Environmental Modelling & Software 66 (2015) 78–86

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Tutorials as a flexible alternative to GUIs: An example for advanced model calibration using Pilot Points

Christian Moeck*, Daniel Hunkeler, Philip Brunner

Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, 2000 Neuchâtel, Switzerland

ARTICLE INFO

Article history:
Received 17 November 2014
Received in revised form
18 December 2014
Accepted 21 December 2014
Available online 21 January 2015

Keywords:
GUI
Tutorial
Model calibration
Pilot Points
HydroGeoSphere

ABSTRACT

Environmental modelling software with graphical user interfaces (GUIs) are user friendly and help to focus on the aspects of modelling rather than on the technicality of the underlying code. This is a fundamental advantage compared to codes without a GUI, as the absence of a GUI can prohibit the widespread application of a software tool. However, there is a downside to GUIs, too. They commonly lag behind the newest software development. We argue that tutorials are a flexible but undervalued alternative to GUIs, and are convinced that tutorials can help to make complex software accessible to an increased number of users. As an example to demonstrate our point, we have written a tutorial that illustrates the use of Pilot-Point based calibration made available through PEST in the HydroGeoSphere modelling environment. We hope that this example will encourage the modelling community to develop tutorials and make them available to the wider public.

1. Introduction

Many environmental modelling software packages such as FEFLOW (Diersch, 2005), GMS or Visual Modflow (Guiguer and Franz, 1996) are equipped with graphical user interfaces (GUIs). GUIs provide a user-friendly environment that helps to focus on the aspects of modelling rather than on the technicality of the underlying code (Seibert and Vis, 2012). For example, FEFLOW provides all required functions to elaborate and post-process an entire, georeferenced flow and transport model within a GUI. In principal, no additional software such as text-editors or GIS packages for the manipulation of spatial data are required in the process (Trefry and Muffels, 2007). On the other hand, the absence of a GUI leaves it to the user to resolve all the technical challenges related to running the software.

Given the user-friendliness of models equipped with a GUI, software without a GUI are in a strategic disadvantage, independent of their capabilities or level of sophistication. In fact, the absence of a GUI might even prevent the wide-spread application of a code. A good illustrative example is the model independent calibration software PEST (Doherty, 2010), that provides a large amount of calibration tools, including Pilot Points calibration. A range of hydrological models with a GUI (e.g. MODFLOW (Harbaugh, 2005), MT3DMS (Zheng, 1990), SEAWAT (Langevin et al., 2008), FEFLOW (Diersch, 2005), MicroFEM (Hemker and de Boer, 2009) and RSM (South Florida Water Management District, 2005)) provide an interface to make easy use of PEST capabilities within the modelling environment. Pilot Points calibration through PEST has been used with all of these codes (e.g. Dausman et al., 2010, Doherty, 2003, Wiese and Nutzmann, 2011). Other hydrological models such as PARFLOW (Ashby and Falgout, 1996), InHM (VanderKwaak and Loague, 2001), OpenGeoSys (Kolditz et al., 2012) or HydroGeoSphere (Therrien et al., 2010) do not have a GUI or do not support PEST or Pilot Points calibration in a GUI environment. The absence of a GUI seems to preclude the application of the Pilot Points options available in PEST. In fact, we are not aware of a single published Pilot Points approach through PEST for a model without the GUI-support for these functionalities. This example is in line with Yamamoto (2008) who states that software without GUIs limit and hurdle users to run sophisticated numerical simulation or calibration approaches. Despite all advantages of GUIs, however, there are drawbacks. For example, the capabilities of GUIs often lag behind the software development of the underlying code. Also, GUIs often do not provide an interface to all the available capabilities. Moreover, GUIs are a potential source of bugs and invite inexperienced modellers to run simulations without having a clear understanding of

---

* Corresponding author. Present address: Department of Water Resources and Drinking Water, Eawag Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland.
E-mail address: Christian.moeck@eawag.ch (C. Moeck).

http://dx.doi.org/10.1016/j.envsoft.2014.12.018
1364-8152/© 2015 Elsevier Ltd. All rights reserved.

the underlying model concept. In addition, compatibly problems might occur if the operating system of the computer is changed or updated. Finally, the development of a GUI is expensive.

A flexible alternative to GUIs are tutorials. Tutorials guide beginners on their first steps on model use (Heistermann et al., 2013) and are part of a learning process. Compared to a GUI, tutorials are far less time-consuming to develop and can be rapidly adapted to changes and extensions of the software. However, despite the advantages, only a relatively small number of tutorials has been published. For instance, Zambrano-Bigiarini and Rojas (2013) provide a model-independent R (R Development Core Team, 2011) package (hydroPSO) used for model calibration. For the package hydromad (Andrews et al., 2011), a modelling framework for water balance accounting and flow routing in spatially aggregated catchments, a tutorial as well as a web page are provided describing theory and model use (http://hydromad.catchment.org/). Käser et al. (2014) provided tutorials for the interpolation of channel cross-sectional data and the refinement of a mesh along a stream in areas of high topographic variability.

Despite these examples we believe that the value of tutorials is underestimated in the environmental modelling community. However, tutorials have the potential to bring non-supported software without a GUI environment easily accesable to the wider community. To demonstrate our point, we have written a tutorial that illustrates the use of Pilot-Point based calibration made available through PEST in the HydroGeoSphere (HGS) modelling environment. Linking the Pilot Points method with HGS is especially important because HGS is a powerful state of the art tool used to reproduce many environmental processes. As stated by Brunner and Simmons (2012), the model has been applied to a wide range of hydrological subjects such as groundwater – surface water interaction (Irvine et al., 2012; Partington et al., 2012, 2013; Bartsch et al., 2014), flow in fractured rocks (Vujevic and Graf, 2012; Graf and Therrien, 2008), groundwater response to climate change (Goderniaux et al., 2011, 2009) and hydrological processes at catchment scale (Li et al., 2008; Sciuto and Diekkrueger, 2010). Representation of spatial heterogeneity is highly important for all of the above mentioned application of HGS. Pilot Points provide the means to account for spatial heterogeneity during model development and calibration.

We discuss the application of the PEST Pilot Points capabilities in a general way, in order to ensure that our tutorial is also useful in modelling environments other than HydroGeoSphere. We also demonstrate the advanced capabilities of PEST for Cross-validation (CV) and Linear Uncertainty Analysis (LUA). CV and LUA is applied to validate alternative calibrated hydraulic conductivity fields and to identify the importance of observations to parameter estimates and predictions. These powerful features of PEST are not supported by GUIs and have not been employed in the context of Pilot Point calibration.

Our paper is organized as follows: In Section 2, we provide some background information on model calibration and Pilot Points. In Section 3, we explain how the pilot point method implemented in PEST can be linked to a code without GUI (in our case HydroGeoSphere) according to our tutorial. Also, the theory of CV and LUA is briefly discussed. In Section 4 an example is presented. Results and conclusions of both, Pilot Points calibration and CV as well as LUA are presented in Sections 5 and 6. Beside results and conclusions, the value and philosophy of tutorials are discussed. The appendix contains a detailed step-by-step tutorial to reproduce the example.

2. Context: model calibration, Pilot Points and Uncertianty analysis

Spatial variations of hydraulic properties play an important role in controlling flow and solute movement in the subsurface (e.g. Sudicky and Huyakorn, 1991; Sudicky et al., 2010; Zheng and Gorelick, 2003). Virtually every hydrogeological investigation requires estimates of hydraulic conductivity (Butler, 2005). However, heterogeneity can typically not be investigated in all details. Nevertheless, model parameters must be provided to the employed models and therefore, calibration is required.

During model calibration, parameters are adjusted until the model output fits historical field measurements (Moore and Doherty, 2006; 2005). The classical calibration approach is based on the principle of parsimony. It consists of subdividing the model domain into zones of piecewise constancy of the hydraulic properties that are then calibrated. The strengths and weaknesses of this approach have been described by Hill and Tiedeman (2006) and are subject to an ongoing debate in the scientific community (Doherty, 2010, 2009, Doherty and Hunt, 2009; Hill, 2010). Pilot Points are an alternative to the zonation approach. Pilot Points introduce great flexibility to calibrate heterogeneous systems without neglecting expert knowledge (Doherty, 2003). Renard (2007) provided a brief history of Pilot Points methods, which was proposed first by de Marsily (1978). The method was further developed by de Marsily et al. (1984), Lavenue, et al. (1995) and Ramarao et al. (1995) and has been implemented into the automatic parameter estimation software PEST (Doherty, 2010).

Pilot Points have been combined with various numerical models of different conceptual complexity to simulate a wide range of environmental problems. For instance, Dausman et al. (2010) used SEAWAT (Langevin and Guo, 2006) to simulate the effect of variable density flow and transport and developed an optimization approach for efficient data acquisition. MODFLOW (Harbaugh, 2005) was used for an investigation of the potential error in predictions made by highly parameterized models calibrated using regularized inversion (Tonkin et al., 2007). Herckenrath et al. (2011) used a "Null-Space-Monte-Carlo" approach to quantify predictive uncertainty for a saltwater intrusion problem. The impact of Pilot Points positions on inversions results was investigated by Kowalsky et al. (2012) and the importance of intraborehole flow in solute transport by Ma et al. (2011). However, using the Pilot Points capabilities of PEST is not straightforward because the numerical model and the calibration software have to interact throughout the calibration process, and further Pilot Points have to be assigned to specific mesh-locations while the interpolation between the Pilot Points has to be mapped to the model mesh.

3. Pilot Points calibration and Uncertainty analysis using PEST

Calibration of a model using Pilot Points involves the following steps. First, Pilot Points are distributed over the model domain. The distribution can be both regularly spaced or denser for locations of interest. Note that any distributed model parameter can be calibrated with Pilot Points. Once the Pilot Points have been defined, PEST interpolates the model parameters (starting from initial values based on the expert knowledge) between the Pilot Points based on a user defined method (e.g. Kriging). PEST offers the option to define multiple geostatistical models that can be assigned to predefined zones in the model domain. The model is then run with the given parameter field, and an objective function is calculated by comparing measured and simulated observations. In the subsequent calibration, Pilot Points values are varied and the corresponding objective functions calculated. PEST modifies the Pilot Points values to minimize the objective function. All options of PEST to minimize the objective function can be used for this task, including mathematical regularisation approaches (e.g. Tikhonov regularization and/or Singular value decomposition (SVD)). An extensive description of the Pilot Points theory can be found in Christensen and Doherty (2008), Doherty (2009, and 2003).

3.1. Pilot Points calibration in PEST to the HydroGeoSphere environment

The following workflow guides users through the aforementioned calibration process in HGS according to our tutorial. The application of Pilot Points requires coordinate transformations between the model and PEST. In the “PEST domain” values for Pilot Points are estimated through the inverse process and the interpolation between them is carried out through “ppk2facg” and “fac2g”. These two programs are part of the PEST suite (Doherty, 2010). A re-transformation into the model specific file structure has to be carried out previous to the model run. Fig. 1 and the following points illustrates this workflow for HGS in greater detail.

[1] In a first step, the HGS mesh must be written to a file by using the HGS command “Mesh to Tecplot”, which causes the pre-processor of HGS (grok) to write all available mesh information.

[2] The program “R2Cord” (available as supplementary material) makes this information readable to the PEST program “ppk2facg” (Doherty, 2010).

[3] Then, executing “ppk2facg” generates a set of kriging factors through which the interpolation can take place from a set of Pilot Points. To run “ppk2facg”, two additional files are required. The first file (Pilot Points file) provides name, easting and northing of every Pilot Point as well as the predefined zone and the assigned value. These values are used for the spatial interpolation to the elements of the model mesh. The second file contains the geostatistical structure defined through a variogram. This information is required for kriging between the Pilot Points. In addition to generating a set of kriging factors, “ppk2facg” writes regularization information, which can optionally be used by “ppkreg” to add prior information to the PEST control file. Prior information used for regularization can help to avoid overfitting and numerical instability caused by the heterogeneity of the generated model parameters. All tasks outlined in this paragraph have to be carried out only once for a Pilot Points calibration approach (see top panel of Fig. 1).

[4] After completing the preparation of the input files, the program “fac2g” (Doherty, 2010) undertakes the interpolation from Pilot Points to elements of the HGS mesh. This interpolation tends to generate smoothed projections of the true hydraulic conductivity. If the number of Pilot Points or observations is too small, the degree of heterogeneity is likely to be underestimated.

[5] The interpolated field is mapped on the elements of the HGS-mesh by using the program “K2HGS” (available as supplement). The pre-processor of HGS (grok) is then initialized by PEST, followed by the model run itself. After achieving the predefined convergence criteria, HGS generates all model specific output data.

[6] PEST reads these data and calculates the objective function. Based on the updated objective function, PEST generates a new set of values for the Pilot Points. This procedure is repeated until PEST aborts the calibration process following the predefined convergence criteria. The sequential calling of the executables as described above must be defined in a batch file.

3.2. Implementation of Pilot Points for codes without GUI support

The provided workflow can be applied to other models by changing two steps. Firstly, coordinates of each mesh element must

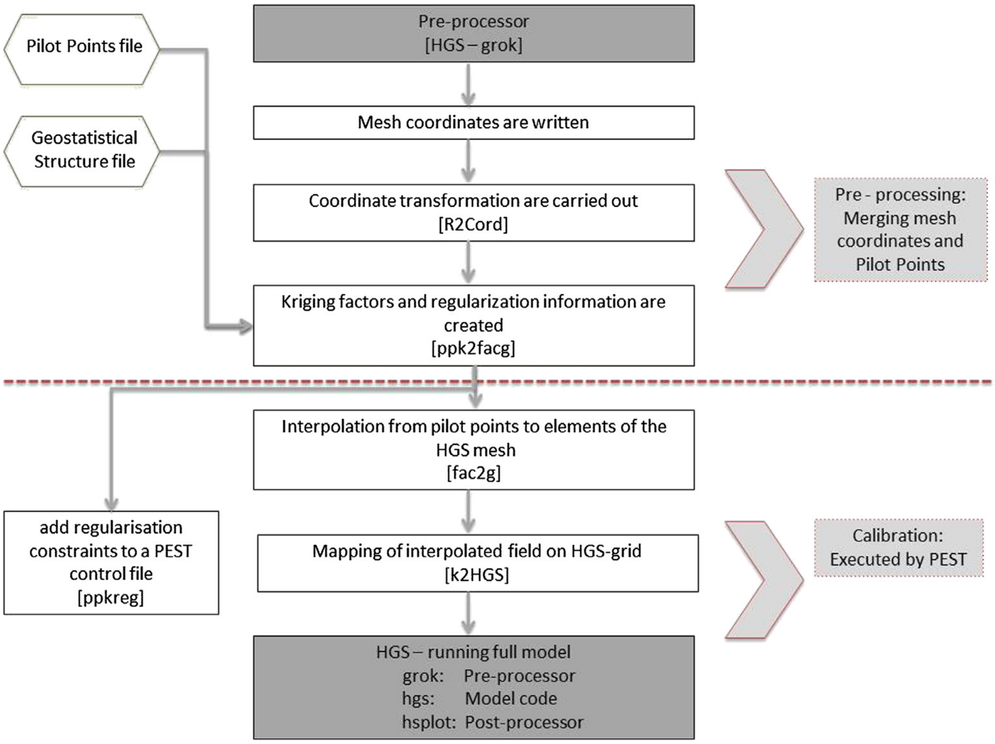

Fig. 1. Flowchart of the methodology to combine Pilot Points calibration using PEST with HGS. In the top panel, the pre-processing and the preparation of the input files are shown. The lower panel illustrates the calibration procedure. The rectangles in gray colour show the specific HGS software.

be provided in a text file. Secondly, the “R2Cord” program (source code available as supplement) must be modified to ensure that the mesh structure can be imported by PEST. From this point on, the steps of step **[3]** are carried out until the execution of “K2HGS”. “K2HGS” has to be modified (source code available as supplement) to import and create a file readable by the numerical model.

3.3. Model performance criteria, Cross-Validation and Linear Uncertainty Analysis

We use a scatter and residual plot to evaluate the model performance as well as the Percent of bias (Pbias), Root Mean Square Error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE) and Kling-Gupta efficiency (KGE). The theoretical background to these performance criteria are given in the tutorial. The application of such multiple performance criteria are generally recommended because a single criterion evaluates only specific aspects of model performance (e.g. Krause et al., 2005). As mentioned in Bennett et al. (2013) model performance criteria which are applicable for one specific model application may not be sufficient for another. The performance evaluation is, however, just one iterative step of model development (Jakeman et al., 2006). Although it is impossible to give standard techniques for model evaluation, a practical general five step procedure to characterise model performance can be found in Bennett et al. (2013).

Here, Cross-Validation (CV) and Linear Uncertainty Analysis (LUA) are applied. The target of CV and LUA is to identify the dependence of the model fit and estimated parameter values on each observation (Foglia et al., 2007; Hill and Tiedeman, 2006). CV is a computationally demanding method, which accounts for model nonlinearity due to the model re-calibration. For each re-calibration one observation after the other observations is omitted.

Using LUA, parameter identifiability of each parameter depending on the information content of the available observations can be calculated. The contribution to the pre- and post-calibration error variance and uncertainty of the different parameters can be computed. Also the worth of different observations in lowering the error variance and uncertainty by selectively removing observations can be quantified. A quite important factor, which demonstrates the attractiveness of the present Linear Uncertainty Analysis, is that parameter or observation values are not required and the analysis can be carried out even before model calibration. Detailed descriptions about the concepts can be found in the tutorial or in the listed references (CV: e.g. Brunner et al. (2012), James et al. (2009), Moore and Doherty (2006, 2005) and LUA: e.g. Foglia et al. (2007), Hill and Tiedeman (2006)).

4. Example

In the following example the hydraulic conductivity field of a finite element 2-D HGS model will be calibrated with Pilot Points using regularization to illustrate the application of the tutorial. Observations for the calibration were taken from a reference model and the two hydraulic conductivity fields (reference and calibrated) are compared. After the calibration, we apply CV by omitting single or groups of observations and re-calibrate the model. Subsequently, we apply different types of statistics following Cook and Weisberg (1982). We rate how much the estimated parameter and the simulated values vary during the calibration when observations are omitted. The influences of each observation on the Pilot Points values and on model predictions are calculated. Then the LUA analysis is carried out. We use the PEST utility “genlinpred” (Doherty, 2010), a batch program running the PEST utilities required for the linear uncertainty analysis to LUA.

4.1. Reference model

A synthetic reference model of a unconfined porous aquifer was created. The 2-D steady state model has a stationary, spatially variable hydraulic conductivity field throughout the entire model domain and is described by a log exponential variogram with a range of 200 m and a sill of 0.29. The mean hydraulic conductivity is 5.8 E^{-4} m s^{-1} (Fig. 2).

Steady state groundwater flow is induced with constant head boundaries of 1 m at the southern border and 5 m at the northern border. “No flow” boundaries are imposed on the remaining borders. A total of 12 observation wells for head measurements are distributed in the model domain (Fig. 3). No random noise was added to the head observations in order to simplify the data estimation.

4.2. Model calibration

130 Pilot Points were distributed in the model domain, regularly spaced. This number was chosen to ensure that sufficient Pilot

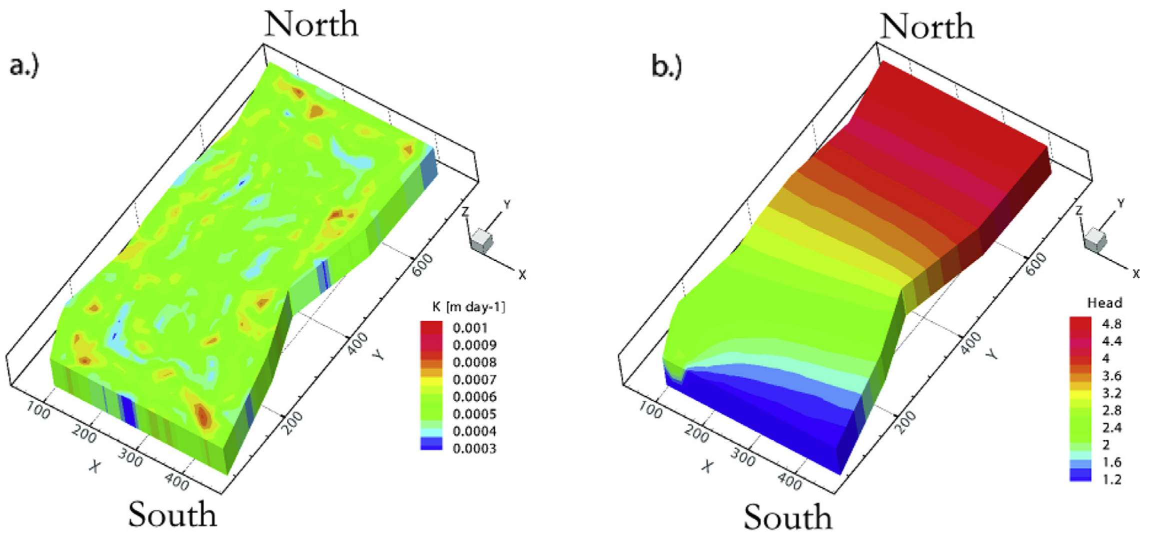

Fig. 2. (a) Distribution of reference hydraulic conductivity [m day^{-1}] within the finite element model domain. (b) Simulated heads within the model domain.

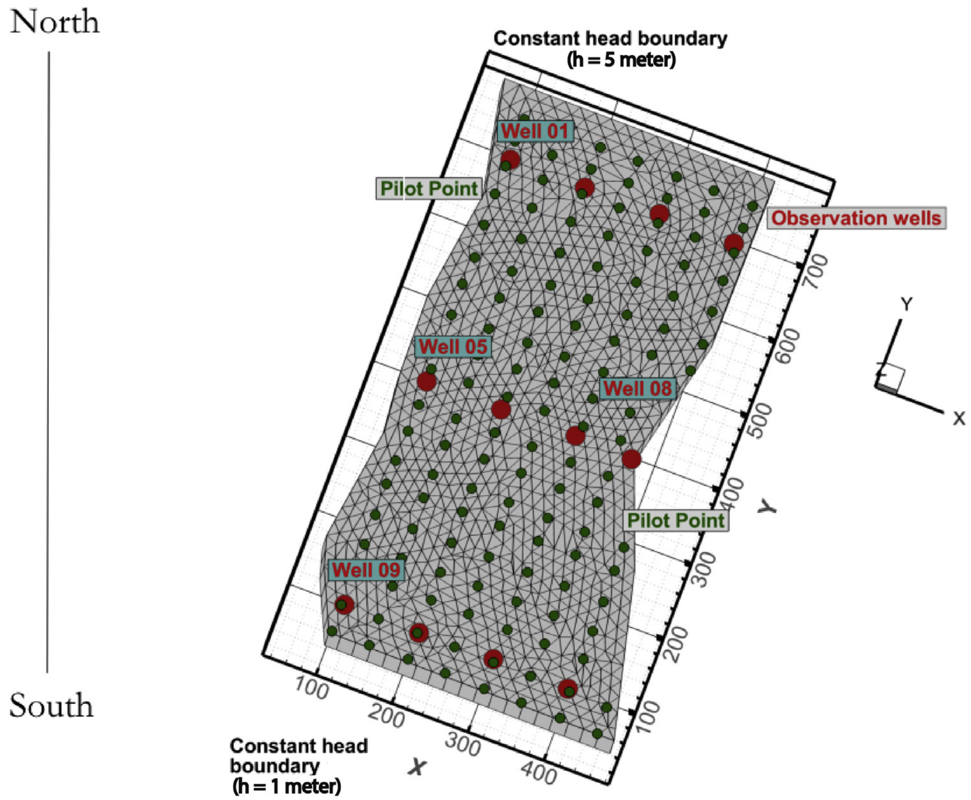

Fig. 3. Model domain with locations of 12 observation wells (red points), regularly distributed 130 Pilot Points (small green points). Constant head boundary conditions are imposed at the northern and southern border whereas the lateral border consist of no flow boundaries. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Points exist to reproduce the reference hydraulic conductivity field. Typically this number is uncertain due to the unknown degree of subsurface heterogeneity and therefore as many Pilot Points as possible should be used. A high number of Pilot Points leads to a good identification of the heterogeneity, good geostatistical characterization (Doherty, 2003) and less sensitivity to the location of the Pilot Points (LaVenue and Pickens, 1992). Observations of hydraulic head of the reference wells are used for the calibration (Fig. 3).

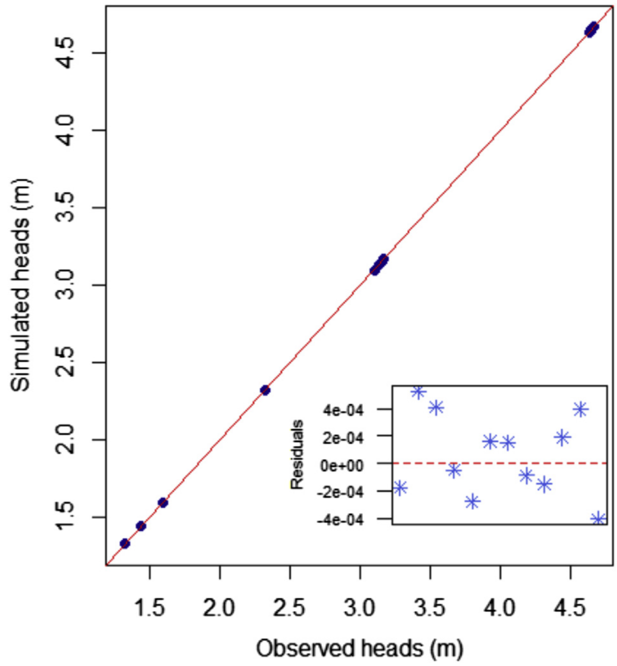

Fig. 4. Simulated versus observed heads. Residuals of all observation wells are displayed in the panel in the lower right.

In the calibration process, initial log transformed hydraulic conductivity of 1.4 E^{-5} m s^{-1} are assigned to all Pilot Points as well as the lower and upper bounds of 1.0 E^{-10} and 1.0 E^{10} m sec^{-1}. This large parameter range was chosen to demonstrate how well PEST estimates realistic parameter values using mathematical regularization. In our synthetic example all observations were equally reliable without any assumed measurement errors and therefore equal weights were applied to all observations. Singular value decomposition (SVD) as well as Tikhonov regularization are applied in the calibration process.

5. Results and discussion

A very good fit was obtained between model output heads and head observations from the synthetic reference model through six iteration steps in the calibration process (Fig. 4). In the residual plot (Fig. 4, smaller panel) a uniform spread of residuals can be observed which indicates that no systematic model bias is introduced during the conceptual model formulation and calibration. The calculated values for Pbias, RMSE, NSE and KGE show that model performance is very good under all criteria (calculated values are given in the tutorial).

Comparing the hydraulic conductivity fields between reference and calibrated model shows large differences in their structure (Fig. 5). The calibrated hydraulic conductivity field shows only a fraction of the real heterogeneity of the reference model. This can

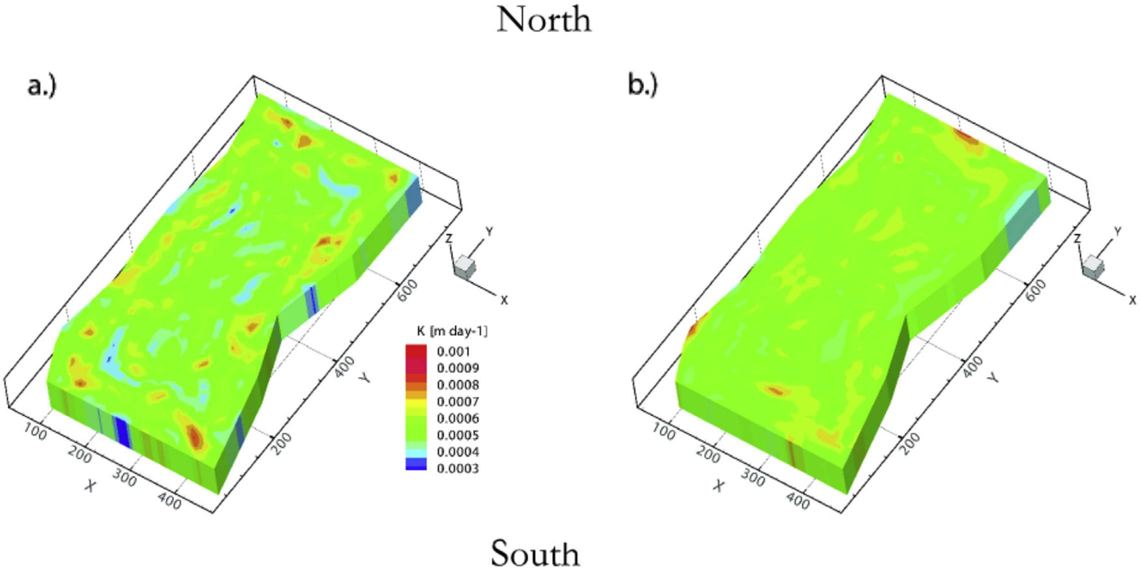

Fig. 5. (a) Reference and (b) calibrated hydraulic conductivity field [m day^{-1}].

result in considerable predictive uncertainties, espically in contaminant transport, as demonstrated by Moore and Doherty (2006, 2005). Post-processing allows to quantify the uncertainty of the head predictions.

To subsequently calculate the influence of each observation on predicted heads, CV is used. This analysis indicates that the head observations close to the southern border (*well 9 to 12*) contain the highest information content for the calibration from all observations (Fig. 6, see left axis). In this particular area, the head change (gradient) is highest compared to the other areas in the model domain. The strongest effect from omitting an observation occurs at observation *well 9*. Reproducing the piezometric head field without observation *well 9* increases the uncertainty. Omitting an observation in the northern part of the model domain (see for instance piezometric head in Fig. 2) has no or only a small effect on the head predictions. The horizontal head differences are small between the observation wells in the northern part (e.g. the 4 wells close to the northern border).

Similar results can also be obtained with the LUA (Fig. 6, see right axis). Omitting observation *well 9* in the calibration would increase the predictive uncertainty variance in the calibration process for reproducing the reference piezometric head field. Omitting any other observation only has a small effect for the head predictions based on the LUA. Although both methods indicate similarity in the results, the computation time is different (Only 131 model runs were needed for LUA whereas for CV 1180 runs carried out). For CV a re-calibration is required for each omitted parameter or parameter group whereas for the LUA it is not required. In addition, calibrated parameter values are also not required. The LUA is only based on parameter and observation sensitivities. Therefore, LUA can be already applied albeit the model is not yet calibrated. However, it should be noted that CV shows the actual effect of omitted observations for the calibration in contrast to LUA.

The influence of omitting individual or groups of observations for the Pilot Points hydraulic conductivity values obtained by CV are shown in Fig. 7 (Equation 21, see tutorial). For most omitted observations only a small change in the parameter values can be observed (blue to light yellow colours (in web version)). Only for the omitted observation *well 9* and observation *group lower* (which in fact includes observation *well 9*) large changes in the Pilot Points values occur (dark yellow to red colours).

In summary, observation *well 9* provides the highest information content for the calibration of the hydraulic conductivity field. Omitting this observation increases the predictive uncertainty. Parameter identifiability for each Pilot Point (Figure 9 in the tutorial) indicates that the data worth of the hydraulic heads is insufficient to constrain the hydraulic conductivity field, even if head observation 9 is included in the calibration. To reduce uncertainty and increase the parameter identifiability additional types of observation, such as contaminant concentration could be used. Additional methodologies to explore the predictive uncertainty have been developed and applied (e.g. Brunner et al., 2012; Moore and Doherty, 2006; 2005; Christensen and Doherty, 2008; Dausman et al., 2010; Doherty and Christensen, 2011; M Gallagher and Doherty, 2007; MR Gallagher and Doherty, 2007; Herckenrath et al., 2011; Tonkin et al., 2007; Schilling et al., 2014). These approaches are readily applicable to models calibrated through Pilot Points methods. For instance, the contribution of the parameter null space can be investigated by using Pilot Points (Hunt et al., 2007) and/or in combination with stochastic field generation to apply a Monte Carlo analysis (Tonkin and Doherty, 2009). Especially the combination of Pilot Points calibration with a stochastic field generator gives the possibility to generate many different stochastic realizations of, for instance hydraulic conductivity fields. Subsequently running the model on all generated fields provides an additional approach to represent uncertainty of predictions. Starting PEST from a succession of different starting values helps to overcome local minima as demonstrated by Skahill and Doherty (2006). Combing these approaches equips modellers with a powerful means to explore uncertainty within satisfying stochastic calibration constrains and can avoid that only local minima are found during the calibration process.

6. Summary and conclusions

In our example we illustrated the application of Pilot Points calibration using the program PEST in combination with the physically based finite element model HGS. Additionally CV and LUA were used for post-processing. Both CV and LUA show similar results, but the latter with much lower computational cost. The fact that the model does not have to be calibrated before an

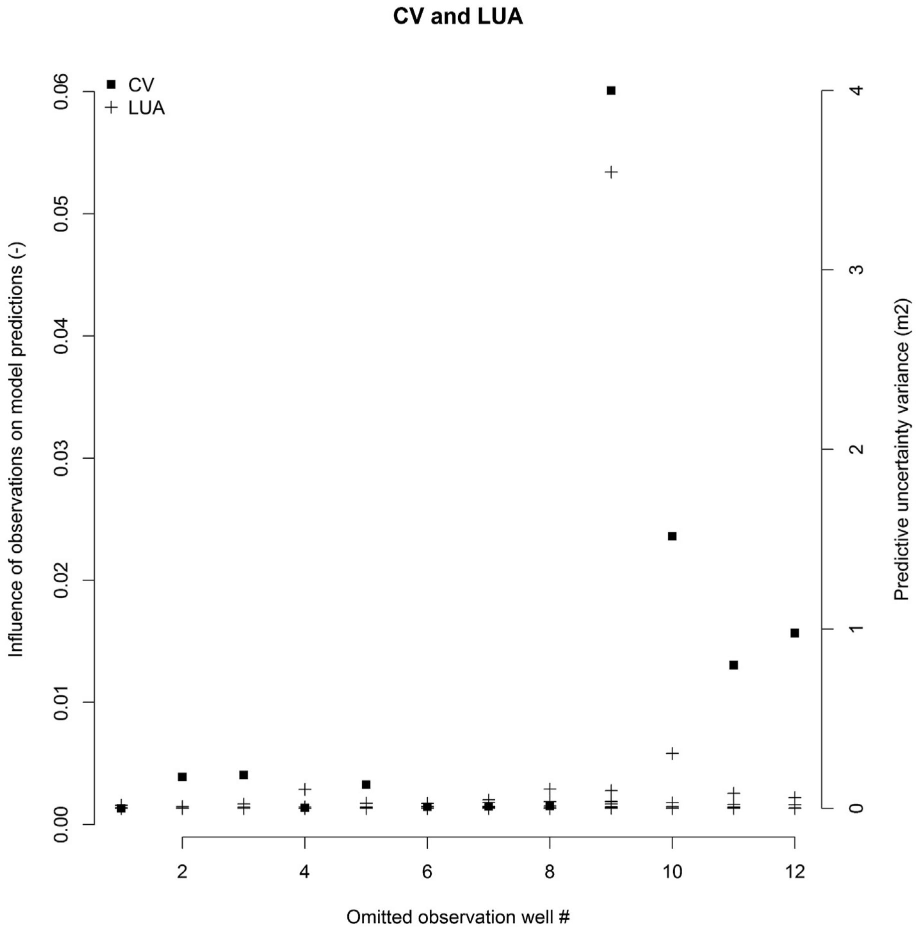

Fig. 6. Left axis: Influence of observations on model predictions (Equation (2), see tutorial) by CV. On the x-axis the omitted observation is shown. The predictions are the simulated head if the chosen observation is omitted. Right axis: The increase of predictive uncertainty variance for each head due to the loss of observation is shown (Equation (3), see tutorial).

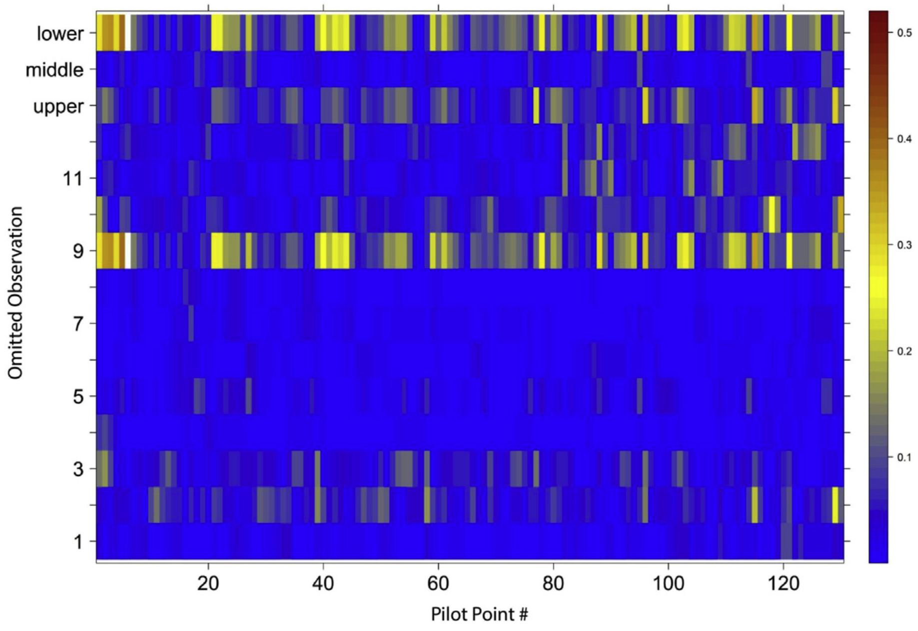

Fig. 7. Parameter influence statistics (Equation (1), see tutorial) from the CV experiment. Omitted individual observations are labeled as observation 1 to 12. Omitted groups of observations are labeled upper (the 4 observations close to the northern border), lower (the 4 observations close to the southern border) and middle (the 4 observations between down and top). All 130 Pilot Points used in the calibration are displayed along the x-axis. The statistic shows the differences between a calibration with all observation and the re-calibration with omitted observation(s) for the Pilot Points hydraulic conductivity values.

uncertainty analysis can be applied makes LUA very attractive tool in a modelling process.

The presented workflow and comprehensive tutorial guides users through the advanced calibration process and helps to overcome the technical challenges associated with employing the Pilot Points method using PEST without a GUI. Although the Pilot Points method is an attractive tool in model calibration, often only a fraction of the actual subsurface heterogeneity can be represented and the risk of over-fitting due to over-parameterisation of the inverse problem exists. Therefore, we have explored the capability of PEST to evaluate model results. However, it is important to know that many other model evaluation criteria exist and generally multiple performance criteria should be applied (Bennett et al., 2013).

Additionally to the tutorial we provided two executable files (source code available as supplement) that are required to use PEST with HGS. The program "R2Cord" transfers mesh coordinates and makes the information readable for further steps using the PEST programs. The mapping of the undertaken interpolation from Pilot Points to elements of the HGS mesh is done by the program "K2HGS". The provided workflow and tutorial can be easily applied to any other numerical models by adjusting the two executable files.

We hope that our provided example will encourage the modelling community to develop tutorials and make them available in order to make the latest software developments and model features more accessible to the wider public.

Acknowledgements

The authors gratefully acknowledge the financial assistance provided by the Swiss National Science Foundation, Project NFP 61, Sustainable Water Management. We are grateful for the discussion provided by Anthony Jakeman and Alexey Voinov. The authors thank the reviewer for providing valuable feedback. We thank Maria Herold, Oliver Schilling and Mehdi Ghasemizade for testing the tutorial and helpful discussions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2014.12.018.

References

Andrews, F.T., Croke, B.F.W., Jakeman, A.J., 2011. An open software environment for hydrological model assessment and development. Environ. Model. Softw. 26, 1171–1185. http://dx.doi.org/10.1016/j.envsoft.2011.04.006.

Ashby, S.F., Falgout, R.D., 1996. A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. Nucl. Sci. Eng. 124, 145–159.

Bartsch, S., Frei, S., Ruidisch, M., Shope, C.L., Peiffer, S., Kim, B., Fleckenstein, J.H., 2014. River-aquifer exchange fluxes under monsoonal climate conditions. J. Hydrol. 509, 601–614. http://dx.doi.org/10.1016/j.jhydrol.2013.12.005.

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20.

Brunner, P., Simmons, C.T., 2012. HydroGeoSphere: a fully integrated, physically based hydrological model. Ground Water 50. http://dx.doi.org/10.1111/j.1745–6584.2011.00882.x.

Brunner, P., Doherty, J., Simmons, C.T., 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. Water Resour. Res. 48 http://dx.doi.org/10.1029/2011wr011342.

Butler, J.J., 2005. Hydrogeological methods for estimation of spatial variations in hydraulic conductivity hydrogeophysics. In: Rubin, Y., Hubbard, S.S. (Eds.), Water Science and Technology Library, pp. 23–58.

Christensen, S., Doherty, J., 2008. Predictive error dependencies when using Pilot Points and singular value decomposition in groundwater model calibration. Adv. Water Resour. 31, 674–700. http://dx.doi.org/10.1016/j.advwatres.2008.01.003.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. In: Monogr. Stat. Appl. Probability 18. Chapman and Hall, New York.

Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010. Quantifying data Worth toward reducing predictive uncertainty. Ground Water 48, 729–740. http://dx.doi.org/10.1111/j.1745-6584.2010.00679.x.

de Marsily, G., 1978. De l'identification des systems hydrogeologiques. Doctorat d'etat thesis Ecole des Mines de Paris. Centre d'Informatique Geologique, Paris.

Diersch, H.J.G., 2005. FEFLOW finite element subsurface flow and transport simulation system. WASY GmbH Inst. Water Resour. Plan. Syst. Research 292.

Doherty, J., 2003. Ground water model calibration using Pilot Points and regularization. Ground Water 41, 170–177. http://dx.doi.org/10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J., 2010. PEST Groundwater Data Utilities. Brisbane, Australia: Watermark Numerical Computing. Downloadable from. www.pesthompage.org.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 http://dx.doi.org/10.1029/2011wr010763. W12534.

Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. J. Hydrol. 366, 119–127. http://dx.doi.org/10.1016/j.jhydrol.2008.12.018.

Foglia, L., Mehl, S.W., Hill, M.C., Perona, P., Burlando, P., 2007. Testing alternative ground water models using cross-validation and other methods. Ground Water 45, 627–641. http://dx.doi.org/10.1111/j.1745-6584.2007.00341.x.

Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. Environ. Model. Softw. 22, 1000–1020. http://dx.doi.org/10.1016/j.envsoft.2006.06.007.

Gallagher, M.R., Doherty, J., 2007. Parameter interdependence and uncertainty induced by lumping in a hydrologic model. Water Resour. Res. 43 http://dx.doi.org/10.1029/2006wr005347.

Goderniaux, P., Brouyere, S., Fowler, H.J., 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. http://dx.doi.org/10.1016/j.jhydrol.2009.04.017.

Goderniaux, P., Brouyere, 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 http://dx.doi.org/10.1029/2010wr010082. Artn W12516.

Guiguer, N., Franz, T., 1996. Visual Modflow: User's Manual.

Harbaugh, A.W., 2005. MODFLOW-2005—The U.S. Geological Survey Modular Ground-water Model—The Ground-water Flow Process. US Geological Survey Techniques and Methods. Book 6, chap A-16.

Heistermann, M., Jacobi, S., Pfaff, T., 2013. Technical Note: an open source library for processing weather radar data (wradlib). Hydrol. Earth Syst. Sci. 17, 863–871. http://dx.doi.org/10.5194/hess-17-863-2013.

Hemker, C.J., de Boer, R.G., 2009. MicroFEM User's Guide. Dr CJ (Kick) Hemker, Amsterdam, The Netherlands, 25 pp.

Herckenrath, D., Langevin, C.D., Doherty, J., 2011. Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo. Water Resour. Res. 47 http://dx.doi.org/10.1029/2010wr009342.

Hill, 2010. Comment on "Two statistics for evaluating parameter identifiability and error reduction" by John Doherty and Randall J. Hunt. J. Hydrol. 380 (3), 481–488.

Hill, Mary C., Tiedeman, Claire R., 2006. Effective Groundwater Model Calibration: with Analysis of Data, Sensitivities, Predictions, and Uncertainty. John Wiley & Sons.

Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. Ground Water 45, 254–262. http://dx.doi.org/10.1111/j.1745-6584.2007.00316.x.

Irvine, D.J., Brunner, P., Franssen, H.J.H., Simmons, C.T., 2012. Heterogeneous or homogeneous? implications of simplifying heterogeneous streambeds in models of losing streams. J. Hydrol. 424, 16–23. http://dx.doi.org/10.1016/j.jhydrol.2011.11.051.

Jakeman, A.J., Letcher, J.P., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environ. Model. Softw. 21 (5), 602–614. May 2006.

James, S.C., Doherty, J.E., Eddebbarh, A.A., 2009. Practical postcalibration uncertainty analysis: Yucca Mountain, Nevada. Ground Water 47, 851–869. http://dx.doi.org/10.1111/j.1745-6584.2009.00626.x.

Käser, D., Brunner, P., Graf, T., Cochand, F., McLaren, R., Therrien, R., 2014. Channel representation in models coupling groundwater and stream water: pitfalls and how to avoid Them. Ground Water 52, 827–836. http://dx.doi.org/10.1111/gwat.12143.

Kolditz, O., Bauer, S., Bilke, L., Bottcher, N., Delfs, J.O., Fischer, T., Gorke, U.J., Kalbacher, T., Kosakowski, G., McDermott, C.I., Park, C.H., Radu, F., Rink, K., Shao, H., Shao, H.B., Sun, F., Sun, Y.Y., Singh, A.K., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, Y., Xie, M., Xu, W., Zehner, B., 2012. OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media. Environ. Earth Sci. 67, 589–599. http://dx.doi.org/10.1007/s12665-012-1546-x.

Kowalsky, M.B., Finsterle, S., Williams, K.H., Murray, C., Commer, M., Newcomer, D., Englert, A., Steefel, C.I., Hubbard, S.S., 2012. On parameterization of the inverse problem for estimating aquifer properties using tracer data. Water Resour. Res. 48 http://dx.doi.org/10.1029/2011wr011203. W06535.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 89–97.

Langevin, C.D., Guo, W., 2006. MODFLOW/MT3DMS–Based simulation of variable-density ground water flow and transport. Ground Water 44, 339–351. http://dx.doi.org/10.1111/j.1745-6584.2005.00156.x.

Langevin, C.D., Thorne, D.T., Dausman, A.M., Sukop, M.C., Guo, W., 2008. SEAWAT Version 4—A Computer Code for Simulation of Multi-species Solute and Heat Transport. US Geological Survey Techniques and Methods. Book 6, chap A-22, 39 pp.

LaVenue, A.M., Pickens, J.F., 1992. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resour. Res. 28 (6), 1543–1569.

Lavenue, A.M., Ramarao, B.S., Demarsily, G., Marietta, M.G., 1995. Pilot points methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields .2. Appl. Water Resour. Res. 31, 495–516. http://dx.doi.org/10.1029/94wr02259.

Li, Q., Unger, A.J.A., Sudicky, E.A., Kassenaar, D., Wexler, E.J., 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. http://dx.doi.org/10.1016/j.jhydrol.2008.05.024.

Ma, R., Zheng, C., Tonkin, M., Zachara, J.M., 2011. Importance of considering intra-borehole flow in solute transport modeling under highly dynamic flow conditions. J. Contam. Hydrol. 123, 11–19. http://dx.doi.org/10.1016/j.jconhyd.2010.12.001.

Moore, C., Doherty, J., 2005. Role of the calibration process in reducing model predictive error. Water Resour. Res. 41 http://dx.doi.org/10.1029/2004wr003501. W05020.

Moore, C., Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. Adv. Water Resour. 29, 605–623. http://dx.doi.org/10.1016/j.advwatres.2005.07.003.

Partington, D., Brunner, P., Simmons, C.T., Werner, A.D., Therrien, R., Maier, H.R., Dandy, G.C., 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. J. Hydrol. 458, 28–39. http://dx.doi.org/10.1016/j.jhydro1.2012.06.029.

Partington, D., Brunner, P., Frei, S., Simmons, C.T., Werner, A.D., Therrien, R., Fleckenstein, J.H., 2013. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resour. Res. 49 (9), 5501–5519.

Ramarao, B.S., Lavenue, A.M., Demarsily, G., Marietta, M.G., 1995. Pilot points methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields .1. Theory and computational experiments. Water Resour. Res. 31, 475–493. http://dx.doi.org/10.1029/94wr02258.

Renard, P., 2007. Stochastic hydrogeology: what professionals really need? Ground Water 45, 531–541. http://dx.doi.org/10.1111/j.1745-6584.2007.00340.x.

Schilling, O.S., Doherty, J., Kinzelbach, W., Wang, H., Yang, P.N., Brunner, P., 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. J. Hydrol. 519, 2258–2271. Part B, 27 November 2014. ISSN 0022–1694 doi.org/10.1016/j.jhydrol.2014.08.063.

Sciuto, G., Diekkrueger, B., 2010. Influence of soil heterogeneity and spatial discretization on catchment water balance modeling. Vadose Zone J. 9, 955–969. http://dx.doi.org/10.2136/vzj2009.0166.

Seibert, J., Vis, M.J.P., 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. Hydrol. Earth Syst. Sci. 16, 3315–3325. http://dx.doi.org/10.5194/hess-16-3315-2012.

Skahill, B.E., Doherty, J., 2006. Efficient accommodation of local minima in watershed model calibration. J. Hydrol. 329, 122–139. http://dx.doi.org/10.1016/j.jhydrol.2006.02.005.

South Florida Water Management District, 2005. Theory Manual—Regional Simulation Model (RSM). SFWMD, Office of Modeling, West Palm Beach, Fla, 308 pp.

Sudicky, E.A., Huyakorn, P.S., 1991. Contaminat migration in imperfectlz known heterogeneous groundwater systems. Rev. Geophys. 29, 240–253.

Sudicky, E.A., Illman, W.A., Goltz, I.K., Adams, J.J., McLaren, R.G., 2010. Heterogeneity in hydraulic conductivity and its role on the macroscale transport of a solute plume: from measurements to a practical application of stochastic flow and transport theory. Water Resour. Res. 46 http://dx.doi.org/10.1029/2008wr007558. W01508.

Therrien, R., McLaren, R.G., Sudicky, E.A., 2010. HydroGeoSphre-a Three Dimensional Numerical Model Describing Fully Integrated Subsurface and Surface Flow and Solute Transport. Groundwater Simulations Group, University of Waterloo.

Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resour. Res. 45 http://dx.doi.org/10.1029/2007wr006678. Artn W00b10.

Tonkin, M., Doherty, J., Moore, C., 2007. Efficient nonlinear predictive error variance for highly parameterized models. Water Resour. Res. 43 http://dx.doi.org/10.1029/2006wr005348. W07429.

Trefry, M.G., Muffels, C., 2007. Feflow: a finite-element ground water flow and transport modeling tool. Ground Water 45, 525–528. http://dx.doi.org/10.1111/j.1745-6584.2007.00358.x.

VanderKwaak, J.E., Loague, K., 2001. Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model. Water Resour. Res. 37, 999–1013. http://dx.doi.org/10.1029/2000wr900272.

Vujevic, K., Graf, T., 2012. Modes of free convective flow in fractured-porous rock. Models Repos. Knowl. 355, 280–285.

Wiese, B., Nutzmann, G., 2011. Calibration of spatial aquitard distribution using hydraulic head changes and regularisation. J. Hydrol. 408, 54–66. http://dx.doi.org/10.1016/j.jhydrol.2011.07.015.

Yamamoto, Hajime, 2008. PetraSim: a graphical user interface for the TOUGH2 family of multiphase flow and transport codes. Ground Water 46.4, 525–528.

Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent particle swarm optimisation software for model calibration. Environ. Model. Softw. 43, 5–25. http://dx.doi.org/10.1016/j.envsoft.2013.01.004.

Zheng, C.M., 1990. MT3D—A Modular Three-dimensional Model for Simulation of Advection, Dispersion, and Reactions of Contaminants in Groundwater Systems. US Environmental Protection Agency, Ada, Okla, 170 pp.

Zheng, C.M., Gorelick, S.M., 2003. Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. Ground Water 41, 142–155.