Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 27 (2015) 53 – 57

Procedia
Environmental Sciences

Spatial Statistics 2015: Emerging Patterns

Spatial Variation of Drivers of Agricultural Abandonment with Spatially Boosted Models

Max Schneider[a,b]*, Gilles Blanchard[a], Christian Levers[b], Tobias Kuemmerle[b]

[a]*Institute of Mathematics, University of Potsdam, Germany*

[b]*Institute of Geography, Humboldt University of Berlin, Germany*

Abstract

Agricultural abandonment (AA) is a significant land use process in the European Union (EU) and modeling its driving factors has great scientific and policy interest. Past studies of drivers of AA in Europe have been limited by their restricted geographic regions and their use of traditional statistical methods, which fail to consider the spatial variation in both predictors and AA itself. In this study, we implement a modeling framework based on boosted classification with spatially-varying terms, choosing the squared loss function and P-splines as base learners for their mathematically superior properties. By comparing models containing both constant and spatially-varying coefficients, we find telling spatial trends in the relationship between the drivers and AA that can be used to inform international policies for agriculture.

**Corresponding author: Max Schneider: maxsch03@uni-potsdam.de*

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee

Keywords: boosted regression; spatial modeling; agricultural abandonment; model selection; land-use change

1. Motivation and Background

The abandonment of agricultural land is a key process of land use and management over the recent history of the European Union. Here, the extent of agricultural land has been declining for numerous reasons, e.g., increasing yields on productive lands, conservation policies or urban pull factors [Cramer et

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee
doi:10.1016/j.proenv.2015.07.114

al. (2008)]. This shift in land use has both negative and positive effects on, e.g., local biodiversity, carbon sequestration, occurrence of fire and cultural changes in rural areas [Benayas et al. (2007), Baumann et al. (2011)]. There is thus pressing scientific and policy interest in better understanding the process of agricultural abandonment (AA), which stems from improved models of its leading drivers.

A number of studies have previously modeled AA; however, two key limitations remain. Available studies focus on spatially limited regions in Europe, counterintuitive to the many EU-wide regulations and programs related to AA [Baumann et al. (2011), Gellrich et al. (2007)]. In addition, the majority of studies use basic forms of logistic regression to model the binary response variable (agriculture abandoned or not?) with a set of regressors driving the response. This traditional approach ignores the spatial trend that inevitably exists both within AA and the corresponding explanatory variables.

In this study, we identify drivers of agricultural abandonment in a wall-to-wall fashion for the first 27 member states of the EU (the EU27). Not only do we cover a broader spatial region (and thus, a far larger dataset) than previous work, but we also explicitly modelize spatial characteristics of the hypothesized drivers and AA itself. For this aim, we build a series of models using a modified form of boosted classification as proposed by Hothorn et al. (2011). This results not only in variable selection for the top drivers of AA but also model selection between a set of candidate models, which isolate the importance of the spatial variation of these drivers. Comparing across models, we find noticeable differences in the behavior of top drivers over the spatial domain.

2. Spatial Boosting as a Classification Framework

In classical boosted classification, a set of d regressor variables $x = (x^{(1)}, ..., x^{(d)})$ is related to a binary response variable, y_i ($i = 1, ..., n$), using a modeling technique referred to as a base learner and denoted $h(x, \theta)$. The parameters for the base learner, θ, are chosen such that they minimize the empirical risk corresponding to a loss function, $L(y, f(x))$, where $f(x)$ is the model predicting y. We compute the negative gradient of this empirical risk function, evaluated with f as the best-fit model given the empirical dataset. The base learner is then used to construct a model to fit x onto this negative gradient vector, minimizing the model's residual sum of squares. This process is repeated iteratively over subsequent empirical risk negative gradients and in each step, a weighted version of the parameter values is added to the previous iteration's parameter values. Algorithm 1 describes the steps for the specific procedure we implement.

The loss function used in our case is the squared loss, which means that its negative gradient is simply the residual, $u_i = y_i - \hat{f}_m(x)$. This method is called L2Boost and has been proven to improve the mean squared error (MSE) over a standard linear learner, the regression tool commonly found in the literature [Buehlmann and Yu (2003)]. It was further shown that smoothing splines will produce minimax optimal MSE when taken as the base learners in L2Boost; however, these have a penalty term which is difficult to compute and so we follow past studies which use P(enalty)-splines instead [Kneib et al. (2009)]. Their smoothing parameter λ is taken as constant for all regressors and is derived from the literature [Hothorn et al. (2011)].

Algorithm 1:

Step 1. Given data (x_i, y_i), $i = 1, ..., n$, fit an (initial) smoothing spline: $\hat{f}_0(x) = h(x, \theta)$. Set $m = 0$.

Step 2. Calculate residuals $u_i = y_i - \hat{f}_m(x)$. Fit a P-spline to the current iteration's residuals. The current iteration's fit is denoted $\hat{h}(\bullet)$. Then, update $\hat{f}_{m+1}(\bullet) = \hat{f}_m(\bullet) + \hat{w}_{m+1} * \hat{h}_{m+1}(\bullet)$, where $\hat{w}_{m+1}$ is the weight given to that (m+1)-th iteration's fit.

Step 3 (iteration). Raise the iteration index $m = m+1$ and repeat step 2. Continue until reaching the stopping iteration, m_{stop}.

Another critical parameter in the boosting procedure is the stopping iteration m_{stop}, as allowing for too many iterations results not only in over-fitting but also increased computational cost. Theoretical study of