英100654

Monographs on Soil Survey

Quantitative and numerical methods in soil classification and survey

R. Webster

北京市农业科学院 书刊资料

Quantitative and numerical methods in soil classification and survey

R. WEBSTER

CLARENDON PRESS · OXFORD
1977

Oxford University Press, Walton Street, Oxford OX2 6DP
OXFORD LONDON GLASGOW NEW YORK
TORONTO MELBOURNE WELLINGTON CAPE TOWN
IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
KUALA LUMPUR SINGAPORE JAKARTA HONG KONG TOKYO
DELHI BOMBAY CALCUTTA MADRAS KARACHI

Oxford University Press 1977

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Oxford University Press

British Library Cataloguing in Publication Data

Webster, R
Quantitative and numerical methods in soil classification and survey. – (Monographs on soil survey).
1. Soil-surveys – Statistical methods
2. Soils – Classification – Statistical methods
I. Title II. Series
631.4'7 S591

ISBN 0-19-854512-6

Typeset at the Alden Press Oxford London and Northampton
Printed in Great Britain
by J.W. Arrowsmith Ltd., Bristol

Foreword

Robin Clarke's classic *A study of soil in the field* still has much to say to field pedologists and soil surveyors after forty years.

Even so, no science remains static. As our understanding of the soil in the field has increased, the aims of soil survey and land evaluation have broadened and new techniques have been developed to achieve them.

Forty years ago even qualitative methods of describing and comparing soils were themselves matters of research, and of course there were no computers to tackle lengthy or complex calculations. Soil scientists wishing to make their work quantitative have for the most part had to adapt the results of statistical theory to their particular circumstances, often with little guidance. Dr. Webster has often found himself in this position, and here he records that part of his experience relevant to other pedologists. This is a textbook on quantitative procedures for soil scientists, written by a soil scientist.

It is the second of a series of new handbooks of which some are intended for reference in the field and some, like this one, are intended to inform the field scientist and guide him in the application of sound modern technique.

Handbooks on remote sensing and air photography, soil description and field records, soil classification, soil survey contracts, and land evaluation are also in preparation. The editors will appreciate any suggestions for further titles.

P.H.T. BECKETT
V.C. ROBERTSON

Preface

This book is addressed to those scientists – pedologists, agronomists, engineers and ecologists – who study the soil by what are broadly speaking survey methods as distinct from experiments. It is written for those who observe, record and analyse information about the soil with its ever-present spatial variation over which they have no experimental control. It describes methods for making survey quantitative, stressing the need for measurement, sensible estimation, and proper planning. It discusses the role of classification, indicating those situations in which it can be helpful. It explains why data should be obtained in particular ways, why certain forms of statistical analysis are appropriate, and illustrates the methods with examples. It aims always to help the reader to choose suitable techniques for tackling his problem.

Some of the examples drawn from my work in the Soil Survey have not previously been published, and I am most grateful to Mr. K.E. Clare for allowing me to use the material here and for his encouragement. I thank Dr. T.R. Harrod, Messrs. S.J. Staines, D.V. Hogan, W.M. Corbett, and Dr. S. Nortcliff whose data I have mapped in Chapter 12, and Mr. M.G. Jarvis who provided the measurements from which Fig. 3.3 was obtained. I also thank Dr. P.A. Burrough for the data used to illustrate regional classification (Chapter 9), analysis of dispersion (Chapter 10), and optimization (Chapter 11).

Over the years several people have discussed the application of statistics and computers to soil survey with me and given me a deeper insight of the subject. I am especially grateful to Dr. F.H.C. Marriott, not only for much helpful advice but also for reading and criticizing the whole of this text. I thank Mrs. B.M. Hersom, Dr. D. Rhind, and Dr. S.W. Bie for suggesting improvements to Chapters 2 and 12, Dr. P.H.T. Beckett for his long and stimulating interest in the subject and for editing my script, and Mrs. M. Cox for typing it. I also thank the authors and publishers for permission to reproduce the following figures and tables: Blackwells, Fig. 3.1; Dr. A.W. Moore and Messrs. Angus and Robertson, Fig. 7.9; Elsevier Scientific Publications, Figs 12.1 to 12.4; John Wiley and Sons, Table 4.1; and Dr. K. Kyuma and the Society of the Science of Soils and Manures of Japan, Tables 8.3 and 8.4.

Oxford
October 1976

R.W.

此为试读,需要完整PDF请访问: www.ertongbook.com

Contents

1. Introduction

A sensible philosophy controlled by a relevant set of concepts saves so much research time that it can nearly act as a substitute for genius.

N.W. PIRIE
Concepts out of context

Classification and measurement

For centuries peasant husbandmen have lived in close harmony with the soil they till. They have learned how the soil responds to their treatment of it and to classify it according to its appearance and behaviour. They have also recognized where the soil changes from one kind to another in the landscape, and divided their land into parcels to be managed more or less differently. Classification is a practical tool by which man traditionally deals with his environment and with the soil in particular. It is also the means by which he communicates information about soil to his neighbours and descendants, matches soil in different places, and predicts behaviour where experience is lacking.

The first soil scientists adopted this approach, both for practical purposes and for more fundamental understanding. There were big differences to be seen and placed on record. The layman's classes and descriptive terms were meaningful and convenient. But once the more obvious distinctions had been made, pedologists turned their attention to finer differences. And in the practical sphere agronomists and engineers needed to describe the soil with which they worked more exactly and consistently, and to predict behaviour more precisely. In both cases the desired consistency and precision could be achieved only by measurement, and so, as in many other branches of science, observation became quantitative. Thus, to describe the soil at some place as 'acid' (classification) was no longer adequate; scientists wanted to say how acid, and they devised methods of measuring its acidity in terms of pH. There is now a large body of literature concerned with individual properties of the soil and how to measure them.

However, there is more to measurement than this. The soil is a more or less continuous mantle. The scientist cannot record what it is like everywhere: he can at best measure properties, whether directly in the field or on material taken into the laboratory, of small portions of the

mantle – that is, from a *sample*. Soil also varies from place to place, often very considerably, so measurements of the soil at one sampling site cannot be used to describe all the soil. In practice, information is usually wanted for areas, and surveys are made in many parts of the world to obtain such information. Fully quantitative information can be obtained by measuring the soil at several, perhaps many, sites.

There is another kind of data that may properly be regarded as quantitative. Surveys are often carried out to determine how much land is of a particular kind (say, suitable for growing rice) or what proportion possesses some attribute (say, waterlogged soil). We might attempt to delimit such land and measure its area by geometry. Alternatively and more economically we could inspect the soil at a number of suitably chosen places and *count* those where the soil possessed the attribute in question. Individual observations would be qualitative, but in total they would assume a quantitative character.

But the matter does not rest here, for we need to know to which of all the soil a particular measurement or set of measurements apply. To what extent may the value obtained at one site be extrapolated? And is it sensible to take an average of several measurements, especially when there are large regional differences? It is often more meaningful to use averages to describe the soil of each region separately; and, of course, the recognition of regions means classification – dividing the area into classes. Similarly prediction can be more secure if it is restricted to individual regions and classes of soil. Further, when measurement is expensive or time consuming it can pay to classify the soil on easily observed characters and sample each class separately for those that are costly to observe. By doing so, unnecessary measurement in large reasonably homogeneous areas can be avoided while at the same time adequate attention is given to smaller but different areas. When a large body of data is collected in the course of an investigation it often needs simplifying to be intelligible: we must be able to see the wood rather than the trees. This can often be achieved by classifying the data. Similarly, when a survey is carried out for planning land-use the sampling sites, though initially described quantitatively, usually need to be grouped. The farmer cannot vary his management continuously in response to continuous variation in the survey records. He has to treat finite tracts of land in a uniform manner as though they were homogeneous. Likewise the engineer cannot easily change his design for every minor fluctuation in soil character.

So although we replace classification by measurements for consistent

description and communicating precise information about soil at particular sites, classification can have an important complementary role in increasing the utility of soil data and enabling us to economize on sampling. Just how important this role is or can be has been a matter of debate, and in recent years classification itself has been the subject of quantitative study. Questions like: how much does my classification improve prediction; how might this population be classified to provide a simple but useful picture; which classification is best for that purpose; to which class should the new individual *I* be allocated; and if classification seems unprofitable is there an alternative, can all be answered quantitatively. Questions such as these are discussed at length later in this book.

Statistical methods

Quantitative description of the soil of different areas and its behaviour involves analysis of more or less large bodies of data. Simple statements about the soil must be seen against the background of ever-present variation, which must be taken into account in the analysis. This is the province of statistics, and most of the techniques described in this book are in some sense statistical. Statistics provides accurate and usable mathematical descriptions of the real world of the soil, both in the laboratory and in the field.

For soil scientists with an agricultural background the subject of statistics often conjures up a picture of experiments carefully designed to compare crop varieties or the effects of fertilizer treatments on yield. The measurements from these are subject to analysis of variance, a mysterious process whose culmination and climax is a test of significance. What joy ensues when an *F* ratio emerges blessed with three stars to confirm our hypothesis, or prejudice! This is unfortunate. Significance tests have a perfectly proper place, but in soil survey they will usually be secondary, superfluous, or even inappropriate to the main purpose of the investigation. In soil survey and soil systematics statistical methods provide means of condensing data into economical descriptions of variable material. They enable us to predict and to estimate the confidence that we may place on prediction. They enable us to identify relationships and structure in our data and to display these. Thus some tasks are probabilistic while others are not. If a significance test is appropriate we should apply it since it could prevent our drawing unwarranted conclusions from sample evidence. But the precise level of probability at which we test is less important, and is to some extent a matter of personal choice.

Aims and means

Much of the first half of the book concerns soil survey. But the reader who has become accustomed to regard soil survey as the recognition of soil types and their display on maps might find himself on unfamiliar ground. He will discover that soil classification is not a necessary prerequisite to mapping nor mapping essential to the purpose of survey. Soil survey is presented simply as the means by which information about the soil of areas of land is obtained. Soil classification is very often a means by which survey can be carried out more economically or its findings presented. It is a convenient tool. Neither survey nor classification is an end in itself, and the investigator is urged to decide just what he wants to know before considering how to go about finding it out. Once an investigator is clear about his goals, and only then, he is in a position to choose the means for reaching them in the most direct way, by designing a survey specifically for the purpose. In some instances this will entail classification and mapping; in others these will be unnecessary or even unhelpful.

In this connection it has seemed to me that sampling is the weakest feature of current practice in many soil surveys and much field research. Much of the data that are obtained only with difficulty in laboratory and field are of little use because the original sampling was unsatisfactory. If the reader finds the emphasis on sampling obstrusive, I offer no apology.

The latter part of the book deals with multivariate methods. Although these can serve very specific purposes, they are also exploratory tools: for identifying relationships between soil profiles and classes of soil; for experimental classification; and for suggesting hypotheses that can be tested by other techniques. These methods have become practicable only with the advent of computers to perform the complex or very long calculations involved. Soil scientists have had little experience with the methods, and most studies reported to date have been made on familiar data collected for other purposes. This has been necessary to give confidence, but with that experience behind us, we are ready to apply the methods in unfamiliar situations.

In addition to these two main groups of topics I have included a chapter introducing computers and programming, and other reviewing quantitative mapping.

Scope

A book of this size cannot deal with all the mathematical techniques that a soil surveyor or systematist might need. I have had to make a

此为试读,需要完整PDF请访问: www.ertongbook.com

selection. In doing so I have tried to cover those topics that soil scientists most often ask about and that I have found useful. Although most are described in statistical texts, soil scientists find it difficult for one reason or another to link the statistical theory to their work. I have tried to provide that link here, especially with the examples. It seems that few students of soil science have studied mathematics since leaving school, and I have constantly borne this in mind in presenting the material. For this reason, and for compactness, I have omitted most derivations and proofs. But the reader who fights shy of symbols must discipline himself to mastering them.

I have assumed that the reader is thoroughly *au fait* with the physical or biological aspects of his work and the methods of measurement in the laboratory or field.

The reader who wishes to apply the methods in his own investigations will often need to consult one or other statistical table. Many standard statistical texts include tables and it has not seemed necessary to duplicate them here. The statistical tables compiled by Fisher and Yates, first published in 1938 and now in their 6th edition (Fisher and Yates 1963), are available to soil scientists in agricultural institutes and research laboratories all over the world. For this reason I have referred to them freely. However, Lindley and Miller (1953) include all the elementary tables, while the Biometrika tables (Pearson and Hartley 1966, 1972) are more comprehensive and provide more advanced tests.

The reader will notice that I have chosen most of the illustrative examples from my own experience. This is not because they are necessarily more suitable than the work of others, but simply that I already had the data and intermediate workings to hand.

A note on terminology and symbols

A number of common English words have a somewhat restricted or special meaning in statistics. Some of these occur frequently throughout this book, and are introduced here.

Attributes and variables. When a property of the soil (or anything else) is recorded qualitatively it constitutes an *attribute.* If, for example, we record calcium carbonate as either *present* or *absent* in the soil at a number of sites, calcium carbonate is an attribute of the soil at those sites. We may also use the term to refer to properties that can occur in more than two qualitative states – the shape of structural aggregates, for example, can be platy, prismatic, blocky, granular, and so on. On the other hand, when a property varies from one place to another and

is measured it is termed a *variable*. If, instead of recording calcium carbonate only as *present* or *absent*, we measure how much there is, calcium carbonate becomes a variable. It is convenient to be able to use the word *character* to embrace both attribute and variable. A random variable is termed a *variate*. When applied to soil, the term usually means a set of measured values of some property in which there is more or less random variation.

Parameter. A *parameter* is a quantity that is constant in the case being considered, though it may differ in other cases. In statistics it is generally reserved for quantities, such as means and standard deviations, of whole populations (q.v.), and is distinguished from estimates of them made by sampling, for which the word *statistic* is used. In computing, a parameter is usually a quantity that is held constant for a particular run of a program. Parameter is not synonymous with variable or character, and the current fashion for using the term with this sense is to be deprecated.

Population. The whole set of individuals or material under study in a particular instance is referred to as a *population*. A population can be either finite – for example, a set of described soil profiles – or infinite – for example, the soil of a district, which can be considered as made up of an infinite number of soil profiles.

Sample. A *sample* is a small set of individuals or a collection of material taken from a larger population about which information is wanted. In soil laboratories the term sample often refers to a single bag (disturbed) or core (undisturbed) of soil taken from the field. In the statistical sense it more often refers to several bags or cores, and applies equally to a set of sites where the soil has been or is to be described or measured without any soil necessarily being collected.

Survey. In the statistical sense a survey simply means a type of investigation in which a situation is observed as it is without alteration other than that unavoidably incurred in sampling. It differs from an *experiment*, in which the investigator changes some feature of the situation on purpose so that he can study effects of the change. The meaning of survey in statistics also differs somewhat from its meaning in soil survey, in which it has implications of mapping and classification.

Classification. This is the act of dividing a population, or agglomerating individuals, into groups. It can also be the set of classes that result from such action. The nouns *class* and *group* are treated as synonyms in this

book. *Allocation* is distinguished from classification and is used to mean the assignment of individuals to classes or the *identification* of individuals as belonging to those classes.

The common mathematical symbols will be familiar to the reader, but there are several other conventions that might need explaining.

Summation. Summation is one of the most frequent operations in statistics. If there are n values, $x_1, x_2, x_3, \ldots, x_n$, their sum can be written as $\sum_{i=1}^{n} x_i$, where

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n.$$

Mean. The arithmetic mean is also much needed, and is usually signified by placing a bar over the symbol for the variate concerned. Thus for the n values of variate X, the mean $\bar{x}$, which is read 'x bar' is

$$\bar{x} = \sum_{i=1}^{n} x_i/n.$$

Product. The product of a set of values is occasionally needed and is symbolized by capital Π. The product of n values of X are thus

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times x_3 \times \ldots \times x_n.$$

Combination. The number of ways of choosing r items out n is

$$\frac{n!}{r!(n-r)!}$$

where $n!$ (factorial n) is $1 \times 2 \times 3 \times \ldots \times n$. It is conventionally abbreviated to $\binom{n}{r}$.

Several lower case Greek letters are widely used: μ (mu), σ (sigma), and ρ (rho) indicate the mean, standard deviation, and correlation coefficient of populations respectively, while the sampling estimates of the last two are denoted by Roman equivalents s and r. The letter χ (chi), as χ^2 (chi square), is used for the distribution with this name. Capital lambda, Λ, is used for Wilks' criterion, while in the lower case it usually refers to latent roots.

Matrix notation. Many multivariate procedures are most economically and clearly expressed in matrix form, and this practice is followed in Chapters 7 to 11. A brief account of matrix algebra and symbolism is given in the Appendix.

2. Data handling and computing

Whatever task lies to your hand,
do it with all your might.
ECCLESIASTES 9:10

The investigator who embarks on any kind of quantitative survey of soil or study of soil classification will want to collect information, and having collected it, to sort it and perform calculations on it. At its simplest a survey might cover two or three soil properties at a few tens of sites only, and calculation involve no more than adding, counting, and finding means. Pencil and paper will suffice to arrange the information and to record the steps in the arithmetic. A small desk or pocket calculator might be a useful aid. More complex studies can involve measuring fifty or more soil properties at hundreds or thousands of sites and examining relations among them; and the same data might need to be screened, analysed, and presented several times in different ways. Here a computer is essential, and many valuable techniques of analysis and display have become practicable only with the advent of computers.

Tasks of intermediate size can often be carried out on a desk calculator provided it has at least one memory or register to hold sums of squares or products. However, it is tedious; and though undue repetition can be avoided by using a programmable calculator, an increasing proportion of work is being done on general-purpose computers as they become cheaper and more readily accessible to the soil scientist. The computer is fast becoming the general work horse for information processing, and we should expect it to be in everyday use by soil surveyors in the not too distant future. So this chapter will outline the workings of the modern computer and the measures a scientist must take in order to use it.

The computer

The modern computer is electronic and digital, in the sense that it works by discrete electronic impulses on information presented to it and stored by it in the form of discrete electrical states. Information from almost every field of study or walk of life can be handled in this way, and it is this that makes the computer so generally useful. There

are analog computers, but they are usually designed for specific purposes and we shall not consider them further.

A computer consists of a number of functional parts, known as *hardware*, whose relations to one another are symbolized in Fig. 2.1. The most important is perhaps the central processor, which contains an *arithmetic* or *logic unit*, and a *control unit*. Closely associated with the central processor is the *memory unit* or *high-speed store*. This consists of many small elements, each of which can be set in either of two states corresponding to the two values, 0 and 1, of a binary digit. Until recently the memory was built of ferrite rings or cores, each of which could be magnetized in either direction, and so it became known as the *core store* or just *core*, and this name has stuck. The elements of the core, each of which can be made to represent a binary digit, or *bit*, are linked into groups to form *words*. The number of bits per word varies from one machine to another, and can be as few as 12 or as many as 96. In general, the smaller machines have fewer bits per word than the larger ones. A small computer might have a store of only 4000 words, a large one over 100 000 words. In some machines bits are grouped in series of eight to form *bytes*, which can uniquely represent all the alphabetic and numeric characters, and special symbols for arithmetic and punctuation. The word is the most important combination of elements to the scientist, since each will hold one number of reasonable size or one command. Computer languages are written in such a way that each word may be identified and addressed individually.

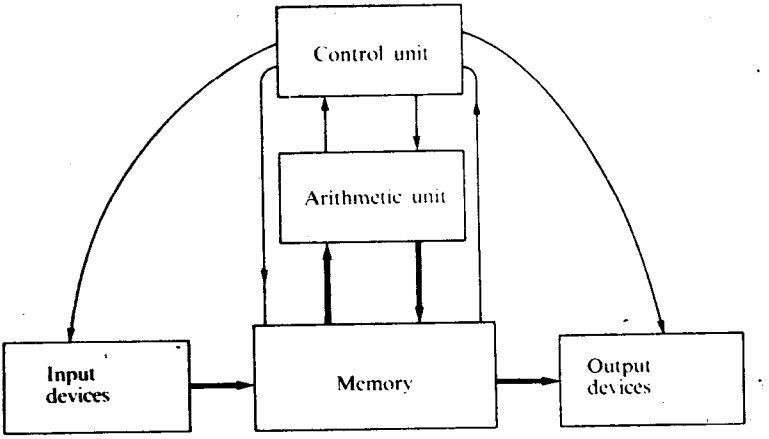

FIG. 2.1. The functional parts of a computer. Broad arrows indicate the transfer of the user's information. The thinner lines show how control is maintained.

Attached to the central processor are a number of peripheral devices that enable data, programs, and other instructions to be entered (*input*), results to be obtained from the machine (*output*), and provide auxiliary storage.

The arithmetic unit consists of a set of circuits, *logic circuitry*, that performs arithmetic and logical tests. The arithmetic units in some machines also have special circuits to determine functions such as square roots and logarithms. The purpose of the control unit is to control the operations within the central processor itself, as well as in the core store and peripheral units, by interpreting programmed instructions and ensuring that they are carried out in their proper sequence.

The machine has a console, usually consisting of an array of switches and a typewriter keyboard, from which the operator instructs the computer. Other input devices include paper-tape and card readers, magnetic tape and disc units, and remote terminals. The most familiar output is that produced on the line printer, which assembles and prints one line of print at a time. Each line may contain a maximum of 120 to 160 characters, and may be printed at the rate of 600 to 1000 lines per minute. Other output devices include paper-tape and card punches, graph plotters, cathode ray tubes, magnetic tape and disc, and remote terminals.

The amount of immediately accessible store, the core, is strictly limited even in the largest computers, so large quantities of data must be stored on an alternative and cheaper medium. Data, and also frequently-used programs, intermediate results, and information that is to be transmitted from one program to another, are usually held in an auxiliary store or *backing store*. Backing store can consist of magnetic tape, disc, or drum. Of these magnetic tape is the cheapest but also the slowest. Nevertheless all allow much faster transfer of information between them and the main store than could be achieved with cards or paper tape.

Files and keyboard terminals

In many modern computers jobs are not run directly from cards or paper tape. Instead, data and programs presented to the computer in this form are read and converted to *files*, which are then held on disc. The information in such files usually retains the same format as that on the cards or tape, and such files are said to have *card-image format*. Jobs are run from the disc.

A machine that operates in this fashion has one or more keyboard

此为试读,需要完整PDF请访问: www.ertongbook.com