Statistical Post-Processing of Daily Precipitation Sums over Complex Terrain Using Censored Standardized Anomalies

Reto Stauffer, Georg J. Mayr, Jakob W. Messner, Nikolaus Umlauf, Achim Zeileis
Contact: `reto.stauffer@uibk.ac.at`

Introduction

Numerical Weather Forecasts

- Weather forecasts are typically provided by numerical forecast models which use the current observed state of the atmosphere to simulate the future weather by solving basic physical equations.
- Numerical forecasts often have errors due to simplifications and imperfect initial conditions.
- Ensemble systems provide several independent weather forecasts based on slightly modified initial states to obtain additional information about the forecast uncertainty (large spread $\rightarrow$ high uncertainty).
- Ensemble systems are not able to capture all possible error sources and are typically underdispersive (too low uncertainty).

Statistical Post-Processing: Distributional regression models correct for systematic errors in location (mean), and scale (uncertainty).

Goals & Challenges

Goals:

- **improve** forecast **skill** for **precipitation**
- computationally **inexpensive model**
- **full-distributional spatial** predictions

Challenges:

- **large fraction of zero-observations** (dry days)
- data **physically limited** to non-negative values
- distribution shows strong **positive skewness**
- **high** spatial **variability**

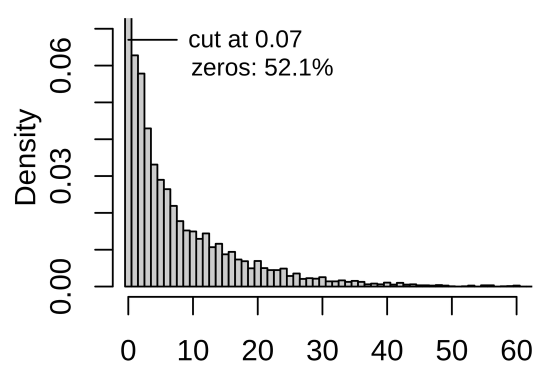

Figure 1: Histogram of observed precipitation sums for one station in [mm day^{-1}].

Standardized Anomaly Model Output Statistics

Standardized Anomalies

Use climatological background information to remove site specific characteristics and to bring all data to a comparable level.

$$y = \frac{obs^{\frac{1}{p}} - \mu_{obs,clim}}{\sigma_{obs,clim}}, \qquad x = \frac{ens^{\frac{1}{p}} - \mu_{ens,clim}}{\sigma_{ens,clim}} \tag{1}$$

obs, *ens*: Precipitation observations and forecasts (daily sums)
p: Parameter of the power transformation
y: Observations (response) on standardized anomaly scale
x: Ensemble forecasts on standardized anomaly scale
$\mu_{obs,clim}$, $\mu_{ens,clim}$: Climatological mean of observations (*obs*) and the ensemble model (*ens*)
$\sigma_{obs,clim}$, $\sigma_{ens,clim}$: Climatological standard deviation, corresponding to $\mu_{obs,clim}$, $\mu_{ens,clim}$

SAMOS

Fit one regression model for all stations at once. Distributional regression model with a power-transformed left-censored logistic distribution ($\mathcal{L}_{cens}$) to account for the zero-observations and skewness.

$$y \sim \mathcal{L}_{cens}(\mu, \sigma) \tag{2}$$
$$\mu = \beta_0 + \beta_1 \cdot (1 - z) + \beta_2 \cdot z \cdot \text{mean}(x) \tag{3}$$
$$\sigma = \gamma_0 + \gamma_1 \cdot z \cdot \log\left(\text{stdv}(x)\right) \tag{4}$$

μ, σ: Location and scale on the standardized anomaly scale
z: Binary split covariate. $z = 0$ if all ensemble forecasts $= 0$, and $z = 1$ otherwise
β, γ: Regression coefficients

Prediction

$$obs^{\frac{1}{p}} \sim \mathcal{L}_{cens}\left(\mu \cdot \sigma_{obs,clim} + \mu_{obs,clim}, \;\; \sigma \cdot \sigma_{obs,clim}\right) \tag{5}$$

In the limiting case that the ensemble would not provide any information: $\mu \rightarrow 0$, $\sigma \rightarrow 1$ and therefore the forecast would exactly be the climatology, the most reliable information in this case.

Study Area & Data

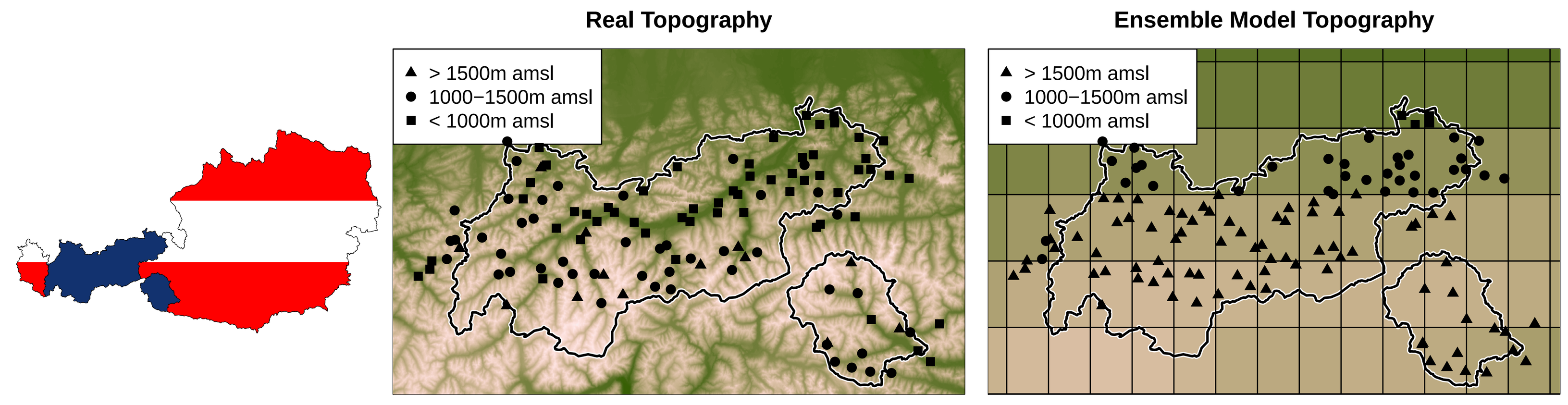

Figure 2: Real (90 m) and ensemble model topography ($\sim$ 40 km) of Tyrol, Austria. In addition: observation sites, marker types indicate the altitude with respect to the underlying topography.

Study Area: Tyrol, Austria; very complex topography (465–3800 m a.m.s.l.).

Observations: Daily precipitation sums of 117 stations, 1971–2012.

Ensemble Model Data: ECMWF reforecasts (training), ECMWF ensemble (prediction). Training data set length for spatial SAMOS up to 9360 using most recent four reforecast runs. Horizontal resolution $\sim$ 40 km.

Climatologies: Observed climatology based on Stauffer et al. (2016), ensemble model climatology for the ECMWF reforecasts (empirical mean/stdv).

Verification: 10-fold cross-validated spatial SAMOS results (out of sample) against two baseline methods: station-wise SAMOS (STN), and raw ensemble (ENS); verification includes the years 2010–2012.

Model Results

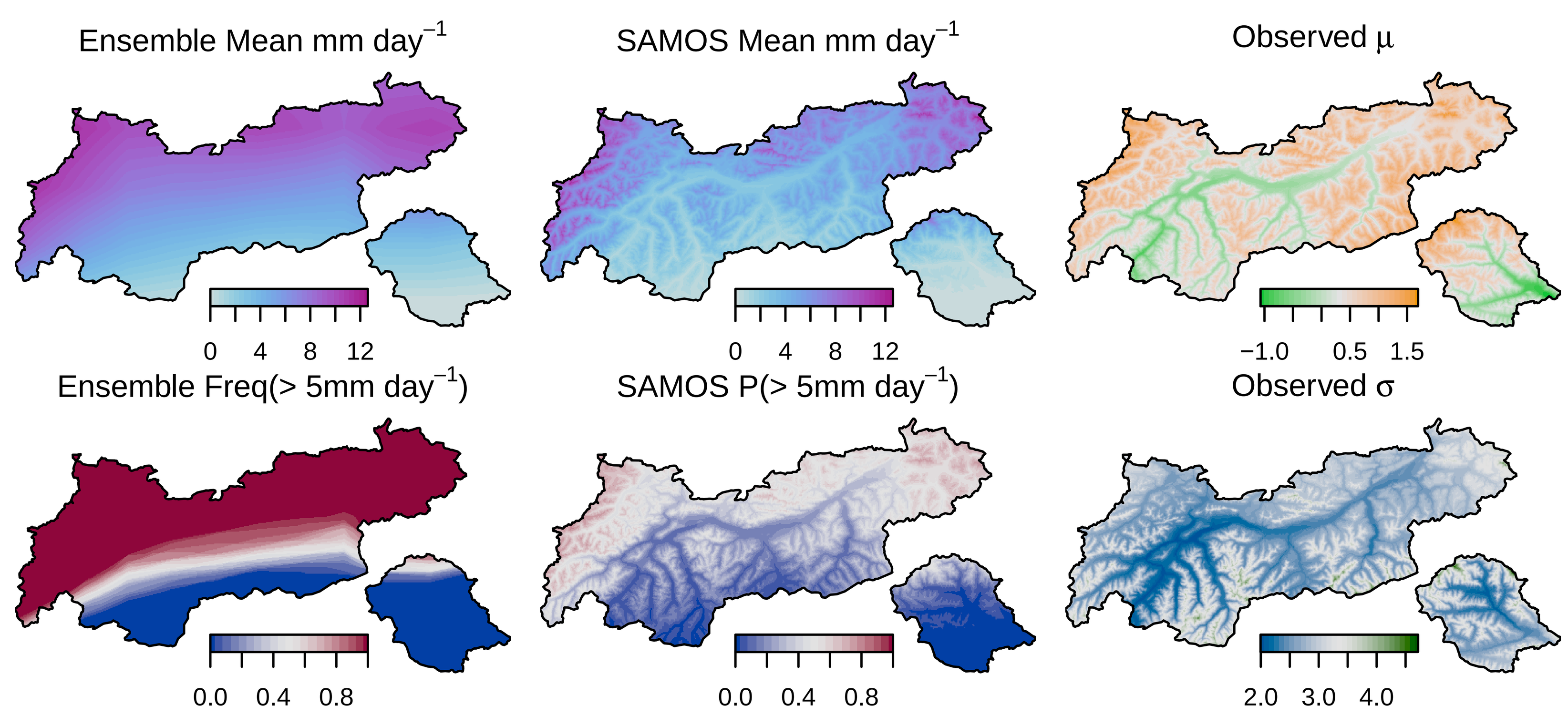

Figure 3: Raw ensemble forecast (left), post-processed SAMOS forecast (center), and observed climatology (right) for May 18, 2010. Forecasts: mean precipitation amount in mm day^{-1} (top), frequency/probability receiving more than 5 mm day^{-1} (bottom). Climatology: latent location $\mu_{obs,clim}$ (top), and latent scale $\sigma_{obs,clim}$ (bottom).

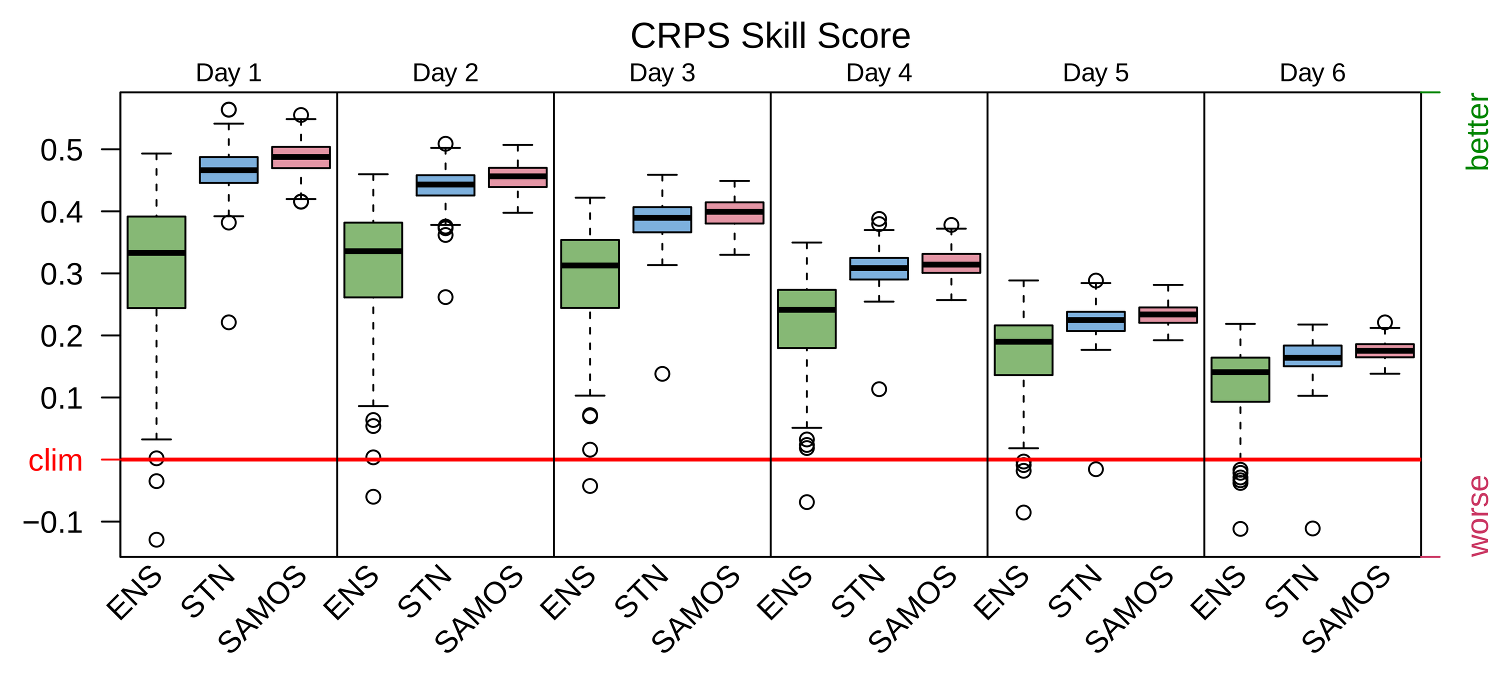

Figure 4: Verification scores for the raw ensemble (ENS), station-wise SAMOS (STN), and spatial SAMOS (SAMOS). Continuous Rank Probability Skill Score using the climatology as reference. Scores for daily precipitation sums, one-day-ahead to six-day-ahead forecasts.

Summary

- Allows for **fully probabilistic spatial** ensemble **correction**
- **Improved probabilistic forecast**
- Anomalies **preserve small-scale** features
- **Simple** and computationally **inexpensive** model

Download

Acknowledgments:
Ongoing project funded by the **Austrian Science Fund (FWF)**: TRP 290-N26.

References:
Dabernig, M., G. J. Mayr, J. W. Messner, and A. Zeileis, 2016: Spatial ensemble post-processing with standardized anomalies. Econ working papers, Atmospheric and Cryospheric Institute, University of Innsbruck.
Stauffer, R., J. W. Messner, G. J. Mayr, N. Umlauf, and A. Zeileis, 2016: Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. Econ working papers, Faculty of Economics and Statistics, University of Innsbruck.