https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences Discussions

Sensitivity of snow models to the accuracy of meteorological forcings in mountain environment

Silvia Terzago[1], Valentina Andreoli[2], Gabriele Arduini[3], Gianpaolo Balsamo[3], Lorenzo Campo[4], Claudio Cassardo[2], Edoardo Cremonese[6], Daniele Dolia[4], Simone Gabellani[4], Jost von Hardenberg[1], Umberto Morra di Cella[6], Elisa Palazzi[1], Gaia Piazzi[4,5], Paolo Pogliotti[6], and Antonello Provenzale[7]

[1]Institute of Atmospheric Sciences and Climate, National Research Council, Torino, Italy
[2]Department of Physics and Natrisk center, University of Torino, Italy
[3]European Centre for Medium-Range Weather Forecasts, Reading, UK
[4]CIMA Research Foundation - International Centre on Environmental Monitoring, Savona, Italy
[5]IRSTEA, Hydrology Research Group, UR HYCAR, 92761 Antony, France
[6]Environmental Protection Agency of Aosta Valley, Aosta, Italy
[7]Institute of geosciences and earth resources, National Research Council, Pisa, Italy

Correspondence: Silvia Terzago (s.terzago@isac.cnr.it)

Abstract. Snow models are usually evaluated at sites providing high-quality meteorological data, so that the uncertainty in the meteorological input data can be neglected when assessing the model performances. However, high-quality input data are rarely available in mountain areas and, in practical applications, the meteorological forcing to drive snow models is typically derived from spatial interpolation of the available in-situ data or from reanalyses, whose accuracy can be considerably lower. In order to fully characterize the performances of a snow model, the model sensitivity to errors in the input data should be quantified.

In this study we test the ability of six snow models to reproduce snow water equivalent, snow density and snow depth when they are forced by meteorological input data with gradually lower accuracy. The SNOWPACK, GEOTOP, HTESSEL, UTOPIA, SMASH and S3M snow models are forced, first, with high-quality measurements performed at the experimental site of Torgnon, located at 2160 m a.s.l. in the Italian Alps (control run). Then, the models are forced by data at gradually lower temporal and/or spatial resolutions, obtained i) by sampling the original Torgnon 30-minute time series at 3, 6, and 12 hours, ii) by spatially interpolating neighboring in-situ station measurements and iii) by extracting information from GLDAS, ERA5, ERA-Interim reanalyses at the gridpoint closest to the Torgnon station. Since the selected models are characterized by different degrees of complexity, from highly sophisticated multi-layer snow models to simple, empirical, single-layer snow schemes, we also discuss the results of these experiments in relation to the model complexity.

Results show that when forced by accurate 30-min resolution weather station data the single-layer, intermediate-complexity snow models HTESSEL and UTOPIA provide similar skills as the more sophisticated multi-layer model SNOWPACK, and these three models show better agreement with observations and more robust performances over different seasons compared to the lower complexity models SMASH and S3M. All models forced by 3-hourly data provide similar skills as the control run while with 6- and 12-hourly temporal resolution forcings we generally observe a reduction in model performances, except for the SMASH model which shows low sensitivity to the temporal degradation of the input data. Spatially interpolated data from

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Discussions

neighboring stations and reanalyses result to be adequate forcings, provided that temperature and precipitation variables are not affected by large biases over the considered period. A simple bias-adjustment technique applied to ERA-Interim temperatures, however, allowed all models to achieve similar performances as in the control run. All models irrespectively of their complexity show weaknesses in the representation of the snow density.

Copyright statement. TEXT

1 Introduction

A wide range of snow models with different degrees of complexity have been developed for hydrological applications, avalanche risk forecasting or climate studies. Some of them are also integrated within modelling chains for numerical weather forecasts or climate modelling. The degree of complexity of the snow schemes depends on the specific purpose for which they have been developed (Magnusson et al., 2015). Simple temperature-index snow models are employed in applications requiring a coarse estimate of snow depth or snow water equivalent. Physical, energy-balance, but still rather simple snow models are often used in complex modelling chains, i.e. in numerical weather prediction systems and in Earth System models, to limit the computational costs. Sophisticated multi-layer snow models are typically used to reconstruct the vertical structure of the snowpack with a high level of detail and high accuracy, as needed for avalanche warning applications.

Snow models are generally evaluated at a number of sites providing high-quality forcing and verification data. Extensive literature documents the underlying physics and the performaces of single snow models (e.g. Dutra et al., 2010; Vionnet et al., 2012; Bartelt and Lehning, 2002), and several studies compare a limited number of snow models to each other (Boone and Etchevers, 2001; Kumar et al., 2013). A few large intercomparison studies benchmarked multiple snow models, including the PILPS2d, PILPS2e, Rhone-Agg, SNOWMIP and SNOWMIP2 coordinated intercomparison projects.

PILPS2d (Slater et al., 2001; Schlosser et al., 2000) and PILPS2e (Bowling et al., 2003) aimed at evaluating snow water equivalent (SWE) simulations provided by different land surface schemes (LSS) in a Russian and a Swedish snow-dominated catchment, respectively. PILPS2d evaluated twentyone land surface schemes forced by 18 years of observed meteorological data from a grassland catchment in Russia, with the aim to investigate the reasons for models scatter in the output snowpack variables. Weaknesses in reproducing mid-season ablation were shown to produce systematic scatter among the models. Albedo and fractional snow cover were both key variables for an accurate representation of the amount of energy absorbed by the snowpack. Indeed, the ablation during the early snow season is another major source of divergence among models: in early winter a thin snow cover is highly sensitive to changes in the forcings and the resulting differences in snowpack conditions tend to persist throughout the whole snow season owing to internal feedback processes.

PILPS2e showed the difficulty of reproducing spring melting. Errors in winter snow sublimation mainly impacted the runoff simulations, while the retention of meltwater within the snowpack affected the timing of the peak in runoff rather than its

Hydrology and Earth System Sciences
Open Access
Discussions

magnitude. For both PILPS2d and PILPS2e the differences in model complexity did not fully explain differences in model results.

The Rhône-AGG experiment (Boone et al., 2004) employed 15 LSSs to address the impact of the model structure and of the spatial resolution of the forcing data on the simulations of the water balance. LSSs with an explicit (bulk or multi-layer) snow scheme provided better SWE simulations than LSSs with a composite snow scheme (i.e with a mixed snow-soil-vegetation layer). LSSs with composite snow schemes showed too early snow ablation and early run-off peaks compared to observations, owing to missing representation of key processes such as ripening and owing to inadequate representation of albedo and thermal conductivity in a mixed snow-soil/vegetation layer. SWE was strongly affected by the spatial resolution of the meteorological forcing. In fact, when high-resolution meteorological forcings were aggregated from 8 km to a coarser grid of 1° (about 69 km) the simulated SWE was reduced by 25-60% in 13 out of 15 LSSs. A single model explicitely considering subgrid elevation effects on the forcing was found to minimize the impact of scaling on the simulated snow water equivalent.

SnowMIP (Etchevers et al., 2002, 2004) performed an intercomparison among snow models of different complexity, used for different applications, including hydrology, global circulation models, snow monitoring, snow physics, avalanche forecasting, with the aim of identifying key processes for each application. Model complexity was found to have a strong impact on the simulation of the net longwave radiation, which strongly affects snow melt dynamics. Models relying on the explicit simulation of the internal snow processes better represented snow surface temperature and the longwave radiation budget. On the contrary, model complexity had smaller impact on the net shortwave radiation, whose accuracy was dependent on the simulation of albedo. Complex models taking into account snow microstructure were able to properly represent the albedo variability (as a function of grain size and type), but also simple snow models with an appropriate parameterization for albedo dynamics guaranteed reliable estimates of this variable.

SnowMIP2 (Rutter et al., 2009) built upon SnowMIP and focused on the simulation of snowpack properties in forested areas compared to open air sites, across different climatic conditions. Single models showed low correlations between different years in forested sites, and low correlations also between forested and open sites, suggesting that no best model for all years and all sites could be easily identified. Calibration can reduce root mean square error (RMSE) in forested sites, but similar improvements did not apply to non-forested sites.

The mentioned studies shed light on the critical snow processes that produce the largest differences between LSS simulations. However they could not clearly define an optimal set of parameterizations for a given application, as for example numerical weather predictions and climate simulations, or the minimum level of model complexity needed to achive satisfactory skills in a given application (Slater et al., 2001). A step forward in this direction has been made by employing a single model with several options to represent each of the most snow-relevant processes, and then testing the effect of parameterizations with different degrees of complexity on the skill of the model (Essery et al., 2013; Clark et al., 2011). Best results were obtained with models with prognostic representation of snow albedo and density, with at least a simple representation of water retention and refreezing in the snowpack. Major outcomes on the identification of the key-processes to be represented in global climate models to improve snow simulations at the large scale are expected from the ongoing coordinated modelling initiative ESM-SnowMIP (Krinner et al., 2018).

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Discussions

A common characteristic among past model intercomparison initiatives is the interest in testing the skills of the models in experimental sites where high-quality meteorological forcings are available, in order to perform a controlled evaluation of the models performances. However, such context does not represent the typical conditions occurring in practical applications, where snow models are run over large climate model grid cells, and they are coupled to atmospheric models that likely provide biased driving data (Essery et al., 2013). Moreover, reliable modelling of snowpack dynamics in mountain regions is hindered by the high spatial and temporal variability of the meteorological forcings, entailing that observations and reanalysis data at a given location are scarcely representative of the conditions of the surrounding area. A recent review paper on the European mountain cryosphere (Beniston et al., 2018) states that disentangling the uncertainties related to the model structure from those related to the meteorological input data is one of the major challenges for the snow modelling at the catchment scale, namely the scale relevant for hydrological applications. A sensitivity analysis performed on a single, physiscally-based snow model showed that the uncertainty on snow simulations due to the forcing can be comparable to or even larger than the uncertainty due to the model structure (Raleigh et al., 2015). The analysis also showed that biases in the forcing data have a larger effect than random errors, with the magnitude of the biases resulting more important that their probability distribution. Building on the results of previous studies we now consider an ensemble of snow models and we investigate their sensitivity to the quality of the meteorological forcing, with the aim of providing practical information on their performances when they are forced with inputs at gradually lower temporal and/or spatial resolution.

We devised a set of experiments with six different snow models with different degrees of complexity in the Alpine measurement site of Torgnon, located at 2160 m a.s.l. in Aosta Valley, Italy. This site has been selected because it provides high-quality meteorological measurements together with a characterization of snowpack properties in terms of depth, mass and surface temperature. First, we evaluate each model forced by accurate station measurements at 30-minute temporal resolution (we refer to this as "optimal" forcing). Second, we evaluate each model forced by data at gradually lower temporal resolution and/or lower accuracy, by employing data from spatial interpolation of neighboring station measurements and from three gridded global reanalyses, and extracting the meteorological time series at the gridpoint closest to the Torgnon station. The set of multi-model, multi-forcing simulations finally allows to discuss the relationship between model complexity and model sensitivity to errors/inconsistencies in the meteorological forcings.

This paper is structured as follows: Sect. 2 presents the snow models employed in the study while Sect. 3 describes the station of Torgnon and the datasets employed for the experiments. Section 4 describes in detail the set of 12 devised experiments and, for each experiment, the method employed to derive the forcing. Section 5 focuses on the evaluation of snow model outputs against observations, finally Sections 6 and 7 discuss the results and draw the conclusions.

2 Snow models

The six models considered in this study, together with a compact overview of their characteristics, are listed in Table 1 and summarized in the following.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Discussions

SNOWPACK is a highly sophisticated, multi-purpose snow and land-surface model, with a detailed description of the mass and energy exchange between the snow, the atmosphere and optionally with the vegetation cover and the soil. It provides a detailed description of snow properties including weak layer characterization (Stoessel et al., 2009), phase changes and water transport in snow (Hirashima et al., 2010). A particular feature is the treatment of soil and snow as a continuum with a choice of a few up to several hundred layers (Bartelt and Lehning, 2002).

GEOTOP 2.0 is a sophisticated, snow and hydrological process-based model. Its strength is an integrated approach that takes into account the interactions between hydrological, cryospheric and geomorphological processes (Endrizzi et al., 2014). The snowpack evolution is dynamically managed by the model through a snow layering scheme which splits and merges the layers depending on their mass. Moreover, the model takes into account snow metamorphism and water percolation into the snowpack.

HTESSEL is the land-surface model of the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS), controlling the evolution of the snow and soil fields and the exchanges of heat and moisture between the land surface and the atmosphere above (Balsamo et al., 2009). HTESSEL includes a process-based single-layer snow scheme to represent the grid cell fraction (tile) that is covered by the snow (Dutra et al., 2010). In this scheme, the snowpack is characterized by a prognostic temperature, mass, density and albedo, updated at each time step. The liquid water content is diagnosed based on the other snow fields (temperature, density and mass), allowing to represent the interception of rainfall by the snowpack and internal melting/refreezing processes (Dutra et al., 2012).

UTOPIA is a land-surface process model representing the physical processes at the interface between surface, vegetation and soil layers, including a scheme which accounts for the main processes occurring in the snowpack (Cassardo, 2015). The snowpack is considered as a single homogenous layer placed upon land surface and its mass, thermal and hydrological balances are analyzed. The model takes into account the partition of soil coverage fractions (bare soil, vegetated soil, soil or vegetation covered by snow) and is able to simulate snow water equivalent, depth, density, albedo and coverage. Snow metamorphism is parameterized.

SMASH is a two-layer snow model that reproduces some of the main physical processes occurring within the snowpack, including accumulation, density dynamics, melting, sublimation, radiative balance, heat and mass exchanges (Piazzi et al., 2019). The model can be coupled with multivariable data assimilation schemes (Piazzi et al., 2018, 2019) allowing the joint assimilation of several snow-related observations to produce SWE and runoff estimates. With the aim of facilitating the implementation of the assimilation algorithms, the complexity of the modelling scheme is accordingly limited (e.g., liquid water storage and refreezing process are neglected). It is noteworthy that in the present study no assimilation scheme has been implemented in SMASH (open-loop configuration).

S3M is a spatially distributed, empirical snow model requiring only few input variables (precipitation, temperature, incoming shortwave radiation and air humidity) to compute the water mass conservation equation and to produce a first estimate of SWE (Boni et al., 2010). A second, optional, independent estimate of the SWE field, obtained combining spatial interpolation of surface snow depth observations and MODIS snow cover, is assimilated into the snow model using a nudging scheme. The result of the data assimilation is an updated SWE map exploiting different sources of information, modeling, remote sensing

Table 1. Features of snow models in terms of model complexity following Slater et al. (2001), snow albedo (α) parameterization, explicit representation of meltwater retention and refreezing in the snowpack (M_w) and a main reference.

Snow model	Complexity	α*	M_w	Reference
SNOWPACK	multi-layer	111	Yes	Bartelt and Lehning (2002)
GEOTOP	multi-layer	011	Yes	Endrizzi et al. (2014)
HTESSEL	single-layer	110	Yes	Dutra et al. (2012)
UTOPIA	single-layer	110	Yes	Cassardo (2015)
SMASH	up to 3 layers	110	No	Piazzi et al. (2018, 2019)
S3M	single layer	010	No	Boni et al. (2010)

*The three digits 1 and 0 represent the dependence or not of the albedo parameterization on surface temperature, snow age, grainsize. 000 means fixed albedo.

and surface stations network measurements. In the use of the model for the experiments proposed in this paper the assimilation scheme is switched off and the model runs in open loop configuration.

In the proposed experiment all the models are used in their default configurations, so no special tuning of the model parameters is done to improve the results over the Torgnon site. All models calculate snow water equivalent and snow density as primary variables, while snow depth is derived from them.

3 Study site and data

3.1 Torgnon station data

Meteorological forcing data are provided by a fully-equipped weather observation station located at Torgnon, 2160 m a.s.l. (45°50’ N, 7°34’E) in the Aosta Valley, Western Italian Alps. The experimental site belongs to the ICOS (IT-Tor, www.icos-ri.eu/) and LTER (lter_eu_it_077, www.data.ltereurope.net/deims/site) networks and it is described in detail by Galvagno et al. (2013), Filippa et al. (2015) and Piazzi et al. (2019). The location is a subalpine grassland, an abandoned pasture located a few kilometres from the village of Torgnon. The site is characterized by an intra-alpine semi-continental climate, with mean annual temperature and precipitation of 3.1°C and 880 mm respectively (Galvagno et al., 2013). During the cold season most of precipitation falls as snow and, on average, from the end of October to late May, the site is snow covered with snow depths reaching 90-120 cm (Galvagno et al., 2013). Wind-induced phenomena are limited in this site, since it experiences low winds, with an average half-hourly wind speed of 1.6 ± 1.3 m/s over the 2012-2014 period.

The station measures all the input variables needed to force the snow models, including air temperature, total precipitation, shortwave (SWIN) and longwave (LWIN) incoming radiation, wind direction and speed, relative humidity, surface pressure, surface temperature. These variables are measured at high frequency, and then aggregated at 30-minute temporal resolution. Precipitation measurements are performed with an OTT Pluvio2 Weighing Rain Gauge, which employs a weight-based tech-

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Discussions

nique to measure both liquid and solid fractions. This is a consolidated technique that provides higher confidence on the reliability of precipitation data than standard rain gauges (Kochendorfer et al., 2017). As the OTT pluviometer has been operational since mid-2012, in our analysis we consider the dataset spanning the period from October 1^{st}, 2012 to June 30^{th}, 2017, covering five complete snow seasons.

The Torgnon station provides also snow-related variables useful for model evaluation, including snow depth measurements, obtained by an ultrasonic distance sensor, surface temperature, snow and soil temperatures at different depths, outgoing shortwave and longwave radiation, all of them available at 30-minute resolution. Snow density and snow water equivalent are measured manually in snow pits several times per snow season during dedicated field campaigns. During the analysis period 20 manual measurements of snow density and snow water equivalent are available. Additionally, since January 2016 snow water equivalent is automatically monitored by a Campbell CS725 sensor, that passively measures the attenuation of naturally existing electromagnetic radiation (Potassium-40 and Thallium-208) emitted from the soil or bedrock below the sensor. The higher the water content of the snow pack, the higher the attenuation of the radiation. The measure is performed every 6 hours and averages the SWE over an area of about 100 m^2. Combining automatic snow water equivalent measurements and the corresponding snow depth measurements, additional daily snow density estimates useful for model validation have been derived for the last two snow seasons.

3.2 Spatial interpolation of meteorological forcings from neighboring stations

The spatial interpolation of ground meteorological observations represents one of the most commonly used practices in the operational applications of hydrological models. In order to test the performances of the models in this condition, an interpolated dataset has been generated for the Torgnon monitoring site by using the MeteoIO library (Bavay and Egger, 2014). Meteorological data from six neighboring stations have been interpolated over a squared digital elevation model of 16 km^2 with a grid resolution of 50 meters centered on the coordinates of Torgnon (Fig. A1 and Tab. A1). The algorithm used for the interpolation is the inverse distance weight (IDW) as first choice for all the meteorological variables. The interpolation accounts also for vertical gradients of both temperature and precipitation. Further details are provided in Appendix A.

3.3 Reanalysis data

In many remote mountain areas in-situ observations to force snow models are unavailable. In this study we explore the use of reanalysis datasets extracted at the Torgnon gridoint.

GLDAS (Global Land Data Assimilation System) is a global dataset exploiting satellite and ground-based observational data combined with advanced modelling and data assimilation techniques in order to generate optimal fields of surface variables (Rodell et al., 2004). In particular, the GLDAS-2.1 archive used in this study contains 36 land surface fields from January 2000 to present time at $0.25°$ (lon/lat) spatial and 3-hour temporal resolutions (Rui and Beaudoing, 2018).

ERA-Interim (Dee et al., 2011) is a global reanalysis including a variety of 3-hourly surface parameters describing atmospheric and land-surface conditions, and 6-hourly upper-air parameters covering the troposphere and stratosphere. ERA-Interim has spatial resolution of $0.75°$, at the latitude of Torgnon corresponding to about 59 km in the zonal and 83 km in the merid-

Table 2. Overview of the experiments and their characteristics in term of forcing data, temporal and spatial resolutions and gap-filling data employed where necessary.

Experiment	Forcing	Temporal resolution	Spatial Resolution	Gapfilling
CTL	Torgnon station	30’	point	ERAI
RAD-ERAI	CTL except SWIN and LWIN from ERAI in case of snowfall	30’	point	ERAI
SWIN-CLS	CTL except SWIN from Clearsky algorithm	30’	point	ERAI
TIME-3h	Torgnon station	3h	point	ERAI
TIME-6h	Torgnon station	6h	point	ERAI
TIME-12h	Torgnon station	12h	point	ERAI
MeteoIO	Six stations close to Torgnon (see Appendix A)	1h	point	none
GLDAS	GLDAS-2.1	3h	25 km	none
ERA5	ERA5	1h	30 km	none
ERAI	ERA-Interim	3h	80 km	none
ERAI-LR	ERAI, lapse-rate correction of air temp.	3h	80 km	none
ERAI-BIAS	ERAI, bias adjustment of air temp.	3h	80 km	none

ional direction. This coarse grid, which is comparable to those of state-of-the-art global climate models, implies a smooth representation of the topography and coarse information on climate variables.

ERA5 (Hersbach and Dee, 2016) is the latest ECMWF global reanalysis product, providing data at higher resolution than ERA-Interim, both in space (30 km) and in time (1 hour). ERA-5 uses one of the most recent versions of the Earth system model and data assimilation methods applied at ECMWF and modern parameterizations of Earth processes compared to older versions used in ERA-Interim. With respect to ERA-Interim, ERA5 has also an improved global hydrological and mass balance, reduced biases in precipitation, and refinements of the variability and trends of surface air temperature (Hersbach and Dee, 2016).

4 Experimental design

We devised a set of twelve experiments at the Torgnon site employing snow models in stand-alone mode, i.e. in which the meteorological forcing is prescribed. The list of experiments is summarized in Table 2. The first experiment is a control run (CTL) in which the models are forced by optimal input data provided by the Torgnon station at 30-minute temporal resolution. This run allows to test the accuracy of the models in describing the temporal evolution of the snow-related variables in optimal conditions, namely when high-quality, high-frequency point measurements are available.

Experiments LWIN-ERAI and SWIN-CLS assess the sensitivity of the models to the radiation input. As most stations, Torgnon site is equipped with an unheated radiation sensor, which is likely to provide unreliable measurements when getting obstructed by snow during snowfall events. Therefore, in the experiment RAD-ERAI we take into account the shading of the

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

radiation sensor in case of snowfall by replacing radiometer measurements with ERA-Interim reanalysis data. In the third experiment, SWIN-CLS, we employ external SWIN data resulting from the clear sky radiation (Yang et al., 2001, 2006) attenuated through the cloud masks from the Meteosat Second Generation satellite in the following way. For each of the 34 radiometers in the Aosta Valley an averaged attenuation factor F is computed as:

$$F = \frac{1}{N}\sum_{i=1}^{N}\frac{R^{i}{}_{st}}{SWIN^{i}} \tag{1}$$

where N is the number of cloud-covered stations, $R^{i}{}_{st}$ is the measured radiation at the ith station and $SWIN^{i}$ is the corresponding modeled radiation in clear sky condition. The incident solar radiation in cloudy conditions at the location j is given by:

$$SWIN^{j}{}_{inc} = R^{j}F \tag{2}$$

Experiments TIME-3h, TIME-6h and TIME-12h investigate the sensitivity of the models to the temporal resolution of the meteorological forcing, since the temporal resolution of many available datasets is coarser than that employed in the CTL run. We basically employ Torgnon data every 3, 6 and 12 hours since October 1st, 2012 time 00:00 UTC, and linearly interpolate them at the 30 minute time step for all variables except for total precipitation. Precipitation is accumulated over 3, 6 or 12 hour time periods and the totals are equally distributed among the corresponding 30 minutes subperiods.

Four additional experiments, namely MeteoIO, GLDAS, ERA5 and ERAI test the case in which no surface station measurement is available and one has to rely on external data. The MeteoIO experiment employs a forcing dataset obtained through the spatial interpolation of data provided by the neighboring stations (see Sect. 3.2 and Appendix A). GLDAS, ERA5 and ERAI experiments use different reanalysis products described in Sect. 3.3, namely GLDAS-2.1, ERA5 and ERA-Interim. Both MeteoIO and reanalysis data required to be rearranged and interpolated to 30-minute resolution in order to be used as forcing for snow models. In the case of ERA-Interim, for example, forecasts are initialized only twice a day at 00:00 UTC and 12:00 UTC and accumulated fluxes of total precipitation, surface solar and thermal downward radiation are available as forecasts at 3-hour intervals for the following 12 hours. From these forecasts we derive the average fluxes over 3-hour intervals and we assume the fluxes to be constant during each interval. For the other ERA-Interim parameters, namely 2-meter temperature, dew-point temperature, surface pressure, 10 metre U and V wind components, we consider the analyses at 00:00, 06:00, 12:00, 18:00 UTC and the forecasts at +3 hours. These data are linearly interpolated in time to the integration time step (30 minutes) of the snow models. Some calculations are necessary to obtain all the variables required by the models. For example, ERA-Interim does not directly provide relative humidity, which we derive using the Magnus formula from the dewpoint temperature and the 2-meter air temperature (Lawrence, 2005).

The last two experiments, ERAI-LR and ERAI-BIAS, investigate if it is possible to improve the performances of snow models when they are forced by reanalyses, for instance ERA-Interim, by applying two simple bias-correction methods to adjust air temperature and hence the amount of solid precipitation with respect to the total one. In ERAI-LR we take into account the fact that ERA-Interim has a smoothed topography and the altitude of the gridpoint closest to the Torgnon station is 680 m lower than the actual elevation of the station. In ERAI-LR experiment we adjust the temperature data assuming a fixed

Hydrology and Earth System Sciences
Open Access
Discussions

moist lapse rate of 6.5°C/km. This correction consists in a cooling of 4.4°C with respect to the original temperature data. In the ERAI-BIAS experiment we correct ERA-Interim air temperature using the difference in the climatological averages between ERA-Interim data and the Torgnon station observations, which was found to be 2.1°C. This bias is assumed to be constant in time and it is subtracted from the original ERA-Interim temperature time series.

A desirable feature of each experiment is that the differences in the model outputs are mainly due to the internal model characteristics rather than to the different parameterizations used by the models to derive the solid and liquid precipitation from the total precipitation input. To this end, we calculate externally the rainfall and the snowfall amounts using a fixed threshold on wet-bulb temperature. Specifically, precipitation is considered as snowfall when wet-bulb temperature is lower than or equal to 1°C and as rainfall otherwise. A slightly different approach was used for GEOTOP which requires precipitation totals (rather than solid and liquid precipitation separately) and then it separates rainfall and snowfall through an internal parameterization based on a fixed threshold on dew-point temperature. In this case the dew-point temperature threshold has been calibrated to obtain approximately the same seasonal accumulated snowfall as that obtained with the method based on wet-bulb temperature. This condition is satisfied with a dew-point temperature threshold of 1.2°C. Both approaches relies on the fact that the temperature interval where rain and snow cohexists is more narrow for wet-bulb temperature and dew-point temperature than for air temperature, thus using the wet-bulb or dew-point temperature allows to reduce the range for which the precipitation phase is uncertain (Sims and Liu, 2015; Endrizzi et al., 2014). With this procedure all the models are driven with the same rainfall and snowfall inputs and the differences in the model simulations are assumed to depend maily on the model structure and on the estimated snow ablation through melting, evaporation and direct air-snow sublimation (Slater et al., 2001).

5 Results

5.1 CTL - impacts of the snow model structure

We run the six models driven by the best forcing available for the Torgnon site, namely the Torgnon station measurements at 30-minute resolution. Figure 1 shows the simulated SWE, snow density (ρ) and snow depth (SD) time series provided by each model compared to the observations, over the period 2012-2017.

All the models are able to reproduce the overall variability of snow characteristics, although with different accuracy. The agreement between simulations and observations is evaluated in terms of centered pattern root mean square error, standard deviation and temporal correlation, and the resulting statistics are summarized through Taylor diagrams (Taylor, 2001) in Fig. 2. Evaluation metrics are calculated over simulated and observed pairs when at least one of the two values exceeds a minimum threshold, namely SWE > 0.005 m, SD > 0.01 m. Snow density pairs are compared in case both the corresponding values of SWE are greater than 0.005 m. The upper panels of Fig. 2 refer to the period 2016-01-01 to 2017-06-30, when continuous measurements of all the three variables are available. Bottom panels refer to the full period of analysis (since 2012-10-01) for which continuous observations are available for snow depth only.

Hydrology and Earth System Sciences
Open Access
Discussions

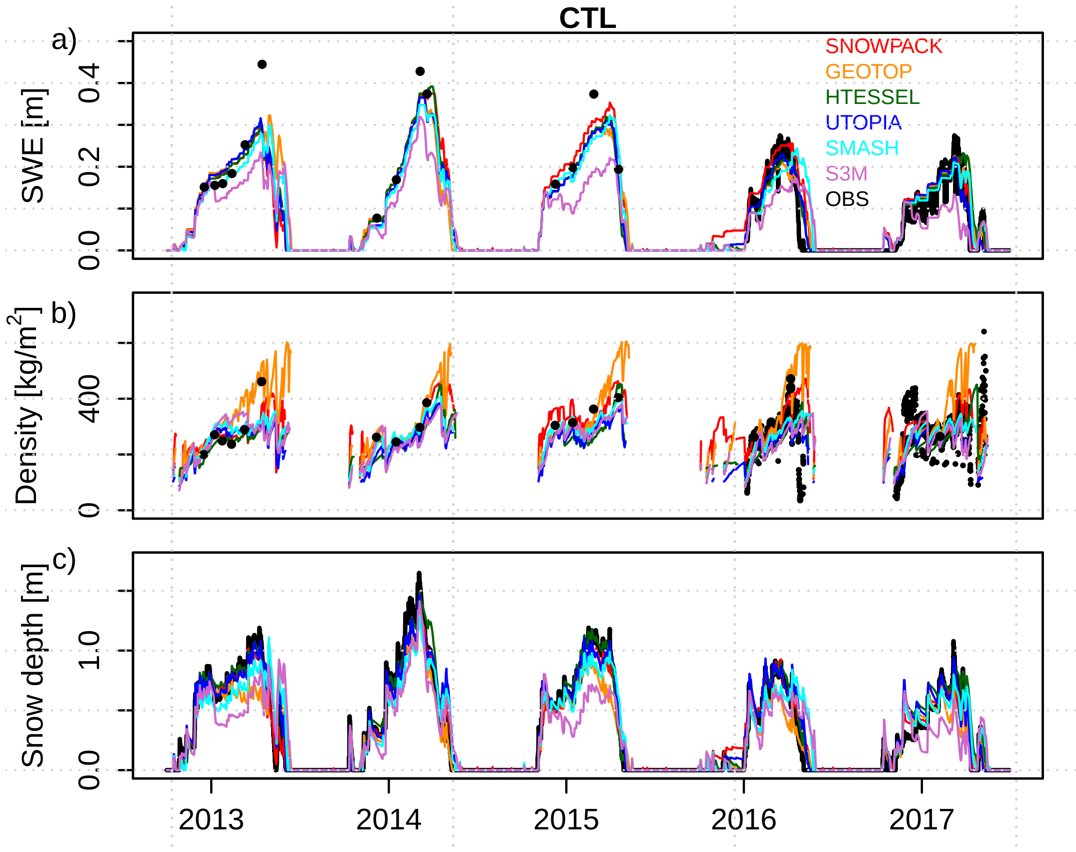

Figure 1. Results of the CTL experiment: a) snow water equivalent (SWE), b) snow density and c) snow depth simulated by the six models considered in the analysis, driven by optimal forcing, i.e. Torgnon station measurements at 30 minute resolution, over the period 2012-2017, compared to observations (black).

Snow water equivalent simulations are in overall good agreement with observations over the period 2016-2017 (Fig. 2a), although with some differences between the models. The best agreement is found with the SNOWPACK, HTESSEL, UTOPIA and GEOTOP models, showing the lowest errors (below 0.04 m SWE) and the highest correlations (above 0.85) with observations. SMASH and S3M are characterized by higher RMSE and lower correlation with observations with respect to the best performing models.

Snow density is simulated with lower skills compared to SWE for all models (Fig. 2b). The agreement between model simulations and observations is rather low for all models, with limited added value from highly sophisticated models. A weak correlation (lower than 0.6) and large errors (above 70 kg/m^2) are found for both SNOWPACK and S3M, namely the most sophisticated and the simplest model, respectively. The GEOTOP model has clear deficiencies in representing spring snow density, in fact it exhibits an overestimation error increasing with time till the end of the snow season.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

The ability of the models in reproducing the temporal evolution of snow depth is related to their skills in reproducing both snow mass and density. The SNOWPACK model reproduces with high scores all the three variables, namely SWE, snow density and snow depth. In the case of GEOTOP, the overestimation of spring snow density is reflected in overall lower skills in reproducing snow depth compared to the other intermediate-complexity models (Fig. 2c). In the case of HTESSEL, instead, small errors in SWE and snow density are compensated and the model skill in reproducing snow depth results slighly higher compared to that of the SNOWPACK model.

Globally the high- and intermediate-complexity models SNOWPACK, HTESSEL and UTOPIA, show similar and good performances in the simulation of both SWE and snow depth and they can be considered the best performing models. Compared to them, SMASH and S3M are characterized by higher RMSE and lower correlation with observations, with the simplest model, S3M, showing the lowest agreement with observations. In this experiment the model complexity is broadly reflected in the model performances, with the most sophisticated model performing best and the simplest model performing worst, likely owing to difficulties of the latter in representing snow melting (Fig. 1a). HTESSEL and UTOPIA, which are single-layer intermediate-complexity snow models performing almost as well as the most sophisticated model SNOWPACK, seem a good trade-off between model complexity and model accuracy when accurate meteorological forcing is employed.

We extend this analysis to a longer period of five complete snow seasons, from 2012 to 2017, limited to the snow depth variable. The relative skills of the models in reproducing snow depth over the full five-seasons period are very similar to those found for the last two-seasons period (Fig 2c,d). The RMSE of the models remains almost unchanged, while the correlation with observations slightly improves over the longer period. The behavior of the models is robust whether considering all the five seasons or only the last two seasons.

Figure 2e allows to investigate the variability of the model performaces in the different snow seasons compared to the whole period. SNOWPACK, HTESSEL and UTOPIA show similar skills across different snow seasons, meaning robustness in reproducing a variety of conditions. Common simulation errors for several models are a positive SWE and a positive snow depth bias in the season 2015-2016 (Fig. 1a,c), when several conditions which are challenging to simulate occurred. First, autumn isolated snowfalls occurred with snow-free conditions in-between: mainly the SNOWPACK model, and to a lesser extent UTOPIA model, failed to reproduce the rapid melting and they continued accumulating snow in time. Second, at the end of the snow season a very rapid melting occurred, which was not captured by any of the models. All the models simulate a meltout date delayed by several days with respect to observations. Di Mauro et al. (2018) demonstrated that the accelerated snowmelt, observed in 2015-2016 season, was caused by the deposition of mineral dust from Sahara: light absorbing impurities in snow, resulting from several dust deposition events, induce albedo reduction that alters the melting dynamics of the snowpack hence advancing snowmelt. As none of the model used in this study accounts for the impact of impurities on snow dynamics, modeled snow melt dates in 2016 were not surprisingly significantly delayed.

The GEOTOP, SMASH and S3M models show different skills depending on the snow season (Fig. 2e) and they provide a wider range of variability in their agreement with observations compared to SNOWPACK, HTESSEL and UTOPIA. For example, a season which is relatively simple to reproduce by all models is 2013-2014. An abundant but ephemeral snow cover was properly accumulated and melted by all models. After a snow-free period, the onset of a persistent snow cover

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

was sustained by heavy snowfalls which led to the highest snow peak in the study period. After this peak, the melting has been quite steady, with few spring snowfall events. These conditions allow all models, even the simplest one, to accurately reproduce the snowpack evolution in terms of snow mass and depth. As a result, for this season the differences between the models in terms of RMSE, standard deviation and temporal correlation with observations are smaller than for other seasons. On the contrary, the season 2012-2013 is more difficult to reproduce for some models, namely GEOTOP, SMASH and S3M, than for SNOWPACK, HTESSEL and UTOPIA. This season was characterized by many snowfall episodes of moderate and light intensity, with moderate melting in-between. In the second half of May 2013 a series of late snowfalls restored a temporary snow cover with more than 0.5 m depth that gradually melted in a couple of weeks. In this conditions, SNOWPACK, HTESSEL and UTOPIA are able to represent quite accurately the changes in the snow depth, while GEOTOP, SMASH and S3M generally tend to overestimate snow depth.

GEOTOP sistematically overestimates snow density with increasing errors from late winter to the end of the snow season. These errors are reflected in the snow depth simulations: spring snow depth and the snow depth peak are underestimated in each snow season of the study period. SMASH, for the 2012-2013 season as well as for the 2015-2016 season, misrepresents the correct timing of the snow depth and snow mass peaks and delays them in the snow season. The delay in the representation of the snow peaks is almost fully compensated by a too rapid spring melting which allows to keep the date of ablation relatively close to the observed one. S3M sistematically underestimates both snow depth and snow water equivalent during all the snow seasons, while the snow density is within the range of variability of the model ensemble. It follows that for S3M the critical variable to improve is SWE.

In conclusion, an added value of sophisticated and intermediate-complexity models compared to lower-complexity models emerges especially during snow seasons that have a more complex temporal behavior.

5.2 RAD-ERAI, SWIN-CLS - model sensitivity to the radiation input

A typical problem occurring in case of snowfall is that when the radiation sensors get covered by snow they record inaccurate or even wrong data. To take into account this issue and test how it affects snow simulations, in the experiment RAD-ERAI we use incoming longwave and shortwave radiation data from the Torgnon station except in case of snowfall, when we employ external LWIN and SWIN data derived from ERA-Interim. In the other experiment, SWIN-CLS, we replace observed incoming shortwave radiation data with the external data described in Sect. 4. The results of these simulations are reported in Table 4. Although the difference between external data and Torgnon data can be high at the time step of the model (not shown), their overall impact on snow simulations is low. In fact, for each model we obtain values of RMSE close to those obtained in the CTL experiment. In particular, model skills does not improve using ERA-Interim or interpolated incoming radiation forcing.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

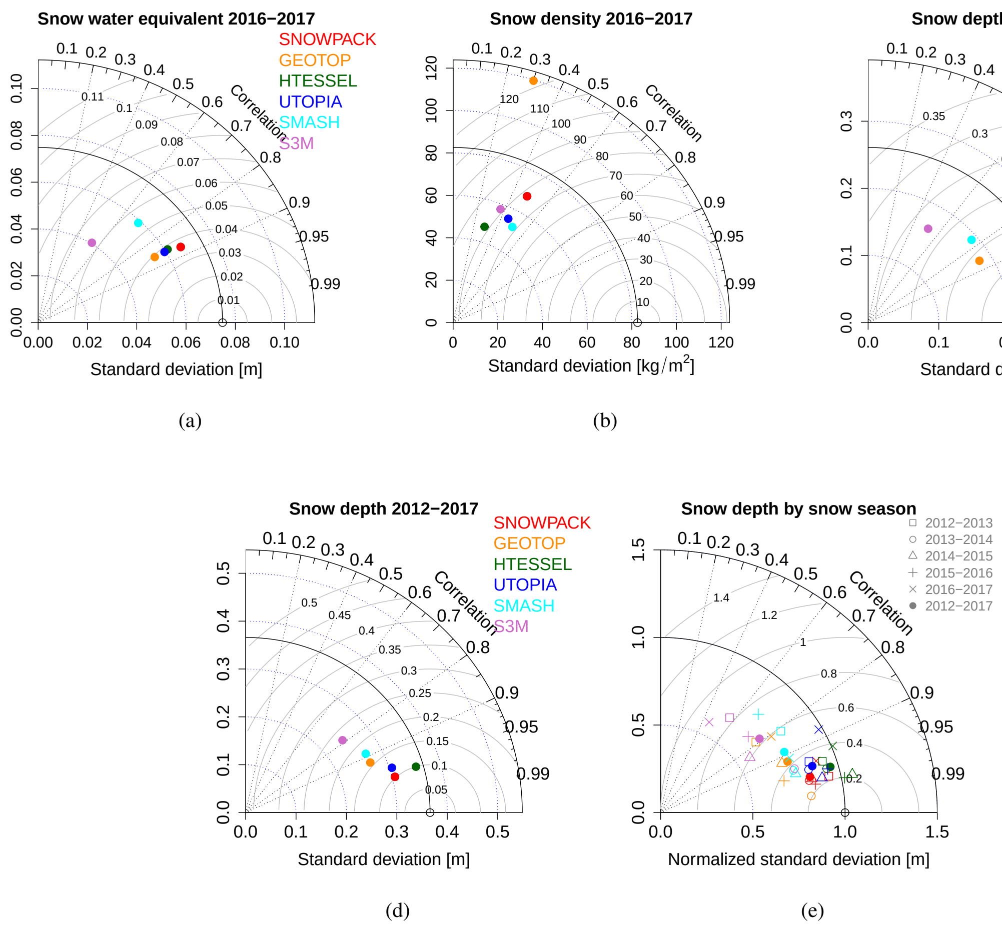

Figure 2. Taylor diagrams of the modeled vs. observed a) snow water equivalent, b) snow density and c) snow depth in the control experiment (CTL) for the period 2016-01-01 to 2017-06-30. Bottom panels represent the statistics of snow depth d) for the whole period 2012-2017 and e) for each snow season in the same period. Please note that, differently from other panels, in panel e) model standard deviations are normalized with respect to the observed ones.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Table 3. RMSE, bias and Pearson correlation of snow depth simulations with respect to observations for each model and each snow season in the control experiment (CTL).

			RMSE				
Model	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	Avg.	Std.
SNOWPACK	0.08	0.12	0.10	0.12	0.08	0.10	0.02
GEOTOP	0.21	0.14	0.18	0.13	0.15	0.16	0.03
HTESSEL	0.11	0.13	0.07	0.08	0.15	0.11	0.03
UTOPIA	0.11	0.14	0.07	0.12	0.13	0.11	0.03
SMASH	0.19	0.19	0.12	0.23	0.12	0.17	0.05
S3M	0.28	0.20	0.28	0.21	0.24	0.24	0.04
			BIAS				
Model	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	Avg.	Std.
SNOWPACK	-0.04	-0.03	-0.05	0.10	0.01	0.00	0.06
GEOTOP	-0.05	-0.11	-0.11	-0.06	-0.01	-0.07	0.04
HTESSEL	0.04	0.04	0.01	0.06	0.11	0.05	0.04
UTOPIA	0.03	-0.01	-0.01	0.09	0.06	0.03	0.04
SMASH	-0.04	-0.09	-0.04	0.05	0.04	-0.02	0.06
S3M	-0.08	-0.12	-0.21	0.01	-0.10	-0.10	0.08
			Pearson Correlation				
Model	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	Avg.	Std.
SNOWPACK	0.98	0.98	0.97	0.98	0.94	0.97	0.01
GEOTOP	0.79	0.99	0.92	0.97	0.81	0.90	0.09
HTESSEL	0.95	0.96	0.98	0.98	0.93	0.96	0.02
UTOPIA	0.94	0.96	0.98	0.96	0.87	0.94	0.04
SMASH	0.81	0.95	0.96	0.69	0.91	0.86	0.11
S3M	0.57	0.95	0.84	0.74	0.45	0.71	0.20

Hydrology and Earth System Sciences
Open Access
Discussions

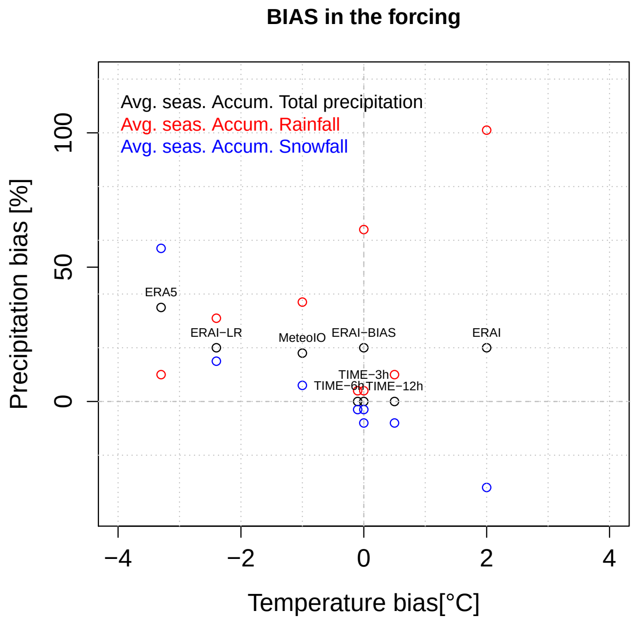

Figure 3. Temperature, total precipitation, rainfall and snowfall average seasonal biases in the forcings exployed in each experiment with respect to the Torgnon station measurements.

5.3 TIME-3h, TIME-6h, TIME-12h - model sensitivity to the temporal resolution of the forcing

A common condition when modeling snowpack evolution in data-sparse areas is the unavailability of meteorological forcings with high temporal resolution, as high as 30 minutes, like that employed in the CTL experiment. In this section we assess the sensitivity of the models to the temporal resolution of the forcing. To this aim, the original 30-minute resolution meteorological observations at Torgnon have been sampled every 3, 6 and 12 hours, and then linearly interpolated at a finer (30-minute) time step, with the only exception of total precipitation that has been accumulated over the 3, 6, 12 hour periods and then equally distributed among the 30-minute sub-periods (see Sect. 4 for details). As expected, the longer the sampling period the smoother are the input time series. For these three (and the other remaining six) experiments, we show in Fig. 3 the biases of air temperature, total precipitation, rainfall and snowfall forcings with respect to the reference forcing of the CTL experiment. Given the method employed to derive TIME-3h, TIME-6h, TIME-12h forcings we expect no bias for total precipitation, while some differences can arise in the rainfall/snowfall partition owing to possible differences in air temperature. According to Fig. 3, TIME-3h and TIME-6h air temperature biases are close to zero, while TIME-12h air temperature bias is about $0.5°C$, with the effect of reducing the amount of the solid precipitation by 10%. We investigate the impact of these biases on the snow simulations in the following.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences Discussions

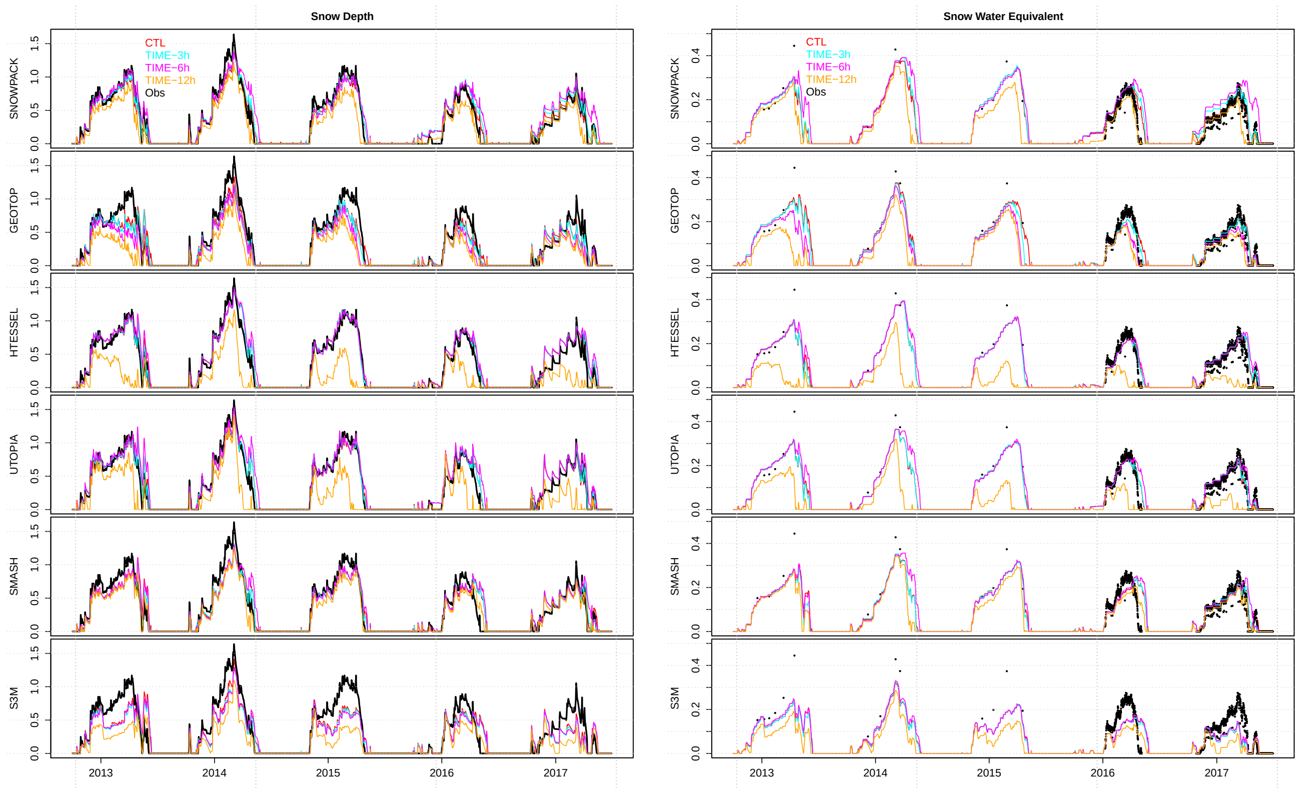

Figure 4. Model simulations of snow depth and SWE when the input is sampled at 3, 6 and 12 hours, compared to the CTL run and to the observations.

Figure 4 represents, for all the models, the simulated snow depth and SWE when input data are sampled (or accumulated, in case of total precipitation) at 3, 6 and 12 hours and then interpolated (or equally distributed for precipitation) over 30-minute time steps, compared to the simulated snow depth obtained with the original 30 minute resolution forcing (CTL) and compared to observations. In addition, Table 4 reports the RMSE associated with these simulations.

The model response to the degradation of the temporal resolution of the forcing can vary remarkably depending on the model and season. A common feature of the models is the small (or null) difference in terms of RMSE between TIME-3h and CTL simulations, indicating that using forcing data at 3-hour temporal resolution generates snow depth simulations almost as accurate as in the case of 30-minute resolution input data.

A second common feature is the general worsening of model performances when using input data at a lower temporal resolution than 3 hours, reflected into an increase of the RMSE values. TIME-6h simulations are usually very close to the CTL in the accumulation period up to the snow peak. Afterwards, in the melting period, some models, mainly SNOWPACK, UTOPIA and to a lesser extent HTESSEL and SMASH, slightly overestimate snow mass/depth in selected seasons, contributing to an increase in the model RMSE. The largest RMSE values are generally obtained with input data at 12 hour resolution: in this case all models represent a thinner snow depth throughout the snow season compared to the CTL run. The snow depth underestimation in TIME-12h expriement can be related to biases in the forcing, namely the positive air temperature bias and the underestimation of snowfall by 10% (Fig. 3) combined with a poor representation of the incoming shortwave radiation. In fact, TIME-12h incoming shortwave radiation forcing exceeds, in average, the CTL values by $+97$ W/m^2. This large overestimation of the incoming shortwave radiation explains most of the underestimation of the surface snow depth. On the contrary, the TIME-6h incoming shortwave radiation forcing shows, a small negative bias (-7 W/m^2) compared to the CTL forcing, which can contribute to the overestimation of the snow depth at the end of the snow season.

The different sensitivities of the models to the bias in the forcing is quite impressive. While for models with higher complexity (SNOWPACK, GEOTOP, HTESSEL and UTOPIA) the error on snow depth increases at increasing temporal degradation of the forcing, for simpler models (SMASH and S3M) the performance is less affected. SMASH, for example, shows little differences between TIME-3h, TIME-6h, TIME-12h and the CTL simulations, revealing to be the less sensitive model to the time resolution of the forcing. In the TIME-12h experiment, SMASH provides the smallest error on snow depth and a similar error on SWE as the most sophisticated model, SNOWPACK. Similarly to SMASH, S3M shows very little differences between TIME-3h, TIME-6h simulations and the CTL simulation, while its perfomances in the TIME-12h experiment falls within the range of values obtained for intermediate-complexity models. In our experiment the lower-complexity models reveal a lower sensitivity to errors in the meteorological forcing, and when these models are driven with 6-hourly and 12-hourly input data they provide similar performances as selected intermediate-complexity models. So, with low temporal resolution data the performances of intermediate and low complexity models get comparable. However, from these experiments an added value of the most sophisticated model SNOWPACK emerges. SNOWPACK forced by the 12-hour resolution forcing still provides lower errors than the simplest model S3M forced by the best available forcing at 30 minute temporal resolution (Table 4).

5.4 MeteoIO, GLDAS, ERA5 and ERA-Interim - model sensitivity to the spatial resolution and bias in the forcing

We consider a rather standard case for which no station measurements are available for the area of interest and one has to rely on gridded datasets, which are generally characterized by lower resolution and lower accuracy with respect to station measurements. To explore a representative range of possible alternatives we employ datasets with different characteristics: the MeteoIO dataset, based on the interpolation of data from neighboring stations, GLDAS, ERA5 and ERA-Interim reanalyses at 25, 31 and 80 km respectively. An overview of the comparison between the meteorological forcing provided by these datasets and the observations in Torgnon is shown in Fig. 3.

MeteoIO forcing is in fairly good agreement with observations. Compared to the meteorological measurements at Torgnon station, MeteoIO shows an average bias of -1°C per snow season and about 20% overestimation of the seasonal total precipitation. However, the effect of these biases on the solid precipitation is weak, and the average seasonal snowfall is very close to the observations. When the MeteoIO forcing is used, the best agreement between simulated and observed SWE and snow depth is obtained with the GEOTOP and SMASH models. Both models provide similar RMSE values as in the CTL runs. Conversely, the SNOWPACK, HTESSEL and UTOPIA model errors are respectively more than twice and three times the corresponding errors in the CTL run, revealing a remarkable sensitivity to relatively small errors in the input data. The S3M model exhibits

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Table 4. RMSE values associated to snow depth and snow water equivalent simulations for all models and all experiments over the periods 2012-2017 and 2016-2017, respectively.

	RMSE snow depth [m]					
Exp	SNOWPACK	GEOTOP	HTESSEL	UTOPIA	SMASH	S3M
CTL	0.10	0.17	0.11	0.12	0.17	0.25
RAD-ERAI	0.12	0.17	0.14	0.13	0.17	0.25
SWIN-CLS	0.11	0.21	0.12	0.13	0.18	0.24
TIME-3h	0.12	0.19	0.11	0.12	0.16	0.26
TIME-6h	0.17	0.26	0.15	0.18	0.19	0.27
TIME-12h	0.21	0.37	0.44	0.38	0.17	0.38
MeteoIO	0.23	0.20	0.38	0.40	0.19	0.31
GLDAS	1.99	2.45	3.61	3.46	1.89	3.55
ERA5	0.74	0.34	0.76	0.80	0.71	0.85
ERAI	0.18	0.45	0.20	0.20	0.27	0.32
ERAI-LR	0.54	0.20	0.58	0.67	0.36	0.46
ERAI-BIAS	0.18	0.27	0.20	0.26	0.13	0.16
	RMSE SWE [m]					
Exp	SNOWPACK	GEOTOP	HTESSEL	UTOPIA	SMASH	S3M
CTL	0.04	0.04	0.04	0.04	0.06	0.08
RAD-ERAI	0.06	0.04	0.05	0.04	0.06	0.08
SWIN-CLS	0.05	0.04	0.04	0.03	0.06	0.07
TIME-3h	0.06	0.03	0.04	0.03	0.06	0.08
TIME-6h	0.09	0.05	0.05	0.05	0.07	0.07
TIME-12h	0.05	0.07	0.13	0.13	0.05	0.10
MeteoIO	0.10	0.04	0.13	0.13	0.07	0.11
GLDAS	1.31	1.68	1.77	1.82	1.19	1.56
ERA5	0.28	0.12	0.22	0.24	0.26	0.24
ERAI	0.05	0.12	0.05	0.05	0.08	0.08
ERAI-LR	0.19	0.04	0.18	0.19	0.13	0.15
ERAI-BIAS	0.05	0.05	0.03	0.05	0.03	0.05

Hydrology and Earth System Sciences
Open Access
Discussions

a moderate decrease in the model performance when driven by MeteoIO, with lower RMSE than the HTESSEL and UTOPIA models.

The GLDAS forcing is affected by strong cold and wet biases. With an average temperature bias of $-3.8°C$ and an average total precipitation bias of 296%, GLDAS lies outside the range displayed in Fig. 3. As expected, the large errors in GLDAS forcings lead to huge errors in simulated snow water equivalent and depth for all models, as confirmed by RMSEs lying out of range in Fig. 6a,c,d.

ERA5 has a large temperature bias (-3.3°C) like GLDAS, but a lower precipitation bias (+35%). The combined effect on the snowfall input is an excess of more than 50% compared to observations (Fig. 3), which clearly affects the snow model output. As expected, all models overestimate snow depth and the duration of the snow cover. The models tend to reproduce a similar evolution of snow depth as in the CTL experiment but with thicker snowpack. In detail, SNOWPACK, HTESSEL and UTOPIA give similar snow depth outputs, consistently with the behavior found in the CTL run. GEOTOP provides the lowest RMSEs for snow water equivalent and snow depth, but this is mainly due to a compensation between the error in the ERA5 forcing (leading to overstimation) and the model error identified in the CTL experiment (leading to underestimation). In general, the difference in performance between models of different complexity is reduced when ERA5 forcing is used. For example the RMSE is similar for SMASH and the most sophisticated model SNOWPACK, as is for S3M and HTESSEL or UTOPIA.

ERA-Interim forcing (ERAI) shows a $+2°C$ temperature bias and a snowfall deficit of about 30% compared to Torgnon observations. When forced by ERA-Interim data, GEOTOP, SMASH and S3M underestimate snow depth in all seasons, while SNOWPACK, HTESSEL and UTOPIA underestimate snow depth mainly during the season 2014-15, when ERA-Interim snowfall is considerably lower with respect to observations throughout this snow season (Fig. 5a). In other snow seasons, for example 2013-14, 2015-16 and 2016-17, SNOWPACK, HTESSEL and UTOPIA snow depth simulations are in fairly good agreement with observations (see for example Fig. 5b). Overall, SNOWPACK, HTESSEL and UTOPIA provide relatively good results when forced by ERA-Interim, with a moderate loss of accuracy with respect to the case of optimal forcing (CTL). In the following we explore the possibility to reduce the RMSE of the other intermediate- and low-complexity models by correcting the main biases in the meteorological forcings.

5.5 Impact of the bias adjustment of ERA-Interim air temperatures

We test the effect of two very simple bias correction techniques applied to ERA-Interim air temperature. In the first approach, in the ERA-LR experiment, we take into account the difference in elevation between ERAI-Interim gridpoint at Torgnon and the true elevation of this site applying a lapse rate correction, i.e. substracting 4.4°C from the original ERAI data. Alternatively, in the ERAI-BIAS experiment, we remove the average bias of ERA-Interim data at the Torgnon site with respect to the station measurements, i.e. substracting 2.6°C from the original ERAI data.

The lapse rate correction reduces too much ERA-Interim temperatures, in fact the average temperature bias shifts from +2 to -2.4°C and the snowfall amount increases from -32 to +15% (Fig. 3).

Hydrology and Earth System Sciences
Open Access
Discussions

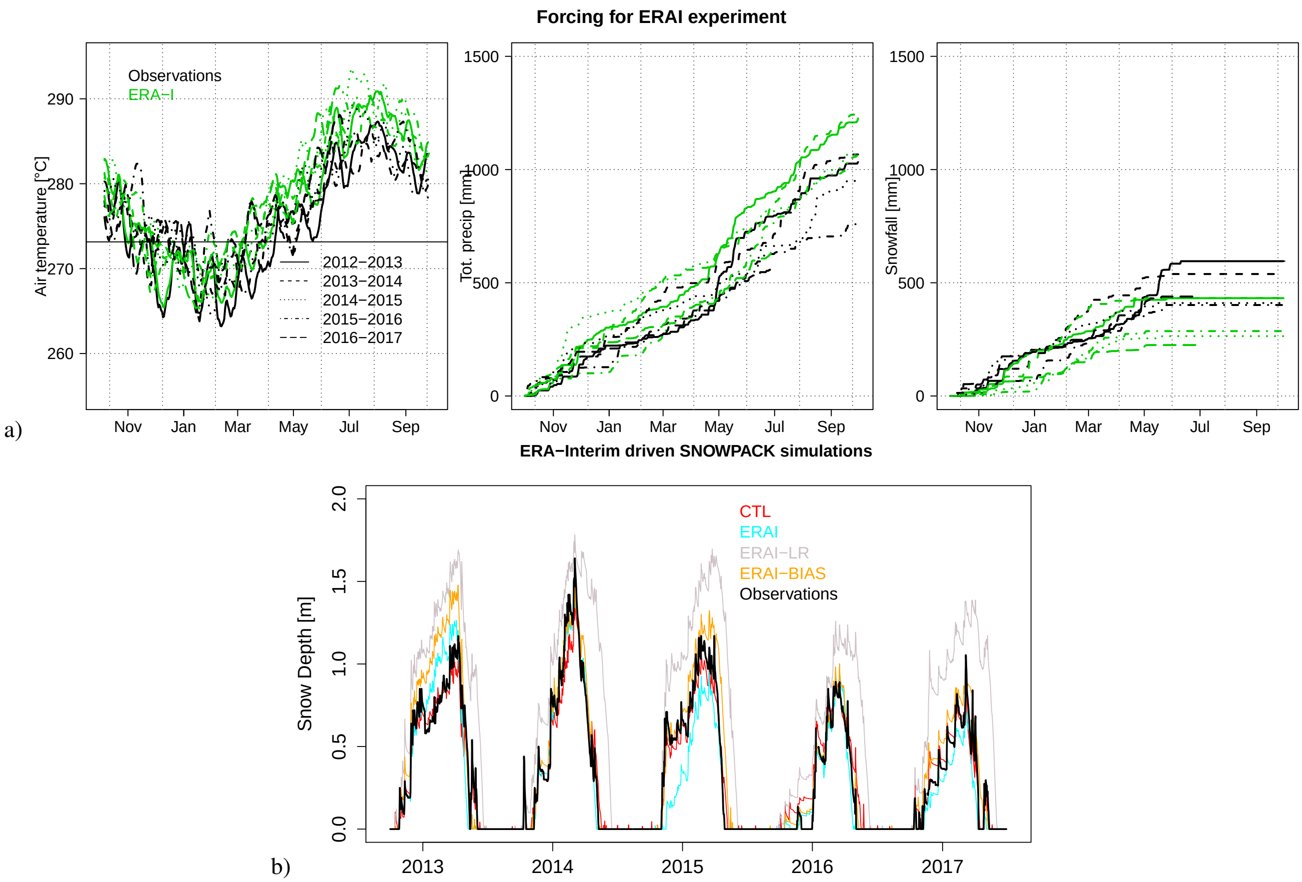

Figure 5. a) ERA-Interim air temperature, total precipitation and snowfall (derived as explained in Sect. 4) at the Torgnon site, compared to the station measurements (black) for each snow season of the period 2012-2017; b) Example of ERA-Interim driven snow depth simulations (ERAI, ERA-LR and ERAI-BIAS experiments), obtained with the SNOWPACK model, compared to CTL run and snow depth observations.

The net effect on the model outputs (ERAI-LR experiment) is an overestimation of snow water equivalent and snow depth. With respect to the ERAI experiment, the RMSE values increase for all models except for GEOTOP, which actually shows a good agreement with observations during the seasons 2013-14 and 2015-16, while it overestimates snow depth in the first half of the other seasons. The GEOTOP underestimation error observed in the ERAI experiment is here compensated by too cold input air temperature, which favor the development and the duration of the snowpack.

The correction based on the adjustment of the mean ERAI temperature bias (ERAI-BIAS experiment) allows to almost remove the snowfall bias. Therefore, this approach guarantees the most effective correction to improve the agreement of the forcing data with Torgnon station measurements. Clearly this approach requires to know at least the average temperature at the site of interest. This correction successfully reduces the RMSE on snow water equivalent and snow depth simulations with respect to the corresponding runs driven by raw ERA-Interim data for GEOTOP, SMASH and S3M models. For the most sophisticated SNOWPACK model, the correction applied to ERA-Interim data has no effects on the RMSE values of snow

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Open Access
Discussions

water equivalent and snow depth simulations, which remain unchanged. In fact, while the simulated snow depth is generally close to observations, improvements gained in the selected seasons (i.e. 2014-15) are compensated by lower performances in other (2012-13) seasons (Fig. 5d), so that, in average, the overall effect on the RMSE is negligible. For the UTOPIA model, the correction applied to ERA-Interim data has no effects on the snow water equivalent, however it slightly increases the error on snow density (lower correlation with available observations) and thus the error on snow depth simulations.

5.6 Discussion

While much work has been done to characterize the performances of snow models when driven by accurate input data, model responses depending on different degrees of accuracy of the input data still needs to be explored in detail. This study offers some hints on this research topic by assessing the simulations of six state-of-art snow models driven by input data with varying accuracy, focusing on the fully-instrumented Torgnon site, in the NW Italian Alps. The snow models selected for the analysis present also different degrees of complexity, from highly sophisticated multi-layer snow models to rather simple single-layer snow models, with the aim of sheding light also on the relations and trade-offs between model complexity and model performances in reproducing snowpack dynamics.

In our experiment, in case of optimal forcing, namely Torgnon station data at 30-minute resolution, the most sophisticated model SNOWPACK and the intermediate-complexity models HTESSEL and UTOPIA show the best agreement with observations. In particular HTESSEL and UTOPIA models, with their single-layer, simpler snow schemes compared to SNOWPACK, can be considered a good trade-off between model complexity and model accuracy. When considering snow depth simulations, for which validation data are available for a longer period than for SWE, an added value of these high- and intermediate-complexity models compared to lower complexity models is evident, especially in the snow seasons that are more difficult to reproduce. SNOWPACK, HTESSEL and UTOPIA, in fact, show similar and good performances across different seasons, revealing robustness in reproducing a variety of conditions, while the simpler snow models SMASH and S3M show larger dispersion of the seasonal scores.

Snow density is more difficult to simulate than SWE and snow depth for all models. The correlation between model simulations and observations is quite low for all models, with no clear added value from highly sophisticated ones (Fig. 2b). In fact, similar RMSE values with respect to snow density observations are found for SNOWPACK, UTOPIA, SMASH and S3M models; GEOTOP, instead, provides a much larger error especially in the spring season, suggesting further checks on the snow density parameterization.

The response of the snow models forced by gradually lower accuracy data is summarized in Fig. 6, showing the model RMSE for all experiments and all variables (upper panels) and the complementary information on the model ranking (bottom panels). No remarkable differences can be detected in the model skills when usign alternative radiation data instead of the Torgnon station measurements, as done in experiments RAD-ERAI and SWIN-CLS. The substantially equivalent results obtained replacing measured data with ERA-Interim data in case of snowfall (RAD-ERAI experiment) can be explained by the combination of two conditions: the intermediate elevation of 2160 m a.s.l. and the orientation of the Torgnon site, both likely contributing to a rapid melting of the snow obstructing the radiometer. This adjustment does not affect model performances.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Open Access
Discussions

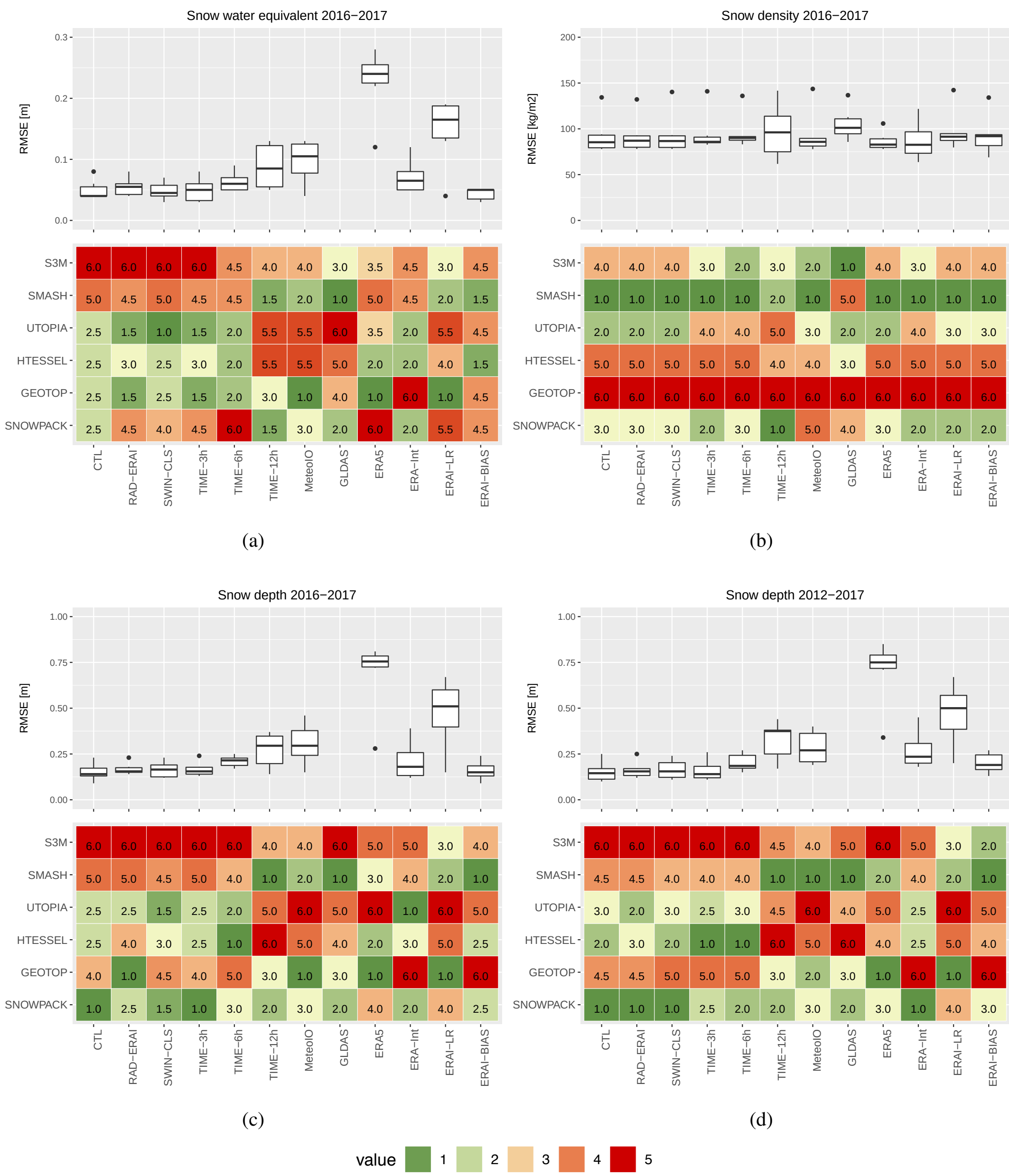

Figure 6. Root Mean Square Error associated to a) snow water equivalent, b) snow density and c) snow depth simulations for each experiment and each model over the period 2016-2017. Panel d) shows the same statistics as c) but on the whole period 2012-2017. Values for GLDAS experiments lie outside the plotted range. Upper panels represent the boxplot statistics, lower panels represent the model rank (1=best model, 5=worst model).

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Similar results are found employing SW radiation estimated as clear-sky radiation attenuated by a factor based on MSG cloud mask and neigbouring station radiation measurements (SWIN-CLS, Sect. 4). Each model shows almost identical RMSE in snow depth in the CTL, RAD-ERAI and SWIN-CLS experiments.

The use of accurate meteorological inputs but at lower temporal resolution, for instance Torgnon station data sampled at 3 hourly time step, does not affect model performances in our experiment. At longer time steps we observe a gradual decay of the snow model skills, which is strongest at the longest time step considered, corresponding to 12 hours. Therefore, at their typical temporal resolution (3 hours), climate and weather forecast model simulations, as well as reanalysis data, can be suitable for driving snowpack models, while lower temporal resolution input data require more in-depth consideration. For example, the sampling of the original 30-minute forcing at 12 hours implied a worsening of the shortwave incoming radiation input, that could be partly overcome with more sophisticated interpolation techniques than the basic linear interpolation employed in this study. In our experiment, when 12-hourly forcing is employed, the intermediate-complexity model SMASH provide comparable or lower errors than any other more sophisticated model. This suggests that with low temporal resolution forcing selected intermediate-complexity models can be employed without reducing the accuracy of the snow outputs.

When only low temporal resolution forcing is available, more sophisticated interpolation techniques compared to the basic linear interpolation employed in this study could improve the agreement with the reference data. When employing low temporal resolution forcing, similar performances are found for intermediate- and low-complexity models, probably because for the low-complexity models, the model error and the forcing error are partly compensated.

Where meteorological station data are not available, spatial interpolation of neighboring stations data or reanalyses can be a valid alternative. In our experiment the best results are obtained with ERA-Interim forcing. Indeed, despite the coarse spatial resolution, ERA-Interim satisfactorily reproduces the meteorological conditions at the Torgnon gridpoint (Fig. 3) and the model errors in terms of snow depth and snow water equivalent are only slighly higher than in the CTL experiment (Fig. 6a,c,d). SNOWPACK, HTESSEL and UTOPIA models again provide the lowest errors compared to intermediate- and low-complexity snow models (GEOTOP, SMASH, S3M). However, also the latter can be an interesting option after applying a simple adjustment of the average ERA-Interim temperature bias with respect to the Torgnon station data, and consequently adjusting also the snowfall amount. In fact, in this way the performances of the intermediate- and lower complexity snow models (GEOTOP, SMASH, S3M) can be substantially improved. The temperature adjustment based on the lapse rate (ERAI-LR), accounting for the difference in elevation between the ERA-Interim gridpoint and the real elevation of the Torgnon station, is found to worsen the model performance. In fact, this correction is blind to the local climatic features and might not be suitable in all situations. For example, in this case the lapse-rate correction is too large and it causes a temperature bias of similar amplitude but opposite sign with respect to the original ERA-Interim data. As a general remark, it is preferable to apply a temperature correction based on local temperature observations or even just climatology, when available, as the correction based on the lapse rate does not ensure a better agreement with the reference data.

Spatial interpolations of neigboring station data, such as the MeteoIO interpolation used here, can be another valid alternative in case of absence of in-situ observations. In our experiment the models RMSE values for snow water equivalent and snow depth are generally comparable to those obtained with Torgnon data at 12-hourly resolution. GLDAS, and to a lesser extent

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Discussions

ERA5, are affected by large temperature and precipitation biases at the Torgnon gridpoint, probably owing to difficulties of these datasets in simulating processes in high-elevation regions. For example, over the Alpine region, ERA5 compared to ERA-Interim is characterized by larger daily precipitation amounts during the season December-to-April, probably in relation to its finer spatial resolution and enhanced orographic processes (average over the period 1980-2014, not shown). ERA5 shows larger precipitation amounts also when compared to the observation-based dataset HISTALP as a reference over the same period (not shown). A careful evaluation of these datasets is thus recommended before using them, especially in mountain areas.

The present analysis allows to straighforwardly evaluate the performances of each model with data of gradually lower accuracy. While, as expected, with accurate forcing the most sophisticated model provides the best agreement with SWE and snow depth observations and the simplest models provide the worst (Fig 6d), more heterogeneous model responses are obtained when lower accuracy data are employed. The most sophisticated model SNOWPACK is not the best performing model throughout all experiments, even though it usually ranks among the best performing ones. The simplest snow model considered in the analysis, S3M, is not always the worst model, especially when low accuracy forcings are employed. SMASH shows an interesting behavior, with no brilliant performances with optimal forcing but outperforming many other models when using lower accuracy inputs. Indeed, SMASH ranks among the best performing models in TIME-12h, MeteoIO, ERA5, ERAI-LR and ERAI-BIAS experiments, suggesting that it can be employed in data-sparse conditions with comparable results as more sophisticated models.

The GEOTOP model provides the best snow depth estimates when forced by MeteoIO, ERA5 and ERAI-LR. However, all these forcing datasets have a cold temperature bias, and GEOTOP model is affected by a sistematic underestimation error on snow depth. These errors offset each other, with the effect that the RMSE on snow depth simulations is smallest compared to the other models. Conversely, when using ERA-Interim forcing, GEOTOP performances are the worst ones due to the positive temperature bias of the reanalysis dataset, which increases the underestimation of snow depth simulations. In this set of experiments GEOTOP model show weaknesses in reproducing both snow water equivalent and snow density, thus calling for a check of its snow scheme.

The UTOPIA and HTESSEL models perform as well as the most sophisticated model SNOWPACK with optimal forcing, but they require less input data, for example they do not need surface temperature. These models can be employed when no information on snowpack internal structure and stratification is needed. UTOPIA and CHTESSEL provide good performances also with low temporal resolution forcings up to 6 hours and with ERA-Interim forcing. However, lower skills are found when employing other low-accuracy input dataset (TIME-12h, MeteoIO), suggesting that the UTOPIA and HTESSEL models can be sensitive to the bias in the meteorological forcing.

In agreement with former studies (e.g. Essery et al., 2013) also in our analysis the best performing models have i) an explicit representation of the meltwater retention and refreezing in the snowpack and ii) an intermediate-complexity representation of the snow albedo as a function at least of surface temperature and snow age. According to our results, the representation of the snowpack as a medium with multiple layers alone does not guarantee improved results compared to models with single-layer snow schemes but taking into account meltwater infiltration and refreezing within the snowpack.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

This intercomparison exercise has been performed on a single mountain site, Torgnon, in the Western Alps, providing ideal conditions to perform the sensitivity study which we aimed to. Further analysis at other test sites would be useful to explore the extent to which our results could be generalized to different situations or models.

6 Conclusions

Relevant issues in snow modelling are the sparseness of meteorological stations providing all the variables required to drive and validate snow models, and the large uncertainties affecting the available measurements. Moreover, in mountain areas the spatial variability of the meteorological parameters is high, and the in-situ stations may be scarcely representative of the conditions in the surrounding areas.

Currently available snow models cover a wide range of complexities, from the most sophisticated schemes that resolve the internal structure of the snowpack to the simplest ones that only provide a coarse estimate of snow depth and snow water equivalent. While several studies evaluate snow models when driven by accurate meteorological data, efforts are still needed to investigate how the models perfom when forced by lower-accuracy meteorological data, as are those typically used in mountain areas.

This study evaluates snow models of different complexities assessing their sensitivity to the accuracy of the input data. An interesting result is that some of the simplest models perform equally well or even better than sophisticated models when input data are poor. For example, the intermediate-complexity model SMASH provides lower RMSE values on SWE simulations than many other higher-complexity models when driven by 12-hourly data, MeteoIO spatially interpolated data, GLDAS, ERA5, or the bias-adjusted ERA-Interim reanalysis. The lowest-complexity model considered in this study, S3M, provides comparable performances as the most sophisticated snow model analyzed here, SNOWPACK, when it is driven by bias-adjusted ERA-Interim data.

On the other hand, this study also shows that sophisticated snow models such as SNOWPACK can successfully reproduce snowpack variability across a wider spectrum of conditions compared to simpler snow models, outperforming them in case of isolated snowfall followed by rapid ablation. Sophisticated models provide good and more stable performances across different seasons. It is worth stressing that the most detailed snow model considered here, SNOWPACK, though not being the best performing model throughout all the experiments with lower accuracy forcings, ranks among the best performing models in all experiments.

Two of the intermediate-complexity snow models, HTESSEL and UTOPIA, provide comparable skills in reproducing SWE and snow depth to the most sophisticated model SNOWPACK in case of optimal forcing. In addition, they show similar skill across different seasons, thus revealing significant robustness in reproducing a variety of conditions. HTESSEL and UTOPIA can thus be considered a good trade-off between model complexity and model accuracy in case of high-quality forcing data, while they are found to be sensitive to biases in the forcing.

Some properties which are common to all models can be highlighted: i) difficulty in reproducing snow density, especially in late spring at the end of the snow season; ii) low model sensitivity to the use of surrogate radiation input data instead of the

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

measured ones, at least for the test site considered here; iii) comparable performances when driven by 3-hourly or 30-minute data, suggesting the possibility to use lower frequency data (up to 3 hours) without loosing accuracy on the snow output; iv) decrease of the models reliability, but not uniformly across the different models, when coarse-grid forcings are employed; v) substantial improvement of the models performances, reducing the differences between models of different complexity, after applying very simple bias adjustment to temperature (and consistently snowfall) forcing.

The present study exploring the relations between snow model complexity, accuracy of the forcing data and model performance has been conceived to set the basis for high-resolution modeling of mountain snow resources at the catchment and regional scales in areas where direct meteorological measurements are insufficient or unavailable and one has to rely on coarse resolution forcing. Such sensitivity experiments pave the way for the production of long-term fine-resolution reanalyses for the alpine snowpack, currently identified as a major gap for cryosphere studies (Beniston et al., 2018; Terzago et al., 2017), as well as of high-resolution future projections of the snowpack conditions. In this case snow models can be employed to downscale climate information provided by regional climate models and achieve information on snowpack characteristics at the scales required by hydrological applications, typically below 1 km. This approach, dedicated to the reconstruction of the mountain snowpack variability at fine scales is complementary to the one pursued by the ongoing ESM-SnowMIP initiative (Krinner et al., 2018) aiming at the improvement of the representation of snow processes and snow-related climate feedbacks in global climate models. Both approaches address issues which have been highlighted as important in cryospheric sciences (Beniston et al., 2018; Terzago et al., 2017) and provide information for a range of applications including the estimation of climate change impacts on the relevant socio-economic and environmental sectors.

Data availability. The data employed in this study are available upon request.

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

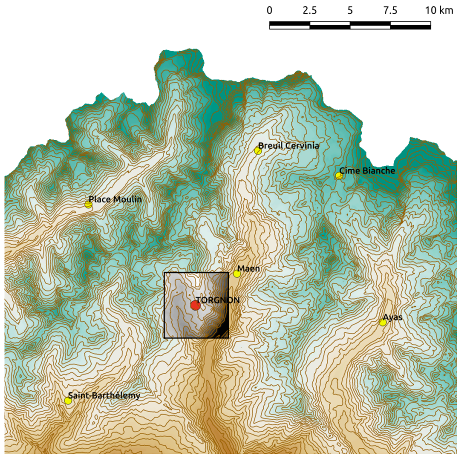

Figure A1. Location of the 6 neighboring stations used for producing the interpolated dataset for the MeteoIO experiment. The grey square represents the extension of the digital elevation model used for the interpolation.

Table A1. Characteristics of the meteorological stations used for the spatial interpolation with MeteoIO library and measured parameters: TA = air temperature; PTOT = precipitation (OTT); SWIN = incoming short wave solar radiation; VW-DW = wind speed and direction; RH = relative humidity. The stations belong to the regional meteorological network of the Aosta Valley.

Station name	Elevation [m a.s.l.]	Distance [km]	TA	PTOT	SWIN	VW-DW	RH
Cime Bianche	3100	12	x	x	x	x	x
Saint-Berthélemy	1675	9.8	x		x	x	x
Place Moulin	1980	9.1	x	x	x	x	x
Breuil Cervinia	2000	10.3	x			x	x
Maen	1310	3.2	x			x	x
Ayas	1566	11.6	x			x	x

Appendix A: Spatial interpolation of meteorological forcings from neighboring stations

In hydrological and snow modeling the spatial interpolation of ground meteorological observations is commonly employed to derive spatially continuous meteorogical forcing to drive the models. In this work, we evaluate the response of snow models with such forcing. An interpolated dataset for Torgnon monitoring site has been prepared exploiting the MeteoIO library (Bavay and Egger, 2014). The meteorological data are interpolated from six neighboring stations, over a squared digital elevation model of 16 km^2 with a grid resolution of 50 meters centered on the coordinate of Torgnon monitoring site (Fig. A1 and Tab. A1).

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences Discussions

Author contributions. ST, AP, CC, EC, SG, UMC, PP conceived the idea of the experiments. All authors partecipitated in the collection of the meteorological datasets the experiments. ST, VA, GA, LC, DD, GP, PP performed the simulations. ST analized the simulations and prepared all figures, all authors provided support in the interpretation of the results. ST wrote the paper with support from all authors.

Competing interests. The authors declare that no competing interests are present.

Disclaimer. TEXT

Acknowledgements. This work received funding from the Italian Project of Interest NextData of the Italian Ministry for Education, University and Research and from the European Union's Horizon 2020 research and innovation program under grant agreement no. 641762 (ECOPOTENTIAL). Part of this work was performed in the framework of the MEDSCOPE (MEDiterranean Services Chain based On climate PrEdictions) ERA4CS project (grant agreement no. 690462) funded by the European Union. ERA5: Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS), date of citation. https://cds.climate.copernicus.eu/cdsapp#!/home

https://doi.org/10.5194/hess-2019-511
Preprint. Discussion started: 25 November 2019

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

References

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, Journal of Hydrometeorology, 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

Bartelt, P. and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model, Cold Regions Science and Technology, 35, 123–145, 2002.

Bavay, M. and Egger, T.: MeteoIO 2.4. 2: a preprocessing library for meteorological data, Geoscientific Model Development, 7, 3135–3151, 2014.

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C.: The European mountain cryosphere: a review of its current state, trends, and future challenges, Cryosphere, 12, 759–794, 2018.

Boni, G., Castelli, F., Gabellani, S., Machiavello, G., and Rudari, R.: Assimilation of MODIS snow cover and real time snow depth point data in a snow dynamic model, in: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, pp. 1788–1791, IEEE, 2010.

Boone, A. and Etchevers, P.: An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an Alpine site, Journal of Hydrometeorology, 2, 374–394, 2001.

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., et al.: The Rhone-aggregation land surface scheme intercomparison project: An overview, Journal of Climate, 17, 187–208, 2004.

Bowling, L. C., Lettenmaier, D. P., Nijssen, B., Graham, L., Clark, D. B., El Maayar, M., Essery, R., Goers, S., Gusev, Y. M., Habets, F., et al.: Simulation of high-latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2 (e): 1: Experiment description and summary intercomparisons, Global and Planetary Change, 38, 1–30, 2003.

Cassardo, C.: UTOPIA: The Manual of Version 2015, Università di Torino, https://doi.org/10.13140/RG.2.2.29664.38404, https://www.researchgate.net/publication/323200198_UTOPIA_The_Manual_of_Version_2015, 2015.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resources Research, 47, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, http://dx.doi.org/10.1002/qj.828, 2011.

Di Mauro, B., Garzonio, R., Rossini, M., Filippa, G., Pogliotti, P., Galvagno, M., Morra di Cella, U., Migliavacca, M., Baccolo, G., Clemenza, M., Delmonte, B., Maggi, V., Dumont, M., Tuzet, F., Lafaysse, M., Morin, S., Cremonese, E., and Colombo, R.: Saharan dust events in the European Alps: role on snowmelt and geochemical characterization, The Cryosphere Discussions, 2018, 1–28, https://doi.org/10.5194/tc-2018-241, https://www.the-cryosphere-discuss.net/tc-2018-241/, 2018.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M., Beljaars, A., Schär, C., and Elder, K.: An improved snow scheme for the ECMWF land surface model: description and offline validation, J. Hydrometeor., 11, 899–916, https://doi.org/10.1175/2010jhm1249.1, 2010.

Dutra, E., Viterbo, P., Miranda, P. M., and Balsamo, G.: Complexity of snow schemes in a climate model and its impact on surface energy and hydrology, Journal of Hydrometeorology, 13, 521–538, 2012.

Endrizzi, S., Gruber, S., Dall'Amico, M., and Rigon, R.: GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects, Geoscientific Model Development, 7, 2831–2857, 2014.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, Advances in Water Resources, 55, 131 – 148, https://doi.org/http://dx.doi.org/10.1016/j.advwatres.2012.07.013, http://www.sciencedirect.com/science/article/pii/S0309170812002011, snow–Atmosphere Interactions and Hydrological Consequences, 2013.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y., Essery, R., Fernandez, A., et al.: SnowMIP, an intercomparison of snow models: first results, in: Proceedings of the International Snow Science Workshop, Penticton, Canada, vol. 29, 2002.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fernandez, A., et al.: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), Annals of Glaciology, 38, 150–158, 2004.

Filippa, G., Cremonese, E., Galvagno, M., Migliavacca, M., Di Cella, U. M., Petey, M., and Siniscalco, C.: Five years of phenological monitoring in a mountain grassland: inter-annual patterns and evaluation of the sampling protocol, International journal of biometeorology, 59, 1927–1937, 2015.

Galvagno, M., Wohlfahrt, G., Cremonese, E., Rossini, M., Colombo, R., Filippa, G., Julitta, T., Manca, G., Siniscalco, C., Morra di Cella, U., et al.: Phenology and carbon dioxide source/sink strength of a subalpine grassland in response to an exceptionally short snow season, Environmental Research Letters, 8, 025 008, 2013.

Hersbach, H. and Dee, D.: ERA5 reanalysis is in production, ECMWF Newsletter 147, ECMWF, 2016.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R., et al.: The quantification and correction of wind-induced precipitation measurement errors, Hydrology and Earth System Sciences, 21, 1973–1989, 2017.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H., and Zhou, W.: ESM-SnowMIP: Assessing models and quantifying snow-related climate feedbacks, Geoscientific Model Development Discussions, 2018, 1–32, https://doi.org/10.5194/gmd-2018-153, https://www.geosci-model-dev-discuss.net/gmd-2018-153/, 2018.

Kumar, M., Marks, D., Dozier, J., Reba, M., and Winstral, A.: Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models, Advances in Water Resources, 56, 77–89, 2013.

Lawrence, M. G.: The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications, Bulletin of the American Meteorological Society, 86, 225–233, https://doi.org/10.1175/bams-86-2-225, http://dx.doi.org/10.1175/bams-86-2-225, 2005.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resources Research, 51, 2707–2723, 2015.

Piazzi, G., Thirel, G., Campo, L., and Gabellani, S.: A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment, The Cryosphere, 12, 2287–2306, 2018.

Piazzi, G., Campo, L., Gabellani, S., Castelli, F., Cremonese, E., di Cella, U. M., Stevenin, H., and Ratto, S. M.: An EnKF-based scheme for snow multivariable data assimilation at an Alpine site, Journal of Hydrology and Hydromechanics, 67, 4–19, 2019.

Raleigh, M., Lundquist, J., and Clark, M.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrology and Earth System Sciences, 19, 3153–3179, 2015.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., et al.: The global land data assimilation system, Bulletin of the American Meteorological Society, 85, 381–394, 2004.

Rui, H. and Beaudoing, H.: README Document for NASA GLDAS Version 2 Data Products, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, USA, 2018.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellström, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W.-P., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T. G., Stähli, M., Stöckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y., and Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), Journal of Geophysical Research: Atmospheres, 114, n/a–n/a, https://doi.org/10.1029/2008JD011063, http://dx.doi.org/10.1029/2008JD011063, 2009.

Schlosser, C. A., Slater, A. G., Robock, A., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., Speranskaya, N. A., and Mitchell, K.: Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS Phase 2 (d), Monthly Weather Review, 128, 301–321, 2000.

Sims, E. M. and Liu, G.: A parameterization of the probability of snow–rain transition, Journal of Hydrometeorology, 16, 1466–1477, 2015.

Slater, A. G., Schlosser, C. A., Desborough, C., Pitman, A., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Entin, J., Mitchell, K., Chen, F., et al.: The representation of snow in land surface schemes: Results from PILPS 2 (d), Journal of Hydrometeorology, 2, 7–25, 2001.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, 2001.

Terzago, S., von Hardenberg, J., Palazzi, E., and Provenzale, A.: Snow water equivalent in the Alps as seen by gridded data sets, CMIP5 and CORDEX climate models, The Cryosphere, 11, 1625–1645, https://doi.org/10.5194/tc-11-1625-2017, https://www.the-cryosphere.net/11/1625/2017/, 2017.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7. 2, Geoscientific Model Development, 5, 773–791, 2012.

Yang, K., Huang, G., and Tamai, N.: A hybrid model for estimating global solar radiation, Solar energy, 70, 13–22, 2001.

Yang, K., Koike, T., and Ye, B.: Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets, Agricultural and Forest Meteorology, 137, 43–55, 2006.