

---

# Functional gradient descent boosting for additive non-linear spatial autoregressive model

Ghislain Geniaux\*<sup>1</sup>

<sup>1</sup>INRAE Ecodeveloppement UR 767 – INRAE – 228 route de l’aérodrome CS 40509 Domaine St Paul - Site Agroparc 84914 AVIGNON Cédex 9, France

## Résumé

I propose in this paper to lay one of the first stones of a bridge between the canonical models of spatial econometrics and machine learning algorithms in order to determine, in bigdata context, which variables to introduce in autoregressive nonlinear models and in which forms? Namely: linear, non-linear, spatially varying, with interactions with other variables? To answer these questions, I propose an extension of the boosting algorithms (Friedman, 2001; Bühlmann et al., 2007) to semi-parametric autoregressive models (SAR, SDM, SEM and SARAR) formulated in the form of additive model with smoothing splines functions. This adaptation is mainly based on an estimation of the spatial parameter by QML following the example of Basile and Gress (2004) and Su and Jin (2010). To avoid the calculation of the spatial multiplier, I propose two extensions of my estimators. One is based on a direct application of the Closed Form Estimator (CFE) recently proposed by Smirnov (2020). I also propose a Flexible Instrumental Variable Approach (Marra and Radice, 2010, Basile et al. 2014) for SAR models with a dynamic construction of the instruments imposed by the way functional gradient descent boosting is working. The proposed estimators can be easily extent to the use of decision trees instead of smoothing splines to allow the identification of more complex variable interactions. Using synthetic data, I demonstrate the good finite sample properties of all my estimators for estimating the non-linear functions and the spatial autoregressive parameter. I also illustrate their interest in term of out-sample accuracy with real data using a large dataset on house prices in France, by comparing the proposed estimators to main of the estimators of statistics, spatial econometrics and machine learning algorithms adapted to regression with spatial data. All these estimators have been made available in a R package entitled `spboost` that will be posted on CRAN at the date of the workshop.

---

\*Intervenant