Probabilistic temperature post-processing using a skewed response distribution

Manuel Gebetsberger[1], Georg J. Mayr[1], Reto Stauffer[2], Achim Zeileis[2]

[1] Institute of Atmospheric and Cryospheric Sciences, Univ. of Innsbruck, Austria
[2] Department of Statistics, Univ. of Innsbruck, Austria

E-mail for correspondence: `Manuel.Gebetsberger@uibk.ac.at`

Abstract: Weather forecasts are typically based on numerical weather prediction models, where possible forecast errors can be corrected by statistical post-processing methods. A common quantity to develop, test, and demonstrate new post-processing methods is the surface air temperature which is frequently assumed to follow a Gaussian distribution. Nevertheless, the classical Gaussian distributional regression models which use only few covariates are not able to account for all local features leading to strongly skewed residuals. The authors demonstrate two approaches to overcome this problem: assuming a skewed response distribution to directly account for skewness, and extending the classical Gaussian distributional regression model by including all available information from the weather prediction model in combination with a boosting variable selection method.

The preliminary findings show satisfying results especially for an Alpine station by either using a generalized logistic type I distribution with few covariates, or using all available information plus an appropriate variable selection procedure. Both approaches are able to improve the predictions with respect to overall performance and calibration. Furthermore, similar results can be achieved for non-Alpine sites although with smaller improvements.

Keywords: Temperature; Ensemble; Forecast; Skewed; Distributional regression

1 Introduction

Weather forecasts are typically generated by numerical weather prediction (NWP) models. Nowadays, ensemble prediction systems (EPSs) are widely used where

multiple NWP runs with slightly perturbed initial conditions and parameterizations try to capture the forecast uncertainty. Nevertheless, it was found that such models often show systematic errors due to simplified physical equations, insufficient resolution, and unresolved processes (Bauer et al. 2015).

One possibility to correct for these errors are statistical post-processing methods. These methods have been tested extensively for various forecast quantities, especially surface air temperature. Distributional regression models (Klein et al. 2015) are one common post-processing strategy. Corrected probabilistic forecasts are obtained by modeling the parameters of a response distribution using linear predictors including covariates provided by an EPS.

Statistical models using the Gaussian assumption are generally able to improve the EPS output and to produce well-calibrated probabilistic temperature forecasts (Gneiting et al. 2015). However, this is not true for all geographical locations. Atmospheric processes associated with unresolved topographical features such as wintry cold pools are per definition not present in the EPS model and often lead to a skewed distribution. If these effects (e.g. wintry cold pools, strong valley heating during summer) cannot be depicted by the regression model, the residuals might be (strongly) skewed leading to inappropriately calibrated probabilistic forecasts.

To overcome possible skewness-related errors, this study assesses two different strategies for an Alpine and two non-Alpine stations. The benefit of a skewed response distribution will be evaluated using different distributional regression models. In addition, it is tested whether additional covariates from the EPS remove the need of a skewed response distribution. Focus will be on the Alpine station where pronounced skewness is visible in observed surface air temperature records.

2   Regression Framework

Distributional regression models can be expressed in a general form:

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

where $\mathcal{D}$ represents the parametric distribution for the response y with θ_k, $k = 1, \ldots, K$ parameters. The parameters are linked to the additive predictors η_k using a monotone link function h_k (e.g., identity, logit, log). Each parameter $h_k(\theta_k)$ can be expressed by an additive predictor of the form:

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

including various (possibly non-linear) functions f_{pk}, $p = 1, \ldots, P$.

3   Data and Model Setup for Comparison

Results for three different sites and $+12h$–$+96h$ forecasts in 6-hourly intervals are shown: Innsbruck located in the European Alps, Munich in the Alpine foreland, and Hamburg in a plain environment. These sites are selected to ensure spatial

independence of observation-forecast data pairs and to investigate the influence of different topographical environments. The covariates are based on forecasts from the European Centre for Medium-Range Weather Forecasts. The EPS provides (discrete) probabilistic forecasts consisting of 50 exchangeable ensemble members for a wide range of atmospheric variables. Two types of regression models are estimated: "simple" models which only use $2m$ temperature forecasts, and "complex" models using all available information (110 covariates) plus a smooth cyclic spline effect representing the seasonality as a function of the day of the year (*season*). Table 1 shows a summary of the models and their specification. We make use of the logistic and generalized logistic distribution in order to test distributions with slightly heavier tails compared to the classical Gaussian distribution. Additionally, the generalized logistic distribution type I (Johnson et al. 1995) is able to account for possible skewness and has the distribution function

$$F(x) = 1/(1 + \exp(-(x - \mu)/\sigma))^\zeta, \tag{3}$$

where ζ defines the additional shape parameter that has to be estimated. Due to the large number of covariates in the "complex" models, a likelihood gradient boosting approach (R package **bamlss**, Umlauf et al. 2017, R package version 0.1-1, https://r-forge.r-project.org/projects/bayesr/) is used for variable selection and coefficient estimation. A 10-fold block-wise cross-validation is further performed to ensure full out-of-sample results.

The overall performance of the models is evaluated on the ignorance score (Ign), and PIT histograms to visually assess calibration (Gneiting et al. 2015). To quantify the information provided by individual PIT histograms the reliability index $RI = \sum_{i=1}^{I} |\kappa_i - \frac{1}{I}|$ is chosen (Feldmann et al. 2015). It accounts for absolute deviations of each PIT-bin from perfect uniform calibration, where I defines the number of bins and κ_i the relative fraction in each bin. Ideally, RI values are close to zero. Additionally, the sharpness of the predictive distributions is verified in terms of average width of predictions intervals which should be as small as possible.

4   Results and Discussion

The response distribution for surface temperature is often assumed to be Gaussian in statistical ensemble post-processing studies. This is obviously not appropriate for the investigated Alpine station used in this study. The simple Gaussian

TABLE 1. Covariates used for distribution parameters μ, σ, ζ for all response distributions. "mean"/"sd": ensemble means and standard deviations, and "*season*": smooth seasonal effect. Covariates in brackets only for "complex" models.

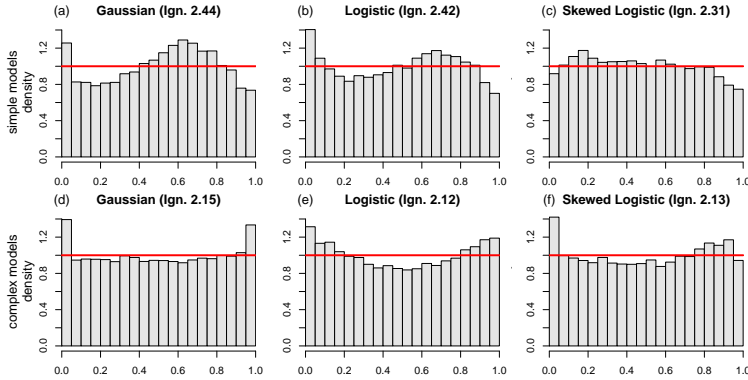| Response Distribution	location μ	scale σ	shape ζ
Gaussian	mean (+*season*)	sd	—
Logistic	mean (+*season*)	sd	—
Generalized Logistic (I)	mean (+*season*)	sd	*season* (+mean)

FIGURE 1. PIT histograms for the Alpine site. Left to right: Gaussian, logistic, and skewed logistic models. Top column shows the "simple" models, bottom column the "complex" models. Ignorance shown in brackets. Horizontal line indicates perfect calibration.

regression model shows uncalibrated and strongly skewed predictions (Fig. 1a). The skewed logistic distribution could clearly improve the predictive performance, indicated by a smaller ignorance and an almost uniformly distributed PIT histogram (Fig. 1c). This improvement is also visible for all individual lead times in terms of ignorance (Fig. 2a) and calibration (Fig. 2d).

A better overall performance in ignorance can be achieved using more complex models which can select from a large number of EPS covariates. Although additional covariates remove the need of a skewed response distribution, the complex Gaussian model still shows large errors in the tails (Fig. 1d), which might result from considerably smaller prediction intervals compared to simple Gaussian model (Fig. 2g).

Overall best ignorance for the Alpine site was achieved using all available covariates and a logistic response distribution (Fig. 1e), however, there is a distinct trade-off between perfect calibration, lowest possible ignorance, and also computational time. While simple models suggest the use of a skewed distribution, complex models perform very similar on all three distributions and imply that the Gaussian assumption might be sufficient. Nevertheless, these complex models using a cross-validated boosting procedure are roughly ten times more expensive in terms of computational time.

Similar results could also be found for the Alpine foreland site (Fig. 2b,e,h) and the topographically plain site (Fig. 2c,f,i), if only with smaller magnitude of improvements due to the two strategies and more visible at shorter lead times. At these two sites, the raw EPS has generally a better forecast performance due to less unresolved distinct local features in the EPS forecasts. Since better EPS forecasts lead to more informative covariates, the overall predictive performance of the statistical models is best at the plain site (Fig. 2c) and worst in the Alpine environment (Fig. 2a).
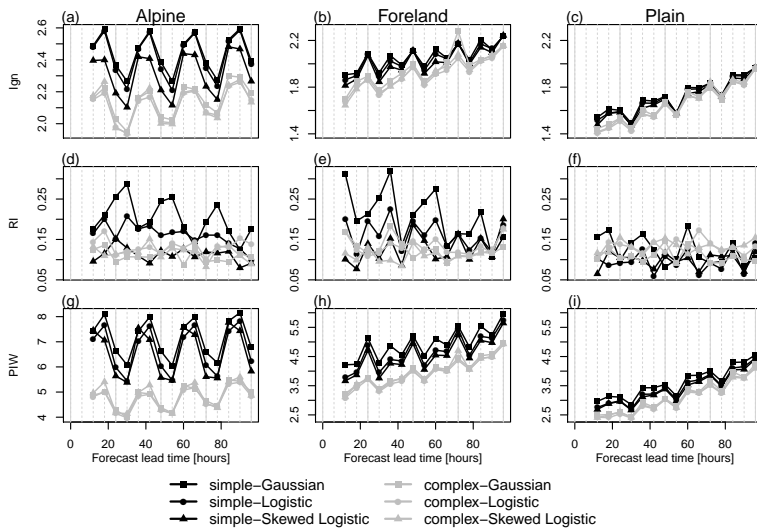
FIGURE 2. Verification measures for an Alpine, a foreland, and a plain site (left to right) using "simple" models (black lines) and "complex" models (grey lines). From top to bottom the ignorance score (Ign), reliability index (RI), and mean width of the 80% prediction interval (PIW) are shown, evaluated at forecasting lead times $+12h$–$+96h$. The 80% PIW defines the width between the predicted 0.1 and 0.9 quantile. Note that the Alpine site has a different range on the y-axis.

5   Conclusion

The results shown highlight the importance of an appropriate distributional assumption for post-processed surface air temperature forecasts, especially when only the corresponding temperature covariate is used in the regression model. A skewed logistic assumption improves calibration and forecast performance for all tested sites, but more pronounced in the Alpine environment.

Compared to this skewed regression model, complex models cannot further improve overall calibration but clearly obtain smaller prediction intervals which also lead to better forecast performance. Generally, the difference in forecast performance between the chosen distributional assumptions for the complex models is rather small. This indicates that the Gaussian assumption is already suitable given the large number of covariates. Nevertheless, these complex models are computationally roughly 10 times more expensive, and hence can be difficult to implement in an operational system.

The authors are currently extending this study by including additional sites to analyze the most influential covariates. A small set of such covariates might already lead to an improvement of similar magnitude, while considerably cutting down the computational costs. While the Gaussian assumption looks suitable for the complex models considered, it has to be investigated whether this is still true if the models only contain the most important covariates, especially for stations located in a complex environment.

References

Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, **525**, 47−55.

Feldmann, K., Scheuerer, M., and Thorarinsdottir, T. L. (2015). Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Monthly Weather Review*, **143**, 955−971.

Gneiting, T., Balabdaoui, F., and Raftery, A.E. (2015). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society*, **69(2)**, 243−268.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions., Volume 2. 2nd edition.* New York: Wiley & Sons.

Klein, N., Kneib, T., Lang, S., and 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.

Umlauf, N., Klein, N., and Zeileis, A. (2017). *BAMLSS: Bayesian additive models for location, scale and shape (and beyond).* Working Papers, Faculty of Economics and Statistics, University of Innsbruck. URL: http://EconPapers.repec.org/RePEc:inn 05.