Reply to Referee #1

Page 4, line 98: This is the methodology you used to address the three concerns as you discussed in the introduction. I would suggest you to show a schematic showing how you address the concerns. This will help readers to get the ideas more straightforward.

Response:

We re-arranged the section 2 in correspondence of the three concerns. In first paragraph, we did an overview of these contents accordingly. Every sub-section is important to the methods developed in this paper. We hope it is straightforward and clear enough to the readers. As given in the revised manuscript (page 4, lines 99-106):

“To address the aforementioned first concern, governing equations for subsurface flow are given at different levels of complexity (section 2.1); numerical solution of these equations are presented (section 2.2); nonlinearity in the soil-water sub-models are reduced by a generalized switching scheme that chooses appropriate forms of the Richards’ equation (RE) according to the hydraulic conditions at each numerical node (section 2.3); then, an iterative feedback coupling scheme is developed to solve the soil-water and groundwater models at independent scales (section 2.4). As for the second concern, a multi-scale water balance analysis is conducted to deal with the scale-mismatching problem at the phreatic surface (section 2.5). To cope with the third concern, a moving Dirichlet boundary at the groundwater table is assigned to the soil water sub-models (see Appendix A.1); the Neumann upper boundary for the saturated model is provided in Appendix A.2.”

Page 6, lines 146-148: What kind of uncertainties could be as such to switch between one and another?

Response:

Switching the form of Richards’ equation only matters with the soil moisture condition. As shown in Figure C1, when it is very dry, the nonlinearity in h-form RE is significant; while when it is near saturation, the θ-form RE is less effective. In our previous study (Zeng et al., 2018), there is a wide range of soil moisture state suitable for both forms of RE. This part of theory was not demonstrated for saving page. Readers can see more details in Zeng et al., (2018). In the revised manuscript (Page 6, lines 150-152), we stated:

“When Se ≥ Se^cr, the soil moisture is closer to saturation, so the h-form RE is chosen as the governing equation; otherwise, when it undergoes dry soil condition, the θ-form RE is preferred. The empirical effective saturation for doing switching varies with soil type and is suggested to Se^cr = 0.4-0.9, the state when both the h- and θ-form REs are stable and efficient.”

Actually, different soil has different ranges of soil moisture suitable for switching of the governing equation. Figure C1 takes a sandy soil as an example. To concise the manuscript, we prefer not to make redundant explanation about how to determine the Se^cr of a certain soil.

Figure C1  The soil moisture state suitable for different forms of Richards’ equation. The van Genuchten model is used as the Constitutive relationship. D(h) is the hydraulic diffusivity, K(h) is the hydraulic conductivity, and C(h) is the soil moisture capacity.

Reference:
Page 7, line 164-165: how? Please offer a more specific description with details, in terms of "Space- and time-splitting strategies"
Response:
The space- and time-splitting strategies in this work, are equivalent to the scale-separation philosophy. That is, models at different scales are recognized as valid tools to describe the sub-systems, and messages at the sub-model interface are transferred within a mathematical framework. An iterative solver is usually developed to resolve the whole system. As the case in this work, the developed iterative feedback coupling “solver” solves the 1D soil water models and a 3D groundwater model at separated spatial and temporal scales. The single-scale methods in contrast, only use upscaling or downscaling approaches to unify sub-models at multiple scales, then solves them as a whole.
An example for doing space- or time-splitting should be the Multi-Scale Finite Element method (MsFEM). Generally speaking, MsFEM conducts the space-splitting mathematically at equation level. While the feedback coupling methods (e.g., the developed model) usually does such splitting physically at reservoir level. Scales are separated between the vadose and saturated zones. To avoid misleading, we corrected some statement as follows (page 7, lines 187-194)
“Space- and time-splitting strategy (see Figure 1) are adopted to separate sub-models at different scales. That is, the soil water models are established by $\Delta z = 10^{-3} \text{ m}$-$10^{-0} \text{ m}$, and $\Delta t = 10^{-5} \text{ d}$-$10^{-0} \text{ d}$; while for the saturated model, the grid sizes are $\Delta x = 10^{0} \text{ m}$-$10^{3} \text{ m}$, and time-step sizes are $\Delta t = 10^{-1} \text{ d}$-$10^{1} \text{ d}$. Water balance at one side of the interface is conserved by scale matching of boundary conditions provided by the sub-model on the other side. For unsaturated flow, the Richards’ equation requires fine discretization of space and time (Miller et al., 2006; Vogel and Ippisch, 2008); while for saturated flow, coarse spatial and temporal grids produce adequate solutions at large scale (Mehl and Hill, 2004; Zeng et al., 2017). To approximate the upper boundary flux of the groundwater flow model, a multi-scale water balance analysis is conducted within each step of the large-scale saturated flow model.”

Page 7, line 174: It is recommended to put this as annex. Page 8, line 187: It is suggested to put this into annex.
Response: We agree. Two of these parts are moved to the Appendix A.

Page 10, line 255: Please use a figure to indicate geometry for this case.
Response:
In the revised manuscript, we tried to clarify test case 1, rather than providing a figure for geometrical discretization. The reason is that the grids for the 1D soil models are uniform and with the same resolution ($\Delta z = 1 \text{ cm}$) and spatial spread ($z = 0$ -1000 cm). The grid for the coupled groundwater flow is quite simple, with $\Delta z = 500 \text{ cm}$ for two of the layers. In the revised manuscript, we added that (page 10, lines 250-251):
“The coupled unsaturated model is discretized into a fine grid with $\Delta z = 1 \text{ cm}$ for solving the Richards’ equation, while the saturated model is discretized into two layers with thickness of 500 cm.”

Page 11, line 262: “alternately” corrected by “alternatively”
Response: Corrected. Thanks.

Page 11, line 267: Is the moving groundwater table also considered in HYDRUS1D model? But, in this case, you only want to test soil water flow?
Response:
The moving table was not considered by a fully-1D (without coupling) HYDRUS1D model. HYDRUS1D model can simulate the unsaturated-saturated flow without coupling of the two parts.
As added in the revised manuscript (page 10, line 239), we use case 1 to “… investigate the benefit brought by
switching the Richards’ equation in the unsaturated zone”. That is, reducing the non-linearity in the soil water models can significantly cut down numerical cost and enhance stability when undergoing rapidly changing atmospheric upper boundaries. “The lower boundary is set non-flux to avoid the extra computational burden caused by variation of the groundwater model. Two scenarios from literature are reproduced with rapidly changing upper boundaries, as well as extreme flow interactions between the unsaturated and saturated zones (page 10, lines 241–244”). In case 1, the moving groundwater table indeed works through the simulation. With groundwater table rising from z = 200 cm to 600 cm, the length of the 1D soil column for unsaturated flow kept reducing from 800 cm to 400 cm. The moving groundwater table was caused by infiltration, rather than groundwater dynamics. The error reduction brought by moving groundwater table was discussed in test case 2.

Page 11, line 269: Please use a figure to explain the geometry you indicated here. And, this case, you only want to test the groundwater flow without considering soil water flow?
Response:
The schematic of case 2 is available (see Fig. 4). We revised the description of case 2, see page 10, lines 255–264. There are mainly two different situations with significant saturated lateral flow, i.e., sudden recharge and dynamic groundwater flow. For a quasi-3D unsaturated–saturated flow model, the unsaturated lateral flow is neglected according to its assumptions. Such assumptions are only applicable for cases with moderate infiltration, or with sudden infiltration while in very large-scale regions. Case 2 is not able to consider a very large region due to the limitation for obtaining reference solutions from a fully 2D/3D unsaturated–saturated model. A smaller region with pumping stresses is practical and demonstrative, as in case 2. The soil water upper boundary is of course applicable to increase complexity of the test case. However, it is not suggested for potentially introducing errors caused by the quasi-3D assumption in such a small-scale test, that is, the absence of unsaturated lateral flow. To better illustrate the benefits brought by using a moving Dirichlet lower boundary, the soil surface is set with the same non-flux boundary to minimize unsaturated lateral flow. In the revised manuscript, we added that (page 10, lines 255–256): “To minimize the unsaturated lateral flow, the soil surface is set with non-flux boundary.” However, this doesn’t necessarily mean that there is no soil water flow. The coupled model, as well as the fully-2D unsaturated-saturated model (VSF), made non-trivial efforts solving the Richards’ equation in the unsaturated zone. Test case 2 successfully demonstrated the necessity for using a moving Dirichlet lower boundary when there is significant saturated lateral flow, which is common for some local events, for example, intensive pumping and autumn irrigation.

Page 11, line 277: This case, you have both irrigation and pumping, so already coupled simulation needed. But you use MODFLOW-VSF model as the “GW Truth” while HYDRUS1D as “SW Truth”?
Response:

Page 12, line 296: Did you test the HYDRUS package for MODFLOW, with MODFLOW-VSF model results? Are there significant difference when compared to CASE2 and CASE3?
Response:
No we didn’t. The original HYDRUS package for MODFLOW has already been tested in literature (Twarakavi et al., 2008). The method was proved to be applicable for most cases without drastic flow interaction at the water table. The advantages of the developed method are illustrated by cases 1, 2, and 3. Under this condition, case 4 reproduced a benchmark synthetic case for regional application, which has already been substantially tested by Twarakavi, N. K. C., Šimůnek, J. and Seo, S.: Evaluating Interactions between Groundwater and Vadose Zone Using the HYDRUS-Based Flow Package for MODFLOW, Vadose Zo. J., 7(2), 757, doi:10.2136/vzj2007.0082, 2008. In their study, such a regional problem was simulated by the REC-ET, MODFLOW-UZF1, and the original HYDRUS package for MODFLOW. It was technically impossible to provide an accurate fully-3D truth solution for such a large region during such a long simulation period. In the same way therefore, we took the results from equivalent methods (original Hydrus package for MODFLOW) as a potential reference solution. Besides, the applicability of the developed method for practical use are presented. The codes, as well as inputs and outputs, are available from paper Twarakavi et al., (2008) and the references therein.

Page 12, line 300: Delete “soil water retention curve”
Response: The nonlinearity in soil water models are one of the concerns we addressed during the discussion. So we changed this sentence into “the non-linearity of the soil water models…” (see page 11, lines 289-290)

Page 12, line 310: Change “allowable” into “acceptable”
Response: Corrected.

Page 25, Fig 1, The figure is blurred not clear to be seen. It is lacking of description in terms of how coupling happens.
Response: We rebuilt the figures and made detailed description of them. See Fig. 1 in page 25.

Page 26, Fig 2: Again, a more detailed explanation is needed, other than just showing the figures. The figure is again blurred and need to be updated with high quality images.
Response: The figures are replaced with higher resolution. Detailed descriptions are presented. See Fig 2. In Page 26.
Reply to Referee #2

This paper developed the switching method of h- and θ-form of Richards equation to lower the non-linearity in the soil–water and groundwater coupling system and the iterative feedback coupling scheme to reduce the coupling errors. Four numerical cases were employed to address three concerns arose using the iterative feedback coupling method.

This work tries to find the tradeoff between the modeling accuracy and the computational cost of the soil–water and groundwater coupling system. The method presented here seems promising in the application of large scale problems. Nevertheless, I have some concerns about the ‘real’ modeling accuracy when using the proposed method. You took the simulation results of HYDRUS1D, MODFLOW-VSF (Thoms et al., 2006), and HYDRUS package for MODFLOW (Seo et al., 2007) as the ‘truth’. Thus, when you compare the simulation results of the proposed method and the “truth” (e.g., Figure 7), it is difficult to determine which method is better.

The paper is well written and structured. Some suggestions are detailed as below.

Response:

Thank you for your comments. The “truth” is not always easy to find. The regional-scale fully-3D solution with high-density discretization is extremely very difficult to obtain (Niswonger and Prudic, 2009). In relevant literature, the 3D Richards’ equation has been regarded as a painstaking truth. That was also the reason we turned to a quasi-3D scheme to approximate regional solution. Although various quasi-3D schemes were developed to reduce the total complexity and numerical difficulty, a tradeoff between cost and benefit was what we mainly concerned. To address three of the problems arose when doing quasi-3D coupling, we carefully designed different scenarios in the fully-1D/2D/3D benchmark problems.

Firstly, the efficiency improvement brought by switching the Richards’ equation was illustrated with rapidly changing atmospheric upper boundaries upon a 1D soil column. To avoid extra CPU time consumption that were caused by more complicated conditions in the saturated domain, only the upper boundary was in function. As a matter of fact, making the benchmark problem more complicated in the saturated part, will inevitably exaggerate the benefits brought by switching the Richards’ equation. The reason is that, when the coupling model suffers from more non-linearity at two sides of the coupling interface, more feedback iterations will be needed. Thus, the percent of CPU cost in the unsaturated zone will increase, which may overstate the advantages in the switching RE method.

Secondly, the error reduction by using the iteratively two-way coupling scheme was demonstrated with the 1D and 3D cases. In the 1D cases, the iterative coupling scheme was regulated by adjusting the closure criteria and maximal number of feedback iterations. Three times of feedback iterations with linearly predicted groundwater table were proved to be reasonable to achieve sound convergence of the problem. The optimum number of 1D soil columns for a 3D regional case was suggested to be as less as possible.

Thirdly, the accuracy improvement by the moving-balancing-domain approach were presented in the 2D case. The dynamically changing groundwater table was designed to better illustrate the significance of the saturated lateral flow in a quasi-3D coupling scheme. That is, when using a moving lower boundary for the balancing domain, the need for extra analysis of the water input from the saturated part was avoided successfully. Such improvement was illustrated by using different lengths of soil columns in the 2D test case. For simplicity, the soil-surface boundary was set non-flux. A more complicated upper boundary in the unsaturated zone, e.g. local infiltration, would of course add to the demonstration of such a pumping test. However, it would increase the solution errors in the unsaturated zone, which originates from the basic quasi-3D assumptions. To avoid the error superposition from the unsaturated and saturated zones, the 2D test case was carefully designed with only saturated stresses.

Finally, to demonstrate the applicability of the developed model for complex regional problems, a synthetic case from literature was revisited. The truth solution was not easy to obtain, so we compared the results with the original HYDRUS package for MODFLOW, as was done in relevant literature (Shen and Phanikumar, 2010; Twarakavi et al., 2008; Xu et al., 2012; Zhu et al., 2012). The purpose of this case study (case 4) was not to tell the accuracy improvement of the developed method, but to show its applicability for practical use.
Based on the above, the purposes and results of the test cases are further explained in the revised manuscript. Specifically, figure 7 of is the validation of the coupling model, other than the advantage of the method against one another.

References:

Title: HYDRYS → HYDRUS
Response:

Line 108: soil capacity → soil water capacity
Response:

All of the misused terms were revised. See lines 111 and 146.

Line 148: “… is suggested to 0.4-0.9” → “… is suggested to be 0.4-0.9”
Response:

It is revised. See line 152.

Line 163-167 (Figure 1): the space- and time-splitting strategies should be illustrated in a more detail way. You said that coupling models at different scales should deal with the inconsistency in spatial and temporal discretization, however, there is not too much context illustrate how such inconsistency was handled in your coupling system.
Response:

Thanks. We agree. The illustrative context for Figure 1 was further provided in lines 186-, as well as in the figure caption (see line 605 in page 26).

Line 253-254: there is no C and Δz in Eqn. 20. The description of C and Δz seems redundant.
Response:

We deleted that. Sorry for misleading.

Figure 6: what is the difference between coupled h-form RE and HYDRUS1D?
Response:

In this case, the h-form RE in the two-way coupling model is used to better illustrate the high performance in the Switching RE. Both forms of RE were incorporated into a two-way coupling scheme. However, it is not necessary and interfering. We removed this part of comparison, see the newly updated Figure 7 in page 31.

Section 4.2: You use both the “Cumulative mass balance errors” and “coupling errors” in the section 4.2 multi-scale water balance analysis, is there any difference between these two terms?
Response:
Sorry for misleading. In this work, the *cumulative mass balance errors* are equivalent to the *coupling errors*, and are exactly what we wanted to reduce with the developed method. We unified it into *coupling errors*, see line 320.

Figure 8: On the basis of the truth, three different methods (stepwise method, iteratively linear method, and non-iteratively linear method) were compared in Figure 8. What kind of method was used for the ‘truth’ (HYDRUS1D)?

Response:
HYDRUS1D is theoretically not a coupling model for unsaturated and saturated flows. It does not need any prediction of the Dirichlet lower boundary, including stepwise, linear extension, and etc. Differences among three of the coupling methods, stepwise, iteratively linear, and non-iteratively linear, were about how the 3D groundwater model (MODFLOW) provided Dirichlet lower boundary for the 1D unsaturated sub-models. To find out the “truth”, the HYDRUS1D solution, which was obtained without any coupling between the unsaturated and saturated sub-domains. So no such coupling method was used in HYDRUS1D. We used the 1D test case to better illustrate how did the (non-)iteratively linear method functioned as an approximation towards the “truth”.

Line 373: $\varepsilon_F = 20 \text{ m/d or } 20\text{cm/d}$?

Response:
We revised it into $\varepsilon_F = +\infty$, see line 364. Different combinations of $\varepsilon_H$ and $\varepsilon_F$ had some effect on the cost-benefit discussion. For example, the group of $(\varepsilon_H, \varepsilon_F) = (0.1 \text{ cm}, 0.01 \text{ cm/d})$ or $(0.01 \text{ cm}, 0.1 \text{ cm/d})$ may lead to similar closure errors. In case of such interference, we opened one of them. That is, only $\varepsilon_F$ was changed from 0.001 cm to 20 cm in the comparative analysis, while $\varepsilon_H$ is set by $+\infty$, which was not possible in real case.

Line 406: (Twarakavi et al., 2008) → Twarakavi et al. (2008)

Response:
Thanks, we revised it.

Line 408: You should clearly present the results of figures in the context, not just say “Figure 13b presents the absolute head difference of the method developed here and the HYDRUS package at the end of stress periods 3, 6, 9, and 12.”

Response:
The discussion of the figures was rephrased. See lines 395-399.

Figure 14: sub-zones 1, 3, 5, 7, 9 or sub-zones 1, 5, 9, 13, 20?

Response:
We corrected it.
Capturing soil-water and groundwater interactions with an iterative feedback coupling scheme: New HYDRUS package for MODFLOW

Jicai Zeng, Jinzhong Yang, Yuanyuan Zha, Liangsheng Shi

State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

Corresponding author: Yuanyuan Zha (zhayuan87@gmail.com)

Abstract. Accurately capturing complex soil-water and groundwater interactions is vital for describing the coupling between subsurface/surface/atmospheric systems in regional-scale models. The non-linearity of the Richards’ equation for water flow, however, introduces numerical complexity to large unsaturated-saturated modeling systems. An alternative is to use quasi-3D methods with a feedback coupling scheme to practically join sub-models with different properties, such as governing equations, numerical scales, and dimensionalities. In this work, to reduce the non-linearity in the coupling system, two different forms of the Richards’ equation are switched according to the soil-water content at each numerical node. A rigorous multi-scale water balance analysis is carried out at the phreatic interface to link the soil water and groundwater models at separated spatial and temporal scales. With a moving-boundary approach at the coupling interface, the non-trivial coupling errors introduced by the saturated lateral fluxes are minimized for problems with dynamic groundwater flow. It is shown that the developed iterative feedback coupling scheme results in significant error reduction, and is numerically efficient for capturing drastic flow interactions at the water table, especially with dynamic local groundwater flow. The coupling scheme is developed into a new HYDRUS package for MODFLOW, which is applicable for regional-scale problems.

Key words: Soil-water-groundwater interaction; Multi-scale water balance; Iterative feedback coupling; Regional-scale modeling; HYDRUS package for MODFLOW
Numerical modeling of the soil-water and groundwater interactions has to deal with both flow components and governing equations at different scales. This adds significant complexity to model development and calibration. Unsaturated soil water and saturated groundwater flows, governed by similar properties in porous media, are usually integrated into a whole modeling system. Although physically consistent and numerically rigorous, methods involving the 3D Richards’ equation \((RE, (Richards, 1931))\) tend to be computationally expensive and numerically unstable due to the large non-linearity and the demand for dense discretization (Kumar et al., 2009; Maxwell and Miller, 2005; Panday and Huyakorn, 2004; Thoms et al., 2006; Zha et al., 2013a), especially for problems with multi-scale properties. In this work, parsimonious approaches, which appear in different governing equations and coupling schemes, are developed for modeling the soil-water and groundwater interactions at regional scale.

Simplifying the soil-water flow details into upper flux boundaries has been widely used to simulate large-scale saturated flow dynamics, such as MODFLOW package and its variants (Langevin et al., 2017; Leake and Claar, 1999; McDonald and Harbaugh, 1988; Niswonger et al., 2011; Panday et al., 2013; Zeng et al., 2017). At local scale in contrast, the unsaturated flow processes are usually approximated with reasonable simplifications and assumptions in the Richards’ equation (Bailey et al., 2013; Liu et al., 2016; Paulus et al., 2013; Šimůnek et al., 2009; van Dam et al., 2008; Yakirevich et al., 1998; Zha et al., 2013b).

The original Richards’ equation, also the mixed-form \(RE\), takes pressure head \((h)\) as the driving force variable, while soil moisture content \((\theta)\) serves as the mass accumulation variable (Krabbenhøft, 2007). To solve the mixed-form \(RE\), either \(h\) or \(\theta\), or a switching of both, is assigned as the primary variable. The \(h\)-form \(RE\) is widely employed for unsaturated-saturated flow simulation, especially in heterogeneous soils, such as the HYDRUS package (Šimůnek et al., 2016). Significant improvement in mass conservation has been achieved with Celia’s modification (Celia et al., 1990), but models based on an \(h\)-form \(RE\) still suffer from high computational cost and low numerical robustness when dealing with rapidly changing atmospheric boundary conditions (Crevoisier et al., 2009; Zha et al., 2017). The \(\theta\)-form \(RE\), addressing the above problems, is inherently mass conservative and less non-linear in the averaged nodal hydraulic diffusivity (Warrick, 1991; Zha et al., 2013b). However, the \(\theta\)-form \(RE\) is not applicable for saturated and heterogeneous soils (Crevoisier et al., 2009; Zha et al., 2013b). In this work, to take advantages of both forms of \(RE\), the governing equations, rather than primary variables (Diersch and Perrochet, 1999; Forsyth et al., 1995; Zha et al., 2013a), are switched at each node according to its saturation degree.

For regional problems, the vadose zone is usually conceptualized into paralleled soil columns without lateral connections. The resulting quasi-3D coupling scheme (Kuznetsov et al., 2012; Seo et al., 2007; Xu et al., 2012; Zhu et al., 2012) significantly reduces the dimensionality and complexity. According to how the messages are transferred across the phreatic interface, the
quasi-3D methods are categorized into (1) the fully coupling scheme, which simultaneously builds the nodal hydraulic connections of models at both sides and implicitly solves the assembled matrices; (2) the one-way coupling scheme, which delivers the soil-water model solutions onto the upper boundary of the groundwater model without feedback mechanism; and (3) the feedback (or two-way) coupling scheme, which explicitly exchanges the head/flux solutions in vicinity of the interface nodes.

The fully coupling scheme (Gunduz and Aral, 2005; Zhu et al., 2012) is numerically rigorous but tends to increase the computational burden for practical conditions. For example, the potential conditional diagonal dominance causes non-convergence for the iterative solvers (Edwards, 1996). Owing to high non-linearity in the soil-water sub-models, the assembled matrices can only be solved with unified small time-steps, which adds to the computational expense. The one-way coupling scheme, as adopted by the MODFLOW-UZF1 package for MODFLOW (Grygoruk et al., 2014; Niswonger et al., 2006), as well as the free drainage mode of SWAP package for MODFLOW-SWAP model (Xu et al., 2012), assumes that the water table depth is of minor influence on flow interactions at the phreatic interface, and is thus problem specific.

The feedback coupling method, in contrast, is widely used (Kuznetsov et al., 2012; Seo et al., 2007; Shen and Phanikumar, 2010; Stoppelenbrug et al., 2005; Xie et al., 2012; Xu et al., 2012) as a compromise between numerical accuracy and computational cost. In a feedback coupling scheme, the soil-water and groundwater sub-models can be built with different governing equations, numerical schemes, and scales of discretization. For flow processes with multi-scale components, such as boundary geometries, parameter heterogeneities, and hydrologic stresses, the scale-separation strategy can be implemented easily. Although the feedback coupling method, either iteratively or non-iteratively, is numerically more rigorous than a one-way coupling method, and tends to reduce the inconsistency of head/flux interfacial boundaries, some concerns arise.

The first concern is the numerical efficiency of the feedback coupling methods. The non-iterative approach (Twarakavi et al., 2008; Xu et al., 2012) usually leads to significant error accumulation when dealing with dynamically fluctuating water table, especially with large time-step sizes. The iterative methods in contrast (Kuznetsov et al., 2012; Stoppelenbrug et al., 2005; Xie et al., 2012), by exchanging head/flux solutions across the interface to meet convergence, are numerically rigorous but computationally expensive, especially when solving the coupled sub-models with a unified time-stepping scheme (Kuznetsov et al., 2012).

The second concern lies in the scale-mismatching problem. For groundwater models (Harbaugh et al., 2017; Langevin et al., 2017; Lin et al., 2010; McDonald and Harbaugh, 1988), the specific yield at the phreatic surface is usually represented by a simple large-scale parameter; while for soil-water models (Niswonger et al., 2006; Šimůnek et al., 2009; Thoms et al., 2006), the small-scale phreatic water release is influenced by the water table depth and the unsaturated soil moisture profile (Dettmann and Bechtold, 2016; Nachabe, 2002). Delivering small-scale solutions of the soil-water models onto the interfacial boundary
of a large-scale groundwater model, as well as maintaining the global mass balance, usually introduce significant non-linearity to the entire coupling system (Stoppelenbrug et al., 2005). Conditioned by this, the mismatch of numerical scales in the coupled sub-models causes significant coupling errors and instability.

The third concern is the non-trivial lateral fluxes between the saturated regions of the vertical soil columns, which are usually not considered in previous study (Seo et al., 2007; Xu et al., 2012). Though rigorous water balance analysis is conducted to address such inadequacy (Shen and Phanikumar, 2010), the lateral fluxes solved with a 2D groundwater model usually require additional effort to build water budget equations in each sub-division represented by the soil columns.

In this work, the $h$- and $\theta$-form of the 1D Richards’ equation ($RE$) are switched at equation level to obtain a new HYDRUS package. To handle three of the aforementioned concerns, a multi-scale water balance analysis is carried out at the phreatic surface to conserve head/flux consistent at the coupling interface. An iterative feedback coupling scheme is developed for linking the unsaturated and saturated flow models at disparate scales. The saturated lateral fluxes between the soil columns are fully removed from the interfacial water balance equation, making it a moving-interface coupling framework. The head/flux solution of MODFLOW-2005 (Harbaugh et al., 2017; Langevin et al., 2017) and of HYDRUS1D (Šimůnek et al., 2009), are relaxed to meet consistency at the phreatic surface.

In this paper, the governing equations at different scales, the multi-scale water balance analysis at the phreatic surface, and the iterative feedback coupling scheme for solving the whole system, are presented in Section 2. Synthetic numerical experiments are described in Section 3. Numerical performance of the developed model is investigated in Section 4. Conclusions are drawn in Section 5.

2 Methodology

To address the aforementioned first concern, governing equations for subsurface flow are given at different levels of complexity (section 2.1); numerical solution of these equations are presented (section 2.2); nonlinearity in the soil-water sub-models are reduced by a generalized switching scheme is presented for choosing appropriate forms of the Richards’ equation ($RE$) according to the hydraulic conditions at each numerical node. A quasi-3D model for regional unsaturated-saturated flow simulation is developed within (section 2.3); then, an iterative feedback coupling scheme is developed to solve the soil-water and groundwater models at independent scales (section 2.4). As for the second concern, a multi-scale water balance analysis is conducted to deal with the scale-mismatching problem at the phreatic surface, a multi-scale water balance analysis is conducted (section 2.5). To cope with the third concern, a moving Dirichlet boundary at the groundwater table is assigned to the soil water sub-models (see Appendix A.1); the Neumann upper boundary for the saturated model is provided in Appendix A.2.
2.1 Governing equations

The mass conservation equation for unsaturated-saturated flow is given by:

\[
\frac{\partial \theta}{\partial t} + \beta \mu_s \frac{\partial h}{\partial t} = \left( C + \beta \mu_s \right) \frac{\partial h}{\partial t} = -\nabla \cdot \mathbf{q}
\]  

(1)

where \( t \) is time [T]; \( \theta \) [L^3L^{-3}] is volumetric moisture content; \( h \) [L] is pressure head; \( \beta \) is one for saturated region while zero for the unsaturated region; \( C \) [L^{-1}] is the soil water capacity \( (C = \partial \theta / \partial h) \) for unsaturated region, while zero for saturated region; \( \mu_s \) [L^{-1}] is specific elastic storage; \( \mathbf{q} \) [LT^{-1}] is Darcian flux calculated by:

\[
\mathbf{q} = -KH
\]  

(2)

where \( K \) [LT^{-1}] is the hydraulic conductivity, \( K = K(\theta) \); \( H \) [L] is the potentiometric head, \( H = h + z \), in which \( z \) is the vertical location with coordinate positive upward. Combining Eqns. (1) and (2) results in the governing equation for groundwater flow

\[
\mu_s \frac{\partial H}{\partial t} = \frac{\partial}{\partial x} \left( K \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left( K \frac{\partial H}{\partial y} \right) + \frac{\partial}{\partial z} \left( K \frac{\partial H}{\partial z} \right)
\]  

(3)

With the assumption that the horizontal unsaturated flows are negligible, the regional vadose zone is usually represented by an assembly of paralleled soil columns. The generalized 1D RE is represented by a switchable format,

\[
\tilde{C} \frac{\partial \psi}{\partial t} = \frac{\partial}{\partial z} \left( \tilde{K} \left( \frac{\partial \psi}{\partial z} + 1 \right) \right)
\]  

(4)

where \( \psi \) is the primary variable. For an \( h \)-form \( RE \), \( \psi = h \), \( \tilde{C} = C \), and \( \tilde{K} = K \); while for a \( \theta \)-form \( RE \), \( \psi = \theta \), \( \tilde{C} = 1 \), \( \tilde{K} = D \), where \( D \) [L^2T^{-1}] is the hydraulic diffusivity, \( D = K/C \).

2.2 Numerical approximation

The governing equation (Eqn. (3)) for the saturated zone is spatially and temporally approximated in the same form with the MODFLOW-2005 model (Harbaugh et al., 2017; Langevin et al., 2017). Celia’s modification (Celia et al., 1990; Šimůnek et al., 2009) is applied to the \( h \)-form 1D RE for temporal approximation. Both forms of \( RE \) are handled with a temporally backward finite difference discretization (Zha et al., 2013b, 2017). Each sub-model is solved by a Picard iteration scheme, which is widely used in some popular codes/software packages (van Dam et al., 2008; Šimůnek et al., 2016).

The spatial discretization of Eqn. (4), as well as the water balance analysis for each node, are based on the nodal flux in element \( i+1/2 \) (bounded by nodes \( i \) and \( i+1 \)), which is

\[
\tilde{q}_{i+1/2}^\psi = -\tilde{K}_{i+1/2}^{j+1,k} \left( \psi_{i+1/2}^{j+1,k+1} - \psi_{i+1/2}^{j+1,k} \right) + \tilde{K}_{i+1/2}^{j+1,k} + \tilde{K}_{i+1/2}^{j+1,k}
\]  

(5)

where the superscripts \( j \) and \( k \) are the levels of time and inner iteration; the subscript \( i \) (or \( i+1/2 \)) is the number of node (or element); \( \Delta z_{i+1/2} \) is the length of the element \( i+1/2 \), \( \Delta z_{i+1/2} = (z_{i+1} - z_i) \). When a soil interface exists at node \( i \) for example, the soil moisture contents in elements \( i-1/2 \) and \( i+1/2 \) are discontinuous at node \( i \), thus dissatisfying the \( \theta \)-form \( RE \). To address this
problem, the correction term \( \varepsilon^{\pm \frac{1}{2}} \), suggested by (Zha et al., 2013b), is employed to handle the heterogeneous interface at nodes \( i \) and \( i+1 \),

\[
\varepsilon_{i+1/2}^{\pm \frac{1}{2}} = \frac{\tilde{K}_{i+1/2}^{\pm \frac{1}{2}}}{\Delta \zeta_{i+1/2}} \left( \psi_{i+1/2}^{\pm \frac{1}{2}} - \psi_{i+1/2}^{\pm \frac{1}{2}} - \psi_{i+1}^{\pm \frac{1}{2}} + \psi_{i+1}^{\pm \frac{1}{2}} \right) \tag{6}
\]

where \( \psi_{i+1/2}^{\pm \frac{1}{2}} \) and \( \psi_{i+1}^{\pm \frac{1}{2}} \) are the continuously distributed \( \psi \) within element \( i+1/2 \), i.e., between the vertices \( i \) and \( i+1 \).

When \( \psi = \psi \) or \( \psi = \theta \) but no heterogeneity occurs, we get \( \psi_{i+1/2}^{\pm \frac{1}{2}} = \psi_{i+1/2}^{\pm \frac{1}{2}} \) and \( \psi_{i+1}^{\pm \frac{1}{2}} = \psi_{i+1}^{\pm \frac{1}{2}} \), so \( \varepsilon_{i+1/2}^{\pm \frac{1}{2}} = 0 \). When \( \psi = \theta \), with soil interfaces at node \( i \) or \( i+1 \), \( \psi_{i+1/2}^{\pm \frac{1}{2}} = \theta \left(h_{i+1/2}^{\pm \frac{1}{2}} - P_{i+1/2} \right) \) and \( \psi_{i+1}^{\pm \frac{1}{2}} = \theta \left(h_{i+1}^{\pm \frac{1}{2}} - P_{i+1} \right) \). It is obvious that

\[
\psi_{i+1/2}^{\pm \frac{1}{2}} \neq \psi_{i+1/2}^{\pm \frac{1}{2}} \text{ or } \psi_{i+1}^{\pm \frac{1}{2}} \neq \psi_{i+1}^{\pm \frac{1}{2}}, \text{ so } \frac{\psi_{i+1/2}^{\pm \frac{1}{2}}}{\Delta \zeta_{i+1/2}} \neq 0.
\]

Hereinafter, \( P_{i+1/2} \) represents the soil parameters in element \( i+1/2 \). For example, in van Genuchten model (van Genuchten, 1980), \( P_{i+1/2} = (\theta_r, \theta_s, n, m, \alpha, k) \), where \( \theta_r \) [L^3] and \( \theta_s \) [L^3] are the residual and saturated soil moisture contents; \( \alpha \) [L], \( n \), and \( m \) are the pore-size distribution parameters, \( m = 1-1/n; k_o \) [LT^{-1}] is the saturated hydraulic conductivity.

### 2.3 Switching the Richards’ equation

Due to lower non-linearity of hydraulic diffusivity (\( D \)) for dry soils (Zha et al., 2013b) and the avoidance of soil water capacity as the storage term, the which inevitably introduces mass balance error, a \( \theta \)-form \( RE \) is more robust than an \( h \)-form \( RE \), especially when dealing with rapidly changing atmospheric boundary conditions—(Zeng et al., 2018). In our work, the \( h \)- and \( \theta \)-form \( REs \) are switched at each node according to its effective saturation \( Se \). The resulting hybrid matrix equation set is solved by Picard iteration. When \( Se > Se^{\text{crit}} \), the soil moisture is closer to saturation, so the \( h \)-form \( RE \) is chosen as the governing equation; otherwise, when it undergoes dry soil condition, the \( \theta \)-form \( RE \) is preferred. The empirical effective saturation for doing switching varies with soil type and is suggested to be \( Se^{\text{crit}} = 0.4-0.9 \), the state when both the \( h \)- and \( \theta \)-form \( REs \) are stable and efficient.

For element \( i+1/2 \), when the governing equations for nodes \( i \) and \( i+1 \) are identical, the spatial approximation of nodal flux is given by Eqn. (5). When the governing equations differ at nodes \( i \) and \( i+1 \), a switched element is produced. Take \( \psi_j = \theta_j \) and \( \psi_{j+1} = h_{j+1} \) for example, the nodal fluxes calculated by Eqn. (5) for different forms of \( RE \) have to be carefully handled by substituting \( \theta_j^{i+\frac{1}{2}} \) with \( \theta_j^{i+\frac{1}{2}} \), while \( h_j^{i+\frac{1}{2}} \) is replaced by \( h_j^{i+\frac{1}{2}} \). When \( \psi_j = \psi_j \) and \( \psi_{j+1} = \theta_{j+1} \), in contrast, \( h_j^{i+\frac{1}{2}} \) is replaced by \( h_j^{i+\frac{1}{2}} \), while \( \theta_j^{i+\frac{1}{2}} \) is replaced by \( \theta_j^{i+\frac{1}{2}} \). The resulting equivalent nodal fluxes \( q^h_{i+1/2} \) and \( \frac{\partial}{\partial x} q^\theta_{i+1/2} q^\theta_{i+1/2} \) are then weighted to obtain an approximation by

\[
q_{i+1/2} = (1-\omega)q^h_{i+1/2} + \omega \cdot q^\theta_{i+1/2} \tag{7}
\]

where \( \omega \) is the weighing weighting factor, \( 0 \leq \omega \leq 1 \). In our work, \( \omega = 0.5 \) is applied to implicitly maintain the unknown variables of both \( h_j^{i+\frac{1}{2}} \) and \( \theta_j^{i+\frac{1}{2}} \). Specifically, when \( \omega = 1 \), the \( h \)-form \( RE \) is used at both of nodes \( i \) and \( i+1 \); when \( \omega = 0 \), the \( \theta-
form RE is employed instead. A detailed study on doing switching of RE between two ends of the soil moisture condition, as well as the description of this approach—the numerical formation—can be found in Zeng et al. (2018). Note that the equation switching method takes full advantages of both the $\theta$- and $h$-forms of RE, which is different from the traditional primary variable switching schemes—(Diersch and Perrochet, 1999; Forsyth et al., 1995; Zha et al., 2013a). The

our work, the switching-RE formation approach is incorporated into a new HYDRUS package.

2.4 Iterative feedback coupling scheme

The Dirichlet and Neumann boundaries are iteratively transferred across the phreatic interface. The groundwater head solution serves as the head-specified lower boundary of the soil columns; while the unsaturated solution is converted into the flux-specified upper boundary of the groundwater model. Due to moderate variation of the groundwater flow, the predicted water-table solution is usually adopted in advance as Dirichlet lower boundary of the fine-scale soil-water flow models (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012), which then in sequence provides the Neumann upper boundary for successively solving the coarse-scale groundwater flow model. Appendix A.1 provides the method for a moving Dirichlet lower boundary, while Appendix A.2 presents the Neumann upper boundary for the 3D groundwater model.

Relaxed feedback iteration is used to accelerate convergence of head and flux at the phreatic surface. The Dirichlet lower boundary head for the soil columns, $z_t$, as well as the Neumann upper boundary fluxes for the phreatic surface, $F_{top}$, are updated within each iterative step

$$
\frac{z_t^{\text{updated}}}{F_{\text{top}}^{\text{updated}}} = \lambda_h \cdot z_t^{\text{new}} + (1 - \lambda_h) \cdot z_t^{\text{old}} \quad \lambda_f \cdot F_{\text{top}}^{\text{new}} + (1 - \lambda_f) \cdot F_{\text{top}}^{\text{old}}
$$

(8)

where superscript old (or new) indicates the previous (or newly calculated) head/flux boundaries at the coupling interface; $\lambda_h$ and $\lambda_f$ are the relaxation factors for head/flux boundaries respectively, $0 < \lambda_h$ and $\lambda_f \leq 1$. The iteration ends when agreements are reached at

$$
|z_t^{\text{updated}} - z_t^{\text{old}}| \leq \varepsilon_H \quad \text{and} \quad |F_{\text{top}}^{\text{updated}} - F_{\text{top}}^{\text{old}}| \leq \varepsilon_F
$$

(9)

where $\varepsilon_H$[L] and $\varepsilon_F$[LT$^{-1}$] are residuals for the feedback iteration of interfacial head and flux.

2.5 Multi-scale coupling approachwater balance analysis

Coupling models at different scales requires dealing with the inconsistency in consistency in their spatial and temporal discretizations at the interface (Downer and Ogden, 2004; Rybak et al., 2015). Space- and time-splitting strategy (see Figure 1) are adopted to conserve massseparate sub-models at two different scales. That is, the soil water models are established by $\Delta z = 10^{-3}$ m-$10^0$ m, and $\Delta t = 10^{-5}$ d-$10^0$ d; while for the saturated model, the grid sizes are $\Delta x = 10^0$ m-$10^3$ m, and time-step sizes are $\Delta t = 10^0$ d-$10^1$ d. Water balance at one side of the scales interface is conserved by scale
matching of boundary conditions provided by the sub-model on the other side. For unsaturated flow simulation, the Richards’ equation requires fine discretization of space and time (Miller et al., 2006; Vogel and Ippisch, 2008); while for saturated flow modeling, coarse spatial and temporal grids produce adequate solutions at large scale (Mehl and Hill, 2004; Zeng et al., 2017).

The Dirichlet and Neumann boundaries at the phreatic interface are iteratively transferred. The groundwater head solution serves as the head-specified lower boundary of the soil columns; while the unsaturated solution is converted into the flux-specified upper boundary of the groundwater model. Due to moderate variation of the groundwater flow, the predicted saturated solution is usually adopted first to approximate the fine-scale soil-water flow (Seo et al., 2007; Chen and Phanikumar, 2010; Xu et al., 2012), which then in sequence provides the Neumann upper boundary for successively solving the coarse-scale groundwater flow model.

(1) Dirichlet boundary prediction

The bottom node of a soil column is adaptively located at the phreatic surface represented by an area-averaged moving Dirichlet boundary

\[ z(T) \] (L) is the elevation of the water table; \( s \) is the control domain of a soil column; \( H(T) \) (L) is potentiometric head solution, as well as the elevation of the phreatic surface, which is obtained by solving the groundwater model; \( s \) is the horizontal area.

To simulate the multi-scale flow process within a macro time step \( \Delta T^{i+1} = T^{i+1} - T^{i} \), the lower boundary head of a soil column is temporally predicted either by stepwise extension of \( z(T) \) (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012) or by linear extrapolation from \( z(T^{i+1}) \) and \( z(T^{i}) \). In Figure 2a, the stepwise extension method (a) causes large deviation from the “truth”. The linear extrapolation is resorted to reduce the coupling errors and to accelerate the convergence of the feedback iteration. The small-scale lower boundary head at time \( (T^{i} < t < T^{i+1}) \) is given by

(2) Neumann boundary feedback

To approximate the upper boundary flux of the groundwater flow model, a multi-scale water balance analysis is conducted within each step of the large-scale saturated flow model. The balancing domain (see b) is bounded by a specific elevation above the phreatic surface, \( z_s \) (L), and the dynamically changing phreatic surface, \( z(t) \) (L).

Assume that the large-scale phreatic aquifer is numerically represented by the activated top layer in a two-dimensional groundwater model.

\[ \text{where} \quad [L] \text{ is the thickness of the phreatic layer, } z_b \text{ is the bottom elevation of the top phreatic layer, } F_{\text{top}} \text{ (LT}^{-1}\text{)} \text{ is the} \]
groundwater recharge into the activated top layer of the phreatic aquifer, $F_{\text{top}} = F_{\text{base}}$ is the leakage into an underlying numerical layer, $F_{\text{base}}$ (positive downward, so as $F_{\text{base}}$). The long-term regional-scale parameter indicating the water yield caused by fluctuation of the water table (Nachabe, 2002), $[-]$, is calculated by

$$R_{\text{top}} = \int_{z_{\text{top}}}^z (w(t') - w(t')) \, dz + \sum_{j=1}^N \int_{t_{j-1}}^{t_j} \left[ q_{\text{top}}(t) - q_{\text{bot}}(t) \right] \, dt$$

where $w$ [L] is the amount of unsaturated water in the moving balancing domain, see Figure 2b, $w(t) = \int_{z_{\text{top}}}^z \theta(t, z) \, dz$; $\Delta z_i = \sum_{j=1}^N \int_{t_{j-1}}^{t_j} z_i(T^{j-1}) - z_i(T^j)$ is the total fluctuation of the phreatic surface during $\Delta T = \sum_{j=1}^N t_j - t_{j-1}$; $\theta_i$ is the saturated soil water content. Approaching a transient state at time $t$, the water balance in a moving water balancing domain (see $z \in [z_i, z_s]$ in Figure 2b) during a small-scale time step $dt$ (defined in Figure 1b) is given by

$$[q_{\text{top}} + l \cdot dz_i / 2 - q_{\text{bot}}] \, dt = w(t) - w(t - dt) + \theta_i \cdot dz_i$$

where $q_{\text{top}}(t)$ and $q_{\text{bot}}(t)$ [LT$^{-1}$] are the nodal fluxes into and out of the moving balancing domain at a fixed top boundary ($z_i$) and a moving bottom boundary ($z_b = \min(z_i(t), z(t-dt))$, $q_{\text{top}} = K(h) \cdot \partial(h+z) / \partial z |_{z=z_i}, q_{\text{bot}} = K(h) \cdot \partial(h+z) / \partial z |_{z=z_b}$) (positive into the balancing domain and negative outside); $dz_i = z(t) - z(t-dt)$ is the transient fluctuation of the phreatic surface during $dt$; $l$ [T$^{-1}$] is the saturated lateral flux into the balancing domain at time $t$, see Figure 2b. Taking $\Gamma$ as the lateral boundary of a sub-domain, the lateral flux

$$l = \int_{x,y,z} \left[ \frac{\partial}{\partial x} \left( K \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left( K \frac{\partial H}{\partial y} \right) \right] \, dx \, dy \, dz$$

$$+ \int_{x,y,z} \left[ \frac{\partial}{\partial x} \left( K \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left( K \frac{\partial H}{\partial y} \right) \right] \, dx \, dy \, dz$$

is supposed to be constant during $\Delta T$; $\Omega$ is the volume of the saturated domain controlled by a soil column, which is horizontally projected into $\Pi$. Temporally integrating Eqn. (11) from time $T'$ to $T^{j-1}$ produces

$$R_{\text{top}} + e_i - R_{\text{bot}} = w(T^{j-1}) - w(T') + \theta_i \cdot \Delta z_i$$

where $R_{\text{top}}$ [L] is the cumulative water flux at $z_s$, $R_{\text{top}} = \int_{T'}^{T^{j-1}} q_{\text{top}}(t) \, dt$, note that $R_{\text{top}}$ is equals $F_{\text{top}}$ in Eqn. (19); $R_{\text{bot}}$ [L] is the cumulative water flux out of the moving balancing domain, $R_{\text{bot}} = \int_{T'}^{T^{j-1}} q_{\text{bot}}(t) \, dt$; $e_i$ [L] is the cumulative lateral input water into the moving balancing domain,

$$e_i = \frac{1}{2} l \cdot \sum_{j=1}^N dt_j \cdot dz_i \ll e_i' = \frac{1}{2} l \cdot \Delta T \cdot \Delta z_i$$

where $N$ is the number of time steps for the small-scale soil-water model within a macro time step $\Delta T$; and $e_i'$ is the non-
trivial saturated later flux produced by a stationary boundary method (Seo et al., 2007; Xu et al., 2012). By taking \( R_{\text{top}} \) as the specific recharge at \( z_t \), the small-scale specific yield \( \bar{S}_y \) is derived from Eqns. (10) and (12) as

\[
\bar{S}_y = \left( R_{\text{top}} + \varepsilon - R_{\text{out}} \right) / \Delta z_t + \mu \cdot M \tag{14}
\]

Suppose \( z_t \) is linearly fluctuating in time, i.e., \( z_t = a \cdot t + b \) (where \( a \) and \( b \) are constants), we get the water table change during a small-scale step \((dt)\) by \( dz_t = a \cdot dt \), thus, \( \varepsilon_l = o(dt^2) \), which means linearly refining the local time-step size \((dt)\) in the soil water model brings about at least quadratic approximation of \( \varepsilon_l \) towards zero. Thus \( \varepsilon_l \) can be neglected from the small-scale mass balance analysis. In the developed model, the large-scale specific yield, \( \bar{S}_y \) in Eqn. (19), represents the water release in the phreatic aquifer; while the small-scale \( \bar{S}_y \) in Eqn. (14), denotes the dynamically changing water yield caused by the fluctuation of the water table. The upper boundary flux \( F_{\text{top}} \) in the phreatic flow equation (Eqn. (19)) is therefore corrected to

\[
F_{\text{top}} = \left[ R_{\text{top}} + (\bar{S}_y - \bar{S}_y) \Delta z_t \right] / \Delta T \tag{15}
\]

Differing from previous studies (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012), a scale-separation strategy is employed in Eqn. (15). The specific yields at two different scales are linked explicitly by \( F_{\text{top}} \). The large-scale properties in the groundwater flow model (MODFLOW-2005) are thus fully remained.

2.5 Relaxation and closure criteria

Relaxed iteration is conducted to exchange head and flux across the phreatic surface. The Dirichlet lower boundary head for the soil columns, \( z_t \), as well as the Neumann upper boundary flux for the phreatic surface, \( F_{\text{top}} \) are updated by

\[
\text{where superscript \textit{old} (or \textit{new}) indicates the previous (or newly calculated) head and flux boundaries at the coupling interface;}
\]

\[
\lambda \text{ is the relaxation factor, } 0 < \lambda \leq 1. \text{ The iteration ends when agreements are reached at}
\]

\[
\text{where } [L] \text{ and } [LT^{-1}] \text{ are residuals for the feedback iteration of interfacial head and flux.}
\]

3 Numerical experiments

In this section, a range of 1D, 2D, 3D, and regional numerical test cases are presented. The 1D tests are benchmarked by the globally refined solutions produced by the HYDRUS1D code (Šimůnek et al., 2009). For the 2D and 3D problems, “truth” solutions are obtained from the MODFLOW-VSF fully-3D unsaturated-saturated flow model VSF (Thoms et al., 2006) and taken as the “truth”. At regional scale, a synthetic case study suggested by (Twarakavi et al., 2008) is reproduced. The codes are run on a 16 GB RAM, 3.6 GHz Intel Core (i3-4160) based personal computer. The reasonable maximal number of
feedback iteration is set at 20. Soil parameters for the van Genuchten model (van Genuchten, 1980) are listed in Table 1.

The root mean square error (RMSE) of the solution $\psi$ at time $t$ is given by

$$RMSE(\psi, t) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left( \psi_i^{ref} - \psi_i(x, t) \right)^2 \right\}^{1/2}$$  

where $\psi$ is the numerical solution of either pressure head or water content; $\psi^{ref}$ is the corresponding reference solution; $i$ is the control volume number of node $i$ (nodes, $i = 1, 2, ..., N$); $\Delta z_i$ is the control volume of node $i$.

### 3.1 Case 1: Rapidly changing atmospheric boundaries

The 1D case is used to investigate the numerical efficiency of the multi-scale coupling scheme of Richards’ equation in the unsaturated zone. A soil column is initialized with hydrostatic water-table depth of 800 cm. That is, $h(t = 0, z) = 200 - z$ cm, with $z = 0$ at the bottom, and $z = 1,000$ cm on the top. Two scenarios from the literature are reproduced with rapidly changing upper boundaries, as well as extreme flow interactions between the unsaturated and saturated zones.

Miller et al.’s problem (Miller et al., 1998) is reproduced in scenario 1. A dry-sandy soil column (see soil #1 in Table 1) experiences a large constant flux infiltration at the soil surface of $q_{top} = 30$ cm/d which ceases at $t = 4$ d.

In scenario 2, Hills et al.’s problem (Hills et al., 1989) is considered. The soils #2 and #3 from Table 1 are alternatively layered with a thickness of 20 cm within the first 80-cm depth. Below 80 cm ($z = 0-920$ cm) is soil #2 with non-flux bottom boundary. The atmospheric upper boundary conditions, rainfall and evaporation change rapidly with time (see Figure 3), over 365 days.

The coupled unsaturated model is discretized into a fine grid with $\Delta z = 1$ cm for solving the Richards’ equation, which is bounded by the moving groundwater table, while the saturated model is discretized into two layers with thickness of 500 cm.

The impact of different numbers of feedback iteration, closure criteria, as well as different forms of 1D Richards’ equation, are investigated. Solutions obtained from the HYDRUS1D model with $\Delta z = 1$ cm, and $\Delta t = 0.05$ d are taken as the “truth”.

### 3.2 Case 2: Dynamic Groundwater flow

A 2D case is analyzed with sharp groundwater flow (see Figure 4). To minimize the unsaturated lateral flow, the soil surface is set with non-flux boundary. The bottom and lateral boundaries are also non-flux. Two pumping stresses are applied to the cross-sectional field with $x \times z = 5,000$ cm $\times 1,000$ cm. Well #1 is located at $(x, z) = (2,500, 0-200)$ cm, with pumping screen at $z = 0-200$ cm, while well #2 is at $(x, z) = (5,000, 0-200)$ cm, with pumping screen of $z = 0-200$ cm. Pumping rates for wells #1 and #2 respectively are $2 \times 10^4$ cm$^2$/d and $1 \times 10^4$ cm$^2$/d per width unit. The initial hydrostatic head of the cross-section is $h_0(x, z) =$
700 cm. Soil #4 in Table 1 fills the entire cross-section. The total simulation lasts 50 days. For the coupled saturated sub-model, as well as the reference model (MODEFLOW-VSF (Thoms et al., 2006)), the cross-section is discretized horizontally into uniform segments with width $\Delta x = 50$ cm, while vertically (top-down/bottom-up) refined into segments with thickness $\Delta z = 5200$ cm, $100$ cm, $50$ cm, $25$ cm, $12.5$ cm, $25$ cm, $50$ cm, $100$ cm, and $2005$ cm (×2001), $100$ cm (×2), $50$ cm (×2), $25$ cm (×2), $12.5$ cm (×4), $25$ cm (×2), $50$ cm (×2), $100$ cm (×2), and $2005$ cm (×1200), where the subscripts hereinafter (×N) are the numbers of discretized segments. The coupled 1D soil-water models are discretized with segmental thickness of $\Delta z = 1$ cm. The fully-2D unsaturated-saturated solutions from VSF model are taken as the “truth”.

### 3.3 Case 3: Pumping and irrigation

In case 3, a phreatic aquifer with $x \times y \times z = 1,000$ m $\times 1,000$ m $\times 20$ m is stressed by constant irrigation and pumping wells. The infiltration rate is 3 mm/d in $(x, y) = (0-440$ m, $560$ m-$1,000$ m), while 5 mm/d in $(x, y) = (560$ m-$1,000$ m, $0-440$ m). Screens for three of the pumping wells locate at $(x, y, z) = (220$ m, $220$ m, $5-10$ m), $(500$ m, $500$ m, $5-10$ m), and $(780$ m, $780$ m, $5-10$ m). The pumping rates are constant at 30 m$^3$/d. The initial hydrostatic head of the aquifer is 18 m. Around and below the aquifer are non-flux boundaries. The aquifer is horizontally discretized with $\Delta x = \Delta y = 40$ m for the coupled saturated model, as well as for the MODFLOW-VSF model for obtaining the “truth” solution. The top-down thicknesses of the fully-3D grid are $\Delta z = 0.10$ m (×30), $0.4$ m (×5), $1$ m (×5), and $2$ m (×5). For the 1D soil columns, $\Delta z = 0.1$ m (×30), and $0.4$ m (×5), which means no soil column reaches the bottom. Different numbers of the sub-zones represented by soil columns, as well as their different geometries, are given in Figure 5. The soil parameters for a sandy loam (soil #5) are given in Table 1. Total simulation lasts 60 days.

### 3.4 Case 4: Complex Synthetic regional flow case study

A hypothetical test case from literature (Niswonger et al., 2006; Prudic et al., 2004; Twarakavi et al., 2008) for large-scale simulation is reproduced here. The overall alluvial basin is divided into uniform grids with $\Delta x = \Delta y = 1,524$ m. The coupled saturated model is conceptualized into a single layer. The initial head, as well as the elevations of land surface and bedrock, are presented in Figure 6a, b, and c. The precipitation, evaporation, and pumping rates for 12 stress periods, each lasted 1/12 of 365 days, are given in Table 2. The infiltration factors (see Figure 6d) are used to approximate the spatial variability of precipitation. The initial head in the vadose zone is set with hydrostatic status. Twenty soil columns, coinciding with the sub-zones in Figure 6d, are discretized separately with a range of gradually refined thicknesses with thickness ($\Delta z$) from 30.48 cm, to 0.3048 cm (bottom-up). A benchmark Comparative analysis is conducted by comparison with the solutions obtained from the original HYDRUS package for MODFLOW (taken as HPM for short) (Seo et al., 2007).
4 Results and discussion

4.1 Reducing the complexity of a feedback coupling system

The numerical difficulty in a coupled unsaturated-saturated flow system originates from the non-linearity of the soil-water hydraulic retention curve models, heterogeneity of the parameters, as well as the variability of the hydrologic stresses (Krabbenhøft, 2007; Zha et al., 2017). In our work, the overall complexity of an iteratively coupled quasi-3D model can be lowered by (1) taking full advantages of the \( h \)- and \( \theta \)-form REs to reduce the nonlinearity in the soil-water models, and (2) smoothing the variability of the exchanged interfacial messages.

Two scenarios in case 1 were selected to address the first point. Sudden infiltration into a dry-sandy soil, and the rapidly altering atmospheric upper boundaries, are tested to illustrate the importance of applying a switching-form RE for lowering the non-linearity in the soil-water models. To evaluate the benefits brought by a switching-form RE, the numerical stability is first considered, as shown in Figure 7. The coupled model in our work is tested with \( h \)-form and switching-form REs. Compared with the HYDRUS1D model (also based on an \( h \)-form RE), the switching-form method is numerically more robust, i.e., with larger minimal time-step sizes (\( \Delta t_{\text{min}} \)) and less computational cost, where minimal time-step size is the allowable acceptable \( 10^{-3} \text{ d} \) for convergence. Notably at the beginning of the sudden infiltration into a dry-sandy soil, in Figure 7a, the \( \Delta t_{\text{min}} \) for a switching method is \( 10^{-3} \text{ d} \), even at early infiltration times, while for the \( h \)-form methods, including HYDRUS1D and the coupled \( h \)-form method, \( \Delta t_{\text{min}} \) is constrained to \( 10^{-8} \text{ d} \) before reaching a painstaking convergence. In Figure 8, the soil water content solution by the coupled switching-form method and the HYDRUS1D method (taken as the “truth”) are compared at depth of 0, 50 cm, and 200 cm. To finish the entire calculation, the coupled switching-form RE method took 17 s seconds, while it was 41 s seconds for the HYDRUS code. When solving the same problem, the matrix equation set is solved 4,903 times with the switching scheme, while 10,925 times for the HYDRUS1D code. Reducing the non-linearity in the switched governing equations contributes to cutting the computational cost by half for problems with rapidly changing upper boundary conditions.

Reducing the complexity of a coupling system can also be attained by smoothing the exchanged information in space and time. As suggested by Stoppelenbrug et al. (2005), a time-varying specific yield calculated by the small-scale soil-water models, \( \tilde{S}_y \), in Eqn. (14), introduces significant variability to the large-scale groundwater model, thus causes extra iterations. A large-scale \( \bar{S}_y \) reduces the non-linearity of the storage term in the groundwater equation. In case 1, using an \( \bar{S}_y \) of 0.1-0.2 in the groundwater model produces best numerical stability for the sandy soil with dramatically uprising water table. With a large-scale \( \bar{S}_y \), the non-linearity introduced by the small-scale soil-water models can be quickly smoothed, as shown in Eqn. (14).
4.2 Multi-scale water balance analysis

The traditional non-iterative feedback coupling methods cannot maintain sound mass balance near the phreatic surface, especially for problems with drastic flow interactions.

One reason is that, to launch a new step of a sub-model at either side of the phreatic interface, the non-iterative feedback methods usually employ a predicted interfacial boundary without correction, which inevitably introduces cumulative mass balance coupling errors. In traditional non-iterative methods (Seo et al., 2007; Xu et al., 2012), such shortcomings can be alleviated by refining the macro time step size ($\Delta T$). However, the Dirichlet head predicted for the soil columns with a stepwise extension method (see Figure 2a), is easy to implement but tends to suffer from significant coupling error. In this work, we proposed a linear extrapolation method for the lower boundary head prediction for the soil water models, see Eqn. (18). Here, we use Niter to indicate the maximal number of feedback iteration. Compared with a traditional stepwise method, the solution obtained by a linear method, either iteratively (with Niter = 3) or non-iteratively (Niter = 0), is easier to approach the truth, see Figure 9. Even with refined macro time step sizes ($\Delta T$ from 0.2 d to 0.005 d), the stepwise method makes a thorough effort to minimize the coupling errors. Notably, three feedback iterations (Niter = 3) are sufficient to reduce the coupling error significantly. Such a one-dimensional case with constant upper boundary flux, avoiding interference from lateral fluxes, illustrates the importance of a temporal scale-matching analysis for coupling the soil-water and groundwater models.

The other factor contributing to the coupling errors in the traditional method lies in neglecting the saturated lateral flux between adjacent soil columns (Seo et al., 2007; Stoppelenbrug et al., 2005; Xu et al., 2012). In practical applications, the fluxes in and out of the saturated parts of the soil columns differ, which adds to the complexity of the coupling scheme. Although a strict water balance equation is established (Shen and Phanikumar, 2010), the concern centers on the spatial scale-mismatching problem. That is, when the coarse-grid groundwater flow solutions are converted into the vertically distributed fine-scale source/sink terms for the soil columns, an extra down-scaling approach is needed to ensure their accuracy. Here we carried out a multi-scale water balance analysis above the phreatic surface. The fine-scale saturated lateral flows are thus excluded from Eqn. (12). The benefits of the moving-boundary approach, can be seen in case 2 which produces significant saturated lateral flux. We have carried out a comparative analysis against the traditional stationary-boundary methods (Seo et al., 2007; Xu et al., 2012). The 2D solution of MODFLOW-VSF is taken as the “truth”. Figure 10 presents the effectiveness of the moving-boundary method. Five stationary soil columns with three different lengths ($L = 1,000$ cm, $500$ cm, and $300$ cm) are compared with an adaptively moving soil column within the iterative feedback coupling scheme. The cross-sectional RMSE of the phreatic surface and the head at bottom layer ($z = 0$), are presented in Figure 10a and b. The soil columns with bottom nodes fixed deeply into the aquifer, instead of moving with the phreatic surface, can introduce large coupling errors. This is caused by the non-trivial saturated lateral fluxes between the adjacent soil columns. With a traditional stationary-boundary method,
such problems can be alleviated by avoiding large saturated lateral fluxes between the soil columns. However, for some spatiotemporally varying local events in a regional aquifer (e.g., flooding or pumping irrigation), such problems increase the burden for sub-zone partitioning. A moving-boundary method instead, is numerically more efficient for minimizing the size of the matrix equation and reducing the coupling errors.

4.3 Regulating the feedback iterations

In coupling two complicated modeling system, a common agreement has been reached that, feedback coupling, either iteratively (Markstrom et al., 2008; Mehl and Hill, 2013; Stoppelenbrug et al., 2005; Xie et al., 2012) or non-iteratively (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012), is numerically more rigorous than a one-way coupling scheme. The main difference between the above two methods lies in the ability to conserve mass within a single step for back-and-forth information exchange. In an iterative method, the head/flux boundaries are iteratively exchanged. There is a cost-benefit tradeoff to obtain higher numerical efficiency.

During the late stages of the recharge in scenario 1 of case 1, the groundwater table rises quickly, which increases the burden on the coupling scheme. In our work, feedback iteration is conducted to eliminate the coupling error with the back-and-forth boundary exchange. To investigate how the feedback iteration influences the numerical accuracy as well as computational cost, solutions are compared with different closure criteria, instead of different maximal numbers of feedback iterations. For this purpose, scenario 1 in case 1 is tested with a range of closure criteria indicated by Closure = 0.001, 0.01, 0.1, 1, 5, and 20. Specifically, Closure = 20 (i.e., ε_H = 20 cm) is too large to regulate any feedback iteration, and is thus labelled by “non-iterative”. The ε_F indicating the closure of the Neumann boundary feedback iteration, is usually related to the phreatic Darcian flux. To avoid its impact on the discussion below, we assume ε_F = 20 m/d +∞, which means no regulation from the flux boundary exchange. However, their relaxation factors are both set by 1.0 to have straightforward update of the interfacial boundaries.

When the wetting front approaches the phreatic surface (at t = 2.4 d), the number of feedback iteration increases dramatically, see Figure 11a. This is caused by the dramatic rise of the water table within each macro time step ΔT. The head/flux interfacial boundaries are thus not easy to approximate the “truth”. With several attempts to exchange the head/flux boundaries, the head solution is effectively drawn towards the “truth”, see Figure 11b. With Closure < 2, i.e., ε_H < 2 cm, the coupling errors are significantly reduced, see Figure 11c. The cost-benefit curve, which is quantified by the number of feedback iteration instead of CPU cost, is indicative to problems with larger scales, and higher dimensionalities.

4.4 Parsimonious decision making

The feedback coupling schemes, either iteratively or non-iteratively, increase the degree of freedom for the users to manage...
the sub-models with different governing equations, numerical algorithms, as well as the heterogeneities in parameters and variabilities in hydrologic stresses. For practical purposes, a significant concern is how to efficiently handle the complicated and scale-disparate systems.

For problems with rapid changes in groundwater flows, as in case 2, the hydraulic gradient at the phreatic surface is large. Using a single soil column usually introduces significant coupling errors at the water table, see Figure 12a. Although portioning more sub-zones means higher accuracy for the coupling method, five or more soil columns are adequate enough to approximate the “truth”. Furthermore, for the saturated nodes deep in the aquifer, such coupling errors are of minor influence, see Figure 12b.

In case 3, a simple pumped and irrigated region was simulated with different numbers of soil columns. A range of tests with total numbers of 16, 12, 9, 5, and 3 soil columns are carried out to obtain a cost-benefit curve shown in Figure 13c. When partitioning the vadose zone into more than 12 soil columns, there is a slight reduction in solution errors (RMSE) and a significant increase in computational cost caused by solving more 1D soil water models. Although the expense can be reduced by using paralleled computation among the soil columns, representing the vadose zone with 3 soil columns can achieve acceptable accuracy, as presented in Figure 13a and b. The computational cost for obtaining the fully-3D solution with MODFLOW-VSF is 15.561 s, which is more than 11 times larger than an iterative feedback coupling method with soil-water models sequentially solved. Problems in more complicated real-world situations can thus be simplified to achieve higher numerical efficiency.

4.5 Regional application

The Prudic et al.’s problem was originally designed to validate a streamflow routing package (Prudic et al., 2004). Stressed by soil-surface infiltration, pumping wells, and general head boundary, the synthetic case was used to evaluate several unsaturated flow packages for MODFLOW (Twarakavi et al., 2008). Based on their studies, in case 4, we compared the developed iterative feedback coupling method with the HYDRUS package for MODFLOW. In case 4, the saturated hydraulic conductivity, as well as its heterogeneity, are forced to be consistent with that in the vadose zone, which is different from the case in Twarakavi et al (2008). Figure 14a gives the contours for the final phreatic head solutions, indicating a good match of the phreatic surface with the HYDRUS package. Figure 14b presents the absolute head difference of the method developed here and the HYDRUS package at the end of stress periods 3, 6, 9, and 12. The dark color blocks indicate the largest difference in head solution. According to Figure 6d, the saturated grid cells controlled by the soil columns of #3, #9, #10, #15, #19 are suffering largest deviation, although with the same horizontal partitioning of the unsaturated zone. The strict iteratively two-way coupling contributes to such accuracy improvement.

For unsaturated-saturated flow situations, the vadose zone flow is important. Figure 15 presents the water content profiles at
sub-zones 1, 3, 5, 7, and 9 as examples. The solutions obtained from the unsaturated models match the original HYDRUS package well. For practical purpose, the manually controlled stress periods for the unsaturated models are no longer a constraint.

In our method, the soil water models run at disparate numerical scales, which makes it possible to handle daily or hourly observed information rather than a stress period lasting 2 or more days in traditional groundwater models.

5 Summary and conclusions

Fully-3D numerical models are available but are numerically expensive to simulate the regional unsaturated-saturated flow. The quasi-3D method presented here, in contrast, with horizontally disconnected adjacent unsaturated nodes, significantly reduces the dimensionality and complexity of the problem. Such simplification brings about computational cost-saving and flexibility for better manipulation of the sub-models. However, the non-linearity of the soil-water retention curve, as well as the variability of realistic boundary stresses of the vadose and saturated zones, usually result in a scale-mismatching problem when attempting numerical coupling. In this work, the soil-water and groundwater models are coupled with an iterative feedback (two-way) coupling scheme. Three concerns about the multi-scale water balance at the phreatic interface are addressed using a range of numerical cases in multiple dimensionalities. We conclude:

1) A new HYDRUS package for MODFLOW is developed by switching the $\theta$ and $h$ forms of Richards’ equation (RE) at each numerical node. The switching-RE circumvents the disadvantages of the $h$- and $\theta$-form REs to achieve higher numerical stability and computational efficiency. The one-dimensional switch-form RE is applied to simulate the rapid infiltration into a dry-sandy soil, and the swiftly altering atmospheric upper boundaries in a layered soil column. Compared with the $h$-form RE, the switching-RE uses $10^5$ times larger minimal time-step size ($\Delta t_{\text{min}}$) and conserves mass better. Lowering the non-linearity of soil-water models with this switching scheme is promising for coupling complex flow modeling systems at regional scale.

2) Stringent multi-scale water balance analysis at the water table is conducted to handle scale-mismatching problems and to smooth information delivered back-and-forth across the interface. In our work, the errors originating from inadequate phreatic boundary predictions are reduced firstly by a linear extrapolation method, and then by an iterative feedback. Compared with the traditional stepwise extension method, the linear extrapolation significantly reduces the coupling errors caused by the scale-mismatching. For problems with severe soil-water and groundwater interactions, the coupling errors are significantly reduced by using an iterative feedback coupling scheme. The multi-scale water balance analysis mathematically maintains numerical stabilities in the sub-models at disparate scales.

3) When a moving phreatic boundary is assigned to soil columns, it avoids coupling errors caused by excluding saturated lateral fluxes from the phreatic water balance analysis. In practical applications for regional problems, the fluxes in and out of the saturated parts of the soil columns differ, which adds to the complexity and phreatic water balance error of the coupling
scheme. With a moving Dirichlet lower boundary, the saturated regions of soil-water models are removed. The coupling error is significantly reduced for problems with major groundwater flow. Extra cost-saving is achieved by minimizing the matrix sizes of the soil-water models.

Future investigation will focus on regional solute transport modeling based on the developed coupling scheme. Surface flow models, as well as the crop models, which appears to be less non-linear than the sub-surface models, will be coupled in an object-oriented modeling system. The RS- and GIS-based data class can then be resorted to handle more complicated large-scale problems.

*Data/code availability:* All the data used in this study can be requested by email to the corresponding author Yuanyuan Zha at zhayuan87@gmail.com.

**Appendix A**

### A.1 The moving Dirichlet lower boundary

The bottom node of a soil column is adaptively located at the phreatic surface, which makes it an area-averaged moving Dirichlet boundary

\[
z_t(T) = \int_{s,1} H(T) \, ds / \int_{s,1} ds
\]

where \( z_t(T) \) [L] is the elevation of the water table; \( \Pi \) is the control domain of a soil column; \( H(T) \) [L] is potentiometric head solution, as well as the elevation of the phreatic surface, which is obtained by solving the groundwater model; \( s \) is the horizontal area.

To simulate the multi-scale flow process within a macro time step \( \Delta T^{j+1} = T^{j+1} - T^j \), the lower boundary head of a soil column is temporally predicted either by stepwise extension of \( z_t(T^j) \) (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012) or by linear extrapolation from \( z_t(T^{j+1}) \) and \( z_t(T^j) \). In Figure 2a, the stepwise extension method (\( z_t'(T^j) \)) potentially causes large deviation from the "truth." In our study, the linear extrapolation is resorted to reduce the coupling errors and to accelerate the convergence of the feedback iteration. The small-scale lower boundary head at time \( t \) (\( T^j < t \leq T^{j+1} \)) is given by

\[
z_t(t) = \frac{(t - T^{j-1}) \cdot z_t(T^j) - (t - T^j) \cdot z_t(T^{j+1})}{T^j - T^{j-1}}
\]

### A.2 The Neumann upper boundary

The moving Dirichlet boundary introduces the need for water balance of a moving balancing domain above the water table (see Figure 2b), which is bounded by a specific elevation above the phreatic surface, \( z_s \) [L], and the dynamically changing phreatic surface, \( z_t(t) \) [L].
Assume that the activated top layer in a three-dimensional groundwater model is conceptualized into a phreatic aquifer, the governing equation for this layer is given by

$$\nabla \cdot (\vec{M} \nabla \vec{H}) = \frac{\partial}{\partial x} \left( K \vec{M} \nabla \vec{H} \right) + \frac{\partial}{\partial y} \left( K \vec{H} \nabla \vec{H} \right) + F_{\text{top}} - F_{\text{base}} \quad (19)$$

where \( \vec{M} \) [L] is the thickness of the phreatic layer, \( \vec{M} = \vec{z} - \vec{z}_0 \); \( \vec{z}_0 \) is the bottom elevation of the top phreatic layer, \( \vec{z}_0 \ll \vec{z} \); \( F_{\text{top}} \) [LT\(^{-1}\)] is the groundwater recharge into the activated top layer of the phreatic aquifer, \( F_{\text{top}} = (K \cdot \nabla \vec{H}/\vec{z})_{\vec{z}=\vec{z}_0} \); \( F_{\text{base}} \) is the leakage into an underlying numerical layer, \( F_{\text{base}} = (K \cdot \nabla \vec{H}/\vec{z})_{\vec{z}=\vec{z}_0} \) (positive downward, so as \( F_{\text{top}} \)). The long-term regional-scale parameter indicating the water yield caused by fluctuation of the water table (Nachabe, 2002), \( \vec{S}_y \) [-], is calculated by

$$\vec{S}_y = V_w \Delta \vec{H} / (A \Delta \vec{H}) \quad (20)$$

where \( V_w \) [L\(^3\)] is the amount of water release caused by fluctuation of the phreatic surface (\( \Delta \vec{H} \) [L]); \( A \) [L\(^2\)] is the area of interest.

Author contribution: Jicai Zeng, Yuanyuan Zha and Jinzhong Yang developed the new package for soil water movement based on a switching Richards’ equation; Jicai Zeng and Yuanyuan Zha developed the coupling methods for efficiently joining the sub-models. Four of the co-authors made non-negligible efforts preparing the manuscript.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgments. This work was funded by the Chinese National Natural Science (No. 51479143 and 51609173). The authors would thank Prof. Ian White and Prof. Wenzhi Zeng for his their laborious revisions and helpful suggestions to the paper.

References


<table>
<thead>
<tr>
<th>#</th>
<th>Soil</th>
<th>$\theta_r$ (cm$^3$cm$^{-3}$)</th>
<th>$\theta_s$ (cm$^3$cm$^{-3}$)</th>
<th>$\alpha$ (1/cm$)$</th>
<th>$n$</th>
<th>$k_s$ (cm/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sand</td>
<td>0.093</td>
<td>0.301</td>
<td>0.0547</td>
<td>4.264</td>
<td>504</td>
</tr>
<tr>
<td>2</td>
<td>Berino loamy fine sand</td>
<td>0.029</td>
<td>0.366</td>
<td>0.028</td>
<td>2.239</td>
<td>541</td>
</tr>
<tr>
<td>3</td>
<td>Glendale clay loam</td>
<td>0.106</td>
<td>0.469</td>
<td>0.010</td>
<td>1.395</td>
<td>13.1</td>
</tr>
<tr>
<td>4</td>
<td>Loam</td>
<td>0.078</td>
<td>0.430</td>
<td>0.036</td>
<td>1.560</td>
<td>24.96</td>
</tr>
<tr>
<td>5</td>
<td>Sandy loam</td>
<td>0.065</td>
<td>0.410</td>
<td>0.075</td>
<td>1.890</td>
<td>106.1</td>
</tr>
</tbody>
</table>
Table 2  The precipitation, evaporation, and pumping rates in 12 stress periods.

<table>
<thead>
<tr>
<th>Stress period</th>
<th>Precipitation (mm/d)</th>
<th>ET (mm/d)</th>
<th>Pumping rate (m$^3$/d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.21</td>
<td>1.32</td>
<td>4078</td>
</tr>
<tr>
<td>2</td>
<td>1.69</td>
<td>1.32</td>
<td>4078</td>
</tr>
<tr>
<td>3</td>
<td>2.11</td>
<td>1.32</td>
<td>2039</td>
</tr>
<tr>
<td>4</td>
<td>4.21</td>
<td>1.32</td>
<td>2039</td>
</tr>
<tr>
<td>5</td>
<td>1.05</td>
<td>1.32</td>
<td>6116</td>
</tr>
<tr>
<td>6</td>
<td>2.11</td>
<td>1.32</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0.63</td>
<td>1.32</td>
<td>4078</td>
</tr>
<tr>
<td>8</td>
<td>1.05</td>
<td>1.32</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0.63</td>
<td>1.32</td>
<td>2039</td>
</tr>
<tr>
<td>10</td>
<td>0.42</td>
<td>1.32</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0.21</td>
<td>1.32</td>
<td>6116</td>
</tr>
<tr>
<td>12</td>
<td>0.21</td>
<td>1.32</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 1: Schematic of the space- and time-splitting strategy for coupling models at two independent scales. $J$ (or $j$), $T$ (or $t$), and $\Delta T$ (or $dt$) are time level, time, and time-step size at coarse (or fine) scale. $\Delta x$ is the resolution for spatial discretization. For a groundwater model, spatial discretization is expected to be large ($\Delta x = 10^3 \text{ m} - 10^0 \text{ m}$); while for soil water models, it occurs to be small ($\Delta x = 10^{-3} \text{ m} - 10^0 \text{ m}$). Multiple levels of temporal discretization are common for regional problems. For groundwater model, the stress periods ($SP$) and macro time step sizes ($\Delta T$) appear by months and days ($10^0 \text{ d} - 10^1 \text{ d}$). For soil water models, the time step sizes are about $10^{-5} \text{ d} - 10^0 \text{ d}$. 
(a) Prediction of Dirichlet boundary for soil water models

(b) Water balance analysis of a moving domain

Figure 2: The Dirichlet-Neumann coupling of the soil-water and groundwater flow models at different scales. (a) Linear or stepwise prediction of Dirichlet lower boundary for the soil water flow model. (b) Water balance analysis based on a balancing domain with moving lower boundary. Blue dash line is the linearly extrapolated groundwater table as an alternative for prediction of Dirichlet lower boundary. $J$ (or $i$), $T$ (or $t$), and $\Delta T$ (or $dt$) are the time level, time, and time-step size at coarse (or fine) scale. At any of the transient state ($t$), the balancing domain is bounded by a user-specified top elevation ($z_{top}$), and the moving phreatic surface ($z_t$). The saturated lateral flux of the moving domain is indicated by $l(t)$, while the unsaturated lateral flux is neglected as the assumption of quasi-3D models. The water flux into and out of the balancing domain is indicated by $q_{top}$ and $q_{bot}$. 

$\begin{array}{ll}
& \\
\end{array}$
Figure 3: Rapidly changing atmospheric upper boundary conditions ($q_{\text{top}} = \text{Precipitation} - \text{Evaporation}$) for scenario 2, case 1.
Figure 4: Different number Schematic of sub-zones partitioned the cross-sectional for the quasi-3D simulations in Case 3-test case 2. Two pumping wells with screens of $z = 0-200$ cm are located at $x = 2,500$ cm and 5,000 cm. The overall vadose zone is partitioned into 16, 12, 9, 5, and 3 sub-zones. Pumping rates per unit width at well #1 and #2 are respectively $2 \times 10^4$ cm$^2$/d and $1 \times 10^4$ cm$^2$/d, respectively.
Figure 5: Input different number of the synthetic regional problem including (a) land surface elevation, (b) initial head, (c) bedrock elevation of the aquifer, and (d) the sub-zones and boundaries partitioned for the quasi-3D simulations in Case 3. The vadose zone is partitioned into 16, 12, 9, 5, and 3 sub-zones.
Figure 6: The time-step sizes through input of the simulation of synthetic regional problem including (a) sudden infiltration into a dry sandy soil column and surface elevation, (b) initial head, (c) bedrock elevation of the aquifer, and (d) rapidly changing atmospheric upper boundary conditions with a layered soil column) the sub-zones and boundaries.
Figure 7: The observed soil moisture content at $z = 0$ cm, 50 cm, and 200 cm for the layered soil column with (b) rapidly changing atmospheric upper boundary conditions (Scenario 2, Case 1), with a layered soil column.
Figure 8: Water table changing with time for different macro time step sizes ($\Delta T$). The observed soil moisture content at $z = 0.005 \text{ d, 0.05 d, 0.1 d cm, 50 cm, and 0.2 d}$, in scenario 1, case 1. The HYDRUS1D solution is taken as 200 cm for the “truth”. Compared with the stepwise extended method, the cumulative mass balance error is significantly reduced by a linear prediction layered soil column with rapidly changing upper boundary conditions (Scenario 2, Case 1).
Figure 9: Comparison of RMSE of (a) the phreatic surface water table changing with time for different macro time step sizes ($\Delta T = 0.005 \, \text{d}, 0.05 \, \text{d}, 0.1 \, \text{d}$, and $0.2 \, \text{d}$), in scenario 1, case 1. The HYDRUS1D solution is taken as the “truth”. Compared with the stepwise extended method (Seo et al., 2007) at $z = 0$ between the moving-boundary and the stationary-boundary methods. Three different lengths of the stationary soil columns, $L = 1,000 \, \text{cm}, 500 \, \text{cm}$, and $300 \, \text{cm}$, are considered. The coupling error is significantly reduced by a linear prediction.
Figure 10: Comparison of RMSE of (a) the phreatic surface and (b) the head solution (at \( z = 0 \)) between the moving-boundary and the stationary-boundary methods. Three different lengths of the stationary soil columns, \( L = 1,000 \text{ cm}, 500 \text{ cm}, \) and \( 300 \text{ cm} \), are considered.
Figure 11: (a) The number of feedback iterations and (b) phreatic surface solution changing with different closure criteria. The legend “Closure = 0.001” means $\varepsilon_H = 0.001$ cm and $\varepsilon_F = 0.001$ cm/d, is used to regulate the feedback iteration. The HYDRUS1D solution is taken as “truth”. Tested in scenario 1, case 1.
Figure 12: Comparison of (a) Phreatic surface water table and (b) head solution (at z = 0) that are changing by the number of soil columns. Solutions obtained with a moving-boundary method in case 2.
Figure 13: (a) Comparison of contours of the phreatic surface solution obtained with the fully-3D and quasi-3D methods; (b) Comparison of the phreatic surface at A-A’ cross-section; (c) computational cost and RMSE changing by the number of total soil columns.
Figure 14: (a) Comparison of elevation of the water table calculated by the HYDRUS package for MODFLOW (Seo et al., 2007) and the developed method ($t = 365$ d); (b) The absolute head difference of the phreatic head solution by the method developed here and HYDRUS package at the end of stress periods 3, 6, 9, and 12. (Case 4).
Figure 15: Comparison of water content profiles obtained from the HYDRUS package for MODFLOW (Seo et al., 2007) and the developed iterative feedback coupling method. Sub-zones 1, 3, 5, 9, 137, and 209 are shown as an example. $(t = 365 \text{ d in Case 4})$. 