

SAMPLING DESIGN FOR QUANTITATIVE INVENTORY OF THE IRRIGATED COTTON SOIL

I.O.A. Odeh and A. B. McBratney
CRC for Sustainable Cotton Production, Department of Agricultural Chemistry
and Soil Science, University of Sydney, Sydney NSW 2006

INTRODUCTION

There is a growing need for soil information for modelling soil processes, land use planning and environmental protection. The merit of decisions regarding these issues, especially the environmental impact analysis, is influenced by accuracy and quality of the available information. Such quality soil information for environmental management is often lacking for most of the arable land of Australia.

In the cotton-growing region of the northern New South Wales and southern Queensland, soil information with quantified accuracy is unavailable or exists in scattered and uncoordinated patterns. Because of this inadequacy, a new field survey is needed to efficiently sample the soil of the region in order that unbiased estimates of the current soil condition can be made with known precision. The estimates are required for a broad statement of baseline status of the soil in the region. This in turn will enable consideration of environmental protection and future assessment of environmental impacts, cotton production activities in particular. Thus, explicitly defined sampling scheme, which is easily repeatable, will be designed and implemented.

The specific aims of this project are to:

- 1) establish a repeatable field survey scheme of key soil variables influencing cotton production in the lower Namoi, Gwydir and Macintyre valleys;
- 2) obtain a quantitative statement of the current status of these variables;
- 3) develop spatial prediction models for optimal estimation of the variables in each of the valleys;
- 4) build up a soil data base incorporated into a Geographical Information System (GIS).

METHODOLOGY

Preliminary analysis

In a project like this that involves enormous amount of field sampling and laboratory analysis, proper planning at the preliminary stage is very essential. In collaboration with the Dutch Scientist, Dr. J.J. de Gruijter (who was on an CRC exchange visit in mid-March to early April, 1994), a survey scheme is being designed for the project. Considering the aims and goals of the project, two types of request (i.e., what the project aims to satisfy): "how much" and "where" (Domburg et al., 1994) will need to be met. By the "how much" request, the variables determined will describe the current status of the soil and, "where" requires that separate spatial means of the variables are estimated for each valley and/or combinations of strata. The accuracy of the estimates is dependent on the sampling scheme and actual sampling to be carried out from a common budget for the three valleys. It is therefore pertinent that the project starts with an analysis of the request types for the survey and the available possibilities that there exist to implement the project. Among the possibilities are the statistical analysis to be adopted and the methods of measurements of the variables in consideration. The next stage is to design the survey scheme which can then be executed meticulously following the laid down plan. Having decided the types of request, we have initialised their analysis and the sampling scheme will be designed and implementation.

Soil variables to be determined

The soil variables to be determined are those that describe the current status of the soil (the degree of chemical and physical degradation of the soil) which connotes the request "how much." Based on this criterion a number of variables have been selected for determination:

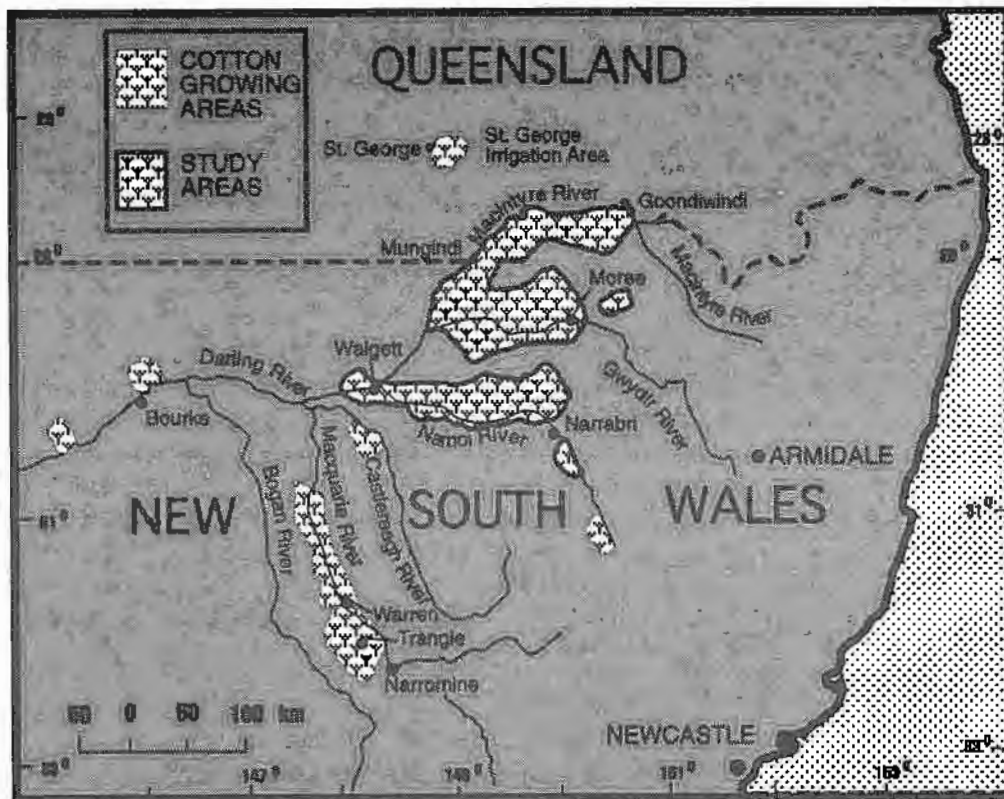


Figure 1. Location of the study areas in the cotton-growing region of New South Wales and the southern Queensland

Electrical conductivity and soil pH:

pH is a measure of soil reaction; it affects chemical activities in the soil, hence availability of some plant nutrients. Electrical conductivity (EC) is a measure of salt concentration in the soil which, in turn, is determined by intensity of salinisation processes; the amount of salts in the soil has some physiological effects on certain crops to a varying degree.

Bicarbonate extracted phosphorus (P):

This is a measure of soluble P which connotes the available P, and hence P fertility.

Exchangeable basic cations and Cation Exchange Capacity (CEC):

The amount of basic cations is a measure of their availability to plants; CEC influences physico-chemical properties of the soil and, in conjunction with *exchangeable sodium*, can be used to quantify sodicity which affects structural and physical processes in the soil.

Organic carbon (OC):

The amount of OC is important as a structural

constituent in addition to its usefulness as a nutrient source for plants.

Particle-size fractions (clay and silt):

The percent clay and silt affect the physical and chemical processes of the soil and hence the latter's overall fertility and degradability.

Sampling Design

Target variable for the design of sampling scheme

Before proceeding with the design of sampling scheme, a target variable, which to some extent is correlated with the variables above and with other soil properties that are of direct agronomic and environmental importance, will need to be chosen. The target variable thus selected will be used to decide details of the chosen sampling design and for modelling the accuracy and cost of the sampling design, done prior to commencement of actual sampling.

Target parameter

The target parameter is the spatial means or averages of the variables determined for each of the three valleys. The spatial means of the soil variables will describe the soil condition or land qualities important for sustainable production of cotton in the study areas. Furthermore, the sample data can be used to compute sample variances of the estimated means, spatial variances, quantiles, areal proportions and any of these statistics within each subregion or a stratum or combinations of strata.

Preliminary Plan of Field Survey

We have initiated a survey plan starting with the analysis of the requests that need to be fulfilled. Based on the aims of the project, we have identified the request type as "how much" and "where" for each of the cotton growing areas in the Namoi, Gwydir and Macintyre valleys.

Prior information and soil spatial variability

Existing information that will be useful for the preliminary analysis have been collected. The

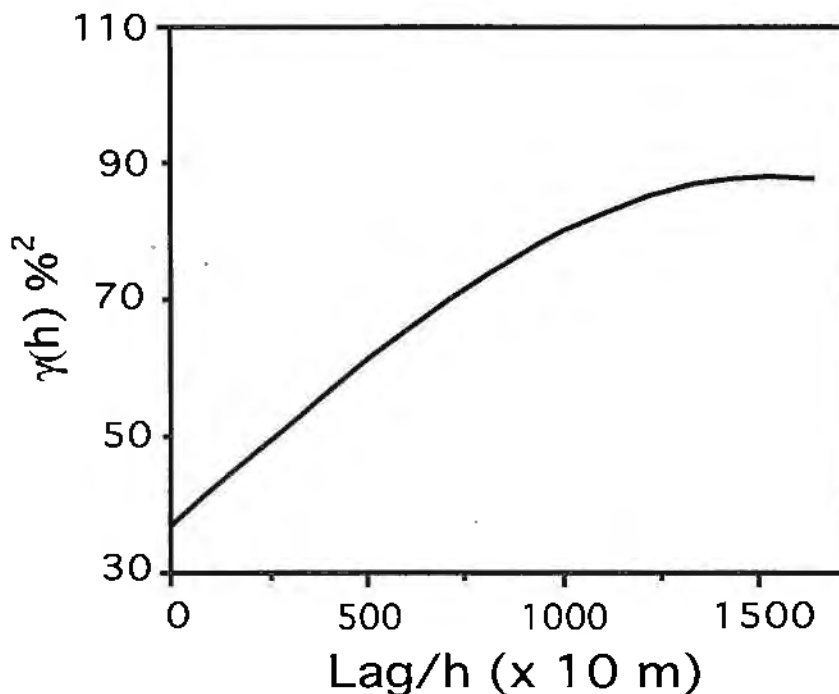


Figure 2. Model semivariogram of the target variable, % clay at 30-40 cm depth

Use of existing information

Given the aims and existing information, an appropriate sampling design will be chosen. The prior information will be used for stratification, if necessary, by dividing the heterogeneous population in each of the subregions into subpopulations that are relatively more homogeneous. This will improve the accuracy of estimates of the entire population in each subregion.

information includes: the Edgeroi soil data that cover the Edgeroi 1:50,000 Topographic Map Sheet (McGarry, et al., 1989); 1:50,000 Topographic maps presumably covering the cotton growing areas in the three valleys; and annual average rainfall data in stations within the region encompassing the three valleys. Interpolation of the rainfall data will be used for stratification for sampling purpose. Additional information that needs to be obtained is the specific farmlands that are or have been used for irrigated cotton production in the three valleys. Satellite

imagery and Shire Maps have been obtained for identification and delineation of cotton farms.

Having perused the Edgeroi soil data, percent clay content at 0.30-0.40 m depth has been selected as the target variable to be used for prediction of accuracy of the survey scheme at a given cost. Clay content at this depth affects many of the pedological and edaphic processes that occur in the immediate depths above and below; it has considerable correlation with many of the variables mentioned above. Based on variation of the percent clay (a graphic representation of the average variances of a variable and the series of distance measures, see Figure 2), total time (three years), the available operational funds to cover the three valleys and the logistic constraints (field accessibility, distance between sampling sites, etc.), maximum sampling size is determined. The budgets for travel and operations are based on this analysis.

Soil variability

After obtaining the calculated variation of the target variable, we fitted a mathematical (spherical) model (Figure 2) to obtain parameters that are useful for designing sampling strategies. The model is in the form:

$$\begin{aligned} \gamma(h) &= 37 + 0.0049h & \text{for } h > 15484 \text{ dam} \\ \gamma(h) &= 37 & \text{for } h < 15484 \text{ dam,} \end{aligned}$$

h being the spatial distance in multiples of 10 m, and $\gamma(h)$ is the half the average squared differences of values of a variable over distance, h .

These parameters were estimated from 125 point data from a cotton-growing area in Edgeroi within the Lower Namoi valley. Assuming similar variation, these parameters will be used as estimates for the Gwydir and Macintyre valleys. The parameters are required for initial prediction of accuracy and cost of survey.

Derivation of maximum sample size given some constraints

Time constraint:

We assume that time available for both the fieldwork and laboratory analysis is approximately 240 days over three years. The logical relations

are as in the following:

$$\begin{aligned} n_{\max}(\text{time}) &= \min\{T_f/t_f, T_l/t_l\} \\ T_f + T_l &= 240 \end{aligned}$$

where:

n_{\max} = maximum sample size due only to time constraint; t_f = average time in field per sample point, assuming help of one assistant (days); t_l = average time in lab per sample point, assuming help of one assistant (days); T_f = total time spent in field (days); T_l = total time spent in lab (days).

(a) Estimation of t_f

From Edgeroi survey experience in drilling to 2 m, t_f can be calculated as

$$t_f = 1/6 \text{ days}$$

(b) Estimation of t_l :

Time needed per sub-sample (including preparation) = 1 hour;
Number of sub-samples per sample point = 4
Available time per day (including 1 assistant and assuming 7 hr. day) = $2 \times 7 = 14$ hours

$$t_l = 4/14 = 1/3.5 \text{ days}$$

(e) Maximize n_{\max} by equating the number of sample points visited in the field to the number of points processed in the laboratory:

$$\begin{aligned} T_f/t_f &= T_l/t_l = 6 T_f = 3.5 (240 - T_f) \\ T_f &= 88 \text{ days} \\ n_{\max}(\text{time}) &= 6 \times 88 = 528 \end{aligned}$$

Cost constraint:

Money available over three years:
= \$ 40 000

(a) Cost of fieldwork (c_f)

Accommodation =	\$ 90 /day
Transport =	\$ 65 /day
Sampling equipment =	\$ 23 /day
Assistant (hired half-time) =	\$ 70 /day
Total =	\$ 248 /day

Total cost of field work (c_f):

$$c_f = 248/6 = \$ 42 \text{ /point}$$

(b) Cost of laboratory work (c_1):
Material (\$ 8 per sub-sample, including \$ 2 for
CEC by external lab):

\$ 32 /point

Assistant (hired on half-time bases):

70/3.5 = \$ 20 /point

Total cost of laboratory work (c_1):

$c_1 = \$ 52$ /point

(d) Total cost per point: $c_f + c_1 = \$ 94$

(e) $n_{\max}(\text{cost}) = 40,000/94 = 425$

Accessibility constraint:

(a) Time field accessible per year = 2
months, given a total of 120 days

(b) $n_{\max}(\text{access}) = 120 \times 6 = 720$

Maximum sample size given the bottleneck constraint

$n_{\max} = \min\{ n_{\max}(\text{time}), n_{\max}(\text{cost}), n_{\max}(\text{access}) \} = \min\{ 528, 425, 720 \} = 425$

The bottleneck constraint is thus the cost. This implies that keeping the costs as low as possible and/or increasing the budget has a direct effect on the survey scheme.

Outline of Survey Scheme

From the design information and the maximum sample size the following outline of survey plan was generated as a first approximation.

Method of determination:

The method of determination for the target variable (clay percentage) will be the standard laboratory analysis used for routine soil survey, i.e., the hydrometer method. Field methods such as "feel" will give too large measurement errors; the other methods e.g. pipette method would be too costly. The estimated standard error of measurement by this method is about 2-5 %. The consequence of this choice is that the estimated model of variation (Figure 2) need not be adjusted because the data on which it was based

were determined using the same method. Accuracies predicted from this model will thus automatically take into account the correct measurement error.

Type of sampling design

From preliminary analysis, stratified simple random sampling (SSRS) (Webster and Oliver, 1990) with equal probabilities of selection and replacement has been chosen for the project. Delineation of the strata within the external boundaries of the irrigated cotton in the three valleys is being done using a satellite imagery and Shire Maps. Hopefully, this will be completed by the end of August 1994 and the final sampling plan will be ready soon after.

Stratification:

In principle, strata will be square in shape, but where a square is intercepted by the outer boundary of the region, the part inside the region is added to an adjacent square to form a larger stratum. The main reason for selecting SSRS is because the maximum sample size in each of the subregions is too large (> 100) for simple random sampling over each of the subregions to be efficient. Stratification variables will be geographical coordinates. No other correlated variables seem to be readily available for stratification. The existing soil map that covers the entire region does not differentiate soil types. Stratification according to irrigation or landuse (cotton, no cotton) seems too cumbersome (the exact boundaries are not available at present and they will be partly outdated at the time of re-sampling).

Stratum boundaries and size:

Square strata will be used because of simplicity of delineation on topographic maps and reduced variance within strata under the assumption of equal variation in all directions. The sampling size should be as small as possible for greater efficiency and accuracy. At least 4 points per stratum will be sampled. Expensive variables, e.g. bulk density, which are time consuming, requiring the use of a drilling rig, can be measured in a subset (1/2 or 3/4) of the total sample points while still having the same type of sampling design and stratification for that subset.

Given the minimum of 4 points per stratum, the maximum number of strata is $425/4 = 106$;

Total areal size of the combined subregions
 $\cong 15600 \text{ km}^2$

The minimum stratum size is calculated as:
 $15600/106 = 147 \text{ km}^2$, or approximately
 $12 \times 12 \text{ km}$.

Sampling within strata:

Simple Random sampling (SI) will be adopted for sampling within the strata. SI is based on equal chance of drawing all the sampling points. The design is chosen because of its robustness against a possible forced reduction of the sample size in case actual field or laboratory work may require that cost and/or time constraints will need to be readjusted. At present there is relatively little experience with fieldwork for random sampling under these circumstances. The laboratory is not normally involved in routine work at this scale.

An additional advantage of stratification by relatively small squares is that with the simple random sampling within strata, there is the potential usefulness of the sample data for mapping at a later stage, both for modelling soil variation and prediction of soil at unvisited sites. Sample points at short mutual distances will be available for modelling soil variation while, together with supplementary sampling, the spatial configuration as a whole will be sufficiently evenly spread to be efficient for prediction.

Using other sampling designs, such as cluster or two-stage stratified sampling, may result in a significant reduction in average access-time (for travelling from point to point). However, such tight spatial clustering of sample points would, as expected, lead to a decrease in accuracy more than compensated by the greater economy of fieldwork.

POTENTIAL BENEFITS

The explicitly defined sampling design can be repeated with known precision and hence, will/ can be adopted for monitoring environmental

impact of irrigation and other management practices. Soil information of known accuracy (in comparison with conventional methods which frequently produce data of unknown quality) will be provided. The basic soil data will be available to other subprograms of the CRC for Sustainable Cotton Production (e.g., Chemical leaching and residue, nutrient balance, and irrigation and salinity management). A soil database will be developed and incorporated into a GIS for information storage, retrieval and analysis. Agro-technological transfer, in terms of potential use of the innovative sampling strategy in other areas of Australia and in terms of soil information transfer through land managers, extension workers and Growers Associations to the farmers, will be achieved.

REFERENCES

- Domburg, P., De Gruijter, J.J. and Brus, D.J. (1994). A structured approach to designing soil survey schemes with prediction of sampling error from variograms. *Geoderma*, 62: 151-164.
- Cochran, W.G. (1977). *Sampling Techniques*. 3rd ed. Wiley, New York, 428 pp.
- McDonald, R.C. et al. (1984). *Australian Soil and Land Survey Field Handbook*. Inkata Press, Melbourne.
- McGarry, D., Ward, W.T. and McBratney, A.B. (1989). *Soil Studies in the Lower Namoi Valley- Methods and Data Set*. CSIRO, Australia Division of Soils.
- Webster, R. and Oliver, M.A. (1990). *Statistical Methods in Soil and Land Resource Survey*. Oxford Univ. Press, Oxford, 316 pp.

ACKNOWLEDGEMENT

The sampling plan was done in collaboration with Dr J.J. de Gruijter of the DLO Winand Staring Centre for Integrated Land, Soil and Water Research, Wageningen, The Netherlands, whilst on an CRC exchange visit during March - April, 1994.