

Statistical Models for wildlife contact networks – confronting theory with data

Shrupa Shah

*Supervised by: Dr. Stephen Davis and Dr. Babak Abbasi
RMIT University*



© Warren Photographic

1

¹ <http://www.warrenphotographic.co.uk/photography/bigs/26631-Field-Vole-white-background.jpg>

Table of Contents

| | |
|--|----|
| 1. Introduction..... | 3 |
| 2. Methodology | 4 |
| 2.1 How the networks were formed..... | 4 |
| 2.2 Model development..... | 6 |
| 2.3 Different types of models..... | 6 |
| 2.4 Estimation of parameters | 8 |
| 3. Results | 10 |
| 3.1 How the different groups were formed | 10 |
| 3.2 How competing models were compared..... | 11 |
| 3.3 Discussion | 12 |
| 4. Conclusion | 14 |
| Acknowledgement | 15 |

The rate of contact plays a significant role in modeling the spread of infectious diseases. This report describes how spatial constraints and individual heterogeneity were incorporated to model the rate of contact of field voles (*Microtus agrestis*). Voles contact networks were formed from mark recapture data collected over 7 years. Every vole trapped represents a node in the network. And since edges were not directly observed, they were inferred from trap sharing. Incorporating the spatial constraints into a simple rate of contact model, which assumes every vole has an equal chance of having contact with every other vole, provides sufficient evidence to suggest an improvement over the random mixing of hosts. Individual characteristics such as mass and gender were investigated to incorporate the individual heterogeneity of voles to have contact with each other into the models. Upon conducting mass versus degree analyses for different networks, there is strong evidence to support the grouping of voles into three classes; BIG male, female, SMALL male. It is found that male voles with higher degree tend to have higher mass and the male voles with lower degree tend to have lower mass. A significant drop in the AIC values was observed for networks classified into the breeding season when both the spatial constraints and individual heterogeneity were combined to model the rate of contact. There is significant supporting evidence that the underlying structure of contact networks is far more complex than the simple random mixing of hosts that most models for the rate of contact assume.

1. Introduction

Mathematical models for infectious disease typically use contact rates (e.g. the number of other voles that a vole encounters over a specified time interval) as one of their main elements in predicting the outcomes of an epidemic. However, few studies have been conducted to determine the patterns of contact in wildlife populations.

This project aims to investigate the importance of spatial constraints and individual heterogeneity on the rate of contact of voles. Spatial constraints are simply the importance that spatial distance plays in the rate of contact. Individual heterogeneity is taking into account characteristics such as mass and sex in determining the rate of contact.

If these dominant hubs, with specific characteristics, can be targeted for vaccination or removal then an epidemic might be brought under control faster or an outbreak prevented.

2. Methodology

2.1 How the networks were formed

Before moving on to talk about models that were fitted to the data to model both the spatial constraints and the individual heterogeneity, this section briefly explains how the networks were formed.

The models were fitted to mark recapture data collected for 7 years from 4 sites. For each of the 4 sites there were 64 primary trapping sessions from May 2001 to March 2007 approximately a month apart. Each primary session had 5 secondary trapping sessions. The data for two consecutive trapping sessions were combined to form a network. The network is formed such that each vole trapped is a node in the network. Contacts are not directly observed and are inferred from trap sharing. An edge is formed between two voles if, and only if, they are caught in the same trap at least once. The spatial location of each vole was calculated as the average position of the traps it was caught in.

During each primary session, 100 traps were laid out, 5 m apart, as shown in Figure 1. For instance, if a vole V1 was caught in A4, B4, C6 and E7 during either of the two consecutive trapping sessions, then its spatial location would be, the average of all the traps it was caught in, x-coordinate=2.75 and y-coordinate=5.25. Similarly, if another vole V2 was caught in H6, H8 and I7 then its spatial location would have x-coordinate=8.33 and y-coordinate=7.

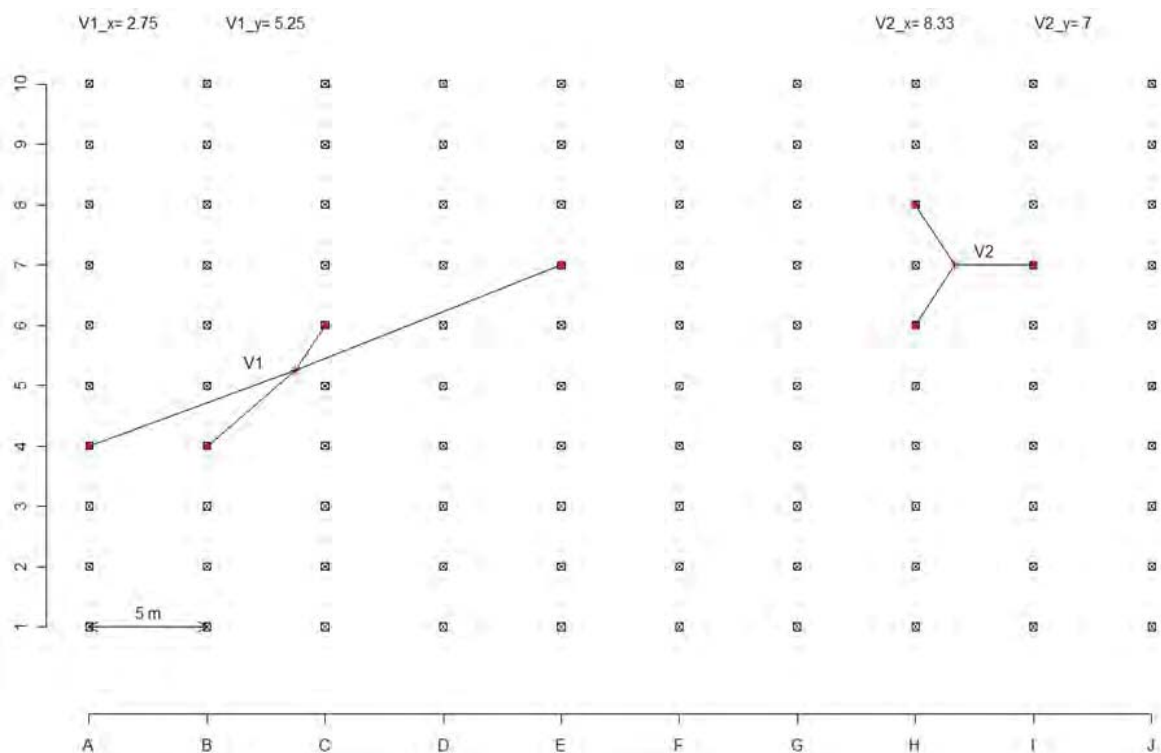


Figure 1 - A visualisation of the trap layout, illustrating how spatial locations for individual voles were derived from averaging the respective trap locations for individual voles caught during either of the two consecutive trapping sessions.

Suppose another vole V3, was caught in E7, F9 and H6 (refer to Figure 2), then its spatial location would be x-coordinate=6.33 and y-coordinate=7.33. An edge would be inferred between V1 and V3 because they were both caught in E7. Similarly, an edge would be inferred between V3 and V2 because they were both caught in H6. Therefore, this simple network would have three nodes and two edges. And each node will have an x-coordinate and y-coordinate.

$$N = \{V_1, V_2, V_3\}$$

$$L = \{e_1, e_2\}$$

$$f = \{e_1: V_1 \sim V_3, e_2: V_2 \sim V_3\}$$

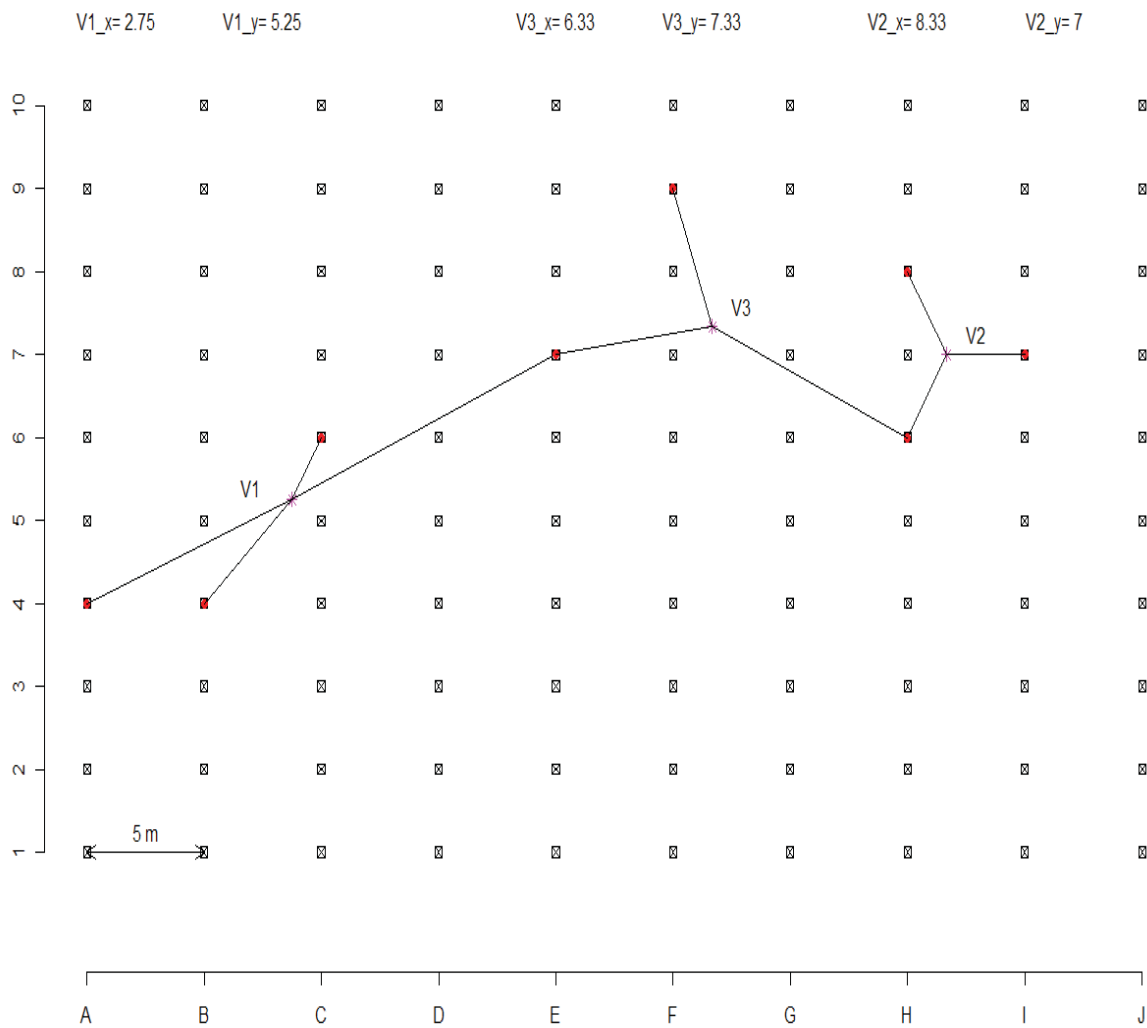


Figure 2 - A visualisation of how edges are inferred between voles caught in the same trap at least once during either of the two consecutive trapping session.

2.2 Model development

Let the number of contacts between vole i and vole j , follow a Poisson distribution with intensity k_{ij} , over some finite time interval.

i.e.

$$\Pr(X = x) = \frac{e^{-k_{ij}}(k_{ij})^x}{x!}$$

Therefore, the probability of observing no contacts between vole i and vole j is when $x = 0$,

$$\Pr(X = 0) = \frac{e^{-k_{ij}}(k_{ij})^0}{0!} = e^{-k_{ij}}$$

Thus the probability of observing at least one contact between vole i and vole j and therefore an edge existing in the observed network is:

$$p_{ij} = 1 - \Pr(X = 0) = 1 - e^{-k_{ij}}$$

2.3 Different types of models

- Two parameter model

Firstly, the following model was considered where the rate of contact only depended on the spatial constraints

$$k_{ij} = ce^{-\lambda s_{ij}}$$

i.e. k_{ij} is the rate of contact, s_{ij} is the Euclidean distance between vole i and vole j , and c and λ are constants to be estimated. The magnitude of λ determines the scale over which the spatial constraints operate.

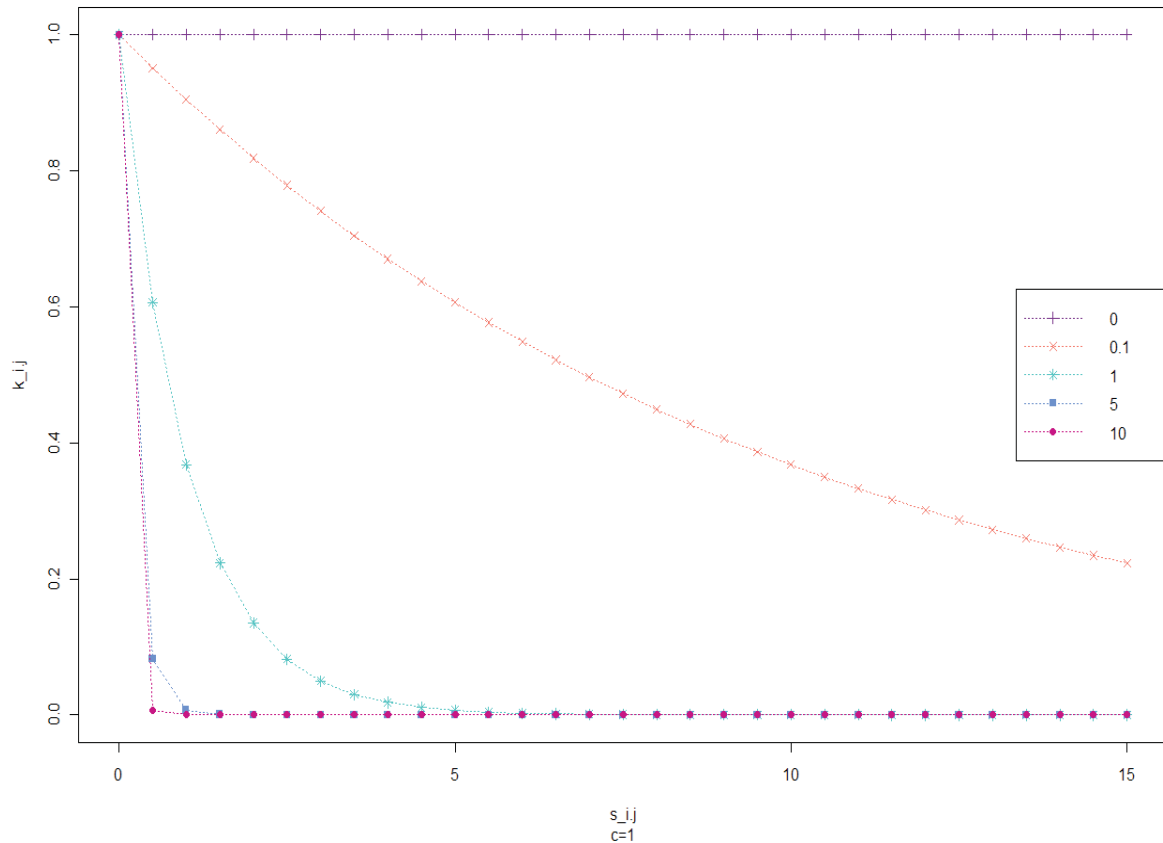


Figure 3 - Illustration of how varying the distance, s_{ij} , and the parameter, λ , can affect the rate of contact, k_{ij} . Assuming $c = 1$.

Suppose the distance, s_{ij} , between two voles is taken on the interval $[0:1:15]$, and lambda is allowed to take values $\{0, 0.1, 1, 5, 10\}$ as shown in Figure 3. For example when $\lambda=0.1$ the rate of contact declines exponentially as the distance between vole i and vole j increases. When $\lambda=1$ the rate of contact declines much faster as we move right along the x-axis than when $\lambda=0.1$. Therefore, for larger values of λ the decline in the rate of contact is much faster such that the rate of contact asymptotically approaches zero for smaller values of s_{ij} .

On the other hand, when $\lambda=0$ the rate of contact, k_{ij} , is constant and does not depend on distance. Therefore, every vole is equally likely to have contact with every other vole and the model resembles the rate of contact model that assumes random mixing of hosts.

This project mainly concentrated on fitting models that take into account the individual heterogeneity of voles. Four different models were fitted, where the contact rate depended on both spatial constraints and individual characteristics such as gender and mass. Refer to “3.1 How the different groups were formed” for a demonstration of how the different classes that (x_i, x_j) can belong to are derived.

- Five parameter model - Model 1

$$k_{ij} = ce^{-\lambda s_{ij}} x_i x_j \quad (x_i, x_j) \in \{M, F, m\}$$

- Five parameter model - Model 2

$$k_{ij} = ce^{-\lambda s_{ij}} (x_i + x_j) \quad (x_i, x_j) \in \{M, F, m\}$$

- Five parameter model - Model 3

$$k_{ij} = ce^{-\lambda s_{ij}} (x_i + x_j) x_i x_j \quad (x_i, x_j) \in \{M, F, m\}$$

where x_i , is a measure of eagerness of vole i , to have contact with another vole. Similarly x_j , is a measure of eagerness of vole j , to have contact with another vole.

- Eight parameter model

$$k_{ij} = ce^{-\lambda s_{ij}} x_{ij} \quad (x_{ij}) \in \{MM, MF, Mm, FF, Fm, mm\}$$

where x_{ij} , is a measure of eagerness of vole i , to have contact with another vole j , such that:

$$(x_{ij}) \in \{MM, MF, Mm, FF, Fm, mm\}$$

2.4 Estimation of parameters

$$\Omega = \{(i, j) | a_{ij} = 1\} \text{ and } \Omega' = \{(i, j) | a_{ij} = 0\}$$

There will be a total of $n(n - 1)$ terms in the likelihood function.

The first product will comprise of m terms.

Whereas, the second product will comprise of $n(n - 1) - m$ terms.

- Two parameter model

$$p(i, j) = (1 - e^{-ce^{-\lambda s_{ij}}}) (a_{ij}) + (e^{-ce^{-\lambda s_{ij}}}) (1 - a_{ij})$$

$$\forall (i, j) \in n, j > i$$

$$L(\lambda, c | p(i_1, j_2), p(i_1, j_3), \dots, p(i_{n-2}, j_n), p(i_{n-1}, j_n)) = \prod_{(i,j) \in \Omega} (1 - e^{-ce^{-\lambda s_{ij}}}) \prod_{(i,j) \in \Omega'} (e^{-ce^{-\lambda s_{ij}}})$$

The values for c and λ were estimated using maximum likelihood estimation.

- Five parameter model - Model 1

$$p(i, j) = \left(1 - e^{-ce^{-\lambda s_{ij} x_i x_j}}\right) (a_{ij}) + \left(e^{-ce^{-\lambda s_{ij} x_i x_j}}\right) (1 - a_{ij})$$

$$\forall (i, j) \in n, j > i$$

$$L(\lambda, c, M, F, m | p(i_1, j_2), p(i_1, j_3), \dots, p(i_{n-2}, j_n), p(i_{n-1}, j_n)) = \prod_{(i,j) \in \Omega} \left(1 - e^{-ce^{-\lambda s_{ij} x_i x_j}}\right) \prod_{(i,j) \in \Omega'} \left(e^{-ce^{-\lambda s_{ij} x_i x_j}}\right)$$

- Five parameter model - Model 2

$$p(i, j) = \left(1 - e^{-ce^{-\lambda s_{ij} (x_i + x_j)}}\right) (a_{ij}) + \left(e^{-ce^{-\lambda s_{ij} (x_i + x_j)}}\right) (1 - a_{ij})$$

$$\forall (i, j) \in n, j > i$$

$$L(\lambda, c, M, F, m | p(i_1, j_2), p(i_1, j_3), \dots, p(i_{n-2}, j_n), p(i_{n-1}, j_n)) = \prod_{(i,j) \in \Omega} \left(1 - e^{-ce^{-\lambda s_{ij} (x_i + x_j)}}\right) \prod_{(i,j) \in \Omega'} \left(e^{-ce^{-\lambda s_{ij} (x_i + x_j)}}\right)$$

- Five parameter model - Model 3

$$p(i, j) = \left(1 - e^{-ce^{-\lambda s_{ij} (x_i + x_j) x_i x_j}}\right) (a_{ij}) + \left(e^{-ce^{-\lambda s_{ij} (x_i + x_j) x_i x_j}}\right) (1 - a_{ij})$$

$$\forall (i, j) \in n, j > i$$

$$L(\lambda, c, M, F, m | p(i_1, j_2), p(i_1, j_3), \dots, p(i_{n-2}, j_n), p(i_{n-1}, j_n)) = \prod_{(i,j) \in \Omega} \left(1 - e^{-ce^{-\lambda s_{ij} (x_i + x_j) x_i x_j}}\right) \prod_{(i,j) \in \Omega'} \left(e^{-ce^{-\lambda s_{ij} (x_i + x_j) x_i x_j}}\right)$$

The values for c , λ , M , F and m were estimated using maximum likelihood estimation.

- Eight parameter model

$$p(i, j) = \left(1 - e^{-ce^{-\lambda s_{ij} x_{ij}}}\right) (a_{ij}) + \left(e^{-ce^{-\lambda s_{ij} x_{ij}}}\right) (1 - a_{ij})$$

$$\forall (i, j) \in n, j > i$$

$$L(\lambda, c, MM, MF, Mm, FF, Fm, mm | p(i_1, j_2), p(i_1, j_3), \dots, p(i_{n-2}, j_n), p(i_{n-1}, j_n)) = \prod_{(i,j) \in \Omega} \left(1 - e^{-ce^{-\lambda s_{ij} x_{ij}}}\right) \prod_{(i,j) \in \Omega'} \left(e^{-ce^{-\lambda s_{ij} x_{ij}}}\right)$$

The values for c , λ , MM , MF , Mm , FF , Fm , and mm were estimated using maximum likelihood estimation.

3. Results

3.1 How the different groups were formed

The two parameter model only takes into account the spatial constraints. Therefore, to incorporate the individual heterogeneity of voles to have contact with each other's individual characteristics such as mass and gender were investigated. Figure 4 shows a plot of mass versus degree for network 24, formed by combining data from trapping sessions 47 and 48. The blue triangles represent the male voles and the red dots represent the female voles. From this graph it is evident that male voles with lower mass tend to have lower degree and male voles with higher mass tend to have higher degree. Sometimes, there are outlier cases like the blue triangle circled in black, which has low mass and high degree. This distinction helped classify the male voles into two groups; those with mass less than 25 were classified into the SMALL MALE group whereas those with mass greater than or equal to 25 were classified into the BIG MALE group.

On the other hand, no such clear distinction is evident in the female voles. Therefore, the female voles were classified into a separate group but did not subdivide this group further on the basis of mass. Yet on average the females had lower contact rate than the males as shown in Figure 4.

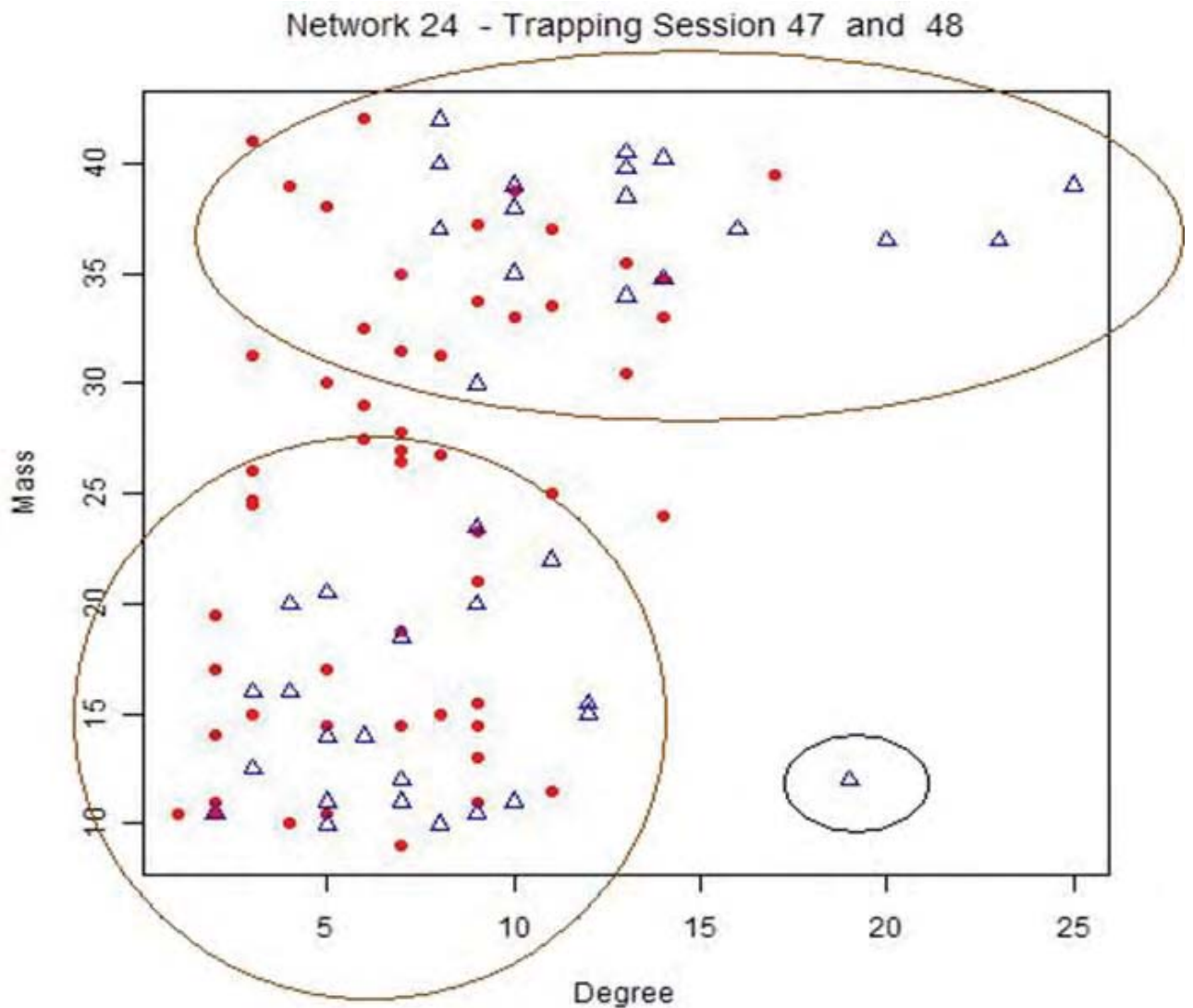


Figure 4 - Mass versus degree plot for individual voles in Network 24, caught in either trapping sessions 47 and/or 48. The blues triangles represent the male voles whereas the red dots represent the female voles.

3.2 How competing models were compared

The Akeike's Information Criterion (AIC) is used to rank competing models. AIC penalizes models for too many parameters. A model is considered to have significantly improved if the AIC values drop by 2 or more.

Furthermore, the networks formed were classified into breeding, non-breeding and 'in-between' season. A network was classified into the breeding season if the trappings occurred between April and October. Otherwise it was classified into the non-breeding season. If trappings occurred in the months of, April and October, they were neither classified into the breeding season nor the non-breeding season, but into the 'in-between' season instead.

Therefore, the choice of the model and hence the number of parameters was based primarily on the minimization of the AIC statistics.

3.3 Discussion

The maximum likelihood estimates of c and λ or c, λ, M, F and m or $c, \lambda, MM, MF, Mm, FF, Fm,$ and mm are the values that minimize the AIC for the two parameter model, five parameter models or eight parameter model respectively, because the estimates maximize the log likelihood for the model. Also larger numbers of parameters in the model do not result in a better fitted model, because estimation of too large a number of parameters introduces estimation errors that adversely affect the use of the fitted model.

Figure 5a) shows the change in the AIC values for the 3 different five parameter models fitted to the data. This is relative to the two parameter model for the networks that were formed from the trappings that occurred in the breeding season.

The drop in the AIC values for the five parameter models relative to the two parameter model is significant for most of the networks during the breeding season. Furthermore, for those networks where the AIC value drops, on most occasions, the five parameter (Model 3) appears to outperform the other two five parameter models.

Contrary to the above, there is no significant improvement in the AIC values for the eight parameter model for the networks that were formed from the trappings that occurred in the breeding season.

Figure 5b) shows the change in the AIC values for the various models fitted to the data, relative to the two parameter model for the networks that were formed from the trappings that occurred in the non-breeding season. There was no significant improvement in the AIC values from the two parameter model for the non-breeding season, except once when the five parameter (Model 3) performed better relative to the two parameter model in November.

Similar to the non-breeding season, there was not much improvement in the AIC values over the two parameter model for the 'in-between' season.

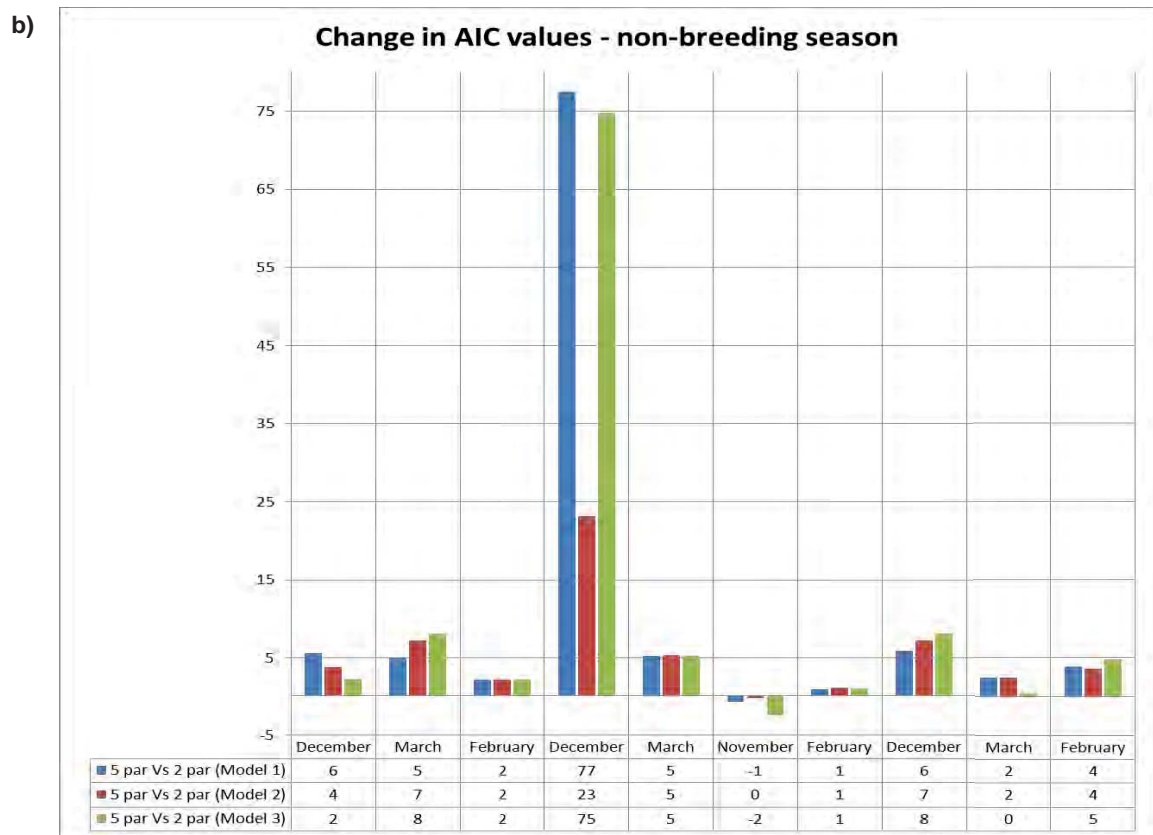
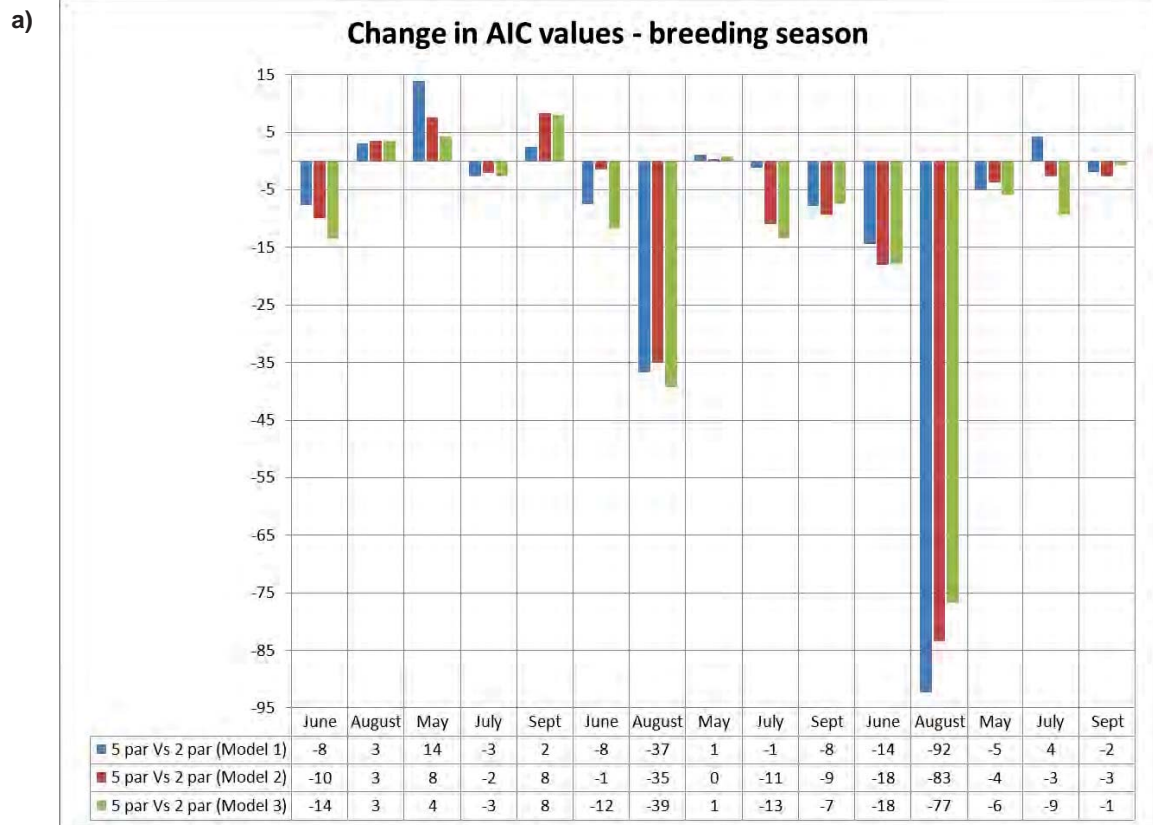


Figure 5 - change in the AIC values, for the 3 different five parameter models fitted to the data, relative to the two parameter model, for the networks that were formed from the trappings that occurred in the a) breeding season and b) non - breeding season.

4. Conclusion

The aim of incorporating spatial constraints and individual heterogeneity is to better understand the contact networks on which infectious disease spreads. Most rate of contact models assume random mixing in the contact process of wildlife. However, significant supporting evidence suggests that the underlying structure of contact networks is far more complex than the simple random mixing of hosts' model for the rate of contact. Such models will likely make wrong or misleading predictions and hence may hamper management of wildlife disease rather than improve it. If the node is for instance a Big Male vole, is it more mobile than the rest of the nodes in the network.

There is significant evidence that by incorporating the spatial constraints and individual heterogeneity into the model to obtain the rate of contact, there is an improvement to the constant rate of contact.

In the future, it may be advantageous to apply the same models to the remaining sites to obtain affirmation that models fit well. The study of the reproductive status of voles may also provide further improvement. In addition, studying the assortativity of the networks will enable us to understand how well connected the hubs are in the network.

Acknowledgement

I wish to thank AMSI for funding this project. I also wish to thank CSIRO for organising the Big Day In and arranging my accommodation in Sydney in conjunction with AMSI. My at most gratitude goes to the best supervisors Dr. Stephen Davis and Dr. Babak Abbasi for their continued support, help, advise and encouragement throughout my project. I extend my gratitude to Dr. Mike Begon and Dr. Sandra Telfer for the data I used in my project and their expert advice. Last but not least RMIT University for providing this wonderful opportunity.

