Potential of snow data to improve the consistency and robustness of a semi-distributed hydrological model using the SAFRAN input dataset

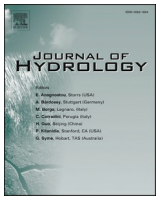
Denis Ruelland

Journal of Hydrology 631 (2024) 130820

Contents lists available at ScienceDirect

Journal of Hydrology

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



Potential of snow data to improve the consistency and robustness of a semi-distributed hydrological model using the SAFRAN input dataset

Denis Ruelland

CNRS, HydroSciences Laboratory, University of Montpellier, Hydropolis, 15 Avenue Flahault, 34093 Montpellier Cedex 5, France

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief.

Keywords:
Degree-day models
Multi-objective calibration
MODIS snow-covered area
SWE
Sensitivity analysis

ABSTRACT

This paper compares different calibration strategies for using snow data combined with streamflow records to constrain model optimisation in mountain catchments. In particular, it assesses to what extent the use of snow observations makes it possible to improve the consistency, identifiability and robustness of the calibrated parameters. To answer this question, a semi-distributed snow and ice model was used on top of a rainfall-runoff model using the SAFRAN meteorological reanalysis as input dataset on several catchments in the Alps and Pyrenees. Model calibration and control were based on streamflow observations, remotely-sensed snow-covered areas and in-situ snow water equivalent (SWE) measurements. The snow and rainfall-runoff parameters were calibrated sequentially and simultaneously against objective functions integrating different combinations of runoff, snow cover and SWE criteria. Statistical assessment of model performances in independent evaluation periods showed that sequential calibration of the snow parameters gives too much weight to snow compared to runoff. Instead, incorporating snow data in the simultaneous calibration of the parameters improves snow simulations without impairing runoff performance. This can be achieved by limiting the weight of the snow criteria to 25% in the objective function. Although local SWE measurements were found to be more useful than satellite observations for identifying more consistent parameters, it is advisable to include both in the calibration process through a compromise objective function. However, the improved model consistency was not accompanied by a significant reduction in equifinality and optimisation times. On the other hand, improving internal consistency made it possible to reduce the interdependence between the parameters of the snow model and those of the rainfall-runoff model. This also made it possible to identify the least sensitive snow parameters in order to fix them at general values without impairing model performance while reducing equifinality with a more parsimonious model. Finally, there was no evidence that using snow observations in the calibration process improves model robustness with respect to climate variability.

1. Introduction

1.1. On the need to calibrate snow-hydrological parameters

Temperature-index models (Hock, 2003) are widely used to estimate snowmelt runoff in operational hydrological models at catchment scale. They rely on parameters that are not directly measurable and must therefore be estimated through model calibration, i.e. by adjusting the model parameters to fit the simulated outputs of the model to the observed variables of the catchment. Temperature-index models are often calibrated at the same time as a rainfall-runoff model, which can pose equifinality issues (Beven, 2006) and hamper the transferability of model parameters in space and over time. Although the main objective is to obtain streamflow forecasts that are as reliable as possible, such a modelling approach also produces estimates of the water stored in the snowpack. The latter information can be compared to remotely-sensed or field measurements, which encourages the use of additional data to further constrain parameter calibration. Such models are consequently usually calibrated against observed runoff, sometimes against satellite snow cover data and, more rarely, against local snow measurements.

1.2. Data for model calibration in mountain catchments

Catchment discharge is often the only observed data available for use in the calibration process (Franz and Karsten, 2013). Because observations of snow conditions are lacking in this case, snow models are calibrated jointly with rainfall–runoff models using observed streamflow series (e.g. Klok et al., 2001; Luo et al., 2013; Valéry et al., 2014). The

E-mail address: denis.ruelland@umontpellier.fr.

https://doi.org/10.1016/j.jhydrol.2024.130820
Received 1 September 2023; Received in revised form 8 January 2024; Accepted 16 January 2024
Available online 3 February 2024

disadvantage is that discharge does not reflect the spatial distribution of snow throughout the catchment (He et al., 2014). As a result, calibrating the model only on an integrated response variable at the catchment outlet (streamflow) does not necessarily enable optimal model performance at inland points (Ajami et al., 2004; Khakbaz et al., 2012; Smith et al., 2004), which is a serious constraint to the reliability of distributed hydrological models. As noted by Franz and Karsten (2013), discharge data may not contain the necessary information to correctly identify all parameter values during calibration. This suggests it may be useful to include auxiliary information to improve parameter identifiability while also providing further insights into the reliability of the processes that have been modelled.

Snow model parameters can also be estimated from local measurements of snow (e.g. Parajka et al., 2007; Besic et al., 2014; Bormann et al., 2014; Magnusson et al., 2014; Tuo et al., 2018; Nemri and Kinnard, 2020; Ruelland, 2023). Snow water equivalent (SWE) and snow depth (SD) are typically monitored through manual observations as well as by automatic measurement stations to estimate snowpack mass and volume on the ground. SWE is a function of snow density (ρ_S) and SD. It can be defined as the depth of the liquid water layer that would be produced if all the snow in the snowpack melted (Rees, 2005). Because it is easier to measure, SD is more often monitored in the field than SWE. However, Magnusson et al. (2015) showed that observed SD was not a critical measure for evaluating snow models and that snow mass information provided by SWE observations was more valuable for testing the reliability of snow models for hydrological applications. Nevertheless, the relation between SWE and SD is complex, in particular because ρ_S generally increases with increasing snow age. As direct measurements of SWE or ρ_S are rarely available, a snow density model is needed to convert SD into SWE (e.g. Magnusson et al., 2014; Bormann et al., 2014; Tuo et al., 2018), which tends to lead to additional parametric uncertainty. Direct SWE measurements are therefore more reliable. However, such point-scale measurements are typically very rare, particularly at the highest elevations. Even when they are available, in-situ SWE observations are only representative of a small subset of the spatial domain (He et al., 2014), which may mean that, despite point-scale accuracy, they are perceived as providing insufficient information on the snowpack at catchment scale (Dong, 2018). While these snow data have already been used to control SWE simulated using conceptual hydrological models (e.g. Garavaglia et al., 2017), few studies have evaluated their benefits for model calibration in both streamflow and SWE simulations. For example, Tuo et al. (2018) found that using SD measurements converted to SWE with density parametrisation in the calibration process greatly improved SWE simulations and slightly improved streamflow simulations. Nemri and Kinnard (2020) also showed that the additional information provided by SWE measured at snow course transects improved the simulation of SWE without significantly impairing streamflow simulations in non-mountainous catchments.

Remotely-sensed snow cover data are an alternative source of data to constrain model parameters. These observations only concern the extent of snow on the ground and provide no information on the mass and volume of snow. But as they record any spatial and temporal variations in snow cover within the catchment, their use is particularly appealing for calibrating and evaluating distributed models. They can be considered as complementary to discharge time series, which are spatially integrated but provide quantitative information on the water balance (Finger et al., 2011). Many authors have thus used satellite-based snow-covered area (SCA) to help calibrate snow model parameters (for a review see Parajka and Blöschl, 2012). Most of these studies used the SCA observed since 2000 by the MODIS (moderate-resolution imaging spectroradiometer) sensor (Hall and Riggs, 2002) at a spatial resolution of about 500 m and at a daily time step. MODIS snow products have been found to be of high quality with an overall accuracy of about 93 %, which varies depending on the type of land cover and on the condition of snow (Hall and Riggs, 2007). In addition to streamflow records, MODIS snow products have therefore been widely used to calibrate and control models in mountain catchments (e.g. Parajka and Blöschl, 2008; Pellicciotti et al., 2012; Finger et al., 2011; Franz and Karsten, 2013; Riboust et al., 2019; Ruelland, 2020; 2023). The authors generally reported that including MODIS snow data in the model calibration improved both model performance (Parajka and Blöschl, 2008) and internal consistency (Finger et al., 2011; Pellicciotti et al., 2012). For instance, Finger et al. (2011) suggested that although MODIS data provide no information on the actual snow amount, their predictive power is similar to that of ground-based mass balance information.

1.3. Multi-objective calibration using snow data in addition to runoff

Approaches designed to integrate information in addition to runoff in the calibration process are generally referred to as multi-objective calibration approaches, which allow different objectives (related to runoff, snow cover, SWE, etc.) to be combined in different ways. For example, snow data can be calibrated separately and sequentially in a hierarchical manner (e.g. Ragettli and Pellicciotti, 2012; Kelleher et al., 2017) or simultaneously (e.g. Finger et al., 2011; Pellicciotti et al., 2012; Ruelland, 2020; 2023) as part of multi-objective and multi-variable strategies aimed at finding acceptable compromises between different modelling objectives. Multi-objective calibration is typically carried out using Monte Carlo sampling (e.g. Finger et al., 2011; Duethmann et al., 2014; Kelleher et al., 2017), evolutionary algorithms based on the concept of Pareto-optimality (e.g. Parajka et al., 2007; Pellicciotti et al., 2012; Duethmann et al., 2014; Hublart et al., 2015; Nemri and Kinnard, 2020), or global optimisation algorithms focussed on composite multi-objective functions (e.g. Parajka and Blöschl, 2008; Riboust et al., 2019; Ruelland, 2020, 2023). These works highlighted the interest of using snow data in addition to runoff to constrain hydrological models but at the same time, raise questions about the best calibration strategies to use in order to reduce parameter uncertainties.

In situ and remotely sensed snow observation data can be used independently or combined with streamflow data to calibrate snow parameters. Calibrating the snow model using only snow observations aims to eliminate dependence on an associated hydrological model for successful identification of snow model parameters, as “snow simulations could be evaluated directly without the need to first pass the output through another model” (Franz and Karsten, 2013). However, it remains unclear whether such a sequential approach is suitable for calibrating snow-hydrological models due to the probable interdependence of their parameters. For instance, Franz and Karsten (2013) tested a multi-step approach by calibrating to MODIS SCA and runoff in separate steps with the aim of optimising only the parameters of the model likely to be sensitive to each type of data. They found that this approach could improve SCA simulations but did not necessarily result in more accurate discharge simulations. In a study based on a snow-hydrological model in non-mountain catchments, Nemri and Kinnard (2020) reported that calibrating snow parameters separately from SWE observations resulted in the best SWE simulations but impaired the streamflow simulations due to overfitting of snow parameters on SWE observations. Hypothesising that the interdependence of snow hydrological parameters is inevitable, as an alternative, other authors (e.g. Parajka and Blöschl, 2008; Ruelland, 2020) suggested to jointly calibrate snow and runoff parameters using multi-criteria objective functions involving observed discharge and snow cover data. The combination of point-scale SWE measurements and remotely-sensed snow observations is much less common. For instance, Ruelland (2023) used a multi-criteria composite function focusing on variations in local SWE, fractional snow-covered area, and different streamflow features to identify the degrees of freedom and complexity warranted in temperature-index models. This objective function, which gives more weight to the runoff criterion than to snow criteria, was judged to be a good compromise for jointly calibrating SWE, SCA and runoff dynamics.

However, calibration may also be significantly affected by the nature of the objective functions “which reflect the joint effects of all the model

parameters in a holistic way", as suggested by He et al. (2014). Optimisation procedures can thus lead to marked uncertainties in parameter estimates, especially if not enough attention is paid to the different ways of weighting the observational data sets in the objective function (Kirchner, 2006; Pellicciotti et al., 2012). Sometimes the default choice is to set all weights equal for all the observation sources used for calibration (e.g. Finger et al., 2015). However, different weights could be chosen to account for the reliability and sampling frequency of the different variables or to reflect the relative importance of agreement for the different variables, with streamflow often being the most important variable (Bergström et al., 2002). However, only a few authors have explicitly explored trade-offs between snow and discharge performance depending on the weights attributed to the various criteria. Duethmann et al. (2014) found limited trade-offs between satisfactory simulations in terms of runoff and SCA, but also reported that "good model performance with respect to discharge did not rule out poor performance in terms of SCA". Parajka et al. (2007) analysed model performance trade-off curves for discharge and SCA interpolated from observed snow depth data. By attributing a relative importance of 80 % (respectively, 20 %) to the streamflow (resp. SCA) criterion in the objective function, they found that using SCA significantly improved model performance in terms of snow cover, but slightly reduced performance in terms of runoff. Parajka and Blöschl (2008) also used a weighted sum approach to analyse the effects of varying the weights to combine satellite-derived SCA and discharge in a multi-objective criterion. Their results suggest that a 90 % weight attributed to the SCA criterion was a representative trade-off between the runoff and snow objectives. On the other hand, Riboust et al (2019) reported that a weighting of the optimization criterion of 75 % allocated to the runoff criterion and of 25 % allocated to the SCA criterion seemed to be a reasonable compromise between the performance of the model in terms of runoff and SCA.

While using snow data for multi-objective calibration of hydrological models has proven useful in producing more reliable simulations and better internal consistency (Parajka et al., 2007; Parajka and Blöschl, 2008; Finger et al., 2011; Pellicciotti et al., 2012; Duethmann et al., 2014; Finger et al., 2015; Kelleher et al., 2017; Tuo et al., 2018; Nemri and Kinnard, 2020), few studies have investigated and demonstrated that such an approach can also reduce model equifinality and uncertainty. For instance, Parajka et al. (2007) showed that using snow cover data in addition to runoff significantly reduced the uncertainty of the parameters representing snow cover dynamics, but on the other hand, tended to increase the uncertainty of other parameters. Although Finger et al. (2011) demonstrated the benefits of using multiple data (discharge, MODIS SCA, and seasonal glacier mass balances) to constrain the parameters of a distributed hydrological model, they did not fully succeed in demonstrating that "these data contained the necessary amount of spatial, volumetric and temporal information for an equifinality-free model calibration". Duethmann et al. (2014) reported that including a snow cover criterion in addition to runoff in the calibration process generally led to "a shift in the parameter distribution, with only small effects on further constraining the distribution". Kelleher et al. (2017) also showed that, although using auxiliary data made it possible to improve internal predictive capacity by selecting more behavioural parameter sets, these additional constraints had a limited impact on narrowing parameter ranges, suggesting that equifinality was not significantly reduced.

Finally, given that the impacts of climate change on mountain water resources are of great concern for water management (Duethmann et al., 2014), hydrological models must be able to produce reliable simulations when climate conditions shift beyond the range of prior experience. Coron et al. (2014) demonstrated that the transferability of water balance adjustments made during calibration in mountain catchments could be very limited, with potentially huge impacts in the case of studies conducted under non-stationary conditions. However, in this study, the tested models were calibrated on runoff alone. Few authors have investigated this question using auxiliary snow data. Sleziak et al. (2020) reported that including snow data into the objective function can improve the temporal stability of snow-related parameters. Duethmann et al. (2020) showed that changes to the objective function to improve internal consistency of the model did not lead to a better performance in a changing climate. Riboust et al. (2019) reported that their model jointly calibrated to MODIS-SCA and runoff tended to produce a slightly more stable performance in the case of streamflow simulations over past periods far from the calibration period. Nemri and Kinnard (2020) recommended using conceptual models calibrated jointly on SWE and runoff data to assess the impacts of climate change on snow cover and spring flow. These recent suggestions are not sufficient to determine whether models calibrated using snow data in addition to runoff are more robust for studies of the impact of climate change.

1.4. Objectives

Despite many attempts to incorporate in-situ measurements and remote sensing products in snow hydrology (see Dong, 2018 for a complementary review), guidelines are still lacking on the use of snow data in addition to runoff to calibrate hydrological models in mountain catchments. In particular, the real merits of information provided by different combinations of SWE and MODIS-SCA have, to our knowledge, never been evaluated in different catchments in the context of water balance modelling. Given the challenges outlined above, comparative studies are needed to answer the following questions:

- Should the snow model and the rainfall-runoff model be calibrated sequentially or simultaneously?
- Is it possible to optimise snow and runoff criteria simultaneously during calibration? If not, how should the snow criteria be weighted in the objective function to improve snow simulations without impairing runoff simulations?
- Is there a best compromise multi-criteria objective function? For example, is it more useful to incorporate point-scale SWE measurements, MODIS-based SCA or both in the calibration process?
- How are the parameters affected by the different objective functions and does the use of snow auxiliary data in a multi-objective calibration improve their identifiability while reducing computation time and equifinality?
- Is a model calibrated on snow data more robust with respect to climatic variability than the same model calibrated on runoff alone?

This article investigates these questions in order to provide recommendations on how to use snow and runoff data jointly to constrain parameter inference of hydrological models in mountain catchments. It compares different calibration strategies combining SWE, SCA and runoff criteria in an attempt to improve the consistency and robustness of snow-hydrological simulations. The paper is organised as follows. Section 2 presents the catchments and data used for the comparative assessment. Section 3 details the modelling framework, i.e. it describes the snow- and ice- accounting routine used on top of a rainfall-runoff model and the general evaluation strategy. Section 4 summarises sensitivity analyses on: (i) the sequential versus simultaneous calibration of the snow model and the rainfall-runoff model; (ii) the weight of the snow criteria in the objective function; (iii) parameter identifiability depending on the objective function; and (iv) model robustness to long-term climate variability. Section 5 discusses the main results in relation to the existing literature and some questionable sources of uncertainty. Finally, Section 6 concludes with recommendations and remarks on the advantages and limits of snow data in reducing model uncertainty.

2. Material: catchment sample and data

2.1. Selection of catchments

The modelling experiment was applied to a set of 13 mountainous,

snow-dominated catchments located in the French Alps and Pyrenees (Table 1). This set of catchments was selected to avoid catchment-specific results. It thus represents geographical and hydrological features that are as diverse as possible with no major human influence on runoff. However, the selection was limited by the availability of high-quality snow and runoff data for model calibration and evaluation.

2.1.1. SWE measurements

To accurately estimate the snowmelt inflows of its dams, the French hydro-power company EDF (*Electricité de France*) runs a network of 31 Cosmic-Ray Snow Sensors (CRS) in French mountainous regions. This network was built in the early 2000s. The majority of CRS were installed at sites which had generally been monitored by manual surveys for several decades, and so the historical snow data available at these sites is significant, and allows robust statistical processing of measurements that were made for calibration of the instruments (Paquet and Laval, 2006). In addition to being the standard used to characterise soil moisture (Zreda et al., 2012), the CRS gauges provide real-time measurement of the snow water equivalent (in mm) by measuring absorption of the cosmic-ray neutrons by the snowpack water (Kodama et al., 1979). A « snow-free » reference signal is necessary to account for the natural variations of the cosmic ray. The local calibration of each device with snow gauge measurements is also essential to account for the site effects and deviations from the sensor to the generic calibration curve. With those precautions, the accuracy and the reliability of the CRS measurements are adequate to estimate SWE. According to Paquet and Laval (2006), 90 % of relative errors between CRS and survey measurements are less than ±20 %. For water values greater than 200 mm, error values fall to ±12 %. The largest differences in the survey can be explained by particular snow conditions such as transport by the wind or a “rotten” coat that evolves very rapidly. Measurements made by CRS gauges can therefore be used for calibration or assimilation in hydrological models.

For the present study, only daily in-situ SWE measurements available in the selected catchments were used. These measurements were acquired over the period 2004–2016 by 15 CRS gauges at elevations ranging from 1350 to 2420 m a.s.l. (Table 1).

2.1.2. Remotely-sensed snow cover

Snow cover area data for the period 2000–2016 were extracted from the MOD10A1 (Terra) and MYD10A1 (Aqua) snow products Collection 5 (C5, Hall et al., 2007a,b) provided free of charge by the National Snow and Ice Data Center (NSIDC, http://nsidc.org). These daily snow products are derived from a Normalised Difference Snow Index (NDSI) calculated from the near-infrared and green wavelengths, and for which a threshold (0.4) was defined for the detection of snow. A complete description of the snow mapping algorithm can be found in Hall et al. (2002). It is worth noting that MODIS Collection 6 (C6, Hall et al., 2016a,b) represents the most recent release of global snow-cover mapping algorithms and could further increase the already high accuracy of previous collections. In MODIS C6, snow cover is reported by its NDSI values instead of in the form of simple binary information about snow cover, as was the case in previous products. This allows more flexibility in using the datasets for specific regions. Some authors (e.g. Da Ronco et al., 2020; Tong et al., 2020) recently investigated the sensitivity and accuracy of different NDSI thresholds used for snow cover mapping and compared the results with former snow cover classification based on a fixed NDSI threshold. They found that the NDSI thresholds vary seasonally, decrease with increasing elevation, and are lower in forested than open land cover settings. For instance, Tong et al. (2020) found that the NDSI thresholds fitted to different elevation and land cover classes improved snow cover mapping by 3–10 % in forested regions above 900 m a.s.l. in Austria. This sensitivity to the NDSI threshold thus makes its choice a critical step for the use of MODIS C6 snow products in hydrological studies. Given the limited improvement enabled by more complex parameterisation of the NDSI threshold in the C6 products, the C5 snow cover maps were used in the present study.

Cloud obscuration represents a significant limit for these products, which are generated from near-infrared and optical sensors. Consequently, the grid cells were gap filled to produce daily cloud-free snow cover maps using a method described and assessed in detail in Ruelland (2020). Briefly, the different classes in the original Terra and Aqua

Table 1
Streamflow gauging stations and main catchment characteristics, as well as associated SWE ground stations. Percentages of glacierised area were estimated from the Randolph glacier inventory version 6 (https://www.glims.org/RGI). Elevations were extracted from a digital elevation model (DEM) at 90 m spatial resolution from the Shuttle Radar Topographic Mission (SRTM, Jarvis et al., 2008).

#	Code	River	Hydrometric station	EDF SWE ground station (m a.s.l.)	Area (km^2)	Glacier cover (%)	Elevations(m a.s.l.)		
							Min	Mean	Max
1	X043401001	Ubaye	Barcelonnette	Passaur (2002)	549	0	1135	2214	3308
2	X045401001	Ubaye	Lauzet-Ubaye	Passaur (2002)	946	0	793	2083	3308
3	X001001001	Durance	Val-des-Pré	Chardonnet (2438)	207	0	1361	2221	3059
4	X001561301	Durance	Briançon	Chardonnet (2438) Izoard (2275)	548	1	1194	2187	3572
5	X013001001	Durance	Argentière	Chardonnet (2438) Izoard (2275) Cézanne (1877)	984	3	960	2178	4017
6	X031001001	Durance	Embrun	Chardonnet (2438) Izoard (2275) Les Marrous (2685) Cézanne (1877)	2170	1	783	2109	4017
7	X050551301	Durance	Espinasses	Chardonnet (2438) Izoard (2275) Les Marrous (2685) Passaur (2002) Cézanne (1877)	3580	1	749	2029	4017
8	W100000101	Isère	Val-d'Isère	Sous les Barmes (2350)	46	10	1851	2659	3538
9	W100000201	Avérole	Bessans	Plan Séti (2640)	45	14	1990	2871	3670
10	Q476102001	Gave de Pau	Lourdes	Barrada-Bugarret (2560) Migouelou (2260)	1070	0	378	1686	3251
11	O020002001	Garonne	Saint-Gaudens	Prat Long (1870) Spijeoles (2560)	2230	0	360	1456	3166
12	O125251001	Ariège	Foix	Font Negre (2200) Les Songes (2030)	1340	0	380	1542	3086
13	X220202201	Verdon	Demandoix	Maison Forestière (2020)	655	0	855	1644	2992

products were first merged into three classes: no-snow (no snow or lake), snow (snow or lake ice), no-data (clouds, missing data, no decision, or saturated detector). The missing values were then filled in three sequential steps: (i) Aqua–Terra merging to combine observations separated by several hours on the same day; (ii) temporal deduction by sliding the time filter up to 9 days; and (iii) spatial deduction by elevation and neighbourhood filter to gap-fill the remaining no-data grid cells. The resulting database consists of binary (snow/no-snow) daily maps at 500 m spatial resolution for the period 2000–2016. This gap-filling algorithm made it possible to reconstruct images with an average accuracy of 94 % estimated from confusion matrices (Ruelland, 2020).

2.1.3. Streamflow data

Daily observed streamflow measurements were extracted from the French hydrological database (Leleu et al., 2014) which gathers a permanent network of hydrometric stations and provides quality-controlled discharge data series in France. All the catchments available in the database, and where SWE observations existed within the limits of the catchments were initially retained. It was subsequently decided to select only catchments whose observed runoff was nearly natural (e.g. was not significantly influenced by dams), with good quality measurements according to the hydrological reports (quality is assessed for low, medium and high flows and classified as poor, medium or good), with less than 10 % daily missing values over the period 2004–2016. Finally, 13 catchments fulfilled these conditions and were retained for the present study (Table 1). The size of the catchments ranges from 45 to 3580 km^2, and their mean elevations from 1456 to 2871 m a.s.l. In these catchments, the streamflow regime is dominated by snowmelt with maximum discharges in spring and early summer.

2.2. Meteorological data

The SAFRAN (*Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige*) analysis scheme developed by Météo-France (Vidal et al., 2010) provides regionalised climate forcings using data from atmospheric and surface observations as well as data from climate models over France. SAFRAN is thus an atmospheric reanalysis which assimilates in-situ observations to calculate the energy fluxes as well as temperature, precipitation, humidity and wind on a regular 64 km^2 (8 km × 8 km) grid. These reanalysis data are available at hourly time steps from 1958 on. For the needs of the present study, SAFRAN daily precipitation and temperature data were used after being spatially aggregated at catchment area scale for use in the modelling experiment (see Section 3.1.1. for extrapolation of these mean areal inputs within each catchment based on elevation gradients).

Mean potential solar radiation (MJ/m^{-2} d^{-1}, i.e. direct and diffuse radiation under clear-sky conditions, hereafter *SR*) was computed based on a digital elevation model (DEM) from the Shuttle Radar Topographic Mission (SRTM, Jarvis et al., 2008), and using the approach developed by Kumar et al. (1997) to model topographic and seasonal variations in solar radiation. This approach uses a simple unweighted gradient of the four nearest neighbours for slope and calculates clear sky radiation corrected for the incident angle (terrain shading). Insolation depends on the day of the year, as well as on the latitude, elevation, slope and aspect, and was computed over the entire study domain based on the DEM resampled at 500 m spatial resolution.

3. Methods

This section presents the modelling framework used to estimate multi-objective and multi-variable parameters. It describes the hydrological model and the general assessment strategy used. Note that some additional and specific experimental details are deliberately presented in the results section to make the protocol easier to read and understand.

3.1. Hydrological model

3.1.1. Snow- and ice-accounting model

The snow- and ice-accounting routine (SIAR) developed by Ruelland (2023) was used. SIAR is an enhanced temperature index model derived from a degree-day scheme with additional processes. It was designed specifically to assist rainfall-runoff models in simulating mountain streamflow in a parsimonious way while including processes deemed necessary to represent the daily dynamics in SWE, SCA and streamflow.

SIAR can be run in fully distributed mode or according to elevation bands. As distributed (or semi-distributed) inputs, SIAR requires the daily liquid equivalent water depth of total precipitation (*P*), daily mean air temperature (*T*) and daily mean potential solar radiation (*SR*).

Following the recommendations made in Ruelland (2023), an adaptable number of elevation bands according to the catchment hypsometry was used in the present study. This distribution mode makes it possible to adapt the number of elevation bands to approximate the elevations of the local SWE measurements, thus facilitating their use for calibration and control. Additionally, a number of equal-area elevation bands according to the catchment hypsometry proved to be a good compromise as it allowed snow and runoff simulations whose accuracy was similar to that in full distribution mode, while limiting calculation times (Ruelland, 2023). In each catchment, the number of elevation bands (*EB*) of equal area was thus defined as $EB = (H_{max} - H_{min}) / 100$, where *H* represents elevation. *EB* ranged from 16 to 32 depending on the catchment. A unique parameter set is used in the catchment to apply the SIAR functions in each elevation band. The snowpack is represented by two internal states: the snow water equivalent (*SWE*) and the snowpack thermal state (*STS*). In each elevation band, these states vary independently according to differences in the input values.

Mean areal *SR* was extracted for each elevation band *i* in each catchment from the 500-m gridded *SR* dataset covering the entire study domain. Mean areal catchment *P* and *T* were extrapolated to each elevation band based on elevation gradients. Following the findings of Ruelland (2023), the temperature lapse rate (*TLR*) was set to a constant value of -0.62 °C (100 m)$^{-1}$ with no spatial and temporal variation linked to solar radiation (coefficient of variation, *CV*, set to zero) while no

Table 2
Parameters of the snow- and ice-accounting routine (SIAR) and their associated fixed values or ranges tested in the modelling experiment.

Par.	Meaning	Unit	Fixed values or ranges tested
PLR	Precipitation lapse rate	% (km)$^{-1}$	0
TLR	Temperature lapse rate	°C (100 m)$^{-1}$	-0.62
CV	Coefficient of variation applied to *TLR*	–	0
T_{50}	50 % rain–snow temperature threshold	°C	1
T_R	Thermal range for the phase separation around T_{50}	°C	4
SFA	Snowfall adjustment	%	[0; 150]
RLR	Rainfall lapse rate	% (km)$^{-1}$	0
K_{sub}	Fraction of *PE* retrieved when sublimation occurs	%	100
T_M	Temperature threshold for melt (melting point)	°C	0
θ	Weighting coefficient for the thermal state of the snowpack	–	[0; 1]
Kf	Constant part of the degree-day snow melt factor	mm °C^{-1} d^{-1}	[0; 4]
Ks	Variable part of the degree-day snow melt factor	mm °C^{-1} d^{-1}	[0; 4]
Kg	Ice degree-day factor indexed on solar radiation	mm °C^{-1} d^{-1}	3.8
T_{acc}	SWE threshold value used to compute FSC during accumulation	mm	60
Rsp	Fraction of the mean annual snowfall to compute FSC during melt	–	1
Number of free parameters authorised in this study			4

orographic gradient for total precipitation (*PLR* parameter set to 0 % $(km)^{-1}$) was applied (see Table 2 for parametrisation). Note, however, that solid precipitation (and therefore total precipitation) was finally made to increase with elevation after the separation phase with the *SFA* parameter (see following paragraph). Potential evaporation (*PE*) was computed in each elevation band using *T* and *SR* based on the formulation proposed by Oudin et al. (2005).

The accumulation phase in SIAR is determined by distinguishing the liquid and solid fractions of total precipitation. The liquid and solid fractions are thus calculated from a linear separation between −1 °C and +3 °C. Once the separation between snowfall (*S*) and rainfall (*R*) is computed, SIAR makes it possible to adjust *S* using a snowfall adjustment (*SFA*) which is a key parameter in addressing both snowfall undercatch and the orographic increase in solid precipitation (and hence total precipitation). For snow-hydrological simulations at operational scales in the Alps and the Pyrenees, Ruelland (2023) recommended adjusting solid precipitation (*SFA*) rather than applying an elevation gradient for total precipitation (*PLR*) before the separation phase and/or for rainfall (*RLR*) after the separation phase. Calibration of the *SFA* parameter was indeed both essential and sufficient to ensure the water balance at catchment scale (streamflow simulations) in addition to satisfactory SWE and snow cover simulations. This is discussed in more detail in Section 5.2.3. *PLR* and *RLR* can thus be ignored by fixing their values to 0 % (km^{-1}) (see Table 2), while snowfall (*S*) is adjusted with *SFA* as follows:

$$Sadj_t^i = S_t^i \times (1 + SFA) \tag{1}$$

Ablation is ensured by sublimation and melt processes. Sublimation is only possible when the mean air temperature is below the melting point T_M (see Ruelland (2023) for further details). Melting can be delayed by considering the thermal inertia of the snowpack: in this case. The snowpack thermal state (*STS*) is simulated from a parameter that gives more or less weight to the air temperature in the preceding time steps, based on a formulation proposed by Valéry et al. (2014) in which *STS* is calculated from a weighting coefficient θ between the value of the internal variable STS_{t-1}^i of the previous time step and the air temperature T_t^i of the day considered (see Eq. (2)). θ is typically calibrated between 0 (no thermal inertia) and 1 (maximum thermal inertia).

$$STS_t^i = \min\left(\left(\theta \times STS_{t-1}^i + (1-\theta) \times T_t^i\right), 0\right) \tag{2}$$

When *STS* reaches a value of 0 °C (i.e. when enough heat has been added to the snowpack to bring its heat content to 0 °C) and the mean air temperature T_t^i is above the melting point T_M (set to 0 °C), snowmelt is calculated from a degree-day melt factor (*MF*) that combines a constant part (*Kf*) and a variable part (*Ks*) indexed on potential solar radiation (without considering cloud cover) so that the melt factor can vary in space and over time:

$$MF_t^i = \begin{cases} \left(Kf + \left(Ks \times SI_t^i \times 2\right) \times \left(T_t^i - T_M\right)\right) & if\ T_t^i > T_M \\ 0 & otherwise \end{cases} \tag{3}$$

with

$$SI_t^i = \frac{SR_t^i - SR_{min}}{SR_{max} - SR_{min}} \tag{4}$$

where SI_t^i is a solar index with values ranging from 0 to 1, SR_t^i is the potential solar radiation $(MJ/m^{-2}\ d^{-1})$ which depends on the day of the year, the latitude, and the topographic characteristics in the elevation band *i*, SR_{min} and SR_{min} are, respectively, minimum and maximum solar radiation in the catchment.

A glacier component based on a degree-day formulation also indexed on potential solar radiation is activated when catchments contain glacierised areas (see Ruelland (2023) for further details).

In each elevation band, SIAR simulates a uniform distribution of SWE which can be compared to SWE measurements when ground stations are available in the elevation bands. Comparing the SWE model simulations with MODIS snow-covered area is less straightforward. The gridded representation of MODIS snow cover maps first requires computing the fractional snow cover (FSC) at the scale of each elevation band. Moreover, while the model simulates the volume of water stored as snow, FSC only indicates the proportion of snow-covered area in each elevation band. However, the relationship between SWE and FSC is subject to hysteresis. During accumulation, snow tends to fall over the entire catchment and snow-covered areas increase quickly and evenly throughout the catchment, whereas during the melting stage, they decrease gradually because snow remains longer at higher elevations and on north-facing slopes (Luce and Tarboton, 2004).

In the SIAR model, the simulated SWE is therefore transformed into FSC based on a bilinear depletion function which allows a more rapid increase in FSC during accumulation and a smoother decrease during ablation (see Ruelland (2023) for equations). The shape of the snow depletion curve is controlled by two parameters (T_{acc} and *Rsp*, see Table 2) and thus remains constant over time from year to year. A maximum snow-covered area is simulated at a T_{acc} value of SWE (set to 60 mm) during accumulation and at a T_{melt} value of SWE (set to *Rsp* × mean annual solid precipitation on elevation band *i*, with *Rsp* = 1) during melting. This is an efficient way of approximating the value of FSC from simulated SWE in which FSC increases linearly from 0 to 100 % for values of SWE under T_{acc} (T_{melt}) mm during accumulation (melting), and reaches 100 % at higher values. The T_{acc} and *Rsp* parameters were set to median values because these values produced good performances for the SWE/FSC relationship in the study catchments regardless of the values of the other parameters. T_{acc} was indeed relatively insensitive between values of 40 and 80 mm because of the rapid increase in snowfall during accumulation. Moreover, the SWE/FSC relationship during melting depends on the mean annual solid precipitation in each elevation band, which itself depends on the snowfall adjustment applied with the *SFA* parameter (a key parameter for simulations, see Ruelland, 2023). Consequently, a value of *Rsp* set to 1 was found to represent the SWE/FSC relationship well, whatever the snowfall adjustment applied and the values of the other calibrated parameters.

In the present study, SIAR was used with four free parameters identified as being the most sensitive for snow accumulation and melt (see Table 2): the snowfall adjustment (*SFA*), the weighting coefficient for the snowpack thermal state (θ), the constant part of the degree-day snow melt factor (*Kf*), and the variable part of the degree-day snow melt factor (*Ks*). All the other parameters were set to physical or general (median) values (see Table 2) that were evaluated during the development of the model (Ruelland, 2023).

3.1.2. Associated rainfall-runoff model

To ensure production and routing to simulate runoff at the catchment outlet, SIAR was associated with the well-known GR4J rainfall-runoff model (Perrin et al., 2003). This model was chosen for its low data requirements and parsimony (only four parameters to calibrate, see Table 3). The model was run at a daily time step and used in lumped mode with SIAR on top. The SIAR outputs (rainfall, melt and *PE*) of each elevation band were thus averaged at the catchment scale to feed GR4J.

Table 3
Parameters of the GR4J model and their ranges tested.

Parameter	Meaning	Unit	Ranges tested
X1	Maximum capacity of the production store *S*	mm	[0; 3000]
X2	Inter-catchment groundwater flows	mm d^{-1}	[-5; 2]
X3	Maximum capacity of the non-linear routing store *R*	mm	[0; 500]
X4	Unit hydrograph (UH) time base	d	[0.5; 2]

3.2. Model calibration and evaluation

3.2.1. Split-sample test

The simulation period was based on the hydrological years (counting from the 1st of September to the 31st of August) 2000–2016. The period when snow and runoff data were fully available (2004–2016) was split into two periods used alternately either for model calibration or evaluation. The second period (2010–2016) was wetter and warmer than the first (2004–2010) with differences in mean annual values of +9 % for precipitation, +0.4 °C for temperature and +16 % for runoff, on average in the catchments. Consistently, the maximum accumulated annual SWE was 11 % higher on average in the 15 CRS gauges over the second period. A 4-year spin-up period was used to initialise the model stores (before the start of the calibration or evaluation periods) and was consequently not used to compute the performance metrics.

3.2.2. Optimisation algorithm and objective functions

The parameters of SIAR and GR4J were optimised sequentially or simultaneously (see results section) using the shuffled complex evolution (SCE) algorithm (Duan et al., 1992). The algorithmic parameters of SCE were set to the values suggested by Duan et al. (1994) to avoid the risk of SCE converging to local optimal solutions. Optimisation was stopped based on either of the following two criteria: if the best objective function value had not improved by 0.01 % over the last five shuffling loops; if the total number of trials had reached 20,000.

Model calibration was carried out by calculating different efficiency criteria, each adapted to the nature of the observations with which the model simulations were compared. Optimisation was thus based on three main criteria C_Q, C_F and C_S. C_Q is a multi-purpose criterion focused on different runoff features (high flows, low flows and daily regime); C_F gives an overview of the reproduction of the FSC in each elevation band (notably the gradual increase in snow cover with elevation) as well as of their mean daily dynamics at the catchment scale; C_S evaluates the reproduction of the local SWE dynamics (accumulation and melting) within the catchment. Each criterion relied on the Nash-Sutcliffe efficiency (*NSE*) metric computed according to various aspects as follows:

$$C_Q = \overline{NSE_Q + NSE_{1/Q} + NSE_{regQ}} \tag{5}$$

$$C_F = \overline{NSE_{FSC} + NSE_{regFSC}} \tag{6}$$

$$C_S = NSE_{SWE} \tag{7}$$

with:

$$NSE_Q = 1 - \frac{\sum_{t=1}^{N}(Q_{sim,t} - Q_{obs,t})^2}{\sum_{t=1}^{N}(Q_{obs,t} - \overline{Q_{obs}})^2} \tag{8}$$

$$NSE_{1/Q} = 1 - \frac{\sum_{t=1}^{N}(1/Q_{sim,t} - 1/Q_{obs,t})^2}{\sum_{t=1}^{N}(1/Q_{obs,t} - 1/\overline{Q_{obs}})^2} \tag{9}$$

$$NSE_{regQ} = 1 - \frac{\sum_{j=1}^{366}(Q_{sim,j} - Q_{obs,j})^2}{\sum_{j=1}^{366}(Q_{obs,j} - \overline{Q_{obs}})^2} \tag{10}$$

$$NSE_{FSC} = \sum_{i=1}^{EZ}\left(\left[1 - \frac{\sum_{t=1}^{N}(FSC^i_{sim,t} - FSC^i_{obs,t})^2}{\sum_{t=1}^{N}(FSC^i_{obs,t} - \overline{FSC^i_{obs}})^2}\right] \times AB_i\right) \tag{11}$$

$$NSE_{regFSC} = 1 - \frac{\sum_{j=1}^{366}(FSC_{sim,j} - FSC_{obs,j})^2}{\sum_{j=1}^{366}(FSC_{obs,j} - \overline{FSC_{obs}})^2} \tag{12}$$

$$NSE_{SWE} = \frac{1}{CRS}\sum_{St=1}^{CRS}\left(1 - \frac{\sum_{t=1}^{N}(SWE^i_{sim,t} - SWE^{St}_{obs,t})^2}{\sum_{t=1}^{N}(SWE^{St}_{obs,t} - \overline{SWE^{St}_{obs}})^2}\right) \tag{13}$$

where $Q_{obs,t}$ and $Q_{sim,t}$ are the observed and simulated streamflows at daily time step t; N is the total number of time steps; $Q_{obs,j}$ and $Q_{sim,j}$ are the mean observed and simulated streamflow of Julian day j; $FSC^i_{obs,t}$ and $FSC^i_{sim,t}$ are the observed and simulated fractional snow-covered area (FSC) in elevation band i at daily time step t; $\overline{FSC^i_{obs}}$ is the mean observed FSC in elevation band i over the test period; AB_i is the relative surface area of elevation band i in the catchment; $FSC_{obs,j}$ and $FSC_{sim,j}$ are the mean observed and simulated FSC of Julian day j at the catchment scale; $SWE^{St}_{obs,t}$ and $SWE^i_{sim,t}$ are the observed and simulated snow water equivalents at the ground station *St* (one to five cosmic-ray snow sensors (*CRS*) are available in each catchment) and in the corresponding elevation band i, respectively.

Model calibration was based on the following three objective functions (hereafter named *FQ*, *SQ* and *SFQ*) which combine the runoff criterion (C_Q) with the criterion on fractional snow cover (C_F) and/or the criterion on local SWE dynamics (C_S), as follows:

$$FQ = (1-\alpha) \times (1 - C_F) + \alpha \times (1 - C_Q) \tag{14}$$

$$SQ = (1-\alpha) \times (1 - C_S) + \alpha \times (1 - C_Q) \tag{15}$$

$$SFQ = (1-\alpha) \times [0.5 \times (1 - C_S) + 0.5 \times (1 - C_F)] + \alpha \times (1 - C_Q) \tag{16}$$

where α is the weight with the following values tested: 100 % (optimisation on runoff only), 75 %, 50 %, 25 % and 0 % (optimisation on snow data only). For example, *FQ100*, *SQ100* and *SFQ100* correspond to objective functions where all the weight was given to the runoff criterion (optimisation on runoff only, hereafter also named *Q100*), whereas *FQ00*, *SQ00* and *SFQ00* correspond to objective functions where all the weight was given to the snow criterion (optimisation on snow data only).

For each objective function, a score of 1 indicates a perfect agreement between the observed and simulated values. The traditional NSE formulation (Eq. (8)) for runoff and Eq. (13) for SWE) is known to emphasise high values since it is based on squared errors. Hence model performance will tend to be biased towards maximum springtime snow accumulation, which was deemed to be a desirable property of the C_S criterion on SWE (Eq. (7)) since the amount of accumulated snow (and consequently snowmelt) has a direct impact on the annual peak flow in most years and catchments. On the other hand, to avoid this bias towards the highest values, the C_Q criterion on runoff (Eq. (5)) was designed to give the same weight to different runoff features (high flows, low flows and daily regime). Concerning the calculation of the C_F criterion on FSC, we followed the recommendations of Riboust et al. (2019) who analysed the impact of the weight given to each elevation band for the evaluation of FSC dynamics and concluded that equal weight should be given to each elevation band for the computation at the catchment scale (NSE_{FSC}, Eq. (11)). However, given the variable number of elevation bands (from 16 to 32 *EB* depending on the catchment hypsometry in the present study, whereas Riboust et al. (2019) limited the catchment distribution to five equal-area elevations bands), it was decided to also include a criterion on the average daily dynamics of FSC at the catchment scale (NSE_{regFSC}, Eq. (12)) in the calculation of the C_F criterion.

3.2.3. Evaluation criteria

The C_S, C_F and C_Q criteria were used to assess model performance over the independent evaluation periods. For each criterion, the maximum possible value is 1, which indicates perfect agreement between the observed and simulated variables. Additionally, a Friedman statistical test (Friedman, 1937) was applied to compare the performance distributions for each criterion. The model performances were ranked from 1 (considered to be the best performance) to values greater than 1, if the performances were judged by the test to be worse (with 95 % confidence).

4. Results

In the following sections, different sensitivity analyses are presented in response to the questions posed in the introduction. Snow and runoff performances are systematically analysed based on the statistical assessment of the entire dataset and, where appropriate, examples extracted from the dataset are also shown. It should be noted that all the results presented were obtained in independent evaluation periods (i.e. not in the calibration periods) with the exception of Sections 4.2.3 and 4.4, which deal specifically with the sensitivity of parameters to the objective function during optimisation.

4.1. Sequential versus simultaneous calibration of the snow model and rainfall-runoff model

Here the aim is to assess whether it is advisable to calibrate the snow model completely independently of the rainfall-runoff model. The specific purpose is to determine if the snow data are sufficient to calibrate the snow- and ice- accounting routine (SIAR), i.e. constraining its parameters without using runoff. To this end, a sequential calibration was tested. In this strategy, the four parameters of the SIAR model were first calibrated only on the FSC observations (*FQ00*), SWE observations (*SQ00*) or both FSC and SWE observations (*SFQ00*), and subsequently prescribed as fixed parameters in the coupled SIAR-GR4J model. Next, the four parameters of the GR4J rainfall-runoff model were calibrated only on runoff observations (*Q100*). The model performances obtained from the sequential calibration were compared to those obtained from simultaneous calibration of the snow model and the rainfall-runoff model using the *FQ50*, *SQ50* and *SFQ50* objective functions to enable a rigorous comparison because these functions give as much weight to the criteria of snow and runoff. The calibrations applied for the comparison between the sequential and simultaneous strategy are summarised in Table 4 and the results are shown in Fig. 1.

Runoff simulations were consistently better (best rank in C_Q with *FQ50*, *SQ50* and *SFQ50*) when the snow and rainfall-runoff models were calibrated simultaneously than when they were calibrated sequentially. On the other hand, the best simulations in SWE and FSC were achieved based on sequential calibration, with the best rank in C_S obtained when using *SQ00-Q100* and in C_F obtained when using *FQ00-Q100*. However, SWE and FSC simulations were not strongly altered by simultaneous calibration of the snow and runoff parameters. Indeed, the performance distributions in C_S and C_F did not differ much between the sequentially calibrated model and the jointly calibrated model. These findings show that sequential calibration of the snow parameters gives too much weight to snow compared to runoff. Integrating snow information though joint calibration appears to be more appropriate since it did not critically impair the quality of the SWE and FSC simulations, while leading to higher performance in runoff.

Table 4
Modelling tests to analyse the independence of the snow model from the rainfall-runoff model.

Mode	Code	Meaning
Sequential calibration	*FQ00 – Q100*	SIAR was first calibrated only on MODIS FSC (*FQ00*) and GR4J was then calibrated on runoff (*Q100*) using SIAR on top with the parameters optimised during the first calibration with *FQ00*.
	SQ00 – Q100	SIAR was first calibrated only on SWE (*SQ00*) and GR4J was then calibrated on runoff (*Q100*) using SIAR on top with the parameters optimised during the first calibration with *SQ00*.
	SFQ00 – Q100	SIAR was first calibrated only on SWE and MODIS FSC (*SFQ00*) and GR4J was then calibrated on runoff (*Q100*) using SIAR on top with the parameters optimised during the first calibration with *SFQ00*.
Simultaneous calibration	*FQ50*	The parameters of SIAR-GR4J were jointly calibrated on MODIS FSC and runoff using the objective function *FQ50*.
	SQ50	The parameters of SIAR-GR4J were jointly calibrated on SWE and runoff using the objective function *SQ50*.
	SFQ50	The parameters of SIAR-GR4J were jointly calibrated on SWE, MODIS FSC and runoff using the objective function *SFQ50*.

4.2. Sensitivity of the snow and runoff performances to the multi-criteria objective functions

4.2.1. Weight of the snow criteria in the objective function

If simultaneous calibration is a recommendable strategy, we still need to know what weight should be given to snow observations in the objective function. Fig. 2 shows the performance distributions obtained with the snow model and the rainfall-runoff model calibrated simultaneously based on the three objective functions (*FQ*, *SQ* and *SFQ*) and different weights (α) tested.

Compared with the model only calibrated against runoff (*Q100*), integrating snow information in the calibration process significantly improved the snow performances (both in C_S and C_F) and did not impair the runoff performances (in C_Q) with *FQ75*, *SQ75* and *SFQ75*. Indeed, the Friedman rank for runoff (C_Q) was best when the runoff weight was set to 75 % in the objective function and the snow simulations (C_S and C_F) were improved compared to calibration against runoff alone. Snow simulations can be further improved by reducing the weight of the runoff criterion in the objective function but at the expense of the runoff simulations. Hence, better performances in C_S and C_F were reached by reducing the weight of the runoff criterion to 25 % in the objective function (with *FQ25*, *SQ25* and *SFQ25*) but the performances in runoff were negatively affected with a Friedman rank only in the third position. Finally, using only the snow criterion in the objective function (with *FQ00*, *SQ00* and *SFQ00*) clearly led to poor runoff simulations. Not surprisingly, this shows that runoff data are required for calibration and that calibrating the model only against snow data is not sufficient.

As a hydrologist, considering that runoff is the most important variable to simulate and that snow data such as local SWE measurements and MODIS-based snow cover are only auxiliary data, the *FQ75*, *SQ75* and *SFQ75* objective functions can be considered as good compromises to constrain the model by integrating snow information. These multi-criteria functions appeared to be appropriate as they improved the model realism with no negative consequences for runoff performance. To paraphrase Kirchner (2006), they made it possible to get “the right answer” (here, runoff simulations of similar accuracy) for “the right reason” (here, more reliable representation of underlying processes as evidenced by improved snow simulations).

4.2.2. Looking for the best-compromise objective function

To go further, one may wonder if there is a best-compromise approach among the objective functions retained: *Q100*, *FQ75*, *SQ75* and *SFQ75*. Fig. 3 compares the performance distributions obtained in evaluation with the model calibrated using the four objective functions.

Overall, there were no significant differences in the runoff performances (C_Q) between *Q100*, *FQ75*, *SQ75* and *SFQ75*. On the other hand, snow simulations differed depending on the objective function. Hence, logically, the best performances in C_S (local SWE simulations) and C_F (snow cover simulations) were achieved with the *SQ75* and *FQ75* objective functions, respectively. This shows that trade-offs exist between the different criteria included in the objective functions and that it is difficult to optimise all the criteria simultaneously. However, the differences in performances between *FQ75* and *SQ75* were much more significant for the C_S criterion (mean value of 0.70 with *FQ75 vs*. 0.81 with *SQ75*) than for the C_F criterion (mean value of 0.84 with *FQ75 vs*. 0.82 with *SQ75*). This clearly shows that the SWE simulations were improved by using the C_S criterion in the objective function with only

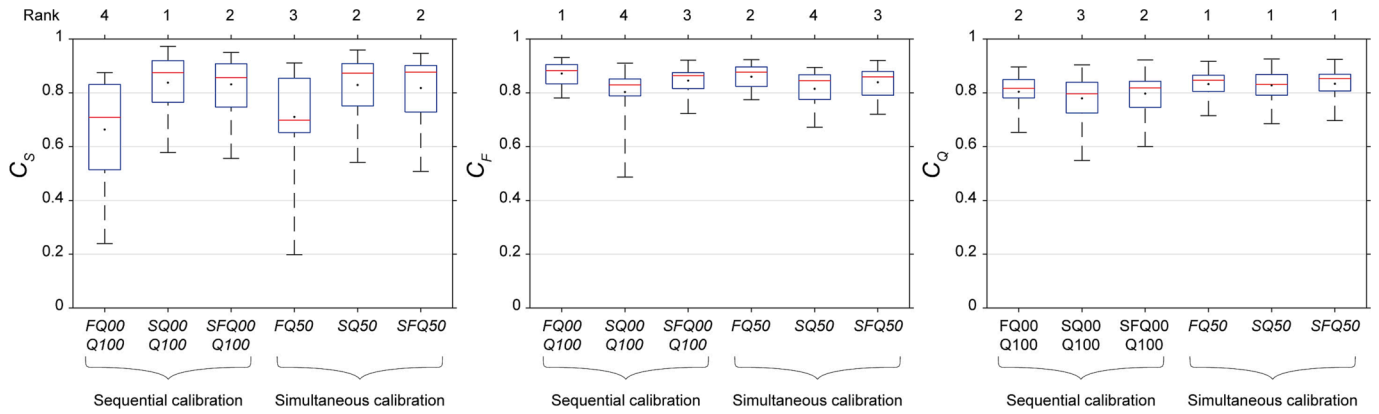


Fig. 1. Boxplots (showing 0, 0.25, 0.50, 0.75 and 1 percentiles) of the performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments using the SIAR/GR4J model with sequential calibration (*FQ00-Q100*, *SQ00-Q100*, *SFQ00-Q100*) and joint calibration (*FQ50*, *SQ50*, *SFQ50*). In the sequential mode, the snow model (SIAR) was first calibrated only against the FSC (*FQ00*) observations, or only against SWE observations (*SQ00*) or against both (*SFQ00*), and the rainfall-runoff model (GR4J) was then calibrated against runoff observations (*Q100*) using the snow parameters (*SFA*, θ, *Kf* and *Ks*, see Table 2) optimised during the first calibration. In the simultaneous mode, the parameters of SIAR-GR4J were calibrated jointly using the *FQ50*, *SQ50* and *SFQ50* objective functions (see Eqs. (14) to (16)).

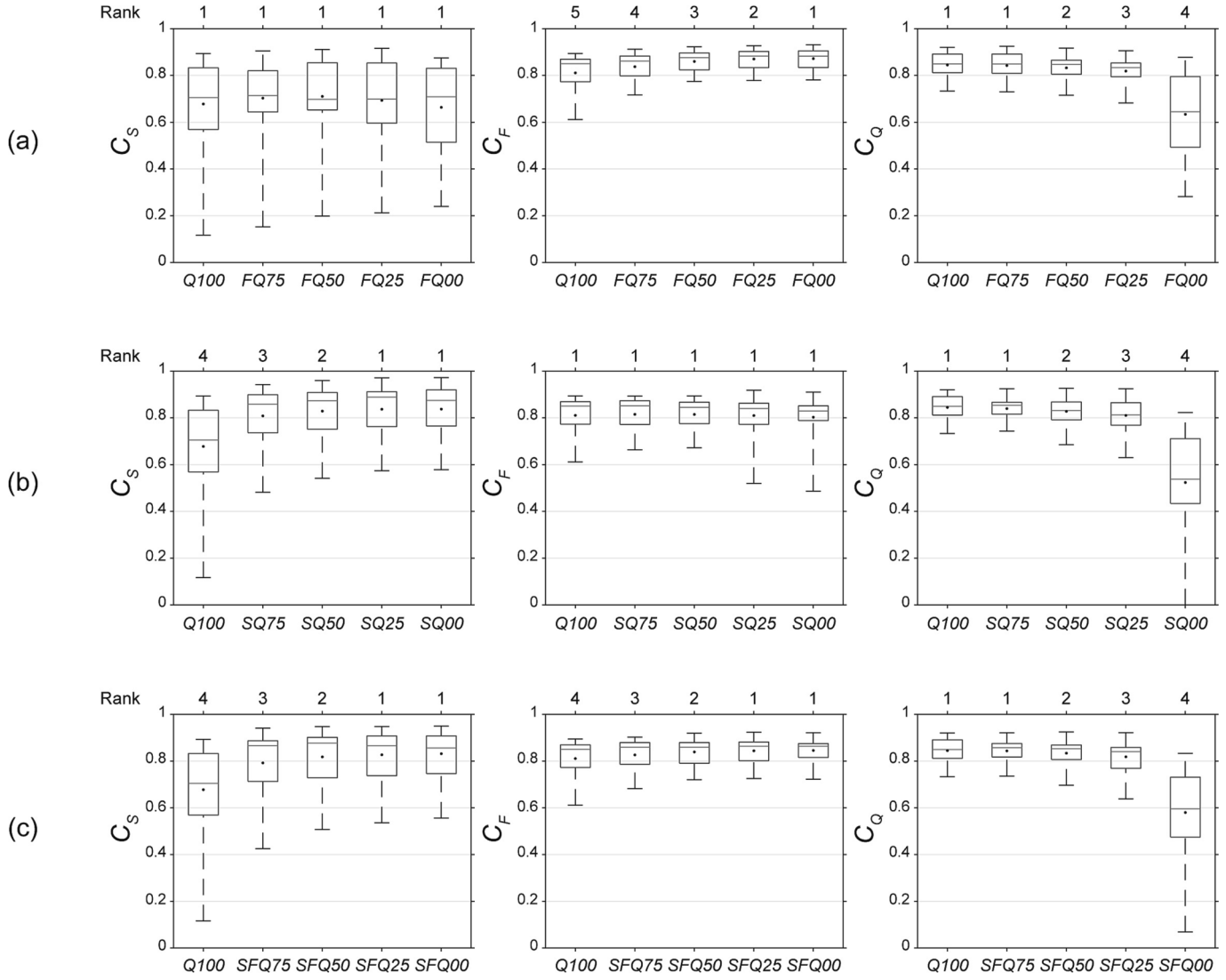


Fig. 2. Performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments using the GR4J model combined with SIAR according to the three objective functions (a) *FQ*, (b) *SQ* and (c) *SFQ* and their associated weight α tested: 100 % (optimisation on runoff only, indicated as *Q100*), 75 %, 50 %, 25 %, and 0 % (optimisation on snow data only).

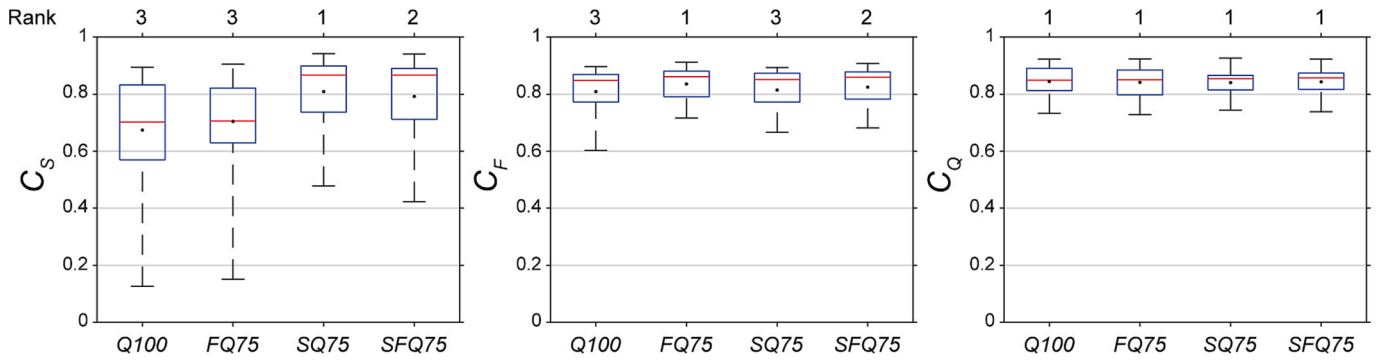


Fig. 3. Performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments using the GR4J model combined with SIAR according to four objective functions: *Q100*, *FQ75*, *SQ75* and *SFQ75*.

limited impairment of the snow cover simulations, whereas using the C_F criterion alone led to less realistic SWE simulations and only a slight improvement in snow cover simulations. The *SQ75* objective function therefore appears to be more useful than the *FQ75* objective function for calibrating the model. Additionally, the *SFQ75* objective function is the best compromise to jointly simulate SWE, snow cover and runoff.

4.2.3. *Trade-offs within a multi-objective framework*

These findings call for additional analysis of the trade-offs between the criteria (C_S, C_F and C_Q) used in the objective functions. It is worth noting that the weight of C_Q was set to 75 % in the objective functions to avoid altering the runoff simulations when using snow data. Trade-offs between the different criteria can be further highlighted by mapping Pareto fronts in the space of performance measures. A representative example to illustrate this point is presented in Fig. 4. The figure shows the Pareto-optimal solutions obtained for the Avérole catchment at Bessans during the 2010–2016 calibration period, and plotted in two dimensions for different combinations of two of the three criteria (C_S, C_F and C_Q) used in the objective functions. Trade-offs between the runoff and snow cover criteria (C_Q *vs.* C_F graph) and between the snow cover and local SWE criteria (C_F *vs.* C_S graph) are clearly less important than trade-offs between the runoff and SWE criteria (C_Q *vs.* C_S graph). This means that the model is less efficient at simultaneously reproducing runoff and local SWE than in reproducing the other pairs of observed variables, and also that the snow cover criterion (C_F) is less helpful in identifying the parameters during optimisation. This is because the C_F criterion presents less variance (the daily fractional snow cover ranges from 0 to 100 %) than the other criteria. Considering the optimal solutions among the 3D Pareto front according to *FQ75*, *SQ75* and *SFQ75*, it is clear that the calibration performances in runoff (C_Q) are very similar and reach the best C_Q values whatever the objective function. This is explained by the fact that the runoff weight was set to 75 % in the objective functions. On the other hand, considering the criteria C_F and C_S (C_F *vs.* C_S graph), this weight leads to solutions that are not Pareto-optimal in order to avoid negative consequences in the runoff simulations.

4.3. *Improvement or impairment of the performance for each individual catchment*

Fig. 5 shows the differences in performance between the model calibrated using snow data (*FQ75*, *SQ75* and *SFQ75*) and the model calibrated using only runoff data (*Q100*) for each catchment and each independent evaluation period. For each criterion (C_S, C_F and C_Q), the differences are expressed by delta values ranging from −0.08 to + 0.36. A positive value (in blue) indicates an improvement in performance while a negative value (in red) indicates an impairment. Compared to *Q100* (calibration on runoff only), 54 % (46 %) of the runoff simulations

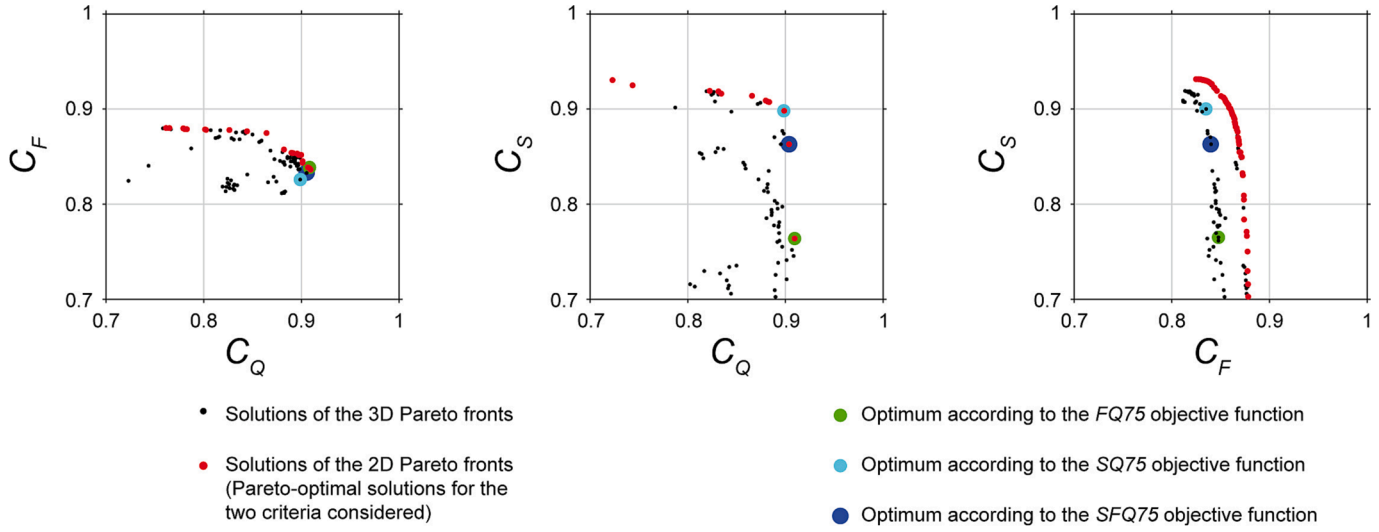


Fig. 4. Projections of the Pareto fronts of the model calibration onto three possible two-dimensional subspaces of the objective space. Example of the calibration in the Avérole catchment at Bessans (Alps) over the 2010–2016 period within a multiple-hypothesis framework against C_Q, C_F and C_S using the non-dominated sorted genetic algorithm II (NSGAII, see Deb (2002) for further details on the algorithm). Back dots: solutions of the 3D Pareto fronts. Red dots: solutions of the 3D Pareto front which remain Pareto-optimal (i.e. non-dominated) for the two criteria considered. The green, cyan and blue dots correspond to the optimum solution according to the *FQ75*, *SQ75* and *SFQ75* objective functions, respectively.

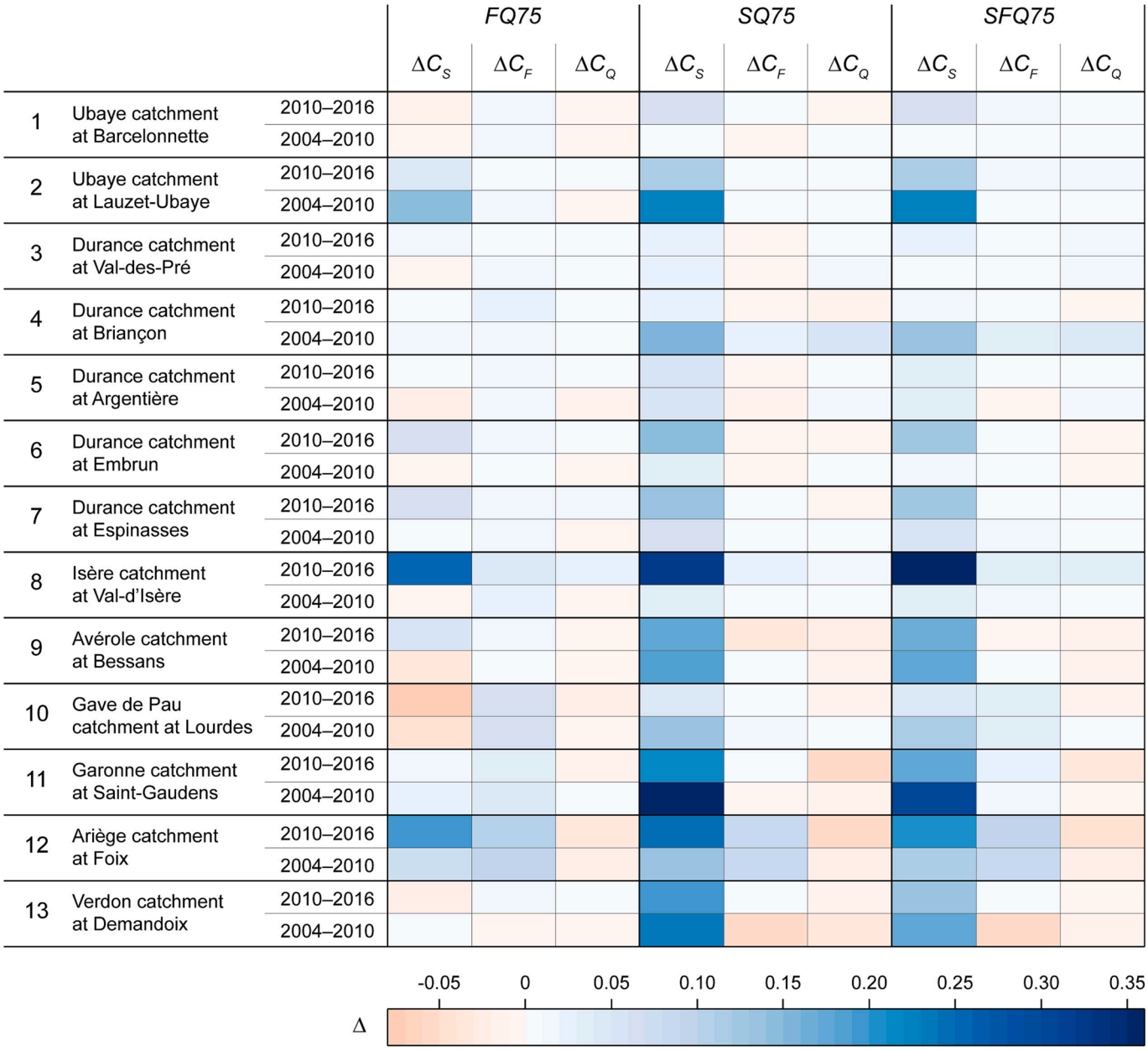


Fig. 5. Improvement (or impairment) of the performance criteria (C_S, C_F and C_Q) for the 13 catchments studied and the two evaluation periods (2004–2010 and 2010–2016). The indicated value ΔC is the difference between the performance of the model calibrated using snow data (with *FQ75*, *SQ75* or *SFQ75*) minus the performance of the model calibrated using only runoff data (*Q100*).

(C_Q) were improved (impaired) by using *SFQ75*. It is important to note that this improvement (impairment) is not statistically significant as already shown in Fig. 3. As far as the C_Q criterion is concerned, comparable results were obtained using *FQ75* and *SQ75*. This explains why the runoff simulations have performances comparable to those obtained when the model was only calibrated against runoff. Ninety-six percent of the FSC simulations (C_F) were very slightly improved by using *FQ75* (combination of FSC and runoff in the objective function). However, the most noticeable improvement was obtained for local SWE simulations with the use of the *SQ75* and *SFQ75* objective functions: 100 % of the SWE simulations (C_S) were improved by integrating the SWE dynamics in the objective function. The improvement was generally highly significant with an average increase in the C_S criterion of 0.13 (0.11) points with *SQ75* (*SFQ75*). It is interesting to note that the results are consistent regardless of differences in climatic and physiographic conditions as well as the size of the catchments studied. The *SFQ75* objective function can still be seen as a good compromise for calibrating the model with snow data since it improved respectively, 100 %, 88 % and 54 % of the SWE, FSC and runoff simulations.

Representative examples of SWE, FSC and runoff simulations are shown in Fig. 6 for the evaluation period 2004–2010 (i.e. calibration was performed using the period 2010–2016). The five catchments presented are the Ubaye at Lauzet-Ubaye (946 km^2, Alps), the Durance at Briançon (548 km^2, Alps), the Isère at Val-d'Isère (46 km^2, Alps), the Avérole at Bessans (45 km^2, Alps), and the Gave de Pau at Lourdes (1070 km^2, Pyrenees). Depending on the catchment, one or two SWE gauges are available. Compared to the objective function based only on runoff (*Q100*), there was a clear improvement in the SWE simulations in the three catchments when the calibration was based on the objective function integrating the C_S criterion (*SFQ75*). Including SWE data during calibration thus made it possible to reproduce peak snow accumulation and changes in local snow volumes more accurately. Interestingly,

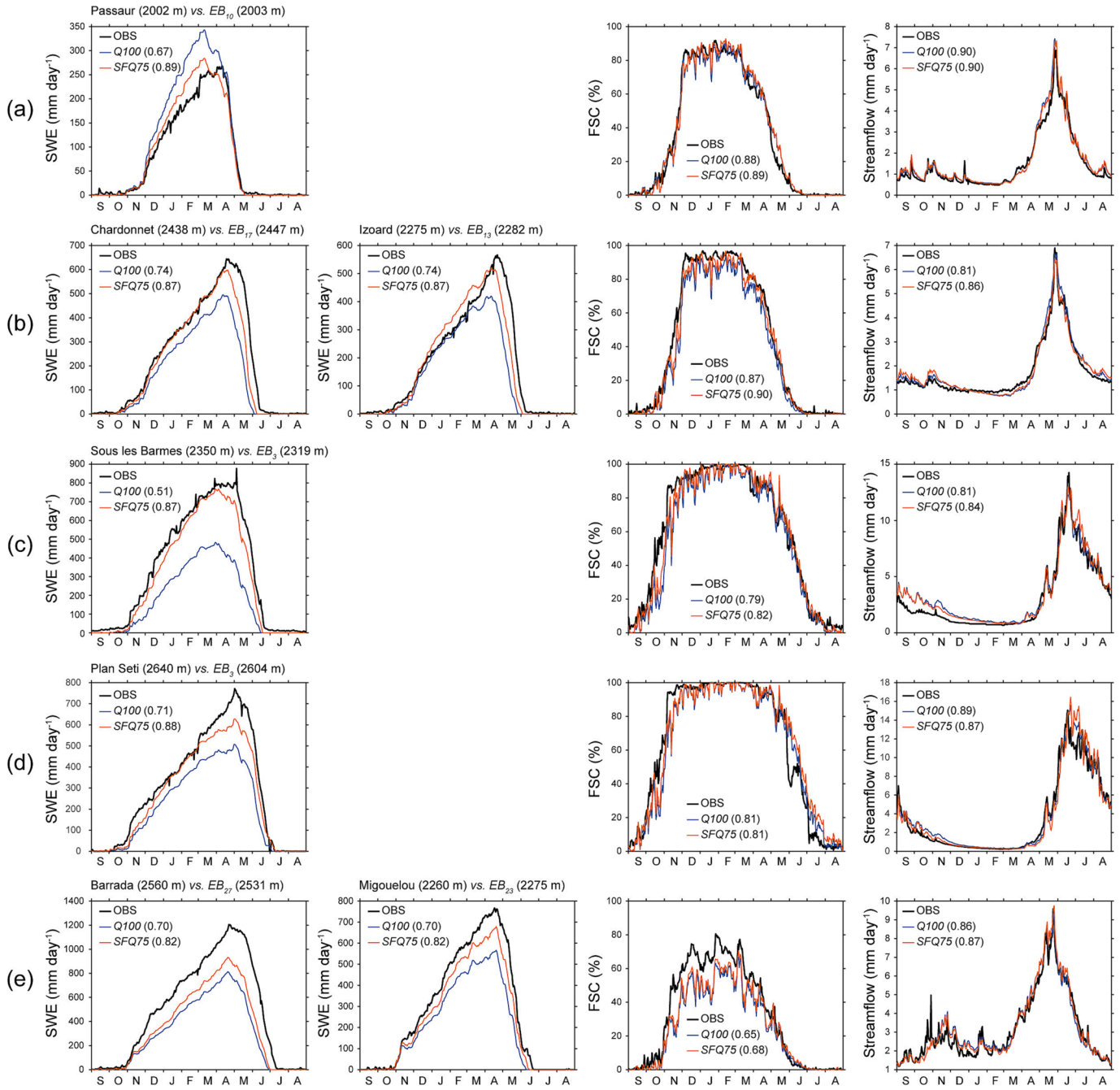


Fig. 6. Comparison of SWE, FSC and streamflow simulations for the evaluation period 2004–2010 using *Q100* and *SFQ75* as the objective function for model calibration: (a) in the Ubaye catchment at Lauzet-Ubaye (946 km^2, Alps), (b) in the Durance catchment at Briançon (548 km^2, Alps), (c) in the Isère catchment at Val-d'Isère (46 km^2, Alps), (d) in the Avérole catchment at Bessans (45 km^2, Alps), and (e) in the Gave de Pau catchment at Lourdes (1070 km^2, Pyrenees). For practical reasons, the simulations are presented here according to mean daily dynamics. Note, however, that the evaluation metrics were calculated according to the interannual daily series. Hence, the numbers in the graphs represent the values obtained for each evaluation criterion, i.e. C_S for SWE (Eq. (5)), C_F for fractional snow cover (Eq. (6)) and C_Q for runoff (Eq. (7)). Note also that, for a given objective function, the value of C_S is the same for all the SWE gauges in the catchment, since C_S is computed as the mean NSE_{SWE} of the SWE gauges in the catchment (see Eq. (13)). EB_i stands for elevation band *i* whose median elevation (in m a.s.l.) in the closest to the elevation (in m a.s.l.) of the SWE station used to compute the model performance in C_S.

this improvement involved a less significant overestimation of the cumulated SWE in the Ubaye catchment (Fig. 6a) and a less significant underestimation of the SWE in the other catchments (Fig. 6b, 6c, 6d and 6e). Depending on the catchment, calibrating the model on runoff alone then led to over- or under-estimation of the contribution of snow to streamflow (see Section 5.1 for further discussion of this point). The *SFQ75* objective function generally also led to FSC and runoff simulations of similar (if not slightly better) accuracy compared to the *Q100* objective function. This shows how useful it is to integrate in-situ SWE observations in the calibration to achieve better internal consistency of the model without negatively affecting simulated streamflow at the catchment outlet.

4.4. Sensitivity of the parameters to the objective function

Here, the aim is to analyse how the parameters are affected by the

different objective functions and whether the use of snow data in a multi-objective optimisation makes them easier to identify. The sensitivity of the parameters to different objective functions is shown in Fig. 7.

Certain parameters appear to be almost unaffected by the objective function: this is the case of the maximum capacity of the production store (*X1*) and the unit hydrograph time base (*X4*), which have similar distributions and median values whatever the objective function. In contrast, some parameter values changed considerably depending on the objective functions. Hence, the cold-content parameter (θ) values are higher with the model calibrated using FSC observations (median equal to 0.49 with *FQ75* and to 0.30 with *SFQ75*) than with the model calibrated without FSC observations (median equal to 0.01 with *Q100* and to 0.02 with *SQ75*). This shows that, to simulate the FSC dynamics accurately in most catchments, thermal inertia should not be neglected, whereas it needs to be more restricted for SWE and runoff simulations. Moreover, using SWE observations in the objective function (as was the case with *SQ75* and *SFQ75*) led to higher values of the snowfall adjustment (*SFA*), lower values of the variable part of the degree-day snow melt factor (*Ks*), lower values of the maximum capacity of the non-linear routing store (*X3*), and higher values of the inter-catchment groundwater flows (*X2*). It is difficult to comment further on the values of these parameters which have limited physical significance because they cannot be assessed from direct measurements. Nevertheless, concerning the *X2* parameter of GR4J, the values obtained with *SQ75* and *SFQ75* can be considered as being more satisfactory because they tend to avoid reaching the lower bound of the range recommended by the authors of the model (Perrin et al., 2003). The differences in the parameter values mainly show that the parameters can offset each other depending on the objective functions as a result of the trade-off between the SWE, FSC and runoff simulations. As demonstrated by the performance distributions analysed above, the parameter sets optimised using each of the objective functions are equally acceptable for the runoff simulations but not for the SWE and FSC simulations. In particular, it is possible to identify sets of parameters that lead to more accurate snow simulations with no negative effect on runoff simulations. This suggests that considering snow constraining data sets in the calibration phase leads to more plausible parameter values.

However, this gain in internal realism was not accompanied by improved computational efficiency since the model calibrated jointly on snow and runoff data did not converge faster. Indeed, the number of SCE trials for optimisation was almost the same (around 8 000 trials on average to reach convergence for each calibration) whatever the objective function used (see Table 5). Moreover, the median frequency of the optimised parameter values (blue values in Fig. 7) did not significantly differ among the objective functions. Parameter identifiability was only enhanced for parameter θ (the weighting coefficient for the thermal state of the snowpack), which had a higher frequency value (54 %) with the *FQ75* objective function. This suggests that parameter identifiability was not necessarily enhanced by integrating snow data in the objective function.

To further illustrate this point, Fig. 8 shows a representative example of parameter sensitivity to the objective functions *Q100* and *SFQ75* during optimisation. Certain values of the optimised parameters (in red) differ depending on whether they were constrained on runoff alone (*Q100*) or jointly on snow and runoff data (*SFQ75*). However, the mean frequency of the optimised parameter values is similar regardless of the objective function used (58 % with *Q100* versus 57 % with *SFQ75*). In this example, the *SFQ75* objective function only improved the identifiability of the parameters *SFA* and *X4* while it altered the identifiability of the other parameters. This again shows that equifinality was not necessarily reduced by using auxiliary data in the calibration, although it did allow for more reliable snow simulations.

In an attempt to better understand the problem of equifinality, parameter interdependence was investigated. Fig. 9 shows a cross-correlation analysis of the calibrated parameters with the objective functions *Q100* and *SFQ75*. When the model was calibrated on runoff alone (*Q100*), statistically significant correlations (p-values < 0.05) were found between the rainfall-runoff parameters (*X1/X2*, *X2/X4*, *X1/X3*) as well as between the snow and rainfall-runoff parameters (*SFA/X1*, *Kf/X3*, *Ks/X2*, *Ks/X3*), but not among the snow parameters. When the model was calibrated jointly with snow and runoff data (*SFQ75*),

Table 5
Mean number of SCE trials to reach convergence during calibration using either the NSE or the KGE metric to compute the objective functions.

	Q100	*FQ75*	*SQ75*	*SFQ75*
NSE	8838	8986	7869	7759
KGE	10526	10182	9445	9673

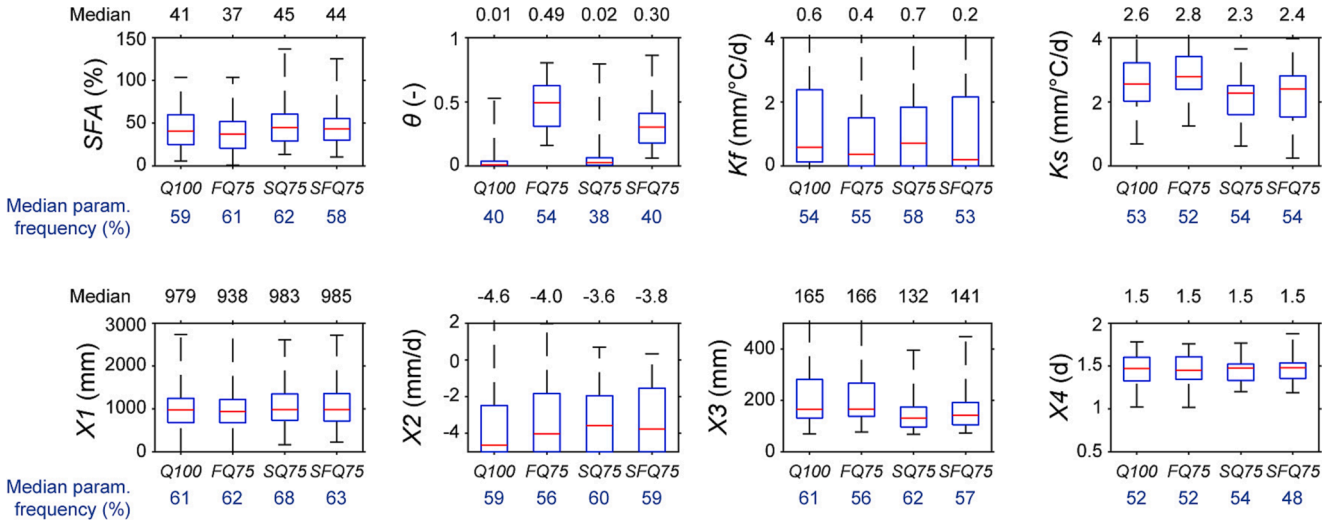


Fig. 7. Parameter distributions obtained during calibration over the 2004–2016 period in the 13 catchments tested ($n = 26$, i.e. 13 catchments × 2 calibration periods) according to four objective functions (*Q100*, *FQ75*, *SQ75* and *SFQ75*). The identifiability of the parameters (in blue) refers to the median frequency (in %) of the optimised parameter values obtained for each parameter among the SCE trials during calibration (see Fig. 8 for more details on the computation). The parameter identifiability is an indicator on the homogeneity of the parameter sets: the higher the value, the lower the equifinality and the easier it is to identify the optimised parameter.

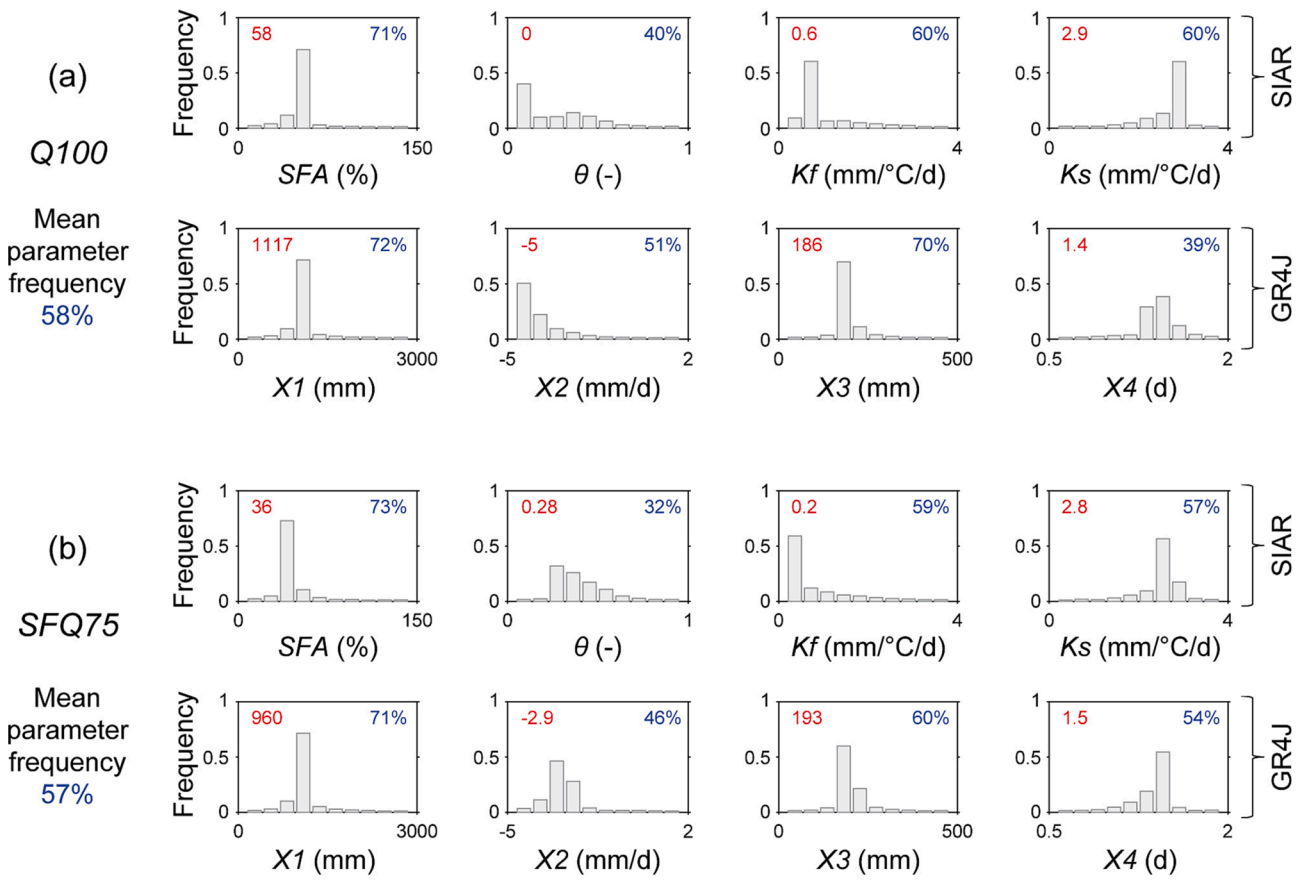


Fig. 8. Identifiability of the parameters with the objective functions (a) *Q100* and (b) *SFQ75* during the simultaneous calibration of the SIAR/GR4J model over the 2004–2010 period in the Durance catchment at Espinasses (3580 km^2, Alps). For each parameter, the bars represent the frequency of 10 classes of parameter values obtained among the SCE trials during calibration. The values in red represent the optimised calibrated parameters which minimise the objective function. The values in blue represent the frequency (in %) of the class corresponding to the optimised parameters. They summarize the identifiability of the parameters: the higher the value, the lower the equifinality and the easier it is to identify the optimised parameter.

parameter correlation was reduced among the rainfall-runoff parameters (with only the *X1/X2* correlation remaining), while the correlation increased significantly between the snow parameters (*θ/Kf*, *θ/Ks*, *Kf/Ks*). This counterbalancing effect explains why equifinality was not significantly reduced between the two calibration strategies. It also suggests that certain snow model parameters (*θ*, *Ks* and *Kf*) could be set to general values to limit equifinality without affecting the quality of the simulations (see Section 5.2.3). On the other hand, whatever the calibration strategy used, a significant covariation persisted between the snowfall adjustment parameter (*SFA*) and certain parameters of the rainfall-runoff model. If the use of snow data in the calibration made it possible to slightly reduce the correlation with the production store capacity (*X1*), it led to an additional dependency with the routing store capacity (*X3*). This interdependence cannot be avoided because the *SFA* parameter makes it possible to adjust snowfall (and hence total precipitation) thus resulting in an adjustment of melt volumes which has to be offset by the production and routing processes.

4.5. Model robustness to long-term climate variability

To complete the analysis of the calibration of the model with snow observations, it is interesting to know whether the model calibrated against snow data is more robust to long-term climatic variability than the model calibrated on runoff alone. To answer this question, the two strategies were evaluated over four past independent 10-year periods using the parameter sets calibrated over the periods 2004–2010 and 2010–2016. The four past evaluation periods used were 1960–1970, 1970–1980, 1980–1990 and 1990–2000. Compared to the recent period used for calibration (2004–2016), the four evaluation periods were wetter and colder, with differences in mean annual values of +7.3 %, +6.7 %, +1.3 % and +7.6 % for total precipitation and −1.0 °C, −1.1 °C, −0.3 °C and −0.6 °C for temperature, on average in the catchments. Only catchments with runoff measurements for these four time periods were used (between 85 % and 100 % of the sample used in this article depending on the period). Fig. 10 presents the performance distributions in C_Q using the model calibrated only against runoff (*Q100*) and the model calibrated against snow data (*FQ75*, *SQ75* and *SFQ75*) for each evaluation period. The C_S and C_F evaluation criteria could not be computed because SWE gauges and MODIS observations were not available for the past periods.

The results show that the model calibrated on snow data (*FQ75*, *SQ75* and *SFQ75*) did not produce better runoff simulations far from the calibration period than the model calibrated on runoff alone (*Q100*). This suggests that using snow observations in the calibration process does not improve the robustness of the model to long-term climate variability, although it did lead to greater internal consistency in the snow simulations over the recent period.

5. Discussion

5.1. Main findings and recommendations

This study assessed the value of local SWE measurements and MODIS-based snow cover in addition to observed streamflow to improve the consistency and robustness of a semi-distributed hydrological model in mountain catchments. The results were obtained here with a single

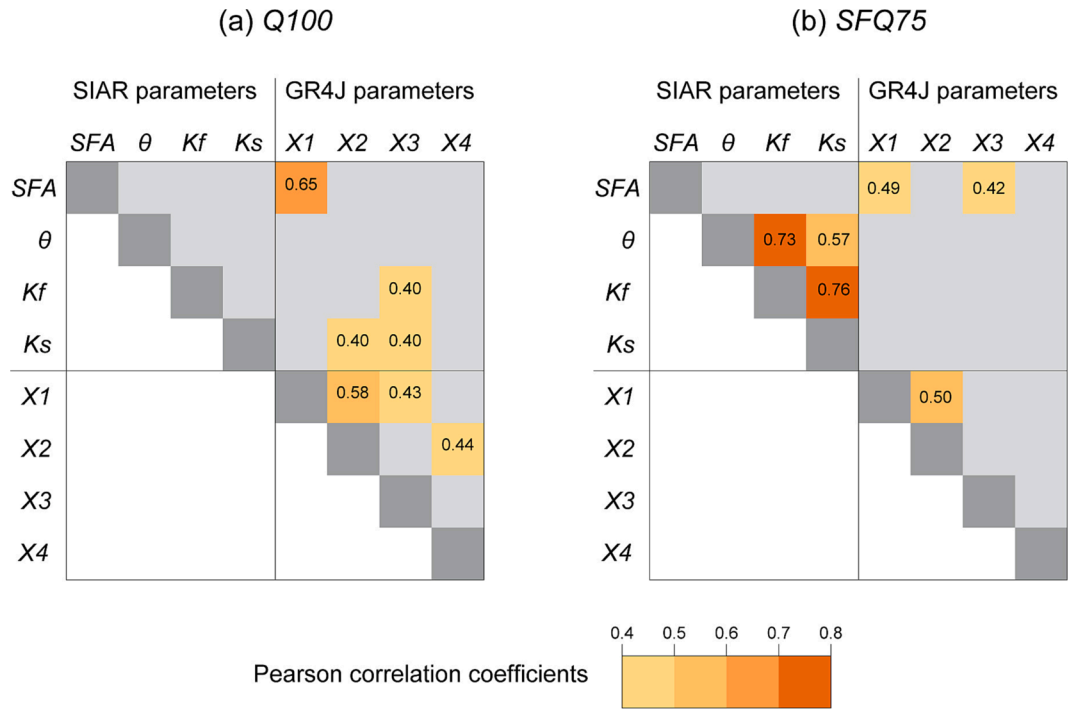


Fig. 9. Pearson correlation coefficients between the optimised parameters obtained using the simultaneous calibration of the SIAR/GR4J model with the objective functions (a) *Q100* and (b) *SFQ75* over the 2004–2016 period in the 13 catchments tested. Note that only statistically significant correlations (*p*-values < 0.05) are shown.

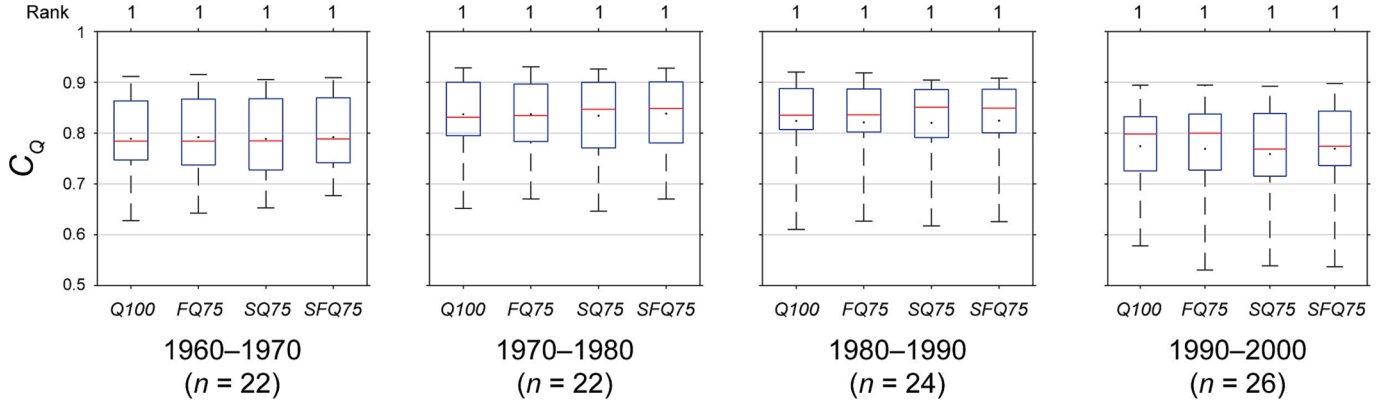


Fig. 10. Performance distributions (in C_Q) obtained over four past periods (1960–1970, 1970–1980, 1980–1990 and 1990–2000) in the tested catchments using the model (SIAR-GR4J) calibrated against runoff only (*Q100*) and partly against snow data (*FQ75*, *SQ75* and *SFQ75*). The parameters used here come from the calibrations over the 2004–2010 and 2010–2016 periods. *n* refers to the number of evaluation exercises (i.e. two, corresponding to the parameter sets optimised over the 2004–2010 and 2010–2016 periods, respectively) multiplied by the number of catchments in which runoff observations were available in the decade concerned.

model and using a catchment sample in the French Alps and Pyrenees at operational scales (13 catchments whose surface area varies from 45 to 3580 km^2). Further tests may be necessary to extrapolate these results to other models and other hydro-climatic conditions. Nevertheless, they already provide guidelines for the use of snow data combined with runoff in the parameter estimation process when using hydrological models to simulate the internal behaviour of mountain catchments.

The sequential calibration of the snow parameters appeared inappropriate because it gave too much weight to snow compared to runoff, in agreement with the results obtained by Nemri and Kinnard (2020) using SWE measurements in addition to runoff in non-mountain catchments. Instead, incorporating snow data for simultaneous calibration of the snow model and the rainfall-runoff model is recommended, because this strategy improved snow simulations without impairing streamflow performance compared to the model calibrated against runoff alone. Results thus show that runoff simulations of similar accuracy can be obtained with fewer snow errors if snow data (SWE measurements and/or MODIS snow cover) are included in the calibration, in line with the results of previous studies using snow cover (Parajka and Blöschl, 2008; Finger et al., 2011; Franz and Karsten, 2013; Duethmann et al., 2014) or local SWE measurements (Tuo et al., 2018; Nemri and Kinnard, 2020). The results then show that using snow observations during calibration of the model can be very effective to achieve greater internal consistency.

With this improvement, a new calibration method was developed using in-situ SWE, snow cover and runoff observations combined. This was made possible by limiting the weight of the snow criteria to 25 % in the objective functions tested. Similar results were reported by Parajka et al. (2007) and Riboust et al. (2019) on the weight to be attributed to snow cover used in combination with runoff in the calibration process. Interestingly, this weight also appears to be suitable for the joint use of

SWE and runoff data as well as of SWE, snow cover and runoff data. Snow simulations could be further improved by increasing the weight of the snow criteria in the objective function but to the detriment of runoff simulations. Indeed, as shown by the statistical evaluation of the model performance in independent evaluation periods as well as the analysis of Pareto-optimal solutions within a multi-objective framework, there were trade-offs between the different criteria included in the objective functions, making it difficult to simultaneously optimise all the criteria. An objective function focusing on SWE, snow cover, and runoff (*SFQ75*) was identified as the best compromise to constrain the model parameters in the catchments studied here. Incorporating SWE measured in situ in the calibration process appeared particularly useful. Indeed, using local SWE measurements (with or without snow cover observations) combined with streamflow records enabled us to reproduce much more realistic local SWE dynamics without significantly altering the snow cover simulations in the catchment or the simulated runoff at the catchment outlet. Conversely, using snow cover observations (without SWE measurements) during optimisation significantly impaired local SWE simulations with only a slight improvement in the snow cover simulations. This shows that snow covers are easier to reproduce than local SWE dynamics whatever the model parametrisation, in particular because the snow cover criterion presents less variance (values remain between 0 and 100 %) than the SWE criterion. This also suggests that spatial information on snow cover is less valuable than volumetric information on local accumulated SWE in identifying parameters that are important for the overall water balance of the catchment. Despite the fact that such point-scale snow observations have limited spatial representativeness, their use was clearly beneficial to simulate catchment-scale snow dynamics even without a dense network of stations (in the present study, only between one and five SWE gauges were available in each study catchment).

It is worth noting that the use of snow measurements in the proposed multi-objective calibration made it possible to reproduce streamflow properly, not only by better fitting the observations in terms of statistical metrics for SWE and snow cover, but also by simulating discharge with a more realistic contribution of snowmelt and its influence on the hydrological cycle. Fig. 11 compares the estimates of three main hydrological components which determine runoff (rainfall, snowmelt and ice melt) between the single variable procedure (*Q100*) and the multi-variable procedure (*SFQ75*) in each catchment. The absolute values of liquid precipitation (rainfall, Fig. 11a) remained unchanged in the two procedures since no orographic gradient was applied to total precipitation before the separation phase or to rainfall after the separation phase (see Section 3.1.1). However, due to the differences in the optimised parameter values between the two objective functions, the contribution of snowmelt to streamflow differed in the model calibrated on runoff only (*Q100*) or on snow data in addition to runoff (*SFQ75*). For certain catchments, this contribution is greater with the multi-variable procedure while for others, on the contrary, it is lower. This is explained in particular by the differences in optimised values for the snowfall adjustment (*SFA*) parameter (see Fig. 7) after the separation phase. Depending on the catchment, the *SFA* values ranged from 6 % to 103 % (median 41 %) with *Q100,* and from 11 % to 125 % (median 44 %) with *SFQ75*. Snowfall (and hence total precipitation) therefore increased systematically in each catchment but the proportions varied depending on the procedure. Differences in the quantity of accumulated snow thus led to differences in the quantity of snowmelt contributing to streamflow (Fig. 11b). It can be assumed that the multi-variable

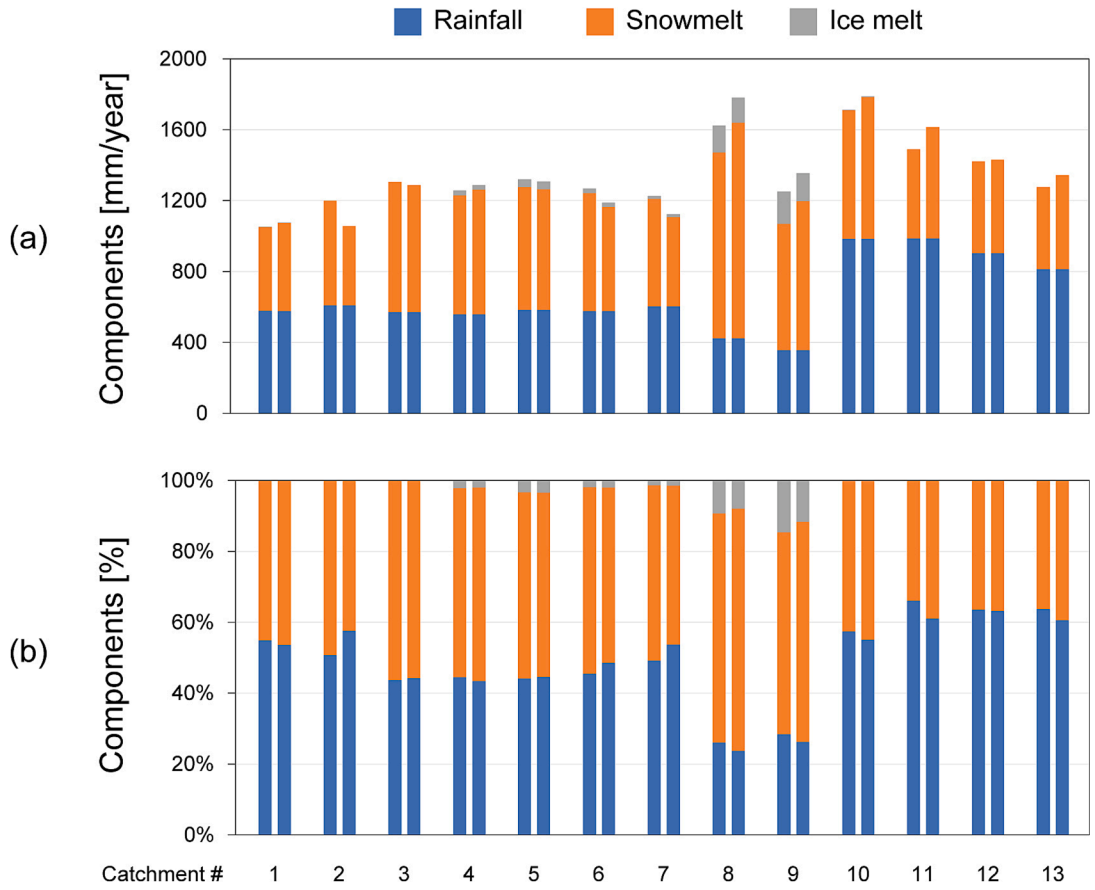


Fig. 11. The main components (rainfall, snowmelt and ice melt) that contribute to the total runoff simulated by the model for the 13 catchments studied (Table 1), when using *Q100* (left bars) and *SFQ75* (right bars) as the objective function for calibration: annual average of (a) absolute and (b) relative contributions over the period 2004–2016.

procedure provides the most realistic distribution, as supported by the analysis of the model performances in terms of SWE, snow cover and runoff. Although similar performances can be obtained for runoff with the single variable objective function (*Q100*), the intermediate states are better represented with the multi-variable objective function (*SFQ75*), giving more credibility to the simulations.

Although it was possible to identify sets of parameters leading to more accurate snow simulations with no negative effect on runoff simulations, the gain in internal realism was not accompanied by a significant reduction in optimisation time and equifinality. Hence, the model calibrated jointly on snow and runoff data did not converge significantly faster than when calibrated on runoff alone. Moreover, an analysis of the parameter frequency during optimisation showed that the use of snow data in the objective function had a very limited impact on narrowing parameter ranges for most parameters. Including snow criteria in the objective function generally led to shifts in the parameter distribution with only marginal effects on the additional constraint of the distribution due to compensation between parameters. This finding suggests that the main benefit of incorporating snow data in the calibration process is improving model consistency by identifying more consistent parameters in terms of the control data rather than reducing parameter uncertainty, which was also reported in Duethmann et al. (2014) and Kelleher et al. (2017).

A final analysis also failed to demonstrate that parameter sets calibrated jointly against snow and runoff are more robust for runoff simulations over past independent periods than parameter sets calibrated against runoff alone. This result should nevertheless be interpreted with caution insofar as the four 10-year periods used as an independent control represent wetter and colder climatic conditions than the recent period used for calibration. With this test, the transferability of the model to drier and warmer conditions was therefore not evaluated, which would be necessary given the expected climate projections in many mountain regions. On the other hand, the model was also evaluated with a split-sample test over the recent period with differences in mean annual values of 9 % for precipitation and 0.4 °C for temperature between the two independent periods. This test also failed to demonstrate that the model calibrated jointly with snow and runoff data was more robust since similar runoff simulations were obtained despite better internal consistency compared with calibration on runoff alone. A more demanding assessment based on a differential split-sample test across climatically contrasting subperiods of discontinuous years (not shown here for the sake of brevity) led to the same conclusions.

5.2. Sources of uncertainties

Some remaining uncertainties that could affect the results are worth discussing. The following sections thus evaluate and discuss the sensitivity of the model performance to the input data, the basic metric in the objective functions, the number of free parameters and the quality of the control data.

5.2.1. Input data

As reported by some authors (e.g. Pellicciotti et al., 2012; Ruelland, 2020), major uncertainty, which is likely to compromise any modelling attempt in mountain catchments, can be found in the input data used to drive the models. For example, Ruelland (2020) showed that the models can be very sensitive to the precipitation and temperature lapse rates used for extrapolation of precipitation and air temperature from point measurements to catchment scale, thus underlining the importance of correctly extrapolating the input meteorological variables. We can therefore question the quality of the precipitation and temperature inputs provided by SAFRAN reanalysis whose spatial resolution could be too coarse for many of the selected catchments. Fig. 12 compares the performance distributions for the 13 catchments using the SAFRAN (8x8 km, Vidal et al., 2010), SPAZM (1x1 km, Gottardi et al., 2012), and STAT (0.5x0.5 km, Ruelland, 2023) databases as the precipitation and temperature input dataset. These databases are based on a similar network of precipitation and temperature stations located in the study areas but are distinguished by how the authors accounted for the precipitation and temperature elevation gradients. While the mesh size (64 km^2) of the SAFRAN reanalysis does not make it possible to accurately reproduce the effects of topography on the variables (even though they are taken into consideration), SPAZM reanalysis allows daily precipitation and temperature fields to be interpolated at a resolution of 1 km^2, considering these effects based on weather patterns (i.e. by type of general atmospheric circulation). A precipitation correction is also applied to compensate for snow under-catch of rain gauges. In the STAT database, the daily climate series and the gauge elevations at each time step are simply interpolated using the inverse distance weighting technique at a 500-m spatial resolution. In so doing, the dependence of precipitation and temperature inputs on elevation as well as the correction for snowfall under-catch are deliberately only considered via model parameters.

As shown in Fig. 12, the best modelling performances in both SWE, snow cover and runoff were obtained with the STAT database. Using the SAFRAN database made it possible to obtain similar performances for SWE and runoff while the snow cover simulations were very slightly less satisfactory. If the use of the SPAZM database also made it possible to obtain reasonable simulations for snow cover and runoff (although slightly less efficient), SWE simulations were clearly less satisfactory than when the SAFRAN and STAT databases were used. These results show that spatial resolution has a limited impact on model performance. The reason the STAT database enabled a better performance is because no elevation gradients were pre-designed off-line of the model, thus

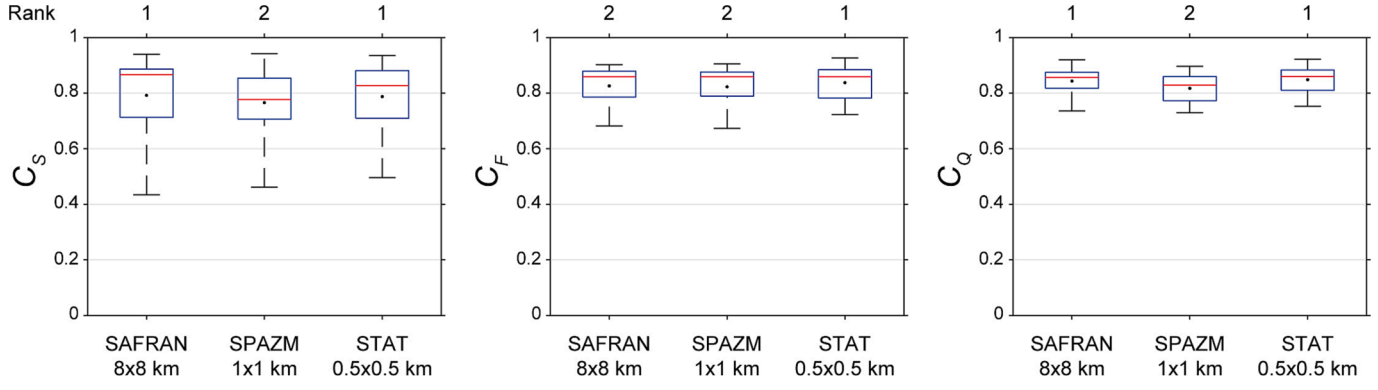


Fig. 12. Comparison of performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments according to the *SFQ75* objective function using the SAFRAN (8x8 km, Vidal et al., 2010), SPAZM (1x1 km, Gottardi et al., 2012), and the STAT (0.5 × 0.5 km, Ruelland, 2023) databases as the precipitation and temperature input dataset for the SIAR/GR4J model.

leaving full management of the orographic dependency to model parameters by means of an inverse approach using the control data that bear the signature of this dependency as well as snowfall under-catch. Consequently, before its extrapolation to the elevation bands within each catchment (see Section 3.1.1), total precipitation is lower than in the SPAZM database. On average over the catchments, the SPAZM database provides 37 % more precipitation than the STAT database with a mean elevation signal of 2100 m a.s.l., while in the STAT database, this signal is about 1400 m a.s.l. linked to the average gauge line. Similarly, temperature is on average 1.4 °C higher in the STAT database with a mean elevation signal of around 1500 m a.s.l. Regardless of the database used, the temperature and precipitation forcings were extrapolated to the elevation bands by applying a fixed constant lapse rate of −0.62 °C (100 m^{-1}) for temperature and no precipitation elevation gradient before the separation phase (see Table 2). On the other hand, calibration of the *SFA* parameter makes it possible to correct snow under-catch while also accounting for any orographic increase in solid precipitation after the separation phase. As evidenced by Fig. 12, this approach led to better modelling performance when applied to forcings without prior consideration of elevation gradients. This suggests that precipitation and temperature estimates obtained with preliminary extrapolation of orographic gradients from point observations are not necessarily appropriate for high-elevation catchments because low-to-mid-elevation meteorological stations are not sufficiently representative of the upper areas, as already demonstrated by Ruelland (2020). To conclude, the coarse spatial resolution of the SAFRAN database limits the impact of the preliminary application of elevation gradients, making it almost as suitable for the modelling experiment as the STAT database where no preconceived lapse rates were taken into account.

Another well-known source of uncertainty is evaporation that can be estimated with more or less complex formulas, among which two stand out. The Oudin formula (Oudin et al., 2005) estimates potential evaporation based solely on mean air temperature and solar radiation. This formulation was used in the present study because it is recognised as the most suitable for rainfall-runoff models. The Penman-Monteith formula (Monteith, 1965) is more sophisticated and includes the use of many variables (mean air temperature, wind speed, specific humidity, atmospheric radiation, visible radiation, atmospheric pressure) to compute potential evaporation. As such, the Penman-Monteith formula is often considered by the scientific community as the most physically realistic.

In addition to the precipitation and temperature data used in the present study, the SAFRAN atmospheric reanalysis also provides an estimate of all the variables required to calculate Penman-Monteith evaporation which is used to drive physically-based models in France. The modelling performances using the Oudin and Penman-Monteith formula were compared (see Fig. 13). Although Penman-Monteith's formulation is more sophisticated, Oudin's formulation clearly led to better modelling performances both in SWE and runoff. The reasons for this better performance are beyond the scope of the present study and would require a more detailed analysis. Nevertheless, it can be noted that if the seasonal dynamics of evaporation were similar in the two formulations, the Penman-Monteith formula led to greater daily variability as well as higher annual evaporation. It is also interesting to note that the same results were obtained using the SPAZM database, although this database provided larger precipitation volumes as input to the model. These findings show that more complex formulas do not necessarily guarantee better performance in hydrological modelling and suggest that simple formulas should be favoured if they significantly improve the simulations, as already demonstrated by Oudin et al. (2005).

5.2.2. *Basic metric used in the objective functions*

Performance criteria aim to assess the goodness-of-fit of a model to observed data. The criteria are generally expressed as a score, for which the best value corresponds to a perfect match between simulations and observations. In hydrology, the Nash-Sutcliffe efficiency (NSE) is still one of the most commonly used criteria, although the past decade has seen a gain in popularity of alternatives such as the Kling–Gupta efficiency (KGE) which has been proposed to address some limitations of the NSE criterion (see Gupta et al., 2009 for more details). However, the KGE criterion also has its own issues. For instance, Cinkus et al. (2023) showed that the score of the KGE can be increased by concurrent overestimation and underestimation of discharge. These counterbalancing errors may favour bias and variability parameters, therefore preserving an overall high score of the performance criteria. As bias and variability parameters generally account for two-thirds of the weight in the equation of KGE, this can lead to an overall higher criterion score without an associated increase in model relevance.

Although the detailed evaluation of the NSE and KGE criteria is beyond the scope of this paper, it may be useful to compare the modelling efficiency obtained using the two metrics. Fig. 14 thus presents the performance distributions obtained over the evaluation periods according to four objective functions (*Q100*, *FQ75*, *SQ75* and *SFQ75*) using the KGE metric as the basis for the computation of the objective functions and evaluation criteria used in the present paper (see Section 3.2.2). Fig. 14 can be compared to Fig. 3 which shows the performance distributions using NSE as basic metric. As reported by Knoben et al. (2019), it is worth noting that NSE and KGE scores should not be compared directly as KGE has no inherent benchmark. Also note that Table 5 shows the mean number of SCE trials for optimisation to reach convergence during calibration using either the NSE or KGE metric to calculate the objective functions. Compared to the KGE criterion, the NSE criterion led to faster optimisation (smaller number of trials for optimisation) and similar modelling results over independent control periods (evaluation criteria ranked the same). These tests suggest that the results of the present study are not affected by the choice of the basic criterion in the objective functions used. The tests also suggest that the KGE criterion is more subject to equifinality than the NSE criterion in the

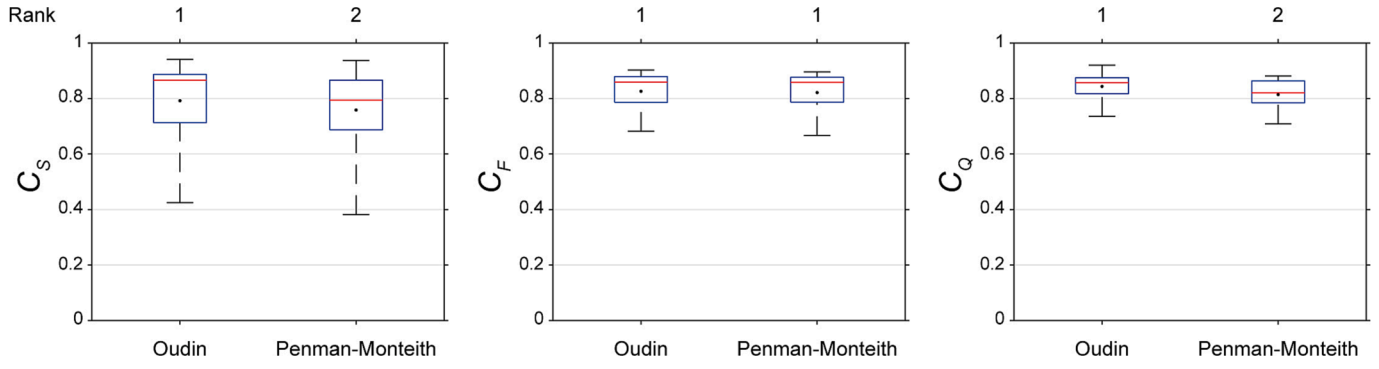


Fig. 13. Comparison of performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments according to the *SFQ75* objective function using the Oudin (Oudin et al., 2005) and Penman-Monteith (Monteith, 1965) formulations to compute potential evaporation as the input dataset.

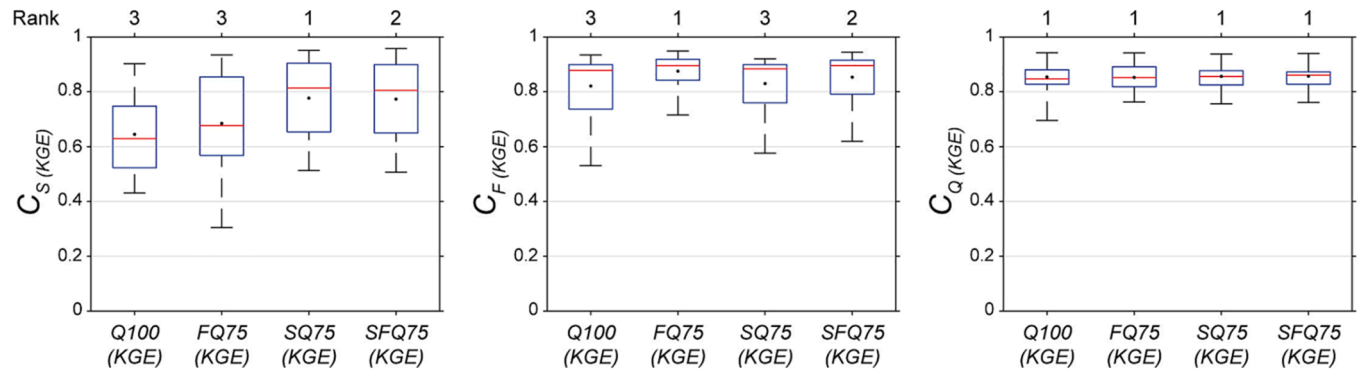


Fig. 14. Performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments according to four objective functions (*Q100*, *FQ75*, *SQ75* and *SFQ75*) using the KGE metric as base in the objective functions and evaluation criteria.

search for an optimal parameter set and that it does not necessarily provide better robustness over independent control periods. This is why we chose the NSE criterion as the basis for the calculation of the objective functions and the evaluation criteria.

5.2.3. *Number of free parameters*

Another question is related to the parametrisation that may be justified in the SIAR model by the snow and runoff control data used. This question would require an extensive discussion which is also beyond the scope of this paper, and the reader can refer for instance to Ruelland (2023) for more details on the degrees of freedom and complexity warranted in temperature-index models in mountainous areas at operational scales. However, as a continuation of this study, the impact of an increase or reduction in the number of free parameters on modelling performance can be briefly analysed here. Fig. 15 compares the performance distributions using the SIAR model with 4 free parameters (SIAR4, like in this paper) with alternative versions with 11 free parameters (SIAR11) and 1 free parameter (SIAR1). Interestingly, performance distributions were not significantly improved by increasing the number of free parameters (SIAR11) nor altered by decreasing their number (SIAR1). This shows that increasing the number of free parameters (hence exacerbating equifinality problems) does not improve model performance, whereas reducing the degree of freedom in the model can be very effective. This finding is not entirely surprising given the existence of interdependencies between certain snow parameters (coefficient for the thermal state of the snowpack as well as constant and variable parts of the degree-day snow melt factor) when the model was calibrated jointly with snow and runoff data (see Fig. 9). This calls for reducing the number of parameters requiring calibration since these parameters can be set to general values without impairing model performance. However, a catchment-specific adjustment of snowfall with the *SFA* parameter is required in order to avoid impairing the local SWE simulations, as already demonstrated by Ruelland (2023).

Applying an adjustment factor only to solid precipitation based on a constant catchment-specific snowfall adjustment can also be discussed. Previous studies in the Alps and Pyrenees (Ruelland, 2020, 2023) showed that annual precipitation recorded by low- to mid-elevation gauges can typically be up to twice as high in elevation catchment areas. For example, Ruelland (2020) estimated that the mean annual precipitation lapse rate was in the order of 30 % $(km)^{-1}$ in Alpine catchments. On the other hand, Ruelland (2023) demonstrated that the optimised values of the precipitation and rainfall lapse rates (*PLR* and *RLR*, see Table 2) clearly tended towards 0 % (km^{-1}) as long as a snowfall adjustment (*SFA*) was applied in various catchments in the Alps and Pyrenees. This suggests that the orographic increase in precipitation mainly originates from solid precipitation in these mountainous areas, in agreement with Dessens and Bücher (1997) who reported that the precipitation lapse rate in the Pyrenees was twice as large in winter as in summer. This can be explained by the fact that the time of year and weather conditions interact. For instance, summer corresponds more to stormy periods with generally poorly defined precipitation-elevation relationships, while winter is more conducive to prolonged episodes of solid precipitation with more easily identifiable orographic effects. The increase in solid precipitation is also consistent with the need to correct systematic errors associated with snowfall under-catch (Kochendorfer et al., 2022), particularly in windy conditions, which typically results in a mass deficit. By tackling both snowfall under-catch and the orographic increase in precipitation, *SFA* is among the parameters that have the most impact on model outputs, and is indispensable (and sufficient) for better internal consistency and for more accurate estimation of the contribution of snow to catchment runoff (see Section 5.1). It controls both the extrapolation of solid precipitation to the spatial scale of the

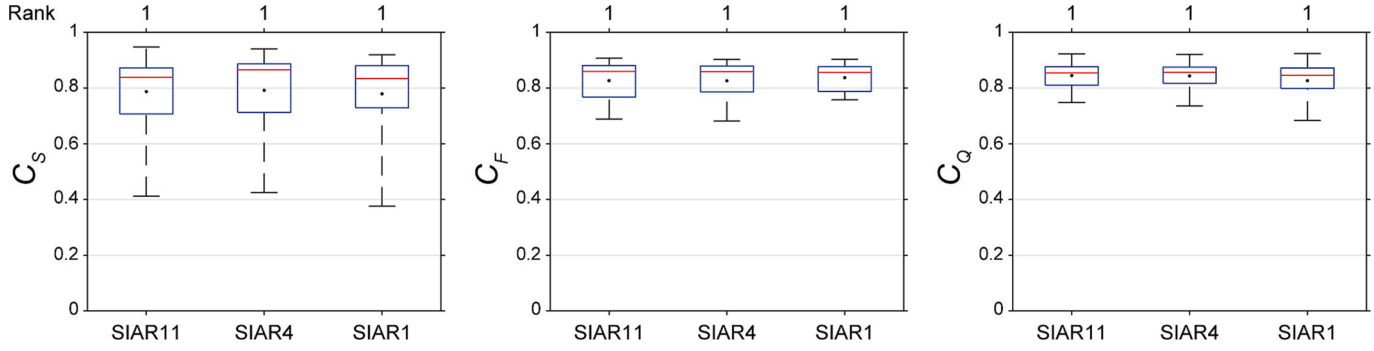


Fig. 15. Comparison of performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments according to the *SFQ75* objective function using the GR4J model at the bottom of the SIAR model with 11 free parameters (*PLR*, *TLR*, *CV*, *SFA*, *RLR*, *θ*, *Kf*, *Ks*, *Kg*, T_{acc} and *Rsp*), with 4 free parameters (*SFA*, *θ*, *Kf* and *Ks*) and with 1 free parameter (*SFA*). With SIAR1, the parameters (*θ*, *Kf* and *Ks)* were fixed to the median values obtained using SIAR4 with the *SFQ75* objective function (see Fig. 7).

catchment, the spatial extent of melt processes, and the amount and timing of streamflow. As such, *SFA* is a key parameter for maintaining an accurate water balance in the model while influencing the amount of snow that must melt before bare ground appears. Although this parameter can be considered to have physical significance at catchment scale, its value cannot be identified based on field experiments because it simultaneously adjusts for snow under-catch and need for orographic increase in solid precipitation in the catchments.

5.2.4. Quality of control data

The quality of the control data that were used to calibrate and control the model can also be questioned. It is difficult to assess the quality of SWE and streamflow records, which, although like any measurement, contain errors, can be considered to have the highest possible precision (see Section 2.1). On the other hand, it is easier to investigate to what extent the gap-filling technique applied to the MODIS data (see Section 2.1.2) affects the results. It is indeed worth noting that in the present study, we chose to fill in the missing data in the chronicles of snow-covered areas provided by the MODIS sensor on board the Terra and Aqua satellites so as to provide cloud-free binary maps. This choice was motivated by having complete spatial and temporal information available on each catchment and because the gap-filling technique was previously shown to have high reconstruction accuracy (Ruelland, 2020).

In order to analyse the impact of the gap-filling procedure, the modelling performances were also tested using raw data from the original Terra and Aqua maps without applying temporal and spatial filtering to remove cloud obscuration. After combining the Aqua and Terra data, 44 % (41 %) of no-data (mainly due to cloud cover) were still present on average over the French Alps (Pyrenees) in the 2000–2016 period (see Ruelland (2020) for more details). Cloud cover is particularly marked during the winter months when the snow cover is most extensive. Although more incomplete, the raw SCA data are theoretically more accurate since they have not been subject to gap-filling using temporal and spatial filters. On the other hand, this considerably limits the possibilities of aggregating MODIS raw data at the scale of the different elevation bands in each catchment since the weighted spatial averages are affected by the missing values. To deal with the inevitable trade-off between reliability and availability, observed snow cover values were only used for model comparison if more than 60 % of the elevation band had no missing data on a given day. When this threshold was not reached, the data concerned were not accounted for in the calibration and evaluation procedures (see Section 3.2). Note that Parajka and Blöschl (2008) reported that a 60 % cloud threshold was a reasonable compromise between SCA data availability and reliability, and Riboust et al. (2019) applied the same threshold.

Fig. 16 shows the performance distributions obtained in evaluation with the model calibrated using four objective functions (*Q100*, *FQ75*, *SQ75* and *SFQ75*) and MODIS data without the application gap-filling filters. As shown by comparison with Fig. 3 using cloud-free snow cover after applying the gap-filling technique, the performance distributions are very similar, as is Friedman ranking between the objective functions. The same goes for the other analyses presented in this paper (not shown here for the sake of brevity). This shows that the cloud removal procedure applied to the remotely-sensed snow covers has a limited impact on the accuracy and associated uncertainty of snow cover estimation and that it may be advisable to use a completely cloud-free dataset for model calibration to avoid the problem of missing data in the MODIS raw snow products.

6. Conclusion

This paper provides guidelines on the use of snow data in addition to runoff to calibrate hydrological models in mountain catchments, based on a sample of 13 snow dominated catchments in the French Alps and Pyrenees. The method evaluates the adequacy of different observational datasets (local SWE, MODIS snow cover and catchment streamflow) as well as various combinations of goodness-of-fit metrics to constrain the coupled SIAR/GR4J model in an attempt to improve its consistency and robustness. Sequential calibration of the snow parameters on SWE and/or snow cover data resulted in the best SWE and snow cover simulations but impaired streamflow simulations. While not drastic, this impairment is significant, and is likely caused by over-adjustment of the snow parameters on snow observations. In contrast, simultaneous calibration of the snow and rainfall-runoff parameters based on multi-criteria composite functions was shown to improve snow simulations without reducing runoff efficiency. This technique is therefore recommended for calibrating the parameters of snow-hydrological models provided that the weight of snow observations in the objective function is limited to 25 %. Local SWE measurements proved particularly useful for identifying parameters leading to better internal consistency of the model. As such, the results show that even sparse SWE observations from cosmic-ray snow sensors are a useful source of internal catchment data for model calibration and control in mountain environments. It is therefore essential to maintain and strengthen existing SWE gauge networks by extending the duration of records and increasing the number of measurement points. MODIS-based snow cover data were found to be less useful than local SWE observations because they only provide information on the areal extent of snow cover, but not its depth/volume, thus presenting less variance to constrain models. However, it is recommended to use the MODIS snow products nevertheless because they also help better represent snow processes. It could not be demonstrated that the improvement in the internal consistency of the model through the use of snow observations in the calibration process is accompanied by a significant reduction in optimisation time and equifinality. On the other hand, the improvement in the internal consistency makes it possible to reduce the interdependence between the parameters of the snow model and the rainfall-runoff model. It also allows to identify the least sensitive snow parameters in order to set them at general values without

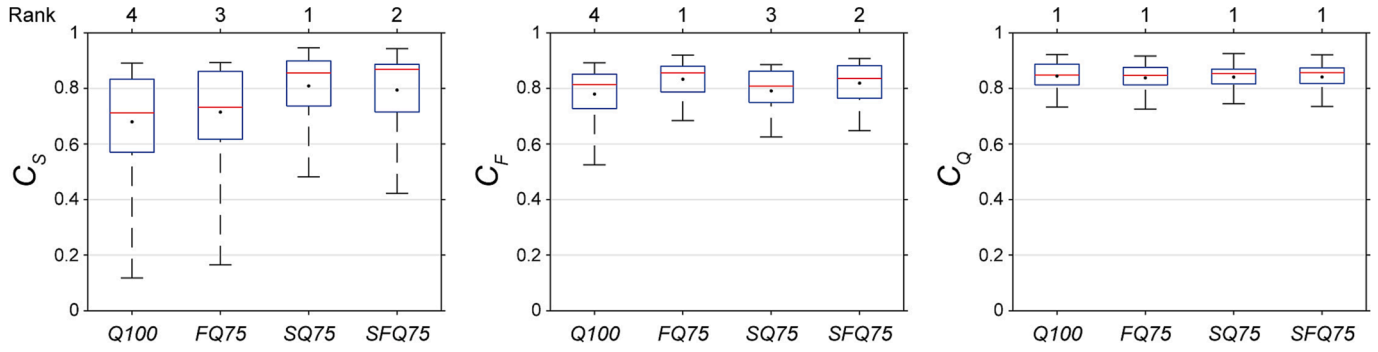


Fig. 16. Performance distributions (in C_S, C_F and C_Q) obtained over the evaluation periods for the 13 catchments with the model calibrated using four objective functions (*Q100*, *FQ75*, *SQ75* and *SFQ75*) and MODIS data without the application of gap-filling filters.

impairing model performance, thereby reducing equifinality with a more parsimonious model. There is also no evidence that the use of snow data improves model robustness to climate variability, although it clearly enhances internal consistency. These results suggest that the main benefit of including snow data in the calibration is to improve model consistency rather than model robustness. Improving model robustness thus remains a complex challenge that cannot be met only by better parametrisation using additional data because it also requires the design of new model structures better adapted to hydrological changes. An interesting avenue of research to this end could consist of better exploiting the dependence of parameter values on climate in order to improve the transferability of the model to contrasted climate conditions.

CRediT authorship contribution statement

Denis Ruelland: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

The author is very grateful to *Eléctricité de France* (*EDF*), *Météo-France* and *Banque Hydro* for providing respectively, the necessary in-situ SWE measurements, meteorological data and runoff for the study. NASA's National Snow and Ice Data Center (NSIDC) Distributed Active Archive Center (DAAC) is also acknowledged for providing MODIS snow products. The author is very grateful to James McPhee and one anonymous reviewer who provided a thorough review of this work, which helped to improve the final manuscript.

References

Ajami, N.K., Gupta, H., Wagener, T., Sorooshian, S., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. J. Hydrol. 29, 112–135. https://doi.org/10.1016/j.jhydrol.2004.03.033.

Bergström, S., Lindström, G., Pettersson, A., 2002. Multi-variable parameter estimation to increase confidence in hydrological modelling. Hydrol. Proc. 16, 413–421. https://doi.org/10.1002/hyp.332.

Besic, N., Vasile, G., Gottardi, F., Gailhard, J., Girard, A., D'Urso, G., 2014. Calibration of a distributed SWE model using MODIS snow cover maps and in situ measurements. Remote Sensing Letters. https://doi.org/10.1080/2150704X.2014.897399.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Bormann, K.J., Evans, J.P., McCabe, M.F., 2014. Constraining snowmelt in a temperature-index model using simulated snow densities. J. Hydrol. 517, 652–667. https://doi.org/10.1016/j.jhydrol.2014.05.073.

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., Goldscheider, N., 2023. When best is the enemy of good – critical evaluation of performance criteria in hydrological models. Hydrol. Earth Syst. Sci. 27, 2397–2411. https://doi.org/10.5194/hess-27-2397-2023.

Coron, L., Andréassian, V., Perrin, C., Bourqui, M., Hendrick, F., 2014. On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrol. Earth Syst. Sci. 18, 727–746. https://doi.org/10.5194/hess-18-727-2014.

Da Ronco, P., Avanzi, F., De Michele, C., Notarnicola, C., Schaefli, B., 2020. Comparing MODIS snow products Collection 5 with Collection 6 over Italian Central Apennines. Int. J. Rem. Sens. 41, 4174–4205. https://doi.org/10.1080/01431161.2020.1714778.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE t. Evolut. Comput. 6, 181–197. https://doi.org/10.1109/4235.996017.

Dessens, J., Bücher, A., 1997. A critical examination of the precipitation records at the Pic du Midi observatory, Pyrenees (France). Climatic Change 36, 345–353. https://doi.org/10.1023/A:1005389721863.

Dong, C., 2018. Remote sensing, hydrological modeling and in situ observations in snow cover research: A review. J. Hydrol. 561, 573–583. https://doi.org/10.1016/j.jhydrol.2018.04.027.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. https://doi.org/10.1029/91WR02985.

Duan, Q., Sorooshian, S., Gupta, V., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4.

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., Güntner, A., 2014. The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. Water Resour. Res. 50, 2002–2021. https://doi.org/10.1002/2013WR014382.

Duethmann, D., Blöschl, G., Parajka, J., 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? Hydrol. Earth Syst. Sci. 24, 3493–3511. https://doi.org/10.5194/hess-24-3493-2020.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., Burlando, P., 2011. The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model. Water Resour. Res. 47, W07519. https://doi.org/10.1029/2010WR009824.

Finger, D., Vis, M., Huss, M., Seibert, J., 2015. The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. Water Resour. Res. 51, 1939–1958. https://doi.org/10.1002/2014WR015712.

Franz, K.J., Karsten, L.R., 2013. Calibration of a distributed snow model using MODIS snow covered area data. J. Hydrol. 494, 160–175. https://doi.org/10.1016/j.jhydrol.2013.04.026.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32, 675–701. https://doi.org/10.1080/01621459.1937.10503522.

Garavaglia, F., Le Lay, M., Gottardi, F., Garçon, R., Gailhard, J., Paquet, E., Mathevet, T., 2017. Impact of model structure on flow simulation and hydrological realism from lumped to semi-distributed approach. Hydrol. Earth Syst. Sci. 21, 3937–3952. https://doi.org/10.5194/hess-21-3937-2017.

Gottardi, F., Obled, C., Gailhard, J., Paquet, E., 2012. Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. J. Hydrol. 432–433, 154–167. https://doi.org/10.1016/j.jhydrol.2012.02.014.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hall, D. K, and. Riggs, G.A.: MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MYD10A1.006, 2016b.

Hall, D. K., and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10A1.006, 2016a.

Hall, D., Riggs, G., and Salomonson, V.: MODIS/Aqua Snow Cover Daily L3 Global 500m SIN Grid, Version 5, National Snow and Ice Data Center, Boulder, Colorado, USA, , 2007b.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products. Remote Sens. Environ. 83, 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0.

Hall, D.K., Riggs, G.A., 2007. Accuracy assessment of the MODIS snow products. Hydrol. Proc. 21, 1534–1547. https://doi.org/10.1002/hyp.6715.

Hall, D., Riggs, G., Salomonson, V., 2007a. MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 5, National Snow and Ice Data Center, Boulder, Colorado. USA. https://doi.org/10.5067/63NQASRDPDB0.

He, Z.H., Parajka, J., Tian, F.Q., Blöschl, G., 2014. Estimating degree-day factors from MODIS for snowmelt runoff modeling. Hydrol. Earth Syst. Sci. 18, 4773–4789. https://doi.org/10.5194/hess-18-4773-2014.

Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282, 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9.

Hublart, P., Ruelland, D., Dezetter, A., Jourde, H., 2015. Reducing structural uncertainty in conceptual hydrological modeling in the semi-arid Andes. Hydrol. Earth Syst. Sci. 19, 2295–2314. https://doi.org/10.5194/hess-19-2295-2015.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-Filled SRTM for the Globe Version 4, Available from the CGIAR-CSI SRTM 90 m Database, Available at: Http://srtm.csi.cgiar.org (last Access: 23 15, 25–54.

Kelleher, C., McGlynn, B., and Wagener, T.: Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding, Hydrol. Earth Syst. Sci., 21, 3325–3352, https://doi.org/10.5194/hess-21-3325-2017, 2017.

Khakbaz, B., Imam, B., Hsu, K., Sorooshian, S., 2012. From lumped to distributed via semi-distributed: calibration strategies for semi-distributed hydrologic models. J. Hydrol. 418–419, 61–77. https://doi.org/10.1016/j.jhydrol.2009.02.021.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42 https://doi.org/10.1029/2005WR004362.
Klok, E.J., Jasper, K., Roelofsma, K.P., Gurtz, J., Badoux, A., 2001. Distributed hydrological modelling of a heavily glaciated Alpine river basin. Hydrol. Sci. J. 46, 553–570. https://doi.org/10.1080/02626660109492850.
Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: inherent benchmark or not? Comparing Nash and Kling efficiency scores. Hydrol. Earth Syst. Sci. 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.
Kochendorfer, J., Earle, M., Rasmussen, R., Smith, C., Yang, D., Morin, S., Mekis, E., Buisan, S., Roulet, Y.-A., Landolt, S., Wolff, M., Hoover, J., Thériault, J., Lee, G., Baker, B., Nitu, R., Lanza, L., Colli, M., Meyers, T., 2022. How well are we measuring snow post-SPICE? Bull. Amer. Meteor. Soc. 103 (370–388), 2022. https://doi.org/10.1175/BAMS-D-20-0228.1.
Kodama, M., Nakai, K., Kawasaki, S., Wada, M., 1979. An application of cosmic-ray neutron measurements to the determination of the snow-water equivalent. J. Hydrol. 41, 85–92. https://doi.org/10.1016/0022-1694(79)90107-0.
Kumar, L., Skidmore, A.K., Knowles, E., 1997. Modelling topographic variation in solar radiation in a GIS environment. Int. J. Geogr. Info. Sys. 11, 475–497. https://doi.org/10.1080/136588197242266.
Leleu, I., Tonnelier, I., Puechberty, R., Gouin, P., Viquendi, I., Cobos, L., Foray, A., Baillon, M., Ndima, P.O., 2014. La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques. La Houille Blanche 1, 25–32. https://doi.org/10.1051/lhb/2014004.
Luce, C.H., Tarboton, D.G., 2004. The application of depletion curves for parameterization of subgrid variability of snow. Hydrol. Processes 18, 1409–1422. https://doi.org/10.1002/hyp.1420.
Luo, Y., Arnold, J., Liu, S.Y., Wang, X.Y., Chen, X., 2013. Inclusion of glacier processes for distributed hydrological modeling at basin scale with application to a watershed in Tianshan Mountains, northwest China. J. Hydrol. 477, 72–85. https://doi.org/10.1016/j.jhydrol.2012.11.005.
Magnusson, J., Gustafsson, D., Hüsler, F., Jonas, T., 2014. Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. Water Resour. Res. 50, 7816–7835. https://doi.org/10.1002/2014WR015302.
Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., Jonas, T., 2015. Evaluating snow models with varying process representations for hydrological applications. Water Resour. Res. 51 https://doi.org/10.1002/2014WR016498.
Monteith, J.L., 1965. Evaporation and Environment. Symposia of the Society for Experimental Biology 19, 205–234.
Nemri, S., Kinnard, C., 2020. Comparing calibration strategies of a conceptual snow hydrology model and their impact on model performance and parameter identifiability. J. Hydrol. 582 https://doi.org/10.1016/j.jhydrol.2019.124474.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2: towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. J. Hydrol. 303, 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.025.
Paquet, E., Laval, M.-T., 2006. Operation feedback and prospects of EDF Cosmic-Ray Snow Sensors. La Houille Blanche 2, 113–119. https://doi.org/10.1051/lhb:200602015.
Parajka, J., and Blöschl, G.: MODIS-based snow cover products, validation, and hydrologic applications: Chapter 9 in multiscale hydrological remote sensing: perspectives and applications, edited by: Chang, N. B. and Hong, Y., CRC Press, Boca Raton, 185–212, https://doi.org/10.1201/b11279, 2012.
Parajka, J., Blöschl, G., 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. J. Hydrol. 358, 240–258. https://doi.org/10.1016/j.jhydrol.2008.06.006.
Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. Hydrol. Processes 21, 435–446. https://doi.org/10.1002/hyp.6253.
Pellicciotti, F., Buergi, C., Immerzeel, W.W., Konz, M., Shrestha, A.B., 2012. Challenges and uncertainties in hydrological modeling of remote Hindu Kush–Karakoram–Himalayan (HKH) basins: suggestions for Calibration Strategies. Mountain Research and Development 32, 39–50. https://doi.org/10.1659/MRD-JOURNAL-D-11-00092.1.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7.
Ragettli, S., Pellicciotti, F., 2012. Calibration of a physically based, spatially distributed hydrological model in a glacierized basin: on the use of knowledge from glaciometeorological processes to constrain model parameters. Water Resour. Res. 48 https://doi.org/10.1029/2011WR010559.
Rees, W.G., 2005. Remote Sensing of Snow and Ice, Boca Raton (USA), CRC Press. Taylor and Francis Group. https://doi.org/10.1201/9780367801069.
Riboust, P., Thirel, G., Le Moine, N., Ribstein, P., 2019. Revisiting a simple degree-day model for integrating satellite data: implementation of SWE-SCA hystereses. J. Hydrol. Hydromech. 67, 70–81. https://doi.org/10.2478/johh-2018-0004.
Ruelland, D., 2020. Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models? Hydrol. Earth Syst. Sci. 24, 2609–2632. https://doi.org/10.5194/hess-24-2609-2020.
Ruelland, D., 2023. Development of the snow- and ice-accounting routine (SIAR). J. Hydrol. 624 https://doi.org/10.1016/j.jhydrol.2023.129867.
Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M., Parajka, J., 2020. The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters. J. Hydrol. 583, 124639 https://doi.org/10.1016/j.jhydrol.2020.124639.
Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model inter-comparison project (DMIP): motivation and experiment design. J. Hydrol. 298, 4–26. https://doi.org/10.1016/j.jhydrol.2004.03.040.
Tong, R., Parajka, J., Komma, J., Blöschl, G., 2020. Mapping snow cover from daily Collection 6 MODIS products over Austria. J. Hydrol. 590, 125548 https://doi.org/10.1016/j.jhydrol.2020.125548.
Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. A multiobjective approach to improve SWAT model calibration in alpine catchments. J. Hydrol. 559, 347–360. https://doi.org/10.1016/j.jhydrol.2018.02.055.
Valéry, A., Andréassian, V., Perrin, C., 2014. As simple as possible but not simpler: What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. J. Hydrol. 517, 1176–1187. https://doi.org/10.1016/j.jhydrol.2014.04.058.
Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the SAFRAN system. Int. J. Climatology 30, 1627–1644. https://doi.org/10.1002/joc.2003.
Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R., 2012. COSMOS: The COsmic-ray Soil Moisture Observing System. Hydrol. Earth Syst. Sci. 16, 4079–4099. https://doi.org/10.5194/hess-16-4079-2012.