Journal of Hydrology 513 (2014) 216–228

Contents lists available at ScienceDirect

Journal of Hydrology

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

Prediction limits of a catchment hydrological model using different estimates of ET_p

R.F. Vázquez *, H. Hampel

Departamento de Recursos Hídricos, Dirección de Investigación (DIUC), Universidad de Cuenca, Av. 12 de Abril S/N, Cuenca, Ecuador

ARTICLE INFO

Article history:
Received 13 September 2013
Received in revised form 13 February 2014
Accepted 22 March 2014
Available online 3 April 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:
Hydrological modelling
MIKE SHE
Evapotranspiration
Model resolution
Uncertainty
Generalised Likelihood Uncertainty Estimator (GLUE)

SUMMARY

A joint deterministic–stochastic protocol based on Monte Carlo simulations (MCS) was applied to the modelling of a medium size catchment using two significantly different sets of potential evapotranspiration (ET_p) data, estimated through a Penman-based method. Modelling error was accounted using the Generalised Likelihood Uncertainty Estimation (GLUE) methodology. Both, water routing and water balance related parameters were considered in the analysis. The research revealed that the hypothesis of the less correct ET_p data set being appropriate for the current modelling could not be rejected since model parameters adjusted to compensate for the use of significantly different ET_p data sets. The GLUE analysis demonstrated that the model predictions exhibited some sensitivity to the hydraulic conductivity of one of the geological layers. Furthermore, the study revealed a considerable error attached to the simulation of both high and low flows, as well as, piezometric levels, for every ET_p data set, which was most likely magnified by the coarse time and spatial scales used in the current modelling.

1. Introduction

Understanding the main physical processes governing the flow dynamics in the land phase of the hydrologic cycle as well as the influence of human activities and climatic changes on the hydrologic cycle plays an important role in integrated catchment management (Feyen and Vázquez, 2011). This is facilitated through a coordinated use of hydrologic-, hydrodynamic-, water optimisation- and water quality-models, in particular, when they are properly applied.

In this framework, hydrologic models are commonly used to make predictions after they have been calibrated and tested. However, these predictions are frequently reported without a thorough assessment of the underlying errors, which arise from the combination of (i) model parameter uncertainties; (ii) input and evaluation data errors; and (iii) model structural errors (Beven and Binley, 1992; Vázquez et al., 2008). The question is then how to assess effectively on underlying prediction errors for distributed models that generally have associated long run times, and for which the sets of effective parameter values that give acceptable fits to the calibration data might be scattered widely in the parameter space (Beven, 1993; Gupta et al., 1998; Beven, 2002; Vázquez et al., 2002; McMichael et al., 2006; Vázquez et al., 2008).

Previous research (Vázquez and Feyen, 2002, 2003, 2004), on the modelling of a medium size catchment in Belgium with the fully distributed physically-based MIKE SHE code (Refsgaard and Storm, 1995), revealed a significant sensitivity of both model performance and effective parameter values to the potential evapotranspiration (ET_p) data, in particular for the parameters of the MIKE SHE module for estimating actual evapotranspiration (ET_{act}). The methodology of the study was based on the independent calibration of the model as a function of the ET_p data set, using a manual trial and error process that inspected as much as possible the parameter space.

In the current research, this sensitivity is further explored within the Generalised Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992; Beven, 2006) that is based on Monte Carlo simulations (MCS). This involves making a large number of runs of the MIKE SHE model, with different parameter sets and different estimates of ET_p. This stochastic–deterministic protocol has the potential of providing a more complete picture of parameter sensitivities and prediction bounds than the previous

* Corresponding author. Tel.: +593 7 405 1000x4496.
E-mail addresses: raulfvazquezz@yahoo.co.uk (R.F. Vázquez), hennihampel@gmail.com (H. Hampel).

http://dx.doi.org/10.1016/j.jhydrol.2014.03.051
0022-1694/

optimisation analyses carried out for the studied catchment. In this context, the use of prediction limits estimation procedures based on MCS is still rarely reported in literature for fully distributed physically-based models

Two ET_p data sets estimated by means of a Penman-based method are considered. $ET_p(A)$ is regarded as significantly overestimating the ET_p in the catchment (Vázquez and Feyen, 2002, 2003, 2004). $ET_p(B)$ is thought to be a more realistic data set for the appropriate modelling of the catchment. The main objective of this study was to explore on whether the "hypothesis" of $ET_p(A)$ being suitable for the modelling of the referred catchment can be rejected (i.e. does the use of $ET_p(A)$ result in many more model rejections?) or whether model parameters can adjust to compensate for the use of wrong data sets (i.e. obtaining similar model performances for significantly different ET_p inputs)?. This research constitutes as such the learning evolution of prior studies (Vázquez and Feyen, 2002, 2003) that were based on manual trial and error model calibration. In this regard, the current stochastic–deterministic approach is likely to better explore the parameter space and as such to provide more sound insides on the modelling effects of using different ET_p data.

The purpose of this research is not exploring on the best approach (i.e. spatial representation/model resolution, length of calibration/validation periods, etc.) to model the dynamics of the study catchment (which was done before for the current study site, i.e., Feyen et al., 2000; Vázquez et al., 2002), but rather, it is to evaluate the previously cited hypothesis on the use of different ET_p data and the respective repercussions on both, predictions as well as parameter values of the MIKE SHE model applied on the study catchment.

Thus, particularly to cope with the complexity of the model of the study catchment and the consequent significant running time associated to every simulation, a coarse modelling resolution and a very limited simulation period were considered to achieve a significant number of MCS, while working with only one MIKE SHE licence available for the current research. This was done for practical reasons, despite the significant uncertainty that these limitations would add to the current hydrological modelling; this research (and manuscript) finally aims at communicating to the reader the need of carrying out suitable modelling tests, such as the one herein described, that estimate modelling bounds (i.e. modelling limitations) prior to the operational use of any numerical model in the context of policy delineation, application and evaluation. Obviously, the reader should apply those tests using more suitable resolutions and simulation periods than the ones herein used for illustrative purposes.

2. Materials

2.1. The study site

The study site, the Gete catchment (586 km^2), which is located in the central part of Belgium (Fig. 1) where the topography is rolling, the soils have a loamy texture, are deep and the groundwater table is generally at a depth of 3 to 10 m below surface. The elevation of the study site varies from approximately 27 m in the northern part to 174 m in the southern part (Fig. 2). Land use in the area is mainly agricultural, including both pasture and cultivated fields, with some local forested areas (Table 1). The spatial distribution of the ten classes corresponds to observations for the period May–August, 1989. Thus, the land use conditions for such period were assumed to be steady throughout the whole modelling process. Nine soil units can be distinguished according to the legend of the Belgian soil map, e.g. loamy soils (Aba, Ada and Adc), sand–loamy soils (Lca, Lda and Ldc), clay soils (Eep and Uep) and soils with stony mixtures (Gbb). The dominant soil type in the study site is the Aba unit. The geology of the study site comprises nine units, some of which occur only in isolated parts of the catchment. The lithostratigraphy of the catchment includes a narrow Quaternarian loamy deposit on top of deeper sandy and clayey units resting on top of a low-permeable Palaeozoic rocky basement (Feyen et al., 2000; Vázquez et al., 2002; Vázquez and Feyen, 2003). The local weather is characterised by moderate humid conditions. For further details on the modelled catchment, see for instance Feyen et al. (2000), Vázquez et al. (2002) or Vázquez (2003).

2.2. The MIKE SHE code

This study was carried out in the scope of a Belgian project focused on the investigation of the performance of integrated fully distributed hydrological codes with respect to their capacity for modelling medium-size catchments under both normal and extreme hydrological conditions. Thus, the MIKE SHE code (Refsgaard and Storm, 1995) was chosen as one of the fully distributed codes to be tested in the context of the project. As such, it was used in the scope of the current study for integrally modelling the flow dynamics of the study site. MIKE SHE is a well-known code that has been used and described in a wide range of applications (Refsgaard, 1997; Jayatilaka et al., 1998; Feyen et al., 2000; Vázquez et al., 2002; Vázquez and Feyen, 2003, 2004; Vázquez et al., 2009). This deterministic distributed hydrological code integrates the entire land phase of the hydrological cycle.

2.3. Main sources of uncertainty

In hydrological modelling data uncertainties exist in part because of: (i) scarcity of accurate spatially distributed parameters and hydrological variables compatible with the code structure; and (ii) the broad (spatial and temporal) modelling scales in relation to data gathering scales. In particular the latter approach, based on the use of coarse modelling grid sizes, is adopted in the current study to compensate for the significant running time associated to the complex model of the study catchment, which made more feasible to run a significant number of Monte Carlo simulations (MCS). Further, the following aspects are likely to affect the suitability of the structure of MIKE SHE for correctly modelling the study catchment:

(i) Assuming that smaller scale equations, derived at a point-scale, are also valid at the larger modelling scale.
(ii) The input time step, during which no change in boundary conditions occurs, was taken as one day, in recognition of the lack of more precise meteorological and calibration data. While the model time step is generally much smaller, this will affect the simulation of some sub-daily processes of the hydrological cycle such as Hortonian flow, drain flows, and saturation processes.

3. Methods

3.1. Estimation of ET_p data sets

In MIKE SHE (DHI, 1998), the calculation of actual evapotranspiration (ET_{act}) in each grid element is based on the Kristensen and Jensen (1975) approach. A detailed description of the MIKE SHE ET_{act} module is given in Vázquez and Feyen (2002, 2003); in what follows, only a brief description of it is given.

ET_{act} and the net rainfall can be modelled as the result of the processes of: (i) interception of rainfall by the canopy that is estimated through a multiplicative function of the interception coefficient (C_{int}) and the Leaf Area Index (LAI); (ii) drainage from

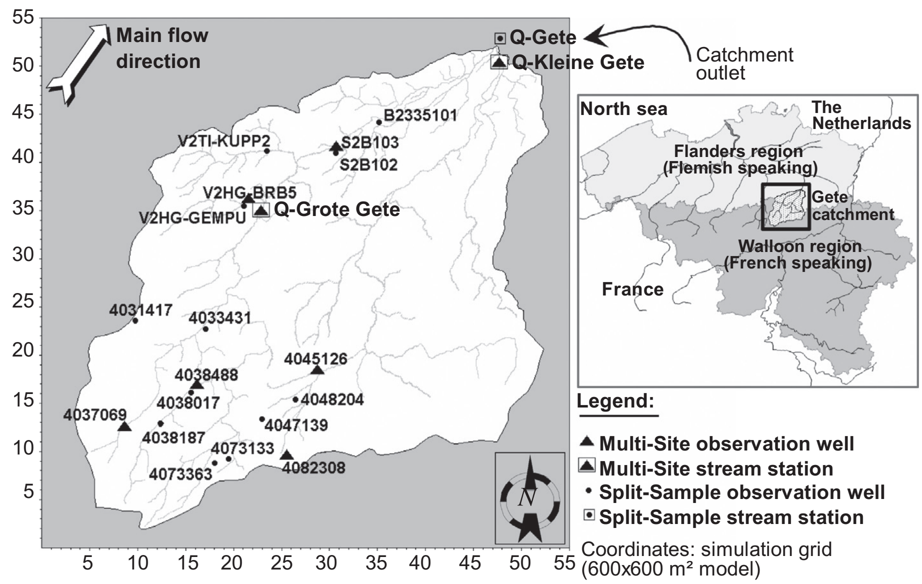

Fig. 1. Location of the study site and gauging stations available for the current modelling.

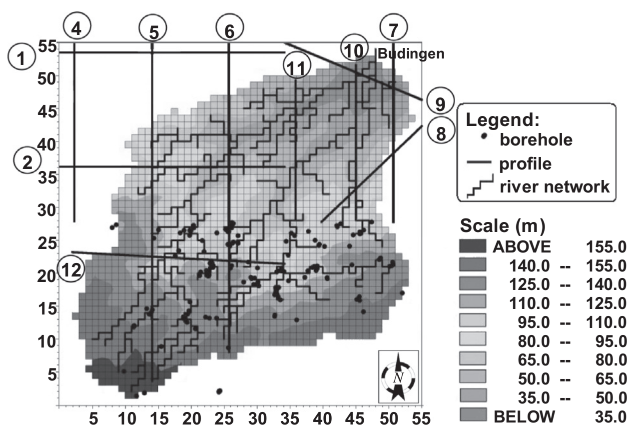

Fig. 2. Digital terrain model (DTM) and river network with the position of boreholes and profiles from geophysical tests (Vázquez, 2003). The map resolution corresponds to the modelling resolution of 600×600 m^2.

Table 1
Land use in the Gete catchment.

Land use description	Class	%
Continuous urban fabric, industrial or commercial units, airports	1	13.0
Port and leisure facilities	2	0.3
Non-irrigated arable lands	3	67.7
Fruit trees and berry plantations	4	1.2
Pastures	5	1.9
Complex cultivation pattern	6	11.1
Land principally occupied by agriculture, with significant areas of natural vegetation	7	3.1
Broad-leaved forest, mixed forest	8	1.1
Water bodies	9	0.1
Paved areas	10	0.4

the canopy; (iii) evaporation from the canopy surface that is approximated as the minimum value of either the maximum interception (I_{max}) or the product of the potential evapotranspiration (ET_p) and the simulation time step (dT); (iv) uptake of water by plant roots and its transpiration; and (v) evaporation from the soil surface.

The plant transpiration (τ_c) and the evaporation from the soil surface (E_s) are estimated as a function of, among other parameters, the empirical parameters C_1, C_2 and C_3, the root mass distribution (A_{rt}), LAI, as well as ET_p. C_1 is purely canopy dependent. C_3 depends on soil type and vegetation. C_2 is defined as a basic evaporation parameter. ET_{act} is then estimated as the contribution of the evaporation from the canopy storage, the transpiration of the vegetation and the soil evaporation (DHI, 1998; Vázquez and Feyen, 2003).

Thus, most of the components of the Kristensen and Jensen (1975) approach depend directly on the estimated ET_p; data that is likely to affect significantly the correct simulation of ET_{act} and as such of the water balance associated to the study catchment, particularly considering that most of the land use (LU) in the catchment is agricultural. Correspondingly, two ET_p data sets were estimated by means of the K_c–ET_o method (Doorenbos and Pruitt, 1977) that uses crop coefficients (K_c) and reference (crop) evapotranspiration (ET_o). The reference crop considered in this study is grass. Two ET_o data sets were estimated with the Food and Agriculture Organisation (FAO) Penman-method 24 (FAO-24). A detailed description of the method can be found at Doorenbos and Pruitt (1977), Vázquez and Feyen (2004). The first ET_o data set (ET_o(A)) was produced with standard parameter values for the various components of the FAO-24 method to emulate previous local ET_o estimation procedures. The second ET_o data set (ET_o(B)) was produced considering locally-applicable parameter values for some of the components of the FAO-24 method (Vázquez and Feyen, 2004) in an attempt to improve previous local ET_o estimation procedures. Specifically, locally applicable parameters were available (Table 2) for the wind function and for the Stefan–Boltzmann equation for estimating the net outgoing long wave radiation (Vázquez and Feyen, 2004).

The estimates produced by the FAO-24 method, using these locally-applicable parameter values for the various components of the method (i.e. ET_o(B)) are comparable (Vázquez and Feyen, 2004) to the estimates obtained with the newer FAO-56 Penman–Monteith method (Allen et al., 1998; http://www.fao.org/docrep/X0490E/X0490E00.htm), parameterised with standard parameter values (i.e. no locally-applicable parameter values were available for the components of the FAO-56 method), which is one of the criteria that encouraged thinking that this ET_o data set is

Table 2
Parameter values used to generate the ET_o data sets (Thom and Oliver, 1977; Groot, 1987; Vázquez and Feyen, 2004).

Parameter	ET_o data set	
	A	*B*
Wind function		
a_w (mm mbar^{-1} day^{-1})	0.270	0.2605
b_w (mm mbar^{-1} km^{-1})	0.0027	0.0016
Stefan–Boltzmann equation		
a_{nl} (–)	0.340	0.560
b_{nl} (mbar$^{-0.5}$)	0.044	0.079
c_{nl} (–)	0.100	0.100

more correct than $ET_o(A)$. An additional criterion that supports considering that $ET_o(B)$ is more correct than $ET_o(A)$ has to do with the results from the Gellens-Meulenberghs and Gellens (1992) assessment that derived a sort of 90% confidence intervals for grass ET_o by considering data from 13 stations, distributed within Belgium (Fig. 3) for a 20-years period ranging from 1967 to 1986. The results of this study are guidelines for the estimation of ET_o adopted by the Royal Meteorological Institute of Belgium (RMIB), and as such were used in this research to assess on the quality of the two ET_o data sets, and as such, implicitly, of the two ET_p sets, since the same K_c values were used for calculating both ET_p data sets. Thus, any conclusion on the quality of the ET_p data can be inferred directly from the ET_o quality assessment, as the proportionality factors (i.e. K_c) are the same in either case and are assumed herein to be accurately assessed.

Herein, it is assumed that the spatial incidence of the meteorological stations (hereby accounted for by Thiessen polygons), located out of the catchment (see Fig. 3), and the spatial variation of land use (defining the geographic variation of the K_c coefficients) are well taken care of in the framework of the current modelling (Vázquez and Feyen, 2002, 2003, 2004).

3.2. Model parameterisation

Here it is provided a brief description of the model paremeterisation. Additional characteristics of the conceptual model of the

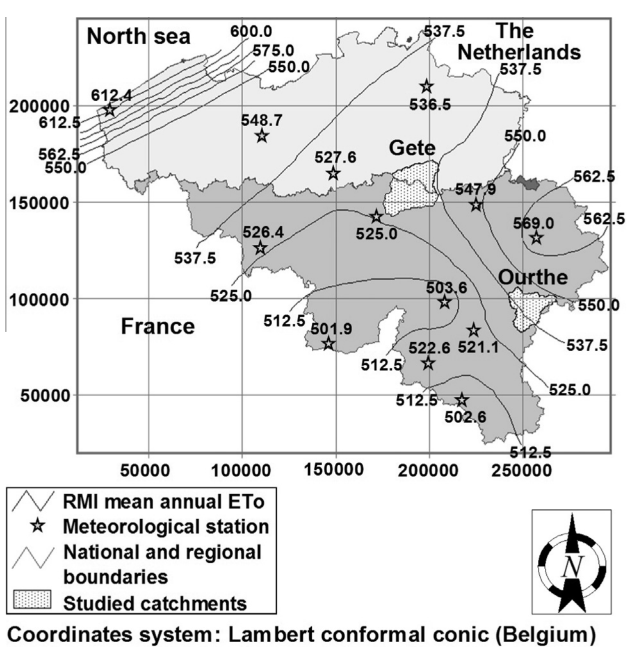

Fig. 3. Meteorological stations used in the research of the Royal Meteorological Institute (RMI) of Belgium (Gellens-Meulenberghs and Gellens, 1992) and resulting spatial distribution of the mean annual ET_o (after Vázquez and Feyen, 2003, 2004).

study site and the corresponding parameterisation of the MIKE SHE structure are given in earlier work, such as, Feyen et al. (2000), Vázquez et al. (2002), Vázquez and Feyen (2003).

The digital terrain model (DTM) was defined by processing available point elevation data by means of a Bilinear (Bi) interpolation method that is accessible as a pre-processor in the MIKE SHE code (DHI, 1998). The definition of the Land Coverage (LC) is based on the classification of LANDSAT satellite imagery available in digital format from the CO-oRdination of INformation on the Environment (CORINE) project via the European Environment Agency (EEA; Vázquez et al., 2008). The pixel resolution of 250×250 m^2 of the LU data was re-sampled to the appropriate modelling resolution by means of the Bilinear (Bi) re-sampling technique that is included in the MIKE SHE software (DHI, 1998; Vázquez, 2003). To complement the description of the agricultural use of the study catchment, Leaf Area Index (LAI) and root depth time series were generated from literature (Allen et al., 1998) on the basis of the crops described in the land use (LU) information. Roughness coefficients for overland flow are based on values taken from literature (Engman, 1986) for different land use types.

Digital information about the topology and elevation of the watercourses was processed with the MIKE SHE river editor (DHI, 1998) to incorporate the river network into the hydrological model of the study site. In the MIKE SHE version (DHI, 1998) used in the current research, the river network is assumed to run along the boundaries of the computational grid squares (Fig. 2). This implies that the resolution of the grid determines the detail of the river in the model set-up (Feyen et al., 2000; Vázquez et al., 2002; Vázquez, 2003). The profile definition of the river tributaries was based on interpolation/extrapolation of a few measured profiles. Drains were specified in the model set-up to improve the simulated hydrograph shape and to account for the small canals and ditches present on a scale smaller than the modelling resolution. The spatial extent of the soil units and their vertical properties could be assessed considering two soil databases of acceptable quality (Vázquez et al., 2008, 2009). Parameters for describing the flow through the soil system were calculated with pedo-transfer functions (PTFs; Vereecken, 1988).

The complexity of the catchment geological system indicated that a three-dimensional (3-D) groundwater model was necessary for simulating the flows and potential heads. The 3-D model was constructed based on 12 geological profiles, digital information about the base of the upper layer (Quaternary period), and 160 borehole descriptions in the Walloon region of the catchment (Fig. 2). Comparison of geological profiles from different origins showed such a wide disparity that the credibility of the geological data was questionable (Vázquez, 2003). Despite the extensive amount of available data, this lack of confidence in the quality of the geological data created a potential source for poor simulation results. In spite of the fact that the geology of the catchment comprises 9 units (some of which occur only in isolated parts of the catchment extent), a prior sensitivity analysis demonstrated that the geological model could be simplified further to include only six units without influencing the global predictions noticeably (Vázquez et al., 2002). Thus, the 3-D geological model includes five upper units on top of the low-permeable basement.

The main grid size used in this study is 600×600 m^2, which has to be considered a coarse discretisation for describing accurately hillslope processes happening in the study catchment. Once more, this is already a resolution that is far too coarse for an accurate modelling of the dynamics of the study catchment, although a prior study demonstrated that acceptable results were still obtained for a model resolution of 600×600 m^2 (Vázquez et al., 2002). Nevertheless, the main objective of this study is different from focusing on the most accurate way of representing the dynamics of the modelled catchment. Furthermore, the number of grid elements in the computational domain (1629) is larger than some previous catchment applications of the MIKE SHE code (Xevi et al., 1997; Christiaens and Feyen, 2002) and other distributed modelling activities (Yu et al., 2001), which further encouraged using this resolution as the main one in the current study. This significant number of grid elements in conjunction with the complex vertical description of the catchment (six geological layers) led to a large computational time associated with every 600-m model run. As a consequence, a short six-months calibration period [1st March 1985–31st August 1985] was chosen. This period was preceded by a six-months spin-up period for attenuating the effects of the initial conditions that were kept the same for every one of the simulations included in the application of the GLUE methodology. A second period [1st September 1985–1st March 1986] was used for model evaluation. Again, these simulation periods are too limited to capture adequately the precipitation-runoff dynamics (responding for instance to climate season variability) occurring in the study catchment. These periods were defined in this study entirely on practical reasons so as to get a reasonable number of MCS under the constraint of having only a single MIKE SHE user licence for the Monte Carlo experiment.

3.3. The GLUE methodology

The joint deterministic–stochastic protocol was based on the application of the Generalised Likelihood Uncertainty Estimator (GLUE) methodology for estimating prediction limits. The fundamental philosophy supporting this methodology is the concept of "equifinality"; i.e. that there may be many model structures and parameter sets that are considered to provide acceptable simulations of available evaluation data (Beven, 1993, 2006).

Multiple sets of parameter values are drawn according to their corresponding assumed prior probability distribution. The simulation results from the Monte Carlo realisations are then evaluated in terms of performance measure(s) against observations. All the predictions (i.e. parameter sets) that exceed a specified threshold of performance, are given a positive likelihood weight and are retained for further consideration. The behavioural likelihood weights are re-scaled such that their sum equals 1, for calculating the distribution functions of both the parameter values and predicted variables. Although subjective decisions such as the choice of a likelihood measure and a threshold value are involved in the application of the GLUE approach, it is transparent and, as such, the decisions are open to debate and justification.

Nonlinearities and parameter interactions can be handled implicitly in the GLUE approach through the likelihood measure, which summarises the non-linear response of a particular model in fitting the available observations. The errors in the input and observation data are also handled implicitly, as the likelihood measure represents the ability of a particular model structure to simulate a particular set of observations given a particular set of inputs.

For distributed models, the main objective of the analysis should be to acquire sufficient behavioural models to sample the different types of model functionality that give rise to good fits in calibration and the potential range of model response in prediction (Beven, 1989; Beven, 1993; Beven, 2001a; Beven and Freer, 2001).

The GLUE methodology enables the integration of additional information into the error assessment by means of a Bayesian type approach to update the likelihood weights and estimated prediction limits. However, Bayesian updating of likelihoods is a choice in GLUE; there are other choices so GLUE need not to be strictly Bayesian (Beven and Binley, 1992; Beven, 2001b). Nevertheless, in the current research, it was decided to use the Bayesian type approach defined by

$$L_p(\Omega_i|\underline{O}) = \frac{L_o(\Omega_i).L_{\underline{O}}(\Omega_i|\underline{O})}{C}, \tag{1}$$

where $L_o(\Omega_i)$ = prior likelihood distribution of the i-th parameter set; $L_{\underline{O}}(\Omega_i|\underline{O})$ = likelihood measure of the i-th parameter set, provided the newer observations ($\underline{O}$) and computed in the newer period of observations; $L_p(\Omega_i|\underline{O})$ = posterior likelihood distribution of the i-th parameter set; and C = scaling constant to ensure that the posterior likelihood measure is unity.

In a first investigation, water routing related parameters were considered; namely, drainage level (z_{dr}), reciprocal time constant (T_{dr}), horizontal saturated hydraulic conductivity (K_x), vertical saturated hydraulic conductivity (K_z) and specific yield (S_y) of the Quaternarian (Kw) and the Landeniaan (Ln) geological layers. These parameters were selected on the basis of prior analyses (Vázquez et al., 2002; Vázquez and Feyen, 2002) that inspected their relevance for simulating the flow dynamics in the study catchment. The parameter subspaces used in this study are bounded by the upper and lower limits, listed in Table 3. The realisations of the hydrogeological parameters were applied to a simple conceptual model of the geological structure of the Kw and Ln layers that prior studies refer to as the most influential for the simulation of the flow dynamics in the study catchment. This simple conceptual model describes the spatial distribution of hydrogeological parameters by considering two distributed zones (i.e. Zone I and Zone II). The parameter values in Zone II are always higher than the parameter values in Zone I (Vázquez et al., 2002).

Table 3
Intervals of the uncertainty process parameters (water routing related parameters).

Model parameter	(Spatial) parameter zone	Geological unit	Lower boundary	Upper boundary
z_{dr} (m)	Entire catchment	–	−1.00	−0.10
T_{dr} (s^{-1})	Entire catchment	–	6.0×10^{-8}	4.0×10^{-7}
K_x (m s^{-1})	Zone I (lower values)	Quaternarian	5.0×10^{-8}	8.0×10^{-7}
		Landeniaan	1.0×10^{-6}	5.0×10^{-5}
K_z (m s^{-1})		Quaternarian	5.0×10^{-9}	8.0×10^{-7}
		Landeniaan	1.0×10^{-8}	1.5×10^{-5}
S_y (–)		Quaternarian	0.050	0.300
		Landeniaan	0.050	0.400
K_x (m s^{-1})	Zone II (higher values)	Quaternarian	5.0×10^{-7}	8.0×10^{-6}
		Landeniaan	1.0×10^{-5}	5.0×10^{-4}
K_z (m s^{-1})		Quaternarian	5.0×10^{-8}	8.0×10^{-6}
		Landeniaan	1.0×10^{-7}	1.5×10^{-4}
S_y (–)		Quaternarian	0.055	0.330
		Landeniaan	0.055	0.440

Legend: z_{dr} = drainage level; T_{dr} = reciprocal time constant; K_x = horizontal saturated hydraulic conductivity; K_z = vertical saturated hydraulic conductivity; S_y = specific yield.

Table 4
Intervals of the uncertainty process for the parameters of the ET_{act} module of MIKE SHE (water balance related parameters).

Model parameter	Lower boundary	Upper boundary
C_{int} (mm)	0.01	1.00
C_1 (–)	0.01	1.00
C_2 (–)	0.05	0.50
C_3 (mm day^{-1})	5.00	40.0
A_{rt} (–)	0.00	5.00
LAI factor (–)	0.50	1.21
z_{rt} factor (–)	0.50	1.50

Table 5
Proportion of runs that failed and that ere behavioural in the calibration period as a function of the Monte Carlo simulations (MCS) analyses.

Analysis	ET_p data set	Grid size (m)	Proportion of runs that	
			Failed	Are behavioural*
MCS(I)	A	600	47/15,000	4234
MCS(II)	B	600	12/15,000	3293
MCS(III)	A	1200	15/25,000	4972
MCS(IV)	B	1200	5/25,000	4641

* *Note:* in the calibration period.

The Quaternarian (Holocene–Pleistocene), Kw, layer is made up of loamy deposits with varying clay content, whilst the Landeniaan, Ln, layer is made up of two formations, namely, the Kortrijk (Lower Eocene) and the Tienen and Landeniaan (Upper Palaeocene) formations. The material of the Kortrijk formation is mainly clay, locally covered by very fine sand, whilst the material of the Tienen and Landeniaan formation is sand with varying clay content; in the lower parts of this formation, it is observed a transition into clayey sand (Vázquez et al., 2002).

Reflecting the lack of prior knowledge on the distribution associated to model parameters, independent uniform distributions were assumed for all of the studied parameters. For each run of the model a new set of parameter values was sampled. A total of 15,000 parameter sets were sampled. The same group of 15,000 parameter sets was tried out in either the first Monte Carlo simulation MCS(I), considering ET_p(A) or the second analysis MCS(II), considering ET_p(B). On average, 25 simulations (600-m model) could be run per day with a personal computer (PC) equipped with a Pentium III processor and having a CPU frequency of 498.5 MHz. Due to software licence constraints, the entire process was carried out with only one PC running 24 h a day.

For the MCS(I) and MCS(II) analyses, the effective values for the parameters of the ET_{act} module of the MIKE SHE code (i.e., water balance related parameters) were assessed from Vázquez and Feyen (2003) and were kept constant for all model parameter realisations considered in the GLUE analysis.

In a second investigation, it was considered not only water routing related parameters but also the water balance related ones, associated to the ET_{act} module of the MIKE SHE code (Kristensen and Jensen, 1975). Thus, a third (i.e. MCS(III)) and fourth (i.e. MCS(IV)) analyses were implemented. The need of these further analyses was stressed in particular by Vázquez and Feyen (2003) that was carried out simultaneously while implementing MCS(I) and MCS(II). Such study, based on a trial and error model calibration, revealed a significant sensitivity of the parameters of

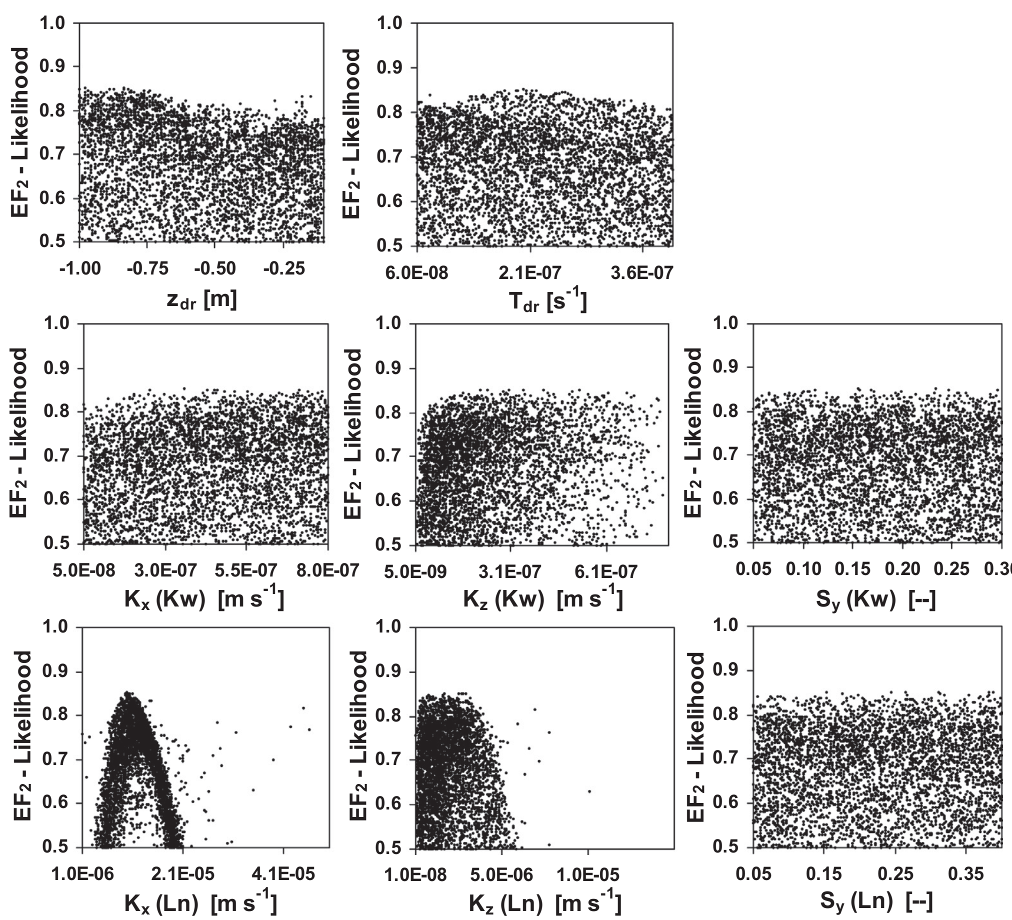

Fig. 4. Scatter plots of behavioural sets of water routing related parameters run through the calibration period, using a 600-m resolution, in relation to the ET_p(A) data set (after Vázquez and Feyen, 2002; Vázquez, 2003).

the MIKE SHE ET_{act} module to the ET_p data set. Thus, correspondingly, the subspaces of the parameters C_{int}, C_1, C_2, C_3 and A_{rt} were inspected in MCS(III) for ET_p(A) and MCS(IV) for ET_p(B). In addition, LAI and z_{rt} factors were also taken into account to multiply the LAI and z_{rt} initial time series, allowing certain compensations during the GLUE analysis for any inadequacies in deriving such time series and estimating grid effective soil parameter values from punctual information. This time, a higher number of parameter sets (25,000) were randomly sampled, as compared to MCS(I) and MCS(II), and this in correspondence with the fact that a higher number of parameters was included this time in MCS(III) and MCS(IV) analyses. The parameter subspaces are bounded by the upper and lower limits listed in Tables 3 and 4. Given the significant number of MCS involved in the investigation, and the high computational time associated with the complex numerical model of the study catchment, it was decided to use a 1200-m modelling resolution, instead of the 600-m resolution that was used for the prior MSC(I) and MSC(II) analyses. This was decided mainly for practical reasons, despite the additional uncertainties attached to the use of even a coarser modelling resolution and, as such, the worse conditions to represent adequately the dynamics of the study catchment that produce unavoidable significant differences between the 600 m model and the 1200 m model. The same warming up, calibration and evaluation periods were considered with regard to the ones used in the MCS(I) and MCS(II) analyses and this with the intention of ensuring a common basis for comparison of results. Table 5 gives an overview of the main characteristics of the MCS analyses considered in this study.

Although the number of sampled parameter sets used in either Monte Carlo analysis seems limited, it is believed that the main functionalities of the hydrological model have been sampled in the parameter space, and as such it is believed that either 15,000 or 25,000 sampled parameter sets is an acceptable number, considering the practical constraints dealing with the complexity of the current model (and as such with the significant simulation time associated to every parameter set); 15,000 (or 25,000) sampled parameter sets was the maximum number of runs that could be achieved in the time, using a commercial model with licence restrictions to a single machine.

The likelihood measure that was computed to characterise the model performance is proportional to the Coefficient of Efficiency that is defined as (Nash and Sutcliffe, 1970; Vázquez et al., 2002; Vázquez and Feyen, 2010):

$$L_{\underline{O}}(\Omega_i|\underline{O}) \propto \left[1 - \frac{\sigma_i^2}{\sigma_{obs}^2}\right] = \mathrm{EF}_2, \qquad (2)$$

where $L_{\Omega}(\Omega_i|\underline{O})$ = likelihood measure for the i-th parameter set (Ω_i) conditioned on the observations $\underline{O}$; σ_{obs}^2 = observed variance; σ_i^2 = error variance for the model, and EF_2 = the Coefficient of Efficiency, which varies between 0 and 1, although it may adopt negative values without a lower limit. It has been used in this research given its common use in the specialised literature, provided that gives an acceptable measure of the combined systematic and random error (Vázquez et al., 2002), despite the fact that it is oversensitive to peak values, although less than other commonly used indexes such as the Pearson type "Coefficient of Determination" R^2 that is the square of the correlation coefficient (Legates and McCabe, 1999; Vázquez et al., 2002).

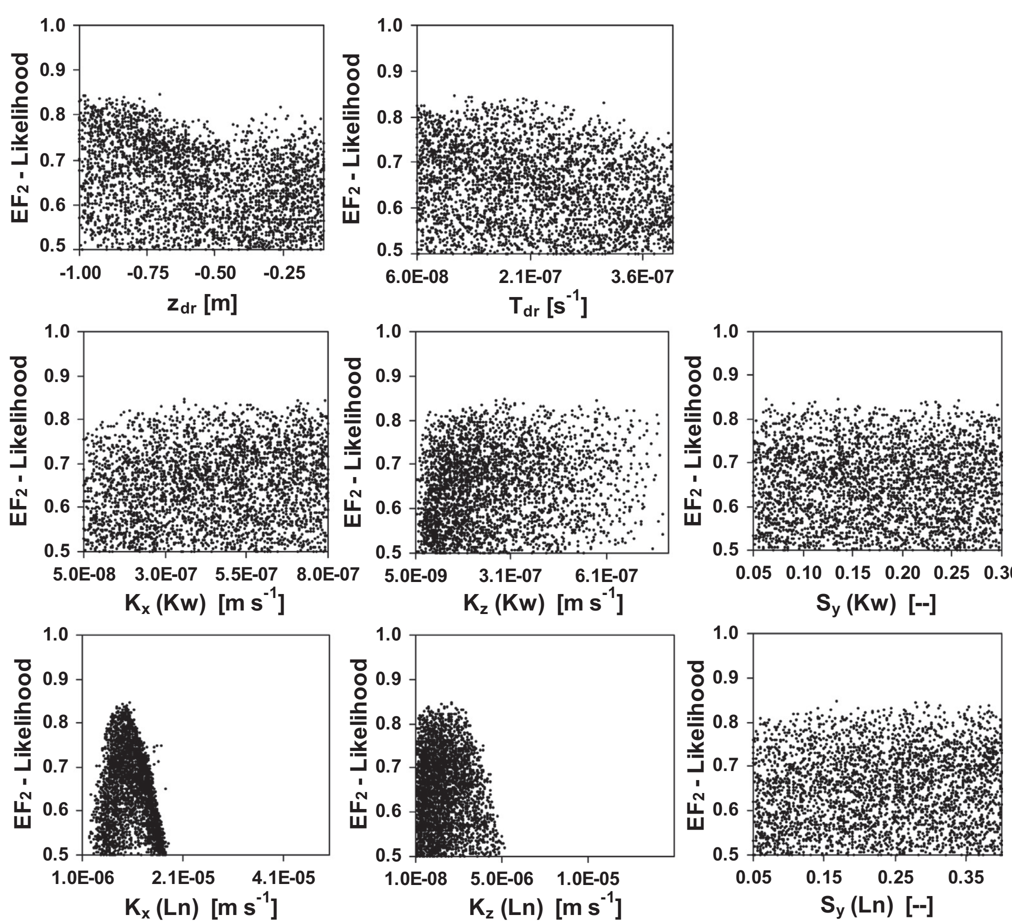

Fig. 5. Scatter plots of behavioural sets of water routing related parameters run through the calibration period, using the ET_p(B) data set and a 600-m resolution.

The behavioural sets were defined by considering an efficiency threshold or cut off criteria ($EF_2 \geqslant 0.5$) for the calibration period [1st March 1985–31st August 1985]. A likelihood weight was then determined for each behavioural model on the basis of the simulation goodness of fit. On the basis of this prior likelihood distribution, a first set of prediction limits were defined. The value of additional information on refining the prediction limits was also inspected in the evaluation period [1st September 1985–1st March 1986] by means of the Bayes type Eq. (1) that was used to update the likelihood distributions. Thus the MIKE SHE code structure was run again to obtain simulation results for this second period considering only the behavioural sets defined in the prior calibration period. The effects of covariation amongst parameters in giving acceptable simulations of the available observations are reflected implicitly in the posterior likelihood weights associated with each behavioural parameter set.

4. Results and discussion

In what follows, only some of the gauging stations that are depicted in Fig. 1 are taken into account (i.e., the outlet streamflow station and the piezometers 4047139 and V2TI-KU.PP2), in agreement with early Monte Carlo simulations studies (i.e., Vázquez, 2003; Vázquez et al., 2009). This in spite of the fact that the remaining gauging stations have been considered in earlier studies that were more focused on the correctness of the simulation of the flow dynamics in the study catchment (see for instance Feyen et al., 2000 or Vázquez et al., 2002). This approach was adopted in the current study (and in the similar studies Vázquez, 2003; Vázquez et al., 2009) to limit the present discussion to a reasonable extent. Also because of the same reason, from the flow simulation viewpoint, the current analysis focuses only in the more influential geological layers, namely the Quaternarian (Kw; piezometer 4047139) and the Landeniaan (Ln; piezometer V2TI-KU.PP2) layers.

Table 5 lists information on the number of runs tried out, the proportion of simulation runs that failed due to instabilities and the proportion of runs that are behavioural in the calibration period, as a function of the MCS analysis.

Fig. 4 (MCS(I), 600-m resolution), Fig. 5 (MC(II), 600-m resolution) and Fig. 6 (MCS(IV), 1200-m resolution) show the scatter plots of the likelihood measures for the set of behavioural models ($EF_2 \geqslant 0.5$) across the sampled parameter ranges. Aiming at maintaining a reasonable extent of the paper, only the results of MCS(IV) are shown with regard to the second modelling investigation. In what follows, the corresponding results of MCS(III) are very similar to the ones of MCS(IV). The likelihood measures were evaluated on the fit to daily discharges, for the Gete station (Fig. 1), in the calibration period [1st March 1985–31st August 1985]. The hydrogeological parameter ranges depicted in the figures (Kw and Ln layers) correspond to the conceptual zone with lower effective values.

The scatter plots of these three figures show that, independently of the MCS analysis, the model performance is most sensitive to the horizontal saturated hydraulic conductivity of the Landeniaan unit (K_x(Ln)), as the respective plots have a peaked band of dots, which implies that the best model performances occurred for parameters sets having K_x(Ln) values approximately between 7.50×10^{-6} m s^{-1} and 1.70×10^{-5} m s^{-1} for either analysis.

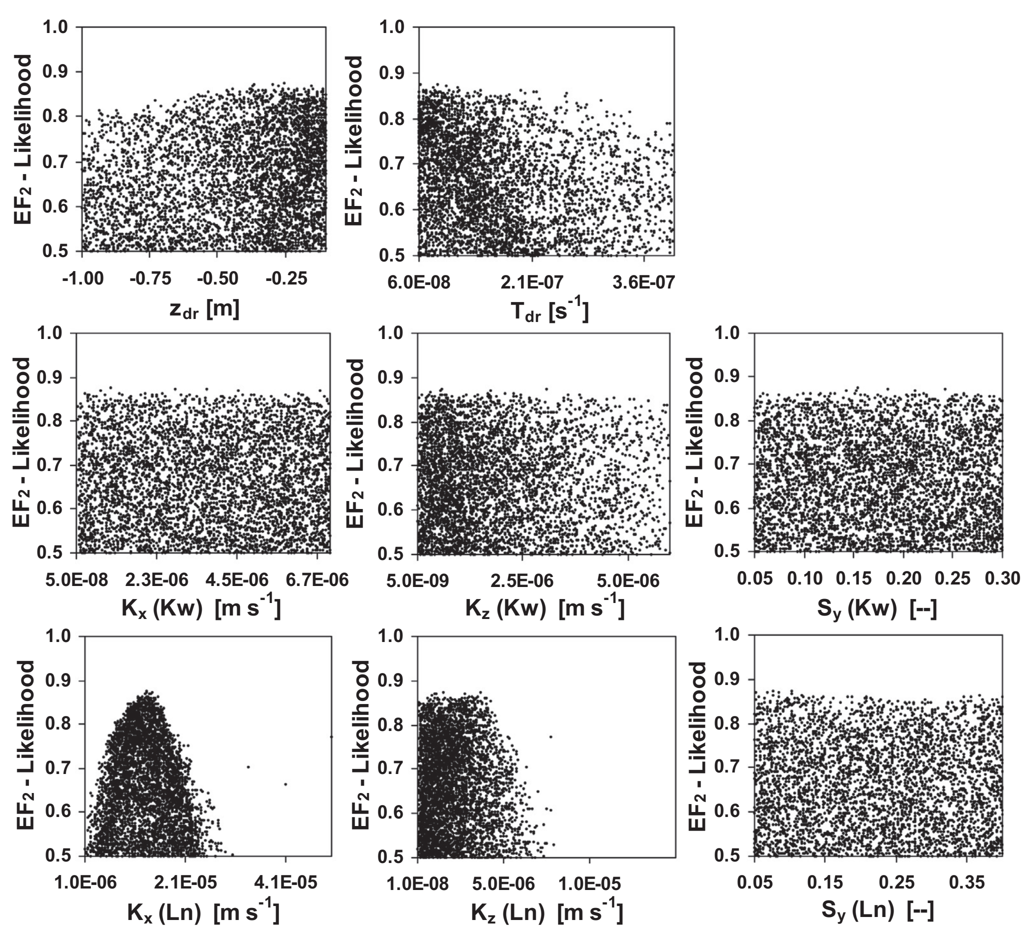

Fig. 6. Scatter plots of behavioural sets of water routing related parameters run through the calibration period, using the ET_p(B) data set and a 1200-m resolution.

However, the scatter plots of K_x(Ln) also show an interesting fact: occasional outliers, suggesting that acceptable performances may also be obtained for parameter realisations falling far away from this parameter interval (in particular for the MCS(I) analysis). The scatter plots of the vertical saturated hydraulic conductivity of the Landeniaan unit (K_z(Ln)) indicate a tendency for achieving better results for values approximately between 1.5×10^{-7} m s^{-1} and 4.0×10^{-6} m s^{-1} for either analysis. However, occasional outliers are also observed for this parameter. The other scatter plots are flat topped across the parameter ranges, suggesting equifinality.

When comparing the scatter plots of K_x(Ln) and K_z(Ln) for the MCS(II) (Fig. 5 and 600-m resolution) and MCS(IV) (Fig. 6 and 1200-m resolution), it is observed wider scattering of the dotty plots for the coarser resolution. Furthermore, the magnitude of best model performances is slightly higher in the calibration period for the coarser resolution (run with parameter sets differing from the ones used in MCS(I) and MCS(II)).

All the MCS analyses, independently of the modelling resolution, confirmed the equifinality of multiple parameter sets for this application of MIKE SHE. Most of the water routing related parameters show scatter plots of model performance that are flat topped across the parameter subspaces, except for K_x(Ln) and K_z(Ln). With respect to the parameters of the MIKE SHE ET_{act} module and the LAI and z_{rt} factors, Fig. 7 shows the scatter plots of model performance for MCS(IV) (i.e. using ET_p(B) and a 1200-m resolution). The figure reveals the insensitivity of the model performance to such parameters and factors, as every plot is flat topped across its corresponding parameter subspace. The same patterns were observed for MCS(III), using ET_p(A). This questions not only the validity of the concept of attaining a single set of optimal parameters but, particularly, the optimality of the parameter sets derived previously through a thorough and systematic (traditional) trial and error calibration (Vázquez and Feyen, 2003).

When comparing MCS(I) (Fig. 4) and MCS(II) (Fig. 5) the magnitudes of best model performances are comparable. Table 5 reveals furthermore that, in the calibration period, there are less model rejections for MCS(I) (using ET_p(A)) than for MCS(II) (the latter, using the likely more correct ET_p(B) data set). This suggest that the hypothesis of the ET_p(A) data set being appropriate for the current modelling could not be rejected so far, in particular because there are less model rejections associated with the use of ET_p(A). Furthermore, the comparable magnitudes of the best model performances (represented herein by the value of the EF_2 index) for both MCS analyses suggest that model parameters have adjusted to compensate for the use of significantly different ET_p data sets. This is particularly in opposition to previous sensitivity analyses (Vázquez and Feyen, 2002, 2003) based on the independent and thorough trial and error calibration of the 600-m model of the study catchment (considering longer simulation periods) as a function of the ET_p data. Clearly, the MCS approach allowed a significantly better exploration of the parameter space than the trial and error approach used in previous work (Vázquez and Feyen, 2002, 2003).

Attempting to limit the extent of this manuscript to a reasonable size, Fig. 8 shows only the prediction limits that were achieved by considering the simulations of the 600-m models. The characteristics of the respective prediction limits for the 1200-m model (i.e. MCS(III) and MCS(IV) analyses) are very similar. Thus, the

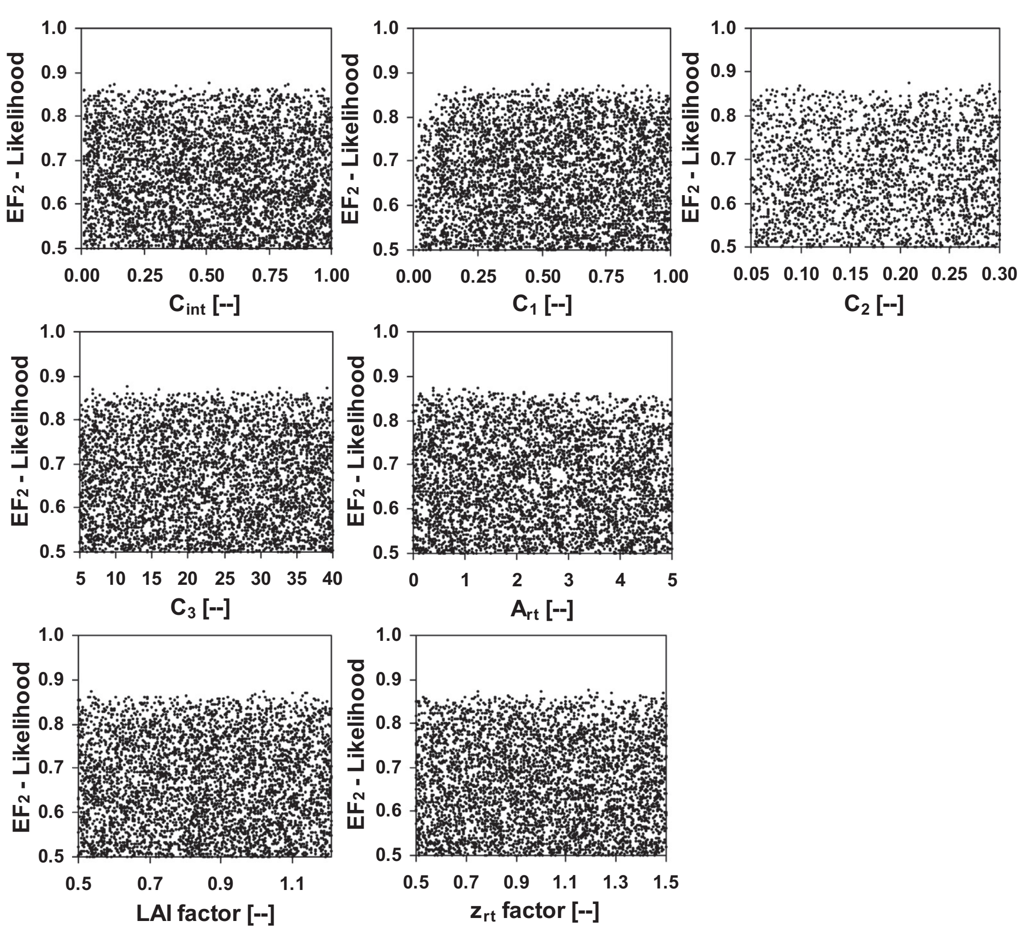

Fig. 7. Scatter plots of behavioural sets of water balance related parameters run through the calibration period, using the ET_p(B) data set and a 1200-m resolution.

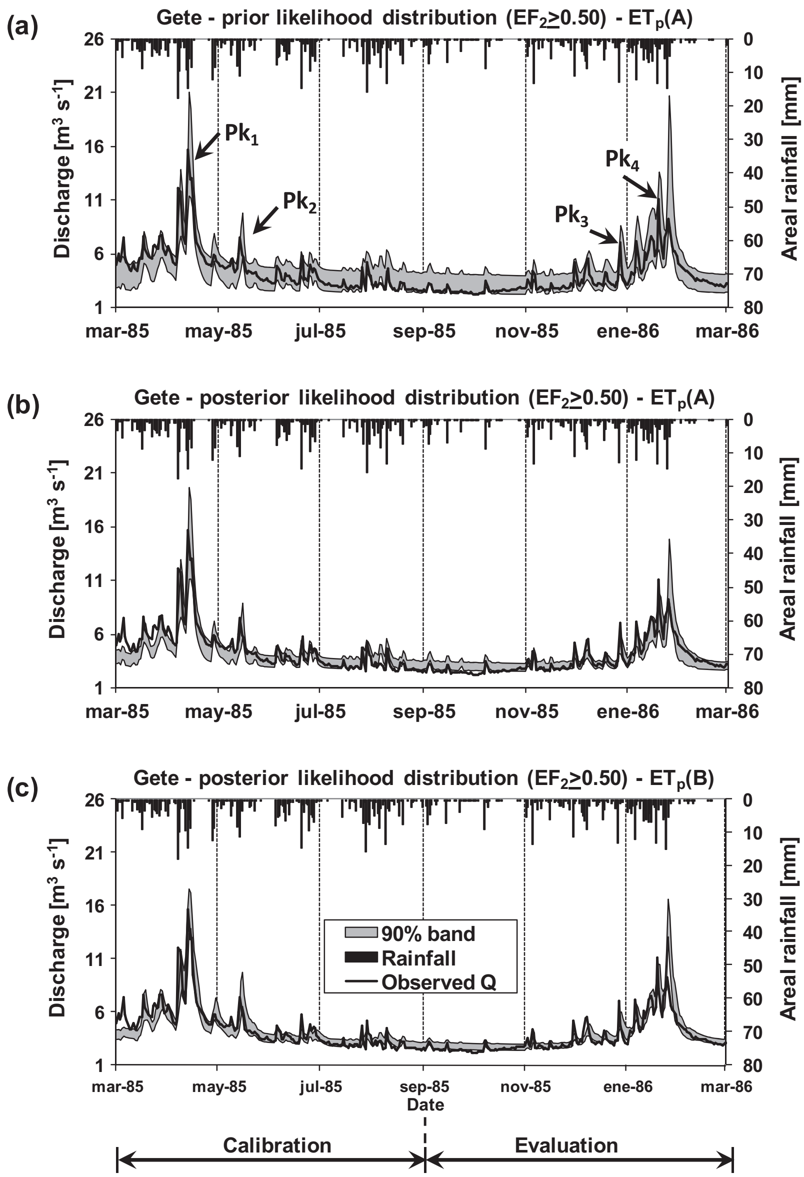

Fig. 8. 90% Streamflow prediction limits in the period [1st of March 1985–1st of March 1986] calculated using a 600-m resolution and (a) the prior likelihood distribution and the $ET_p(A)$ data set; (b) the posterior likelihood distribution and the $ET_p(A)$ data set; and (c) the posterior likelihood distribution and the $ET_p(B)$ data set. The posterior likelihood distribution was obtained after conditioning based on observed streamflow in the period [1st of September 1985–1st of March 1986].

figure depicts 90% prediction bands, calculated in both the calibration as well as the evaluation periods considering a likelihood cutoff value equal to 0.50 (and the 600-m resolution). The figure shows the prediction bounds calculated using (a) the prior likelihood distribution for the MCS(I) analysis using the $ET_p(A)$ data set; (b) the posterior likelihood distribution for the MCS(I) analysis; and (c) the posterior likelihood distribution for the MCS(II) analysis using the $ET_p(B)$ data set. Plots (a and b) illustrate the effects of considering additional information (i.e. during the evaluation period) on the refinement of the prediction limits. Plots (b and c) enable to compare the effects of the use of different ET_p data sets on the results of the GLUE analysis.

Focusing in the calibration period [1st March 1985–31st August 1985], the 90% prediction band, calculated using the prior distribution and $ET_p(A)$ (plot a), is wide for most of the simulation period. The uncertainty is even greater for the main peak events and in some cases the observed discharge crosses over the bounds of the prediction band. The updating of the confidence limits, according to the Bayesian type approach, narrowed the 90% uncertainty band (plot (b)), so that the observed discharge is more often outside the prediction bounds. Plots (b and c) indicate comparable prediction limits, calculated using posterior distributions, as a function of the $ET_p(A)$ and the $ET_p(B)$ data sets, although there is apparently a lower uncertainty attached to the simulation of the first main peak (in April 1985) when using $ET_p(B)$. Fig. 8 depicts furthermore that the models failed in general to represent the low flows, especially in the last part of the calibration period from July 1985 onwards. These deficiencies are probably caused by a combination of significant data uncertainties and/or modelling assumptions, such as the assumption that hydraulic conductivity is constant with depth for the geological layers, rather than being only the effect of inaccurate initial conditions.

Focusing in the evaluation period [1st September 1985–1st March 1986], Fig. 8 shows that the observed discharge crossed over more frequently the lower bound of the 90% prediction bands, depicting in this way that in the evaluation period the models have more difficulties to simulate the low flows despite the different combination of parameters. This is likely to be the direct

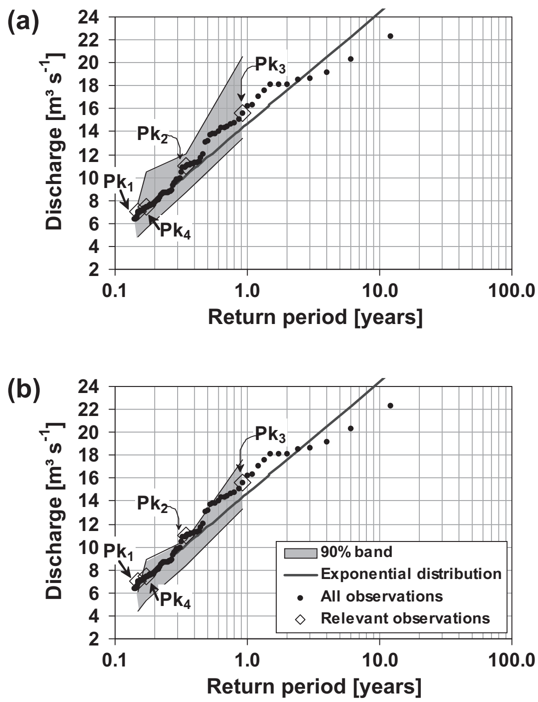

Fig. 9. 90% Peakflow prediction limits in the period [1st of March 1985–1st of March 1986] calculated using a 600-m resolution, the ET_p(B) data set and (a) the prior likelihood distribution; and (b) the posterior likelihood distribution, conditioned on observed streamflow in the period [1st of September 1985–1st of March 1986].

consequence of considering only two of the climate seasons occurring in a 6-month period and as such interfering with a better model assimilation of the complete seasonal variation (i.e. winter, spring, summer and autumn) throughout the model calibration process. Furthermore, these results indicate that the parameterised structure of the MIKE SHE code, for either analysis, had problems to represent correctly the higher peaks in the evaluation period, especially at the end of January 1986. The latter is emphasised in Fig. 9, which for the simulation of peak events shows the 90% prediction bands for MCS(II), calculated after conditioning on the observed catchment-wide stream discharge. The respective results for MCS(I) are comparable to the ones herein illustrated.

The plots in Fig. 9 include the time series of observed independent extreme events and the corresponding Exponential distribution that was derived through an Extreme Value Analysis (EVA), using data in the period [1st of January 1984–31st of December 1995]. The EVA was implemented considering the peak over threshold (POT) algorithm (Pandey et al., 2003). Independent peak daily discharge values were used in the analysis rather than taking into account only one yearly extreme event as conventionally done. In this context, the Generalised Extreme Value theory states that the extreme value distribution $F_{X/X \geqslant x_{th}}$ converges to the Generalised Pareto distribution (GPD), as the threshold x_{th} becomes higher (Pickand, 1975). The analysis showed that, in the period of analysis, an Exponential function fitted reasonably well the observed data above a threshold value of 6.4 $m^3 s^{-1}$ (Vázquez et al., 2009).

Four observed independent peak events were identified in the total simulation period [1st March 1985–1st March 1986]; these are highlighted and labelled in Fig. 9, in correspondence with Fig. 8. The four discharge values simulated at the same dates as the four relevant peak observations were used to obtain the 90% prediction bands. The same empirical return periods associated with the four relevant observations were assigned correspondingly to the simulated events. These four relevant peaks have an empirical return period inferior to one year in the context of the 12-year period of analysis. The plots of Fig. 9 confirm that the peak events in the evaluation period are underestimated, as the peak observations crossed over the upper prediction bound (posterior likelihood distribution).

Fig. 10 depicts the 90% uncertainty bands on the simulation of the piezometric levels in two well locations, calculated after conditioning on the observed catchment-wide discharge. The wells are located in the northern and southern portions of the catchment (Fig. 1) with their screens in the study-relevant Quaternarian (Kw) and Landeniaan (Ln) layers; these wells were chosen to illustrate the simulation results since they are representative of the wells with screens in the Kw and Ln layers. The figure shows that the observations were recorded with different measuring intervals. For well V2TI-KU.PP2 the gathering interval is of about 1 month. For well 4047139 the availability of observations is much scarcer, with measurements done irregularly in time more or less every

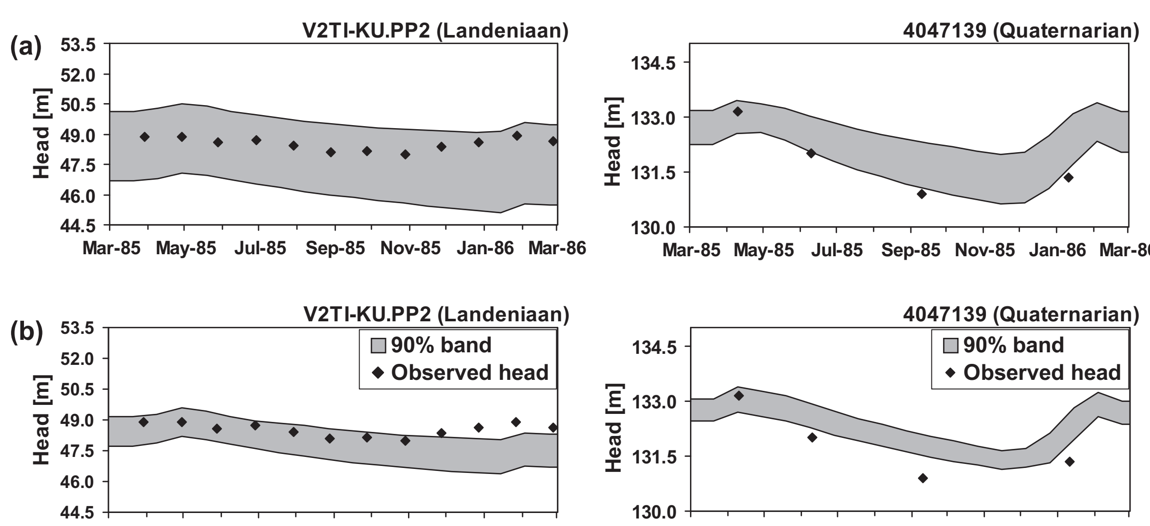

Fig. 10. 90% Piezometric head prediction limits (wells V2TI.KU.PP2 and 4047139) in the period [1st of March 1985–1st of March 1986] calculated using a 600-m resolution, the ET_p(B) data set and (a) the prior likelihood distribution; and (b) the posterior likelihood distribution, conditioned on observed streamflow in the period [1st of September 1985–1st of March 1986].

two months and a half. Further, the figure reveals that there is a discrepancy with respect to the accuracy of the simulations in the two locations, in response to the coarse resolution considered in the analysis, which do not enable a more accurate simulation. Nevertheless, it should be considered that the groundwater module of the current catchment model is also affected by (i) the significant uncertainty attached to the geological data (Feyen et al., 2000; Vázquez and Feyen, 2003); (ii) the significant discrepancies among the elevations of the input data used to built-up the DEM of the catchment, the resulting DEM and the datum used for monitoring the observation wells; and (iii) the incommensurability issues resulting from comparing $600 \times 600\,m^2$ (and even $1200 \times 1200\,m^2$) grid simulation results with point-scale well observations.

5. Conclusions and recommendations

It was found that the ET_p input is not that sensitive to the current modelling, which is contrary to the results of previous work based on traditional calibration (Vázquez and Feyen, 2002, 2003). Throughout the Monte Carlo based calibration process, model parameters compensated for poor ET_p estimates so that comparable, and even better, predictions were obtained for poorer ET_p estimates. This illustrates that despite a multidimensional sensitivity analysis (i.e. "multi-calibration") previously applied (Vázquez and Feyen, 2002, 2003), the limitations of traditional model calibration, resulting particularly from the significant difficulty of a correct inspection of the model parameters space, led to biased conclusions, particularly, on the sensitivity of the parameters of the MIKE SHE ET_p module (Kristensen and Jensen, 1975). The current Monte Carlo based inspection of the model parameters space revealed the insensitivity of those parameters to the current hydrological modelling, besides the fact that parameter combination and interaction compensated for the use of less correct data.

In addition, independently of the ET_p data set, the current model of the study catchment has some problems to perform well as a continuous hydrograph simulator. Particularly, it has difficulties representing the main wet event in the evaluation period and also the low flow during the months of September and October 1985. These seem to be the consequences of errors in boundary conditions, errors in input data, as well as, model structure uncertainties and, particularly, scale discrepancies between model structure and the coarse grid sizes used in this modelling that most likely were magnified by the lumped daily resolution included in it. However, the separate contribution of these data uncertainties to the effective values of model parameters and to the total prediction uncertainty could not be accounted for, owing to different aspects, among which, the lack of access to the structure of the code is the most significant.

The GLUE analysis showed that, for the water routing related parameters and independently of the ET_p data set, the sensitivity of predictions to the horizontal saturated hydraulic conductivity of the Landeniaan layer (K_x(Ln)) is the highest. The vertical saturated hydraulic conductivity of the Landeniaan layer (K_z(Ln)) exhibits also some sensitivity. These aspects support the conclusions of preliminary calibration and sensitivity analyses on the important contribution of the Landeniaan (Ln) layer to the hydrological response of the Gete catchment model (Vázquez et al., 2002; Vázquez and Feyen, 2003). However, the sensitivities of the other process parameters to the model performance are very low and there was no clear optimal parameter set evident in the sample of models run.

Further, marked departures of the observed discharge from the predicted uncertainty bands were noticed, for some parts of the catchment-wide hydrograph. In this respect, it should be observed that the current calibration period is spring and summer, whilst the evaluation period is autumn and winter. However, the same criteria of rejection of parameter sets were used in either simulation period. Thus, it is likely that some parameter sets that do well in the first (i.e. calibration) period are over-conditioned on spring/summer data and, as such, are rejected in the second period (autumn/winter). Nevertheless, several parameter sets survived the rejection criteria in the second period, which enabled assessing the posterior likelihood.

The analysis depicted wide prediction intervals during some peak discharge periods. The 90% prediction bands associated with the simulation of the peak events emphasised the significant uncertainty attached to the simulated peaks. Wide prediction intervals were also obtained for the piezometric levels in the well locations in syntoni with the strong uncertainty associated with the geological and well-monitoring data and the incommensurability issues arising from comparing 600-m (and even 1200-m) grid simulation results and point-scale piezometric observations.

One of the main criticisms, against Monte Carlo (MC) based methods, is the number of simulation runs that are needed to sample appropriately the parameter space. This is still a constraint for MC applications of distributed models, despite the increasing availability of cheap PC-based parallel computing systems (although here the greater limitation was the single MIKE SHE software licence). However, recent research with respect to automatic optimisation of complex distributed models, such as Madsen et al. (2002), Madsen (2003) for the MIKE SHE code involved numerous simulation runs that make these methods comparable to the MC based methods in terms of computational effort, and this to find only one or two solutions of the Pareto front (Madsen, 2003). In general, uncertainty estimation methods (Binley et al., 1991; Klepper et al., 1991; Beven and Binley, 1992) have the potential of overcoming the inadequacies of traditional optimisation processes for spatially distributed models that are multi-input–output, by inspecting broadly the parameter space in a stochastic–deterministic context. The latter has been stressed and illustrated by the current research that, on the basis of a Monte Carlo sampling of the model parameters space, revealed some inadequacies in the conclusions derived from prior studies that included traditional optimisation (i.e. traditional sampling of the model parameters space).

Acknowledgements

This study was made possible by research financial support awarded to the first author from the IUPWARE Programme (Belgium), the Katholieke Universiteit Leuven (Project PDM/03/188) and the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA, Spain). Preparation of this manuscript was possible thanks to (i) the contract of the first author funded by the Universidad de Cuenca (UC; Ecuador); (ii) the working environment created by project "Establecimiento de modelos numéricos para casos selectos de la gestión de recursos hídricos" financed by the Research Directorate of the UC (DIUC); as well as (iii) the support to the second author from the PROMETEO Program of the National Secretary of Higher Education, Science, Technology and Innovation (SENESCYT; Ecuador).

References

Allen, G.R., Pereira, L.S., Raes, D., Martin, S., 1998. Crop evapotranspiration – guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, Rome.

Beven, K.J., 1989. Changing ideas in hydrology the case of physically based models. J. Hydrol. 105, 157–172.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51.

Beven, K.J., 2001a. How far can we go in distributed hydrological modelling? Hydrol. Earth Sys. Sci. 5 (1), 1–12.

Beven, K.J., 2001b. Rainfall-Runoff Modelling: The Primer. John Wiley and Sons, Chichester.
Beven, K.J., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrol. Process. 16, 189–206.
Beven, K.J., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36.
Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6 (3), 279–298.
Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. J. Hydrol. 249, 11–29.
Binley, A.M., Beven, K.J., Calver, A., Watts, L.G., 1991. Changing responses in hydrology: assessing the uncertainty in physically based model predictions. Water Resour. Res. 27 (6), 1253–1261.
Christiaens, K., Feyen, J., 2002. Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework. Hydrol. Process. 16 (2), 373–391.
DHI, 1998. MIKE-SHE v. 5.30 User Guide and Technical Reference Manual. Denmark: Danish Hydraulic Institute.
Doorenbos, J., Pruitt, W.O., 1977. Crop water requirements. FAO Irrigation and Drainage Paper 24, Rome.
Engman, E.T., 1986. Roughness coefficients for routing surface runoff. J. Irrig. Drain. Eng. 112 (1), 39–53.
Feyen, J., Vázquez, R.F., 2011. Modeling hydrological consequences of climate and land use change – progress and challenges. Maskana 2 (2), 83–100.
Feyen, L., Vázquez, R.F., Christiaens, K., Sels, O., Feyen, J., 2000. Application of a distributed physically-based hydrological model to a medium size catchment. Hydrol. Earth Syst. Sci. 4 (1), 47–63.
Gellens-Meulenberghs, F., Gellens, D., 1992. L'evapotranspiration Potentielle en Belgique: variabilite spatiale et temporelle. Internal publication, A-130. Belgian Royal Meteorological Institute, Brussels.
Groot, J.J.R., 1987. Simulation of nitrogen balance in a system of winter wheat and soil. Simulation rep. CABO-TT No.13. Centre for Agrobiological Research and Department of Theoretical Production Ecology. Wageningen Agricultural university, Wageningen, The Netherlands.
Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and non-commensurable measures of information. Water Resour. Res. 34 (4), 751–763.
Jayatilaka, C.J., Storm, B., Mudgway, L.B., 1998. Simulation of water flow on irrigation bay scale with MIKE SHE. J. Hydrol. 208, 108–130.
Klepper, O., Scholten, H., van de Kamer, J.P.G., 1991. Prediction uncertainty in an ecological model of the Oosterschelde estuary. J. Forecast. 10, 191–209.
Kristensen, K.J., Jensen, S.E., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. Nordic Hydrol. 6, 170–188.
Legates, D.R., McCabe, G.J., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrological and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.
Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv. Water Res. 26 (2), 205–216.
Madsen, H., Wilson, G., Ammentorp, H.C., 2002. Comparison of different automated strategies for calibration of rainfall-runoff models. J. Hydrol. 261, 48–59.
McMichael, C.E., Hope, A.S., Loaiciga, H.A., 2006. Distributed hydrological modelling in semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation. J. Hydrol. 317, 307–324.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I: a discussion of principles. J. Hydrol. 10, 282–290.
Pandey, M.D., van Gelder, P.H.A.J.M., Vrijling, J.K., 2003. Bootstrap simulations for evaluating the uncertainty associated with peaks-over-threshold estimates of extreme wind velocity. Environmetrics 14, 27–43.
Pickand, J., 1975. Statistical inference using extreme order statistics. Annals Stat. 3, 119–131.
Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. J. Hydrol. 198, 69–97.
Refsgaard, J.C., Storm, B., 1995. MIKE SHE. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, USA, pp. 809–846.
Thom, A.S., Oliver, H.R., 1977. On Penman's equation for estimating regional evaporation. Q. J. R. Meteorological Soc., Berkshire 103, 345–357.
Vázquez, R.F., 2003. Assessment of the performance of physically based distributed codes simulating medium size hydrological systems. Doctoral dissertation ISBN 90-5682-416-3. Department of Civil Engineering, K.U. Leuven, Belgium, p. 335.
Vázquez, R.F., Feyen, J., 2002. Assessment of the performance of a distributed code in relation to the ET_p estimates. Water Res. Manage. 16 (4), 329–350.
Vázquez, R.F., Feyen, J., 2003. Effect of potential evapotranspiration estimates on effective parameters and performance of the MIKE SHE-code applied to a medium-size catchment. J. Hydrol. 270 (4), 309–327.
Vázquez, R.F., Feyen, J., 2004. Potential evapotranspiration for the distributed modelling of Belgian basins. J. Irrig. Drain. Eng. 130 (1), 1–8.
Vázquez, R.F., Feyen, J., 2010. Rainfall-runoff modelling of a rocky catchment with limited data availability: defining prediction limits. J. Hydrol. 387, 128–140.
Vázquez, R.F., Feyen, L., Feyen, J., Refsgaard, J.C., 2002. Effect of grid-size on effective parameters and model performance of the MIKE SHE code applied to a medium sized catchment. Hydrol. Process. 16 (2), 355–372.
Vázquez, R.F., Willems, P., Feyen, J., 2008. Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteria approach. Hydrol. Process. 22 (13), 2159–2179.
Vázquez, R.F., Beven, K., Feyen, J., 2009. GLUE based assessment on the overall predictions of a MIKE SHE application. Water Resour. Manage. 23 (7), 1325–1349.
Vereecken, H., 1988. Pedotransferfunctions for the generation of hydraulic properties for Belgian soils. Faculty of Agricultural and Applied Biological Sciences. Katholieke Universiteit Leuven (K.U. Leuven), Leuven, Belgium, Doctoral dissertation, p. 254.
Xevi, E., Christiaens, K., Espino, A., Sewnandan, W., Mallants, D., Sorensen, H., Feyen, J., 1997. Calibration, validation and sensitivity analysis of the MIKE-SHE model using the neuenkirchen catchment as case study. Water Res. Manage. 11, 219–239.
Yu, P.Sh., Yang, T.Ch., Chen, Sh.J., 2001. Comparison of uncertainty analysis methods for a distributed rainfall-runoff model. J. Hydrol. 244, 43–59.