PLANS FOR AN ECONOMIC EVALUATION OF A DRYLAND SALINITY PROGRAM IN YORNANING, WA

ANN COWLING

1. AIMS OF THE PROJECT

ABARE proposes to conduct a case study of a dryland salinity amelioration program being carried out by the Western Australian Department of Agriculture in the East Yornaning catchment. The Yornaning catchment is located 170 km south-east of Perth, and has an area of approximately 100 sq km. There are 16 farms in the catchment, which are involved almost exclusively in the production of wheat and wool. Increasing salinity in the area has resulted in decreasing production, and as a result of the salting, there are a number of large bare eroding patches.

The aims of the Yornaning project are to:

1. Develop models for linking economic performance with farm management practices, soil quality and geological structure.
2. Develop models for making these same links over time.
3. Perform a cost-benefit analysis of the Western Australian Department of Agriculture's farm management plan for the Yornaning catchment.

2. DATA

Data are available on a large number of variables describing aspects of economic performance, farm management practices, geological features and soil quality.

The catchment was overflown by a geological survey company several times in 1989–90 using a variety of remote sensing techniques. Hence most of the physical data are available measured on a 10 metre grid, whereas the economic and farm management practice data will be available on a paddock by paddock basis only. Paddocks will therefore be used as the units in the analysis.

3. STATISTICAL ANALYSIS

Measurements made on paddocks that are close together are generally more similar than those made on paddocks that are far apart, so that the data are likely to be positively correlated rather than independent. Spatial methods of analysis will therefore be used with these data in order to estimate mean parameters efficiently.

3.1 SPATIAL ANALYSIS USING VARIOGRAMS

A two-dimensional spatial data set has components

1. locations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n : \mathbf{x}_i \in D \subset R^2\}$;
2. data $\{Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_n)\}$.

In agricultural field trials, the locations at which the data are collected typically form a regular lattice, and ARIMA models can be used for the covariance in each of the directions of the axes as in CULLIS and GLEESON [3].

However, in the Yornaning context, the data come from an irregular lattice, and the above approach cannot be taken. The covariance structure will therefore be modelled using variograms, a method developed in Geostatistics (see, for example, JOURNEL and HUIJBREGTS [4]).

The variogram, $2\gamma(\mathbf{x}_1, \mathbf{x}_2)$, is defined by

$$2\gamma(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{var}[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]$$

and is related to the covariance function, $\mathrm{C}(\mathbf{x}_1, \mathbf{x}_2)$ when the covariance depends only on the distance $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$ between the two locations and not on the actual locations themselves. More formally, a spatial process is said to be stationary of order 2 when:

1. The mathematical expectation E[Z($\mathbf{x}$)] exists and does not depend on the location $\mathbf{x}$:
$$\mathrm{E}[Z(\mathbf{x})] = m, \quad \forall \mathbf{x}.$$
2. For each pair of rv's $\{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})\}$ the covariance exists and depends on the separation distance $\mathbf{h}$,
$$\mathrm{C}(\mathbf{h}) = \mathrm{E}[Z(\mathbf{x}+\mathbf{h}).Z(\mathbf{h})] - m^2, \quad \forall \mathbf{x}.$$

Under second order stationarity, the variogram is related to the covariance as follows:

$$\gamma(\mathbf{h}) = \mathrm{C}(\mathbf{0}) - \mathrm{C}(\mathbf{h}).$$

The hypothesis of second order stationarity assumes that the covariance exists, but the existance of the variogram is a weaker hypothesis allowing infinite variance, which corresponds to many physical phenomena. Consequently, the second order stationarity hypothesis is often slightly weakened to the intrinsic hypothesis:

1. The mathematical expectation $\mathrm{E}[Z(\mathbf{x})]$ exists and does not depend on the location $\mathbf{x}$:
$$\mathrm{E}[Z(\mathbf{x})] = m, \quad \forall \mathbf{x}.$$
2. For all vectors $\mathbf{h}$, the increment $[Z(\mathbf{x}+\mathbf{h})-Z(\mathbf{x})]$ has a finite variance that does not depend on $\mathbf{x}$,
$$\mathrm{var}[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h}), \quad \forall \mathbf{x}.$$

Although the condition $\mathrm{E}[Z(\mathbf{x})] = m$ appears restrictive, more complex mean surfaces can be fitted. CRESSIE [1] uses median polish to estimate the mean surface, then subtracts it off and uses $Z(\mathbf{x})$ to refer to the error terms which all have mean 0.

Moment estimation is used for the variogram:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} [Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]^2$$

where $|N(\mathbf{h})|$ is the number of distinct sample pairs lagged by the vector $\mathbf{h}$, and is estimated in a number of directions. If the variogram is anisotropic, that is, different in different directions, then a linear transformation of the coordinate system can be performed to induce isotropy, or otherwise different variograms have to be fitted in different directions.

A number of different variogram models have been found useful in practice (see, for example, JOURNEL and HUIJBREGTS [4]), two of which are given below:

<u>linear model</u>

$$\gamma(\mathbf{h};\theta) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ c_0 + b\|\mathbf{h}\|, & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

<u>exponential model</u>

$$\gamma(\mathbf{h};\theta) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ c_0 + c_e[1 - \exp(-\|\mathbf{h}\|/a_e)], & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

The variogram parameters can be estimated in any of the usual ways. ZIMMERMAN and ZIMMERMAN [6] performed a Monte Carlo comparison of seven methods of estimation using the linear and exponential models above, and concluded that no estimator was uniformly superior and that all give unbiased predictors of the mean. They therefore suggest that computationally simple methods such as ordinary least squares or weighted least squares be used rather than computationally intensive methods such as restricted maximum likelihood (REML) or maximum likelihood. However, in the case in which more than one mean parameter is being estimated, it is likely that REML would produce superior estimates.

Once the variogram has been estimated, the covariance matrix can be obtained, and the mean is estimated using generalised least squares.

3.2 THE FIRST YEAR'S DATA

At this stage, it is envisaged that a linear model will be fitted, of the form

$$Y(\mathbf{x_i}) = f(P_1(\mathbf{x_i}), \ldots, P_P(\mathbf{x_i}), M_1(\mathbf{x_i}), \ldots, M_M(\mathbf{x_i})) + \epsilon(\mathbf{x_i}),$$

where the locations $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ are the paddock centres, $Y(\mathbf{x_i})$ is an economic variable such as gross margin measured on the ith paddock, $P_1(\mathbf{x_i}), \ldots, P_P(\mathbf{x_i})$ are measurements of physical variables such as conductivity and soil type measured on the ith paddock, $M_1(\mathbf{x_i}), \ldots, M_M(\mathbf{x_i})$ represent farm management practices such as crop planted, tillage practices, fertilizer application rate used on the ith paddock, and $\epsilon(\mathbf{x_i})$ is the error for the ith paddock.

Initial estimates of the mean parameters will be made assuming that $\epsilon(\mathbf{x_1}), \ldots, \epsilon(\mathbf{x_n})$ are independent. The estimated surface will then be subtracted from $Y(\mathbf{x_1}), \ldots, Y(\mathbf{x_n})$ giving residuals $R(\mathbf{x_1}), \ldots, R(\mathbf{x_n})$, which are our estimates of the errors. The errors, which all have zero expectation, will then be modelled using variograms, and the correlation matrix of the errors will be used to re-estimate the mean parameters using generalised least squares. Successive cycles of re-estimation of covariance then mean parameters can be performed if desired.

3.3 LONGITUDINAL DATA

Changes in farm management strategy do not generally show an immediate effect on the water table level. It generally takes at least 5 years for the water table level to drop under a plot of trees. The study in the Yornaning catchment will therefore be continued for at least the next 5 years to monitor the effect of the various management choices on economic performance and salinity level. This information should be of use in identifying management strategies that are effective in reducing salinity, and at the same time are profitable for the farmers.

In modelling over time, a model of the form

$$Y(\mathbf{x_i}, t) = f(P_1(\mathbf{x_i}, t), \ldots, P_P(\mathbf{x_i}, t), M_1(\mathbf{x_i}, t), \ldots, M_M(\mathbf{x_i}, t)) + \epsilon(\mathbf{x_i}, t),$$

where the locations $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ are again the paddock centres, $t \in T = \{1, \ldots, 5\}$ is the year, $Y(\mathbf{x_i}, t)$ is the economic variable such as gross margin measured on the ith paddock in the tth year, $P_1(\mathbf{x_i}, t), \ldots, P_P(\mathbf{x_i}, t)$ are measurements of physical variables such as conductivity and water table height measured on the ith paddock in the tth year, $M_1(\mathbf{x_i}, t), \ldots, M_M(\mathbf{x_i}, t)$ represent farm management practices such as crop planted, tillage practices, fertilizer application rate used on the ith paddock in the tth year, and $\epsilon(\mathbf{x_i}, t)$ is the error for the ith paddock in the tth year. It is expected that the covariance structure over time can be modelled using an ARIMA process, and the spatial covariance can be estimated as in Section 3.1. Then, provided the process is separable (see MARTIN [5] for a definition of separable lattice processes), the covariance variance structure of the whole data set can be easily obtained, and the mean parameters estimated using generalised least squares as in Section 3.3 above.

4. REFERENCES

[1] CRESSIE, N. (1986). Kriging nonstationary data. *JASA*, **81**, 625–634.

[2] CRESSIE, N. (1991). *Statistics for Spatial Data.* Wiley, NY.

[3] CULLIS, B. and GLEESON, A. (1991). Spatial analysis of field experiments—an extension to two dimensions. *Biometrics*, to appear.

[4] JOURNEL, A. and HUIJBREGTS, C. (1978). *Mining Geostatistics.* Academic Press, London.

[5] MARTIN, R. (1990). The use of time series models and methods in the analysis of agricultural field trials. *Communications in Statistics–Theory and Methods*, **19**, 55–81.

[6] ZIMMERMAN, D. and ZIMMERMAN, B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, **33**, 77–91.

Australian Bureau of Agricultural and Resource Economics
GPO Box 1563
Canberra ACT 2601
Australia