6 DATA INTEGRATION AND ANALYSIS

C.J. Pearce, W.J. Boardman, M.A. Green and P.J. Bowen

6.1 Geographic Information System

Every item of data collected in the jarrah inventory, whether from the aerial photographs or from ground-based measurements, is tied to a geographical location. The locations of the measured plots, provided to an accuracy of 20 to 30 m by the GPS, form a more or less uniform rectangular grid of points over the area covered by the inventory. They therefore form the basis of a quasi-continuous map of the resource parameters measured. Such data is handled, and related to other spatial data (in particular other maps), by a geographic information system (GIS).

At the commencement of the jarrah forest inventory, the Department already possessed a pre-existing GIS, or rather the software and hardware tools which go to make up such a system. It also had a large database of thematic map data in digital form, covering the area of the inventory, as well as some high precision digital maps related to tenure and the forest management plan. The task required for the jarrah inventory was to adapt the existing GIS tools to meet the data storage, processing and reporting requirements of that project.

The existing GIS facilities consisted of three main components: Intergraph, a sophisticated proprietary computer-aided mapping and graphics facility; ARC/INFO which is another proprietary system for computer-assisted mapping and map analysis; and FMIS, a forest management information system developed in-house. The Intergraph is used for high precision map input, editing and storage, and for derivative map production. ARC/INFO is intended for spatial analysis (such as buffer generation, or polygon intersection) and some map product generation, and FMIS is a grid-based map overlay system for forest resource inventory and area statements. For the jarrah inventory project, all three components are required, in a way which links them together more closely and systematically than has been necessary in the past.

In the following subsections we describe the requirements imposed on the GIS by the jarrah forest inventory and the overall system design adopted to meet these requirements.

6.1.1 System requirements

It is convenient to discuss the requirements under three main headings: data storage, data analysis and processing, and output or reporting. The essential requirement for data storage is that all the measurement data, with its associated geographical location, is linked to a graphical representation of that location. That is, the three types of data collected - navigational data, photo-measurement data and ground-measurement data - must be stored in a database which links to a graphics design file showing the locations of the photo plots. This is necessary in order for the measurement data to be selected and accessed in relation to other, external, map data or map representations. The system must also be able to store such external reference maps, which themselves may be linked to other databases. Examples of such maps include topographic map and cadastral boundaries, as well as numerous thematic maps relating to forest type, treatment, operations, and Departmental administrative boundaries. Much of this information already exists in the Intergraph and FMIS data bases.

The processing requirements, however, are more demanding and less easy to predict in the long term.

Firstly, the system must have full data editing, updating and archiving facilities, and must provide a secure linkage between the three different types of stored measurement. It must also have the means of transferring graphical data (the 'reference maps' referred to above) between different components of the system (such as FMIS and ARC/INFO) and of combining any graphical data in one component of the system with graphical data provided from some other component.

It must be able to carry out polygon overlay between different reference maps, based on the database or 'attribute' values associated with the polygons (such as the actual year of logging associated with logging history polygons). There must be the means for generating 'buffers' or 'zones', either along linear features or about points, so that the buffers form new maps which may be overlaid with the measurement data. It must also have digital terrain modelling capabilities, so that conditions on slope or elevation, for example, may be imposed when interrogating the data. The central processing requirement, however, is to be able to select all measurement plots within a window defined by a reference map (generated in any way required), and to extract and make available all data from the database applying to those plots. This capability is necessary in order for the resource to be estimated within any pre-defined area.

Finally, it is essential that the data stored in the measurement databases should be accessible (following selection by plot identifier) to FORTRAN programs for analysis and statistical reporting.

The reporting requirements lie in the areas of both graphical output and statistical or tabular statements.

First there are maps required, showing the principal points of photos against a topographic background, at scales of 1:250 000 (for planning purposes) and 1:25 000 (for use in navigation by ground-measurement crews). When interrogation takes place, a map showing the window of interrogation, again with a suitable topographic background for reference, is needed. Following more elaborate interrogations, it may be necessary to generate a high quality map showing a succession of interrogation windows, together with some kind of shadowing, colouring or symbols to indicate resource levels. A perspective view of a three-dimensional representation of the resource level or 'elevation' may also be desirable.

The system must be capable of outputting tabular resource information in virtually any form which may be requested. Since the processing requirements dictate that the measurement data should be accessible to FORTRAN programs, there should be no restriction on the type of statistical summaries which may be compiled by direct extraction from the database. In particular, the system must be capable of generating input files for a statistical analysis package. Another form of output would be an FMIS resource statement, made possible by creating an FMIS 'theme' (or map) of resource density through the use of the digital elevation model facility.

6.1.2 System design

The design adopted for handling the processing of the jarrah forest inventory data consists basically of a set of GIS modules (some pre-existing) and the interfaces between them.

There are five main modules, plus a number of smaller ones intended mainly for data validation, housekeeping and so on.

The central module, in which much of the geographical processing takes place, is ARC/INFO, shortly to be supplanted by ARC/SQL. This system is used for all data storage (in the ORACLE database) and data management. It also contains the graphical representation of the principal points (in an 'ARC coverage'). The generation of the final form of the geographical window of interrogation also takes place in ARC/SQL, as is the intersection between the array of principal points and the window polygon or polygons, in order to generate the set of photo plots within the window.

The generation of the window of interrogation may also involve two other modules: Intergraph and FMIS.

The role of FMIS is to provide that component of the interrogation window which depends on forest thematic maps such as forest type, cutting history, administrative areas and so on.

Intergraph is used to store and supply topographic features, such as roads, or towns, which may be used to define buffers of exclusion or zones for inclusion. In addition, the Intergraph can provide tenure and management-plan map information. The Intergraph is also employed to produce the principal point maps following the aerial photography, and may be involved in the final production of resource maps for management.

The next major module is the product allocation program. This takes a list of products, in a priority order, and allocates those products to the ground-measured plots within the interrogation window. It also reads the relevant photo-measurement data, and generates a composite data file for input to the statistical program, which is the final main module. This process is described in detail in Sections 6.2.2 to 6.2.5.

In addition to the above, special software has been developed to provide efficient and easily operated interfaces between the main modules, so that data may be passed rapidly between them with minimum intervention.

6.2 Processing Geographic Information System and resources data

6.2.1 Processing Geographic Information System data

Once the data are in the system, and have been validated, the processing sequence follows a downward path through the flow diagram in Figure 39.

The first stage is the definition of the area to be interrogated for resource - the window. The specification of this area will in general involve themes from FMIS, topographic information from Intergraph, and buffer generation within ARC/INFO. First, a multiple map overlay is carried out in FMIS to generate a composite map which represents the area satisfied by a subset of the constraints defining the interrogation window. The composite is then transferred to ARC/INFO and a coverage created. Similarly, the appropriate topographic and cultural features (such as roads, rivers, place names or sawmill location) are transferred from Intergraph to ARC/INFO and the required buffers, or zones, created. A polygon overlay is then carried out to intersect any such buffers with the composite map from FMIS to generate the final window for interrogation.

The second stage is to select the sample plots which lie inside the interrogation window. This is achieved by using the point-on-polygon operation in ARC/INFO, and then generating a list of the identifiers of those plots which lie within the window.
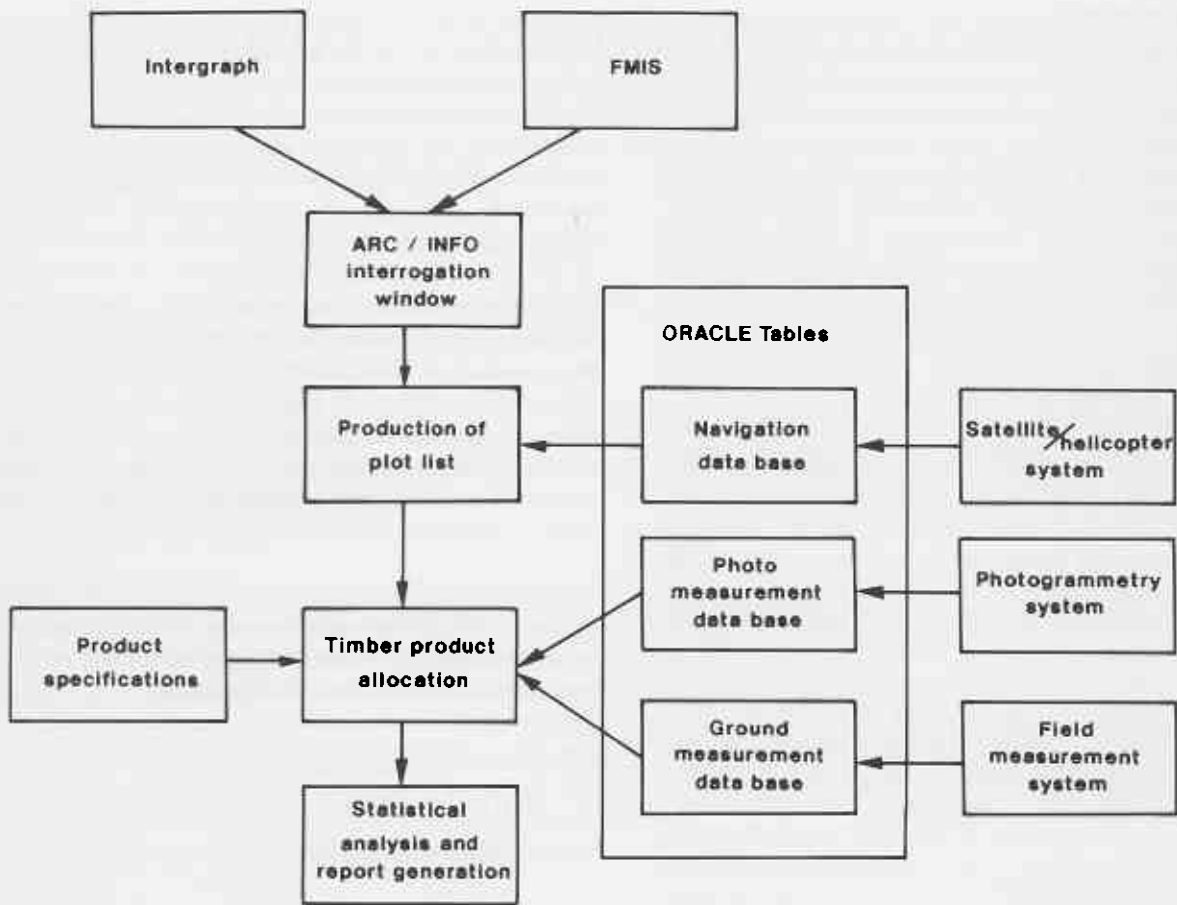
Figure 39
Flow diagram for processing Geographic Information Systems data.

The third stage is to extract and process the measurement data within the window. This is carried out by a program described in Section 6.2.3, which takes the list of plot identifiers, and uses it to read from the ORACLE database all the photo measurements and ground measurements of the plots within the interrogation window.

The program searches for and allocates the required products based on the ground measurement data, and constructs an input file for the statistics module. The latter combines the measurement data and generates the summarized resource statement which is required.

The final processing stage is the construction of a map, showing the window of interrogation, tabulation of resource level information, and topographic background. This currently takes place on the Intergraph, but may in future be done with ARC/INFO.

6.2.2 Processing photo-measurement data

Photo data are collected either directly from the Adam Technology MPS-2 analytical stereodigitizers or via a data entry program running on a personal computer from the parallax measurements on the stereotopes or interpretoskop. These systems include initial validation of the data, which is followed later by more extensive cross-checking.

The parameters recorded for each plot are the plot identifier (which is linked to the location coordinates), photo scale, plot diameter, plot type (full plot or half plot on a forest boundary) and land cover (e.g. native forest, exotic species, water body, mine). Individual tree parameters include the tree species and its silvicultural class (regrowth or old-growth) as well as the tree height. Plot data are checked against navigational data and cadastral boundaries to ensure that the coverage of the photo samples is correct and that there are no duplications or omissions.

An empirical formula, obtained by linear regression (Section 2.4.2), is applied to each tree height to determine gross bole volume. A correction factor is applied to the tree height estimate if it is known that the interpreter's measurements are significantly biased. The volume estimates are then added to the ORACLE database. When ground data are later added to the database, a ground-measurement flag is set to indicate the double sample plots.

6.2.3 Processing ground plot data

The ground plot data contain all of the information describing the wood qualities in each sample tree. The data for each tree are assessed by computer programs to derive the estimates of volume for each of the various product types. The product specifications are not fixed, but can be varied for each run of the program, and the order in which the products are computed can also be changed to match changing priorities in the market.

In the computer assessment, each tree is assessed in turn for the products within it. First the tree is tested to see if it is potentially suitable based on species, diameter at breast height, bole length or condition. The most valuable product is looked for first (based on the user's priorities) and assigned wherever possible. Then the next most valuable product is sought and assigned, if possible, where no other product has already been assigned. This process continues until the tree is completely assigned or until all desired products have been considered, in which case any remainder will be assigned as waste.

To reduce the computational load, each product is defined in terms of acceptable, conditionally-acceptable or unacceptable wood quality types. Most computation time can therefore be devoted to conditionally-acceptable quality types. For these types, a large quantity of defect would disqualify the corresponding section of log from consideration. For example, an acceptable quality is a burl, a conditional quality is rot, and an unacceptable quality is attack by pinhole borers(*Atractocerus kreuslerae*).

For a conditional quality measured as a cross-sectional area, whose quantity is not too large to be unacceptable on its own, there are two considerations to be made when determining acceptability. First, the quality (defect) could occur in a quadrant different from other defects below (Fig. 40); as such it would effectively add to these other defects (qualities) when assessing the suitability of the entire length being considered for assignment to the product. Second, it could be coincident so that there is no increase in cross-section resulting.

If the sum of the various areas exceeds a permitted maximum (such as 50 per cent cross-sectional area) it would disqualify the section from consideration. Where a section is unacceptable, then, if there is sufficient length below, the product would be assigned to that length.

After failure owing to too large a quantity by itself in one section, or through non-coincidence with other qualities, checking would begin again above the failed section. In the case of defect in non-coincident quadrants, the lowest section being considered would be omitted in case it is out of alignment with those above it. This would be done by adjusting the lower height to be at the top of the omitted section. Then the checking process would start over again (Fig. 41).

Some conditional qualities are more difficult to accommodate than others. In particular, these include sweep and bend. In the case of bend there can be an upper limit to acceptability. Where the maximum is reached and, if there is sufficient length below, then that length can be assigned and checking can recommence above the bend.
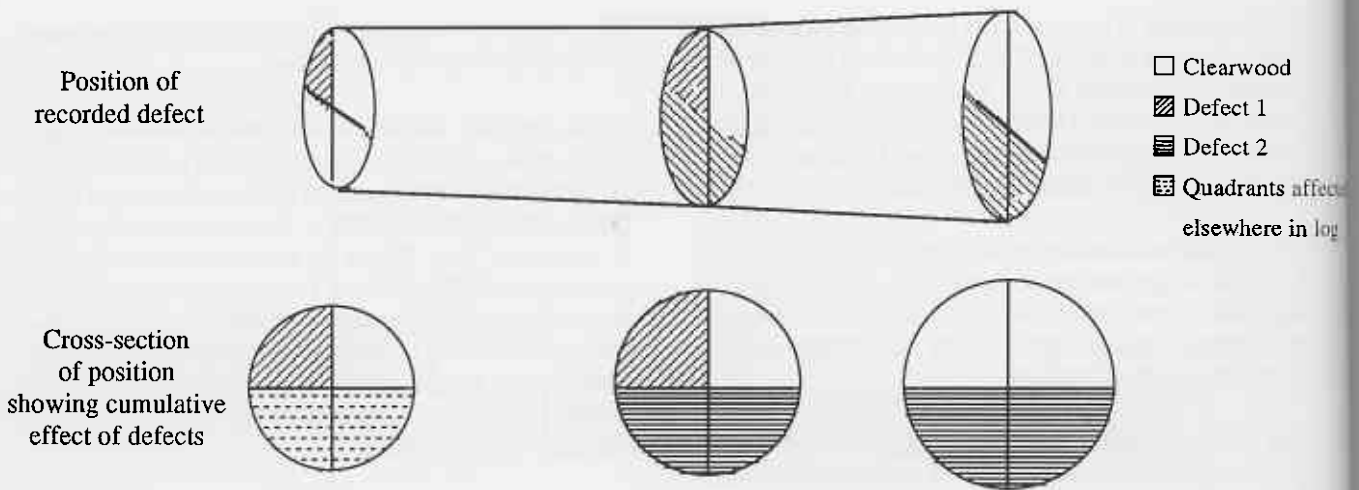
Position of
recorded defect

☐ Clearwood
▨ Defect 1
▤ Defect 2
▦ Quadrants affec...
elsewhere in lo...

Cross-section
of position
showing cumulative
effect of defects

Figure 40
Combined effect of defects in different sections of the same log.

A

Position of
recorded defect

4   3   2   1

B

Cross-section

Combined
defects in log B

Combined
defects in log A

Combined defects
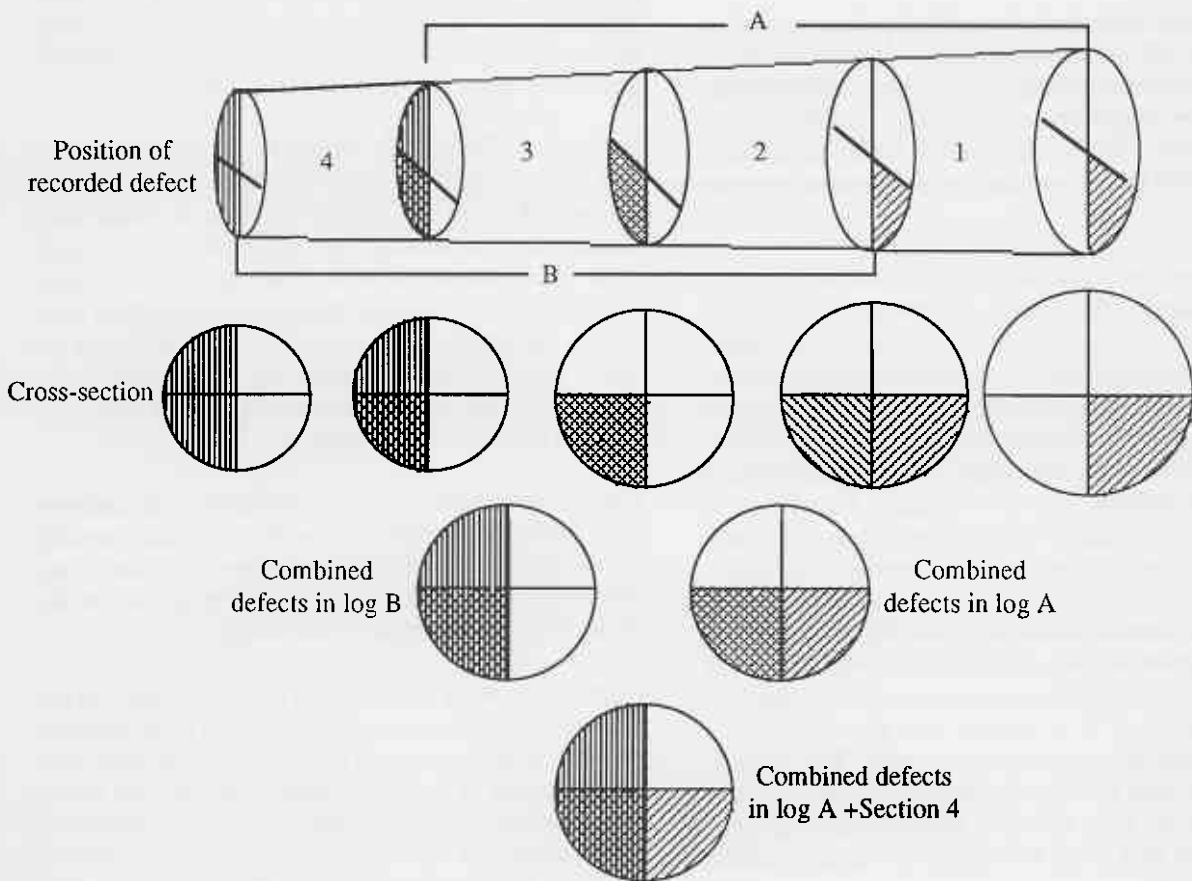in log A +Section 4

Figure 41
Combining acceptible sections to optimize product allocation. If a maximum cross-sectional area of 50 per cent is allowed, then section 4 is unacceptable to Log A since the combined area of defect would be 75 per cent. If Log A is not long enough then Log B (and above) will be considered and section 1 demoted. If log A is long enough then it will be assigned and checking will begin from the beginning of section 4, unless section 4 is unacceptable for reasons other than non-alignment of defects - in which case processing would recommence from the end of section 4.

48

Where there is insufficient length below and the angle of the bend is not too great, the product can extend into the bend until the minimum cross-sectional area is reached, including lower defects. The direction of the bend in relation to the position of other conditional qualities can also be a factor.

In the case of sweep it may be possible to have some shorter lengths of a product within the sweep section. Once again it is necessary to combine the effect of the sweep with other defects in determining the length available for consideration. Sometimes products can be assigned with just a part within the sweep. There is no requirement to assign a product below the sweep so, unless there is a length restriction, the greatest length possible is the desired goal.

Each ground plot tree is assigned a stratum value consistent with the same rules used to assign trees to strata in photo-plot measurement. The strata are based on plot size differences and variations in volume calculation for the various species and for old-growth (which differ in shape). All product volume calculations for the tree are then assigned this stratum value. Since fewer strata are required than are available after interpretation of photographs, rules are applied to photo-plot stratum values to coalesce relevant strata into the same strata as the ground tree strata. The product volumes are accumulated into a total for the tree. By including the *waste* sections, this provides an estimate of gross bole volume.

6.2.4   Volume estimations

For the ground plot data, log volumes are estimated using the integral of a formula used to estimate the diameter 'D(h)' at a height 'h' given the diameter at breast height (d.b.h.), the diameter at crown break (d.c.) and the crown break height. The formula was originally derived for karri but it is accurate for other species as long as the diameter at crown break is measured, and not estimated by a taper function.

To calculate the volume V_{12} of a log within a karri bole of crown diameter D_C (cm), d.b.h.o.b. D_B (cm), Height to Crown H_C (m) from height H_1 (m), up to height H_2 (m) :

$$d.o.b. (H) = D_C + \chi . (\theta . \tau + \beta) \text{ in cm}$$

where H is such that $0 < H = \text{or} < H_C$ in m

$$\theta = A_0 + 0.01 (A_1 . H_C - A_2 . D_B)$$

where $A_0 \approx 0.811$, $A_1 \approx 0.434$, $A_2 \approx 0.25$ (from Karri Volume Table)

$$\tau = (H_C - H)/(H_C - 1.3)$$

$$\beta = (1 - \theta) e^{-\zeta(H - 1.3)}$$

$$\zeta = \lambda / D_B$$

where $\lambda = 75.0747$

$$\rightarrow d.o.b.(H) = a - bH + c\,e^{-\zeta H} \text{ in cm}$$

where
$$a = D_C + \chi . \Phi . H_C$$
$$b = \chi . \Phi .$$
$$c = \chi . \psi$$

and $\chi = D_B - D_C$; $\Phi = \theta/(H_C - 1.3)$ and $\psi = (1-\theta) e^{\zeta(1.3)}$

$$d.u.b.(H) = (\rho) . d.o.b.(H) \quad \text{in cm}$$

$$= a_1 - b_1 H + c_1 e^{-\zeta H}$$

$$V_{12} = R \left\{ \begin{array}{l} a^2 (H_2 - H_1) - (ab)\left(H_2^2 - H_1^2\right) \\[2mm] + \left(\dfrac{b^2}{3}\right)\left(H_2^3 - H_1^3\right) \\[2mm] + \left(\dfrac{2bc}{\zeta^2} - \dfrac{2ac}{\zeta}\right)\left(e^{-\zeta H_2} - e^{-\zeta H_1}\right) \\[2mm] + \left(\dfrac{2bc}{\zeta}\right)\left(H_2 e^{-\zeta H_2} - H_1 e^{-\zeta H_1}\right) \\[2mm] - \left(\dfrac{c^2}{2\zeta}\right)\left(e^{-2\zeta H_2} - e^{-2\zeta H_1}\right) \end{array} \right\} \text{ in m}^3.$$

where $R = (\pi/40000) \rho^2$

At the butt end, standard stump heights are used. These are 0.2 m for regrowth trees and 0.4 m for old-growth. Thus any volume calculation below breast height (1.3 m) would also be reasonably accurate since there is only a small length for the estimated diameter to diverge from the true value. The relationship is less accurate in the butt swell region.

There are some precautions, however, which must be made for trees of very small bole. Where the bole length is 1.3 m or less, then the formula for a cylinder is used with diameter equal to the diameter at crown break. The volume of a cylinder is also assumed in the case of logs on the ground where the midpoint diameter and length are measured directly.

Where the bole is less than 2 m but more than 1.3 m and the diameter at breast height over bark is small at less than 300 mm, then a conic formula is used for volume calculation. The diameter over bark at crown break and diameter at breast height over bark are the parameters used along with the upper and lower heights of the log section.

49

Once the sorting process has determined the bottom and top heights of a section to be assigned to a product, these heights are used to calculate its volume. The volume is then added to that stratum for that species/product.

Volumes are then entered into a statistical system that was developed in-house to process the double sampling estimates and the outputs are generated using SAS programs on the Department's VAX SAS system.

Total volumes are estimated using the double sampling for ratio estimation (Cochran 1977). With this method, the mean volume per hectare of gross bole volume or product volume is defined by :

$$\overline{V}_T = \overline{v}_G / \overline{v}_P * \overline{V}_P$$

where $\overline{V}_T$ = total volume per ha (m³ha⁻¹)
$\overline{v}_G / \overline{v}_P$ = ratio of ground measured volume divided by the photo measured volume for the same photo plots.
$\overline{V}_P$ = the mean volume per hectare predicted from all photo plots (m³ha⁻¹).

Each tree on ground or photo plot measurements is categorized as old-growth or regrowth. This may be used to modify the volume function if the analysis indicates sufficient improvement in the volume model. Old-growth trees have reached their maximum height and continue increasing in diameter, but there is a large variation in the resulting volume. Regrowth on the other hand increases in diameter with height.

Variations in the relationship occur with forest quality. In poorer forest, bole height is a smaller proportion of total height than in good quality forest. Attempts were made to use *top height* of the plot to determine this, but investigations revealed that the sample size of the plot (20 m radius) was often insufficient to reflect a true top height of the site. Many areas of forest are also cut over and the resultant regrowth has not reached its potential height and too few veterans are present. Consequently, the forest areas have been stratified into quality classes and recorded in the GIS. The photo plot position information is used to determine which volume function is used for a particular plot.

6.2.5   Outputs

The jarrah inventory system has been designed as an interrogation type database, though routine resource summaries can be generated. Routine summaries can be made on the basis of current product specifications and values. The real value in the system, however, is the ability that it gives to re-run the data to test options. These options might include changing the amount of defect allowed when evaluating resources availability, or changing priorities to see the effects on supply and revenue.

Such interrogations are often area-specific, hence the need to link the resource data to the Department's GIS. Using this facility an area of interest can be specified by categories, such as management or land-use type, within a boundary or specified distance or any combination thereof (Fig. 42).

The ground and photo-plots within such an area are extracted from the database and run through the sorting routine and calculation process. Outputs consist of various degrees of detail from species/product combinations by strata to total volume summaries, with associated error estimates. The volume summaries can be supported by map products highlighting the areas involved.

Because the inventory has been designed to answer specific questions relating to the resources contained in defined areas under specified utilization criteria, logically the major outputs will be maps showing specified areas accompanied by tables of volume data and description of the utilization criteria.

One map, for example (Fig. 43), might show the boundaries of the *interrogation* area, plus all of the areas which have been excluded from the volume statement, such as private property, conservation areas, road, river and stream reserves, areas currently being logged, areas logged since measurement, and other areas excluded for various purposes, such as bauxite mining. Another overlay or map might show the distribution of samples in the area (Fig. 44).

Tables of volume estimates that might be required could show, for example, volumes of products within various species, silvicultural and condition classes with an accompanying key to explain the criteria used to define each of the products (Table 6).

Accompanying each table could be a statement of risk showing the probabilities associated with resource commitments at various levels (Fig. 45).

50

Figure 42
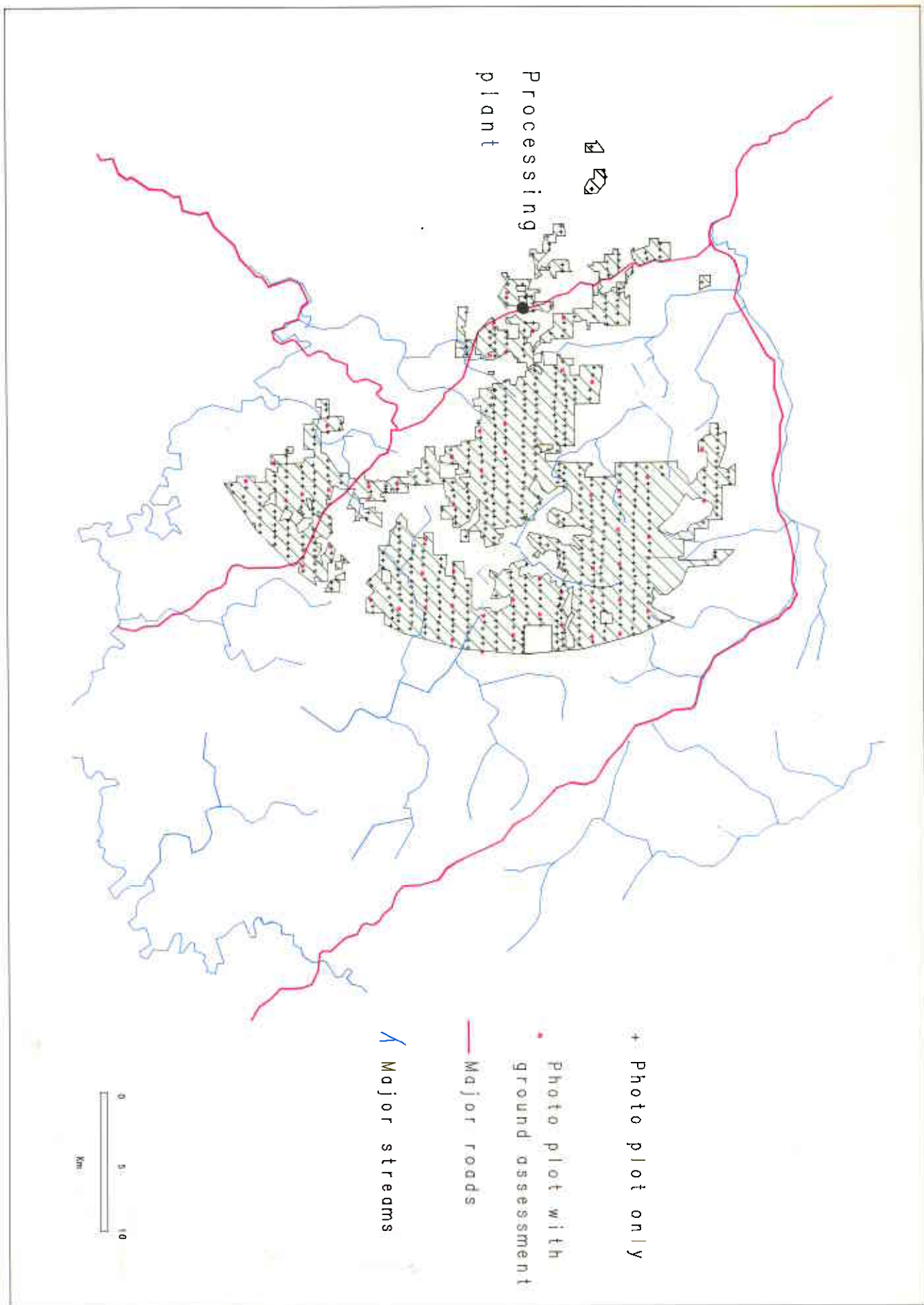Map showing State forest and sample points.

51

Figure 43
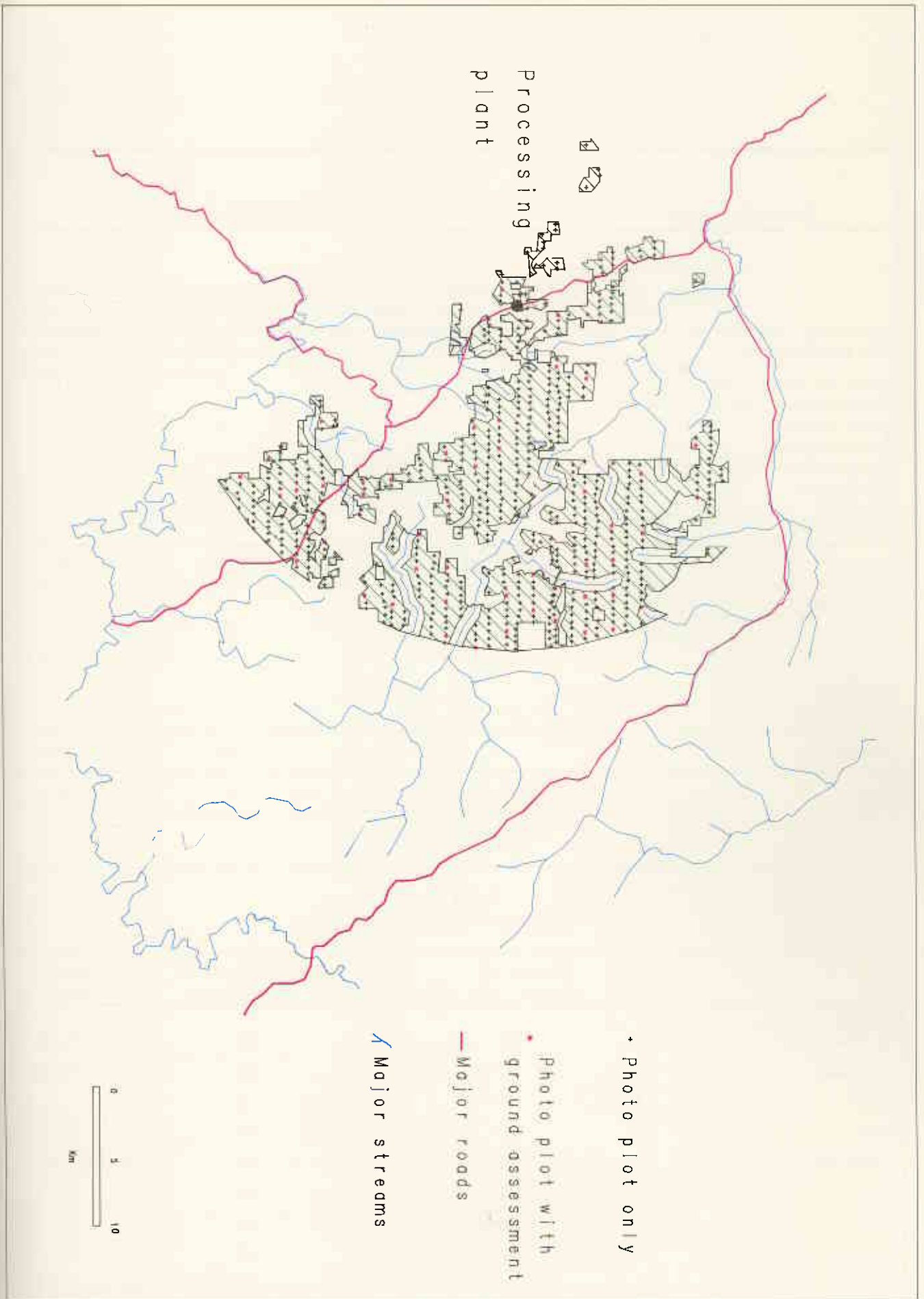Sample points and State forest within 20 km of a processing plant.

Figure 44
Sample points and State forest within 20 km of processing plant but not within stream buffers.

53

Table 6a
Volume output statement.
Hypothetical example of volume output per hectare by species, silvicultural class and product grade

	Sawlogs			Chip wood	Power Poles	Fence Timber	Industrial Firewood	Waste	Total
	1st Grade	2nd Grade	3rd Grade						
Live standing trees									
Jarrah old-growth	10	5	11				7	8	41
Jarrah regrowth	3	3	4		0.5	0.5	4	3	18
Jarrah crop trees	4.5	3.5	9		2		1	8	28
Marri old-growth			1	20	1			8	30
Marri regrowth				10	1			9	20
Wandoo	1	1	2					2	6
Blackbutt	0.5	0.5						1	2
Other species								14	14
Subtotal	19	13	27	30	4.5	0.5	12	53	159
Dead wood									
Jarrah							50	20	70
Marri								20	20
Wandoo								10	10
Blackbutt								5	5
Other species								5	5
Subtotal							50	60	110
Total	19	13	27	30	4.5	0.5	62	113	269

Table 6b
Key to product specifications

	Minimum length (m)	Minimum top diameter (mm)	Sweep percentage (max)	Percentage defect sectional area	Other special constraints
Sawlogs					
1st Grade	2.1	250	3	50	No pinhole borers
2nd Grade	2.1	250	4	70	No pinhole borers
Small	2.1	150	0	5	
Short	1.0	250	3	50	
Chipwood	2.1	100	5	70	0% charcoal
Poles	9.5	150	1	10	
Piles	11.0	200	1	10	
Industrial rounds	1.8	60	3	10	
Industrial firewood	1.0	100	5	80	0% termites

Figure 45

Statement of risk in volume estimation. For jarrah gross bole volume (alive) the odds against the resource being less than a given level are, for example, 200 to 1 against the jarrah resource being less than 28.3 million m³