Extended abstract No. 156

UTILIZATION OF GENERALIZED LINEAR AND GENERALIZED ADDITIVE MODELS TO PREDICT PLANT SPECIES DISTRIBUTIONS IN PEATLANDS.

Rémy Pouliot
Université Laval, Peatland Ecology Research Group, Centre d'Études Nordiques. 2425 rue de l'Agriculture, Quebec City, Canada, G1V 0A6. Telephone number: 1-418-656-2131. E-mail address: remy.pouliot.1@ulaval.ca

Line Rochefort
Université Laval, Peatland Ecology Research Group, Centre d'Études Nordiques.

SUMMARY

Predictive models as generalized linear or generalized additives models (GLMs or GAMs) are tools that provide statistical relationships between species abundances and a series of climatic or chemical variables. The prediction of the probability to find a species in non-surveyed locations and the quantification of species presence are thus possible. GLMs and GAMs were used for the prediction of the distribution of ericaceous shrubs and *Sphagnum* mosses in natural peatlands along many environmental gradients. In addition, we will discuss of the possibility to use GLMs or GAMs results as a tool evaluating the success of peatland ecological restoration or in a goal of species conservation.

KEY WORDS: Bogs; GAM; GLM; Niches; Predictive curves.

INTRODUCTION

Species data based on random or systematic field sampling, such as abundance or presence-absence observations, are frequently used to describe species distribution along environmental gradients at specific locations. However, the graphical representation of these data gives only a picture of a site at a particular moment and comparisons with other sites or geographical regions are limited. In contrast, species distribution models usually provide valuable statistical tools used to describe the relationship between the plant responses (field observations) and a series of predictor variables (the different environmental gradients). In that case, the prediction of the probability to find a species and to quantify its abundance is possible according to environmental conditions of other non-surveyed sites.

Species distribution modeling along many environmental gradients (the predictors) could be used to illustrate their ecological niche. It is generally assumed that a species distribution model represent the realized niche of a species because field observations are already constrained by biotic interactions or limiting resources. Environmental gradients influence species directly or indirectly by many ways but can be classified into three groups: 1) limiting factors, such as temperature, 2) disturbances (natural or anthropogenic) and 3) resources, such as nutrients or water (Guisan and Thuiller, 2005). Environmental gradients with the most distinct niches are usually considered as the most important ones (Gignac, 1992), but they must have a significant effect on studied species (Økland, 1990). In addition, species distribution models have to be robust: a significant model explaining a low proportion of the variance might lead to false conclusions.

Generalized linear and generalized additive models (GLMs and GAMs) are frequently used to create predictive curves for species distributions (Guisan et al., 2002) They are extremely flexible as they allow us to work with the different distributions of ecological including the not normal data. They also fit well with more traditional statistics as linear modeling or analysis of variance. GLMs and GAMs are both based on a link function assuming the relationship between the response variable and the environmental predictor(s). The major difference between GAMs and GLMs is the linear predictor chosen by the user (Barry and Welsh, 2002) and the fact that the link function is smoothed for GAMs. These predictive models can increase our comprehension of ecosystems as they have the capacity to represent the non-linear structures frequently found in field data.

Two theoretical models describe well the species responses to environmental gradients: the Gaussian symmetric unimodal and the skewed unimodal responses. Using artificial data based on explicit theory, both predictive models attain a comparable success to compare these kinds of plant responses (Austin et al., 2006). The authors concluded that GLMs and GAMs performed well with Gaussian responses and direct predictors as radiation or temperature whereas GAMs showed slightly better results for skewed responses and indirect predictors as slope or orientation. Ecological background and the statistics abilities of users have more impact on conclusions made than the chosen predictive model (Austin et al., 2006).

Ecological niches of *Sphagnum* species are well known in western North America (see for example: Gignac, 1992) and in Europe (see for example: Hájková and Hájek, 2007). The situation is less clear in eastern North America where, to our knowledge, ecological niches of *Sphagnum* species have not been fully described up to date. On the other hand, ecological niches of other plant families typical of peatlands, such as Cyperaceae, have been described in Canada (Gignac et al., 2004), but much work is still needed for the comprehension of other peatland vascular plant niches like ericaceous shrubs. Our principal objective is to develop predictive curves using GLMs or GAMs for the distribution of *Sphagnum* and ericaceous species in natural peatlands dominated by *Sphagnum* mosses of Eastern Canada. A secondary objective is to evaluate GLMs and GAMs as predictive tools by comparing the results of both methods. For the purpose of this extended abstract, we will only present few preliminary results and discuss about some possible uses of species distribution models.

MATERIALS AND METHODS

Site selection and plant species variables

In total, 146 natural peatlands dominated by *Sphagnum* mosses in eastern Canada (all bogs, semi-forested or open) were visited once between 1993 and 2008. In each peatland, abundance of all species was noted (cover percentage) in vegetation quadrats. Between three and thirty quadrats per peatlands were done. For the analyses, a total of 744 quadrats were used.

Studied environmental variables (the predictors)

Meaningful environmental variables were selected according to results of another study that compared niche breadth and niche overlap of different peatland species in natural bogs with abandoned or restored peatlands following peat extraction (Pouliot, 2011). Thirteen variables had a significant effect on the abundance of plant species found in natural peatlands dominated by *Sphagnum* mosses (pH; electrical conductivity; Ca, K, Mg, Na and SO_4 concentrations; distance from water table; shade; mean annual temperature; number of days with temperature > 0°C; degree-days and mean annual precipitation). Chemical variables were measured in water samples taken on each natural peatland (around one sample per group of three quadrats) and climatic variables derived from values found in the database of Environment Canada for the nearest meteorological station.

Statistical procedures

For the purpose of this communication, we compared the results along one environmental gradient (mean annual precipitations; from 880 to 1589 mm per year) for two ericaceous shrub species (*Chamaedaphne calyculata* and *Rhododendron groenlandicum*). GLM and GAM procedures of the SAS software were used. For GAMs, two smoothing functions available in the SAS software were tested (SPLINE and LOESS). Ultimately, all significant environmental variables will be used to predict the distribution of the most common *Sphagnum* mosses and ericaceous shrubs found in sampled natural peatlands. To determine to most significant model, we will also test many link functions with both GLMs and GAMs.

RESULTS

Comparison between GLMs and GAMs

Mean annual precipitation had a significant effect for both species (for *Chamaedaphne calyculata*: $p = 0.037$ and for *Rhododendron groenlandicum*: $p = 0.001$; backward selection of variables). For GLMs, the most significant model was the quadratic one for *C. calyculata* ($p < 0.001$; Fig. 1b) whereas that was the linear one for *R. groenlandicum* ($p < 0.001$; Fig. 1e). For GAMs, the most significant was the SPLINE smoothing function for *C. calyculata* ($\chi^2 < 0.001$; Fig. 1c), but that was the LOESS one for *R. groenlandicum* ($\chi^2 < 0.001$; Fig. 1f). The shape of predicted curves was similar whatever the generalized model (GLM or GAM).

DISCUSSION / CONCLUSION

Comparison between GLMs and GAMs

Predictive curves of species abundance along the mean annual precipitation gradient illustrated relatively the same responses for both GLMs and GAMs. That is similar with conclusions of Austin et al. (2006). Even if the predictive values were very similar for both methods, the species distribution model done by GAM for *Chamaedaphne calyculata* predicted lower covers for the higher predicted values. For *Rhododendron groenlandicum*, the linear effect shown with GLM was less evident for GAM even if the same decrease of predicted values was observed. The preliminary results presented here will be complemented with future analyses including more environmental gradients and link functions. They will improve our predictions and help to identify the best method among GAMs or GLMs.

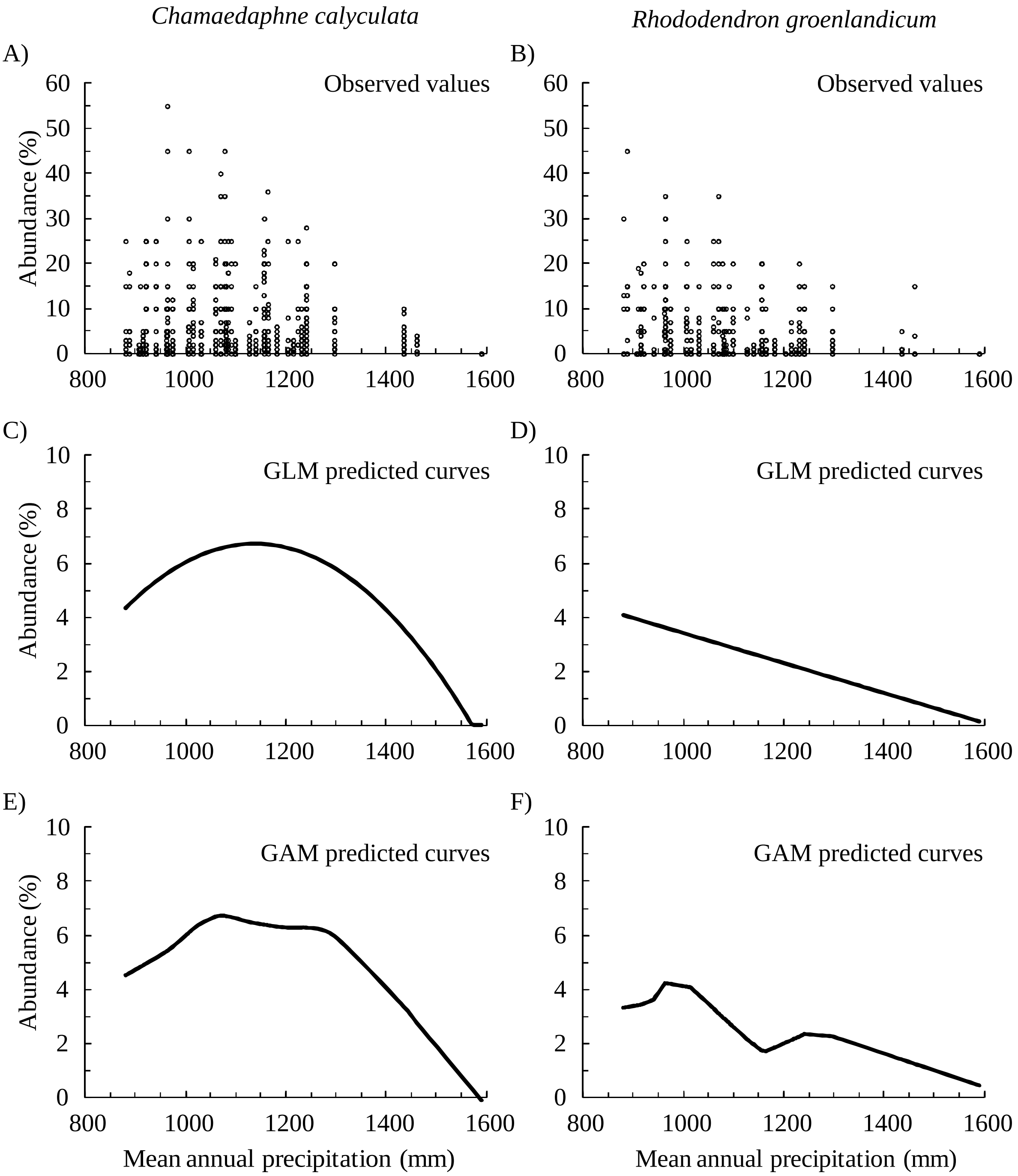

Figure 1. Observed and predicted values of species abundance along mean annual precipitation gradient for *Chamaedaphne calyculata* and *Rhododendron groenlandicum*. Parts A-B represent the observed values following field surveys in natural peatlands dominated by Sphagnum mosses in eastern Canada (n = 744 quadrates). Parts C-D and parts E-F represent respectively the predictive curves for species abundance generated with generalized linear models (GLMs) and generalized additive models (GAMs). Note that X-axis begin at 880 mm.

Distribution along mean annual precipitation gradient

Chamaedaphne calyculata and *Rhododendron groenlandicum* generally co-occur in natural peatlands dominated by *Sphagnum* mosses. However, it seems that *R. groenlandicum* should be more abundant along the dry end of the mean annual precipitation gradient (below 1000 mm). On the other hand, *C. calyculata* should be more present in the middle part of the gradient (between 1080 and 1330 mm). Information along other environmental gradients will be included in the next steps of this project to better predict the distribution of this two species as well as other ericaceous species.

Possible uses of GLMs and GAMs in peatland ecosystems

Predictive curves for species distribution could be used for many purposes. They can illustrate the ecological niches of species in a given ecosystems or along the most important environmental gradients, as it was shown in this short communication for peatlands dominated by *Sphagnum* mosses (and will be done later in more details). Species distribution models can be really useful for species conservation projects as they can predict the potential abundances that different species may achieve in a particular area. Managers would be thus able to take action if the abundance of a given species is lower than expected.

Another interesting application of GLMs or GAMs outputs is for peatland ecological restoration projects or management (or for any other type of ecosystem restoration). For peatland restoration, one could predict the potential species target pool based on the climatic and residual abiotic conditions of the site to be restored. It follows that the evaluation of the vegetation recolonization success would be possible taking the new environmental conditions in the restored sites as a reference rather than using plant abundance found in natural peatlands (the classical approach in the evaluation of the restoration success). Therefore, the absence of a species in a restored peatland would not be necessarily interpreted as a failure of the restoration project but as a consequence of irretrievable alterations in the site environmental conditions. Species distribution models such as GLMs and GAMs could thus help to explain and quantify these differences.

ACKNOWLEDGEMENTS

Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada, the Ministry of Natural Resources and Energy of New Brunswick, and the Canadian Sphagnum Peat Moss Association and its members. Special thanks go to Monique Poulin for the permission to use her data and to Dennis L. Gignac for their explications about GLMs and GAMs. Thanks to all PERG members for the constructive discussions.

REFERENCES

Austin, M. P., Belbin, L., Meyers, J. A., Doherty, M. D. and Luoto, M. (2006). Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling* **199**, 197-216.

Barry, S. C. and Welsh, A. H. (2002). Generalized additive modelling and zero inflated count data. *Ecological Modelling* **157**, 179-188.

Gignac, L. D. (1992). Niche structure, ressource partitioning, and species interactions of mire bryophytes relative to climatic and ecological gradients in Western Canada. *The Bryologist* **95**, 406-418.
Gignac, L. D., Gauthier, R., Rochefort, L. and Bubier, J. (2004). Distribution and habitat niches of 37 peatland Cyperaceae species across a broad geographic range in Canada. *Canadian Journal of Botany* **82**, 1292-1313.
Guisan, A., Edwards Jr., T. C. and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions, setting the scene. *Ecological Modelling* **157**, 89-100.
Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**, 993-1009.
Hájková, P. and Hájek, M. (2007). *Sphagnum* distribution patterns along environmental gradients in Bulgaria. *Journal of Bryology* **29**, 18-26.
Økland, R. H. (1990). A phytoecological study of the more Northern Kisselbergmosen, SE Norway. III. Diversity and habitat niche relationships. *Nordic Journal of Botany* **10**, 191-220.
Pouliot, R. (2011). Initiation du patron de buttes et de dépressions dans les tourbières ombrotrophes boréales. PhD Thesis. Université Laval, Québec, Canada. 247 pp.