A conceptual grey rainfall-runoff model for simulation with uncertainty

Stefano Alvisi, Anna Bernini and Marco Franchini

ABSTRACT

This paper presents an approach based on grey numbers to represent the total uncertainty of a conceptual rainfall-runoff model. Using this approach, once the grey numbers representing the model parameters have been properly defined, it is possible to obtain simulated discharges in the form of intervals (grey numbers) whose envelope defines a band which represents the total model uncertainty. The application to a real case showed that the construction of this band, according to a rigorous application of grey number theory, involves long computational times. However, these times can be significantly reduced using a simplified computing procedure with minimal approximations in the quantification of the simulated grey discharge. Relying on this simplified procedure, the conceptual rainfall-runoff grey model was then calibrated in order to respect a predefined level of model uncertainty, i.e. the band obtained from the envelope of simulated grey discharges had to include an assigned percentage of observed discharges and was at the same time as narrow as possible. Finally, the uncertainty bands were compared with the ones obtained using a well-established approach for characterising uncertainty, the Generalised Likelihood Uncertainty Estimation (GLUE) method. The results of the comparison showed that the proposed approach may represent a valid tool for characterising the total uncertainty of a rainfall-runoff model.

Key words | grey numbers, rainfall-runoff model, uncertainty

Stefano Alvisi (corresponding author)
Anna Bernini
Marco Franchini
Department of Engineering,
University of Ferrara,
Italy
E-mail: *stefano.alvisi@unife.it*

INTRODUCTION

The characterisation/quantification of uncertainty in the context of hydrological modelling today represents a topic of primary importance (Li *et al.* 2010). There is now a broad consensus, both within the scientific community and among engineers/technicians involved in decision-making processes related to territorial management and protection, that an appropriate characterisation of uncertainty connected to a hydrological model is essential for both research and operational purposes, for example for flood management, the definition of alert thresholds, basin and reservoir management, etc. (Montanari 2011).

Generally speaking, in the hydrological realm the characterisation of uncertainty aims to provide a quantification of the credibility/reliability of one or more quantities. It is, however, important to clarify: (1) the context in which the uncertainty is evaluated, (2) which particular causes of error or sources of uncertainty are being considered and, finally, (3) what particular approach/methodology is being used to characterise them.

With respect to context (point 1), it is possible to distinguish two main situations in which the uncertainty connected to the hydrological model is evaluated, namely, (a) in the simulation phase and (b) in the forecasting phase. In the former case, the characterisation of uncertainty is aimed at quantifying the accuracy (reliability) of the outputs of a simulation model at different computational time steps, where all inputs to the model at the different points in time are known (Biondi *et al.* 2010). Forecasting uncertainty, in contrast, characterises the variability of the future value of a given forecast variable (discharge at the outlet of a basin, river stage, etc.), which is conditional on

doi: 10.2166/hydro.2012.069

the information that is available and known up to the time when the forecast is made (Todini 2008).

With respect to the various sources of uncertainty (point 2) it may be observed, on one hand, that some natural phenomena are characterised by an intrinsic variability which cannot be described with a deterministic approach and, on the other hand, the very structure and parameters of hydrological models are a source of error and hence of uncertainty, the models themselves always being a more or less simplified approximation of reality. Also, the data measured and used as inputs to the model for its calibration or validation can be affected by measurement errors. Given the presence of these various causes of error, it is possible to define different types of uncertainty, namely, (a) inherent randomness (e.g. the weather), (b) uncertainty related to the structure of the model, (c) uncertainty related to the model parameters, and (d) uncertainty related to the data. The various causes of error, i.e. these different components of uncertainty, lead to total error, i.e. the formation of total uncertainty associable with the model. It is important to observe, however, that the contribution of the different sources of error to total error is generally not known and, as highlighted by Gupta *et al*. (2005), breaking down total error into its individual components is often difficult, particularly in a hydrological context, where the models are not linear (Shrestha & Solomatine 2008). In practice, while total uncertainty can typically be characterised and estimated *a posteriori* based on the values furnished by the model and the corresponding observed values, identifying the individual components of uncertainty is generally much more complex, if not impossible, unless *a priori* assumptions are made (Montanari 2011).

Finally, regarding the characterisation of uncertainty (point 3), many methods have been developed and proposed in the scientific literature. Based on the indications of Shrestha & Solomatine (2008), these methods can be classified in general into six categories, namely, (1) analytical methods, in which the probability distributions of the inputs are used as the basis for analytically deriving the probability distributions of the output variable (see, for example, Tung 1996), (2) approximation methods, which instead of considering the entire probability distribution concentrate solely on the first and second order statistics of the output variable (First Order Analysis, Dettinger & Wilson 1981), (3) resampling-based methods, generally known as ensemble techniques or the Monte Carlo method, in which the probability distribution of the output variable of a model is obtained through numerous simulations (see, for example, Kuczera & Parent 1998), (4) Bayesian methods that combine Bayes' theorem with various simulation techniques to estimate or update the probability distributions of the parameters of a model and estimate the uncertainty of the model accordingly (Krzysztofowicz 1999), (5) methods based on the analysis of model errors (Montanari & Brath 2004), and (6) grey number or fuzzy set theory-based methods which provide a non-probabilistic approach for representing uncertainty through the concept of vagueness or imprecision (see, for example, Maskey *et al*. 2004; Alvisi & Franchini 2011).

It is important to observe that the above is indeed a possible classification of the methods, but the attribution of a method to one category or another may sometimes be open to discussion. This is the case, for example, with the GLUE (Generalised Likelihood Uncertainty Estimation, Beven & Binley (1992)) method, which can be considered as transversal among the different categories since, given its very general nature, it can be applied with any weight simulation approach, including Bayesian statistical methods using a specific error model, and fuzzy methods.

Furthermore, each method has its own advantages and limitations and frequently presents some parts which are subject to the user's arbitrariness. For instance, the methods belonging to the categories (1) and (3) require an *a priori* definition of the probability distributions of the parameters and inputs. Other methods, such as those in category (5), require the formulation of specific assumptions about errors (e.g. normality and homoscedasticity of errors), while the fuzzy set theory-based methods require the definition of a criterion for characterising the different levels of credibility (h-level) to be used to construct the fuzzy number (Shrestha & Solomatine 2006), or, alternatively, the definition of a criterion for selecting the form of the fuzzy number itself (Bardossy *et al*. 1990). Montanari (2011) develops a deep analysis of these aspects for many of the methods used for representing uncertainty and the reader can make reference to his paper for a general review of the subject.

This study presents a method for characterising uncertainty in the context of *simulation* of the rainfall-runoff process, where the attention is focused on *total* model *uncertainty*, and hence not on the individual components identified and discussed previously. The technique used is based on *grey number theory* (Deng 1982, 1987) and, with reference to the classification proposed above, it thus falls within category (6). This choice was based on the fact that grey number theory represents an appropriate tool for modelling uncertainties that do not originate from randomness but are caused by imprecise (or incomplete) knowledge about a real system (Jacquin & Shamseldin 2007). When developing a rainfall-runoff simulation model, the representation of the process is in general incomplete/simplified, particularly where a *conceptual explicit-soil-moisture-accounting* (ESMA, O'Connell 1991) model is used (examples of conceptual ESMA models are, for instance, the Dawdy & O'Donnell (1965) model, or the ARNO model by Todini (1996); for further examples see Singh (1995)). It is this imprecise representation of the process which is the major source of uncertainty in the simulation phase and not the randomness inside the physical process. Thus, grey theory can be used as a valid tool for modelling the related uncertainties.

In the context of water systems in general, and hydrological ones in particular, the grey number technique has been used in a number of studies in recent years to take account of the uncertainty/vagueness associated with a given process or given input and/or output variables of a model, for example in the modelling of aquifers (Wu *et al.* 2005), characterisation of river floodplains (Karmakar 2009), real-time river stage forecasting (Alvisi & Franchini 2012), characterisation of aspects related to water quality in a river system (Karmakar & Mujumdar 2006, 2007) or pipe roughness calibration in a water distribution system (Alvisi & Franchini 2010).

In the sections that follow, after briefly presenting the structure of the conceptual ESMA rainfall-runoff model adopted, we shall describe the basic notions associated with grey numbers and how they were used to represent the total uncertainty of the model. Finally, we shall present and discuss the results obtained when the proposed method was applied to a specific case study and compare them with those obtained applying the GLUE method (Beven & Binley 1992).

THE ADM MODEL

The rainfall-runoff model considered in this study is of the conceptual ESMA type and is called ADM – A Distributed Model – proposed by Franchini (1996). The model consists of two main blocks, the first representing the water balance and the second the component of transfer to the basin outlet (see Figure 1). Each main block is divided into two subparts.

As regards the water balance, two zones are identified, an upper zone and a lower one. The upper zone has rainfall and potential evapotranspiration as its inputs and produces surface and subsurface runoff. The input of the lower zone is percolation flow from the upper zone and its output is base runoff. The formation of surface runoff is based on the concept of probabilistic distribution of the soil storage capacity as per the Xinanjiang model (Zhao *et al.* 1980), where the surface runoff is the spatial integral of the infinitesimal contribution deriving from the saturated elementary areas.

The transfer of these water balance output components to the basin outlet takes place in two stages. The first regards the transfer along hill slopes towards the channel network and the second the transfer along the channel network towards the basin outlet. Both of these stages are modelled using a transfer function which represents the solution of the convection-diffusion equation assuming lateral uniformly distributed rainfall (Franchini & Todini 1988). The transfer functions associated with the two stages are characterised by two different pairs of parameters.

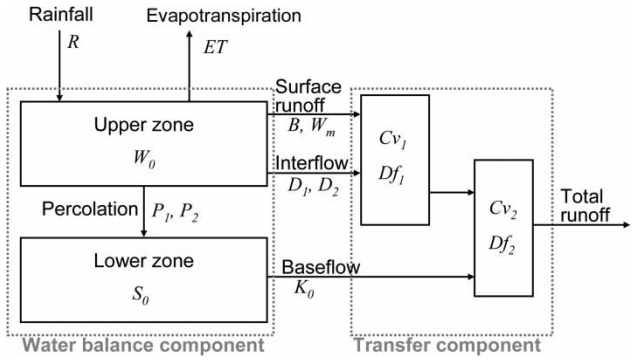

Figure 1 | Schematic representation of the ADM model.

Overall, the block representing the water balance is characterised by seven parameters, whereas the one representing transfer is characterised by four parameters, so that the model has a total of $n_{par} = 11$ parameters (see Table 1).

GREY NUMBERS

A grey number is a number whose exact value is unknown but which falls within an interval that is known (Liu & Lin 1998). With a grey number, therefore, the uncertainty associated with a given numerical quantity is represented by means of an interval whose upper and lower limits are known, whereas its distribution within the interval is not (Yang *et al.* 2004).

Table 1 | Parameters of the ADM model and range of variation defined on the basis of Franchini (1996)

Parameter	Description	Lower limit	Upper limit
W_m (mm)	Average storage capacity of the upper zone	50	300
B (–)	Shape parameter of the storage capacity probability distribution	0	1
D_1 (mm/h)	Maximum interflow value	0.1	6
D_2 (–)	Shape parameter of the interflow curve	2	10
P_1 (mm/h)	Maximum percolation value	0.0001	0.5
P_2 (–)	Shape parameter of the percolation curve	2	8
K_0 (h)	Depletion rate constant of the reservoir representing base runoff	12	144
Cv_1 (m/s)	Convection coefficient of the parabolic hydrograph for transfer along hillslopes	0.5	1
Df_1 (m^2/s)	Diffusion coefficient of the parabolic hydrograph for transfer along hillslopes	100	1,000
Cv_2 (m/s)	Convection coefficient of the parabolic hydrograph for transfer along the network	1	3
Df_2 (m^2/s)	Diffusion coefficient of the parabolic hydrograph for transfer along the network	1,000	5,000

A grey number $x^{\pm}$ can thus be mathematically expressed as (Cheng *et al.* 2002):

$$x^{\pm} = [x^-, x^+] = \{x \in x^{\pm} | x^- \le x \le x^+\} \quad (1)$$

where x indicates a closed and limited set of real numbers and x^- and x^+ are respectively the lower and upper limits of the interval.

In the context of grey numbers it is possible to define the four operations between two grey numbers $x^{\pm} = [x^-, x^+]$ and $y^{\pm} = [y^-, y^+]$ (where $0 \notin [y^-, y^+]$ in the case of division) as follows (Wang & Wu 1998):

$$\text{addition: } [x^-, x^+] + [y^-, y^+] = [x^- + y^-, x^+ + y^+] \quad (2)$$

$$\text{subtraction: } [x^-, x^+] - [y^-, y^+] = [x^- - y^+, x^+ - y^-] \quad (3)$$

$$\begin{aligned}\text{multiplication: } & [x^-, x^+] \times [y^-, y^+] \\ &= [\min\{x^- y^-, x^- y^+, x^+ y^-, x^+ y^+\}, \\ &\quad \times \max\{x^- y^-, x^- y^+, x^+ y^-, x^+ y^+\}]\end{aligned} \quad (4)$$

$$\text{division: } [x^-, x^+] \div [y^-, y^+] = [x^-, x^+] \times \left[\frac{1}{y^+}, \frac{1}{y^-}\right] \quad (5)$$

Furthermore, a function f of m grey numbers $x_1^{\pm}, x_2^{\pm}, \cdots x_m^{\pm}$ can be defined as follows (Yang *et al.* 2004):

$$\begin{aligned}\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^{\pm} = \Big[&\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^-, \\ &\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^+\Big]\end{aligned} \quad (6)$$

where $\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^{\pm}$ is the grey function, and $\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^-$ and $\left[f(x_1^{\pm}, x_2^{\pm}, \ldots, x_m^{\pm})\right]^+$ are respectively the minimum and maximum values of the function. Calculating the function f of m grey numbers $x_1^{\pm}, x_2^{\pm}, \ldots x_m^{\pm}$ involves looking for: (a) the set of real/crisp values $x_1, x_2, \ldots x_m$, where each real value is included within the corresponding grey number, i.e. $x_i^- \le x_i \le x_i^+$, which minimises the function; and (b) the set of real/crisp values, different from the previous one, which maximises it. In other words, it is necessary to solve two optimisation problems, one of minimisation and one of maximisation, of a

function of m variables, where each variable x_i is defined over an assigned interval $x_i^{\pm}$.

SIMULATION AND CALIBRATION OF THE GREY ADM MODEL

Based on the definition of grey numbers and the mathematics associated with them, we will show first of all how grey *inputs, parameters and initial conditions* can be used to perform a simulation with the ADM model and calculate the discharges at the basin outlet in the form of intervals, i.e. grey numbers (direct problem) which, in turn, summarise the total model uncertainty. We shall subsequently present a method for estimating the grey numbers representing (only) *the parameters* from a series of observed crisp/grey values of the output variable (inverse problem), so that the band representing the output produced by such grey parameters (direct problem) can still be interpreted as the total uncertainty of the model.

Simulation of the rainfall-runoff process using the grey ADM model

As described above under The ADM model, using the ADM model it is possible to represent the rainfall-runoff transformation process in a basin in a simplified manner. With reference to Figure 2(a), let T_{sim} be a simulation time window consisting of n_{sim} time intervals Δt; let $Q_{\text{sim},i}$ (where $i = 1{:}n_{\text{sim}}$) indicate the simulated crisp discharges at the basin outlet at the end of each of the time intervals Δt making up the window T_{sim}, i.e. at each of the times t_i where $i = 1{:}n_{\text{sim}}$; finally, let R_i and ET_i respectively represent the areal rainfall and evapotranspiration associated with the ith time interval Δt.

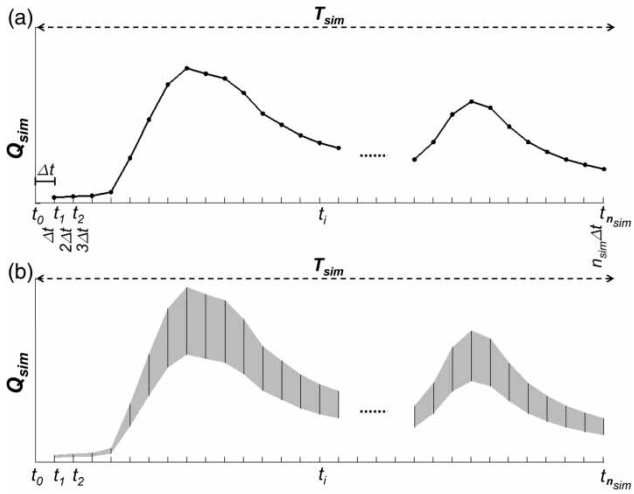

Figure 2 | Example of a time series of simulated discharges over a simulation time window T_{sim} given (a) crisp and (b) grey model parameters.

Formally, where W_m, B, D_1, D_2, P_1, P_2, K_0, $Cv1$, Df_1, Cv_2, Df_2 indicate the $n_{\text{par}} = 11$ crisp parameters of the model (see Table 1), and W_0 and S_0 the initial water content in the upper and lower zones, it follows that

$$Q_{\text{sim},i} = f(R_i, R_{i-1}, ., R_{i-m}, ET_i, ET_{i-1}, .. ET_{i-m}, W_0, S_0, W_m, B, D_1, D_2, P_1, P_2, K_0, Cv_1, Df_1, Cv_2, Df_2) \quad (7)$$

i.e. the simulated crisp discharge $Q_{\text{sim},i}$ (con $i = 1{:}n_{\text{sim}}$) is a function of (a) the inputs, rainfall and evapotranspiration, observed in the time intervals preceding time t_i until the system gains 'memory' (t_i–$m\Delta t$), (b) the initial system conditions (whose effect is reduced until disappearing as i increases, i.e. as we 'move away' from the initial point in time) and (c) the $n_{\text{par}} = 11$ parameters of the model.

Let us now assume that each of the inputs, parameters and initial water content conditions of the model is represented by a grey number, rather than a crisp number. At each computational time step the model output, i.e. runoff at the basin outlet at the generic time t_i, is a function of the inputs, parameters and initial conditions, so if each of these is represented by a grey number, the output will likewise be a grey number. In other words, at each time step the grey model does not furnish a crisp value of simulated discharge, but rather a grey number, i.e. an interval which reflects all the uncertainties of the parameters, inputs and initial conditions, *given the selected model*, and thus the envelope of the lower and upper extremes of the grey discharges defines a band, as shown in Figure 2(b), which can be interpreted as the total model vagueness/uncertainty.

The grey discharges can be computed relying on the definition of the function of grey numbers previously provided (see above under Grey numbers). Practically speaking, with reference to the generic time t_i (where $i = 1{:}n_{\text{sim}}$), we look for a set of real/crisp values of the inputs, parameters and initial conditions, each included in the corresponding grey number, such that the crisp simulation conducted starting

from the initial time t_0 provides the minimum simulated discharge $Q^{-}_{\mathrm{sim},i}$ at time t_i; similarly, we look for a set of real/crisp values of the inputs, parameters and initial conditions such that the crisp simulation provides the maximum simulated discharge $Q^{+}_{\mathrm{sim},i}$ at time t_i. For each simulation time point it is thus necessary to solve two optimisation problems, one of minimisation and one of maximisation. The discharge values obtained through such optimisation processes, $Q^{-}_{\mathrm{sim},i}$ and $Q^{+}_{\mathrm{sim},i}$, respectively represent the lower and upper extremes of the grey number $Q^{\pm}_{\mathrm{sim},i}$ representing the simulated discharge at time t_i.

Formally, using the notation of Equations (6) and (7), where $R_i^{\pm}$ and $ET_i^{\pm}$ respectively indicate the grey numbers representing the areal rainfall and evapotranspiration associated with the ith time interval Δt, $W_m^{\pm}, B^{\pm}, D_1^{\pm}, D_2^{\pm}, P_1^{\pm}, P_2^{\pm}, K_0^{\pm}, Cv_1^{\pm}, Df_1^{\pm}, Cv_2^{\pm}$, and $Df_2^{\pm}$ indicate the grey numbers representing the model parameters and $W_0^{\pm}, S_0^{\pm}$ indicate the grey numbers representing the initial water content conditions, it follows that:

$$Q^{\pm}_{\mathrm{sim},i} = \left[Q^{-}_{\mathrm{sim},i,}, Q^{+}_{\mathrm{sim},i}\right] = f\left(R_i^{\pm}, R_{i-1}^{\pm}, ., R_{i-m}^{\pm}, ET_i^{\pm}, ET_{i-1}^{\pm}, ET_{i-m}^{\pm}, W_0^{\pm}, S_0^{\pm}, W_m^{\pm}, B^{\pm}, D_1^{\pm}, D_2^{\pm}, P_1^{\pm}, P_2^{\pm}, K_0^{\pm}, Cv_1^{\pm}, Df_1^{\pm}, Cv_2^{\pm}, Df_2^{\pm}\right) \quad (8)$$

From an operational standpoint, the problem of direct simulation can be solved by pairing an optimiser with the ADM model as illustrated in Figure 3: with reference to the generic simulation time point t_i (where $i = 1{:}n_{\mathrm{sim}}$), the optimiser produces a generic solution made up of the real/crisp values of the inputs, parameters and initial conditions, where each real value is included in the corresponding grey number. Based on this set of values, the ADM simulation model is used to calculate the value of the simulated discharge $Q_{\mathrm{sim},i}$, which in turn represents the objective function to be minimised (maximised). Incidentally, the simulated discharge $Q_{\mathrm{sim},i}$ is obtained by carrying out the simulation against real/crisp inputs, parameters and initial conditions from the initial time t_0 *up to* time t_i. The minimisation (maximisation) process ends when one of the optimiser stopping criteria is met (e.g. convergence criterion, i.e. the minimisation (maximisation) process is stopped when the near optimal solution does not change significantly over an assigned number of iterations), so that the lower extreme $Q_{\mathrm{sim},i}\left(\text{upper } Q^{+}_{\mathrm{sim},j}\right)$ of the simulated grey discharge is given. It is important to point out that the procedure described must be repeated for each of the time points t_i (where $i = 1{:}n_{\mathrm{sim}}$) making up the simulation time window and twice for each time point, to identify first the minimum and then the maximum of the simulated discharge.

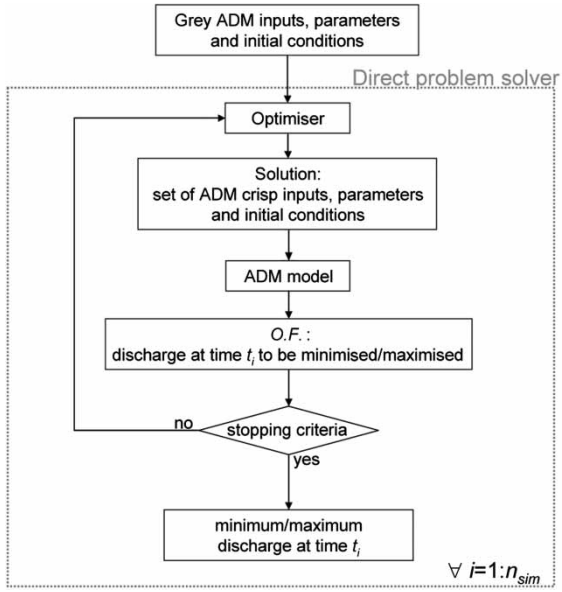

Figure 3 | Procedure for solving the direct problem (simulation) with the grey ADM model.

Calibration of the grey ADM model

In the solution of the direct problem previously described, inputs, parameters and initial conditions of the ADM model are assumed to be grey numbers, thus highlighting a situation where uncertainty is present in all of them. The model output, at each time instant, is a grey discharge which reflects the uncertainties in the parameters, inputs and initial conditions, *given the model*. This uncertainty is thus the *total model uncertainty*. It is clear that if the uncertainty in the inputs, parameters and initial conditions is increased, the grey discharge at each time instant will have a larger amplitude as a consequence of the grey mathematics. Now, in order to calibrate the ADM model we would like to fix the level of model uncertainty, i.e. we would like to fix a level of model uncertainty we consider acceptable. This acceptable level could be defined in such a way that

at least a given percentage (e.g. $PI = 95\%$) of observed discharges are included in the corresponding grey discharges produced by the model and, at the same time, these grey discharges are as narrow as possible. Incidentally, the observed discharges can be either crisp, thus assuming the corresponding measurement uncertainty as negligible, or grey, thus taking into account the corresponding measurement uncertainty, (for instance, quantified through the use of a grey rating curve in turn estimated through a procedure such as that proposed by Alvisi & Franchini (2013)). Incidentally, in this latter case, the grey discharges are said to be included within the band when both their extremes are within the band itself.

Once we have fixed this acceptable level of uncertainty in the model output (model uncertainty), which however does not imply in a univocally way the shape of the uncertainty band, we are interested in defining the grey parameters, inputs and initial conditions producing such an output. This would imply breaking down the global model uncertainty into the different components (input uncertainty, parameter uncertainty, initial condition uncertainty) but, as pointed out in the introduction, this would be very complex, if not impossible (Shrestha & Solomatine 2008; Montanari 2011). However, this is not our purpose since we are not interested in characterising the single uncertainty components but only the total model uncertainty selected. Thus we assume crisp inputs and initial conditions (as usually assumed in practical applications) and we delegate the grey parameters to produce grey discharges which overall respect the uncertainty level requested. In this situation the grey parameters do not represent the parameter uncertainty alone since other sources of uncertainty are included within them. In other words, these grey parameters are those which, *within the framework of grey mathematics,* allow the predefined level of total model uncertainty to be obtained when other sources of uncertainty (inputs and initial conditions) are considered as crisp values. This approach is certainly different from that proposed by Vrugt *et al.* (2003) which allows for the estimate of the parameter uncertainty alone by using a procedure based on the Metropolis algorithm (Metropolis *et al.* 1953). In this regard, it may be worth referring to Figure 11 of the paper by Vrugt *et al.* (2003) where a clear distinction between the uncertainty of the model and the uncertainty of the parameter estimates is made (the figure mentioned refers to the parameterisation of the rainfall-runoff model HYMOD proposed by Moore (1985)).

Formally speaking, the calibration process described above consists of looking for the lower and upper extremes of each of the $n_{\text{par}} = 11$ parameters of the model W_m^-, W_m^+, B^-, B^+, D_1^-, D_1^+, D_2^-, D_2^+, P_1^-, P_1^+, P_2^-, P_2^+, K_0^-, K_0^+, Cv_1^-, Cv_1^+, Df_1^-, Df_1^+, Cv_2^-, Cv_2^+, Df_2^-, Df_2^+, such that:

$$\sum_{i=1}^{n_{\text{cal}}} |Q_{\text{sim},i}^+ - Q_{\text{sim},i}^-| \text{ will be minimum} \tag{9}$$

subject to

$$\left(\frac{1}{n_{\text{cal}}}\sum_{i=1}^{n_{\text{cal}}} \delta_i\right) \cdot 100 \geq PI$$

where (10)

$$\delta_i = \begin{cases} 1 & \text{if } Q_{\text{sim},i}^- \leq Q_{\text{obs},i} \leq Q_{\text{sim},i}^+ \\ 0 & \text{otherwise} \end{cases}$$

where $Q_{\text{obs},i}$ (where $i = 1{:}n_{\text{cal}}$) is the observed crisp discharge (to be substituted by $Q_{\text{obs},i}^{\pm}$ if the observed discharge is grey) at the basin outlet at the end of each of the time intervals Δt making up the calibration time window T_{cal}, i.e. at each of the time points t_i where $i = 1{:}n_{\text{cal}}$ and PI is the pre-assigned percentage of observed values (both crisp or grey) to be included within the simulated grey discharges.

The actual algorithm used to solve the mathematical problem described by Equations (9) and (10) is provided in Appendix A (available online at http://www.iwaponline.com/jh/015/069.pdf). Here it is worth noting that from an operational standpoint, the procedure for solving the problem of calibrating the grey parameters of the model can be schematised as illustrated in Figure 4. The optimiser produces a generic solution in which the upper and lower extremes of the n_{par} grey numbers representing the model parameters are encoded. With that solution the direct problem is solved, that is, the grey discharges at the basin outlet $Q_{\text{sim},i}^{\pm}$ (where $i = 1{:}n_{\text{cal}}$) are calculated. Based on the simulated grey discharges and the corresponding observed discharges, the value of the objective function is calculated. The procedure is stopped when one of the optimiser stopping criteria is met (e.g. convergence criterion), thus providing the optimal values of the grey parameters. It is

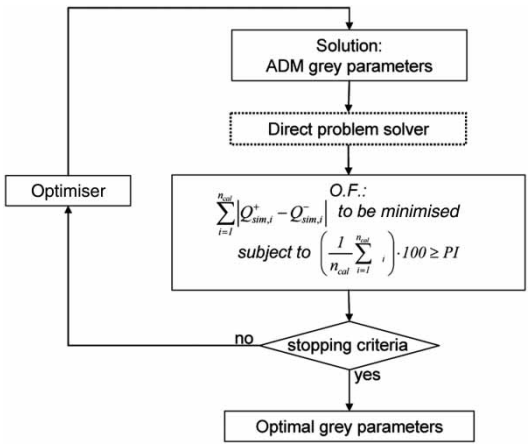

Figure 4 | Procedure for solving the inverse problem, i.e. calibration of the grey parameters of the ADM model.

important to note that the resolution of the inverse problem, for each hypothesised set of grey parameters, entails solving a direct problem, i.e. looking for the upper and lower extremes of the simulated discharge for each of the time points for which the observed discharge values ($\forall i$ where $i = 1{:}n_{cal}$) are available.

It is worth highlighting that it is possible to have a case where the grey discharges generated by the rainfall-runoff model are not able to include the pre-selected percentage of observed discharges, even assuming very large grey parameters. However, this is not a limit of the approach but instead a positive property which can indicate that something is not right, like data inconsistency and/or incorrect model structure assumptions, and thus, as indicated for these situations by Beven *et al.* (2011), deeper investigations are necessary before performing the model calibration.

CASE STUDY

The proposed approach was applied to a real case, the basin of the Sieve river, a tributary of the Arno river, having its outlet at Fornacina. The basin, situated in the Tuscan Apennines, extends for 831 km^2. The main course of the river is 56 km long and the time of concentration is about 10 h. The available data for this basin consists of the hourly time series of discharges at the basin outlet, areal rainfall (calculated using the Thiessen polygon method (Thiessen 1911) on the basis of point rainfall measured by 12 rain gauges within the basin) and evapotranspiration (estimated on the basis of temperature measurements in four different points of the basin, using the method suggested by Todini (1996)) pertaining to two distinct periods: the first extending from 1 December 1959 to 31 March 1960, for a total of 2,928 data items; the second from 1 January 1992 to 31 December 1996, for a total of 43,848 data items. It may be noted that the first period is fairly short. However, it has already been used in several other studies to set up reliable simulation and forecasting rainfall-runoff models (Franchini & Pacciani 1991; Solomatine & Dulal 2003; Shrestha & Solomatine 2006) since it contains a rich sequence of flood events of different entities separated by periods which are long enough to allow for complete development of the depletion phases; for further details concerning the basin and both datasets, reference can be made to the studies by Franchini & Pacciani (1991) and Toth & Brath (2007).

The results obtained when the procedures for solving the direct (simulation) and inverse (calibration) problems were applied to the case study are presented and discussed below. More precisely, first we describe the direct problem, solved by considering as grey numbers the parameters alone (while rainfall and evapotranspiration are represented through crisp numbers as done for the initial conditions), in order to highlight the numerical complexity of the procedure described above under Simulation of the rainfall-runoff process using the grey ADM model, for the grey simulation of discharges. In particular, at the end of this phase we propose a simplified grey simulation procedure useful for accelerating the computing times. Thereafter we describe the calibration procedure which, with the aid of the simplified simulation procedure, allows us to define the grey parameter values (inverse problem) which are able to produce the pre-selected total model uncertainty (see above under Calibration of the grey ADM model).

APPLICATION OF THE PROCEDURE TO SOLVE THE DIRECT PROBLEM

The procedure for solving the direct problem, described above under Simulation of the rainfall-runoff process using the grey ADM model, was applied in the case study with the aim of reproducing the (grey) trend in runoff

at the basin outlet over a time window corresponding to the period between 5/12/1959 and 28/2/1960, for a total of $n_{sim} = 2{,}061$ simulation time intervals. As mentioned above, rainfall, evapotranspiration and initial conditions were represented through crisp numbers whereas the parameters were assumed to be grey and the corresponding values were fixed making reference to the intervals defined for the same parameters in Franchini (1996), as shown in Table 1.

The values W_0 and S_0 of the initial water content in the upper and lower zones were both fixed as equal to 20 mm as in Franchini (1996).

From a computational standpoint, the direct problem was solved by relying on the scheme illustrated in Figure 3, in which the optimiser is the MATLAB^{TM} *fmincon* function based on Sequential Quadratic Programming (Powell 1978, 1983; Schitlowski 1985); at each time step the near optimal parameter set which produced the minimum (maximum) discharge at the previous time step was used as initial guess, thus allowing the optimisation process, which was stopped according to the 'convergence' criterion, to be speeded up significantly.

Figure 5 shows the trend in simulated grey discharges over the period 5/12/1959–28/2/1960; the intervals, or grey numbers representing the simulated discharges at the different points in time, are indicated by vertical black lines, and their envelope by a grey band. Figure 6 shows the crisp values of several model parameters provided by the optimiser to produce, for each t_i, the lower and uppers

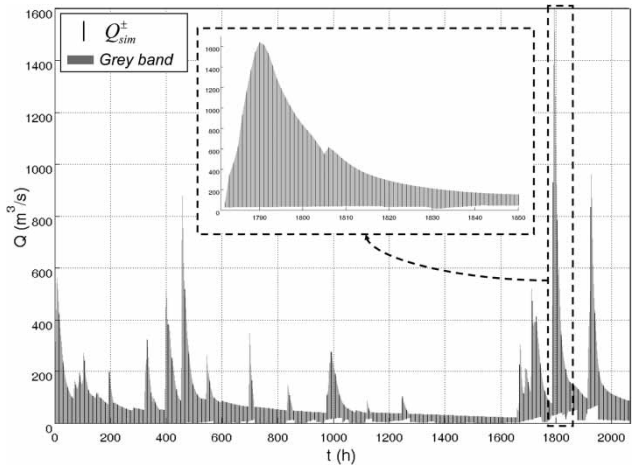

Figure 5 | Band for the period 5/12/1959–28/2/1960, simulated using the procedure for solving the direct problem with crisp initial conditions and inputs and grey parameters.

extremes of the simulated grey discharges of Figure 5. As can be observed, the crisp parameter values which provide the minimum and maximum values of the simulated discharges at the different time points very often coincide with the extremes (lower or upper) of the corresponding grey numbers of the parameters.

Finally, it should be observed that it takes approximately 6.5 h with a Pentium 4, 3.4 GHz processor to obtain the results of the direct grey simulation of Figure 5. Such a long computing time is a consequence of the fact that the simulation time window is made up of $n_{sim} = 2{,}061$ simulation time intervals and, under the procedure illustrated, it is necessary to solve $2n_{sim}$ optimisation problems, where each problem may require a greater or less number of direct crisp simulations (120 on average in the case considered).

This computational burden certainly represents a limit to the application of the proposed procedure for direct simulation and this problem become an even more significant one when we consider the calibration process, whose solution requires solving various direct simulation problems, as illustrated in Figure 4.

Based on these considerations, an expeditious procedure for solving the direct problem was developed as an alternative to the previously illustrated one, with the aim of significantly reducing the computational times.

The development of this procedure is substantially founded on the observation that (a) the crisp parameter values which produce the minimum and maximum values of simulated discharges over time tend to coincide with the extreme values of the corresponding grey numbers (see Figure 6) and (b) some model parameters may be more significant than others for the purpose of characterising the variability of simulated discharges. In particular, let $n^*_{par} < n_{par}$ be the number of the most significant model parameters (which can be identified via the Hornberger–Spear–Young (HSY) method (Hornberger & Spear 1981; Young 1983) as described in Appendix B – available online at http://www.iwaponline.com/jh/015/069.pdf). The expeditious procedure for solving the direct problem can be summed up as follows:

1. All of the possible combinations of parameters that can be obtained are considered, with each of the n^*_{par} most significant parameters being set at the lower or upper

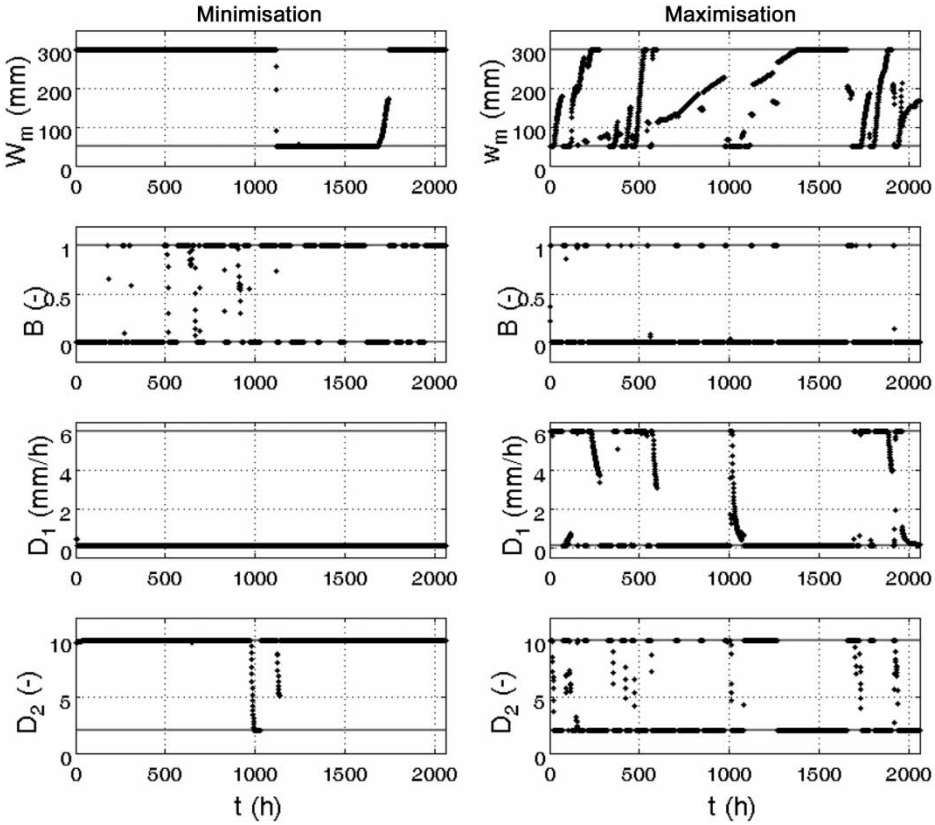

Figure 6 | Crisp values (black dots) of the parameters W_m, B, D_1 and D_2 provided by the optimisation process for identification of the lower (left column) and upper (right column) extremes of the grey discharges for the period 5/12/1959–28/2/1960. The grey lines represent the lower and upper extremes of the grey numbers of the parameters within which the crisp numbers are sought (see Table 1).

extreme of the corresponding grey number. The least significant parameters are set on the middle value of the corresponding grey number (or hypothesised interval of variation). Overall there are $2^{n^*_{\text{par}}}$ different combinations of crisp parameters.

2. For each combination of crisp parameters a direct crisp simulation is performed with the ADM model over the entire simulation time window.
3. For each simulation time point, i.e. for every t_i (where i=1:n_{sim}), the minimum value (which will thus represent $Q^-_{\text{sim},i}$) and maximum value (which will thus represent $Q^+_{\text{sim},i}$) are identified among the $2^{n^*_{\text{par}}}$ simulated discharge values.

This expeditious procedure was applied in order to simulate the grey discharges over the same time window considered earlier, the period between 5/12/1959 and 28/2/1960.

From an operational standpoint, the most significant parameters were identified, as previously mentioned, by using the HSY method (Hornberger & Spear 1981; Young 1983) described in Appendix B, which led to the identification of the following $n^*_{\text{par}} = 8$ most significant parameters: W_m, B, D_1, D_2, P_1, P_2, K_0, Cv_2.

It is worth highlighting that this expeditious procedure, given its enumeration nature, produces solutions where the minimum and the maximum value of the discharge in each simulation time point are univocally identified (given the assumption that only the values on the extreme ends of the grey parameters are considered). These values, in turn, well approximate the corresponding minimum and maximum discharges representing the extremes of the grey discharges identified as near optimal solutions by using the optimiser. In Figure 7 the band obtained with the expeditious procedure described above is compared with the band obtained using the procedure described above under

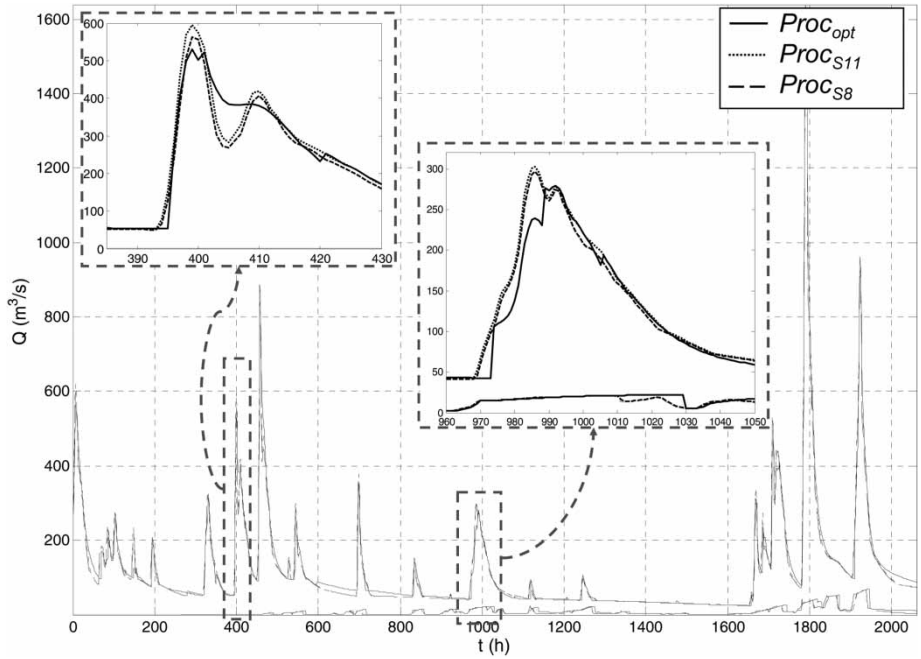

Figure 7 | Comparison between the band obtained using the procedure for solving the direct problem, described in Simulation of the rainfall-runoff process using the grey ADM model ($Proc_{opt}$), and those obtained with the expeditious procedure considering $n_{par} = 11$ parameters and only the $n^*_{par} = 8$ most significant parameters.

Simulation of the rainfall-runoff process using the grey ADM model (based on the optimiser). These bands are also compared with the band that would be obtained if the expeditious procedure were applied without preselecting the most significant parameters, that is, assuming $n^*_{par} = n_{par} = 11$ and thus carrying out 2^{11} simulations.

As may be observed, the application of the expeditious procedure, both with and without preselection of the most significant parameters ($n^*_{par} = 8$ and $n^*_{par} = n_{par} = 11$, respectively), leads to the identification of a discharge band that is *very similar* to the one resulting from the procedure described above under Simulation of the rainfall-runoff process using the grey ADM model, based on solving two optimisation problems for each point in time, assuming that the crisp parameters can take on any value within the corresponding grey number. More specifically, it was observed that the expeditious procedure without preselection of the most significant parameters, only for several times t_i associated with the recession phase of the event, leads to the identification of upper extremes of the grey numbers (slightly) lower than those that would be obtained by solving the maximisation problem through optimisation. In contrast, at the times corresponding to the peak, the upper extremes of the grey numbers obtained using the expeditious procedure with $n^*_{par} = n_{par} = 11$ are often (slightly) higher than those provided by the optimisation process and this can be ascribed to the difficulty that the optimiser (of a local type) has in identifying the maximum without being trapped into local 'extremes' when processing 11 parameters simultaneously. Incidentally, in order to avoid this latter problem, direct simulation tests were carried out with the procedure described above under Simulation of the rainfall-runoff process using the grey ADM model, also using the SCE-UA algorithm (Shuffled Complex Evolution – University of Arizona; Duan *et al.* 1992) as the optimiser; however, the results obtained were substantially equivalent and the computational times even longer. Indeed, the results highlighted an inner difficulty also for this latter algorithm in dealing with such a complicated 'function' represented by the 11-parameter-rainfall-runoff model, though we cannot exclude that modifying some parameters of the optimiser, e.g. significantly increasing the number of complexes or the number of individuals/solutions considered at each iteration, (slightly) more precise results could have been obtained but at the cost of much longer computational time for each optimisation. In

any case, the results of the optimiser, either based on the MATLAB™ *fmincon* function or the SCE-UA algorithm, are only useful to validate the *reliability* of the 'expeditious procedure' (see Figure 7), which is the one to be used within the framework of the calibration phase since it is the only one characterised by very short computational times, without requiring the identification of the exact and univocal solution, at each time step, but instead accepting the good reproduction of the near optimal solution produced by the optimiser.

As regards the computational times, an analysis showed that using the expeditious procedure, both without and above all with a preselection of the most significant parameters, brings considerable benefits. In fact, the computing times go from 23,300 s (about 6.5 h), which is how long it takes to complete the procedure with two optimisations (performed with the MATLAB™ *fmincon* function) at every computing step assuming that the crisp parameters can take on any value within the corresponding grey number, to 140 s for the expeditious procedure without preselection of the most significant parameters, and 12 s for the expeditious procedure considering only the $n^*_{\mathrm{par}} = 8$ most significant parameters.

In short, the evident benefits in terms of decreasing the computational times and the limited approximations with respect to the widths of the grey numbers of simulated discharges show that the expeditious procedure proposed, particularly the one envisaging a preselection of the most significant parameters, represents a valid tool for solving the direct problem in operatively acceptable times. For this reason the calibration process will refer to this procedure for solving its own 'internal' direct problem.

APPLICATION OF THE PROCEDURE FOR SOLVING THE INVERSE PROBLEM

The grey values of the $n^*_{\mathrm{par}} = 8$ most significant parameters of the ADM model previously identified were calibrated according to the procedure described above under Calibration of the grey ADM model, using the expeditious procedure described in the preceding section to solve the direct problem. The remaining three parameters were fixed on the middle value of the defined interval indicated in Table 1. The calibration was repeated twice, the first time using data extracted from the time series for the period between December 1959 and March 1960, and the second time with data extracted from the time series for the period between January 1992 and December 1996. The choice of using the data regarding these periods to perform two different calibrations was based on the consideration that the quality of the data is different; in particular the series relating to the period between January 1992 and December 1996 can be considered of 'lower quality', as significantly discordant patterns can be identified for a number of events as regards the entity of rainfalls and the corresponding runoffs.

In detail, with reference to the time series for the period between December 1959 and March 1960, a calibration time window (T_{cal}^{59-60}) was chosen corresponding approximately to the first month of data (see Table 2), preceded by a warm-up period of about 50 h serving to reduce the effect of the initial values of water content W_0 and S_0 in the upper and lower zones, both fixed as equal to 20 mm. The remaining data, forming a series hereinafter called T_{val}^{59-60}, were used in the procedure validation phase.

With reference to the time series for the period between January 1992 and December 1996, the calibration time window T_{cal}^{92-96} was defined by joining four different time windows (see Table 2) identified in such a way as to contain significant events. Also in this case a prior warm-up period of 300 h was considered for each time window in order to reduce the effect of the initial values of water content W_0 and S_0 in the upper and lower zones, both fixed as equal to 20 mm. The remaining data, aggregated to form a series hereinafter called T_{val}^{92-96}, were used in the procedure validation phase. More precisely, using the grey parameters produced by the calibration process, the simulation was performed over the entire period from January 1992 to December 1996 and the performance indexes were computed with reference to the period represented by the union of the five time windows indicated in Table 2 (in the first time window a warm-up period of 300 h was considered).

In both cases the observed discharges were assumed to be represented by crisp numbers and thus without errors,

Table 2 | Starting and ending date of the calibration and validation time windows for the two periods considered

Period: December 1959–March 1960	
Calibration time period (T_{cal}^{59-60}) (the first 50 h represent the warm-up period)	01/12/1959 h: 0:00–05/01/1960 h: 10:00
Validation time period (T_{val}^{59-60})	05/01/1960 h: 11:00–31/03/1960 h: 23:00
Period: January 1992–December 1996	
Calibration time period (T_{cal}^{92-96}): composed of four time windows (the first 300 h of each period represent the warm-up period)	27/09/1992 h: 20:00–09/01/1993 h: 24:00 24/10/1993 h: 12:00–19/01/1994 h: 0:00 27/12/1994 h: 16:00–25/03/1995 h: 4:00 26/11/1995 h: 00:00–11/04/1996 h: 12:00
Validation time period (T_{val}^{92-96}): composed of five time windows (the first 300 h of the first time window represent the warm-up period)	01/01/1992 h: 0:00–27/09/1992 h: 19:00 10/01/1993 h: 1:00–24/10/1993 h: 11:00 19/01/1994 h: 1:00–27/12/1994 h: 15:00 25/03/1995 h: 5:00–25/11/1995 h: 23:00 11/04/1996 h: 13:00–31/12/1996 h: 23:00

leaving for a future study a grey-based numerical investigation concerning the characterisation of the observed discharge uncertainty and its effects on the grey ADM model parameterisation; the grey values of the $n^*_{\text{par}} = 8$ parameters undergoing calibration were sought within the intervals indicated in Table 1, subject to the constraint that the bands representing the simulated grey discharges had to include at least a percentage $PI = 95\%$ of the observed crisp discharge values associated with the calibration time window.

From a computational standpoint, in order to solve the inverse problem (i.e. identification of the grey parameters) use was made of the SCE-UA (Duan *et al.* 1992) algorithm to provide an initial near optimal solution which was subsequently refined using the MATLAB™ *fmincon* function based on Sequential Quadratic Programming (Powell 1978, 1983; Schitlowski 1985). In both cases the decision variables, i.e. the extreme values of the grey numbers representing the $n^*_{\text{par}} = 8$ parameters undergoing calibration, were encoded within the optimisation solver by using the method described by Alvisi & Franchini (2010, see Appendix, p. 445) and the convergence criterion was used to stop the optimisation process.

Discussion of the results

Table 3 shows the values of the grey parameters obtained from the calibrations performed on the time windows T_{cal}^{59-60} and T_{cal}^{92-96}. It may be immediately observed that the grey numbers obtained for the parameters in the calibration with the time window T_{cal}^{92-96} all correspond to a wider interval and in general include the grey values of the parameters calibrated on the time window T_{cal}^{59-60}; in other words, the parameters calibrated on the time window T_{cal}^{92-96} are characterised by greater vagueness compared to those calibrated on the time window T_{cal}^{59-60}. This is understandable considering the 'inferior quality' of the data set pertaining to the period 1992–1996, mainly due to inconsistencies between rainfall and discharge data and/or presence of data relative to phenomena which cannot be described by the rainfall-runoff model used, given its structure (for example, snowfall and snow-melting events), and hence the need to define wider intervals for the parameters in order to explain the lower representativeness of the simulation performed with the ADM model.

Table 4 shows the average widths of the bands obtained from the envelope of the grey numbers representing the

Table 3 | Grey values of the model parameters obtained from calibrations on the time windows T_{cal}^{59-60} and T_{cal}^{92-96}

	W_m (mm)	B (–)	D_1 (mm/h)	D_2 (–)	P_1 (mm/h)	P_2 (–)	K_0 (h)	Cv_2 (m/s)
Cal. on T_{cal}^{59-60}	[74, 279]	[0.1, 0.8]	[3.9, 5.5]	[6.4,7.5]	[0.37, 0.4]	[3.8, 6.5]	[92,101]	[2.21, 2.23]
Cal. on T_{cal}^{92-96}	[69, 282]	[0.1, 0.9]	[0.5, 5.6]	[2.6, 9.4]	[0.03, 0.5]	[2.5, 7.6]	[12,134]	[1.1, 2.9]

Table 4 | Average width (*AW*) of the bands and corresponding percentages of observed values included (*POC*) obtained using the procedure based on grey numbers (assuming $PI = 95\%$ in the calibration phase) in relation to the different calibration and validation time windows

	Calibration		Validation	
	AW **(m^3/s)**	***POC*** **(%)**	***AW*** **(m^3/s)**	***POC*** **(%)**
Cal. on T_{cal}^{59-60} → Val. on T_{val}^{59-60}	68.4	95.1	51.9	81.0
Cal. on T_{cal}^{92-96} → Val. on T_{val}^{92-96}	74.3	95.0	35.5	75.2
Cal. on T_{cal}^{59-60} → Val. on T_{val}^{92-96}	68.4	95.1	17.8	49.5
Cal. on T_{cal}^{92-96} → Val. on T_{val}^{59-60}	74.3	95.0	96.1	93.3

corresponding simulated grey discharges and the percentage of observed values included within the bands. The average width AW of the bands is defined as (Zhang *et al.* 2009):

$$AW = \frac{1}{n_{\text{obs}}} \sum_{i=1}^{n_{\text{obs}}} |Q_{\text{sim},i}^{+} - Q_{\text{sim},i}^{-}| \tag{11}$$

whereas the percentage of observed values included within the band (percentage of coverage (*POC*)) is defined as (Zhang *et al.* 2009):

$$POC = \frac{1}{n_{\text{obs}}} \sum_{i=1}^{n_{\text{cal}}} \delta_i$$

where (12)

$$\delta_{\text{i}} = \begin{cases} 1 & \text{if } Q_{\text{sim},i}^{-} \le Q_{\text{obs},i} \le Q_{\text{sim},i}^{+} \\ 0 & \text{otherwise} \end{cases}$$

where n_{obs} represents the number of observed data in the time window considered.

In particular, Table 4 shows the results of the two calibrations performed on the time windows T_{cal}^{59-60} and T_{cal}^{92-96} and of the corresponding validations performed respectively on the time windows T_{val}^{59-60} and T_{val}^{92-96}. It also shows the results of the model validation for the time window T_{cal}^{92-96}, which was performed on the basis of the grey parameters calibrated on the time window T_{cal}^{59-60}, and results of the model validation for the time window T_{val}^{59-60}, which was performed on the basis of the grey parameters calibrated on the time window T_{cal}^{92-96}.

Analysing the results obtained in the calibration phase, we may observe that the percentage of observed values actually included corresponds to the one imposed, thus demonstrating the correctness of the calibration procedure adopted. If we consider, on the other hand, the validation results obtained for the time windows T_{val}^{59-60} and T_{val}^{92-96} using the parameters calibrated on the time windows T_{cal}^{59-60} and T_{cal}^{92-96}, respectively, we note that *POC* falls to around 81% for the first data set (relative to 1959–1960) and around 75% for the second data set (relative to 1992–1996, that is the one of 'inferior quality'). A reduction of *POC* with respect to the expected value 95% is not surprising: the fact that a performance index deteriorates in validation phase is indeed always expected when model parameterisation is performed through a calibration process and in fact this is systematically observed also in the case of crisp rainfall-runoff models. Thus, in the case of a grey rainfall-runoff model, the grey parameters calibrated in order to produce grey discharges containing at least 95% of observed discharges (being as narrow as possible) can be considered, in validation phase, as those which are expected to produce grey discharges containing ideally (but usually less than) 95% of observed discharges according to the quality of the data used and the 'flexibility' of the model to capture hydrological conditions not included in the calibration period (of course, this fraction may be even greater than that requested during calibration, if the hydrological conditions dominating the validation data set coincide with those for which the 'calibrated' model has better performances). Figure 8 shows the corresponding trend in the band obtained for the validation time window T_{val}^{59-60} using the grey procedure with parameters calibrated on the time window T_{cal}^{59-60}. Similarly, Figure 9 shows the trend in the band obtained for the validation time window T_{val}^{92-96} using the grey procedure with parameters calibrated on the time window T_{cal}^{92-96}.

With reference to Figure 8, it may be observed that the fact that the *POC* value reported in Table 4, equal to 81%, is lower than the expected 95% is mainly due to the presence of some well-defined time intervals (see intervals

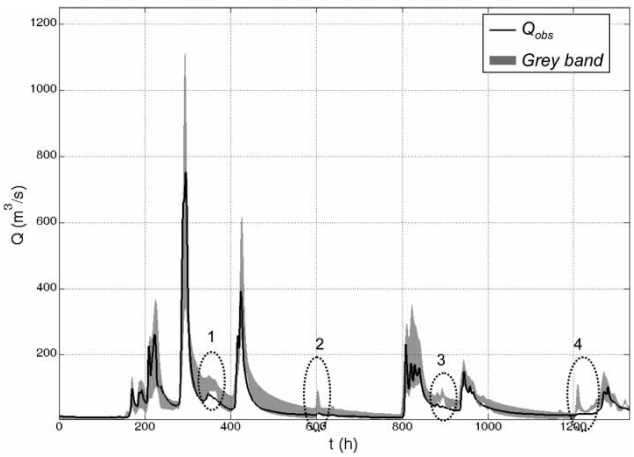

Figure 8 | Validation time window.T_{val}^{59-60} Band produced by the grey ADM model with grey parameters calibrated on the time window T_{cal}^{59-60}.

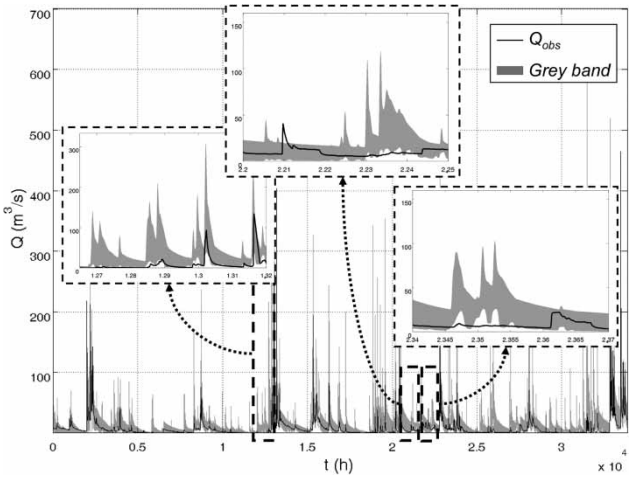

Figure 9 | Validation time window T_{val}^{92-96}. Band produced by the ADM model with grey parameters calibrated on the time window T_{cal}^{92-96}.

highlighted in Figure 8) for which the outputs of the model in terms of grey runoffs deviate significantly from the observed values. These time intervals are concentrated in the recession phases (see for example intervals 1 and 3) or are associated with simulated events, of modest entity, which do not have any corresponding observed/recorded event (see for example intervals 2 and 4). Furthermore, even though the proposed approach does not explicitly impose, in calibration phase, the inclusion of all the (main) peaks, it leads to uncertainty bands which include the peaks of all the major events. This can be explained recalling that the calibration procedure adopted requires a given percentage of observed discharges of the calibration data set to be contained within the grey discharges and that this data set contains many more medium and low flows than peaks, as typically occurs for continuous simulations; thus the constraint of containing at least a given percentage of observed discharges is mainly driven by the medium and low flows. From a numerical point of view, a conceptual ESMA rainfall-runoff model is typically not very sensitive to inputs (rainfall) corresponding to these low flows and thus, in order to produce bands which include a large part of these values, the calibration procedure produces grey parameters whose amplitude is wide enough to generate grey discharges which tend to increase their amplitude as the forcing to the model (rainfall input) increases thus making easier the inclusion of several observed peaks, as indeed happens in Figure 8.

These considerations apply and take on even greater significance if we consider the band associated with the validation time window T_{val}^{92-96} shown in Figure 9, for which the *POC* of the simulated band was found to be 75%. In this case, in fact, it is possible to identify numerous time intervals in which the responses of the simulation model, although grey, are in 'disagreement' with the observed discharges. Figure 9 shows, by way of example, zoom-in views of several of these time intervals. It is important to observe that in this case also the time intervals in question are concentrated in points corresponding to low observed discharge values; however, their high frequency (greater compared to the dataset for the period 1959–1960) results in a significant reduction in *POC* compared to the expected value of 95%.

The effect of the different 'data quality' on *POC* in the validation phase is still more evident if we consider the results of the validation on the time window T_{val}^{92-96}, performed using the calibrated parameters for the time window T_{cal}^{59-60}. As may be observed from Table 4, *POC* falls even further, to around 50%. This is understandable considering that the grey parameters calibrated with a 'good' dataset (i.e. a data set with consistent rainfall-runoff data and without data relative to phenomena which cannot be described by the rainfall-runoff model used) are 'narrow' (see Table 3) and therefore, once used in a simulation with a poorer quality dataset, are able to explain the trend in observed values to a lesser degree (as also shown

by the average width of the corresponding simulated band, which, the time window (T_{val}^{92-96}) being equal, is reduced by about one half, from around 35 m^3/s in the case of parameters calibrated on the time window T_{cal}^{92-96} to around 18 m^3/s in the case of parameters calibrated on the time window (T_{cal}^{59-60}). The opposite applies for the validation results with respect to the time window T_{val}^{59-60}, where the validation was performed using the parameters calibrated on the time window T_{cal}^{92-96}. In this case the average bandwidth obtained in the validation phase (T_{val}^{59-60}), based on parameters calibrated on a dataset (rainfall-discharge) affected by a high degree of uncertainty, increases significantly (from around 52 m^3/s for the parameters calibrated on the time window T_{cal}^{59-60} to around 96 m^3/s for the parameters calibrated on the time window T_{cal}^{92-96}), so that a high percentage of observed values, close to 93% can be included within the band.

In short, percentages of observed values included in the validation periods depend on the 'quality of the data', i.e. on the consistency between rainfall and runoff observed values and on the hydrological conditions that are present in the validation data set, i.e. on the fact that these data are relative to phenomena which can be modelled by the rainfall-runoff model used, given its structure, and/or show similar dynamics already considered in the calibration period; in fact, if the procedure is parameterised with a good dataset, using a dataset of 'inferior quality' in the validation phase will make it difficult to correctly represent the vagueness/uncertainty associated with these latter data.

To conclude the analysis of the results obtained with the proposed procedure, we present, by way of comparison, the results that would be obtained using a method widely employed in the field of hydrology to characterise global uncertainty in the simulation phase, namely, GLUE (Beven & Binley 1992). As it is a very well-known, widely applied method, the main information concerning it is summed up in Appendix C (available online at http://www.iwaponline.com/jh/015/069.pdf). Here we shall only note that with this method it is assumed that, once the structure of a given model has been assigned, there will not exist only one optimal set of crisp model parameters, but rather there may exist a number of equally 'good' sets; this concept, called 'equifinality' (Beven 1993), is implemented by evaluating different sets of parameter values randomly generated within predefined probability distributions and associating with each set a likelihood measure on the basis of which it is possible to derive the uncertainty band (Blasone *et al.* 2008) (see Appendix C). It is worth noting that this uncertainty band is thus formed by behavioural realisations alone while the band produced by the grey ADM model, consistently with the grey paradigm used to derive it, should not be read as an ensemble of single crisp model realisations but instead as a single grey realisation produced by the grey ADM model (i.e. each band is formed by vertical segments representing the grey discharges computed at each time instant).

The application of the GLUE method to the case study considered here entailed looking for $n_{sp} = 2{,}000$ sets of behavioural parameters (see Appendix C), assuming for each significant parameter a uniform probability distribution whose limits correspond with the intervals indicated in Table 1, whereas the non-significant parameters were fixed to the corresponding central value as was done for the application of the grey procedure; the Nash–Sutcliffe (*NS*) coefficient (Nash & Sutcliffe 1970) was used as a likelihood indicator and a threshold value *LT* (see Appendix C) equal to 0.3 was fixed. This procedure was calibrated on the time windows T_{cal}^{59-60} and T_{cal}^{92-96} and validated on the corresponding time windows T_{val}^{59-60} and T_{val}^{92-96}. The corresponding average bandwidths and percentages of coverage assuming the uncertainty level $\alpha = 0.95$ (see Appendix C) are indicated in Table 5. With reference to the calibration phase, it can be observed that both for the period T_{cal}^{59-60}

Table 5 | Average width (*AW*) of the bands and corresponding percentages of observed values included (*POC*) obtained using the GLUE procedure (assuming a threshold value $NS = 0.3$ and $\alpha = 95\%$) in relation to the different calibration and validation time windows

			Calibration		Validation	
	NS	α	***AW*** (m^3/s)	***POC*** (%)	***AW*** (m^3/s)	***POC*** (%)
Cal. on T_{cal}^{59-60} → Val. on T_{val}^{59-60}	0.3	0.95	75.8	94.3	60.3	84.9
Cal. on T_{cal}^{92-96} → Val. on T_{val}^{92-96}	0.3	0.95	48.4	85.2	23.6	58.2

and T_{cal}^{92-96} the *POC* is slightly lower than 95%. In particular, for the calibration period T_{cal}^{59-60} the *POC* index is only slightly lower than 95%, at around 94%, and the average width *AW* is equal to 75.8 m^3/s, i.e. slightly greater than the corresponding *AW* of the grey approach (see Table 4), whereas for the calibration period T_{cal}^{92-96} the *POC* index is definitely lower than 95%, at around 85%, while the average width *AW* is lower than the corresponding one of the grey approach (see Table 4). Indeed there is no expectation with GLUE that the level of uncertainty α of the simulations should represent the coverage of the observations, as observed by Beven (2006). However, since the percentage of observed values actually included within the bands is a good practical indicator to assess the method's capacity to describe the variability of the observed values (Xiong & O'Connor 2008), in order to make a fair comparison with the grey approach, we selected both the level of uncertainty α and the *NS*-threshold values in such a way that the band produced by the GLUE method in the calibration phase had a *POC* equal to that of the grey approach, i.e. 95%. With reference to the period 1959–1960 in Table 6, *POC* and *AW* produced by the GLUE method for different levels of uncertainty α and *NS*-threshold values are summarised. As can be observed, for a given level of uncertainty α (e.g. $\alpha = 0.95$) changes in the *NS*-threshold value produce slight changes in the *POC* index: in fact, as the threshold decreases from $NS = 0.3$ to $NS = 0.1$, the *POC* index slightly increases from 94.3 to 94.8%. This finding is in line with the results of Xiong & O'Connor (2008) concerning the analysis of the effects of the threshold value on *POC*. On the other hand, for a given threshold (e.g. $NS = 0.3$) changes in the level of uncertainty α produce significant changes in *POC*: in fact, for α increasing from 0.95 to 0.99 the *POC* index increases from 94.3 to 97.3%. In practice, as can be observed in Table 6, a *POC* index around 95% for the calibration period T_{cal}^{59-60} could be obtained by setting the threshold value to $NS = 0.3$ and assuming $\alpha = 0.97$.

Figure 10 shows the band obtained for the validation time window T_{val}^{59-60} using the GLUE procedure calibrated on the time window T_{cal}^{59-60} assuming the threshold value $NS = 0.3$ and $\alpha = 0.97$ (i.e. those that guarantee a *POC* almost equal to 95% in the calibration phase). Keeping in mind the previously mentioned different conceptual background characterising the GLUE and grey bands, i.e. in the GLUE case the band is an ensemble of distinct behavioural realisations of the crisp model whereas in the grey approach the band is a single grey realisation produced by the grey ADM model, if we compare the band of Figure 10 obtained using the GLUE procedure with the corresponding band of Figure 8 obtained using the grey model we can observe that the two procedures provide similar results in terms of band trends and widths. In particular, the problems encountered with the grey method in relation to the four numbered small events, already commented with reference to Figure 8,

Table 6 | Average width (*AW*) of the bands and corresponding percentages of observed values included (*POC*) obtained using the GLUE procedure for the period 1959–60 (Cal. on T_{cal}^{59-60} → Val. on T_{val}^{59-60}) assuming different *NS*-threshold values and uncertainty levels α

		Calibration		Validation	
NS	α	***AW*** (m^3/s)	***POC*** (%)	***AW*** (m^3/s)	***POC*** (%)
0.3	0.95	75.8	94.3	60.3	84.9
0.2	0.95	77.3	94.4	61.0	84.8
0.1	0.95	78.2	94.8	61.4	84.2
0.3	***0.97***	***81.9***	***94.9***	***65.3***	***85.9***
0.2	0.97	82.7	95.2	65.9	86.7
0.1	0.97	84.0	95.5	66.6	87.0
0.3	0.99	91.7	97.3	74.0	89.8
0.2	0.99	92.5	97.6	74.8	89.8
0.1	0.99	93.7	97.7	75.3	89.9

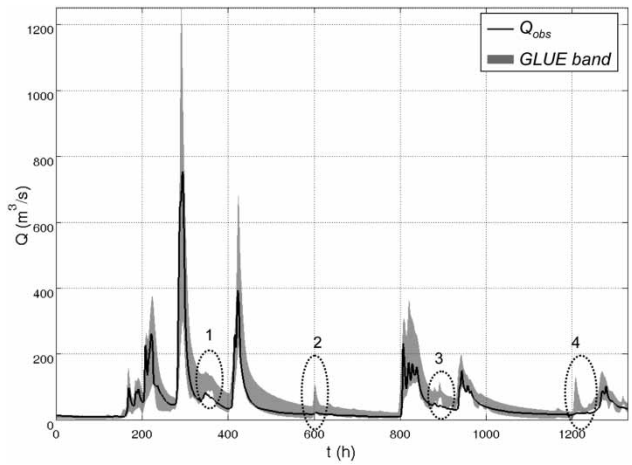

Figure 10 | Validation time window T_{val}^{59-60}. Band produced by the GLUE procedure calibrated on the time window T_{cal}^{59-60} with threshold value $NS = 0.3$ and $\alpha = 0.97$%.

are still present when the GLUE method is applied: in fact the GLUE method also gives bands that do not include the observed data for these events and thus this situation can properly be ascribed to some inconsistencies in the data or to the difficulties of the rainfall-runoff model (due to some internal limitation or inaccuracy) in effectively reproducing the discharges in these time windows.

The average bandwidths and percentages of coverage indicated in Table 6 also show that the GLUE method, with threshold value $NS = 0.3$ and $\alpha = 0.97$, and the grey method (see Table 4) give similar performances. However, with respect to the calibration time window, the GLUE method gives a greater AW (81.9 versus 68.4 m^3/s) (POC being the same), whereas, for the validation time window, the GLUE method gives a slightly higher POC (85.9 versus 81.0%) but also a greater AW (65.3 versus 51.9 m^3/s; see also Figures 8 and 10, peak at $t \cong 300$ h and peak at $t \cong 450$ h).

Finally, as regards the period 1992–1996, we must report that even modifying the threshold value and the level of uncertainty α in order to make a fair comparison with the grey approach, it was not possible to achieve a $POC = 95\%$ in the calibration phase. In fact, even by reducing the threshold value NS to 0.1 and simultaneously increasing the level of uncertainty α to 0.99, the POC index increases to nearly 91%.

CONCLUSIONS

This paper proposes a new procedure for characterising total uncertainty in a (conceptual ESMA) rainfall-runoff simulation model based on using grey numbers to parameterise the model. The model so parameterised provides not a crisp value of the simulated discharge at each time step, but rather an interval representing the total model vagueness/uncertainty. A rigorous application of the procedure for direct simulation through the rainfall-runoff model with grey parameters revealed computational limits as a result of excessive computing times. The use of an alternative expeditious procedure showed that it was possible to significantly reduce the computational times with minimal approximations in the quantification of the grey numbers representing the simulated discharges, thus making the methodology applicable also in real-life contexts and in calibration phase.

The grey parameterisation of the rainfall-runoff model and its subsequent validation on different sets of real data revealed that the procedure leads to the definition of uncertainty bands which, in the calibration phase, perfectly cover a percentage of observed values equal to the one imposed, but which in validation tend to underestimate these percentages. This tendency becomes more marked as the 'quality of the data' used declines, i.e. when the data show inconsistencies and/or are relative to phenomena which cannot be described by the rainfall-runoff model used given its structure, and is particularly accentuated when the procedure is calibrated on datasets of good quality and validated on datasets of poor quality. Finally, a comparison between the bands obtained respectively with the grey procedure and GLUE method shows a strong similarity between the two approaches, although the grey approach produces slightly narrower bands, the POC being the same.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Elena Toth for providing the rainfall, temperature and discharge data of the Sieve river basin pertaining to the period 1992–1996.

The authors are grateful to Professor Keith Beven and to the other two anonymous reviewers for their helpful and constructive comments that helped us to improve the quality and clarity of the paper.

This study was carried out within the framework of Terra & Acqua Tech Laboratory, Axis I activity 1.1 of the POR FESR 2007–2013 project funded by Emilia-Romagna Regional Council (Italy) (http://fesr.regione.emilia-romagna.it/allegati/comunicazione/la-brochure-dei-tecnopoli).

REFERENCES

Alvisi, S. & Franchini, M. 2010 Pipe roughness calibration in water distribution systems using grey numbers. *J. Hydroinform.* **12** (4), 424–445.

Alvisi, S. & Franchini, M. 2011 Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environ. Model. Softw.* **26** (4), 523–537.

Alvisi, S. & Franchini, M. 2012 Grey neural networks for river stage forecasting with uncertainty. *Phys. Chem. Earth* **42–44**, 108–118.

Alvisi, S. & Franchini, M. 2013 A grey-based method for evaluating the effects of rating curve uncertainty on frequency analysis of annual maxima. *J. Hydroinform.* **15** (1), 194–210.

Bardossy, A., Bogardi, I. & Duckstein, L. 1990 Fuzzy regression in hydrology. *Water Resour. Res.* **26** (7), 1497–1508.

Beven, K. J. 1993 Prophesy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resour.* **16**, 41–51.

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**, 279–298.

Beven, K. J., Smith, P. J. & Wood, A. 2011 On the colour and spin of epistemic error (and what we might do about it). *Hydrol. Earth Syst. Sci.* **15**, 3123–3133.

Biondi, D., Versace, P. & Sirangelo, B. 2010 Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: an application to a small catchment. *J. Hydrol.* **386**, 38–54.

Blasone, R. S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Zyvoloski, G. A. & Robinson, B. A. 2008 Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Adv. Water Res.* **31**, 630–648.

Cheng, F. J., Hui, S. H. & Chen, Y. C. 2002 Reservoir operation using grey fuzzy stochastic dynamic programming. *Hydrol. Proc.* **16**, 2395–2408.

Dawdy, D. R. & O'Donnel, T. 1965 Mathematical models of catchment behaviour. *J. Hydraul. Div. Proc. Am. Soc. Civil Eng.* **91** (HY4), 123–127.

Deng, J. L. 1982 Control problems of grey systems. *Syst. Control Lett.* **1** (5), 288–294.

Deng, J. L. 1987 *Basic Methodology of Grey System*. Publishing House of Huazhong, University of Science and Technology, Wuhan.

Dettinger, M. D. & Wilson, J. L. 1981 First order analysis of uncertainty in numerical models for groundwater flows. *Water Resour. Res.* **17** (1), 149–161.

Duan, Q., Sorooshian, S. & Gupta, V. K. 1992 Effective and efficient global optimization for conceptual rainfall runoff models. *Water Resour. Res.* **24** (7), 1163–1173.

Franchini, M. 1996 Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall runoff models. *Hydrol. Sci. J.* **41**, 21–39.

Franchini, M. & Pacciani, M. 1991 Comparative analysis of several conceptual rainfall–runoff models. *J. Hydrol.* **122**, 161–219.

Franchini, M. & Todini, E. 1988 PABL: a parabolic and backwater scheme with lateral inflow and outflow. 5th IAHR International Symposium. Stochastic Hydraulics, Birmingham, UK.

Gupta, H. V., Beven, K. J. & Wagener, T. 2005 Model calibration and uncertainty estimation. In: *Encyclopedia of Hydrological Sciences* (M. G. Anderson, ed.). John Wiley, New York, pp. 2015–2031.

Hornberger, G. M. & Spear, R. C. 1981 An approach to preliminary analysis of environmental systems. *J. Environ. Manage.* **12**, 7–18.

Jacquin, A. P. & Shamseldin, A. Y. 2007 Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modelling. *Water Resour. Res.* **43**, W04425.

Karmakar, S. 2009 *Grey Optimization For Uncertainty Modeling In Water Resources Systems: Grey Optimization Models for Environmental and Water Resources Systems Analysis*. LAP Lambert Academic Publishing, Saarbrucken, Germany.

Karmakar, S. & Mujumdar, P. P. 2006 Grey fuzzy optimization for water quality management of a river system. *Adv. Water Resour.* **29**, 1088–1105.

Karmakar, S. & Mujumdar, P. P. 2007 A two-phase grey fuzzy optimization approach for water quality management of a river system. *Adv. Water Resour.* **30**, 1218–1235.

Krzysztofowicz, R. 1999 Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* **35** (9), 2739–2750.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211**, 69–85.

Li, L., Xia, J., Xu, C. Y. & Singh, V. P. 2010 Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *J. Hydrol.* **390**, 210–221.

Liu, S. & Lin, Y. 1998 *An Introduction to Grey Systems: Foundation, Methodology and Applications*. IIGSS Academic Pub., Slippery Rock, Pennsylvania, USA.

Maskey, S., Guinot, V. & Price, R. K. 2004 Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach. *Adv. Water Resour.* **27**, 889–898.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091.

Montanari, A. 2005 Large sample behaviours of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **41**, W08406.

Montanari, A. 2011 Uncertainty of hydrological predictions. In: *Treatise on Water Science* (P. Wilderer, ed.), vol. 2. Academic Press, Oxford, pp. 459–478.

Montanari, A. & Brath, A. 2004 A stocastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **40**, W01106.

Moore, R. J. 1985 The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* **30** (2), 273–297.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models 1: a discussion of principles. *J. Hydrol.* **10** (3), 282–290.

O'Connel, P. E. 1991 A historical perspective. In: *Recent Advances in the Modeling of Hydrologic Systems* (D. S. Bowles & P. E. O'Connel, eds). Kluwer, Dordrecht, pp. 3–30.

Powell, M. J. D. 1978 A fast algorithm for nonlinearly constrained optimization calculations. In: *Numerical Analysis* (G. A. Watson, ed.). Lecture Notes in Mathematics, vol. 630. Springer Verlag, Berlin, Germany.

Powell, M. J. D. 1983 Variable metric methods for constrained optimization. In: *Mathematical Programming: The State of the Art* (A. Bechem, M. Grostschel & B. Korte, eds). Springer Verlag, Germany, pp. 288–311.

Schitlowski, K. 1985 NLQPL: A Fortran-subroutine solving constrained nonlinear programming problems. *Oper. Res.* **5**, 485–500.

Shrestha, D. L. & Solomatine, D. P. 2006 Machine learning approaches for estimation of prediction interval for the model output. *Neural Netw.* **19** (2), 225–235.

Shrestha, D. L. & Solomatine, D. P. 2008 Data-driven approaches for estimating uncertainty in rainfall runoff modelling. *J. River Basin Manage.* **6** (2), 109–122.

Singh, V. P. 1995 *Computer Models of Watershed Hydrology*. Water Resources Publications, Highlands Ranch, CO.

Solomatine, D. P. & Dulal, K. N. 2003 Model trees as an alternative to neural networks in rainfall runoff modelling. *Hydrol. Sci. J.* **48** (3), 399–411.

Stedinger, J. R., Vogel, R. M., Lee, S. U. & Batchelder, R. 2008 Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res.* **44**, W00B06.

Thiessen, A. H. 1911 Precipitation averages for large areas. *Mon. Weather Rev.* **39** (7), 1082–1084.

Todini, E. 1996 The ARNO rainfall runoff model. *J. Hydrol.* **175**, 339–382.

Todini, E. 2008 A model conditional processor to assess predictive uncertainty in flood forecasting. *Intl. J. River Basin Manage.* **6** (2), 123–137.

Toth, E. & Brath, A. 2007 Multistep ahead streamflow forecasting: role of calibration data in conceptual and neural network modelling. *Water Resour. Res.* **43**, W11405.

Tung, Y.-K. 1996 Uncertainty and reliability analysis. In: *Water Resources Handbook* (L. W. Mays, ed.). McGraw-Hill Book Company, New York, pp. 7.1–7.65.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003 A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **39** (8), 1201.

Wang, Q. & Wu, H. 1998 The concept of grey number and its property. In: *Proceedings of North American Fuzzy Information Processing Society*. NAFIPS, Pensacola Beach, FL, USA, pp. 45–49.

Wu, Q., Zhou, W., Li, S. & Wu, X. 2005 Application of grey numerical model to groundwater resource evaluation. *Environ. Geol.* **47**, 991–999.

Yang, Y., John, R. & Chiclana, F. 2004 Grey sets: a unified model for fuzzy sets and rough sets. In: *Proceedings of the 2004 UK Workshop on Computational Intelligence* (W. Withall & C. Hide, eds). Loughborough, UK, pp. 239–246.

Young, P. C. 1983 The validity and credibility of models for badly defined systems. In: *Uncertainty and Forecasting of Water Quality* (M. B. Beck & G. Van Straten, eds). Springer Verlag, New York.

Xiong, L. & O'Connor, K. M. 2008 An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modelling. *J. Hydrol.* **349**, 115–124.

Zhao, R. J., Zhuang, Y. L., Fang, L. R., Liu, X. R. & Zhang, Q. S. 1980 The Xinanjiang model. *IAHS Publ.* **129**, 351–356.

Zhang, X., Liang, F., Srinivasan, R. & Van Liew, M. 2009 Estimating uncertainty of streamflow simulation using Bayesian neural networks. *Water Resour. Res.* **45**, W02403.

First received 1 June 2011; accepted in revised form 17 April 2012. Available online 21 July 2012