.

5.5. Conclusions on the model diagnostic

Although not calibrated, PUMMA seems to present a few strengths, among which: specification of precipitation (at least for seasonal analysis purpose), specification and calculation of evapotranspiration, and water transfer specification (lateral hydraulic conductivity and river routing).

Observed model weaknesses appear to come mainly from soil water storage underestimation, all over the catchment. Specification of water pathways in the upstream catchment should also be improved. The last direction towards model improvement could be the inclusion of infiltration excess (Horton) runoff. However, this is probably a second order factor as compared to soil water storage and water pathways specifications.

Being the dominant module of PUMMA for the Mercier catchment, as it represents both agricultural and forest land use, these conclusions apply specifically to the HEDGE module, in particular the relevance of the evapotranspiration representation and parameterization, and the underestimation of soil water storage. However, as already pointed out, the data set used in this study cannot bring more quantitative elements. In addition, the soil moisture data were only available in cropped fields. This makes distinctions between parameterization of agricultural land or forested land use difficult. The relevance of the URBS model to represent processes at the cadastral scale was assessed elsewhere (e.g. Rodriguez et al. 2008). In our study, only one water level gauge was draining a large portion of urbanized area (@11) and it showed no significant difference with other sensors. As such, it can be concluded that no specific malfunctioning of the URBS module could be detected according to the available data.

The results of the present study can also be compared with previous work using the same model or modules, in particular the study by Jankowfsky et al. (2014) on the neighbouring Chaudanne catchment. These authors obtained similar results in terms of seasonal variations of discharge at the outlet. They could also identify a problem of overestimation of the summer flood peaks that was found to be due to a non-adequate parameterization of URBS surface components. In our case, this problem was not detected, either because there are less urbanized zones in the Mercier catchment than in the Chaudanne, or more likely because we used parameter values already optimized for the Chaudanne. To some extent, this study can confirm that periurban catchments like Chaudanne or Mercier have to be parameterized differently than more urbanized catchments such as those tested by Rodriguez et al. (2008).

In order to improve the PUMMA model on the Mercier catchment, a sensitivity analysis to soil depth specification (average value, distribution in space, either randomly or according to altitude – see also Jankowfsky et al., 2014) should be performed to see if more realistic simulations can be obtained in terms of simulated discharge volume. It could also be interesting to use discharge recession analysis to infer catchment water storage, as proposed for instance by Vannier et al. (2014). This could provide an independent estimate that can be used to constrain this variable in the model. And it would be interesting to perform the same study using the improved mesh proposed by Sanzana et al. (2017) to check the influence of specification of water pathways.

5.6. Interest of such additional data for model evaluation

The water level network used in this study was initially set up in order to study how the active hydrographic network was changing both seasonally and at event scales in an intermittent catchment (Sarrazin, 2012). The sensors were low cost sensors, conceived at Institute of Fluid Mechanics in Toulouse. They had low accuracy and sometimes delivered quite noisy signals. However, we could derive from these data indicators insights into interesting patterns of hydrological behaviour: intermittency, reaction and response times, response amplitude, correlations with rainfall characteristics and antecedent soil moisture. These indicators helped us to identify the weaknesses of our model and gave us directions towards improving it.

Yet, given the accuracy of the sensors, these indicators are either all related to the dynamics (reaction and response time) or correspond to qualitative information (intermittency, response amplitude). It appears to us that the main problem in the model was volume overestimation during events. The water level sensors network does not provide information that can be used to constrain parameters controlling runoff volume such as soil depth. It would be useful to have some measurement sites inside the catchment where discharge estimation would be feasible with sufficient accuracy as to obtain information on discharge and consequently on volumes. Such data could be used in a calibration process. Calibration is not relevant with the current water level data. Indeed observed water level depends on the very local geometric configuration of the channel, whereas the model assumes a uniform rectangular channel throughout the river reach. As already discussed in 5.3, simulated water level values are not directly comparable to

observations.

In addition, such rough water level measurements are most useful when located in headwater subcatchments, thus reducing the number of hydrological processes potentially involved. On the one hand, downstream sensors may be influenced by propagation or compensation effects, which cannot be easily interpreted. On the other hand, downstream sensors can be used to evaluate the model's flow routing component, by estimating propagation between two sensors. It was not done in this study because the flow routing component was not identified as problematic, but it could be added easily in a generic diagnostic protocol.

Soil moisture data also proved to be interesting and showed that the model could reproduce seasonal dynamics quite well. In the present case, only surface soil moisture was available. In order to get information more directly comparable with soil water storage, monitoring soil moisture on the whole soil profile would be recommended. However, this kind of measurement can only monitor the top soil and not the altered bedrock that may play an important role in controlling catchment water storage (Vannier et al., 2014). Therefore, an integrated estimation of catchment water storage as mentioned above could also be a useful complement in order to better constrain the model.

6. Conclusions and recommendations

The paper proposes a general methodology for a diagnostic evaluation of a complex distributed hydrological model, based on discharge data at the outlet and additional distributed information such as water level and surface soil moisture data. Due to sensor quality issues or representativeness, the proposed hydrological signatures are only able to characterize the dynamics of the hydrological response. They are however useful to assess the relevance of the simulated hydrological response and the underlying model functioning hypotheses with respect to land use, upstream/downstream, rainfall types, preferably, when they are located upstream the catchment. The results also show that information about the dynamics of the response is not sufficient to correct model discrepancies in simulating observed water volume in the river and that distributed quantitative information about catchment soil water storage and discharge are required to improve the simulated response with this respect.

The proposed methodology is illustrated using the PUMMA model in the Mercier sub-catchment (6.6 km^2). Model parameters are specified according to field data and a previous study performed in a neighbouring catchment (Jankowfsky et al., 2014), without calibration. The distributed water level and soil moisture network of sensors were useful in the model evaluation process. They assessed the ability of the model to account for the complexity of the hydrological response according to various factors: land use, upstream or downstream, rainfall type events or antecedent soil moisture. These data helped identify parameters and processes that could be improved or for which a sensitivity analysis would be required. But in their present form, the provided information is only qualitative and cannot be used directly in a calibration process. Direct estimates of catchment water storage and distributed discharge data would be more useful if calibration was to be performed.

Further work to improve the model according to the diagnostic results will consist first of revising specifications of soil depth in the model, and second of assessing results improvement brought by the new model mesh proposed by Sanzana et al. (2017) in order to avoid polygons that are either non-convex or significantly larger. The evaluation process presented in this paper can be repeated and further sensitivity analyses based on various parameters can be conducted.

Declaration of interests

None.

Acknowledgements

The first author thanks Polytechnique Montréal for providing the funding allowing his 6-month stay at Irstea Lyon. Most of the data used in this study were collected during the AVuPUR (Assessing the Vulnerability of PeriUrban Rivers) project funded by the French National Research Agency (ANR) under contract ANR-07-VULN-01, the IRIP project funded by Région Rhône-Alpes and Agence de l'Eau, and in the framework of OTHU (Observatoire de Terrain en Hydrologie Urbaine). We thank Judicaël Dehotin and the IRIP project for providing soil moisture data.

Appendix:. Criteria computed for model evaluation

The following criteria were computed for model evaluation: the Nash-Sutcliffe efficiency (NSE) and the bias (PBIAS) were calculated according to Eqs. (A1) and (A2), where Q stands for discharge, $\bar{Q}$ is the mean discharge, *cal* means calculated, *meas* denotes measured, t the time and n the number of time steps in the evaluated period.

$$NSE = 1 - \frac{\sum_{t=1}^{n} (Q_{cal,t} - Q_{meas,t})^2}{\sum_{t=1}^{n} (Q_{meas,t} - \bar{Q}_{meas})^2} \tag{A1}$$

$$PBIAS = \frac{\sum_{t=1}^{n} (Q_{cal,t} - Q_{meas,t})}{\sum_{t=1}^{n} Q_{meas,t}} * 100 \tag{A2}$$

For the event-based analysis, additional criteria such as difference on peak value (Eq. (A3)), peak time (Eq. (A4)), *NSE* and volume errors were computed. For this comparison we also added the Pearson correlation coefficient R^2 and weighted correlation coefficient ωR^2 defined by Krause et al. (2006) (see Eqs. (A5) and (A6)).

$$Peak\ discharge\ error = (Q_{p\,cal} - Q_{p\,meas})/Q_{p\,meas} \tag{A3}$$

$$Peak\ discharge\ lag\ time = t_{p\,cal} - t_{p\,meas} \tag{A4}$$

where p stands for peak.

$$\omega R^2 = \begin{cases} |a|.\ R^2 if a \leq 1 \\ \frac{1}{|a|}.\ R^2 if a > 1 \end{cases} \tag{A5}$$

with $Q_{cal,t} = a.\ Q_{meas,t} + b$ and

$$R^2 = \left(\frac{\sum (Q_{meas,t} - \bar{Q}_{meas})(Q_{cal,t} - \bar{Q}_{cal})}{\sqrt{\sum (Q_{meas,t} - \bar{Q}_{meas})^2} \sqrt{\sum (Q_{cal,t} - \bar{Q}_{cal})^2} } \right)^2 \quad \text{(A6)}$$

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2018.12.035. These data include Google maps of the most important areas described in this article.

References

Branger, F., Braud, I., Debionne, S., Viallet, P., Dehotin, J., Henine, H., Nédélec, Y., Anquetin, S., 2010. Towards multi-scale integrated hydrological models using the LIQUID framework. Overview of the concepts and first application examples. Environ. Model. Softw. 25 (12), 1672–1681. https://doi.org/10.1016/j.envsoft.2010.06.005.

Braud I., Chancibault K., Debionne S., Lipeme Kouyi G., Sarrazin B, Jacqueminet C., Andrieu H., Béal D., Bocher E., Boutaghane H., Branger F., Breil P., Chocat B., Comby J., Dehotin J., Dramais G., Furusho C., Gagnage M., Gonzalez-Sosa E., Grosprêtre L., Honegger A., Jankowfsky S., Joliveau T., Kermadi S., Lagouy M., Leblois E., Martin J. Y., Mazagol P.O., Michell K., Molines N., Mosini M.L., Puech C., Renard F., Rodriguez F., Schmitt L., Thollet F., Viallet P., 2010. The AVuPUR project (Assessing the Vulnerabiliy of Peri-Urbans Rivers): experimental set up, modelling strategy and first results. InL Proceedings of the 7th Novatech 2010 Conference, June 28-July 1 2010, Lyon, France, 10pp. http://hal.archives-ouvertes.fr/hal-00527564_v1/.

Braud, I., Breil, P., Thollet, F., Lagouy, M., Branger, F., Jacqueminet, C., Kermadi, S., Michel, K., 2013. Evidence of the impact of urbanization on the hydrological regime of a medium-sized periurban catchment in France. J. Hydrol. 485, 5–23. https://doi.org/10.1016/j.jhydrol.2012.04.049.

Chow, V.T., 1973. Open Channel Hydraulics. International students ed. In: International ed. McGrawHill Civil Engineering Series, London, UK, pp. 109–113.

Clark, M., McMillan, H., Collins, D., Kavetski, D., Woods, R., 2011. Hydrological field data from a modeller's perspective: Part 2. Process-based evaluation of model hypotheses. Hydrol. Process. 25, 523–543. https://doi.org/10.1002/hyp.7902.

Constantz, J., Stonestorm, D., Stewart, A.E., Niswonger, R., Smith, T.R., 2001. Analysis of streambed temperatures in ephemeral channels to determine streamflow frequency and duration. Water Resour. Res. 37 (2), 317–328. https://doi.org/10.1029/2000WR900271.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20 (6), 682–690. https://doi.org/10.1029/WR020i006p00682.

Dehotin, J., Breil, P., Braud, I., de Lavenne, A., Lagouy, M., Sarrazin, B., 2015. Detecting surface runoff location in a small catchment using distributed and simple observation method. J. Hydrol. 525, 113–129. https://doi.org/10.1016/j.jhydrol.2015.02.051.

FAO, 1998. Crop evaporation–Guidelines for computing crop water requirements–FAO Irrigation and drainage paper 56. FAO, ISBN, 92-95.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2008. Understanding catchment behavior through stepwise model concept improvement. Water Resour. Res. 44 (1), W01402. https://doi.org/10.1029/2006WR005563.

Flügel, W.A., 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modeling using PRMS/MMS in the drainage basin of the River Bröl, Germany. Hydrol. Process. 9 (3–4), 423–436. https://doi.org/10.1002/hyp.3360090313.

Gnouma, R., 2006. Aide à la calibration d'un modèle hydrologique distribué au moyen d'une analyse des processus hydrologiques: application au bassin versant de l'Yzeron. Inst. National des Sci. Appliquées de Lyon 412.

Gonzalez-Sosa, E., Braud, I., Dehotin, J., Lassabatère, L., Angulo-Jaramillo, R., Lagouy, M., Branger, F., Jacqueminet, C., Kermadi, S., Michel, K., 2010. Impact of land use on the hydraulic properties of the topsoil in a small French catchment. Hydrol. Process. 24 (17), 2382–2399. https://doi.org/10.1002/hyp.7640.

Goutaland, D., 2009. Programme ANR AVuPUR. Prospection géophysique par panneau électrique de trois parcelles d'un sous-bassin de l'Yzeron. Rapport du CETE de Lyon, Juin 2009, 31 pp.

Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modelling 2. Is the concept realistic? Water Resour. Res. 28 (10), 2659–2666. https://doi.org/10.1029/92WR01259.

Gupta, H., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrol. Process. 22, 3802–3813. https://doi.org/10.1002/hyp.6989.

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H.K., Pierrefeu, G., 2018. Impact of stage measurement errors on streamflow uncertainty. Water Resour. Res. 54, 1952–1976. https://doi.org/10.1002/2017WR022039.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183.

Jacqueminet, C., Kermadi, S., Michel, K., Béal, D., Gagnage, M., Branger, F., Jankowfsky, S., Braud, I., 2013. Land cover mapping using aerial and VHR satellite images for distributed hydrological modelling of periurban catchments: Application to the Yzeron catchment (Lyon, France). J. Hydrol. 485, 68–83. https://doi.org/10.1016/j.jhydrol.2013.01.028.

Jankowfsky, S., 2011. Understanding and modelling of hydrological processes in small periurban catchments using an object-oriented and modular distributed approach. Application to the Chaudanne and Mercier sub-catchments (Yzeron catchment, France). University of Grenoble, France.

Jankowfsky, S., Branger, F., Braud, I., Gironas, J., Rodriguez, F., 2013. Comparison of catchment and network delineation approaches in complex suburban environments: application to the Chaudanne catchment, France. Hydrol. Process. 27 (25), 3747–3761. https://doi.org/10.1002/hyp.9506.

Jankowfsky, S., Branger, F., Braud, I., Rodriguez, F., Debionne, S., Viallet, P., 2014. Assessing anthropogenic influence on the hydrology of small peri-urban catchments: Development of the object-oriented PUMMA model by integrating urban and rural hydrological models. J. Hydrol. 517, 1056–1071. https://doi.org/10.1016/j.jhydrol.2014.06.034.

Kenny, F., Matthews, B., 2005. A methodology for aligning raster flow direction data with photogrammetrically mapped hydrology. Comput. Geosci. 31, 768–779. https://doi.org/10.1016/j.cageo.2005.01.019.

Kirchner, J., 2006. Getting the right answers for the right reasons: link measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, W03S04. https://doi.org/10.1029/2005WR004362.

Klemes, V., 1986. Dilettantism in hydrology: transition of destiny? Water Resour. Res. 22 (9), 177–188. https://doi.org/10.1029/WR022i09Sp0177S.

Krause, P., Base, F., Bende-Michl, U., Fink, M., Flugel, W., Pfennig, B., 2006. Multiscale investigations in a mesoscale catchment - hydrological modelling in the Gera catchment. Adv. Geosci. 9, 53–61. https://doi.org/10.5194/adgeo-9-53-2006.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., Le Boursicaud, R., 2014. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. J. Hydrol. 509, 573–587. https://doi.org/10.1016/j.jhydrol.2013.11.016.

Lyon, S.W., Troch, P.A., Broxton, P.D., Molotch, N.P., Brooks, P.D., 2008. Monitoring the timing of snowmelt and the initiation of streamflow using a distributed network of temperature/light sensors. Ecohydrology 1 (3), 215–224. https://doi.org/10.1002/eco.18.

Maréchal, D., 2011. Du drain potentiel au drain réel: utilisation de données satellitales à très haute résolution pour l'étude de l'origine géomorphologique des chemins de l'eau sur des bassins versants soumis aux crues éclair. Thèse de l'Ecole doctorale des. Mines de Saint-Etienne 313, p.

Maréchal, D., Ayral, P.-A., Bailly, J.-S., Puech, C., Sauvagnargues-Lesage, S., 2013. Sur l'origine morphologique des écoulements par l'analyse d'observations hydrologiques distribuées. Application à deux bassins versants cévenols (Gard, France). Géomorphologie 1, 47–62. https://doi.org/10.4000/geomorphologie.10120.

McMillan, H.K., Booker, D.J., Cattoën, C., 2016. Validation of a national hydrological model. J. Hydrol. 541 (Part B), 800–815. https://doi.org/10.1016/j.jhydrol.2016.07.043.

McMillan, H., Clark, M., Bowden, W., Duncan, M., Woods, R., 2011. Hydrological field data from a modeler's perspective: Part 1. Diagnostic tests for model structure. Hydrol. Process. 25, 511–522. https://doi.org/10.1002/hyp.7841.

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., Rupp, D., 2014. Spatial variability of hydrological processes and model structure diagnostic in a 50 km^2 catchment. Hydrol. Process. 28, 4896–4913. https://doi.org/10.1002/hyp.9988.

Miles, J.C., 1985. The representation of flows to partially penetrating rivers using groundwater flow models. J. Hydrol. 82 (3–4), 341–355. https://doi.org/10.1016/0022-1694(85)90026-5.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Am. Soc. Agric. Biol. Eng. 50 (3), 885–900. https://doi.org/10.13031/2013.23153.

Moussa, R., Chahinian, N., Bocquillon, C., 2007. Distributed hydrological modelling of a Mediterranean mountainous catchment – model construction and multi-site validation. J. Hydrol. 337, 35–51. https://doi.org/10.1016/j.jhydrol.2007.01.028.

Paiva, R.C.D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L.G.G., Calmant, S., Getirana, A., Santos da Silva, J., 2013. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. Hydrol. Earth Syst. Sci. 17 (7), 2929–2946. https://doi.org/10.5194/hess-17-2929-2013.

Pereira-Cardenal, S.J., Riegels, N.D., Berry, P.A.M., Smith, R.G., Yakovlev, A., Siegfried, T.U., Bauer-Gottwein, P., 2011. Real-time remote sensing driven river basin modeling using radar altimetry. Hydrol. Earth Syst. Sci. 15 (1), 241–254. https://doi.org/10.5194/hess-15-241-2011.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M.,

Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. J. Appl. Meteorol. Climatol. 47 (1), 92–107. https://doi.org/10.1175/2007JAMC1636.1.

Rawls, W., Brakensiek, D., 1985. Prediction of soil water properties for hydrologic modeling. In: Jones, E., Ward, T. (Eds.), Watershed Management in the Eighties: Proceedings of the American Society of Civil Engineers Symposium, Denver, April 30–May 1. ASCE, New York, pp. 293–299.

Rodriguez, F., Andrieu, H., Morena, F., 2008. A distributed hydrological model for urbanized areas – Model development and application to case studies. J. Hydrol. 351 (3–4), 268–287. https://doi.org/10.1016/j.jhydrol.2007.12.007.

Sanzana, P., Gironás, J., Braud, I., Branger, F., Rodriguez, F., Vargas, X., Hitschfeld, N., Mu-oz, J.F., Vicu-a, S., Meijia, A., Jankowfsky, S., 2017. A GIS-based urban and peri-urban landscape representation toolbox for hydrological distributed modeling. Environ. Modell. Software 91, 168–185. https://doi.org/10.1016/j.envsoft.2017.01.022.

Sanzana, P., Gironás, J., Braud, I., Hitschfeld, N., Branger, F., Rodriguez, F., Fuamba, M., Romero, J., Vargas, X., Muñoz, J.F., Vicuña, S., Mejía, A., 2018. Decomposition of 2D polygons and its effect in hydrological models. J. Hydroinf. https://doi.org/10.2166/hydro.2018.031.

Sarrazin, B., 2012. MNT et observations multi-locales du réseau hydrographique d'un petit bassin versant rural dans une perspective d'aide à la modélisation hydrologique. Ecole doctorale Terre, Univers, Environnement. L'Institut National Polytechnique de Grenoble 269, pp (in French), available at.

Smith, M., Koren, V., Zhang, Z., Moreda, F., Cui, Z., Cosgrove, B., Mizukami, N., Kitzmiller, D., Ding, F., Reed, S., Anderson, E., Schaake, J., Zhang, Y., Andréassian, V., Perrin, C., Coron, L., Valéry, A., Khakbaz, B., Sorooshian, S., Behrangi, A., Imam, B., Hsu, K.-L., Todini, E., Coccia, G., Mazzetti, C., Andres, E.O., Francés, F., Orozco, I., Hartman, R., Henkel, A., Fickenscher, P., Staggs, S., 2013. The distributed model inter comparison project – Phase 2: Experiment design and summary results of the western basin experiments. J. Hydrol. 507, 300–329. https://doi.org/10.1016/j.jhydrol.2013.08.040.

Uhlenbrook, S., Sieber, A., 2005. On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. Environ. Modell. Softw. 20, 19–32. https://doi.org/10.1016/j.envsoft.2003.12.006.

Vannier, O., Braud, I., Anquetin, S., 2014. Regional estimation of catchment-scale soil properties by means of streamflow recession analysis for use in distributed hydrological models. Hydrol. Process. 28, 6276–6291. https://doi.org/10.1002/hyp.10101.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.-M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30 (11), 1627–1644. https://doi.org/10.1002/joc.2003.

Viaud, V., Durand, P., Merot, P., Sauboua, E., Saadi, Z., 2005. Modeling the impact of the spatial structure of a hedge network on the hydrology of a small catchement in a temperate climate. Agr. Water Manag. 74 (2), 135–163. https://doi.org/10.1016/j.agwat.2004.11.010.

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., Karssenberg, D., 2014. The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. Water Resour. Res. 50 (8), 6874–6891. https://doi.org/10.1002/2013WR014639.