Journal of Hydrology 637 (2024) 131411

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Remote sensing evapotranspiration in ensemble-based framework to enhance cascade routing and re-infiltration concept in integrated hydrological model applied to support decision making

Mostafa Gomaa Daoud [a,*], Jeremy T. White [b], Eric D. Morway [c], Christiaan van der Tol [a], Maciek W. Lubczynski [a,*]

[a] Department of Water Resources, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7522 NH Enschede, The Netherlands
[b] Intera, Boulder, CO, USA
[c] U.S. Geological Survey, Nevada Water Science Center, Carson City, NV, USA

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief, with the assistance of Yueqing Xie, Associate Editor

Keywords:
Surface-groundwater interactions
Remote sensing observations
Integrated hydrological modelling
Cascade routing and re-infiltration
Decision support modelling

ABSTRACT

Integrated hydrological models (IHMs) help characterize the complexity of surface–groundwater interactions. The cascade routing and re-infiltration (CRR) concept, recently applied to a MODFLOW 6 IHM, improved conceptualization and simulation of overland flow processes. The CRR controls the transfer of rejected infiltration and groundwater exfiltration from upslope areas to adjacent downslope areas where that water can be evaporated, re-infiltrated back to subsurface, or discharged to streams as direct runoff. The partitioning between these three components is controlled by uncertain parameters that must be estimated. Thus, by quantifying and reducing those uncertainties, next to uncertainties of the other model parameters (e.g. hydraulic and storage parameters), the reliability of the CRR is improved and the IHM is better suited for decision support modelling, the two key objectives of this work. To this end, the remotely sensed MODIS-ET product was incorporated into the calibration process for complementing traditional hydraulic head and streamflow observations. A total of approximately 150,000 observations guided the calibration of a 13-year MODFLOW 6 IHM simulation of the Sardon catchment (Spain) with daily stress periods. The model input uncertainty was represented by grid-cell-scale parameterization, yielding approximately 500,000 unknown input parameters to be conditioned. The calibration was carried out through an iterative ensemble smoother. Incorporating the MODIS-ET data improved the CRR implementation, and reduced uncertainties associated with other model parameters. Additionally, it significantly reduced the uncertainty associated with net recharge, a critical flux for water management that cannot be directly measured and rather is commonly estimated by IHM simulations.

Nomenclature

$\alpha_{i,j}$	Fraction of flow from the cell i to the neighboring j cell
ADAS	Automated data acquisition system
$\beta_{i,j}$	Part of $\alpha_{i,j}$ that is spatio-temporal parameter and can be calibrated
CRR	Cascade routing and re-infiltration
cv	Coefficient of variation representing spatial variability
ΔS	Change in total catchment storage
ΔS_g	Change in groundwater zone storage
ΔS_u	Change in unsaturated zone storage
d_{ext}	Extinction depth
DISV	Discretization by vertices package of MODFLOW 6

(*continued on next column*)

Nomenclature (*continued*)

DRN	Drain package of MODFLOW 6
DSM	Decision support modelling
d_{surf}	Surface depth at which groundwater exfiltration can start
elv	Land surface elevation
E_s	Surface evaporation
E_I	Canopy interception
ET	Evapotranspiration
ET_g	Groundwater evapotranspiration
ET_{ss}	Subsurface evapotranspiration
ET_u	Unsaturated zone evapotranspiration

(*continued on next page*)

* Corresponding authors.
E-mail addresses: m.g.m.daoud@utwente.nl (M.G. Daoud), lubczynskimaciek@gmail.com (M.W. Lubczynski).

https://doi.org/10.1016/j.jhydrol.2024.131411

Available online 25 May 2024

Nomenclature (*continued*)

Exf_{gw}	Groundwater exfiltration – the groundwater discharge to the land surface when the water table rises up or close to the land surface level
Exf_{gw}^{e}	Evaporated groundwater exfiltration
Exf_{gw}^{i}	Re-infiltrated groundwater exfiltration
Exf_{gw}^{s}	Groundwater exfiltration routed to nearest surface water body (e.g. streams or lakes)
h	Percentage of histogram intersection
I	Initial infiltration
I_a	Active infiltration – sum of I and RE^i
IES	Iterative ensemble smoother
IHM	Integrated hydrological model
K	Histogram of the observed MODIS-ET
K_b	Hydraulic conductivity of stream reach's bed
K_h	Horizontal hydraulic conductivity
K_{sat}	Saturated vertical hydraulic conductivity of the unsaturated zone
K_v	Vertical hydraulic conductivity
$l_{i,j}$	Distance between the centers of the connected i and j cells
L	Histogram of the simulated ET
MVR	Mover package of MODFLOW 6
P	Precipitation
P_e	Effective precipitation
PET	Potential evapotranspiration
PET_e	Effective potential evapotranspiration – PET minus E_I
q	Total stream outflow at the catchment outlet
q_B	Baseflow
q_g	Lateral groundwater outflow
r	Pearson correlation coefficient
R^2	Coefficient of determination
$RE^e = (RI^e + Exf_{gw}^e)$	Evaporated water originated from the sum of RI^e and Exf_{gw}^e
$RE^i = (RI^i + Exf_{gw}^i)$	Re-infiltrated water originated from the sum of RI^i and Exf_{gw}^i
$RE^s = (RI^s + Exf_{gw}^s)$	Direct runoff originated from the sum of RI^s and Exf_{gw}^s
RF	Random Forest
R_g	Gross groundwater recharge
RI	Rejected infiltration
RI^e	Evaporated rejected infiltration
RI^i	Re-infiltrated rejected infiltration
RI^s	Rejected infiltration routed to nearest surface water body (e.g. streams or lakes)
R_n	Net groundwater recharge
RS	Remote sensing
SFR	Streamflow routing package of MODFLOW 6
$S_{i,j}$	Slope gradient between cell i and j
SM	Soil moisture
S_s	Specific storage
std	Standard deviation
S_y	Specific yield
θ_{ext}	Extinction water content
θ_{resid}	Residual water content
θ_{sat}	Saturated water content
UZF	Unsaturated-zone flow package of MODFLOW 6
WB	Water balance

1. Introduction

Groundwater is a vital freshwater resource, used globally by humans and natural systems. In many settings, groundwater flow dynamics are strongly influenced by land surface processes, surface water bodies and wetlands through the unsaturated zone. Hereafter, these types of interactions are collectively referred to as "surface–groundwater interactions". As the effects of climate change are already manifested within watersheds, in particular the increased frequency of drought events, surface–groundwater interactions have become even more important (Condon et al., 2021; Camporese and Girotto, 2022). Many recent reviews document the challenges related to water quantity, quality and ecology due to surface–groundwater interactions (Brunner et al., 2017; Haque et al., 2021; Yang et al., 2021; Ntona et al., 2022). For example, high evapotranspiration rates can reduce groundwater recharge and contribute to water table declines, which in turn can harm groundwater-dependent ecosystems and impair water quality. Extensive pumping or prolonged droughts threaten the sustainability of groundwater resources for irrigation and domestic water supply (Gleeson et al., 2020) and reduce baseflow, leading to water supply shortages (Famiglietti et al., 2011). These examples emphasize the usefulness of improved simulations of surface–groundwater interactions (Condon et al., 2021) to inform water management. An improved understanding of surface–groundwater interaction processes equips water managers to better steward water resources for domestic, agricultural, and other public purposes (Haque et al., 2021). Studying the spatial and temporal dynamics of surface–groundwater interactions is often pursued with integrated hydrological models (IHMs).

IHMs are physically-based models, capable of dynamically simulating the interactions between the surface and sub-surface (including both the saturated and unsaturated zones) flow domains within a single model simulation. Even the 'simplified IHMs' (Daoud et al., 2022), that simplify the governing equations of one or more of the flow domains, are complex models. IHMs are still rarely used to support decision-making because they need a long time to run, are less numerically stable, and the process of calibration through history matching is more complicated compared to simple models. However, IHMs are capable of addressing more complicated system behaviors; for example: (a) quantifying the relationship between groundwater recharge and actual evapotranspiration of a certain modelled system (Doble and Crosbie, 2017), (b) accessing the potential of managed aquifer recharge for a given location (Martinsen et al., 2022), (c) discovering how to alter water tables and groundwater flow to enhance and sustain agricultural production (Morway et al., 2013), (d) investigating the link between land use and water resources management (Janus et al., 2023), or (e) exploring the potential effects of climate change on water rights and river basin operations (Kitlasten et al., 2021). Such predictions, that may support decision-making, can only be addressed with more complex, multi-domain models, such as IHMs. Additionally, the complexity included in IHMs helps in quantifying the predictive uncertainty associated with alternative management scenarios.

The IHM numerical conceptualization employed in this study uses the cascade routing and re-infiltration (CRR) concept, recently introduced for use with the MODFLOW 6 numerical code (Langevin et al., 2017) and described in Daoud et al. (2022). The CRR concept (Fig. 1) simulates the transfer of the available water [rejected infiltration (RI) plus groundwater exfiltration (Exf_{gw}), i.e. the groundwater discharge to the land surface when the water table rises up or close to the land surface level] from the upslope areas to the adjacent downslope areas. That water can subsequently be: (a) evaporated (RE^e); (b) re-infiltrated back to the subsurface (RE^i); or (c) discharged as direct runoff (RE^s) into the nearest surface water body (e.g. streams or lakes). RE^e is the sum of the evaporated rejected infiltration (RI^e) and evaporated groundwater exfiltration (Exf_{gw}^e). RE^i is the sum of re-infiltrated rejected infiltration (RI^i) and re-infiltrated groundwater exfiltration (Exf_{gw}^i). Lastly, RE^s is the sum of routed rejected infiltration (RI^s) to the nearest surface water body (e.g. streams or lakes) and routed groundwater exfiltration (Exf_{gw}^s) to the nearest surface water body (e.g. streams or lakes). The mathematical design of the CRR concept is explained in Fig. 1b. RE^i is controlled by the vertical saturated hydraulic conductivity of top soil (K_{sat}). RE^i is the amount of RI + Exf_{gw}, which is less than the $K_{sat} - P_e$ term. This term determines the availability of soil to still infiltrate more water (RE^i) after the initial infiltration (I) into the subsurface from the effective precipitation (P_e). The remaining water (RI + Exf_{gw} $-RE^i$) is then partitioned between RE^s and RE^e through a flow partitioning factor ($\alpha_{i,j}$), which is the fraction of flow from cell *i* to the neighboring *j* cell.

The introduction of the CRR concept in MODFLOW 6 allowed for substantial improvement in the conceptualization of overland flow

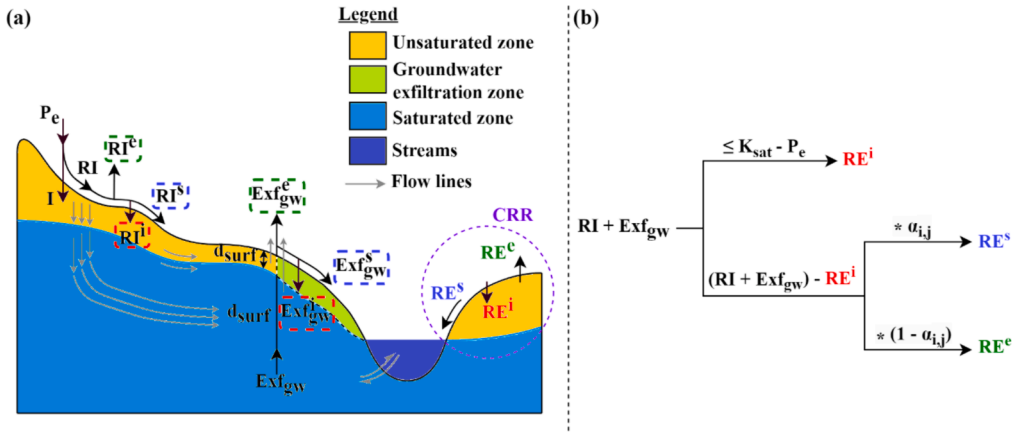

Fig. 1. Schematic diagram of the CRR concept showing: (a) it's three components (highlighted by the dash-line purple large circle); the two components surrounded by green dash-line rectangles are grouped as RE^e (total evaporated water), while the two components surrounded by red dash-line rectangles are grouped as RE^i (total re-infiltrated water), and the two surrounded by blue dash-line rectangles are grouped as RE^s (total direct runoff). P_e is the effective precipitation (precipitation – canopy interception), I is the initial amount of water infiltrated into the subsurface, and d_{surf} is the surface depth – a user-specified depth relative to the land surface where the groundwater exfiltration starts; (b) the mathematical design. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

processes within the numerical simulation. Because the parameters used in the CRR concept affect the fluxes of the water balance, they are an important part of the model calibration. The amount of RE^i is controlled by the K_{sat}, while the water partitioning between RE^e and RE^s is controlled by $\alpha_{i,j}$, which must be estimated. If not informed by observations, $\alpha_{i,j}$ is largely uncertain and can have a deleterious effect on the simulated water balance. The uncertainty of $\alpha_{i,j}$ can be reduced by assimilating additional relevant information, i.e. ideally both evapotranspiration and streamflow observations, effectively conditioning the model water balance. In this study, we use evapotranspiration and the available low-magnitude streamflow observations to condition the $\alpha_{i,j}$ in the CRR concept. This approach is expected to help improve the CRR concept by better informing the flow partitioning between evaporated water and direct runoff and help reduce the uncertainties of other parameters within the MODFLOW 6 IHM.

Reducing the parameter uncertainties in the IHM and the CRR makes it a better decision-support tool. Decision-support modelling (DSM) is undertaken in environmental/management settings when a decision needs to be made related to how humans interact with natural hydrologic systems. DSM aims to elucidate the chances that a given decision will lead to unwanted environmental or economical losses (Doherty and Moore, 2020). Generally, the models used for DSM: (1) provide robust estimates of uncertainties of model predictions, particularly those that pertain to unwanted aspects of system behavior, (2) reduce predictive uncertainties as much as possible through the inclusion of available information, typically involving history matching to the measurement of the system state, and (3) have an appropriate level of complexity (model process, structural, and parameterization complexity) with respect to both, the prediction to be made and to the data to be used (Doherty and Simmons, 2013; Doherty and Moore, 2020). Generally, using hydrological models for DSM and quantifying the model uncertainties requires separate methods beyond the model simulation (Moges et al., 2020). Many reviews highlighted the most commonly applied methods to assess uncertainty (Refsgaard et al., 2007; Moges et al., 2020; Herrera et al., 2022). Among the reviewed methods, there are inverse modelling (i.e. Bayesian statistics), predictive uncertainty, Monte Carlo analysis, sequential data assimilation, sensitivity analysis and scenario analysis. Software such as UCODE (Poeter et al., 2005; Poeter et al., 2014), PEST (Doherty, 2010), and PEST++ (White et al., 2020) are often employed to apply one or more of these methods to a hydrological model investigation (hereafter referred to as "model partners"). Depending on the model and the hydrological settings, tasks of uncertainty quantification can be computationally intensive, and require a model to run many times (up to ten thousands or more).

Given these DSM conditions, simple models tend to be used more often to support decision-making in water resources management and planning, particularly groundwater resources, and their associated studies (Moore and Doherty, 2021). Simple "standalone" hydrological models represent only one hydrological domain. Examples are models that simulate only the surface water system, the unsaturated zone, or the saturated groundwater flow domain. Simple standalone models need less run time than complex models and are more numerically stable, making them suitable tools for decision support analyses such as data assimilation and uncertainty analysis (Hugman and Doherty, 2022). However, these models are not able to fully represent all aspects of a given hydrological system regime, particularly spatial and temporal patterns of surface–groundwater interactions, compared to more complex models. As a result, the uncertainty estimates for model predictions may be biased. Moreover, the processes and parameters of the simple standalone models are not able to fully reproduce the state observations if the simplifications are substantial (Moore and Doherty, 2021). In such cases, selection of a more complex IHM model would better facilitate DSM. However, using IHMs for DSM is computationally demanding, and often precludes the use of a standard desktop computer. Thus, the use of IHMs for DSM has not been widely applied so far. The level of model complexity used by DSM has been addressed in many studies (White et al., 2019; Doherty and Moore, 2020; Doherty and Moore, 2021; Moore and Doherty, 2021; Hugman and Doherty, 2022). However, to our knowledge, the applicability of IHMs as examples of more complex models for DSM has not been demonstrated yet.

The research proposed in this study has two main objectives: (1) to improve the reliability of the CRR concept by reducing its associated uncertainties and also the uncertainties of other parameters of the applied IHM, and (2) to demonstrate that a complex IHM is able to serve decision support modelling in a computationally efficient way. To do so, the Sardon model, first documented in Daoud et al. (2022), was updated by (a) incorporating additional and different (i.e. remote sensing, RS) observations next to those documented in Daoud et al. (2022), and (b) performing ensemble-based calibration with high-dimensional parameterization.

2. Methods

2.1. Study area

The Sardon experimental catchment (~80 km^2) is located in the western part of Spain, ~40 km west of Salamanca city (Fig. 2). The catchment's altitude ranges from 730 m a.s.l. at the northern watershed boundary to 860 m a.s.l. at the southern boundary. The catchment is composed of weathered and fractured granites with sparse occurrence of other rock types. The study area has a semi-arid Mediterranean climate. The precipitation is strongly annually variable, ranging from ~300 mm yr^{-1} (2009) to > 900 mm yr^{-1} (2001). The vegetation is commonly referred to as "savannah type", which in the Mediterranean environment is represented by oak woodland. There are only two types of oak tree species, evergreen oak *Quercus ilex* (*Q.i.*) and broad-leafed deciduous oak *Quercus pyrenaica* (*Q.p.*), with sparse, ~7 % canopy coverage. The remaining area is covered by seasonal grass occurring only for approximately three months per year (typically from Mid-April to Mid-June), while for the rest of a year, the soil is bare except for some patches of evergreen *Cytisus scoparius* (Scotch broom) shrub. A land cover map with 6 classes (i.e. grass/bare soil, outcrop, *Q.i.* on soil, *Q.i.* on outcrops, *Q.p.* on soil, *Q.p.* on outcrops) was defined (Fig. 2 in Daoud et al., 2022).

The Sardon catchment has two automated data acquisition system (ADAS) stations that were implemented to monitor different hydrological variables on an hourly basis (Lubczynski and Gurwin, 2005). The ADAS stations record climatic variables including precipitation, air temperature, wind speed, relative humidity, incoming and outgoing solar radiation, and barometric pressure. Moreover, 14 automated in-situ groundwater monitoring points are spatially distributed over the Sardon catchment (Fig. 2). A flume gaging station, located at the northern catchment outlet (Fig. 2), was operational from 1997 to 2001. The flume gauge focused on low-flow river measurements (maximum discharge capacity of 145 l s^{-1}). After 2001, the low-flow has been estimated by extrapolation, using former flume flows and water levels recorded in the piezometer (before and after 2001), located at the river channel downstream (a few meters away) of the flume (Hassan et al., 2014).

An important benefit of studying the Sardon area is that it is a pilot area representative of a much larger agroforestry area known in Spain as Dehesa and in Portugal as Montado. For the remainder of this report, this region is hereafter referred to as the Dehesa-Montado (DeMo) area [~30,000 km^2, (Sales-Baptista et al., 2016)]. Both, the Sardon and the DeMo share the following characteristics: (a) land cover of oak woodland savannah type, (b) climatic conditions of water-limited environment, (c) presence of hard rocks, and (d) agroforestry, silvo-pastoral land use. The DeMo area supports multiple socio-economic uses such as livestock for high-quality meat production, agriculture, cork and fuelwood production, hunting, etc., having an essential role in the economy of rural areas (Moreno and Cáceres, 2016). However, in recent years, the increasing frequency of droughts, associated with climate change, is threatening this fragile, water-controlled environment (Carpintero et al., 2021), where it's productivity depends on water availability (Campos et al., 2013; González-Dugo et al., 2021). In such conditions, accurate monitoring and modelling of hydrological processes, and their effects on water availability and vegetative health, provide important insights regarding the system's stability. Hence, they can assist in predicting the consequences of climate change effects upon the fragile DeMo ecosystem and its water resources.

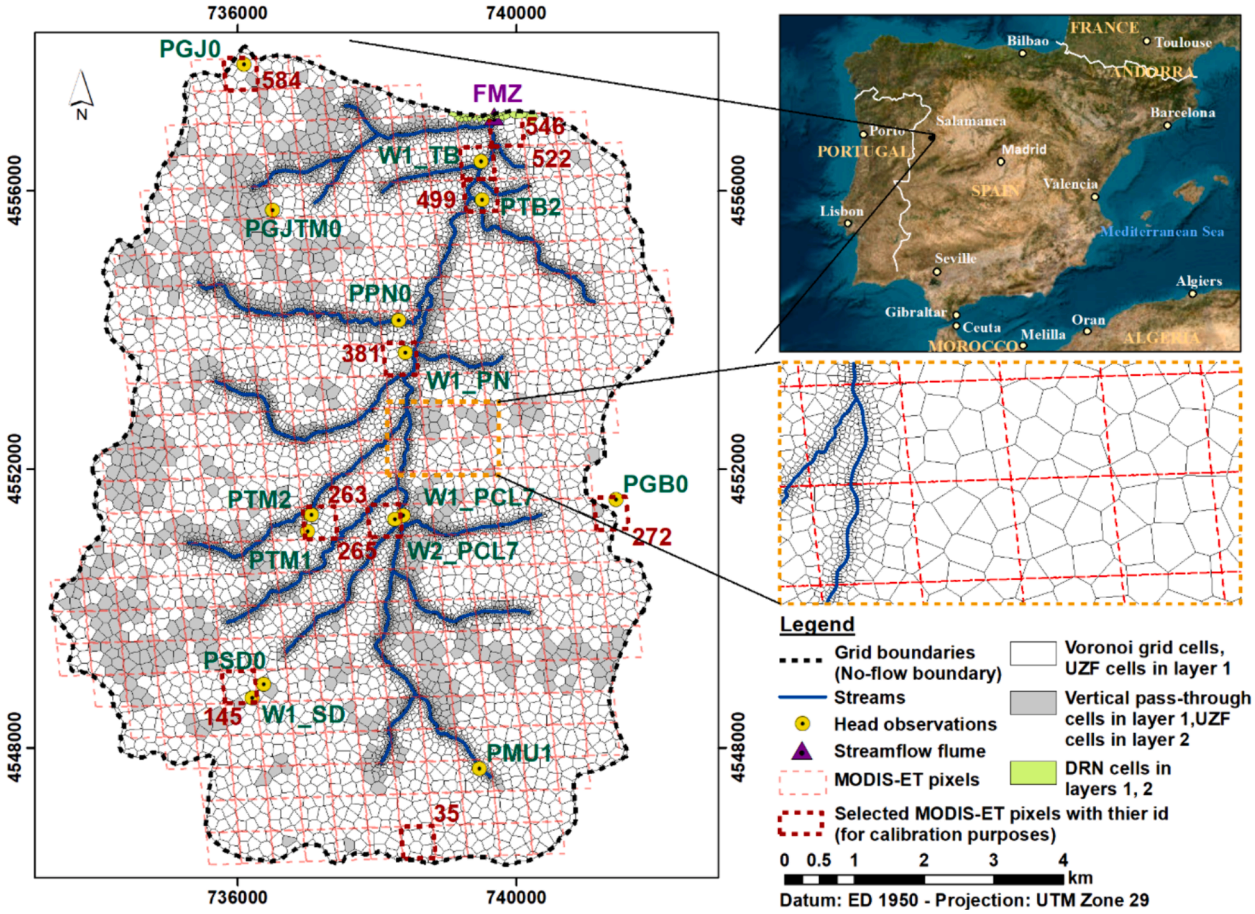

Fig. 2. Base map of the Sardon catchment with the model Voronoi grid, boundary conditions and observation locations.

2.2. Conceptual model

Details about the model conceptualization and setup are described by Daoud et al. (2022). A summary and the alterations that were made to the model conceptualization and setup, as compared to Daoud et al. (2022), are described next. The climatic driving forces are the P_e and the effective potential evapotranspiration (PET_e). P_e is equal to precipitation (P) minus canopy interception (E_I), while PET_e is equal to potential evapotranspiration (PET) minus E_I. PET_e is applied to the subsurface (unsaturated zone and groundwater zone). In this application, the total evapotranspiration (ET) is the sum of the surface evaporation (E_s) and the subsurface evapotranspiration (ET_{ss}). The E_s is the sum of the E_I and the RE^e, while the ET_{ss} is the sum of the unsaturated zone evapotranspiration (ET_u) and the groundwater zone evapotranspiration (ET_g). The numerical MODFLOW 6 code (described in section 2.3) ignores the E_s in the ET simulation. Rather, it simulates only the ET_{ss}, assuming that $ET = ET_{ss}$. Therefore, the PET_e is applied only for the simulated ET_{ss}. Detailed calculations of P, E_I, and PET are in Daoud et al. (2022).

The water balance conceptualization consists of two zones, the unsaturated zone and the groundwater zone [Fig. 3 in Daoud et al. (2022)]. The only water input to the catchment is the P, while the external outputs are the ET, the Sardon River outflow (q), and the lateral groundwater outflow across the permeable section of the northern boundary (q_g). A detailed description of the water balance and the corresponding equations are in Daoud et al. (2022).

2.3. Numerical model setup

The Sardon MODFLOW 6 model was set up using the open-source FloPy Python package (Bakker et al., 2016; Bakker et al., 2024). The model input files are archived in Daoud et al. (2024). The model is transient and uses daily stress periods. The transient simulation is initialized with an arbitrary 1-year "spin-up" period that duplicates the daily stresses of the first year of the simulation (1 October 2007 – 30 September 2008). The spin-up period helps create realistic initial conditions in the model at the start of a 7-year (1 October 2007 – 30 September 2014) calibration period, then followed by a 6-year validation period (1 October 2014 – 30 September 2020). The model uses a Voronoi unstructured grid (Sen, 2016; Fig. 2). The grid was horizontally discretized such that the smallest cells (width ~15–20 m) are located along the streams and become larger the further they are away from the streams (maximum cell width ~200 m, Fig. 2). The grid uses two layers with variable thicknesses defined earlier by Francés et al. (2014). Layer 1 cells, with zero thickness, (Fig. 2) were deactivated in the model simulation by assigning them as vertical pass-through cells (an option to represent discontinuous "pinchout" layers by connecting the cells overlying and underlying the pass-through cells) in the discretization by vertices (DISV) package (Langevin et al., 2017). Layer 2 cells are active throughout the whole model domain. The model perimeter is defined by the topographic catchment boundaries, and was simulated as a no-flow boundary. A small (<1 km) section at the northern boundary, centered on the Sardon River outlet, represents lateral groundwater outflow from the catchment in addition to surface water outflow (Fig. 2). The lateral groundwater outflow section is represented by the drain (DRN) package (Langevin et al., 2017). The unsaturated zone flow is simulated using the unsaturated zone flow (UZF) package, which uses the method of characteristics to solve the kinematic wave approximation equation (Niswonger et al., 2006). The climatic driving forces (P_e and PET_e) are specified in the UZF package. All the active cells of layer 1 and the uppermost cells of layer 2 were simulated with the UZF package to support unsaturated, or a mix of saturated and unsaturated flow conditions (Fig. 2). The Sardon River surface-water network was defined using the streamflow routing (SFR) package (Langevin et al., 2017) with 1 SFR reach per grid cell (Fig. 2). The SFR package simulates surface-water streamflow while accounting for groundwater discharge to the channel, seepage losses from channel, and overland flow discharged to the channel. Moreover, the SFR package facilitates comparison with observed flows collected at the basin outlet that helps to further constrain the calibration. The mover (MVR) package (Morway et al., 2021) is activated to simulate the direct runoff that cascades from cell to cell and from cell to stream reach for implementation of the CRR concept within the MODFLOW simulation.

The CRR concept is implemented in MODFLOW 6 through the "FACTOR" option of the MVR package. The "FACTOR" rule sets the flow fraction specified by the user that controls the transfer of the available water from a provider, in this case a UZF object (within a grid cell), to a receiver, a different downslope (either UZF or SFR) object. The available water is transferred from all providers to all receivers within the given model stress period (1 day), which approximately match the concentration time (~1 day) of the Sardon catchment. A FACTOR rule is specified spatially, for each MVR connection, and may vary within each model time step. The FACTOR rule is defined in the CRR concept as $\alpha_{i,j}$ and is calculated as follows:

$$\alpha_{i,j} = \beta_{i,j} \times S_{i,j} / \sum_{j=1}^{m} S_{i,j} \quad (1)$$

where $S_{i,j} = (elv_i - elv_j)/l_{ij}$ is the slope between the providing grid cell i and each receiver cell j that shares a boundary with cell i, elv_i and elv_j are land surface elevations of cells i and j respectively, l_{ij} is the distance between the center of the cells i and j, respectively, m is the number of j cells connected to cell i and $\beta_{i,j}$ is an additional flow factor used in model calibration, allowing for appropriate partitioning between evaporated water and direct runoff. $\alpha_{i,j}$ and $\beta_{i,j}$ range from 0 to 1. The $\left(S_{i,j} / \sum_{j=1}^{m} S_{i,j}\right)$ term in Eq. is fixed for each i-j connection, which means $\beta_{i,j}$ represents the remaining part of the $\alpha_{i,j}$ that is spatio-temporal parameter and can be calibrated.

2.3.1. Observations

The observation data used in this study (Table 1) are of five types: (1) groundwater hydraulic heads, (2) low-magnitude streamflow ($\leq$0.145 $m^3\ s^{-1}$), (3) observed evapotranspiration, and two inequality observations (4) PET, and (5) land surface levels. The first two are standard in-situ observations and were used in Daoud et al. (2022), while the latter three were not used previously. The groundwater heads of the 14 observation points and the low-magnitude streamflow observation point (Fig. 2) were retrieved from the ADAS in-situ monitoring network (Lubczynski and Gurwin, 2005).

The evapotranspiration observations were obtained from the MODIS evapotranspiration RS product (MOD16, Running et al., 2021). The MOD16 product includes 5 "data" layers, one of which represents the MODIS-estimated evapotranspiration (MODIS-ET). MODIS-ET is generated through an improved Penman-Monteith algorithm (Mu et al., 2011) and temporally aggregated every 8 days using 500 m spatial resolution pixels. The Sardon catchment is covered by 359 MODIS-ET pixels, and all the images for the modelled period, i.e. October 2007–September 2020, were downloaded using the LP DAAC NASA-USGS tool (AppEEARS Team, 2022).

Since the observation data for calibration are of different types, have different units and span multiple orders of magnitude, they were divided into groups (Table 1); each group was assigned a unique weight and an expected noise (error) in the calibration process. The weights and the noise are assigned separately and for different purposes since that separation is an important consideration in applied stochastic calibration processes. The weights are used to ensure that the contribution of each observation group to the initial objective function is approximately balanced (Doherty and Welter, 2010). The explicit representation of observation noise ensures that the calibration process appropriately accounts for the expected quality of the observation data (White, 2018).

Table 1
Observations used in model calibration divided into 5 groups by observation type. For each group, the following details are presented: the number of observation points (locations are shown in Fig. 1), the weight, the expected noise (error), the temporal resolution, the total number of observations during the entire model simulation period, and the purpose. The first three groups are for model calibration, while the last two groups are for ensuring certain physical relationships.

Group	Observation type	Number of points	Weight	Total noise (error)	Temporal resolution	Total number	Purpose
1	Groundwater heads	14	1 m^{-1}	0.2 m	Daily	33,699	model calibration
2	Low streamflow	1	0.0025 s m^{-3}	0.05 m^3 s^{-1}	Daily	1577	model calibration
3	MODIS-ET product	359	0.21 8 days mm^{-1}	5 mm 8 $days^{-1}$	8-days	115,598	model calibration
Total number of observations used for model calibration (group (1, 2, 3)						150,874	
4	PET	359	0.19 day mm^{-1}	1 mm day^{-1}	Daily	917,963	ensure simulated ET ≤PET
5	Land surface levels	385	0.19 m^{-1}	0.2 m	Daily	984,445	ensure simulated groundwater ≤land surface

In this modeling effort, the weights (Table 1) were assigned using the general adjustment options available in the calibration software (section 2.4) that adjust the weights of each observation group, so that each observation group contributed equally to the initial objective function value. The assumed noise associated with each observation group was explicitly represented (Table 1). That noise can be divided into two parts: (a) structural noise – noise due to errors in the implementation of the model's approximation of the natural system, and (b) expected measurement noise – noise due to lack of accuracy in the measurements, sensors calibration, etc. The former should be defined based on the authors' knowledge of the MODFLOW 6 numerical code and understanding of the various approximations that are required to build a numerical representation of a complex natural system, while the latter could be assumed equal to the standard deviation of each observation. During the calibration process, an ensemble of noisy observations was generated to represent the noise. The measurement noise ascribed to each observation was assumed to be the same for all the observations in each observation group. For the groundwater heads and the low streamflow, the standard deviation was assigned based on the author's knowledge about the field collection of the measurements, while for the MODIS-ET product, the standard deviation was assigned based on previous literature studies (Velpuri et al., 2013; Long et al., 2014; Jiang and Ryu, 2016; Michel et al., 2016; Moreira et al., 2019; Barraza et al., 2019; Khan et al., 2020; Zhu et al., 2022; Liu et al., 2023).

2.3.1.1. Inequality observations. In the standard MODFLOW 6 solution, the E_s is not considered as part of total ET, and the simulated ET is equal to the ET_{ss} only. Therefore, the applied PET_e is used to simulate ET_{ss} only. In this application, after adding E_s to ET_{ss}, ET may exceed PET, which makes no physical sense. To overcome this issue, the inequality observation option (White et al., 2020), available in the model partner algorithm used for the calibration (section 2.3), was used. The inequality observation is a special observation type used to make inequality constraints (less than or greater than) to reflect the nature of certain system behavior. As long as the model output satisfies the user-defined inequality constraint (either less than or greater than the corresponding observation value), there is no objective function penalty. If not, an objective function penalty is calculated and added to the total objective function of the inverse solution. The applied PET rates, pre-calculated for every grid cell, were used as inequality observations to ensure that the simulated ET $\leq$ PET.

Another inequality observation, used herein, was the land surface levels. The purpose of using them is to help constrain the simulated groundwater levels not to exceed the land surface. The land surface inequality observations were assigned at selected grid cells, ensuring representative spatial distribution over the model grid. The selected grid cells (385) are the sum of: (a) the 359 cells that match the centers of the MODIS-ET pixels (Fig. 2), which are distributed every 500 m over the study area, and (b) the 26 cells (active cells in both layers) that match the 14 groundwater observation points (Fig. 2).

2.3.2. Model parameters

The model parameters to be conditioned include: hydraulic parameters: (1) horizontal (K_h) and (2) vertical (K_v) hydraulic conductivity of the two aquifers (i.e., two layers); two storage parameters: (3) specific storage (S_s) and (4) specific yield (S_y) of the two aquifers; (5) streambed hydraulic conductivity (K_b); unsaturated zone flow parameters including: (6) K_{sat}, (7) residual water content (θ_{resid}), (8) saturated water content (θ_{sat}), (9) extinction water content (θ_{ext}), (10) extinction depth (d_{ext}); and (11–13) the CRR parameters representing the flow partitioning factors between the connected features ($\beta_{i,j}$). The parameters listed above were assigned on the grid-cell scale basis, resulting in ~500,000 estimated parameters (Table 2). The hydraulic and storage parameters were assigned as spatially-variable but temporally invariant for all active cells in both layers. Similarly, the unsaturated zone parameters also were assigned as spatially-variable but temporally invariant for all UZF cells. However, $\beta_{i,j}$ was assigned as a spatially- and semi-temporally variable for each CRR connected features (represented in MODFLOW 6 by each MVR connection) and further divided into three sets. The first set (β_{i,j_1}) represents dry months (precipitation $\leq$ 20 mm $month^{-1}$), the second set (β_{i,j_2}) represents moderately wet months (precipitation between 20 and 40 mm $month^{-1}$), and the third set (β_{i,j_3}) represents wet months (precipitation $\geq$ 40 mm $month^{-1}$).

2.4. Model calibration

After the inclusion of the MODIS-ET into the Sardon model, which is pixel-size-scale (500 m), the total number of observations was approximately 150,000 (Table 1). Also, the model parameters were assigned on the grid-cell-scale, resulting in approximately 500,000 parameters (Table 2) for analysis of model input uncertainty. For such a highly parameterized system, manual calibration (trial and error) is not possible and the inverse solution, using model partners (specific software) is required. The model inverse solution was accomplished through the iterative ensemble smoother (IES) approach implemented in the model-independent parameter estimation open-source software PEST++ (PESTPP-IES; White, 2018; White et al., 2020). The IES algorithm is a derivative-free ensemble-based data assimilation approach, first put forward by Chen and Oliver (2013). The algorithm uses a Monte Carlo analysis to approximate the first-order relation between model parameters and observations. It uses the well-known Gauss-Levenberg-Marquardt algorithm (Oliver et al., 2008; Doherty, 2015) to simultaneously upgrade the parameter ensemble toward minimizing a stochastic objective function (the weighted sum-of-squared residuals between the observations and the equivalent model-simulated values). Fig. 3 shows the steps followed during the model calibration using the PESTPP-IES algorithm with pre- and post-processing phases. The process model (MODFLOW 6) was linked to the PESTPP-IES algorithm through the open-source pyEMU Python package (White et al., 2016).

2.4.1. Pre-processing phase

On many occasions, there is a need to define inter-relationships between different parameters so that known parameter relationships are appropriately maintained during the inverse solution. To this end, the parameter preprocessor PAR2PAR, from the suite of PEST utilities (Doherty, 2021), was adopted to enforce two important parameter relationships; the first ensures that the K_v of each grid cell does not exceed

Table 2
Model parameterization.

	Parameter abbreviation	Parameter group	Dependency	Unit	Lower bound	Upper bound	Initial value (prior)	Calibrated value (posterior)	Number
1	K_h	Horizontal hydraulic conductivity	–	m day^{-1}	0.0005	15	0.05	0.0005 – 15	46,604
2	K_v	Vertical hydraulic conductivity	$\leq K_h$	m day^{-1}	0.0005	0.5	0.01	0.0005 – 0.5	46,604
3	S_s	Specific storage	–	m^{-1}	10^{-6}	10^{-5}	5×10^{-6}	10^{-6} – 10^{-5}	46,604
4	S_y	Specific yield	–	–	0.03	0.08	0.05	0.03 – 0.08	46,604
5	K_b	Streams' bed hydraulic conductivity	–	m day^{-1}	0.0001	1	0.01	0.0001 – 1	4132
6	K_{sat}	Saturated hydraulic conductivity	$=K_v$	m day^{-1}	0.0005	0.5	0.01	0.0005 – 0.5	23,302
7	θ_{resid}	Residual water content	–	m^3 m^{-3}	0.02	0.09	0.05	0.02 – 0.09	23,302
8	θ_{sat}	Saturated water content	–	m^3 m^{-3}	0.3	0.4	0.35	0.3 – 0.4	23,302
9	θ_{ext}	Extinction water content	$=\theta_{resid}$	m^3 m^{-3}	0.02	0.09	0.05	0.02 – 0.09	23,302
10	d_{ext}	Extinction depth	–	m	0.0001	5	1	0.0001 – 1.5	23,302
11	β_{i,j_1}	Flow partitioning factor during dry months	–	–	0.7	1	1	0.76 – 1	70,436
12	β_{i,j_2}	Flow partitioning factor during moderate months	–	–	0.8	1	1	0.82 – 1	70,436
13	β_{i,j_3}	Flow partitioning factor during wet months	–	–	0.9	1	1	0.92 – 1	70,436
								Total	471,762

its K_h (i.e., $K_v \leq K_h$) and the second, that $\beta_{i,j}$ is adjusted (calibrated) prior to the calculation of $\alpha_{i,j}$ which is passed to the process model.

Another pre-processing phase introduces another ET-related flux: E_I, which was calculated outside MODFLOW 6 by Daoud et al. (2022), using the revised Gash analytical model (Gash et al., 1995). During each PESTPP-IES iteration, E_I was added to the simulated by MODFLOW 6 ET fluxes to define the total ET for the comparison with the MODIS-ET observations.

2.4.2. Model run phase

150 parameter realizations were generated through PESTPP-IES to achieve model calibration. PESTPP-IES is designed in such a way that its computational burden is insensitive to the number of parameters to be conditioned (White, 2018). That allowed increasing the number of the conditioned parameters to ~500,000 with little extra computational cost. The built-in parallel run manager option in PESTPP-IES was used to reduce the waiting time to complete the inverse solution. The realizations were run in parallel using a Linux machine with 128 CPUs. The maximum number of sequential iterations was set equal to 3. Prior to the first iteration, PESTPP-IES generated an ensemble of parameter fields (150) by sampling the parameter values from the user-specified multi-Gaussian prior parameter distribution. The prior parameter distribution was truncated using the applied parameter bounds (Table 2) to avoid extreme (i.e. implausible) parameter values being drawn or sought from the multivariate Gaussian distribution. In the iterations that followed and for each individual realization, the parameters (Table 2) were modified to minimize the misfit between the observations and the model-simulated equivalent values. The outcome of the calibration process undertaken with PESTPP-IES was the posterior parameter ensemble, in which each realization was considered a "calibrated model".

2.4.3. Post-processing phase

The outputs of interest from each model run were the daily time series of: (1) simulated groundwater heads of the 14 model grid cells that host the 14 groundwater monitoring points (Fig. 2); (2) simulated streamflow at the catchment outlet (Fig. 2); and (3) the following four simulated fluxes: RI^e, Exf^e_{gw}, ET_u, and ET_g. For the groundwater heads and the streamflow, direct matching between the simulated and the observed values was possible, while post-processing was needed for comparing the simulated ET with the observed MODIS-ET (Fig. 3). First, the four daily simulated ET fluxes (RI^e, Exf^e_{gw}, ET_u, and ET_g) were summed and added to a fifth flux, E_I calculated outside MODFLOW 6, to estimate the daily total simulated ET. Next, the mosaic of daily simulated ET for all grid cells (Fig. 2) was overlain on top of each MODIS-ET pixel for calculating a spatially averaged (see the zoom window in Fig. 2) ET equivalent that matched the spatial resolution of the MODIS-ET observations. Third, the daily spatially-averaged ET was temporally aggregated every 8 days to match the temporal resolution of the MODIS-ET observations. Finally, the daily total simulated ET (calculated from the second step of the model post-processing) was compared with the corresponding applied PET to ensure that the daily total simulated ET $\leq$ PET. Similarly, the simulated groundwater heads were compared with the corresponding land surface level to ensure that the daily groundwater heads $\leq$ land surface level. PESTPP-IES repeated the pre-processing phase, the forward model run, and the post-processing phase for every realization (150) in each iteration (3).

2.5. Validation assessment

The validation capability of the ensemble of the calibrated realizations, yielded by the calibration process, was assessed by simulating a validation period of unseen observations, which were not used in the calibration. This period encompasses 1 October 2014 to 30 September 2020. The same procedure, used to define the climatic driving forces (P, E_I and PET) during the calibration period, was used in the validation period, except for a gap in the P records that extended for nearly two years (i.e., 16 June 2016–23 May 2018). For model observations, the same procedure as used for the calibration period (section 2.2.3) was followed. Each calibrated parameter realization was used to run the validation period, and the simulation results were recorded.

2.5.1. Data gap filling in the validation period

The 2-year P data gap was filled in using the Random Forest (RF) machine learning algorithm, available in the Scikit-learn Python package (Pedregosa et al., 2011). The variables used to fill in the gap were the P records of 5 meteorological stations neighboring the Sardon catchment. The two ADAS stations (Fig. 2) of the Sardon catchment, which include the P records were established in 1996 and continue through 2022; therefore, for the 5 neighboring stations, the P records of the same 25 years period, i.e. 1 October 1996 – 30 September 2021, were retrieved from the Agencia Estatal de Meteorología (AEMET) website (La Agencia Estatal de Meteorología (AEMET), 2022). The gap-fill-in task was performed in three main steps: (a) build the RF model and tune its parameters, (b) evaluate the model accuracy, and (c) use the model to predict the gap. The RF model was built by splitting the P data into 75 % training data and 25 % testing data. The RF model parameters (e.g. number of trees, maximum tree depth, minimum number of samples split, etc.) were tuned (calibrated) to ensure the best fit between the estimated P of the training set and the corresponding observed P of the target Sardon station. The model fit was then assessed using two different statistical metrics, i.e. (1) the coefficient of determination (R^2) between the two pairs – the estimated P of the testing set and the corresponding observed P of the target Sardon station, and (2) the root mean square (RMSE) of the same two pairs. To ensure stability, the RF

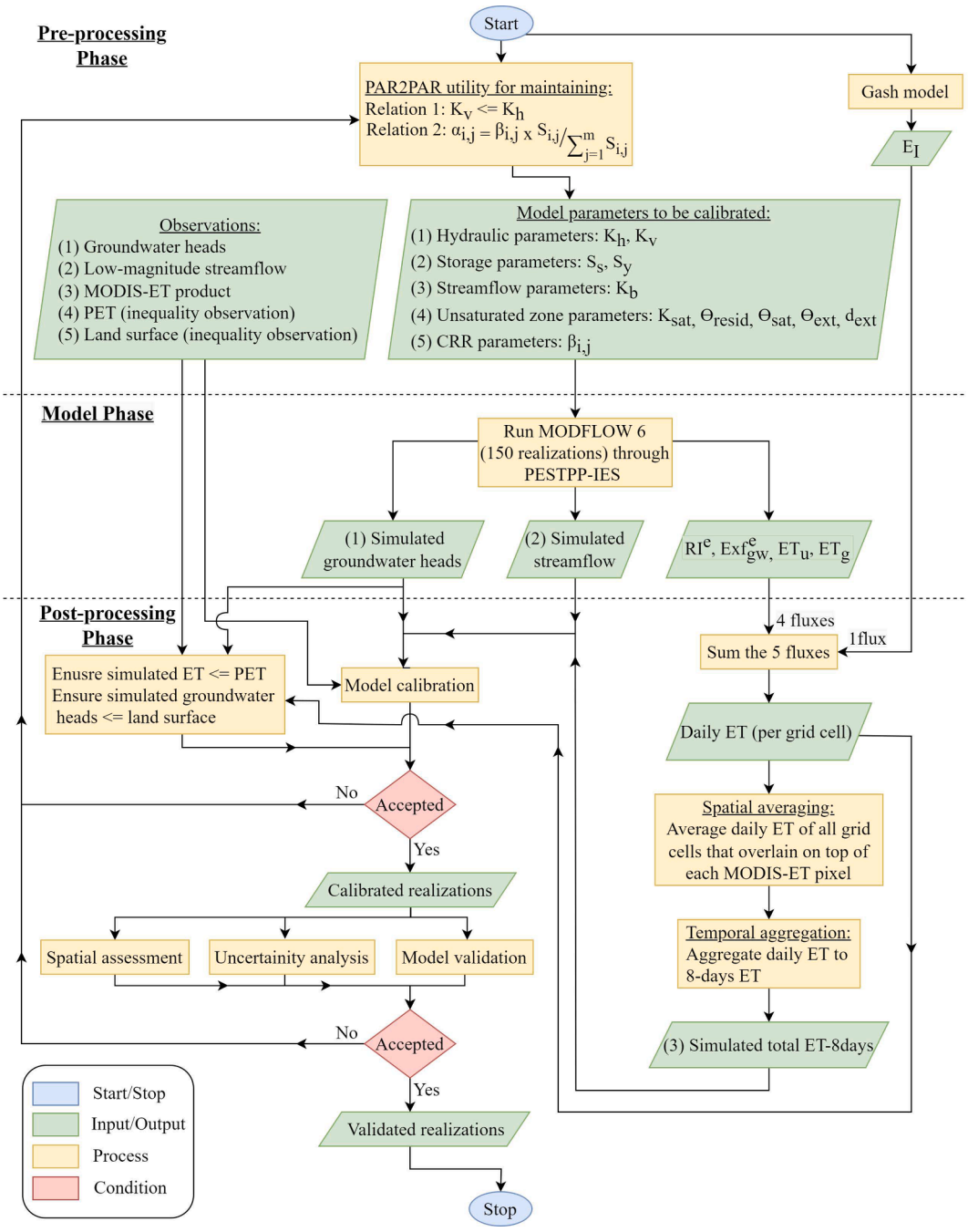

Fig. 3. Methodology flowchart to run the MODFLOW 6 model through PESTPP with pre- and post-processing phases.

model was evaluated by running the model multiple times (e.g. 50, called model cross-validation), each run had different and random P data split into training/testing sets. The model fit was deemed acceptable when the average R^2 of all the RF models runs was $\geq$0.6 and the average RMSE was $\leq$2.5 mm day^{-1}. Finally, the model was used to predict the P data gap of the target Sardon station.

2.6. Spatial assessment

Incorporating RS spatially-distributed observations (e.g. MODIS-ET) into the model, would require assessing the spatial distribution of the model-simulated equivalent values. The model was assessed spatially through the SPAtial EFficiency (SPAEF) matrix (Demirel et al., 2018a; Koch et al., 2018). The SPAEF matrix is calculated as per Eq. (2) (Demirel et al., 2018b) and reflects three weighted components: correlation, coefficient of variation and histogram match. Among the used observations, the SPAEF matrix was calculated for the MODIS-ET since it is the only observation type that is spatially continuous.

$$SPAEF = 1 - \sqrt{(r-1)^2 + (cv-1)^2 + (h-1)^2}$$

$$r = \rho(obs,\ sim),\ cv = \left(\frac{\sigma_{sim}}{\mu_{sim}}\right) / \left(\frac{\sigma_{obs}}{\mu_{obs}}\right),\ \text{and}\ h = \frac{\sum_{j=1}^{n} \min(K_j, L_j)}{\sum_{j=1}^{n} K_j} \quad (2)$$

where r is the Pearson correlation coefficient between the observed MODIS-ET and the simulated ET, cv is the fraction of coefficient of variations representing spatial variability, and h is the percentage of histogram intersection (Swain and Ballard, 1991). The h is calculated for a given histogram K of the observed MODIS-ET and the histogram L of the simulated ET, each one containing number of bins (n), i.e. herein n = 10 bins. The calculations are standardized to a mean (μ) of 0 and a standard deviation (σ) of 1 (z score) to avoid the effect of different units.

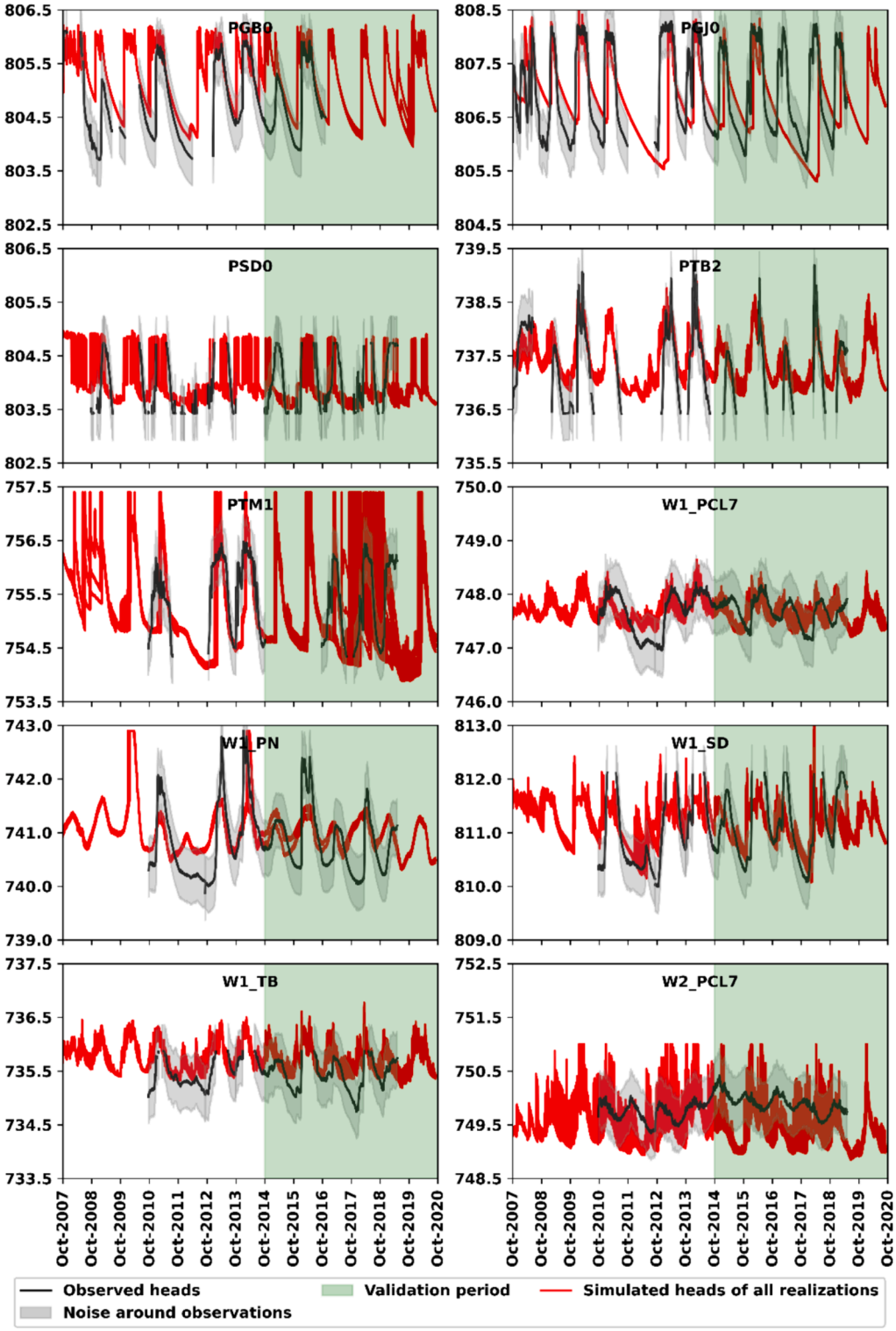

Fig. 4. Groundwater hydrographs showing the daily simulated heads of all realizations (red lines) versus the daily observed heads (black lines) at 10 selected observation points within the calibration period [1 October 2007 – 30 September 2014 (non-shaded)] and the validation period [1 October 2014 – 30 September 2020 (green-shaded)]. The locations of groundwater head observations are shown in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Calibration results

Fig. 4 compares the daily simulated heads for all realizations to the daily observed heads at 10 of the 14 monitoring points. The visual comparison of the results generally show an acceptable match between the observed and the simulated heads at each location, not only during the calibration period but also during the validation period. W1_PCL7 and W1_SD show a very good match between the simulated and the observed values, while the calibrated fits at PTM1 show some misfit. That is, at PTM1, many of the simulated values fall outside the grey region (observed values plus the noise range).

Fig. 5 compares the simulated ET-8 days of every calibrated

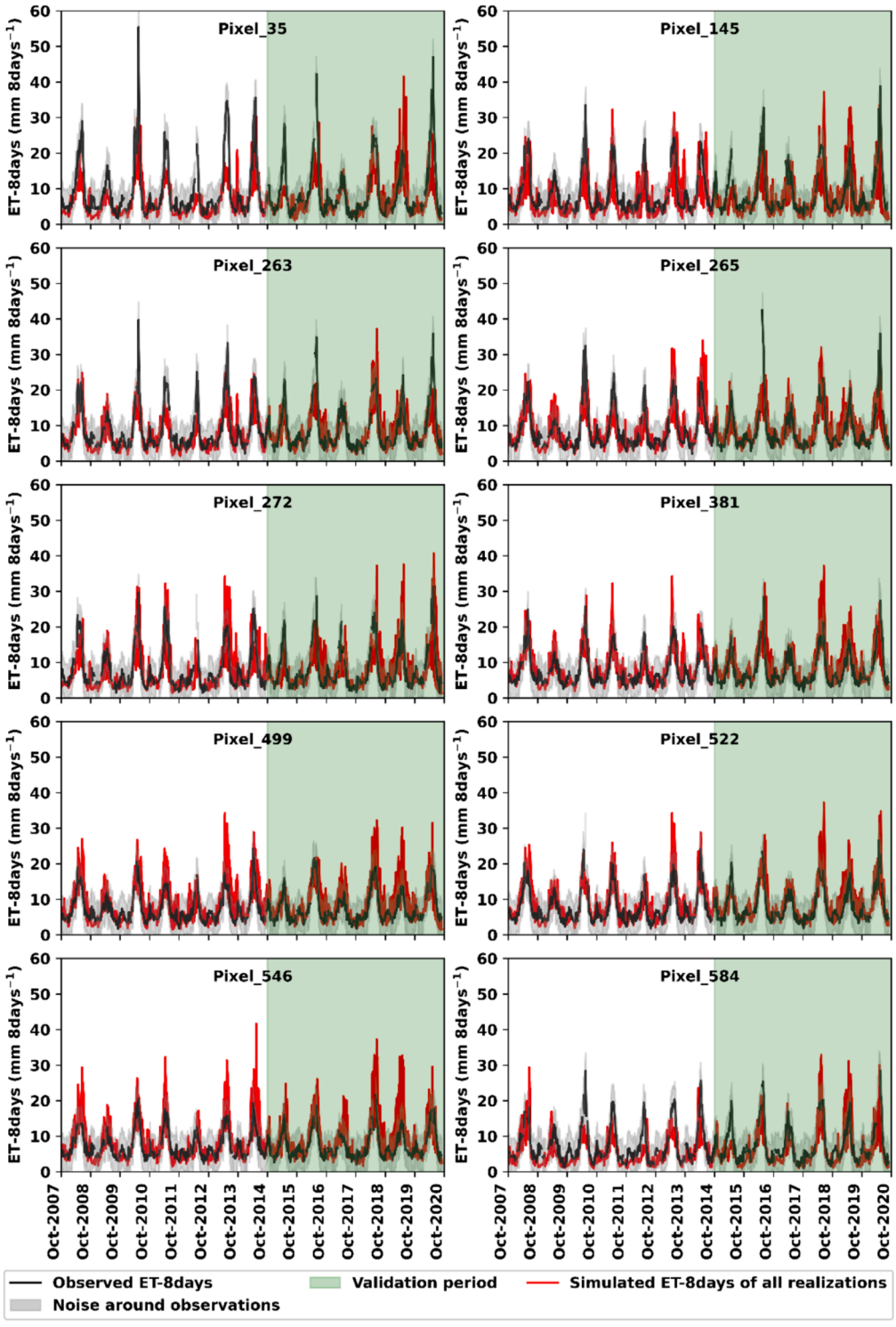

Fig. 5. Evapotranspiration hydrographs showing the simulated ET-8 days of all realizations (red lines) versus the observed ET-8 days (black lines) at 10 selected observation MODIS pixels within the calibration period [1 October 2007 – 30 September 2014 (non-shaded)] and the validation period [1 October 2014 – 30 September 2020 (green-shaded)]. The IDs and locations of the selected ET observation pixels are shown in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

realization to the observed MODIS-ET 8 days at 10 selected observation pixels. In general, the visual comparison of the results show an acceptable match between the observed and simulated ET-8 days across all the observation pixels during both, the calibration period and the validation period. The simulated ET exhibits the best fit with its observed equivalent (compared to simulated groundwater heads and streamflow), since, for most of the pixels, the simulated ET-8 days fall within the grey region (observed values plus the noise range). However, at few pixels, such as pixel 35, during certain simulation periods, there was a failure to capture the observed peaks (below the grey region). In contrast, there were over-estimated peaks (above the grey region) such as the ones at pixel 546.

Fig. 6a shows the daily simulated total streamflow, the daily simulated baseflow of all realizations, and the daily observed low flow at the outlet point. Fig. 6b is a zoom window showing the model fits in 2009 (dry year) and 2010 (wet year) during the calibration period, whereas Fig. 6c highlights the model fits in 2015 (dry year) and 2016 (wet year) during the validation period. The total simulated streamflow is the summation of the simulated direct runoff and the simulated baseflow. While the flume can only measure low flows ($\leq$0.145 m^3 s^{-1}), these observations similar to Daoud et al. (2022), were critical for calibrating the simulation during baseflow recessions, for example, during April 2009 and again in April 2017. In addition, the low flow observations guided assessment of the simulated streamflow when the observed streamflow fell below 0.145 m^3 s^{-1}. However, calibrating the simulated (red line) streamflow outside the above-mentioned streamflow recession periods was not attempted. The true channel outflow at the Sardon catchment outlet consists of both surface flow (measured) and significant subsurface flow along a major fracture system aligned with the surface river channel (Lubczynski and Gurwin, 2005).

The model spatial performance was assessed based on the spatial matrix (SPAEF). The realization with minimum error variance was selected for the spatial assessment. Table 3 presents the SPAEF, with its components, for two different days that each had high ET rates during the calibration and validation periods respectively. In addition, the SPAEF for the entire model period was calculated and is shown in Table 3. In general, the 1-day analyses show an acceptable SPAEF during both the calibration period (0.54) and the validation period (0.47). The day selected during the calibration period (1 May 2009) showed a reasonable correlation of 0.64, a relatively high coefficient of variation of 1.11 indicating a high spatial variability, and a histogram match of 0.76 indicating reasonable spatial details within the separated clusters between high and low values. The day selected during the validation period (1 May 2015) showed a high correlation of 0.92, an even higher coefficient of variation of 1.31 (compared to the day selected during the calibration period), and a reasonable histogram match of 0.59. The SPAEF calculations over the entire model simulation showed smaller values compared to the individual day examples. The combination of a low correlation of 0.36, a high coefficient of variation of 0.96, and a reasonable histogram match of 0.73 led to an overall low SPAEF of 0.3, mainly affected by the low correlation.

3.2. Calibrated parameters

The conditioned parameters were divided into groups (Table 2) assigned on the grid-cell-scale, resulting in approximately 500,000 estimated parameters. The inverse solution, guided by the observations, had a greater effect on specific parameters than others. The calibrated parameter values in the K_h, K_v, θ_{ext}, S_y parameter groups, spanned the range between lower and upper bounds (Table 2), while parameters in the d_{ext} and $\beta_{i,j}$ parameter groups covered a limited part of the available

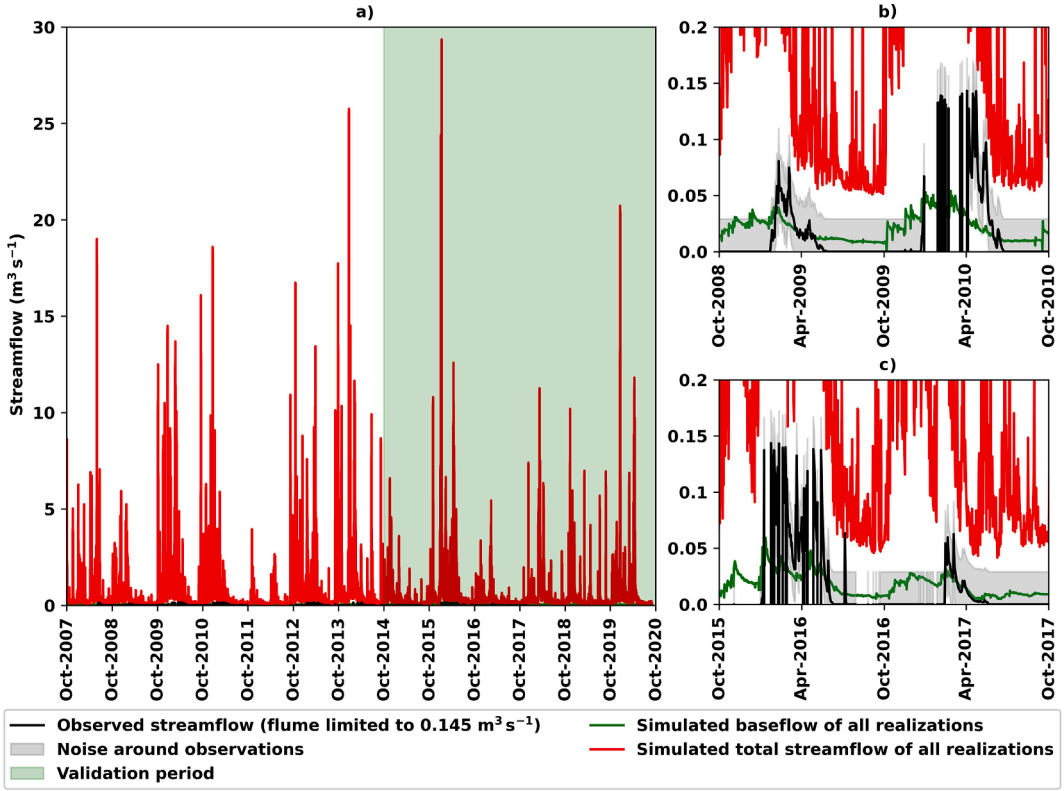

Fig. 6. Daily simulated streamflow (red lines) and simulated baseflow (green lines) of all realizations versus the daily observed low flow at the basin outlet (black lines) within: (a) the calibration period [1 October 2007 – 30 September 2014 (non-shaded)] and the validation period [1 October 2014 – 30 September 2020 (green-shaded)], (b) zoom-window showing the model fit for two years during the calibration period with different hydrologic conditions (dry: 2009; wet: 2010), and (c) zoom-window showing the model fit for two years during the validation period (dry: 2015; wet: 2016). Note that: (1) the detection limit for the flume-observed low flow is 0.145 m^3 s^{-1}, and (2) the simulated total streamflow is graphically restricted in zoom windows to <0.2 m^3 s^{-1} for visualization purposes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
SPAEF calculations, with its components as per Eq. (2), for two different days in the simulated period. Both days correspond to high ET rates, the first is during the calibration period and the second is during the validation period. The SPAEF for the entire model period is shown in the third row. Note that the optimal values for all the components are = 1.

Period	SPAEF components			SPAEF
	Correlation (r)	Coefficient of variation (cv)	Histogram match (h)	
1-May-2009	0.64	1.11	0.76	0.54
1-May-2016	0.92	1.31	0.59	0.47
Entire model period	0.36	0.96	0.73	0.30

parameter range. The parameter groups K_h, S_y, θ_{ext} had the largest changes between their prior and posterior "calibrated" ensemble values. Changes in the K_v group were moderate, while changes in the K_b and $\beta_{i,j}$ groups were the smallest. Fig. 7 shows the mean absolute changes for 6 selected parameter groups, 4 of them with large changes, one with moderate changes (K_v) and one with small changes ($\beta_{i,j}$). The changes are expressed in Fig. 7 by the histogram bins (colored with grey). The changes are large (e.g. Fig. 7a, c, and d) when more bins appear (y-axis) with less percentage (towards the 0 value in the x-axis). The changes are small (e.g. Fig. 7f) when less bins (y-axis) appear with less percentage (towards the 0 value in the x-axis).

Considering the CRR concept, the $\beta_{i,j}$ parameter groups changed little during model calibration. The initial $\beta_{i,j}$ value in the prior ensemble was all set to 1 (which was also the maximum value, Table 2), meaning that all the available water (RI + Exf_{gw}, Fig. 2) was partitioned to be RE^s and $RE^e = 0$. The reasons for low changes (deviation from the initial value = 1) in $\beta_{i,j}$ is likely due to that the occurs during and/or shortly after rainfall periods when the vapor pressure deficit is low so the expected evaporation is also low. The sum of RI and Exf_{gw}, from which RE^e originates, was an order of magnitude higher than the other ET fluxes (ET_u and ET_g). Thus, in order to let RE^e be within the same order of magnitude as the other ET fluxes, $\beta_{i,j}$ should be close to 1, (that was also taken into consideration in the lower bounds of $\beta_{i,j}$ – minimum was 0.7, Table 2). As a result, the changes in the total ET were very sensitive to the changes of $\beta_{i,j}$, and consequently, the deviation of the $\beta_{i,j}$ from the initial value (=1) was minimal.

3.3. Water balance

The daily water balance (WB) of all simulation realizations was exported and aggregated into yearly rates. The realization with minimum error variance was selected for further WB analysis. Table 4 shows the 13-year WB mean values of the entire catchment over the total model simulation period (1 October 2007 – 30 September 2020), presented in percentage of P. The WB equations are described in Daoud et al. (2022). The only input to the catchment is P, while the outputs are: ET = 57.5 % of P, q = 43.0 % of P, and negligible q_g = 0.2 % of P (Table 4). The contribution of the CRR fluxes was substantial in the WB. For the RE^i, the initial infiltration to the unsaturated zone (I = 66.8 % of P) was increased by the RE^i (8.9 % of P) resulting in final active infiltration (I_a = I + RE^i) equal to 75.7 % of P (Table 4). For the RE^s, the q (43.0 % of P) was mainly originated from the RE^s (39.4 % of P) in addition to low q_B (3.6 % of P) (Table 4). While for RE^e, it's contribution (5.5 % of P) to the total ET was not as large as expected and likely due to the large contribution of the other ET fluxes (ET_u and ET_g; 29.9 and 14.6 % of P respectively) (Table 4). Considering the R_n, it is an essential flux of interest for describing the sustainability of groundwater resources (Hassan et al., 2014). The R_n (3.8 % of P) is calculated as the difference between the R_g (46.6 % of P) and the sum of the ET_g (14.6 % of P) and the Exf_{gw} (28.2 % of P) (Table 4). The R_n varied between −1.5 % of P in the driest years (2009 and 2015) to 5.5 % of P in the wettest years (2013 and 2014) and on average was 3.8 % of P over the simulation period.

3.4. Uncertainty analysis

The uncertainties of the fluxes were evaluated through the progress from the prior ensemble to the posterior and the predictive ensembles. Fig. 8 shows histograms of 5 fluxes of interest on 4-day examples, representing dry and wet conditions during the calibration period (Fig. 8a, b) and during the validation period (Fig. 8c, d). The first three fluxes are the CRR fluxes, then the total ET, and finally the R_n. For most of the shown examples, the histogram shape, representing the probability density distribution, was converted from a diffuse shape of the prior ensemble to a narrower (lower variance) and higher peak (more certain) on two days with different hydrological conditions (dry and wet) during the calibration period (Fig. 8a, b). Similarly, the predictive ensemble during the validation period (Fig. 8c, d) was similar to the posterior ensemble, with a high peak and narrower range, indicating a reduction in the uncertainty during the predictions with an acceptable level of certainty.

However, the reduction of the uncertainty was not at the same level

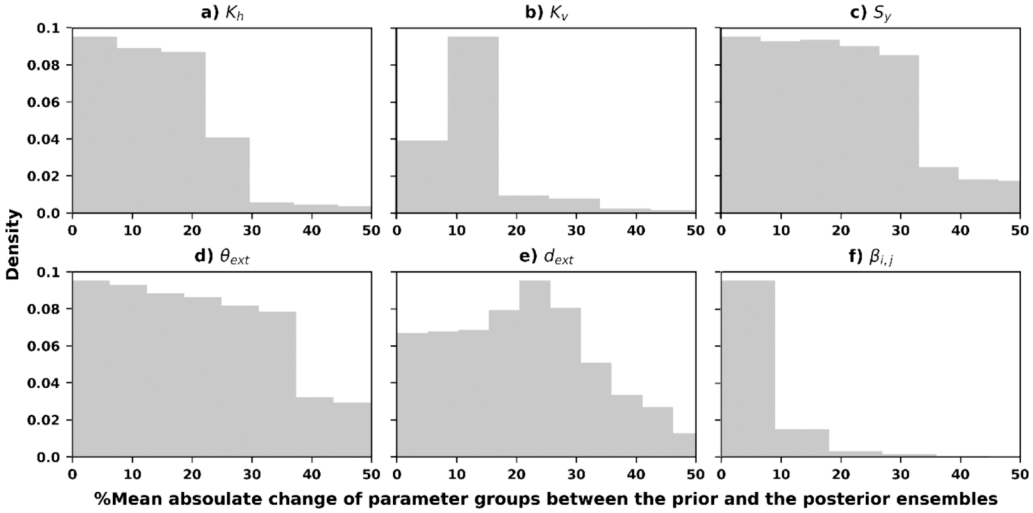

Fig. 7. Mean absolute change in parameter values between their prior and posterior values during calibration process for selected parameter groups.

Table 4
Mean WB of the entire catchment over the total model simulation period (1 October 2007 – 30 September 2020) presented as percentage of P.

Symbol	Flux	Dependency	% of P
P	Precipitation	–	100.0
E_I	Canopy interception	–	7.5
P_e	Effective precipitation	–	92.5
RI	Rejected infiltration	–	25.7
I	Initial infiltration	–	66.8
RE^i	CRR re-infiltrated water	–	8.9
I_a	Active infiltration	$= I + RE^i$	75.7
RE^e	CRR evaporated water	–	5.5
ET_u	Unsaturated zone evapotranspiration	–	29.9
ET_g	Groundwater evapotranspiration	–	14.6
ET	Evapotranspiration	$= E_I + RE^e + ET_u + ET_g$	57.5
q_B	Baseflow	–	3.6
RE^s	CRR direct runoff	–	39.4
q	Stream outflow	$= q_B + RE^s$	43.0
R_g	Gross recharge	–	46.6
Exf_{gw}	Groundwater exfiltration	–	28.2
R_n	Net recharge	$= R_g - ET_g - Exf_{gw}$	3.8
q_g	Lateral groundwater outflow		0.2
ΔS_u	Unsaturated zone storage		0.3
ΔS_g	Groundwater storage		−0.9
ΔS	Total catchment storage	$= \Delta S_u + \Delta S_g$	−0.6
WB of the entire catchment		$P = ET + q + q_g \pm \Delta S$	

in the shown examples. The reduction in the uncertainty can be mathematically expressed by the decrease in the standard deviation and the change in the mean from the prior to the posterior/predictive fluxes. For example, for the CRR fluxes, the decrease in the standard deviation (std) from the prior to the predictive RE^e on the wet day – 15 January 2016 was 0.23 mm day^{-1} (log-based) – all the following values are also log-based. While there was a minor decrease in the std from the prior to the predictive RE^e on the dry day – 15 August 2015 (0.02 mm day^{-1}). Moreover, there was a significant change in the mean RE^s (0.97 mm day^{-1}) from the prior to the predictive on the wet day – 15 January 2016. For ET, the difference between the mean of the posterior ET and the observed MODIS-ET was minimized on the wet day – 15 January 2010 (0.08 mm day^{-1}) as compared to the difference between the mean of the prior ET and the observed MODIS-ET on the same day (0.16 mm day^{-1}). Similarly, the difference between the mean of the predictive ET and the observed MODIS-ET was minimized on the dry day – 15 August 2015 (0.1 mm day^{-1}) as compared to the difference between the mean of the prior ET and the observed MODIS-ET on the same day (0.29 mm day^{-1}). For R_n, the changes in the mean and the decrease in the std from the prior to the predictive were significant on the dry day – 15 August 2015 (0.67, 0.5 mm day^{-1} respectively) and were minor on the wet day – 15 January 2016 (0.07, 0.02 mm day^{-1} respectively).

4. Discussion

4.1. Highly parameterized modelling

In fulfilment of the first objective of this study, namely to improve the implementation of the CRR concept by reducing its associated uncertainties and also the uncertainties of other parameters of the Sardon MODFLOW 6 IHM, two main techniques were adopted. First, a highly parameterized (number of parameters was approximately 500,000, Table 2) approach was adopted during model calibration to: (a) account for a great number of parameters' uncertainties, and (b) provide a robust assimilation of the spatial and temporal characteristics of the RS observations, both leading to improved IHM predictive capability. The highly parameterized and calibrated IHM helped improve the simulation (i.e., improved fit to observations) of the Sardon Catchment. Consequently, it improved the simulated fluxes, particularly the fluxes of interest that cannot be easily observed (e.g. R_n) and/or the fluxes that are originated from the combination of different fluxes (e.g. $ET = E_I + RE^e + ET_u + ET_g$ and $R_n = R_g - ET_g - Exf_{gw}$) where each flux is controlled by different parameter(s). A number of these parameters (e.g. θ_{ext}, d_{ext}, $\beta_{i,j}$, section 3.2) showed changes in their mean from the prior to the posterior ensembles. Consequently, the ET and R_n fluxes (examples presented in section 3.4 and Fig. 8) showed reductions in their standard deviation and differences in the mean between the prior and the posterior/predictive ensembles, indicating a reduction in the uncertainties of these fluxes, and a reduction in the uncertainties of the corresponding parameters.

The MODFLOW 6 IHM, like many other IHMs, includes explicit simulation of several processes (e.g., groundwater flow, surface-water flow) as well as the interaction between these processes across the unsaturated zone. When additional processes are explicitly represented in a model, the total number of parameters, that must be specified, increases. With the application of PESTPP-IES, the additional parameters in the IHM provide sufficient parameter density for obtaining a good fit between simulated state variables and observations and help guarantee that post-calibration predictions, associated with alternative management of interest, are not artificially limited by a lack of parameter insufficiency (Doherty and Moore, 2020). However, the use of temporally long and spatially dense records of observations, such as the MODIS-ET RS product used herein, requires a high level of parameterization to be able to properly reflect their spatial and temporal variability. Sufficiently parameterized models, when coupled with the model partners, i.e. PESTPP-IES, facilitate good replication of the historical behavior of the system, indicating an appropriate level of assimilation of the information in the data (Hugman and Doherty, 2022).

4.2. Use of non-traditional observations

In addition to highly parameterized model calibration, the second technique used to accomplish our first objective was the inclusion of non-traditional observation types (e.g. MODIS-ET) in the Sardon model calibration – in addition to traditional observations, i.e. hydraulic heads and streamflow (low-flow). Doing so helped constrain the unusually high number of parameters used in the IHM with the CRR concept. Benefits from the use of non-traditional MODIS-ET observation data include: (1) constraining the total simulated ET, and (2) proper partitioning of $RI + Exf_{gw}$ between the two water fluxes, i.e. RE^e and RE^s. The changes in the parameters' groups (θ_{ext}, d_{ext}, $\beta_{i,j}$) from the prior to the posterior ensembles (Table 2 and Fig. 7) demonstrate the relative importance of this approach. For example, the sensitivity of θ_{ext}, d_{ext}, and $\beta_{i,j}$ is shown by their significant impact on RE^e, ET_u, and ET_g, which are components of the total simulated ET that was subsequently compared to the MODIS-ET observations.

To highlight the susefulness of the MODIS-ET observations in the WB and in the reliability of the CRR concept, the model was calibrated, following the same calibration methodology (section 2.4; Fig. 3), but without the MODIS-ET observations. Then, the two calibrated models (with and without the MODIS-ET) were compared. Table 5 shows the differences in the CRR fluxes, ET, q, R_g and R_n between the two calibrated models in terms of the contribution to the WB (in % of P). The calibrated model without MODIS-ET observations had a higher RE^e and ET and a lower RE^s and q than the one that used MODIS-ET observations. The likely reason is that the parameters θ_{ext}, d_{ext}, $\beta_{i,j}$, which affect the ET fluxes, were not constrained by any observations, leading to an overestimation of the ET fluxes (RE^e, ET_u, and ET_g) and consequently an underestimation of the RE^s and the q. The RE^i was similar between the two calibrated models. This similarity was expected because the RE^i is conditioned by the K_{sat}, which is not affected by the MODIS-ET observations. The model, calibrated without including MODIS-ET observations in the calibration dataset, predicted negative R_n. The likely reason

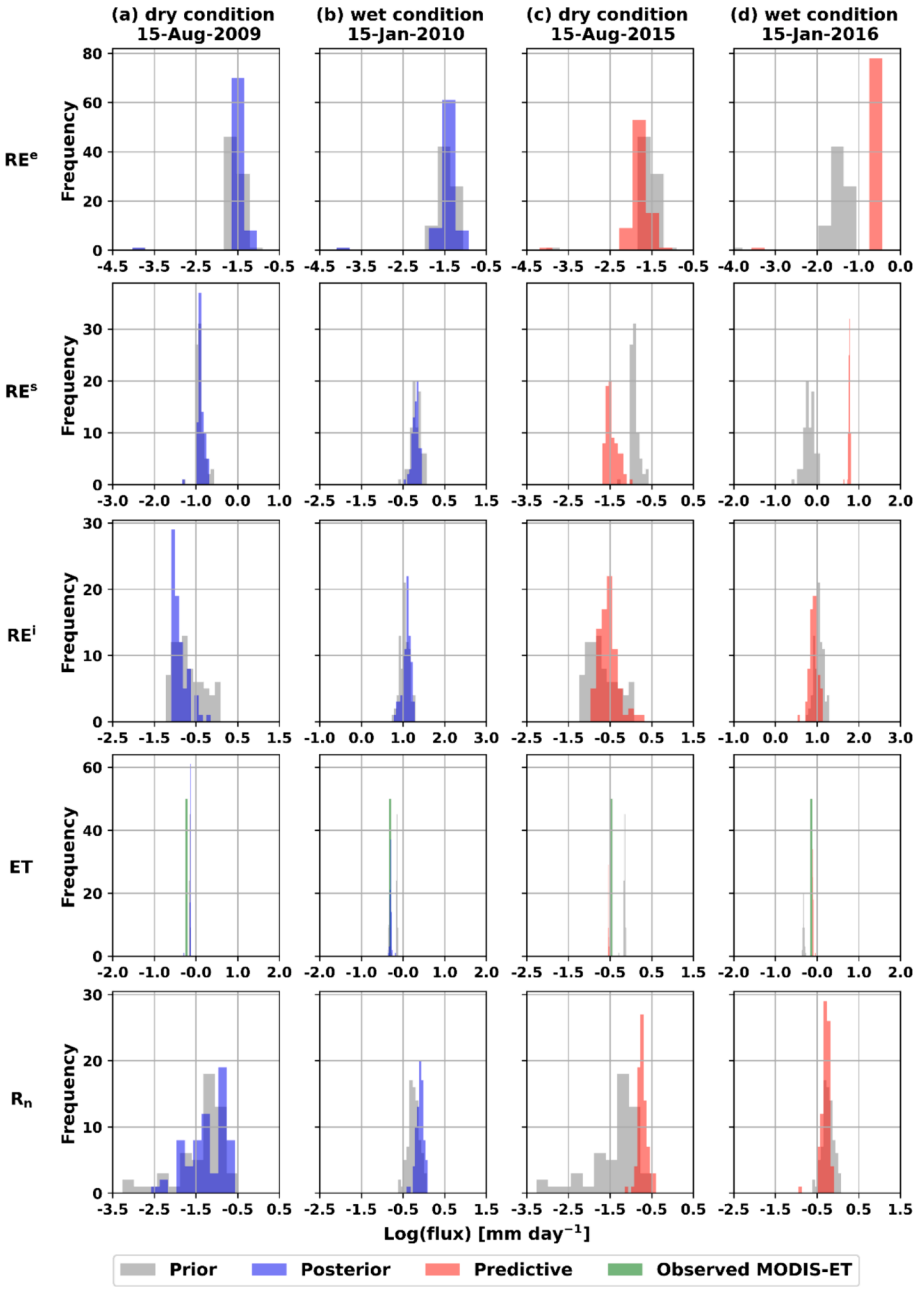

Fig. 8. Uncertainty of five fluxes [the three CRR fluxes (RE^e, RE^s, RE^i), ET and R_n] over the entire catchment in two days during the calibration period representing: (a) dry condition; (b) wet condition; and other two days during the validation period representing: (c) dry condition; and (d) wet condition.

is that due to the overestimated ET, the R_g was reduced and consequently the R_n was also reduced. The negative R_n indicated non-sustainable groundwater resources in the Sardon area, which constrains the possibilities for groundwater abstraction solutions for future agriculture and/or agroforestry activities.

Table 5
Contribution of selected fluxes to the 13-year mean WB (unit is percentage of P) in two different calibrated models: (a) with, and (b) without including the MODIS-ET observations among the calibration dataset.

Symbol	Flux	(a) with MODIS-ET	(b) without MODIS-ET
RE^e	CRR evaporated water	5.5	11.7
RE^i	CRR re-infiltrated water	8.9	8.4
RE^s	CRR direct runoff	39.4	21.3
ET	Evapotranspiration	57.5	72.0
q	Stream outflow	43.0	24.7
R_g	Gross recharge	46.6	29.2
R_n	Net recharge	3.8	−5.0

The adoption of an IHM for DSM, specifically the simulation of additional processes within a catchment, facilitates the inclusion of new and different types of data in the calibration process, and therefore improves the model on which water management decisions are based. An extensive review by Schilling et al. (2019) showed that in many studies, including at least one unconventional observation type (e.g., soil moisture, temperature, tracer concentrations, residence time, and exchange flux such as infiltration, baseflow, or actual evapotranspiration) along with more traditional state observation types, can improve parameter estimation and ultimately reduce the uncertainty of model predictions. Moreover, if the traditional state observations are scarce, using unconventional observations retrieved from satellite RS products can additionally condition uncertain parameters, which in turn could minimize the IHMs' predictive uncertainty (Lubczynski et al., 2024).

One such observation is RS ET. For example, Wambura et al. (2018) used RS ET, as a source of information, to condition the parameters of a SWAT model (Arnold et al., 2012) of the Wami river basin, in Tanzania. In their study, a comparison was made between the performance of the SWAT model in two cases: (a) using only hydrograph state observations, and (b) supplementing hydrograph state observations with the MODIS-ET observations. Incorporation of the additional ET observations led to: (a) further refined parameter values, (b) improved simulation of the mean behavior of ET within the river basin, and (c) reduction of prediction uncertainty. Moreover, Gelsinari et al. (2021) used MODIS-ET in the coupling of the 1D unsaturated SWAP model (Dam et al., 2008; Kroes et al., 2017) with the MODFLOW-2005 groundwater flow model (Harbaugh, 2005). The use of ET observations was justified by the impact of ET on groundwater levels, and subsequently on the net recharge estimates. Thus, constraining the simulated ET with observations of ET improved the simulated net recharge (Doble and Crosbie, 2017; Morway et al., 2023). In addition, Gaur et al. (2022) used two different satellite RS state variables, ET data (Eswar et al., 2017) and soil moisture (SM) data (Tomer et al., 2016) in the MIKE SHE (Graham and Butts, 2005) IHM for the Subarnarekha catchment in Eastern India. The inclusion of these data allowed to compare the observed ET and SM patterns to the simulated patterns, showing improvements in the simulated water balance. Similar conclusions about the benefits of the non-traditional observations were highlighted by Mei et al. (2023). Their analysis emphasized benefits of the use of different combinations of non-traditional observations including SM, ET, and runoff products in addition to traditionally-used observations like streamflow for calibrating a GSFLOW (Markstrom et al., 2008) model near Lake Michigan, USA. Their results showed that the additional observations, especially, SM and ET, resulted in better overall model performance in terms of simulated streamflow and simulated ET.

4.3. Observation noise

The observation noise was taken into consideration in the inverse solution. For the groundwater head observations, the total noise (measurement noise plus structural noise) was assigned a value of 0.2 m. For each observation point (Fig. 2), the time-series of groundwater heads were calculated from the recorded total pressure time-series and the corresponding atmospheric pressure time-series, both measured by the ADAS. The difference between the two is the time-series of the water column fluctuation, further converted to water table depth. The water table depth time-series were subtracted from the altitudes of observation points, surveyed using a geodetic differential GPS, resulting in the time-series of the groundwater heads. Based on the accuracy of the measurements, the measurement noise was estimated as ~0.05 m, while the remaining 0.15 m, was the structural noise (based on the authors' knowledge about the MODFLOW 6 code).

Uncertainty of the MODIS-ET product depends on many factors: meteorological conditions, seasonality, landcover, altitude level, and scaling effects (Velpuri et al., 2013; Long et al., 2014; Jiang and Ryu, 2016; Michel et al., 2016; Moreira et al., 2019; Barraza et al., 2019; Khan et al., 2020; Zhu et al., 2022; Liu et al., 2023). In general, these sources report that MODIS-ET performs better: (a) in semi-humid and semi-arid regions compared to humid or arid climates, (b) in the wet season compared to the dry season, (c) over grasslands and shrubs compared to croplands and forests, (d) at medium altitudes (500–2000 m a.s.l.) compared to high altitudes (>2000 m a.s.l.), and (e) at basin scales compared to point scales. The bias of the MODIS-ET product, based on the literature information cited previously that meets the Sardon catchment conditions [semi-arid grassland located at medium altitudes (700–900 m a.s.l.)], is in the range of 5–20 mm month^{-1}, averaging about 12.5 mm month^{-1}. Given the temporal resolution of the product was every 8 days, an observational noise of 5 mm per 8 days was assigned in the inverse solution.

Streamflow observations were the most uncertain of the observation types for a number of reasons. First, direct streamflow measurements could only be collected during low-flow conditions ($\leq$0.145 m^3 s^{-1}). Second, data was not directly measured during the simulation period but was derived as explained in Hassan et al. (2014) and briefly mentioned in section 2.1. Third, the flume maintenance, completed by a local farmer, was not regularly documented. Finally, significant groundwater flow occurs along the major fault line directly below the surface river channel and goes unmeasured (Lubczynski and Gurwin, 2005).

4.4. IHMs and DSM

Modeling studies that support decisions related to current and future water resource management often rely on simple (standalone) hydrological models as the basis for DSM. Our second objective in this research was to demonstrate the applicability of using IHMs for DSM, which is a recent development. In this application, an inverse model calibration allowed for quantifying the predictive uncertainty using the PESTPP-IES algorithm, implemented in the PEST++ software. The model simulated outputs reasonably matched the observations in both the calibration and the validation periods. Also, the predictive uncertainty was minimized, indicating that the inverse solution had accomplished what was expected.

DSM using IHMs is an intensive task requiring long run times with high computational demand. However, due to ever-increasing computational resources in many research organizations (e.g. high-performance clusters, multiple connected computers and cloud computing platforms) and improved software support for parallelizing the model runs, that complication has been reduced. In this study, the PESTPP-IES inversion solution took ~4–5 days to run on a Linux server with 128 CPUs. In Kitlasten et al. (2021), with the use of the MODSIM-GSFLOW IHM (Morway et al., 2016; Niswonger et al., 2017), also calibrated with PESTPP-IES, the inversion solution took ~5–6 days. So, in both studies, the model run times were considered reasonable.

Most algorithms, used for parameter estimation and uncertainty quantification, are sensitive to the number of adjustable parameters because a Jacobian matrix needs to be filled. Filling the Jacobian matrix is directly related to the number of the conditioned parameters and therefore, needs a number of model runs equal to the number of the conditioned parameters. In contrast, PESTPP-IES is based on an iterative ensemble smoother that uses an empirical Jacobian matrix generated from an ensemble of random parameter values (Chen and Oliver, 2013; White, 2018; White et al., 2020), which removes the relation between number of parameters and number of model runs. This yields more reasonable computational requirements to: (1) convert the Sardon IHM into a highly parameterized model with roughly 500,000 gird-cell-scale-based estimated parameters, and (2) incorporate robust RS observations (e.g. MODIS-ET), next to other observations (groundwater heads and streamflow), which led to a large number of total observations (up to 150,000).

The proposed methods were applied in a relatively small catchment (80 km^2), but it can also be applied to bigger areas using a similarly fine grid discretization as proposed herein, resulting in much larger number of grid cells and consequently larger number of model parameters. The use of an algorithm such as PESTPP-IES for model calibration and uncertainty analysis, would allow for applying the proposed methods also in large areas, since it is insensitive to the number of model parameters. Consequently, it greatly reduces the computational requirements while dealing with large scale catchments with high level of parameterization. The information, obtained from applying such methods, is expected to be the reference in decision-making when government's adaptation to climate change policies is to be implemented at different levels (local, regional, national, etc.) to mitigate impacts of climate change and reduce ecosystem vulnerability.

5. Conclusions

The original Sardon MODFLOW 6 IHM, presented by Daoud et al. (2022), was upgraded by: (a) incorporating additional MODIS-ET observations, and (b) converting the model to a highly parameterized model to improve the implementation of the CRR concept and reduce the parameter uncertainties in order to reduce the predictive uncertainty of the fluxes of interest. Equipped with a large number of observations (~150,000) and parameters (~500,000), the model was calibrated using the PESTPP-IES algorithm, implemented in the PEST++ software. The inverse solution provided a good match between the observations and their simulated equivalents. More importantly, the uncertainty associated with each parameter was reduced from the prior to the posterior ensembles during the calibration process, leading to a reduction in the predictive uncertainty during the validation period. The reduction in parameters' uncertainty also led to an increase in the reliability of the CRR concept by allowing for an accurate partitioning of the CRR fluxes. In addition, the uncertainty in the fluxes of interest (e.g. total evapotranspiration and net recharge) was reduced from the prior to both, the posterior and the predictive ensembles (examples in section 3.4).

This study demonstrates the utility of using highly-parameterized IHMs and non-traditional observations for DSM, while accounting for and reducing model uncertainty. The applied methods significantly improved model predictive capabilities in the Sardon catchment and could therefore be considered for use in other bigger catchments around the world. Using a more complex IHM for DSM better informs decision-making by representing more physical processes (e.g. net recharge and total evapotranspiration) that are not typically available in standalone models. Hence, this information adds value to support decision-making models that inform water resource management.

CRediT authorship contribution statement

Mostafa Gomaa Daoud: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jeremy T. White:** Writing – review & editing, Validation, Methodology, Software. **Eric D. Morway:** Writing – review & editing, Validation, Methodology, Software. **Christiaan van der Tol:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Maciek W. Lubczynski:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Model input files associated with the MODFLOW 6 model described in this manuscript provided at the following location: https://doi.org/10.5066/P94AOSZQ (Daoud et al., 2024).

Acknowledgements

The authors would like to acknowledge Philip Gardner, Sarah Peterson, and Jill Frankforter from the U.S. Geological Survey for their review of the manuscript. We would also like to thank the Editor Huaming Guo, the Associate Editor, and two anonymous Reviewers for their constructive comments, which allowed us to improve the quality of the manuscript.

This research was supported by the faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

AppEEARS Team, 2022. Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). Ver. 3.4. NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA. https://appeears.earthdatacloud.nasa.gov.

Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., Haney, E.B., Neitsch, S.L., 2012. Soil & Water Assessment Tool: Input/Output Documentation Version 2012. https://swat.tamu.edu/media/69296/swat-io-documentation-2012.pdf.

Bakker, M., Post, V., Hughes, J. D., Langevin, C. D., White, J. T., Leaf, A. T., Paulinski, S. R., Bellino, J. C., Morway, E. D., Toews, M. W., Larsen, J. D., Fienen, M. N., Starn, J. J., Brakenhoff, D. A., Bonelli, W. P. (2024). FloPy v3.6.0: U.S. Geological Survey Software Release. https://doi.org/10.5066/F7BK19FH.

Bakker, M., Post, V., Langevin, C.D., Hughes, J.D., White, J.T., Starn, J.J., Fienen, M.N., 2016. Scripting MODFLOW model development using python and FloPy. Groundwater 54 (5), 733–739. https://doi.org/10.1111/gwat.12413.

Barraza, V., Grings, F., Franco, M., Douna, V., Entekhabi, D., Restrepo-Coupe, N., Huete, A., Gassmann, M., Roitberg, E., 2019. Estimation of latent heat flux using satellite land surface temperature and a variational data assimilation scheme over a eucalypt forest savanna in Northern Australia. Agric. For. Meteorol. 268, 341–353. https://doi.org/10.1016/J.AGRFORMET.2019.01.032.

Brunner, P., Therrien, R., Renard, P., Simmons, C.T., Franssen, H.J.H., 2017. Advances in understanding river-groundwater interactions. Rev. Geophys. 55 (3), 818–854. https://doi.org/10.1002/2017RG000556.

Camporese, M., Girotto, M., 2022. Recent advances and opportunities in data assimilation for physics-based hydrological modeling. Frontiers in Water 122. https://doi.org/10.3389/FRWA.2022.948832.

Campos, I., Villodre, J., Carrara, A., Calera, A., 2013. Remote sensing-based soil water balance to estimate Mediterranean holm oak savanna (dehesa) evapotranspiration under water stress conditions. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2013.04.033.

Carpintero, E., Anderson, M. C., Andreu, A., Hain, C., Gao, F., Kustas, W. P., González-Dugo, M. P. (2021). Estimating Evapotranspiration of Mediterranean Oak Savanna at Multiple Temporal and Spatial Resolutions. Implications for Water Resources Management. Remote Sensing 2021, Vol. 13, Page 3701, 13(18), 3701. https://doi.org/10.3390/RS13183701.

Chen, Y., Oliver, D.S., 2013. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput. Geosci. 17 (4), 689–703. https://doi.org/10.1007/S10596-013-9351-5.

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H. J. H., Verhoef, A., Van Loon, A. F., Sulis, M., Abesser, C. (2021). Global Groundwater Modeling and Monitoring: Opportunities and Challenges. Water Resources Research, 57(12), e2020WR029500. https://doi.org/10.1029/2020WR029500.

Dam, J.C. van, Groenendijk, P., Hendriks, R.F.A., Kroes, J.G., 2008. Advances of Modeling Water Flow in Variably Saturated Soils with SWAP. Vadose Zone J. 7 (2), 640–653. https://doi.org/10.2136/VZJ2007.0060.

Daoud, M. G., Morway, E. D., White, J. T., van der Tol, C., Lubczynski, M. W. (2024). MODFLOW 6 models for incorporating remote sensing evapotranspiration in ensemble-based framework to enhance cascade routing and re-infiltration concept in integrated hydrological model applied to support decision making. https://doi.org/10.5066/P94AOSZQ.

Daoud, M.G., Lubczynski, M.W., Zoltan, V., Francés, A.P., 2022. Application of a novel cascade-routing and reinfiltration concept with a Voronoi unstructured grid in MODFLOW 6, for an assessment of surface-water/groundwater interactions in a hard-rock catchment (Sardon, Spain). Hydrgeol. J. 1–27 https://doi.org/10.1007/S10040-021-02430-Z.

Demirel, M. C., Stisen, S., Koch, J. (2018b). SPAEF version 1.0 with histogram match. https://doi.org/10.5281/ZENODO.1158890.

Demirel, M.C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., Stisen, S., 2018a. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. Hydrol. Earth Syst. Sci. 22 (2), 1299–1315. https://doi.org/10.5194/HESS-22-1299-2018.

Doble, R.C., Crosbie, R.S., 2017. Review: Current and emerging methods for catchment-scale modelling of recharge and evapotranspiration from shallow groundwater. Hydrgeol. J. 25 (1), 3–23. https://doi.org/10.1007/S10040-016-1470-3.

Doherty, J., 2015. *Calibration and Uncertainty Analysis for Complex Environmental Models*. Watermark Numerical. Computing.

Doherty, J., Moore, C. (2021). Decision-Support Modelling Viewed through the Lens of Model Complexity. A GMDSI Monograph. https://doi.org/10.25957/p25g-0f58.

Doherty, J., Moore, C., 2020. Decision Support Modeling: Data Assimilation, Uncertainty Quantification, and Strategic Abstraction. Groundwater 58 (3), 327–337. https://doi.org/10.1111/gwat.12969.

Doherty, J., Simmons, C.T., 2013. Groundwater modelling in decision support: reflections on a unified conceptual framework. Hydrgeol. J. 21 (7), 1531–1537. https://doi.org/10.1007/S10040-013-1027-7.
Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res. 46 (5), 5525. https://doi.org/10.1029/2009WR008377.
Doherty, J. (2010). PEST, Model-Independent Parameter Estimation—User Manual: 5th Edition. https://www.epa.gov/sites/default/files/documents/PESTMAN.PDF.
Doherty, J. (2021). PEST Model-Independent Parameter Estimation User Manual Part II: PEST Utility Support Software. https://pesthomepage.org/documentation.
Eswar, R., Sekhar, M., Bhattacharya, B.K., 2017. Comparison of three remote sensing based models for the estimation of latent heat flux over India. Hydrol. Sci. J. 62 (16), 2705–2719. https://doi.org/10.1080/02626667.2017.1404067.
Famiglietti, J.S., Lo, M., Ho, S.L., Bethune, J., Anderson, K.J., Syed, T.H., Swenson, S.C., De Linage, C.R., Rodell, M., 2011. Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophys. Res. Lett. 38 (3) https://doi.org/10.1029/2010GL046442.
Francés, A.P., Lubczynski, M.W., Roy, J., Santos, F.A.M., Mahmoudzadeh Ardekani, M. R., 2014. Hydrogeophysics and remote sensing for the design of hydrogeological conceptual models in hard rocks – Sardón catchment (Spain). J. Appl. Geophys. 110, 63–81. https://doi.org/10.1016/J.JAPPGEO.2014.08.015.
Gash, J.H.C., Lloyd, C.R., Lachaud, G., 1995. Estimating sparse forest rainfall interception with an analytical model. J. Hydrol. 170 (1–4), 79–86. https://doi.org/10.1016/0022-1694(95)02697-N.
Gaur, S., Singh, B., Bandyopadhyay, A., Stisen, S., Singh, R., 2022. Spatial pattern-based performance evaluation and uncertainty analysis of a distributed hydrological model. Hydrol. Process. 36 (5), e14586.
Gelsinari, S., Pauwels, V.R.N., Daly, E., van Dam, J., Uijlenhoet, R., Fewster-Young, N., Doble, R., 2021. Unsaturated zone model complexity for the assimilation of evapotranspiration rates in groundwater modelling. Hydrol. Earth Syst. Sci. 25 (4), 2261–2277. https://doi.org/10.5194/hess-25-2261-2021.
Gleeson, T., Cuthbert, M., Ferguson, G., Perrone, D., 2020. Global groundwater sustainability, resources, and systems in the anthropocene. Annu. Rev. Earth Planet. Sci. 48, 431–463. https://doi.org/10.1146/ANNUREV-EARTH-071719-055251.
González-Dugo, M. P., Chen, X., Andreu, A., Carpintero, E., J. Gómez-Giraldez, P., Carrara, A., Su, Z. (2021). Long-Term water stress and drought assessment of mediterranean oak savanna vegetation using thermal remote sensing. Hydrology and Earth System Sciences, 25(2), 755–768. https://doi.org/10.5194/HESS-25-755-2021.
Graham, D. N., Butts, M. B. (2005). Flexible integrated watershed modeling with mike she. In Watershed models, Eds. V. P. Singh & D. K. Frevert (pp. 245-272 Taylor and Francis). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4e9c3b289c52f7ff21471d7c3e9f357e66378ec9.
Haque, A., Salama, A., Lo, K., Wu, P., 2021. Surface and groundwater interactions: A review of coupling strategies in detailed domain models. Hydrology 8 (1), 35. https://doi.org/10.3390/HYDROLOGY8010035.
Harbaugh, A.W., 2005. MODFLOW-2005: the U.S. Geological Survey modular ground-water model–the ground-water flow process. U.S. Geol. Surv. Tech. Methods 6–A16. https://doi.org/10.3133/TM6A16.
Hassan, S.M.T., Lubczynski, M.W., Niswonger, R.G., Su, Z., 2014. Surface–groundwater interactions in hard rocks in Sardon Catchment of western Spain: An integrated modeling approach. J. Hydrol. 517, 390–410. https://doi.org/10.1016/J.JHYDROL.2014.05.026.
Herrera, P.A., Marazuela, M.A., Hofmann, T., 2022. Parameter estimation and uncertainty analysis in hydrological modeling. Wiley Interdiscip. Rev. Water 9 (1), e1569.
Hugman, R., Doherty, J., 2022. Complex or simple—does a model have to be one or the other? Front. Earth Sci. 10, 705. https://doi.org/10.3389/FEART.2022.867379.
Janus, T., Tomlinson, J., Anghileri, D., Sheffield, J., Kollet, S., Harou, J., 2023. Multicriteria land cover design via coupled hydrologic and multi-sector water management models. J. Hydrol. 620, 129294 https://doi.org/10.1016/J.JHYDROL.2023.129294.
Jiang, C., Ryu, Y., 2016. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). Remote Sens. Environ. 186, 528–547. https://doi.org/10.1016/J.RSE.2016.08.030.
Khan, M.S., Baik, J., Choi, M., 2020. Inter-comparison of evapotranspiration datasets over heterogeneous landscapes across Australia. Adv. Space Res. 66 (3), 533–545. https://doi.org/10.1016/J.ASR.2020.04.037.
Kitlasten, W., Morway, E.D., Niswonger, R.G., Gardner, M., White, J.T., Triana, E., Selkowitz, D., 2021. Integrated hydrology and operations modeling to evaluate climate change impacts in an agricultural valley irrigated with snowmelt runoff. Water Resour. Res. 57 (6), e2020WR027924 https://doi.org/10.1029/2020WR027924.
Koch, J., Demirel, M.C., Stisen, S., 2018. The SPAtial EFficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. Geosci. Model Dev. 11 (5), 1873–1886. https://doi.org/10.5194/GMD-11-1873-2018.
Kroes, J. G., Dam, J. C. van, Bartholomeus, R. P., Groenendijk, P., Heinen, M., Hendriks, R. F. A., Mulder, H. M., Supit, I., Walsum, P. E. V. van. (2017). SWAP version 4. Wageningen Environmental Research. https://doi.org/10.18174/416321.
La Agencia Estatal de Meteorología (AEMET). (2022). https://www.aemet.es/en/portada.
Langevin, C.D., Hughes, J.D., Banta, E.R., Niswonger, R.G., Panday, S., Provost, A.M., 2017. Documentation for the MODFLOW 6 Groundwater Flow Model. *U.S. Geological Survey Techniques and Methods* 6–A55. https://doi.org/10.3133/tm6a55.
Liu, H., Xin, X., Su, Z., Zeng, Y., Lian, T., Li, L., Yu, S., Zhang, H., 2023. Intercomparison and evaluation of ten global ET products at site and basin scales. J. Hydrol. 617, 128887 https://doi.org/10.1016/J.JHYDROL.2022.128887.
Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50 (2), 1131–1151. https://doi.org/10.1002/2013WR014581.
Lubczynski, M.W., Gurwin, J., 2005. Integration of various data sources for transient groundwater modeling with spatio-temporally variable fluxes—Sardon study case, Spain. *J. Hydrol.* 306 (1–4), 71–96. https://doi.org/10.1016/J.JHYDROL.2004.08.038.
Lubczynski, M.W., Leblanc, M., Batelaan, O., 2024. Remote sensing and hydrogeophysics give a new impetus to integrated hydrological models: A review. J. Hydrol. 633, 130901 https://doi.org/10.1016/J.JHYDROL.2024.130901.
Markstrom, S. L., Niswonger, R. G., Regan, R. S., Prudic, D. E., Barlow, P. M. (2008). GSFLOW - Coupled Ground-Water and Surface-Water Flow Model Based on the Integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005). In U.S. Geological Survey Techniques and Methods 6-D1. https://doi.org/10.3133/TM6D1.
Martinsen, G., He, X., Koch, J., Guo, W., Refsgaard, J.C., Stisen, S., 2022. Large-scale hydrological modeling in a multi-objective uncertainty framework – Assessing the potential for managed aquifer recharge in the North China Plain. J. Hydrol.: Reg. Stud. 41, 101097 https://doi.org/10.1016/J.EJRH.2022.101097.
Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H., Eberts, S., Niswonger, R., Regan, R. S., Hunt, R. J. (2023). Can hydrological models benefit from using global soil moisture, evapotranspiration, and runoff products as calibration targets? Water Resources Research, e2022WR032064. https://doi.org/10.1029/2022WR032064.
Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., Mccabe, M.F., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project - Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrol. Earth Syst. Sci. 20 (2), 803–822. https://doi.org/10.5194/HESS-20-803-2016.
Moges, E., Demissie, Y., Larsen, L., Yassin, F. (2020). Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. Water 2021, Vol. 13, Page 28, 13(1), 28. https://doi.org/10.3390/W13010028.
Moore, C.R., Doherty, J., 2021. Exploring the adequacy of steady-state-only calibration. Front. Earth Sci. 9, 777. https://doi.org/10.1007/S10457-014-9785-3.
Moreira, A. A., Ruhoff, A. L., Roberti, D. R., Souza, V. de A., da Rocha, H. R., de Paiva, R. C. D. (2019). Assessment of terrestrial water balance using remote sensing data in South America. J. Hydrol., 575, 131–147. https://doi.org/10.1016/J.JHYDROL.2019.05.021.
Moreno, G., Cáceres, Y. (2016). System report: Iberian Dehesas, Spain; AGFORWARD; Agroforestry for Europe: https://www.agforward.eu/documents/WP2_ES_Dehesa_system_description.pdf.
Morway, E. D., Buto, S. G., Niswonger, R. G., Huntington, J. L. (2023). Assessing potential effects of changes in water use in the middle Carson River Basin with a numerical groundwater-flow model, Eagle, Dayton, and Churchill Valleys, west-central Nevada. In U.S. Geological Survey Scientific Investigations Report 2023–5008. https://doi.org/10.3133/SIR20235008.
Morway, E.D., Gates, T.K., Niswonger, R.G., 2013. Appraising options to reduce shallow groundwater tables and enhance flow conditions over regional scales in an irrigated alluvial aquifer system. J. Hydrol. 495, 216–237. https://doi.org/10.1016/J.JHYDROL.2013.04.047.
Morway, E.D., Niswonger, R.G., Triana, E., 2016. Toward improved simulation of river operations through integration with a hydrologic model. Environ. Model. Softw. 82, 255–274. https://doi.org/10.1016/J.ENVSOFT.2016.04.018.
Morway, E.D., Langevin, C.D., Hughes, J.D., 2021. Use of the MODFLOW 6 Water Mover Package to Represent Natural and Managed Hydrologic Connections. Groundwater 59 (6), 913–924. https://doi.org/10.1111/GWAT.13117.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800. https://doi.org/10.1016/J.RSE.2011.02.019.
Niswonger, R. G., Prudic, D. E., Regan, R. S. (2006). Documentation of the Unsaturated-Zone Flow (UZF1) Package for modeling Unsaturated Flow Between the Land Surface and the Water Table with MODFLOW-2005. In U.S. Geological Survey Techniques and Methods 6-A19. https://doi.org/10.3133/TM6A19.
Niswonger, R.G., Morway, E.D., Triana, E., Huntington, J.L., 2017. Managed aquifer recharge through off-season irrigation in agricultural regions. Water Resour. Res. 53 (8), 6970–6992. https://doi.org/10.1002/2017WR020458.
Ntona, M.M., Busico, G., Mastrocicco, M., Kazakis, N., 2022. Modeling groundwater and surface water interaction: An overview of current status and future challenges. Sci. Total Environ. 846, 157355 https://doi.org/10.1016/J.SCITOTENV.2022.157355.
Oliver, D. S., Reynolds, A. C., Liu, N. (2008). Inverse theory for petroleum reservoir characterization and history matching (Vol. 9780521881). Cambridge University Press. https://doi.org/10.1017/CBO9780511535642.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Machine Learning Research 12 (85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html.
Poeter, E. P., Hill, M. C., Banta, E. R., Mehl, S., Christensen, S. (2005). UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation constructed using the JUPITER API. In U.S. Geological Survey Techniques and Methods 6-A11. https://doi.org/10.3133/TM6A11.
Poeter, E. P., Hill, M. C., Lu, D., Tiedeman, C. R., Mehl, S. W. (2014). UCODE_2014, with new capabilities to define parameters unique to predictions, calculate w eights using simulated values, estimate parameters with SVD, evaluate uncertainty with MCMC,

and more. https://geology.mines.edu/wp-content/uploads/sites/117/2018/11/UCODE_2014_User_Manual-version02.pdf.
Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – A framework and guidance. Environ. Model. Softw. 22 (11), 1543–1556. https://doi.org/10.1016/J.ENVSOFT.2007.02.004.
Running, S. W., Mu, Q., Zhao, M., Moreno, A. (2021). MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500 m SIN Grid V061 [Data set]. In NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD16A2GF.061.
Sales-Baptista, E., D'Abreu, M.C., Ferraz-de-Oliveira, M.I., 2016. Overgrazing in the Montado? The need for monitoring grazing pressure at paddock scale. Agrofor. Syst. 90 (1), 57–68. https://doi.org/10.1007/S10457-014-9785-3.
Schilling, O.S., Cook, P.G., Brunner, P., 2019. Beyond Classical Observations in Hydrogeology: The Advantages of Including Exchange Flux, Temperature, Tracer Concentration, Residence Time, and Soil Moisture Observations in Groundwater Model Calibration. Rev. Geophys. 57 (1), 146–182. https://doi.org/10.1029/2018RG000619.
Sen, Z., 2016. Sampling and Deterministic Modeling Methods. In: Spatial Modeling Principles in Earth Sciences. Springer, Cham, pp. 25–96. https://doi.org/10.1007/978-3-319-41758-5_2.
Swain, M.J., Ballard, D.H., 1991. Color indexing. Int. J. Comput. Vis. 7 (1), 11–32. https://doi.org/10.1007/BF00130487.
Tomer, S.K., Bitar, A.A., Sekhar, M., Zribi, M., Bandyopadhyay, S., Kerr, Y., 2016. MAPSM: A spatio-temporal algorithm for merging soil moisture from active and passive microwave remote sensing. Remote Sens. (Basel) 8 (12), 990. https://doi.org/10.3390/RS8120990.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49. https://doi.org/10.1016/J.RSE.2013.07.013.
Wambura, F.J., Dietrich, O., Lischeid, G., 2018. Improving a distributed hydrological model using evapotranspiration-related boundary conditions as additional constraints in a data-scarce river basin. Hydrol. Process. 32 (6), 759–775. https://doi.org/10.1002/HYP.11453.
White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. Environ. Model. Softw. 109, 191–201. https://doi.org/10.1016/J.ENVSOFT.2018.06.009.
White, J.T., Fienen, M.N., Doherty, J.E., 2016. A python framework for environmental model uncertainty analysis. Environ. Model. Softw. 85, 217–228. https://doi.org/10.1016/j.envsoft.2016.08.017.
White, J.T., Knowling, M.J., Moore, C.R., 2019. Consequences of groundwater-model vertical discretization in risk-based decision-making. Groundwater 58 (5), 695–709. https://doi.org/10.1111/GWAT.12957.
White, J.T., Hunt, R.J., Fienen, M.N., Doherty, J.E., 2020. Approaches to highly parameterized inversion: PEST++ Version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis. *Tech. Methods*. https://doi.org/10.3133/TM7C26.
Yang, X., Hu, J., Ma, R., Sun, Z., 2021. Integrated hydrologic modelling of groundwater-surface water interactions in cold regions. Front. Earth Sci. 9 https://doi.org/10.3389/FEART.2021.721009.
Zhu, W., Tian, S., Wei, J., Jia, S., Song, Z., 2022. Multi-scale evaluation of global evapotranspiration products derived from remote sensing images: Accuracy and uncertainty. J. Hydrol. 611, 127982 https://doi.org/10.1016/J.JHYDROL.2022.127982.