ELSEVIER

Agricultural Economics 27 (2002) 425–443

AGRICULTURAL ECONOMICS

www.elsevier.com/locate/agecon

Spatial dimensions of precision agriculture: a spatial econometric analysis of millet yield on Sahelian coversands

Raymond J.G.M. Florax [a,*], Roelf L. Voortman [b,1], Joost Brouwer [c,2]

[a] *Department of Spatial Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b] *Centre for World Food Studies (SOW–VU), Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[c] *Brouwer Environmental & Agricultural Consultancy, Wildekamp 32, 6721 JD Bennekom, The Netherlands*

Abstract

The identification of local soil variability caused by within-field differences of macronutrients and ecological features is of paramount importance for the effectiveness of precision agriculture. We present several spatial statistical and econometric techniques to capture local differences in soil variation, ecological characteristics, and yield more effectively than the analytical techniques traditionally used in agronomy. The application of these techniques is illustrated in a case study dealing with precision agriculture in the West African Sahel. The production of millet on acid sandy soils constitutes a typical example of low soil fertility areas exhibiting small absolute but large relative differences in crop production conditions over short distances.

JEL classification: C21; C51; Q12; R32

Keywords: Precision agriculture; Spatial econometrics; Millet yield; Sahel; Coversands

1. Introduction

Fertiliser recommendations, specifying the appropriate mix and dose, are often made for large regions (Bullock et al., this issue). In the case of developing countries, they may even pertain to entire countries. Soils are however heterogeneous over space, even over short distances (Eswaran et al., 1992). Uniform recommendations are therefore likely to be inefficient (Bouma, 1997; Clay et al., 1999) and can eventually even cause local yield reductions.

Precision agriculture constitutes the modern management strategy to cope with soil heterogeneity. In a high-tech fashion, precision agriculture is based on four essential ingredients: a spatial representation and analysis of soil and yield variability, a soil-type and hence location-specific optimum fertiliser prescription, a global positioning system (GPS), and farm machinery capable of variable rate applications (Bullock et al., this issue). In theory, the linkage of these four ingredients carries a large potential for improving the efficiency of soil resource use. In less-developed countries, however, fields are often small and the technology of GPS and farm machinery is mostly not available. Differences in crop performance are

* Corresponding author. Tel.: +31-20-4446092; fax: +31-20-4446004.
E-mail addresses: rflorax@feweb.vu.nl (R.J.G.M. Florax), r.l.voortman@sow.vu.nl (R.L. Voortman), brouwbar@bos.nl (J. Brouwer).
URLs: http://www.feweb.vu.nl/re/master-point, http://www.sow.vu.nl/
[1] Tel.: +31-20-4449321; fax: +31-20-4449325.
[2] Tel.: +31-318-413468.

0169-5150/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0169-5150(02)00068-3

nevertheless important, and their relevance is amplified by cash constraints, as is the case for instance for millet production on Sahelian coversands in Africa, covered in our case study.

Precision agriculture is an appealing concept and its principles quite naturally lead to the expectation that farming inputs can be used more effectively, with subsequent improvements in profits and environmentally less burdensome production. Especially in the case of small farmers in developing countries, precision agriculture holds the promise of substantial yield improvement with minimal external input use. However, most research on precision farming is conducted in developing countries, and reveals that increased input efficiencies result in rather modest profitability increases (Kilian, 2000). This may be the main reason for adoption rates being rather low (Cook et al., 2000). The main obstacle is most likely the absence of fertiliser application prescriptions that adequately match individual site characteristics (Stewart and McBratney, 2000; Bullock et al., this issue).

Soil variability and appropriate fertiliser doses in precision farming can be investigated using soil and crop yield maps. Soil units can be entered as grids into a geographic information system (GIS). However, soil unit boundaries are usually arbitrarily determined and the soil-mapping units allow a good deal of variation of soil properties. This problem can be overcome by interpolation of chemical soil properties, sampled in a spatial point pattern, to the regular grid structure. The visual interpretation of multiple layers of soil attributes is then, however, rather cumbersome, even if complemented by exploratory data analysis in a GIS environment. In addition, the application of different methods of spatial interpolation (inverted distance weighting, spline functions, spatial polynomials, kriging) usually has a considerable impact on the obtained results, and univariate interpolations may very well be inappropriate because crop yield responses evidently depend on systematic variation in other variables (Nielsen et al., 1999).

Input prescriptions can also be derived from regression analyses using crop yield as the dependent variable and inputs, soil and climate variables or soil types as the explanatory variables. However, the method of regression analysis gives rise to complications of its own. Some of these complications are inherent to any econometric analysis. Variable selection and specification of functional form are difficult, also because agronomy and soil science do not always provide adequate theoretical guidance. For instance, the literature typically focuses on the effect of applied macronutrients nitrogen (N), phosphorus (P), and potassium (K), although it is well known that a multitude of interacting soil and climatic factors affect crop yields and input efficiencies (FAO, 1983). Regression models are often specified as simple linear or quadratic relationships although nutrient interaction models allowing for plateau growth have been derived (so-called Mitscherlich and Mitscherlich–Baule specifications). These are, however, much more difficult to estimate.

The above may be an important cause for the rather modest performance of regression models. Applications of regression analysis have usually not been capable of explaining more than 30% of crop yield variation at the field level (Nielsen et al., 1997, 1999; Wendroth et al., 1999). However, another important reason for a relatively poor fit is likely to be the prevalent soil heterogeneity, causing crop yield response to a single variable to vary depending on the varying total soil constellation in combination with differences in external input treatments. Moreover, the inherent spatial nature of soil and crop data should not be ignored.

The inherent spatial feature of the analysis implies that agricultural data exhibit differences depending on the exact location in the field, and spatial patterns or clustering is likely to occur. The commonly used ordinary least squares (OLS) regressions are inappropriate for this purpose and lead to misleading results. Although appropriate techniques are available in spatial statistics and spatial econometrics, they have to date not been used extensively in agronomic research. There are a few exceptions. Long et al. (1992) demonstrate inconsistencies between spatial models and classical OLS because of the erroneous assumption of independence. Kessler and Lowenberg-DeBoer (1999) conclude that spatial regression techniques are more appropriate. Bongiovanni and Lowenberg-DeBoer (2000, 2001) use spatial techniques and show that for their particular example, the spatial autoregressive models consistently point to the profitability of N application, whereas the OLS model does not.

This paper extends this spatial line of research. We review several exploratory as well as explanatory techniques from spatial statistics and spatial econometrics, and empirically illustrate their relevance and impact by means of a case study pertaining to millet yield production on the acid sandy soils of the Sahelian coversands. Specifically, Section 2 describes precision agriculture requirements in the low-tech setting of smallholders in the Sahel, and summarises the state-of-the-art knowledge regarding crop yield variation in agronomic research. In Section 3, the prevalent environmental conditions in the Sahel are spelled out, and we introduce the data for the case study. Section 4 describes spatial data analysis techniques, and the use of these techniques is illustrated in Section 5. Section 6 provides a summary and conclusions.

2. Precision farming and agronomic knowledge about yields

Although high-tech and low-tech precision farming are vastly different, the basic goals behind the two are identical and aimed at optimisation of resource use. High-tech precision agriculture is capital-intensive. It is pervasively dominated by technological hardware: GPS, variable-rate application machinery, and GIS. Low-tech precision agriculture is much more dependent on visual observation of topography, topsoil characteristics, local vegetation (arable weeds) and crop performance, and on labour as opposed to capital as a main production input. However, in both variants of precision agriculture the focus is on identification and use of location-specific resource applications, for instance fertiliser or manure and herbicide or weeding labour, that contribute to yield optimisation and in turn, depending on market circumstances, to profit optimisation.

The case study in this article represents a low-tech situation, in which precision farming is a traditional phenomenon that can be developed further. In southwest Niger, peasant farmers grow millet as their staple food on acid and sandy Aeolian deposits (coversands). They cultivate mainly with hand-tools and external input levels are currently low or non-existent. The capital-extensive production mode is not necessarily a negative impediment for a profitable application of precision agriculture, although critical presuppositions such as an adequate financing system to be able to buy sufficient amounts of fertiliser are not yet satisfied in many instances. As the local farmers know very well, precision agriculture is equally relevant to the application of scarce local resources and of external inputs. For both types of resources, the crucial issue is for instance *which* fertiliser to apply *where* in order to achieve optimal fertiliser efficiency and optimal returns to labour and cash investment. Consequently, the development of analytical techniques determining efficient and effective external fertiliser use are crucial for both high-tech and low-tech precision agriculture.

Spatial dimensions of agricultural production are particularly relevant on Sahelian coversands, because yield variation is substantial within short distances (Scott-Wendt et al., 1988a,b; Geiger and Manu, 1993; Hermann et al., 1994; Lamers and Feil, 1995; Manu et al., 1996; Brouwer and Bouma, 1997; Rockström and De Rouw, 1997; Krogh, 1999; Rockström et al., 1999). The causes for the extreme and very localised variation in yield are still poorly understood (Buerkert, 1995), but they can likely be derived from equally variable soil and hydrological conditions (Brouwer and Powell, 1998). Prevailing surface soils are the result of various episodes of Aeolian deposition and erosion. As a result, they vary in origin (parent material) and age, both of which affect soil chemistry. Moreover, the coversands vary in grain-size distribution, which influences the susceptibility to crust formation on the soil surface, and consequent moisture infiltration and seedling emergence (Sombroek and Zonneveld, 1971; Zonneveld et al., 1971; Voortman et al., 2002). Under unfertilised conditions, the spatial distribution of these soil differences leads to rather abrupt crop yield changes within a few meters. Within a 1 ha field, yields measured on a 5 m × 5 m regular grid, varied between 0 and 2885 kg ha^{-1} (Brouwer and Bouma, 1997). This suggests that efficient and effective fertiliser treatment should be and can be spatially differentiated.

In their attempt to develop appropriate fertiliser technologies, agronomists have struggled with the problem of localised millet yield variation (Moorman and Kang, 1978; Scott-Wendt et al., 1988b; Wendt et al., 1993; Hermann et al., 1994; Buerkert, 1995; Manu et al., 1996). The analysis of variance (ANOVA) of results of conventional block experiments accounts for differences between blocks, but highly localised soil variability violates the requirement that soil conditions within individual blocks should be homogeneous. Lack of data for different levels of spatial aggregation limits the possibilities of choosing an appropriate spatial scale for the analysis of crop yield variation. Traditional experimental research suggests that the macronutrients, N, P and K, limit millet yield, although not uniformly (for an overview, see Buerkert and Hiernaux, 1998; Voortman and Brouwer, 2002).

As an alternative to multi-factorial experimental treatment with repetitions, a nexus of regression analysis methods has been applied to empirically observed yield patterns. Scott-Wendt et al. (1988a,b) investigate a transect of poorly to well growing millet, and observe low yields to be correlated with high aluminum (Al) saturation levels and lower cation levels (calcium (Ca), magnesium (Mg), and K). Manu et al. (1996) perform a pair-wise comparison of good and poor spots in terms of growth, and identify Al saturation to be higher in poor spots, while at the same time the pH value of the soil is lower. Stein et al. (1997) and Gandah et al. (1998) use linear regression models to investigate soil chemistry in relation to millet yields. The explanatory power of such a model is, however, rather low. Rockström et al. (1999) also use a linear regression model and report a slightly better statistical fit. However, in their case a large part of the variation is explained by manure or fertiliser treatment, and in two out of three years, native soil chemistry does not seem to play a significant role. In the remaining year, organic carbon and base saturation are significant, but these compound variables are difficult to interpret with respect to fertiliser requirements. Voortman and Brouwer (2002), and Voortman et al. (2002) use non-parametric kernel density regression (Keyzer and Sonneveld, 1997) to identify explanatory variables and functional forms, and subsequently apply parametric regression techniques. They observe very modest yield impacts of the macronutrients N, P and K, but significant contributions of cation ratios, aluminum saturation, total exchangeable bases (TEB), the soil pH value and manure levels, in addition to the abovementioned relevance of crusting and local topography and local hydrology.

The overview of the West African literature shows that persistent problems exist. The issue of spatial scale is difficult to solve due to limited data availability for smaller grid sizes. The choice of explanatory variables is complicated because of a lack of theoretical guidance from agronomy and soil science. Functional forms range from simple additive linear relations to Cobb–Douglas, and more complicated non-linear specifications allowing for factor substitution and plateau growth. Almost without exception, however, the spatial dimension of the data is largely ignored.

In order to improve this situation, we concentrate on incorporating the inherent spatial structure and variation in a regression framework. Because of the focus on spatial effects, we use a simple Cobb–Douglas specification, but the principles and most of the techniques can be applied with a more complex function as well. The selection of explanatory variables is based on the above literature review, and includes the macronutrients commonly applied as external inputs, manure levels and geo-physical characteristics of the plot, and chemical interactions.

3. Data and variable definition

The empirical example is concerned with 1992 data, sampled from a 1 ha field cultivated by a local African farmer. The field is located just outside the village of Bellaré, near the ICRISAT Sahelian Centre, 40 km southeast of Niamey, Niger. The altitude of the location is approximately 240 m above mean sea level, the average annual temperature is 29 °C, and the average rainfall is 545 mm in a well-defined rainy season, lasting from May to September.

The crop of pearl millet (*Pennisetum glaucum* (L.) R. Br., labelled Millet) was planted in two batches: 80% of the field after the first sizeable rains on May 16, and the remainder after follow-up rains on May 26, 1992. At harvest (September 15–16), yields were measured for 5 m × 5 m regularly spaced plots of the 1 ha field. These measurements are rescaled to the 10 × 10 grid level (through averaging), because soil chemistry data are only available at this level of aggregation. The average standardised millet yield per grid cell equals 649 kg ha^{-1}, with a coefficient of variation of 0.49.

Climatological conditions hardly vary within a 1 ha field, but details of the climate during the growth season provide important background information. The rainy season of 1992 was generally considered "good," and no intra-season drought periods were observed. An earlier analysis indeed confirmed that, at the field level, yield reductions due to moisture stress are unlikely in 1992 (Voortman and Brouwer, 2002). Potential effects of moisture availability on crop yield must therefore be attributable to local differences in infiltration and/or overland flow, caused by localised variation in topography and surface crusting.

Surface crusting (labelled Crust) was recorded directly in the field on a discrete five-point scale. Topographic variation (Topovariation) was quantified using altitude relative to neighbouring plots, derived from a topographic survey. It provides a measure of the curvature of the terrain and can be considered as a proxy for overland flow processes (for details see Voortman and Brouwer, 2002). Both variables are again rescaled to the 10×10 grid level by means of averaging. Soil samples were taken from the centroids of the 5×5 grids, at 0–0.2, 0.2–0.4, and 0.7–0.9 m depth. For the chemical analysis, the original samples of four adjacent plots were combined for each sampling depth, so the chemical characteristics represent average values at the 10×10 level.[3] To facilitate the empirical analysis, and imitate what many soil fertility specialists do, only the topsoil data (0–0.2 m) are used in this article.

Farmers and agronomist are traditionally concerned with the three macronutrients (N, P and K) affecting biomass production. From the various soil nitrogen measures available, N-total (Kjeldahl; simply labelled N) is used, because this is commonly analysed and hence increases comparability with other studies. For phosphorus, levels of P-Bray, P-total and P-H_2O are available because of the crucial role of phosphorus in this region (Pieri, 1985; Bationo et al., 1990, 1991; Klaij et al., 1994). In the empirical example we use P-Bray (labelled P for simplicity), since it explains millet yield best (Bationo et al., 1991). For K, the only available measurement is exchangeable K in cmol kg^{-1}. The values for K are converted to parts per million (ppm) to allow for comparison with N and P, and are labelled K from here on. The N, P and K variables refer to native soil chemistry as no artificial inputs were used.

Finally, manure levels of cattle and small ruminants, originating from animals resting in, or passing through, the field during the dry season, are taken into account. The spatial distribution is haphazard and uneven, because it is determined by animal behaviour. Manure of cattle (Cattle manure) and small ruminants (Sheep manure) were measured in kg ha^{-1} at the beginning of the growing period, at surface level.

[3] The aggregation was not ideal, because it results in a loss of information about spatial variation, but it was unavoidable because of lack of financial resources for the analysis of all soil samples.

4. An introduction to techniques for spatial data analysis

In this section, we cover various areas of spatial statistics and spatial econometrics relevant for the analysis of our spatial yield data. Section 4.1 introduces the notion of spatial effects. In Section 4.2, we review exploratory spatial data analysis techniques facilitating the detection of spatial yield clusters. Section 4.3 deals with modelling spatial dimensions, in particular, specification testing and estimation in spatial econometric models, providing the background for the millet yield models in Section 5.

4.1. Spatial processes, data, and effects

Cressie (1993) concisely defines the notion of a spatial process as the realisation of a random variable at a particular location, and distinguishes three different situations depending on the type of spatial data available. First, in the case of *spatial points*, one can define a spatial point pattern, such as a crop disease. *Geo-statistical data* are needed to describe an attribute that varies continuously over space, such as the chemical properties of soils. Third, *lattice data* provide a fixed collection of a finite, countable number of grids for which, for instance, the total yield is observed (see Anselin, this issue, for more details).

Regardless of the type of data at hand, two different types of spatial effects are relevant: spatial heterogeneity and spatial dependence. Spatial heterogeneity concerns the uniqueness of attribute values at specific locations. If, for instance, yields are not randomly distributed over space, clusters of high or low values can be observed that coincide with specific locations close to each other on the surface. The typical feature of spatial dependence or spatial autocorrelation[4] is that

[4] A formal definition of spatial autocorrelation is $\mathrm{Cov}(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) \neq 0$, for $i \neq j$, pointing to the coincidence of attribute similarity expressed in y and location similarity for locations i and j. Spatial dependence or spatial autocorrelation are used interchangeably from here on, although strictly speaking spatial autocorrelation is a stronger assumption than spatial dependence (Anselin, 2001).

it is two-dimensional and multidirectional. An observation of an attribute at one location is correlated with the value of the same attribute at a different location, and vice versa, and this correlation can extend in different directions.

Although spatial heterogeneity does not have severe implications for the information that can be obtained from a spatial data sample, spatial autocorrelation does because an observation is partly predictable from neighbouring observations. Spatial heterogeneity and spatial dependence usually concur as meaningful interpretations of a spatial process because the uniqueness or heterogeneity of an attribute observed for a subset of the data can coincide with spatial proximity and hence autocorrelation for that attribute among the same observations (see also Anselin, this issue).

4.2. Exploratory spatial data analysis

The seminal work of Cliff and Ord (1981) has induced substantial research on the statistical properties of spatial data. Much attention has been given to assessing the degree to which data are spatially autocorrelated or clustered. For ordinal or interval data the univariate Moran's I statistic is frequently used. It is given by:

$$I = \frac{n}{S_0}\frac{\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}, \qquad (1)$$

where $\boldsymbol{x}$ is a $(n \times 1)$ vector of observations defined in deviations from the mean, $\boldsymbol{W}$ a spatial weight matrix with $(n \times n)$ elements w_{ij} representing the topology of the spatial system, and S_0 is the sum of the elements of the spatial weights matrix. The weight matrix can be defined on the basis of contiguity, distance, or complex general weights (Cliff and Ord, 1981, pp. 17–18; see also Anselin, this issue). Statistical inference can be based on the standardised value $[I - E(I)]/\text{S.D.}(I)$, but crucially depends on the stochastic assumptions. The moments can be derived analytically assuming that the x follow a normal distribution, or that the distribution is unknown but can be approximated in a nonparametric framework using a randomisation approach (Cliff and Ord, 1981, pp. 42–46).

Although Moran's I gives rise to a test on spatial dependence it can also be used to detect spatial heterogeneity. Anselin (1996) denotes that Eq. (1) can be rearranged as:

$$I = \boldsymbol{x}'(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{W}\boldsymbol{x}, \qquad (2)$$

omitting the standardisation term. This shows that Moran's I is formally equivalent to the estimated parameter of a regression of $\boldsymbol{W}\boldsymbol{x}$ on $\boldsymbol{x}$. This result can be visualised in a Moran scatterplot. The scatterplot can be used to visually identify spatial clusters, outliers and local non-stationarity. The latter can also be assessed statistically by means of local indicators of spatial association (see Anselin, 1995).

It is common practice to interpret Moran's I as a correlation coefficient, although its value is strictly speaking not restricted to the $[-1, +1]$ interval. High positive values signal the occurrence of similar attribute values over space (either high or low values), and hence spatial clustering. Negative values indicate the joint occurrence of high and low attribute values in nearby locations. A value close to zero (more precisely, $-1/(n-1)$, the expected value of Moran's I in the absence of spatial correlation) can be taken as evidence of a random allocation of attribute values over space.

4.3. Modelling spatial dimensions

The development of spatial econometrics goes back to the late 1970s. It centres on the explicit consideration of spatial dimensions in statistical modelling. Based on early work in regional economics and in geography, a considerable body of diagnostic tests, specification strategies, and estimation techniques has been developed (for an overview, see Anselin and Bera, 1998; Anselin, 2001).

The issues of spatial heterogeneity and spatial dependence have received considerable attention in a regression framework. Spatial heterogeneity and spatial dependence are not easily distinguished in an observational sense (Anselin, 2001). Spatial clusters of similar observations may, at the same time, be indicative of spatial heterogeneity and of spatial autocorrelation. In addition, it is obvious that one of the most common specifications, the so-called spatial autoregressive error model, induces a spatial form of heteroskedasticity. The spatial (autoregressive) error model reads as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1}\boldsymbol{\varepsilon}, \qquad (3)$$

where $\boldsymbol{y}$ is the $(n \times 1)$ vector with observations on the dependent variable, $\boldsymbol{X}$ the $(n \times k)$ design matrix containing the explanatory variables, $\boldsymbol{\beta}$ the $(k \times 1)$ vector with parameters, $\boldsymbol{\varepsilon}$ a $(n \times 1)$ vector of iid errors, and λ the spatial autoregressive parameter. The error covariance matrix is given by $\sigma^2[(\boldsymbol{I} - \lambda \boldsymbol{W})'(\boldsymbol{I} - \lambda \boldsymbol{W})]^{-1}$, showing that heteroskedasticity is present even if the error terms are homoskedastic.

This applies likewise to the spatial lag model,

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4}$$

where ρ is the spatial autoregressive lag parameter, and for which the error covariance is the same as in the autoregressive error model assuming otherwise homoskedastic errors.

The standard spatial process models introduced above illustrate two quite different interpretations of spatial dependence. Dependence can either be modelled as a substantive process or as a nuisance (Anselin and Rey, 1991; Anselin and Florax, 1995). In the spatial lag model, a substantive theoretical interpretation can be given to the spatial interaction. In the spatial error model, spatial dependence is caused either by (erroneously) omitted spatially correlated variables, and is hence reflected in the error term, or it is caused by boundaries of regions that do not coincide with actual behavioural units (see also Anselin and Bera, 1998). The different specifications of spatial dependence have divergent implications for estimation and statistical inference. The spatial error model is an example of the more general class of models with a non-spherical variance–covariance matrix, although due to the multidirectional nature of spatial dependence the estimation is more difficult than for time series (in particular, estimated generalised least squares estimators are inconsistent). The spatial lag model exhibits endogeneity that can be taken into account by instrumental variable or general methods of moments techniques, but should preferably be solved using an appropriate maximum likelihood estimator (see Anselin, 1988, for details).[5] In contrast to the time series case, where OLS remains consistent if the errors are not serially correlated and its use is asymptotically warranted, OLS estimators for the spatial lag model are biased and inconsistent, irrespective of the properties of the error term. Nuisance dependence in the error term is less serious because OLS remains unbiased, but it is inefficient.

Rearranging Eq. (4) shows that the spatial error model is equivalent to an extended spatial lag model comprising both the spatially lagged dependent and spatially lagged exogenous variables:

$$\boldsymbol{y} = \lambda W \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{W}\breve{\boldsymbol{X}}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\breve{\boldsymbol{X}}$ is the original design matrix $\boldsymbol{X}$ except for the constant. The formal equivalence only holds if $(k-1)$ nonlinear constraints are satisfied, specifically $\lambda\boldsymbol{\beta} = -\boldsymbol{\gamma}$. This model is generally referred to as the "spatial Durbin" or "common factor" model. The equivalence of the spatial error and the common factor model shows that tests with either the spatial error or the spatial lag model as the alternative hypothesis are likely to have power against the other alternative as well (Anselin et al., 1996; Anselin, 2001).

Apart from focused tests with an informative alternative hypothesis various diffuse tests merely reflecting whether the residuals are spatially correlated have been developed. The oldest and best known is Moran's I for regression residuals, given by:

$$I = \frac{n}{S_0} \frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}}, \tag{6}$$

where $\boldsymbol{e}$ is the $(n \times 1)$ vector of OLS residuals. As with the univariate Moran's I, statistical inference can be based on the assumption of asymptotic normality, or alternatively, assuming that the distribution is unknown, on a theoretical randomisation or empirical permutation approach. Moments and estimation details are given in Cliff and Ord (1981), and Anselin (1988).

Focused tests have a clear alternative hypothesis and have been developed in a maximum likelihood framework. In particular, the Lagrange Multiplier tests LM_λ and LM_ρ explicitly have the spatial error or the spatial lag model as their respective alternative hypothesis. The LM-error test (Burridge, 1980) is identical to a scaled squared Moran coefficient, and reads as:

$$\mathrm{LM}_\lambda = \frac{1}{T}\left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{s^2}\right)^2, \tag{7}$$

where s^2 is the maximum likelihood variance $\boldsymbol{e}'\boldsymbol{e}/n$, and $T = \mathrm{tr}(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}^2)$ where tr is the matrix trace

[5] If the spatially lagged variables are exogenous (a model that is not shown here) OLS retains its desirable properties.

operator. The test asymptotically follows a χ^2 distribution with one degree of freedom. The LM-lag test has the same asymptotic distribution, and looks similar:

$$\mathrm{LM}_\rho = \frac{1}{nJ_{\rho\cdot\beta}}\left(\frac{e'Wy}{s^2}\right)^2, \tag{8}$$

where $J_{\rho\cdot\beta} = [(WXb)'M(WXb) + Ts^2]/ns^2$ is a part of the estimated information matrix, b the OLS estimated parameter vector, and M the projection matrix $(I - X(X'X)^{-1}X')$.

The finite sample performance of the abovementioned tests is very well documented in the literature (see Florax and De Graaff, 2003, for an overview). The power of the Lagrange Multiplier spatial error (lag) test against a spatial lag (error) model seriously complicates finding a statistically rigorous specification strategy that identifies the correct underlying (but unknown) data generating process. The derivation of specification tests with locally misspecified alternatives resulted in robust spatial tests (Anselin et al., 1996), which can be used to distinguish a helpful specification strategy (Florax et al., 2003). The test for a spatial error process robust to the local presence of a spatial lag is:

$$\mathrm{LM}^*_\lambda = \frac{1}{T - T^2(nJ_{\rho\cdot\beta})^{-1}} \times \left(\frac{e'We}{s^2} - T(nJ_{\rho\cdot\beta})^{-1}\frac{e'Wy}{s^2}\right)^2. \tag{9}$$

This clearly shows the subtraction of a correction factor that accounts for the local misspecification of a spatial lag process. The test for a spatial lag process robust to the local presence of a spatial error is given by:

$$\mathrm{LM}^*_\rho = \frac{1}{nJ_{\rho\cdot\beta} - T}\left(\frac{e'Wy}{s^2} - \frac{e'We}{s^2}\right)^2. \tag{10}$$

The robust tests are asymptotically distributed following a χ^2 distribution with one degree of freedom. With these tests performing well in a finite sample setting, it is easy to see that the following specification rule can be fruitfully applied: if both the spatial error and the spatial lag test are significant and only one of the robust variants, then the significant robust test points the correct alternative (Anselin et al., 1996; Florax et al., 2003).

The simplest form of heterogeneity is a non-constant variance of the errors or heteroskedasticity, which may be either spatially induced (as above) or a-spatial. Both forms are easily tested by means of, for instance, the Breusch–Pagan test, which reads as:

$$\mathrm{BP} = \tfrac{1}{2} f'Z(Z'Z)^{-1}Z'f, \tag{11}$$

where the elements of f are defined by $f_i = (e_i/s)^2 - 1$, and Z is a $(n \times k)$ matrix containing the variables thought to influence the heteroskedasticity. The BP test is asymptotically distributed as χ^2 with k degrees of freedom.[6]

A combination of misspecifications, in particular spatial dependence and heteroskedasticity, affects the power of both spatial dependence and heteroskedasticity tests. Specifically, spatial dependence tests seem to over-reject the null hypothesis of no spatial dependence when heteroskedasticity is present. The power of heteroskedasticity tests, on the contrary, is substantially lower when (particularly positive) spatial autocorrelation is present (Anselin and Griffith, 1988; Anselin and Rey, 1991).

Increasingly complex forms of spatial heterogeneity occur in the case of discrete or continuous spatial variation. In the discrete case, the spatial observations can be grouped in such a way that the variation pertains to different spatial subsamples, where each group can be treated as homogeneous. This can be easily modelled by means of spatial regimes. In the continuous case, substantially more complex specifications are needed (Anselin, 1988).

In most real-world applications, an appropriate specification of spatial dependence and spatial heterogeneity is needed simultaneously (De Graaff et al., 2001). This is also apparent from the millet yield application discussed in the next section.

5. Spatial analysis of millet yield in South West Niger

The millet yields are measured in $\mathrm{kg\,ha^{-1}}$ per hectare for 10×10 regular grids of the 1 ha field.

[6] The Breusch–Pagan test is sensitive to non-normality, in which case studentized versions are appropriate. The functional form used for the tests implies additive heteroskedasticity under the random coefficient specification (see Anselin, 1992, for details).

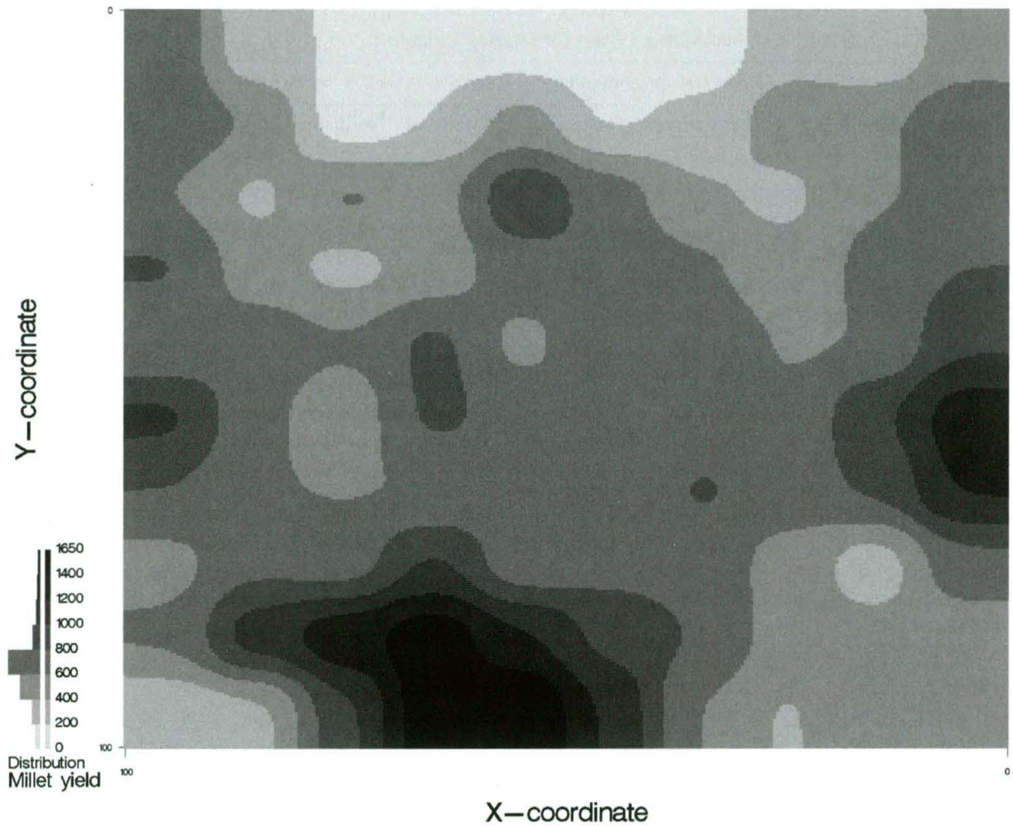

Fig. 1. Mapping of spatially interpolated values and a histogram of millet yield.

Fig. 1 shows the spatial distribution of millet yield on an interpolated grid taking the yield data as point data referring to the centroid of each grid cell. We use multivariate non-parametric kernel density regression methods for the interpolation (Keyzer and Sonneveld, 1997), which are less restrictive than parametric techniques, such as kriging, because the latter impose an exogenously defined spatial structure. The interpolation techniques are merely used for the visual presentations. The numerical analyses are strictly based on grid cell data. Visual inspection of the map shows that there is an obvious clustering of similar attribute values: relatively high yields are clustered in the bottom–center and middle right-hand side of the map, and there is clustering of very low yields in the top part and relatively low yields in both bottom corners as well.

The significance of the spatial clustering pattern can be assessed statistically by means of Moran's I. In order to do so, the spatial structure has to be exogenously defined. We use a contiguity matrix accounting for direct neighbours according to the queen concept, implying that cells are neighbours if they have a common border in the horizontal, the vertical, or the diagonal direction.[7] This so-called first-order queen concept is appropriate because spatial interdependencies are likely to be of relevance only at short distances, and the small sample performance of this interaction pattern has been shown to perform well (Anselin and Florax, 1995).

Table 1 presents several descriptive statistics, a Wald test on normality, and the results of Moran's I test under the normality and the randomisation assumption. The Wald test on normality is important to assess the

[7] The weights matrix is standardized, implying that all row sums are scaled to unity. As a result, a spatially lagged variable, $\boldsymbol{Wx}$, contains the average value in neighboring locations.

Table 1
Descriptive statistics, Wald normality test, and Moran's I test for selected variables[a]

	Descriptive statistics		Wald test on normality		Moran's I				
	Mean	S.D.	Test statistics	Prob.	Test statistics	Normality assumption		Randomisation assumption	
						z-value[b]	S.D.	z-value[b]	S.D.
Millet	649.13	320.25	12.04	0.00	0.48	–	–	10.32***	0.053
N	112.59	17.72	2.94	0.22	0.25	3.97***	0.053	–	–
P	2.62	1.29	4345.19	0.00	0.09	–	–	2.45**	0.044
K	44.89	14.38	388.13	0.00	0.32	–	–	5.67***	0.051
Crust	3.23	0.55	0.95	0.62	0.29	5.91***	0.053	–	–
Topovariation	30.00	7.03	26.68	0.00	0.03	–	–	0.14	0.053
Cattle manure	310.42	323.77	38.37	0.00	0.63	–	–	12.14***	0.053
Sheep manure	27.66	23.52	36.62	0.00	0.56	–	–	9.76***	0.053

[a] The definition and measurement units of the variables are: millet: millet yield in $\mathrm{kg\,ha^{-1}}$; N: total nitrogen (Kjeldahl) in parts per million (ppm); P: phosphorus (Bray) in ppm; K: exchangeable potassium, originally measured in $\mathrm{cmol\,kg^{-1}}$ but transformed to ppm for reasons of comparability; Crust: a semi-quantitative field measurement of crust resistance on a scale of 1–5, where 1 is strong resistance/severe crusting and 5 low resistance/no crusting; Topovariation: the sum of the differences between the altitude of a grid cell and all neighbours defined using the queen criterion of adjacency, expressed in decimeters and rescaled to positive values; a high value refers to concave low-lying locations and a low value to convex high locations; Cattle manure: manure of cattle in $\mathrm{kg\,ha^{-1}}$; Sheep manure: manure of small ruminants in $\mathrm{kg\,ha^{-1}}$. The number of observations is 100.

[b] Significance is indicated with ***, ** and * for the 1, 5 and 10% level, respectively.

validity of using the asymptotic normal distribution for Moran's I.[8] Table 1 shows that for all variables, except for N and Crust, the null hypothesis of normally distributed observations is rejected. We, therefore, use the randomisation assumption to assess the degree of spatial autocorrelation for those variables. Table 1 shows that, except for Topovariation, the spatial distribution of attribute values is spatially autocorrelated. As the sign of the test statistic is positive, high (low) values are surrounded by high (low) values in neighbouring grids. The spatial clustering is strongest for millet yield (already visually presented above), and for the two types of manure. Manure is mainly restricted to the bottom half of the field, due to the presence of a deep well used for watering livestock just outside the field on that side.

For the modelling analysis, we use a traditional agro-economic yield function, the Cobb–Douglas specification:

$$y = \alpha + C\gamma + \varepsilon, \tag{12}$$

where y is the $(n \times 1)$ vector with logarithmic values of the quantity of output, C the $(n \times k)$ matrix containing the logarithmic values of the geophysical attributes, γ the $(k \times 1)$ vector with coefficient parameters, and ε a $(n \times 1)$ vector of iid errors. Specifications that are more appropriate from an economic and/or agronomic perspective, such as quadratic or flexible functional form specifications or plateau models, can be used as well, but we concentrate the analysis on demonstrating the relevance of accounting for spatial effects.

The first column of Table 2 refers to straightforward OLS estimation of a linearised (doublelog) version of the yield function, and the diagnostic tests show that there is ample evidence for heteroskedasticity in this specification. The combination of Lagrange Multiplier tests indicates that a spatial lag model is likely to be the correct specification, because the LMERR and LMLAG tests are both significant but the significance of the robust LMERR test is notably smaller whereas the robust LMLAG test remains significantly different from zero. These test results should be interpreted with caution because the asymptotic properties of tests for spatial dependence are crucially dependent on the assumption of normally distributed error terms, an assumption that is rejected (see the Jarque–Bera test in Table 2). Caution regarding the accuracy of the parameter estimates is also needed because of the relatively high degree of multicollinearity (see

[8] The SpaceStat software is used throughout the analysis (see Anselin, 1992, and http://www.spacestat.com).

Table 2
Estimation results for the loglinear Cobb–Douglas yield function using the OLS estimator, the maximum likelihood spatial lag estimator (MLLAG), and the maximum likelihood spatial lag estimator with groupwise heteroskedasticity (MLLAG + GHET)[a]

	OLS	MLLAG	MLLAG + GHET
Constant	−6.431***	−6.422***	−3.036**
	(−3.643)	(−4.703)	(−2.452)
N	1.936***	1.380***	0.752***
	(4.581)	(4.141)	(2.866)
P	0.256*	0.121	0.024
	(1.776)	(1.084)	(0.317)
K	0.896***	0.476**	0.323**
	(3.671)	(2.476)	(2.226)
W·MILLET		0.684***	0.723***
		(8.405)	(9.301)
R^2–adjusted	0.40	0.55	.39
F	23.209***		
Likelihood	−87.978	−69.107	−57.014
n	100	100	100
CN	85		
JB	68.794***		
(Spatial) BP/KB	20.721***	58.038***	24.186***[b]
Moran's I	5.030***		
LMERR	19.921***	1.107	
Robust LMERR	3.067*		
LMLAG	43.132***		
Robust LMLAG	26.279***		

[a] In parentheses t-values for OLS, and z-values for the ML estimators are presented. Significance is indicated with ***, ** and * for the 1, 5 and 10% level, respectively. The meaning of the abbreviations for the misspecification diagnostics is: CN is the condition number providing an indication for multicollinearity, JB a test on normality of the errors, (Spatial) BP/KB the regular (or alternatively spatial) variant of the Breusch–Pagan or Koenker–Bassett tests for heteroskedasticity, and Moran's I, LMERR, Robust LMERR, LMLAG, and Robust LMLAG are defined in Eqs. (7)–(11), respectively (see Anselin, 1992, for details).

[b] Likelihood Ratio test on groupwise heteroskedasticity.

the condition number, CN), although this is a factor endemic to these kinds of traditional specifications.

The magnitude and significance of the estimated parameter value for W·MILLET reveals that there is substantial positive correlation among yield values (see the second column of Table 2). There is no remaining spatial error correlation, but heteroskedasticity is still apparent. The positive and significant coefficient of the spatially lagged dependent variable (W·MILLET) indicates that millet yield observations for different grid cells cannot be considered in isolation. Exogenous changes in one grid cell are intrinsically linked to millet yield changes in neighbouring grid cells.

Subsequently, we use Moran scatterplots of millet yield to investigate whether a relevant spatial "cause" can be found for this heterogeneity. Fig. 2 (top) shows that most millet yield observations are within two standard deviations of the mean. Taking logarithms of yields suggest a bipartition, because grid cells with below average yield neighbours show a much higher dispersion than those with above average yield neighbours (Fig. 2 bottom). The last column of Table 2 therefore shows the results for a spatial lag model with groupwise heteroskedasticity, with two groups distinguished according to yield clusters defined by quadrant I and II, and quadrant III and IV in the Moran scatterplot, respectively. The Likelihood Ratio test on groupwise heteroskedasticity shows that the distinction in variance between the two groups is significantly different from zero. The results of the spatial lag model and the spatial lag model with groupwise heteroskedasticity are (except for N) very similar in terms of magnitude of the estimated coefficients, but the estimated standard errors are greater in the latter (resulting in lower z-values, which conforms to expectations).

The dramatic shift in coefficients when switching from an a-spatial or traditional to a spatial model and the high significance of almost all test statistics in Table 2 illustrates the resulting bias if one relies on a simple a-spatial OLS model instead of a spatial lag model (eventually with a correction for heteroskedasticity). It should be noted though, that the coefficients of the spatial lag model do no longer represent total marginal elasticities, as a change in one of the exogenous variables will "filter through" the whole spatial system given the dependence of local yields on yields of the neighbours. This is obvious from the reduced form of the spatial lag model $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \beta \boldsymbol{x} + \boldsymbol{\varepsilon}$ that reads as $\boldsymbol{y} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}(\beta \boldsymbol{x} + \boldsymbol{\varepsilon})$. The first derivative with respect to the input equals $\partial \boldsymbol{y}/\partial \boldsymbol{x} = [(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}]'\beta = [(\boldsymbol{I} - \rho \boldsymbol{W})']^{-1}\beta$, which is a matrix of elasticities containing the direct production elasticity on the diagonal and indirect elasticities generated through the linkages within the spatial system in the off-diagonal positions. In the spatial lag model, the elasticities are therefore cell-specific, and equal to the estimated β weighted by the column sum of the spatial transformation matrix $(\boldsymbol{I} - \rho \boldsymbol{W})^{-1}$.

The extremely poor soil chemistry of the field—witness the relatively low levels of N, P and K in

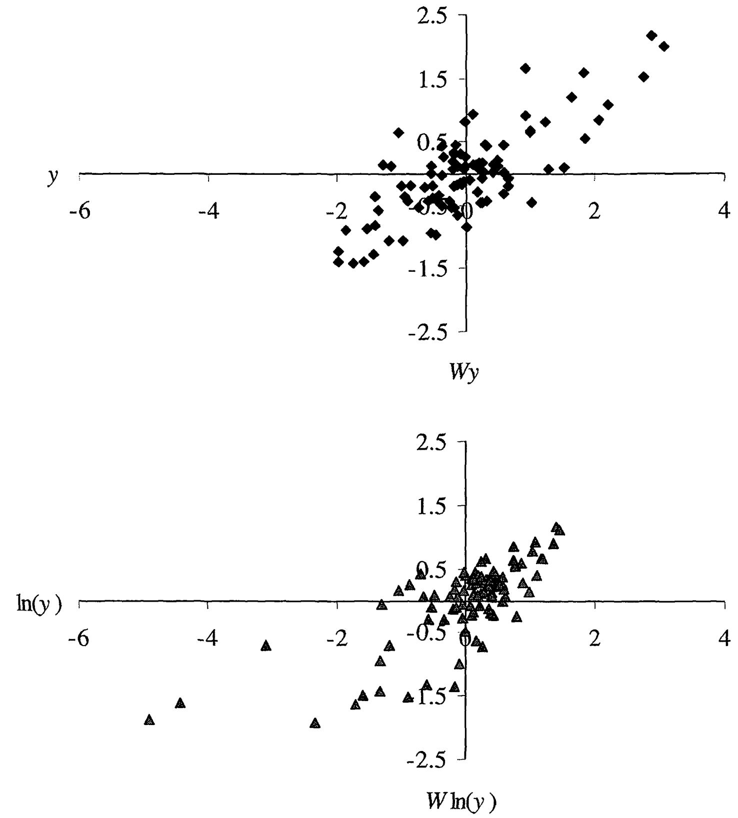

Fig. 2. Moran scatterplots with standardised grid yield on the horizontal axis and standardised grid yield of queen-based neighbours on the vertical axis (top; bottom for logarithmic yield).

Table 1—leads one to expect that the elasticity levels may be relatively high, and eventually exhibiting decreasing returns. Table 2 shows that the direct elasticities for the macronutrients are substantially lower for the spatial models. However, the average of the total elasticities, including the spatial spill-over effects of the neighbours,[9] show increasing returns for nitrogen and potassium (the average spatial elasticities are 2.715 for N, 0.087 for P, and 1.166 for K, for the spatial lag model with groupwise heteroskedasticity). Increasing returns may seem counterintuitive, but they are reasonable for this specific application since in particular for N and K the current levels of macronutrients are extremely low (see also Table 1).

[9] The above derivation for the total elasticity contains an inverse term that causes the elasticity for each grid cell to depend on all grid cells within the spatial system (see Anselin and Florax, 1995).

So far, the specification of the model has been driven by a rather mechanic approach, without taking into account substantive factors that may explain the spatial variation. Table 3, therefore, presents a model with a more extensive representation of spatial differences. In addition to N, P and K we include variables related to crusting, local topography, and manure levels. Column 1 of Table 3 shows that these spatially varying factors are very relevant, but they do not entirely remove the heteroskedasticity and the spatial autocorrelation. We, therefore, proceed by estimating a spatial lag model because the combination of Lagrange Multiplier tests points in that direction (see Column 2). The spatial lag model, however, still exhibits heteroskedasticity.

A more extensive exploratory spatial data analysis provides additional insight into the nature of heterogeneity and dependence. Based on stratigraphical evidence, the proportions of cations at the exchange

Table 3
Estimation results for a loglinear Cobb–Douglas specification using the OLS, the maximum likelihood spatial lag estimator (MLLAG), the maximum likelihood spatial lag estimator with groupwise heteroskedasticity (MLLAG + GHET), and the maximum likelihood spatial lag estimator with two regimes and groupwise heteroskedasticity (MLLAG + REG)[a]

	OLS	MLLAG	MLLAG + GHET	MLLAG + REG[b]		
				Soil type A	Soil type B	Chow
Constant	−6.150***	−6.243***	−3.868***	−7.803***	−1.216	6.490**
	(−4.371)	(−5.237)	(−3.405)	(−3.752)	(−0.830)	
N	1.336***	1.066***	0.722***	1.070*	0.513*	0.709
	(3.929)	(3.666)	(3.151)	(1.758)	(2.007)	
P	0.234*	0.152	0.049	0.875	0.023	2.003
	(1.975)	(1.507)	(.779)	(1.461)	(0.351)	
K	0.408**	0.219	0.078	0.630	0.065	1.530
	(2.024)	(1.247)	(.636)	(1.427)	(0.512)	
Crust	1.880***	1.562***	1.457***	1.000*	1.343***	0.354
	(6.332)	(6.020)	(7.366)	(1.852)	(6.262)	
Topovariation	0.316**	0.259**	0.101	0.252	−0.072	1.917
	(2.302)	(2.225)	(.902)	(1.334)	(−0.520)	
Cattle manure	0.114**	0.049	0.063**	−0.026	0.066*	1.110
	(2.415)	(1.199)	(1.972)	(−0.321)	(1.882)	
Sheep manure	0.197***	0.110*	0.045	0.337*	0.046	2.262
	(2.699)	(1.721)	(0.984)	(1.795)	(0.933)	
W·MILLET		0.524***	0.629***	0.495***		
		(5.746)	(8.062)	(5.233)		
R^2–adjusted	0.64	0.71	0.57	0.79		
Likelihood	−60.830	−50.902	−34.402	−26.299		
F	25.883***					
n	100	100	100	100		
CN	123					
JB	232.487***					
(Spatial) BP/KB	12.073*	66.665***	33.001***[e]	26.716***[e]		
Moran's I	2.220**					
LMERR	2.313	1.912[c]				
Robust LMERR	3.521*					
LMLAG	17.366***	19.856***[d]				
Robust LMLAG	18.574**					
Chow				17.470**		

[a] In parentheses t-values are reported for OLS, and z-values for the maximum likelihood estimators. Significance is indicated with ***, ** and * for the 1, 5 and 10% level, respectively. See footnote b to Table 2 for the meaning of the misspecification diagnostics.
[b] The coefficient for W·MILLET and the regression diagnostics refer to the whole sample (i.e., both soil types).
[c] Likelihood Ratio test on spatial error process.
[d] Likelihood Ratio test on spatial lag process.
[e] Likelihood Ratio test on groupwise heteroskedasticity.

complex and the related Al saturation profile, we distinguish three different soil types that derive from shallow layers of different coversand types (see also Voortman et al., 2002). Fig. 3 shows the spatial distribution of the different soil types. In the econometric analyses, we aggregate the different soil types to two categories, because Type C only contains two grid-cells. The spatial lag model is subsequently estimated with differing variances for those two groups. The estimation results are shown in Column 3 of Table 3.

In addition, we explore whether the heterogeneity can be given a substantive interpretation in the sense that two different spatial regimes apply for the different soil types. The last three columns of Table 3 show the results for this specification. The overall Chow test rejects the null hypothesis that the coefficients of

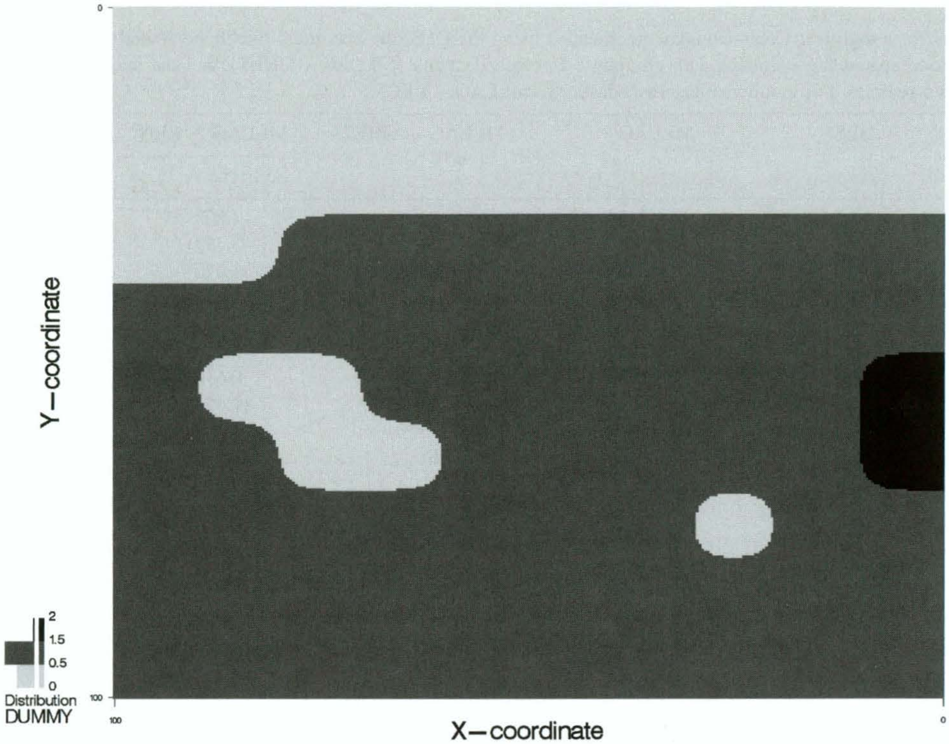

Fig. 3. Three-coloured choropleth map with three different types of parent material (coversands) in light gray (Type A), dark gray (Type B), and black (Type C).

the two regimes are the same, although Chow tests on the individual coefficients are not significantly different from zero. It is also striking that the estimated direct elasticities for N, P and K are substantially higher for Type A soils, where yields are very low, as compared to Type B soils. The effect of crusting is higher for Type B soils. Overall, there is evidence for significant spatial heterogeneity and dependence, and the spatial lag specification accounting for soil differences achieves the best fit in terms of the value of the Likelihood, although the greatest improvement in fit relates to the change from the a-spatial to the spatial lag model. The dependence is rather strong, and indicates significant spatial clustering of similar values. Clearly, the explanatory power of this specification is substantially better than the abovementioned 30% that is usually obtained in traditional (a-spatial) regressions.

Fig. 4 concisely shows the impact of taking into account spatial effects. The figure presents both the OLS and "spatial" elasticity for N. The OLS elasticity is given in Column 1 of Table 3, and the spatial elasticity is calculated for the spatial lag model with regimes and groupwise heteroskedasticity (the last three columns of Table 3). The irregularities at each 10-grid-cell-interval for the spatial elasticity show the impact of so-called edge effects. It is evident that the OLS elasticities are biased: they are overestimated for one part of the field, and underestimated for the other. This is also apparent from a comparison of the OLS-based elasticities reported in Table 3 and the averages of the spatial elasticities, which are 1.424, 0.670 and 0.543 for N, P and K, respectively. Fig. 4 further shows that although the OLS-based elasticity is constant, the implied yield changes of applying one more unit N vary spatially. However, there is a clear distinction to the yield changes implied by the spatial model, both in terms of heterogeneity (one can clearly see the two spatial regimes) and in terms of size (OLS overestimates the yield change in one part of the field and underestimates it in the other part).

The elasticities do not provide a usable treatment advice for the farmer. Because the inherent soil chemistry

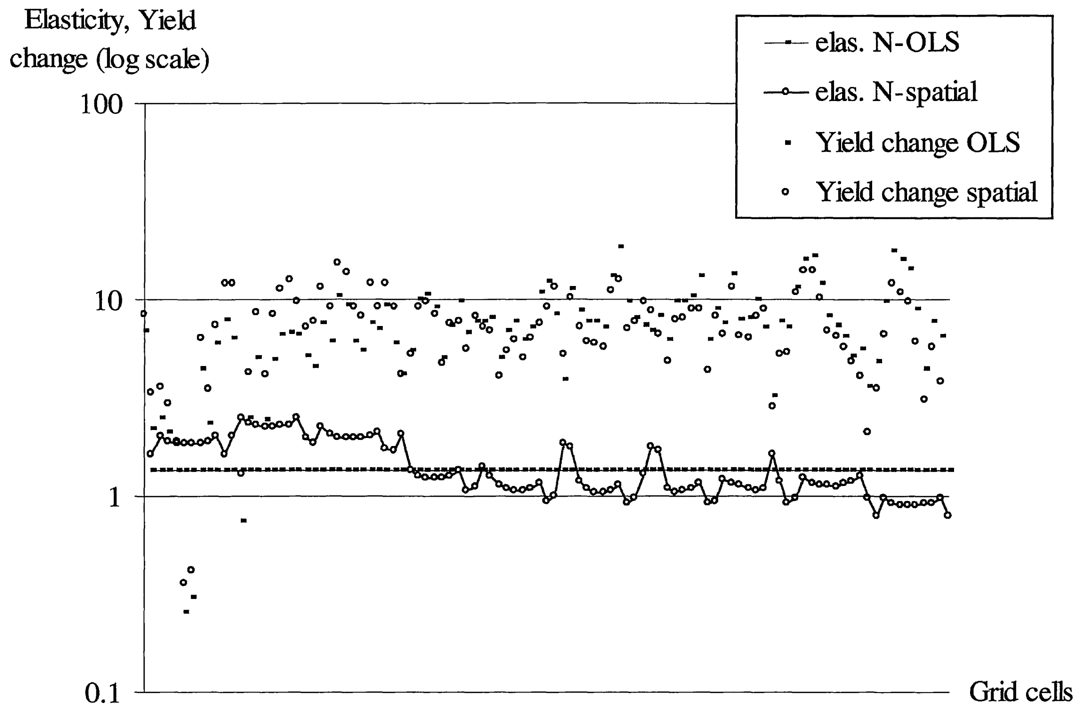

Fig. 4. Output elasticities and yield changes for N, based on OLS and the spatial lag model with two regimes and groupwise heteroskedasticity, on a log scaled vertical axis.

shows differing average levels of N, P and K (for instance, approximately 112 ppm for N, 3 for P, and 45 for K) vastly different amounts of fertiliser have to be applied to attain similar yield changes, even if the elasticities are the same. Taking into account the tight cash-constraints of local farmers, the relevant issue is to identify the area that is most effective in producing extra yield resulting from fertiliser use. We, therefore, choose to plot in Fig. 5 the yield changes associated with a one-unit increase in the macronutrients for the spatial regimes specification in Table 3 (i.e. the specification that allows for parameter variation between different soil types). Fig. 5 shows that N is generally least effective since its level expressed in ppm is the highest of the three macronutrients. Its use should be avoided in the core area of field part A, which includes a highly sealed remnant of a termite mound. In general, adding N will be most effective on grid cells in part B and C where currently very high yields are obtained (see also Fig. 1). The response to P is generally much higher than to N and K, but in part A even more so than in part B. However, the high effects of P are critically dependent on the assumption that applied P behaves as P-Bray in the soil. This may not be entirely the case, since the Bray method of analysis measures relatively easily available P. The response to K is greatest in part A (again with the exception of its core area) and the transition zone towards part B. In all three cases, the application of fertiliser to the core area of part A and the lower corners in part B prove to be wasteful if cash available for fertiliser is limited.

The suggested fertiliser treatment apparent from Figs. 4 and 5 is different from the advice that results from a traditional agronomic analysis (the OLS equation). Column 1 in Table 3 can be viewed as providing the basis for a traditional agronomic fertiliser prescription. The use of spatial statistical and econometric techniques results in spatially explicit treatment prescriptions that may vary over different spatial clusters, and they are based on unbiased estimates, because the bias related to ignoring inherent spatial linkages is avoided.

One can wonder what the potential relevance of the above findings is for poor farmers in the Sahel.

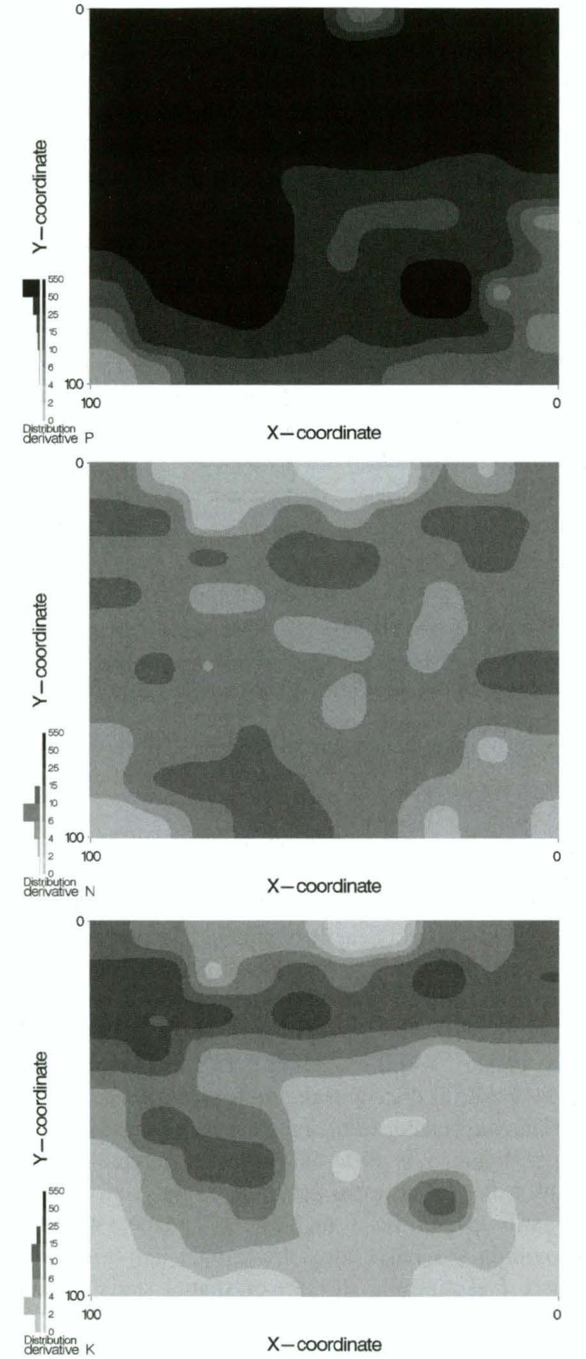

Fig. 5. Yield changes in $kg\,ha^{-1}$ per unit ppm increase of P (top), N (middle), and K (bottom) in the topsoil. Note: the yield change is least for the light gray areas and increases towards darker areas; the classes are: 0–2, 2–4, 4–6, 6–10, 10–15, 15–25, 25–50 and $>50\,kg\,ha^{-1}$ per unit ppm change.

Certainly, they cannot afford the expensive data gathering and subsequent labour-intensive laboratory and econometric analyses. Even if a governmental institution picks up this task, it will be impossible to cover the fields of all farmers. However, the soils of this area have a common formation history. They have developed in different kinds of coversand, but over substantial distances there appear to be unifying principles with respect to the factors that determine plant growth and crop yield (Voortman et al., 2002). These principles are likely to find expression in remnant of natural vegetation and arable weed communities (Bannink et al., 1974), which farmers obviously recognise. Well-targeted research on a limited number of representative sites across climatologic zones can thus provide information of relevance for research and extension for large areas.

Finally, the limiting factors in developing countries are generally labour and nutrients. As a rule there are no or only very limited resources available for additional manure or mineral fertiliser, or due to unreliable rainfall farmers consider it too risky to invest in such additional inputs (Brouwer and Bouma, 1997; Gandah, 1999). It is, therefore, not so much a matter of return to capital or return to land, as well as return to labour and return to local inputs that matters. In our case study the manure and urine had been deposited by livestock resting around a well, before and after being watered. There was no intentional pattern in the spatial distribution, other than that the animals preferred to stay close to the well. There was no addition of other nutrients. The nutrient contents of the soil (as measured in this study) did not include the nutrients stored in the manure still on the surface, but did include the effects on the soil of previous years. Our data and analyses show that there is a considerable over-supply of manure in the part of the field close to the well, while in other parts addition of manure will significantly increase yield. The same inefficient application pattern is often found on fields where livestock are camped at night (Gandah, 999). Provided the required labour is available at the right time of the year, farmers can considerably increase the efficiency of the more or less haphazard manure application at very low cost, by redistributing some of the manure to other parts of the field. Any additional inputs they can afford can be applied similarly.

6. Conclusions

In modern agricultural production systems precision agriculture is advocated because it constitutes an important means of handling local variations in soil characteristics when one wants to improve crop yields and increase input efficiencies. Precision agriculture depends on the use of a global (GPS) or local positioning system to generate spatial data, statistical and econometric tools to analyse such data in order to come up with highly localised optimal fertiliser prescriptions, and farm machinery capable of variable rate applications. In both high-tech and low-tech precision agriculture the method to determine soil variability, and the subsequent use of spatial analysis techniques to attain spatially optimised fertiliser mixes and doses, is important.

A review of the agronomy literature shows that ANOVA of block experiments and more extensive regression analyses of multi-factorial experimental treatment with repetitions has not yet resulted in a thorough understanding of yield variation. Much of the research has focused on macronutrients (N, P and K) in the topsoil, although recently other physical and environmental aspects are also considered and a start has been made with incorporating the influence of space. This paper further extends the analysis of the inherent spatial nature of crop yield variation. We discuss several techniques from spatial statistics and spatial econometrics, and apply those techniques to a case study concerned with millet production on acid sandy soils of the West African Sahel.

Spatial analysis techniques focus on detection and specification of spatial effects that may either show up in spatial heterogeneity or as spatial dependence (clustering). An exploratory analysis of crop yield and various other characteristics of the relevant parcel facilitates the detection of spatial effects. The use of Moran's I statistic, scatterplots, and GIS-based mapping are useful to assess the importance of intrinsic spatial linkages within the field and to investigate the extent of spatial heterogeneity in soil, ecology and yield. Subsequently, spatial effects can be assessed in a modelling context as well. Various test statistics are available, and the use of Lagrange Multiplier tests ensures that proper guidance to alternative specifications is safeguarded. The use of spatial econometric modelling techniques as compared to the traditional agronomic approach based on OLS regressions, contributes to a more effective use of precision agriculture for two main reasons. First, accounting for spatial dependence among different subsections of the field avoids the bias or inefficiency that would result from the use of the simpler OLS estimator. Second, the spatial heterogeneity can be explicitly modelled resulting in an analysis and subsequent fertiliser prescription that is much more accurate. Accounting for spatial effects also contributes to enhancing the explanatory power of agronomic yield models.

The application of spatial statistical and econometric techniques to millet yield data of a 1 ha plot in the Sahel has clearly demonstrated the advantages of using these techniques. Given our focus on the use of spatial techniques, several aspects have not yet been treated in depth in this case study. In particular, the specification of the functional form of the yield function and the selection of explanatory variables (regarding, for instance, physical and environmental characteristics, and varying depth of measurement) merit further attention. The specification of relevant chemical and physical processes can also be further refined, for instance tendencies to soil destabilisation and surface sealing induced by the soil chemistry can result in lower water infiltration and poor seedling establishment.

The current analyses clearly show that ignoring spatial effects results in considerable bias with respect to local yield variability in relation to macronutrients. The attainment of appropriate management strategies and the efficiency and effectiveness of precision agriculture therefore crucially depends on the application of state-of-the-art spatial analysis.

Acknowledgements

We would like to thank Andreas Buerkert and John Lamers for providing the crop yield and manure data, as well as Stephen Gaze, Diafaro Amadou, Idé Sanda, Djibo Soumaila, Zurkafili Amadou and Moussa Mahamane for their assistance in data collection. The field data were gathered when Joost Brouwer was at the present Laboratory of Soil Science and Geology, Wageningen University and Research Centre (WUR), based at ICRISAT Sahelian Centre in Niamey, Niger. Funding of his research by DGIS, Netherlands Ministry of Foreign Affairs, is gratefully acknowledged.

We have very much profited from comments by two anonymous reviewers and Gerald C. Nelson on an earlier version of the paper.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, 284 pp.

Anselin, L., 1992. SpaceStat: A Program for the Analysis of Spatial Data. National Center for Geographic Information and Analysis. University of California, Santa Barbara, 250 pp.

Anselin, L., 1995. Local indicators of spatial association (LISA). Geograph. Anal. 27, 93–115.

Anselin, L., 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer, M., Scholten, H., Unwin, D. (Eds.), Spatial Analytical Perspectives on GIS. Taylor and Francis, London, pp. 111–125.

Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell (Basil), Oxford, pp. 310–330.

Anselin, L., Griffith, D.A., 1988. Do spatial effects really matter in regression analysis? Papers of the Regional Sci. Assoc. 65, 11–34.

Anselin, L., Rey, S., 1991. Properties of tests for spatial dependence in linear regression models. Geograph. Anal. 23, 112–131.

Anselin, L., Florax, R.J.G.M., 1995. Small sample properties of tests for spatial dependence in regression models: some further results. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Directions in Spatial Econometrics. Springer, Berlin, pp. 21–74.

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 237–289.

Anselin, L., Bera, A.K., Florax, R.J.G.M., Yoon, M.J., 1996. Simple diagnostic tests for spatial dependence. Regional Sci. Urban Econ. 26, 77–104.

Bannink, J., Leijs, H.N., Zonneveld, I.S., 1974. Weeds as Environmental Indicators, Especially for Soil Conditions. Bodemkundige Studies 11, STIBOKA, Wageningen.

Bationo, A., Christianson, C.B., Baethgen, W.E., 1990. Plant density and nitrogen effects on pearl millet production in Niger. Agronomy J. 82, 290–294.

Bationo, A., Christianson, C.B., Baethgen, W.E., Mokwunye, A.U., 1991. Comparison of five soil testing methods to establish phosphorus sufficiency levels in soil fertilized with water-soluble and sparingly soluble-P sources. Fertil. Res. 28, 271–279.

Bongiovanni, R., Lowenberg-DeBoer, J., 2000. Nitrogen Management in Corn Using Site-Specific Crop Response Estimates from a Spatial Regression Model. In: Robert, et al. (Eds.).

Bongiovanni, R., Lowenberg-DeBoer, J., 2001. Precision agriculture: economics of nitrogen management in corn using site-specific response estimates from a spatial regression model. Proceedings of the Paper Presentation at the American Agricultural Economists Association Annual Meeting, Chicago, 5–8 August 2001, http://agecon.lib.umn.edu/cgi–bin/view.pl, 26 pp.

Bouma, J., 1997. Precision agriculture: introduction to the spatial and temporal variability of environmental quality. Proceedings of the CIBA Foundation Symposium 210 on Precision Agriculture: Spatial and Temporal Variability of Environmental Quality. Wiley, Chichester, pp. 231–235.

Brouwer, J., Bouma, J., 1997. Soil and Crop Growth Variability in the Sahel: Highlights of Research (1990–1995) at ICRISAT Sahelian Centre. Information Bulletin 49, ICISAT, Patancheru, Andrah Pradesh, India, 42 pp.

Brouwer, J., Powell, J.M., 1998. Microtopography and leaching: possibilities for making more efficient use of nutrients in African agriculture. In: Smaling, E.M.A. (Ed.), Nutrient Balances as Indicators of Productivity and Sustainability in sub-Saharan African Agriculture. Agric., Ecosyst. Environ. 71, 231–241.

Buerkert, A., 1995. Effects of Crop Residues, Phosphorus and Spatial Soil Variability on Yield and Nutrient Uptake of Pearl Millet (*Pennisetum glaucum* L.) in Southwest Niger. Verlag Ulrich E. Grauer, Stuttgart, 272 pp.

Buerkert, A., Hiernaux, P., 1998. Nutrients in the Sudano-Sahelian zone: losses, transfers and role of external inputs. Zeitschrift für Pflanzenernährung und Bodenkunde 161, 365–383.

Burridge, P., 1980. On the Cliff–Ord test for spatial correlation. J. R. Stat. Soc. B 42, 107–108.

Clay, D.E., Carlson, C.G., Chang, J., Clay, S.A., Malo, D.D., Ellsbury, M.M., Lee, J., 1999. Systematic Evaluation of Precision Farming Soil Sampling Requirements. In: Robert, et al. (Eds.), pp. 253–265.

Cliff, A.D., Ord, J.K., 1981. Spatial Processes: Models and Applications. Pion, London, 266 pp.

Cook, S.E., Adams, M.L., Bramley, R.G.V., 2000. What is Obstructing the Wider Adaptation of Precision Agriculture. In: Robert, et al. (Eds.).

Cressie, N., 1993, Statistics for Spatial Data, rev. ed. Wiley, New York, 900 pp.

De Graaff, T., Florax, R.J.G.M., Nijkamp, P., Reggiani, A., 2001. A general misspecification test for spatial regression models: heteroskedasticity, dependence, and nonlinearity. J. Regional Sci. 41, 255–276.

Eswaran, H., Beinroth, F.H., Kimble, J., Cook, T., 1992. Soil diversity in the tropics: implications for agricultural development. In: Lal, R., Sanchez, P.A. (Eds.), Myths and Science of Soils in the Tropics. SSSA Special Publication No. 29, SSSA/ASA, Madison, pp. 1–16.

FAO, 1983. Guidelines: Land Evaluation for Rainfed Agriculture. FAO Soils Bulletin 52, FAO, Rome, 237 pp.

Florax, R.J.G.M., De Graaff, Th., 2003. The performance of diagnostics for spatial dependence in linear regression models: a meta-analysis of simulation studies. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Berlin (forthcoming).

Florax, R.J.G.M., Folmer, H., Rey, S.J., 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. Regional Sci. Urban Econ. (forthcoming).

Gandah, M., 1999. Spatial variability and farmer resource allocation in millet production in Niger. Unpublished Ph.D. thesis. Laboratory of Soil Science and Geology, Wageningen Agricultural University, Wageningen, 115 pp.

Gandah, M., Bouma, J., Brouwer, J., van Duivendooden, N., 1998. Use of a scoring technique to assess the effect of field variability on yield of pearl millet grown on three Alfisols in Niger. Netherlands J. Agric. Sci. 46, 39–51.

Geiger, S.C., Manu, A., 1993. Soil surface characteristics and variability in the growth of millet in the plateau and valley region of Western Niger. Agric. Ecosyst. Environ. 45, 203–211.

Hermann, L., Hebel, A., Stahr, K., 1994. Influence of microvariability in sandy Sahelian soils on millet growth. Zeitschrift für Pflanzenernährung und Bodenkunde 157, 111–115.

Kessler, M.C., Lowenberg-DeBoer, J., 1999. Regression Analysis of Yield Monitor Data and Its Use in Fine Tuning Crop Decisions. In: Robert, et al. (Eds.), pp. 821–828.

Keyzer, M.A., Sonneveld, B.G.J.S., 1997. Using the Mollifier method to characterize datasets and models: the case of the universal soil loss equation. ITC J. 1997, 265–272.

Kilian, B., 2000. Economic Aspects of Precision Farming: A German Viewpoint. In: Robert, et al. (Eds.).

Klaij, M.C., Renard, C., Reddy, K.C., 1994. Low input technology options for millet-based cropping systems in the Sahel. Exp. Agric. 30, 77–82.

Krogh, L., 1999. Soil fertility variability and constraints on village scale transects in Northern Burkina Faso. Arid Soil Res. Rehabil. 13, 17–38.

Lamers, J.P.A., Feil, P.R., 1995. Farmers knowledge and management of spatial soil and crop growth variability in Niger, West Africa. Netherlands J. Agric. Sci. 43, 375–389.

Long, D.S., DeGloria, S.D., Griffith, D.A., Carlson, G.R., Nielsen, G.A., 1992. Spatial regression analysis of crop and soil variability within an experimental research field. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of Soil Specific Crop Management: A Workshop on Research and Development Issues, 14–16 April, ASA/CSSA/SSSA, Madison, pp. 365–366.

Manu, A., Pfordresher, A.A., Geiger, S.C., Wilding, L.P., Hossner, L.R., 1996. Soil parameters related to crop growth variability in Western Niger, West Africa. Soil Sci. Soc. Am. J. 60, 283–288.

Moorman, F.R., Kang, B.T., 1978. Microvariability of soils in the tropics and its agronomic implications with special reference to West Africa. In: Stelly, M. (Ed.), Diversity of Soils in the Tropics. American Society of Agronomy Special Publication No. 34, Madison, Wisconsin, pp. 29–43.

Nielsen, D.R., Wendroth, O., Jürschik, P., Kühn, G., Hopmans, J.W., 1997. Precision agriculture: challenges and opportunities of instrumentation and field measurement. In: Cruvinel, P.E., et al. (Eds.), Simposio Nacional de Instrumentação Agropecuaria. SIAGRO, EMBRAPA, CNPDIA, São Carlos, Brazil, pp. 65–80.

Nielsen, D.R., Wendroth, O., Pierce, F.J., 1999. Emerging Concepts for Solving the Enigma of Precision Farming Research. In: Robert, et al. (Eds.), pp. 303–318.

Pieri, C., 1985. Food crop fertilization and soil fertility: the IRAT experience. In: Ohm, H.W., Nagy, J.G. (Eds.), Appropriate Technologies for Farmers in Semi-Arid West Africa. Purdue University, West Lafayette, IN.

Rockström, J., de Rouw, A., 1997. Water, nutrients and slope position in on-farm millet cultivation in the Sahel. Plant Soil 195, 311–327.

Rockström, J., Barron, J., Brouwer, J., Galle, S., de Rouw, A., 1999. On-farm spatial and temporal variability of soil and water in pearl millet cultivation. Soil Sci. Soc. Am. 63, 1308–1319.

Scott-Wendt, J., Hossner, L.R., Chase, R.G., 1988a. Variability in pearl millet (*Pennisetum americanum*) fields in semiarid West Africa, Arid Soil Research. Arid Soil Res. Rehabil. 2, 49–58.

Scott-Wendt, J., Chase, R.G., Hossner, L.R., 1988b. Soil chemical variability in sandy ustalfs in semiarid Niger, West Africa. Soil Sci. 145, 414–419.

Stewart, C.M., McBratney, A.B., 2000. Development of a Methodology for the Variable-Rate Application of Fertiliser in Irrigated Cotton Fields. In: Robert, et al. (Eds.).

Sombroek, W.G., Zonneveld, I.S., 1971. Ancient dune fields and fluviatile deposits in the Rima–Sokoto river basin (N.W. Nigeria). Soil Survey Papers No. 5. Netherlands Soil Survey Institute, Wageningen, 109 pp.

Stein, A., Brouwer, J., Bouma, J., 1997. Methods for comparing spatial variability patterns of millet yield and soil data. Soil Sci. Soc. Am. J. 61, 861–870.

Voortman, R.L., Brouwer, J., 2002. An empirical analysis of the simultaneous effects of nitrogen, phosphorus and potassium in millet production on spatially variable fields in SW Niger. Nutrient Cycling in Agroecosystems (in press).

Voortman, R.L., Brouwer, J., Albersen, P.J., 2002. Characterization of Spatial Soil Variability and Its Effect on Millet Yield on Sudano–Sahelian Coversands in SW Niger. SOW–VU Working Paper 02–02, SOW–VU, Amsterdam, 29 pp.

Wendroth, O., Jürschik, P., Giebel, A., Nielsen, D.R., 1999. Spatial Statistical Analysis of On-Site Crop Yield and Soil Observations for Site-Specific Management. In: Robert, et al. (Eds.), pp. 159–170.

Wendt, J.W., Berrada, A., Gaoh, M.G., Schulze, D.G., 1993. Phosphorus sorption characteristics of productive and unproductive Niger soils. Soil Sci. Soc. Am. J. 57, 766–773.

Zonneveld, I.S., de Leeuw, P.N., Sombroek, W.G., 1971. An Ecological Interpretation of Aerial Photographs in a Savanna Region in Northern Nigeria. ITC publication, Series B, No. 63, ITC, Enschede, 41 pp.

Further reading

Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), 1999. Precision agriculture. Proceedings of the 4th International Conference, 19–22 July 1998, St Paul, MN. American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, Madison, WI, 1938 pp.

Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), 2000. Proceedings of the 5th International Conference on Precision Agriculture and Other Resource Management, 16–19 July, Bloomington, MN. American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, Madison, WI, CD-Rom.