The Use of the Geoadditive Model with Interactions in a Precision Agriculture Context: a Comparison of Different Spatial Correlation Structures

Barbara Cafarelli, Corrado Crocetta
Università degli Studi di Foggia, b.cafarelli@unifg.it

Annamaria Castrignanò
CRA, Bari

Abstract: Accelerated land degradation is mostly human induced and occurs in all eco-regions regardless of social, economic and political conditions. Precision Agriculture is an ecological management strategy based on the use of several sources of information in order to support decisions concerning the agricultural practice. In this context, the use of methodologies, taking into account spatial and temporal variability associated to every aspect of agricultural production processes, can improve crop yields and environmental quality. In this paper, a geoadditive model with interactions is proposed to analyse the nonlinear relations between an indicator of durum wheat production with other crop features with the aim of considering explicitly the spatial dependence and the temporal variation in production.

Keywords: Geoadditive Model, Matérn family, Spatial Correlation Structures, Cross-validation, Additive model with Interactions

1. Introduction

Precision Agriculture or site-specific crop management is a means of managing spatial and temporal variability of different data types: edaphic (i.e. soil related), anthropogenic, topographic, biological and meteorological factors which are deemed to affect crop yield. The target of Precision Agriculture is to increase crop productivity, optimise inputs, increase farmer's profitability and reduce environmental impact, through the application of variable rate inputs on the basis of the actual local requirements of crop rather than an estimation averaged over the whole field. In this context, defining reliable methods for assessing and predicting within-field variations in soil and crop properties is very important. Effects of the soil's physical and chemical properties on crop yield are predictable and can be mapped relatively easily, whereas effects due to climatic conditions, nutrient deficiency, pests and diseases, being time-dependent, are more difficult to predict. The application of proper statistical models, to assess spatial and temporal variation and predict crop response to site-specific environmental conditions, is then crucial in the perspective of Precision Agriculture. In particular, a geoadditive model with interactions is proposed to analyse the spatial distribution of the harvest index (White, Wilson, 2006), a commonly used indicator of commercial wheat production, and its nonlinear relations with other crop features over two years. The model adopted here is a further development of an extension (Cafarelli and Castrignanò, 2011) of the original geoadditive model of Kammann and Wand (2003), that explicitly considers data stratified according to crop in two different years. Two geoadditive models with interactions, considering the same response variable and covariates, for the same linear and non-linear relationships between response variable and covariates over time, but differing for spatial correlation structures, are fitted. The selection among the fitted models is done by cross-validation (Carroll and Cressie, 1996).

2. Materials and Methods

The trial was carried out on a 12-ha field cropped with durum wheat (Triticum durum DESF), located at the CER-CRA research centre for cereals, Foggia (41° 27' N, 15° 36' E, 90 m above sea level), south-eastern Italy. The soil was a deep, silty-clay Vertisol of alluvial origin, classified as fine, mesic, Typic, Chromoxerert. The climate was characterized by hot and dry summers and rains concentrated mostly in the winter months. The agricultural trial was carried out during two crop seasons: 2005-2006 and 2007-2008. One-hundred georeferenced measurements of the harvest index (*HI*), number of fertile plants (*FP*) and electrolytic weight (*EW*) were taken for each year. The samples with more than one missing value were discarded leaving only ninety-three and ninety-one georeferenced soil samples to be considered for the first and the second wheat season, respectively. that *HI* had different spatial distributions in the two years, which share a marginal bell shaped distribution. This consideration was supported by a graphical check, which led us to adopt a semi-parametric approach, based on a geoadditive model with interactions. A full representation of the geoadditive model with interactions is:

$$HI_i = \beta_0 + \beta_1 EW_i + g(FP_i) + f_{year}(location\_of\_year_i) + \boldsymbol{\beta}_{\mathbf{x}}\mathbf{x}_i + S(\mathbf{x}_i) + \varepsilon_i, \quad (1)$$

where $i = 1, \ldots, 184$ represents the spatial-temporal observation, g and f are smooth functions, $\mathbf{x}_i = (X_i, Y_i)$, in UTM WGS84 coordinate system, is the spatial location of the i-th observation and $S(\mathbf{x}) \sim N(0, \sigma^2_{\mathbf{x}} h_0(r, \nu))$, where $\sigma^2_{\mathbf{x}}$ is the sill, r is the range, ν is the smoothing parameter and $h_0(r, \nu)$ is a Matérn family covariance function used to specify the spatial correlation structures. The exponential and the Gaussian covariance structures were used in the fitted models. This occurred by setting $\nu = 3/2$ or $\nu \to \infty$, respectively, in the function $h_0(r, \nu)$ (Minasny and McBratney, 2005). Independently of the specification of $h_0(\cdot)$, the Gaussian spatial process $S(\cdot)$ is independent of the error term $\boldsymbol{\varepsilon}$ and the additive components. In model 1, the term $f_{year}(\cdot)$ corresponds to the number of spatial locations within a particular year and represents the interaction between the year factor and the overall spatial effect. The relatively small sample size permitted the use of the parsimonious low rank parameterization of model 1 (Hastie, 1996). The choice of linear components was done according to approximated Z-values given by `lme`, while the significance of nonlinear effects, identified with the exploratory data analysis, was assessed by restricted likelihood ratio tests (Kammann and Wand, 2003; Greven et al., 2008; Crainiceanu, 2008; Ruppert et al., 2009). Independently of the spatial correlation structure adopted, the number of nodes for representing the nonlinear *FP* effect was 15 and was obtained as in Ngo et al. (2004), whereas the number of nodes in the low-rank formulation of the spatial component was obtained by CLARA algorithm (Kaufman et al., 1990). The coordinates of the 23 spatial nodes were obtained by a space-filling algorithm implemented in function `default.knots.2D` within the R library SemiPar. The low rank formulation of model 1 was estimated by REML using function `lme` of the R library `nlme` (Pinheiro and Bates, 2000). The three cross-validation techniques CV_1, CV_2 and CV_3 suggested by Carroll and Cressie (1996) were used to compare the accuracy and the precision of estimates of the two models. In particular CV_1 was used to assess the unbiasedness of the predictor (optimal value: CV_1=0), CV_2 was used to assess the accuracy of the mean squared prediction error (optimal value: CV_2=1) and CV_3 was used to check the goodness of prediction (small value of CV_3 indicates a good fit).

3. Results

The result comparison suggests that the two fitted models have good and similar performances and are very useful for analyzing the relationship between *HI* and the covariates during the two crop years (Table 1). For this reason the most generally used exponential covariance structure was chosen. The fitted geoadditive model, obtained by using the exponential correlation structure to specify the spatial dependence of the geographical component, is reported in Table 2. From the table inspection, one sees that both agronomical variables (*EW*, *FP*) impact significantly on *HI*, however the relationship with *FP* is more complex, due also to the higher uncertainty in *FP* measurement. All nonlinear components of the geoadditive model are significant on the basis of the degrees of freedom (Table 2) estimates that confirmed the appropriateness of including the nonlinear effects of *FP*, the spatial component and the interaction between the factor year and the overall spatial effect in the fitted model.

Spatial correlation structure	CV_1	CV_2	CV_3
Exponential	-0.61	1.29	8.44
Gaussian	-0.62	1.31	8.42

4. Concluding remarks

The proposed approach is a quick and effective method of predicting the spatial distribution of the harvest index using standard agronomic measurements over two years. The great advantages of geoadditive models lie mainly in the possibility to jointly analyse spatial and temporal variations and to treat the complex interactions, quite often non linear, between production process and several different variables (soil, crop, atmosphere, management). Moreover, these models allow us to predict agronomical variables in specific locations of the field and this piece of information is crucial for Precision Agriculture. These considerations and the possibility of estimating linear effects and variance components of non linear effects and error term by REML, using mixed effects model procedures routinely implemented in statistical software, lead us to recommend a wider use of geoadditive models with interactions in the presence of spatial dependence and temporal variation.

Table 1: Cross-validation errors with two different spatial correlation structures

Linear component			
Covariates	**Coefficients**	**Std.Error**	**p-value**
EW	0.092	0.028	**<0.05**
Non-linear component			
Covariates	**df**	**N° knots**	
FP	**9.35**	**15**	
locations of year$_{2006}$	**8.02**	**15**	
locations of year$_{2008}$	**8.02**	**15**	
X, Y	**7.02**	**23**	

Table 2: Summary of the REML based fit of the model with exponential correlation structure.

References

Cafarelli B., Castrignanò A. (2011). The use of geoadditive models to estimate the spatial distribution of grain weight in an agronomic field: a comparison with kriging with external drift, *Environmetrics* 22.

Carroll SS, Cressie N. (1996). A comparison of geostatistical methodologies used to estimate snow water equivalent. *Water Resour. Bull.* 32: 267– 278.

Greven S, Crainiceanu CM, Kuechenhoff H, Peters A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17: 870-891

Kammann EE, Wand MP. (2003). Geoadditive models, *Applied Statistics*, 52: 1-18.

Minasny B. and McBratney A.B. (2005). The Matérn function as a general model for soil variograms, Geoderma 128, 192-207.

Pinheiro JC and Bates DM. (2000). *Mixed-Effects Models in S and S-Plus*, Springer Verlag: New York

Ruppert D, Wand MP, Carroll RJ. (2009). Semiparametric Regression During 2003-2007. *Electronic Journal of Statistics*, 3: 1193-1256

Stein ML. 1999. *Interpolation of Spatial Data – Some Theory for Kriging*. Springer Verlag, Ney York.

White E.M., Wilson F.E.A. (2006). Responses of grain yield, biomass and harvest index and their rates of genetic progress to nitrogen availability in ten winter wheat varieties. *Irish Journal of Agricultural and Food Research* **45**: 85–101.