

## Towards automatic calibration of 2-D flood propagation models

P. Fabio<sup>1,2</sup>, G. T. Aronica<sup>2</sup>, and H. Apel<sup>3</sup>

<sup>1</sup>Dipartimento di Ingegneria Idraulica e Applicazioni Ambientali, Università di Palermo, Palermo, Italy

<sup>2</sup>Dipartimento di Ingegneria Civile, Università di Messina, Messina, Italy

<sup>3</sup>GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

Received: 21 October 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 6 November 2009

Revised: 10 March 2010 – Accepted: 20 May 2010 – Published: 7 June 2010

**Abstract.** Hydraulic models for flood propagation description are an essential tool in many fields and are used, for example, for flood hazard and risk assessments, evaluation of flood control measures, etc. Nowadays there are many models of different complexity regarding the mathematical foundation and spatial dimensions available, and most of them are comparatively easy to operate due to sophisticated tools for model setup and control. However, the calibration of these models is still underdeveloped in contrast to other models like e.g. hydrological models or models used in ecosystem analysis. This has two primary reasons: first, lack of relevant data against which the models can be calibrated, because flood events are very rarely monitored due to the disturbances inflicted by them and the lack of appropriate measuring equipment in place. Second, 2-D models are computationally very demanding and therefore the use of available sophisticated automatic calibration procedures is restricted in many cases. This study takes a well documented flood event in August 2002 at the Mulde River in Germany as an example and investigates the most appropriate calibration strategy for a simplified 2-D hyperbolic finite element model. The model independent optimiser PEST, that enables automatic calibrations without changing model code, is used and the model is calibrated against over 380 surveyed maximum water levels. The application of the parallel version of the optimiser showed that (a) it is possible to use automatic calibration in combination of 2-D hydraulic model, and (b) equifinality of

model parameterisation can also be caused by a too large number of degrees of freedom in the calibration data in contrast to a too simple model setup. In order to improve model calibration and reduce equifinality, a method was developed to identify calibration data, resp. model setup with likely errors that obstruct model calibration.

### 1 Introduction

Floods are serious events and may have severe socio-economic impacts on vulnerable areas. Thus a reliable evaluation of the inundation extent and depths of a given flood scenario is a very important support for strategic food risk management. Different models which simulate the hydraulic behaviour of a river system are available and they should be calibrated and tested with care before exploitation.

Model calibration is the process whereby model parameters are adjusted until a satisfactory match between model response and historical data is achieved. The geometry and the roughness parameters are considered to be the most important elements affecting predicted inundation extent and flow characteristics, as elaborated with wide bibliography by Pappenberger et al. (2005). The roughness parameter will, in part, compensate the sources of errors related to these elements (Romanowicz and Beven, 2003; Marks and Bates, 2000), thus the calibration becomes a crucial issue.

One should be cautious in the selection of roughness coefficients based on the nature of the channel and floodplain surface only even if literature offers many sources of guidance. In fact, roughness coefficients in models do not represent



Correspondence to: P. Fabio  
(fabio@idra.unipa.it)

surface roughness only, but also turbulent momentum losses not explicitly modelled (Werner et al., 2005a). In addition, roughness coefficients often have to compensate for insufficient model assumption, e.g. depth averaged flow, and model setup, thus becoming what is known as effective roughness parameters. In general practice, calibration and estimation are performed manually, mostly in a “trial-and-error” fashion. This is difficult, complex, subjective, time-consuming and depends much on the expertise of the modellers. However, parameter estimation algorithms can significantly improve and facilitate this task, as shown in many other areas of environmental modelling. Here, an objective function that measures the discrepancy between observations and model outputs is defined, and the algorithm adjusts the parameter values to minimise the objective function until a convergence criterion is reached.

In general, automatic calibration procedures can be divided in two main families: Global optimisers like the popular population-evolution-based algorithms, such as the Shuffled Complex Evolution model developed by the University of Arizona (SCEUA) (Duan et al., 1992, 1993; Sorooshian et al., 1993), and the classical gradient-based approaches, like the Gauss-Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). These optimisation families have previously been mostly applied to hydrological models, where parameters can be less well defined (i.e. less physically based). Global methods are more robust in finding the global minimum in the parameter space of the objective function but they are computationally very demanding. In fact, in order to explore the whole parameter space they require a much higher number of model runs to converge compared to gradient-based methods. Gradient-based methods on the other hand are computationally very efficient but the solution can be dependent on the initial parameter values and they might get trapped in a local minimum, if the response space of the objective function is highly complex. However, these methods may be the only possibility to automatically calibrate CPU time demanding models, like the one presented here. In this case the selection of the initial parameter values has to be taken with care and should be checked by multiple optimisation runs with different starting points, if model run time allows.

In the present study the Gauss-Levenberg-Marquardt method implemented in the parameter estimation tool PEST (acronym for Parameter ESTimation) is used (Doherty, 2004, 2008). PEST is considered the most efficient method compared to other gradient based methods (Doherty, 2004, 2008) and it is successfully applied in many scientific fields such as groundwater, hydrological and water quality modelling. While PEST also offers global optimisation routines, this study explores the gradient-based method in a first step for the reasons given above.

Using the gradient-based approach this study focuses on different calibration strategies for a 2-D hydraulic model. It is calibrated against a serious flood event that occurred in

August 2002 on the river Mulde in the city of Eilenburg in Saxony, Germany. In the different calibration strategies four aggregation levels of the spatially distributed surface roughness were considered: (a) a single roughness value for the channel and the whole floodplain; (b) two roughness values attributed to the channel and the whole floodplain; (c) four roughness values attributed to the channel and three land use classes in the floodplain; (d) five roughness values related to the channel and four land use classes in the floodplain.

Being certain that a computer-based model is an imperfect representation of a physical system, a perfect match is not expected from a calibration to the available field measurements. This inability may be due to the presence of errors in both data and in the model (Gupta et al., 1998). We assume that the mathematical structure of the model is predetermined and fixed and that the upper and lower bounds of parameter ranges can be specified a priori.

In most cases the calibration of flood inundation models is constrained by insufficient data against which the model can be calibrated. This is especially the case for distributed floodplain data, causing the often observed dilemma that a spatially distributed model is calibrated against bulk flow data in the channel allowing for a high degree of equifinality in model realisations (e.g. Aronica et al., 2002; Bates, 2004; Pappenberger et al., 2005). However, in this study area the flood event occurred in August 2002 was well documented: flood depths were recorded from a large number of water marks, the maximum inundation extent was surveyed from satellite imaging and water marks, and the flood hydrograph was recorded at the upstream flowgauge. This data set enabled not only the automatic calibration, it will also show that a large amount of data or information do not assure an improvement in the identification of the parameters. It is not only the number, but also the quality of the information contained in the data that is important. Increasing the amount of data does not necessarily improve the parameter estimation (Sorooshian et al., 1993). A wide literature analyses simulation errors of inundation model discussing of the possible sources in order to obtain acceptable models even at local scale (Bates and De Roo, 2000; Horritt, 2006; Mignot et al., 2006; Schumann et al., 2007; Neal et al., 2009 are just a few examples). In this paper a procedure to remove potentially erroneous data is presented.

## 2 Methodology

### 2.1 2-D hydrodynamic model

In order to model the flow regime in an urban area, a simplified 2-D model, which is able to consider the hydraulically important features like streets, buildings, channels, etc., is the favoured option. In this study, the model of Aronica et al. (1998a) was applied. It is a hyperbolic model based on de Saint-Venant equations for 2-D shallow-water flow

(DSV), where convective inertial terms are neglected in order to eliminate the related numerical instabilities. The conservative mass and momentum equations for 2-D shallow-water flow can be written as

$$\frac{\partial H}{\partial t} + \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = 0 \quad (1)$$

$$\frac{\partial p}{\partial t} + gh \frac{\partial H}{\partial x} + gh J_x = 0; \quad \frac{\partial q}{\partial t} + gh \frac{\partial H}{\partial y} + gh J_y = 0 \quad (2)$$

where  $H(t, x, y)$  = free surface elevation;  $p(t, x, y)$  and  $q(t, x, y) = x$ - and  $y$ -components of the unit discharge (per unit width);  $h$  = water depth;  $g$  = gravitational acceleration; and  $J_x$  and  $J_y$  = hydraulic resistances in the  $x$ - and  $y$ -directions. The hydraulic resistance is parameterised by the Manning-Strickler formulation, the Strickler roughness coefficient  $k$  ( $\text{m}^{1/3} \cdot \text{s}^{-1}$ ) is related to Manning coefficient  $n$  ( $\text{m}^{1/3} \cdot \text{s}^{-1}$ ) through

$$k = \frac{1}{n}, \quad (3)$$

Equations (1) and (2) were solved by using a finite element technique with triangular elements. The free surface elevation is assumed to be continuous and piece-wise linear inside each element, where the unit discharges in the  $x$  and  $y$  directions are assumed to be piece-wise constant.

The finite element approach allows a more detailed description of hydraulic behaviour of flow in the flooded areas, in fact unstructured meshes are able to reproduce the complex topography of built-up and urban areas. High hill-top, building blocks, houses and other obstacles are treated as internal islands within the triangular mesh covering the entire flow domain. Being explicitly 2-D the model also allows for spatially varying roughness coefficients in the whole model domain. For proper model setup detailed topographic information are required: topographical map preferably with a scale of 1:10 000 and lower, a high spatial resolution DEM and data set about the river topography (a number of cross sections with bed elevations, channel widths and roughness coefficients are useful to improve the mesh descriptive capability in those parts of floodplains; Horritt and Bates, 2001).

## 2.2 Model calibration

The 2-D model was calibrated using the model independent optimiser PEST (Doherty, 2004, 2008). It gives the possibility of an automatic calibration without the necessity to change the model at all, but only the values of the parameters in one or more model input files. Applying this concept to hydraulic modelling is a step towards a more objective and encompassing model calibration compared to the general manual practice.

PEST adjusts model parameters to obtain the best match between model generated values and the correspondent measurements in a weighted least squared sense. Given the number of degrees of freedom present in the calibration procedure of 2-D models, where individual friction parameters

can theoretically be assigned at each computational node and at each time-step (Marks and Bates, 2000) some difficulties may arise because of the equifinality problem encountered during the calibration procedure (cf. e.g. Beven, 1993, 1996, 2006; Beven and Binley 1992; see also bibliography in Beven and Freer, 2001).

The parameter estimation software PEST implements the Gauss-Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963) for parameter estimation and uncertainty analysis. The method is a combination of gradient descent and Newton's method. Parameter estimation is an iterative process linearizing the relationship between model parameters and model outputs. The linearization is conducted by formulating a Taylor expansion of the actual parameter set. At every iteration the partial derivatives of each model output with respect to every parameter are calculated using finite differences. The technique follows the steepest gradient of the objective function until the gradient becomes small with respect to a certain tolerance limit. In the steepest region of the objective function the search for the minimum is performed slowly (with small step size), in shallow regions the movement is quicker (with large steps). Moreover, Marquardt improved the method considering each component of the gradient according to the curvature, i.e. the search moves further in the directions in which the gradient is smaller in order to speed up the convergence (this is very important for example when the solution space of the objective function presents a long and narrow valley). The parameter upgrade vector given by the Levenberg-Marquardt method is written as

$$\mathbf{p} - \mathbf{p}_0 = (\mathbf{J}^t \cdot \mathbf{Q} \cdot \mathbf{J} + \lambda \text{diag}[\mathbf{J}^t \cdot \mathbf{Q} \cdot \mathbf{J}])^{-1} \cdot \mathbf{J}^t \cdot \mathbf{Q} \cdot \boldsymbol{\varepsilon} \quad (4)$$

where  $\mathbf{p}$  = subsequent parameter vector,  $\mathbf{p}_0$  = current parameter values,  $\mathbf{J}$  = the Jacobian matrix,  $\mathbf{Q}$  = a diagonal matrix such that the inverse is proportional to the covariance matrix of the observations,  $\lambda$  = the Marquardt lambda.

The data used in the calibration of flood inundation models are chiefly observed flood extends and depths. The inundation extends are derived during the flood from geo-referenced aerial photographs of the flood event or from remote sensing in combination with a detailed DEM, or derived from ground surveys of inundation marks in the aftermath of the flood. Observed depths (usually maximum inundation depths) are less common, because they require ground surveys of inundation marks in the inundated area after the flood. Instrumental records are very rare for inundation depths on floodplains. The recorded depths are compared to the simulated to calculate the residuals. Some examples of application of such data are given in a few studies (Apel et al., 2009; Aronica et al., 1998b; Werner et al., 2005b; Pappenberger et al., 2005). Also Hunter et al. (2005) show that predictions of stage offer considerable potential for reducing uncertainty over effective parameter specification. The errors between the observed and predicted outputs, i.e. the residuals, are formulated as

$$\varepsilon_i = h_{i,\text{obs}} - h_{i,\text{sim}}(\boldsymbol{\theta}) \quad (i = 1, \dots, m) \quad (5)$$

where  $m$  = number of observations;  $\theta$  = vector of model parameters (i.e. roughness coefficients);  $h_{i,\text{obs}}$  = observed water depth at  $i$ -th site;  $h_{i,\text{sim}}(\theta)$  = simulated water depth at the same site generated using the parameter values  $\theta$ .

The aim of PEST is to minimize the Sum of Squares SQ objective function  $F(\theta)$  given by

$$SQ = F(\theta) = \sum_{i=1}^m w_i \cdot (\varepsilon_i)^2 \quad (6)$$

where  $w_i$  are weights that can be assigned to individual observations, resp. errors. The weights in the objective function allow to focus on some observations in the optimisation, thus, trustworthy observations can have a greater weight than those which cannot be trusted as much. In the present study the unit value has been assigned to all residual weights and the measurements errors have been assumed to be independent. In fact, it is very difficult to deal with correlations in observed data, especially when there is no a priori information about them. Under the assumption of independence of the observations, the matrix  $\mathbf{Q}$  defined in Eq. (4) has diagonal elements only and the elements of  $\mathbf{Q}$  are the observation weights.

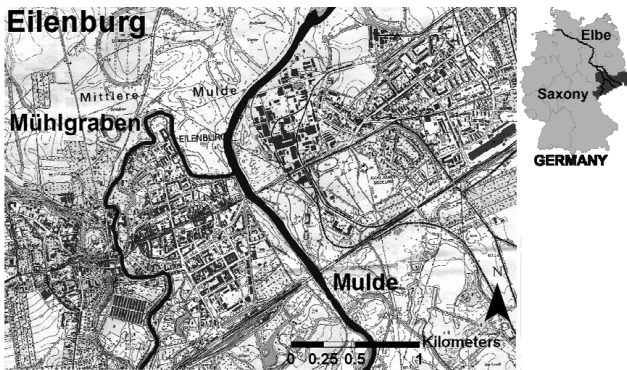
As already pointed out, the Gauss-Levenberg-Marquardt method is gradient based and uses a local search method to find the minimum of the objective function. It can be criticized because it can be trapped in local objective function minima, so that the solution is dependent on the starting point. Global search methods could be used but they require a much greater number of model runs and depending on the particular case of study the cost in terms of time could become prohibitive. However, even the number of model evaluations in gradient based methods in combination with long model run times may also prohibit the application of gradient based methods in many hydraulic model calibrations. For this reason PEST offers a parallel version of the optimiser that considerably decreases the time required for the calibration, if multicore computing facilities are available. Also, the search for the minimum of the objective function using PEST is achieved using fewer model runs than any other parameter estimation method (Doherty, 2004, 2008), which favours the usage of PEST additionally. Computational clusters or office grid solution can be used with PEST. However, at present PEST is restricted to Windows-based operation systems, thus the usual Linux-based large computational clusters cannot be used unless Windows emulating software is installed. The parallel automatic calibration has been tested in this case study, where a single model runtime was approximately 4 h. On a workstation with 8 Intel Xeon X5355 processors at 2.66 GHz CPU speed, the algorithm converged after 4–9 optimisation iterations, which took 4–5 days for each optimisation run.

After the parameter estimation process PEST calculates the 95% confidence limits of the adjustable parameters if the covariance matrix has been calculated. It should be noted

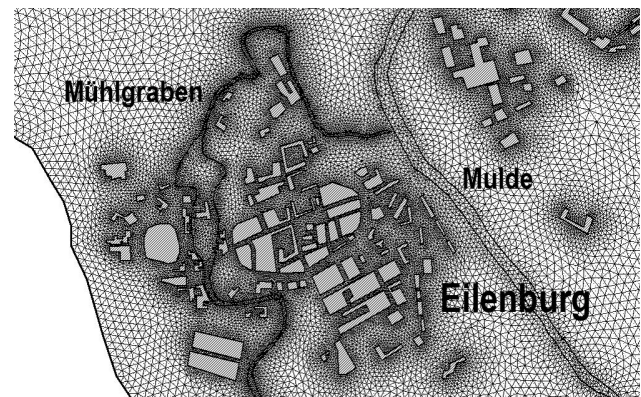
that parameter confidence limits are calculated on the basis of the same linearity assumption which was used to derive the equations for parameter improvement underlying each PEST optimisation iteration. Moreover no account is taken of parameter upper and lower bounds in the calculation of 95% confidence intervals. For example they are not truncated at the parameter domain boundaries so as not to provide a misleading impression of parameter certainty. Thus confidence limits provide only an indication of uncertainty but they are useful to compare different calibration strategies.

### 2.3 Model performance evaluation

Traditionally, hydraulic models have been calibrated using water levels of discharges recorded at the downstream outflow of the model. However, calibration using these data is not the most desirable approach because distributed model performance cannot be tested and stream gauge data can be affected by an error even more than 20% during extreme floods (Bales and Wagner, 2009). Recently a number of authors have therefore used spatial information about the extent of the inundation area derived from post-event flood line surveys, aerial photos, satellite or airborne radar imagery (SAR) data, or LIDAR survey (Hunter et al., 2007) for calibration. Using spatially explicit information of the inundation extent (and in very few cases also inundation depths) the 2-D models can be better constraint, i.e. parameter equifinality reduced. Only a single satellite image of the inundation extent is usually available and this does not allow adequate checking of the inundation dynamics simulated by the model (Woodhead, 2009). An ideal data set would include spatially detailed data like gauged bulk flow data internal to model reach or point maximum water depths or elevations observed e.g. as high-water marks on surviving structures. Internal gauging stations produce data highly resolved in time but they are very uncommon and maximum water depths or elevations cannot be used to test the ability of a model to simulate dynamic flooding (Hunter et al., 2007). Any single data type will only test some aspects of model performance and moreover can be potentially erroneous. However, by combining more than one type of observational data in the calibration process all these limitations should hopefully be overcome, but to date limited applications still exist chiefly because data sets for historical events are quite rare. In this study we used the observed maximum inundation depths only. Other useful calibration criteria like the observed maximum inundation extent were not used, despite having the potential to assess the quality of the spatial inundation prediction of the model (e.g. Bates, 2004). The reason is that due to the specific morphology of the flood plain, which is a rather flat valley confined with steep hill slopes on both sides, the valley wide inundation during this event, and the resolution of the DEM, the information content of the inundation map comes close to zero. The simulated inundation depth at the valley sides could differ several meters without



**Fig. 1.** Investigation area overview and topographical map of Eilenburg.



**Fig. 2.** Layout of the mesh of the simplified 2-D-finite element model.

greatly changing the inundation extent and thus the flood area index comparing the simulated and mapped inundation extend. The inappropriateness of the inundation extend for the present case study was already shown in a previous study using different hydraulic models and manual calibration on the event, the model results of estimating the inundation extend were published in Apel et al. (2009). Therefore, in this case more meaningful indexes for the comparison of the different calibrations are, besides the objective function used in PEST: the mean absolute error (MAE), the root mean square error (RMSE) and the average error (BIAS) of the simulation results from the measured maximum inundation depths calculated as follows

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |\varepsilon_i|; \quad \text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (\varepsilon_i)^2}; \quad (7)$$

$$\text{BIAS} = \frac{1}{m} \sum_{i=1}^m \varepsilon_i$$

where  $m$  = number of data points.

### 3 Case study

The urban area of Eilenburg, located in the federal state of Saxony, Germany, was used in this study. The city is crossed by the Mulde River, a tributary of Elbe, and the Mühlgraben bypass, diverted from the main stream approx. 10 km upstream of Eilenburg (cf. Fig. 1).

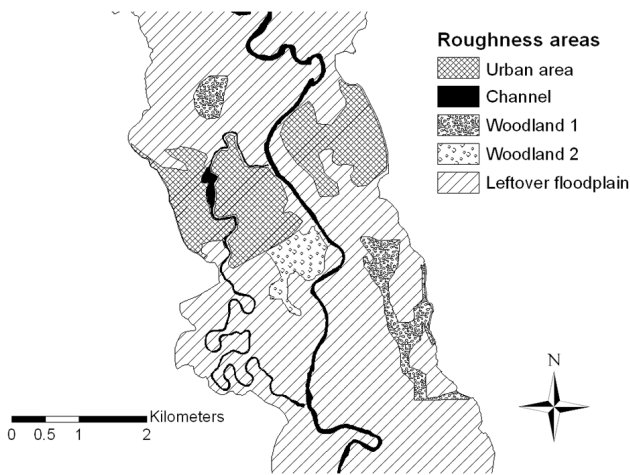
In August 2002 a severe flood event hit many European countries along the Elbe and the Danube rivers and many of their tributaries. Germany was affected, and Saxony was German federal state that suffered the most damage. In particular, the city of Eilenburg and the surroundings were completely flooded; inundation depths up to 5 m in the vicinity of the river and 3 m in the town were reached. Because of the severity of the event, the flooding was well documented:

flood depths were surveyed by laser reflectometry of water marks on buildings above ground at 390 points (data provided by Schwarz et al., 2005), thus yielding detailed point information of inundation depths in the town. The accuracy of the laser measurements is in the mm-order. Some higher errors are introduced when the ground reference level was hard to determine, e.g. by bushes growing in front of the building or unclear water marks. No detailed information is available about this, but the errors of the surveyed maximum inundation depths can be assumed at below 0.2 m. The location of the survey points was determined by handheld GPS. Thus the errors in the location of the survey points are below 6 m. This extensive data set has been used for the calibration of the inundation model.

For the hydraulic model, boundary conditions are always given by the incoming unit flux along the upstream boundary and the water surface elevation at the downstream boundary (Aronica et al., 1998b). Upstream boundary conditions were given by the measured hydrograph at the gauge Golzern, which is the closest gauging station and is located on the Mulde approximately 20 km upstream of Eilenburg. The data recorded by the next downstream gauging station of the Mulde River in Bad Düben could not be used for model calibrations, because the water levels largely exceeded the rating curve. Moreover the gauge was also considerably influenced by the floods of the Elbe River, both from overland flow and the nearby junction of the rivers, and it is located at a considerable distance to the model domain (about 25 flow km). In the model version used in the calibration normal water depth for the channel and zero water depths for the downstream nodes located in the floodplain have been assumed as lower boundary condition.

The 2-D-model operated on an unstructured mesh of 46 417 nodes and 87 945 triangular elements shown in Fig. 2. Floodplain and river topography were derived from a DEM with 25 m spatial resolution, that was generated and issued from the German Federal Authority of Geodesy and





**Fig. 3.** Layout of the spatial roughness distribution in the computational domain considered for the case of study (woodland 1: deciduous forest, woodland 2: low forest interspersed with agriculture).

Cartography (BKG) and is based on topographical maps with a scale of 1:25 000. These maps, especially the elevation information are mainly based on terrestrial surveys supported by remote sensing information in some parts. The errors of the 25 m resolution DEM are estimated by BKG to be below 1 m in the flood plains and up to several meters on steep hillslopes and mountainous areas. Additionally some channel and bank node elevations were taken from channel bathymetric surveys and linearly interpolated between 18 cross sections available for the model domain. The large railway dam and bridge directly upstream of the town (cf. Fig. 1) are represented in the model, whereas the smaller downstream bridge is not. Channel, floodplain and the extent of the domain were digitised from the 1:25 000 maps of the area, the length of the modelled reach is about 8.5 km. The reach was restricted to this stretch in order to keep model runtimes at a tolerable level, which are about 4 h for a single model run simulating five days in the version used for the calibration study.

The Strickler roughness coefficient is the unique parameter involved in the calibration, which is spatially distributed. The model structure allows one coefficient for each triangular element to be used, but based on land use, the domain was divided into five principal regions (Fig. 3). Using the official CORINE land use classification, which is a European standard, four areas were distinguished in the floodplain: the urban area of Eilenburg, two woodlands and the leftover floodplain. The roughness parameterization was consequently performed on the land use classification.

### 3.1 Calibration outline

The model calibration was performed against the water depths, not water elevations. It is often argued that

inundation models perform better when calibrated against water elevations than depths. In case of independently derived water elevations this may be the cause, because the models can compensate for DEM errors through the roughness parameterization. However, we generally argue that calibration against the actual observed variable, the water depth, gives insights into model behaviour and should thus be preferred. Also, in the presented case where no independent observation of water or ground observations are available, the use of water elevations do not improve model performance and calibration. Here the water elevation have to be calculated from the simulated, resp. observed inundation depths and the elevation of the underlying DEM. Since PEST optimises the model based on the sum of squares that in this case evaluates to the same value (cf. Eq. 5):

$$\varepsilon_i = h_{i,obs} - h_{i,sim}(\theta) = (h_{i,obs} + DEM) - (h_{i,sim}(\theta) + DEM)$$

Different calibration strategies were adopted according to different aggregation levels of the roughness regions. In each region one roughness coefficient has been defined uniformly distributed. In the first level, a single and uniform roughness coefficient was adopted for the whole floodplain and the river. In the second level, the river and the floodplain were considered separately. For the third, four roughness areas were considered: the channel and the city areas aggregated in a single region (on the assumption that flow through a mostly paved urban environment, especially with high flow depths as in this case, is similar to open channel flow), the two woodlands and the leftover floodplain. Finally, the last level considers five separate regions according to the CORINE land use distribution shown in Fig. 3.

To overcome the problem of how to choose the range of parameter space, a large range including physical realistic values was considered, thus giving space for the estimation of effective parameters. For both main channel and floodplain the lower value for the Strickler roughness coefficient was set equal to  $5 \text{ m}^{1/3} \cdot \text{s}^{-1}$  (equivalent Manning coefficient  $n$  is  $0.2 \text{ m}^{-1/3} \cdot \text{s}$ ), corresponding to dense wood, and the upper one to  $90 \text{ m}^{1/3} \cdot \text{s}^{-1}$  (equivalent Manning coefficient  $n$  is  $0.011 \text{ m}^{-1/3} \cdot \text{s}$ ), corresponding to concrete. In other studies similar ranges were defined: prior ranges used by Werner et al. (2005b) were loosely based on those given by Chow (1959): the range between  $0.02 \text{ m}^{-1/3} \cdot \text{s}$  and  $0.1 \text{ m}^{-1/3} \cdot \text{s}$  for the main channel, and between  $0.02 \text{ m}^{-1/3} \cdot \text{s}$  and  $0.3 \text{ m}^{-1/3} \cdot \text{s}$  for the floodplain. Pappenberger et al. (2007) adopted a sampling range  $0.01\text{--}0.2 \text{ m}^{-1/3} \cdot \text{s}$  for the channel and  $0.05\text{--}0.3 \text{ m}^{-1/3} \cdot \text{s}$  for the floodplain, imposing channel friction always lower than floodplain friction. Bates and Townley (1988) used  $0.01\text{--}0.05 \text{ m}^{-1/3} \cdot \text{s}$  as main channel Manning values, and the condition  $n_{fl} = 3n_{ch} + 0.01$  for the floodplain. Through the utility PAR2PAR, PEST gives the opportunity to manipulate the parameters before providing them to the model by defining relations between different parameters through mathematical formulations. Thus, in some calibrations, we constraint the

**Table 1.** Different calibration strategies and parameter aggregation used in the study ( $k$  = Strickler roughness coefficients).

Calibration	Number of $k$ parameters	Aggregated areas	Constrains description
A	1	All	unconditioned
B	2	Urban area, woodlands and leftover floodplain as a uniform floodplain	roughness of the river always lower than roughness of the floodplain
C	2	Urban area, woodlands and leftover floodplain as a uniform floodplain	unconditioned
D	4	Urban area and channel as a single land use class/roughness area	roughness of the river always lower than roughness of the floodplain
E	5	All CORINE land use classes	roughness of the river always lower than roughness of the floodplain
F	5	All CORINE land use classes	unconditioned

roughness of the river to be always lower than the roughness of the floodplain. In the context of PAR2PAR this is achieved by defining e.g. for the floodplain a “new” Strickler roughness parameter  $k_{fp\_ratio}$  that is defined by the ratio of the Strickler roughness coefficient in the channel to the Strickler roughness coefficient roughness in the floodplain, and imposing as lower bound of this parameter the unit value:

$$k_{fp\_ratio} = \frac{k_{ch}}{k_{fp}} > 1 \quad (8)$$

However, the way PAR2PAR is implemented limits in the definition of parameter bounds, because only bounds for  $k_{fp\_ratio}$  can be defined, not for the actual parameters roughness parameters  $k_{ch}$  and  $k_{fp}$ . In Table 1 the calibration outline adopted in the study is summarised.

## 4 Results and discussion

### 4.1 Comparison among calibrations

All automatic calibrations performed with PEST converged and successfully improved the objective function. The results in terms of final estimated roughness parameters and confidence bounds are collected in Table 2. Initial values were set following criteria suggested by previous experience gained in manual model calibration (Apel et al., 2009). Except for the roughness coefficient related to the channel, each calibration gave quite low Strickler roughness values, i.e. high hydraulic resistance, for the floodplain compared to literature values. Some remarks on this finding can be made. First, roughness values present in literature are usually referred to results of

**Table 2.** Final roughness estimations and confidence bounds for the calibration scenarios given by PEST. Note: in case of conditioned parameters no confidence bounds can be derived.

$k$ parameter	Initial value	Estimated value	PEST 95% percent confidence limits lower limit	upper limit
A				
all	10	12.00	11.04	12.95
B				
floodplain	20	7.68	–	–
channel	30	35.78	23.69	47.86
C				
floodplain	8	5.59	4.10	7.09
channel	36	50.54	38.44	62.64
D				
floodplain	20	6.46	–	–
channel-urban	30	20.35	15.10	25.60
woodland 1	20	4.07	–	–
woodland 2	20	4.07	–	–
E				
floodplain	20	6.33	–	–
channel	30	31.64	1.40	61.88
woodland 1	20	6.33	–	–
woodland 2	20	6.33	–	–
urban area	15	12.94	–	–
F				
floodplain	8	5.06	1.85	8.27
channel	30	46.60	20.14	73.05
woodland 1	8	5.00	–20.31	30.31
woodland 2	8	9.81	–30.13	49.74
urban area	13	8.62	–2.34	19.59

**Table 3.** Goodness of fit indexes of the different calibration runs using all data points.

Number of parameters	Constrains description	Calibration	SQ (m <sup>2</sup> )	BIAS (m)	MAE (m)	RMSE (m)
1	unconditioned	A	282.1	0.145	0.669	0.851
2	conditioned	B	268.3	0.109	0.622	0.829
2	unconditioned	C	264.9	0.083	0.594	0.824
4	conditioned	D	268.4	0.111	0.614	0.830
5	conditioned	E	265.7	0.106	0.609	0.825
5	unconditioned	F	263.4	0.089	0.589	0.822

one-dimensional model applications, while in this case study a 2-D model code is applied. Second, usual tabular data are referred to micro-roughness condition that is unrealistic for the floodplain surface. Third, the high roughness may have to compensate for errors introduced by the short reach under study and the lower boundary condition imposed (cf. Sect. 3) in absence of surveyed discharges for the lower boundary, as illustrated in Horritt et al. (2007). Further exploring Table 2 it is worth noting that the roughness in the river is in some cases very high. This has to attributed to the intensity of flood event with large discharge and flow depth over the whole floodplain compared to channel width and depth. In this case the influence of the channel flow on maximum floodplain inundation becomes marginal, because the bulk of the flow is overbank.

When parameters are not conditioned, PEST gives 95%-confidence limits associated to the estimated parameter values. These are statistically derived from the distribution of the parameters during the optimisation. As we can see, the larger the number of involved parameters, the larger become the confidence intervals. Moreover, when the calibration considers five parameters the lower limits are even negative. This does not mean that negative roughness parameters were actually used in the calibration. This is rather the result of the statistical derivation of the confidence limits in case of low mean and high variance, where the lower 95%-confidence bounds can eventually evaluate to negative values. Negative roughness values do not make any sense physically, however, the width of the confidence limits indicate how uncertain the estimation of the parameter is: narrow bands imply higher confidence and thus higher importance of the parameter on the results and vice versa. It also indicates the equifinality of model parameterisations that arise by the increasing number of possible parameter combinations able to match the observation data satisfyingly. Following this rationale the inundation depths are dominated by the floodplain area, respectively the floodplain roughness showing the narrowest confidence limits, especially in scenarios C and F. This means that the floodplain surrounding the urban area influence significantly the flow propagation in the urban area. The sensitivity of the model to floodplain roughness is explainable through the

comparatively large extension of that land use class and the fact that it provides the internal boundary conditions for the flow in the urban area.

In case of four conditioned parameters (D), the estimated roughness values for the woodlands fall slightly outside the settled range of parameter space as a result of the conditioning of the parameters. The utility PAR2PAR required the manipulation of the parameters, so that transformed parameters and related bounds are included in the PEST input files, but unfortunately it is very difficult to incorporate and/or control parameter-dependent bounds. Also, the confidence limits are provided for the transformed parameters and not for the roughness parameters of interest.

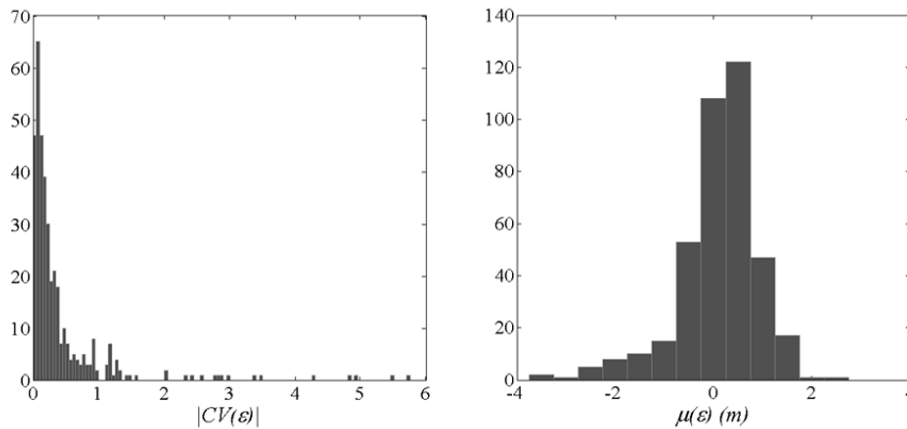
#### 4.1.1 Model performance

As illustrated in Sect. 2.3 we assessed the model performance additionally to the objective function of PEST by the MAE, RMSE and BIAS. The values of these indexes and of the objective function are reported in Table 3. All the models seem to perform equally well with all indexes, with only a slight preference of the non-conditioned parameter sets (calibrations C and F). In general the performance is already satisfying for all scenarios, with little BIAS and RMSE and MAE in the range of the expected DEM error of <1 m in the floodplain. However, this performance assessment additionally emphasizes the equifinality in model parameterisations as already discussed in the previous section. The causes of the equifinality are the distribution and spatial coverage of the and use classes in the simulation domain, but also the distribution of the calibration data. Most of the calibration data points are located in the urban area thus putting an additional emphasis on the roughness of the floodplain and the urban area. However, also erroneous data points or DEM errors at the data points can have a considerable influence on parameter equifinality and thus the automatic calibrating process. If they are not identified and removed or weighed accordingly, as presented up to this point, they can obstruct the search for an optimal solution, because they may dominate the objective function. Another possible reason is the mismatch between model complexity and information content of the



**Table 4.** BIAS calculated with removed data points for the different calibrations.

CV  threshold	No. points	Calibrations						CV(BIAS)
		A	B	C	D	E	F	
0	390	0.145	0.109	0.083	0.111	0.106	0.089	0.200
<b>0.05</b>	<b>343</b>	<b>0.059</b>	<b>0.018</b>	<b>-0.011</b>	<b>0.020</b>	<b>0.014</b>	<b>-0.006</b>	<b>1.578</b>
0.10	278	0.157	0.112	0.080	0.110	0.105	0.083	0.259
0.15	231	0.128	0.076	0.040	0.079	0.072	0.046	0.425
0.20	192	0.111	0.064	0.031	0.062	0.056	0.034	0.483
0.25	162	0.086	0.041	0.012	0.037	0.033	0.013	0.728
<b>0.30</b>	<b>143</b>	<b>0.066</b>	<b>0.021</b>	<b>-0.011</b>	<b>0.016</b>	<b>0.011</b>	<b>-0.009</b>	<b>1.783</b>

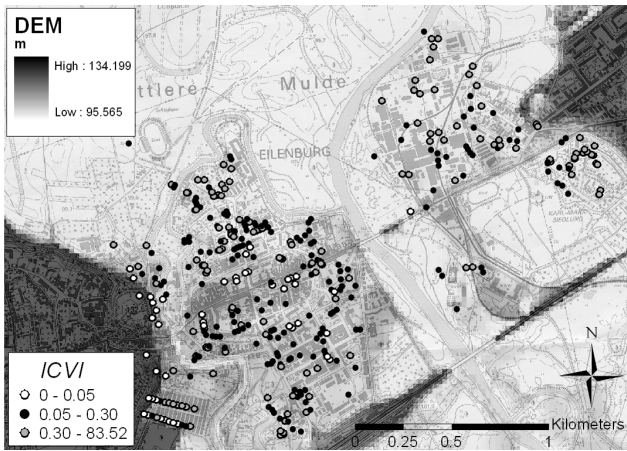


**Fig. 4.** Histograms of absolute coefficient variation and mean of the errors of different calibration runs for every calibration point.

calibration data, the usual cause of equifinality. However, the mismatch in this case is just opposite of the normal case: usually a complex model is calibrated with just a few data points, which are often bulk measurements, e.g. a two dimensional hydraulic model and a downstream discharge hydrograph. In the present case we deal with many data points, whereas the roughness in the respective land use class is not further distinguished, as well as a fairly coarse DEM. This means we use a comparatively simple model setup which is not sufficient to explain the information content of the calibration data properly.

In order to explore the reasons for the equifinality and to possibly reduce it, we take the advantage of the different calibration strategies applied and search for erroneous data points utilizing the different simulation results. We also test whether the calibration is sensitive to a reduced number of calibration points, i.e. an adjustment of data complexity to model complexity. To this end the variance of the residuals of the different calibration results was examined. The idea was to identify and remove calibration points with likely errors (DEM, survey, etc.) and then check for sensitivity of the goodness of fit criteria for the calibration runs without running the calibration again in first place. The criterion adopted was the coefficient of variation (CV) of different calibration

runs for every calibration point. The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean and is a normalized measure of dispersion of a probability distribution, Fig. 4 shows the histograms for the absolute value  $|CV(\epsilon)|$  and the mean,  $\mu(\epsilon)$ , of the errors for all calibration points of all calibration runs. Based on the CV we defined points as erroneous based on the following rational: points with high mean absolute difference of all calibration runs and low variation (standard deviation) caused by different model parameterization were removed as erroneous, because they cannot be explained by model parameterization (or with the current model setup). In terms of CV, these are the points with low CV's (low standard deviation/high mean). Thus points in the calculation with absolute coefficients of variation lower than a threshold were removed. For the selection of the threshold, however, no objective measure can be defined. Therefore several thresholds values were selected:  $|CV|=0, 0.05, 0.1, 0.15, 0.2, 0.25$  and  $0.3$ . After simply removing points with  $|CV|$  less than each threshold the BIAS for every calibration was computed again with the current calibrated model results (Table 4). Inspecting Table 4 it is possible to identify two particular thresholds (0.05 and 0.30, corresponding a 343 and 143 remaining data points) where the BIAS of all calibrations is very low and



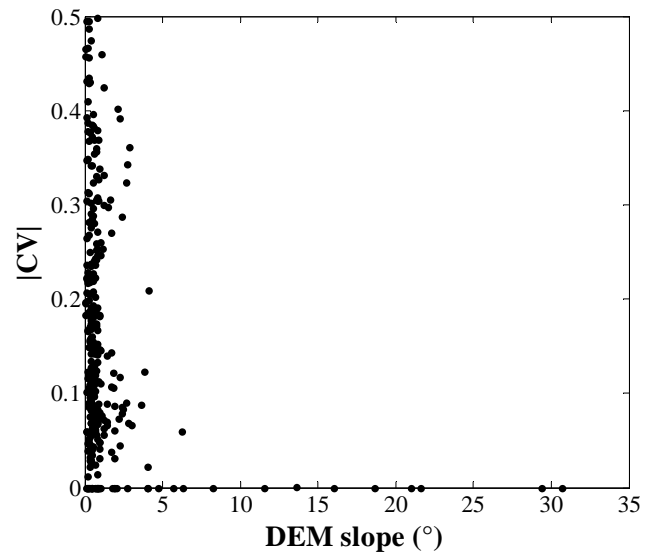
**Fig. 5.** Spatial distribution of absolute coefficient of variation of different calibration runs for every calibration point.

the coefficient of variation of BIAS ( $CV(\text{BIAS})$ ) between the different calibrations is very high. This means that in these cases we see a clear response in the goodness of fit criteria to the different calibration strategies.

In order to find explanations for the possible errors or justifications for the removal of these points, we plotted the spatial distribution of the absolute coefficient of variation grouped according to these two thresholds in Fig. 5. From the spatial distribution of points with  $|CV|$  in the range 0–0.05 we can argue that the quality of the DEM has to be questioned, rather than the quality of the simulation results. Many of these points are situated along the transition of the flat floodplain to the steep valley slopes, where the errors in the DEM are typically the largest, especially with this resolution. For the remaining points in the urban area, especially the old city centre where some small hillocks exist, it is also quite likely that the DEM does not contain the required information about the micro-topography. This qualitative finding is further corroborated by the scatter plot in Fig. 6, where the slope of the DEM is plotted against  $|CV|$ . It can be seen that with high DEM slopes only low absolute coefficients of variance are associated. While high slope values may already hint to possible DEM errors, the lacking sensitivity to different calibration strategies gives a more definite hint.

For the exclusion of points with  $|CV|$  in the range 0.5–0.3 it is hard to find a plausible justification. In general we are now at the point where DEM errors are still likely, but errors in model setup and structure exist at the same time. However, with the available current data we can not distinguish between DEM and model errors any further. On an abstract level it can be argued that with this number of points we reach a match between model and data complexity, but this is hard to prove.

In a next step we ran the so far most successful calibrations C and F (2 and 5 unconditioned roughness classes) in

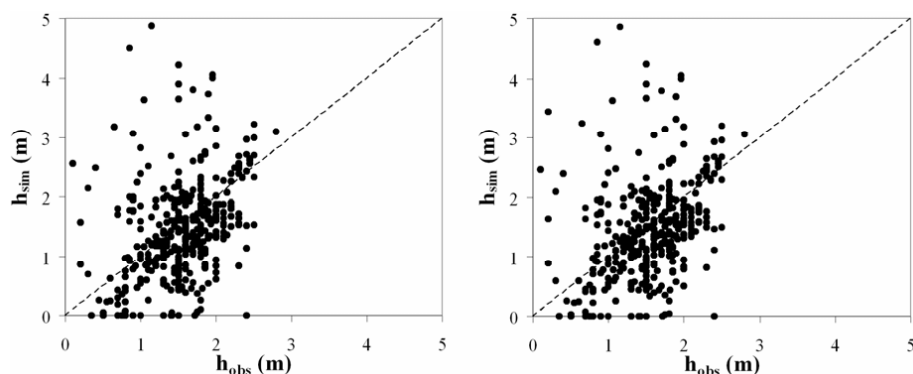


**Fig. 6.** Scatter plot of DEM slope against absolute coefficient of variance  $|CV|$  between the different calibrations for each calibration data point.

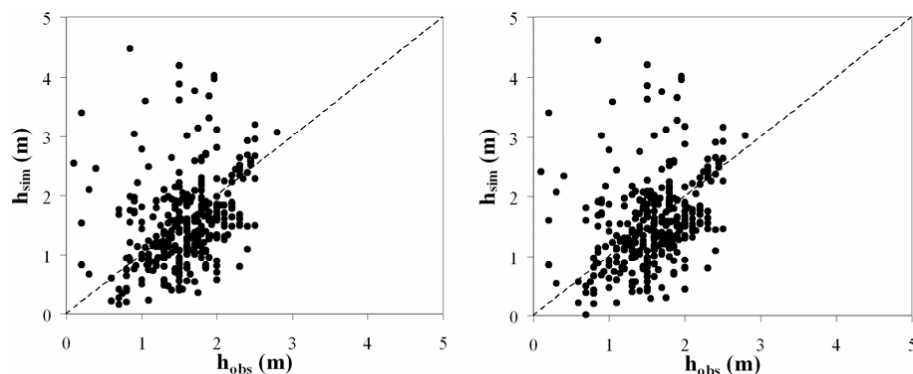
PEST again using 343 and 143 data points only, i.e. points with  $|CV|$  larger than 0.05 and 0.3 respectively. Table 5 shows results in terms of objective function SQ, BIAS, MAE and RMSE. After the second cycle of calibrations (with 343 data points) BIAS is significantly lower as in the calibrations using all data points, but not as low as expected by just calculating the BIAS after removing points, especially for calibration strategy F (cf. Table 4). RMSE and MAE also decreased, but not as significantly as the BIAS.

After the third cycle of calibrations (with 143 data points) all indexes decreased significantly. The BIAS is negligible, but also the MAE and RSME reduced drastically, as well as the objective function. But as mentioned before, at this level it is hard to explain or justify the removal of the points with the available data sets. A thorough inspection and ground survey of the elevation of the points in question could help, but at this point we cannot tell if there are errors in the data points itself, the underlying DEM, the model setup or if we indeed reach a match in data and model complexity.

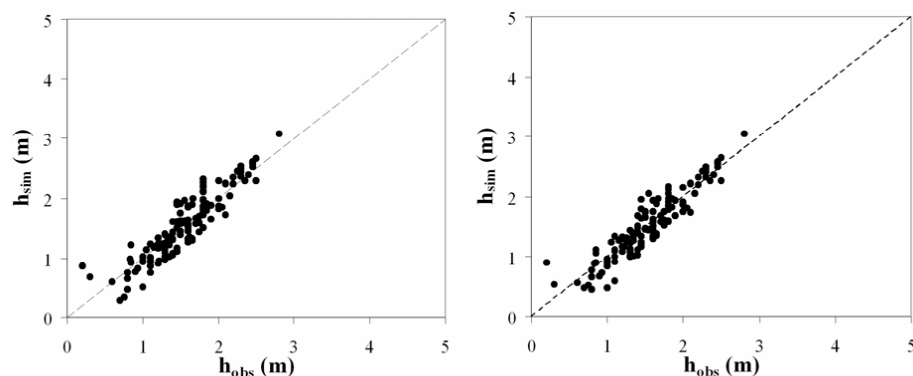
Further investigations of the effects of removing insensitive calibrations points are made by comparing dotted plots shown in Figs. 7, 8 and 9. Due to the high visual similarity between couples of figures, they confirm how the two model setups (with 2 or 5 roughness regions) perform equally well for each calibration procedure (390, 343 and 143 data points), underlining the equifinality theory. These graphs offer the opportunity to explore the reasons for the high RMSE values in Table 5. The high RMSE (especially for 390 data points) is due to the dominance of the outliers, where the simulated water depths drastically under- and overestimate the surveyed. These failures cannot be explained or rectified



**Fig. 7.** Dotty plots related to the calibrations with 390 data points and calibration strategies C (left) and F (right).



**Fig. 8.** Dotty plots related to the calibrations with 343 data points and calibration strategies C (left) and F (right).



**Fig. 9.** Dotty plots related to the calibrations with 143 data points and calibration strategies C (left) and F (right).

by roughness parameterization, this is rather a model error, i.e. the DEM. The most severe errors are removed when using 343 data points, as illustrated in Fig. 8, where especially all severe underpredictions are removed thus causing the improved BIAS. Also, MAE and RMSE improve in the same rate as the BIAS. Using only 143 calibration points the BIAS is removed almost completely, as shown in Fig. 9. Corresponding to the reduced scatter also the MAE and RMSE drop significantly.

Comparing the actual estimated roughness values and the associated confidence intervals for the different number of data sets used in the calibrations given in Tables 6 and 7, it can be observed that the actual roughness estimates do not differ much between the different data sets, except for class woodland 2. However, it has to be noted that this area is located directly upstream of the cities railway station and track, which crosses the valley orthogonally and has the highest elevation in the floodplain. Therefore it can be reasoned that

**Table 5.** Goodness fit criteria of calibrated model results after removing points with low |CV|.

Number of parameters	Calibration	No. points	SQ (m <sup>2</sup> )	BIAS (m)	MAE (m)	RMSE (m)
Calibrations with 390 observed depths						
2	C	390	264.9	0.083	0.594	0.824
5	F	390	263.4	0.089	0.589	0.822
Calibrations with 343 observed depths						
2	C	343	189.4	0.019	0.530	0.743
5	F	343	187	0.018	0.522	0.738
Calibrations with 143 observed depths						
2	C	143	7.2	0.003	0.177	0.224
5	F	143	5.2	0.005	0.058	0.202

**Table 6.** Calibration results as given by PEST considering 343 data points.

<i>k</i> parameter	Initial value	Estimated value	PEST 95% percent confidence limits	
			lower limit	upper limit
C				
floodplain	8	5.74	4.08	7.39
channel	36	51.42	39.92	62.92
F				
floodplain	8	5.00	2.41	7.59
channel	30	45.36	18.70	72.58
woodland 1	8	5.00	-20.41	30.41
woodland 2	8	31.17	-65.05	127.38
city	13	9.83	-2.12	21.77

**Table 7.** Calibration results as given by PEST considering 143 data points.

<i>k</i> parameter	Initial value	Estimated value	PEST 95% percent confidence limits	
			lower limit	upper limit
C				
floodplain	8	5.52	4.79	6.25
channel	36	52.17	47.03	57.31
F				
floodplain	8	5.00	3.93	6.07
channel	30	47.41	37.96	56.85
woodland 1	8	8.63	-6.60	23.85
woodland 2	8	5.00	-7.75	17.75
city	13	8.81	5.12	12.50

the influence of this particular area on the inundation process is largely overruled by the barrier imposed by the railways tracks directly downstream of it. This means in consequence that the simulation results are insensitive to the roughness parameterization in this land use class.

In contrast to the actual estimated roughness values the confidence intervals associated to the values differ considerably between the two calibrations with reduced data sets. Whereas with 343 remaining data points the confidence intervals hardly change compared to the original data set, they are significantly reduced using only 143 data points. Following the rationale applied above, that the confidence intervals can serve as an indicator of the equifinality of the model parameterisation, it can be reasoned that equifinality is reduced in this calibration approach. This, in turn, would also point into the direction of a match in model and data complexity.

## 5 Conclusions

In the present study an automatic calibration procedure has been applied to a 2-D hydraulic model utilising a

comprehensively data set of maximum inundation depths for a flood event occurred in August 2002 in Eilenburg, Germany. The optimiser used was the Parameter ESTimation tool PEST implementing a gradient based minimum search method of the objective function. The method proved to be effective in calibrating the model for different parameterisation strategies. However, by applying different parameterisations and the confidence intervals computed for the estimated roughness values equifinality of model parameterisations could be detected. In particular, with the proposed calibration layout two model setups that perform equally well could be identified. Contrary to the usual case of complex models with large degrees of freedom in the parameter space, which are calibrated against just a few or bulk data, we could illustrate that the opposite situation may also cause equifinality: large degrees of freedom in the data in contrast to a comparatively simple model setup/parameterisation.

A large amount of data or information does not assure an improvement in the identification of the parameters, also the quality of data is very important. This leads to the question whether and how the equifinality can be explained and reduced: is the mismatch in data and model complexity

responsible alone or can we detect errors in data or model setup. To find answers to this question a method for the identification of possible erroneous data points was developed based on the coefficient of variance between the different calibration strategies for every calibration point. This method proved to be successful in improving the model calibration by removing data points with low coefficient of variations from the calibration data set. Dotty plots of observed and simulated water depths confirmed the effectiveness of this approach. Moreover, the removal of the points could partly be explained and justified by DEM errors and resolution. In line with the findings that improving accuracy of DEM data could improve the reliability of flood inundation models (Werner et al., 2005b), that a model that gives a good overall fit to the available data may not give locally good results (Pappenberger et al., 2007), and that the quality of the calibration data is essential for results, the proposed methods helps in identifying erroneous calibration data points that otherwise obstruct proper model calibration. Also, the selection of an appropriate model parameterisation can be supported by the presented method.

However, it has to be noted that above a certain number of removed points, i.e. a certain level of coefficient of variance, no unique or plausible explanation for the removal can be given. However, there are indications that by the removal of about half of the data points a match in model and data complexity is reached, thus enabling a significantly better model calibration and a reduction of equifinality.

While this study gives first insights in the possibilities of automatic calibration of 2-D hydraulic models and the detection of equifinality and erroneous calibration data points, a number of questions remain open: Is the gradient based method efficient in finding the global maximum? How to implement multi-objective optimisations considering e.g. maps of flood inundation extends and time series of discharge and stage at various points in the simulation domain? How to determine the optimal match between data and model complexity? How to consider the uncertainty in calibration data in the automatic calibration? These questions will be the challenges in research for the scientific community in the coming years.

Edited by: J. Freer

## References

- Apel, H., Aronica, G. T., Kreibich, H., and Thielen, A. H.: Flood risk analyses – how detailed do we need to be?, *Nat. Hazards*, 49, 79–98, 2009.
- Aronica, G., Bates, P. D. and Horritt, M. S.: Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE, *Hydrol. Process.*, 16(10), 2001–2016, 2002.
- Aronica, G., Hankin, B., and Beven, K.: Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, *Adv. Water Resour.*, 22(4), 349–365, 1998b.
- Aronica, G. T., Tucciarelli, T., and Nasello, C.: 2-D multi-level model for flood Wave propagation in flood-affected areas, *J. Water Res. Pl.-ASCE*, 124(4), 210–217, 1998a.
- Bales, J. D. and Wagner, C. R.: Sources of uncertainty in flood inundation maps, *J. Flood Risk Manag.*, 2(2), 139–147, 2009.
- Bates, P. D.: Remote sensing and flood inundation modeling, *Hydrol. Process.*, 18(13), 2593–2597, 2004.
- Bates, B.C., and De Roo, A.P.J.: A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, 236(1-2), 54–77, 2000.
- Bates, B. C. and Townley, L. R.: Nonlinear, discrete flood event models, 1. Bayesian estimation of parameters, *J. Hydrol.*, 99, 61–76, 1988.
- Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16(1), 41–51, 1993.
- Beven, K. J.: Equifinality and uncertainty in geomorphological modelling, in: *The Scientific Nature of Geomorphology*, edited by: Roads, B. L. and Thorn, C. E., Wiley, Chichester, 289–313, 1996.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Beven, K. J. and Binley, A. M.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, 6(3), 279–298, 1992.
- Beven, K. J. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29, 2001.
- Chow, V.: *Open Channel Hydraulics*, McGraw-Hill, New York, 1959.
- Doherty, J.: *PEST: Model Independent Parameter Estimation*. Fifth edition of user manual, Watermark Numerical Computing, Brisbane, Australia, 2004.
- Doherty, J.: *Addendum to the PEST Manual*, Watermark Numerical Computing, Brisbane, Australia, 2008.
- Duan, Q., Gupta, V. K., and Sorooshian, S.: Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, *J. Optimiz. Theory Appl.*, 76(3), 501–521, 1993.
- Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and Efficient Global Optimization jbr Conceptual Rainfall-Runoff Models, *Water Resour. Res.*, 28(4), 1015–1031, 1992.
- Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751–763, 1998.
- Horritt, M. S.: a methodology for the validation of uncertain flood inundation models, *J. Hydrol.*, 326(1–4), 153–165, 2006.
- Horritt, M. S., Di Baldassare, G., Bates, P. D. and Brath, A.: Comparing the performance of a 2-D finite element and a 2-D finite volume model of floodplain inundation using airborne SAR imagery, *Hydrol. Process.*, 21, 2745–2759, 2007.
- Horritt, M. S. and Bates, P. D.: Predicting floodplain inundation: raster-based modelling versus the finite element approach, *Hydrol. Process.*, 15, 825–842, 2001.



- Hunter, N. M., Bates, P. D., Horritt, M. S., De Roo, P. J., and Werner, M. G. F.: Utility of different data types for calibrating flood inundation models within a GLUE framework, *Hydrol. Earth Syst. Sc.*, 9(4), 412–430, 2005.
- Hunter, N. M., Bates, P. D., Horritt, M. S., and Wilson, M. D.: Simple spatially-distributed models for predicting flood inundation: A review, *Geomorphology*, 90, 208–225, 2007.
- Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.*, 2, 164–168, 1944.
- Marks, K. and Bates, P. D.: Integration of high-resolution topographic data with floodplain flow models, *Hydrol. Process.*, 14, 2109–2122, 2000.
- Marquardt, D.: An algorithm for least-squares estimation of non-linear parameters, *J. Soc. Indust. Appl. Math.*, 11(2), 431–441, 1963.
- Mignot, E., Paquier, A., and Haider, S.: Modelling floods in a dense urban area using 2D shallow water equations, *J. Hydrol.*, 327(1–2), 186–199, 2006.
- Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., and Horritt, M. S.: Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations, *J. Hydrol.*, 368(1–4), 42–55, 2009.
- Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., and Matgen, P.: Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach, *J. Hydrol.*, 333, 275–287, 2007.
- Pappenberger, F., Beven, K., Horritt, M., and Blazkova, S.: Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, *J. Hydrol.*, 302, 46–69, 2005.
- Romanowicz, R. and Beven, K.: Estimation of flood inundation probabilities as conditioned on event inundation maps, *Water Resour. Res.*, 39(3), W1073, doi:10.1029/2001WR001056, 2003.
- Schumann, G., Matgen, P., Hoffmann, L., Hostache, R., Pappenberger, F., and Pfister, L.: Deriving distributed roughness values from satellite radar data for flood inundation modelling, *J. Hydrol.*, 344(10–2), 96–111, 2007.
- Schwarz, J., Maiwald, H., and Gerstberger, A.: Quantifizierung der Schäden infolge Hochwassereinwirkung: Fallstudie Eilenburg, *Bautechnik*, 82(12), 845–856, 2005.
- Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model, *Water Resour. Res.*, 29(4), 1185–1194, 1993.
- Werner, M. G. F., Blazkova, S., and Petr, J.: Spatially distributed observations in constraining inundation modelling uncertainties, *Hydrol. Process.*, 19, 3081–3096, 2005a.
- Werner, M. G. F., Hunter, N. M., and Bates, P. D.: Identifiability of distributed floodplain roughness values in flood extent estimation, *J. Hydrol.*, 314, 139–157, 2005b.
- Woodhead, S.: Evaluation of Inundation Models, FLOODsite Project, T08-07-01, available at: [http://www.floodsite.net/html/partner\\_area/project\\_docs/T08\\_07\\_01\\_Inundation\\_Model\\_Evaluation\\_M8\\_1\\_V1\\_7\\_P15.pdf](http://www.floodsite.net/html/partner_area/project_docs/T08_07_01_Inundation_Model_Evaluation_M8_1_V1_7_P15.pdf), last access: 2 February 2010, 2009.