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

Real Topography

Model Topography

IBK

IBK
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)*

\[
y \sim \mathcal{N}(\mu, \sigma)
\]

\[
\mu = \beta_0 + \beta_1 \cdot \overline{T}_m
\]

\[
\log(\sigma) = \gamma_0 + \gamma_1 \cdot \log(\text{sd}(T_m))
\]

\(y\): 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)

\[
y \sim \mathcal{N}(\mu, \sigma)
\]

\[
\mu = \beta_0 + \beta_1 \cdot T_m
\]

\[
\log(\sigma) = \gamma_0 + \gamma_1 \cdot \log(\text{sd}(T_m))
\]

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

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

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

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

\[ y^{1/2} \sim \mathcal{N}_0(\mu, \sigma) \]
\[ \mu = \beta_0 + \beta_1 \cdot \frac{P_m^{1/2}}{m} \]
\[ \log(\sigma) = \gamma_0 + \gamma_1 \cdot \log\left(\text{sd}\left(\frac{P_m^{1/2}}{m}\right)\right) \]

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

\[ y^* \sim \mathcal{N}_{\text{var}}(\mu^*, \sigma^*) \]
\[ \mu^* = \beta_0^* + \beta_1^* \cdot P_m^* \]
\[ \log(\sigma^*) = \gamma_0^* + \gamma_1^* \cdot \log(\text{sd}(P_m^*)) \]

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
\[ \mathcal{N}_{\text{var}}: \] censored Gaussian distribution; varying censoring point
\[ \mu^*/\sigma^*: \] distributional parameters, anomaly scale
\[ m: \] ensemble member \( m \in \{1 \ldots M\} \)
Spatial Postprocessing

**Background Climatology**

\[ y^{\frac{1}{2}} \sim \mathcal{N}_0(\tilde{\mu}_y, \tilde{\sigma}_y) \]

\[ \tilde{\mu}_y = f_1(\text{alt}) + f_2(\text{doy}) + f_3(\text{lon}, \text{lat}) + f_4(\text{doy}, \text{lon}, \text{lat}) \]

\[ \log(\tilde{\sigma}_y) = g_1(\text{alt}) + g_2(\text{doy}) + g_3(\text{lon}, \text{lat}) + g_4(\text{doy}, \text{lon}, \text{lat}) \]
Ensemble Postprocessing

Parameter Estimation

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

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

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

Raw Ensemble Forecast

2018-11-19 06 UTC to 2018-11-20 06 UTC (+78h forecast)

precipitation [mm d⁻¹]
Ensemble Postprocessing

Postprocessed Forecast

2018−11−19 06 UTC to 2018−11−20 06 UTC (+78h forecast)

precipitation [mm d⁻¹]
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: MySQL, SQLite, GRIB, NetCDF.

## Data wrangling
- Spatio-temporal data: ncdf4, raster, sp, zoo, ecCodes.

## Statistical postprocessing
- Probabilistic regression models: mgcv, crch, bamlss.

## Visualization
- Forecast maps: raster, PROJ.4, colorspace.

## Deployment
- Web-app with R interface: MySQL, jQuery, bootstrap, leaflet.


Thank you for your attention!

https://retostauffer.org