

https://dspace.org/

A hierarchical state space model of greater Kudu (Tragelaphus Strepsiceros) population dynamics in the Kruger National Park

Gatawa, Tatenda

2012-01-16
http://hdl.handle.net/10539/10981

Downloaded from a DSpace repository. https://dspace.org/

A HIERARCHICAL STATE SPACE MODEL OF GREATER KUDU (*TRAGELAPHUS STREPSICEROS*) POPULATION DYNAMICS IN THE KRUGER NATIONAL PARK

A Research Report Submitted to the Higher Degrees Committee, University of the Witwatersrand, Johannesburg, in Partial Fulfillment of the Requirements for the Degree of Master of Science.

By

Tatenda Gatawa

Supervisor: Professor Jacqueline Galpin, School of Statistics and Actuarial Science

Co- Supervisor: Professor Norman Owen-Smith, School of Animal Plant and Environmental Sciences

20 October 2010

DECLARATION

I, Tatenda Bernard Machona Gatawa, hereby declare that this Research Report is my own work and that it has not been submitted for examination for the award of a degree at any other university.

Signed: ……..

TATENDA BERNARD MACHONA GATAWA

ABSTRACT

Abundance counts of ungulate species which are carried out using an aerial census are susceptible to measurement errors. These measurement errors result from environmental factors such as vegetation cover and the resultant effect that factors, such as the annual rainfall pattern, have on the accuracy of the count. Given these measurement errors, models of population abundance that are based on these counts should not only account for the population dynamic process - they need to model the measurement error process simultaneously in order to produce plausible estimates. In addition to this, the recorded counts from these censuses do not give a breakdown of the total count into the different gender and life-history stages of the animals counted. This research report investigates the usefulness of a hierarchical Bayesian hidden process modelling approach at explicitly including the measurement error process and a sub-model for the gender and life-history stage of the animals counted into a population dynamics model based on the aerial census counts. The data used are aerial counts of Kudu in the Kruger National Park from 1983 to 1993. The result is a model which estimates the measurement error in each year of the census and also breaks down the overall count into the numbers in each gender and life-history stage. The usefulness of the model is evaluated based on statistical model diagnostics.

ACKNOWLEDGEMENTS

I am grateful to many people whose assistance has enabled me to carry out this research. I would like to make special mention of my supervisor, Professor JS Galpin, for answering all my questions, providing continual guidance and above all for her patience. I am also grateful to Professor Norman Owen-Smith, whose dedication to the study of ungulates in the Kruger National Park highlighted the research question, for providing the support which made it possible to carry out this study.

DEDICATION

For my parents, Judith and Bernard Gatawa, who gave me unwavering support and provided me with countless opportunities to grow throughout my studies; and my darling wife, Polinah Gatawa, who encouraged me throughout and helped me to bring this research report to completion.

Table of Contents

Chapter 1 ..... **1**

Introduction **1**

1.1 Social Organisation and Distribution of Kudu Populations *3*

1.2 Kudu abundance in the KNP *3*

1.3 Modelling Population Abundance *4*

1.4 Aim and objectives and limitations of the report *7*

1.5 Data *8*

1.6 Organisation of the report *9*

Chapter 2 ..... **10**

Literature Review **10**

2.1 Population Abundance Dynamics *10*

2.2 Sources of Uncertainty *13*

2.3 Models for Population Abundance *15*

2.4 Summary of the literature review *20*

Chapter 3 ..... **22**

Methodology **22**

3.1 Definition of SSMs *22*

3.1.1 Representation of the State and Observation Processes *25*

3.2 Bayesian modelling *26*

3.2.1 Brief Overview of Bayesian Inference *27*

3.2.2 Hierarchical Bayes Models *28*

3.3 Markov Chain Monte Carlo Techniques and Bayesian Inference *29*

3.3.1 Bayesian Computation *29*

3.3.1.1 The Gibbs Sampler *30*

3.3.2 Bayesian Inference for Hidden Process Models *32*

3.4 The WinBUGS modelling Software and Model Output *33*

3.5 Model Diagnostics *34*

Chapter 4 ..... **37**

Results **37**

4.1 Model Definition *37*

4.1.1 The State Process *37*
4.1.1.1 Defining Priors *37*
4.1.1.2 Hyperdistribution pdfs (Model Covariates) *39*
4.1.1.3 Initial Values *41*
4.1.1.4 Prior pdfs for parameters *42*
4.1.1.5 Observation Process pdf *43*
4.2 The Estimated Total Population of Kudu in The KNP *43*
4.2.1 Running Quantile Plots Total Population of Kudu in The KNP *47*
4.2.2 Kernel Density Estimates Marginal Posterior Distribution Function of Total Population *48*
4.2.3 Autocorrelation Functions Estimated Total Population *49*
4.2.4 Estimated Bias in Total Census Counts *49*
4.3 Estimated Number of Juveniles *51*
4.3.1 Running Quantiles Estimated Number of Juveniles *51*
4.3.2 Kernel Density Estimates Number of Juveniles *51*
4.3.3 Autocorrelation Functions Number of Juveniles *52*
4.4 Estimated Juvenile Survival Rate *53*
4.4.1 Running Quantiles Estimated Juvenile Survival Rate *53*
4.4.2 Kernel Density Estimate: Marginal Posterior Distribution Function of Juvenile Survival Rate *53*
4.4.3 Autocorrelation Function of Juvenile Survival Rate *54*
4.5 The Estimated Number of Male Yearlings *55*
4.5.1 Running Quantiles Number of Male Yearlings *55*
4.5.2 Kernel Density Estimate: Marginal Posterior Distribution Number of Male Yearlings *55*
4.5.3 Autocorrelation Function Number of Male Yearlings *56*
4.6 Estimated Male Yearling Survival Rate *57*
4.6.1 Running Quantiles Male Yearling Survival Rate *57*
4.6.2 Kernel Density Estimates Male Yearling Survival Rate *57*
4.6.3 Autocorrelation Functions of Male Yearling Survival Rate *58*
4.7 Estimated Number of Female Yearlings *59*
4.7.1 Running Quantiles of Number of Female Yearlings *59*
4.7.2 Kernel Density Function: Marginal Posterior Distribution Number of Female Yearlings *59*
4.7.3 Autocorrelation Functions of Number of Female Yearlings *60*

4.8 Estimated Female Yearling Survival Rate 61
4.8.1 Running Quantiles Female Yearling Survival rate 61
4.8.2 Kernel Density Estimates Female Yearling Survival RateS 61
4.8.3 Autocorrelation Functions of Female Yearling Survival Rate 62
4.9 Estimated Number of Male Adults 63
4.9.1 Running Quantiles Number of Male Adults 63
4.9.2 Kernel Density Estimate: Marginal Posterior Distribution Number of Adult Males 63
4.9.3 Autocorrelation Functions Number of Male Adults 64
4.10 Estimated Adult Male Survival Rate 64
4.10.1 Running Quantiles Adult Male Survival Rate 64
4.10.2 Kernel Density Estimate: Adult Male Survival Rate 65
4.10.3 Autocorrelation Functions Adult Male Survival Rate 66
4.11 Estimated Number of Adult Females 66
4.11.1 Running Quantiles Number of Adult Females 66
4.11.2 Kernel Density Estimates: Number of Adult Females 67
4.11.3 Autocorrelation Function Number of Adult Females 68

Chapter 5 69

Discussion of Results, Conclusion and Recommendations 69

5.1 Discussion of Results 69
5.2 Conclusion and recommendations 72

References 74

Appendix A: WinBUGS Code 82

Appendix B: Running Quantile Plots 85

Appendix C: Kernel Density Estimates for Marginal Posterior Distribution Functions 108

Appendix D: Autocorrelation Functions 132

Appendix E: Summary Statistics (Final Model – Model 2) 143

List of Tables

Table 1: EAS count and rainfall figures 1983-1993 ..... *9*

Table 2: Estimated regression models for Kudu survival by Life Stage (Owen –Smith et al, 2005) *40*

Table 3: Initial Values of Model Parameters *41*

Table 4: Initial values- juveniles (model 2) *46*

Table 5: Initial values – female yearlings (model 2) *46*

Table 6: Initial Values of the Adult Female Survival Rate - Model 2 ..... *46*

Table 7: Difference in estimated EAS Counts Model 1 and Model 2 ..... *50*

Table 8: Estimated Bias in EAS Counts *50*

List of Figures

Figure 1: Map of The KNP *4*

Figure 2: Estimated Survey Error *70*

Figure 3: Estimated Survival Rates for the different Gender and Life-History Stages *71*

Chapter 1

Introduction

Population abundance is the variable of interest in most ecological models. The majority of studies, and models built, focus on factors involved in population dynamics and biological interactions which are sensitive to the number of animals present. The major population dynamics in ecological models include birth, progression to the different age classifications and death. Without precise knowledge of the size and spatial distribution of the population it is not possible to study such density-dependent relationships (Williams, Nicholas & Conroy, 2002). Abundance estimates are also used in measuring the performance of population models. Abundance is further emphasized in line with the goals of management, which concentrate on population size. It is, typically, of interest to increase the population size for species which are rare, or seen to be advantageous; whilst efforts must be made to control population growth of pestilent species. Lastly, good estimates of population abundance are essential for the evaluation of the various conservation, wildlife, fisheries and pest management programs (Williams *et al*, 2002). As a result researchers and other users of population abundance counts rely heavily on the utilisation of appropriate statistical sampling and census procedures to ensure the accuracy of the resultant abundance counts.

Estimates of the number of animals in an area are needed for understanding the species dynamics and for management. An aerial survey is the most feasible means by which large animals inhabiting an extensive area can be counted (Seber, 1982). It is recognised that the resultant abundance counts are typically imprecise and usually biased downwards (Jachmann, 2002). A number of approaches to improve the

precision of estimates of the true numbers of animals from the raw counts are discussed in the literature. It is often found that these approaches result in estimates which are implausible against the background of previous data and knowledge of population dynamics (Shenk, White & Burnham, 1998). Several studies (Maunder, Skaug & Hoyle, 2008; Newman & Lindley, 2006; Newman, Buckland, Lindley, Thomas & Fernandez, 2006; Clark & Bjørnstad, 2004) have investigated the effectiveness of addressing this issue via modelling by including covariates in the model that can be used to ensure plausible results.

This research report has two aims:

1. To model the population dynamics of the Greater Kudu (*Tragelaphus strepsiceros*) species in the Kruger National Park (KNP) in order to correct for possible measurement errors in the census count data, and
2. To reconcile the overall census counts to a model of population structure which allows us to estimate the number of animals and the yearly survival rates in each gender and age classification. The total census count does not provide this information.

The model is based on a time series of aerial survey data and predictor variables. Modelling requires corrections taking into account measurement error in aerial survey data, as well as the specification of a model to describe the overall system (which includes gender and life- history stage survival rates). It is of interest to evaluate the properties of the model built in terms of the plausibility of assumptions made and the statistical model diagnostics. This research report will investigate a modelling approach that allows for the explicit modelling of measurement errors and the inclusion of sub-models for the population dynamics within a single model for the total census counts.

1.1 Social Organisation and Distribution of Kudu Populations

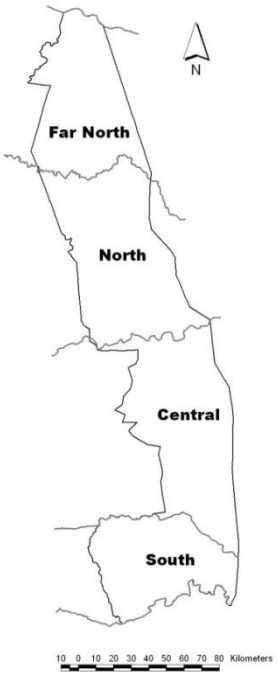
The Greater Kudu is an herbivorous antelope. It is the second largest antelope, after the Eland (*Taurotragus oryx*), and is classified as an ungulate because it has hooves. The female of the species live in groups of 2-15 in a home range of 7-12 sq. miles, and the males live in a range of about 11 sq. miles that can overlap the home ranges of 2-3 female groups (Perrin, 1999). Kudu are distinguished by the 6-10 stripes that run across their sides and the males' long twisted horns. They can live for up to 23 years in captivity and are indigenous to Angola, Botswana, Central African Republic, Chad, Democratic Republic of the Congo, Djibouti, Eritrea, Ethiopia, Kenya, Malawi, Mozambique, Namibia, Somalia, South Africa, Sudan, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe (International Union for Conservation of Nature, 2002) .

1.2 Kudu abundance in the KNP

The KNP is the biggest game reserve in South Africa. It spreads over 20,000 square kilometres and covers 350 km from north to south and 60 km from east to west (South African National Parks, 2008) see fig 1. below. In the KNP, aerial censuses of Kudu have been done since 1977 as part of the parks' "Ecological Aerial Surveys" (EAS) program. The aim of the program is to provide annual estimates of total abundance for several large herbivorous species in order to track changes in the population trend of the species and to link such changes to variation in environmental factors (Viljoen & Retief, 1994).

The EAS program consists of annual aerial surveys of 66 census blocks (subdivisions of the KNP). Parallel strips of width 800m (400m on each side of the aircraft) are observed from a height of 65-70m at a speed of 166-185 km/hr. The six member survey team consists of a pilot, four observers (two on each side of the plane) and a data recorder. Prior to 1993 animals sighted were recorded onto a map. This

procedure was then replaced by the real-time recording of sighted animals onto a Global Positioning System (GPS) coupled to a hand-held palmtop computer. Counts were conducted during the dry winter months of May to August. Between 1980 and 1993 treatment of the park was complete however, in 1977 only the northern half was covered, in 1996 only the central region and in all the other years one of the four major regions was excluded.



Prior to the count in 1993, the KNP was completely fenced; however, in 1993 the western boundary fence along the parks' Central Region was removed. The possibility of two-way movement between the KNP and the adjoining private wildlife reserves that resulted did not have a significant effect on count totals, as evidenced by ungulate counts conducted on the private reserves. (Owen-Smith & Mills, 2006).

Several of the species monitored by the EAS, including kudu, experienced a progressive decline in abundance after 1986 (Ogutu & Owen-Smith, 2003).

Figure 1: Map of The KNP

1.3 Modelling Population Abundance

In building a model for Kudu population abundance, we need to take into consideration that population levels change mainly in response to a number of factors such as annual rainfall, resource availability, competition with other species at the same trophic level in the food chain, disease, their own population levels and predation (Owen-Smith & Mills, 2006) . In addition, it is necessary to take into consideration that modelling uncertainty arises from a number of sources which

include *inherent randomness*, *model uncertainty* (which includes *measurement error*, *systematic error* and *subjective judgment*) and *natural variation* (Regan, Colyvan & Burgman, 2002) see Section 2.2 below for details. In particular it is necessary to take into consideration the measurement errors that arise as a result of the aerial survey enumeration procedure. More details of the inherent weaknesses of aerial surveys as an enumeration procedure are given in Section 2.2 below. The numerous population ecology models that have been developed, including the models of Bulmer (1975) and Dennis and Taper (1994), do not explicitly account for measurement error and therefore are not suitable for aerial survey data. An alternative modelling approach which does enable the inclusion of measurement errors into the modelling process is investigated in this research report.

Although only the total count is observed in each year, the total count is made up of subgroups – juveniles, yearlings and adults. The number observed in any year is made up of

- recruits (new juveniles)
- yearlings (survivors of the previous year's juveniles), and
- adults (survivors of the previous year's yearlings plus the survivors of the previous year's adults).

The recruits are the offspring of the adult females, and are a function of the number of these, as well as the average number of offspring for a female. The numbers for each year build on the numbers of previous years, hence forming a chain. The model is based on the true numbers (state) of the different groups in each year, plus the survival rates and reproductive rate. These rates are variable, and so can be specified as following a distribution with unknown parameters, which are themselves subject

to random variability. The estimated model relates the observed counts to this model of the dynamics, including the distributions, and also includes a term allowing for the random variability of the observations. The mathematical definition of the model is given in Section 3.4. The research follows the work of Owen-Smith and Mason (2005) and Owen-Smith, Mason and Ogutu (2005) in that the possible models estimated in those studies are used to define the population dynamic process for Kudu in this study. This study uses the possible models estimated as input into a new approach based on a hierarchical Bayesian State Space Model (SSM) (see Section 3.4.1.1).

A Bayesian SSM has been investigated in this research report because it provides a framework for including multiple sources of uncertainty in the model - including the measurement error process and also enables us to reconcile the total counts to a population structure. In this framework the system is defined in terms of two processes. The actual population of the animals is defined by the "state" process and the observations made on the state process (the counts) are defined by the "space" process. The relationship between the actual process (the state) and the counts (the space) is defined by a set of probability density functions that can be used to incorporate multiple sources of uncertainty in determining the relationship between the two (see Section 3.1 below). The SSM will be estimated using Markov Chain Monte Carlo (MCMC) estimation methods. MCMC is a computational technique used to fit Bayesian SSMs. MCMC computation is discussed in more detail in Section 3.3 below.

1.4 Aim and objectives and limitations of the report

The aim of this study is to build and fit a Bayesian SSM of population dynamics which explicitly accounts for the measurement errors in the time series of aerial survey data and which also reconciles the total count data to an estimated population structure.

The model fit will be assessed using the following convergence diagnostics

1 kernel density estimates of the model parameters

2 sample trace plots of the estimated parameter values

3 quantile plots of estimated parameter values

4 the estimated Markov Chain error, and

5 plots of the autocorrelation function for each variable in the chain.

The convergence diagnostics are discussed in Section 3.4. The research problem is to investigate the usefulness of a model that accounts for plausible population dynamics and the measurement errors in the data, giving a partition of the variance of the census counts into that which can be attributed to the true population dynamic process and that resulting from observation error.

Apart from the ability of the model to take the hidden process variables that affect count efficiency into account, it also provides statistical measures of the quality of the model fit and allows possible areas of misfit to be identified. The time period being investigated has been chosen because it allows us to focus on the underlying process, without having to include additional parts to the model to account for the effect of different areas being surveyed in different years and for the removal of the fence between KNP and the private game reserves. The focus of the study is not on getting the full abundance estimates over the period of the surveys, but rather on

identifying the usefulness of the model, which can then be extended to take these aspects into account.

A possible limitation is that these aspects have not been considered as they were beyond the scope of the study. Another possible limitation is that no comparison is made between the results of the model specified to the results of other models specified in the literature (as discussed in Section 2.3). Such a comparison was also beyond the scope of this Research Report.

The focus of the study is on highlighting how to add a specification for measurement error or any other hidden process to the systematic model for population abundance which is derived from ecological theory in a single statistical model. Distributional assumptions for the process and the measurement error can be easily modified and evaluated within the general Bayesian hierarchical modelling framework described and investigated in this Research Report.

The research report covers counts over the period from 1983 to 1993 since for these consecutive years counts covering all the regions of the KNP were carried out. This means that, in assessing the model fit, issues of data missing for some sections of the park, the change from recording on maps to using GPS, and any possible changes related to the removal of the fence, need not be considered.

1.5 Data

The report covers counts over the 11 year period from 1983 to 1993. For these consecutive years counts covering all the regions of the KNP were carried out. The time period chosen for investigation allows us to focus on the underlying process, without having to include additional parts in the model to account for the effect of

different areas being surveyed in different years and for the removal of the fence between KNP and the private game reserves after 1993. The model data also includes measurements of mean wet and dry season rainfall over the KNP over the 11 year study period as shown in table 1.

Year	EAS Count	Wet Season Rainfall	Dry Season Rainfall
1983	7005	0.767	0.4
1984	8661	0.078	0.491
1985	10,432	0.259	0.164
1986	10,760	0.413	0.358
1987	8,786	0.371	0.380
1988	8,214	0.008	0.543
1989	6,935	0.296	1.107
1990	5,967	0.1	0.94
1991	4,706	0	0.907
1992	3,976	0.874	0.919
1993	3,172	0.092	0.612

Table 1: EAS count and rainfall figures 1983-1993

1.6 Organisation of the report

The rest of this report is organised as follows: Chapter 1 contains the Introduction, Aims, Objectives and Limitations of the report. Chapter 2 is a Literature Review. Chapter 3 covers the Methodology used and Chapter 4 presents Results. Chapter 5 includes a Discussion of Results, Conclusions and Recommendations.

Chapter 2

Literature Review

In this chapter we look at the literature concerning possible explanatory variables related to population dynamics, sources of uncertainty and possible models in the separate sections below.

2.1 Population Abundance Dynamics

Animal population levels change in response to the direct and or lagged influence of resource availability, their own population size, environmental conditions and direct or lagged changes in their interaction with, or position relative to, other animals which compete with them in the food chain (Owen-Smith & Mills, 2006).

Density dependent growth is population change that is mediated through factors that are affected by the population size and that become more effective as the population grows (United States Environmental Protection Agency, 2006). The availability of resources is an example - as the population of a species grows the importance of *availability or absolute levels of natural resources* as a determinant of population size also grows. The assumption is that there is an equilibrium level at which the population stabilises; however that level will change as a result of background changes in resource availability, predation and climatic conditions (Turchin, 1995). The exact mechanism by which density dependence operates will determine whether its effect is felt immediately or after a time lag. Density dependence will, for instance, be felt immediately as a result of heightened competition for scarce resources and heightened predation in response to increased abundance. The effects of density dependence mediated through changing predator abundance in response to

changing prey levels, recovery times of vegetation and changing populations of species at the same trophic level will be felt after a time lag (Turchin, 2003).

Density dependence is an internal cause of population change. External factors can also directly influence population levels; for instance, Grenfell, Wilson, Finkenstädt, Coulson, Murray, Albon, Pemberton, Clutton-Brock and Crawley (1998) found high synchronicity in the population dynamics of Soay sheep on two islands 3.5km apart, implicating large scale weather variations as a major factor in population changes. Ogutu and Owen-Smith (2003) investigated the effect of the large scale climatic conditions governed by the El Niño Southern Oscillation (ENSO) and the effects of wet and dry season rainfall on the abundance of the 11 species covered by EAS surveys.

The relative significance of the possible influences will differ from case to case. Populations of moose (*Alces alces*) monitored over a 40 year period in the Isle Royale National Park, Lake Superior, USA, responded mostly to the abiotic influences of winter precipitation, the North Atlantic Oscillation Index and average spring and summer temperatures, and least to the influence of their predator the wolf (*Canis lupus*) (Vucetich & Peterson, 2004). On the contrary, predation was the major cause for the persistently low densities of four ungulate species monitored over 34 years in the Etosha National Park, Namibia, (Gasaway, Gasaway & Berry, 1996). In the KNP, populations of Roan antelope (*Hippotragus equinus*) fell largely as a result of increased predation. Kudu populations (and all other ungulate populations that experienced a persistent decline after 1986) were more susceptible to variations in wet and dry season rainfall (Owen-Smith & Mason, 2005; Owen-Smith, Mason & Ogutu, 2005).

In addition, the effect of the different factors may not be readily discernible. Changes in vegetation which have an effect on an animal's general health and survival, also affect the likelihood of its succumbing to predation (Smuts, 1978). Post, Stenseth, Peterson, Vucetich and Ellis (2002) studied the predator-prey relationship between a closed population of wolves and moose on Isle Royale (Michigan, USA) and noted a confounding of the effects of predation with the presence of higher loads of parasites, and other factors such as nutrition and weather which affected the rate of predation but also affected population levels. Owen-Smith and Mills (2006) cite heightened predation as a factor in population declines which came about through augmentation of water points in the KNP and the resultant increase in the food base of predators.

Lastly, the separate age and sex segments of the population respond differently to the various influences on population dynamics, necessitating that models be refined to reflect gender and life-history stage. For instance, in red deer (*Cervus Elaphus*) males are more susceptible to low winter temperatures, food shortages and adverse climatic conditions in general (Clutton-Brock, Major & Guinness, 1985). Sexual dimorphism is a systematic difference in the physiological characteristics of the male and female members of a particular species, and it has been investigated as a possible cause of the differing survival rates between adult male and female Kudu (Owen-Smith, 1993). Owen-Smith (1990) found that juvenile, yearling and old female Kudu survival rates were affected by lagged annual rainfall and lagged biomass density, whilst only a weak relation to both variables was found for younger adult females. Studies of ungulates in the KNP (Owen-Smith *et al*, 2005; Owen-Smith & Mason, 2005) have considered changes in the survival rates of the different life-history stages and the resultant effect on population trend and have also considered the

different factors most affecting survival at each life-history stage. Coulson, Catchpole, Albon, Morgan, Pemberton, Clutton-Brock, Crawley and Grenfell (2001) investigated the effects of density, climatic variations and demographic structure on the population dynamics of Soay sheep found in the St. Kilda archipelago in the UK, and found that irregular fluctuations occurred because the age and sex structure of the population was independent of population size and that the different age and sex segments responded differently to density and weather changes such that the effect of weather changes on population dynamics was different even though the overall population size was similar. This underscores the importance of including the gender and age structure of the population in a model of population dynamics.

2.2 Sources of Uncertainty

Knowledge of the possible explanatory variables discussed above is not sufficient to build a plausible model of animal population dynamics. There are 3 classes into which uncertainty concerning data and models in ecology can be divided; *natural variation, inherent randomness and model uncertainty* (which includes *measurement error*, *systematic error* and *subjective judgment*) (Regan *et al*, 2002).

1 Natural variation occurs when ecological processes change in time or space in a way that is unpredictable.

2 Inherent randomness in an ecological process is when it varies, not because of any errors made in our observations of it, but because it is not deterministic in nature.

3 Model uncertainty arises because of the imperfect nature of the model building process (imperfect answers to questions such as: Which variables to include and

in what mathematical form?) and the limited knowledge of the model builders. These first three categories of uncertainty largely encompass the factors discussed in the previous section. In addition to these we have also to consider:

4 Measurement error; defined as error that occurs because of the equipment used in the observation process or operator error.

5 Systematic error, which is similar to measurement error as regards its sources, however measurements that are subject to systematic error will be biased; even repeated measurements from such a process will converge to a mean that is different from the true mean. Systematic errors are very difficult to detect and the only way to remedy them is to recognize them and build them into the model. It is necessary for us to understand the causes of systematic error and make adjustments for them, which implies knowledge (or at least accurate estimates) of their size and direction.

6 Subjective uncertainty, which arises as a result of the different interpretations that can be given to data and results (more so if they are error prone).

Owen-Smith and Mills (2006) note the following problems with aerial survey data.

1 Aerial survey data are biased undercounts (Jachmann, 2002), with the level of bias changing from year to year according to weather conditions. For instance, counts of brown animals in drought years are highly biased (undercounts) as a result of animals being camouflaged in dry grass.

2 Weather conditions affect both counting efficiency and population dynamics, making it difficult to separate spurious from systematic influences on population change. For instance; count efficiency is a function of weather and habitat

conditions which are largely functions of rainfall. Thus rainfall (a possible process variable) has both a process and spurious effect on abundance as counted.

3 Juvenile animals are less visible therefore population changes following a change in the juvenile survival rate may be under or over-estimated.

4 An undercount in one year will spuriously magnify the apparent increase in the following year generating impracticable survival rate estimates in excess of 1 for some species and negative autocorrelation of error between successive observations.

5 Smoothing transformations introduced to mitigate the effects of deviant counts result in serial autocorrelation of errors (Chatfield, 2001). Such transformations also cause difficulties in separating the effect of predictors acting over the long term from the effects of short term factors. Ogutu and Owen-Smith (2003) experienced difficulties distinguishing the effects of seasonal rainfall which is a short-term influence and changing habitat and predator abundance which act over the longer term.

6 It should be noted that as a result of changes in rainfall from year to year or over groups of years and the effects of this on counting efficiency, the data is likely to exhibit a non-constant variance or heteroscedasticity as well as serial correlation of errors (an undercount in one year spuriously magnifies the increase in the next).

2.3 Models for Population Abundance

Several models for population abundance have been proposed. Bulmer (1975) developed a random walk model of population abundance. Ives, Abbot and Ziebarth (2010) investigate an autoregressive moving average model. Pollard, Lakhani and

Rothery (1987) investigated a distribution free model of population dynamics that utilises the correlation coefficient between observed population changes and the absolute population size to test hypotheses on the determinants of population dynamics. Dennis and Taper (1994) use a discrete time stochastic logistic model as a descriptive model for population dynamics and estimate the parameters of this model using maximum likelihood techniques. Shenk *et al* (1998) showed in an extensive review of population dynamics models that these models break down in the presence of sampling variability and, therefore, are not suitable for use in the presence of measurement errors.

The EAS surveys data are the result of aerial counts. Counts viewed in isolation provide limited information (Williams *et al*, 2002). They show some unknown proportion of the true population size, and require more information about this proportion to allow estimation of population parameters and inference about abundance, or change in abundance. Assume that we count over an area inhabited by a population and that all animals should be counted. It follows that the only animals which are not counted are those which have not been detected.

Let,

C = abundance count

β = detection probability given that the animal is present in the population.

Then, the expected value of the count is;

$E[C] = \beta \mathrm{N}$, where N is the true population and

$1-\beta$ = undetected proportion.

Detectability, which results in measurement error, is a major problem with aerial survey data. Analyses which take into account the uncertainty that results from population dynamics not being deterministic in nature (natural variation and inherent randomness as above) and uncertainty resulting from a possible misspecification of the model used (model uncertainty), ignore a further source of uncertainty which arises from errors in the observations made on the process (measurement error). Such models underestimate the uncertainty about population processes and abundance (Calder, Lavine, Muller & Clark, 2003; Newman *et al*, 2006). It is this problem which this report attempts to address. An attempt is made to formulate a model which accounts for all the sources of variation simultaneously.

Newman *et al* (2006) propose using hidden process models (HPMs) to account for the stochasticity in the process (process variation) and also stochasticity in the measurements made on the process (observation variation). It is assumed that the true data on the ecological process is a latent (unobserved) variable, whose development is modelled by a process model which accounts for process error. The observed data are taken to be functions of the true data with appropriate probability distributions describing observation error. The HPM, which gives the relationship between the process model (called the state process) and the actual data, takes into account the observation process with observation error; called the space model.

An HPM is an extension of the State Space Model (SSM) described by Harvey (1989). In an SSM it is assumed that the current state depends only on the previous state, that is, has the Markov property. This assumption is relaxed in the special case of the HPM, such that the current state can be dependent on the state going back any

number of previous periods and not only on the immediately preceding period, as in the case of the SSM.

A hierarchical HPM is one in which the random parameters are themselves dependent on other parameters which are called hyperparameters (see Section 3.2.2). Hierarchical modelling allows for the inclusion of an additional level to the model building that enables more sources of variation to be included. For instance one of the random parameters in a model of population dynamics could be the juvenile survival rate, and this can be modelled in a hierarchical model as being dependent on hyperparameters which include factors such as rainfall, competition and predation. Buckland, Anderson, Burnham, Laake, Borchers and Thomas (2004: pp 97) state that "Such generalized state-space models are sometimes referred to as hidden Markov or hidden process models." Buckland *et al* (2004: pp 97) further note that "Environmental stochasticity … could be added by allowing one or more biological process parameters to be random variables. Such models are called hierarchical state-space models." These models will be referred to as HPMs in this report.

The approach to SSMs and HPMs in the recent literature is via Bayesian as opposed to classical methods of inference. Probability density functions that link the observations to the process can take on many forms reflecting different sampling or estimating procedures such as mark-recapture and attempted censuses using line transect methods such as aerial survey.

Using SSMs and HPMs it is possible to model the development of the process, as well to forecast future values. SSMs have been applied extensively in statistical research; see, for example, Jones (1993) on serially correlated longitudinal data and Durbin and Koopman (2001) on state space approaches to time series analysis.

Under assumptions of linearity and normality of errors, Bayesian methods yield fully analytical results. However assumptions can be relaxed to accommodate non-linearity and non-normality and analysis pursued via computer intensive posterior simulation methods such as MCMC simulation (Durbin & Koopman 2001; Buckland, Newman, Thomas & Koesters, 2004; Newman *et al*, 2006; Calder *et al*, 2003; Millar & Meyer, 2000).

The flexibility of Bayesian SSMs in terms of allowing for non-linearity and non-normality of error terms has resulted in increasing applications in recent times. In ecological studies, researchers in fisheries applications have shown the most interest in incorporating observation error into their models using SSMs. Because they have been easier to calculate using the Kalman filter, early studies have assumed normal models for the observations and the process, as well as a linear relationship between the observations and true abundances (Buckland *et al*, 2004). Such models are called *normal dynamic linear models* (West & Harrison, 1997). Sullivan (1992) used a Kalman filter state space model to incorporate observation error and system stochasticity into a model of fish population dynamics based on observed data which were length categorised catches. The Kalman filter approach to SSMs for fisheries has also been used by Pella (1993), Gudmundsson (1994), Schnute (1994), Freeman and Kirkwood (1995), Kimura, Balsiger and Ito (1996), Red and Simons (1996) and Newman (1998). Millar and Meyer (2000), use an age-structured Bayesian SSM to analyse catch-at-age data recorded from research vessel surveys. Rivot, Prévost, Parent and Bagliniere (2004); extend the model to salmon stage-structured populations and Newman and Lindley (2006), use a hierarchical Bayesian SSM for a model of fish population dynamics that, in addition to process and observation errors, includes environmental covariates for the parameters. Bayesian SSMs have been

used for wild animal populations for red deer (*Cervus Elaphus*) in Scotland (Trenkel, Elston & Buckland, 2000), moose in Bialowieza Primeval Forest in Poland (Clark & Bjørnstad, 2004), British lapwings (*Vanellus vanellus*) (Brooks, King & Morgan, 2004) and British grey seals (*Halichoerus grypus*) (Thomas, Buckland, Newman & Harwood, 2005) amongst others. Dennis, Ponciano and Taper (2010) investigate the usefulness of replicated sampling to increase population count efficiency in a state-space framework.

Convergence diagnostics for Bayesian models give a measure of the goodness of fit of the estimated model and include measures based on examination of the kernel density estimate of the estimated posterior distribution function, examination of trace plots of estimated parameters, autocorrelation functions of sample values of parameter estimates generated by the MCMC estimator and monitoring of the Markov Chain error percentage and standard deviation of the parameter estimates (Ntzoufras, 2009). The convergence diagnostics for Bayesian HPMs are defined and discussed in more detail in Chapter 4.

2.4 Summary of the literature review

Wildlife population levels have been shown to respond mostly to resource availability, predation, overall population size, environmental conditions and relative position to other species occupying a similar position in the food chain. Population levels are also affected by large scale weather variations, climatic conditions and disease. The gender and life-history stage of the animal also affects how animals respond to the various influences necessitating that these variables be included in models of wildlife population abundance.

The literature review has shown that possible sources of variability to consider in a model for population dynamics should include natural variation, inherent randomness, measurement error and subjective judgement. Aerial surveys in particular are often biased undercounts which are very sensitive to weather conditions. Other issues with aerial surveys include the fact that juveniles are usually less visible and the effects of erroneous counts in years affected by extreme weather conditions for example spuriously affect the observed increase in subsequent years.

Several models to estimate the parameters of population dynamics models have been proposed, however these models break down in the presence of sampling variability making them unsuitable for use with data that include measurement errors.

Bayesian HPMs account for stochasticity in the process and stochasticity in the measurements which makes them suitable for use with aerial survey data. The model convergence diagnostics give an indication of the goodness of fit of the estimated model. It is our intention in this study to formulate a Bayesian hierarchical HPM for Kudu population dynamics in the KNP, and to fit it using an MCMC simulation based on the Gibbs Sampler.

Chapter 3

Methodology

This chapter discusses general issues around the estimation of HPMs. The general definition of state space models is given in Section 3.1 to 3.4, giving details of the specification of the general SSM, general issues concerning the estimation of the model, hierarchical Bayesian modelling of the processes using HPMs, MCMC estimation of the distributions and diagnostics to evaluate the adequacy of the model.

3.1 Definition of SSMs

The presentation below follows Calder *et al,* (2003), Buckland *et al, (* 2004), Thomas *et al* (2005) and Newman *et al* (2006).

Firstly, HPMs and inference procedures are described and then an outline of how the model is defined for the KNP data is given.

The model described below spells out how the observed count data and the assumptions concerning the initial distribution of the true state of the process are iteratively combined to estimate the true abundance at each time point. The parameters of the model are the factors affecting observation (some measures of observation error) and the factors affecting population dynamics (such as predation rates and competition rates).

The HPM is defined as follows. At time point t, we define two parallel time series initialized at time point 0;

- y_t = vector of observed counts from time $t = 0$ up to T, called the space

vector.

- n_t = vector of actual abundance from time $t = 0$ up to T, called the state vector. This vector is assumed to be unobservable. Note that n_t can be of higher dimension or have more components than y_t. For instance, whilst y_t might represent an overall animal count total, we might wish to model n_t as counts broken down by age (juvenile/adult) and/or sex. Depending on the extent of the difference in dimensionality, this will affect the identifiability of the model.

Following the specification of HPMs in Newman *et al* (2006), three classes of probability density functions (pdfs) are identified. The first gives the initial state vector, the second describes the development of the state vector from one time period to the next and the last gives the relationship of the observation or space vector to the state vector. Formally

1 initial state pdf = $g_0(n_0 \mid \theta)$,

2 state t pdf = $g_t(n_t \mid n_{t-1}, n_{t-2}, \ldots, n_0; \theta)$, summarised; $g_t(n_t \mid n_{0:t-1}; \theta)$,

3 observation t pdf = $f_t(y_t \mid n_t; \theta)$,

where θ is the vector of parameters.

The maximum likelihood estimate of θ is that which maximizes the likelihood obtained by integrating the joint distribution of the state n_t and the observations y_t over the state n_t. Following Newman *et al* (2006) and Buckland *et al* (2004) the likelihood is given by

$$p\left(\mathrm{y}^{\mathrm{T}}|\theta\right)=\int_{\mathrm{n}_0\ldots\mathrm{n}_t}\left\{\prod_{\mathrm{t}=1}^{\mathrm{T}}\mathrm{f}_{\mathrm{t}}\left(\mathrm{y}_t|\mathrm{n}_t;\theta\right)g_t\left(\mathrm{n}_t|\mathrm{n}_{t-1};\theta\right)\mathrm{dn}_t\right\}g_0\left(\mathrm{n}_0;\theta\right)\mathrm{dn}_0 \tag{1}$$

where y^{T} is a transposed vector of observations from period 1 to T and

$$\int_{n_0\ldots n_t}=\int_{n_0}\ldots\int_{n_t}.$$

The expected value of the state n_t, given θ and y_t, is obtained by integrating over the states:

$$E\left[\mathrm{n}_t\,|\,\mathrm{y}^{\mathrm{T}};\theta\right]=\frac{\int_{n_0\ldots n_t}\mathrm{n}_t\left\{\prod_{t=1}^{T}f_t\left(\mathrm{y}_t\,|\,\mathrm{n}_t;\theta\right)g_t\left(\mathrm{n}_t\,|\,\mathrm{n}_{t-1};\theta\right)\mathrm{dn}_t\right\}g_0\left(\mathrm{n}_0;\theta\right)\mathrm{dn}_0}{p\left(\mathrm{y}^{T}\,|\,\theta\right)} \tag{2}$$

This procedure of calculating the expectation of n_t given the full vector of observations y^{T} is known as *smoothing*. If inference is made given only the observations up to time point t, that is using only the vector of observations up to time point *t* or y^{t}, the procedure is referred to as filtering, whilst inference made given $\mathrm{y}^{\mathrm{t-1}}$ is referred to as one-step ahead prediction.

Bayesian inference involves obtaining the posterior distribution for the vector of parameters θ. The posterior distribution is given by:

$$p\left(\theta\,|\,\mathrm{y}^{T}\right)=\frac{p\left(\mathrm{y}^{T}\,|\,\theta\right)p\left(\theta\right)}{f\left(\mathrm{y}^{T}\right)} \tag{3}$$

where $f\left(\mathrm{y}^{T}\right)$ is obtained by integrating $p\left(\mathrm{y}^{T}\,|\,\theta\right)p\left(\theta\right)$ over θ.

3.1.1 Representation of the State and Observation Processes

Following Newman *et al* (2003) and Buckland *et al* (2004), we present the state process in a modular form. The pdf for the state process variation g_t is "modularised" into a *sequence* of sub-processes that define the overall process evolution. This allows us to model the pdf g_t using a set of linked pdfs representing each sub-process, enabling us to focus attention on individual sub-processes and to test theories concerning their nature. Possible sub-processes might be birth and survival.

The state process is modelled as follows. After the r th sub-process, the population is in state $\mathrm{u}_{r,t}$. Given k such sub-processes occurring between time intervals $t-1$ and t the sequential development of the process is described below:

$$\mathrm{u}_{1,t} \sim G_{1,t}\left(\mathrm{n}_{0:t-1,}\theta\right),\ \mathrm{u}_{2,t} \sim G_{2,t}\left(\mathrm{u}_{1,t}\theta\right),\ \ldots,\ \mathrm{n}_t = \mathrm{u}_{k,t} \sim G_{k,t}\left(\mathrm{u}_{1,t},\ldots,\mathrm{u}_{k-1,t},\theta\right) \tag{4}$$

The $G_{r,t}$ are conditional distribution functions for the sub-processes. Thus, for example, $\mathrm{u}_{2,t}$ has conditional pdf $g_{2,t}\left(\mathrm{u}_{2,t} \mid \mathrm{n}_{0:t-1},\mathrm{u}_{1,t},\theta\right)$. To obtain $g_t\left(\mathrm{n}_t \mid \mathrm{n}_{0:t-1},\theta\right)$, it is necessary to integrate over the sub-processes;

$$g_t\left(\mathrm{n}_t \mid \mathrm{n}_{0:t-1},\theta\right) = \int_{u_{1:k-1,t}} \prod_{r=1}^{k} g_{r,t}\left(\mathrm{u}_{r,t} \mid \mathrm{n}_{0:t-1},\mathrm{u}_{1:r-1},t,\theta\right) d\mathrm{u}_{1:k-1,t}, \tag{5}$$

where $\mathrm{u}_{1:r,t} = \mathrm{u}_{1,t},\ldots,\mathrm{u}_{\mathrm{r,t}}$. Note: $\mathrm{n}_t \equiv \mathrm{u}_{k,t}$; the state of the population after the k th sub-process.

The full model for the states, observations and parameters is:

$$\Pr\left(\mathrm{n}_0, \mathrm{u}_{1:k,1}, \ldots, \mathrm{u}_{1:k,T}, \mathrm{y}_{1:T}, \theta\right) =$$

$$g_0\left(\mathrm{n}_0 \mid \theta\right) g_0\left(\theta\right) \prod_{t=1}^{T} \left[f_t\left(\mathrm{y}_t \mid \mathrm{n}_t, \theta\right) \prod_{r=1}^{k} g_{r,t}\left(\mathrm{u}_{\mathrm{r,t}} \mid \mathrm{n}_{\mathrm{o}:t-1}, \mathrm{u}_{1:r\text{-}1}, t, \theta\right) \right], \quad \mathrm{n_t} = \mathrm{u_{k,t}} \tag{6}$$

In defining the pdf of the observation process as $f_t\left(\mathrm{y}_t \mid \mathrm{n}_t, \theta\right)$, we are assuming that the observation process given the present abundance is independent of previous counts. This can be relaxed, for instance when the y_t are weighted combinations of current and past survey data and thus become functions of n_t and previous y.

3.2 Bayesian modelling

Evaluating the distributions above requires integration. If the observation and state processes are linear in their parameters, and in some other special cases, analytical solutions to the integrations may be possible. In the case of linear and Gaussian distributed models, the Kalman filter (Harvey, 1989) can be used to obtain solutions (see Besbeas, Lebreton & Morgan, 2003, for an ecological example).

The cases described above are examples of classical inference in state space and HPMs. The approach to HPMs and SSMs in the recent literature is via Bayesian as opposed to classical methods of inference. The reason for this is that computational Bayes methods allow inference for a wide range of non-normal distributions; thus allowing more reasonable distributional assumptions to be made in the modelling process. This is the approach taken in this study. Thus model parameters are considered to be random variables and the data are utilized to update our beliefs concerning the distributions of these random variables. The goal is to estimate $\mathrm{n}_{1:T}$ and θ given $\mathrm{y}_{1:T}$. Bayesian inference using computer intensive procedures, referred

to as Monte Carlo inference, allows investigation of the posterior distributions of $\mathrm{n}_{1:T}$ and parameter vector θ through simulation. We obtain simulated sample values of the states $\mathrm{n}_{1:T}$ and the parameters θ, given the observations $\mathrm{y}_{1:T}$. Monte Carlo inference can be carried out using either MCMC (Gilks, Richardson & Spiegelhalter, 1996) or Sequential Importance Sampling (Doucet, de Freitas & Gordon, 2001) approaches. MCMC techniques are utilized in this study because of the computational advantages described above. The methodology allows us to carry out model evaluation and convergence checking using summary statistics such as the Markov Chain Error and a number of diagnostic plots discussed in Chapter 4.

3.2.1 Brief Overview of Bayesian Inference

Bayesian inference involves a reversal of the traditional roles assigned in the classical statistical inference framework, whereby; given observed data *y* the classical inference problem is to make inferences concerning the distribution of the observed data given parameters θ , $f\left(y \mid \theta\right)$. The data are assumed to be random realiz`ations of a stochastic process with constant parameters θ. Conversely, in the Bayesian approach inference is made concerning $f\left(\theta \mid y\right)$; the unknown parameters are treated as random variables. The Bayesian approach can be summarised in four key steps

1. Specification of the likelihood model $f\left(y \mid \theta\right)$. This model describes the best model for the data and is similar to the model postulated in classical inference.
2. Determination of a prior distribution for the parameters θ. The prior distribution embodies our beliefs or expert knowledge regarding θ before

observing the data. They are largely subjective in nature.

3. Calculation of the posterior distribution $f(\theta \mid y)$ using Bayes Theorem.
4. Making inferences based on the posterior distribution.

The data update our initial model via Bayes Theorem:

$$f(\theta \mid y) = \frac{f(\theta) f(y \mid \theta)}{\int f(\theta) f(y \mid \theta) d\theta} \tag{7}$$

3.2.2 Hierarchical Bayes Models

Hierarchical models arise when we wish to include a further level to our model. This level consists of covariates that determine the values of our models' random parameters.

First, we specify the distribution, $f(\mathrm{y} \mid \theta)$, of the observations $\mathrm{y} = (y_1, \ldots, y_n)$ given the unknown parameters $\theta = (\theta_1, \ldots, \theta_k)$; and assume θ is a random realisation from the *prior distribution*, $\pi(\theta \mid \lambda)$ and λ is a vector of *hyperparameters* that determine our model parameters.

Suppose λ is known. Inference for θ is given by the *posterior distribution*;

$$p(\theta \mid \mathrm{y}, \lambda) = \frac{p(\mathrm{y}, \theta \mid \lambda)}{p(\mathrm{y} \mid \lambda)} = \frac{p(\mathrm{y}, \theta \mid \lambda)}{\int p(y \mid \theta) d\theta} = \frac{f(\mathrm{y} \mid \theta) \pi(\theta \mid \lambda)}{\int f(\mathrm{y} \mid \theta) \pi(\theta \mid \lambda) d\theta} \tag{8}$$

In practice λ is not known and a *hyperprior distribution*, $h(\lambda)$, is required and (8)

becomes $p\left(\theta \middle| \mathrm{y}\right) = \dfrac{p(\mathrm{y},\theta)}{\mathrm{p}(\mathrm{y})} = \dfrac{\int f(\mathrm{y}|\theta)\pi\left(\theta \middle| \lambda\right)h(\lambda)d\lambda}{\int f(\mathrm{y}|\theta)\pi\left(\theta \middle| \lambda\right)h(\lambda)d\theta d\lambda}$ (9).

Another approach is to estimate λ with $\hat{\lambda}$; which is that value which maximizes the marginal distribution $p(\mathrm{y}|\lambda) = \int f(\mathrm{y}|\theta)\pi(\theta|\lambda)d\theta$ taken as a function of λ.

Inference is given by the estimated posterior distribution $p\left(\theta \middle| \mathrm{y},\hat{\lambda}\right)$ given by inserting $\hat{\lambda}$ into (1). This is the method known as *empirical Bayes analysis*.

Computational difficulties can arise in calculation of the normalising integral in the denominator of Bayes formula. These are addressed using MCMC techniques.

3.3 Markov Chain Monte Carlo Techniques and Bayesian Inference

Use of Bayesian methods has in the past been inhibited by the high dimensional integrations required to calculate the posterior distribution. More recently, the problem has been addressed by simulation methods which are used to draw samples which converge in distribution to the required posterior distribution (see below).

3.3.1 Bayesian Computation

A major reason why Bayesian analysis could not be carried out for most realistic problems in the past was that, in many cases, the integrations required to carry out inference using Bayes Theorem were intractable in closed form. There are certain forms for the prior, called *conjugate priors*, which allow at least part of the solution to be obtained analytically, but generally the presence of nuisance parameters such as unknown variances results in intractable integrations remaining. Bayesian analysis is enjoying renewed interest because the availability of cheap high speed computing in

more recent times has allowed the use of MCMC methods. Using these methods it is no longer necessary to carry out complex integrations. An example of these methods which is used in this study, the Gibbs sampler, is presented below.

3.3.1.1 The Gibbs Sampler

In this section we describe the Gibbs Sampler, which is one of the MCMC integration techniques that are used in Bayesian analysis. MCMC integration techniques originated from the Metropolis algorithm (Metropolis, Rosenbluth A, Rosenbluth M, Teller A & Teller H, 1953). The Metropolis algorithm was developed by physicists in an attempt to resolve complex integrals by first expressing them as expectations of a distribution and then estimating this expectation by drawing samples from that distribution. As mentioned above the use of Bayesian approaches was limited by the high-dimensional integrations required to obtain the posterior distribution. MCMC integration techniques in general attempt to simulate direct samples from some complex distribution. One particular MCMC integration technique, the Gibbs Sampler, was found to be applicable to a broad class of Bayesian problems resulting in a significant increase of interest in Bayesian approaches to statistical problems (Gelfand & Smith, 1990).

The Gibbs sampler, named after the physicist J. W. Gibbs, is an algorithm which generates a sequence of samples from the joint probability distribution of two or more random variables. The sequence obtained is used to approximate the joint distribution or to compute an integral, such as an expected value. The Gibbs sampler can be used when the closed form of the joint distribution is not known, but the conditional distribution of each variable is known. It is used to draw a sample from the distribution of each variable in turn, conditioning on the current values of the other variables. These methods have gained popularity because they allow inference

from posterior distributions of large dimensions, not by giving closed form solutions but by giving a sample $\left\{\theta^{(g)}, g=1,\ldots,G\right\}$ from the posterior distribution. These samples are generally adequate for reliable inference (Clark & Gelfand, 2006). The methods have become popular because the samples drawn (correlated recursive draws from a stationary Markov chain, the stationary distribution of which is equivalent to the posterior distribution required) can be made arbitrarily accurate by increasing the Monte Carlo sample size G (Clark & Gelfand, 2006). Convergence of the Markov chain to the posterior can be guaranteed for an extremely wide range of posterior distributions. There are, however, side issues with the determination of the point at which the Markov chain has converged and with the estimation of variances.

There are several versions of the Gibbs sampler, such as the slice Gibbs sampler (Ntzoufras, 2009). This is used when the full conditional distributions cannot be written in a convenient form. The standard version of the Gibbs sampler, as used in this research, is described below. Given a model of k parameters $\theta=\left(\theta_1,\ldots,\theta_k\right)$, we define the *full conditional distribution* of a parameter θ_i as $\left\{p\left(\theta_i \mid \theta_{j\neq i}, \mathrm{y}\right),\ i=1,\ldots,k\right\}$. The full conditional distribution is usually determined from the product of the prior distribution of the parameter of interest and the likelihood function, i.e. $p\left(\theta_i \mid \theta_{j\neq i}, y\right) \propto \pi\left(\theta_i\right) f\left(y \mid \theta\right)$.

The Gibbs sampler consists of the following steps.

Begin with arbitrary starting values, $\left(\theta_2^{(0)},\ldots,\theta_k^0\right)$

For $\left(t \in 1:T\right)$, do:

1 Draw θ_1^t from $p\left(\theta_1 \mid \theta_2^{t-1}, \theta_3^{t-1}, \ldots \theta_k^{t-1}, \mathrm{y}\right)$

2 Draw θ_2^t from $p\left(\theta_2 \mid \theta_1^t, \theta_3^{t-1}, \ldots \theta_k^{t-1}, \mathrm{y}\right)$.

.

k Draw θ_k^t from $p\left(\theta_k \mid \theta_1^t, \theta_2^t, \ldots \theta_{k-1}^t, \mathrm{y}\right)$

It has been shown (Gelman, Carlin, Stern & Rubin, 1995) that the distribution of the resultant k-tuple obtained at iteration t, converges to the actual joint posterior distribution, $p\left(\theta_1, \ldots, \theta_k \mid \mathrm{y}\right)$. For large enough t, say greater than t_0, $\left\{\theta^t,\ t = t_0 + 1, \ldots, T\right\}$ is a correlated sample from the true posterior and can be used to estimate any posterior properties of interest. The period from $t = 0$ to $t = t_0$ is called the *burn-in* period; which is the amount of time it takes for the Markov chain to converge.

There are number of possible ways in which to determine the point at which the sample has converged. These include an inspection of the autocorrelation function of the samples. The point at which this autocorrelation reduces to zero indicates the point at which the Markov Chain has converged. The number of iterations required for the Markov Chain to converge can be greatly reduced by making an appropriate choice of starting values which would ensure that the chain is initiated as close to the centre of the distribution as possible. In the present research, the starting values for the key parameters are actually values for these parameters that have been estimated using other research methods (Owen-Smith and Mason, 2005).

3.3.2 Bayesian Inference for HPMs

Three steps are involved

1. We quantify prior knowledge of θ, n_0 and $\mathrm{n}_{1:T}$ by *prior* pdf $g\left(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta\right)$

given by;

$$g(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta) = g_0(\mathrm{n}_0 \mid \theta)\, g_0(\theta) \times \prod_{t=1}^{T} g_t(\mathrm{n}_t \mid \mathrm{n}_{1:t-1}, \mathrm{n}_0, \theta) \tag{10}$$

2. We define the likelihood function; which is the observed data as a function of the states $\mathrm{n}_{1:T}$ and parameters θ, i.e.

$$L(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta \mid \mathrm{y}_{1:T}) \equiv \prod_{t=1}^{T} f_t(\mathrm{y}_t \mid \mathrm{n}_t, \theta) \tag{11}$$

3. Bayes theorem is used to update the prior distribution using the observed data; yielding a *posterior* distribution,

$$g(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta \mid \mathrm{y}_{1:T}) = \frac{g(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta) \times L(\mathrm{n}_0, \mathrm{n}_{1:T}, \theta \mid \mathrm{y}_{1:T})}{f(\mathrm{y}_{1:T})} \tag{12}$$

Information on the states and parameters can be obtained by sampling from the posterior distribution and measures of location, spread, percentiles and correlations between states and parameters obtained. It is our intention in this study to formulate a Bayesian hierarchical state space model for Kudu population dynamics in the KNP, and to fit it using MCMC simulation based on the Gibbs Sampler.

3.4 The WinBUGS modelling Software and Model Output

The model is estimated using the WinBUGS (Lunn, Thomas, Best & Spiegelhalter, 2000) program. WinBUGS is a Bayesian modelling program that implements powerful MCMC algorithms that are well suited to a wide range of target distributions and that are capable of estimating complex models. The WinBUGS software is freely available at http://www.mrc-bsu.cam.ac.uk/bugs.

The following steps are required in order to estimate a hierarchical Bayesian model using WinBUGS:

1 The user specifies the model in the form of the prior distributions of the parameters to be estimated.
2 Initial values for all random parameters have to be provided, as well as data for all model inputs.
3 The software then generates MCMC simulations which have a stationary distribution equal to the posterior distribution of the parameters.
4 The WinBUGS algorithm generates a sample of observations from the posterior distribution of the parameters from which it is possible to draw an estimate of the posterior distribution using kernel density estimates and also to provide descriptive statistics based on this sample which include estimated mean values of the parameters and posterior probability intervals.
5 Posterior Probability Intervals are the Bayesian equivalent of frequentist confidence intervals. They differ from frequentist confidence intervals in that frequentist confidence intervals are determined by the data alone, whilst posterior probability intervals are also determined by research problem specific "expert" or prior information which is contained in the prior probability distribution and are thus more subjective by definition.

The WinBUGS code for the estimated model is shown in Appendix A below.

3.5 Model Diagnostics

Model diagnostics are used to assess whether or not the model has converged. There are many ways to monitor convergence of the model, however the simplest is to visually inspect a number of diagnostic plots (Ntzoufras, 2009, p 41). Several

statistical tests have also been developed to test for convergence (Ntzoufras, 2009, p 41- 42), however the visual inspection approach has been preferred for this application. These are also the standard methods used in the papers in the ecological field, as discussed in the literature review.

The following output is discussed in Chapter 4.

1 The summary statistics show the mean, median, standard deviation and a posterior probability interval of the sample values simulated from the estimated posterior distribution of the parameters. The posterior probability interval is a range of values which has a posterior probability of 0.95 of containing the parameter value. The Monte Carlo (MC) error is a measure of the variability in each estimate drawn from the MCMC which is due to simulation error. The MC error must be low in order to calculate the parameter of interest with precision. The graphs show an index ranging from 1 to 11. 1 corresponds to 1983, 2 to 1984 and so on with 11 corresponding to 1993.
2 The running quantile plots show the running posterior median of the simulations against iteration number with 95% Confidence Intervals for the parameter estimates. These can be used to assess convergence of the Markov Chain to the posterior distribution of interest. If the Markov chain has converged the plot will show a straight line without perturbations or high variability.
3 The kernel density plot. This is a smoothed kernel density estimate of the marginal posterior distribution of each parameter. This gives an indication of the expected range of values of the estimated parameters and also gives an estimate of the shape of the univariate posterior distribution of each parameter.
4 The autocorrelation function shows the correlation between sequential estimates of a parameter generated by the Markov Chain. Autocorrelation occurs when

sequential estimates of a parameter from the conditional distribution are correlated. This may be caused by highly correlated model parameters resulting in the Gibbs Sampler becoming slow in exploring the full posterior distribution. An autocorrelation function showing high autocorrelations between sequential estimates of a parameter indicates slow mixing within chains and, usually, slow convergence. This is an important model diagnostic which is used to evaluate the goodness of fit of the estimated model. In the case of high autocorrelation between sample values drawn from the MCMC, one of the remedies is to apply "thinning" to the sampling process. Thinning means that not all sample values are recorded, rather values are recorded at a regular interval determined by the researcher; say every 10th value. The order of thinning applied may be determined following an inspection of autocorrelation functions from an initial run. If autocorrelation up to lag 5 is observed then this could be mitigated by sampling every 6th value (for instance) from the MCMC. This process reduces the autocorrelation between sample values and increases the independence of the sample.

Chapter 4

Results

4.1 Model Definition

4.1.1 The State Process

The model defined for the state process is hierarchical. Survival rates are assumed to be random and are modelled as functions of environmental covariates.

Following the life history model of Owen-Smith & Mason (2005), the state vector $\mathbf{n}_t$ is of length 5 and comprises:

1 the number of juveniles (< 1 year) in year t, $\{J_t\}$,

2 the number of female yearlings (1-2 years) in year t, $\{YF_t\}$ and the number of male yearlings (1-2 years) in year t, $\{YM_t\}$,

3 the number of female adults (> 2 years) in year t, $\{AF_t\}$ and the number of male adults (> 2 years) in year t, $\{AM_t\}$.

Thus $n_t' = (J_t, YF_t, YM_t, AF_t, AM_t)$.

The development of the state vector is governed by two sub-processes which generate n_t from n_{t-1}; stage-specific survival and recruitment/birth. A non-stochastic sub-process, age-incrementation, is included between survival and recruitment.

4.1.1.1 Defining Priors

The research uses the parameters estimated by Owen-Smith *et al*, (2005) using

regression and time-series models in the initial priors, and looks at the usefulness of including the effect of covariates that affect measurement error using HPM methodology.

The realised juvenile survival is the number of juveniles that survive into the next year. The prior distribution of realised juvenile survival is defined as RJ_t ~ Binomial $\left(J_t, \phi_{J_t}\right)$ and the survival rate ϕ_{J_t} is random and reflects the influence of environmental effects as described in Owen-Smith *et al*, (2005) (see below).

Based on Owen-Smith *et al*, (2005), we assume an even sex distribution in the juvenile segment and, thus, the number of female yearlings present in year t is modelled as $\left(YF_t\right)$~Binomial$\left(RJ_{t-1}, 0.5\right)$. Realised female yearling survival is modelled as $\left(RYF_t\right)$~Binomial$\left(YF_t, \phi_{YF_t}\right)$. The number of male yearlings present in year t $\left(YM_t\right)$ is modelled as (RJ_{t-1} - YF_t). Realised male yearling survival $\left(RYM_t\right)$~Binomial$\left(YM_t, \phi_{YM_t}\right)$, where ϕ_{YF_t} and ϕ_{YM_t} are random survival rates for female and male yearlings respectively, with prior distributions reflecting the influence of environmental effects (see below). The number of adult female animals present in year t is modelled as $\left(AF_t\right) = \left(RYF_{t-1}\right) + \left(RAF_{t-1}\right)$ and the number of male adult animals present in year t $\left(AM_t\right) = \left(RYM_{t-1}\right) + \left(RAM_{t-1}\right)$; where $\left(RAF_{t-1}\right)$ and $\left(RAM_{t-1}\right)$ are the realised female and male survivors respectively. Realised adult female and male survival are distributed as, $\left(RAF_t\right)$ ~ Binomial$\left(RYF_{t-1} + RAF_{t-1}, \phi_{AF_t}\right)$ and $\left(RAM_t\right)$ ~ Binomial$\left(RYM_{t-1} + RAM_{t-1}, \phi_{AM_t}\right)$, where ϕ_{AF_t} and ϕ_{AM_t} are random survival rates for female and male adults respectively,

with distributions reflecting the possible influence of environmental, density dependent and species specific characteristics (see below). The number of juveniles born, juvenile recruitment, is modelled as $J_t \sim \text{Poisson}(\kappa * AF_t * \pi)$ where, κ is the mean number of offspring per adult female, and π is the proportion of adult females that breed.

4.1.1.2 Hyperdistribution pdfs (Model Covariates)

Survival rates are modelled as random parameters that have prior distributions that incorporate the effects of environmental covariates. Environmental effects on the survival rate are incorporated into the model based on regression models that were fitted to the survival rate of the different life-history stages of Kudu in the KNP (Owen-Smith *et al*, 2005). The fitted regression models including relative population abundance, seasonal rainfall components, predation and prior abundance are shown below (rainfall measurements were $\log_e$ transformed). Wet and dry season rainfall were the only covariates with statistically significant estimates and as a result these were the only covariates included in the model.

The covariates and the estimated regression models below are used to define upper and lower bounds for the survival rate and are treated as "prior information" or *a priori* knowledge in the Bayesian context in the model build. The prior distribution of the survival rate is a Beta distribution with lower bound given by the regression estimate below less its standard deviation and upper bound given by the regression estimate below plus its standard deviation.

Stage	Intercepts +/- Std Error	Wet Season Rainfall	Dry Season Rainfall	R^2
Juvenile	0.463 +/- 0.024	0.160 +/- 0.057	0.123 +/- 0.036	0.407
Yearling	0.847 +/- 0.026	0.162 +/- 0.068		0.290
Adult	0.875 +/- 0.017	0.123 +/- 0.046		0.341

Table 2: Estimated regression models for Kudu survival by Life Stage (Owen –Smith *et al*, 2005)

The parameters of the Beta distributions are defined below.

Digression on the Beta Distribution

The probability density function of the Beta Distribution is given by,

$$f(\phi) = \frac{(\phi - a)^{p-1}(b-\phi)^{q-1}}{\beta(p,q)(b-a)^{p+q-1}} \quad , \quad a \le \phi \le b;\ p,q > 0 \tag{13}$$

where,

$$\beta(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt\,.$$

The parameters a and b are the lower and upper bounds respectively of the distribution, and p and q are the shape parameters. Given that a and b are known, the method of moments estimators for p and q are given by,

$$p = \left(\frac{\bar{x}-a}{b-a}\right)\left(\frac{(\bar{x}-a)(b-\bar{x})}{s^2} - 1\right) \text{and}$$

$$q = \left(1 - \frac{\bar{x}-a}{b-a}\right)\left(\frac{(\bar{x}-a)(b-\bar{x})}{s^2} - 1\right).$$

The lower and upper bound parameters for the survival rates (a and b above) are obtained from the fitted regression models above. The lower bound of the survival

rate a in the prior distribution for the survival rate is given by substituting into the fitted regression model; the values of the estimated coefficients (given above) less their standard error and the upper bound is given by the values of the estimated coefficients (also given above) plus their standard error. The shape parameters of the Beta distribution (p and q above) are found by applying the method of moments estimators above, where $\bar{x}$ is simply the mean survival rate for each group and s^2 is the variance of the survival rate. The means and variances of the survival rates of the different life-history stages of Kudu in the KNP were estimated in an earlier study (Owen-Smith & Mason, 2005) and are used as $\bar{x}$ and s^2 in the formula above.

4.1.1.3 Initial Values

Owen-Smith *et al* (2005) derived estimates of the survival rates associated with zero population growth for each population segment. A vector of this constant value for each life-history stage was used as the initial value for survival over the 11 year period. The model derived in Owen-Smith *et al* (2005) also showed that adult male survival was consistently about 0.85 of adult female survival and this was reflected in the initial values and in the modelling process. The initial values are summarised in the table below.

Parameter	**Initial Value**
Juvenile survival rate ϕ_J	0.55
Female yearling survival rate ϕ_{YF}	0.70
Male yearling survival rate ϕ_{YM}	0.70
Adult female survival rate ϕ_{AF}	0.875
Adult male survival rate ϕ_{AM}	0.75

Table 3: Initial Values of Model Parameters

The total population count for Kudu in 1983 was 7005. The mean ratio (over the

period 1983-1993) of juveniles to adult females was 0.340, the mean ratio of yearlings to adult females was 0.239 and the mean ratio of adult males to adult females was 0.408. From this we derive that the mean proportion of adult females in the population was 0.503. This gives the following initial values

AF = (0.503 X 7005) = 3525,

AM = (0.408 X 3525) = 1438,

J = (0.340 X 3525) = 1199,

Y = (0.239 X 3525) = 843.

4.1.1.4 Prior pdfs for parameters

Because they are determined by random quantities derived in the previous period, the number of adult females, adult males, female yearlings and male yearlings *at time period 1* are assumed to follow uniform distributions narrowly ranging about their initial values. The values below are 1 standard deviation around the estimated mean values of the population counts in 1983 (Owen-Smith *et al*, 2005).

Thus,

AF_1 ~ uniform (3503, 3543),

AM_1 ~ uniform (1417, 1457),

(YF_1) ~uniform (816, 866), and

$(YM_1) = (YF_1)$.

Uniform priors are chosen for the counts as a non-informative benchmark against which to test more informative or specific distributions for the process if necessary. Thereafter, the model is as described above. The number of juveniles born, juvenile recruitment, is modelled as $J_t \sim \text{Poisson}(\kappa * AF_t * \pi)$ where, κ is the mean number of offspring per adult female, and π is the proportion of adult females that breed. Kudu generally bear only one offspring per season therefore $\kappa_i = 1$ and from above we see that $\pi_i = 0.340$.

4.1.1.5 Observation Process pdf

The observation process is assumed to follow a normal distribution with mean equal to the true mean and a standard deviation of 400 around the true count:

$$f(y_t \mid n_t, \theta) \sim Normal(n_t, \sigma_i^2).$$

The counts are modelled using binomial distributions. The observation errors in the counts were modelled using a normal distribution, as there is no reason to believe that these observation errors would arise from a skew distribution. As shown in Table 1, the counts recorded are large numbers, the minimum overall count being 3,172, making the use of the normal distribution theoretically viable, as well as being computationally convenient.

4.2 The Estimated Total Population of Kudu in The KNP

The initial model fitted was as described above, however the results showed signs of model misfit (discussed below) and as a result some adjustments were made to the initial value assumptions and model update parameters in order to improve the

overall model fit. The initial model fitted is referred to as model 1 below and the updated model is referred to as model 2. The results of both models are discussed with respect to the diagnostics obtained, in order to illustrate the sensitivity of the results to the set burn-in period, sample size and thinning rate. For ease of comparison only limited output is shown and discussed in this chapter and the rest is contained in the appendices. The initial model fitted to the data was updated until the Markov Chain had passed the burn - in period (*see Section 3.3.1.1*). The burn – in period was evaluated based on an inspection of the autocorrelation function of the estimates. Burn –in occurred after 80,000 iterations for model 2. The burn – in period should be sufficiently long to ensure that the MCMC has converged and that all samples that are collected after discarding information from the burn-in period are samples from the posterior distribution of interest. In the case of model 1 (the first model estimated) 250 sample observations were drawn at a thinning rate of 50 observations after the burn-in period of 50,000 iterations had been observed. Note that the total number of iterations used in the updating process was 250*50 or 12,500 iterations.

The initial values and prior distribution assumptions used in building the model are described in Sections 4.1.1.3, 4.1.1.4 and 4.1.1.5 above. The following initial values were used for the different gender and life-history stage counts. These initial values were the estimated mean counts over the sample period, 1983 to 1993. Based on the model described in Section 4.1, initial values were required for the juvenile survival rate, the female yearling survival rate, the adult female survival rate, the number of juveniles, the number of female yearlings and the number of adult females. The initial values for the survival rates and the counts are shown in Section 4.1.1.4 above.

The following changes were made to the initial model specification and update parameters in the second model run:

1. The burn-in period required to start recording sample values was increased from 50,000 iterations to 150,000 iterations to ensure that the MCMC had converged before sample values were recorded.
2. The initial sample size was 250 observations at a thinning rate of 50 observations; this was increased to 2100 observations at a thinning rate of 200 observations. This was done to ensure that the sample drawn was independent and to resolve the problem of high autocorrelation of estimated parameters present in the first model output.
3. Changes were made in the initial values used in the model updating process. The initial model used constant values for the initial assumptions on count and survival rate. These constant values were based on the estimated mean survival rates required for zero population growth (Owen-Smith *et al,* 2005). For the second model run, Model 2 below, the estimated values for survival rate and count in each year derived from the first model run, Model 1 below, were used as initial values or prior knowledge. These were used as initial values because although the model output indicated some problems with the model diagnostics, the estimates calculated for each year were considered to be more informative than the constant mean value used to initiate the model updating process in the initial specification. The priors used to initialise the second model run were an improvement on the constant mean values used in the initial model run because they are from a theoretically more appropriate model.

The following initial values were used for the second model run:

Year	Number of Juveniles	Juvenile survival rate
1983	1451	0.9175
1984	1522	0.7896
1985	1661	0.7688
1986	1725	0.7161
1987	1556	0.8104
1988	1430	0.6029
1989	1244	0.5341
1990	1050	0.6402
1991	833.8	0.6983
1992	722.8	0.7253
1993	472	0.9758

Table 4: Initial values- juveniles (model 2)

Year	Number of Female Yearlings	Female Yearling Survival Rate
1983	512	0.4669
1984	664.6	0.6219
1985	601.1	0.4716
1986	639.1	0.3577
1987	617	0.3544
1988	625.6	0.421
1989	428.6	0.3843
1990	333.9	0.3784
1991	332.1	0.4253
1992	289.7	0.3694
1993	264	0.3947

Table 5: Initial values – female yearlings (model 2)

Year	Number of Adult Females	Adult Female Survival Rate
1983	4274	0.9864
1984	4458	0.9933
1985	4845	0.9875
1986	5069	0.8565
1987	4565	0.8735
1988	4201	0.8142
1989	3672	0.7977
1990	3082	0.7561
1991	2454	0.8149
1992	2132	0.6051
1993	1383	0.1172

Table 6: Initial Values of the Adult Female Survival Rate - Model 2

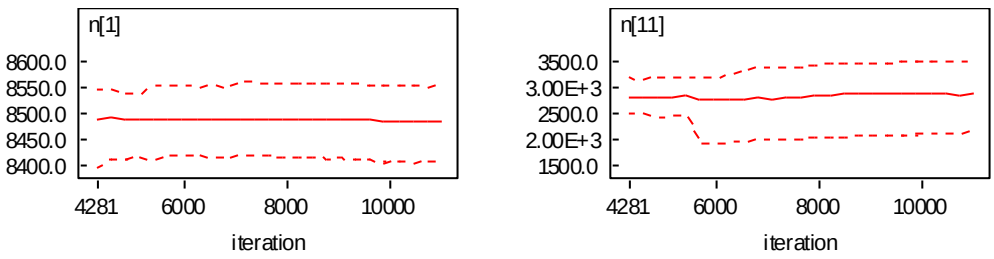
The initial values for the female yearling survival rate and the number of female yearlings are shown below. The number of male yearlings is determined as the difference between the number of juvenile survivors in the previous year and the number of female yearlings in this year. The male yearling survival rate is assumed to be equal to the female yearling survival rate.

The initial values for the adult female survival rate and the number of adult females are shown below. The number of male adults is assumed to be a proportion of 0.408 of the total number of adult females based on the mean ratio of adult males to adult females and adult male survival is assumed to be 0.85 of adult female survival, as in model 1.

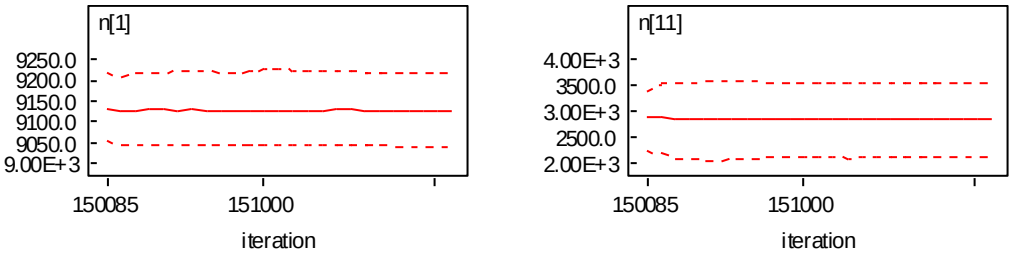
The diagnostics obtained from the second model run were highly satisfactory, indicating a significant reduction in MC error, better convergence diagnostics and also showing that the problem of autocorrelation of estimated sample values had been resolved. The model output is presented below. Because of the high number of graphs only the graphs for year 1 - 1983 (denoted by n[1] in the figures) and year 11 - 1993 (denoted by n[11]) are shown in this chapter, the rest are shown in Appendix D.

4.2.1 Running Quantile Plots Total Population of Kudu in The KNP

Model 1



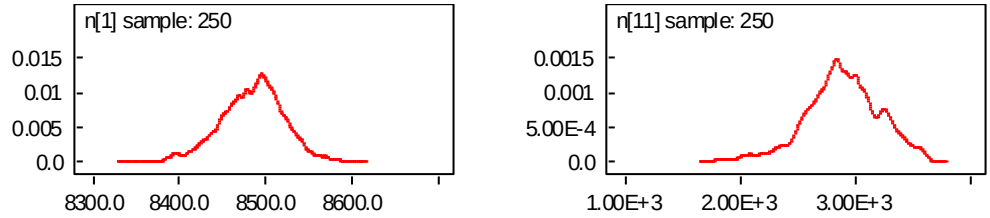
Model 2



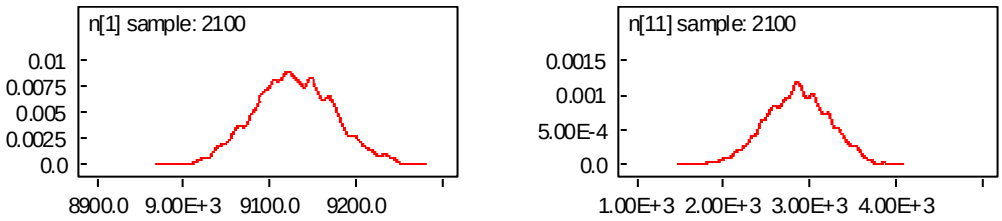
The straight running quantile plots for the total model count for model 2 indicate that model had converged after the burn-in period. The 95% confidence intervals for the initial model run were not as stable.

4.2.2 Kernel Density Estimates Marginal Posterior Distribution Function of Total Population

Model 1



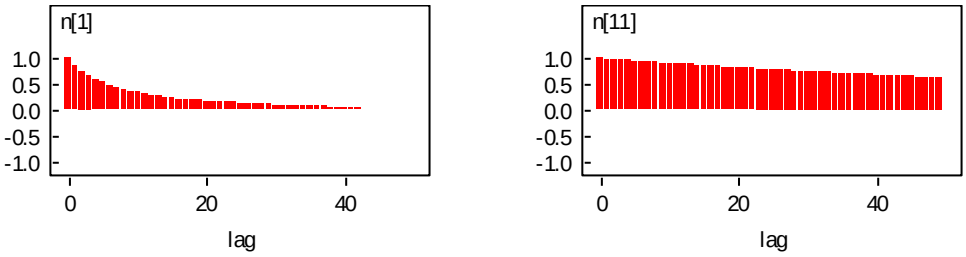
Model 2



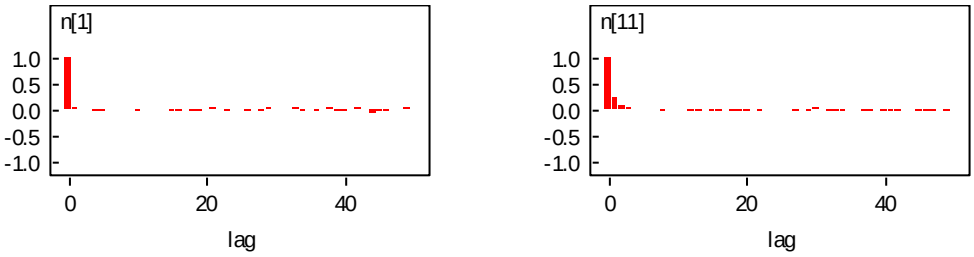
The kernel density estimate for the total count is a uni-modal bell shaped distribution which seems to be normally distributed around the estimated sample mean.

4.2.3 Autocorrelation Functions Estimated Total Population

Model 1



Model 2



The autocorrelation functions of the estimated counts in each year in Model 1 indicated severe autocorrelation of estimates, which showed that the model had not converged successfully and that the samples were not independent. The autocorrelation functions obtained in model 2 indicate that the model for the total count has converged as there is no significant autocorrelation between successive sample values from the Markov Chain. The mean estimated Markov Chain error throughout the sample period was 10, which is a relatively small magnitude of variance in comparison to the total population size counted. Summary statistics for the estimated count are shown in Appendix E. The estimated total count is 9128 in 1983 and 2856 in 1993.

4.2.4 Estimated Bias in Total Census Counts

The tables below show the estimated measurement error in the EAS survey counts

based on a comparison of the EAS survey counts to the estimated totals from the model. This output shows the estimated sampling error that occurred in each census year based on our model assumptions (either an over-count or under-count) and indicates what the estimated true total is based on the model results. This output can be used to correct the census counts in each year.

Year	Estimated Total Count from Model 1	Estimated Total Count from Model 2	Difference
1983	8493	9128	635
1984	9028	9193	165
1985	9566	9702	136
1986	9928	9995	67
1987	8945	8844	-101
1988	8288	8129	-159
1989	6996	7030	34
1990	5791	5867	76
1991	4720	4710	-10
1992	4102	4015	-87
1993	2824	2856	32

Table 7: Difference in estimated EAS Counts Model 1 and Model 2

Year	EAS Count	Estimated Total Count from Model 2	Estimated Survey Error
1983	7005	9128	-2123
1984	8661	9193	-532
1985	10432	9702	730
1986	10760	9995	765
1987	8786	8844	-58
1988	8214	8129	85
1989	6935	7030	-95
1990	5967	5867	100
1991	4706	4710	-4
1992	3967	4015	-48
1993	3127	2856	271

Table 8: Estimated Bias in EAS Counts

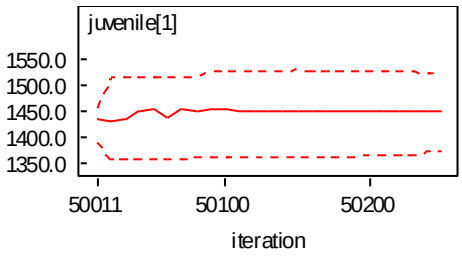
The estimated total in each year from the model is made up of the sum of estimated juveniles, male and female yearlings and male and female adults. The results for

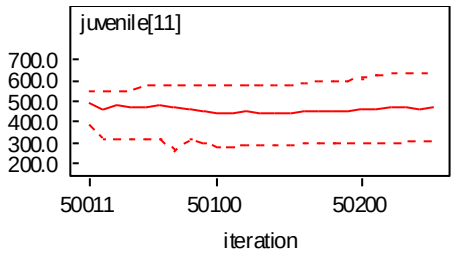
these sub-models for the counts and survival rates which contribute to the total count are shown in the sections below.

4.3 Estimated Number of Juveniles

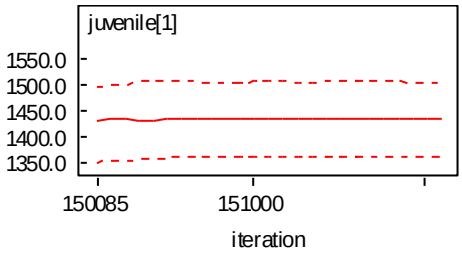
4.3.1 Running Quantiles Estimated Number of Juveniles

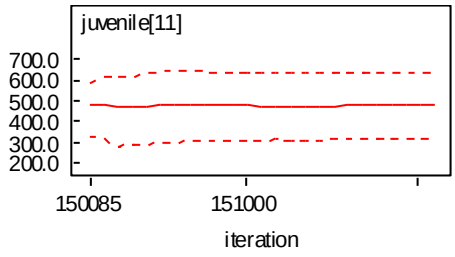
Model 1





Model 2

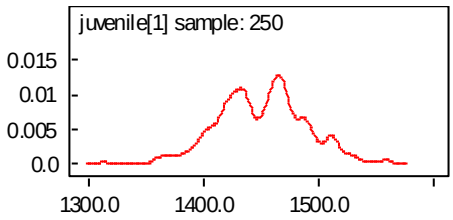


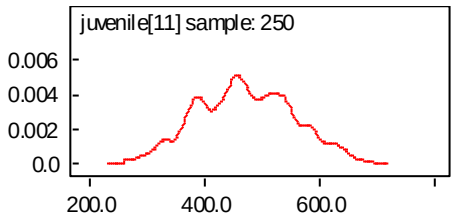


The initial quantile plots for the number of juveniles were unstable indicating that the sub-model had not converged. Model 2 quantile plots show a straight line which indicates convergence.

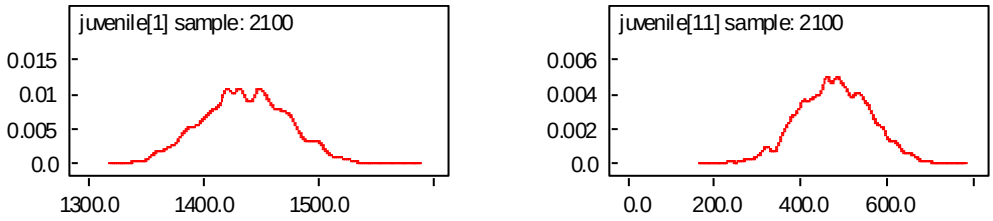
4.3.2 Kernel Density Estimates Number of Juveniles

Model 1





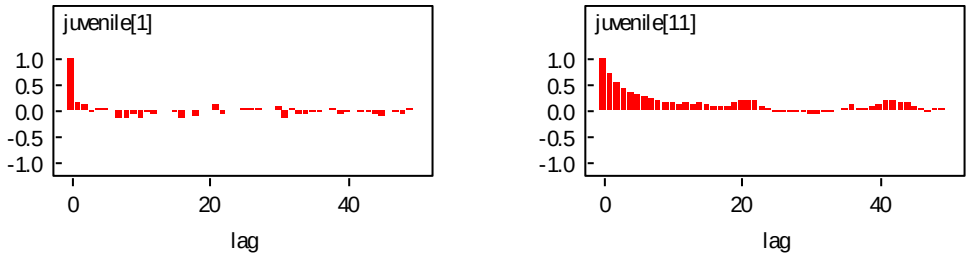
Model 2



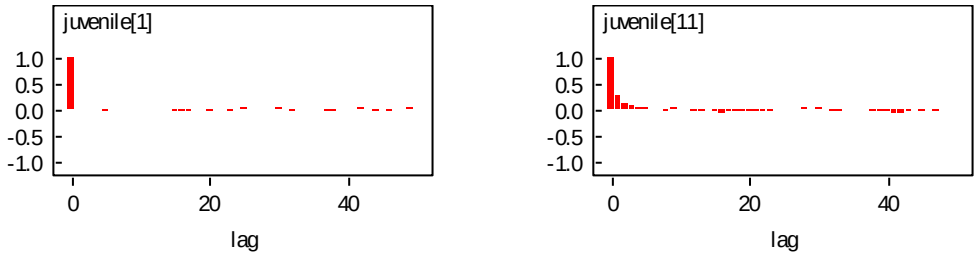
The kernel density estimates for model 2 also improve on model 1 kernel density estimates in that they have fewer peaks and a smoother outline.

4.3.3 Autocorrelation Functions Number of Juveniles

Model 1



Model 2

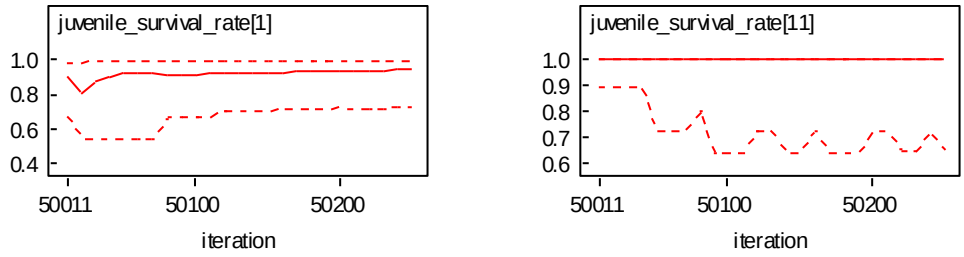


The relatively high autocorrelations derived in the first model run are resolved in the second model run. The model diagnostics indicate that this sub-model has converged. The estimated count in 1983 was 1436 and 477 in 1993. The maximum estimated model simulation error was 9 which is small relative to the total counts.

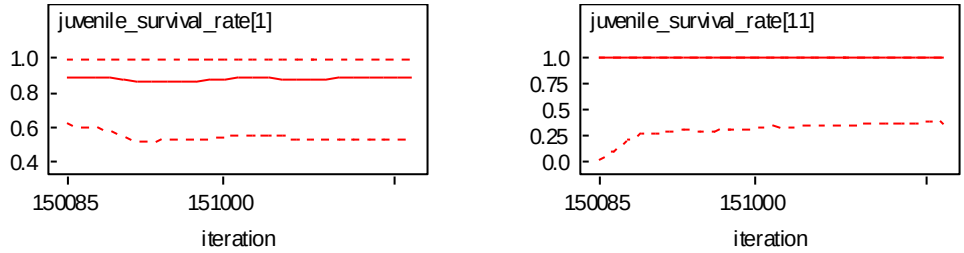
4.4 Estimated Juvenile Survival Rate

4.4.1 Running Quantiles Estimated Juvenile Survival Rate

Model 1



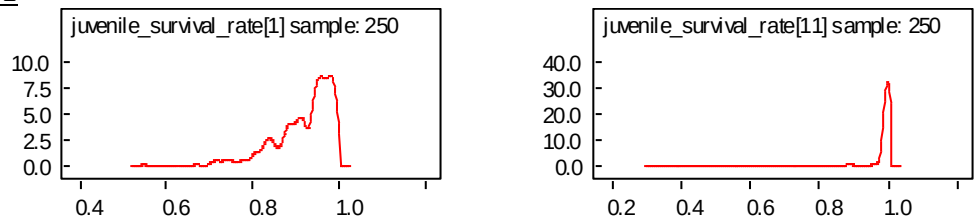
Model 2



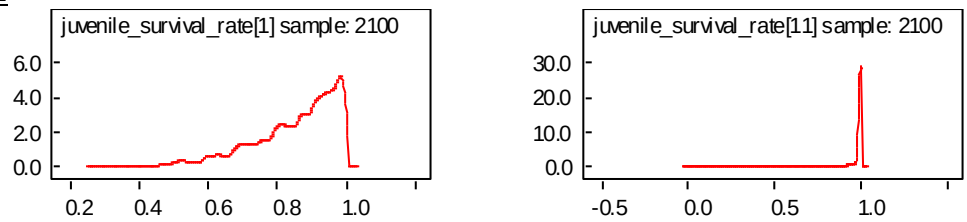
The high levels of instability in the quantile plots in model 1 are improved upon significantly in the re-specified model.

4.4.2 Kernel Density Estimate: Marginal Posterior Distribution Function of Juvenile Survival Rate

Model 1



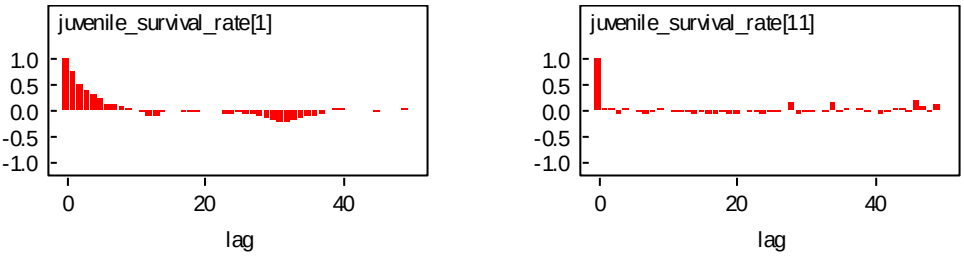
Model 2



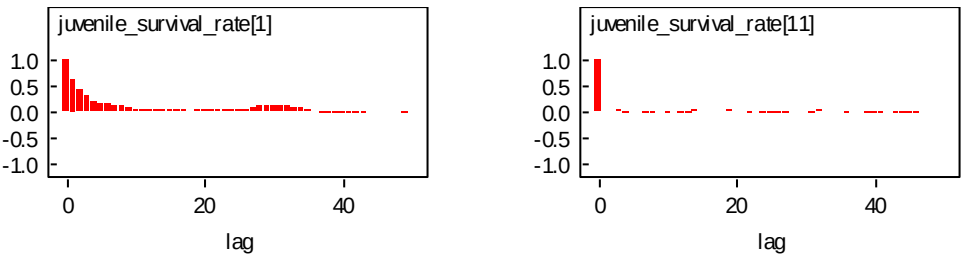
The kernel density estimates for juvenile survival did not change significantly between model 1 and 2, although the model 2 density estimate is more stable than the initial estimate.

4.4.3 Autocorrelation Function of Juvenile Survival Rate

Model 1



Model 2

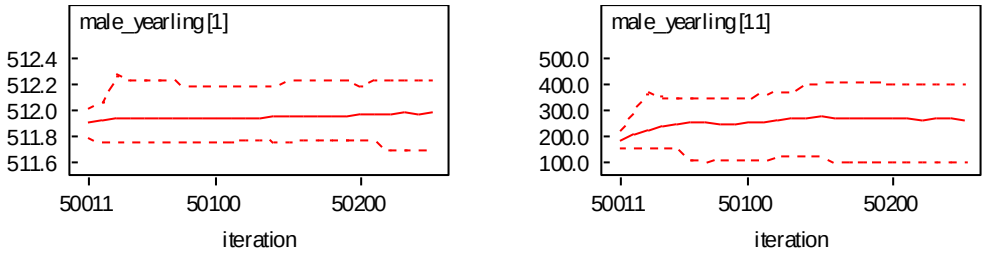


Some of the years have slightly elevated autocorrelation functions however these are not at a level that would indicate major problems with model convergence, as this autocorrelation rapidly decreases to zero.

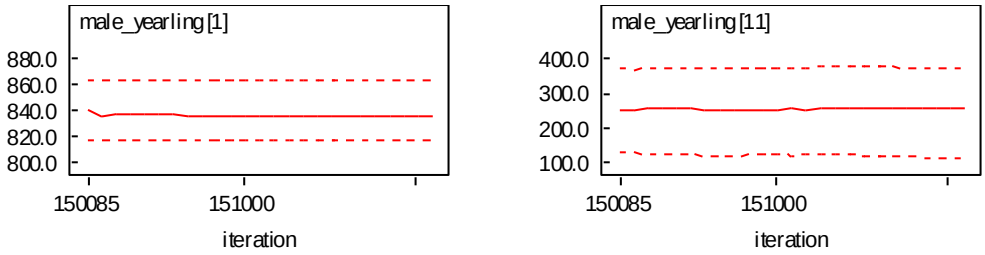
4.5 The Estimated Number of Male Yearlings

4.5.1 Running Quantiles Number of Male Yearlings

Model 1



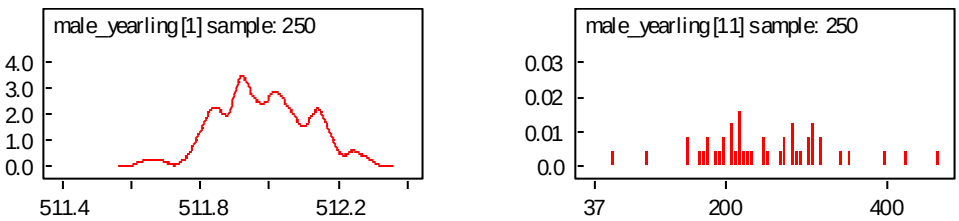
Model 2



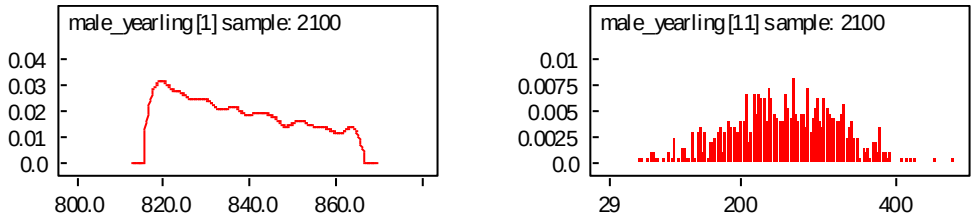
Model 2 quantile plots for male yearling survival indicate that the sub-model converged.

4.5.2 Kernel Density Estimate: Marginal Posterior Distribution Number of Male Yearlings

Model 1



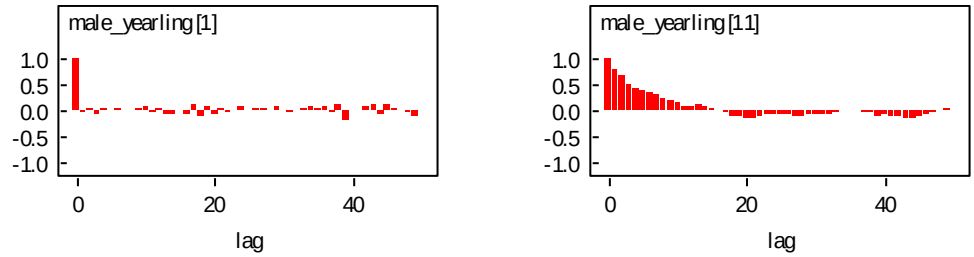
Model 2



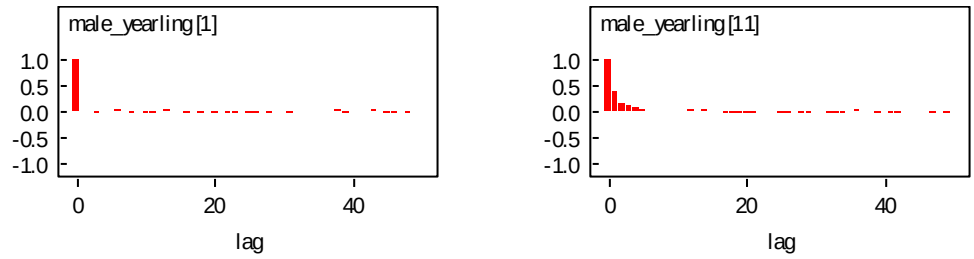
The kernel density estimates derived from model 2 are more stable than the estimates obtained from the initial model run.

4.5.3 Autocorrelation Function Number of Male Yearlings

Model 1



Model 2

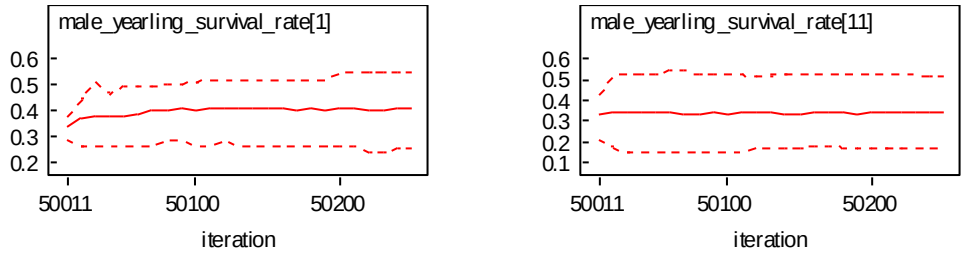


Autocorrelation functions indicate that convergence and sample independence improved for this sub-model for the re-specified model. The estimated number of male yearlings is 837 in 1983 and 251 in 1993. The maximum estimated error due to simulation was 19.

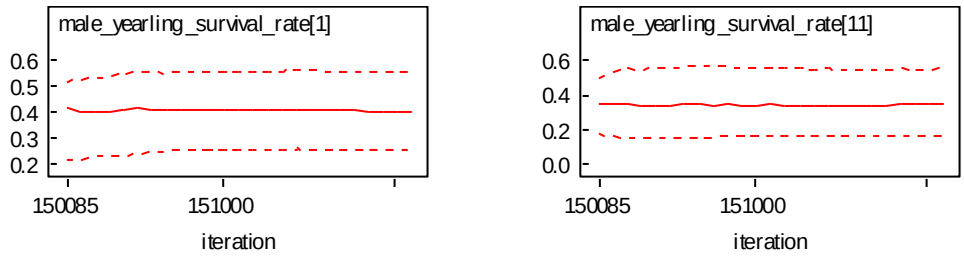
4.6 Estimated Male Yearling Survival Rate

4.6.1 Running Quantiles Male Yearling Survival Rate

Model 1



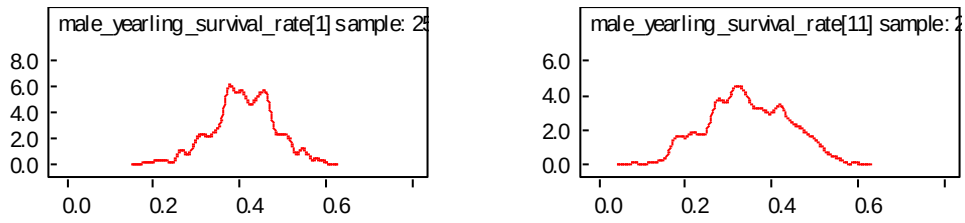
Model 2



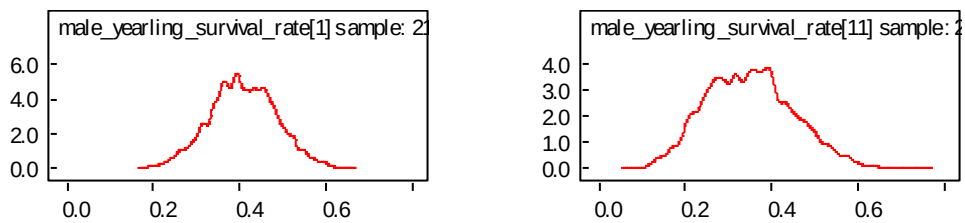
The quantile plots indicate that the model has converged to the posterior distribution of interest

4.6.2 Kernel Density Estimates Male Yearling Survival Rate

Model 1



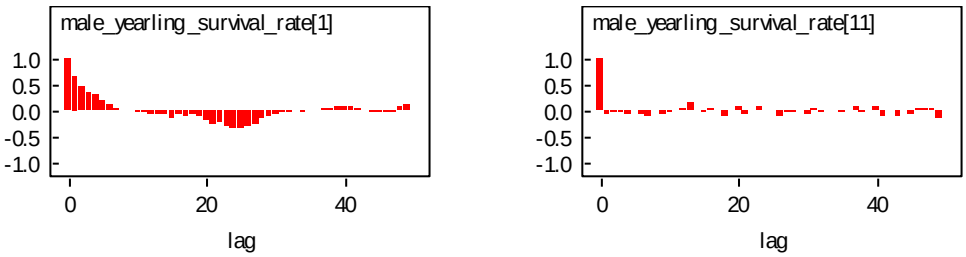
Model 2



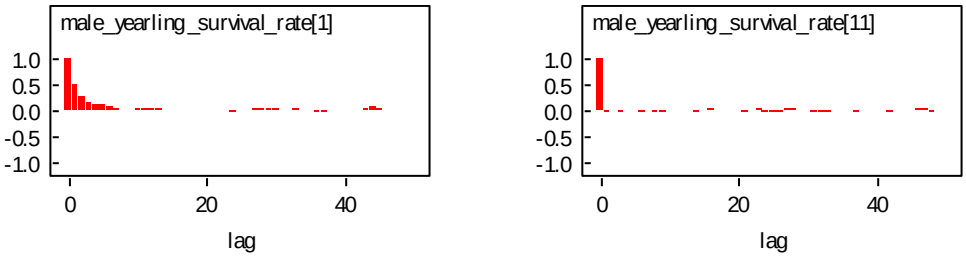
The kernel density estimates for the male yearling survival rate are stable and single-peaked which shows that the distribution has been successfully estimated.

4.6.3 Autocorrelation Functions of Male Yearling Survival Rate

Model 1



Model 2

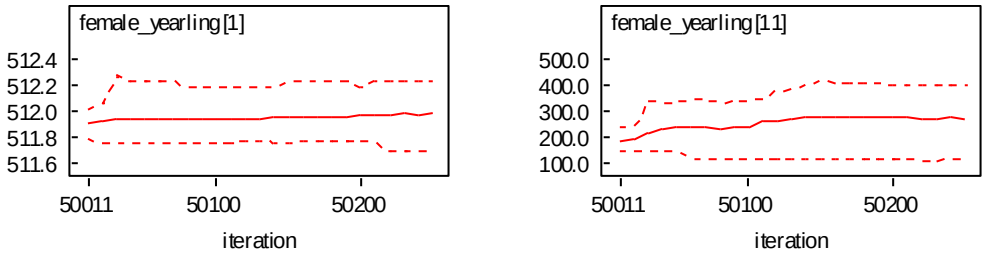


The autocorrelation between estimated sample values was significantly improved in the re-specified model run. The diagnostics indicate that this sub-model has converged. Estimates of the survival rates for each year and the estimated simulation errors are shown in appendix E.

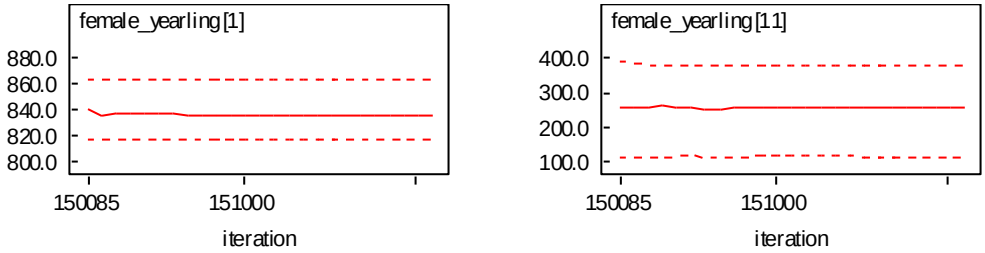
4.7 Estimated Number of Female Yearlings

4.7.1 Running Quantiles of Number of Female Yearlings

Model 1



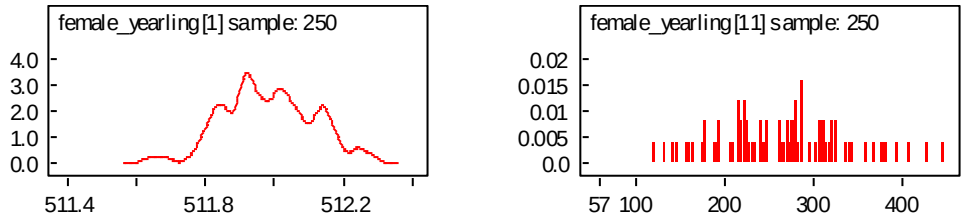
Model 2



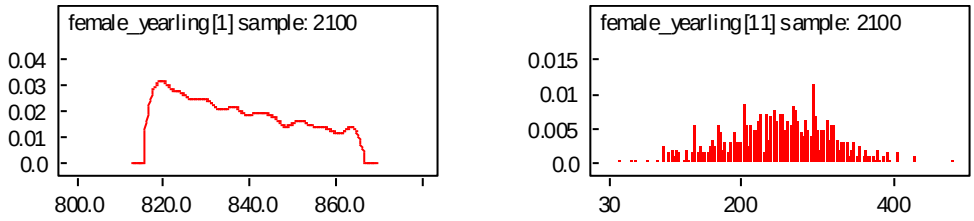
The running quantile plots for the number of female yearlings indicate that the sub-model has converged.

4.7.2 Kernel Density Function: Marginal Posterior Distribution Number of Female Yearlings

Model 1



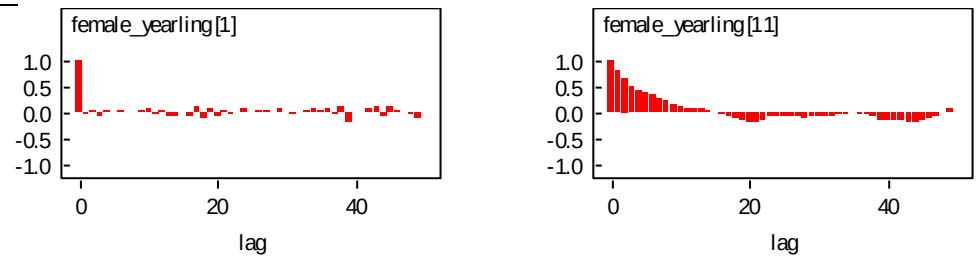
Model 2



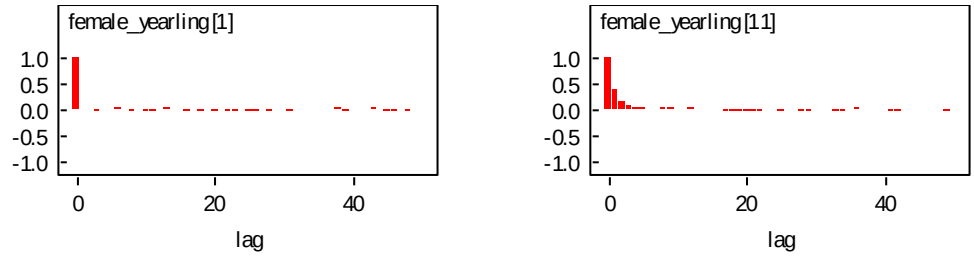
The kernel density estimates obtained from the second model run are more stable than the estimates obtained from the initial model run.

4.7.3 Autocorrelation Functions of Number of Female Yearlings

Model 1



Model 2

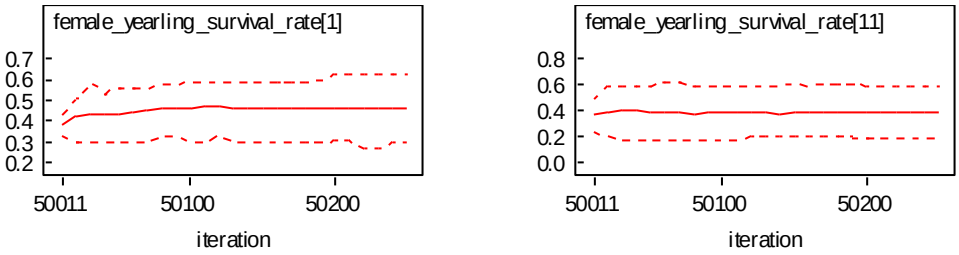


The estimated number of female yearlings was 837 in 1983 and 252 in 1993. The maximum estimated simulation error was 19 animals in 1987 when the mean estimated count was 609 animals. The model diagnostics do not indicate any significant problems with the model fit.

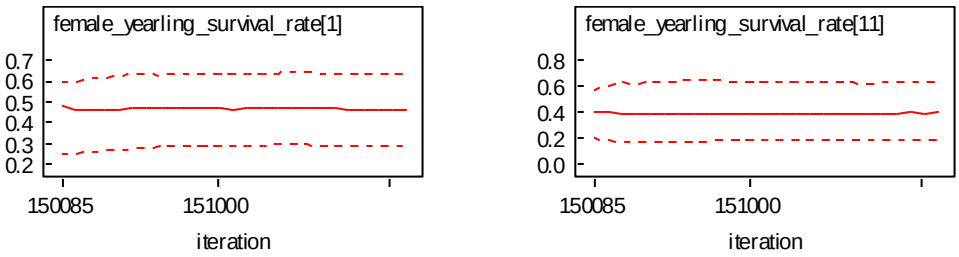
4.8 Estimated Female Yearling Survival Rate

4.8.1 Running Quantiles Female Yearling Survival rate

Model 1



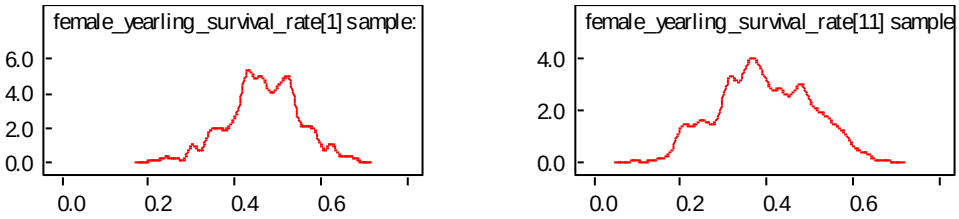
Model 2



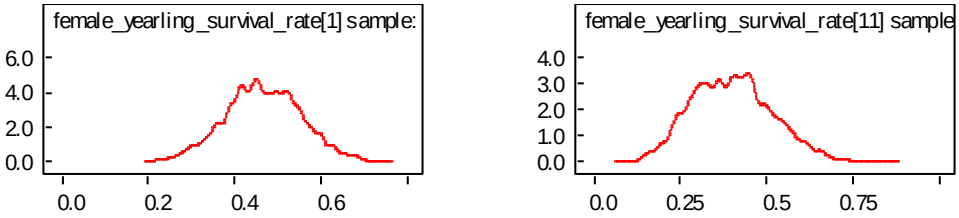
The running quantile plots are straight indicating that the sub-model has converged.

4.8.2 Kernel Density Estimates Female Yearling Survival Rate

Model 1



Model 2

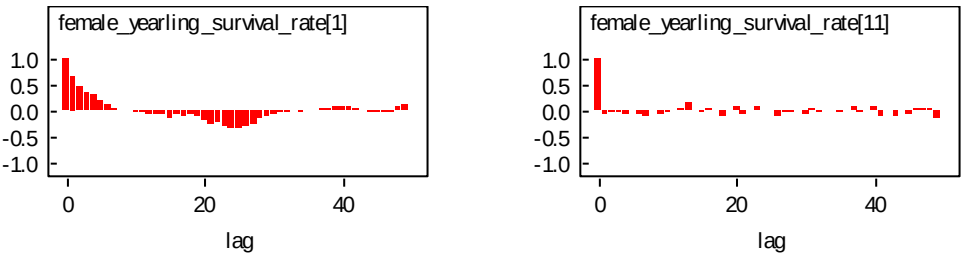


The kernel density estimates are smooth and uni-modal indicating that the posterior

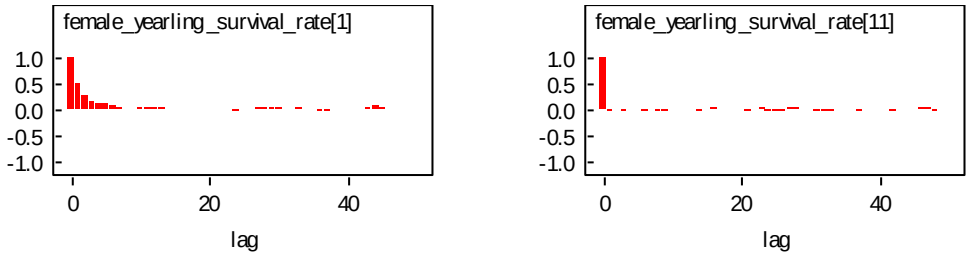
distribution function of the female yearling survival rate has been successfully estimated.

4.8.3 Autocorrelation Functions of Female Yearling Survival Rate

Model 1



Model 2

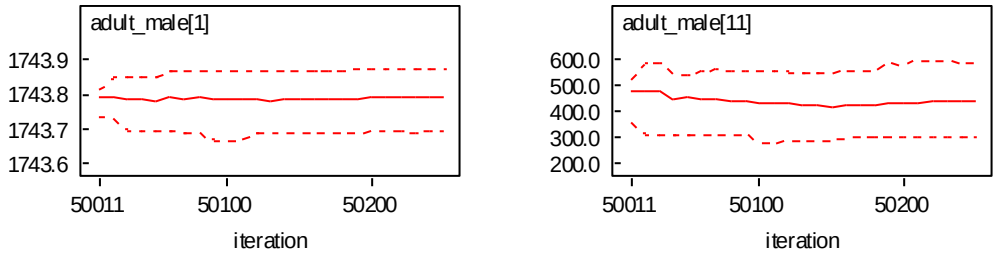


The problem of autocorrelation in estimated sample values was significantly improved in the re-specified model. The estimated survival rate was 0.47 in 1983 and 0.39 in 1993. The maximum estimated survival rate was 0.57 in 1984 and the minimum was 0.36 in 1992. The model diagnostics do not indicate any significant problems with the model fit.

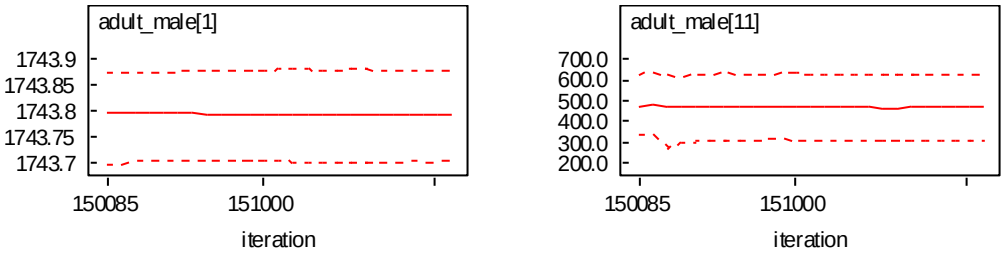
4.9 Estimated Number of Male Adults

4.9.1 Running Quantiles Number of Male Adults

Model 1



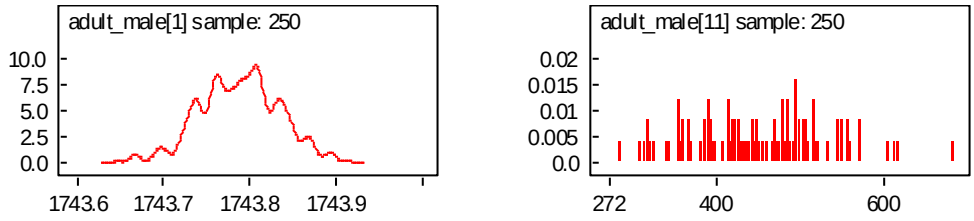
Model 2



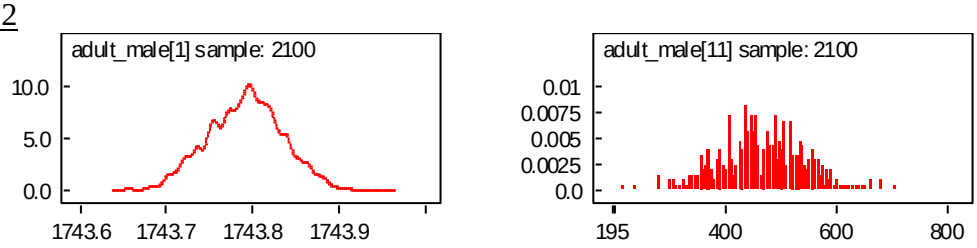
The running quantile plots indicate that the model has converged.

4.9.2 Kernel Density Estimate: Marginal Posterior Distribution Number of Adult Males

Model 1



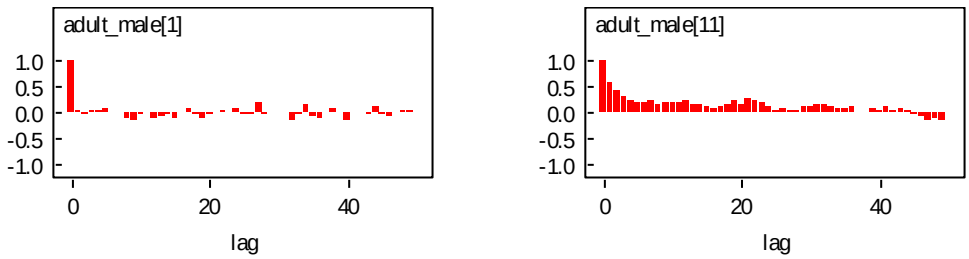
Model 2



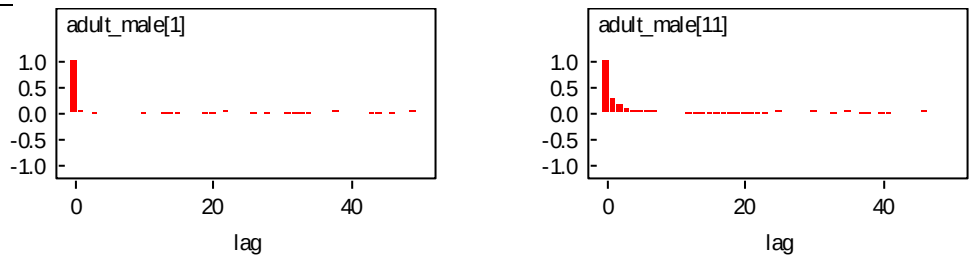
The kernel density estimates derived from the re-specified model are more stable.

4.9.3 Autocorrelation Functions Number of Male Adults

Model 1



Model 2

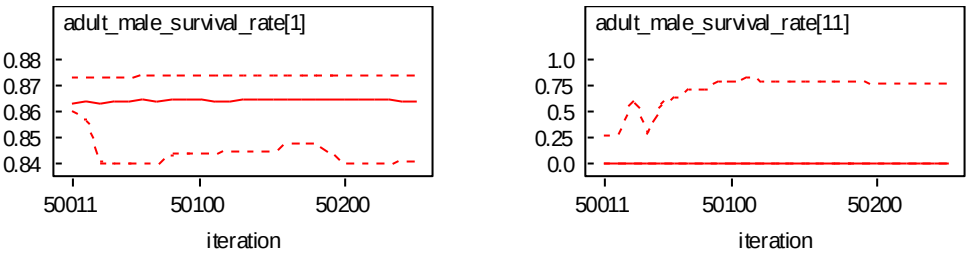


The slightly elevated correlations between successive estimated counts in some years are resolved in the re-specified model. The estimated number of adult males was 1477 in 1983 and 470 in 1993. The maximum estimated simulation error was 8 animals in 1987 when the total count was 1589. The model diagnostics indicate that the model has successfully converged to the posterior distribution of interest.

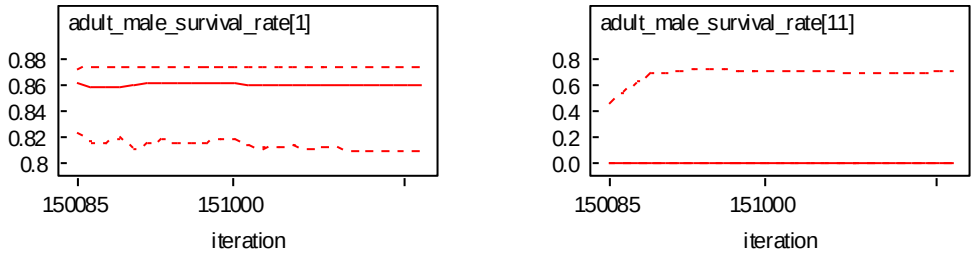
4.10 Estimated Adult Male Survival Rate

4.10.1 Running Quantiles Adult Male Survival Rate

Model 1



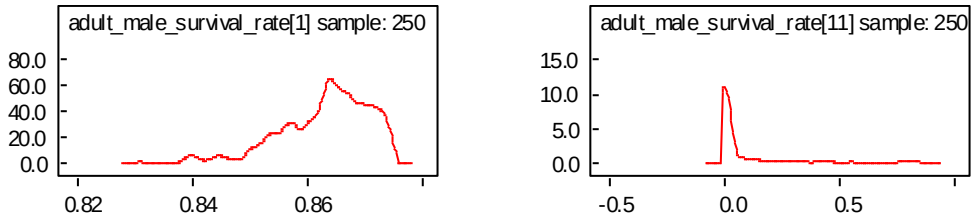
Model 2



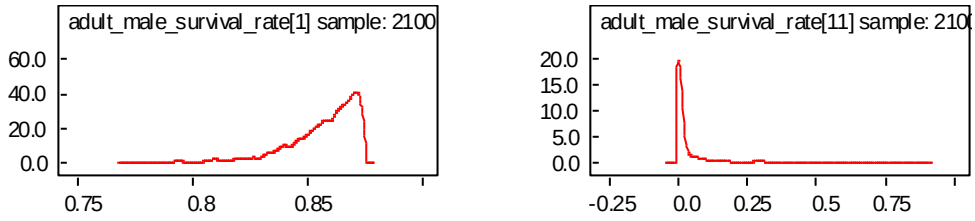
The highly unstable quantile plots in the initial model run are resolved in the re-specified model.

4.10.2 Kernel Density Estimate: Adult Male Survival Rate

Model 1



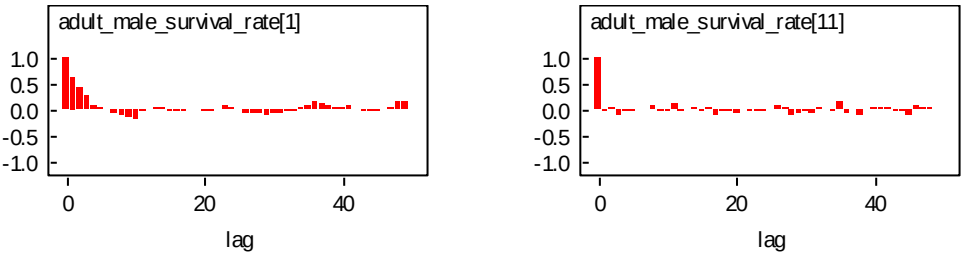
Model 2



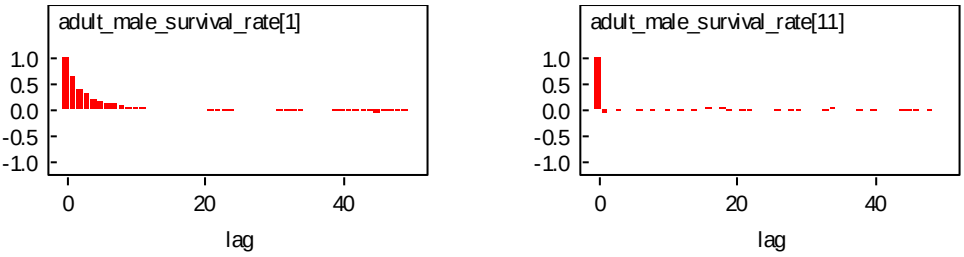
The kernel density estimates from the re-specified model are more stable.

4.10.3 Autocorrelation Functions Adult Male Survival Rate

Model 1



Model 2

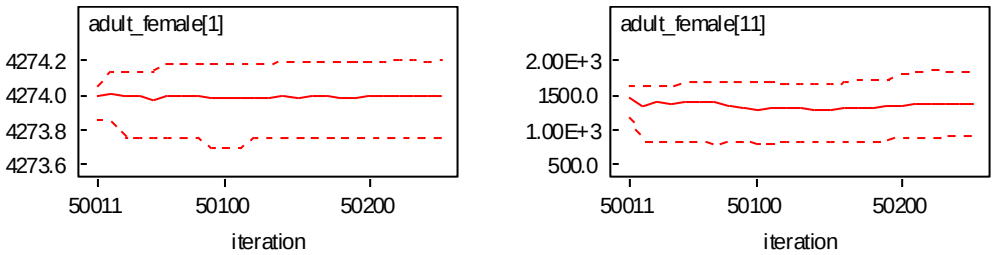


The mean estimated adult male survival rate was 0.86 in 1983 and 0.1 in 1993. The mean estimated Markov Chain error was less than 0.01. The model diagnostics did not indicate any major problems with model fit.

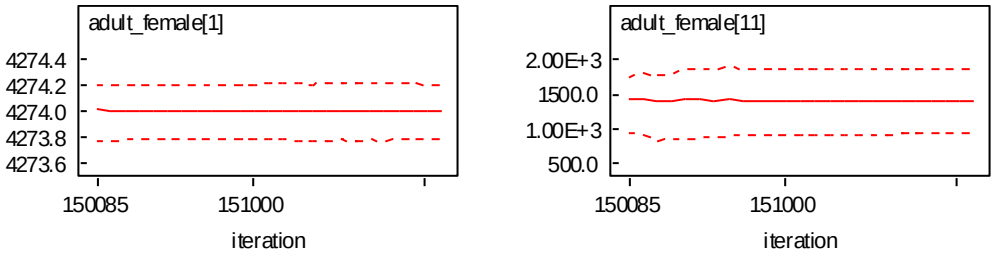
4.11 Estimated Number of Adult Females

4.11.1 Running Quantiles Number of Adult Females

Model 1



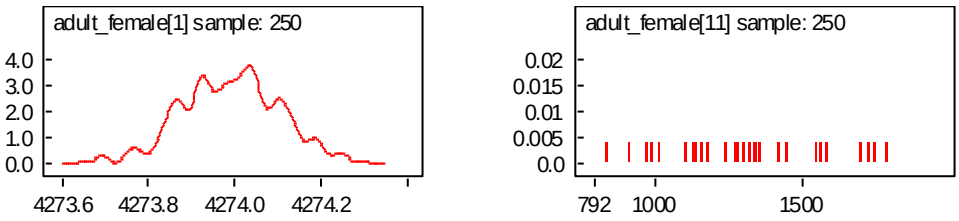
Model 2



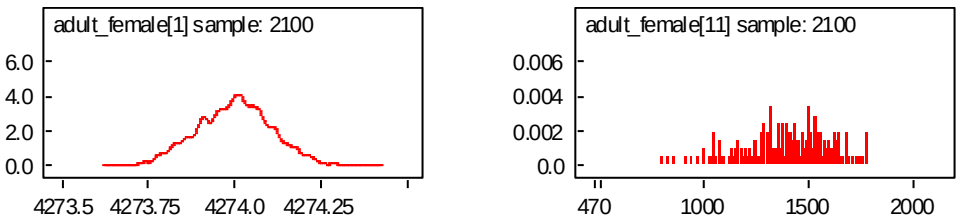
The running quantile plots indicate that the model has converged.

4.11.2 Kernel Density Estimates: Number of Adult Females

Model 1



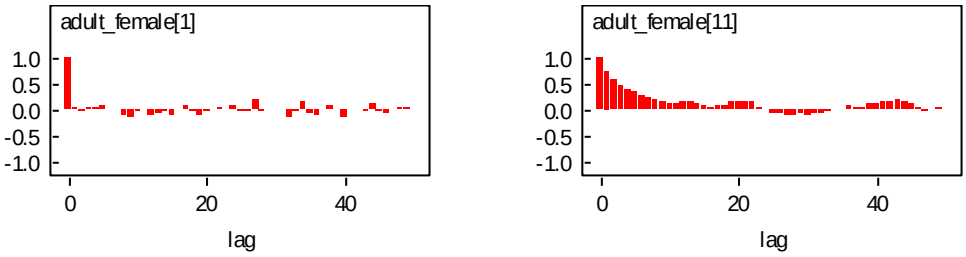
Model 2



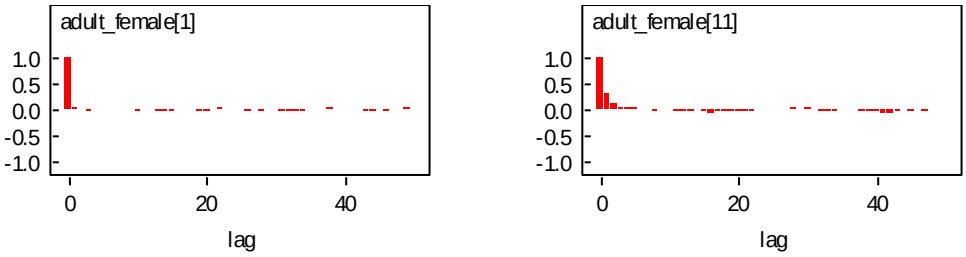
The kernel density estimates from the re-specified model are more stable than the estimates derived from the initial model run.

4.11.3 Autocorrelation Function Number of Adult Females

Model 1



Model 2



The problem of autocorrelation of estimates was resolved by the re-specified model. The estimated number of adult females was 4274 in 1983 and 1404 in 1993. The maximum count of adult female was in 1986 when the estimated total count grew to 5109 animals. The model diagnostics do not indicate any significant problems with the model fit. The full model summaries and diagnostics are shown in Appendix B to E.

Chapter 5

Discussion of Results, Conclusion and Recommendations

In this chapter we discuss the results and make conclusions and recommendations.

The MCMC chain specified in Model 2 was updated for a burn in period of 150,000 iterations. After this 2100 iterations with a "thinning" of 200 lags were observed in order to select the sample used to estimate the model parameters. Thinning means that not all sample values are recorded, rather values are recorded at a regular interval determined by the researcher; in this case every 200th value. This process reduces the autocorrelation between sample values and increases the independence of the sample. Estimating a model of this size is quite onerous in terms of the computing power required. The model took 9 days to complete updating.

5.1 Discussion of Results

The research problem investigated was, firstly, to formulate a model for the population dynamics of Kudu in the KNP that would account for the measurement errors that arise from the aerial survey enumeration procedure used to conduct the EAS and, secondly, to reconcile the overall census counts to a model of population structure which allows us to estimate the number of animals and the yearly survival rates in each gender and age classification from the overall census count which does not provide this information. The model obtained was re-specified based on the diagnostics obtained, in order to illustrate the sensitivity of the results to the set burn-in period, sample size and thinning rate.

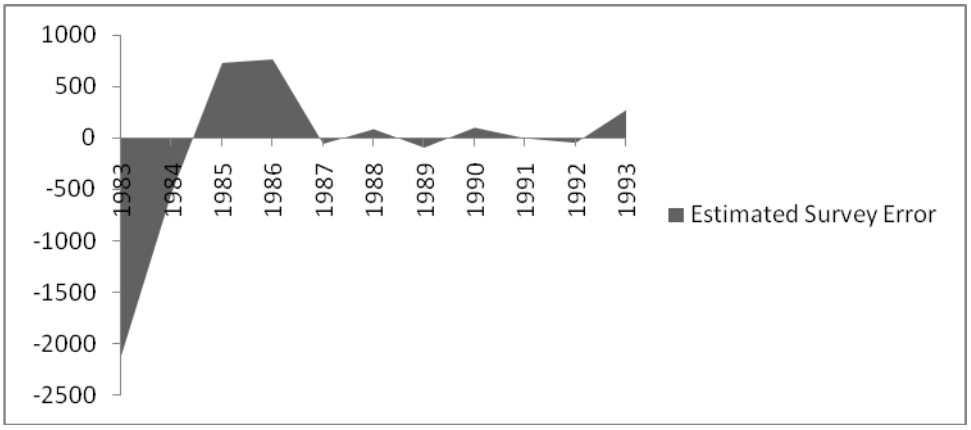


Figure 2: Estimated Survey Error

Figure 2 above shows the estimated census errors derived from the model. The data indicate that the largest errors in the total EAS counts most likely occurred in 1983 and again in 1985 and 1986. This was based on the model for the error process that took into account wet and dry season rainfall as environmental covariates. The errors shown are cyclical with under-counts being followed by over-counts. This is in accordance with the effect described in Jachmann (2002), whereby under-counts in one year spuriously inflate the growth recorded in the next year and an over-count leads to a spurious population decrease in the following census. The model for the population structure that was investigated using the hierarchical model allowed us to estimate the survival rates below for the different gender and life-history stages within the population. According to these estimates survival was lowest amongst yearlings and juvenile survival was increasing towards the end of the period.

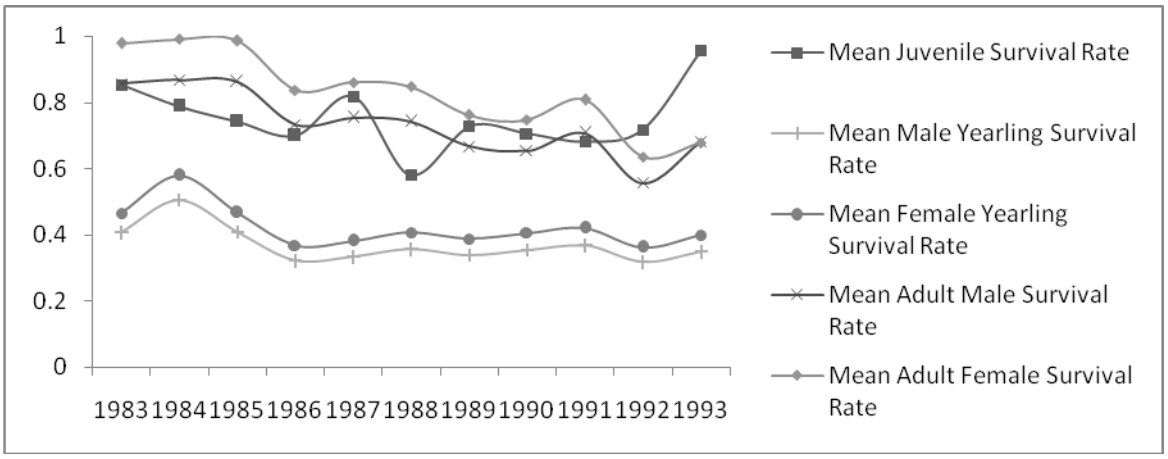


Figure 3: Estimated Survival Rates for the different Gender and Life-History Stages

The initial counts were recognised as being highly susceptible to bias for a number of reasons including factors which affect count efficiency such as weather and habitat conditions and the difficulty of observing juvenile animals from the air. We included information in the model that would further reduce the variability in the estimated counts by including sub-models for the gender and life-stage history of the animals including the appropriate covariates that would determine the counts and survival rates of the different genders and life-history stages. The research made extensive use of the research carried out by Professor Norman Owen-Smith and it would not have been possible to carry out this research without the data he provided on the actual EAS counts and the recorded values of the covariates used such as rainfall in the KNP over the census period. The choice of covariates to include in the model was also guided by previous studies by Owen-Smith *et al* (2005), which investigated the significant determinants of ungulate survival during the period covered by this research report. Wet and dry season rainfall were the only covariates that were included in the model as they were the only parameters with significant coefficients in a regression analysis of possible determinants of survival that included density dependence and predation among other variables (Owen-Smith *et al*, 2005).

The estimated regression equations relating wet and dry season rainfall to survival were used to determine bounds for the survival rate of each gender and life-history stage in a Beta-distribution model for the survival rates - and thus constituted a part of our prior information on the process. From a series of total counts spanning the period from 1983 to 1993 we were able to estimate an adjusted total count for each year which corrected for estimated measurement error; in addition we were also able to estimate the counts and survival rates for juveniles, male and female yearlings and male and female adults.

The model diagnostics, which included 95% posterior confidence intervals for the parameter estimates, estimated simulation error, kernel density estimates and autocorrelation functions indicate that the re-specified model for the total count converged and that the model fit was very good. Generally the model fit for all the sub-models was also very good. Some of the autocorrelation functions for estimated counts and survival rates for the sub- models such as the autocorrelation for the juvenile survival rate in 1986 and 1987, and the autocorrelations for number of juveniles in 1986 and also the number of male yearlings 1985 and 1986, do show slightly elevated levels. However these autocorrelations decrease to zero quite rapidly by the 10^{th} lag or so in the chain at most and given that 420 000 iterations of the chain have been performed and that the slightly elevated autocorrelations are not corroborated by any of the other diagnostics we can conclude that even these sub-models have converged.

5.2 Conclusion and recommendations

In general, the WinBUGS model has shown how the Bayesian framework can be

successfully utilised to build a model that can incorporate environmental covariates and that can explicitly include a theorised relationship between the estimated counts and the true population in order to correct for measurement errors in a census count of ungulates. The model estimates a high number of parameters simultaneously and achieves a better model fit (based on autocorrelations and kernel density estimates) at estimating the survival rates as compared to the actual counts. There are no definitively correct estimates of the model parameters in the ecological literature, but the estimates from this model do not conflict with those in the literature. The final model diagnostics indicate that this model was fitted to the data successfully. It is possible that other assumptions on the distributions could give models as good as the model estimated. The statistical framework provided is very flexible and allows further researchers to easily define distributional assumptions and relationship equations to test on the data based on the current ecological theories.

References

Besbeas P, Lebreton JD & Morgan BJT. (2003). *The Efficient Integration of Abundance and Demographic Data*. Applied Statistics 52, pp 95-102.

Brooks SP, King R & Morgan BJT. (2004). *A Bayesian approach to combining animal abundance and demographic data*. Animal Biodiversity and Conservation, 27(1), pp 515-529.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL & Thomas I. (2004) Advanced Distance Sampling: Estimating abundance of biological populations. Oxford University Press, Oxford. 416 pages.

Buckland ST, Newman KB, Thomas L & Koesters NB. (2004). *State-space models for the dynamics of wild animal populations*. Ecological Modelling 171, pp 157-175.

Bulmer MG. (1975). *The statistical analysis of density dependence.* Biometrics, 31 (4), pp 901-911.

Calder C, Lavine M, Muller P & Clark JS. (2003). *Incorporating multiple sources of stochasticity into dynamic population models*. Ecology 84, pp 1395-1402.

Chatfield C. (2001). "Time-Series Forecasting". Chapman & Hall/CRC, London, 267 pages.

Clark JS & Bjørnstad ON. (2004). *Population Time Series: Process Variability, Observation Errors, Missing Values, Lags and Hidden States*. Ecology 85 (11), pp 3140-3150.

Clark JS & Gelfand AE. (2006). “Hierarchical modelling for the environmental sciences: statistical methods and applications”. Oxford University Press, Oxford, 216 pages.

Clutton-Brock T, Major HM & Guinness FE. (1985). *Population regulation in male and female red deer*. Journal of Animal Ecology 54, pp 831-846.

Coulson T, Catchpole EA, Albon SD, Morgan JT, Pemberton JM, Clutton-Brock TH, Crawley MJ & Grenfell BT. (2001). *Age, sex, density, winter weather, and population crashes in Soay sheep*. Science 292, pp 1528-1531.

Dennis B & Taper ML. (1994). *Density dependence in time series observations of natural populations: estimation and testing*. Ecological Monographs 64, pp 205- 224.

Dennis B, Ponciano JM & Taper ML. (2010) *Replicated sampling increases efficiency in monitoring biological populations.* Ecology 91 (2), pp 610-620

Doucet A, de Freitas N & Gordon N. (2001). *An introduction to sequential Monte Carlo methods*, in “Sequential Monte Carlo Methods in Practice” Doucet A, de Freitas N, and Gordon N, Eds. New York: SpringerVerlag, 624 pages.

Durbin J & Koopman SJ. (2001). "Time series analysis by state space methods”. Oxford Statistical Series 24, Oxford University Press. New York, 280 pages.

Freeman SN & Kirkwood GP. (1995). *On a structural time series method for estimating stock biomass and recruitment from catch and effort data*. Fisheries Research 22 (1), pp 77-98.

Gasaway WC, Gasaway KT & Berry HH. (1996). *Persistent low densities of plains ungulates in Etosha National Park, Namibia: testing the food regulation hypothesis*.

Canadian Journal of Zoology 74, pp 1556-1572.

Gelfand AE & Smith FM. (1990). *Sampling-based approaches to calculating marginal densities*. Journal of the American Statistical Association 85, pp 398–409.

Gelman A, Carlin JB, Stern HS & Rubin DB. (1995). "Bayesian Data Analysis". Chapman and Hall, 696 pages.

Gilks W, Richardson S & Spiegelhalter D. (1996). "Markov chain Monte Carlo in practice". Chapman and Hall, 486 pages.

Grenfell BT, Wilson K, Finkenstädt BF, Coulson TN, Murray S, Albon SD, Pemberton JM, Clutton-Brock TH & Crawley MJ. (1998). *Noise and determinism in synchronized sheep dynamics*. Nature 394, pp 674-677.

Gudmundsson G. (1994). *Time series analysis of catch-at-age data.* Applied Statistics 43 (1), pp 117-126.

Harvey AC (1989). "Forecasting, Structural Time Series Models and the Kalman Filter." Cambridge University Press, Cambridge, 574 pages.

Ives AR, Abbott KC & Ziebarth NL. (2010) *Analysis of ecological time series with ARMA(p,q) models*. Ecology 91 (3), pp 858-871.

IUCN (International Union for Conservation of Nature and Natural Resources). (2002). *2002 IUCN Red List of Threatened Species*.

Jachmann H. (2002). *Comparison of aerial counts with ground counts for large African herbivores.* Journal of Applied Ecology 39, pp 841-852.

Jones, RH. (1993). "Longitudinal data with serial correlation: a state-space approach". Chapman CRC/Hall. United States of America, 225 pages.

Kimura DK, Balsiger JW & Ito DH. (1996). *Kalman filtering the delay-difference equation: practical approaches and simulations*. Fishery Bulletin 94 (4), pp 678-691.

Lunn DJ, Thomas A, Best N & Spiegelhalter D. (2000). *WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility.* Statistics and Computing 10, pp 325-337.

Maunder MN, Skaug HJ & Hoyle SD (2008). *Comparison of estimators for mark-recapture models: random effects, hierarchical Bayes, and AD Model Builder*. In: "Modeling Demographic Processes in Marked Populations". Thomson DL, Cooch EG, Conroy MJ (eds.) Environmental and Ecological Statistics, Springer, New York, 1131 pages.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller A, and Teller H (1953). *Equations of state calculations by fast computing machines*. Journal of Chemical Physics 21, pp 1087–1091.

Millar RB & Meyer R. (2000). *Bayesian State-Space Modeling of Age-Structured Data: Fitting a Model is Just the Beginning*. Canadian Journal of Fisheries and Aquatic Sciences 57(1), 43-50.

Newman, KB. (1998). *State-space modeling of animal movement and mortality with application to salmon*. Biometrics 54, pp 1290-1314.

Newman KB & Lindley ST. (2006). *Accounting for Demographic and Environmental Stochasticity, Observation Error and Parameter Uncertainty in Fish Population Dynamics Models*. Journal of Fisheries Management 26, pp 685-701.

Newman KB, Buckland ST, Lindley ST, Thomas L & Fernandez C. (2006). *Hidden process models for animal population dynamics*. Ecological Applications 16 (1), pp

74 -86.

Ntzoufras I. (2009). “Bayesian Modelling Using WinBUGS”. John Wiley & Sons. New Jersey, 492 pages.

Ogutu JO and Owen-Smith N. (2003). *ENSO, rainfall and temperature influences on extreme population declines among African savanna ungulates*. Ecological Letters 6, pp 412–419.

Owen-Smith N. (1990). *Demography of a large herbivore, the Greater Kudu Tragelaphus strepsiceros, in Relation to Rainfall.* The Journal of Animal Ecology 59 (3), pp 893-913.

Owen-Smith N. (1993). *Comparative mortality rates of male and female Kudus: The costs of sexual size dimorphism.* Journal of Animal Ecology 62(3), pp 428-440.

Owen-Smith N & Mason DR. (2005). *Comparative changes in adult vs. Juveniles survival affecting population trends of African Ungulates*. Journal of Animal Ecology 74, pp 762-773.

Owen-Smith N, Mason DR & Ogutu JO. (2005*). Correlates of survival rates for 10 African ungulate populations: density, rainfall and predation*. Journal of Animal Ecology 74, pp 774-788.

Owen-Smith N & Mills MGL. (2006). *Manifold interactive influences on the population dynamics of a multispecies assemblage.* Ecological Monographs 76 (1), pp 73 -92.

Pella JJ. (1993) Utility of Time Series Models and the Kalman Filter for predicting consequences of fishery actions. In “Proceedings of the international symposium on

management strategies for exploited fish populations." Edited by Kruse G, Eggers DM, Marasco RJ, Pautzke C. & Quinn II TJ. Alaska Sea Grant College Program Report 93-02, University of Alaska, Fairbanks. pp 571-593.

Perrin MR. (1999). *The social organisation of the greater Kudu Tragelaphus strepsiceros (Pallas 1766).* Tropical Zoology 12(2), pp 169-208.

Pollard EKH, Lakhani & Rothery P. (1987). *The detection of density dependence from a series of annual censuses*. Ecology 68, pp 2046-2055.

Post E, Stenseth NC, Peterson RO, Vucetich JA & Ellis AM. (2002). *Phase dependence and population cycles in a large-mammal predator-prey system.* Ecology 83, pp 2997- 3002.

Red WJ, & Simons CM, (1996). *Analyzing catch-effort data by means of the Kalman filter*. Canadian Journal of Fisheries and Aquatic Sciences 53, pp 2157-2166.

Regan HM, Colyvan M & Burgman MA. (2002). *A taxonomy and treatment of uncertainty for ecology and conservation biology.* Ecological Applications 12, pp 618-628.

Rivot E, Prévost E, Parent E & Bagliniére JL. (2004). *A Bayesian state-space modelling framework for fitting a salmon stage-structured population model to multiple time series of field data*. Ecological Modelling 179, pp 463-485.

Seber GAF. (1982). "The estimation of animal abundance and related parameters", Charles Griffin, London, 654 pages.

Schnute JT. (1994). *A general framework for developing sequential fisheries models.* Canadian Journal of Fisheries and Aquatic Sciences 51, pp 1676-1688.

Shenk TM, White GC & Burnham KP. (1998). *Sampling variance effects on detecting density dependence from temporal trends in natural populations.* Ecological Monographs. 68(3), pp 445–463.

Smuts GL. (1978). *Interrelations between predators, prey, and their environment*. Bioscience 2, pp 316-320.

South African National Parks. (2008). *Maps.* Available: http://www.sanparks.org.za/parks/kruger/. Last accessed 20 September 2009.

Sullivan PJ. (1992). *A Kalman filter approach to catch-at-length analysis.* Biometrics 48, pp 237-257.

Thomas L, Buckland ST, Newman KB & Harwood J. (2005). *A unified framework for modelling wildlife population dynamics*. Australian and New Zealand Journal of Statistics 47(1), pp 19-34.

Trenkel VM, Elston DA & Buckland ST. (2000). *Calibrating population dynamics models to count and cull data using sequential importance sampling*. Journal of the American Statistical Association 95, pp 363-374.

Turchin P. (1995). *Population regulation: old arguments and a new synthesis.* Pages 19-40 in Cappucino N & Price PW, editors. "Population dynamics: new approaches and synthesis." Academic Press, San Diego, California, USA.

Turchin P. (2003). "Complex population dynamics: a theoretical/empirical synthesis." Princeton Univ. Press, 474 pages.

United States Environmental Protection Agency. (2006). Glossary Retrieved November 20, 2009 from http://www.epa.gov/waterscience/biocriteria/glossary.html.

Viljoen PC & Retief PF. (1994). *The use of the global positioning system for real time data collecting during ecological aerial surveys in the Kruger National Park.* Koedoe 37, pp 149-157.

Vucetich JA & Peterson RO (2004). *The influence of top-down, bottom-up, and abiotic factors on the moose (Alces alce) population of Isle Royale*, Proceedings of the Royal Society of London, B 271, pp 183-189.

West M & Harrison J. (1997). "Bayesian forecasting and dynamic models." Second edition. Springer-Verlag, New York, New York, USA, 701 pages.

Williams BK, Nicholas JD & Conroy MJ. (2002). "Analysis and management of animal populations" Academic Press, 817 pages.

Appendix A: WinBUGS Code

```
#winbugs code for hierarchical hidden process model

model

{

        female_yearling[1] ~  dunif(816,866)

        male_yearling[1]   <-  (female_yearling[1])

        adult_female[1]      ~ dunif(3503,3543)

        adult_male[1]        ~  dunif(1417,1457)

        adult_male[1] <- 0.408*adult_female[1]


    for (j in 2 : T)  {

                female_yearling[ j ] ~ dbin(0.5, realised_juvenile_survival[j-1])

                male_yearling[ j ] <-  (realised_juvenile_survival[j-1] -female_yearling[j])

                adult_female[ j ] <- realised_female_yearling_survival[j -1] + realised_adult_female_survival[j -1]

                adult_male[ j ] <- realised_male_yearling_survival[j-1] + realised_adult_male_survival[j-1]

               }

  for ( i in 1:T)   {


        recruit[i]<- 0.34*adult_female[i]

        juvenile[i] ~ dpois(recruit[i])


        realised_juvenile_survival[i] ~ dbin(juvenile_survival_rate[i], juvenile[i])

        realised_female_yearling_survival[i] ~ dbin(female_yearling_survival_rate[i], female_yearling[i])

        realised_male_yearling_survival[i] ~ dbin(male_yearling_survival_rate[i], male_yearling[i])


         realised_adult_female_survival[i] ~ dbin(adult_female_survival_rate[i], adult_female[i])

         realised_adult_male_survival[i] ~ dbin(adult_male_survival_rate[i], adult_male[i])


#hyperdistribution pdfs

       juvenile_survival_rate[i] ~  dBeta(p_juv_surv[i] ,q_juv_surv[i])

          a_juv_surv[i] <-  abs(0.439 + 0.217*wet_season_rain[i] + 0.159*dry_season_rain[i])

          b_juv_surv[i] <-  abs(0.487 + 0.103*wet_season_rain[i] + 0.087*dry_season_rain[i])
```

```
p_juv_surv[i] <- abs(((0.55 - a_juv_surv[i] )/(b_juv_surv[i] - a_juv_surv[i]))*(((0.55-a_juv_surv[i])*(b_juv_surv[i] - 0.55)/0.02722) - 1))

q_juv_surv[i] <- abs((1- (0.55 - a_juv_surv[i] )/(b_juv_surv[i] - a_juv_surv[i]))*(((0.55-a_juv_surv[i])*(b_juv_surv[i] - 0.55)/0.02722) - 1))


female_yearling_survival_rate[i] ~ dBeta(p_fyearl_surv[i] , q_fyearl_surv[i])

a_fyearl_surv[i] <- abs(0.821 + 0.094*wet_season_rain[i])

b_fyearl_surv[i] <- abs(0.873 + 0.23*wet_season_rain[i] )

p_fyearl_surv[i] <- abs(((0.70 - a_fyearl_surv[i] )/(b_fyearl_surv[i] - a_fyearl_surv[i]))*(((0.70-a_fyearl_surv[i])*(b_fyearl_surv[i] - 0.70)/0.010816) - 1))

q_fyearl_surv[i] <- abs((1- (0.70 - a_fyearl_surv[i] )/(b_fyearl_surv[i] - a_fyearl_surv[i]) )*(((0.70-a_fyearl_surv[i])*(b_fyearl_surv[i] - 0.70)/0.010816) - 1))



male_yearling_survival_rate[i] <- abs(0.875*female_yearling_survival_rate[i])


adult_female_survival_rate[i] ~ dBeta(p_fadult_surv[i] ,q_fadult_surv[i])

a_fadult_surv[i] <- abs(0.821 + 0.046*wet_season_rain[i] + 0.061*dry_season_rain[i])

b_fadult_surv[i] <- abs( 0.871+ 0.202*wet_season_rain[i] + 0.175*dry_season_rain[i])

p_fadult_surv[i] <- abs(((0.875 - a_fadult_surv[i] )/(b_fadult_surv[i] - a_fadult_surv[i]))*(((0.875 - a_fadult_surv[i])*(b_fadult_surv[i] - 0.875)/0.0134) - 1))

q_fadult_surv[i] <- abs((1- (0.875 - a_fadult_surv[i] )/(b_fadult_surv[i] - a_fadult_surv[i]))*(((0.875 - a_fadult_surv[i])*(b_fadult_surv[i] - 0.875)/0.0134) - 1))



adult_male_survival_rate[i] <- 0.875* adult_female_survival_rate[i]

#total abundance

n[i] <- juvenile[i] + female_yearling[i] + male_yearling[i] + adult_female[i] + adult_male[i]



#observation process pdf

count[i] ~ dnorm(n[i], 0.00000625)

}}

#data

list(T = 11, count = c(7005, 8661, 10432, 10760, 8786, 8214, 6935, 5967, 4706, 3967, 3127), wet_season_rain =c(0.767, 0.078, 0.259, 0.413, 0.371, 0.008, 0.296, 0.1, 0, 0.874, 0.092),

dry_season_rain =c(0.4, 0.491, 0.164, 0.358, 0.380, 0.543, 1.107, 0.94, 0.907, 0.919, 0.612))
```

```
#initial values

list(juvenile_survival_rate = c(0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55),

    female_yearling_survival_rate = c(0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7) ,

    adult_female_survival_rate = c(0.875, 0.875, 0.875, 0.875, 0.875, 0.875, 0.875, 0.875, 0.875, 0.875, 0.875) ,

juvenile = c(1454, 1454, 1454, 1454, 1454, 1454, 1454, 1454, 1454, 1454, 1454),

female_yearling = c(422, 422, 422, 422, 422, 422, 422, 422, 422, 422, 422),

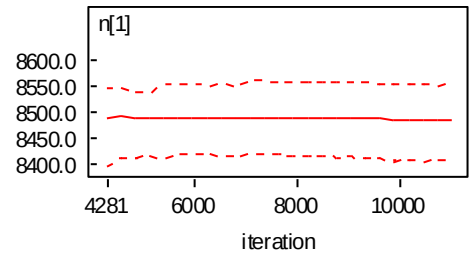
realised_juvenile_survival = c(1199, 1199, 1199, 1199, 1199, 1199, 1199, 1199, 1199, 1199, 1199),

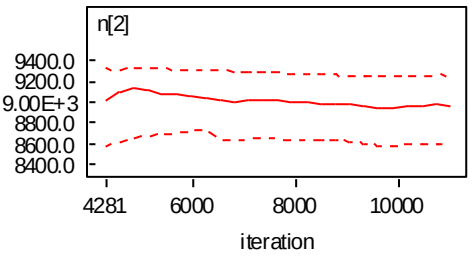
realised_female_yearling_survival = c(422, 422, 422, 422, 422, 422, 422, 422, 422, 422, 422),

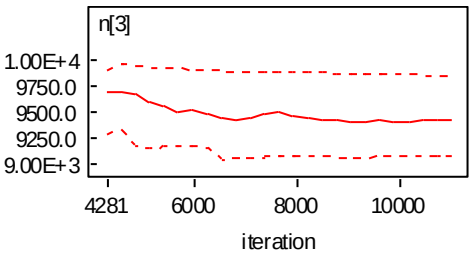
realised_male_yearling_survival = c(422, 422, 422, 422, 422, 422, 422, 422, 422, 422, 422), realised_adult_female_survival =
c(3525, 3525, 3525, 3525, 3525, 3525, 3525, 3525, 3525, 3525, 3525), realised_adult_male_survival =  c(1438, 1438, 1438,
1438, 1438, 1438, 1438, 1438, 1438, 1438, 1438))
```

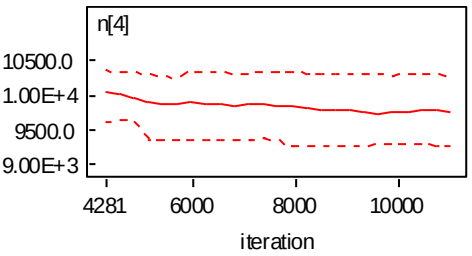
Appendix B: Running Quantile Plots

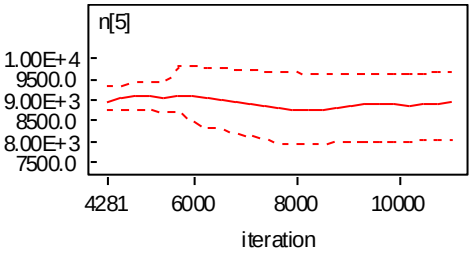
Running Quantiles: Total Count – Model 1

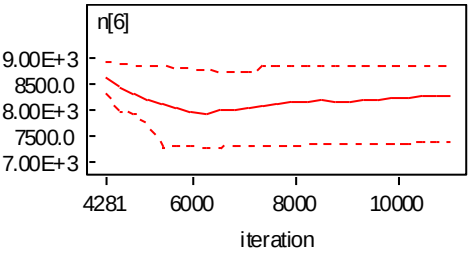


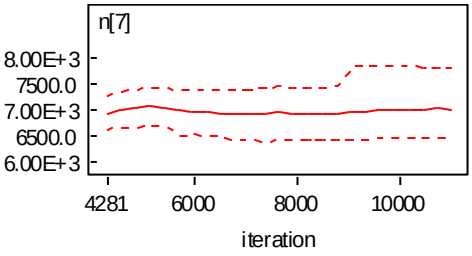


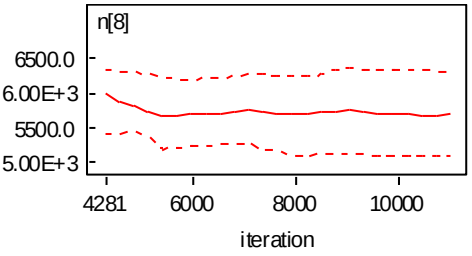


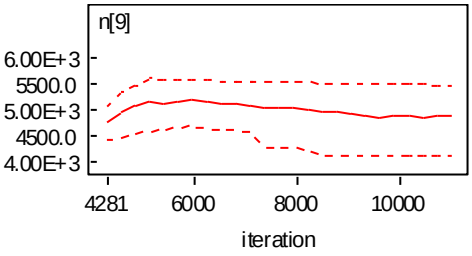


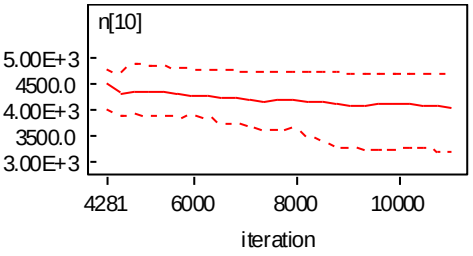


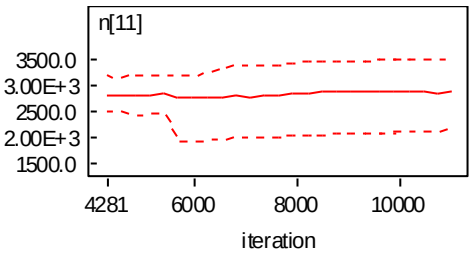




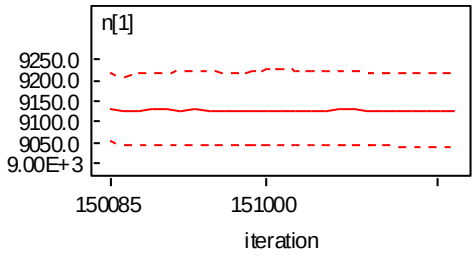


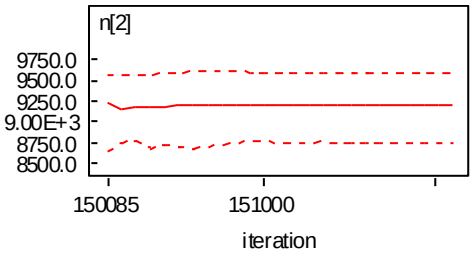


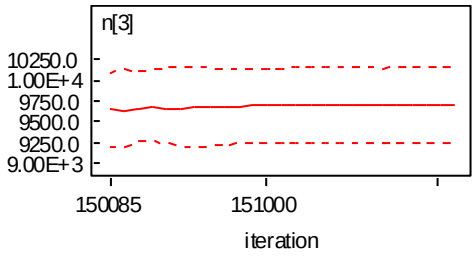


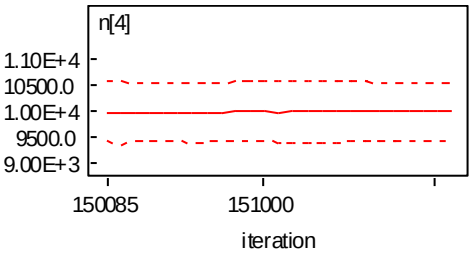

n[11]
3500.0
3.00E+3
2500.0
2.00E+3
1500.0
4281
6000
8000
10000
iteration

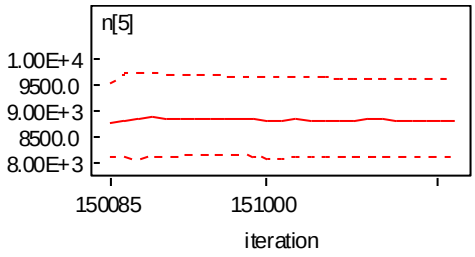
Running Quantiles: Total Count – Model 2

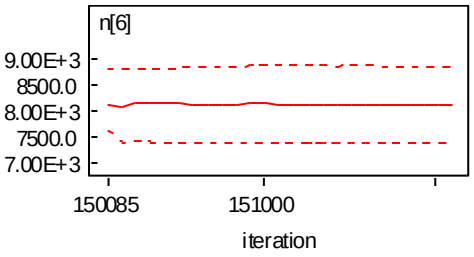


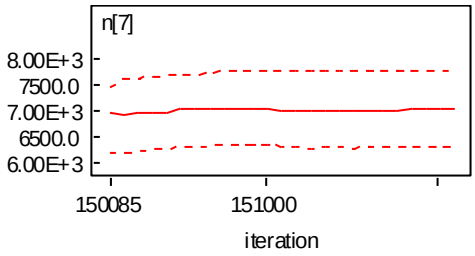


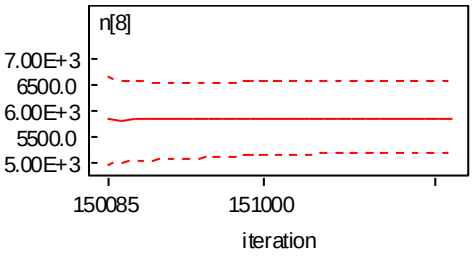


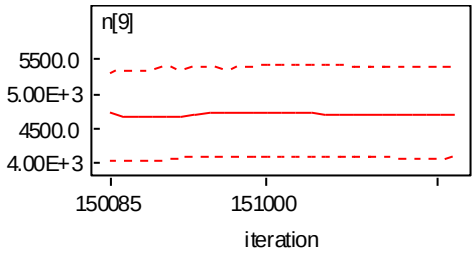


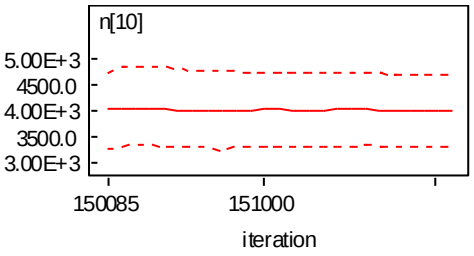


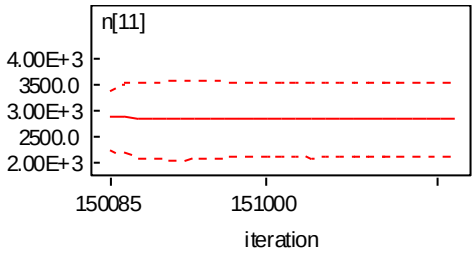




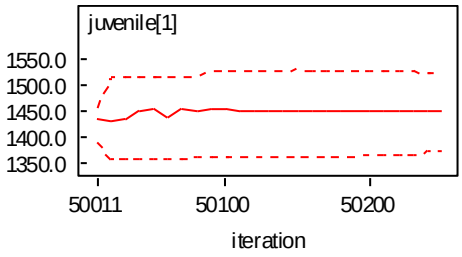


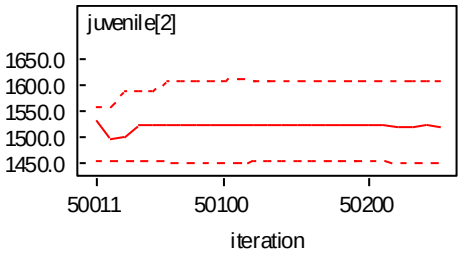


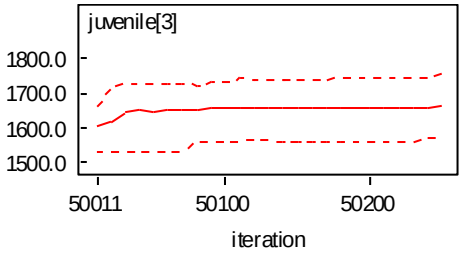


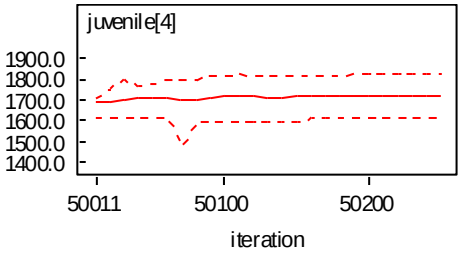


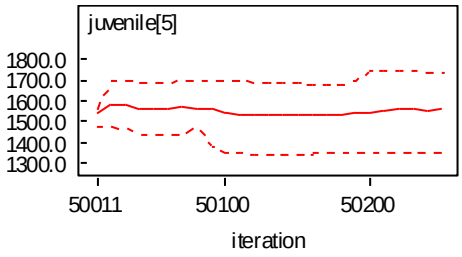
Running Quantiles: Estimated Number of Juveniles – Model 1

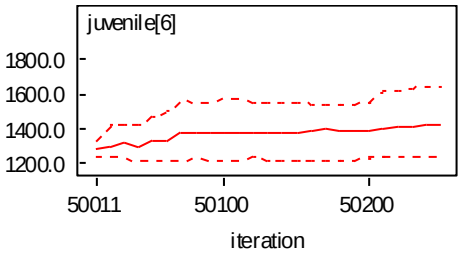


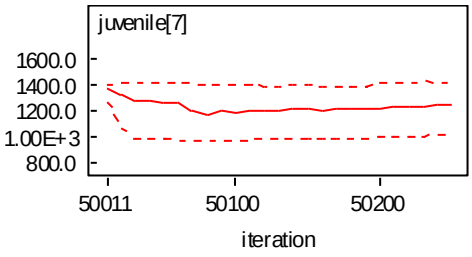


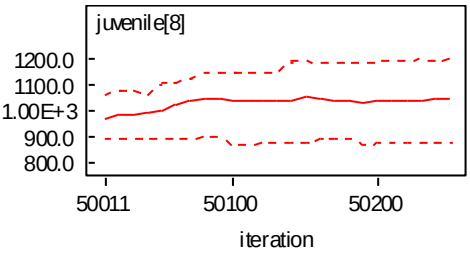


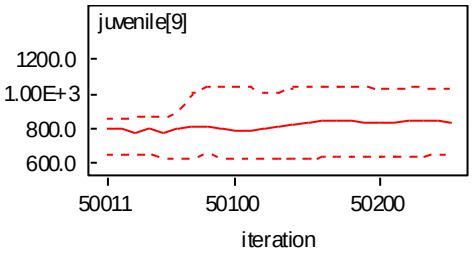


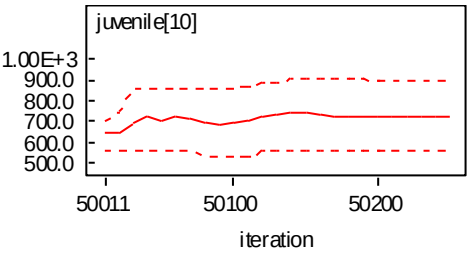


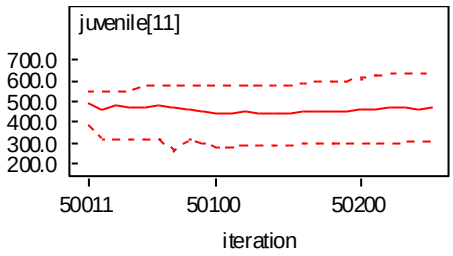




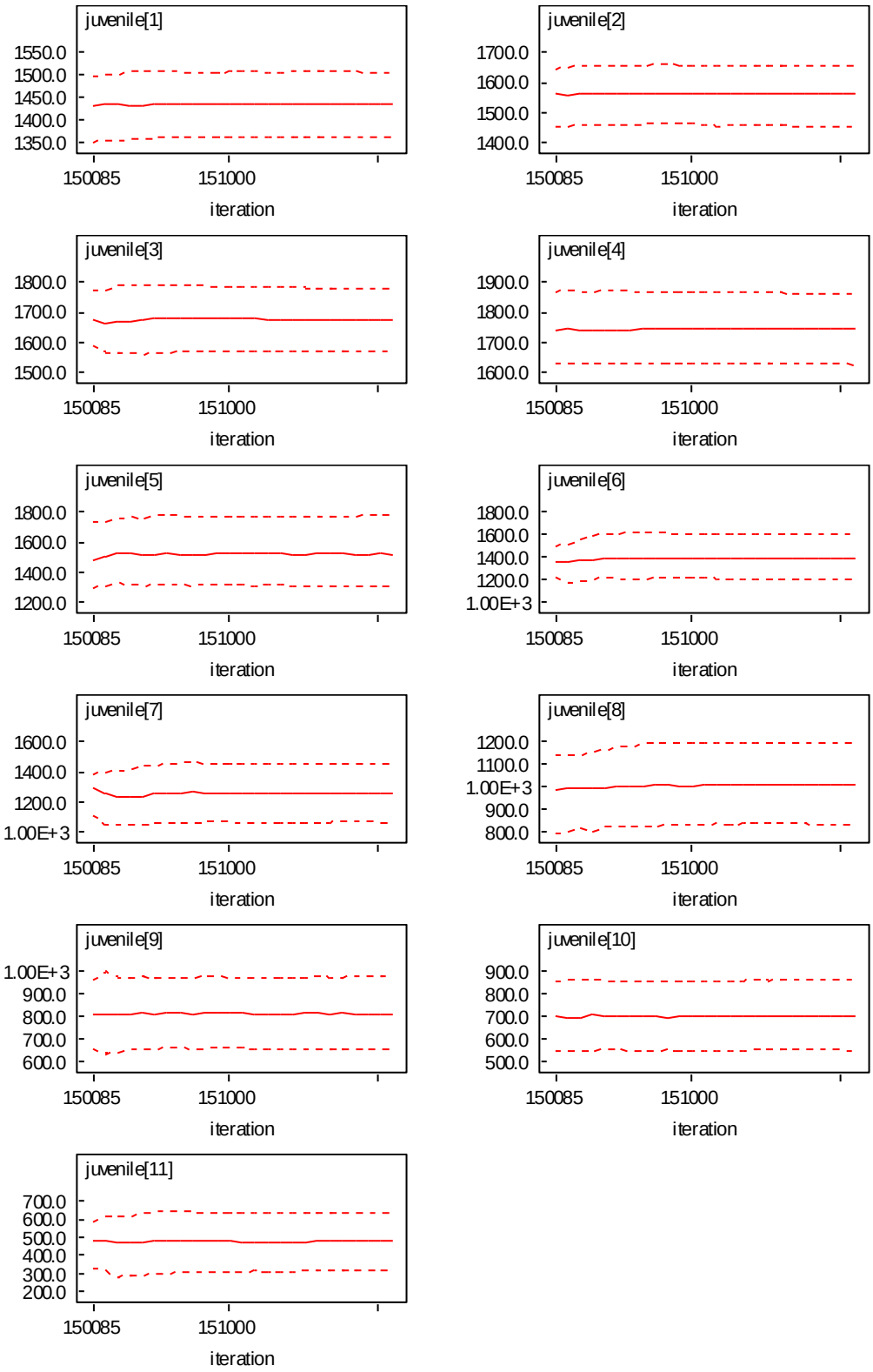




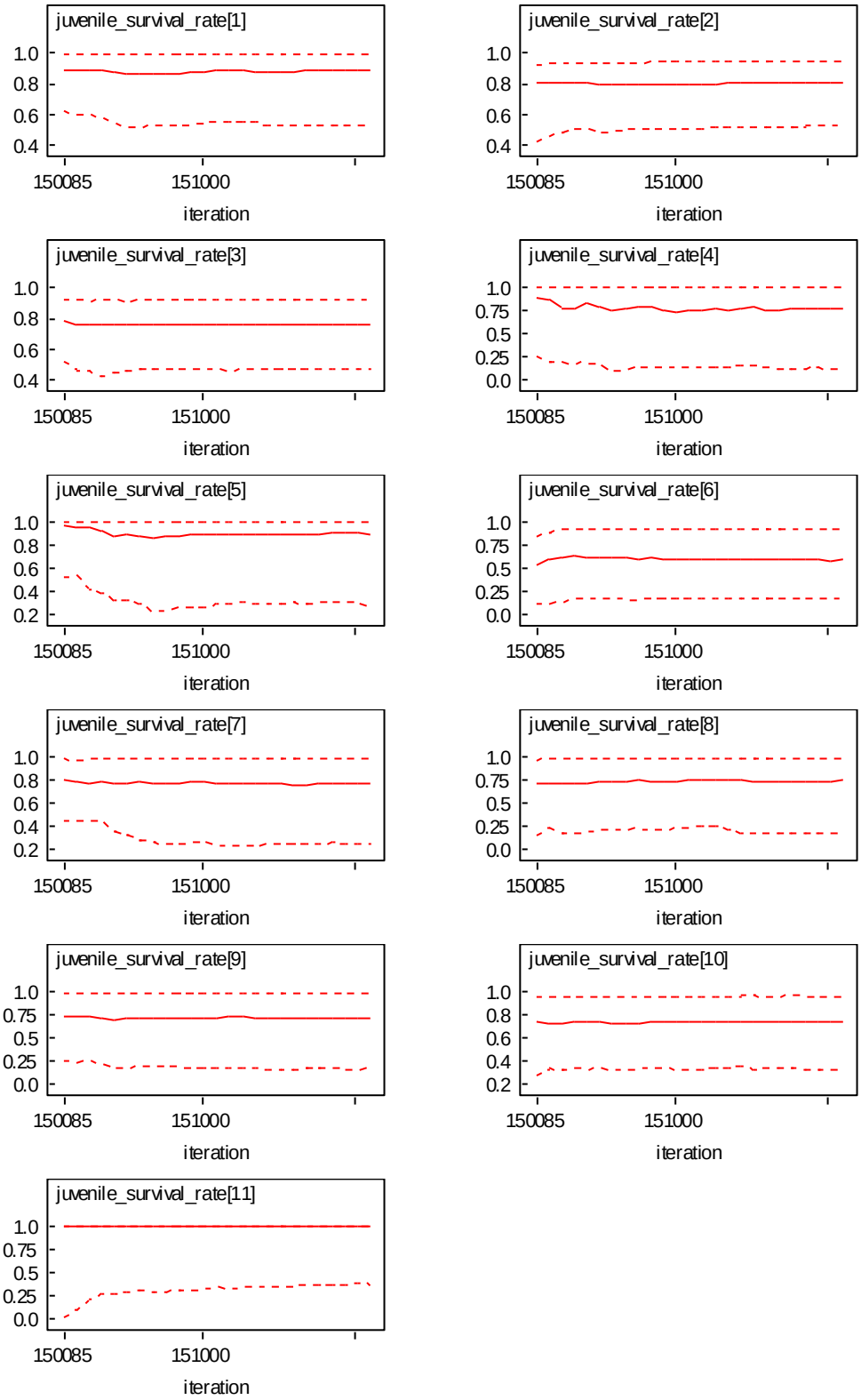



juvenile[11]
700.0
600.0
500.0
400.0
300.0
200.0
50011
50100
50200
iteration

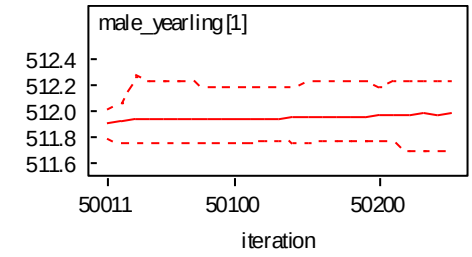
Running Quantiles: Estimated Number of Juveniles – Model 2

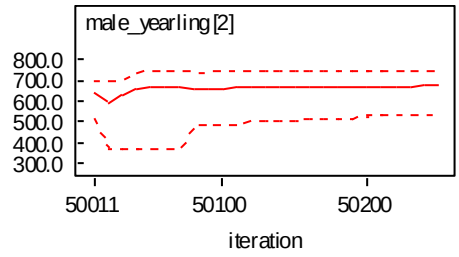


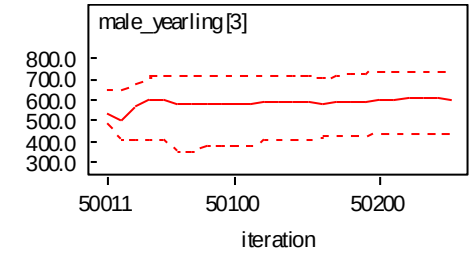
Running Quantiles: Estimated Juvenile Survival Rate – Model 1

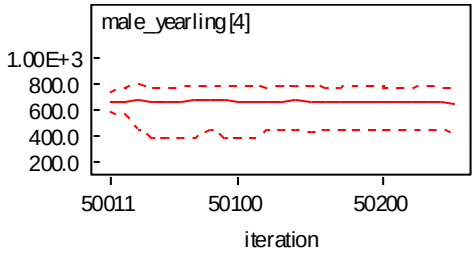


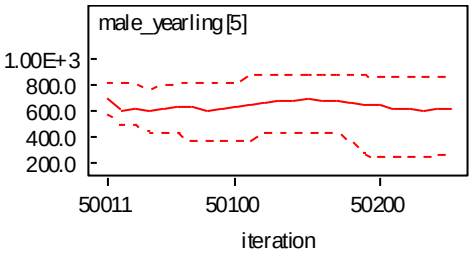
Running Quantiles: Number of Male Yearlings – Model 1

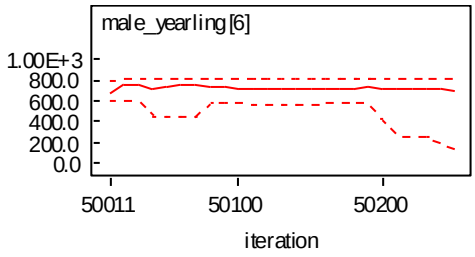


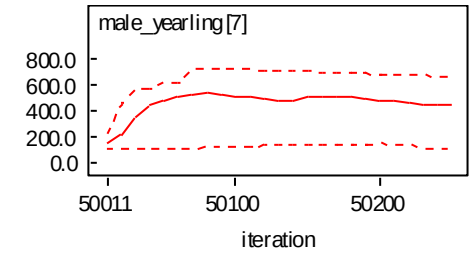


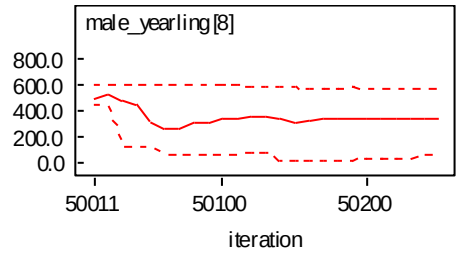


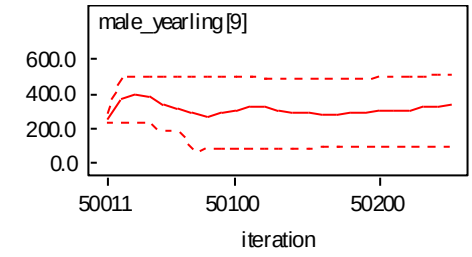


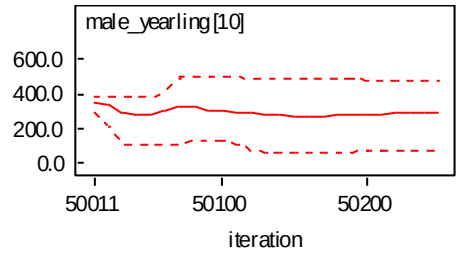


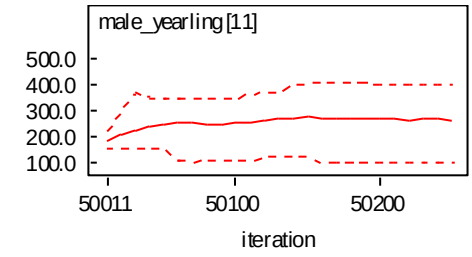




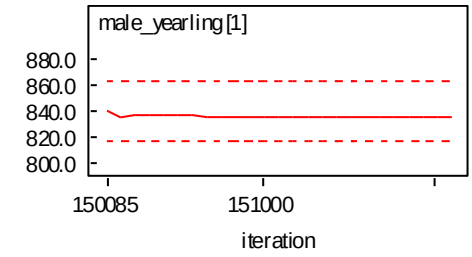


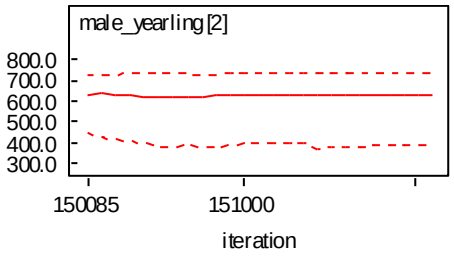


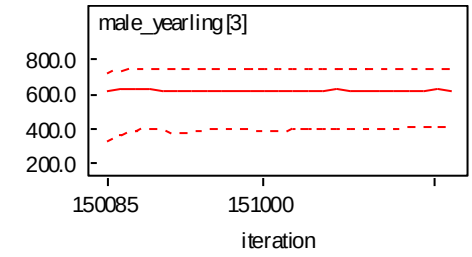


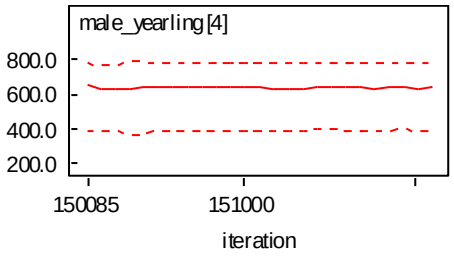


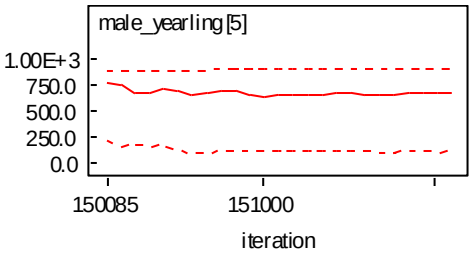
Running Quantiles: Number of Male Yearlings – Model 2

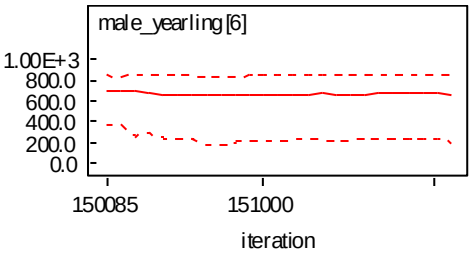


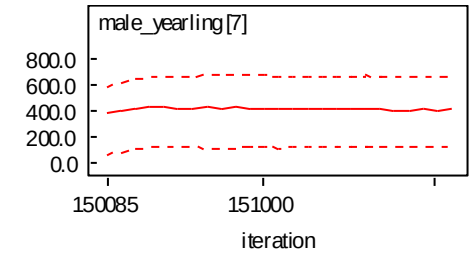


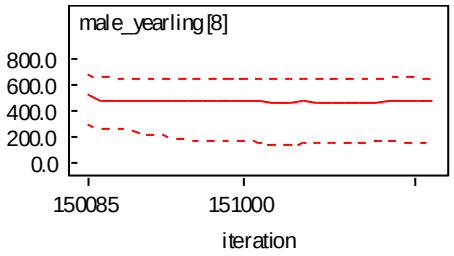


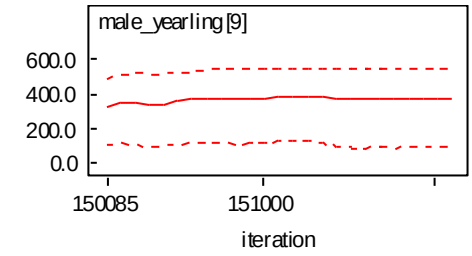


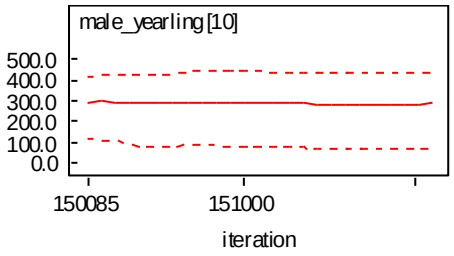


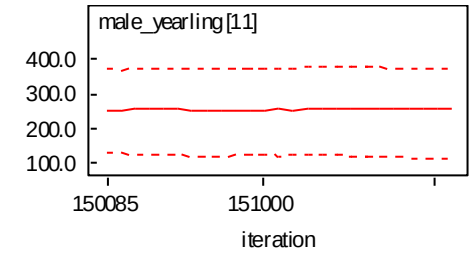




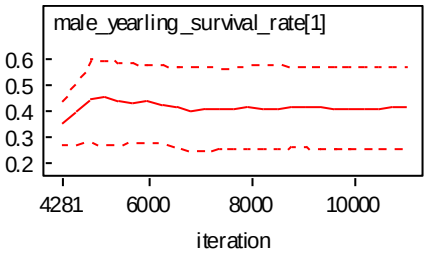


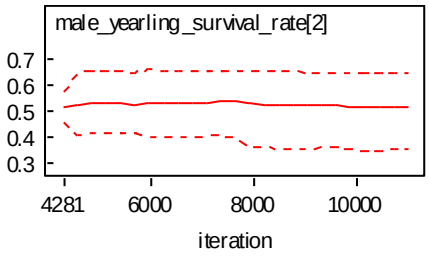


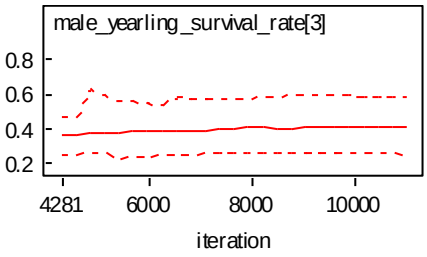


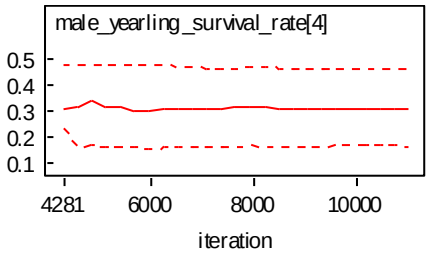


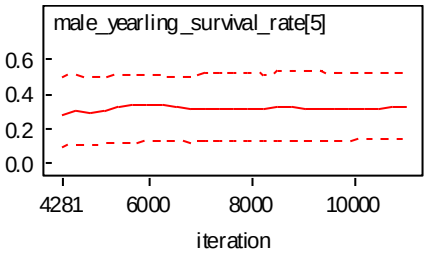
Running Quantiles: Male Yearling Survival Rate – Model 1

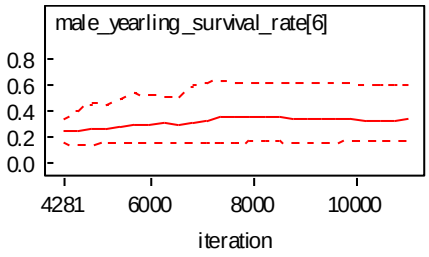


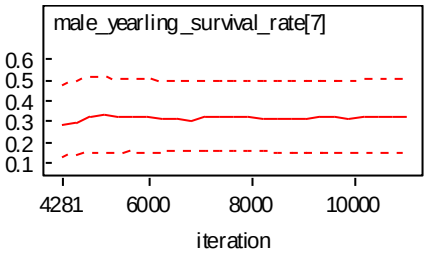


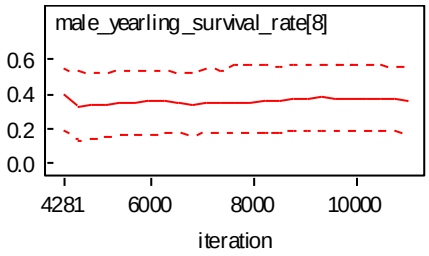


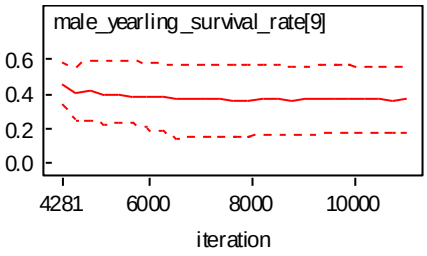


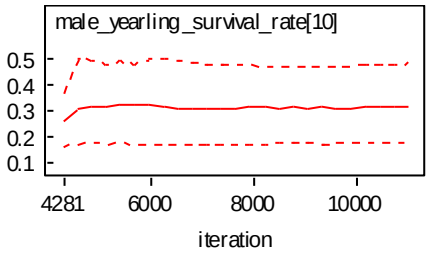


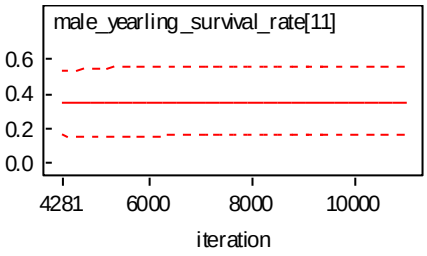




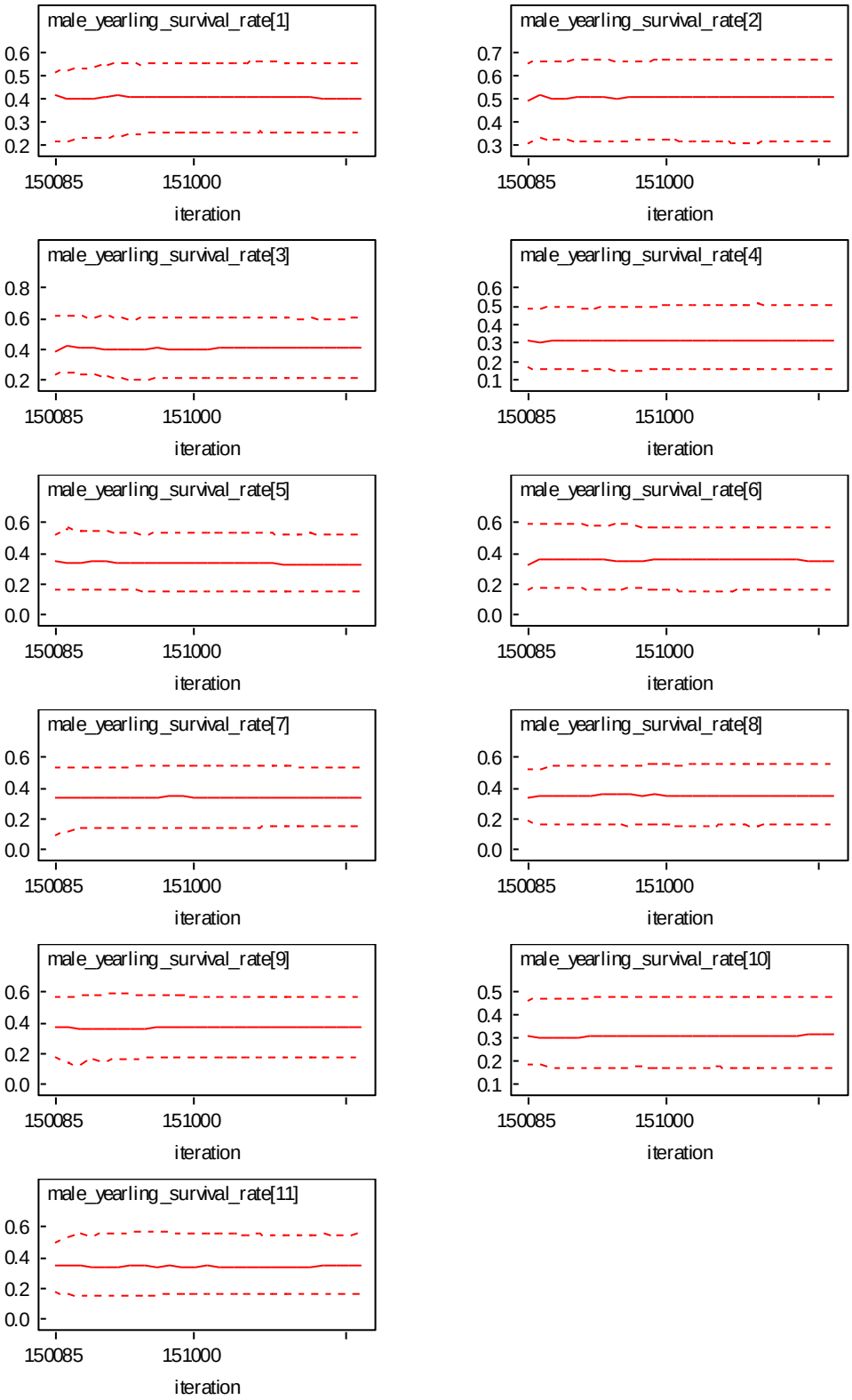




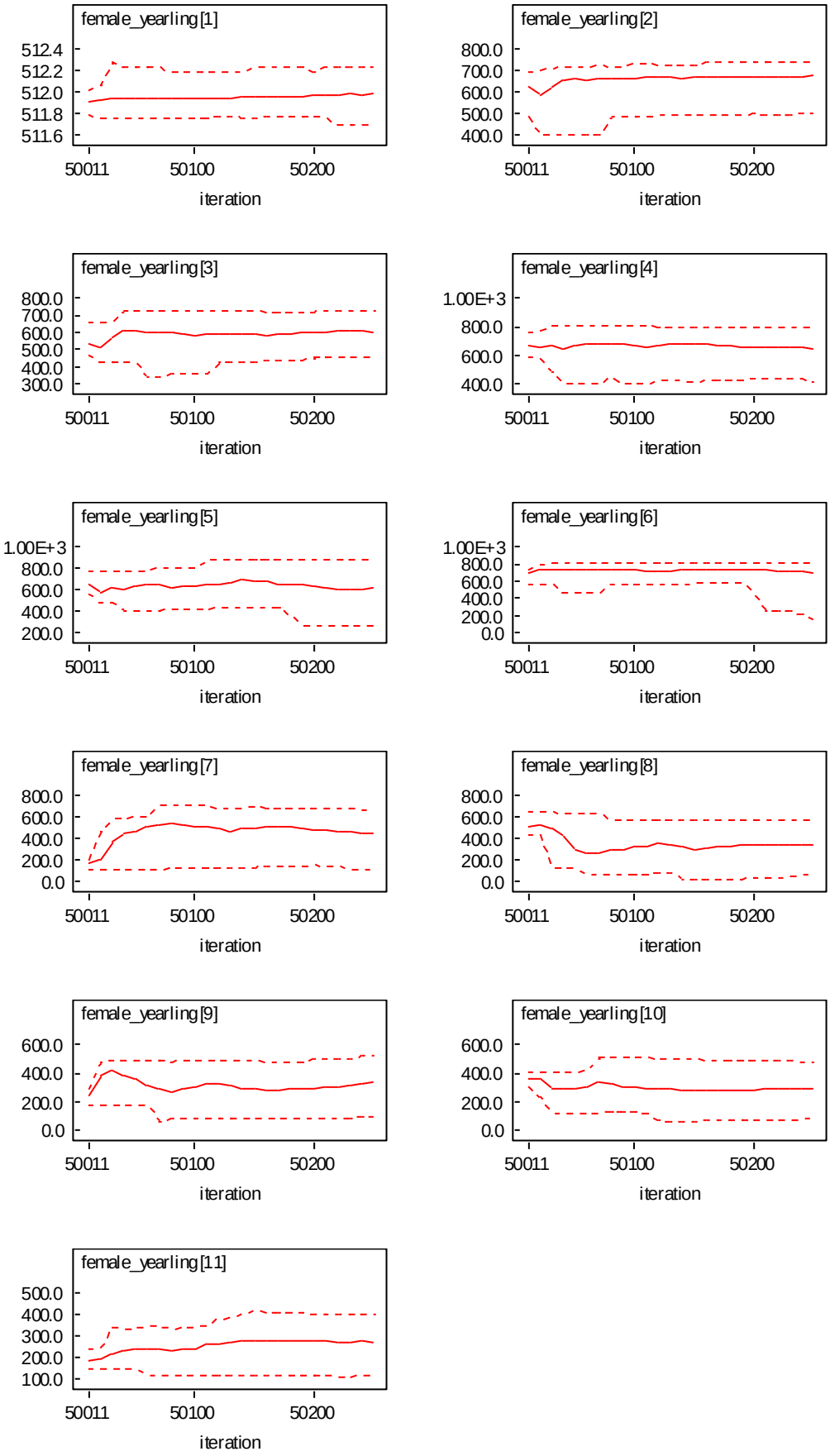




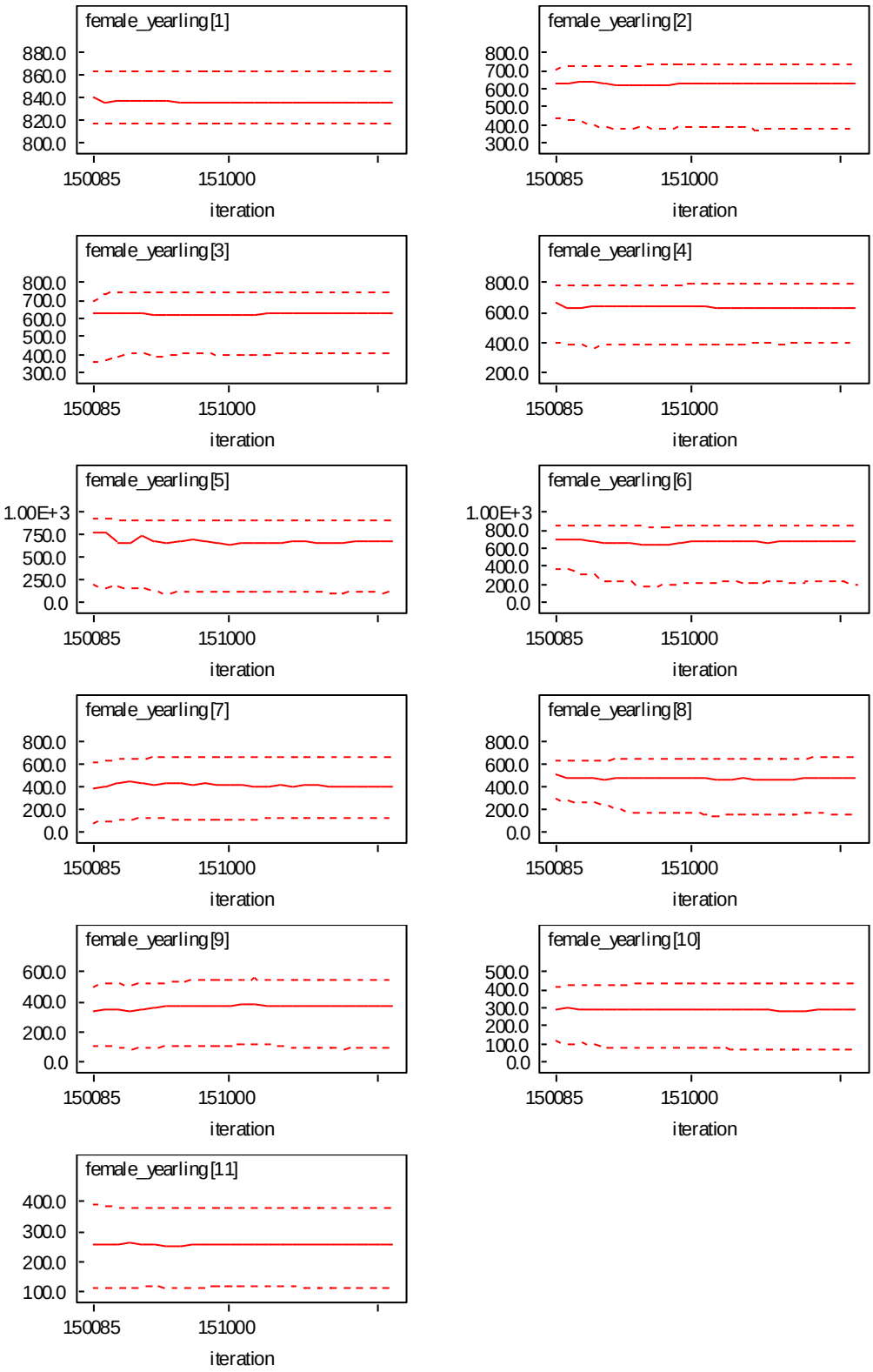
Running Quantiles: Male Yearling Survival Rate – Model 2



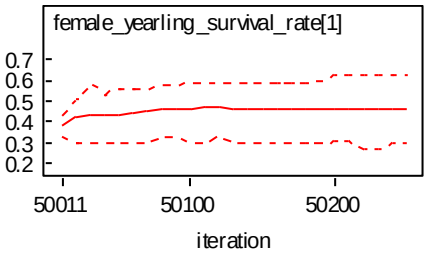
Running Quantiles: Number of Female Yearlings – Model 1

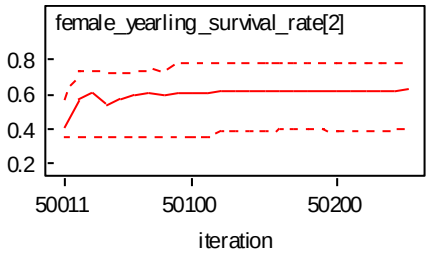


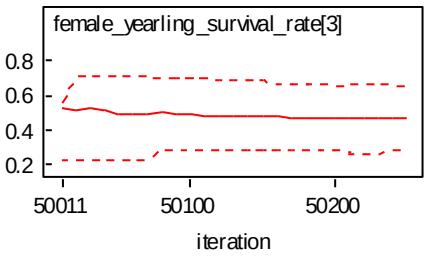
Running Quantiles: Number of Female Yearlings – Model 2

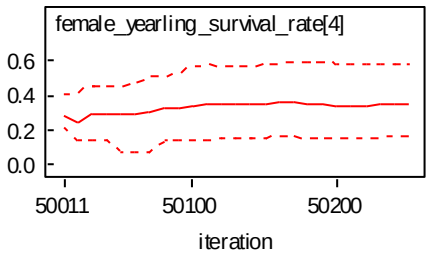


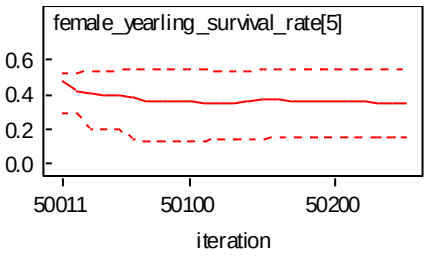
Running Quantiles: Female Yearling Survival Rate – Model 1

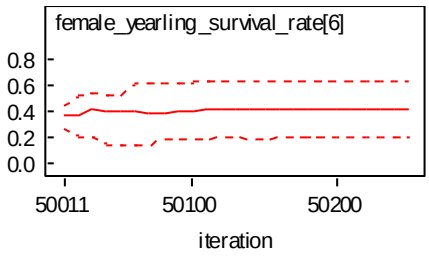


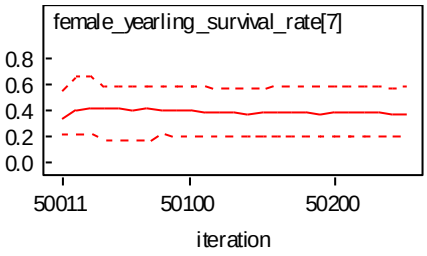


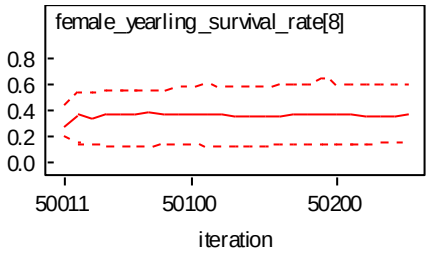


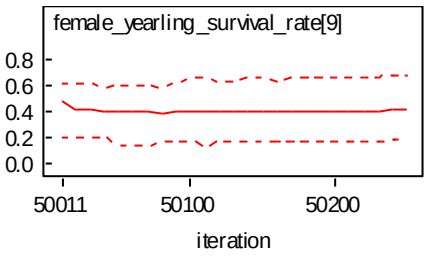


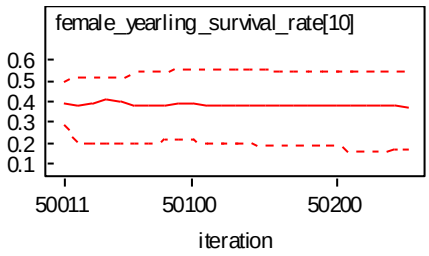


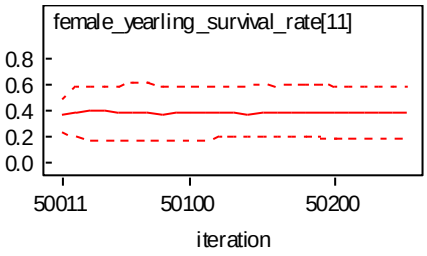




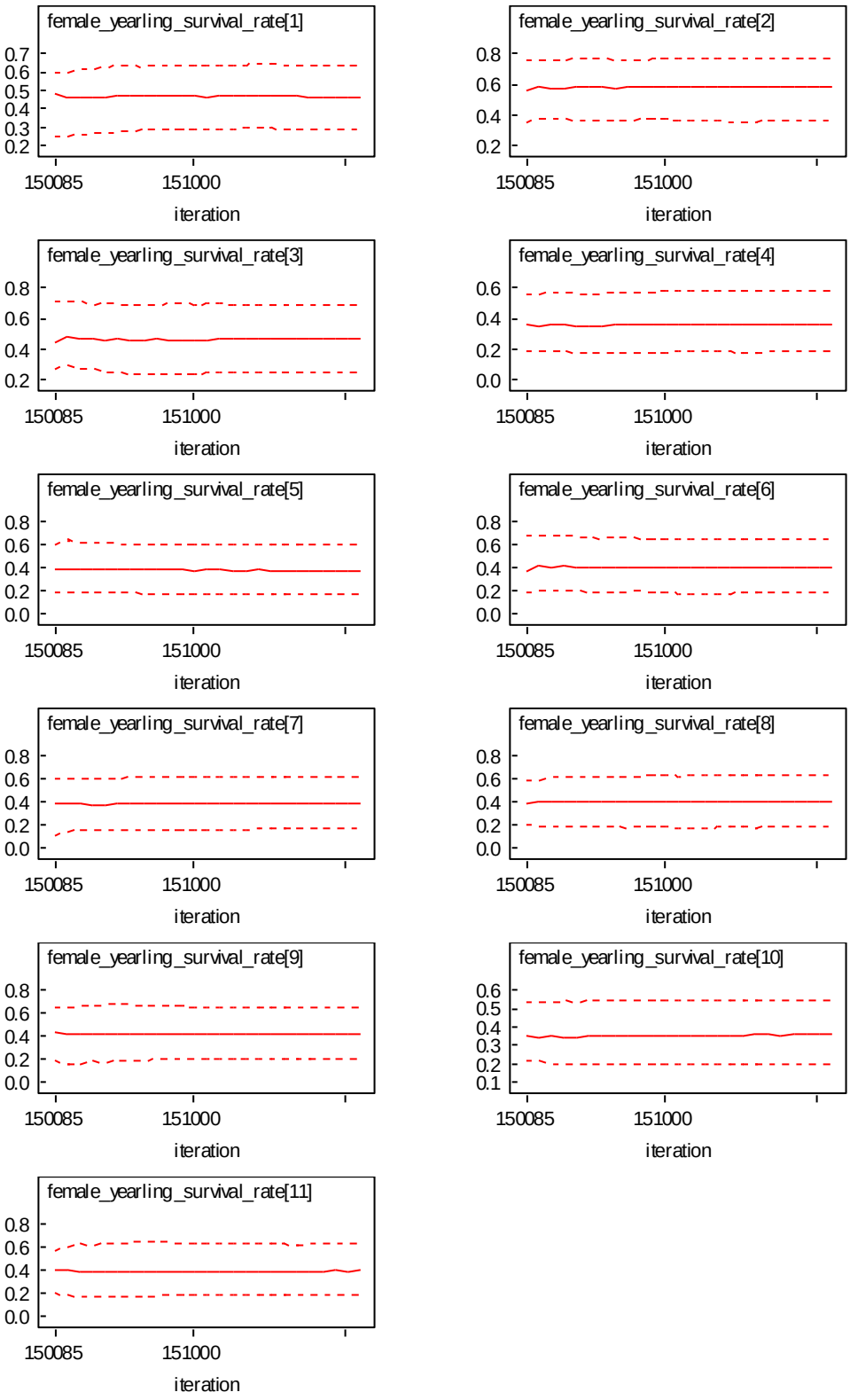




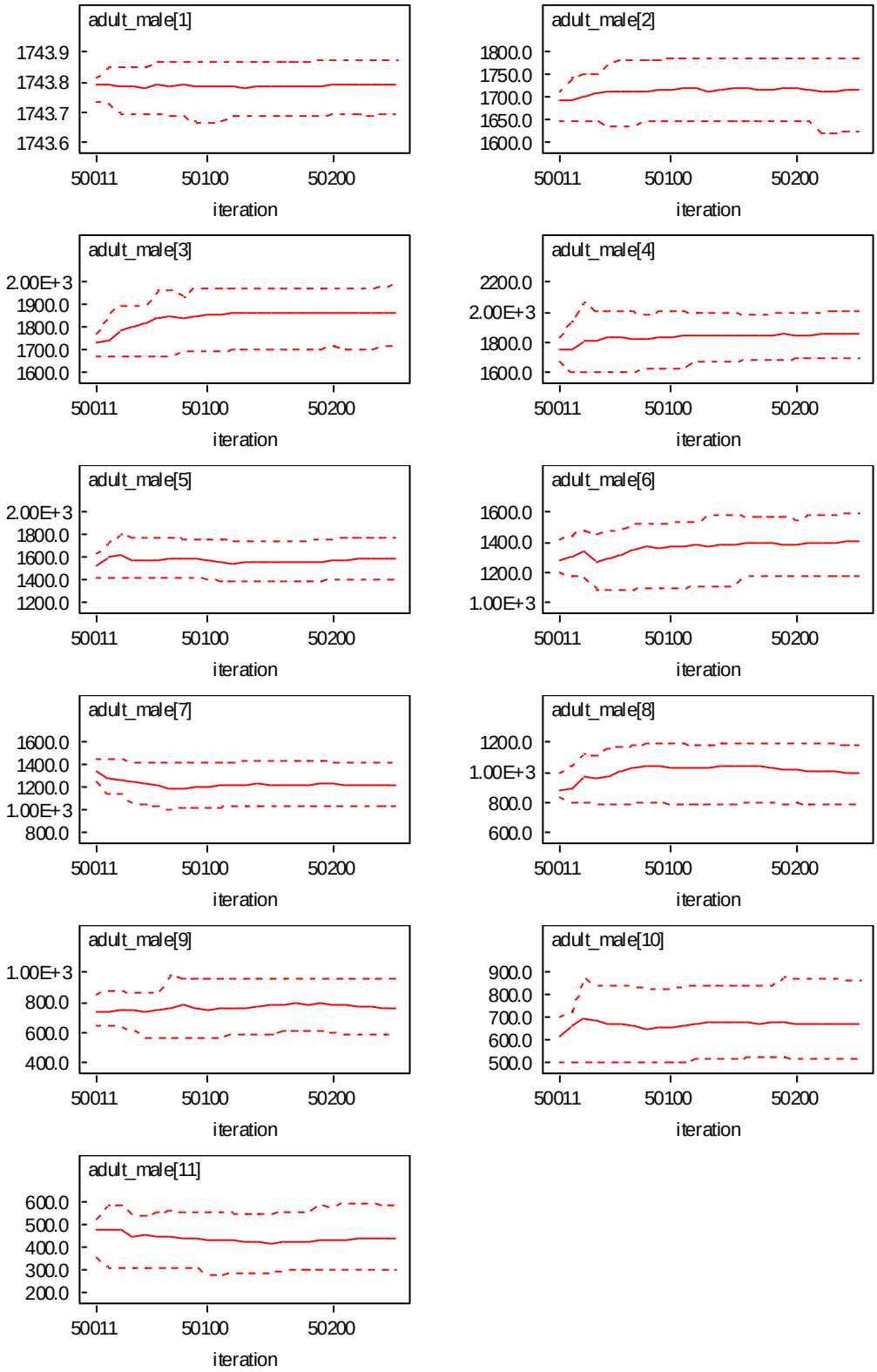




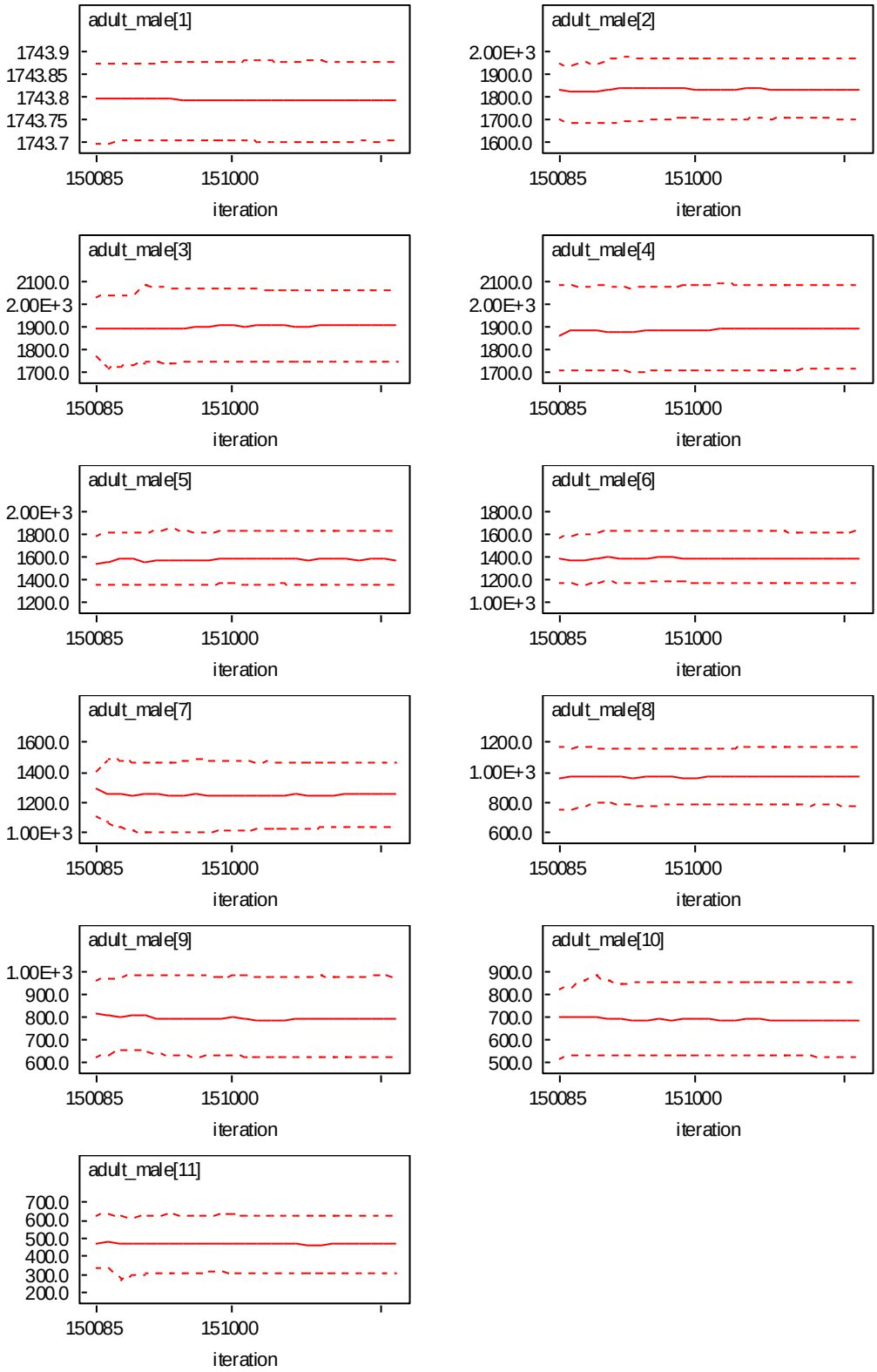
Running Quantiles: Female Yearling Survival Rate – Model 2



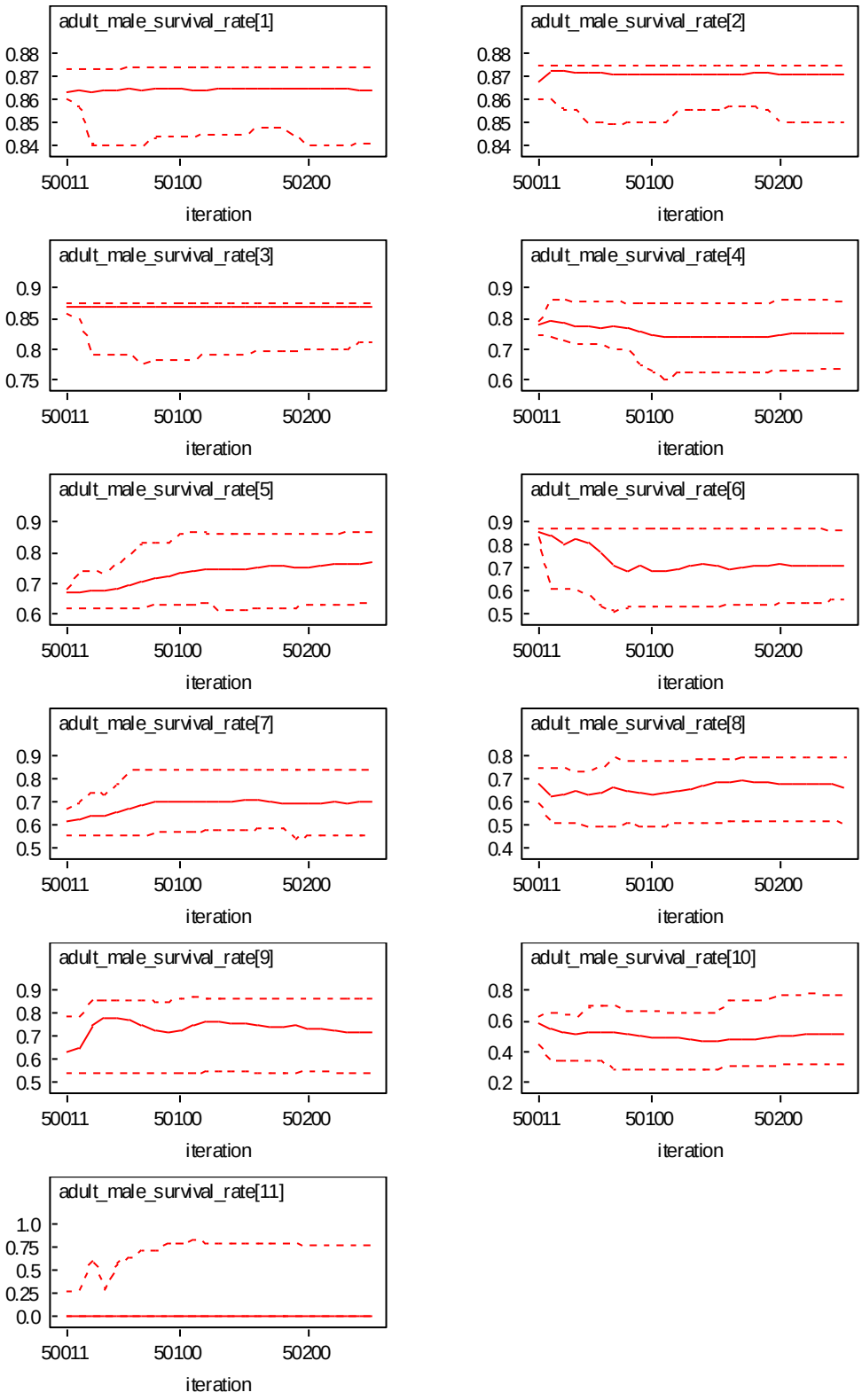
Running Quantiles: Number of Male Adults - Model 1



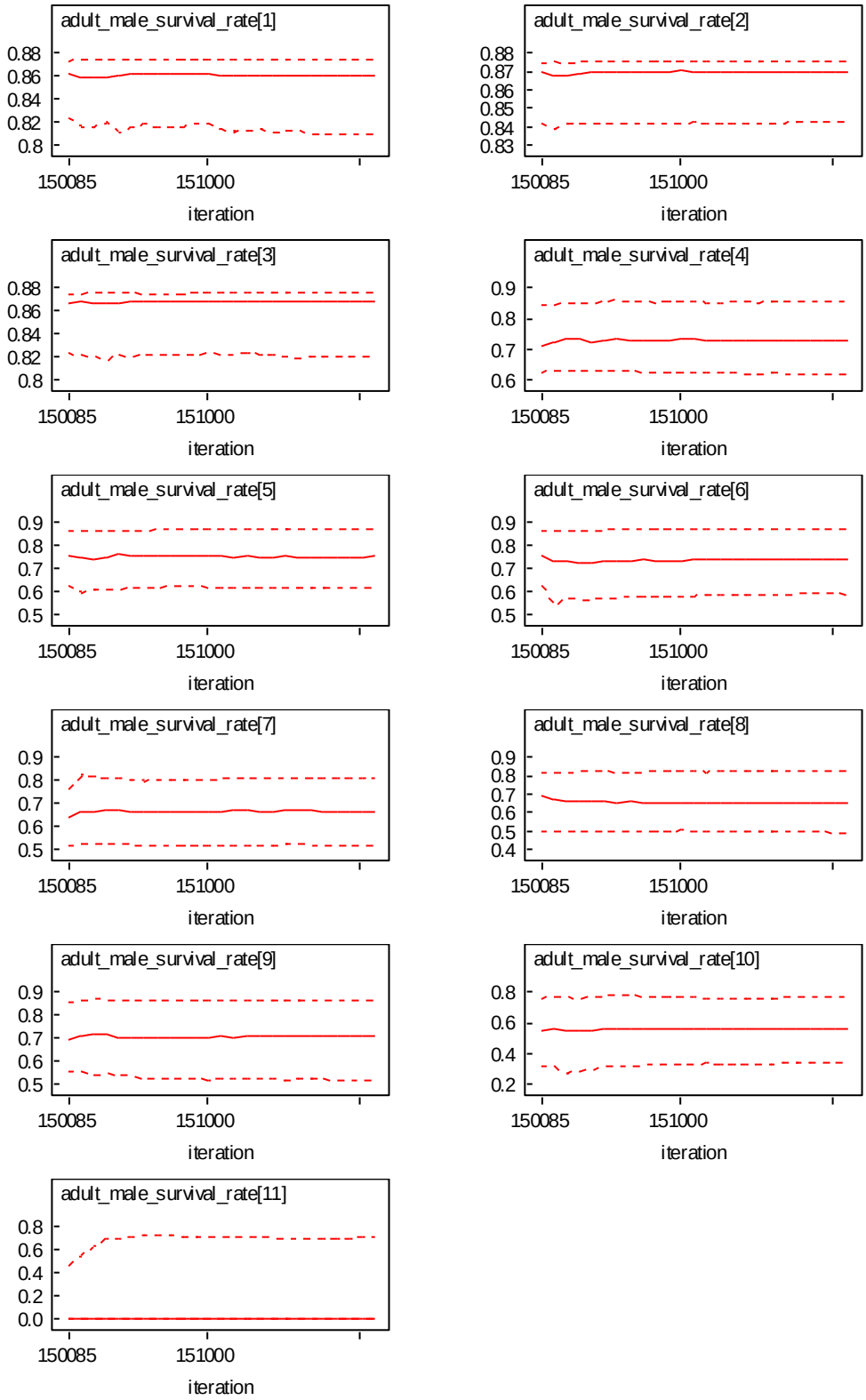
Running Quantiles: Number of Male Adults - Model 2



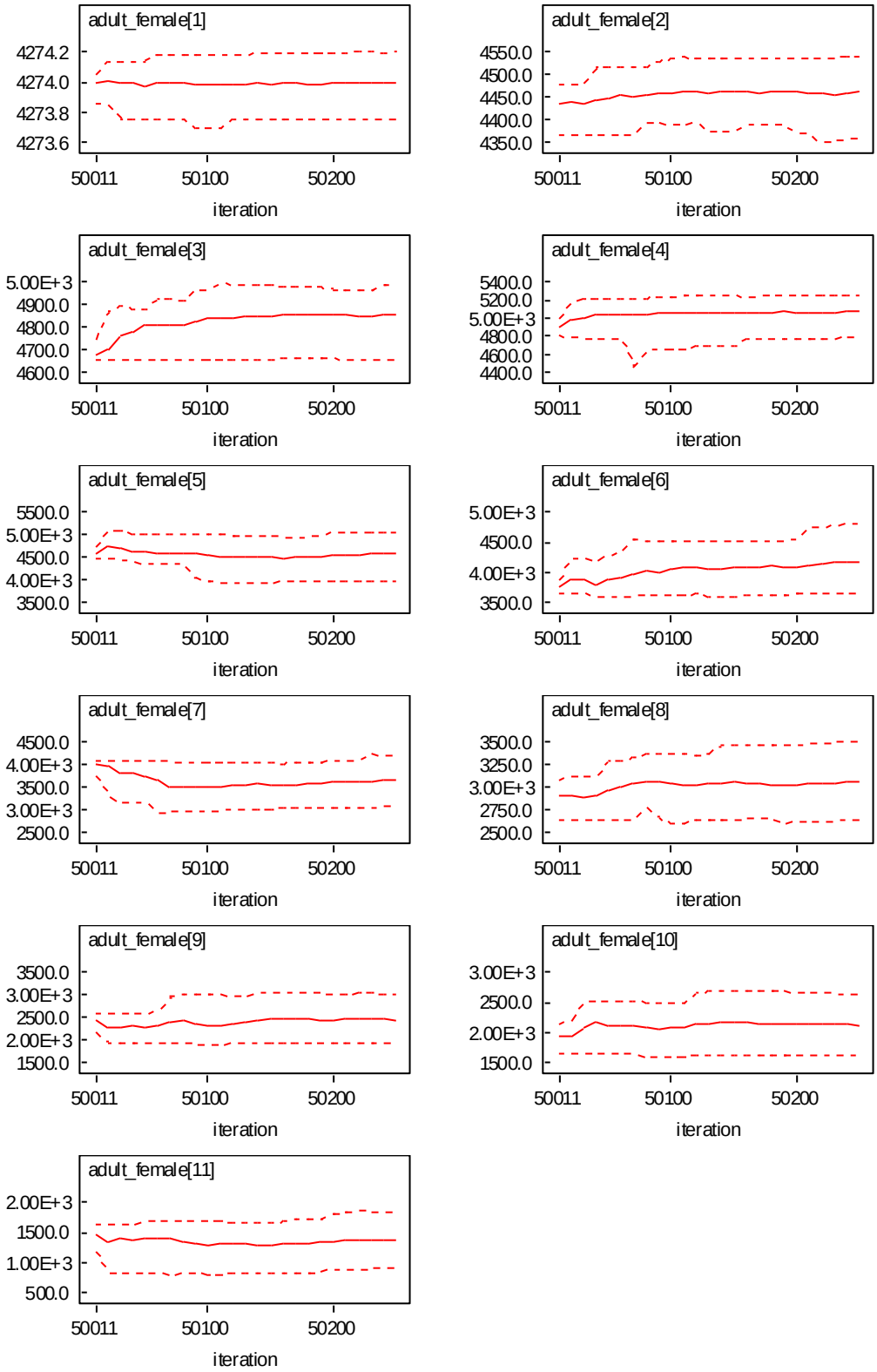
Running Quantiles: Adult Male Survival Rate – Model 1



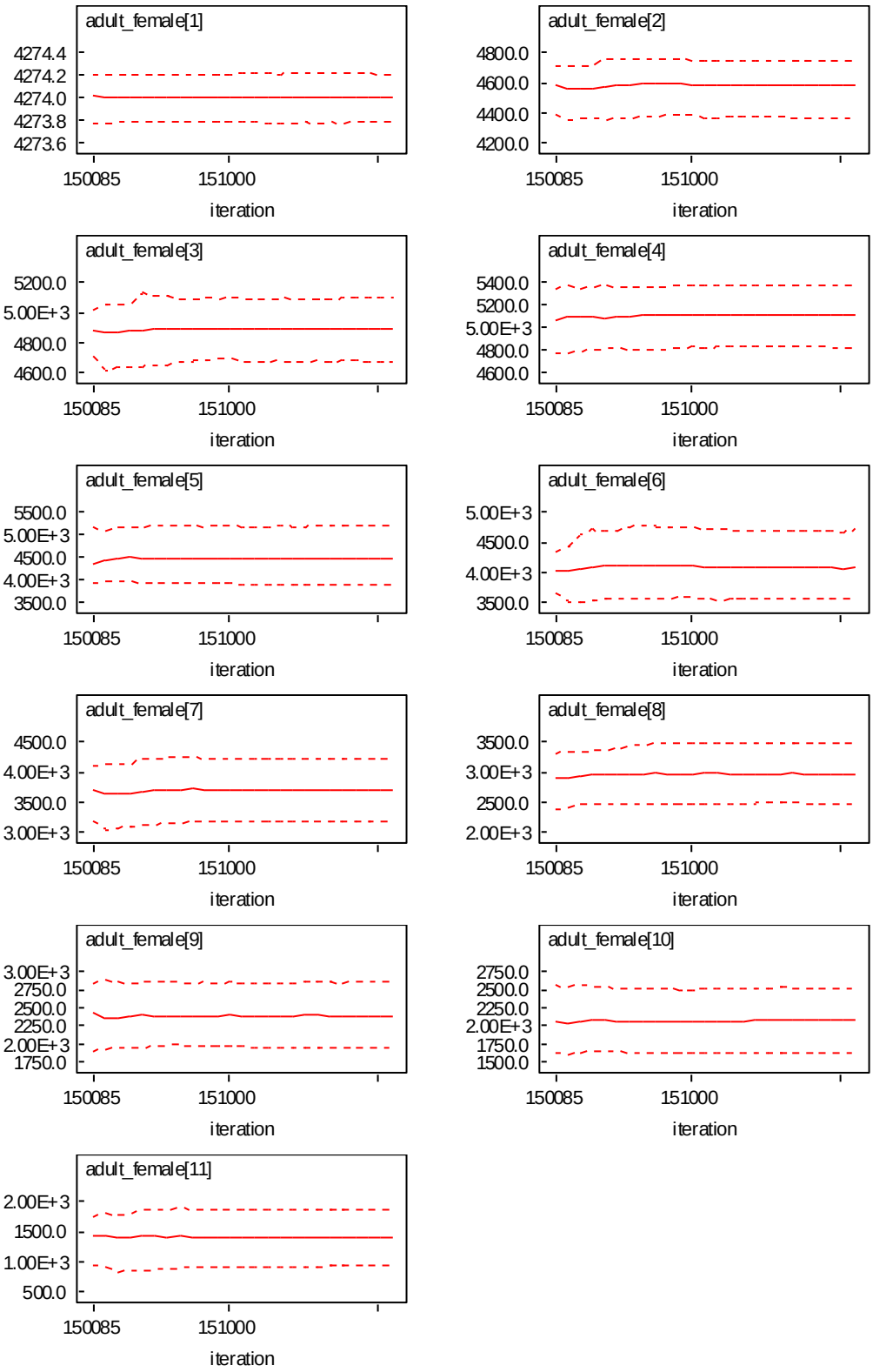
Running Quantiles: Adult Male Survival Rate – Model 2



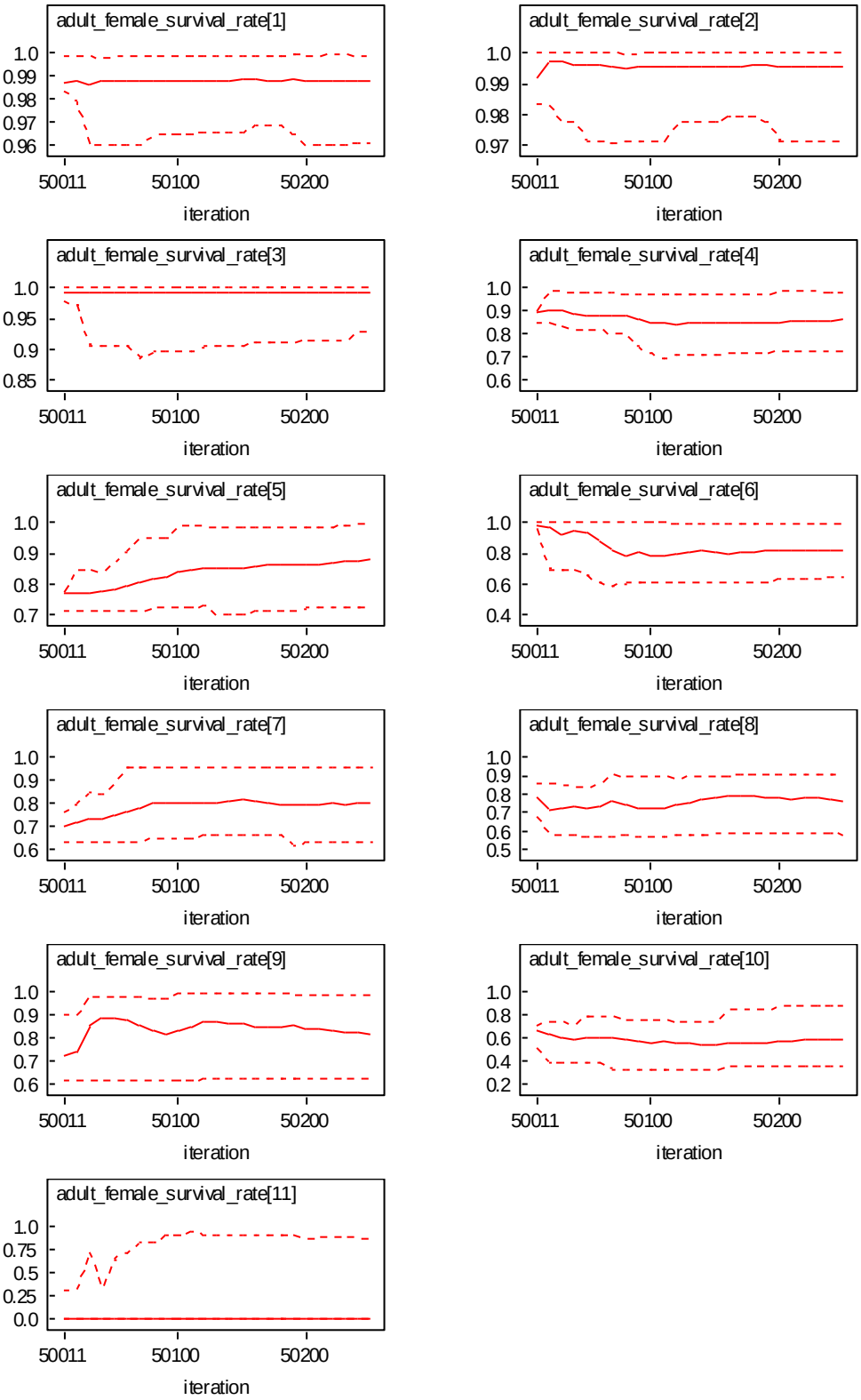
Running Quantiles: Number of Adult Females – Model 1



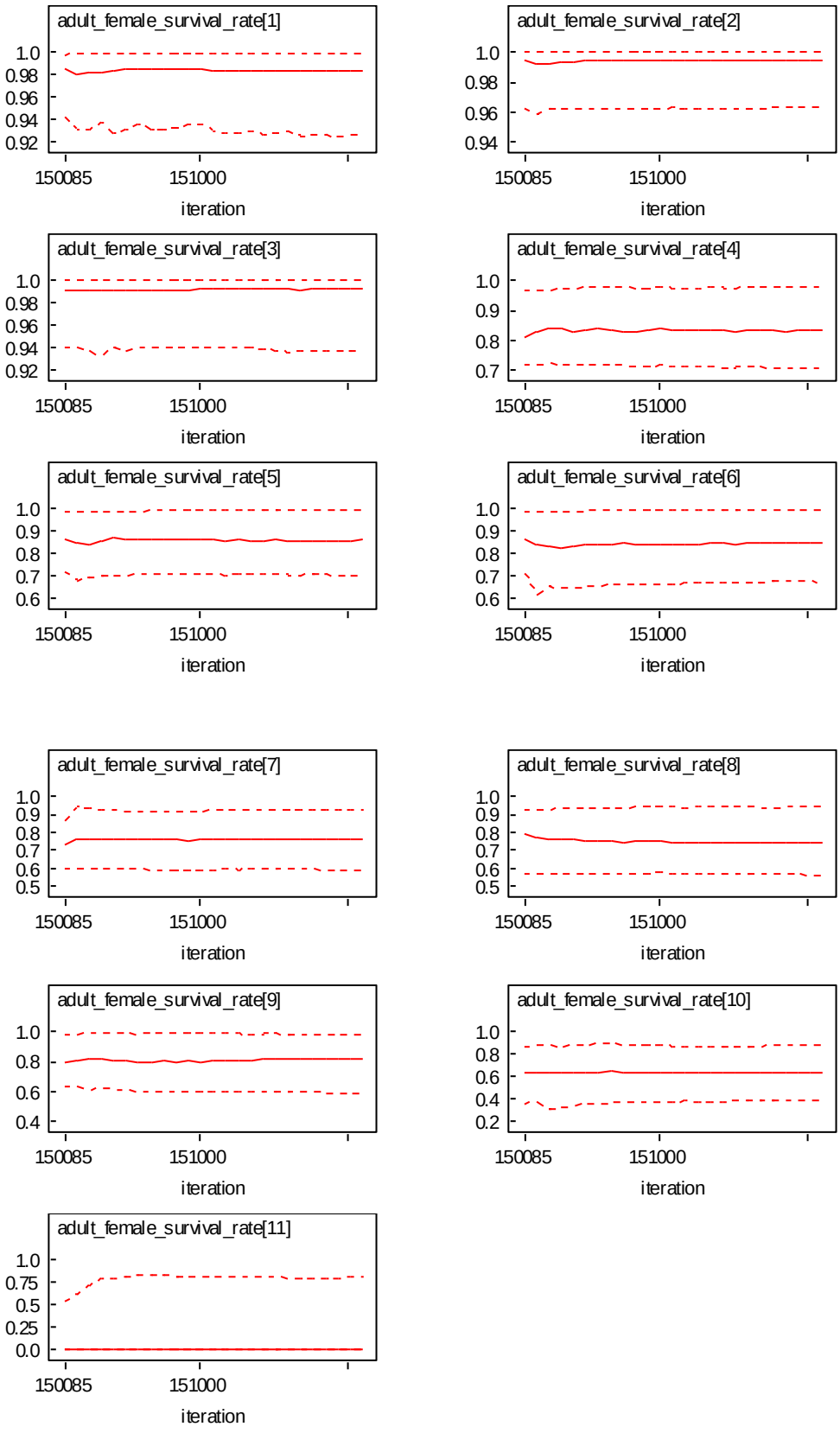
Running Quantiles: Number of Adult Females – Model 2



Running Quantiles: Adult Female Survival Rate – Model 1

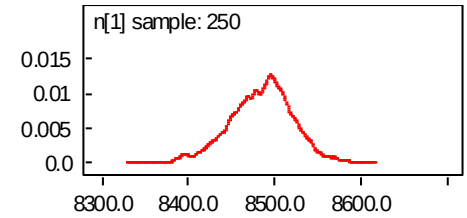


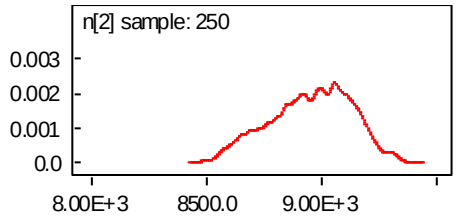
Running Quantiles: Adult Female Survival Rate – Model 2

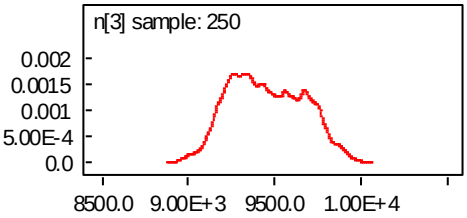


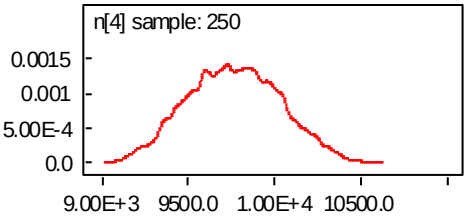
Appendix C: Kernel Density Estimates for Marginal Posterior Distribution Functions

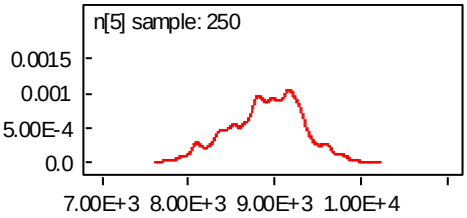
Kernel Density Estimates: Total Count – Model 1

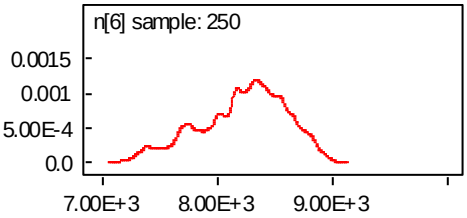


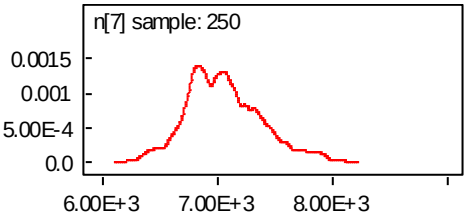


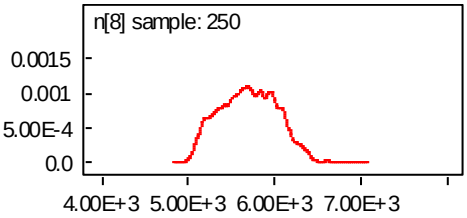


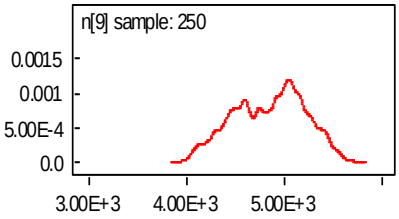


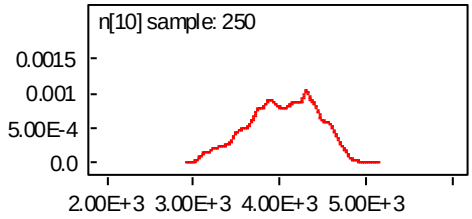


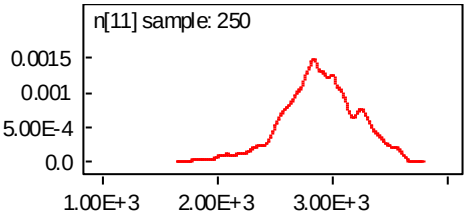




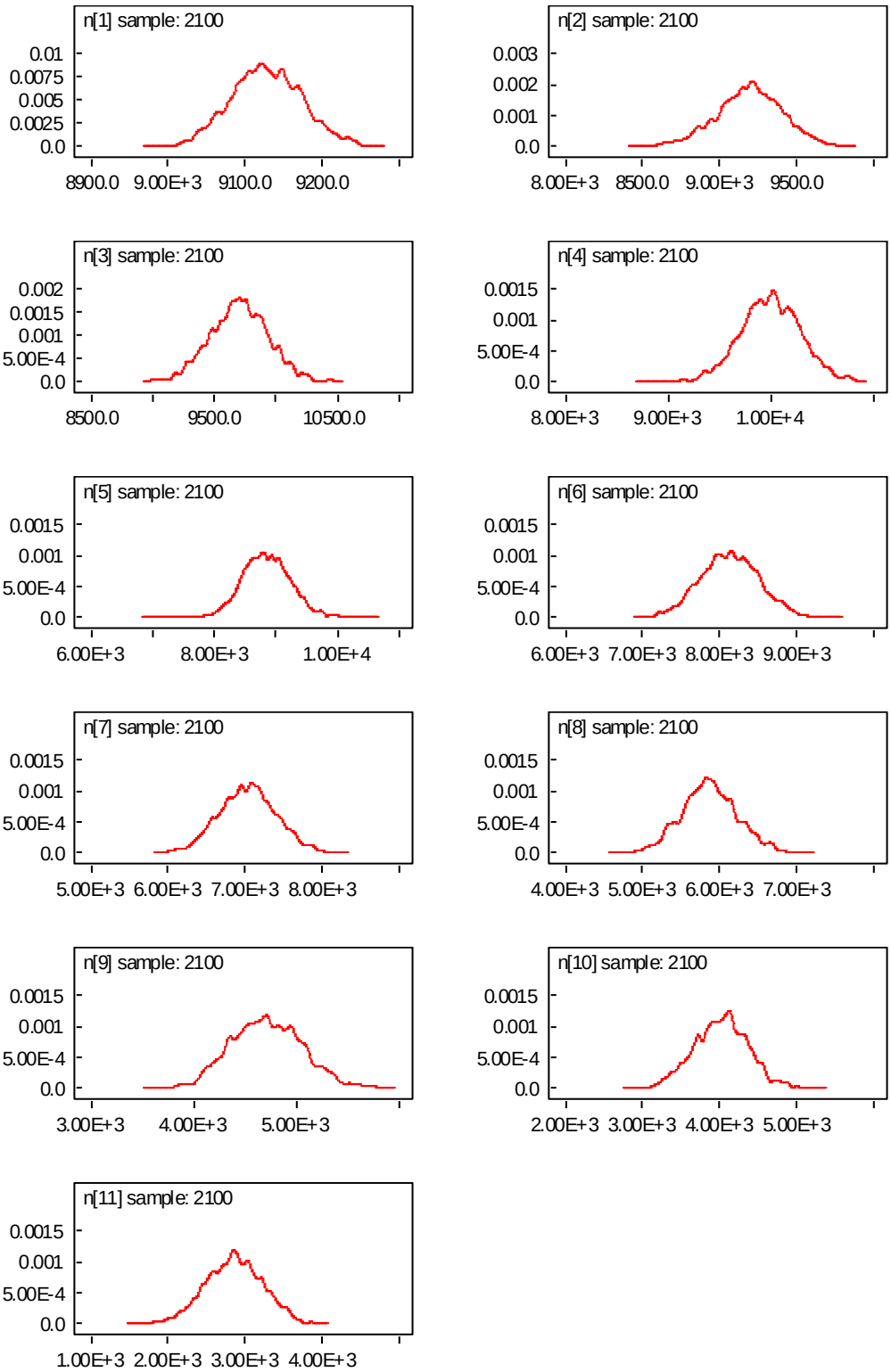




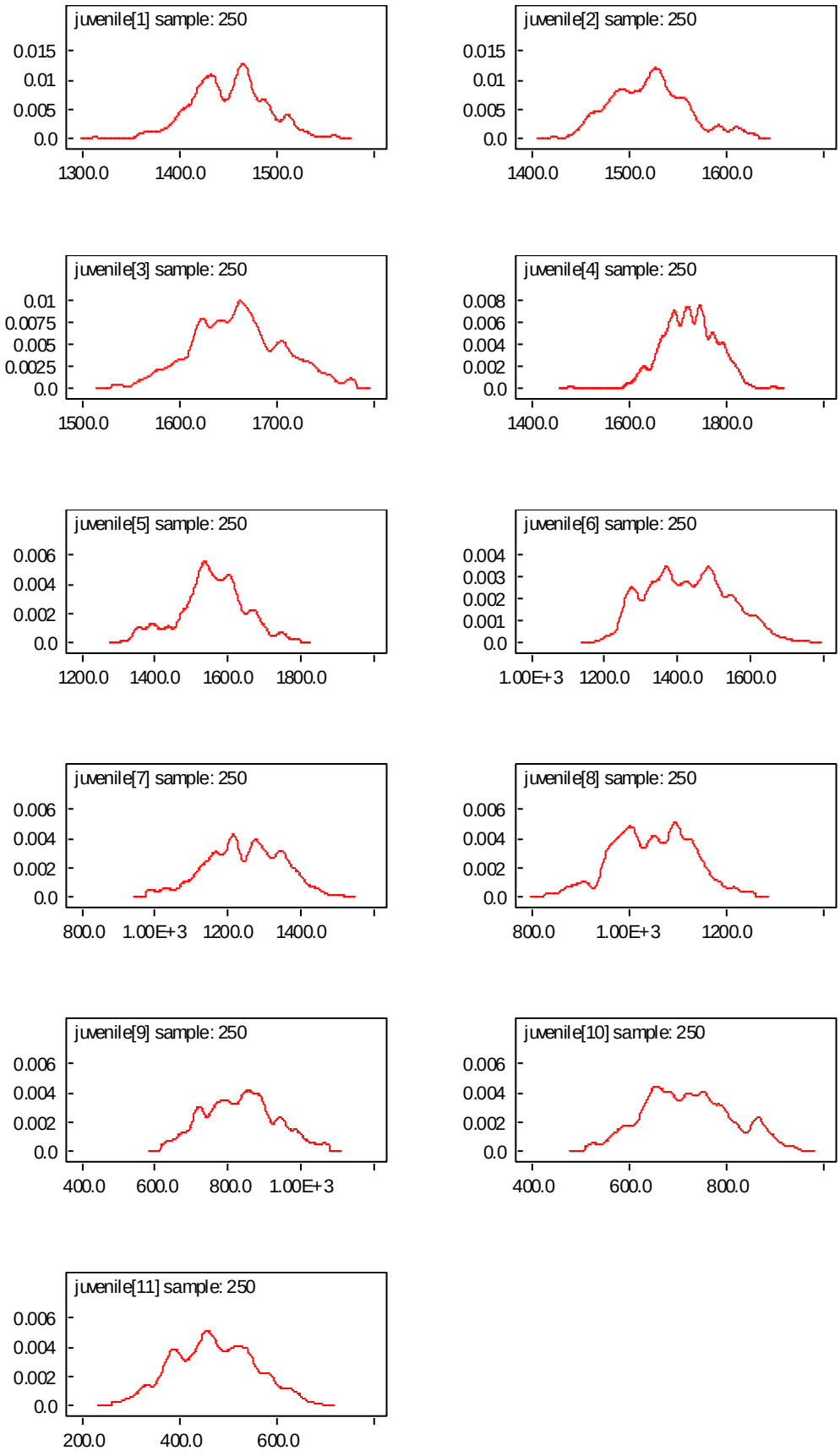



n[11] sample: 250
0.0015
0.001
5.00E-4
0.0
1.00E+3
2.00E+3
3.00E+3

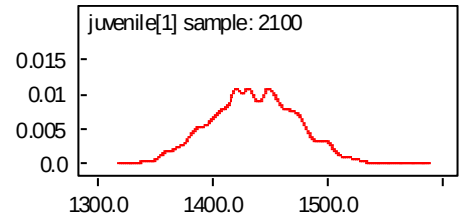
Kernel Density Estimates: Total Count – Model 2

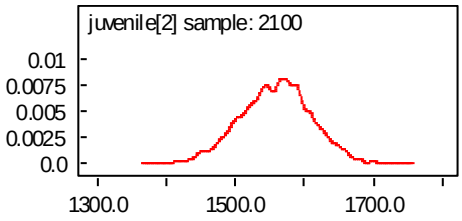


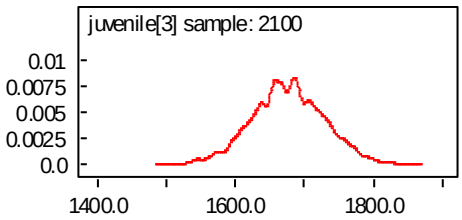
Kernel Density Estimates: Number of Juveniles – Model 1

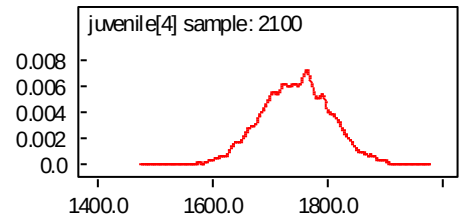


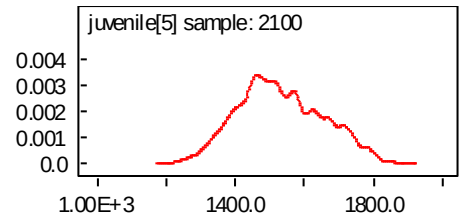
Kernel Density Estimates: Number of Juveniles – Model 2

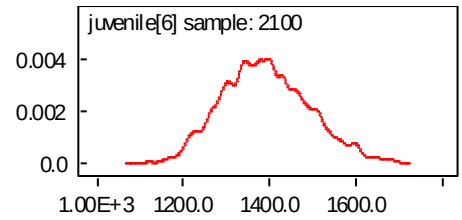


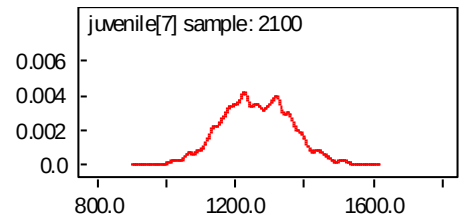


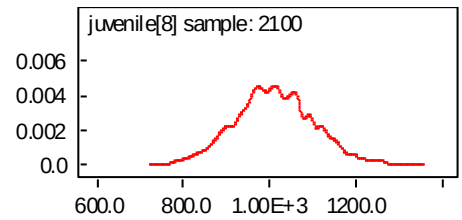


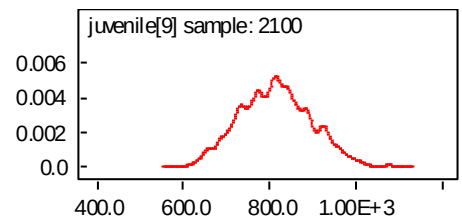


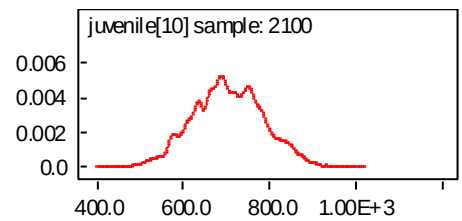


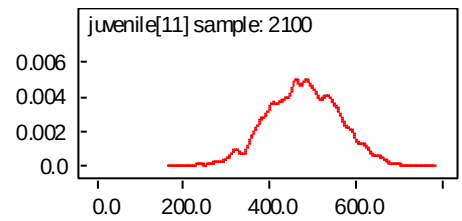




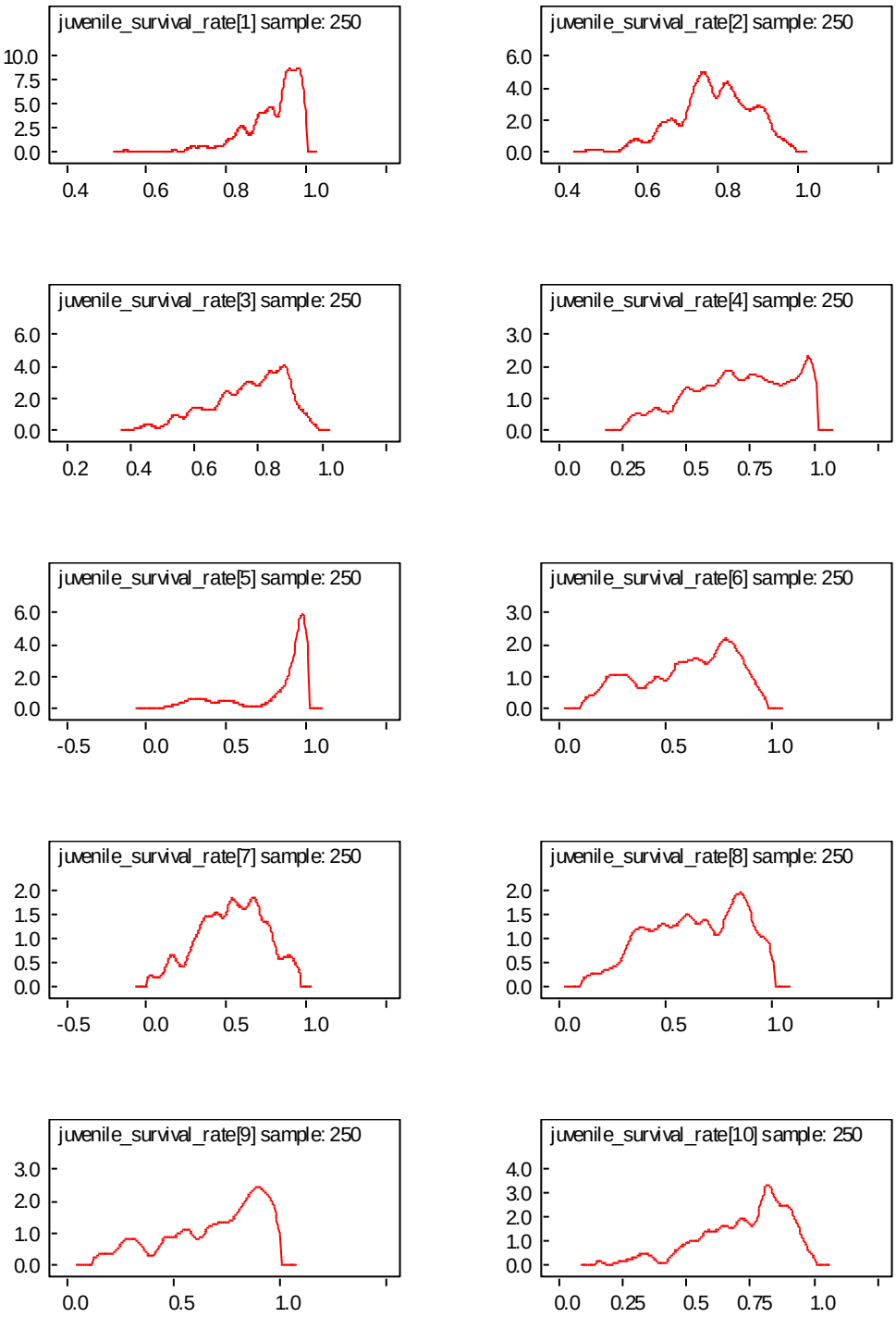


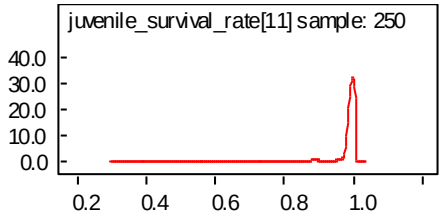




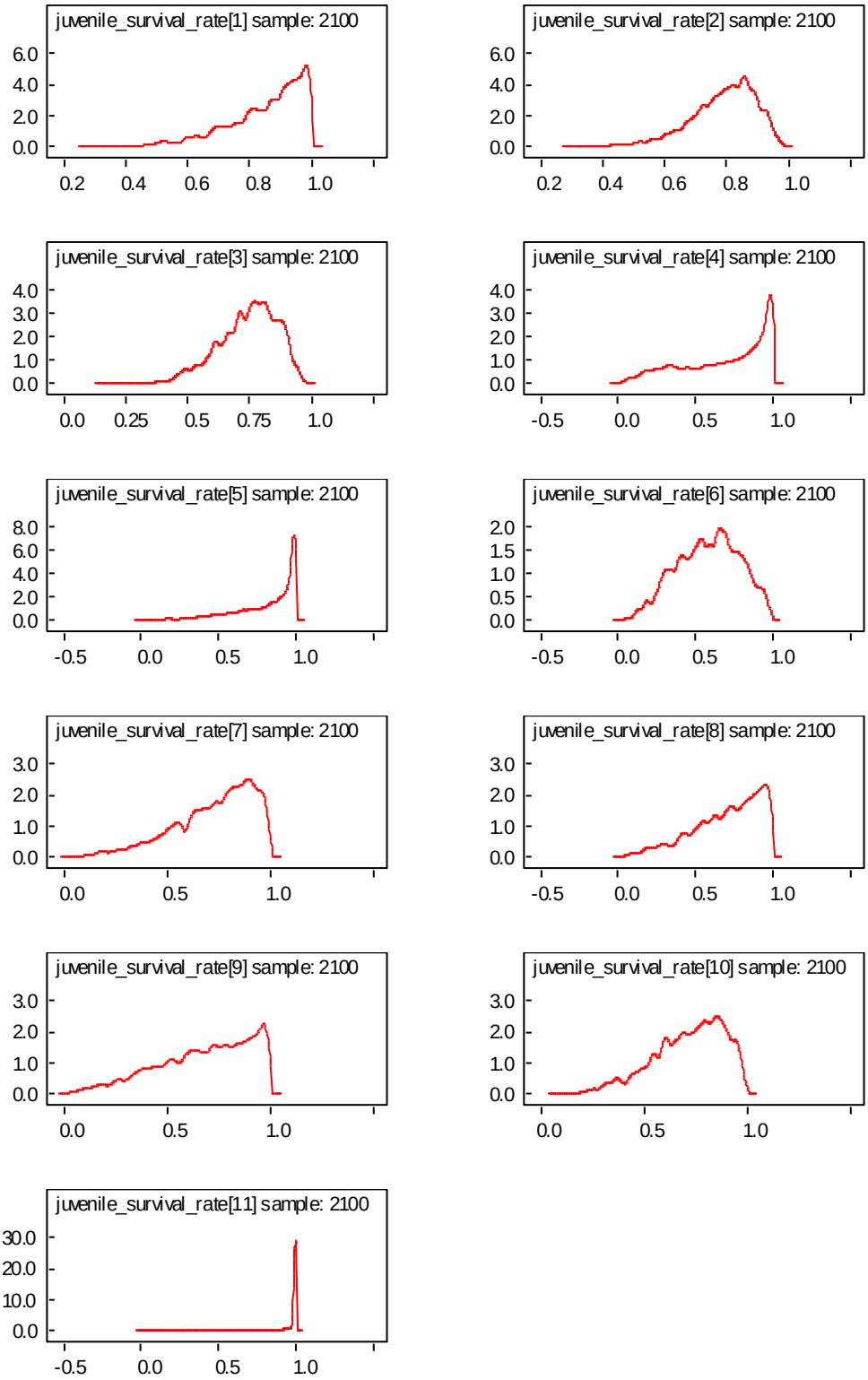


Kernel Density Estimates: Juvenile Survival Rate – Model 1

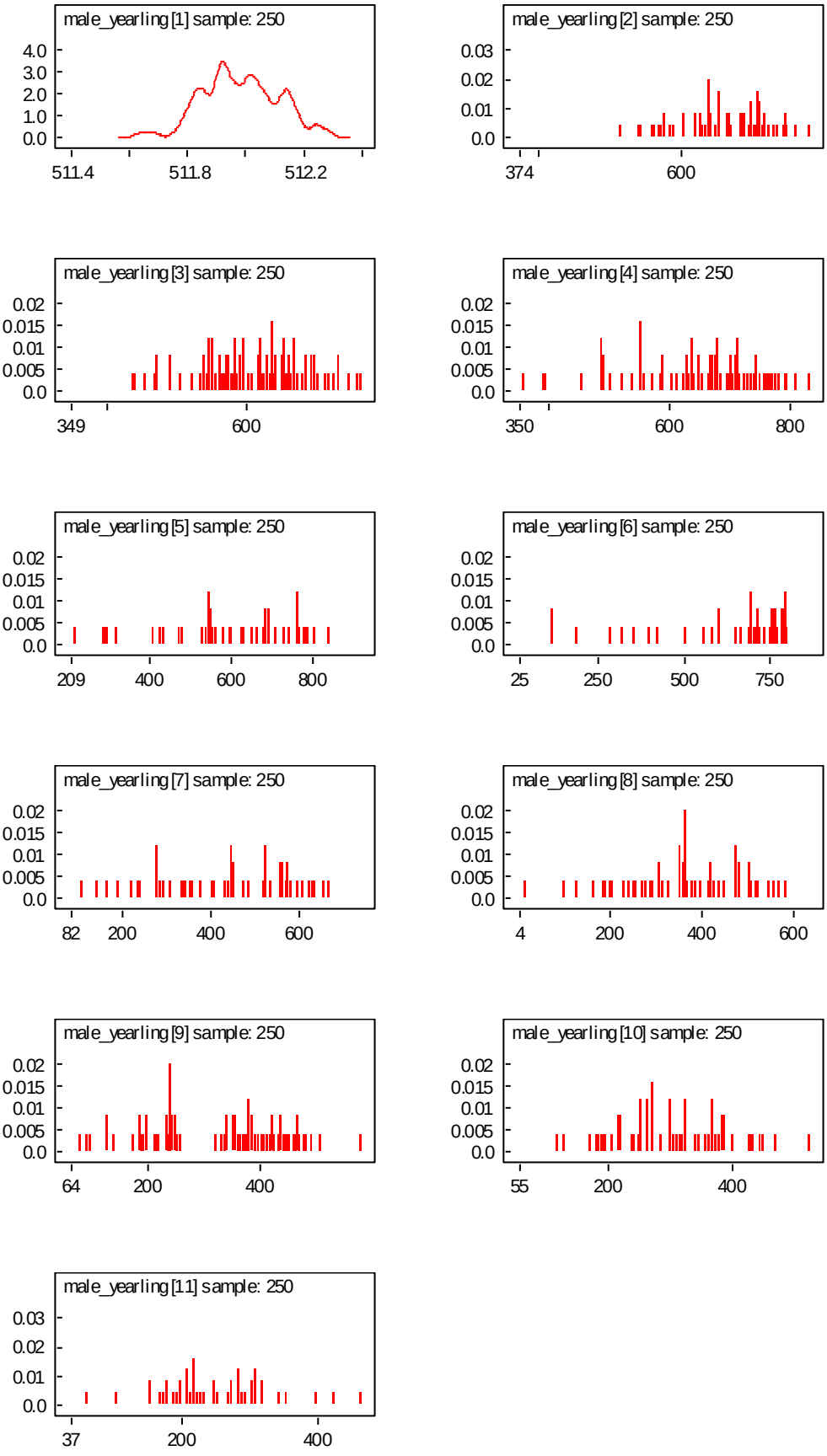



juvenile_survival_rate[11] sample: 250
40.0
30.0
20.0
10.0
0.0
0.2
0.4
0.6
0.8
1.0

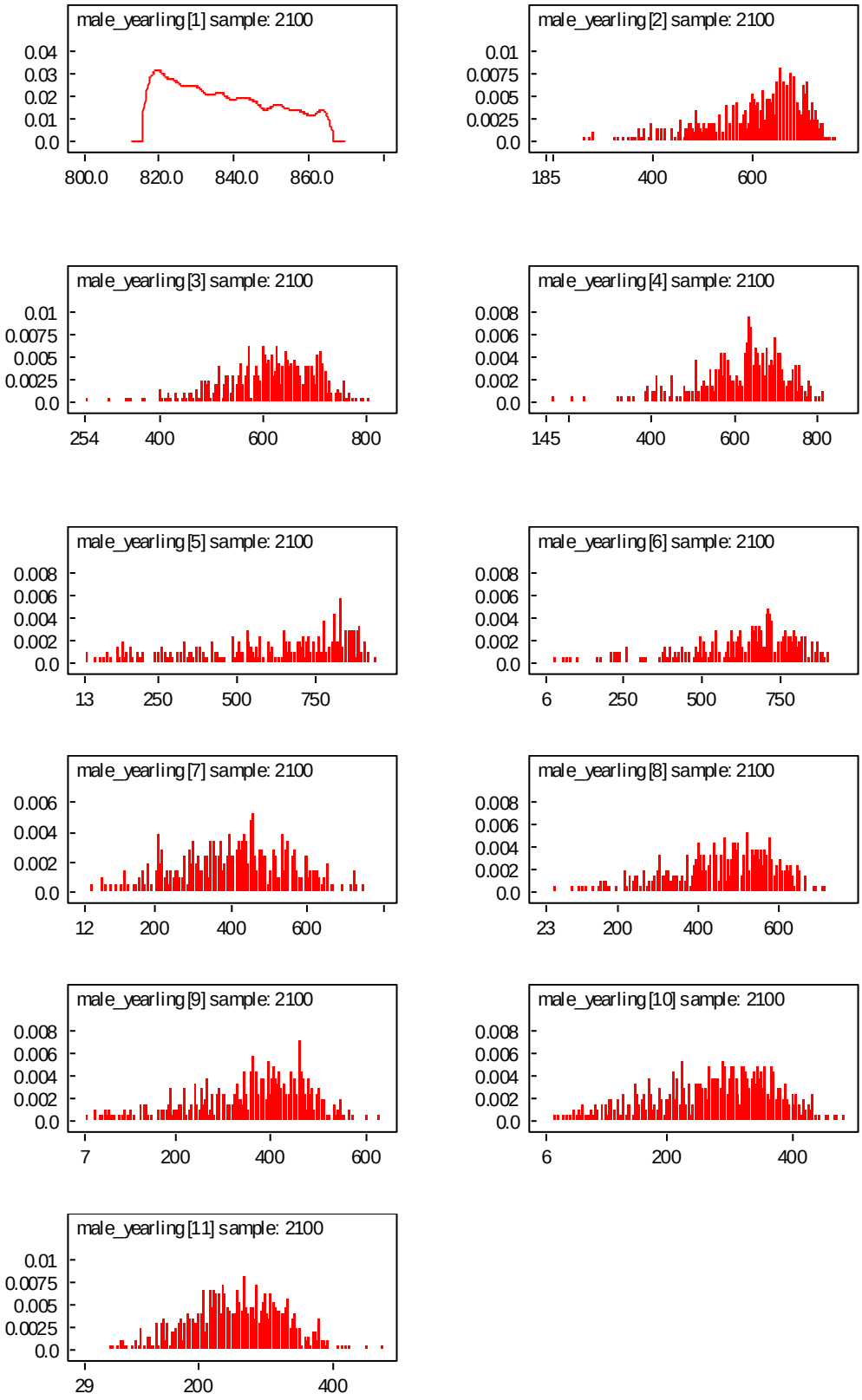
Kernel Density Estimates: Juvenile Survival Rate – Model 2



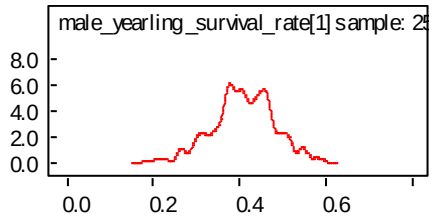
Kernel Density Estimates: Number of Male Yearlings - Model 1

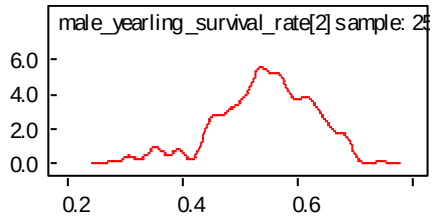


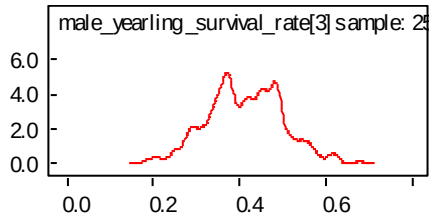
Kernel Density Estimates: Number of Male Yearlings - Model 2

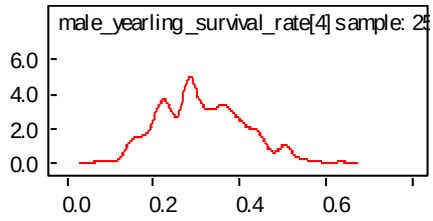


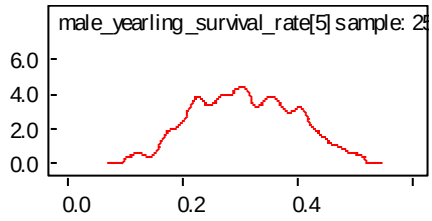
Kernel Density Estimates: Male Yearling Survival Rate – Model 1

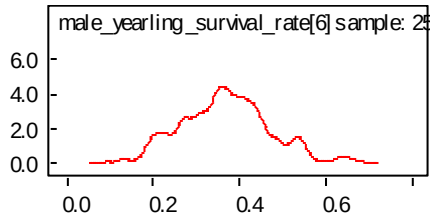


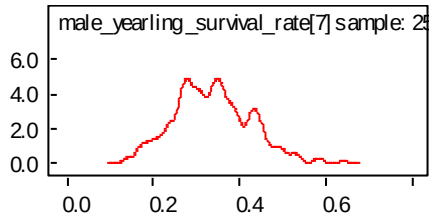


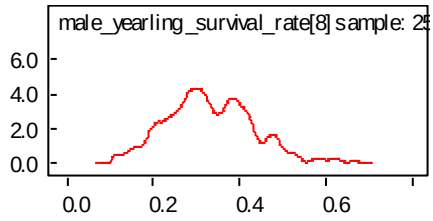


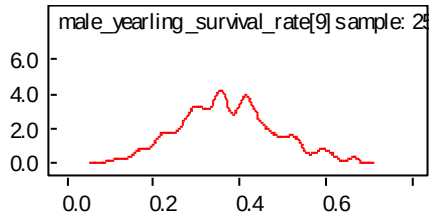


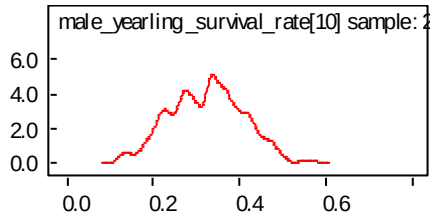


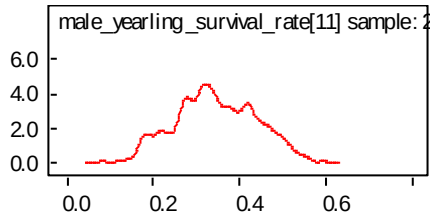




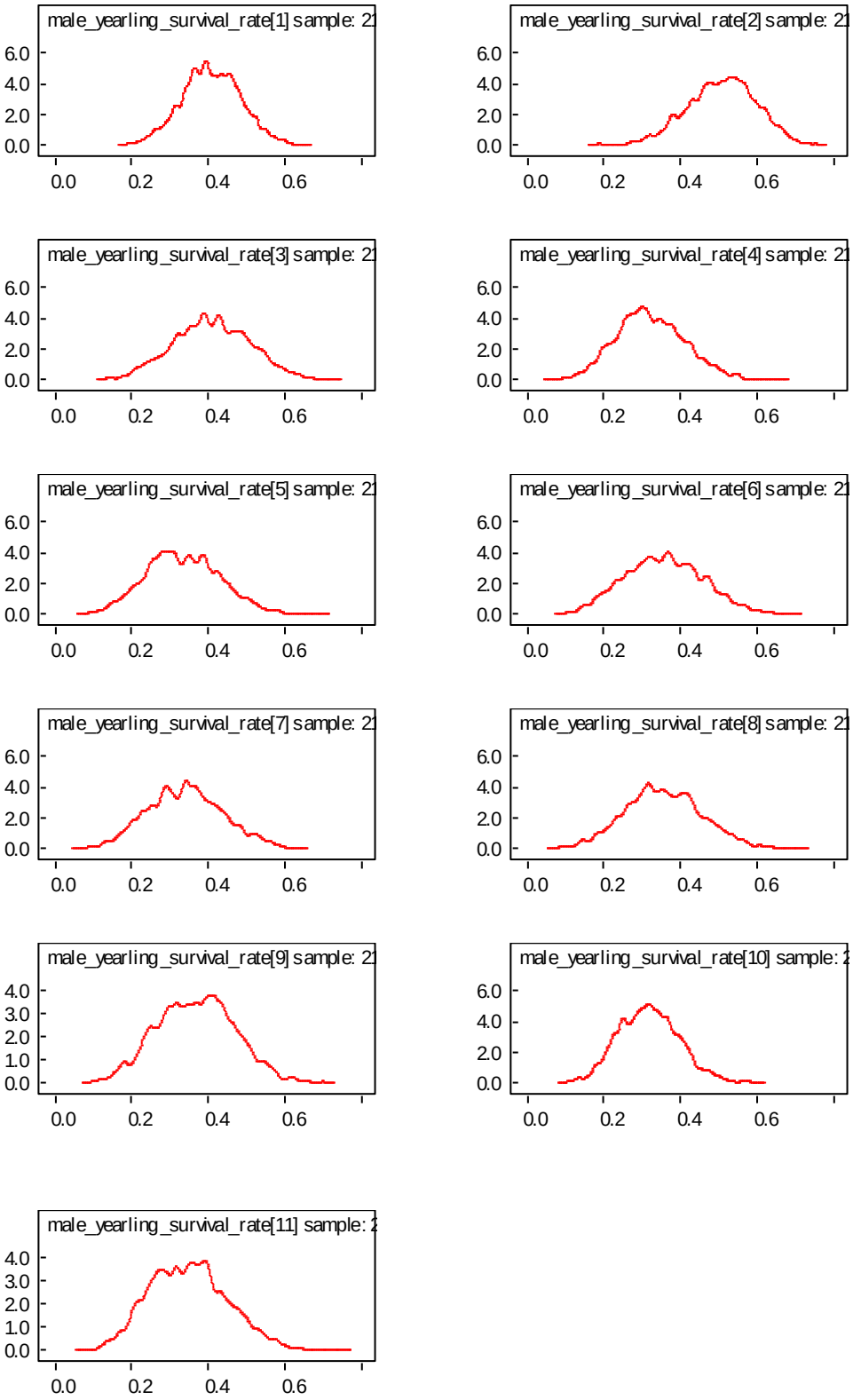




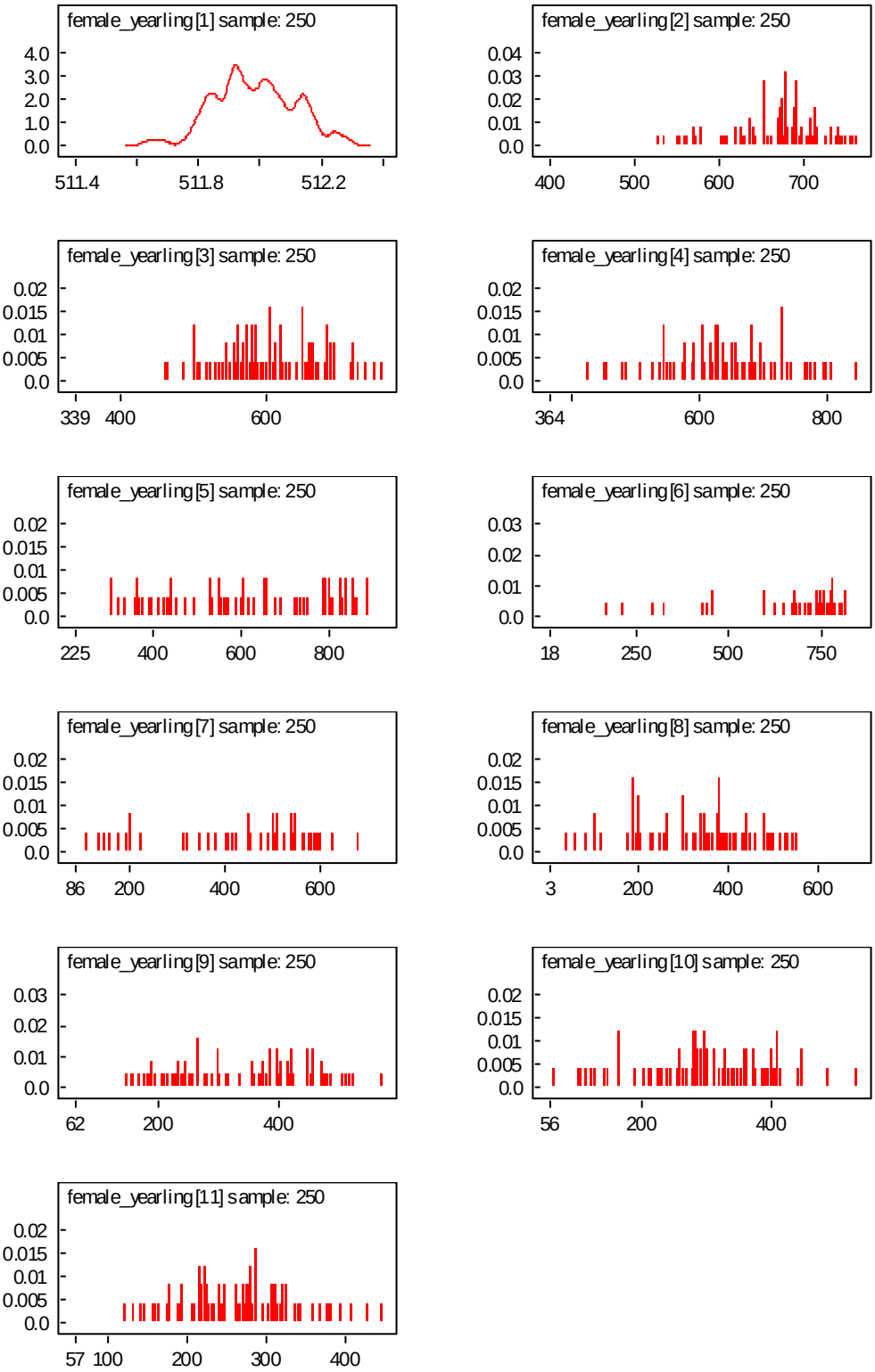




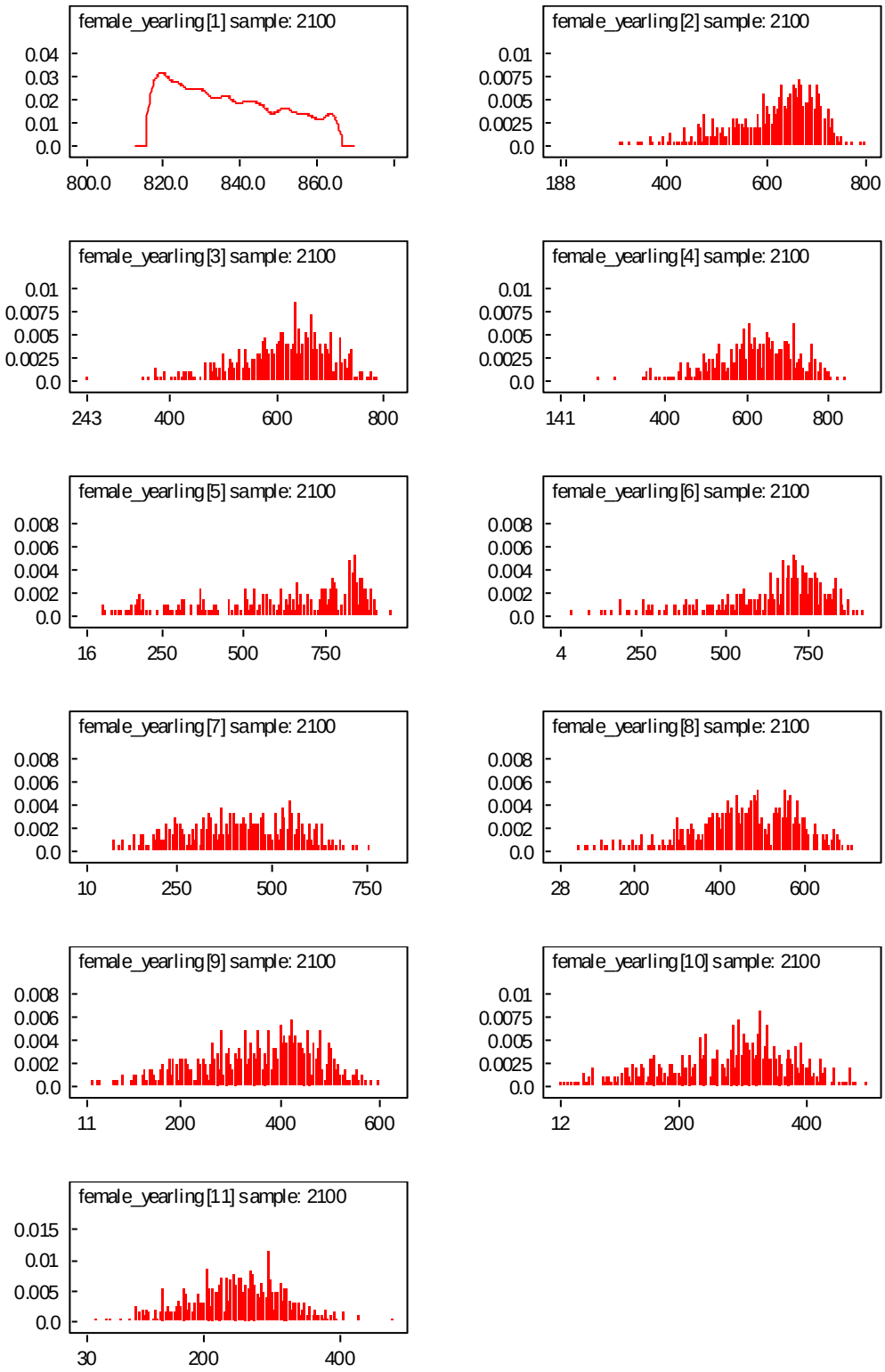
Kernel Density Estimates: Male Yearling Survival Rate – Model 2



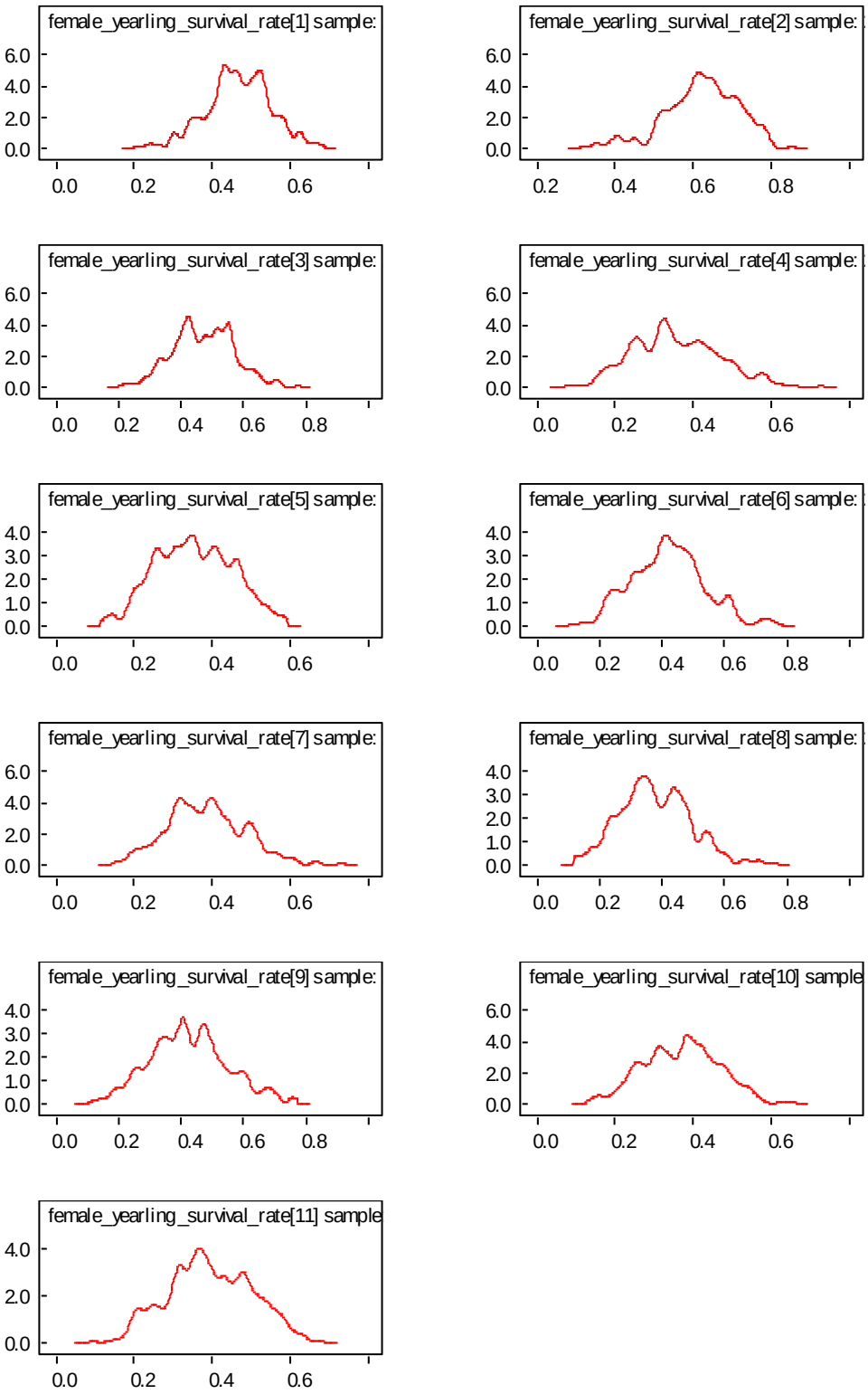
Kernel Density Function: Number of Female Yearlings – Model 1



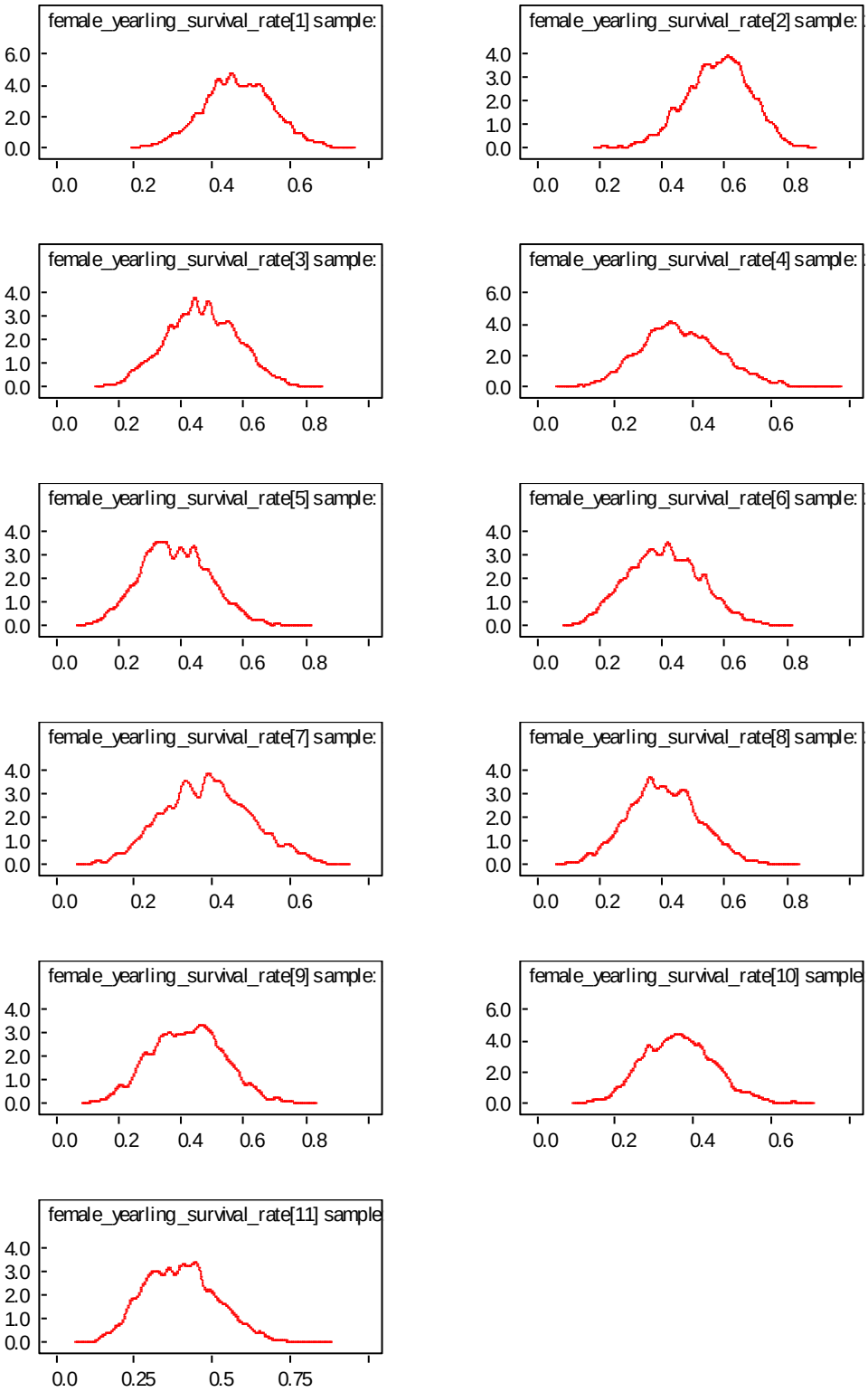
Kernel Density Function: Number of Female Yearlings – Model 2



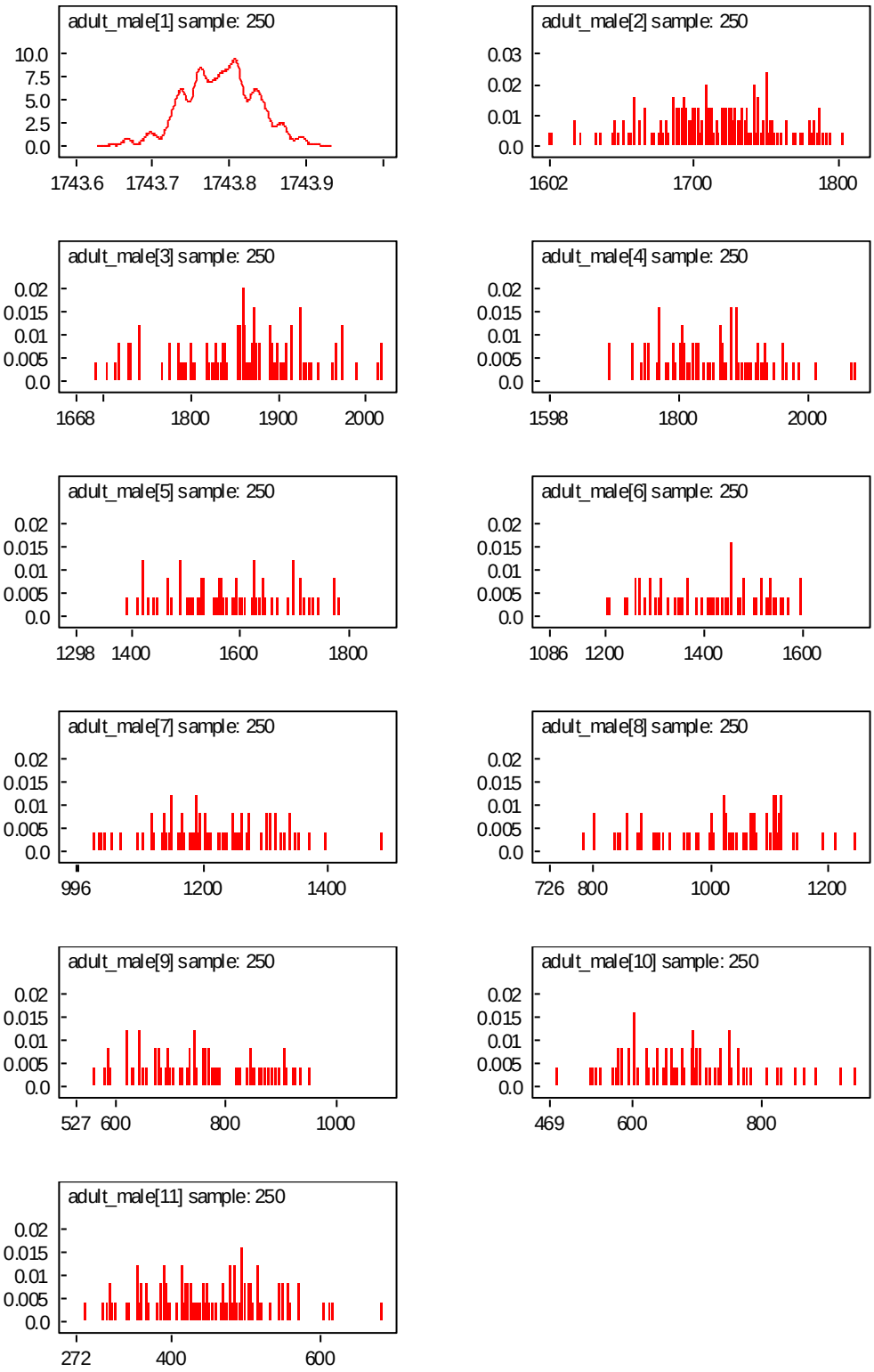
Kernel Density Estimates: Female Yearling Survival Rate – Model 1



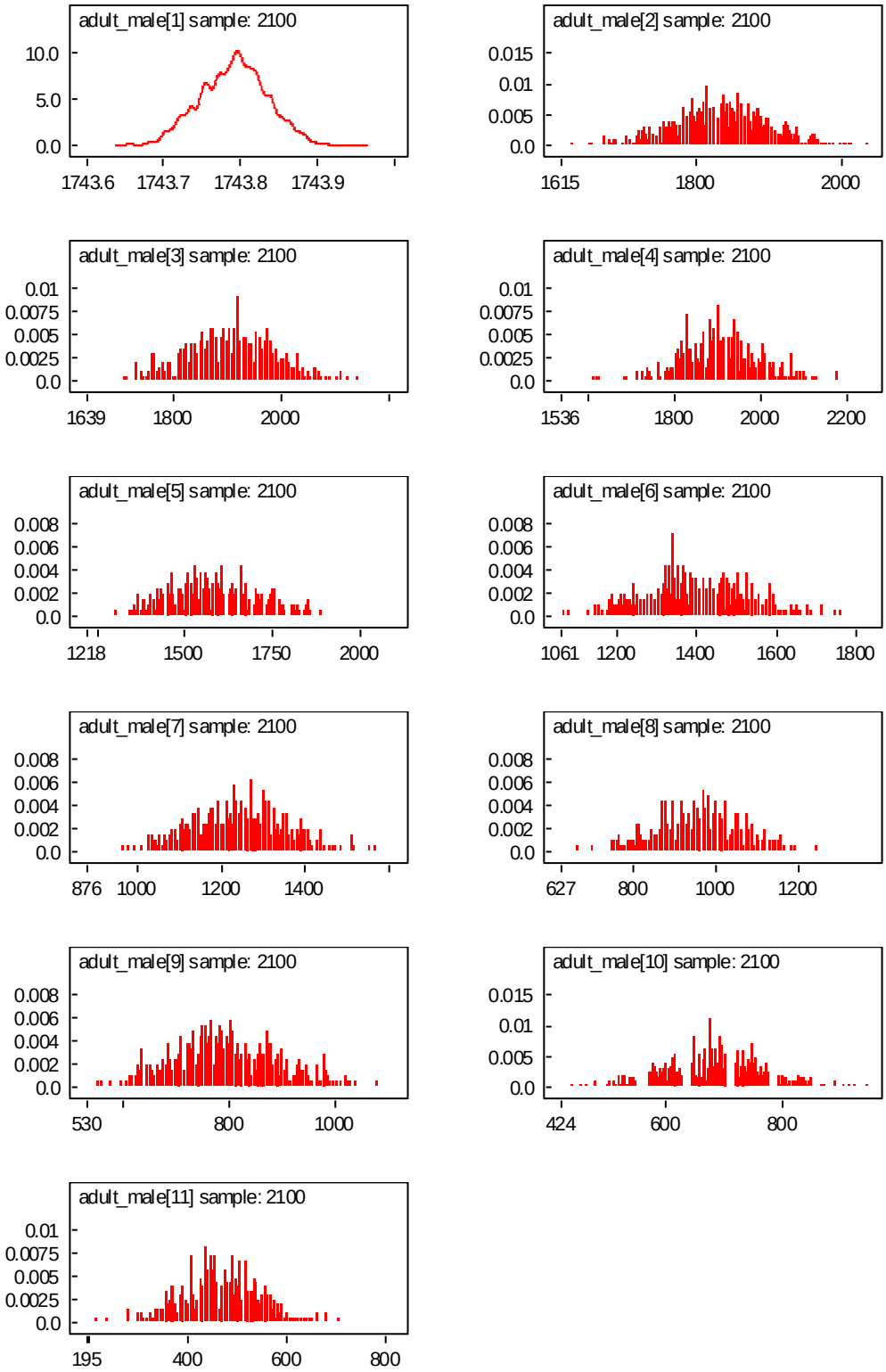
Kernel Density Estimates: Female Yearling Survival Rate – Model 2



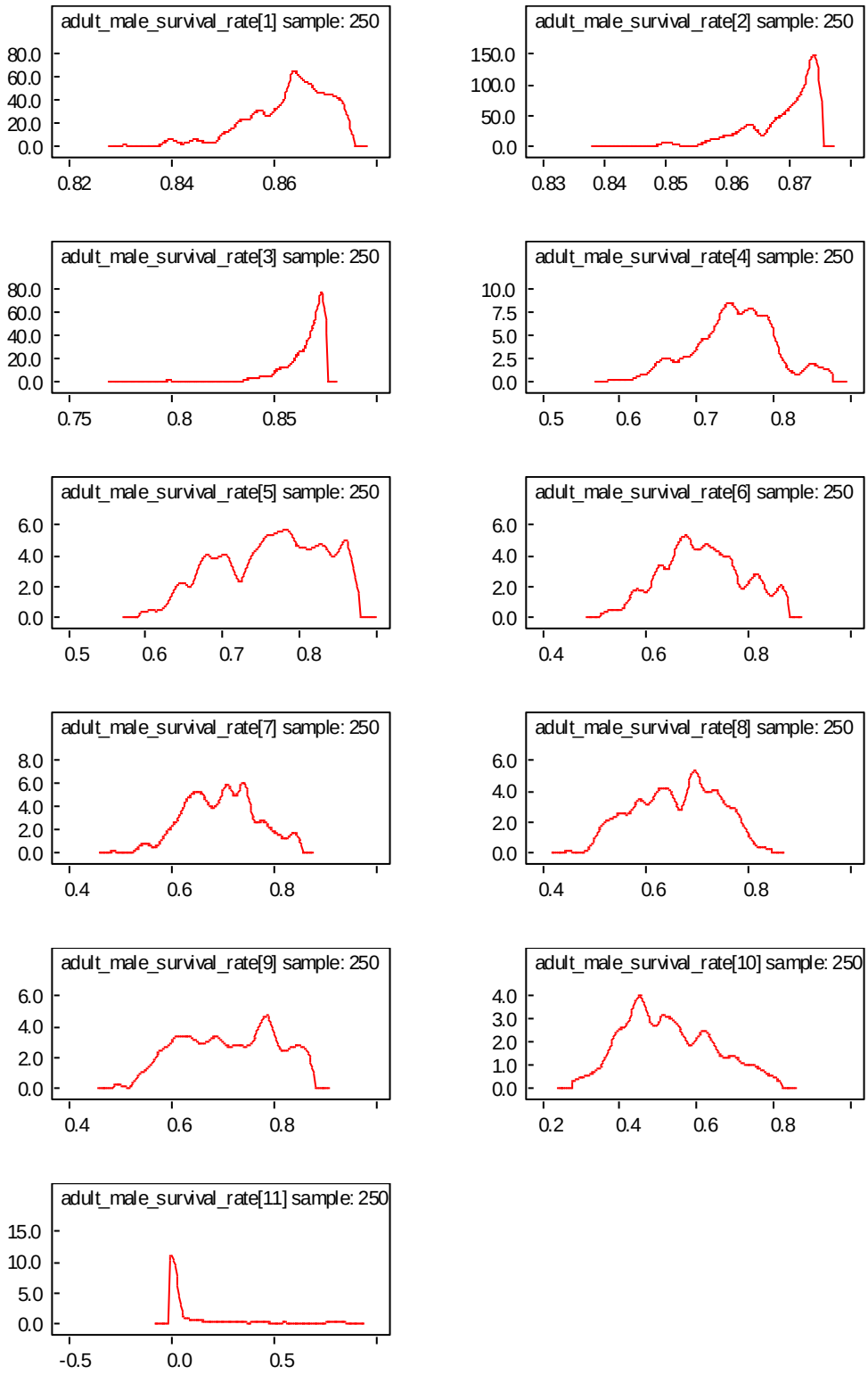
Kernel Density Estimate: Number of Adult Males – Model 1



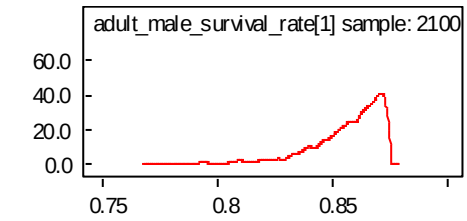
Kernel Density Estimate: Number of Adult Males – Model 2

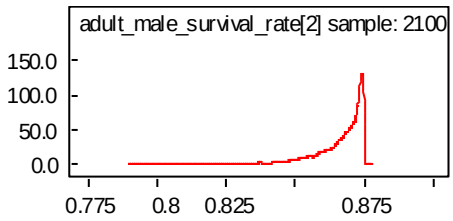


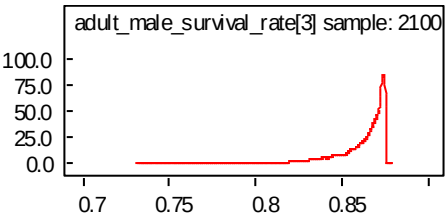
Kernel Density Estimate: Adult Male Survival Rate – Model 1

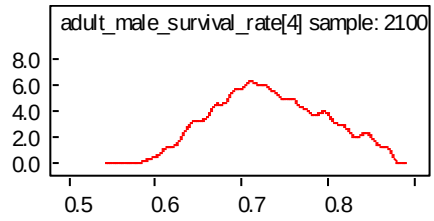


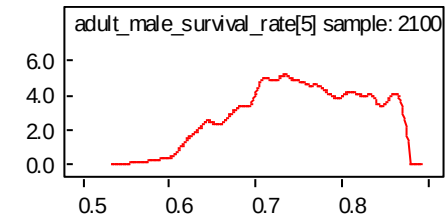
Kernel Density Estimate: Adult Male Survival Rate – Model 2

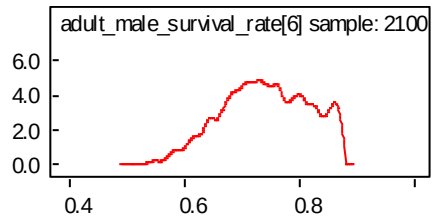


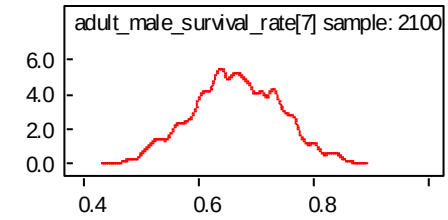


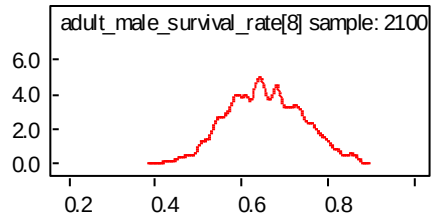


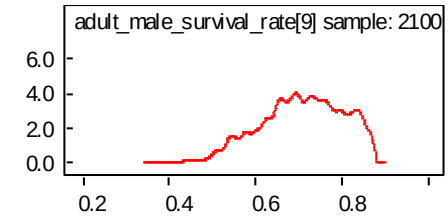


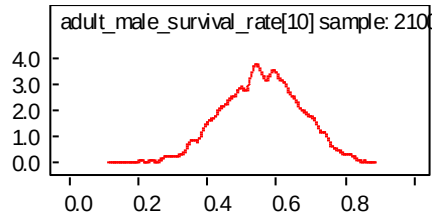


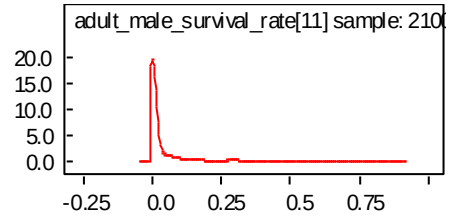




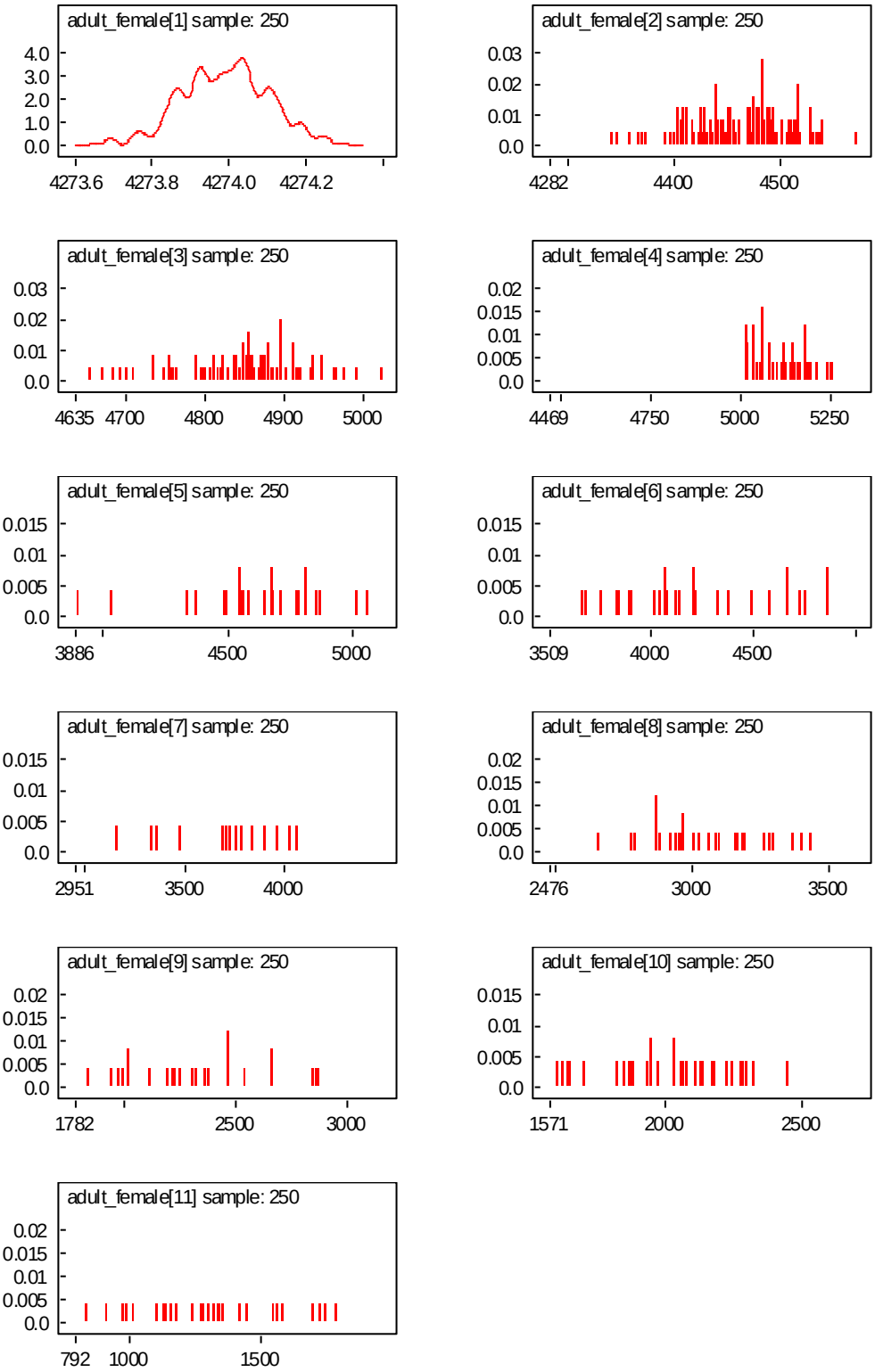




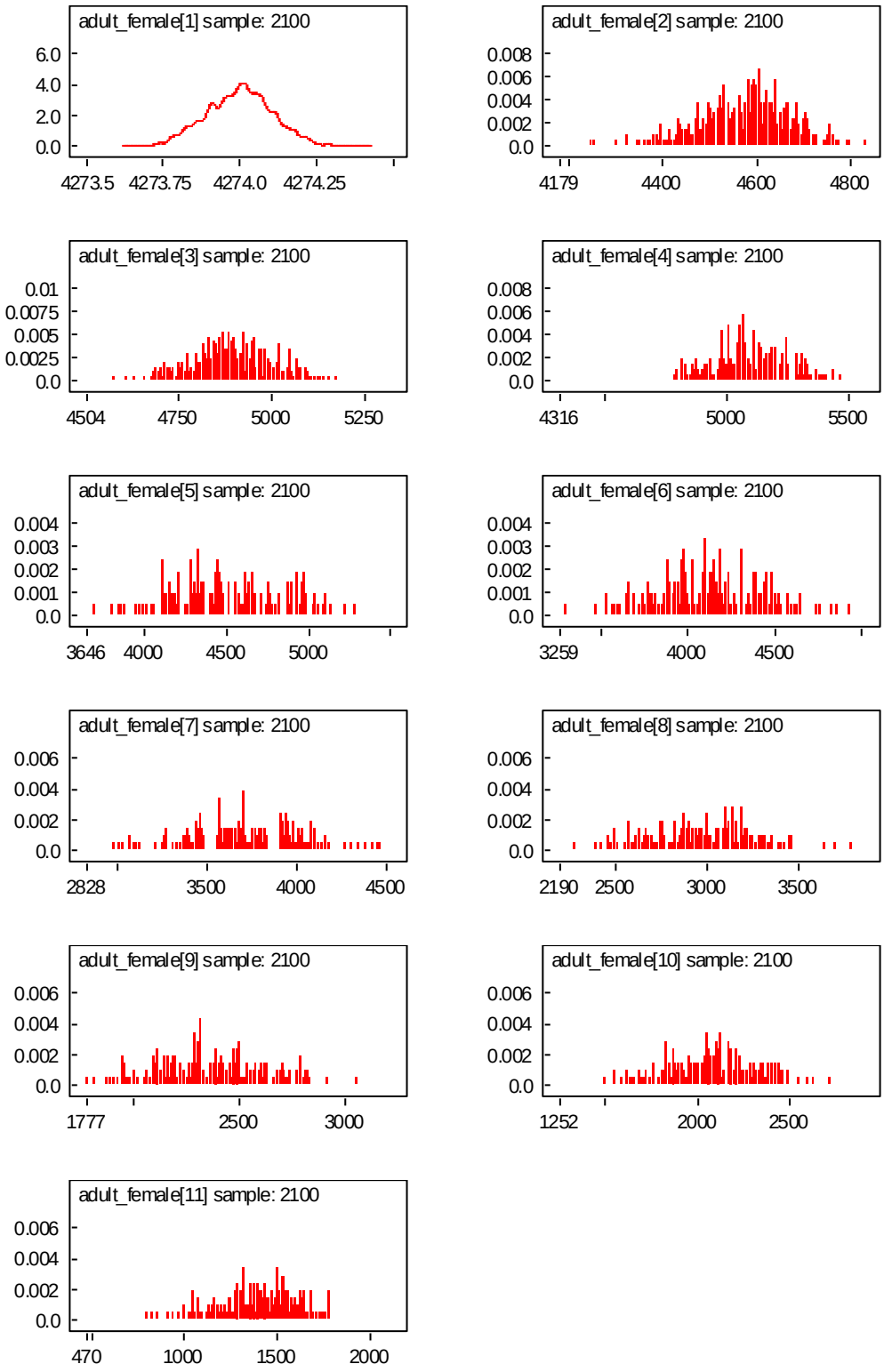




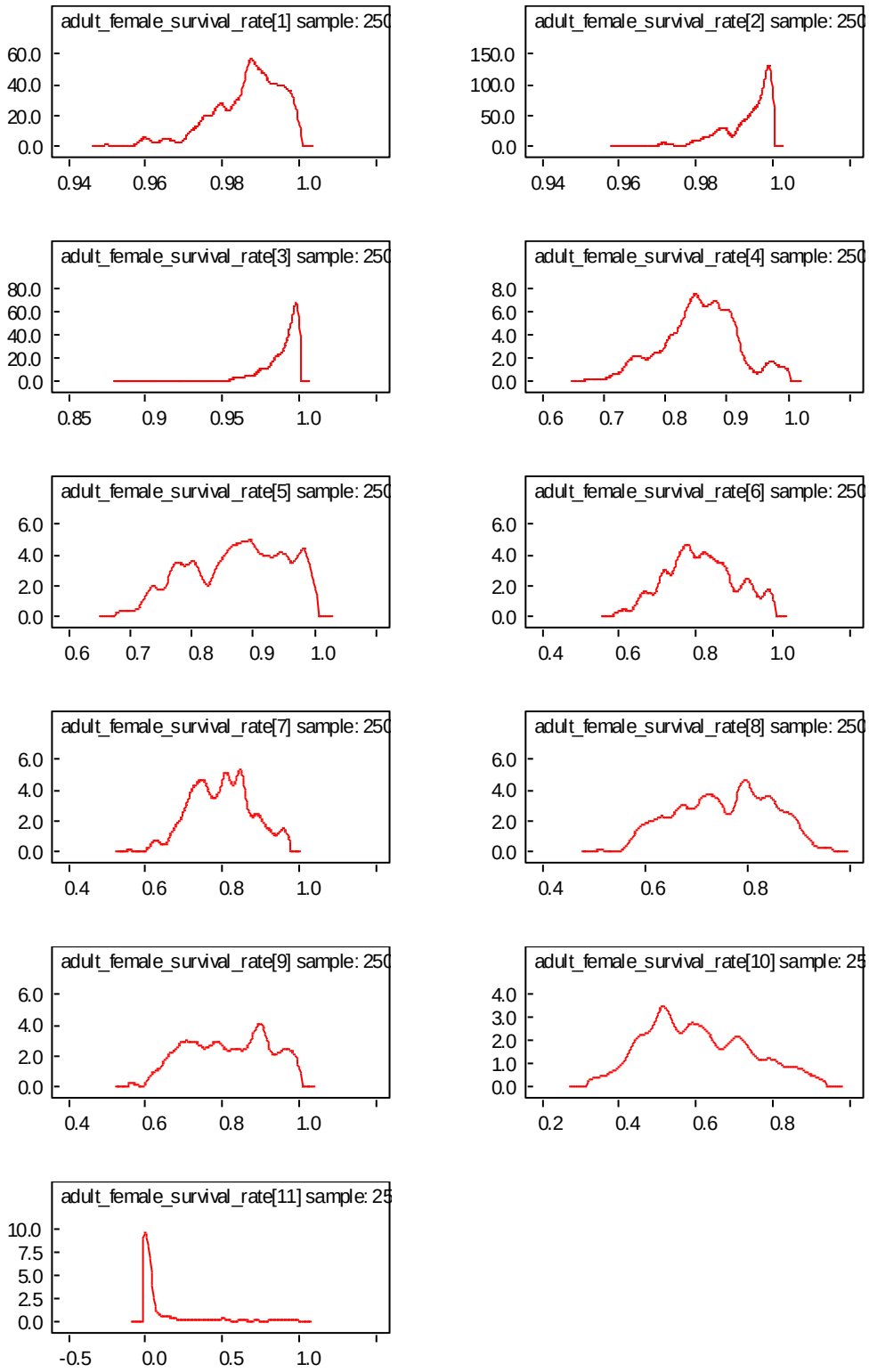
Kernel Density Estimate: Number of Adult Females – Model 1



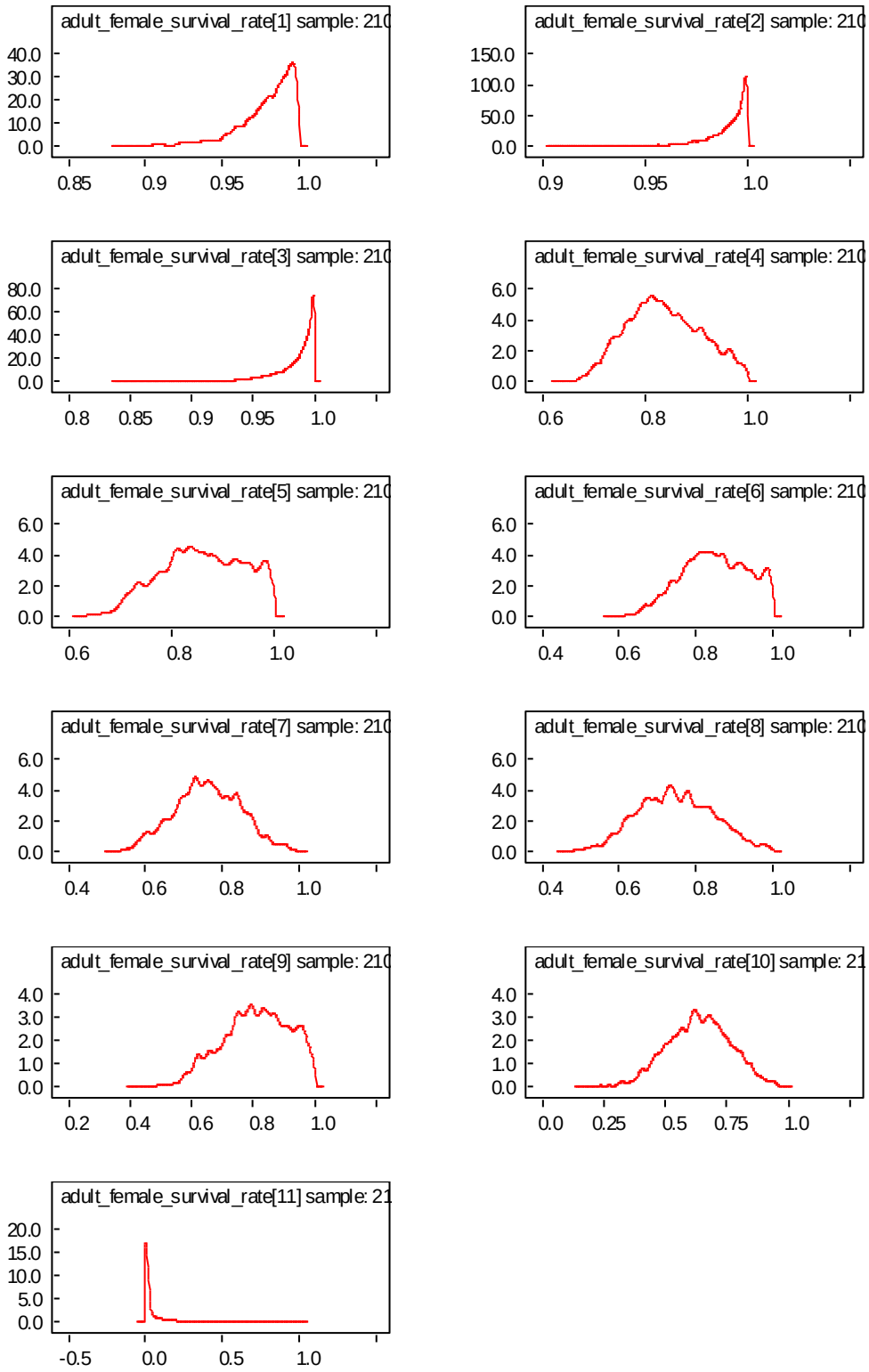
Kernel Density Estimate: Number of Adult Females – Model 2



Kernel Density Estimates: Adult Female Survival Rate – Model 1

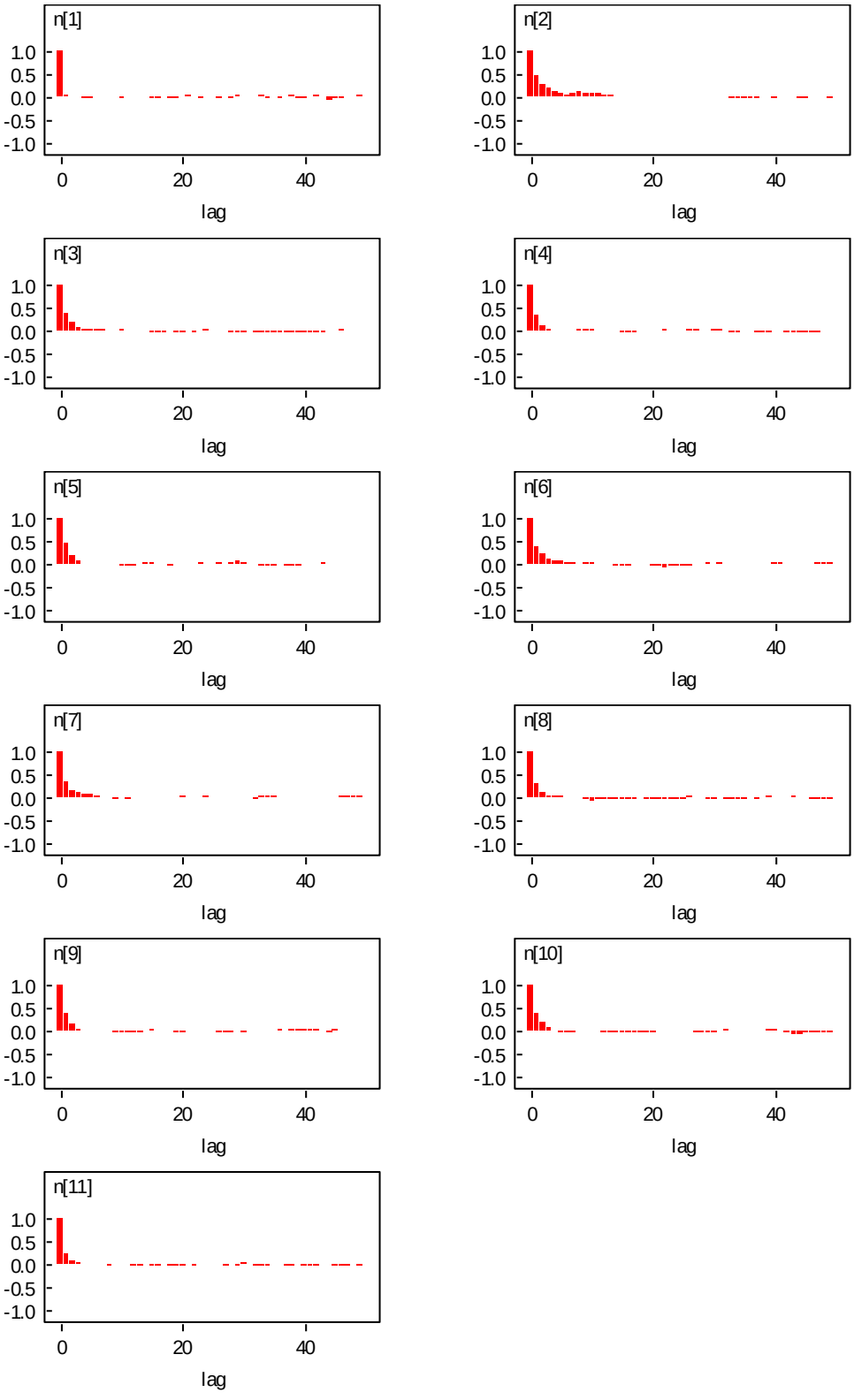


Kernel Density Estimates: Adult Female Survival Rate – Model 2

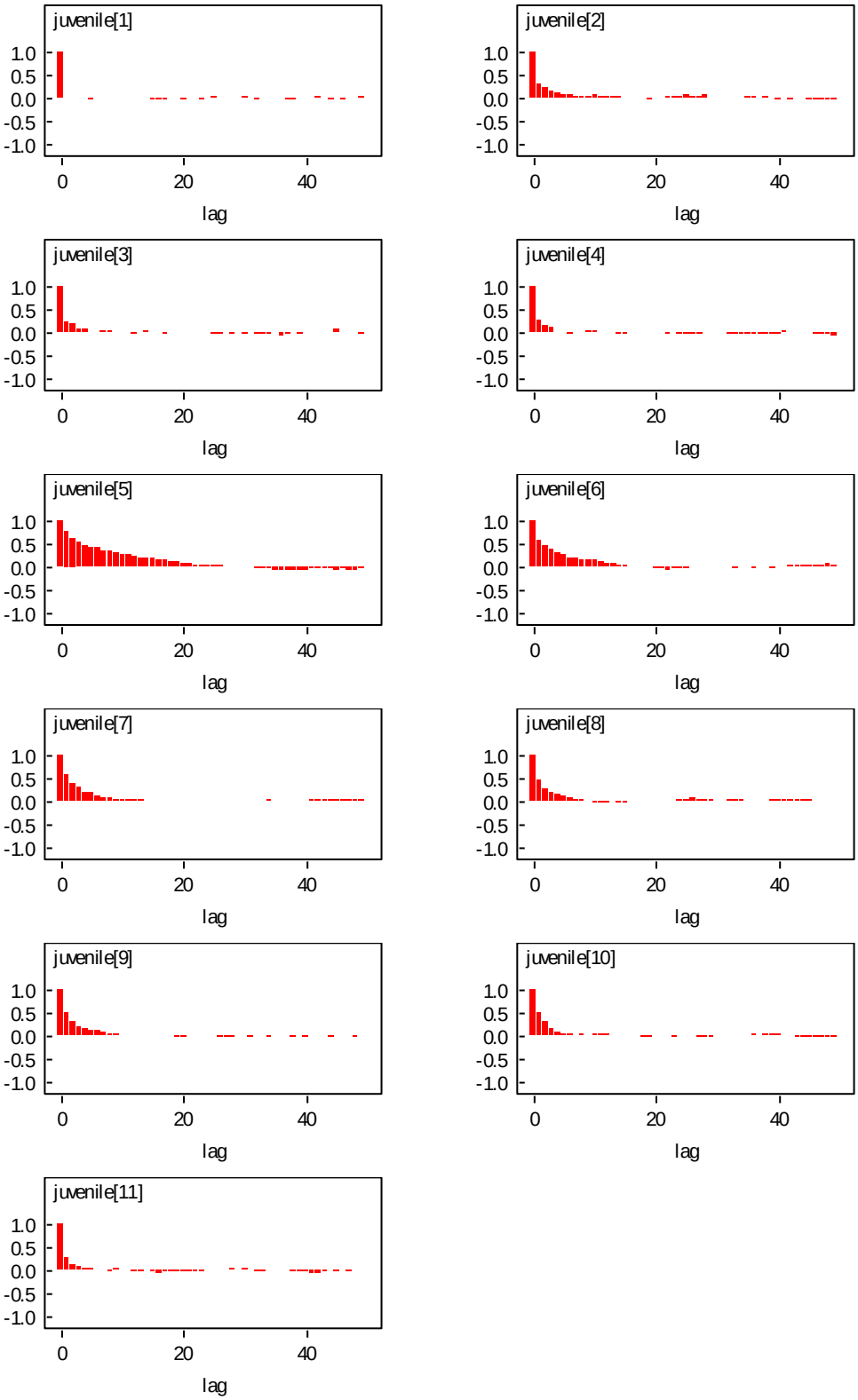


Appendix D: Autocorrelation Functions

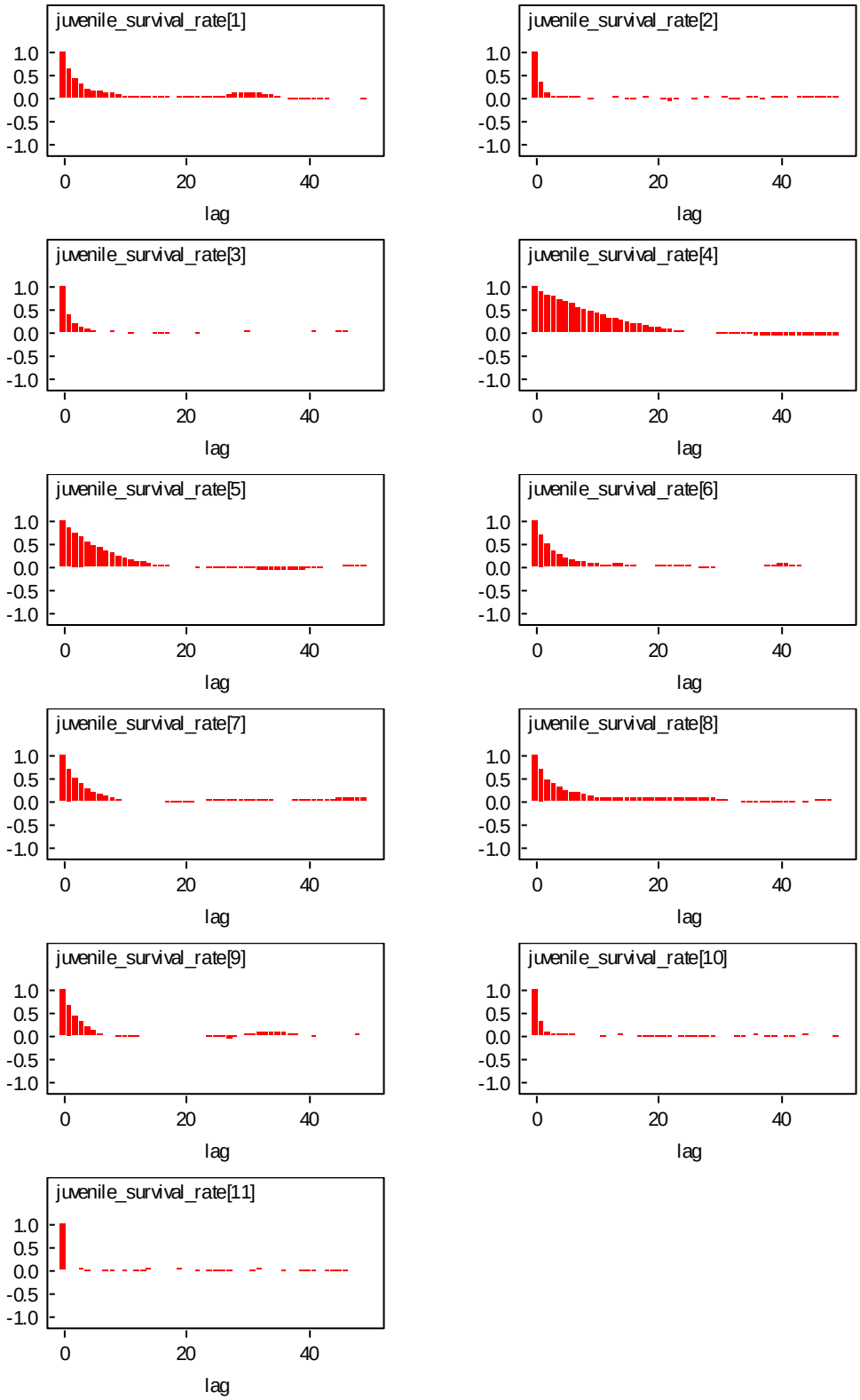
Autocorrelation Functions: Total Count



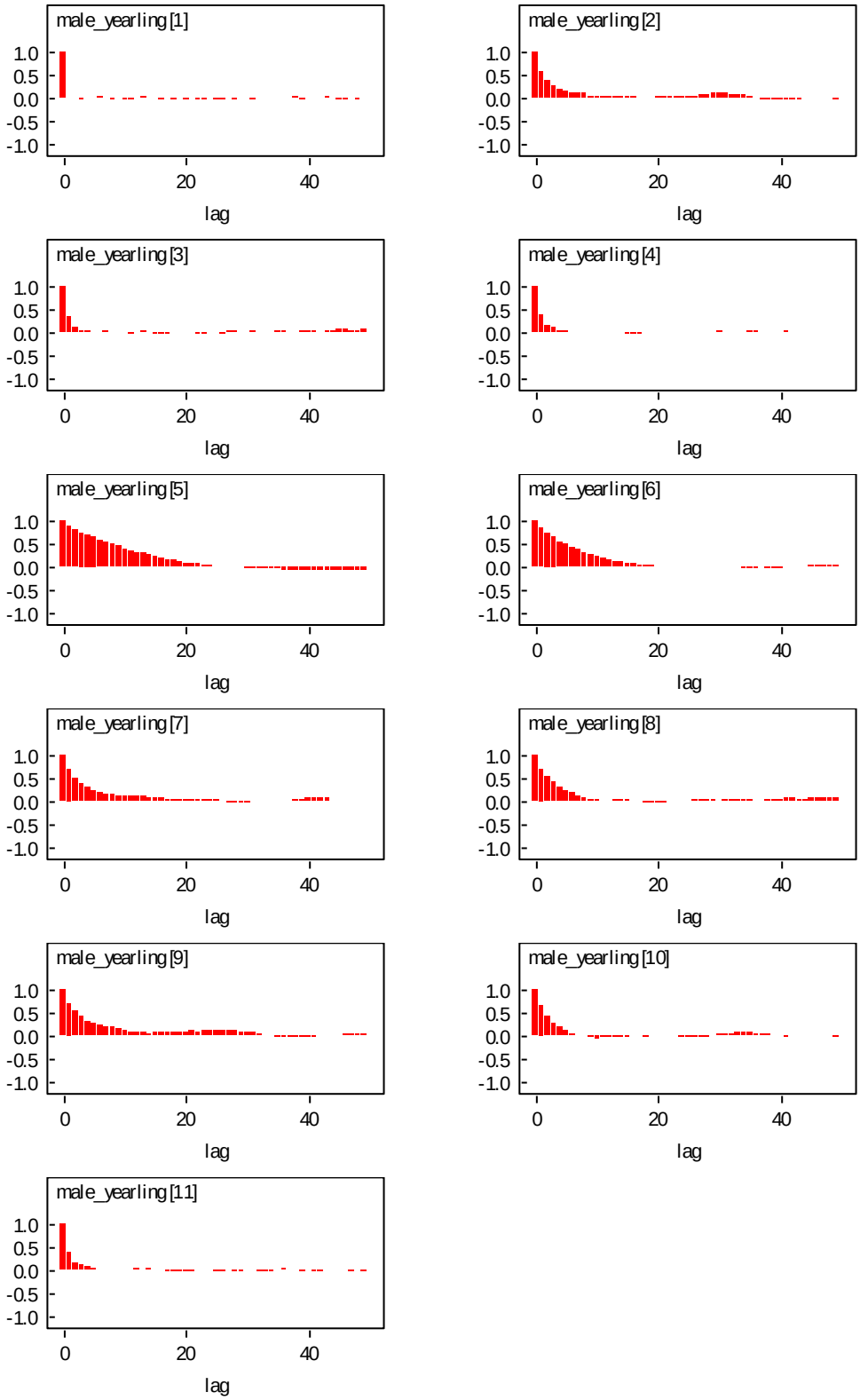
Autocorrelation Functions: Estimated Number of Juveniles



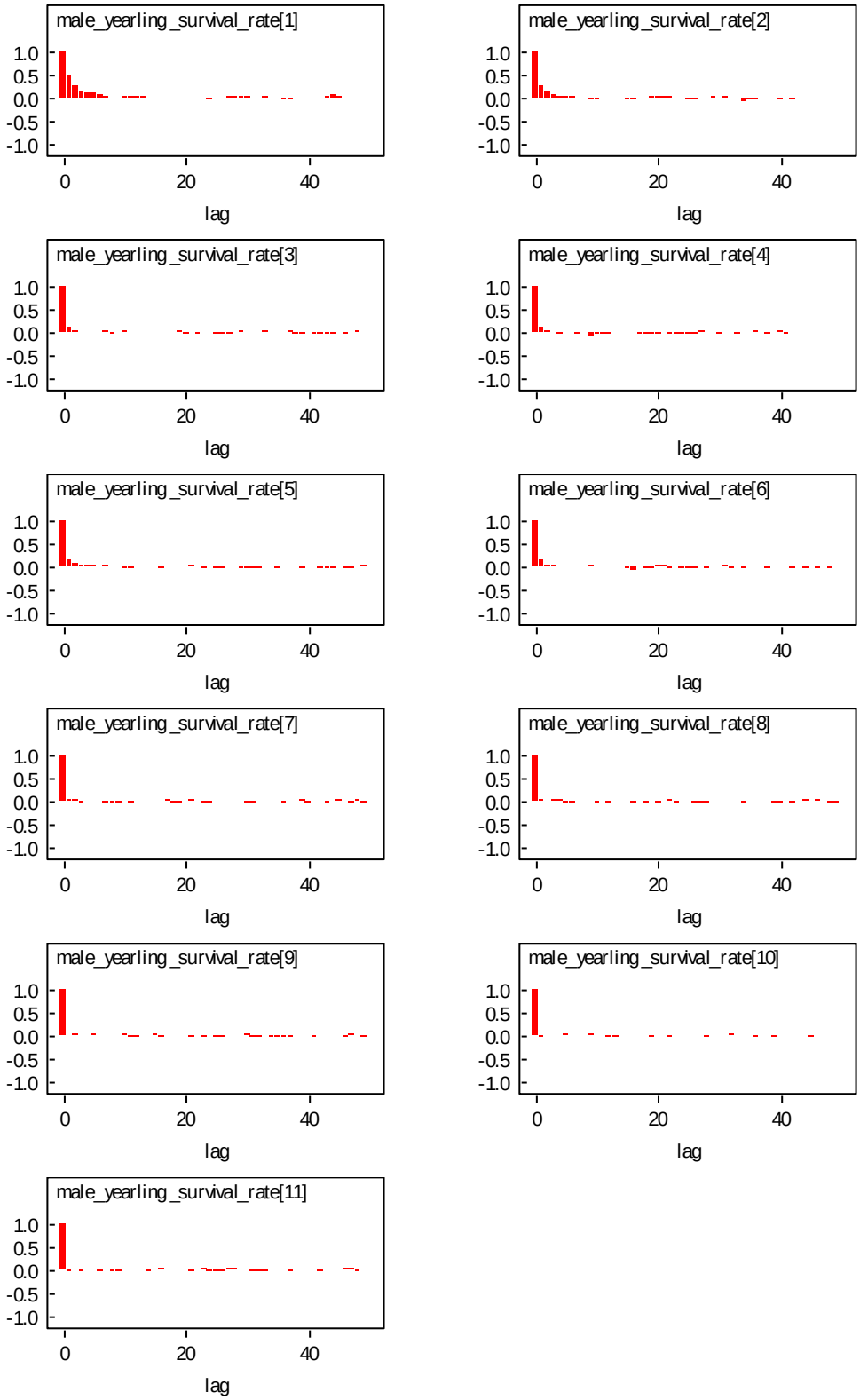
Autocorrelation Function: Juvenile Survival Rate



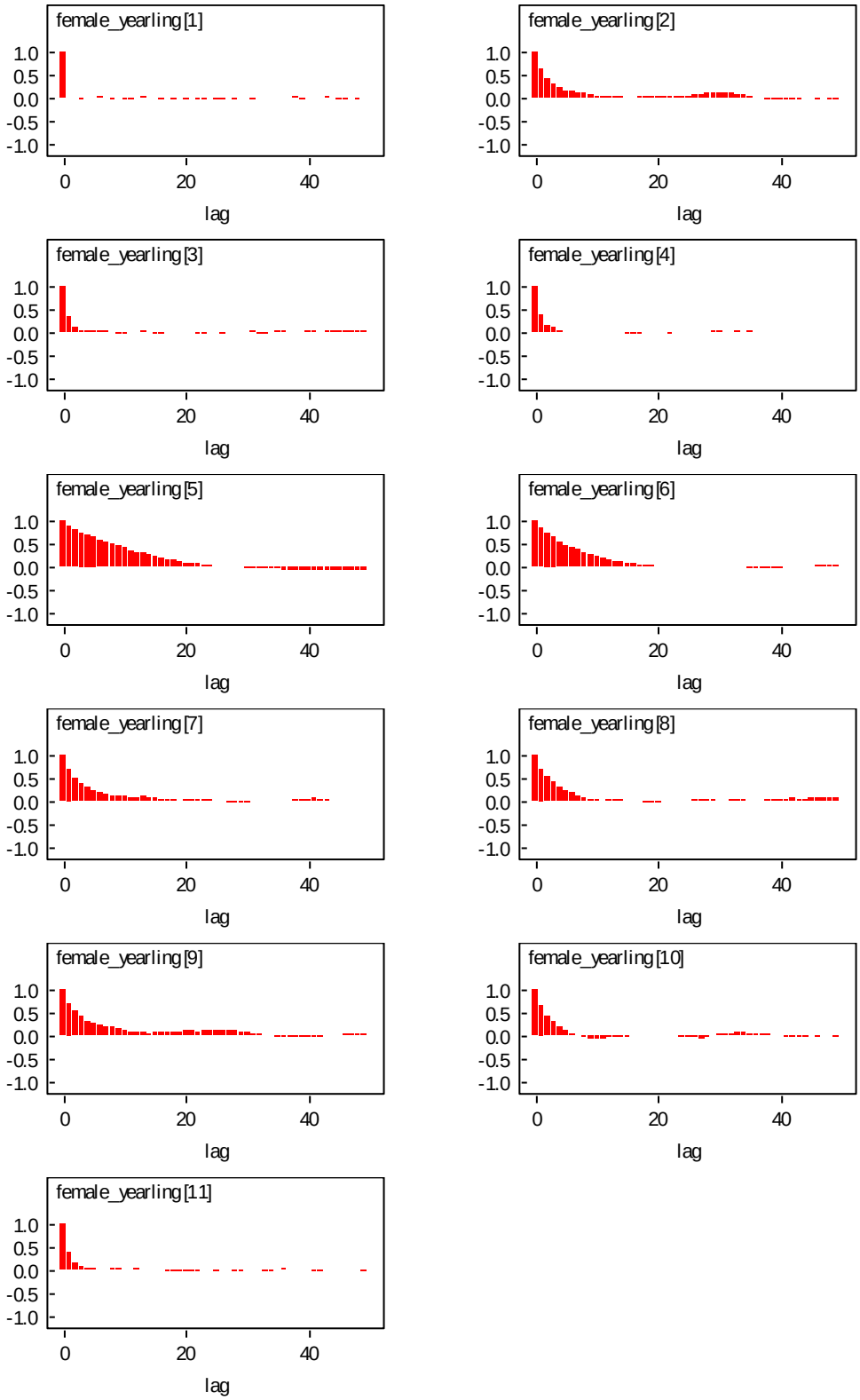
Autocorrelation Function: Number of Male Yearlings



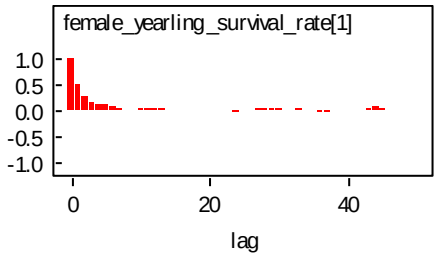
Autocorrelation Functions: Male Yearling Survival Rate

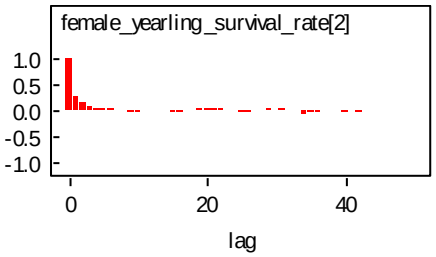


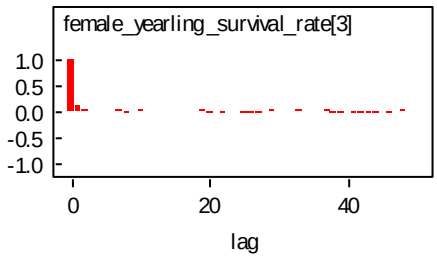
Autocorrelation Functions: Number of Female Yearlings

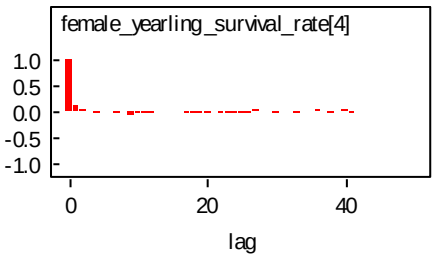


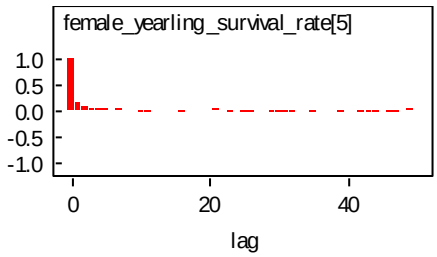
Autocorrelation Functions: Female Yearling Survival Rate

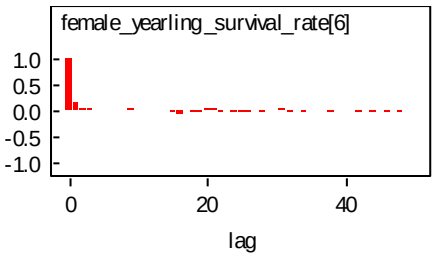


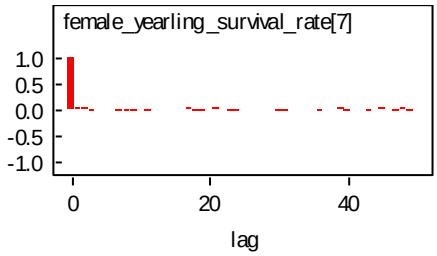


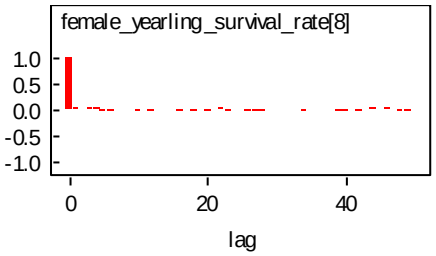


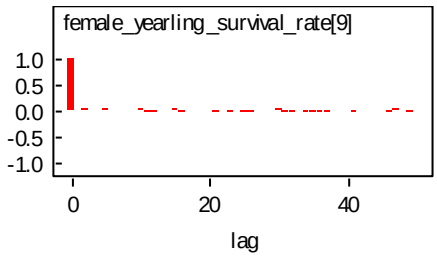


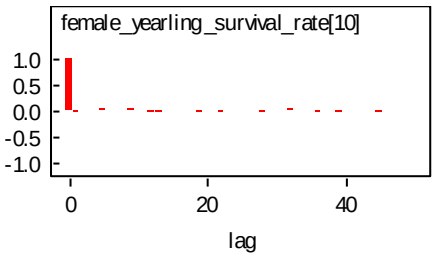


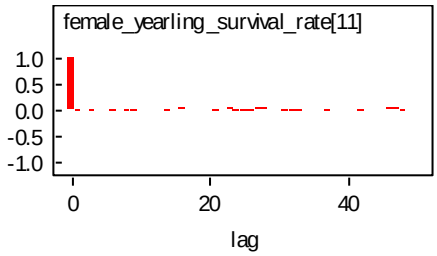




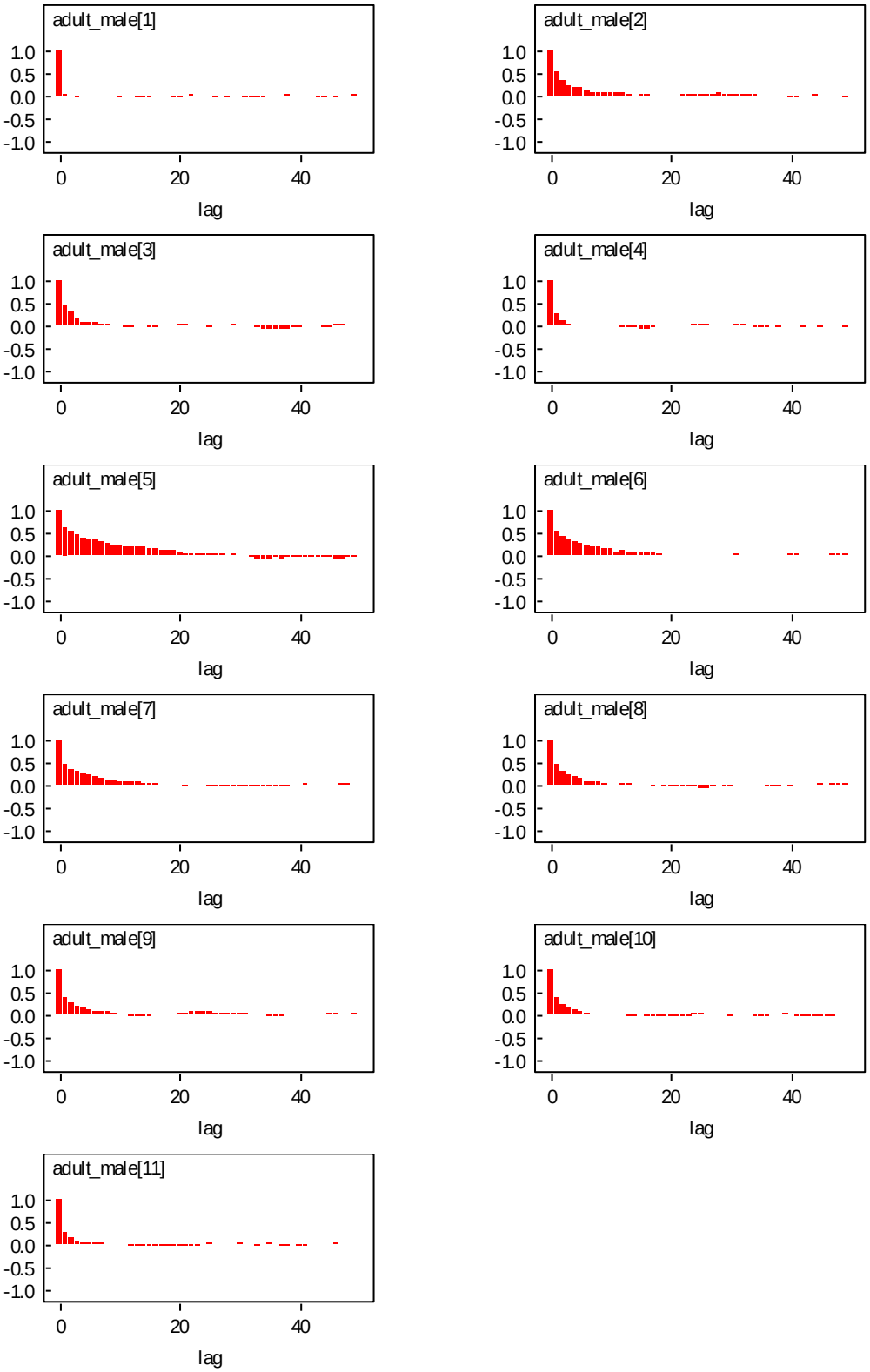




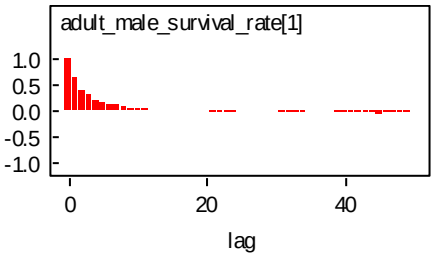


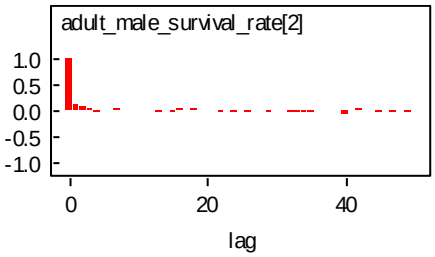


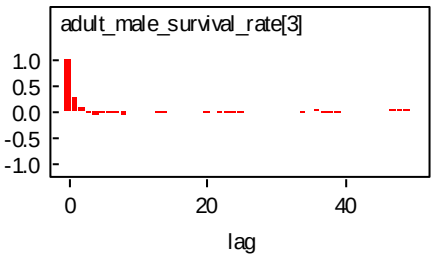
Autocorrelation Functions: Number of Male Adults

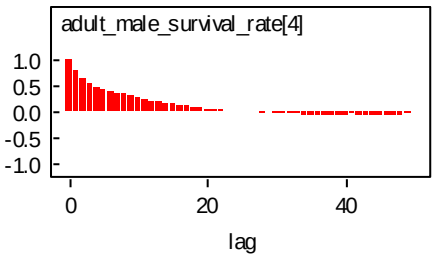


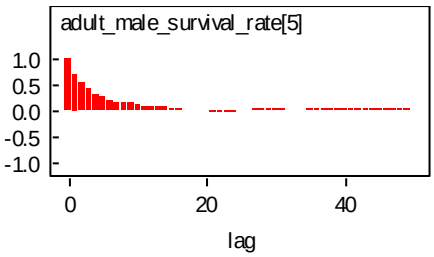
Autocorrelation Functions: Adult Male Survival Rate

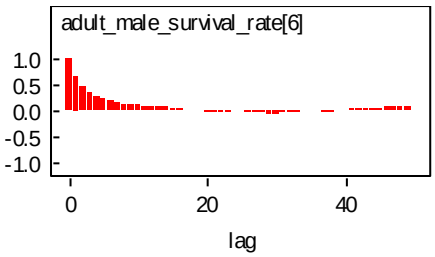


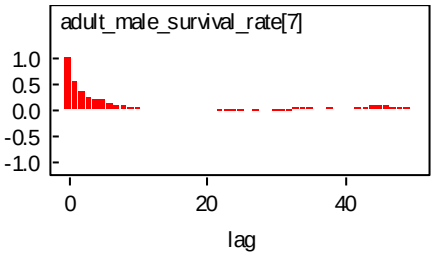


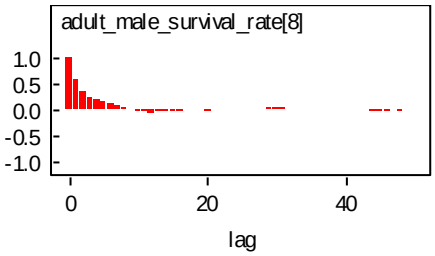


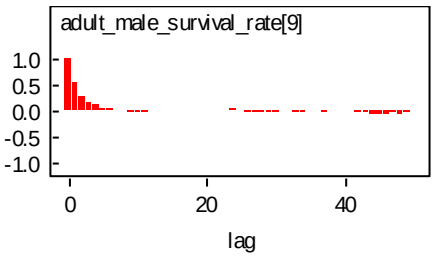


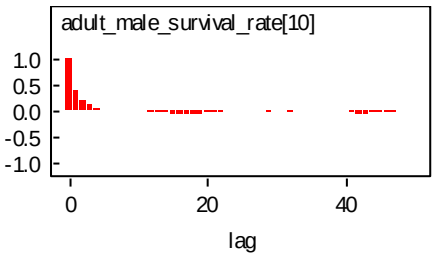


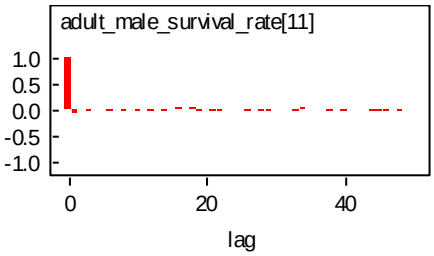




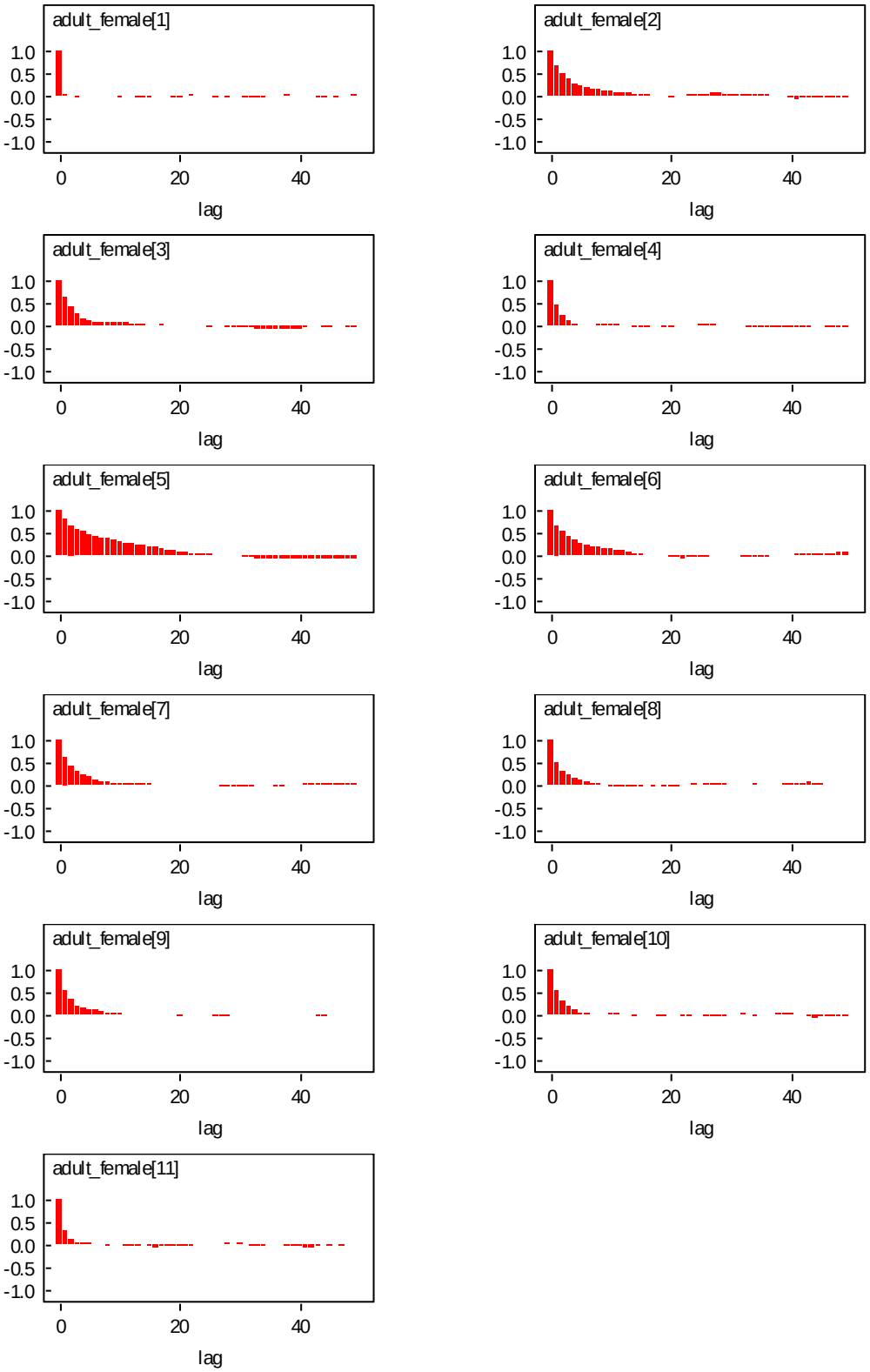




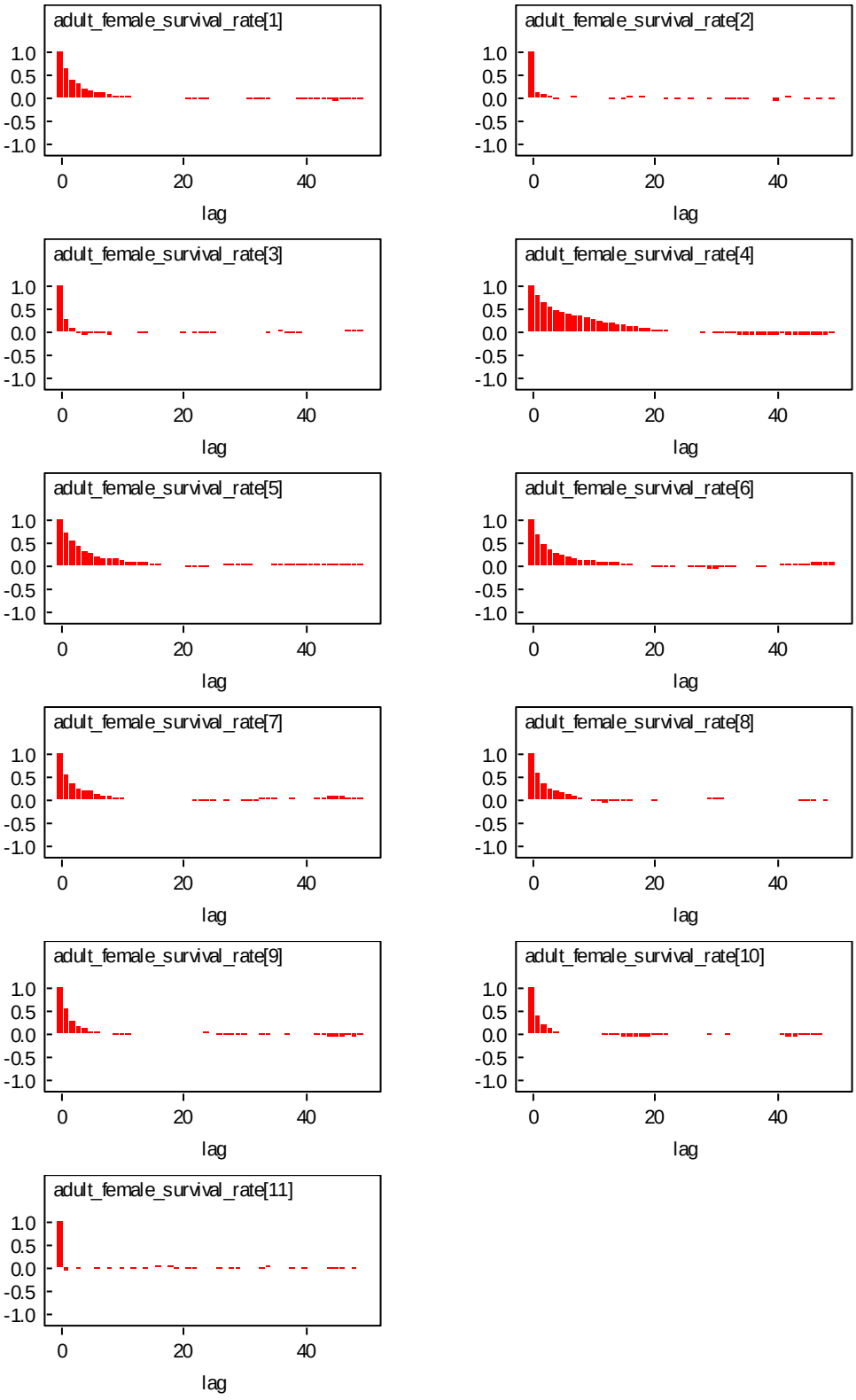




Autocorrelation Functions: Number of Adult Females



Autocorrelation Functions: Adult Female Survival Rate



Appendix E: Summary Statistics (Final Model – Model 2)

Summary Statistics for the Estimated Total Count of Kudu in the KNP

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5%
1983	9128	45.48	0.9923	9042	9127	9220
1984	9193	213.5	8.526	8749	9199	9597
1985	9702	235.1	7.178	9249	9704	10150
1986	9995	290.4	9.228	9420	9997	10570
1987	8844	384.7	13.87	8118	8830	9625
1988	8129	375	13.65	7406	8128	8869
1989	7030	371.8	14.33	6312	7036	7770
1990	5867	357.9	9.704	5187	5860	6605
1991	4710	347.8	10.81	4082	4701	5395
1992	4015	349.4	10.48	3325	4027	4712
1993	2856	366.7	9.699	2138	2859	3553

Summary Statistics for the Estimated Number of Juveniles

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	1436	37.13	0.9012	1364	1436	1507
1984	1558	51.12	2.167	1454	1561	1657
1985	1676	53.16	1.707	1570	1676	1781
1986	1745	60.55	1.867	1628	1746	1862
1987	1531	123.6	8.995	1309	1520	1778
1988	1391	101.5	5.862	1210	1385	1600
1989	1260	98.46	4.896	1067	1259	1455
1990	1011	91.62	3.535	834.7	1009	1198
1991	813.3	83.08	3.607	655.7	812.1	978.6
1992	704.1	81.69	3.189	546.8	701.3	864.6
1993	477.6	82.64	2.479	315.9	478.4	637.7

Summary Statistics Estimated Juvenile Survival Rate

Year	**Mean**	**Standard Deviation**	**MC error**	**Lower 2.5% CI**	**median**	**Upper 97.5% CI**
1983	0.8527	0.1245	0.006227	0.5354	0.8845	0.9964
1984	0.7882	0.1061	0.003796	0.5313	0.8049	0.943
1985	0.7414	0.1205	0.004753	0.475	0.7581	0.9246
1986	0.6995	0.2745	0.0228	0.1322	0.7831	0.9991
1987	0.8167	0.2092	0.01512	0.2673	0.9007	0.9997
1988	0.58	0.2012	0.01081	0.1853	0.5927	0.9317
1989	0.7274	0.1951	0.00951	0.2477	0.7718	0.9824
1990	0.7044	0.2202	0.01373	0.1879	0.7456	0.9893
1991	0.6813	0.23	0.01005	0.1759	0.718	0.9892
1992	0.7167	0.1712	0.005444	0.3215	0.7438	0.9642
1993	0.9568	0.1443	0.003525	0.3823	1	1

Summary Statistics Number of Male Yearlings

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	837	14.15	0.3119	816.9	835.2	864.1
1984	611.6	91.92	4.476	385	634	737
1985	613.8	85.14	2.788	411	624	749
1986	622.2	102.3	3.902	395	636	783
1987	609.9	240.5	19.71	116	681	903
1988	624.7	169.3	12.7	208	667	854
1989	403.3	144.8	8.343	126	413	672
1990	458.2	128.2	6.434	166	474	656
1991	356.5	117.3	7.409	95	373	543
1992	275.9	95.33	3.763	71	287	432
1993	251.9	67.48	2.324	113	255	373

Summary Statistics Male Yearling Survival Rate

Year	Mean	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	0.4073	0.0768	0.003029	0.2559	0.405	0.5573
1984	0.5065	0.09079	0.003198	0.3179	0.5112	0.6704
1985	0.4085	0.09942	0.002886	0.2161	0.4076	0.6022
1986	0.3217	0.08944	0.001781	0.1609	0.3152	0.5071
1987	0.3339	0.09648	0.002708	0.1542	0.3294	0.5282
1988	0.3566	0.1025	0.002671	0.1687	0.3548	0.5643
1989	0.3398	0.09792	0.002499	0.1588	0.3396	0.5366
1990	0.355	0.0999	0.002183	0.1665	0.3517	0.5576
1991	0.3681	0.1018	0.002627	0.1752	0.369	0.565
1992	0.3171	0.07885	0.00183	0.1747	0.3159	0.48
1993	0.349	0.1009	0.001958	0.1676	0.3477	0.5565

Summary Statistics Number of Female Yearlings

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	837	14.15	0.3119	816.9	835.2	864.1
1984	613	92.87	4.686	380	634	739
1985	615.5	85.35	3	415	626	749
1986	621.9	104.6	4.035	399	632	793
1987	609.7	241.1	19.94	116	678	903
1988	624.8	169.3	12.58	196	670	856
1989	403.3	144.7	8.183	124	408	665
1990	457.4	127.5	6.417	165	474	659
1991	356.1	116.6	7.468	98	373	544
1992	275.8	95.55	3.725	72	288	434
1993	252.4	67.85	2.333	111	256	379

Summary Statistics Female Yearling Survival Rate

Year	Mean	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	0.4655	0.08777	0.003462	0.2925	0.4628	0.6369
1984	0.5789	0.1038	0.003655	0.3633	0.5843	0.7662
1985	0.4669	0.1136	0.003298	0.247	0.4658	0.6882
1986	0.3677	0.1022	0.002036	0.1838	0.3602	0.5795
1987	0.3816	0.1103	0.003095	0.1762	0.3764	0.6037
1988	0.4076	0.1172	0.003053	0.1928	0.4055	0.6449
1989	0.3883	0.1119	0.002856	0.1815	0.3882	0.6132
1990	0.4057	0.1142	0.002495	0.1903	0.402	0.6373
1991	0.4207	0.1163	0.003003	0.2002	0.4217	0.6457
1992	0.3623	0.09011	0.002091	0.1996	0.361	0.5486
1993	0.3988	0.1153	0.002238	0.1916	0.3973	0.636

Summary Statistics Number of Adult Males

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	1744	0.04465	0.000961	1744	1744	1744
1984	1835	66.73	3	1706	1834	1968
1985	1905	80.38	2.943	1748	1907	2062
1986	1897	92.7	2.868	1717	1894	2085
1987	1589	127.9	8.806	1362	1581	1839
1988	1396	118	6.864	1173	1391	1633
1989	1254	111.1	5.421	1036	1253	1468
1990	971.3	102.6	3.942	781	969	1173
1991	794.2	90.98	3.954	626	792	983
1992	689.8	85.08	3.008	528	690	857
1993	469.5	80.67	2.675	312	470	630

Summary Statistics Adult Male Survival Rate

Year	Mean Survival	Standard	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	0.8562	0.01656	0.000828	0.8103	0.8607	1744
1984	0.8669	0.00917	0.000215	0.8427	0.8701	1968
1985	0.8623	0.01568	0.000406	0.8204	0.868	2062
1986	0.7326	0.06342	0.004448	0.6186	0.7282	2085
1987	0.7528	0.07099	0.004082	0.6177	0.7531	1839
1988	0.7415	0.07557	0.003975	0.5894	0.7415	1633
1989	0.6657	0.07548	0.003283	0.5183	0.6649	1468
1990	0.6535	0.08651	0.003666	0.495	0.6498	1173
1991	0.7069	0.09527	0.003581	0.5157	0.7115	983
1992	0.5565	0.1126	0.003652	0.3385	0.5579	857
1993	0.681	0.1822	0.003532	6.087E-20	0.0002	630

Summary Statistics Number of Adult Females

Year	Mean Count	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	4274	0.1094	0.002357	4274	4274	4274
1984	4574	95.52	5.175	4367	4580	4744
1985	4891	106.7	4.74	4673	4892	5099
1986	5109	138.2	4.75	4827	5115	5368
1987	4505	350.3	26.68	3885	4468	5208
1988	4093	289.5	16.99	3568	4076	4707
1989	3710	273.4	13.61	3176	3704	4230
1990	2969	260.3	9.935	2469	2966	3477
1991	2389	235.4	10.71	1951	2382	2855
1992	2070	232.9	9.287	1622	2075	2526
1993	1404	237.6	6.928	938	1407	1869

Summary Statistics Adult Female Survival Rate

Year	Mean	Standard Deviation	MC error	Lower 2.5% CI	median	Upper 97.5% CI
1983	0.9785	0.01892	0.000947	0.9261	0.9836	4274
1984	0.9908	0.01048	0.000246	0.963	0.9944	4744
1985	0.9854	0.01792	0.000464	0.9376	0.992	5099
1986	0.8373	0.07248	0.005083	0.707	0.8323	5368
1987	0.8603	0.08113	0.004665	0.706	0.8607	5208
1988	0.8474	0.08636	0.004542	0.6736	0.8474	4707
1989	0.7608	0.08626	0.003752	0.5923	0.7599	4230
1990	0.7469	0.09887	0.00419	0.5657	0.7427	3477
1991	0.8078	0.1089	0.004093	0.5894	0.8131	2855
1992	0.636	0.1287	0.004174	0.3868	0.6376	2526
1993	0.678	0.2083	0.004036	6.956E-20	0.0002	1869