

Monotonicity-constrained species distribution models

BENJAMIN HOFNER,¹ JÖRG MÜLLER,^{2,3} AND TORSTEN HOTHORN^{4,5}

¹Institut für Medizininformatik, Biometrie und Epidemiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg,
Waldstraße 6, 91054 Erlangen, Germany

²Nationalparkverwaltung Bayerischer Wald, Freyunger Straße 2, 94481 Grafenau, Germany

³Chair for Terrestrial Ecology, Department of Ecology and Ecosystem Management, Technische Universität München,
Hans-Carl-von-Carlowitz-Platz 2, 85350 Freising-Weihenstephan, Germany

⁴Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539 München, Germany

Abstract. Flexible modeling frameworks for species distribution models based on generalized additive models that allow for smooth, nonlinear effects and interactions are of increasing importance in ecology. Commonly, the flexibility of such smooth function estimates is controlled by means of penalized estimation procedures. However, the actual shape remains unspecified. In many applications, this is not desirable as researchers have a priori assumptions on the shape of the estimated effects, with monotonicity being the most important. Here we demonstrate how monotonicity constraints can be incorporated in a recently proposed flexible framework for species distribution models. Our proposal allows monotonicity constraints to be imposed on smooth effects and on ordinal, categorical variables using an additional asymmetric L_2 penalty. Model estimation and variable selection for Red Kite (*Milvus milvus*) breeding was conducted using the flexible boosting framework implemented in R package **mboost**.

Key words: *boosting; Milvus milvus; monotonic P-splines; ordinal variables; Red Kite breeding distribution, Bavaria, Germany; spatial autocorrelation; species distribution model; structured additive model; variable selection.*

INTRODUCTION

Species distribution models estimate the regression relationship between the environment and the distribution of a species. The environment is characterized by a potentially large number of variables that can be used to model occurrence or abundance of species and typically include climatic variables or variables characterizing the habitat. While simple models, e.g., an ordinary linear model, appeal to researchers and practitioners, more complex models, e.g., regression trees, may uncover structures that simpler models will miss (Elith and Graham 2009). Recently, Hothorn et al. (2011) developed a framework that allows the influence of the environment on species distributions to be described in a very flexible manner (i.e., as a combination of linear effects, nonlinear effects, spatial effects, and so forth) while maintaining interpretability of the regression effects. Within this framework, ecologists can specify local components (such as spatial autocorrelation, spatiotemporal effects, and nonstationary effects) along with global model components. The global model component describes the effects of the environmental variables as a sum of smooth functions.

Manuscript received 30 November 2010; revised 18 April 2011; accepted 8 May 2011. Corresponding Editor: J. Elith.

⁵ Corresponding author. E-mail: Torsten.Hothorn@stat.uni-muenchen.de

In many situations, researchers have prior knowledge of or assumptions about the shape of these smooth effects. A very strong assumption is the common linearity assumption. This assumption can be relaxed if one switches to (generalized) additive models (GAM) or structured additive regression (STAR) models. The latter extend GAMs so that a mixture of effects might be used, as, for example, linear effects, smooth effects, spatial effects, random effects, and varying coefficient terms. The linearity assumption implicitly imposes a monotonicity assumption, but without restrictions on the direction of monotonicity, i.e., monotonically increasing or decreasing. It is often more appropriate to relax the linearity assumption but to maintain the monotonicity assumption. This is the case in our application, namely modeling the breeding distribution of the Red Kite (*Milvus milvus*) in the German federal state of Bavaria. Hothorn et al. (2011) used a very flexible STAR model for describing this breeding distribution. In certain aspects, the resulting model leads to biologically implausible effects, such as non-monotonic estimates for certain environmental variables that are difficult to justify and might be due to artifacts in the data. Particularly at the extremes of environmental gradients, observational data, e.g., in atlases, often represent low numbers of observations (e.g., very high or very low elevation). This can lead to artificial bumps in the estimated regression curves that lack any biological background. Also a low amount of information in the middle of an environmental gradient may

induce such effects. Here, we extend the modeling framework suggested in an earlier work of our research group (Hothorn et al. 2011) and allow for monotonicity constraints to be incorporated on the smooth additive regression curves in the global model component. In doing so, we add a Bayesian flavor to species distribution models while technically adhering to a frequentist approach. In this sense, allowing for monotonicity constraints builds a bridge between the two worlds of statistics.

In the following section we provide a short, practical introduction to the model-estimation algorithm used here, namely boosting. We then introduce a novel method to estimate monotonicity-constrained, smooth effects in boosting for species distribution models. Subsequently, we demonstrate how this idea can also be applied to estimate monotonic effects for ordinal environmental variables. In *Methods: Red Kite breeding distribution* (below) we describe the species distribution model for Red Kite breeding in detail. The improved model results obtained under monotonicity of certain effects are discussed in *Results* (below).

METHODS

Boosting species distribution models

We assume that the probability of confirmed Red Kite breeding at some point s (here given as longitude and latitude in Bavaria, Germany) and time t under environmental conditions as characterized by environmental variables $\mathbf{x} = (x^{(1)}, \dots, x^{(p)})$ is given by the inverse logistic transformation of the regression function f to be estimated:

$$\begin{aligned} & \text{Prob(presence of Red Kite breeding} | \mathbf{x}, s, t) \\ &= \text{logit}^{-1}[f(\mathbf{x}, s, t)]. \end{aligned} \quad (1)$$

Hothorn et al. (2011) decompose the contributions of environmental variables, space, and time via a structured additive regression model (Fahrmeir et al. 2004) in an additive fashion:

$$f(\mathbf{x}, s, t) = \underbrace{f_{\text{env}}(\mathbf{x})}_{\text{global}} + \underbrace{f_{\text{ns}}(\mathbf{x}, s) + f_s(s) + f_{\text{st}}(s, t)}_{\text{local}}. \quad (2)$$

The local model terms f_{ns} , f_s and f_{st} , which capture nonstationary effects, and spatial and spatiotemporal autocorrelation are discussed in detail by Hothorn et al. (2011). Here, we mainly focus on the global model term f_{env} , which describes the influence of environmental variables on Red Kite breeding as the sum of functions in the manner of STAR models, i.e., $f_{\text{env}}(\mathbf{x}) = \sum_{j=1}^p f_j(x^{(j)})$ but take the local components into account in the model. The term $f_j(x^{(j)})$ represents the contribution of the j th environmental variable to the global model component.

Hothorn et al. (2011) propose to fit species distribution models of the form (1) by component-wise functional gradient descent boosting (Bühlmann and

Yu 2003), minimizing the empirical negative log-likelihood of the Bernoulli distribution. The algorithm is especially attractive owing to its intrinsic variable selection properties (Kneib et al. 2009, Hofner et al. 2011). Hothorn et al. (2011) describe the algorithm in more detail in the context of species distribution models. One begins with a constant model $\hat{f} = 0$ and computes the residuals $\mathbf{u} = [u_1, \dots, u_n]^\top$ (defined by the negative gradient of the loss function, which corresponds to the log-likelihood evaluated at each observation; see Bühlmann and Yu 2003, Bühlmann and Hothorn 2007, Hothorn et al. 2011). Each model component f_1, \dots, f_p , f_{ns} , f_s , and f_{st} is fitted separately to the residuals \mathbf{u} by penalized least squares, and only the model component that describes these residuals best is updated by adding a small percentage of the fit (e.g., 10%) to the current model fit. New residuals are computed, and the whole procedure is iterated until a fixed number of iterations is reached. The final model \hat{f} is defined as the sum of all models fitted in this process. As we update only one modeling component in each boosting iteration, variables are selected by stopping the boosting procedure after an appropriate number of iterations. This is usually done using cross-validation techniques.

For each of the model components, a corresponding regression model that is applied to fit the residuals has to be specified, the so-called “baselearner.” The choice of these baselearners is the key to estimate the decomposed species distribution model (2). The baselearners determine which functional form each of the components can take. Therefore, imposing monotonicity constraints on specific baselearners will lead to species distribution models whose estimated effects are monotone. In the following, we concentrate on one baselearner only, i.e., on the procedure for fitting the contribution of one single environmental variable. For the sake of simplicity, we drop the subscripts $j = 1, \dots, p$ and refer to the observations of variable x as $[x_1, \dots, x_n]^\top$, with corresponding model component $f(x)$ in the following section. Our primary aim is to introduce constraints in the fitting algorithm such that the resulting estimate is monotone, for example, monotonically increasing: $\hat{f}(x_i) \leq \hat{f}(x_{i'})$ for two observations $x_i \leq x_{i'}$.

Estimating monotonic smooth effects

The approaches to flexible, yet smooth modeling considered in this paper are based on B-splines with difference penalties on adjacent knots (i.e., P-splines; Eilers and Marx 1996). Consider observations $\mathbf{x} = [x_1, \dots, x_n]^\top$ of a single variable x , then a nonlinear function $f(x)$ can be approximated as

$$f(x) = \sum_{m=1}^M \beta_m B_m(x; l) = \mathbf{B}(x)^\top \boldsymbol{\beta}$$

where $\boldsymbol{\beta} = [\beta_1, \dots, \beta_M]^\top$ is the $(M \times 1)$ vector of regression coefficients, $B_m(x; l)$ is the m th B-spline basis function of degree l and $\mathbf{B}(x) = [B_1(x), \dots, B_M(x)]^\top$ is

the $(M \times 1)$ vector of B-splines (for simplicity, we dropped l). The basis functions are defined on a grid of $K = M - (l - 1)$ inner knots ξ_1, \dots, ξ_K with additional boundary knots (and usually a knot expansion in the boundary knots). For more details on the construction of B-splines, we refer to Eilers and Marx (1996). The function estimates can be written in matrix notation as $\hat{f}(\mathbf{x}) = \mathbf{B}\hat{\beta}$, where the $(n \times M)$ design matrix $\mathbf{B} = [\mathbf{B}(x_1), \dots, \mathbf{B}(x_n)]^\top$ comprises the B-spline basis vectors $\mathbf{B}(x)$ evaluated for each observation x_i .

The function estimate might adapt the data too closely and might become too erratic. To enforce smoothness of the function estimate, an additional penalty is used that penalizes large differences of the coefficients of adjacent knots. Hence, for a continuous vector of residuals $\mathbf{u} = [u_1, \dots, u_n]$, we can estimate the function via a penalized least-squares criterion,

$$Q(\beta) = (\mathbf{u} - \mathbf{B}\beta)^\top (\mathbf{u} - \mathbf{B}\beta) + \lambda J(\beta; d)$$

with a quadratic difference penalty on the coefficients:

$$J(\beta; d) = \sum_{m=d+1}^M (\Delta^d \beta_m)^2 = \beta^\top \mathbf{D}_{(d)}^\top \mathbf{D}_{(d)} \beta \quad (3)$$

where d is the order of the difference penalty for P-splines and λ is the smoothing parameter that governs the smoothness. The difference operator Δ is defined such that $\Delta\beta_m = \Delta^1\beta_m = (\beta_m - \beta_{m-1})$, $\Delta^2\beta_m = \Delta(\Delta\beta_m) = \Delta\beta_m - \Delta\beta_{m-1} = \beta_m - 2\beta_{m-1} + \beta_{m-2}$, and so on. The $(M \times M)$ difference matrices $\mathbf{D}_{(c)}$ are constructed such that they lead to the appropriate differences (see, e.g., Eilers and Marx 1996).

To achieve a smooth, yet monotonic function estimate, Eilers (2005) introduced P-splines with an *additional asymmetric difference penalty*. The penalized least squares criterion becomes

$$Q(\beta) = (\mathbf{u} - \mathbf{B}\beta)^\top (\mathbf{u} - \mathbf{B}\beta) + \lambda J(\beta; d) + \lambda_2 J_{\text{asym}}(\beta; c) \quad (4)$$

with asymmetric penalty

$$J_{\text{asym}}(\beta; c) = \sum_{m=c+1}^M v_m (\Delta^c \beta_m)^2 = \beta^\top \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)} \beta \quad (5)$$

where c is the order of the difference penalty. The choice of c implies the type of additional constraints: monotonicity for $c = 1$ or convexity/concavity for $c = 2$. In the remainder of the paper, we restrict our attention to monotonicity constraints; however, one can always think of convex or concave constraints as well. The asymmetric penalty looks very much like the P-spline penalty (Eq. 3) with the important distinction of weights v_m , which are specified as

$$v_m = \begin{cases} 0 & \text{if } \Delta^c \beta_m > 0 \\ 1 & \text{if } \Delta^c \beta_m \leq 0. \end{cases} \quad (6)$$

The weights are collected in the $(M \times M)$ diagonal

matrix $\mathbf{V} = \text{diag}(\mathbf{v})$. With $c = 1$, this enforces monotonically *increasing* functions. Changing the direction of the inequalities in the distinction of cases leads to monotonically *decreasing* functions. As the weights (Eq. 6) depend on the coefficients β , a solution to Eq. 4 can only be found by iteratively minimizing $Q(\beta)$ with respect to β , where the weights \mathbf{v} are updated in each iteration. This is equivalent to repeatedly solving

$$\hat{\beta} = (\mathbf{B}^\top \mathbf{B} + \lambda_1 \mathbf{D}_{(d)}^\top \mathbf{D}_{(d)} + \lambda_2 \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)})^{-1} \mathbf{B}^\top \mathbf{u} \quad (7)$$

with updated weights. The estimation converges if no further changes in the weight matrix \mathbf{V} occur. In our experience, the algorithm converges very quickly within a few steps. The penalty parameter λ_2 associated with the additional constraint (Eq. 5) could be chosen, for example, by cross-validation techniques. However, this would increase the computational costs dramatically, and, more importantly, λ_2 resembles the researcher's a priori belief in monotonicity. Hence, the proposed procedure is located inbetween pure frequentist and Bayesian approaches. Eilers (2005) proposes that the penalty parameter chosen should be quite large (e.g., 10^6), where larger values are associated with a stronger impact of the monotonic constraint on the estimation. In the limit of $\lambda_2 = 0$, no monotonicity constraint is applied to the estimation procedure. For small values of λ_2 , the monotonicity constraint is a suggestion rather than a constraint. However, if the penalty parameter chosen is large enough (as suggested), the actual size is negligible. The effect of the monotonicity constraint can be observed in Fig. 1a, where the blue curve is fitted via P-splines (corresponding to $\lambda_2 = 0$) and the red curve with monotonic P-splines ($\lambda_2 = 10^6$). As the corresponding derivatives (dashed lines) show, the constrained fit is monotonically increasing (derivative always >0), while the unconstrained fit is not monotonic. A detailed simulation study can be found in the Appendix. In short, we simulated artificial data where the true effect of the covariates on the response is monotonic. We could show that effect estimates of unconstrained models regularly violate monotonicity and often result in models with a higher mean squared error compared to monotonicity-constrained models. Hence, in this setting, monotonicity-constrained models are clearly superior.

Estimating monotonic categorical effects

As Eilers' (2005) asymmetric difference penalty relies only on the neighborhood information, i.e., implies an ordering of β_m (for P-splines induced by the knots), we can extend the estimation scheme to *ordered factors* x . The ordering of categories is reflected in an ordering of the effects for the respective categories. Let x be an ordered, categorical variable with categories $1, \dots, k$. We use dummy-coded variables $x^{(2)}, \dots, x^{(k)}$ (the first category is the reference category), where $x^{(m)} = 1$ if $x = m$ and $x^{(m)} = 0$ otherwise. With the intercept $\mathbf{1} = [1, \dots, 1]^\top$, we obtain the dummy-coded $(n \times k)$ design matrix \mathbf{X}

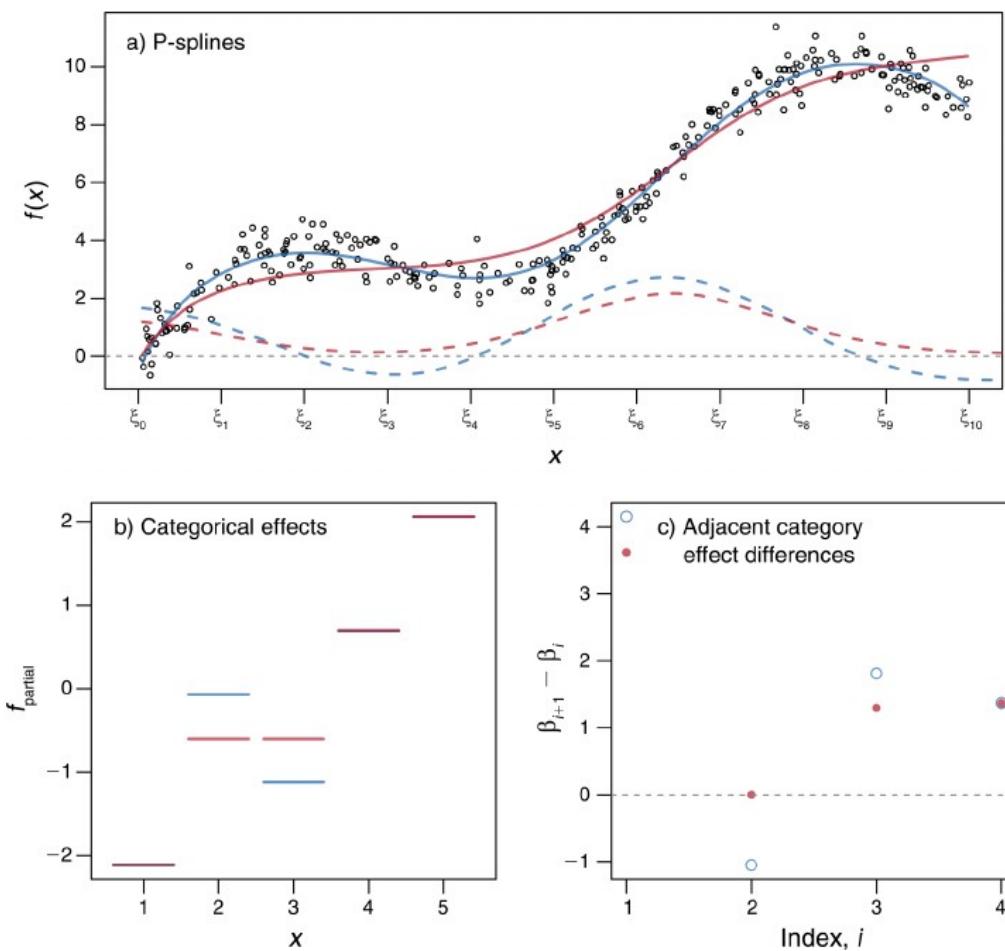


FIG. 1. Monotonically constrained effects. (a) P-splines (blue) and monotonic P-splines (red) with corresponding derivatives (dashed lines), where ξ_1, \dots, ξ_{10} denote the knots of the P-splines. The x -axis is a numeric value; the y -axis denotes the effects. (b) Unconstrained (blue) and constrained (red) estimates for categorical effects. The x -axis is a categorical variable at levels 1, 2, 3, 4, and 5; the y -axis shows the effects of the partial contribution of x . (c) Differences of effects of adjacent categories, where blue open circles represent unconstrained estimates and red solid dots represent constrained estimates.

$= (\mathbf{1}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)})$ of the ordinal variable. The estimation problem can then be written as a penalized least-squares problem:

$$Q(\beta) = (\mathbf{u} - \mathbf{X}\beta)^T(\mathbf{u} - \mathbf{X}\beta) + \lambda_2 J_{\text{asym}}(\beta; c) \quad (8)$$

with a $(k \times 1)$ vector of regression coefficients and an asymmetric penalty defined essentially equal to Eq. 5. The only difference is that the difference matrix \mathbf{D} is replaced by a difference matrix where the element $D_{1,1}$ is set to 0 as the intercept β_1 is not penalized and as $\beta_2 - 0 = \beta_2$ must fulfill the constraint (as the effect of the reference category is set to 0). For $c = 1$, this enforces monotonically increasing or decreasing effect estimates.

Fig. 1b shows the effect of the monotonicity constraint. The effects of the second and third category clearly violate monotonicity, which is eliminated with the monotonicity constraint. For monotonically increasing effects, the differences of adjacent effect estimates should always be positive (cf. derivative for smooth functions). This is enforced by the penalty as Fig. 1c shows.

As in monotonic P-splines, the asymmetric monotonicity penalty can be combined with a smoothness penalty. We can use a difference penalty of order d for coefficients of adjacent categories (see Gertheiss and Tutz 2009, Hofner et al. 2011):

$$J(\beta, d) = \sum_{m=d+1}^M (\Delta^d \beta_m)^2 = \beta^T \mathbf{D}_{(d)}^T \mathbf{D}_{(d)} \beta.$$

This further penalty ensures that the jumps between adjacent categories are small, i.e., neighboring categories are assumed to be similar and thus estimates should not be too erratic. In the combination,

$$Q(\beta) = (\mathbf{u} - \mathbf{X}\beta)^T(\mathbf{u} - \mathbf{X}\beta) + \lambda_1 J(\beta; d) + \lambda_2 J_{\text{asym}}(\beta; c) \quad (9)$$

we get a monotonicity-constrained, “smooth” estimate for the categorical variable. Both estimation criteria (Eqs. 8 and 9) can be minimized by repeatedly solving Eq. 7, where \mathbf{B} is replaced by \mathbf{X} .

Red Kite breeding distribution

We extracted the Red Kite breeding data on grid cells of the topographic map (1:25 000 for Bavaria) with an average cell area of 33.9 km² from the Bavarian breeding atlas in the periods 1979–1983 (Nitsche and Plachter 1987) and 1996–1999 (Bezzel et al. 2005). A minimum of one record in a quadrant was sufficient to define the presence of breeding; all other squares were taken as absence of Red Kite breeding. Environmental variables describing climate and landscape on the grid are given along with summary statistics in the Appendix; for a more detailed description of these data, we refer the reader to Hothorn et al. (2011).

Hothorn et al. (2011) compared species distribution models of different complexities for Red Kite breeding. Their results indicate that a model with smooth additive effects of the environmental variables along with a nonstationary effect of altitude and spatiotemporal autocorrelation fits the data best (model “(add/vary)” in their notation). We used this model as a starting point, i.e., we specified smooth effects for all continuous variables, linear effects for categorical variables, an additional smooth, spatial component to account for spatial correlation (possibly varying for study period), and a spatially varying effect for altitude. Furthermore, we applied monotonic restrictions (all decreasing) on the effects of Coniferous Forest, Cities and Villages, Precipitation of the Wettest Month, and Precipitation of the Wettest Quarter. The effects of these environmental variables as estimated by Hothorn et al. (2011) are difficult to interpret because of local extremes (minima and maxima) and erratic fluctuations. For example, small coniferous forests are sufficient and often used for breeding. The surrounding non-forest land is the major foraging area, and thus its composition is the main limitation factor of a grid cell (Mebs and Schmidt 2006). Thus, it is appropriate to assume a monotonic decreasing influence of coniferous forests on Red Kite breeding. Similarly for climate variables, such as the wettest quarter, one has to expect a monotonic decrease of the influence because this variable determines the density of ground vegetation. With an increasing density caused by increasing precipitation, the availability of the major prey will decrease. We therefore expect a positive effect only with lower levels of precipitation and an increasing negative effect with higher levels. Again, here the observational data underlying the model may lead to artificial bumps in the estimated functions. One should note that it is of particular importance to specify monotonic constraints for both variables of precipitation as these variables are highly correlated and hence share a lot of information. The reasons for our belief in monotonicity are the same for both variables. Furthermore, if one effect remains unconstrained, it may capture some of the erratic behavior left over from the constrained estimate. In this case, the model and its interpretation could be misleading. For the proportion of cities and villages, we also have to expect a

monotonically decreasing effect because the species may forage even in the vicinity of small villages, but will clearly avoid larger cities, possibly because of human disturbance and a decrease in foraging habitats, such as meadows, and nesting habitats such as forests (Mebs and Schmidt 2006).

Here, we estimated two different models: (1) the additive model with monotonicity constraints on the effect of four variables “(mono)” and (2) the unconstrained model “(add/vary).” The latter model is used as a competitor to assess how monotonicity constraints of some variables affect the estimates of these and other effects. We selected variables intrinsically in the applied component-wise boosting algorithm. The optimal model, i.e., the optimal number of boosting iterations, was estimated by the bootstrap separately for (mono) and (add/vary). More details on both models are given in the Appendix.

RESULTS

The resulting effect estimates of the two models are depicted in Fig. 2. The upper row depicts the effects estimated with monotonicity constraints. We can conclude from model (mono) that Red Kites prefer dry areas with broad-leaf forests or mixed forests and avoid especially coniferous forests for breeding. Areas with a mixture of meadows and trees are more likely to be used as Red Kite breeding habitats as indicated by our model. At the same time, we found a large spatial autocorrelation in Red Kite breeding that cannot be attributed to the measured variables. Moreover, a spatially varying effect for altitude was identified, along with a change in the spatial pattern over the observed time periods (see Appendix for details).

Comparing the two models (see Fig. 2), we can conclude that the effect estimate of Coniferous Forests shows a decreasing trend for model (add/vary); however, model (mono) shows an even smoother, monotonic fit and keeps the main functional form. The ordinal variable Cities and Villages, which represents the population density, had a monotonically decreasing effect in the original model. To keep this effect, we specified a monotonic constraint. In this case, the effect was still monotonically decreasing, and thus the additional penalization hardly affected the estimation. In the case of Precipitation of Wettest Month, we observed a small, rather erratic effect in model (add/vary), which completely vanished when we restricted the effect to be monotonically decreasing. At the same time, the effect of Precipitation of Wettest Quarter was estimated as a small but clearly monotonically decreasing effect. With respect to the remaining, “unconstrained” effects, the two models were very similar. The spatial, spatial-temporal, and spatially varying effects almost coincided (see Appendix). The unconstrained smooth and categorical effects were practically identical (Fig. 2). This shows that the model estimation in boosting is very stable with respect to the uncon-

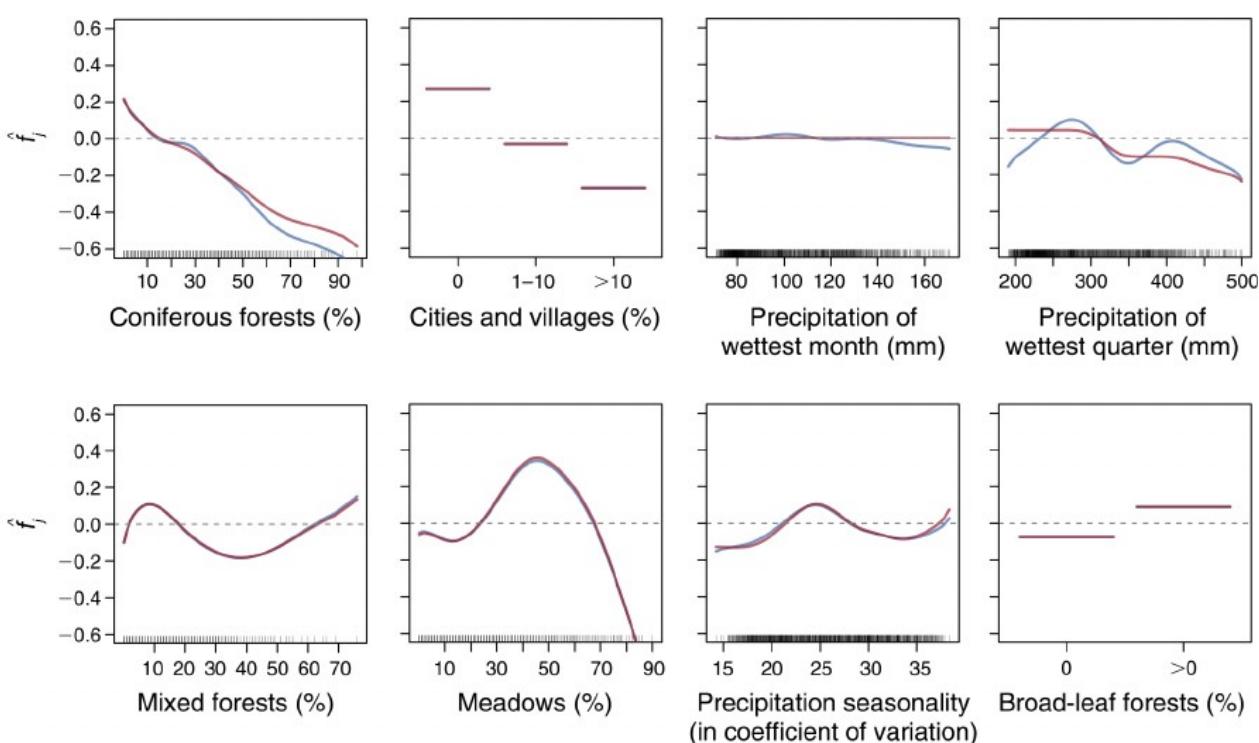


FIG. 2. Estimated partial effects (i.e., the other effects are set to zero in each graph) of the environmental variables on Red Kite (*Milvus milvus*) breeding. The estimates from the unconstrained model (add/vary) are given in blue. The effect estimates from the model with monotonicity constraints (mono) are given in red, where the effects depicted in the upper row of graphs were subjected to a (decreasing) monotonicity constraint. The y -axis, \hat{f}_i , denotes the estimated partial effects of the corresponding value on the x -axis.

Reports

strained effects, i.e., constraining some of the effects where it is ecologically reasonable hardly influences the effects and the interpretation of other variables. Nevertheless, monotonic effects are useful as they allow models that are easier to interpret and reflect the subject-matter knowledge of ecologists better than unconstrained-effect estimates.

DISCUSSION

We extended the flexible framework for species distribution modeling based on boosting methods (Hothorn et al. 2011) to allow monotonic constraints for certain variables. The monotonicity constraint on continuous variables leads to monotonic, yet smooth effects. Many other approaches to monotonic modeling result in non-smooth function estimates (e.g., Dette et al. 2006, de Leeuw et al. 2009, Fang and Meinshausen 2010). Most of these approaches have appealing theoretical properties. In ecology, however, we feel that smooth effect estimates are biologically more plausible and hence preferable. In the limit, i.e., if the ‘true effect’ is monotonic, our approach leads to P-splines. Hence, in this case, it does not matter whether monotonic constraints are used or not.

Monotonicity constraints might be especially useful in (but are not necessarily restricted to) data sets with relatively few observations or noisy data as supported by

our simulation study (see Appendix). Another example where monotonic effects were recently applied can be found in Leathwick et al. (2006). There, the survey method was correlated with the outcome in a way that led to counterintuitive results. The introduction of monotonicity constraints helped to estimate more appropriate models.

Despite the ease of use, monotonicity restrictions should be utilized with care: knowledge of the subject matter should govern the choice of variables that should be modeled monotonically. Monotonicity constraints should be used with great care as, for example, other predominant influences such as strong competitors might govern the species distribution (Austin 2002). Furthermore, *a priori* assumptions might not always reflect the truth and should be carefully reconsidered if a monotonic effect seems inappropriate. In this context, one should keep in mind that other models also have assumptions and constraints, such as linearity, but these are usually ignored or forgotten.

Monotonic P-splines integrate seamlessly in the proposed boosting approach. This allows one to have a single framework to fit possibly complex models. The idea of asymmetric penalties for adjacent coefficients was transferred to ordinal factors. The resulting novel approach to monotonic effect estimation for ordinal variables was also integrated in the boosting framework.

Because of the built-in selection step in each boosting iteration, the algorithm allows one to decide whether the monotonic effect is present or not (see Fig. 2: Wettest Month for Red Kites). The boosting algorithm informs us whether the monotonic effect is present in the data, such that a truly non-monotonic effect will not be selected by the algorithm and hence be set to 0. However, one should be aware that constrained and unconstrained effects for one variable in the same model cannot be specified. This would lead to a preference of the unconstrained effect, as it can better adapt to the data in those cases where small violations of the constraint occur. To compare monotonic and non-monotonic effects, it is necessary to estimate two separate models, as we did for the Red Kite breeding distribution. These models can now be easily compared with respect to the stability and interpretability of the results. The goodness of fit could also be assessed, but one should keep in mind that the unconstrained model might fit the data better yet might be less easy to interpret. The possibility to constrain certain effects in highly flexible statistical methods allows ecologists to start with the most flexible and complex model that, in a stepwise refinement process, is then simplified by restricting certain parts of the model to monotonic or even linear functions without losing too much model accuracy. The resulting simpler model will be easier to interpret and to understand, and, finally, to accept.

Implementation

All proposed baselearners are implemented in the R package **mboost** (Hothorn et al. 2010a, b). The R code for the Red Kite distribution model is given in the Supplement.

ACKNOWLEDGMENTS

The authors thank the three anonymous referees for their helpful comments, Jochen Hölzinger and the Bavarian Environment Agency for providing the Red Kite breeding data from Bavaria, and Karen A. Brune for carefully editing the manuscript. B. Hofner and T. Hothorn were supported by Deutsche Forschungsgemeinschaft, grant HO 3242/1-3.

LITERATURE CITED

- Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157:101–118.
- Bezzel, E., I. Geiersberger, G. V. Lossow, and R. Pfeifer. 2005. Brutvögel in Bayern, Verbreitung 1996 bis 1999. Eugen Ulmer Verlag, Stuttgart, Germany.
- Bühlmann, P., and T. Hothorn. 2007. Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* 22:477–505.
- Bühlmann, P., and B. Yu. 2003. Boosting with the L_2 loss: regression and classification. *Journal of the American Statistical Association* 98:324–339.
- de Leeuw, J., K. Hornik, and P. Mair. 2009. Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software* 32.
- Dette, H., N. Neumeyer, and K. F. Pilz. 2006. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12:469–490.
- Eilers, P. H. C. 2005. Unimodal smoothing. *Journal of Chemometrics* 19:317–328.
- Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11:89–121.
- Elith, J., and C. Graham. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32:66–77.
- Fahrmeir, L., T. Kneib, and S. Lang. 2004. Penalized structured additive regression: A Bayesian perspective. *Statistica Sinica* 14:731–761.
- Fang, Z., and N. Meinshausen. 2010. LASSO ISOtone for high dimensional additive isotonic regression. Technical report, ArXiv e-prints. <<http://arxiv.org/archive/stat.ME/1006.2940>>
- Gertheiss, J., and G. Tutz. 2009. Penalized regression with ordinal predictors. *International Statistical Review* 77:345–365.
- Hofner, B., T. Hothorn, T. Kneib, and M. Schmid. 2011. A framework for unbiased model selection based on boosting. *Journal of Computational and Graphical Statistics*, *in press*. [doi: 10.1198/jcgs.2011.09220]
- Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner. 2010a. mboost: Model-based boosting. <<http://r-forge.r-project.org/projects/mboost/>>. R package version 2.1-0.
- Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner. 2010b. Model-based boosting 2.0. *Journal of Machine Learning Research* 11:2109–2113.
- Hothorn, T., J. Müller, B. Schröder, T. Kneib, and R. Brandl. 2011. Decomposing environmental, spatial, and spatiotemporal components of species distributions. *Ecological Monographs* 81:329–347.
- Kneib, T., T. Hothorn, and G. Tutz. 2009. Variable selection and model choice in geoadditive regression models. *Biometrika* 65:626–634.
- Leathwick, J. R., J. Elith, M. P. Francis, T. Hastie, and P. Taylor. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* 321:267–281.
- Mebs, T., and D. Schmidt. 2006. Greifvögel Europas, Nordafrikas und Vorderasiens: Biologie. Bestandsverhältnisse. Bestandsgefährdung. Franckh-Kosmos Verlags, Stuttgart, Germany.
- Nitsche, G., and H. Plachter. 1987. Atlas der Brutvögel Bayerns 1979–1983. Bayerisches Landesamt für Umweltschutz, München, Germany.

APPENDIX

Further details on the habitat-selection model of the Red Kite together with empirical results for monotonic effects from a simulation study (*Ecological Archives* E092-161-A1).

SUPPLEMENT

R code for conducting statistical analysis of Red Kite distribution and for the simulation study (*Ecological Archives* E092-161-S1).

