Universität Innsbruck

Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance

Jakob W. Messner (jakob.messner@uibk.ac.at), Georg J. Mayr, Daniel S. Wilks, and Achim Zeileis

Motivation

Extended logistic regression:

- Popular extension of logistic regression to statistically calibrate numerical weather prediction ensembles.
- Previous studies: No improvements when using ensemble spread.

Question: No predictive information in ensemble spread?

Extended logistic regression (XLR)

Logistic regression:

Probabilistic forecast of binary events (e.g., wind speed y below threshold q_j).

$$P(y < q_j|\mathbf{x}) = \frac{\exp(\mathbf{x}^\top\beta)}{1+\exp(\mathbf{x}^\top\beta)} = \Lambda(\mathbf{x}^\top\beta) \quad (1)$$

$\mathbf{x}$: vector of predictor variables (e.g., ensemble mean M; $\mathbf{x} = (1, M)^\top$)

β: vector of regression coefficients ($\beta = (\beta_0, \beta_1)^\top$)

Λ: Cumulative distribution function of standard logistic distribution

Disadvantage:

- Several thresholds: Predicted probabilities not constrained to be mutually consistent (i.e., negative probabilities to fall between thresholds)

Extended logistic regression:

Idea (Wilks 2009): Include threshold q_j as additional predictor variable

$$P(y < q_j|\mathbf{x}) = \Lambda(\alpha q_j + \mathbf{x}^\top\beta) = \Lambda\left(\frac{q_j + \mathbf{x}^\top\beta/\alpha}{1/\alpha}\right) \quad (2)$$

α: additional regression coefficient

Advantages:

- Mutual consistency.
- Single equation for several thresholds $\rightarrow$ less coefficients.
- Replace q_j with arbitrary value $\rightarrow$ full continuous predictive distribution.

Previous studies:

No improvements with ensemble spread S as additional predictor variable (i.e., $\mathbf{x} = (1, M, S)^\top$).

Probable reason:

- Predictor variables only affect location but not the variance of predictive distribution (Figure 1).
- Ensemble spread: mainly uncertainty information $\rightarrow$ should affect variance.

XLR Effects

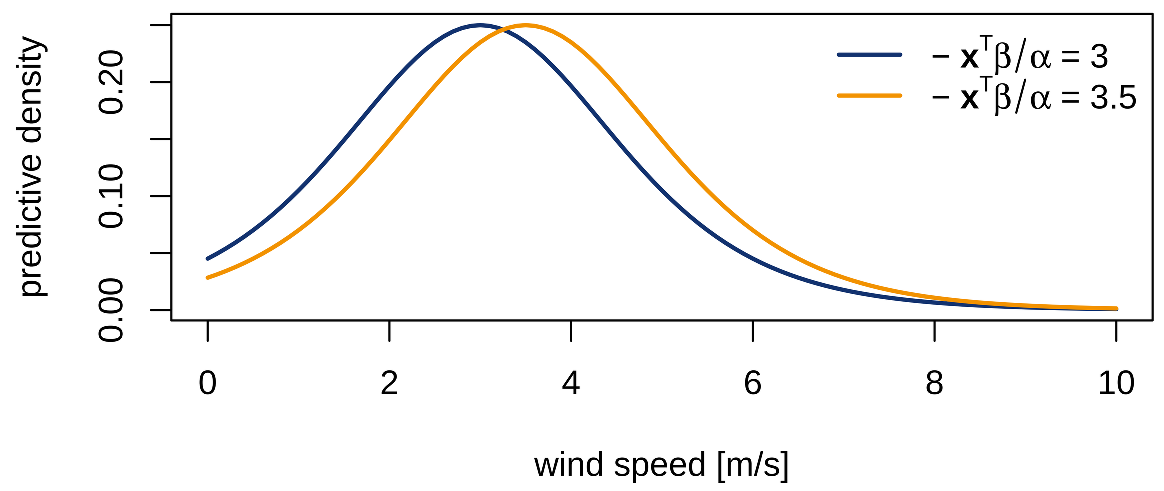

Figure 1 : Effect of predictor variables $\mathbf{x}$ (ensemble mean and/or standard deviation) on predictive distribution of XLR or XLR:S. E.g., blue: high ensemble spread, orange: low ensemble spread.

Note: Variance of predictive distribution determined by $1/\alpha$ (constant).

Heteroscedastic XLR (HXLR)

Idea:

Include ensemble spread as predictor variable for the variance of predictive distribution.

$$P(y < q_j|\mathbf{x}) = \Lambda\left(\frac{q_j - \mathbf{x}^\top\gamma}{\exp(\mathbf{z}^\top\delta)}\right) \quad (3)$$

$\mathbf{x} = (1, M)^\top$

$\mathbf{z} = (1, S)^\top$

M, S: ensemble mean and standard deviation

γ, δ: vectors of regression coefficients ($\gamma = -\beta/\alpha$)

exp: to ensure positive denominator

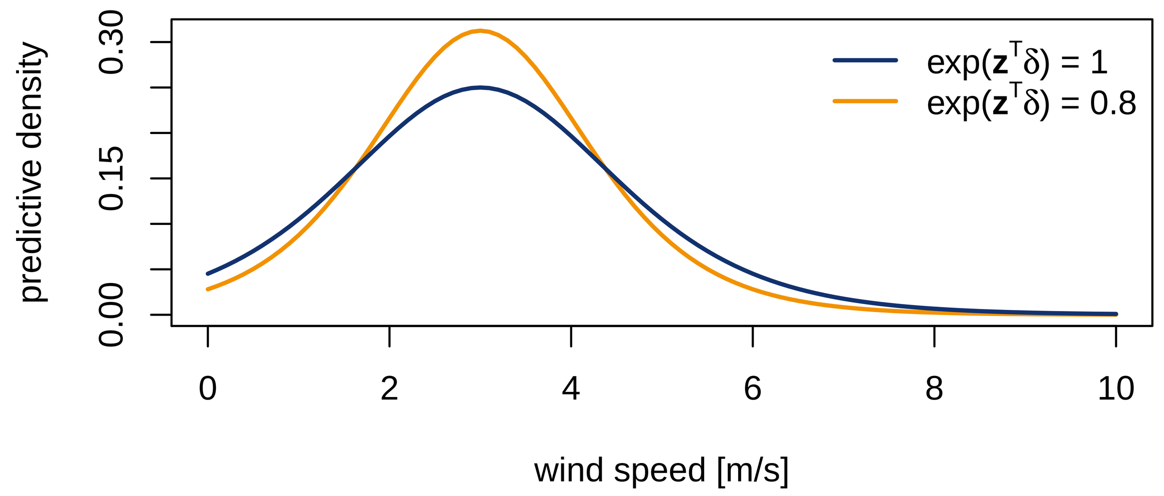

Figure 2 : Effect of predictor variables $\mathbf{z}$ on predictive distribution of HXLR. Blue: high ensemble spread, orange: low ensemble spread.

Case study

- 10 meter wind speed
- 11 European locations (*Amsterdam*, *Berlin*, *Brussels*, *Copenhagen*, *Frankfurt*, *London*, *Lisbon*, *Madrid*, *Paris*, *Rome*, *Wien*)
- April 2010–December 2012
- ECMWF ensemble forecasts

Models:

XLR: $\mathbf{x} = (1, M)^\top$

XLR:S: $\mathbf{x} = (1, M, S)^\top$

HXLR: $\mathbf{x} = (1, M)^\top, \mathbf{z} = (1, S)^\top$

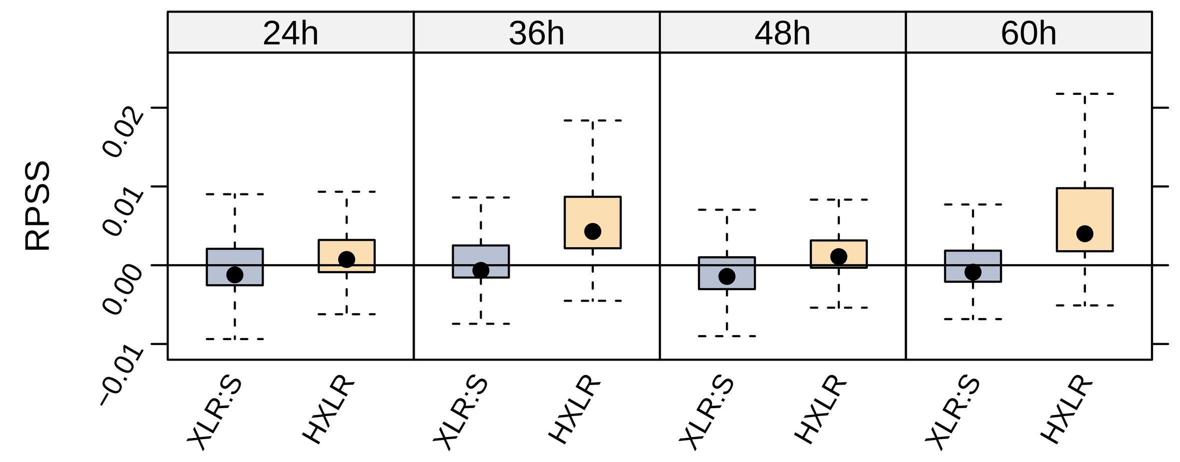

Figure 3 : Ranked probability skill score (*RPSS*) relative to extended logistic regression (*XLR*) aggregated over 11 European locations for lead times 24, 36, 48, and 60 hours. 9 climatological deciles as thresholds. Positive values indicate improvements over *XLR*.

Conclusion

Extended logistic regression:

Constant variance of predictive distribution

$\rightarrow$ ensemble spread information cannot be used effectively.

Heteroscedastic extended logistic regression:

Ensemble spread directly controls the variance of predictive distribution.

Case study:

Spread information of ECMWF ensemble effectively used with HXLR.

References:

Messner, J. W. and A. Zeileis, 2013: *crch: Censored Regression with Conditional Heteroscedasticity*. URL `http://CRAN.R-project.org/package=crch`, R package version 0.1-0.

Messner, J. W., A. Zeileis, G. J. Mayr, and D. S. Wilks, 2014: Heteroscedastic extended logistic regression for post-processing of ensemble guidance. *Monthly Weather Review*, **142**, 448–456.

Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, **368 (March)**, 361–368.