Distributional Trees for Circular Data

Lisa Schlosser[1], Moritz N. Lang[1,2], Torsten Hothorn[3], 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
[3] Epidemiology, Biostatistics and Prevention Institute, Universität Zürich, Zürich, Switzerland

E-mail for correspondence: Lisa.Schlosser@uibk.ac.at

Abstract: For probabilistic modeling of circular data the von Mises distribution is widely used. To capture how its parameters change with covariates, a regression tree model is proposed as an alternative to more commonly-used additive models. The resulting distributional trees are easy to interpret, can detect non-additive effects, and select covariates and their interactions automatically. For illustration, hourly wind direction forecasts are obtained at Innsbruck Airport based on a set of meteorological measurements.

Keywords: Distributional Trees; Circular Response; Von Mises Distribution.

1 Motivation

Circular data can be found in a variety of applications and subject areas, e.g., hourly crime rate in the social-economics, animal movement direction or gene-structure in biology, and wind direction as one of the most important weather variables in meteorology. Circular regression models were first introduced by Fisher and Lee (1992) and further extended by Jammalamadaka and Sengupta (2001) and Mulder and Klugkist (2017) among others. While most of the already existing approaches are built on additive regression models, we propose an adaption of regression trees to circular data by employing distributional trees.

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.

2 Methodology

Distributional trees (Schlosser et. al, 2019) fuse distributional regression modeling with regression trees based on the unbiased recursive partitioning algorithms MOB (Zeileis et. al, 2008) or CTree (Hothorn et. al, 2006). The basic idea is to partition the covariate space recursively into subgroups such that an (approximately) homogeneous distributional model can be fitted to the response in each resulting subgroup. To capture dependence on covariates, the association between the model's scores and each available covariate is assessed using either a parameter instability test (MOB) or a permutation test (CTree). In each partitioning step, the covariate with the highest significant association (i.e., lowest significant p-value, if any) is selected for splitting the data. The corresponding split point is chosen either by optimizing the log-likelihood (MOB) or a two-sample test statistic (CTree) over all possible partitions.
In this study distributional trees are adapted to circular responses by employing the von Mises distribution, also known as "the circular normal distribution". Based on a location parameter $\mu \in [0, 2\pi]$ and a concentration parameter $\kappa > 0$ the density for $y \in [0, 2\pi]$ is given by:

$$f_{\mathrm{vM}}(y; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \; e^{\kappa \cos(y-\mu)} \tag{1}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order 0 (see, e.g., Jammalamadaka and Sengupta 2001, for a more detailed overview).
In each subgroup maximum likelihood estimators $\hat{\mu}$ and $\hat{\kappa}$ are obtained by maximizing the corresponding log-likelihood $\ell(\mu, \kappa; y) = \log(f_{\mathrm{vM}}(y; \mu, \kappa))$. The model scores are given by $s(y; \mu, \kappa) = (\partial_\mu \ell(\mu, \kappa; y), \partial_\kappa \ell(\mu, \kappa; y))$. In a subgroup of size n, evaluating the scores at the individual observations and parameter estimates $s(y_i; \hat{\mu}, \hat{\kappa})$ yields an $n \times 2$ matrix that can be employed as a kind of residual, capturing how well a given observation conforms with the estimated location $\hat{\mu}$ and precision $\hat{\kappa}$, respectively. Hence MOB or CTree can assess whether the scores change along with the available covariates. If so, by maximizing a partitioned likelihood the parameter instabilities are incorporated into the model. This procedure is repeated recursively until there are no significant parameter instabilities or until another stopping criterion is met (e.g., subgroup size or tree depth).

3 Application

Wind is a classical circular quantity and accurate forecasts of wind direction are of great importance for decision-making processes and risk management, e.g., in air traffic management or renewable energy production. This study employs circular regression trees to obtain hourly wind direction forecasts at Innsbruck Airport. Innsbruck lies at the bottom of a deep

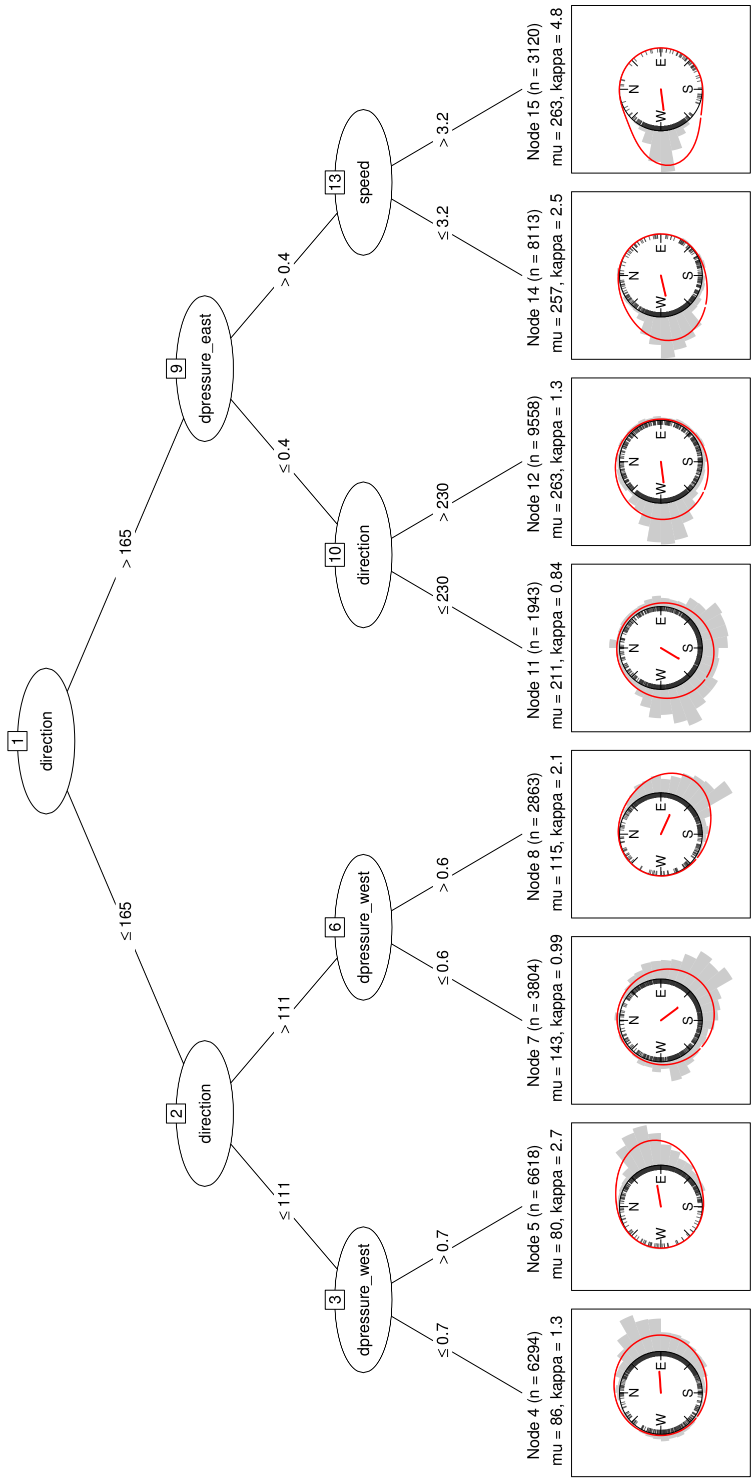

FIGURE 1. Fitted tree based on the von Mises distribution for wind direction forecasting. In each terminal node the empirical histogram (gray) and fitted density (red line) are depicted along with the estimated location parameter (red hand). The covariates employed are wind direction (degree), wind speed (ms^{-1}), and pressure gradients (dpressure; hPa) west and east of the airport, all lagged by one hour.

valley in the Alps. Topography channels wind along the west-east valley axis or along a tributary valley intersecting from the south. Hence, pressure gradients to which valley wind regimes react both west and east of the airport are considered as covariates along with other meteorological measurements at the airport (lagged by one hour), such as wind direction and wind speed at Innsbruck Airport. Note that in the meteorological context wind direction is defined on the scale $[0, 360]$ degree and increases clockwise from North (0 degree).

Figure 1 depicts the resulting distributional tree, including both the empirical (gray) and fitted von Mises (red) distribution of wind direction in each terminal node. Based on the fitted location parameters $\hat{\mu}$, the subgroups can be distinguished into the following wind regimes: (1) Up-valley winds blowing from the valley mouth towards the upper valley (from east to west, nodes 4 and 5). (2) Downslope winds blowing across the Alpine crest along the intersecting valley towards Innsbruck (from south-east to north-west, nodes 7 and 8). (3) Down-valley winds blowing in the direction of the valley mouth (from west to east, nodes 12, 14, and 15). Node 11 captures observations with rather low wind speeds that cannot be distinguished clearly into wind regimes and consequently are associated with a very low estimated concentration $\hat{\kappa}$. In terms of covariates, the lagged wind direction ("persistence") is mostly responsible for distinguishing the broad wind regimes listed above while the pressure gradients and wind speed separate between subgroups with high vs. low precision.

4 Discussion and outlook

Distributional trees for circular responses are established by coupling model-based recursive partitioning with the von Mises distribution. The resulting trees can capture nonlinear changes, shifts, and potential interactions in covariates without prespecification of such effects. This is particularly useful for modeling wind direction in mountainous terrain where wind shifts can occur due to turns of the pressure gradients along a valley.

4.1 Ensembles and random forests

A natural extension are ensembles or forests of such circular trees that can improve forecasts by regularizing and stabilizing the model. Random forests introduced by Breiman (2001) average the predictions of an ensemble of trees, each built on a subsample or bootstrap of the original data. A generalization of this strategy is to obtain weighted predictions by adaptive local likelihood estimation of the distributional parameters (Schlosser et. al, 2019). More specifically, for each possibly new observation x a set of "nearest neighbor" weights $w_i(x)$ is obtained that is based on how often x is assigned to the same terminal node as each learning observation $y_i, i \in \{1, \ldots, n\}$.

The parameters μ and κ are then estimated for each (new) observation x by weighted maximum likelihood based on the adaptive nearest neighbor weights:

$$\underset{\mu,\kappa}{\text{argmax}} \sum_{i=1}^{n} w_i(x) \cdot \ell(\mu, \kappa; y_i). \tag{2}$$

Therefore, the resulting parameter estimates can smoothly adapt to the given covariates x whereas $w_i(x) = 1$ would correspond to the unweighted full-sample estimates and $w_i(x) \in \{0, 1\}$ corresponds to the abrupt splits from the tree.

4.2 Splits in circular covariates

In order to obtain more parsimonious and more stable trees another possible extension for *circular covariates* (with or without a *circular response*) is to consider their circular nature when searching the best split into two segments. In general, searching the best separation of a covariate into a "left" and "right" daughter node tries to maximize the segmented log-likelihood:

$$\max \left(\sum_{y \in left} \ell(\hat{\mu_1}, \hat{\kappa_1}; y) + \sum_{y \in right} \ell(\hat{\mu_2}, \hat{\kappa_2}; y) \right) \tag{3}$$

where $\hat{\mu_1}$, $\hat{\kappa_1}$, $\hat{\mu_2}$, $\hat{\kappa_2}$ are the estimated parameters of the von Mises distribution in the two daughter nodes. Searching a single split point ν in a circular covariate $\in [0, 2\pi)$ only considers linear splits into the intervals *left* $= [0, \nu]$ and *right* $= (\nu, 2\pi)$, thus enforcing a potentially unnatural separation at zero. This can be avoided by searching for two split points ν and τ considering a split into one interval *left* $= [\nu, \tau]$ and its complement *right* $= [0, \nu) \cup (\tau, 2\pi)$, encompassing zero. The latter strategy is invariant to the (often arbitrary) definition of the direction at zero.
When one split point ν is sufficiently close to zero and the other τ sufficiently far away, a simple linear split typically suffices to capture such a split (as seen for the lagged wind direction in Figure 1). If both ν and τ differ clearly from zero, two linear splits should also lead to a reasonable (but less parsimonious) fit. However, if both ν and τ are rather close to zero, a linear split strategy might miss such a pattern.
The required test statistic to maximally select two split points simultaneously is straightforward to accommodate in the CTree framework by providing all binary indicators corresponding to the splits into *left*/*right* intervals. However, this will become increasingly slow for larger sample sizes but it might be possible to speed up computations by exploiting the particular covariance structure similar to Hothorn and Zeileis (2008). In the MOB framework the extension is not quite as straightforward but one strategy could be to adapt double maximum tests à la Bai and Perron (2003).

Hence, the splitting idea can be naturally extended to a two-point search, however, for an unbiased and inference-based selection the corresponding testing strategies might need further adaption.

Computational details: R packages implementing the proposed methods are currently under development at `https://R-Forge.R-project.org/projects/partykit/`.

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

References

Bai, J., and Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, **18**, 1 – 22.

Breiman, L. (2001). Random Forests. *Machine Learning*, **45**, 1, 5 – 32.

Fisher, N. I., and Lee, A. J. (1992). Regression Models for an Angular Response. *Biometrics*, **48**, 3, 665 – 677.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, **15**, 3, 651 – 674.

Hothorn, T., and Zeileis, A. (2008). Generalized Maximally Selected Statistics. *Biometrics*, **64**, 4, 1263 – 1269.

Jammalamadaka, S. R., and Sengupta, A. (2001). *Topics in Circular Statistics*. World Scientific.

Mulder, K., and Klugkist, I. (2017). Bayesian Estimation and Hypothesis Tests for a Circular Generalized Linear Model. *Journal of Mathematical Psychology*, **80**, 4 – 14.

Schlosser, L., Hothorn, T., and Zeileis, A. (2019). Distributional Regression Forests for Probabilistic Precipitation Forecasting in Complex Terrain. arXiv:1804.02921, *arXiv.org E-Print Archive*.

Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-Based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, **17**, 2, 492 – 514.