ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gebetsberger, Manuel; Stauffer, Reto; Mayr, Georg J.; Zeileis, Achim

**Working Paper**

# Skewed logistic distribution for statistical temperature post-processing in mountainous areas

Working Papers in Economics and Statistics, No. 2018-06

**Provided in Cooperation with:**
Institute of Public Finance, University of Innsbruck

*Suggested Citation:* Gebetsberger, Manuel; Stauffer, Reto; Mayr, Georg J.; Zeileis, Achim (2018) : Skewed logistic distribution for statistical temperature post-processing in mountainous areas, Working Papers in Economics and Statistics, No. 2018-06, University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon), Innsbruck

This Version is available at:
https://hdl.handle.net/10419/184984

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

universität innsbruck

Research Platform
Empirical and Experimental Economics

# Skewed logistic distribution for statistical temperature post-processing in mountainous areas

**Manuel Gebetsberger, Reto Stauffer, Georg J. Mayr, Achim Zeileis**

Working Papers in Economics and Statistics

2018-06

University of Innsbruck
https://www.uibk.ac.at/eeecon/

**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 71022
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

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

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

# Skewed logistic distribution for statistical temperature post-processing in mountainous areas

**Manuel Gebetsberger**
University of Innsbruck

**Reto Stauffer**
University of Innsbruck

**Georg J. Mayr**
University of Innsbruck

**Achim Zeileis**
University of Innsbruck

**Abstract**

Non-homogeneous post-processing is often used to improve the predictive performance of probabilistic ensemble forecasts. A common quantity to develop, test, and demonstrate new methods is the near-surface air temperature frequently assumed to follow a Gaussian response distribution. However, Gaussian regression models with only few covariates are often not able to account for site-specific local features leading to strongly skewed residuals. This residual skewness remains even if many covariates are incorporated. Therefore, a simple refinement of the classical non-homogeneous Gaussian regression model is proposed to overcome this problem by assuming a skewed response distribution to account for possible skewness. This study shows a comprehensive analysis of the performance of non-homogeneous post-processing for 2 m temperature for three different site types comparing Gaussian, logistic, and skewed logistic response distributions. Satisfying results for the skewed logistic distribution are found, especially for sites located in mountainous areas. Moreover, both alternative model assumptions but in particular the skewed response distribution, can improve on the classical Gaussian assumption with respect to overall performance, sharpness, and calibration of the probabilistic predictions.

*Keywords*: Statistical post-processing; Probabilistic temperature forecast; Skewed distribution; Distributional regression.

## 1. Introduction

Probabilistic weather forecasts have become state-of-the-art over the last years (Gneiting and Katzfuss 2014). As such, they are important to address the chaotic nature of the atmosphere, and to express the uncertainty of a specific forecast (Lorenz 1963). The expected uncertainty is typically provided by an ensemble prediction system (EPS, Leith 1974) where multiple forecasts are produced by a numerical weather prediction (NWP) model with slightly perturbed initial conditions, model physics, and parameterizations. However, it was found that these forecasts often show systematic errors for both the expectation and the uncertainty due to required simplified physical equations, insufficient resolution, and unresolved processes (Bauer, Thorpe, and Brunet 2015).

One possibility to correct for these errors are statistical post-processing techniques (Gneiting

and Katzfuss 2014) such as Gaussian ensemble dressing (GED, Roulston and Smith 2003), non-homogeneous Gaussian regression (NGR or EMOS, Gneiting, Raftery, Westveld, and Goldman 2005), Bayesian model averaging (BMA, Raftery, Gneiting, Balabdaoui, and Polakowski 2005), or logistic regression (Wilks 2009; Messner, Mayr, Wilks, and Zeileis 2014). These methods have been tested extensively for air temperature forecast and other quantities where NGR (with various extensions) represents one of the most popular approaches.

The two most important properties of probabilistic forecasts are sharpness and calibration (Gneiting, Balabdaoui, and Raftery 2007). Previous studies show that extensions of the classical NGR (Schefzik, Thorarinsdottir, and Gneiting 2013; Scheuerer and Büermann 2014; Möller and Groß 2016; Dabernig, Mayr, Messner, and Zeileis 2017) and other temperature post-processing methods (Hagedorn, Hamill, and Whitaker 2008; Verkade, Brown, Reggiani, and Weerts 2013; Feldmann, Scheuerer, and Thorarinsdottir 2015; Wilks 2017) are able to improve the predictive performance of the classical NGR with respect to specific predictive performance measures such as sharpness and calibration.

However, in recent publications the presented PIT histograms (Dawid 1984) often do not show the desired perfectly uniform distribution to confirm calibration (cf., Figure 5c,g in Scheuerer and Büermann 2014, Figure 4c in Möller and Groß 2016, or Figure 7 in Messner, Mayr, and Zeileis 2017). More specifically, the histograms indicate skewness in the residual distribution. Since a marginal Gaussian model without covariates can already exhibit skewness for temperature data (Toth and Szentimrey 1990; Warwick and Curran 1993; Harmel, Richardson, Hanson, and Johnson 2002), skewness is supposed to vanish if covariates are incorporated. Nevertheless, the residual distribution is still found skewed even after adjusting by covariates (Scheuerer and Büermann 2014; Möller and Groß 2016; Messner *et al.* 2017). Since covariates are based on output of NWP models, a remaining skewness is likely to originate in small-scale or local atmospheric processes insufficiently or not at all resolved by the NWP models. One example are locations in regions where topography is only coarsely resolved in the model. Then many thermally induced slope and valley wind systems as well as subsidence / lifting zones (Steinacker 1984; Whiteman 1990; Zängl 2004) will be absent and lead to the residual skewness of the post-processed temperature distributions.

So far, most studies assume a Gaussian response distribution for their temperature post-processing methods (Gneiting *et al.* 2005; Hagedorn *et al.* 2008; Verkade *et al.* 2013; Scheuerer and Büermann 2014; Möller and Groß 2016; Gebetsberger, Messner, Mayr, and Zeileis 2017; Dabernig *et al.* 2017). As the Gaussian distribution is symmetric it is not able to account for possible skewness by itself. Hence, this article proposes an extension of the non-homogeneous Gaussian regression framework (Gneiting *et al.* 2005) using a skewed rather than a symmetric response distribution in order to obtain sharp and calibrated probabilistic temperature forecasts. To examine the need of the asymmetry, probabilistic temperature forecasts are presented for a set of stations with different characteristics including sites in the European Alps and in topographically plain areas across Central Europe.

## 2. Methods and data

This section briefly describes the regression framework, the distributions used, data, the statistical models employed, and the verification methodology.

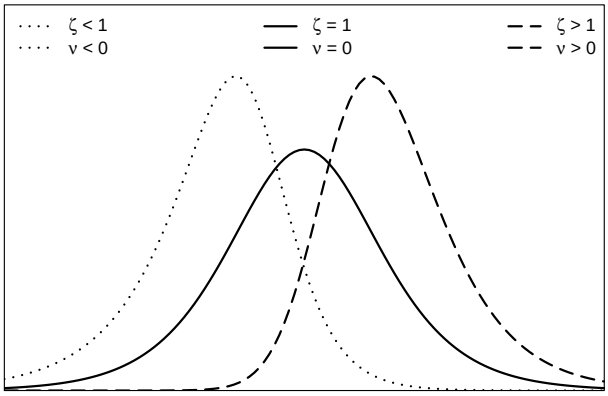


Figure 1: Density function of the skewed logistic distribution, illustrating the $3^{rd}$ moment ($\nu$, skewness) depending on the chosen shape parameter $\zeta$.

## 2.1. Non-homogeneous regression framework

The non-homogeneous Gaussian regression framework as proposed by Gneiting *et al.* (2005) is a special case of a distributional regression model (Klein, Kneib, Lang, and Sohn 2015) and can be expressed in its general form as:

$$y \sim \mathcal{D}\big(h_1(\theta_1) = \eta_1, \ \ldots, \ h_K(\theta_K) = \eta_K\big). \tag{1}$$

A response variable $y$ is assumed to follow some distribution $\mathcal{D}$ with distribution parameters $\theta_k$, $k = 1, \ldots, K$. Each parameter is linked to an additive predictor $\eta_k$ using a monotone link function $h_k$. In this article we use the identity-link $h_k(\eta_k) = \eta_k$ for the location parameters and a log-link for scale and shape parameters to ensure positivity. Each linear predictor can be expressed by a set of additive predictors of the form:

$$\eta_k = \eta_k(\mathbf{x}, \beta_k) = f_{1k}(\mathbf{x}_1, \beta_{1k}) + \cdots + f_{Pk}(\mathbf{x}_P, \beta_{Pk}) \tag{2}$$

including various (possibly non-linear) functions $f_{pk}$, $p = 1, \ldots, P$. Hereby, $\mathbf{x}_p$ defines a matrix of covariates used, and $\beta_{pk}$ is the vector of regression coefficients to be estimated.

## 2.2. Response distributions

This study compares three different distributions for temperature post-processing: (i) the frequently-used Gaussian distribution, (ii) the symmetric logistic distribution, and (iii) the generalized logistic distribution type I. The logistic distribution is used to assess the impact of having slightly heavier tails (Gebetsberger *et al.* 2017). The generalized logistic distribution type I is particularly of interest as it allows to account for possible skewness in the data. For simplicity, it will further be referred as skewed logistic distribution.

The skewed logistic distribution has the cumulative distribution function (CDF):

$$CDF(x) = \frac{1}{(1 + \exp(-\frac{x-\mu}{\sigma}))^{\zeta}} \tag{3}$$

with location parameter $\mu$, scale parameter $\sigma$, and shape parameter $\zeta$. The first derivation of Equation 3 leads to the probability density function (PDF):

$$PDF(x) = \frac{\zeta \cdot \exp(-\frac{x-\mu}{\sigma})}{\sigma \cdot (1 + \exp(-\frac{x-\mu}{\sigma}))^{2\zeta}}. \tag{4}$$

The additional shape parameter $\zeta$ is responsible for the skewness. Figure 1 sketches the PDF for three different shape parameter values $\zeta$ and corresponding skewness $\nu$. $\zeta$ has to be positive definite, where values below 1 create negative skewness (heavier left tail, $\nu < 0$), while values above 1 produce positive skewness (heavier right tail, $\nu > 0$). For $\zeta \equiv 1$ the skewed logistic distribution describes the symmetric logistic distribution.

As an example, values for $\zeta = \{0.50, 1, 3.82\}$ produce a skewness of $\nu = \{-0.85, 0, 0.85\}$ as illustrated in Figure 1. Details regarding skewness calculation can be found in Appendix A.

## 2.3. Data and statistical models

### *Data*

Results are presented for forecasts at 27 different sites in Central Europe (Figure 2) for forecasts +12 h to +96 h at 6-hourly intervals. The sites were selected to investigate the influence of different topographical environments. Therefore, the stations are clustered into three distinct groups representing Alpine sites located in inner-Alpine regions (12), foreland sites in the peripheral area close to the Alps (6), and plain sites in topographically flat areas (9).

Temperature observations are provided by automatic weather stations (10 minute mean values). As input, 2 m temperature forecasts of the $50 + 1$ member EPS of the European Centre for Medium-Range Weather Forecasts (ECMWF) are used. For this study only EPS forecasts initialized at 0000 UTC are considered. The data set covers the time period January 1, 2012 through December 31, 2016 resulting in five years of data which yields a sample size of approximately 1800 for each individual station and forecast lead time.

In this article, detailed case studies will be shown for Innsbruck, Austria (Alpine site), and Hamburg, Germany (plain site; cf., Figure 2), which differ particularly in their topographical environments. While the Alpine site is located in a narrow Alpine valley surrounded by high mountainous exceeding 2500 m of altitude, the plain site is characterized by its vicinity to the sea (100 km) and only few hills with an altitude below 160 m. Due to the necessary simplifications the NWP model topography is missing large parts of the topographical structures, especially for the Alpine site (Figures 1 and 3 in Stauffer, Umlauf, Messner, Mayr, and Zeileis (2017)).

### *Statistical models*

Similar to previous works (cf., Scheuerer and Büermann 2014; Feldmann *et al.* 2015; Möller and Groß 2016; Dabernig *et al.* 2017) we only utilize the ensemble mean ($\overline{ens}$) and ensemble standard deviation ($SD_{ens}$) of the 2 m temperature forecasts from the ECMWF EPS in this study. In the following model specification, the ensemble mean is used for the linear predictor of the location parameter $\mu$ while the ensemble standard deviation is used for the linear predictor of the corresponding scale parameter $\sigma$.

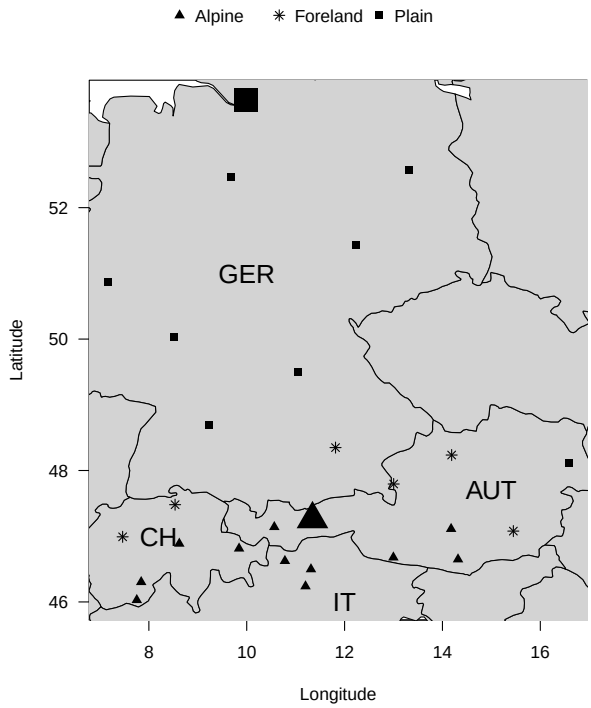


Figure 2: Study area and selected stations in Germany (GER), Switzerland (CH), Italy (IT), and Austria (AUT). The markers indicate stations classified as Alpine (triangle), foreland (star), and plain (square). Large symbols represent stations which are discussed in detail in this article: Innsbruck, Austria (large triangle), and Hamburg, Germany (large square).

While the Gaussian and the logistic distribution have only two parameters ($\mu$ and $\sigma$), the skewed logistic distribution has an additional shape parameter $\zeta$. To be able to capture possible skewness over the season, a smooth cyclic spline $f$ depending on the day of the year (DOY) is used in the linear predictor for the shape parameter. The model specification for the study presented can be summarized as follows:

$$
\begin{aligned}
y &\sim \mathcal{D}(\mu, \sigma, \zeta), && (5)\\
\mu &= \beta_0 + \beta_1 \cdot \overline{ens}, && (6)\\
\log(\sigma) &= \gamma_0 + \gamma_1 \cdot \log(SD_{ens}), && (7)\\
\log(\zeta) &= f(DOY), && (8)
\end{aligned}
$$

| Name / Distribution | $\mu$ | $\log(\sigma)$ | $\log(\zeta)$ |
|---|---|---|---|
| Gaussian | $\overline{ens}$ | $SD_{ens}$ | — |
| Logistic | $\overline{ens}$ | $SD_{ens}$ | — |
| Skewed logistic | $\overline{ens}$ | $SD_{ens}$ | $f$(DOY) |

Table 1: Covariates used in the linear predictors of the distributional parameters $\mu$, $\sigma$, and $\zeta$ for all response distributions. $\overline{ens}$; $SD_{ens}$: ensemble mean and standard deviations of ensemble 2 m temperature, respectively; $f$(DOY): smooth cyclic seasonal effect.

for which the additional parameter $\zeta$ is solely used for models using the skewed logistic response distribution. Table 1 shows a comprehensive overview all models and the covariates used in the corresponding linear predictors.

### 2.4. Verification methodology

Different scores are used to assess the predictive performance of the models tested. The overall performance is evaluated by the logarithmic score (LS; Wilks 2011) and the continuous ranked probability score (CRPS; Hersbach 2000). Of particular interest for this study is the performance of the post-processing models in terms of sharpness and calibration (Gneiting *et al.* 2007).

The sharpness of the probabilistic forecasts is verified using the average prediction interval width (PIW). Results for three different intervals are shown in this article: 50%, 80%, and 95%. As an example: the 80% PIW describes the range between the 10%-percentile and the 90%-percentile of the probabilistic forecast. The smaller the PIW, the sharper the forecast.

Calibration is visually evaluated using PIT histograms. In addition, the reliability index (RI; Delle Monache, Hacker, Zhou, Deng, and Stull 2006) and prediction interval coverage (PIC) are shown. The RI allows to analyze an aggregated measure over a large number of individual PIT histograms. RIs are defined as $\sum_{i=1}^{I} |\kappa_i - \frac{1}{I}|$ where $I$ defines the number of individual bins in a PIT histogram and $\kappa_i$ the observed relative frequency in each bin. In this study we use a binning of 5%. The RI describes the sum of the absolute deviation from each bin in a specific PIT histogram from perfect calibration. Thus, perfectly calibrated forecasts would show an RI of zero. PICs show the calibration for a specific interval. As for PIW, PICs are shown for the 50%, 80%, and 95% interval together with theoretical PICs of 50%, 80% and 95%. The closer the empirical PIC to the theoretical PIC, the better the calibration.

## 3. Results and discussion

This section presents a detailed analysis for a typical Alpine site and compares the performance measures to a topographically plain site followed by a comprehensive analysis of all sites used for this study.

All results presented are out-of-sample results using 5-fold block-wise cross-validation. For each model, station, and lead time, five regression models have been estimated on 4 years of data while one full year (2012, 2013, 2014, 2015, or 2016) is used as test data set.

### 3.1. Alpine case study

To show the performance of the proposed approach, the analysis for one selected site with a distinct Alpine character is shown (large triangle, Figure 2). The left column of Figure 3 presents the verification for this Alpine site. Figure 3 (top down) shows LS, CRPS, 80% PIW, and RI for all forecast lead times. A dominant diurnal cycle for LS, CRPS, and 80% PIW can be seen for all three models, where smallest (best) scores are obtained during nighttime (0000/0600 UTC) and largest during daytime (1200/1800 UTC). Overall, only a small decrease in the forecast performance can be seen with increasing lead time which implies comparable skills between the first and fourth forecast day.

When comparing the logistic model with the benchmark Gaussian model, the logistic model

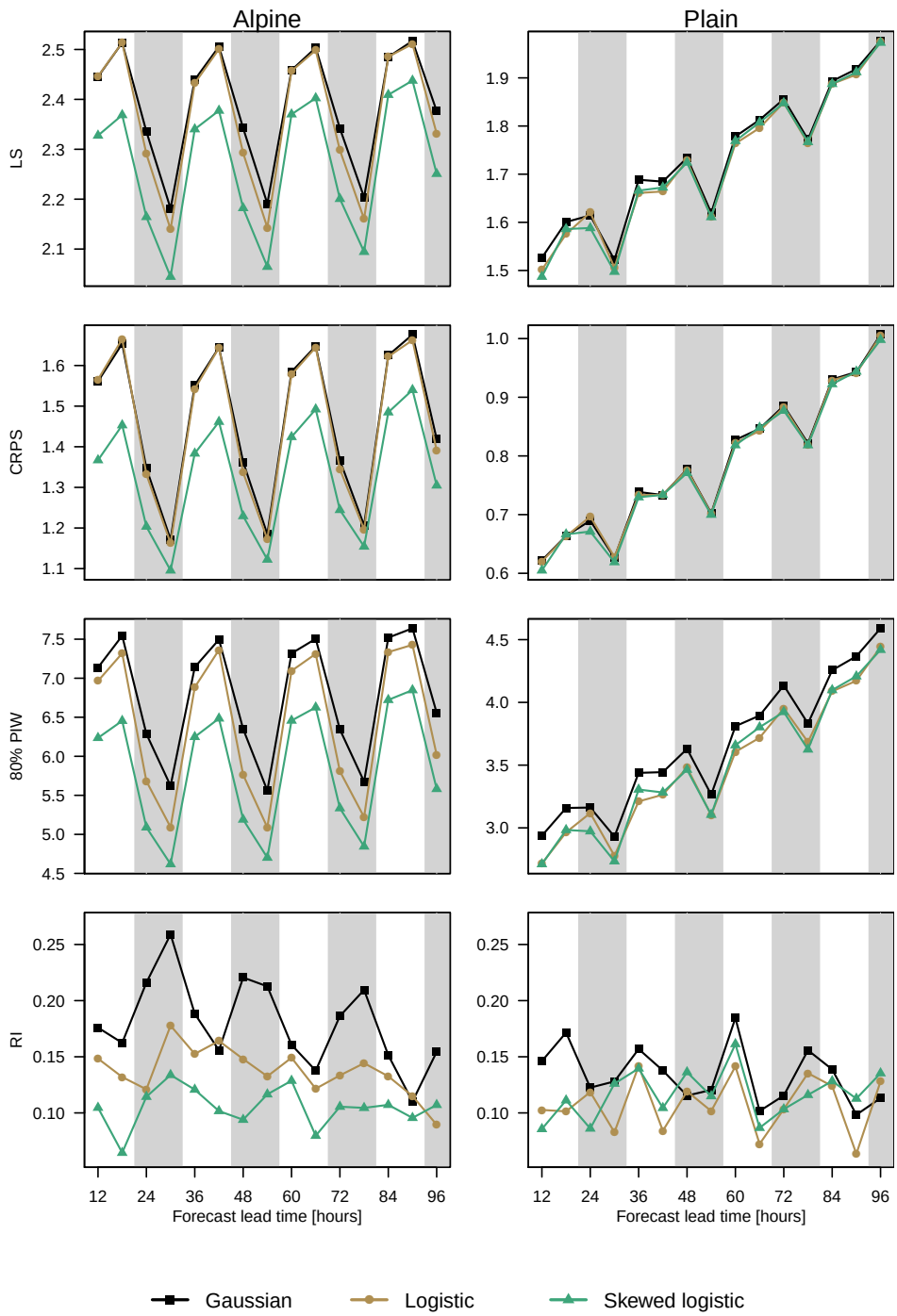


Figure 3: Performance measures at the selected Alpine site (left) and plain site (right) for all three models (Gaussian: squares; logistic: circles; skewed logistic: triangles). Top down LS, CRPS, 80% PIW, and RI are shown, evaluated for all forecasts +12 h up to +96 h ahead. Night time forecasts (0000/0600 UTC) are highlighted by vertical gray bars. Please note that the displayed range on the ordinate differs between the left and right column, except for RI.

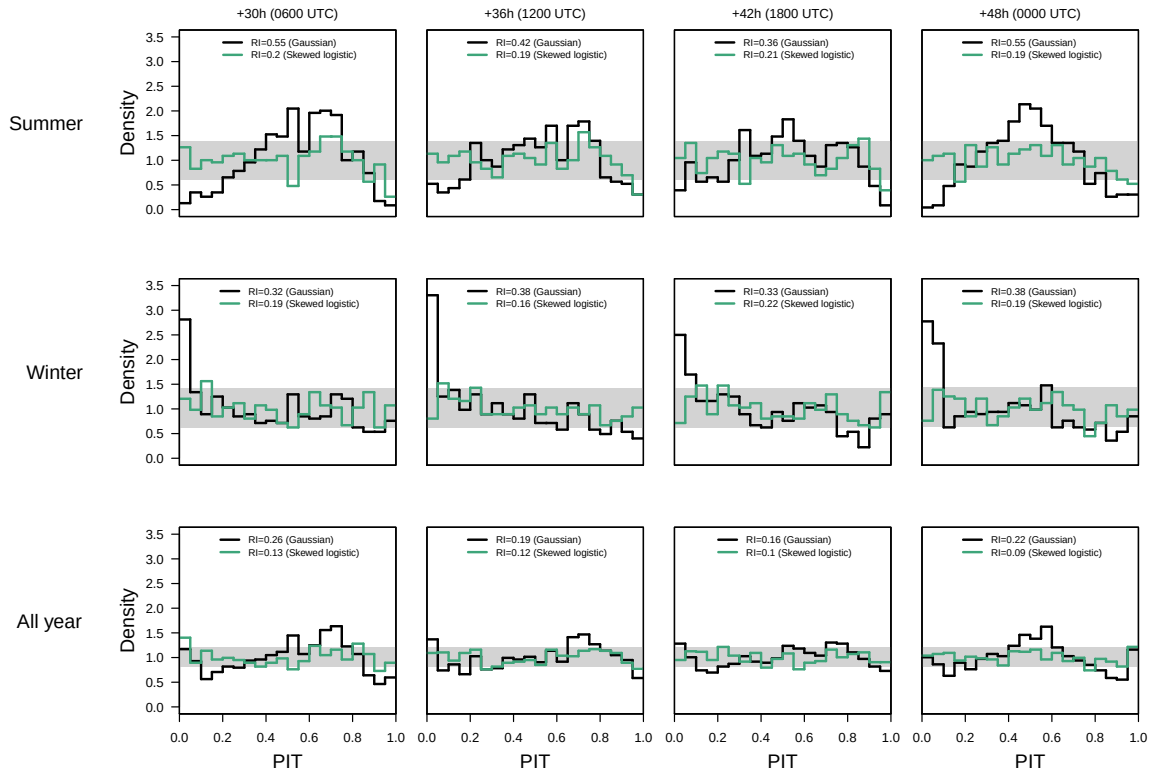


Figure 4: PIT histograms at the Alpine site for the Gaussian (black/dark line) and skewed logistic (green/bright line) models for the two-day ahead forecasts (left to right: 0600/1200/1800/0000 UTC) corresponding to forecasts +30/+36/+42/+48 hours ahead. Top down, the PIT histograms are shown for summer only (Jun/Jul/Aug), winter only (Dec/Jan/Feb), and for the whole year. The gray horizontal bar shows the pointwise 95% confidence interval around 1 for perfect calibration.

shows small improvements in LS and CRPS, especially during nighttime. A similar picture can be seen for the sharpness (80% PIW) where strongest improvements can be achieved during nighttime, but with an overall stronger improvement for all lead times. Furthermore, the logistic model is able to remove large parts of the diurnal pattern in terms of calibration showing a smoother RI over time compared to the Gaussian model.

When comparing the skewed logistic model with the two models using a symmetric response distribution, a considerably large improvement for all lead times can be seen for LS and CRPS, but also in terms of sharpness (80% PIW) and calibration (RI).

Figure 4 shows PIT histograms for the two-day ahead forecasts. To increase readability, only the Gaussian and skewed logistic models are shown where biggest improvements can be seen. PITs are shown for 0600 UTC, 1200 UTC, 1800 UTC, and 0000 UTC to check the characteristics for different times of the day. Top down PITs for summer season, winter season, and the full year are shown to highlight seasonal differences in calibration. Forecasts for day one, three, and four show a very similar picture (not shown).

When focusing on the Gaussian model several features can be identified. During summer (top

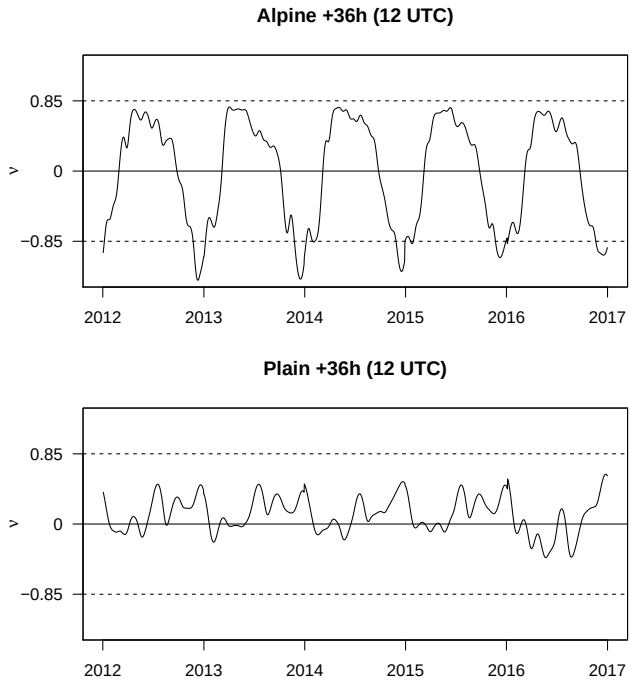


Figure 5: Time series of the empirical skewness $\nu$ of the forecasted skewed logistic distribution for lead time +36 h for the Alpine station (top) and plain site (bottom). Symmetric forecasts would show a value of 0.

row) a pronounced peak can be seen around $0.5 - 0.8$ with varying position over the time of the day. While approximately centered at midnight ($\approx 0.5$), this peak is slightly moved towards higher values for all other times of the day. The convex shape of the Gaussian results indicate overdispersion, while the asymmetry indicates residual skewness. This is likely caused by the not yet resolved topography in the NWP. Due to the volume effect (Whiteman 1990; Steinacker 1984), the heating of the valley atmosphere is stronger than predicted by the EPS, so that more observations fall into the upper part of the forecasted distribution.

The opposite occurs during winter (middle row). A strong peak in the observed frequency is shown at the lower tail ($< 0.1$) for all four lead times, strongest for 1200 UTC. An observed frequency of 3 (e.g., for the lowest 5% bin for winter time, 1200 UTC) means that observations falling into this interval are observed three times more often than expected. This is most likely associated with cold pools often formed within the not yet fully resolved valley during the cold season. Furthermore, an overall decrease of the frequencies can be seen towards higher bins.

These two distinct patterns can also be seen in the all-year PITs (bottom row) even if strongly damped due to the seasonal averaging. Overall, the results indicate that the Gaussian regression model is not able to capture all necessary characteristics of the data.

In contrast, the skewed logistic model (Figure 4) shows overall better calibration. For all seasons (top down) and all times of the day (left to right) the majority of all observed frequencies lie within the confidence interval. Only during summer season (top row) the frequencies seem to be slightly too low for high values ($> 0.9$).

Due to its properties the skewed logistic distribution is able to capture the lower-tail characteristics during winter and, at the same time, allows to properly adjust the distribution for the warm season. Top panel of Figure 5 shows a time series of the empirical skewness for the skewed logistic models for all +36 h forecasts (1200 UTC) over the whole validation time period. During the warm season the predictions are positively skewed with values around 0.5. Contrarily, strong negative skewness with values of −1.0 and below can be seen during the cold season. The consideration of this seasonally-dependent skewness yields an overall better performance compared to the Gaussian model.

### 3.2. Alpine vs. plain site

To see the benefits of a non-symmetric response distribution in a different environment, the same study is shown for a selected plain site (large square in Figure 2 and right column in Figure 3).

Similar to the Alpine site, a pronounced diurnal cycle is visible for all models in terms of LS and CRPS (Figure 3) with better scores for nighttime. In contrast to the Alpine site, a clear decrease in forecast performance with increasing lead time can be seen. Moreover, almost all scores (LS, CRPS, PIW) are smaller than at the Alpine site even at the highest lead time. However, the two heavy-tailed models (logistic and skewed logistic) improve sharpness (80% PIW) and calibration (RI) even more. Residual skewness is also smaller than at the Alpine site as Figure 5 shows. Additionally, the change in sign of skewness between warm and cold season is nearly absent. Skewness is still present, however, the amplitude is way lower than for the Alpine site and shows values relatively close to 0 (symmetric). Even if the improvements over the symmetric logistic models are only minor, the additional skewness still yields slightly better results.

In comparison to the Alpine site, the plain site shows an overall better forecast performance measures except for RI where both stations show similar scores. This is mainly due to the overall better performance of the NWP for regions with no or only little topography. In such situations the overall performance of the NWP is already adequate and the EPS provides covariates with more information content. Thus, benefit of the statistical post-processing methods is much smaller compared to sites in complex terrain. In this example the Gaussian assumption seems to be an appropriate choice and the improvements of the logistic or skewed logistic distribution is only minor.

### 3.3. Comparison for all sites

The following section extends the comparison to the entire study area including all 27 sites for all three groups (Figure 2).

Figure 6 shows averaged scores for LS, CRPS, the mean 80% PIW, and RI for the three different groups. Each box-whisker contains the mean score for the individual stations and all 15 lead times. This yields $12 \times 15$ values for group 'Alpine', $6 \times 15$ for group 'foreland', and $9 \times 15$ for group 'plain'. In addition, the numeric values of all medians are provided in Table 2 alongside with median values for two alternative PIWs (50% and 95%) and the prediction interval coverage (PIC) for the same three intervals. The validation shows increasing forecast performance with decreasing topographical complexity (top down) independent of the statistical model.

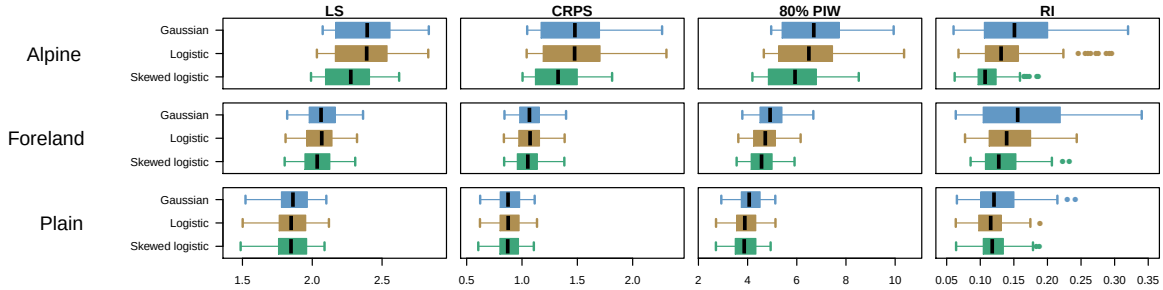


Figure 6: Performance measures in terms of LS, CRPS, 80% PIW, and RI (left to right), clustered for Alpine, foreland, and plain sites (top to bottom). The box-whiskers are based on average scores for each station and lead time.

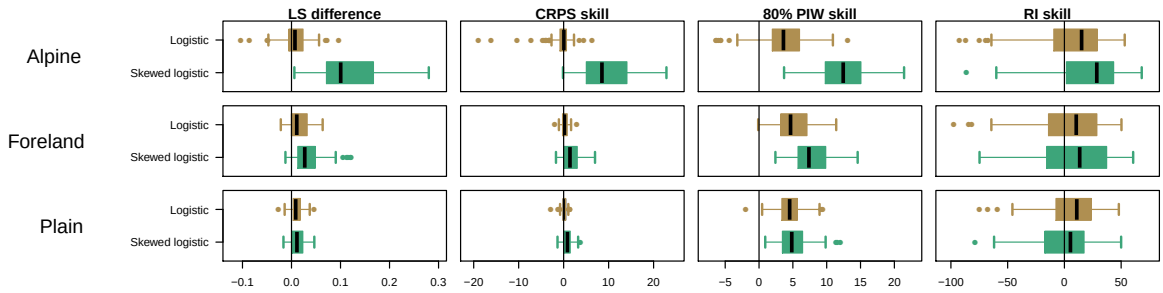


Figure 7: As in Figure 6, but showing values of improvements against the classical Gaussian model. Note that improvements are reported by positive values. Differences are shown for LS, whereas skill scores in % are shown for CRPS, the 80% PIW, and RI.

| | **Model** | **LS** | **CRPS** | **RI** | **PI 50%** | | **PI 80%** | | **PI 95%** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PIW | PIC | PIW | PIC | PIW | PIC |
| | Gaussian | 2.39 | 1.48 | 0.15 | 3.5 | 52.9 | 6.7 | 81.3 | 10.2 | 94.2 |
| **Alpine** | Logistic | 2.39 | 1.48 | 0.13 | 3.2 | 49.6 | 6.5 | 79.8 | 10.8 | 94.8 |
| | Skewed logistic | 2.28 | 1.33 | 0.11 | 3.0 | 48.8 | 5.0 | 79.7 | 9.9 | 94.9 |
| | Gaussian | 2.06 | 1.07 | 0.16 | 2.6 | 52.3 | 4.9 | 82.2 | 7.5 | 94.7 |
| **Foreland** | Logistic | 2.07 | 1.07 | 0.14 | 2.4 | 48.8 | 4.7 | 80.1 | 7.9 | 95.2 |
| | Skewed logistic | 2.04 | 1.05 | 0.13 | 2.3 | 47.7 | 4.6 | 79.3 | 7.6 | 95.0 |
| | Gaussian | 1.86 | 0.87 | 0.12 | 2.0 | 52.9 | 4.1 | 82.1 | 6.2 | 95.0 |
| **Plain** | Logistic | 1.85 | 0.88 | 0.12 | 1.9 | 49.0 | 3.9 | 80.2 | 6.5 | 95.6 |
| | Skewed logistic | 1.85 | 0.87 | 0.12 | 1.9 | 48.9 | 3.9 | 80.2 | 6.0 | 95.4 |

Table 2: Median of (left to right) the logarithmic score (LS), continuous ranked probability score (CRPS), reliability index (RI), and three prediction intervals (PI) reporting width (PIW) and prediction interval coverage (PIC) for Alpine, foreland, and plain sites (top to bottom), evaluated for each model type (Gaussian, logistic, skewed logistic).

Figure 7 shows the improvements from using non-Gaussian distributions, relative to the Gaussian reference model. Positive values indicate that the alternative model show a profit over the Gaussian one. The models using a logistic or skewed logistic distribution are able to outperform the Gaussian model in all scores. The results from the models using the symmetric logistic distribution show only minor improvements in terms of LS and CRPS but can significantly reduce the 80% PIW by $3.6 - 4.8\%$ (median). Smaller improvements can also be seen in the RI (cf. Section 3.1 and 3.2).

On the other hand, even stronger improvements can be found for the skewed logistic models. In median, the skewed logistic models outperform the Gaussian models in all scores and have the highest skill, except for RI at the plain sites where the logistic models performed best. Improvements are particularly large at the Alpine sites where improvements of about 8% for CRPS (median) up to 23% for certain stations and lead times can be achieved. The calibration (RI) is also improved by about 30% in median, although for some stations and lead times RI values become worse than for the Gaussian model. A similar picture with an overall smaller magnitude is shown for the foreland and plain sites with a median CRPS skill score of 1.4%/0.9% and a median 80% PIW skill score of 7.4%/4.8%. Most of the improvements can be attributed to the increased sharpness (PIW), which also yields overall smaller LS and CRPS.

## 4. Summary and conclusion

Non-homogeneous regression is a widely used statistical method for post-processing numerical ensemble forecast. It was originally developed to improve probabilistic air temperature forecasts and assumes a Gaussian response distribution.

However, several studies have shown that marginal temperature distributions can be skewed or non-symmetric, respectively (Warwick and Curran 1993; Harmel *et al.* 2002). This marginal skewness can result from topographically induced effects such as cold pools during winter or a strong valley bottom heating within narrow valleys during hot summer days. Thus, skewness is much stronger for locations surrounded by complex terrain than for sites in plain regions.

Moreover, skewness is supposed to decrease if additional covariates are included in the Gaussian model (see e.g., Feldmann *et al.* 2015; Möller and Groß 2016; Messner *et al.* 2017). However, the calibration of the results presented in these articles indicate that residual skewness remains, even by including more variables than only the ensemble temperature covariate. Thus, the skewness might need to be included using an appropriate response distribution.

In this study, the skewed logistic distribution was used and compared to the (symmetric) logistic and Gaussian distributions for probabilistic post-processing of 2 m air temperature at 27 sites in Central Europe for stations in three different environments: Alpine, foreland close to the Alps, and sites located in plain regions. The skewed logistic distribution furthermore allows to directly handle possible skewness in the data, if needed.

The two logistic distributions better perform for one-day up to four-day ahead forecasts for the majority of all stations and lead times. Improvements from using the skewed logistic distribution are largest for Alpine sites for the two most important probabilistic properties 'sharpness' and 'calibration'. The amount of improvement decreases with decreasing complexity of the topography.

When PIT histograms are used to check for calibration, they have to be checked for different

seasons, lead times, and hours of day. Averaging over the whole year or multiple times of the day may mask shortcomings especially in complex terrain and the distinct patterns as shown in the results might easily get overseen.

In conclusion, the Gaussian assumption for probabilistic temperature post-processing may be appropriate for regions where the ensemble provides sufficient information about the marginal distribution of the response. However, if the covariates used in the regression model miss some features, residual skewness is often found in the results. An alternative response distribution, such as the proposed skewed logistic distribution, allows to directly address unresolved skewness and is able to (strongly) increase the predictive performance of the probabilistic forecasts.

## 5. codeavailability

The results of this study have been achieved using the *R* package **bamlss** (Umlauf, Klein, and Zeileis 2018), where a new family for the generalized logistic type I distribution has been implemented and is now available on R-forge using the distributional properties from the *R* package **glogis** (Zeileis and Windberger 2014). Model estimation is performed using a gradient boosting approach with a 10-fold cross-validation to find the optimal stopping iteration for the boosting based on RMSE.

## 6. Acknowledgments

Results partly achieved using the HPC infrastructure LEO of the University of Innsbruck. This project was partially funded by Doctoral funding of the University of Innsbruck, Vizerektorat für Forschung, and the Austrian Research Promotion Agency (FFG), project 'ProfCast', grant No. 858537.

## A. Skewness of the skewed logistic distribution

The $3^{rd}$ moment (skewness, $\nu$) is a function of the shape parameter $\zeta$:

$$\nu(\zeta) = \frac{\Psi''(\zeta) - \Psi''(1)}{(\Psi'(\zeta) + \Psi'(1))^{\frac{3}{2}}} \tag{9}$$

where $\Psi'$ and $\Psi''$ denote the first and second derivative of the polygamma function $\Psi(x)$ (Abramowitz and Stegun 1965, Section 6.4.1, page 260) defined as:

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}. \tag{10}$$

Herein, $\Gamma(x)$ denotes the Gamma function (Abramowitz and Stegun 1965, Section 6.1.1, page 255) and $\Gamma'(x)$ its first derivative. The Gamma function is defined as:

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt \tag{11}$$

## References

Abramowitz M, Stegun IA (1965). "Handbook of mathematical functions with formulas, graphs and mathematical tables (National Bureau of Standards Applied Mathematics Series No. 55)." *Journal of Applied Mechanics*, **32**.

Bauer P, Thorpe A, Brunet G (2015). "The quiet revolution of numerical weather prediction." *Nature*, **525**, 47–55. doi:10.1038/nature14956.

Dabernig M, Mayr GJ, Messner JW, Zeileis A (2017). "Spatial ensemble post-processing with standardized anomalies." *Quarterly Journal of the Royal Meteorological Society*, **143**(703), 909–916. doi:10.1002/qj.2975.

Dawid A (1984). "Present position and potential developments: Some personal views: Statistical theory: The prequential approach." *Journal of the Royal Statistical Society: Series A (General)*, **147**(2), 278–292. doi:10.2307/2981683.

Delle Monache L, Hacker JP, Zhou Y, Deng X, Stull RB (2006). "Probabilistic aspects of meteorological and ozone regional ensemble forecasts." *Journal of Geophysical Research*, **111**(24), 1–15. doi:10.1029/2005JD006917.

Feldmann K, Scheuerer M, Thorarinsdottir TL (2015). "Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression." *Monthly Weather Review*, **143**(2005), 955–971. doi:10.1175/MWR-D-14-00210.1.

Gebetsberger M, Messner JW, Mayr GJ, Zeileis A (2017). "Estimation methods for nonhomogeneous regression models – Minimum continuous ranked probability score vs. maximum likelihood." *Working papers*, Faculty of Economics and Statistics, University of Innsbruck. URL https://www2.uibk.ac.at/downloads/c4041030/wpaper/2017-23.pdf.

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, Katzfuss M (2014). "Probabilistic forecasting." *Annual Review of Statistics and Its Application*, **1**, 125–151. doi:10.1146/annurev-statistics-062713-085831.

Gneiting T, Raftery AE, Westveld AH, Goldman T (2005). "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation." *Monthly Weather Review*, **133**, 1098–1118. ISSN 1520-0493. doi:10.1175/MWR2904.1.

Hagedorn R, Hamill T, Whitaker J (2008). "Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: two-meter temperatures." *Monthly Weather Review*, **136**(7), 2608–2619. doi:10.1175/2007MWR2410.1.

Harmel RD, Richardson CW, Hanson CL, Johnson GL (2002). "Evaluating the adequacy of simulating maximum and minimum daily air temperature with the normal distribution." *Journal of Applied Meteorology*, **41**(7), 744–753. doi:10.1175/1520-0450(2002)041<0744:Etaosm>2.0.Co;2.

Hersbach H (2000). "Decomposition of the continuous ranked probability score for ensemble prediction systems." *Weather and Forecasting*, **15**, 559–570. doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Klein N, Kneib T, Lang S, Sohn A (2015). "Bayesian structured additive distributional regression with an application to regional income inequality in Germany." *The Annals of Applied Statistics*, **9**(2), 1024–1052. doi:10.1214/15-AOAS823.

Leith C (1974). "Theoretical skill of Monte Carlo forecasts." *Monthly Weather Review*, **102**, 409–418. doi:10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

Lorenz EN (1963). "Deterministic nonperiodic flow." *Journal of the Atmospheric Sciences*, **20**, 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Messner JW, Mayr GJ, Wilks DS, Zeileis A (2014). "Extending extended logistic regression: Extended versus separate versus ordered versus censored." *Monthly Weather Review*, **142**, 3003–3014. doi:10.1175/MWR-D-13-00355.1.

Messner JW, Mayr GJ, Zeileis A (2017). "Nonhomogeneous boosting for predictor selection in ensemble postprocessing." *Monthly Weather Review*, **145**(1), 137–147. doi:10.1175/MWR-D-16-0088.1.

Möller A, GroßJ (2016). "Probabilistic temperature forecasting based on an ensemble autoregressive modification." *Quarterly Journal of the Royal Meteorological Society*, **142**(696), 1385–1394. doi:10.1002/qj.2741.

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005). "Using Bayesian model averaging to calibrate forecast ensembles." *Monthly Weather Review*, **133**, 1155–1174. ISSN 0027-0644. doi:10.1175/MWR2906.1.

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

Schefzik R, Thorarinsdottir TL, Gneiting T (2013). "Uncertainty quantification in complex simulation models using ensemble copula coupling." *Statistical Science*, **28**(4), 616–640.

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**, 405–422. doi:10.1111/rssc.12040.

Stauffer R, Umlauf N, Messner JW, Mayr GJ, 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.

Steinacker R (1984). "Area height distribution of a valley and its relation to the valley wind." *Beiträge zur Physik der Atmosphäre*, **57**, 64–71.

Toth Z, Szentimrey T (1990). "The binormal distribution: A distribution for representing asymmetrical but normal-like weather elements." *Journal of Climate*, **3**(1), 128–136. doi:10.1175/1520-0442(1990)003<0128:TBDADF>2.0.CO;2.

Umlauf N, Klein N, Zeileis A (2018). "BAMLSS: Bayesian additive models for location, scale and shape (and beyond)." *Journal of Computational and Graphical Statistics*. doi:10.1080/10618600.2017.1407325. Forthcoming.

Verkade JS, Brown JD, Reggiani P, Weerts H (2013). "Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales." *Journal of Hydrology*, **501**, 73–91. doi:10.1016/j.jhydrol.2013.07.039.

Warwick G, Curran E (1993). "A binormal model of frequency distributions of daily maximum temperature." *Australian Meteorological Magazine*, **42**, 151–161.

Whiteman C (1990). *Observations of thermally developed wind systems in mountainous terrain*. Atmospheric processes over complex terrain. Meteorological Monographs. Blumen W. doi:10.1007/978-1-935704-25-6_2.

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

Wilks DS (2011). *Statistical methods in the atmospheric sciences*. 3 edition. Academic Press.

Wilks DS (2017). "On assessing calibration of multivariate ensemble forecasts." *Quarterly Journal of the Royal Meteorological Society*, **143**(702), 164–172. doi:10.1002/qj.2906.

Zängl G (2004). "A reexamination of the valley wind system in the Alpine Inn Valley with numerical simulations." *Meteorology and Atmospheric Physics*, **87**(4), 241–256. doi:10.1007/s00703-003-0056-5.

Zeileis A, Windberger T (2014). *glogis: Fitting and testing generalized logistic distributions*. URL https://CRAN.R-project.org/package=glogis.

## Affiliation:

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

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

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

https://www.uibk.ac.at/eeecon/wopec/

2018-06 **Manuel Gebetsberger, Reto Stauffer, Georg J. Mayr, Achim Zeileis:** Skewed logistic distribution for statistical temperature post-processing in mountainous areas

2018-05 **Reto Stauffer, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Hourly probabilistic snow forecasts over complex terrain: A hybrid ensemble postprocessing approach

2018-04 **Utz Weitzel, Christoph Huber, Florian Lindner, Jürgen Huber, Julia Rose, Michael Kirchler:** Bubbles and financial professionals

2018-03 **Carolin Strobl, Julia Kopf, Raphael Hartmann, Achim Zeileis:** Anchor point selection: An approach for anchoring without anchor items

2018-02 **Michael Greinecker, Christopher Kah:** Pairwise stable matching in large economies

2018-01 **Max Breitenlechner, Johann Scharler:** How does monetary policy influence bank lending? Evidence from the market for banks' wholesale funding

2017-27 **Kenneth Harttgen, Stefan Lang, Johannes Seiler:** Selective mortality and undernutrition in low- and middle-income countries

2017-26 **Jun Honda, Roman Inderst:** Nonlinear incentives and advisor bias

2017-25 **Thorsten Simon, Peter Fabsic, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Probabilistic forecasting of thunderstorms in the Eastern Alps

2017-24 **Florian Lindner:** Choking under pressure of top performers: Evidence from biathlon competitions

2017-23 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Estimation methods for non-homogeneous regression models: Minimum continuous ranked probability score vs. maximum likelihood

2017-22 **Sebastian J. Dietz, Philipp Kneringer, Georg J. Mayr, Achim Zeileis:** Forecasting low-visibility procedure states with tree-based statistical methods

2017-21 **Philipp Kneringer, Sebastian J. Dietz, Georg J. Mayr, Achim Zeileis:** Probabilistic nowcasting of low-visibility procedure states at Vienna International Airport during cold season

2017-20 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior

2017-19 **Martin Geiger, Richard Hule:** The role of correlation in two-asset games: Some experimental evidence

2017-18 **Rudolf Kerschbamer, Daniel Neururer, Alexander Gruber:** Do the altruists lie less?

2017-17 **Meike Köhler, Nikolaus Umlauf, Sonja Greven:** Nonlinear association structures in flexible Bayesian additive joint models

2017-16 **Rudolf Kerschbamer, Daniel Muller:** Social preferences and political attitudes: An online experiment on a large heterogeneous sample

2017-15 **Kenneth Harttgen, Stefan Lang, Judith Santer, Johannes Seiler:** Modeling under-5 mortality through multilevel structured additive regression with varying coefficients for Asia and Sub-Saharan Africa

2017-14 **Christoph Eder, Martin Halla:** Economic origins of cultural norms: The case of animal husbandry and bastardy

2017-13 **Thomas Kneib, Nikolaus Umlauf:** A primer on bayesian distributional regression

2017-12 **Susanne Berger, Nathaniel Graham, Achim Zeileis:** Various versatile variances: An object-oriented implementation of clustered covariances in R

2017-11 **Natalia Danzer, Martin Halla, Nicole Schneeweis, Martina Zweimüller:** Parental leave, (in)formal childcare and long-term child outcomes

2017-10 **Daniel Muller, Sander Renes:** Fairness views and political preferences - Evidence from a large online experiment

2017-09 **Andreas Exenberger:** The logic of inequality extraction: An application to Gini and top incomes data

2017-08 **Sibylle Puntscher, Duc Tran Huy, Janette Walde, Ulrike Tappeiner, Gottfried Tappeiner:** The acceptance of a protected area and the benefits of sustainable tourism: In search of the weak link in their relationship

2017-07 **Helena Fornwagner:** Incentives to lose revisited: The NHL and its tournament incentives

2017-06 **Loukas Balafoutas, Simon Czermak, Marc Eulerich, Helena Fornwagner:** Incentives for dishonesty: An experimental study with internal auditors

2017-05 **Nikolaus Umlauf, Nadja Klein, Achim Zeileis:** BAMLSS: Bayesian additive models for location, scale and shape (and beyond)

2017-04 **Martin Halla, Susanne Pech, Martina Zweimüller:** The effect of statutory sick-pay on workers' labor supply and subsequent health

2017-03 **Franz Buscha, Daniel Müller, Lionel Page:** Can a common currency foster a shared social identity across different nations? The case of the Euro.

2017-02 **Daniel Müller:** The anatomy of distributional preferences with group identity

2017-01 **Wolfgang Frimmel, Martin Halla, Jörg Paetzold:** The intergenerational causal effect of tax evasion: Evidence from the commuter tax allowance in Austria

University of Innsbruck

Working Papers in Economics and Statistics

2018-06

Manuel Gebetsberger, Reto Stauffer, Georg J. Mayr, Achim Zeileis

# Skewed logistic distribution for statistical temperature post-processing in mountainous areas

**Abstract**
Non-homogeneous post-processing is often used to improve the predictive performance of probabilistic ensemble forecasts. A common quantity to develop, test, and demonstrate new methods is the near-surface air temperature frequently assumed to follow a Gaussian response distribution. However, Gaussian regression models with only few covariates are often not able to account for site-specific local features leading to strongly skewed residuals. This residual skewness remains even if many covariates are incorporated. Therefore, a simple refinement of the classical non-homogeneous Gaussian regression model is proposed to overcome this problem by assuming a skewed response distribution to account for possible skewness. This study shows a comprehensive analysis of the performance of non-homogeneous post-processing for 2m temperature for three different site types comparing Gaussian, logistic, and skewed logistic response distributions. Satisfying results for the skewed logistic distribution are found, especially for sites located in mountainous areas. Moreover, both alternative model assumptions but in particular the skewed response distribution, can improve on the classical Gaussian assumption with respect to overall performance, sharpness, and calibration of the probabilistic predictions.

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