University of Innsbruck

Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis

Working Papers in Economics and Statistics

2016-28

Research Platform eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

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 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

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

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

Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

Manuel Gebetsberger
Universität Innsbruck

Jakob W. Messner
Universität Innsbruck

Georg J. Mayr
Universität Innsbruck

Achim Zeileis
Universität Innsbruck

Abstract

Raw ensemble forecasts display large errors in predicting precipitation amounts and its forecast uncertainty, especially in mountainous regions where local effects are often not captured. Therefore, statistical post-processing is typically applied to obtain automatically corrected weather forecasts where precipitation represents one of the most challenging quantities. This study applies the non-homogenous regression framework as a start-of-the-art ensemble post-processing technique to predict a full forecast distribution and improves its forecast performance with three statistical tricks. First of all, a novel split-type approach effectively accounts for perfect ensemble predictions that can occur. Additionally, the statistical model assumes a censored logistic distribution to deal with the heavy tails of precipitation amounts. Finally, the optimization of regression coefficients for the scale parameter is investigated with suitable link-functions. These three refinements are tested for stations in the European Alps for lead-times from +24h to +48h and accumulation periods of 24 and 6 hours. Results highlight an improvement due to a combination of the three statistical tricks against the default post-processing method which does not account for perfect ensemble predictions. Probabilistic forecasts for precipitation amounts as well as the probability of precipitation events could be improved, especially for 6 hour sums.

Keywords: non-homogeneous regression, censored logistic distribution, log-link, probabilistic precipitation forecasts, operational forecasting.

1. Introduction

Physically-based ensemble forecasts define a standard in operational weather forecasting nowadays. Decades ago, they were developed to capture atmospheric forecast uncertainty (Leith 1974). Starting from slightly perturbed initial conditions, a dynamic model integrates a certain number of ensemble members forward in time. The individual ensemble members can then be seen as a random draw from a future state of the atmosphere. If an ensemble were perfect, the resulting differences in the member forecasts would only come from differences in their initial conditions. Hence, the true observation would be a random drawing from an

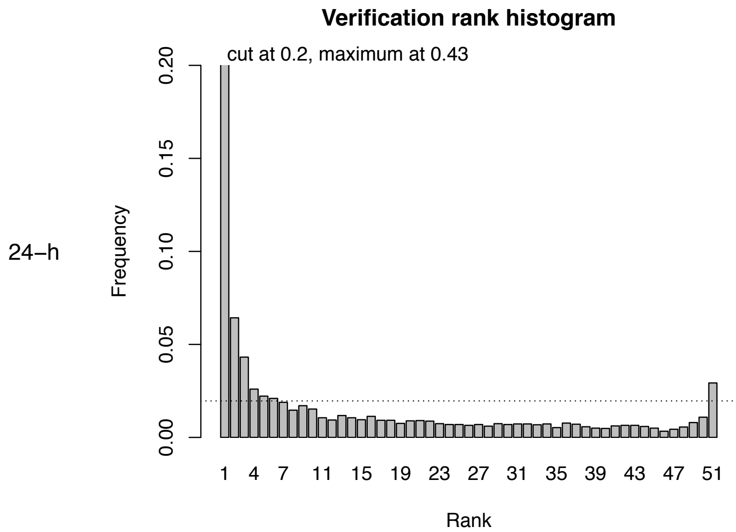

Figure 1: Rank histogram for 24-hour precipitation sums of the raw ECMWF ensemble forecasts for the region of Bolzano, Italy.

underlying distribution as well (Wilks 2011).

This can be analyzed in rank histograms (Hamill and Colucci 1998; Anderson 1995; Talagrand, Vautard, and Strauss 1997), where studies have shown that ensembles of precipitation forecasts have a systematic bias and not enough uncertainty (underdispersion) in their forecasts, especially in short-range forecasting (Hamill and Colucci 1998; Mullen and Buizza 2002). This results from the fact, that numerical weather prediction models mainly suffer from imperfect representation of real world physics, a lack of initial observational data, and missing topographical effects (Bauer, Thorpe, and Brunet 2015).

The European Alps represent a region with an extraordinary complex topography. Due to insufficient horizontal grid spacing, unresolved valleys and mountain ridges cause missing local effects for precipitation that occur in reality. Most of the precipitation is rained out before it reaches inner alpine valleys, leading to drying ratios of about 35% (Smith, Jiang, Fearon, Tabary, Dorninger, Doyle, and Benoit 2003).

Therefore, systematic errors and an underdispersion can be observed for this region, as illustrated in Figure 1 (see Section 3.3.1). This rank histogram highlights a strong bias due to the peak at rank 1, where precipitation amounts are strongly overestimated by the raw ensemble. Additionally, an underdisperion is visible since observations are mostly below the lowest and above the highest member forecast (on rank 1 and 51).

In order to correct for these errors and to supply automatically-corrected forecasts to weather services, the raw ensemble has to be post-processed. Numerous approaches for statistically post-processing ensembles exist that simultaneously correct for ensemble mean and ensemble dispersion (Roulston and Smith 2003; Gneiting, Raftery, Westveld III, and Goldman 2005; Raftery, Gneiting, Balabdaoui, and Polakowski 2005; Sloughter, Raftery, Gneiting, and Fraley 2007).

These methods have been tested extensively for different variables (e.g., temperature, mean

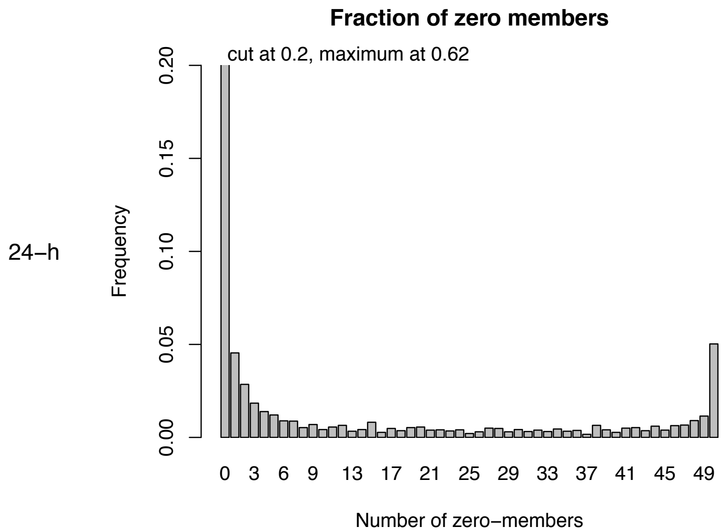

Figure 2: Frequency [%] of ensemble forecasts containing a certain number of members being zero (0-50), evaluated for the region of Bolzano, Italy for 24-hour sums.

sea level pressure, wind, and precipitation), and appropriate distributions: Gaussian (Gneiting *et al.* 2005), truncated normal (Thorarinsdottir and Gneiting 2010), gamma (Wilks 1990), generalized extreme value (GEV, Scheuerer 2014), or censored Gauss and logistic (Wilks 2009; Messner, Mayr, Zeileis, and Wilks 2014b; Messner, Mayr, Wilks, and Zeileis 2014a). Gneiting and Katzfuss (2014) nicely review suitable distributions for certain variables, and statistical ensemble post-processing and verification techniques in general.

However, only a few studies use more than ensemble mean and standard deviation as regressor variables in their statistical models. For precipitation, the fraction of ensemble members being zero, used as an additional regressor, can improve the post-processed forecasts (Sloughter *et al.* 2007; Bentzien and Friederichs 2012; Scheuerer 2014).

A closer look at this fraction for the region of Bolzano in the European Alps illustrates a special characteristic. Figure 2 tells us, that the complete ensemble is mostly positive (peak at zero) or all members are zero (peak at 50). In order to produce sharp forecasts for precipitation occurrence, this is what we suppose from the ensemble to do: forecasting probabilities close to 0 and 100%. Furthermore, Figure 3 shows the observed precipitation amount conditional on the fraction of ensemble members being zero, displayed for a representative station. The forecasts of all ensemble members predicting no precipitation are perfect in the sense that no precipitation has been observed. This figure also illustrates, that "perfect" cases can occur on lower split levels for the fraction of zeros as well (fraction larger than 0.1).

In order to account for those existing perfect cases within this Alpine region, we will introduce a new split-type approach as the first statistical trick and apply it to the non-homogeneous regression framework (NHR) of Gneiting *et al.* 2005. As a state-of-the-art post-processing method, this approach is generally able to provide a full probability distribution for precipitation amounts.

Nevertheless, a classical Gaussian distribution is not appropriate for precipitation data which

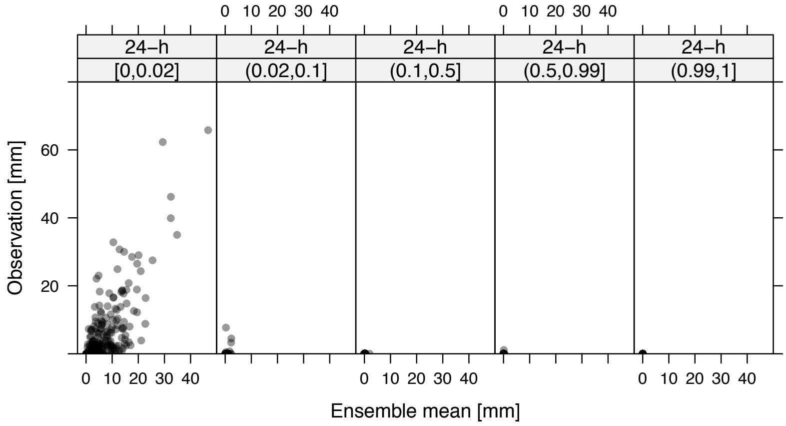

Figure 3: Ensemble mean values against observed precipitation at Bolzano for 24-hour sums: Columns show cases with a different fraction of zero EPS-members (0-0.02, 0.02-0.1, 0.1-0.5, 0.5-0.99, 0.99-1). x- and y-axis are in [mm].

have a physical limit at zero. Furthermore, events with large precipitation amounts can be underestimated on the Gaussian tails that are typically too weak. To overcome this, our second statistical trick assumes a heavy-tailed distribution which deals with those precipitation characteristics.

Additionally, the non-negativity of the distribution's dispersion parameter has to be ensured. Numerical optimization of regression coefficients can lead to negative dispersion parameters that should be avoided. According to established studies, this requirement can be achieved by squaring the optimization value (Gneiting *et al.* 2005), or by applying a link function to the dispersion sub-model (Messner *et al.* 2014a). A comparison of these concepts has not been made so far and will be performed in this study as the third statistical trick.

This article is structured as followed: In Section 2 we will explain our statistical tricks in detail. Further, Section 3 describes our study area and comparison setup. Section 4 presents our results, which will be concluded and summarized in Section 5.

2. Tricks

In this section we briefly describe the basic non-homogeneous regression (NHR) framework, followed by our three statistical tricks: split approach for perfect predictions, heavy tails, and link functions.

2.1. Non-homogeneous regression

Non-homogenous regression (NHR) was initially developed for a Gaussian response, e.g., temperature (Gneiting *et al.* 2005). This approach uses a linear model framework, where two linear equations are defined to correct for the location part (Eq. 1) and the scale part (Eq. 2), respectively. Typically, the Gaussian parameters for location and scale (μ_i, σ_i) are linearly linked to ensemble mean ($\overline{ens_i}$) and ensemble standard deviation ($SD_{ens,i}$) for each event i:

$$\mu_i = \beta_0 + \beta_1 \cdot \overline{ens_i} \tag{1}$$

$$\sigma_i = \gamma_0 + \gamma_1 \cdot SD_{ens,i} \tag{2}$$

The four coefficients ($\beta_0, \beta_1, \gamma_0, \gamma_1$) can be estimated simultaneously by numerically optimizing the log-likelihood function:

$$logLik = \sum_{i=1}^{N} \log(f(precip_i)) \tag{3}$$

which is defined as the sum over the logarithmic densities of the probability density function (PDF) f, evaluated at the observed value $precip_i$. Concerning the classical NHR approach, f defines the Gaussian PDF.

Since precipitation data are non-negatively defined and skewed to the right, this Gaussian NHR has to be modified. A simple approach is given by the concept of censoring at a certain threshold (Cohen 1959). Regarding precipitation, this threshold is typically defined at zero. Similar to the classical NHR, we assume a latent Gaussian process y which is allowed to become negative. As a result, this latent process has to be censored at zero to obtain sensible values for precipitation:

$$precip_i = \begin{cases} 0 & y_i \leq 0 \\ y_i & y_i > 0 \end{cases} \tag{4}$$

The log-likelihood function which has to be optimized differs to Eq. 3 by distinguishing between events on the censoring level ($precip_i = 0$) and above the censoring level ($precip_i > 0$):

$$logLik_i = \begin{cases} \log(F(0)) & precip_i = 0 \\ \log(f(precip_i)) & precip_i > 0 \end{cases} \tag{5}$$

where F represents the cumulative distribution function (CDF) and f the PDF, evaluated at the censoring level zero or the observed value $precip_i$, respectively.

2.2. Split approach

In the introduction we have already implied the importance to use the fraction of ensemble members being zero. Scheuerer (2014) already used this information for probabilistic precipitation forecasts in Germany. He added a new regressor variable *frac* into the location part of Eq. 1, which accounts for the fraction of K members being zero: $frac = \frac{1}{K}\sum_{k=1}^{K} 1_{member_k=0}$. Defined as a numeric value between 0 and 1, this additional regressor could improve the forecasts.

Concerning the European Alps, some specialities can be observed for *frac*: Figure 3 suggests that if (nearly) all ensemble members are zero (i.e., *frac* is high), precipitation may take very low (or even zero) values and the regression relationship almost collapses. Therefore, the influence of *frac* is better not captured by an additive regressor (as in Scheuerer 2014) but by an interaction term that can also be interpreted as splitting the data at a certain split level ν, further referred as "split approach".

More specifically we derive a binary variable z_i indicating whether (almost) all ensemble members are zero:

$$z_i = \begin{cases} 1 & \text{if } \mathit{frac} \geq \nu \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

If $\nu = 1$ then z_i indicates whether *all* ensemble members are zero but choosing lower split levels might also be useful (see below).

Subsequently, this new regressor enters the NHR equations in form of an interaction:

$$\mu_i = \beta_0 + \beta_1 \cdot \overline{ens_i} \cdot (1 - z_i) + \beta_2 \cdot z_i \tag{7}$$

$$\sigma_i = \gamma_0 + \gamma_1 \cdot SD_{ens,i} \cdot (1 - z_i) \tag{8}$$

which can be interpreted as follows: The usual censored NHR with slopes β_1 and γ_1, respectively, is only estimated for the cases with $z_i = 0$, i.e., where at least a certain fraction $(1 - \nu)$ of ensemble members indicate precipitation. Conversely if $z_i = 1$, i.e., (almost) all ensemble members indicate no precipitation, the regression collapses to a simple climatological mean $\mu_i = \beta_0 + \beta_2$ and standard deviation $\sigma_i = \gamma_0$.

Typically, the coefficient β_2 will be negative leading to lower predicted precipitation. Note that due to the censoring, the probability for positive precipitation may become arbitrary small if β_2 becomes increasingly negative. For this reason $\sigma_i = \gamma_0$ is also kept fixed to avoid that both mean and standard deviation collapse to zero.

The choice for the "best" split point between the NHR regression and simple climatological fit is not obvious and explored in more detail below. Considering Figure 2 it seems that $\nu = 1$ should be sufficient because there are only few observations with large fractions but below 1. However, from Figure 3 for station Bolzano it might also be reasonable to switch from the proper NHR regression to the climatological mean at a lower split point, e.g., $\nu = 0.5$ or even $\nu = 0.1$.

2.3. Heavy tails

Although the censored Gaussian distribution is able to capture precipitation characteristics (non-negativity, many observations at zero), more suitable distributions regarding rare events with large amounts of precipitation exist.

In this work we will additionally investigate the censored logistic distribution. By having a more pronounced tail than the Gaussian, the logistic distribution possibly accounts better for extreme cases on those tails and was found to be meaningful (Messner *et al.* 2014b). Censoring and log-likelihood maximization can be performed as in the Gaussian case described previously, but using the logistic PDF (Eq. 9) and CDF (Eq. 10). Note that σ defines the scale parameter, and μ the location parameter of the logistic distribution to be consistent with the censored NHR framework of Eq. 7 and 8.

$$f(y, \mu, \sigma) = \frac{e^{-\frac{y-\mu}{\sigma}}}{\sigma \cdot (1 + e^{-\frac{y-\mu}{\sigma}})^2} \tag{9}$$

$$F(y, \mu, \sigma) = \frac{1}{1 + e^{-\frac{y-\mu}{\sigma}}} \tag{10}$$

Clearly, there might be other suitable distributions accounting for rare events, e.g., generalized extreme value (GEV, Scheuerer 2014), censored shifted Gamma (Scheuerer and Hamill 2015), that can be used within this split approach as well.

Until now, we have defined a censored NHR-framework using the new split approach for a certain split level ν. Our third, and last refinement will now focus on the dispersion sub-model in Eq. 8.

2.4. Link functions

Since the scale parameter in Eq. 8 is non-negatively defined, we have to ensure that individual predictions are kept non-negative during optimization. This can be achieved in two ways: by parameter constraints for γ_0, γ_1 (e.g. squaring these coefficients, Gneiting *et al.* 2005), or by using a suitable link function (e.g. log-link, Messner *et al.* 2014b). Therefore, we will investigate possible differences of using different link-functions g for the scale sub-model:

$$g(\sigma) = \gamma_0 + \gamma_1 \cdot g(SD_{ens}) \cdot (1 - z) \tag{11}$$

In this study we compare the logarithmic-link (Messner *et al.* 2014b) and two other link-functions that use the parameter constraint of Gneiting *et al.* (2005):

- quadratic (quad): $g(\sigma) = \sigma^2$ (Gneiting *et al.* 2005)
- logarithmic (log:) $g(\sigma) = log(\sigma)$ (Messner *et al.* 2014b)
- identity (id): $g(\sigma) = \sigma$ (Scheuerer 2014)

Additionally, we use the same transformation g for the standard deviation on the right site of Eq. 11. Those link-functions mentioned here will be applied in our split approach.

3. Data and setup

This section defines the data for our research area and the comparison setup for the statistical models.

3.1. Data

As mentioned in the introduction, raw ensemble forecasts for precipitation amounts suffer from large location and dispersion errors (Figure 1). In order to correct for errors in a complex environment, our region of interest is embedded in the Central European Alps, in the North of Italy (Figure 4). A special concern for this region focuses on accurate probabilistic forecasts of precipitation amounts and the probability of precipitation for agricultural purposes. This region is famous for wine and fruit growing, where precipitation events and precipitation amounts can highly influence the evolution of plant pathogens (Löpmeier, Wittich, Frühauf, and Schittenhelm 2013; Carisse, Bacon, Lefebvre, and Lessard 2009).

The ensemble forecasts we want to correct for this region, are based on the operational ensemble prediction system (EPS) of the ECMWF, which consists of 50 ensemble members

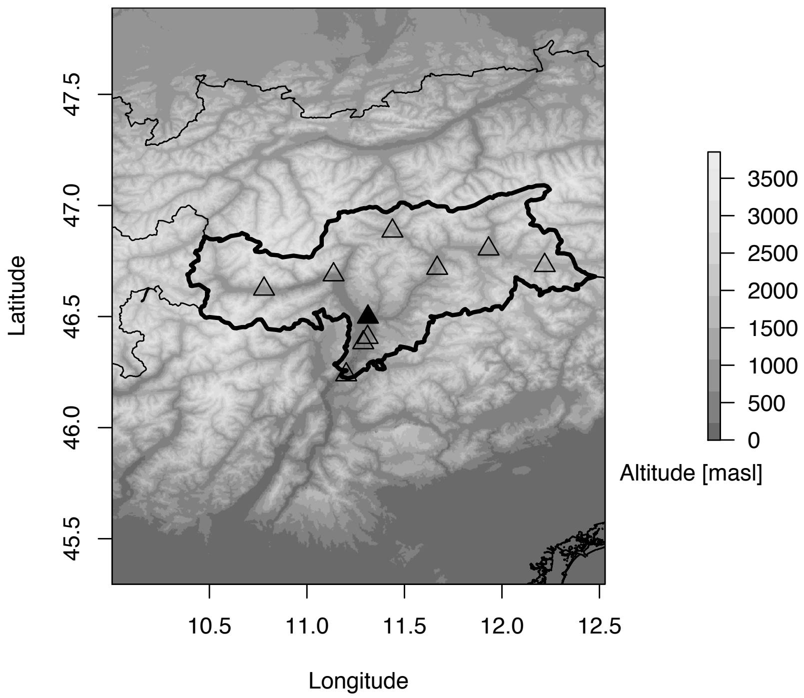

Figure 4: Region of interest: the bold area defines the region of Bolzano, Italy. Triangles represent our station sites, where the filled triangle denotes the station in Bolzano, that is used for demonstrative examples.

with a horizontal grid size of 32 km. The model output is then bilinearly interpolated to 10 station sites of interest.

Observed precipitation amounts are based on 10 minute measurements of automated weather stations, which are owned by the local weather service. The datasets cover the period from 2011-01-01 to 2014-01-01.

The forecast horizon of interest is for the second day from +24 h to +48 h, and based on the 00 UTC run of the ECMWF EPS. In this study, we will focus on different aggregation periods of 6- and 24-hour precipitation amounts.

3.2. Comparison setup

Table 1 gives an overview about the statistical models used in this article. In order to quantify the quality of our new split approach, we will use a reference model. The reference approach uses the quadratic link with a parameter constraint for the estimated scale coefficients (*quad*), as used by Gneiting *et al.* (2005) for temperature forecasts. This model is extended by using the fraction of members being zero (*quad_frac*), as proposed by Scheuerer 2014. Finally, we use our split approach with the quadratic-link (*quad_split*), the logarithmic-link (*log_split*), and the identity-link (*id_split*). Except for the log-link split model, all models use the parameter constraint of squaring the coefficients in the scale sub-model.

To have a fair comparison, a ten-fold cross-validation is performed for each individual case and results are evaluated on the left-out data.

The optimization itself is performed in R by using the *crch* package, which performs maximum

Model name	Zero information	Link function in scale-sub-model	Parameter constraint on γ_0, γ_1
quad	-	quadratic	power of 2
quad_frac	*frac*	quadratic	power of 2
quad_split approach	z	quadratic	power of 2
log_split	z	logarithmic	-
id_split	z	identity	power of 2

Table 1: Overview of statistical models used for comparison: zero information describes if the fraction of members being zero (*frac*) is used, the split approach with the z regressor, or no information used.

likelihood optimization (R Core Team 2016; Messner, Mayr, and Zeileis 2016).

4. Results

This section is structured as followed: first, we will briefly compare the statistical models to the raw ensemble, followed by the quantification of our three statistical tricks against the reference post-processing method.

4.1. Comparison to raw ensemble forecasts

It is essential that post-processing has to improve the raw ensemble forecasts. We therefore perform a brief ensemble evaluation with the continuous ranked probability score (CRPS, Hersbach 2000; Gneiting *et al.* 2005; Wilks 2011) for the probabilistic forecasts, and the Brier score (BS, Brier 1950) to check the probability of precipitation (PoP), both described in the following. Model performance is evaluated on the CRPS:

$$CRPS = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} (F_i^{fcst} - F_i^{obs})^2 \tag{12}$$

by comparing the forecasted and observed cumulative distribution function F.

In order to compare the performance of different statistical models (Table 1), we further compute the continuous ranked probability skill score (CRPSS):

$$CRPSS = 1 - CRPS_{mod}/CRPS_{ref} \tag{13}$$

where $CRPS_{mod}$ is each model score and $CRPS_{ref}$ our reference approach. Since the reference has no skill against itself, the CRPSS is zero.

Furthermore, our needed forecasts for PoP in our region are analyzed by the BS:

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2 \tag{14}$$

which is a mean squared difference between the forecast probabilities p_i and the binary value of precipitation yes or no o_i.

Although the ensemble does not provide a full continuous probability distribution, it is possible to verify the empirical CDF due to Hersbach (2000). Additionally, the percentage of ensemble

		CRPS		BS	
Type	name	24-h	6-h	24-h	6-h
Raw ensemble	EPS	1.53907	0.52670	0.46050	0.43435
Gaussian models	quad	1.17707	0.40141	0.10704	0.08531
	quad_frac	1.17659	0.39179	0.10679	0.08447
	quad_split	1.17685	0.40099	0.10681	0.08528
	log_split	1.16565	0.38180	0.10723	0.08430
	id_split	1.17077	0.39816	0.10639	0.08444
Logistic models	quad	1.16456	0.39369	0.10277	0.08319
	quad_frac	1.16462	0.38085	0.10278	0.08272
	quad_split	1.16439	0.39148	0.10284	0.08323
	log_split	1.16257	0.37852	0.10290	0.08270
	id_split	1.16136	0.38371	0.10246	0.08245

Table 2: Median CRPS and Brier scores (BS) for the raw ensemble, Gaussian and logistic models, evaluated separately for different accumulation periods (24,6 hour) over the research area.

members predicting precipitation can be used to verify the probability of precipitation (PoP). CRPS and BS values are summarized in Table 2.

Clearly, censored Gaussian and censored logistic models show lower CRPS values than the raw ensemble, both improving the raw forecasts significantly by a value of 25% on median. The CRPS is generally smaller for 6-hour sums, resulting from smaller precipitation amounts that are observed.

Regarding the PoP, the raw ensemble could also be improved significantly by all statistical models. 24-hour sums obtain a Brier score of 0.46 on median, and 6-hour sums a score of 0.42 on median. Compared to the raw ensemble, the post-processed forecasts of all statistical models improve about 75%.

4.2. Split approach and split levels

Since we have clarified a clear improvement against the raw ensemble, we will focus on the statistical models in the following.

Figure 5 summarizes CRPSS values for censored Gaussian and logistic models, relative to our reference approach where the squared scale parameter is optimized without additional information of members being zero (*quad*). The boxplots represent individual cases (lead-times) for each station, which are between +24h and +48h forecasts in advance : 24-hour sums include 10, and 6-hour 40 CRPSS values.

All split models (using $\nu = 1$) show an increased forecast skill which is even more dominant for 6-hour accumulation periods. Median values are highest for split models using the log-link. This pattern is similar for censored logistic models , especially for 6-hour sums. The model, which uses the fraction of zeros (*quad_frac*) leads to a similar magnitude of improvement as the split model with id-link, both for censored Gaussian and logistic simulations.

The previously described split models have all been performed on the split level of $\nu = 1$, where all members are zero. Since split models using the log-link performed best in the previous validation, we refitted this split model for different split levels ($\nu = 0.02, 0.1, 0.5$). The validation shows a decreasing CRPSS median for split models that perform the split at

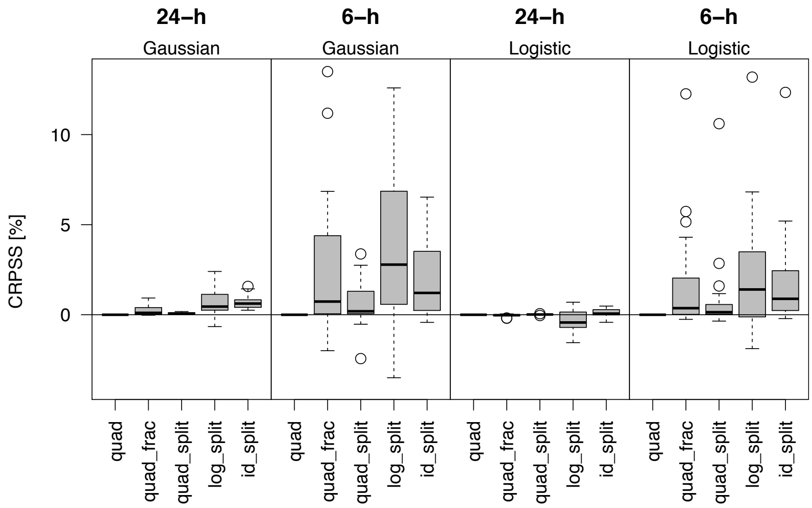

Figure 5: Continuous ranked probability skill score (CRPSS) for censored Gaussian and logistic models in reference to the *quad*-model without splitting. Results from ten-fold cross-validation of data for 10 stations and different lead-times (+24h to +48h), evaluated separately for different accumulation periods.

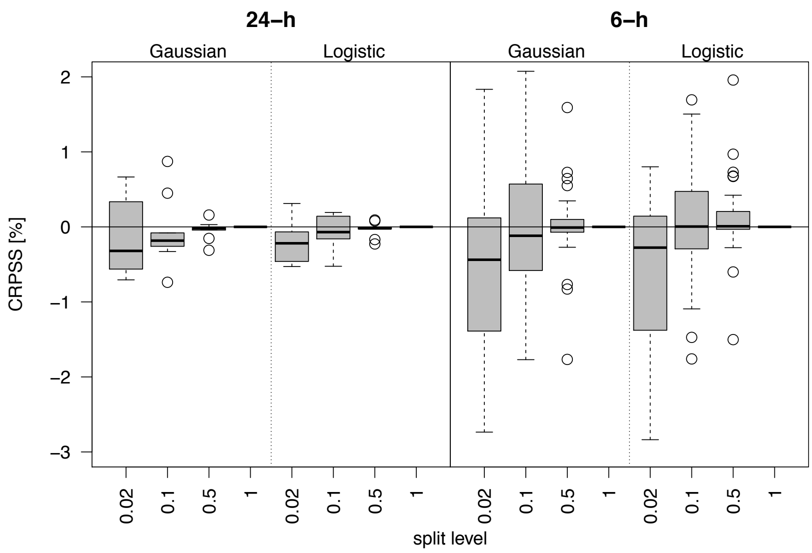

Figure 6: Continuous ranked probability skill score (CRPSS) for censored Gaussian and logistic split models using the log-link, in reference to the split model splitted at $\nu = 1$. Results from ten-fold cross-validation of data for 10 stations and different lead-times (+24h to +48h), evaluated separately for different accumulation periods. The reference model in each accumulation period has a skill of zero.

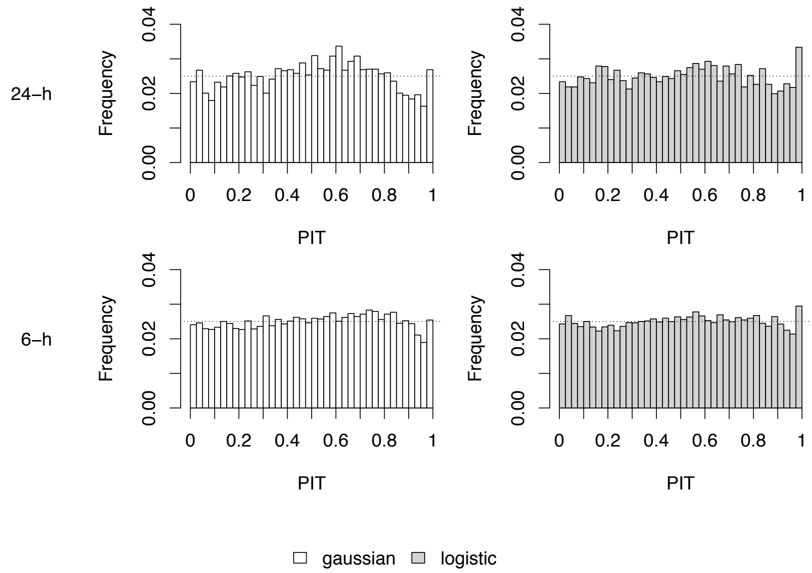

Figure 7: Probability Integral Transform (PIT) analysis for log-link models with censored Gaussian predictions (left column) and censored logistic predictions (right column), respectively. Results from ten-fold cross-validation over 10 stations and different lead-times (+24h to +48h), evaluated separately for different accumulation periods (24, 6 hour). Bin-width is 0.025.

$\nu = 0.02$ (Figure 6). Higher split levels, e.g. 0.5, illustrate no skill on median.

This result is equal for censored Gaussian and censored logistic split models using the log-link. Although an optimum split level could be found on split levels in between, we could not identify a significant benefit of using other split levels for validation data (results not shown).

4.3. Heavy tails

Calibration is one of the most important characteristics which should be achieved by probabilistic forecasts. We therefore compute the probability integral transform (PIT), which is similar to rank histograms (Hamill and Colucci 1998; Anderson 1995; Talagrand *et al.* 1997). It bins the forecasted cumulative probability density function and counts where the observed value falls into. If the model is well calibrated, the bins should all have the same number of observations.

Figure 7 shows PIT histograms for different accumulation periods. For simplicity, only splitted log-link models are shown, since they performed best in terms of CRPSS values. The remaining models generate very similar histograms (results not shown). Both, Gaussian and logistic models are more calibrated if short accumulation periods are forecasted. Logistic models generally produce histograms that are more uniformly distributed.

This improvement by the logistic tail is also visible for PoP forecasts, where we additionally compute the BS decomposition based on Murphy and Winkler (1987). The BS and its probabilistic attributes of reliability, resolution, and sharpness are illustrated in Figure 8. Brier scores are very similar among the models and decrease for short observation intervals in general. This is related to the number of zeros, which increases for shorter accumulation

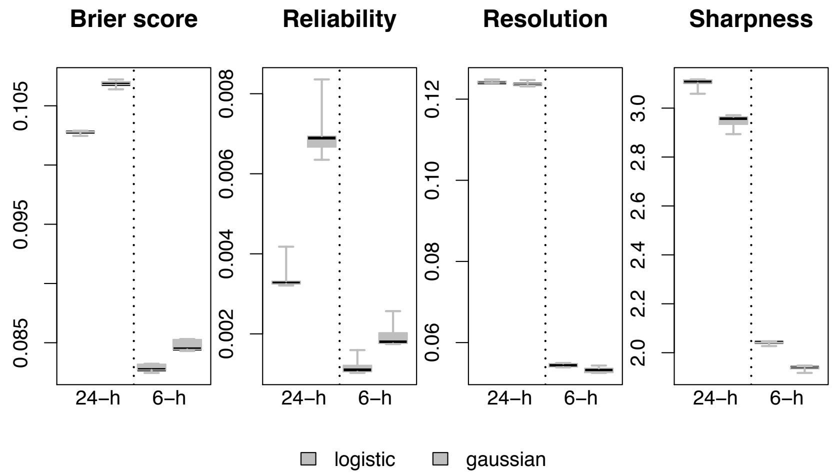

Figure 8: Brier score and its decomposition into reliability, resolution, and sharpness for probability of precipitation (PoP): Results from ten-fold cross-validation over 10 stations and different lead-times (+24h to +48h), evaluated separately for different accumulation periods. Each boxplot illustrate the 5 models (Table 1), both for censored Gaussian (right boxplot) and logistic models (left boxplot) respectively.

periods. Clearly, logistic models obtain the smallest scores for all periods. The decomposition also clarifies a smaller reliability, slightly larger resolution, and larger sharpness of censored logistic forecasts. An additional overview about the numerical values of this decomposition for each individual model can be found in the appendix (Table 3).

4.4. Link functions

So far we have seen an improved skill of split models and by using the logistic tail. CRPS differences in the split models (Figure 5) might be understood by looking at the regression fits for different link-functions. Figure 9 gives an example fit for censored Gaussian models for cases where the raw standard deviation was larger than 0. Although the linear fits for the latent mean value (left graphic) do not vary a lot, the fit for the scale parameter (right graphic) highlights larger differences. If the ensemble would already be perfect, the fitted curves would follow the dashed black line. Since this is not the case, ensemble mean values are corrected to lower values (fits below the black line) and ensemble standard deviations to higher values (fits above the black line).

Furthermore, differences in the predicted scale parameter can be seen for small values of the ensemble standard deviation (e.g., $SD_{ens} = 0.5$), where the log-link predictions are largest. This pattern was found for other cases as well (results not shown).

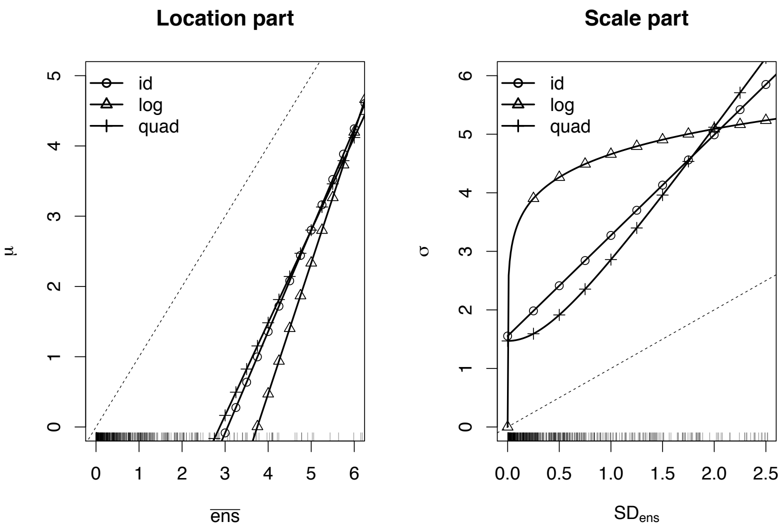

Figure 9: Link-functions for censored Gaussian models showing location- and scale predictions at Bolzano, 6-hour sum for lead-time +36h: identity-link (circle), log-link (triangle), quadratic-link (cross); x-axis denotes the ensemble mean $\overline{ens}$ for location-models and the ensemble standard deviation ens_SD for scale-models. The rug bars on the x-axis illustrate the raw ensemble values used for fitting. Y-axis describe the predictions of latent mean and standard deviation.

5. Summary and conclusion

In this study we have refined the non-homogeneous regression (NHR) approach (Gneiting *et al.* 2005) for precipitation forecast and applied it to a study area in the European Alps. A combination of three statistical tricks is able to improve post-processed probabilistic forecasts for precipitation amounts, and the probability of precipitation: a new split approach accounting for perfect cases in the raw ensemble, the censored logistic distribution for heavy tails of precipitation data, and a suitable link-function for uncertainty predictions.

Our newly introduced split approach is able to account for perfect ensemble forecasts that can occur. Especially for non-precipitation events, the raw ensemble was found to be perfect if all members are zero. The estimation of a separate climatological mean for cases where a certain fraction of zeros occur could clearly outperform the reference method, that does not make use of the information of zeros. Additionally, the general expression of the split approach allows us to use other split levels, which were found to be worse for the lowest split-level (0.1) and skillful for higher levels (larger than 0.5). It is also possible to use the split approach for different definitions of *frac*, which we have defined as the number of members being zero. Contrary, *frac* can be also the number of members below a certain value (e.g. 0.001) to overcome the problem of noise in the bilinear interpolation of ensemble data for instance.

Furthermore, using the censored logistic distribution increased the forecast skill compared to censored Gaussian models. The pronounced tail of the logistic distribution was able to capture rare events, and produced improved CRPS and BS values, as well as calibration verified in PIT histograms. Clearly, distributions with skewness or an even more pronounced tail, e.g. Gamma or GEV, represent another good distribution assumption as well (Wilks 1990; Scheuerer 2014; Scheuerer and Hamill 2015), since a peak of very extreme events was still visible in the PIT evaluation.

Our third refinement has been the investigation of different link-functions for the dispersion sub-model in the NHR-approach. Depending on the used forecast distribution, distribution parameters can require positive values during numerical optimization. We could find notable differences in forecast skill for different link-functions, especially for short accumulation periods of 6-hour sums. The best performance has been achieved by usage of the log-link (*log*) for the scale sub-models, which optimizes the logarithmic standard deviation (Messner *et al.* 2014a). This approach outperformed simulations where the squared scale (*quad*) or scale parameter (*id*) is estimated and is an attractive candidate for this optimization issue. Differences are visible if the estimated scale parameters are back-transformed into the scale of the original standard deviation. Although all link-functions could correct for the raw ensembles underdispersion, *id*- and *quad*-models produced too less uncertainty. This especially takes place in the range for smaller regressor values of ensemble standard deviation, where most of the cases occurred. Additionally, the combination of log-link and the split approach allows us to use the logarithmic standard deviation as regressor. Otherwise, the logarithmic standard deviation could not be used (due to infinities occurring if there is no standard deviation in the ensemble).

To summarize the overall forecast performance for our study area, all statistical models could improve the raw ensemble forecasts by 25% on median, in terms of CRPSS. Our results imply that an untransformed censored logistic assumption is adequate for short accumulation periods and in predicting the probability of precipitation. Our proposed split models highlight the importance of using the information of zeros predicted by the raw ensemble. The differences in daily sums seemed to be negligible small. Results also showed differences in link-functions where the logarithmic-link performed best. Our three statistical tricks illustrate the largest benefit for short accumulation periods (6h).

6. Acknowledgments

This work is an ongoing PhD - project and funded by Autonome Provinz Bozen - Abteilung Bildungsförderung, Universität und Forschung (Nr. ORBZ110725). We thank the Austrian weather Service (ZAMG) for their supply of ECMWF EPS data, and the weather service of South Tyrol for their observational data. Many thanks also to Reto Stauffer, who has done a lot of data handling.

Appendix

	Gaussian models					**Logistic models**				
	quad	quad_frac	quad_split	log_split	id_split	quad	quad_frac	quad_split	log_split	id_split
24-REL	0.00694	0.00635	0.0069	0.00836	0.00666	0.00325	0.00321	0.00327	0.00418	0.00332
6-REL	0.00176	0.00175	0.0018	0.00257	0.00203	0.0011	0.00102	0.00105	0.00159	0.00122
24-RES	0.12348	0.12314	0.12367	0.12471	0.12385	0.12406	0.12401	0.12402	0.12486	0.12444
6-RES	0.05253	0.05335	0.05259	0.05434	0.05367	0.05398	0.05438	0.0539	0.05497	0.05484
24-SHARP	2.97099	2.93331	2.96455	2.89358	2.95825	3.11742	3.10234	3.11576	3.05869	3.10605
6-SHARP	1.94735	1.94153	1.93967	1.9173	1.9342	2.04687	2.04229	2.04243	2.02699	2.03926

Table 3: Brier score decomposition for probability of precipitation as illustrated in Figure 8. Rows show values for reliability (REL), resolution (RES), and sharpness (SHARP) for different accumulation periods (24, 6 hour) for censored Gaussian and logistic models as described in Table 1.

References

Anderson JL (1995). "A method for producing and evaluating probabilistic forecast from ensemble model integration." *Journal of Climate*, **9**, 1518–1530. ISSN 0894-8755. `doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2`.

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

Bentzien S, Friederichs P (2012). "Generating and calibrating probabilistic quantitative precipitation forecasts from the high-resolution NWP model COSMO-DE." *Weather and Forecasting*, **27**, 988–1002. `doi:10.1175/WAF-D-11-00101.1`.

Brier GW (1950). "Verification of forecasts expressed in terms of probability." *Monthly Weather Review*, **78**, 1–3. ISSN 0036-8075. `doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2`.

Carisse O, Bacon R, Lefebvre A, Lessard K (2009). "A degree-day model to initiate fungicide spray programs for management of grape powdery mildew [Erysiphe necator]." *Plant Pathology*, **31**, 186–194.

Cohen AJ (1959). "Simplified estimators for the normal distribution when samples are singly censored or truncated." *Technometrics*, **1**, 217–237. `doi:10.2307/1266442`.

Gneiting T, Katzfuss M (2014). "Probabilistic forecasting." *Annual Review of Statistics and Its Application*, **1**, 125–151. ISSN 2326-8298. `doi:10.1146/annurev-statistics-062713-085831`.

Gneiting T, Raftery AE, Westveld III 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.

Hamill TM, Colucci SJ (1998). "Evaluation of Eta – RSM ensemble probabilistic precipitation forecasts." *Monthly Weather Review*, **126**, 711–724.

Hersbach H (2000). "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting*, **15**, 559–570.

Leith C (1974). "Theoretical skill of Monte Carlo forecasts." *Monthly Weather Review*, **102**, 409–418.

Löpmeier F, Wittich K, Frühauf C, Schittenhelm S (2013). "Entwicklungen und Stand der Aktivitäten in der Agrarmeteorologie." *promet*, **38**(1/2), 2–10.

Messner JW, Mayr GJ, Wilks DS, Zeileis A (2014a). "Extending extended logistic regression: Extended versus separate versus ordered versus censored." *Monthly Weather Review*, **142**, 3003–3014. ISSN 0027-0644. doi:10.1175/MWR-D-13-00355.1.

Messner JW, Mayr GJ, Zeileis A (2016). "Heteroscedastic Censored and Truncated Regression with crch." *The R-Journal*, **17**(11), in press. Accessed October 5 2016, URL http://journal.r-project.org/archive/2016-1/messner-mayr-zeileis.pdf.

Messner JW, Mayr GJ, Zeileis A, Wilks DS (2014b). "Heteroscedastic extended logistic regression for postprocessing of ensemble guidance." *Monthly Weather Review*, **142**, 448–456. ISSN 0027-0644. doi:10.1175/MWR-D-13-00271.1.

Mullen SL, Buizza R (2002). "The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system." *Weather and Forecasting*, **17**, 173–191.

Murphy A, Winkler R (1987). "A general framework for forecast verification." *Monthly Weather Review*, **115**, 1130–1338.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Accessed October 5 2016, URL http://www.R-project.org/.

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.

Scheuerer M (2014). "Probabilistic quantitative precipitation forecasting using ensemble model output statistics." *Quarterly Journal of the Royal Meteorological Society*, **140**, 1086–1096. ISSN 00359009. doi:10.1002/qj.2183.

Scheuerer M, Hamill TM (2015). "Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions." *Monthly Weather Review*, **143**, 4578–4596. doi:10.1175/MWR-D-15-0061.1.

Sloughter JM, Raftery AE, Gneiting T, Fraley C (2007). "Probabilistic quantitative precipitation forecasting using Bayesian model averaging." *Monthly Weather Review*, **135**, 3209–3220. ISSN 0027-0644. doi:10.1175/MWR3441.1.

Smith RB, Jiang Q, Fearon MG, Tabary P, Dorninger M, Doyle JD, Benoit R (2003). "Orographic precipitation and air mass transformation: An Alpine example." *Quarterly Journal of Meteorological Society*, **129**, 433–454. doi:10.1256/qj.01.212.

Talagrand O, Vautard R, Strauss B (1997). "Evaluation of probabilistic prediction systems." In *Proceeding of workshop on predictability, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK*, pp. 1–25.

Thorarinsdottir TL, Gneiting T (2010). "Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **173**, 371–388. doi:10.1111/j.1467-985X.2009.00616.x.

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 (1990). "Maximum likelihood estimation for the gamma distribution using data containing zeros." *Journal of Climate*, **3**, 1495–1501.

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

Affiliation:

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

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

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

http://eeecon.uibk.ac.at/wopec/

2016-28 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

2016-27 **Michael Razen, Matthias Stefan:** Greed: Taking a deadly sin to the lab

2016-26 **Florian Wickelmaier, Achim Zeileis:** Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

2016-25 **Michel Philipp, Carolin Strobl, Jimmy de la Torre, Achim Zeileis:** On the estimation of standard errors in cognitive diagnosis models

2016-24 **Florian Lindner, Julia Rose:** No need for more time: Intertemporal allocation decisions under time pressure

2016-23 **Christoph Eder, Martin Halla:** The long-lasting shadow of the allied occupation of Austria on its spatial equilibrium

2016-22 **Christoph Eder:** Missing men: World War II casualties and structural change

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *forthcoming in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

2015-16 **Niall Flynn, Christopher Kah, Rudolf Kerschbamer:** Vickrey Auction vs BDM: Difference in bidding behaviour and the impact of other-regarding motives

2015-15 **Christopher Kah, Markus Walzl:** Stochastic stability in a learning dynamic with best response to noisy play

2015-14 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** Measuring regional innovation in one dimension: More lost than gained?

2015-13 **Christoph Hauser, Gottfried Tappeiner, Janette Walde:** The roots of regional trust

2015-12 **Christoph Hauser:** Effects of employee social capital on wage satisfaction, job satisfaction and organizational commitment

2015-11 **Thomas Stöckl:** Dishonest or professional behavior? Can we tell? A comment on: Cohn et al. 2014, Nature 516, 86-89, "Business culture and dishonesty in the banking industry"

2015-10 **Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, Henk Kelderman:** Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees

2015-09 **Martin Halla, Gerald Pruckner, Thomas Schober:** The cost-effectiveness of developmental screenings: Evidence from a nationwide programme *forthcoming in Journal of Health Economics*

2015-08 **Lorenz B. Fischer, Michael Pfaffermayr:** The more the merrier? Migration and convergence among European regions

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city *forthcoming in European Economic Review*

2015-06 **Martin Geiger, Wolfgang Luhan, Johann Scharler:** When do Fiscal Consolidations Lead to Consumption Booms? Lessons from a Laboratory Experiment *forthcoming in Journal of Economic Dynamics and Control*

2015-05 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket payments in the Austrian healthcare system - a distributional analysis

2015-04 **Rudolf Kerschbamer, Matthias Sutter, Uwe Dulleck:** How social preferences shape incentives in (experimental) markets for credence goods *forthcoming in Economic Journal*

2015-03 **Kenneth Harttgen, Stefan Lang, Judith Santer:** Multilevel modelling of child mortality in Africa

2015-02 **Helene Roth, Stefan Lang, Helga Wagner:** Random intercept selection in structured additive regression models

2015-01 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures for pharmaceuticals: Lessons from the Austrian household budget survey

University of Innsbruck

Working Papers in Economics and Statistics

2016-28

Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis

Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

Abstract
Raw ensemble forecasts display large errors in predicting precipitation amounts and its forecast uncertainty, especially in mountainous regions where local effects are often not captured. Therefore, statistical post-processing is typically applied to obtain automatically corrected weather forecasts where precipitation represents one of the most challenging quantities. This study applies the non-homogenous regression framework as a start-of-the-art ensemble post-processing technique to predict a full forecast distribution and improves its forecast performance with three statistical tricks. First of all, a novel split-type approach effectively accounts for perfect ensemble predictions that can occur. Additionally, the statistical model assumes a censored logistic distribution to deal with the heavy tails of precipitation amounts. Finally, the optimization of regression coefficients for the scale parameter is investigated with suitable link-functions. These three refinements are tested for stations in the European Alps for lead-times from +24h to +48h and accumulation periods of 24 and 6 hours. Results highlight an improvement due to a combination of the three statistical tricks against the default post-processing method which does not account for perfect ensemble predictions. Probabilistic forecasts for precipitation amounts as well as the probability of precipitation events could be improved, especially for 6 hour sums.

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