AUTOREGRESSIVE SPATIAL INTERPOLATION
AND AGRICULTURAL MODELLING:
A PROGRAMMING APPROACH

Michiel A. KEYZER[*]

Abstract - *Research in the field of agricultural policy often calls for an explicit representation of geographical aspects of the country or region under consideration. Furthermore, in connection with environmental issues one has to account for spatial interactions, for example in connection with water flows or erosion processes. However, this inevitably leads to the problem that all data needed for estimation are not available at all points. The paper addresses this issue and presents, for a spatial model, an interpolation procedure by maximum likelihood estimation. In this model, the spatial interactions are represented through an autoregressive scheme whereby the value of a given endogenous variable at one point depends among others on the value of the same variable at adjacent points. This procedure is referred to as an (autoregressive spatial) interpolation because the model yields, among others interpolation 'forecasts' for missing data. To effectuate the estimation, the paper suggests the use of a constrained mathematical program, rather than of a standard statistical routine, as this enables the user to keep a more clear distinction between likelihood function, the autoregressive model, the constraints on parameters and the restrictions implied by the available data.*

---

[*] Centre for World Food Studies (SOW-VU), *Vrije Universiteit* Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands.

INTRODUCTION

There is an increasing recognition of the need for an explicit representation of geographical aspects within agricultural policy studies. One reason for this is that land utilization and land cover change are perceived as key elements in the process of global climate change, another that the analysis of environmental issues e.g. those related to excessive fertilization and water pollution needs spatial referencing because the damages strongly depend on the natural conditions at micro-level (IGBP/HDP 1995).

The quantitative models used in agricultural policy studies usually introduce the geographical dimension by describing crop and livestock production at the national, regional or farm level. The standard practice of making agricultural policy models more explicit geographically is to refine the level of regional aggregation from say, the provincial level to the district or village level. In such studies, spatial interaction is at best incorporated on the economic side, via transportation flows of commodities, migration of labour and market clearing processes. At the same time these models do not incorporate any spatial interaction at the biophysical level: the biophysical relationships for crop and livestock production that apply in one location do not contain any (endogenous or exogenous) variables from other locations.

Indeed, as long as the natural conditions remain constant and do not have to be quantified explicitly, there is little need to account for such interdependence, as agricultural output in one location will only depend on the inputs applied by the farmers of that same location. However, once pollution or climate change are thought to affect the natural conditions under which crops and livestock have to grow, via winds, water flows etc., the spatial interactions in the biophysical sphere have to be addressed and this calls for a representation within a spatial model, not only a spatial database. However, this type of model will be very demanding in terms of data requirements, and this requires methods of model estimation that can deal with spatially incomplete data sets, i.e. that can deal with spatial interpolation.

It will be the objective of the present paper to specify a procedure for estimating, by a maximum likelihood (ML-) method, a spatial model of physical interactions, for such a case of missing observations. While this spatial model will be cast in a general form, it will cover the special case of a yield function for a given crop, that relates harvestable output per hectare at a particular location to both the inputs that are applied by the farmer and the natural conditions that

prevail at this location, which are not independent from those in the neighbourhood. The general model will be given an autoregressive form, where the value of a given variable at one location depends among others on the value of the same variable at adjacent locations.

To solve the optimization problem of this ML-estimation, we work directly with a constrained mathematical program rather than reducing the problem to a set of first-order conditions, which are eventually, after elimination of as many variables as possible, solved as a system of nonlinear equations with the parameters to be estimated as unknowns. Such a 'concentrated' approach would have the merit of reducing the dimensionality of the problem but has as it drawback a in modularity and sparseness of representation. Concentration yields formulas of great complexity, where matrix inverses abound (see e.g. Cressie, 1993, chapter 3) and where sparseness of representation is lost because the inverse of a sparse matrix is usually non-sparse. In contrast, a programming approach makes it possible to maintain a clear distinction between the likelihood function, the autoregressive model, the constraints on parameters and the restrictions implied by the available data. It will be shown that a wide array of model specifications can be accommodated in this manner.

The paper proceeds as follows. In section 1 we introduce the general form of the autoregressive spatial model we shall be concerned with. Section 2 introduces four extensions: (i) monotonic (e.g. logarithmic) transformations of the endogenous variable, (ii) the case where the covariance of error terms is unknown, (iii) the treatment of aggregate information say, on the regional average over a set of points, and (iv) the representation of endogenous variables that are not, or only in part, spatially correlated. Section 3 is devoted to a comparison with two competing approaches, co-kriging and generalized least squares which are widely used in geo-statistics. Section 4 sketches a possible application to the modelling of agricultural supply within a spatial setting. Section 5 concludes. Three aspects of practical implementation are considered in the Annex: imposing restrictions on parameters, solving the program and computing the asymptotic covariance matrix.

1. GENERAL FORMULATION

1.1. The spatial model

Suppose that we are given a set of observations on various variables sampled at geographically referenced points (by latitude and longitude). These variables measure altitude, temperature, radiation, length of growing season, agricultural yields, fertilizer applications etc. Their coverage over the map of the

country or region under study is uneven and differs across variables.

We distinguish R sites, indexed r and J exogenous site-related variables x_{jr}, indexed j and r. Among the site-related variables we take $j = 1, 2, 3$ to denote latitude, longitude and altitude, respectively and $j = 4$ as the unit element for the regression constant. We assume that data are available for all x_{jr}. We also define K endogenous variables q_{kr} at every point r as well as the observation set G and observations y_{kr} of q_{kr} that are supposed to be available for all kr $\square$ G. Finally, we introduce the error variable u_{kr} defined for kr $\square$ G. The simplest model will abstract from spatial autocorrelation and be written as:

$$q_{kr} = \Sigma_j B_{kj}x_{jr} \qquad\qquad \text{for all kr} \qquad\qquad (1.1a)$$
$$y_{kr} = q_{kr} + u_{kr} \qquad\qquad \text{for all kr} \square \text{ G} \qquad\qquad (1.1b)$$

The estimation problem is now to find a numerical value for the parameters B that provides a good fit to the data which are available   (say, minimizes $\Sigma_{kr\square G}\ u_{kr}^2$). Once this value has been computed, interpolation merely amounts to calculating q_{kr} for all kr $\square$ G, using (1.1a).

Model (1.1) is common in trend surface analysis, where the set of exogenous variables usually includes various nonlinear transformations of the original data as well as higher order polynomials. However, this type of regression is often found unsatisfactory because a good fit requires such a large number of parameters that significance of the estimates becomes very low. Autoregressive models tend to offer greater flexibility for fewer parameters. The basic spatial autoregressive model may be viewed as an extension of (1.1) and reads as:

$$q_{kr} = \Sigma_{k'r'}A_{kr,k'r'}q_{k'r'} + \Sigma_j B_{kj}x_{jr} \qquad\qquad \text{for all kr} \qquad\qquad (1.2a)$$
$$y_{kr} = q_{kr} + u_{kr} \qquad\qquad\qquad \text{for all kr} \square \text{ G} \qquad (1.2b)$$

where, for the diagonal, $A_{kr,kr} = 0$. For the autoregressive term $\Sigma_{k'r'}A_{kr,k'r'}q_{k'r'}$ we shall assume that every q_{kr} depends on $q_{k'r}$ for k′-values other than k and on $q_{kr'}$ for selected r′-values (neighbouring locations). Further details will follow in section 2.1 below. Note that, due to the autoregression, the value of q_{kr} depends on the missing values for endogenous variables. Therefore, the issues of parameter estimation and interpolation will have to be addressed simultaneously. We can now write (1.2) in matrix form as:

$$q = Aq + Bx \tag{1.3a}$$
$$y = Tq + u \tag{1.3b}$$

where A is the $KR\infty KR$-matrix with elements $A_{kr,k'r'}$ and B the $KR\infty JR$-matrix with elements B_{kj} for every r. Note that the variables in the constraint do not carry the subscript r.[1] This enables us to represent cross effects among variables at a given point, as well as cross effects among points (spatial effects). The known matrix T describes the data structure by mapping endogenous variables to observations. In its simplest form, it has: $T_{kr,kr} = 1$ for $kr \square G$ and 0 otherwise. In section 3.3 we give a more general interpretation. The row dimension of T and the dimension of the vectors u and y correspond to the number of members of the set G. If all observations are missing for a given variable, the variable is said to be latent. We complete the specification of the basic model by allowing for a more general specification of errors:

$$y = Tq + Pu \tag{1.3b'}$$

where P is a given square nonsingular matrix. It is important to note that in this model the error Pu only bridges the gap between the observation y and the model generated value Tq. In (1.3a), there is no error on q itself. Assuming [I-A] to be nonsingular, we may write this equation in explicit form as:

$$q = [I-A]^{-1}Bx \tag{1.4}$$

Note that the matrix A can contain up to $(K\infty R)\infty(K\infty R)$ unknown coefficients, which exceeds by far the number of observations that are available. Hence, some structure has to be imposed on this matrix. This can be achieved by requiring equality among coefficients, or by treating the effect as a function of distance, or by allowing for effects from direct neighbours only. Thus, A and B will be treated as given functions of parameters θ. Yet if these functions are not well specified, some of the coefficients in θ and possibly some of the interpolated values q_{kr} will remain unidentified.

---

[1] The autoregression problem in space differs from the one in time in at least three respects: (i) the time-axis is one-dimensional but space is two- or three-dimensional; (ii) along the time-axis observations are usually available at regular intervals (weeks, months, years), whereas in space the grid of points where data are collected is irregular and differs across variables; (iii) since the time-axis has a natural orientation from past to future, points for which observations are available are located in one subset (the past) that is well separated from the points where this is not the case (the future); in a spatial context the two are in general intermixed.

1.2. Least squares estimation

Least squares estimation of the parameters θ would lead to the program:

$$\min_{q,\, u,\, \theta} \tfrac{1}{2} u'u$$

subject to   (1.5)

$$[I-A(\theta)]q = B(\theta)x$$
$$y = Tq + Pu$$
$$\theta \in \Theta$$

where $'$ denotes the transpose and the variables below the maximand are the choice variables; y does not appear in the list, since it is given. When fitting a model to available data, several criteria other than the sum of squared errors could be chosen, for example the sum of absolute values of errors. Yet least squares has the particular virtue of yielding parameter estimates that, under narrowly defined conditions [i.c. u is normally distributed as $u \sim N(0, \sigma^2 I)$], coincide with those obtained from maximum likelihood (ML-) estimation.

1.3. Maximum likelihood estimation

However, if these conditions do not hold, as is often the case in applications, one has to deal with the maximum likelihood problem itself, which may also be written as a mathematical program but has a more complicated objective function. We review the main steps of the ML-approach in relation to autoregression and interpolation (see e.g. Greene, 1991) for a more general treatment). In the standard case, the ML-approach starts from:

(i) a given observation y,
(ii) a deterministic model $y^s = F(x, \theta)$ with unknown parameters $\theta \in \Theta$,
(iii) an error term $e = y - y^s$ with a postulated likelihood density function $L(e; \theta)$ which assigns a likelihood to an error e, given a value θ for the parameters).

If the assumed distribution for e is normal with mean zero and given covariance matrix Ω, the θ in $L(e; \theta)$ can be dropped. In terms of (1.5), one may write $e = Pu$, where $u \sim N(0, \sigma^2 I)$, and, therefore, $e \sim N(0, \sigma^2 PP')$. The aim of ML-estimation is then to find a value for θ that maximizes the likelihood of the observation y i.e. maximizes the density $L(y; y^s, \theta)$ which corresponds to the density $L(e; \theta)$ of e. Since in (iii) above, the error term was assumed to be additive,

the relation between the two is straightforward and $\mathrm{L}(y; y^s, \theta) = L(y - y^s; \theta)$. Thus, the program for ML-estimation may be written as:

$$\max_{y^s, e, \theta} L(e; \theta)$$

subject to   (1.6)

$$y^s = F(x, \theta)$$
$$y = y^s + e$$
$$\theta \in \Theta$$

Besides the difference in objective and error term, program (1.5) differs from (1.6) in two respects. First, (1.5) has *autoregression*. This can be incorporated in the ML-program by specifying the (deterministic) model as an implicit function $H(y^s, x, \theta) = 0$; it will be assumed that, at all values $\theta \in \Theta$, the implicit function $H(\supseteq)$ meets the requirements of the implicit function theorem and defines the explicit, function $y^s(x, \theta)$, which is continuously differentiable in θ. Such a requirement was already reflected in the nonsingularity assumption for $[I-A(\theta)]$. Secondly, in (1.5) observations may be missing, as represented via the matrix T. To incorporate this in the deterministic model, we replace $H(y^s, x, \theta)$ by $H(q, x, \theta)$ and the ML-program becomes:

$$\max_{q, y^s, e, \theta} L(e; \theta)$$

subject to   (1.7)

$$H(q, x, \theta) = 0$$
$$y^s = Tq$$
$$y = y^s + e$$
$$\theta \in \Theta$$

As a third extension, let us briefly investigate the implications of accounting for *errors in independent variables*. So far, the vector x denotes the exogenous or independent variables, i.e. the variables not explained by the model, and is assumed to be nonstochastic. Clearly, for items such as latitude, longitude or altitude, such an assumption seems appropriate but for other independent variables, errors of measurement may have to be considered.

The situation with errors in the independent variables may be represented by treating q as referring to all stochastic variables, both the dependent and the independent ones. Now y^s will no longer be determined uniquely by the deterministic model (for given values of x and θ), as the number of stochastic

variables will for a well specified form exceed the number of equations. This makes it impossible to derive the explicit form of the likelihood function for y from the assumed likelihood of u and the constraints. Two approaches may be followed to overcome this difficulty. The first and most common one is to restore uniqueness by defining a data generating process (additional equations) for the independent variables that have errors. This is usually implemented either by specifying some autocorrelation process (with y_{kr} at point r assumed to be dependent on its value at other points), or via instrumentalization (treating y_{kr} as dependent on x_{jr} or on $y_{k'r}$ at the same point). The second approach treats the systematic part of the independent variables as a parameter that is to be estimated. A distinction is often made between dependent variables y, observed independent variables with errors z and given, nonstochastic values x. The error term v on z is usually assumed to be uncorrelated with the error term e on y. Program (1.7) may now be rewritten with the sum of the log-likelihoods $L_e(e; \theta)$ of e and $L_v(v; \theta)$ of u as objective:

$$\max \ L_e(e; \theta) + L_v(v; \theta)$$
$$q, v, y^s, z^s, e, \theta$$

subject to   (1.8)

$$H(q, z^s, x, \theta) = 0$$
$$y^s = Tq$$
$$y = y^s + e$$
$$z = z^s + v$$
$$\theta \ \Box \ \Theta$$

Note that for any given feasible triple (z^s, x, θ), the constraint $H(q, z^s, x, \theta) = 0$ of this program will define the continuously differentiable function $y^s(z^s, x, \theta)$. Thus, the likelihood function satisfies $\mathsf{L}(y; y^s, \theta) = L(y-y^s; \theta)$ as before. In practical terms, program (1.8) treats the independent variable z as 'soft', allowing some deviation of z^s from z, so as to improve the likelihood $L_e(e; \theta)$, as long as this outweighs the deterioration in $L_v(v; \theta)$. Yet we should mention that, in spite of all these refinements, the above formulations all postulate a given functional form for the likelihood functions $L_e(e; \theta)$ and $L_v(v; \theta)$, without indicating how to arrive at a correct specification. This would require a study of the data generating processes, which falls outside the scope of this paper.

2. THREE EXTENSIONS

2.1. Monotonic transformations

Generalizations of (1.5) such as in programs (1.7) and (1.8) fit relatively easily within the approach discussed so far, because they maintain an additive error term. Modifications in the specification of this error term require an adjustment in the likelihood function. Consider the monotonic transformation $g(y_{kr}, \zeta)$ with unknown parameter ζ, which we denote in vector-form by $g(y, \zeta)$, which may for example be the exponential mapping $\exp(\zeta y)$. The relation between model and data as in (1.3b) is then represented by:

$$g(y, \zeta) = Tq + e \qquad (2.1)$$

This change leads to a modification of the likelihood function for y, since the identity $L(y; y^s, \theta) = L(y - y^s; \theta)$ no longer holds and least squares will not be ML. Let $g'(y_{kr}, \zeta)$ denote the absolute value of derivative of $g(y_{kr}, \zeta)$ with respect to y_{kr}. The concentrated log-likelihood (Greene, 1991, p. 344) now becomes:

$$L_c = \Sigma_{kr \square G} \ln |g'(y_{kr}, \zeta)| - N/2 \, (1+\ln(2\pi)) - N/2 \, \ln(1/N \, u'u) \qquad (2.2)$$

where N is the number of elements kr in G and |.| denotes the absolute value. This function can readily be treated as maximand of a nonlinear program.

2.2. Unknown covariance

Least squares also ceases to be ML if the covariance among error terms is a function $\Omega(\theta)$ with unknown parameters. For $e = P(\theta)u$, one gets $\Omega(\theta) = \sigma^2 \, P(\theta)P(\theta)'$ and:

$$\ln L = - N/2\ln(2\pi) + 1/2\ln \sigma^2 - \ln|P(\theta)| - 1/(2\sigma^2) \, u'u \qquad (2.3)$$

where the term |.| denotes a determinant, which is a cumbersome, nonconcave function. The following linear approximation avoids it:

$$\ln|P(\theta)| \, V \, -(P'(\hat{\theta})\partial P(\hat{\theta})/\partial\theta)(\theta-\hat{\theta}) + \ln|P(\hat{\theta})| \qquad (2.4)$$

since $\partial\ln|P|/\partial P = -\partial\ln|P^{-1}|/\partial P = -P'$. In a procedure like (2.3), this approximation can be substituted in the objective as one proceeds through a sequence of

programs.

2.3. Combining point data with regional averages

So far, we were only concerned with situations where the data on y_{kr} are either available ($y_{kr} \square$ G) or not available and this led to a particularly simple structure for the matrix T. In practice, the situation is often more complex because additional data are available say, on aggregates (over k and/or over r) say, regional averages. For example, one map may provide information on land cover by forest while the other distinguishes particular types of forest. One statistical source may give average rainfall in a region, while the other reports on the rainfall at a particular point. In principle all these data can be expressed in terms of the most detailed elements via the matrix T that links variables q to observations y and the specification (1.5) actually covers this case already. In this context error specification deserves special attention, as the variance of the average can hardly be taken to be the same as for isolated points. We add that it is easy to extend the specification so as to allow for inequality constraints on q, so as to define acceptable ranges. This is of special relevance at boundary points, where the autoregressive structure is somewhat artificial.

2.4. Restrictions on autocorrelation

Restrictions on parameter sets may be used to rule out specific autocorrelation patterns, either spatially or across variables at the same point (see also section A.1). Yet this is not a very flexible approach. Suppose for example that the soil type is known to differ significantly between two neighbouring points. Ruling out all spatial correlation say, between the crop yields attainable under identical input use (e.g. fertilizer and labour) at these two points may be as unrealistic as postulating a spatial correlation according to model (1.3a). What one would like in this case is to correct the effect of autocorrelation for the variation in soil type. For this, one may modify equation (1.3a) into:

$$[I-A(\theta)](q-M(\theta)x) = B(\theta)x \tag{2.7}$$

where $M(\theta)x$ is the correction term, such that $(q-M(\theta)x)$ is the effect that can be transposed to other points (say, the yield corrected for soil specificity). If $M(\theta)$ is linear in θ, the bilinear structure will be maintained and computation can proceed as in (2.3).

3. A COMPARISON WITH ALTERNATIVE METHODS

3.1. Co-kriging

We compare program (1.5) with the formulation of the co-kriging model. Co-kriging is widely used in spatial statistics, mainly to explore the structure of multivariate spatial information and to perform interpolation, rather than a means of estimating a particular simulation model. For example, Goovaerts (1992) applies a principal components method to a spatial data set. The co-kriging model in Stein (1991) uses the approach for interpolation and considers a partition of the index set K into K_1 and K_2, where K_1 refers to the variable to be predicted, for which the observations set is G_1, and K_2 to co-variables with observation set G_2. The model is:

$$y = Cx + e \tag{3.1}$$

where e has mean zero and given covariance Ω, and the elements of depend on the distance between the points. For given covariance, the maximum likelihood estimator C can be obtained by generalized least-squares as:

$$\min_{e, \theta} \frac{1}{2} e'\Omega^{-1}e$$

subject to   (3.2)

$$y = C(\theta)x + e$$

or

$$\hat{\theta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$$

where X is the $R\infty J$ matrix with x_j as the j-th column and θ is the vector consisting of the elements of the matrix C. We can rewrite (3.2) into a form that more closely resembles (1.5). Let P be such that $\Omega = PP'$. For positive definite Ω, it is possible to derive P from Ω and write:

$$\min_{q, u, \theta} \frac{1}{2} u'u$$

subject to   (3.3)

$$q = C(\theta)x$$
$$y = Tq + Pu$$

Here also, the use of the possibly sparse matrix P enables us to avoid matrix inversion. As long as P is taken as given, least-squares will yield a maximum likelihood estimator but if P is treated as a function of say, θ, this property will be lost and maximization of the (possibly concentrated) log-likelihood function must

be the objective, as in section 2.2. It may be added that the form (3.3) automatically yields the interpolations. The form also makes it clear that whenever interpolation is to take place at a new point, the full system will have to be re-estimated (as for (1.5) unless a particular recursive structure is assumed to hold for P, e.g. if it has zero off-diagonal elements in the column corresponding to the new point. Such an assumption is implicit in Stein (1991) and various other applications of co-kriging. In applications, the variogram matrix P^{-1} (or the covariance Ω) is usually obtained from preliminary calculations with a model for the spatial variability of y. In this respect practices are similar to those of time-series analysis, where the type of autoregressive scheme is often determined by inspection of the data prior to ML-estimation. Such techniques may also be applicable to obtain initial estimates of error term distributions under ML-estimation.

3.2. Generalized Least Squares

A second alternative is Generalized Least Squares (GLS), for which the program may be written as:

$$\min_{e, \theta} 1/2\ e'\Omega^{-1}e$$

subject to (3.4)

$$y = A(\theta)y + B(\theta)x + e$$

Solution could proceed by elimination of the constraints via out-substitution of e in the objective. The GLS approach differs from the program (1.5). A comparison will enable us to summarize the main characteristics of the approach proposed in this paper. We note the following differences:

− *The covariance matrix.* GLS uses the inverse covariance matrix, while (1.5) has Pu in the constraint. This difference is only apparent since working with $e'\Omega^{-1}e$ in the objective and error e in the constraint is, for positive definite Ω, equivalent to having u'u in the objective and Pu in the constraint, for $\Omega=PP'$.

− *Endogenous variables as choice variables of the program.* The endogenous variable y no longer appears in (3.4). The variable q (= y) is treated as exogenous. This was possible because all data are supposed to be available.

− *Specification of the error term*. The specification of the error term is different, and this reflects the treatment of y as an exogenous variable. Consequently, the model cannot be used for obtaining an (unbiased) estimate of y-

values for x-values outside the sample.

Note that the form y = A(θ)(y-e) + B(θ)x + e, as in (1.5) makes it possible to maintain the equation y = q + e and estimate. However, by the same token it becomes impossible to eliminate the constraint without having to deal with the large-scale, non-sparse inverse $[I-A(\theta)]^{-1}$. This explains the need to maintain the constraints of the optimization program, rather than eliminating them via substitution, not only when formulating the estimation problem but also at the stage of computation.

4. AGRICULTURAL MODELLING

The specifications discussed so far are very general. The model (1.3a) is a linear form of the even broader formulation in (1.7). A large number of practical applications would fit within this framework. As an example, we consider a modelling framework for agricultural production applicable say, at provincial level, articulating it as follows:

(i) - *Climate.* Climatic conditions are represented through a spatially correlated autoregressive system as in (1.3a). This generates a complete set of values q_{kr} for temperature, rainfall and other climatic variables that affect the conditions for agricultural production. These values would also depend on the exogenous variables x_{jr} such a latitude, longitude, altitude and possibly also slope, etc:

$$[I-A(\theta)]q = B(\theta)x \qquad (4.1)$$

(ii)- *Land inventory.* It is assumed that at location r, an area L_{ru} is available for land of quality u, where u = 1, ..., U. Hence, we associate to every point r with coordinates x_{1r} and x_{2r}, the surfaces L_{ru}, which may, for example, have (x_{1r}, x_{2r}) as their point of gravity.

(iii) - *Crop yield.* Let crops be indexed c, with c = 1, ..., C. The yield of crop c will depend on three factors: climatic conditions, soil conditions and cultivation practices. We disregard any spatial interdependence, as the physiology of the crop depends purely on local conditions which are supposed to be reflected fully in the values of q_r and x_r and the soil type u. Hence, the yield relation may be written as:

$$s_{ucr} = G_{uc}(q_r, x_r, v_{ucr}) \qquad (4.3a)$$

where v_{ucr} is the vector with elements v_{iucr} denoting the quantity of input i (e.g. labour or fertilizer) applied per hectare for cultivation of crop c at location; s_{ucr} is the supply per hectare (or yield) of crop c on land type u at location r.

(iv) - *Crop output and demand for inputs*. Total crop output and input demand follow as:

$$S_{ucr} = s_{ucr} L_{ucr} \qquad (4.3b)$$
$$V_{ucr} = v_{ucr} L_{ucr} \qquad (4.3c)$$

under the requirement that the land constraint has to hold:

$$\Sigma_c L_{ucr} \leq L_{ur} \qquad (4.3d)$$

Equations (4.3a-d) typically represent the constraints of a standard farm model for say, a revenue maximizing producer. Furthermore, to avoid full specialization, additional restrictions will be required (see e.g. Folmer et al., 1995).

In this setup, two properties stand out. First, climatic variables are not affected by cultivation practices but through a spatial interdependence that may result from explicit unidirectional linkages e.g. through winds an water flows, as well as from multidirectional proximity effects (dissipation). The matrix $A(\theta)$ can be specified to account for such effects. Secondly, the assumption that crop yield in function (4.3a) only depends on local conditions and inputs applied indicates that we treat the 'farm' at location r as a non-spatial entity. This is appropriate if we want to develop a model for a large geographical unit where the size of the individual farm does not matter. In such a model the land availability L_{ur} actually refers to a large number of farms that are treated as identical. This is the relevant scale for representing spatial variation in natural conditions.

In contrast, if we want to model the individual farm, the variation in climatic conditions becomes irrelevant. Still, the spatial model may prove useful to account for dissipation of inputs V_{ucr} at various locations r on the same farm. The incidence of such inputs on the farm then becomes part of the vector q_r, while the initial application is given in the vector x_r in (4.3a). This makes it possible to represent flows of water and nutrients both on the surface and within the soil. If in addition, these variables are given a time-subscript, this also provides a way of representing the spread of diseases, or a sequential harvesting of contiguous plots. The autoregressive (4.3a) equation will now appear among the constraints of the farm

model and may, moreover, be helpful in avoiding full specialization.

5. CONCLUSION

The main point of this paper has not been to advocate a particular spatial model but to show how the constrained programming formulation makes it possible to maintain a clear distinction between the likelihood function, the postulated autoregressive model, the data structure and the parameter set. Applied work to implement this approach is currently in progress. Estimation on a nationwide climatic dataset for Nigeria has been completed and estimation of an agricultural model similar to (3.3) is underway. The approach is also to be applied shortly in a model that describes land utilization and land cover for Russia and China (Ermoliev et al., 1996). In these applications the main practical advantages appear to be that (i) the points where data are available do not have to lie on some given grid, (ii) data on regional aggregates can be made use of, (iii) missing observations can be estimated though interpolation and (iv) due to the flexibility of the mathematical program, there is no need to remain within the confines of some standard estimation routine and this make it relatively easy to incorporate nonlinear forms, price dependence and other generalizations.

We conclude that, while it would obviously be wrong to state as a rule that an agricultural model should be spatially explicit and account for spatial interaction, the spatial approach seems promising in several respects. It provides an opportunity to tap a wealth of geographical data without losing the geo-referencing and the representation of spatial interactions after aggregations to provincial or state level.

ANNEX:

IMPLEMENTATION OF THE ESTIMATION PROCEDURE

At least three steps must be taken to reach a practical implementation of program (1.5) as an ML-estimation: (i) the functions $A(\theta)$ and $B(\theta)$ must be specified; (ii) program (1.5) must be solved; (iii) procedures must be formulated to characterize the optimum, in particular the reliability of estimates.

A.1. Restrictions on parameters

We propose to specify $A(\theta)$ via a function that uses the values of the same variable at neighbouring points as well as those of related variables at the same point, as follows:

$$A_{krk'r'} = \kappa_{kr}\, w_{rr'} + (1-\kappa_{kr})\, \alpha_{kk'} \qquad\qquad (A.1)$$

The first term relates to the spatial effect of the same variable k at other sites r′ and the second to the local effect of other variables k′ at the same site. This specification reflects interpolation in a geometric sense: a value y_{kr} may be seen as a weighted average of an interpolation $\breve{y}_{kr} = \Sigma_{r'}\, w_{rr'}\, y_{kr'}$ and a 'local' estimate $\tilde{y}_{kr} = \Sigma_{k'}\, \alpha_{kk'}\, y_{k'r} + \Sigma_j\, B_j\, x_{jr}$. There are thus two weights: the weight κ_{kr} between local and neighbouring effects and the weights $w_{rr'}$ among neighbours.

We also impose numerical restrictions on these parameters:

(i) zero diagonal: $\alpha_{kk} = 0$ and $w_{rr} = 0$ for all k and r;

(ii) spatial weight: $\Sigma_{r'}\, w_{rr'} = 1$ for all r;

(iii) contraction: $\Sigma_{k'}\, |\alpha_{kk'}q_{k'}| < q_k$ for all k and q_k positive.

Condition (i) is an obvious requirement to avoid 'self-regression'. Condition (ii) follows from the interpretation as a weight. The contraction condition (iii) guarantees that the dominant eigenvalue of the matrix A will not exceed unity. It is imposed to ensure that spatial effects dampen out. We may assume that the effect of differing units of measurements across k has been eliminated either by a change in units or by taking the logarithms of all variables (see also section 3.1) and for convenience we also require that condition (iii) holds for $q_k = 1$, to maintain linearity of the parameter set.

A.2. Computing the optimum

Program (1.5) will typically contain various nonconvexities and this makes it necessary to devise practical ways of solving it iteratively, because a direct calling of some mathematical programming algorithm will seldom prove effective. Here we propose to exploit the bilinear structure of the model (the $A(\theta)y$ is a bilinear form due to the linearity of $A(\theta)$ and solve a sequence of simpler programs, which read as:

$$\min \tfrac{1}{2}\, u'u$$

q, u, θ
subject to   (A.2)
$[I-A(θ)]\hat{q}$ - B(θ)x + $[I-A(\hat{θ})]q$ - B($\hat{θ}$)x = 0
y = Tq + Pu
θ ☐ Θ

where $\hat{q}$ and $\hat{θ}$ are given estimated values. For a linear specification of A(θ) and B(θ), program (A.2) will have linear constraints and a quadratic objective, a form that can be solved easily. It yields an update for θ as well as values for u and q. A fixed point will have been obtained when the new value of θ will coincide with the $\hat{θ}$ that was entered into (A.2). Next, when a value $\left\| θ - \hat{θ} \right\| ≤ e$ has been reached, we solve for given $\hat{θ}$, the program:

min 1/2 u′u
q, u
subject to   (A.3)
$[I-A(\hat{θ})]q = B(θ)x$
y = Tq + Pu

This will yield a feasible solution for program (1.5). The algorithm should converge quickly when started from there, since at θ = $\hat{θ}$ an optimum of (A.2) is also a stationary point of (1.5). More generally, if the model is nonlinear, as may be the case in programs (1.7) and (1.8), it is practical to proceed via a sequence of quadratic programs, that perform a quadratic approximation for the log-likelihood and a linear approximation for the function H(⊇).

a.3. The covariance matrix

If we assume that after appropriate testing, the constraint on θ is nonbinding i.e. lying in the strict interior of Θ, we may start from equation (1.7) and differentiate it with respect to θ and calculate the associated derivative for y^s:

$∂H/∂q \; dq/dθ + ∂H/∂θ = 0$ (A.4.a)
$dy^s/dθ = Tdq/dθ$ (A.4.b)
$du/dθ = -P^{-1}dy^s/dθ$ (A.4.c)

After defining the matrix F = -du/dθ, we can calculate the asymptotic covariance matrix V for θ (see e.g. Davidson and MacKinnon, 1993, chapter 2) as:

$$V = \hat{\sigma}^2 \, [F'F]^{-1} \tag{A.5}$$

where, $\hat{\sigma}^2 = 1/N \, u'u$, (for N equal to the dimension of y) or possibly $1/(N-H) \, u'u$, for H equal to the dimension of θ, to correct for degrees of freedom. From there, we can compute the covariance matrix of the interpolated values as:

$$W = (\partial q_2(\theta)/\partial\theta) \, \text{covar}(\theta) \, (\partial q_2(\theta)/\partial\theta)'$$

or

$$W = \hat{\sigma}^2 \, F_2[F'F]^{-1}F_2' \tag{A.6}$$

where the rows of the matrices $\partial q(\theta)/\partial\theta$ and F are assumed to have been partitioned in two blocks, the first corresponding to kr-values for which observations are available and the second, with subscript 2, corresponding to the interpolated values. It must be emphasized that W is the covariance of $q(\theta)$ only, without allowance for errors at points where no data are available. If we assume that P is the unit matrix and that there is an error of measurement with estimated variance $\hat{\sigma}^2$, for every kr $\square$ G, the variance becomes:

$$W = \hat{\sigma}^2 \, (I + F_2[F'F]^{-1}F_2') \tag{A.7}$$

Now if we assume that some constraints on θ were binding (say, those associated with condition (iii) in section A.1), the set Θ will consist of the constraint:

$$M\theta = m \tag{A.8}$$

To compute the covariance one should partition the matrix M and the vector θ into 'basic' and 'nonbasic' parts θ_B and θ_N, and treat the basic variables as a function of the nonbasic ones to obtain a function $\tilde{\theta}(\theta_N)$:

$$\tilde{\theta}_B(\theta_N) = M_B^{-1}(m - M_N\theta_N) \tag{A.9.a}$$
$$\tilde{\theta}_N(\theta_N) = \theta_N \tag{A.9.b}$$

Relation (A.4.a) will now become:
$$\partial H/\partial q \, dq/d\theta_N + \partial H/d\theta \, \partial\tilde{\theta}(\theta_N)/\partial\theta_N = 0$$
$$\tag{A.10}$$

Further calculations proceed as in (A.4)-(A.6) but with F_N instead of F. The linear matrix equation (A.4.a) or (A.10) and the inverse V can be obtained by solving a linear program that treats the matrix equation as constraint and minimizes the sum of slacks on these equations. While such a procedure may appear like a roundabout way for solving a simple problem, it enables us to take advantage of the sparse and numerically stable implementations that are available for various linear programming algorithms. We have seen that this is of special relevance in the context of spatial analysis, where the autocorrelation patterns tend to generate large and sparse matrices. If space limitations restrict the size of the linear program that can be handled, one may process groups of elements of θ in sequence.

REFERENCES

Brooke A., Kendrick D. and Meeraus A., 1992, "*GAMS: a User's Guide*", Release 2.25, Boyd and Fraser, San Francisco, CA.

Cressie N.A.C., 1993, "*Statistics for Spatial Data*", Wiley, New York.

Davidson R. and MacKinnon J.M., 1993, "*Estimation and Inference in Econometrics*", Oxford University Press, Oxford.

Folmer C., Keyzer M.A., Merbis M.D., Stolwijk H.J.J. and Veenendaal P.J.J., 1995, "*The Common Agricultural Policy beyond the MacSharry Reform*", Amsterdam, North-Holland.

Fischer G., Ermoliev Y., Keyzer M.A. and Rosenzweig C., 1996, "*Simulating the Socio-economic and Biogeographical Driving Forces of Land-use and Land-cover Change; The IIASA Land-use Change Model*", WP-96-010, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Goovaerts P., 1992, "Factorial Kriging Analysis: a Useful Tool for Exploring the Structure of Multivariate Spatial Oil Information", *Journal of Soil Sciences*, 43, p. 597-619.

Greene W.H., 1991, "*Econometric Analysis*", Macmillan, New York.

Heuvelink G.B.M., 1993, "Error Propagation in Quantitative Spatial Modelling", *Netherlands Geographical Studies 163*.

IGBP/HDP, 1995, "*Land-use and Land-cover Change; Science/Research Plan*", IGBP Report N° 35, HDP Report N° 7.

Isaaks E.H. and Srivasta R.M., 1989, "*An Introduction to Applied Geostatistics*", Oxford University Press, Oxford.

Stein A., 1991, "De ruimtelijke variabiliteit van bodem en water", *Ruimtelijke statistiek van bodem en water,* Commissie voor hydrologisch onderzoek TNO, Rapporten en Nota's, N° 24.

Résumé

La recherche en matière de politique agricole exige souvent un traitement plus ou moins détaillé de données sur la géographie du pays ou de la région en cause et si cette recherche a trait à l'environnement, il faudra également tenir compte de certaines interactions entre les unités géographiques, par exemple de débits fluviaux ou d'érosion. Toutefois, plus l'analyse contiendra de détail géographique, plus il sera difficile, voire impossible, d'obtenir une base de données complète. Cet article s'adresse à cette question et décrit, pour un modèle spatial, une procédure d'estimation par interpolation à vraisemblance maximale. Dans ce modèle, les interactions spatiales sont représentées par des équations autorégressives dans lesquelles la valeur d'une certaine variable dans une localité dépend entre autres de la valeur de cette même variable dans les localités adjacentes. On appellera cela une interpolation (spatiale autoregressive) pour indiquer que ce modèle calcule entre autres des 'prévisions' interpolées pour les données non disponibles. Pour préserver une distinction nette entre la fonction de probabilité, le modèle autoregressif, les contraintes sur les paramètres et les restrictions qui résultent de la non-disponibilité de certaines données, la méthode proposée fait appel à un programme mathématique plutôt qu'à une routine statistique ordinaire.

Resumen

El estudio en materia de política agrícola exige a menudo, un tratamiento màs o menos detallado de los datos sobre la geografía del país o de la región en causa; y si este estudio trata del medio ambiente, tendremos que tener en cuenta ciertas interacciones entre las unidades geográficas, por ejemplo los caudales fluviales o la erosión. De todas maneras, mientras màs detalles geográficos

contiene el análisis, màs difícil será, ver imposible, obtener una base de datos completos.

Este articulo se refiere a esta pregunta y describe, para un modelo espacial, un proceso de estimación por interpolación a máxima verosimilitud. En este modelo, las interacciones espaciales son representadas por las ecuaciones autoregresivas en las cuales, el valor de cierta variable en una localidad depende del valor de esta misma variable en las localidades adyacentes.

Lo denominaremos interpolación (espacial autoregresiva) para indicar que este modelo calcula las previsiones de interpolación para datos aùn, no disponibles. Para preservar una clara distinción entre la función de probabilidad, el modelo autoregresivo, las obligaciones sobre los parámetros y las restricciones que resultan de la indisponibilidad de ciertos datos, el método propuesto nos hace pensar màs bien a un programa matemático qué a una rutina estadística ordinaria.