Integrating GIS and RDBMS Technologies During Construction of a Regional Groundwater Model

Ross S. Brodie
Australian Geological Survey Organisation (AGSO), Canberra, 2601, Australia

Abstract Finite difference groundwater flow models like MODFLOW require cell-by-cell averages for a myriad of parameters. In reality, the data a modeller uses comes from many sources in a variety of formats. Point measurements from boreholes are a critical dataset and are combined with lines (potentiometric & structural contours), polygons (geological & land use maps) and rasters (Landsat imagery). The modeller needs a working environment to store, integrate and analyse these datasets and to derive the cell-by-cell model input. Secondly, the model output needs to be compared with the original source data that describes the real world. Borehole information is typically stored in a relational database management system (RDBMS) and geographical information systems (GIS) are designed for managing spatial information. These technologies have been used as the working environment for the Lower Darling model, which is a large regional groundwater flow model within the Murray Basin, southeast Australia. Different strategies were developed to manipulate the available data into MODFLOW input files and also for the modelled heads and flows to be compared with field observations.

1. INTRODUCTION

The Lower Darling groundwater model covers nearly 80,000 km^2 of the northwest quadrant of the Murray Basin, in southeast Australia. The model is one of five regional groundwater flow models subdividing the basin, constructed to allow prediction of the long term groundwater response to natural resource management strategies (Figure 1). The imperative is the rising watertable in the Cainozoic basin sediments that has caused waterlogging and land salinisation as well as increased salt loads to the Murray River. This watertable rise is due to enhanced recharge caused by removal of deep rooted native vegetation for pasture, cropping and irrigation development. At this stage, the Lower Darling model is a MODFLOW-based, steady state, three dimensional model simulating groundwater conditions observed in 1988.

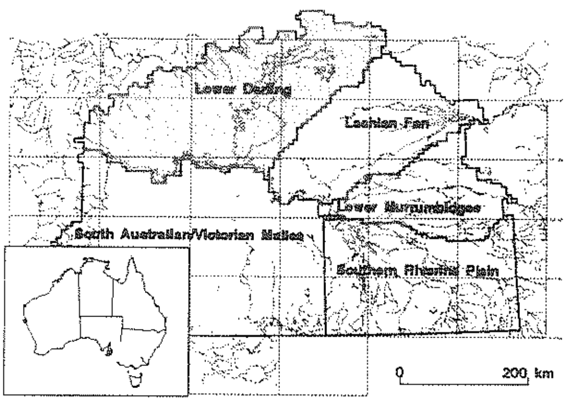

Figure 1: Murray Basin regional groundwater models.

2. THE WORKING ENVIRONMENT

The model was constructed using a series of linkable Unix-based software systems, allowing the disparate datasets for the model area to be stored, integrated and manipulated (Figure 2). This includes a relational database management system (RDBMS) to store key borehole data, a geographical information system (GIS) to maintain and analyse spatial data, image processing software and a model pre/post processor. In this way, the inherent strengths of each of these systems can be maximised.

The borehole record is stored in a series of data tables in a RDBMS [Brodie, 1993]. This includes general information such as bore name, location, elevation and depth, downhole stratigraphy, information on the well intake and pump tests, standing water level measurements and water chemistry. Authority tables are used to store codes and definitions, to ensure consistency in coding and that queries return the appropriate records. The unique identifier for each borehole is stored in the *boreid* field which is common to most of the data tables. It is used to link the data tables together, providing the functionality of a relational database. The identifier is also stored in the GIS coverage containing the points representing the boreholes. In this way, results of a query using the Structured Query Language (SQL) of the borehole database can be displayed in their appropriate geographical position. This capability was used extensively in the compilation of model input, whereby relevant data were processed and extracted from the borehole database and integrated with spatial information within the GIS.

An image processor was used to rectify, clip and manipulate imagery for use within the GIS. Landsat MSS imagery (to indicate land use and vegetation clearing) and rasterised gravity and aeromagnetic data (useful for defining basin structure) were processed. The software generated standard BIL files which could readily be incorporated into GIS displays.

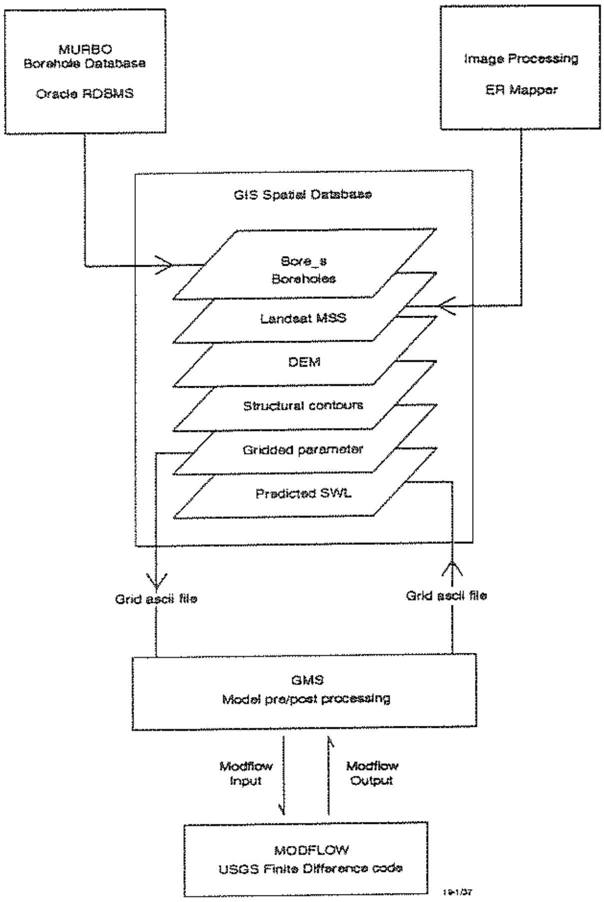

Figure 2: Working environment for compiling the Lower Darling Model.

The bulk of the data manipulation occurred within the GIS. Model parameters were averaged for each 7.5km x 7.5km model cell by creating a 2-D grid of the same dimensions and extent as the model domain. How this was achieved varied with the format of the source data (Figure 3). When a parameter had been mapped as a continuous surface by contours and point measurements (eg. topographic contours and spot heights), a triangular irregular network (TIN) was constructed. This involved forming the points which represent the vertices of the contours into a series of interconnecting triangles. Additional information such as spot heights, extra contours, or arcs depicting axes of maxima (ridges) or minima (valleys) were added to better define the data surface. This was essentially an iterative process where contours generated off the TIN were verified against the original contours for goodness of fit. The source data were then modified or added to and a new TIN was generated. When the TIN adequately represented the data surface, it was resampled on the basis of the model grid.

Other datasets such as soils and surface geology are categorical and are mapped as discrete polygons. For the model, layer hydraulic conductivities and leakances were mapped in this way. These polygon GIS coverages were simply resampled on the basis of the 2-D model grid. Model parameters were also generated within the GIS by algebraically combining grids. Examples include thickness grids where the layer base was subtracted from the layer top, and transmissivity grids where thickness was multiplied by the horizontal hydraulic conductivity.

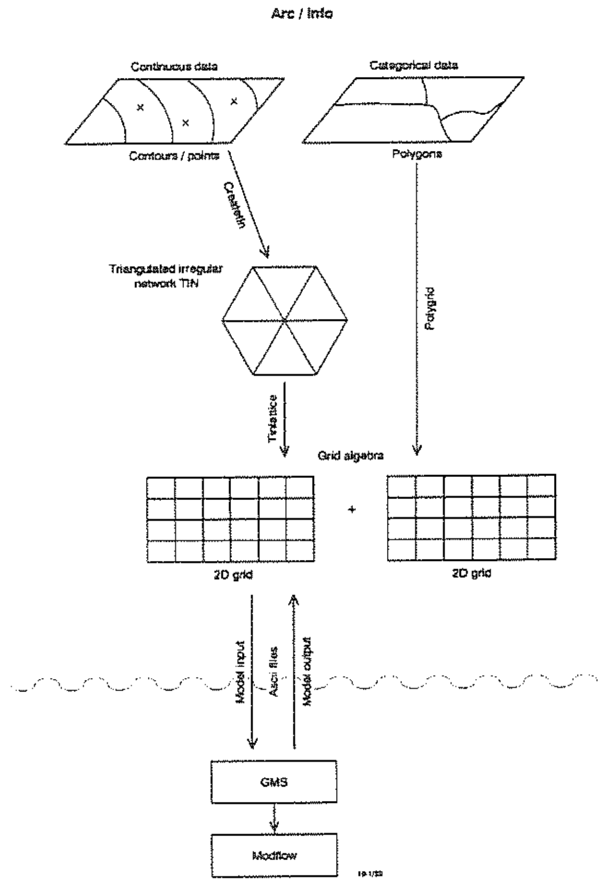

Figure 3: GIS data processing for creating MODFLOW input.

Once the 2-D grid matching the dimensions of the model grid was generated in the GIS, it was exported as an ASCII file. In turn, this file was imported into the model pre/post processor so that the model input files could be constructed. The processor was used to validate and visualise model input. The data pathway was also operated in the reverse direction (Figure 2). Once the MODFLOW simulation was run, the predicted heads could be imported into the pre/post processor. An ASCII file could then be generated for import into the GIS. In this way, the model predictions could be integrated with the spatial database established for the model.

3. CREATING THE MODEL INPUT

The functionality of both the RDBMS and GIS was used in the derivation of model input. This can be best described by giving examples of methodologies developed during model construction.

3.1 Hydraulic Conductivity

According to Anderson and Woessner [1992], if the thickness of a model layer (B) is much larger than the thickness of an isotropic lithological interval (b) then the hydrologically equivalent horizontal hydraulic conductivity (Kh) for the model layer can be calculated using:

$$(Kh)_{i,j} = \sum_{k=1}^{m} \frac{K_{i,j,k} b_{i,j,k}}{B_{i,j}}$$

where

$$B_{i,j} = \sum_{k=1}^{m} b_{i,j,k} \qquad (1)$$

This approach was applied to the 20,000+ borehole lithology intervals recorded in the relational database (Figure 4). A nominal hydraulic conductivity (K) needed to be assigned to each of the lithological intervals. In the lithology data table, the lithological description of each downhole interval is defined by a rock type field (*name*) and six descriptor fields (*desc1* to *desc6*). All of the 3728 unique combinations of these seven fields were loaded into a separate data table so that a nominal hydraulic conductivity could be assigned to each lithological description. These assigned values were then updated into the K value for each of the borehole intervals in the lithology table.

The depth to top and thickness (B) for each of the model layers is required for each borehole. A data table (*layer*) was created to store this information and was populated by using the stratigraphy table (*strat*). The layer top is the top of the uppermost stratigraphic unit making up that model layer, and the layer thickness is the sum of the stratigraphic units comprising the model layer.

With the essential elements of the horizontal hydraulic conductivity equation (1) available, the weighted average of Kh was calculated for the model layers intersected downhole. Any potential mismatch between the lithology intervals and the model layer intervals were accounted for in the SQL statement. For instance, if the top of the model layer fell within a particular lithology interval, then only the difference between the model layer top and the base of the lithology interval was used as a thickness term (b). The final product is a data table (*layer*) defining the model layers intersected in each borehole, with a horizontal hydraulic conductivity (Kh) based on the lithological record assigned to these model layers.

The *layer* data table was accessed from within the GIS to allow the Kh estimates to be placed in their spatial context. Boreholes which only partially penetrated the model layer and therefore had a potentially misleading Kh estimate were highlighted in the display. Mapping such as aquifer yield, geological structure and aquifer boundaries were overlain to assist in the manual interpolation of Kh. Polygonal zones were mapped and subsequently rasterised for input into the model.

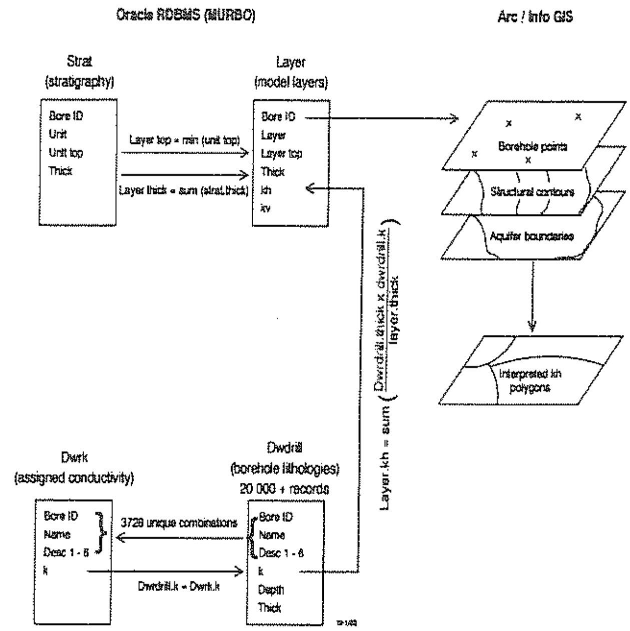

Figure 4: Estimation of Kh from lithological descriptions.

Vertical hydraulic conductivities were calculated in a similar way [Anderson & Woessner, 1992]. Using the thickness (b) and assigned conductivity (K) of assumed isotropic lithological layers within the model layer, the hydrologically equivalent vertical hydraulic conductivity (Kv) for the model layer of thickness B was calculated by using:

$$(K_v)_{i,j} = \frac{B_{i,j}}{\sum_{k=1}^{m} b_{i,j,k} / K_{i,j,k}} \qquad (2)$$

As such, the averaged Kv for a model layer is largely defined by the conductivity assigned to the finest grained lithological unit. In MODFLOW, the vertical hydraulic conductivities are used to define a leakance term between model layers. This deals with the prism of aquifer between a model node and the node of the underlying cell, rather then the entire thickness of the model layer. Hence, lithological records for the borehole interval defined by the midpoint of a model layer as the top and the midpoint of the underlying model layer as the base is used in equation (2).

3.2 River Package

The River package describes the interaction between the river system (consisting of reaches of the Murray and Darling Rivers) and the shallow aquifer. It is also a good example of a MODFLOW module which dealt with a variety of datasets in a variety of formats within the GIS environment. The input file requires for each designated river cell, the model layer, row and column number, the elevation of the river stage (*stage*), the river bed hydraulic conductance (*cond*) and the elevation of the base of the river bed (*rbot*).

The configuration of river cells was maintained in a GIS polygon coverage, created by combining a line coverage depicting the main drainage with a polygon coverage representing the model cells. This allowed automatic calculation of the row and column identifiers for each river cell. The layer attribute was assigned to the river cells by overlaying the boundary polygons of the model layers.

To assist in the estimation of the remaining input parameters, the river cells were classified into the following categories: (i) confluence cells where significant river channels meet; (ii) end cells at the margins of the active model domain; (iii) gauge cells containing a river gauging station; and (iv) lock cells where a lock and weir pool was present. These particular river cells were important as they became the end points with the required information to interpolate the parameters for the intervening river cells.

For instance, a median river stage value was estimated from the gauging record of ten stations located on the Darling river system. Using GIS functionality, the stage of the river between these gauge cells were calculated by linear interpolation. In contrast, the Murray River is regulated by a series of locks which creates a stepped effect in the river stage. Hence, the linear interpolation between the lock cells used the upstream modal average of the lower lock and the downstream modal average of the upper lock as the stage limits.

The river bed conductance parameter requires the geometry of the river bed (length, width and thickness) and an estimate of its vertical hydraulic conductivity. By intersecting the drainage line coverage with the model cell polygon coverage, the river arcs are split on the basis of the model cells. As the length of each arc is a system-generated attribute, it was a simple matter of summing the lengths of all the river reaches belonging to a particular model cell to obtain the channel length parameter.

The estimation of the channel width was based on 5m resolution orthophoto imagery along the rivers. This was displayed as a background image and a series of points created where the channel width was measured using the distance functions of the GIS. These measuring points were then allocated to the nearest river reach using a nearest neighbour analysis. The width of each river reach was calculated by averaging the measuring points allocated to that particular arc. In turn, the average width of the river channel for each model cell was calculated as the average for the river reaches within that cell, weighted on the basis of length. Due to the absence of field data, both the river bed thickness and the hydraulic conductivity were assigned during model calibration.

The GIS methodology to estimate the elevation of the river bed depended on the availability of river bed surveys. Data varied from high density sampling to sporadic measurements, necessitating different strategies for calculating the parameter. Along the South Australian reach of the Murray River, a very detailed river survey made up of a multitude of transects zigzagging across and up the river was available. This resulted in a point coverage measuring the top elevation of the river bed both laterally across the channel and longitudinally upstream. The average elevation of the deepest part of the river channel for the reach within each model cell was derived from this survey data. This required complex GIS processing to split up the survey into individual cross-river transects, to find the minima for each of these transects, and to average these minimum survey points allocated to each individual model cell.

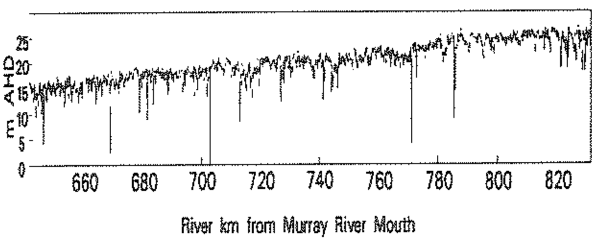

Figure 5: Profile of bed of Murray River from Wentworth to SA-Vic border (Rural Water Commission of Victoria, 1989).

In contrast, a longitudinal profile of the Victorian part of the Murray River had been measured by an echo sounding following the centre line of the channel. The data was stored within the GIS in the vertical plane, with river distance from the Murray Mouth as the horizontal axis and elevation as the vertical (Figure 5). To link this dataset to the model cells, the minimum and maximum river distance bounding each model cell was calculated. This was done within the GIS by locating geographical features which had designated river distances and adding (or subtracting) the length increment to the end points of the river reaches at the boundaries of each cell. The vertices of the arcs defining the longitudinal profile (Figure 5) were then converted into points and the maximum and minimum river distances that described each river cell used as the criteria to select subsets of these points. To reduce the influence of anomalies caused by local obstructions during the survey (refer Figure 5), any profile data beyond two standard deviations were removed from the subset, before a mean value was calculated for the model cell.

The Darling river bed has poor survey control, limited to elevations estimated at the gauging stations. Like the process to calculate the river stage, these gauge cells were used as end points for linear interpolation of values for the intervening cells.

3.3 Recharge

Recharge estimation was critical to the model, as one of the objectives was to predict the long term hydrological effects of regional clearing of mallee vegetation. Recharge has been measured to increase from <0.1% of rainfall under native vegetation to 2-13% post-clearing [Cook et al, 1996].

Vegetation and soil clay content are the dominant controls in the model area in determining the proportion of rainfall that recharges groundwater. These factors were incorporated into a raster based GIS methodology to define the recharge array (Figure 6). Basically, recharge was defined as a function of land systems mapping as a de facto soils map, combined with vegetation and rainfall, with consideration for accessions from other sources.

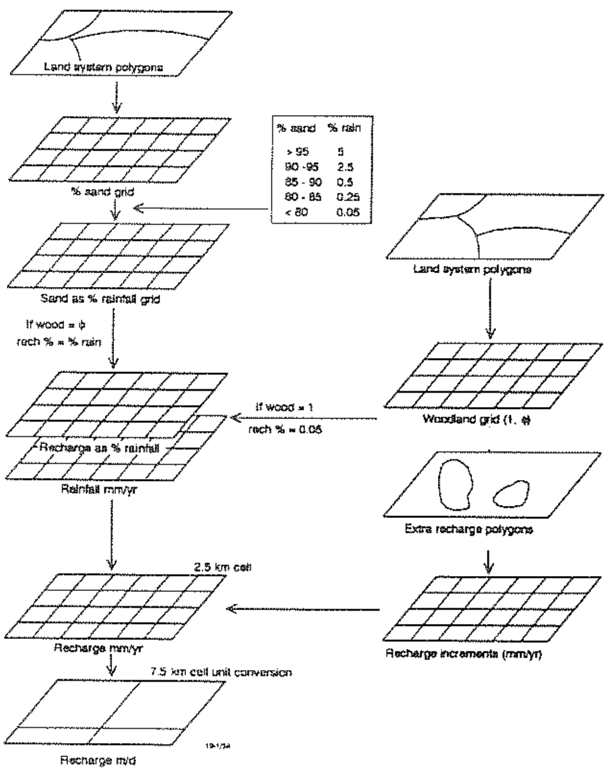

Figure 6: GIS methodology to estimate areal recharge.

Land systems mapping [Walker, 1991; Laut et al, 1977] was used to map regional changes in soil sand content. Each land system unit was assigned a nominal percentage sand based on its dominant soil description. This value was high (>80%) for aeolian landforms with characteristically sandy soils, and low (<20%) for the clayey soils of lacustrine landforms. These values were stored as an attribute in the land system polygon coverage, and used in the translation to a % sand grid. In turn, another classification scheme was used to convert the % sand grid into a recharge rate expressed as a percentage of rainfall. The thresholds used in this translation were based on the published literature on field estimates of recharge in the basin. For example, land systems with 85 to 90% sand, were allocated a recharge rate of 0.5% of rainfall.

The land systems mapping was also categorised on the basis of the density of dominant vegetation. This produced a grid with a simple binary classification where land systems containing dense woodland vegetation such as mallee, belah-rosewood or black box were assigned a value of 1, the remaining land systems being zero. The woodland grid was used as a condition to create the grid of recharge as a percentage of rainfall. If the grid cell contained woodland (value = 1) then recharge was set as 0.05% of rainfall, regardless of soil type. Otherwise (value = 0) the percentage rainfall as designated by the sand content was used. The resulting % rainfall grid was multiplied by the average rainfall grid, to calculate rainfall recharge in mm/yr.

Additional recharge sources such as irrigation accessions were delineated as polygons with an assigned recharge rate, ranging from 0.03 to 20 mm/yr. This polygon coverage was converted to an equivalent recharge grid in mm/yr, and added to the rainfall recharge grid generated from the vegetation and soil classifications. The resulting grid was resampled into a cell size corresponding to the model and a conversion factor used to derive the final recharge grid in m/d. This was the grid which was exported into the model pre/post processor for use in the MODFLOW recharge package.

4. ANALYSING THE MODEL OUTPUT

Following a model run, the MODFLOW output of simulated heads, drawdowns and cell-to-cell flow terms were imported into the pre/post processor. This enabled the display functions such as contouring, colour fill and vector arrows, typical of these software packages to be used. Hence, the model output could be displayed in the context of the model input.

In addition, the model output could be reformatted for import into the GIS. This allowed comparison of the modelled heads and flow terms in the context of the spatial database representing the model area, giving added flexibility during the calibration process.

For example, the modelled heads for each layer could be translated into a 2-D grid within the GIS and algebraically combined with other 2-D grids. The modelled head grid was subtracted from the observed head grid to derive a head residual, highlighting problem areas. The modelled heads of one layer were subtracted from the modelled heads of the overlaying layer to present the vertical flow regime. A range of statistics on the grids useful as calibration indicators were derived. These functions allowed comparison of the model output directly with the source data which was used in the compilation of model input. Changes to the input parameters were made and the MODFLOW files updated.

The simulated flux between the river and aquifer could be assigned to each designated river cell stored in the river polygon coverage. Hence, the river-groundwater interaction could be displayed with features such as locks, gauging stations and river confluences. By combining the groundwater salinity of the unconfined aquifer interpolated by the GIS for each cell with the modelled flux, the salt load in the cells where the river is gaining groundwater could be estimated. Hence, reaches of the Murray River with high accessions of saline groundwater were identified.

The same methods could be used for estimating salt loads associated with water flow between the model layers and external sources/sinks. These include the large water storages along the rivers (such as the Menindee Lakes), as well as the Lower Cretaceous aquifer underlying the Cainozoic sediments, which are represented as general head boundary cells in the model.

5. CONCLUSIONS

The combination of a relational database, GIS, image processing software and a MODFLOW pre/post processor was used to successfully calibrate a regional groundwater model of the northwestern quadrant of the Murray Basin. Over 90% of model cells have simulated heads within five metres of observed heads, which is a reasonable error criterion considering the paucity and poor quality of hydrographic data in the area. In addition, the simulated salt loads due to groundwater accessions into the Murray River are consistent with previous estimates. The combination of software allowed the inherent strengths of each of these packages to be used to advantage. Data in various formats from various sources could be properly managed. Data processing strategies could be tailor-made to derive the model input from the available data. Likewise, the model output could be directly compared with the spatial database which represents the real world.

6. ACKNOWLEDGMENTS

Heather Rennie, Martyn Moffat and Evert Bleys (AGSO) helped prepare the figures for this paper. Jim Kellett (AGSO) reviewed the original manuscript. The Lower Darling groundwater model is part of Natural Resource Management Strategy (NRMS) Project D5039 - *Murray Darling Basin Groundwater Modelling*, funded by the Murray Darling Basin Commission.

7. REFERENCES

Anderson, M.P., and W.W Woessner, *Applied groundwater modelling - simulation of flow and advective transport,* Academic Press, 381pp., London, 1992.

Brodie, R.S., Documentation for MURBO: The Murray Basin Borehole Database. *AGSO Record 1993/77.* Australian Geological Survey Organisation, Canberra. 1993.

Cook, P.G., A.K. Kennett-Smith, G.R. Walker, G.R. Budd, R.M. Williams, and R. Anderson, Impact of dryland agriculture on land and river salinisation in the western lands, New South Wales. *CSIRO Technical Memorandum 96.16*, 1996.

Laut, P., P.C. Heyligers, G. Keig, E. Loffler, C. Margules, R.M. Scott, and M.E. Sullivan, *Environments of South Australia, Province 2 Murray Mallee, Province 5 Eastern Pastoral, Province 6 Flinders Ranges.* Division of Land Use Research, CSIRO,1977.

Rural Water Commission of Victoria, *Murray River Survey - South Australian Border to Wentworth*, Survey & Hydrographic Branch, RWC, 1989.

Walker, P.J., Land systems of Western New South Wales. *Technical Report No 25*, Soil Conservation Service of NSW, 1991.