

Efficient calibration of a distributed *pde*-based hydrological model using grid coarsening



D. von Gunten^a, T. Wöhling^{a,c}, C. Haslauer^a, D. Merchán^b, J. Causapé^b, O.A. Cirpka^{a,*}

^a University of Tübingen, Center for Applied Geoscience, Hölderlinstr. 12, 72076 Tübingen, Germany

^b Geological Survey of Spain IGME, C/Manuel Lasala no. 44, 9B, Zaragoza 50006, Spain

^c Lincoln Agritech Ltd., Ruakura Research Centre, Hamilton, New Zealand

ARTICLE INFO

Article history:

Received 19 June 2014

Received in revised form 5 October 2014

Accepted 8 October 2014

Available online 22 October 2014

This manuscript was handled by Peter K. Kitanidis, Editor-in-Chief, with the assistance of Wolfgang Nowak, Associate Editor

Keywords:

Integrated hydrological model

HydroGeoSphere

Model calibration

Grid hierarchy

Spatial discretization

SUMMARY

Partial-differential-equation based integrated hydrological models are now regularly used at catchment scale. They rely on the shallow water equations for surface flow and on the Richards' equations for subsurface flow, allowing a spatially explicit representation of properties and states. However, these models usually come at high computational costs, which limit their accessibility to state-of-the-art methods of parameter estimation and uncertainty quantification, because these methods require a large number of model evaluations. In this study, we present an efficient model calibration strategy, based on a hierarchy of grid resolutions, each of them resolving the same zonation of subsurface and land-surface units. We first analyze which model outputs show the highest similarities between the original model and two differently coarsened grids. Then we calibrate the coarser models by comparing these similar outputs to the measurements. We finish the calibration using the fully resolved model, taking the result of the preliminary calibration as starting point. We apply the proposed approach to the well monitored Lerma catchment in North-East Spain, using the model HydroGeoSphere. The original model grid with 80,000 finite elements was complemented with two other model variants with approximately 16,000 and 10,000 elements, respectively. Comparing the model results for these different grids, we observe differences in peak discharge, evapotranspiration, and near-surface saturation. Hydraulic heads and low flow, however, are very similar for all tested parameter sets, which allows the use of these variables to calibrate our model. The calibration results are satisfactory and the duration of the calibration has been greatly decreased by using different model grid resolutions.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Recently, partial-differential-equation (*pde*) based hydrological models, that couple the shallow-water equations for surface flow and the Richards' equations for subsurface flow, have been successfully applied in various settings, from catchment scale (e.g., Condon et al., 2013; Goderniaux et al., 2011; Li et al., 2008; Shao et al., 2013) to continental scale (Lemieux et al., 2008). They are regarded as useful tools to represent hydrological processes, especially when studying spatially distributed surface–subsurface interactions or catchments driven by climatic or irrigation changes (Pérez et al., 2011), two problems difficult to analyze with simpler “bucket”-type models. However, *pde*-based models are usually computationally very demanding (Blasone et al., 2008) and sometimes require days of CPU time for a single forward run (Goderniaux et al., 2009) on a current desktop computer.

As a result, calibration of these models, which typically requires a large number of model evaluations, can be a slow, tedious, and subjective process. To reduce the number of simulations needed, model calibration of *pde*-based models is often limited to a trial-and-error process (e.g., Bonton et al., 2012; Calderhead et al., 2011; Goderniaux et al., 2009; Li et al., 2008; Pérez et al., 2011; Xevi et al., 1997), even though Blasone et al. (2008) and McMichael et al. (2006) have also proposed ensemble-based approaches. In the latter approaches, multiple parameters sets are generated by Monte-Carlo methods and weighted or modified depending on the likelihood of model outcomes in comparison to measurement (Beven and Binley, 1992). However, the large number of simulations needed renders approaches involving Monte-Carlo methods almost impossible for large *pde*-based models because of their long simulation times. The importance of sensitivity analysis to decrease the number of calibration parameters, and thus the required number of simulation runs, has also been recognized (e.g., Muleta and Nicklow, 2005; Christiaens and Feyen,

* Corresponding author.

E-mail address: olaf.cirpka@uni-tuebingen.de (O.A. Cirpka).

2002). Nevertheless, even for the most efficient calibration methods available today, the number of simulations needed would still be too large to apply these methods to integrated *pde*-based models. Therefore, model reduction is currently the only feasible option to calibrate such models.

Simulation time is known to greatly depend on spatial discretization (e.g., Vazquez et al., 2002). Typically, finer grids cause a non-linear increase of computational costs compared to coarser ones due to the larger number of unknowns. However, a sufficiently fine mesh is needed to realistically represent the topography of the catchment, which is important to properly simulate run-off, infiltration, and surface–subsurface exchange fluxes, or to characterize zones with large changeability in state variables. Therefore, to trade off model accuracy and simulation time, the spatial discretization should be chosen carefully in distributed models. Various studies have been conducted to find the minimum spatial discretization needed to adequately represent the catchment under consideration (e.g., Bruneau et al., 1995; Carrera-Hernandez et al., 2012; Chaplot, 2014; Chaubey et al., 2005; Cotter et al., 2003; Dutta and Nakayama, 2009; Kuo et al., 1999; Moglen and Hartman, 2001; Molnár and Julien, 2000). In general, the grid cell size must be smaller for catchments with highly uneven relief than for those with smooth topography (Chaplot, 2014). In addition, problems as the modeling of erosion (Hessel, 2005), spatially varying evapotranspiration (Sciuto and Diekkrüger, 2010), or reactive transport (Chaplot, 2005) are more sensitive to grid size than the simulation of hydraulic heads or discharge, especially low flow. The choice of spatial discretization therefore depends also on the simulation objectives (Cotter et al., 2003).

The importance of the resolution of spatial discretization has been recognized before, particularly in studies on reactive transport (Mehl and Hill, 2002). In this field, grid telescoping, i.e., the modeling of a reactive transport problem using two grids, namely a coarse grid representing the whole catchment and a fine one representing the surrounding area of the contaminated plume, is relatively common (Mehl and Hill, 2002; Mehl et al., 2006). However, grid telescoping requires a well defined inner domain of interest and an outer domain from which conditions can be extracted and used as boundary conditions for the inner domain. This is not suitable in all situations, e.g., it cannot be used for large non-point contamination problems, such as agricultural nitrate leaching from cultivated land.

Recently, more attention has been given to the influence of spatial discretization on model calibration of *pde*-based hydrological model. For example, Wildemeersch et al. (2014) analyzed parameter sensitivity, i.e., the influence of model parameters on the simulation output, and the linearized confidence interval for various spatial discretizations. These quantities were found to be very similar for all grid sizes, in this synthetic case study based on a Belgian catchment of approximately 300 km².

In the present study, we also focus on the links between model calibration and grid resolution. We propose a methodology to accelerate calibration in fully coupled *pde*-based hydrological models. Our main objective is to reduce simulation time, while obtaining a final model with a precise description of topography. To reach this objective, we vary the resolution of spatial discretization during the calibration. Moreover, we analyze how the changing grid resolution affects the model outcome and test the validity of our method in a case study in North-East Spain.

A prerequisite of the present analysis is that the subsurface structure and land-use at the surface is represented by zonation, which – in principle – is represented on all grid levels. Important questions of coarse-graining fine-scale information onto the resolution scale of larger grids are beyond the scope of the present study. Calibrating systems that account for internal heterogeneity

would require a multi-scale representation of the domain, robust coarse-graining rules, and the provision of fine-scale proxy data used in the calibration, which has been done in conceptual hydrological modeling (Samaniego et al., 2010) but not yet in *pde*-based catchment-scale models.

The remainder of this paper is structured as follows: First, we describe the principles of the proposed calibration method. Then, we present the governing equations of the numerical model HydroGeoSphere (Therrien, 2006), used in this study, and the study area. This is followed by the construction of the conceptual model for the test case. Afterward, we compare the outputs of the model when using different computational grids. Finally, we report the results of the sensitivity analysis and the calibration of our case study. We conclude with an evaluation of our calibration method.

2. Proposed calibration strategy

The proposed method to accelerate calibration utilizes a set of coarser grids on which simulations run faster than on the original fine grid. The coarser grids should be coarse enough to noticeably decrease computation time while capturing enough system behavior to be useful for calibration. We suggest to use two auxiliary grids, a coarse grid and an intermediate one.

The coarse grid is used to largely constrain model parameters in the full parameter space. The model outputs based on simulations using this grid and the fine grid should be comparable but may still show large differences, for example a consistent bias in model prediction along the parameters sets. Model results from simulations with the intermediate grid should be more comparable to those of the fine grid than the coarse-grid results, but differences may prevail. Indeed, if the intermediate grid would yield identical results in comparison to the fine grid, the latter grid would be unnecessary. We will show in the following that, upon grid refinement, hydraulic head and low flow can be adequately represented in a coarse grid while peak flow, evapotranspiration and saturation needs to be modeled on a finer grid, at least in the catchment under consideration in this study.

Conceptually, the proposed calibration method consists of the following seven steps, summarized in Fig. 1:

1. Set up of three computational grids:
 - a fine grid used in the final model,
 - an intermediate grid used to restrict the possible parameter space,
 - and a coarse grid used to estimate the possible parameter space.
2. Systematic comparison of the simulation results using different computational grids.
3. Parameter sensitivity analysis on the intermediate grid.
4. Constraining the feasible parameter space using the coarse grid.
5. Calibration of model parameters on the intermediate grid.
6. Transfer of the model parameters to the fine grid and eventual final parameter adjustments.
7. Model validation and evaluation on the fine grid.

The choice of the discretization for the coarse and intermediate grids depends on the impact of the grid coarsening on the model outputs used for calibration, such as stream flow hydrographs or hydraulic heads. This therefore also depends on catchment characteristics such as the topography or soil-hydraulic parameters. However, the size of the grid cells used in our case study, described in Section 6, may serve as a starting point in other applications. The method can be adapted to use more or less than three grids, depending on the complexity of the problem.

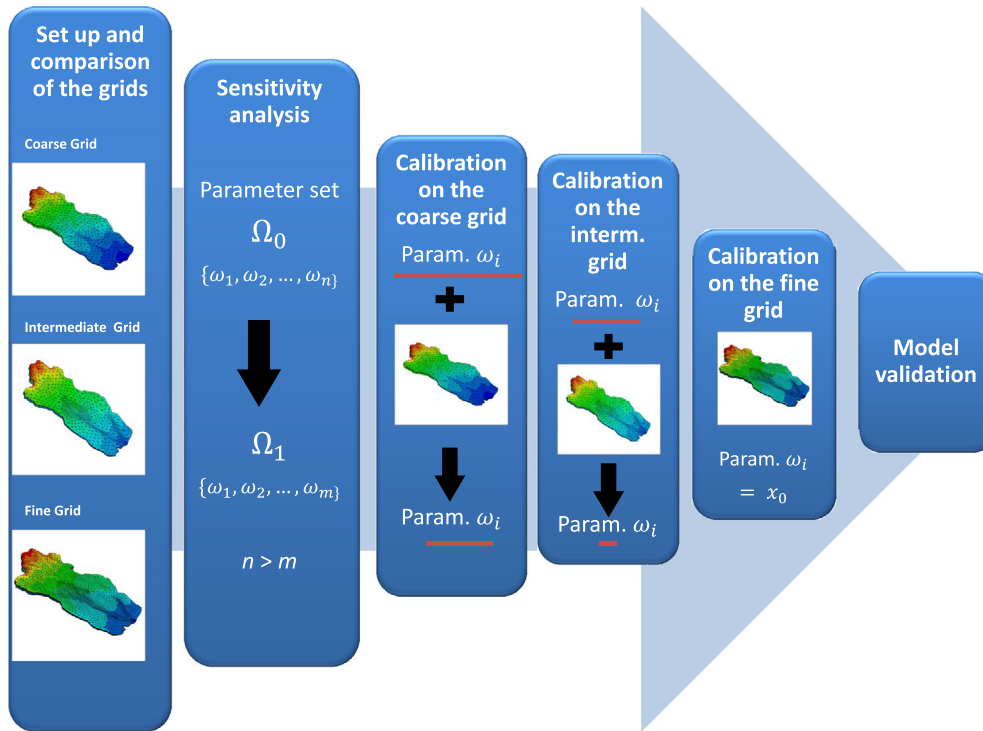


Fig. 1. A summary of the proposed calibration method.

In addition to determining the spatial discretization, the comparison between the grids also gives information about what part of the measured data should be used in each calibration step. It may be beneficial to use only part of the measured data on the coarse or intermediate grids and the whole data set at a later stage, when the fine grid is used. For example, in our case study, low flow is better represented on the coarse grid than peak flow. As a result, low flow can be used earlier during calibration, when the model is run with the coarse grid, while peak flow might be better considered on the fine grid.

3. Mathematical and numerical model

The calibration method described above could be applied to many grid-based distributed hydrological models. In this work, HydroGeoSphere has been selected, as it is a well established fully coupled *pde*-based numerical model used to simulate ground- and surface-flow processes at catchment scale (Therrien, 2006). In this section, the underlying assumptions and equations of HydroGeoSphere are briefly summarized. A more precise description is given by Therrien et al. (2010).

3.1. Subsurface flow

HydroGeoSphere uses the Richards' equation to model variably saturated sub-surface flow in porous media:

$$\frac{\partial(S_w \theta_s)}{\partial t} - \nabla \cdot (\mathbf{K} \mathbf{k}_r \nabla h) = \sum \Gamma_{sub} \pm Q \quad (1)$$

in which S_w [–] represents the degree of water saturation, θ_s [–] is the saturated water content, assumed to be identical with the porosity, t [s] is time, $\mathbf{K}$ [m s^{-1}] is the saturated hydraulic conductivity tensor, $\mathbf{k}_r$ [–] represents the relative permeability, h [m] denotes the hydraulic head, Γ_{sub} [s^{-1}] is the fluid-exchange rate with the surface or the canopy layer and Q [s^{-1}] denotes the source/sink term from the outside of the domain.

To model the unsaturated flow field, constitutive relationships between hydraulic head, relative permeability and water saturation are needed, in addition to Eq. (1). For this purpose, we use the well-known van Genuchten parameterization (van Genuchten, 1980).

3.2. Surface flow

Surface flow is simulated using the diffusive-wave approximation of the two-dimensional Saint Venant equations (Moussa and Bocquillon, 2000). This approximation neglects local and convective acceleration terms of the momentum equation. It is generally applicable to catchments with a mild slope, thus lacking supercritical flow (Therrien et al., 2010). It can be written as followed:

$$\frac{\partial \phi_s h_s}{\partial t} - \frac{\partial}{\partial x} \left(k_{mx} d \frac{\partial h_s}{\partial x} \right) - \frac{\partial}{\partial y} \left(k_{my} d \frac{\partial h_s}{\partial y} \right) + d \Gamma_s \pm Q_s = 0 \quad (2)$$

where h_s [m] is the surface hydraulic heads, x and y [m] are the spatial coordinates, k_{mx} and k_{my} [m s^{-1}] are the surface conductances of the Manning equation in direction of x and y , d [m] is the water depth, Γ_s [s^{-1}] is the fluid exchange with the subsurface and the canopy layer and Q_s [m s^{-1}] are the external sources or sinks. ϕ_s [–] is the surface “porosity”, introduced by Panday and Huyakorn (2004), which accounts for soil depression and obstruction, for example due to vegetation.

3.3. Surface–subsurface coupling

To model exchange between the surface and the subsurface, a Darcy-like equation is used. This so-called dual-node approach models the exchange flux as if a thin layer of a porous material controlled the infiltration/exfiltration of water (Therrien et al., 2010):

$$\Gamma_s d = \frac{k_r K_{zz}}{l_{ex}} (h - h_s) \quad (3)$$

where Γ_s [s^{-1}] represents the water exchange flux, d [m] is the water depth in the surface domain, K_{zz} [$m\ s^{-1}$] denotes the saturated vertical conductivity of the porous medium close to the surface, $h - h_s$ [m] is the difference in hydraulic head between the surface and the subsurface nodes, l_{ex} [m] is the coupling length or the depth of the conceptualized layer, separating surface and subsurface, and k_r [–] denotes the relative permeability. When water infiltrates, k_r is the relative permeability of the porous media. If the water is flowing from the subsurface to the surface, k_r depends on the water depth and the obstruction length, i.e., the height of the vegetation or the obstacles which hinder the surface flow (Therrien et al., 2010).

3.4. Interception and evapotranspiration

HydroGeoSphere models interception, transpiration and evaporation separately, following the model of Kristensen and Jensen (1975) (Therrien et al., 2010). Interception is modeled as a reservoir whose size depends on the leaf area index (LAI). This reservoir is filled by rain events and depleted by evaporation. Transpiration is modeled as a function of soil moisture, potential evapotranspiration (PET), evapotranspiration from the canopy layer, root depth, and LAI. Actual evaporation is a function of soil moisture and PET minus transpiration and evaporation from the canopy layer. Transpiration and evaporation are assumed to be zero below a certain saturation limit chosen by calibration, and transpiration stops as well in nearly water-saturated soils. The spatial distribution of evapotranspiration depends on available moisture, and is a function of the root distribution function, which decreases with depth.

4. Study area

We tested the calibration approach at the Lerma catchment (centered at 42.06°N, 1.14°W), which is a sub-catchment of the

Arba catchment within the Ebro basin in North-East Spain. The surface catchment area is about 7.3 km² large and the altitude ranges between 330 and 490 m.a.s.l. Land use is predominantly agriculture (Pérez et al., 2011). While agriculture depended originally on natural rainfall, about half of the catchment area has been transformed into irrigated crop land starting in April 2006. The volume of irrigation water has changed from 0 m³/year in 2005 to $2.1 \cdot 10^6$ m³/year in 2011 (Merchán et al., 2013).

The climate in the Lerma catchment is semi-arid with a mean precipitation of 402 mm/year (2004–2011) and a mean potential evapotranspiration of 1301 mm/year (2004–2011, Merchán et al., 2013). Winter and summer are the driest seasons while spring is the wettest. The geology of the Lerma catchment is composed of two hydrogeologically relevant layers: The glacis layer on top and the so-called “buro” layer underneath. The glacis is made of clastic, permeable, and unconsolidated deposits of Quaternary age while the buro is a Tertiary bedrock made of lutite and marlstones. The glacis forms an aquifer whereas the buro can be considered as an aquitard even though the top first meters are likely weathered and play a role in the water circulation of the catchment. The Lerma soils are thin and composed of inceptisols (Pérez, 2011). No production wells are used in this catchment for groundwater abstraction.

The Lerma catchment has been extensively studied (Abraham et al., 2011; Merchán et al., 2013; Pérez et al., 2011; Skhiri and Dechimi, 2011; Urdanoz and Aragüés, 2011) and data coverage is comprehensive, especially concerning surface properties and agricultural management. Irrigation data, regularly checked for plausibility, are available (Pérez et al., 2011). Crop type and planting dates are also known. Since 2002, daily meteorological data (precipitation, radiation, relative humidity and wind) have been measured at the station Ejea de los Caballeros, about 5 km away from the center of the Lerma catchment. Topographic information is available in the form of a digital elevation model with a horizontal resolution of 5 m (I.G.N., 2012). Surface flow discharge has been

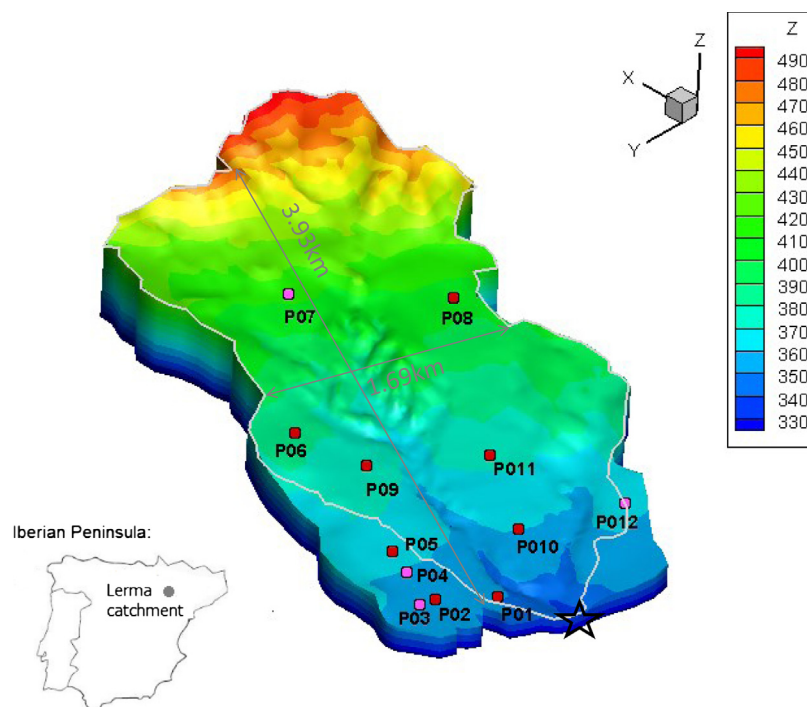


Fig. 2. Altitude of the Lerma catchment (m.a.s.l.) and position of the wells. The wells installed in 2008 are indicated by red dots, the ones installed in 2010 by purple dots, and the catchment outlet (42.06°N, 1.14°W) is indicated by a star. Vertical exaggeration is 5:1. The gray line represents the limits of the surface flow domain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measured at the catchment outlet with a temporal resolution of 15 min, starting in 2006. Since May 2008, groundwater levels in the glacia layer have been measured in eight observation wells, usually with a monthly frequency. In addition, water tables in four additional wells have been measured since March 2010 (Fig. 2). Geological and soil type data are scarce even though a soil characterization campaign was conducted (Pérez, 2011). In general, the material forming the glacia consists of approximately 60% sand, 20% silt, and 20% clay. However, as measurements were taken at only 10 locations, this campaign probably does not reflect the heterogeneity of the catchment. In addition, the depth of the buro layer has been measured during the drilling of 12 wells as well as estimated at 63 additional locations using electrical sounding (Plata-Torres, 2012). Soil depth is more uncertain but is estimated to range between 30 and 45 cm over the buro and 50 and 90 cm over the glacia (Beltrán, 1986).

5. Conceptual model

An initial modeling study of the Lerma catchment was conducted by Pérez et al. (2011). With new information becoming available, a more detailed model has been developed for the present study. The updated model differs from that presented by Pérez et al. (2011) in the following aspects: The depth of the buro is now better constrained as a result of a geoelectrical-sounding campaign (Plata-Torres, 2012). Consequently, the depth of the unconsolidated materials, forming the aquifer, are now shallower in the model. In addition, feedbacks between soil moisture and evapotranspiration were neglected by Pérez et al. (2011), but are accounted for in the present study. Finally, we use an updated, more accurate digital elevation model and discriminate between soil and aquifer materials. This set of changes caused the need of a new calibration.

Pérez et al. (2011) performed a manual calibration of his model. However, this approach was slow. In addition, only very few parameter sets could be tested and uniform parameter values were used for the creation of the initial conditions. Consequently, we decided to develop a new calibration method in addition to a new conceptual model.

We apply a distributed *pde*-based model to the Lerma catchment because of the onset of irrigation in the catchment between 2006 and 2009. This spatially-variable water source influences the other hydrological processes such as evapotranspiration and is difficult to represent in “bucket-type” models (Pérez et al., 2011). In addition, we will use this model in the future to study climate-change impact where distributed modeling of recharge and evapotranspiration is an important advantage. The possibility of studying the unsaturated zone is another benefit of choosing a *pde*-based model for this study.

In the present model, the subsurface of the Lerma model is separated into six zones as shown in Fig. 3. The lower zone represents the “buro” material. The intermediate layer is divided into two zones. The first one represents the part of the domain where the buro is close to the surface and more permeable than in the rest of the buro layer. The second zone represents the glacia, i.e., the unconsolidated materials forming the aquifer in the lower part of the domain. The top layer represents the soil (top panel of Fig. 3). Three zones are considered therein: one directly above the buro, one representing the agricultural soils, and one standing for the bare soil in the rest of the domain. The zones are assumed to be homogeneous and horizontally isotropic, while the vertical hydraulic conductivity is assumed to be one order of magnitude smaller than the horizontal one. The domain is vertically subdivided into 22 computational layers. The thickness of these layers decreases in the vicinity of the surface.

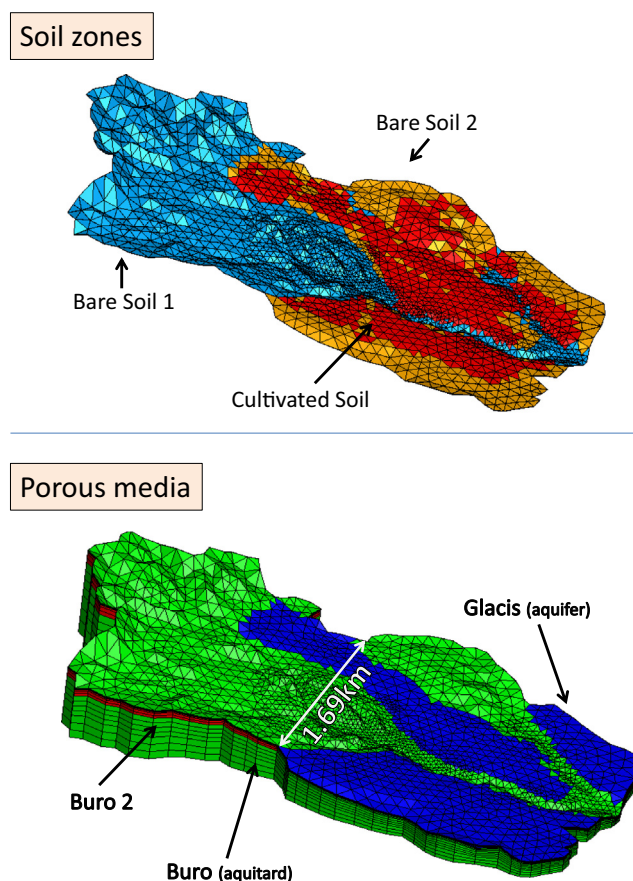


Fig. 3. Conceptual model of the catchment: soil and hydrogeological units. Extent of the cultivated soil is shown for the year 2008. Vertical exaggeration is 5:1.

The surface domain is separated into 56 sub-domains representing the different cropping and irrigation schemes. The known crop types consist mainly of corn, cereals, sunflower, and tomato. The actual evapotranspiration and the Manning coefficient depend on crop rotation patterns in the individual plots. Manning coefficients and other surface-flow parameters are taken from Pérez (2011). Actual evapotranspiration parameters are derived from a literature study, summarized in Section 7 and Table 5.

No-flow boundary conditions are assumed at the bottom and on the lateral sides of the subsurface model. For the surface domain, we assume a critical flow depth at the lateral limits (Pérez, 2011). The extent of the surface catchment (i.e. the watershed) is defined by the topography while the size of the aquifer is estimated based on the electrical sounding survey (Plata-Torres, 2012). The subsurface catchment is somewhat larger than the watershed constructed from topography. Groundwater flow and surface flow are nevertheless computed on the whole domain (Fig. 2). As a result of the boundary conditions, the fraction of the surface run-off that is generated

Table 1

Characteristics of the grids used. The simulation duration is based on an average of the four parameter sets used during the grid comparison, apart from the very fine grid case where only the second parameter set was used. The simulations were performed over the time period of one year on one core with a desktop computer Intel Core i7-2600 CPU @ 3.40 Ghz.

Grid name	No. of elements	No. of layers	Simulation time
Very fine grid	217,872	24	37 h 05 min
Fine grid	79,332	22	27 h 54 min
Intermediate grid	16,200	14	1 h 27 min
Coarse grid	10,448	14	40 min

outside of the watershed of the river gauge freely leaves the domain at the lateral surface boundaries.

To create the initial conditions, we started with a fully saturated model and repeatedly simulated one year with the meteorological forcing of 2005 until the hydrograph and the hydraulic heads achieved dynamic steady state (i.e., the temporal fluctuations were nearly identical from one simulation year to the next). Initial conditions were created separately for all grids, but for the fine grid the interpolated initial conditions of the intermediate grid were used as a starting point instead of a fully saturated model.

Irrigation and potential evapotranspiration are calculated using the FAO Penman–Monteith method (Allen et al., 1998) at daily time steps. Daily values of precipitation are used as model input as well, with the exception of days with intensive rainfall (higher than 25 mm/day). In this case, the total daily rain is considered to fall during 3 h in summer and spring. A duration of 9 h is assumed in winter and autumn. This approach aims to represent intensive convective events which are frequent in this catchment during spring and summer. The duration of smaller precipitation events has less impact on surface run-off and they have been kept as daily average to improve the model efficiency. The duration of the intense rain events is estimated using data with a resolution of 15 min from the meteorological station of Cola del Saso, which is situated about 10 km from the center of the catchment. Cola del Saso has a climate that is very similar to that of the Lerma catchment even though daily precipitation can show important short-term differences.

6. Comparison between the computational grids

As explained in Section 2, a coarse and an intermediate grid were used during calibration in addition to the fine grid, which

forms the final model grid. In addition, a fourth refined grid was created and compared to the fine grid to test whether the spatial resolution of the final model, which we aim to calibrate, was adequate.

The size of the elements in each grid has been chosen to minimize the trade-off between time gain and modeling accuracy. The model using the coarse grid aims to generally represent the hydrological processes, notably the mean hydraulic heads, while the model using the intermediate grid has the additional goal of representing the low flow at the catchment outlet. The spatial resolution of the fine grid should be sufficient to model more complex hydrological processes as peaks flows and evapotranspiration. More precisely, the number of elements is decreased by a factor of eight for the coarse grid, compared to the grid of the original model. The element size is increased by about 50% and the vertical discretization is reduced to 12 layers, instead of 22 in the fine grid. For the intermediate grid, we use the coarse grid with some refinements in the cells along the streams. More detailed characteristics of all grids and their mean simulation time are presented in Table 1 and Fig. 4.

In this section, we compare the outputs of the Lerma model for all grids. Our goal is to assess the similarities and differences among the grids to optimize their use during the calibration process. Discharge at the catchment outlet, hydraulic heads in the eight piezometers installed in 2008, soil saturation, and evapotranspiration at the surface are analyzed. To imitate the calibration process, we do not use the parameters or the initial conditions used in the final model. Three parameter sets, whose values are coherent with the soil properties of the catchment, and two artificial initial conditions are arbitrarily chosen and listed in Table 2. Meteorological and irrigation data from the year 2009 are used for this comparison.

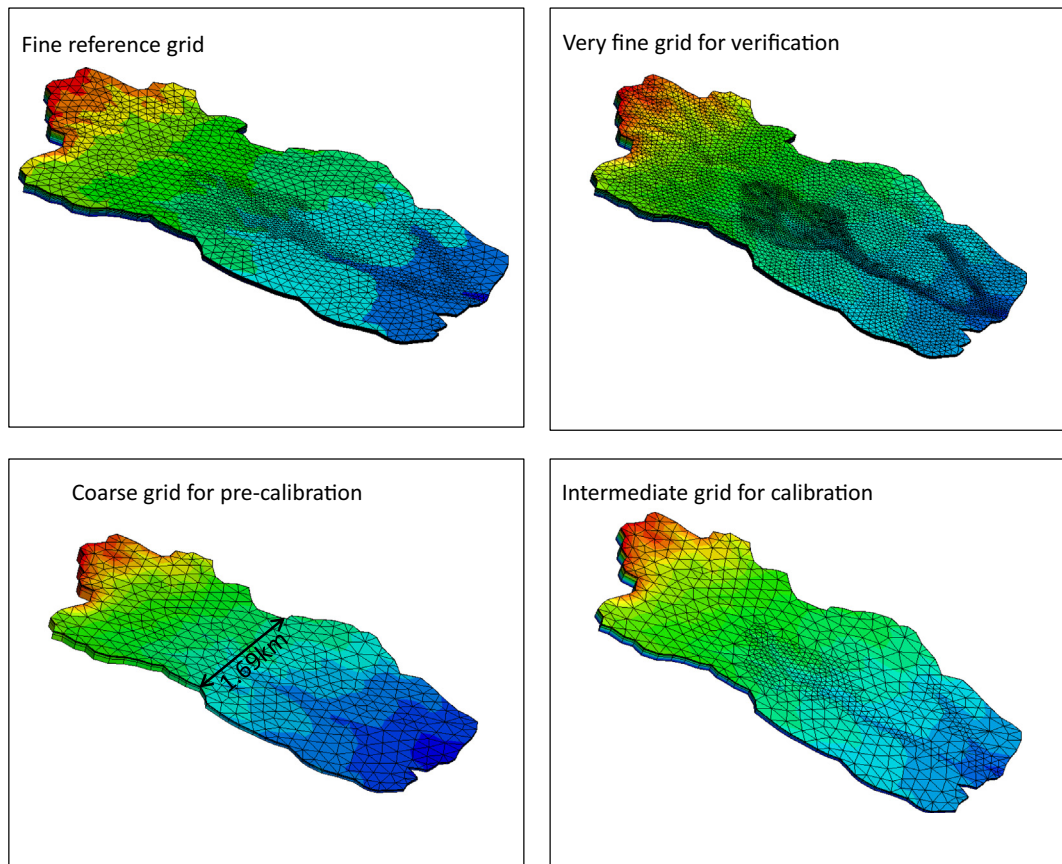


Fig. 4. Fine, very fine, coarse, and intermediate grids.

Table 2

Parameters set and initial conditions used during grid comparison. K is the hydraulic conductivity [m/s], α and β are the van-Genuchten parameters. All other parameters are kept constant at the value presented in Tables 4 and 5.

Parameter set	Param. 1	Param. 2	Param. 3	Param. 4
Bare soil 1				
K	1.5×10^{-5}	5×10^{-5}	1.5×10^{-5}	1.5×10^{-5}
α	3	5	3	3
β	2.25	2	1.5	2.25
Bare soil 2				
K	10^{-5}	10^{-4}	1.5×10^{-5}	10^{-5}
α	2	5	3	2
β	1.35	2	1.5	1.35
Cultivated soil				
K	2×10^{-5}	4×10^{-5}	1.5×10^{-5}	2×10^{-5}
α	2	5	3	2
β	1.25	2	1.5	1.25
Glacis				
K	0.0002	2×10^{-6}	1.5×10^{-5}	0.0002
α	2	5	3	2
β	1.5	2	1.5	1.5
Buro 2				
K	10^{-6}	10^{-5}	10^{-5}	10^{-6}
α	5	5	3	5
β	1.4	2	1.5	1.4
Buro				
K	5×10^{-7}	10^{-7}	10^{-7}	10^{-7}
α	3	5	3	3
β	1.8	2	1.5	1.8
Initial head	Surface	Surface	West: 380 m East: Surface	West: 380 m East: Surface

6.1. Discharge

Fig. 5 shows the hydrograph obtained with the fine, intermediate and coarse grids for parameter set 2. The results are very similar for the other parameters sets, as shown in Table 3. To quantitatively compare these hydrographs, two goodness-of-fit indicators are considered in Table 3 and during model calibration: The Nash-Stucliff efficiency (NSE) and the root mean square error (RMSE). NSE is defined in Eq. (4) and compares the relative magnitude of the squared residuals to the variance of the measured data (Nash and Sutcliffe, 1970). RMSE is given in Eq. (5) and is constructed as the square-root of the mean squared difference between measured and modeled values, assuming a perfectly unbiased model.

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{mes} - Q_{obs})^2}{\sum_{i=1}^N (Q_{mes} - \overline{Q_{mes}})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{obs} - Q_{mes})^2} \quad (5)$$

where Q_{obs} is the observed discharge, Q_{mes} the measured discharge, and N the number of observations.

To compare discharge between the different grids, we separate peak and low flow. In this paper, low flow is defined as the total flow on days without precipitation on that day and the day before. Because of the small size of the catchment and because of the reasonably efficient irrigation management (Abraham et al., 2011), significant run-off is not usually observed in periods without precipitation. Our definition of low flow is therefore similar to the base flow in dry days. More complex methods, which take advantage of the distributed nature of HydroGeoSphere (Partington et al., 2012), could be applied to separate base flow and run-off during wet days. However, we wanted to compare the full extent of peak flow and we therefore did not apply

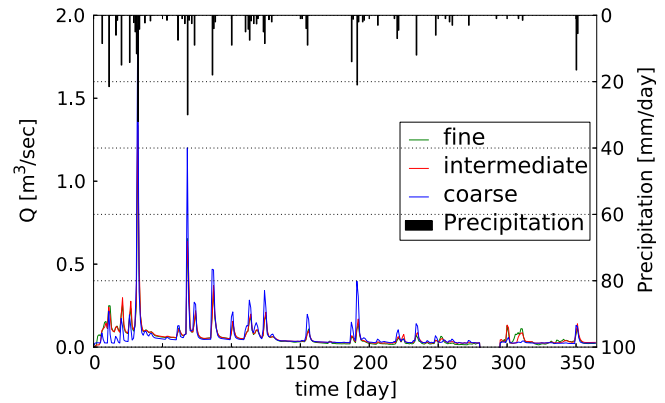


Fig. 5. Discharge comparison between the model results using the coarse, intermediate and fine grids for the parameter set 1. Meteorological input from 1st October 2008 to 30 September 2009.

Table 3

Goodness-of-fit measures of the hydrograph between the intermediate and fine grids.

Parameter set	NSE	RMSE [%]
Param. 1	0.98	1.43
Param. 2	0.93	2.6
Param. 3	0.96	1.9
Param. 4	0.98	1.5

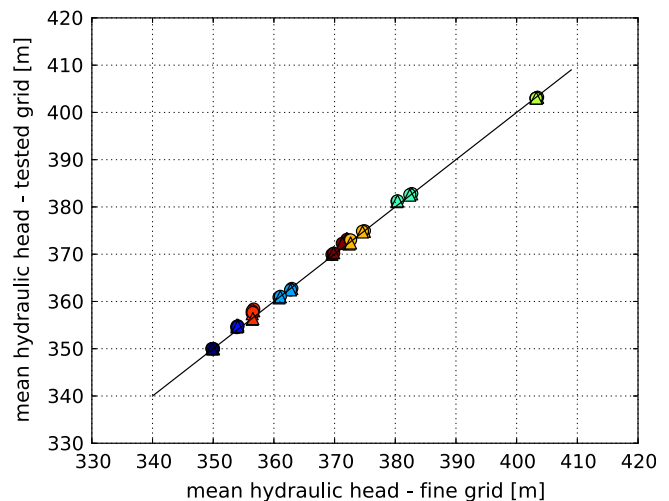


Fig. 6. Comparison of hydraulic heads computed on the various grids. The triangles show the hydraulic head comparison between the fine grid and the coarse grid, and the circles show the comparison between the fine grid and the intermediate grid for the four parameter sets. Average for one year. The wells are from the lowest to the highest hydraulic head: P01, P02, P010, P05, P011, P09, P06, P08.

base-flow separation method. Using the coarse grid results in less low flow compared to the fine grid (Fig. 5). We think this is caused by the coarse grid having a smaller surface slope than the fine grid because of the larger grid cells, which average elevation. This leads to a mismatch of run-off pathways (Chaplot, 2014) because flow is slower and more evenly distributed on the surface. However, we use this grid only to constrain the feasible parameter space. For this purpose, the difference between the model results is acceptably small. For the intermediate grid, the modeled discharge is very similar to that of the fine grid for all parameter sets (Table 3). As a consequence, the intermediate grid can be used for parameter calibration.

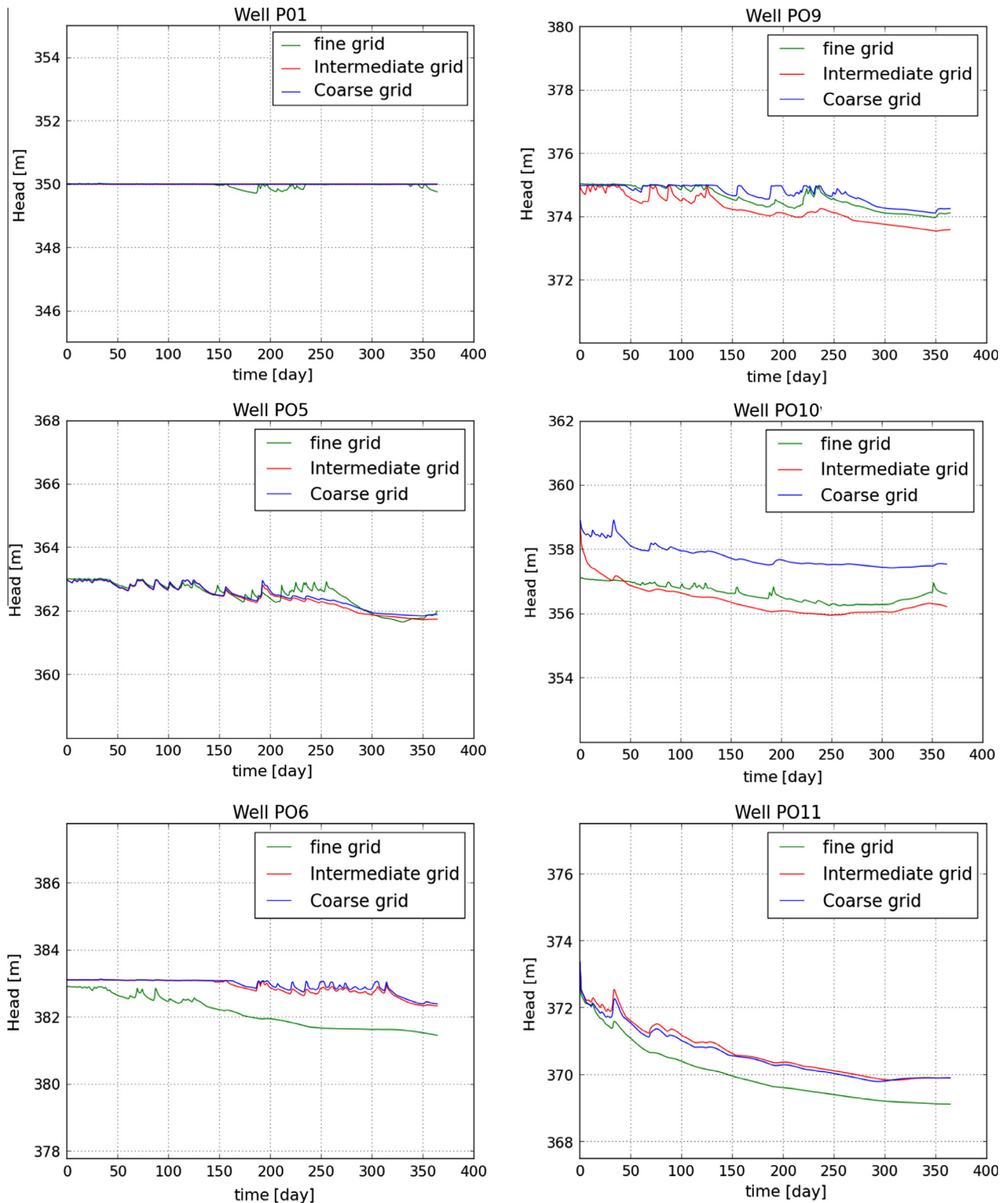


Fig. 7. Hydraulic-head variability for the hydrological year 2009 for the fine, intermediate, and coarse grids using parameter set 1.

Fig. 5 shows that the high-discharge peaks have a tendency to be lower on the fine grid than on the coarse one. A similar behavior has previously been observed (e.g., [Moglen and Hartman, 2001](#); [Wildemeersch et al., 2014](#)), even though the opposite behavior has been reported as well ([Chaplot, 2014](#)). In our model this response is not consistent either, as peaks simulated using the fine

grid are sometimes higher. Our model was tested with artificial precipitation input, and these simple tests show that peak height depends on the previous soil water content. A high water content before a precipitation event results in high peaks in the coarse or intermediate grid. On the contrary, a dry soil generally results in higher peaks on the fine grid. A possible explanation for this

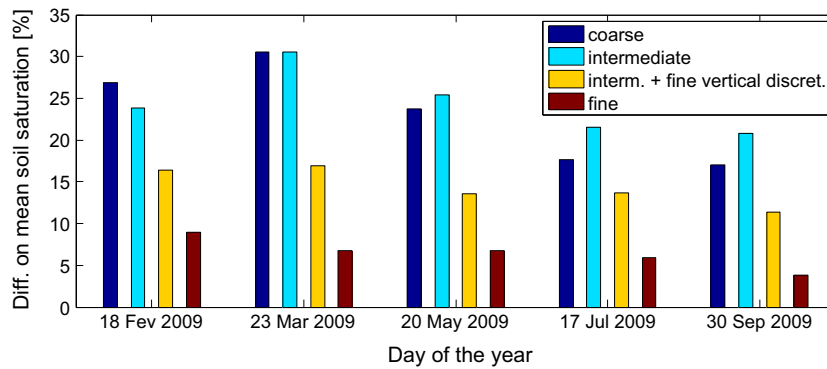


Fig. 8. Mean soil saturation (upper 40 cm) of each grid compared to the mean soil saturation of the very fine grid using parameter set 2.

difference is the soil saturation which is usually higher in the coarse and intermediate grids, as discussed in Section 6.3. Consequently, infiltration is higher in the coarse and intermediate grids when the soil is dry as the relative permeability is higher. On the contrary, when the soil is wet, the higher soil water content in the coarse and intermediate grids decreases infiltration. However, local topography and spatial difference in soil saturation, among other influences, complicates this simple picture.

6.2. Hydraulic heads

Fig. 6 presents the comparison between the fine, intermediate, and coarse grids with respect to hydraulic heads. In general, the simulation using the three different grids shows a close agreement for the average hydraulic head. Differences are mostly the result of the initial conditions as the chosen initial conditions depend on the surface elevation, which differs among the grids. Seasonal variability is similar between the grids as well (an example is shown in Fig. 7). However, during calibration on the intermediate and coarse grids, the average hydraulic heads were used as the target variables. Consequently, the variability of the hydraulic heads are of less importance for calibration.

6.3. Soil saturation and evapotranspiration

As mentioned in Section 5, the future purpose of the model is to study hydrological states and processes in a warmer climate, notably water saturation and evapotranspiration. These states and processes are sensitive to the vertical and horizontal spatial discretization of the model (Sciuto and Diekkrüger, 2010; Boone and Wetzel, 1996; Tiktak and Bouten, 1992). Therefore, it is necessary to ensure that the fine grid is accurate in simulating these states and processes, i.e., that this grid has the spatial precision needed to represent actual evapotranspiration and water saturation. To test this, a grid with even finer elements was developed and the results of the fine grid were compared to those of the very fine grid.

Because of the influence of evapotranspiration and the large variations in water content at the top of the unsaturated zone, the most important differences between the simulation results of the grids are found close to the surface. In addition, the strong relationship between (near-surface) soil-moisture and evapotranspiration has been found a key feature to represent water fluxes in coupled, integrated models, as large-scale land surface model (Gayler et al., 2014). Consequently, in Fig. 8, we compare the water saturation in the top soil (up to 40 cm depth) from each grid with the one obtained on the very fine grid at five specific dates. The soil water content of the soil was found to be constantly between 30% and 20% lower when using the very fine grid compared to the

intermediate grid. This bias is significant and explains why we could not use the intermediate grid in our final model. The mean water saturation on the fine grid is only between 9% and 3% higher than in the very fine grid, which is acceptable for our purpose. Indeed, this difference is significantly smaller than measured water saturation variability at small (10 m²) scale (e.g., Schmitz and Sourell, 2000).

To compare the impact of vertical and horizontal discretization, we tested the intermediate grid with the vertical discretization of the fine grid. Fig. 8 shows that the vertical discretization explains about half of the bias. The horizontal discretization is therefore important. In addition, comparisons between the intermediate and the fine grid after 6 years of simulation show that the difference in soil water content is approximately constant with time. Indeed, the higher soil water content results in most cases in a lower infiltration in the model using the intermediate grid compared to the one using the fine grid as mentioned in Section 6.1. However, the variability in the soil water content was not studied in detail to confirm this finding.

The bias in soil water content is probably related to evapotranspiration. Indeed, mean daily evapotranspiration is about 7% lower in the intermediate grid than in the fine and very fine grids. Similar observations have been made before in HydroGeoSphere, notably by Sciuto and Diekkrüger (2010) who explained that spatial discretization influences total evapotranspiration in the following way: Transpiration is a function of soil moisture. Under and above a certain saturation, no transpiration occurs. Transpiration increases linearly from the low saturation limit to the next threshold, where transpiration is at maximum. However, larger elements average soil moisture and therefore decrease gradients in soil saturation.

In other words, in a dry climate, θ_{min} , the lower saturation threshold where no transpiration occurs anymore, is reached more rapidly in a coarse grid than in a fine one. As an example, let's examine a zone with the soil saturation θ_a , where $\theta_a < \theta_{min}$. If this zone is represented by a single model cell, transpiration will be zero. By contrast, if this zone is separated into n model cells, it is possible to have $\theta_i > \theta_{min}$ while satisfying the condition of $\theta_a = \frac{1}{n} \sum_{i=1}^n \theta_i \frac{A_i}{A_{tot}}$, where θ_i is the soil saturation of cell i , A_i represents the area of this cell and A_{tot} is the area of the entire zone. As cells with $\theta_i > \theta_{min}$ exist, the overall transpiration will be larger than zero when using the finer discretization.

6.4. Conclusion of the grid comparison

In summary, the simulation results using the intermediate and coarse grids show important differences to those of the fine grid, notably when surface saturation or evapotranspiration are examined. Nevertheless, sufficient similarities in simulated hydraulic heads and flow discharge, especially low flow, allow the use of the

coarse and intermediate grids as a proxy model during the first steps of calibration. Similar results between the simulations using the very fine grid and the fine grid show that the spatial resolution of our final model is sufficient to capture the hydrological processes of interest.

7. Sensitivity analysis

A *pde*-based model uses a large number of parameters. Some of them have a large impact on the output and others only a very

small one. Therefore, to effectively calibrate a hydrological model, it is necessary to determine the parameters with the highest sensitivity to the model results (Christiaens and Feyen, 2002). This sensitivity analysis is done on the intermediate grid for the year 2008 because the intermediate grid is the one used most intensively during calibration. For each parameter, the model is run twice, once with a higher and once with a lower value than the original parameter value. The other parameters are kept constant. Description of the parameters, initial and modified values are shown in Tables 4 and 5. The choice of the parameters depends on the

Table 4
Parameters used during the sensitivity analysis, related to surface and subsurface flow.

Parameter name	Units	Value	Higher value	Lower value	Role
K_{bs1}	m/s	10^{-5}	10^{-4}	10^{-6}	Hydraulic conductivity – Bare Soil 1
K_{bs2}	m/s	10^{-5}	10^{-4}	10^{-6}	Hydraulic conductivity – Bare Soil 2
K_{cs}	m/s	10^{-4}	10^{-3}	10^{-5}	Hydraulic conductivity – Cultivated soil
K_{be}	m/s	10^{-7}	10^{-6}	10^{-8}	Hydraulic conductivity – Buro 2
K_{g1}	m/s	10^{-4}	10^{-3}	10^{-5}	Hydraulic conductivity – Glacis
K_b	m/s	10^{-7}	10^{-6}	10^{-8}	Hydraulic conductivity – Buro
α_{bs1}	1/m	4	5	3	Van Genuchten parameter – Bare Soil
α_{bs2}	1/m	4	5	3	Van Genuchten parameter – Bare Soil 2
α_{cs}	1/m	4	5	3	Van Genuchten parameter – Cultivated soil
α_{be}	1/m	4	5	3	Van Genuchten parameter – Buro 2
α_{g1}	1/m	4	5	3	Van Genuchten parameter – Glacis
α_b	1/m	4	5	3	Van Genuchten parameter – Buro
β_{bs1}	–	1.5	2	1.2	Van Genuchten parameter – Bare Soil 1
β_{bs2}	–	1.5	2	1.2	Van Genuchten parameter – Bare Soil 2
β_{cs}	–	1.5	2	1.2	Van Genuchten parameter – Cultivated soil
β_{be}	–	1.5	2	1.2	Van Genuchten parameter – Buro 2
β_{g1}	–	1.5	2	1.2	Van Genuchten parameter – Glacis
β_b	–	1.5	2	1.2	Van Genuchten parameter – Buro
S_{st}	1/m	10^{-4}	10^{-3}	10^{-5}	Specific storage for all soil zones
θ_s	–	0.3	0.4	0.2	Porosity for all soil zones
S_r	–	0.1	0.15	0.05	Residual Saturation for all soil zones
k_{bmx} , k_{bmy}	m/s	0.01	0.015	0.005	Manning parameter for bare soil (x and y direction)
k_{cmx} , k_{cmy}	m/s	0.04	0.05	0.03	Manning parameter for cultivated soil (x and y direction)
r_s	m	0.001	0.0015	0.0005	Average height of the soil depressions
o_s	m	0.05	0.07	0.03	Average height of the vegetation or other obstacles
l_{ex}	m	0.001	0.01	0.0001	Coupling between surface and subsurface

Table 5
Parameters used during the sensitivity analysis, related to evapotranspiration.

Param. name	Units	Value	Higher value	Lower value	Role	Reference
W_p	m	–150	–100	–200	Wilting point for all crops	Tolk (2003)
θ_a	–	0.9	0.95	0.85	Highest saturation where transpiration happens	Panday and Huyakorn (2004)
θ_o	–	0.8	0.87	0.73	Highest saturation where transpiration is not limited	Panday and Huyakorn (2004)
R_c	m	1	0.5	1.5	Root depth for corn	Canadell et al. (1996), Breuer et al. (2003)
R_w	m	1	0.5	1.5	Root depth for winter cereal	Breuer et al. (2003)
R_t	m	0.2	0.1	0.3	Root depth for tomato	Rosário et al. (1996)
C_1	–	0.31	0.5	0.1	1st Transpiration fit. param.	Li et al. (2008)
C_2	–	0.2	0.4	0.1	2nd Transpiration fit. param.	Li et al. (2008)
C_3	–	3.7	6.4	1	3rd Transpiration fit. param.	Li et al. (2008), Therrien et al. (2010)
E_d	m	0.2	0.3	0.1	Max. evaporation depth	Therrien et al. (2010)
e_1	–	0.83	0.73	0.93	Lowest saturation where evaporation happens	Panday and Huyakorn (2004)
LAI_c	–	Variable	6.5	4	LAI for corn	Gardioli et al. (2003), Mailhol et al. (1997), Howell et al. (1996), Breuer et al. (2003)
LAI_w	–	Variable	3	1.5	LAI for winter cereal	Breuer et al. (2003)
LAI_t	–	Variable	3.5	2	LAI for tomato	Koning (1994), Heuvelink (1995)
LAI_b	–	Variable	0.5	0.05	LAI for “bare soil” (i.e. for shrub and grass)	–
C_s	m	0.0035	0.0025	0.0045	Maximum height of the canopy storage	Kozak et al. (2007)

estimated limits of the parameter space, which are based on a literature study and previous work from Pérez (2011). The initial value is the mean of the limits of the respective parameter space. The modified values are half of the difference between the mean and the higher or lowest limits of the parameter space. The initial conditions are taken from the calibrated model.

Previously, the same sensitivity analysis has been done with artificial initial conditions, yielding similar results. In addition, we ran the same sensitivity analysis on the fine grid to confirm that parameter sensitivity is not greatly influenced by the spatial discretization (Wildemeersch et al., 2014). Results were again similar.

In this sensitivity analysis, the flow hydrograph at the outlet and the hydraulic heads measured at five observation wells (P02, P06, P09, P010 and P011) were analyzed. Based on the different simulations of these variables, the impact of each parameters on the model output was quantified as follows:

$$S_i = \frac{I_i}{\Delta b_i} \sum_{j=1}^{N_o} w_j \frac{RMSE(y_{ij}, y_{0j})}{\Delta_{max,j}} \quad (6)$$

where S_i is the sensitivity index of the parameter i , I_i is the estimated size of the parameter space for this parameter and Δb_i is the difference between the original and modified parameter value. By design, $\frac{I_i}{\Delta b_i} = 4$ in this analysis. N_o represents the type of outputs, i.e., the five observations wells and the hydrograph. y_{0j} is the j th output of the unmodified model while y_{ij} is the j th output of the model with the modified parameter b_i . The RMSE is calculated by Eq. (5). The weights w_j are 0.5 for the hydrograph and 0.1 for each well. $\Delta_{max,j}$ is the maximum of the absolute difference between all modified models and the unmodified model for the output j . A high S_i indicates a parameter with a high influence on the model output.

Fig. 9 shows the S_i values for all parameters with a $S_i > 7\%$. These are 12 parameters out of 44 parameters. The saturated hydraulic conductivities have the largest impact on the simulated hydraulic heads and on the hydrograph, as found in other sensitivity analyses related to surface–subsurface modeling (Bonton et al., 2012). The van-Genuchten parameters of the zones close to the surface and the porosity ranked second in sensitivity. There are numerous parameters related to evapotranspiration but they have a small impact on the overall simulation results for hydraulics heads and hydrographs. Annual mean PET in the Lerma catchment is about three or four times the actual evapotranspiration. Consequently, soil water content controls actual evapotranspiration, a fact which limits the importance of the parameters related to evapotranspiration. The first and third fitting parameter of the

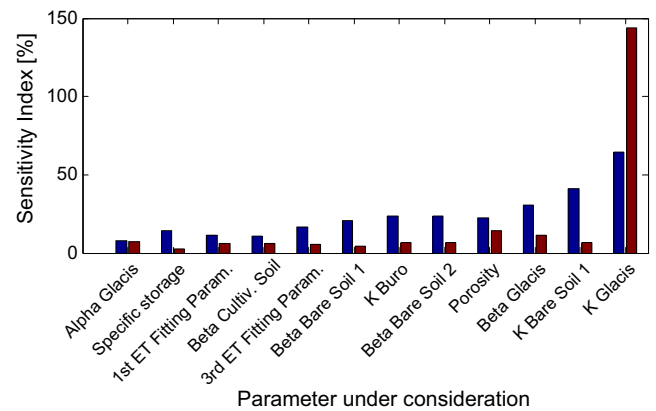


Fig. 9. Sensitivity indices of all parameters in the sensitivity analysis (Tables 4 and 5) with a mean sensitivity index above 7%. For each parameter, we tested an higher and lower value than the base case value. The higher value test is indicated in blue and the lower one in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transpiration function C_1 and C_3 are the most sensitive parameters of all evapotranspiration parameters, when hydraulic heads or the flow hydrograph is considered.

The S_i value of LAI is difficult to compare to the other parameters because it varies with time but, based on tests with different constant LAI, S_i of LAI seems to be in the lower range ($<7\%$).

8. Calibration

Our model was calibrated using data of the years 2006–2009 and validated using the data of the years 2010–2011. Based on the results of the sensitivity analysis, the hydraulic conductivity for each zone, the van-Genuchten parameters and the porosity were the only parameters modified during calibration. The calibration was performed in five steps:

- First, we tested 200 randomly-chosen parameters sets using the coarse grid and artificial initial conditions to explore the most sensitive parameters space. In this first calibration step, we chose the best parameter set by comparing the modeled and measured mean hydraulic heads and the low flow of the hydrograph. Low flow was simply defined as the total flow on days without precipitation at that day and the day before, as explained in Section 6.1.

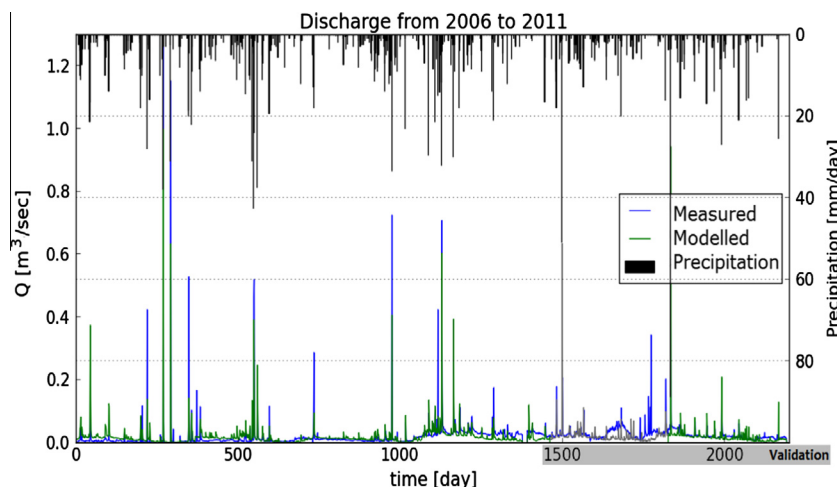


Fig. 10. Measured and modeled hydrograph for 2006–2011. Please notice the missing data at the end of 2009. Because of snow and breakage of the irrigation pipe, 2010 is not taken into account in our analysis.

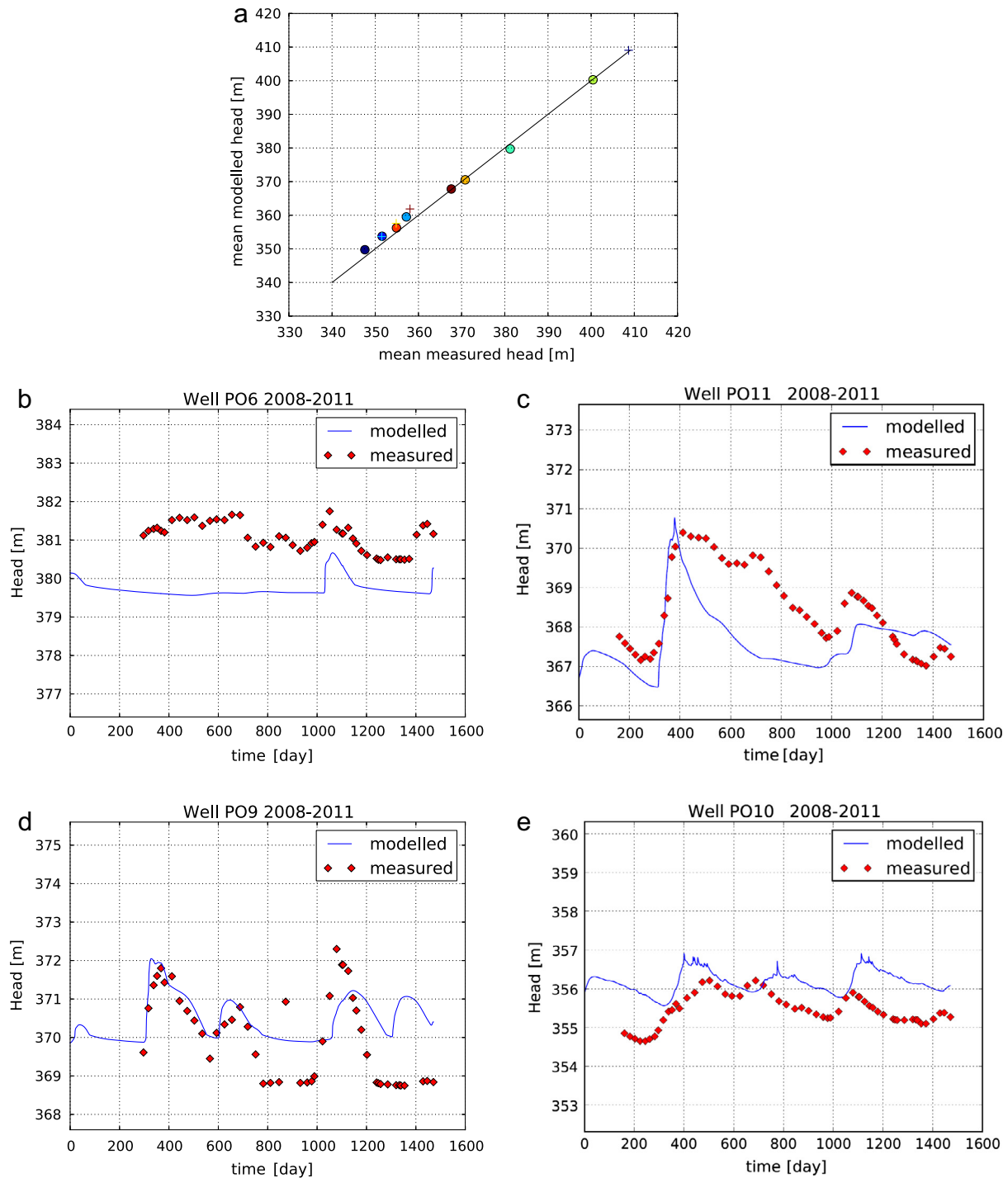


Fig. 11. Average modeled and measured value for all wells and transient hydraulic head for the wells P06, P09, P10 and P11. On the figure at the top, the wells are from the lowest to the highest hydraulic head: P01, P02, P03, P010, P04, P05, P012, P011, P09, P06, P08, P07. The mean modeled heads are the averages of the daily values from October 1, 2007 to September 30, 2011. The mean measured heads are the averages of the measured monthly values for the years 2008–2011.

- Using the selected best parameter set, we updated the initial condition by running the intermediate grid 100 times with the meteorological input of the year 2005 to reach dynamical steady state.
- Then, we manually calibrated the intermediate grid by trial and error, using the daily hydraulic heads and the full hydrograph. We tested about 70 parameter sets and we selected the parameter set which resulted in the highest NSE value and lowest

Table 6

Goodness-of-fit measure for calibration and validation periods.

	2006–2009	2011
NSE	0.74	0.92
RMSE (%)	3.16	1.36

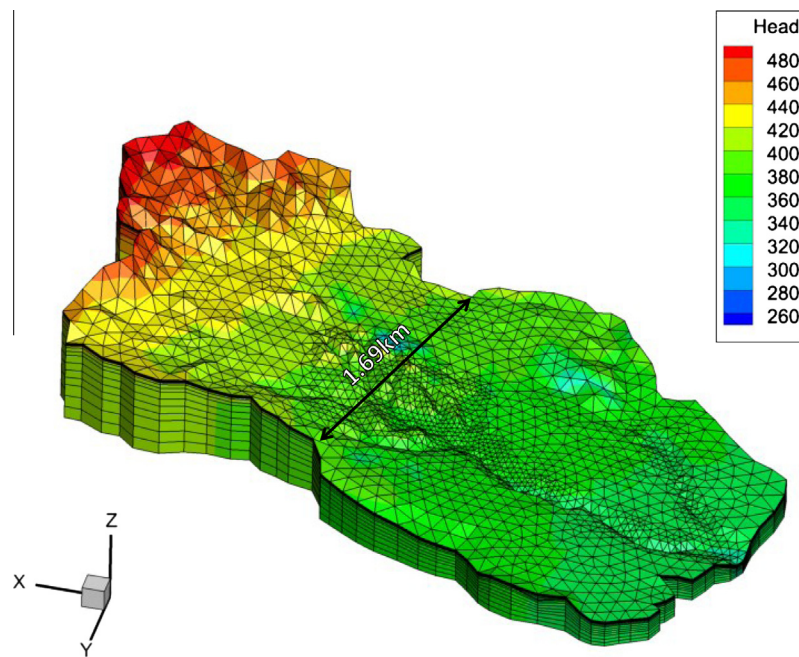


Fig. 12. Calibrated hydraulic head in m.a.s.l. for the Lerma catchment on the 4th November 2009. Vertical exaggeration is 5:1.

RMSE for the outflow. We estimated the quality of the modeled hydraulic heads by visually comparing this output with the measurements.

- Subsequently, the initial conditions were updated on the intermediate grid with the best parameter set found. The initial conditions were then transferred to the fine grid, using a linear interpolation scheme. The final model with the fine grid was then run three times with the meteorological input from 2005 and the parameters from the intermediate grid to finalize the initial conditions.
- Afterwards, we finished the calibration on the fine grid by manually adjusting the surface hydraulic conductivity to better match the peak flows at the outlet.

Fig. 10 shows the calibrated hydrograph and Fig. 11 the calibrated hydraulic heads. Table 6 lists the values of NSE and RMSE. Hydraulic-head values for the whole catchment are presented in Fig. 12 for the 4th November 2009. The calibrated parameters are listed in Table 7. All other parameters are identical to those given in Tables 4 and 5.

Based on the high NSE value of 0.74 and the low RMSE of 3.16%, we assess that our model can reproduce the flow hydrograph and the hydraulic heads well. Visual examination also confirms that the hydrograph is reproduced well even if the model has a tendency to underestimate discharge during the irrigation period (July–September), notably in 2009. Modeled mean hydraulic heads are close to the measured ones. Indeed, there is less than 2 m of difference between calibrated and measured mean hydraulic heads. This is on the same order of magnitude than the annual variability in most wells. In addition, the model simulations capture most of the variability of the hydraulic head measurements, even though a tendency to underestimate it was observed, notable for well P09 (Fig. 11). This might be explained by the position of well P09, situated close to a road whose filling material might disturb groundwater flow. Overall, we judged that the calibration was successful.

However, in 2010, the modeled hydrograph differs significantly from the measured one. An explanation may be that 2010 is one of the rare years in which snow was present in the Lerma catchment. Unfortunately, our model does not account for snow. In addition,

Table 7

Calibrated parameters for hydraulic conductivity, porosity and van-Genuchten parameters. See Table 4 for the definition of the parameters.

Parameter	Units	Calibrated Value
K_{bs1}	m/s	3×10^{-6}
K_{bs2}	m/s	10^{-6}
K_{cs}	m/s	8×10^{-5}
K_{be}	m/s	3×10^{-6}
K_{g1}	m/s	1.7×10^{-4}
K_b	m/s	10^{-7}
α_{bs1}	1/m	4.95
α_{bs2}	1/m	4
α_{cs}	1/m	4
α_{g1}	1/m	4
β_{bs1}	–	1.35
β_{bs2}	–	1.4
β_{cs}	–	1.35
β_{g1}	–	2
$\theta_{s,bs1}$	–	0.25
$\theta_{s,bs2}$	–	0.25
$\theta_{s,cs}$	–	0.25
$\theta_{s,g1}$	–	0.2

the main irrigation pipe broke in spring. This event caused a flood in this relatively small catchment, which is not captured by the model simulations. The year 2010 was therefore omitted from our analysis. However, the model was able to reproduce the following year adequately, with a NSE value 0.92 and a RSME of 1.36%. In addition, the model satisfactorily reproduces hydraulic heads during the whole validation period in all observation wells, including the wells drilled in 2010 which were not included in the model calibration.

During calibration, working with different grids was very useful. To quantify this advantage, we may examine the number of years that the model needs to run during calibration. We tested 200 parameter sets during two years for our first estimation. Then, we simulated 2×100 years for the initial condition and approximately 70 parameter sets over four years for manual calibration. To finish the calibration, three parameter sets over four years were tested on the fine grid, after a final update to the initial conditions.

Consequently, the equivalent of approximately 890 years were computed during model calibration. With the fine grid, this would have taken us more than two years of computational time. The creation of the initial conditions alone would have lasted more than 3 months. With our choice of grids, the simulation time was greatly reduced. We needed about ten days on one computational core to process the 200 first simulations, about six days to create the initial conditions and between five and eight hours to manually test one parameter set. Consequently, the calibration would last approximately a month if no adjustments to the model were needed.

9. Conclusions

To simulate hydrological processes at catchment scale, coupled *pde*-based models have the advantage that they simulate spatially distributed, measurable state variables such as hydraulic heads, while accounting for explicit coupling between surface and subsurface processes, thus allowing to predict changes of fluxes at compartmental interfaces. However, calibration of large distributed models is time-consuming and difficult, which limits their usefulness.

The method presented here is a significant step toward solving this problem. Using grid coarsening is a practical solution which reduces the time needed to test simulations with various parameter sets. Hence, it allows the modeler to better explore the parameter space in a shorter amount of time. We tested this approach under realistic conditions, using a catchment-scale case study, and we showed that this method has decreased calibration time. We suggest that our method can successfully be applied to other catchments as well. However, guidelines for the choice of the grid precision should be investigated further, as it is now principally based on empirical knowledge.

In addition, the proposed method could be easily modified to perform an automatic calibration. In addition, it can be extended to allow for simple uncertainty analysis similar to the one presented in Wildemeersch et al. (2014). Reactive transport modeling could also profit from a coarser grid for calibration of flow, as a finer spatial discretization is usually needed to resolve transport than flow (Chaubey et al., 2005).

Acknowledgments

The authors wish to thank Julia Knapp for her help during the calibration and the sensitivity analysis as well as David Rudolph from the University of Waterloo for his continuous support during this research. We are thankful to Juan Plata, Felix Rubio and Antonio Azcón, from the Geological Survey of Spain, for making the geophysical data freely available and providing information about the aquifer. The support of the Spanish meteorological national agency (AEMET) is here gratefully acknowledged too. This study was performed within the International Research Training Group “Integrated Hydrosystem Modelling” under the Grant GRK 1829/1 of the DFG (Deutsche Forschungsgemeinschaft). Research in the Lerma catchment is supported by the Grant CGL-2012-32395 (Spanish Ministry of Economy and Competitiveness and European Union, FEDER funds). Daniel Merchán is sponsored by the BES2010-034124 grant of the Spanish Ministry of Economy and Competitiveness.

References

Abraham, R., Causapé, J., García-Garzábal, I., Merchán, D., 2011. Implementing irrigation: water balances and irrigation quality in the Lerma basin (Spain). *Agric. Water Manage.* 102, 97–104.

Allen, R., Pereira, L., Raes, D., Smith, M., 1998. Crop evapotranspiration (guidelines for computing crop water requirements). FAO Irrigation and Drainage Paper 56.

Beltrán, A., 1986. Estudio de los suelos de la zona regable de Bardenas II. Sectores VIII, IX, X, XII y XIII. Instituto Nacional de Reforma y Desarrollo Agrario, Ministerio de Agricultura, Pesca y Alimentación.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* 6, 279–298.

Blasone, R.S., Madsen, H., Rosbjerg, D., 2008. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *J. Hydrol.* 353, 18–32.

Bonton, A., Bouchard, C., Rouleau, A., Rodriguez, M.J., Therrien, R., 2012. Calibration and validation of an integrated nitrate transport model within a well capture zone. *J. Contam. Hydrol.* 128, 1–18.

Boone, A., Wetzzel, P.J., 1996. Issues related to low resolution modeling of soil moisture: experience with the PLACE model. *Global Planet. Change* 13, 161–181.

Breuer, L., Eckhardt, L., Frede, H., 2003. Plant parameter values for models in temperate climates. *Ecol. Modell.* 169, 237–293.

Bruneau, P., Gascuel-Oudoux, C., Robin, P., Merot, P., Beven, K., 1995. Sensitivity to space and time resolution of a hydrological model using digital elevation data. *Hydrol. Process.* 9, 69–81.

Calderhead, A., Therrien, R., Rivera, A., Martel, R., Garfias, J., 2011. Simulating pumping-induced regional land subsidence with the use of insar and field data in the Toluca valley, Mexico. *Adv. Water Resour.* 34, 83–97.

Canadell, J., Jackson, R., Ehleringer, J., Mooney, H., Sala, O., Schulze, E., 1996. Maximum root depth of vegetation types at global scale. *Oecologia* 108, 583–595.

Carrera-Hernandez, J., Smerdon, B., Mendoza, C., 2012. Estimating groundwater recharge through unsaturated flow modelling: sensitivity to boundary conditions and vertical discretization. *J. Hydrol.* 452, 90–101.

Chaplot, V., 2005. Impact of DEM mesh size and soil map scale on SWAT runoff, sediment, and NO₃-N loads predictions. *J. Hydrol.* 312, 207–222.

Chaplot, V., 2014. Impact of spatial input data resolution on hydrological and erosion modeling: recommendations from a global assessment. *J. Phys. Chem. Earth* 69, 23–35.

Chaubey, I., Cotter, A.S., Costello, T.A., Soerens, T.S., 2005. Effect of DEM data resolution on SWAT output uncertainty. *Hydrol. Process.* 19, 621–628.

Christiaens, K., Feyen, J., 2002. Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE-SHE model. *Water Resour. Res.* 38, 1169–1185.

Condon, L.E., Maxwell, R.M., Gangopadhyay, S., 2013. The impact of subsurface conceptualization on land energy fluxes. *Adv. Water Resour.* 60, 188–203.

Cotter, A.S., Chaubey, I., Costello, T.A., Soerens, T.S., Nelson, M.A., 2003. Water quality model output uncertainty as affected by spatial resolution of input data. *JAWRA – J. Am. Water Resour. Assoc.* 39, 977–986.

Dutta, D., Nakayama, K., 2009. Effects of spatial grid resolution on river flow and surface inundation simulation by physically based distributed modelling approach. *Hydrol. Process.* 23, 534–545.

Gardioli, J., Serio, L., Maggiora, A.D., 2003. Modelling evapotranspiration of corn (zea mays) under different plant densities. *J. Hydrol.* 271, 188–196.

Gayler, S., Wöhling, T., Grzeschik, M., Ingwersen, J., Wizemann, H.D., Warrach-Sagi, K., Högy, P., Attinger, S., Streck, T., Wulfmeyer, V., 2014. Incorporating dynamic root growth enhances the performance of Noah-MP at two contrasting winter wheat field sites. *Water Resour. Res.* 50, 1337–1356.

van Genuchten, M., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Am. Soc. Soil Sci.* 44, 892–898.

Goderniaux, P., Brouyère, S., Blenkinsop, S., Burton, A., Fowler, H.J., Orban, P., Dassargues, A., 2011. Modeling climate change impacts on groundwater resources using transient stochastic climatic scenarios. *Water Resour. Res.* 47.

Goderniaux, P., Brouyère, S., Fowler, H., Blenkinsop, S., Therrien, R., Orban, P., Dassargues, A., 2009. Large scale-surface subsurface hydrological model to assess climate change impacts on groundwater reserves. *J. Hydrol.* 373, 122–138.

Hessel, R., 2005. Effects of grid cell size and time step length on simulation results of the Limburg soil erosion model (LISEM). *Hydrol. Process.* 19, 3037–3049.

Heuvelink, E., 1995. Growth, development and yield of a tomato crop: periodic destructive measurements in a greenhouse. *Sci. Horticul.* 61, 77–99.

Howell, T., Evett, S., Tolk, J., Schneider, A., Steiner, J., 1996. Evapotranspiration of corn – southern high plains. In: *Evapotranspiration and Irrigation Scheduling*, Proceedings of the International Conference, American Society of Agricultural Engineers.

I.G.N., 2012. Modelo Digital del Terreno, hoja 282 del Mapa Topográfico Nacional (in Spanish). Instituto Geográfico Nacional.

Koning, A., 1994. Development and dry matter distribution in glasshouse tomato: a quantitative approach. Ph.D. thesis, Agricultural University of Wageningen.

Kozak, J., Ahuja, L.R., Green, T., Ma, L., 2007. Modelling crop canopy and residue rainfall interception effects on soil hydrological components for semi-arid agriculture. *Hydrol. Process.* 21, 229–241.

Kristensen, K., Jensen, S., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. *Nordic Hydrol.* 6, 170–188.

Kuo, W.L., Steenhuis, T.S., McCulloch, C.E., Mohler, C.L., Weinstein, D.A., DeGloria, S.D., Swaney, D.P., 1999. Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model. *Water Resour. Res.* 35, 3419–3428.

Lemieux, J.M., Sudicky, E.A., Peltier, W.R., Tarasov, L., 2008. Simulating the impact of glaciations on continental groundwater flow systems: 2. Model application to

- the Wisconsinian glaciation over the Canadian landscape. *J. Geophys. Res.* 113, F03018.
- Li, Q., Unger, A., Sudicky, E., Kassenaar, D., Wexler, E., Shikaze, S., 2008. Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. *J. Hydrol.* 357, 317–336.
- Mailhol, J.C., Olufayo, A., Ruelle, P., 1997. Sorghum and sunflower evapotranspiration and yield from simulated leaf area index. *Agric. Water Manage.* 35, 167–182.
- McMichael, C., Hope, A., Loaiciga, H., 2006. Distributed hydrological modelling in California semi-arid shrublands: MIKE-SHE model calibration and uncertainty estimation. *J. Hydrol.* 317, 307–324.
- Mehl, S., Hill, M.C., 2002. Development and evaluation of a local grid refinement method for block-centered finite-difference groundwater models using shared nodes. *Adv. Water Resour.* 25, 497–511.
- Mehl, S., Hill, M.C., Leake, S.A., 2006. Comparison of local grid refinement methods for MODFLOW. *Groundwater* 44, 792–796.
- Merchán, D., Causapé, J., Abrahao, R., 2013. Impact of irrigation implementation on hydrology and water quality in a small agricultural basin in Spain. *Hydrol. Sci. J.* 58, 1400–1413.
- Moglen, G., Hartman, G., 2001. Resolution effects on hydrologic modeling parameters and peak discharge. *J. Hydrol. Eng.* 6, 490–497.
- Molnár, D., Julien, P., 2000. Grid-size effects on surface runoff modeling. *J. Hydrol. Eng.* 5, 8–16.
- Moussa, R., Bocquillon, C., 2000. Approximation zones of the saint-venant equations for flood routing with overbank flow. *Hydrol. Earth Syst. Sci.* 4, 251–261.
- Muleta, M., Nicklow, J., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* 306, 127–145.
- Nash, J., Sutcliffe, V., 1970. River flow forecasting through conceptual models, part I – A discussion of principles. *J. Hydrol.* 10, 282–290.
- Panday, S., Huyakorn, P., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Adv. Water Resour.* 27, 361–382.
- Partington, D., Brunner, P., Simmons, C., Werner, A., Therrien, R., Maier, H., Dandy, G., 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. *J. Hydrol.* 459, 28–39.
- Pérez, A., 2011. Physics-based numerical modeling of surface-groundwater flow and transport at catchment scale. Ph.D. thesis, Universität Tübingen.
- Pérez, A., Abrahao, R., Causapé, J., Cirpka, O., Bürger, C., 2011. Simulating the transition of a semi-arid rainfed catchment towards irrigation agriculture. *J. Hydrol.* 409, 663–681.
- Plata-Torres, J., 2012. Informe sobre la campaña de sondeos eléctrico verticales efectuados en el barranco de Lerma (Zaragoza). Grupo de Geofísica del Instituto Geológico y Minero de España.
- Rosário, M.D., Calado, A.M., Oliveira, G., Portas, C., 1996. Tomato root distribution under drip irrigation. *J. Am. Hort. Soc.* 121, 644–648.
- Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* 46.
- Schmitz, M., Sourell, H., 2000. Variability in soil moisture measurement. *Irrigat. Sci.* 19, 147–151.
- Sciuto, G., Diekkrüger, B., 2010. Influence of soil heterogeneity and spatial discretization on catchment water balance modeling. *Vadose Zone J.* 9, 955–969.
- Shao, H., Kosakowski, G., Berner, U., Kulik, D.A., Mäder, U., Kolditz, O., 2013. Reactive transport modeling of the clogging process at Maqarin natural analogue site. *J. Phys. Chem. Earth* 64, 21–31.
- Skhiri, A., Dechimi, F., 2011. Irrigation return flows and phosphorus transport in the Middle Ebro River valley (Spain). *Spanish J. Agric. Res.* 9, 938–949.
- Therrien, R., 2006. HydroGeoSphere – A Three-Dimensional Numerical Model Describing Fully-Integrated Subsurface and Surface Flow and Solute Transport. Ph.D. thesis, Université Laval and University of Waterloo.
- Therrien, R., McLaren, R., Sudicky, E., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport – User Manual, University of Waterloo.
- Tiktak, A., Bouten, W., 1992. Modelling soil water dynamics in a forested ecosystem. III: Model description and evaluation of discretization. *Hydrol. Process.* 6, 455–465.
- Tolk, J., 2003. Soils, permanent wilting points. *Encycl. Water Sci.*, 927–929.
- Urdanoz, V., Aragüés, R., 2011. Pre- and post-irrigation mapping of soil salinity with electromagnetic induction techniques and relationships with drainage water salinity. *Am. Soc. Soil Sci.* 75, 207–215.
- Vazquez, R.F., Feyen, L., Feyen, J., Refsgaard, J.C., 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrol. Process.* 16, 355–372.
- Wildemeersch, S., Goderniaux, P., Orban, P., Brouyère, S., Dassargues, A., 2014. Assessing the effects of spatial discretization on large-scale flow model performance and prediction uncertainty. *J. Hydrol.* 510, 10–25.
- Xevi, E., Christiaens, K., Espino, A., Sewnandan, W., Mallants, D., Sorensen, H., Feyen, J., 1997. Calibration, validation and sensitivity analysis of the MIKE-SHE model using the Neuenkirchen catchment as case study. *Water Resour. Manage.* 11, 219–242.