Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas*

STEFANIE JÖRG-HESS

Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

NENA GRIESSINGER

WSL Institute for Snow and Avalanche Research (SLF), Davos, Switzerland

MASSIMILIANO ZAPPA

Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

(Manuscript received 9 October 2014, in final form 4 June 2015)

ABSTRACT

Good initial states can improve the skill of hydrological ensemble predictions. In mountainous regions such as Switzerland, snow is an important component of the hydrological system. Including estimates of snow cover in hydrological models is of great significance for the prediction of both flood and streamflow drought events. In this study, gridded snow water equivalent (SWE) maps, derived from daily snow depth measurements, are used within the gridded version of the conceptual hydrological model Precipitation Runoff Evapotranspiration Hydrotope (PREVAH) to replace the model SWE at initialization. The ECMWF Ensemble Prediction System (ENS) reforecast is used as meteorological input for 32-day forecasts of streamflow and SWE. Experiments were performed in several parts of the Alpine Rhine and the Thur River. Predictions where modeled SWE estimates were replaced with SWE maps could successfully enhance the predictability of SWE up to a lead time of 25 days, especially at the beginning and the end of the snow season. Additionally, the prediction of the runoff volume was improved, particularly in catchments where the snow accumulation, and thus the runoff volume, had been greatly overestimated. These improvements in predictions have been made without affecting the ability of the forecast system to discriminate between the different runoff volumes observed. A spatial similarity score was first used in the context of SWE forecast verification. This confirmed the findings of the time series analysis and yielded additional insight on regional patterns of extended range SWE predictability.

1. Introduction

Probabilistic forecasts have become more popular in hydrology (Dale et al. 2014; Ramos et al. 2013; Fundel and Zappa 2011). Over the last few years the value of ensembles for runoff forecasts has been shown primarily in studies that have focused on flood forecasting (Marty

et al. 2013; Addor et al. 2011; Cloke and Pappenberger 2009; Zappa et al. 2008; Verbunt et al. 2007; Pappenberger et al. 2005). Recently, droughts and low flows have been studied using hydrological ensemble forecasts (Demirel et al. 2013; Fundel et al. 2013). Their main advantage is that the uncertainty of the meteorological input can then be represented (Verbunt et al. 2007; Pappenberger et al. 2005). This is the main source of uncertainty in flood predictions (Zappa et al. 2011), while parameter uncertainty is the main source of uncertainty in low-flow forecasts (Demirel et al. 2013).

Ensemble predictions for a Swiss catchment benefited from good initial states for a lead time of up to 32 days in the low-flow regime (Fundel et al. 2013). Possible approaches to improve the initial states could be the pre-processing of the meteorological variables (Fundel et al.

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JHM-D-14-0193.s1>.

Corresponding author address: Stefanie Jörg-Hess, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Zürcherstr. 111, CH-8903 Birmensdorf, Switzerland.
E-mail: s.joerg.hess@gmail.com

2010) or the assimilation of runoff, snow cover, and soil moisture. The use of preprocessed meteorological variables can also improve runoff predictions (Demargne et al. 2014; Shukla et al. 2012). It is often sufficient to preprocess precipitation and take the other meteorological input variables from “raw” numerical weather prediction (NWP) models (Jörg-Hess et al. 2015).

In France runoff has been successfully assimilated for hydrological predictions in several studies (Thirel et al. 2010a,b; Aubert et al. 2003). In Switzerland, however, many catchments are influenced by human activities such as hydropower systems, lake regulation, and environmental flows (Tharme 2003), which means this procedure is less likely to produce enhanced runoff predictions.

Several studies have found that including remotely sensed or observed soil moisture data (Brocca et al. 2010; Aubert et al. 2003) successfully improves runoff predictions. The main difficulty with assimilating soil moisture is that satellite-based soil moisture information has a coarse resolution and only represents the soil moisture of the top soil. Moreover, long time series of in situ soil moisture measurements are rarely available (Seneviratne et al. 2012).

Snow-cover information has usually been assimilated into hydrological models from satellite observations or point measurements of the snow cover (Thirel et al. 2013; He et al. 2012; Moradkhani 2008; Su et al. 2008). This improves the streamflow near the end of the snowmelt season (Clark et al. 2006). Satellite snow-cover information is available for the snow-covered area (SCA) or, alternatively, the snow water equivalent (SWE). SCA provides the spatial information snow/no snow, which is insensitive to changes in the snow amount (Magnusson et al. 2014). For remotely sensed SWE data, which provide much more information than the binary SCA product, the horizontal resolution of about 25 km is too coarse for hydrological modeling in mountainous terrain (Dietz et al. 2012). Point measurements of the snow cover, on the other hand, often require spatial interpolation prior to data assimilation to be useful for hydrological models (Slater and Clark 2006).

In Switzerland, gridded SWE maps have been available for several years with a horizontal resolution of $1 \text{ km} \times 1 \text{ km}$ (Jörg-Hess et al. 2014; Magnusson et al. 2014). During the snow season (November–May), SWE maps are produced weekly, based on measured snow depth and estimated snow bulk densities (Jonas et al. 2009). A horizontal resolution of 1 km is reached by interpolating the SWE values with an inverse distance weighting approach based on two Gaussian filters and a nonlinear trend of SWE with elevation. The temporal dynamics in the SWE maps agrees with the absolute daily SCA and provides more useful information for

snow mass balance than the SCA (Hüsler et al. 2014). These maps provide an alternative way of representing snow-cover information from satellite observations and point measurements for hydrological modeling.

In this study, we assessed the quality of monthly forecasts of ensemble runoff and SWE for different domains and seasons with and without replacing the SWE maps of the hydrological model Precipitation Runoff Evapotranspiration Hydrotope (PREVAH; Viviroli et al. 2009a,b,c; Zappa et al. 2003) with SWE maps estimated from observations. Snow greatly influences the hydrological cycle, especially in mountainous regions such as Switzerland, as it acts as a temporary buffer for precipitation. In Switzerland, for example, the percentage of the solid precipitation is about 30%, and snowmelt contributes to about 40% to the Swiss runoff (Blanc and Schaedler 2014). Therefore, indicating accurate information on the initial state of the snow storage is expected to enhance runoff predictions and especially SWE predictions (e.g., Koster et al. 2010; Slater and Clark 2006).

In this study we introduce several new aspects into the used hydrological model, including the simultaneous verification of ensemble runoff and SWE predictions, and we investigate the benefit of using good initial snow storage data for both these variables. Daily predictions of runoff and SWE up to a lead time of one month were carried out using meteorological observations and the ENS ensemble reforecast (Vitart et al. 2008) from the European Centre for Medium-Range Weather Forecasts (ECMWF) as forcing for the hydrological model PREVAH (Viviroli et al. 2009c). Daily SWE maps ($1 \text{ km} \times 1 \text{ km}$) were imported into PREVAH at initialization. This resulted in a higher temporal and spatial resolution than in many other studies that inserted SWE to hydrological models (e.g., Andreadis and Lettenmaier 2006; Slater and Clark 2006; Su et al. 2008). The daily mean catchment runoffs and gridded SWE forecasts with a 32-day lead time were produced from 1991 to 2008. We also evaluated the SWE predictions by measuring map similarities. The similarity measure “mapcurves” (Hargrove et al. 2006) quantifies the similarity between the SWE maps and the forecasted SWE fields by measuring the spatial overlap of the two maps.

Potential users who may be interested in enhanced SWE predictions for large areas with a lead time of 32 days include stakeholders from the hydropower sector; ski resorts that need to plan artificial snow production; and the government, which is responsible for flood and drought predictions (Zappa et al. 2014; Kruse and Seidl 2013).

In the following sections we describe the procedure that was used to import SWE maps into the hydrological model PREVAH for several alpine regions in

FIG. 1. Study domains with gauging stations (points 1–11) with the inset showing the altitudinal distribution (<1000, 1000–1500, 1500–2000, 2000–2500, and >2500 m MSL) of each domain. SWE maps are not available in Bod200.

Switzerland. [Section 2a](#) describes the study areas, and [section 2b](#) describes the hydrological model PREVAH and the experimental setup. [Section 3](#) presents the data and the experimental setup. [Section 4](#) introduces the performance measures for the different simulations, and [section 5](#) describes the sensitivity analysis for the runoff and SWE predictions, with the general conclusions and summary in [section 6](#).

2. Study domains and hydrological modeling

a. Study domains

The spatial information of the study domains has a horizontal resolution of $200\text{ m} \times 200\text{ m}$. The study domains are the small pre-alpine catchment of the Thur River (Thu200) located in the northeastern part of Switzerland and the Alpine Rhine that covers the region from the Rhine sources to Lake Constance ([Fig. 1](#)). The Alpine Rhine is divided into the subregions (Table 1) Lake Constance (Bod200) and the mountainous domains Hinter Rhine (HiR200), Vorder Rhine (VoR200), Vorarlberg (VoA200), and Landquart Plessur (LaP200). The catchment Thu200 is not covered by glaciers. The mountainous catchments of the Alpine Rhine have a glacialization between 3.8% (VoR200) and 1.4% (LaP200). At the river gauge in Neuhausen downstream from Lake Constance, the glacialization of the catchment is only 0.7%.

The runoff for these domains within and across regions was computed with a simplified routing to roughly take into consideration hydropower systems (including artificial dams, intakes, diversion, and rough estimates of hydropeaking activity) and major lakes. Water is diverted where artificial dams and intakes are present. Larger regulated lakes are defined as linear storages and the inflow to dammed lakes as hydrograph curves. Translation of flood waves from an upstream basin to a downstream gauge was defined by analyzing the travel time of observed peak runoff events. This procedure of routing and water redistribution allows for storage of meltwater in spring, modeling hydropeaking, damping floods, and capturing the main influence of large lakes.

Runoff gauging stations in the domains were selected from available stations to assign altitudinal gradients and the propagation of snow accumulation and melt with runoff. From PREVAH reference simulations, we calculated that in the highest domains of the Alpine Rhine more than 60% of the runoff originates from snowmelt. In the lower catchments Thu200 and Bod200, snowmelt contributes about 20% to runoff ([Zappa et al. 2012](#)).

The different seasonal cycles of runoff and snow storage are illustrated for two catchments ([Fig. 2](#)). The Thur is a rain-fed catchment where the runoff shows little seasonality. Snow plays a role in the annual cycle

TABLE 1. Size, mean elevation, and the contribution of snowmelt to runoff in the study domains. Abbreviations (Abbr.) used in the figures are listed. Some of the river gauges evaluated are fully embedded in one region (boldface), and others are the results of runoff routing from region to region (italic).

Domain	Abbr.	Size (km ²)	Elevation (m MSL)	Snowmelt to runoff (%)	River gauges included
Thur	Thu200	1725	762	26	Thur, Andelfingen (11) Halden (10)
Lake Constance	Bod200	5364	644	20	Rhine, <i>Neuhausen</i> (9)
Hinter Rhine	HiR200	1639	2058	63	Hinter Rhine (3) Landwasser (1) Albula (2)
Vorder Rhine	VoR200	1517	1955	62	Vorder Rhine (4)
Vorarlberg	VoA200	1948	1354	50	Ill (7) Rhine, <i>Diepoldsau</i> (8)
Landquart Plessur	LaP200	1238	1672	62	Landquart (6) Rhine, <i>Domat-Ems</i> (5)

but its contribution to runoff is small. The Landquart River, in contrast, is mostly governed by snow accumulation and melt. The average runoff hydrograph shows a strong seasonal cycle with the highest runoff volume in June.

Hereafter, the abbreviations presented in the second column of Table 1 are used when results refer to the study domains declared in Fig. 1. The results concerning the river gauges within the study domains (last column in Table 1) are declared either with the name of the station or with the label of the point presented in Fig. 1.

b. The hydrological model PREVAH

Runoff predictions were generated using the hydrological model PREVAH (Viviroli et al. 2009c) forced with the ECMWF Ensemble Prediction System (ENS) or meteorological observations. To run PREVAH, spatial information (i.e., land use, aspect, and elevation) and gridded meteorological variables are required. The meteorological input variables are precipitation, 2-m temperature, 2-m dewpoint, sunshine duration, surface albedo, and solar radiation. The use of a fully distributed version (200 m × 200 m) of this conceptual model allowed the redistribution of SWE to lower altitudes (Gruber 2007). Major conceptual differences between the hydrological response units (HRUs) and the fully distributed version of PREVAH are as follows:

- 1) The representation of local precipitation events in the gridded version is geographically better.
- 2) Redistribution of snow in steep regions according to the algorithms presented in Gruber (2007) is considered. This feature removes the accumulated masses of eternal snow in the highest and steepest peaks of the catchment area that are included in simulations with PREVAH and other hydrological models (e.g., Zappa and Kan 2007; Verbunt et al. 2003).
- 3) The HRU version of PREVAH can be forced with hourly meteorological input and should be preferred

in real-time applications that focus on floods (Liechti et al. 2013; Addor et al. 2011).

- 4) The gridded version of PREVAH is intended for low-flow and water resource assessments studies and has so far only been forced with daily meteorological fields. It allows for the transient assimilation of land-cover scenarios (Kobierska et al. 2013; Schattan et al. 2013).

In PREVAH, the snow module is temperature-index based (Zappa et al. 2003). This allows snow accumulation to be modeled with few input variables and few free parameters. The degree-day approach as introduced by Hock (1999) is used for modeling snowmelt. This approach uses a variable degree-day factor that has a seasonal cycle between minimum and maximum temperatures (Viviroli et al. 2009c). Similar to the snow accumulation and snowmelt modules, the formulation of the other main hydrological process implemented in PREVAH is conceptual (Viviroli et al. 2009c). The runoff generation module is a tailored adaptation of the equations first used in the Hydrologiska Byråns Vattenavdelning (HBV) model (e.g., Lindström et al. 1997). Differences are in the additional consideration of a slow- and fast-responding groundwater component (Gurtz et al. 2003). The storage coefficients for runoff generation are catchment-specific parameters that need to be estimated by calibration. Evapotranspiration is calculated following the Penman–Monteith approach, as presented in Gurtz et al. (1999). Glacial melt is computed with the same equations used for snowmelt and separate tunable parameters (Hock 1999; Zappa et al. 2003). All water fluxes are computed separately for each raster element. The only lateral communication between cells is represented by the snow redistribution routine (Gruber 2007).

The baseline runoff simulations are divided into a calibration period (1984–96) and a verification period (1980–83 and 1997–2009). The regionalized calibration

FIG. 2. (top) Seasonal cycle of the median (black line) and interquartile (gray shading; 25–75th percentile) daily runoff and (bottom) SWE for the (a),(b) Thur and (c),(d) Landquart catchments for the period 1991–2008. Note that the SWE axis of the Landquart region is 10 times larger than that for the Thur region. For the observed runoff, a moving average of 5 days is illustrated. SWE is displayed with a moving average of 3 days.

parameters were taken from a project that focused on climate change impacts on water resources (Bernhard and Zappa 2012; Speich et al. 2015) and are based on the work of Köplin et al. (2010). Further details about the PREVAH input data, physics, and parameterization are given in Viroli et al. (2009c).

3. Data and experimental setup

a. Meteorological forcing

For the reference simulations of the runoff hydrograph (P.href), data from weather stations (Begert et al. 2005) provided by the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) were used. The interpolation and downscaling of these variables for PREVAH is described in Jörg-Hess et al. (2015), Schattan et al. (2013), and Viroli et al. (2009c). Data from 1980 to 2009 have been processed for calibrating and validating the hydrological model and to generate the initial conditions for the experiments with numerical forecasts.

The meteorological forecast data have been obtained from the unified variable resolution ENS (Virt

et al. 2008) operated by the ECMWF. Each week a 51-member forecast with a lead time of 32 days is provided. The horizontal resolution is 32 km for the first 10 days and then changes to 64 km for days 11–32. After day 10 the atmospheric model is coupled to an oceanic model. With each operational ENS forecast, a 5-member reforecast of weekly 32-day forecasts over the past 18 years is started with the model version from March 2011. The global ERA-Interim provides the boundary conditions to force the five ensemble members. For this study, the reforecast over the period 1991–2008, as used in Fundel et al. (2013) for the Thur catchment, is adopted. The downscaling of the meteorological variables to $200\text{ m} \times 200\text{ m}$ and the adjustment of precipitation is completed as presented in Fundel et al. (2013) and Addor et al. (2011).

b. Snow water equivalent

In recent decades the density of monitoring stations in Switzerland has increased and the accuracy of mapped data has consequently improved. To produce a set of daily SWE maps with constant high accuracy for the winter months 1991–2008, we used two

climatologies of SWE maps based on different station densities and on two different methods of SWE mapping [see sections 3b(1) and 3b(2)]. The quantile mapping statistical calibration method was applied to refine the set of maps, which are based on fewer stations but cover a longer time span (Jörg-Hess et al. 2014). The months November–May are defined as winter months.

1) SWE MAPPING METHOD FOR CLIMATOLOGY 1971–2012

Here the SWE climatology is based on daily snow depth measurements (HS, hereafter) from 106 stations in Switzerland. Elevations between 200 and 2500 m MSL are covered, but most stations are situated below 2000 m MSL. We applied a mapping model that first calculates the daily nonlinear trend of SWE over elevation. The offset of each station's SWE from the trend is calculated and these detrended SWE values are then interpolated over a digital elevation model with a resolution of $1 \text{ km} \times 1 \text{ km}$. For spatial interpolation, we used a three-dimensional distance-weighted Gaussian filter. This method is described in detail in Jörg-Hess et al. (2014). The method of detrending was optimized and validated for the adopted station density.

2) SWE MAPPING METHOD FOR CLIMATOLOGY 2001–12

These SWE values are based on daily HS measurements at 344 stations, which are part of the different snow monitoring networks of MeteoSwiss, the Institute for Snow and Avalanche Research (SLF), and Land Vorarlberg (Austria). The stations are evenly spread across Switzerland and parts of Tyrol and Vorarlberg, covering elevation bands between 200 and 2950 m MSL, although the station density decreases above 2200 m MSL. Accumulation and ablation were first calculated separately and each interpolated over a digital elevation model with a resolution of $1 \text{ km} \times 1 \text{ km}$. Daily SWE grids were then calculated on the basis of the ablation and accumulation grids. Jörg-Hess et al. (2014) validated these maps and showed that concerning correlation best results are obtained in the high-elevation areas, while the bias against point observations is better for low-altitude areas.

3) SWE HOMOGENIZATION WITH QUANTILE MAPPING

The statistical approach called quantile mapping was applied to transfer the spatial pattern of the shorter, more accurate SWE map climatology (based on 344 stations) to the long-term SWE map climatology (based

on 106 stations). The method was applied to each grid cell for each day of the winter seasons 1971–2012. The exceedance probability of a specific SWE value in one climatology corresponds to the quantile of the same exceedance probability in the other climatology. We calculated the empirical cumulative distribution functions for a time window, for which SWE maps of both climatologies exist (here the winters 2001–12). We then calibrated all SWE values of the long-term climatology by using these distribution functions. For example, in a given pixel, for a date before 2001 (e.g., 1 March 1992), an SWE of 100 mm has been interpolated. This represents a nonexceedance probability of 80% over the period 2001–12. The same nonexceedance probability in the “short but more accurate” climatology applies to SWE of 115 mm. So we replace the 100 mm in the long climatology with the 115 mm. The method is described in detail in Jörg-Hess et al. (2014). Finally, we used the calibrated SWE maps for the years 1991–2000 and the SWE maps based on 344 stations for the years 2001–08 to replace the initial SWE conditions of the hydrological model in some of the completed experiments (see later in the manuscript). SWE maps for the Bod200 domain could not be processed because of missing SWE and snow depth information in the German areas of this study domain. However, it is interesting to see the propagation on the imported SWE maps of the upstream catchments on the runoff at the station located at the outlet of Lake Constance, most of which comes from the Bod200 domain (Fig. 1).

c. Experimental setup

The meteorological variables and SWE maps are scaled to a resolution of 200 m to meet the resolution of the spatial information used to configure PREVAH. The scaling to the 200-m resolution of the hydrological model occurs by adopting bilinear interpolation. In the case of air temperature, the downscaling procedure also considers a constant temperature lapse rate of $-0.7 \text{ K (100 m)}^{-1}$. This allows the consideration of the finer topography used by PREVAH with respect to the coarse topography available from ENS. PREVAH's SWE values can easily be replaced by SWE maps at the initialization of a monthly model run. We evaluated the impact of such replacement on the quality of PREVAH discharge and SWE simulations. The hydrometeorological forecast chain that consists of forcing PREVAH with the five members' ENS reforecasts used for this experiment is similar to what was presented in Fundel et al. (2013). The application was expanded to include not only the Thur but also the Alpine Rhine basin with its main tributaries (Fig. 1). Unlike Fundel et al. (2013), who used the HRU version of PREVAH (Gurtz et al.

FIG. 3. Setup for the simulations of runoff and SWE. Predictions with the hydrological model PREVAH are forced with ENS or with observations. Four sets of simulations have been calculated: (i) the hydrological reference simulation (P.href), (ii) the reference simulation with updated SWE (SWE.href) and the two correspondent sets of reforecast, (iii) P.ens uses initial conditions from P.href, and (iv) SWE.ens uses initial conditions from SWE.href. The colored regions are the SWE maps and the dark gray lines are the observations, the red circles are the initialization of the 30 days of model runs, and the dashed arrows are the Variable Resolution Ensemble Prediction System (VarEPS) members.

1999), we used a gridded version of PREVAH (see section 2b). See the supplemental material for a comparison of the predictions of the HRU with those of the gridded PREVAH version for the Thur basin. The initial states of the first forecast on 1 January 1991 are taken from reference simulations, which are run with meteorological observations without SWE replacement. We call this hydrological reference run “P.href.” In the chain, each week during the period 1991–2008, a PREVAH forecast with a lead time of 32 days was started. Four sets of simulations were performed (Fig. 3):

- 1) A deterministic hydrological reference run forced by interpolated precipitation (P.href) yielding a simulated discharge and SWE time series.
- 2) A second hydrological reference run (SWE.href) forced by the same weather forcing, and using the same calibration, where SWE is replaced with SWE maps elaborated from observed data (see section 3b) at initialization of a 32-day deterministic run. All other internal states of the model are modeled as in the case of P.href. In SWE.href the weekly SWE replacements are propagated. This guarantees that the nonlinear effects of replacing SWE on further hydrological processes considered (most prominently infiltration, snow age, albedo, and runoff generation) are also propagated during all the 18 years of the experiment.

- 3) A reforecast in quasi-operational mode producing five discharge and SWE projections every week with a lead time of 32 days using five-member ENS forcing (P.ens). These projections use initial conditions from P.href and can be compared to the reference simulation of the first step.
- 4) A set similar to P.ens, but in addition to this, the SWE is also updated weekly from the available SWE maps data before launching the forecasts. This last set is labeled as SWE.ens.

According to the outcomes of these four sets of simulations, we evaluated the effect of imported SWE at initialization for runoff and SWE predictions. The effect of the numerical weather prediction model ENS (P.ens and SWE.ens) was compared to the reference simulations forced with meteorological observations (P.href and SWE.href).

4. Performance measures

The runoff and SWE predictions were evaluated with the verification scores for different runoff quantiles, lead times, and months of the year. For a better overview of the evaluation with lead time, the scores are illustrated in 5-day steps for the lead times 1, 6, 11, 16, 21, 26, and 31 days. Fundel et al. (2013) investigated many different

scores but the general conclusions were similar for all of them. Therefore, for this study we focused on one score, the two alternatives forced choice score (2AFC; see section 4b).

a. Nash–Sutcliffe coefficient, benchmark efficiency, and volumetric deviation

The verification evaluated the model performance with the commonly used hydrological score Nash–Sutcliffe coefficient (Nash and Sutcliffe 1970) for normal (NSE) and logarithmic runoff (logNSE) and two benchmark approaches (Schaeffli and Gupta 2007). We evaluated the NSE for comparability with other studies and the logNSE because the focus is on low flow (Krause et al. 2005).

To account for the strong seasonality in most of the catchments investigated, we also used the benchmark efficiency (BE), which employs a seasonally varying benchmark instead of the mean runoff used as the benchmark in the NSE. We calculated the seasonally varying benchmark as the interannual mean runoff for each calendar day with a window of 30 days around this day.

The water balance of the simulations was evaluated with the volumetric deviation (VOL%; Zappa and Kan 2007). A VOL% close to 0 indicates that the runoff volume of the observations and the simulations correspond well (Viviroli et al. 2009c). VOL% refers to the duration of the declared calibration and validation periods. In case of the evaluation of the experiments with ENS, VOL% is assessed for partial time series with identical lead time.

b. Two alternatives forced choice score

The 2AFC is a score that evaluates the ability of the model to distinguish between different observations (Fundel et al. 2013; Weigel and Mason 2011; Mason and Weigel 2009). The score is very flexible and can be used on a range of forecasts from simple yes/no forecasts of dichotomous outcomes to continuous forecasts for deterministic and probabilistic forecasts. The score is based on comparing two observation–forecast pairs and then evaluating whether the forecast can be used to successfully distinguish between the observations. In the supplemental material the idea of 2AFC is illustrated.

For dichotomous observations with probabilistic forecasts, this means that the probabilities define the number of categories. In our example with five ensemble members, six probability bins are used. Let n_0 be the number of nonevents and n_1 the number of events that occurred. The j th forecast probability of an event when an event occurred is $p_{1,j}$, and $p_{0,i}$ is the i th forecast

probability when no event occurred. The 2AFC score for dichotomous observations with a probabilistic forecast is defined as

$$2\text{AFC} = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I(p_{0,i}, p_{1,j}), \quad (1)$$

where $I(p_{0,i}, p_{1,j})$ is defined as

$$I(p_{0,i}, p_{1,j}) = \begin{cases} 0.0 & \text{if } p_{1,j} < p_{0,i} \\ 0.5 & \text{if } p_{1,j} = p_{0,i} \\ 1.0 & \text{if } p_{1,j} > p_{0,i} \end{cases} \quad (2)$$

Continuous observations with probabilistic forecasts are treated similarly, except that there are more than two possible outcomes of the observations. The number of observed categories are the n observations. The 2AFC score for continuous observations with a probabilistic forecast is defined as

$$2\text{AFC} = \frac{2}{n(n-1)} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I(p_{0,i}, p_{1,j}). \quad (3)$$

Usually, the added precision of continuous observations, rather than dichotomous observations, decreases the score of the forecast.

The primary weakness of the 2AFC score is that it illustrates the potential of a forecast system after proper calibration. Systematic errors of the forecast system are not considered and therefore the score gives no indication if the forecasts can be taken at face value. The 2AFC score is a measure of how often a forecast is correct. Thus, a 2AFC of 50% means that the forecast is correct half of the time and a simple guess based on climatology would give a similar result. A perfect forecast has a 2AFC of 100% and skillful forecasts have a 2AFC larger than 50%. As the 2AFC score measures the discriminative power of a forecast system, it is also dependent on the ensemble size. The 5-member ENS reforecast we used has a larger sampling uncertainty than the full 51-member ENS forecast that makes it more difficult for the system to discriminate. The skill of this 5-member system is considered as the lower bound of the effective skill (Weigel and Mason 2011; Mason and Weigel 2009).

c. The similarity measure mapcurves

To quantify the similarity between SWE maps and forecasted SWE fields, the similarity measure mapcurves (Hargrove et al. 2006), which measures the spatial overlap of two maps, is used. We evaluated the SWE forecasts with other similarity measures (Goodman and Kruskal's λ , Theil's uncertainty coefficient U , and

Cramér's V ; see Rees 2008) and found they all produced similar results. This is similar to what Speich et al. (2015) evaluated when using these criteria in the context of evaluating spatial and temporal patterns of climate change impacts on water resources. We therefore decided to illustrate the validation only with mapcurves here.

Mapcurves is a goodness-of-fit (GOF) method that shows the degree of similarity between categorical maps. For each pair of classes (class A and class B), one from each of the two maps being compared, the algorithm calculates a GOF measure. This measure is determined by the proportion of class A that overlaps with class B, weighted by the proportion of class B that overlaps with class A. This GOF measure ranges from zero, meaning that the categories do not overlap, to one, in the case of a perfect fit. The scores are then summed up for each class of map A and displayed as a power curve. The x axis is divided into equal intervals between zero and 100%, and the y axis indicates the percentage of GOF scores equal to or more than the percentage shown on the x axis. The area under this curve, lying between zero and one, is the measure of similarity between the two maps. For the full formulation of mapcurves, please refer to Hargrove et al. (2006). A compact description and related application in hydrological context is presented in Speich et al. (2015).

The value of mapcurves depends greatly on the classification of the compared values and can therefore not be interpreted on their own (Hargrove et al. 2006). Conclusions about the similarity measure should be drawn in the sense that simulated SWE and SWE maps are more similar for lead time 1 than for lead time 6. SWE was classified using 67 categories, with small intervals for smaller SWE values and larger intervals for larger SWE values. We found, on comparing SWE maps with predicted SWE, that the changes in the mapcurves' similarity with ongoing lead time is best represented if 50 categories are used. Including more categories led to no extra support in interpreting the results for this example.

5. Results

In this section, the forecast quality for discharge in different flow regimes and SWE were evaluated on a daily basis for several domains in the Alpine Rhine region and the Thur region.

First, the reference predictions with the hydrological model PREVAH (P.href) were evaluated for the calibration period and the verification period (Fig. 4). The NSE was approximately 0.8 at most stations and ranges

FIG. 4. Verification scores for the PREVAH discharge simulations P.href for the gauging stations in the Thur and the Alpine Rhine catchment. (from left to right) The gauging stations are ordered according to their catchment size (area) with gray being the largest; NSE for normal (NSE) and log of the runoff (logNSE), and the BE and VOL%; except for the area, values are shown for the calibration period (labeled c; 1984–96) and the verification period (labeled v; 1980–83 and 1997–2008). The gauging stations with asterisks are uninfluenced catchments.

from 0.56 to 0.96. The NSE for logarithmic runoff was slightly lower. We had some simulation problems in the Hinter and Vorder Rhine, probably because of overly rough assumptions about the operation of the hydropower stations. Not only do hydropower activities cause some simulation problems, but the BE indicated that small catchments with a strong seasonal runoff cycle behave similarly (Albula Tiefenkaastel and Landwasser Davos). Problems with small alpine catchments may be due to the fact that snowmelt in the mountains is a dynamic process that requires a more physically based model than the degree-day approach, which does not consider subdaily variations. The high NSE and logNSE values measured at Rhine Neuhausen, immediately downstream of Lake Constance are most likely due to the damping effect of the lake.

As regionalized model parameters (i.e., not catchment specific) are used, the VOL% (volumetric deviation) is not zero for the calibration period at most stations. The verification scores for the calibration and verification periods were in the same order of magnitude.

a. Runoff forecast verification

We first evaluated the effect of using imported SWE maps for the discharge predictions with PREVAH (P.ens and SWE.ens). For short lead times, the runoff predictions were good at discriminating between different runoff volumes (2AFC_Q), detecting low-flow events (2AFC_Q15), and predicting the runoff volume (Fig. 5). Q15 has been calculated as daily mean runoff nonexceedance probability of 15% utilizing a window of 31 days around each day of the year (Fundel et al. 2013). Note that deviations of VOL% in Figs. 4 and 5 arise

FIG. 5. Verification scores for the PREVAH discharge forecasts (left three columns) P.ens and (right three columns) SWE.ens for the gauging stations in the Thur and the Alpine Rhine catchment during the period 1991–2008. The 2AFC_Q was calculated for the continuous runoff ensemble, and 2AFC_Q15 is a probabilistic verification of the low-flow events based on the 15% runoff quantile. The lt1, lt6, lt16, and lt31 are the forecasts with a lead time of 1, 6, 16, and 31 days. The gauging stations with asterisks are uninfluenced catchments.

from the different color scales. For short lead times, the 2AFC_Q was around 0.9 for most catchments. The decrease in the 2AFC_Q with ongoing lead time was small. Importing SWE maps at each weekly reinitialization did not increase the success rate of the predictions in discriminating between the different runoff volumes observed (2AFC_Q) when the scores are averaged over all forecasts. A low-flow event could be discriminated from a nonevent using the 15th quantile (2AFC_Q15) less well than 2AFC_Q for lead times greater than lt6. Also in this case, SWE.ens did not result in improved skill in detection of low-flow events when compared to P.ens. For the runoff simulations around the disappearance of the snow cover imported SWE maps are expected to have a more positive effect on the predictions. Switching the focus to the deviations in volume error (VOL%, Fig. 5), the analysis indicates that the under- or overestimations of the runoff volume increase with ongoing lead time. The evaluation of VOL% in the case of SWE.ens leads to improved agreement as compared to P.ens for most of the evaluated basins. The estimation of snow resources at initialization of the forecast with the SWE maps seems more adequate than estimating SWE by the mean of the interpolated precipitation and temperature as realized in the case of P.ens. A closer look at particular catchments offers further insights. The 2AFC_Q15 estimations for long lead times are poorest for the Thur catchment. In this basin, snow accumulation is frequently interrupted by snowmelt in the areas below 1000 m (see Fig. 1). In Thu200 (gauges 10 and 11), runoff volume was mostly underestimated, while in Hir200 (gauges 1, 2, and 3) runoff was mostly overestimated. The overestimated runoff volume could be reduced by importing SWE maps at initialization.

Fundel et al. (2013) found that, for the Thur, the predictability of runoff decreases with ongoing lead time. The model performed skillfully for lower runoff quantiles until day 32, but for higher runoff quantiles only until day 10, after which the skill was lost. We extended this experiment to the Alpine Rhine and obtained similar indications for river basins with different area and elevation (Fig. 6). Here it can be demonstrated that the predictability of Q15 for up to 32 days lead time is achievable in high mountainous areas and in the large downstream areas collecting waters from domains located at different elevation ranges. Generally, the 2AFC scores for P.ens and SWE.ens were very similar. The score was slightly larger for the higher quantiles at lead times shorter than 10 days and for the lower quantiles at longer lead times.

Comparing the 2AFC patterns in the Thur to those in the catchments of the Alpine Rhine (Landquart) showed that, in the Thur, the decrease in the 2AFC with ongoing lead time was faster. The memory effect yielded by the higher SWE influence in the Landquart basin is the cause of this. At the gauging station Neuhausen, the 2AFC scores were high (between 0.8 and 1). Large lakes (here the Lake of Constance) act as big reservoirs and dampen the effects of flood peaks and integrate the effects of streamflow drought. Like SWE, lake storage also has a long memory effect. This improves the quality of both the reference simulations (calibration–verification period) and the predictions made using ENS.

b. SWE verification

We first compared the predicted SWE of P.ens and SWE.ens with SWE maps at the beginning (lt1) and the end (lt31) of the forecast (Fig. 7). SWE predicted with P.ens was overestimated compared to the SWE

FIG. 6. The 2AFC score for (left) Thur, (middle) Landquart, and (right) Rhine Neuhausen during the period 1991–2008. (top) P.ens and (bottom) SWE.ens daily mean runoff verified against the observed runoff in the study period 1991–2008. The probabilities that the ensemble forecast would exceed different, quantile-based, varying thresholds were verified for selected lead times of up to 32 days.

maps, especially above 2000 m MSL. The overestimation of SWE from P.ens was nearly constant for the different lead times. With the Wilcoxon–Mann–Whitney test (Wilks 2006), P.ens differed significantly from the SWE maps, with a critical value of $p < 0.05$ for all the catchments and elevation ranges investigated.

When SWE maps were imported to PREVAH at the initialization (SWE.ens), the predicted SWEs correspond with the SWE maps at lead time 1. This similarity was maintained over the entire forecast period of 32 days. Figure 7 represents an integral evaluation for different elevation bands and might therefore be affected by compensation effects concerning spatial biases. The “mapcurves” analyses presented in a later section will complement this information with a quantification of spatial agreement. In the region above 2000 m MSL the overestimation of SWE with P.ens was largest. Although the SWE with SWE.ens is visually very similar to the SWE maps, the difference between the two is statistically significant. As the snowmelt contributes to about 60% of the total runoff, the snow storage is very important for runoff predictions. The free model parameters were calibrated for the snow module according to the temperature index. This large overestimation of SWE was therefore not that obvious in the runoff predictions, which indicates the assimilation can potentially be improved. In addition to the model calibration, which optimized the runoff simulation with runoff observations,

the fact that the effect of SWE import to the runoff is evaluated for the yearly average over the considered 18 years can increase this impression.

Our results showed that the overestimation of SWE from P.ens increased from November to April (Fig. 8). To illustrate this aspect, two years with a slightly different evolution of snow cover were selected in Hir200. In 1998, melting events in February and April were observed. In 2000, however, the snow cover grew steadily until May, when the melting period began. The overestimated SWE from P.ens with respect to the SWE maps was eliminated with SWE.ens for all lead times and over the entire season.

The ensemble spread at lt1 was narrow and grew slightly with ongoing lead time. This narrow ensemble band indicates that the uncertainty of the meteorological input variables was damped in the prediction of the snow storage with respect to spread, which is what usually happens when the focus is on streamflow only (Fundel et al. 2013; Addor et al. 2011). The differences between SWE from P.ens and SWE.ens grew constantly over the season in 1998, while in 2000 they were small at the beginning of the winter season with a sharp increase in April. Despite these differences, at the beginning of the melting period the overestimated SWE of P.ens compared to SWE.ens was similar in both years. The melting period in February 1998 was simulated in P.ens and SWE.ens.

FIG. 7. Distribution of SWE in the five elevation ranges (<1000, 1000–1500, 1500–2000, 2000–2500, and >2500 m MSL) for (from top to bottom) the Thu200, LaP200, and HiR200 domains during the period 1991–2008. Boxes display the median and the interquartile range and are drawn for SWE from P.ens, SWE.ens for lead time 1 (no filling), and lead time 31 (gray filling). The whiskers extend to the maximum SWE. SWE maps are the observed values of the dates that correspond to the dates of the forecasts at lead times 1 and 31.

We expected the imported SWE maps to influence SWE predictions more than the runoff predictions. The ability of the forecast to discriminate between different SWE values was smallest at the beginning and end of the snow season (Fig. 9). This demonstrates that the ensemble has problems in capturing SWE in the phases where air temperature is close to zero and therefore short periods with and without snow can occur. As soon as air temperatures are clearly below zero degrees, predictability improves. Similarly to runoff, the predictability of SWE decreased with ongoing lead time. The decrease in the predictability with lead time was largest in Thu200, smaller in LaP200, and smallest in HiR200. Generally, the SWE predictability was larger in LaP200 and HiR200 than in Thu200. The import of SWE

maps at initialization increased the model performance of the prediction up to a lead time of approximately 20 days in Thu200 and approximately 25 days in LaP200 and HiR200. The white areas in the figure indicate periods without snow cover (May in Thu200) or periods where at initialization no SWE maps were available. In LaP200, for a lead time of 21 days, the SWE.ens predicted an SWE of zero in all forecasts, which explains the white panel.

The increased predictability of SWE occurred without decreasing the performance of the runoff prediction. This might mean that the exact distribution and quantity of snow resources do not affect the quality of streamflow simulation at the scale of the basins we considered. This is a classic paradox in models when they are calibrated and evaluated against discharge (integral of the catchment response; [Kirchner 2006](#)). The use of local SWE maps might help to reduce the uncertainty for small ungauged areas. Generally, even if some scores might suffer from importing SWE maps, the overall outcome of the model was better at representing the actual dynamics of water resources in the target area. It has also to be acknowledged that errors in SWE estimations affect runoff only when model and reality have different status concerning snow coverage (yes/no) and/or the SWE is in the same magnitude of the potential snowmelt at a particular day.

The spatial verification is illustrated visually and with the similarity measure mapcurves for a randomly chosen day (5 February 2001, Fig. 10). This day was in the middle of the snow season of a year with above average snow-cover depths. P.ens overestimated SWE at high elevation 100% more than the SWE maps. This was also represented in a bias larger than 300 mm and mapcurves similarities of less than 10%. Neither visually nor with the two scores was a considerable change in the ongoing lead time detected in P.ens. The extremely high SWE values simulated were accumulated on the glaciers.

As expected, the SWE simulated with SWE.ens corresponded well with the SWE maps, and the over- or underestimations were in the range of 40 mm. The mapcurves similarity increased by up to 80% at lead time 1. At lead time 1 SWE.href and SWE.ens were identical with the SWE maps, except for the water surfaces, which were snow-free by definition in PREVAH. We are currently analyzing this limitation. With ongoing lead time, SWE.href overestimated SWE at the highest elevations in the southwestern part and underestimated SWE in the valleys. The ensembles mean (SWE.ens) overestimated SWE in the northeastern part.

The difference between SWE.href (forced by interpolated precipitation) and SWE.ens (forced by ENS) may be due to problems in the representation of temperature

FIG. 8. Evolution of forecast SWE and SWE maps for different lead times for the hydrological years (left) 1998 and (right) 2000 in HiR200. Plots for other regions can be found in the supplemental material. The black and gray-shaded areas represent the spread of the ensemble for the lead times 1, 6, 16, and 31. (top) The SWE predictions (red dots and lines) from P.ens, (middle) those from SWE.ens, and (bottom) their differences. The vertically shaded areas indicate the season when PREVAH SWE was replaced by SWE maps at initialization.

gradients and precipitation in ENS, which uses a very coarse topography. Probably pools of cold air and other fine resolution local effects were not well represented either by the interpolation tools adopted (Viviroli et al. 2009c) or by ENS. A new temperature dataset has been recently made available by MeteoSwiss (Frei 2014), which might alleviate such problems. An assessment of this additional dataset is beyond the goals of this paper, but will be addressed in follow-up investigations. In addition, there is potential for improvement in the downscaling method of the ENS reforecast. Note that this was an example of the spatial verification of one randomly chosen 32-day forecast to visualize the SWE predictions.

For a general spatial verification, mapcurves was averaged over all forecasts for different lead times and over the winter season (Fig. 11). The mean mapcurves values for the ensemble were used because the spread of the ensemble was small. As the classification intervals have a range from 5 to 1000 mm, the probability that two ensembles will be in the same class is high. The mapcurves similarity was high at Lt1 and decreased with ongoing lead time, except for the comparison of P.ens

with the SWE maps (Fig. 11c). The decrease in the similarity was fastest during the melting process. Mapcurves for the comparison of SWE.maps with SWE.href and with P.href is very similar to the values of the comparison of SWE.maps with SWE.ens (Fig. 11a), or to P.ens (Fig. 11c). For the evolution of the snow cover in PREVAH, the initial state is therefore more important than the quality of the meteorological parameters. The mapcurves verification for additional domains can be found in the supplemental material.

6. Conclusions

In the present study, gridded SWE maps derived from snow depth measurements were imported into a hydrological model at initialization to provide ensemble predictions of streamflow and SWE. The predictions have a lead time of 32 days and were tested in selected alpine domains in Switzerland, where the contribution of snowmelt to runoff differs. The focus was on the influence of imported SWE maps on the predictability of runoff and SWE. We evaluated the effect of replacing

FIG. 9. The 2AFC score for ensemble SWE forecasts for (left) Thu200, (middle) LaP200, and (right) HiR200 domains for the period 1991–2008. (top) P.ens and (bottom) SWE.ens daily SWE forecasts verified against SWE maps in the study period 1991–2008. Probabilities that the ensemble forecasts would exceed different, quantile-based, varying thresholds were verified for selected lead times of up to 32 days.

model SWE with SWE maps at initialization using different validation scores.

The results reveal that imported SWE maps increased the predictability of SWE without decreasing the performance of the runoff predictions. With respect to a constant or seasonally varying benchmark and the ability of the model to distinguish between different runoff volumes, model performance was not greatly enhanced by importing SWE maps, but the error of the streamflow volume could be reduced. This suggests that importing SWE maps could be a way of improving estimations of precipitation adjustment factors (Sevruk 1986) in high alpine areas. As this experiment was a first analysis of the response of a runoff–precipitation model to changes in SWE information, no new model calibration was performed. The parameters of the model were calibrated in a way to optimize the runoff predictions with the SWE simulated with the PREVAH snow module. Although runoff predictability was not increased by importing SWE maps, it is reasonable to include this information to guarantee a better transferability. As can be seen in the decrease of the volumetric runoff deviation (VOL%), the state of the snow storage is improved. This enables a better representation of the snowmelt process that is beneficial for runoff predictions outside the calibration period. A better initial snow state is important for long-range

predictions and is beneficial for predicting the timing of the snowmelt.

In the Alpine Rhine catchments, the damping effect of Lake Constance was illustrated. Major lakes act as large reservoirs and can therefore absorb the effects of flood peaks and relieve drought and low-flow events. This improves the quality of the predictions at the gauging stations downstream of the lake. Thus, the simulation of streamflow was of the lowest quality achieved for basins with areas between 500 and 2000 km². Such basins cover a large range of elevation and are mostly affected by hydropowering. However, in this study, the operations of hydropower facilities were greatly idealized. We anticipate that early prediction of streamflow drought and water resources may be more successful if a close collaboration with the hydropower companies could be established.

For the SWE predictions, including the SWE maps was very valuable to improve the SWE forecasts up to a lead time of 25 days. This improvement in SWE predictions was not very surprising as the modeled SWE is very prone to errors in the predicted or observed precipitation and temperature fields. In Fig. 11 of the supplemental material, we present an evaluation of the intrinsic predictability of discharge. Such predictability is limited to ten days for most of the year, with a peak of 25 days during the snowmelt season.

FIG. 10. (from left to right) SWE maps and differences between predicted SWE and SWE maps for HiR200 for the randomly chosen forecast from 5 Feb 2001 for lead times (from top to bottom) 1, 11, and 21. The numbers in the plots are the bias and the mapcurves similarity (mc).

Thus, SWE is consistently more predictable than what one would expect from the input uncertainty. SWE is highly persistent up to a lag of 2 weeks, especially in the higher catchments. Apart from the meteorological information that is used in the PREVAH forecasts, the uncertainty representation from the ensemble forecast provides additional value to the case where the initial SWE would be used for the entire forecast period. Further, the predicted SWE fields were validated with the same product that had been imported to the hydrological model. We could not avoid this, because no other SWE product with a similar quality was available for the validation.

It should also be noted that the subgrid variability differs between the SWE maps and the SWE predictions. The SWE maps were estimated by considering subgrid variability that depended on the variability of land cover and slopes. PREVAH, however, corrects subgrid temperature according to slope and aspect. Therefore, PREVAH takes into consideration some of the topographic effects of snowmelt (Gurtz et al. 2003;

Zappa et al. 2003). The effects of snow interception and snow in forests are not taken into account in the SWE predictions in PREVAH.

In-depth analyses indicated that the absolute and relative error of predicted SWE is small (± 5 cm) below 2000 m MSL. Above 2000 m MSL, larger differences between the predicted and observed SWE were found at the end of the winter season (March and April). Modeling the snowmelt correctly is difficult partly because the parameterization of the snow model optimized for the HRU version of PREVAH does not allow a redistribution of SWE. The interpolation of precipitation and temperature also potentially causes problems in snowmelt modeling.

The similarity measure mapcurves showed a large increase in the overlapping region from SWE predictions compared with the original simulations, where SWE originated from the PREVAH snow module and predicted SWE with SWE maps at initialization. The mapcurves values decreased relatively fast in the first forecast week, but thereafter the decrease was slow.

FIG. 11. Similarity of (from left to right) predicted SWE and SWE maps with the validation score mapcurves for the months November–April and different lead times (x axis) in the domain HiR200 for the period 1991–2008. In the supplemental material, plots for other regions are given.

We have shown that importing SWE maps can improve the SWE prediction for a considerable lead time without decreasing the predictability of the runoff. This is an important achievement, because an improved state of SWE helps to provide more accurate information of the available water for snowmelt, which is crucial for runoff prediction in spring. The assimilation of SWE maps into a hydrological model should provide superior SWE estimates because it combines information from observations and modeling predictions, ideally taking into consideration the limitations of each (Andreadis and Lettenmaier 2006). Frequently used methods to assimilate SWE into hydrological models are the ensemble Kalman filter, which considers the measurement uncertainty (Evensen 2007; Clark et al. 2006; Slater and Clark 2006), and the particle filter, which conserves the mass balance in the system (Thirel et al. 2013).

In a next step, the assimilation of SWE maps could be tested to take into account the modeled information and the total water balance of the hydrological volume. This could help to investigate the hypothesis of better outcomes for internal (ungauged) subareas in follow-up studies. In addition, model parameters should be calibrated with changed SWE initialization to be able to use the SWE maps in operational hydrological predictions. Soon medium-range (32 days) forecasts are going to be delivered to the Swiss prototype platform for early recognition of droughts and water shortages (www.drought.ch; Zappa et al. 2014; Seneviratne et al. 2013).

Acknowledgments. This study was supported by the Swiss National Research Program Sustainable Water

Management (NRP 61 project DROUGHT-CH). We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ENS reforecast and the Federal Office for the Environment (FOEN) for providing the river runoff observations. We thank MeteoSwiss for providing the meteorological observations. The data analysis and visualization were performed with R. We thank Mischa Schmid, Felix Fundel, Matthias Speich, Tobias Jonas, Florian Pappenberger, and Sonia Seneviratne for their preliminary work, helpful comments, and suggestions for improving the manuscript and Silvia Dingwall for reviewing the language of this article. The comments of three anonymous reviewers helped to clarify the structure of the article and are greatly appreciated.

REFERENCES

- Addor, N., S. Jaun, F. Fundel, and M. Zappa, 2011: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.*, **15**, 2327–2347, doi:10.5194/hess-15-2327-2011.
- Andreadis, K. M., and D. P. Lettenmaier, 2006: Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.*, **29**, 872–886, doi:10.1016/j.advwatres.2005.08.004.
- Aubert, D., C. Loumagne, and L. Oudin, 2003: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. *J. Hydrol.*, **280**, 145–161, doi:10.1016/S0022-1694(03)00229-4.
- Begert, M., T. Schlegel, and W. Kirchhofer, 2005: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int. J. Climatol.*, **25**, 65–80, doi:10.1002/joc.1118.
- Bernhard, L., and M. Zappa, 2012: Schlussbericht CCHydrologie: Natürlicher Wasserhaushalt der Schweiz und ihrer bedeutendsten Grosseinzugsgebiete. WSL Tech. Rep., 84 pp.

- Blanc, P., and B. Schaedler, 2014: Water in Switzerland—An overview. *Swiss Hydrological Commission Rep.*, 28 pp.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer, 2010: Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrol. Earth Syst. Sci.*, **14**, 1881–1893, doi:[10.5194/hess-14-1881-2010](https://doi.org/10.5194/hess-14-1881-2010).
- Clark, M. P., A. G. Slater, A. P. Barrett, L. E. Hay, G. J. McCabe, B. Rajagopalan, and G. H. Leavesley, 2006: Assimilation of snow covered area information into hydrologic and land-surface models. *Adv. Water Resour.*, **29**, 1209–1221, doi:[10.1016/j.advwatres.2005.10.001](https://doi.org/10.1016/j.advwatres.2005.10.001).
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, doi:[10.1016/j.jhydrol.2009.06.005](https://doi.org/10.1016/j.jhydrol.2009.06.005).
- Dale, M., J. Wicks, K. Mylne, F. Pappenberger, S. Laeger, and S. Taylor, 2014: Probabilistic flood forecasting and decision-making: An innovative risk-based approach. *Nat. Hazards*, **70**, 159–172, doi:[10.1007/s11069-012-0483-z](https://doi.org/10.1007/s11069-012-0483-z).
- Demargne, J., and Coauthors, 2014: The science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bull. Amer. Meteor. Soc.*, **95**, 79–98, doi:[10.1175/BAMS-D-12-00081.1](https://doi.org/10.1175/BAMS-D-12-00081.1).
- Demirel, M. C., M. J. Booij, and A. Y. Hoekstra, 2013: Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resour. Res.*, **49**, 4035–4053, doi:[10.1002/wrcr.20294](https://doi.org/10.1002/wrcr.20294).
- Dietz, A. J., C. Kuenzer, U. Gessner, and S. Dech, 2012: Remote sensing of snow—A review of available methods. *Int. J. Remote Sens.*, **33**, 4094–4134, doi:[10.1080/01431161.2011.640964](https://doi.org/10.1080/01431161.2011.640964).
- Evensen, G., 2007: *Data Assimilation: The Ensemble Kalman Filter*. Springer, 279 pp.
- Frei, C., 2014: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *Int. J. Climatol.*, **34**, 1585–1605, doi:[10.1002/joc.3786](https://doi.org/10.1002/joc.3786).
- Fundel, F., and M. Zappa, 2011: Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts. *Water Resour. Res.*, **47**, W09520, doi:[10.1029/2010WR009996](https://doi.org/10.1029/2010WR009996).
- , A. Walser, M. A. Liniger, C. Frei, and C. Appenzeller, 2010: Calibrated precipitation forecasts for a limited-area ensemble forecast system using reforecasts. *Mon. Wea. Rev.*, **138**, 176–189, doi:[10.1175/2009MWR2977.1](https://doi.org/10.1175/2009MWR2977.1).
- , S. Jörg-Hess, and M. Zappa, 2013: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. *Hydrol. Earth Syst. Sci.*, **17**, 395–407, doi:[10.5194/hess-17-395-2013](https://doi.org/10.5194/hess-17-395-2013).
- Gruber, S., 2007: A mass-conserving fast algorithm to parameterize gravitational transport and deposition using digital elevation models. *Water Resour. Res.*, **43**, W06412, doi:[10.1029/2006WR004868](https://doi.org/10.1029/2006WR004868).
- Gurtz, J., A. Baltensweiler, and H. Lang, 1999: Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. *Hydrol. Processes*, **13**, 2751–2768, doi:[10.1002/\(SICI\)1099-1085\(19991215\)13:17<2751::AID-HYP897>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1099-1085(19991215)13:17<2751::AID-HYP897>3.0.CO;2-O).
- , M. Zappa, K. Jasper, H. Lang, M. Verbunt, A. Badoux, and T. Vitvar, 2003: A comparative study in modelling runoff and its components in two mountainous catchments. *Hydrol. Processes*, **17**, 297–311, doi:[10.1002/hyp.1125](https://doi.org/10.1002/hyp.1125).
- Hargrove, W. W., F. M. Hoffman, and P. F. Hessburg, 2006: Map-curves: A quantitative method for comparing categorical maps. *J. Geogr. Syst.*, **8**, 187–208, doi:[10.1007/s10109-006-0025-x](https://doi.org/10.1007/s10109-006-0025-x).
- He, M., T. S. Hogue, S. A. Margulis, and K. J. Franz, 2012: An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. *Hydrol. Earth Syst. Sci.*, **16**, 815–831, doi:[10.5194/hess-16-815-2012](https://doi.org/10.5194/hess-16-815-2012).
- Hock, R., 1999: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. *J. Glaciol.*, **45**, 101–111.
- Hüsler, F., T. Jonas, M. Riffler, J. P. Musial, and S. Wunderle, 2014: A satellite-based snow cover climatology (1985–2011) for the European Alps derived from AVHRR data. *Cryosphere*, **8**, 73–90, doi:[10.5194/tc-8-73-2014](https://doi.org/10.5194/tc-8-73-2014).
- Jonas, T., C. Marty, and J. Magnusson, 2009: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *J. Hydrol.*, **378**, 161–167, doi:[10.1016/j.jhydrol.2009.09.021](https://doi.org/10.1016/j.jhydrol.2009.09.021).
- Jörg-Hess, S., F. Fundel, T. Jonas, and M. Zappa, 2014: Homogenisation of a gridded snow water equivalent climatology for alpine terrain: Methodology and applications. *Cryosphere*, **8**, 471–485, doi:[10.5194/tc-8-471-2014](https://doi.org/10.5194/tc-8-471-2014).
- , S. B. Kempf, F. Fundel, and M. Zappa, 2015: The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland. *Meteor. Appl.*, **22**, 444–458, doi:[10.1002/met.1474](https://doi.org/10.1002/met.1474).
- Kirchner, J. W., 2006: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.*, **42**, W03S04, doi:[10.1029/2005WR004362](https://doi.org/10.1029/2005WR004362).
- Kobierska, F., T. Jonas, M. Zappa, M. Bavay, J. Magnusson, and S. M. Bernasconi, 2013: Future runoff from a partly glacierized watershed in central Switzerland: A two-model approach. *Adv. Water Resour.*, **55**, 204–214, doi:[10.1016/j.advwatres.2012.07.024](https://doi.org/10.1016/j.advwatres.2012.07.024).
- Köplin, N., D. Viviroli, B. Schdler, and R. Weingartner, 2010: How does climate change affect mesoscale catchments in Switzerland? A framework for a comprehensive assessment. *Adv. Geosci.*, **27**, 111–119, doi:[10.5194/adgeo-27-111-2010](https://doi.org/10.5194/adgeo-27-111-2010).
- Koster, R. D., S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle, 2010: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nat. Geosci.*, **3**, 613–616, doi:[10.1038/ngeo944](https://doi.org/10.1038/ngeo944).
- Krause, P., D. Boyle, and F. Bäse, 2005: Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.*, **5**, 89–97, doi:[10.5194/adgeo-5-89-2005](https://doi.org/10.5194/adgeo-5-89-2005).
- Kruse, S., and I. Seidl, 2013: Social capacities for drought risk management in Switzerland. *Nat. Hazards Earth Syst. Sci.*, **13**, 3429–3441, doi:[10.5194/nhess-13-3429-2013](https://doi.org/10.5194/nhess-13-3429-2013).
- Liechti, K., M. Zappa, F. Fundel, and U. Germann, 2013: Probabilistic evaluation of ensemble discharge nowcasts in two nested alpine basins prone to flash floods. *Hydrol. Processes*, **27**, 5–17, doi:[10.1002/hyp.9458](https://doi.org/10.1002/hyp.9458).
- Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström, 1997: Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.*, **201**, 272–288, doi:[10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3).
- Magnusson, J., D. Gustafsson, F. Hüsler, and T. Jonas, 2014: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water Resour. Res.*, **50**, 7816–7835, doi:[10.1002/2014WR015302](https://doi.org/10.1002/2014WR015302).
- Marty, R., I. Zin, and C. Obled, 2013: Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of probabilistic quantitative precipitation forecasts: Flash flood case studies in the Cevennes–Vivarais region (southern France). *Hydrol. Processes*, **27**, 33–44, doi:[10.1002/hyp.9543](https://doi.org/10.1002/hyp.9543).

- Mason, S. J., and A. P. Weigel, 2009: A generic forecast verification framework for administrative purposes. *Mon. Wea. Rev.*, **137**, 331–349, doi:10.1175/2008MWR2553.1.
- Moradkhani, H., 2008: Hydrologic remote sensing and land surface data assimilation. *Sensors*, **8**, 2986–3004, doi:10.3390/s8052986.
- Nash, J., and J. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.
- Pappenberger, F., K. J. Beven, N. M. Hunter, P. D. Bates, B. T. Gouweleeuw, J. Thielen, and A. P. J. de Roo, 2005: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European flood forecasting system (EFFS). *Hydrol. Earth Syst. Sci.*, **9**, 381–393, doi:10.5194/hess-9-381-2005.
- Ramos, M. H., S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, **17**, 2219–2232, doi:10.5194/hess-17-2219-2013.
- Rees, W. G., 2008: Comparing the spatial content of thematic maps. *Int. J. Remote Sens.*, **29**, 3833–3844, doi:10.1080/01431160701852088.
- Schaeffli, B., and H. V. Gupta, 2007: Do Nash values have value? *Hydrol. Processes*, **21**, 2075–2080, doi:10.1002/hyp.6825.
- Schattan, P., M. Zappa, H. Lischke, L. Bernhard, E. Thuring, and B. Diekkruiger, 2013: An approach for transient consideration of forest change in hydrological impact studies. *IAHS Publ.*, **359**, 311–319.
- Seneviratne, S. I., and Coauthors, 2012: Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resour. Res.*, **48**, W06526, doi:10.1029/2011WR011749.
- , and Coauthors, 2013: Droughts in Switzerland (in German). *Aqua Gas*, **9**, 38–47.
- Sevruk, B., Ed., 1986: *Correction of Precipitation Measurements*. ETH/IASH/WMO Workshop on the Correction of Precipitation Measurements, Vol. 23, ETH Zürich, 288 pp.
- Shukla, S., N. Voisin, and D. P. Lettenmaier, 2012: Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill. *Hydrol. Earth Syst. Sci.*, **16**, 2825–2838, doi:10.5194/hess-16-2825-2012.
- Slater, A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. *J. Hydrometeorol.*, **7**, 478–493, doi:10.1175/JHM505.1.
- Speich, M. J. R., B. Bernhard, A. J. Teuling, and M. Zappa, 2015: Application of bivariate mapping for hydrological classification and analysis of temporal change and scale effects in Switzerland. *J. Hydrol.*, **523**, 804–821, doi:10.1016/j.jhydrol.2015.01.086.
- Su, H., Z. L. Yang, G. Y. Niu, and R. E. Dickinson, 2008: Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter. *J. Geophys. Res.*, **113**, D08120, doi:10.1029/2007JD009232.
- Tharme, R. E., 2003: A global perspective on environmental flow assessment: Emerging trends in the development and application of environmental flow methodologies for rivers. *River Res. Appl.*, **19**, 397–441, doi:10.1002/rra.736.
- Thirel, G., E. Martin, J.-F. Mahfouf, S. Massart, S. Ricci, and F. Habets, 2010a: A past discharges assimilation system for ensemble streamflow forecasts over France—Part 1: Description and validation of the assimilation system. *Hydrol. Earth Syst. Sci.*, **14**, 1623–1637, doi:10.5194/hess-14-1623-2010.
- , —, —, —, —, F. Regimbeau, and F. Habets, 2010b: A past discharge assimilation system for ensemble streamflow forecasts over France—Part 2: Impact on the ensemble streamflow forecasts. *Hydrol. Earth Syst. Sci.*, **14**, 1639–1653, doi:10.5194/hess-14-1639-2010.
- , P. Salamon, P. Burek, and M. Kalas, 2013: Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. *Remote Sens.*, **5**, 5825–5850, doi:10.3390/rs5115825.
- Verbunt, M., J. Gurtz, K. Jasper, H. Lang, P. Warmerdam, and M. Zappa, 2003: The hydrological role of snow and glaciers in alpine river basins and their distributed modeling. *J. Hydrol.*, **282**, 36–55, doi:10.1016/S0022-1694(03)00251-8.
- , A. Walser, J. Gurtz, A. Montani, and C. Schär, 2007: Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies. *J. Hydrometeorol.*, **8**, 897–909, doi:10.1175/JHM594.1.
- Vitart, F., and Coauthors, 2008: The new VarEPS-monthly forecasting system: A first step towards seamless prediction. *Quart. J. Roy. Meteor. Soc.*, **134**, 1789–1799, doi:10.1002/qj.322.
- Viviroli, D., H. Mittelbach, J. Gurtz, and R. Weingartner, 2009a: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland—Part II: Parameter regionalisation and flood estimation results. *J. Hydrol.*, **377**, 208–225, doi:10.1016/j.jhydrol.2009.08.022.
- , M. Zappa, J. Gurtz, and R. Weingartner, 2009b: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environ. Modell. Software*, **24**, 1209–1222, doi:10.1016/j.envsoft.2009.04.001.
- , —, J. Schwanbeck, J. Gurtz, and R. Weingartner, 2009c: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland—Part I: Modelling framework and calibration results. *J. Hydrol.*, **377**, 191–207, doi:10.1016/j.jhydrol.2009.08.023.
- Weigel, A. P., and S. J. Mason, 2011: The generalized discrimination score for ensemble forecasts. *Mon. Wea. Rev.*, **139**, 3069–3074, doi:10.1175/MWR-D-10-05069.1.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- Zappa, M., and C. Kan, 2007: Extreme heat and runoff extremes in the Swiss Alps. *Nat. Hazards Earth Syst. Sci.*, **7**, 375–389, doi:10.5194/nhess-7-375-2007.
- , F. Pos, U. Strasser, P. Warmerdam, and J. Gurtz, 2003: Seasonal water balance of an alpine catchment as evaluated by different methods for spatially distributed snowmelt modeling. *Nord. Hydrol.*, **34**, 179–202.
- , and Coauthors, 2008: MAP D-PHASE: Real-time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Lett.*, **9**, 80–87, doi:10.1002/asl.183.
- , S. Jaun, U. Germann, A. Walser, and F. Fundel, 2011: Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmos. Res.*, **100**, 246–262, doi:10.1016/j.atmosres.2010.12.005.
- , L. Bernhard, F. Fundel, and S. Jörg-Hess, 2012: Forecasts and scenarios of snow and water resources in alpine environments (in German). *Forum für Wissen 2012*, Federal Institute for Forest, Snow and Landscape Research WSL, 19–27.
- , —, C. Spirig, K. Stahl, I. Seidl, and M. Stähli, 2014: A prototype platform for water resources monitoring and early recognition of critical droughts in Switzerland. *IAHS Publ.*, **364**, 492–498.