Mathematisch-Naturwissenschaftliche Fakultät

Carlos Antonio Fernandez-Palomino | Fred F. Hattermann | Valentina Krysanova | Fiorella Vega-Jacome | Axel Bronstert

Towards a more consistent eco-hydrological modelling through multi-objective calibration

Journal article | Version of record

Terms of use:
This work is licensed under a Creative Commons License. This does not apply to quoted content from other authors. To view a copy of this license visit https://creativecommons.org/licenses/by/4.0/.

Towards a more consistent eco-hydrological modelling through multi-objective calibration: a case study in the Andean Vilcanota River basin, Peru

Carlos Antonio Fernandez-Palomino [a,b], Fred F. Hattermann[a], Valentina Krysanova[a], Fiorella Vega-Jácome [c] and Axel Bronstert [b,d]

aResearch Department II – Climate Resilience, Potsdam Institute for Climate Impact Research, Potsdam, Germany; bInstitute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany; cHidrología – Estudios e Investigaciones Hidrológicas, Servicio Nacional de Meteorología e Hidrología del Perú, LIMA, Peru; dWater Science Group, Potsdam Institute for Climate Impact Research, Potsdam, Germany

ABSTRACT

Most hydrological studies rely on a model calibrated using discharge alone. However, judging the model reliability based on such calibration is problematic, as it does not guarantee the correct representation of internal hydrological processes. This study aims (a) to develop a comprehensive multi-objective calibration framework using remote sensing vegetation data and hydrological signatures (flow duration curve – FDC, and baseflow index) in addition to discharge, and (b) to apply this framework for calibration of the Soil and Water Assessment Tool (SWAT) in a typical Andean catchment. Overall, our calibration approach outperformed traditional discharge-based and FDC signature-based calibration strategies in terms of vegetation, streamflow, and flow partitioning simulation. New hydrological insights for the region are the following: baseflow is the main component of the streamflow sustaining the long dry-season flow, and pasture areas offer higher water yield and baseflow than other land-cover types. The proposed approach could be used in other data-scarce regions with complex topography.

1 Introduction

The understanding of hydrological processes and vegetation dynamics within a basin is crucial for better water resources management. For this purpose, hydrological models that integrate hydrological processes, vegetation, and biogeochemical cycles (carbon, nitrogen, and phosphorus) have been used, often called eco-hydrological river basin models (Krysanova and Arnold 2008). Examples of well-supported models are the Soil and Water Assessment Tool (SWAT; Arnold *et al.* 1998), Soil and Water Integrated Model (SWIM) (Krysanova *et al.* 1998), Variable Infiltration Capacity (VIC) (Liang *et al.* 1994), and Hydrological Predictions for the Environment (HYPE) (Lindström *et al.* 2010). In this study, SWAT is applied since it is internationally accepted as a robust tool for interdisciplinary modelling of basin water resources (Gassman *et al.* 2007, Abbaspour *et al.*, 2017) and ecosystem services (Francesconi *et al.* 2016). It has been applied and evaluated in diverse climates from arid and semi-arid regions (Brouziyne *et al.* 2017, Jajarmizadeh *et al.* 2017) to wet and tropical areas (Strauch and Volk 2013, Alemayehu *et al.* 2017).

Regardless of the choice of hydrologic model, one primary task in any hydrological modelling is the determination of model parameters during the model calibration procedure, owing to the mismatch between model complexity and available data (Devak and Dhanya 2017, Razmkhah *et al.* 2017). Estimation of model parameters is commonly performed using manual and automatic calibration approaches, with discharge-related measures (e.g. Nash-Sutcliffe efficiency, NSE) most commonly used as the objective function because discharge at the basin outlet integrates all hydrological processes upstream. However, it has been argued that model calibration based solely on discharge does not guarantee the credibility of a hydrological model since the water balance components can be misrepresented despite the performance statistics being accurate (Hattermann *et al.* 2005, Pokhrel and Yilmaz 2012, Guse *et al.* 2016, Pfannerstill *et al.* 2017, Larabi *et al.* 2018, Acero Triana *et al.* 2019). Recognizing this deficiency, other strategies have been focused on improving the calibration to better represent hydrological processes and system dynamics. For this, hydrological signatures mostly derived from streamflow time series, e.g. the flow duration curve (FDC), have been used. The application of signatures related to FDC provides more information about the hydrological behaviours of the modelled basin (Hrachowitz *et al.* 2014) and their underlying processes (Yilmaz *et al.* 2008, Gupta *et al.* 2009). FDC has often been used for model evaluation (Yilmaz *et al.* 2008, Pokhrel and Yilmaz 2012, Hrachowitz *et al.* 2014, Pfannerstill *et al.* 2014, 2017) and lately as an objective in model calibration (Shafii and Tolson 2015, Chilkoti *et al.* 2018, Sahraei *et al.* 2020). These studies have demonstrated that signature-based calibration approaches (using discharge and FDC) lead to a more accurate discharge simulation and the reduction of predictive uncertainty. However, Shafii *et al.* (2017) argued that these approaches do not necessarily guarantee correct flow partitioning among the different flowpaths, which is critical when modelling hydrology-related processes like solute

transport, erosion, surface runoff, and baseflow contribution. Therefore, to properly reproduce flow partitioning, the inclusion of another hydrological signature such as the baseflow index (the ratio of long-term mean base flow to total streamflow) in the multi-objective calibration framework is proposed in this study. The baseflow index represents the baseflow component of streamflow, which is critical for regulating seasonal distribution of river flows, and is associated with climatic and physiographic (e.g. soil type, geology, topography, and vegetation) characteristics of the basin (Beck *et al.* 2013, Mohammed and Scholz 2018, Singh *et al.* 2019). Baseflow, or the baseflow index, is crucial to develop appropriate water resources management strategies, such as aquatic ecosystem preservation, hydropower generation, and low-flow forecasting (Beck *et al.* 2013, Singh *et al.* 2019). Previously mentioned and proposed calibration approaches have been focused mainly on the identification of physical parameters related to streamflow, evapotranspiration, and flow components. However, additional identification of vegetation parameters is crucial for models integrating vegetation dynamics. Indeed, leaf area index (LAI) is a key driver of the water balance of a landscape and is considered in SWAT for subsequent estimation of other processes, such as evapotranspiration, biomass accumulation, sediment, baseflow, and surface runoff (Ma *et al.* 2019). Only a few hydrological studies in general and SWAT-related publications in particular have considered the combined model calibration of LAI and streamflow dynamics and proved that this calibration leads to improved streamflow and evapotranspiration simulations (Strauch and Volk 2013, Alemayehu *et al.* 2017, Ha *et al.* 2018, Rajib *et al.* 2018). These studies showed also the utility of remotely sensed LAI data to calibrate the SWAT LAI-related parameters (i.e. plant parameters) in data-scarce basins. It is important to mention that the SWAT LAI estimation is based on heat units (Neitsch *et al.* 2011), and for that, the total number of heat units needed to bring the plant to maturity (PHU_PLT) must be estimated by the user for the vegetation growth simulation. Previous studies have used satellite-based LAI to identify SWAT plant parameters considering a constant PHU_PLT value for each plant type throughout the basin. However, this could be critical, especially in basins with complex topography – in particular high altitudinal differences – since the air temperature (biophysical variable) that controls the PHU_PLT (Neitsch *et al.* 2011) depends on the altitude. To overcome this, we used the satellite-based LAI data to identify not only plant parameters but also PHU_PLT for each plant in each hydrological response unit (HRU). Moreover, we investigated the relationship between the elevation and PHU_PLT, which could be useful for application in other data-scarce regions with complex topography.

In general, some previous studies have reported the benefits of incorporating hydrological signatures (e.g. FDC signatures) and remote sensing data (e.g. LAI) in the calibration of hydrological models (Shafii and Tolson 2015, Chilkoti *et al.* 2018, Ha *et al.* 2018, Rajib *et al.* 2018, Sahraei *et al.* 2020). However, to the best of our knowledge, no previous study has taken into account both datasets for hydrological model calibration. We consider that inclusion of these variables, in addition to streamflow, in the calibration can improve the model reliability in representing the hydrological system. Hence, one of the objectives of this study is to develop a multi-objective calibration framework that exploits the benefits of using both hydrological signatures and satellite-based LAI data for eco-hydrological models. This study is the first of its kind considering these benefits for a more realistic hydrological modelling not only of streamflow but also of vegetation dynamics and flow partitioning. As such, we hope to contribute to hydrological modelling science with a new way of understanding the eco-hydrological processes of basins with complex topography for efficient water resource management.

We conducted our study in the tropical Andes of Peru where there is a paucity of research. Most existing studies related to Andes hydrology have focused mainly in Andean basins dominated by Páramo ecosystems, which span the Andean region of Venezuela, Colombia, Ecuador, and northern Peru (e.g. Buytaert *et al.* 2007, Guzmán *et al.* 2015, Mosquera *et al.* 2015, Hill *et al.* 2018, Carrillo-Rojas *et al.* 2019). Only a few studies have focused on the hydrology of Peruvian Andean basins, and most of these in small catchments dominated by a glacier (Somers *et al.* 2019), forest (Clark *et al.* 2014), páramo, and puna biome (Ochoa-Tocachi *et al.* 2016). Therefore, another objective of this study is to contribute to the basic understanding of hydrological processes of the tropical Andes of Peru. To this end, we selected the Vilcanota River basin (VRB), which is dominated by land uses such as pasture of the puna biome, forest, and agriculture, to better understand its hydrology (e.g. water budget) and to analyse the hydrological services offered by each land-use type regarding the water yield and baseflow.

For this purpose, our proposed novel multi-objective calibration framework for eco-hydrological models such as SWAT, and for basins with complex terrain such as the VRB, consists of a step-wise calibration scheme. First, SWAT LAI-related parameters are calibrated using satellite LAI data. In the second step, parameters related to streamflow, evapotranspiration, and flow components are calibrated. For the latter step, we propose the inclusion of the baseflow index in addition to discharge-related performance metrics and signatures based on FDC within a multi-objective calibration approach to better constrain the flow partitioning during the calibration process. This approach is compared to conventional discharge-based and signature-based calibration approaches in order to test the model's ability to simulate vegetation dynamics, streamflow, and flow partitioning in the Andean VRB. Furthermore, we address how streamflow calibration strategies impact parameter identifiability and equifinality.

2 Materials and methods

2.1 Study area

The VRB is located in the southern Andes of Peru in the Cuzco region (Fig. 1). Its drainage area is 9617 km^2, and the topography within the basin is characterized by a terrain with steep slopes and elevations that range from 2136 to 6301 m a.s.l. Predominant soils are Lithosols (67%) and Kastanozems (23%) (FAO-UNESCO 1988). The land use is dominated by natural pasture (68%) with a minor contribution of mixed forest
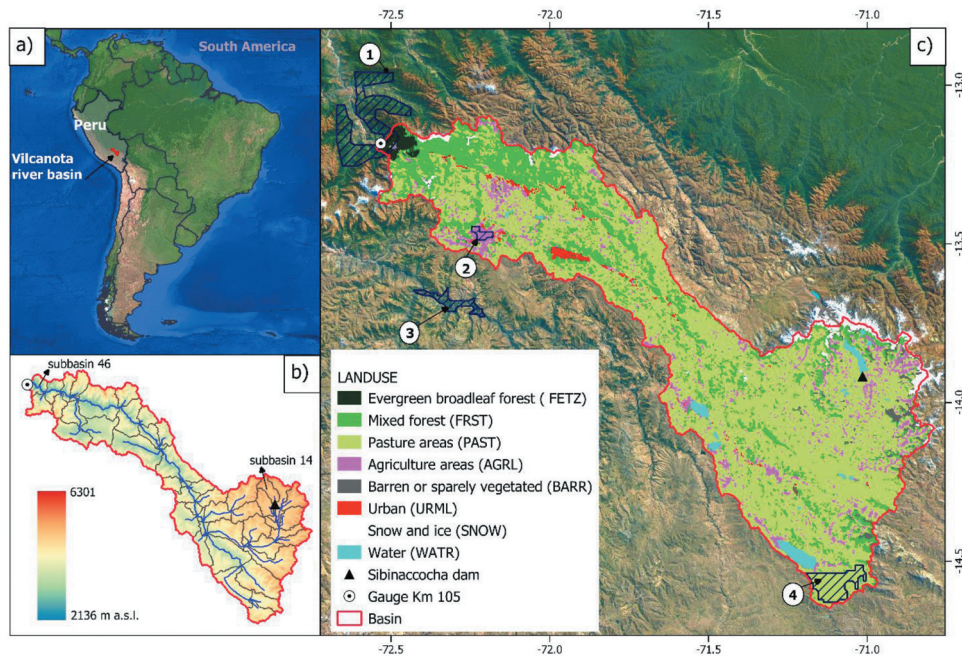
Figure 1. (a) Location of the Vilcanota River Basin (VRB); (b) Terrain elevation, sub-basins, and river network ; and (c) land-use map. Dashed areas represent sample site locations for the major vegetation classes (1: FETZ, 2: AGRL, 3: FRST and 4:PAST) which were used to mask the MODIS (Moderate Resolution Imaging Spectroradiometer) LAI.

(15%), agriculture (8%) and evergreen forest (4%). The latter mainly spans the area which is close to the basin outlet (Fig. 1) and experiences higher amounts of precipitation. The annual mean precipitation is 748.5 mm/year (1985–2015), with more than 80% of the rainfall occurring during the rainy season (October–March). Hydrologically, VRB shows high variability of daily discharge, ranging from 20 m³/s in the dry season to 1100 m³/s in the rainy season, and an average daily discharge of 120 m³/s (1985–2015 period). In VRB there are several natural lakes, but since 1996, Lake Sibinaccocha (Fig. 1) has been dammed for water storage during the wet season and to supply water to the Machu Picchu Hydroelectric Power Plant (EGEMSA company) during the dry season (Catacora-Acevedo 2008). This dam has 120 hm³ active storage volume with a small sub-drainage basin surface (137 km² being 1.4% of the VRB area). The Vilcanota River provides water for drinking, irrigation, and energy production, hence understanding the hydrologic system and accurate simulation of streamflow are important for appropriate water resources management.

2.2 SWAT model

The SWAT model is a process-oriented, semi-distributed and time-continuous river basin model used to simulate hydrological processes as well as vegetation dynamics, nutrients, pesticides, and sediment loads within a basin (Arnold et al. 1998, Neitsch et al. 2011). It is also possible to include water management activities such as reservoirs for hydrological simulation of managed basins (Neitsch et al. 2011). SWAT divides a basin into sub-basins, which are then further subdivided into HRUs representing unique combinations of land use, soil type, and slope classes (Neitsch et al. 2011). SWAT differentiates between the land phase that controls the water, sediment,

and solute loads to the main channel in each sub-basin and the routing phase that defines water, solute, and sediment, movement through the channel network to the basin outlet (Arnold et al. 1998, Neitsch et al. 2011). The water balance computation is performed at the HRU level considering four water storage types (snow, soil profile, and shallow and deep aquifers), as follows:

$$\Delta S = \sum_{i=1}^{N}(P - \text{WYLD} - \text{ET} - \text{losses}) \tag{1}$$

where ΔS is the change in water storage (mm); N is the time in days; and P, WYLD, ET, and losses are the amounts of precipitation (mm), water yield (mm), evapotranspiration (mm), and groundwater losses (mm), respectively. Water yield (WYLD) is given by the contribution of surface runoff (Qsurf), lateral flow (Qlat), and return flow (Qgw) from the aquifers (shallow and deep).

In SWAT, the vegetation dynamics (e.g. LAI) are simulated based on the simplified version of the Environmental Policy Integrated Climate (EPIC) plant growth model (Neitsch et al. 2011). The vegetation growth is simulated based on daily cumulative heat units (plant heat requirements) reflecting the fact that plant growth only occurs on the days when daily mean temperature exceeds the base temperature for growth (Neitsch et al. 2011). This means that temperature is the main vegetation growth controlling factor in SWAT but is restricted by temperature, water, and nitrogen or phosphorous stress (Neitsch et al. 2011). For a detailed description of the SWAT vegetation module, readers are advised to read Neitsch et al. (2011).

In this study, we used SWAT-T (Alemayehu et al. 2017), a modified SWAT version based on SWAT 2012 (Rev. 627), which provides an improved vegetation growth module for a

better simulation of plant growth dynamics in tropical regions. SWAT-T uses the soil moisture index – a quotient of rainfall and potential evapotranspiration – as an indicator to initiate a new growth cycle within a predefined period as the months when the rainy season starts, e.g. October to November for Andean basins. This SWAT-T feature was introduced to overcome SWAT's shortcomings in simulating the seasonal growth cycles for trees and perennials in the tropics, where rainfall rather than temperature is the dominant plant growth controlling factor (Strauch and Volk 2013, Alemayehu *et al.* 2017). Moreover, SWAT-T uses a logistic function to simulate the LAI curve during the senescence stage (Strauch and Volk 2013), instead of the linear decreasing LAI curve which could underestimate evapotranspiration (Wei *et al.* 2018). SWAT-T is referred to as SWAT in this paper.

2.3 Input data

The inputs (e.g. topography, land use, soil, and meteorology) and their sources are summarized in Table 1. As geographical input data, a digital elevation model (DEM) of 90 m resolution, a land-use map obtained from European Space Agency and Climate Change Initiative – Land Cover Project (ESA CCI-LC), and a soil map from the Harmonized World Soil Database (HWSD) that includes soil properties were used for the hydrological model. The daily gridded PISCO (Peruvian Interpolated data of SENAMHI's Climatological and Hydrological Observations) meteorological forcing data

Table 1. Data type, resolution, and data source.

Data type	Resolution	Description/source
Topography	90 m	Digital elevation of the Shuttle Radar Topography Mission (SRTM v. 4.1) product (Jarvis *et al.* 2008) (http://srtm.csi.cgiar.org/)
Land use	300 m	Land-use classification representative for the year 2010 obtained from ESA CCI-LC (http://maps.elie.ucl.ac.be/CCI/viewer/)
Soil	1000 m	Horizon-specific soil properties for each soil type based on the HWSD (Abbaspour and Ashraf Vaghefi 2019)
Soil thickness	1000 m	Gridded global data of soil thickness (Pelletier *et al.* 2016) used to implement variable soil thicknesses at HRUs
Hydrologic soil group	250 m	Global gridded hydrologic soil group data (Ross *et al.* 2018) used to update the curve number (CN) parameter at HRUs after the model creation; this step helped to identify appropriate CN values, mainly in current urban areas
Temperature	Daily/10 km (1981–2016)	Gridded temperature (maximum and minimum) dataset for Peru (PISCO temperature V1.1, Huerta *et al.* 2018) provided by SENAMHI (ftp://publi_dgh2:123456@ftp.senamhi.gob.pe/)
Precipitation	Daily/10 km (1981–2016)	Gridded rainfall dataset for Peru (PISCO precipitation V2.1, Aybar *et al.* 2019) provided by SENAMHI (ftp://publi_dgh2:123456@ftp.senamhi.gob.pe/)
Reservoir	Daily (1996–2015)	Controlled outflow data from Sibinaccocha Dam obtained from Electricity Generation Company of Machupicchu (EGEMSA)
Discharge	Daily (1958–2015)	Flow data at km-105 hydrological station (EGEMSA)
LAI	8 d/1000 m (2000–2016)	Improved MODIS LAI data based on the MODIS collection 5 LAI product (MOD15A2) (Yuan *et al.* 2011) (http://globalchange.bnu.edu.cn/research/lai)

(precipitation, and maximum and minimum temperature) for driving SWAT model simulations was used, as provided by the National Service of Meteorology and Hydrology of Peru (SENAMHI). Controlled outflow data from the Sibinaccocha Dam was used to consider its impact on downstream runoff since 1996.

2.4 Reference data for model calibration and verification

2.4.1 Leaf area index

The satellite-based Moderate Resolution Imaging Spectroradiometer (MODIS) LAI product (Yuan *et al.* 2011) was used as a reference to calibrate LAI dynamics of perennial plants. MODIS LAI has been proven capable of reproducing vegetation timely and accurately (Yuan *et al.* 2011, Ma *et al.* 2019) and has been used successfully in the calibration and/or validation of SWAT plant parameters (e.g. Strauch and Volk 2013, Alemayehu *et al.* 2017, Ha *et al.* 2018, Rajib *et al.* 2018).

To derive the reference LAI data, we selected 200 pixels for pasture, 100 for the mixed forest, 405 for the evergreen forest, and 35 for agriculture, using as a mask the corresponding representative area (polygon) defined in Fig. 1 with the help of the land-use map and Google Earth images. Then, from these subsets, we derived the 8-day median LAI time series for each of the land-use types. Note that the identified representative areas for evergreen forest and mixed forest span areas outside of the basin (Fig. 1). This was necessary to obtain MODIS LAI pixels with better LAI temporal patterns, because forest areas located inside the basin mostly present noisy LAI time series with breaks, which could be attributed to the cloud contamination of the MODIS LAI in those areas.

2.4.2 Streamflow

Daily observed discharge series (1958–2015) at the km-105 gauging station located at the VRB basin outlet was used (Fig. 1) for model calibration and validation of streamflow simulation.

2.5 SWAT model set-up

The SWAT model was set up for the VRB based on the available input data listed in Table 1. The model includes one reservoir, 53 sub-basins, and 320 HRUs (Fig. 1). The modified Soil Conservation Service Curve Number method, the Hargreaves method, and the variable storage method were used to simulate surface runoff and infiltration, potential evapotranspiration, and river flow routing, respectively. The Sibinaccocha Reservoir outflow was simulated using a predefined daily outflow option in SWAT to account for the effects of this dam on downstream runoff. For more details about the SWAT model configuration, see Neitsch *et al.* (2011).

The periods considered for model warm-up, calibration, and validation for simulation of LAI were 2001–2004, 2005–2010, and 2011–2015, and those for streamflow were 1981–1984, 1985–1990, and 1991–2015, respectively. Note that the model calibration for streamflow was carried out in the predamming period of the river, while the LAI calibration and validation periods were constrained based on LAI data availability.

2.6 SWAT calibration and evaluation framework

We propose the following step-wise framework for SWAT model calibration for basins with complex terrain (Fig. 2).

First, SWAT LAI-related parameters are calibrated using 8-day MODIS LAI as a reference for each perennial tropical vegetation type (e.g. pasture, mixed forest, forest evergreen, and agriculture) at the HRU level. To do so,

(a) The simulated LAI is calibrated against MODIS LAI to identify SWAT LAI-related parameters (see Section 3.2) for a specific HRU. This procedure helps to build a SWAT plant database (plant.dat) for the study area, which is used for the next steps.

(b) As the total number of heat units needed to bring the plant to maturity parameter (PHU_PLT) depends on temperature data (Neitsch *et al.* 2011), PHU_PLT is calibrated for each HRU using MODIS LAI as a reference since the temperature varies with height in mountainous regions. This step is important for better modelling of vegetation LAI dynamics in HRUs defined by trees and perennials.

In this study, NSE is used as the objective function for LAI and PHU_PLT calibration.

The second part of the calibration strategy deals with the calibration of the SWAT parameters that mainly control streamflow, evapotranspiration, and discharge components. For this purpose, we propose a multi-objective calibration approach that includes the baseflow index to constrain the model calibration in addition to discharge-related performance metrics and hydrological signatures related to FDC. This approach is compared to approaches applied before, where only hydrograph goodness-of-fit metrics (e.g. NSE or the log NSE, lNSE) and signatures related to FDC were used. Table 2 shows the single- and multi-objective optimization approaches tested in this study, and Table 3 shows the objective metrics. In the formulation of multi-objective calibration approaches, we selected lNSE instead of NSE as part of the objectives for flow timing and magnitude to avoid overfitting discharge peaks, since NSE is

Table 2. Optimization problems. In the optimization process, NSE and lNSE (log NSE) were maximized while the absolute values of FDCsign (FDC signature) and BIAS_BFI (bias of baseflow index) were minimized.

Number	Approach	Formulation	Optimization problem
1	Applied before	Single-objective	NSE
2	Applied before	Single-objective	lNSE
3	Applied before	Bi-objective	lNSE, FDC$_{sign}$
4	Proposed in this study	Multi-objective	lNSE, FDC$_{sign}$, BIAS_BFI

more sensitive to larger errors that often occur in high-flow periods (Krause *et al.* 2005, Gupta *et al.* 2009). To quantify catchment overall behaviour and flashiness, several aspects of FDC were considered to evaluate the model performance in emulating the FDC. Following Chilkoti *et al.* (2018), FDC is divided into four segments, of peak flow volume (0–2%), high flow volume (2–20%), mid-segment slope (20–70%), and low flow volume (70–100%). The respective FDC partitioning represents peak flow events occurring rarely, quick runoff (due to snow-melt and/or rainfall), the quickness of a basin's response, and baseflow components of the streamflow (Yilmaz *et al.* 2008, van Werkhoven *et al.* 2009, McMillan *et al.* 2017, Chilkoti *et al.* 2018). To properly reproduce flow partitioning, the baseflow index is used to constrain the slow-flow component.

For the model evaluation, discharge statistics and individual hydrological signatures defined in Table 3 are used. The conventional hydrograph goodness-of-fit metrics (NSE, lNSE, and Percent bias (PBIAS)) are used to evaluate the model performance in streamflow simulation (Krause *et al.* 2005, Moriasi *et al.* 2007). Note that in this study, PBIAS is not included as an objective function for model calibration but is only used to evaluate model performance. PBIAS measures the average tendency of the simulated discharge, which can be larger or smaller than the measured values. The optimal value of PBIAS is 0.0, with low values indicating accurate model simulation. A positive (negative) value of this measure indicates overestimation (underestimation). FDC signatures (Speak, Shigh, Smid, and Slow) are used to assess the model performance through biases in the flow distributional response (Chilkoti *et al.* 2019). PBIAS_BFI (introduced in this study) is used to evaluate the model capability in the simulation of flow partitioning in terms of the baseflow index. The overall goal of
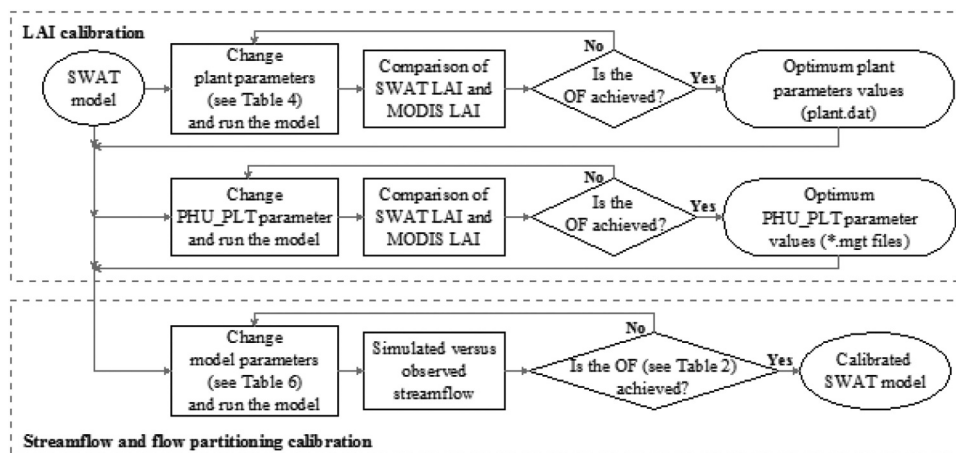
Figure 2. Flowchart of the SWAT model calibration framework. OF: objective function, PHU_PLT: total number of heat units needed to bring the plant to maturity. In the case of LAI calibration, the Nash-Sutcliffe efficiency (NSE) is the OF. The OF for streamflow and flow partitioning calibration is defined in Table 2 for each calibration strategy applied in this study.

Table 3. Mathematical formulation of goodness-of-fit metrics and hydrological signatures. O and S are observed and simulated flow, respectively (in m^3/s); EP is exceedance probability; P, H, and L are the indices of the minimum flow of the peak-flow, high-flow and low-flow segments, respectively.

Criterion (reference)	Equation	Description
Discharge-related performance measures		
Nash-Sutcliffe efficiency (Nash and Sutcliffe 1970)	$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - O_a)^2}$	O_a is the average of the observed flow and n is the number of observations under evaluation
Nash-Sutcliffe efficiency-log (Krause et al. 2005)	$lNSE = 1 - \frac{\sum_{i=1}^{n}(\ln(S_i) - \ln(O_i))^2}{\sum_{i=1}^{n}(\ln(O_i) - \ln(O_p))^2}$	
Percent bias (Gupta et al. 1999)	$PBIAS = \frac{\sum_{i=1}^{n}(S_i - O_i)}{\sum_{i=1}^{n} O_i} \times 100$	
Signature measures		
Percent bias in FDC peak-segment volume (Yilmaz et al. 2008)	$S_{peak} = \frac{\sum_{p=1}^{P}|S_p - O_p| \times 100}{\sum_{p=1}^{P} O_P}$	$p = 1, 2, \ldots, P$ are flow indices located within the FDC peak-flow segment (EP lower than 2%)
Percent bias in FDC high-segment volume (Yilmaz et al. 2008)	$S_{high} = \frac{\sum_{h=1}^{H}|S_h - O_h| \times 100}{\sum_{h=1}^{H} O_h}$	$h = 1, 2, \ldots, H$ are flow indices located within the high-flow segment (2–20% flow EP)
Percent bias in FDC mid-segment slope (van Werkhoven et al. 2009; Yilmaz et al. 2008)	$S_{mid} = \frac{|(S_{m1} - S_{m2}) - (O_{m1} - O_{m2})| \times 100}{(O_{m1} - O_{m2})}$	$m1$ and $m2$ are the lowest and highest flow EP, respectively, within the mid-segment (20–70%)
Percent bias in FDC low-segment volume (Yilmaz et al. 2008)	$S_{low} = \frac{\sum_{l=1}^{L}|S_l - O_l| \times 100}{\sum_{l=1}^{L} O_l}$	$l = 1, 2, \ldots, L$ are flow indices located within the low-flow segment (70–100% flow EP)
FDC signature (Chilkoti et al. 2018)	$FDC_{sign} = \frac{1}{4}\left(S_{peak} + S_{high} + S_{mid} + S_{low}\right)$	FDC_{sign} is the aggregated FDC signature
Bias of baseflow index	$BIAS\_BFI = SWAT\_BFI - BFI$	$SWAT\_BFI$ is the simulated baseflow index and BFI the reference

this multi-criteria evaluation design is to assess how realistically the model represents the hydrologic system response, which is crucial both for hydrological models intended to operate in a predictive mode and for projecting climate change impacts (Krysanova et al. 2018).

2.7 Baseflow index estimation

The reference baseflow index (BFI) was estimated from streamflow using two baseflow separation techniques. The baseflow filter program (BFLOW, http://swat.tamu.edu/) (Arnold and Allen 1999) and the Eckhardt filter using the Web GIS-based Hydrograph Analysis Tool system were applied (WHAT, https://engineering.purdue.edu/mapserve/WHAT/) (Lim et al. 2005). For a detailed description of the filter methods, the reader is referred to Arnold et al. (1995), Arnold and Allen (1999), and Lim et al. (2005). BFLOW and WHAT have been used successfully in many studies related to the SWAT model (e.g. Luo et al. 2012, Meaurio et al. 2015, Yesuf et al. 2016, Jang et al. 2018).

Following Jang et al. (2018), the simulated baseflow index (SWAT_BFI) was computed as follows:

$$SWAT\_BFI = [Qlat + Qgws + Qgwd] / [Qsurf + Qlat + Qgws + Qgwd] \qquad (2)$$

where Qsurf is the surface runoff, Qlat is the lateral flow, Qgws is the return flow from the shallow aquifer, and Qgwd is the return flow from the deep aquifer.

2.8 Multi-objective optimization algorithm

We applied the Borg Multi-Objective Evolutionary Algorithm (Borg MOEA) (Hadka and Reed 2013) to achieve the optimum solutions for SWAT parameters based on the calibration strategies defined in Section 2.6, as Borg MOEA has superior performance

when compared with a range of state-of-the-art multi-objective algorithms (Hadka and Reed 2012, 2013). Moreover, it was applied successfully in the calibration of SWAT in which hydrograph goodness-of-fit metrics and signatures related to FDC were included in the objective function (Chilkoti et al. 2018). The Borg MOEA is an auto-adaptive optimization algorithm that uses a population-based search to find the archived non-dominated solutions (Pareto approximate set) at the end of the optimization. The Borg MOEA parameterization was based on its default recommended parameter values (Hadka and Reed 2013). The initial population size was set to 100, generated based on random parameter sampling. To achiev a reasonable trade-off between objectives, the ε-precision level was set to 0.01 for the NSE family and a difference of 1% for FDC_sign and BIAS_BFI. The total number of objectives for the evaluation was set to 500 for LAI calibration, 500 for PHU_PLT, and 1000 for streamflow. For more details on Borg MOEA theory and features, readers are advised to see Hadka and Reed (2013).

3 Results and discussion

3.1 BFI estimation

Baseflow index estimation (the ratio of long-term mean baseflow to total streamflow) was conducted from daily streamflow data recorded at the km-105 hydrological station for 1964–1990 since this period does not include the potential effects of the Sibinaccocha Dam on runoff (from 1996 onwards). The baseflow indices estimated by BFLOW and WHAT were 0.76 and 0.78, respectively. The mean of these values was considered the reference baseflow index (BFI = 0.77); it means that around 77% of the river discharge at gauge km-105 can be attributed to baseflow. The latter is consistent with the baseflow index reported for the neighbouring Andean Kosñypata basin with similar geology, topography, and vegetation, where 77% of annual flow was attributed to baseflow (Clark et al.

2014). In addition, BFLOW estimated the flow recession constant from the shallow aquifer (ALPHA_BF equal to 0.0351) which was used to replace the default ALPHA_BF value of the model.

3.2 Performance of LAI simulation

Table 4 presents the optimized plant parameter values for mixed forest (FRST), evergreen broadleaf forest (FETZ), pasture (PAST), and agricultural areas (AGRL). Using these optimal parameter values, the PHU_PLT (head unit) parameter was calibrated for all HRUs covered by perennial plants so that the SWAT-simulated LAI mimics the MODIS 8-day LAI as close as possible. Note that AGRL was considered perennial in this study to simulate the LAI decline during senescence using a logistic function (a feature of the SWAT version used here), since the AGRL MODIS LAI curve follows a logistic curve instead of a linear curve during the senescence stage (Fig. 3(b,c)). This is in agreement with Wei et al. (2018), who recommended the use of the logistic LAI curve during senescence for agricultural crops.

Figure 3(a) shows the strong negative relationship ($R^2 \geq 0.83$) between calibrated PHU_PLT values and elevation for PAST, AGRL, and FRST. This means that the total number of heat units required for a plant to reach maturity (PHU_PLT) decreases with altitude, and this can be used as a descriptor variable to estimate PHU_PLT in basins showing high elevation gradients. As a result, the plant parameters and the relationship between PHU_PLT and elevation derived in this study can be used in other Andean basins.

Comparing the temporal variation of LAI dynamics (Fig. 3(b)), in general, SWAT-simulated LAI corresponds well with MODIS LAI data during both calibration and validation periods for all perennial plants. This observation is supported by good model performance statistics (NSE ≥ 0.63, $R^2 \geq 0.76$, and PBIAS within reasonable limits ±15%) in simulating LAI dynamics for both evaluation periods as shown in Table 5. In addition, Figs 3(c) and A1 (see Appendix) show that the spatio-temporal variability of the seasonal LAI simulated by SWAT agrees well with MODIS LAI for all perennials. Since LAI in SWAT influences the simulation of hydrological and vegetation processes such as evapotranspiration, biomass accumulation, streamflow, and sediments (Strauch and Volk 2013, Alemayehu et al. 2017, Ha et al. 2018, Rajib et al. 2018, Ma et al. 2019), the good model performance in

LAI simulation found here increases the quality of the simulation of these processes.

Figure 3(c) shows a similar seasonal LAI pattern for PAST, FRST, and AGRL, and it follows the seasonal rainfall pattern of the Andes. However, the onset/end (February–March/October–November) of LAI development for FETZ is delayed regarding the onset/end (October/March) of the rainy season in the rainforest area. This behaviour was also observed in other tropical regions with natural ecosystems (e.g. Alemayehu et al. 2017).

We found in the literature that only a few SWAT-related studies have reported the calibration of plant parameters, and most of these have only considered a constant PHU_PLT value for each plant type. To simulate appropriately the vegetation dynamics and hydrological processes that depend on LAI, however, the calibration of plant parameters is crucial, particularly the parameter that controls plant development such as the PHU_PLT, which varies with altitude in mountain basins, as demonstrated in this study. We believe that our results and proposed LAI calibration strategy can support modellers for a better simulation of vegetation dynamics.

## 3.3 Model performance in streamflow simulation

The Borg MOEA approach (see Section 2.8) was used to calibrate the SWAT model according to the aforementioned calibration strategies (see Table 2) and the parameters obtained are shown in Table 6. These model parameters were chosen to correct the deficiencies of the uncalibrated model (e.g. systematic flow underestimation being higher during low-flow periods, strong simulated peak discharges, etc.), for which the authors' knowledge about basin characteristics, model structure, and how each model parameter influences the hydrological processes was important.

For single-objective, bi-objective, and multi-objective calibration, the Pareto front solution consisting of one, three, and six sets of parameters were obtained, respectively. Figure 4 summarizes the results of different calibration strategies. Furthermore, the hydrographs, the FDCs, and the mean seasonal flow dynamics of the different objective calibrated simulations are shown in Figs 5(a,b), and 6, respectively.

In terms of the temporal variation of discharge dynamics (Fig. 5(a)), FDCs (Fig. 5(b)), and seasonal discharge dynamics

Table 4. Calibrated SWAT plant parameter values for HRUs with perennial plants: mixed forest (FRST), evergreen broadleaf forest (FETZ), pasture (PAST), and agricultural areas (AGRL).

Parameter	Parameter description	Calibrated values			
		PAST	AGRL	FRST	FETZ
BIO_E	Radiation-use efficiency ((kg/ha)/(MJ/m^2))	17.04	13.92	1.10	0.56
BLAI	Maximum potential leaf area index (m^2/m^2)	1.10	2.74	1.70	5.30
FRGRW1	Fraction of PHU corresponding to the first point on the optimal leaf area development curve	0.06	0.07	0.02	0.10
LAIMX1	Fraction of BLAI corresponding to the first point on the optimal leaf area development curve	0.02	0.17	0.10	0.20
FRGRW2	Fraction of PHU corresponding to the second point on the optimal leaf area development curve	0.49	0.38	0.44	0.50
LAIMX2	Fraction of BLAI corresponding to the second point on the optimal leaf area development curve	0.90	0.92	0.98	0.90
DLAI	Fraction of total PHU when leaf area begins to decline	0.48	0.59	0.40	0.48
ALAI_MIN	Minimum leaf area index for plant during dormant period (m^2/m^2)	0.31	0.58	0.32	0.90
T_BASE	Minimum temperature for plant growth (℃)	2.07	3.46	2.00	0.05
T_OPT	Optimal temperature for plant growth (℃)	18.16	10.00	14.50	13.18
PHU_PLT	Total number of heat units needed to bring plant to maturity	Variable for each HRU			
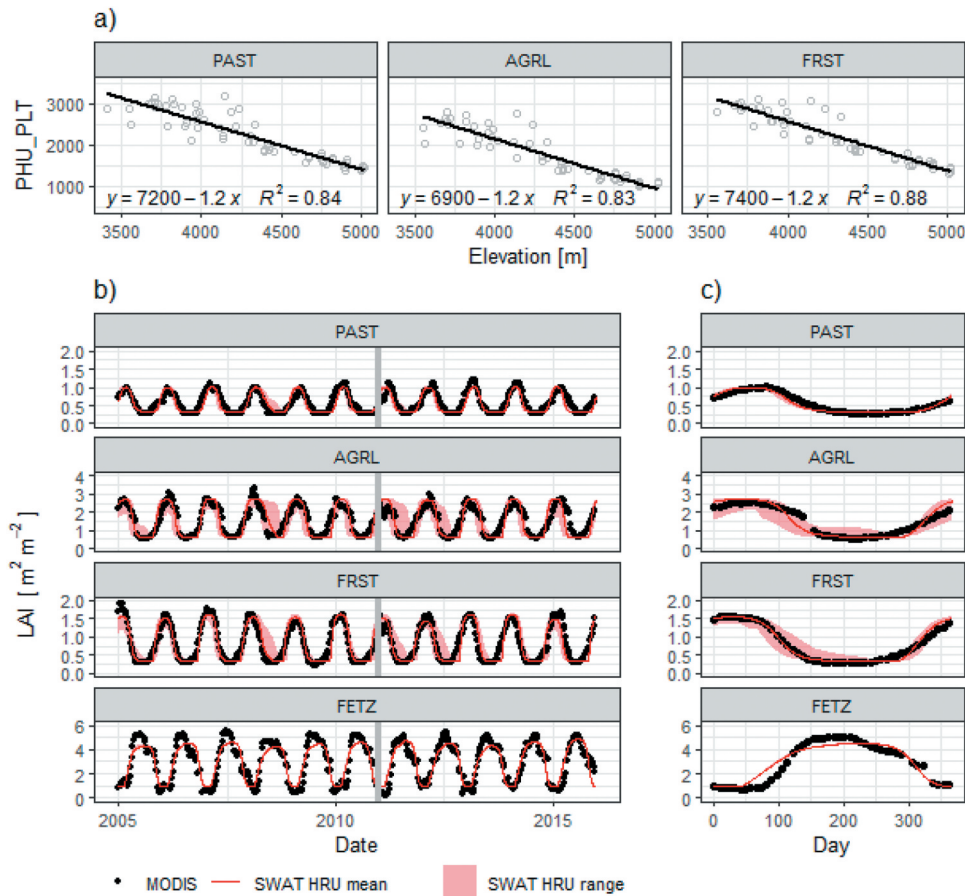
Figure 3. (a) (a) Scatterplot of PHU_PLT against elevation. (b) Simulated daily LAI (range over HRUs and area weighted HRU mean) and the 8-day MODIS LAI. (c) Long-term (2005–2015) average daily (8-day) LAI based on SWAT (MODIS). The vertical grey line marks the end of the calibration period and the beginning of the validation period. R^2 is the coefficient of determination.

Table 5. Performance metrics for the SWAT for simulating LAI in the calibration (validation) period. Note that performance refers to 8-day aggregated data. R^2 is the coefficient of determination.

	PAST	AGRL	FRST	FETZ
NSE	0.78 (0.72)	0.77 (0.63)	0.91 (0.83)	0.81 (0.81)
R^2	0.79 (0.76)	0.81 (0.76)	0.91 (0.85)	0.82 (0.81)
PBIAS %	−0.90 (−6.20)	0.20 (−1.10)	1.90 (0.00)	1.00 (2.20)

(Fig. 6), the NSE calibrated simulation corresponds well with observed daily discharge during high discharge season but fails (flow underestimation) during the low discharge season. This observation is supported by the higher (lower) performance value for NSE (lNSE) as shown in Fig. 4. Moreover, hydrological signatures based on FDC (Fig. 4) show that this calibration approach is primarily focused on the peak, high, and mid flows at the expense of improvements to the low-flow predictions. This finding is congruent with previous studies (Krause *et al.* 2005, Chen *et al.* 2018, Zhang *et al.* 2018b).

Nevertheless, lNSE calibrated simulations match well with observed discharge in all aspects of the hydrograph (Fig. 5) although discharges are moderately underestimated during high-discharge season (Fig. 6). Hence, lNSE is used in this study as part of the multi-objective calibration approaches to drive the model in simulating all hydrograph aspects in the calibration.

The results for bi-objective and multi-objective approaches show that the observed and simulated hydrographs match each

Table 6. Selected parameters and their ranges for model calibration for streamflow. In the "Change type" column, R refers to a relative change of parameter values during the calibration, and V to absolute change. "Adjusted value" refers to the parameter mean values associated with the Pareto set obtained in the multi-objective scenario.

Parameter	Description	Range	Change type	Adjusted value
CN2	Runoff curve number for moisture condition II	[−0.15, 0.15]	R	0.06
SURLAG	Surface runoff delay coefficient	[0.1, 2]	V	0.11
SOL_BD	Wet bulk density	[−0.25, 0.25]	R	0.09
SOL_K	Soil hydraulic conductivity	[−0.25, 0.25]	R	−0.15
SOL_AWC	Available water capacity of the soil layer	[−0.5, 0.25]	R	−0.31
GW_DELAY	Groundwater delay time	[1, 100]	V	42.16
RCHRG_DP	Deep aquifer percolation fraction	[0.05, 1]	V	0.52

other to a high degree (Fig. 5), and all performance measures (goodness of fit and FDC signature metrics; Fig. 4) show that these formulations are superior to the results of single-objective approaches. This finding demonstrates that the inclusion of FDC signatures in addition to discharge-based performance measures within a multi-objective calibration leads to improved discharge simulation, which is in agreement with previous studies (Pokhrel and Yilmaz 2012, Hrachowitz *et al.*
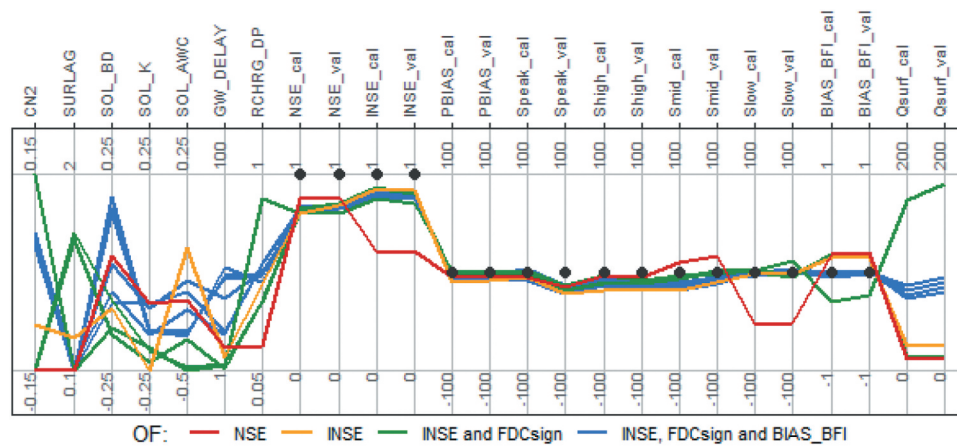
Figure 4. Parallel coordinates plot of the Pareto front optimal solutions obtained by different calibration strategies. For each solution, optimal parameter values, discharge-based performance measures (NSE, INSE, and PBIAS), and hydrological signatures (Speak, Shigh, Smid, Slow, and BIAS_BFI) are displayed. _cal (_val) indicates the measurements for the calibration (validation) period. Qsurf is the mean surface runoff in mm. A description of the SWAT parameters is provided in Table 6. A description of objective functions (OFs) is provided in Table 3.
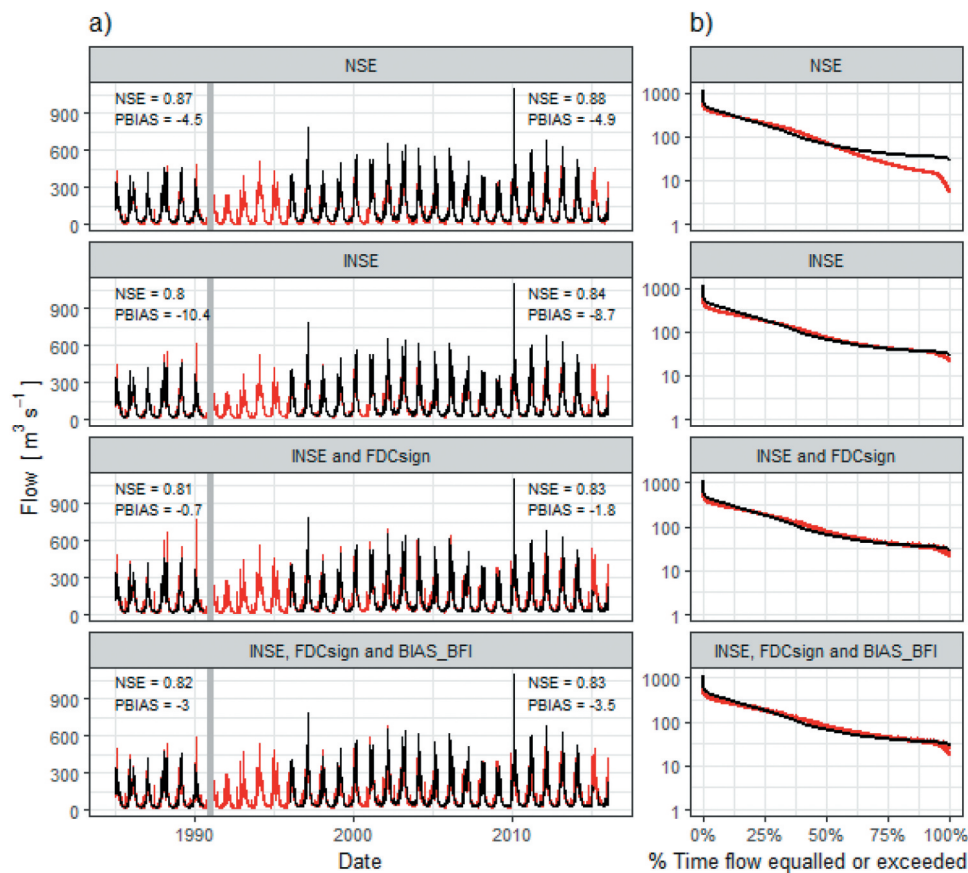

Figure 5. Comparison of the observed (black line) and simulated (red line/area) (a) daily discharges and (b) simulated and observed FDC for four calibration strategies. In the case of the bi-objective (INSE and FDCsign) and multi-objective (INSE, FDCsign, and BIAS_BFI) optimization, the red area indicates optimal Pareto solutions and NSE and PBIAS are the mean of Pareto solutions. The grey line marks the end of the calibration period and the beginning of the validation period.

2014, Shafii and Tolson 2015, Pfannerstill *et al.* 2017, Chilkoti *et al.* 2018, Sahraei *et al.* 2020).

Regarding the model performance in simulating the flow components, Fig. 4 shows the smaller values of PBIAS_BFI for the multi-objective calibration approach, which demonstrates that our approach leads to a more accurate representation of the flow partitioning into surface runoff and baseflow. Whereas approaches applied before (single-objective and bi-objective approaches) fail in simulating the partitioning into flow components according to the baseflow index, this agrees with the findings of Shafii *et al.* (2017), who reported that traditional signature-based
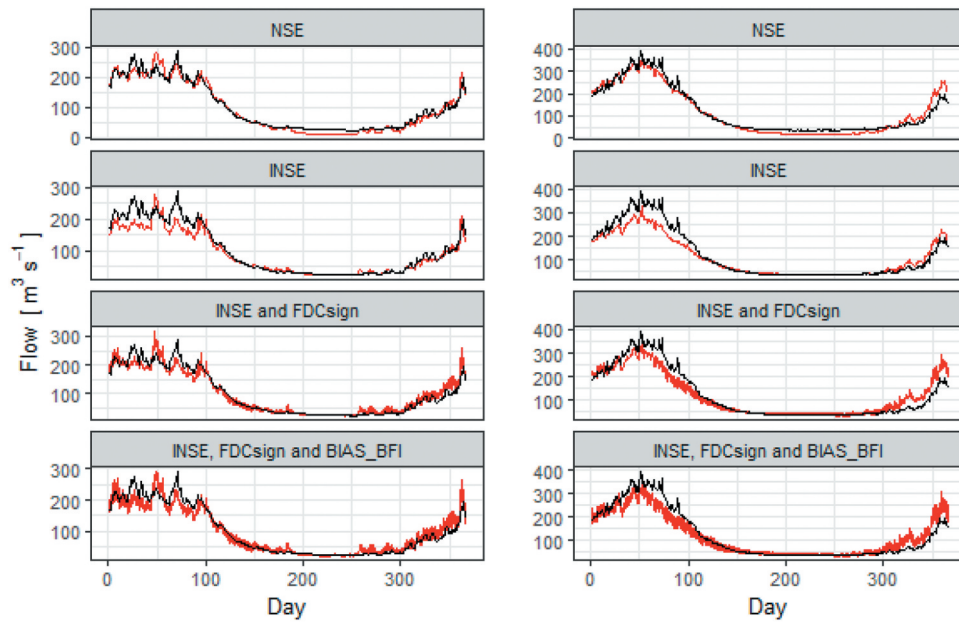
Figure 6. The mean seasonal dynamics of simulated discharge (red line/area) and observed discharge (black line) for each calibration strategy in the calibration period (left) and validation period (right).

calibration (using discharge and FDC) does not necessarily guarantee correct flow partitioning in a Hydrologic Model (HYMOD) hydrology model application.

Considering the rating performance criteria of Moriasi *et al.* (2007), NSE greater than 0.75 and PBIAS less than 10% are indicative of very good model performance for streamflow simulation; therefore, model performance was very good in both calibration (NSE ≥ 0.8, |PBIAS| ≤ 10%) and validation (NSE ≥ 0.8, |PBIAS| ≤ 8.7%) for daily streamflow simulation in all calibration strategies performed in this study. The results of this study (Fig. 4), however, clearly demonstrate that practitioners must be careful judging model credibility using only these discharge-based metrics, since good model performance for streamflow and FDC simulation does not guarantee internal consistency of all simulated processes (e.g. surface runoff and baseflow).

In this study, additional calibration strategies (Appendix Fig. A2) that do not include FDC signatures were performed, and their results suggest that our proposed multi-objective approach is much more robust.

3.4 Parameter identifiability

The identification of physically plausible, representative, and robust model parameter sets for the basin under investigation is an important task in hydrologic modelling (Wagener *et al.* 2001, Shafii and De Smedt 2009). For this purpose, we focused on the range of parameter values associated with the Pareto optimal solutions, as suggested by Gupta *et al.* (1998), and particularly on the parameters retained by bi-objective and multi-objective calibration strategies. A parameter becomes more identifiable when the parameter range is narrower and/ or its optimum values are located in a particular region of the feasible range.

Figure 4 shows that in the bi-objective calibration scenario, four parameters (Wet bulk density (SOL_BD), Soil hydraulic conductivity (SOL_K), Available water capacity of the soil layer (SOL_AWC), and Groundwater delay time (GW_DELAY)) are identifiable and three parameters (Runoff curve number for moisture condition II (CN2), Surface runoff delay coefficient (SURLAG), and Deep aquifer percolation fraction (RCHRG_DP)) are barely or not identifiable, since different values of these parameters give similar results in combination with the other parameters. Otherwise, six parameters (CN2, SURLAG, SOL_K, SOL_AWC, GW_DELAY, and RCHRG_DP) – influencing main hydrologic processes such as surface runoff, lateral flow, evapotranspiration, and return flow from aquifers – are well identifiable, and only the parameter SOL_BD is hardly identifiable in the multi-objective calibration strategy. The larger number of identifiable parameters in the latter approach is due to the inclusion of the baseflow index in the multi-objective calibration, which guides the optimization algorithm to identify parameters related to processes that impact the baseflow index. Therefore, when more information (objectives) is fed (required) into model calibration, the number of identifiable parameters will also increase.

Comparing the robustness of parameter sets obtained by each calibration strategy, the parameter sets of models calibrated based on single-objective and bi-objective optimization performed satisfactorily in simulating streamflow but for the unrealistic representation of surface runoff (Fig. 4) as well as of the basin baseflow index. On the contrary, the parameter sets obtained by a multi-objective strategy led to appropriate representation of the baseflow index as well as streamflow and FDC simulation, and hence these parameter sets can be catalogued as representative for the study basin. The superiority of the multi-objective strategy in simulating flow partitioning is related to the better identification of the CN2 parameter,

which directly impacts surface runoff and infiltration partition and alters water balance components (Arnold *et al.* 2012, Qi *et al.* 2020). Here, only representative parameters are described, and the narrower ranges of CN2, SURLAG, and RCHRG_DP obtained by the multi-objective approach indicate that the basin response is very sensitive to surface runoff and deep aquifer contribution. The obtained optimum values for CN2 (Fig. 4) must lead to the increase of surface runoff, and consequently to simulated high peaks. Hence, SURLAG values contributed to smoothing the simulated hydrograph in the channel due to the delay in surface runoff release from the HRUs (Neitsch *et al.* 2011). The resulting optimal values of SOL_K (SOL_AWC) were identified to overcome the initial issue of the fast water movement through the soil (flow underestimation). Optimal values of RCHRG_DP (mean value 0.52) indicate that from the total water for aquifer recharge, approximately 52% (48%) recharges the deep aquifer (shallow aquifer). The high percentage of water reaching the deep aquifer is important for return flow from this aquifer to improve the streamflow simulation in the low-flow period. This agrees with the finding of Clark *et al.* (2014), who demonstrated the importance of return flow from deep aquifers to explain the sustained dry season flow in the neighbouring Andean Kosñypata basin.

3.5 Equifinality

Figure 4 shows that each calibration strategy produces a model or several models with good performance in streamflow simulation despite each model having a different set of parameters. This result demonstrates that model outputs are subjected to the effects of equifinality or the non-uniqueness issue (different sets of parameters in the calibration procedure resulting in similar simulations – see Beven 2006). This issue is common in complex nonlinear models such as SWAT (Shen *et al.* 2012, Ficklin and Barnhart 2014, Her and Chaubey 2015, Zhang *et al.* 2018a), which presents interactions among its parameters as reported by Zhang *et al.* (2018a).

Controlling the equifinality to arrive at meaningful parameter sets and solutions is a challenge. For instance, Fig. 4 shows that conventional calibration strategies based on hydrograph goodness-of-fit optimization (approaches 1 and 2 with NSE or lNSE only) produce pseudo-accurate models (with unrealistic parameter values), showing accurate performance statistics in streamflow simulation while incorrectly representing some internal basins processes. This study also reveals that even the bi-objective calibration strategy including FDC signature as criterion in addition to lNSE does not improve the equifinality. However, our proposed multi-objective calibration strategy improves parameter identifiability and reduces the equifinality because the inclusion of the baseflow index as part of the objective function leads to better identification of the CN parameter, which controls the flow partitioning into surface runoff and baseflow.

Overall, model calibration using only discharge is not sufficient to judge the validity of a model in representing the hydrologic system. Therefore, we suggest including more variables (e.g. LAI, evapotranspiration, snow, baseflow, and hydrological signatures) to better constrain the calibration process (which was also suggested by Krysanova *et al.* 2018). Likewise,

we suggest the use of multi-objective evolutionary algorithms that search for acceptable trade-offs between objectives, since these methods can help to mitigate the parameter uncertainty partly due to equifinality during the calibration. Otherwise, practitioners must be careful using single-objective optimization algorithms, since no sampling design schemes used in these algorithms consider the interactions among the parameters (Song *et al.* 2015, Devak and Dhanya 2017, Razmkhah *et al.* 2017), and hence the solution can be subjected to equifinality.

3.6 Basin water balance and vegetation response

For a better understanding of the water budget for VRB, Table 7 shows the long-term average annual values (1985–2015 period) for the most relevant water balance and flow components. The results show that annual precipitation in the basin is 748 mm, of which about 50% is lost from the system by evapotranspiration (ET = 375 mm), and 50% is the water yield of the basin (WYLD = 373 mm). Similar ratios (44.6–51%) of ET over precipitation in Ecuadorian Andean basins were reported (Guzmán *et al.* 2015, Mosquera *et al.* 2015, Carrillo-Rojas *et al.* 2019). Regarding the WYLD components, the contribution of baseflow (BF = 291 mm, 78%) is higher than that of surface runoff (Qsurf = 82 mm, 22%). The main component of baseflow is the lateral flow (Qlat = 137 mm, 47%), followed by return flow from the deep aquifer (Qgwd = 79 mm, 27%) and the return flow from the shallow aquifer (Qgws = 75 mm, 26%). Our results indicate that the large baseflow contribution plays a key role in modulating the flow regime of VRB, and the important groundwater contribution explains the dry-season baseflow. This outcome is consistent with the findings for Peruvian Andean basins draining into the Amazon River, such as the Kosñypata (Clark *et al.* 2014) and upper Marañon River basin (Hill *et al.* 2018), where the substantial dry-season discharge was attributed to return flow from deep aquifers.

To evaluate the role of tropical Andean vegetation on water yields and baseflow, the baseflow index estimated by SWAT – SWAT_BFI (ratio of water yield over precipitation) for each land-use type and its spatial variability at the HRU level is shown in Fig. 7e,c (Fig. 7d,b). The results show that the greatest ratio (~0.63) of water yield over precipitation is produced in pasture (PAST) areas which span the middle and upper part of the basin mainly (Fig. 1). Urban (URML) and barren (BARR) areas can produce water

Table 7. Long-term average annual water balance (1985–2015 period) for the VRB.

Water balance components	Value
Precipitation, *P* (mm)	748
Evapotranspiration, ET (mm)	375
Water yield, WYLD (mm)	373
Surface runoff, Qsurf (mm)	82
Baseflow, BF = Qlat + Qgws + Qgwd (mm)	291
Lateral flow, Qlat (mm)	137
Return flow from the shallow aquifer, Qgws (mm)	75
Return flow from the deep aquifer, Qgwd (mm)	79
Checks	
BFI	0.77
SWAT_BFI	0.79
BIAS_BFI	0.02
ET/P	0.50
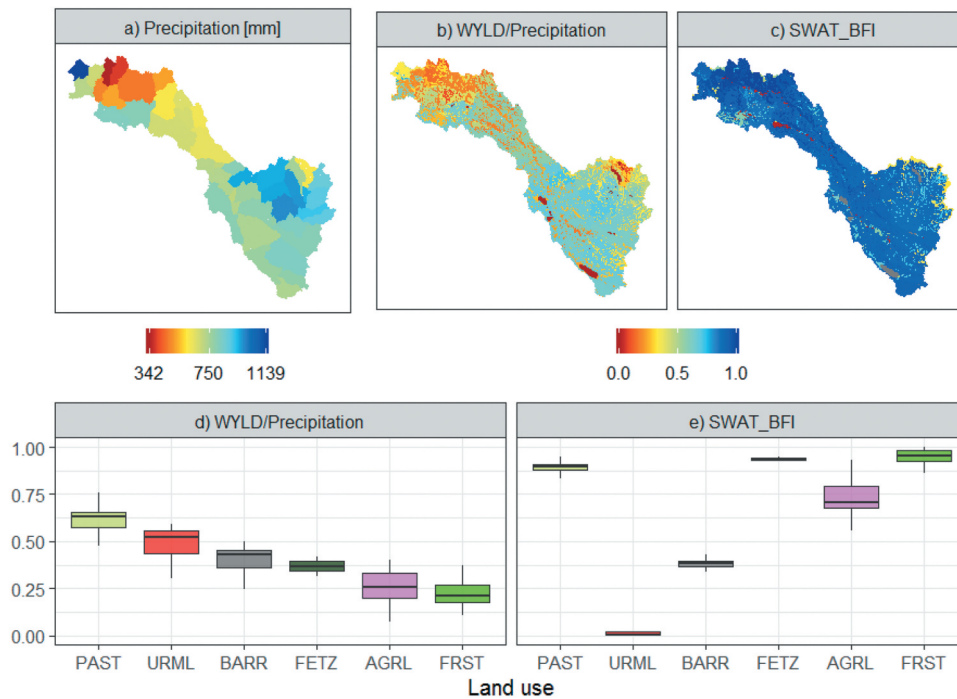
Figure 7. Above is (a) the spatial variability at HRU level of precipitation, (b) ratio of water yield (WYLD) over precipitation, and (c) simulated baseflow index (SWAT_BFI). Below is (d) the ratio of WYLD over precipitation and (e) SWAT_BFI for each land-use type.

yields of ~43% and ~52% of precipitation, respectively, contributed by surface runoff mainly, as the baseflow index (SWAT_BFI) values are very low in these land uses. Although the evergreen broadleaf forest (FETZ) areas located around the basin outlet experience the greatest amount of precipitation, the water yield in those areas is lower (< 37% of precipitation), as for example in agricultural areas (AGRL) and mixed forest (FRST).

Figure 7(e) shows lower values of baseflow indices for agricultural areas, compared to pasture for instance, which may indicate the poor hydrological regulation capacity of cultivated areas. Otherwise, mixed forest, evergreen broadleaf forest, and pasture exhibit higher rates (> ~0.89) of baseflow to water yield, which highlights the features of these land uses in improving the infiltration and subsurface processes. We observed, however, that pasture presents important hydrological services such as greater water yields and higher rates of baseflow simultaneously in comparison to forest, which shows higher baseflow but poor water yields in VRB. The negative impacts of forest on water yield are consistent with the findings of previous studies in Andean basins (e.g. Buytaert *et al.* 2007, Ochoa-Tocachi *et al.* 2016).

Finally, the verification of BIAS_BFI being equal to 0.02 (2%; Table 7) highlights the SWAT model's capability in simulating surface runoff and baseflow, which is essential to help the local water resources management including water supply services, identification of critical areas for soil conservation intervention, hydroelectric energy production, and preventing floods and droughts.

3.7 Limitations and perspectives

Despite the efforts of this study to reduce parameter uncertainty during the calibration procedure, we note that our results are still subject to uncertainties in the input data (e.g.

climate, soil, and land use), data used for model calibration (e.g. discharge, BFI, and LAI), and model structure due to the simplification of hydrologic processes as well as to uncertainties in remaining model parameters. For instance, despite the utility of the climate data used in this study to drive hydrologic model in a basin with data scarcity, such as VRB, gridded climate data are subject to uncertainties in the observed data and spatial interpolation procedures. In particular, precipitation may have systematic bias caused by wind, which is inherent in precipitation measurements and introduces an unquantified error (Pollock *et al.* 2018). We assumed that BFI, which was estimated using digital filter methods based on daily discharge and used to constrain the flow partitioning, gives a physically plausible result. However, future work should involve the use of tracers and/or stable isotopes to estimate BFI and validate the indirect methods used in this study. We rely on discharge data from the km-105 hydrometric station for model calibration and validation. However, discharge data may have errors because of inherent uncertainties in flow measurement and rating curves (Tomkins 2014). Uncertainties up to ±20% in flow measurement using the traditional area-velocity method and the current meter were reported in Andean basins of Colombia (Parra *et al.* 2016). Hence, future studies are needed to quantify uncertainties in hydrologic modelling owing to errors in observed discharge data in Andean basins. Finally, in this study, only one streamgauge was used for model calibration and validation because of data scarcity, but we are confident in the robustness of our methodology which can be used in instrumented basins to perform the calibration/validation at multiple sites within the catchment, and even in ungauged basins where FDC and baseflow index can be obtained through regionalization approaches (e.g. Beck *et al.* 2013, Atieh *et al.* 2017).

4 Summary and conclusions

This study developed a step-wise, multi-objective calibration framework applied to the SWAT model for the simulation of vegetation dynamics, streamflow, and flow partitioning. The first part of the framework deals with model calibration of leaf area index dynamics, for which the SWAT LAI-related parameters of perennial plants were calibrated using MODIS LAI data as a reference and following the LAI calibration scheme proposed in this study (see Section 2.6). The second part of the calibration strategy deals with model calibration of streamflow and flow partitioning, for which the inclusion of the baseflow index as well as discharge and FDC signatures within a multi-objective calibration approach is proposed. This approach is compared to discharge-based (single-objective, e.g. NSE) and signature-based (bi-objective, which includes criterion for discharge and FDC signatures) calibration strategies. The data-scarce Vilcanota River basin located in the Peruvian Andes served as a case study to demonstrate the advantages of the proposed model calibration framework, with a view to providing a better understanding of the basin's internal hydrological processes. The following conclusions can be drawn from the study:

- The LAI calibration scheme led to good model performance in the simulation of LAI when compared to MODIS LAI. Moreover, our findings shed light on the fact that in basins with high elevation gradients, heat units change with altitude; therefore, the SWAT parameter that controls the plant growth (PHU_PLT) decreases with height, and its calibration is crucial for correct LAI simulation in mountainous regions.
- Our results also show that better model performance in streamflow, FDC, and flow partitioning simulation is achieved when the model is calibrated using the proposed multi-objective calibration approach, whereas calibration approaches applied previously led to an unrealistic representation of flow partitioning even though good model performance for streamflow simulation is achieved with these strategies.
- The proposed methodology was observed to increase the identifiability of SWAT parameters related to evapotranspiration, streamflow, and flow partitioning, whereas the parameter values obtained by previous calibration approaches were unrealistic.
- This study shows that the solution of the SWAT model, which presents interactions among its parameters (Zhang *et al.* 2018a), using previous calibration approaches is subjected to equifinality since these approaches produced pseudo-accurate models, showing good model performance for streamflow simulation while incorrectly representing some internal basin processes. In contrast, the proposed multi-objective calibration, which includes the baseflow index, was observed to reduce the parameter equifinality.
- Regarding the eco-hydrology of the Andean Vilcanota River basin, it was found that evapotranspiration represents 50% of the average annual precipitation. The baseflow is the main component of the long-term streamflow (78% of it, on average) with an important contribution from deep aquifers that sustains the dry-season baseflow. Our findings further illustrate that areas covered by

pasture offer better hydrological services regarding the water yield and baseflow in comparison to other land uses. The ability of the SWAT model to realistically simulate vegetation dynamics, streamflow and baseflow can contribute to improving water resources management of the VRB and similar water catchments.

Overall, our proposed calibration and validation framework for hydrologic models such as SWAT increases the chances of obtaining the right answer for the right reason in hydrologic modelling, which is a crucial step toward more realistic hydrological applications. Examples include a better understanding of basin hydrology and water resources and an evaluation of the impacts of land-use changes and climate change. The proposed calibration framework can be applied in any mountain basin and can be adapted to the calibration of other physically or process-based hydrological models.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Carlos Antonio Fernandez-Palomino http://orcid.org/0000-0002-0573-4025
Fiorella Vega-Jácome http://orcid.org/0000-0003-4580-9324
Axel Bronstert http://orcid.org/0000-0002-6369-8536

References

Abbaspour, K. and Ashraf Vaghefi, S., 2019. *Harmonized world soil database in SWAT format*. doi:10.1594/PANGAEA.901309

Abbaspour, K., Vaghefi, S., and Srinivasan, R., 2017. A guideline for successful calibration and uncertainty analysis for soil and water assessment: a review of papers from the 2016 international SWAT conference. *Water*, 10, 6. doi:10.3390/w10010006

Acero Triana, J.S., *et al.*, 2019. Beyond model metrics: the perils of calibrating hydrologic models. *Journal of Hydrology*, 578, 124032. doi:10.1016/j.jhydrol.2019.124032

Alemayehu, T., *et al.*, 2017. An improved SWAT vegetation growth module and its evaluation for four tropical ecosystems. *Hydrology and Earth System Sciences*, 21 (9), 4449–4467. doi:10.5194/hess-21-4449-2017

Arnold, J.G. and Allen, P.M., 1999. Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association*, 35 (2), 411–424. doi:10.1111/j.1752-1688.1999.tb03599.x

Arnold, J.G., *et al.*, 1995. Automated base flow separation and recession analysis techniques. *Ground Water*, 33, 1010–1018. doi:10.1111/j.1745-6584.1995.tb00046.x

Arnold, J.G., *et al.*, 2012. Swat: model use, calibration, and validation. *Transactions of the ASABE*, 55, 1491–1508. doi:10.13031/2013.42256

Arnold, J.G., *et al.*, 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association*, 34 (1), 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x

Atieh, M., *et al.*, 2017. Prediction of flow duration curves for ungauged basins. *Journal of Hydrology*, 545, 383–394. doi:10.1016/j.jhydrol.2016.12.048

Aybar, C., *et al.*, 2019. Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 1–16. doi:10.1080/02626667.2019.1649411

Beck, H.E., *et al.*, 2013. Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, 49, 7843–7863. doi:10.1002/2013WR013918

Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320 (1–2), 18–36. doi:10.1016/j.jhydrol.2005.07.007

Brouziyne, Y., *et al.*, 2017. SWAT manual calibration and parameters sensitivity analysis in a semi-arid watershed in North-western Morocco. *Arabian Journal of Geosciences*, 10 (19), 427. doi:10.1007/s12517-017-3220-9

Buytaert, W., Iñiguez, V., and Bièvre, B.D., 2007. The effects of afforestation and cultivation on water yield in the Andean páramo. *Forest Ecology and Management*, 251 (1–2), 22–30. doi:10.1016/j.foreco.2007.06.035

Carrillo-Rojas, G., *et al.*, 2019. The breathing of the Andean highlands: net ecosystem exchange and evapotranspiration over the páramo of southern Ecuador. *Agricultural and Forest Meteorology*, 265, 30–47. doi:10.1016/j.agrformet.2018.11.006

Catacora-Acevedo, E.A., 2008. *Predicciones del comportamiento de caudales de la C.H. Machupicchu mediante análisis ARIMA de series temporales*. Lima, Peru: Universidad Nacional de Ingeniería. Available from: http://cybertesis.uni.edu.pe/handle/uni/585

Chen, Y., *et al.*, 2018. Toward improved calibration of SWAT using season-based multi-objective optimization: a case study in the Jinjiang Basin in Southeastern China. *Water Resource Management*, 32, 1193–1207. doi:10.1007/s11269-017-1862-8

Chilkoti, V., Bolisetti, T., and Balachandar, R., 2018. Multi-objective autocalibration of SWAT model for improved low flow performance for a small snowfed catchment. *Hydrological Sciences Journal*, 63 (10), 1482–1501. doi:10.1080/02626667.2018.1505047

Chilkoti, V., Bolisetti, T., and Balachandar, R., 2019. Diagnostic evaluation of hydrologic models employing flow duration curve. *Journal of Hydrologic Engineering*, 24 (6), 05019009. doi:10.1061/(ASCE)HE.1943-5584.0001778

Clark, K.E., *et al.*, 2014. The hydrological regime of a forested tropical Andean catchment. *Hydrology and Earth System Sciences*, 18 (12), 5377–5397. doi:10.5194/hess-18-5377-2014

Devak, M. and Dhanya, C.T., 2017. Sensitivity analysis of hydrological models: review and way forward. *Journal of Water and Climate Change*, 8 (4), 557–575. doi:10.2166/wcc.2017.149

FAO-UNESCO, 1988. Soil map oF world, revised legend. Rome: FAO, World Soil Resources Reports 60.

Ficklin, D.L. and Barnhart, B.L., 2014. SWAT hydrologic model parameter uncertainty and its implications for hydroclimatic projections in snowmelt-dependent watersheds. *Journal of Hydrology*, 519, 2081–2090. doi:10.1016/j.jhydrol.2014.09.082

Francesconi, W., *et al.*, 2016. Using the Soil and Water Assessment Tool (SWAT) to model ecosystem services: a systematic review. *Journal of Hydrology*, 535, 625–636. doi:10.1016/j.jhydrol.2016.01.034

Gassman, P.W., *et al.*, 2007. The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASAE*, 50, 1211–1250. doi:10.13031/2013.23637

Gupta, H.V., *et al.*, 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1–2), 80–91. doi:10.1016/J.JHYDROL.2009.08.003

Gupta, H.V., Sorooshian, S., and Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research*, 34 (4), 751–763. doi:10.1029/97WR03495

Gupta, H.V., Sorooshian, S., and Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4 (2), 135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)

Guse, B., *et al.*, 2016. Demasking the integrated information of discharge: advancing sensitivity analysis to consider different hydrological components and their rates of change. *Water Resources Research*, 52 (11), 8724–8743. doi:10.1002/2016WR018894

Guzmán, P., *et al.*, 2015. Comparative analysis of baseflow characteristics of two Andean catchments, Ecuador. *Hydrological Processes*, 29 (14), 3051–3064. doi:10.1002/hyp.10422

Ha, L., *et al.*, 2018. Calibration of spatially distributed hydrological processes and model parameters in SWAT using remote sensing data and an auto-calibration procedure: a case study in a Vietnamese River Basin. *Water*, 10, 212. doi:10.3390/w10020212

Hadka, D. and Reed, P., 2012. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evolutionary Computation*, 20 (3), 423–452. doi:10.1162/EVCO_a_00053

Hadka, D. and Reed, P., 2013. Borg: an auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, 21 (2), 231–259. doi:10.1162/EVCO_a_00075

Hattermann, F.F., *et al.*, 2005. Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment - Validation and uncertainty analysis. *Hydrological Processes*, 19, 693–714. doi:10.1002/hyp.5625

Her, Y. and Chaubey, I., 2015. Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. *Hydrological Processes*, 29 (19), 4220–4237. doi:10.1002/hyp.10487

Hill, A.F., Stallard, R.F., and Rittger, K., 2018. Clarifying regional hydrologic controls of the Marañón River, Peru through rapid assessment to inform system-wide basin planning approaches. *Elementa: Science of the Anthropocene*, 6, 37. doi:10.1525/elementa.290

Hrachowitz, M., *et al.*, 2014. Process consistency in models: the importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, 50 (9), 7445–7469. doi:10.1002/2014WR015484

Huerta, A., Aybar, C., and Lavado-Casimiro, W., 2018. *PISCO temperatura v.1.1*. Lima-Perú: SENAMHI - DHI-2018.

Jajarmizadeh, M., *et al.*, 2017. Optimal calibration and uncertainty analysis of SWAT for an arid climate. *Air, Soil and Water Research*, 10, 117862211773179. doi:10.1177/1178622117731792

Jang, W.S., Engel, B., and Ryu, J., 2018. Efficient flow calibration method for accurate estimation of baseflow using a watershed scale hydrological model (SWAT). *Ecological Engineering*, 125, 50–67. doi:10.1016/j.ecoleng.2018.10.007

Jarvis, A., *et al.*, 2008. *Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database*. CGIAR Consortium for Spatial Information (CGIAR-CSI).

Krause, P., Boyle, D.P., and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97. doi:10.5194/adgeo-5-89-2005

Krysanova, V. and Arnold, J.G., 2008. Advances in ecohydrological modelling with SWAT-a review. *Hydrological Sciences Journal*, 53, 939–947. doi:10.1623/hysj.53.5.939

Krysanova, V., *et al.*, 2018. How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal*, 63 (5), 696–720. doi:10.1080/02626667.2018.1446214

Krysanova, V., Müller-Wohlfeil, D.I., and Becker, A., 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. *Ecological Modelling*, 106, 261–289. doi:10.1016/S0304-3800(97)00204-4

Larabi, S., *et al.*, 2018. Multi-criteria process-based calibration using functional data analysis to improve hydrological model realism. *Water Resources Management*, 32 (1), 195–211. doi:10.1007/s11269-017-1803-6

Liang, X., *et al.*, 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99 (D7), 14415–14428. doi:10.1029/94jd00483

Lim, K.J., *et al.*, 2005. AUTOMATED WEB GIS BASED HYDROGRAPH ANALYSIS TOOL, WHAT. *Journal of the American Water Resources Association*, 41 (6), 1407–1416. doi:10.1111/j.1752-1688.2005.tb03808.x

Lindström, G., et al., 2010. Development and testing of the HYPE (hydrological predictions for the environment) water quality model for different spatial scales. Hydrology Research, 41 (3–4), 295–319. doi:10.2166/nh.2010.007

Luo, Y., et al., 2012. Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. Hydrology and Earth System Sciences, 16 (4), 1259–1267. doi:10.5194/hess-16-1259-2012

Ma, T., et al., 2019. Enhancing SWAT with remotely sensed LAI for improved modelling of ecohydrological process in subtropics. Journal of Hydrology, 570, 802–815. doi:10.1016/J.JHYDROL.2019.01.024

McMillan, H., Westerberg, I., and Branger, F., 2017. Five guidelines for selecting hydrological signatures. Hydrological Processes, 31, 4757–4761. doi:10.1002/hyp.11300

Meaurio, M., et al., 2015. Evaluation of SWAT models performance to simulate streamflow spatial origin. The case of a small forested watershed. Journal of Hydrology, 525, 326–334. doi:10.1016/J.JHYDROL.2015.03.050

Mohammed, R. and Scholz, M., 2018. Flow–duration curve integration into digital filtering algorithms for simulating climate variability based on river baseflow. Hydrological Sciences Journal, 63, 1558–1573. doi:10.1080/02626667.2018.1519318

Moriasi, D.N., et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50, 885–900. doi:10.13031/2013.23153

Mosquera, G.M., et al., 2015. Runoff from tropical alpine grasslands increases with areal extent of wetlands. CATENA, 125, 120–128. doi:10.1016/J.CATENA.2014.10.010

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10 (3), 282–290. doi:10.1016/0022-1694(70)90255-6

Neitsch, S.L., et al., 2011. Soil & water assessment tool theoretical documentation version 2009. Texas Water Resource Institute, TR-406 1–647.

Ochoa-Tocachi, B.F., et al., 2016. Impacts of land use on the hydrological response of tropical Andean catchments. Hydrological Processes, 30, 4074–4089. doi:10.1002/hyp.10980

Parra, B.G., et al., 2016. Uncertainty of discharge estimation in high-grade Andean streams. Flow Measurement and Instrumentation, 48, 42–50. doi:10.1016/j.flowmeasinst.2016.02.005

Pelletier, J.D., et al., 2016. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. Journal of Advances in Modeling Earth Systems, 8, 41–65. doi:10.1002/2015MS000526

Pfannerstill, M., et al., 2017. How to constrain multi-objective calibrations of the SWAT model using water balance components. JAWRA Journal of the American Water Resources Association, 53 (3), 532–546. doi:10.1111/1752-1688.12524

Pfannerstill, M., Guse, B., and Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. Journal of Hydrology, 510, 447–458. doi:10.1016/J.JHYDROL.2013.12.044

Pokhrel, P. and Yilmaz, K.K., 2012. Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. Journal of Hydrology, 418–419, 49–60. doi:10.1016/J.JHYDROL.2008.12.004

Pollock, M.D., et al., 2018. Quantifying and mitigating wind-induced undercatch in rainfall measurements. Water Resources Research, 54, 3863–3875. doi:10.1029/2017WR022421

Qi, J., et al., 2020. Effects of surface runoff and infiltration partition methods on hydrological modeling: A comparison of four schemes in two watersheds in the Northeastern US. Journal of Hydrology, 581. doi:10.1016/j.jhydrol.2019.124415

Rajib, A., et al., 2018. Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. Journal of Hydrology, 567, 668–683. doi:10.1016/j.jhydrol.2018.10.024

Razmkhah, H., AkhoundAli, A.-M., and Radmanesh, F., 2017. Correlated parameters uncertainty propagation in a rainfall-runoff model, considering 2-copula; case study: Karoon III River Basin. Environmental Modeling & Assessment, 22, 503–521. doi:10.1007/s10666-017-9569-z

Ross, C.W., et al., 2018. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. Scientific Data, 5, 180091. doi:10.1038/sdata.2018.91

Sahraei, S., Asadzadeh, M., and Unduche, F., 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: application in flood forecasting for Canadian Prairies. Journal of Hydrology, 588, 125095. doi:10.1016/j.jhydrol.2020.125095

Shafii, M., et al., 2017. A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. Water Resources Research, 53 (4), 3279–3301. doi:10.1002/2016WR019736

Shafii, M. and De Smedt, F., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. Hydrology and Earth System Sciences, 13, 2137–2149. doi:10.5194/hess-13-2137-2009

Shafii, M. and Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resources Research, 51 (5), 3796–3814. doi:10.1002/2014WR016520

Shen, Z.Y., Chen, L., and Chen, T., 2012. Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. Hydrology and Earth System Sciences, 16, 121–132. doi:10.5194/hess-16-121-2012

Singh, S.K., et al., 2019. Towards baseflow index characterisation at national scale in New Zealand. Journal of Hydrology, 568, 646–657. doi:10.1016/j.jhydrol.2018.11.025

Somers, L.D., et al., 2019. Groundwater buffers decreasing glacier melt in an Andean watershed—but not Forever. Geophysical Research Letters, 46, 13016–13026. doi:10.1029/2019GL084730

Song, X., et al., 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. Journal of Hydrology, 523, 739–757. doi:10.1016/j.jhydrol.2015.02.013

Strauch, M. and Volk, M., 2013. SWAT plant growth modification for improved modeling of perennial vegetation in the tropics. Ecological Modelling, 269, 98–112. doi:10.1016/J.ECOLMODEL.2013.08.013

Tomkins, K.M., 2014. Uncertainty in streamflow rating curves: methods, controls and consequences. Hydrological Processes, 28, 464–481. doi:10.1002/hyp.9567

van Werkhoven, K., et al., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, 32, 1154–1169. doi:10.1016/J.ADVWATRES.2009.03.002

Wagener, T., et al., 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5 (1), 13–26. doi:10.1038/sdata.2018.91

Wei, Z., et al., 2018. The application of a modified version of the SWAT model at the daily temporal scale and the hydrological response unit spatial scale: a case study covering an irrigation district in the Hei River Basin. Water, 10, 1064. doi:10.3390/w10081064

Yesuf, H.M., et al., 2016. Streamflow prediction uncertainty analysis and verification of SWAT model in a tropical watershed. Environmental Earth Sciences, 75, 806. doi:10.1007/s12665-016-5636-z

Yilmaz, K.K., Gupta, H.V., and Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. Water Resources Research, 44. doi:10.1029/2007WR006716

Yuan, H., et al., 2011. Reprocessing the MODIS leaf area index products for land surface and climate modelling. Remote Sensing of Environment, 115 (5), 1171–1187. doi:10.1016/J.RSE.2011.01.001

Zhang, J., Lei, X., and Li, Q., 2018a. Two model performance comparisons with multisite observations based on uncertainty methods for modeling hydrologic dynamics. Journal of Irrigation and Drainage Engineering, 144 (1), 04017060. doi:10.1061/(ASCE)IR.1943-4774.0001284

Zhang, R., et al., 2018b. Can multi-objective calibration of streamflow guarantee better hydrological model accuracy? Journal of Hydroinformatics, 20, 687–698. doi:10.2166/hydro.2018.131

Appendix.
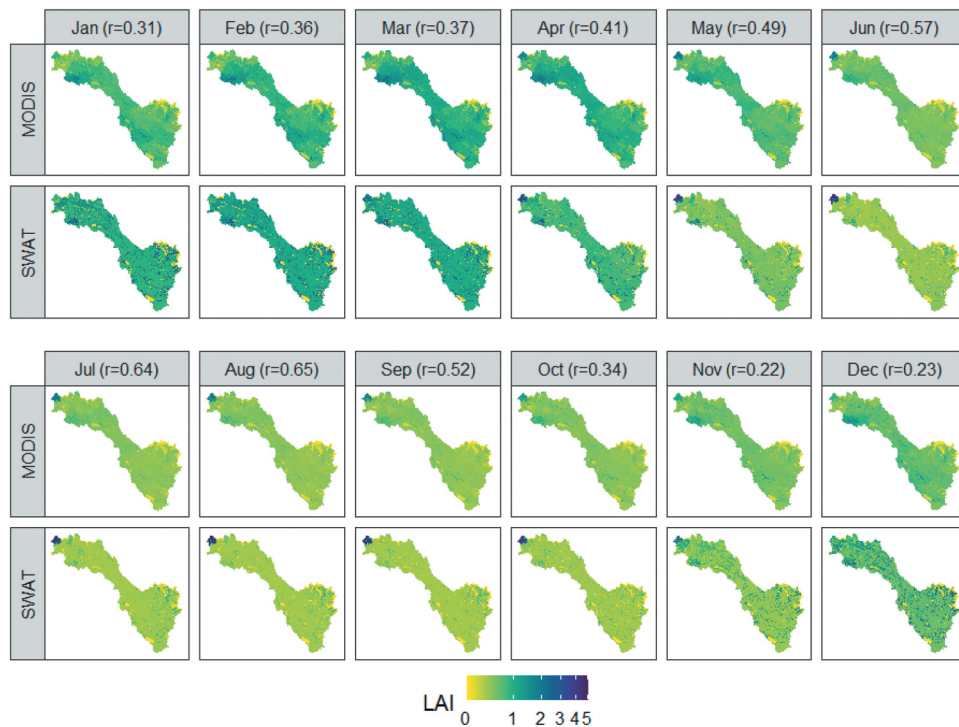
Figure A1. Spatio-temporal variability of average monthly LAI values for the period 2005–2015 estimated by MODIS and SWAT at HRU level. The spatial correlation (r) between MODIS and SWAT LAI is shown in parentheses.
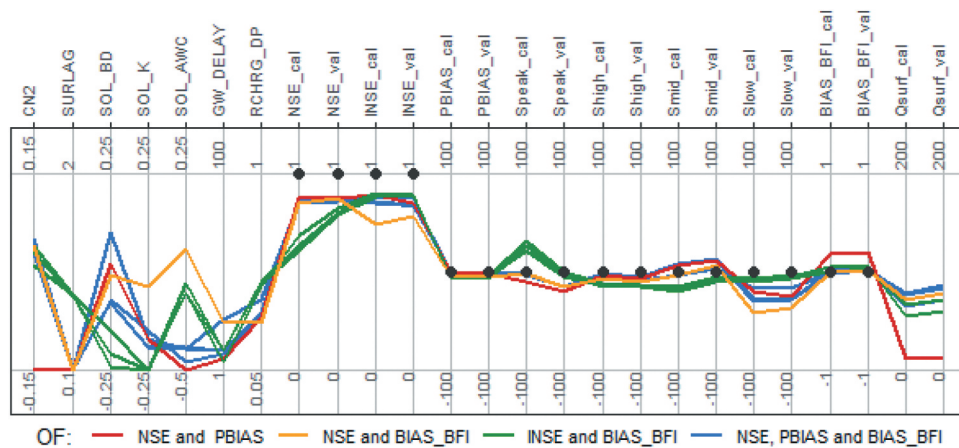
Figure A2. Parallel coordinates plot of the Pareto front optimal solutions obtained by additional calibration strategies: the traditional biobjective function based on discharge only (NSE and PBIAS) and the objectives defined by discharge measures and baseflow index, such as "NSE and BIAS_BFI," "INSE and BIAS_BFI," and "NSE, PBIAS, and BIAS_BFI." For each solution, optimal parameter values, discharge-based performance measures (NSE, INSE and PBIAS), and hydrological signatures (Speak, Shigh, Smid, Slow, and BIAS_BFI) are displayed. _cal (_val) indicates the measurements for the calibration (validation) period. Qsurf is the mean surface runoff in mm. A description of SWAT parameters is provided in Table 6. A description of objective functions (OFs) is provided in Table 3. All calibration strategies shown in this figure perform worse than those calibration strategies that include INSE, FDC signatures, and baseflow index within a multiobjective calibration framework as shown in Fig. 4.