Application of the Vandermeer-Moloney Algorithm for Determination of Category Size to Australian Native Plants

[1]**Rosenberg, K,** [1]**J. Boland,** [2]**F. Tiver and** [3]**I. Watson**

[1]Centre for Industrial and Applied Mathematics, University of South Australia,
Mawson Lakes, SA, 5095. E-Mail: Kathrine.rosenberg@unisa.edu.au
[2]School of Natural and Built Environments , University of South Australia, Mawson Lakes, SA, 5095.
[3] Centre for Management of Arid Environments, Department of Agriculture, Western Australia
PO Box 483, Northam, WA, 6401

Keywords: *Vandermeer-Moloney, stage classes, matrix population models, transition probabilities*

ABSTRACT

Matrix population models are a useful tool to analyse plant and animal demography. They have a relatively simple structure and provide a straightforward tool to assess life history.

One problem however is determining the stage classes that individuals can be divided into. Assigning discrete stage classes to a continuous variable (e.g. age or density), that does not always have natural divisions, is difficult. For Australian native plants, Tiver and Andrew (1997) have defined a set of nine stage classes, from juvenile, through to senescent.

Our present data is simply in terms of physical measurements such as height and canopy size, with no possible means of estimating age and without stage classification being undertaken. The task presented in this paper is to implement a methodology for taking this raw data (in particular the canopy size) and splitting it into classes.

Vandermeer (1978), with modifications by Moloney (1986), developed a numerical approach to assign individuals to classes. This is done by minimising two possible sources of error. If the stage classes are too wide, much information will be lost because individuals will be growing within each stage class and secondly if the classes are too narrow then this can lead to problems with parameter estimation, since each new class requires a new set of parameters to specify its mortality, fecundity and state transition rates.

We apply the Vandermeer-Moloney algorithm to data on candidate species *Ptilotus obovatus* (cotton bush) and *Acacia aneura* (mulga) for a very lightly grazed grazed site at Yerilla Station in the arid shrublands of Western Australia.

We obtain not only a stage distribution, but since our data set contains values over a number of years, we have also estimated transition probabilities from stage to stage. These include transitions from a stage to a lower stage, reflecting regression through grazing or environmental effects.

Once the stage classes and transition matrices have been found we can then use them to examine the impact on the population given a range of grazing and rainfall scenarios.

The motivation for this work followed on from the research that resulted in the paper by McArthur, Boland and Tiver (2006), where it became apparent that not all our plant species had stage classes assigned to them, nor could they be easily defined. For this reason the Vandermeer-Moloney algorithm was investigated so a numerical approach could be used, to split a sample of plants into stage classes.

1 INTRODUCTION

Matrix population models are a useful tool to analyse plant and animal demography. They have a relatively simple structure and provide a straightforward tool to assess life history.

One problem however is determining the stage classes into which to divide the individuals into. Assigning discrete stage classes to a continuous variable (e.g. age or density), that does not always have natural divisions, is difficult. For Australian native plants, Tiver and Andrew (1997) have defined a set of nine stage classes, from juvenile, through to senescent.

Vandermeer (1978), with modifications by Moloney (1986), developed a numerical approach to assign individuals to classes. This is done by minimising two possible sources of error. If the stage classes are too wide, much information will be lost because individuals will be growing within each stage class and secondly if the classes are too narrow then this can lead to problems with parameter estimation, since each new class requires a new set of parameters to specify its mortality, fecundity and state transition rates.

We apply the Vandermeer-Moloney algorithm to data on candidate species *Ptilotus obovatus* (cotton bush) and *Acacia aneura* (mulga) for a very lightly grazed site.

We obtain not only a stage distribution, but since our data set contains values over a number of years, we have also estimated transition probabilities from stage to stage. These include transitions from a stage to a lower stage, reflecting regression through grazing or environmental effects.

2 SITE DESCRIPTION

We have data that was collected yearly from 1986 until 1990 as part of a study conducted by the Department of Agriculture Western Australia. The experimental site was on Yerilla Station, 150 km north of Kalgoorlie, in the arid shrublands of Western Australia (mean annual rainfall of 238 mm). The area pre-1986 had not been stocked with sheep for 15 years, for further details see Fletcher (1995).

As outlined by Fletcher (1995) the plant community consisted of mulga (*Acacia aneura*) and bowgada (*Acacia ramulosa*) with a shrub understorey of cotton bush (*Ptilotus obovatus*), flannel bush (*Solanum lasiophyllum*) and a range of poverty bushes (*Eremophila forrestii*, *E. georgei*, *E. glandulifera*, *E. latrobei*) which in places formed dense thickets. The ground storey was comprised of the perennial grasses woolly butt wanderrie (*Eragrostis eriopoda*) and broad leafed wanderrie (*Monachather poradoxa*). Annual herbs and grasses appeared in response to rainfall events.

The trial area was subdivided into seven paddocks which provided for five stocking rates of goats, one with sheep, and an area with all grazing excluded. We will be considering a paddock which was very lightly grazed (about 1 goat per 20 ha) in order to categorise the shrub populations under close to natural conditions.

2.1 Plant species

Of the species at Yerilla Station our focus will be on two species, the mulga and cotton bush.

Mulga (*Acacia aneura*) is a small tree or large shrub ranging in height from 2 to 8 m, often multi-stemmed with a spreading crown. It is suited to a range of soils including clay but needs good drainage. Mulga is a very common and a significant component of arid zone vegetation, it is found over a wide area of all Australian mainland states except Victoria. The mulga has a varied response to palatability and grazing, at Yerilla, mulga has low palatability.

Cotton bush (*Ptilotus obovatus*) is a dense, greyish shrub growing to around 60 cm high. It has multiple stems which originate at ground level. Cotton bush occurs naturally in arid inland areas of all Australian mainland states. It is a palatable species and decreases in both density and plant size under heavy grazing.

3 MATRIX POPULATION MODELS

Matrix population models provide the means of estimating growth rates and other measures associated with the viability of the population, such as extinction. The first step in formulating a matrix population model is to determine what the stage or age classes should be. In some cases this division may be natural, but often this is not the case and the investigator has to choose these classes.

The Leslie matrix model was developed by P.H. Leslie and uses mortality and fecundity rates to give a projection of an organism's population distribution based on the initial population distribution of age groups (Bruce and Shernock 2002). However, this formulation was extended to stages by Lefkovitch (1965). The Lefkovitch model is necessary because of the longevity of many Australian native plants, particularly those of the semi-arid and arid lands, which are the areas we are focussing upon.

We have data from a number of studies wherein the stage classes were estimated by the data collectors (or investigators), with the stage specifications determined according to the work of Tiver and Andrew (1997). They have defined a set of nine stage classes, from juvenile, through to senescent. However, data cannot always be put in these stage classes. The present data that will be used is in terms of canopy size, with no possible means of estimating age and without stage classification being undertaken. The task then was thus to implement a methodology for taking this raw data and splitting it into stages. The methodology we used was developed by Vandermeer (1978) and extended by Moloney (1986).

3.1 Determining stage classes

We have taken the liberty of delineating the Vandermeer-Moloney algorithm as given in Caswell (2001) in order to completely define the methodology for the ease of the readers. It considers sample populations which may have different underlying transition probabilities at different census periods or within different subpopulations. We consider a simplified version that does not consider subpopulations, see Caswell (2001).

Vandermeer (1978) initially developed this algorithm to minimise two kinds of errors. The first is the distribution error (DE), whereby if the stage classes are too wide then much information will be lost because individuals will be growing within each stage class. The second error is the sampling error (SE), if the classes are too narrow then this can lead to problems with parameter estimation, since each new class requires a new set of parameters to specify its mortality, fecundity and state transition rates. The Moloney (1986) extension to this model accounts for the differences in transition probabilities among census periods.

4 VANDERMEER-MOLONEY ALGORITHM

Assume the censuses are taken at times $t = 1, \cdots, T$. Let $m_i(t)$ be the size of individual i at time t, where $i = 1, \cdots I$ and I is the total number of individuals. Let the growth increment of individual i from time t to time $t+1$ be given by

$$d_i(t) = m_i(t+1) - m_i(t).$$

Define a size interval

$$\Omega = [M_{min}, M_{max}]$$

with a midpoint $M_{mid} = (M_{min} + M_{max})/2$. We want to evaluate the sampling error (SE) and distribution error (DE) given by the size of the interval Ω. To do this we start by defining two indicator variables

$$s_i(t) = \begin{cases} 1 & \text{if } m_i(t) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

$$r_i(t) = \begin{cases} 1 & \text{if } s_i(t) = 1 \text{ and } m_i(t+1) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

for each $i \in I$. The variable $s_i(t)$ indicates whether individual i starts in Ω, and $r_i(t)$ indicates whether individual i remain in Ω. We use these to estimate the probability P that an individual stays in Ω, such that

$$\hat{P} = \frac{\sum_t \sum_i r_i(t)}{\sum_t \sum_i s_i(t)},$$

and then let $\hat{Q} = 1 - \hat{P}$, which is the probability of an individual leaving Ω. If there was no DE, $\hat{P}$ would be independent of where in the interval Ω the individual started. So, assume that everyone starts at M_{mid}, and define the new size of the individual at time $t+1$ as

$$m_i^*(t+1) = M_{mid} + d_i(t)$$

and the new indicator variable

$$r_i^*(t) = \begin{cases} 1 & \text{if } s_i(t) = 1 \text{ and } m_i^*(t+1) \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Then we calculate new estimates $\hat{P}_{mid}(t)$ and $\hat{Q}_{mid}(t)$, where

$$\hat{P}_{mid}(t) = \frac{\sum_i r_i^*(t)}{\sum_i s_i(t)}$$

and $\hat{Q}_{mid}(t) = 1 - \hat{P}_{mid}(t)$. The estimate $\hat{P}_{mid}(t)$ is the probability of staying in the interval Ω between census period t and $t+1$.

Moloney measures DE by the mean squared proportional deviation of $\hat{P}_{mid}(t)$ from the expected value $\hat{P}$ of staying in the interval and $\hat{Q}_{mid}$ from the expected value $\hat{Q}$ of leaving the interval. This is calculated as

$$DE = \frac{1}{T-1} \sum_t \frac{1}{2} \left(\left(\frac{\hat{P}_{mid}(t) - \hat{P}}{\hat{P}} \right)^2 + \left(\frac{\hat{Q}_{mid}(t) - \hat{Q}}{\hat{Q}} \right)^2 \right). \quad (1)$$

The DE will approach a value of zero as $M_{max} - M_{min}$ approaches zero since there will be little or no difference between m^* and m. If DE was the only criteria used in choosing the category size we would always choose the smallest category possible. However, this is clearly inappropriate as the sample sizes used in estimating transition probabilities decline as category sizes are smaller, hence the estimates for P and Q will become increasingly inaccurate. The error introduced by this inaccuracy will be referred to as the SE.

To calculate SE we use a resampling method. Consider the individuals that start out in Ω; those individuals with $s_i(t) = 1$. Draw a random sample of the same number of individuals, with replacement, from this set. Define $m^{\dagger}_{ik}(t+1)$ as the size of the ith individual in the kth random sample at time $t+1$. Calculate the probabilities of staying in Ω in sample k as

$$\hat{P}^{\dagger}_{k}(t) = \frac{\sum_i r^{\dagger}_{ik}(t)}{\sum_i s_i(t)}$$

and $\hat{Q}^{\dagger}_{k}(t) = 1 - \hat{P}^{\dagger}_{k}(t)$. The larger the sample size in Ω, the more tightly the resampled estimates should cluster around the original estimates, we use $K = 200$. Moloney measures SE by the mean squared proportional deviation of the resampled and the original expected values, such that

$$\begin{aligned} SE &= \frac{1}{K(T-1)} \sum_{t=1}^{T-1} \sum_{k=1}^{K} \frac{1}{2} \left(\left(\frac{\hat{P}^{\dagger}_{k}(t) - \hat{P}}{\hat{P}} \right)^2 \right. \\ &\qquad \left. + \left(\frac{\hat{Q}^{\dagger}_{k}(t) - \hat{Q}}{\hat{Q}} \right)^2 \right). \end{aligned} \tag{2}$$

Given the above calculations of DE and SE in equations (1) and (2) Maloney suggests choosing category sizes according to the following algorithm. First fix the lower bound of the first category (at zero, or the smallest size that makes sense) and find the upper bound that minimises $SE+DE$. Fix that as the new lower bound of the second category and find the upper bound that minimises $SE+DE$ for the second category. Repeat until you reach the maximum size.

In the following subsections we will apply this algorithm to both the cotton bush and mulga.

4.1 Application of the algorithm to cotton bush

In determining the stage classes for cotton bush, data from the lightly grazed paddock at Yerilla Station was used. The census periods were taken yearly from 1986 to 1990 corresponding to $t = 1, 2, 3, 4$ and 5, prior to this period the paddock was not grazed for 15 years. Within the paddock 91 individuals were recorded and measured at each census period, giving $i = 1, 2, \cdots 96$.

Initially, for the first iteration of the Vandermeer-Moloney algorithm, we set $M_{min} = 0$, the resulting output is shown in the first graph of Figure 1. It can be observed that the smallest value of SE+DE occurs when the upper boundary is 15. Therefore, with the second iteration $M_{min} = 15$, and the resulting SE+DE is displayed in the second graph of Figure 1, where the next upper boundary is calculated to be 40.

As the Vandermeer-Moloney algorithm is repeated, the final stage classes are

(0-15]	(15-40]	(40-55]	(55-80]	>80

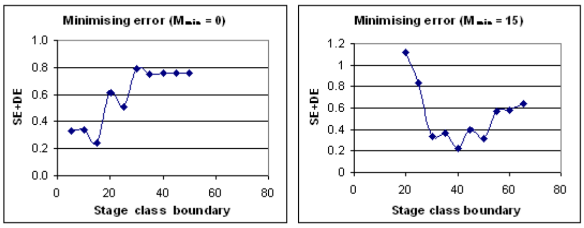

Figure 1. First two iterations of the V-M algorithm (Cotton Bush)

These stage classes can then be used to find the resulting transition matrices, giving the probability of transition from stage to stage. This is done through empirical estimation, where for example, if 20 plant species started in stage I and 10 moved to Stage II in a given time period then the transition probability from stage I to stage II would be 0.5. The transition matrix given below is read from row to column, a row may not add to one due to rounding errors.

$$\begin{array}{c|ccccc} & I & II & III & IV & V \\ \hline I & 0.67 & 0.20 & 0.14 & 0 & 0 \\ II & 0.03 & 0.75 & 0.22 & 0 & 0 \\ III & 0 & 0.15 & 0.41 & 0.43 & 0 \\ IV & 0 & 0 & 0.36 & 0.5 & 0.14 \\ V & 0 & 0 & 0.17 & 0.08 & 0.75 \end{array}$$

From the transition matrix it can be seen that cotton bush can sometimes move either backwards or forwards more than one stage in a given time period, as would happen due to grazing or dry seasonal conditions.

4.2 Application of the algorithm to mulga

At Yerilla Station, the mulga plants were measured yearly from 1986 to 1990. We used the canopy size to determine the stage classes, which ranged from 5 cm to 210 cm in diameter. Using the parameters $I = 1, 2, \cdots 156$, $t = 1, 2, 3 \cdots, 5$, $K = 200$ and an initial $M_{min} = 0$ the first iteration of the Vandermeer-Moloney algorithm is shown in the first graph of Figure 2, where SE+DE is plotted against M_{max} (the upper stage boundary). It can be seen that the minimum SE+DE occurs when $M_{max} = 20$. This will be used as M_{min} in the next iteration.
The second plot in Figure 2 shows SE+DE, when $M_{min} = 20$, it can be seen that the next upper boundary will be 35.

As the Vandermeer-Moloney algorithm is repeated, the final stage classes are

(0-20]	(20-35]	(35-60]	(60-75]	(75-90]
(90-105]	(105-115]	(115-135]	(135-160]	>160

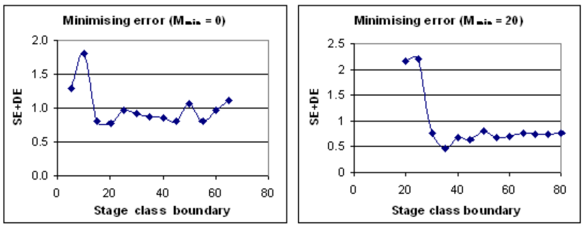

Figure 2. First two iterations of the V-M algorithm (Mulga)

These stage classes ($I, II, III, \cdots, IX, X$) can then be used to find the resulting transition matrix (a matrix row may not add to one due to rounding errors)

$$\begin{bmatrix} 0.79 & 0.03 & 0.17 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.04 & 0.58 & 0.38 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 0.78 & 0.22 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.07 & 0.59 & 0.34 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.03 & 0.61 & 0.36 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.50 & 0.50 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.13 & 0.43 & 0.43 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.08 & 0.62 & 0.31 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.68 & 0.26 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

4.3 Discussion of results and notes of caution

Two features of the resulting stages in the chosen species should be emphasised. One is that there is a significantly lower number of discernable stages in the cotton bush, but this is to be expected since cotton bush is a smaller shrub. The stages for cotton bush are similar to those that would be used if expert knowledge alone was being used.

The other feature is that the number of stages we found for mulga is 10, although this may seem high the number of breakdowns seem reasonable and are not dissimilar to the stage class definition of 9 given in Tiver and Andrew (1997). Moloney's original paper demonstrated some small variation in the number of classes with different samples and our work may well have demonstrated a similar situation.

By using the Vandermeer-Moloney algorithm we now have the capabilities to work with the model by McArthur, Boland and Tiver (2006), such that we will use the transition matrix to see the effects of grazing on Australian native plants, in particular arid rangeland species across Australia. Without this numerical approach our data could not be grouped into classes, and the transitions impossible to calculate. In addition applications can be used to examine a range of grazing management and climate scenarios such as those in Watson, McKeon and Wilcox (2004).

Caswell (2001) does mention that some caution should be used when applying the Vandermeer-Moloney algorithm. These are given by the following three reasons

1. The algorithm considers only if an individual remains or leaves the category Ω, it does not consider where the individual goes.
2. Temporal variation in the vital rates, if it does exist, appears as variability in $\hat{P}_{mid}(t)$ and $\hat{P}_k^{\dagger}(t)$, but not in $\hat{P}$. Thus temporal variance contributes to both SE and DE, although it is neither distribution error or sampling error.
3. Defining size class boundaries progressively from smallest to largest is not guaranteed to minimise SE+DE over all possible choices of category sizes.

We have taken the above three points into consideration, and although valid we do not think that they are enough to outweigh the reasons of why we are using the algorithm in the first place. It is unclear at this time how restrictive these points are and whether they lead to significant inaccuracies, but it is an area for further research.

5 CONCLUSIONS

The motivation for this work is the incorporation of the resulting transition probabilities from stage to stage in a dynamical population matrix model of the type delineated in McArthur, Boland and Tiver (2006). The goal of the modelling is to investigate management options to ensure the long-term survivability of plant species in the rangeland areas of Australia. The raw data needs to be transformed so that it is expressible in this form for the model to operate.

The application of the Vandermeer-Moloney algorithm allows stage classes of either plants or animals to be determined via a numerical approach. This algorithm is particularly useful when there is no natural division into stages.

Another benefit to the Vandermeer-Moloney algorithm is its accountability to handle different census periods having different transition matrices. Our data from Yerilla Station having this structure consisted of 5 census periods, so the algorithm suited our purposes.

Once the stage classes and transition matrices have been found we can then use them to examine the viability of the population using theory from matrix population models.

6 ACKNOWLEDGEMENT

This research has been done with the input and support of all our Desert Knowledge CRC project team. Also thanks to Wayne Fletcher and Phil Thomas for use of their unpublished data.

7 REFERENCES

Bruce, M and E. Shernock (2002), Stage Based Population Projection Matrices and Their Biological Applications, *http://online.redwoods.cc.ca.us/instruct/darnold/laproj/Fall2002/MikeEm/paper.pdf*

Caswell, H. (2001), Matrix Population Models (2nd Edition), Sinauer Associates, Inc. Publishers.

Fletcher, W. (1995), Yerilla Goat Grazing Study:-1989-1990, *Miscellaneous publication*, **20/95**, Department of Agriculture Western Australia, South Perth.

Lefkovitch, L.P. (1965), The Study of Population Growth in Organisms Grouped by Stages, *Biometrics* **21/1**.

McArthur, L., J. Boland, and F. Tiver (2006-in press), The effects of grazing on *Myoporum platycarpum* in the arid rangelands of South Australia, *Ecological Modelling*.

Moloney, K. (1986) A Generalized Alorithm for Determining Category Size, *Oecologia*, **69**, 176-180.

Tiver, F. and M.H. Andrew (1997), Relative Effects of Herbivory by Sheep, Rabbits, Goats and Kangaroos on Recruitment and Regeneration of Shrubs and Trees in Eastern South Australia, *Journal of Applied Ecology*, **34**, 903-914.

Watson, I.W., G.M. McKeon and D.G. Wilcox (2004), Modelling climate and management effects on shrub populations in the Gascoyne area of Western Australia and the North East District of South Australia, Book chapter in: *Pasture degradation and recovery in Australia's rangelands* (Editors: G.M. McKeon, W. Hall, B.H. Henry, G. Stone and I.W. Watson). Published by Queensland Department of Natural Resources, Mines and Energy, Brisbane, Australia, 174-190.

Vandermeer, J. (1978), Choosing Category Size in a Stage Projection Matrix, *Ocecologia*, **32**, 79-84.