Response to comments raised by Edwin Sutanudjaja

We thank Dr. Sutanudjaja very much for his excellent reviews. We also acknowledge him directly in the updated manuscript. Below you could find the original comments (marked in blue color) and point-by-point response (marked in black color).

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)
I have few minor points/suggestions for the latest version of the paper:

Page 3 lines 3-5 (P3L3-5): I suggest to remove the sentence “Although the methods for subsurface flow ..., the approaches for surface flow are inevitably based on some approximations and conceptualizations (e.g., diffusive-wave approximation, 1-D rill flow).”, which is not relevant for the manuscript. I am not sure that the approach used by the authors for modeling surface water flow (e.g. the “Muskingum algorithm”, Eqs. 5-8) and their simplified conceptualization for surface water-groundwater interaction (see my comment for Equation 11) are better than the ones used in some well-known ISSHMs.

Response: Thank you very much for the insightful comment. I agree that both ISSHMs and conceptual models rely on some conceptualizations and assumptions. Therefore, we deleted the statements from the manuscript accordingly. Please check it out at P2, L30-34.

Figure 1: Based on your response to my previous review, the term K is always zero. Could you please expand the caption of Figure 1 to explain it? Alternatively, you may want to remove K from Figure 1 such that it avoids the confusion with the symbol Ks used for hydraulic conductivities.

Response: Thank you again for this insightful comments. We followed the reviewer’s advice and removed the K from Figure. 1. Please check the updated Figure. 1.

Figure 5: What does the color represent? Please provide the legend. In the caption, please provide your threshold value (used for removing the original stream network).

Response: The color represent the value of runoff. We do not provide the legend
because the conceptualization of streams are different in mHM and in OGS, thus a unified legend may lead to misunderstanding. Nevertheless, we add this sentence to the caption of Figure 5:

“In the left figure, small tributaries where the runoff rates are below the threshold (0.145 m3/m) have been removed from the original stream network.”

Please check it out at the caption of Figure 5.

Equation 11: I think that the fluxes Q4 are always positive. Therefore, in terms of surface water and groundwater exchange, you conceptualize that water can only flow from groundwater to surface water, not vice versa. Please clarify this assumption/weakness in the manuscript.

Response: Thank you for this comment. We clarified this assumption in the manuscript already:

“Consequently, we do not account for any feedback between river and groundwater head fluctuations. This approach is parsimonious and numerically efficient, meanwhile fully preserves the well-tested parameterization algorithm in mHM, and has been evaluated by groundwater head measurements in the mesoscale test site.”

Please check it out at P23, L12-15.

P14L14-16: Could you please clarify and explain with what you meant by “... a Dirichlet boundary condition is assumed at the northwestern and northeastern edges”? Is it a fixed head boundary condition? It seems that this is inconsistent with your previous sentence stating that “... no flow boundaries are set at the outer perimeters ...”.

Response: Yes, we set a fixed head BC at the northwestern and northeastern edges. In order to avoid misconception, we slightly modified this sentence. Please check it out at P14, L3-5.

Figure 7b: Please provide the y-axis label, including its unit (m?)

Response: We have added the label as well as the unit. Please check the updated Figure 7.

Figure 8: As suggested in my previous review, please use the same maximum and minimum values (and the same color scales) for all sub-figures so that they can easily be compared among each other.

Response: Thank you for this suggestion. I have changed the colormap to a unified one. Please check the updated Figure 8.

Figures 12a-d: Please use the same maximum and minimum values/colors for
all sub-figures so that they can easily be compared among each other.

Figures 12e-f: Please use the same maximum and minimum values/colors for all sub-figures so that they can easily be compared among each other.

Response: Thank you for this suggestion. We agree that the label should be unified in the sub-figures in Figure 12. We also followed the reviewer’s suggestion and unify the color range in Figure 12. Please check it out in the updated manuscript.
Response to the referee review of Prof. Dr. Cirpka

We thank Prof. Dr. Cirpka so much for his efforts devoted in reviewing our manuscript. It is our great honor to exchange ideas with him, through which our manuscript has also been significantly improved. We are aware that Prof. Cirpka has put tremendous efforts in reviewing our manuscript, so we acknowledged him directly in the manuscript. We followed all of his constructive comments, aiming to meet his concerns and to improve the reader-friendliness of our manuscript. We have revised the manuscript carefully based on comments from two reviews. We replied (in black) every single question (in blue) raised by the reviewers and itemized them in below.

1 Major Remarks

This is the second time that I review this paper (which seems to have a somewhat longer history). In general, the paper has improved significantly, and the major coupling approach has become clearer.

1. While the general approach is explained clearly in section 2.3, the normal naïve reader would not think that the approach works the way as described after reading the abstract and even the introduction. From the abstract, it should be clear that the groundwater storage is NOT taken out of mHM; instead fluxes into and out of the groundwater body are evaluated by mHM and used as Neumann boundary condition for the groundwater model. That is not stated explicitly.

Response: We agree with the reviewer that it is necessary to state the coupling strategy clearly in the abstract. Thus, we added the following statements into the abstract:

“The two models are one-way coupled through model interfaces GIS2FEM and RIV2FEM, by which the grid-based fluxes of groundwater recharge and the river-groundwater exchange generated by mHM, are converted to fixed-flux boundary conditions of the groundwater model OGS. Specifically, the grid-based vertical reservoirs in mHM are completely preserved for the estimation of land-surface fluxes, while OGS acts as a plug-in to the original mHM modeling framework for groundwater flow and transport modeling.”

Please check the modified manuscript at L5-10, P1.

2. Likewise, the introduction reads as if the approach was towards two-way coupling, which is not done in the presented method. While I agree that shallow groundwater tables influence actual ET in reality (pae 2, lines 8-12), this is not the case in the chosen approach where ET belongs to mHM which does not know the groundwater table. You really need to explain the problem that you DO solve by the
approach, not the problems that you do not address. The motivation should be:

a. Conceptual hydrological models are good to get water balances right (essentially meeting hydrographs), they are quick to run and can be calibrated in stochastic frameworks. But these models cannot handle groundwater tables as that is not a state variable of the models.

b. Pde-based subsurface models are good to get hydraulic heads and travel-time distributions. In the unsaturated zone, they often fail getting quick flow components right because the relevant features are not resolved. Adding pde-based surface models is possible in general, but requires the resolution of fine-scaled features (which are sometimes not known), a tremendous number of uncertain parameters, and takes ages to run. Thus calibrating these models is cumbersome, and doing this in a stochastic framework is computationally unfeasible. Without pde-based surface models, the subsurface models still need boundary conditions at the surface.

c. The general idea of the paper is to perform the hydrological calculations with mHm, including the simplified linear groundwater storage, extract fluxes into- and out of groundwater from mHm, and use this as Neumann BC for a pde-based groundwater model using OGS. By this, mHm is “pimped up” to predict hydraulic heads in groundwater.

d. The model can thus NOT consider feedbacks between land-surface processes and the groundwater table. It is also clear that an aquifer system that inherently acts completely different than a linear storage cannot be simulated right. The approach builds on the assumption that the groundwater storage calibrated within mHm really represents groundwater and is not an alias of other storage mechanisms.

I highly recommend having a discussion on thinkable coupling strategies upfront and justifying the general choice already in the introduction. Do definitely not argue with features that you deem necessary but that you can’t deliver.

Response: thank you again for this important comment. According to the reviewer’s concern, the statements irrelevant to this paper should be eliminated. We totally agree with him, and eliminated the all statements on the feedback from groundwater model to land surface model. We also restructured the introduction according to the reviewer’s suggestion. For example, we added a discussion on the choice of one-way coupling strategy. Please check it out in the updated introduction section.

3. Running a groundwater model that has only Neumann boundary conditions (recharge, no flow, exchange with streams) can cause non-uniqueness problems. In the confined case, it is obvious that this is an ill-posed forward problem. In the
phreatic case it’s somewhat better, but definitely not optimal (specifically not at steady state). I teach my students that they need at least one Dirichlet boundary to fix the head. This should be discussed somewhere. (In the previous version I thought that the rivers either form a fixed-head or leaky boundary conditions).

Response: We are also aware that the exclusive use of Neumann BC could lead to nonuniqueness of numerical solution. Thus we set specified head BC at the northwestern and northeastern edges in the example application such that a good solution can be attained. We added the statements below into the manuscript:

“The exclusive use of Neumann BC is not recommended and may lead to nonuniqueness of solutions. At least one specified head boundary should be set at the perimeter or internal nodes to constrain the model solution.”

Please check it out at P10, L19-21.

4. The Pearson correlation coefficient is NOT a metric for the goodness of a fit, as it tests linearity between two data sets rather than identity. The correct metric would be the root mean square error or the mean absolute error. Please exchange all correlation coefficients by RMSE or MAE.

Response: We agree that RMSE and MAE are proper summery statistics to measure the model performance. Accordingly, we replaced all correlation coefficients by MAE. Please check it out in the updated manuscript.

Detailed Comments:

1. Page 1, lines 5-6: “by which grid-based fluxes of groundwater recharge and river-groundwater exchange generated by mHm are converted to fixed-flux boundary conditions of the groundwater model OGS.”
2. Page 1, line 8: “Specifically, the grid-based vertical reservoirs in mHM, including a linear groundwater storage, are completely preserved”
3. Page 1, line 11: “using discharge for mHm and long-term mean of groundwater head measurements for OGS, respectively.”
4. Page 1, line 12 and several times thereafter: Eliminate the correlation coefficient as metric for the goodness of a model fit.
5. Page 1, line 14: “so that the one-way coupled model”
6. Page 1, line 16: “provide an adequate representation”

Response: Those comments are about syntax errors. We have modified those statements accordingly.

7. Page 2, lines 8ff.: While this is true, it is not resolved at all by the presented approach. The arguments given here speak against using the method presented.
8. Page 2, line 24: I have no clue what you mean by “parametrization of topographical features” in the context of a groundwater model.

Response: We have modified the statements on coupling strategies accordingly. Please check it out at P2, L30-34.

9. Page 3, lines 3-5: Confusing arguments. What you really need to say here is that these models use pde's for both subsurface and surface flow, requiring a fine resolution and many parameters. There is also no “unsaturated groundwater”.

10. Page 3, line 6: replace “field-scale” by “hillslope-scale”

Response: Those comments are about tiny syntax errors. We have modified those statements accordingly.

11. Page 3, line 7: Warning! Both Reed Maxwell and Ed Sudicky sell their pde-based models with applications on the continental scale. Whether the parameters in these models have any physical meaning is a different story, but you can’t say that the pde-based models are restricted to small-scale applications. Sabine Attinger is co-author on a paper submitted to GMD using TerrSysMP on the scale of the state of Baden-Württemberg.

Response: We have deleted this sentence accordingly.

12. Page 3, line 9: Sorry, but it’s conceptual models that neglect the momentum balance altogether. Darcy’s law is derived from a momentum balance, and the shallow-water equations even more so.

Response: We have modified those statements accordingly.

13. Page 3, lines 23-24: Yes, but you can of course use more complex formulations for land-surface processes in pde-based models (see TerrSysMP coupling CLM to PARFLOW). The true difficulty is that, unless you resolve all relevant fine-scale features, you don't get quick flow components right with pde-based approaches. This is an inherent scaling problem. The conceptual hydrological models overcome this difficulty by introducing quick flow out of the blue, which is conceptually impossible in the pde-based approaches.

Response: We have modified those statements accordingly. Please check it out in the revised manuscript.

14. Page 3, line 30: Remove “therefore”

15. Page 3, lines 31-32: If the overall aim was to model regional-scale groundwater flow dynamics, you would choose a stand-alone groundwater model. So that cannot be the overall goal. What you really want is to extend a regional-scale conceptual
hydrological model to also predict hydraulic heads (and eventually groundwater transport and a proceeding step).

Response: We have modified those statements accordingly.

16. Page 4, line 25: eliminate “by applying the MPR methodology”.
17. Page 4, line 31: “we only listed the equations needed for the coupling”

Response: We have modified those statements accordingly.

18. All explanations related to equations: In the mHm part, the variable symbols are followed by units (e.g., “[mm]”), in the OGS part they are followed by dimensions (e.g., “[L]”). Either way is OK, but it should be consistent. (I’d prefer the dimensions over the units)

Response: We have modified those statements accordingly.

19. Page 5, line 11: beta_5 is not a rate, it’s a rate coefficient. And it has a dimension (of an inverse time, which explains why it is NOT a rate but a rate coefficient).

Response: We have modified those statements accordingly. Please check it out at P5, L5-10.

20. Page 5, line 14: same arguments for beta_6

Response: We have modified those statements accordingly.

21. Page 6, line 25: In which direction is z oriented? According to equation (10), where z is subtracted from the pressure head, it must be downwards. Please specify.

Response: We have modified those statements to “z is the vertical coordinate (positive upward) [L]”. Please check it out at P7, L15.

22. Page 6, lines 26 & 27: “the OGS source code” and “the OGS mesh”
23. Page 8, line 1: “OGS is used to simulate groundwater flow for prescribed fluxes at all boundaries” rather than “hydrological processes”
24. Page 8, line 14: “mHM is run independently of OGS to calculate land surface fluxes including exchange fluxes of the groundwater storage”
25. Page 8, line 21: “is transformed to distributed river discharges along the streams and spatially distributed exchange rates between streams and groundwater needed in OGS”
26. Equation 11: What happens if the stream nodes are irregularly spaced?
Response: Thank you for this observation. We have changed the Eq. 11 to the form of normalized flux such that the irregularly spaced nodes can be well treated. Please check it out at P9, L22-26.

27. Page 9, line 15: “For each node” rather than “each of the nodes”

Response: We have modified those statements accordingly.

28. Equations (12) and (13). This is still not clear. I understand equation (13), provided that $W_i(x)$ is the FEM weighting function. But what exactly is done in equation (12)? $W_j(x)$ seems to be a shape function of the mHM grid, but that is not stated. Actually, my understanding is that the index $j$ relates to an mHM grid cell and that $W_j(x)$ is unity within the mHM cell and zero elsewhere. But that is not written in the text accompanying the equation.

Response: The difference between Eq.12 and Eq.13 is that Eq.12 calculates the normalized flux (flux divided by surface area) and Eq.13 converted the normalized flux to volumetric flux through face integration. The recharge file obtained from mHM is a raster file, which means mHM does not have a shape function. Eq. 12 calculated the normalized recharge by weighting the nodal flux obtained from mHM using shape function. We use different indexing ($i$ and $j$) to distinguish local indexing system ($j$) from global indexing system ($i$). Indeed, $W_i(x)$ is identical to $W_j(x)$ as they are both shape functions of OGS surface mesh. To avoid possible confusion caused by indexing, we unify the indexing system in the revised manuscript by replacing $j$ with $i$. Please check it out at P10, L9-12.

29. Page 9, line 30: “the groundwater model simulates groundwater flow to obtain hydraulic heads. The groundwater model may also be used to compute travel times and solute transport within the groundwater domain, requiring additional boundary conditions; but this is not described in the present paper.”

30. Page 11, line 7: “3.2 Meteorological and surface properties”

31. Page 11, line 12: “kriged onto”

32. Page 11, lines 12ff.: “Potential ET was estimated by the method of Hargreaves and Samani (1985). Other datasets used in mHM are the digital elevation data ... soil and geological maps, and derived properties such as sand and clay contents, bulk density, CORINE land-cover information…” (geology appeared twice, I have no clue what meta-data you meant)

Response: Those comments are about syntax errors. We have modified those statements accordingly.

33. Page 11, section 3.3: It’s not “stratified aquifer model”, it’s “stratigraphic” (as you relate aquifer properties to stratigraphic units, which need not be stratified).
34. Section 3.6 and also 4.1, 4.3: See my remarks related to the Pearson correlation coefficient. Calculate and report RMSE or MAE instead.
Response: We have added the statistics MAE into the manuscript. Please have a look at the updated Figure 11, and the corresponding statements in Section 3.6, 4.1 and 4.3.

35. Page 14, line 25: “the wells” rather than “those wells”

36. Page 15, line 1: “which is due to the unknown local or eve sub-grid-scale properties”.

37. Page 16, line 3: “where the Muschelkalk crops out”

38. Page 16, line 9: Which figure is “the figure”

39. Page 20, line 16: The model was not calibrated in the transient mode. Thus, you did not calibrate the specific yield and the specific storativity, which largely determine the dynamic behavior of hydraulic head.

Response: We have modified those statements accordingly. Please check it out in the revised manuscript.

40. Page 20, lines 30-31: This again is the trap of arguing with something that you do not resolve at all. Your hydrological model does not consider feedbacks of groundwater to surface hydrology at all.

41. Page 22, line 16: Double “derived recharge rates”

42. Page 22, line 25: “capability” rather than “ability”

43. Page 22, line 29: “computational effort” rather than “consumption”

44. Page 22, line 34: “can explicitly be modeled” (word order)

Response: We have modified those statements accordingly. Please check it out in the revised manuscript.
Improved regional scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwater model OpenGeoSys (OGS)

Miao Jing\(^1\), Falk Heße\(^1\), Rohini Kumar\(^1\), Wenqing Wang\(^2\), Thomas Fischer\(^2\), Marc Walther\(^2,3\), Matthias Zink\(^1\), Alraune Zech\(^1\), Luis Samaniego\(^1\), Olaf Kolditz\(^2,4\), and Sabine Attinger\(^1,5\)

\(^1\)Department of Computational Hydrosystems, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
\(^2\)Department of Environmental Informatics, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
\(^3\)Institute of Groundwater Management, Technische Universität Dresden, Bergstr. 66, 01069 Dresden, Germany
\(^4\)Applied Environmental Systems Analysis, Technische Universität Dresden, Dresden, Germany
\(^5\)Institute of Earth and Environmental Sciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

Abstract.

Most large-scale hydrologic models fall short in reproducing groundwater head dynamics due to their over-simplified representation of groundwater flow. In this study, we aim to extend the applicability of the mesoscale Hydrologic Model (mHM v5.7) to subsurface hydrology by coupling it with the porous media simulator OpenGeoSys (OGS). The two models are one-way coupled through model interfaces GIS2FEM and RIV2FEM, by which the grid-based vertical fluxes of groundwater recharge and the river-groundwater exchange generated by mHM, are converted into upper-surface to fixed-flux boundary conditions of the groundwater model OGS. Specifically, the grid-based vertical reservoirs in mHM are completely preserved for the estimation of land-surface fluxes, while OGS acts as a plug-in to the original mHM modeling framework for groundwater flow and transport modeling. The applicability of the coupled model (mHM#OGS v1.0) is evaluated by a case study in the central European meso-scale river basin - Nägelstedt. Different time steps, i.e. daily in mHM and monthly in OGS, are used to account for fast surface flow and slow groundwater flow. Model calibration is conducted following a two-step procedure using discharge for mHM and long-term mean of groundwater head measurements for OGS, respectively. Based on the model summary statistics, namely the Nash–Sutcliffe model efficiency (NSE), the Pearson correlation coefficient, the mean absolute error (MAE), and the inter-quartile range error (QRE), the coupled model is able to satisfactorily represent the dynamics of discharge and groundwater heads at several locations across the study basin. Our exemplary calculations show that the one-way coupled model can take advantage of the spatially explicit modeling capabilities of surface and groundwater hydrologic models, and provide us with an adequate representation of the spatio-temporal behaviors of groundwater storage and heads, thus making it a valuable tool for addressing water resources and management problems.

1 Introduction

1.1 Background

...
Historically, large-scale hydrologic models had been primarily developed to predict river discharge. Most of these models use simple representations of underlying hydrological processes, usually bucket-type expressions combined with several vertical water storage layers to describe near-surface water flow to describe water storage and flow inside the catchment (Refsgaard and Storm, 1995; Wood et al., 1997; Koren et al., 2004; Samaniego et al., 2010; Niu et al., 2011). Moreover, due to the limitation in computational capability, all traditional hydrologic models simplify water flow processes by ignoring lateral groundwater flow, especially at a large scale. Thus, such models inevitably fall short of characterizing subsurface groundwater dynamics (Beven et al., 1984; Liang et al., 1994; Clark et al., 2015), where such lateral flows are dominant. In addition, estimates of groundwater storage and heads are particularly error prone due to the low sensitivity of groundwater (storage) to river flows.

The implicit groundwater representations in traditional hydrologic models are inadequate in many aspects. Depth to groundwater has a strong influence on near-surface water processes such as evapotranspiration (Chen and Hu, 2004; Yeh and Eltahir, 2005; Koirala et al., 2014). Moreover, fluctuations in the water table are known as a factor that affects runoff generation and thus their adequate representation in hydrologic models influences the prediction ability of catchment runoff (Liang et al., 2003; Chen and Hu, 2004; Koirala et al., 2014). Typical hydrologic models are also demonstrably inadequate at simulating groundwater representation in these conceptual (bucket-type) hydrologic models is consequently not adequate in several aspects. First, these models aggregate the heterogeneity of typically stratified groundwater aquifers, and fall short in adequately representing groundwater heads and low flow conditions (Ameli et al., 2016; Hale and McDonnell, 2016). Second, these models often do not properly capture the dynamics of solute transport and retention at the catchment scale. For example, Van Meter et al. (2017) found that current nitrogen fluxes in rivers can be dominated by groundwater legacies. An over-simplified groundwater representation is inadequate for understanding travel-time distributions (TTDs) at a catchment scale and is therefore incapable of describing such legacy behavior (Botter et al., 2010; Benettin et al., 2015, 2017). Moreover, stream subsurface water interactions may be significant in modulating the human and environmental effects of nitrogen pollution (Azizian et al., 2017). Finally, to assess the response of groundwater to climate change, a more accurate groundwater representation including lateral subsurface flow is urgently needed to predict the response of groundwater to climate change (Scibek and Allen, 2006; Green et al., 2011; Ferguson et al., 2016).

In parallel, numerous groundwater models have been developed, which allow for both steady-state and transient groundwater flow in three dimensions, with complex boundaries and a complex representation of strata, accounting for the representation of sub-surface heterogeneity and varying degree of sources and sinks. A variety of numerical codes are available such as MODFLOW (Harbaugh et al., 2000), FEFLOW (Diersch, 2013), and OpenGeoSys (Kolditz et al., 2012). Groundwater models usually represent subsurface flow by Darcy’s law substituted into the continuity equation, but fall short in providing good
representation of surface and shallow soil processes. For example, models for predicting groundwater storage changes under either climate change (e.g., global warming) or human-induced scenarios (e.g., agricultural pumping) often use a constant or linear expression to represent spatially distributed recharge (Danskin, 1999; Selle et al., 2013). Moreover, parameterization of topographical and geological parameters is a big challenge due to the strong spatial and temporal heterogeneity and lack of data (Moore and Doherty, 2006; Arnold et al., 2009).

In recent years, many integrated surface-groundwater models have been developed. ISSHMs commonly focus on the comprehensive treatment of both surface flow processes (e.g., 1-D or 2-D overland flow) and subsurface flow processes (e.g., 1-D or 3-D Richards flow) using a two-way coupling procedure (Paniconi and Putti, 2015). Some well-recognized ISSHMs are models are available such as: InHM (VanderKwaak and Loague, 2001; Smerdon et al., 2007), Parflow (Maxwell and Miller, 2005; Maxwell et al., 2015), OpenGeoSys (Delfs et al., 2012; Kolditz et al., 2012), tRIBS (Ivano et al., 2004), CATHY (Camporese et al., 2010), GSFLOW (Markstrom et al., 2008; Hunt et al., 2013), HydroGeoSphere (Therrien et al., 2010; Hwang et al., 2014), MIKE SHE (Graham and Butts, 2005), MODHMS (Panday and Huyakorn, 2004; Phi et al., 2013), GEOtop (Rigon et al., 2006), IRENE (Spanoudaki et al., 2009), CAST3M (Weill et al., 2009), PIHM (Kumar et al., 2009; Qu and Duffy, 2007), and PAWS (Shen and Phanikumar, 2010). Although the methods for subsurface flow in ISSHMs are commonly based on saturated/unsaturated groundwater flow equations, the approaches for surface flow are inevitably based on some approximations and conceptualizations. PDE-based hydrologic models usually represent subsurface flow by accounting for both saturated and unsaturated groundwater flows. The flow fields can be directly computed on the basis of spatial gradients of the modeled primary variable, e.g., the hydraulic head. The PDE-based models are flexible in coping with subsurface heterogeneity by means of proper characterization of aquifer system (e.g., diffusive-wave approximation, 1-D rill flow). The applications of these ISSHMs in the literature mainly focus on the field–and small watershed-scale, while assessments of modeled groundwater head dynamics at larger scales can only be found in very few publications (Goderniaux et al., 2009; Sutanudjaja et al., 2011). At this larger, stratification or geostatistical approach, and thus are able to reduce aggregation errors caused by geological heterogeneity (de Marsily et al., 2005; Cirpka and Attinger, 2003; Zech et al., 2009).

Furthermore, these models can explicitly compute flow pathlines and provide direct estimates of travel times of water and solute particles. These properties of PDE-based models provide a significant advantage over (bucket-type) conceptual models especially in complex real-world applications (Park et al., 2008; Engdahl and Maxwell, 2015; Danesh-Yazdi et al., 2018).

However, despite these advantages in modeling the deeper subsurface flows, PDE-based models are not without problems, in particular in capturing the near-surface flow dynamics i.e., regional, scale most of the ISSHMs are based on a continuity of pressure and flux on the surface water/groundwater (SW/GW) interface, while the momentum balance condition is always neglected (Paniconi and Putti, 2015). ISSHMs relying on shallow water and the Richards equation assumptions in shallow portions of the subsurface. For example, the PDE-based models often encounter problems in simulating quick flow dependent on essentially unknown sub-grid scale topographic variability and subsurface structure (Paniconi and Putti, 2015). Nevertheless, these models are capable of simulating the dynamic interaction of different processes within SW/GW components (the unsaturated zone for simulating the quick flow components, which are mainly dependent on subgrid heterogeneities of topographic variation...
as well as soil and land-cover characteristics (Panconi and Putti, 2015). Using a complex PDE-based surface hydrologic model to simulate near-surface processes is possible in general, but it does require a model implementation at a fine spatial resolution to resolve the sub-grid features (e.g., the interaction of soil moisture and groundwater head (Rihani et al., 2010; Cuthbert et al., 2013; Sutanudjaja et al., 2014; Maxwell et al., 2015) as well as the storage-runoff correlation (VanderKwaak and Loague, 2001; Liang et al., 2003; Huntington and Niswonger, 2012; Koirala et al., 2014) and include a tremendous number of uncertain model parameters. Furthermore these models have dramatically increased numerical complexity and computation time, and thus calibrating these models is a cumbersome task, and doing this in a stochastic framework is computationally not feasible.

To summarize these points, bucket-type hydrologic models, such as mHM (Samaniego et al., 2010; Kumar et al., 2013b), VIC (Liang et al., 1994), and HBV (Lindström et al., 1997), are good at predicting quantities such as discharge but, as mentioned above, are highly conceptual and their model results are difficult to interpret with respect to certain processes (e.g., groundwater storage and heads). The output of more mechanistic ISSHMs—outputs of PDE-based hydrologic models, such as Parflow, CATHY, and HydroGeoSphere, are highly interpretable but show consistently worse performance than typical hydrologic models when predicting runoff discharge dynamics (Gulden et al., 2007; Panconi and Putti, 2015). The differing capabilities of these models result in the more mechanistic ISSHMs are a two types of models are the result of the different challenges that are posed by the various compartments of the terrestrial water cycle. One of the main challenges in modeling surface and near-surface storage is process uncertainty, since processes such as evapotranspiration, land use, land cover, snow pack are extremely complex and dynamic. The process uncertainty decreases as one goes deeper into subsurface storage. In is caused by the strong non-linearities of hydrological processes and the fine-scale variability of land-surface features. Thus it can be hardly solved by PDE-based models, but can be well handled by bucket-type models through the parameterization process. In the deeper subsurface storage, hydrological processes are relatively well understood and therefore conceptually simpler the temporal and spatial scale of groundwater process is significantly larger than the shallow storage, and the flow is governed by linear PDE (Darcy’s law), thus makes the PDE-based model standard at a large scale (Dagan, 2012). Meanwhile, the data uncertainty becomes more significant in the deep subsurface storage in comparison to shallow storage due to the spatially sparse hydrogeological data. Moreover, a recent study reveals the strong spatial and temporal heterogeneity of processes and properties at the surface/groundwater (SW/GW) interface, and underlines the importance of quantifying variability across several scales at the SW/GW interface and its significance to water resources management (McLachlan et al., 2017). It therefore stands to reason, that the use of a hybrid of both these model frameworks is a good choice for a joint representation of surface and subsurface water storages and fluxes.

Therefore, the coupling between the bucket-type land-surface model and the subsurface saturated/unsaturated groundwater model is highly relevant. Several well-tested coupled models have been developed in recent years, including ParFlow-CLM (Maxwell and Miller, 2005; Maxwell et al., 2015; Kurtz et al., 2016), GSFLOW (Markstrom et al., 2008; Hunt et al., 2013), PCR-GLOBWB (Sutanudjaja et al., 2011, 2014), and CP(v1.0) (Bisht et al., 2017).

In this study, we therefore coupled the mesoscale Hydrologic Model (mHM v5.7) (Samaniego et al., 2010; Kumar et al., 2013b) with the porous media simulator OpenGeoSys (OGS)
with the overall aim of modelling regional-scale groundwater flow dynamics. The overall aim is to provide a proper representation of groundwater flows and storages and enabling the coupled model to provide reliable estimates of groundwater heads. mHM has demonstrated its pre-eminence in coping with process uncertainty in the near-surface zone while providing excellent discharge prediction (Huang et al., 2017) near-surface process uncertainty while providing reliable representation of observed discharge behavior across range of scales and locations (Samaniego et al., 2010; Kumar et al., 2013b; Huang et al., 2012; Selle et al., 2013). Using these The general idea behind the coupling is to use the hydrological simulation of with mHM, including the simplified linear groundwater storage, extract fluxes into and out of groundwater from mHM, and use this as Neumann boundary condition for the PDE-based groundwater model OGS. By doing so, we augment the predictive power of mHM to also predict hydraulic heads in groundwater. The one-way coupling approach considered here has a number of advantages. First, the one-way coupling can be regarded as a conservative approach, such that the parametrization process, which is one of the most significant features of mHM, remains fully intact. In particular, this means that the whole body of confidence in the predictive power of mHM, that has been build up over the years can be fully relied on. Second, using such a one-way coupling will allow users of mHM to simply extend currently established catchment models and enhance their abilities in the aforementioned way. Using a more sophisticated two-way coupling, would entail users having to rebuild their models almost entirely. Third, a one-way coupling allows for ready future expansion of the functionality of the coupled model, e.g., legacy of solutes in groundwater, should the need arises. Finally, one-way coupling takes less computational efforts and achieves better numerical stability than two-way coupling.

By coupling two well-tested model codes, we want to answer the following scientific questions: (1) Can spatially distributed groundwater heads and their dynamics be reasonably captured by expanding the capabilities of a surface hydrologic model, such as mHM at the regional-scale, while conserving its excellence in predicting discharge? (2) Can spatially resolved groundwater recharge estimates, provided by mHM, improve the prediction of head measurements of groundwater models such as OGS? To answer these questions, we applied the coupled model mHM#OGS v1.0 in a central German meso-scale catchment (850 km²), and evaluated the model skills using measurements of streamflow and groundwater heads from several wells located in the study area. The coupled (surface) hydrologic and groundwater model (mHM#OGS v1.0) presented in this paper is our first attempt toward the development of a large-scale coupled modeling system with the aim to analyze the spatio-temporal variability of groundwater flow dynamics at a regional scale.

The paper is structured as follows. In the next section, we describe the model concept, model structure, and the coupling scheme. In Section 3.1, the study area and model setup used for illustration in this study are comprehensively described. In Section 4, we present the simulation results of mHM#OGS v1.0 in a catchment in the application. In the Section 5, we discuss the model results as well as advantages and limitations of current modeling approach.
2 Model description

2.1 mesoscale Hydrologic Model (mHM)

The mesoscale Hydrologic Model (mHM, www.ufz.de/mhm) is a spatially explicit distributed hydrologic model that uses grid cells as a primary modeling unit, and accounts for the following processes: canopy interception, snow accumulation and melting, soil moisture dynamics, infiltration and surface runoff, evapotranspiration, subsurface storage and discharge generation, deep percolation, baseflow, discharge attenuation, as well as flood routing (Figure 1). The runoff generation applies a robust scheme which routes runoff in upstream cells along river networks using the Muskingum-Cunge algorithm. The model is driven by daily meteorological forcings (e.g., precipitation, temperature), and utilizes observable physical properties or signals of the basin (e.g., soil textural, vegetation, and geological properties) to infer the spatial variability of the required parameters. mHM is an open-source project written in Fortran 2008. Parallel versions of mHM using OpenMP concepts are available.

A unique feature of mHM is the application of Multiscale Parameter Regionalization (MPR). The MPR method accounts for subgrid variability in physical characteristics of the catchment such as topography, soil and vegetation. The MPR methodology facilitates the flexibility of the model for hydrological simulations at various spatial scales by applying the MPR methodology (Samaniego et al., 2010; Kumar et al., 2013a, b; Rakovec et al., 2016a, b; Samaniego et al., 2017). mHM differentiates three levels to better represent the spatial variability of state and input variables. The effective parameters at different spatial scales are dynamically linked by a physically-based upscaling scheme. A detailed description of MPR, as well as the formulations governing hydrological processes, are given by Samaniego et al. (2010) and Kumar et al. (2013b).

Below, we list the equations that describe near-surface processes in the deep soil and groundwater layers. The comprehensive system of equations of mHM can be found in Samaniego et al. (2010). Here, we only listed the equations needed for the coupling to OpenGeoSys. In the subsurface reservoir, which is the second vertical layer ($x_5$ in Figure 1), interflow is partitioned.
into fast interflow \(q_2\) and slow interflow \(q_3\):

\[
q_2(t) = \max\{I(t) + x_5(t - 1) - \beta_1(z_2 - z_1), 0\} \beta_2 \\
q_3(t) = \beta_3(x_5(t - 1))^{\beta_4}
\]

where \(q_2(t)\) is fast interflow at time \(t\) \[\text{mm-d\-L}^{-1}\], \(I\) is the infiltration capacity \[\text{mm}\], \(x_5\) is the water depth of water storage in the deep soil reservoir \[\text{mm}\], \(\beta_1\) is the maximum holding capacity of the deep soil reservoir [-], \(z_i\) is depth of subsurface layer \(i\) [L], \(\beta_2\) is the fast recession constant [\text{fast-recession constant \text{T}^{-1}}], \(q_3(t)\) is slow interflow at time \(t\) \[\text{mm-d\-L}^{-1}\], \(\beta_3\) is the slow concession constant, and \(\beta_4\) is the exponent that quantifies the degree of non-linearity of the cell response.

The groundwater recharge is equivalent to the percolation to the groundwater reservoir (the third vertical layer, see \(x_6\) in Figure 1). The groundwater recharge \(C(t)\) can be expressed by

\[
C(t) = \beta_5 x_5(t - 1)
\]

where \(C(t)\) is the groundwater recharge in cell \(i\) \[\text{mm-d\-L}^{-1}\], and \(\beta_5\) is the effective percolation rate coefficient [\text{T}^{-1}].

In the groundwater reservoir, baseflow is generated following a linear relationship between storage and runoff:

\[
q_4(t) = \beta_6 x_6(t - 1)
\]

where \(q_4(t)\) is the baseflow \[\text{mm-d\-L}^{-1}\], \(\beta_6\) is the baseflow recession rate coefficient [\text{T}^{-1}], and \(x_6\) is depth of groundwater reservoir \[\text{mm}\].

The runoff from upstream grid cells and the internal runoff in cell \(i\) are routed into streams using the Muskingum algorithm:

\[
Q^1_i(t) = Q^1_i(t - 1) + c_1(Q^0_i(t - 1) - Q^1_i(t - 1)) + c_2(Q^0_i(t) - Q^0_i(t - 1))
\]

with

\[
Q^0_i(t) = Q_{i'}(t) + Q^1_i(t) \\
c_1 = \frac{\Delta t}{\kappa(1 - \xi) + \frac{\Delta t}{2}} \\
c_2 = \frac{\Delta t}{\kappa(1 - \xi) + \frac{\Delta t}{2}}
\]

where \(Q^0_i\) and \(Q^1_i\) denote the runoff entering and leaving the river reach located in cell \(i\), respectively \[\text{mm-d\-L}^{-1}\], \(Q_{i'}\) is the contribution from the upstream cell \(i'\) \[\text{mm-d\-L}^{-1}\], \(\kappa\) is the Muskingum travel time parameter [T], \(\xi\) is the Muskingum attenuation parameter [-], \(\Delta t\) is the time step-size [h\text{T}], and \(t\) is the time index for each \(\Delta t\) interval.

### 2.2 OpenGeoSys (OGS)

OpenGeoSys (OGS) is an open-source project with the aim of developing robust numerical methods for the simulation of Thermo-Hydro-Mechanical-Chemical (THMC) processes in porous and fractured media. OGS is written in C++ with a focus
on the finite element analysis of coupled multi-field problems. Parallel versions of OGS based on both MPI and OpenMP concepts are available (Wang et al., 2009; Kolditz et al., 2012; Wang et al., 2017). To date, two OGS versions are available: OGS5 (https://github.com/ufz/ogs5) and OGS6 (https://github.com/ufz/ogs). In this study, the term “OpenGeoSys (OGS)” represents OGS5 if not stated otherwise.

OGS has been successfully applied in different fields, such as water resources management, hydrology, geothermal energy, energy storage, CO₂ storage, and waste disposal (Kolditz et al., 2012; Shao et al., 2013; Gräbe et al., 2013; Wang et al., 2017). In the field of hydrology / hydrogeology, OGS has been applied to regional groundwater flow and transport (Sun et al., 2011; Selle et al., 2013), contaminant hydrology (Beyer et al., 2006; Walther et al., 2014), reactive transport (Shao et al., 2009; He et al., 2015), and sea water intrusion (Walther et al., 2012), among others.

Saturated groundwater flow follows the continuity equation and Darcy’s law:

\[ S \frac{\partial \psi_p}{\partial t} = -\nabla \cdot \mathbf{q} + q_s \]  

\[ \mathbf{q} = -K_s \nabla (\psi_p + z) \]

where \( S \) is specific storage coefficient in confined aquifers, and the specific yield in unconfined aquifers \([1/L]\), \( \psi_p \) is the pressure head in the porous medium \([L]\), \( t \) is time \([T]\), \( \mathbf{q} \) is the specific discharge or Darcy velocity \([LT^{-1}]\), \( q_s \) is the volumetric general source/sink term \([T^{-1}]\), \( K_s \) is the saturated hydraulic conductivity tensor \([LT^{-1}]\), and \( z \) is the vertical coordinate (positive upward) \([L]\).

The stream network is normally represented by a set of polylines in the geometry file of OGS. In the case of a 3-D model, a common way to set up the polyline system is to utilize the mapping tool embedded in OGS source codes, by which the shape file obtained from GIS software representing streams can be easily mapped onto the upper surface of OGS mesh and converted into a set of polylines. Each reach of the stream network can be represented by one polyline or several continuous polylines, depending on the demand of the user. Each polyline consists of a set of continuous mesh nodes, upon which Dirichlet, Neumann or Robin boundary conditions can be applied.

### 2.3 Coupling mechanism

The coupled model mHM#OGS v1.0 is developed to simulate SW/GW flow in one or more catchments by simultaneously calculating flow across the land surface and within the groundwater. mHM#OGS v1.0 simulates flow within three hydrological regions. The first region is limited by the upper boundary of the plant canopy and the lower boundary of the soil zone bottom. The second region includes open-channel water, such as streams. The third region is the water-saturated aquifer. mHM is used to simulate the processes in the first and second regions, while OGS is used to simulate the hydrological processes simulate groundwater flow for prescribed fluxes at all boundaries in the third region.

The coupling initiative aims to add additional predictive capability of groundwater heads, which is achieved by OGS, to the existing capability of predicting discharges that is achieved by mHM. mHM is used to estimate step-wise and component-wise a water budget through model calibration against discharge. In contrast, OGS serves as a post-processor to obtain groundwater heads by using mHM simulated recharge and baseflow as driving forces. Two model interfaces, namely GIS2FEM and
Figure 2. Schematic of the coupled model mHM#OGS v1.0. a) Original structure of the vertically layered reservoir of mHM. b) Structure of the coupled model (mHM#OGS v1.0). c) Illustration of data interpolation and transformation through the coupling interface GIS2FEM. d) Scheme of the river interface RIV2FEM. For the sake of simplicity, the figure only displays mHM layers relevant to this study and neglects the other mHM layers (i.e. $x_1 - x_4$). In Figure 2c, the grid-based mHM fluxes (e.g., recharge) are linearly interpolated to the top surface of the OGS mesh, and further transferred into volumetric values and directly assigned to the surface mesh nodes of the OGS grid.

RIV2FEM, have been developed to link the two models by transferring recharge and baseflow from mHM to Neumann boundary conditions in OGS. The two models are executed separately and sequentially, typically with different temporal (e.g., daily in mHM and weekly or monthly in OGS) and spatial resolutions (e.g., rectangular, structured grids with coarse resolution in mHM and smaller, potentially unstructured grids with fine resolution in OGS). The original vertically layered reservoirs in mHM, namely the soil-zone reservoir and the subsurface reservoir are preserved, implying that all well-tested features of mHM (e.g., MPR, infiltration-runoff partitioning) are retained in the coupled model.

To illustrate the coupling mechanism in detail, we itemized the coupling workflow below.
1. mHM is run independently of OGS to calculate land surface fluxes including exchange fluxes of the groundwater storage. Using gridded meteorological forcings (precipitation, temperature, and potential evapotranspiration), the grid-based infiltration rates (e.g., groundwater recharge) and runoff components (e.g., interflow, baseflow) are estimated and saved as mHM output files. The original linear groundwater reservoir (depth $x_0$ in Figure 1) is used to estimate baseflow.

Moreover, MPR is used in the calibration process such that subgrid variabilities can be validly calculated. The spatially distributed groundwater recharge and total routed baseflow are written into raster files for later use.

2. After the mHM run has finished, the step-wise routed baseflow estimated by mHM is transformed to distributed river discharges along streams as represented and spatially distributed exchange rates between streams and groundwater needed in OGS.

Most physically-based ISSHMs PDE-based models characterize river-groundwater interaction based on either first-order flow exchange or boundary condition switching (Paniconi and Putti, 2015). However, this approach inevitably relies on a parameter set—these approaches inevitably introduce extra parameters describing geometric, topographic, and hydraulic properties of the stream channel (e.g., river bed conductance, river bed and drain elevations, channel width). Unfortunately, these parameters are essentially unknown at a large scale due to the lack of data and the subgrid-scale variability of these parameters. Due to these limitations, we use an alternative approach which is based on the routed baseflow estimated by mHM.

mHM and OGS conceptualize streams differently: streams in mHM are implicitly defined based on pre-processing of digital elevation model (DEM) data and a routing scheme, while OGS uses an explicit predefined river geometry. In OGS, each reach of the stream network is defined by a polyline in the OGS geometry file. To coordinate the two different approaches, we developed a model interface, RIV2FEM, to convert the routed baseflow estimated by mHM to Neumann boundary conditions assigned at stream nodes of the OGS mesh (Figure 2d). Via RIV2FEM, the routed baseflow estimated by mHM is transferred to the uniformly disaggregated discharges by distributing it uniformly along the predefined stream network in OGS (Figure 2d):

$$q_i(t) = \frac{Q_4(t)}{N} \sum_{i=1}^{N} A_i$$

(11)

where $q_i(t)$ denotes the disaggregated discharge assigned at every stream node in OGS at normalized flux of disaggregated discharge at time $t$ \([L^3T^{-1}L^{-1}]\), $Q_4(t)$ denotes the routed baseflow at the outlet of catchment at time $t$ \([L^3T^{-1}]\), $A_i$ is the nodal area of the $i$th stream node \([L^2]\), $N$ denotes the total number of stream nodes in OGS. The uniformly disaggregated discharges are then assigned to every stream node in OGS to serve as the Neumann boundary condition (Figure 2d). This approach significantly reduces the number of parameters, avoids the uncertainty caused by the unknown river properties, and is suitable for many real-world applications that suffer from scarce data. Moreover, as recharge and baseflow are directly taken from mHM, the mass conservation criterion is naturally satisfied in this approach.
3. The distributed groundwater recharge generated by mHM is fed to the coupling interface GIS2FEM, and then transferred to the upper surface boundary conditions of the OGS model.

The coupling interface GIS2FEM is used to interpolate and transfer mHM grid-based recharge to OGS nodal recharge values. GIS2FEM interpolates the flux value to the top surface elements of the OGS mesh. The detailed workflow is:

- GIS2FEM reads the raster file generated by mHM and the mesh file of OGS.
- In the case of a 3-D mesh, GIS2FEM extracts the upper surface of the OGS mesh. For each of the nodes on this surface, GIS2FEM searches for the mHM grid cell that the node is located in, and assigns the recharge value of this grid cell to the corresponding node (marked as $C^m$).
- After all top surface elements have been processed, GIS2FEM undertakes the face integration calculation, by which the specific recharge $C^m$ [LT$^{-1}$] calculated by mHM is converted into volumetric recharge $C^{in}$ [L$^3$T$^{-1}$] and assigned to the corresponding OGS mesh nodes (Figure 2c). Specifically, the specific recharge $C$ in a certain element is calculated as:

$$C(x) = \sum_{j=1}^{N} W_{j}(x)C^m_{j},$$  \hspace{1cm} (12)

where $x$ is the spatial coordinate on the surface, $N$ is the total number of nodes in a surface element, $W_j-W_i$ is the weighting function of node $j$, $C^m_{j}$ is the specific recharge at node $j$th node of OGS surface element, $C^m_{j}$ is the specific recharge at node $j$th node of OGS surface element (calculated by mHM) [LT$^{-1}$]. Then the volumetric recharge $C^{in}_i$ at node $i$th node (here $i$ is the global node index) is calculated by the face integration calculation:

$$C^{in}_i = -\int_{\partial\Omega} W_i(x)C(x)d(x),$$  \hspace{1cm} (13)

where $C^{in}_i$ is the volumetric recharge of node $i$ [L$^3$T$^{-1}$], $\partial\Omega$ is the surface boundary of the FEM domain, $W_i$ is the weighting function of node $i$th node.

4. After the mHM-generated recharge and baseflow have been transferred into boundary conditions at the upper surface of the OGS mesh, the groundwater model is run to simulate the groundwater flow and transport.

In this step, additional boundary conditions can be set up in OGS mesh on the basis of expert knowledge. The exclusive use of Neumann BC is not recommended and may lead to nonuniqueness of solutions. At least one specified head boundary should be set at the perimeter or internal nodes to constrain the model solution. The groundwater model simulates groundwater flow to obtain hydraulic heads in the example application. The groundwater model may also be used to compute travel times and solute transport within the groundwater domain, requiring additional boundary conditions; but this is not described in the present paper.
3 Example application

3.1 Study area and model setup

We use a meso-scale catchment (about 850 km$^2$) upstream of the Nägelstedt gauge located in central Germany to test our coupled model (Figure 3). The Nägelstedt catchment comprises the headwaters of the Unstrut River, a tributary of the River Saale. We selected this study area because many of the groundwater monitoring wells in the area are operated by the Thuringian State Office for the Environment and Geology (TLUG) and the Collaborative Research Center AquaDiva (Küsel et al., 2016). The elevation within the catchment ranges between 164 m and 516 m, whereby the higher regions are in the west and south and belong to the forested hill chain of the Hainich (Figure 3). The Nägelstedt catchment is one of the most intensively used agricultural regions in Germany. In terms of drinking water supply, about 70% of the water requirement is satisfied by groundwater (Wechsung, 2005). About 17% of the land in this region is forested area, 78% is covered by crop and grassland, and 4% is urban and transport area. The mean annual precipitation in this area is about 660 mm.
In this study, mHM runs were executed for a time period of 35 years (from January 1, 1970 to December 30, 2004), with the period 1970 - 1974 being used for spin-up. OGS was run for the period from January 1, 1975 to December 30, 2005. mHM was run with a daily time step, while OGS was run with a monthly time step. The resolution of mHM grid cells is 500 m × 500 m. OGS uses a structured, hexahedral 3-D mesh, with a spatial resolution of 250 m × 250 m in the horizontal direction and 10 m in the vertical direction over the whole domain. The detailed input data and parameter-set to run both models are detailed in the following sections.

3.2 Meteorological forcings and morphological surface properties

We started the modeling by performing the daily simulation of mHM to calculate near-surface hydrological processes. The mHM model is forced by daily meteorological conditions, including distributed precipitation and atmospheric temperature. The spatial patterns of precipitation and atmospheric temperature were based on point measurements of precipitation and atmospheric temperature at weather stations from the German Meteorological Service (DWD). The point data at weather stations were subsequently kriged into a 4 km precipitation field, and then downscaled to mHM grid cells. Moreover, the potential ET was quantified based on the method from Hargreaves and Samani (1985). Other datasets used in mHM are the DEM data, which is the basis for deriving properties such as slope, river beds, and flow direction; soil and geological maps, and derived properties such as sand and clay contents, bulk density, and dominant geological types; CORINE land-cover information (in the years 1990, 2000 and 2005); and discharge data at the outlet of the catchment.

3.3 Aquifer properties

We used a stratified aquifer stratigraphic model to explicitly represent the heterogeneous distribution of hydraulic properties (hydraulic conductivity, specific yield, and specific storage). The stratified aquifer stratigraphic model is based on well log data and geophysical data obtained from the Thuringian State Office for the Environment and Geology (TLUG). We used the workflow developed by Fischer et al. (2015) to convert the data format, by which the complex 3-D geological model was converted into the open-source VTK format file that can be directly read by OGS.

The major stratigraphic units in the study site are the Muschelkalk (Middle Triassic) and the Keuper (Upper Triassic). Younger Tertiary and Quaternary deposits are less important for the large-scale hydrogeology of the basin. The Keuper deposits mainly lie in the center of the Unstrut Basin and act as permeable shallow aquifers. In the Nägelstedt catchment, the Keuper deposits are further subdivided into two geological sub-units: Middle Keuper (km) and Lower Keuper (ku) (see Figure 4). The Muschelkalk is marked by a prevailing marine environment and is subdivided into three sub-units the Upper Muschelkalk (mo), Middle Muschelkalk (mm, dolomites and residues of eroded salt layers) and Lower Muschelkalk (mu, limestones). According to previous geological surveys (Seidel, 2004), the sub-units of the Muschelkalk have varying hydraulic properties depending on their positions and depths. They are further divided into sub-units with higher permeabilities (mo1, mm1 and mu1) and sub-units with lower permeabilities (mo2, mm2 and mu2) (Figure 4). The Upper Muschelkalk (mo) has been widely considered as a karstified formation. Recent research by Kohlhepp et al. (2017) has revealed that in the Hainich Critical Zone, the intense karstification and the conduit are limited at the base of the mo formation. Accordingly, we use the equivalent porous medium
Figure 4. Three-dimensional and cross-sectional views of the hydrogeologic zonation in the Nägelstedt catchment. The upper left-hand figure highlights the distribution of alluvium and soil zones. The upper right-hand figure shows two vertical geological cross-sections. The lower map shows the detailed zonation of geological sub-units beneath the soil zone and alluvium.

approach to characterize the Upper Muschelkalk. The uppermost layer with a depth of 10 m is set as a soil layer (Figure 4). A high-permeability alluvium layer is set along the mainstream and major tributaries to represent granite and stream deposits (Figure 4).

3.4 Boundary conditions

Based on the steep topography along the watershed divides, groundwater is assumed to be naturally separated and unable to pass across the boundaries of the watershed. In general, no-flow boundaries are set at the outer perimeters surrounding the basin as well as at the lower aquitard, except for the northwestern and northeastern edges. On the basis of the measurements, a Dirichlet boundary condition is assumed at the northwestern and northeastern edges.

The stream network was delineated by processing a grid-based runoff raster file generated by mHM. The grid-based runoff was converted to a valid stream network compatible with OGS. The necessity of transferring the mHM runoff raster file to the OGS stream network has been elaborated in Section 2.3. Particularly in this case study, we removed the small intermittent
Figure 5. Illustration of the stream network used in this study. a) Original stream network based on the streamflow routing algorithm of mHM; b) Processed stream network that was used in this study. The small tributaries where the runoff rates are below the threshold (0.145 m$^3$/s) as shown in the left figure, have been removed from to form the original stream network right figure.

Tributaries by setting a threshold value of long-term averaged routed runoff. Only streams with a runoff rate higher than the threshold (in this case study, 0.145 m$^3$/s) are delineated as valid streams. In other words, we neglected the intermittent streams to the upper stream reaches (Figure 5). The preprocessed stream network consists of a main stream and four tributaries (Figure 5b). The reach of each stream is defined as a polyline in a geometry file. As illustrated in Section 2.3, uniformly disaggregated groundwater discharges processed by the interface RIV2FEM were assigned to every OGS mesh node within the stream network.

3.5 Calibration procedure

The calibration of the coupled model follows a two-step procedure. In the first step, mHM was calibrated independent of OGS for the period from 1970 to 2005 by matching the observed runoff at the outlet of the catchment. The first 5 years were used as spin-up period to set up initial conditions in the near-surface soil zone. The calibration quality is quantified by the Nash-Sutcliffe coefficient of efficiency (NSE):

$$NS E = 1 - \frac{\sum_{i=1}^{n} |(q_m - q_s)|^2}{\sum_{i=1}^{n} |(q_m - \bar{q}_m)|^2} $$

(14)

where $q_s$ is the simulated discharge [L$^3$T$^{-1}$], $q_m$ is the measured discharge [L$^3$T$^{-1}$], and $\bar{q}_m$ is the mean of measured discharge [L$^3$T$^{-1}$].

In the second step, the steady-state groundwater model in OGS was calibrated to match the long-term mean of observed groundwater levels. The long-term mean of recharge and baseflow estimated by mHM were fed to the steady-state groundwater model as Neumann boundary conditions. The calibration was performed using the software package PEST (Doherty et al., 1994). The model parameters were adjusted within a fixed interval until the value of objective function, which is the sum of weighted squared residuals of modeled and observed groundwater heads, was minimized. Specifically, the intervals of adjustable parameters were taken from the literature (Wechsung, 2005; Seidel, 2004), and the weights assigned to each observation were set uniformly to 1. The calibration result is assessed by the root-mean-square error (RMSE).
3.6 Model evaluation and sensitivity analysis

We used the time series of groundwater levels in 19 monitoring wells to evaluate the predictive capability of the transient model. In the transient model, hydraulic conductivities are obtained from the calibrated steady-state model. Meanwhile, the initial condition of the groundwater head is directly taken from the result of steady-state model. The Pearson correlation coefficient \( R_{\text{cor}} \) and the inter-quartile range error \( QRE \) are used as two summary statistics to evaluate the predictive capability.

The (relative) inter-quartile range error \( QRE \) is defined by:

\[
QRE = \frac{IQ_{\text{md}}^{7525} - IQ_{\text{dt}}^{7525}}{IQ_{\text{dt}}^{7525}}
\]  

where \( IQ_{\text{md}}^{7525} \) and \( IQ_{\text{dt}}^{7525} \) are the inter-quartile ranges of simulations and observations, respectively.

We sought to quantify the sensitivity of groundwater flows to the different spatial pattern of recharge. For this purpose, a uniform recharge scenario was established as the reference scenario. The sensitivity analysis follows a two-step workflow. First, we calibrated the steady-state groundwater models for the two recharge scenarios independently. Second, we conducted transient simulations by assigning the same values of storage parameters, and then observed their corresponding performances in two recharge scenarios. With the exception of recharge scenario and hydraulic-conductivity values, all model parameters (e.g., specific yield and specific storage) and inputs are set to be identical in both scenarios. The Pearson correlation coefficient \( R_{\text{cor}} \), mean absolute error (MAE), and the inter-quartile range error \( QRE \) are used as two summary skill scores to assess model performances in the two recharge scenarios.

4 Results

4.1 Calibration

As the first part of calibration, mHM is calibrated against discharge. The calibration results demonstrate the predictive capability of mHM in reproducing the time series of catchment discharge (Figure 6). The Nash–Sutcliffe model efficiency coefficient (NSE) is 0.88, while the Pearson correlation coefficient is 0.96. Other fluxes, such as evapotranspiration measured at eddy-covariance stations inside this area, also shows quite reasonable correspondence to the modeled estimate (Heße et al., 2017).

In the second step, the steady-state groundwater model is calibrated against the long-term mean of groundwater heads. Table 1 shows the calibrated hydraulic conductivities in each of the geological units. The objective function of calibration, which is the sum of squared weighted residuals, converged from an initial value of 8625 m\(^2\) to 464.74 m\(^2\) after a total of 114 model runs. Broadly speaking, the steady-state model can plausibly reproduce the finite numbers of observed groundwater heads in the catchment. Figure 7 shows the one-to-one plot of simulated and observed groundwater heads (locations of those wells are shown in Figure 3). In general, the model is capable of reproducing spatially-distributed groundwater heads over a wide range, with an overall RMSE of 6.45 m. Most of the discrepancies between individual observations and simulations are within a reasonable range (i.e., less than 6 m). Nevertheless, some monitoring wells show larger discrepancies between observations and simulations (i.e., greater than 6 m), which is due to the unknown local geological properties and or even subgrid-scale
topographies properties. For the sake of simplicity, no further attempt was made to add more model complexity to improve the model fit.

The simulated depth to groundwater over the whole catchment using the calibrated hydraulic-conductivity values is shown in Figure 7c. Broadly speaking, the calibrated model reasonably reproduces the spatial groundwater table distribution. Groundwater depth varies between greater than 40 m in the higher southwestern and northern mountainous areas, to less than 5 m in the central lowlands. The plausibility of steady-state simulation results can be assessed through regionalized observations of groundwater heads (Wechsung, 2005).

4.2 Spatio-temporal patterns of recharge and baseflow

Groundwater recharge has a spatially variable and dynamic behavior depending on the sporadic, irregular, and complex features of precipitation, geological structure, and morphological features. The temporal and spatial variability of groundwater recharge and baseflow is estimated by mHM over a period of 30 years from 1975 to 2005.
Figure 8 shows the spatial variability of groundwater recharge in three months: March (Figure 8a), May (Figure 8b), and January (Figure 8c). The results indicate that the location of highest recharge rate is in the upstream mountainous areas where the Muschelkalk aquifer outcrops, but varies in different seasons. The maximum value of monthly groundwater recharge varies from 26 mm in March, to 51 mm in May and 14 mm January. We also evaluated the plausibility of groundwater recharge simulated by mHM through comparison to other reference datasets. At the large-scale, the groundwater recharge simulated by mHM agrees quite well with estimates from the Hydrological Atlas of Germany (Zink et al., 2017).

Figure 9a shows the distribution of monthly groundwater recharge and monthly baseflow. Over the entire year, groundwater inflow (recharge) and outflow (baseflow) are balanced, exhibiting a mean value of 8 mm/month. The difference between the two values is merely 2%. The figure, however, indicates that the distribution of monthly groundwater recharge is skewed to the right, whereas the distribution of monthly baseflow is more peaked. Figure 9b depicts the time series of groundwater recharge and baseflow, which further demonstrates that the deviation of monthly groundwater recharge is larger than the baseflow. This phenomenon further reveals the significant buffering effect of the linear groundwater storage in mHM.
Table 1. Main hydraulic properties used in the case study under the default mHM-generated recharge scenario.

<table>
<thead>
<tr>
<th>Geological units</th>
<th>Hydraulic conductivity [m/s]</th>
<th>Specific yield [-]</th>
<th>Specific storage ([\text{m}^{-1}])</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower limit</td>
<td>Upper limit</td>
<td>Calibrated value [m/s]</td>
</tr>
<tr>
<td>km</td>
<td>1.0 × 10⁻⁶</td>
<td>5.5 × 10⁻³</td>
<td>1.844 × 10⁻⁵</td>
</tr>
<tr>
<td>ku</td>
<td>1.0 × 10⁻⁷</td>
<td>3.4 × 10⁻⁴</td>
<td>2.848 × 10⁻⁵</td>
</tr>
<tr>
<td>mo1</td>
<td>8.0 × 10⁻⁸</td>
<td>2.0 × 10⁻³</td>
<td>3.570 × 10⁻⁵</td>
</tr>
<tr>
<td>mm1</td>
<td>1.0 × 10⁻⁷</td>
<td>9.0 × 10⁻⁴</td>
<td>3.594 × 10⁻⁵</td>
</tr>
<tr>
<td>mu1</td>
<td>5.0 × 10⁻⁹</td>
<td>2.0 × 10⁻⁴</td>
<td>6.202 × 10⁻⁶</td>
</tr>
<tr>
<td>mo2</td>
<td>1.0 × 10⁻⁸</td>
<td>5.0 × 10⁻⁴</td>
<td>3.570 × 10⁻⁶</td>
</tr>
<tr>
<td>mm2</td>
<td>3.0 × 10⁻⁸</td>
<td>9.0 × 10⁻⁵</td>
<td>3.594 × 10⁻⁶</td>
</tr>
<tr>
<td>mu2</td>
<td>5.0 × 10⁻¹⁰</td>
<td>2.0 × 10⁻⁵</td>
<td>6.202 × 10⁻⁷</td>
</tr>
<tr>
<td>soil</td>
<td>5.0 × 10⁻⁵</td>
<td>1.0 × 10⁻²</td>
<td>6.017 × 10⁻⁵</td>
</tr>
<tr>
<td>alluvium</td>
<td>4.0 × 10⁻⁵</td>
<td>1.0 × 10⁻²</td>
<td>3.219 × 10⁻⁴</td>
</tr>
</tbody>
</table>

4.3 Model evaluation against dynamic groundwater heads

In this subsection, the head observations of several monitoring wells in the catchment were used to evaluate the model performance. We analyzed discrepancies between the modeled and observed groundwater heads by subtracting long-term mean values, \(h_{\text{mod}}\) and \(h_{\text{obs}}\). Four model-skill scores including the mean value, the median value, the Pearson correlation coefficient \(R_{\text{cor}}\), and the inter-quartile range error, QRE, are used to evaluate the model performance.

Six wells with different geological and morphological properties were chosen as samples to exhibit the model performance (Figure 10). Specifically, well W10 is located in the northern uplands and is near the main stream, whereas well W1 is located in the southwestern lowlands. As can be observed from Figure 10, they provide good fits between simulated and observed heads, with a \(R_{\text{cor}}\) of 0.87 and 0.76, and QRE of -23.34% and -1.65%, respectively. Well W17 is located in the Lower Keuper unit, while well W16 is located in the upper Muschelkalk formation. In these two monitoring wells, the simulations are highly correlated with observations with high values of \(R_{\text{cor}}\) (0.71 and 0.82), in spite of their different geological properties (Figure 10). The simulation results at monitoring well W13 (located in the northern mountainous area) also exhibit a high correlation with the observation with a \(R_{\text{cor}}\) of 0.85 and W7 (located at the northern upland) also exhibit good correspondence with the observations (Figure 10). In general, the model is capable of capturing the historical trends of groundwater dynamics, even though the mean values of simulations and observations may deviate to some extent. Due to the limited spatial resolution and complex hydrogeological structure, this degree of discrepancy is acceptable.

4.4 Model sensitivity to different recharge scenarios

As described in Section 3.6, a reference recharge scenario (RR), i.e., a spatially uniform recharge scenario, is set up to assess the effect of spatial patterns of recharge on groundwater heads. In this uniform recharge scenario, RR, the steady-state
Figure 10. Comparison between measured (green dashed line) and simulated groundwater head anomalies (blue solid line). (a) W10 is located in uplands, near a stream. (b) W17 is located in a mountainous area. (c) W16 is located at a hillslope in the northern uplands. (d) W1 is located in the lowland. (e) W13 is located in the northern mountains. (f) W7 is located in the upland, northern upland.
Table 2. Hydraulic properties used in the uniform-recharge scenario (RR).

<table>
<thead>
<tr>
<th>Geological units</th>
<th>km</th>
<th>ku</th>
<th>mol</th>
<th>mm1</th>
<th>mu1</th>
<th>mo2</th>
<th>mm2</th>
<th>mu2</th>
<th>soil</th>
<th>alluvium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydraulic conductivity</td>
<td>5.023</td>
<td>6.216</td>
<td>8.608</td>
<td>2.990</td>
<td>5.316</td>
<td>8.604</td>
<td>2.997</td>
<td>5.317</td>
<td>5.239</td>
<td>7.302</td>
</tr>
</tbody>
</table>

Figure 11. Barplots of a) the Pearson correlation coefficient $R_{cor}$, mean absolute error MAE and b) the absolute inter-quartile range error $|QRE|$ in all monitoring wells in two recharge scenarios.

The groundwater model, was re-calibrated using the long-term mean of spatially uniform recharge (Table 2). For the purpose of showing discrepancies between two recharge scenarios, we compared the values of $R_{cor}$, MAE and $|QRE|$ at each monitoring well between the spatially distributed recharge, $mR$, and the uniform recharge, $RR$ (Figure 11). The mean value and the median value of $R_{cor}$ and $|QRE|$ were also calculated and are shown in Figure 11. Figure 11a indicates that the correlation between observations and simulations MAE using the spatially distributed recharge $mR$ is higher (4.04 m) is lower than that using the uniform recharge $RR$, with mean values of 0.704 and 0.677, respectively. The standard deviations are nearly the same in both scenarios (4.61 m). Considering that the only difference between the two recharge scenarios is their spatial patterns, we conclude that accounting for spatially-distributed recharge provides a moderate improvement in the model.

Figure 11b shows the absolute values of inter-quartile range error (|QRE|) in simulations using the two recharge scenarios ($mR$ and $RR$). We found that the deviation of |QRE| is significantly larger than $R_{cor}$, i.e., the |QRE| in two wells are abnormally higher than the other wells. The higher values of |QRE| at W8 and W18 may be caused by their proximity to model boundaries, as the two wells are located either near a river or near the catchment perimeter. This deviation indicates that accurate quantification of the amplitude of head fluctuations at certain locations is difficult, which may be due to the proximity of boundaries or complex local topography and geology. Nevertheless, 16 out of 19 wells exhibit low inter-quartile range errors, with the values of |QRE| in a range of ±40% in the spatially distributed $mR$ scenario. We also observe a smaller mean and standard deviation.
Figure 12. Seasonal variation of spatially-distributed groundwater heads by their anomalies after removing the long-term mean groundwater heads (unit: m). a) Long-term mean groundwater head distribution in spring; b) Long-term mean groundwater head distribution in summer; c) Long-term mean groundwater head distribution in autumn; d) Long-term mean groundwater head distribution in winter; e) Monthly mean groundwater head distribution in the wet season (August 2002); f) Monthly mean groundwater head distribution in the dry season (August 2003).

The 19 chosen monitoring wells cover the geological units of the alluvium, Keuper, and Muschelkalk, and range from high mountains to lowlands across the catchment. These results demonstrate the promising modeling capability of the model and highlight the moderately better historical matching when using a spatially distributed pattern of groundwater recharge.

Figure 12 illustrates the seasonality of groundwater heads by showing the spatial distribution of groundwater heads averaged over the spring, summer, autumn, and winter seasons, respectively. A strong spatial variability can be observed. For example, the fluctuation amplitudes of groundwater heads in the northern, eastern, and southeastern mountainous areas are larger than...
in the central plains area. In order to illustrate predicted groundwater levels and droughts caused by extreme climate events, we selected a meteorologically wet month (August 2002) and a meteorologically dry month (August 2003), and show the corresponding variations of groundwater heads in Figures 12e and 12f. In general, the groundwater heads in the wet season are higher than the long term mean values (Figure 12e). The variation of groundwater heads in the dry season, however, shows a strong spatial variability. Such a strong spatial variability of groundwater heads variation has also been reported by Kumar et al. (2016).

5 Discussion and conclusions

Our simulation results demonstrate that the coupled model mHM#OGS v1.0 can generally reproduce groundwater-head dynamics very well. It is also able to reasonably reproduce fluctuation amplitudes of groundwater heads, although with less accuracy. The simulation results also reveal that the stochastical and physically-based representations of groundwater dynamics can be intrinsically linked on the condition that the geometry and geological structure of groundwater aquifer are reasonably characterized. Compared to the good predictive capability of capturing the general trend behavior, the amplitude of head time series is hard to reproduce. This might be because local geological formations in the vicinity of monitoring wells may significantly alter local groundwater flow behavior, and thus further affect groundwater head fluctuations.

The results of this study demonstrate the successful application of the well-established hydrologic model, mHM, in estimating spatially heterogeneous groundwater recharge and baseflow at a regional scale. At a spatial scale of $10^3$ km$^2$ (the scale in this study), the distributed recharge estimated by mHM is clearly superior to using homogeneous recharge. mHM has been successfully applied at the continental scale covering entire Europe (Thober et al., 2015; Kumar et al., 2013b; Rakovec et al., 2016b; Zink et al., 2017). The successful application of the coupled model in this study suggests a huge potential for extending the applicability of mHM#OGS v1.0 to a larger-scale (e.g., $10^4 - 10^6$ km$^2$) or even a global scale.

The results of this study demonstrate a viable strategy for improving classic meso- to large-scale distributed hydrologic models, such as the current version of mHM (Samaniego et al., 2010; Kumar et al., 2013b), VIC (Liang et al., 1994), PCR-GLOBWB (Van Beek and Bierkens, 2009), WASMOD-M (Widén-Nilsson et al., 2007). These distributed hydrologic models do not calculate spatio-temporal groundwater heads and are therefore unable to represent groundwater head dynamics in their groundwater compartment. The physical representation of groundwater flow is, however, relevant in future regional-scale and possibly global hydrologic models to accurately determine travel times, solute export from catchments, and water quality in rivers (Botter et al., 2010; Benettin et al., 2015; Van Meter et al., 2017). The coupled model mHM#OGS v1.0 also offers the potential for predicting groundwater drought in analyzing the dynamic behavior of groundwater heads. Thus, it could be a useful tool for understanding groundwater anomalies under extreme climate conditions (Kumar et al., 2016; Marx et al., 2017).

For example, building on previous work of Heße et al. (2017), who calculated Travel Time Distributions (TTDs) using mHM, we can now expand the range of their work to the complete critical zone, which is important for comprehensively understanding particle (e.g., pollutant) transport behavior and the historical legacy in soil zone and groundwater storage (Basu et al., 2010; Beniston et al., 2014). mHM#OGS v1.0 fits well with the long-term simulation of nitrogen transport in the terrestrial water
cycle. The coupled model is also able to evaluate surface water and groundwater storage changes under different meteorological forcing conditions, which allows the comprehensive evaluation of hydrologic response to climate changes (e.g., global warming). Additionally, OGS demonstrates its capability in addressing thermo-hydro-mechanical-chemical (THMC) coupling processes in large-scale hydrologic cycles (not reflected in this study), which is significant for a wide range of real-world applications, including nutrient circulation, salt water intrusion, drought, and heavy metal transport (Kalbacher et al., 2012; Selle et al., 2013; Walther et al., 2014, 2017).

In addition to improving the predictive abilities of mHM, we can also demonstrate some improvements for the groundwater model OGS. Our results showed a modest improvement using mHM generated recharge compared to a simpler, uniform recharge rate. We currently gain a strong advantage for the description of the top boundary condition, i.e. the recharge, which is temporal and spatially variable through the input of mHM. Even more, the recharge fluxes provided are based on mHM’s phenomenological process description, which significantly better describes the surface level recharge fluxes than common approaches through recharge rates derived by empirical relations. 

In this study, we have focused our efforts on extending the applicability of mHM from surface hydrology to subsurface hydrology by a simple one-way coupling. Consequently, we do not account for any feedback between river and groundwater head fluctuations. While being a simplification of reality, this approach has certain advantages. First, the one-way coupling can be regarded as a conservative approach, such that the parametrization process, which is one of the most significant features of mHM, remains intact. In this way, we do not compromise any of its well-established features, such as the calibration of model parameters at different scales and accurate runoff prediction, while obtaining in addition very good estimates of groundwater heads, flow paths, and travel times. The inability of mHM to provide good estimates for these quantities has been noted in the past (see, e.g., Rakovec et al. (2016b); Heße et al. (2017)) and our work therefore extends the predictive abilities of mHM. Second, using such a one-way coupling will allow users of mHM to simply extend currently established catchment models and enhance their abilities in the aforementioned way. Using a more sophisticated approach is parsimonious and numerically efficient, meanwhile fully preserves the well-tested parameterization algorithm in mHM. Unlike two-way coupling, would entail users having to rebuild their models almost entirely. Third, a one-way coupling allows for ready future expansion of the predictive power of an mHM catchment model, should the need arise. Finally, one-way coupling takes less computational consumption and achieves better numerical stability than two-way coupling. In short, unlike two-way coupling, the one-way coupling described here allows the user to expand the abilities of mHM without sacrificing any of its well-known and well-established properties. However, in a next step, we will devote to incorporate a full, two-way coupling using the next version of the mHM#OGS model. The main limitation of one-way coupling is that the effects of a shallow depth to groundwater on actual ET, maintained by lateral groundwater flow, cannot be explicitly addressed. However, the dynamic interactions between overland flow and groundwater flow, as well as between soil moisture dynamics and groundwater dynamics can be explicitly modeled and investigated using a full coupling scheme. This approach is open to a broader spectrum of calibration options, such as calibration by remotely sensed soil moisture data.

In conclusion, we can state that the coupled model mHM#OGS v1.0 retains the predictive capability of mHM for discharge volumes. In addition, it is capable of reproducing groundwater head dynamics. The simulation results indicate a promising
predictive ability, confirmed by calibration and comparison to observed discharge and groundwater heads. Based on the historical match of discharge and groundwater heads in the case study, we conclude that the coupled model mHM#OGS v1.0 is a valuable tool for addressing many challenging problems in the field of water management, including pollutant transport and legacy, climate change, and groundwater drought.

**Code and data availability.** The mesoscale Hydrologic Model mHM (current release: 5.7) is an open-source community software and can be accessed from several mirrored repositories: SVN: http://www.ufz.de/index.php?en=40114; GitLab: https://git.ufz.de/mhm; GitHub: https://github.com/mhm-ufz. The modified source code of OGS5 can be freely acquired via the following link: https://github.com/UFZ-MJ/OGS_mHM.git. The model interface GIS2FEM and RIV2FEM can be freely acquired via the following link: https://github.com/UFZ-MJ/OGS_mHM/tree/master/UTL/GIS2FEM and https://github.com/UFZ-MJ/OGS_mHM/tree/master/UTL/RIV2FEM.

The input files of the case study in Nägelstedt catchment can be found in the Github repository: https://github.com/UFZ-MJ/OGS_mHM/tree/master/test_case. The dataset used in the case study can be found in the Github repository: https://github.com/UFZ-MJ/OGS_mHM/tree/master/data.

**Acknowledgements.** This research received funding from the Deutsche Forschungsgemeinschaft via Sonderforschungs- bereich CRC 1076 AquaDiva. We kindly thank Sabine Sattler from Thuringian State office for the Environment and Geology (TLUG) for providing the geological data. We kindly thank our data providers: the German Weather Service (DWD), the Joint Research Center of the European Commission, the Federal Institute for Geosciences and Natural Resources (BGR), the European Environmental Agency (EEA), the European Water Archive (EWA), and the Global Runoff Data Centre (GRDC). **We thank the two reviewers Prof. Olaf Cirpka and Dr. Edwin Sutanudjaja very much for their comprehensive reviews.**
References


Cirpka, O. A. and Attinger, S.: Effective dispersion in heterogeneous media under random transient flow conditions, Water resources research,
Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S.,
Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models,

Cuthbert, M., Mackay, R., and Nimmo, J.: Linking soil moisture balance and source-responsive models to estimate diffuse and preferential


Danskin, W. R.: Evaluation of the hydrologic system and selected water-management alternatives in the Owens Valley, California, vol. 2370,


Business Media, 2013.


Engdahl, N. B. and Maxwell, R. M.: Quantifying changes in age distributions and the hydrologic balance of a high-mountain watershed


Fischer, T., Naumov, D., Sattler, S., Kolditz, O., and Walther, M.: GO2OGS 1.0: A versatile workflow to integrate complex geological

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


