Spatio-Temporal Censored Model of Precipitation Climatology

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

reto.stauffer@uibk.ac.at
Estimated expectation of precipitation, 8th of July in [mm/day]
Estimated probability of precipitation, 8th of July

Kufstein
●
Innsbruck
●
St.Anton a.A.
●
Galtuer
●
Lienz
●
Soelden
●
Mayrhofen
●
Introduction
Introduction

![Map of Jungholz region]

**Precipitation Data**

- **Observed**
  - Median
  - $q=0.90$
  - $q=0.75$

- **Model**
  - Median
  - $q=0.90$
  - $q=0.75$

**Day of the Year vs. Precipitation [mm/day]**

- 0 50 100 150 200 250 300 350
- 0 10 20 30 40

---

*Graph showing observed and model data for precipitation at Jungholz during Summer.*
Introduction
Introduction

Nikolsdorf

Summer

Observed
q=0.90
q=0.75
median

Model
q=0.90
q=0.75
median

precipitation [mm/day]

day of the year

0 50 100 150 200 250 300 350
0 10 20 30 40
Introduction

Our goal: precipitation climatology with . . .
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station ⇔ grid)
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station ⇔ grid)
- **daily** resolution
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station $\Leftrightarrow$ grid)
- **daily** resolution
- **full** climatological **distribution**
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station ⇔ grid)
- **daily** resolution
- **full** climatological **distribution**
- **no extensive** or **manual tuning** for new areas
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station ⇔ grid)
- **daily** resolution
- **full** climatological **distribution**
- no extensive or manual tuning for new areas

Applications

- alpine risk assessment
- tourism
Introduction

Our goal: precipitation climatology with . . .

- scalable spatial resolution (station ⇔ grid)
- daily resolution
- full climatological distribution
- no extensive or manual tuning for new areas

Applications

- alpine risk assessment
- tourism
- improving snow forecasts
Strategy
Strategy

Model setup

- generalized additive model for $\mu$ and $\sigma^2$ (GAMLSS)
Strategy

Model setup

- **generalized additive model** for \( \mu \) and \( \sigma^2 \) (GAMLSS)
- **altitude**: linear effect

Open question which distribution?

\[ y \sim N(\mu, \sigma^2) \]

not suitable for all aggregation levels
Strategy

Model setup

- generalized additive model for $\mu$ and $\sigma^2$ (GAMLSS)
- altitude: linear effect
- season: cyclic cubic spline
Strategy

Model setup

- **generalized additive model** for $\mu$ and $\sigma^2$ (GAMLSS)
- **altitude**: linear effect
- **season**: cyclic cubic spline
- **spatial**: two dimensional thin-plate splines

Open question: which distribution?

$y \sim N(\mu, \sigma^2)$

not suitable for all aggregation levels
Strategy

Model setup

- **generalized additive model** for $\mu$ and $\sigma^2$ (GAMLSS)
- **altitude**: linear effect
- **season**: cyclic cubic spline
- **spatial**: two dimensional thin-plate splines
- **spatially variable season**: spatial variability on seasonal pattern
Strategy

Model setup

- **generalized additive model** for $\mu$ and $\sigma^2$ (GAMLSS)
- **altitude**: linear effect
- **season**: cyclic cubic spline
- **spatial**: two dimensional thin-plate splines
- **spatially variable season**: spatial variability on seasonal pattern

Open question

- which distribution?
  
  \[ y \sim \mathcal{N}(\mu, \sigma^2) \] not suitable for all aggregation levels
The data: monthly sums

Monthly sums

precipitation \([\sqrt{\text{mm/month}}]\)

<table>
<thead>
<tr>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
</tr>
<tr>
<td>0.04</td>
</tr>
<tr>
<td>0.08</td>
</tr>
<tr>
<td>0.12</td>
</tr>
</tbody>
</table>

precipitation \([\sqrt{\text{mm/month}}]\)
The data: monthly sums

Monthly sums

<table>
<thead>
<tr>
<th>Density</th>
<th>0.00</th>
<th>0.04</th>
<th>0.08</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation [(\sqrt{\text{mm}/\text{month}})]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[\text{mean} = 9.36\]
\[\text{sdev} = 3.33\]
The data: daily amounts

Daily amounts

-5 0 5
0.00 0.05 0.10 0.15 0.20 0.25 cut @ 0.25

Observed

Precipitation $[\sqrt{\text{mm/day}}]$
The data: daily amounts

Daily amounts

-5 0 5

0.00 0.05 0.10 0.15 0.20 0.25

Observed
First guess

precipitation [\sqrt{\text{mm}/day}]
Censoring

Can be seen as censored if . . .

- **limited** to a threshold
- values exceeding threshold **cannot occur** or **not observable**
Censoring

Can be seen as censored if . . .

- **limited** to a threshold
- values exceeding threshold cannot occur or not observable

Examples

- survival in 5 yr clinical study: **right** $y_i \leq 5$
- hours worked this week: **two sided** $0 \leq y_i \leq 168$

Precipitation:

$0 \leq y_i \leq 100$

Left censored Gaussian distribution

$y^* \sim N(\mu, \sigma^2), y = \max(\tau = 0, y^*)$
Censoring

Can be seen as censored if . . .

- limited to a threshold
- values exceeding threshold cannot occur or not observable

Examples

- survival in 5 yr clinical study: right \( y_i \leq 5 \)
- hours worked this week: two sided \( 0 \leq y_i \leq 168 \)
- precipitation: left \( 0 \leq y_i \)
Censoring

Can be seen as censored if . . .

- limited to a threshold
- values exceeding threshold cannot occur or not observable

Examples

- survival in 5 yr clinical study: right \( y_i \leq 5 \)
- hours worked this week: two sided \( 0 \leq y_i \leq 168 \)
- precipitation: left \( 0 \leq y_i \)

Left censored Gaussian distribution

\[ \begin{align*}
    y^* &\sim \mathcal{N}(\mu, \sigma^2), \\
    y &= \max(\tau = 0, y^*)
\end{align*} \]
The data: daily amounts

Daily amounts

Cut @ 0.25

Mean = −0.37
Sdev = 2.43

Precipitation [√mm/day]
The data: daily amounts

- Daily amounts of precipitation (mm/day)
- Density distribution
- Observed data
- Censored distribution
- Fraction of zeros: 0.57
- Parameter $\mu = 1.04$
Model setup

Assumptions

\[ y \sim D(\mu, \sigma^2), \quad \mu = \eta_\mu, \quad \log(\sigma) = \eta_\sigma \]

- \( y \): response; daily precipitation (transformed)
- \( D(\ldots) \): left censored Gaussian distribution on 0
- \( \mu/\sigma \): location/scale
Model setup

Assumptions

\[ y \sim \mathcal{D}(\mu, \sigma^2), \quad \mu = \eta \mu, \quad \log(\sigma) = \eta \sigma \]

Linear predictor

\[ \eta = \beta_0 + \beta_1 \text{alt} + f_1(\text{yday}) + f_2(\text{long, lat}) + f_3(\text{yday, long, lat}) \]

- y: response; daily precipitation (transformed)
- \( \mathcal{D}(\ldots) \): left censored Gaussian distribution on 0
- \( \mu/\sigma \): location/scale
- \( \eta \): linear predictor; identical for \( \mu \) and \( \log(\sigma) \)
Model and Results
Model and Results

Model estimation

- R package **bamlss** (Umlauf et al. (2015); Poster 72)
- censored Gaussian distribution
Model and Results

Model estimation

- **R package `bamlss`** (Umlauf et al. (2015); Poster 72)
- censored Gaussian distribution

Data set

- 110 stations
- 41 years
- daily rain/snow amounts
- $\sim 1.5$ million observations

Stations used
Model and Results

\[ \eta = \beta_0 + \beta_1 \text{alt} + f_1(\text{yday}) + f_2(\text{long, lat}) + f_3(\text{yday, long, lat}) \]
Model and Results

\[ \eta = \beta_0 + \beta_1 \text{alt} + f_1(\text{yday}) + f_2(\text{long, lat}) + f_3(\text{yday, long, lat}) \]

\[ \text{season} \]
Model and Results

\[ \eta = \beta_0 + \beta_1 \text{alt} + f_1(y\text{day}) + f_2(\text{long, lat}) + f_3(y\text{day, long, lat}) \]

season
\[ \eta = \beta_0 + \beta_1 \text{alt} + f_1(\text{yday}) + f_2(\text{long}, \text{lat}) + f_3(\text{yday}, \text{long}, \text{lat}) \]
Model and Results

Expectation = \( E( y_{ij} \mid \mu_{ij}, \sigma_{ij} ) \)
Model and Results

$$\text{Expectation} = E(y_{ij} | \mu_{ij}, \sigma_{ij})$$
**Model and Results**

Expectation = \( E(y_{ij} | \mu_{ij}, \sigma_{ij}) \)

![Graph showing the expectation of rainfall variation over months.](image)
Model and Results

Expectation = \( E(y_{ij} \mid \mu_{ij}, \sigma_{ij}) \)
Model and Results

Expectation = \( E(\gamma_{ij} \mid \mu_{ij}, \sigma_{ij}) \)

full spatial model
Model and Results

\[ \text{Expectation} = E(y_{ij} \mid \mu_{ij}, \sigma_{ij}) \]

- **full spatial model**
- **single station**

![Graph showing the expectation of precipitation over time with different lines for full spatial model and single station.](image-url)
Summary

- “simple”, generalized setup
- full spatio-temporal model with censored response
- spatial/temporal resolution arbitrary scalable
- accurate estimate at station level
Summary

- “simple”, generalized setup
- full spatio-temporal model with censored response
- spatial/temporal resolution arbitrary scalable
- accurate estimate at station level

- handling zero-observations
- full climatological distribution
- quantiles
- probability of precipitation
Thank you for your attention!

Further details:
Scientific article in progress.

Contact:
Reto Stauffer, Department of Statistics, University of Innsbruck.
reto.stauffer@uibk.ac.at.

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).
Data set provided by the “Ministerium für ein lebenswertes Österreich”, hydrographical service Tyrol (ehyd.gv.at).
References


Umlauf, N., N. Klein, A. Zeileis, and S. Lang, 2015: *bamlss: Bayesian Additive Models for Location Scale and Shape (and Beyond)*. URL https://r-forge.r-project.org/scm/?group_id=865, r package version 0.1-1/r1358.
Model and Results

Model on monthly sums

- R package mgcv
- \( y = (\text{monthly sums})^{\frac{1}{2}} \)
- \( y \sim \mathcal{N}(\mu, \sigma^2) \)
- \( \mu = \eta_\mu, \quad \log(\sigma) = \eta_\sigma \)
Model and Results

Model on monthly sums
- \( \text{R package } \text{mgcv} \)
- \( y = (\text{monthly sums})^{\frac{1}{2}} \)
- \( y \sim \mathcal{N}(\mu, \sigma^2) \)
- \( \mu = \eta \mu, \log(\sigma) = \eta \sigma \)

Model on daily amounts
- \( \text{R package } \text{bamlss} \)
- \( y = (\text{daily amounts})^{\frac{1}{1.5}} \)
- \( y^* \sim \mathcal{N}(\mu, \sigma^2) \)
- \( \mu = \eta \mu, \log(\sigma) = \eta \sigma \)

Error statistics ([mm/month])

<table>
<thead>
<tr>
<th></th>
<th>in-sample</th>
<th></th>
<th>out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mgcv</td>
<td>bamlss</td>
<td>mgcv</td>
</tr>
<tr>
<td>BIAS</td>
<td>-0.13</td>
<td>0.58</td>
<td>-0.21</td>
</tr>
<tr>
<td>MAE</td>
<td>37.62</td>
<td>37.37</td>
<td>38.41</td>
</tr>
<tr>
<td>RMSE</td>
<td>50.44</td>
<td>50.17</td>
<td>51.42</td>
</tr>
</tbody>
</table>