Using a Bayesian Network in a GIS to Model Relationships and Threats to Koala Populations Close to Urban Environments

D.V. Pullar and T.H. Phan

Geography, Planning and Architecture, University of Queensland, Australia
Email: d.pullar@uq.edu.au

Keywords: *Bayesian networks, ecological modelling, GIS*

EXTENDED ABSTRACT

Predicting native species occurrence on the urban fringe is problematic due to the high level of human disturbance. None-the-less there are many ecologically important remnant patches and other mixed habitats that support native wildlife in these areas. Urban communities also value these habitats for their existence value. Therefore it is important to have a predictive model to help inform planners on what supporting environmental conditions they need to conserve and inform decisions on how to reduce threats and sustain wildlife populations. We have found traditional statistical distribution models do not work well in these environments due to gaps in the data and accounting for irregularities from human disturbance. As a solution we use probabilistic cause-effect models to capture the relevant relationships. These models are developed as a Bayesian network to: i) deal with issues of missing data, and ii) use a model-based methodology for analysis and discovery of non-linear relationships.

Our study is concerned with koala conservation in a coastal regional in Australian. The Redlands Shire in the state of Queensland is colloquially referred to as the Koala Coast due to the abundance of its koala populations, but these are under significant threat from urban expansion. While background disturbances from roads and urban development do represent the main threat, other controllable threat variables such as dog ownership have an impact on koala populations. A Bayesian network model has been developed to understand these relationships and threats (Figure 1).

Habitat conditions for koalas are relatively well understood. Preferred areas have specific eucalypt species, high soil fertility, low acidic and well drained soils. Apart from direct degradation of habitat from land clearing, the major disturbance threats to koalas are from diseases, vehicle movement and domestic pests. Studies focussed on the first two threats show that levels of urbanisation are generally a good indicator of koala population disturbance, but very little has been done on domestic animals. Next to cars, domestic dogs are the highest human-induced impact on koala mortality. Our Bayesian model separates the affects of urbanisation and dog presence to consider what management actions could be taken to reduce the threat of dog attacks.

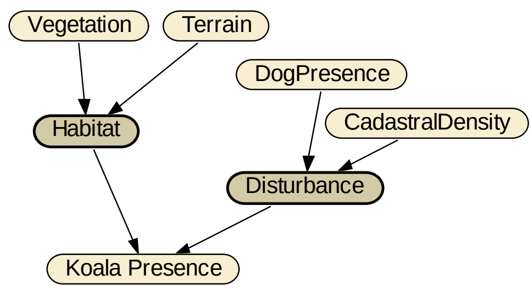

Figure 1. Bayesian network predicting koala presence in relation to threats for domestic dogs.

Data was available on many of the key variables, however it was not complete. Mapping was available for vegetation and urban settlement patterns. We also obtained data on dog presence for this study for the year 2005, but recent data for the response variable was sparse. Data from koala surveys carried in 1996 to 1999 was available, but it was difficult to relate these variables because of the incongruent time periods between the koala surveys and domestic dog data. Attempts to statistically model this relationship gave very poor correlations. The decision was made to use a Bayesian network model that was initially parameterised from a subjective survey conducted with ecologists and environmental managers. The survey gave an initial network that was then trained from the available data.

Outputs from the model show predictions of koala distributions, scenarios on effects of management actions to control dogs, and the uncertainties in predictions.

1. INTRODUCTION

This study is conducted to gain an understanding of koala distributions and how different processes, such as habitat quality and disturbance influence distribution patterns of koala. The context for this research is a region along the east coast of Australia; it is aptly named the Koala Coast due the high koala populations. Parts of this region is located close to highly populated urban areas. The study site is the Redlands which is a coastal shire bordering the city of Brisbane. Studies have documented the influence of urbanisation on koala populations (Dique et al., 2004). The direct threat from vehicle traffic has been surveyed and studied in detail (Dique et al., 2003), but few studies have been done on domestic dog attacks. A better understanding of the relationship between dog ownership and koala attacks is also important to better inform koala conservation policy in urban areas.

The paper reviews the key habitat requirements for koalas in the next section. This identifies the main datasets required for their study; including urbanisation, vegetation, topography and soil. Models to relate habitat and koala presence are discussed in the same section. Typically logistic regression models are used, but these models have low predictive significance for our study area. This is attributed to gaps in completeness of data sets and the difficulty of finding truly representative data to model koala occurrence. Section 4 describes the application of a Bayesian network model to overcome some of the previous difficulties. This includes an expert survey to provide a prior structure and parameterisation of the model. Results for this model are shown. The conclusion discusses outcomes from the study and suggests future directions.

2. KOALA MODELLING

2.1. Koala Ecology

Koalas naturally occur in large numbers along coastal plains, tableland slopes and plains in eastern Australian forests. Further inland, their population is dispersed among intervening woodlands and trees fringing watercourses (Lee and Martin, 1988). Koala occurrence is strongly associated with specific eucalyptus species located on fertile soils, but it does vary geographically and data is rarely available at an appropriate level of detail. As an alternative to using detailed soil-vegetation mapping data, Dique et al. (2004) based their analysis on general land cover data. Land types were classified as urban, remnant, bushland and other to estimate koala density and abundance estimation. The classification still requires in situ vegetation collection for model validation.

Disturbances to koala population affect their distribution and dispersal. Apart from direct loss of habitat via land clearing; the three major disturbance factors are diseases, vehicle movement and domestic pests. Vehicle movement is a major threat with approximately 300 koala injuries reported from 1995 to 2001; mainly in the breeding season (Dique et al. 2003). Road usage and vehicle speed are significant factors. Domestic dog attacks are also a major threat with 85 koala injuries and 66 deaths recorded annually (Dique et al. 2003).

2.2. Models

Species distribution models are commonly modelled with predictive statistical or spatially explicit models based on field data (Guisan and Zimmermann, 2000). There are problems with regard to data uncertainty not accounting for biotic interactions and other environmental causal relationships to the modelled species. The effects of these problems may be reduced by developing models in line with ecological knowledge of the species (Austin 2002a). Species ecology should be well understood, such as home range, movement and dispersal to assist model development. Rhodes et al. (2006) shows there is a high degree of spatial variability for koala distributions in semiurban landscapes. Spatially varying attributes that have a physiological influence on species subsistence include species-specific habitat, topography, soil moisture, nutrient contents and pH, etc. (Austin 2002a,b). The inclusion of human disturbance in these models has had mixed success. At larger landscape scales habitat quality is the main defining variable and other factors are insignificant. At finer scales environmental characteristics within a certain distance of species presence are significant. A recent comparison of modelling methods (Elith et al., 2006) concludes that substantial improvements could be obtained through better ability to accommodate trade-offs between variables and fitting complex responses in model fitting. This points to the need for more expressive models build to capture the specific needs of a species.

Another key consideration for distribution modelling is whether the field data records presence only, or if absences are also observed. With the exception of Poisson statistics, most statistical models require absence observations to model distribution; but this data is rarely collected. Guisan and Zimmermann (2000) and Elith et al.

(2005) discuss this at some length. A reasonable solution is to generate pseudo-absences for randomly selecting points in the landscape.

3. BAYESIAN NETWORK MODEL

This section describes the development of a koala distribution model that accounts for human disturbances. Our aim is to use the model to inform policy on options for koala protection.

Vegetation and terrain datasets were available to correlate with habitat quality for koalas. However there were gaps in data relating to dogs and their impact on koala populations. Table 1 summarises the main data issues. Extensive koala surveys were available in the period 1996-1999, but only occasional koala sighting data was available after this date. We obtained data on dog ownership for 2005, but historical data was not available.

Table 1. Summary of variables for koala model

Variables	1996-1999	2000-current
koala density	known	sparse
dog ownership	unknown	2005 only

Attempts to predict koala distributions using statistical approaches, such as logistic and Poisson regression, gave very poor associations, e.g. a coefficient of correlation less than 0.1. Bayesian networks provide an interesting alternative to deal with data gaps and for building more structured models. Bayesian networks emphasise a form of knowledge representation to capture the main causal relationships of interest, and rigorous analysis with probabilities to assess the strength of casual relationships.

This section describes the Bayesian networks model development and results.

3.1. Model development

A Bayesian Belief Network (BBN) is a graphical model describing a set of probabilistic variables (shown as nodes) and identified causal relationships (shown as links) that specify the joint and conditional probability between variables (Jensen, 2001). The probabilities in BBN’s may be interpreted as subjective degrees of belief and may include variables that have conceptual definitions. These are referred to as latent variables meaning that they are not directly measured but are instead inferred from other variables.

We used latent variables in our BBN model to represent the concepts for habitat quality and disturbance (Figure 1). The combinations of these variables indicated koala occurrence.

Habitat quality was inferred from observed data on vegetation and terrain. Disturbance was inferred from observed data on urban density and dog presence. All the observable data was integrated to a 250 metre analysis grid. This was the minimal area for koala home range and gave sufficient resolution to detect patterns across the landscape.

Vegetation. The Australian Koala Foundation produce a 1:100,000 altlas with map classes for koala habitat (AKF 2006). The classes are derived from patches of remnant vegetation which identify Eucalyptus species that koalas feed on and utilise for sheltering or foraging. It also considers soil characteristics and local community knowledge to categorize habitats as primary, secondary, other, water and unknown habitats. We used the primary and secondary classes as providing high quality habitat and other as moderate quality.

Terrain. Soil moisture content is an important consideration in combination with vegetation for koalas. Soils located in lowland areas have high water content and higher primary productivity for leaf production. However soil moisture data was unavailable, so a surrogate landform index was used instead. The index was derived from a combination of topography and aspect. A 25 metre DEM was used to compute topographic position as the level of landscape exposure (Guisan and Zimmerman, 2000). When combined with slope it is possible to develop a landform index with values signifying valley bottom floors, bottom swales or toe slopes, and ridges, side slopes and hills. Higher moisture content was assumed in valley bottoms, moderate for the toe slopes and poor moisture along ridges.

Urban density. Land use, road and cadastral parcel boundary data was analysed to develop an indicator for urban influence on koala populations. A number of measures were tested for their statistical correlation with koala occurrence. Noting that the correlation was weak due to data gaps; we found the density of cadastral boundaries provided as good a measure as any of the more complex urban characterisations. The line density of urban boundaries intersecting the analysis grid was computed for data available in 1996 and 2005.

Dog presence. Geocoded locations for dog licences were obtained from the local council for 2005, and were spatially associated to the ownership parcels. This was converted to dog presence for grid cells when greater than 20% of the area of a cell had a dog. Note that we did not have any property level data on enclosures for dogs and assumed they could access the full extent of a property. We did not have access to historical dog licence data, so dog presence for 1996 was recorded as unknown in our model.

Koala occurrence. Koala sighting data was obtained from the Queensland Environmental Protection Agency from their WildNet database (EPA 2006). As mentioned, extensive data comprising 2019 koala sightings was available for surveys conducted between 1996 to 1999 by the Queensland Environmental Protection Agency. After the year 2000 only occasional sighting data was available with 48 cases. Koala presence was arbitrarily classed as likely if 4 or more sightings were reported for a cell, and possible if at least one sighting was reported.

3.2. Bayesian model knowledge elicitation

Figure 2 shows the Bayesian model with states for variables. The states for koala presence were probabilistic, i.e. marginal probabilities for likely, possible and unlikely which sum to 1; and all other states were deterministic.

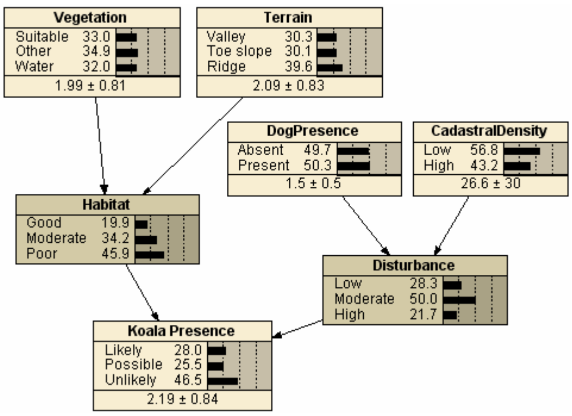

Figure 2. Bayesian network predicting koala presence in relation to threats for domestic dogs.

Interviews were carried out with three experts from the University of Queensland and the Queensland Environmental Protection Agency to classify the network. The network structure and the classification of data for the parent nodes, i.e. vegetation, terrain, dog presence and cadastral density, was explained to respondents. They were then asked to assign values for the conditional probabilities for habitat, disturbance and koala presence. The averaged results for the survey responses are given in Table 2.

Table 2. Conditional probability table for koala presence node in Bayesian network from expert classification

Habitat	Disturbance	Likely	Possible	Unlikely
Good	Low	75.000	20.000	5.000
Good	Moderate	60.000	30.000	10.000
Good	High	50.000	25.000	25.000
Moderate	Low	50.000	25.000	25.000
Moderate	Moderate	33.000	33.000	34.000
Moderate	High	30.000	30.000	40.000
Poor	Low	10.000	25.000	65.000
Poor	Moderate	5.000	25.000	70.000
Poor	High	5.000	10.000	85.000

3.3. Bayesian model learning

A procedure, known as Bayesian learning (Jensen, 2001), was applied to the expert classified network to update the conditional probabilities with the additional observed data for koala sightings. For 1996-1999 and after 2000 there were 265 and 33 cases respectively for grid cells having koala presence as likely or possible. As discussed, no directly observed absence data was available so pseudo absences were generated by randomly picking grid cells that did not have koala occurrences. An equal number of pseudo absences were generated, i.e. 265 and 33, for the two periods.

The expectation-maximization algorithm (Netica, 2006) was used to update the conditional probabilities for koala presence. Bayesian networks learning finds the maximum likelihood for the variables, that is the states which are most likely given the data. If more cases are entered that support an observed state then this is given greater influence on the result. Hence it weights the higher number of case observations from 1996-1999 appropriately. A similar approach is used to factor in the prior classification from the expert interviews. We treated the interviews as providing equal evidence as the observed case data, hence 586 prior cases, e.g. (265 + 33) × 2, were randomly sampled from the expert classified network. Table 3 shows the results for the updated conditional probabilities.

Robust diagnostics of the results are beyond the scope of this paper; however a comparison between Tables 2 and 3 shows similarity in values indicating consistency between the expert derived classification and data learning. A sensitivity analysis of the network shows that much greater variation (20:1) is explained by habitat quality as compared to disturbance. Figure 3 shows the locality of the study area and Figure 4 shows the classified prediction for 2005. It shows the pressure of urbanisation skirting land with high likelihood of koala occurrence. Conservation efforts should be targeted at these areas.

Table 3. Conditional probability table for koala presence node in Bayesian network with prior expert classification trained from observed case data.

Habitat	Disturbance	Likely	Possible	Unlikely
Good	Low	70.709	23.200	6.091
Good	Moderate	54.551	33.331	12.119
Good	High	49.845	25.048	25.107
Moderate	Low	45.955	24.773	29.272
Moderate	Moderate	28.075	31.232	40.693
Moderate	High	28.530	29.036	42.434
Poor	Low	9.993	24.569	65.438
Poor	Moderate	5.123	24.793	70.084
Poor	High	5.133	10.381	84.485

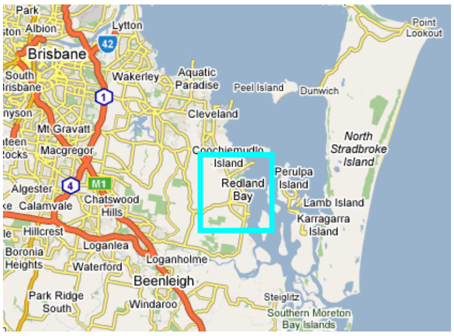

Figure 3. Locality map for study area, Redlands Shire near Brisbane, Australia.

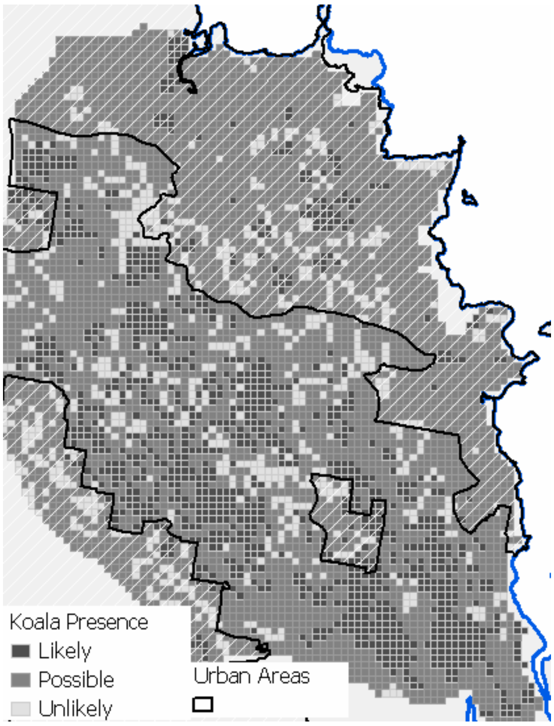

Figure 4. Map for study area showing classification of koala occurrence as likely (dark shade), possible or unlikely (light shade).

4. CONCLUSION

The paper has described the use of a Bayesian network to model koala occurrence. The aim is to relate habitat quality and disturbances from urbanisation and dog presence to koala conservation. If this link can be demonstrated then action can be taken to better protect koalas in line with community values. Modelling koala occurrence with statistical analysis proved difficult because of gaps in the data and the complex relationships that needed to be represented. A Bayesian network provides an efficient encoding of the problem. It allows data with unknown values, e.g. the missing data on dog ownership from between 1996-1999, and is partly explained away from the other data.

The next stage of the project will be to extend the model to relate koala injuries and mortalities. Data is available over both time periods for this analysis and another node will be added to the Bayesian network to evaluate this relationship.

5. REFERENCES

Australian Koala Foundation 2006, SEQ Bioregion Koala Habitat Atlas. [URL: www.savethekoala.com/kha.html, Accessed July 2007]

Austin, M.P. (2002a) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, *Ecological Modelling* (157),101-18.

Austin, M.P. (2002b) Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In: Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, CA, 73-82.

Dique, D.S., H.J. Preece, J. Thompson and D.L. de Villiers (2004) Determining the distribution and abundance of a regional koala population in south-east Queensland for conservation management, *Wildlife Research* 31(2), 109 – 117.

Dique, D.S., J. Thompson, H. J. Preece, G. C. Penfold, D. L. de Villiers and R. S. Leslie (2003) Koala mortality on roads in south-east Queensland: the koala speed-zone trial, *Wildlife Research* 30(4), 419 – 426.

Elith, J., C.H. Graham, et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography 29(2): 129-151.

EPA (2004) Wildnet (Database), Environmental Protection Agency, Brisbane, Australia.

Guisan, A. and N.E. Zimmermann (2000) Predictive habitat distribution models in ecology, *Ecological Modelling* 135(2-3), 147-86.

Jensen, F.V. (2001) Bayesian Networks and Decision Graphs, Springer, New York.

Lee, A. and R. Martin (1988) The Koala a Natural History, News South Wales University Press, Sydney.

Netica (2006) Norsys Software Corp., Vancouver, Canada. [URL: www.norsys.com]

Rhodes, J.R., T. Wiegand, C.A. McAlpine, J. Callaghan, D. Lunney, M. Bowen and H.P. Possingham (2006) Modeling species distributions to improve conservation in semiurban landscapes: koala case study, *Conservation Biology* 20(2), 449-59.

6. ACKNOWLEDGEMENTS

We gratefully acknowledge the provision of registered dog ownership data from the Redlands Shire Council, Queensland, Australia; and an extraction of WildNet data for the study area from the Queensland Environmental Protection Agency.