Calibration of physically based models: back to basics?

Vincent Guinot and Philippe Gourbesville

ABSTRACT

The modelling of extreme hydrological events often suffers from a lack of available data. Physically based models are the best available modelling option in such situations, as they can in principle provide answers about the behaviour of ungauged catchments provided that the geometry and the forcings are known with sufficient accuracy. The need for calibration is therefore limited. In some situations, calibration (seen as adjusting the model parameters so that they fit the calculation as closely to the measurements as possible) is impossible. This paper presents such a situation. The MIKE SHE physically based hydrological model is used to model a flash flood over a medium-sized catchment of the Mediterranean Alps (2820 km^2). An examination of a number of modelling alternatives shows that the main factor of uncertainty in the model response is the model structure (what are the dominant processes). The second most important factor is the accuracy with which the catchment geometry is represented in the model. The model results exhibit very little sensitivity to the model parameters, and therefore calibration of these parameters is found to be useless.

Key words | extreme events, hydrological modelling, model calibration, ungauged catchments

Vincent Guinot (corresponding author)
Université Montpellier 2,
Maison des Sciences de l'Eau,
34095 Montpellier Cedex 5,
France
E-mail: *guinot@msem.univ-montp2.fr*

Philippe Gourbesville
University of Nice-Sophia Antipolis,
98 boulevard Edouard Herriot,
06204 Nice,
France
E-mail: *gourbesv@unice.fr*

INTRODUCTION: NEED AND ABSENCE OF NEED FOR CALIBRATION OF PHYSICALLY BASED MODELS

What is a model? Definitions and classification

There are probably as many definitions of what is a physically based model as there are physically based models. The following three definitions are quoted extensively:

> Goal-oriented, simplified mathematical description of a system [. . . that is] a time-varying entity of interest separated from its environment. (Schilling 1996)

> A simplified representation of the natural system. (Refsgaard 1996)

> A complex assembly of several, if not many, constituent hypotheses. (Beck *et al*. 1997)

A common underlying notion in these definitions is that a model is a simplified representation of reality, built to serve a number of well-defined purposes, under a certain set of assumptions as to 'how the real world behaves'. It is, in the most general terms, 'a collection of signs that serves as a sign' (Abbott 1992; see also Abbott 2002). A model may ignore many (natural or artificial) processes which are actually occurring in reality if they are not considered to be relevant or influential. For instance, most river models neglect the influence of evaporation, seepage through the river-bed or interactions with aquifers because these phenomena have different time scales and spatial scope from that of processes such as floods, that occur in most river flows.

The literature usually distinguishes between three types of models (Refsgaard 1996).

1. Empirical, or 'black box', models which aim to relate a given set of 'input' variables (such as rainfall intensity, river discharge at earlier times) to 'output variables' (such as runoff, river discharges at future times), without attempting to identify and describe the physical processes that relate these inputs and outputs. Typical examples are regression models,

artificial neural networks and chaos theory-based models. Such models have frequently been applied to catchment modelling in the past, in particular for rainfall–runoff modelling (Minns 1998). These models must be calibrated. This is done by adjusting their internal parameters in order to minimize the difference between the measured variables and the calculated outputs.

2. Conceptual, or ‘grey-box’, models describe the system under study in terms of storage entities that exchange mass via fluxes. In general, these fluxes are simple functions of the volumes stored in the various compartments of the model. Typical examples are reservoir models, such as the NAM (Nielsen & Hansen 1973).
3. Physically based models describe the system to be modelled in terms of processes that are accounted for by laws expressing general principles (e.g. continuity, momentum and/or energy conservation). The state and evolution of the system is described using state variables that are functions of both space and time. These variables bear a physical meaning and are measurable. The principles used in physically based models are assumed to be valid for a wide range of situations, including situations that have not yet been observed. Therefore, their range of validity should be expected to be wider than that of the other two types of model. A typical example of a physically based, hydrological modelling system is the SHE/MIKE SHE system (Abbott *et al*. 1986*a*, *b*) used in the present study.

Model types and data requirements

It has become customary to say that any model needs data. This is true, of course, but the three types of model described above need *different types* of data. Moreover, the nature and amount of data needed depends on what the model is to be used for.

For instance, empirical models need no geometrical description of the morphology (topography, topology and dimensions of the river network, geometry of the geological layers, etc.) of the system to be modelled, while such data is of primary importance to a physically based model. Conversely, an empirical model needs extensive time series for all inputs and outputs in order to allow for calibration of its internal relationships, while the time series is, or should be, considered as essential to validation when used by a physically based model.

These considerations clearly delineate the fields of application of the various model types. Empirical models can be applied only to events that fall within the range of events recorded in the past. In other words, they are applicable only to systems of which the physical characteristics do not change with time, i.e. for the modelling of common events. However, when applied adequately, they may give very good results (see, e.g., Minns & Hall 1996; Minns 1998). In contrast, physically based models are useful tools to assess the impact of changes in the physics of the catchment (urbanization, aquifer withdrawal, etc.), or for the assessment of extreme events. They usually give less reliable results than empirical models in the field of short-term forecasting because a short-term forecast with a physically based model would require huge amounts of data about the initial state of the model (e.g. for a rainfall–runoff model, initial water depth, soil moisture content and depth to the aquifer water table at all points of a catchment). Most of the time, however, such data are not available.

Unknown parameters and phenomena observable with difficulty

From the considerations described above, it is clear that physically based models are possibly the most suitable type of models to address the modelling of extreme events. However, the modelling attempt often suffers from several drawbacks.

- Usually, very little accurate historical data exist about extreme events because there are no measurement devices (e.g. flooding of urban areas), because any such devices have been destroyed or because measurements were impossible during the events.
- The physical characteristics of the modelled domain may be subject to large spatial variability, which

results in considerable uncertainty in the parameters of the laws that describe the flow. A typical example is that of groundwater flow and contaminant transport modelling, where the soil hydraulic parameters may change dramatically within a few metres.
- There may be a lack of data on the geometry and the physical characteristics of the system under study. For instance, it is very difficult, if not impossible, to obtain extensive data on the geometry of a multi-layered aquifer because of the cost of sampling.

Nevertheless, it is customary to build up models using an approximate characterization of the geometry, on the grounds that the missing information can be supplemented by model calibration, i.e. by adjusting the unknown or uncertain parameters of the model that have not been measured extensively.

Differences between direct parameter measurement and calibration

Calibration is very often used as a way of adjusting – if not of determining completely – the parameters of a model from measured data. The differences between the direct measurement of a parameter and calibration are described below.

The direct measurement of a parameter (e.g. soil porosity or hydraulic conductivity) is often done under ‘controlled’ conditions, i.e.:

- in a reasonably well-defined space;
- a limited number of phenomena are involved at the same time, so that the inverse problem (inferring the parameter value from the measurements) is well-posed;
- the measured parameter should be the determinant factor in the accuracy of the mathematical relationships that describe the phenomena occurring during the measurement (e.g. one would not calibrate the specific yield from a steady pumping test, because the specific yield is not a parameter of steady groundwater flow equations);
- the forcings (such as the infiltration or pumping rate) are controlled;
- these forcings correspond to the range of validity of the laws that involve the parameter to be measured.

The calibration of a model is the opposite in that very few conditions can be controlled by the observer:

- at the time, the region of space covered by the calibration is undefined (in general, it is the whole model);
- the numbers of most of the phenomena involved and their interactions are not controlled by the modeller (e.g. in the field of rainfall–runoff modelling, surface runoff, subsurface runoff and river–aquifer exchange occur at the same time, and they are coupled);
- many parameters may exert a competing influence on the phenomena, or some of them may turn out not to be influential at all;
- the forcings are not controlled, in as much as they may be subject to huge spatial and temporal variability (e.g. the spatial and temporal variability of rainfall);
- moreover, the forcings do not necessarily lie in the range of validity of the mathematical laws used by the model.

This makes calibration a very hazardous operation if the modeller has no knowledge or experience of the physics occurring in the system being modelled. Moreover, as stated by Ewen & Parkin (1996) and Parkin *et al.* (1996), calibration, as for any model operation, should be carried out bearing in mind the specific objectives of the modelling attempt. Since the phenomena involved may be different for different flow regimes, it is most likely that certain parameters should be calibrated using only some of the available records (e.g. only low flows or high flows) because they are influential only for this specific range of the records.

Purpose of the present paper

The purpose of the present paper is to substantiate the following statements.

1. In a physically based modelling attempt, the uncertainty induced by the schematization of the processes and of the geometry is much larger than that induced by the lack of knowledge of the parameters.
2. Therefore, calibration (or even worse, automatic calibration) may be useless and/or dangerous in many cases, and so is the belief that the outputs of the model should be adjusted as closely as possible to the measurements.
3. A model should be operated only by a modeller (i.e. by someone who has experience and knowledge of the physics and understands how reality is described in the model). Therefore, the belief that models can be operated by anybody is not justified.

These statements are illustrated by the case study of the flash flood of 5 November 1994 that occurred on the Var catchment (France). In the present paper, we restrict our considerations to physically based rainfall–runoff modelling for extreme events, but the arguments developed here are applicable to many other situations, such as one-or two-dimensional free-surface flow modelling. Some of these arguments are indeed not arguments but facts, e.g. to many specialists such as coastal modellers, who know that the accuracy of two- or three-dimensional coastal models is not conditioned by the model parameters, but by the accuracy of the bathymetry and boundary conditions. The modelling attempts presented here can be seen as a preliminary step to the blind validation methodology proposed by Ewen & Parkin (1996) and Parkin *et al.* (1996).

APPLICATION EXAMPLE: THE VAR CATCHMENT (FRANCE)

Description of the watershed

The Var catchment is located in the southern part of France, near the city of Nice. It discharges directly to the Mediterranean Sea (Figure 1). The Var is the largest catchment in the French Mediterranean Alps, with a surface of 2,822 km. The spring which is the source of the river is located at Esing, at an elevation of 1,800 m, in a geological area characterized mainly by metamorphic formations. The ground levels range from 0 m (sea level) to above 3,000 m. The shape of the catchment is roughly rectangular, and it is tilted to the north–north–east. Its dimensions are 70 km and 75 km in the east–west and north–south directions, respectively. The ground slopes may reach large values (more than 20%) in some parts of the catchment. The major tributaries of the Var are la Tinée, la Vésubie and l'Estéron (see Figure 1), with a total length of about 125 km.

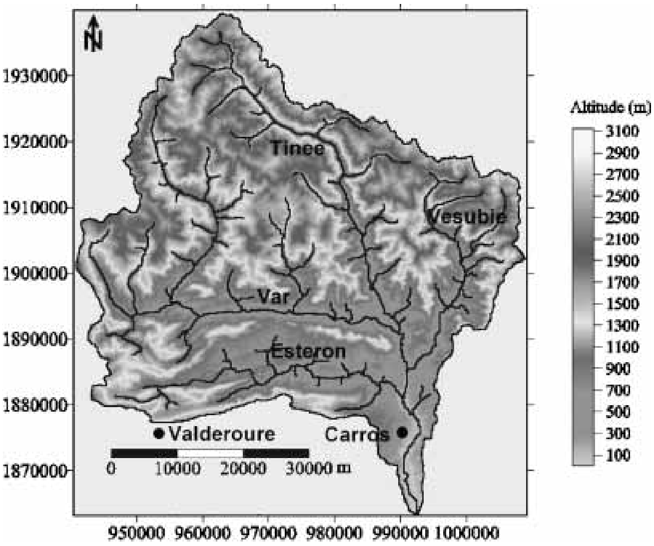

Figure 1 | Topography of the Var watershed (from Geoman 1997).

Geological background

The Var basin is characterized by a strongly heterogeneous geology. Four main geological families can be identified.

1. A large crystalline and metamorphic zone around the Argentera–Mercantour mountains. Outcrops from this formation can be seen over the eastern part of the basin.
2. The Mesozoic and Cenozoic successions of the Castellane arch, that are represented by series of calcareous or marl–calcareous formations, and by the Cenozoic deposits characteristics of the Annot

sandstone. These formations cover the south-eastern part of the basin.

3. The Permian successions, mainly pelitic and sometimes plasteric, in the middle part of the catchment.
4. The Pliocenic Quaternary successions of the lower valley of the Var, formed by puddings of pebble-stones and by blue marls.

All these geological formations are associated with thin soils that are very quickly saturated during rainfall events, leading to large runoff production.

Rainfalls and discharges

Rainfalls in the Var catchment are typical of the Mediterranean Alps. The yearly total precipitation is about 815 mm, concentrated mainly in 65–80 days per year. The surface hydrology of the catchment is governed by snow-melt and intense rainfall events. Floods occur mainly in spring owing to the combination of snow-melt and rainfall, and in autumn owing to intense precipitation. In the summer, the discharges in the downstream area are limited to a few cubic metres per second ($10\ \mathrm{m^3\ s^{-1}}$ is a common value). The discharges of the Var and its tributaries are not well known, since measurements are available only for the periods 1974–1976 and 1985–1994. Consequently, there is a lack of knowledge about the hydrological aspects of the catchment, and more particularly about the dynamics of floods.

The lower reaches of the Var have been subjected to many man-induced transformations over the past three centuries. Eleven weir-type artificial controls were built between 1701 and 1986 in order to maintain the exchanges between the river and the aquifer in the lower valley, as well as to prevent the bed from being eroded. A number of small hydroelectric power plants were built on the weirs between 1985 and 1989. These structures limit the conveyance of the river and induce particular hydraulic behaviour.

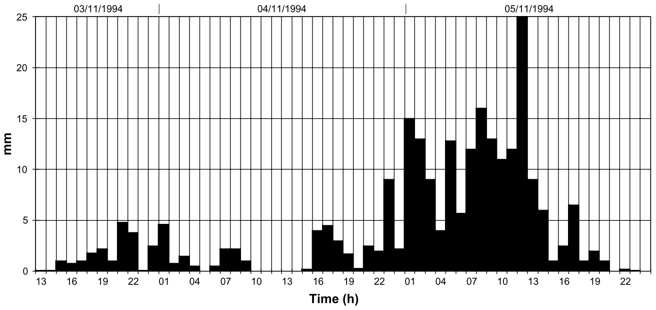

Figure 2 | Rainfalls recorded in the Valderoure–Estéron basin (from Météo France and CEMAGREF 1996).

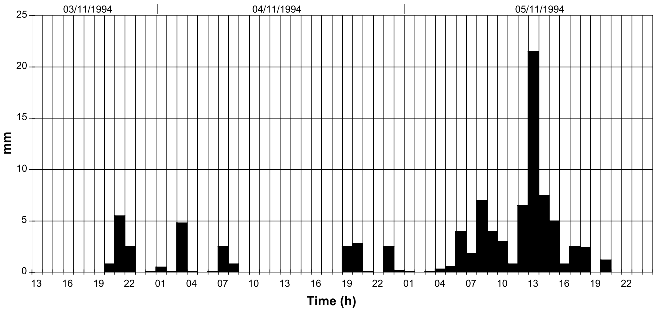

Figure 3 | Rainfalls recorded in the Carros–Var basin (from Météo France and CEMAGREF 1996).

The flash flood of November 1994

The flood of 5 November 1994 was one of the most spectacular hydrological events recorded in the Var valley. It occurred after a rainfall event that was triggered by a large disturbance moving over the whole south-east part of France. The average accumulated rainfall over the catchment between 2 and 5 November was about 200 mm, with values up to 283 mm at some points. The rainfall was moderate but continuous from 2 to 4 November, which led to a progressive saturation of the soils all over the catchment. The whole of 5 November 1994 was characterized by large rainfall intensities, and half of the accumulated rainfall over the whole period fell on this day. Conversely, the rainfall was limited to less than 100 mm along the coastline (Nice and Carros). Figure 2 shows the recorded hourly rainfalls at Valderoure, which is located in the Estéron sub-basin. Figure 3 shows the hourly rainfalls at Carros, along the coastline.

High precipitation intensities over almost fully saturated soils triggered intense runoff. Owing to the exceptional intensity of the flood, only estimates of the discharges are available at a limited number of stations.

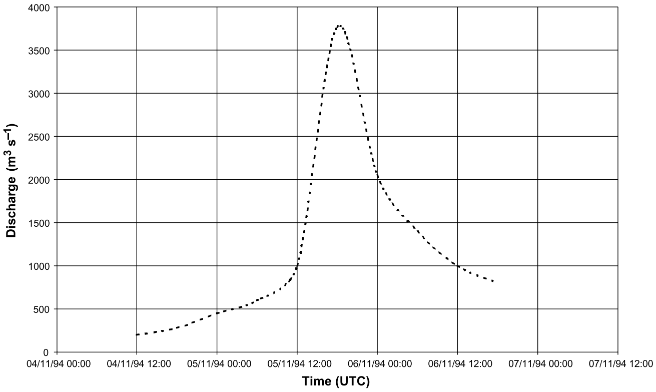

Figure 4 | Hydrograph recorded at the Napoléon III Bridge, Nice (from DIREN PACA & CEMAGREF 1996).

Figure 5 | Napoléon III Bridge, Nice, 5 November 1994, at 1900 hours (picture: Nice Matin).

Figure 6 | Highway A8, Nice, 5 November 1994 (picture: Nice Matin).

Figure 7 | Highway A8, Nice, 5 November 1994 (picture: Nice Matin).

The peak value of the discharges in the downstream area (Napoléon III Bridge) was estimated to have been 3,770 $m^3 s^{-1}$ at 1900 hours (Figures 4 and 5). This value must be examined critically.

- The last point used to build the rating curve corresponds to a discharge of 600 $m^3 s^{-1}$.
- Two major weirs, built in the river bed and dedicated to bed stabilization, were destroyed by the event. Clearly, the breakdown of these structures induced an artificial increase in the discharge. The artificial increase has been assessed to be about 500 $m^3 s^{-1}$.

The dikes were overtopped at several points in the lower valley, most of which was inundated (Figures 6 and 7). The phenomenon was completely new and had never before been seen by the inhabitants. A number of national and departmental roads were destroyed (Figures 8 and 9), some bridges were completely washed away and numerous buildings were severely damaged. Luckily no lives were lost, but the damage was estimated at around 23×10^6 Euros. The international airport at Nice was closed for 1 week. Of particular interest is the fact that the locally operated centre for flood crisis management was located in the basement of an administrative building

Figure 8 | National road 202, Var valley, 6 November 1994 (picture: Nice Matin).

Figure 9 | Downstream area of the Var valley, 6 November 1994 (picture: Nice Matin).

next to the river. Consequently, it was flooded almost immediately.

MODELLING ATTEMPTS AND RESULTS

Seven models were built and run to simulate the flood event of 5 November 1994. These models were based on different structures, geometries and parameter values. The main objective of the simulations, which are described below, was to assess the relative importance of model structure, geometry and parameter value with respect to the simulation results. Since one of the goals was to assess the influence of process schematization on the simulation results, the modular MIKE SHE modelling system (Abbott *et al*. 1986*a*, *b*; DHI 2000) was used.

The purpose of the present study was to assess the predictive capability of a physically based rainfall–runoff model and the respective importance of various other factors (i.e. schematization, geometry, parameter values). Consequently, the catchment was considered to be ungauged and no calibration was attempted. The few available measurements were used only for an evaluation of the predictive capability of the model.

Assumptions and data

Three assumptions were made.

1. Owing to the continuous rainfall of 3 and 4 November 1994, the soils were assumed to be fully saturated over the catchment. Consequently, the infiltration rate was assumed to represent only 10% of the rainfall intensity. The infiltration coefficient was assumed to be uniform over the whole catchment.
2. Considering the time scale of the simulation, evapotranspiration was neglected.
3. Because the base flow was negligible compared with the magnitude of the flood, the interactions between the river and the aquifer system were neglected.

Therefore, the only processes which were taken into account in the model were infiltration (which was proportional to the rainfall intensity), the two-dimensional runoff on the ground surface and the one-dimensional flow in the river system. The destruction of the weirs on 5 November could not be represented in the model.

The data available for the study are listed below.

- Two digital elevation models (DEMs) (Geoman 1997) were available at two different resolutions (75 m and 300 m).
- The topology of the river network was obtained from topographical maps with a scale of 1/50,000.
- The geometry of the river network (longitudinal profile and cross sections, including flood plain and

valley geometry) was deduced from the DEMs. In addition to this, two cross sections, including the outlet of the catchment, were available from previous topographical studies.

- The rainfall intensity was available at six rainfall stations for the period from 1300 hours on 3 November to 2400 hours on 5 November, with a recording time step of 1 h.
- The hydrograph at the outlet of the catchment (Napoleon Bridge), available at successive time steps of 1 h, was used only for model evaluation. This hydrograph was obtained by converting the measured water levels to discharges using a rating curve. It should be noted that, owing to the exceptional character of the event, the upper part of the hydrograph is outside the range of the (measured) data used for the construction of the rating curve. Therefore, the value of the peak discharge that results from an extrapolation of the rating curve cannot be considered to be reliable. However, the peak time can be considered to be reasonably accurate.

Model and simulation characteristics

As mentioned above, several models were built, based on different schematizations of the physical processes, as well as different geometries and parameter values. The list below details the characteristics of these models.

- Model 300 used a DEM with a resolution of 300 metres for the topography. The value of the overland Strickler coefficient was assumed to be 20. The river network was not represented in this model, the objective being to check whether a topographical resolution of 300 m is sufficient to represent the dynamics of the river network for a large catchment.
- Model 300a was based on the same DEM and overland Strickler coefficient as Model 300, but incorporated the river network, the elevations and geometry of which were based on the DEM with a resolution of 300 metres. The Strickler coefficient in the river was taken to be equal to 20 from previous modelling studies (CEMAGREF 1996).
- Model 300c was the same as Model 300a except that the overland Strickler coefficient was assumed to be 10 instead of 20.
- Model 300d was the same as Model 300a but with the river geometry (elevations and cross sections) corrected on the basis of a more accurate DEM with a resolution of 75 m.
- Model 75 used a DEM with a resolution of 75 m. As in Model 300, the overland Strickler coefficient was 20 and the river system was not incorporated.
- Model 75a was the same as Model 75, but the one-dimensional river model was incorporated. The topography was based on the DEM with a resolution of 75 m. A Strickler coefficient of 20 was assumed in the river.
- Model 75c was the same as Model 75a except that the overland Strickler coefficient was set equal to 10.

Table 1 | Characteristics of the rainfall–runoff models.

Model	Resolution (m)	*K* overland	Resolution of the river network model (m)
300	300	20	No river
300a	300	20	300
300c	300	10	300
300d	300	10	75
75	75	20	No river
75a	75	20	75
75c	75	10	75

Table 1 summarizes the characteristics of the seven models.

The models were run over the period 3–6 November 1994. Although the main goal was to reproduce the flash flood of 4 November, preliminary runs had shown that

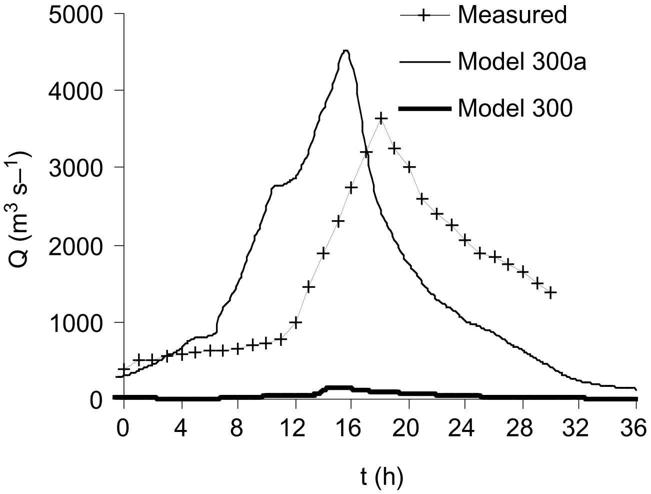

Figure 10 | Event of 5 November 1994 modelled using a DEM resolution of 300 m, with (Model 300a) and without the river network (Model 300). The time is counted from 0000 hours on 5 November 1994.

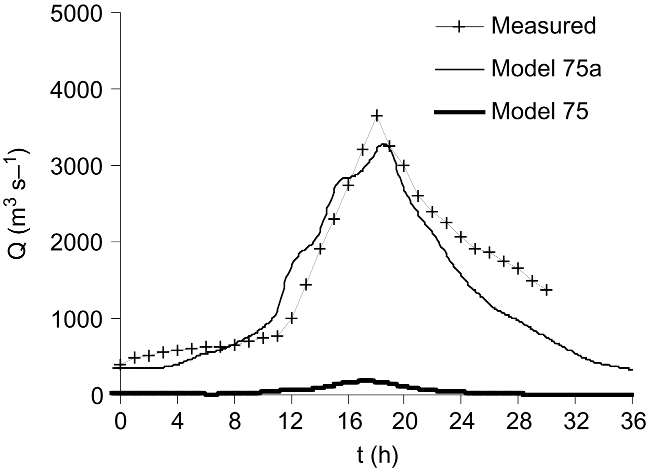

Figure 11 | Event of 5 November 1994 modelled using a DEM resolution of 75 m, with (Model 75a) and without the river network (Model 75). The time is counted from 0000 hours on 5 November 1994.

since the concentration time of the catchment was about 8 h, it was necessary to start the simulation at least 12 h before the period of interest. This precaution eliminates possible biases introduced by an incorrect assumption of the initial state.

A point of particular importance is that DEMs should never be used directly in such models without prior verification, particularly when their resolution is too coarse to represent the topographical features accurately. In such a case, non-existing depressions and artificial pits may be present in the model that do not correspond to any physical reality, resulting in an artificial increase in surface storage and runoff deficit. A fundamental precaution, prior to any simulation, is to check the DEM topography in order to make sure that the pits mentioned above (if any) correspond to actual features of the catchment. In the present case, it was necessary to correct the DEMs at a number of points before starting the simulations.

RESULTS

Figure 10 shows the simulated hydrographs at the outlet of the catchment for models 300 and 300a, compared with the measured one (bearing in mind the uncertainty associated with the measured peak value). It can be seen that the absence of the river network leads to a dramatic underestimation of the hydrograph. The explanation for this is straightforward: in reality, the channel network that concentrates runoff water is responsible for the fast response of the catchment. If the river network is not represented in the model, a cell width of 300 m is insufficient to represent the geometry in the river accurately, especially in the upstream part of the catchment. Consequently, the overland runoff process remains very slow, and most of the surface water is still stored in the catchment at the end of the simulation because it did not have time to reach the outlet. Although it improves the model results, the introduction of the river network does not yield very satisfactory results in that the peak discharge of model 300a is too large and too early compared with the measured one.

Figure 11 shows the simulation results for models 75 and 75a. The very small accumulated runoff volume in model 75 can again be accounted for by the absence of the river network. Because the two-dimensional runoff is very slow, most of the water is still stored on the ground surface at the end of the simulation. A DEM resolution of 75 m is still insufficient to represent the geometry of the river network. The one-dimensional river model is needed in

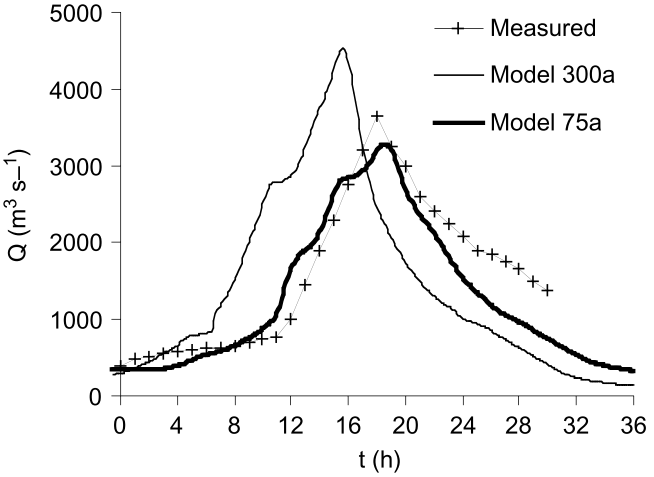

Figure 12 | Event of 5 November 1994 modelled using a DEM resolution of 300 m (Model 300a) and 75 m (Model 75a). The time is counted from 0000 hours on 5 November 1994.

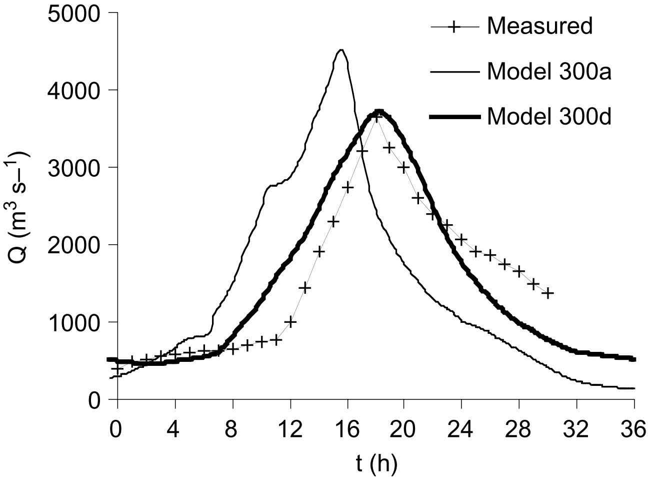

Figure 13 | Event of 5 November 1994 modelled using a DEM with a resolution of 300 m for a river geometry based on 300 m (Model 300a) and 75 m (Model 300d) DEM resolutions. The time is counted from 0000 hours on 5 November 1994.

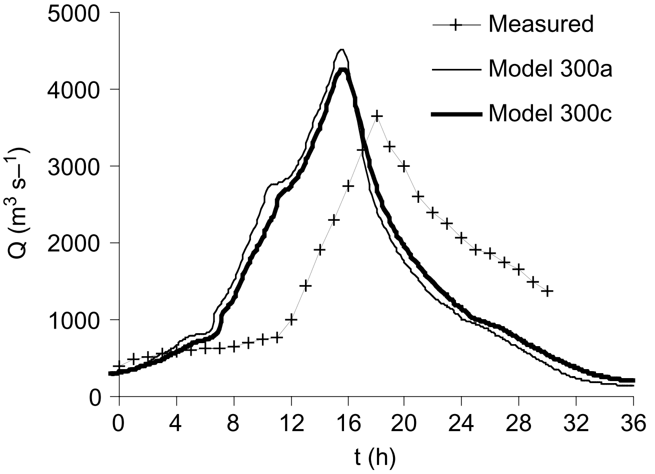

Figure 14 | Event of 5 November 1994 modelled using a DEM resolution of 300 m, with an overland Strickler coefficient of 20 (Model 300a) and 10 (Model 300c). The time is counted from 0000 hours on 5 November 1994.

order to achieve satisfactory results. Its introduction in the model (model 75a) is seen to lead to a dramatic improvement in the accuracy of the simulation results. On the basis of the reservations expressed in the previous subsections about the value of the peak discharge, model 75a is probably the most accurate of all the variants presented here in terms of peak value.

The influence of the DEM resolution (in the presence of the one-dimensional river network) is clearly illustrated by Figure 12, which shows the results of models 300a and 75a. The peak discharge is reached at 1530 hours in Model 300a, whereas it was recorded at around 1800 hours in reality.

The essential influence of the geometry is further illustrated by Figure 13, which compares the results given by models 300a and 300d. Although the rising limb is not as good in Model 300d as in Model 75a, the peak time is much more accurate than that given by Model 300a.

The comparisons between Models 300a and 300c (for a DEM resolution of 300 m) and between Models 75a and 75c (for a DEM resolution of 75 m) allows the influence of the overland Strickler coefficient to be assessed. From Figures 14 and 15, it appears that the value of the overland Strickler coefficient has a much smaller influence on the simulation results than the schematization of the model (i.e. the nature of the processes represented in the model, namely with or without one-dimensional river flow) and the resolution of the DEM. This influence is larger when the resolution of the DEM is refined, because the proportion of areas with mild slopes (mainly in the valleys) increases when the resolution is refined. Still, the influence of the overland Strickler coefficient is not as strong as that of the model geometry or of model schematization.

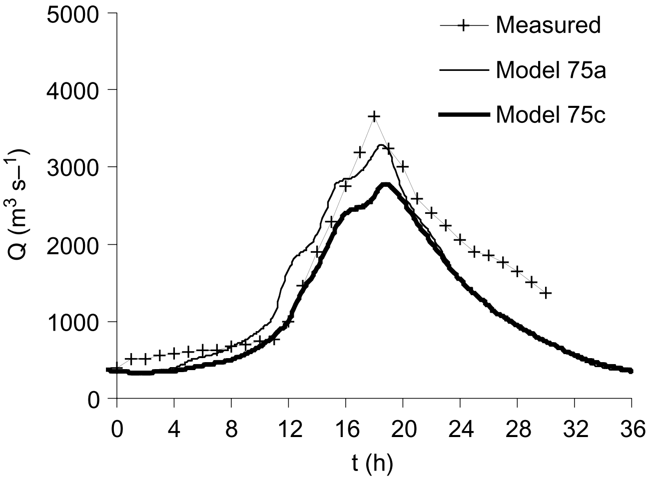

Figure 15 | Event of 5 November 1994 modelled using a DEM with a resolution of 75 m for both the DEM and the river. Results obtained with an overland Strickler of 20 (Model 75a) and 10 (Model 75c). The time is counted from 0000 hours on 5 November 1994.

DISCUSSION AND CONCLUSIONS

Several conclusions can be drawn from the experiments described above.

1. The predominant factor for the accuracy of the present model is seen to be the schematization of the hydrological processes. Indeed, neglecting the process of fast, one-dimensional river routing induces an overestimation of the catchment response time, with a subsequent overestimation of stored volumes on the ground surface and a tremendous underestimation of the peak discharge.
2. The second most important factor is the geometry. This is illustrated by the comparison between Models 300a and 75a. It should be stressed that, as shown by both Figures 12 and 13, the geometry has to be accurate for both the DEM and the river network in order to yield satisfactory results. A good description of the river alone is not sufficient; neither is a good DEM sufficient if the geometry of the river is not represented with great accuracy.
3. The values of the parameters classically subject to calibration in most models exert very little influence on the quality of the present simulation results.

As mentioned earlier in this paper, points 1–3 summarize a well-known fact in many modelling disciplines, such as coastal modelling.

1. Consequently, in this case, calibration is not only useless, but is also potentially harmful to the predictive power of the model. Only specialists in the physics involved can tell whether a given parameter can and should be calibrated.
2. Moreover, it does not make sense to try and fit models up to the last decimal place if the geometrical features of the system under study are not represented with sufficient accuracy.
3. The results given above militate in favour of the use of physically based models for the assessment of extreme events in ungauged catchments. As outlined by a number of studies on infiltration mechanisms and runoff production (see, e.g., Julien & Molgen 1990; Ogden & Julien 1993; Merz & Plate 1997), the uncertainty and spatial variability in the parameters does not have a strong influence on the simulation results for extreme events, the duration of which exceeds the characteristic response time of the system under study.

Of course, one should not infer universal rules of good modelling practice from only one modelling experiment. However, the authors believe that the conclusions of the present study can be generalized to situations that have similarities with the present one, i.e. situations where the predominant phenomena are well identified (in the present case, overland flow is the predominant mechanism), and if these phenomena are represented accurately by the sets of equations solved by the model (i.e. no phenomenon is 'emulated' or 'simulated' or 'replaced' by another one).

Since people learn best from their mistakes, it is believed that the present study could serve as an illustrative example in a possible (yet to come!) code of 'good model calibration practices'. We are convinced that the following four guidelines – which are also the last conclusions – should be incorporated to such a hypothetical code.

1. Always check that the model is not sensitive to data and grid resolution before attempting any calibration.
2. Never calibrate a parameter before you have eliminated all other sources of error and inaccuracy.
3. Never try to calibrate a parameter to which the model results are not sensitive.
4. Avoid calibration. Carry out an uncertainty analysis instead and simply validate your model by checking it against the measurements. Validating means analysing and explaining the differences between the model results and the measurements. A modeller should be knowledgeable enough to admit, and diplomatic enough to explain to their client(s), that all simulations are approximate and that *all* model results should be provided with an uncertainty interval.

ACKNOWLEDGEMENTS

The idea behind the present study stemmed from workshops on physically based hydrological modelling given at the university of Nice–Sophia Antipolis (France) in the framework of an ERASMUS project, and at IHE Delft (The Netherlands).

This work could not have been carried out without the support of the Danish Hydraulic Institute (DHI) Water and Environment that develops and markets the MIKE SHE software used in the present study. The models based on a resolution of 75 m contain more than 500,000 cells and 7,000 river links, and a special version of MIKE SHE with extended memory was compiled specifically for the present study. The authors would like to express their gratitude to Thomas Clausen, from DHI, for his patience and dedication in taking care of all the informatics aspects.

The authors would also like to express their gratitude to Professors M.J. Hall and J.A. Cunge from IHE for their comments and suggestions for improvements to the paper.

REFERENCES

Abbott, M. B. 1992 The theory of the hydrologic model, or: the struggle for the soul of hydrology. In *Advances in Theoretical Hydrology. A Tribute to James Dooge*. Elsevier, Amsterdam, pp. 237–254.

Abbott, M. B. 2002 On definitions. *J. Hydroinformatics*, (**4**)2, electronic version only.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. 1986a An introduction to the European Hydrological System – Système Hydrologique Européen, ''SHE''. 1. History and philosophy of a physically based, distributed modelling system. *J. Hydrol.* **87**, 45–59.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. 1986b An introduction to the European Hydrological System – Système Hydrologique Européen, ''SHE''. 2. Structure of a physically based, distributed modelling system. *J. Hydrol.* **87**, 61–77.

Beck, M. B., Ravetz, J. R., Mulkey, L. A. & Barbnwell, T. O. 1997 On the problem of model validation for predictive exposure assessment. *Stoch Hydrol. Hydraul* **B11,** 229–254.

CEMAGREF 1996 Etude de la crue du Var du 5 novembre 1994, Aix-en-Provence. 67 pp.

DHI 2000 *MIKE SHE Water Movement: User Manual*.

Ewen, J. & Parkin, G. 1996 Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *J. Hydrol.* **175**, 583–594.

Geoman (ed.) 1997 Visual DEM France. CD-ROM DEM. Editions Geoman, Quimper, France.

Julien, P. Y. & Molgen, G. E. 1990 Similarity and length scale for spatially varied overland flow. *Wat. Resources Res.* **26**, 1819–1832.

Merz, B. & Plate, E. 1997 An analysis of the effects of spatial variability of soil and soil moisture on runoff. *Wat. Resources Res.* **33**, 2909–2922.

Minns, A. W. 1998 Artificial neural networks as subsymbolic process descriptors. *PhD thesis*, Delft University of Technology and IHE Delft, 124 pp.

Minns, A. W. & Hall, M. J. 1996 Artificial neural networks as rainfall–runoff models. *Hydrol. Sci. J.* **41**, 399–417.

Nielsen, S. A & Hansen, E. 1973 Numerical simulation of the rainfall–runoff process on a daily basis. *Nordic Hydrol.* **4**, 171–190.

Ogden, F. L. & Julien, P. Y. 1993 Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales. *Wat. Resources Res.* **29**, 2589–2597.

Parkin, G., O'Donnell, G., Ewen, J., Bathurst, J. C., O'Connell, P. E. & Lavabre, J. 1996 Validation of catchment models for predicting land-use and climate change impacts. 2. Method. *J. Hydrol.* **175**, 595–613.

Refsgaard, J. C. 1996 Terminology, modelling protocol and classification of hydrological model codes. In *Distributed Hydrological Modelling* (ed. Abbott, M. B. & Refsgaard, J. C.). Kluwer, Dordrecht.

Schilling, W. 1996 Concepts of model building. NATO advanced study institute: hydroinformatics tools for planning. *Design, Operation and Rehabilitation of Sewer Systems.* Harrachow, Czech Republic.