WESTERN AUSTRALIAN HERBARIUM RESEARCH NOTES

No. 6, 1982: 173-180

CARTOGRAPHIC METHODOLOGY AND COMPUTER TECHNIQUES IN THE PRODUCTION OF SPECIES DISTRIBUTION MAPS

By H.R. Foston and K.R. Butson

Surveyor General's Division of the Department
of Lands and Surveys, Western Australia

ABSTRACT

This paper outlines the design of a computer-based system for storing information on the Australian geographic distribution of plant species and describes an instance of its use in the production of distribution maps for 729 Australian species of Acacia. *Included also is an outline of the cartographic design and production principles of these maps.*

INTRODUCTION

The task

The objective was to produce for each of the 729 species of *Acacia* (Maslin and Pedley, 1982a) a separate, single colour map of Australia, bearing 1^{o} x 1.5^{o} grid-cells, with species occurrences indicated by a number within the appropriate cells. Six maps were to be printed to each A4 size page and each map was to be titled as to species name and author. The grid system was to be a modified version of the Australia 1:250,000 map series, each cell being assigned a number according to the system of Brook (1977). The occurrence of a species within a cell was to be indicated by printing the Brook number (i.e. the Australian Biogeographical Integrated Grid System number, ABIGS) in that cell.

Selection of production method

It was decided to employ computer graphics technology rather than a conventional manual map production system because of future benefits likely to accrue from multiple use of the stored data. Such usages include a variety of graphical and other data analyses, as well as the potential use of the technique in similar studies in other plant (or animal) groups. The computerizing of distribution information merely for the production of a "one off" mapping exercise, as an alternative to a manual process, is not cost effective.

Development of the project

The project was developed in five stages: cartographic design and specifications; design and implementation of the computer system; preparation of the base maps; digital capture of species distribution data; and grid-cell numbering by automated plotter.

CARTOGRAPHIC DESIGN AND SPECIFICATIONS

(a) The Brook numbering system

This uses a 3-digit number in grid-cells based on the 1:250,000 National Mapping series sheets, an identifying initial digit being used for each State (see Figure 1 in Maslin and Pedley, 1982a). The horizontal rows of cells in each State are numbered sequentially and in most cases this sequence is maintained onto the next row. However, as the numbering system is not arranged as a logical co-ordinated matrix, it is not immediately adaptable to a digital co-ordinate system (see below). Therefore, the design had to incorporate into the map-sheet construction, a co-ordinate referencing system for each cell to which the appropriate Brook numbers could be referred.

(b) Production scale

This was governed by the maximum size of sheet able to be accepted by the automated plotter and also by the minimum type and pen size of the plotter. These constraints necessitated a minimum grid-cell size of 4.5 x 3 mm to accommodate the 3 digit Brook numbers and represented a 3:2 enlargement of the final production scale. Each quarter imposition sheet (see below) was approximately 700 mm (X) by 1000 mm (Y).

(c) Construction of the imposition sheets

An imposition sheet consists of 4 subsheets called quarter impositions; each quarter imposition consists of 4 pages and each page bears 6 maps (Figure 1). Thus to cover the 729 Australian species of Acacia 8 impositions or 32 quarter impositions or 128 pages were required. The construction of a four page quarter imposition was the same for all 32 sheets involved and was designed as a homogeneous map sheet network forming a single grid with all relevant data referenced to a primary origin. This arrangement required that two of the four pages be inverted i.e. pairs of pages head to head.

The projection employed was pseudo-cylindrical, displaying the meridians and parallels as straight lines, forming a rectangular grid system to which co-ordinate referencing was applicable.

Points such as page corner, map corner and grid-cell corner were referenced from the primary origin as co-ordinates expressed in multiples of the grid-cell dimensions, i.e. 4.5 (X) by 3 mm (Y), and converted to distances. For example the SW corner of the SW map of page 30 in Figure 1 would be referenced by 8 Columns (X) and 18 Rows (Y) thereby producing plotting scale co-ordinates 36 mm (X) and 54 mm (Y). Similarly in the case of the inverted pages, i.e. pages 22 and 27 in Figure 1, a secondary origin was established enabling the relative co-ordinates to be rotated through 180° (X and Y then became negated).

(d) Accuracy

The method of co-ordinate referencing by columns and rows created the probability of cumulative error over the length of the sheet. To minimise this factor, and because the plotter operated on imperial measurement, the dimensions of a grid cell were converted from metric units to 6 decimal places of an inch.

Precise positioning of the Brook number in each cell required the consideration of height, width and spacing of digits and the specification of the co-ordinate reference for the first digit to the nearest 0.001 of an inch.

(e) Page-to-Imposition Arrangement

The pages in imposition form are not placed in numerical order: the correct sequence is obtained when the flat sheets, printed on both sides, are folded in a standard procedure to book page size (Figure 2 - "printing logic"). To facilitate the programming of the correct species data to its appropriate map on a specific page, each of the 729 species names was sequentially numbered on an alphabetical listing, and assigned its relevant page and imposition number. Provision was made for introductory text to occupy the first 5 pages.

An algorithm was later deduced to automate the selection of correct page content to imposition number (i.e. the "printing logic").

METHODS

Two basic outputs were required from this computing system: development of the necessary computer files to accomplish the computerized plotting of Brook numbers for the 729 *Acacia* species distributions; and development of a separate computer file containing the occurrences of each species, i.e. the Brook numbers pertinent to each species.

(a) The Necessary Computer Files

(i) The Brook File. The Brook numbering system for grid cells $1^{\circ} \times 1.5^{\circ}$ is declared and not mathematically derived from a co-ordinate referencing system, i.e. it does not have a series of horizontal and vertical references from which the Brook number is derived. The absence of such a mathematical numbering system necessitated the creation of a separate computer file where all numbers were referenced to a common base, namely row and column values related to a primary grid origin (see Figure 1). This common base or co-ordinate system performed the dual purpose of: providing the position of each Brook number; and providing a link or key into the Occurrences File (see below).

Therefore this file consisted of records indexed by cell position, i.e. row and column number. Each cell position had its corresponding Brook number.

(ii) The Occurrences File. This File stored distribution data for each species derived from the digitizing of cell centroids (the visual centres of the grid cells). These digitized centroids were converted from local co-ordinates (millimetres) into positional co-ordinates of "row" and "column" numbers, i.e. so many cells up and so many across. Thus each record on the Occurrences File, indexed on species number, i.e. the number according to the species alphabetical position, had its occurrences represented as "row" and "column" numbers. Therefore the Occurrences File contained no Brook numbers but rather co-ordinates linking into the Brook File. Consequently if the Brook numbering system were altered, e.g. to allow extension to include islands, the Brook file would be simply updated and the Occurrences File not affected at all. In fact the Occurrences File has been designed to be totally independent of the numbering system and of the names of species (the Names File). The row and column co-ordinates link into the Brook File whilst the species number links into the Names File (see below).

(iii) The Names File. The species names and corresponding authors were encoded onto computer disk and sorted into alphabetical order. The Names File was then created with each record indexed by species number, i.e. its alphabetical position. Each species number had its corresponding species name and author.

(b) Plotting Method

The plotting program involved the linking of the main Occurrences File with its subsidiary dual links: the Brook File and the Names File. This linking was done in conjunction with an algorithm for selecting the appropriate species for the appropriate page and selection of the appropriate pages for the appropriate imposition (see above). This algorithm could be referred to as the "printing logic".

The recurring cycle of events for each species is as follows:

(i) A record (species number) was extracted from the Occurrences File. The selection of the record is governed by the printing logic.

(ii) Each occurrence (row and column) was then extracted from that record. The computer then drove the plotter to each of these occurrence positions (via row and column) and almost simultaneously fetched the appropriate Brook number from the Brook File (again via row and column) which was then plotted.

(iii) For each record extracted from the Occurrences File the computer also fetched the corresponding name and author from the Names File and plotted them (name in italics and author in Roman).

The continual traversing of all three files necessitated these files being of random access type.

The plot tape referred to in Figure 2 consists of a full imposition, being 4 plots, one for each quarter imposition. Each quarter imposition plot was designed to maximise the dimensional capacity of the plotter (see above).

(c) Computerised Acacia File

This file was designed for future computer analysis and each record was to contain the species name as the index and the Brook numbers for that name as the rest of the record. The file was to be in alphabetical order. The method of production was again a matter of linking the main Occurrences File with its subsidiary dual links: the Brook File and the Names File. However the linking in this case was not done in conjunction with the printing logic but simply in sequential (alphabetical) order. This file was subsequently processed (actually "inverted") in order to produce the species lists contained in the second paper in this series (Maslin and Pedley, 1982b); for each record in those lists the Brook number became the index and the occurring species names became the rest of the record.

It is anticipated that further computer analyses will use this Acacia File and that the file will be kept current as new distribution records come to hand.

(d) Data Revision

The receipt of additional data necessitated the provision of an updating program for the three main files and a display program for verification of the added data.

As a consequence the files that were created have the low cost capabilities of immediate data revision at any future date.

For example, an average batch of transactions for revision, requiring 10 seconds of computer time, is estimated to cost $1.00 and a subsequent computer listing, summarizing the new contents of the Occurrences File, $3.50.

(e) Preparation of the Base Maps

The 729 species maps involved the preparation of 32 x 4 page quarter imposition sheets (6 maps per page) on transparent stable based drafting film for the addition of running title, species names, authors, page numbers and Brook numbers by automated plotter. A master sheet was initially prepared to conform to the cartographic design specifications referred to above. From this master the final 32 copies were produced by photography. The master was produced firstly in two components: (1) the coastline and State borders, and (2) the map surround and grid network.

The coastline and State borders were photographically processed as a 40% screened image prior to combining with the surround and grid. The purpose of this was to reduce the intensity of the coastline and borders to ensure the legibility of any Brook numbers overprinting these lines.

(f) Digital Capture of Species Distribution Data

Rough working maps, used to score species distribution data (one map to each species), were the actual documents that were digitized.

Being paper copies, the working maps had pronounced and varying scale distortions. To ensure that the digitizing referenced the correct cells, each working map was related to a true-to-scale master and the data digitized from the master copy.

Digitizing was carried out using firstly a Gradicon Digitizer linked to a card punch, and later, owing to equipment breakdown, completed on a Summagraphics Digitizer linked to a CYBER disk. The technique involved recording the centre point of those cells in which a species occurred. These centroid points were automatically converted to "row" and "column" co-ordinate values relative to the primary origin and comprised the contents of the Occurrences File (see above). The digitizing of the distributions for 729 *Acacia* species, representing the recording of approximately 10,000 points, was completed in 6 man days.

(g) Grid Cell Numbering by Automated Plotter

The Plotting Program treated each imposition, i.e. 4 plots of 24 maps each, as a separate plot. The production of 8 impositions or 32 quarter impositions necessitated the computer generation of 8 plot tapes to plot the 729 species occurrences on a Xynetics 1101A flat bed plotter.

The plotter was off-line to the computer and was thus driven by the plot tape via a tape drive.

Unfortunately some problems in image density were experienced during the running of the plot tapes. The very small pen (0.1 mm diam.) used for the Brook numbers demanded a precise ink mixture and also required a reduction of plotter speed to ensure a satisfactory density.

All base sheets were subjected to thorough preparation to ensure a grease free surface. The base sheets were laid on the flat bed, registered with the plotter and held in position by a vacuum during the numbering process. The plot tapes for each imposition were loaded into the plotter tape drive and the plotting controlled from the operators console. The small inherent scale differences, always experienced in reproduced images, and the fixed scale of the plot tape created some minor problems of adjustment in the registration process.

CONCLUSION

The present project illustrates how recent developments in computer-assisted cartography can aid the processing of biological distribution data so as to produce species maps and lists, with a high degree of automation. Though hardly cost-effective in a single application, the system for *Acacia*, described above, has the potential to be extended to other groups of plants or animals, covering a wide range of precision and scale. Moreover, the same computerized data base could be made the subject of an extended series of analyses for many purposes.

Provided software is designed for maximum applicability, as in the present instance, its cost would be amortized over a period, according to the number of users. However, amortization of a system for multi-user costing must preclude the software remaining the property of the developing organization.

REFERENCES

Brook, A.J. (1977). Australian Biogeographical Integrated Grid System. ABIGS Technical Manual. (Dept. of Zoology: University of Melbourne.)

Maslin, B.R. and Pedley, L. (1982a). The distribution of *Acacia* (Leguminosae: Mimosoideae) in Australia. Part 1. Species distribution maps. W. Austral. Herb. Res. Notes No. 6: 1-127.

Maslin, B.R. and Pedley, L. (1982b). The distribution of *Acacia* (Leguminosae: Mimosoideae) in Australia. Part 2. Lists of species occurring in 1^{o} x 1.5^{o} grid cells. W. Austral. Herb. Res. Notes No. 6: 129-171.

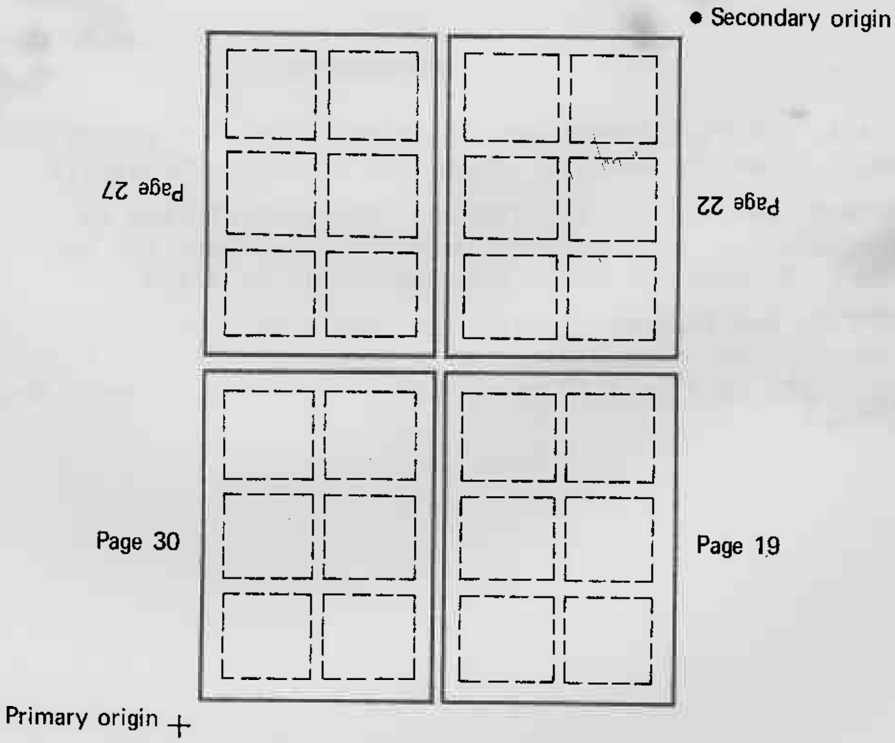

Fig. 1. Example of a 4 page quarter imposition and the non-sequential page layout.

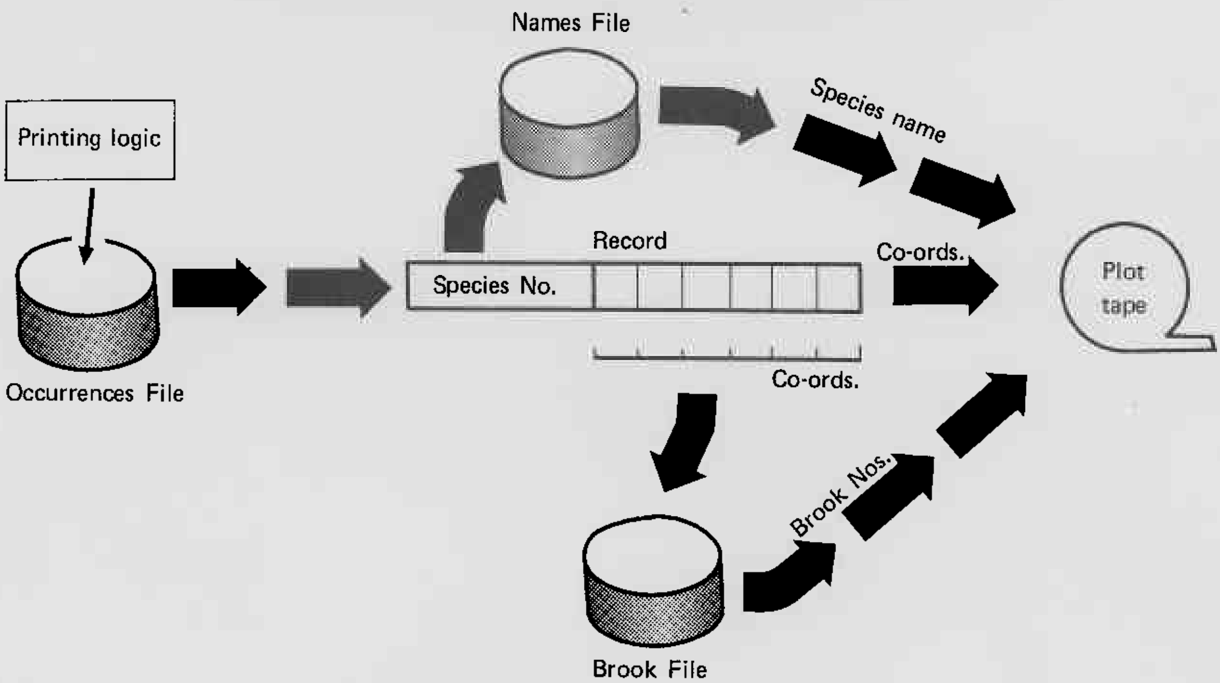

Fig. 2. Computerized plotting - the recurring cycle of events for each species.

16918/1/82—1 200—L/3166 WILLIAM C. BROWN, Government Printer, Western Australia