Estimating Groundwater Recharge in Fully Integrated *pde*-Based Hydrological Models

Bastian Waldowski[1], Emilio Sánchez-León[2], Olaf A. Cirpka[3], Natascha Brandhorst[1], Harrie-Jan Hendricks Franssen[4,5], Insa Neuweiler[1]

[1]Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, University of Hannover, Germany
[2]Bayerisches Landesamt für Umwelt, Hof/Saale, Germany
[3]Department of Geosciences, University of Tübingen, Germany
[4]Forschungszentrum Jülich GmbH, Agrosphere (IBG-3), Germany
[5]Centre for High-Performance Scientific Computing in Terrestrial Systems, Jülich, Germany

Key Points:

- We introduce a method to calculate groundwater recharge efficiently and consistently in fully integrated hydrological models.
- We estimate contributions to groundwater recharge based on groundwater-storage changes and the flux crossing the water table.
- We emphasize the importance of numerical models to capture the spatially and temporally complex nature of groundwater recharge.

Corresponding author: Bastian Waldowski, waldowski@hydromech.uni-hannover.de

Abstract

Groundwater recharge is the main forcing of regional groundwater flow. In traditional partial-differential-equation (*pde*)-based models that treat aquifers as separate compartments, groundwater recharge needs to be defined as a boundary condition or it is a coupling condition to other compartments. Integrated models that treat the vadose and phreatic zones as a continuum allow for a more sophisticated calculation of subsurface fluxes, as feedbacks between both zones are captured. However, they do not contain an explicit groundwater-recharge term so it needs to be estimated by post-processing. Groundwater recharge consists of the flux crossing the water table, which can be calculated based on hydraulic gradients, and of changes in groundwater storage. We introduce a method to evaluate the change of groundwater storage by a time-cumulative water balance over the depth section of water table fluctuations, avoiding the use of a specific yield. We demonstrate the approach first by a simple 1-D vertical model that does not allow for lateral outflow and illustrates the ambiguity of computing groundwater recharge by different methods. We then apply the approach to a 3-D model with a complex topography and subsurface structure. The latter example shows that groundwater recharge is highly variable in space and time with notable differences between regional and local estimates. Local heterogeneity of topography or subsurface properties results in complex redistribution patterns of groundwater. In fully integrated models, river-groundwater exchange flow may severely bias the estimate of groundwater recharge. We therefore advice masking out groundwater recharge at river locations.

1 Introduction

Groundwater is the largest available freshwater resource on earth. Aquifer-scale groundwater flow critically depends on groundwater recharge. Its accurate estimation is thus crucial for quantitative and qualitative groundwater studies. While groundwater recharge is highly variable in time and space (Healy, 2010), it is rarely possible to capture both scales with a high resolution. Water-resources-assessment studies (e.g., Kearns & Hendrickx, 1998; Höglauer et al., 2019) commonly focus on the temporal behavior of groundwater recharge while its small-scale spatial variability is often neglected. Detailed information about spatial patterns of groundwater recharge can, however, also be of interest in quantitative studies, for instance when planning the placement of extraction wells. In groundwater-quality studies, the accurate assessment of the spatial distribution of recharge is indispensable. High-recharge areas are of particular interest because here the aquifers are most vulnerable to contamination (US National Research

Council, 1993; De Vries & Simmers, 2002; Healy, 2010). Such information can be used in the delineation of water-protection zones or in the designation of waste-disposal sites. Temporal patterns of groundwater recharge also matter when evaluating groundwater quality, since temporal extremes of groundwater recharge are crucial with respect to aquifer vulnerability.

Groundwater recharge occurs at the transition between the *vadose zones* (parts of the subsurface holding *soil water*) and *phreatic zones* (parts of the subsurface holding *groundwater*). These zones differ with respect to flow behavior as well as applications. In contrast to groundwater, soil water is not directly accessible by a well, but it is essential with respect to water availability to plants. Water movement in the vadose zone is predominantly vertical, with typical mean specific-discharge values that are a fraction of mean precipitation (on the order of decimeters per year), whereas groundwater flow is predominantly horizontal with typical specific-discharge values on the order of decimeters per day. Biogeochemical processes also differ significantly because of inter-phase mass transfer with soil gas in the vadose zone, which is lacking in groundwater altogether. As these two subsurface compartments differ quite a lot, it is often useful to consider them as separate compartments in the assessment of both water quantity and quality. For example, there is a significant time delay between a water parcel infiltrating at the land surface and its arrival in the groundwater body, which matters in water-quality assessment. At the transition between the phreatic and vadose zones is the capillary fringe, a fully saturated layer with negative pressure heads. While it shows a flow behavior similar to groundwater, the capillary fringe is often not considered as part of groundwater as it is not detected by a piezometer when measuring the position of the water table. In the absence of a capillary fringe, the phreatic zone is synonymous with the saturated zone.

Despite a common intuitive agreement on what is meant by groundwater recharge, it lacks a universal definition (e.g., Barthel, 2006). Freeze (1969, p. 153-154) defined it as the ”entry into the saturated zone of water made available at the water table surface, together with the associated flow away from the water table within the saturated zone” and Lerner et al. (1990, p. 6) defined it as ”the downward flow of water reaching the water table, forming an addition to the groundwater reservoir”. Similar definitions were made by Meinzer and Stearns (1929), Sophocleous (1991) and Healy (2010). These definitions indicate that groundwater recharge represents the downward flow of water at the water table, as well as storage changes of the groundwater reservoir. We emphasize that downward flow at the water table and changes of the groundwater storage are not synonyms. Changes in groundwater storage at a certain location may show a highly variable and seasonal behavior, affecting recharge estimates on small

time scales, but are often negligible in the annual average (Szilagyi et al., 2003; Healy, 2010). Therefore, methods that estimate recharge based on groundwater-storage changes are often conducted on an event basis (Crosbie et al., 2005).

Long-term averages of groundwater recharge are balanced by groundwater outflow, which refers to the lateral onward flow of groundwater inside the aquifer. Ultimately, this water may flow into another aquifer or into surface-water bodies. In several case studies, groundwater outflow is restricted to the flow into rivers, often called base-flow (e.g., Szilagyi et al., 2003; Lee et al., 2006) or groundwater runoff (e.g., Meinzer & Stearns, 1929). Long-term recharge studies (e.g., Szilagyi et al., 2003; Lee et al., 2006) usually focus on this steady component of groundwater recharge, neglecting storage changes in the groundwater reservoir. For a complete assessment of groundwater recharge, both this steady recharge process and changes in groundwater storage need to be considered.

We define *groundwater recharge* as the flux that affects groundwater storage and causes groundwater outflow at a specific location. In case of upwards directed flow, the sign of this flux would be negative. Since the concept of negative groundwater recharge is not common, we refer to such fluxes as *capillary rise* throughout this work. This includes also upwards flow in cases where the water table is at the land surface so that the flow would rather be addressed as free seepage. We define downward fluxes occurring at depths above the water table as *potential recharge*. These fluxes are sometimes denoted drainage (Healy, 2010) or net infiltration (Heilweil et al., 2006). In surface hydrology and soil science, the distinction between potential recharge and actual groundwater recharge is sometimes not made (e.g., Fayer et al., 1996; Keese et al., 2005), which can cause an overestimation of groundwater recharge (e.g., Yin et al., 2011).

1.1 Measurement of Groundwater Recharge

Measuring groundwater recharge has been considered a difficult task (Healy, 2010). Direct measurements at the water table are often unfeasible and spatial coverage is commonly low. Lysimeters are stationary and expensive (Lerner et al., 1990) and provide only an estimate of potential recharge (Barthel, 2006; Healy, 2010).

Tracer methods for estimating groundwater recharge are often based on strategies of groundwater-age dating, which in turn usually rely on isotopes (such as oxygen-18 or tritium) or gases that can be traced back to industrial/anthropogenic activities (such as chlorofluorocarbons or sulfur hexafluoride). One may differentiate between unsaturated-zone and groundwater tracer meth-

ods. Unsaturated-zone tracer methods (e.g., Brunner et al., 2004; Wang et al., 2008; Li et al., 2017) potentially provide high spatial coverage, but often obtain only estimates of potential recharge, which in turn rely on simplifications regarding the movement of water and tracers in the vadose zone (Healy, 2010). Groundwater-tracer methods (e.g., Schlosser et al., 1989; Solomon & Sudicky, 1991; McMahon et al., 2011; Qin et al., 2011) give a more direct estimate of groundwater recharge, but are usually not applicable at a high spatial resolution (Healy, 2010).

Another common approach of estimating groundwater recharge consists of measuring water budget components. Here groundwater recharge is commonly approximated either as a residual of the water budget or as the change in groundwater storage over time. Estimating groundwater recharge as a residual of the water budget is prone to high uncertainty as the variance of the residual terms equals the sum of variances of all other contributions to the water budget.

Changes of groundwater storage to estimate groundwater recharge have been derived from measurements of the water table depth for almost a century (e.g., Meinzer & Stearns, 1929) and are still used today (e.g., Labrecque et al., 2020). While such approaches benefit from their simplicity and a relatively small number of input parameters, they are usually based on simplifying assumptions regarding the vadose zone (Childs, 1960) and lateral groundwater outflow is usually neglected (Lerner et al., 1990). Furthermore, the spatial coverage of such estimates is limited by the number and distribution of observation wells.

If adequate observation methods are available, more reliable estimates of groundwater recharge can be obtained if multiple independent methods are applied (Lerner et al., 1990; De Vries & Simmers, 2002; Hendricks Franssen et al., 2008). In addition to measurement-based estimations, groundwater recharge can be computed by numerical models. While such models simplify reality, they allow a more detailed analysis of the water movement than measurements and provide high temporal and spatial resolution given the necessary inputs.

1.2 Numerical Simulation of Groundwater Recharge

In numerical models that consider the vadose and phreatic zones as separate compartments, groundwater recharge is usually defined as the flux leaving the vadose zone. This flux is required for coupling of the compartments and it may result in an increase in groundwater storage as well as lateral groundwater outflow, so it is consistent with our definition of ground-

water recharge. While a modeling strategy that separates the vadose and phreatic zones benefits from computational efficiency and simplicity, "the focus on isolated components within what we know to be an interconnected hydrologic system is a limitation that can only be addressed with an integrated approach" (Maxwell et al., 2015, p. 924).

Fully integrated partial-differential-equation (*pde*)-based models such as ParFlow (Jones & Woodward, 2001; Kollet & Maxwell, 2006; Maxwell, 2013), HydroGeoSphere (Therrien et al., 2010; Brunner & Simmons, 2012), CATHY (Camporese et al., 2010) or OpenGeoSys (Kolditz et al., 2012) consider the water-saturated and -unsaturated parts of the subsurface as a continuum (in which water flow can be described by Richards' equation) and explicitly account for exchange fluxes between the land surface and the subsurface. These models require no explicit evaluation of groundwater recharge because the entire subsurface is treated as a single continuum and a distinction between the vadose and phreatic zones is not required. Still, in post-processing, this distinction is often made due to the different applications depending on the compartment the water is stored in. We would argue that these models are predestined to estimate groundwater recharge since they give a physically based estimate and can simulate feedbacks between the vadose and phreatic zones in a more sophisticated manner than multi-compartment models of subsurface flow. If groundwater recharge should be analyzed in an integrated model, this needs to be done by post-processing. As we will show, this is not straightforward.

In contrast to multi-compartment models, the flux from the vadose into the phreatic zone of an integrated model is almost exclusively indicative of lateral groundwater outflow. Neglecting storage changes due to compression, the flux crossing the water table cannot cause an increase in groundwater storage, as its destination (the phreatic zone) is already fully saturated. This statement might appear counterintuitive, as the water table often does fluctuate in such models. However, those fluctuations are caused when parts of the vadose zone reach full saturation, a process that takes place above rather than at the water table. Estimating groundwater recharge in integrated models requires assessing both lateral groundwater outflow and groundwater-storage changes. If surface flow is considered, rivers and local accumulation of surface water further complicate the process of estimating groundwater recharge.

Fully integrated models are at the edge of hydrogeological modeling and contain all information needed to make predictions about groundwater recharge on a high temporal and spatial resolution, yet the focus when results of these models are analyzed often lies on quanti-

ties closer to the land surface (e.g. surface runoff, infiltration or soil moisture in upper layers). The number of studies estimating groundwater recharge based on simulations of fully integrated models is scarce, and the respective methods for estimating groundwater recharge are usually not explained in detail. A consensus on a method for estimating groundwater recharge from these models is still missing. Frei et al. (2009) studied the temporal and spatial behavior of river-aquifer exchange fluxes in a domain with a water table between $2.5\,m - 12.5\,m$ below the surface using ParFlow. They estimated recharge by subtracting the change of water storage in the vadose zone from the flux of water infiltrating the soil without elaborating on how they quantified the changes in the vadose-zone storage. Guay et al. (2013) examined groundwater/surface-water interactions using both a weakly coupled model with a fixed water table and the fully integrated model CATHY. They referred to groundwater recharge from CATHY as "the sum of vertical fluxes that cross a dynamically changing water table" (Guay et al., 2013, p. 2268). They observed that recharge estimates from the fully integrated model were higher and more spatially variable than those obtained with the weakly coupled model. Maxwell et al. (2015) studied the surface and subsurface flow behavior over large parts of North America ($\sim 6.3 \cdot 10^6\,km^2$) at a spatial resolution of $1\,km$ using ParFlow. They estimated potential recharge by subtracting evapotranspiration from precipitation.

The high amount of information written out by fully integrated models allows an analysis of subsurface flow and states at a high level of complexity and detail but also poses a challenge to filter and process the information in order to properly analyze and visualize it. In this work, we present a novel approach for estimating groundwater recharge in fully integrated models. We claim that an analysis of the deeper subsurface layers of these models can be an important contribution to assess water quality and quantity in the future. We discuss the different contributions to the recharge estimate, as well as potential misconceptions and pitfalls. Our approach accounts for lateral groundwater outflow (and the steady component of groundwater recharge that balances it), and changes in groundwater storage based on concepts of water-table-fluctuation methods (e.g., Sophocleous, 1991; Healy & Cook, 2002; Crosbie et al., 2005). Even though the basic approaches have long been defined, this specific implementation for integrated subsurface-flow models is, to the best of our knowledge, unique.

This work is structured as follows. The upcoming section is dedicated to the underlying theory and governing equations describing variably saturated flow and groundwater recharge in integrated models. In section 3 we present two considerably different test cases. In section 4 we show our results for the two test cases, which consist of a detailed analysis of the dif-

ferent contributions to groundwater recharge in a fully integrated model, and in section 5 we draw conclusions and discuss the limitations of the proposed methodology.

2 Groundwater Recharge in Fully Integrated Models

Weakly coupled subsurface-flow models explicitly evaluate groundwater recharge as a coupling term between the vadose and phreatic zones, which are conceptualized as separate compartments (see Figure 1.a). In these models, the water leaving the vadose zone defines groundwater recharge. One-way coupled models would not allow capillary rise from groundwater to the vadose zone, as they only consider a flux from the vadose to the phreatic zone. Feedbacks between the compartments are not taken into account.

In contrast, integrated models consider the entire subsurface down to the bottom of the aquifer as a continuum (see Figure 1.b) and estimating groundwater recharge becomes a post-processing step. In these models, the water table rises as parts of the vadose zone become fully water-saturated. In contrast to weakly coupled models, there is no flux leaving the vadose zone that causes a rise in the water table. Instead, water percolates into the lower parts of the vadose zone, causing an increase in saturation. The flux crossing the water table ($F_{VZ\rightarrow GW}$ in Figure 1.b) is merely a measure of lateral groundwater outflow ($F_{GW,lat}$) and compression.

In the following, we show the equations for variably saturated flow and methods for estimating groundwater recharge in integrated models. Afterwards, we discuss the nuances of estimating groundwater recharge in fully integrated models, which additionally consider overland flow.

2.1 Governing Equations

Variably saturated flow problems are generally solved by the Richards' equation (Richards, 1931):

$$S_s \cdot S_w(\psi) \cdot \frac{\partial \psi}{\partial t} + \phi \cdot \frac{\partial S_w(\psi)}{\partial t} = \nabla \cdot [K(\psi)\nabla(\psi + z)] + Q_s \tag{1}$$

with the specific storage S_s $[L^{-1}]$, water saturation S_w $[-]$, pressure head ψ $[L]$, time t $[T]$, porosity ϕ $[-]$, hydraulic conductivity K $[LT^{-1}]$, geodetic elevation z $[L]$ and the source/sink term Q_s $[T^{-1}]$. In this study, the dependence of the water saturation S_w and hydraulic conductivity K on the pressure head ψ is described by the parameterization of van Genuchten (1980):

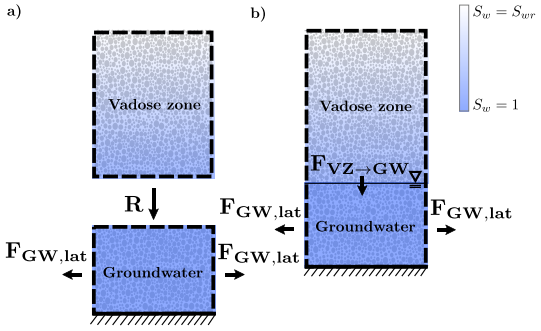

Figure 1. Differences in the conceptualization of the subsurface between weakly coupled and integrated approaches. S_w is water saturation and $F_{GW,lat}$ is lateral groundwater outflow. a) The flux leaving the control volume of a separate vadose-zone compartment is groundwater recharge R. b) In integrated subsurface-flow models, $F_{VZ\rightarrow GW}$ is the flux leaving the vadose zone by crossing the water table.

$$S_w(\psi) = \begin{cases} S_{wr} + \frac{(S_{ws}-S_{wr})}{[1+(\alpha\cdot|\psi|)^n]^m} & \text{if } \psi < 0 \\ S_{ws} & \text{otherwise} \end{cases} \tag{2}$$

$$K(\psi) = \begin{cases} K_s \cdot \frac{\left[1-\frac{(\alpha\cdot|\psi|)^{(n-1)}}{(1+(\alpha\cdot|\psi|)^n)^m)}\right]^2}{[1+(\alpha\cdot|\psi|)^n]^{(m/2)}} & \text{if } \psi < 0 \\ K_s & \text{otherwise} \end{cases} \tag{3}$$

with the residual water saturation S_{wr} $[-]$, the maximum water saturation S_{ws} $[-]$, the saturated hydraulic conductivity K_s $[LT^{-1}]$ and the van-Genuchten parameters α $[L^{-1}]$, n $[-]$, and $m = 1-(1/n)$.

Equation 1 is subject to initial conditions throughout the domain and boundary conditions at all boundaries of the integrated subsurface domain.

2.2 Contributions to Groundwater Recharge

Figure 2.a shows different flow components in the subsurface of an integrated model. $F_{I,E}$ is the exchange flux at the land surface, F_T is transpiration, $F_{VZ,lat}$ is lateral in- and outflow in the vadose zone, $F_{GW,lat}$ is lateral groundwater in- and outflow, and $F_{VZ\rightarrow GW}$ is the flux crossing the water table. $\dot{S}_{GW}$ is the change in groundwater storage between times t_0 and t_1. We neglect the compressibility of the fluid and pore structure in Figure 2 to keep it concise. The arrows in Figure 2.a show the classic groundwater recharge setting (water percolating downwards and then laterally leaving) and are defined to indicate the positive flow direction with

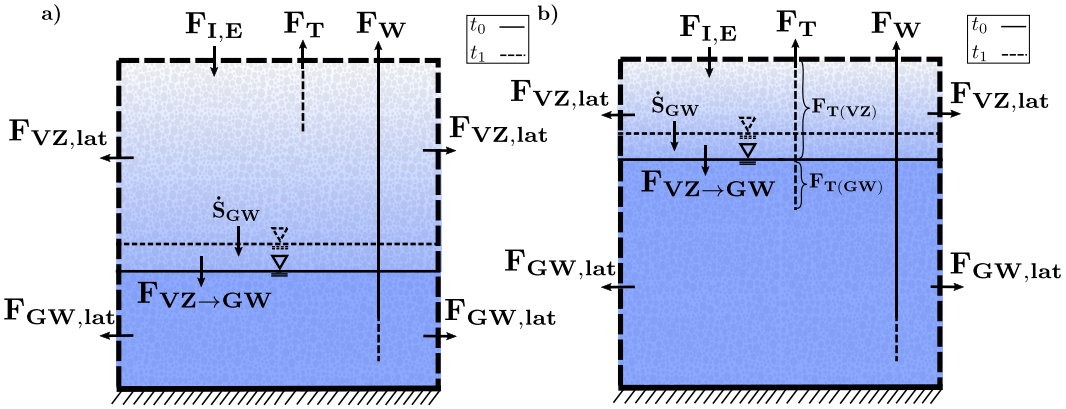

Figure 2. Fluxes in an integrated subsurface-flow model at two different times t_0 and t_1. Illustrated are: The exchange flux at the land surface $F_{I,E}$, transpiration F_T, lateral in- and outflow from the vadose $F_{VZ,lat}$ and phreatic $F_{GW,lat}$ zones, the flux crossing the water table $F_{VZ\rightarrow GW}$, temporal groundwater-storage changes $\dot{S}_{GW}$, and the flux caused by an extraction well F_W. Arrows indicate the positive flow direction. Compressibility is not shown. a) The water table lies below the root zone. Transpiration only drains the vadose zone. b) The water table lies within the root zone. Transpiration is divided into transpiration of groundwater $F_{T(GW)}$ and transpiration of soil water $F_{T(S)}$.

respect to the equations following in this section. All of these flow components could also flow in the opposite direction. Then they would contribute to the following equations with a negative sign. All contributions to groundwater recharge are considered as specific volumetric fluxes related to the horizontal cross section. While such a flux may not necessarily imply a direct physical meaning, this procedure allows for an easy water balance with the typical dimension of groundwater recharge $[LT^{-1}]$.

Without additional sources or sinks in the groundwater body, almost all water entering the phreatic domain by crossing the water table ($F_{VZ\rightarrow GW}$) must flow out as $F_{GW,lat}$ (the small difference results from groundwater compressibility). To clarify the different contributions one could consider the case of no groundwater leaving the aquifer ($F_{GW,lat} = 0$). Neglecting compressibility, the flux crossing the water table $F_{VZ\rightarrow GW}$ then has to be zero. Still, the water table could rise due to downward flow in the vadose zone, resulting in groundwater-storage changes $\dot{S}_{GW}$. Groundwater recharge needs to include groundwater-storage changes as well as outflow and can thus be defined as:

$$R = \dot{S}_{GW} + F_{GW,lat} + C_{GW} + F_W = \dot{S}_{GW} + F_{VZ\rightarrow GW} \tag{4}$$

with $F_{VZ \rightarrow GW}$ as the flux crossing the water table. C_{GW} is a flux equivalent to the groundwater compressibility (term 1 of Equation 1 multiplied by the vertical depth of the control volume) and F_W is the extraction by a well (introduced to generalize the methodology and not further examined within this work). Alternatively, recharge can be calculated as a residual of the water budget in the vadose zone:

$$R = F_{I,E} - F_T - F_{VZ,lat} - \dot{S}_{VZ} - C_{VZ} \tag{5}$$

in which $\dot{S}_{VZ}$ is the storage change in the unsaturated zone, C_{VZ} is the vadose zone compressibility and $F_{I,E}$, F_T and $F_{VZ,lat}$ are the water budget components shown in Figure 2.b. Depending on the position of the water table, regions of the vadose zone may be reassigned as belonging to groundwater, or vice versa, at each time. Estimating $\dot{S}_{VZ}$ is thus similar to estimating $\dot{S}_{GW}$, as it also requires accounting for storage changes due to water table fluctuations.

In cases with a shallow water table (see Figure 2.b), transpiration can act as a sink term to the vadose zone (VZ) and to the groundwater (GW). If plants extract groundwater, the flux crossing the water table must balance outflow, compressibility, the extraction by a well, and root extraction from groundwater: $F_{VZ \rightarrow GW} = F_{GW,lat} + C_{GW} + F_W + F_{T(GW)}$. This raises the question if water being transferred to groundwater, only to be immediately transpired, should be considered groundwater recharge or not. Amongst others, Doble and Crosbie (2017) addressed this issue by distinguishing gross groundwater recharge, R_{gross}, from net groundwater recharge, R_{net}:

$$R_{gross} = R_{net} + F_{ET(GW)} \tag{6}$$

in which $F_{ET(GW)}$ is the evapotranspiration from groundwater. Assuming that evaporation only occurs at the land surface (included in $F_{I,E}$ in Figure 2.b), equation 6 becomes:

$$R_{gross} = R_{net} + F_{T(GW)} \tag{7}$$

in which $F_{T(GW)}$ is transpiration from groundwater via root extraction.

In this shallow-water-table case (so if $F_{T(GW)}$ is greater than 0), the second equal sign in Equation 4 does not hold anymore and Equation 4 should be split up into the following two equations instead:

$$R_{gross} = \dot{S}_{GW} + F_{VZ \to GW} \tag{8}$$

$$R_{net} = \dot{S}_{GW} + F_{GW,lat} + C_{GW} + F_W \tag{9}$$

2.3 Estimating the Components of Groundwater Recharge

The integral assessment of groundwater recharge needs to include both the steady component that balances lateral groundwater outflow $F_{GW,lat}$ and the highly variable component $\dot{S}_{GW}$. $F_{GW,lat}$ represents the lateral movement of water inside the phreatic zone and can be expressed as:

$$F_{GW,lat} = \frac{1}{A_h} \int_{A_v|_{\psi>0}} \vec{q} \cdot \vec{n} \cdot dA \tag{10}$$

with $\vec{q} = K(\psi)\nabla(\psi + z)$ as filter velocity, $A_v|_{\psi>0}$ as all vertical interfaces of the control volume (a column in the domain where $\psi > 0$), $\vec{n}$ as normal vector and A_h as the horizontal cross section of the regarded area. Please note that the partitioning into multiple horizontal areas is necessary to acquire a spatial pattern of groundwater recharge. Within this work, A_h is a horizontal grid cell of the numerical model ($\Delta x \cdot \Delta y$), but it could also be chosen otherwise.

In a discretized model, calculating $F_{GW,lat}$ involves summing up multiple lateral fluxes (calculated according to the rules of the model) from bottom to top. It may be preferable to estimate lateral groundwater outflow from $F_{VZ \to GW}$ since that generally requires fewer calculations. $F_{VZ \to GW}$ is the flux crossing the water table from the vadose into the phreatic zone and it can be expressed as:

$$F_{VZ \to GW} = \frac{1}{A_h} \int_{A|_{\psi=0}} \vec{q} \cdot \vec{n} \cdot dA \tag{11}$$

with $A|_{\psi=0}$ as the water table surface. Discretized, this means calculating fluxes from unsaturated to saturated cells and usually, fewer fluxes need to be calculated this way compared to $F_{GW,lat}$. Note that both $F_{GW,lat}$ and $F_{VZ \to GW}$ are specific volumetric fluxes related to the same horizontal cross section A_h. For more details on our implementations to calculate these fluxes in our model, please see our published routines (Waldowski, 2022).

Estimating changes in groundwater storage is more ambiguous. It requires analyzing the fluctuations of the water table. The simplest way to do so is by applying the water-table-fluctuation

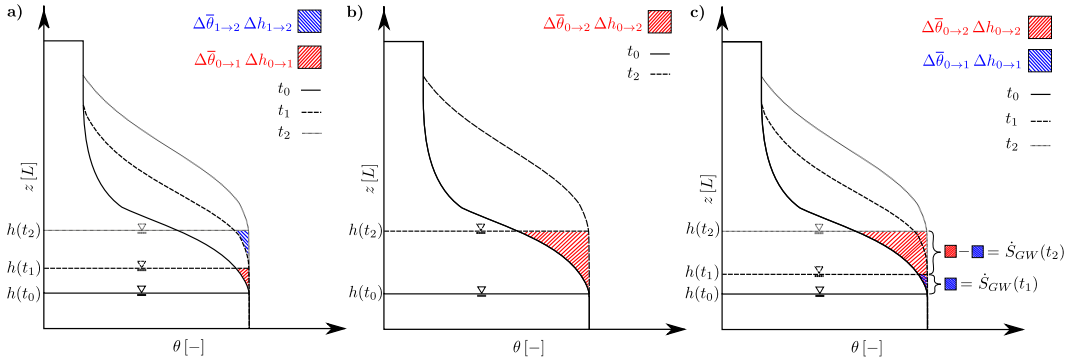

Figure 3. Exemplary relationship between the water content θ and the depth z. h is the water-table height. Hatched areas represent depth-integrated changes in the water content. In a) three different points in time are considered (t_0, t_1, t_2) and in b) two (t_0, t_2). c) Cumulative estimate of $\dot{S}_{GW}$.

method: $\dot{S}_{GW} = \frac{\Delta h}{\Delta t} \cdot S_y$ (e.g., Healy & Cook, 2002). S_y is the specific yield, a factor that accounts for the dampening behavior of the vadose zone when water infiltrates into or exfiltrates from groundwater. Estimating S_y is typically based on simplifying assumptions regarding the vadose zone. These might include the instantaneous drainage of the pore space or a soil-water profile that moves vertically with the water table but remains constant in shape (Childs, 1960). In the following, we restrain from such assumptions and consequently avoid using the term *specific yield* altogether.

The total rate of change of groundwater storage equals $\frac{dh \cdot \overline{\theta_s}}{dt}$, with $h\,[L]$ being the groundwater head, $t\,[T]$ time and $\overline{\theta_s}$ the saturated water content, averaged over the vertical depth which contained the water table in the considered time period. The term $\frac{dh \cdot \overline{\theta_s}}{dt}$ includes water that percolates downwards towards the water table, as well as soil water that is redefined as groundwater as the domains change with the moving water table. For a physically meaningful estimate of groundwater recharge, the latter should be excluded. The volume of water per unit area that is actually added to groundwater during a water table rise or withdrawn when the water table falls theoretically equals $dh \cdot d\overline{\theta}$. Here, $d\overline{\theta}\,[-]$ is the change in water content averaged over the depth that the water table rose/fell.

Considering discrete time steps Δt and discrete changes Δh and $\Delta\overline{\theta}$, we can show that $\Delta h \cdot \Delta\overline{\theta}$ is not a reliable estimate of $\dot{S}_{GW}$. Figure 3.a shows soil moisture profiles of an exemplary domain with a rising water table at three different points in time. The hatched areas between the retention curves illustrate $\Delta h \cdot \Delta\overline{\theta}$ for each discrete time step. In Figure 3.b, the same exemplary water table rise is shown, but the time step size is doubled. Even though the wa-

ter table in Figures 3.a and 3.b rises by the same total amount, it is apparent that $\Delta\overline{\theta}_{0\to1}\,\Delta h_{0\to1} + \Delta\overline{\theta}_{1\to2}\,\Delta h_{1\to2} \neq \Delta\overline{\theta}_{0\to2}\,\Delta h_{0\to2}$.

This mismatch stems from the underlying concept that each part of the domain below full saturation is assigned to the vadose zone at every point in time, resulting in $\lim_{\Delta t\to 0}\sum_{i=1}^{n}\Delta\overline{\theta}_{i-1\to i}\,\Delta h_{i-1\to i} = 0$. To overcome this issue, we introduced a cumulative term $S_{GW,cum}$ that considers the total gain or loss of groundwater, from a starting time t_0 to time t_n:

$$S_{GW,cum}(t_n) = \Big(\overline{\theta}(t_n) - \overline{\theta}(t_0)\Big) \cdot \Big(h(t_n) - h(t_0)\Big) \tag{12}$$

Please note that $\overline{\theta}(t_0)$ does change over time, as it is the initial water content $\theta(t_0)$ averaged over the range that the water table rose/fell: $h(t_n) - h(t_0)$. We can determine the change in groundwater storage $\dot{S}_{GW}$ for a particular time step Δt from $S_{GW,cum}$ as:

$$\dot{S}_{GW}(t_n) = \frac{S_{GW,cum}(t_n) - S_{GW,cum}(t_n - \Delta t)}{\Delta t} \tag{13}$$

Changes in groundwater storage $\dot{S}_{GW}$ depend on the moisture profile above the water table. Determining $\dot{S}_{GW}$ based on this cumulative estimate ensures that $\sum_{t=t_0}^{t=t_n}\dot{S}_{GW}(t)$ does not depend on Δt (see Figure 3.c). Still, $\dot{S}_{GW}$ needs to refer to a unique fixed state at the initial time to exclude groundwater gained due to the initial pre-saturation of the vadose zone.

2.4 Estimating Groundwater Recharge in Fully Integrated Models

Fully integrated models introduce additional challenges for recharge estimation as they simultaneously solve water-balance equations for the subsurface and the land surface. This implies the integration of the vadose and phreatic zones as well as the coupling of the land surface and the subsurface.

For a description of different fully integrated models and their respective approaches of coupling the subsurface and the land surface, the interested reader is referred to Maxwell et al. (2014). The approaches described there include first-order exchange (e.g. HydroGeoSphere), boundary condition switching (CATHY), or assuming continuity of pressure (e.g. ParFlow). The surface-flow equations vary between different models. Amongst others are the kinematic-wave approximation of the Saint Venant equation used by ParFlow (Kollet & Maxwell, 2006) and the diffusive-wave approximation implemented in HydroGeoSphere (Therrien et al., 2010).

In such models, infiltration and groundwater recharge may be caused by precipitation, rivers, and intermittent water bodies (local ponding). Amounts of recharge resulting from these different sources may differ in orders of magnitude.

In this study, we consider recharge caused by intermittent water bodies as a contribution to groundwater recharge, while neglecting recharge caused by rivers (R_{river}). The exclusion of river-groundwater exchange has been discussed by Lerner et al. (1990) who defined *localized recharge* as recharge "in the absence of well-defined channels" (Lerner et al., 1990, p. 6). The rather vague term "well-defined" suggests that an objective and clear distinction between recharge caused by rivers and by other components is hardly ever possible.

Generally, a distinction between river-groundwater exchange and groundwater recharge can be based on the location where the recharge occurs. Estimating these *river-groundwater exchange locations* requires critical examination of the case study at hand since locations of rivers might change over time and rivers might influence neighboring parts of the domain. We suggest that the river location, which is used to distinguish river-groundwater exchange from groundwater recharge, should include a buffer zone extending beyond the actual rivers. It should be noted that, when analyzing different recharge sources, further distinctions into *net* and *gross recharge* can still be made.

To summarize: In the sections 2.2 to 2.4 we have shown the equations and procedures of how we calculate groundwater recharge in a fully integrated model. We now exemplarily present the key steps for estimating R_{gross}. For the detailed procedure, please see our published routines (Waldowski, 2022).

Divide the horizontal surface into distinct areas (spatial resolution of recharge)

Do the following at each time and each distinct horizontal area

If the grid cell is at a river

- $R_{gross} = 0$

Else

- Read vertical ψ-profile
- Determine depth to watertable (depth at which $\psi = 0$) via linear interpolation
- Calculate $F_{VZ \rightarrow GW}$ based on vertical ψ-gradients and Equation 11
- Calculate $\dot{S}_{GW}$ according to Equations 12 and 13
- Calculate R_{gross} according to Equation 8

3 Illustrative Test Examples

We studied the applicability of our method of estimating groundwater recharge using two substantially different test cases. Test case 1 is a vertical soil column with no lateral groundwater outflow, constraining the effects of groundwater recharge to changes in groundwater storage. We use this simple test case to demonstrate that the flux $F_{VZ\rightarrow GW}$ that crosses the water table in an integrated model cannot capture all facets of groundwater recharge. Furthermore, we compare our recharge estimates with those from the water-table-fluctuation method. Test case 2 is a more realistic scenario, in which we investigate spatial patterns of groundwater recharge, as well as influences of lateral subsurface flow, vegetation, and rivers.

3.1 Test Case 1: Vertical Soil Column

Test case 1 consists of a vertical soil column of $2\,m$ height with no-flow boundaries at the sides and the bottom (see Figure 4.a). In the initial state, the domain is in hydrostatic equilibrium with a water table depth of $1\,m$. At the top, a daily varying flux is applied for a total simulation time of 100 days. Surface water might accumulate, but surface runoff does not occur. The porous medium is homogeneous and sandy, with the following parameters: $\phi = 0.43$, $K_s = 3 \cdot 10^{-6}\,\frac{m}{s}$, $\alpha = 3.6\,\frac{1}{m}$, $n = 1.6$, $S_{wr} = 0.1$, $S_{ws} = 1.0$. The problem is simulated with ParFlow, version 3.6.0 (Ashby & Falgout, 1996; Jones & Woodward, 2001; Kollet & Maxwell, 2006; Maxwell, 2013).

3.2 Test Case 2: Groundwater Recharge in a Complex Three-Dimensional Domain

To test the method in a more realistic setting, we used a model based on the coupled land surface-subsurface model presented by Erdal et al. (2019). This model was constructed with the Terrestrial Systems Modeling Platform TSMP (Gasper et al., 2014; Shrestha et al., 2014), which couples the community land-surface model (CLM) (Oleson et al., 2004) to ParFlow. The coupling of ParFlow and CLM is described by Maxwell and Miller (2005) and Kollet and Maxwell (2008).

The model domain covers a rectangular area of $1\,km \times 5\,km$, and has a uniform depth of $50\,m$. The grid has a horizontal resolution of $40\,m$ and layers of variable thickness. The area is characterized by a floodplain with an adjacent hillslope, enclosed by three rivers running along the northern, southern and western boundaries (enforced by topographic gradients). All lateral boundaries in the subsurface are no-flow boundaries so that water can leave the domain

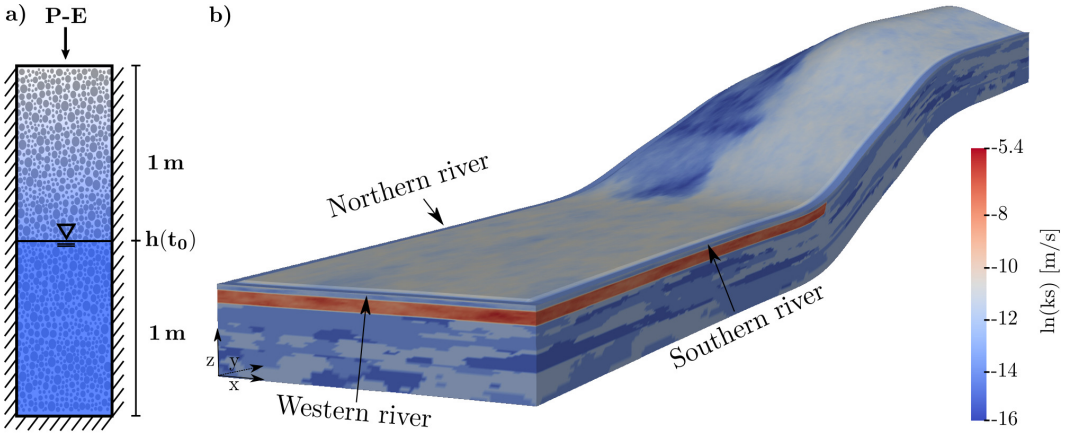

Figure 4. a) Domain setup for test case 1: $2\,m$ deep soil column with impermeable boundaries at the sides. b) Domain setup for test case 2 (see also Erdal et al., 2019): Spatial distribution of the hydraulic conductivity K_s in the model domain (vertical exaggeration: 5).

only via evapotranspiration or a river outlet defined at the upper southwestern corner of the model.

The subsurface is divided into soil, gravel (only in the floodplain), bedrock, and riverbed units. The soil units are further subdivided into three different soil types. Each soil unit is defined by two layers and covers a specific area of the model domain with a single plant functional type: corn in the floodplain, grass on the southern hillslope, needleleaf evergreen temperate trees on the northern hillslope. The soil texture of all upper layers is heterogeneous, with varying percentages of sand, clay, and silt. To avoid sharp transitions between soil layers, we generated spatially correlated soil texture fields using the conditional-points method of Baroni et al. (2017). Soil parameters were derived using a set of pedotransfer functions defined by Cosby et al. (1984) and Tóth et al. (2015).

All atmospheric forcings used for this model are heterogeneous distributions that were generated with a bigger and more complex model presented by Schalge et al. (2021).They are typical for the climate in Central Europe. We spun up our model by repeatedly applying the same precipitation data until the absolute difference in depth to the water table between consecutive years was below $0.01\,m$. Afterwards we applied a different set of precipitation fields to obtain a dynamic response of the system.

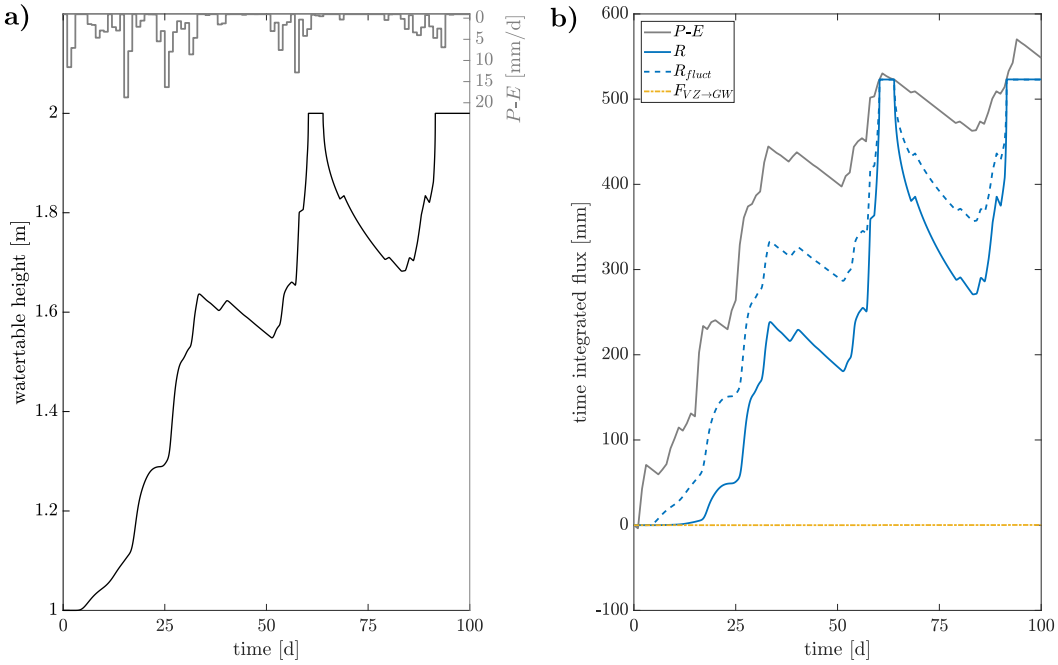

Figure 5. a) Bottom to top: water-table height as a function of time in days. Top to bottom: Net forcings (precipitation minus evaporation). Evaporation is constant ($1.37\,mm/d$) and precipitation is variable. b) Net forcings *P-E*, groundwater recharge from our proposed method *R* (see section 2), groundwater recharge based on the water-table-fluctuation method R_{fluct} and the flux crossing the water table $F_{VZ\rightarrow GW}$ integrated over time.

4 Results and Discussion

4.1 Test Case 1: Vertical Soil Column

In this test case, we applied a daily varying flux (see Figure 5.a, top to bottom) to the top of a laterally impermeable vertical soil column. Evaporation (E) was applied only at the land surface and set to a constant value of $1.37\,mm/d$, introducing periods of net evaporation at days without rain. We decided for this simplistic design to precisely address and show which implications an integrated subsurface has when estimating groundwater recharge.

The water table rises from the initial position of $1\,m$ and reaches the land surface at $2\,m$ twice: $\sim$60 days and $\sim$90 days after the start of simulations. When the entire column is water-saturated, additional water may accumulate as surface water, but cannot run off. Figure 5.b shows fluxes integrated over time, with *P-E* $[LT^{-1}]$ denoting net forcing, $R\,[LT^{-1}]$ being groundwater recharge according to equations 4, 11 and 13, $R_{fluct}\,[LT^{-1}]$ denoting groundwater recharge calculated with the water-table-fluctuation method ($R_{fluct} = \frac{\Delta h}{\Delta t} \cdot S_y$), and $F_{VZ\rightarrow GW}\,[LT^{-1}]$ be-

ing the flux crossing the water table. S_y is the specific yield, which we assume constant in space and time, evaluated by $S_y = \phi - \int_{\psi=-1\,m}^{\psi=0\,m} \theta(\psi)\, d\psi = 0.13$ considering the initial state. There is no transpiration from groundwater so that $R_{gross} = R_{net} = R$.

In its original formulation, the water-table-fluctuation method neglects $F_{GW,lat}$ and therefore, R_{fluct} is a reliable estimation only when lateral flow is negligible. We use R_{fluct} as a reference for comparing the results of our approach with test case 1, where no lateral flow occurs.

As lateral groundwater outflow $F_{GW,lat}$ is inhibited, the flux crossing the water table $F_{VZ\rightarrow GW}$ is almost zero (equals aquifer compressibility) throughout the simulation (see Figure 5.b). Differences between *P-E* and R or R_{fluct} can be attributed to the difference between water that enters the subsurface, affecting the soil moisture above the water table, and water that actually recharges into groundwater. Groundwater recharge (in blue) increases only after the water table rises (~ 4 days). Close to full saturation (~ 50 days), the vadose zone disappears and the time-integrated fluxes *P-E*, R and R_{fluct} intersect, which means that the water balance closes. Water added afterwards remains as surface water in the system until evaporation removes all ponded water and the water table declines. At this time, capillary rise becomes larger than evaporation, meaning that water previously labeled as groundwater is now counted as water of the vadose zone.

These results show that R and R_{fluct} close the groundwater balance by defining an initial state and tracking the water table position (see Figure 5.b) and only differ in instantaneous values for this test case. The mimicking of the water-table height by the time-integrated groundwater recharge estimates is a necessity for this simplified test case, because the domain has closed boundaries and thus $F_{GW,lat} = 0$. In more complex flow scenarios, this simple relationship between water-table height and groundwater recharge is not to be expected. While R considers the initial soil moisture profile (and its shape) (see Figure 3.c), R_{fluct} uses a specific yield representative of the average initial soil moisture profile. By integrating over the whole vadose zone, S_y is overestimated initially since in reality, the water content near the water table is higher than further away and hence, the storage capacity is reduced. This means that R_{fluct} overestimates groundwater recharge at the start of the simulation and must underestimate it when approaching full saturation to meet the water balance.

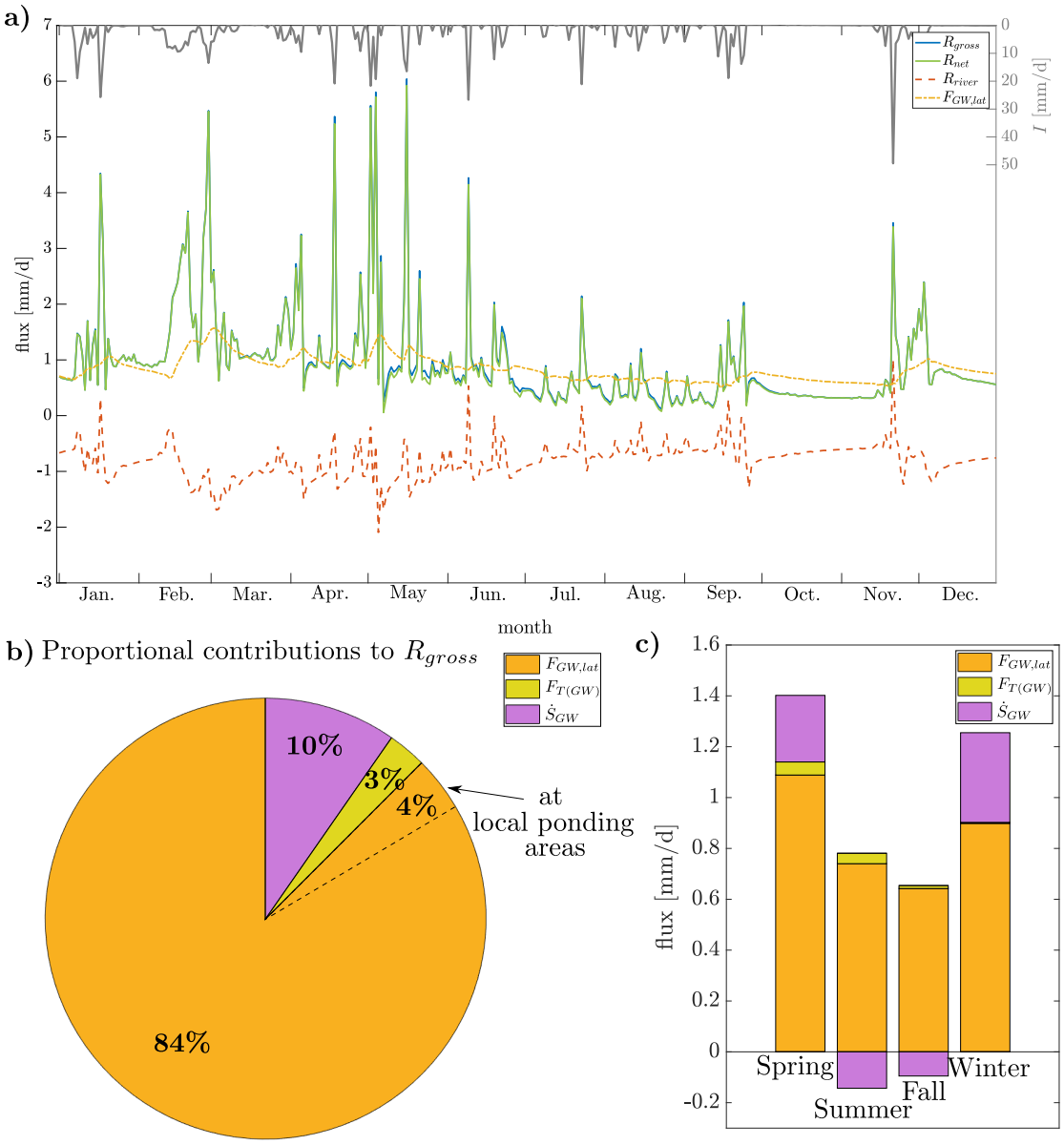

Figure 6. a) Daily mean fluxes averaged over the whole domain. Illustrated are: Gross recharge R_{gross}, net recharge R_{net}, river-groundwater exchange R_{river} and lateral groundwater outflow $F_{GW,lat}$ which is mathematically equivalent to a component of groundwater recharge (see equation 4). On the second y-axis, the net inflow I (CLM-variable: 'QINFL') is plotted from top to bottom. b) Percentage of recharge components concerning the total annual average. Shown are the amount of recharge that is needed to balance lateral groundwater outflow $F_{GW,lat}$, groundwater-storage changes $\dot{S}_{GW}$ and the recharge needed to balance groundwater transpiration $F_{T(GW)}$. c) Distribution of $F_{GW,lat}$, $F_{T(GW)}$ and $\dot{S}_{GW}$ for spring, summer, fall and winter (meteorological seasons).

4.2 Test Case 2: Groundwater Recharge in a Complex Three-Dimensional Domain

The complexity of test case 2 is generated by variable topography, vegetation, and subsurface properties, as well as surface flow. These more realistic settings generate a dynamic system in which groundwater recharge is affected by lateral groundwater outflow $F_{GW,lat}$, river-groundwater exchange, local ponding, groundwater transpiration $F_{T(GW)}$, and complex subsurface flow patterns. We do not address recharge due to compression in the following analysis, since it only represents $\sim 0.01\%$ of the total annual groundwater recharge in our test case.

Figure 6.a shows daily averages of gross recharge R_{gross} (Equation 8), net recharge R_{net} (Equation 9), river-groundwater exchange R_{river} (which is not part of groundwater recharge), lateral groundwater outflow $F_{GW,lat}$ (Equation 10) and net inflow I (calculated by CLM). Groundwater recharge estimated at the rivers plus a buffer zone of $40\,m$ (one grid cell) is regarded as R_{river}. In our test case, R_{river} is mainly negative, indicating flow from groundwater into the rivers. Note that the results in Figure 6.a are averaged over the whole domain, and rivers (including their buffer zone) cover less than $20\,\%$ of the area, so that local values of river-groundwater exchange are much higher in absolute terms. In our test case, groundwater discharge to the rivers is enhanced by the boundary conditions, which do not allow groundwater to flow laterally out of the domain. The temporal patterns are discussed in detail to analyze the relevance of different contributions to recharge at different timescales.

While showing some distinct patterns, average estimates of groundwater recharge (R_{gross} and R_{net}) generally reflect the imposed forcings (see Figure 6.a). Averages of groundwater recharge are above zero throughout the simulated year, even during October and November when no additional water is added to the system by the forcings.

Differences between R_{gross} and R_{net} (blue and green lines in Figure 6.a) are small throughout the simulated year. In the vegetation period from April to September, R_{gross} is systematically higher than R_{net}. In general, changes in groundwater storage are more sensitive to precipitation than lateral groundwater fluxes, which can be seen by the more dynamic response of R_{gross} and R_{net} compared to $F_{GW,lat}$.

Figure 6.b shows the contributions of different components to gross recharge in the annual average. We estimated a total of $351.5\,mm$ of gross recharge for the simulated year, from which 2.8% balances $F_{T(GW)}$. As expected, in areas with shallower water tables the amount

of groundwater recharge needed to balance $F_{T(GW)}$ increases (e.g., $\sim 5.0\,\%$ in the floodplain + at the transition to the hillslope).

Figure 6.a further shows that the spatial average of $F_{GW,lat}$ does not considerably vary throughout the year. R_{gross} and R_{net} show more temporal fluctuation, as $\dot{S}_{GW}$ (difference between R_{net} and $F_{GW,lat}$) is highly variable and more sensitive to precipitation than $F_{GW,lat}$. Responses to precipitation peaks can often be seen almost immediately in R_{gross} and R_{net}, whereas $F_{GW,lat}$ tends to be smoother. Lateral groundwater outflow $F_{GW,lat}$ demands $87.5\,\%$ of R_{gross}, with $4.3\,\%$ occurring in areas with local ponding. This implies that most of the groundwater recharge moves on laterally. We also estimated that $\dot{S}_{GW}$ constitutes 10% of the annual groundwater recharge, which means that the water table rises in the simulated year. A net rise of the water table was caused by the heavier rainfall imposed during the simulated year than during spinup. The contribution of $\dot{S}_{GW}$ would have been negative if the simulated year had been dryer (lower $P-E$) than the spinup year.

Figure 6.c shows the composition of gross recharge over the different seasons. We see the most distinct seasonal behavior in $\dot{S}_{GW}$, with spring and winter months causing an increase in groundwater storage and summer and fall months decreasing it. $F_{T(GW)}$ correlates with the vegetation season and water table depth, causing a higher demand for gross recharge in spring and summer. $F_{GW,lat}$ also shows a seasonal trend, with lower values in summer and fall and consequently a lower demand for gross recharge.

The recharge estimates at four distinct locations are discussed in detail to address the question if recharge in the whole domain is well represented locally and where a location at which groundwater recharge is indicative for the whole domain could be.

Figure 7.a shows the spatial distribution of the temporal average of gross recharge, and the location of four different points selected to show the local temporal evolution (Figure 7.b). In general, we observe low spatial variability of gross recharge in the floodplain and at the hills, with values close to $1.5\,mm/d$ and $1.0\,mm/d$, respectively. In contrast, spatial variability is high in the regions close to the outlet and at the transition between the hillslope and the floodplain. The extreme recharge values in this transition zone (dark red and blue in Figure 7.a) are caused mainly by the topographic variability.

Figure 7.b compares the temporal evolution of gross recharge at four specific locations with the spatial mean. We observe the highest water table depths at the top of the hills (see

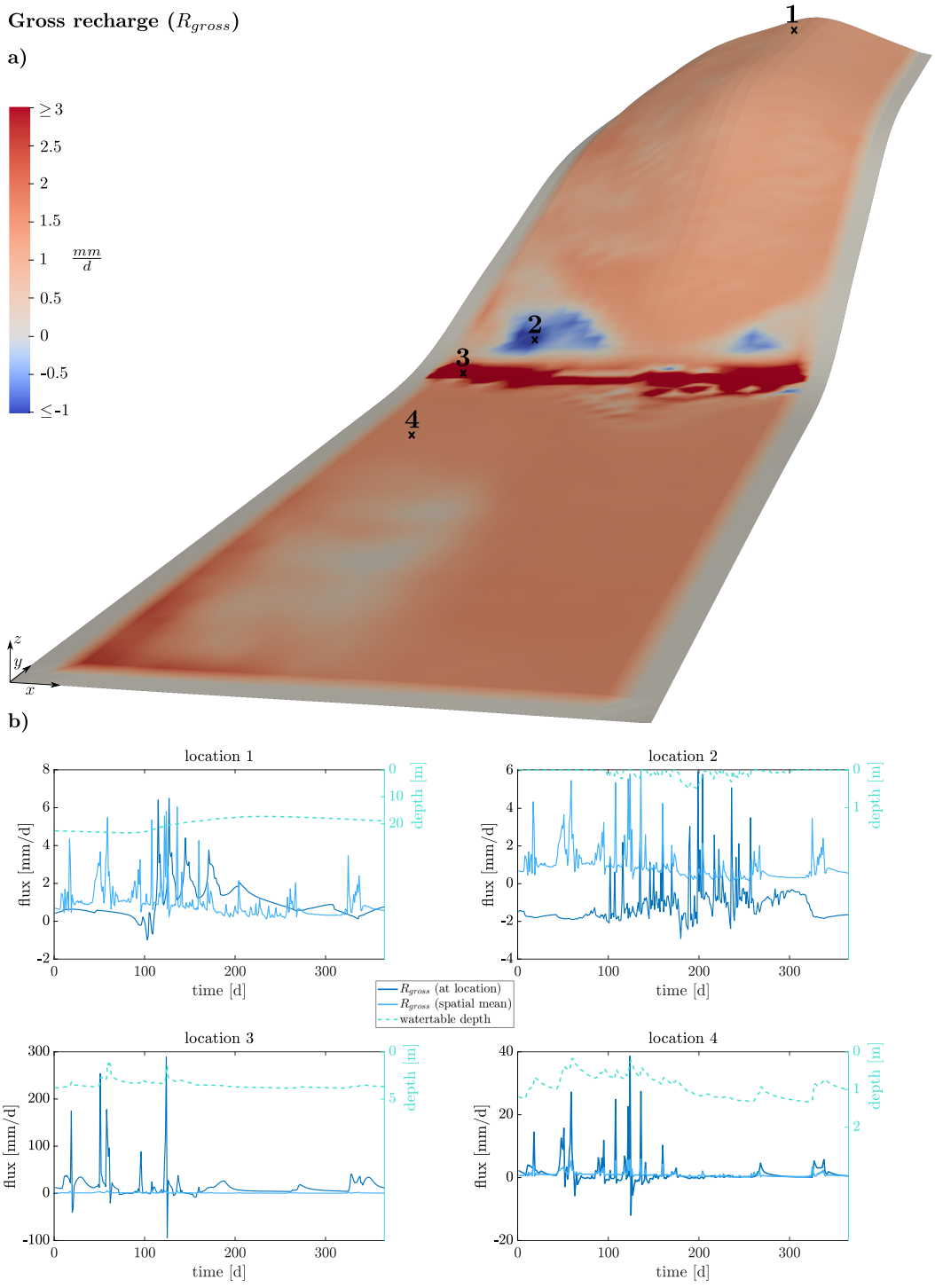

Figure 7. a) Spatial distribution of gross groundwater recharge R_{gross} (temporal mean). The maximum and minimum of the color bar are adjusted for better visibility. b) Comparison of the temporal evolution of gross recharge R_{gross} at four different locations (dark blue) and the spatial mean (light blue). The local water table depths are shown in cyan.

location 1 in Figure 7.a and b). Because the water table lies deep within the bedrock, travel times are high leading to a less dynamic response to the atmospheric forcings. The water table at this location rises by $3.5\,m$ during the simulated year. However, this is not translated to high recharge rates due to the low porosity. While the temporal average of R_{gross} at this location is similar to the spatial average over the whole domain, the system's response is less dynamic than the average behavior with practically no correlation between the two signals ($r = -0.06$). It is worth noting that the shift of R_{gross}-values from below to above the spatial mean occurs exactly when the water table rises (after ~ 100 days). This highlights the impact of the ground depth on groundwater recharge estimates and how the temporal behavior is influenced by changes of the water table.

Location 2 is located on the hillslope right before transitioning to the floodplain. Here, the negative values observed (in blue) are related to high lateral groundwater outflow $F_{GW,lat}$ upstream caused by rapid changes in topographic elevations and high water tables. Due to the high downward slopes, a notable amount of groundwater crosses the water table into the vadose zone and/or land surface, resulting in capillary rise. Most of this water recharges again into the groundwater at lower elevations (see below). Between days 100 and 250, the local response shows similar patterns to those of the spatially averaged R_{gross}. In the first 90 days, the water table depth at this location does not change and the value of R_{gross} shows a less variable temporal behavior than the spatial mean. At around 325 days, there is a rise in the spatial mean and a fall in the local value of R_{gross}.

At location 3, which is located at the foot of the hillslope, we observe the highest temporal mean value of gross recharge (R_{gross} = $14.4\,mm/d$). The high values of groundwater recharge observed here are driven by the abrupt changes in the topographic gradient and subsurface material. In general, water flows down the hillslope and recharges into the gravel body buried in the floodplain. While at this location lateral flow takes place in both the vadose zone as well as on the land surface, we found that the extreme peaks of groundwater recharge at this location are mainly caused by upstream surface flow and strong groundwater-surface water interactions. Note that this behavior reflects conditions in natural landscapes, whereas the (re-)infiltration of the surface flow may be inhibited in managed landscapes by drainage channels along the foot of the hillslope, which we have not considered in our test case. A total of 85% of gross recharge R_{gross} at this location is demanded by lateral groundwater outflow $F_{GW,lat}$. This water enters the phreatic domain at this location and flows either downstream and/or to the northern river. Due to the large magnitude of lateral flow, water tables change rapidly and

sometimes reach the land surface. Location 3 does not only demonstrate extremely high ($288.7\,mm/d$), but also extremely low ($-94.3\,mm/d$) values of gross recharge, often related to rapid declines of the water table. This stresses the complexity of groundwater-recharge dynamics, depending on the topographic conditions.

Finally, location 4 is located within the floodplain and further away from the hillslope. Gross recharge in this region has the highest correlation to the spatial mean ($r = 0.61$). While both cover a similar range of values, we can see that peaks in both positive and negative direction are more pronounced at this location compared to the spatial mean.

These results show, along with the spatial distribution in 7.a, that there is not a single location in our domain which is representative of the spatial mean and stress the need to consider spatial variability when estimating groundwater recharge for a larger domain.

5 Conclusions

In the course of this work, we have discussed the challenges and nuances of estimating groundwater recharge from the results of fully integrated surface-/subsurface flow models, showing two distinct test cases. With our first test case (1-D vertical soil column in ParFlow), we demonstrated that for integrated models, the flux crossing the water table $F_{VZ\rightarrow GW}$ is only a partial estimate of groundwater recharge as it misses changes in groundwater storage ($\dot{S}_{GW}$). We estimate $\dot{S}_{GW}$ cumulatively to avoid a numerical time step size dependency (visualized in Figure 3). Due to the fully integrated modeling approach, we do not have to rely on simplifying assumptions such as a constant specific yield. In this simplified testcase, results from our approach (R) were basically similar to predictions by the water-table-fluctuation method (R_{fluct}), but they showed a more realistic temporal behavior that better aligns with the dynamic nature of the specific yield.

Our second test case was a more realistic scenario (3-D heterogeneous domain, with rivers at the sides, a hillslope and a floodplain). We estimated a total annual and spatial average groundwater recharge of approximately $1\,mm/d$, which is high but not uncommon in the related region (e.g., Neumann & Wycisk, 2003). Locally and temporally finer resolved recharge rates deviated from this average. In time series at specific locations of the domain, we have observed peaks of groundwater recharge that lie in ranges of tens, sometimes even hundreds of mm/d. Similarly high values were observed by others that estimated local time series of groundwater recharge (e.g., O'Reilly, 2004; Crosbie et al., 2005).

We found lateral groundwater outflow $F_{GW,lat}$ to be the key flux concerning the annual average of groundwater recharge, whereas changes in groundwater storage $\dot{S}_{GW}$ played the major role in terms of temporal variability.

By analyzing its spatial distribution, we have shown that groundwater recharge is highly heterogeneous, and that hardly any point of the domain can be considered representative of the entire domain. This emphasizes the benefit of numerical models when estimating groundwater recharge if necessary data is available since they allow for a spatial coverage that could not be achieved by measurements. We also found surface flow to be the major cause of local extremes of groundwater recharge and promoted the idea to mask out groundwater recharge at river locations (R_{river}) and consider it separately.

The flux crossing the water table $F_{VZ\rightarrow GW}$ can be seen as a global contribution to groundwater recharge. That is because it takes water into account that recharges into the aquifer and then moves laterally ($F_{GW,lat}$). We also demonstrated that it does not directly indicate groundwater-storage changes at a single location, as shown in test case 1. The component $F_{VZ\rightarrow GW}$ may be taken as a measure of groundwater recharge only in long-term studies when it is expected that water table fluctuations average out over time, which means that groundwater-storage changes $\dot{S}_{GW}$ can be neglected.

The high temporal variability of groundwater recharge, which in our case has shown to be mainly influenced by $\dot{S}_{GW}$, is relevant on both the seasonal scale and on an event basis. In applications of water availability, accurately capturing seasonal fluctuations of groundwater recharge can be relevant. In applications regarding contaminant and pesticide transport, we see a particular need to accurately capture extreme recharge events.

While we expect the general tendencies that we have observed ($\dot{S}_{GW}$ is more indicative of instantaneous values and $F_{VZ\rightarrow GW}$ depicts long-term averages) to be transferable to other sites, we recognize that the share of $\dot{S}_{GW}$ on the groundwater recharge estimate is highly specific to the test case at hand. It should also be kept in mind that predicting only $\dot{S}_{GW}$ (and thus neglecting $F_{VZ\rightarrow GW}$) is insufficient in most applications, especially if spatial distributions of groundwater recharge and longer time periods are of interest. That is because a potentially high proportion of groundwater recharge does not cause water table fluctuations but moves on laterally instead, as we have shown in test case 2. We would claim that having information about this proportion of recharge matters for planning e.g. groundwater extraction. If extraction wells were placed in areas where $F_{VZ\rightarrow GW}$ is high, a high yield of the well would be expected and

621 one could extract groundwater before it laterally moves on (potentially into a river or the sea,
622 compromising its purity).

623 A distinction between groundwater recharge R and river-groundwater exchange R_{river} is
624 a challenge unique to fully integrated models which we have tackled by differentiating between
625 river and non-river locations. A binary distinction between river and non-river locations is in-
626 evitably accompanied by assumptions since rivers may spontaneously occur, dry out and me-
627 ander and their locations might be quite ambiguous. In fully integrated models with 2-D rep-
628 resentations of rivers, the latter are not defined in advance but occur based on pressure con-
629 ditions. This implies that in our case, the definition of river locations was a post-processing
630 task. We cannot give universal instructions on how to define such river locations since that de-
631 pends on the test case at hand as well as on the specific requirements on the recharge estimate.
632 Our choice was made on subjectively chosen measures. Finding general rules how to distin-
633 guish river-groundwater exchange from groundwater recharge is beyond the scope of this work.

634 A limitation of our estimate of $\dot{S}_{GW}$ is that it depends on a fixed initial time as a ref-
635 erence for the later states. We would claim that this is inevitable if one aims to distinguish stor-
636 age changes of the groundwater and the vadose zone within an integrated subsurface in a phys-
637 ically meaningful way.

638 While the distinctions we make between the phreatic and the vadose zones as well as
639 the distinctions between local ponding and rivers are common and practical with respect to
640 applications, it should be kept in mind that they are not included in the fundamental concept
641 of a fully integrated model.

642 This work aims to make predictions of fully integrated models about groundwater flow
643 more accessible and easier to portray. We have shown a post-processing method that has de-
644 livered robust predictions in two fundamentally different and uniquely challenging test cases.
645 We see our essential claim, that one needs to account for both the flux crossing the water ta-
646 ble ($F_{VZ\rightarrow GW}$) as well as groundwater-storage changes ($\dot{S}_{GW}$) for an all-embracing estimate of
647 groundwater recharge in a continuum to be validated throughout this work. Test case 1 em-
648 phasized the importance of $\dot{S}_{GW}$ whereas test case 2 showed the relevance of $F_{VZ\rightarrow GW}$. Ways
649 how we determined these contributions (e.g. the cumulative estimate of groundwater-storage
650 changes) should only be seen as suggestions that remain open for discussion.

Groundwater recharge is essential for the planning of water resources and fully integrated models are powerful tools that enable a spatial and temporal analysis while respecting the interconnectedness of the hydrological compartments. Our results suggest that their ability to simulate lateral flow in the entire subsurface and on the land surface is key to identify extreme values and distinct spatial and temporal patterns of groundwater recharge. To plan withdrawals, assess aquifer vulnerability, or acquire an upper boundary condition for a groundwater model, such a detailed analysis of groundwater recharge is desirable. There are different points of view on groundwater recharge depending on the underlying scientific discipline and fully integrated models are capable of building bridges between these different disciplines. This work should act as an overview for fully integrated modelers and inspire more research on groundwater recharge in the future. With respect to the remaining high interest and recent development in large-scale simulations of fully integrated models (e.g., Yang et al., 2020; Naz et al., 2020; Belleflamme et al., 2021; Oloruntoba et al., 2021; O'Neill et al., 2021), we see high potential for our method to extract further information from these computationally expensive model runs.

Data Availability

Simulation results (raw outputs and postprocessed data) and post-processing routines are archived in the Zenodo repository (`https://dx.doi.org/10.5281/zenodo.6244618`). Our ParFlow post-processing routines are also publicly accessible on GitHub under this link: https://github.com/Waldowski/Gr

Acknowledgments

This research was performed in the Research Unit FOR2131 "Data Assimilation for Improved Characterisation of Fluxes across Compartmental Interfaces", funded by Deutsche Forschungsgemeinschaft (grants: NE 824/12-2, Ci 26/13-2, and HE 6239/4-2).

References

Ashby, S. F., & Falgout, R. D. (1996). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear science and engineering*, *124*(1), 145–159.

Baroni, G., Zink, M., Kumar, R., Samaniego, L., & Attinger, S. (2017). Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. *Hydrology and Earth System Sciences*, *21*(5), 2301–2320.

Barthel, R. (2006). Common problematic aspects of coupling hydrological models with

groundwater flow models on the river catchment scale. *Advances in Geosciences*, *9*, 63–71.

Belleflamme, A., Görgen, K., Kollet, S., & Wagner, N. (2021). Forecasting soil moisture at 600m resolution over germany usingthe hydrologic model parflow/clm with ecmwf atmospheric forcing. In *Joint virtual workshop on "connecting global to local hydrological modelling and forecasting: scientific advances and challenges"*.

Brunner, P., Bauer, P., Eugster, M., & Kinzelbach, W. (2004). Using remote sensing to regionalize local precipitation recharge rates obtained from the chloride method. *Journal of Hydrology*, *294*(4), 241–250.

Brunner, P., & Simmons, C. T. (2012). Hydrogeosphere: a fully integrated, physically based hydrological model. *Groundwater*, *50*(2), 170–176.

Camporese, M., Paniconi, C., Putti, M., & Orlandini, S. (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, *46*(2).

Childs, E. (1960). The nonsteady state of the water table in drained land. *Journal of Geophysical Research*, *65*(2), 780–782.

Cosby, B., Hornberger, G., Clapp, R., & Ginn, T. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water resources research*, *20*(6), 682–690.

Crosbie, R. S., Binning, P., & Kalma, J. D. (2005). A time series approach to inferring groundwater recharge using the water table fluctuation method. *Water Resources Research*, *41*(1).

De Vries, J. J., & Simmers, I. (2002). Groundwater recharge: an overview of processes and challenges. *Hydrogeology Journal*, *10*(1), 5–17.

Doble, R. C., & Crosbie, R. S. (2017). Current and emerging methods for catchment-scale modelling of recharge and evapotranspiration from shallow groundwater. *Hydrogeology Journal*, *25*(1), 3–23.

Erdal, D., Baroni, G., Sánchez-León, E., & Cirpka, O. A. (2019). The value of simplified models for spin up of complex models with an application to subsurface hydrology. *Computers & geosciences*, *126*, 62–72.

Fayer, M., Gee, G., Rockhold, M., Freshley, M., & Walters, T. (1996). Estimating recharge rates for a groundwater model using a gis. *Journal of Environmental Quality*, *25*(3), 510–518.

Freeze, R. A. (1969). The mechanism of natural ground-water recharge and discharge: 1. one-dimensional, vertical, unsteady, unsaturated flow above a recharging or discharging ground-water flow system. *Water Resources Research*, *5*(1), 153–171.

Frei, S., Fleckenstein, J., Kollet, S., & Maxwell, R. M. (2009). Patterns and dynamics of river–aquifer exchange with variably-saturated flow using a fully-coupled model. *Journal of Hydrology*, *375*(3-4), 383–393.

Gasper, F., Görgen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., & Kollet, S. (2014). Implementation and scaling of the fully coupled terrestrial systems modeling platform (terrsysmp v1. 0) in a massively parallel supercomputing environment–a case study on juqueen (ibm blue gene/q). *Geosci. Model Dev*, *7*(5), 2531–2543.

Guay, C., Nastev, M., Paniconi, C., & Sulis, M. (2013). Comparison of two modeling approaches for groundwater–surface water interactions. *Hydrological Processes*, *27*(16), 2258–2270.

Healy, R. W. (2010). *Estimating groundwater recharge*. Cambridge University Press.

Healy, R. W., & Cook, P. G. (2002). Using groundwater levels to estimate recharge. *Hydrogeology journal*, *10*(1), 91–109.

Heilweil, V. M., Solomon, D. K., & Gardner, P. M. (2006). Borehole environmental tracers for evaluating net infiltration and recharge through desert bedrock. *Vadose Zone Journal*, *5*(1), 98–120.

Hendricks Franssen, H.-J., Brunner, P., Makobo, P., & Kinzelbach, W. (2008). Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images. *Water Resources Research*, *44*(1).

Höglauer, S., Corcoran, F., Bresinsky, L., Sauter, M., Renard, P., Gebel, M., & Engelhardt, I. (2019). Quantification of large-scale and long-term ground-water recharge and water resources in karst aquifers under mediterranean climate: deterministic versus stochastic approaches. In *Mid-term conference–frankfurt am main, germany 20-21 february 2019* (p. 24).

Jones, J. E., & Woodward, C. S. (2001). Newton–krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, *24*(7), 763–774.

Kearns, A. K., & Hendrickx, J. M. (1998). Temporal variability of diffuse groundwater recharge in new mexico. *NASA*(19980218685).

Keese, K., Scanlon, B. R., & Reedy, R. C. (2005). Assessing controls on diffuse groundwater

recharge using unsaturated flow modeling. *Water Resources Research*, *41*(6).

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.-O., Fischer, T., ... others (2012). Opengeosys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (thm/c) processes in porous media. *Environmental Earth Sciences*, *67*(2), 589–599.

Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, *29*(7), 945–958.

Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, *44*(2).

Labrecque, G., Chesnaux, R., & Boucher, M.-A. (2020). Water-table fluctuation method for assessing aquifer recharge: application to canadian aquifers and comparison with other methods. *Hydrogeology Journal*, *28*(2), 521–533.

Lee, C.-H., Chen, W.-P., & Lee, R.-H. (2006). Estimation of groundwater recharge using water balance coupled with base-flow-record estimation and stable-base-flow analysis. *Environmental Geology*, *51*(1), 73–82.

Lerner, D. N., Issar, A. S., & Simmers, I. (1990). *Groundwater recharge: a guide to understanding and estimating natural recharge* (Vol. 8). Heise Hannover.

Li, Z., Chen, X., Liu, W., & Si, B. (2017). Determination of groundwater recharge mechanism in the deep loessial unsaturated zone by environmental tracers. *Science of the Total Environment*, *586*, 827–835.

Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, *53*, 109–117.

Maxwell, R. M., Condon, L., & Kollet, S. (2015). A high-resolution simulation of groundwater and surface water over most of the continental us with the integrated hydrologic model parflow v3. *Geoscientific model development*, *8*(3), 923.

Maxwell, R. M., & Miller, N. L. (2005). Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*, *6*(3), 233–247.

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., ... others (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water resources research*, *50*(2),

1531–1549.

McMahon, P., Plummer, L., Böhlke, J., Shapiro, S., & Hinkle, S. (2011). A comparison of recharge rates in aquifers of the united states based on groundwater-age data. *Hydrogeology Journal*, *19*(4), 779–800.

Meinzer, O., & Stearns, N. D. (1929). *A study of ground water in the pomperaug basin, connecticut: with special reference to intake and discharge* (Tech. Rep.). Govt. Print. Off.,.

Naz, B. S., Sharples, W., Goergen, K., & Kollet, S. (2020). High-resolution pan-european multi-model simulations of hydrologic states and fluxes. In *Egu general assembly conference abstracts* (p. 10652).

Neumann, J., & Wycisk, P. (2003). Mittlere jährliche grundwasserneubildung. *Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit: Bonn, Germany*.

Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P., ... others (2004). Technical description of the community land model (clm). *Tech. Note NCAR/TN-461+ STR*.

Oloruntoba, B. J., Kollet, S. J., Montzka, C., Vereecken, H., & Franssen, H.-J. H. (2021). High resolution land surface modelling over africa: The role of spatial and temporal resolution. In *Agu fall meeting 2021*.

O'Neill, M. M., Tijerina, D. T., Condon, L. E., & Maxwell, R. M. (2021). Assessment of the parflow–clm conus 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous united states. *Geoscientific Model Development*, *14*(12), 7223–7254.

O'Reilly, A. M. (2004). *A method for simulating transient ground-water recharge in deep water-table settings in central florida by using a simple water-balance/transfer-function model* (No. 2004). US Geological Survey.

Qin, D., Qian, Y., Han, L., Wang, Z., Li, C., & Zhao, Z. (2011). Assessing impact of irrigation water on groundwater recharge and quality in arid environment using cfcs, tritium and stable isotopes, in the zhangye basin, northwest china. *Journal of Hydrology*, *405*(1-2), 194–208.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, *1*(5), 318–333.

Schalge, B., Baroni, G., Haese, B., Erdal, D., Geppert, G., Saavedra, P., ... others (2021). Presentation and discussion of the high-resolution atmosphere–land-surface–

subsurface simulation dataset of the simulated neckar catchment for the period 2007–2015. *Earth System Science Data*, *13*(9), 4437–4464.

Schlosser, P., Stute, M., Sonntag, C., & Münnich, K. O. (1989). Tritiogenic 3he in shallow groundwater. *Earth and Planetary Science Letters*, *94*(3-4), 245–256.

Shrestha, P., Sulis, M., Masbou, M., Kollet, S., & Simmer, C. (2014). A scale-consistent terrestrial systems modeling platform based on cosmo, clm, and parflow. *Monthly weather review*, *142*(9), 3466–3483.

Solomon, D., & Sudicky, E. (1991). Tritium and helium 3 isotope ratios for direct estimation of spatial variations in groundwater recharge. *Water Resources Research*, *27*(9), 2309–2319.

Sophocleous, M. A. (1991). Combining the soilwater balance and water-level fluctuation methods to estimate natural groundwater recharge: practical aspects. *Journal of hydrology*, *124*(3-4), 229–241.

Szilagyi, J., Harvey, F. E., & Ayers, J. F. (2003). Regional estimation of base recharge to ground water using water balance and a base-flow index. *Ground Water*, *41*(4), 504–513.

Therrien, R., McLaren, R., Sudicky, E., & Panday, S. (2010). Hydrogeosphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. *Groundwater Simulations Group, University of Waterloo, Waterloo, ON*.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., & Tóth, G. (2015). New generation of hydraulic pedotransfer functions for europe. *European journal of soil science*, *66*(1), 226–238.

US National Research Council. (1993). *Ground water vulnerability assessment: Predicting relative contamination potential under conditions of uncertainty.* Washington, DC: The National Academies Press.

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil science society of America journal*, *44*(5), 892–898.

Waldowski, B. (2022). *Estimating groundwater recharge in fully integrated hydrological models.* doi: https://doi.org/10.5281/zenodo.6244618

Wang, B., Jin, M., Nimmo, J. R., Yang, L., & Wang, W. (2008). Estimating groundwater recharge in hebei plain, china under varying land use practices using tritium and bromide tracers. *Journal of Hydrology*, *356*(1-2), 209–222.

Yang, Z., Kurtz, W., Gebler, S., Schüler, L., Kollet, S., Vereecken, H., & Hendricks-Franssen, H.-J. (2020). Long term hydrologic simulations for the meso-scale catchments rur and bode in germany by tsmp. In *Egu general assembly conference abstracts* (p. 5696).

Yin, L., Hu, G., Huang, J., Wen, D., Dong, J., Wang, X., & Li, H. (2011). Groundwater-recharge estimation in the ordos plateau, china: comparison of methods. *Hydrogeology Journal*, *19*(8), 1563–1575.