]RemkoUijlenhoet

Unsaturated zone model complexity for the assimilation of evapotranspiration rates in groundwater modeling

Simone Gelsinari[1,2], Valentijn R.N. Pauwels[1], Edoardo Daly[1], Jos van Dam[3], [6[4], Nicholas Fewster-Young[5], and Rebecca Doble[2]

[1]Department of Civil Engineering, Monash University, Clayton, Victoria, Australia
[2]CSIRO Land and Water, Waite Campus, Glen Osmond, South Australia, Australia
[3]Soil physics and Land Management, Wageningen University & Research, Wageningen, The Netherlands
[4]Hydrology and Quantitative Water Management Group, Wageningen University & Research, Wageningen, The Netherlands
[5]UniSA STEM, University of South Australia, Adelaide, South Australia, Australia
[6]Department of Water Management, Delft University of Technology, Delft, The Netherlands

Correspondence: Simone Gelsinari (simone.gelsinari@monash.edu)

Abstract. The bio-physical processes occurring in the unsaturated zone have a direct impact on the water table (WT) dynamics. Representing these processes through the application of Unsaturated Zone Models (UZMs) of different complexity has an impact on the estimates of[SG:] ~~recharge rates. In coupled configurations with UZMs, these recharge rates are often used as input for groundwater models and drive the water table dynamic. Because recharge estimates are always affected by uncertainty, model-data fusion methods, such as data assimilation, can be used to reduce the uncertainty in the model results. In this study, the required complexity of the UZM to update groundwater models through the assimilation of evapotranspiration (ET) rates is assessed for a water-limited site in South Australia. ET rates are assimilated because they have been shown to be related to the groundwater table dynamics, and thus form the link between remote sensing data and the deeper parts of the soil profile. It has been found, under the test site conditions, that a conceptual UZM can be used to improve groundwater model results through the assimilation of ET rates~~ the volumes of water flowing between the the unsaturated zone and the groundwater region. These values, known as net-recharge, are often used as the shared variable that couples UZMs to groundwater models. However, as recharge estimates are always affected by a degree of uncertainty, model-data fusion methods, such as data assimilation, can be used to inform these coupled models and reduce uncertainty. This study assesses the effect of UZM complexity (conceptual versus physically-based) to update groundwater model outputs, through the assimilation of actual evapotranspiration (AET) rates, for a water-limited site in South Australia. AET rates are assimilated because they have been shown to be related to the WT dynamics, and thus form the link between remote sensing data and the deeper parts of the soil profile. Results have been quantified using standard metrics, such as RMSE and r, and reinforced by calculating the $CRPS$, which is specifically designed to determine a more representative error in stochastic models. It has been found that, once properly calibrated to reproduce the AET-WT dynamics, a simple conceptual model may be sufficient for this purpose, thus using one configuration over the other should be motivated by the specific purpose of the simulation and the information available.

1 Introduction

Actual evapotranspiration (AET) and SG:groundwater recharge to the water table (WT) are two SG:related major components of the water cycle. SG:This is because AET is a function of the soil water content within the root zone, as the root water uptake is distributed along the entire root system (Grinevskii, 2011; Neumann and Cardon, 2012). Improving AET estimates, by means of a detailed modeling of the soil water transport, can SG:~~lead to better~~enhance the simulation of SG:net recharge (i.e. recharge to the WT minus transpiration from WT) and, in turn, WT dynamics. This is particularly important when the WT is within the reach of the roots, as is common in Australian semi-arid catchments (Banks et al., 2011), SG:~~where the direct transpiration from the WT is a major contribution to the total ET~~ because the root water uptake from groundwater and the capillary fringe can largely contribute to AET (Mensforth et al., 1994; Orellana et al., 2012).

AET is often simulated through SG:a variety of numerical models that reproduce the soil water-vegetation interaction SG:with different level of details. Advanced integrated surface water-groundwater models (e.g. Hydrogeosphere (Therrien et al., 2006), CATHY (Camporese et al., 2010), PARFLOW (Jones and Woodward, 2001)) or coupled saturated-unsaturated zone models (Facchi et al., 2004; Simunek et al., 2009; Zhu et al., 2012; Van Walsum and Veldhuizen, 2011; Grimaldi et al., 2015) are able to account for the direct groundwater-vegetation interaction. In general, the representation of the unsaturated zone is obtained SG:~~with~~from simple conceptual water balance models or detailed physically-based models.

Conceptual unsaturated zone models (UZMs) simplify the processes occurring in the unsaturated zone, and are widely used for spatially distributed hydrological simulations (Teuling and Troch, 2005). An example is Batelaan and De Smedt (2007), who successfully applied a coupled surface-groundwater balance model at the regional scale focusing on the assessment of recharge rates. Conceptual water balance models have been found to be flexible as they usually require shorter run times and fewer parameters, and are suitable when stochastic simulations based on Monte-Carlo techniques are applied (Kim and Stricker, 1996; Fatichi et al., 2016). However, for more detailed simulations, such as in ecohydrology or agricultural modeling, simple UZMs may fail to accurately simulate important processes SG:~~(i.e. water stress, root growth)~~such as water stress or root growth (Krysanova and Arnold, 2008), thus physically-based models are preferred. SG:~~These models commonly~~Commonly, physically-based models solve the Richards equation for water flow in porous media, relying on SG:~~the~~ relationships between SG:soil volumetric water content, hydraulic conductivity and soil water pressure head (van Dam et al., 2008; Scheerlinck et al., 2009). SG:~~Therefore~~Physically-based models have SG:thus the ability to account for specific effects SG:~~(e.g. capillary rise)~~ that affect the calculation of ETSG:, such as capillary rise, SG:~~thus~~thereby impacting recharge estimates. The latter is particularly important when UZMs are coupled to saturated models as recharge acts as the link between both models (Doble et al., 2017).

SG:~~Because of the number and~~The spatial variability SG:and number of parameters (e.g. the water retention curve and detailed vegetation characteristics) required by physically-based models, their application, particularly in data scarce areas, can be challenging (Simmons and Meyer, 2000). On the other hand, conceptual models may require fewer input data, SG:~~but~~however, their recharge estimates may be less reliable. This SG: ~~is~~occurs because they are affected by both structural uncertainty, induced by the simplification of the model (Renard et al., 2010), and the epistemic and aleatory uncertainty of the forcing inputs (Khatami et al., 2019). Accurate model parameters and meteorological inputs are far from always available, especially at large spatial scales. Therefore, the use of remote sensing data can provide vital information for these models (Entekhabi and Moghaddam, 2007; Carroll et al., 2015; Lu et al., 2020).

One way to make use of the remote sensing observations is through data assimilation, which SG:~~is a method to~~ combines model results with independent observations to reduce model uncertainty. ~~There is a plethora of studies on the assimilation of diverse observations (e.g. soil moisture (SM), leaf area index, streamflow and groundwater levels) in hydrology,~~In the field of hydrology, there is a plethora of studies on the assimilation of diverse observations such as soil moisture (SM), leaf area index, streamflow and groundwater levels) SG: ~~which have been reviewed several times in the last decade by various authors~~ (Liu et al., 2012; Li et al., 2016). SG:~~Many of these studies are based on~~ Satellite remote sensing data SG:~~, which~~ have been proven to be a valid alternative when field-based observations cannot provide sufficiently accurate measurements. Remotely sensed SM values are a function of the water content of the upper few centimeters of the soil (Pipunic et al., 2014). Consequently, models using remotely sensed SM assimilation extrapolate the update for the upper soil layer to the entire modeled soil column through the covariance between the upper and lower layer modeled SM values. On the other hand, remotely sensed ET rates are a function of the modelled water content of the soil column up to the rooting depth. SG:In consideration of this, the assimilation of remotely sensed AET SG:~~assimilation thus directly updates~~has the potential of directly updating the water content of the entire modeled soil column. In recent years, the assimilation of satellite-based ET observations has been recognised to be beneficial for the reduction of the uncertainty of several hydrogeological products (e.g. recharge and depth to WT), especially for data-scarce areas (Entekhabi and Moghaddam, 2007; Doble et al., 2017; Gelsinari et al., 2020)).

SG:~~However,~~ All satellite observations present a trade-off between accuracy, time-frequency and spatial coverageSG:~~, and~~. In addition, no satellite retrievals are SG:~~not~~ free from errorsSG:~~. For instance~~, as exposed in Long et al. (2014), who analyzed and compared the uncertainty in the ET estimates from different sources including the Moderate-Resolution Imaging Spectroradiometer (MODIS). They concluded that ET derived from land surface models had a lower uncertainty than the MODIS based ET (5 mm/month vs 12.5 mm/month, respectively), and suggested a hybrid approach for taking advantage of the integration of land surface models and remotely sensed products SG:~~to reduce uncertainty~~. Droogers et al. (2010), SG:~~using a physically-based UZM,~~ applied an inverse modeling approach (i.e. forward-backwards optimization) SG:using a physically-based UZM, and found that improvements were obtained when the frequency of the ET observations was finer than a 15-day interval. SG:~~Therefore, an assimilation algorithm that correctly accounts for the observation errors when assimilating remotely sensed ET observations into UZMs should be used for this purpose.~~It appears that, for the purpose of proficiently using ET retrievals, the assimilation framework should allow frequent updates (< 15-days interval), and account for observation errors. This was also shown synthetically by Gelsinari et al. (2020) who SG:who improved the model outputs, using the Ensemble Kalman Filter (EnKF) for the sequential assimilation of the averaged 8-days ET into a conceptual UZM coupled to MODFLOW (Harbaugh, 2005) SG:~~, improving the model outputs~~. SG:~~Unlocking the potential~~

~~of using ET observations to inform models, with the aim of reducing the uncertainty in the outputs, is a currently active area of research.~~ The assimilation of satellite ET observations have been shown to be a feasible way to constrain hydrologic models, SG:~~but~~however this has yet to be validated against experimental data. Furthermore, it is known that UZMs of different complexity can yield different ET estimates, producing distinct recharge values and, in turn, a diverse dynamics of the WT.

This study aims to perform the validation of the AET assimilation framework proposed SG:~~synthetically~~ in Gelsinari et al. (2020) and to assessSG: ~~the UZM complexity required for the assimilation to positively update groundwater models~~ the use of a conceptual and a physically-based UZM within a data assimilation framework to improve WT estimates.. SG:~~In particular,~~ The quantities of interest are the temporal WT fluctuation dynamics and the modeled AET. A conceptual and a physically-based UZM are coupled to MODFLOW, and applied to a water-limited study site in the south-east of South Australia. Remotely sensed ET observations are assimilated into both SG:of these coupled models, and SG:~~an assessment of the improvements in the model results is made~~assessments of the improvements in the results of the model are made. Based on this assessment, a number of recommendations regarding the required UZM complexity to obtain a positive impact on the quantities of interest are made.

2 Methods

2.1 Study Area and Data

The study area is situated in the south-eastern part of South Australia, north of the city of Mount Gambier (See figure 1[a],[b]). This region has a Mediterranean climate with cool wet winters and warm dry summers. SG:The climatic forcing inputs are rainfall and potential ET (PET) obtained from the Bureau of Meteorology SG:~~(BOM) station number 26021~~(BOM - station n. 26021). The historical data for this station report an average annual rainfall and potential ET of approximately 710 and 980 mm$\cdot$year^{-1}, respectively, calculated over the period 1942-2017. The Morton equation (Donohue et al., 2010) and the Budyko-curve (Donohue et al., 2007) SG:~~thus~~ classify the area as dominated by evapotranspiration SG:~~,~~ or water-limited (Jackson et al., 2009; Benyon et al., 2006).

The study site is a *Pinus Radiata* plantation next to the Mount Gambier airport (Figure 1[c]). The area was originally planted in July 1996 with a density of 1225 trees/ha SG:~~,~~ and there was no thinning of the plantation during the observations. The survey performed by Benyon et al. (2006) classified the soil as duplex. This type of soil presents a contrast between the upper part, which features sandy-loam characteristics with high hydraulic conductivity, and the lower part, classified as clay, with a finer texture and lower hydraulic conductivity. The average WT depth, from the observations at one bore, is reported at approximately 6 meters below the surface. SM observations were taken SG:with a neutron probe, about every 4 weeks from August 2000 to January 2005, SG: up to a depth of 3 meters at an interval of 30 cm. SG:~~The campaign was conducted from August 2000 to January 2005 with an average measurement frequency of 4 weeks.~~ SG:~~Because in the area~~In this region more than 90% of the available groundwater is

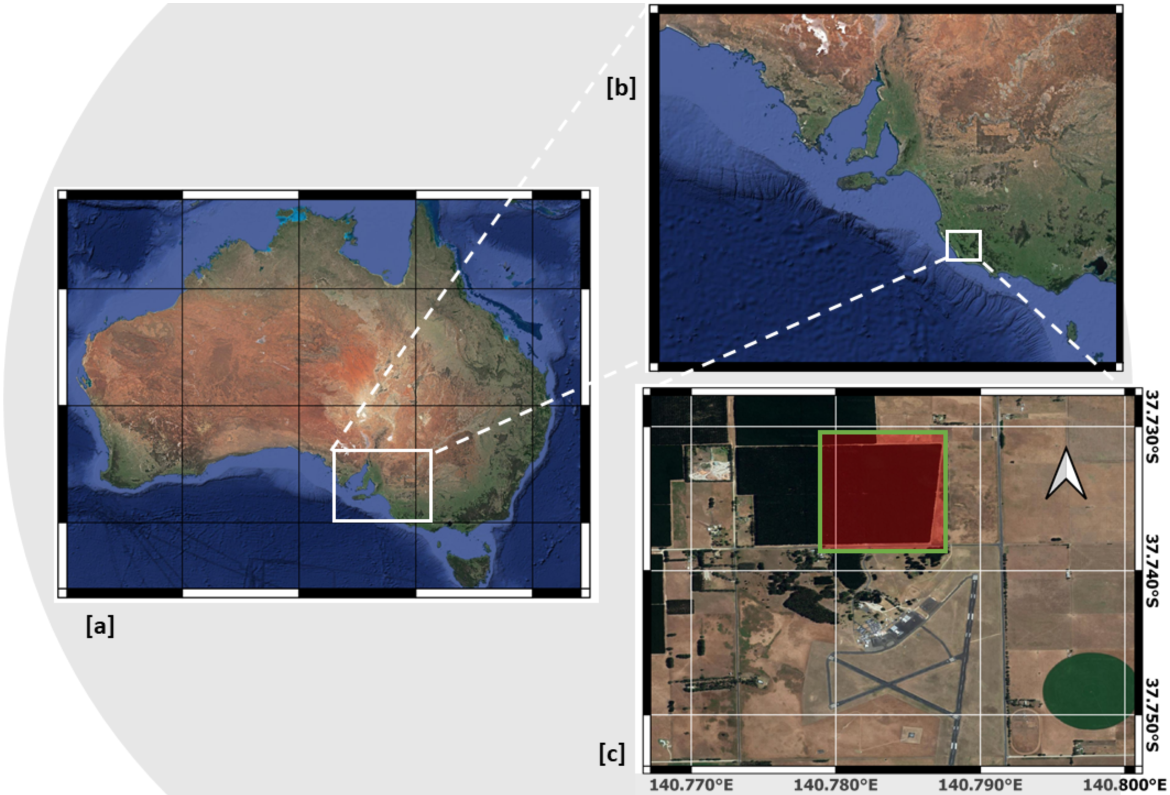

Figure 1. Localization of the study area within Australia [a], The South East of South Australia [b], and a detail of the forest plantation [c]. The red square indicates the CMRSET tile. © Google Maps

in shallow aquifers[SG:]~~,~~ and these plantations have been shown to have direct access to groundwater (Benyon and Doody, 2004).

[SG:]~~Remotely sensed data of actual ET from the CSIRO MODIS reflectance-based scaling evapotranspiration (CMRSET) algorithm were used~~AET data are derived from the remotely sensed CSIRO MODIS reflectance-based scaling evapotranspiration (CMRSET) algorithm (Guerschman et al., 2009). These values are obtained by rescaling the PET rates calculated with the Penman-Monteith algorithm using the Enhanced Vegetation Index and Global Vegetation Moisture Index obtained from the MODIS spectroradiometer (Swaffer et al., 2020). The observations are available every 8 days with [SG:]~~a finest~~a spatial resolution of 250 by 250 m.

2.2 Model Description

[SG:]~~Two different configurations of coupled groundwater-unsaturated zone models were tested. The UZMs conceptualization and the coupling to the groundwater model are shown in Figure 2. This section introduces the models and the coupling framework.~~The tests presented in this study used two different configurations of coupled groundwater-unsaturated zone models, which are depicted in Figure 2. The following sections describe the models as well as the coupling framework.

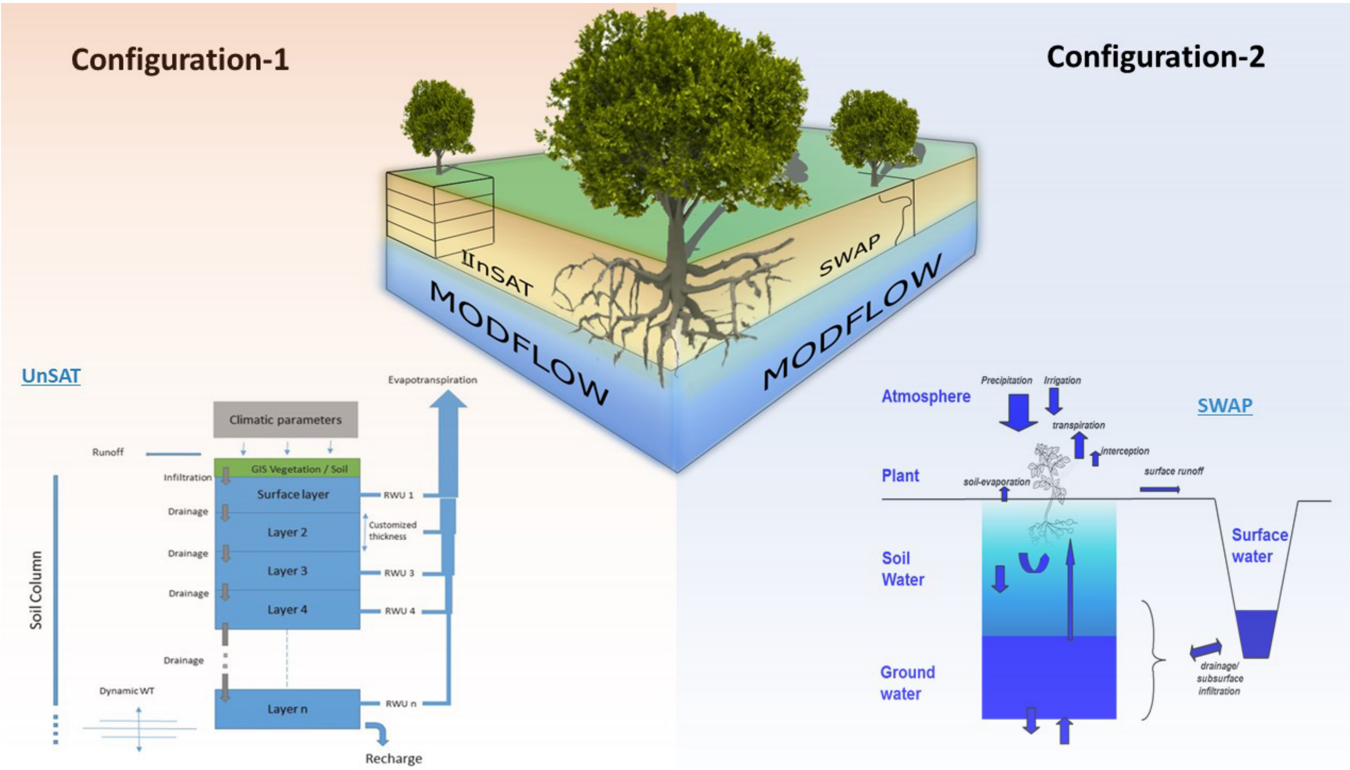

Figure 2. Coupled models representation. Left: UnSAT conceptualization coupled to MODFLOW. Right: SWAP conceptualization coupled to MODFLOW.

2.2.1 UnSAT - UZM

The UnSAT (**Un**saturated zone & **SAT**ellite) UZM is a one-dimensional soil water balance model. The unsaturated zone is divided into layers and the water balance of each layer is solved at every time step. SG: Water flows downward from the top layer to the last and the latter delivers recharge (see Figure 2). The model uses climate forcing data (i.e. precipitation [mm$\cdot$hr^{-1}] and PET [mm$\cdot$hr^{-1}]) on a raster distributed basis as inputs, and returns values of actual ET [mm$\cdot$hr^{-1}], runoff [mm$\cdot$hr^{-1}], recharge and soil water content (θ [mm^3 mm^{-3}]). The soil is parameterized using the porosity (θ_s), a critical soil-water content to define water stress (θ_* [mm^3 mm^{-3}]), residual soil-water content (θ_r [mm^3 mm^{-3}]) (as in Laio et al. (2001)), hydraulic conductivity (K$_s$ [mm$\cdot$hr^{-1}]), and an empirical value for drainage (b [-]); the root system is defined using root SG: ~~length~~depth (lr [mm]) and

[SG:] ~~the~~a root density distribution parameter (Vr [-]) [SG:] ~~as explained in~~ (Vrugt et al., 2001a).

[SG:] The size of the layers (Δz) remains constant while their number changes according to the depth to WT, which is provided by the groundwater model. Along the soil profile, the model accounts for water extraction due to root water uptake. For each layer, the water balance equation is solved using an explicit finite difference approximation, solved with an hourly time step (Δt). The water balance of the layer at soil surface is calculated as

$$\theta_1^{t+1} = \theta_1^t + \frac{P^t - AET_1^t - Q^t - D_1^t}{\Delta z_1} \cdot \Delta t \,, \tag{1}$$

[SG:] where θ [mm^3 mm^{-3}] is soil volumetric water content, P [mm h^{-1}] is precipitation, AET [mm h^{-1}] is actual ET, D [mm h^{-1}] is percolation, and Q [mm h^{-1}] is runoff. In Eq. 1, [SG:] the subscripts 1 refer to the the soil layer at the surface, and the superscripts refer to the time step. The percolation D_1^t which [SG:] proceeds to the lower layer, is defined by the Clapp–Hornberger (Clapp and Hornberger, 1978) [SG:] modification of the Brooks-Corey model.

[SG:] AET is calculated as:

$$AET = PET \cdot \beta(z) \cdot \alpha(\theta), \tag{2}$$

where $\beta(z)$ is the root distribution function as in Vrugt et al. (2001b), and α [SG:] is a water stress reduction function (Laio et al., 2001; Feddes et al., 1976).

[SG:] For the layers below the first, including the last layer which delivers recharge to the groundwater model, the water balance equation is

$$\theta_n^{t+1} = \theta_n^t + \frac{D_{n-1}^t - (AET_n^t + D_n^t)}{\Delta z_n} \cdot \Delta t \,. \tag{3}$$

[SG:] For a more detailed description of the UnSAT model see Gelsinari et al. (2020).

2.2.2 SWAP - UZM

The Soil Water Atmosphere Plant (SWAP v. 4.0) model, developed by Alterra, is one of the most used physically-based UZMs (van Dam et al. (2008); Kroes et al. (2017)). This agro-hydrological model [SG:] ~~applies the Richards equation and~~ is able to simulate the water, heat and solute flow in heterogeneous, variably saturated soils. [SG:] Water flow is simulated using the Richards equation. In addition, it has the potential of accounting for a detailed soil water vegetation interaction as it specifically simulates the dynamics of the crop growth cycle.

[SG:]~~SWAP has a long history of applications for climate change studies Droogers2008,Farkas2014, fire hazard evaluation Taufik2019, impact of land-use change studies Bennett2013, water use management Droogers2000, groundwater exploitation Li2019, and holistic assessment of the soil hydraulic properties Pinheiro2019.~~

In SWAP, the Richards equation is solved for the pressure head using finite differences. The soil hydraulic retention functions are based on the analytical formulations proposed by van Genuchten (1980). The model requires the van Genucthen soil parameters and a number of vegetation specific parameters (Feddes et al., 1976). In this study, the drought stress parameters are a result of the calibration. These are the pressure head below which water uptake reduction starts (i.e. -3000 mm) and the pressure head triggering no further water extraction (i.e. -30000 mm). [SG:]For this experiment, the standard forest root density distribution is [SG:]~~used~~applied. [SG:]~~SWAP has the ability to represent an internal saturated part of the soil column that is controlled by a specified head (simulating drains in the original conceptualization) at the boundary of the domain. In this study, the internal SWAP saturated function is neglected and replaced by the MODFLOW model.~~

2.2.3 Groundwater Model

The groundwater model chosen for the study is MODFLOW 2005 (Harbaugh, 2005). This modular flexible model has packages dedicated to the calculation of ET and the application of recharge to the groundwater. In this study, the ET package of MODFLOW (EVT) was replaced with the UZMs (UnSAT and SWAP), and the recharge (RCH) package was used to apply the UZMs calculated net-recharge to the cell-specific head.

[SG:]~~The aquifer saturated hydraulic conductivity (K_h) and specific yield (S_y) as well as model discretization are defined through FloPy, a library that allows MODFLOW to run in a Python environment~~ Flopy (Bakker et al., 2016) [SG:], a library that allows MODFLOW to run in a Python environment, was used to generate the saturated model and specify parameters, such as the aquifer saturated hydraulic conductivity (K_h) and specific yield (S_y). The model runs at an 8-day time-step, which is considered adequate for [SG:]~~GW~~WT dynamics. This choice was made to synchronize the models and assimilation time frequencies as the CMRSET data are available with a temporal resolution of 8 days.

2.2.4 Coupling

[SG:]~~The UZMs were coupled to MODFLOW according to their time steps.~~ UZMs require a shorter time step [SG:]than MODFLOW as the water content varies at a higher frequency than the depth to the WT in the groundwater model (Xu et al., 2012). Variation of the WT at regional scales usually can only be observed at temporal scales in the order of moths or years. Thus, applying a larger time step for the saturated zone model is a valuable option to reduce the computational time (Facchi et al., 2004). At large spatial scales, dimensional simplification to 1D unsaturated zone flow simulations has been shown to be sound because the direction of the unsaturated zone flow is predominantly vertical (Zhu et al., 2011).

Configuration-1 (Figure 2 left side) features the UnSAT model coupled to MODFLOW SG:~~through recharge~~. This configuration specifically accounts for plant transpiration from the WT by calculating the balance between recharge entering the WT (positive) and transpiration (negative). UnSAT runs at an hourly time step while MODFLOW runs with an 8-day time step, matching the SG: ~~MODIS~~CMRSET time step. Once MODFLOW has performed the calculation of the WT levels, these are fed back on a raster basis to UnSAT, which uses them to recalculate the number of layers in which the unsaturated zone is discretised. This dynamic scheme, defined in Zeng et al. (2019) as the non-iterative feedback coupling method, is considered a valuable trade-off between the computational cost of fully coupled or iterative schemes and numerical accuracy.

For Configuration-2 (Figure 2 right side), the unsaturated zone is simulated through the SWAP model, with the pressure head along the soil column as state variable. The model has been coupled to MODFLOW through the recharge similar to the coupling methodology reported by Xu et al. (2012). This way of coupling SG:~~the model~~ requires caution in the definition of the S_y parameter, which becomes part of the SG:~~deterministic calibration and is further explained in~~calibration (See Section 2.3).

2.3 Model Domain and Calibration

The SG:~~coupled~~ model configurations were applied to a domain of 1 x 5 cells of 1 km^2 each and a single vertical unconfined layer (Figure 3) SG:The domain discretisation was chosen as a result of a sensitivity analysis conducted on a range of model domains varying from fine (1 x 20 cells) to coarse (1 x 5 cells). The boundary cells were set to a constant head obtained via calibration (i.e. 3.5 m below the surface). SG:The location chosen allows to keep the model configuration as simple as possible, and setting the boundary conditions of the saturated model as constant head. This is due to the site being in the centre of a forestry block, more than two kilometres from any groundwater extraction, originally selected for a previous study to specifically look at the impacts of forestry on groundwater (Benyon et al., 2006). SG:For this region, where WT is 6 m deep or shallower, it has been shown that forestry transpiration from groundwater is around 2 orders of magnitude (i.e. 435 ML yr for a 1 km by 1 km fully forested cell) larger than the maximum groundwater extraction rate from a single bore. To further reinforce the selection of constant head boundary conditions, an analysis of the WT fluctuations was conducted on bores in proximity of the study area but outside of the forest, showing that for a WT level of 4.4 m below the surface with the standard deviation was low (i.e. 0.12 m). This supports the assumption that the if higher WT table fluctuations are observed (such as the investigated location), these are dependent on the local net recharge.

UnSAT can account for the decrease of K_s along the soil column, whereas SWAP is capable of explicitly modeling the heterogeneity of the soil column, as described in Section 2.1. Thus, for Configuration-1, SG:~~the decay of K_s is a result of the calibration, while other~~ soil parameters are homogeneous along the soil column length (i.e. 10 m)SG:~~.~~, while in Configuration-2, the first (Upper) 1.5 meters of soil is classified as "Sandy-Loam" soil and the second (Lower) is a "Loam-Clay" soil spanning the rest

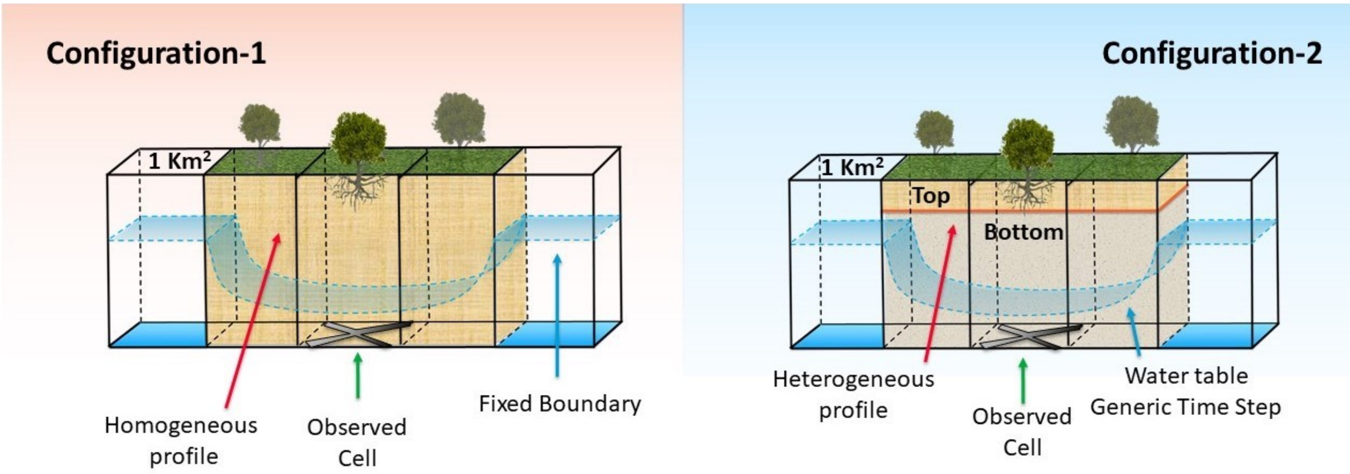

Figure 3. Schematic of the simulation domain. Configuration-1 models the unsaturated zone as a homogeneous profile with UnSAT. Configuration-2 models the soil heterogeneity by accounting for the change in soil properties with SWAP. SG: ~~The WT is represented at initial conditions on the left-hand side, and at a generic simulation time, showing the depression caused by the root water extraction, on the right-hand side~~The representation of WT levels refers to the generic simulation time, and is conceptually showing the depression caused by the root water extraction.

of the simulated soil column (i.e. 8.5 m).

SG: ~~Preliminary analyses of this study (not shown) indicated that, in order to get significant improvements in the model outputs, the link between WT depth and ET had to be accurately reproduced. For both configurations, attempting to assimilate ET fluxes without reproducing the interdependence between WT and actual ET yielded poor filter performances.~~In order for the system to be observable, the link between WT levels and AET has to be accurately reproduced. It should be noted that this link has been described in the literature (Shah et al., 2007; Xie et al., 2018; Zha, 2020). To account for this interdependence, SG: ~~and reduce the order of freedom of the ill-posed problem of calibration,~~ a multi-objective function (MOF) which combines WT depths and actual ET values, was introduced. Then, SM observations were used for refinement and to set boundaries to the soil parameters. The algorithm Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995; Shi and Eberhart, 1998) was used for calibration minimising the specifically defined MOF:

$$\mathrm{MOF} = \frac{\mathrm{RMSE}(WT)}{\sigma(WT)} + \frac{\mathrm{RMSE}(ET)}{\sigma(ET)}, \tag{4}$$

Table 1. Calibrated parameter values used for the simulations and their SG:~~perturbation fraction~~coefficient of variation.

Model Parameter	Configuration 1 UnSAT + MODFLOW Homogeneous	Configuration 2 SWAP + MODFLOW Top | Bottom	SG:~~Perturbation~~Coefficient SG:~~fraction~~of variation %
Hydraulic conductivity - K_s [mm·hr^{-1}]	25	24 | 41	10
Drought Stress (Reduction) [mm]	-	-3000	-
Drought Stress (No extraction) [mm]	-	-30000	-
Oxygen Stress (Reduction) [mm]	-	-100	-
Oxygen Stress (No extraction) [mm]	-	+ 5	-
Soil porosity [mm^3·mm^{-3}]	0.35	0.36 | 0.36	-
Critical transpiration SM (θ_*) [mm^3·mm^{-3}]	0.12	-	-
Residual SM (θ_r) [mm^3·mm^{-3}]	0.03	0.01 | 0.02	-
Drainage empirical value [-]	2.50	-	-
Root depth (lr) [mm]	8000	2900	10
Root distribution parameter (Vr) [-]	0.5	-	-
MODFLOW K_h [m·d^{-1}]	10.0	8.0	10
MODFLOW S_y [-]	0.12	0.11	10

where RMSE is the Root Mean Square Error, and σ is standard deviation. PSO searches the n-dimensional solution space, where n is the number of parameters given, in order to minimise equation 4. The calibrated parameters are listed in Table 1.

SG:~~For each configuration, the observation data set was divided into two periods used for calibration and validation~~By applying a calibration/validation approach, the observation datasets were divided into two periods. For calibration, 46 8-day time steps covering roughly the year 2001 was used, while the rest of the data set (4.5 years in total) were used for validation.

2.4 Assimilation

The EnKF (Evensen, 1994) was used because of its SG:~~ability to deal~~ reduced computational burden when dealing with highly non-linear systems. The filter initially requires the establishment of a number of ensemble members, generated by perturbing the forcing inputs of precipitation and PET. SG:After having tested other ensemble sample sizes (i.e. 16, 32, 64), the ensemble SG:~~population~~sample size was set to SG:M = 32, a size which has been widely used for a number of EnKF applications (Mitchell et al., 2002; Pauwels et al., 2013) SG:, and represented the best trade-off between computational time and accuracy for the case tested. To verify the spread and accuracy of the ensemble, a number of statistical variables, originally developed for numerical weather prediction by Talagrand et al. (1997), were calculated on the ensemble population (see Section 2.4.1).

Usually, in data assimilation studies, the assimilated observations are model states (also called prognostic variables) such as SM, pressure head[SG:], and WT levels. This paper uses AET flux observations, which are diagnostic variables [SG:]~~of the coupled configurations~~. Therefore, the interaction between AET and model states occurs in the UZM, of which AET is a model result. Following [SG:]~~the findings of~~ Gelsinari et al. (2020), AET data from [SG:]~~MODIS (CMRSET data set)~~CMRSET are assimilated into the coupled model configurations.

The two configurations apply a similar scheme of the EnKF, the difference lying in the composition of the [SG:]aggregated state vector, as the state variables of the UZMs are different. [SG:]~~More s~~Specifically, the state vector of Configuration-1, for a single ensemble member ($i = 1, \cdots, M$) is composed of the WT level h and [SG:]a vector of the SM values at time step s, [SG:]~~reading~~represented as:

$$\mathbf{z}_{[\mathbf{1}]_s}^{i,f} = [\theta_1 \quad \theta_2 \cdots \theta_n] \tag{5}$$

where $\theta_1, \theta_2, \cdots, \theta_n$ are the [SG:]values of SM [SG:]~~values of the UZM layers~~content for each layer of the UZM, for the i-th ensemble member, and f representing forecast.

For Configuration-2, the [SG:]~~state vector is similarly composed and reads~~vector of soil water pressure heads is

$$\mathbf{z}_{[\mathbf{2}]_s}^{i,f} = [p_1 \quad p_2 \cdots p_n] \tag{6}$$

where, for the i-th ensemble member, $p_1, p_2, \cdots, p_n$ are the pressure head values for each layer of the UZM. The filter scheme is then similarly applied for both configurations as follows. [SG:]Here, only the aggregated state vector of Configuration-1 (composed in the same fashion for both configurations) for the assimilation time step k and the ensemble member i is reported. This is

[SG:]~~The aggregated state vector for the assimilation time step k and the ensemble member i is then composed in the same fashion for both configurations. The aggregated vector of Configuration-1 is~~

$$\mathbf{x}_k^{i,f} = [h^{i,f}, \quad \mathbf{z}_{[1]_1}^{i,f}, \mathbf{z}_{[1]_2}^{i,f}, \cdots, \mathbf{z}_{[1]_t}^{i,f}]^T, \tag{7}$$

where t is the number of times the UZM model is applied [SG:]~~during the assimilation time step~~between two applications of the filter (i.e 8-days), [SG:]~~which is different in the two configurations,~~ T indicates the transposed vector, and h is the WT level, [SG:]~~permanent~~constant during the t time steps, simulated by MODFLOW.

SG: The average state vector reads:

$$\overline{\mathbf{x}}_k^f = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_k^{i,f}\,. \tag{8}$$

SG: To compose the state deviation matrix, the value of $\overline{\mathbf{x}}_k^f$ is subtracted from the elements of the state vector as:

$$\mathbf{X}_k^f = [\mathbf{x}_k^{1,f} - \overline{\mathbf{x}}_k^f \quad \mathbf{x}_k^{2,f} - \overline{\mathbf{x}}_k^f \quad \mathbf{x}_k^{3,f} - \overline{\mathbf{x}}_k^f \cdots \mathbf{x}_k^{M,f} - \overline{\mathbf{x}}_k^f] \tag{9}$$

The observation from the CMRSET for the k time step is the vector

$$\mathbf{y}_k = AET_k \tag{10}$$

SG: which is a scalar value because of the choice of matching observation and assimilation frequencies.

Because of the 8-days frequency of the observations, the average AET over 8 days simulated by the models is

$$\hat{\mathbf{y}}_k^{i,f} = \frac{1}{8}\sum_{s=1}^{t} ET_s^{i,f}\,, \tag{11}$$

SG: with s being the individual UZM steps. The average over the ensemble population (M) of equation 11 reads

$$\overline{\mathbf{y}}_k^f = \frac{1}{M}\sum_{i=1}^{M}\hat{\mathbf{y}}_k^{i,f}\,. \tag{12}$$

SG: The matrix for observation-simulation deviation is composed as:

$$\mathbf{Y}_k^f = [\hat{\mathbf{y}}_k^{1,f} - \overline{\mathbf{y}}_k^f \quad \hat{\mathbf{y}}_k^{2,f} - \overline{\mathbf{y}}_k^f \quad \hat{\mathbf{y}}_k^{3,f} - \overline{\mathbf{y}}_k^f \cdots \hat{\mathbf{y}}_k^{M,f} - \overline{\mathbf{y}}_k^f]\,. \tag{13}$$

SG: Combining the matrices calculated above it is possible to calculate the background state covariance matrix

$$\mathbf{PH}^T = \frac{1}{M-1}\mathbf{X}_k^f\mathbf{Y}_k^{fT} \tag{14}$$

SG: and the observation-simulation error covariance matrix

$$\mathbf{HPH}^T = \frac{1}{M-1}\mathbf{Y}_k^f \mathbf{Y}_k^{fT}. \tag{15}$$

SG: These lead to the formulation of the Kalman Gain as:

$$\mathbf{K}_k = \frac{\mathbf{PH}^T}{\mathbf{HPH}^T + \mathbf{R}_k} \tag{16}$$

where $\mathbf{R}_k$ SG: is the observation error covariance matrix. The Kalman gain transfers the difference between the observed and simulated ET to the state variables with the updating equation

$$\mathbf{x}_k^{i,a} = \mathbf{x}_k^{i,f} + \mathbf{K}_k[\mathbf{y}_k - \hat{\mathbf{y}}_k^{i,f} + \mathbf{v}_k^i], \tag{17}$$

where $\mathbf{v}_k^i$ SG: is a random number with mean 0 and standard deviation as the observation error (i.e. 0.2 mm/day).

According to SG: ~~the findings in~~ Gelsinari et al. (2020), the state variable update had to be SG: ~~limited~~constrained to preserve numerical stability. This was equally true for both models and applies to the WT levels SG: and SM. A limitation of $\pm 50\%$ of the prior values is applied for the SM content of Configuration-1 and, similarly, to the pressure head variable of Configuration-2. This avoids the convergence problem in physically-based models reported in Zhang et al. (2018).

2.4.1 Ensemble Generation

The generation of a statistically meaningful ensemble, which preserves the relationship between ET and WT levels obtained during the calibration, is crucial for the application of the EnKF (Gelsinari et al., 2020). A number of ensemble generation techniques were applied to the two configurations, and a consistent approach for both configurations was adopted. SG: The average over the verification period of the ratios between ensemble skill and ensemble spread, which should tend to 1, and between ensemble skill and mean squared error, which should tend to $\sqrt{(M+1)/2M}$ (Talagrand et al., 1997; De Lannoy et al., 2006)SG: , were calculated on the modeled ET values as in Gelsinari et al. (2020). First, a simple perturbation of forcing inputs, by adding a random number sampled from Gaussian distributions with different standard deviations, as performed by Gelsinari et al. (2020), was tested.

SG: By perturbing the forcing inputs alone, the ensemble spread was not reaching appropriate values. SG: ~~Then~~Thus, a mixed method involving the perturbation of both inputs and parameters, with the latter perturbed by adding a random number proportionally to the calibrated value, was applied. For the UZMs, the parameters selected for the perturbation were K_s and root depth, and for MODFLOW the saturated K_h and S_y. Initial conditions of WT levels were also perturbed to induce a good spread in the ensemble from the early stages of the simulation. [SG:] The Talagrand et al. (1997) verification skills were applied to the ensembles generated with the aforementioned approach, and the most adequate ensembles for the two configurations were retained. The scores obtained for the two ratios were comparable to others found in the literature (e.g. De Lannoy et al. (2006); Pauwels and De Lannoy (2009); Gelsinari et al. (2020)). [SG:] ~~This ensemble of simulation is~~These ensembles are defined as the open loop, which represents the "prior" distribution. After applying the filter, the resulting distribution is called the assimilation run and represents the "posterior".

[SG:] ~~In such a nonlinear configuration, it is a challenge to generate ensembles that maintain the statistical accuracy, and simultaneously preserve the ET - WT relationship. The most adequate ensembles for the two configurations, obtained by calculating the ensemble validation skills on the modeled ET based on the method explained in, were retained. Results are shown in Section 3.3.~~

2.5 Verification Skills

In this section, [SG:] ~~the performance of the filter applied to the assimilation results and the respective open loop runs are assessed~~the results for the open loop and assimilation runs are assessed. This is conducted by analysing the error between the predicted models and the observations. The common quantifiable measures used to assess the overall errors in these models are the Root Mean Square Error (RMSE), the Pearson correlation coefficient (r), and the Continuous Ranked Probability Score (CRPS) (Hersbach, 2000).

[SG:] ~~The assimilation skills are evaluated using the Root Mean Square Error (RMSE) and the Pearson correlation coefficient defined as: two common metrics, the Root Mean Square Error (RMSE) and the Pearson correlation coefficient , jointly with the Continuous Ranked Probability Score (CRPS), which presents a good quantification of the difference between the predicted and the observed cumulative distributions.~~

The RMSE and r are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{L}\sum_{k=1}^{L}(o_k - f_k)^2}, \tag{18}$$

[SG:] where o_k are the observations, f_k are the modelled variables at time step k and L is the size of the data set. The RMSE is an important metric which measures the square of the difference in the errors and is presented to convey a possible board overall error between the observations and the model. A disadvantage of RSME is the excessive weight of large outliers.

[SG:]~~where o_k is the observation and f_k is the modeled variable at time k and L is the size of the data set. The filter performance are assessed comparing these metrics, applied to the assimilation results to the respective open loop run.~~

[SG:]The next component in analysing the performance in verifying the results is the Pearson correlation coefficient to understanding the relationship between the observed values and the predicted model values. The Pearson correlation coefficient (r) is defined as

$$r = \frac{\sum_{k=1}^{L}(o_k - \overline{o})(f_k - \overline{f})}{\sqrt{\sum_{k=1}^{L}(o_k - \overline{o})^2 \cdot \sum_{k=1}^{L}(f_k - \overline{f})^2}}, \tag{19}$$

[SG:]In particular, this is investigating the strength of the linear relationship between the predicted and observed values as they proceeds through time. A value of $r > 0$ implies a positive relationship; the closer the value of r is to 1 the stronger and more accurate the relationship.

The CRPS is a measure to quantify the difference between the predicted value and the observed cumulative distribution in terms of the probabilistic distributions for each time step. The CRPS is calculated, at a specific time step, from the cumulative distribution function given by the ensemble simulation of the variable of interest x (i.e. ET and WT levels) as follows:

$$\mathrm{CRPS}_k = \int_{-\infty}^{+\infty} (P(x)_k - P_0(x)_k)^2 \, dx \, , \tag{20}$$

[SG:]where P_0 is the observation distribution at the time step (k), and $P(x)$ is the cumulative distribution function. As the observation (x_0) is usually a single value, P_0 is formulated as $P_0 = H(x - x_0)$, with H being the Heaviside function. The CRPS is applied to reinforce the value of the result as it is specifically designed to assess probabilistic simulations and is increasingly being used in hydrologic ensemble simulations. The reasons are it intrinsically weighs errors by assigning a lower weight to the largest residuals (Schneider et al., 2020) [SG:]thus accounting for observations that in other cases are defined as outliers. Therefore, it is the preferable measure in determining the error in forecasting models since it is more robust to outliers producing a more representative result. Note that a value of the CRPS of zero is only possible in case of a perfect deterministic forecast. Thus, the lower the value of the CPRS the better the model performance.

This is calculated over the entire simulation period and the average CRPS is defined as follows:

$$\overline{CRPS} = \frac{1}{L}\sum_{k=1}^{L} CRPS_k\,, \tag{21}$$

where L is the number of observations. This permits an analysis of the measures and a comparison of the RMSE and the CRPS. In this experiment, the models are naturally stochastic processes allowing the CPRS to produce a more representative value and robust measure for the errors compared to the RMSE. In conclusion of this section, all three measures will be used to assess the performance of the filter.

3 Results and Discussion

3.1 Deterministic Runs

During the calibration with the PSO, the dynamics of the parameter optimization algorithm was monitored, showing that the MODFLOW saturated hydraulic conductivity (K_h) had a consistent tendency towards high values (100 m·d^{-1} or higher) in order to minimise Equation (4). This was interpreted as an effect of ET component on the objective function SG:~~ET component~~, which was inducing the UZMs to transpire water directly from the WT to compensate for the low ET values. The boundary conditions for the groundwater model were thus modified by imposing a constant head boundary with shallower WT depth, which maintained K_h at a plausible order of magnitude. SG:~~Conceptually, these boundary conditions represent the water supplied from the regional aquifer to the plantation, and induce the WT depression shown on the right hand side of Figure~~

SG:~~The calibration technique proposed in (Section) was able to reproduce the link between the WT dynamics and AET for both configurations~~With the calibration technique proposed in Section 2.3, the coupled models were able to simultaneously reproduce the dynamics of both the WT and ET for the two configurations (See Figure 4). Configuration-1 performs better overall in the representation of the WT dynamics with a RMSE of 0.23 m, while the RMSE of Configuration-2 is slightly larger being 0.36 m. Configuration-1 also shows a higher correlation coefficient (0.790 vs 0.400) for the WT. Configuration-1 shows a lower temporal variability than Configuration-2, but the latter better matches the temporal evolution of the WT. There is a time lag between groundwater observations and model WT fluctuation for Configuration-2, which also explains the higher RMSE and lower correlation. This lag may be induced by preferential flow that the Richards equation does not account for, or to a slower response of the WT to the meteorological input that is discussed later in this section.

The SG:~~capillary fringe and~~ soil heterogeneity is represented differently by the two configurations. The physically-based Configuration-2, SG:~~the detail of the capillary fringe is represented in Figure [d] by the blurred area above the saturated zone (i.e. dark blue). Configuration-2 is also able to represent~~ can represent the heterogeneity of the soil column, as shown in Figure 5 [d] where a sharp variation of the SM content at 1.7 m depth is caused by the different soil parameters. Configuration-1 has no ability to SG:~~represent the capillary fringe effect, and it~~

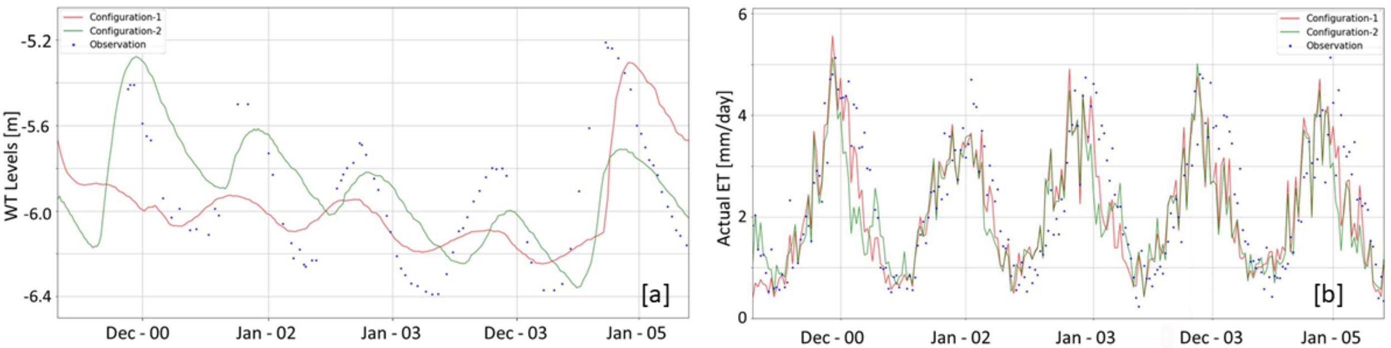

Figure 4. Observed and modelled [a] WT fluctuations and [b] AET after calibration.

Table 2. Results for the calibrated runs.

Variable	Configuration	RMSE	*r*
WT	1	0.230 [m]	0.790
Levels	2	0.360 [m]	0.400
SM	1	0.049 [$\mathrm{mm}^3\ \mathrm{mm}^{-3}$]	0.410
Upper	2	0.045 [$\mathrm{mm}^3\ \mathrm{mm}^{-3}$]	0.610
SM	1	0.085 [$\mathrm{mm}^3\ \mathrm{mm}^{-3}$]	0.592
Lower	2	0.018 [$\mathrm{mm}^3\ \mathrm{mm}^{-3}$]	0.850
AET	1	0.791 [mm day^{-1}]	0.811
	2	0.870 [mm day^{-1}]	0.788

~~does not~~ explicitly account for duplex soil[SG:], thus the soil profile does not show sharp variations of the water content. However, it can account for a decay of the hydraulic conductivity along the soil column. Because of these reasons, the modeled SM from Configuration-2 shows a good agreement with the observations, especially in the lower soil (Figure 5 [f]). Configuration-1 has a low SM RMSE (0.049 $\mathrm{mm}^3 \cdot \mathrm{mm}^{-3}$) and a reasonable agreement in terms of the Pearson correlation coefficient *r* (0.410) for the upper soil [b], but the resulting SM is consistently below the observed values in the bottom soil (panel [c]), with an RMSE of 0.137 $\mathrm{mm}^3 \cdot \mathrm{mm}^{-3}$. Both configurations report a higher correlation for the lower soil.

For AET, Configuration-1 yields good results with a lower RMSE and similar correlation when compared to Configuration-2. In particular, the physically-based Configuration-2 underestimates the simulated ET for the Southern hemisphere late summer/early autumn as shown in Figure 4 [b]. In this period, the soil water content is low [SG:]~~, as shown in~~ (Figure 5 [d]), and the [SG:]~~system is actively transpiring~~ the roots take up water directly from [SG:]~~the~~ groundwater. This can be interpreted as an effect of the coupling to the groundwater model. The conceptually based Configuration-1, with a rooting depth of 8.0 m, is able to extract water directly from the water table and immediately transforms it into AET. Configuration-2, with a rooting depth of 2.9 m, achieves this by reducing the pressure head along the soil column. Thus water has to flow across a part of [SG:]the unsaturated

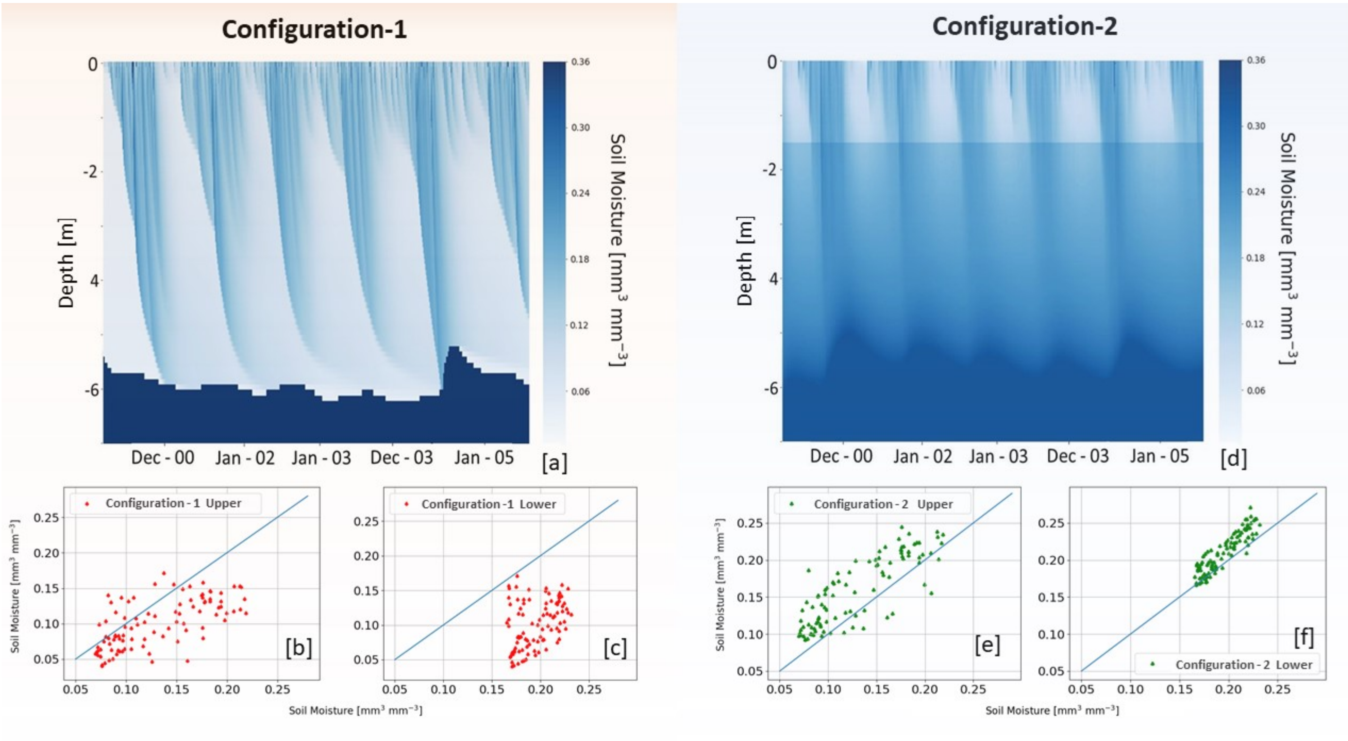

Figure 5. Temporal evolution of the SM contents and WT levels. Panels [a] and [d] show the entire modeled column, including the fluctuation of the WT (i.e. the dark blue area). Panel [b] and [e] represent the modeled and observed water content for the upper soil (averaged over 0-300 mm depth). Panel [c] and [f] show these results for the lower soil (averaged over the interval 1500-1800 mm depth).

zone before becoming available for direct plant transpiration, reducing the rapid response of the model to the forcing inputs. This also explains the lag in the WT dynamics previously described. Another possible reason for the underestimation of AET are the two oxygen stress parameters that reduce transpiration in conditions close to saturation (Table 1). These parameters are calibrated and kept constant during the simulation period. Configuration-2 has shown to be highly sensitive to these parameters, while Configuration-1 does not include this process.

3.2 Ensemble simulations

The generation of the ensemble is *SG:* ~~also~~ found to be a key step of the method. The simple perturbation of forcing inputs was not able to generate a sufficiently broad ensemble spread, particularly for Configuration-2. For both configurations, the combined perturbation of parameters and forcing inputs was found to produce *SG:* ~~adequate~~more accurate ensembles*SG:* ~~. This is obtained by applying the ensemble validation , as discussed in~~, in accordance with the ensemble validation skills calculated on the first year of the data set, excluding the *SG:* ~~first~~ 10 *SG:* first time steps to avoid the influence of the initial conditions*SG:* ~~(i.e.~~; the validation is thus applied from the 10th to the 45th time step. For the meteorological data, the best ensembles are obtained by perturbing the input with a random number sampled from a Gaussian distribution having a standard deviation proportional to the value of the forcing inputs (i.e. 50% for Configuration-1 and 10% for Configuration-2). For [SG:] ~~the~~ parameters, [SG:] the last column of Table 1 lists the [SG:] ~~perturbation fractions~~coefficient of variation. Additionally, for Configuration-2, S_y has a lower limit of 0.1 to preserve numerical stability of the coupled models.

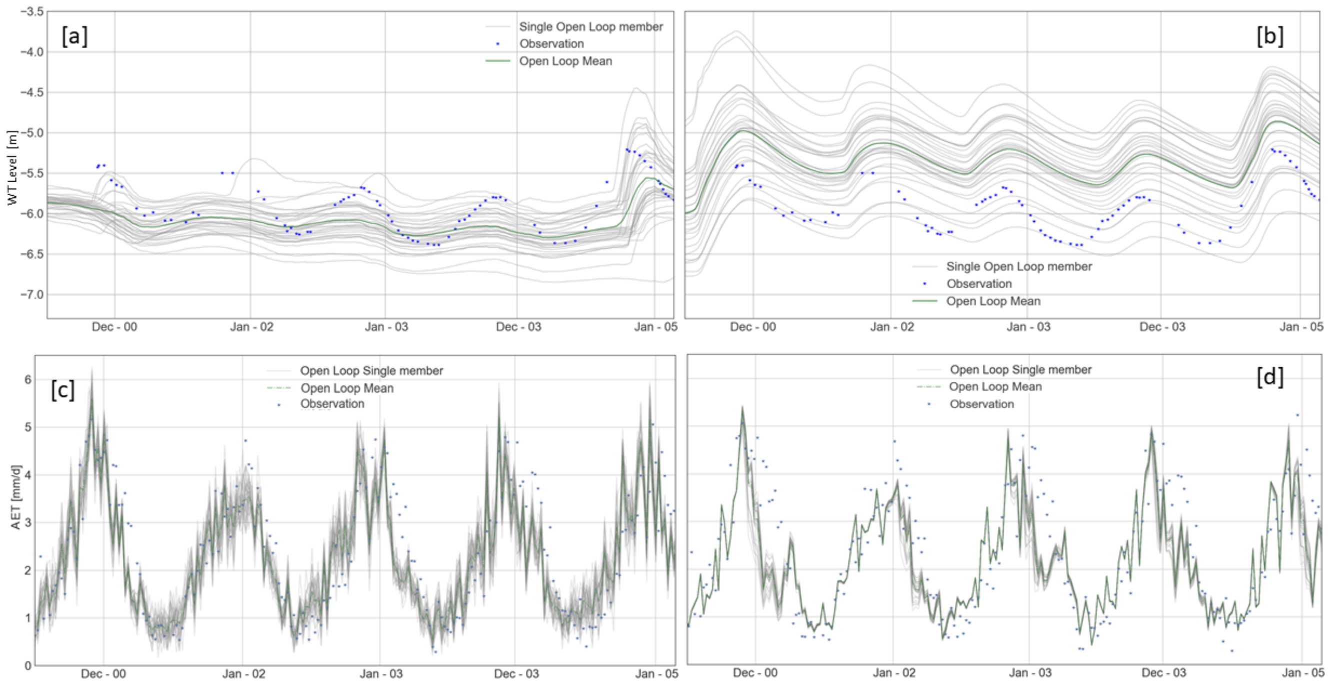

Figure 6. WT levels and AET and spread of the open loop ensembles for Configuration-1 [a,c] and Configuration-2[b,d]

In the case of the conceptual Configuration-1, the WT level spread of the open loop ensemble is consistently covering the observations (Figure 6[a]). The mean of the ensemble is close to the observations, but does not follow the seasonal variability appropriately. The associated spread of the AET for Configuration-1 is wider than that of Configuration-2. More specifically, the latter is narrow during wet periods (i.e. April to November) and becomes wider for the dry period (Figures 6[c] and 6[d]). A similar effect, with a larger magnitude, was reported during the ensemble generation phase and led to the [SG:] ~~double perturbation of~~perturbation of both the meteorological inputs and the parameters as explained in Section 2.4.1.

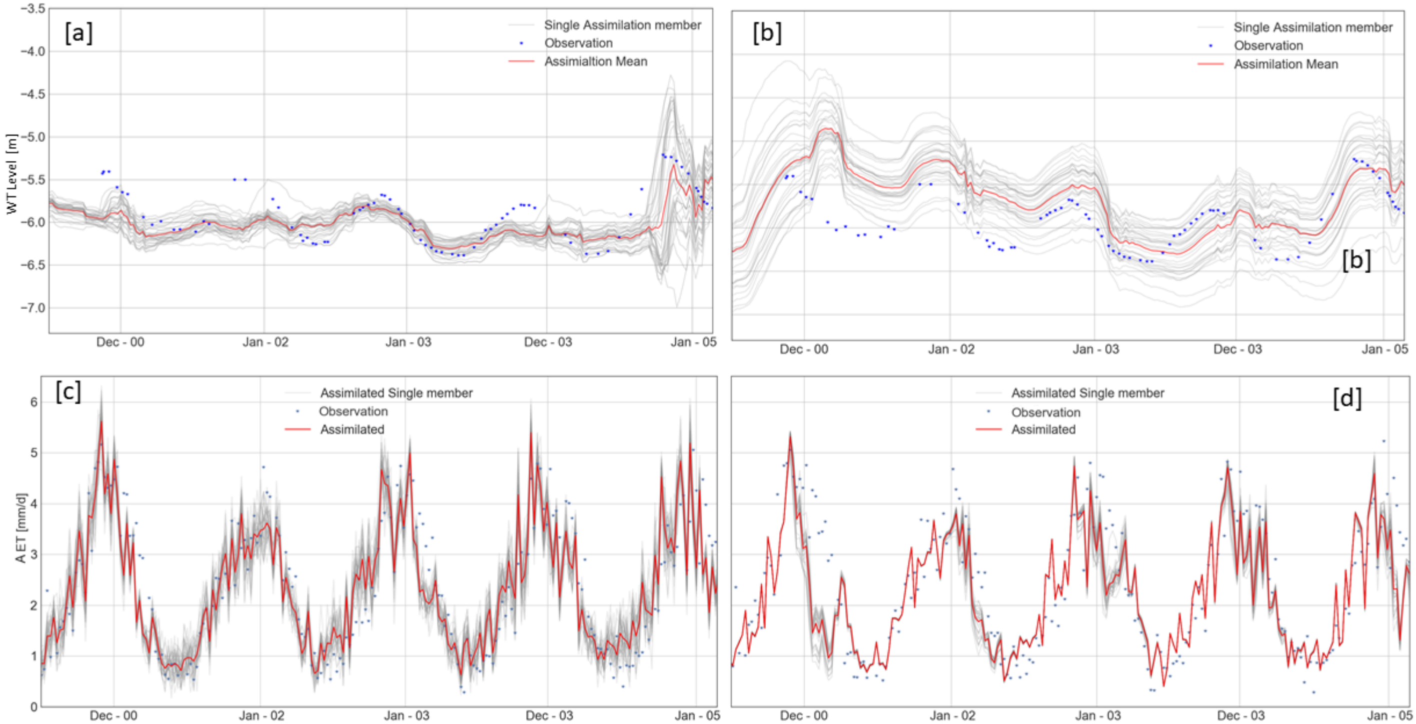

Figure 7. WT levels and AET and spread of the assimilation run for Configuration-1 [a,c] and Configuration-2 [b,d]

The spread of the WT levels for Configuration-2 (see 6[b]) covers the WT observations for most of the simulations and is wider than for Configuration-1. The mean represents the amplitude of the seasonal fluctuations better as compared to Configuration-1, but leads to a shallower WT as a result of the perturbation of the forcing inputs.

Table 3. RMSE, correlation $^{SG:}$and $\overline{CRPS}$ for three variables between the open loop and the assimilation.

Config	Type	AET			WT Levels			SM Upper Soil			SM Lower Soil		
		RMSE	r	$\overline{CRPS}$	RMSE	r	$\overline{CRPS}$	RMSE	r	$\overline{CRPS}$	RMSE	r	$\overline{CRPS}$
1	Open loop	0.760	0.820	0.541	0.280	0.730	0.161	0.045	0.497	0.204	0.102	0.468	0.078
	Assimilation	0.730	0.830	0.508	0.236	0.734	0.134	0.044	0.498	0.206	0.098	0.428	0.077
2	Open loop	0.830	0.810	0.624	0.626	0.880	0.426	0.041	0.888	0.037	0.019	0.940	0.013
	Assimilation	0.810	0.820	0.597	0.307	0.675	0.229	0.042	0.864	0.036	0.015	0.900	0.012

Table 3 summarizes [SG:]~~and compares the results between the open loop and the assimilation for AET, WT levels, and SM contents of the upper and lower soil layers.~~ the RMSE, r and $\overline{CRPS}$ values for AET, WT levels, and SM contents (upper and lower soil layers), and compares the results of the assimilation run to the open-loop. The table allows for a quick comparison between the results given by the RMSE, standardly used and understood across the modeling community, and the CRPS averaged over the simulation period (i.e. $\overline{CRPS}$), which allows for an appropriate and representative analysis of the ensemble distributions. For both configurations, the AET assimilation slightly decreases the RMSE, the $\overline{CRPS}$ and improves the correlations. In particular, the RMSE of AET for Configuration-1 reduces from 0.76 mm$\cdot$day^{-1} for the open loop to 0.73 mm$\cdot$ day^{-1}. The RMSE of AET for Configuration-2 reduces from 0.83 mm$\cdot$day^{-1} for the open loop to 0.81 mm day^{-1}. [SG:]A similar patter is observed for the $\overline{CRPS}$, with the relative percentage change improvements varying from 4.3% to 6.1%. In this case, the $\overline{CRPS}$ reinforces the relevance of the RMSE results. The correlation also improves[SG:], although marginally, for both configurations (i.e. + 0.01). However, these are non-trivial results as the data assimilation, through the EnKF, is designed to [SG:]only improve the model states. [SG:]~~Therefore, the reduction of the AET errors suggests that the improved state variables are contributing to a better modeling of other hydrological quantities.~~Therefore, the observed reduction in AET errors suggests that the model states (i.e. WT, SM) updated by the filter are contributing to better modeling of other hydrological quantities (e.g. AET).

[SG:]In particular there are instances in Configuration-2 [SG:]where the assimilation is not able to improve AET in the [SG:]~~Summer~~first quarter of 2001 and [SG:], to a lesser extent, the beginning of 2003. This causes poorer WT simulations performances during these periods, [SG:] as seen in Figure 7[b], [SG:]and highlighted by the higher values in Figure 8[b]. Here, the filter is trying to increase the amount of water in the system to match the higher assimilated observation, which is a correct application of the methodology. Thus, the WT is made shallower by the filter but this does not reflect in a higher modeled AET. The reason for this is the behaviour of the SWAP vegetation parameter oxygen stress. The filter is increasing the pressure head of the system, in an attempt to provide more water to transpire, but the actual transpiration from the plant is hindered by SWAP, which recognises the soil to be too saturated for the plant to transpire. The EnKF then causes the WT to rise, and increases the amount of recharge entering the groundwater. When the observed AET is lower than the simulations, the filter reduces the pressure head and the model allows the plant to transpire. Therefore, in the two time steps after this effect, the modeled AET is higher than the observation, after which this phenomenon disappears. This artefact is not seen in Configuration-1 as the oxygen stress is not accounted for.

In Figure 8, [SG:]the single bars represent the $CRPS_k$ computed for each time the observations (WT levels above, AET below) are available over the entire simulation period. For this reason the length of these datasets is different, as the CMRSET observations temporal interval is every eight days while the WT levels observations are more infrequent and present gaps. Generally, lower CRPS values are seen in Configuration-1 for both WT levels and AET. CRPS values for the WT level are substantially reduced in the first part of the simulation for both configurations, with Configuration-1 performing particularly well between August 2002 through July 2003 and in reducing the prior errors around the end of 2004. Analyzing the CRPS for AET (See Figure 8[c])[SG:], in most cases the assimilation improves the CRPS values, with the exemption of the central part

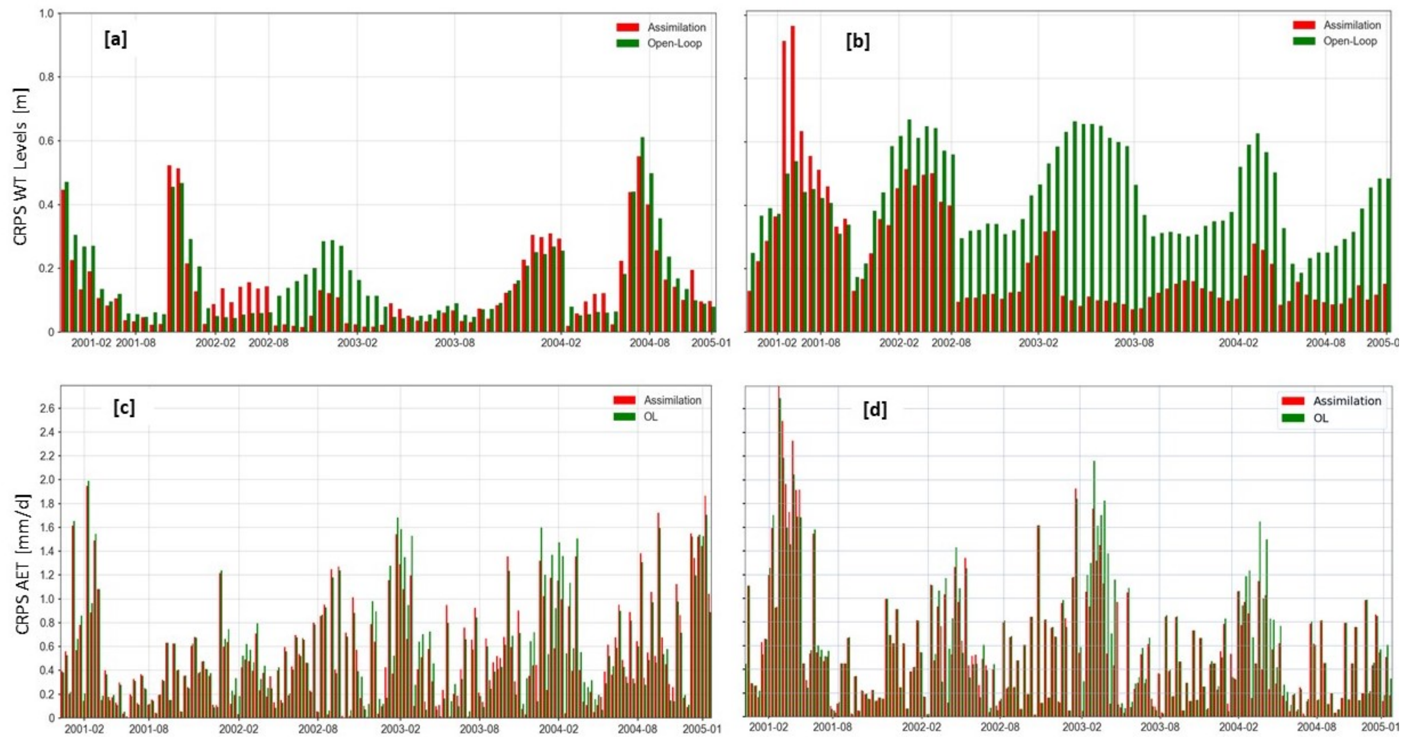

Figure 8. CRPS values WT levels and AET for Configuration-1 [a,c] and Configuration-2 [b,d]

of 2004. This is also seen in Figure 7 [SG:] when the assimilation fails to reduce the AET in winter of 2004. For Configuration-2, the WT level CRPS starts with higher values, and apart from the spikes of 2001 (discussed earlier), it presents continuous, constant improvements, outperforming Configuration-1 in the last part of the simulation. This pattern is similarly observed for the CRPS calculated for AET.

[SG:] ~~Figures show the observations, the mean of the open loop (blue dash-dotted line) and the mean of the assimilation runs (red dot line), for Configuration-1 and Configuration-2, respectively.~~ For both configurations, the assimilation improves the RMSE and $\overline{CRPS}$ when compared to the open loop runs. The best results are obtained for Configuration-1, showing a RMSE of 0.236 m with a 15% error reduction compared to the open loop [SG:] ; this result is endorsed by the $\overline{CRPS}$ of 0.134 (error reduction of 16%). Configuration-2 resulted in a substantial RMSE reduction of 38.9% as compared to the open loop.[SG:] The magnitude of these improvements is corroborated by the $\overline{CRPS}$, with a value after the assimilation of 0.229 from the prior of 0.426, which translates in an improvement of around 46%. However, RMSE [SG:] and $\overline{CRPS}$ values (0.307 and 0.229 m, respectively) are still higher in Configuration-1 than in Configuration-2. Apart from the oxygen stress artefacts explained above, the assimilation run of Configuration-2 is consistently better than the open loop. This is not always the case for Configuration-1, where the open loop was already performing well. The correlation remains largely unchanged for Configuration-1, and reduces for Configuration-2 mainly due to the updates during the SG: ~~summers of 2000/2001 and 2002/2003~~first quarter of 2001 and beginning of 2003.

SG: ~~Figures 8 and 9 show the observations, the mean of the open loop (blue dash-dotted line) and the mean of the assimilation runs (red dot line), for Configuration-1 and Configuration-2, respectively. The two violin plots shown in the insets to Figure 8 and 9 provide a visual representation of the magnitude of uncertainty before (prior) and after (posterior) the assimilation. In general, the spread of the WT levels for Configuration-1 is narrower than the equivalent for Configuration-2. Even when the mean of the open loop is closer to the observation, as in the first violin plot of figure 8, the assimilation helps in reducing the uncertainty around the WT levels. The second violin plot shows an ideal situation, where the assimilation mean is very close to the observed value and the uncertainty interval is narrow. This combination was not obtained for Configuration-2. As shown in the violin plots of figure 9, the posterior covariances (i.e. the red violin plot) are still large after the assimilation. This means a lower uncertainty reduction compared to Configuration-1.~~

SG: ~~Figure presents the scatter plots of the SM in the top (at a depth of 300 mm) and bottom (1800 mm) parts of the soil for each configuration.~~ For SM, the results are reported in Table 3, divided into the upper and lower soil. The open loop of Configuration-1 presents a RMSE of 0.045 $mm^3 \cdot mm^{-3}$ for the upper soil and 0.102 $mm^3 \cdot mm^{-3}$ for the lower soil. In the latter, the simulated water contents are consistently lower than the observations. This is mainly due to the model's inability to represent capillary rise. The assimilation only marginally improved the SM content, with slightly better results for the bottom part of the soil, where the RMSE was reduced to 0.098 $mm^3 \cdot mm^{-3}$. The open loop of Configuration-2 has a lower RMSE, 0.041 and 0.017 $mm^3 \cdot mm^{-3}$ for the top and bottom part of the soil, respectively. However, it is slightly overestimating the SM content for the entire column. This is consistent with the shallower WT (i.e. more water in the system) observed for the WT levels in the open loop (Figure 6[d]). The assimilation did not improve the top layer SM content, with an RMSE of 0.042 $mm^3 \cdot mm^{-3}$. However, SG: ~~the assimilation improved the SM content of the bottom part~~effects of the assimilation are seen for the SM content of the lower soil, SG: ~~(Figures [g] and [h])~~ for which the best results are obtained (i.e. 0.015). SG: For SM, the $\overline{CRPS}$ is unable to show significant variations and does not add valuable information. Although the improvements for SM are limited and cannot be considered conclusive, the feasibility for this framework of updating the entire soil column is a positive result of the assimilation of AET rates, as opposed to the assimilation of remotely sensed SM values. The latter usually results in stronger updates in the upper parts of the soil, because of the reduced correlation between the SM contents in the upper and deeper parts of the soil column SG: (Pipunic et al., 2014).

Generally, these results consolidate the synthetic approach in Gelsinari et al. (2020), and further confirm that the assimilation framework is not only able to update and improve the WT level, which is a prognostic variable of the coupled models, but also the modeled AET, and consequently the recharge to the WT. In addition, albeit marginally, the filter improves the unsaturated zone state variables regardless of the manner in which the SM content is calculated (volumetric SM or pressure head).

4 Conclusions

This study validates the assimilation of the satellite-based actual evapotranspiration (AET) data set (CMRSET) into two SG:~~coupled~~ unsaturated zone SG:models coupled to MODFLOW SG:~~groundwater configurations~~forming two configurations, one using a conceptual water balance model (UnSAT) and the other using a physically-based agro-hydrological model (SWAP). SG:~~Specifically, These configurations are composed by a conceptual water balance model (UnSAT) and a physically-based agro-hydrological model (SWAP), respectively, coupled to MODFLOW and~~These configurations are applied to a semi-arid pine plantation in the south-east of South Australia SG:where the WT is within reach of the trees' root system.

The most important findings can be summarized as:

Calibration. This study shows the need to calibrate the model using a multi-objective function, with normalised components of water table (WT) and AET. In this way, both configurations are representing the WT-AET SG:~~relationship in an appropriate manner and benefit~~dynamics, and are thus able to benefit from the assimilation of AET observations.

Configuration-1. The assimilation of AET values through the Ensemble Kalman Filter (EnKF) using a conceptual unsaturated zone model produced the best results for the prognostic variable WT levels and the diagnostic fluxes of AET. SM values were SG:~~also slightly improved~~updated in both the upper and lower parts of the soil columnSG:, although only to a minor extent. SG:~~]However~~In addition, because of the model conceptualization, the mismatch in the lower part of the soil is considerably larger than for Configuration-2. The reduced number of parameters of this configuration allows for a simpler calibration, which is able to represent the WT dynamics. Similarly, the generation of an appropriate ensemble is more straightforward, mostly due to the model conceptualization, which allows the WT to respond quickly to direct root water extraction by transpiration.

Configuration-2. The AET assimilation into a physically-based unsaturated zone model, based on the Richards equation, produced the largest improvements to the WT levels with SG:~~a larger uncertainty reduction and an adequate representation of the capillary fringe~~a better representation of the soil heterogeneity. Improvements to AET fluxes were similar to Configuration-1. For SM, SG:~~generally~~ the impact of the assimilation algorithm was small, with a positive update for the lower soil layers, and a negative update for the upper layers. Here, the calibration involved a larger number of parameters and produced a good representation of the SM dynamics. However, due to the non-linearity introduced with the coupling SG:~~(e.g. capillary fringe)~~, errors in the WT levels and AET fluxes are higher. In addition, the ensemble generation is constrained by the high model parameterization, making it more difficult to produce an appropriate ensemble that preserves the ET-WT relationship.

AET information. The updating of the entire soil column is an advantage of the assimilation of remotely sensed AET over satellite SM retrievals. AET rates express the moisture status of the entire root zone. Thus, assimilating AET SG:has the potential to overcome the SM assimilation tendency to produce stronger updates in the most superficial part of the soil because of the reduced correlation between the upper and lower SM contents. SG:This experiment only showed the feasibility for the proposed assimilation framework to improve SM contents. Preliminary results indicated that Configuration-2 is preferred to conduct more experiments in order to quantify the significance of the SM updates.

In conclusion, [SG:]~~it is possible to use either a conceptual or a physically-based unsaturated zone model in the assimilation of satellite-based AET estimates to inform hydrogeological models. Both model coupling configurations reduce the uncertainty related to state variables (such as WT and SM) and fluxes of AET.~~the experiment explored the added value of AET information for constraining unobservable estimates (i.e. net-recharge) calculated by hydrogeological models. Improving the AET fluxes led to better recharge estimates. Thus, as recharge is a key quantity driving the WT dynamics, the link between AET and WT in the model is strengthened. It was shown that it is possible to use either a conceptual or a physically-based unsaturated zone model in the assimilation of satellite-based AET estimates to inform hydrogeological models. [SG:]The assimilation results have been quantified using standard metrics, such as RMSE and r, and reinforced by calculating the $CRPS$, which is a specifically designed metric for ensemble simulations. The $CRPS$ is applied as it is a measure to determine a more representative error, since is more robust in accounting for uncertainty in stochastic models. The findings indicate that a simple conceptual model may be sufficient for this purpose, thus using one configuration over the other should be motivated by the specific purpose of the simulation and the information available.

[SG:]This study contributes to unlock the potential of using AET observations to inform hydrological models, with the aim of reducing the uncertainty in the outputs, and it [SG:]~~This study~~ represents a step towards the use of satellite-based AET retrievals for water resources management. For future applications at larger scales, more research is to be conducted in areas with different groundwater, vegetation and soil conditions, with the intent of prioritizing regions where the AET assimilation is more effective.

Data availability. Forcing input http://www.bom.gov.au/
Results: https://figshare.com/s/222b9874a7ff328a8a24
(Would become https://doi.org/10.26180/5ec4cda0b2612 upon publication)
CMRSET: https://doi.org/10.25919/5c36d76d2ac2f/

Author contributions. SG performed the modelling work and wrote the manuscript. VP supervised the implementation of the EnK, ED supervised the implementation of the UnSAT model, JvD supervised the implementation and coupling of the SWAP model, [SG:]NFY contributed to the results analysis, and RD supervised the entire project. All co-authors also provided input to the writing of the manuscript.

Competing interests. No competing interests were found

Acknowledgements. Funding for this project has been provided by the CSIRO Land & Water Effective Floodplain Management Project. SG acknowledges the financial support by the Faculty of Engineering at Monash University through the Graduate Research International Travel Award and thanks the chair group of Hydrology and Quantitative Water Management at Wageningen University & Research for the support during his visit. SG also thanks Karina Gutierrez Jurado for her support and suggestions during the preparation of this manuscript

References

Estimation of Groundwater Evapotranspiration Using Diurnal Groundwater Level Fluctuations under Three Vegetation Covers at the Hinterland of the Badain Jaran Desert, Advances in Meteorology, 2020, https://doi.org/10.1155/2020/8478140, 2020.

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., and Fienen, M. N.: Scripting MODFLOW Model Development Using Python and FloPy, Groundwater, 54, 733–739, https://doi.org/10.1111/gwat.12413, http://doi.wiley.com/10.1111/gwat.12413, 2016.

Banks, E. W., Brunner, P., and Simmons, C. T.: Vegetation controls on variably saturated processes between surface water and groundwater and their impact on the state of connection, Water Resour. Res., 47, 1–14, https://doi.org/10.1029/2011WR010544, 2011.

Batelaan, O. and De Smedt, F.: GIS-based recharge estimation by coupling surface-subsurface water balances, J. Hydrol., 337, 337–355, https://doi.org/10.1016/j.jhydrol.2007.02.001, 2007.

Benyon, R. G. and Doody, T. M.: Water Use by Tree Plantations in South East South Australia. CSIRO Forestry and Forest Products, Tech. rep., CSIRO, http://www.ffp.csiro.au/http://www.dwlbc.sa.gov.au/http://www.secatchment.com.au/, 2004.

Benyon, R. G., Theiveyanathan, S., and Doody, T. M.: Impacts of tree plantations on groundwater in south-eastern Australia, Aust. J. Bot., 54, 181, https://doi.org/10.1071/BT05046, http://www.publish.csiro.au/?paper=BT05046, 2006.

Camporese, M., Paniconi, C., Putti, M., and Orlandini, S.: Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data, Water Resour. Res., 46, https://doi.org/10.1029/2008WR007536, 2010.

Carroll, R. W. H., Pohll, G. M., Morton, C. G., and Huntington, J. L.: Calibrating a Basin-Scale Groundwater Model to Remotely Sensed Estimates of Groundwater Evapotranspiration, J. Am. Water Resour. Assoc., https://doi.org/10.1111/jawr.12285, 2015.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, https://doi.org/10.1029/WR014i004p00601, 1978.

De Lannoy, G. J., Houser, P. R., Pauwels, V. R., and Verhoest, N. E.: Assessment of model uncertainty for soil moisture through ensemble verification, J. Geophys. Res. Atmos., 111, https://doi.org/10.1029/2005JD006367, 2006.

Doble, R. C., Pickett, T., Crosbie, R. S., Morgan, L. K., Turnadge, C., and Davies, P. J.: Emulation of recharge and evapotranspiration processes in shallow groundwater systems, J. Hydrol., 555, 894–908, https://doi.org/10.1016/j.jhydrol.2017.10.065, https://doi.org/10.1016/j.jhydrol.2017.10.065https://linkinghub.elsevier.com/retrieve/pii/S0022169417307424, 2017.

Donohue, R. J., Roderick, M. L., and McVicar, T. R.: On the importance of including vegetation dynamics in Budyko's hydrological model, Hydrol. Earth Syst. Sci., 11, 983–995, https://doi.org/10.5194/hess-11-983-2007, 2007.

Donohue, R. J., McVicar, T. R., and Roderick, M. L.: Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate, J. Hydrol., 386, 186–197, https://doi.org/10.1016/j.jhydrol.2010.03.020, http://linkinghub.elsevier.com/retrieve/pii/S0022169410001460, 2010.

Droogers, P., Immerzeel, W. W., and Lorite, I. J.: Estimating actual irrigation application by remotely sensed evapotranspiration observations, Agric. Water Manag., 97, 1351–1359, https://doi.org/10.1016/j.agwat.2010.03.017, 2010.

Entekhabi, D. and Moghaddam, M.: Mapping recharge from space: Roadmap to meeting the grand challenge, Hydrogeol. J., https://doi.org/10.1007/s10040-006-0120-6, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10 143, https://doi.org/10.1029/94JC00572, 1994.

Facchi, A., Ortuani, B., Maggi, D., and Gandolfi, C.: Coupled SVAT–groundwater model for water resources simulation in irrigated alluvial plains, Environ. Model. Softw., 19, 1053–1063, https://doi.org/10.1016/j.envsoft.2003.11.008, http://linkinghub.elsevier.com/retrieve/pii/S1364815203002603https://linkinghub.elsevier.com/retrieve/pii/S1364815203002603, 2004.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, J. Hydrol., pp. 45–60, https://doi.org/10.1016/j.jhydrol.2016.03.026, 2016.

Feddes, R. A., Kowalik, P., Kolinska-Malinka, K., and Zaradny, H.: Simulation of field water uptake by plants using a soil water dependent root extraction function, J. Hydrol., 31, 13–26, https://doi.org/10.1016/0022-1694(76)90017-2, 1976.

Gelsinari, S., Doble, R., Daly, E., and Pauwels, V. R.: Feasibility of improving groundwater modeling by assimilating evapotranspiration rates., Water Resour. Res., https://doi.org/10.1029/2019WR025983, https://onlinelibrary.wiley.com/doi/abs/10.1029/2019WR025983, 2020.

Grimaldi, S., Orellana, F., and Daly, E.: Modelling the effects of soil type and root distribution on shallow groundwater resources, Hydrol. Process., 29, 4457–4469, https://doi.org/10.1002/hyp.10503, 2015.

Grinevskii, S. O.: Modeling root water uptake when calculating unsaturated flow in the vadose zone and groundwater recharge, Moscow Univ. Geol. Bull., 66, 189–201, https://doi.org/10.3103/s0145875211030057, 2011.

Guerschman, J. P., Van Dijk, A. I., Mattersdorf, G., Beringer, J., Hutley, L. B., Leuning, R., Pipunic, R. C., and Sherman, B. S.: Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia, J. Hydrol., 369, 107–119, https://doi.org/10.1016/j.jhydrol.2009.02.013, https://linkinghub.elsevier.com/retrieve/pii/S002216940900105X, 2009.

Harbaugh, A. W.: MODFLOW-2005 , The U . S . Geological Survey Modular Ground-Water Model — the Ground-Water Flow Process, U.S. Geol. Surv. Tech. Methods, p. 253, https://doi.org/U.S. Geological Survey Techniques and Methods 6-A16, 2005.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Jackson, R. B., Jobbágy, E. G., and Nosetto, M. D.: Ecohydrology in a human-dominated landscape, Ecohydrology, 2, 383–389, https://doi.org/10.1002/eco.81, http://www3.interscience.wiley.com/journal/122653919/abstracthttp://doi.wiley.com/10.1002/eco.81, 2009.

Jones, J. E. and Woodward, C. S.: Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems, Adv. Water Resour., 24, 763–774, https://doi.org/10.1016/S0309-1708(00)00075-0, 2001.

Kennedy, J. and Eberhart, R.: Particle Swarm Optimization, Proc. 1995 IEEE Int. Conf. Neural Networks. Part 1 (of 6), 4, 1942–1948, https://doi.org/10.1007/s11721-007-0002-0, http://www.scopus.com/inward/record.url?eid=2-s2.0-0029535737{&}partnerID=40{&}md5=e6bf04ae50f3268ae545d88ed91d1fc5, 1995.

Khatami, S., Peel, M. C., Peterson, T. J., and Western, A. W.: Equifinality and Flux Mapping: A New Approach to Model Evaluation and Process Representation Under Uncertainty, Water Resour. Res., pp. 1–20, https://doi.org/10.1029/2018WR023750, 2019.

Kim, C. P. and Stricker, J. N.: Influence of spatially variable soil hydraulic properties and rainfall intensity on the water budget, Water Resour. Res., 32, 1699–1712, https://doi.org/10.1029/96WR00603, 1996.

Kroes, J., van Dam, J., Bartholomeus, R., Groenendijk, P., Heinen, M., Hendriks, R., Mulder, H., Supit, I., and van Walsum, P.: SWAP version 4, Tech. rep., Wageningen University & Recharge, https://doi.org/10.18174/416321, http://library.wur.nl/WebQuery/wurpubs/522980, 2017.

Krysanova, V. and Arnold, J. G.: Advances in ecohydrological modelling with SWAT - A review, Hydrol. Sci. J., 53, 939–947, https://doi.org/10.1623/hysj.53.5.939, 2008.

Laio, F., Porporato, A., Fernandez-Illescas, C. P., and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: Active role in hydrologic processes and responce to water stress IV. Discussion of real cases, Adv. Water Resour., 24, 745–762, https://doi.org/10.1016/S0309-1708(01)00007-0, 2001.

Li, Y., Grimaldi, S., Walker, J. P., and Pauwels, V. R.: Application of remote sensing data to constrain operational rainfall-driven flood forecasting: A review, https://doi.org/10.3390/rs8060456, 2016.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., Van Dijk, A. I. J. M., Van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863–3887, https://doi.org/10.5194/hess-16-3863-2012, 2012.

Long, D., Longuevergne, L., and B. R. Scanlon: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites., Water Resour. Res., 50, 1131–1151, https://doi.org/10.1002/2013WR014581, 2014.

Lu, Y., Steele-Dunne, S. C., and De Lannoy, G. J.: Improving soil moisture and surface turbulent heat flux estimates by assimilation of SMAP brightness temperatures or soil moisture retrievals and GOES land surface temperature retrievals, J. Hydrometeorol., 21, 183–203, https://doi.org/10.1175/JHM-D-19-0130.1, 2020.

Mensforth, L. J., Thorburn, P. J., Tyerman, S. D., and Walker, G. R.: Sources of water used by riparian Eucalyptus camaldulensis overlying highly saline groundwater, Oecologia, 100, 21–28, https://doi.org/10.1007/BF00317126, 1994.

Mitchell, H. L., Houtekamer, P. L., and Pellerin, G.: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter*, Mon. Weather Rev., 130, 2791–2808, https://doi.org/10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2, 2002.

Neumann, R. B. and Cardon, Z. G.: The magnitude of hydraulic redistribution by plant roots: a review and synthesis of empirical and modeling studies., New Phytol., 194, 337–52, https://doi.org/10.1111/j.1469-8137.2012.04088.x, http://doi.wiley.com/10.1111/j.1469-8137.2010.03195.xhttp://www.ncbi.nlm.nih.gov/pubmed/22417121, 2012.

Orellana, F., Verma, P., Loheide, S. P., and Daly, E.: Monitoring and modeling water-vegetation interactions in groundwater-dependent ecosystems, Rev. Geophys., 50, https://doi.org/10.1029/2011RG000383, 2012.

Pauwels, V. R. N. and De Lannoy, G. J. M.: Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, Water Resour. Res., https://doi.org/10.1029/2008WR007590, 2009.

Pauwels, V. R. N., De Lannoy, G. J. M., Hendricks Franssen, H.-J., and Vereecken, H.: Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter, Hydrol. Earth Syst. Sci., 17, 3499–3521, https://doi.org/10.5194/hess-17-3499-2013, https://www.hydrol-earth-syst-sci.net/17/3499/2013/, 2013.

Pipunic, R. C., Ryu, D., and Walker, J. P.: Assessing Near-Surface Soil Moisture Assimilation Impacts on Modeled Root-Zone Moisture for an Australian Agricultural Landscape, Remote Sens. Terr. Water Cycle, 9781118872, 305–317, https://doi.org/10.1002/9781118872086.ch18, 2014.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, 1–22, https://doi.org/10.1029/2009WR008328, 2010.

Scheerlinck, K., Pauwels, V. R. N., Vernieuwe, H., and De Baets, B.: Calibration of a water and energy balance model: Recursive parameter estimation versus particle swarm optimization, Water Resour. Res., 45, https://doi.org/10.1029/2009WR008051, 2009.

Schneider, R., Henriksen, H. J., and Stisen, S.: A robust objective function for calibration of groundwater models in light of deficiencies of model structure and observations, Hydrol. Earth Syst. Sci. Discuss., pp. 1–26, https://doi.org/10.5194/hess-2019-685, 2020.

Shah, N., Nachabe, M., and Ross, M.: Extinction Depth and Evapotranspiration from Ground Water under Selected Land Covers, Groundwater, 45, 329–338, https://doi.org/https://doi.org/10.1111/j.1745-6584.2007.00302.x, https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.2007.00302.x, 2007.

Shi, Y. and Eberhart, R.: A modified particle swarm optimizer, in: 1998 IEEE Int. Conf. Evol. Comput. Proceedings. IEEE World Congr. Comput. Intell. (Cat. No.98TH8360), pp. 69–73, IEEE, https://doi.org/10.1109/ICEC.1998.699146, http://ieeexplore.ieee.org/document/699146/, 1998.

Simmons, C. S. and Meyer, P. D.: A simplified model for the transient water budget of a shallow unsaturated zone, Water Resour. Res., 36, 2835–2844, https://doi.org/10.1029/2000WR900202, http://doi.wiley.com/10.1029/2000WR900202, 2000.

Simunek, J., Sejna, M., Saito, H., Sakai, M., and van Genuchten, M.: "The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 4.08. HYDRUS Softw. Ser. 3., Dep. Environ. Sci., Univ. Calif., Riverside., p. 332, 2009.

Swaffer, B. A., Habner, N. L., Holland, K. L., and Crosbie, R. S.: Applying satellite-derived evapotranspiration rates to estimate the impact of vegetation on regional groundwater flux, Ecohydrology, 13, 1–14, https://doi.org/10.1002/eco.2172, 2020.

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of Probabilistic Prediction Systems, Tech. rep., Meteo-France, Illkirch, France, 1997.

Teuling, A. J. and Troch, P. A.: Improved understanding of soil moisture variability dynamics, Geophysical Research Letters, 32, https://doi.org/10.1029/2004GL021935, 2005.

Therrien, R., McLaren, R., Sudicky, E., and Panday, S.: Hydrogeosphere—a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport., Tech. rep., Groundwater Simul. Group,, Waterloo, Ont., Canada., 2006.

van Dam, J. C., Groenendijk, P., Hendriks, R. F., and Kroes, J. G.: Advances of modeling water flow in variably saturated soils with SWAP, Vadose Zo. J., 7, 640, https://doi.org/10.2136/vzj2007.0060, 2008.

van Genuchten, M. T.: A closed-form equation for predicting the Hydraulic conductivity of unsaturated zone, Soil Sci. Soc. Am. J., 44, 892–898, 1980.

Van Walsum, P. E. V. and Veldhuizen, A. A.: Integration of models using shared state variables: Implementation in the regional hydrologic modelling system SIMGRO, J. Hydrol., 409, 363–370, https://doi.org/10.1016/j.jhydrol.2011.08.036, http://dx.doi.org/10.1016/j.jhydrol.2011.08.036, 2011.

Vrugt, J., Hopmans, J., and Simunek, J.: Calibration of a two-dimensional root water uptake model, Soil Sci. Soc. Am. J., 65, 1027–1037, https://doi.org/10.2136/sssaj2001.6541027x, https://dl.sciencesocieties.org/publications/sssaj/abstracts/65/4/1027, 2001a.

Vrugt, J. A., Van Wijk, M. T., Hopmans, J. W., and Šimunek, J.: One-, two-, and three-dimensional root water uptake functions for transient modeling, Water Resour. Res., 37, 2457–2470, https://doi.org/10.1029/2000WR000027, 2001b.

Xie, Y., Crosbie, R., Yang, J., Wu, J., and Wang, W.: Usefulness of Soil Moisture and Actual Evapotranspiration Data for Constraining Potential Groundwater Recharge in Semiarid Regions, Water Resources Research, 54, 4929–4945, https://doi.org/https://doi.org/10.1029/2018WR023257, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023257, 2018.

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

Zeng, J., Yang, J., Zha, Y., and Shi, L.: Capturing soil-water and groundwater interactions with an iterative feedback coupling scheme : new HYDRUS package for MODFLOW, Hydrol. Earth Syst. Sci., pp. 637–655, 2019.

Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., and Franssen, H. J. H.: Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model, Adv. Water Resour., 111, 224–238, https://doi.org/10.1016/j.advwatres.2017.11.003, https://doi.org/10.1016/j.advwatres.2017.11.003, 2018.

Zhu, Y., Zha, Y.-y., Tong, J.-x., and Yang, J.-z.: Method of coupling 1-D unsaturated flow with 3-D saturated flow on large scale, Water Sci. Eng., 4, 357–373, https://doi.org/10.3882/j.issn.1674-2370.2011.04.001, http://www.sciencedirect.com/science/article/pii/S167423701530168X, 2011.

Zhu, Y., Shi, L., Lin, L., Yang, J., and Ye, M.: A fully coupled numerical modeling for regional unsaturated-saturated water flow, J. Hydrol., 475, 188–203, https://doi.org/10.1016/j.jhydrol.2012.09.048, http://dx.doi.org/10.1016/j.jhydrol.2012.09.048, 2012.