

SCGLR: a component-based multivariate regression method to model species distributions.

F. Mortier<sup>a</sup>, X. Bry<sup>b</sup>, G. Cornu<sup>a</sup>, and C. Trottier<sup>b</sup>

(a) UR « Biens et Services des Ecosystèmes forestiers » CIRAD, Montpellier, France,  
[fmortier@cirad.fr](mailto:fmortier@cirad.fr) and [guillaume.cornu@cirad.fr](mailto:guillaume.cornu@cirad.fr)

(b) Institut de Mathématiques et de Modélisation de Montpellier (I3M) CNRS : UMR5149 –  
Université Montpellier II - Sciences et techniques,  
[catherine.trottier@univ-montp2.fr](mailto:catherine.trottier@univ-montp2.fr) and [xavier.bry@univ-montp2.fr](mailto:xavier.bry@univ-montp2.fr)

**Keywords:** Tropical forests, species distributions, generalized linear models, regression, partial least square, Fisher scoring algorithm

**Abstract:** Understanding how environmental factors spatially impact species coexistence is of primary importance in ecology. Species distribution models are statistical or mathematical models combining observations of species occurrence or abundance with environmental factors. They are used to gain insights into ecological and evolutionary processes and to predict distributions across landscapes, sometimes requiring extrapolation in space and time. Generalized linear models (GLM) are frequently used for modeling species distributions. Nevertheless GLM can suffer from different limitations: (i) the classical approaches do not allow the simultaneous modeling of multiple species (ii) the level of model realism and the prediction robustness can be poor if relevant predictors are omitted. We propose a regularizing component-based technique, supervised component generalized linear regression (SCGLR), which extends the Fisher scoring algorithm by combining partial least square (PLS) regression with GLM estimation in the multivariate context. The proposed method can handle different types of dependent variables (occurrence, abundance, growth) and explanatory variables mixing different types, and account for additional covariates. We illustrate SCGLR and the associated R-package using a dataset built from the abundance of 27 common tree genera in the tropical moist forest of the Congo-Basin and 40 geo-referenced environmental variables collected in the CoForChange project<sup>1</sup>.

## References:

Bry, X. and Trottier, C. and Verron, T. and Mortier, F. 2013. Supervised component generalized linear regression using a PLS-extension of the Fisher scoring algorithm. *Journal of Multivariate Analysis*. 47—60

Mortier F. and Trottier C. and Cornu G. and Bry X. 2013. SCGLR: Supervised Component Generalized Linear Regression (SCGLR). R package version 1.1, <http://CRAN.R-project.org/package=SCGLR>

Mortier F. and Trottier C. and Cornu G. and Bry X. 2013. SCGLR - An R package for Supervised Component Generalized Linear Regression. *Journal of Statistical Software*. submitted

---

<sup>1</sup> <http://www.coforchange.eu/project>