

Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance

JAKOB W. MESSNER AND GEORG J. MAYR

Institute of Meteorology and Geophysics, University of Innsbruck, Innsbruck, Austria

ACHIM ZEILEIS

Department of Statistics, University of Innsbruck, Innsbruck, Austria

DANIEL S. WILKS

Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York

(Manuscript received 21 August 2013, in final form 4 October 2013)

ABSTRACT

To achieve well-calibrated probabilistic forecasts, ensemble forecasts are often statistically postprocessed. One recent ensemble-calibration method is extended logistic regression, which extends the popular logistic regression to yield full probability distribution forecasts. Although the purpose of this method is to postprocess ensemble forecasts, usually only the ensemble mean is used as the predictor variable, whereas the ensemble spread is neglected because it does not improve the forecasts. In this study it is shown that when simply used as an ordinary predictor variable in extended logistic regression, the ensemble spread affects the location but not the variance of the predictive distribution. Uncertainty information contained in the ensemble spread is therefore not utilized appropriately. To solve this drawback a new approach is proposed where the ensemble spread is directly used to predict the dispersion of the predictive distribution. With wind speed data and ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) it is shown that by using this approach, the ensemble spread can be used effectively to improve forecasts from extended logistic regression.

1. Introduction

Weather forecasts are very important for many parts of social and economic life. For example, they are used for severe weather warnings, for decision making in agriculture and industry, or for planning of leisure activities. Generally these forecasts are based on numerical weather prediction (NWP) models. Unfortunately, because of uncertainties in the initial conditions and unknown or unresolved atmospheric processes these models are always subject to error. Luckily some of these errors are systematic and can be corrected with statistical postprocessing, often also referred to as model output

statistics (MOS; Glahn and Lowry 1972). However, not all errors can be corrected and for many customers it is important to get additional information about the remaining forecast uncertainty. For this purpose many forecasting centers provide ensemble forecasts. These are multiple NWP forecasts with slightly perturbed initial conditions and sometimes also different model formulations. The idea is that these different forecasts should represent the range of possible outcomes (Lorenz 1996). Large ensemble spreads are then presumably associated with high forecast uncertainties and small spreads that signify low uncertainties. However, in practice the initial ensemble members do not represent initial-condition uncertainty (Hamill et al. 2003; Wang and Bishop 2003). Furthermore ensemble forecasts exhibit the same model errors as single integration forecasts. Thus, to achieve unbiased and calibrated uncertainty forecasts, statistical postprocessing is needed.

In the past decade much research has gone into finding appropriate methods to postprocess ensemble forecasts.

 Denotes Open Access content.

Corresponding author address: Jakob W. Messner, Institute of Meteorology and Geophysics, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria.
E-mail: jakob.messner@uibk.ac.at

DOI: 10.1175/MWR-D-13-00271.1

For example, Roulston and Smith (2003) proposed dressing the ensemble members with historical model errors and Raftery et al. (2005) suggested Bayesian model averaging for this purpose. Gneiting et al. (2005) proposed to use linear regression with error variances depending on the ensemble spread, and for binary predictands Hamill et al. (2004) proposed to use logistic regression. Comparisons of these and other methods (Wilks 2006a; Wilks and Hamill 2007) showed that logistic regression is one of the better approaches. A very promising extension of logistic regression has been proposed recently (Wilks 2009). By including the predictand threshold in the regression equations this extended logistic regression allows derivation of full predictive distributions. The extended logistic regression method has been used in several studies for probabilistic precipitation forecasts (Schmeits and Kok 2010; Ruiz and Saulo 2012; Roulin and Vannitsem 2012; Hamill 2012; Ben Bouallègue 2013; Scheuerer 2013) and was shown to perform very well compared to standard logistic regression (Wilks 2009; Ruiz and Saulo 2012) and other ensemble postprocessing methods (Schmeits and Kok 2010; Ruiz and Saulo 2012; Scheuerer 2013). In all of these studies, extended logistic regression is used to post-process ensemble forecasts, but usually the ensemble mean was used as the only predictor variable. There were also several attempts to additionally include the ensemble spread, but with the exception of Hamill (2012) it was always disregarded because it did not improve the forecasts.

In this study we show that the predictive distribution of the transformed predictand is logistic and that the predictor variables only affect the location (mean) but not the dispersion (variance) of this logistic distribution. So far the ensemble spread was always included as ordinary predictor variable in extended logistic regression so that its information was only used to predict the location but not the dispersion of the forecast distribution. However, the ensemble spread is generally expected to mainly contain information about the forecast uncertainty, which in turn should be directly related to the dispersion of the predictive distribution. Hence, the uncertainty information contained in the ensemble spread cannot be utilized properly by extended logistic regression, so that it is not surprising that no improvements could be found.

To solve this drawback of extended logistic regression, we therefore propose a simple new approach in which the ensemble spread can be directly used as predictor for the dispersion of the forecast probability distribution. To illustrate our findings and test if improvements can be achieved with this new approach, we compare different approaches to include the ensemble

spread in extended logistic regression on wind speed data from 11 European locations and ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF).

The remainder of the paper is organized as follows: in section 2 we describe the extended logistic regression model and show the problems when including the ensemble spread as ordinary predictor variable. Our new approach is introduced in section 3. Results from the case study are shown in section 4 and a summary and conclusions can be found in section 5.

2. Extended logistic regression

Originally, logistic regression is a regression model from the generalized linear model framework (Nelder and Wedderburn 1972) to model the conditional probability of binary events. As such it is also a well-suited MOS method for binary predictands (Hamill et al. 2004). For example, the probability of a continuous variable y to fall below a certain threshold q can be predicted with

$$P(y < q | \mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T \boldsymbol{\beta})} = \Lambda(\mathbf{x}^T \boldsymbol{\beta}), \quad (1)$$

where \mathbf{x} is a vector of predictor variables [e.g., NWP forecasts; $\mathbf{x} = (1, x_1, x_2, \dots)^T$] and $\boldsymbol{\beta}$ is a vector of regression coefficients [$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots)^T$] that is generally estimated with maximum likelihood estimation (see appendix A). The regression function has the same mathematical form as the cumulative distribution function of the standard logistic distribution Λ , which is indicated by the final equality in Eq. (1).

Often, more than one threshold is of interest and separate logistic regressions are fitted for each of these thresholds. This approach has the disadvantage that the predicted probabilities are not constrained to be mutually consistent. In other words, for two thresholds q_a and q_b with $q_a < q_b$ it can occur that $P(y < q_a | \mathbf{x}) > P(y < q_b | \mathbf{x})$, which would imply nonsense negative probabilities for $P(q_a \leq y < q_b | \mathbf{x})$.

To avoid these inconsistencies, Wilks (2009) extended logistic regression by including (a transformation of) the thresholds q_j as additional predictor variable:

$$P(y < q_j | \mathbf{x}) = \Lambda[\alpha g(q_j) + \mathbf{x}^T \boldsymbol{\beta}]. \quad (2)$$

Here $g(q_j)$ is a nondecreasing function of q_j and α is an additional coefficient that has to be estimated. In addition to avoiding negative probabilities, this extended logistic regression has the advantage that fewer coefficients have to be estimated (instead of different vectors $\boldsymbol{\beta}$ for each threshold, α and $\boldsymbol{\beta}$ are the same for

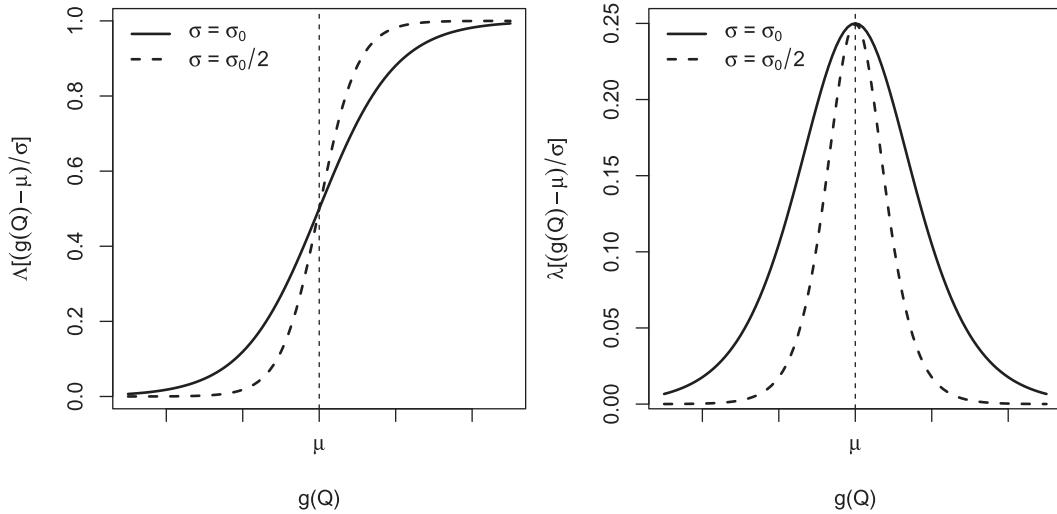


FIG. 1. (left) Cumulative distribution function and (right) probability density function of the logistic distribution for different scale parameters σ . Here $\lambda(x) = d\Lambda(x)/dx$ is the probability density function of the standard logistic distribution.

all thresholds), which is especially advantageous for small training datasets (Wilks 2009). Furthermore, the probability to fall below any arbitrary value Q can be easily computed by replacing q_j with Q :

$$P(y < Q | \mathbf{x}) = \Lambda[\alpha g(Q) + \mathbf{x}^T \boldsymbol{\beta}]. \quad (3)$$

Equation (3) can also be interpreted as continuous cumulative distribution function, which implies that full continuous probability distributions can be provided.

Since $g(\cdot)$ has to be a nondecreasing function, the equation

$$P(y < Q | \mathbf{x}) = P[g(y) < g(Q) | \mathbf{x}] \quad (4)$$

is always fulfilled. With Eq. (4) and some rearrangements, Eq. (3) can also be written as

$$P[g(y) < g(Q) | \mathbf{x}] = \Lambda \left[\frac{g(Q) + \mathbf{x}^T \boldsymbol{\beta} / \alpha}{1/\alpha} \right] \quad (5)$$

and upon setting $\mu = -\mathbf{x}^T \boldsymbol{\beta} / \alpha$ and $\sigma = 1/\alpha$ we obtain the following:

$$P[g(y) < g(Q) | \mathbf{x}] = \Lambda \left[\frac{g(Q) - \mu}{\sigma} \right]. \quad (6)$$

This notation allows one to easily see that the conditional probability distribution of the transformed predictand $g(y)$ given the predictor variables \mathbf{x} is a logistic distribution with location parameter μ and scale parameter σ . Cumulative distribution functions and probability density functions of this distribution with different

scale parameters σ are shown in Fig. 1. The shape of the logistic distribution is very similar to that of the normal distribution but with somewhat heavier tails. The mean of this distribution is μ and in terms of the scale parameter the variance is $\sigma^2 \pi^2/3$ (Johnson et al. 1995).

Note that the scale parameter $\sigma = 1/\alpha$ is constant so that the predictor variables in \mathbf{x} only affect the mean, not the variance of the logistic predictive distribution. Hence, when included as additional predictor variable in \mathbf{x} , the ensemble spread has no effect on the dispersion of the predictive distribution. However, usually large ensemble spreads are associated with high forecast uncertainties, which in turn should be related to wider predictive distributions. In contrast the level of uncertainty should generally have no effect on the location of the forecast probability distribution.

3. Heteroscedastic extended logistic regression

In the previous section we showed that when using the ensemble spread as an ordinary predictor variable in extended logistic regression, uncertainty information is not utilized appropriately. As a more effective approach we therefore propose to use the ensemble spread directly as predictor for the *dispersion* of the predictive distribution. Therefore, we simply replace μ and σ in Eq. (6) with

$$\mu = \mathbf{x}^T \boldsymbol{\gamma} \quad (7)$$

and

$$\sigma = \exp(\mathbf{z}^T \boldsymbol{\delta}), \quad (8)$$

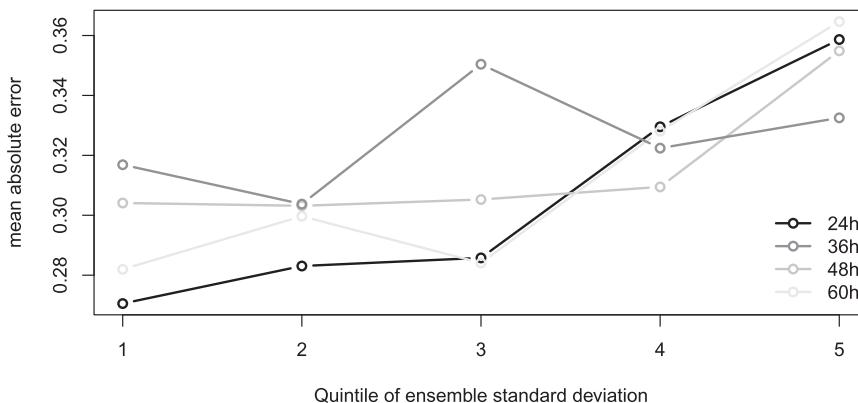


FIG. 2. Mean absolute error of ensemble median for different ensemble standard deviations and lead times computed for Wien-Hohe-Warte (33 months of data). Quintiles are used to divide the ensemble standard deviation into different levels. Note that for this plot all wind speeds are square root transformed.

respectively. Here \mathbf{z} is an additional vector of input variables (i.e., the ensemble spread) and $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are coefficient vectors that have to be estimated. The exponential function is used here as a simple method to ensure positive values for σ .

Note that with $\mathbf{z} = 1$ this model is completely equivalent to the original extended logistic regression [Eq. (2)] with $\alpha = 1/\exp(\boldsymbol{\delta})$ and $\boldsymbol{\beta} = -\boldsymbol{\gamma}/\exp(\boldsymbol{\delta})$.

The idea of using the ensemble spread as predictor for the dispersion is not completely new. For Gaussian linear regression models, Gneiting et al. (2005) proposed a similar approach, which has been proven to perform well in several studies (e.g., Wilks 2006a; Wilks and Hamill 2007).

4. Case study

In this section, we apply the findings from the previous sections on real data. We use 10-m wind speed observations (mean over last 10 min) from the following 11 European weather stations: Amsterdam Airport Schiphol in Amsterdam, Netherlands (52.3°N, 4.783°E); Berlin Tegel Airport in Berlin, Germany (52.55°N, 13.3°E); National Airport in Brussels, Belgium (50.9°N, 4.533°E); Copenhagen Airport in Copenhagen, Denmark (55.6°N, 12.633°E); Frankfurt Main in Frankfurt, Germany (50.033°N, 8.583°E); Heathrow in London, United Kingdom (51.467°N, -0.45°E); Geof in Lisbon, Portugal (38.767°N, -9.133°E); Barajas in Madrid, Spain (40.467°N, -3.55°E); Orly in Paris, France (48.717°N, 2.383°E); Fiumicino in Rome, Italy (41.8°N, 12.233°E); and Wien-Hohe-Warte in Vienna, Austria (48.249°N, 16.356°E), from April 2010 to December 2012. As NWP forecasts we use ensemble wind speed forecasts bilinearly interpolated to the instrument location from the ECMWF

(Molteni et al. 1996), initialized at 0000 UTC for the lead times 24, 36, 48, and 60 h.

Figure 2 shows a clear positive correlation between ensemble spread and forecast error for Wien-Hohe-Warte (similar for most other locations). This positive spread–skill relationship suggests that the ensemble spread contains potentially useful uncertainty information. To investigate how this information might be used most effectively, we compare different extended logistic regression models.

For all models we use the square root function for $g(\cdot) = \sqrt{\cdot}$. This function gave good results for precipitation forecasts in several studies (Wilks 2009; Schmeits and Kok 2010; Roulin and Vannitsem 2012; Ruiz and Saulo 2012; Ben Bouallègue 2013; Scheuerer 2013) and also improves our wind speed forecasts compared to the identity function. As potential regressors we use the ensemble mean M and standard deviation S of the square root–transformed ensemble wind speed forecasts. Furthermore, we selected $J = 9$ climatological quantiles with probabilities $1/10, 2/10, \dots, 9/10$ as thresholds q_j for each location separately.

Table 1 lists the models that are used in the following. In addition to the extended logistic regression model with the ensemble mean as single predictor variable (XLR) there are four models that use the ensemble standard deviation. The models XLR:S and XLR:SM are standard extended logistic regression models with the ensemble standard deviation as additional predictor variable, either alone (XLR:S) or multiplied with the ensemble mean (XLR:SM). In the heteroscedastic extended logistic regression model (HXLR) the ensemble standard deviation is only included as predictor variable for the scale and in HXLR:S it is additionally also used as predictor variable for the location of the predictive distribution.

TABLE 1. List of different extended logistic regression models. The \mathbf{x} and \mathbf{z} are vectors of predictor variables for the location and scale of the predictive distribution, respectively. The M and S are the mean and standard deviation of square root transformed wind speed ensemble forecasts, respectively.

| Model | | \mathbf{x} | \mathbf{z} |
|--------|--|------------------------|--------------|
| XLR | Extended logistic regression | $(1, M)^T$ | 1 |
| XLR:S | Extended logistic regression | $(1, M, S)^T$ | 1 |
| XLR:SM | Extended logistic regression | $(1, M, S \times M)^T$ | 1 |
| HXL | Heteroscedastic extended logistic regression | $(1, M)^T$ | $(1, S)^T$ |
| HXL:S | Heteroscedastic extended logistic regression | $(1, M, S)^T$ | $(1, S)^T$ |

Before reporting the forecast quality of these different models it is interesting to investigate the effect of the ensemble spread on the predicted probability distributions. Figure 3 shows predicted probability density functions of the XLR:S and HXL models for different ensemble standard deviations. For the XLR:S model it can be seen that contrary to the desired effect, larger ensemble standard deviations are related to slightly sharper distributions. In contrast, the HXL model uses the ensemble standard deviation more appropriately and larger ensemble standard deviations are clearly related to wider distributions.

Next we compare the performance of the different models. Since extended logistic regression can provide multicategorical probabilistic forecasts, the ranked probability score (Epstein 1969; Wilks 2006b) is a well-suited measure of forecast quality:

$$\text{RPS} = \sum_{j=1}^J [P(y < q_j | \mathbf{x}) - I(y < q_j)]^2. \quad (9)$$

Here $J = 9$ is the number of thresholds and $I(\cdot) = 1$ if the argument in brackets is true and 0 if it is not. To get independent training and test datasets we estimate and verify the models with tenfold cross validation. With this cross validation we get one RPS value for each event in the dataset. From these individual RPS values, 250 estimates of the mean (RPS) are computed on 250 bootstrap samples. This is all done separately for each model, location, and lead time. Since we are mainly interested in improvements that can be achieved with the ensemble standard deviation we finally compute skill scores (RPSS) with the standard extended logistic regression model (XLR) as a reference:

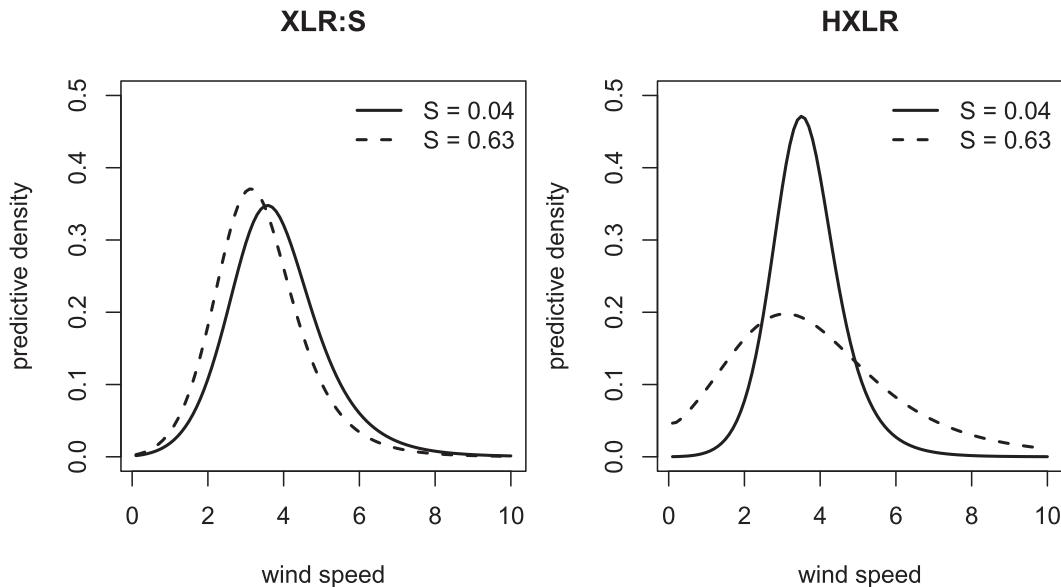


FIG. 3. Predicted probability density functions of (left) XLR:S and (right) HXL for small and large ensemble standard deviations, respectively. The models (see Table 1 for details) are fitted for Wien-Hohe-Warte and 36-h lead time. For all curves the ensemble mean is $M = 2$, which is approximately the mean ensemble mean of the dataset. The ensemble standard deviations 0.04 and 0.63 are approximately the minimum and maximum ensemble standard deviation in the dataset, respectively.

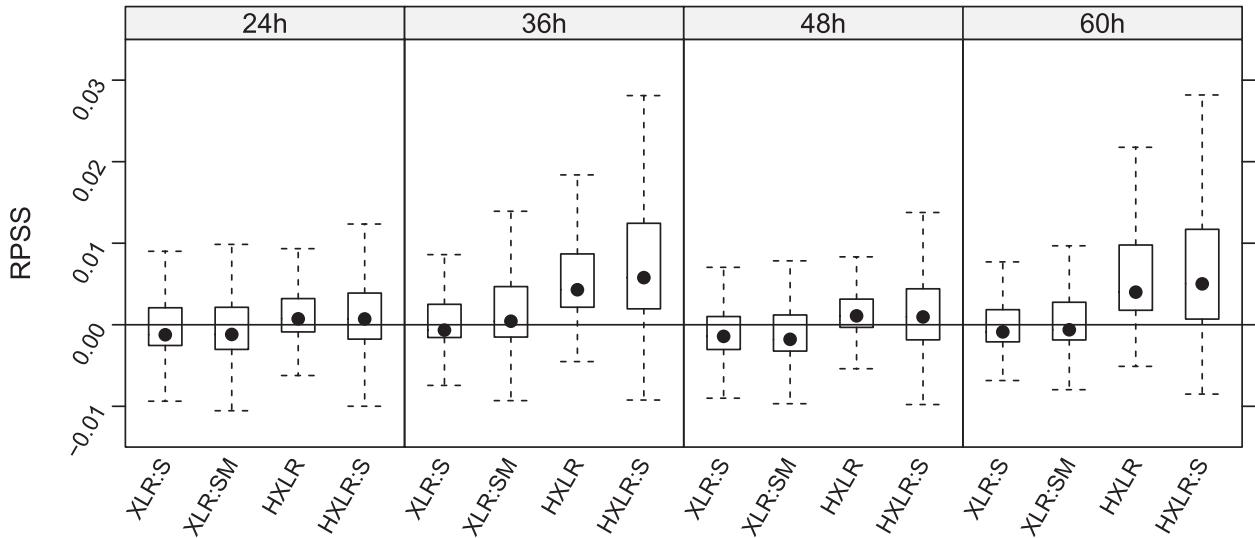


FIG. 4. Ranked probability skill score (RPSS) relative to extended logistic regression (XLR) for different lead times and models (see Table 1 for details) aggregated over 11 European locations (see text for details). Nine climatological deciles are used as thresholds. Positive values indicate improvements over XLR. The boxes indicate the interquartile ranges of the 11×250 values from the bootstrapping approach and the whiskers show the most extreme values that are less than 1.5 times the length of the box away from the box (farther outliers have been omitted).

$$RPSS = 1 - \frac{\overline{RPS}}{RPS_{XLR}}. \tag{10}$$

Note that here positive values indicate improvements over the standard extended logistic regression model.

Figure 4 shows the RPSS of the different models and lead times aggregated over the 11 locations. It can be seen that including the ensemble standard deviation simply as ordinary predictor variable (XLR:S, XLR:SM) does not improve forecast quality of extended logistic regression. However, the reason is not the absence of predictive information in the ensemble standard deviation since using it with our new approach (HXLR) clearly improves the forecast quality, especially for day time forecasts (36- and 60-h lead time). Since the ensemble standard deviation seems not to contain any predictive information on the location it is also not advantageous to include it additionally as predictor variable for the location (HXLR:S). The effect of the lead time on the RPSS is only weak but for daytime forecasts (12 and 36 h) the superiority of HXLR is more pronounced. Note that we also tested longer lead times (up to 96 h) and shorter training data lengths (down to 6 months), but results were similar and are therefore not shown.

Figure 5 shows the RPSS for selected locations aggregated over lead times 24–60 h. While most of the locations show similar patterns as in Fig. 4 (e.g., Amsterdam, Wien) there are also some locations (e.g., Berlin) where including the ensemble spread as ordinary predictor variable (XLR:S, XLR:SM) is superior to heteroscedastic

extended logistic regression (HXLR). This suggests that for these locations the ensemble spread also contains predictive information on the location. For nonnegative predictands like wind speed, large observed values are generally related to large ensemble spreads. Therefore, it is indeed conceivable that the ensemble spread contains some predictive information on the location that is not yet covered by the ensemble mean. However, additional improvements can be achieved when including the ensemble spread as a predictor for both location and scale of the predictive distribution (HXLR:S).

Finally, Fig. 6 shows reliability diagrams for 36-h forecasts of the first climatological decile [$P(y < q_1 | \mathbf{x})$] and the climatological median [$P(y < q_5 | \mathbf{x})$] for the models XLR:S and HXLR. Both models are fairly reliable with only little differences between each other. For the lower decile both models are slightly overforecasting (points below diagonal). The logistic predictive distribution of extended logistic regression involves a point mass at zero (i.e., positive predictive density for negative wind speeds; Schefzik et al. 2013). Because zero wind speeds occur relatively rarely, this might be the reason for the overestimated probabilities to fall below the lower decile.

5. Summary and conclusions

The inclusion of the ensemble spread in extended logistic regression has been shown in several studies not to improve the forecast skill. As we have shown in this paper this is not surprising because when the ensemble

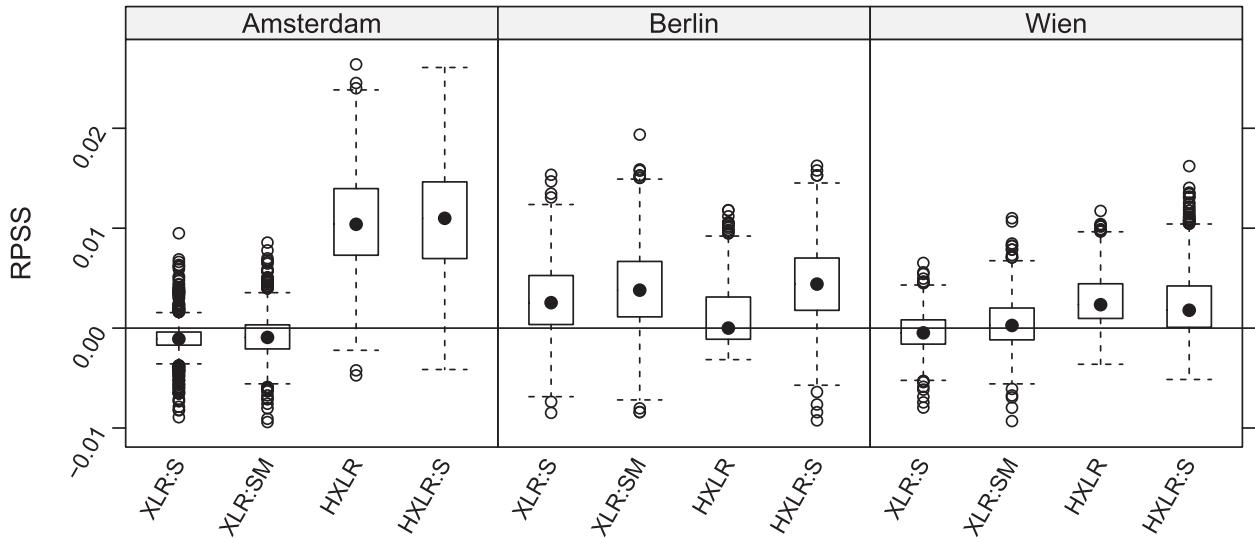


FIG. 5. Ranked probability skill score (RPSS) relative to extended logistic regression (XLR) for selected locations, aggregated over lead times 24, 36, 48, and 60 h. Nine climatological deciles are used as thresholds. The boxes indicate the interquartile ranges of the 4×250 values from the bootstrapping approach and the whiskers show the most extreme values that are less than 1.5 times the length of the box away from the box. Farther outliers are plotted as circles.

spread is included as an ordinary predictor variable it modifies only the location but not the dispersion of the forecast distribution. Uncertainty information contained in the ensemble spread is therefore not used appropriately. To solve this problem we proposed a new approach called heteroscedastic extended logistic regression where the ensemble spread is directly used as predictor for the *scale* of the predictive distribution.

To illustrate the advantages of this new approach we used wind speed observations from 11 European

locations and ensemble forecasts from ECMWF. Consistent with our findings and with results from previous studies, the inclusion of the ensemble standard deviation as an ordinary predictor variable has no clear positive effects on forecast quality. In contrast, with our new approach the uncertainty information in the ensemble standard deviation is used effectively to achieve clear improvements.

An additional single case study with precipitation data showed similar results. We therefore expect that our

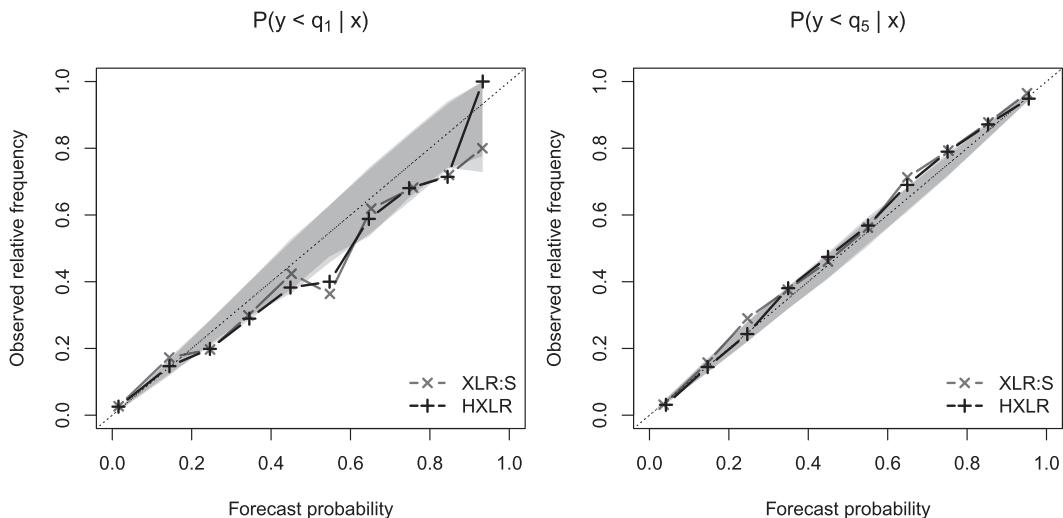


FIG. 6. Reliability diagrams for the 36-h probability forecasts (left) $P(y < q_1 | x)$ (first climatological decile) and (right) $P(y < q_5 | x)$ (climatological median) from XLR:S (gray) and HXLR (black). Forecasts are pooled for all locations and aggregated in 0.1 intervals. The gray areas show the 95% consistency intervals for XLR:S and HXLR (with alpha blending) derived from consistency resampling (Bröcker and Smith 2007).

results can be transferred to other variables and/or locations. However, this still has to be tested.

Hamill (2012) got better forecasts when using the ensemble variance multiplied with the ensemble mean as an additional predictor variable. This suggests that in his data, the ensemble spread also contained predictive information on the location of the predictive distribution. Consistent with these findings, we also found individual weather stations where including the ensemble spread as ordinary predictor variable is even superior to heteroscedastic extended logistic regression. However, further improvements could be achieved when including the ensemble spread as predictor variable for both location and spread of the predictive distribution.

To enhance the flexibility of extended logistic regression, Ben Bouallègue (2013) proposed the use of interaction terms between the threshold and the predictor variables. An interaction term between threshold and ensemble spread could also be used to control the dispersion of the predictive distribution. Contrary to heteroscedastic extended logistic regression such a model can be easily implemented with standard binary logistic regression software. However, with interaction terms the ensemble spread also has some undesired effects on the distribution location.

Extended logistic regression has been shown in several studies to perform well compared to other ensemble postprocessing algorithms (e.g., Schmeits and Kok 2010; Ruiz and Saulo 2012; Scheuerer 2013). However, a major drawback of this method was that uncertainty information contained in the ensemble spread could not be utilized effectively. Heteroscedastic extended logistic regression is therefore a very attractive extension of extended logistic regression to further enhance its competitiveness.

Acknowledgments. We thank Tom Hamill, Tilmann Gneiting, Constantin Junk, and an anonymous reviewer for their valuable comments that helped to improve this manuscript. This study was supported by the Austrian Science Fund (FWF): L615-N10. The first author was also supported by a Ph.D. scholarship from the University of Innsbruck, Vizerektorat für Forschung. Data from the ECMWF forecasting system were obtained from the ECMWF Data Server.

APPENDIX A

Likelihood Function

To estimate the coefficients α and β (extended logistic regression) or γ and δ (heteroscedastic extended logistic regression) maximum likelihood estimation is

used. The general log-likelihood function for logistic regression models is

$$l = \sum_{i=1}^N \log(\pi_i), \tag{A1}$$

where N is the length of the training dataset and π_i is the predicted probability for the i th observed outcome. For binary logistic regression there are two possible outcomes, so that

$$\pi_i = \begin{cases} P(y_i < q | \mathbf{x}_i) & y_i < q \\ 1 - P(y_i < q | \mathbf{x}_i) & y_i \geq q \end{cases}. \tag{A2}$$

In previous studies the sum of this binary log-likelihood over all thresholds is used as objective function that is maximized to estimate the regression coefficients. However, the predicted probability of the i th outcome actually is

$$\pi_i = \begin{cases} P(y_i < q_1 | \mathbf{x}_i) & y_i < q_1 \\ P(y_i < q_j | \mathbf{x}_i) - P(y_i < q_{j-1} | \mathbf{x}_i) & q_{j-1} \leq y_i < q_j, \\ 1 - P(y_i < q_M | \mathbf{x}_i) & y_i \geq q_M \end{cases} \tag{A3}$$

so that the correct maximum likelihood estimator is given by the maximization of Eqs. (A1) and (A3). In this study we employ this maximum likelihood estimator to take advantage of all standard asymptotic inference in the maximum likelihood framework. However, the concepts presented in this paper do not depend on the objective function and results should also not differ significantly when using the sum of binary log-likelihoods [Eq. (A2)] to estimate the coefficients.

APPENDIX B

Computational Details

Our results were obtained on Ubuntu using R 2.15.2 (R Core Team 2012). A function to fit (heteroscedastic) extended logistic regression models is included in the package crch 0.1-0 (Messner and Zeileis 2013).

REFERENCES

Ben Bouallègue, Z., 2013: Calibrated short-range ensemble precipitation forecasts using extended logistic regression with interaction terms. *Wea. Forecasting*, **28**, 515–524.
 Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Wea. Forecasting*, **22**, 651–661.
 Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987.

- Glahn, H., and D. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118.
- Hamill, T. M., 2012: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Mon. Wea. Rev.*, **140**, 2232–2252.
- , C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741–1758.
- , J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447.
- Johnson, N. L., S. Kotz, and N. Balakrishnan, 1995: *Continuous Univariate Distributions*. Vol. 2. Wiley, 752 pp.
- Lorenz, E., 1996: Predictability: A problem partly solved. *Proc. ECMWF Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.
- Messner, J. W., and A. Zeileis, cited 2013: crch: Censored Regression with Conditional Heteroscedasticity. R package version 0.1-0. [Available online at <http://CRAN.R-project.org/package=crch>.]
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroligis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Nelder, J., and R. Wedderburn, 1972: Generalized linear models. *J. Roy. Stat. Soc.*, **135A**, 370–384.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.
- R Core Team, cited 2012: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [Available online at <http://www.R-project.org/>.]
- Roulin, E., and S. Vannitsem, 2012: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, **140**, 874–888.
- Roulston, M. S., and L. A. Smith, 2003: Combining dynamical and statistical ensembles. *Tellus*, **55A**, 16–30.
- Ruiz, J. J., and C. Saulo, 2012: How sensitive are probabilistic precipitation forecasts to the choice of calibration algorithms and the ensemble generation method? Part I: Sensitivity to calibration methods. *Meteor. Appl.*, **19**, 302–313.
- Schefzik, R., T. Thorarindottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, in press.
- Scheuerer, M., 2013: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quart. J. Roy. Meteor. Soc.*, in press.
- Schmeits, M. J., and K. J. Kok, 2010: A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Mon. Wea. Rev.*, **138**, 4199–4211.
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.
- Wilks, D. S., 2006a: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteor. Appl.*, **13**, 243–256.
- , 2006b: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- , 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Appl.*, **368**, 361–368.
- , and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390.