retostauffer

# Spatio-Temporal Ensemble Postprocessing over Complex Terrain

## High-Resolution Precipitation Forecasts for Tyrol

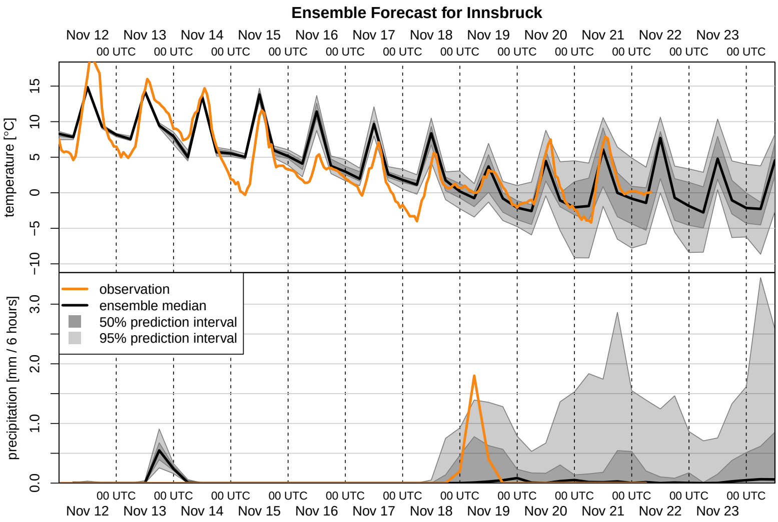
https://retostauffer.org

# Weather Forecasting

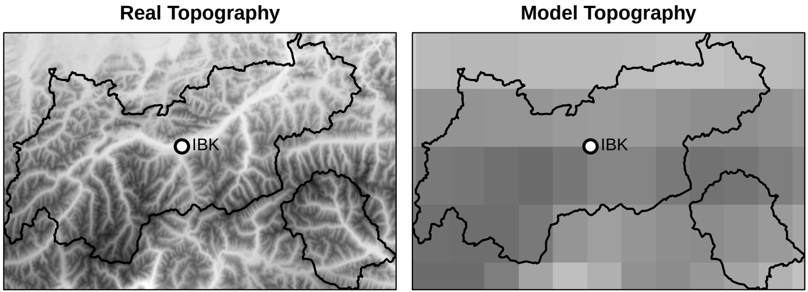
**Numerical Weather Forecast Models**

- estimate current state (analysis)
- predict future state (prognosis)
- ensemble systems: quantify uncertainty

# Weather Forecasting



# Weather Forecasting



# Ensemble Postprocessing

## Concept

- based on historical forecasts and observations
- identify systematic errors in both, mean and variance
- apply correction to **new** forecasts

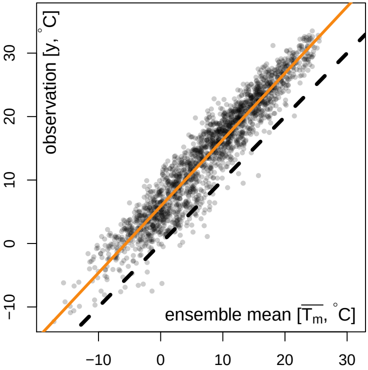
## Statistical Models

- distributional regression models
- distributional forests
- neuronal networks

# Ensemble Postprocessing

**Ensemble Model Output Statistics** (EMOS; Gneiting 2005)

$$
\begin{aligned}
y &\sim \mathcal{N}(\mu, \sigma) \\
\mu &= \beta_0 + \beta_1 \cdot \overline{T_m} \\
\log(\sigma) &= \gamma_0 + \gamma_1 \cdot \log\left(\mathrm{sd}(T_m)\right)
\end{aligned}
$$

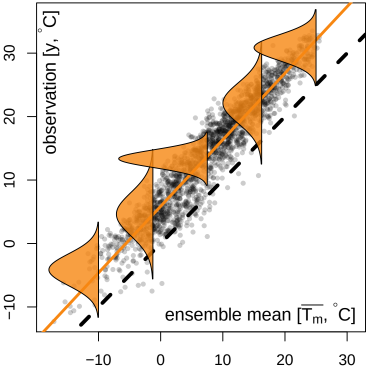



$y$**:** temperature observation
$\mathcal{N}$**:** Gaussian distribution
$T_m$**:** temperature forecast
$m$**:** ensemble member $m \in \{1 \ldots M\}$

# Ensemble Postprocessing

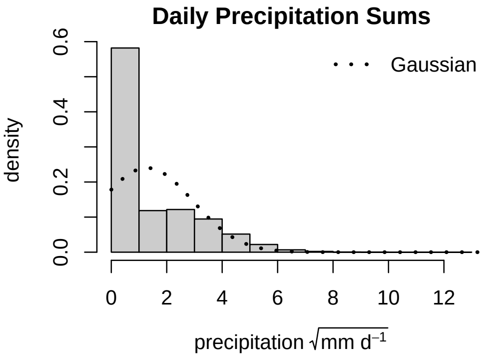
**Ensemble Model Output Statistics** (EMOS; Gneiting 2005)

$$
\begin{aligned}
y &\sim \mathcal{N}(\mu, \sigma) \\
\mu &= \beta_0 + \beta_1 \cdot \overline{T_m} \\
\log(\sigma) &= \gamma_0 + \gamma_1 \cdot \log\left(\mathrm{sd}(T_m)\right)
\end{aligned}
$$




$y$**:** temperature observation
$\mathcal{N}$**:** Gaussian distribution
$T_m$**:** temperature forecast
$m$**:** ensemble member $m \in \{1 \ldots M\}$

# Ensemble Postprocessing



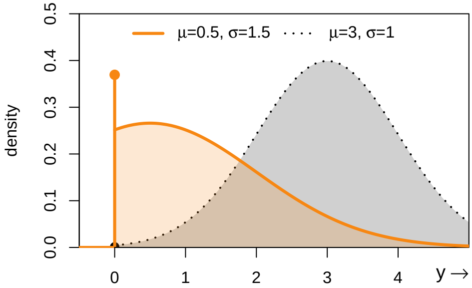

**Precipitation**

- limited to positive values
- large fraction of days without rain
- Gaussian assumption inappropriate

# Censoring

## Left-Censored Gaussian Distribution ($\mathcal{N}_0$)

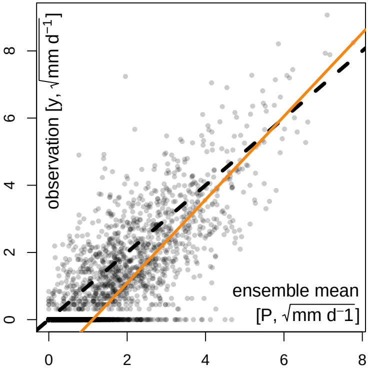
$$\text{PDF: } \phi_0(y_i|\mu_i, \sigma_i) = \begin{cases} \Phi(0|\mu_i, \sigma_i) & \text{if } y_i = 0 \\ \phi(y_i|\mu_i, \sigma_i) & \text{else} \end{cases}$$



# Censoring

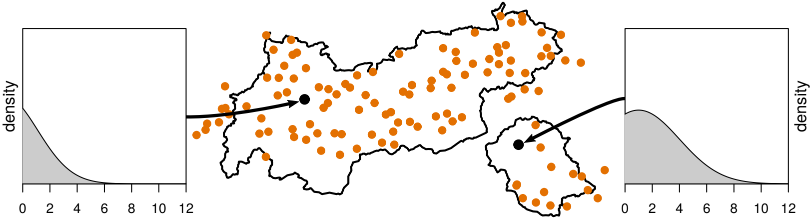
## Censored EMOS for Precipitation

$$
\begin{aligned}
y^{\frac{1}{2}} &\sim \mathcal{N}_0(\mu, \sigma) \\
\mu &= \beta_0 + \beta_1 \cdot \overline{P_m^{1/2}} \\
\log(\sigma) &= \gamma_0 + \gamma_1 \cdot \log\left(\mathrm{sd}(P_m^{1/2})\right)
\end{aligned}
$$




**$y$:** observed precipitation sum
**$\mathcal{N}_0$:** censored Gaussian distribution
**$P_m$:** precipitation forecast
**$m$:** ensemble member $m \in \{1 \ldots M\}$

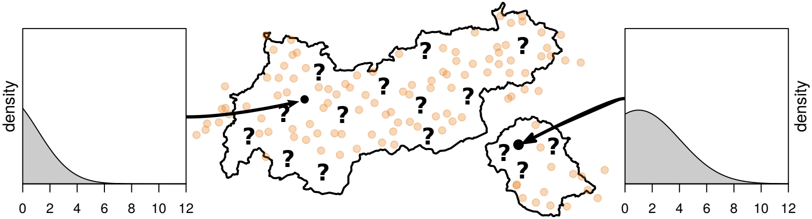
# Ensemble Postprocessing




**Station-Wise Postprocessing**

- relatively simple
- interpolate ensemble forecasts
- apply non-homogeneous censored EMOS
- **one** model for **each** station

# Ensemble Postprocessing




**Station-Wise Postprocessing**

- relatively simple
- interpolate ensemble forecasts
- apply non-homogeneous censored EMOS
- **one** model for **each** station

# Spatial Postprocessing

**Standardized Anomaly Model Output Statistics (SAMOS)**

- remove location and season dependent characteristics from the data
- pool all stations
- estimate **one global** regression **model**

**Standardized Anomalies**

$$y^* = \frac{y^{\frac{1}{2}} - \tilde{\mu}_{obs}}{\tilde{\sigma}_{obs}}; \quad P_m^* = \frac{P_m^{\frac{1}{2}} - \tilde{\mu}_P}{\tilde{\sigma}_P}$$

# Spatial Postprocessing

## SAMOS Model Specification

$$\begin{aligned} y^* &\sim \mathcal{N}_{\text{var}}\left(\mu^*, \sigma^*\right) \\ \mu^* &= \beta_0^* + \beta_1^* \cdot \overline{P_m^*} \\ \log(\sigma^*) &= \gamma_0^* + \gamma_1^* \cdot \log\left(\text{sd}(P_m^*)\right) \end{aligned}$$

## De-standardization

$$y^{\frac{1}{2}} \sim \mathcal{N}_o\left(\mu^* \cdot \tilde{\sigma}_y + \tilde{\mu}_y, \;\; \sigma^* \cdot \tilde{\sigma}_y\right)$$

$y^*$/$P^*$**:** observed/forecasted standardized anomalies
$\mathcal{N}_{var}$**:** censored Gaussian distribution; varying censoring point
$\mu^*$/$\sigma^*$ distributional parameters, anomaly scale
$m$**:** ensemble member $m \in \{1 \dots M\}$

# Spatial Postprocessing

**Background Climatology**

$$
\begin{aligned}
\mathrm{y}^{\frac{1}{2}} &\sim \mathcal{N}_0\left(\tilde{\mu}_y, \tilde{\sigma}_y\right) \\
\tilde{\mu}_y &= f_1(\mathrm{alt}) + f_2(\mathrm{doy}) + f_3(\mathrm{lon}, \mathrm{lat}) + f_4(\mathrm{doy}, \mathrm{lon}, \mathrm{lat}) \\
\log(\tilde{\sigma}_y) &= g_1(\mathrm{alt}) + g_2(\mathrm{doy}) + g_3(\mathrm{lon}, \mathrm{lat}) + g_4(\mathrm{doy}, \mathrm{lon}, \mathrm{lat})
\end{aligned}
$$

# Ensemble Postprocessing

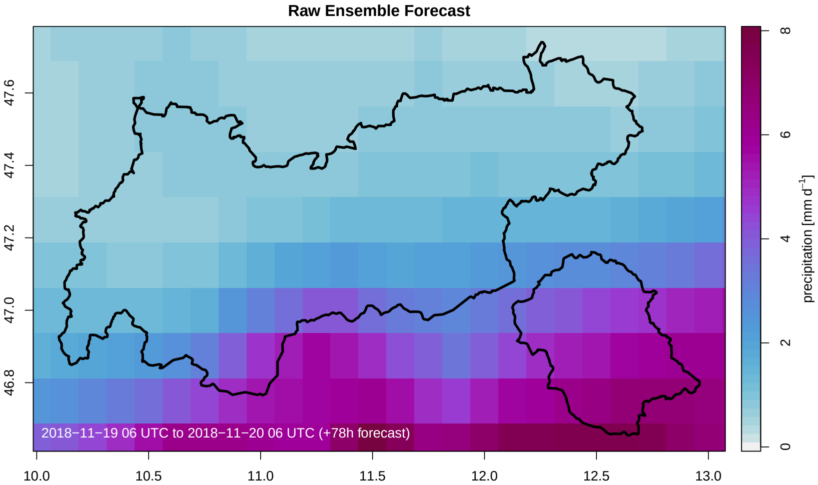
## Parameter Estimation

$$\ell(\theta \mid y) = \sum_{i=1}^{N} \log \phi_0(y_i, \theta); \quad \theta = (\beta, \gamma)$$

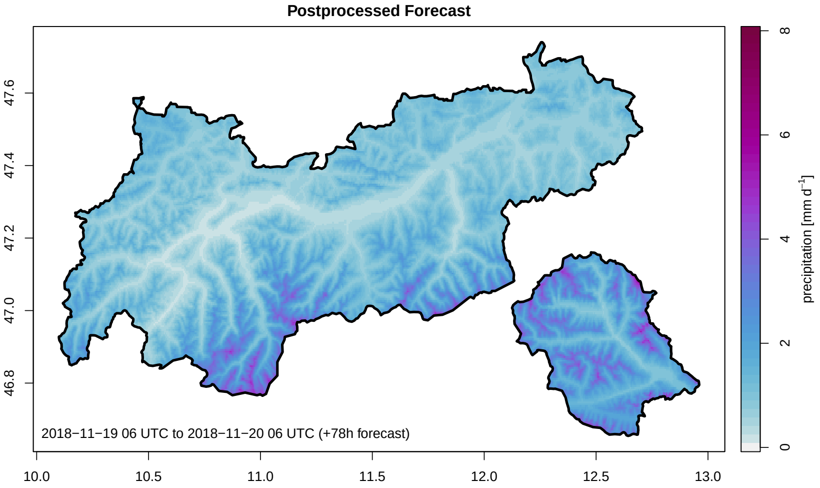
$$\hat{\theta} = \underset{\theta \in \mathbb{R}}{\text{argmax}} \Big( \ell(\theta \mid y) \Big)$$

- iterative weighted least squares
- Markov chain Monte Carlo (MCMC)
- gradient boosting
- distributional forests

# Ensemble Postprocessing



# Ensemble Postprocessing



# Forecast Verification

## Valuable Probabilistic Forecasts

- Are unbiased,
- as sharp as possible,
- but as wide as necessary.

## Typical Scores

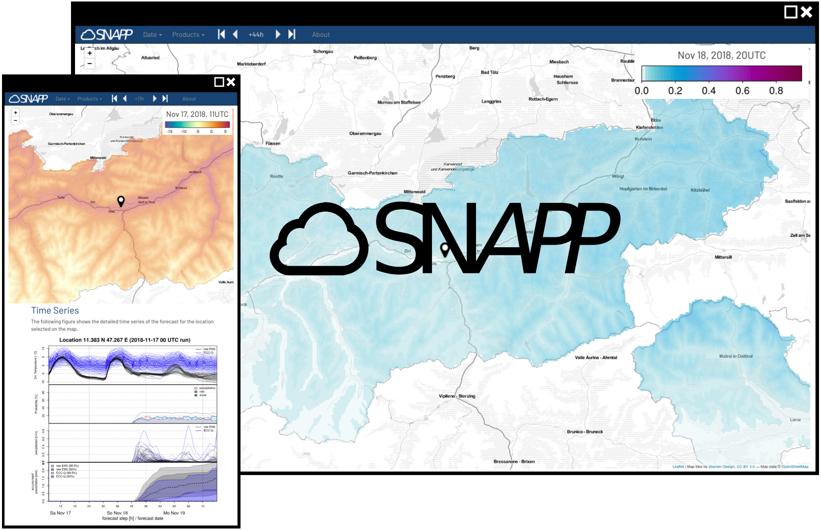
- univariate forecasts
  - log-score, ignorance
  - continuous ranked probability score (CRPS)
  - probability integral transform histograms (PIT)
- multivariate forecasts
  - energy score (ES)
  - variogram score (VS)
- economic value score

# Snow and Snowfall Forecasts

**Methodology**

- standardized anomaly model output statistics (SAMOS)
  - daily precipitation forecasts
  - hourly temperature forecasts
- novel re-weighting scheme for temporal downscaling
- ensemble copula coupling

# Snow and Snowfall Forecasts



# Conclusion

| | |
|---|---|
| **Input** | Observations and global forecasts:<br>`MySQL`, `SQLite`, `GRIB`, `NetCDF`. |
| **Data wrangling** | Spatio-temporal data:<br>`ncdf4`, `raster`, `sp`, `zoo`, `ecCodes`. |
| **Statistical postprocessing** | Probabilistic regression models:<br>`mgcv`, `crch`, `bamlss`. |
| **Visualization** | Forecast maps:<br>`raster`, `PROJ.4`, `colorspace`. |
| **Deployment** | Web-app with *R* interface:<br>`MySQL`, `jQuery`, `bootstrap`, `leaflet`. |

# References

**Stauffer, R.**, Mayr, G. J., Messner, J. W. and Zeileis, A. (2018). Hourly Probabilistic Snow Forecasts over Complex Terrain: A Hybrid Ensemble Postprocessing Approach. *Advances in Statistical Climatology, Meteorology and Oceanography*, accepted for publication.

**Stauffer, R.**, Umlauf N., Messner, J. W., Georg, G. J., Zeileis A. (2017). Ensemble Postprocessing of Daily Precipitation Sums over Complex Terrain Using Censored High-Resolution Standardized Anomalies. *Monthly Weather Review* **145(3)**, 955–969. doi:10.1175/MWR-D-16-0260.1

**Stauffer, R.**, Mayr, G. J., Messner, J. W., Umlauf N., and Zeileis, A. (2017). Spatio-Temporal Precipitation Climatology Over Complex Terrain Using a Censored Additive Regression Model. *International Journal of Climatology* **37(7)**, 3264–3275. doi:10.1002/joc.4913

Schlosser, L., Hothorn, T., **Stauffer, R.**, and Zeileis A. (2018). Distributional Regression Forests for Probabilistic Precipitation Forecasting in Complex Terrain. *Annals of Applied Statistics*, under review.

Gebetsberger, M., **Stauffer, R.**, Mayr, G. J., and Zeileis, A. (2018). Skewed Logistic Distribution for Statistical Temperature Post-processing in Mountainous Areas. *Advances in Statistical Climatology, Meteorology and Oceanography*, revise and resubmit.

**Stauffer, R.**, Mayr, G. J., Dabernig M., Zeileis, A. (2015). Somewhere Over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations. *Bulletin of the American Meteorological Society* **96(2)**, 203–216. doi:10.1175/BAMS-D-13-00155.1

Gneiting, T., Raftery, A. E., Westveld, A. H. and 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.

Thank you for your attention!

https://retostauffer.org