"About predictions in spatial autoregressive models: Optimal and almost optimal strategies"

Christine Thomas-Agnan, Thilbault Laurent and Michel Goulard

About predictions in spatial autoregressive models: Optimal and almost optimal strategies

December 18, 2013

Abstract

We address the problem of prediction in the classical spatial autoregressive lag model for areal data. In contrast with the spatial econometrics literature, the geostatistical literature has devoted much attention to prediction using the Best Linear Unbiased Prediction approach. From the methodological point of view, we explore the limits of the extension of BLUP formulas in the context of the spatial autoregressive lag models for in sample prediction as well as out-of-sample prediction simultaneously at several sites. We propose a more tractable "almost best" alternative. From an empirical perspective, we present data-based simulations to compare the efficiency of the classical formulas with the best and almost best predictions.

1   Introduction

Whereas prediction is a basic concern in geostatistics (Cressie, 1990), it has not been paid as much attention in the econometrics literature. Bivand (2002) recognizes the importance of the question: "Prediction for new data ... is a challenge for legacy spatial econometric models, raising the question of what a BLUP (best linear prediction) would look like". Kato (2008) explores the best linear prediction problem in the framework of spatial error models. In the context of spatial lag models, other authors (LeSage and Pace (2004), Pace and LeSage (2008), Kelejian and Prucha (2007), Pace and LeSage (2008), Bennet, Griffith and Haining (1989)) have addressed some aspects of this question and we will summarize their contribution in section 2.

We first present the different types of prediction situations encountered according to whether we predict at a sample unit or an out-of-sample one and to whether

one or several points are predicted simultaneously. To motivate the need for out-of-sample prediction, let us present the context of a case study in Lesne et al. (2008). Until 1999, the French population census was exhaustive and realized by the French statistical institute (INSEE) approximately every ten years. Since 2004, this exhaustive census has been replaced by a census survey which consists in annual samples and delivers an up-to-date information. In particular, the communes with less than 10000 inhabitants at the 1999 census (called *small communes*) are sampled exhaustively every five year at the rate of one fifth per year. The sampling design of these small communes is stratified by region and inside each region, the small communes are partitioned into five rotational groups by using a balanced sample design taking into account some auxiliary socio-economics variables given by the 1999 census. Between 2004 and 2009, polling organizations needed an estimate of the population for all the small communes and of its evolution since the previous complete census of 1999. The population of all the small communes would not be delivered by the INSEE before 2009 but data sets containing the population of the two first rotational groups, corresponding to 2004 and 2005, were already known and could be used to predict the population of the other three rotational groups. In that case, out-of-sample prediction formulae were necessary for spatial models. Figure 1 presents the positions of the spatial units where population data was available at the time of this case study.
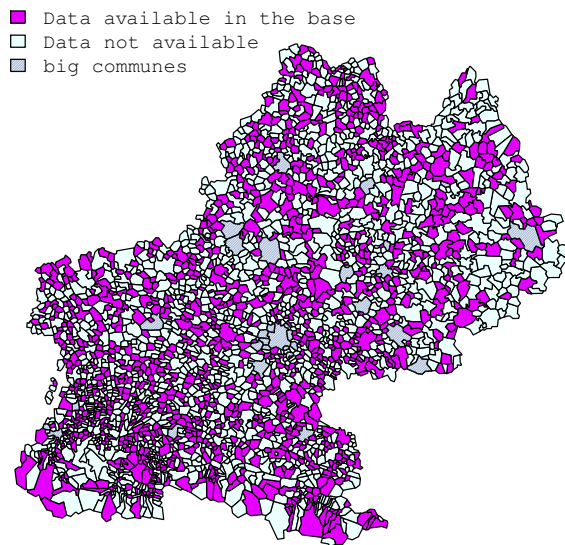

Figure 1: Spatial units where population data was available at the time of this study

We first review the classical prediction formulae encountered in the literature for the spatial simultaneous autoregressive (SAR or LAG depending on authors) models. Then we recall how best linear unbiased prediction (BLUP) can be done in the framework of these models using an adapted formulation of the Goldberger formula.

We introduce several alternatives to this formula and finally demonstrate that the simple formulas classically implemented in usual softwares can thus be improved upon substantially.

2 State of the art about best prediction in spatial autoregressive lag models

2.1 Models and prediction situations

We consider prediction in the classical homoscedastic spatial autoregressive lag model (LAG model hereafter). Given a spatial weight matrix W, this model can be written

$$Y = \rho W Y + X\beta + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. In reduced form, this is equivalent to

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\epsilon. \tag{2}$$

Let us recall a few classical facts about this model. The mean of Y in this model is given by

$$\mu = (I - \rho W)^{-1} X\beta$$

and its covariance structure by

$$\Sigma = [(I - \rho W')(I - \rho W)]^{-1}\sigma^2, \tag{3}$$

The precision matrix Q is then easily derived

$$Q = \Sigma^{-1} = \frac{1}{\sigma^2}(I - \rho W')(I - \rho W) \tag{4}$$

If ρ is known, the best linear unbiased estimator (BLUE) of $\mu = (I - \rho W)^{-1} X\beta$ is $\hat{\mu} = (I - \rho W)^{-1} X\hat{\beta}$.

We will distinguish two types of prediction situations: the in-sample and out-of-sample cases. In the in-sample prediction problem, we have n spatial units for which we observe the dependent variable Y as well as the independent variables X and we want to predict the value of Y at the observed sites after fitting the model which is the same as computing the fitted value of Y. These predicted values can be used for example to compute a goodness of fit criterion. This situation is illustrated in the left part of Figure 2. In the out-of-sample case, we have two types of spatial units: the in-sample units for which we observe the dependent variable Y_S as well as the independent variable X_S and the out-of-sample units for which we only observe the independent variable X_0 and we want to predict the variable Y_O from the knowledge of Y_S, X_S and X_O. This situation is illustrated in the right part of Figure 2. In the out-of-sample case, we will further distinguish according to the number of spatial units to be predicted simultaneously: if there is only one such unit, we will talk about a single out-of-sample prediction case, otherwise about a multiple out-of-sample prediction case.
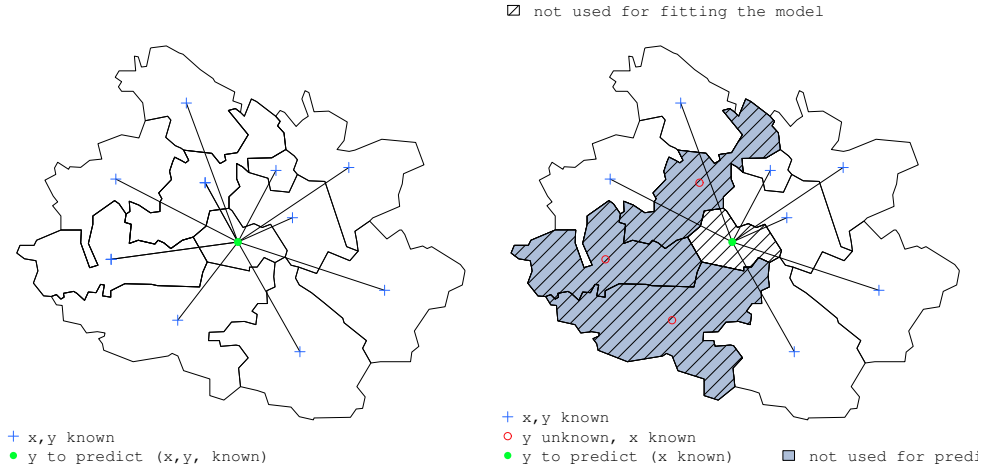
Figure 2: In-sample (left) and and out-of-sample (right) single prediction problem. Shaded areas are sample units which are not used at the model fitting stage. Crosses are in-sample units, empty circles are out-of-sample units and full circle is the point to predict.

2.2 Submodels for in-sample and out-of-sample units

We need to relate the model driving the in-sample units to the out-of-sample ones and we assume there is an overall model driving the in- and out-of sample units. As in Kato (2008), we partition X and Y in $X = (X_S, X_0)$ and $Y = (Y_S, Y_0)$ where X_S (resp Y_S) denote the vector of components of X corresponding to in-sample spatial units and X_O (resp Y_O) denote the vector of components of X corresponding to out-of-sample spatial units. Let n_O and n_S denote respectively the number of out-of sample and in-sample units with $n = n_O + n_S$. Similarly, we partition the spatial weights matrix W as follows

$$W = \begin{pmatrix} W_S & W_{SO} \\ W_{OS} & W_O \end{pmatrix}, \tag{5}$$

where

- W_S is the $n_S \times n_S$ submatrix corresponding to the neighborhood structure of the n_S in-sample sites,

- W_O the $n_O \times n_O$ submatrix corresponding to the neighborhood structure of the n_O out-of-sample sites,

- W_{OS} the $n_O \times n_S$ submatrix indicating the neighbors of the out-of-sample units among the in-sample units

- W_{SO} the $n_S \times n_O$ submatrix indicating the neighbors of the in-sample units among the out-of-sample units.

4

We assume that there is a model corresponding to (1) with a normalized matrix W for the n observations of (X, Y) and that the sub-model M_S driving the vector X_S, Y_S corresponding to the sample units has the same expression but using sub-matrix W_S renormalized. This natural assumption leads to two constraints : one comes from the mean $((I - \rho W)^{-1} X)_S = (I - \rho W_S)^{-1} X_S$ and the other comes from the variance $(var(Y))_S = var(Y_S)$. It is important to note that while a corresponding decomposition of the precision matrix is easily derived from (4), the covariance matrix for sub-model M_S on the other hand is not an extraction of Σ because of the inversion in formula (3).

2.3 Classical prediction formulas

2.3.1 Goldberger formula

Goldberger (1962) proposed a formula for prediction in the framework of a general linear model $Y = \mu + V$ with **known** V. It is well known that

$$\mathbb{E}(Y_O \mid Y_S) = \mathbb{E}(Y_O) + Cov(Y_O, Y_S) Var(Y_S)^{-1} (Y_S - \mathbb{E}(Y_S))$$

and the Golberger formula (1962) gives the BLUP

$$Y_O^* = \hat{\mu}_O + Cov(Y_O, Y_S) Var(Y_S)^{-1} (Y_S - \hat{\mu}_S),$$

where $Y_O^* = \lambda' Y_S$ minimizes $\mathbb{E}(Y_O^* - Y_O)^2$ under the constraint that $\mathbb{E}(Y_O^* - Y_O) = 0$ and where $\hat{\mu}_O$ and $\hat{\mu}_S$ are estimators of respectively $E(Y_O)$ and $E(Y_S)$. In practice, one does not know V (i.e. ρ in the LAG model) and needs to replace it in the formula by an estimator. To simplify, by a slight abuse of language, we will call the ensuing predictor BLUP as well since the theoretical BLUP has no practical interest. It is the application of this formula which has given rise to the famous Kriging predictor in geostatistics. In fact Golberger (1962) gave the formula for a set O reduced to a point but the formula remains true for a set of points O . In that case the problem is to find $Y_O^* = \Lambda' Y_S$ minimizing $\text{Tr}(\mathbb{E}(Y_O^* - Y_O)(Y_O^* - Y_O)')$ under the constraint that $\mathbb{E}(Y_O^* - Y_O) = 0$ where Λ is a matrix. Note that the matrix formulation is equivalent to applying the Goldberger formula one point at a time.

2.3.2 In-sample prediction

In an ordinary linear model which is model (1) for $\rho = 0$, the best linear unbiased predictor (BLUP) of Y_S coincides with the best linear unbiased estimator (BLUE) of μ and is given by

$$\hat{Y}_S^T = X_S \hat{\beta} = \hat{\mu}, \tag{6}$$

where $\hat{\beta}$ is the estimator of β calculated by fitting the model with in-sample units. One could imagine using this predictor in a LAG model and we would call it the "trend predictor".

Based on the equality between BLUE and BLUP for the OLS model, it is then easy and natural to imagine a predictor for the general case $\rho \neq 0$ which we will call the

"trend corrected predictor" given by

$$\hat{Y}_S^{TC} = (I - \hat{\rho}W)^{-1}X_S\hat{\beta}, \tag{7}$$

where $\hat{\beta}$ and $\hat{\rho}$ are the estimators of β and ρ calculated by fitting the model with in-sample units. This predictor is used for example in the LeSage matlab toolbox for computing the in-sample predicted values. Note however that this one does not possess any kind of optimality property.

Another predictor introduced by Haining (1990) and detailed by Bivand (2004) is given by

$$\hat{Y}_S^{TS} = X_S\hat{\beta} + \hat{\rho}WY_S \tag{8}$$

Thereafter, we call this predictor the "trend-signal-noise" predictor. This one is used in the Bivand R package **spdep**. Note that if $\hat{\rho} = 0$, then these three predictors are all equal.

If we had $\hat{\rho} = \rho$ and $\hat{\beta} = \beta$, we would get

$$\mathbb{E}(\hat{Y}_S^{TC}) = \mathbb{E}(\hat{Y}_S^{TS}) = \mathbb{E}(Y_S)$$

Another version of the Goldberger formula used in the framework of conditional autoregressive CAR models (see for example Guyon and Gaetan (2008) for in-sample prediction is the following

$$\hat{Y}_S = \hat{\mu}_S - \text{Diag}(Q_S)^{-1}[Q_S](Y - \hat{\mu}) \tag{9}$$

where $\text{Diag}(Q_S)$ denotes the diagonal matrix containing the diagonal of the precision matrix Q_S and $[Q_S] = Q_S - \text{Diag}(Q_S)$. Again, in practice, ρ is unknown and must be substituted by $\hat{\rho}$.

In the framework of the LAG model, the same arguments yield the following version of Goldberger formula

$$\hat{Y}_S^{BP} = (I - \rho W_S)^{-1}X_S\hat{\beta} - \text{Diag}(Q_S)^{-1}[Q_S](Y - (I - \rho W_S)^{-1}X_S\hat{\beta}), \tag{10}$$

where $Q_S = \frac{1}{\sigma^2}(I - \rho W_S')(I - \rho W_S)$. Note that since this second version of Goldberger is based on the precision matrix rather than the covariance matrix, it should be preferred to the first one for the LAG model. Using a coordinate formulation rather than a matrix form, this formula is equivalent to

$$\hat{Y}_i^{BP} = \hat{\mu}_i - \sum_{j=1, j \neq i}^{n} \frac{q_{ij}}{q_{ii}}(Y_j - \hat{\mu}_j), \tag{11}$$

where q_{ij} is the (i, j) element of matrix Q_S and $\hat{\mu}_i$ are the components of $\hat{\mu}$ given by (6).

2.3.3 Out-of-sample prediction

The trend-signal-noise predictor $\hat{Y}^{TS}$ cannot be defined in the case of out-of-sample prediction since it requires some values of Y_0 which are unobserved. However in the case of a single prediction it is possible to compute it because of the zeros on the diagonal of W.

The trend-corrected strategy can be applied here because it only involves the values of X for the out-of-sample units

$$\hat{Y}^{TC} = (I - \hat{\rho}W)^{-1}\hat{\beta}X = \begin{pmatrix} \hat{Y}_S^{TC} \\ \hat{Y}_O^{TC} \end{pmatrix} \tag{12}$$

and

$$\hat{Y}_O^{TC} = -(D - CA^{-1}B)^{-1}CA^{-1}X_S\hat{\beta} + (D - CA^{-1}B)^{-1}X_O\hat{\beta} \tag{13}$$

for $(I - \hat{\rho}W) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_{n-p} - \hat{\rho}W_S & -\hat{\rho}W_{SO} \\ -\hat{\rho}W_{OS} & I_p - \hat{\rho}W_O \end{pmatrix}$

Kelejian and Prucha (2007) use Goldberger formula for single out-of-sample prediction in the particular case when $O = \{i\}$ and for $W_{i.}Y$ replacing Y_S. Griffith (2010) proposes an EM procedure combining estimation of spatial parameters and imputation of missing values in the framework of the spatial filtering method (Griffith, 2003).

3 Out-of-sample prediction: extensions and new proposals

3.1 Another formulation of Goldberger formula for LAG models

For out-of-sample best prediction, if we first concentrate on the case of single prediction, formula (11) can be applied with the precision matrix Q corresponding to the sample units augmented with the point to predict.

In the case of out-of-sample best prediction, Harville (1997) derives a Goldberger formula written in terms of inverse matrix Q, similar to the prediction formula for markov gaussian vector field of Rue (2005, page 31). As Pace and LeSage (2008) point out, it is based on the fact that $Cov(Y_O, Y_S)Var(Y_S)^{-1} = -Q_O^{-1}Q_{OS}$, which arises from expressing that the partitioned matrix Q is the inverse of the partitioned matrix $Var(Y)$. The Goldberger formula can thus be expressed in terms of precision matrices as follows

$$\hat{Y}_O^{BP} = \hat{Y}_O^{TC} - Q_O^{-1}Q_{OS} \times (Y_S - \hat{Y}_S^{TC}) \tag{14}$$

with

$$Q = I - \rho(W' + W) + \rho^2 W'W = \begin{pmatrix} Q_S & Q_{SO} \\ Q_{OS} & Q_O \end{pmatrix}$$

Let us note that the matrix to invert is Q_O and has the size of the number of out-of-sample units whereas in the first version of the Goldberger formula, the size of the matrix to invert is equal to the number of in-sample units. If the size of the matrix to be inverted is a crucial point, then using the precision formula instead of the variance one can help.

3.2 Extension of the Kelejian-Prucha predictor

We first propose to generalize the Kelejian-Prucha approach to multiple prediction where Y_O is predicted by linear combination of $W_{OS}Y_S$ instead of Y_S. Golberger formula gives the best predictor as

$$\hat{Y}_O^{BP_W} = \hat{Y}_O^{TC} + \Sigma_{OS}W'_{OS}(W_{OS}\Sigma_S W'_{OS})^{-1}(W_{OS}Y_S - W_{OS}\hat{Y}_S^{TC}) \qquad (15)$$

However we believe that it is unlikely in practical situations that one has the information about the linear combination of neighboring values $W_{OS}Y_S$ without having the entire knowledge of Y_S. Moreover, formula (15) is not simpler to compute than the best prediction given by formula (14): the size of the matrix to invert is equal to the number of out-of-sample units.

For this reason, we propose the following alternative which consists in using the Harville formula for a case where the set S is replaced by N where N is the set of all sites in S which are neighbors in the sense of W of at least one site in O. The idea is to use only the neighbors of the out-of-sample sites (the ones in in O) in order to predict. Let J be the set of such indices and n_J its size. Let $W^{\{J,O\}}$ be the neighborhood matrix for sites which are in S or J.

$$W^{\{J,O\}} = \begin{pmatrix} \overset{\overset{n_J}{\longleftrightarrow}}{W_J} & \overset{\overset{n_O}{\leftrightarrow}}{W_{JO}} \\ \hline W_{OJ} & W_O \end{pmatrix} \begin{matrix} \updownarrow n_J \\ \\ \updownarrow n_O \end{matrix}$$

The corresponding partition of the precision matrix corresponding to sites in $\{J, O\}$ is

$$Q_{\{J,O\}} = I_{n_J+p} - \hat{\rho}(W_{\{J,O\}} + W'_{\{J,O\}}) + \hat{\rho}^2(W'_{\{J,O\}}W_{\{J,O\}}) = \begin{pmatrix} Q_J & Q_{JO} \\ Q_{OJ} & Q_O \end{pmatrix}$$

and thus we get the following predictor

$$\hat{Y}_O^{BP_N} = \hat{Y}_O^{TC} - Q_O^{-1}Q_{OJ}(Y_J - \hat{Y}_J^{TC})_J. \qquad (16)$$

The advantage of this predictor lies in the fact that it reduces the computational burden since the size of the matrix $Q_{OJ}(Y_J - \hat{Y}_J^{TC})_J$ is $n_O \times n_J$ instead of $n_O \times n_S$. If we were using the Goldberger formula, the new predictor would be written

$$Y_O^{BP_N} = \hat{Y}_O^{TC} + Cov(Y_O, Y_J)Var(Y_J)^{-1}(Y_J - \hat{Y}_J^{TC}),$$

Clearly the new predictor is not optimal, but one can hope it has some almost optimality behavior. Our proposition can be related to the classical "kriging with moving neighborhood" which is often used in geostatistics. In the same spirit, Kato (2008) considers the feasible generalized least square predictor in the framework of SEM models. Note that because of the links between W and Q, if we consider $W'W$-neighbouring, that is order 2 W-neighbouring, the predictor will be optimal and is equal to the predictor with Q-neighbours. Indeed the reason is that if we look at prediction with a set of Q-neighbours then it means that Q can be written :

$$Q = \left(\begin{array}{c|c} Q_{S \setminus J} & 0 \\ \hline 0 & Q_{\{J,O\}} \end{array} \right)$$

and thus $Q_O^{-1} Q_{OS} \times (Y_S - \hat{Y}_S^{TC})$ is equal to $Q_O^{-1} Q_{\{J,O\}} \times (Y_{\{J,O\}} - \hat{Y}_{\{J,O\}}^{TC})$ and therefore is optimal.

3.3 Alternative: back to single prediction

Because the single prediction formulas are simpler, when p out-of-sample units have to be predicted, we propose to apply the "single out-of-sample" formula to each of the out-of-sample unit separately, ignoring at each stage the remaining $p - 1$ units. This allows also to include the Trend-signal strategy which exists out-of-sample only in the single prediction case. This leads us to defining alternatives of each of the five predictors $\hat{Y}^{TC}$, $\hat{Y}^{TS}$, $\hat{Y}^{BP}$, $\hat{Y}^{BP_W}$ and $\hat{Y}^{BP_N}$ which will be denoted respectively by $\hat{Y}^{TC^1}$, $\hat{Y}^{TS^1}$, $\hat{Y}^{BP^1}$, $\hat{Y}^{BP_W^1}$ and $\hat{Y}^{BP_N^1}$. These formulae of course do not apply if an out-of-sample point has no neighbors among the sample units but in that situation a non-spatial formula is doing just as well.

4 Comparing the predictors by simulation

4.1 Simulation framework

In order to compare the different predictors, we design a simulation study. We use the Midi-Pyrénées region divided into $n = 283$ cantons for our study region. We construct a weight matrix W using the 10 nearest neighbors scheme (distance is based on the distance between centroids of the cantons).
We simulate three explanatory variables as follows:

- $X_1 \sim \mathcal{N}(15, 3)$

- $X_2 \sim \mathcal{B}(100, 0.45)/100$

- $X_3 \sim \log(\mathcal{U}_{[0,283]})$

In order not to restrict attention to gaussian distributions, the choice of the second distribution is motivated by its bounded support and the choice of the third by its

right skewness. We use the following spatial autoregressive lag regression model to generate the dependent variable

$$Y = (I - \rho W)^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon) \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2) \qquad (17)$$

The parameter β and σ are fixed to $\beta = (0, 1/4, 6, 1)$ and $\sigma = 1$. For the in-sample comparison, ρ takes a range of values $\rho = 0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.9$. For the out-of-sample comparison, ρ is equal to 1.

4.2   In-sample prediction simulation results

In this section, the sample contains the 283 initial sites described in section 4.1. For each choice of ρ and σ, we draw 500 samples of the model and we compute the maximum likelihood estimates of the parameters based on the in-sample locations and the corresponding predictions. We use the total mean square error of prediction $MSE_k = \frac{1}{n}\sum_i^n (y_i - Y_i^k)^2$ for each method $k = TS, TC, BP$ to compare the quality of the predictors. Note that this criterion includes the statistical error due to parameter estimation. The results of the in-sample comparison are in Table 1. The

	$M\tilde{S}E_{BP}$	$M\tilde{S}E_{TS}$	BP/TS	$M\tilde{S}E_{TC}$	BP/TC
$\rho = 0.05$	0.9707	0.9720	0.9986	0.9754	0.9952
	(0.0832)	(0.0833)		(0.0838)	
$\rho = 0.2$	0.9850	0.9884	0.9966	1.0006	0.9844
	(0.0832)	(0.0835)		(0.0852)	
$\rho = 0.35$	0.9646	0.9756	0.9897	1.0192	0.9464
	(0.0847)	(0.0841)		(0.0896)	
$\rho = 0.5$	0.9597	0.9814	0.9779	1.0890	0.8813
	(0.0799)	(0.0803)		(0.1039)	
$\rho = 0.65$	0.9494	0.9883	0.9606	1.2531	0.7576
	(0.0790)	(0.0799)		(0.1450)	
$\rho = 0.8$	0.9308	0.9871	0.9429	1.6571	0.5660
	(0.0844)	(0.0848)		(0.2738)	
$\rho = 0.9$	0.9152	0.9878	0.9265	2.8981	0.3158
	(0.0784)	(0.0812)		(0.9635)	

Table 1: MSE for different predictors and comparison with Best predictor when the parameter ρ takes different values from 0.05 (mild correlation) to 0.9 (strong correlation).

mean error is stable across values of ρ for TS, is increasing for TC and decreasing for BP. Variances are stable. The efficiency ratio BP/TS is decreasing with spatial correlation but remains close to 1 whereas the efficiency ratio BP/TC decreases dramatically with ρ. We do not report results for different values of σ because they do not reveal any variation with respect to this parameter.

4.3 Out-of-sample prediction simulation results

To evaluate the performance of the different predictors for the out-of-sample case, we use the same model as before to generate the samples. The number of replications is 1000 and we report the average mean square error of prediction over the out-of-sample units.

We choose at random a given number of sites (27 or 54) which will be declared out-of-sample (in O). We predict the Y variable on the out-of-sample locations based on the sample S constituted by the remaining sites. We consider several situations depending upon the number of out-of-sample units and upon the aggregation level of the out-of-sample units. The corresponding configurations of out-of-sample units are shown in Figures 3 and 4 and the level of aggregation is increasing from left to right.
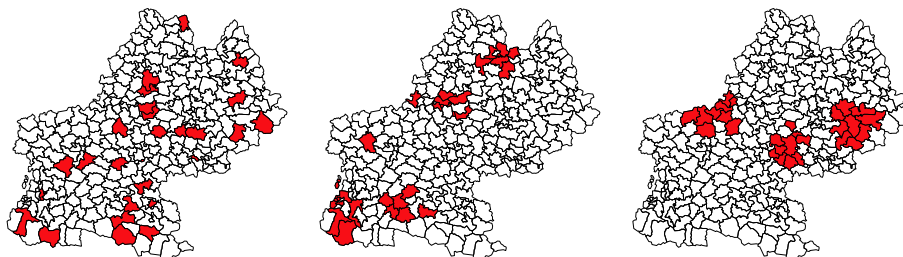

Figure 3: The three configurations for 27 out-of-sample units positions
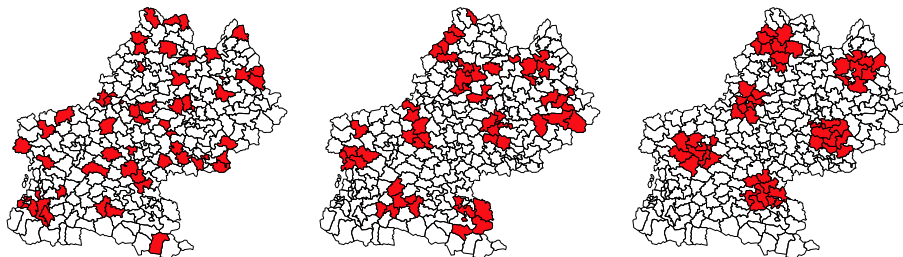
Figure 4: The three configurations for 54 out-of-sample units positions

The results for the case of 27 out-of-sample units are reported in table 2 and those for the case of 54 out-of-sample units are reported in table 3.

Aside BP1 whose rank changes, whatever configurations and number of sites to predict, we obtain the following ranking between methods in decreasing order of efficiency

$$BP \ < \ BP_N \ < \ BP_W \ < \ BP_N^1 \ < \ BP_W^1 \ < \ TS^1 \ < \ TC \ < \ TC^1$$

	BP	BP^1	$\frac{BP}{BP^1}$	TC	$\frac{BP}{TC}$	TC^1	$\frac{BP}{TC^1}$	TS^1	$\frac{BP}{TS^1}$
Conf. 1	0.998	1.000	0.998	1.126	0.886	1.131	0.883	1.026	0.973
	(0.275)	(0.276)		(0.303)		(0.305)		(0.280)	
Conf. 2	1.031	1.048	0.983	1.145	0.900	1.184	0.870	1.069	0.964
	(0.296)	(0.302)		(0.326)		(0.340)		(0.305)	
Conf. 3	1.038	1.060	0.979	1.129	0.919	1.144	0.908	1.064	0.976
	(0.280)	(0.285)		(0.307)		(0.309)		(0.285)	

	BP_W	$\frac{BP}{BP_W}$	$BP_W{}^1$	$\frac{BP}{BP_W{}^1}$	BP_N	$\frac{BP}{BP_N}$	$BP_N{}^1$	$\frac{BP}{BP_N{}^1}$
Conf. 1	1.003	0.996	1.007	0.992	0.999	1.000	1.003	0.996
	(0.276)		(0.277)		(0.275)		(0.277)	
Conf. 2	1.035	0.996	1.057	0.975	1.032	0.999	1.055	0.977
	(0.298)		(0.305)		(0.297)		(0.305)	
Conf. 3	1.041	0.997	1.065	0.975	1.039	1.000	1.063	0.977
	(0.281)		(0.286)		(0.281)		(0.285)	

Table 2: Simulation results for the 27 out-of-sample units case

Note that the worst ratio is around 0.88.

Because the reported prediction errors are averages over out-of-sample units, we suspected it may hide different situations depending on the number of missing neighbors of a given out-of-sample unit. The following table 4 reports the prediction errors as a function of the number of missing neighbors. This number k ranges from 0 to 9 and for each k, we repeat 1000 times the following process

- choose a site i at random

- remove k neighbors at random from the neighbors of i, their set is N

- the in-sample set of sites becomes $S \setminus N$ and the out-of-sample set of sites is N

- simulate the vector Y for all the sites

- predict the Y on the sites in N and compute the prediction error.

The first column of the table contains the predictive mean square error (PMSE) of the BP predictor and the remaining ones report the ratio of the optimal PMSE with the PMSE of all the other methods.

We observe that the BP predictive mean square error indeed slightly increases with the number of missing neighbors. The efficiency of BP1 and TC1 with respect to BP decreases with the number of missing neighbors. The efficiency of TC with respect to BP increases with the number of missing neighbors which we interpret as revealing the fact that when the information gets poor in the neighborhood, it is just as well to use the mean to predict (the correction is inefficient). The efficiency of BP_W with respect to BP remains stable.

	BP	BP^1	$\frac{BP}{BP^1}$	TC	$\frac{BP}{TC}$	TC_1	$\frac{BP}{TC^1}$	TS^1	$\frac{BP}{TS^1}$
Conf. 1	1.009	1.012	0.997	1.130	0.893	1.140	0.886	1.035	0.975
	(0.196)	(0.196)		(0.221)		(0.225)		(0.201)	
Conf. 2	1.029	1.036	0.992	1.137	0.905	1.152	0.893	1.054	0.975
	(0.199)	(0.199)		(0.226)		(0.230)		(0.204)	
Conf. 3	1.037	1.061	0.978	1.136	0.913	1.158	0.896	1.069	0.970
	(0.205)	(0.213)		(0.234)		(0.240)		(0.214)	

	BP_W	$\frac{BP}{BP_W}$	$BP_W{}^1$	$\frac{BP}{BP_W{}^1}$	BP_N	$\frac{BP}{BP_N}$	$BP_N{}^1$	$\frac{BP}{BP_N{}^1}$
Conf. 1	1.012	0.997	1.017	0.992	1.010	0.999	1.015	0.994
	(0.196)		(0.196)		(0.196)		(0.196)	
Conf. 2	1.031	0.998	1.042	0.987	1.029	1.000	1.039	0.989
	(0.200)		(0.201)		(0.200)		(0.201)	
Conf. 3	1.040	0.997	1.070	0.970	1.038	0.999	1.068	0.971
	(0.206)		(0.214)		(0.206)		(0.215)	

Table 3: Simulation results for the 54 out-of-sample units case

5   Conclusion

At least in the case of this particular model, the performance of BP_N, BP_W,BP_N1,BP_W1 are very close to that of the best prediction and much better than that of TC, TS, TC1,TS1. We did not consider a larger variety of parameter values because a few attempts have shown that the results were quite stable.

For the in-sample case, the performance of the trend-signal-noise predictor is not so bad and it is very easy to compute. BP_N is better than BP_W in terms of efficiency but BP_W is closer to BP in terms of projection coefficients. BP_W is better than TC, less good than TS.

We developed our study on the case of the LAG model. The conclusions would apply for the Spatial Durbin model :

$$y = \rho W y + \alpha 1 + X\beta + WX\gamma + \varepsilon$$

with $\varepsilon \sim \mathrm{N}(0,\ \sigma^2\ \mathrm{I})$ because the structure of the variance matrix is the same as in the LAG case and the structure of the mean is similar $(I - \rho W)^{-1} Z\beta_1$.

For the case of the spatial error model SEM which is a linear model with LAG residuals, we refer the reader to Kato (2008).

nb mis. neib.	$PM\bar{S}E_{BP}$	BP/BP^1	BP/TC	BP/TC^1	BP/TS^1	BP/BP_W
0	0.979 (1.374)	1.000	0.898	0.898	0.984	0.998
1	0.987 (1.377)	0.989	0.905	0.887	0.972	0.997
2	0.986 (1.364)	0.987	0.902	0.884	0.973	1.000
3	1.015 (1.420)	0.981	0.924	0.901	0.978	0.998
4	0.996 (1.379)	0.971	0.909	0.871	0.958	0.989
5	1.002 (1.409)	0.959	0.913	0.850	0.947	0.991
6	1.015 (1.401)	0.930	0.923	0.850	0.930	0.995
7	1.042 (1.456)	0.902	0.944	0.844	0.914	1.002
8	1.036 (1.435)	0.798	0.939	0.756	0.806	0.989
9	1.059 (1.434)	0.694	0.957	0.706	0.709	0.993

Table 4: Prediction errors as a function of number of missing neighbors

References

[1] N. Cressie (1990), The origins of kriging, Mathematical Geology Volume 22, Number 3 (1990), 239-252.

[2] R. Bennet, D.Griffith and R. Haining (1989), Statistical analysis of spatial data in the presence of missing observations: an application to urban census data, *Environment and Planning A*, 21: 1511-1523.

[3] R. Bivand (2002) Spatial econometrics functions in R: Classes and methods. *Journal of Geographical Systems*, 4, 405-421.

[4] C. Gaetan and X. Guyon (2008), Modlisation et statistique spatiales, Springer Verlag.

[5] A.S. Goldberger (1962) Best linear unbiased prediction in the generalized linear regression model. *JASA*, 57,298, 369-375.

[6] Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*, Springer.

[7] D. Griffith (2010) Spatial filtering and missing georeferenced data imputation: a comparison of the Getis and Griffith methods, in *Perpectives on spatial data analysis,* L. Anselin and S.J. Rey (eds), Springer-Verlag.

[8] R.P. Haining (1990) *Spatial data analysis in the social and environmental sciences*, Cambridge university Press.

[9] Harville D.(1997) Matrix algebra from a statistician's perspective, Springer-Verlag, New-York.

[10] T. Kato (2008) A further exploration into the robustness of spatial autocorrelation specifications, *Journal of Regional Science*, 48 (3), 615-638.

[11] H.H. Kelejian and I.R. Prucha (2007) The relative efficiencies of various predictors in spatial econometric models containing spatial lags, *Regional Science and Urban Economics*, 37, 363-374.

[12] J.P. LeSage and R.K. Pace (2004) Models for spatially dependent missing data, *Journal of Real Estate Finance and Economics*, 29:2, 233-254.

[13] R. Kelley Pace and James P. LeSage (2008), Spatial Econometric Models, Prediction, Encyclopedia of Geographical Information Science, Shashi Shekhar and Hui Xiong (eds.), Springer-Verlag, 10.1007/978-0-387-35973-1-1266.

[14] J.P. Lesne, H. Tranger, A. Ruiz-Gazen and C. Thomas-Agnan (2008) Predicting population annual growth rates with spatial models, Preprint.

[15] H. Rue and L. Held (2005), Gaussian Markov Random Fields, Theory and Applications, Chapman & Hall/CRC.

[16] Wainwright, M., J. and Jordan, M., I. (2008), Graphical Models, Exponential Families, and Variational Inference, Foundations and Trends R in Machine Learning, Vol. 1, Nos. 162 (2008) 16305.