INTEGRATING HISTORICAL DATASETS TO PRIORITISE AREAS FOR BIODIVERSITY MONITORING?

Anita Smyth [1], *Vanessa Chewings* [1], *Gary Bastin* [1], *Simon Ferrier* [2], *Glen Manion* [2] *and Bretan Clifford* [3]

[1] CSIRO Sustainable Ecosystems, PO Box 2111, Alice Springs, NT 0871
[2] Department of Environment and Conservation, PO Box 402, Armidale, NSW 2350
[3] Department of Infrastructure, Planning and Environment, PO Box 2130, Alice Springs, NT 0871

ABSTRACT

We conducted a 'proof of concept' study to assess the feasibility of using historical survey and environmental datasets for prioritising areas for monitoring biodiversity. Our specific interest was to test the conventional wisdom that existing survey data on plants and animals, and environmental products from satellite imagery collected for different purposes could be readily used to identify areas in the landscape where monitoring of biodiversity was a priority. If this proved to be true, we believe that the commercial benefits would be substantial as the costs of data gathering would be less, the large investment already associated with obtaining these datasets would be value-added and the maps of priority produced from the study would help land managers to target areas for monitoring biodiversity. In the late 80s and early 90s, Parks and Wildlife Commission of the Northern Territory (PWCNT) and CSIRO invested in the Central Australian Ranges Geographical Information System (CARGIS). We used the comprehensive datasets from this study, and spatially coincident airborne and satellite remote-sensing products from other studies as our data sources for this case study. In this paper, we report the value of using historical datasets for prioritising areas by highlighting the benefits and deficiencies with this approach and suggest improvements.

INTRODUCTION

Government natural resource management (NRM) agencies are calling for environmental performance reporting on the condition of biodiversity and other natural resources in Australia's rangelands. Whilst land managers are becoming increasingly interested in taking up such initiatives, they remain hamstrung by the lack of policy frameworks and tools to accomplish it. As most of us know, monitoring biodiversity is not an easy task whether you are a biodiversity expert or land manager. First, available resources for biodiversity are driven by erratic climate conditions. Another difficulty is land use pressures are uneven across landscapes and without years of on-ground experience it can be difficult to know confidently where damage is likely to be irreversible. Finally, knowing what to measure (e.g. Faith 2003), how and where it should be done for practical outcomes are daunting challenges indeed.

Ideally, to identify where to monitor biodiversity, we need a comprehensive and detailed knowledge of the distribution of all entities of biodiversity (genes, populations, species, communities, ecosystems) occurring within a region (Ferrier 2002, Faith 2003). Although considerable investment has been put into bioregional surveys, it remains unrealistic to ever expect such complete knowledge. Instead the common practice is to use those entities for which we do have information as surrogates for biodiversity as a whole (e.g. species, assemblage of species or ecosystems). There are many different types of surrogates that could be used from individual species to complements of species, ecosystems and communities (see review by Faith 2003). There is also a diversity of modelling techniques for deriving these surrogates ranging from coarse-filter ones using environmental domain modelling to the more intermediate filters of ecosystem mapping and general dissimilarity modelling (GDM) to the fine-filter modelling of individual species / species assemblage-environment modelling (Ferrier 2002). More importantly, they rely on a reasonable amount of biophysical data. With the development of the National Land and Water Resources Audit, State and Territory bioregional surveys and environmental data products, an opportunity exists to explore how readily available datasets can be used to identify where best to monitor biodiversity. There are notable benefits in using existing datasets for this purpose such as value-adding NRM investments and reducing R&D costs. However, these datasets often differ in their quality of measurement and spatial resolution, which raises concern about their utility for underpinning biodiversity monitoring. Datasets may need to be broken into subsets to accomplish our purpose and this may then compromise results for statistical modelling. Knowing their value will increase and improve their use for NRM planning and influence future design of, and investment in, bioregional surveys.

We are presently completing a 'proof of concept' study to assess the feasibility of identifying priority areas for monitoring biodiversity. Historical biological survey data and environmental products derived from a digital elevation model, and airborne and satellite datasets for the central Australian ranges region are used to create maps of compositional dissimilarity (i.e. collective biodiversity) and landscape condition classes to prioritise areas for monitoring. We chose this region because of the large number of datasets generated for the Central Australian Ranges Geographical Information System (CARGIS) project undertaken by CSIRO and Parks and Wildlife Commission of the Northern Territory (now the Department of Infrastructure, Planning and Environment) during the late 1980s and early 1990s (http://www.cse.csiro.au). This study used an assemblage or individual species approach to map the predicted distributions and likelihood of occurrence of plant and animal species for specific localities at a fine scale (Griffin and Duguid 1997). We used the CARGIS datasets and also introduced new climate and satellite datasets to create the initial models and carry out prioritisation analyses.

In this paper, we report on the benefits and deficiencies of using historical biological and environmental data to create an initial set of models for mapping compositional dissimilarity and then suggest some improvements. We specifically do not discuss the modelling aspects as these will be covered in the client's report and in another manuscript.

STUDY AREA

The study area was the region of the digital elevation model (DEM) created in the previous CARGIS study (Fig. 1). This covers some 53,000 sq km in central Australia (extending from the Strangeways Range in the north, east along the Harts Range, south to the James Range and to the western extent of the MacDonnell Ranges). It has parts of five bioregions – Burt, MacDonnell Ranges, Simpson-Strzelecki Dunefields, Finke and Great Sandy Desert. The study area is referred to hereafter as the DEM. Mismatch in the spatial extent of some datasets meant that the DEM was subset into a smaller area called sub-DEM.

SPATIAL STATISTICAL MODELLING AND DATA REQUIREMENTS

We used the non-linear, multivariate general dissimilarity modelling technique for modelling compositional dissimilarity as the first step in identifying priority areas mainly because it uses compositional dissimilarity (or complementarity) of survey areas as a surrogate of biodiversity. This measure indicates a larger amount of diversity in species occurrence than species richness alone. It is the mix of two components of biodiversity – the difference in composition of species between different habitats/environments and the difference in composition of species between spatially isolated occurrences of the same habitats/environments in the landscape (Pressey *et al.* 1993). The GDM was developed by Ferrier and colleagues (2002) to predict the compositional dissimilarity of biodiversity as it changes in response to environmental gradients at pairs of survey sites in the landscape.

Another value of the GDM is that it gives reasonable models with relatively small amounts of biological survey data. This is an important consideration as most rangeland regions are data-poor with few and possibly biased samples both spatially and taxonomically of biodiversity. An exception to this is the central Australian ranges, which are data-rich having CARGIS and, more recently, other bioregional surveys. However, if we need to subset datasets to adjust for anomalies, and thereby decrease the number of survey sites for modelling, then we have an opportunity to examine the performance of the GDM to predict compositional dissimilarity.

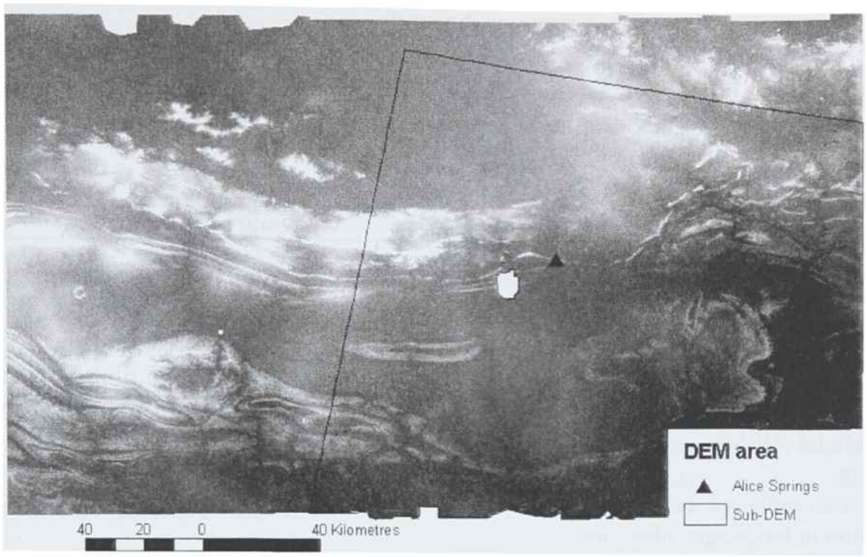

Figure 1. Map of the DEM (53,112 sq km) and sub-DEM regions (22,600 sq km) of the study area showing mountain ranges and plains. The distinctive ridges in the middle of the photo are the MacDonnell ranges.

The GDM requires spatial data on species occurrence and continuous data for environmental attributes. Our first step was to model predicted compositional dissimilarity and then extrapolate it for the entire study area, followed by a numerical classification of these predicted values in order to derive a map of differences in compositional dissimilarity across the entire study region. The next step will be to combine this information with a spatial index for landscape condition and use the environmental diversity approach of Faith and Walker (1996a,b), to prioritise and select areas for monitoring biodiversity.

READILY AVAILABLE DATASETS – GDM MODELLING

Two biological survey and sixty environmental datasets were used to conduct the GDM work. These are summarised in Table 1.

Biological Survey Data

The flora dataset contained information on the presence/absence of all perennial woody species for 4,683 sites. The data were collected from 248 transects which were 1 km long with up to 20 sites spaced 50 m apart. All species were recorded within a 20 m radius of the site. A GPS reading was taken for each transect at the first survey site and stepped out for the other 19 sites (G. Griffin, *pers. comm.*). To avoid a lack of independence of occurrence of plant species amongst sites, we used the single occurrence of species for the whole transect (25 ha) and assigned the GPS location recorded at the first site. This reduced the number of sites to 248 but the total area surveyed for all 20 m-radii was 62 sq km. No grasses or other non-woody, annual plants species were recorded. A total of 122 species were used to predict compositional dissimilarity for the study area.

The fauna dataset had presence/absence records from surveys of birds, mammals and reptiles for 117 sites throughout the DEM area. Only residents and non-irruptive species were used in the analysis. No invertebrate data was available for this study. The total number of species used for predicting compositional dissimilarity was 35, 110 and 76 species of mammals, birds and reptiles respectively.

These data were the response variables for the GDM modelling and were not aggregated into assemblages or taxonomic groups at this stage of the study.

Table 1. Description of datasets used in general dissimilarity modelling with spatial resolutions covering the Central Australian Geographical Information System (Griffin and Duguid 1997) and the digital elevation model (DEM) within CARGIS.

Layer	Source	Coverage
Flora records	Griffin (1997b,c)	CARGIS
Fauna records	Hobbs and Reid (1997a,b,c) Reid *et al.* (1997)	CARGIS
Geochemistry (iron oxide, calcium oxide, manganese oxide, magnesium oxide, silicon oxide, potassium oxide)	Griffin (1997a)	CARGIS
Digital Elevation Model (DEM)	Tier (1997)	DEM
DEM derived layers • slope, aspect, solar radiation, catchment size, wetness, position in landscape, elevation diversity, log catchment size, north-south slope.	Griffin and Chewings (1997) Tier and Chewings (1997)	DEM
Climate layers of temperature, radiation and moisture	ANUCLIM/BIOCLIM http://cres.anu.edu.au/outputs/anuclim.php	DEM
Satellite data • Landsat TM 1989 mosaic bands 1,2,3,4,5,7 • Probability of woody and non-woody vegetation cover derived from Landsat TM 1989 mosaic.	Furby (2002)	DEM
• PD54 (index of vegetation cover) variance and contrast (derived using Landsat ETM 2000 data)	Pickup *et al* (1993)	
Radiometrics – potassium, thorium, uranium, total count	Clifton (2003)	CARGIS

DEM Derived Layers

The resolution of the DEM was 100 x100 m pixels and all grid layers were calculated during the CARGIS study. Slope, north-south aspect, wetness, position in the landscape, elevation diversity, radiation and catchment size were the predictors for GDM modelling and are described in the references listed in Table 1.

Climate layers

Climate data for this study were derived using ANUCLIM 5.1 (http://cres.anu.edu.au/outputs/anuclim.php) and the radiation layers corrected for slope and aspect. We ran BIOCLIM (http://cres.anu.edu.au/outputs/anuclim.php) using both weekly and monthly calculations. Thirty-five predictors describing temperature, radiation, moisture and the seasonality of each were used in the GDM as predictors (see references for website listing of 35 variables).

Satellite data

Three types of satellite data were used to derive ten vegetation predictors: (1) Landsat 1989 TM raw bands 1, 2, 3, 4, 5 and 7; (2) derived images of the probability of woody and non-woody cover using the Australian Greenhouse Office (AGO) mosaic for 1989; and (3) textural measures of variance and contrast in vegetation cover calculated from the Year 2000 PD54 index. We used these predictors in an exploratory manner with incomplete knowledge of their possible relationships to the flora and fauna variables.

Radiometrics

A recent mosaic of radiometrics for the Northern Territory was provided by the Government Survey Office in the Department of Business, Industry and Resource Development (Clifton 2003). This was broken down into subsets and masked for the DEM area and five predictors were used: potassium, thorium, uranium, a total count of all three elements and a high-pass filtered total count.

GDM MODELS OF FLORISTIC AND FAUNAL DIVERSITY

Sixty-one predictors were used to model the floristic and faunal diversity as measured by compositional dissimilarity. Although we ran GDM models for weekly and monthly climate data, there was little difference in performance as measured by "explained deviance" and the following section reports modelling results for only the monthly climate data. A preliminary flora GDM with geochemistry as a surrogate for geology was run but this predictor was not selected as being influential. The most likely reason for this is that geochemistry measures rock and not soil chemistry and it is the latter which is more likely to influence plant diversity. We therefore did not use geochemistry in any additional GDM.

The performance of the historical datasets to model biodiversity varied noticeably between the sub-DEM and the DEM areas. For the DEM area, 31 of the 60 predictors best explained floristic diversity (Table 2) and the model was a reasonably good fit. The best model of faunal diversity used 18 predictors but it performed more poorly explaining just a third of the differences in animal diversity throughout the DEM area (Table 2). When we ran the same models for the sub-DEM, the best model for floristic diversity explained only 27% of the variation, a very poor fit. The fit of the fauna model was worse still with only 9% of the deviance explained (Table 2).

The most influential predictors of plant compositional dissimilarity in the DEM region were species turnover in different habitats (beta diversity), elevation, wetness, Landsat TM bands 1 and 7, total count of radiometric elements, slope, vegetation texture (variance index), seasonality of temperature, isothermality (mean diurnal range/annual temperature range), moisture of coldest quarter, moisture for lowest quarter, moisture for highest quarter and seasonality of moisture.

Table 2. Results of general dissimilarity modelling of flora and fauna compositional dissimilarity.

Model	Unexplained Deviance (%)	Null Deviance	Best Deviance	Number of Samples
Flora (DEM)	49.76	368.00	184.87	248
Fauna (DEM)	32.53	129.22	87.18	117
Flora (sub-DEM)	27.47	354.95	257.43	178
Fauna (sub-DEM)	9.24	344.89	313.02	40

PERFORMANCE OF HISTORICAL DATASETS

The performance of historical datasets is assessed using the fit of the GDM, their appropriateness relative to pre-processing effort and opportunities for improvement, and their ecological relevance.

Model fit

The historical datasets have performed surprisingly well at predicting plant diversity in the study area as a whole but not for the sub-DEM. Although the number of survey sites is smaller in the latter, the density of survey sites is similar (DEM – 0.005 and sub-DEM – 0.007 sites per sq km) and therefore this is unlikely to be the reason for the poor fit. A more plausible explanation is the location and smaller area of the sub-DEM (22,600 sq km). The DEM has proportionally more sites in the mountain ranges than the sub-DEM which has a balance of sites in lowland, slope and mountainous areas.

Given the small scale of the sub-DEM, it is likely that no gradients could be detected in the environmental predictors whereas they could be measured at the larger scale of the DEM (53,112 sq km). We believe that the sparse coverage of data points in the climate datasets, which represented just over half the predictors in the GDM, may have also contributed to the poor fit. A promised revision of the climate profiles in BIOCLIM for inland Australia may remedy this shortcoming (M. Hutchinson, *pers. comm.*). However, the fact that the landscape condition datasets were only available for the sub-DEM does present us with a challenge. We will either need to be more selective in our choice of climate predictors or develop a new approach for extrapolating the landscape condition datasets to the DEM in order to show how priority monitoring areas can be identified.

The performance of the historical datasets to model faunal diversity was very unsatisfactory. We interpret this poor fit to be caused by inappropriate and/or insufficient predictors of their habitat requirements. Climate, topography, radiation, the raw Landsat TM bands, radiometrics and the cover of woody and non-woody vegetation and its textural characteristics may describe broad patterns in potential habitat for animals (e.g. Coops and Catling 1997) but it's the temporal changes in the quality of that habitat over time that influence animal distributions at larger scales. Our historical datasets did not reflect temporal changes in environmental predictors. Even so, our results highlight that we need to explore further our choice of environmental predictors to understand exactly what aspects of habitats they represent especially the TM bands, radiometrics and the textural indices.

Pre-processing and improvements

Approximately 65 person-days were required for pre-processing the data. This included dataset selection, accuracy checks, projecting imagery from GDA94 to AGD66, conversion of categorical data to continuous data, derivation of new indices from existing datasets, creating climate layers, and masking datasets to the DEM and sub-DEM areas. Some anomalies existed that needed improvement, e.g. the vegetation field data were collected in transects up to 1 km in length, or 10 pixels of the environmental data. These transects often crossed gradients of slope and aspect, but we used the value of a single grid cell (the start of the transect) for each of the environment layers. This method of sampling will impact on the value of some layers more than others. The climate layers were generalised average surfaces and it is unlikely that they reflect conditions around the period of the ground sampling. Also, a DEM was essential for the derivation of climate layers and correction of radiation layers using slope and aspect made the products more convincing. The DEM was produced from streams, contours and spot heights in the early 1990s and took about a year's effort. A number of the field samples were also not used as they were located outside the DEM area. Ten years later, other available techniques could reduce the amount of resources needed to create a suitable DEM and it may therefore be possible to extend the area of the DEM at some stage.

We used available Landsat TM imagery from 1989 (around the time of the field work) and 2000 but the mosaics are composites of data acquired on various dates, so there are seasonal variations across the imagery, particularly in the wetter year of 2000. The layers showing probability of woody cover were derived using algorithms designed to map cover greater than 20% (Furby 2002), but much of the DEM area has sparse cover, possibly reducing the accuracy of this product.

A constraint of the modelling technique was the need for continuous data. This meant that we were unable to use some of the original CARGIS datasets (e.g. geology and land system mapping). We did explore using geochemistry (% oxides) as a surrogate but it was only available for some of the area.

Ecological relevance

The datasets used for the GDM were selected on what was available at the time. Ideally, we would have liked large amounts of continuous spatial data describing local climate, geology, soil and vegetation attributes that could be used to better describe the habitat requirements of rangeland biodiversity. Instead we have explored the possibility of radiometrics, raw Landsat TM bands, woody and non-woody vegetation and its derived textural characteristics as possible surrogates. We recognise that we were being exploratory and making a considerable leap of faith. Nevertheless, some of these were significant predictors and that allows further study of the implications of this for developing biodiversity indicators. For example, Landsat TM bands 1 and 7, total count of radiometric elements and vegetation texture (variance measure only) were some of the most significant predictors of the distribution of plant species occurrences. At this stage, we cannot explain why other TM bands, or their combinations, that indicate vegetation attributes did not feature. For example, TM 3 is a useful indicator of cover on predominantly red soils (Graetz *et al.* 1982), TM 2 and 3 are used to produce the PD54 index, TM 1 can indicate litter (Pickup *et al.* 2000) and TM 3 combined with one of the infrared bands (TM4-5, 7) indicates vegetation greenness (i.e. NDVI). Further ground-truthing studies are required to understand the ecological significance of the prominence of remotely-sensed layers in the GDM modelling and the distribution of plant species.

CONCLUSIONS

We believe that there are opportunities for using existing data to predict the patterns of biodiversity in the landscape at regional scales. Because we will never have information on the distributions of all plants and animals in the rangelands, there is an increasing demand by NRM and land managers to deliver practical indicators for monitoring biodiversity (see James, these proceedings). This means that we will need to model surrogates of environmental attributes to predict biodiversity patterns and know which environmental attributes consistently represent those patterns the best. Testing the efficacy of historical biological-survey datasets and environmental data is a critical first step. Having said this, we need to be cautious with their use and where possible, develop guidelines for their use in biodiversity modelling.

ACKNOWLEDGEMENTS

We gratefully thank Craig James and Andrew Ash for their helpful comments on the manuscript.

REFERENCES

Coops, N.C. and Catling, P.C. (1997). Predicting the complexity of habitat in forests from airborne videography for wildlife management. *International Journal of Remote Sensing* 18: 2677-2686.

Clifton, R. (2003): Radiometric Map of the Northern Territory 1:2,500,000. Northern Territory Geological Survey.

Faith, D.P. (2003). "Biodiversity", *The Stanford Encyclopaedia of Philosophy* (Summer 2003 Edition), Edward N. Zalta (ed.) http://plato.stanford.edu/entries/biodiversity/

Faith, D.P. and Walker, P.A. (1996a). Environmental diversity: on the best possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity and Conservation* 5: 399-415.

Faith, D.P. and Walker, P.A. (1996b). How do indicator groups provide information about the relative biodiversity of different sets of areas? On hotspots, complementarity and pattern-based approaches. *Biodiversity Letters* 3: 18-25.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology* 51: 331-363.

Ferrier, S., Drielsma, M., Mannon, G. and Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation* 11: 2309-2338.

Furby, S. (2002). *Landcover Change: Specification for Remote Sensing Analysis*. Technical report No. 9. National Carbon Accounting System. Australian Greenhouse Office.

Graetz, R.D. and Gentle, R.R. (1982). The relationships between reflectance in the Landsat wavebands and the composition of an Australian semi-arid shrub rangeland. *Photogrammetric Engineering and Remote Sensing* 48: 1721-1730.

Griffin, G.F. (1997a). *Geochemistry of rocks in the Central Australian Ranges.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Griffin, G.F. (1997b). *Vegetation sampling in the Central Australian Ranges.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Griffin, G.F. (1997c). *Plant species distribution in the Central Australian Ranges - Part 1: Introduction and methods.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Griffin, G.F. and Chewings, V.H. (1997). *Solar radiation: Calculation and biological significance in the Central Australian Ranges.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Griffin, G.F. and Duguid, A.W. (1997). *Central Australian Ranges ecosystems - Overview.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Hobbs, T.J. and Reid, J.R.W. (1997a). *Bird fauna of the Central Australian Ranges - Species summaries and models.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Hobbs, T.J. and Reid, J.R.W. (1997b). *Mammal fauna of the Central Australian Ranges - Species summaries and models.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Hobbs, T.J. and Reid, J.R.W. (1997c). *Reptile fauna of the Central Australian Ranges - Species summaries and models.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

James, Craig, D. (this volume). Biodiversity monitoring in rangelands.

Pickup, G., Chewings, V.H. and Nelson, D.J. (1993). Estimating changes in vegetation cover over time in arid rangelands using Landsat MSS data. *Remote Sensing Environments* 43: 243-263.

Pressey, R.L., Humphries, C.J., Margules, C.R., Vane-Wright, R.I. and Williams, P.H. (1993). Beyond opportunism: key principles for systematic reserve selection. *Trends in Ecology and Evolution* 8: 124-128.

Reid, J.R.W., Hobbs, T.J., Morton, S.R. and Duguid, A.W. (1997). *Vertebrate fauna of the Central Australian Ranges.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Tier, G. (1997). *Digital elevation model of the Central Australian Ranges.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.

Tier, G. and Chewings, V.H. (1997). *Derivation of terrain layers from the digital elevation model of the Central Australian Ranges GIS.* Consultancy report to the Parks and Wildlife Commission of the Northern Territory. CSIRO, Alice Springs.