University of Innsbruck

Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis

Working Papers in Economics and Statistics

2016-21

Research Platform eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

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

Reto Stauffer
Universität Innsbruck

Georg J. Mayr
Universität Innsbruck

Jakob W. Messner
Universität Innsbruck

Nikolaus Umlauf
Universität Innsbruck

Achim Zeileis
Universität Innsbruck

Abstract

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

Keywords: ensemble post-processing, distributional regression, precipitation, complex terrain, censoring.

1. Introduction

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

Precipitation forecasts, or weather forecasts in general, are typically provided by numerical

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

To correct for systematic errors and to correct the uncertainty provided by the EPS, post-processing methods are often applied. A variety of ensemble post-processing methods for precipitation are available nowadays, such as analog methods (Hamill, Whitaker, and Mullen 2006; Hamill, Scheuerer, and Bates 2015), ensemble dressing (Roulston and Smith 2003), Bayesian model averaging (BMA; Sloughter, Raftery, Gneiting, and Fraley 2007; Fraley, Raftery, and Gneiting 2010), extended logistic regression (Wilks 2009; Ben Bouallègue and Theis 2014; Messner, Mayr, Zeileis, and Wilks 2014b), or non-homogeneous regression (Gneiting, Raftery, Westveld III, and Goldman 2005). Several extensions exist for non-normally distributed variables (Thorarinsdottir and Gneiting 2010; Lerch and Thorarinsdottir 2013; Scheuerer 2014; Scheuerer and Hamill 2015). For precipitation, Messner, Mayr, Wilks, and Zeileis (2014a) shows that a censored logistic regression fits well, while Scheuerer (2014) and Scheuerer and Hamill (2015) use a left-censored generalized extreme value distribution (GEV), or a left-censored shifted gamma distribution, respectively.

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

In this article, a new spatial statistical post-processing method for daily precipitation sums over complex terrain is presented. Due to topographical influences large differences can be observed between two neighboring stations. These differences are driven by sub-grid processes which cannot yet be resolved by global EPS models. To account for these small-scale spatial variabilities among all stations we are using an adapted version of the anomaly approach first published by Scheuerer and Büermann (2014) and further extended by Dabernig, Mayr, Messner, and Zeileis (2016). Observations and ensemble forecasts are transformed into standardized anomalies by subtracting the long-term climatological mean and dividing by the climatological standard deviation. This allows to remove station-dependent characteristics from the data and to fit one single regression model for all stations at once. As the model does not rely on site-specific characteristics anymore the corrections can be applied to future ensemble forecasts and to create probabilistic forecasts for any arbitrary location within the area of interest.

Following Dabernig *et al.* (2016) we use the *Standardized Anomaly Model Output Statistics* (SAMOS) approach and extend the framework to fulfill all requirements needed for precipitation post-processing. SAMOS offers a simple and computationally efficient framework for fully probabilistic spatial post-processing and is applied to the ECMWF ensemble in combination with the ECMWF reforecasts. The approach presented qualifies for an operational system, as no extensive archive of historical forecasts is required, and it automatically adapts to the

latest ECMWF ensemble model version or any EPS model providing similar reforecasts.

2. Area of Interest and Data

2.1. Study Area

To develop and validate the new method presented in this study we focus on the governmental area of Tyrol, Austria. Tyrol has a size of about 12500 km^2 and is home to approximately 740000 inhabitants (Statistik Austria 2016) living in the two separated parts with North Tyrol on the north side of the main Alpine ridge, and East Tyrol south of the main Alpine ridge. The study area is located in the Eastern Alps showing a highly complex topography. Figure 1 shows the state borders of Tyrol and the topography reaching from 465–3798 m a.m.s.l. including some of the highest mountains in Austria. Due to the high population density and the strong economic focus on tourism ($>$ 10 million tourists in 2014; Land Tirol 2014) there is a high demand for accurate weather forecasts.

2.2. Observational Data

The local hydrographical service provides a dense precipitation measurement network, from which 117 stations in Tyrol and its surrounding will be used for model training and validation spanning September 1971 trough the end of 2012. The mean distance to the four closest stations in the surrounding is only about 10 kilometers. Locations of the observation sites are highlighted in Figure 1. The hydrographical service performs rigorous quality controls on the observations and makes them freely available for any non-commercial use on the maintainres website (BMLFUW 2016).

2.3. Numerical Weather Forecast Data

The numerical forecasts are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) including the operational ensemble (ENS; 00 UTC initial) which consists of 50+1 individual forecasts based on perturbed initial conditions (50 forecasts plus control run), and the ECMWF reforecast data set. The ECMWF reforecast data set exists since February 18, 2010 and was slightly extended over the years. Until June 14, 2012 the reforecast was computed once a week, providing ensemble reforecasts consisting of 4+1 members for the most recent 18 years. From June 21, 2012 through the end of 2012 the number of years was extended to 20. The ECMWF reforecast is designed to always provide the model climate of the latest ECMWF ENS version, often used for model calibration (e.g., Hamill, Hagedorn, and Whitaker 2008; Hamill 2012).

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

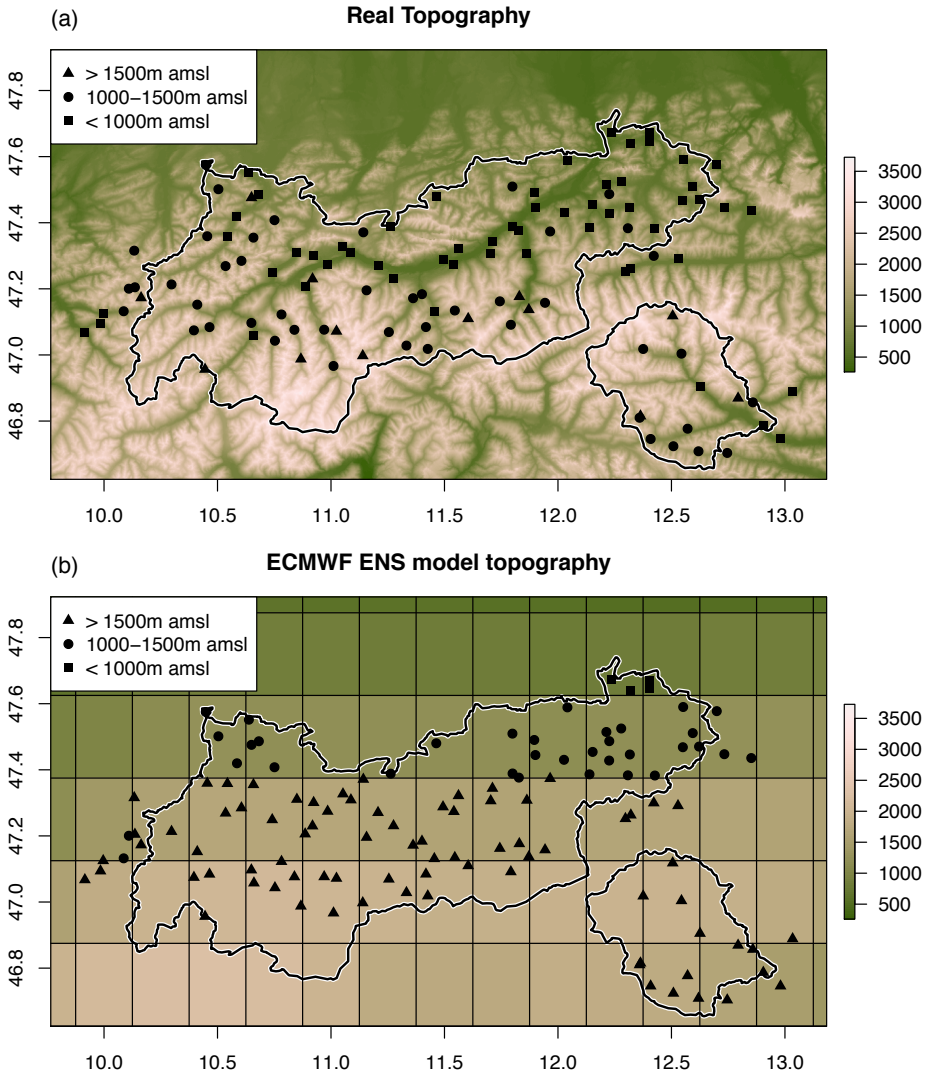

Figure 1: The black line shows the state borders of Tyrol, Austria. Each marker represents an observation site (totally 117), the marker type indicates the altitude: square ($\leq 1000m$ a.m.s.l.), bullet ($1000 - 1500m$ a.m.s.l.), and triangle ($\geq 1500m$ a.m.s.l.) with respect to the underlying topography. The background shows the **(a)** real topography (CGIAR-CSI 2016), and the **(b)** ECMWF EPS model topography with a 0.5 degree resolution as used between February 2010 and December 2012.

2.4. Training- and Verification Data Set

The ECMWF reforecasts are used to (i) compute the climatology of the ECMWF ensemble which will be used as a background information, and to (ii) train the statistical post-processing including the most recent 4 reforecast runs centered around the current date. Therefore, the model climatology is based on 4 runs $\times$ 5 members $\times$ 20 years = 400 individual forecasts (details in Sec. 3.3). For the training data set the reforecasts are bi-linearly interpolated to each of the 117 observation sites. Out of each interpolated reforecast ensemble (day-wise, 4+1 members) the mean and standard deviation is then used to estimate the regression coefficients. We use the most recent four 00 UTC forecast runs yielding to a training sample of up to 4 runs$\times$20 years = 80 data pairs per station, or 4 runs$\times$20 years$\times$117 stations = 9360 observation/reforecast pairs for the full spatial SAMOS (details in Sec. 4.2).

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

Due to the availability of the observations (Sec. 2.2), and the ECMWF reforecasts (Sec. 2.3), the time period between February 26, 2010, and December 31, 2012 will be used for verification with an overall data availability of 99.4% and roughly 120500 unique observation/forecast pairs.

3. Methodology

3.1. Censored Non-homogeneous Logistic Regression (CNLR)

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

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

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

on.

Both parameters (μ, σ) are expressed by a linear predictor including the covariates, or explanatory variables. As suggested by Gneiting *et al.* (2005), the mean of the ensemble forecast drives the location μ, and the standard deviation of the ensemble drives the scale σ. For this study, we only use the forecasted daily accumulated total precipitation from the ensemble (Sec. 2.3) as covariate. In Equation 1 "m" denotes the mean, "s" the standard deviation of the forecasted power-transformed daily total precipitation amounts of the ensemble members.

In addition, a second covariate z has been included. z is a binary split variable which takes 1 if all forecast-members in the training data set predict less than 0.1 $mm\ day^{-1}$ ($z = 1$, "no" precipitation), and 0 otherwise. This allows to handle dry and wet cases differently and has a positive impact on the results. It furthermore solves the problem of taking the log of $s = 0$. The log-transformation on the scale σ is used to ensure non-negative scale values during optimization. The full *CNLR* assumptions can then be written as:

$$\begin{aligned} precipitation &= \begin{cases} 0 & \text{if } \ y \leq 0 \\ y^p & \text{else} \end{cases} \\ y &\sim \mathcal{L}(\mu, \sigma) \\ \mu &= \beta_0 + \beta_1 \cdot z + \beta_2 \cdot m \cdot (1 - z) \\ \log(\sigma) &= \gamma_0 + \gamma_1 \cdot \log(s) \cdot (1 - z) \end{aligned} \tag{1}$$

In case of a dry ensemble forecast ($z = 1$) the linear predictors collapse to $\mu = \beta_0 + \beta_1$ and $\log(\sigma) = \gamma_0$ such that the model only consists of two estimated constants. For wet cases ($z = 0$) the linear predictors yield to $\mu = \beta_1 + \beta_2 \cdot m$ and $\log(\sigma) = \gamma_0 + \gamma_1 \cdot \log(s)$, which corresponds to the NGR model proposed by Gneiting *et al.* (2005). These assumptions allow to correct the bias, but also a possible overdispersion or underdispersion of the ensemble as the scale σ depends on the predicted ensemble standard deviation.

The model as specified in Equation 1 can be applied at every arbitrary location where both, historical observations and historical ensemble forecasts are available. For point-wise ensemble post-processing, *one CNLR model* has to be fitted at *each observation site*. In this case all *CNLR* models are independent and have its own regression coefficients $\beta_\bullet$ and $\gamma_\bullet$. As these coefficients are site-specific, spatial predictions are not directly possible and would require an additional interpolation method which allows to account for supplementary covariates, such as terrain or surface properties.

Instead of a two-step approach of performing station-wise estimates and interpolate/extrapolate the resulting coefficients afterwards, we will extend the model to include the training data of all stations at once, and fit one simple and computationally efficient model for fully probabilistic spatial estimates.

3.2. Standardized Anomaly Model Output Statistics (SAMOS)

The statistical method presented in this article is based on the anomaly approach first published by Scheuerer and Büermann (2014) and further extended by Dabernig *et al.* (2016) focusing on temperature forecasts across Germany and Northern Italy, respectively. We extend the Standardized Anomaly Model Output Statistics (SAMOS) approach by Dabernig *et al.* (2016) yielding to a censored SAMOS version for precipitation post-processing.

Climatological properties between two precipitation observation sites may vary in mean (location) and variability (scale). This is especially true over complex terrain where only a few kilometers between a valley and a mountain station can result in very large climatological differences (Frei and Schär 1998; Isotta *et al.* 2014; Stauffer, Messner, Mayr, Umlauf, and Zeileis 2016). These small scale features influence daily precipitation sums, but are not yet fully resolved by global numerical ensembles (ENS). Therefore, a high-resolution *spatio-temporal climatology* is used as *background information* to provide small-scale features at any location within the study area. Instead of modeling the relationship between past observations and past numerical weather forecasts directly, the statistical model uses high-resolution *standardized anomalies*. Anomalies are defined as the short term deviation from the local long-term climate. These anomalies can be divided by the local climatological variability to obtain *standardized* anomalies. Standardized anomalies of the observations (*precipitation*) are defined as:

$$y^* = \frac{precipitation^{\frac{1}{p}} - \mu_{obs,clim}}{\sigma_{obs,clim}} \tag{2}$$

Where $\mu_{obs,clim}$ and $\sigma_{obs,clim}$ describe the long-term climatological properties of daily observations, and will be discussed in detail in the next section. y^* denotes the resulting latent response on the standardized anomaly scale which follow a standard logistic distribution $\mathcal{L}(0,1)$. Equivalent to Equation 2, standardized anomalies of the ensemble forecasts (*ens*) can be computed using the climatological properties $\mu_{ens,clim}$, and $\sigma_{ens,clim}$ of the ensemble:

$$ens^* = \frac{ens^{\frac{1}{p}} - \mu_{ens,clim}}{\sigma_{ens,clim}} \tag{3}$$

The ensemble climatology ($\mu_{ens,clim}$, $\sigma_{ens,clim}$) is described in the next section.

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

Combining the *NCLR* model from Equation 1 with the concept of standardized anomalies (Eqns. 2 & 3) leads to the full specification of the *Standardized Anomaly Model Output Statistics* (SAMOS) model with a left-censored logistic response:

$$\begin{aligned} y^* &\sim \mathcal{L}(\mu^*, \sigma^*) \\ \mu^* &= \beta_0 + \beta_1 \cdot z + \beta_2 \cdot m^* \cdot (1 - z) \\ \log(\sigma^*) &= \gamma_0 + \gamma_1 \cdot \log(s^*) \cdot (1 - z) \end{aligned} \tag{4}$$

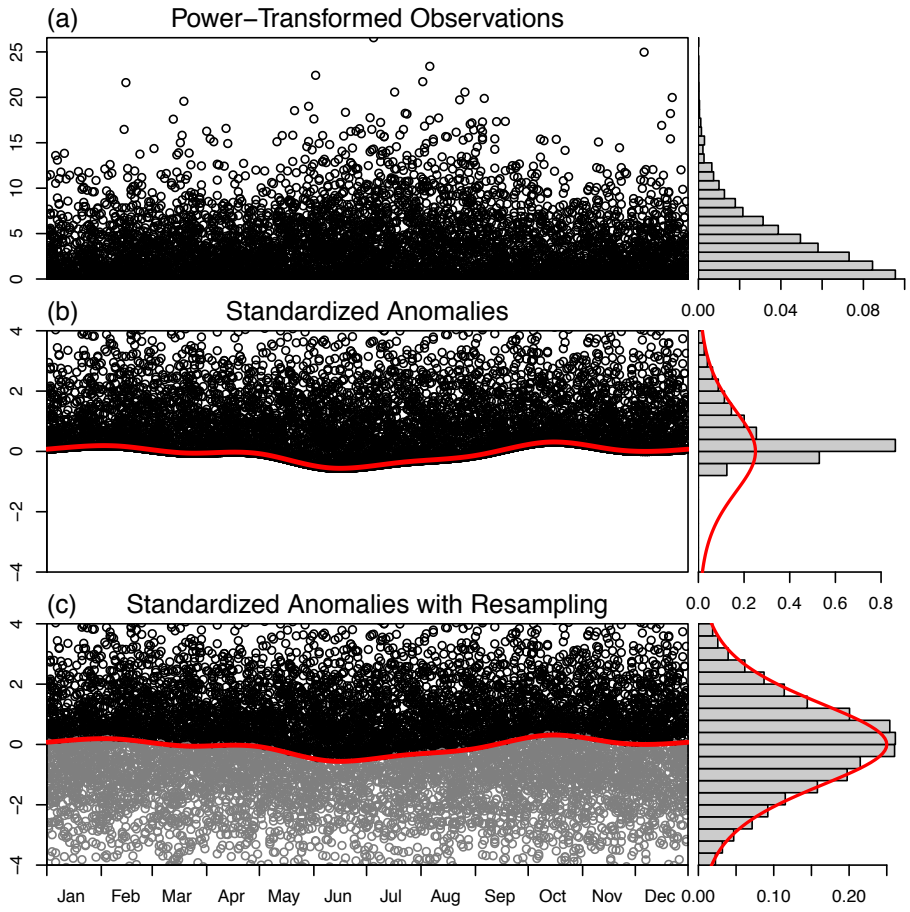

Figure 2: Example of standardized anomalies for one specific station (Bromberg) with roughly 8500 unique daily observations between 1987 and 2013. **(a)** daily observations on power-transformed scale ($mm^{\frac{1}{p}}\ day^{-1}$ with $p = 1.35$); **(b)** standardized anomalies; **(c)** standardized anomalies with resampled censored data (for visual justification only). Left column: data plotted against the day of the year. The red solid line in **(b)** & **(c)** shows the shifted censoring point due to standardization. Resampled censored observations are shown in gray in Figure **(c)**. Right column: density histograms. For **(b)** & **(c)** the standard logistic distribution is shown in red ($\mathcal{L}(0, 1)$).

The standardized anomalies y^* are still assumed to follow a logistic distribution. The linear predictors for location μ^*, and $\log(\sigma^*)$ on the standardized anomaly scale depend on the standardized ensemble anomalies (ens^*) and the binary split indicator z. Covariates "m^*" and "s^*" are the empirical mean and standard deviation of the standardized ensemble forecast anomalies (Eqn. 3).

Once the standardized anomalies are known, the regression coefficients of the SAMOS model given by Equations 2–4 can be estimated using maximum likelihood optimization, as offered by the *R* package **crch** (Messner, Mayr, and Zeileis 2016), or similar software. The climatological estimates required to create the standardized anomalies are explained in detail in Section 3.3.

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

$$\mathcal{L}_0\Big(\underbrace{\mu^* \cdot \sigma_{obs,clim} + \mu_{obs,clim}}_{\text{location}}, \quad \underbrace{\sigma^* \cdot \sigma_{obs,clim}}_{\text{scale}} \Big) \tag{5}$$

The descaled zero left-censored distribution $\mathcal{L}_0(\dots)$ describes the full post-processed ENS forecast distribution on the power-transformed scale. Since the SAMOS regression coefficients are location independent, the post-processed predictions can be computed at any location within the study area where both, ENS forecasts and climatological estimates ($\mu_{\bullet,clim}$, $\sigma_{\bullet,clim}$) are available. We are using spatio-temporal climatologies (details in the next section), wherefore the only limitation for the post-processed ENS forecasts is the horizontal grid spacing of the spatial climatology, which itself only depends on the resolution of the available digital elevation model (see Stauffer *et al.* 2016). From the full-probabilistic SAMOS forecasts different properties can then be derived such as the mean or expectation, quantiles, probability of precipitation, or probabilities to exceed a certain threshold. To retrieve the corrected forecasts on the original scale in $mm\ day^{-1}$, the inverse power-transformation has to be taken into account. Details can be found in Appendix A.

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

3.3. Climatological Estimates

The climatological properties $\mu_{\bullet,clim}$ and $\sigma_{\bullet,clim}$ for both the observations and the ensemble forecasts have to be specified to be able to derive the standardized anomalies y^* and ens* (Eqns. 2 and 3). The computation of the observed climatology is based on Stauffer *et al.* (2016) but uses a left-censored logistic instead of Gaussian distribution and consequently a modified power-transformation parameter. Using the same optimization method as in Stauffer *et al.* (2016) led to an optimal power parameter of $p = 1.35$ for the logistic distribution for this study.

The observed spatio-temporal climatology is based on all 117 stations (Fig. 1) and uses daily precipitation measurements from 1971 through the end of 2009 yielding to roughly 1.5 million individual observations. Data from the years 2010–2013 are set aside for verification.

The climatology is based on a non-homogeneous regression model similar to the SAMOS method. In contrast to Equations 1 and 4, the linear predictors of the climatological model include smooth one-dimensional and multi-dimensional spline effects to depict all features of the climatology. In addition to the global intercepts (β, γ) an altitudinal effect (s_1, t_1), an effect to describe the seasonality based on the day of the year (s_2, t_2), a spatial effect on dependent longitude and latitude (s_3, t_3), and a three-dimensional effect to describe spatial variations in the seasonal pattern (s_4, t_4) are included. Further details can be found in Stauffer *et al.* (2016). The full model specification of the observation-climatology can be expressed as:

$$\begin{aligned} \mathit{precipitation} &= \begin{cases} 0 & \text{if } \; y \leq 0 \\ y^p & \text{else} \end{cases} \\ y &\sim \mathcal{L}\big(\mu_{obs,clim}, \sigma_{obs,clim}\big) \\ \mu_{obs,clim} &= \beta + s_1(\text{alt}) + s_2(\text{yday}) + s_3(\text{lon}, \text{lat}) + s_4(\text{yday}, \text{lon}, \text{lat}) \\ \log(\sigma_{obs,clim}) &= \gamma + t_1(\text{alt}) + t_2(\text{yday}) + t_3(\text{lon}, \text{lat}) + t_4(\text{yday}, \text{lon}, \text{lat}) \end{aligned} \tag{6}$$

Again, both parameters of the power-transformed left-censored logistic distribution (location $\mu_{obs,clim}$ and scale $\sigma_{obs,clim}$) are modeled. This is required, as they are used for the standardization of the SAMOS model. Although the climatology model (Sec. 3.3) is quite complex, estimation only takes about 30 hours and has to be done only rarely, e.g., once a year.

In addition to climatological estimates of the observations, climatological estimates $\mu_{ens,ens}$, and $\mu_{ens,ens}$ are required to compute standardized anomalies of the ensemble forecasts as in Equation 3. The two parameters represent the long-term climatology of the ECMWF EPS system (Sec. 2.3) and are computed from the ECMWF reforecast data set. Mean and standard deviation are based on up to 400 individual forecasts provided by the most recent four reforecast runs (Sec. 2.4):

$$\begin{aligned} \mu_{ens,clim} &= \text{mean}(\text{reforecast}) \\ \sigma_{ens,clim} &= \text{stdv}(\text{reforecast}) \cdot C \end{aligned} \tag{7}$$

The climatological location $\mu_{ens,clim}$ is simply the empirical mean, the climatological scale $\sigma_{ens,clim}$ is the "standard deviation" of the reforecast used. The factor $C = \sqrt{3}/\pi$ is used to get the empirical scale of a logistic distribution, to be on the same scale as the observed climatological scale $\sigma_{obs,clim}$ (Eqn. 6).

4. Results and Verification

4.1. SAMOS Results

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

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

The example shows that SAMOS is able to add interpretable and meaningful features to the ENS during the post-processing procedure. However, the performance cannot be evaluated with a single case alone. The next section therefore contains a detailed analysis and verification on a three year independent data set.

4.2. Verification

For verification, the predictions of four different methods will be compared with unused (out-of-sample) data between February 2010 and December 2012. As two baseline methods, the climatologies (CLIM; Section 3.3), and the raw ECMWF ensemble predictions (ENS) will be used. Furthermore, a station-wise post-processing (STN) is included, based on Equation 1. For STN a separate *CNLR* model is estimated for each of the 117 stations in the data set.

The predictions of all methods are out-of-sample such that the data used for verification are not included in the training data set which is used to estimate the regression coefficients. CLIM is based on all available observations, except that the years 2010–2013 are excluded (Sec. 3.3). Therefore, CLIM predictions are spatially in-sample, but temporally out-of-sample. STN is using the latest four available reforecast runs yielding to spatially in-sample, but temporally out-of-sample predictions. SAMOS is the only method whose predictions can be verified both spatially and temporally out-of-sample. Therefore a leave-one-out cross validation is performed. For each station, the SAMOS regression coefficients were estimated based on the most recent four reforecast runs but excluding this one specific station. The

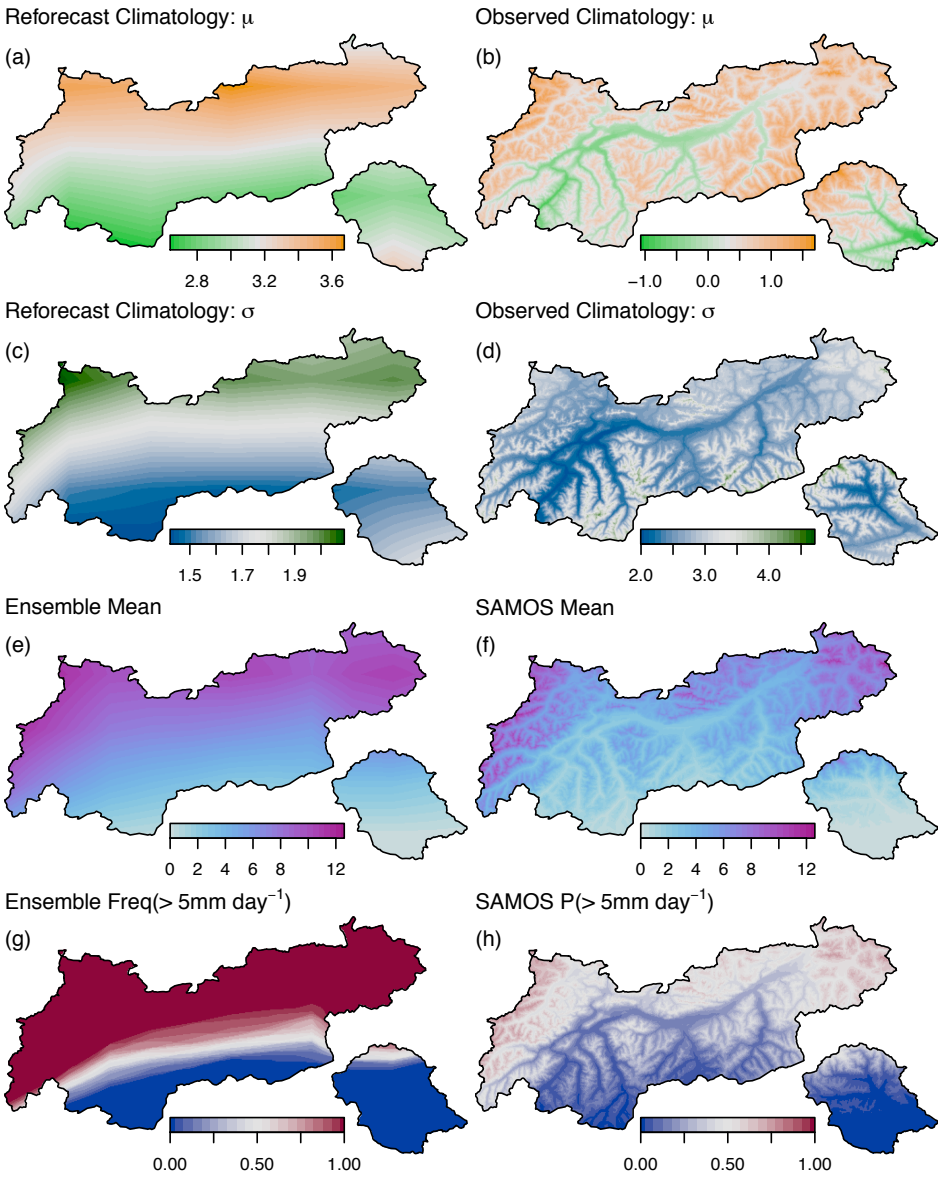

Figure 3: Example prediction for May 18, 2010, one-day-ahead forecast. Left column shows reforecast climatologies and the raw ensemble forecast; right column shows the observed climatology, and the post-processed SAMOS predictions. Top down: climatological location (μ; **(a)**, **(b)**), climatological scale (σ; **(c)**, **(d)**), forecast mean (**(e)**, **(f)**), and frequency **(g)** and probability **(h)** of exceeding 5 $mm\ day^{-1}$. Location μ and scale σ on the latent power-transformed scale in $mm^{\frac{1}{p}}$ with $p = 1.35$. Note that **(a)**/**(b)** and **(c)**/**(d)** use different color scales regarding the range of the data.

Method Description and **Abbreviation**	(i) Spatially o.o.s.	(ii) Temporally o.o.s.	(iii) Spatial Prediction
Climatology: **CLIM**	*no*	yes	yes
ECMWF ensemble: **ENS**	*no*	yes	yes
Station-wise: **STN**	*no*	yes	*no*
Spatial SAMOS: **SAMOS**	yes	yes	yes

Table 1: Summary of all 4 methods used for verification in Section 4.2. Column 2 & 3 indicate whether the results in the verification are (i) Spatially out-of-sample, and/or (ii) Temporally out-of-sample (o.o.s.). The last column shows whether (iii) the method provides spatial predictions, or not.

forecasts were then made for the excluded station only. Table 1 contains a summary of all four methods and shows their sample behavior.

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

In addition to the CRPS, Figure 5 shows the Brier scores for three different thresholds. The most right box-whisker shows the Brier score of CLIM. CLIM does not depend on the forecast horizon, but only on the day of the year and is therefore valid for all forecast periods shown in the boxes left of it. A Brier score of 0 would indicate a perfect forecast, while 0.5 is as good as a coin flip. For threshold 0 $mm\ day^{-1}$ (precipitation yes/no) it can be seen that the ENS performs poorly, even worse than the climatology. This indicates a wet bias in the ENS. Both post-processing methods perform significantly better than the climatology. Overall, SAMOS shows the best performance, even for long forecast horizons. Figure 5*b* & 5*c* show the same verification for 1 $mm\ day^{-1}$, and 10 $mm\ day^{-1}$ respectively. For these thresholds ENS is better than CLIM but is outperformed by the post-processing methods. For large thresholds (Fig. 5*c*) and large forecast horizons all methods are similar. Differences between them are no longer significant.

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

The verification rank histogram assesses the calibration of discrete distributions as provided

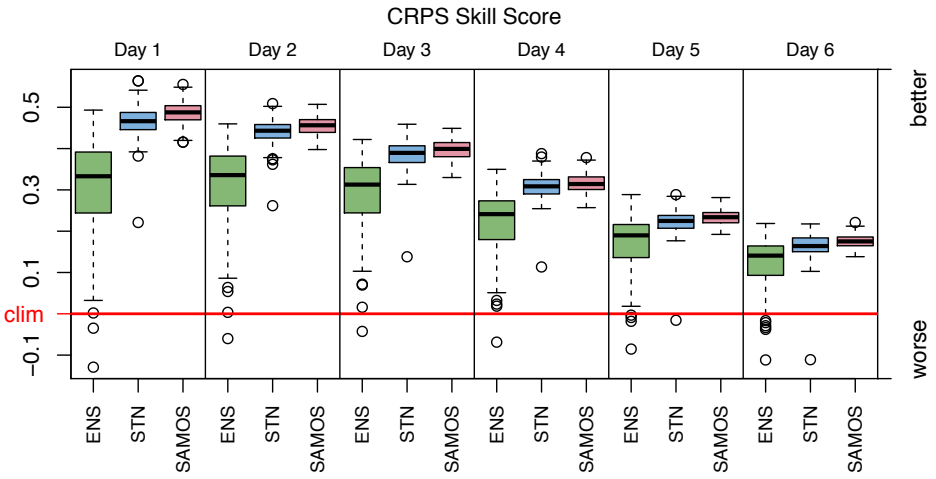

Figure 4: Continuous rank probability score (CRPS) shown as a skill score with climatology (from Eqn. 6) as reference. The boxes from left to right show the model performance for one-day-ahead to six-day-ahead forecasts. Each box contains three box-whisker plots for the raw **ENS** (green; left), and the two post-processing methods **STN** (blue; middle) and **SAMOS** (red, right). Each one contains 117 station-wise mean skill scores. The box shows the upper and lower quartile, the whiskers the 1.5 interquartile range. Additionally, the median (black bar) and the outliers (circles) are plotted. Values below 0 indicate stations with less skill than the climatology. The higher the values, the better the performance of the method.

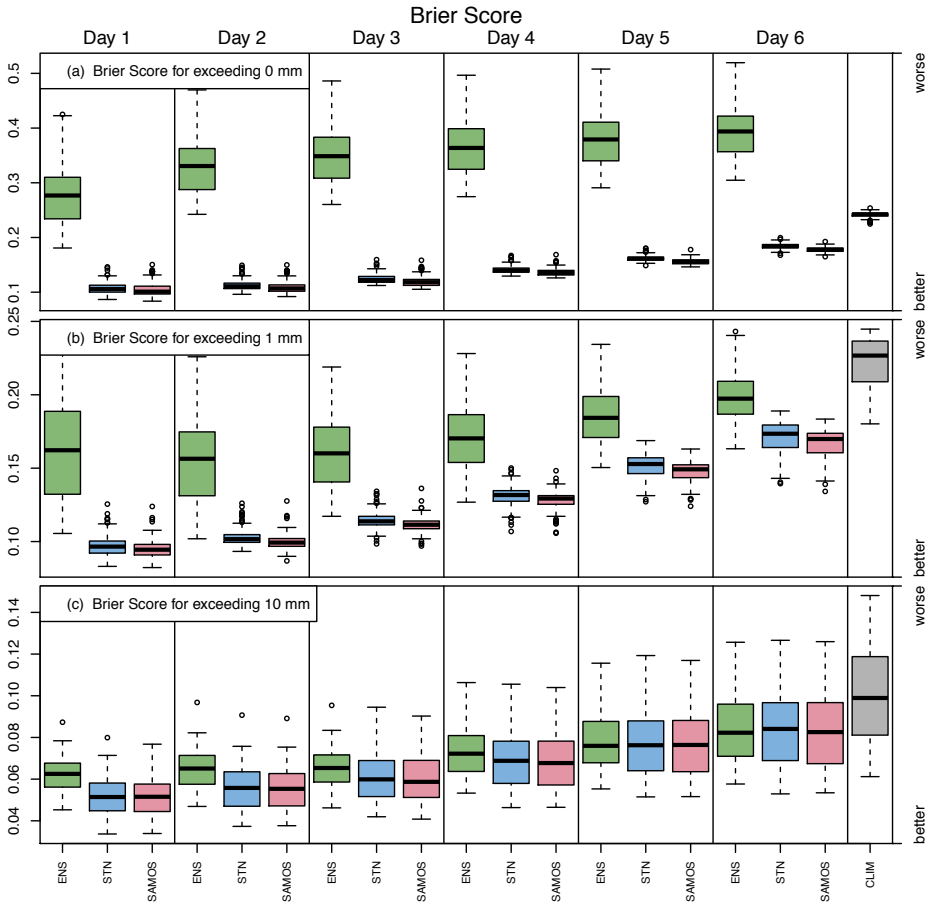

Figure 5: Brier scores (BS) for three different thresholds (top down): 0 $mm\ day^{-1}$, 1 $mm\ day^{-1}$, and 25 $mm\ day^{-1}$. The specification of the box-whiskers as in Figure 4. For the ENS (green) the frequency is used, for the two post-processing methods STN (blue) and SAMOS (red) the probabilities are derived from the predicted distribution. From left to right: scores for one-day-ahead to six-day-ahead forecasts, the last box shows the climatological forecast (gray) depending on the day of the year only, but not on the forecast horizon. Lower BS indicates a better performance.

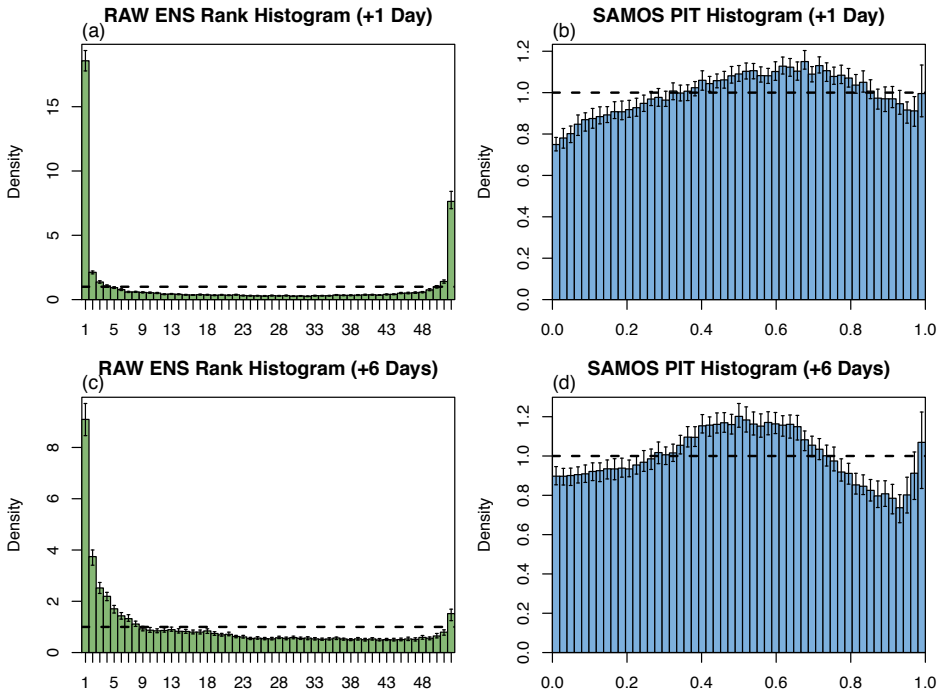

Figure 6: Rank histograms of the RAW ensemble (**(a)**, **(c)**), PIT histograms of the SAMOS forecasts (**(b)**, **(d)**). Top row: one-day-ahead forecasts; bottom row: six-day-ahead forecast. The error bars show the 95% confidence intervals of a 100-times day-wise random bootstrap. Rank histogram: 52 ranks (50+1 ensemble members). The *concave* shape indicates underdispersion. To have a similar look, the PIT histogram shows 52 bins, each of width $\frac{1}{52}$ (first bin: $(\frac{0}{52} - \frac{1}{52}\%]$, second $(\frac{1}{52} - \frac{2}{52}\%]$, and so on). The *convex* shape indicates slight overdispersion.

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

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

5. Discussion and Conclusion

In this study, the SAMOS model has been extended and applied to daily precipitation sums. It has been shown that the concept of using standardized anomalies (Scheuerer and Büermann 2014; Dabernig *et al.* 2016) can be used to correct precipitation forecasts of numerical ensemble forecast models. The Standardized Anomaly Model Output Statistics (SAMOS) post-processing method is able to create accurate spatial predictions of daily precipitation sums over complex terrain. SAMOS uses high-resolution spatial climatologies as background information to transform the data (observations and ensemble forecasts) into standardized anomalies. This (i) removes location-dependent climatological features from the data, and (ii) brings all data to a comparable level, which allows to account for the small-scale features in the study area which are not yet resolved by the ensemble model.

To create the standardized anomalies, daily estimates of the climatological mean (location $\mu_{\bullet,clim}$) and variability (scale $\sigma_{\bullet,clim}$) are required. For the observations, we use high-resolution spatio-temporal climatologies as presented by Stauffer *et al.* (2016). The climatology of the ECMWF ensemble model is provided by the ECMWF reforecast data set.

Once both climatologies are known, the observations and the ensemble forecasts can be converted into standardized anomalies such that all data follow a standard logistic distribution. As all location-dependent characteristics are removed this allows to apply one simple regression model including all data at once. This model directly returns fully probabilistic predictions for any arbitrary location within the study area, even for regions without observational sites.

The results show that the spatial SAMOS outperforms the station-wise non-homogeneous regression models (STN), even if the SAMOS predictions are (unlike STN) spatially out-of-sample. This is mainly related to the training data set. While STN only includes interpolated forecasts of one location, the SAMOS training data set includes the data of all stations leading to more robust estimates. The SAMOS calibration indicates that the assumed response distribution is not optimal. A different distribution might improve the skill and remove the need of the power transformation (Scheuerer 2014; Hamill *et al.* 2015).

The goal of this study is to use the SAMOS approach proposed by (Dabernig *et al.* 2016) and to extend the method for the application of precipitation sums, or censored responses in

general. While only focusing on daily precipitation sums up to day six in this study it would be worth to extend the forecast horizon and the study area, but also to include additional covariates and to apply the SAMOS approach to other meteorological parameters.

As the estimation of the SAMOS requires only little computational time, the SAMOS can easily be re-fitted as soon as a new reforecast run is available. This ensures that the SAMOS automatically adapts itself to the latest ECMWF ensemble model version within a very short transition period. Nowadays, the ECMWF reforecast (ECMWF 2016) is run twice a week providing 10+1 members, which could further improve the performance of the SAMOS. However, since the observation data set was only available through the end of 2012 this could not have been tested yet.

6. Acknowledgements

Ongoing project funded by the Austrian Science Fund (FWF): TRP 290-N26. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC). Observation data set provided by the hydrographical service Tyrol (`ehyd.gv.at`).

Appendix A

Properties of the Power-Transformed Left-Censored Logistic Distribution

The probability density function λ and the cumulative distribution function Λ of a non-censored logistic distribution are defined as:

$$\lambda(x|\mu,\sigma) = \frac{\exp(-\frac{x-\mu}{\sigma})}{\sigma\big(1+\exp(-\frac{x-\mu}{\sigma})\big)^2} \tag{8}$$

$$\Lambda(x|\mu,\sigma) = \frac{1}{1+\exp(-\frac{x-\mu}{\sigma})} \tag{9}$$

The density λ_0 and distribution function Λ_0 of a zero left-censored logistic distribution including the power-transformation $\frac{1}{p}$ can then be written as:

$$\lambda_0(x_i|\mu_i,\sigma_i,p) = \begin{cases} 0 & \text{for } \; x_i < 0 \\ \Lambda(0 \mid \mu_i,\sigma_i) & \text{for } \; x_i = 0 \\ \lambda(x_i^{1/p} \mid \mu_i,\sigma_i) & \text{else} \end{cases} \tag{10}$$

$$\Lambda_0(x_i|\mu_i,\sigma_i,p) = \begin{cases} 0 & \text{for all: } \; x_i < 0 \\ \Lambda(x_i^{1/p} \mid \mu_i,\sigma_i) & \text{else} \end{cases} \tag{11}$$

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

$$P(x_i > 0 \mid \mu_i,\sigma_i) = 1 - P(x_i \leq 0) = 1 - \Lambda(0|\mu_i,\sigma_i) \tag{12}$$

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

$$P(x_i > \kappa \mid \mu_i,\sigma_i,p) = 1 - P(x_i \leq \kappa) = 1 - \Lambda(\kappa^{\frac{1}{p}}|\mu_i,\sigma_i) \tag{13}$$

Furthermore, the expectation of the distribution on the original scale has to be evaluated. The expectation on the original scale "x" in $mm\ day^{-1}$ can be retrieved using:

$$E[x|\mu_i,\sigma_i,p] = \int_0^\infty x \cdot \lambda(x^{\frac{1}{p}}|\mu_i,\sigma_i) \cdot \frac{x^{(\frac{1}{p}-1)}}{p} dx \tag{14}$$

A last property of interest is the median of the distribution, again on the original scale "x" in $mm\ day^{-1}$. Parameter μ in Equations 1 & 4 describes the latent unobservable location. The median is then given as:

$$\text{median}(x|\mu,p) \begin{cases} 0 & \text{for } \; \mu \leq 0 \\ \mu^p & \text{else} \end{cases} \tag{15}$$

Appendix B

Error Measures used for Verification

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

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{0}^{\infty}\Big(\Lambda_0(x^{\frac{1}{p}}|\mu_i,\sigma_i) - H(x)\Big)^2 dx \tag{16}$$

Where x is the response variable on the original scale in $mm\ day^{-1}$, N is the number of forecasts included, Λ_0 the CDF of the forecasted distribution (Eqn. 11), and H the CDF of the observation represented by a heavy-side step function which takes 0 for all $x <$ observation, and 1 otherwise. While x is on the original scale ($mm\ day^{-1}$), both distributional parameters, location μ and scale σ, are on the power-transformed scale. Therefore, the power transformation $\frac{1}{p}$ is required to evaluate the CDF Λ_0. As no analytical solution has been found, the CRPS is evaluated by quantile sampling with $n = 2000$.

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

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}}\ , \tag{17}$$

where the CRPS of the method to test is in the numerator, the CRPS of the reference method in the denominator. Values below 0 indicate that the tested method performs worse than the reference. CPRSS values can take values in the range of $[-\infty, 1]$.

As a second measure the Brier score (BS) is shown to verify the skill of the forecast probabilities. One of the most frequently used threshold for precipitation forecasts is 0 mm, also known as the probability of precipitation. As an ensemble system does not provide a fully probabilistic forecasts, the frequency is used as an estimator of the probability. As an example: if half of all ensemble members predict no precipitation, the other half does, the frequency is 0.5 and can be seen as a probability of $\sim 50\%$ if the number of ensemble members is sufficiently large. The Brier score can then be written as:

$$BS(\kappa) = \frac{1}{N}\sum_{i=1}^{N}\Big(P(x_i > \kappa \mid \mu_i,\sigma_i,p) - o_i(t)\Big)^2 \tag{18}$$

where N is again the number of forecasts included, P_i the predicted probability that an event exceeds threshold κ (Eqn. 13), and o_i the binary observation which takes 0 for all observations $x_i \leq \kappa$, and 1 otherwise.

References

Ben Bouallègue Z, Theis SE (2014). "Spatial Techniques Applied to Precipitation Ensemble Forecasts: from Verification Results to Probabilistic Products." *Meteorological Applications*, **21**(4), 922–929. doi:10.1002/met.1435.

BMLFUW (2016). "Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW), Abteilung IV/4 – Wasserhaushalt." Available at http://ehyd.gv.at. Accessed February 29 2016.

Box GE, Cox DR (1964). "An analysis of transformations." *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252. URL http://www.jstor.org/stable/2984418.

Buizza R, Houtekamer PL, Pellerin G, Toth Z, Zhu Y, Wei M (2005). "A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems." *Monthly Weather Review*, **133**(5), 1076–1097. doi:10.1175/MWR2905.1.

CGIAR-CSI (2016). "SRTM 90m Digital Elevation Database v4.1." Available at http://srtm.csi.cgiar.org. Accessed February 29 2016.

Dabernig M, Mayr GJ, Messner JW, Zeileis A (2016). "Spatial Ensemble Post-Processing with Standardized Anomalies." *Working papers*, Atmospheric and Cryospheric Institute, University of Innsbruck. URL http://EconPapers.repec.org/RePEc:inn:wpaper:2016-08.

ECMWF (2016). "Re-Forecast for Medium and Extended Forecast Range." Available at http://www.ecmwf.int/en/forecasts/documentation-and-support/re-forecast-medium-and-extended-forecast-range. Accessed June 9 2016.

Epstein ES (1969). "Stochastic Dynamic Prediction." *Tellus*, **21**(6), 739–759. doi:10.1111/j.2153-3490.1969.tb00483.x.

Fraley C, Raftery AE, Gneiting T (2010). "Calibrating Multimodel Forecast Ensembles with Exchangeable and Missing Members Using Bayesian Model Averaging." *Monthly Weather Review*, **138**(1), 190–202. doi:10.1175/2009MWR3046.1.

Frei C, Schär C (1998). "A Precipitation Climatology of the Alps from High-Resolution Rain-Gauge Observations." *International Journal of Climatology*, **18**, 873–900. doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.

Gneiting T, Balabdaoui F, Raftery AE (2007). "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**(2), 243–268. doi:10.1111/j.1467-9868.2007.00587.x.

Gneiting T, Raftery AE, Westveld III AH, Goldman T (2005). "Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation." *Monthly Weather Review*, **133**(5), 1098–1118. doi:10.1175/MWR2904.1.

Hagedorn R, Buizza R, Hamill TM, Leutbecher M, Palmer TN (2012). "Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts." *Quarterly Journal of the Royal Meteorological Society*, **138**(668), 1814–1827. ISSN 1477-870X. doi:10.1002/qj.1895.

Hamill TM (2012). "Verification of TIGGE Multimodel and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States." *Monthly Weather Review*, **140**(7), 2232–2252. doi:10.1175/MWR-D-11-00220.1.

Hamill TM, Hagedorn R, Whitaker JS (2008). "Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation." *Monthly Weather Review*, **136**(7), 2620–2632. doi:10.1175/2007MWR2411.1.

Hamill TM, Scheuerer M, Bates GT (2015). "Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses." *Monthly Weather Review*, **143**(8), 3300–3309. doi:10.1175/MWR-D-15-0004.1.

Hamill TM, Whitaker JS, Mullen SL (2006). "Reforecasts: An Important Dataset for Improving Weather Predictions." *Bulletin of the American Meteorological Society*, **87**(1), 33–46. doi:10.1175/BAMS-87-1-33.

Hutchinson MF (1998). "Interpolation of Rainfall Data with Thin Plate Smoothing Splines – Part I: Two Dimensional Smoothing of Data with Short Range Correlation." *Journal of Geographic Information and Decision Analysis*, **2**, 168–185.

Isotta FA, Frei C, Weilguni V, Perčec Tadić M, Lassègues P, Rudolf B, Pavan V, Cacciamani C, Antolini G, Ratto SM, Munari M, Micheletti S, Bonati V, Lussana C, Ronchi C, Panettieri E, Marigo G, Vertačnik G (2014). "The Climate of Daily Precipitation in the Alps: Development and Analysis of a High-Resolution Grid Dataset from Pan-Alpine Rain-Gauge Data." *International Journal of Climatology*, **34**(5), 1657–1675. ISSN 1097-0088. doi:10.1002/joc.3794.

Land Tirol (2014). "Statistisches Handbuch Bundesland Tirol 2014." Available at https://www.tirol.gv.at/fileadmin/themen/statistik-budget/statistik/downloads/Statistisches_Handbuch_2014.pdf. Accessed June 6 2016.

Lerch S, Thorarinsdottir T (2013). "Comparison of non-homogeneous regression models for probabilistic wind speed forecasting." *Tellus A*, **65**. ISSN 1600-0870. doi:10.3402/tellusa.v65i0.21206.

Messner JW, Mayr GJ, Wilks DS, Zeileis A (2014a). "Extending Extended Logistic Regression: Extended versus Separate versus Ordered versus Censored." *Monthly Weather Review*, **142**(8), 3003–3014. doi:10.1175/MWR-D-13-00355.1.

Messner JW, Mayr GJ, Zeileis A (2016). "Heteroscedastic Censored and Truncated Regression with crch." *The R-Journal*, **17**(11), in press. URL https://journal.r-project.org/archive/accepted/messner-mayr-zeileis.pdf.

Messner JW, Mayr GJ, Zeileis A, Wilks DS (2014b). "Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance." *Monthly Weather Review*, **142**(1), 448–456. doi:10.1175/MWR-D-13-00271.1.

Mullen SL, Buizza R (2001). "Quantitative Precipitation Forecasts over the United States by the ECMWF Ensemble Prediction System." *Monthly Weather Review*, **129**(4), 638–663. doi:10.1175/1520-0493(2001)129<0638:QPFOTU>2.0.CO;2.

Roulston MS, Smith LA (2003). "Combining dynamical and statistical ensembles." *Tellus A*, **55**(1), 16–30. doi:10.1034/j.1600-0870.2003.201378.x.

Scheuerer M (2014). "Probabilistic Quantitative Precipitation Forecasting Using Ensemble Model Output Statistics." *Quarterly Journal of the Royal Meteorological Society*, **140**(680), 1086–1096. doi:10.1002/qj.2183.

Scheuerer M, Büermann L (2014). "Spatially Adaptive Post-Processing of Ensemble Forecasts for Temperature." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **63**(3), 405–422. doi:10.1111/rssc.12040.

Scheuerer M, Hamill TM (2015). "Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions." *Monthly Weather Review*, **143**(11), 4578–4596.

Sloughter JML, Raftery AE, Gneiting T, Fraley C (2007). "Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging." *Monthly Weather Review*, **135**(9), 3209–3220.

Statistik Austria (2016). "Bevölkerung." Available at http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/index.html. Accessed June 22 2016.

Stauffer R, Messner JW, Mayr GJ, Umlauf N, Zeileis A (2016). "Spatio-Temporal Precipitation Climatology over Complex Terrain Using a Censored Additive Regression Model." *Working papers*, Faculty of Economics and Statistics, University of Innsbruck. URL http://EconPapers.repec.org/RePEc:inn:wpaper:2016-07.

Stidd CK (1973). "Estimating the Precipitation Climate." *Water Resources Research*, **9**(5), 1235–1241. doi:10.1029/WR009i005p01235.

Thorarinsdottir TL, Gneiting T (2010). "Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **173**(2), 371–388. ISSN 1467-985X. doi:10.1111/j.1467-985X.2009.00616.x.

Wilks DS (2009). "Extending logistic regression to provide full-probability-distribution MOS forecasts." *Meteorological Applications*, **16**(3), 361–368. ISSN 1469–8080. doi:10.1002/met.134.

Affiliation:

Reto Stauffer, Jakob W. Messner
Department of Statistics
Faculty of Economics and Statistics
Universität Innsbruck
Universitätsstraße 15
6020 Innsbruck, Austria, *and*
Institute of Atmospheric and Cryospheric Sciences
Faculty of Geo- and Atmospheric Sciences
Universität Innsbruck
Innrain 52
6020 Innsbruck, Austria
E-mail: Reto.Stauffer@uibk.ac.at, Jakob.Messner@uibk.ac.at

Nikolaus Umlauf, Achim Zeileis
Department of Statistics
Faculty of Economics and Statistics
Universität Innsbruck
Universitätsstraße 15
6020 Innsbruck, Austria
E-mail: Nikolaus.Umlauf@uibk.ac.at, Achim.Zeileis@uibk.ac.at

Georg J. Mayr
Institute of Atmospheric and Cryospheric Sciences
Faculty of Geo- and Atmospheric Sciences
Universität Innsbruck
Innrain 52
6020 Innsbruck, Austria
E-mail: Georg.Mayr@uibk.ac.at

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *forthcoming in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

2015-16 **Niall Flynn, Christopher Kah, Rudolf Kerschbamer:** Vickrey Auction vs BDM: Difference in bidding behaviour and the impact of other-regarding motives

2015-15 **Christopher Kah, Markus Walzl:** Stochastic stability in a learning dynamic with best response to noisy play

2015-14 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** Measuring regional innovation in one dimension: More lost than gained?

2015-13 **Christoph Hauser, Gottfried Tappeiner, Janette Walde:** The roots of regional trust

2015-12 **Christoph Hauser:** Effects of employee social capital on wage satisfaction, job satisfaction and organizational commitment

2015-11 **Thomas Stöckl:** Dishonest or professional behavior? Can we tell? A comment on: Cohn et al. 2014, Nature 516, 86-89, "Business culture and dishonesty in the banking industry"

2015-10 **Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, Henk Kelderman:** Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees

2015-09 **Martin Halla, Gerald Pruckner, Thomas Schober:** The cost-effectiveness of developmental screenings: Evidence from a nationwide programme

2015-08 **Lorenz B. Fischer, Michael Pfaffermayr:** The more the merrier? Migration and convergence among European regions

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city *forthcoming in European Economic Review*

2015-06 **Martin Geiger, Wolfgang Luhan, Johann Scharler:** When do Fiscal Consolidations Lead to Consumption Booms? Lessons from a Laboratory Experiment *forthcoming in Journal of Economic Dynamics and Control*

2015-05 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket payments in the Austrian healthcare system - a distributional analysis

2015-04 **Rudolf Kerschbamer, Matthias Sutter, Uwe Dulleck:** How social preferences shape incentives in (experimental) markets for credence goods *forthcoming in Economic Journal*

2015-03 **Kenneth Harttgen, Stefan Lang, Judith Santer:** Multilevel modelling of child mortality in Africa

2015-02 **Helene Roth, Stefan Lang, Helga Wagner:** Random intercept selection in structured additive regression models

2015-01 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures for pharmaceuticals: Lessons from the Austrian household budget survey

University of Innsbruck

Working Papers in Economics and Statistics

2016-21

Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis

Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

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

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)