Ensemble Postprocessing of Daily Precipitation Sums over Complex Terrain Using Censored High-Resolution Standardized Anomalies

RETO STAUFFER
Department of Statistics, and Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria

NIKOLAUS UMLAUF
Department of Statistics, University of Innsbruck, Innsbruck, Austria

JAKOB W. MESSNER
Department of Statistics, and Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria

GEORG J. MAYR
Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria

ACHIM ZEILEIS
Department of Statistics, University of Innsbruck, Innsbruck, Austria

(Manuscript received 16 July 2016, in final form 18 October 2016)

ABSTRACT

Probabilistic forecasts provided by numerical ensemble prediction systems have systematic errors and are typically underdispersive. This is especially true over complex topography with extensive terrain-induced small-scale effects, which cannot be resolved by the ensemble system. To alleviate these errors, statistical post-processing methods are often applied to calibrate the forecasts. This article presents a new full-distributional spatial postprocessing method for daily precipitation sums based on the standardized anomaly model output statistics (SAMOS) approach. Observations and forecasts are transformed into standardized anomalies by subtracting the long-term climatological mean and dividing by the climatological standard deviation. This removes all site-specific characteristics from the data and makes it possible to fit one single regression model for all stations at once. As the model does not depend on the station locations, it directly allows the creation of probabilistic forecasts for any arbitrary location. SAMOS uses a left-censored power-transformed logistic response distribution to account for the large fraction of zero observations (dry days), the limitation to nonnegative values, and the positive skewness of the data. ECMWF reforecasts are used for model training and to correct the ECMWF ensemble forecasts with the big advantage that SAMOS does not require an extensive archive of past ensemble forecasts as only the most recent four reforecasts are needed, and it automatically adapts to changes in the ECMWF ensemble model. The application of the new method to the central Alps shows that the new method is able to depict the small-scale properties and returns accurate fully probabilistic spatial forecasts.

Denotes content that is immediately available upon publication as open access.

Corresponding author e-mail: Reto Stauffer, reto.stauffer@uibk.ac.at

This article is licensed under a Creative Commons Attribution 4.0 license (http://creativecommons.org/licenses/by/4.0/).

DOI: 10.1175/MWR-D-16-0260.1

© 2017 American Meteorological Society.
1. Introduction

In mountainous regions, large amounts of precipitation can lead to severe floods and landslides during spring and summer and to dangerous avalanche conditions during winter. Accurate and reliable knowledge about the expected precipitation can therefore be crucial for strategic planning and to raise awareness among the public.

Precipitation forecasts, or weather forecasts in general, are typically provided by numerical weather prediction models. Nowadays, most forecast centers also compute probabilistic forecasts based on numerical ensemble prediction systems (EPSs; Epstein 1969; Buizza et al. 2005) as probabilistic information can be crucial, for example, for strategic planning or decision-makers. An ensemble consists of several (independent) forecast runs with slightly different initial conditions, model physics, and/or parameterizations. The goal of an EPS is not only to provide one single forecast but also to provide additional information about the weather-situation-dependent forecast uncertainty. Although EPSs are undergoing constant improvements, they are not able to provide fully reliable forecasts and are typically underdispersive (Mullen and Buizza 2001; Hagedorn et al. 2012).

To correct for systematic errors and to correct the uncertainty provided by the EPS, postprocessing methods are often applied. A variety of ensemble postprocessing methods for precipitation are available nowadays, such as analog methods (Hamill et al. 2006, 2015), ensemble dressing (Roulston and Smith 2003), Bayesian model averaging (BMA; Sloughter et al. 2007; Fraley et al. 2010), extended logistic regression (Wilks 2009; Ben Bouallégue and Theis 2014; Messner et al. 2014b), or nonhomogeneous regression (Gneiting et al. 2005). Several extensions exist for nonnormally distributed variables (Thorarinsdottir and Gneiting 2010; Lerch and Thorarinsdottir 2013; Scheuerer 2014; Scheuerer and Hamill 2015). For precipitation, Messner et al. (2014a) show that a censored logistic regression fits well, while Scheuerer (2014) and Scheuerer and Hamill (2015) use a left-censored generalized extreme value (GEV) distribution or a left-censored shifted gamma distribution, respectively.

These postprocessing methods are often applied on a station or gridpoint level such that for each location, one set of regression coefficients is estimated to correct the ensemble forecasts. However, for a wide range of applications, predictions for locations between observational sites are of great interest. Therefore, the regression models have to be extended such that spatial probabilistic predictions can be made.

In this article, a new spatial statistical postprocessing method for daily precipitation sums over complex terrain is presented. Even on a small spatial scale, two neighboring stations can show very different characteristics in terms of observed precipitation sums. These differences can be caused by topographically induced flow regimes, orographic lifting and shading effects, convective regimes, and many other factors. Most of these processes cannot yet be resolved by global EPS models. To account for these small-scale spatial variabilities among all stations, we are using an adapted version of the anomaly approach first published by Scheuerer and Bürermann (2014) and further extended by Dabernig et al. (2017). Observations and ensemble forecasts are transformed into standardized anomalies by subtracting the long-term climatological mean and dividing by the climatological standard deviation. This removes the station-dependent characteristics from the data and makes it possible to fit one single regression model for all stations at once. As the model does not rely on site-specific characteristics anymore, the corrections can be applied to future ensemble forecasts to create probabilistic forecasts for any arbitrary location within the area of interest.

Following Dabernig et al. (2017), we use the standardized anomaly model output statistics (SAMOS) approach and extend the framework to fulfill all requirements needed for precipitation postprocessing. SAMOS offers a simple and computationally efficient framework for fully probabilistic spatial postprocessing and is applied to the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble in combination with the ECMWF reforecasts. The approach presented qualifies for an operational system as no extensive archive of historical forecasts is required. SAMOS uses a rolling 4-week time window as a training dataset so that only the reforecasts of the most recent month from the operational ECMWF data dissemination have to be retained, which currently (in 2016) consist of eight independent reforecast runs covering the previous 20 years. Because of this rolling training dataset, SAMOS automatically adapts itself to the latest ensemble model version within a very short time period.

2. Area of interest and data

a. Study area

To develop and validate the new method presented in this study, we focus on the governmental area of Tyrol, Austria. Tyrol has a size of about 12,500 km² and is home to approximately 740,000 inhabitants (Statistik Austria 2016) living in the two separated parts, with North Tyrol on the north side of the main Alpine ridge and East Tyrol south of the main Alpine ridge. The study area is located in the eastern part of the Alps, showing a highly complex topography. Figure 1 shows the state borders of Tyrol and the topography reaching from 465 to 3798 m MSL, including some of the highest mountains in
Austria. Because of the high population density and the strong economic focus on tourism (>10 million tourists in 2014; Kaiser et al. 2014), there is a high demand for accurate weather forecasts.

**b. Observational data**

The local hydrographical service provides a dense precipitation measurement network, whereof 117 stations in Tyrol and its surroundings will be used for model training and validation spanning September 1971 through the end of 2012. The mean distance to the four closest stations in the surroundings is only about 10 km. Locations of the observation sites are highlighted in Fig. 1. The hydrographical service performs rigorous quality controls on the observations and makes them freely available for any noncommercial use on the maintainers’ website (Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft 2016).

**c. Numerical weather forecast data**

The numerical forecasts are obtained from the ECMWF, including the operational ensemble (ENS; 0000 UTC initial), which consists of 50 + 1 individual forecasts based on perturbed initial conditions (50 forecasts plus control run) and the ECMWF reforecast dataset. The ECMWF reforecast dataset has existed since 18 February 2010 and was slightly extended over the years. Until 14 June 2012, the reforecast was computed once a week, providing ensemble reforecasts consisting of 4 + 1 members for the most recent 18 years.
From 21 June 2012 through the end of 2012, the number of years was extended from 18 to 20. This reforecast is designed to provide the model climate of the latest ECMWF ENS version and is often used for model calibration (e.g., Hamill et al. 2008; Hamill 2012).

In this study, the time period from February 2010 to December 2012 is used. Every Thursday, the reforecasts for the same date two weeks in advance have been computed, including a 4 + 1 member ensemble for the most recent 18–20 years. As an example, on Thursday 1 November 2012, the reforecast for 15 November has become available for the most recent 20 years, namely 15 November 2011, 15 November 2010, . . ., 15 November 1992, with 4 + 1 members each.

d. Training and verification dataset

The ECMWF reforecasts are used to compute the climatology of the ECMWF ensemble, which will be used as background information and to train the statistical postprocessing, including the most recent four reforecast runs centered around the current date (computed every Thursday; section 2c). Therefore, the model climatology is based on 4 runs × 5 members × 20 yr = 400 individual forecasts (details in section 3c). For the training dataset, the reforecasts are bilinearly interpolated to each of the 117 observation sites. Out of each interpolated reforecast ensemble (daywise, 4 + 1 members), the mean and standard deviation is used later to build the training dataset. We use the most recent four 0000 UTC reforecast runs yielding to a training sample of up to 4 runs × 20 yr = 80 data pairs per station, or 4 runs × 20 yr × 117 stations = 9360 observation–forecast pairs for the full spatial SAMOS (details in section 4b).

Once the regression coefficients are estimated, the correction can be applied to future EPS forecasts using the mean and standard deviation of the 50 + 1 members of the ECMWF ENS.

Because of the availability of the observations (section 2b) and the ECMWF reforecasts (section 2c), the time period between 26 February 2010 and 31 December 2012 will be used for verification, with an overall data availability of 99.4% and roughly 120,500 unique observation–forecast pairs.

3. Methodology

a. Censored nonhomogeneous logistic regression (CNLR)

The distribution of precipitation observations at a particular observation site shows three main properties: it is limited to nonnegative values, has a large fraction of 0 observations (dry days), and is strongly positively skewed. We take the nonhomogeneous Gaussian regression (NGR; Gneiting et al. 2005) as our base model and extend the NGR framework to suit spatial precipitation postprocessing.

In contrast to the original NGR, a logistic response distribution is assumed. The logistic distribution shows a similar bell shape as the Gaussian distribution but has slightly heavier tails. The logistic distribution is defined by two parameters: the location \( \mu \) describing the mean and the scale \( \sigma \) describing the width of the distribution. To remove the positive skewness, a power transformation \( 1/p \) is applied to the observations and to every ensemble member (Box and Cox 1964). Different power parameters \( p \) have already been suggested in the literature for precipitation applications such as \( p = 3 \) (Stidd 1973) or \( p = 2 \) (Hutchinson 1998). However, the optimal power parameter is a function of the data, the model assumptions, and the application. For this study, the power parameter \( p \) has been set to \( p = 1.35 \), which turned out to fit best for the dataset and distribution used (details in section 3c).

Furthermore, the response is assumed to be left censored at 0 to account for the nonnegative observations and the large fraction of 0 observations. The concept of left censoring assumes that there is an underlying latent (unobservable) process driving the observable response, which can be described by a linear predictor. While the latent response \( y \) is allowed to become negative, the observable response “precipitation” is simply 0 if the latent response \( y \) is below zero or the inverse power-transformed latent response \( y^p \) otherwise. For simplicity, the zero left-censored nonhomogeneous logistic regression will be denoted as CNLR from now on.

Both distributional parameters (\( \mu, \sigma \)) are expressed by a linear predictor including the covariates or explanatory variables. As suggested by Gneiting et al. (2005), the mean of the ensemble forecast drives the location \( \mu \), and the standard deviation of the ensemble drives the scale \( \sigma \). For this study, we only use the forecasted daily accumulated total precipitation from the ensemble (section 2e) as the meteorological predictor variable. In Eq. (1), \( m \) denotes the mean, and \( s \) denotes the standard deviation of the forecasted power-transformed daily total precipitation amounts of the ensemble members.

Following the idea of Gebetsberger et al. (2016), a second covariate \( z \) has been included. The term \( z \) is a binary split variable, which takes 1 if all forecast members in the training dataset predict less than 0.01 mm day\(^{-1} \) (\( z = 1 \); “no” precipitation) or 0 otherwise. This allows us to handle dry and wet cases differently and has a positive impact on the results. It furthermore solves the problem of taking the logarithm of the ensemble standard deviation if all members predict 0 mm, which leads to \( s = 0 \). The log transformation
on the scale $\sigma$ is used to ensure nonnegative-scale values during optimization. The full CNLR assumptions can then be written as

$$
\text{precipitation} = \begin{cases} 
0 & \text{if } y \leq 0 \\
y^p & \text{else}
\end{cases}.
$$

$$
y \sim \mathcal{D}(\mu, \sigma),
$$

$$
\mu = \beta_0 + \beta_1 \times z + \beta_2 \times m \times (1 - z),
$$

$$
\log(\sigma) = \gamma_0 + \gamma_1 \times \log(s) \times (1 - z). 
$$

(1)

In case of a dry ensemble forecast ($z = 1$), the linear predictors collapse to $\mu = \beta_0 + \beta_1$ and $\log(\sigma) = \gamma_0$ such that the model only consists of two estimated constants describing the climatological distribution of the response conditional on all cases where $z = 1$. The variable $\beta_1$ typically becomes strongly negative, which leads to a strongly negative latent location $\mu$ and overall small expected amounts of precipitation for the case $z = 1$. For wet cases ($z = 0$), the linear predictors become $\mu = \beta_1 + \beta_2 \times m$ and $\log(\sigma) = \gamma_0 + \gamma_1 \times \log(s)$, which corresponds to the NGR model proposed by Gneiting et al. (2005). These assumptions allow us to correct the bias but also a possible overdispersion or underdispersion of the ensemble as the scale $\sigma$ depends on the predicted ensemble standard deviation. Even if the two cases are not independent and connected via the scale part, discontinuities occur at the transition where $z$ goes from 0 to 1. As this only happens in regions with very small predicted amounts of precipitation, the effect on the results is marginal.

The model as specified in Eq. (1) can be applied at every arbitrary location where both historical observations and historical ensemble forecasts are available. For pointwise ensemble postprocessing, one CNLR model has to be fitted at each observation site. In this case, all CNLR models are independent and have their own regression coefficients $\beta_s$ and $\gamma_s$. As these coefficients are site specific, spatial predictions are not directly possible and would require an additional interpolation method, which allows us to account for supplementary covariates, such as terrain or surface properties.

Instead of a two-step approach of performing stationwise estimates and interpolating/extrapolating the resulting coefficients afterward, we extend the model to include the training data of all stations at once and fit one simple and computationally efficient model for fully probabilistic spatial estimates.

b. SAMOS

The statistical method presented in this article is based on the anomaly approach first published by Scheuerer and Büermann (2014) and further extended by Dabernig et al. (2017), focusing on temperature forecasts across Germany and northern Italy, respectively. We extend the SAMOS approach by Dabernig et al. (2017), yielding to a censored SAMOS version for precipitation postprocessing.

Climatological properties between two precipitation observation sites may vary in mean (location) and variability (scale). This is especially true over complex terrain where only a few kilometers between a valley and a mountain station can result in very large climatological differences (Frei and Schär 1998; Isotta et al. 2014; Stauffer et al. 2017). These small-scale features influence daily precipitation sums but are not yet fully resolved by global numerical ENSs. Therefore, a high-resolution spatiotemporal climatology is used as background information to provide small-scale features at any location within the study area. Instead of modeling the relationship between past observations and past numerical weather forecasts directly, the statistical model uses high-resolution standardized anomalies. Anomalies are defined as the short-term deviation from the local long-term climate. These anomalies can be divided by the local climatological variability to obtain standardized anomalies. Standardized anomalies of the observations (precipitation) are defined as

$$
y^* = \frac{\text{precipitation}^{1/p} - \mu_{\text{obs,clim}}}{\sigma_{\text{obs,clim}}},
$$

where $\mu_{\text{obs,clim}}$ and $\sigma_{\text{obs,clim}}$ describe the long-term climatological properties of daily observations and will be discussed in detail in section 3c. The term $y^*$ denotes the resulting latent response on the standardized anomaly scale, which follows a standard logistic distribution $\mathcal{D}(0, 1)$. Equivalent to Eq. (2), standardized anomalies of the ensemble forecasts (ens) can be computed with the climatological properties $\mu_{\text{ens,clim}}$ and $\sigma_{\text{ens,clim}}$ of the ensemble using

$$
\text{ens}^* = \frac{\text{ens}^{1/p} - \mu_{\text{ens,clim}}}{\sigma_{\text{ens,clim}}}. 
$$

The ensemble climatology $(\mu_{\text{ens,clim}}, \sigma_{\text{ens,clim}})$ is described in section 3c.

Because of standardization, the censoring point on the anomaly scale becomes a function of the observed climatology. While the censoring point is on 0 (no precipitation) on the original or power-transformed scale [Eq. (1)], the censoring threshold becomes $-\mu_{\text{obs,clim}}/\sigma_{\text{obs,clim}}$ after standardizing the data. Figure 2a shows the power-transformed observations with a constant censoring threshold of 0 throughout the whole year. Figure 2b shows all standardized anomalies and
the shifted censoring threshold indicated by the solid line. As observations below the censoring threshold never occur, all data points lie on or above this line. Figure 2c is an extension of Fig. 2b, where all observations on the censoring threshold (0 mm day$^{-1}$ on the original scale) were simulated from the standard logistic distribution for visual justification. As shown in the density plot, the standardized anomalies now follow a latent standard logistic distribution $L(0, 1)$. As each of the 117 stations is standardized using its specific climatological properties $\mu_{\text{obs,clim}}$ and $\sigma_{\text{obs,clim}}$, the standardized anomalies of all stations show the same distribution $L(0, 1)$. Thus, the standardization removes site-specific features from the data and brings the data of all stations onto a comparable level.

Combining the CNLR model from Eq. (1) with the concept of standardized anomalies [Eqs. (2) and (3)] leads to the full specification of the SAMOS model with a left-censored logistic response:

$$ y^* \sim L(\mu^*, \sigma^*), $$

$$ \mu^* = \beta_0 + \beta_1 \times z + \beta_2 \times m^* \times (1 - z), $$

$$ \log(\sigma^*) = \gamma_0 + \gamma_1 \times \log(s^*) \times (1 - z). $$

As on the power-transformed scale, the standardized anomalies $y^*$ are still assumed to follow a logistic distribution. The linear predictors for location $\mu^*$ and $\log(\sigma^*)$ on the standardized anomaly scale depend on the standardized ensemble anomalies (ens*) and the binary split indicator $z$. In this study, total precipitation forecasts are used as the only meteorological variable. The covariates $m^*$ and $s^*$ therefore correspond to the empirical mean and standard deviation of the standardized total precipitation forecast anomalies [Eq. (3)].
Once all covariates are known, the regression coefficients of the SAMOS model given by Eqs. (2)–(4) can be estimated using censored maximum likelihood optimization as offered by the R package crch (Messner et al. 2016) or similar software. The climatological estimates required to create the standardized anomalies are explained in detail in section 3c.

Given all regression parameters $\beta_*$ and $\gamma_*$ of the SAMOS model [Eq. (4)], the correction can be applied to future ensemble forecasts. As the SAMOS model returns both parameters on the standardized anomaly scale, they have to be destandardized with respect to the spatial climatology:

$$\mathcal{L}_0 \left( \mu^* \times \sigma_{\text{obs,clim}} + \mu_{\text{obs,clim}} \cdot \sigma^* \times \sigma_{\text{obs,clim}} \right).$$  

(5)

The destandardized zero left-censored distribution $\mathcal{L}_0(\cdot)$ describes the full postprocessed ENS forecast distribution on the power-transformed scale. Since the SAMOS regression coefficients are location independent, the postprocessed predictions can be computed at any location within the study area where both ENS forecasts and climatological estimates ($\mu_{\text{clim}}$, $\sigma_{\text{clim}}$) are available. As spatiotemporal climatologies are used (details in section 3c), the only limitation for the postprocessed ENS forecasts is the horizontal grid spacing of the spatial climatology, which itself only depends on the resolution of the available digital elevation model (see Stauffer et al. 2017). From the full-probabilistic SAMOS forecasts, different properties can then be derived, such as the mean or expectation, quantiles, probability of precipitation, or probabilities exceeding a certain threshold. To retrieve the corrected forecasts on the original scale in millimeters per day, the inverse power transformation has to be taken into account. Details can be found in appendix A.

In the limiting case that the ensemble would not provide any information at all, $\mu^*$ approaches 0 and $\sigma^*$ approaches 1, resulting in $\mu = \mu_{\text{obs,clim}}$ and $\sigma = \sigma_{\text{obs,clim}}$, which corresponds to the underlying high-resolution climatology—the most reliable information available in this case.

c. Climatological estimates

The climatological properties $\mu_{\text{clim}}$ and $\sigma_{\text{clim}}$ for both the observations and the ensemble forecasts have to be specified to be able to derive the standardized anomalies $y^*$ and ens$^*$ [Eqs. (2) and (3)]. The computation of the observed climatology is based on Stauffer et al. (2017) but uses a left-censored logistic instead of Gaussian distribution and consequently a modified power-transformation parameter. As in Stauffer et al. (2017), the optimal power parameter was chosen using a power-adjusted maximum likelihood approach optimizing 117 stationwise climatologies. Since the optimal power parameters did not show a distinct spatial or altitudinal dependency, the median among all 117 estimates was selected using a constant $p = 1.35$ in this study.

The observed spatiotemporal climatology is based on all 117 stations (Fig. 1) and uses daily precipitation measurements from 1971 through the end of 2009, yielding to roughly 1.5 million individual observations. Data from the years 2010–13 are set aside for verification.

The climatology is based on a nonhomogeneous regression model similar to the SAMOS method. In contrast to Eqs. (1) and (4), the linear predictors of the climatological model include smooth one-dimensional and multidimensional spline effects to depict all features of the climatology. In addition to the global intercepts ($\beta, \gamma$), an altitudinal effect ($s_1, t_1$), an effect to describe the seasonality based on the day of the year ($s_2, t_2$), a spatial effect on dependent longitude and latitude ($s_3, t_3$), and a three-dimensional effect to describe spatial variations in the seasonal pattern ($s_4, t_4$) are included. Further details can be found in Stauffer et al. (2017). The full model specification of the observation climatology can be expressed as

$$\text{precipitation} = \begin{cases} 0 & \text{if } y \leq 0, \\ \exp(y) & \text{else,} \end{cases},$$

$$y \sim \mathcal{L}(\mu_{\text{obs,clim}}, \sigma_{\text{obs,clim}}),$$

$$\mu_{\text{obs,clim}} = \beta + s_1(\text{alt}) + s_2(y\text{day}),$$

$$+ t_1(\text{lon}, \text{lat}) + t_2(y\text{day}),$$

$$\log(\sigma_{\text{obs,clim}}) = \gamma + t_3(\text{alt}) + t_4(y\text{day}),$$

$$+ t_5(\text{lon}, \text{lat}) + t_6(y\text{day}, \text{lon}, \text{lat}).$$

(6)

Again, both parameters of the power-transformed left-censored logistic distribution (location $\mu_{\text{obs,clim}}$ and scale $\sigma_{\text{obs,clim}}$) are modeled. This is required as they are used for the standardization of the SAMOS model. Although the climatology model (section 3c) is quite complex, estimation only takes about 30 h and has to be done rarely, for example, once a year.

In addition to climatological estimates of the observations, climatological estimates $\mu_{\text{ens,clim}}$ and $\sigma_{\text{ens,clim}}$ are required to compute standardized anomalies of the ensemble forecasts as in Eq. (3). The two parameters represent the long-term climatology of the ECMWF EPS (section 2c) and are computed from the ECMWF reforecast dataset. The mean and standard deviation are
based on up to 400 individual forecasts provided by the most recent four reforecast runs (section 2d):

\[
\mu_{\text{ens,clim}} = \text{mean(reeforecast)},
\]

\[
\sigma_{\text{ens,clim}} = \text{stdv(reeforecast)} \times C.
\]  

(7)

The climatological location \(\mu_{\text{ens,clim}}\) is simply the empirical mean; the climatological scale \(\sigma_{\text{ens,clim}}\) is the “standard deviation” of the reforecast used. The factor \(C = \sqrt{3/\pi}\) is used to get the empirical scale of a logistic distribution to be on the same scale as the estimated scale of the observation climatology \(\sigma_{\text{obs,clim}}\) [Eq. (6)].

4. Results and verification

a. SAMOS results

Figure 3 shows an example of the climatologies used for 18 May 2010 and the resulting spatial SAMOS predictions. It can be seen in all climatological estimates (Figs. 1a–d) that the altitudinal dependency is the most dominant effect for this day (cf. Fig. 1). The ENS with a horizontal grid spacing of \(\sim 40\) km \(\times 40\) km is only able to resolve the main Alpine ridge leading to the smooth north–south transitions in the left column of Fig. 3. The ensemble climatology correctly shows larger location \(\mu\) (Fig. 3a) and scale \(\sigma\) (Fig. 3c) toward the prealpine flatland to the north and the south; however, this is only a very rough approximation of what is actually observed (Figs. 3b,d).

Figures 3e–h show the predictions for 18 May 2010, when a cold front hit the Alps from the north driven by a strongly pronounced low pressure system east of the study area. As a result, the forecasts show larger precipitation amounts north of the area due to orographic lifting and blocking. As the ENS is only able to represent the topography as one smooth ridge (Fig. 1), the only feature that can be identified in the ENS prediction is a gradual decrease of precipitation from north to south over the main Alpine ridge. In reality, a first mountain ridge alongside the northern boundary of the study area is blocking the air mass. Larger amounts of precipitation are typically observed in southern Germany north of Tyrol, while the well-marked Alpine valleys in Tyrol typically receive less precipitation. This can be seen in the observed climatology (Fig. 3b) but also for this particular day in the corrected SAMOS forecasts (Figs. 3f,h). South of the largest valley with a west–east orientation, increased forecasted amounts and probabilities can be seen in the corrected SAMOS predictions related to a secondary lifting of the air masses at the high mountains close to the main Alpine ridge.

The example shows that SAMOS is able to add interpretable and meaningful features to the ENS during the postprocessing procedure. However, the performance cannot be evaluated with a single case alone. Section 4b therefore contains a detailed analysis and verification on a 3-yr independent dataset.

b. Verification

For verification, the predictions of four different methods will be compared with unused (out of sample) data between February 2010 and December 2012. As two baseline methods, the climatologies (CLIM; section 3c) and the raw total precipitation predictions from the ECMWF ENS will be used. The empirical frequency of the \(50 + 1\) ensemble members is used as probability to compute the Brier scores shown in the results. Furthermore, a stationwise postprocessing (STN) is included based on Eq. (1). For STN, a separate CNLR model is estimated for each of the 117 stations in the dataset.

The predictions of all methods are out of sample such that the data used for verification are not included in the training dataset, which is used to estimate the regression coefficients. CLIM is based on all available observations except that the years 2010–13 are excluded (section 3c). Therefore, CLIM predictions are spatially in sample but temporally out of sample. STN is using the latest four available reforecast runs yielding to spatially in-sample but temporally out-of-sample predictions. SAMOS is the only method whose predictions can be verified both spatially and temporally out of sample. Therefore, a leave-one-out cross validation is performed. For each station, the SAMOS regression coefficients were estimated based on the most recent four reforecast runs excluding this one specific station. Forecasts were then made for the excluded station only. Table 1 contains a summary of all four methods and shows their sample behavior. Full in-sample SAMOS results are omitted as hardly any differences can be seen compared to the cross-validated out-of-sample results.

The continuous ranked probability score (CRPS; appendix B) of all predictions is shown as a continuous ranked probability skill score (CRPSS) in Fig. 4 using CLIM as reference. Values below zero indicate less predictive skill than CLIM. The higher the score, the better the performance of the corresponding method. As the CRPS is a fully probabilistic score, it penalizes for a possible dislocation of the predicted distribution but also for the wrongly predicted width or sharpness. The scores show an overall decrease with increasing forecast horizon for all three methods, slowly approaching the skill of the climatology. The two postprocessing methods STN and SAMOS show a significant
improvement with respect to the ENS up to the 6-day-ahead forecasts. SAMOS outperforms the STN method, even if it is verified fully out of sample. The differences between STN and SAMOS are small but all significant (paired two-sided t test, 5% significance level; not shown).

In addition to the CRPSS, Fig. 5 shows the Brier skill scores (BSSs) for three different thresholds using CLIM as the reference method again. Positive BSSs show that the method has more predictive skill than the reference; values below zero show less skill than CLIM. For threshold 0 mm day$^{-1}$ (precipitation yes/no), it can be
seen that the ENS performs poorly, even worse than the climatology. This is mainly caused by a wet bias in the ENS (not shown), which depends on the design of the ENS predicting an average over a relatively large grid cell. Both postprocessing methods perform significantly better than the climatology. Overall, SAMOS shows the best performance even for long forecast horizons. Figures 5b and 5c show the same verification for 1 and 10 mm day\(^{-1}\), respectively. For these thresholds, ENS is better than CLIM but outperformed by the post-processing methods. For large thresholds (Fig. 5c) and large forecast horizons, all methods become very similar. Differences between them are no longer significant.

As last measure of performance, verification rank histograms and probability integral transform (PIT) histograms are shown in Fig. 6 for the ENS and SAMOS 1-day-ahead and 6-day-ahead forecasts to assess the calibration (Gneiting et al. 2007). In general, a more uniformly distributed histogram shows better calibration. A concave shape indicates that the forecasted distribution is too tight (underdispersive); a convex shape indicates that the distribution is too wide (overdispersive).

The verification rank histogram assesses the calibration of discrete distributions as provided by the 50 + 1 members of the ENS, yielding to 52 possible ranks. For each pair of total precipitation forecasts and observations, the rank is evaluated. Observations falling below the lowest ensemble member forecast are assigned to rank 1; observations falling above the highest ensemble member forecast are assigned to rank 52. All others are assigned to the ranks 2–51 with respect to the ensemble distribution as shown in Figs. 6a and 6c. The pronounced concave shape of the rank histogram indicates a strong underdispersion of the raw ENS such that a large fraction falls into the tails of the distribution or even outside.

The PIT histogram shows a similar measure for probabilistic forecasts. For each observation/forecast pair, the quantile conditional on the observed value is evaluated (\([0.0 – 1.0]\)) and pooled into equidistant bins. For easy comparison with the rank histogram, we have chosen 52 uniformly distributed bins as shown in Figs. 6b and 6d. SAMOS is much better calibrated than the ENS, but the convex shape indicates that the distribution of the SAMOS is slightly wider than what is observed (overdispersive).

5. Discussion and conclusions

In this study, the standardized anomaly model output statistics (SAMOS) model has been extended and

![Fig. 4. CRPSS with climatology from Eq. (6) as reference. (from left to right) The boxes show the model performance for 1-day-ahead to 6-day-ahead forecasts. Each box contains three box-and-whisker plots for the (left) raw ENS, and the two postprocessing methods (middle) STN and (right) SAMOS. Each one contains 117 stationwise-mean skill scores. The boxes show the upper and lower quartile, and the whiskers show the 1.5 interquartile range. Additionally, the median (black bar) and the outliers (circles) are plotted. Values below 0 indicate stations with less skill than the climatology. The higher the values, the better the performance of the method.](image-url)
applied to daily precipitation sums. It has been shown that the concept of using standardized anomalies (Scheuerer and Büermann 2014; Dabernig et al. 2017) can be used to correct precipitation forecasts of numerical ensemble forecast models. The SAMOS post-processing method is able to create accurate spatial predictions of daily precipitation sums over complex terrain. SAMOS uses high-resolution spatial climatologies as background information to transform the data (observations and ensemble forecasts) into standardized anomalies. This (i) removes location-dependent climatological features from the data and (ii) brings all data to a comparable level to account for the small-scale features in the study area, which are not yet resolved by the ensemble model. SAMOS returns fully probabilistic predictions for any arbitrary location within the study area, even for regions without observational sites.

To create the standardized anomalies, daily estimates of the climatological mean (location $\mu_{\text{clim}}$) and variability (scale $\sigma_{\text{clim}}$) are required. The observed climatology estimate is based on the method presented by Stauffer et al. (2017) using a censored logistic rather than censored Gaussian response distribution. The censored logistic distribution has been chosen for this study as the spatiotemporal climatology showed slightly better calibration. Both distributions are very similar except that the logistic distribution has somewhat heavier tails, which is partly compensated by the

**FIG. 5.** BSSs for three different thresholds using climatology from Eq. (6) as reference: (a) 0, (b) 1, and (c) 10 mm day$^{-1}$. The specifications of the box-and-whisker plots are as in Fig. 4. The frequency of the daily total precipitation is used for ENS, whereas the probabilities for the two postprocessing methods STN and SAMOS are derived from the predicted distribution. (from left to right) Scores for 1-day-ahead to 6-day-ahead forecasts. The higher the values, the better the performance of the method.
The results show that the spatial SAMOS outperforms the STNs even if the SAMOS predictions are (unlike STNs) spatially out of sample. This is mainly related to the training dataset. While the STN only includes interpolated forecasts of one location, the SAMOS training dataset includes the data of all stations, leading to more robust estimates. The SAMOS calibration indicates that the assumed response distribution is not optimal. A different distribution might improve the skill and remove the need of the power transformation (Scheuerer 2014; Hamill et al. 2015).

The goal of this study is to use the SAMOS approach proposed by (Dabernig et al. 2017) and to extend the method for the application of precipitation sums or censored responses in general. While only focusing on daily precipitation sums up to day 6 in this study, it would be worthwhile to extend the forecast horizon and the study area but also to include additional covariates and to apply the SAMOS approach to other meteorological variables.
parameters. A further SAMOS extension to account for spatiotemporal correlation structures would be of great interest. Because of the standardization, SAMOS corrects for a possible underprediction or overprediction of the ensemble over long time scales but not on a single event as only the spatial correlation structure of the EPS is considered at this stage.

As the estimation of the SAMOS requires only little computational time, the SAMOS can easily be refitted as soon as a new reforecast run is available. This ensures that the SAMOS automatically adapts itself to the latest ECMWF ensemble model version within a very short transition period. Nowadays, the ECMWF reforecast (ECMWF 2016) is run twice a week, providing 10 + 1 members, which could further improve the performance of the SAMOS but could not have been tested.

Acknowledgments. This research is part of an ongoing project funded by the Austrian Science Fund (FWF), Grant TRP 290. The computational results presented in this article have been achieved in part using the Vienna Scientific Cluster (VSC). The observation dataset was provided by the Tyrol hydrographical service (http://ehyd.gv.at).

APPENDIX A

Properties of the Power-Transformed Left-Censored Logistic Distribution

The probability density function \( \lambda \) and the cumulative distribution function \( \Lambda \) of a noncensored logistic distribution are defined as

\[
\lambda(x | \mu, \sigma) = \frac{\exp\left(\frac{x - \mu}{\sigma}\right)}{\sigma \left[1 + \exp\left(\frac{x - \mu}{\sigma}\right)\right]^2}, \quad (A1)
\]

\[
\Lambda(x | \mu, \sigma) = \frac{1}{1 + \exp\left(\frac{x - \mu}{\sigma}\right)}. \quad (A2)
\]

The density \( \lambda_0 \) and distribution function \( \Lambda_0 \) of a zero left-censored logistic distribution including the power transformation \( 1/p \) can then be written as

\[
\lambda_0(x_i | \mu, \sigma, p) = \begin{cases} 0 & \text{for } x_i < 0, \\ \Lambda(0 | \mu, \sigma) & \text{for } x_i = 0, \\ \lambda(x_i^{1/p} | \mu, \sigma, p) & \text{else} \end{cases} \quad (A3)
\]

\[
\Lambda_0(x_i | \mu, \sigma, p) = \begin{cases} 0 & \text{for all: } x_i < 0, \\ \Lambda(x_i^{1/p} | \mu, \sigma, p) & \text{else} \end{cases}. \quad (A4)
\]

where both are set to zero below the censoring point at 0. For \( x_i^{1/p} \geq 0 \), both follow the density and distribution function of the noncensored logistic distribution (\( \lambda \) and \( \Lambda \) respectively) except that the density \( \lambda_0(x_i = 0 | \mu, \sigma) \) depicts the point mass at the censoring point, which conforms the distribution function evaluated at 0. This also directly specifies the probability that precipitation will be observed at a certain location/time:

\[
P(x_i > 0 | \mu, \sigma) = 1 - P(x_i \leq 0) = 1 - \Lambda(0 | \mu, \sigma). \quad (A5)
\]

The probability of exceeding a certain threshold can be derived for any threshold \( \kappa \geq 0 \):

\[
P(x_i > \kappa | \mu, \sigma, p) = 1 - P(x_i \leq \kappa) = 1 - \Lambda(x^{1/p} \geq \kappa | \mu, \sigma). \quad (A6)
\]

Furthermore, the expectation of the distribution on the original scale has to be evaluated. The expectation on the original scale \( x \) in millimeters per day can be retrieved using

\[
E[x | \mu, \sigma, p] = \int_0^\infty \lambda(x^{1/p} | \mu, \sigma) \frac{x^{(1/p)-1}}{p} \, dx. \quad (A7)
\]

A last property of interest is the median of the distribution, again on the original scale \( x \) in mm \( \text{day}^{-1} \). Parameter \( \mu \) in Eqs. (1) and (4) describes the latent unobservable location. The median is then given as

\[
\text{median}(x | \mu, p) = \begin{cases} 0 & \text{for } \mu \leq 0, \\ \mu \text{ else} \end{cases}. \quad (A8)
\]

APPENDIX B

Error Measures Used for Verification

As a fully probabilistic score, the CRPS is shown in the verification section of this article. The mean CRPS of a zero left-censored power-transformed logistic distribution (see appendix A) can be written as

\[
\text{CRPS} = \frac{1}{N} \sum_{i=1}^{N} \int_0^\infty \left[ \Lambda_0(x_{i}^{1/p} | \mu, \sigma) - H(x) \right]^2 \, dx, \quad (B1)
\]

where \( x \) is the response variable on the original scale in millimeters per day, \( N \) is the number of forecasts included, \( \Lambda_0 \) is the CDF of the forecasted distribution [Eq. (A4)], and \( H \) is the CDF of the observation represented by a Heaviside step function, which takes 0 for all \( x < \) observations and 1 otherwise. While \( x \) is on the original scale \( \text{mm \text{day}^{-1}} \), both distributional
parameters, location $\mu$ and scale $\sigma$, are on the power-transformed scale. Therefore, the power transformation $1/p$ is required to evaluate the CDF $\Lambda$. As no analytic solution has been found, the CRPS is evaluated by quantile sampling with $n = 2000$.

The CRPS is shown as a skill score (CRPSS) in this article. A skill score shows the performance against a reference method. As the CRPS can only take non-negative values, the CRPSS can be written as

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{\text{ref}}},$$  \tag{B2}

where the CRPS of the method to test is in the numerator, and the CRPS of the reference method is in the denominator. Values below 0 indicate that the tested method performs worse than the reference. CPRSS values can take values in the range of $[-\infty, 1]$. As a second measure, the BSS is shown to verify the skill of the forecast probabilities. One of the most frequently used thresholds for precipitation forecasts is 0 mm, also known as the probability of precipitation. As an EPS does not provide fully probabilistic forecasts, the frequency of the daily total precipitation sum is used as an estimator of the probability. As an example, if half of all ensemble members predict no precipitation, the frequency is 0.5 and can be seen as a probability of $\sim 50\%$ if the number of ensemble members is sufficiently large. The Brier score can then be written as

$$\text{BS}(\kappa) = \frac{1}{N} \sum_{i=1}^{N} \left[ P(x_i > \kappa | \mu_i, \sigma_i, p) - o_i(t) \right]^2,$$  \tag{B3}

where $N$ is again the number of forecasts included, $P_i$ the predicted probability that an event exceeds threshold $\kappa$ [Eq. (A6)], and $o_i$ the binary observation which takes 0 for all observations $x_i \leq \kappa$ and 1 otherwise. Correspondingly, the Brier skill score is defined as

$$\text{BSS} = 1 - \frac{\text{BSS}}{\text{BSS}_{\text{ref}}},$$  \tag{B4}

REFERENCES

Ben Bouallégue, Z., and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. *Meteor. Appl.*, 21, 922–929, doi:10.1002/met.1435.

Box, G. E. P., and D. R. Cox, 1964: An analysis of transformations. *J. Roy. Stat. Soc.*, 26B, 211–252.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, 133, 1076–1097, doi:10.1175/MWR2905.1.

Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft, 2016: Abteilung IV—Wasserhaushalt. Accessed 29 February 2016. [Available online at http://ehyd.gv.at.]

Dabernig, M., G. J. Mayr, J. W. Messner, and A. Zeileis, 2017: Spatial ensemble post-processing with standardized anomalies. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2975, in press.

ECMWF, 2016: Re-forecast for medium and extended forecast range. ECMWF, accessed 9 June 2016. [Available online at http://www.ecmwf.int/en/forecasts/documentation-and-support/re-forecast-medium-and-extended-forecast-range.]

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, 21, 739–759, doi:10.1111/j.2153-3490.1969.tb04853.x.

Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, 138, 190–202, doi:10.1175/2009MWR3046.1.

Frei, C., and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, 18, 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.

Gebetsberger, M., J. W. Messner, G. J. Mayr, and A. Zeileis, 2016: Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions. University of Innsbruck Working Papers in Economics and Statistics 2016-28, 25 pp. [Available online at http://EconPapers.repec.org/RePEc:inn:wpaper:2016-28.]

Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 133, 1098–1118, doi:10.1175/MWR2904.1.

——, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecast calibration and sharpness. *J. Roy. Stat. Soc.*, 69B, 243–268, doi:10.1111/j.1467-9868.2007.00587.x.

Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 138, 1814–1827, doi:10.1002/qj.1895.

Hamill, T. M., 2012: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Mon. Wea. Rev.*, 140, 2232–2252, doi:10.1175/MWR-D-11-00220.1.

——, J. S. Whitaker, and S. L. Mullen, 2006: Reforecast: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, 87, 33–46, doi:10.1175/BAMS-87-1-33.

——, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, 136, 2620–2632, doi:10.1175/2007MWR2411.1.

——, M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS reforecasts and climatology-calibrated precipitation analyses. *Mon. Wea. Rev.*, 143, 3300–3309, doi:10.1175/MWR-D-15-0004.1.

Hutchinson, M. F., 1998: Interpolation of rainfall data with thin plate smoothing splines—Part I: Two dimensional smoothing of data with short range correlation. *J. Geogr. Inf. Decis. Anal.*, 2, 168–185.

Isotta, F. A., and Coauthors, 2014: The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *Int. J. Climatol.*, 34, 1657–1675, doi:10.1002/joc.3794.
Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara, 2008: SRTM 90m digital elevation database, version 4.1. Consultative Group on International Agricultural Research Consortium for Spatial Information, accessed 29 February 2016. [Available online at http://srtm.csi.cgiar.org.]

Kaiser, M., and Coauthors, 2014: Statistisches Handbuch Bundesland Tirol 2014. Land Tirol Rep., 422 pp. [Available online at https://www.tirol.gv.at/fileadmin/themen/statistik-budget/statistik/downloads/Statistisches_Handbuch_2014.pdf.]

Lerch, S., and T. L. Thorarinsdottir, 2013: Comparison of non-homogeneous regression models for probabilistic wind speed forecasting. Tellus, 65, 21206, doi:10.3402/tellusa.v65i0.21206.

Messner, J. W., G. J. Mayr, D. S. Wilks, and A. Zeileis, 2014a: Extending extended logistic regression: Extended versus separate versus ordered versus censored. Mon. Wea. Rev., 142, 3003–3014, doi:10.1175/MWR-D-13-00355.1.

——, ——, A. Zeileis, and D. S. Wilks, 2014b: Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. Mon. Wea. Rev., 142, 448–456, doi:10.1175/MWR-D-13-00271.1.

——, ——, and ——, 2016: Heteroscedastic censored and truncated regression with crch. R J., 8, 173–181. [Available online at https://journal.r-project.org/archive/2016-1/messner-mayr-zeileis.pdf.]

Mullen, S. L., and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. Mon. Wea. Rev., 129, 638–663, doi:10.1175/1520-0493(2001)129<0638:QPFOTU>2.0.CO;2.

Roulston, M. S., and L. A. Smith, 2003: Combining dynamical and statistical ensembles. Tellus, 55, 16–30, doi:10.1034/j.1600-0870.2003.201378.x.

Scheuerer, M., 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. Quart. J. Roy. Meteor. Soc., 140, 1086–1096, doi:10.1002/qj.2183.

——, and L. Büermann, 2014: Spatially adaptive post-processing of ensemble forecasts for temperature. J. Roy. Stat. Soc., 63C, 405–422, doi:10.1111/rssc.12040.

——, and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. Mon. Wea. Rev., 143, 4578–4596, doi:10.1175/MWR-D-15-0061.1.

Sloughter, J. M. L., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Mon. Wea. Rev., 135, 3209–3220, doi:10.1175/MWR3441.1.

Statistik Austria, 2016: Bevölkerung. Accessed 22 June 2016. [Available online at https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/index.html.]

Stauffer, R., G. J. Mayr, J. W. Messner, N. Umlauf, and A. Zeileis, 2017: Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. Int. J. Climatol., doi:10.1002/joc.4913, in press.

Stidd, C. K., 1973: Estimating the precipitation climate. Water Resour. Res., 9, 1235–1241, doi:10.1029/WR009i005p01235.

Thorarinsdottir, T. L., and T. Gneiting, 2010: Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. J. Roy. Stat. Soc., 173A, 371–388, doi:10.1111/j.1467-985X.2009.00616.x.

Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteor. Appl., 16, 361–368, doi:10.1002/met.134.