Hydrol. Earth Syst. Sci. Discuss., 6, 6833–6864, 2009
www.hydrol-earth-syst-sci-discuss.net/6/6833/2009/

Hydrology and Earth System Sciences Discussions

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio[1,2], G. T. Aronica[2], and H. Apel[3]

[1]Dipartimento di Ingegneria Idraulica e Applicazioni Ambientali, Università di Palermo, Palermo, Italy
[2]Dipartimento di Ingegneria Civile, Università di Messina, Messina, Italy
[3]GFZ German Research Centre for Geoscience, Section 5.4 Hydrology, Potsdam, Germany

Received: 21 October 2009 – Accepted: 23 October 2009 – Published: 6 November 2009

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Hydraulic models for flood propagation description are an essential tool in many fields, e.g. civil engineering, 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 basically two reasons: first, the lack of relevant data against 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. Secondly, especially the two-dimensional 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 full 2-D hyperbolic finite element model. The model independent optimiser PEST, that gives the possibility of automatic calibrations, is used. The application of the parallel version of the optimiser to the model and calibration data 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 with likely errors that obstruct model calibration.

1 Introduction

Floods are serious events and may have severe socioeconomic impacts on vulnerable areas. Thus a reliable evaluation of the inundation extent and depths of a given flood

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

scenario is a very important support for strategic food risk management. Different models to simulate the hydraulic behaviour of a river system are available to date 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. (2004). 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 the models do not represent surface roughness only, but also turbulent momentum losses not explicitly modelled (Werner et al., 2005a). In addition, roughness coefficient often has to compensate insufficient model setup as well, thus becoming what is kown as "efficient" roughness parameters. In general praxis, 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 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 gradient-based approaches, like the Gauss-Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). Global methods are more robust in finding the global minimum in the parameter space of the objective function but they are computationally very demanding because they require a large number of model runs to explore the whole parameter space. Gradient-based methods on the other hand are computational 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 non-linear. 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 times allows.

In the present study the Gauss-Levenberg-Marquardt method implemented in the parameter estimation tool PEST is used (Doherty, 2004, 2008). PEST is considered the most efficient method compared to other gradient based methods (Doherty, 2004, 2008) and it is successful applied in many fields such as groundwater, hydrological and water quality models. This study focuses on different calibration strategies for a two-dimensional hydraulic model. It is calibrated against a serious flood event that occurred on 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.

HESSD
6, 6833–6864, 2009

For flood propagation models hardly ever sufficient calibration data exist, but in the study area the historical event was well documented in the aftermath of the flood. This study will 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 certainly improve the parameter estimation (Sorooshian et al., 1993). Here a procedure to remove potentially erroneous data is also presented.

2 Methodology

2.1 Two-dimensional model

In order to model the flow regime in an urban area, a detailed full two-dimensional 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. (1998b) was applied. It is a hyperbolic model based on de Saint-Venant equations for two-dimensional 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 two-dimensional shallow-water flow can be written as follows

$$\frac{\partial H}{\partial t}+\frac{\partial p}{\partial x}+\frac{\partial q}{\partial y}=0, \tag{1}$$

$$\frac{\partial p}{\partial t}+gh\frac{\partial H}{\partial x}+ghJ_x=0; \qquad \frac{\partial q}{\partial t}+gh\frac{\partial H}{\partial y}+ghJ_y=0, \tag{2}$$

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 ($\mathrm{m}^{1/3} \cdot \mathrm{s}^{-1}$) is related to Manning coefficient n ($\mathrm{m}^{1/3} \cdot \mathrm{s}^{-1}$) through

$$k = \frac{1}{n}. \tag{3}$$

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 hilltop, blocks and other obstacles are treated as internal islands within the triangular mesh covering the entire flow domains. Moreover, the finite element method allows defining spatially explicit roughness coefficients for the floodplain inundation. For these reasons, the model requires detailed topographic information: 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 (acronym for Parameter ESTimation) (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. An automatic calibration offers the advantage of using the speed and the power of computer, resulting in time saving for modellers. But even more, applied in hydraulic modelling, it is a step towards a more objective and encompassing model calibration compared to the general practice.

PEST adjusts model parameters to obtain the best match between model generated values and the correspondent measurements in the weighted least squared sense. Given the number of degrees of freedom present in the calibration procedure of two-dimensional models, whereby separate friction parameters can theoretically be assigned at each computational node and at each time-step (Marks and Bates, 2000), this task is both relatively easy to accomplish, but also quite difficult because of the equifinality problem (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 follows

$$\boldsymbol{p} - \boldsymbol{p_0} = (\mathbf{J}^t \cdot \mathbf{Q} \cdot \mathbf{J} + \lambda\, \mathrm{diag}[\mathbf{J}^t \cdot \mathbf{Q} \cdot \mathbf{J}])^{-1} \cdot \mathbf{J}^t \cdot \mathbf{Q} \cdot \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{p_0}$=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, λ=the Marquardt lambda.

HESSD
6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

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, i.e. manpower. Instrumental records are very rare for inundation depths on floodplains. The recorded depths are compared to the simulated to evaluate the residuals. Some examples of application of such data are given in a few studies (Apel et al., 2009; Aronica et al., 1998a; Werner et al., 2005b). 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 can be written as follows

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

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

The aim is to minimize the Least Squared object function given by

$$F(\boldsymbol{\theta}) = \sum_{i=1}^{m} w_i \cdot (\varepsilon_i)^2, \tag{6}$$

where w_i=weight that can be assigned to individual errors. In the present study the unit value was assigned to all residuals.

As already pointed out, the Gauss-Levenberg-Marquardt method is gradient based and use a local search method to find the minimum of the objective function. It can be criticized because too easily it can be trapped in local objective function minima, so that the solution is dependent on the starting point. Global 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, 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 (in this case study, the model runtime was approximately 4 h and the computational cost of every PEST calibration was 4–5 d on a workstation with 8 Intel Xeon X5355 processors with 2.66 GHz CPU speed, reached after 4–9 optimisation iterations).

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. I.e. 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.

3 Case study

As case study the urban area of Eilenburg, located in Saxony, Germany, has been selected. 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 (in Fig. 1 a topographical map is shown).

In August 2002 a severe flood event hit many European countries along Elbe, Danube rivers and some their tributaries. Germany was affected, and Saxony was the German federal state the most damaged. 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 enormous extent, the flooding was well documented: flood depths were recorded from water marks at 390 buildings in the city centre thus yielding detailed point information of inundation depths in the town and were provided by Schwarz et al. (2005). This extensive data set has been used for the calibration of the inundation model. Upstream boundary conditions were given by the measured hydrograph at the gauge Golzern, which is the closest gauging station. 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 be 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. 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 25 m DEM, moreover some channel and bank node elevations were taken from channel surveys and linearly interpolated between 18 cross sections. Channel plan form and the extent of the domain were digitised from 1:25 000 maps of the area.

In the flood propagation model considered for this study, the Strickler roughness coefficient is the unique parameter involved, which is, however, 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). Four areas were distinguished in the floodplain: the urban area of Eilenburg, two woodlands and the leftover floodplain.

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

4 Results and discussion

4.1 Calibration outline

Different calibration strategies were adopted according to different aggregation levels of the roughness regions, where an ensemble average roughness coefficient was assumed. 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, the two woodlands and the leftover floodplain. Finally, the last level considers five separate regions according to the 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 $5\,\mathrm{m}^{1/3}\cdot\mathrm{s}^{-1}$ (equivalent Manning coefficient n is $0.2\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$), corresponding to dense wood, and the upper one to $90\,\mathrm{m}^{1/3}\cdot\mathrm{s}^{-1}$ (equivalent Manning coefficient n is $0.011\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$), 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 $0.02\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ and $0.1\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ for the main channel, and between $0.02\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ and $0.3\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ for the floodplain. Pappenberger et al. (2007) adopted a sampling range $0.01–0.2\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ for the channel and $0.05–0.3\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ for the floodplain, imposing channel friction always lower than floodplain friction. Bates and Townley (1988) used $0.01–0.05\,\mathrm{m}^{-1/3}\cdot\mathrm{s}$ 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. Thus, in some calibrations, we constraint the roughness of the river to be always lower than the roughness of the floodplain. In Table 1 the calibration outline adopted in the study is summarised.

4.2 Comparison among calibrations

The results given by PEST 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 estimated Strickler values, i.e. high hydraulic resistance, for the floodplain. Some remarks can be made. First, roughness values present in literature are usually referred to results of one-dimensional models applications, while in this case study a full two-dimensional model code is applied. Second, usual tabular data are referred to micro-roughness condition that is unrealistic for the floodplain surface. Exploring Table 2, roughness in the river is in some cases very high, in the given event with large discharge and flow depth over the whole floodplain compared to channel width and depth, the influence of the channel on maximum floodplain inundation becomes marginal. When parameters are not conditioned, PEST gives us the 95% percent confidence limits. As we can see, the larger the number of involved parameters the larger become the confidence interval. Moreover, when the calibration considers five parameters the lower limits are even negative. This does not make any sense physically, but gives us an idea of how uncertain the estimation of the parameter set is. 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.

In case of four conditioned parameters, the estimated roughness values for the woodlands fall slightly outside the settled range of parameter space, even so the calibration has been considered still acceptable. 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. Even the confidence limits are provided for the transformed parameters and not for the roughness parameters of interest.

4.2.1 Assessing model performance

In this study we used the observed maximum inundation depths only. Other useful calibration criteria like the observed maximum inundation extend 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 changing the inundation extent and thus the flood area index comparing the simulated and mapped inundation extend. 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

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

where m=number of data points.

The values of these indexes and of the objective function as calculated by PEST with Eq. (6) (unit error weights were used) are reported in Table 3. All the models seem to perform equally well, with only a slight preference of the non-conditioned parameter sets (calibrations B and F), giving an idea of equifinal model parameterisations. This behaviour could be attributed to different reasons. First, because of the extraordinary magnitude of the flood the influence of different roughness areas in the floodplain is apparently suppressed. Second, even more likely the equifinality is caused by the spatial location and the quality of the calibration data. Most of the points are situated in the urban area, therefore a clear distinction among the different roughness areas

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

cannot be made, especially because some of them are not large compared to the others. Third, also erroneous data points or DEM errors at the data points can have a considerable influence on 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. Fourth, the mismatch between model complexity and calibration data, the usual cause for equifinality, surely has an influence here. 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 have many data points, all within just one assumed roughness class, i.e. 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 possibly reduced 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. For 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 $|CV(\varepsilon)|$ and the mean, $\mu(\varepsilon)$, 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 the coefficient of variation of BIAS (CV(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 chronically 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 doesn't contain the required information about the micro-topography.

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 one could argue 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 B and F (2 and 5 unconditioned roughness classes) in 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 F, 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.

Comparing the actual estimated roughness values and the associated confidence interval 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 orthogonal and has the highest elevation in the floodplain. Therefore it can be reasoned that 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.

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 rational 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. 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.

This lead 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. The removal of the points could partly be explained and justified by DEM errors. 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.

References

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H.: Flood risk analyses – how detailed do we need to be?, Nat. Hazards, 49, 79–98, 2009.

Aronica, G. T., 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, 1998a.

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, 1998b.

Bates, P. D.: Remote sensing and flood inundation modeling, Hydrol. Process., 18(13), 2593–2597, 2004.

HESSD

6, 6833–6864, 2009

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 App., 76(3), 501–521, 1993.
Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimization for 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. 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, A. P. J., and Werner, M. G. F.: Utility of different data types for calibrating flood inundation models within a GLUE framework, Hydrol. Earth Syst. Sci., 9, 412–430, 2005, http://www.hydrol-earth-syst-sci.net/9/412/2005/.
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. Ind. Appl. Math., 11(2), 431–441, 1963.

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, 2004.

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.

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 observationsin 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.

Table 1. Different calibration strategies and parameter aggregation used in the study.

Calibration	Number of k parameters	Aggregated areas	Constrains description
A	1	All	unconditioned
B	2	Urban area, woodlands and leftover floodplain	roughness of the river always lower than roughness of the floodplain
C	2	Urban area, woodlands and leftover floodplain	unconditioned
D	4	City and channel	roughness of the river always lower than roughness of the floodplain
E	5	None	roughness of the river always lower than roughness of the floodplain
F	5	None	unconditioned

HESSD
6, 6833–6864, 2009

Table 2. Calibration results as given by PEST.

k parameter	Initial value	Estimated value	PEST 95% percent confidence limits lower limit	upper limit
		1 parameter		
all	10	12.00	11.04	12.95
		2 conditioned parameters		
floodplain	20	7.68	–	–
channel	30	35.78	23.69	47.86
		2 free parameters		
floodplain	8	5.59	4.10	7.09
channel	36	50.54	38.44	62.64
		4 conditioned parameters		
floodplain	20	6.46	–	–
channel-city	30	20.35	15.10	25.60
woodland 1	20	4.07	–	–
woodland 2	20	4.07	–	–
		5 conditioned parameters		
floodplain	20	6.33	–	–
channel	30	31.64	1.40	61.88
woodland 1	20	6.33	–	–
woodland 2	20	6.33	–	–
city	15	12.94	–	–
		5 free parameters		
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	49.75
city	13	8.62	−2.34	19.59

HESSD

6, 6833–6864, 2009

Title Page

Abstract Introduction

Conclusions References

Tables Figures

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 3. Indexes calculated for calibration comparison.

Number of parameters	Constrains description	Calibration	F (m^2)	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

HESSD
6, 6833–6864, 2009

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

|CV| threshold	No. points	Calibrations A	B	C	D	E	F	CV(BIAS)
0	390	0.145	0.109	0.083	0.111	0.106	0.089	0.200
0.05	**343**	**0.059**	**0.018**	**−0.011**	**0.020**	**0.014**	**−0.006**	**1.578**
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
0.30	**143**	**0.066**	**0.021**	**−0.011**	**0.016**	**0.011**	**−0.009**	**1.783**

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

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

Number of parameters	Calibration	No. points	$F(m^2)$	BIAS (m)	MAE (m)	RMSE (m)
Calibrations with 390 observed depths						
2	B	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	B	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	B	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.

k parameter	Initial value	Estimated value	PEST 95% percent confidence limits lower limit	upper limit
		2 free parameters		
floodplain	8	5.74	4.08	7.39
channel	36	51.42	39.92	62.92
		5 free parameters		
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

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

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

k parameter	Initial value	Estimated value	PEST 95% percent confidence limits lower limit	upper limit
		2 free parameters		
floodplain	8	5.52	4.79	6.25
channel	36	52.17	47.03	57.31
		5 free parameters		
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

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

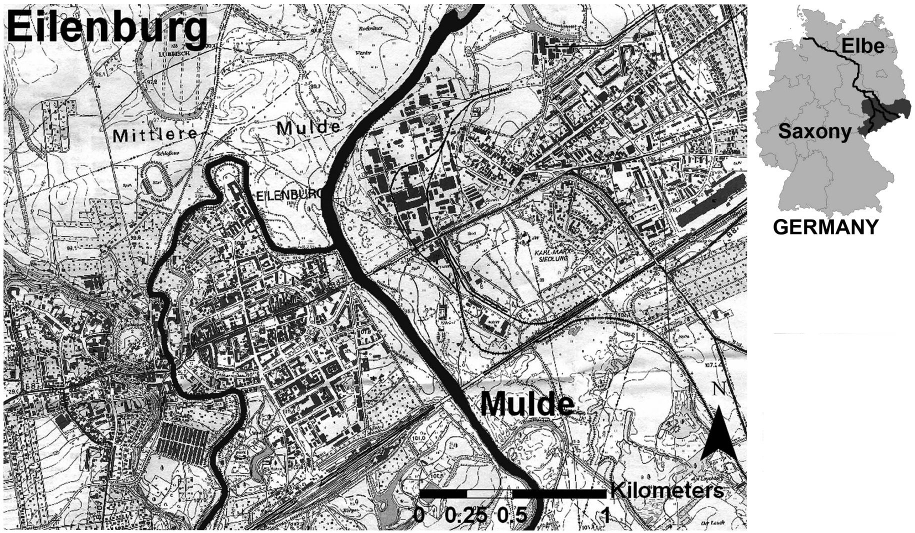

Fig. 1. Investigation area overview and topographical map of Eilenburg.

HESSD
6, 6833–6864, 2009

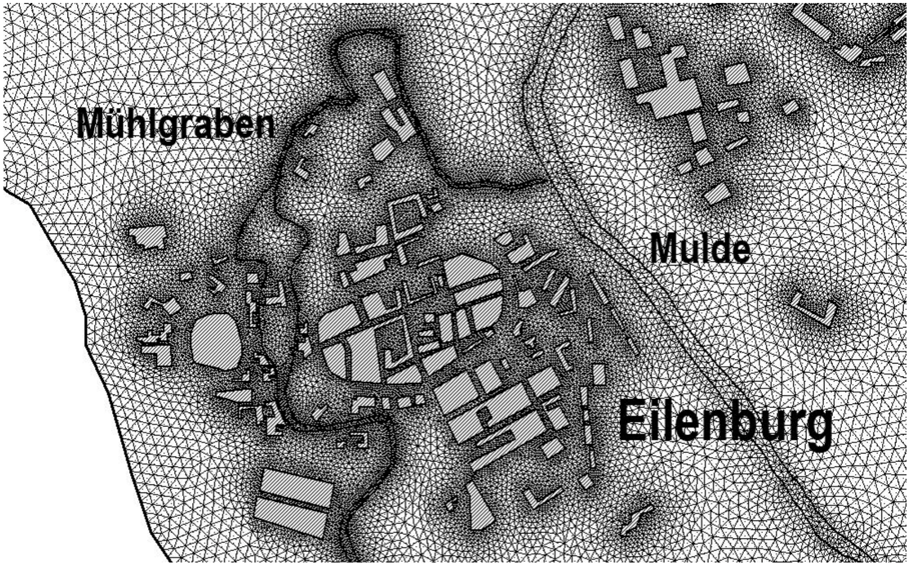

Fig. 2. Layout of the mesh of the full 2-D-finite element model.

HESSD
6, 6833–6864, 2009

HESSD

6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.

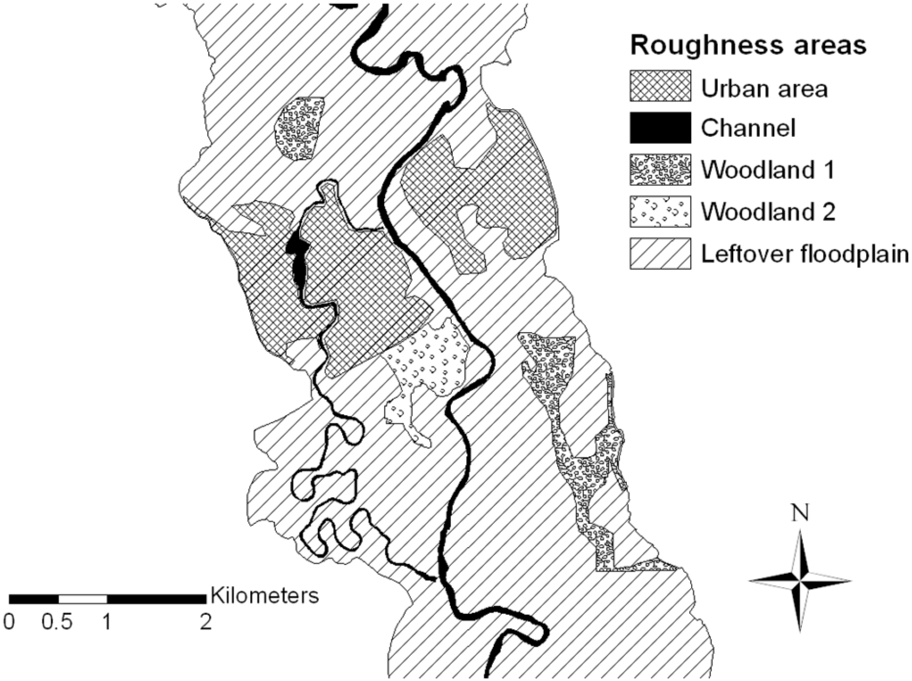

Fig. 3. Layout of the spatial roughness distribution in the computational domain considered for the case of study (Woodland 1: deciduous forest, woodland 2: low forest interspersed with agriculture).

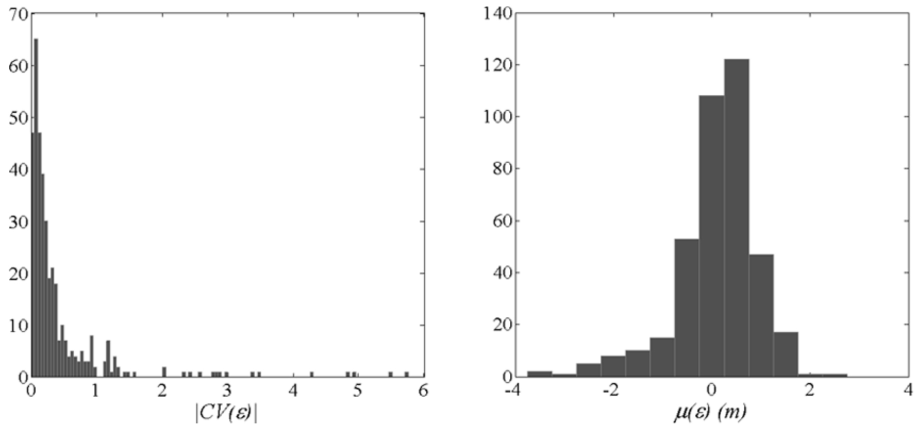

Fig. 4. Histograms of absolute coefficient variation and mean of the errors of different calibration runs for every calibration point.

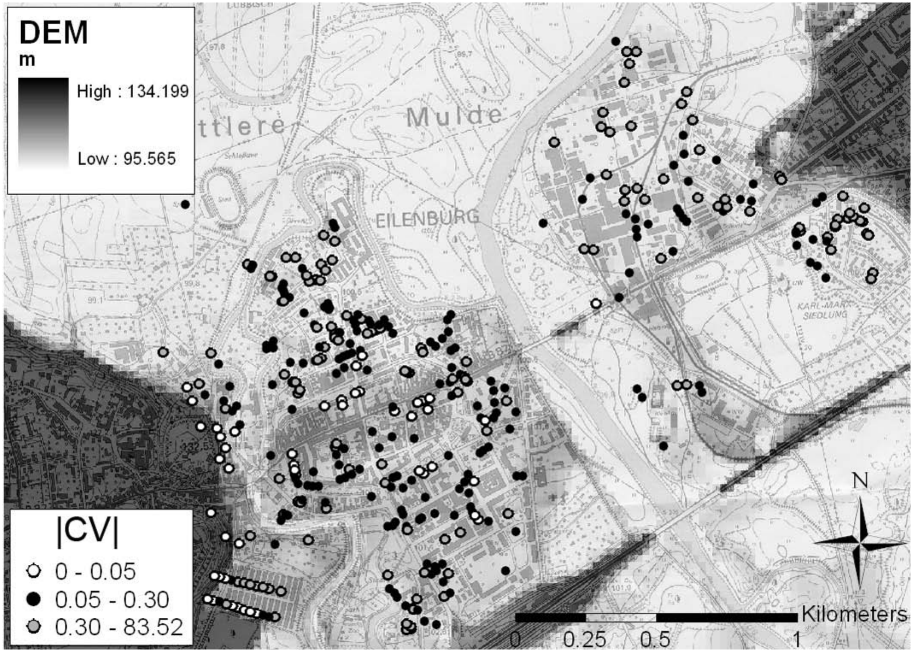

Fig. 5. Spatial distribution of absolute coefficient variation of different calibration runs for every calibration point.

HESSD
6, 6833–6864, 2009

Towards automatic calibration of 2-dimensional flood propagation models

P. Fabio et al.