Adv. Stat. Clim. Meteorol. Oceanogr., 4, 65–86, 2018
https://doi.org/10.5194/ascmo-4-65-2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

ASCMO
Open Access

Hourly probabilistic snow forecasts over complex terrain: a hybrid ensemble postprocessing approach

Reto Stauffer[1], Georg J. Mayr[2], Jakob W. Messner[3], and Achim Zeileis[1]

[1]Department of Statistics, Faculty of Economics and Statistics, Universität Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria
[2]Institute of Atmospheric and Cryospheric Sciences, Faculty of Geo- and Atmospheric Sciences, Universität Innsbruck, Innrain 52, 6020 Innsbruck, Austria
[3]Department of Electrical Engineering, Technical University of Denmark, Elektrovej, Building 325, 2800 Kgs. Lyngby, Denmark

Correspondence: Reto Stauffer (reto.stauffer@uibk.ac.at)

Received: 27 March 2018 – Revised: 24 August 2018 – Accepted: 12 October 2018 – Published: 14 December 2018

Abstract. Accurate and high-resolution snowfall and fresh snow forecasts are important for a range of economic sectors as well as for the safety of people and infrastructure, especially in mountainous regions. In this article a new hybrid statistical postprocessing method is proposed, which combines standardized anomaly model output statistics (SAMOS) with ensemble copula coupling (ECC) and a novel re-weighting scheme to produce spatially and temporally high-resolution probabilistic snow forecasts. Ensemble forecasts and hindcasts of the European Centre for Medium-Range Weather Forecasts (ECMWF) serve as input for the statistical postprocessing method, while measurements from two different networks provide the required observations.

This new approach is applied to a region with very complex topography in the eastern European Alps. The results demonstrate that the new hybrid method allows one not only to provide reliable high-resolution forecasts, but also to combine different data sources with different temporal resolutions to create hourly probabilistic and physically consistent predictions.

1 Introduction

Large parts of our daily social and economic life strongly rely on weather forecasts. In this article we focus on the governmental area of Tyrol, Austria, which is located in the eastern Alps and consists of a large number of narrow valleys surrounded by high mountains. The economic backbone of Tyrol is tourism with more than 5.3 million visitors and more than 25 million overnight stays recorded during the winter season 2013/14 (Amt der Tiroler Landesregierung, 2014). In winter tourism strongly focuses on Alpine outdoor sports such as skiing and back-country skiing, for which resorts and skiing areas need sufficient amounts of snow and good snow conditions. On the other hand, the "white gold" can also cause hazardous situations. During the winter seasons 2009–2016 145 people died in avalanche accidents in Austria (Lawinenwarndienst Tirol, 2009–2017), of which more than half of all events and deaths occurred in Tyrol. Furthermore, severe snow events can obstruct traffic on roads, on train tracks and at airports. Accurate and reliable forecasts of fresh snow and snowfall for the region of Tyrol are therefore of high importance for the public and also for decision makers or warning services (see, e.g., Zhu et al., 2002; Palmer, 2002; Neal et al., 2014; Knox et al., 2015; Raftery, 2016).

Weather forecasts are typically provided by numerical weather prediction (NWP) models predicting the future atmospheric state on a global or regional scale. Due to different influencing factors such as the model resolution, necessary approximations and parameterizations but also imperfect initial conditions and the chaotic behavior of the atmosphere, these forecasts are never fully exact. Ensemble prediction systems (EPSs) address these issues by running sev-

eral independent forecasts for the same day using different and slightly perturbed initial conditions and model formulations to provide valuable additional information about the uncertainty of a specific weather forecast. Due to the spatial discretization of the underlying NWP model the EPS can only depict information on a grid-scale level and is not able to provide reliable information on the point scale. Thus, EPS forecasts typically show too little spread (Hagedorn et al., 2012; Mullen and Buizza, 2001) and require additional correction of the EPS uncertainty to enhance the predictive skill for specific locations. One widely accepted procedure to reduce possible forecast errors and to adjust the uncertainty information is statistical ensemble postprocessing. Statistical postprocessing methods use historical weather forecasts and the corresponding observations to detect and correct possible systematic EPS errors.

A wide range of different ensemble postprocessing methods have been proposed, including analog methods (Hamill et al., 2006, 2015), ensemble dressing methods (Roulston and Smith, 2003), extended logistic regression (Wilks, 2009; Bouallègue and Theis, 2014; Messner et al., 2014b), a non-homogeneous mixture model approach with similarities to Bayesian model averaging (BMA; Sloughter et al., 2007; Fraley et al., 2010), or distributional regression methods. Distributional regression models optimize the parameters of a pre-specified response distribution to correct for both errors in the mean and errors in the uncertainty, given a set of covariates. One of the earliest and most well-known approaches is the ensemble model output statistics (EMOS) approach first published by Gneiting et al. (2005) and applied to near-surface temperature. This approach has further been extended by Thorarinsdottir and Gneiting (2010), Lerch and Thorarinsdottir (2013), Scheuerer (2014), Scheuerer and Hamill (2015), Messner et al. (2014a), Scheuerer (2014), Scheuerer and Hamill (2015) and many others for different meteorological quantities using different response distributions and optimization approaches.

Originally, distributional regression was only applied to specific locations, but has also been extended for spatial and even spatio-temporal corrections of the ensemble forecasts. Many of these extensions are based on anomalies (Scheuerer and Büermann, 2014) or standardized anomalies (Dabernig et al., 2017; Stauffer et al., 2017b) to account for location-specific characteristics in mean and variance and create corrected and fully probabilistic spatial predictions of temperature and daily precipitation sums over potentially complex terrain.

In terms of snow prediction several difficulties have to be considered. The availability and quality of good and reliable snow observations are sparse, even in the region of Tyrol. Measuring snow can be tricky due to possible snow drift, melting processes, or liquid water input (rain) between two observation times, which can yield large measurement errors (Rasmussen et al., 2012). Overall, the amount and quality of snow measurements make it very difficult to train reliable spatial postprocessing models.

An alternative approach to predict fresh snow amounts is to make use of precipitation and temperature forecasts rather than directly to predict snow. The postprocessed temperature and precipitation forecasts can then be used as a proxy to retrieve fresh snow amounts and snowfall forecasts. The temperature forecasts are on the one hand required to determine whether precipitation reaches the ground as rain or snow and on the other hand to estimate the snow density. Snow density and its alteration are affected by the prevalence of inversions, additional cooling effects due to melting and evaporation of hydrometeors, and other local effects, and are thus an extremely complex issue itself. For simplicity we will only regard the problem of whether precipitation occurs as snow or rain and assume that precipitation will fall as snow as soon as the 2 m dry air temperature falls below $+1.2\,^{\circ}\mathrm{C}$, a threshold used in the literature for the European Alps (Rohregger, 2008; Bellaire et al., 2011).

Major challenges of converting probabilistic precipitation and temperature forecasts into fresh snow predictions are the very different temporal resolutions of ensemble predictions, temperature observations, and precipitation observations. European Centre for Medium-Range Weather Forecast (ECMWF) hindcast and EPS forecasts, which we use in this study, have a temporal resolution of 6 and 1 h, respectively, temperature observations are usually available hourly, and precipitation or snow heights are often only measured once or a few times a day.

In this article we propose a new hybrid approach that combines standardized anomaly model output statistics (SAMOS; Dabernig et al., 2017; Stauffer et al., 2017b) with ensemble copula coupling (ECC; Schefzik et al., 2013) and a novel re-weighting scheme to combine these data to

i. create full probabilistic spatial predictions,

ii. provide probabilistic temperature and precipitation forecasts on an hourly temporal scale, and

iii. create a physically consistent copula (pair of temperature and precipitation) which can be used to

iv. create spatially and temporally high-resolution snowfall and fresh snow amount forecasts.

The structure of this article is as follows. Section 2 introduces the different statistical methods required to achieve the desired goal. The methods section is followed by the description of the different data sets used in this study (Sect. 3) and the explicit specification of the statistical models (Sect. 4) used in the results section (Sect. 5). At the end the results and limitations of this approach will be discussed (Sect. 6).

2 Methods

This section contains the three methodological blocks required to create probabilistic snow forecasts. Distributional regression is explained in Sect. 2.1 followed by the required extensions for the SAMOS in Sect. 2.2. Section 2.3 shows the ensemble copula coupling (ECC) approach to generate a postprocessed ensemble followed by the re-weighting procedure in Sect. 2.4 which is required to transform daily precipitation sums into hourly predictions. Finally, hourly temperature and precipitation sums will be converted into probabilities of snowfall and fresh snow amounts in Sect. 2.5.

2.1 Distributional regression

Statistical methods considering all parameters of a specific response distribution can be summarized as "distributional regression models". The EMOS for temperature using a normal response distribution as originally suggested by Gneiting et al. (2005) can be seen as a classical example of this family.

Imagine a time series of 2 m temperature observations $\boldsymbol{y} = \{y_i\}_{i=1,\ldots,N}$ at a specific site and the corresponding ensemble forecasts of the 2 m temperature from the EPS $\mathbf{x} = \{x_{\text{im}}\}_{i=1,\ldots,N}^{m=1,\ldots,M}$ where N denotes the total sample size of the data set and M the number of ensemble members. x_{im} is the individual 2 m temperature prediction of the NWP for date/time i of member m. The EMOS, which slightly differs from the original EMOS as proposed by Gneiting et al. (2005), is specified as

$$
\begin{aligned}
y_i &\sim \mathcal{N}(\mu_i, \sigma_i), && (1)\\
\mu_i &= \beta_0 + \beta_1 \cdot \overline{\boldsymbol{x}}_i, && (2)\\
\log(\sigma_i) &= \gamma_0 + \gamma_1 \cdot \langle \boldsymbol{x}_i \rangle. && (3)
\end{aligned}
$$

The response y_i is assumed to follow a normal distribution $\mathcal{N}$ with the two distributional parameters μ_i (location or mean) and σ_i (scale or standard deviation). Both parameters are expressed by a linear predictor including an intercept (β_0/γ_0) and a slope coefficient (β_1/γ_1) for a covariate. While the location μ_i depends on the ensemble mean $\overline{\boldsymbol{x}}_i$ over all members $m = 1, \ldots, M$ for each individual sample i, the log scale depends on the logarithm of the corresponding ensemble standard deviation denoted as $\langle \boldsymbol{x}_i \rangle$. The log link on σ_i ensures positive variance in predictions.

The coefficients $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_0, \gamma_1)$ can be estimated by using an appropriate M estimator such as the maximum-likelihood estimator maximizing the likelihood:

$$
\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmax}} \left(\prod_{i=1}^{N} \phi\left(\frac{y_i - \mu_i}{\sigma_i} \right) \right), \qquad (4)
$$

where $\phi\left(\frac{y_i-\mu_i}{\sigma_i}\right)$ denotes the standard normal probability density function (PDF) evaluated at each individual $i = 1, \ldots, N$ in the data set.

For the daily precipitation sums the model shown in Eqs. (1)–(3) can be improved by replacing the response distribution and adding an additional covariate z which allows one to account for EPS forecasts where the majority of all EPS members predict no precipitation. Following the work of Gebetsberger et al. (2017) and Stauffer et al. (2017a), the model specification can be written as follows:

$$
\begin{aligned}
y_i^{1/p} &= \mathcal{L}_0(\mu_i, \sigma_i), && (5)\\
\mu_i &= \beta_0 + \beta_1 \cdot \overline{\boldsymbol{x}_i^{1/p}} \cdot (1 - z_i) + \beta_2 \cdot z_i, && (6)\\
\log(\sigma_i) &= \gamma_0 + \gamma_1 \cdot \langle \boldsymbol{x}_i^{1/p} \rangle \cdot (1 - z_i). && (7)
\end{aligned}
$$

The power-transformed observations y_i are assumed to follow a left-censored logistic distribution $\mathcal{L}_0$ censored at 0 and a power parameter $p = 1.35$. The additional covariate z_i takes 1 if 80 % or more of all ensemble members predict less than 0.05 mm over 24 h and 0 otherwise and is used to handle unanimous predictions (cf. Gebetsberger et al., 2017). The corresponding M estimator can be written as

$$
\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmax}} \left(\prod_{i=1}^{N} f\left(\frac{y_i^{1/p} - \mu_i}{\sigma_i} \right) \right)
$$

$$
\text{with } f = \begin{cases} \Lambda\left(\frac{-\mu_i}{\sigma_i}\right) & \text{if } y_i = 0 \\ \lambda\left(\frac{y_i^{1/p}-\mu_i}{\sigma_i}\right) & \text{else} \end{cases}, \qquad (8)
$$

where λ is the PDF and Λ the cumulative distribution function (CDF) of the standard logistic distribution.

2.2 SAMOS

While the model specifications in Eqs. (1)–(3) and (5)–(7) work well for single stations, an extension is required for spatial and/or spatio-temporal ensemble postprocessing. In the following, we will employ the SAMOS approach (Dabernig et al., 2017; Stauffer et al., 2017b) for this purpose. Its basic idea is to remove location- and time-specific characteristics from the observation and EPS data by transforming them into standardized anomalies. This transformation then allows one to fit a single postprocessing model that is valid for the whole area and all season and can thus be applied to any location and time.

Standardized anomalies of the observations ($\boldsymbol{y}^*$) and EPS forecasts ($\mathbf{x}^*$, for each member $m \in M$) will be characterized by a superscript asterisk from here on and are defined as

$$
y_i^* = \frac{y_i - \widetilde{\mu}_{y,i}}{\widetilde{\sigma}_{y,i}} \quad \text{and} \quad x_{\text{im}}^* = \frac{x_{\text{im}} - \widetilde{\mu}_{x,i}}{\widetilde{\sigma}_{x,i}}. \qquad (9)
$$

$\widetilde{\mu}_{\bullet,i}$ and $\widetilde{\sigma}_{\bullet,i}$ are the estimates of the climatological location and scale for each required quantity and depend on the location and season respectively. A comprehensive description of how these climatologies are specified can be found

in Appendix A. Once the climatologies and thus the standardized anomalies are known, the SAMOS regression coefficients can be estimated using Eqs. (1)–(8) by simply replacing y_i and $\boldsymbol{x}_i$ by their corresponding standardized anomalies y_i^* and $\boldsymbol{x}_i^*$ (except in the condition in Eq. (8) where $y_i = 0$ is not replaced). Given a new EPS forecast, the postprocessed predictions can be obtained by applying the SAMOS correction. As the regression coefficients $\hat{\boldsymbol{\theta}}$ are time and location independent, the correction can be performed on the EPS grid scale. Spatial predictions can be retrieved by bilinearly interpolating the resulting location ($\boldsymbol{\mu}^*$) and scale ($\boldsymbol{\sigma}^*$) parameters to the desired spatial resolution and transforming the results back to the original scale (e.g., °C or mm). Algorithm 1 contains the pseudo-code for the SAMOS procedure as used for this article.

Algorithm 1 SAMOS postprocessing procedure. Detailed description and graphical representation in Appendix A.

1. Compute standardized anomalies of observations:

   Input. Observations $\boldsymbol{y}$ (temperature: hourly, precipitation 24-hourly; Sect. 3.3). Each y_i is an observation at a given station and time point.

   i. Estimate spatio-temporal climatology $\tilde{\mu}_{y,i}$, $\tilde{\sigma}_{y,i}$ for response $\boldsymbol{y}$ including seasonal and station location characteristics. Separate climatologies are estimated for temperature and precipitation (Appendix A). *Note.* This allows to obtain climatological parameters $\tilde{\mu}_{y,i}$, $\tilde{\sigma}_{y,i}$ not only at observed locations/time points i but also at new locations/time points j.

   ii. Compute standardized anomalies $y_i^* = \left(y_i - \tilde{\mu}_{y,i}\right)/\tilde{\sigma}_{y,i}$ separate for temperature and precipitation (Sect. 2.2).

   Output. Spatio-temporal climatologies and standardized anomalies y_i^* of the observations at station level.

2. Calculate standardized anomalies of NWP forecasts:

   Input. Gridded hindcasts (temperature: 6-hourly, precipitation: 24-hourly; Sect. 3.2) separately for each required covariate $\boldsymbol{x} \in \mathbf{X}$. Each $\boldsymbol{x}_\mathrm{g}$ is one hindcast at a given grid point for a specific time and forecast lead time with $M = 10 + 1$ members.

   i. Estimate gridded model climatologies $\tilde{\mu}_{x,\mathrm{g}}$, $\tilde{\sigma}_{x,\mathrm{g}}$ separately for each $\boldsymbol{x} \in \mathbf{X}$ (Appendix A).

   ii. Compute standardized anomalies $x^*_\mathrm{gm} = \left(x_\mathrm{gm} - \tilde{\mu}_{x,\mathrm{g}}\right)/\tilde{\sigma}_{x,g}$ for all required covariates $\boldsymbol{x} \in \mathbf{X}$ for each $m \in M$ (Sect. 2.2).

   iii. Bilinearly interpolate standardized anomalies x^*_gm to all station locations to obtain x^*_im.

   Output. Model climatologies and standardized anomalies x^*_im at station level (temperature: 6-hourly, precipitation: 24-hourly).

3. Estimate SAMOS models:

   Inputs. Standardized anomalies (y_i^*, x^*_im) from Steps 1 and 2 (temperature: 6-hourly, precipitation 24-hourly). As y_i^* and $\boldsymbol{x}_i^*$ are no longer location and season dependent: pool all data (space and time) into one combined training data set (separately for temperature and precipitation).

   Estimate the statistical models $\boldsymbol{y}^* \sim \mathcal{D}(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$ to get the required regression coefficients $\hat{\boldsymbol{\theta}}$, separate models for temperature and precipitation (Sects. 2.1 and 2.2).

   Output. Two sets of estimated regression coefficients $\hat{\boldsymbol{\theta}}$ for temperature and precipitation postprocessing, respectively.

4. Predict temperature and precipitation given a new NWP forecast:

   Inputs. Gridded EPS forecasts (Sect. 3.1), observation climatologies (Step 1i) and gridded model climatologies (Step 2i).

   i. Compute standardized anomalies at grid level for covariates $\boldsymbol{x} \in \mathbf{X}$ of each member $m \in M$ of the new EPS forecast with respect to model climatology: $x^*_\mathrm{gm} = \left(x_\mathrm{gm} - \tilde{\mu}_{x,\mathrm{g}}\right)/\tilde{\sigma}_{x,\mathrm{g}}$.

   ii. Correct forecast anomalies for each $g \in G$ using the estimated coefficients $\hat{\boldsymbol{\theta}}$ from Step 3 to get μ^*_g and σ^*_g

   iii. Interpolate parameters (μ^*_g, σ^*_g) of the postprocessed standardized anomalies to obtain μ_j^* and σ_j^* where each j corresponds to a given (arbitrary) location within the study area and a specific time point and forecast lead time.

   iv. Transform corrected μ_j^* and σ_j^* to physical scale: $\hat{y}_j \sim \mathcal{D}\left(\mu_j^* \cdot \tilde{\sigma}_{y,j} + \tilde{\mu}_{y,j},\ \sigma_j^* \cdot \tilde{\sigma}_{y,j}\right) = \mathcal{D}\left(\hat{\mu}_j,\ \hat{\sigma}_j\right)$

   Output. Postprocessed full parametric predictions at each target location (temperature: hourly, precipitation: 24-hourly; Sect. 5.6).

2.3 Ensemble copula coupling

The SAMOS procedure (Sect. 2.2) provides postprocessed probabilistic predictions for 2 m temperature as well as corrected probabilistic forecasts for 24 h precipitation sums. Due to the model specification, SAMOS allows one to retrieve predictions for any arbitrary location within the area of interest (spatial prediction) and even for all forecast steps covered by the training data set (temporal predictions) with one set of regression coefficients. This allows one to create forecasts for $+30/+54/+78$ h for the 24 h precipitation sums, and hourly forecasts for 2 m temperature for the whole study area.

In order to retrieve probabilistic snowfall forecasts from the SAMOS predictions, the marginal predictive distributions of temperature and precipitation have to be combined such that correlations between them are considered. This can be achieved by using ensemble copula coupling (ECC) proposed by Schefzik et al. (2013). The basic idea is to restore the physical coupling between two or more quantities based on the rank order structure of the raw EPS. As numerical predictions are based on physically consistent prognos-

tic equations, each EPS member provides a distinct physically meaningful combination of temperature and precipitation. This property is lost during the SAMOS postprocessing since both quantities are corrected independently. However, the coupling can be restored by drawing a sample of the postprocessed predictive distributions and rearranging the sampled values in the rank order structure of the original EPS forecasts. ECC is applied to each target location individually to restore the spatial correlation structure of the EPS.

There are different ways to draw a new sample from the postprocessed distributions. It turned out (not shown) that the quantile mapping approach with equidistant probabilities (ECC-Q; Schefzik et al., 2013) yields the best and most stable results for this application, which supports the findings of Schefzik et al. (2013). For ECC-Q, a set of $M = 50 + 1$ ensemble members is drawn from the postprocessed distribution based on equidistant probabilities. In the case of the 2 m temperature SAMOS returns hourly estimates for location $\hat{\mu}_j$ and scale $\hat{\sigma}_j$ of a Gaussian distribution (Eq. 3; Algorithm 1 step 4iv). Using the inverse Gaussian CDF $\Phi^{-1}(\boldsymbol{\pi} \mid \hat{\mu}_j, \hat{\sigma}_j)$ with equidistant probabilities $\boldsymbol{\pi} = \frac{1}{M+1}, \ldots, \frac{M}{M+1}$ a new $50 + 1$-member temperature ensemble can be retrieved from the postprocessed distribution.

The very same can be done for the daily precipitation sums using the inverse distribution function of the power-transformed left-censored logistic distribution:

$$\Lambda_0^{-1}\left(\boldsymbol{\pi} \mid \hat{\mu}_j, \hat{\sigma}_j, p\right) = \max\left(0,\ \Lambda^{-1}(\boldsymbol{\pi} \mid \hat{\mu}_j, \hat{\sigma}_j)\right)^p, \qquad (10)$$

where Λ^{-1} is the inverse CDF of the uncensored logistic distribution. Due to the left-censoring at 0, some of the M quantiles can fall on the censoring point, with an increasing number of 0s with decreasing location $\hat{\mu}_j$ and vice versa. For situations where precipitation is very unlikely $\hat{\mu}_j$ might be highly negative, which yields a postprocessed ensemble where all M members predict exactly $0\,\mathrm{mm}\,24\,\mathrm{h}^{-1}$. However, there is still the problem that our two quantities are not available on the same temporal scale. To be able to restore the full EPS rank order structure on an hourly temporal resolution the postprocessed daily precipitation sums first have to be downscaled to an hourly interval.

2.4 Precipitation re-weighting

Temperature and precipitation observation data are based on two different observational networks with different temporal resolutions. The 2 m temperature observations are available hourly, while precipitation sums are only reported once a day (details in Sect. 3.3). This temporal resolution is maintained by the SAMOS postprocessing so that it also differs for the forecasts of the different quantities.

As temperature shows a clear diurnal cycle, it is crucial to know at which time of day precipitation is expected to be observed, as the timing can highly affect the precipitation phase and thus the total fresh snow amount. Therefore, the precipitation forecasts have to be temporally downscaled before they can be combined with the temperature forecasts. For this purpose, we extend ECC (Sect. 2.3) with a novel re-weighting scheme where the daily precipitation sums are allocated to the hours of the day according to the time series of the raw EPS predictions. For example, if an EPS member predicts 10 % of its daily precipitation to fall between 10:00 and 11:00, 10 % of the corresponding precipitation forecast is allocated to this hour. This allows one to downscale each of the $M = 50 + 1$ draws from the marginal precipitation to an hourly temporal resolution and to combine the hourly precipitation predictions afterwards with the respective draws from the marginal temperature distribution. Algorithm 2 shows the temporal downscaling procedure to generate hourly precipitation copulas from the postprocessed daily precipitation sums.

Algorithm 2 Re-weighting pseudo-code for temporal downscaling of probabilistic precipitation forecasts to generate a new $50 + 1$ member copula with an hourly temporal resolution from postprocessed probabilistic daily precipitation sum forecasts provided by SAMOS.

Inputs. Gridded EPS forecasts of 24 h accumulated precipitation sums and postprocessed probabilistic 24 h precipitation sums returned by SAMOS. Index j denotes a specific target location, season, and forecast lead time.

1. Bilinearely interpolate forecasted 24 h precipitation sums of each of the $50 + 1$ EPS members to target location $j \in J$ to receive $(\mathrm{tp}_{j1}, \ldots, \mathrm{tp}_{jm}, \ldots, \mathrm{tp}_{jM})$.
2. Draw a copula $(\hat{y}_{j1}, \ldots, \hat{y}_{jm}, \ldots, \hat{y}_{jM})$ of 24 h postprocessed precipitation sums using ECC-Q drawing from the full probabilistic predictive distribution $\mathcal{L}_0(\hat{\mu}_j, \hat{\sigma}_j)^{1.35}$ returned by SAMOS (Sect. 2.2).
3. Compute correction weighting factors $\boldsymbol{\omega}_j = (\hat{y}_{j1}/\mathrm{tp}_{j1}, \ldots, \hat{y}_{jm}/\mathrm{tp}_{jm}, \ldots, \hat{y}_{jM}/\mathrm{tp}_{jM})$.
4. Correct hourly EPS time series forecasts of each member using the weights $\boldsymbol{\omega}_j$ such that the sum over 24 1-hourly precipitation sums of a specific copula member m is equal to $\hat{y}_{jm}$ $(= \omega_{jm} \cdot \mathrm{tp}_{jm})$.

Output. A $50 + 1$ member postprocessed ensemble with hourly precipitation sums for each target location.

For stability reasons, the weights $\boldsymbol{\omega}$ are computed using values for $\hat{y}_{jm}$ and tp_{jm} rounded to two digits $(\frac{1}{100}\,\mathrm{mm\,d}^{-1})$ to avoid weights close to infinity. If $\hat{y}_{jm}$ or tp_{jm} is 0, the corresponding weight is set to 0 as well. After re-weighting, the precipitation forecasts are at the very same temporal resolution as the temperature forecasts and the rank order structure can be restored with respect to the underlying EPS (Sect. 2.3). This procedure is repeated for each target location, e.g., on a regular grid with a much finer resolution than the underlying NWP, to create high-resolution spatial predictions.

Due to the ensemble copula (Sect. 2.3) and the re-weighting procedure the full probabilistic predictions as returned by SAMOS are reduced to a $50+1$-member ensemble.

This is necessary as the precipitation postprocessing uses a censored response distribution and a parametric decomposition is not possible (central limit theorem). As a side note it has to be mentioned that the ranks of the hourly copulas are no longer strictly preserved and might sometimes differ from the original rank structure of the EPS.

2.5 New snow amount and probability of snow

Once ECC-Q and re-weighting are applied to the marginal distributions, bi-variate time series of calibrated hourly precipitation sums and 2 m temperatures are available for each of the M ensemble members. For each individual pair of member m and forecast step s the "snow indicator function" SI_{ms} can be retrieved.

$$\mathrm{SI}_{ms} = \begin{cases} \text{“dry” if} \\ \quad \text{precipitation}_{ms} \leq 0.05\,\mathrm{mm\,h^{-1}} \\ \text{“rain” if} \\ \quad \text{precipitation}_{ms} > 0.05\,\mathrm{mm\,h^{-1}} \wedge T_{2\,\mathrm{m}} > 1.2\,^{\circ}\mathrm{C} \\ \text{“snow” if} \\ \quad \text{precipitation}_{ms} > 0.05\,\mathrm{mm\,h^{-1}} \wedge T_{2\,\mathrm{m}} \leq 1.2\,^{\circ}\mathrm{C} \end{cases} \tag{11}$$

The threshold of $0.05\,\mathrm{mm\,h^{-1}}$ has been chosen as the smallest recorded value of the rain gauges used for validation is 0.1 mm. To distinguish between rain and snow we use a fixed threshold of 1.2 °C as a rough approximation, following Bellaire et al. (2011, p. 1121). The empirical probabilities π_{cs} for each of the three classes (snow, rain, and dry, which are mutually exclusive for each individual member and forecast time step) or for combinations can be computed using

$$\pi_{cs} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}(\mathrm{SI}_{ms} = c), \tag{12}$$

where s is a specific forecast step and c is the desired class (e.g., snow, rain, rain $\vee$ snow). $\mathbf{1}(\cdot)$ is an indicator function which takes 1 if the argument in brackets is true or 0 otherwise. The conditional expectation can be derived similarly:

$$E[c] = \frac{\sum_{i=1}^{M} \text{precipitation}_{ms} \cdot \mathbf{1}(\mathrm{SI}_{ms} = c)}{\sum_{i=1}^{M} \mathbf{1}(\mathrm{SI}_{ms} = c)}. \tag{13}$$

If one is interested in the snow height of fresh snow (E[snow] in centimeters), the snow density has to be taken into account. A rule of thumb is the "1 : 10 rule" where 1 mm of liquid water equivalent, the quantity forecasted by the postprocessing, corresponds to 1 cm of fresh snow. This is equivalent to a fresh snow density of $100\,\mathrm{kg\,m^{-3}}$. In reality, fresh snow densities can vary strongly between 10 and $526\,\mathrm{kg\,m^{-3}}$ given location and prevailing conditions (e.g., Meister, 1985; Judson and Doesken, 2000; Roebber et al., 2003). As reliable fresh snow height or density observations with the desired temporal resolution are not available for this study, a detailed verification cannot be performed. For visual representation we simply assume a mean density of $100\,\mathrm{kg\,m^{-3}}$.

3 Data

This section presents the data sets used for this study. These consist of two different EPS forecast data sets (ECMWF hindcast and operational EPS) and three different sources of observation data for model training and verification.

3.1 Numerical weather prediction data: forecast data

All predictions presented in this article are based on the ECMWF EPS. The ECMWF EPS consists of 50 perturbed ensemble members and 1 control run ($50 + 1$) and is initialized four times a day every 6 h. For this study, the control run is treated the same way as the 50 perturbed members. We will solely focus on the 00:00 UTC forecast run of EPS model version *43r1*. This version became operational on 22 November 2016 and the output is available at an hourly temporal resolution up to +90 h ahead on a $\sim 16\,\mathrm{km} \times 16\,\mathrm{km}$ regular longitude–latitude grid. A visual representation of the grid is shown in Fig. 1.

The presented application will focus on the winter season 2016/17 (1 December 2016 through 15 April 2017) and on predictions from +6 h to +78 h in advance, spanning the first 3 days after EPS initialization (06:00 to 06:00 UTC of 3 consecutive days).

3.2 Numerical weather prediction data: training data

To train the SAMOS models we use ECMWF hindcasts, similar to the approach of Stauffer et al. (2017b). ECMWF hindcasts become available twice a week (Mondays and Thursdays), providing a $10+1$ member ensemble for the same date over the previous 20 years, initialized at 00:00 UTC. For example: on Monday 2 January 2017 hindcasts for 2 January 2016, 2015, …, 1998, and 1997 become available. As for the EPS, the hindcast control run is treated as an additional member to increase the ensemble sample size. The hindcasts are available at the same spatial resolution as the EPS, but at a 6-hourly temporal resolution only. To create the training data set for the statistical postprocessing models all hindcasts are bilinearly interpolated to each of the measurement sites (see Sect. 3.3). Overall, 235 different grid points from the numerical model are involved in the interpolation for all 199 sites.

For the statistical postprocessing methods of 2 m temperature, all 6-hourly intervals from +6 to +78 h will be used. Besides the forecasted 2 m temperature the 2 m dew point temperature, 850 hPa temperature, and surface pressure forecasts are used as additional covariates (see Sect. 4). For precipitation, 24 h total precipitation sum hindcasts are used for the forecast time steps +30, +54, and +78 h.

3.3 Observational data

Two major different observation networks will be used in the following. As in Stauffer et al. (2017b), daily liquid water equivalent observations from the Tyrol network of hydrographical services (EHYD; BMLFUW, 2018) are used for the postprocessing of daily precipitation sums. In comparison to other networks in the area, the hydrographical service maintains the highest density of stations (number of stations) with very long historical records (up to 47 years of data). The observation sites are well distributed up to an altitude of about 1800 m a.m.s.l. However, observations are only made once a day (manually) at 06:00 UTC. In the following, the observations from 110 sites in and around Tyrol are used to train the precipitation SAMOS models.

The second network consists of 89 automated weather stations operated by the national weather service (TAWES network; Zentralanstalt für Meteorologie und Geodynamik). Seventy-five out of these 89 stations provide at least 6 years of data. Observations are recorded every 10 min, of which all observations at every full hour are used for training and validation of the 2 m air temperature SAMOS models.

The TAWES network also provides automated precipitation measurements at a 10 min resolution. However, the length of historical records is much shorter compared to the time series provided by EHYD data set. Furthermore, the measurement errors of the automated rain gauges are expected to be larger than the errors from the daily manual records provided by the hydrographical service, especially during winter. Thus, we decided to not use the TAWES precipitation observations for model training and for the estimates of the spatio-temporal climatologies. Nevertheless, since observations from the hydrographical service are only available up to 2012 at this time (2018), we do use TAWES precipitation observations for validation. Therefore, daily precipitation sums are generated by taking the sum over all 10 min intervals between 06:10 and 06:00 UTC of the following day (yields 144 10 min values). Periods for which more than four 10 min values are missing are eliminated.

In addition to the temperature and precipitation observations from the hydrographical service and the TAWES network, meteorological aerodrome reports (METARs) from Innsbruck Airport are used in the verification section as it is the only longer-term source of temporally high-resolution *precipitation-phase* observations available. The weather conditions from the METARs are classified as "snow" (if the report contains SN, SG, IC, PL, SNRA, or RASN), "rain" (if the message contains DZ, RA, SNRA, or RASN), and "dry" (else). Conditions with sleet (mixed rain/snow; SNRA/RASN) are attributed to both "snow" and "rain". METARs are available every 30 min, created by either a human observer or an automated procedure if the airport is closed over night. These observations have been aggregated to an hourly temporal resolution and will be used to validate the forecasted probabilities of snowfall. Overall, 3318 observations are available for the time period of interest, with 333 cases reporting rain or sleet (10 %), 246 cases snow or sleet (7.5 %), and 2786 cases dry conditions (84 %).

Figure 1 shows an overview of the area of interest. The markers show the locations of the observational sites from the two networks (TAWES, EHYD) and the location of the airport (581 m a.m.s.l.). To the right the height distribution of the stations from the two networks is shown.

4 Statistical models

This section presents the specifications of the models that will be compared and tested in Sect. 5. During the preparation of this paper, a variety of slightly different model formulations were tested and the presented models are only a subset that was selected because they performed well or showed interesting results.

All the models follow the approaches presented in Sect. 2 but differ in their input variables and whether the data are transformed to standardized anomalies. Four models will be used for 2 m temperature and three for daily precipitation sums. The training data set to estimate the regression coefficients is composed of all forecast steps provided by the ECMWF hindcasts from +6 up to +78 h on a regular 6 h interval. For precipitation, these forecasts are aggregated to 24 h sums, resulting in forecast steps +30, +54, and +78 h. The power parameter was set to $p = 1.35$, found to have the best predictive cross-validated performance in Stauffer et al. (2017b).

Table 1 shows the different model assumptions and naming. The first two models named *EMOS* correspond to Eqs. (1)–(8) operating on the physical scale (not on standardized anomalies). One crucial modification has to be made for the 2 m temperature: interactions with factors for the time of day (*hour*; 00:00/06:00/12:00/18:00 UTC) and the station (*station*) are included to capture spatial and diurnal differences, yielding separate (and independent) coefficients for each station and each time of day. For daily precipitation sums, this extension has not been made as only 06:00 UTC observations are included (no diurnal effect required) and station-wise regression models partially returned highly unstable estimates due to the low number of observations for each individual site. Please note that the *EMOS* models are *not designed for spatial or spatio-temporal predictions* even if spatial predictions would be possible in the case of precipitation. These two models serve as a reference for the performance of the SAMOS models.

All other models are spatio-temporal (in the case of 2 m temperature) and spatial (in the case of daily precipitation sums) SAMOS models operating on the standardized anomaly scale. Thus, the spatial and temporal characteristics among all stations and for all lead times are already removed from the data and do not have to be considered in the linear predictors for location $\boldsymbol{\mu}^*$ and scale $\boldsymbol{\sigma}^*$.

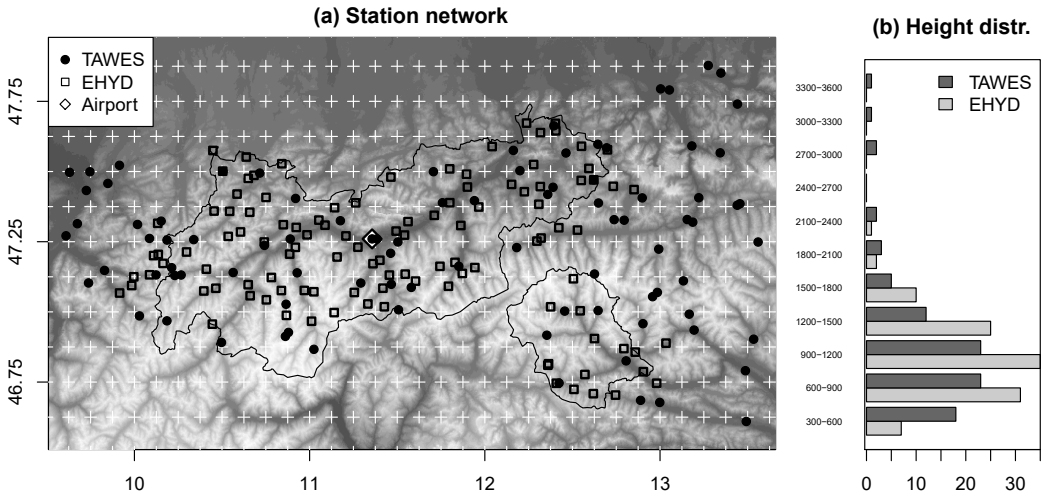

Figure 1. Panel **(a)** shows the topography of the area of interest. Overlays: center of the grid cells of the NWP model data (white crosses), governmental area of Tyrol (black outline), location of the TAWES stations (89; circles) and EHYD stations (110; squares). The airport is indicated by a diamond in the center of the map. Panel **(b)** shows the height distribution of the stations grouped into 300 m intervals: number of stations (abscissa) and altitude intervals (ordinate; meters a.m.s.l.).

The second and third pairs of models, named *SAMOS_hom* and *SAMOS_het*, are two SAMOS variations, both solely using the corresponding quantity from the EPS as a covariate (i.e., 2 m temperature and total precipitation, respectively). While *SAMOS_het* is a full heteroscedastic model including the ensemble standard deviation in the linear predictor for the scale $\boldsymbol{\sigma}^*$, *SAMOS_hom* is a homoscedastic model where the scale does not depend on any covariates. These two models allow one to quantify the improvement in the predictive performance by including the ensemble spread information in the postprocessing methods. For 2 m temperature, a fourth model called *xSAMOS_het* (*x* for *extended*) is used, which includes additional covariates for both location $\boldsymbol{\mu}^*$ and scale $\boldsymbol{\sigma}^*$. A set of multilinear models (not shown) has been tested that includes different interactions and nonlinear effects in the linear predictors, but no major improvements have been found. Thus, a relatively simple model specification for *xSAMOS_het* is included in this article to demonstrate that SAMOS can easily be extended. The multilinear *xSAMOS_het* contains three additional covariates as linear main effects for both location $\boldsymbol{\mu}^*$ and scale $\boldsymbol{\sigma}^*$. For each of the covariates separate regression coefficients are estimated during model optimization which, in this case, yields 10 coefficients in total (one intercept and four covariates in each linear predictor).

5 Results

The first two subsections show the performance of the full predictive distributions of the 2 m temperature (Sect. 5.1) and daily precipitation forecasts (Sect. 5.2). Section 5.3 shows a example of the spatial coherence restored via ECC followed by a detailed verification of hourly predictions and hourly precipitation-type classification based on the postprocessed ensembles. Last but not least, spatial forecasts for a specific forecast are shown in Sect. 5.6 to demonstrate the feasibility of high-resolution areal predictions.

5.1 Temperature (6 h intervals)

Figure 2 shows bias, continuous rank probability score (CRPS; Gneiting and Raftery, 2007), mean width of the prediction interval between the 10 % and 90 % percentiles, and CRPS skill scores, all based on the full predictive distribution returned by the statistical models. All results are temporally out-of-sample and validated on the TAWES network for all forecast steps $+6/+12/\ldots/+72/+78$ h as used to train the statistical models on hindcasts. The box-and-whiskers show station-wise mean scores for the spatio-temporal climatology (CLIM; Eq. A1), the raw EPS, and the four statistical postprocessing models (cf. Table 1).

The raw EPS performs poorly for the area of interest as the NWP model with its current spatial resolution is not able to represent the local topography. It performs even worse than the underlying climatology in terms of bias and CRPS. All statistical postprocessing models perform significantly better and are essentially bias-free. As expected, the station-wise statistical *EMOS* model performs best since it has separate model coefficients for each station location and is thus more flexible than the spatial models. In terms of CRPS, the spatial models lose about 7 %–12 % of skill (Fig. 2d; *SAMOS_hom*: −12.3 %; *SAMOS_het*: −12.3 %; *xSAMOS_het*: −6.9 %), but allow one to predict at any arbitrary location within the area of interest and any desired time between +6 and +78 h. The two models *SAMOS_hom* and *SAMOS_het* perform very

Table 1. Statistical model specification for 2 m temperature (left) and 24 h precipitation sums (right). For each model the linear predictors for μ and $\log(\sigma)$ are shown. Superscript asterisk indicate variables on the standardized anomaly scale (SAMOS). $T_{2\,\mathrm{m}}$, $\mathrm{Td}_{2\,\mathrm{m}}$, T_{850}, P, and tp are the 2 m temperature, 2 m dew point temperature, temperature in 850 hPa, surface pressure, and total precipitation ensemble forecasts respectively. $\overline{X}$ are ensemble means, $\langle X \rangle$ denote ensemble log-standard deviations. X / hour and X / station are interactions between X and the "time of the day" and/or the "station".

Models for 2 m temperature using a Gaussian response distribution.		Models for 24 h precipitation sums using a power-transformed left-censored logistic response distribution.	
Heteroscedastic EMOS models (EMOS; cf. Eqs. 1–3 and 5–7) These models *are not designed to provide spatial or spatio-temporal predictions.*			
μ	= hour / station $+\overline{T_{2\,\mathrm{m}}}$ / hour / station	μ	$= \overline{\mathrm{tp}^{1/p}} \cdot (1-z) + z$
$\log(\sigma)$	= hour / station $+\langle T_{2\,\mathrm{m}} \rangle$ / hour / station	$\log(\sigma)$	$= \langle \mathrm{tp}^{1/p} \rangle \cdot (1-z)$
Homoscedastic SAMOS models (SAMOS_hom)			
μ^*	$= \overline{T^*_{2\,\mathrm{m}}}$	μ^*	$= \overline{\mathrm{tp}^{1/p^*}} \cdot (1-z) + z$
$\log(\sigma^*)$	= constant	$\log(\sigma^*)$	= constant
Heteroscedastic SAMOS models (SAMOS_het)			
μ^*	$= \overline{T^*_{2\,\mathrm{m}}}$	μ^*	$= \overline{\mathrm{tp}^{1/p^*}} \cdot (1-z) + z$
$\log(\sigma^*)$	$= \langle T^*_{2\,\mathrm{m}} \rangle$	$\log(\sigma^*)$	$= \langle \mathrm{tp}^{1/p^*} \rangle \cdot (1-z)$
Extended Heteroscedastic SAMOS models (xSAMOS_het)			
μ^*	$= \overline{T^*_{2\,\mathrm{m}}} + \overline{\mathrm{Td}^*_{2\,\mathrm{m}}} + \overline{T^*_{850}} + \overline{P^*}$	–	
$\log(\sigma^*)$	$= \langle T^*_{2\,\mathrm{m}} \rangle + \langle \mathrm{Td}^*_{2\,\mathrm{m}} \rangle + \langle T^*_{850} \rangle + \langle P^* \rangle$	–	

similarly, indicating that the uncertainty information from the EPS 2 m temperature forecast provides barely any additional information. Small improvements can be achieved by including additional covariates (model *xSAMOS_het*).

Overall, all statistical models show promising values in terms of CRPS (median 1.45–1.65 °C) and mean absolute error (median 2.0–2.3 °C; not shown) across all four methods. The median of the mean prediction interval width for the 10 %–90 % interval is around 6.0 °C for the station-wise *EMOS* model and around 6.9–7.2 °C for the SAMOS models.

5.2 Daily precipitation sums

Figure 3 shows the verification of the daily precipitation sum predictions for the forecast steps $+30/+54/+78$ h. Again, this analysis is based on the full predictive distribution returned by the statistical models. Here, the validation is done on different stations (TAWES) than used for model fitting (EHYD; Sect. 3), so that these results are spatially and temporally out of sample. The box-and-whiskers show station-wise mean scores for the spatio-temporal climatology (CLIM; Eq. A2), the raw daily accumulated total precipitation from the ECMWF EPS (raw EPS), and the three postprocessing methods shown in Table 1.

The top row of Fig. 3 shows bias, CRPS, and the Brier score for probability of precipitation ($\mathrm{BS}_{0\,\mathrm{mm}}$). The row below shows skill scores with the raw EPS as reference. The two SAMOS models (*SAMOS_hom* and *SAMOS_het*) show the best bias among all methods but less predictive skill in terms of MAE, CRPS, and $\mathrm{BS}_{0\,\mathrm{mm}}$ than the *EMOS* model not using standardized anomalies. The distinct improvements in the $\mathrm{BS}_{0\,\mathrm{mm}}$ are expected due to the well-known wet bias of the EPS when comparing interpolated data (spatial scale) to a specific site (point scale). As for 2 m temperature, the use of the forecasted EPS uncertainty in the heteroscedastic model (*SAMOS_het*) brings barely any improvement. The performance of the *EMOS* model requires special attention. Even if this model is not designed to create spatial predictions the results show a slightly better performance than the two SAMOS models.

5.3 Spatial coherence (ensemble copula coupling)

Sections 5.1 and 5.2 examine the predictive skill of the full probabilistic predictions (SAMOS; Sect. 2.2). The next step is to apply ECC-Q based on the postprocessed hourly 2 m temperature forecasts and daily precipitation sums (Sect. 2.3) to restore the spatial structure of the forecasts.

To illustrate the effect of ECC-Q, Figs. 4 and 5 show forecasts for both 2 m temperature and daily precipitation sums, of one random member (member 38) of the forecast for 10 March 2017. Both figures show the actual forecasts of this specific member and the deviation of this member from the median of the full underlying ensemble. The latter one is used to highlight the spatial coherence which is less perceptible in the forecasts itself due to the superimpo-

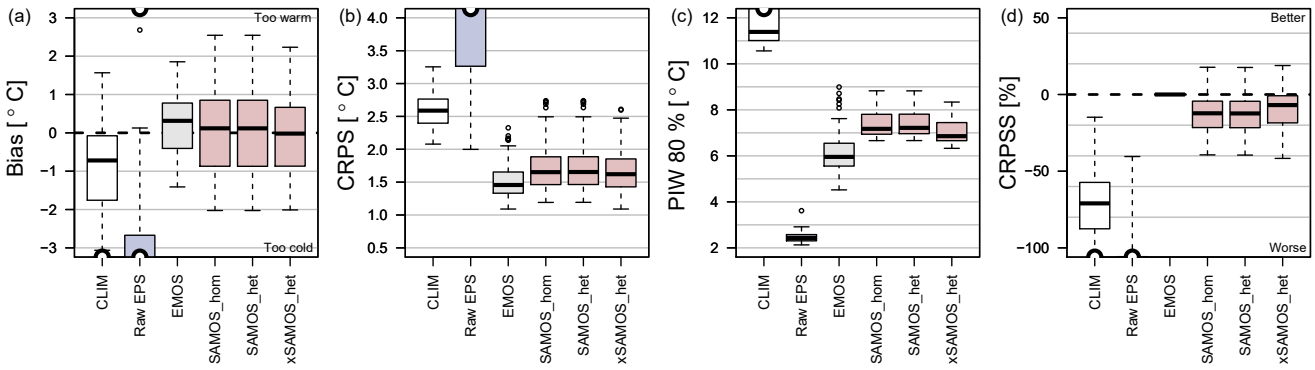

Figure 2. Scores for 2 m temperature forecasts based on the full predictive distribution based on $6/+12/\ldots/+72/+78$ h forecasts as used for model training. The box-and-whisker shows station-wise means for **(a)** bias (observation minus forecast), **(b)** CRPS, **(c)** width of the 80 % prediction interval, and **(d)** CRPS skill scores with *EMOS* as reference. Scores are shown for the climatology (CLIM), the raw EPS, and the four postprocessing models (cf. Table 1). Abscissa are set to manually specified ranges; the "semi-sphere" marker (top/bottom) indicates data outside the plotted range.

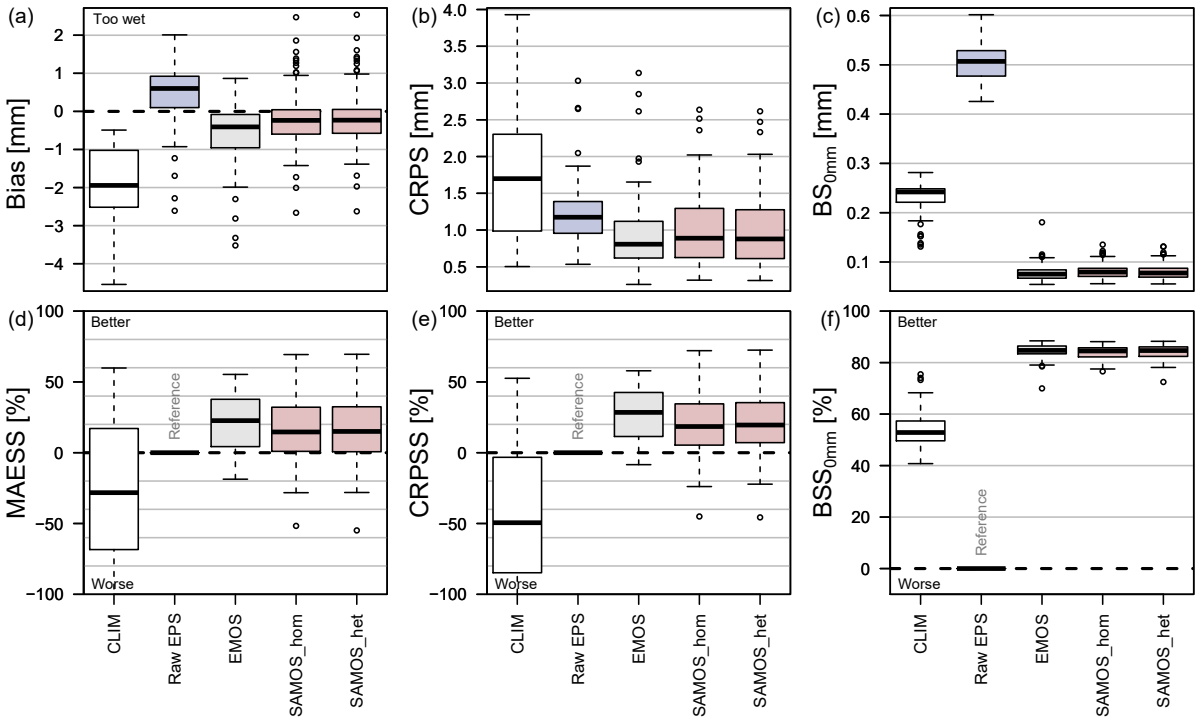

Figure 3. Scores for 24 h precipitation sums based on the full predictive distribution for +30, +54, and +78 h forecasts as used for model training. Box-and-whiskers of station-wise mean scores for **(a)** bias (observation minus forecast), **(b)** CRPS, and **(c)** Brier scores for probability of precipitation. The scores are shown for the climatology (CLIM), raw EPS, and the three postprocessing models (cf. Table 1). The lower row shows skill scores for **(d)** mean absolute error, **(e)** CRPS, and **(f)** Brier score for probability of precipitation with the *raw EPS* as reference. Positive skill scores indicate an improvement over the *raw EPS*.

sition of location- and elevation-dependent effects. Forecasts and deviations are shown for the raw ensemble, the quantiles drawn from the full probabilistic predictions, and ECC-Q after restoring the rank order structure of the EPS.

As ECC-Q uses quantiles based on equidistant probabilities, the quantiles drawn from the full probabilistic distribution are ordered. Thus, the forecasts of member 38 ($\pi = 38/52$) are always higher than the median of the ensemble

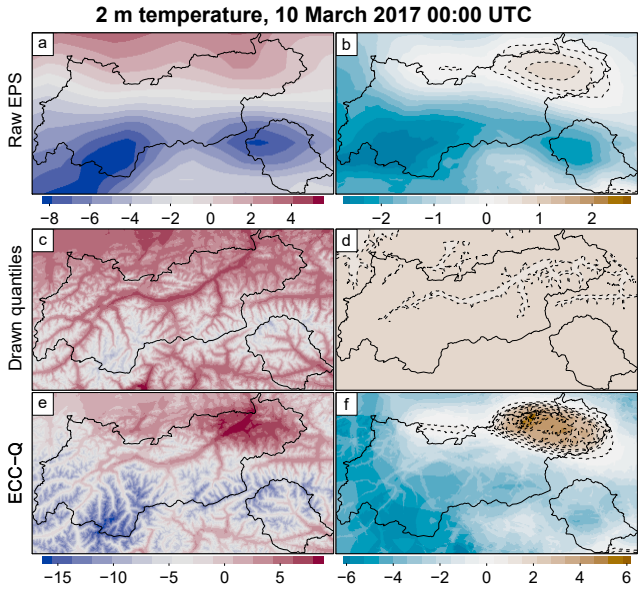

Figure 4. 2 m temperature forecasts of member 38 for 10 March 2017 00:00 UTC. Top–down: raw EPS **(a, b)**, unsorted quantile **(c, d)**, and ECC-Q **(e, f)** after restoring the rank order structure. Forecast (**a, c, e**; °C) and deviation of this forecast from the median of the corresponding ensemble (**b, d, f**; °C) are shown. Overlays: contours for positive deviation (dashed) and the borders of the governmental area of Tyrol (solid). Please note that the color scale of the top row differs from the scale of the two lower rows.

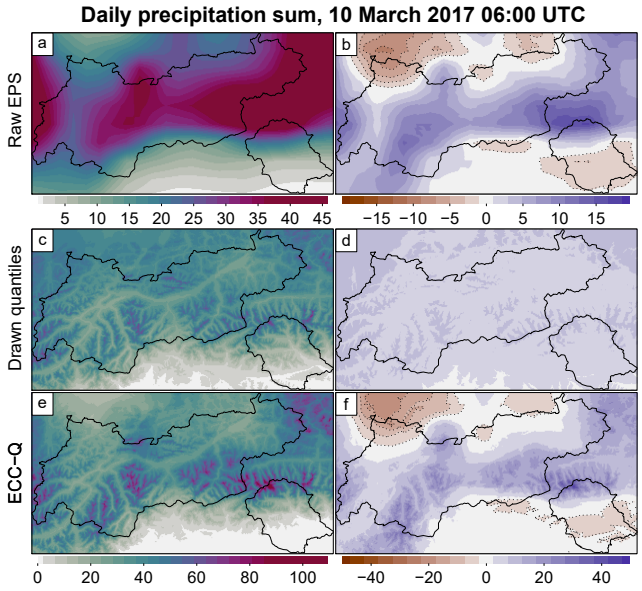

Figure 5. As Fig. 4 but for daily precipitation sums valid 9 March 2017 06:00 UTC to 10 March 2017 06:00 UTC. Forecasts **(a, c, e)** and deviations from the ensemble median **(b, d, f)** are shown in mm $24\,h^{-1}$. In contrast to Fig. 4 contours are plotted for negative deviations.

($\pi = 0.5$) before the rank order structure is restored. This can be seen in Figs. 4d and 5d, where the deviation against the ensemble median is plotted. In this case the deviation is (more or less) a constant positive offset across the whole domain with only little spatial structure. These small spatial features are induced by the SAMOS procedure where the data are transformed into the standardized anomaly scale and back to the physical scale (Sect. 2.2) and are not associated with the spatial coherence of the EPS (cf. Figs. 4b and 5b). To restore the spatial structure, the quantiles have to be reordered given the rank order structure of the raw EPS at each of the target locations. The bottom rows of Figs. 4 and 5 show the forecasts after rearranging the quantiles. In contrast to the non-rearranged forecasts (middle row) the postprocessed forecasts with restored rank-order structure exhibit a very similar spatial coherence to the raw EPS (top row of Figs. 4 and 5). The coherence of the EPS is maintained unchanged in large parts, but is not identical as it is slightly modified by the postprocessing procedure.

5.4 Hourly temperature and precipitation sums

Sections 5.1 and 5.2 show that the postprocessing models are able to improve the predictive performance of the raw EPS for temperature and daily precipitation sums. The main goal of this study is to provide accurate and reliable snow predictions by combining hourly 2 m temperature and precipitation forecasts. Thus, an hourly temporal resolution for both temperature and precipitation forecasts is required. This section therefore shows the verification of hourly forecasts. For temperature, the hourly forecasts are based on the spatio-temporal SAMOS model *xSAMOS_het* as it shows the overall best performance among all tested spatial models. The hourly precipitation sums are based on the predictions from the *SAMOS_het* model downscaled to the desired temporal resolution using the re-weighting approach presented in Sect. 2.4. Since the re-weighted precipitation forecasts are only available as ensembles but not as full predictive distributions, ensemble verification methods are employed in the following.

Figure 6a–d show ensemble rank histograms (Hamill, 2001) for hourly temperature predictions and hourly precipitation sums for the raw EPS and the postprocessed forecasts. Each observation is assigned to a rank where observations falling below the lowest member get rank 1 and observations higher than the highest member get rank 52 (50+1 members, 52 possible ranks). A perfectly uniform distribution would indicate perfect calibration. For temperature (Fig. 6a, b), the postprocessing strongly improves calibration compared to the raw EPS. However, the pronounced U-shape indicates that the predicted uncertainty is lower than in reality (underdispersion). A similar picture can be seen for the hourly precipitation sums plotted as "stacked ensemble rank histograms" (Fig. 6c, d). The total height of the bars given the

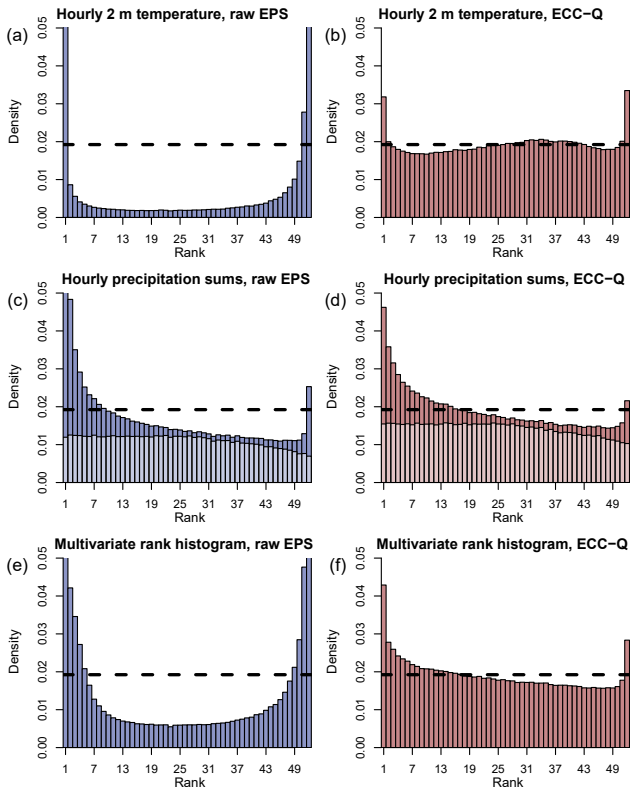

Figure 6. (Stacked) ensemble rank histograms for hourly 2 m temperature **(a, b)** and hourly precipitation sum forecasts **(c, d)** plus multivariate rank histogram **(e, f)** of the raw EPS **(a, c, e)** and postprocessed copula **(b, d, f)** with $50 + 1$ members each. The rank histograms contain all available forecasts for all stations and forecast steps $+7$, $+8$, …, $+78$ h in advance. For precipitation, the faded colors show the rank histogram for all forecasts where 50 % or more of all members predicted $0\,\mathrm{mm\,h^{-1}}$. Please note that the y axis is cut at 0.05 for all histograms.

rank shows the rank histogram of the full verification data set. The faded colors show the calibration for all forecasts where at least 50 % of all members forecasted $0\,\mathrm{mm\,h^{-1}}$ (dry cases). It can be seen that the dry cases are relatively well calibrated and that the majority of the underdispersion results from the wet cases. Nevertheless, the asymmetry (decreasing density with increasing rank) indicates a small wet bias also for the dry cases.

To score the multivariate skill of the combined temperature and precipitation forecasts, the bottom row of Fig. 6 shows multivariate (bivariate) rank histograms (Gneiting et al., 2008). In contrast to the univariate rank histograms the multivariate rank histogram takes the rank order structure between the two quantities into account. As for the univariate rank histograms the multivariate rank histogram shows much better calibration of the postprocessed predictions but shows very similar patterns to the two univariate histograms (Fig. 6a–c).

To investigate the univariate predictive performance of hourly predictions for different forecast horizons, Fig. 7 shows CRPS skill scores for all individual lead times. Each box-and-whisker contains station-wise mean skill scores over the verification period. While always on a high level, the 2 m temperature forecasts for morning hours ($+7$ to 12, $+31$ to 36, $+55$ to 60, corresponding to 07:00–12:00 UTC) show slightly less skill. For precipitation, the skill scores are overall positive but clearly decreasing with increasing forecast horizon. The lowest skill scores are found for early morning hours ($+26$ to 30, $+50$ to 54, $+74$ to 78; 02:00–06:00 UTC).

5.5 Fresh snow amounts and probability of snowfall

This section shows the verification for the main target variable. Due to the limited availability of temporally high-resolution and reliable observations this can only be done for one site, the regional airport in Innsbruck (Fig. 1). Figure 8 shows reliability diagrams (Bröcker and Smith, 2007) for the probability of precipitation (rain $\vee$ snow), rain, and snow. As Sect. 5.4 indicates that large parts of the improvements are expected to come from temperature postprocessing, three different models will be compared: the raw EPS, the full ECC-Q, and a mixed version. The mixed version uses the raw hourly precipitation forecasts from the EPS but the postprocessed temperature predictions to examine the contribution of the precipitation postprocessing. The validation for all three methods is based on the classification described in Sect. 2.5 and the aggregated METAR observations as described in Sect. 3.3.

For all three precipitation classes ECC-Q is able to outperform the raw EPS (less off-diagonal) and shows lower Brier scores and lower numbers for reliability while losing some resolution. ECC-Q is also beneficial over the mixed version using uncorrected precipitation sums. For snow the two methods using postprocessed temperature forecasts (mixed and ECC-Q) perform very similarly but show different biases. While the mixed model exhibits a wet bias (observed frequencies larger than forecasted probabilities), ECC-Q shows a dry bias. The results for snow should not be overinterpreted as snowfall is relatively rare at this station (7.5 % of all cases). The raw EPS again shows the well-known wet bias in all three classes.

Next, Fig. 9 shows a forecast time series example for a random station and a day when the temperature is just around 1.2 °C, the threshold used to decide whether the forecasted precipitation will fall as snow or rain. As no fresh snow measurements are available, a validation of the forecasted fresh snow amounts cannot be performed for this case.

What can be seen is that the ECC-Q temperature predictions (Fig. 9a) show a much larger spread than the raw EPS. The postprocessed temperature uncertainty dominates the variation of the observed temperature over the whole forecast period (days 1–3). The observations, however, nicely fall into this interval, which yields the overall well-calibrated

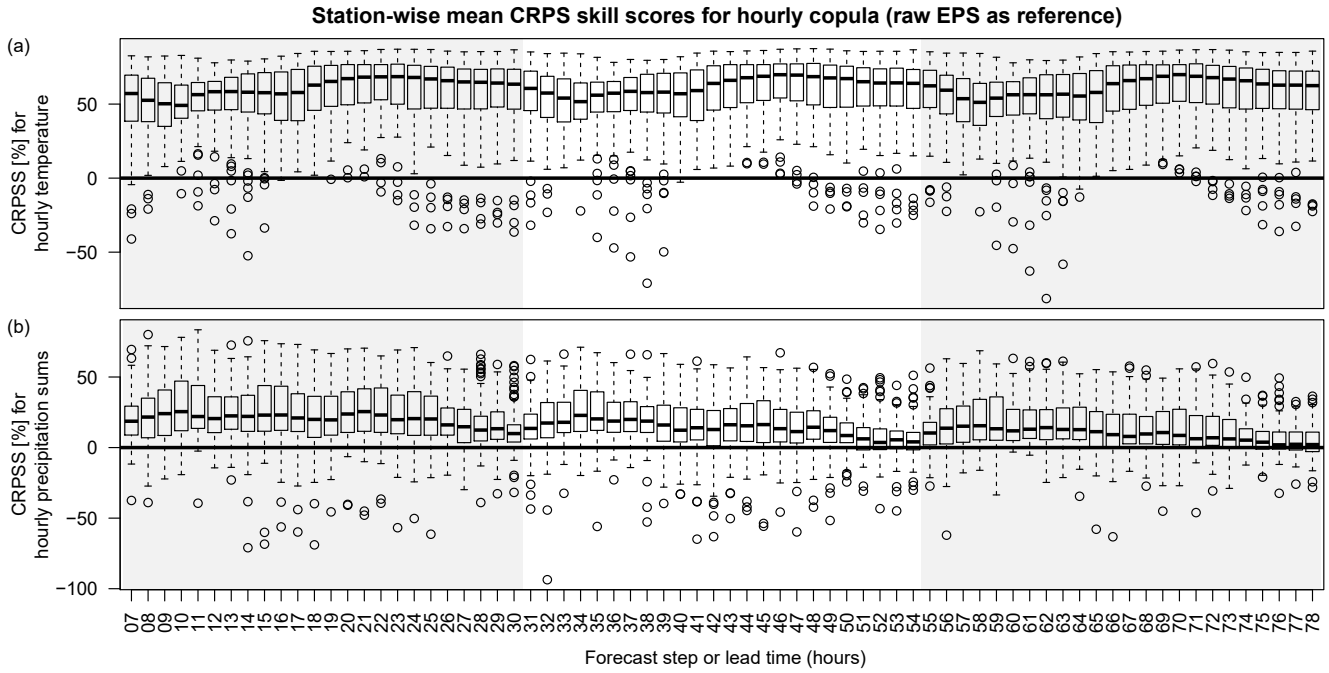

Figure 7. Continuous ranked probability skill scores (CRPSS) for 2 m temperature **(a)** and hourly precipitation sums **(b)** based on station-wise mean empirical CRPS values ($50 + 1$ members). The raw EPS is used as a reference. CRPSSs are shown for each individual forecast step from +7 to +78 after model initialization. CRPSSs above 0 (bold black line) show that the postprocessed hourly forecasts outperform the raw EPS.

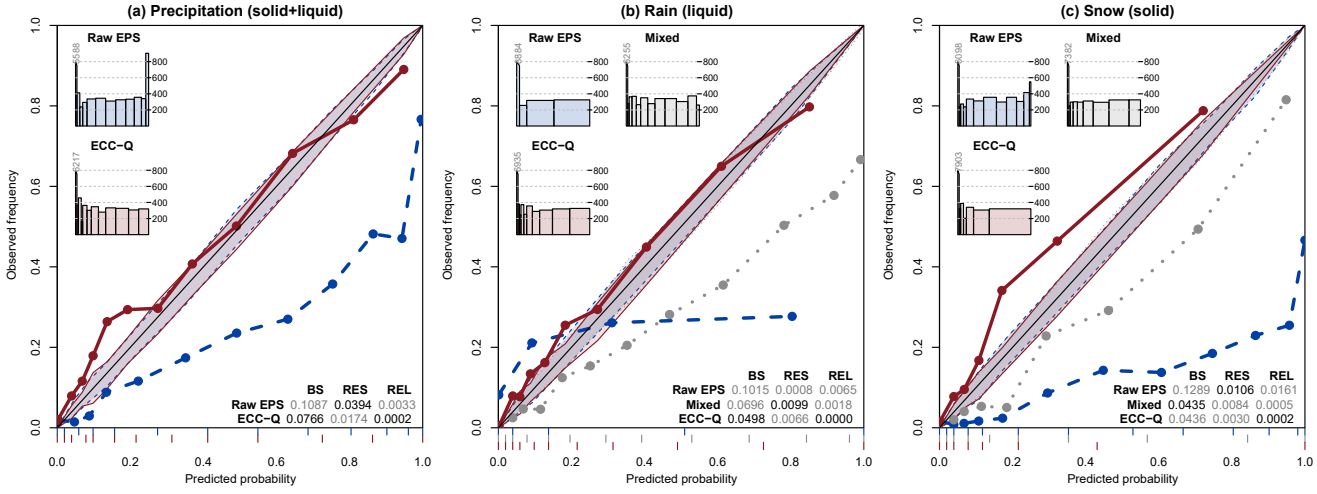

Figure 8. Reliability diagrams for hourly predictions of precipitation (snow ∨ rain; **a**), snowfall **(b)** and rain **(c)** at Innsbruck Airport based on meteorological aerodrome reports (METARs) for the raw EPS (dashed) and the postprocessed forecasts (solid). Binning based on empirical quantiles to ensure a similar number of observations per bin (bins indicated along the x axis). The shaded area shows the 90 % confidence interval. Histograms: counts of the number of observations in each bin in the reliability diagram. The analysis is based on ≈ 9700 observation–forecast pairs for each precipitation type. Mean Brier score (BS), as well as mean resolution (RES) and reliability (REL) from a BS decomposition (Murphy, 1973), are shown in the lower right corner.

forecasts (see Fig. 6). For precipitation (Fig. 9c), the differences between the raw EPS and the postprocessed copula are less pronounced. Fig. 9b shows the probability of snow ∨ rain (precipitation), rain, and snow as defined by Eq. (12). The expected amounts of snow ∨ rain (precipitation) and snow from the postprocessed forecasts are plotted in Fig. 9d. Rather than plotting each individual ECC member, the median and two confidence intervals are shown. For this specific date and

location, the median shows 30.5 mm of precipitation (rain and/or snow liquid water equivalent) accumulated over the 3 consecutive days, of which 8.4 mm is expected to fall as snow. When assuming the 1 : 10 rule (Sect. 2.5) and not taking the alteration of the aging snow into account, this corresponds to a median of 8.4 cm of fresh snow within 3 days.

5.6 Spatial forecast example

As a last result, Figs. 10 and 11 show a spatial forecast example to demonstrate the ability to create high-resolution spatial predictions. These results show the +48 h forecast initialized 8 March 2017 on an approximately 500 m × 500 m grid (corresponds to the +48 h forecast shown in Fig. 9).

While Fig. 10 shows the probability of precipitation (snow ∨ rain), rain, and snow, Fig. 11 shows the expected amount of precipitation for the period > 47 to +48 h. The color coding represents the dominant precipitation type based on $\pi_{\text{snow, +48 h}}$ and $\pi_{\text{rain, +48 h}}$ (cf. Eq. 12). In addition, the snow line ($\pi_{\text{snow, +48 h}} > \pi_{\text{rain, +48 h}}$) is shown. For visual purposes the spatial predictions are plotted for the whole domain even if parts of the area are already outside the area covered by the stations used to create the underlying observation climatologies and to train the statistical models. Thus, forecasts outside the dashed line (Fig. 10a) should be interpreted with caution. The individual EPS and ECC-Q members used to derive probabilities and the expectation can be found in Appendix B; one specific member is shown in more detail in Sect. 5.3.

6 Discussion

This article presents a new hybrid approach to combine standardized anomaly output statistics (SAMOS) with ensemble copula coupling (ECC) and a novel re-weighting scheme for probabilistic snow forecasts. The results demonstrate that the new approach provides a framework for accurate high-resolution spatio-temporal probabilistic forecasts for 2 m temperature, precipitation, and snowfall over complex terrain.

The use of ECMWF hindcasts for model training and ECMWF EPS for prediction offers a computationally efficient way to get the required inputs for the SAMOS method (see Appendix A). Rather than estimating a complex spatio-temporal climatology for each covariate (as in Dabernig et al., 2017), only empirical moments (mean and standard deviation) of an appropriate hindcast subset have to be derived. The latest eight hindcast runs (4 weeks) centered around the date of interest are used to capture the seasonality. As this processing step is very cheap in terms of computational costs, one can easily derive hindcast climatologies for a range of possible covariates, which allows for a simple and low-cost multilinear extension of the SAMOS approach. Furthermore, due to the use of a rolling 4-week training period, the post-processing procedure automatically adapts itself to possible changes in the underlying NWP model within a few weeks. However, the rank histograms (Fig. 6) for both the 2 m temperature and daily precipitation sums show a pronounced U-shape. The same characteristics can be seen for all tested postprocessing models (not shown) whether or not standardized anomalies are used. The rank histograms for in-sample predictions based on the training data set itself (not shown) do not show this distinct pattern. A possible reason could be that the forecasted uncertainty of the hindcasts and the uncertainty information from the current EPS seem to differ. If the EPS overall provided sharper forecasts than the hindcast on which the regression coefficients are estimated, this would also yield underdispersive predictions after postprocessing. A detailed analysis of this specific issue was performed (beyond this article; not shown), but a clear statement to prove or falsify the hypothesis cannot be given.

The additional ensemble copula coupling (ECC-Q; Sect. 2.3) and re-weighting strategy yield satisfying results and are able to restore the spatial coherence based on the spatial structure of the raw EPS (Sect. 5.3, Appendix B). However, the bivariate verification (Fig. 6) shows distinct underdispersion. Additional tests have been performed to verify the improvement by restoring the multivariate rank order structure. Therefore, the multivariate rank histogram has been computed using random correlation by drawing a random rank order structure from the ensemble. It turns out (not shown) that the multivariate rank histogram with the random rank order structure only differs marginally from the one shown in Fig. 6 for both the raw EPS and ECC-Q. In other words: the correlation between 2 m temperature and hourly precipitation sums is negligible, at least for this study. Thus, the impacts of the cases where the rank order structure is not strictly preserved due to the re-weighting (Sect. 2.4) are not further investigated as no verifiable effect is expected.

Nevertheless, the method is still able to strongly improve calibration and reliability of the forecasts, especially for 2 m temperature, even though the sharpness is rather low. The mean 80 % prediction interval width for temperature is between 6.9 and 7.2 °C for the SAMOS methods. On a rainy/snowy day this interval is quite likely wider than the overall diurnal temperature variation. The relatively wide predictive intervals are a result of the input data. Due to the current spatial resolution, the EPS is not able to represent the area of interest in all its details. Consequently, a wide range of local features are not yet included. To mention one specific feature: the EPS shows a far-too-strong near-surface cooling over night, especially over snow. Errors of 15 °C between the forecasted 2 m temperature and the corresponding observation are relatively frequent for Alpine grid points. Furthermore, the forecasted EPS uncertainty does not seem to be very informative as almost no improvements can be seen when including it in the statistical models.

To improve the temperature forecasts, we include the temperature from the 850 hPa level as an additional covariate, which can be seen as a "free atmosphere" prediction over the

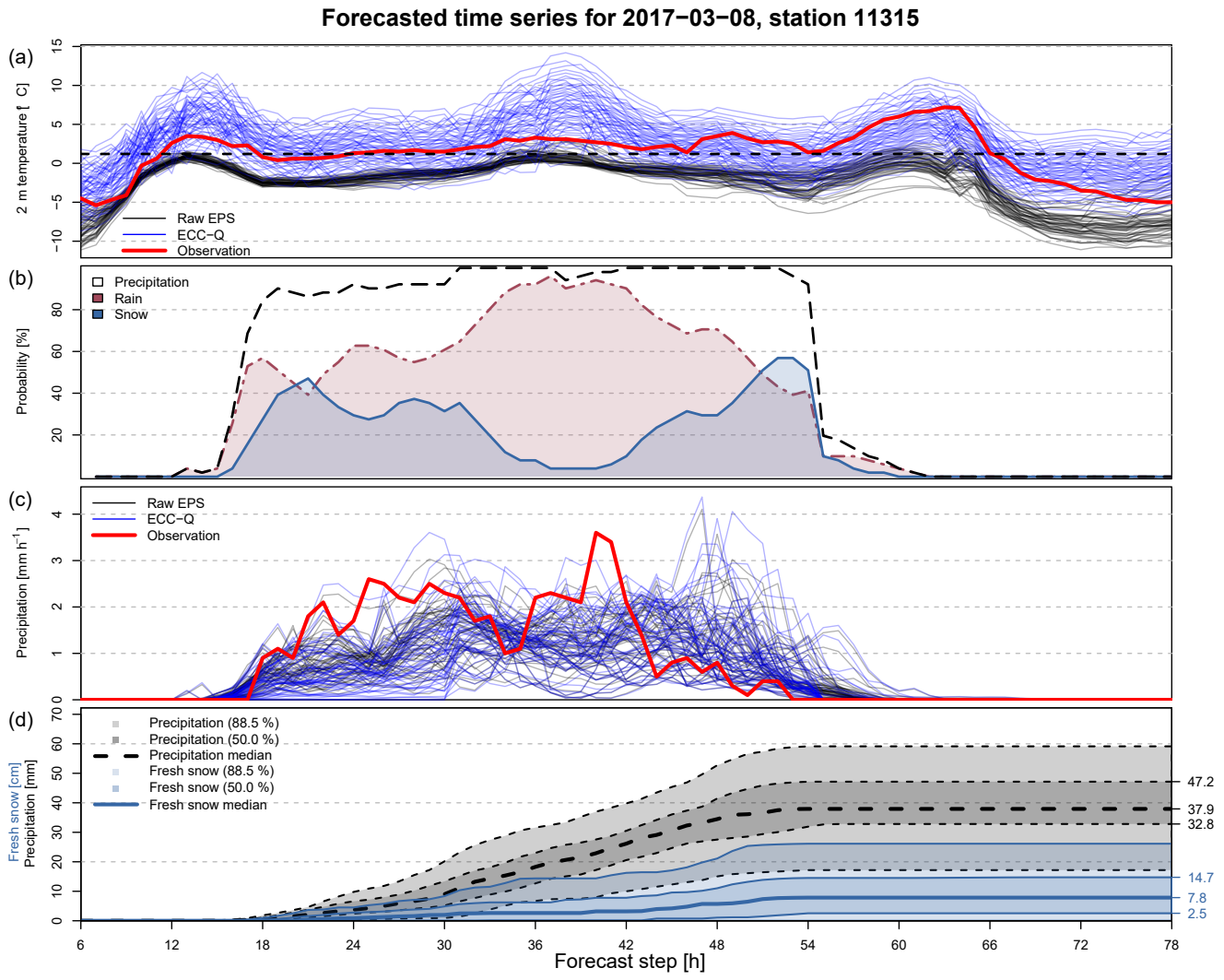

Figure 9. Example prediction for 8 March 2017 (station 11315, Holzgau) for the whole forecast horizon +6 up to +78 h ahead. **(a)** Raw EPS forecast (black), postprocessed copula (blue), and observation (red; bold) for 2 m temperature. The black dashed line is the 1.2 °C line used for precipitation type classification. **(b)** Probability of snow (blue solid), rain (red dotdash), and precipitation (snow ∨ rain; black dashed). **(c)** Hourly precipitation forecasts and observations as in panel **(a)**. **(d)** Postprocessed forecasts for precipitation sum (dashed; mm), and fresh snow height (solid; cm) using the 1 : 10 rule (snow density of 100 kg m^{-3}). Predicted medians, predicted 50 % intervals, and predicted 88.5 % intervals are shown.

area of interest. Furthermore, the 850 hPa temperature is a prognostic quantity which should be less strongly affected by possibly unrealistic surface processes (cooling/heating effects). In addition, surface pressure and 2 m dew point temperature are included to correct for weather-situation-dependent errors and very dry/wet conditions. The model shown in this article only includes the additional covariates as linear main effects and is more a proof of concept. We have also tested derived covariates such as 2 m potential temperature and nonlinear mixtures of 2 m temperature and 850 hPa temperatures to allow high-elevation stations to take the information from an elevated air mass ("free atmosphere") rather than from the near surface. As none of these models showed large improvements, and for simplicity, we decided not to show the results of these more complex models in this article. However, the "extended heteroscedastic SAMOS model" demonstrates that the SAMOS model can easily be extended by including additional covariates which do not necessarily have to be linear. As shown, this allows one to further improve the predictive performance, even with this simple model. A more flexible SAMOS model might bring further improvements, e.g., by including a larger set of covariates, including interactions between the different covariates, or by using more flexible effects such as multi-dimensional effects which can be used to represent elevation-

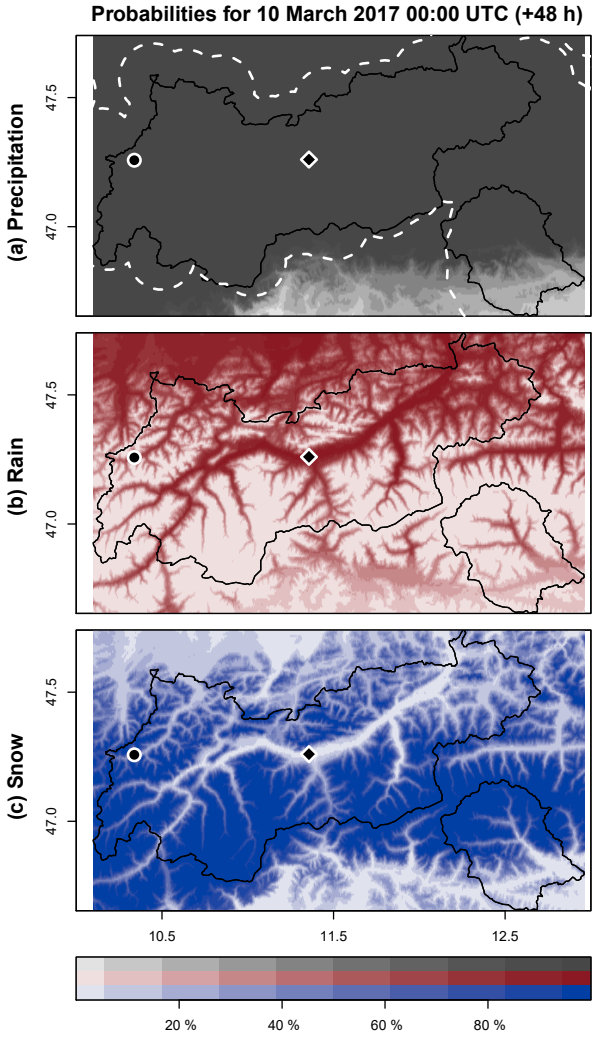

Figure 10. Top–down: 1 h probability of precipitation (rain ∨ snow), rain, and snow for 10 March 2017 00:00 UTC (+48 h forecast initialized 8 March 2017). Overlays: the governmental area of Tyrol (solid line), Innsbruck Airport (diamond), and the location of the example station used in Fig. 9 (circle). The white dashed line outlines the area not further away than 10 km from the closest measurement site.

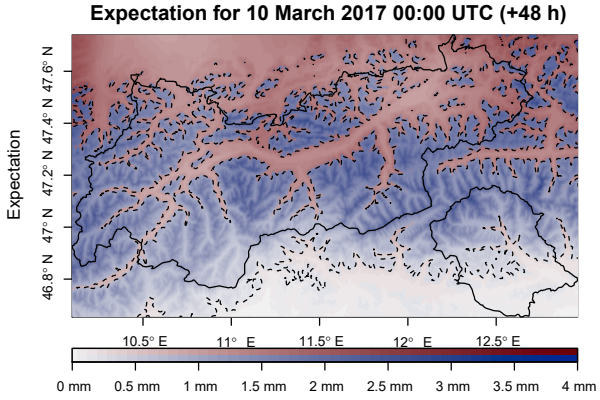

Figure 11. Expected 1 h amount of liquid water content for 10 March 2017 00:00 UTC (+48 h forecast initialized 8 March 2017). Areas with a higher chance of observing snow are shown in blue, those with a higher chance of observing rain in red. The dashed line (top) shows the forecasted snow line with an equal chance of observing snow or rain (Eq. 12). Overlay: governmental area of Tyrol (solid line).

dependent effects and which will be worth investigating in more detail in the future.

As the results show (Fig. 3), the *EMOS* model for daily precipitation sums slightly outperforms the SAMOS models, which is somehow unpleasant. A possible reason is that the overall (not location-dependent) bias and slope correction is of most importance and that this simple model is better able to correct for it. A second reason could be that the underlying observation climatology (which is an all-year climatology; Appendix A) might not perfectly capture the cold season and causes the slightly worse predictive performance of the SAMOS models. Further improvements of the underlying climatology might be beneficial for the predictive skill of the SAMOS results.

One of the biggest advantages of the proposed hybrid approach is that forecasts can be produced on the same temporal scale as the current EPS even if the underlying data sets used for model training (hindcasts and observations) are available on coarser temporal scales or even different timescales for different variables. This allows one to combine the best information from (location-)independent sources to get the most reliable probabilistic predictions possible. For the present study, two observation networks have been combined, one providing long-term daily precipitation records, and one providing temporally highly resolved temperature measurements.

Overall the 2 m temperature and precipitation forecasts serve as a good proxy for probabilistic snowfall forecasts, which is the main target variable of this study. The results show very promising results in terms of calibration and reliability of both the expected amount of precipitation and fresh snow, but also the probability of observing snowfall at an hourly temporal resolution.

Code and data availability. The main parts of this study are based on *R* package bamlss (Umlauf et al., 2017) to compute the spatio-temporal observation climatologies and *R* package crch (Messner et al., 2016) to estimate the (censored) non-homogeneous regression models. The continuous ranked probability scores are based on *R* package scoringRules (Jordan et al., 2018).

Observations from the hydrographical service (BMLFUW, 2018) can be downloaded from the website of the Bundesministerium für Land und Forstwirtschaft und Wasserwirtschaft (http://ehyd.gv.at, last access: 14 November 2018).

Appendix A: Standardized anomaly model output statistics (SAMOS)

For spatio-temporal ensemble postprocessing we followed the approach of Dabernig et al. (2017) and Stauffer et al. (2017b), which we summarize in the following. In contrast to other statistical postprocessing methods, SAMOS uses standardized anomalies for both the response and the covariates. This allows one to remove location-specific and time-specific characteristics from the data and to estimate one single regression model for all stations and forecast lead times at once. For this study we closely follow the original articles (Dabernig et al., 2017; Stauffer et al., 2017b) but slightly modify the specification, especially for the temperature SAMOS, to adapt to the different study area.

Observation climatologies. Two separate spatio-temporal models have been estimated for 2 m air temperature observations and daily precipitation sums. Both models have effects to capture seasonal, altitudinal, and spatial climatological features represented by (multi-dimensional) nonlinear functions. The 2 m temperature observations are available at an hourly temporal resolution. Therefore, additional nonlinear cyclic effects have to be included to capture the diurnal effects in the climatological estimates.

The spatio-temporal model for the 2 m temperature uses the geographical location (longitude lon, latitude lat, and altitude alt), the "hour of the day" (hour), and the "day of the year" (doy) as covariates and is specified as follows:

$$
\begin{aligned}
&\text{temperature} \sim \mathcal{N}\left(\widetilde{\boldsymbol{\mu}}_y, \widetilde{\boldsymbol{\sigma}}_y\right),\\
&\widetilde{\boldsymbol{\mu}}_y = f_1(\text{hour}, \text{doy}, \text{alt})\\
&\quad + f_2(\text{hour}, \text{doy}) + f_3(\text{doy}, \text{lon}, \text{lat})\\
&\quad + f_4(\text{hour}) + f_5(\text{doy}) + f_6(\text{doy}, \text{alt}) + f_7(\text{alt}),\\
&\log\left(\widetilde{\boldsymbol{\sigma}}_y\right) = g_1(\text{hour}, \text{doy}, \text{alt}) + g_2(\text{hour}, \text{doy})\\
&\quad + g_3(\text{doy}, \text{lon}, \text{lat})\\
&\quad + g_4(\text{hour}) + g_5(\text{doy}) + g_6(\text{doy}, \text{alt}) + g_7(\text{alt}),
\end{aligned}
\tag{A1}
$$

where $f_\bullet$ and $g_\bullet$ are up to three-dimensional smooth spline effects. Cyclic P-splines are used for all effects depending on the "day of the year" or the "hour of the day"; all other effects use penalized thin plate splines with a varying number of possible degrees of freedom. Following the same concept, the spatio-temporal model for daily precipitation sums is defined as

$$
\begin{aligned}
&\text{precipitation}^{1/p} \sim \mathcal{L}_0\left(\widetilde{\boldsymbol{\mu}}_y, \widetilde{\boldsymbol{\sigma}}_y\right),\\
&\widetilde{\boldsymbol{\mu}}_y = f_1(\text{alt}) + f_2(\text{doy}) + f_3\left(\text{lon}, \text{lat}\right)\\
&\quad + f_4(\text{doy}, \text{lon}, \text{lat}),\\
&\log\left(\widetilde{\boldsymbol{\sigma}}_y\right) = g_1(\text{alt}) + g_2(\text{doy}) + g_3(\text{lon}, \text{lat})\\
&\quad + g_4(\text{doy}, \text{lon}, \text{lat}).
\end{aligned}
\tag{A2}
$$

As for Eq. (A1), cyclic P splines are used for effects which depend on the "day of the year", while all others use penalized thin plate splines. The major difference to the temperature climatology (Eq. A1) is that a left-censored logistic response distribution $\mathcal{L}_0$ is used on power-transformed observations of precipitation$^{1/p}$ ($p = 1.35$; cf. Stauffer et al., 2017b). The complexity of the linear predictors in Eq. (A2) is lower than in Eq. (A1) as no effects for diurnal variation have to be considered.

Model climatology. Similar spatio-temporal climatologies as for the observations could be estimated for all quantities from the EPS which are used as covariates in the SAMOS models. This would have to be done for each quantity separately using a reasonably large data set of historical EPS forecasts. However, we instead extract the model climatologies directly from ECMWF hindcasts. These hindcasts are produced operationally twice a week and consist of $10 + 1$ members using the same model version and model specification as the current EPS. For each hindcast run the forecasts for the same date over the most recent 20 years are computed. The hindcasts are designed to represent the climatology of the current EPS model and are used to calibrate EPS forecasts and as input for postprocessing applications (e.g., Hagedorn et al., 2012, 2008). For our SAMOS approach we can thus simply derive the empirical mean and empirical standard deviation over a set of hindcasts to get the climatological estimates $\widetilde{\boldsymbol{\mu}}_x$ and $\widetilde{\boldsymbol{\sigma}}_x$ required to compute the standardized anomalies for covariate $\mathbf{x}$ (Eq. 9). Climatologies for lead times when no hindcast output is available (between the regular 6 h interval) are created using simple grid-point-wise linear interpolation.

Hindcasts are produced every Monday and Thursday (available Tuesday/Friday), computed 2 weeks in advance. Taking hindcasts for ± 2 weeks around the date of interest yields eight independent hindcast runs with 11 members and 20 years of (re-)forecasts each, which yields $8 \cdot 11 \cdot 20 = 1760$ forecasts. With this large number of independent predictions these climatological estimates are fairly robust. Due to the 4-week centered rolling window the climatologies automatically adapt themselves to the prevailing season. Separate climatologies for each forecast step are required to capture diurnal cycles (for temperature) and to account for changes in the model climate with increasing forecast horizon such as drifting means or increasing ensemble standard deviation. Thus, for this study, 13 separate climatologies for the temperature models ([+6 h, +12 h, …, +72 h, +78 h]) and 3 climatologies for the precipitation forecasts ([+30 h, +54 h, +78 h]) are required.

Estimation of the SAMOS models (see Table 1). Equations (1)–(3) and (5)–(7) show the basic heteroscedastic models used for *SAMOS_hom* and *SAMOS_het*. The only modification is to replace the response $\boldsymbol{y}$ and the covariate $\mathbf{x}$ with the corresponding standardized anomaly $\boldsymbol{y}^*$ and $\mathbf{x}^*$ (Eq. 9). For

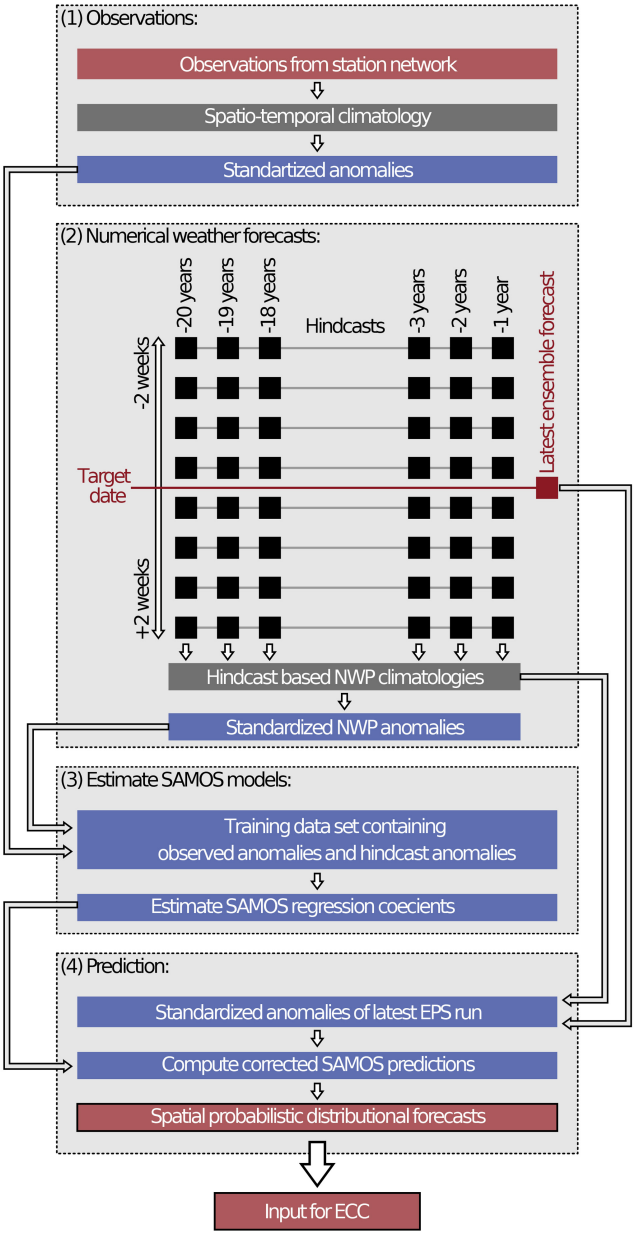

Figure A1. Schematic concept of the SAMOS postprocessing based on ECMWF hindcasts (black), ECMWF EPS forecasts (red), and observations (orange). Background climatologies (gray) are used to convert the data from the physical scale into standardized anomalies (blue) used to estimate the regression coefficients of the SAMOS postprocessing method. The SAMOS correction can be applied to the standardized anomalies of a new EPS forecast to obtain spatial or spatio-temporal probabilistic forecasts (full distribution). These results are used as input for the ECC approach.

the *xSAMOS_het* model the linear predictors in Eqs. (1)–(3) are extended by simply adding additional covariates, resulting in a multilinear SAMOS model.

Once the regression coefficients of the SAMOS model have been estimated, future ensemble forecasts can be corrected by first computing standardized anomalies using the same model climatology as for model training and correcting the standardized anomalies of the ensemble forecast using the estimated SAMOS models. As the outcomes (μ_i^* and σ_i^*) are on the standardized anomaly scale, they have to be rescaled with respect to the observation climatology to obtain physical values (e.g., °C or mm). The final predictive distribution is thus

$$y_i \sim \mathcal{D}\left(\mu_i^* \cdot \widetilde{\sigma}_{y,i} + \widetilde{\mu}_{y,i},\ \sigma_i^* \cdot \widetilde{\sigma}_{y,i}\right)^p, \tag{A3}$$

where $\mathcal{D}$ represents the normal distribution $\mathcal{N}$ in the case of 2 m temperature postprocessing with $p = 1$ and $\mathcal{L}_0$ in the case of the power-transformed daily precipitation sums' postprocessing with $p = 1.35$.

Algorithm 1 presents pseudo-code for all steps. The same is shown in Fig. A1 as a graphical representation of this procedure, visualizing the required data sets, the required steps, and their dependencies.

Appendix B: Individual copula members

Figures B1 and B2 show the individual EPS members (Fig. B1) and the corresponding re-weighted ensemble copula coupling members (Fig. B2) for the +48 h forecast for 10 March 2017 10:00 UTC as used to derive the probabilities and expectation plotted in Figs. 10 and 11. For easier comparison the NWP forecasts are bilinearly interpolated to $\sim 500 \times 500\,\mathrm{m}^2$ to match the resolution of the postprocessed predictions.

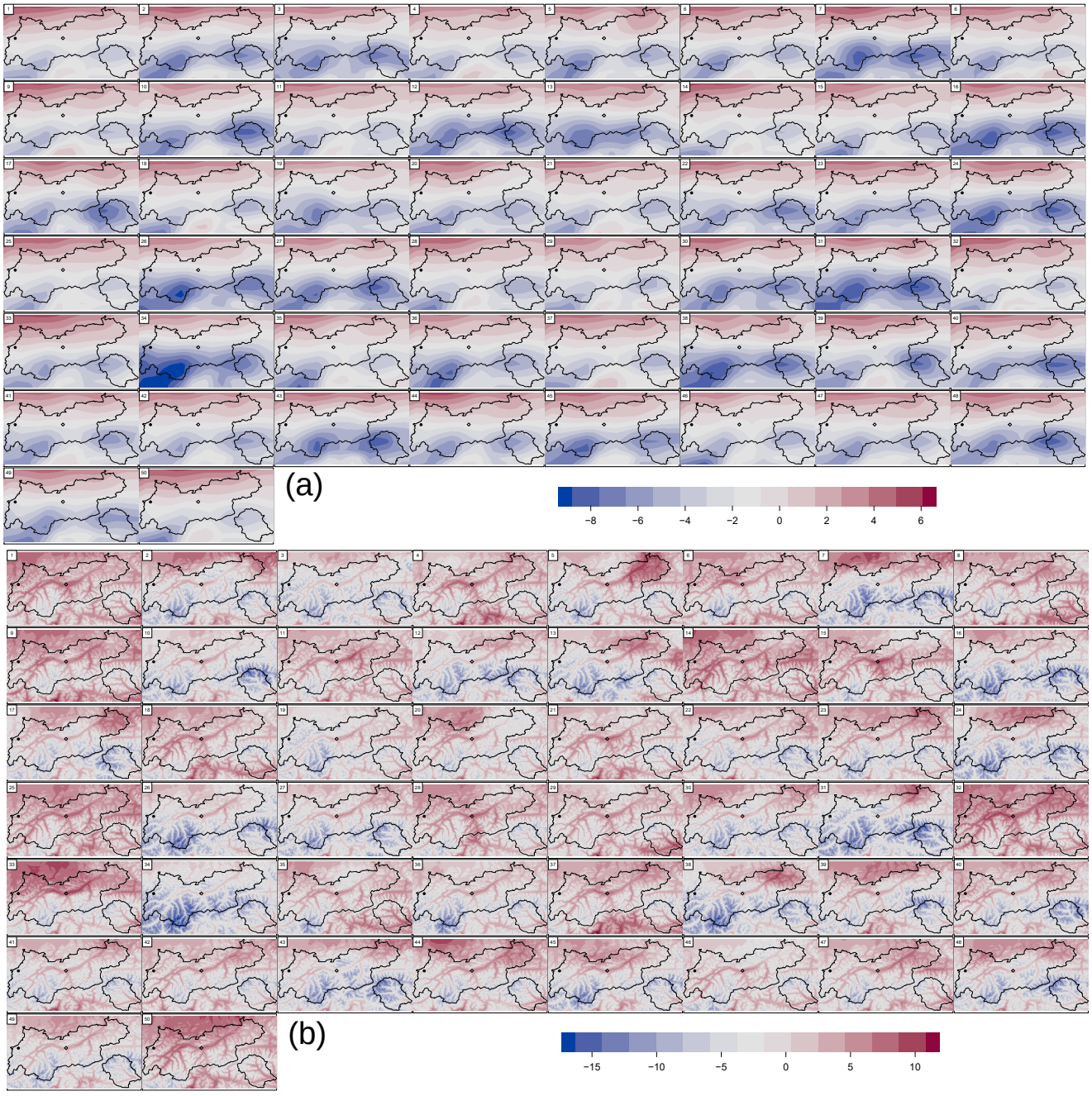

Figure B1. Stamps for +48 h forecasts initialized 8 March 2017 00:00 UTC (valid for 10 March 2017 00:00 UTC). Individual EPS members for 2 m temperature **(a)** and the corresponding copula members **(b)**. Please note that the color scale for all members of one type (EPS/copula) is identical, but the scales between the raw EPS and the results from the postprocessing differ.

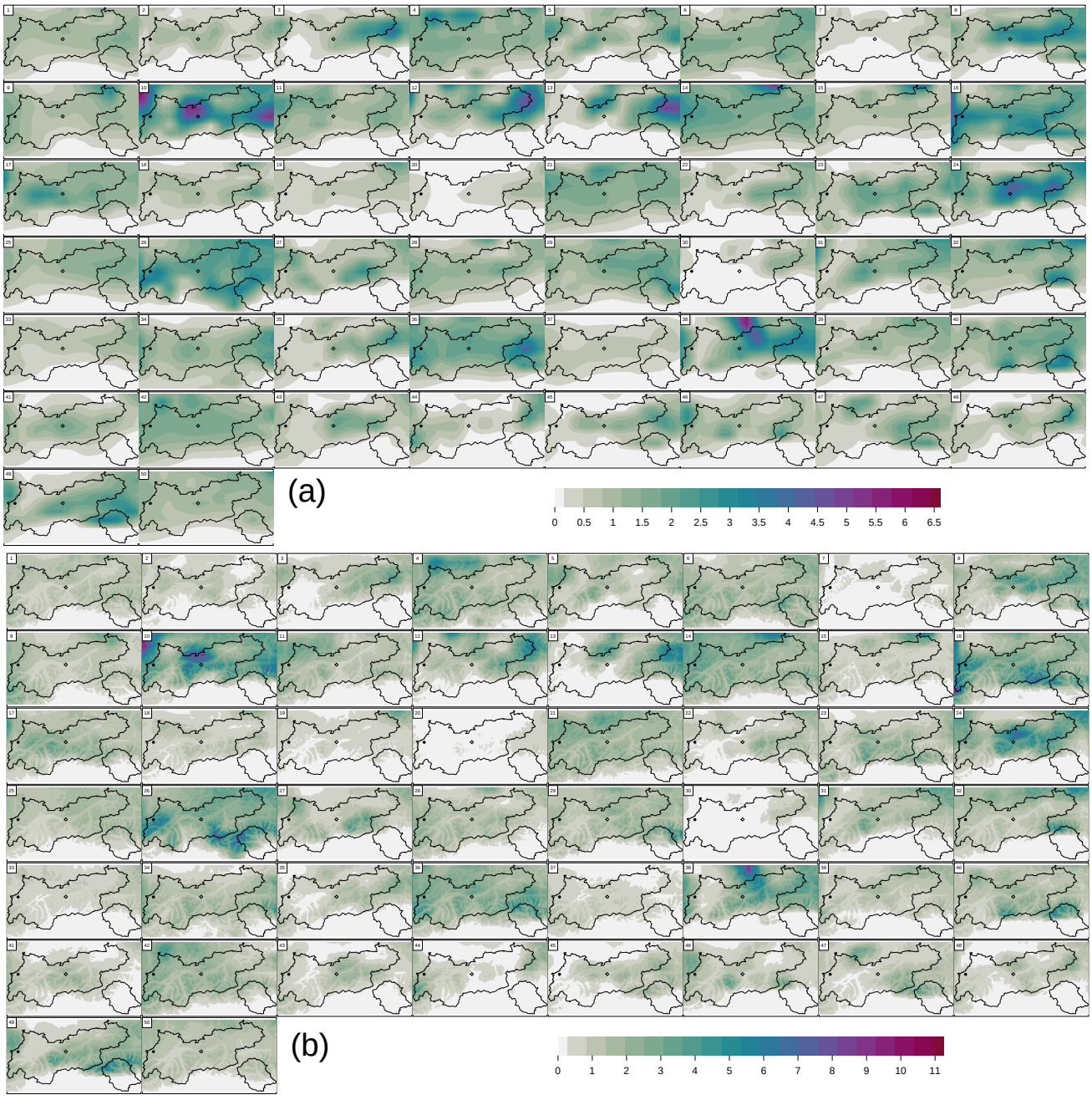

Figure B2. Stamps for +48 h forecasts initialized 8 March 2017 00:00 UTC (valid for 10 March 2017 00:00 UTC). Individual EPS members for 1 h precipitation sums **(a)** and the corresponding copula members **(b)**. Please note that the color scale for all members of one type (EPS/copula) is identical, but the scales between the raw EPS and the results from the postprocessing differ.

Author contributions. This study summarizes the ideas developed within our most recent research project by all the members, including RS, GJM, JWM, and AZ. The majority of the work for this study was performed by RS. The statistical models are, to a large extent, based on the two R packages bamlss and crch developed by JWM and AZ (and others). All the authors closely worked together discussing the results and findings and commented on this paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This project was partially funded by the Austrian Science Fund (FWF), grant TRP 290, and the Austrian Research Promotion Agency (FFG), grant no. 858537. The data sets are provided by the Zentralanstalt für Meteorologie und Geodynamik Vienna (ZAMG; https://zamg.ac.at, last acccess: 14 November 2018) and the Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW), Abteilung IV/4 – Wasserhaushalt (http://ehyd.gv.at, last acccess: 14 November 2018).

Edited by: Christopher Paciorek
Reviewed by: two anonymous referees

References

Amt der Tiroler Landesregierung: Statistisches Jahrbuch Bundesland Tirol, https://www.tirol.gv.at/fileadmin/themen/statistik-budget/statistik/downloads/Statistisches_Handbuch_2014.pdf (last access: 15 July 2016), 2014.

Bellaire, S., Jamieson, J. B., and Fierz, C.: Forcing the snow-cover model SNOWPACK with forecasted weather data, The Cryosphere, 5, 1115–1125, https://doi.org/10.5194/tc-5-1115-2011, 2011.

BMLFUW: Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW), Abteilung IV/4 – Wasserhaushalt, available at: http://ehyd.gv.at, last access: 14 November 2018.

Bouallègue, Z. B. and Theis, S. E.: Spatial Techniques Applied to Precipitation Ensemble Forecasts: from Verification Results to Probabilistic Products, Meteorol. Appl., 21, 922–929, https://doi.org/10.1002/met.1435, 2014.

Bröcker, J. and Smith, L. A.: Increasing the Reliability of Reliability Diagrams, Weather Forecast., 22, 651–661, https://doi.org/10.1175/WAF993.1, 2007.

Dabernig, M., Mayr, G. J., Messner, J. W., and Zeileis, A.: Spatial Ensemble Post-Processing with Standardized Anomalies, Q. J. Roy. Meteor. Soc., 143, 909–916, https://doi.org/10.1002/qj.2975, 2017.

Fraley, C., Raftery, A. E., and Gneiting, T.: Calibrating Multimodel Forecast Ensembles with Exchangeable and Missing Members Using Bayesian Model Averaging, Mont. Weather Rev., 138, 190–202, https://doi.org/10.1175/2009MWR3046.1, 2010.

Gebetsberger, M., Messner, J. W., Mayr, G. J., and Zeileis, A.: Fine-Tuning Non-Homogeneous Regression for Probabilistic Precipitation Forecasts: Unanimous Predictions, Heavy Tails, and Link Functions, Mon. Weather Rev., 145, 4693–4708, https://doi.org/10.1175/MWR-D-16-0388.1, 2017.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, J. Am. Stat. Assoc., 102, 359–378, https://doi.org/10.1198/016214506000001437, 2007.

Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T.: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation, Mon. Weather Rev., 133, 1098–1118, https://doi.org/10.1175/MWR2904.1, 2005.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A.: Assessing Probabilistic Forecasts of Multivariate Quantities, with an Application to Ensemble Predictions of Surface Winds, Test, 17, 211–235, https://doi.org/10.1007/s11749-008-0114-x, 2008.

Hagedorn, R., Hamill, T. M., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures, Mon. Weather Rev., 136, 2608–2619, https://doi.org/10.1175/2007MWR2410.1, 2008.

Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., and Palmer, T. N.: Comparing TIGGE Multimodel Forecasts with Reforecast-Calibrated ECMWF Ensemble Forecasts, Q. J. Roy. Meteor. Soc., 138, 1814–1827, https://doi.org/10.1002/qj.1895, 2012.

Hamill, T. M.: Interpretation of Rank Histograms for Verifying Ensemble Forecasts, Mon. Weather Rev., 129, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2, 2001.

Hamill, T. M., Whitaker, J. S., and Mullen, S. L.: Reforecasts: An Important Dataset for Improving Weather Predictions, B. Am. Meteorol. Soc., 87, 33–46, https://doi.org/10.1175/BAMS-87-1-33, 2006.

Hamill, T. M., Scheuerer, M., and Bates, G. T.: Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses, Mon. Weather Rev., 143, 3300–3309, https://doi.org/10.1175/MWR-D-15-0004.1, 2015.

Jordan, A., Krüger, F., and Lerch, S.: Evaluating Probabilistic Forecasts with scoringRules, J. Stat. Softw., https://jstatsoft.org, forthcoming, 2018.

Judson, A. and Doesken, N.: Density of Freshly Fallen Snow in the Central Rocky Mountains, B. Am. Meteorol. Soc., 81, 1577–1587, https://doi.org/10.1175/1520-0477(2000)081<1577:DOFFSI>2.3.CO;2, 2000.

Knox, T., Gerhold, L., and Ulbrich, U.: Perception and Use of Uncertainty in Severe Weather Warnings by Emergency Services in Germany, Atmos. Res., 158–159, 292–301, https://doi.org/10.1016/j.atmosres.2014.02.024, 2015.

Lawinenwarndienst Tirol: Winterberichte 2009/2010 bis 2015/2016, available at https://lawine.tirol.gv.at/archiv/winterberichte/ (accessed: 14 November 2017), 2009–2017.

Lerch, S. and Thorarinsdottir, T.: Comparison of Non-Homogeneous Regression Models for Probabilistic Wind Speed Forecasting, Tellus A, 65, 21206, https://doi.org/10.3402/tellusa.v65i0.21206, 2013.

Meister, R.: Density of New Snow and its Dependence on Air Temperature and Wind, Zürcher Geographische Schriften, 23, 73–79, 1985.

Messner, J. W., Mayr, G. J., Wilks, D. S., and Zeileis, A.: Extending Extended Logistic Regression: Extended versus Separate versus Ordered versus Censored, Mon. Weather Rev., 142, 3003–3014, https://doi.org/10.1175/MWR-D-13-00355.1, 2014a.

Messner, J. W., Mayr, G. J., Zeileis, A., and Wilks, D. S.: Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance, Mon. Weather Rev., 142, 448–456, https://doi.org/10.1175/MWR-D-13-00271.1, 2014b.

Messner, J. W., Mayr, G. J., and Zeileis, A.: Heteroscedastic Censored and Truncated Regression with crch, The R Journal, 8, 173–181, https://journal.r-project.org/archive/2016-1/messner-mayr-zeileis.pdf (last access: 14 November 2018), 2016.

Mullen, S. L. and Buizza, R.: Quantitative Precipitation Forecasts over the United States by the ECMWF Ensemble Prediction System, Mon. Weather Rev., 129, 638–663, https://doi.org/10.1175/1520-0493(2001)129<0638:QPFOTU>2.0.CO;2, 2001.

Murphy, A. H.: A New Vector Partition of the Probability Score, J. Appl. Meteorol., 12, 595–600, https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2, 1973.

Neal, R. A., Boyle, P., Grahame, N., Mylne, K., and Sharpe, M.: Ensemble Based First Guess Support Towards a Risk-based Severe Weather Warning Service, Meteorol. Appl., 21, 563–577, https://doi.org/10.1002/met.1377, 2014.

Palmer, T. N.: The Economic Value of Ensemble Forecasts as a Tool for Risk Assessment: From Days to Decades, Q. J. Roy. Meteor. Soc., 128, 747–774, https://doi.org/10.1256/0035900021643593, 2002.

Raftery, A. E.: Use and Communication of Probabilistic Forecasts, Stat. Anal. Data Min., 9, 397–410, https://doi.org/10.1002/sam.11302, 2016.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, B. Am. Meteorol. Soc., 93, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1, 2012.

Roebber, P. J., Bruening, S. L., Schultz, D. M., and Cortinas Jr., J. V.: Improving Snowfall Forecasting by Diagnosing Snow Density, Weather Forecast., 18, 264–287, https://doi.org/10.1175/1520-0434(2003)018<0264:ISFBDS>2.0.CO;2, 2003.

Rohregger, J. B.: Methoden zur Bestimmung der Schneefallgrenze, Master's thesis, Universität Wien, Austria, 2008.

Roulston, M. S. and Smith, L. A.: Combining Dynamical and Statistical Ensembles, Tellus A, 55, 16–30, https://doi.org/10.1034/j.1600-0870.2003.201378.x, 2003.

Schefzik, R., Thorarinsdottir, T. L., and Gneiting, T.: Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling, Stat. Sci., 28, 616–640, https://doi.org/10.1214/13-STS443, 2013.

Scheuerer, M.: Probabilistic Quantitative Precipitation Forecasting Using Ensemble Model Output Statistics, Q. J. Roy. Meteor. Soc., 140, 1086–1096, https://doi.org/10.1002/qj.2183, 2014.

Scheuerer, M. and Büermann, L.: Spatially Adaptive Post-Processing of Ensemble Forecasts for Temperature, J. R. Stat. Soc. C-Appl., 63, 405–422, https://doi.org/10.1111/rssc.12040, 2014.

Scheuerer, M. and Hamill, T. M.: Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions, Mon. Weather Rev., 143, 4578–4596, https://doi.org/10.1175/MWR-D-15-0061.1, 2015.

Sloughter, J. M. L., Raftery, A. E., Gneiting, T., and Fraley, C.: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging, Mon. Weather Rev., 135, 3209–3220, https://doi.org/10.1175/MWR3441.1, 2007.

Stauffer, R., Mayr, G. J., Messner, J. W., Umlauf, N., and Zeileis, A.: Spatio-Temporal Precipitation Climatology over Complex Terrain Using a Censored Additive Regression Model, Int. J. Climatol., 37, 3264–3275, https://doi.org/10.1002/joc.4913, 2017a.

Stauffer, R., Umlauf, N., Messner, J. W., Mayr, G. J., and Zeileis, A.: Ensemble Postprocessing of Daily Precipitation Sums over Complex Terrain Using Censored High-Resolution Standardized Anomalies, Mon. Weather Rev., 145, 955–969, https://doi.org/10.1175/MWR-D-16-0260.1, 2017b.

Thorarinsdottir, T. L. and Gneiting, T.: Probabilistic Forecasts of Wind Speed: Ensemble Model Output Statistics by Using Heteroscedastic Censored Regression, J. R. Stat. Soc. A Stat., 173, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x, 2010.

Umlauf, N., Klein, N., and Zeileis, A.: BAMLSS: Bayesian Additive Models for Location, Scale and Shape (and Beyond), J. Comput. Graph. Stat., 27, 612–627, https://doi.org/10.1080/10618600.2017.1407325, 2017.

Wilks, D. S.: Extending Logistic Regression to Provide Full-Probability-Distribution MOS Forecasts, Meteorol. Appl., 16, 361–368, https://doi.org/10.1002/met.134, 2009.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Myln, E. K.: The Economic Value of Ensemble-Based Weather Forecasts, B. Am. Meteorol. Soc., 83, 73–83, https://doi.org/10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2, 2002.