Correlation Effects in a Bivariate Gaussian Regression Model for Wind Vectors

Moritz N. Lang[1,2], Georg J. Mayr[2], Reto Stauffer[1], Achim Zeileis[1]

[1] Department of Statistics, Universität Innsbruck, Innsbruck, Austria
[2] Department of Atmospheric and Cryospheric Science, Universität Innsbruck, Innsbruck, Austria

E-mail for correspondence: `Moritz.Lang@uibk.ac.at`

Abstract: A new probabilistic post-processing method for wind vector forecasts in a distributional regression framework is presented employing a bivariate Gaussian distribution. In contrast to previous studies all parameters of the distribution are simultaneously modeled, namely the location and scale parameters for both surface wind components plus the correlation coefficient between them. Here flexible specifications of the correlation effects are explored in a distributional regression setup, using numerical weather predictions of wind direction and wind speed as regressors. The resulting model is illustrated for bivariate probabilistic wind forecasts at a station in the plains and within complex terrain.

Keywords: Distributional Regression; Wind Vectors; Ensemble Forecasts.

1 Introduction

Wind is one of the classical circular quantities in atmospheric sciences, and wind speed and wind direction are mutually dependent. To gain accurate probabilistic forecasts of both wind speed and direction, wind vectors of an ensemble prediction system (EPS) are often post-processed in a distributional regression framework. In contrast to previous studies (e.g., Schuhen et al. 2012), Lang et al. (2019) propose to model all distribution parameters simultaneously using a flexible distributional regression model based on EPS forecasts of wind direction and wind speed as explanatory variables. They show that capturing correlation appropriately can improve the model and its predictions, especially when the EPS has lower predictive quality for the location and scale of the wind vectors. To gain a better understanding

This paper was published as a part of the proceedings of the 34th International Workshop on Statistical Modelling (IWSM), University of Minho, Portugal, 7-12 July 2019.

of the estimated correlation, we complement the basic linear model from Lang et al. (2019) with their advanced correlation effect specification. This naturally extends the frequently-used component-wise post-processing of both wind vectors to a full distributional regression model.

2 Bivariate Gaussian Models

The zonal and meridional components of the horizontal wind vector are represented by a bivariate Gaussian distribution with likelihood function

$$L(\boldsymbol{\mu}, \Sigma | \boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^2 |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right), \tag{1}$$

where $\boldsymbol{y} = (y_1, y_2)^\top$ are bivariate observations and $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ the distributional location parameters. The subscript asterisk acts as a placeholder for the zonal and meridional wind component from here on. The covariance matrix is defined as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{2}$$

The location parameters $\mu_\star \in \mathbb{R}$, the scale parameters $\sigma_\star > 0$, and the correlation parameter $\rho \in [-1, 1]$ are linked to additive predictors by an identity, logarithmic and rhogit link, respectively (Klein et al. 2014).
For the location and scale part, the corresponding wind component forecasts are used as covariates, namely EPS-forecasted zonal wind information (vec_1) to model the zonal component of the bivariate response, and EPS-forecasted meridional wind information (vec_2) to model the meridional component:

$$\begin{aligned} \mu_\star &= \alpha_{\star 0} + f_{\star 0}(\text{doy}) + (\alpha_{\star 1} + f_{\star 1}(\text{doy})) \cdot \overline{\text{vec}}_\star \\ \log(\sigma_\star) &= \beta_{\star 0} + g_{\star 0}(\text{doy}) + (\beta_{\star 1} + g_{\star 1}(\text{doy})) \cdot \log(\text{sd}(\text{vec}_\star)), \end{aligned} \tag{3}$$

where $\alpha_\bullet$ and $\beta_\bullet$ are regression coefficients, and $f_\bullet(\text{doy})$ and $g_\bullet(\text{doy})$ employ cyclic regression splines conditional on the day of the year (doy). The covariates $\overline{\text{vec}}_\star$ and $\log(\text{sd}(\text{vec}_\star))$ refer to the mean and log standard deviation of the EPS wind components, respectively.
The correlation structure is estimated conditional on the mean EPS wind direction ($\overline{\text{dir}}$) and speed ($\overline{\text{spd}}$) by modeling a linear interaction between these two covariates:

$$\text{rhogit}(\rho) = \gamma_0 + h_0(\text{doy}) + h_1(\overline{\text{dir}}) + (\gamma_1 + h_2(\overline{\text{dir}})) \cdot \overline{\text{spd}}, \tag{4}$$

with $\text{rhogit}(\rho) = \rho / \sqrt{(1 - \rho^2)}$; γ_0 is the global and $h_0(\text{doy})$ the seasonally varying intercept. The effect $h_1(\overline{\text{dir}})$ estimates the dependence of the correlation given the wind direction and $(\gamma_1 + h_2(\overline{\text{dir}})) \cdot \overline{\text{spd}}$ employs a varying effect of wind speed conditional on the wind direction. Note that in the meteorological context wind direction is defined on the scale $[0, 360]$ degree and increases clockwise from North.

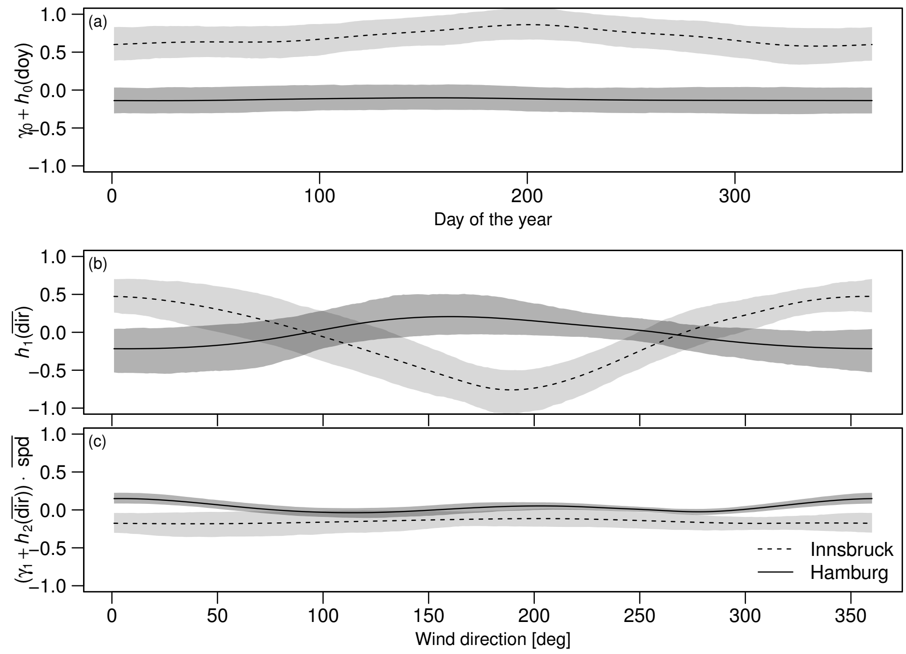

FIGURE 1. Cyclic effects for linear predictor terms of ρ on the rhogit scale: (a) Seasonal variation, (b) wind direction main effect, (c) interaction effect of wind direction and wind speed, for wind speed fixed at $1\,\mathrm{ms}^{-1}$. Effects for both Innsbruck (dashed) and Hamburg (solid) are shown on the rhogit scale along with 95% credible intervals based on MCMC sampling.

3 Results and Conclusion

To study the correlation structure as specified in Eq. (4), this section shows a detailed analysis of the estimated effects for a site in the plains (Hamburg) and an alpine site (Innsbruck) at forecast step +12 h (valid at 12 UTC).
Figure 1a shows a strong positive intercept effect for the correlation parameter for Innsbruck with little seasonal variability; the positive correlation describes stronger uncertainty along the valley than across the valley. This effect is increased when the EPS predicts northerly wind and decreased for southerly wind forecasts unconditional on the wind speed (Fig. 1b, c). For Hamburg, a slightly negative intercept effect exists unconditional on the day of the year (Fig. 1a). The effect on wind direction is negative except for southerly winds (Fig. 1b), whereas the varying effect of wind speed conditional on the wind direction is positive with the lowest coefficients for southerly winds (Fig. 1c).
Figure 2a shows the mean estimated correlation for the two stations conditional on the EPS wind direction and speed for the whole training data set. For Innsbruck, northerly EPS wind forecasts typically correspond to observed winds along the east-west orientated valley axis. Therefore, for northerly EPS wind forecasts the positive correlation describes a stronger uncertainty along the valley than across the valley (see Fig. 1b). For moderate to strong southerly EPS wind forecasts, the observed wind is typically blowing from an intersecting valley from the South and therefore the uncertainty is higher across the valley axis (negative correlation). For Hamburg,

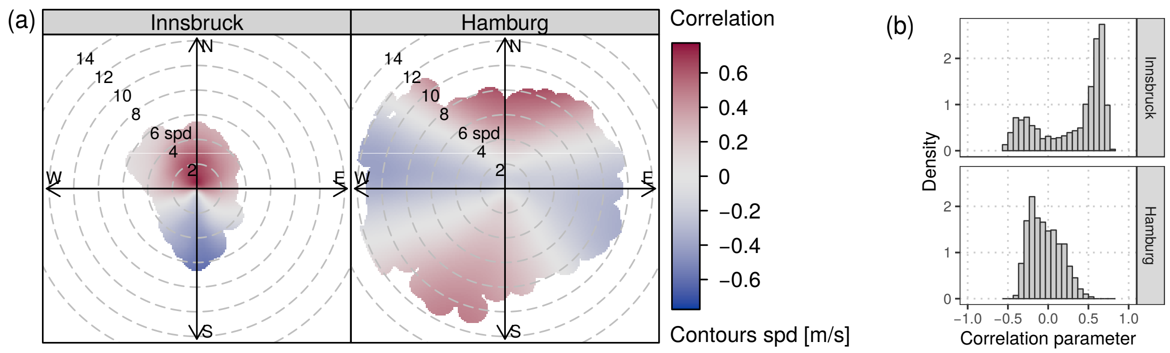

FIGURE 2. (a) Circular plot of the mean estimated correlation parameter ρ and (b) the distribution of the correlation parameters for Innsbruck and Hamburg. Correlation is shown on the parameter scale for the full training data set.

positive correlation exists for northerly and southerly EPS wind forecasts, whereas negative correlation exists for westerly and easterly wind forecasts. For all wind directions, the effects are higher with increasing wind speed (see Fig. 1c). As Hamburg lies in the plains, the distinct estimated correlation suggests that local shadowing effects due to e.g., local building and vegetation seem to have an impact. However, the estimated correlation for Hamburg is on average smaller than for Innsbruck where either rather large positive values or negative values exist (Fig. 2b). This is in accordance with Lang et al. (2019) who found that the explicit estimation of correlation is mainly important when the estimates of the location and scale parameters have only little skill due to a low information content of the EPS wind forecasts. Lang et al. (2019) show that the lack of relevant information in the EPS can be mitigated by using a more flexible location and scale specification. This dampens the correlation effects but their qualitative patterns remain unchanged.

Acknowledgments: This project was funded by the Austrian Research Promotion Agency (FFG), grant no. 858537.

References

Klein, N., Kneib, T., Klasen, S., and Lang, S. (2014). Bayesian Structured Additive Distributional Regression for Multivariate Responses. *Journal of the Royal Statistical Society C.* **64**(4), 569 – 591.

Lang, M.N., Mayr, G.J., Stauffer, R., and Zeileis, A. (2019). Bivariate Gaussian Models for Wind Vectors in a Distributional Regression Framework. *arXiv 1904.01659*, arXiv.org E-Print Archive.

Schuhen, N. , Thorarinsdottir, T.L., and Gneiting, T. (2012). Ensemble Model Output Statistics for Wind Vectors. *Monthly Weather Review.* **140**(10), 3204 – 3219.