Capturing soil-water and groundwater interactions with an iterative feedback coupling scheme: New HYDRYS 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

5   *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 join practically 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

1 Introduction

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,

25 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.

30 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.,

35 2013b).

The original Richards' equation, also the *mixed-form RE*, takes pressure head (*h*) as the driving force variable, while soil moisture content (*θ*) serves as the mass accumulation variable (Krabbenhøft, 2007). To solve the *mixed-form RE*, either *h* or *θ*, 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

40 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 *θ-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 *θ-form RE* is not applicable for saturated and heterogeneous soils (Crevoisier et al., 2009; Zha et al., 2013b). In this work, to

45 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

50   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.

55   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 (Grygoruk et al., 2014; Niswonger et al., 2006), as well as the free

60   drainage mode of 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

65   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.,

70   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).

75   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

80 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

85 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 *θ-form* of the 1D *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

90 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

95 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

The governing equations for subsurface flows are given at different levels of complexity. A generalized switching scheme is

100 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 an iterative feedback

coupling scheme. To deal with the scale-mismatching problem at the phreatic surface, a multi-scale water balance analysis is

conducted.

2.1 Governing equations

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

$$\frac{\partial \theta}{\partial t} + \beta \mu_s \frac{\partial h}{\partial t} = \left(C + \beta \mu_s \right) \frac{\partial h}{\partial t} = -\nabla \cdot \mathbf{q} \tag{1}$$

where *t* is time [T]; θ [L^{3}L^{-3}] is volumetric moisture content; *h* [L] is pressure head; β is one for saturated region while zero

for the unsaturated region; C [L^{-1}] is the soil capacity ($C = \partial\theta/\partial h$) for unsaturated region, while zero for saturated region; μ_s

[L^{-1}] is specific elastic storage; $\mathbf{q}$ [LT^{-1}] is Darcian flux calculated by:

110
$$\mathbf{q} = -K\nabla H \qquad (2)$$

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

$$\mu_s \frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(K\frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y}\left(K\frac{\partial H}{\partial y} \right) + \frac{\partial}{\partial z}\left(K\frac{\partial H}{\partial z} \right) \qquad (3)$$

115 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,

$$\hat{C}\frac{\partial \psi}{\partial t} = \frac{\partial}{\partial z}\left(\hat{K}\left(\frac{\partial \psi}{\partial z} + 1 \right) \right) \qquad (4)$$

where ψ is the primary variable. For an *h-form RE*, $\psi = h$, $\hat{C} = C$, and $\hat{K} = K$; while for a *θ-form RE*, $\psi = \theta$,

$\hat{C} = 1$, $\hat{K} = D$, where D [L^{2}T^{-1}] is the hydraulic diffusivity, $D = K / C$.

120 **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,

125 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

$$q^{\psi}_{i+1/2} = -\frac{\hat{K}^{j+1,k}_{i+1/2}}{\Delta z_{i+1/2}}\left(\psi^{j+1,k+1}_{i+1} - \psi^{j+1,k+1}_{i} \right) - K^{j+1,k}_{i+1/2} + \varepsilon^{j+1,k}_{i+1/2} \qquad (5)$$

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

130 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 *θ-form RE*. To address this

problem, the correction term $\varepsilon^{j+1,k}_{i+1/2}$, suggested by (Zha et al., 2013b), is employed to handle the heterogeneous interface at

nodes i and $i+1$,

$$\varepsilon_{i+1/2}^{j+1,k} = \frac{\hat{K}_{i+1/2}^{j+1,k}}{\Delta z_{i+1/2}}\left(\psi_{i+1}^{j+1,k} - \psi_i^{j+1,k} - \vec{\psi}_{i+1}^{j+1,k} + \bar{\psi}_{i,\Omega}^{j+1,k}\right) \tag{6}$$

where $\vec{\psi}_{i+1}^{j+1,k}$ and $\bar{\psi}_i^{j+1,k}$ are the continuously distributed ψ within element $i+1/2$, i.e., between the vertices i and $i+1$.

When $\psi = h$, or when $\psi = \theta$ but no heterogeneity occurs, we get $\psi_{i+1}^{j+1,k} = \vec{\psi}_{i+1}^{j+1,k}$ and $\psi_i^{j+1,k} = \bar{\psi}_i^{j+1,k}$, so $\varepsilon_{i+1/2}^{j+1,k} = 0$. When

$\psi = \theta$, with soil interfaces at node i or $i+1$, $\vec{\psi}_{i+1}^{j+1,k} = \theta\left(h_{i+1}^{j+1,k}, \mathbf{p}_{i+1/2}\right)$ and $\bar{\psi}_i^{j+1,k} = \theta\left(h_i^{j+1,k}, \mathbf{p}_{i+1/2}\right)$. It is obvious that

$\psi_i^{j+1,k} \neq \bar{\psi}_i^{j+1,k}$ or $\psi_{i+1}^{j+1,k} \neq \vec{\psi}_{i+1}^{j+1,k}$, so $\varepsilon_{i+1/2}^{j+1,k} \neq 0$.

Hereinafter, $\mathbf{P}_{i+1/2}$ represents the soil parameters in element $i+1/2$. For example, in van Genuchten model (van Genuchten,

1980), $\mathbf{P}_{i+1/2} = (\theta_r, \theta_s, n, m, \alpha, k_s)$, where θ_r [L³L⁻³] and θ_s [L³L⁻³] are the residual and saturated soil moisture contents; α [L⁻¹],

n, and m are the pore-size distribution parameters, $m = 1−1/n$; k_s [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 capacity as

the storage term, the θ-form RE is more robust than an h-form RE, especially when dealing with rapidly changing atmospheric

boundary conditions. In our work, the h- and θ-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 $Se \geq Se^{crit}$, the h-form RE is chosen as the

governing equation; otherwise, the θ-form RE is preferred. The empirical effective saturation for doing switching is suggested

to 0.4-0.9.

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 $\psi_i = \theta_i$

and $\psi_{i+1} = h_{i+1}$ for example, the nodal fluxes calculated by Eqn. (5) for different forms of RE have to be carefully handled

by substituting $\theta_{i+1}^{j+1,k+1}$ with $\theta_{i+1}^{j+1,k}$, while $h_i^{j+1,k+1}$ is replaced by $h_i^{j+1,k}$. When $\psi_i = h_i$ and $\psi_{i+1} = \theta_{i+1}$, in contrast,

$h_{i+1}^{j+1,k+1}$ is replaced by $h_{i+1}^{j+1,k}$, while $\theta_i^{j+1,k+1}$ is replaced by $\theta_i^{j+1,k}$. The resulting equivalent nodal fluxes $q_{i+1/2}^h$ and

$q_{i+1/2}^\theta$ are then weighted to obtain an approximation by

$$q_{i+1/2} = (1-\omega)q_{i+1/2}^\theta + \omega \cdot q_{i+1/2}^h \tag{7}$$

where ω is the weight factor, $0 \leq \omega \leq 1$. In our work, $\omega = 0.5$ is applied to implicitly maintain the unknown variables of both

$h_{i+1}^{j+1,k+1}$ and $\theta_i^{j+1,k+1}$. Specifically, when $\omega = 1$, the h-form RE is used at both of nodes i and $i+1$; when $\omega = 0$, the θ-form RE

is employed instead. A detailed description of this approach can be found in Zeng et al. (2018).

Note that the equation switching method takes full advantages of both forms of *RE*, which is different from the traditional

160   primary variable switching schemes   (Diersch and Perrochet, 1999; Forsyth et al., 1995; Zha et al., 2013a). The switching-

RE formation is incorporated into a new HYDRUS package.

2.4 Multi-scale coupling approach

Coupling models at different scales requires dealing with the inconsistency in spatial and temporal discretization (Downer and

Ogden, 2004; Rybak et al., 2015). Space- and time-splitting strategies (see **Figure 1**) are adopted to conserve mass at two of

165   the scales. 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-

170   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; 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.

(1) Dirichlet boundary prediction

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

boundary

$$z_t(T) = \int_{s \in \Pi} H(T) ds \, / \int_{s \in \Pi} ds \tag{8}$$

where $z_t(T)$ [L] is the elevation of the water table; Π 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

180   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 2**a, the stepwise extension method ($z_t'(T^J)$) causes large deviation

from the "truth". The linear extrapolation is resorted to reduce the coupling errors and to accelerate the convergence of the

185   feedback iteration. The small-scale lower boundary head at time t ($T^J < t < T^{J+1}$) is given by

$$z_t(t) = \frac{(t - T^{J-1}) \cdot z_t(T^J) - (t - T^J) \cdot z_t(T^{J-1})}{T^J - T^{J-1}} \qquad (9)$$

(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 **Figure 2**b) 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 large-scale phreatic aquifer is numerically represented by the activated top layer in a two-dimensional groundwater model,

$$\overline{S}_y \frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(K\overline{M} \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y}\left(K\overline{M} \frac{\partial H}{\partial y} \right) + F_{top} - F_{base} \qquad (10)$$

where $\overline{M}$ [L] is the thickness of the phreatic layer, $\overline{M} = z_s - z_0$, z_0 is the bottom elevation of the top phreatic layer, $z_0 \ll z_s$; F_{top} [LT^{-1}] is the groundwater recharge into the activated top layer of the phreatic aquifer, $F_{top} = \left(K \cdot \partial H / \partial z \right)_{z=z_s}$; F_{base} is the leakage into an underlying numerical layer, $F_{base} = \left(K \cdot \partial H / \partial z \right)_{z=z_0}$ (positive downward, so as F_{base}). The long-term regional-scale parameter indicating the water yield caused by fluctuation of the water table (Nachabe, 2002), $\overline{S}_y$, [-], is calculated by

$$\overline{S}_y = V_w / (A \cdot \Delta H) \qquad (11)$$

where V_w [L^3] is the amount of water release caused by fluctuation of the phreatic surface (ΔH [L]); A [L^2] is the area of interest. At small spatial and temporal scales, e.g., within a macro step time $\Delta T = T^{J+1} - T^J$ and a local area of interest with thickness of $\overline{M} = z_s - z_0$, the specific storage term in Eqn. (1) is vertically integrated into a transient one-dimensional expression (Dettmann and Bechtold, 2016),

$$\tilde{S}_y = \left[w(T^{J+1}) - w(T^J) + \theta_s \cdot \Delta z_t \right] / \Delta z_t + \mu_s \cdot \overline{M} \qquad (12)$$

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

$$[q_{top} + l \cdot dz_t - q_{bot}] \cdot dt = w(t) - w(t - dt) + \theta_s \cdot dz_t \qquad (13)$$

210   where $q_{top}(t)$ and $q_{bot}(t)$ [LT^{-1}] are the nodal fluxes into and out of the moving balancing domain at a fixed top boundary (z_s)

and a moving bottom boundary ($z_b = \min(z_t(t), z_t(t-dt))$), $q_{top} = K(h)\cdot\partial(h+z)/\partial z\big|_{z=z_s}$, $q_{bot} = K(h)\cdot\partial(h+z)/\partial z\big|_{z=z_b}$

(positive into the balancing domain and negative outside); $dz_t = z_t(t) - z_t(t-dt)$ 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 2**b. Taking Γ as the lateral boundary

of a sub-domain, the lateral flux $l = \iiint\limits_{x,y,z\in\Omega}\left[\dfrac{\partial}{\partial x}\left(K\dfrac{\partial H}{\partial x}\right) + \dfrac{\partial}{\partial y}\left(K\dfrac{\partial H}{\partial y}\right)\right]dxdydz \Big/ \iiint\limits_{x,y,z\in\Omega} dxdydz$ is supposed to be

215   constant during ΔT; Ω is the volume of the saturated domain controlled by a soil column, which is horizontally projected into

Π. Temporally integrating Eqn. (13) from time T^J to T^{J+1} produces

$$R_{top} + \varepsilon_l - R_{bot} = w(T^{J+1}) - w(T^J) + \theta_s \cdot \Delta z_t \tag{14}$$

where R_{top} [L] is the cumulative water flux at z_s, $R_{top} = \int_{T^J}^{T^{J+1}} q_{top}(t)dt$, note that R_{top} is equals F_{top} in Eqn. (10); R_{bot} [L] is

the cumulative water flux out of the moving balancing domain, $R_{bot} = \int_{T^J}^{T^{J+1}} q_{bot}(t)dt$; ε_l [L] is the cumulative lateral

220   input water into the moving balancing domain,

$$\varepsilon_l = \frac{1}{2}l\cdot\sum_{j=1}^{N} dt^j dz_t^j \ll \varepsilon_l' = \frac{1}{2}l\cdot\Delta T\cdot\Delta z_t \tag{15}$$

where N is the number of time steps for the small-scale soil-water model within a macro time step ΔT; and ε_l' is the non-

trivial saturated later flux produced by a stationary boundary method (Seo et al., 2007; Xu et al., 2012). By taking R_{top} as the

specific recharge at z_s, the small-scale specific yield $\tilde{S}_y$ is derived from Eqns. (12) and (14) as

225   $$\tilde{S}_y = \left(R_{top} + \varepsilon_l - R_{bot}\right)/\Delta z_t + \mu_s \cdot M \tag{16}$$

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 ε_l towards zero. Thus ε_l can be neglected from the small-

scale mass balance analysis. In the developed model, the large-scale specific yield, $\overline{S}_y$ in Eqn. (11), represents the water

230   release in the phreatic aquifer; while the small-scale $\tilde{S}_y$ in Eqn. (16), denotes the dynamically changing water yield caused

by the fluctuation of the water table. The upper boundary flux F_{top} in the phreatic flow equation (Eqn. (10)) is therefore

corrected to

$$F_{top} = \left[R_{top} + (\overline{S}_y - \tilde{S}_y)\Delta z_t\right]/\Delta T \tag{17}$$

Differing from previous studies (Seo et al., 2007; Shen and Phanikumar, 2010; Xu et al., 2012), a scale-separation strategy is

235   employed in Eqn. (17). The specific yields at two different scales are linked explicitly by F_{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_{top}, are updated by

240

$$z_t^{updated} = \lambda \cdot z_t^{new} + (1-\lambda) \cdot z_t^{old}$$
$$F_{top}^{updated} = \lambda \cdot F_{top}^{new} + (1-\lambda) \cdot F_{top}^{old}$$

$$(18)$$

where superscript *old* (or *new*) indicates the previous (or newly calculated) head and flux boundaries at the coupling interface;

λ is the relaxation factor, $0 < \lambda \leq 1$. The iteration ends when agreements are reached at

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

$$(19)$$

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

245   **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 HYDRUS1D code (Šimůnek et al., 2009). For 2D and 3D problems, solutions by the

MODFLOW-VSF model (Thoms et al., 2006) are 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

250   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 ψ at time t is given by

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

$$(20)$$

where ψ is the numerical solution of either pressure head or water content; ψ^{ref} is the corresponding reference solution; C_i is

the control volume of node i ($i = 1, 2, \ldots, N$), Δz_i is the control volume of node i.

255   **3.1 Case 1: Rapidly changing atmospheric boundaries**

The 1D case is used to investigate the numerical efficiency of the multi-scale coupling scheme. 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.

260   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 alternately 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.

265   The coupled unsaturated model is discretized into a fine grid with Δz = 1 cm for solving the Richards' equation, which is bounded by the moving groundwater table. 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 Δz = 1 cm, and Δt = 0.05 d are taken as the "truth".

3.2 Case 2: Dynamic Groundwater flow

270   A 2D case is analyzed with sharp groundwater flow. Two pumping stresses are applied to the cross-sectional field with $x \times z$ = 5,000 cm × 1,000 cm. Well #1 is located at (x, z) = (2,500 cm, 0-200 cm), while well #2 is at (x, z) = (5,000 cm, 0-200 cm). The initial hydrostatic head of the cross-section is 700 cm. Soil #4 in Table 1 fills the entire cross-section. The total simulation lasts 50 days. For the coupled saturated model, as well as the reference model (MODFLOW-VSF (Thoms et al., 2006)), the cross-section is discretized horizontally into Δx = 50 cm, while vertically (top-down) into Δz = 5 cm$_{(\times 200)}$, 12.5 cm$_{(\times 4)}$, 25 cm$_{(\times 2)}$,

275   50 cm$_{(\times 2)}$, 100 cm$_{(\times 2)}$, and 200 cm$_{(\times 1)}$, where the subscripts hereinafter ($\times N$) are the numbers of discretized segments. The coupled 1D soil-water models are discretized with Δz = 1 cm.

3.3 Case 3: Pumping and irrigation

In case 3, a phreatic aquifer with $x \times y \times z$ = 1,000 m × 1,000 m × 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).

280   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 m³/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 Δx = Δy = 40 m for the coupled saturated model, as well as for the MODFLOW-VSF model for obtaining "truth" solution. The top-down thicknesses of the full-3D grid are Δz = 0.10 m$_{(\times 30)}$, 0.4 m$_{(\times 5)}$, 1 m$_{(\times 5)}$, and 2 m$_{(\times 5)}$. For the 1D soil columns, Δz = 0.1 m$_{(\times 30)}$, and 0.4 m$_{(\times 5)}$, which means no soil

285   column reaches the bottom. Different numbers of the sub-zones represented by soil columns, as well as their differing geometries, are given in **Figure 4**. The soil parameters for a sandy loam (soil #5) are given in Table 1. Total simulation lasts 60 days.

3.4 Case 4: Complex regional flow

A hypothetical test case from literature (Niswonger et al., 2006; Prudic et al., 2004; Twarakavi et al., 2008) for large-scale
290   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 5**a, 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 5**d) 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-
295   zones in **Figure 5**d, are discretized separately with a range of gradually refined thicknesses (Δz) from 30.48 cm, to 0.3048 cm
(bottom-up). A benchmark analysis is conducted by comparison with the solutions obtained from the original HYDRUS
package for MODFLOW (Seo et al., 2007).

4 Results and discussion

4.1 Reducing the complexity of a feedback coupling system

300   The numerical difficulty in a coupled unsaturated-saturated flow system originates from the non-linearity of the soil-water
hydraulic retention curves, 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 *θ-form REs* to reduce the nonlinearity in the soil-water models, and (2) smoothing the
variability of the exchanged interfacial messages.

305   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 6**. 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
310   larger minimal time-step sizes (Δt_{min}) and less computational cost, where minimal time-step size is the allowable 10^{-3} d for
convergence. Notably at the beginning of the sudden infiltration into a dry-sandy soil, in **Figure 6**a, the Δt_{min} for a switching
method is 10^{-3} d, even at early infiltration times, while for the *h-form* methods, including HYDRUS1D and the coupled *h-form*
method, Δt_{min} is constrained to 10^{-8} d before reaching a painstaking convergence. In **Figure 7**, 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, and
315   200 cm. To finish the entire calculation, the coupled *switching-form RE* method took 17 s, while it was 41 s 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.

320 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. (12), introduces significant variability to the large-scale groundwater model. A large-scale $\overline{S}_y$ reduces the non-linearity of the storage term in the groundwater equation. In case 1, using an $\overline{S}_y$ 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 $\overline{S}_y$, the non-linearity introduced by the small-scale soil-water models can be quickly smoothed, as shown in Eqn. (17).

325 **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 330 balance 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 (ΔT). However, the Dirichlet head predicted for the soil columns with a stepwise extension method (see **Figure 2**a), 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. (9). Here, we use Niter to indicate the maximal number of feedback iteration. Compared with a traditional stepwise method, the solution obtained by 335 a linear method, either iteratively (with Niter = 3) or non-iteratively (Niter = 0), is easier to approach the truth, see **Figure 8**. Even with refined macro time step sizes (ΔT 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.

340 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 345 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. (13). 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 9** 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 9**a 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_H for indicating the closure of the Neumann boundary feedback iteration, is usually related to the phreatic

Darcian flux. To clear 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 straight forward 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 10**a. 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 10**b. With Closure < 2, i.e., $\varepsilon_H < 2$ cm, the coupling errors are

380 significantly reduced, see **Figure 10**c. 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

385 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 11**a. Although portioning more sub-zones means higher accuracy for the coupling method, five or more soil columns are adequate enough to approximate

390 the "truth". Furthermore, for the saturated nodes deep in the aquifer, such coupling errors are of minor influence, see **Figure 11**b.

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 12**c. When partitioning the vadose zone into more than 12 soil columns, there is a slight reduction in solution errors (RMSE) and a

395 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 12**a 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

400 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

405   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 13**a gives the contours for the final phreatic head solutions, indicating a good match of the phreatic surface

        with the HYDRUS package. **Figure 13**b 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.

410   For unsaturated-saturated flow situations, the vadose zone flow is important. **Figure 14** presents the water content profiles at

        sub-zones 1, 3, 5, 7, and 9 as examples. The solutions obtained from the unsaturated models match 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

420   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 θ and h forms of Richards' equation (*RE*) at each

425   numerical node. The *switching-RE* circumvents the disadvantages of the h- and θ-*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 (Δt_{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.

430   (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

435   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

440   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

445   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.

Author contribution: Jicai Zeng, Yuanyuan Zha and Jinzhong Yang developed the new package for soil water movement based

450   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 for his laborious revisions and helpful suggestions to the paper.

455   References

Bailey, R. T., Morway, E. D., Niswonger, R. G. and Gates, T. K.: Modeling variably saturated multispecies reactive
     groundwater solute transport with MODFLOW-UZF and RT3D, Groundwater, 51(5), 752–761, doi:10.1111/j.1745-
     6584.2012.01009.x, 2013.

Celia, M. A., Bouloutas, E. T. and Zarba, R. L.: A general mass-conservative numerical solution for the unsaturated flow

460   equation, Water Resour. Res., 26(7), 1483–1496, doi:10.1029/WR026i007p01483, 1990.

Crevoisier, D., Chanzy, A. and Voltz, M.: Evaluation of the Ross fast solution of Richards' equation in unfavourable conditions for standard finite element methods, Adv. Water Resour., 32(6), 936–947, doi:10.1016/j.advwatres.2009.03.008, 2009.

van Dam, J. C., Groenendijk, P., Hendriks, R. F. A. and Kroes, J. G.: Advances of Modeling Water Flow in Variably Saturated Soils with SWAP, Vadose Zo. J., 7(2), 640, doi:10.2136/vzj2007.0060, 2008.

465   Dettmann, U. and Bechtold, M.: One-dimensional expression to calculate specific yield for shallow groundwater systems with microrelief, Hydrol. Process., 30(2), 334–340, doi:10.1002/hyp.10637, 2016.

Diersch, H. J. G. and Perrochet, P.: On the primary variable switching technique for simulating unsaturated-saturated flows, Adv. Water Resour., 23(3), 271–301, doi:10.1016/S0309-1708(98)00057-8, 1999.

Downer, C. W. and Ogden, F. L.: Appropriate vertical discretization of Richards' equation for two-dimensional watershed-
470   scale modelling, Hydrol. Process., 18(1), 1–22, doi:10.1002/hyp.1306, 2004.

Edwards, M. G.: Elimination of Adaptive Grid Interface Errors in the Discrete Cell Centered Pressure Equation, J. Comput. Phys., 126(2), 356–372, doi:10.1006/jcph.1996.0143, 1996.

Forsyth, P. A., Wu, Y. S. and Pruess, K.: Robust numerical methods for saturated-unsaturated flow with dry initial conditions in heterogeneous media, Adv. Water Resour., 18(1), 25–38, doi:10.1016/0309-1708(95)00020-J, 1995.

475   van Genuchten, M. T.: A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1, Soil Sci. Soc. Am. J., 44(5), 892, doi:10.2136/sssaj1980.03615995004400050002x, 1980.

Grygoruk, M., Batelaan, O., Mirosław-Świltek, D., Szatyłowicz, J. and Okruszko, T.: Evapotranspiration of bush encroachments on a temperate mire meadow - A nonlinear function of landscape composition and groundwater flow, Ecol. Eng., 73, 598–609, doi:10.1016/j.ecoleng.2014.09.041, 2014.

480   Gunduz, O. and Aral, M. M.: River networks and groundwater flow: A simultaneous solution of a coupled system, J. Hydrol., 301(1–4), 216–234, doi:10.1016/j.jhydrol.2004.06.034, 2005.

Harbaugh, A. W.: MODFLOW-2005, the U.S. Geological Survey modular ground-water model: The ground-water flow process, US Department of the Interior, US Geological Survey Reston, VA, USA., 2005.

Hills, R. G., Porro, I., Hudson, D. B. and Wierenga, P. J.: Modeling one-dimensional infiltration into very dry soils: 1. Model
485   development and evaluation, Water Resour. Res., 25(6), 1259–1269, doi:10.1029/WR025i006p01259, 1989.

Krabbenhøft, K.: An alternative to primary variable switching in saturated-unsaturated flow computations, Adv. Water Resour., 30(3), 483–492, doi:10.1016/j.advwatres.2006.04.009, 2007.

Kumar, M., Duffy, C. J. and Salvage, K. M.: A Second-Order Accurate, Finite Volume–Based, Integrated Hydrologic Modeling (FIHM) Framework for Simulation of Surface and Subsurface Flow, Vadose Zo. J., 8(4), 873, doi:10.2136/vzj2009.0014,

490   2009.

Kuznetsov, M., Yakirevich, A., Pachepsky, Y. A., Sorek, S. and Weisbrod, N.: Quasi 3D modeling of water flow in vadose zone
         and groundwater, J. Hydrol., 450–451, 140–149, doi:10.1016/j.jhydrol.2012.05.025, 2012.

Langevin, C. D., Hughes, J. D., Banta, E. R., Provost, A. M., Niswonger, R. G. and Panday, S.: MODFLOW 6 Groundwater
         Flow (GWF) Model Beta version 0.9.03, U.S. Geol. Surv. Provisional Softw. Release, doi:10.5066/F76Q1VQV, 2017.

495 Leake, S. A. and Claar, D. V: Procedures and computer programs for telescopic mesh refinement using MODFLOW, Citeseer.
         [online] Available from: http://az.water.usgs.gov/MODTMR/tmr.html, 1999.

Lin, L., Yang, J.-Z., Zhang, B. and Zhu, Y.: A simplified numerical model of 3-D groundwater and solute transport at large
         scale area, J. Hydrodyn. Ser. B, 22(3), 319–328, doi:10.1016/S1001-6058(09)60061-5, 2010.

Liu, Z., Zha, Y., Yang, W., Kuo, Y. and Yang, J.: Large-Scale Modeling of Unsaturated Flow by a Stochastic Perturbation
500   Approach, Vadose Zo. J., 15(3), 20, doi:10.2136/vzj2015.07.0103, 2016.

Markstrom, S. L., Niswonger, R. G., Regan, R. S., Prudic, D. E. and Barlow, P. M.: GSFLOW—Coupled Ground-Water and
         Surface-Water Flow Model Based on the Integration of the Precipitation-Runoff Modeling System (PRMS) and the
         Modular Ground-Water Flow Model (MODFLOW-2005), Geological Survey (US)., 2008.

Maxwell, R. M. and Miller, N. L.: Development of a coupled land surface and groundwater model, J. Hydrometeorol., 6(3),
505   233–247, doi:10.1175/JHM422.1, 2005.

McDonald, M. G. and Harbaugh, A. W.: A modular three-dimensional finite-difference ground-water flow model. [online]
         Available from: http://pubs.er.usgs.gov/publication/twri06A1, 1988.

Mehl, S. and Hill, M. C.: Three-dimensional local grid refinement for block-centered finite-difference groundwater models
         using iteratively coupled shared nodes: a new method of interpolation and analysis of errors, Adv. Water Resour., 27(9),
510   899–912, doi:10.1016/j.advwatres.2004.06.004, 2004.

Mehl, S. and Hill, M. C.: MODFLOW–LGR—Documentation of ghost node local grid refinement (LGR2) for multiple areas
         and the boundary flow and head (BFH2) package, 2013th ed., 2013.

Miller, C. T., Williams, G. A., Kelley, C. T. and Tocci, M. D.: Robust solution of Richards' equation for nonuniform porous
         media, Water Resour. Res., 34(10), 2599–2610, doi:10.1029/98WR01673, 1998.

515 Miller, C. T., Abhishek, C. and Farthing, M. W.: A spatially and temporally adaptive solution of Richards' equation, Adv. Water
         Resour., 29(4), 525–545, doi:10.1016/j.advwatres.2005.06.008, 2006.

Nachabe, M. H.: Analytical expressions for transient specific yield and shallow water table drainage, Water Resour. Res.,
         38(10), 11-1-11–7, doi:10.1029/2001WR001071, 2002.

Niswonger, R. G., Prudic, D. E. and Regan, S. R.: Documentation of the Unsaturated-Zone Flow (UZF1) Package for Modeling

520   Unsaturated Flow Between the Land Surface and the Water Table with MODFLOW-2005, US Department of the

        Interior, US Geological Survey., 2006.

    Niswonger, R. G., Panday, S. and Ibaraki, M.: MODFLOW-NWT, a Newton formulation for MODFLOW-2005, US Geol.

        Surv. Tech. Methods, 6, A37, 2011.

    Panday, S. and Huyakorn, P. S.: A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface

525   flow, Adv. Water Resour., 27(4), 361–382, doi:10.1016/j.advwatres.2004.02.016, 2004.

    Panday, S., Langevin, C. D., Niswonger, R. G., Ibaraki, M. and Hughes, J. D.: MODFLOW–USG version 1: An unstructured

        grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume

        finite-difference formulation, US Geological Survey., 2013.

    Paulus, R., Dewals, B. J., Erpicum, S., Pirotton, M. and Archambeau, P.: Innovative modelling of 3D unsaturated flow in

530   porous media by coupling independent models for vertical and lateral flows, J. Comput. Appl. Math., 246, 38–51,

        doi:10.1016/j.cam.2012.07.032, 2013.

    Prudic, D. E., Konikow, L. F. and Banta, E. R.: A New Streamflow-Routing (SFR1) Package to Simulate Stream-Aquifer

        Interaction with MODFLOW-2000. [online] Available from: http://pubs.er.usgs.gov/publication/ofr20041042, 2004.

    Richards, L. A.: Capillary conduction of liquids through porous mediums, J. Appl. Phys., 1(5), 318–333,

535   doi:10.1063/1.1745010, 1931.

    Rybak, I., Magiera, J., Helmig, R. and Rohde, C.: Multirate time integration for coupled saturated/unsaturated porous medium

        and free flow systems, Comput. Geosci., 19(2), 299–309, doi:10.1007/s10596-015-9469-8, 2015.

    Seo, H., Šimůnek, J. and Poeter, E.: Documentation of the HYDRUS package for MODFLOW-2000, the US Geological Survey

        modular ground-water model, Gr. Water Model. Ctr., Color. Sch. Mines, Golden, (1980), 1–98, 2007.

540 Shen, C. and Phanikumar, M. S.: A process-based, distributed hydrologic model based on a large-scale method for surface -

        subsurface coupling, Adv. Water Resour., 33(12), 1524–1541, doi:10.1016/j.advwatres.2010.09.002, 2010.

    Šimůnek, J., Van Genuchten, M. T. and Sejna, M.: The HYDRUS-1D software package for simulating the one-dimensional

        movement of water, heat, and multiple solutes in variably-saturated media, Univ. California-Riverside Res. Reports, 3,

        1–240, 2005.

545 Šimůnek, J., Šejna, M., Saito, H., Sakai, M. and van Genuchten, M. T.: The HYDRUS-1D software package for simulating the

        movement of water, heat, and multiple solutes in variably saturated media, version 4.0, HYDRUS software series 3,

        Dep. Environ. Sci. Univ. Calif. Riverside, Riverside, California, USA, 315, 2008.

    Šimůnek, J., Sejna, M., Saito, H., Sakai, M. and van Genuchten, M. T.: The HYDRUS-1D software package for simulating the

        one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 4.08., 2009.

550   Šimůnek, J., Van Genuchten, M. T. and Šejna, M.: Recent Developments and Applications of the HYDRUS Computer Software
        Packages, Vadose Zo. J., 15(7), doi:10.2136/vzj2016.04.0033, 2016.

Stoppelenbrug, F. J., Kovar, K., Pastoors, M. J. H. and Tiktak, A.: Modelling the interactions between transient saturated and
        unsaturated groundwater flow. Off-line coupling of LGM and SWAP, RIVM Rep., 500026001, 70, 2005.

Thoms, R. B., Johnson, R. L. and Healy, R. W.: User's guide to the variably saturated flow (VSF) process to MODFLOW.
555   [online] Available from: http://pubs.er.usgs.gov/publication/tm6A18, 2006.

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.

Vogel, H.-J. and Ippisch, O.: Estimation of a Critical Spatial Discretization Limit for Solving Richards' Equation at Large
        Scales, Vadose Zo. J., 7(1), 112, doi:10.2136/vzj2006.0182, 2008.

560   Warrick, A. W.: Numerical approximations of darcian flow through unsaturated soil, Water Resour. Res., 27(6), 1215–1222,
        doi:10.1029/91WR00093, 1991.

Xie, Z., Di, Z., Luo, Z. and Ma, Q.: A Quasi-Three-Dimensional Variably Saturated Groundwater Flow Model for Climate
        Modeling, J. Hydrometeorol., 13(1), 27–46, doi:10.1175/JHM-D-10-05019.1, 2012.

Xu, X., Huang, G., Zhan, H., Qu, Z. and Huang, Q.: Integration of SWAP and MODFLOW-2000 for modeling groundwater
565   dynamics in shallow water table areas, J. Hydrol., 412–413, 170–181, doi:10.1016/j.jhydrol.2011.07.002, 2012.

Yakirevich, A., Borisov, V. and Sorek, S.: A quasi three-dimensional model for flow and transport in unsaturated and saturated
        zones: 1. Implementation of the quasi two-dimensional case, Adv. Water Resour., 21(8), 679–689, doi:10.1016/S0309-
        1708(97)00031-6, 1998.

Zeng, J., Zha, Y., Zhang, Y., Shi, L., Zhu, Y. and Yang, J.: On the sub-model errors of a generalized one-way coupling scheme
570   for linking models at different scales, Adv. Water Resour., 109, 69–83, doi:10.1016/j.advwatres.2017.09.005, 2017.

Zeng, J., Zha, Y. and Yang, J.: Switching the Richards' equation for modeling soil water movement under unfavorable
        conditions, J. Hydrol., doi:10.1016/j.jhydrol.2018.06.069, 2018.

Zha, Y., Shi, L., Ye, M. and Yang, J.: A generalized Ross method for two- and three-dimensional variably saturated flow, Adv.
        Water Resour., 54, 67–77, doi:10.1016/j.advwatres.2013.01.002, 2013a.

575   Zha, Y., Yang, J., Shi, L. and Song, X.: Simulating One-Dimensional Unsaturated Flow in Heterogeneous Soils with Water
        Content-Based Richards Equation, *Vadose Zo. J.*, 12(2), 13, doi:10.2136/vzj2012.010, 2013b.

Zha, Y., Yang, J., Yin, L., Zhang, Y., Zeng, W. and Shi, L.: A modified Picard iteration scheme for overcoming numerical
        difficulties of simulating infiltration into dry soil, *J. Hydrol.*, 551, 56–69, doi:10.1016/j.jhydrol.2017.05.053, 2017.

Zhu, Y., Shi, L., Lin, L., Yang, J. and Ye, M.: A fully coupled numerical modeling for regional unsaturated-saturated water

580   flow, *J. Hydrol.*, 475, 188–203, doi:10.1016/j.jhydrol.2012.09.048, 2012.

582 **Table 1**   Soil parameters used in the test cases.

#	Soil	θ_r	θ_s	α [1/cm]	n	k_s [cm/d]
1	Sand	0.093	0.301	0.0547	4.264	504
2	Berino loamy fine sand	0.029	0.366	0.028	2.239	541
3	Glendale clay loam	0.106	0.469	0.010	1.395	13.1
4	Loam	0.078	0.430	0.036	1.560	24.96
5	Sandy loam	0.065	0.410	0.075	1.890	106.1

583

584 **Table 2**   The precipitation, evaporation, and pumping rates in 12 stress periods.

Stress period	Precipitation (mm/d)	ET (mm/d)	Pumping rate (m^3/d)
1	0.21	1.32	4078
2	1.69	1.32	4078
3	2.11	1.32	2039
4	4.21	1.32	2039
5	1.05	1.32	6116
6	2.11	1.32	0
7	0.63	1.32	4078
8	1.05	1.32	0
9	0.63	1.32	2039
10	0.42	1.32	0
11	0.21	1.32	6116
12	0.21	1.32	0

585
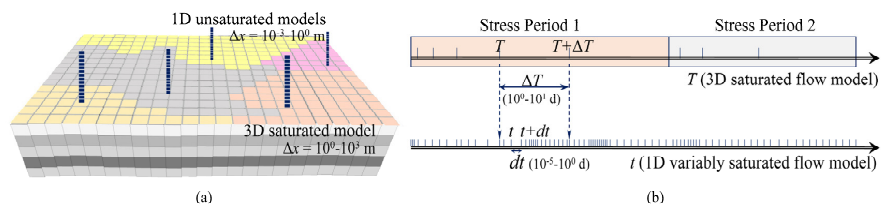
Figure 1: Schematic of the space- and time-splitting strategy for coupling models at two independent scales. J (or j), T (or t), and ΔT (or dt) are time level, time, and time-step size at coarse (or fine) scale. Δx is the resolution for spatial discretization.

586
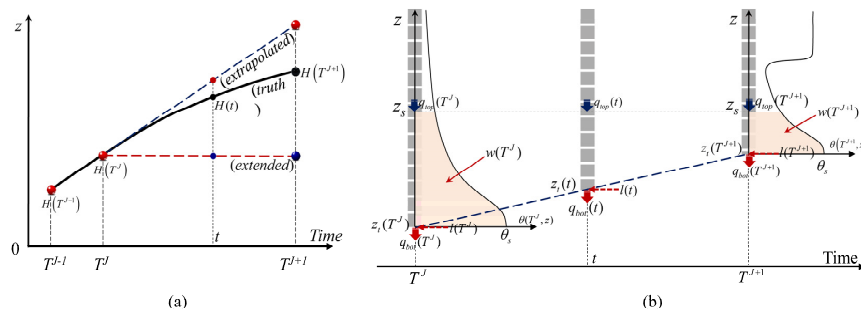
Figure 2: The Dirichlet-Neumann coupling of the soil-water and groundwater flow models at different scales. (a) Dirichlet lower boundary of the soil water flow model. (b) Water balance analysis based on a balancing domain with moving lower boundary.

587
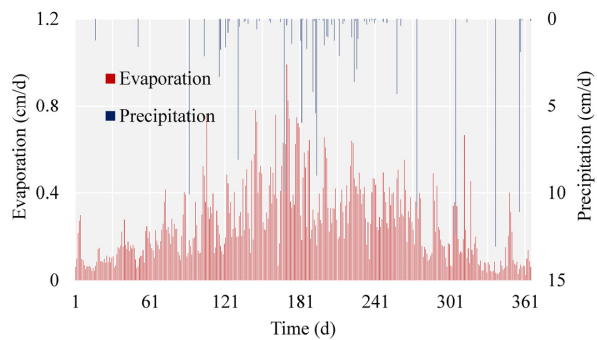
Figure 3: Rapidly changing atmospheric upper boundary conditions (q_{top} = Precipitation − Evaporation) for scenario 2, case 1.

588
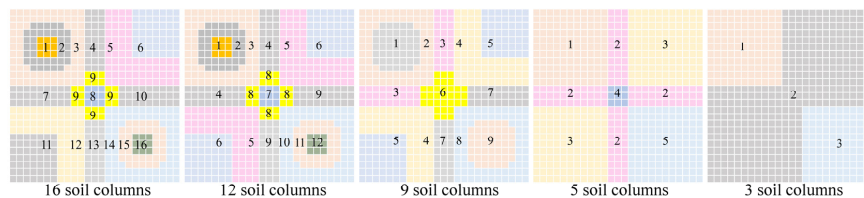
Figure 4: Different number of sub-zones partitioned for the quasi-3D simulations in Case 3. The overall vadose zone is partitioned into 16, 12, 9, 5, and 3 sub-zones.

589
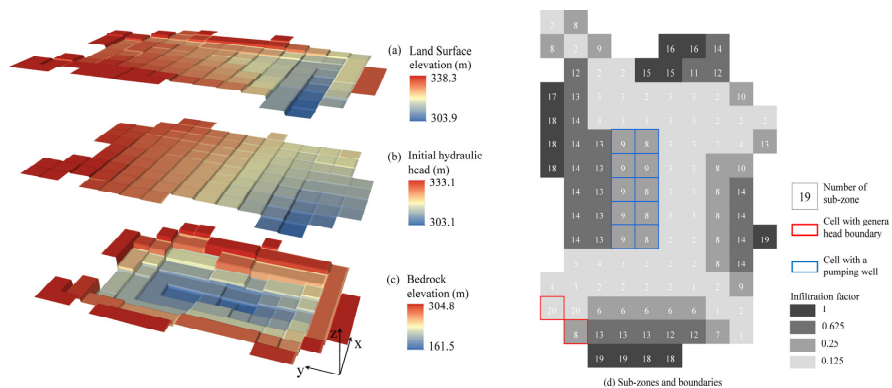
Figure 5: Input of the synthetic regional problem including (a) land surface elevation, (b) initial head, (c) bedrock elevation of the aquifer, and (d) the sub-zones and boundaries.

590
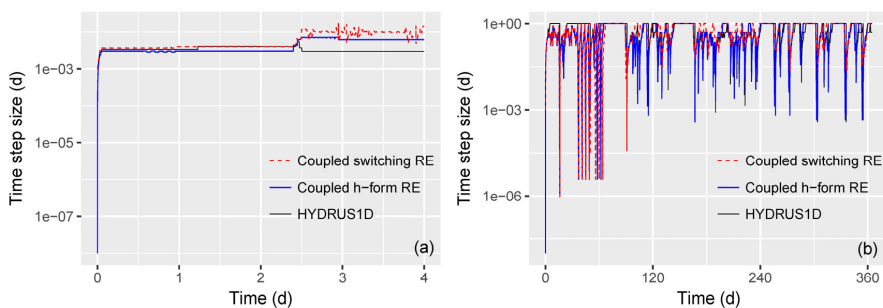
Figure 6: The time-step sizes through the simulation of (a) sudden infiltration into a dry-sandy soil column, and (b) rapidly changing atmospheric upper boundary conditions with a layered soil column.

591
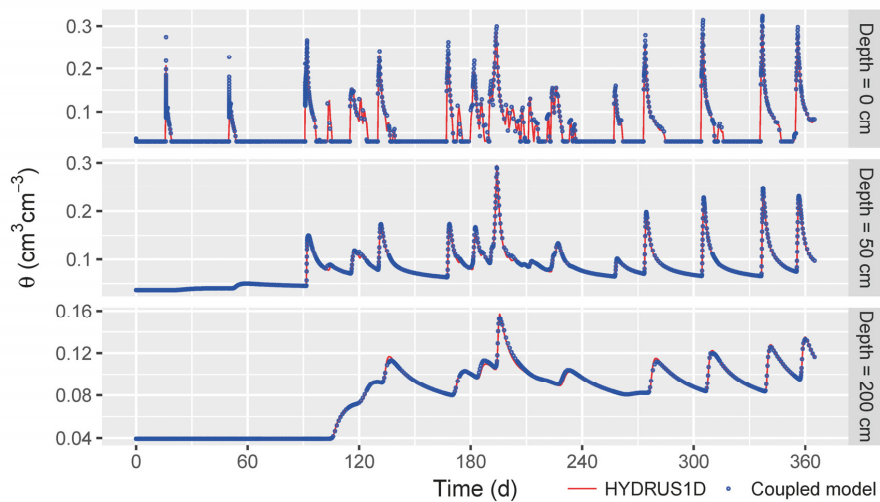
Figure 7: The observed soil moisture content at $z = 0$ cm, 50 cm, and 200 cm for the layered soil column with rapidly changing upper boundary conditions (Scenario 2, Case 1).

592
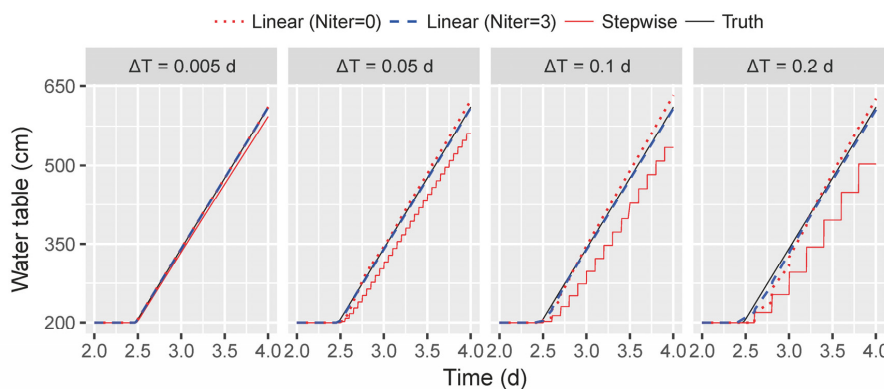
Figure 8: Water table changing with time for different macro time step sizes (ΔT = 0.005 d, 0.05 d, 0.1 d, and 0.2 d), in scenario 1, case 1. The HYDRUS1D solution is taken as the "truth". Compared with the stepwise extended method (Seo et al., 2007), the cumulative mass balance error is significantly reduced by a linear prediction.

593
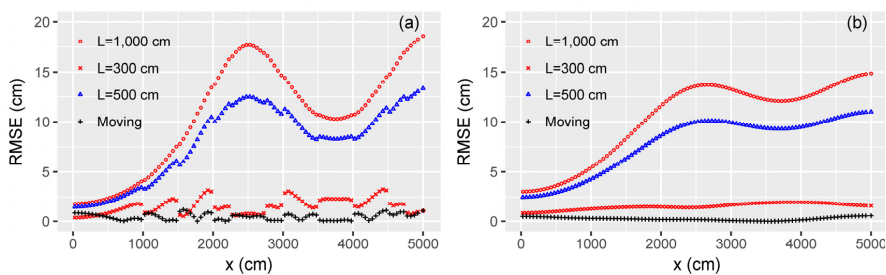
Figure 9: Comparison of RMSE of (a) the phreatic surface and (b) the head solution (at $z = 0$) between the moving-boundary and the stationary-boundary methods. Three different lengths of the stationary soil columns, L = 1,000 cm, 500 cm, and 300 cm, are considered.

594
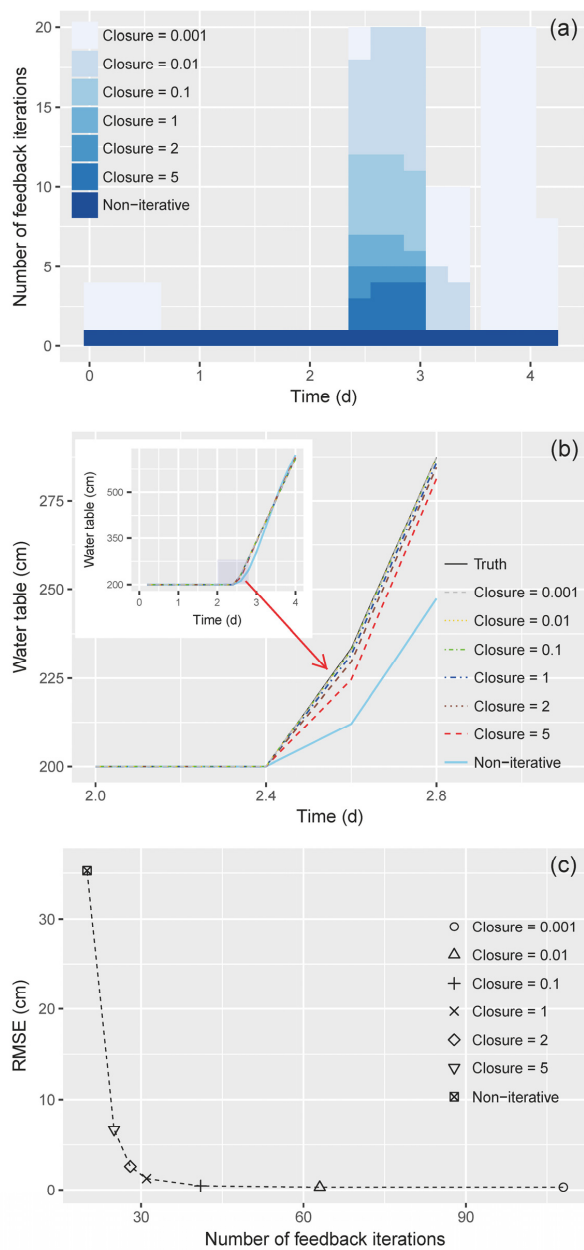
Figure 10: (a) The number of feedback iterations and (b) phreatic surface solution changing with different closure criteria. The legend "Closure = 0.001" means $\varepsilon_H = 0.001$ cm, and $\varepsilon_F = 0.001$ cm/d. The HYDRUS1D solution is taken as "truth". Tested in scenario 1, case 1.

595
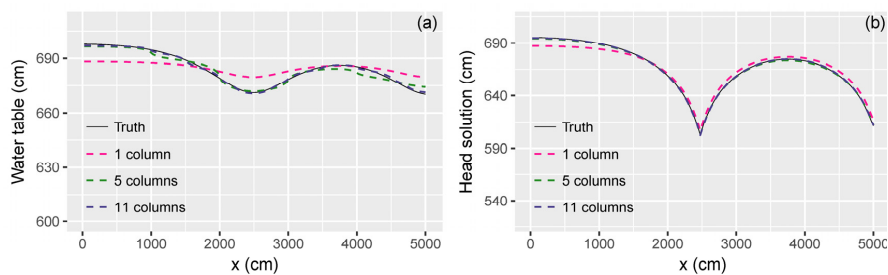
Figure 11: Comparison of (a) Phreatic surface and (b) head solution (at $z = 0$) changing by the number of soil columns. Solutions obtained with a moving-boundary method in case 2.

596
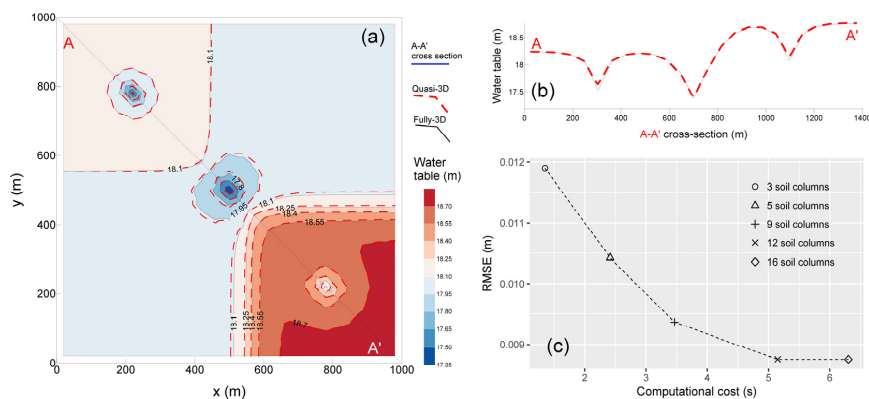
Figure 12: (a) Comparison of contours of the phreatic surface solution obtained with the fully-3D and quasi-3D methods; (b) Comparison of the phreatic surface at A-A' cross-section; (c) computational cost and RMSE changing by the number of total soil columns.

597
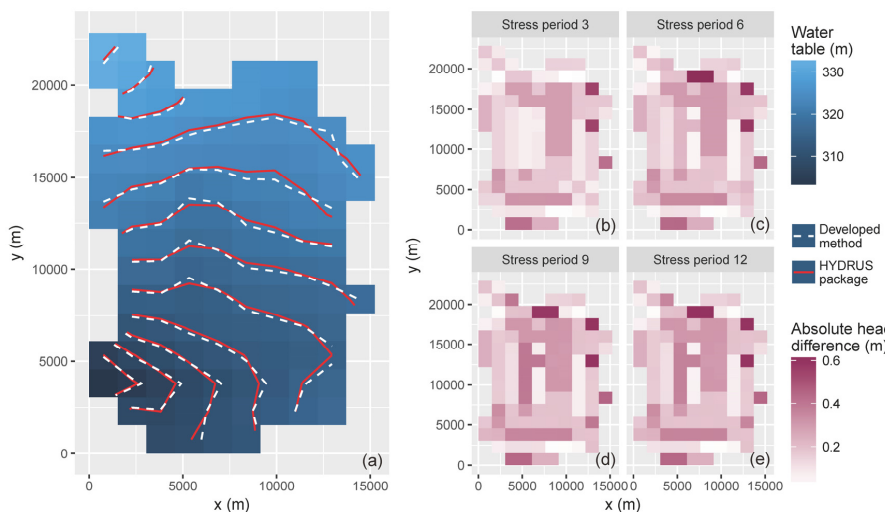
Figure 13: (a) Comparison of elevation of the water table calculated by the HYDRUS package for MODFLOW (Seo et al., 2007) and the developed method ($t = 365$ d); (b) The absolute head difference of the phreatic head solution by the method developed here and HYDRUS package at the end of stress periods 3, 6, 9, and 12. (Case 4).

598
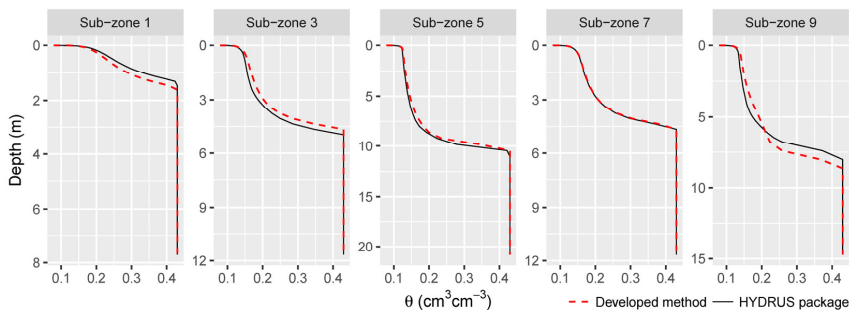
Figure 14: Comparison of water content profiles obtained from the HYDRUS package for MODFLOW (Seo et al., 2007) and the developed iterative feedback coupling method. Sub-zones 1, 5, 9, 13, and 20 are shown as an example. (t = 365 d in Case 4).

599