universität
innsbruck

Research Platform
Empirical and Experimental Economics

Distributional regression forests for probabilistic precipitation forecasting in complex terrain

Lisa Schlosser, Torsten Hothorn, Reto Stauffer, Achim Zeileis

Working Papers in Economics and Statistics

2018-08

University of Innsbruck
https://www.uibk.ac.at/eeecon/

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 71022
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

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

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

Distributional Regression Forests for Probabilistic Precipitation Forecasting in Complex Terrain

Lisa Schlosser Universität Innsbruck
Torsten Hothorn Universität Zürich
Reto Stauffer Universität Innsbruck
Achim Zeileis Universität Innsbruck

Abstract

To obtain a probabilistic model for a dependent variable based on some set of explanatory variables, a distributional approach is often adopted where the parameters of the distribution are linked to regressors. In many classical models this only captures the location of the distribution but over the last decade there has been increasing interest in distributional regression approaches modeling all parameters including location, scale, and shape. Notably, so-called non-homogenous Gaussian regression (NGR) models both mean and variance of a Gaussian response and is particularly popular in weather forecasting. More generally, the GAMLSS framework allows to establish generalized additive models for location, scale, and shape with smooth linear or nonlinear effects. However, when variable selection is required and/or there are non-smooth dependencies or interactions (especially unknown or of high-order), it is challenging to establish a good GAMLSS. A natural alternative in these situations would be the application of regression trees or random forests but, so far, no general distributional framework is available for these. Therefore, a framework for distributional regression trees and forests is proposed that blends regression trees and random forests with classical distributions from the GAMLSS framework as well as their censored or truncated counterparts. To illustrate these novel approaches in practice, they are employed to obtain probabilistic precipitation forecasts at numerous sites in a mountainous region (Tyrol, Austria) based on a large number of numerical weather prediction quantities. It is shown that the novel distributional regression forests automatically select variables and interactions, performing on par or often even better than GAMLSS specified either through prior meteorological knowledge or a computationally more demanding boosting approach.

Keywords: parametric models, regression trees, random forests, recursive partitioning, probabilistic forecasting, GAMLSS.

1. Introduction

In regression analysis a wide range of models has been developed to describe the relationship between a response variable and a set of covariates. The classical model is the linear model (LM) where the conditional mean of the response is modeled through a linear function of the covariates (see the left panel of Figure 1 for a schematic illustration). Over the last decades this has been extended in various directions including:

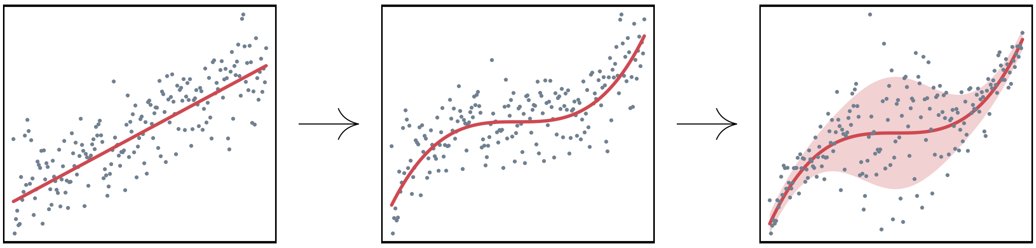

Figure 1: Parametric modeling developments. (Generalized) linear models (left), generalized additive models (middle), generalized additive models for location, scale, and shape (right).

- *Generalized linear models* (GLMs, Nelder and Wedderburn 1972) encompassing an additional nonlinear link function for the conditional mean.
- *Generalized additive models* (GAMs, Hastie and Tibshirani 1986) allowing for smooth nonlinear effects in the covariates (Figure 1, middle).
- *Generalized additive models for location, scale, and shape* (GAMLSS, Rigby and Stasinopoulos 2005) adopting a probabilistic modeling approach. In GAMLSS, each parameter of a statistical distribution can depend on an additive predictor of the covariates comprising linear and/or smooth nonlinear terms (Figure 1, right).

Thus, the above-mentioned models provide a broad toolbox for capturing different aspects of the response (mean only vs. full distribution) and different types of dependencies on the covariates (linear vs. nonlinear additive terms).

While in many applications conditional mean regression models have been receiving the most attention, there has been a paradigm shift over the last decade towards distributional regression models. An important reason for this is that in many fields forecasts of the mean are not the only (or not even the main) concern but instead there is an increasing interest in probabilistic forecasts. Quantities of interest typically include exceedence probabilities for certain thresholds of the response or quantiles of the response distribution. Specifically, consider weather forecasting where there is less interest in the mean amount of precipitation on the next day. Instead, the probability of rain vs. no rain is typically more relevant or, in some situations, a prediction interval of expected precipitation (say from the expected 10% to 90% quantile). Similar considerations apply for other meteorological quantities and hence attention in the weather forecasting literature has been shifting from classical linear deterministic models (Glahn and Lowry 1972) towards probabilistic models such as the non-homogeneous Gaussian regression (NGR) of Gneiting, Raftery, Westveld III, and Goldman (2005). The NGR typically describes the mean of some meteorological response variable through the average of the corresponding quantity from an ensemble of physically-based numerical weather predictions (NWPs). Similarly, the variance of the response is captured through the variance of the ensemble of NWPs. Thus, the NGR considers both the mean as well as the uncertainty of the ensemble predictions to obtain probabilistic forecasts calibrated to a particular site.

In summary, the models discussed so far provide a broad and powerful toolset for parametric distributional fits depending on a specified set of additive linear or smooth nonlinear terms.

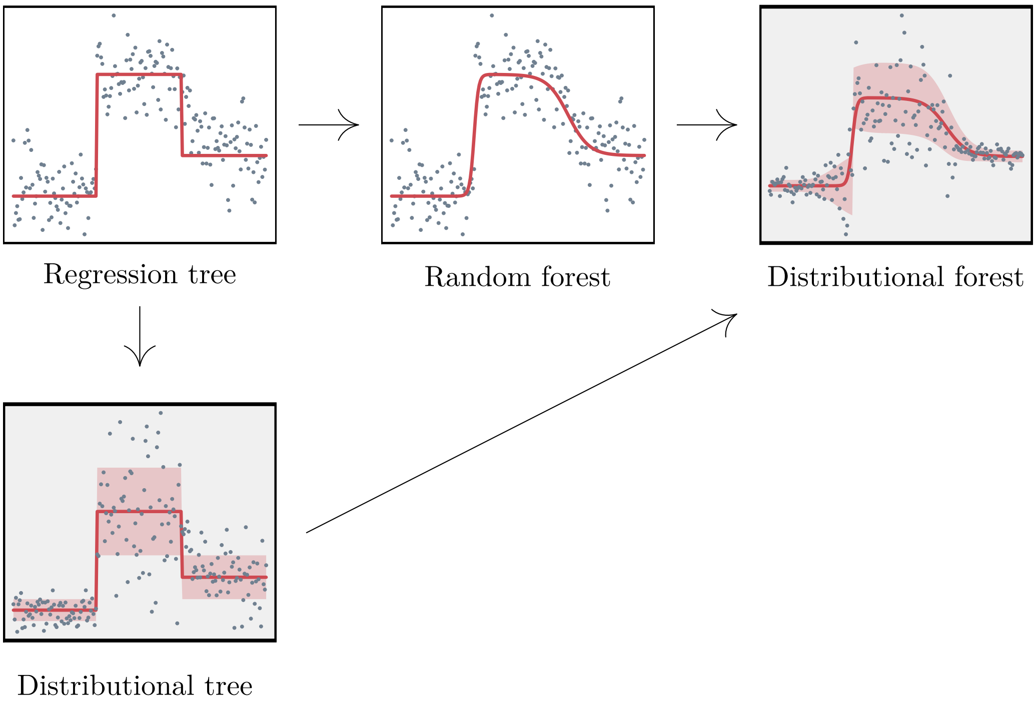

Figure 2: Tree and forest developments. Regression tree (top left), distributional tree (bottom left), random forest (top middle), and distributional forest (top right).

A rather different approach to capturing the dependence on covariates are tree-based models.

- *Regression trees* (Breiman, Friedman, Olshen, and Stone 1984) recursively split the data into more homogeneous subgroups and can thus capture abrupt shifts (Figure 2, top left) and approximate nonlinear functions. Furthermore, trees automatically carry out a forward selection of covariates and their interactions.
- *Random forests* (Breiman 2001) average the predictions of an ensemble of trees fitted to resampled versions of the learning data. This stabilizes the recursive partitions from individual trees and hence better approximates smooth functions (Figure 2, top middle)

While classical regression trees and random forests only model the mean of the response we propose to follow the ideas from GAMLSS modeling – as outlined in Figure 1 – and combine tree-based methods with parametric distributional models, yielding two novel techniques:

- *Distributional regression trees* (for short: *distributional trees*) split the data into more homogeneous groups with respect to a parametric distribution, thus capturing changes in any distribution parameter like location, scale, or shape (Figure 2, bottom left)
- *Distributional regression forests* (for short: *distributional forests*) utilize an ensemble of distributional trees for obtaining stabilized and smoothed parametric predictions (Figure 2, top right).

In the following, particular focus is given to distributional forests as a method for obtaining probabilistic forecasts by leveraging the strengths of random forests: the ability to capture

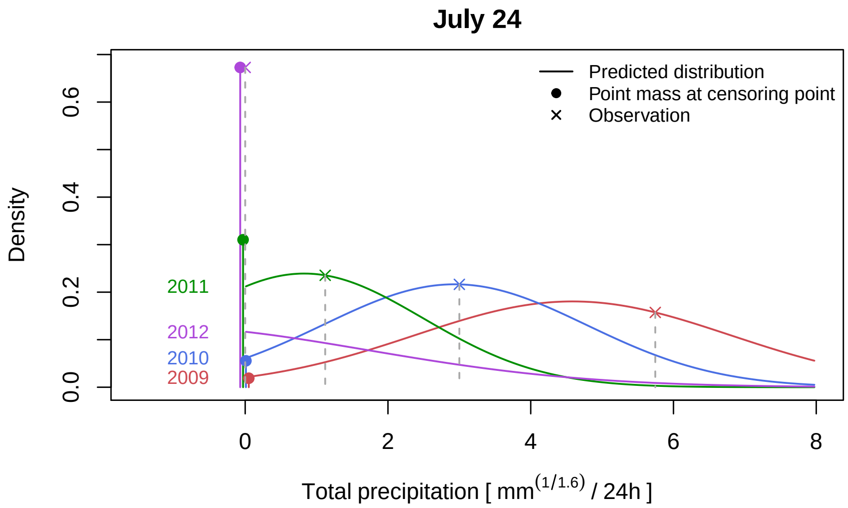

Figure 3: Total precipitation predictions by a distributional forest at station Axams for July 24 in 2009, 2010, 2011 and 2012 learned on data from 1985–2008. Observations are left-censored at 0. The corresponding predicted point mass is shown at the censoring point (0).

both smooth and abruptly changing functions along with simultaneous selection of variables and possibly complex interactions. Thus, these properties make the method particularly appealing in case of many covariates with unknown effects and interactions where it would be challenging to specify a distributional regression model like GAMLSS.

In weather forecasting, these properties are especially appealing in mountainous regions and complex terrain where a wide range of local-scale effects are not yet resolved by the NWP models. Thus, effects with abrupt changes and possibly nonlinear interactions might be required to account for site-specific unresolved features. To illustrate this in practice, precipitation forecasts are obtained with distributional forests at 95 meteorological stations in a mountainous region in the Alps, covering mainly Tyrol, Austria, and adjacent areas (see the map in Figure 8). More specifically, a Gaussian distribution, left-censored at zero, is employed to model 24-hour total precipitation so that the zero-censored point mass describes the probability of observing no precipitation on a given day (see Figure 3). Forecasts for July are established based on data from the same month over the years 1985–2012 including 80 covariates derived from a wide range of different NWP quantities. As Figure 3 shows, the station-wise forests yield a full distributional forecast for each day – here for one specific day (July 24) at one station (Axams) over four years (2009–2012) – based on the previous 24 years as learning data. The corresponding observations conform reasonably well with the predictions. In Section 3 we investigate the performance of distributional forests in this forecasting task in more detail. Compared to three alternative zero-censored Gaussian models distributional forests perform at least on par and sometimes clearly better while requiring no expert knowledge for the model specification. The three alternatives are: a standard ensemble model output statistics approach (EMOS, Gneiting *et al.* 2005) based on an NGR, a GAMLSS with regressors prespecified based on meteorological expertise (following Stauffer, Umlauf, Messner, Mayr, and Zeileis 2017b), and a boosted GAMLSS (Hofner, Mayr, and Schmid 2016) using non-homogeneous boosting (Messner, Mayr, and Zeileis 2017) as an alternative technique for variable selection among all 80 available regressors.

2. Methodology

To embed the distributional approach from GAMLSS into regression trees and random forests, we proceed in three steps. (1) To fix notation, we briefly review fitting distributions using standard maximum likelihood in Section 2.1. (2) A recursive partitioning strategy based on the corresponding scores (or gradients) is introduced in Section 2.2, leading to distributional trees. (3) Ensembles of distributional trees fitted to randomized subsamples are employed to establish distributional forests in Section 2.3.

The general distributional notation is exemplified in all three steps for the Gaussian distribution, left-censored at zero. The latter is employed in the empirical case study in Section 3 to model power-transformed daily precipitation amounts.

2.1. Distributional fit

A distributional model $\mathcal{D}(Y, \boldsymbol{\theta})$ is considered for the response variable $Y \in \mathcal{Y}$ using the distributional family $\mathcal{D}$ with k-dimensional parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and corresponding log-likelihood function $\ell(\boldsymbol{\theta}; Y)$. The GAMLSS framework (Rigby and Stasinopoulos 2005) provides a wide range of such distributional families with parameterizations corresponding to location, scale, and shape. Furthermore, censoring and/or truncation of these distributions can be incorporated in the usual straightforward way (see e.g., Long 1997, Chapter 7.2).

To capture both location and scale of the probabilistic precipitation forecasts while accounting for a point mass at zero (i.e., dry days without rain), a Gaussian distribution left-censored at zero (for short: zero-censored Gaussian) with location parameter μ and scale parameter σ is employed. Therefore, the corresponding log-likelihood function with parameter vector $\boldsymbol{\theta} = (\mu, \sigma)$ is

$$\ell(\mu, \sigma; Y) = \begin{cases} \log\left\{\frac{1}{\sigma} \cdot \phi\left(\frac{Y-\mu}{\sigma}\right)\right\}, & \text{if } Y > 0 \\ \log\left\{\Phi\left(\frac{-\mu}{\sigma}\right)\right\}, & \text{if } Y = 0 \end{cases}$$

where ϕ and Φ are the probability density function and the distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. Other distributions $\mathcal{D}$ and corresponding log-likelihoods $\ell(\mu, \sigma; Y)$ could be set up in the same way, e.g., for censored shifted gamma distributions (Scheuerer and Hamill 2015) or zero-censored logistic distributions (Gebetsberger, Messner, Mayr, and Zeileis 2017).

With the specification of the distribution family and its log-likelihood function the task of fitting a distributional model turns into the task of estimating the distribution parameter $\boldsymbol{\theta}$. This is commonly done by maximum likelihood (ML) based on the learning sample with observations $\{y_i\}_{i=1,\dots,n}$ of the response variable Y. The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \ell(\boldsymbol{\theta}; y_i).$$

Equivalently, this can be defined based on the corresponding first-order conditions

$$\sum_{i=1}^{n} s(\hat{\boldsymbol{\theta}}, y_i) = 0,$$

where $s(\boldsymbol{\theta}; y_i)$ is the associated score function

$$s(\boldsymbol{\theta}; y_i) = \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}; y_i).$$

The latter is subsequently employed as a general goodness-of-fit measure to assess how well the distribution with parameters $\boldsymbol{\theta}$ fits one individual observation y_i.

2.2. Distributional tree

Typically, a single global model $\mathcal{D}(Y, \boldsymbol{\theta})$ is not sufficient for reasonably representing the response distribution and covariates $\mathbf{Z} = Z_1, \dots, Z_m \in \mathcal{Z}$ are employed to capture differences in the distribution parameters $\boldsymbol{\theta}$. In weather forecasting, these covariates typically include the output from numerical weather prediction systems and/or lagged meteorological observations.

To incorporate the covariates into the distributional model, they are considered as regressors in additive predictors $g_j(\theta_j) = f_{j,1}(\mathbf{Z}) + f_{j,2}(\mathbf{Z}) + \dots$ in GAMLSS. Link functions $g_j(\cdot)$ are used for every parameter θ_j $(j = 1, \dots, k)$ based on smooth terms $f_{j,k}$ such as nonlinear effects, spatial effects, random coefficients, or interaction surfaces (Klein, Kneib, Lang, and Sohn 2015). However, this requires specifying the additive terms and their functional forms in advance which can be challenging in practice and potentially require domain knowledge, especially if the number of covariates m is large.

Regression trees generally take a different approach for automatically including covariates in a data-driven way and allowing for abrupt changes, nonlinear and non-additive effects, and interactions. In the context of distributional models the goal is to partition the covariate space $\mathcal{Z}$ recursively into disjoint segments so that a homogenous distributional model for the response Y can be found with segment-specific parameters. More specifically, the B disjoint segments $\mathcal{B}_b$ $(b = 1, \dots, B)$ partition the covariate space

$$\mathcal{Z} = \dot{\bigcup_{b=1,\dots,B}} \mathcal{B}_b,$$

and a local distributional model $\mathcal{D}(Y, \boldsymbol{\theta}^{(b)})$ (i.e., with segment-specific parameters $\boldsymbol{\theta}^{(b)}$) is fitted to the response Y in each segment.

To find the segments $\mathcal{B}_b$ that are (approximately) homogenous with respect to the distributional model with given parameters, the idea is to use a gradient-based recursive-partitioning approach. In a given subsample of the learning data this fits the model by ML (see Equation 2.1) and then assesses the goodness of fit by assessing the corresponding scores $s(\hat{\boldsymbol{\theta}}; y_i)$ (see Equation 2.1).

To sum up, distributional trees are fitted recursively via:

1. Estimate $\hat{\boldsymbol{\theta}}$ via maximum likelihood for the observations in the current subsample.
2. Test for associations or instabilities of the scores $s(\hat{\boldsymbol{\theta}}, y_i)$ and $Z_{l,i}$ for each partitioning variable Z_l $(l = 1, \dots, m)$.
3. Split the sample along the partitioning variable Z_l^* with the strongest association or instability. Choose the breakpoint with the highest improvement in the log-likelihood or the highest discrepancy.

4. Repeat steps 1–3 recursively in the subsamples until these become too small or there is no significant association/instability (or some other stopping criterion is reached).

Different inference techniques can be used for assessing the association between scores and covariates in step 3. In the following we use the general class of permutation tests introduced by Hothorn, Hornik, Van de Wiel, and Zeileis (2006a) which is also the basis of conditional inference trees (CTree, Hothorn, Hornik, and Zeileis 2006b). Alternatively, one could use asymptotic M-fluctuation tests for parameter instability (Zeileis and Hornik 2007) as in model-based recursive partitioning (MOB, Zeileis, Hothorn, and Hornik 2008). More details are provided in Appendix A.

For obtaining probabilistic predictions from the tree for a (possibly new) set of covariates $\mathbf{z} = (z_1, \dots, z_m)$, the observation simply has to be "sent down" the tree and the corresponding segment-specific MLE has to be obtained. This can also be understood as a weighted MLE where the weights select those observations from the learning sample that fall into the same segment:

$$w_i^{\text{tree}}(\mathbf{z}) = \sum_{b=1}^{B} \mathbf{1}((\mathbf{z}_i \in \mathcal{B}_b) \wedge (\mathbf{z} \in \mathcal{B}_b)),$$

where $\mathbf{1}(\cdot)$ is the indicator function. The predicted distribution for a given $\mathbf{z}$ is then fully specified by the estimated parameter $\hat{\boldsymbol{\theta}}(\mathbf{z})$ where

$$\hat{\boldsymbol{\theta}}(\mathbf{z}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{i=1}^{n} w_i^{\text{tree}}(\mathbf{z}) \cdot \ell(\boldsymbol{\theta}; y_i).$$

2.3. Distributional forest

While the simple recursive structure of a tree model is easy to visualize and interpret, the abrupt changes are often too rough, instable, and impose steps on the model even if the true underlying effect is smooth. Hence, ensemble methods such as bagging or random forests (Breiman 2001) are typically applied to smooth the effects, stabilize the model, and improve predictive performance.

While Breiman and Cutler's random forests (Breiman 2001) grow ensembles of trees that pick up changes in the location of the response across the covariates, *distributional forests* employ an ensemble of T *distributional trees*. These pick up changes in the "direction" of any distribution parameter by considering the full score vector for choosing splitting variables and split points. Each of the distributional trees is grown on a different data set obtained through bootstrap sampling (or subsampling) and in each node only a random subset of the covariates $\mathbf{Z}$ is considered. As usual in random forests, this reduces the correlation among the trees and stabilizes the variance of the model.

To obtain probabilistic predictions from a distributional forest, it still needs to be specified how to compute the parameter estimates $\hat{\boldsymbol{\theta}}(\mathbf{z})$ for a (potentially new) set of covariates $\mathbf{z}$. Following Hothorn and Zeileis (2017) we interpret random forests as adaptive local likelihood estimators using the averaged "nearest neighbor weights" (Lin and Jeon 2006) from the T trees in the forest

$$w_i^{\text{forest}}(\mathbf{z}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{b=1}^{B^t} \frac{\mathbf{1}((\mathbf{z}_i \in \mathcal{B}_b^t) \wedge (\mathbf{z} \in \mathcal{B}_b^t))}{|\mathcal{B}_b^t|},$$

where $|\mathcal{B}_b^t|$ denotes the number of observations in the b-th segment of the t-th tree. Thus, these $w_i^{\text{forest}}(\mathbf{z}) \in [0, 1]$ whereas $w_i^{\text{tree}}(\mathbf{z}) \in \{0, 1\}$. Hence, weights cannot only be 0 or 1 but change more smoothly, giving high weight to those observations i from the learning sample that co-occur in the same segment $\mathcal{B}_b^t$ as the new observation $\mathbf{z}$ for many of the trees $t = 1, \dots, T$. Consequently, the parameter estimates may, in principle, change for every observation and can be obtained by

$$\hat{\boldsymbol{\theta}}(\mathbf{z}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{i=1}^{n} w_i^{\text{forest}}(\mathbf{z}) \cdot \ell(\boldsymbol{\theta}; y_i).$$

In summary, this yields a parametric distributional regression model (through the score-based approach) that can capture both abrupt effects and high-order interactions (through the trees) and smooth effects (through the forest).

Distributional forests share some concepts and algorithmic aspects with other generalizations of Breiman and Cutler's random forests. Nearest neighbor weights are employed for aggregation in survival forests (Hothorn, Lausen, Benner, and Radespiel-Tröger 2004), quantile regression forests (Meinshausen 2006), transformation forests (Hothorn and Zeileis 2017), and generalized random forests for causal inferences (Athey, Tibshirani, and Wager 2017). These procedures aggregate over trees fitted to specific score functions (e.g., log rank scores in survival trees, model residuals in transformation or generalized forests). Distributional forests, in contrast to these nonparametric approaches, provide a compromise between model flexibility and interpretability: The parameters of a problem-specific distribution (zero-censored Gaussian for precipitation) have a clear meaning but may depend on external variables in a quite general way.

3. Probabilistic precipitation forecasting in complex terrain

Many statistical weather forecasting models leverage the strengths of modern numerical weather prediction (NWP) systems (see Bauer, Thorpe, and Brunet 2015). One frequently used method based on distributional regression models is the ensemble model output statistics (EMOS) approach first proposed by Gneiting *et al.* (2005) to produce high-quality forecasts for specific quantities and sites. In case of precipitation forecasting EMOS typically uses the ensemble mean of "total precipitation" (*tp*) forecasts as predictor for the location parameter and the corresponding ensemble standard deviation for the scale part of the statistical model to correct for possible errors of the ensemble in both, the expectation but also the uncertainty of a specific forecast.

While this approach alone is already highly effective in the plains, it typically does not perform as well in complex terrain due to unresolved effects in the NWP system. For example, in the Tyrolean Alps – considered in the following case study – the NWP grid cells of 50×50 km^2 are too coarse to capture single mountains, narrow valleys, etc. Therefore, it is often possible to substantially improve the predictive performance of the EMOS by including additional predictor variables, either from local meteorological observations or an NWP model. Unfortunately, it is typically unknown which variables are relevant for improving the predictions. Simply including all available variables may be computationally burdensome and can lead to overfitting but, on the other hand, excluding too many variables may result in a loss of valuable information. Therefore, selecting the relevant variables and interactions among all possible covariates is crucial for improving the statistical forecasting model.

In the following, it is illustrated how distributional forests can solve this problem without requiring prior expert knowledge. For fitting the forest only the response distribution and the list of potential predictor variables need to be specified (along with a few algorithmic details) and then the relevant variables, interactions, and potentially nonlinear effects are determined automatically in a data-driven way. Here, we employ a zero-censored Gaussian distribution and 80 predictor variables computed from ensemble means and spreads of various NWP outputs. The predictive performance of the forest is compared to three other zero-censored Gaussian models: (a) a standard basic EMOS, (b) a GAMLSS with prespecified effects and interactions based on meteorological knowledge/experience, and (c) a boosted GAMLSS with automatic selection of smooth additive terms based on all 80 predictor variables.

3.1. Data

Learning and validation data consist of observed daily precipitation sums provided by the National Hydrographical Service (BMLFUW 2016) and numerical weather forecasts from the U.S. National Oceanic and Atmospheric Administration (NOAA). Both, observations and forecasts are available for 1985–2012 and the analysis is exemplified using July, the month with the most precipitation in Tyrol.

Observations are obtained for 95 stations all over Tyrol and surroundings, providing 24-hour precipitation sums measured at 0600 UTC and rigorously quality-checked by the National Hydrographical Service. NWP outputs are obtained from the second generation reforecast data set of the global ensemble forecast system (GEFS, Hamill *et al.* 2013). This data set consists of an 11-member ensemble based on a fixed version of the numerical model and a horizontal grid spacing of about 50×50 km^2 initialized daily at 0000 UTC from December 1984 to present providing forecasts on a 6-hourly temporal resolution. Each of the 11 ensemble members uses slightly different perturbed initial conditions to predict the situation-specific uncertainty of the atmospheric state.

From the GEFS, 14 basic forecast variables are considered with up to 12 variations each such as mean/maximum/minimum over different aggregation time periods. A detailed overview is provided in Table 1, yielding 80 predictor variables in total.

To remove large parts of the skewness of precipitation data, a power transformation (Box and Cox 1964) is often applied, e.g., using cubic (Stidd 1973) or square root (Hutchinson 1998) transformations. However, the power parameter may vary for different climatic zones or temporal aggregation periods and hence we follow Stauffer, Mayr, Messner, Umlauf, and Zeileis (2017a) in their choice of 1.6^{-1} as a suitable power parameter for the region of Tyrol. The same power transformation is applied to both the observed precipitation sums and the NWP outputs "total precipitation" (*tp*) and "convective available potential energy" (*cape*).

3.2. Models and evaluation

The following zero-censored Gaussian regression models are employed in the empirical case study, see Table 2 for further details:

- *Distributional forest:* All 80 predictor variables are considered for learning a forest of 100 trees. Bootstrap sampling is employed for each tree using a third of the predictors in each split of the tree ("mtry"). Parameters are estimated by adaptive local likelihood based on the forest weights as described in Section 2.

Basic covariates	#	Variations
tp: total precipitation, power transformed (by 1.6^{-1}) *cape*: convective available potential energy, power transformed (by 1.6^{-1})	12	ensemble mean of sums over 24h, ensemble std. deviation of sums over 24h, ensemble minimum of sums over 24h, ensemble maximum of sums over 24h all for 6–30 ensemble mean of sums over 6h for 6–12, 12–18, 18–24, 24–30 ensemble std. deviation of sums over 6h for 6–12, 12–18, 18–24, 24–30
dswrf: downwards short wave radiation flux (“sunshine”) *msl*: mean sea level pressure *pwat*: precipitable water *tmax*: 2m maximum temperature *tcolc*: total column-integrated condensate *t500*: temperature on 500 hPa *t700*: temperature on 700 hPa *t850*: temperature on 850 hPa	6	ensemble mean of mean values, ensemble mean of minimum values*, ensemble mean of maximal values, ensemble std. deviation of mean values, ensemble std. deviation of minimum values*, ensemble std. deviation of maximal values, all over 6–30
tdiff500850: temperature difference 500 to 850 hPa *tdiff500700*: temperature difference 500 to 700 hPa *tdiff700850*: temperature difference 700 to 850 hPa	3	ensemble mean of difference in mean, ensemble minimum of difference in mean, ensemble maximum of difference in mean all over 6–30
msl_diff: mean sea level pressure difference	1	$msl\_mean\_max - msl\_mean\_min$ over 6–30

Table 1: Basic covariates together with the number (#) and the type of variations. Time periods indicate aggregation time periods in hours after NWP model initialization (e.g., 6–30 corresponds to +6 h to +30 h ahead forecasts, 0600 UTC to 0600 UTC of the next day). *Minimum values of *dswrf* over 24 h are always zero and thus neglected.

- *EMOS:* Standard ensemble model output statistics models use the ensemble mean of total precipitation as regressor in the location submodel and the corresponding ensemble standard deviation in the scale submodel. The parameters are estimated by maximum likelihood, using an identity link for the location part and a log link for the scale part (following the advice of Gebetsberger *et al.* 2017).

- *Prespecified GAMLSS:* Smooth additive splines are selected for the most relevant predictors based on meteorological expert knowledge following Stauffer *et al.* (2017b). More specifically, based on the 80 available variables, eight terms are included in the loca-

Model	Type	Location (μ)	Scale ($\log(\sigma)$)
Distributional forest	recursive partitioning	all	all
EMOS	linear	*tp_mean*	*tp_sprd*
Prespecified GAMLSS	spline in each	*tp_mean*, *tp_max*, *tp_mean1218* $*$ *cape_mean1218*, *dswrf_mean_mean*, *tcolc_mean_mean*, *pwat_mean_mean*, *tdiff500850_mean*, *msl_diff*	*tp_sprd*, *dswrf_sprd_mean*, *tp_sprd1218* $*$ *cape_mean1218*, *tcolc_sprd_mean*, *tdiff500850_mean*
Boosted GAMLSS	spline in each	all	all

Table 2: Overview of models with type of covariate dependency and included covariates for each distribution parameter. $A * B$ indicates an interaction between covariate A and B.

tion submodel and five in the scale submodel. Both involve an interaction of *tp* and *cape* in the afternoon (between 1200 UTC and 1800 UTC) to capture the potential for thunderstorms that frequently occur in summer afternoons in the Alps. The model is estimated by maximum penalized likelihood using a backfitting algorithm (Stasinopoulos and Rigby 2007).

- *Boosted GAMLSS:* Smooth additive splines are selected automatically from all 80 available variables, using non-cyclic boosting for parameter estimation (Hofner *et al.* 2016; Messner *et al.* 2017). This updates the predictor terms for the location or scale submodels iteratively by maximizing the log-likelihood only for the variable yielding the biggest improvement. The iteration stops early – before fully maximizing the in-sample likelihood – based on a (computationally intensive) out-of-bag bootstrap estimate of the log-likelihood. The grid considered for the number of boosting iterations ("mstop") is: $50, 75, \dots, 975, 1000$.

The predictive performance in terms of full probabilistic forecasts is assessed using the continuous ranked probability score (CRPS, Hersbach 2000). For each of the models this assesses the discrepancy of the predicted distribution function F from the observation y by

$$\mathrm{CRPS}(y, F) = \int_{-\infty}^{\infty} (F(z) - \mathbf{1}(y \leq z))^2 dz$$

where $\mathbf{1}(\cdot)$ is the indicator function. In the subsequent applications, the mean CRPS is always evaluated out of sample, either using cross-validation or a hold-out data set (2009–2012) that was not used for learning (1985–2008). CRPS is a proper scoring rule (Gneiting and Raftery 2007) often used within the meteorological community. Lower values indicate better performance.

To assess differences in the improvement of the forests and GAMLSS models over the basic

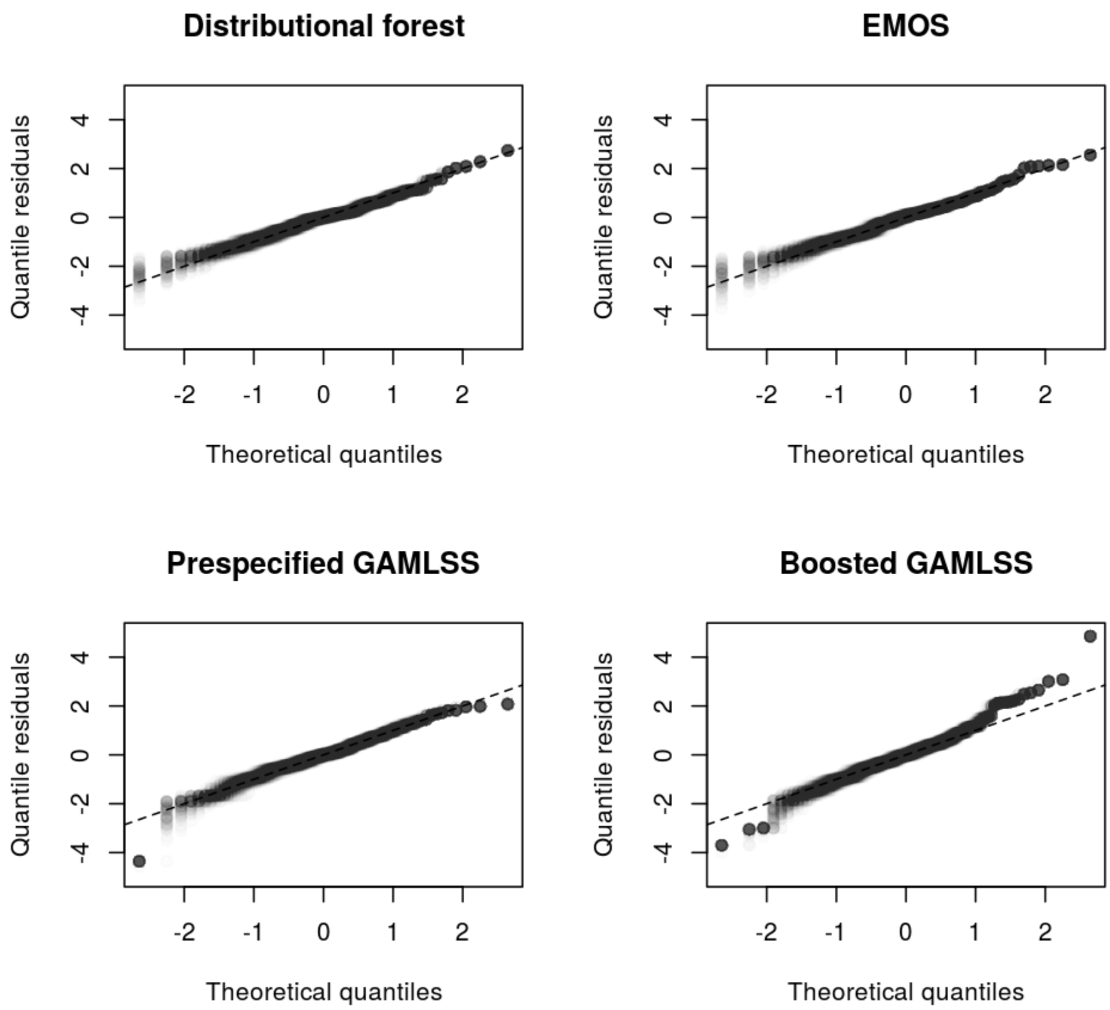

Figure 4: Out-of-sample residual QQ plots (2009–2012) for station Axams based on models learned on data from 1985–2008.

EMOS, a CRPS-based skill score with EMOS as the reference method is computed:

$$\mathrm{CRPSS}_{\mathrm{method}} = 1 - \frac{\mathrm{CRPS}_{\mathrm{method}}}{\mathrm{CRPS}_{\mathrm{EMOS}}}.$$

3.3. Application for one station

In a first step, we show a detailed comparison of the competing models for one observation site, Axams in Tyrol (in the center of the study area, see Figure 8). As for all other stations, daily precipitation observations and numerical weather predictions are available for the month of July from 1985 through 2012. In Figure 3 in the introduction the probabilistic forecasts from the distributional forest, trained on 1985–2008, for July 24 in 2009–2012 have already been shown as a motivational example. The figure shows that the model properly depicts the

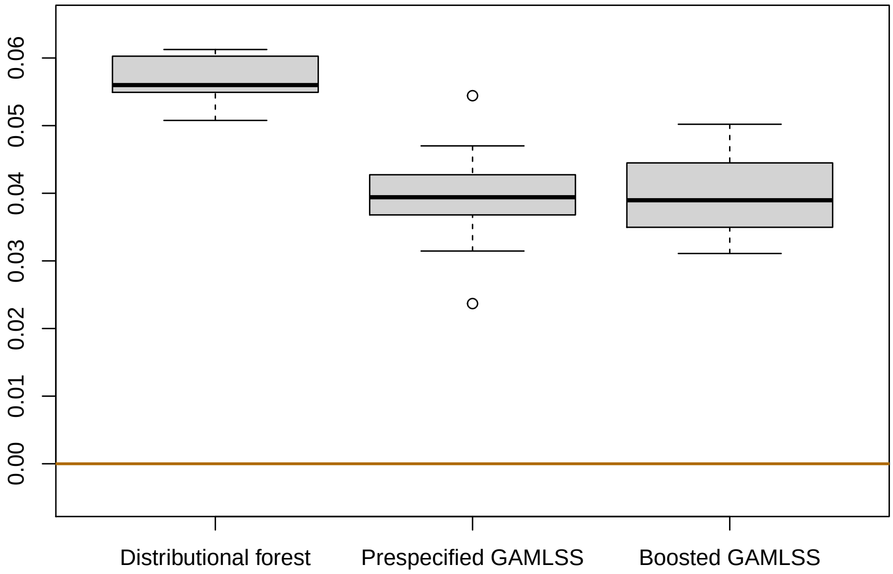

Figure 5: CRPS skill score from the 10 times 7-fold cross validation at station Axams (1985–2012). The horizontal orange line pertains to the reference model EMOS.

point masses at zero (i.e., the probability of a dry day) and the forecasted probability density function for the total amount of precipitation. The four sample forecasts differ considerably in location μ, scale σ, and the amount of censoring while conforming quite well with the actual observations from these days. While this is a nice illustrative example we are interested in the overall predictive performance and calibration of the distributional fits.

To assess calibration Figure 4 shows residual QQ plots for out-of-sample predictions (2009–2012) from the different models trained on 1985–2008. Due to the point masses at zero 100 draws from the randomized quantile residuals (Dunn and Smyth 1996) are plotted in semi-transparent gray. Overall, the randomized quantile residuals conform quite well with the theoretical standard normal quantile (i.e., form a straight line close to the diagonal), indicating that all four models are sufficiently well calibrated. This is also supported by the corresponding probability integral transform (PIT, Gneiting, Balabdaoui, and Raftery 2007) histograms in Appendix B.

To assess the predictive performance, a full cross-validation is carried out rather than relying on just the one fixed test set for the years 2009–2012. To do so, a 10 times 7-fold cross-validation is carried out where each splits the available 28 years into 7 subsets of 4 randomly selected years. The models are learned on 6 folds (= 24 years) and evaluated on the 7-th fold (= 4 years) using the average CRPS across all observations. The resulting 10 CRPS skill scores are displayed by boxplots in Figure 5 using EMOS as the reference model (horizontal line at a CRPSS of 0). Both GAMLSS models and the distributional forest perform distinctly better than the EMOS model. While the two GAMLSS lead to an improvement of around 4 percent, the distributional forest has a slightly higher improvement of around 5.5 percent in median.

Finally, it is of interest how this improvement in predictive performance by the distributional forest is accomplished, i.e., which of the 80 covariates are selected in the trees of the forest.

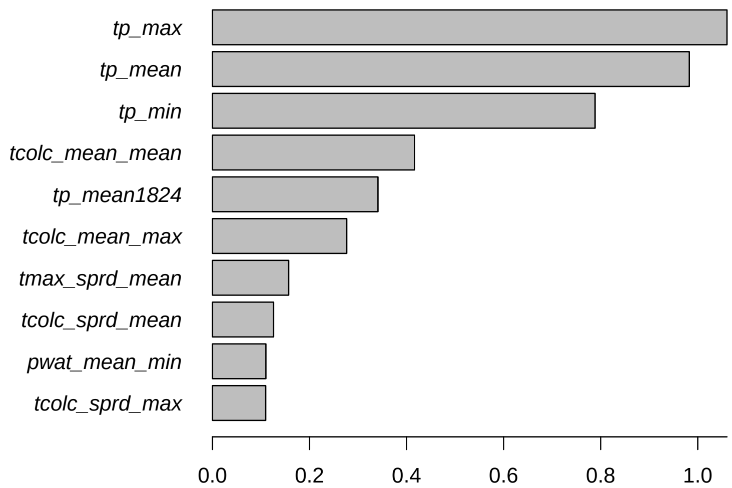

Figure 6: CRPS-based variable importance for the top 10 covariates in the distributional forest. Based on data for station Axams, learning period 1985–2008 and assessed in 2009–2012.

As the 100 trees of the forest do not allow to simply assess the variables' role graphically, a common solution for random forests in general is to consider variable importance measures. Here, this is defined as the amount of change in CRPS when the association between one covariate and the response variable is artificially broken through permutation (and thus also breaking the association to the remaining covariates).

Figure 6 shows the 10 covariates with the highest permutation importance (i.e., change in CRPS) for station Axams. As expected the NWP outputs for total precipitation (*tp*) are particularly important along with total column-integrated condensate (*tcolc*). Also, both variables occur in various transformations such as means (either of the full day or certain parts of the afternoon), spreads, or minima/maxima. Thus, while the covariates themselves are not surprising, selecting a GAMLSS with a particular combination of all the transformations would be much more challenging.

3.4. Application for all stations

After considering only one observational site up to now, a second step evaluates and compares the competing methods on all 95 available stations. As in the previous section, all models are learned on the first 24 years and evaluated by the average CRPS on the last 4 years. More specifically, the CRPS skill score against the EMOS model is computed for the out-of-sample predictions at each station and visualized by parallel coordinates plots with boxplots superimposed in Figure 7. Overall, distributional forests have a slightly higher improvement in CRPSS compared to the two GAMLSS which is best seen by looking at the boxplots and the green line representing the results for station Axams. The underlying parallel coordinates additionally bring out that the prespecified GAMLSS sometimes performs rather differently (sometimes better, sometimes worse) compared to the two data-driven models. Values below zero show that, for some stations, EMOS performs better than the more complex statistical methods.

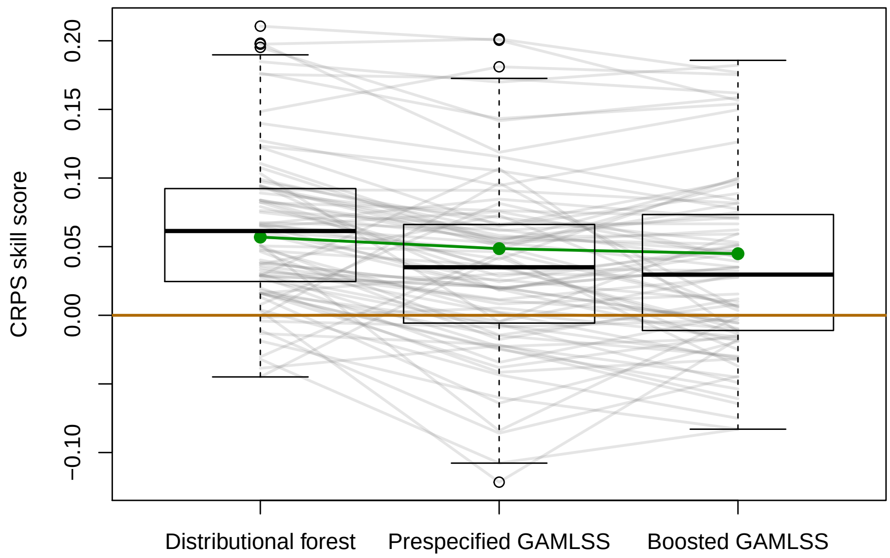

Figure 7: CRPS skill score for each station (gray lines with boxplots superimposed). Station Axams is highlighted in green and the horizontal orange line pertains to the reference model EMOS. The models are learned on 1985–2008 and validated for 2009–2012.

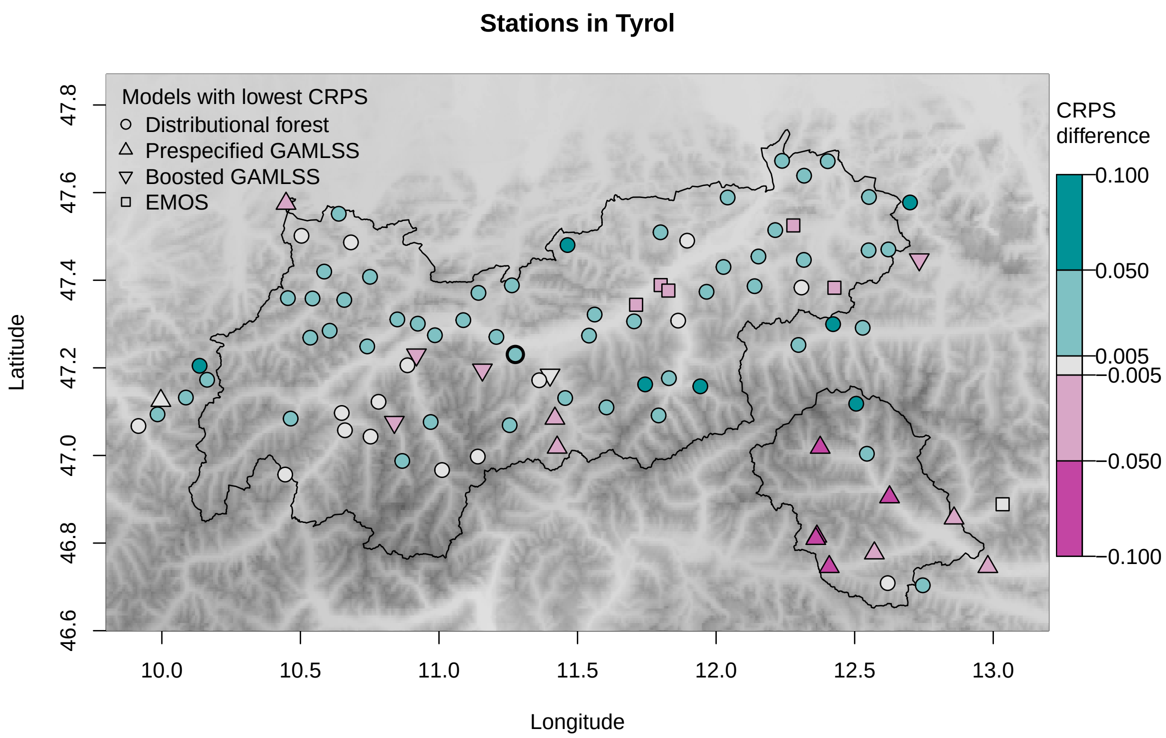

Figure 8: Map of Tyrol coding the best-performing model for each station (type of symbol). The color codes whether the distributional forest had higher (green) or lower (red) CRPS compared to the best of the other three models. The background shows the local topography (Robinson *et al.* 2014). Station Axams is highlighted in bold.

To assess whether these differences in predictive performance are due to differences in the topography, Figure 8 shows a brief spatial summary of all stations. Each station is illustrated by a symbol that conveys which model performed best in terms of CRPS on the last 4 years of the data. Additionally, the color of the symbol indicates the CRPS difference between distributional forest and the best-performing other model. Green signals that the distributional forest performs better than the other models whereas red signals that another model performs better. Overall the distributional forest performs on par (gray) or better (green) for the majority of stations. Only for a few stations in the north-east EMOS performs best, and in East Tyrol the prespecified GAMLSS performs particularly well.

4. Discussion

Distributional regression modeling is combined with tree-based modeling to obtain a novel and flexible method for probabilistic forecasting. The resulting distributional trees and forests can capture abrupt and nonlinear effects and interactions in a data-driven way. By basing the split point and split variable selection on a full likelihood and corresponding score function, the trees and forests can not only pick up changes in the location but also the scale or shape of any distributional family.

Distributional forests are an attractive alternative when prespecifying or boosting all possible effects and interactions in a GAMLSS model is challenging. Distributional forests are rather straightforward to specify requiring only little prior subject matter knowledge and also work well in the presence of many potential covariates. The application to precipitation forecasting in complex terrain illustrates that distributional forests often perform on par or even better than their GAMLSS counterparts. Hence, they form a useful addition to the already available toolbox of probabilistic forecasts for disciplines such as meteorology.

Computational details

The proposed methods are in the R package **disttree** (version 0.1.0) based on the **partykit** package (version 1.2.1), both available on R-Forge at (`https://R-Forge.R-project.org/projects/partykit/`). The function `distfit` fits distributional models by maximum likelihood, which is used as the basis for the tree-building function `disttree`, upon which the `distforest` is built. All functions can either be used with GAMLSS family objects from the R package **gamlss.dist** (Stasinopoulos and Rigby 2007, version 5.0.3) or with custom lists containing all required information about the distribution family.

In addition to **disttree**, Section 3 employs R package **crch** (Messner, Mayr, and Zeileis 2016, version 1.0.0) for the EMOS models, **gamlss** (Stasinopoulos and Rigby 2007, version 5.0.5) for the prespecified GAMLSS, and **gamboostLSS** (Hofner *et al.* 2016, version 2.0.0) for the boosted GAMLSS.

The fitted distributional forest for July 24 and observation station Axams (including Figure 3) is reproducible using `demo("RainAxams", package = "disttree")`. This also includes fitting the other zero-censored Gaussian models considered in this paper and generating the corresponding QQ plots (Figure 4) and PIT histograms (Figure 9). Full replication of all results can be obtained with `demo("RainTyrol", package = "disttree")` requiring the companion R package **RainTyrol** (version 0.1.0), also available within the R-Forge project.

References

Athey S, Tibshirani J, Wager S (2017). "Generalized Random Forests." *arXiv:1610.01271*, arXiv.org E-Print Archive. URL https://arxiv.org/abs/1610.01271.

Bauer P, Thorpe A, Brunet G (2015). "The Quiet Revolution of Numerical Weather Prediction." *Nature*, **525**(7567), 47–55. doi:10.1038/nature14956.

BMLFUW (2016). "Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW), Abteilung IV/4 – Wasserhaushalt." Available at http://ehyd.gv.at/. Accessed: 2016–02–29.

Box GEP, Cox DR (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society B*, **26**(2), 211–252.

Breiman L (2001). "Random Forests." *Machine Learning*, **45**(1), 5–32. doi:10.1023/a:1010933404324.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984). *Classification and Regression Trees*. Wadsworth, California.

Dunn PK, Smyth GK (1996). "Randomized Quantile Residuals." *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. doi:10.2307/1390802.

Gebetsberger M, Messner JW, Mayr GJ, Zeileis A (2017). "Fine-Tuning Non-Homogeneous Regression for Probabilistic Precipitation Forecasts: Unanimous Predictions, Heavy Tails, and Link Functions." *Monthly Weather Review*, **145**(11), 4693–4708. doi:10.1175/MWR-D-16-0388.1.

Glahn HR, Lowry DA (1972). "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting." *Journal of Applied Meteorology*, **11**(8), 1203–1211. doi:10.1175/1520-0450(1972)011<1203:tuomos>2.0.co;2.

Gneiting T, Balabdaoui F, Raftery AE (2007). "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society B*, **69**(2), 243–268. doi:10.1111/j.1467-9868.2007.00587.x.

Gneiting T, Raftery AE (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association*, **102**(477), 359–378. doi:10.1198/016214506000001437.

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**(5), 1098–1118. doi:10.1175/mwr2904.1.

Hamill TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau Jr TJ, Zhu Y, Lapenta W (2013). "NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset." *Bulletin of the American Meteorological Society*, **94**(10), 1553–1565. doi:10.1175/bams-d-12-00014.1.

Hastie T, Tibshirani R (1986). "Generalized Additive Models." *Statistical Science*, **1**(3), 297–310.

Hersbach H (2000). "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting*, **15**(5), 559–570. doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hofner B, Mayr A, Schmid M (2016). "**gamboostLSS**: An R Package for Model Building and Variable Selection in the GAMLSS Framework." *Journal of Statistical Software*, **74**(1), 1–31. doi:10.18637/jss.v074.i01.

Hothorn T, Hornik K, Van de Wiel MA, Zeileis A (2006a). "A Lego System for Conditional Inference." *The American Statistician*, **60**(3), 257–263. doi:10.1198/000313006X118430.

Hothorn T, Hornik K, Zeileis A (2006b). "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, **15**(3), 651–674. doi:10.1198/106186006x133933.

Hothorn T, Lausen B, Benner A, Radespiel-Tröger M (2004). "Bagging Survival Trees." *Statistics in Medicine*, **23**, 77–91. doi:10.1002/sim.1593.

Hothorn T, Zeileis A (2017). "Transformation Forests." *arXiv 1701.02110*, arXiv.org E-Print Archive. URL http://arxiv.org/abs/1701.02110.

Hutchinson MF (1998). "Interpolation of Rainfall Data with Thin Plate Smoothing Splines – Part II: Analysis of Topographic Dependence." *Journal of Geographic Information and Decision Analysis*, **2**(2), 152–167.

Klein N, Kneib T, Lang S, Sohn A (2015). "Bayesian Structured Additive Distributional Regression with an Application to Regional Income Inequality in Germany." *The Annals of Applied Statistics*, **9**(2), 1024–1052. doi:10.1214/15-aoas823.

Lin Y, Jeon Y (2006). "Random Forests and Adaptive Nearest Neighbors." *Journal of the American Statistical Association*, **101**(474), 578–590. doi:10.1198/016214505000001230.

Long JS (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks.

Meinshausen N (2006). "Quantile Regression Forests." *Journal of Machine Learning Research*, **7**, 983–999. doi:10.1109/icmlc.2014.7009082.

Messner JW, Mayr GJ, Zeileis A (2016). "Heteroscedastic Censored and Truncated Regression with **crch**." *The R Journal*, **8**(1), 173–181. doi:https://journal.R-project.org/archive/2016-1/messner-mayr-zeileis.pdf.

Messner JW, Mayr GJ, Zeileis A (2017). "Non-Homogeneous Boosting for Predictor Selection in Ensemble Post-Processing." *Monthly Weather Review*, **145**(1), 137–147. doi:10.1175/MWR-D-16-0088.1.

Nelder JA, Wedderburn RWM (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society A*, **135**(3), 370–384. doi:10.2307/2344614.

Rigby RA, Stasinopoulos DM (2005). "Generalized Additive Models for Location Scale and Shape (with Discussion)." *Journal of the Royal Statistical Society C*, **54**, 507–554. doi:10.1111/j.1467-9876.2005.00510.x.

Robinson N, Regetz J, Guralnick RP (2014). "EarthEnv-DEM90: A Nearly-Global, Void-Free, Multi-Scale Smoothed, 90m Digital Elevation Model From Fused ASTER and SRTM Data." *ISPRS Journal of Photogrammetry and Remote Sensing*, **87**, 57–67. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2013.11.002.

Scheuerer M, Hamill TM (2015). "Statistical Post-Processing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions." *Monthly Weather Review*, **143**(11), 4578–4596. doi:10.1175/MWR-D-15-0061.1.

Stasinopoulos DM, Rigby RA (2007). "Generalized Additive Models for Location Scale and Shape (GAMLSS) in R." *Journal of Statistical Software*, **23**(7), 1–46. doi:10.18637/jss.v023.i07.

Stauffer R, Mayr GJ, Messner JW, Umlauf N, Zeileis A (2017a). "Spatio-Temporal Precipitation Climatology over Complex Terrain Using a Censored Additive Regression Model." *International Journal of Climatology*, **37**(7), 3264–3275. doi:10.1002/joc.4913.

Stauffer R, Umlauf N, Messner JW, Mayr GJ, Zeileis A (2017b). "Ensemble Post-Processing of Daily Precipitation Sums over Complex Terrain Using Censored High-Resolution Standardized Anomalies." *Monthly Weather Review*, **45**(3), 955–969. doi:10.1175/MWR-D-16-0260.1.

Stidd CK (1973). "Estimating the Precipitation Climate." *Water Resources Research*, **9**(5), 1235–1241. doi:10.1029/wr009i005p01235.

Strasser H, Weber C (1999). "On the Asymptotic Theory of Permutation Statistics." *Mathematical Methods of Statistics*, **8**, 220–250.

Zeileis A, Hornik K (2007). "Generalized M-Fluctuation Tests for Parameter Instability." *Statistica Neerlandica*, **61**(4), 488–508. doi:10.1111/j.1467-9574.2007.00371.x.

Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, **17**(2), 492–514. doi:10.1198/106186008x319331.

A. Tree algorithm

In the following, the tree algorithm applied in the empirical case study discussed in this paper is explained. For notational simplicity, the testing and splitting procedure is described for the root node, i.e., the entire learning sample with observations $\{y_i\}_{i=1,\dots,n}$, $n \in \mathbb{N}$. In each child node the corresponding subsample depends on the foregoing split(s).

After fitting a distributional model $\mathcal{D}(Y, \boldsymbol{\theta})$ to the learning sample with observations $\{y_i\}_{i=1,\dots,n}$ as explained in Section 2.1 the resulting estimated parameter $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$, $k \in \mathbb{N}$ can be plugged in the score function $s(\boldsymbol{\theta}, Y)$. In that way a goodness-of-fit measurement is obtained for each parameter θ_j and each observation y_i. To use this information, statistical tests are employed to detect dependencies between the score values

$$s(\hat{\boldsymbol{\theta}}, y) = \begin{pmatrix} s(\hat{\boldsymbol{\theta}}, y_1)_1 & s(\hat{\boldsymbol{\theta}}, y_1)_2 & \dots & s(\hat{\boldsymbol{\theta}}, y_1)_k \\ \vdots & \vdots & \ddots & \vdots \\ s(\hat{\boldsymbol{\theta}}, y_n)_1 & s(\hat{\boldsymbol{\theta}}, y_n)_2 & \dots & s(\hat{\boldsymbol{\theta}}, y_n)_k \end{pmatrix}$$

and each variable $Z_l \in \{Z_1, \dots, Z_m\}$.

$$H_0^l : s(\hat{\boldsymbol{\theta}}, Y) \qquad \perp \qquad Z_l$$

These hypotheses are assessed with permutation tests based on the multivariate linear statistic

$$T_l = \sum_{i=1}^{n} v_l(z_{li}) \cdot s(\hat{\boldsymbol{\theta}}, y_i).$$

The type of the transformation function v_l depends on the type of the split variable Z_l. If Z_l is numeric then v_l is simply the identity function $v_l(z_{li}) = z_{li}$. If Z_l is a categorical variable with H categories then $v_l(z_{li}) = e_H(z_{li}) = (\mathbf{I}(z_{ji} = 1), \dots, \mathbf{I}(z_{li} = H))$ such that v_l is a unit vector where the element corresponding to the value of z_{li} is 1. Observations with missing values are excluded from the sums.

With the conditional expectation μ_l and the covariance Σ_l as derived by Strasser and Weber (1999) the test statistic can be standardized. The observed multivariate linear statistic t is mapped onto the real line by a univariate test statistic c. In the application of this paper a quadratic form is chosen, such that

$$c_{\text{quad}}(t, \mu, \Sigma) = (t - \mu)\Sigma^{+}(t - \mu)^{\top}$$

where Σ^{+} is the Moore-Penrose inverse of Σ. Alternatively, the maximum of the absolute values of the standardized linear statistic can be considered.

The smaller the p-value corresponding to the standardized test statistic $c(t_l, \mu_l, \Sigma_l)$ is the stronger the discrepancy from the assumption of independence between the scores and the split variable Z_l. After Bonferroni-adjusting the p-values it has to be assessed whether any of the resulting p-values is beneath the selected significance level. If so, the partitioning variable Z_{l^*} with the lowest p-value is chosen as splitting variable.

The breakpoint that leads to the highest discrepancy between score functions in the two resulting subgroups is selected as split point. This is measured by the linear statistic

$$T_{l^*}^{r} = \sum_{i \in \mathcal{B}_{1r}} s(\hat{\boldsymbol{\theta}}, y_i)$$

where $\mathcal{B}_{1r}$ is the first of the two new subgroups that are defined by splitting in split point r of variable Z_{l^*}. The split point is then chosen as follows:

$$r^* = \underset{r}{\operatorname{argmin}}\, c(t^r_{l^*}, \mu^r_{j^*}, \Sigma^r_{l^*}).$$

One repeats the testing and splitting procedure in each of the resulting subgroups until some stopping criterion is reached. This criterion can for example be a minimal number of observations in a node or a minimal p-value for the statistical tests. In that way pre-pruning is applied in order to find right-sized trees and hence avoid overfitting.

This permutation test based tree algorithm is presented in Hothorn *et al.* (2006b) as the CTree algorithm. A different framework to build a tree is provided by the MOB algorithm which is based on M-fluctuation tests (Zeileis *et al.* 2008).

B. PIT histograms

As an alternative to QQ plots based on randomized quantile residuals, probability integral transform (PIT, Gneiting *et al.* 2007) histograms are a commonly used tool to assess how well-calibrated distributional fits are. Figure 9 shows out-of-sample PIT histograms for the same probabilistic forecasts as Figure 4, confirming that despite some deviations the overall calibration of all models is acceptable.

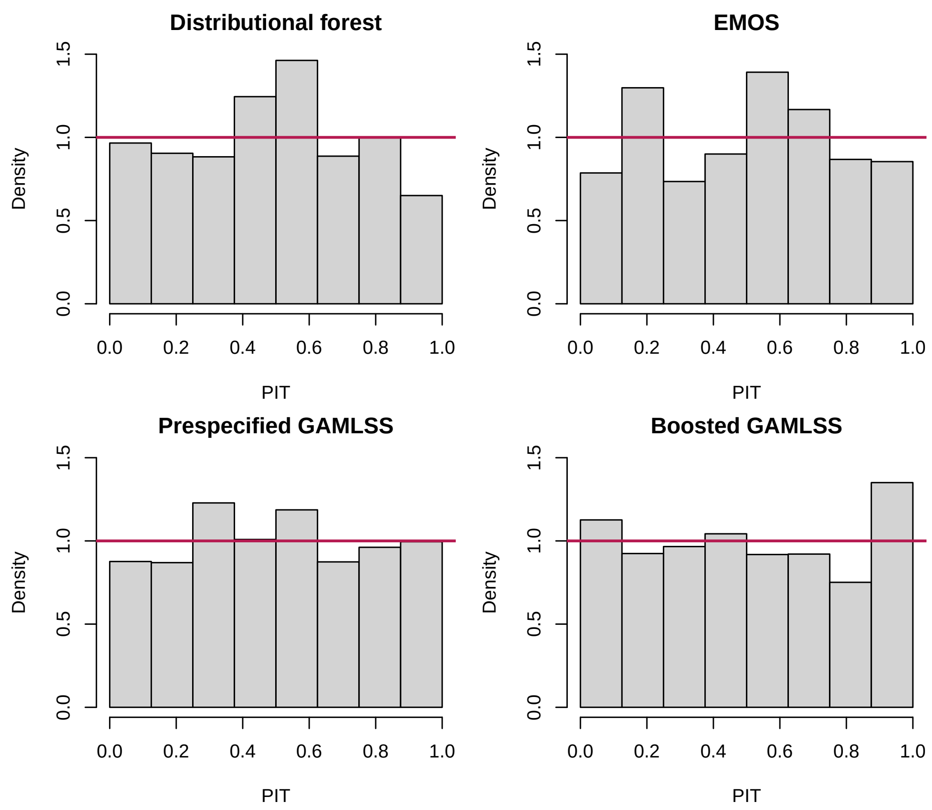

Figure 9: Out-of-sample PIT histograms (2009–2012) for station Axams and models learned on data from 1985–2008.

Affiliation:

Lisa Schlosser, Reto Stauffer, Achim Zeileis
Universität Innsbruck
Department of Statistics
Faculty of Economics and Statistics
Universitätsstr. 15
6020 Innsbruck, Austria
E-mail: Lisa.Schlosser@uibk.ac.at, Reto.Stauffer@uibk.ac.at,
Achim.Zeileis@R-project.org
URL: https://www.uibk.ac.at/statistics/personal/schlosser-lisa/,
https://retostauffer.org/,
https://eeecon.uibk.ac.at/~zeileis/

Torsten Hothorn
Universität Zürich
Institut für Epidemiologie, Biostatistik und Prävention
Hirschengraben 84
CH-8001 Zürich, Switzerland
E-mail: Torsten.Hothorn@R-project.org
URL: http://user.math.uzh.ch/hothorn/

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

https://www.uibk.ac.at/eeecon/wopec/

2018-08 **Lisa Schlosser, Torsten Hothorn, Reto Stauffer, Achim Zeileis:** Distributional regression forests for probabilistic precipitation forecasting in complex terrain

2018-07 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Delegated decision making and social competition in the finance industry

2018-06 **Manuel Gebetsberger, Reto Stauffer, Georg J. Mayr, Achim Zeileis:** Skewed logistic distribution for statistical temperature post-processing in mountainous areas

2018-05 **Reto Stauffer, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Hourly probabilistic snow forecasts over complex terrain: A hybrid ensemble postprocessing approach

2018-04 **Utz Weitzel, Christoph Huber, Florian Lindner, Jürgen Huber, Julia Rose, Michael Kirchler:** Bubbles and financial professionals

2018-03 **Carolin Strobl, Julia Kopf, Raphael Hartmann, Achim Zeileis:** Anchor point selection: An approach for anchoring without anchor items

2018-02 **Michael Greinecker, Christopher Kah:** Pairwise stable matching in large economies

2018-01 **Max Breitenlechner, Johann Scharler:** How does monetary policy influence bank lending? Evidence from the market for banks' wholesale funding

2017-27 **Kenneth Harttgen, Stefan Lang, Johannes Seiler:** Selective mortality and undernutrition in low- and middle-income countries

2017-26 **Jun Honda, Roman Inderst:** Nonlinear incentives and advisor bias

2017-25 **Thorsten Simon, Peter Fabsic, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Probabilistic forecasting of thunderstorms in the Eastern Alps

2017-24 **Florian Lindner:** Choking under pressure of top performers: Evidence from biathlon competitions

2017-23 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Estimation methods for non-homogeneous regression models: Minimum continuous ranked probability score vs. maximum likelihood

2017-22 **Sebastian J. Dietz, Philipp Kneringer, Georg J. Mayr, Achim Zeileis:** Forecasting low-visibility procedure states with tree-based statistical methods

2017-21 **Philipp Kneringer, Sebastian J. Dietz, Georg J. Mayr, Achim Zeileis:** Probabilistic nowcasting of low-visibility procedure states at Vienna International Airport during cold season

2017-20 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior

2017-19 **Martin Geiger, Richard Hule:** The role of correlation in two-asset games: Some experimental evidence

2017-18 **Rudolf Kerschbamer, Daniel Neururer, Alexander Gruber:** Do the altruists lie less?

2017-17 **Meike Köhler, Nikolaus Umlauf, Sonja Greven:** Nonlinear association structures in flexible Bayesian additive joint models

2017-16 **Rudolf Kerschbamer, Daniel Muller:** Social preferences and political attitudes: An online experiment on a large heterogeneous sample

2017-15 **Kenneth Harttgen, Stefan Lang, Judith Santer, Johannes Seiler:** Modeling under-5 mortality through multilevel structured additive regression with varying coefficients for Asia and Sub-Saharan Africa

2017-14 **Christoph Eder, Martin Halla:** Economic origins of cultural norms: The case of animal husbandry and bastardy

2017-13 **Thomas Kneib, Nikolaus Umlauf:** A primer on bayesian distributional regression

2017-12 **Susanne Berger, Nathaniel Graham, Achim Zeileis:** Various versatile variances: An object-oriented implementation of clustered covariances in R

2017-11 **Natalia Danzer, Martin Halla, Nicole Schneeweis, Martina Zweimüller:** Parental leave, (in)formal childcare and long-term child outcomes

2017-10 **Daniel Muller, Sander Renes:** Fairness views and political preferences - Evidence from a large online experiment

2017-09 **Andreas Exenberger:** The logic of inequality extraction: An application to Gini and top incomes data

2017-08 **Sibylle Puntscher, Duc Tran Huy, Janette Walde, Ulrike Tappeiner, Gottfried Tappeiner:** The acceptance of a protected area and the benefits of sustainable tourism: In search of the weak link in their relationship

2017-07 **Helena Fornwagner:** Incentives to lose revisited: The NHL and its tournament incentives

2017-06 **Loukas Balafoutas, Simon Czermak, Marc Eulerich, Helena Fornwagner:** Incentives for dishonesty: An experimental study with internal auditors

2017-05 **Nikolaus Umlauf, Nadja Klein, Achim Zeileis:** BAMLSS: Bayesian additive models for location, scale and shape (and beyond)

2017-04 **Martin Halla, Susanne Pech, Martina Zweimüller:** The effect of statutory sick-pay on workers' labor supply and subsequent health

2017-03 **Franz Buscha, Daniel Müller, Lionel Page:** Can a common currency foster a shared social identity across different nations? The case of the Euro.

2017-02 **Daniel Müller:** The anatomy of distributional preferences with group identity

2017-01 **Wolfgang Frimmel, Martin Halla, Jörg Paetzold:** The intergenerational causal effect of tax evasion: Evidence from the commuter tax allowance in Austria

University of Innsbruck

Working Papers in Economics and Statistics

2018-08

Lisa Schlosser, Torsten Hothorn, Reto Stauffer, Achim Zeileis

Distributional regression forests for probabilistic precipitation forecasting in complex terrain

Abstract
To obtain a probabilistic model for a dependent variable based on some set of explanatory variables, a distributional approach is often adopted where the parameters of the distribution are linked to regressors. In many classical models this only captures the location of the distribution but over the last decade there has been increasing interest in distributional regression approaches modeling all parameters including location, scale, and shape. Notably, so-called non-homogenous Gaussian regression (NGR) models both mean and variance of a Gaussian response and is particularly popular in weather forecasting. More generally, the GAMLSS framework allows to establish generalized additive models for location, scale, and shape with smooth linear or nonlinear effects. However, when variable selection is required and/or there are non-smooth dependencies or interactions (especially unknown or of high-order), it is challenging to establish a good GAMLSS. A natural alternative in these situations would be the application of regression trees or random forests but, so far, no general distributional framework is available for these. Therefore, a framework for distributional regression trees and forests is proposed that blends regression trees and random forests with classical distributions from the GAMLSS framework as well as their censored or truncated counterparts. To illustrate these novel approaches in practice, they are employed to obtain probabilistic precipitation forecasts at numerous sites in a mountainous region (Tyrol, Austria) based on a large number of numerical weather prediction quantities. It is shown that the novel distributional regression forests automatically select variables and interactions, performing on par or often even better than GAMLSS specified either through prior meteorological knowledge or a computationally more demanding boosting approach.

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