

**HIERARCHICAL BAYESIAN MODELLING FOR THE  
ANALYSIS OF THE LACTATION OF DAIRY  
ANIMALS**

by

CAROLINA SUSANNA LOMBAARD (née VILJOEN)

THESIS

Submitted in fulfillment of the requirements for the degree

PHILOSOPHIAE DOCTOR

IN

THE FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES

DEPARTMENT OF MATHEMATICAL STATISTICS

UNIVERSITY OF THE FREE STATE

BLOEMFONTEIN

MARCH 2006

PROMOTER: PROF. DR. P.C.N. GROENEWALD

## **ACKNOWLEDGEMENTS**

I wish to express my most sincere thanks to my promoter, Professor P.C.N. Groenewald, for his advice, encouragement, numerous suggestions and, above all, his patience during the preparation of this thesis.

I also wish to acknowledge both the Animal Improvement Institute at the Agricultural Research Council in Pretoria and the South African Stud Book and Livestock Improvement Association in Bloemfontein for making available the data used for illustration purposes in this thesis.

A word of thanks also goes to Professor T. McDonald and Doctor I. Garisch, both of whom assisted in extracting data from the initial Jersey database.

I also wish to thank my friends and family for their encouragement and understanding during the years it took to complete this thesis, especially my parents – without their support this would never have been possible.

Special thanks go to my husband and daughter who had to put up with me through so much of this!

TO: My parents, husband and Susan

# CONTENTS

## DEFINING NOTATION

N.1	Mathematical	viii
N.2	Probabilistic	viii
N.3	Distributional	ix
N.4	Statistical	ix

## CHAPTER 1: INTRODUCTION

1.1	What is lactation	1
1.2	Milk composition	2
1.3	Important characteristics of lactation	3
1.4	Why is it important to model lactation	5
1.5	Factors influencing lactation	7

## CHAPTER 2: HISTORICAL DEVELOPMENT OF LACTATION CURVES

2.1	Where did it all start	10
2.2	Typical lactation curve shapes	23
2.3	Curves fitted to composition traits	25
2.4	Persistency	26

## CHAPTER 3: METHODS APPLIED IN FITTING LACTATION CURVES

3.1	Interpolation to obtain total milk production	29
3.2	IRIS	30
3.3	Linear and non-linear methods of curve fitting	32
3.4	Bayesian approach to lactation curve fitting	34
3.5	Assumption about errors	36

## **CHAPTER 4: A BAYESIAN APPROACH**

4.1	The generalised Bayesian model	41
4.2	The Gibbs sampler	51
4.3	The characteristics of the lactation curve	53
4.4	Handling incomplete data	54
4.5	The Metropolis-Hastings algorithm	58

## **CHAPTER 5: THE WOOD MODEL AND THE JERSEY DATA**

5.1	The Jersey data	62
5.2	Fitting the Wood model	64
5.3	The covariate matrix $Z_i$	67
5.4	The results of the Gibbs sampler	68
5.5	The lactation trait milk yield	73
5.5.1	Expected total milk yield	80
5.5.2	Expected peak milk yield	84
5.5.3	Expected time of peak milk yield	87
5.5.4	Expected persistency of lactation	90
5.5.5	Milk yield: In summary	93
5.6	The lactation composition traits	96
5.6.1	Fat content of milk	96
5.6.2	Protein content of milk	103
5.6.3	Milk composition traits fat and protein: In summary	108
5.7	Predictions based on the Wood model	109

## **CHAPTER 6: ALTERNATIVE MODELS FITTED TO JERSEY DATA**

6.1	Other lactation models	112
6.2	The general exponential model	113
6.2.1	The lactation trait milk yield	116
6.2.2	The lactation trait fat content of milk	120
6.2.3	The lactation trait protein content of milk	123
6.2.4	Predictions based on the general exponential model	126

6.3	The 4-parameter Morant model	127
6.3.1	The lactation trait milk yield	129
6.3.2	The lactation trait fat content of milk	132
6.3.3	The lactation trait protein content of milk	136
6.3.4	Predictions based on the 4-parameter Morant model	139
6.4	The adapted Wilmink model	140
6.4.1	The lactation trait milk yield	145
6.4.2	The lactation trait fat content of milk	149
6.4.3	The lactation trait protein content of milk	152
6.4.4	Predictions based on the adapted Wilmink model	155
6.5	Alternative lactation models in summary	156

## **CHAPTER 7: BAYES FACTORS AND MODEL COMPARISON**

7.1	Bayes factors	158
7.2	Marginal likelihood from Gibbs output	160
7.3	Bayes factors in application	161
7.4	Model comparison using Bayes factors	163

## **CHAPTER 8: FURTHER PREDICTION BASED ON MODELS FITTED**

8.1	Introduction	166
8.2	Prediction for further data: The general case	167
8.3	Prediction and the Wood model	170
8.4	Prediction and the general exponential model	176
8.5	Prediction and the 4-parameter Morant model	181
8.6	Prediction and the adapted Wilmink model	186
8.7	Prediction of further data in summary	190

## **CHAPTER 9: THE WOOD MODEL AND THE INCOMPLETE SAANEN DATA**

9.1	The Saanen data	194
9.2	The Wood model fitted to incomplete data	195

9.3	Setting the covariates contained in $\mathbf{Z}_i$	198
9.4	The results of the Gibbs sampler	200
9.5	The lactation trait milk yield	201
9.6	The Wood lactation curve characteristics	205
9.7	The lactation composition traits	212
9.8	Prediction of missing lactation records	218
9.9	Saanen data results: A summary	220

## **CHAPTER 10: EPILOGUE**

10.1	The results in summary	221
10.2	Contribution to lactation research in perspective	222

<b>REFERENCES</b>		225
-------------------	--	-----

## **APPENDIX**

A1	An example of MATLAB <sup>®</sup> code: Finding marginal posterior distributions through Gibbs sampling in the case of complete data	235
A2	An example of MATLAB <sup>®</sup> code: Determining the marginal likelihood required for calculation of Bayes factors	239
A3	An example of MATLAB <sup>®</sup> code: Finding marginal posterior distributions through Gibbs sampling in the case of incomplete data	248

<b>SUMMARY</b>		253
----------------	--	-----

<b>OPSOMMING</b>		255
------------------	--	-----

# DEFINING NOTATION

## N.1 Mathematical

$\text{diag}\{x\}$	the elements of the vector $x$ written on the main diagonal of the formed matrix and all other elements of the matrix taking on the value zero.
$\text{etr}[A]$	$\exp[\text{trace of the matrix } A]$
$f(t) \propto g(t)$	the functions of $f$ and $g$ are proportional
$\Gamma(x)$	gamma function ( $x > 0$ )
$\mathbf{h} = [h_1 \quad h_1 \quad \dots \quad h_n]$	boldface signifies a vector
$\mathbf{H} = [h_{ij}]$	uppercase signifies a matrix
$\mathbf{I}_h$	identity matrix of order ( $h \times h$ )
$\ln x$	natural logarithm of $x$ or $\log_e(x)$
$\text{tr}(A)$	trace of the matrix $A$
$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} \begin{matrix} (a) \\ (b) \end{matrix}$	$R$ is a partitioned matrix such that partition $\mathbf{R}_{11}$ is of order $a$ rows by $a$ columns, partition $\mathbf{R}_{12}$ is of order $a$ rows by $b$ columns, etc.
$\text{vec } A'$	the elements of matrix $A$ stacked row-wise into one column
$\mathbf{X} (n \times v)$	matrix $\mathbf{X}$ of order $n$ rows by $v$ columns

## N.2 Probabilistic

i.i.d.	independent and identically distributed
$f(x/?)$	density of variable $X$ , conditional on parameter $\mathbf{q}$
$Y \sim f(y/?)$	$Y$ is distributed with density $f(y/?)$



### N.3 Distributional

$IW_a(P, \mathbf{d})$	inverse Wishart distribution of order $a$ with parameter $P$ and $\mathbf{d}$ degrees of freedom
$N(\mathbf{q}, \Sigma)$	multivariate normal distribution
$N(\mathbf{m}, \mathbf{s}^2)$	univariate normal distribution
$TN(\mathbf{m}, \mathbf{s}^2)$	truncated univariate normal distribution
$W_a(P^{-1}, \mathbf{d})$	Wishart distribution of order $a$ with parameter $P^{-1}$ and $\mathbf{d}$ degrees of freedom

### N.4 Statistical

$B_{12}$	Bayes factor in favour of Model 1 relative to Model 2
$f(\mathbf{q}   x)$	generic posterior density of $\mathbf{q}$ given data $x$
$f(\mathbf{r}_{11}, \mathbf{r}_{22}, \mathbf{r}_{12}   \mathbf{M}, \mathbf{B})$	joint conditional distribution of $\mathbf{r}_{11}$ , $\mathbf{r}_{22}$ and $\mathbf{r}_{12}$
$H_1$	hypothesis under model 1
$H_2$	hypothesis under model 2
$p(\mathbf{q})$	generic prior density of $\mathbf{q}$

# CHAPTER 1

## INTRODUCTION

### 1.1 WHAT IS LACTATION

The term lactation may be defined as the secretion of milk manufactured from simple blood nutrients by the milk-synthesising cells of the mammary glands, together with the removal thereof from the mammary gland (Hurley, 2003). These special glands, also known as mammae, are found in the udder of the females of mammal species and have the same simple and relatively homogeneous basic structure for all mammal species.

In mammals the purpose of milk is to feed the newborn and therefore lactation usually begins at the end of pregnancy. The stimulation of the mammary gland to start producing milk around the time of birth or parturition is controlled by changes in the levels of certain hormones, some of which are involved in the action process of parturition. It is the removal of milk from the udder, usually but not necessarily by the newborn, that initiates lactation.

A fluid, known as colostrum, is secreted before true milk is produced. This first secretion from the udder occurs shortly after or sometimes even before parturition of the infant. Colostrum is a relatively clear fluid containing, amongst others, serum, white blood cells, and protective antibodies and is mainly responsible for immunity transfer during the first few hours of life, making it vital to the survival of the newborn. The composition of the colostrum secretion gradually changes over a period of 2 to 3 days after parturition, depending on the mammal species under consideration, to that of mature milk. Both colostrum and milk are secreted in response to the sucking action of the infant on the nipple or teat. This sucking action can also be simulated by artificial means, such as milking by hand and the milk machines used in the dairy industry (Mephram, 1976).

Lactation is controlled by hormones resulting in different amounts of milk secreted at various stages of lactation. It is generally accepted that as a result of the influence of hormones, together with the stimulus of milk removal, milk yield rises to a peak, where after for the rest of lactation milk yield is in continual decline. The daily milk yield, length of time until an animal reaches peak milk yield, as well as the duration of lactation differs from mammal

species to mammal species and sometimes even for different breeds within a certain species. Total milk yield is the result of the combination of the amount of daily milk produced and the duration of lactation. Secretion of milk stops when the infant is no longer allowed to suckle or when artificial stimulation of the mammary glands end (Whittemore, 1980).

## **1.2 MILK COMPOSITION**

Milk may be described as a white liquid designed for the nourishment of the infant, of which water constitutes over 80% by weight in most mammal species (Mephram, 1976). Cow milk, for example, contains approximately 87,4% water (Whittemore, 1980). The remaining constituents are solids in the form of lipids, carbohydrate and proteins, as well as various vitamins and minerals. Milk is secreted as a complex mixture of these components. The composition of milk, however, varies considerably between species and even within the same species, as well as during lactation – with the major changes usually occurring soon after the start of lactation.

Milk is synthesised in specialised secretory cells of the mammary glands from substances absorbed from the blood of the mother. Through a process known as biosynthesis lipids, commonly known as milk fat, are synthesised mainly from triglycerides that are derivatives of glycerol, but also from other fatty acids and glucose in the blood. Milk fat droplets form the cream of milk. Milk fat is the most variable component of milk and ranges from a little over 1% to greater than 50%, depending on the mammal species under consideration. Considerable variation in milk fat content may also occur within mammal species.

Lactose, present in the milk of most mammals, is unique to the mammary gland and plays an important role in milk synthesis. It also forms the major carbohydrate in milk. Lactose is defined as the sugar of milk synthesised from blood glucose of the mother. As a result of the close relationship between lactose synthesis and the amount of water drawn into milk through the process of osmosis, the lactose content of milk is the least variable component of milk.

Several types of proteins are found in milk, but the major milk proteins are unique to milk. The major milk proteins may be divided into two main groups, caseins and whey proteins. Whey proteins found in milk are mainly  $\beta$ -lactoglobulin,  $\alpha$ -lactalbumin, serum albumin and immunoglobulins, although long list of enzymes, hormones, growth factors, and other protein

components also form part of the whey protein content of milk. Milk proteins are synthesised from amino acids. The protein content of milk also varies considerably among species and sometimes even within species (from 1% to 14%), although not as much as milk fat. Hurley (2003) points out that it is generally accepted that protein percentage is positively correlated with milk fat percentage.

Because milk is the only source of food in mammal infants, the nutritional composition thereof is important with respect to skeletal and soft tissue development. Milk minerals provide these necessary components in the form of calcium and phosphorus, secreted into milk in a complex with caseins. Milk also contains most other minerals found in the body; these include sodium, iron and potassium, to name a few. In order to contribute to the general health of the infant, milk also contains all the major vitamins. Fat-soluble vitamins contained primarily in milk fat are vitamins A, D, E and K. The B vitamins are found in the aqueous phase of milk, as these are water-soluble. In addition to all of the above, milk also contains a number of other biochemical components, including bioactive factors such as growth hormones, enzymes and various others (Hurley, 2003).

### **1.3 IMPORTANT CHARACTERISTICS OF LACTATION**

The purpose of milk produced by mammals in nature is to feed the infants of these mammals, which are at parturition totally dependent on the mother and unable to find food by themselves. Therefore, milk production commences at a relatively high rate at parturition. As the infant grows, the amount of milk secreted continues to increase over a period of time in order to satisfy the needs of the growing infant until a peak production level is reached. Once this peak level is reached, the mother can no longer fulfil the growing nutrient requirements of the infant. The duration of this increase in milk production until peak level, as well as the level of peak yield, differ for different species of mammals. After peak production is attained, milk production gradually declines. This decline is generally associated with the infant becoming more independent from its mother, resulting in the development of the ability of the infant to feed by itself. Subsequently, weaning of the infant by its mother takes place (Lee et al., 1991). Hurley (2003) refers to the rate of decline in milk production as the persistency of milk production. In the dairy industry, where the infant is removed from the mother a few days after parturition, machine milking simulates the same effect.

The variation in the milk production during lactation produces what is termed a lactation curve. According to Ruvuna et al. (1995) lactation curves represent the relationship between milk yield and time after parturition. The shape of the standard lactation curve may be described as increasing, at a relatively high rate, up to the point where peak production is obtained, after which it declines at a slower rate until the end of the milk production cycle. Standard curves of this form are often referred to as type I curves. Variations from this standard pattern can however sometimes occur. In some red deer, for example, a continuously decreasing curve has also been found. Lactation curves of this nature are generally referred to as so-called type II curves (Landete-Casitllejos and Gallego, 2000), but may sometimes also be referred to as atypical curves (Ferris et al., 1985). These atypical curves are commonly found in cases where no lactation records prior to peak are observed.

Since the first research on lactation has taken place, a variety of functions have been used to model lactation. The majority of these functions, however, have the two important characteristics in common. Milk yield as a function of time is firstly peaked and secondly skewed to the right, to represent a lactation curve that is desirable with respect to the biological progression of the process (Tozer and Huffaker, 1999).

Whittemore (1980) noted that at the start of lactation, the first milk or colostrum contains twice the normal concentration of solids, five times the protein, approximately twice the fat and half the lactose. Once this composition has settled down and true milk is produced, a certain pattern in both yield and composition becomes apparent. Fat and protein content usually vary inversely to yield, while lactose in most cases goes into steady decline over the whole lactation. There is little day-to-day variation in protein and lactose content of milk and any changes that occur are gradual. The fat content of milk, however, does vary considerably from day to day.

To make provision for the production of colostrum, almost all studies of lactation consider changes in milk production and composition only from the point in time that true milk is produced. This means that the study of lactation in mammals only commences 2 to 3 days after parturition, depending on the mammal species under consideration.

According to Hurley (2003), fat is the most variable constituent of milk, while lactose is the least variable, but differences among individuals within a breed are often greater than differences among breeds. Although it is generally accepted that production of fat and protein

are correlated, the production of these composition constituents seem to be inversely related to lactation yield (Groenewald and Viljoen, 2003).

## **1.4 WHY IS IT IMPORTANT TO MODEL LACTATION**

Describing lactation in mammals using a lactation curve aims at providing a concise summary of the pattern of milk yield. The shape of the resulting lactation curve provides valuable information about the biological and economic efficiency (Grossman and Koops, 1988) of the animal or herd under consideration.

Milk is extracted from dairy animals for the purpose of feeding people. Around the world, mainly cow milk is used for this purpose, although a small volume of goat and sheep milk is also consumed. From a management point of view, knowledge of the lactation curve of lactating dairy animals is required for feeding, breeding and economic management of a dairy herd. Lactation curves are especially important when making decisions that are time-dependent. Knowing when to expect an animal to reach peak yield, would affect the feeding strategy followed, allowing economic management of feed to the extent that would satisfy the animal's requirement during various stages of lactation, reduce cost, and possibly maintaining peak yield for as long as possible (Tozer and Huffaker, 1999).

Lactation curves also allow for the identification of animals with a relatively constant yield throughout lactation, as well as animals with a high peak yield, but sudden decline thereafter. Information provided by lactation curves could also assist management, where decisions concerning aspects such as culling and milking strategies are concerned. It may for instance not be worthwhile to carry on milking an animal for an extended period of time, if it yields most of its milk early in lactation and then shows a sudden decline with respect to yield thereafter (Sakul and Boylan, 1992).

Lactation models may also be used in prediction of future milk yields of an individual animal or a herd. The objective when using a lactation curve in prediction, is to predict yield on each day of lactation with minimum error in the presence of variation as a result of environmental and other factors, in order to determine the underlying pattern of milk yield. The extent of the usefulness of a lactation model depends on how well it succeeds in imitating the biological lactation process and how well it adjust for environmental and other factors that could

influence production (Olori et al., 1999). Management decisions may also be made based on yield early in lactation, together with prediction of remaining yield. For example, in identifying sick animals before the manifestation of clinical signs and in identifying animals with special dietary needs (Gipson and Grossman, 1989). Prediction information may also be of value in deciding on culling or keeping breeding stock (Sherchand et al., 1995).

In general, two possible payment schemes are applicable to the dairy industry. The producer may be paid for his milk based on quantity alone, or paid according to quantity once it has been adjusted for quality. Depending on the ultimate use of the milk as intended by the purchaser, premium payments for milk could be related to the content of the milk with respect to milk-fat, protein or total milk solids and therefore it is often important to also consider curves fitted to the lactation traits, fat and protein, as well as curves fitted to milk yield when studying lactation. Selection of contributors to a dairy herd could therefore be based on all of milk, protein and fat yield.

In some countries milk quotas have been introduced, resulting in an increase in yield above the specified quota not being desirable. For this reason it might be more beneficial to include animals in a dairy herd that peak at a lower, but more sustainable yield level, i.e. animals that produce milk at a greater level of persistency (Ferris et al., 1985). Animals with a high level of peak yield followed by a sharp decline in production thereafter would be undesirable. Tekerli et al. (2000) points out that cows with flatter lactation curves, seem to be less prone to incidences of metabolic and reproductive disorders, which often occurs as a result of the physiological stress of high levels of yield. Lactation that follows a flatter curve may, however, result in a slight reduction in total milk yield (Varona et al., 1998).

One should not only be focusing on the dairy industry when considering lactation curves. For example, milk yield of the dam is the single most influential factor in the weaning weight of a beef calf and for this reason it is important when managing beef cattle to understand the shape of the lactation curve in a beef cow. The pattern of milk production would impact on the feeding and weaning strategies followed, so that economically beneficial decisions may be taken (Kim et al., 1998).

Whittemore (1980) warns that the use of lactation curves in both research and farm management should be approached with caution. The idea is not that a herd should follow the predetermined curve, but that once such a curve has been set up for a herd, it should act as a

reference point from which deviations may be observed and causes for such deviations be investigated.

By fitting a parametric curve to the pattern of milk yield and to the yields of the traits of milk production, statistical analysis of the parameter estimates are also made possible.

## **1.5 FACTORS INFLUENCING LACTATION**

As early as 1969, Wood already commented on the fact that the shape of the lactation curve in cows is affected by not only biological factors such as age and fertility of the cow, but also by various environmental factors. He specifically noted that season of parturition probably has the most drastic effect on the shape of the lactation curve.

According to Whittemore (1980), environmental and seasonal changes all bring about compositional changes in milk. The so-called comfort zone for most breeds of dairy cattle for instance, is between 5°C and 25°C, with temperatures below or above this generally being responsible for a reduction in yield. With respect to composition traits, low temperatures may increase the fat content of milk, while high temperatures are usually associated with a decline in milk fat. Hurley (2003) points out that at elevated temperatures the reason for this is that milk production and feed consumption are reduced automatically in an effort to counter the production of heat associated with these metabolic processes. Reduced milk yields are the result of depressed appetite. Heat stress is especially harmful to peak milk production.

Season of parturition is also expected to have a significant effect on total milk production. For cows, milk yields over the entire lactation seems to be higher when parturition takes place in autumn and decreases progressively when parturition occurs in winter, spring or summer. The reason for this is probably related to both temperature and the quality and availability of digestible feeds. Ferris, Mao and Anderson (1985) reported that season of parturition affected initial yield, peak yield, rise to peak and decline thereafter, and time of peak yield in dairy cows. Tekerli et al. (2000) specifically found that peak yield in dairy cows is higher when parturition takes place in autumn or winter. In the case of dairy goats, Ruvuna et al. (1995) noted that the greatest yield was obtained from does kidding in the hot dry season and the lowest yield from does kidding in the cold dry season. Gipson and Grossman (1990) confirm that season of kidding in dairy goats affects both initial and peak yield.



In cows, milk yields increase, be it at a decreasing rate, until about 8 years of age and then decrease at an increasing rate (Hurley, 2003). The production of mature cows is about 25% more than that of 2-year-old heifers, with approximately one fifth of this increase attributed to body weight while the remaining 80% results from increased udder development as a result of recurring parturitions. Although a large cow generally produces more milk than a small cow, the relationship between body weight and milk production is not directly proportional. Freeze and Richards (1992) also confirmed the effect of age on the lactation curve of Holstein dairy cows, but in their study maximum yield was attained at an age of roughly 6½ years. With respect to the composition traits, their study showed that the fat content increased with age, but that protein content starts declining as soon as the total yield starts declining, that is after about 6½ years. Mostert, Theron and Kanfer (2001) found that during the first two parities, younger dairy cows have lower total milk yield. Franci et al. (1999) reported that in the case of Massese sheep ewes, only total milk yield was affected by the age of an ewe.

Batra (1986), however, found that the effect of age of dairy cows on the lactation curve was not significant. Factors that he found to have a significant effect on the lactation curve were the station at which the herd is located, both year and month of parturition, and the number of days since the end of previous lactation. Jamrozik and Schaeffer (1997) mention that test day yields for Holstein dairy cows are affected by factors such as breed, region, how the herd is managed, day of the year (including weather conditions), parity, age at calving, month of calving, days in milk, pregnancy status, medical treatments and number of milking times per day.

Other studies considered parity number, which is the number of a particular parturition when considered in sequential order, rather than the age of the mother as having a significant effect on the lactation curve. Rowlands et al. (1982) notes that peak yield in dairy cows occurs later during first parity than is the case for second parity. Portolano et al. (1996) in their study of the lactation of Comisana sheep found a positive correlation between parity and peak yield, while parity and time of peak yield were negatively correlated.

Gipson and Grossman (1989) reported that in dairy goats, time of peak yield was later for first than for third parity does and also that initial yield, peak yield and total yield were lower in first parity does than in third parity does. They also mention that breed has little effect on the shape of the lactation curve. In 1990, however, Gipson and Grossman reported that the breed of dairy goat did affect both the level and time of peak yield, but their finding on parity

remained that initial and peak yield were lower for first than for later parity does and that time of peak yield was later for first parity does. Apka et al. (2001) found that in Red Sokoto goat herds, season of parturition and parity affected the shape of the lactation curve, with highest yield also occurring in the third parity. Groenewald and Viljoen (2003) found that in dairy goats, peak yield increased with increasing parity up until about the third or fourth parity and time of peak yield was later for first than for later parity does. Total milk yield was affected by the time during the season that kidding occurred, with higher yields occurring in does kidding earlier in the season, and that year of kidding significantly affected both total yield and peak yield. These results corresponded with that of others studying lactation in dairy goats (Mavrogenis, Constantinou, and Louca, 1984; Kala and Prakash, 1990; Rabasco et al., 1993; Kominakis et al., 2000).

Wood (1970) noted that differences in management of herds, including the intervals between milk extraction sessions, did not seem to really affect the shape of the lactation curve. He also noted that parity and season of calving were the two factors with the greatest influence on the lactation curve and as a result, inclusion of these two factors in a model would lead to more accurate prediction.

Tozer and Huffaker (1999) pointed out that almost all research to that point in time had been carried out on lactation records of animals that roam in the northern hemisphere, where environmental conditions and management practices are very different from that which occur in the southern hemisphere. They found that in the case of Australian Holstein-Friesian dairy cows, the resulting lactation curve shapes and yield characteristics differ from the results obtained from studies of dairy cows in Europe, North America and the United Kingdom. Our study of lactation was carried out on data acquired under South African conditions and will therefore make a valuable contribution to the knowledge of lactation curves under southern hemisphere conditions.

# CHAPTER 2

## HISTORICAL DEVELOPMENT OF LACTATION CURVES

### 2.1 WHERE DID IT ALL START

In all the models described it is assumed that  $y_t$  denotes daily milk yield,  $t$  denotes time in days after parturition, and  $a, b, c, d, m, k$  and  $w$  denote model parameters.

The first attempt at the development of a mathematical model to describe the lactation curve was as early as 1923. Brody, Ragsdale and Turner (1923) used an exponential decline function of the following form for this purpose:

$$y_t = a \exp(-ct). \quad (2.1)$$

Although this model<sup>1</sup> resulted in a good attempt to describe the declining phase of lactation, it was unable to model the initial rise in production to peak yield. To overcome this limitation, Brody, Turner and Ragsdale presented an improved version of their model in 1924. This time the model made provision for the initial rise to peak production by incorporating an inclining function into the model:

$$y_t = a \exp(-bt) - a \exp(-ct). \quad (2.2)$$

This model meant that increase in yield to peak production took place at a rate of  $\ln(\frac{c}{b})/(c-b)$ . Although this was a great improvement on their first model, later researchers such as Cobby and Le Du (1978) found on fitting this model to lactation data of cows, that it resulted in underestimation of milk yield in mid-lactation and overestimated milk yield in late lactation.

This was followed by a parabolic exponential function introduced by Sikka (1950) to model milk yield. This model:

$$y_t = a \exp(bt - ct^2) \quad (2.3)$$

---

<sup>1</sup> Note that Wood attributes this model to Gaines, but work by Gaines in this field was only published in 1927, whereas Brody et al. already published their paper in 1923.

resulted in a bell shaped truncated curve that, as a result of the curve symmetry around peak yield, only fitted milk yield reasonably during first lactation.

In 1958 Fischer attempted to improve on the Brody, Ragsdale and Turner model in (2.2), by substituting the exponential decline built into this model with a linear decline:

$$y_t = a - bt - a \exp(-ct). \quad (2.4)$$

The result was a model that tends to the straight line  $a - bt$  after peak yield has been obtained. Peak yield for this model occurs at  $c^{-1} \ln(ac/b)$  and the ratio  $a/b$  estimates the duration of lactation. This model underestimated peak milk yield and also peaked relatively early (Rowlands, Lucey and Russell, 1982).

Vujicic and Bacic (1961) attempted a modification of the model in (2.1):

$$y_t = tc^{-a} \exp(-ct). \quad (2.5)$$

This model seems to be the first attempt at developing a model that varies both directly and exponentially with time.

In an effort to improve on all models that existed at the time, Nelder (1966) suggested an inverse polynomial model be fitted to lactation data:

$$y_t = t / (a + bt + ct^2). \quad (2.6)$$

For this model peak milk yield of  $(2/(ac)+b)^{-1}$  occurs at time  $\sqrt{(a/c)}$ . The result was a model with a good fit when lactation started at a relatively low initial yield and peaked relatively early.

This was followed by what has been described as one of the major advances in modelling lactation - the model suggested by Wood (1967). Wood proposed a gamma function of the following form be used:

$$y_t = at^b \exp(-ct). \quad (2.7)$$

In this model, the parameter  $a$  approximates the level at which production of milk commences at parturition. According to Shanks et al. (1981) the parameter  $b$  is an index of the ability of a cow to make effective use of energy in producing milk, but mathematically according to Wood (1972) the parameter  $b$  represents the rate at which the rise to peak yield takes place and the parameter  $c$  in turn represents the rate of decline after peak yield was attained. Cobby and Le Du (1978) states that these interpretations of the parameters  $b$  and  $c$  "is a considerable over-simplification and could be misleading". In 1977 Wood tried to justify the use of his

model from a physiological point of view, but still does not interpret the parameters  $b$  and  $c$  from a practical biological point of view. From this model Wood (1967) also defined the following lactation curve characteristics:

$$\text{Total milk yield:} \quad y = \frac{a}{c^{b+1}} \Gamma(b+1) \quad (2.8)$$

$$\text{Yield to time } n: \quad y_n = a \int_0^n n^2 \exp(-cn) dn \quad (2.9)$$

$$\text{Time of peak milk yield:} \quad t = b/c \quad (2.10)$$

$$\text{Peak milk yield:} \quad y_{\max} = a(b/c)^b \exp(-b) \quad (2.11)$$

Wood also realised the need for a measure that could describe the ability of an animal to maintain peak production or the so-called measure of persistency of lactation. As a first attempt Wood noted that, because total yield or  $y$  is a function of  $c^{-(b+1)}$ , this may be used as a measure of persistency. It was reasoned that variation in total yield was almost entirely as a result of variation in  $a$  and in  $c^{-(b+1)}$ , and because  $a$  describes the level at which lactation commences then, for all lactations commencing at the same level, variation in  $c^{-(b+1)}$  would describe the remaining variation or the extent to which peak yield is maintained. In 1970 Wood used this measure of persistency in log form so that the measure of persistency of lactation became:

$$S = -(b+1) \ln c. \quad (2.12)$$

Rowland et al. (1982) mentions that because the Wood persistency measure is dimensionless, it is a valuable measure in comparison of persistency among both the various lactations of the same cow and lactations of different cows within a herd. Grossman et al. (1999) criticised the Wood persistency measure as being “difficult to interpret biologically”.

To this day Wood’s equation is still widely used and generally regarded by animal scientists as one of the best models that exist for modelling lactation. There is, however, one justifiable criticism of the Wood model in that it implicitly results in a production level of zero at time  $t = 0$ , which is known not to be true in most mammal species. Tozer and Huffaker (1999) do, however, state that a cow initially yields colostrum instead of true milk, which is not considered to have any economic value, and therefore a cow only enters a dairy herd as contributor once it comes into true milk. From an economic and management point of view, fixing milk yield at zero therefore does not represent a significant problem. Some studies of the Wood model (Scott et al., 1996) found that this model has the tendency to overestimate milk yield prior to peak yield and in late lactation. Underestimation of milk yield in mid-

lactation sometimes also occurs. This has resulted in the ongoing search for an even better lactation model.

In the study of lactation curves in dairy goats Fuller (1969) used grafted polynomials to obtain a lactation model of the following form:

$$y_t = a + bt + ct^2 + dr_1^2 + mr_2^2 \quad (2.13)$$

where  $r_1$  has the value  $t - 52$  when  $t > 52$  and 0 otherwise and  $r_2$  has the value  $t - 85$  when  $t > 85$  and 0 otherwise. No reference could however be found of this model being used in animals other than dairy goats and therefore the reason for the choice of the values 52 and 85 in the above could not be ascertained.

In 1971 the search for an improvement on the then existing lactation models continued with the proposal of the quadratic model by Dave:

$$y_t = a + bt + ct^2. \quad (2.14)$$

The next attempt at finding a good lactation model was that of Madalena, Martinez and Freitas (1979) with the use of a simple linear regression model:

$$y_t = a - bt. \quad (2.15)$$

The ingenuity of this model is questionable as it only represents a straight line with declining slope and is therefore unable to model the initial rise to peak yield.

To improve on the previous effort Molina and Boschini (1979) proposed the combination of two straight lines of equal but opposite slopes that intersect at peak yield at time  $t_0$ :

$$y_t = \begin{cases} a + bt & t < t_0 \\ a + b(2t_0 - t) & t \geq t_0 \end{cases}. \quad (2.16)$$

The idea of equal but opposite slopes may be questionable, because milk yield in most mammal species rises to peak at a faster rate than the subsequent decline after peak yield has been reached.

Working from a popular model as base, Dhanoa (1981) attempted to reparameterise the gamma function proposed by Wood in (2.6), with the following result:

$$y_t = at^{mc} \exp(-ct) \quad (2.17)$$

where,  $m$  = time until peak milk yield is reached. This model resulted in a lower correlation between parameters  $m$  and  $c$ , than was the case between parameters  $b$  and  $c$  in the original Wood model.

In 1982 Singh and Gopal proposed two new models. The first was the so-called linear cum log model:

$$y_t = a - bt + c \ln t \quad (2.18)$$

and the second the quadratic cum log model:

$$y_t = a + bt + ct^2 + d \ln t. \quad (2.19)$$

In linear cum log model time of peak yield is at  $c/b$ . Time of peak yield for the quadratic cum log model is at  $\frac{-b \pm \sqrt{b^2 - 8cd}}{4c}$ . Unfortunately at time  $t = 0$  both these models are undefined, because  $\ln t = -\infty$ .

In an attempt to include a seasonality effect in the Wood's model, Goodall (1983) proposed the inclusion of a categorical variable  $D$  that takes on the value 0 in the colder 6 month period from October to March in the northern hemisphere and the value 1 from April to September, resulting in the model:

$$y_t = at^b \exp(-ct + dD) \quad (2.20)$$

where  $d$  then estimates the seasonality factor. This technique allowed for quantitative assessment of the effect of seasonal changes on yield.

Another modification of the gamma function proposed by Wood was attempted by Jenkins and Ferrell (1984) by setting the exponent of  $t$ , which is the value of  $b$  in the Wood model, equal to 1:

$$y_t = at \exp(-ct). \quad (2.21)$$

This model has one important limitation in that the rise to peak yield is relatively slow, rendering this model of little use in practise (Landete-Castillejos and Gallego, 2000).

In 1987 Ali and Scheaffer suggested a polynomial regression model of the following form be used to model lactation:

$$y_t = a + bt + ct^2 + d \ln t + k(\ln t)^2 \quad (2.22)$$

This model only adds one term,  $k(\ln t)^2$ , to the model fitted by Singh and Gopal (1982).

Working from the model introduced by Fischer as basis, Wilmink (1987) introduced two lactation models. In the first model a quadratic term was added and an adjustment made to the exponential term, resulting in

$$y_t = a + bt + c \exp(-wt) + dt^2. \quad (2.23)$$

This model was then adjusted to obtain the second model by dropping the quadratic term from his initial function:

$$y_t = a + bt + c \exp(-wt). \quad (2.24)$$

In both (2.23) and (2.24)  $a$  may be interpreted as the level at which production commences,  $b$  as the decrease after peak yield is reached and  $c$  as the initial rise to peak. The factor  $w$  was set equal to 0,05 and is related to the time of peak yield, which for the data on Dutch Friesians used in this study was approximately 50 days after parturition. The model in (2.24) was again applied by Olori et al. in 1999, but they estimated this factor to be  $w = 0,61$ .

In 1988 Papajcsik and Bodero searched for a better performing model by combining into functional pairs combinations of certain increasing functions  $t^b$ ,  $1 - \exp(-t)$ ,  $\ln(t)$  and  $\arctan(t)$ , and decreasing functions  $\exp(-t)$  and  $1/\cosh(t)$ , where  $\arctan$  and  $\cosh$  respectively refers to the arctangent and hyperbolic cosine functions. This resulted in the following six models:

$$y_t = at^b / \cosh(ct) \quad (2.25)$$

$$y_t = a(1 - e^{-bt}) / \cosh(ct) \quad (2.26)$$

$$y_t = a \arctan(bt) / \cosh(ct) \quad (2.27)$$

$$y_t = a \ln(bt) \exp(-ct) \quad (2.28)$$

$$y_t = a \ln(bt) / \cosh(ct) \quad (2.29)$$

and 
$$y_t = a \arctan(bt) \exp(-ct). \quad (2.30)$$

Their study compared these six models to those described in (2.1) to (2.3), (2.6) and (2.7), (2.14) to (2.16), and (2.18). The models of Wood and (2.25) were found to be the best representations of the lactation curve for the data on Holstein cows considered.

Next a novel approach was introduced by Grossman and Koops (1988) in that they suggested that lactation could be viewed as a multiphasic biological process. Although at this point it was not uncommon to view lactation as a two stage process, usually divided into incline until peak yield as first stage and decrease after peak yield as second, nobody thought of suggesting



lactation to have more than two stages. The suggested multiphasic logistical function determines the total milk yield by obtaining the sum of the yield resulting from each of the lactation phases:

$$y_t = \sum_{i=1}^n \{a_i b_i [1 - \tanh^2(b_i(t - c_i))]\} \quad (2.31)$$

where  $n$  is the number of lactation phases considered and  $\tanh$  is the hyperbolic tangent. For each phase  $i$ , peak yield equals  $a_i b_i$  and occurs at time  $c_i$ . The duration of each phase is related to  $2b_i^{-1}$  that represents the time it takes to acquire 75% of asymptotic total yield during that phase. This model was applied as a two stage or diphasic and a three stage or triphasic model only, with a better fit resulting from the triphasic model due to smaller and less correlated residuals. This is attributed to the fact that early in lactation the diphasic model results in a poor fit because the hyperbolic tangent requires symmetry in both phases, and when only two phases are considered a symmetric curve does not fit the possible steep rise that occurs early in lactation. Gipson and Grossman (1989) noted that, although more research in this regard is required, for a diphasic model the first phase could possibly be considered as the so-called “peak” phase because of its “proximity to overall peak and short duration”. Similarly the second phase could perhaps be referred to as the “persistency” phase. This model is criticised by Rook et al. (1993), because, although it seems to behave well when fitted to lactation data, no justification could at the time be given for why lactation may be viewed as a multiphasic process.

In 1989 Morant and Gnanasakthy considered curves that resulted from the study of the proportional rates of changes in lactation yield. Mathematically the pattern of the proportional changes is defined by  $(dy/dt)/y$ . Lactation curves were obtained from these proportional rates of change by determining the integral of  $(dy/dt)/y$  which then results in the natural logarithm of yield at time  $t$ , or  $\ln y_t$ . The result:

$$y_t = a \exp(-bt + d \exp(-kt)/k) \quad (2.32)$$

$$y_t = at^b \exp(-(c - dt)t) \quad (2.33)$$

$$y_t = a \exp(-bt + ct^2 + d \exp(-kt)/k) \quad (2.34)$$

$$y_t = a \exp(-bt + ct^2 + d/t) \quad (2.35)$$

$$y_t = a \exp(-bt + ct^2 + d/(t+k)). \quad (2.36)$$

They experienced problems in fitting the models denoted in (2.32), (2.34) and (2.36). Fitting the curves in (2.33) and (2.35) were, however, straight forward, with the smallest correlation among parameters obtained for (2.35). This led to the reparameterisation of (2.35) resulting in the model:

$$y_t = a \exp(-bt' + ct'^2 + d/t) \quad (2.37)$$

where  $t' = (t - t_0)/100$  and  $t_0$  is a constant, which for the purpose of their study was fixed at the value 150 days. Later researchers of lactation models, such as Williams (1993), referred to this model as the 4-parameter Morant model. Morant and Gnanasakthy then rearranged this model so that it became

$$y_t = a \exp(-bt'(1 + rt') + ct'^2 + d/t) \quad (2.38)$$

where  $r$  is a constant determined as the slope of the regression for the estimates of parameter  $c$  on the estimates of parameter  $b$  as obtained from (2.37). It was noted that as a result of the increase in the value of  $t$  as time goes by, the parameter  $d$  only really affects the shape of the curve in the early days of lactations with its effect becoming more and more negligible as time goes by. A major advantage of this model is that the parameters have relatively simple interpretations. The logarithm of parameter  $a$  represents the expected yield on day  $t_0$ . Parameter  $b$  is defined as the rate of change in yield at  $t_0$  and is the main shape-affecting parameter of the curve. The parameter  $c$  also affects the shape of the lactation curve and is said to measure “the extent to which persistency changes during lactation”. They, however defined persistency as the extent to which day-to-day yields at any stage of lactation are maintained and may be determined using

$$1 - [b + 2(br - c)t']/100 - d/t^2. \quad (2.39)$$

The parameter  $d$  provides the rate at which yield increase during early lactation. This model was also adopted by Williams (1993) to fit lactation curves to British dairy goats, but he comments on the fact that effective estimation of  $d$  is problematic as a result of the drastic effect of errors in observation early in lactation, as well as because of the small amount of data available prior to peak yield when compared to that after peak yield. To overcome this he used the mean of the estimated  $d$ 's in the model instead of calculating a unique estimate for every lactation curve fitted. This, however, still does not imply that  $d$  is a constant. Gipson and Grossman (1990) in a review of lactation curves fitted to lactation data obtained from dairy goats noted that the model in (2.33) may be referred to as the general exponential, and that by setting both  $b$  and  $d$  equal to zero the model becomes the exponential function suggested by Brody, Ragsdale and Turner (1923), by setting  $b$  equal to zero the model results

in the parabolic exponential function suggested by Sikka (1950), and by setting  $d$  equal to zero the incomplete gamma function fitted by Wood (1967) is obtained.

In order to limit the number of parameters to be estimated in case of a multiphasic model, a diphasic model is preferred over one with more than two phases. Weigel et al. (1992) attempted to improve the diphasic version of the Grossman and Koops model by means of a power transformation of time and replacing  $a_i b_i$  by a new parameter  $d_i$ . This was the applied to lactation data as both a single stage or monophasic model :

$$y_t = d \left[ 1 - \tanh^2 \left( b \left( t^k - c \right) \right) \right] \quad (2.40)$$

and a diphasic model:

$$y_t = d_1 \left[ 1 - \tanh^2 \left( b_1 \left( t^k - c_1 \right) \right) \right] + d_2 \left[ 1 - \tanh^2 \left( b_2 \left( t - c_2 \right) \right) \right]. \quad (2.41)$$

In 1993 Rook, France and Dhanoa again attempted to model lactation as the product of a constant,  $A$ , a monotonically increasing function of time,  $f_1(t)$ , and a monotonically decreasing function of time,  $f_2(t)$ . The following six monotonically increasing functions were considered:

the Mitscherlich function  $1 - a \exp(-bt)$ , (2.42)

the Michaelis-Menten function  $1/[1 + a(b + t)]$ , (2.43)

the generalised saturation kinetic function  $1/[1 + a(b + t^c)]$ , (2.44)

the logistic function  $1/(1 + a \exp(-bt))$ , (2.45)

the Gompertz function  $a \exp[(-\ln a)(1 - \exp(-bt))]$  (2.46)

and the hyperbolic tangent  $[1 + \tanh(a + bt)]/2$ . (2.47)

Only two monotonically decreasing functions were considered:

the exponential function  $\exp(-dt)$  (2.48)

and the inverse straight line  $1/(1 + ct)$ . (2.49)

This resulted in twelve lactation curves of the form  $y_t = A f_1(t) f_2(t)$ , that were fitted, together with the model proposed by Wood, to lactation data obtained from dairy cows. It was found that the Wood model together with the following function combinations fitted the data well: Mitscherlich  $\times$  exponential, Michaelis-Menten  $\times$  exponential, logistic  $\times$  exponential and logistic  $\times$  inverse straight line.

Williams (1993) suggested that the 4-parameter Morant model be extended to a six parameter model to make it more comparable to the diphasic model suggested by Grossman and Koops in (2.31) with its six parameters. The result was a model Williams referred to as the 6 parameter Morant model:

$$y_t = a \exp(-bt' + ct'^2 + d/t + kt'^3 + mt'^4) \quad (2.50)$$

where  $t' = (t - t_0)/100$  and  $t_0$  is a constant that Williams also fixed at the value 150 days. As a result of the large number of parameters in this model, it provided a good fit to lactation data of white British dairy goats.

In an effort to overcome the underestimation of peak yield and overestimation of yield later in lactation that occurs as a result of using the Wood model, Cappio-Borlino, Pulina and Rossi (1995) introduced a non-linear modification of the Wood model:

$$y_t = at^{b \exp(-ct)} \quad (2.51)$$

Although a lot more complex than the Wood model, this model reduced the extent of both underestimation early in lactation and overestimation in the final stage of lactation for the data used. Franci et al. (1999) refers to this model as the bi-exponential function. They also found that this model was well suited to describe lactation with an initial sharp rise in milk production.

Guo and Swalve (1995) introduced a model, referred to as the mixed log model, of the following form:

$$y_t = a + bt^{\frac{1}{2}} + c \ln t. \quad (2.52)$$

This model differs from that suggested by Singh and Gopal (1982), referred to as the linear cum log model, in that the square root of  $t$  is obtained in the second term. The model, however, tends to underestimate peak yield, while overestimating the post-peak yield (Olori et al., 1999).

In 1999 Grossman, Hartz and Koops in their research on the persistency of lactation yield also introduced a novel approach to modelling lactation. They viewed the lactation curve of an average cow as the result of three intersecting straight lines. The first of these lines is said to describe the initial rise in yield to peak, the second line has a slope of zero and represents the peak yield over the period for which it is sustained, while the third line represents subsequent

decline after peak. As a result, two very similar models were suggested, the first the so-called lactation persistency model:

$$y_t = y_P + b_1(t - t') - a_1 b_1 \ln \left[ \frac{e^{(t/a_1)} + e^{(t'/a_1)}}{1 + e^{(t'/a_1)}} \right] + a_2 b_3 \ln \left[ \frac{e^{(t/a_2)} + e^{((t'+P)/a_2)}}{1 + e^{((t'+P)/a_2)}} \right] \quad (2.53)$$

where  $y_P$  is the level of constant yield during the peak phase;  $b_1$  is the slope of the straight line during the initial inclining phase,  $b_3$  is the slope of the straight line during the final declining phase,  $t'$  is the transition time from the slope of the first straight line to the slope of the second straight line,  $a_1$  and  $a_2$  are the durations of transition from the slopes of the first to the second, and from the second to the third straight line, and  $P$  is the number of days during which the level of constant yield of the peak phase is maintained. The second model is simply a reduced form of the first model and referred to as the reduced lactation persistency model:

$$y_t = \frac{y_P}{t'} t - \frac{y_P}{t'} \ln \left[ \frac{e^t + e^{t'}}{1 + e^{t'}} \right] + b_3 \ln \left[ \frac{e^t + e^{t'+P}}{1 + e^{t'+P}} \right] \quad (2.54)$$

with the same parameter interpretations as in the lactation persistency model above. The main advantage of both these models is that persistency,  $P$ , forms part of the model in the form of a parameter.

As commented by Tozer and Huffaker (1999), a wide variety of different mathematical equations for modelling lactation are found in the literature and these have been applied to lactation data from a variety of different mammals. Some of these resulting lactation curves perform better in certain studies than others, but so far no single lactation model has emerged as a consistent best performer in all cases. Rekaya, Carabaño and Toro (2000) point out that the most desirable model would be one with a limited number of parameters and a biological interpretation that is of value from a practical point of view.

Recently the focus in the study of lactation has moved away from attempts to find a standard robust model. Instead researchers are now more concerned with the methods used to fit the existing models.

Table 2.1 contains a summary of the above lactation models and the different animals to which these lactation models have been applied, together with the reference numbers of the

literature in the References where data on these animals have been used in the application of the various models. Models that have the same origin or that are similar in structural nature have been grouped together in chronological order within blocks.

**Table 2.1: Lactation model and application summary**

<p>Brody et al. (1923) - exponential decline function:  <math>y_t = a \exp(-ct)</math>            Dairy cows: [5]            Red deer: [55]</p>	<p>Sikka (1950) - parabolic exponential function:  <math>y_t = a \exp(bt - ct^2)</math>            Dairy cows: [87]            Red deer: [55]</p>
<p>Fischer (1958):  <math>y_t = a - bt - a \exp(-ct)</math>            Württemberg Spotted Mountain cows: [20]            British Friesian cows: [78]            Dairy cows: [13] [95]            Holstein cows: [86]</p>	<p>Wood (1967) - gamma function:  <math>y_t = at^b \exp(-ct)</math>            Friesian cows: [103] [104] [105]            British Friesian cows: [57] [78] [106]            Dutch Friesian cows: [42]            Holstein-Friesian cows: [70] [75] [94]            Holstein cows: [19] [83] [86] [91]            Simmental cows: [89]            Spanish dairy cows: [96]            Brown Swiss cows: [82]            Dairy cows: [13] [36] [37] [77] [85] [95] [107] [109]            Hanwoo Korean beef cows: [53]            US Sheep breeds: [81]            Massese sheep: [21]            Dairy sheep: [9] [10]            Comisana sheep: [73]            Merino sheep: [38] [39]            Crossbred sheep: [93]            Saanen dairy goats: [40]            White British dairy goats: [100]            Red Sokoto goats: [2]            Goats: [79]            Red deer: [55]</p>
<p>Vujicic and Bacic (1961):  <math>y_t = tc^{-a} \exp(-ct)</math>            Dairy cows: [97]</p>	<p>Dhanoa (1981) - reparameterise Wood:  <math>y_t = at^{mc} \exp(-ct)</math>            Friesian cows: [18]</p>
<p>Wilmink I (1987):  <math>y_t = a + bt + c \exp(-wt) + dt^2</math> for (<math>w = 0,05</math>)            Dutch Friesian cows: [101]</p>	<p>Goodall (1983) - seasonally adjusted Wood:  <math>y_t = at^b \exp(-ct + dD)</math>            British Friesian cows: [34] [35] [57]</p>
<p>Fischer (1958):  <math>y_t = a - bt - a \exp(-ct)</math>            Württemberg Spotted Mountain cows: [20]            British Friesian cows: [78]            Dairy cows: [13] [95]            Holstein cows: [86]</p>	<p>Jenkins and Ferrell (1984) - adjustment of Wood:  <math>y_t = at \exp(-ct)</math>            Dairy cows: [49]            Red deer: [55]</p>
<p>Wilmink II (1987):  <math>y_t = a + bt + c \exp(-wt)</math>            South African Holsteins cows: [68] (<math>w = 0,05</math>)            South African Jerseys cows: [68] (<math>w = 0,05</math>)            Holstein-Friesian cows: [70] (<math>w = 0,61</math>)            Dutch Friesian cows: [101] (<math>w = 0,05</math>)            Dairy cows: [95] (<math>w = 0,05</math>)</p>	<p>Cappio-Borlino et al. (1995) - bi-exponential function:  <math>y_t = at^{b \exp(-ct)}</math>            Sardinian dairy sheep: [7]            Massese sheep: [21]</p>
<p>Nelder (1966) - inverse polynomial:  <math>y_t = t / (a + bt + ct^2)</math>            Red deer: [55]            Dairy cows: [4]            Holstein-Friesian cows: [70] [94]            Holstein cows: [83] [86]</p>	<p>Morant and Gnanasakthy (1989) - general exponential:  <math>y_t = at^b \exp(-(c - dt)t)</math>            Friesian heifers: [67]</p>
<p>Fuller (1969) - grafted polynomials:  <math>y_t = a + bt + ct^2 + dt_1^2 + mr_2^2</math></p>	
<p>Dave (1971) - quadratic function:  <math>y_t = a + bt + ct^2</math>            Dairy sheep: [10]            Indian water buffalo: [14]</p>	

<p>Madalena et al. (1979) - simple linear regression:  <math>y_i = a - bt</math>            Holstein-Friesian cows: [15]            Holstein-Friesian × Gir cows: [15]</p>	<p>Morant and Gnanasakthy (1989) – 4-parameter Morant:  <math>y_i = a \exp(-bt + ct^2 + d/t)</math>            Friesian heifers: [38] [67]</p>
<p>Molina and Boschini (1979) - straight lines of equal, but opposite slopes:  <math display="block">y_i = \begin{cases} a + bt &amp; t &lt; t_0 \\ a + b(2t_0 - t) &amp; t \geq t_0 \end{cases}</math>            Holstein cows: [66]</p>	<p>Morant and Gnanasakthy (1989):  <math>y_i = a \exp(-bt'(1 + rt') + ct^2 + d/t)</math>            where <math>t' = (t - t_0)/100</math> and <math>t_0 = 150</math> days.            Friesian heifers: [67]            Dairy cows: [95]            British Friesian cows: [57]            White British dairy goats: [100]</p>
<p>Singh and Gopal (1982) - linear cum log:  <math>y_i = a - bt + c \ln t</math>            Indian dairy buffalo: [88]            Holstein cows: [86]</p>	<p>Williams (1993) – 6-parameter model Morant:  <math>y_i = a \exp(-bt' + ct'^2 + d/t + kt'^3 + mt'^4)</math>            where <math>t' = (t - t_0)/100</math> and <math>t_0 = 150</math> days.            White British dairy goats: [100]</p>
<p>Singh and Gopal (1982) - quadratic cum log:  <math>y_i = a + bt + ct^2 + d \ln t</math>            Indian dairy buffalo: [88]            Holstein cows: [86]</p>	<p>Morant and Gnanasakthy (1989):  <math>y_i = a \exp(-bt + d \exp(-kt)/k)</math>  <math>y_i = a \exp(-bt + ct^2 + d \exp(-kt)/k)</math>  <math>y_i = a \exp(-bt + ct^2 + d/(t + k))</math>            Friesian heifers: [67]</p>
<p>Ali and Scheaffer (1987) - polynomial regression model:  <math>y_i = a + bt + ct^2 + d \ln t + k(\ln t)^2</math>            Holstein-Friesian cows: [70]            Dairy cows: [1]</p>	<p>Grossman and Koops (1988) - multiphasic logistical function:  <math display="block">y_i = \sum_{i=1}^n \{a_i b_i [1 - \tanh^2(b_i(t - c_i))]\}</math>            Dutch Friesian cows: [42]            Holstein-Friesian cows: [94]            Israeli Holstein cows: [15]            Dairy cows: [95]            White British dairy goats: [100]            Dairy goats: [31]</p>
<p>Guo and Swalve (1995) - mixed log model:  <math>y_i = a + bt^{\frac{1}{2}} + c \ln t</math>            Dairy cows: [43]            Holstein-Friesian cows: [70]</p>	
<p>Papajcsik and Boderó (1988):  <math>y_i = at^b / \cosh(ct)</math>  <math>y_i = a(1 - e^{-bt}) / \cosh(ct)</math>  <math>y_i = a \arctan(bt) / \cosh(ct)</math>            Holstein cows: [86]            Friesian cows: [72]</p>	<p>Weigel et al. (1992) - adapted monophasic function:  <math>y_i = d[1 - \tanh^2(b(t^k - c))]</math>            Dairy cows: [99]            Holstein cows: [86]</p>
<p>Papajcsik and Boderó (1988):  <math>y_i = a \ln(bt) \exp(-ct)</math>  <math>y_i = a \ln(bt) / \cosh(ct)</math>  <math>y_i = a \arctan(bt) \exp(-ct)</math>            Friesian cows: [72]</p>	<p>Weigel et al. (1992) - adapted diphasic function:  <math>y_i = d_1[1 - \tanh^2(b_1(t^k - c_1))] + d_2[1 - \tanh^2(b_2(t - c_2))]</math>            Dairy cows: [99]            Holstein cows: [86]</p>

<p>Rook et al. (1993):  Michaelis-Menten×exponential:  <math display="block">y_t = A [1 + a(b + t)]^{-1} \exp(-dt)</math> Generalised saturation kinetic×exponential:  <math display="block">y_t = A [1 + a(b + t^c)]^{-1} \exp(-dt)</math> Logistic×exponential:  <math display="block">y_t = A (1 + a \exp(-bt))^{-1} \exp(-dt)</math> Gompertz×exponential:  <math display="block">y_t = A (a \exp[(-\ln a)(1 - \exp(-bt))]) \exp(-dt)</math> Hyperbolic×exponential:  <math display="block">y_t = A ([1 + \tanh(a + bt)]/2) \exp(-dt)</math> Mitscherlich×inverse straight line:  <math display="block">y_t = A (1 - a \exp(-bt)) (1 + ct)^{-1}</math> Michaelis-Menten×inverse straight line:  <math display="block">y_t = A [1 + a(b + t)]^{-1} (1 + ct)^{-1}</math> Generalised saturation kinetic  × inverse straight line:  <math display="block">y_t = A [1 + a(b + t^c)]^{-1} (1 + ct)^{-1}</math> Logistic×inverse straight line:  <math display="block">y_t = A (1 + a \exp(-bt))^{-1} (1 + ct)^{-1}</math> Gompertz×inverse straight line:  <math display="block">y_t = A (a \exp[(-\ln a)(1 - \exp(-bt))]) (1 + ct)^{-1}</math> Hyperbolic×inverse straight line:  <math display="block">y_t = A ([1 + \tanh(a + bt)]/2) (1 + ct)^{-1}</math> Dairy cows: [77]</p>	<p>Grossman et al. (1999) - lactation persistency model:  <math display="block">y_t = y_p + b_1(t - t') - a_1 b_1 \ln \left[ \frac{e^{(t/a_1)} + e^{(t'/a_1)}}{1 + e^{(t'/a_1)}} \right]</math> <math display="block">+ a_2 b_3 \ln \left[ \frac{e^{(t/a_2)} + e^{((t'+p)/a_2)}}{1 + e^{((t'+p)/a_2)}} \right]</math> Dairy cows: [41] [95]</p> <hr/> <p>Grossman et al. (1999) - reduced lactation persistency model:  <math display="block">y_t = \frac{y_p}{t'} t - \frac{y_p}{t'} \ln \left[ \frac{e^t + e^{t'}}{1 + e^{t'}} \right] + b_3 \ln \left[ \frac{e^t + e^{t'+p}}{1 + e^{t'+p}} \right]</math> Dairy cows: [41] [95]</p> <hr/> <p>Rook et al. (1993):  Mitscherlich × exponential:  <math display="block">y_t = A (1 - a \exp(-bt)) \exp(-dt)</math> Holstein-Friesian cows: [94]  Dairy cows: [77] [95]</p>
---	--

## 2.2 TYPICAL LACTATION CURVE SHAPES

It is generally accepted that the standard lactation curve that applies to most mammal species increases up to the point where peak yield is attained, whereafter a gradual decline takes place until the end of the lactation process is reached. Standard lactation curves of this nature are often also referred to as type I curves (Landete-Castillejos and Gallego, 2000). In fitting the various theoretical lactation curve models to data researcher have, however, also come across another lactation curve shape, one that has no peak and graphically represents a curve that is in continual decline. Lactation curves of this form are generally referred to as atypical (Ferris et al., 1985) or type II (Landete-Castillejos and Gallego, 2000) lactation curves.

There are a number of possible reasons for the occurrence of these atypical or type II lactation curves. In almost all cases where lactation is study, test day records are obtained by weighing



milk yield and measuring milk composition traits at various intervals after parturition. These intervals are often based on certain days preferred by management, and do not occur at the same stage of lactation for all animals. As a result, it sometimes happens that for a particular animal the first observation is made after peak yield had already been reached. If this is the case, when fitting a lactation curve to this data, the illusion is created that no rise to peak yield and subsequent peak have occurred, resulting in an atypical lactation curve as best fit. The same kind of result could also occur in cases where the intervals between test day records collected during the early days of lactation is so great that it is unable to successfully depict peak yield, for example where the first observation occurs during the rise to peak and the second during the declining phase after peak, but the resulting weight is lower for the second observation than for the first observation. Observational errors when weighing milk yield could also attribute to fitting atypical lactation curves. In a small number of non-dairy mammals such as Red deer (Arman et al., 1974) and Iberian Red deer (Loudan et al., 1983; Garcia et al., 1999; Landete-Castillejos and Gallego, 2000) atypical or type II lactation curves have been found as the applicable lactation curve. Landete-Castillejos and Gallego (2000) mentions the possibility that atypical or type II lactation curves might not be uncommon in wild ungulates.

Researchers differ in opinion on how such atypical or type II lactation curves should be handled. Some researchers feel that these curves imply that peak yield occurred prior to parturition and that for this reason these curves represent an unrealistic picture and should not be allowed in the study of lactation in dairy animals (Ferris et al., 1985; Tekerli et al., 2000).

Another school of thought is that when fitting a parametric lactation model to data the possible occurrence of atypical or type II lactation curves should be limited by placing restriction on the values the parameters of the curve can take and thereby forcing the lactation curve to be typical. In such cases the parameter values that would lead to atypical or type II curves are determined, restrictions on the values of these parameters formulated and appropriate estimation methods employed so that these values are not obtained. In this respect Bayesian analysis using the Gibbs sampler, together with an acceptance-rejection step, is a useful tool to reduce the occurrence of atypical or type II lactation curves (Varona et al., 1998; Rekaya et al., 2000; Groenewald and Viljoen, 2003). Varona et al. (1998) reported a reduction in atypical or type II lactation curves from 29% to 0,25% in Spanish dairy cows using Wood's model and a Bayesian analysis of this nature.

A third approach to handling atypical or type II lactation curves was considered by Landete-Castillejos and Gallego (2000) after determining that an atypical or type II lactation curve was most often the lactation curve of choice in Iberian Red deer. They pointed out that one should carefully inspect your data before selecting the model(s) to fit to ensure that the model(s) fitted do indeed have the ability to accommodate atypical or type II lactation curves when required. Models such as the inverse polynomial in (2.6) developed by Nelder (1966) and the model in (2.21) developed by Jenkins and Ferrel (1984) are unable to produce atypical or type II lactation curves.

Schneeberger (1981) found that in dairy cows the percentage of atypical or type II lactation curves decreased with increasing parity number. This result makes it plausible that atypical or type II curves may also occur in the ordinary course of lactation in dairy animals. For this reason using a Bayesian analysis that would reduce, but not totally eliminate the occurrence of this curve type would possibly be the golden mean.

### **2.3 CURVES FITTED TO COMPOSITION TRAITS**

As pointed out earlier depending on the intended use of milk produced, the composition traits of milk could also be important. When this is the case a composition trait curve may be fitted to the milk fat, protein and lactose content of the milk. The percentages of milk fat and protein are typically convex functions over time, which means that they vary inversely to yield. Considerable day-to-day variation has been noted in the fat content of milk (Whittemore, 1980). The percentage of lactose, however, does not vary in relation to yield, but goes into a steady decline over the whole lactation. As a result of this both milk fat and protein are often also modelled with the same functions that is used to model milk yield, on condition that they are able to take on a convex form. Lactose, on the other hand, requires a function that has the ability to model decline, as no rise to peak is present. Very little work has been done on modelling the lactose content of milk.

The earliest reference found on lactation models fitted to composition traits, was that of Goodall (1986) who fitted the seasonally adjusted Wood model of (2.20) to both the milk fat and the protein percentages of milk.

In 1987 Wilmink fitted both the models that he suggested in (2.23) and (2.24) to model milk yield, to the milk fat and protein content of milk as measured in terms of weight. The factor  $w$  in these models was again set equal to 0,05 for estimating fat and protein content curves.

Morant and Gnanasakthy (1989) use the rearranged reparameterisation of their model as given in (2.38) to model milk fat, protein and lactose content as measured in kilograms. This was the first reference found in which an attempt was made to also model lactose. The model used seemed to fit protein and lactose content, but in the case of fat a larger residual variation was obtained. This should be expected due to the day-to-day variation in fat content of milk.

De Boer et al. (1989) only attempted to model the fat content of milk by fitting a diphasic function, i.e. the model in (2.31) with  $n = 2$ , to the fat content of milk as measured by weight.

Sakul and Boylan (1992) fitted the Wood model of (2.7) to the proportions of fat, protein and lactose contained in milk. The Wood model provided a satisfactory fit with respect to the fat and protein content of milk, but was unable to adequately model the proportion of lactose contained in milk.

In 2001 Mostert, Theron and Kanfer again used the reduced Wilmink model provided in (2.24) and fitted it to the milk fat and protein content of milk measured in kilograms. They also used 0,05 as value of the factor  $w$  in this study

Groenewald and Viljoen (2003) also fitted the Wood model in (2.7) to the fat and protein content of milk measured in kilograms.

## **2.4 PERSISTENCY**

The earliest reference to the persistency of lactation found, was that of Wood (1967) in which he defined persistency as “the extent to which peak yield is maintained”. Wood then also developed a measure of persistency to correspond to this definition (see (2.12)). Since then a number of other researchers have also attempted to not only define persistency, but also to develop a measure of persistency. No single consistent definition of persistency, however, exists within the framework of lactation research and as a result the defined measures of persistency are also inconsistent. Morant and Gnanasakthy (1989) defined persistency as “the extent to which yield is maintained from day to day at any stage of lactation” and then

developed the measure given in (2.39), which is greater than one when yield increases, equal to one when yield is constant, and less than one in case of declining yield. The most widely used definition seems to be that persistency is the rate of decline in milk yield after peak, and the measures of persistency developed as a result of this definition include ratios of yield during various stages of lactation, variation in test day yields and formulations of measures from lactation model parameters (Grossman, Hartz and Koops, 1999).

The first and to date only attempt at developing model that contains persistency as a model parameter was that of Grossman, Hartz and Koops (1999). They first defined persistency as the number of days during which the level of milk yield remains constant and then developed the models in (2.52) and (2.53) referred to respectively as the lactation persistency and reduced lactation persistency models.

Various factors seem to influence persistency, but parity seems to have the greatest influence. In studies of the lactation of dairy cows it was reported that the Wood measure of persistency is significantly higher in first lactation than in later lactations (Rowlands et al., 1982). Sölkner and Fuchs (1987) obtained the same result for dairy cows by using both ratios of yields during different stages of lactation, as well as measures expressed as ratios of variation that occurred in test day yields. They also found that the season of parturition, as well as the total milk yield had a significant effect on measures of persistency obtained through these two methods. Tekerli et al. (2000) employed three persistency measures in their study of dairy cows; Wood measure of persistency, ratios of yields during different stages of lactation, and ratios of variation that occurred in test day yields; and in all cases persistency during first lactation proved greatest. In the study of lactation in dairy goats using parameter based persistency formulas (Gipson and Grossman, 1990; Ruvuna et al., 1995), it was found that persistency was significantly affected by parity, season of kidding and level of total production, with higher yielding does being less persistent than lower yielding does. Persistency was greatest during cooler seasons and during first parity, declining there after. Portolano et al. (1996) found that for Comisana sheep significant positive correlation existed between persistency and time of peak yield, while persistency and peak yield, and persistency and parity were negatively correlated. A more persistent ewe would therefore have a curve with lower peak that occurred later relative to less persistent ewes. Schneeberger (1981) also found negative genetic correlation between initial yield and persistency when studying Swiss Brown dairy cows. Gipson and Grossman (1989) also noted a negative relationship between

both persistency and parity, and persistency and yield for dairy goats. Lactation curve of dairy goats were found to in general be flatter and as a result more persistent than those of dairy cows (Ruvuna et al., 1995; Apka et al., 2001).

To summarise, four groups of persistency measures have, been identified: 1) measures derived from the lactation model, containing estimated parameters from the afore mentioned model; 2) measures that express persistency as the ratio of yields during different stages of lactation; 3) measures derived from variation that occurs in test day milk yield; and 4) the duration (in number of days) for which the level of milk yield remains constant (Grossman, Hartz and Koops, 1999). Because of these differences in the nature of persistency measures, it is not possible to directly compare the persistency values obtained by the various measures, although there seems to be general consensus with respect to the effect of different biological and environmental factors on the various persistency measures.

Persistency of lactation is considered a very important characteristic of the lactation curve. As was mentioned in Chapter 1, dairy animals producing milk at a greater level of persistency are preferred in the dairy industry, because this would reduce and stabilise production cost and ultimately maximise profit. For example, cows with a flatter lactation curve is considered more persistent and would have an increased proportion of roughage in their rations, which would reduce their production cost (Tekerli et al., 2000).

# CHAPTER 3

## METHODS APPLIED IN FITTING LACTATION CURVES

### 3.1 INTERPOLATION TO OBTAIN TOTAL MILK PRODUCTION

Often in the management of a dairy herd interest is focussed on the total milk yield of an individual animal over the lactation period. As a result of both the time and labour cost factor involved in weighing milk produced during every extraction session, the milk yield of an animal is weighed at intervals determined by management. The result is test day records that do not provide a complete picture of the animal's lactation process, but from which the total yield have to be obtained. One method commonly used in the dairy industry approximates the total yield of an animal from the incomplete test day records using various interpolation procedures, sometimes also referred to as standardisation and projection procedures (Serrano et al., 2001).

One interpolation method commonly used is the so-called Test Interval Method (TIM) that makes use of a form of linear interpolation (Olori and Galesloot, 1999). This method assumes that the yield on any test day is the mean yield for the interval of time from halfway between this test day and the immediately preceding test day up to halfway between this test day and the one immediately thereafter. To make provision for the periods before the first day and after the last test day, however, the yield on the first test day is assumed to be the daily yield since calving, and the yield on the last test day is assumed to be the daily yield until termination of lactation. The total yield over the lactation is then determined through

summation to be as follows:

$$y = p_1 y_1 + \sum_{i=2}^n \{0.5 p_i (y_{i-1} + y_i)\} + p_{n+1} y_n$$

where,  $y$  = total yield

$y_i$  = test day yield as measured on test day  $i$

$p_1$  = number of days in the interval from calving date to first test date

$p_i$  = number of days in the interval between test day  $i$  and the preceding test day

$p_{n+1}$  = number of days in the interval between test day  $i$  and termination day

Olori and Galesloot (1999) pointed out that for this method to produce valid results, the first test day has to occur before day 50 in the cows lactation cycle. This method is also criticised for being unable to take into account the pattern of lactation that occurs around peak yield because of its dependence on the interval between test days.

## **3.2 IRIS**

IRIS is a computer software package that was developed in the Netherlands for the collection and sensible organisation of data for management purposes on cattle farms. One module of this package makes provision for milk recording, and is referred to as IRIS-Dairy. IRIS also includes modules on herd book registration, type classification and statistics on cattle data. Olori and Galesloot (1999) fully describe the procedure employed by IRIS-Dairy, which is summarised below.

IRIS-Dairy makes provision for prediction of the lactation of a cow very early in lactation. Using IRIS-Dairy it is already possible to predict lactation as soon as the first test-day record becomes available. Predicting the next lactation cycle based on the yield in the immediately preceding lactation is also possible on condition that the cow and herd of which it forms part participates in milk performance testing so that information on calving date and age, as well as production level of the herd crucial to these predictions are available.

This program uses what is known as the Standard Lactation Curve (SLAC) method to make such predictions. Standard lactation curves (SLAC) are obtained from the mean yield of a comparable group of cows referred to as contemporaries. A contemporary group consists of cows of similar age, with the same herd production level, calving in the same season. Based on the assumption that cows from the same contemporary group have similar production patterns, these standard lactation curves indicate the expected pattern of production for a cow similar to the contemporaries and guides the prediction of daily yields in the future part of lactation. Predictions are then based on the stage of lactation, and the yield of the cow on the last test day, together with the cumulative yield from its previous lactation as indicators of that cow's potential. The lactation curve of an individual cow is then obtained from the

standard lactation curve by incorporating the deviation of the observed yields from the expected yields.

Standard lactation curves relevant to the specific environmental and biological factors influencing a particular breed of dairy cow have been derived. SLAC curves for a breed had to be developed for class combinations of twenty herd level classes, eighteen parity/age classes and six calving seasons, which resulted in 2160 standard lactation curves for each of the lactation traits milk yield, fat content and protein content for that breed. This was done by using the test day records of the group of contemporaries from which the standard lactation curve had to be constructed and determining their yield at 15 fixed days, from the 10<sup>th</sup> day in intervals of 20 days up to day 290, during the lactations of these animals. To derive the yield on fixed days prior to the first test day and after the last test day the lactation model developed by Wilmink (1987) as denoted in (2.24), together with  $w = 0,05$  was used. The yields of fixed days between the first and last test days were obtained by means of interpolation between successive test day yields. The expected or mean yields on these fixed days were then obtained for the group of contemporary cows and, once these yields had been adjusted for the effect of herd level, age and calving season, they then represented (when connected) the standard lactation curve for such a contemporary group of cows. A fixed effects model with herd level, age class and calving season as fixed effects was used to estimate the effect of these factors.

To predict a daily yield level after the last actual test day using the standard lactation curve method, the following calculation is employed:

$$\hat{Y}_t = E(Y_t) + b_1[Y_{305_{PL}} - E(Y_{305_{PL}})] + b_2[X_i - E(X_i)]$$

where,

$\hat{Y}_t$  = predicted yield on day  $t$  of lactation

$E(Y_t)$  = expected yield on day  $t$  of lactation obtained from SLAC

$Y_{305_{PL}}$  = total yield of preceding lactation

$E(Y_{305_{PL}})$  = expected tot yield of preceding lactation obtained from SLAC

$X_i$  = yield measured on last test day of lactation in progress

$E(X_i)$  = expected yield for last test day of lactation in progress obtained from SLAC

$b_1, b_2$  = prediction factors



The prediction factors  $b_1$  and  $b_2$  are determined for each interval of 20 days from the 10<sup>th</sup> day of lactation for three parity groups (parities 1, 2, and 3 or higher) for each of the lactation traits milk yield, fat content and protein content. This is done by estimating the deviation of the observed yield from the expected yield for each cow included in the derivation of the standard lactation curves. Least squares regression of the model

$$Y_i - E(Y_i) = \beta_0 + \beta_1 [Y_{305_{PL}} - E(Y_{305_{PL}})] + \beta_2 [X_i - E(X_i)]$$

is used and the resulting estimates of the regression coefficients  $b_1$  and  $b_2$  provide the projection factor values of  $b_1$  and  $b_2$ .

This method of prediction is used in predicting milk yield, fat content and protein content curves. As may be expected, as lactation progresses the accuracy of prediction using this method improves. This method of prediction has two main advantages: 1) the lactation yield at the start of lactation (i.e. at time  $t = 0$ ) is not assumed equal to zero; and 2) it is relatively easy to use on farm management level, because all information required is either already available to management from historic or current lactation records or built into the program in the form of the standard lactation curves required.

### **3.3 LINEAR AND NON-LINEAR METHODS OF CURVE FITTING**

As alternative to the approximation methods above, the lactation curve models discussed in Chapter 2, together with the test day records of an animal, may be used to model the lactation of the animal. The total yield over the duration of the lactation may then be acquired using the method relevant to the lactation model fitted. Both non-linear and linear estimation methods have been employed to fit lactation curve models, where the method employed is determined by the nature of the model to be fitted. Some lactation models may be fitted by means of both non-linear and linear estimation.

Some of the lactation models are linear by nature, while others may be transformed into linear models. Wood already noted in 1967 that the gamma function converts to a simple linear regression model by performing a natural log transformation, making it possible to determine the values of  $a$ ,  $b$  and  $c$  by means of least squares estimation. This principle was applied by Cobby and Le Du (1978), Torres-Hernandez and Hohenboken (1980), Sakul and Boylan (1992), Tekerli et al. (2000) and others when fitting the Wood model. This approach to model

estimation by performing a natural log transformation to obtain a linear model may also be applied to other models, such as Brody's exponential decline function (Brody et al., 1923), Sikka's parabolic exponential function (Sikka, 1950), Goodall's seasonally adjusted Wood model (Goodall, 1983; Lennox et al., 1992), the Jenkins' adjustment of the Wood model (Jenkins and Ferrell, 1984), the 4-parameter Morant model (Lennox et al., 1992, Groenewald et al., 1995) and other models suggested by Morant and Gnanasakthy (Morant and Gnanasakthy, 1989), and the 6-parameter Morant model (Williams, 1993). Nelder's inverse polynomial model (1966) may also be transformed to a multiple regression model, but this time using an inverse transformation (Batra, 1986). Models that are intrinsically already linear without a natural log transformation are the grafted polynomial (Fuller, 1969), the quadratic model (Dave, 1971), the simple linear regression model of Madalena et al. (1979), the model of Molina and Boshini (1979), the linear cum log and quadratic cum log models (Singh and Gopal, 1982), the polynomial regression model (Ali and Scheaffer, 1987), and the mixed log model (Guo and Swalve, 1995).

Non-linear regression by means of the iterative Marquardt's compromise method (Marquardt, 1963) as available in SAS, has been the most common method of fitting non-linear lactation models to data thus far, and has been used to fit Nelder's inverse polynomial function, the Mitscherlich-exponential function, the multiphasic model with  $n = 1$  or monophasic fit (Tozer and Huffaker, 1999) and the Wood model (Ferris et al., 1985; Cappio-Borlino et al., 1995; Ruvuna et al., 1995; Portolano et al., 1996; Franci et al., 1999; Tozer and Huffaker, 1999). Another method frequently employed is non-linear regression using the Gauss-Newton iterative method also available in SAS. This was done for the Wood model, and the multiphasic function fitted with  $n = 2$  as diphasic to milk yield (Grossman and Koops, 1988; De Boer et al., 1989) and with  $n = 3$  as triphasic to fat (De Boer et al., 1989). Other models for which unspecified methods of non-linear fitting were used include the 1924-Brody model; the Fischer model (Cobby and Le Du, 1978; Rowlands et al., 1982), the Wilmink I model (Wilmink, 1987; Olori et al. 1999), the Wilmink II model (Wilmink, 1987; Olori et al. 1999; Vargas et al. 2000; Mostert et al. 2001), the Mitscherlich-exponential model (Rook et al., 1993; Vargas et al., 2000), the 4-parameter Morant model (Williams, 1993; Vargas et al., 2000), and both the lactation persistency and reduced lactation persistency models (Grossman et al., 1999; Vargas et al., 2000).

Often, in studies where multiple regression and least squares estimation were employed to fit lactation models, further analysis of the regression coefficients were performed by means of analysis of variance to study the effect of various environmental and other cofactors on the lactation model. This was done using a generalised linear model that combines both fixed effects, representing the cofactors at their various levels, and a random error in order to describe either a function of the estimates of the parameters of the model fitted or the actual observed data. Interaction effects were very seldom included in such a model, because these were mostly not significant and random effects were only found in models where genetic effects were included. Analysis of variance then indicated the presence of significant effects. (Batra, 1986; Grossman and Koops, 1988; Gipson and Grossman, 1989; Olori et al., 1999; Akpa et al., 2001)

### **3.4 BAYESIAN APPROACH TO LACTATION CURVE FITTING**

Although using a Bayesian approach in statistical research of animal genetics is a common practise, a Bayesian approach was applied in a very small proportion of research work done on lactation models thus far. The first attempt at modelling lactation by means of an approach that also forms part of the Bayesian framework, was that of Goodall and Sprevak (1984), where a Monte Carlo simulation procedure was employed to generate simulated values of milk yield in dairy cows. A year later in 1985 Goodall and Sprevak formally used the Kalman filter as a Bayesian estimation procedure that included prior information on the herd, as well as observed lactation records, to estimate the parameters of the lactation curve of dairy cows. In both these studies the lactation model proposed by Wood was fitted.

In 1996 Groenewald et al., to fit the Wood model to lactation data of Merino sheep, assumed a hierarchical Bayesian model and applied the Gibbs sampler to obtain the marginal posterior distributions of functions of parameters, including lactation curve characteristics such as peak yield, time of peak yield, persistency and total milk yield for the Wood model.

Jamrozik et al. (1997) required information on lactation test day records for the purpose of *genetic evaluation* of first lactation Holstein cows. For this purpose the Wilmink model described in (2.24) with  $w = 0,05$  and a random regression model with both fixed and random effects was used to analyse test day records of dairy cows. They wanted to then compare three different random regression models using combinations of the Wilmink model and a

linear function of five covariates suggested by Ali and Schaeffer (1987) as linear functions of covariates for fixed and random regression coefficients. Model variances and covariances were estimated by means of Bayesian methods utilising the Gibbs sampler to generate samples from the marginal posterior distribution.

In 1998 Varona et al. analysed test day records of Spanish dairy cows using the Wood model and a hierarchical Bayesian procedure. They specifically point out that the main advantage of this approach is that posterior marginal distributions of covariance components, *breeding values* and systematic effects for Wood's model are easily obtained and comment on the usefulness of this form of analysis in reducing the occurrence of type II or atypical lactation curves by restricting the values of parameter estimates in the model. Restricting the values of the parameter estimates are made possible by including an acceptance-rejection step in the Gibbs sampler. Special reference is also made to the "solid estimates of systematic effects and breeding values" obtained for the parameters of the lactation curve through the Bayesian procedure.

Rekaya et al. (2000) employed a three-stage hierarchical Bayesian approach, together with the Wood function, to model lactation in Holstein-Friesian cows. Restrictions were placed on two of the parameters in the Wood model, which required the use of the adaptive rejection Metropolis within Gibbs scheme. In this case the ultimate objective again was estimating *genetic correlation* between parameters of the Wood lactation curve and functions thereof such as total yield, persistency, and peak yield. The obtained estimates indicated that modifying the shape of the lactation curve through genetic selection is possible.

Chang et al. (2001) also used the Wood model and a three-stage hierarchical Bayesian approach, with Gibbs sampling and the Metropolis-Hastings algorithm to draw samples from the posterior distributions of the model parameters to model lactation in dairy sheep. As in preceding studies where the Bayesian approach was applied, the objective was the investigation of *genetic variation* of features of lactation curves. In 2002 Chang et al. revisited this approach to the investigation of *genetic variation* to compare results from the quadratic function of (2.14) proposed by Dave in 1971 and that of the Wood function.

Groenewald and Viljoen (2003) used a hierarchical Bayesian approach, together with Wood's model to model not only lactation trait curves for milk yield of Saanen dairy goats, but also for the protein and fat composition of milk. Marginal posterior distributions were again

obtained through Markov chain Monte Carlo methods. Various covariates and environmental factors were also built into the Bayesian model to enable the integrated study of the effects of these factors on lactation curves.

It is evident from the above discussion that very little work has been done on prediction of lactation records from a Bayesian point of view, and that the majority of the research in this field to date was done for the purpose of *genetic evaluation and selection*.

### 3.5 ASSUMPTION ABOUT ERRORS

When the general form of a function is  $e^{\mathbf{x}_i\boldsymbol{\beta}}$ , where  $\mathbf{x}_i$  ( $1 \times p$ ) is some vector function of time  $t$ , the estimation and inferences made dependent crucially on the assumptions made about the errors. Here are three possible approaches:

1) Often the model is fitted by non-linear least squares methods. This is basically non-parametric and implies no distributional assumptions. However, no statistical inferences (such as variance estimation, confidence intervals, etc.) made from the results are strictly valid. From a statistical point of view, such a method of fitting the model implies an additive error,

$$y_i = e^{\mathbf{x}_i\boldsymbol{\beta}} + \mathbf{e}_i \quad , \quad i = 1, \dots, n \quad (3.1)$$

where  $\mathbf{e}_i \sim i.i.d.N(0, \mathbf{S}^2)$ , which in turn implies that  $-\infty < y_i < \infty$ , while  $y_i$  is a strictly positive variable. This can lead to nonsensical inferences. The only way (3.1) can be used as a statistical model is if the error is bounded from below so that  $-e^{\mathbf{x}_i\boldsymbol{\beta}} < \mathbf{e}_i < \infty$ . This makes statistical inferences extremely complicated, as the sample space for the error term is a function of the unknown parameter, as with a truncated normal distribution.

2) The most common method of fitting the model is to assume a multiplicative error, so that

$$y_i = e^{\mathbf{x}_i\boldsymbol{\beta}} \mathbf{e}_i \quad , \quad i = 1, \dots, n \quad (3.2)$$

and

$$\ln y_i = \mathbf{x}_i\boldsymbol{\beta} + \ln \mathbf{e}_i \quad (3.3)$$

Then the least squares estimate of  $\mathbf{b}$  is also the maximum likelihood estimator if we assume

that  $\ln \mathbf{e}_i \sim i.i.d.N(0, \mathbf{s}^2)$ , i.e.  $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{z}$ , with  $X (n \times p)$  and  $\mathbf{z} = \ln(\mathbf{y})$ . This means that  $\mathbf{e}_i$  has a lognormal distribution with mean  $E[\mathbf{e}_i] = e^{\frac{1}{2}\mathbf{s}^2}$  and variance  $\text{Var}[\mathbf{e}_i] = e^{\mathbf{s}^2} (e^{\mathbf{s}^2} - 1)$ . This is a reasonable model and works well, but there are a couple of points that are often overlooked when analysing the error terms. If  $\mathbf{e}_i$  is estimated by

$$\hat{\mathbf{e}}_i = y_i e^{-\mathbf{x}_i \hat{\boldsymbol{\beta}}}, \quad i = 1, \dots, n \quad (3.4)$$

then  $E(\ln \hat{\mathbf{e}}_i) = E(\ln y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}) = 0$ , but researchers often use the log transformation to estimate  $\mathbf{b}$  and then define the error estimate as the direct deviation,

$$\hat{\mathbf{e}}_i = y_i - e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}}, \quad i = 1, \dots, n. \quad (3.5)$$

Since

$$E[Y_i] = e^{\mathbf{x}_i \boldsymbol{\beta}} E[\mathbf{e}_i] = e^{\mathbf{x}_i \boldsymbol{\beta} + \frac{1}{2}\mathbf{s}^2} \quad (3.6)$$

and

$$\begin{aligned} E[e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}}] &= \int_{-\infty}^{\infty} e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}} n(\hat{\boldsymbol{\beta}} : \mathbf{b}, (X'X)^{-1}\mathbf{s}^2) d\hat{\boldsymbol{\beta}} \\ &= e^{\mathbf{x}_i \boldsymbol{\beta} + \frac{1}{2}\mathbf{s}^2 \mathbf{x}_i (X'X)^{-1} \mathbf{x}_i'} \end{aligned}$$

it follows that

$$E[\hat{\mathbf{e}}_i] = e^{\mathbf{x}_i \boldsymbol{\beta} + \frac{1}{2}\mathbf{s}^2} \left( 1 - e^{-\frac{1}{2}\mathbf{s}^2 (\mathbf{x}_i (X'X)^{-1} \mathbf{x}_i' - 1)} \right). \quad (3.7)$$

Since  $\mathbf{x}_i (X'X)^{-1} \mathbf{x}_i' < 1$ , it means that  $E[\hat{\mathbf{e}}_i] > 0$  always. So  $e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}}$  is not an unbiased predictor of  $y_i$  and tends to underfit the true function  $e^{\mathbf{x}_i \boldsymbol{\beta}}$ . Yet  $\hat{\mathbf{e}}_i$  is often used to compare models and to determine goodness of fit.

The variance of  $\hat{\mathbf{e}}_i$  is given by

$$\text{Var}[\hat{\mathbf{e}}_i] = e^{2\mathbf{x}_i \boldsymbol{\beta}} \left[ e^{\mathbf{s}^2} (e^{\mathbf{s}^2} - 1) + e^{a_{ii}\mathbf{s}^2} (e^{a_{ii}\mathbf{s}^2} - 1) + 2e^{\frac{1}{2}\mathbf{s}^2 (a_{ii} + 1)} \right] - 2e^{\frac{1}{2}\mathbf{s}^2 (a_{ii} + 1)} \quad (3.8)$$

where  $a_{ii} = \mathbf{x}_i (X'X)^{-1} \mathbf{x}_i'$ .

Secondly, the expectation as well as the variance of  $\hat{\mathbf{e}}_i$  are monotonic functions of the function  $e^{\mathbf{x}_i \boldsymbol{\beta}}$  itself. So a study of the residuals  $\hat{\mathbf{e}}_i$  after fitting a linear model *should* show a pattern,

and that is not evidence of a lack of fit or autocorrelation between residuals. The residuals in (3.5) resulting from the fitting as in (3.3) should show discernable pattern

The estimator (3.4) of  $\mathbf{e}_i$  is also not an unbiased estimator of  $E[\mathbf{e}_i] = e^{\frac{1}{2}\mathbf{s}^2}$  since

$$\begin{aligned} E[\hat{\mathbf{e}}_i] &= E[y_i e^{-\mathbf{x}_i \hat{\boldsymbol{\beta}}}] \quad , \quad i = 1, \dots, n \\ &= E[e^{z_i} e^{-\mathbf{x}_i (X'X)^{-1} X'z}] \\ &= E[e^{-(a_{ii}-1)z_i}] \prod_{j=1}^n E[e^{-a_{ij}z_j}] \end{aligned}$$

where  $a_{ij} = \mathbf{x}_i (X'X)^{-1} \mathbf{x}_j'$  and  $z_j \sim N(\mathbf{x}_j \boldsymbol{\beta}, \mathbf{s}^2)$ .

Then

$$E[\hat{\mathbf{e}}_i] = e^{\frac{1}{2}\mathbf{s}^2(1-a_{ii})} \quad i = 1, \dots, n \quad (3.9)$$

which is smaller than  $E[\mathbf{e}_i]$ , since  $0 < a_{ii} < 1$ , and so tends to underestimate the true error.

3) In paragraph (2) the condition on the error is that  $P[\varepsilon_i > 1] = 0,5$ . Since the lognormal distribution is skewed, this condition is achieved when  $E[\mathbf{e}_i] = e^{-\frac{1}{2}\mathbf{s}^2}$ . Instead, we could impose the condition that  $E[\mathbf{e}_i] = 1$ ,  $i = 1, \dots, n$ . This means that we assume that  $\ln \mathbf{e}_i \sim N(-\frac{1}{2}\mathbf{s}^2, \mathbf{s}^2)$  so that  $\mathbf{e}_i \sim \ln N(1, e^{\mathbf{s}^2} - 1)$ . Then

$$z_i = \ln y_i \sim N(\mathbf{x}_i \boldsymbol{\beta} - \frac{1}{2}\mathbf{s}^2, \mathbf{s}^2). \quad (3.10)$$

Now  $y_i$  is an unbiased estimator of the function value  $e^{\mathbf{x}_i \boldsymbol{\beta}}$ , i.e.

$y_i \sim \ln N(e^{\mathbf{x}_i \boldsymbol{\beta}}, e^{2\mathbf{x}_i \boldsymbol{\beta}}(e^{\mathbf{s}^2} - 1))$ . The variances of  $y_1, \dots, y_n$  are not homogeneous. The maximum likelihood estimator of  $\mathbf{b}$  in (3.10) is given by

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1} X'(z + \frac{1}{2}\mathbf{s}^2 \mathbf{I}_n), \quad (3.11)$$

which only increases the constant term of the usual estimator,  $\hat{\boldsymbol{\beta}}_0 = (X'X)^{-1} X'z$ , by  $\frac{1}{2}\mathbf{s}^2$ .

Interestingly, the variance estimator remains the same as before,

$$\hat{\mathbf{s}}^2 = \frac{1}{n} (z - X\hat{\boldsymbol{\beta}}_0)'(z - X\hat{\boldsymbol{\beta}}_0). \quad (3.12)$$

So in paragraph (2) the assumption is that the median of the error distribution is equal to one, while in paragraph (3) the mean is assumed equal to one. When  $\mathbf{s}^2$  is relatively small, there is little difference between the fitted functions from the two approaches. From a convenience point of view the method of paragraph (2) is preferable.

The expectation of the estimated error is

$$E[\hat{\mathbf{e}}_i] = e^{-\frac{1}{2}\mathbf{s}^2 a_{ii}}, \quad (3.13)$$

which makes  $\hat{\mathbf{e}}_i$  again a biased estimator of  $E[\mathbf{e}_i]$ .

4) The Wilmlink model

$$y_t = a + bt + ce^{wt} = f(t) \quad (3.14)$$

is often used for the estimation of the lactation curves of cows. However, it has the flaw that a fitted curve can be negative. Least squares fitting implies a statistical model with additive errors that are normally distributed,

$$y_t = f(t) + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(0, \mathbf{s}^2). \quad (3.15)$$

Inferences from this model can often yield unrealistic results. The only way to use this model as a statistical model is to (a) assume that  $\ln y_t = f(t) + \mathbf{e}_t$ ,  $\mathbf{e}_t \sim N(0, \mathbf{s}^2)$ , so that we are back with the general form of paragraph 2 for  $w$  known, or (b) assume that  $\mathbf{e}_t$  has a truncated distribution, i.e.

$$\mathbf{e}_t \sim TN(0, \mathbf{s}^2), \quad -f(t) \leq \mathbf{e}_t \leq \infty \quad (3.16)$$

and  $f(t) \geq 0$ . Then

$$f(\mathbf{e}_t) = \frac{1}{f(f(t))\sqrt{2p\mathbf{s}^2}} e^{-\frac{1}{2\mathbf{s}^2}\mathbf{e}_t}, \quad -f(t) \leq \mathbf{e}_t \leq \infty, \quad (3.17)$$

but the expectation of  $\mathbf{e}_t$  is no longer zero.

In summary, the multiplicative model (3.2) is in general preferable to the additive model (3.1) for fitting lactation models for the following reasons:



1) As mentioned before, (3.1) implies that the production can be negative unless restrictions are placed on the distribution of the error term. This greatly complicates statistical inferences.

2) Usually when model (3.1) is employed the assumption of a common error variance is made, which means a constant variation in production  $y$  over the lactation period. For lactation data it is reasonable to assume that the error variance will decrease with production, due to both less variation in production and smaller measurement errors. The multiplicative model (3.2) naturally allows for this, as the variance in production is an increasing function of the expected production.

3) When the lactation model can be linearised by a log transformation, a log-normal distribution on the error term is appropriate, and normal theory is available for inferences.

It should be remembered that a maximum likelihood curve fitted after a log (or any other) transformation is not the best non-linear least squares fit in the sense of (3.5). This, however, is not a major drawback as the maximum likelihood estimate will in most cases be the more reasonable fit.

Care should also be taken when creating a linear model with a transformation other than a log transformation. For example, the inverse polynomial model (Nelder (1966)),

$$y_t = t/(a + bt + ct^2), \quad (3.18)$$

can be linearised as

$$1/y_t = b + a/t + ct, \quad (3.19)$$

but least squares estimation of the parameters can give nonsensical answers when applied to the original model, with the estimated lactation curve tending to infinity at the roots of the denominator. The problem is that if a multiplicative error is assumed for the original model, then

$$1/y_t = (b + a/t + ct) / \mathbf{e}_t, \quad (3.20)$$

and no distributional assumptions about  $\mathbf{e}_t$  can make maximum likelihood estimation feasible. If this model is to be used, non-linear least square estimation is the best option, but with the drawback that no statistical inference is possible.

# CHAPTER 4

## A BAYESIAN APPROACH

### 4.1 THE GENERALISED BAYESIAN MODEL

Any lactation model to which the assumption of multiplicative errors is applied and that may be re-written in linear form by means of a transformation in such a manner that, as a result, additive errors for the transformed linear form there-of holds, may be analysed in the manner as discussed below. Keeping in mind the discussion on the error assumptions in section 3.5, this basically implies that a log transformation of the original model (with multiplicative errors) is required to transform the model to linear (with resulting additive errors). Once a model has been obtained in generalised linear form, a hierarchical model similar to that introduced by Lindley and Smith (1972) is assumed and a Bayesian approach is used to model lactation. In general the methodology used is as follows:

Suppose that for every animal  $i$  in a herd of  $k$  animals the lactation observations are denoted by  $W_{ijs}(t_{ijp})$ , where  $j$  serves as an index of only those lactation cycles under consideration,  $s$  denotes the relevant lactation trait among those that were observed (e.g. milk yield, protein content, fat content, or lactose content),  $p$  refers to the index of the test day during the lactation cycle on which the observation was made, and  $t_{ijp}$  is the time of the  $p$ -th test day. Note that initially it is assumed that over a specified period the same number of consecutive lactation cycles has been observed for each animal in the herd. As a result,  $i = 1, \dots, k$ ,  $j = 1, \dots, q$  where for every animal in the herd a total of  $q$  lactation cycles are observed,  $s = 1, \dots, u$  where  $u$  is the number of lactation traits observed, and  $p = 1, \dots, n_{ij}$  where  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ . If a non-linear model with the assumption of multiplicative errors has to be fitted to the original data and by performing a natural log transformation on the data the result is:

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) \quad (4.1)$$

to which an observation model of generalised linear model form, with additive  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ , may now be fitted. If it is assumed that  $a_{ijs}$ ,  $b_{ijs}$ , ...,  $d_{ijs}$  are the  $v$  regression coefficients of the generalised linear form of the lactation model for the  $i$ -th animal

during its  $j$ -th lactation cycle for the lactation trait indicated by  $s$ , then the model for lactation trait  $s$  of animal  $i$  during lactation cycle  $j$  is:

$$\mathbf{y}_{ijs} = \mathbf{X}_{ij} \mathbf{m}_{ijs} + \mathbf{e}_{ijs} \quad (4.2)$$

where  $\mathbf{y}_{ijs}$  is the vector of transformed data values for the lactation trait  $s$  of animal  $i$  in its  $j$ -th observed lactation cycle,  $\mathbf{m}_{ijs}$  is the vector containing the  $v$  regression coefficients for the generalised linear form of the lactation model fitted for this lactation trait of the  $i$ -th animal in lactation cycle  $j$  and when multiplied with the design matrix  $\mathbf{X}_{ij}$  it returns the appropriate generalised model form of the fitted lactation model.

If this model is applied to milk yield and all of the composition traits contained in the data simultaneously, then the model for animal  $i$  in the  $j$ -th observed lactation cycle is:

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij} \mathbf{M}_{ij} + \mathbf{E}_{ij} \quad (4.3)$$

where  $\text{vec } \mathbf{E}'_{ij} (un_{ij} \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_{ij}})$ . Note that  $\text{vec } \mathbf{A}'$  denotes the elements of matrix  $\mathbf{A}$  when stacked row-wise into one column. The  $r$ -th row of  $\mathbf{X}_{ij} (n_{ij} \times v)$  contains the elements of the  $r$ -th row of a design matrix that would return the appropriate generalised linear model when multiplied with the matrix of coefficients  $\mathbf{M}_{ij}$ , where  $v$  denotes the number of coefficients that every lactation trait in the generalised linear model contains. The matrix of coefficients has  $u$  columns, each containing the coefficients of the generalised linear model for one of the lactation traits, milk yield, protein contents and fat content etc., therefore:

$$\mathbf{M}_{ij} (v \times u) = \begin{bmatrix} a_{ij1} & a_{ij2} & \cdots & a_{iju} \\ b_{ij1} & b_{ij2} & \cdots & b_{iju} \\ \vdots & \vdots & \cdots & \vdots \\ d_{ij1} & d_{ij2} & \cdots & d_{iju} \end{bmatrix}, \quad (4.4)$$

$$\mathbf{s}_s^2 (1 \times u) = [\sigma_1^2 \quad \sigma_2^2 \quad \cdots \quad \sigma_u^2] \quad (4.5)$$

and

$$\mathbf{F} (u \times u) = \text{diag}\{\mathbf{s}_s^2\}. \quad (4.6)$$

To further generalise this model, all of the  $q$  consecutive lactation cycles of animal  $i$  as indexed by  $j$  are included in the model. As a result for animal  $i$  over all  $q$  of its observed lactation cycles, the model becomes:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_{i1} \\ \vdots \\ \mathbf{Y}_{iq} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (4.7)$$

where  $\mathbf{Y}_i (n_i \times u)$ ,  $\text{vec } \mathbf{E}'_i (un_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^q n_{ij}$  the number of test days for animal  $i$  during all  $q$  lactation cycles observed,  $\mathbf{X}_i (n_i \times vq) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i (vq \times u) = \begin{bmatrix} \mathbf{M}_{i1} \\ \vdots \\ \mathbf{M}_{iq} \end{bmatrix}. \quad (4.8)$$

The model can similarly be extended to include all records observed for all animals in the herd, resulting in the following complete model:

$$\mathbf{Y} = \mathbf{X}\mathbf{M} + \mathbf{E} \quad (4.9)$$

where  $\mathbf{Y} (n \times u)$ ,  $\text{vec } \mathbf{E}' (un \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_n)$   $n = \sum_{i=1}^k n_i$  the total number of test day over all  $k$  animals in the herd during all observed lactation cycles,  $\mathbf{X} (n \times vqk) = \text{diag}\{\mathbf{X}_i\}$  and

$$\mathbf{M} (vqk \times u) = \begin{bmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_k \end{bmatrix}. \quad (4.10)$$

Often the lactation data available for analysis purposes do not only include information on the milk yield and composition traits of production on each test day, but also information on other factors such as parturition date, parity number, season of parturition, age of the mother and region in which the herd is located. From the discussion in section 1.5 other researchers have found that some these factors could significantly influence lactation. As a result it is therefore necessary to make provision for this information in the formulation of a lactation model. To include the factors considered to be possible significant covariates, it was decided to include a covariate matrix as part of the prior information of the model. Therefore, let  $\mathbf{Z}_i$  denote the  $(m \times q)$  matrix of  $m$  covariates for animal  $i$ .

Because the interests of lactation research focuses on the lactation curves and lactation characteristics of the individual animals within the herd, this analysis is conducted with respect to the model of animal  $i$ , in other words by considering the model in the form  $\mathbf{Y}_i = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i$ . For analysis purposes the vector form of this model and the covariates

applicable to the data is considered, with  $\mathbf{z}_i = \text{vec } \mathbf{Z}_i$  denoting the  $(mq \times 1)$  column vector formed by arranging the columns of  $\mathbf{Z}_i$  into one long column. Similarly for  $\mathbf{m}_{ij}(uv \times 1) = \text{vec } \mathbf{M}_{ij}$  which then gives the parameters of the model for animal  $i$  during lactation cycle  $j$  as one column, with the first set of  $v$  rows containing the parameters of the first trait, the second set of  $v$  rows that of the second trait and so forth up until set  $u$ . As a result of earlier work by Lindley and Smith (1972), Groenewald et al. (1996), Chang et al. (2001), and Groenewald and Viljoen (2003), the prior distribution of  $\mathbf{m}_{ij}$  is assumed to be

$$\mathbf{m}_{ij} | \mathbf{B}, \mathbf{S} \sim N(\mathbf{B}\mathbf{z}_{ij}, \mathbf{S}) \quad (4.11)$$

where  $\mathbf{B}$  ( $uv \times m$ ) is a matrix of regression coefficients and  $\mathbf{z}_{ij}(m \times 1)$  is the  $j$ -th column of  $\mathbf{Z}_i$ .

Further, if  $\mathbf{m}_i(uvq \times 1) = \left[ \mathbf{m}'_{i1} \cdots \mathbf{m}'_{iq} \right]'$  represents the column vector of the parameters of animal  $i$  over all the lactation cycles considered, then

$$\mathbf{m}_i | \mathbf{B}, \mathbf{S}, \mathbf{R} \sim N\left((\mathbf{I}_q \otimes \mathbf{B})\mathbf{z}_i, \mathbf{R} \otimes \mathbf{S}\right). \quad (4.12)$$

The  $q \times q$  matrix  $\mathbf{R} = \{\rho_{jr}\}$  represents the covariances between the model parameters of the same animal in successive lactation cycles when  $q > 1$ .

The prior distributions on the other parameters are as follows:

$$\mathbf{p}(\mathbf{B}) \propto 1, \quad (4.13)$$

$$\pi(\mathbf{F}^{-1}) \propto \left( \prod_{s=1}^u \sigma_s^2 \right)^{-1}, \quad (4.14)$$

$$\mathbf{R}^{-1} \sim W_q\left((\delta\mathbf{P})^{-1}, \delta\right) \quad (4.15)$$

$$\text{and } \mathbf{S}^{-1} \sim W_{uv}\left((g\mathbf{G})^{-1}, g\right) \quad (4.16)$$

where  $\mathbf{d}$ ,  $\mathbf{P}$ ,  $g$  and  $\mathbf{G}$  are assumed known. Priors (4.13) and (4.14) are the standard Jeffreys priors for the location and scale parameters, while priors (4.15) and (4.16) have to be proper, but noninformative. The choice of their parameters will be discussed in Section 5.2.

The full conditional distributions of all the model parameters of interest from which the marginal posterior distributions (conditional on the observed data only) will be obtained by means of Gibbs sampling procedure, are derived as follows:

Suppose  $\mathbf{y}_i = \text{vec } \mathbf{Y}_i \zeta$  then, for  $i = 1, \dots, k$  :

$$\begin{aligned}
& f(\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto f(\mathbf{y}_i | \mathbf{m}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{m}_i | \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto \exp\left[-\frac{1}{2}(\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i) \mathbf{m}_i)' (\mathbf{F} \otimes \mathbf{I}_{n_i})^{-1} (\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i) \mathbf{m}_i)\right] \\
& \quad \cdot \exp\left[-\frac{1}{2}(\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i)' (\mathbf{R} \otimes \mathbf{S})^{-1} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i)\right] \\
& \propto \exp\left[-\frac{1}{2} \left\{ \mathbf{m}_i' (\mathbf{I}_u \otimes \mathbf{X}_i') (\mathbf{F}^{-1} \otimes \mathbf{I}_{n_i}) (\mathbf{I}_u \otimes \mathbf{X}_i) \mathbf{m}_i - 2 \mathbf{m}_i' (\mathbf{I}_u \otimes \mathbf{X}_i') (\mathbf{F}^{-1} \otimes \mathbf{I}_{n_i}) \mathbf{y}_i \right. \right. \\
& \quad \left. \left. + \mathbf{m}_i' (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \mathbf{m}_i - 2 \mathbf{m}_i' (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i \right\} \right] \\
& \propto \exp\left[-\frac{1}{2} \left\{ \mathbf{m}_i' \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right] \mathbf{m}_i \right\} - 2 \mathbf{m}_i' \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right] \right]
\end{aligned}$$

Completing the square:

$$\begin{aligned}
& f(\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto \exp\left[-\frac{1}{2} \left\{ \left( \mathbf{m}_i - \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right]^{-1} \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right] \right)' \right. \right. \\
& \quad \left. \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right] \right. \\
& \quad \left. \left( \mathbf{m}_i - \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right]^{-1} \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right] \right)' \right. \\
& \quad \left. - \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right]' \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right]^{-1} \right. \\
& \quad \left. \left. \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right] \right\} \right]
\end{aligned}$$

As a result the conditional distribution of  $\mathbf{m}_i$  is:

$$\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R} \sim N \left( \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right]^{-1} \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i') \mathbf{y}_i + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i \right], \left[ (\mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i) + (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \right]^{-1} \right)$$

$$\begin{aligned}
& f(\mathbf{B} \mid \mathbf{Y}, \mathbf{M}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(y_i \mid \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{M}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{B} \mid \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(\mathbf{m}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{B}) \\
& \propto \prod_{i=1}^k \exp \left[ -\frac{1}{2} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i)' (\mathbf{R} \otimes \mathbf{S})^{-1} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i) \right] \cdot 1 \\
& \propto \prod_{i=1}^k \exp \left[ -\frac{1}{2} \left\{ \mathbf{m}_i' (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) \mathbf{m}_i - 2 \mathbf{m}_i' (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i \right. \right. \\
& \quad \left. \left. + \mathbf{z}_i' (\mathbf{I}_q \otimes \mathbf{B}') (\mathbf{R}^{-1} \otimes \mathbf{S}^{-1}) (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i \right\} \right] \\
& \propto \prod_{i=1}^k \exp \left[ -\frac{1}{2} \left\{ \mathbf{z}_i' (\mathbf{R}^{-1} \otimes \mathbf{B}' \mathbf{S}^{-1} \mathbf{B}) \mathbf{z}_i - 2 \mathbf{z}_i' (\mathbf{R}^{-1} \otimes \mathbf{B}' \mathbf{S}^{-1}) \mathbf{m}_i \right\} \right]
\end{aligned}$$

where  $\mathbf{z}_i (mq \times 1) = \text{vec } \mathbf{Z}_i$  is the column vector formed by arranging the columns of  $\mathbf{Z}_i (m \times q)$  into one long column. Similarly  $\mathbf{m}_{ij} (uv \times 1) = \text{vec } \mathbf{M}_{ij}$ , where  $\mathbf{M}_{ij} (v \times u)$ . Defining  $\tilde{\mathbf{M}}_i (uv \times q) = [\mathbf{m}_{i1} \ \cdots \ \mathbf{m}_{iq}]$  and reverting back to  $\mathbf{Z}_i (m \times q)$ :

$$\begin{aligned}
& f(\mathbf{B} \mid \mathbf{Y}, \mathbf{M}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \\
& \propto \exp \left[ -\frac{1}{2} \text{tr} \left\{ \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \mathbf{B}' \mathbf{S}^{-1} \mathbf{B} - 2 \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \tilde{\mathbf{M}}_i' \mathbf{S}^{-1} \mathbf{B} \right\} \right] \\
& \propto \exp \left[ -\frac{1}{2} \text{tr} \left\{ \mathbf{B}' \mathbf{S}^{-1} \mathbf{B} \sum_{i=1}^k (\mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i') - 2 \mathbf{B}' \mathbf{S}^{-1} \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right\} \right] \\
& \propto \exp \left[ -\frac{1}{2} \text{tr} \left\{ \left( \mathbf{B} - \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}^{-1} \mathbf{Z}_i' \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1} \right)' \mathbf{S}^{-1} \right. \right. \\
& \quad \left. \left. \left( \mathbf{B} - \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}^{-1} \mathbf{Z}_i' \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1} \right) \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right) \right\} \right]
\end{aligned}$$

So that:

$$\mathbf{B} \mid \mathbf{M}_i, \mathbf{S}, \mathbf{R} \sim N \left( \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}^{-1} \mathbf{Z}_i' \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1}, \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1} \otimes \mathbf{S} \right).$$

Also

$$\begin{aligned}
& f(\mathbf{F}^{-1} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{S}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(\mathbf{y}_i \mid \mathbf{M}, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{F}^{-1} \mid \mathbf{M}_i, \mathbf{B}, \mathbf{S}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(\mathbf{y}_i \mid \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{F}^{-1}) \\
& \propto \prod_{i=1}^k \left| \mathbf{F} \otimes \mathbf{I}_{n_i} \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i) \mathbf{m}_i)' (\mathbf{F} \otimes \mathbf{I}_{n_i})^{-1} (\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i) \mathbf{m}_i) \right] \cdot \frac{1}{\sigma_1^2 \cdot \sigma_2^2 \dots \sigma_s^2} \\
& \propto \left| \mathbf{I}_{n_i} \right|^{-\frac{s}{2}} \left| \mathbf{F} \right|^{\frac{n}{2}} \text{etr} \left[ -\frac{1}{2} \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' \mathbf{F}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right] \cdot \frac{1}{\sigma_1^2 \cdot \sigma_2^2 \dots \sigma_s^2} \\
& \propto (\sigma_1^2 \cdot \sigma_2^2 \dots \sigma_s^2)^{-\left(\frac{n+1}{2}\right)} \text{etr} \left[ \mathbf{F}^{-1} \left\{ -\frac{1}{2} \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right\} \right]
\end{aligned}$$

where,  $\mathbf{F} = \text{diag}\{\mathbf{s}_1^2, \mathbf{s}_2^2, \dots, \mathbf{s}_s^2\}$  so that:

$$\mathbf{F} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{S}, \mathbf{R} \sim IW_s \left( \frac{1}{2} \text{diag} \left\{ \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right\}, \frac{n}{2} \right)$$

and therefore,

$$\mathbf{F}^{-1} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{S}, \mathbf{R} \sim W_s \left( \left( \frac{1}{2} \text{diag} \left\{ \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right\} \right)^{-1}, \frac{n}{2} \right).$$

Similarly,

$$\begin{aligned}
& f(\mathbf{S}^{-1} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{F}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(\mathbf{y}_i \mid \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{M}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{S} \mid \mathbf{B}, \mathbf{F}, \mathbf{R}) \\
& \propto \prod_{i=1}^k f(\mathbf{M}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{S}) \\
& \propto \prod_{i=1}^k \left| \mathbf{R} \otimes \mathbf{S} \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i)' (\mathbf{R} \otimes \mathbf{S})^{-1} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i) \right] \\
& \quad \cdot \text{etr} \left[ -\frac{1}{2} (g\mathbf{G}) \mathbf{S}^{-1} \right] \left| \mathbf{S} \right|^{-\frac{1}{2}(g-u\nu-1)}
\end{aligned}$$



$$\begin{aligned}
&\propto |\mathbf{R}|^{-\frac{uvk}{2}} |\mathbf{S}|^{-\frac{kg}{2}} \text{etr} \left[ -\frac{1}{2} \sum_{i=1}^k \mathbf{R}^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i)' \mathbf{S}^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i) \right] \\
&\quad \cdot \text{etr} \left[ -\frac{1}{2} (g\mathbf{G})\mathbf{S}^{-1} \right] |\mathbf{S}|^{-\frac{1}{2}(g-uv-1)} \\
&\propto |\mathbf{S}|^{-\frac{1}{2}(kq+g-uv-1)} \text{etr} \left[ \mathbf{S}^{-1} \left( -\frac{1}{2} \left\{ \sum_{i=1}^k (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i) \mathbf{R}^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i)' + g\mathbf{G} \right\} \right) \right]
\end{aligned}$$

so that,

$$\mathbf{S}^{-1} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{F}, \mathbf{R} \sim W_{uv} \left( \left\{ \sum_{i=1}^k (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i) \mathbf{R}^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B}\mathbf{Z}_i)' + g\mathbf{G} \right\}^{-1}, kq+g \right).$$

And finally,

$$\begin{aligned}
&f(\mathbf{R}^{-1} \mid \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{F}, \mathbf{S}) \\
&\propto \prod_{i=1}^k f(\mathbf{y}_i \mid \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{M}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{R} \mid \mathbf{B}, \mathbf{F}, \mathbf{S}) \\
&\propto \prod_{i=1}^k f(\mathbf{M}_i \mid \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{R}) \\
&\propto \prod_{i=1}^k |\mathbf{R} \otimes \mathbf{S}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B})\mathbf{z}_i)' (\mathbf{R} \otimes \mathbf{S})^{-1} (\mathbf{m}_i - (\mathbf{I}_q \otimes \mathbf{B})\mathbf{z}_i) \right] \\
&\quad \cdot \text{etr} \left[ -\frac{1}{2} (d\mathbf{P})\mathbf{R}^{-1} \right] |\mathbf{R}|^{-\frac{1}{2}(d-q-1)} \\
&\propto |\mathbf{R}|^{-\frac{uvk}{2}} |\mathbf{S}|^{-\frac{kg}{2}} \text{etr} \left[ -\frac{1}{2} \sum_{i=1}^k \mathbf{R}^{-1} (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i)' \mathbf{S}^{-1} (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i) \right] \\
&\quad \cdot \text{etr} \left[ -\frac{1}{2} (d\mathbf{P})\mathbf{R}^{-1} \right] |\mathbf{R}|^{-\frac{1}{2}(d-q-1)} \\
&\propto |\mathbf{R}|^{-\frac{1}{2}(uvk+d-q-1)} \text{etr} \left[ \mathbf{R}^{-1} \left( -\frac{1}{2} \left\{ \sum_{i=1}^k (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i)' \mathbf{S}^{-1} (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i) + d\mathbf{P} \right\} \right) \right]
\end{aligned}$$

so that,

$$\mathbf{R}^{-1} \mid \mathbf{y}_i, \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S} \sim W_q \left( \left\{ \sum_{i=1}^k (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i)' \mathbf{S}^{-1} (\mathbf{M}_i - \mathbf{B}\mathbf{Z}_i) + d\mathbf{P} \right\}^{-1}, uvk+d \right).$$

To summarise:

$$\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R} \sim N \left( \left\{ \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \right) \mathbf{y}_i + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B} \right) \mathbf{z}_i \right] \right\}, \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \right), \quad i = 1, \dots, k \quad (4.17)$$

$$\mathbf{B} | \mathbf{M}, \mathbf{S}, \mathbf{R} \sim N \left( \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}^{-1} \mathbf{Z}_i' \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1}, \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}^{-1} \mathbf{Z}_i' \right)^{-1} \otimes \mathbf{S} \right) \quad (4.18)$$

$$\mathbf{F}^{-1} | \mathbf{Y}, \mathbf{M} \sim \mathbf{W}_s \left( \left( \frac{1}{2} \text{diag} \left\{ \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right\} \right)^{-1}, \frac{n}{2} \right) \quad (4.19)$$

$$\mathbf{S}^{-1} | \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{R} \sim \mathbf{W}_{uv} \left( \left\{ \sum_{i=1}^k (\tilde{\mathbf{M}}_i - \mathbf{B} \mathbf{Z}_i) \mathbf{R}^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B} \mathbf{Z}_i)' + g \mathbf{G} \right\}^{-1}, kq + g \right) \quad (4.20)$$

$$\mathbf{R}^{-1} | \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{S} \sim \mathbf{W}_q \left( \left\{ \sum_{i=1}^k (\mathbf{M}_i - \mathbf{B} \mathbf{Z}_i)' \mathbf{S}^{-1} (\mathbf{M}_i - \mathbf{B} \mathbf{Z}_i) + d \mathbf{P} \right\}^{-1}, uvk + d \right) \quad (4.21)$$

Estimation of the model parameters for an individual animal follows from (4.17), while predicting the model parameters of one or more of the  $q$  lactation cycles under consideration for an individual animal  $i$ , or  $\mathbf{m}_{if}^{(2)}$ , given the performance in the  $h < q$  previous lactation cycle, for which the parameters are  $\mathbf{m}_i^{(1)}$ , may now be derived as follows. From (4.12) it is known that:

$$\mathbf{m}_i | \mathbf{B}, \mathbf{S}, \mathbf{R} \sim N \left( (\mathbf{I}_q \otimes \mathbf{B}) \mathbf{z}_i, \mathbf{R} \otimes \mathbf{S} \right).$$

Therefore the  $\mathbf{m}_i$  may be partitioned as follows

$$\mathbf{m}_i = \begin{pmatrix} \mathbf{m}_i^{(1)} \\ \mathbf{m}_{if}^{(2)} \end{pmatrix} \begin{matrix} (uvh \times 1) \\ (uv(q-h) \times 1) \end{matrix} \quad \text{where } h = 1, \dots, (q-1),$$

so that  $h$  represents the number of previous lactation cycles considered in the prediction of the remaining  $q - h$  lactation cycles and

$$(\mathbf{I}_q \otimes \mathbf{B})\mathbf{z}_i = \begin{pmatrix} (\mathbf{I}_h \otimes \mathbf{B})\mathbf{z}_i^{(1)} \\ (\mathbf{I}_{q-h} \otimes \mathbf{B})\mathbf{z}_i^{(2)} \end{pmatrix},$$

$$\mathbf{R} \otimes \mathbf{S} = \begin{pmatrix} \mathbf{R}_{11}^{(h)} \otimes \mathbf{S} & \mathbf{R}_{12}^{(h)} \otimes \mathbf{S} \\ \mathbf{R}_{21}^{(h)} \otimes \mathbf{S} & \mathbf{R}_{22}^{(h)} \otimes \mathbf{S} \end{pmatrix} \begin{pmatrix} (uvh) \\ (uv(q-h)) \end{pmatrix}$$

Note that if:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \sim N(\boldsymbol{\mu}, \mathbf{S}) \text{ then}$$

$$\mathbf{X}^{(2)} | \mathbf{X}^{(1)} = \mathbf{x}^{(1)} \sim N(\boldsymbol{\mu}^{(2)} + \mathbf{S}_{21}\mathbf{S}_{11}^{-1}(\mathbf{x}^{(1)} - \boldsymbol{\mu}^{(1)}), \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}).$$

Therefore

$$\begin{aligned} \mathbf{m}_{if}^{(2)} | \mathbf{m}_i^{(1)} = \mathbf{m}_i^{(1)*}, \mathbf{B}, \mathbf{S}, \mathbf{R} \\ &\sim N\left(\left(\mathbf{I}_{q-h} \otimes \mathbf{B}\right)\mathbf{z}_i^{(2)} + \left(\mathbf{R}_{21}^{(h)} \otimes \mathbf{S}\right)\left(\mathbf{R}_{11}^{(h)} \otimes \mathbf{S}\right)^{-1}\left(\mathbf{m}_i^{(1)*} - \left(\mathbf{I}_h \otimes \mathbf{B}\right)\mathbf{z}_i^{(1)}\right), \right. \\ &\quad \left. \left(\mathbf{R}_{22}^{(h)} \otimes \mathbf{S}\right) - \left(\mathbf{R}_{21}^{(h)} \otimes \mathbf{S}\right)\left(\mathbf{R}_{11}^{(h)} \otimes \mathbf{S}\right)^{-1}\left(\mathbf{R}_{12}^{(h)} \otimes \mathbf{S}\right)\right) \\ &\sim N\left(\left(\mathbf{I}_{q-h} \otimes \mathbf{B}\right)\mathbf{z}_i^{(2)} + \left(\mathbf{R}_{21}^{(h)}\mathbf{R}_{11}^{(h)-1} \otimes \mathbf{I}_{uv}\right)\left(\mathbf{m}_i^{(1)*} - \left(\mathbf{I}_h \otimes \mathbf{B}\right)\mathbf{z}_i^{(1)}\right), \right. \\ &\quad \left. \left(\mathbf{R}_{22}^{(h)} \otimes \mathbf{S}\right) - \left(\mathbf{R}_{21}^{(h)}\mathbf{R}_{11}^{(h)-1}\mathbf{R}_{12}^{(h)} \otimes \mathbf{S}\right)\right) \\ &\sim N\left(\left(\mathbf{I}_{q-h} \otimes \mathbf{B}\right)\mathbf{z}_i^{(2)} + \left(\mathbf{R}_{21}^{(h)}\mathbf{R}_{11}^{(h)-1} \otimes \mathbf{I}_{uv}\right)\left(\mathbf{m}_i^{(1)*} - \left(\mathbf{I}_h \otimes \mathbf{B}\right)\mathbf{z}_i^{(1)}\right), \left(\mathbf{R}_{22,1}^{(h)} \otimes \mathbf{S}\right)\right) \end{aligned} \quad (4.22)$$

In the Markov chain Monte Carlo simulation  $\mathbf{m}_i^{(1)*}$  is drawn from  $\mathbf{m}_i^{(1)} | \mathbf{y}_i^{(1)}, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}$  in (4.17) with  $\mathbf{X}_i^{(1)}$ ,  $\mathbf{y}_i^{(1)}$  and  $\mathbf{z}_i^{(1)}$  as applicable to the lactation cycles assumed known. Then  $\mathbf{m}_{if}^{(2)*}$  is drawn from (4.22).

Once  $\mathbf{m}_{if}^{(2)} = \mathbf{m}_{if}^{(2)*}$  is drawn, the predictive density follows from the fact that

$$\mathbf{Y}_i | \mathbf{M}_i, \mathbf{F} \sim N(\mathbf{X}_i \mathbf{M}_i, \mathbf{F} \otimes \mathbf{I}_{n_i}).$$

So the predictive distribution of  $(q-j)$  future lactation cycles becomes:

$$\mathbf{y}_{if}^{(2)} | \mathbf{m}_{if}^{(2)} = \mathbf{m}_{if}^{(2)*}, \mathbf{F} \sim N\left(\left(\mathbf{I}_u \otimes \mathbf{X}_{if}^{(2)}\right)\mathbf{m}_{if}^{(2)*}, \mathbf{F} \otimes \mathbf{I}_{n_{if}(q-h)}\right) \quad (4.23)$$

where  $\mathbf{y}_{if}^{(2)}$  is of order  $(un_{if}(q-h) \times 1)$ .

## 4.2 THE GIBBS SAMPLER

The Gibbs sampler is a highly efficient, computer intensive sampling algorithm that provides a technique to indirectly generate random variables in the absence of the density using elementary Markov chain properties as base. A sample,  $Y^{(1)}, \dots, Y^{(n)}$ , with the same distributional form as that of the unknown density is generated provided that all the full conditional distributions are known and available. When this sample simulated through Gibbs sampling becomes large enough, i.e. when  $n$  becomes large enough, any of the characteristics of the marginal distribution, such as the mean, variance and others, may be calculated with great accuracy. It is, however, important to note that although the method generates a sample, the end result of calculations based on such a sample actually yield population values. In 1984 Geman and Geman showed that the distribution of such a sample of  $n$  values generated by means of the Gibbs sampler under reasonably general conditions, converges to the marginal distribution of the variable for which the sampling was performed as  $n \rightarrow \infty$ . In other words, if we for example consider the mean of  $f(y)$  from a simulated sample  $Y^{(1)}, \dots, Y^{(n)}$  obtained by means of the Gibbs sampler then:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y^{(i)} = \int_{-\infty}^{\infty} y f(y) dy = E(Y).$$

If the sample size is large enough the characteristics obtained from it would be that of the population itself. In this manner even the actual density may be determined.

To explain the Gibbs sampling procedure, consider the two variables  $U$  and  $V$  with their respective conditional distributions  $f(u|v)$  and  $f(v|u)$ . Through Gibbs sampling the following sequence of random variables is generated:  $U^{(0)}, V^{(0)}, U^{(1)}, V^{(1)}, U^{(2)}, V^{(2)}, \dots, U^{(n)}, V^{(n)}$ . This is done by first specifying a starting value for one of the variable, say  $U^{(0)} = u^{(0)}$ , and then iteratively sampling from the two full conditional distributions alternatively as follows until  $n$  sets of random variables have been obtained to form the required sequence of random variables:

$$V^{(i)} \sim f(v|U^{(i)} = u^{(i)}) \quad \text{and} \quad U^{(i+1)} \sim f(u|V^{(i)} = v^{(i)})$$

Casella and George (1992) noted that the rate of convergence (in distribution) of the sequence generated determines the so-called “efficiency of the Gibbs sampler”. This convergence may be handled in a number of different ways in order to find an approximate sample from the marginal distributions of interest. One method would be to generate  $k$  samples of size  $n$  by means of the Gibbs sampler and then using the  $n^{\text{th}}$  simulated value from each of the  $k$  samples,

i.e.  $Y_1^{(n)}, Y_2^{(n)}, \dots, Y_k^{(n)}$ , which will then, on condition that  $n$  is sufficiently large, form an independent sample from the marginal distribution,  $f(y)$ . If  $k$  is chosen to be large enough,  $f(y)$  itself or any characteristic there-of may be calculated to the desired degree of accuracy (Gelfand and Smith, 1990). Gelfand, Hills, Racine-Poon and Smith (1990) noted the appropriate values of  $n$  and  $k$  required for convergence would depend on the application considered at the time and that these values cannot be pre-specified unilaterally. They also suggested checking for convergence graphically. This is done by simulating a large number of values through Gibbs sampling and then plotting the data generated for each of the random variables univariately to see when the simulated values seem to stabilise. Convergence of the Gibbs sampler was also considered by Wang, Rutledge and Gianola (1993), Roberts and Polson (1994), Zellner and Min (1995) and many others.

In 1991 Geyer suggested that instead of sampling the  $n^{\text{th}}$  or last simulated value from each of  $k$  independently generated sequences, one long sequence may be generated and every  $r^{\text{th}}$  value from such a sequence may then be extracted to form, on condition that  $r$  is large enough, an independent sample from the marginal distribution,  $f(y)$ . MacEachern and Berliner (1994) described another popular approach in using the Gibbs sampler to generate a marginal distribution. A long sequence of dependent variables are generated by means of the Gibbs sampler and then the sets variables generated during the first few iterations are discarded as the so-called “burn-in” period of the process. The result of the remaining iterations is then used, without any further subsampling, as the marginal distributions for the variables under consideration.

As mentioned earlier, the Gibbs sampler is used to obtain the posterior distributions of the model parameters from the full conditional distributions found in equations (4.17) to (4.21). To apply the Gibbs sampler to equations (4.17) to (4.21), starting values in the form of the least squares estimators were specified for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , and then the observations  $\mathbf{m}_i$ ,  $i = 1, \dots, k$ , were generated from (4.17). These values of  $\mathbf{M}$ , together with the starting values of the other parameters, were then used in turn in (4.18) to (4.21) and then in (4.17) to generate the second set of observations in the sequence. The result of this second set of observations in the sequence were then again used in another iteration of the process to generate the third set of observation in the sequence, and so forth. This is repeated  $r$  times, after which the results of first  $l$  iterations were discarded as “burn-in” period, from which it then follows that in the case of  $\mathbf{B}$ , for example:

$$f(\mathbf{B} | \mathbf{Y}) \approx \frac{1}{(r-1)} \sum_{i=t+1}^r f_i(\mathbf{B} | \mathbf{Y}, \mathbf{M}, \mathbf{F}, \mathbf{S}, \mathbf{R})$$

The posterior distributions of any function of these parameters, such as those of the characteristics of the production curve itself may now be obtained very easily.

### 4.3 THE CHARACTERISTICS OF THE LACTATION CURVE

In the case of the Wood model of (2.7) as described in Chapter 2, certain lactation curve characteristics have been written in terms of functions of model parameters. These characteristics of the lactation curve of interest are:

total milk yield (kg),

$$\mathbf{f} = \int_0^{300} t^b \exp(a + ct) dt, \quad (4.24)$$

peak milk yield (kg),

$$\mathbf{h} = \left(-\frac{b}{c}\right)^b \exp(a - b), \quad (4.25)$$

time of peak yield (days),

$$\mathbf{t} = -\frac{b}{c}, \quad (4.26)$$

and persistency of production (Wood, 1968),

$$\mathbf{y} = -(b + 1) \ln(-c). \quad (4.27)$$

For the above quantities the conditional distributions are unspecified, but because they are functions of model parameters of which the posterior distributions are easily obtained through the use of the Gibbs sampler, they are still easily obtainable. Note that the above quantities as given in (4.24) to (4.27) are of very little value when considering the composition trait curves such as that of fat and protein, and are therefore only considered in term of milk production. As a result, the reduced model only containing the parameters relevant to milk yield have to be considered.

The marginal posterior densities of the mean of each of these characteristics are obtained as follows. For each set of simulated values of  $\mathbf{m}_{ij}$  values,  $i = 1, \dots, k$  and  $j = 1, \dots, q$  for milk yield only as obtained through Gibbs sampling, the value of the quantity under consideration is calculated. A histogram for the values of this quantity is then constructed. After

standardising such a histogram to make the area under the histogram equal to one, it is smoothed by means of Pearson curve fitting. Note that the system of Pearson curves makes use of the first four moments about the mean to describe the shape of the distribution according to its location (mean), dispersion (variance), skewness and kurtosis, and then fits the resulting curve based on these descriptive measures. The resulting curve provides a good approximation to the posterior of the expected value of the characteristic under consideration.

#### 4.4 HANDLING INCOMPLETE DATA

Initially it was assumed that over a specified period a total of  $q$  consecutive lactation cycles were observed for each animal in the herd. This is, however, often not the case, resulting in what could be considered as incomplete lactation cycle data. In such a case it might happen that, over the same specified period, test day data might be recorded for some animals over all  $q$  lactation cycles under consideration, while for other animals test day data on only some of these lactation cycles are available.

To illustrate the simplest possible case where this could occur, assume that lactation data over a period that would include at most two consecutive lactation cycles are available. Included in such a data set are three possible recording schemes of lactation cycle data; 1) lactation data for animals that were recorded during the first of the two possible consecutive lactation cycles only; 2) lactation data for animals that were recorded only during the second of the two possible consecutive lactation cycles, and 3) lactation data for animals that were recorded during both the consecutive lactation cycles. Therefore, instead of using as notation  $q$  lactation cycles for all animals in the data set,  $q_i = 1, 2$  now denotes the number of lactation cycles under consideration for animal  $i$ .

As a result the matrix  $\mathbf{R}$ , which represents the covariances between the model parameters of the same animal in successive lactation cycles only when  $q > 1$ , is affected by this recording scheme. All other parameters and their full conditional distributions are only affected with respect to dimension as  $q$  is now replaced by  $q_i$ , while  $q = \sum_{i=1}^k q_i$  are the total number of lactation cycles over all animals contained in the data. This affects the degrees of freedom of

the distribution of  $\mathbf{S}^{-1}$  so that it becomes  $q + g$ , to still reflect the number of lactation cycles in the data set plus the degrees of freedom reflected by  $g$ .

The full conditional distributions therefore are:

$$\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}_i \sim N \left( \left\{ \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}_i^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \right) \mathbf{y}_i + \left( \mathbf{R}_i^{-1} \otimes \mathbf{S}^{-1} \mathbf{B} \right) \mathbf{z}_i \right] \right\}, \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}_i^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \right), \quad i = 1, \dots, k \quad (4.28)$$

$$\mathbf{B} | \mathbf{M}, \mathbf{S}, \mathbf{R}_i \sim N \left( \sum_{i=1}^k \tilde{\mathbf{M}}_i \mathbf{R}_i^{-1} \mathbf{Z}_i' \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}_i^{-1} \mathbf{Z}_i' \right)^{-1}, \left( \sum_{i=1}^k \mathbf{Z}_i \mathbf{R}_i^{-1} \mathbf{Z}_i' \right)^{-1} \otimes \mathbf{S} \right) \quad (4.29)$$

$$\mathbf{F}^{-1} | \mathbf{Y}, \mathbf{M}_i \sim \mathbf{W}_s \left( \left( \frac{1}{2} \text{diag} \left\{ \sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i)' (\mathbf{Y}_i - \mathbf{X}_i \mathbf{M}_i) \right\} \right)^{-1}, \frac{n}{2} \right) \quad (4.30)$$

$$\mathbf{S}^{-1} | \mathbf{Y}, \mathbf{M}, \mathbf{B}, \mathbf{R}_i \sim \mathbf{W}_{uv} \left( \left\{ \sum_{i=1}^k (\tilde{\mathbf{M}}_i - \mathbf{B} \mathbf{Z}_i) \mathbf{R}_i^{-1} (\tilde{\mathbf{M}}_i - \mathbf{B} \mathbf{Z}_i)' + g \mathbf{G} \right\}^{-1}, q + g \right) \quad (4.31)$$

The conditional distribution of  $\mathbf{R}$  now becomes slightly more complex due to the fact that the matrix  $\mathbf{R}_i$  contains the covariances between the model parameters of the same animal in successive seasons only when  $q_i = 2$ . Therefore when an animal was recorded over two seasons  $\mathbf{R}_i$  is of dimension  $(2 \times 2)$ , but when an animal was recorded only in one season  $\mathbf{R}_i$  is a scalar.

$$f(\mathbf{R}^{-1} | \mathbf{y}_i, \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S})$$

$$\propto \prod_{i=1}^k f(\mathbf{y}_i | \mathbf{M}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}_i) \cdot f(\mathbf{M}_i | \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{R} | \mathbf{B}, \mathbf{F}, \mathbf{S})$$

$$\propto \prod_{i=1}^k f(\mathbf{M}_i | \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(\mathbf{R})$$

$$\begin{aligned} &\propto \prod_i^{k_2} |\mathbf{R} \otimes \mathbf{S}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left\{ \left( \mathbf{m}_i^{(2)} - (\mathbf{I}_{q_i=2} \otimes \mathbf{B}) \mathbf{z}_i^{(2)} \right)' (\mathbf{R} \otimes \mathbf{S})^{-1} \left( \mathbf{m}_i^{(2)} - (\mathbf{I}_{q_i=2} \otimes \mathbf{B}) \mathbf{z}_i^{(2)} \right) \right\} \right] \\ &\cdot \prod_i^{k_1} \mathbf{r}_{(1)}^{-\frac{uv}{2}} \exp \left[ -\frac{1}{2} \mathbf{r}_{(1)}^{-1} \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{B} \mathbf{z}_{(i)}^{(1)} \right)' \mathbf{S}^{-1} \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{B} \mathbf{z}_{(i)}^{(1)} \right) \right] \\ &\cdot \text{etr} \left[ -\frac{1}{2} (d\mathbf{P}) \mathbf{R}^{-1} \right] |\mathbf{R}|^{-\frac{1}{2}(d-q_i-1)} \end{aligned}$$



$$\begin{aligned}
&\propto |\mathbf{R}|^{-\frac{uvk_2}{2}} \text{etr} \left[ -\frac{1}{2} \mathbf{R}^{-1} \sum_i^{k_2} \left\{ \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right)' \mathbf{S}^{-1} \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right) \right\} \right] \\
&\cdot \mathbf{r}_{(1)}^{-\frac{uvk_1}{2}} \exp \left[ -\frac{1}{2} \mathbf{r}_{(1)}^{-1} \sum_i^{k_1} \left\{ \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{Bz}_{(i)}^{(1)} \right)' \mathbf{S}^{-1} \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{Bz}_{(i)}^{(1)} \right) \right\} \right] \\
&\cdot \text{etr} \left[ -\frac{1}{2} (d\mathbf{P}) \mathbf{R}^{-1} \right] |\mathbf{R}|^{-\frac{1}{2}(d-q_i-1)} \\
&\propto |\mathbf{R}|^{-\frac{uvk_2+d-q_i-1}{2}} \text{etr} \left[ -\frac{1}{2} \mathbf{R}^{-1} \left\{ d\mathbf{P} + \sum_i^{k_2} \left[ \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right)' \mathbf{S}^{-1} \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right) \right] \right\} \right] \\
&\cdot \mathbf{r}_{(1)}^{-\frac{uvk_1}{2}} \exp \left[ -\frac{1}{2} \mathbf{r}_{(1)}^{-1} \sum_i^{k_1} \left\{ \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{Bz}_{(i)}^{(1)} \right)' \mathbf{S}^{-1} \left( \mathbf{m}_{(i)}^{(1)} - \mathbf{Bz}_{(i)}^{(1)} \right) \right\} \right]
\end{aligned}$$

where  $\tilde{\mathbf{M}}_{(i)}^{(2)}$  ( $uv \times 2$ ),  $\mathbf{m}_{(i)}^{(1)}$  ( $uv \times 1$ ),  $\mathbf{Z}_{(i)}^{(2)}$  ( $m \times 2$ ),  $\mathbf{z}_{(i)}^{(1)}$  ( $m \times 1$ ) and  $\rho_{(1)}$  is a scalar. Also note that  $k_2$  denotes the number of animals recorded in both lactation cycles under consideration, while  $k_1$  denotes the number of animals recorded in only one lactation cycle, i.e. either the first or the second lactation cycle. Similarly, the superscripts in parenthesis to  $\mathbf{M}$ ,  $\mathbf{m}$ ,  $\mathbf{Z}$  and  $\mathbf{z}$  refer to the parameters and covariate matrices of the animals recorded in both, or in only one cycle. The subscript in parenthesis to  $\rho$  indicates that this scalar value is applicable to an animal recorded in only one lactation cycle. However, when  $q_i = 1$ , animal  $i$  under consideration may have been recorded during either the first or the second of the two possible lactation cycle and therefore the full conditional distribution of  $\mathbf{R}^{-1}$  is of the following form:

$$\begin{aligned}
f(\mathbf{R}^{-1} | \mathbf{M}, \mathbf{B}, \mathbf{S}) &\propto |\mathbf{R}|^{-\frac{1}{2}(uvk_2+\delta)} \text{etr} \left[ -\frac{1}{2} \mathbf{R} \left\{ \delta\mathbf{P} + \sum_i^{k_2} \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right)' \mathbf{S}^{-1} \left( \tilde{\mathbf{M}}_{(i)}^{(2)} - \mathbf{BZ}_{(i)}^{(2)} \right) \right\} \right] \\
&\cdot \rho_{(11)}^{-\frac{uvk_{11}}{2}} \exp \left[ -\frac{1}{2} \rho_{(11)}^{-1} \sum_i^{k_{11}} \left( \mathbf{m}_{(i)}^{(11)} - \mathbf{Bz}_{(i)}^{(11)} \right)' \mathbf{S}^{-1} \left( \mathbf{m}_{(i)}^{(11)} - \mathbf{Bz}_{(i)}^{(11)} \right) \right] \\
&\cdot \rho_{(12)}^{-\frac{uvk_{12}}{2}} \exp \left[ -\frac{1}{2} \rho_{(12)}^{-1} \sum_i^{k_{12}} \left( \mathbf{m}_{(i)}^{(12)} - \mathbf{Bz}_{(i)}^{(12)} \right)' \mathbf{S}^{-1} \left( \mathbf{m}_{(i)}^{(12)} - \mathbf{Bz}_{(i)}^{(12)} \right) \right] \quad (4.32)
\end{aligned}$$

where  $k_2$  still denotes the number of animals recorded in both the first and second lactation cycle, while  $k_{11}$  denotes the number of animals recorded in only one cycle, the first lactation cycle, and  $k_{12}$  the animals recorded in only the second lactation cycle. As before the same principle applies to the superscripts in parenthesis for  $\mathbf{M}$ ,  $\mathbf{m}$ ,  $\mathbf{Z}$  and  $\mathbf{z}$ . In the case of the subscripts in parenthesis to  $\rho$ , the first subscript indicates that only one lactation cycle has been recorded, while the second subscript makes reference to which of the two consecutive lactation cycles is applicable to the animal under consideration.

However,

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{R}^{-1} = \frac{1}{\rho_{11}\rho_{22} - \rho_{12}^2} \begin{bmatrix} \rho_{22} & -\rho_{12} \\ -\rho_{12} & \rho_{11} \end{bmatrix}.$$

For the hyperparameter  $\mathbf{P}$  the expansion is

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}.$$

The joint conditional distribution on the elements of  $\mathbf{R}$  may now be derived as follows:

$$\begin{aligned} & f(\rho_{11}, \rho_{22}, \rho_{12} \mid \mathbf{M}, \mathbf{B}, \mathbf{S}) \\ & \propto (\rho_{11}\rho_{22} - \rho_{12}^2)^{-\frac{1}{2}(uvk_2 + \delta)} \text{etr} \left[ -\frac{1}{2} \frac{1}{\rho_{11}\rho_{22} - \rho_{12}^2} \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{bmatrix} \right. \\ & \quad \left. \left\{ \delta \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} + \sum_i^{k_2} \left\{ \begin{bmatrix} \mathbf{m}_{i1}^{(2)'} & -\mathbf{B}'\mathbf{z}_{i1}^{(2)'} \\ \mathbf{m}_{i2}^{(2)'} & -\mathbf{B}'\mathbf{z}_{i2}^{(2)'} \end{bmatrix} \mathbf{S}^{-1} \begin{bmatrix} \mathbf{m}_{i1}^{(2)} - \mathbf{B}\mathbf{z}_{i1}^{(2)} & \mathbf{m}_{i2}^{(2)} - \mathbf{B}\mathbf{z}_{i2}^{(2)} \end{bmatrix} \right\} \right\} \right] \\ & \quad \cdot \rho_{11}^{-\frac{uvk_{11}}{2}} \exp \left[ -\frac{1}{2} \rho_{11}^{-1} \sum_i^{k_{11}} (\mathbf{m}_{(i)}^{(1)} - \mathbf{B}\mathbf{z}_{(i)}^{(1)})' \mathbf{S}^{-1} (\mathbf{m}_{(i)}^{(1)} - \mathbf{B}\mathbf{z}_{(i)}^{(1)}) \right] \\ & \quad \cdot \rho_{22}^{-\frac{uvk_{12}}{2}} \exp \left[ -\frac{1}{2} \rho_{22}^{-1} \sum_i^{k_{12}} (\mathbf{m}_{(i)}^{(2)} - \mathbf{B}\mathbf{z}_{(i)}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{(i)}^{(2)} - \mathbf{B}\mathbf{z}_{(i)}^{(2)}) \right] \\ & \propto (\rho_{11}\rho_{22} - \rho_{12}^2)^{-\frac{1}{2}(uvk_2 + \delta)} \exp \left[ -\frac{1}{2} \frac{1}{\rho_{11}\rho_{22} - \rho_{12}^2} \right. \\ & \quad \left( \rho_{22} \left\{ \delta P_{11} + \sum_i^{k_2} (\mathbf{m}_{i1}^{(2)} - \mathbf{B}\mathbf{z}_{i1}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i1}^{(2)} - \mathbf{B}\mathbf{z}_{i1}^{(2)}) \right\} \right. \\ & \quad - 2\rho_{12} \left\{ \delta P_{12} + \sum_i^{k_2} (\mathbf{m}_{i2}^{(2)} - \mathbf{B}\mathbf{z}_{i2}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i1}^{(2)} - \mathbf{B}\mathbf{z}_{i1}^{(2)}) \right\} \\ & \quad \left. \left. + \rho_{11} \left\{ \delta P_{22} + \sum_i^{k_2} (\mathbf{m}_{i2}^{(2)} - \mathbf{B}\mathbf{z}_{i2}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i2}^{(2)} - \mathbf{B}\mathbf{z}_{i2}^{(2)}) \right\} \right\} \right] \\ & \quad \cdot \rho_{11}^{-\frac{uvk_{11}}{2}} \exp \left[ -\frac{1}{2} \rho_{11}^{-1} \sum_i^{k_{11}} (\mathbf{m}_{(i)}^{(1)} - \mathbf{B}\mathbf{z}_{(i)}^{(1)})' \mathbf{S}^{-1} (\mathbf{m}_{(i)}^{(1)} - \mathbf{B}\mathbf{z}_{(i)}^{(1)}) \right] \\ & \quad \cdot \rho_{22}^{-\frac{uvk_{12}}{2}} \exp \left[ -\frac{1}{2} \rho_{22}^{-1} \sum_i^{k_{12}} (\mathbf{m}_{(i)}^{(2)} - \mathbf{B}\mathbf{z}_{(i)}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{(i)}^{(2)} - \mathbf{B}\mathbf{z}_{(i)}^{(2)}) \right]. \end{aligned} \tag{4.33}$$

When using the information from all records in obtaining marginal posterior distributions, a rather complicated sampling scheme is required. The elements of  $\mathbf{R}^{-1}$  are obtained individually conditional on certain restrictions based on the values of the rest of the elements of  $\mathbf{R}^{-1}$  using the Metropolis Hastings algorithm.

In the above discussion the simplest of such incomplete data cases was described, but the principle explained here may be expanded to any number of incomplete lactation cycles in a data set for a herd observed over a specified period.

## 4.5 THE METROPOLIS-HASTINGS ALGORITHM

The Metropolis-Hastings algorithm, developed in 1953 by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, is a Markov chain Monte Carlo (MCMC) method used to generate non-standard and often complex multivariate distributions through sampling. This method was initially only used in field of physics, but in 1970 Hastings generalised it to be more applicable in a statistical setting. It has, however, only really become popular in the area of Bayesian statistics since the 1990's with the increasing availability of computing power through the use of faster and more powerful computers. The Gibbs sampler described earlier is a special application of the Metropolis-Hastings algorithm (Gelman, 1992). The most important feature of the Metropolis-Hastings algorithm is that it can simulate values from any distribution for which the analytical form of the distribution is known, without having any knowledge of the normalising constant of the distribution.

A sample is required from a target posterior distribution,  $\pi(\mathbf{q} | y)$ , also known as a probability mass function. To acquire such a sample, values of  $\mathbf{q}$  are drawn from approximate distributions with the same restrictions as  $\pi(\mathbf{q} | y)$  and then corrected to obtain a sample that is closer in distribution to the target posterior distribution. Sample draws form a Markov chain as a result of sequential sampling, form a distribution that depends on the last sample value drawn and converges to a unique stationary posterior distribution. The success of this method is based on the improvement of the approximate distributions at each simulation step so that the result converges to the target distribution (Gelman et al., 1995). The algorithm may also be used in an acceptance-rejection application when the applicable density is in a non-identifiable form.

To obtain a sample from the target distribution,  $\pi(\mathbf{q} | y)$ , by means of the Metropolis-Hastings algorithm, the first step is to select an arbitrary starting value  $\mathbf{q}^{(0)}$  such that  $\pi(\mathbf{q}^{(0)} | y) > 0$ . Next for  $t = 1, 2, \dots$  the following iterative procedure is followed to each time update from  $\mathbf{q}^{(t-1)}$  to  $\mathbf{q}^{(t)}$ :

1) A candidate point  $\mathbf{q}^*$  is sampled from a proposal density at time  $t$ ,  $\mathbf{q}^* \sim q_t(\mathbf{q}^* | \mathbf{q}^{(t-1)})$ .

2) Calculate the following: 
$$r = \min \left\{ 1, \frac{p(\mathbf{q}^* | y) q_t(\mathbf{q}^{(t-1)} | \mathbf{q}^*)}{p(\mathbf{q}^{(t-1)} | y) q_t(\mathbf{q}^* | \mathbf{q}^{(t-1)})} \right\}$$

3) Set:

$$q^t = \begin{cases} \mathbf{q}^* & \text{with probability } r \\ \mathbf{q}^{(t-1)} & \text{otherwise} \end{cases}$$

Every time steps 1 to 3 above are performed, one iteration of the procedure has been performed. In principle the above steps imply that when an observation has been drawn from the proposal density  $q(\cdot | \cdot)$ , it is assumed that it originated from  $\pi(\cdot | y)$  with probability  $r$ .

An infinite number of possible proposal densities exist that would result in a Markov chain that converges to the target density. As long as the proposal density is symmetric it is evident that  $q_t(\mathbf{q}^{(t-1)} | \mathbf{q}^*) = q_t(\mathbf{q}^* | \mathbf{q}^{(t-1)})$  and therefore the probability with which the new value,  $\mathbf{q}^*$ , is accepted is the ratio of the density ordinate at the new value and the density ordinate at the current value. The result is that if the new value is “more likely” than the current value the new value is accepted, else the current value is kept with probability  $1 - r$ . These retained values ultimately constitute a sample from the target distribution  $\pi(\cdot | y)$ .

Many options are available when constructing a proposal density, but the most frequent one used is the random-walk proposal where the candidate value is equal to the current value plus noise:

$$\mathbf{q}^* = \mathbf{q}^{(t-1)} + \mathbf{e},$$

where  $\mathbf{e}$  is random variable symmetrically distributed around zero (Chib and Greenberg, 1995). This implies that the candidate value is selected by perturbing the current value of the chain at random, while still staying in the neighbourhood of the current value. Once the candidate value has been generated the probability that it comes from the target distribution is assessed. The chain will remain at the current value or  $\mathbf{q}^{(t-1)}$  with probability  $1 - r$  if the posterior value or  $\pi(\mathbf{q}^{(t-1)} | y)$  is higher, i.e. it will then never generate candidate values where  $\pi(\mathbf{q}^* | y) = 0$ .

Standard options for use as proposal densities are the uniform, normal and  $t$ -distributions, with the choice usually determined by the parameter space. If the parameter space is, for example, bounded, the uniform distribution is used, but for the real line as parameter space the normal or  $t$ -distribution are used. One characteristic of the proposal density that deserves

careful consideration is its variance. If the variance is too large, too many of the proposal values will not be accepted, resulting in a chain that will remain in one place for too long. However, if the variance is too small the chain will, as a result of a high acceptance ratio, move very slowly through the sample space, with high autocorrelation and a small effective sample size. An acceptable acceptance rate for the proposal values is considered to usually be between 40% and 70% depending on the scenario to which it is applied. The variance of the proposal density may be adjusted by first running samples of chains to investigate and if necessary correct the variance of the proposal density in order to adjust the acceptance rate.

Earlier it was mentioned that, for the incomplete data case, the elements of  $\mathbf{R}^{-1}$  are obtained individually conditional on certain restrictions based on the values of the rest of the elements of  $\mathbf{R}^{-1}$  using the Metropolis Hastings algorithm. To now obtain the marginal posterior distributions of each of the elements of  $\mathbf{R}$ , i.e.  $\mathbf{r}_{11}$ ,  $\mathbf{r}_{22}$  and  $\mathbf{r}_{12}$ , from (4.33) the Metropolis-Hastings algorithm is employed with the following restrictions that are placed on the values of these parameters:

$$\mathbf{r}_{11} > \frac{\mathbf{r}_{12}^2}{\mathbf{r}_{22}} \quad (4.34)$$

$$\mathbf{r}_{22} > \frac{\mathbf{r}_{12}^2}{\mathbf{r}_{11}} \quad (4.35)$$

$$-\sqrt{\mathbf{r}_{11}\mathbf{r}_{22}} < \mathbf{r}_{12} < \sqrt{\mathbf{r}_{11}\mathbf{r}_{22}} \quad (4.36)$$

These restrictions are necessary to ensure that  $\mathbf{R}$  remains positive definite and are used to obtain the specified probability mass functions used in the application of the Metropolis-Hastings algorithm. For example, when determining the marginal distribution

$$\begin{aligned} & f(\rho_{11} | \rho_{22}, \rho_{12}, \mathbf{M}, \mathbf{B}, \mathbf{S}) \\ & \propto (\rho_{11}\rho_{22} - \rho_{12}^2)^{\frac{1}{2}(uvk_2 + \delta)} \exp \left[ -\frac{1}{2} \frac{1}{\rho_{11}\rho_{22} - \rho_{12}^2} \left( \rho_{22} \left\{ \delta P_{11} + \sum_i^{k_2} (\mathbf{m}_{i1}^{(2)} - \mathbf{Bz}_{i1}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i1}^{(2)} - \mathbf{Bz}_{i1}^{(2)}) \right\} \right. \right. \\ & \quad \left. \left. - 2\rho_{12} \left\{ \delta P_{12} + \sum_i^{k_2} (\mathbf{m}_{i2}^{(2)} - \mathbf{Bz}_{i2}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i1}^{(2)} - \mathbf{Bz}_{i1}^{(2)}) \right\} \right. \right. \\ & \quad \left. \left. + \rho_{11} \left\{ \delta P_{22} + \sum_i^{k_2} (\mathbf{m}_{i2}^{(2)} - \mathbf{Bz}_{i2}^{(2)})' \mathbf{S}^{-1} (\mathbf{m}_{i2}^{(2)} - \mathbf{Bz}_{i2}^{(2)}) \right\} \right) \right] \\ & \cdot \rho_{11}^{\frac{9k_{11}}{2}} \exp \left[ -\frac{1}{2} \rho_{11}^{-1} \sum_i^{k_{11}} (\mathbf{m}_{(i)}^{(11)} - \mathbf{Bz}_{(i)}^{(11)})' \mathbf{S}^{-1} (\mathbf{m}_{(i)}^{(11)} - \mathbf{Bz}_{(i)}^{(11)}) \right] \end{aligned} \quad (4.37)$$

using this algorithm with the restrictions set out in (4.34) to (4.36), a candidate value,  $\mathbf{r}_{11}^c$ , of the following form is obtained:

$$\mathbf{r}_{11}^c = \frac{\mathbf{r}_{12}^2}{\mathbf{r}_{22}} + \left( \mathbf{r}_{11} - \frac{\mathbf{r}_{12}^2}{\mathbf{r}_{22}} \right) \exp(z) \quad (4.38)$$

where  $z$  is randomly sampled from a uniform distribution in the interval  $-0.5$  to  $0.5$ . This candidate value is then accepted with probability

$$\min \left( 1, \frac{f(\rho_{11}^c | \rho_{22}, \rho_{12}, \mathbf{M}, \mathbf{B}, \mathbf{S})}{f(\rho_{11} | \rho_{22}, \rho_{12}, \mathbf{M}, \mathbf{B}, \mathbf{S})} \right). \quad (4.39)$$

The marginal distributions of  $\mathbf{r}_{12}$  and  $\mathbf{r}_{22}$  are obtained in a similar manner.

In the incomplete data case the full conditional distributions of the model parameters  $\mathbf{m}_i$  where  $i = 1, \dots, k$ ,  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$  and  $\mathbf{S}^{-1}$  as specified in (4.28) to (4.31), as well as the joint conditional distribution of the elements of  $\mathbf{R}$  as found in (4.33) from which the marginal distribution of each element of  $\mathbf{R}$  is determined, is used to obtain the marginal distributions of the parameters by means of MCMC methods. For  $\mathbf{m}_i$ ,  $i = 1, \dots, k$ ,  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$  and  $\mathbf{S}^{-1}$  the Gibbs sampler is used, while the marginal posterior distributions of each of the elements of  $\mathbf{R}$  is obtained by using the Metropolis-Hastings algorithm.

# CHAPTER 5

## THE WOOD MODEL AND THE JERSEY DATA

### 5.1 THE JERSEY DATA

The Jersey is one of the oldest dairy cattle breeds in the world. It originated on the British Island of Jersey in the English Channel in the late 1700's, but has since as a result of their adaptability to a wide range of climatic and geographical conditions, spread all over the world. They are described not only as excellent grazers, but also as more tolerant of heat than the larger dairy breeds. One outstanding characteristic of this breed of dairy cattle is that it produces more milk in relation to its own body weight than any other breed. In most cases milk production in excess of thirteen times the bodyweight of the animal is recorded during any lactation cycle. On average an adult female would weigh approximately 400 kg. The breed is not only considered favourable as a result of its milk production ability, but also because of its butterfat production. (Department of Animal Science, Oklahoma State University, 1997).

The Jersey data used in this study forms part of a larger data set on the lactations of Jersey dairy cows that was obtained from the Animal Improvement Institute at the Agricultural Research Council, Irene, Pretoria. The original data set contained lactation information on 209 274 lactation cycles. This data file consisted of almost 52 million entries in a total of 31 fields (columns) that included information such as farmer member number, animal identification number, calving date, age of cow at calving, season of calving, parity number, test day date, number of times (or days) tested during lactation cycle, number of days from calving to test day, milk weight in kilograms as measured during each of either two or three milking session on each test day, the percentage of fat in milk composition, and the percentage of protein in milk composition. Of the 31 fields in the file, 7 were duplicated, which leaves 24 unique fields. Given the excessively large size of the data set available, a more manageable subset was obtained from this data set for Bayesian analysis purposes by performing several edits. This would result in a reduction of the time spent on obtaining results through the computer intensive simulation techniques discussed in the previous chapter.

The subset from the above data considered in this study, only include cows recorded during all four years from 1995 to 1998, with at least eight test day records per lactation cycle for all four these cycles under consideration. The resulting data set, referred to as the Jersey data from this point onwards, contains lactation data for 1141 cows recorded on at least eight, but no more than ten test days during four consecutive lactation cycles from 1995 to 1998. This resulted in a total of 37 163 test day records for each of milk yield, protein percentage and fat percentage in milk composition. Note that each test day milk measure is the sum of the milk yields measured during either two or three milking sessions on that test day, while for both fat and protein content in milk composition the average of the milking sessions, averaged by respective milk yields are considered. The number of days from calving to test day was used as indicator of at which stage of lactation a lactation record was observed. Additional information also included in the Jersey data for analysis purposes are parity number, region, calving year and calving season.

Parities identified in the data ranged from first parity to eleventh parity, of which cows in 515 first parities, 825 second parities, 987 third parities, 1066 fourth parities, 599 fifth parities, 306 sixth parities, 153 seventh parities, 75 eighth parities, 29 ninth parities, 8 tenth parities and only one eleventh parity occurred in the 4564 lactation cycles observed. To obtain a better balance among the number of lactation cycles observed in each parity, the seventh to eleventh parities were combined into one group referred to as parity seven or greater. Animals from seven different regions were included in the data; these were referred to as regions one to five, seven and eight. No cows in region six were recorded. Although it is known that all seven these regions are within South Africa, the Animal Improvement Institute considered the geographical location of each of these regions to be confidential. The year of parturition, 1995 to 1998, served as an indicator of the lactation cycle under consideration. The season of parturition was also denoted as belonging to one of six possible two-month seasons, but it was decided to reduce this to only two seasons by grouping together as season one, the warmer months October to March, while the colder months April to September occur as season two.

Cows calve throughout the year usually at intervals of at least thirteen months. The lactation cycle lasts 305 days and starts four days after parturition. Colostrum is secreted during the first three days and only on the fourth day after parturition the secretion of true milk commences. To make provision for this, the start of lactation is adjusted so that day one of the lactation cycle is considered to be on the fourth day after parturition by subtracting 3 days



form the number of days from calving to test day that is used to determine at which stage of lactation a lactation record was observed.

## 5.2 FITTING THE WOOD MODEL

From the discussion on the Wood model in Chapter 2 it is evident that this model is the model most often used to model lactation in dairy animals, including dairy cows such as the Jersey. The Wood model assumes that the expected milk yield (in kg/day) of an animal at time  $t$  can be represented over the lactation period by

$$E(W_t) = t^b \exp(a + ct) \quad (5.1)$$

where  $-\infty < a < \infty$ ,  $b > 0$  and  $c < 0$ . The parameters  $a$ ,  $b$  and  $c$  are unknown and may differ from one animal to another.

In 1976 Wood showed that this model could also be used to estimate the various milk composition traits such as fat and protein content in milk yield of dairy cattle. Sakul and Boylan (1992) found the Wood model to be useful in describing almost all milk composition traits in sheep breeds, except for lactose content. As a result it was decided to also use the Wood model to model milk composition traits fat and protein content on which data are available in the Jersey data. The percentages of fat and protein contained in milk are typically convex functions over time and therefore the restrictions placed on the model parameters in order to make provision, when modelling these composition traits, for the convex nature of these functions over time are  $b < 0$  and  $c > 0$ , while  $-\infty < a < \infty$  remains.

By assuming multiplicative errors for the model in (5.1) and after performing a natural log transformation, the observation model for the  $i^{\text{th}}$  animal is written as:

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) = a_{ijs} + b_{ijs} \ln(t_{ijp}) + c_{ijs} t_{ijp} + e_{ijps} \quad (5.2)$$

where  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, q$  where for every animal in the herd a total of  $q$  lactation cycles are observed,  $s = 1, \dots, u$  where  $u$  is the number of lactation traits observed, and  $p = 1, \dots, n_{ij}$  where  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ . For the Jersey data the number of animals is  $k = 1141$ , for which  $q = 4$  lactation cycles each have been observed for the  $u = 3$  lactation traits milk yield, percentage of fat and

percentage of protein in milk composition and the number of test days for animal  $i$  during lactation cycle  $j$  is  $8 \leq n_{ij} \leq 10$  for all values of  $i$  and  $j$ .

The generalised linear model form of the model for animal  $i$  during lactation cycle  $j$  as described in (4.3) for the general case is:

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij} \mathbf{M}_{ij} + \mathbf{E}_{ij} \quad (5.3)$$

where  $\text{vec } \mathbf{E}'_{ij} (3n_{ij} \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_{ij}})$ . For the Wood model the  $v = 3$  regression coefficients of the generalised linear form of the lactation model for the  $i^{\text{th}}$  animal during its  $j^{\text{th}}$  lactation cycle for the lactation trait indicated by  $s = 1, 2, 3$  are  $a_{ijs}, b_{ijs}, c_{ijs}$ , so that

$$\mathbf{M}_{ij} (3 \times 3) = \begin{bmatrix} a_{ij1} & a_{ij2} & a_{ij3} \\ b_{ij1} & b_{ij2} & b_{ij3} \\ c_{ij1} & c_{ij2} & c_{ij3} \end{bmatrix}. \quad (5.4)$$

The  $r$ -th row of design matrix  $\mathbf{X}_{ij} (n_{ij} \times 3)$  contains the elements that would return the Wood model in generalised linear form when multiplied with the matrix of coefficients  $\mathbf{M}_{ij}$ ,

$$\mathbf{X}_{ij}^{(r)} = [1 \quad \ln t_{ijr} \quad t_{ijr}]. \quad (5.5)$$

Note that to make provision for the initial secretion of colostrum, time  $t$  would be measured from day 4 after parturition, i.e. on day 4 after parturition  $t = 1$ .

For animal  $i$  over all  $q = 4$  lactation cycles observed,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_{i1} \\ \mathbf{Y}_{i2} \\ \mathbf{Y}_{i3} \\ \mathbf{Y}_{i4} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (5.6)$$

where  $\text{vec } \mathbf{E}'_i (3n_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^4 n_{ij}$ ,  $\mathbf{X}_i (n_i \times 12) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i (12 \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \\ \mathbf{M}_{i3} \\ \mathbf{M}_{i4} \end{bmatrix}. \quad (5.7)$$

From the earlier discussion it follows that the additional information available in the Jersey data, which are the factors parity number, region, calving year and calving season, has to also be included in the model. This is done by means of the covariate matrix  $\mathbf{Z}_i (17 \times 4)$  for the  $i^{\text{th}}$

animal, which is then used as described in section 4.1 of Chapter 4. A full description of how this covariate matrix  $\mathbf{Z}_i$  is obtained for each animal follows in section 5.3 below.

As was mentioned earlier,  $q = 4$  lactation cycles, one for each of the years 1995 to 1998 were observed for all 1141 animal in the Jersey data. As a result the matrix  $\mathbf{R}$ , containing the covariances between the model parameters of the same animal in successive lactation cycles now has the following form:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{r}_{13} & \mathbf{r}_{14} \\ \mathbf{r}_{12} & \mathbf{r}_{22} & \mathbf{r}_{23} & \mathbf{r}_{24} \\ \mathbf{r}_{13} & \mathbf{r}_{23} & \mathbf{r}_{33} & \mathbf{r}_{34} \\ \mathbf{r}_{14} & \mathbf{r}_{24} & \mathbf{r}_{34} & \mathbf{r}_{44} \end{bmatrix}. \quad (5.8)$$

Through the use of the Gibbs sampler as described in section 4.2 of Chapter 4 and the full conditional distributions of all the model parameters of interest as set out in equations (4.17) to (4.21), the marginal posterior distributions conditional on the observed data only were obtained. The hyperparameters  $\mathbf{P}$  and  $\mathbf{G}$  required to generate marginal distributions for  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , were taken as the moments estimators from the sampling distributions of  $\mathbf{S}$  and  $\mathbf{R}$ , with degrees of freedom as small as possible,  $d = 4$  and  $g = 10$ .

Applying the sampling approach suggested by MacEachern and Berliner (1994) the Gibbs sampler was put through a “burn-in” period of 2000 simulation iterations, after which 10 000 sets of parameters were generated and kept using equations (4.17) to (4.21). This required simulating from normal and Wishart distributions only. The distributions of  $\mathbf{M}_i$  and  $\mathbf{B}$  were obtained by sampling from two normal distributions, which is a relatively simple procedure in Matlab. This required generating samples from the standard normal distribution that were then transformed to be normally distributed with the appropriate mean and standard deviation. The distributions of  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  were generated from their respective Wishart distributions. To explain how this was done, it is assumed that the variable of interest has a Wishart distribution with parameter  $\mathbf{A}$  and degrees of freedom  $\phi$ . To sample from this distribution a random matrix  $\mathbf{H}$  from a standard normal distribution with the same number of rows as the parameter matrix  $\mathbf{A}$  and number of columns equal to  $\phi$ , is generated. The matrix square root of  $\mathbf{A}$  is then multiplied with this random normal matrix  $\mathbf{H}$  just generated in the previous step. The result of this,  $\mathbf{A}^{\frac{1}{2}}\mathbf{H}$ , is then multiplied with its transpose to obtain a

simulated parameter matrix,  $A^{\frac{1}{2}}\mathbf{H}\mathbf{H}'A^{\frac{1}{2}}$ , from a Wishart distribution with parameter  $\mathbf{A}$  and degrees of freedom  $\phi$ .

Due to both speed and storage constraints posed by Matlab, only the 10 000 sets of parameters generated for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  were stored. This was done because it is possible to regenerate the parameters contained in  $\mathbf{M}_i$ ,  $i = 1, \dots, 1141$ , with relative ease by either obtaining the product  $\mathbf{B}\mathbf{Z}_i$ , where  $\mathbf{Z}_i$  is the covariate matrix applicable to animal  $i$  or by regenerating through Gibbs sampling the parameters of  $\mathbf{M}_i$  only using the 10 000 stored parameter matrices  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ . This resulted in storing just over 41 million less parameter values than what otherwise would have been the case.

### 5.3 THE COVARIATE MATRIX $\mathbf{Z}_i$

In the Jersey data additional information available on various other factors that could possibly have a significant influence on lactation, is provided. These factors, referred to as cofactors, are parity number, region, calving year and calving season. The parity numbers in the data range from 1 to 7 or greater, regions under consideration are 1 to 5, 7 and 8, calving years are 1995 to 1998, and calving season are warm season and colder season. To include this information on the cofactors in the model a covariate matrix  $\mathbf{Z}_i$  for the  $i^{\text{th}}$  animal has to be constructed.

The additional information on these cofactors is translated into a covariate matrix  $\mathbf{Z}_i(17 \times 4)$  for each animal  $i$  with a total of  $m = 17$  covariates for each of the  $q = 4$  lactation cycles. To explain how this covariate matrix is structured, we consider  $\mathbf{Z}_i = [\mathbf{z}_{i1} \ \mathbf{z}_{i2} \ \mathbf{z}_{i3} \ \mathbf{z}_{i4}]$  where  $\mathbf{z}_{ij}(17 \times 1)$  for  $j = 1, \dots, 4$  form the column vectors of covariates for animal  $i$  over each of its 4 lactation cycles. In the vector  $\mathbf{z}_{i1}$ , for example, the first element or  $z_{i1.1}$  is always a constant, 1. Elements two to seven of this vector are used to identify the parity number. If the animal is in the first parity all seven these elements are set at zero, for an animal in second parity  $z_{i1.2}$  is equal to 1 and  $z_{i1.3}$  to  $z_{i1.7}$  are set at zero, for an animal in third parity  $z_{i1.3}$  is equal to one, but  $z_{i1.2}$  and  $z_{i1.4}$  to  $z_{i1.7}$  are set at zero and so forth. Elements eight to thirteen of  $\mathbf{z}_{i1}$  are used to identify the region in which the animal is found. As was the case with parity number  $z_{i1.8}$  to  $z_{i1.13}$  will all be set at zero if the applicable region is region 1, for an animal in region 2  $z_{i1.8}$

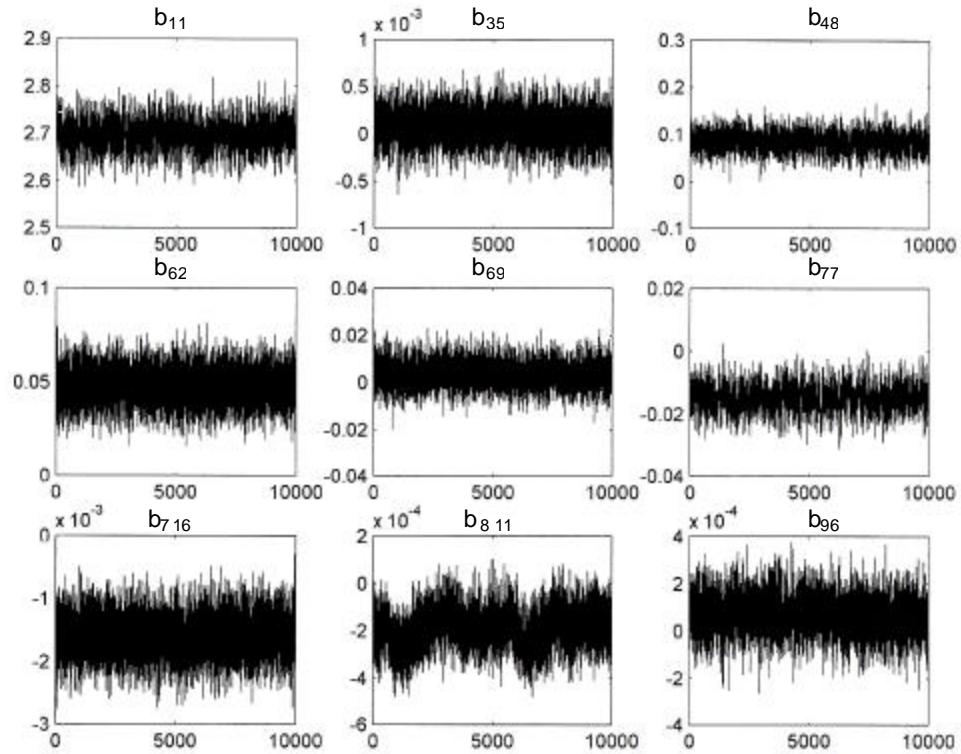
would equal one, while  $z_{i1.9}$  to  $z_{i1.13}$  are all set at zero, and so forth. Similarly, elements 14 to 16 are coded as one and zeros to identify the calving year, with all elements set at base level zero for 1995. The final element of such a vector is used to distinguish between the two seasons, with  $z_{i1.17}$  equal to zero if the calving season for the animal under consideration is during the warmer months and  $z_{i1.17}$  equal to one otherwise. Note that in each case the number of covariates required to distinguish among the different levels of each possibly significant cofactor is one less than the number of possible cofactor levels. For a particular cofactor, by setting all covariates equal to zero, the so-called base level of this cofactor is obtained. The base levels of these cofactors are parity 1 for parity number, region 1 for region of occurrence, 1995 for year of calving and warmer season for season of calving. A typical example of such a covariate matrix would be:

$$\mathbf{Z}_i(17 \times 4) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad (5.9)$$

from which it is evident that this animal, which occurred in region 3, was in its first parity during 1995, when it calved in the warmer season; in its second parity during 1996, when it also calved in the warmer season; in its third parity during 1997, when it this time round calved in the colder season; and in its fourth parity during 1998, when it again calved in the colder season.

## 5.4 THE RESULTS OF THE GIBBS SAMPLER

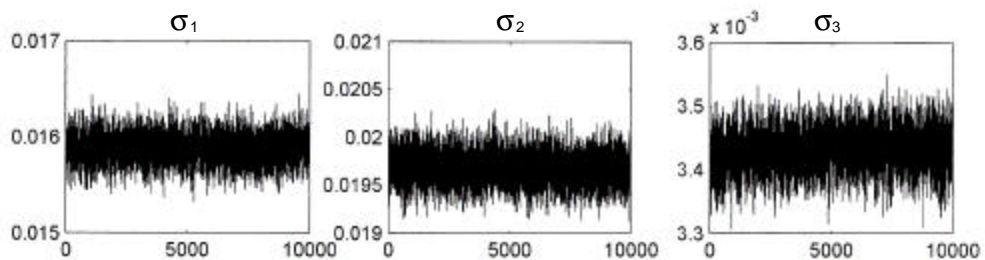
To illustrate the stability of the result obtained, after convergence, when 10 000 sets of parameters have been generated through Gibbs sampling, the following scans of randomly selected elements of  $\mathbf{B}$ ,  $\mathbf{F}$ ,  $\mathbf{S}$  and  $\mathbf{R}$  are considered.



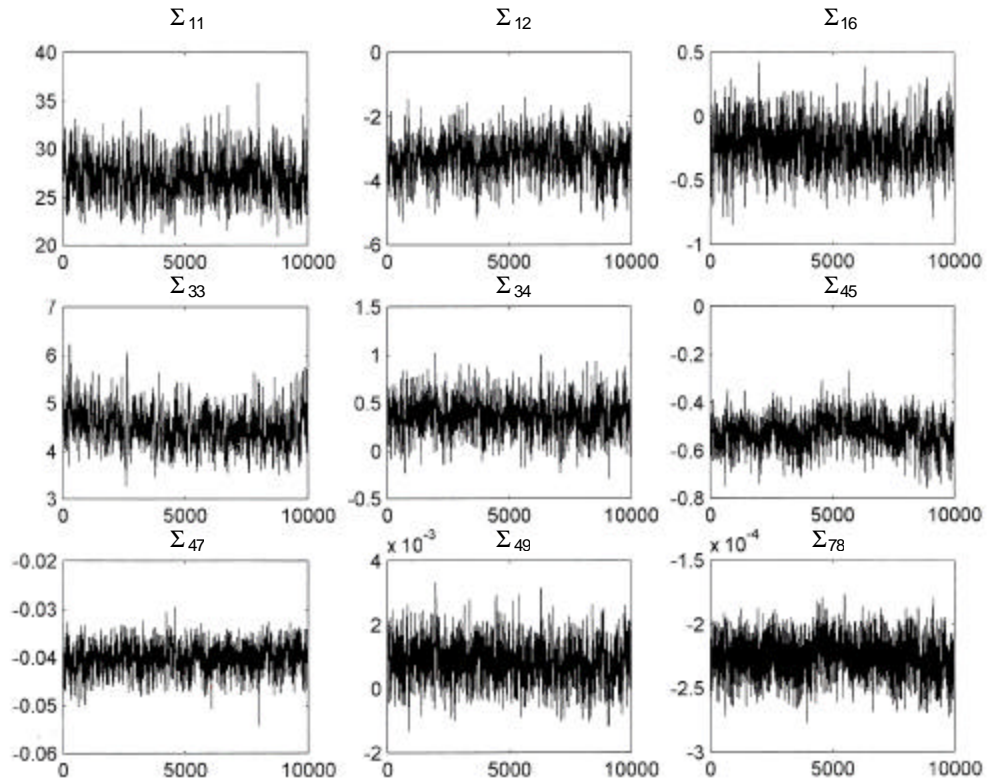
**Figure 5.1: Scans of nine randomly selected elements of the regression matrix  $B$  for the 10 000 simulations retained.**

From the result presented in Figure 5.1 it is evident that after convergence the parameters generated are stable. Some of the parameter values, such as  $b_{811}$ , display slower cyclical movement through the simulated sample space, but are still relatively stable with respect to the interval in which it varies.

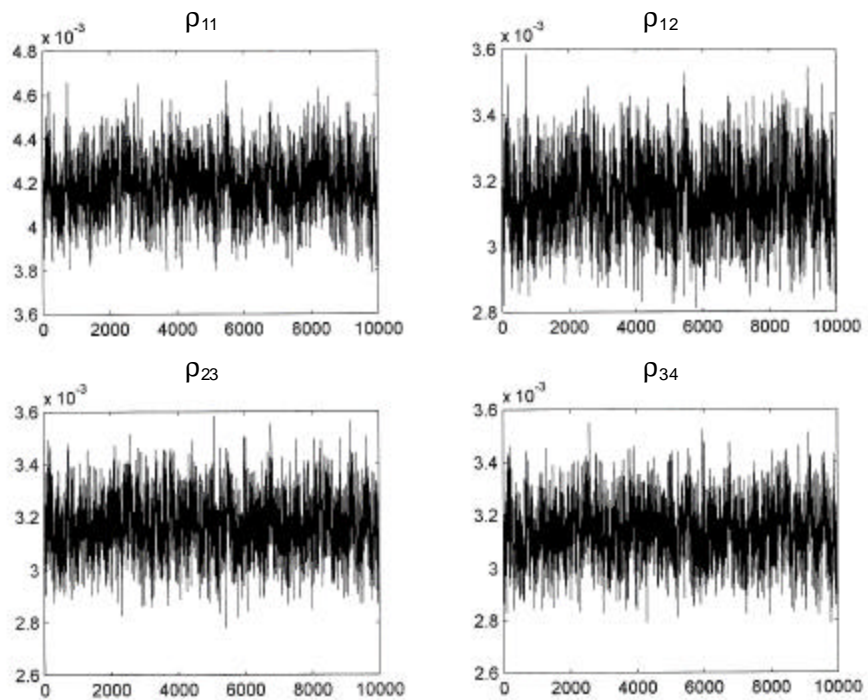
From Figure 5.2 that follows it is also clear that the elements on main diagonal of the diagonal matrix  $F$  are stable for the 10 000 sets of parameters retained from the Gibbs sampler.



**Figure 5.2: Scans of the three elements on the main diagonal of the matrix  $F$  for the 10 000 simulations retained.**



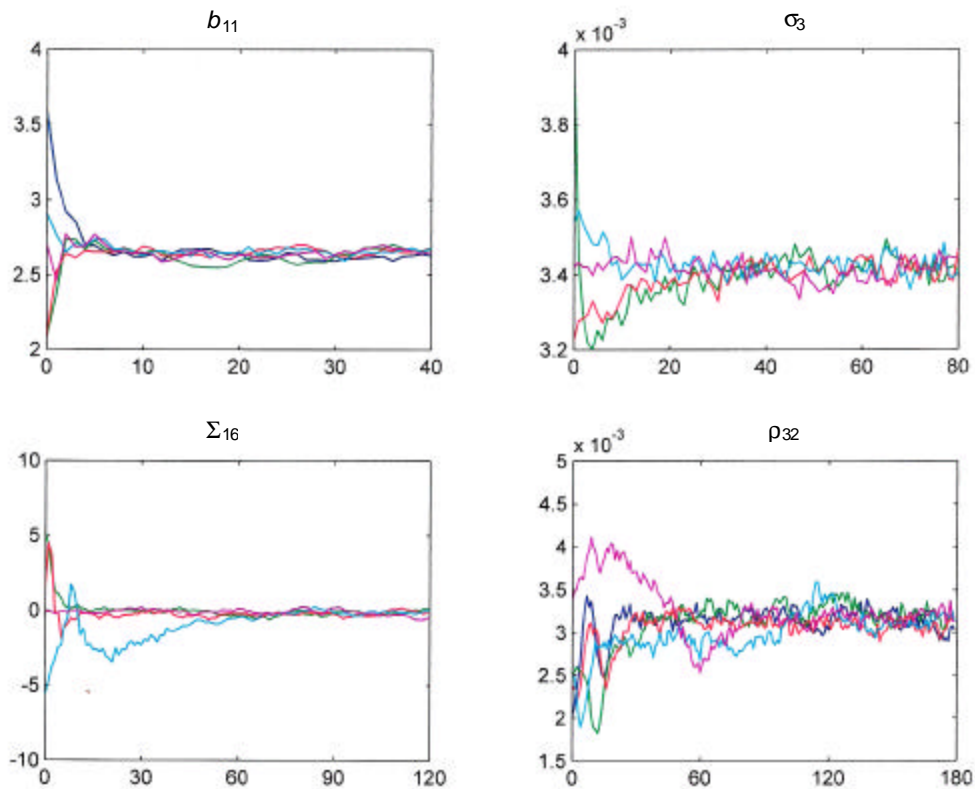
**Figure 5.3: Scans of nine randomly selected elements of the covariance matrix  $S$  for the 10 000 simulations retained.**



**Figure 5.4: Scans of four randomly selected elements of the covariance matrix  $R$  for the 10 000 simulations retained.**

From both Figures 5.3 and 5.4 it is seen that, although some of the parameter values display slower cyclical movement through the simulated sample space, all results of parameters contained in these covariance matrices are still relatively stable with respect to the interval within which they vary.

As was described in Section 4.2, to apply the Gibbs sampler, starting-point values in the form of the least squares estimators were specified for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ . These were considered to result in good starting-point values for the Gibbs sampler. Upon further investigation, however, it was discovered that, as a result of the high efficiency of convergence of the method, “worse” starting-point values also resulted in relatively quick convergence of the parameters obtained through Gibbs sampling. To illustrate scans of the initial simulations from starting-point values of the four parameters,  $b_{11}$ ,  $\sigma_3$ ,  $\Sigma_{16}$  and  $\rho_{32}$ , are considered in Figure 5.5.



**Figure 5.5: Convergence obtained for the parameters  $b_{11}$ ,  $\sigma_3$ ,  $\Sigma_{16}$  and  $\rho_{32}$  using different starting point values**

For each of the parameters in Figure 5.5 either four or five different starting-point values, all “worse” than their least squares estimators, were used to illustrate the rate of convergence. For  $b_{11}$  convergence occurred very quickly in all cases, while for the other parameters



considered convergence took slightly longer. Depending on the quality of the starting-point value for the same parameter some simulation cycles converged faster than others as is evident in the case of  $\Sigma_{16}$  where the smallest starting values required a greater number of simulations (approximately 70 iterations) than any of the others to converge. The other starting-point values used for  $\Sigma_{16}$  seem to already converge after about 20 iterations. The slowest convergence in the above was obtained for the parameter  $\rho_{32}$ .

From the marginal posterior distributions of the elements of the regression matrix  $\mathbf{B}$  ( $9 \times 17$ ) that represents the effects of the covariates on the parameters of the three lactation trait curves, milk yield, percentage of fat and percentage of protein in milk composition, the effects of the covariates on these parameters can be assessed. It is important to remember that not all covariates apply to all animals. The matrix  $\mathbf{Z}_i$  is responsible for the identification of the covariates relevant to the  $i^{\text{th}}$  animal. If we consider the example of the matrix  $\mathbf{Z}_i$  given in (5.9) above, it is evident that only the effect of those covariates to which a one was awarded in the covariate matrix is relevant to this animal. By multiplying the regression matrix  $\mathbf{B}$  and the matrix of covariates  $\mathbf{Z}_i$ , the effects of relevant regression matrix elements applicable to animal  $i$  is extracted and added. The first column of the regression matrix  $\mathbf{B}$  contains the base level effects on the parameters of the three lactation trait curves, while the remaining columns contain the “added” effect on the parameters of the three lactation traits as a result of each of the covariates. The base level effect would be the effect on the parameters resulting from base level settings of the covariate matrix  $\mathbf{Z}_i$  as described in Section 5.3 above. By determining the product  $\mathbf{BZ}_i$ , the sum of the base level effect and the additional effects of the covariates on the nine parameters of the three traits for animal  $i$  for the covariates identified by  $\mathbf{Z}_i$ , is determined.

From the 90% highest posterior density (HPD) intervals, 92 of the 153 the elements of the regression matrix  $\mathbf{B}$  significantly affect the parameters of the three lactation trait curves milk yield, percentage of fat and percentage of protein in milk composition. All nine base level effects contained in  $\mathbf{B}$  are significant. It is also worth noting that the effects of all other covariates on the parameters of the trait curves are significant for 2 or more parameters. The matrix  $\mathbf{B}$  given below contains the mean of the 10 000 simulated matrices, with all values indicated in red representing the mean elements of this regression matrix for which the 90% HPD intervals indicate that the corresponding covariate significantly affects the parameters of the lactation trait curves.

$$B' = \begin{bmatrix} 2,69719 & 1,33862 & 1,24774 & 0,02520 & 0,01180 & -0,01328 & -0,00144 & 0,00068 & 0,00078 \\ 0,20146 & 0,04852 & 0,06972 & 0,00507 & -0,00748 & -0,01462 & -0,00078 & -0,00003 & 0,00013 \\ 0,22380 & 0,06645 & 0,07129 & 0,02434 & -0,01452 & -0,01482 & -0,00114 & -0,00001 & 0,00012 \\ 0,21055 & 0,08431 & 0,09155 & 0,03666 & -0,02213 & -0,02057 & -0,00132 & 0,00003 & 0,00016 \\ 0,15873 & 0,06926 & 0,08485 & 0,05264 & -0,02124 & -0,01794 & -0,00147 & 0,00002 & 0,00011 \\ 0,22507 & 0,10066 & 0,07582 & 0,03940 & -0,02881 & -0,01559 & -0,00147 & 0,00000 & 0,00007 \\ 0,12498 & 0,02311 & 0,09938 & 0,05711 & -0,01315 & -0,02314 & -0,00158 & -0,00004 & 0,00013 \\ -0,10789 & 0,05630 & -0,00462 & 0,04777 & -0,01594 & -0,00334 & -0,00043 & 0,00001 & -0,00005 \\ 0,01491 & 0,03874 & -0,00029 & 0,02974 & -0,00874 & 0,00368 & -0,00030 & -0,00001 & -0,00018 \\ -0,21668 & 0,01849 & -0,01374 & 0,06927 & -0,01317 & -0,00251 & -0,00046 & 0,00001 & -0,00011 \\ -0,09654 & -0,02324 & -0,02018 & 0,06980 & 0,00963 & 0,00345 & -0,00079 & -0,00018 & -0,00015 \\ -0,14731 & -0,15891 & -0,03259 & 0,07252 & 0,06206 & 0,01413 & -0,00031 & -0,00093 & -0,00038 \\ 0,02310 & 0,05038 & -0,03588 & 0,01279 & -0,03021 & 0,00733 & 0,00016 & 0,00019 & -0,00015 \\ 0,02474 & 0,01182 & -0,00206 & -0,01793 & 0,00064 & -0,00608 & 0,00032 & -0,00008 & 0,00012 \\ 0,04081 & 0,00889 & 0,02543 & -0,00725 & 0,00867 & -0,00805 & 0,00019 & -0,00008 & 0,00008 \\ 0,07293 & 0,04018 & 0,00992 & -0,01134 & 0,00041 & -0,01697 & 0,00008 & -0,00008 & 0,00025 \\ -0,24694 & 0,13489 & 0,10385 & 0,09104 & -0,03034 & -0,02288 & -0,00109 & 0,00011 & 0,00004 \end{bmatrix}$$

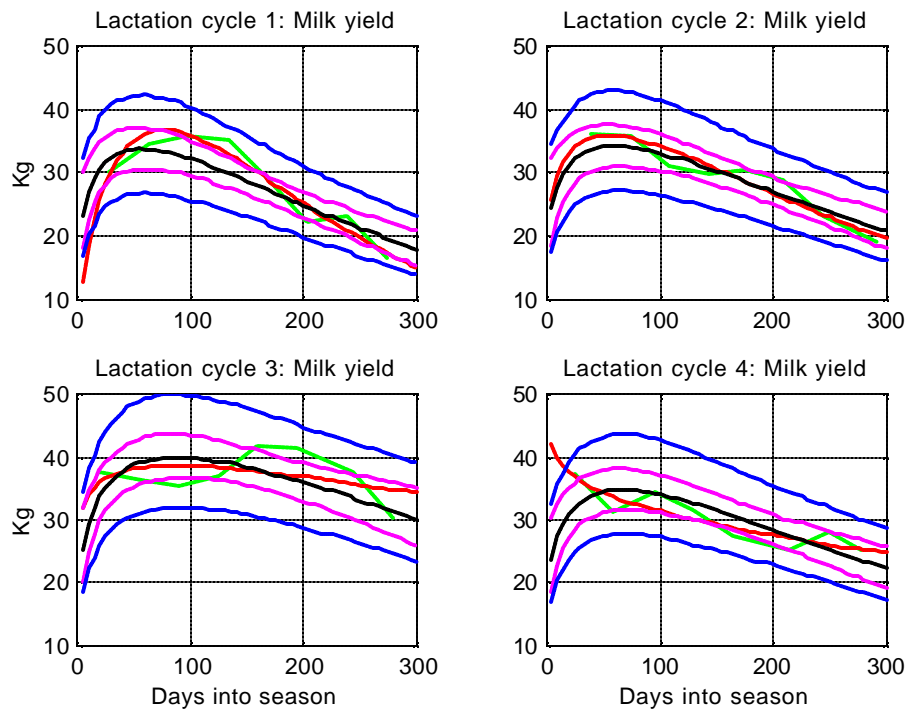
No single covariate could therefore be removed from the model, although it is possible that, for one or more of the parameters of a specific animal, the “added” effects as a result of the covariates added to the base level effect could be very small or in some cases even insignificant.

Once the Wood model had been fitted to the Jersey data and the marginal distributions of the parameters obtained by means of Gibbs sampling, the various aspects of interest in the lactation data of the Jerseys may be investigated. One of the method employed as part of the investigation is the construction of highest posterior density (HPD) intervals, *not* credibility intervals, for the expected lactation trait curves. It should be noted that, in some cases, the posterior distributions are very diffused so that 95% HPD intervals are of little practical value. In those cases the 90% HPD intervals are reported.

## 5.5 THE LACTATION TRAIT MILK YIELD

The Wood model, which is of the form  $E(W_i) = t^b \exp(a + ct)$ , was fitted to the Jersey data to model the lactation process. Once the 10 000 simulated  $M_i$  matrices for each animal  $i$  over all four its lactation cycles have been obtained, the lactation curve or milk yield curve for any animal  $i$  during each of its four lactation cycles for each of the 10 000 simulations can be

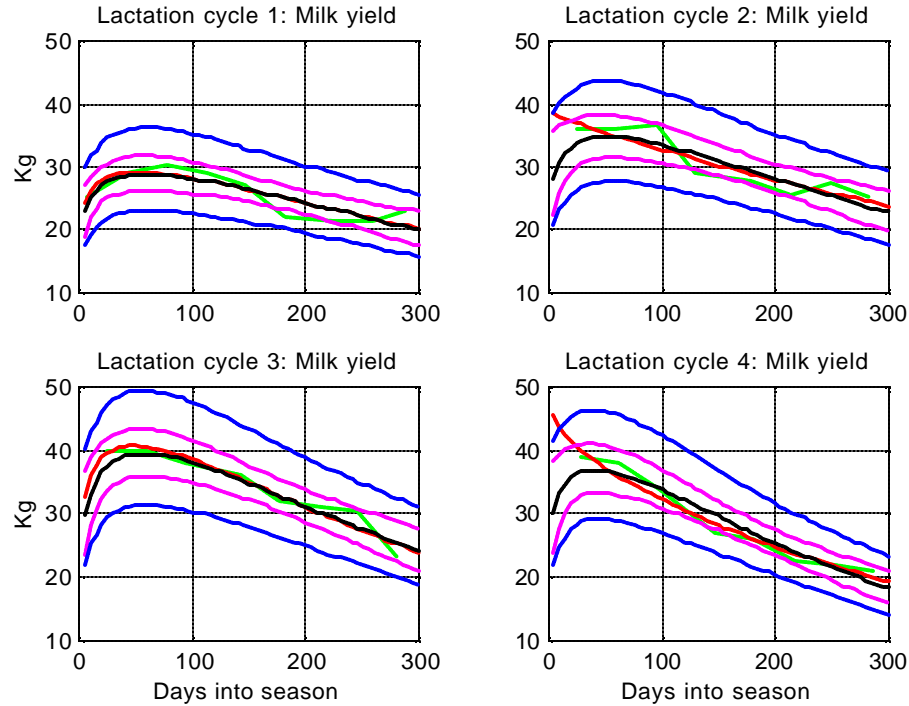
obtained. The mean of the 10 000 lactation curves generated for each lactation cycle in this manner gives the expected lactation curve for animal  $i$  in each lactation cycle. As a result of not being able to store the 10 000 sets of the parameters contained in  $\mathbf{M}_i$  ( $12 \times 3$ ) for  $i = 1, \dots, 1141$ , these values have to be redetermined. Two alternative methods to do this exist. Alternative 1, the fastest and simplest of the two, is to estimate the parameters contained in  $\tilde{\mathbf{M}}_i$  ( $9 \times 4$ ) through calculating  $\mathbf{B}\mathbf{Z}_i$  for each of the 10 000 simulated  $\mathbf{B}$  ( $9 \times 17$ ) matrices, using the  $\mathbf{Z}_i$  ( $17 \times 4$ ) relevant to animal  $i$ . Alternative 2, slightly more cumbersome although probably more accurate, is to use each of the 10 000 simulated matrices  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , together with the matrix  $\mathbf{Z}_i$  ( $17 \times 4$ ) relevant to animal  $i$ , to again simulate by means of the Gibbs sampler the elements of  $\mathbf{M}_i$  for 10 000 simulation iterations. This means that only the matrix  $\mathbf{M}_i$  has to be generated using the Gibbs sampler described in section 4.2 of Chapter 4, as all other parameter matrices from the 10 000 original simulation iterations are available.



**Figure 5.6:** For each of the four lactation cycle of Animal 135 the expected lactation curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

Using the second alternative to regenerate 10 000  $\mathbf{M}_i$  matrices through Gibbs sampling for two animals ( $i = 135$  and  $i = 511$ ), the estimates of the model parameters for milk yield contained in the 10 000 simulated  $\mathbf{M}_i$  matrices for each of  $i = 135$  and  $i = 511$  were used to

determine 10 000 lactation curves for each animal. The mean of the 10 000 lactation curves for each of the animals  $i = 135$  and  $511$  gave the expected lactation curve for that animal. These expected lactation curves for each of the four lactation cycles are represented as a black solid line in Figures 5.6 and 5.7 for the two animals under consideration.



**Figure 5.7:** For each of the four lactation cycle of Animal 511 the expected lactation curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 511 is given by —.

In Figures 5.6 and 5.7 the green line represents the observed lactation yield values connect by straight-line segments for the test day records of the four lactation cycles. The red line is the result of least squares estimation of the lactation curve fitted to the data in a frequentistic context through linear least squares regression, after a log transformation had been performed on the data. The magenta coloured broken lines provide the 90% HPD intervals for the expected lactation curve during each lactation cycle. The blue broken lines in turn represent a 90% prediction interval for the expected lactation curve during each lactation cycle. The 90% prediction intervals were obtained by first finding the 10 000 lactation yield curves that would result when obtaining the product  $X_i M_i$  for each of the 10 000 simulated  $M_i$  matrices and the design matrix  $X_i$  relevant to animal  $i$ . The predictive density then follows from the fact that

$$Y_i | M_i, F \sim N(X_i M_i, F \otimes I_{n_i}). \quad (5.10)$$

As a result a predictive matrices  $Y_i^f$  is simulated from a normal distribution of the form denoted in (5.10) for each of the 10 000 means,  $X_i M_i$ , and variances,  $F \otimes I_{n_i}$ . The predictive matrices in  $Y_i^f$  referring to lactation yield only are then used to construct 90% HPD intervals for these predictions. These intervals are graphed as the 90% prediction intervals for the animals under consideration.

The least squares estimate of the milk yield curve for the fourth lactation cycle of animal 135 takes on an atypical form, so too does the least squares estimate of the milk yield curve for both the second and fourth lactation cycle of animal 511, but in all three the afore mentioned cases, however, the Bayesian curve estimate of the expected milk yield is *not* atypical. This illustrates the ability of this method to reduce the number of atypical curves fitted to the herd. This reduction of atypical curves follows from the incorporation of information from the rest of the herd as a result of the inclusion of the other parameters estimated for all animals in the estimation process through the Gibbs sampler. Although the matrix  $M_i$  estimates the regression coefficients relevant to animal  $i$  only, to estimate this matrix the Gibbs sampler requires that information from the whole herd be included in the conditional distribution of  $M_i$  through the parameters matrices  $B$ ,  $F^{-1}$ ,  $S^{-1}$  and  $R^{-1}$  contained in it. The values of these parameter matrices are also continually updated with every iteration of the Gibbs sampler.

However, it most often happens that we are interested not in the expected lactation or milk yield curve of a particular animal, but rather in the expected lactation curve of the herd with respect to one of the levels of a certain cofactor. For example, interest may be focused on the expected lactation curve for one of the seven parity groups. Instead of calculating  $BZ_i$  for each of the 10 000 simulated  $B$  matrices using the  $Z_i(17 \times 4)$  relevant specifically to animal  $i$  as was suggested in Alternative 1 above to obtain  $M_i$  for that specific animal  $i$ ,  $Z_i$  is replaced with  $z^*(17 \times 1)$  which only considers the various levels of the covariates for the one cofactor of interest and eliminates all other cofactors by taking the averages over the levels of these other cofactors as the values of their covariates in the vector  $z^*$ . No distinction among the four lactation cycles are now required because interest is now focused on a level of one of the cofactors only irrespective of lactation cycle, allowing, therefore, for a reduction in the number of columns contained in  $z^*$  to only one instead of the four contained in any of the matrices  $Z_i$ ,  $i = 1, \dots, 1141$ .

If, for example, the expected milk yield curve of parity 3 has to be determined, the cofactors regions, calving year and calving season have to be averaged over their respective levels.

Because the cofactor “region” has 7 levels, the average effect of each of its cofactor levels would be  $\frac{1}{7}$ , and therefore the levels of the 6 covariates used to identify the 7 cofactor levels are set equal to  $\frac{1}{7}$  in order to average out its effect. Similarly for the cofactor “calving year”, which has 4 cofactor levels identified by 3 covariates, the average effect of  $\frac{1}{4}$  is assigned to each covariate level to average out the effect of this cofactor. For the cofactor “calving season” with its 2 cofactor levels, the covariate used to identify its levels has to be assigned a  $\frac{1}{2}$  to average out its effect. To identify parity 3 in the covariate vector  $\mathbf{z}^*$  a one is still, as before, assigned to the second covariate, while all other parity covariates are set at 0. This implies that to determine the expected lactation or milk yield curve of parity 3:

$$\mathbf{z}^{*'} = [1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{7} \ \frac{1}{7} \ \frac{1}{7} \ \frac{1}{7} \ \frac{1}{7} \ \frac{1}{7} \ \frac{1}{4} \ \frac{1}{4} \ \frac{1}{4} \ \frac{1}{2}].$$

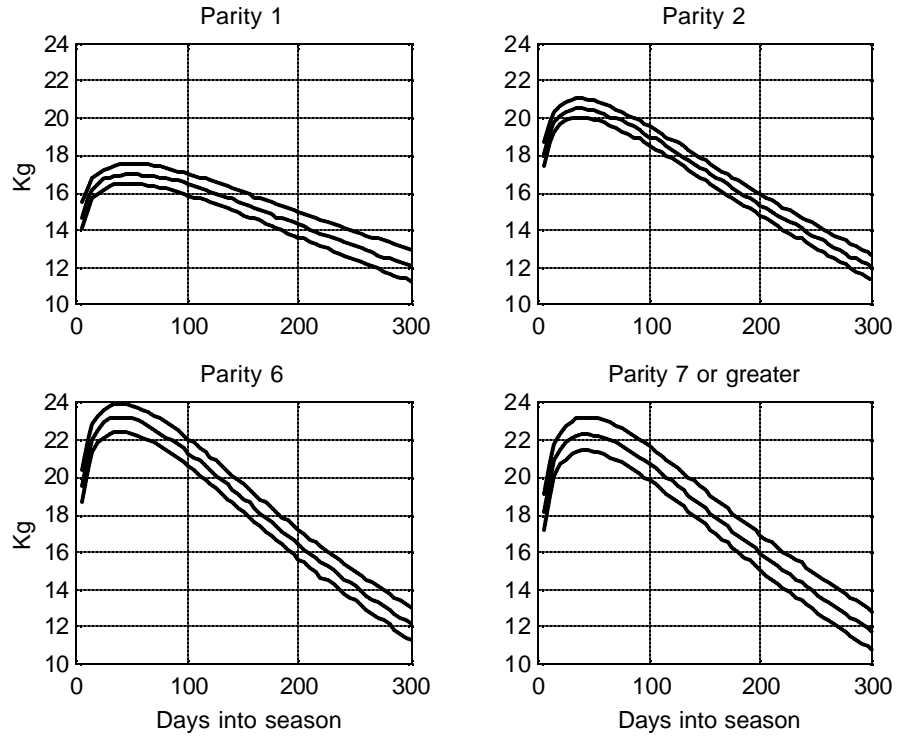
Using  $\mathbf{z}^*$  and each of the 10 000 simulated  $\mathbf{B}$  matrices, 10 000 vectors

$$\mathbf{m}^c = \mathbf{B}\mathbf{z}^* \tag{5.11}$$

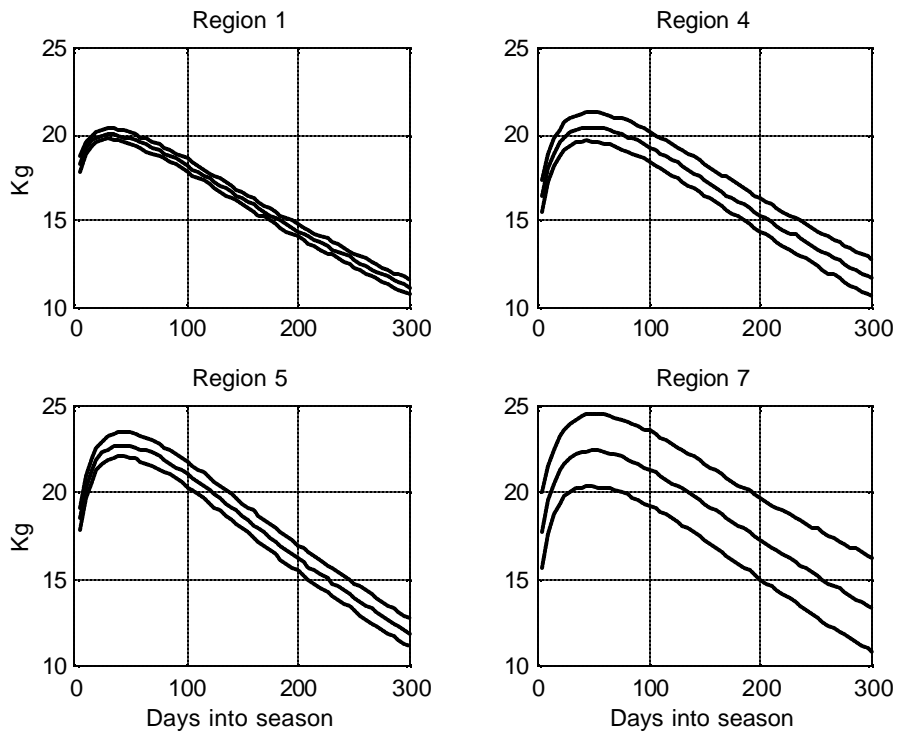
are calculated. The parameter values contained in  $\mathbf{m}^c$  related to milk yield are then used to construct the 10 000 lactation curves for the cofactor of interest. The mean of this set of 10 000 lactation curves then provides the expected lactation curve for a particular level of the cofactor of interest, in this case parity 3 of the cofactor parity, after all other cofactors have been eliminated.

In Figure 5.8 below the expected lactation curves together with the 95% HPD intervals for four of the seven parities when the other cofactors, “region”, “calving year” and “calving season”, have been averaged out, are given. Note that these curves are now valid with respect to the herd under consideration and do not refer to a specific animal. From this result it seems as if the lowest level of peak milk yield in this herd is attained in the first parity, with a sharp increase in peak milk yield from first to second parity, after which a steady increase is maintained until about the sixth parity, when peak milk again seems to start declining. The flattest lactation curve with lowest total milk yield seems to occur during the first parity, raising the suspicion that this parity is most persistent. It also appears as if the time of peak milk yield is later during the first parity than during subsequent parities.

When considering the expected lactation curves together with their 95% HPD intervals in Figure 5.9 for regions 1, 4, 5 and 7 when the other cofactors have been averaged out, it seems evident that there are differences among the regions with respect to level of peak milk yield and time of peak milk yield. Total yield in region 7 also seems higher with greater variability because of the much wider 95% HPD interval boundaries when compared to other regions.

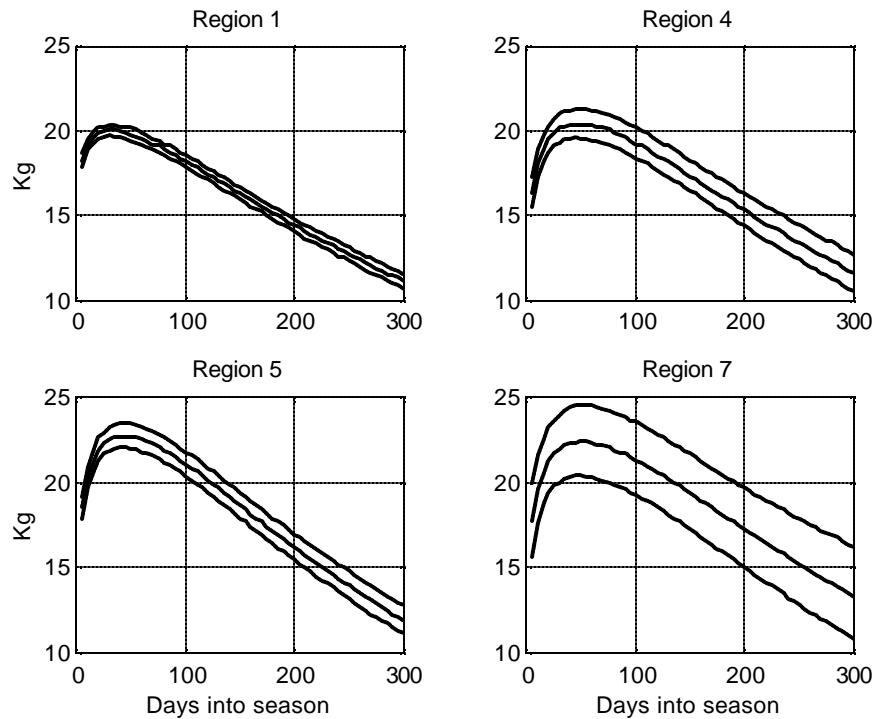


**Figure 5.8: Expected lactation curves with 95% HPD intervals for parities 1, 2, 6 and 7 or greater when other cofactors have been averaged out.**



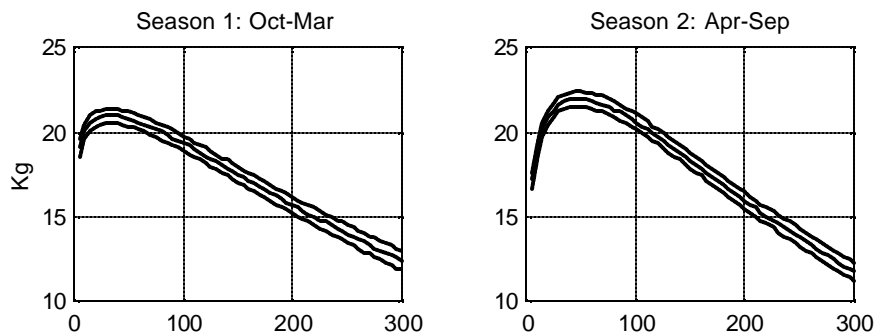
**Figure 5.9: Expected lactation curves with 95% HPD intervals for regions 1, 4, 5 and 7 when other cofactors have been averaged out.**

When comparing the expected lactation curves in Figure 5.10 for the four years under consideration after the other cofactors have been averaged out, it is evident that peak yield in 1996 was much lower than during any other year, but it seems as if very little differences occurred among the four years with respect to time of peak yield.



**Figure 5.10:** Expected lactation curves with 95% HPD intervals for 1995, 1996, 1997 and 1998 when other cofactors have been averaged out.

From Figure 5.11 in which the expected lactation curves with 95% HPD intervals for the two seasons are displayed, it is clear that during the cooler months of Season 2 peak milk yield, although attained slightly later, occurs at a higher level than is the case in Season 1. Total milk yield would be greater during the cooler months.



**Figure 5.11:** Expected lactation curves with 95% HPD intervals for seasons 1 and 2 when all other cofactors have been averaged out.

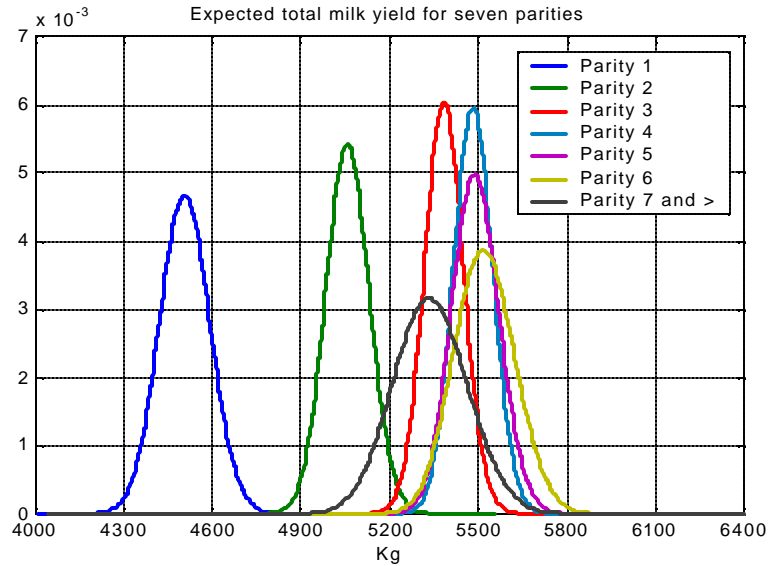


To confirm the above and other suspected results, it is necessary to further investigate the lactation curve characteristics as specified by Wood (1967) and given in equations (4.24) to (4.27) in Chapter 4.

### 5.5.1 Expected total milk yield

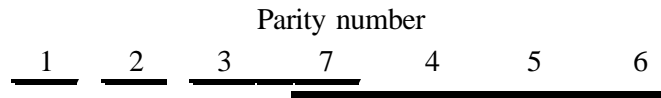
For the Wood model total milk yield was defined as  $f = \int_0^{300} t^b \exp(a + ct) dt$ . As explained in Section 4.3 in Chapter 4, by using the 10 000 simulated  $M_i$  matrices for each animal  $i$  over all four its lactation cycles, it would be possible to calculate this value  $f$  for each animal  $i$  during each of its four lactation cycles for each of the 10 000 simulations. As mentioned before not being able to store the 10 000 sets of the parameters  $M_i$ ,  $i = 1, \dots, 1141$ , requires either as first alternative estimating the parameters contained in  $\tilde{M}_i$  through calculating  $BZ_i$  for each of the 10 000 simulated  $B$  matrices, using the  $Z_i$  relevant to animal  $i$  or as second alternative regenerating  $M_i$  through Gibbs sampling as described earlier in Section 5.5. After reconstructing the 10 000 sets of the parameters contained in  $\tilde{M}_i$ , the total milk yield for an animal  $i$  may be determined for each of these sets of parameters. Once 10 000 values of total milk yield is obtained for animal  $i$  during a particular lactation cycle, a histogram is constructed for these values of  $f$  and a Pearson curve fitted to the histogram to determine the marginal posterior density of expected total milk yield.

As was the case with the expected lactation or milk yield curve, it most often happens that we require the expected total milk yield with respect to one of the levels of a certain cofactors, rather than the expected total milk yield of a particular animal. If this is the case, using  $z^*$  constructed as explained earlier in order to average out the other cofactors, together with each of the 10 000 simulated  $B$  matrices, 10 000 vectors  $m^c$  are obtained using equation (5.11). The parameter values for the lactation trait milk yield only as contained in  $m^c$  are then used to obtain 10 000 values of total milk yield,  $f$ , where only one level of the particular cofactor of interest applies to the result, while all other cofactors have been averaged out through the use of  $z^*$ . As before, a histogram is constructed for the values of  $f$  and a Pearson curve fitted to the histogram to determine the marginal posterior density of expected total milk yield for that cofactor level of interest. If this is done for each of the seven parity groups, the marginal posterior distribution of expected total milk yield for all seven parities are as follows:



**Figure 5.12:** Expected total milk yields for the seven parities when all other cofactors have been averaged out.

To investigate the differences in expected total milk yield for the seven parities when the other cofactors have been averaged out, 90% HPD intervals for the differences in expected total milk yield when parities are compared two at a time, were constructed. Such HPD intervals are obtained by sorting the results from the calculation of the differences in total milk yield for two levels, say  $i$  and  $j$ , of the cofactor, or  $f_i - f_j$ , in ascending order and then finding, as the lower and the upper limit of the 90% HPD interval respectively, the values in positions 500 and 9 500 for the 10 000 sorted values of  $f_i - f_j$ . The result of this investigation is summarised in Figure 5.13 below, where solid lines under parity numbers indicate groups of parities for which no significant differences in expected total milk yield were found.



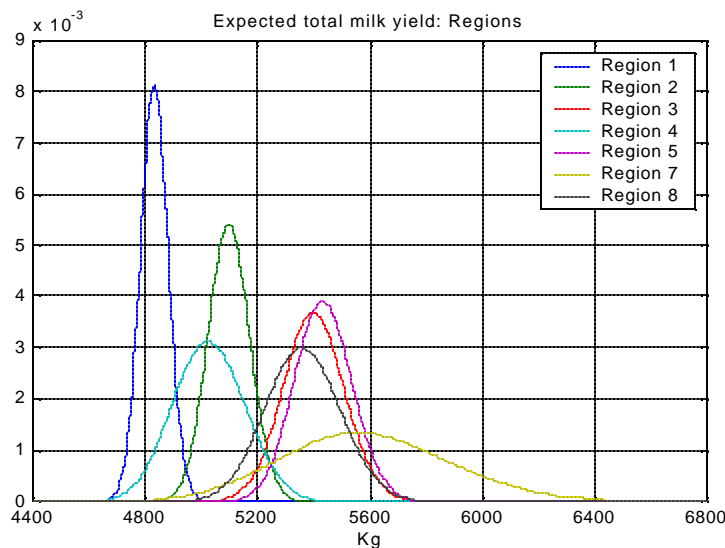
**Figure 5.13:** Summarised result of the comparison of expected total milk yields for the seven parities using 90% HPD intervals for differences.

Usually, when such comparison through the use of HPD intervals for differences is graphically illustrated in this manner, the sequence in which the compared characteristics occur is the same as that of the means of the marginal posterior distributions when listed in ascending order. If this had been the case for the parities in Figure 5.13, the sequence would have been parity 1, 2, 7 or >, 3, 4, 5, and 6. This is, however, not the case here. The

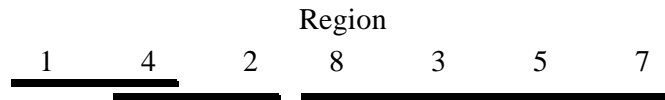
difference in the order in which the parities are listed in this illustration occurs as a result of the influence of the variances of the marginal posterior distributions of the expected total milk yield for the seven parities.

From Figures 5.12 it is evident that total milk yield in the first parity is the lowest as suspected earlier when the expected lactation curves were considered, but that it gradually increases with increasing parity number up until about the fourth parity, after which total milk yield remains relatively stable until the sixth parity. A decline in total milk yield was then recorded for parity seven or greater. As could be expected Figure 5.13 shows that no significant differences occur among the total milk yield levels of parities 4, 5 and 6, but surprisingly at the 90% HPD level parity group 7 or > also do not significantly differ from the afore mentioned parities. Although the total milk yield in parity 3 differs significantly from that in parities 4, 5, and 6, there is no significant difference on the 90% HPD level between parities 3 and 7 or >. Parities 1 and 2 differ significantly from all other parities on this level.

Using the same method as explained above, the expected total milk yields for the different levels of the other cofactors were also investigated. In each of these cases the levels of the cofactor that were of no interest were averaged out in order to obtain the expected total milk yield for the level of the particular cofactor of interest. To average out the effect of the cofactor “parity” with its 7 levels, the 6 covariates used to identify its levels had to be assigned a  $\frac{1}{7}$ .



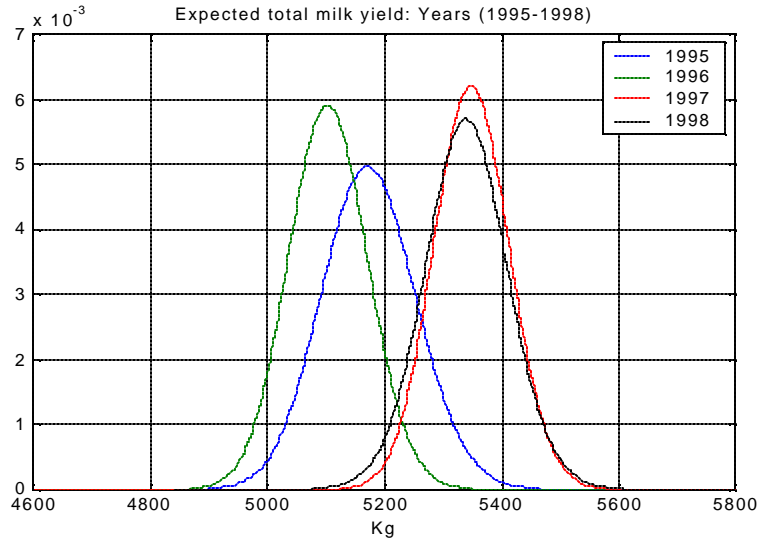
**Figure 5.14:** Expected total milk yields for the seven regions when all other cofactors have been averaged out.



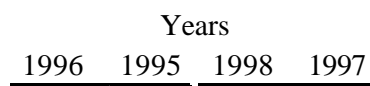
**Figure 5.15:** Summarised result of the comparison of expected total milk yields for the seven regions using 90% HPD intervals for differences.

The graph in Figure 5.14 also shows that the expected total milk yields are not the same in the seven regions. The lowest expected total milk yield resulted in region 1, while not only the highest expected total milk yield, but also the one with the greatest variability was found in region 7 as was suspected when the expected lactation curves for the regions were considered. From Figure 5.15 it is seen that no significant difference in expected total milk yield occurred between regions 1 and 4, regions 2 and 4, and among regions 3, 5, 7 and 8.

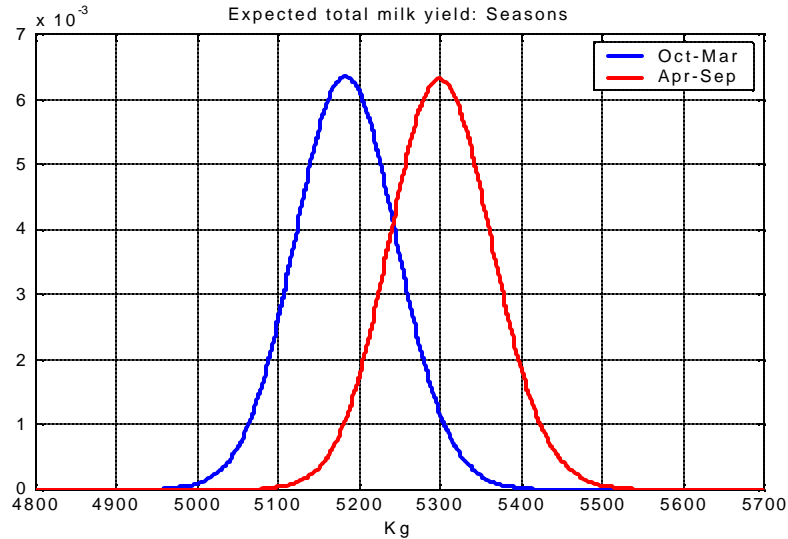
When considering the expected total milk yields for the four calving years, Figure 5.16 shows the lowest expected total milk yield to be that of 1996, while 1997 resulted in the largest value, with 1998 much the same as that of 1997. Figure 5.17 confirmed that 1997 and 1998 do not significantly differ with respect to expected total milk yield, but also that 1995 and 1996 are not significantly different from one another in this respect.



**Figure 5.16:** Expected total milk yields for the four calving years when all other cofactors have been averaged out.



**Figure 5.17:** Summarised result of the comparison of expected total milk yields for the four calving years using 90% HPD intervals for differences.



**Figure 5.18: Expected total milk yields for the two calving seasons when all other cofactors have been averaged out.**

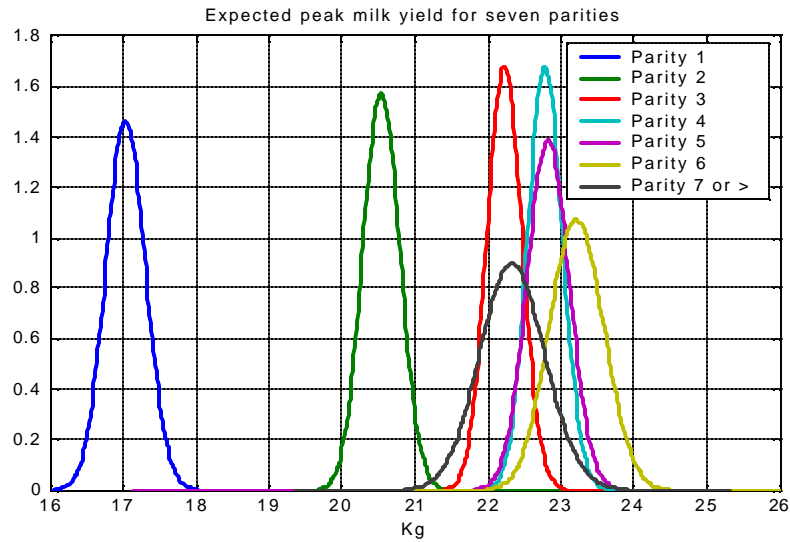
The densities in Figure 5.18 show, as suspected when the expected lactation curves of the seasons were considered, that higher expected total milk yield was found in the cooler months than in the warmer months. The result of the 90% HPD interval for the difference between the two seasons was [57,2038 ; 178,2050], indicating therefore that a significant difference in expected total milk yield during the two lactation season considered does exist.

### 5.5.2 Expected peak milk yield

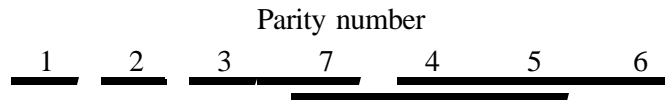
As mentioned earlier, Wood defined peak milk yield as  $h = \left(-\frac{b}{c}\right)^b \exp(a - b)$ . To investigate expected peak milk yield for the various levels of a cofactor, the same approach as explained above for expected total milk yield where the effect of the other cofactors are averaged out, is used.

Expected peak milk yield for the seven parities in Figure 5.19 was found to indeed be lowest in the first parity after which there is an increase in expected peak milk yield with every parity up to the sixth parity. Only in parity 7 or > did the expected peak milk yield again decrease to more or less the same level as that of parity 3, although with greater variability. Figure 5.20 shows that parities 1 and 2 are significantly different from all other parities with respect to expected peak milk yield, while parities 3 and 7 or >, parities 4, 5 and 7 or >, and parities 4, 5

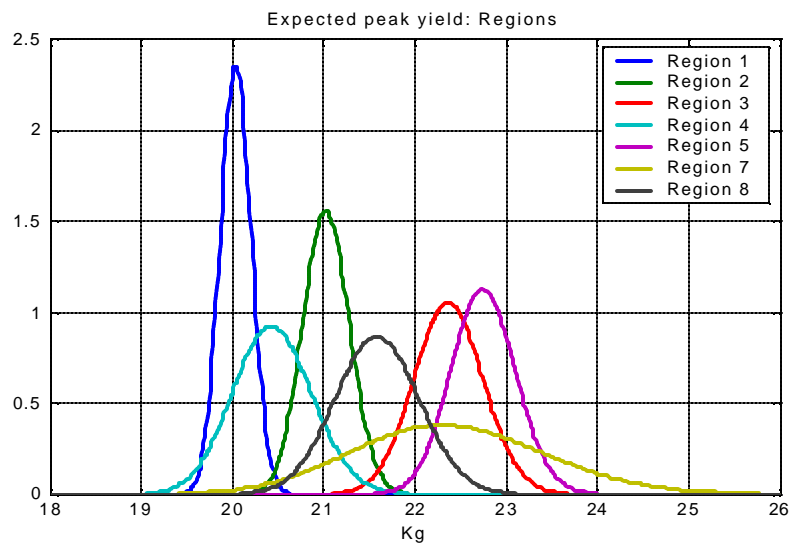
and 6 do not differ significantly when the 90% HPD intervals for the differences between parities are considered.



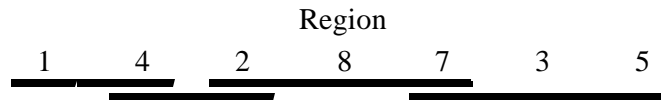
**Figure 5.19:** Expected peak milk yields for the seven parities when all other cofactors have been averaged out.



**Figure 5.20:** Summarised result for the comparison of expected peak milk yields of the seven parities using 90% HPD intervals for differences.



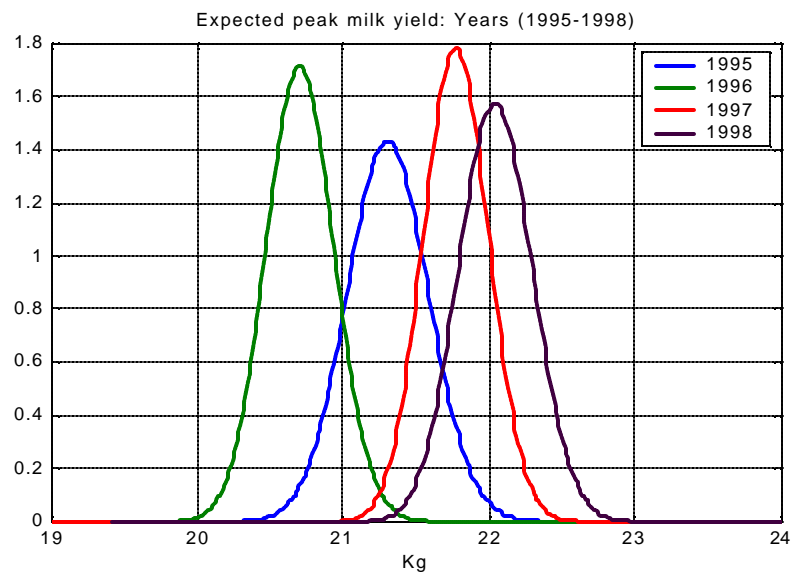
**Figure 5.21:** Expected peak milk yields for the seven regions when all other cofactors have been averaged out.



**Figure 5.22:** Summarised result of comparison of expected peak milk yields for the seven regions using 90% HPD intervals for differences.

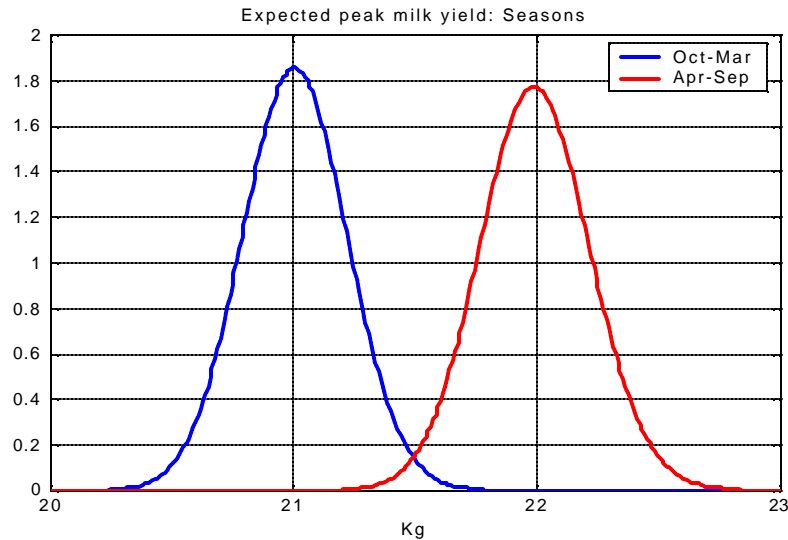
In Figure 5.21 it is evident that there are the differences among the regions with respect to expected peak milk yield. The lowest and least variable of the expected peak milk yields occurred in region 1, while the highest expected peak milk yield resulted in region 5. Region 7 a relatively high expected peak milk yield with the largest of variance all regions. From Figure 5.22 it is seen that the expected peak milk yield in regions 1 and 4, regions 2 and 4, regions 2, 7 and 8, and regions 3, 5 and 7 do not differ significantly.

From Figure 5.23 that considers the expected peak milk yield for the four years it is seen that there are considerable differences among the four years with respect to peak milk yield, 1996 having the lowest expected peak milk yield. Upon further investigation of these differences using 90% HPD intervals for differences between expected peak milk yields of years, in order to compare expected peak milk yield for the four years in pairs, all differences between years proved to be significant.



**Figure 5.23:** Expected peak milk yields for the four calving years when all other cofactors have been averaged out.

The result obtained when comparing the expected milk yields of the two seasons for which the marginal densities are given in Figure 5.24 also indicated significant difference between the two seasons, with April to September indeed having higher expected peak milk yield.



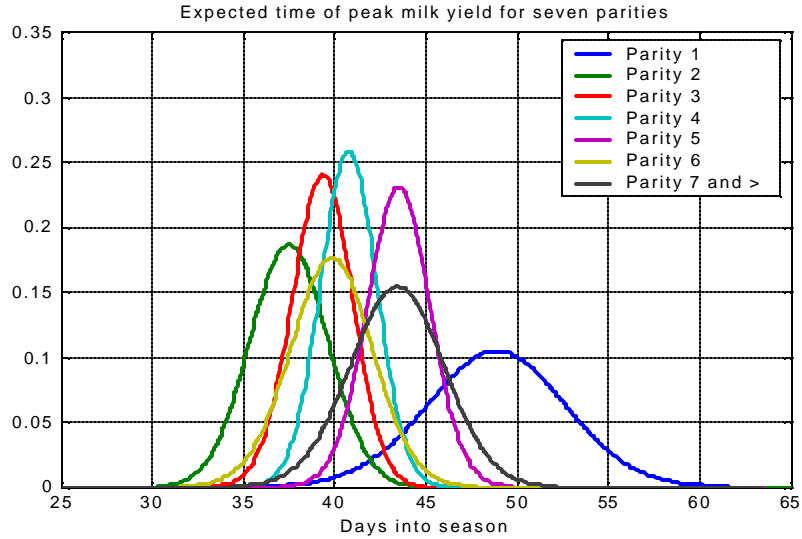
**Figure 5.24:** Expected peak milk yields for the two calving seasons when all other cofactors have been averaged out.

### 5.5.3 Expected time of peak milk yield

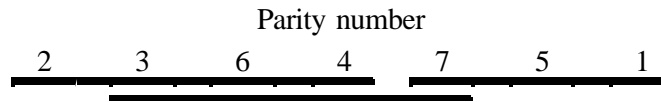
From the Wood model it follows that time of peak yield is defined as  $t = -\frac{b}{c}$ . To study expected time of peak yield for the herd at various levels of a certain cofactor, the approach explained earlier for expected total milk yield was again used. The effect of cofactors that are of no interest at a particular time is, as before, averaged out through the use of the covariate vector  $z^*$  as explained earlier.

Expected time of peak milk yield was found in Figure 5.25 to be latest in the first parity, as mentioned when the expected lactation curves for parity were considered, and earliest in the second parity after which it becomes gradually later with each parity up until the fifth parity. Expected time of peak yield for parity 6 was midway between that of parities 3 and 4, while that of parity 7 or > is similar to parity 5. Note that parity 1 has the greatest variability with respect to expected time of peak milk yield. From Figure 5.26 it is seen that parities 2, 3, 4 and 6, parities 3, 4, 6 and 7 or >, and parities 1, 5 and 7 or > do not differ significantly with respect to expected time of peak milk yield.

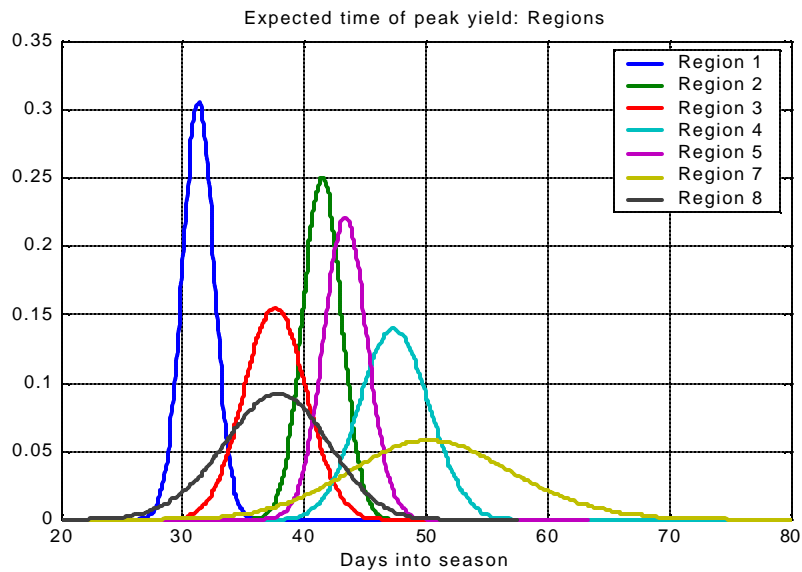




**Figure 5.25:** Expected time of peak milk yields for the seven parities when all other cofactors have been averaged out.

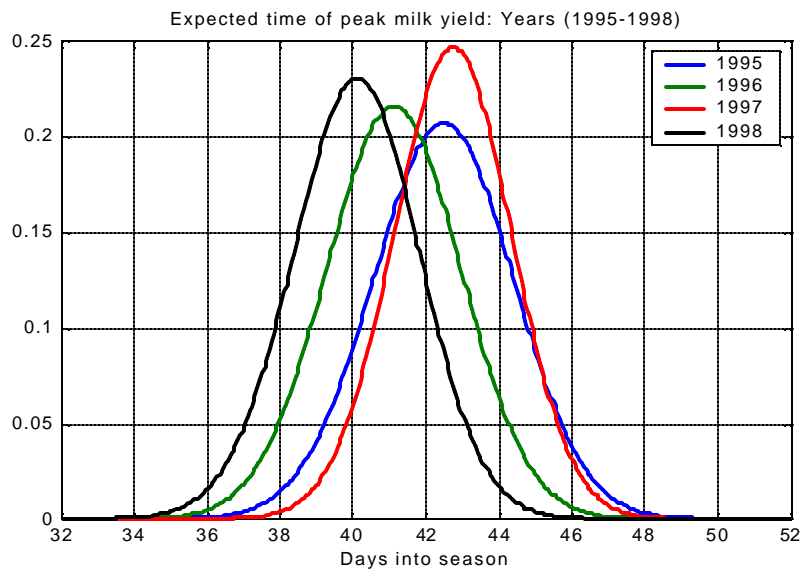


**Figure 5.26:** Summarised result of comparison of expected time of peak milk yields for the seven parities using 90% HPD intervals for differences.

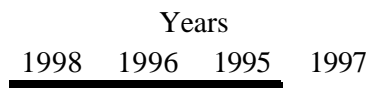


**Figure 5.27:** Expected time of peak milk yields for the seven regions when all other cofactors have been averaged out.

From the comparison of the marginal posterior distributions of time of peak milk yield using 90% HPD intervals of differences between time of peak milk yield for regions, it follows that region 1 differs significantly from all regions except region 8. Further more, as could be expected regions 2, 3, and 8, regions 2, 5, and 7, regions 4, 5, and 7, and regions 5, 7 and 8 do not differ significantly with respect to expected time of peak milk yield. Region 8 only differs significantly from region 4 on the 90% HPD level, which is more as a result of the variance of the distribution of region 8, than because of its location parameter. It was not possible to graphically summarise the result of comparison of expected time of peak milk yields for the seven regions by means of a line diagram as was done above, because of the influence of the variances of the above marginal posterior distributions. As was mentioned earlier the sequence of regions in such a diagram would normally be the same as that of the means of the marginal distributions when sorted in ascending order. Here large differences in variances of the marginal posterior distributions make this of form of illustration impossible, as it is not possible to arrange the regions in any illustratable order.

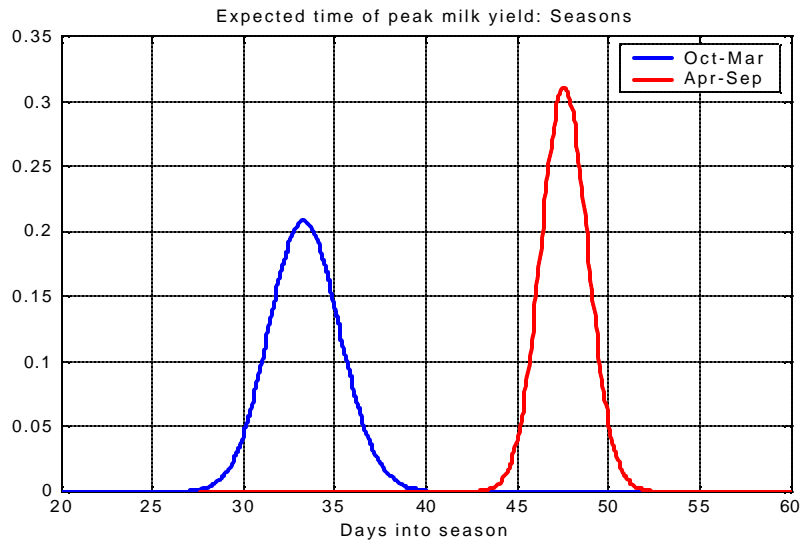


**Figure 5.28:** Expected time of peak milk yields for the four calving years when all other cofactors have been averaged out.



**Figure 5.29:** Summarised result of comparison of expected time of peak milk yields for the four calving years using 90% HPD intervals for differences.

From Figures 5.28 and 5.29 it follows that only 1997 and 1998 differ significantly with respect to time of peak milk yield. Furthermore, when Figure 5.30 is also considered greater distributional differences seem to occur between the two calving seasons than among calving years with respect to expected time of peak milk yield. As was suspected when the expected lactation curves for the two seasons were considered, expected peak milk yield occurs later during the cooler months of April to September.



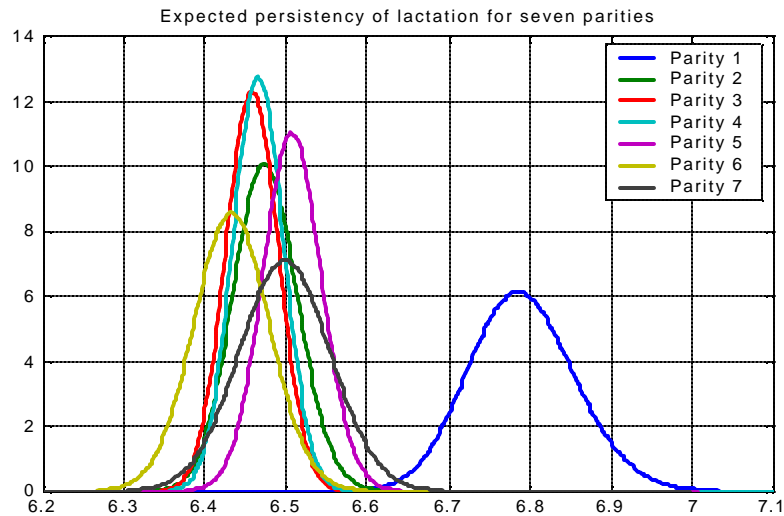
**Figure 5.28:** Expected time of peak milk yields for the two calving seasons when all the r cofactors have been averaged out.

### 5.5.4 Expected persistency of lactation

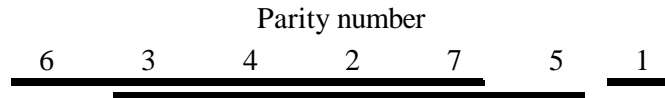
The Wood definition of persistency of lactation is  $y = -(b + 1) \ln(-c)$  and is said to describe the ability of an animal to maintain peak milk yield. The expected persistency of lactation for the various levels of a cofactor is again investigated using the same approach as described earlier for expected total milk yield, where the effect of the other cofactors are averaged out using the appropriate form of the covariate vector  $z^*$  as previously explained.

Figure 5.31 shows the level of expected persistency to be much greater during the first parity than is the case for later parities. This was expected because of the flatter expected lactation curve that resulted during the first parity when compared to later parities. When the summary of the 90% HPD intervals of differences in persistency for parities in Figure 5.32 is considered we see that parity 1 differs from all other parities with respect to expected

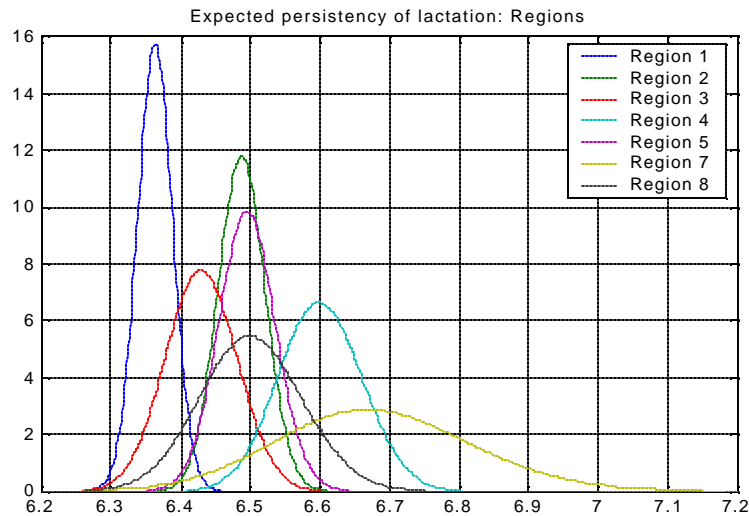
persistence. The only other significant difference that occurs is found between parities 5 and 6, with parity 6 having the lowest persistency of all.



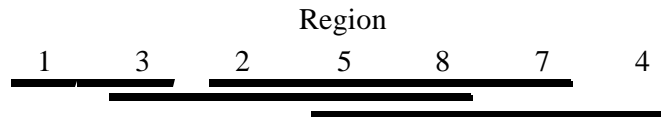
**Figure 5.31:** Expected persistency of lactation for the seven parities when all other cofactors have been averaged out.



**Figure 5.32:** Summarised result of comparison of expected persistency of lactation for the seven parities using 90% HPD intervals for differences.

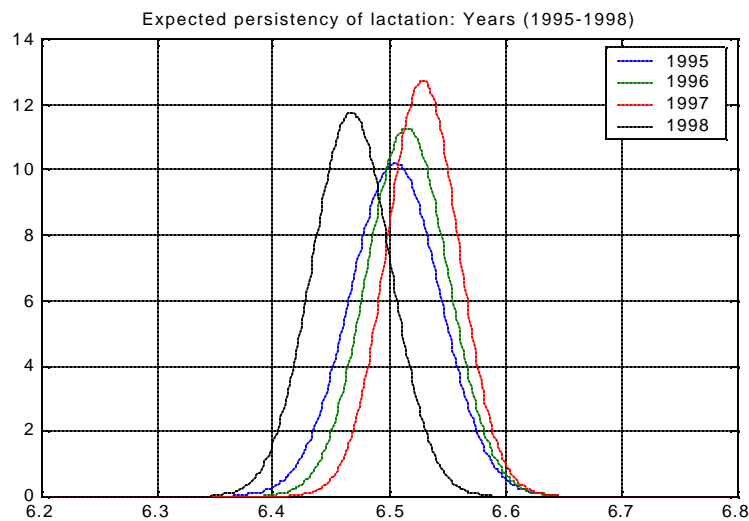


**Figure 5.33:** Expected persistency of lactation for the seven regions when other cofactors have been averaged out.

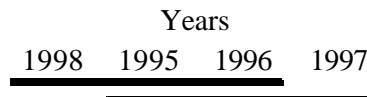


**Figure 5.34:** Summarised result of comparison of expected persistency of lactation for the seven regions using 90% HPD intervals for differences.

When considering Figures 5.33 and 5.34 it is seen that the region with the lowest expected persistency is region 1, while region 7 not only has the highest level of expected persistency, but also the greatest variability in persistency. Because of this large variance in region 7, the sequence in which the regions are listed in the summary of results for the 90% HPD intervals of differences in persistency for regions in Figure 5.34 do not follow the sequence of the means of the marginal posterior distributions when listed in ascending order. From Figure 5.34 it is also evident that regions 1 and 3, regions 2, 3, 5 and 8, regions 2, 5, 7 and 8, and regions 4, 5, 7 and 8 do not differ significantly with respect to expected persistency.

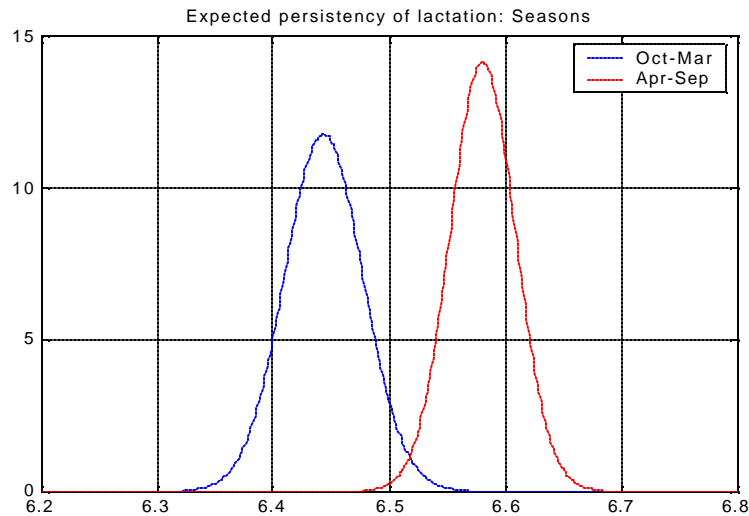


**Figure 5.35:** Expected persistency of lactation for the four calving years when other cofactors have been averaged out.



**Figure 5.36:** Summarised result of comparison of expected persistency of lactation for the four calving years using 90% HPD intervals for differences.

Similar to the results of expected time of peak yield, it follows from Figures 5.35 and 5.36 that only 1997 and 1998 differ significantly with respect to expected persistency of yield.



**Figure 5.37: Expected persistency of lactation for the two calving seasons when other cofactors have been averaged out.**

In Figure 5.37 the expected persistency of lactation differs considerably for the two calving seasons, with greater persistency occurring during the cooler season that spans from April to September. The result of the 90% HPD interval for the difference between the two seasons was  $[-0,1737 ; -0,1008]$ , also indicating that a significant difference in expected persistency between the two seasons does exist.

### 5.5.5 Milk yield: In summary

It should be noted that the expected lactation or milk yield curves generated for each of the levels of the various cofactors when the effects of the other cofactors are eliminated through averaging, provide the Bayesian equivalent to the Standard lactation curves (SLAC's) of milk yield for that level of the cofactor, as discussed in section 3.2 of Chapter 3, with respect to this particular herd. The expected lactation curve for parity 3, for example, should therefore also be viewed as the herd standard with respect to lactation for all animals in their third parity.

The isolated effects of the various levels of the four cofactors parity number, region, calving year and calving season, on the four characteristics total milk yield, peak milk yield, time of

peak milk yield and persistency of the lactation trait milk yield is summarised here. How these results compare to those of other researchers of lactation in cattle are also mentioned.

For the seven parities under consideration the following were found. Parity 1 had the lowest expected total and expected peak milk yield, together with the latest expected time of peak yield and the greatest variance with respect to expected time of peak yield. Parity 1 had the flattest lactation curve as a result of having the lowest peak and total milk yield, and the latest time of peak yield and therefore had a greater expected persistency than any other parity. Expected persistency in parity 1, however, also had the greatest variance. Parity 1 differed significantly from all other parities with respect to expected total and expected peak milk yield, as well as expected persistency. Wood (1969), Rowlands et al. (1982), Sölkner and Fuchs (1987), Stanton et al. (1992) and Tekerli et al. (2000) all found that, in dairy cows, persistency of lactation is greater during the first parity than during later parities. Tekerli et al. (2000) also reported lower total and peak milk yields during first than during later parities. For the Jersey data parity 2 had the second lowest expected total and expected peak milk yields, with the earliest expected time of peak yield. Parity 2 differed significantly from all other parities with respect to both expected total and expected peak milk yield. Other studies of lactation by Rowlands et al. (1982), Keown et al. (1986) also found time of peak yield to be earlier in the second parity than in the first, while Keown et al. (1986) extended this result to time of peak in parity 3 being later than in parity 2. For the Jersey data both the characteristics expected total and expected peak milk yield increased with parity increasing parity number up to parity 6. Expected total milk yield in parity 7 or greater was less than that of parity 3. The expected peak yield of parity 7 or greater was just higher than parity 3, while expected time of peak yield for parity 7 or greater was similar to that of parity 5. Expected persistency of parities 5 and 7 or greater were similar although these parities were a lot less persistent than the first parity. Parity 6 was the least persistent, while parities 2, 3 and 4 were similar with respect to persistency. Furthermore, significant differences also exist between groups of parities with respect to all characteristics of lactation as described earlier.

The effects of the seven regions on lactation indicated that region 1 has the lowest expected total and expected peak yield, the earliest expected time of peak yield and, as a result of this, the lowest expected persistency. In contrast to region 1, region 7 had the highest expected total milk yield, relatively high expected peak milk yield and the latest expected time of peak yield and therefore the highest expected persistency. For all four the afore mentioned

characteristics, however, region 7 had the greatest variance. Region 4 also performed well with respect to expected persistency. It had the second highest level of persistency, that followed from the second lowest expected total and expected peak milk yield levels, and the second latest expected time of peak yield after region 7. Therefore, although region 4 is not the highest producing region, it produces milk at a highly persistent level. Significant differences exist between regions with respect to all characteristics of lactation as described earlier. Batra (1986) also noted that the influence of different stations as location of herds had a significant effect on lactation.

As a result of some years being dry and others wet, it was expected that calving years would vary with respect to their influence on lactation. The lowest expected total and expected peak yields, and second latest expected time of peak yield were recorded in 1996. With respect to expected persistency, however, 1996 was second best. The years 1997 and 1998 had the highest and second highest expected total milk yield, and the second highest and highest expected peak yield respectively. However, 1998 had the earliest expected time of peak yield, while 1997 had the latest time of peak yield. As a result, 1998 was the least persistent of the four years, while 1997 was most persistent. The year 1995 never fared the best or worst with respect to any of these characteristics. The years 1997 and 1998 differed significantly with respect to all the considered characteristics except expected total milk yield. Other significant differences for the characteristics under consideration were as discussed earlier. Other researchers of lactation, such as Batra (1986) and Tekerli (2000), also found calving year to have a significant influence on lactation.

Season 2, representing the cooler months from April to September, performed better with respect to all characteristics of lactation. This season had greater levels of expected total and expected peak milk yields, peaked later and had greater expected persistency of yield. Differences between the two seasons with respect to all characteristics considered were all significant. This concurs with results published by, amongst others, Sölkner and Fuchs (1987), Stanton et al (1992) and Tekerli et al. (2000) who found calving season to significantly influence lactation. Tekerli et al. (2000) found fall and winter to have higher total milk yield, while both Stanton et al. (1992) and Tekerli et al. (2000) found higher levels of peak milk yield during colder periods. Sölkner and Fuchs (1987) and Tekerli et al. (2000) reported season of calving to be of considerable influence on persistency of lactation.



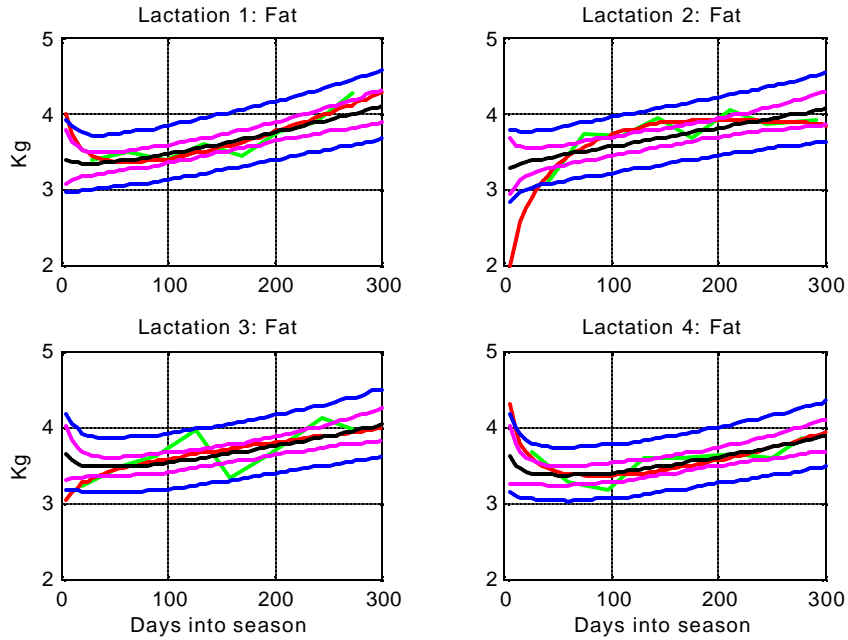
## 5.6 THE LACTATION COMPOSITION TRAITS

Both fat and protein content of milk in the Jersey data were measured as the percentage contained in milk composition and are typically a convex function over time. In 1976 Wood showed that the model  $E(W_i) = t^b \exp(a + ct)$  could also be used to estimate various milk composition traits such as fat and protein content in the milk of dairy cows. The reason for this may be found in the fact that the Wood model becomes convex over time when the value of  $b$  is negative and that of  $c$  is positive. Sakul and Boylan (1992) found that the Wood model is useful in describing almost all milk composition traits, with the exception of lactose content, in sheep breeds. In this study the Wood model was used to *simultaneously* model not only milk yield, but also the milk composition traits fat content and protein content. For each animal  $i$  the matrix  $M_i(12 \times 3)$  therefore contains the parameters of the Wood model for all three lactation traits over all four lactation cycles under consideration. Each column of  $M_i$  contains the parameters of one of the lactation traits for all four lactation cycles, where the parameters of the lactation cycles are listed as column vectors, one below the other.

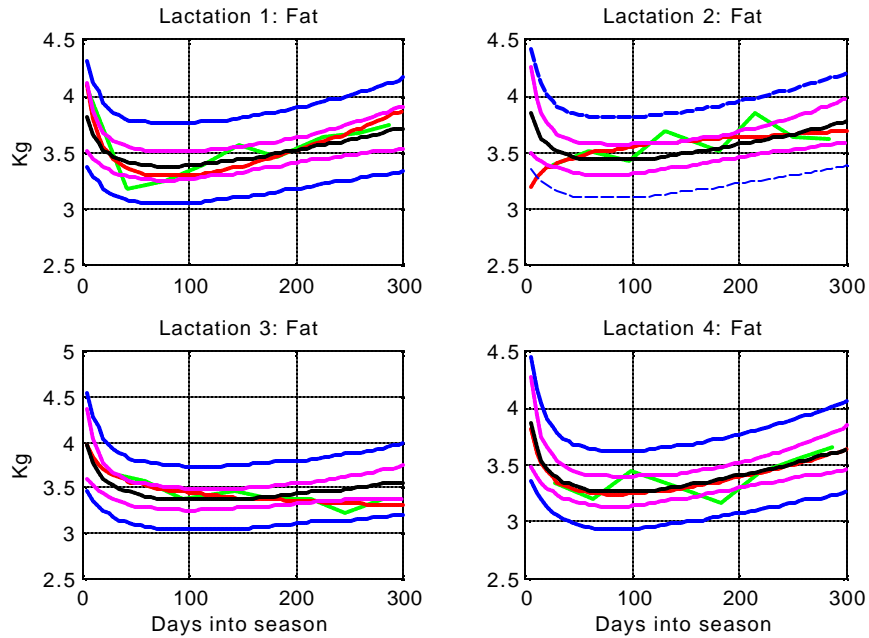
### 5.6.1 Fat content of milk

Once the 10 000 simulated  $M_i$  matrices for each animal  $i$  over all four its lactation cycles have been obtained, 10 000 fat content curves for animal  $i$  during each of its four lactation cycles can be generated. The mean of the 10 000 fat content curves for animal  $i$  in each lactation cycle then gives the expected fat content curve for animal  $i$  in that lactation cycle. Because, as a result of storage constraints, no  $M_i$  matrices were stored, either one of the two alternatives described earlier in section 5.5 may again be used to regenerate  $M_i$ . In Alternative 1 the values of the parameters contained in  $\tilde{M}_i(9 \times 4)$  are estimated by calculating  $BZ_i$  for each of the 10 000 simulated  $B(9 \times 17)$  matrices, using the  $Z_i(17 \times 4)$  relevant to animal  $i$ . While using Alternative 2, 10 000  $M_i$  matrices are obtained through Gibbs sampling using the 10 000 previously simulated  $B, F^{-1}, S^{-1}$  and  $R^{-1}$  matrices.

Regenerating 10 000  $M_i$  matrices by means of the second alternative for two animals ( $i = 135$  and  $i = 511$ ) and using the parameters in  $M_i$  relevant to fat content only to find 10 000 fat content curves, the expected fat content for each of the four lactation cycles for animals 135 and 511 were obtained as the mean of the 10 000 fat content curves and graphed as black solid lines in Figures 5.38 and 5.39.



**Figure 5.38:** For each of the four lactation cycle of Animal 135 the expected fat content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.



**Figure 5.39:** For each of the four lactation cycle of Animal 511 the expected fat content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 511 is given by —.

In Figures 5.38 and 5.39 the green line represents the observed fat content values of the test day records connected by straight-line segments for the four lactation cycles of the two animals. The red line is again the result of least squares estimation of the fat content curve fitted to the data. The magenta coloured broken lines provide the 90% HPD intervals for the expected fat content curve during each lactation cycle. The blue broken lines in turn represent a 90% prediction interval for the expected fat content curve during each lactation cycle obtained as the 90% HPD intervals for predictions resulting from the use of the predictive density that follows from (5.10) to simulate a predictive matrices  $Y_i^f$  from the normal distribution denoted in (5.10) for each of its 10 000 means,  $X_i M_i$ , and variances,  $F \otimes I_{n_i}$  as explained for milk yield in section 5.5 above.

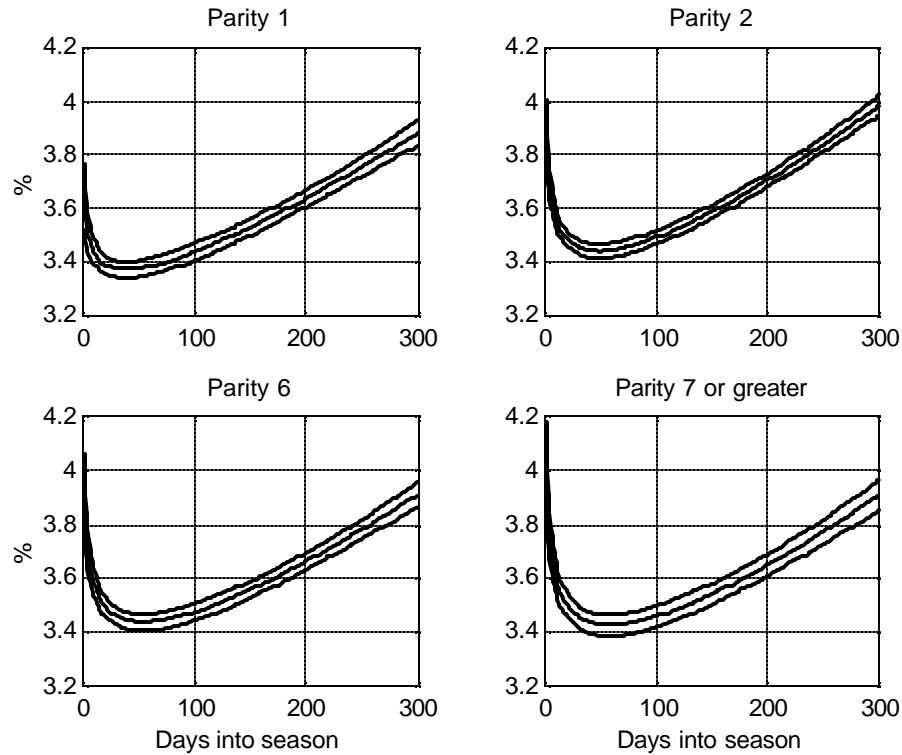
In the case of fat and protein content in milk composition when measured as percentages of composition, these functions are typically convex over time, and therefore curves for these traits of non-convex form are considered atypical. For this reason when considering the least squares estimate of the fat content curve for the second and third lactation cycles of animal 135 and the second lactation cycle of animal 511 these estimates are said to take on atypical forms. In contrast to this the Bayesian curve estimate of the expected fat content for both third lactation cycle of animal 135 and the second lactation cycle of animal 511 are *not* atypical, while that of the second lactation cycle of animal 135 is considered “less” atypical than its least squares counterpart. This again illustrates the ability of the Bayesian method to reduce the occurrence of curves of atypical nature as a result of the inclusion of information from the whole herd through the use of the parameter matrices  $B$ ,  $F^{-1}$ ,  $S^{-1}$  and  $R^{-1}$  in the conditional distribution of  $M_i$  in the Gibbs sampler. These parameter matrices,  $B$ ,  $F^{-1}$ ,  $S^{-1}$  and  $R^{-1}$ , are updated in every iteration of the Gibbs sampler.

It is worthwhile to note that atypical least squares estimates of the fat and protein content curves sometimes occur as a result of the first test day record occurring relatively late in the lactation cycle. A good example of this would be first test day record of the second lactation cycle of animal 135 that only occurs on day 39 of the cycle. Because a large portion of the lactation cycle that precedes this point in time is unknown or, more to the point, the portion where the greatest changes in fat content occurs is unknown, this is most probably the reason for the least squares estimate of the lactation curve being atypical in form. The least squares estimate of the curve is only based on the actual observations, while the Bayesian estimate of

the expected fat content curve is supported by the information from the rest of the herd that, to a large extent, corrects the atypical result obtained by least squares estimation.

As was the case with milk yield, interest is most often not focussed on the fat content curve of an individual animal, but on the expected fat content curve of the herd with respect to one of the levels of a certain cofactor. For this reason the same procedure explained earlier for of milk yield is again employed. This means calculating the 10 000 vectors  $m^c = Bz^*$  where  $z^*(17 \times 1)$  only considers one particular level of the cofactor of interest and eliminates all other cofactors by taking their averages as the values of their respective covariates in the vector  $z^*$ . The parameter values contained in  $m^c$  related to fat content are then used to construct the 10 000 fat content curves of the level of interest for the cofactor under consideration. The mean of this set of 10 000 fat content curves then provides the expected fat content curve for a particular level of the cofactor of interest after all other cofactors have been eliminated through averaging.

The expected fat content curves together with the 95% HPD intervals for four of the seven parities when the other cofactors, “region”, “calving year” and “calving season”, have been averaged out, are given in Figure 5.40. From this result the convex nature of the fat content curves are evident. Parity 1 had the lowest expected percentage of fat through out lactation, while for parities 2 through to 6 the expected percentage of fat in milk composition remained much the same until approximately day 70 when slight differences began to occur. By day 300 milk in parity 2 contained the highest expected percentage of fat. For parity 7 or greater the expected minimum percentage of fat occurred the latest, only on day 58. The remainder of the results for expected percentage of fat in milk composition for the parities are summarised in Table 5.1 below.



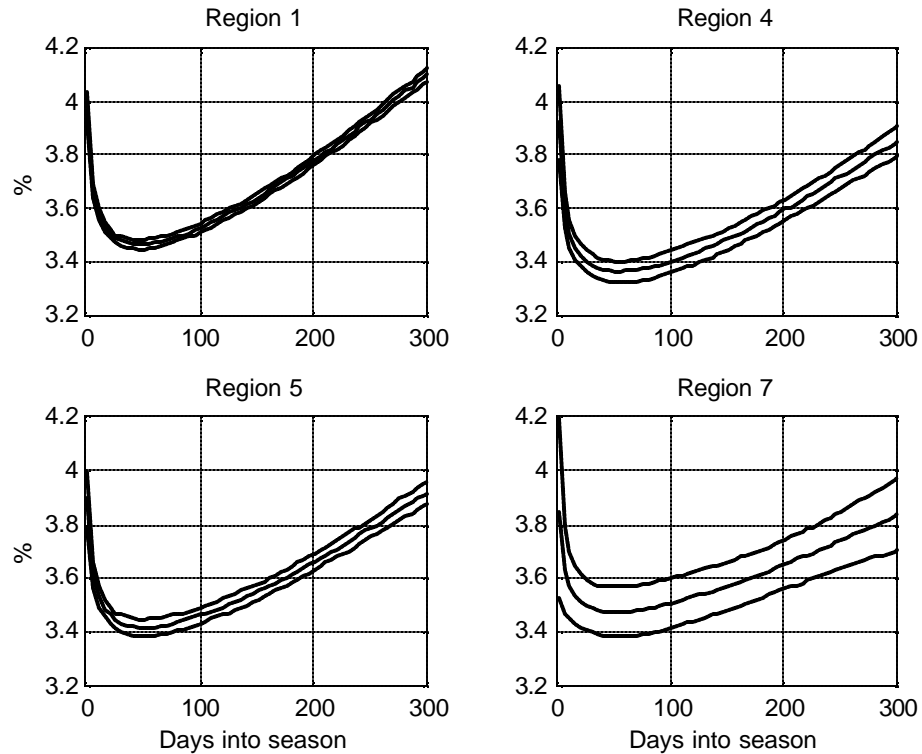
**Figure 5.40:** Expected fat content curves with 95% HPD intervals for parities 1, 2, 6 and 7 or greater when all other cofactors have been averaged out.

**Table 5.1:** Summary of the expected fat content for the seven parities

Parity	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1	3,6458	3,3719	38	3,8821
2	3,9093	3,4412	49	3,9848
3	3,9153	3,4420	50	3,9731
4	3,9956	3,4409	54	3,9718
5	3,9688	3,4434	54	3,9370
6	3,9332	3,4377	54	3,9109
7 or >	4,0274	3,4265	58	3,9088

In Figure 5.41 the expected fat content curves together with the 95% HPD intervals for four of the seven regions are given. For all regions the level of expected fat content at on day 1 and day 300 of lactation are very similar. Minimum expected fat content for all regions occurred

between day 46 and 56, with the greatest difference between regions with respect to minimum expected fat percentage only 0,1130%. The variance in expected fat content was smallest for region 1 and largest for region 7, which was also the case for milk yield in these regions.

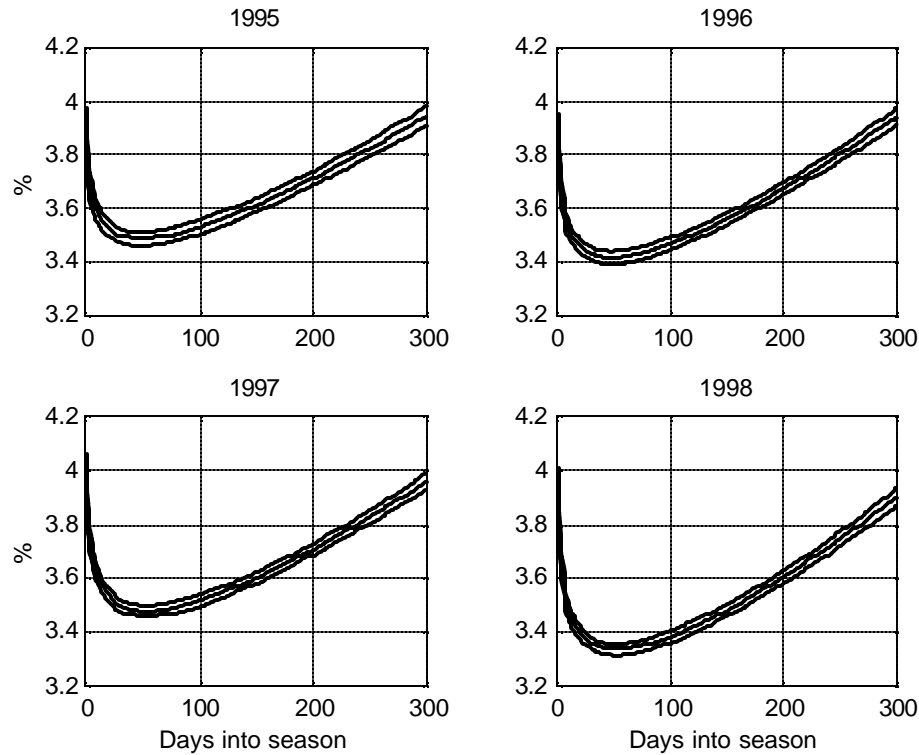


**Figure 5.41:** Expected fat content curves with 95% HPD intervals for regions 1, 4, 5 and 7 when all other cofactors have been averaged out.

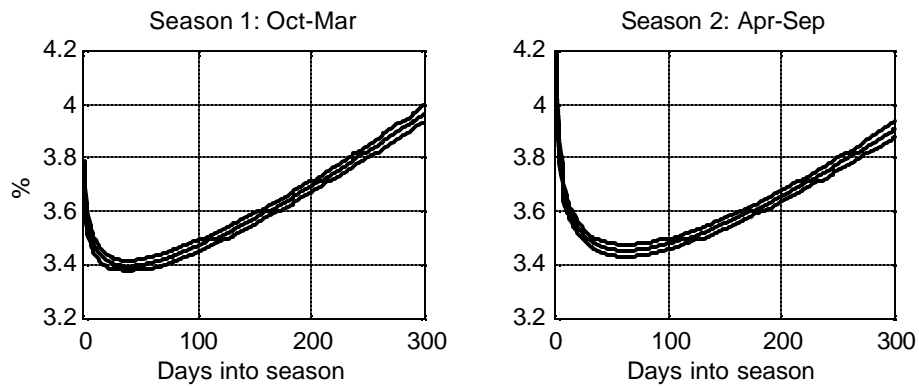
Figure 5.42 and Table 5.2 considers the expected fat content for the four years. The day of minimum expected fat content only varied by 5 days between the earliest and latest day for the four years, while the difference in minimum expected fat percentage was less than 0,15%.

**Table 5.2: Summary of the expected fat content for the four years**

Year	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1995	3,8789	3,4869	48	3,9468
1996	3,8713	3,4184	49	3,9437
1997	3,9790	3,4792	53	3,9592
1998	3,9185	3,3371	53	3,9032



**Figure 5.42: Expected fat content curves with 95% HPD intervals for 1995, 1996, 1997 and 1998 when all other cofactors have been averaged out.**



**Figure 5.43: Expected fat content curves with 95% HPD intervals for seasons 1 and 2 when all other cofactors have been averaged out.**

From Figure 5.43 it is evident that the expected fat content with respect to the two seasons when all other cofactors have been averaged out, differed quite considerably with respect to day of minimum fat content. For season 1 this minimum occurred on day 39, while for season 2 it occurred only on day 63. The expected minimum fat percentage, however, did not differ as much. Minimums of 3,3982% and 3,4535% were recorded for seasons 1 and 2 respectively. The expected fat content on day 300 were similar for the two seasons, 3,9671

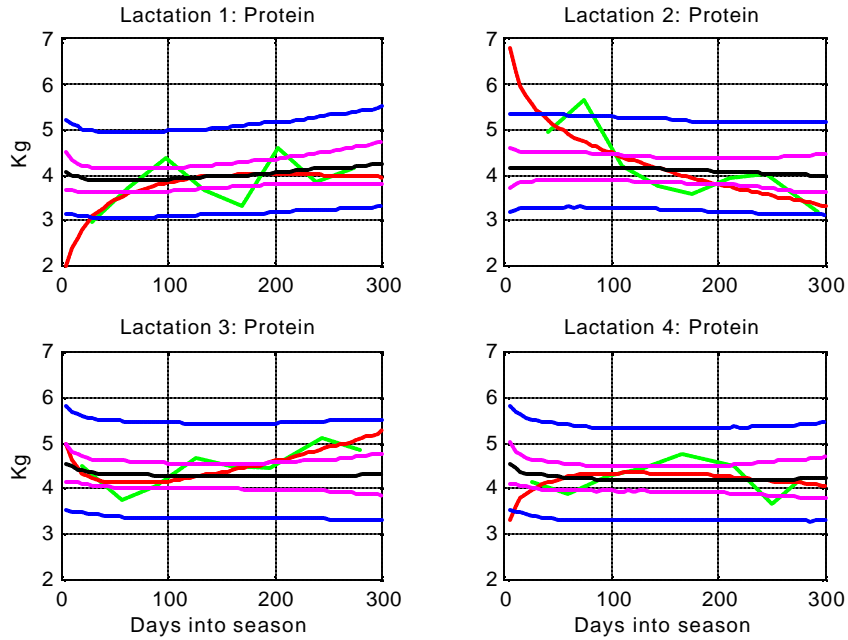
for season 1 and 3,9094 for season 2, although it differed more on day 1, with 3,7136% in season 1 and 4,1202% in season 2.

### 5.6.2 Protein content of milk

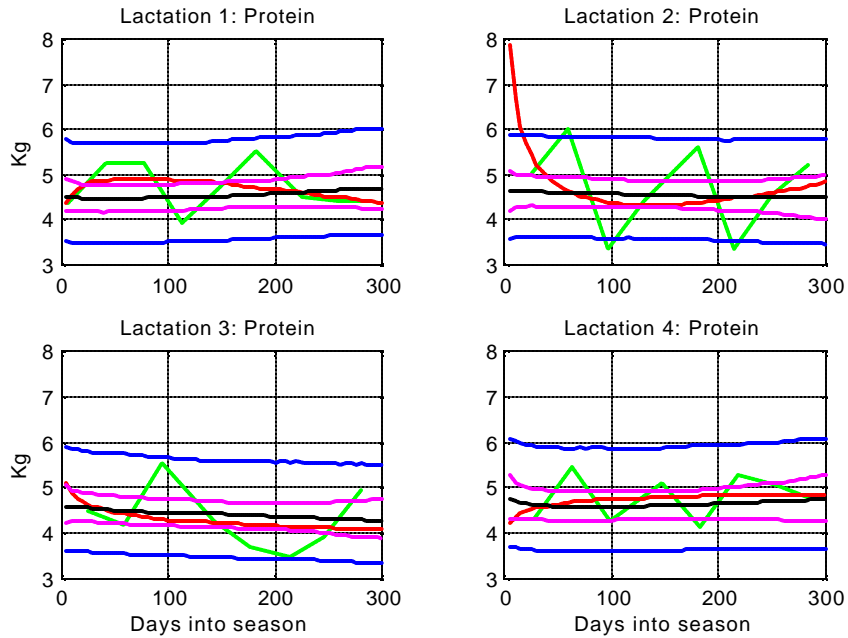
Using the same procedure as described for fat content in section 5.6.1, it is possible to obtain the expected protein content curves for any animal  $i$ . Therefore the expected protein content curves for each of the four lactation cycles of animals 135 and 511 were again obtained as the mean of 10 000 protein content curves that resulted from the parameters relevant to protein for the 10 000  $M_i$  matrices regenerated using the second alternative. These expected protein content curves are graphed as black solid lines in Figures 5.44 and 5.45 below. The green line represents the observed protein content values of the test day records connected by straight-line segments for the four lactation cycles of the two animals. The red line is again the result of least squares estimation of the protein content curve fitted to the data. The magenta coloured broken lines provide the 90% HPD intervals for the expected protein content curve during each lactation cycle. The blue broken lines in turn represent a 90% prediction interval for the expected protein content curve during each lactation cycle, obtained as the 90% HPD intervals for predictions resulting from the use of the predictive density through simulation as explained for milk yield in section 5.5 above. When considering the first and fourth lactation cycles of both animals 135 and 511 the ability of the Bayesian method to reduce the occurrence of atypical curves is again evident.

Using the same procedure explained earlier for milk yield and fat content curves, the expected protein content curve for a particular level of the cofactor of interest when all other cofactors have been averaged out is again obtained. The expected protein content curves together with the 95% HPD intervals for four of the seven parities after the other cofactors, “region”, “calving year” and “calving season”, have been averaged out, are given in Figure 5.46. From this result it is evident that protein content curves are mostly convex in nature. When considering Figure 5.46, together with the results of the expected percentage of protein in milk composition for the seven parities as summarised in Table 5.3, the following is noted. In parity 1 lowest expected percentage of protein occurs when lactation commences, while for all other parities the lowest expected percentage of protein occur later in lactation.



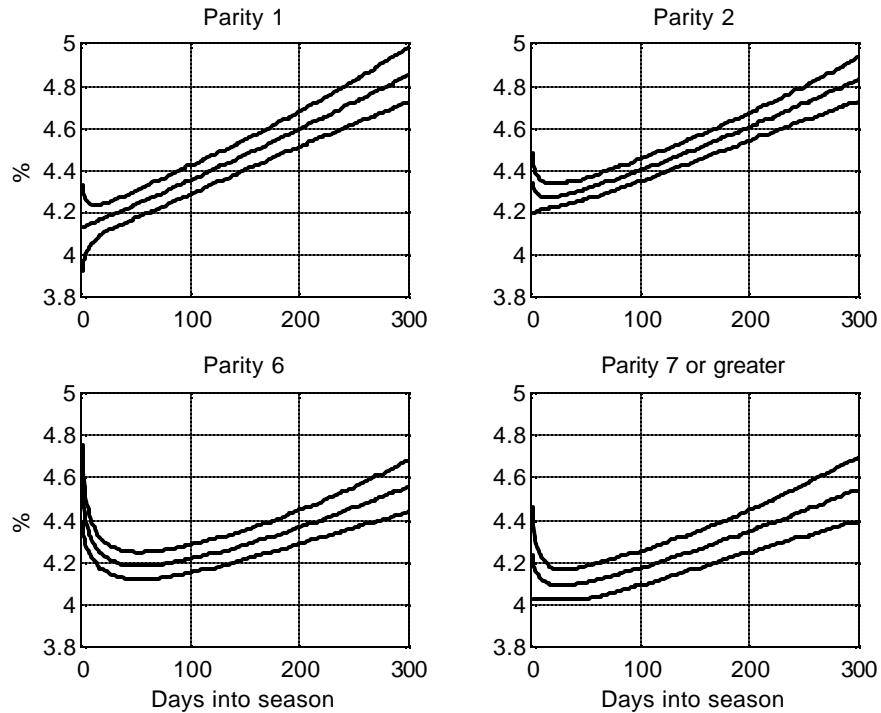


**Figure 5.44:** For each of the four lactation cycle of Animal 135 the expected protein content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the protein curve is given by —. The observed data for Animal 135 is given by —.



**Figure 5.45:** For each of the four lactation cycle of Animal 511 the expected protein content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the protein curve is given by —. The observed data for Animal 511 is given by —.

It seems as if greater variation occur with respect to the expected protein content of the various parities than was the case for the expected fat content of the seven parities. This is especially evident when considering the various times at which the minimum expected protein percentage in the composition of milk for the different parities occur. The lower boundaries of the 95% HPD intervals of parities 1 and 2 seem to take on “atypical” form. This is attributed to the method used to obtain these intervals and not to the shape of the 10 000 protein content curves used to obtain the expected protein content curve.

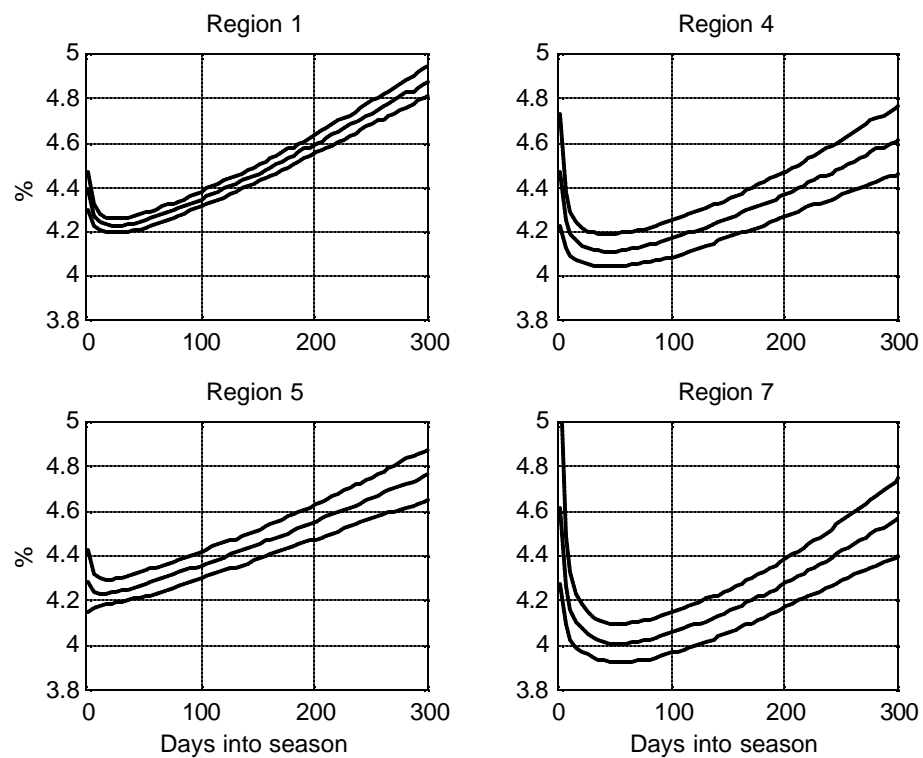


**Figure 5.46:** Expected protein content curves with 95% HPD intervals for parities 1, 2, 6 and 7 or greater when all other cofactors have been averaged out.

**Table 5.3:** Summary of the expected protein content for the seven parities

Parity	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1	4,1353	4,1353	1	4,8503
2	4,3400	4,2779	16	4,8302
3	4,4184	4,2639	28	4,7523
4	4,4982	4,2320	39	4,6919
5	4,4309	4,1807	39	4,6307
6	4,5726	4,1883	53	4,5574
7 or >	4,2318	4,0983	27	4,5449

Figure 5.47 provides expected protein content curves together with the 95% HPD intervals for four of the seven regions. Minimum expected protein content for region 5 occurred earliest (day 16) and at the highest level of the minima (4,2351%), and that of region 7 latest (day 56) and at the lowest level of the minima (4,0098%). The difference between these two regions with respect to minimum expected protein percentage is 0,2253%. The variance in expected protein content is again smallest for region 1 and largest for region 7, which was also the case for milk yield and fat content in these regions.

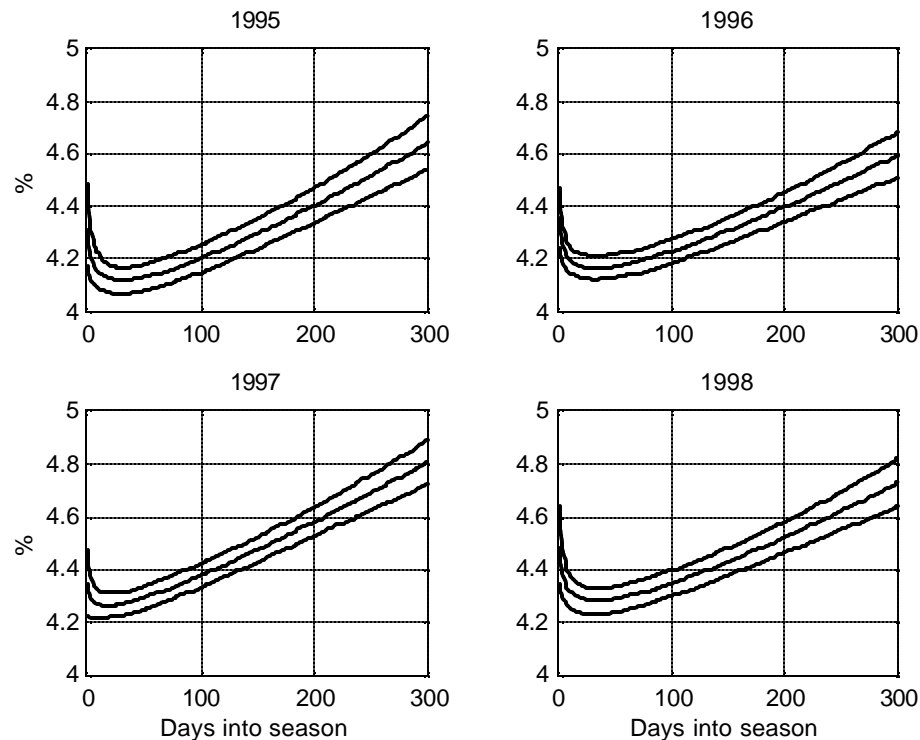


**Figure 5.47:** Expected protein content curves with 95% HPD intervals for regions 1, 4, 5 and 7 when all other cofactors have been averaged out.

Figure 5.48 and Table 5.4 considers the expected protein content for the four years. The day of minimum expected protein content varied by 16 days between the earliest and latest time for the four years, while the difference in minimum expected protein percentage was less than 0,17%.

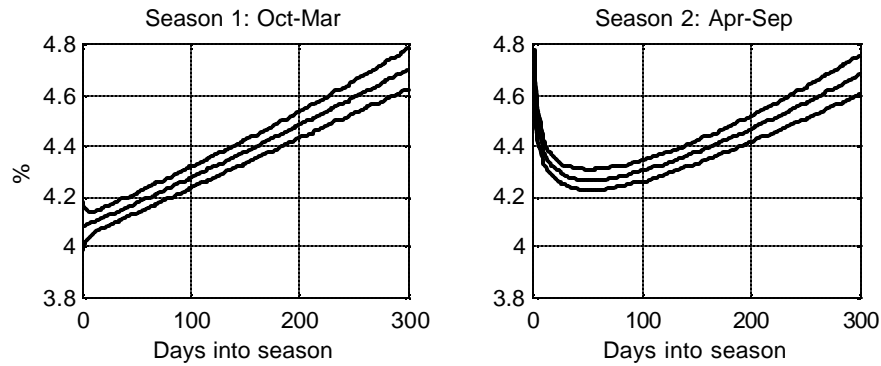
**Table 5.4: Summary of the expected protein content for the four years**

Year	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1995	4,3071	4,1188	31	4,3699
1996	4,3576	4,1660	34	4,5977
1997	4,4350	4,2643	18	4,8063
1998	4,4833	4,2829	34	4,7289



**Figure 5.48: Expected protein content curves with 95% HPD intervals for 1995, 1996, 1997 and 1998 when all other cofactors have been averaged out.**

The expected protein content curves with 95% HPD interval for the two seasons are provided in Figure 5.49. The minimum expected protein percentage for season 1 occurred on day 1 at a level of 4,0871%, while this minimum for season 2 was at 4,2664% on day 52. There are considerable differences in the levels at which of expected protein content commences and in the shape of the curves of expected protein content for the two seasons, but on day 300 the expected protein content of the two seasons are very similar.



**Figure 5.49:** Expected protein content curves with 95% HPD intervals for seasons 1 and 2 when all other cofactors have been averaged out.

### 5.6.3 Milk composition traits fat and protein: In summary

The expected fat content and protein content curves generated for each of the levels of the various cofactors when the effects of all other cofactors have been eliminated by averaging out these cofactors, provide the Bayesian equivalent to the Standard lactation curves (SLAC's) of fat content and protein content, as discussed in section 3.2 of Chapter 3, with respect to this particular herd. For the Jersey data the expected fat content curve of region 1, for example, should therefore also be viewed as the herd standard for the Jerseys in region 1 with respect to fat content. Therefore all Jersey cows in region 1 that may be considered similar to those animals included in the Jersey data are expected to have similar expected fat content curves.

Whittemore (1980) noted the existence of an inverse relationship between the level of milk yield and the percentage of fat contained in the milk. This is also true for the Jersey data, although the levels of minimum expected fat content generally occur later and that of protein content generally occur earlier than the level of peak milk yield for all cofactors. Hurley (2003) noted that it is generally accepted that the percentage of fat and the percentage of protein contained in milk are positively correlated. This was found to be true generally for the Jersey data as well. Upon investigation it was found that correlation between expected fat content and expected protein content for the 7 parities were all above 90% with the highest value of 97.42% being that of parity 4. According to Hurley (2003) milk fat is usually the most variable component in the composition of milk. This however was not the case for the Jersey data, where the expected protein content varied considerably more with respect to the form of the marginal posterior distributions for the various cofactor levels considered and with respect to the width of the 95% HPD intervals as a result of the variances.

Stanton et al. (1992) found that for cows calving in winter the protein percentage was slightly lower than for others. This is not the case in the Jersey data. Cows calving during the cooler months generally have higher levels of protein throughout lactation. Very little published work on lactation with respect to dairy cows model fat and protein contained in milk composition.

## 5.7 PREDICTIONS BASED ON THE WOOD MODEL

In section 4.1 of Chapter 4 the procedure for the prediction of one or more lactation cycle of an animal  $i$  based on the results from one or more the preceding lactation cycles is given for the general case. For the Jersey data each cow has a total of 4 observed lactation cycles in the data for which lactation modelling has already been performed. Using the method as described for the general case, it is possible to predict the results of the fifth lactation cycle for the  $i^{\text{th}}$  animal based on the four preceding recorded lactation cycles.

As first step the model parameters of the fifth lactation cycle, i.e. the lactation cycle to be predicted, or  $\mathbf{m}_{if}^{(2)}$  have to be determined if the model parameters of the four previously recorded lactation cycle are given by  $\mathbf{m}_i^{(1)} = \mathbf{m}_i^{(1)*}$ . This is done through MCMC simulation using the conditional distribution of  $\mathbf{m}_{if}^{(2)}$  specified in equation (4.22). In equation (4.22) when predicting the fifth lactation cycle for an animal in the Jersey data we use  $q = 5$  as the total number of lactation cycles under consideration and  $h = 4$  as the number of lactation cycles that are known. The conditional distribution for predicting the fifth year's model parameters for an individual or  $\mathbf{m}_{if}^{(2)}$ , given the performance in the previous four years then becomes:

$$\mathbf{m}_{if}^{(2)} | \mathbf{m}_i^{(1)} = \mathbf{m}_i^{(1)*}, \mathbf{B}, \mathbf{S}, \mathbf{R} \\ \sim N\left(\mathbf{B}\mathbf{z}_i^{(2)} + \left(\mathbf{R}_{21}^{(h)}\mathbf{R}_{11}^{(h)-1} \otimes \mathbf{I}_9\right)\left(\mathbf{m}_i^{(1)*} - \left(\mathbf{I}_4 \otimes \mathbf{B}\right)\mathbf{z}_i^{(1)}\right), \left(\mathbf{R}_{22}^{(h)} \otimes \mathbf{S}\right) - \left(\mathbf{R}_{21}^{(h)}\mathbf{R}_{11}^{(h)-1} \mathbf{R}_{12}^{(h)} \otimes \mathbf{S}\right)\right) \quad (5.12)$$

where  $\mathbf{z}_i^{(1)}$  is as applicable to the four known lactation cycles and  $\mathbf{z}_i^{(2)}$  contain the covariates expected to apply to the lactation cycle to be predicted. For the Jersey data where the covariates contained in a covariate vector identify the cofactor levels of the cofactors parity number, region, calving year and calving season for one lactation cycle, these covariates in  $\mathbf{z}_i^{(2)}$  for the fifth year to be predicted, are set as follows. The cofactor parity number is

increased by one from that in the fourth lactation cycle and the covariates identifying parity number is then set according to this value. The cofactors region and calving season are kept as they were in the fourth of the recorded lactation cycles, while the calving year is averaged out by using  $\frac{1}{4}$  as the value of each of the covariates used to identify it. It is necessary to average out the effect of calving year, because the fifth lactation cycle to be predicted here is expected to occur in 1999 for which no effect is included in the regression matrix  $\mathbf{B}$ . The years for which effects are include in matrix  $\mathbf{B}$  are 1995 to 1998 and therefore their effects are removed by averaging out the cofactor calving year in the covariate vector of the fifth year. Remember that the covariate vector,  $\mathbf{z}_i^{(2)}$ , serves as identifier of the effects in the regression matrix  $\mathbf{B}$  relevant to animal  $i$ . The marginal posterior distribution of  $\mathbf{m}_{if}^{(2)}$  is obtained through 10 000 iterations of the Gibbs sampler employing the results of parameter matrices  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  from the original Gibbs sampling used to model the four recorded lactation cycles.

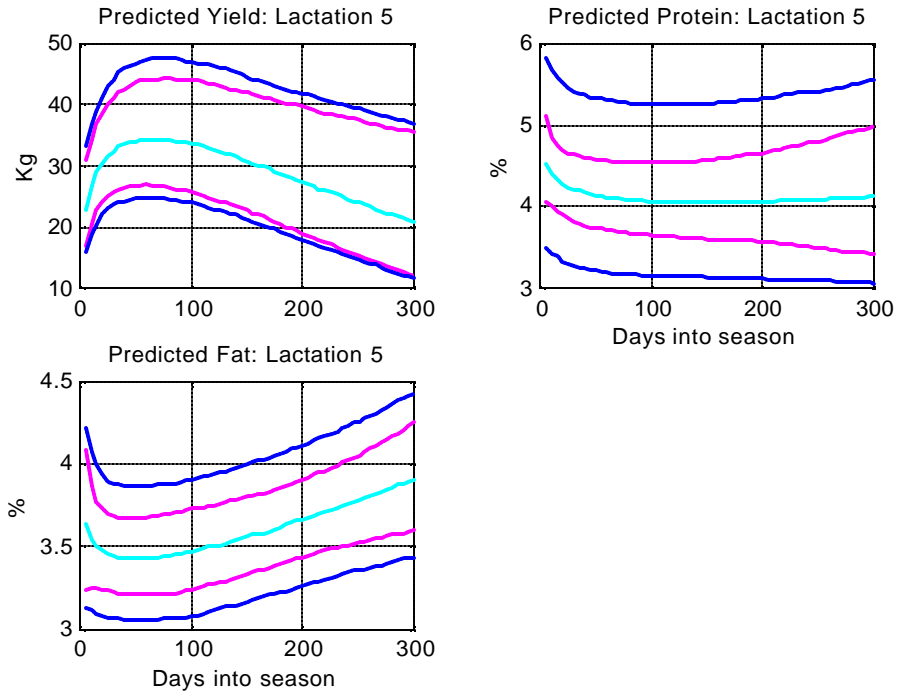
The predictive distribution of the fifth lactation cycle is:

$$\mathbf{y}_{if}^{(2)} | \mathbf{m}_{if}^{(2)} = \mathbf{m}_{if}^{(2)*}, \mathbf{F} \sim N\left(\left(\mathbf{I}_3 \otimes \mathbf{X}_{if}^{(2)}\right) \mathbf{m}_{if}^{(2)*}, \mathbf{F} \otimes \mathbf{I}_{n_{if}}\right) \quad (5.13)$$

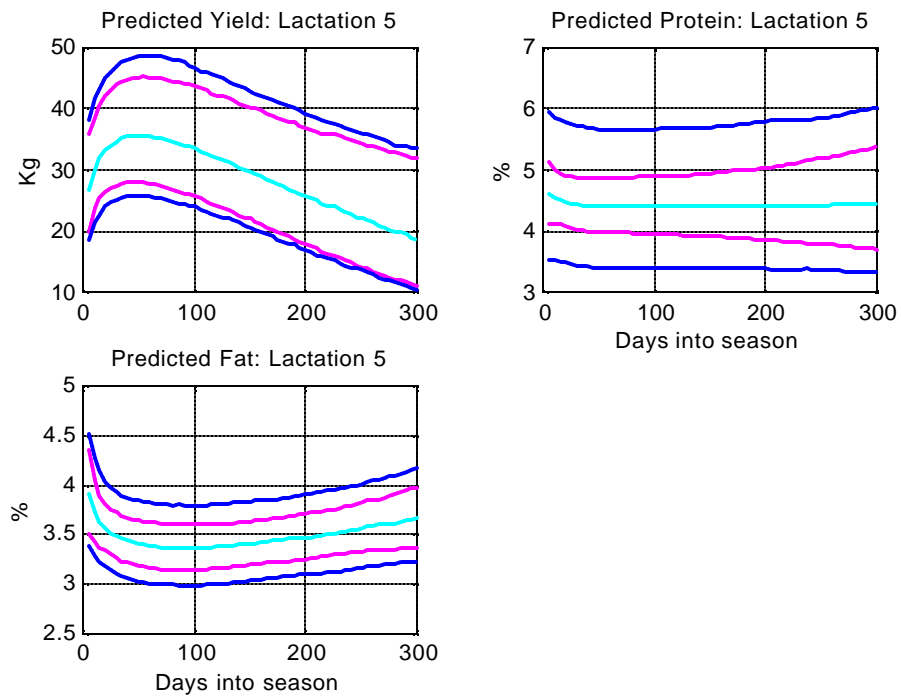
where  $\mathbf{y}_{if}^{(2)}$  is of order  $(3n_{if} \times 1)$ . The mean of the result of  $\left(\mathbf{I}_3 \otimes \mathbf{X}_{if}^{(2)}\right) \mathbf{m}_{if}^{(2)*}$  calculated for the 10 000 simulated values of  $\mathbf{m}_{if}^{(2)}$  then gives the prediction for the fifth lactation cycle and, using the 10 000 means of the predictive distribution, the 90% HPD interval for the prediction of the fifth lactation cycle is obtained.

To find a prediction interval for the prediction for the fifth lactation cycle, the predictive vector  $\mathbf{y}_{if}^{(2)}$  is simulated from a normal distribution of the form denoted in (5.13) for each of the 10 000 means,  $\left(\mathbf{I}_3 \otimes \mathbf{X}_{if}^{(2)}\right) \mathbf{m}_{if}^{(2)*}$ , and variances,  $\mathbf{F} \otimes \mathbf{I}_{n_{if}}$ . The predictive vectors in  $\mathbf{y}_{if}^{(2)}$  referring to each of the lactation traits milk yield, fat content and protein content are then used to construct 90% HPD intervals for the predictions of each of these traits. These intervals are graphed as the 90% prediction intervals for the animal under consideration.

To illustrate the result of this prediction procedure, the predicted lactation or milk yield curve, as well as the predicted fat and protein content curves for the fifth lactation cycle given the performance of the animal in preceding four recorded lactation cycles were graphed for animals 135 and 511 in the Jersey data. In each of following graphs, the predicted curves are represented by a turquoise line, while the 90% HPD intervals are given as magenta broken lines and the 90% prediction intervals as blue broken lines. The result follows.



**Figure 5.50:** For the fifth lactation cycle of Animal 135 the predicted milk yield, fat content and protein content curves are given by —, their 90% HPD interval by - - and 90% prediction interval by - - .



**Figure 5.51:** For the fifth lactation cycle of Animal 511 the predicted milk yield, fat content and protein content curves are given by —, their 90% HPD interval by - - and 90% prediction interval by - - .



# CHAPTER 6

## ALTERNATIVE MODELS FITTED TO JERSEY DATA

### 6.1 OTHER LACTATION MODELS

The Wood model had been an obvious choice when selecting models to test the approach used in this study, because it has thus far been used in most studies of lactation modelling as the benchmark model to which the performance of other models have been compared. Of the other models discussed in Chapter 2 only those that, with the assumption of multiplicative errors and after a log transformation as discussed in section 3.5 of Chapter 3, are of linear form are appropriate to be used in the approach suggested in this research. The condition of linearity of the resulting model after assuming multiplicative errors and performing a log transformation, limits the possible number of lactation models that may be used in the application of the Bayesian approach suggested here to 13, excluding the Wood model. Of the 13 possible models some, such as Brody's exponential decline function inherently do not have the ability to usefully describe lactation in dairy cows. Some of the other possible models, such as Dhanoa's reparameterised Wood model and the Jenkins and Ferrel adjustment of the Wood model, are so close to the Wood model applied in Chapter 5 that it did not make much sense from an illustration point of view to use these. The 6-parameter Morant model suggested by Williams in 1993 was also considered of little use, as the high number of parameters in this model would fit the data well, but required a larger number of test day records per animal and probably represented an over-parameterisation of the model. Because of Morant and Gnanasakthy (1989) criticism of three of their own models, as well as similarities between some of their models, some of their suggested models were also eliminated as possible models to use in illustration. Ultimately, based on all of the preceding, the general exponential model and 4-parameter Morant model were chosen for further illustration.

One other model, the Wilmink II model, also caught interest. Although the Wilmink II model did not adhere to the condition of linearity after assuming multiplicative errors and performing a log transformation, it warranted further investigation. The reason for this followed from the good result that was obtained by Olori et al. (1999) using the Wilmink II

model in comparison to the other model fitted, as well as because the Wilmink II model had in recent years been the model preferred by researchers of the Agricultural Research Council in South Africa to model lactation data of both South African Holstein and Jersey cows (Mostert et al., 2001). It was therefore decided to use an adapted form of this model in this study.

## 6.2 THE GENERAL EXPONENTIAL MODEL

The general exponential model was introduced in 1989 by Morant and Gnanasakthy to model lactation in Friesian heifers. This model assumes that the expected milk yield (in kg/day) of an animal at time  $t$  can be represented over the lactation period by

$$E(W_t) = At^b \exp((c + dt)t) \quad (6.1)$$

where  $A = \exp(a)$ . The parameters  $a, b, c$  and  $d$  are unknown and may differ from one animal to another. This model is, because of its ability to take on a convex form, also used to estimate the various milk composition traits such as fat and protein content in milk.

With the assumption of multiplicative errors for the model in (6.1), by performing a natural log transformation, the observation model for animal  $i$  becomes:

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) = a_{ijs} + b_{ijs} \ln(t_{ijp}) + c_{ijs} t_{ijp} + d_{ijs} t_{ijp}^2 + e_{ijps} \quad (6.2)$$

where,  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ ,  $i = 1, \dots, k, j = 1, \dots, q, s = 1, \dots, u, p = 1, \dots, n_{ij}$  and  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ . The Jersey data contains  $k = 1141$  animals, all of whom have been observed over  $q = 4$  lactation cycles for the  $u = 3$  lactation traits milk yield, and percentage of fat and percentage of protein in milk composition, with from 8 to 10 test day records observed for each animal during each of the four lactation cycles.

The generalised linear model form of the model for animal  $i$  during lactation cycle  $j$  is, as described in (4.3) for the general case:

$$Y_{ij} = X_{ij} M_{ij} + E_{ij} \quad (6.3)$$

where  $\text{vec } E'_{ij} (3n_{ij} \times 1) \sim N(\text{vec } 0, F \otimes I_{n_{ij}})$ . In the case of the general exponential model there are  $v = 4$  regression coefficients in the generalised linear form of the lactation model for the  $i^{\text{th}}$

animal during its  $j^{\text{th}}$  lactation cycle for the lactation trait indicated by  $s = 1, 2, 3$ . These are  $a_{ijs}$ ,  $b_{ijs}$ ,  $c_{ijs}$  and  $d_{ijs}$ , so that

$$\mathbf{M}_{ij}(4 \times 3) = \begin{bmatrix} a_{ij1} & a_{ij2} & a_{ij3} \\ b_{ij1} & b_{ij2} & b_{ij3} \\ c_{ij1} & c_{ij2} & c_{ij3} \\ d_{ij1} & d_{ij2} & d_{ij3} \end{bmatrix}. \quad (6.4)$$

Row  $r$  of design matrix  $\mathbf{X}_{ij}(n_{ij} \times 4)$  contains the elements that would return the general exponential model in generalised linear form when multiplied with the matrix of regression coefficients  $\mathbf{M}_{ij}$  in (6.4) above,

$$\mathbf{X}_{ij}^{(r)} = [1 \quad \ln t_{ijr} \quad t_{ijr} \quad t_{ijr}^2]. \quad (6.5)$$

The model extended for animal  $i$  to all  $q = 4$  lactation cycles then is,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_{i1} \\ \mathbf{Y}_{i2} \\ \mathbf{Y}_{i3} \\ \mathbf{Y}_{i4} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (6.6)$$

where  $\text{vec } \mathbf{E}'_i(3n_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^4 n_{ij}$ ,  $\mathbf{X}_i(n_i \times 16) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i(16 \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \\ \mathbf{M}_{i3} \\ \mathbf{M}_{i4} \end{bmatrix}. \quad (6.7)$$

The matrix  $\mathbf{R}$ , containing the covariances between the model parameters of the same animal in successive lactation cycles remains of the form:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{r}_{13} & \mathbf{r}_{14} \\ \mathbf{r}_{12} & \mathbf{r}_{22} & \mathbf{r}_{23} & \mathbf{r}_{24} \\ \mathbf{r}_{13} & \mathbf{r}_{23} & \mathbf{r}_{33} & \mathbf{r}_{34} \\ \mathbf{r}_{14} & \mathbf{r}_{24} & \mathbf{r}_{34} & \mathbf{r}_{44} \end{bmatrix}, \quad (6.8)$$

because  $q = 4$  lactation cycles were observed for all 1141 animal in the Jersey data.

Additional information on the cofactors parity number, region, calving year, and calving season available in the Jersey data have to again also be included in the model. To do this a covariate matrix  $\mathbf{Z}_i(17 \times 4)$  is, as before, constructed for each animal  $i$  as explained in section

5.3 of Chapter 5. This covariate matrix is then used as described in section 4.1 of Chapter 4 for the generalised Bayesian model.

Once the model for the generalised exponential model had been written in the above form, the Gibbs sampler as described in section 4.2 of Chapter 4 and the full conditional distributions of all the model parameters of interest as set out in equations (4.17) to (4.21) are used to obtain the marginal posterior distributions conditional on the observed data only. The hyperparameters  $\mathbf{G}$  and  $\mathbf{P}$  required to generate marginal distributions for  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , were again set equal to the moments estimators from the sampling distributions of  $\mathbf{S}$  and  $\mathbf{R}$  with degrees of freedom still as small as possible,  $d = 4$  and  $g = 13$ .

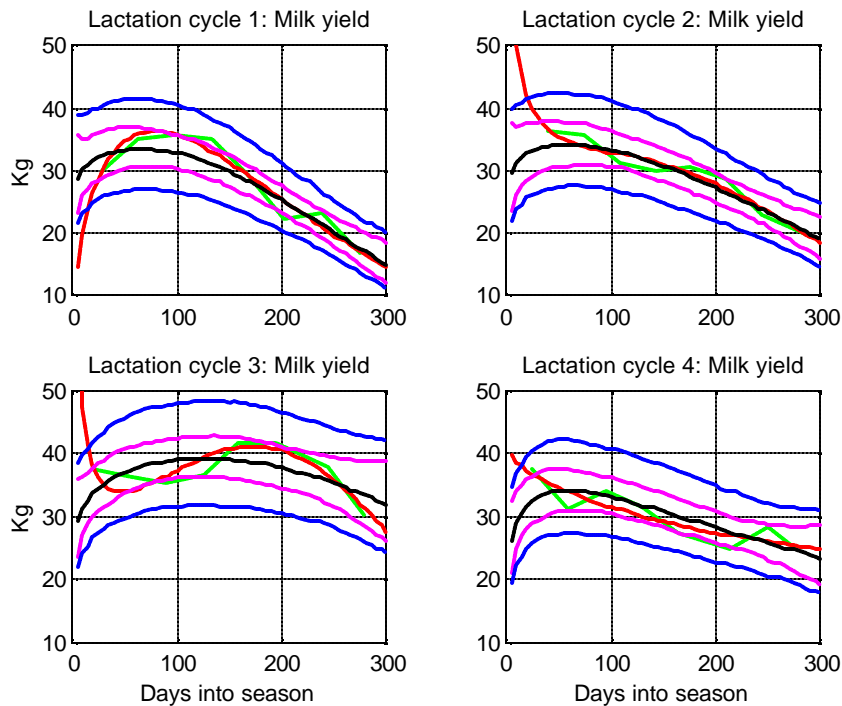
As in Chapter 5 the Gibbs sampler was put through a “burn-in” period of 2000 simulation iterations, after which 10 000 sets of parameters were generated and kept using equations (4.17) to (4.21). This again required simulating from normal and Wishart distributions only, as described in Chapter 5. Speed and storage constraints posed by Matlab, again resulted in only the 10 000 sets of parameters generated for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  being stored. To regenerate the parameters contained in  $\mathbf{M}_i$ ,  $i = 1, \dots, 1141$ , either of the two procedures referred to as Alternatives 1 and 2 respectively and explained in full in section 5.5 of Chapter 5 may again be used.

Using the 10 000 resulting regression matrices  $\mathbf{B}$  ( $9 \times 17$ ) to construct 90% HPD intervals for the elements of  $\mathbf{B}$ , it was found that only 85 of the 204 elements of this matrix significantly affect the parameters of the three lactation trait curves, milk yield, percentage of fat and percentage of protein in milk composition. The affect of both the base level and the levels of the covariates on the parameter  $d$  of the milk yield curve are not significant. The base level effects on all 11 the other parameters of the model are however significant. With respect to the effects of all covariates on the parameters of the traits curves, all covariates significantly affect at least one of the 12 parameters and therefore all cofactors were kept in the model. It is noted though that it is possible that for a specific animal, the contributions of all included covariates to the parameters for that animal could be very small or perhaps even insignificant.

Further investigation of the Jersey data when modelled by means of the general exponential model followed once the marginal distributions of the parameters had been obtained through Gibbs sampling.

## 6.2.1 The lactation trait milk yield

Using the 10 000  $M_i$  ( $16 \times 3$ ) matrices for an animal  $i$  over all four lactation cycles regenerated through alternative 2 of the two possible methods to do so, the lactation or milk yield curves for each lactation cycle of each of the 10 000 simulations for such an animal can be obtained. Note that only the parameters in  $M_i$  ( $16 \times 3$ ) relevant to milk yield in each of the four lactation cycles are used to obtain milk yield curves. The mean of the 10 000 lactation curves for each lactation cycle gives the expected lactation curve for animal  $i$  during each lactation cycle. To illustrate such a result the expected lactation curves for each of the four lactation cycles of animal  $i = 135$  are given as solid black lines in Figure 6.1 below.



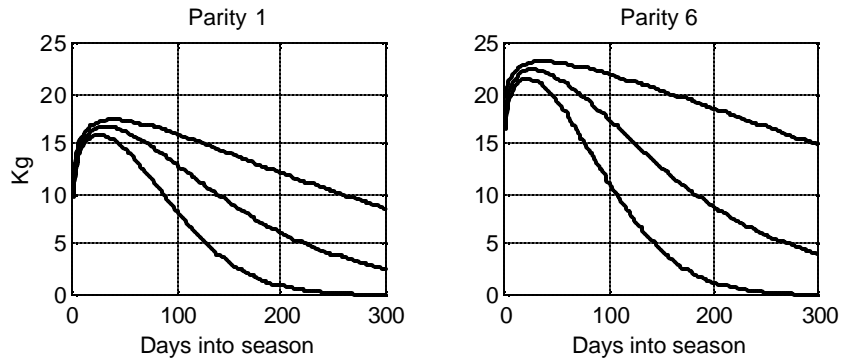
**Figure 6.1:** For each of the four lactation cycle of Animal 135 the expected lactation curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

In Figure 6.1 the green line represents the observed milk yield values during each lactation cycle connected by straight-line segments. The red line is the least square estimate fitted to the data using the general exponential model. The magenta broken lines provide the 90% HPD intervals for the expected lactation curves during each lactation cycle and the blue

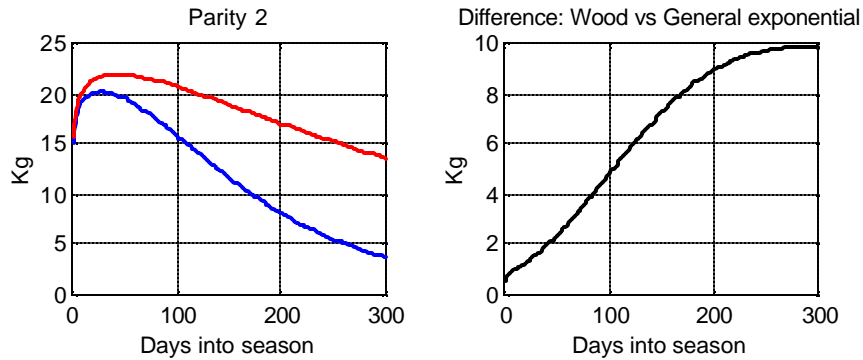
broken lines represent a 90% prediction intervals for the expected lactation curve obtained using the 90% HPD intervals of the predictions that result from the predictive density as explained in section 5.5 of Chapter 5. The least squares estimates of the milk yield curve of lactation cycles 2 and 4 were atypical, while that of lactation cycle 3 commences at an unrealistic point and has 2 turning points. However, the Bayesian curve estimated in all three these cases are of typical lactation curve form, again demonstrating the ability of the method to reduce the occurrences of such unpractical results even when a lactation model other than that of Wood is used.

The expected lactation or milk yield curve of the herd with respect to one of the levels of a certain cofactor may also be obtained as before by constructing a  $z^*$  vector that only considers the level of the cofactor of interest and averages out the effect of all other cofactors. The 10 000 sets of parameter values that apply when this is done results, as before, from  $m^c = Bz^*$  for each of the 10 000 simulated  $B$  matrices. The parameter values contained in  $m^c$  applicable to milk yield are then used to construct 10 000 lactation curves for the particular level of the cofactor of interest. The expected lactation curve for such a particular level of the cofactor of interest is then obtained as the mean of these 10 000 lactation curves. The procedure was explained in greater detail in section 5.5 of Chapter 5.

In Figure 6.2 below the expected lactation curves, together with their 95% HPD intervals for parities 1 and 6 when the other cofactors, “region”, “calving year” and “calving season”, have been averaged out, are given. These curves are now valid with respect to the herd under consideration and are the Bayesian equivalent of the Standard lactation curves (SLAC’s), as discussed in section 3.2 of Chapter 3, when the general exponential model is used to model lactation. In Figure 6.3 the expected lactation curve of parity 2 when the general exponential model is used to model lactation is compared to the result for the same parity obtained from the Wood model in Chapter 5. In the figure on the left it is seen that there are considerable differences in shape between the two models, with the expected lactation curve for parity 2 of the Wood model always above that of general exponential model. The figure on the right provides the differences in milk yield between the two curves on the left plotted over time. Initially the expected yields are very similar, but towards the end of the lactation cycle these differences increase up to almost 10 kilograms.



**Figure 6.2: Expected lactation curves with 95% HPD intervals resulting from the general exponential model for parities 1 and 6 when the other cofactors have been averaged out.**



**Figure 6.3: (1) On the left the expected lactation curves for parity 2 using the general exponential model (—) and the Wood model (---) when other cofactors have been averaged out. (2) On the right the differences in milk yield between the two curves in (1) plotted over time.**

The results with respect to the expected lactation curves of all seven parity groups are summarised in Table 6.1. The lowest level of peak milk yield occurs in first parity, with a sharp increase to second parity. With the exception of parity 5, peak milk yield increases from first to sixth parity and then declines in parity 7 or >. Time of expected peak yield is again latest in parity 1, but when fitting the general exponential model it is modelled to occur earlier than the mean time of peak yield that resulted on day 48 for the Wood model. The expected lactation curves for all parities end in a lower expected yield on day 300 than what was the case for the Wood model.

**Table 6.1: Summary of the expected milk yield for the seven parities using the general exponential model.**

Parity	Expected kg's milk on day 1	Maximum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1	10,6888	16,6740	32	2,6184
2	15,1086	20,2038	28	3,7242
3	17,0189	21,7202	29	5,0600
4	17,0379	22,1419	30	5,2354
5	16,4348	22,0090	31	4,7181
6	17,9320	22,3382	26	4,0648
7 or >	15,5584	21,2362	28	3,0995

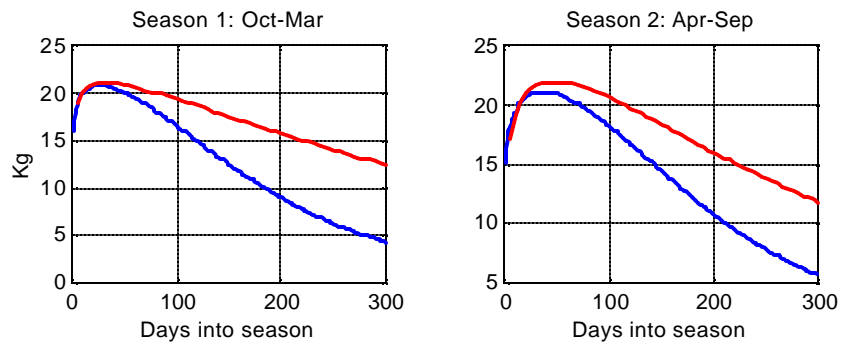
The results with respect to the expected lactation curves of the four years are summarised in Table 6.2. When using the general exponential model, the expected lactation curves of the four calving years commence, peak and end at lower levels than was the case for the Wood model. The general exponential model also models the expected time of peak milk yield earlier for the four calving years than what is the case when using the Wood model for this purpose.

**Table 6.2: Summary of the expected milk yield for the four years using the general exponential model.**

Year	Expected kg's milk on day 1	Maximum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1995	16,8993	20,4512	28	4,8500
1996	14,8709	20,2076	31	4,2768
1997	15,0523	21,2616	33	4,7667
1998	15,2319	21,6145	31	4,1225

In Figure 6.4 the expected lactation curves for the two seasons when the general exponential model is used to model lactation is compared to the result obtained from the Wood model in Chapter 5. Expected milk yield for the first season is modelled to commence at a lower level, with slightly earlier peak at a lower peak level than expected milk yield for this season when employing the Wood model. The result of this comparison for the second season is similar only with respect to expected peak and time of expected peak milk yield. The largest difference in expected milk yield for the two models occurs at the end of lactation.

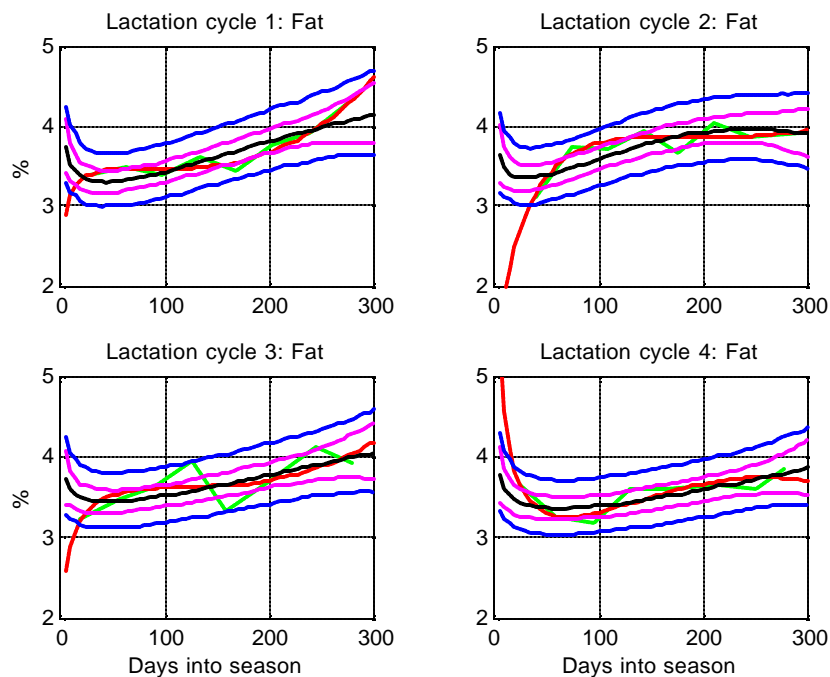




**Figure 6.4:** The expected lactation curves for seasons 1 and 2 using the general exponential model (—) and the Wood model (---) when other cofactors have been averaged out.

## 6.2.2 The lactation trait fat content of milk

The expected fat content curve for an animal during each lactation cycle may be obtained using the same approach described for milk yield in section 6.2.1.

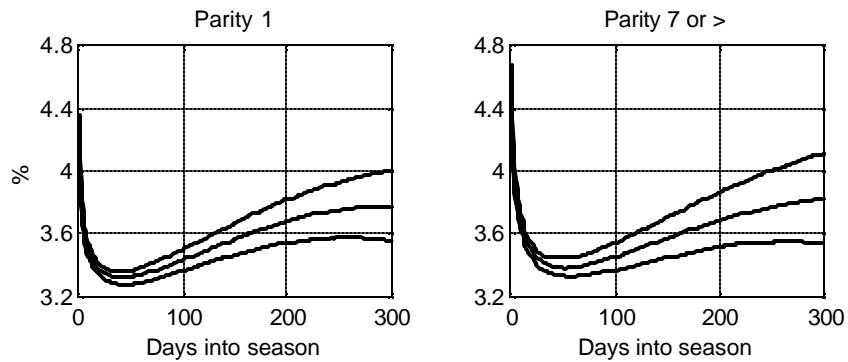


**Figure 6.5:** For each of the four lactation cycle of Animal 135 the expected fat content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - -. The least squares estimate of the fat content curve is given by —. The observed data for Animal 135 is given by —.

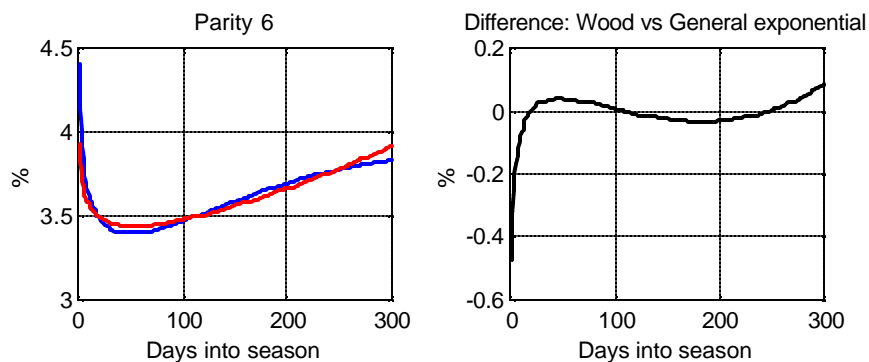
In Figure 6.5 the expected fat content curves for each of the four lactation cycles of animal  $i = 135$  are given as solid black lines, the green lines represent the observed data, the red lines

the least squares estimated obtained using the general exponential model and the magenta lines the 90% HPD intervals for the expected fat content curves. The blue lines represent the 90% prediction intervals for the expected lactation curves obtained as explained before. The least squares estimate of the fat content curves in the first three lactation cycles are atypical, but the expected fat content curves estimated using the Bayesian approach again successfully addresses this problem.

Expected fat content curves for the herd with respect to one of the levels of a cofactor of interest were again obtained using the procedure briefly explained above with respect to milk yield. As part of the obtained results, Figure 6.6 gives the expected fat content curves together with 95% HPD intervals for parities 1 and 7 or greater when the other cofactors in the model have been averaged out.



**Figure 6.6: Expected fat content curves with 95% HPD intervals resulting from the general exponential model for parities 1 and 7 or greater when the other cofactors have been averaged out.**



**Figure 6.7:(1) On the left the expected fat content curves for parity 6 using the General exponential model (—) and the Wood model (---) when other cofactors have been averaged out. (2) On the right the differences in fat content between the two curves in (1) the plotted over time.**

In Figure 6.7 the expected fat content curves of parity 6 resulting from the general exponential and Wood models are compared. In the figure on the left it is seen that there are very little differences in shape between the two models, while the figure on the right confirms that the scale of these differences is indeed small, with the largest differences at the start and end of the lactation cycle. The results with respect to the expected fat content curves of all seven parities are summarised in Table 6.3. On average expected fat content of the parities is 0,5% higher on the first day of the lactation cycle for the general exponential model than for the Wood model, while the level of the minimum expected fat content is 0,05% lower and occurs from 4 to 6 days later. For the Wood model the expected fat content of the parities on day 300 is on average 0,11% higher than for the general exponential model.

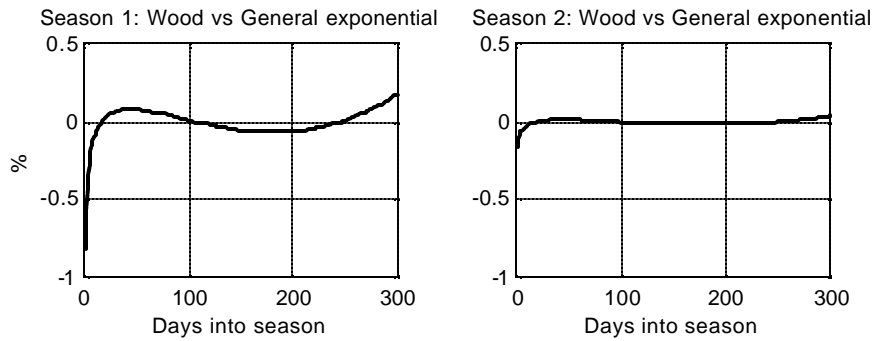
**Table 6.3: Summary of the expected fat content for the seven parities using the general exponential model.**

Parity	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1	4,1748	3,3208	42	3,7759
2	4,3921	3,3923	46	3,8769
3	4,4801	3,3876	47	3,8542
4	4,4476	3,3979	49	3,8702
5	4,4763	3,3969	49	3,8270
6	4,4055	3,3975	50	3,8253
7 or >	4,4923	3,3845	52	3,8162

The results with respect to the expected fat content curves of the four years are summarised in Table 6.4. When comparing the results for the expected fat content of the four years to that obtained using the Wood model the results are on average similar to that of the parities above, except for the time of minimum fat content which is now for 1 to 6 days later for the general exponential model than for the Wood model.

**Table 6.4: Summary of the expected fat content for the four years using the general exponential model.**

Year	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1995	4,4791	3,4371	47	3,8294
1996	4,2983	3,3770	47	3,8525
1997	4,4743	3,4264	47	3,8440
1998	4,3844	3,2930	49	3,8121

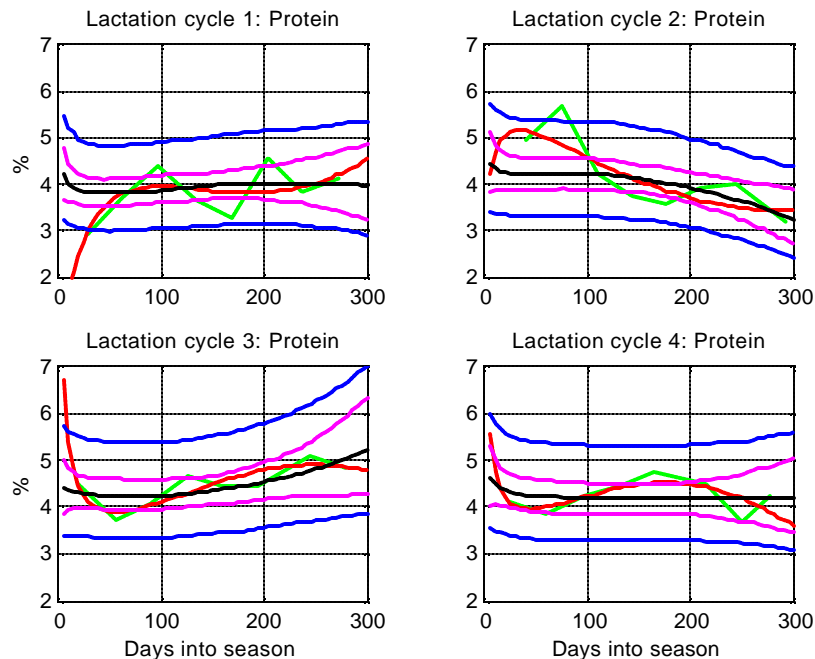


**Figure 6.8:** The differences in expected fat content for seasons 1 and 2 when comparing the expected fat content curves that resulted from Wood model to that of the general exponential model over time.

From Figure 6.8 it follows that the expected fat content curve that resulted from the Wood and general exponential models are more similar during the second calving season than the first.

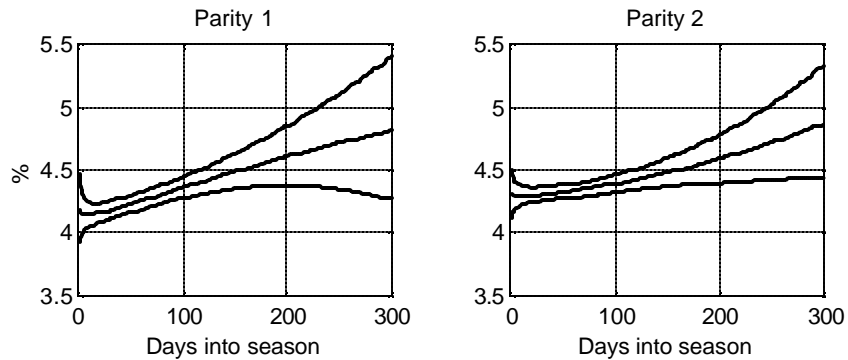
### 6.2.3 The lactation trait prote in content of milk

The expected protein content curve for an animal during each lactation cycle may also be obtained using the approach described for milk yield in section 6.2.1. In Figure 6.9 the expected protein content curves for each of the 4 lactation cycles of animal  $i = 135$  are given.



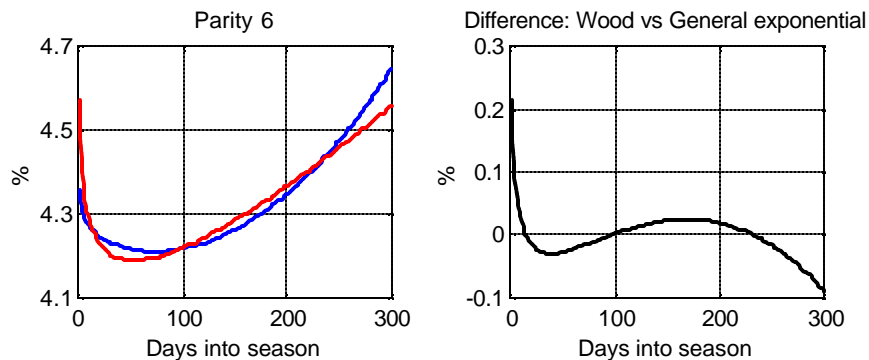
**Figure 6.9:** For each of the four lactation cycle of Animal 135 the expected protein content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the protein curve is given by —. The observed data for Animal 135 is given by —.

To illustrate the result for the expected protein content, Figure 6.10 gives the expected protein content curves together with 95% HPD intervals for parities 1 and 2 when the effects of all other cofactors have been eliminated by taking averages.



**Figure 6.10:** Expected protein content curves with 95% HPD intervals resulting from the general exponential model for parities 1 and 2 when other cofactors have been averaged out.

In Figure 6.11 the expected protein content curves of parity 3 resulting from the general exponential and Wood models are compared. The greatest differences between the two models occur at the start and at the end of the lactation cycle. The results with respect to the expected protein content of all seven parities are summarised in Table 6.5.



**Figure 6.11:** (1) On the left the expected protein content curves for parity 6 using the general exponential model (—) and the Wood model (---) when other cofactors have been averaged out. (2) On the right the differences in protein content between the two curves in (1) the plotted over time.

The greatest variation in differences between the result of the Wood model and general exponential model occurs with respect to the protein content of milk. This is clear when considering the results of protein content for the seven parities as no fixed pattern in the differences of the results for the two models may be discerned. For example, when considering the expected protein percentages on day 1, day 300 and at minimum we find that

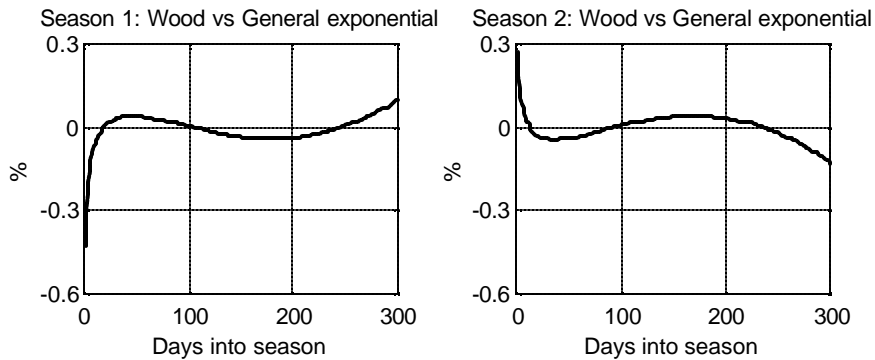
for one parity the values of the Wood model may be greater than that of the general exponential model, while for another parity it is the other way around. The same kind of result also occurs with respect to the day on which minimum protein content is attained. The same result is also observed with respect to the expected protein content for the four years when comparing the summarised results of Table 6.6 to that of the Wood model.

**Table 6.5: Summary of the expected protein content for the seven parities using the general exponential model.**

Parity	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1	4,1786	4,1427	8	4,8111
2	4,3148	4,2955	11	4,8725
3	4,7031	4,2516	40	4,7306
4	4,6475	4,2271	46	4,6976
5	4,6192	4,1753	46	4,6367
6	4,3552	4,2099	74	4,6485
7 or >	4,4148	4,1010	38	4,5695

**Table 6.6: Summary of the expected protein content for the four years using the general exponential model.**

Year	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1995	4,6517	4,1006	41	4,6193
1996	4,3094	4,1825	38	4,6390
1997	4,2714	4,2649	6	4,8360
1998	4,6117	4,2789	41	4,7337

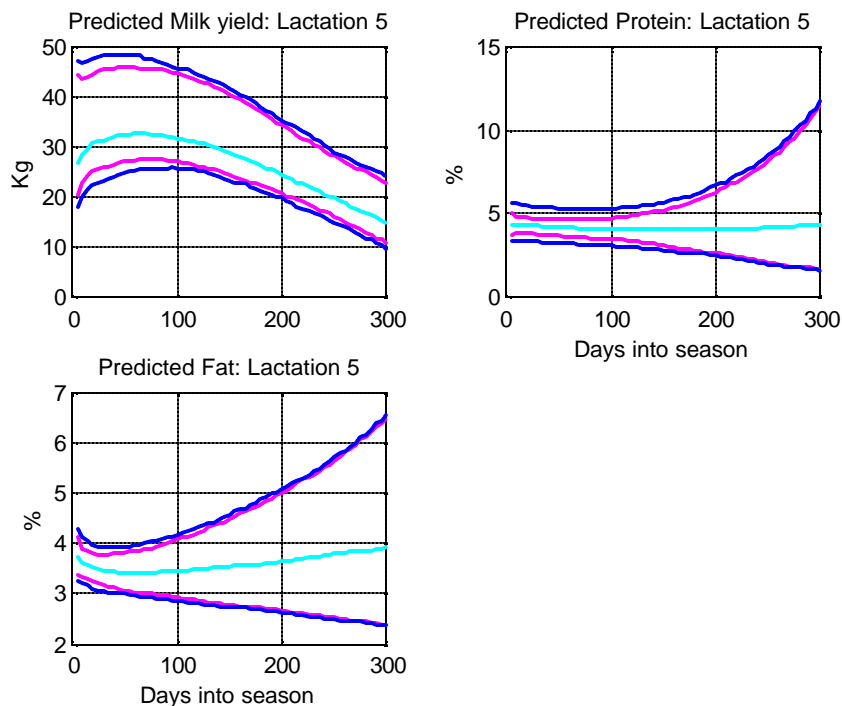


**Figure 6.12: The differences in expected protein content for seasons 1 and 2 when comparing the expected protein content curves that resulted from Wood model to that of the general exponential model over time.**

Figure 6.12 supports the statement that no fixed pattern in the differences of the results for the two models with respect to protein content may be discerned.

## 6.2.4 Predictions based on the general exponential model

In section 4.1 of Chapter 4 the procedure for the prediction of one or more lactation cycles of an animal  $i$  based on the results from one or more preceding lactation cycles is discussed for the general case. In section 5.7 of Chapter 5 this method was applied in order to predict a fifth lactation cycle of an animals based on the results from the preceding four lactation cycles making use of the Wood model. This same prediction procedure will now be used here to predict the milk yield, fat content and protein content curves for the fifth lactation cycle of animal  $i = 135$  given the performance in the preceding four lactation cycles modelled by means of the general exponential model. The results are provided in Figure 6.13 where the solid turquoise line represents the predicted curve, while the 90% HPD intervals are given as magenta broken lines and the 90% prediction intervals as blue broken lines.



**Figure 6.13:** For the fifth lactation cycle of Animal 135 the predicted milk yield, fat content and protein content curves resulting from the general exponential model are given by —, their 90% HPD interval by - - and 90% prediction interval by - - .

### 6.3 THE 4-PARAMETER MORANT MODEL

The 4-parameter Morant model also resulted from research in 1989 by Morant and Gnanasakthy in an effort to model lactation in Friesian heifers. This model assumes that the expected milk yield (in kg/day) of an animal at time  $t$  can be represented over the lactation period by

$$E(W_t) = A \exp(bt + ct^2 + d/t) \quad (6.9)$$

where  $A = \exp(a)$ . The parameters  $a, b, c$  and  $d$  are unknown and may differ from one animal to another. This model also has the ability to take on a convex form and is therefore also used to estimate the milk composition traits fat and protein content of milk for the Jersey data.

With the assumption of multiplicative errors for the model in (6.9) and after performing a natural log transformation, the observation model for animal  $i$  is:

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) = a_{ijs} + b_{ijs}t_{ijp} + c_{ijs}t_{ijp}^2 + \frac{d_{ijs}}{t_{ijp}} + e_{ijps} \quad (6.10)$$

where  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ ,  $i = 1, \dots, 1141$ ,  $j = 1, \dots, 4$ ,  $s = 1, 2, 3$ , and  $p = 1, \dots, n_{ij}$  where  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ . From 8 to 10 test day records were observed for each animal during each of the four lactation cycles.

As described in (4.3) for the general case, the model for animal  $i$  during lactation cycle  $j$  in generalised linear model form is:

$$Y_{ij} = X_{ij}M_{ij} + E_{ij} \quad (6.11)$$

where  $\text{vec } E'_{ij}(3n_{ij} \times 1) \sim N(\text{vec } 0, F \otimes I_{n_{ij}})$ . The 4-parameter Morant model has  $v = 4$  regression coefficients in the generalised linear form of the lactation model for animal  $i$  in its  $j^{\text{th}}$  lactation cycle for the lactation trait indicated by  $s = 1, 2, 3$ . These regression coefficients are denoted by  $a_{ijs}$ ,  $b_{ijs}$ ,  $c_{ijs}$  and  $d_{ijs}$ , so that

$$M_{ij}(4 \times 3) = \begin{bmatrix} a_{ij1} & a_{ij2} & a_{ij3} \\ b_{ij1} & b_{ij2} & b_{ij3} \\ c_{ij1} & c_{ij2} & c_{ij3} \\ d_{ij1} & d_{ij2} & d_{ij3} \end{bmatrix}. \quad (6.12)$$

In the design matrix  $X_{ij}(n_{ij} \times 4)$  row  $r$  contains the elements that would return the generalised linear form of the 4-parameter Morant model when multiplied with the matrix of regression coefficients  $M_{ij}$  in (6.12) above,



$$\mathbf{X}_{ij}^{(r)} = \begin{bmatrix} 1 & t_{ijr} & t_{ijr}^2 & t_{ijr}^{-1} \end{bmatrix}. \quad (6.13)$$

The model extended for animal  $i$  to all  $q = 4$  lactation cycles then is,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_{i1} \\ \mathbf{Y}_{i2} \\ \mathbf{Y}_{i3} \\ \mathbf{Y}_{i4} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (6.14)$$

where  $\text{vec } \mathbf{E}'_i (3n_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^4 n_{ij}$ ,  $\mathbf{X}_i (n_i \times 16) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i (16 \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \\ \mathbf{M}_{i3} \\ \mathbf{M}_{i4} \end{bmatrix}. \quad (6.15)$$

Because  $q = 4$  lactation cycles were observed for all 1141 animals in the Jersey data, the matrix  $\mathbf{R}$ , containing the covariances between the model parameters of the same animal in successive lactation cycles remains of the form denote in (6.8). As before the covariate matrix  $\mathbf{Z}_i (17 \times 4)$  for each animal  $i$  contains the information on the cofactors.

Next the Gibbs sampler as described in section 4.2 of Chapter 4 using the full conditional distributions of all the model parameters of interest as set out in equations (4.17) to (4.21) is employed to obtain the marginal posterior distributions conditional on the observed data only. The hyperparameters  $\mathbf{G}$  and  $\mathbf{P}$  required to generate marginal distributions for  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , were again set equal to the moments estimators from the sampling distributions of  $\mathbf{S}$  and  $\mathbf{R}$  with degrees of freedom still as small as possible,  $\mathbf{d} = 4$  and  $g = 13$ . As before the Gibbs sampler was put through a “burn-in” period of 2000 simulation iterations, after which 10 000 sets of parameters were generated. Again only the 10 000 sets of parameters generated for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  were stored, because to regenerate the parameters contained in  $\mathbf{M}_i$ , either of the two alternatives explained in full in section 5.5 of Chapter 5 may be used.

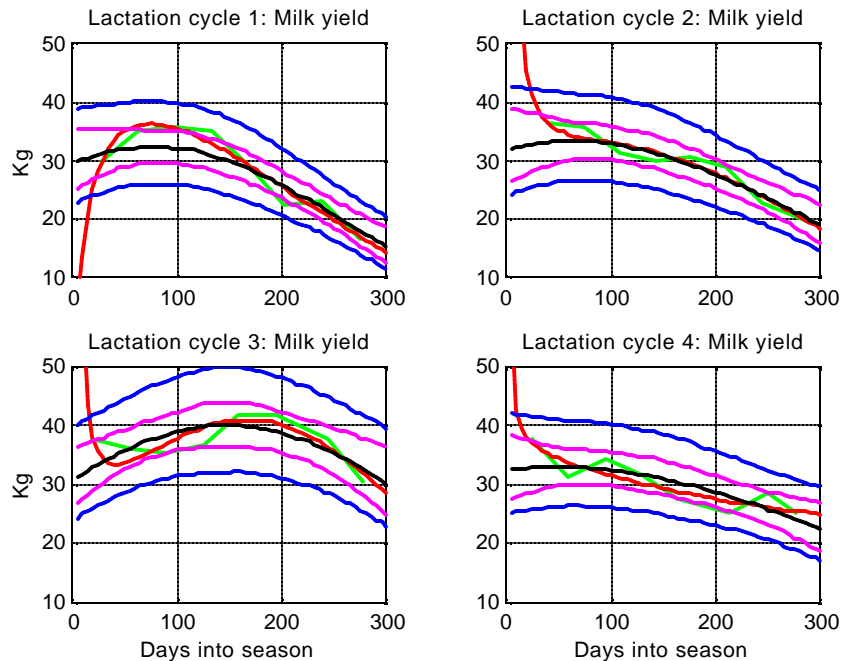
The 90% HPD intervals for  $\mathbf{B}$ , constructed from the 10 000 regression matrices  $\mathbf{B}$  ( $9 \times 17$ ) resulting from the Gibbs sampler indicated that 109 of the 204 elements of this matrix significantly affect the parameters of the three lactation trait curves, milk yield, percentage of fat and percentage of protein in milk composition. The affects of all levels of this matrix on the parameter  $c$  of the protein content curve are not significant. However, the base level effects on all 11 the other parameters of the model are significant. The effects of all

covariates on the parameters of the traits curves are significant for at least two of the 12 parameters in the model and therefore all cofactors were kept.

Once the marginal distributions of the parameters of the general exponential model had been obtained through Gibbs sampling, further investigation of the results for Jersey data based on this lactation model follows.

### 6.3.1 The lactation trait milk yield

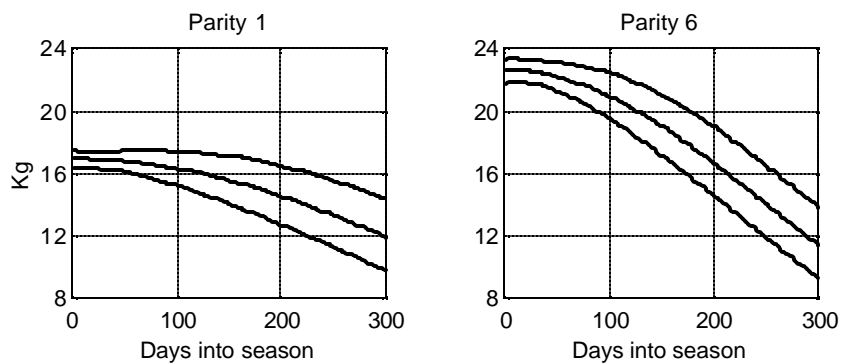
Figure 6.14 shows the results, obtained as before, of the expected lactation curves for the four lactation cycles of animal 135 as black lines, while the observed data are given in green. The red lines are the least squares estimates of the 4-parameter Morant model fitted to the data. The magenta broken lines represent the 90% HPD intervals of the expected lactation curve and the blue broken lines the 90% prediction intervals for the expected lactation curve obtained using the 90% HPD intervals of the predictions resulting from the predictive density as explained in section 5.5 of Chapter 5.



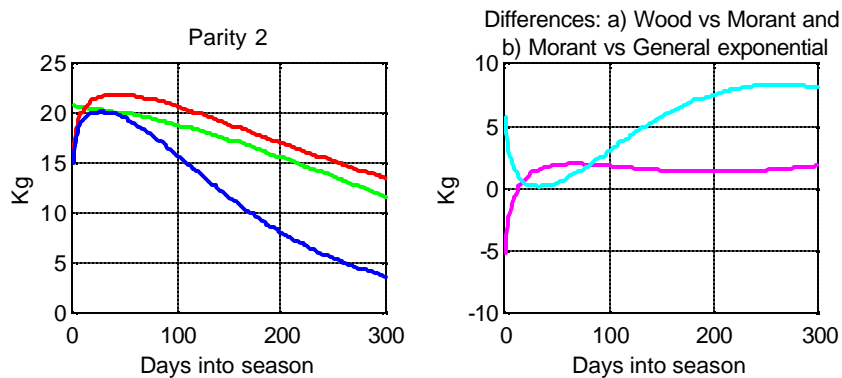
**Figure 6.14:** For each of the four lactation cycle of Animal 135 the expected lactation curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

The expected lactation or milk yield curve of the herd with respect to one of the levels of a cofactor is again determined using the method applied in section 5.5 of Chapter 5 and section 6.2.1 above. The results of these curves are the Bayesian equivalents of the Standard lactation curves (SLAC's) discussed in section 3.2 of Chapter 3 when the 4-parameter Morant model is used to model lactation.

Figure 6.15 below shows the expected lactation curves and their 95% HPD intervals for parities 1 and 6 when the other three cofactors (region, calving year and calving season) have been averaged out.



**Figure 6.15:** Expected lactation curves with 95% HPD intervals resulting from the 4-parameter Morant model for parities 1 and 6 when other cofactors have been averaged out.



**Figure 6.16:** (1) On the left the expected lactation curves for parity 2 using the 4-parameter Morant model (—), the general exponential model (---) and the Wood model (- - -) when all other cofactors have been averaged out. (2) On the right the differences in milk yield between the expected curves for parity 2 the plotted over time for (a) Wood vs Morant model (- - -) and (b) Morant vs general exponential model (—).

The graph on the left in Figure 6.16 compares the expected lactation curves of parity 2 obtained from the 4-parameter Morant model in green, the general exponential model in blue and the Wood model in red, while in the graph on the right the difference over time between the expected lactation curves of parity 2 for the Wood against the 4-parameter Morant model is plotted in magenta and that for the 4-parameter Morant model against the general exponential model is plotted in turquoise. From this is clear that the resulting expected lactation curves of the Wood and 4-parameter Morant models are more similar in scale than the expected lactation curves of the 4-parameter Morant and general exponential models. This result also applies to the expected lactation curves of all other parities.

The results with respect to the expected lactation curves of all seven parity groups are summarised in Table 6.7. When compared to the expected lactation curves of the Wood model it is noted that the time of peak milk yield is considerably earlier in the curves resulting from the 4-parameter Morant model and even earlier than in the general exponential model, with the first three parities attaining peak yield on the first day of the lactation cycle. From about day 50 in the lactation cycle the difference between expected lactation curves of the Wood and 4-parameter Morant model remains more or less constant for all parities, with the expected curves resulting from the Wood model always above that of the 4-parameter Morant model.

**Table 6.7: Summary of the expected milk yield for the seven parities using 4-parameter Morant model.**

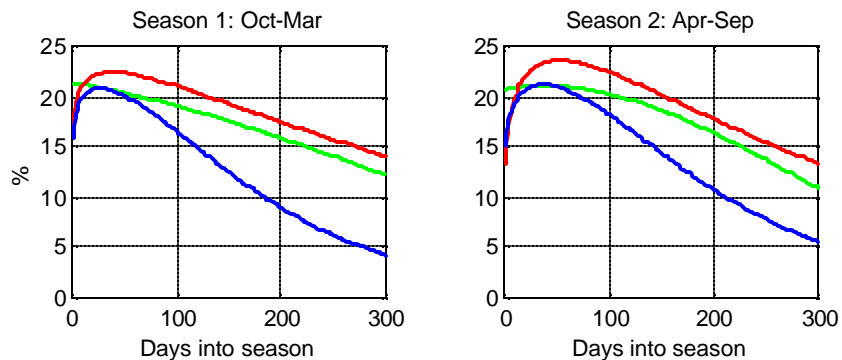
Parity	Expected kg's milk on day 1	Maximum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1	16,9478	16,9478	1	11,9704
2	20,7641	20,7641	1	11,7074
3	22,0742	22,0742	1	11,7733
4	22,1666	22,2834	8	11,6906
5	21,8394	21,9862	21	11,5007
6	22,5006	22,5823	6	11,4014
7 or >	21,0783	21,2422	17	10,8625

A summary of the results of the expected lactation curves for the four calving years is provided in Table 6.8. When comparing the expected lactation curves of the 4-parameter Morant, Wood and general exponential models for the four calving years results similar to that of parity are obtained.

**Table 6.8: Summary of the expected milk yield for the four years using the 4-parameter Morant model.**

Year	Expected kg's milk on day 1	Minimum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1995	20,4941	20,6620	15	11,0226
1996	20,2700	20,3097	5	11,4851
1997	21,1837	21,1981	10	11,9187
1998	21,9733	21,9733	1	11,7412

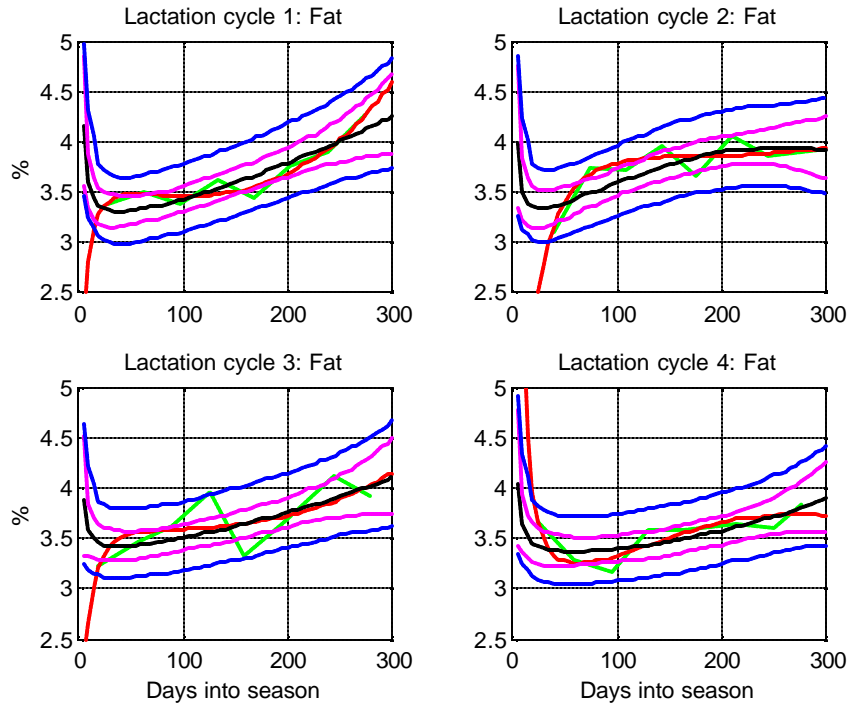
In Figure 6.17 the expected lactation curves of the two calving seasons for the 4parameter Morant, Wood and general exponential models are compared. From this is evident that the results are similar to the comparison for parity commented on earlier.



**Figure 6.17:** The expected lactation curves for seasons 1 and 2 using the 4-parameter Morant model (—), the general exponential model (---) and the Wood model (---) when other cofactors have been averaged out.

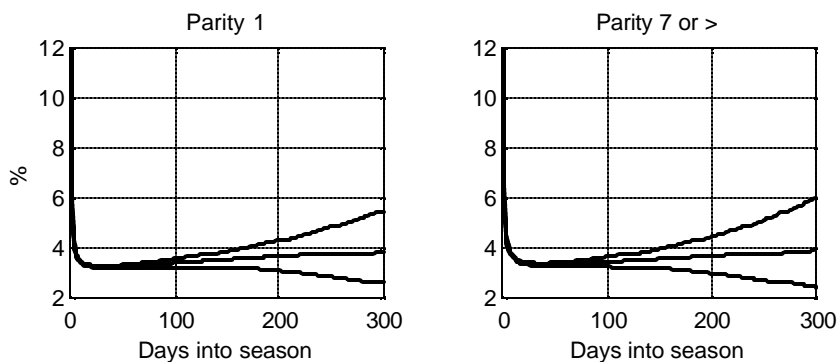
### 6.3.2 The lactation trait fat content of milk

The expected fat content curve of an animal  $i$  for each lactation cycle may again be obtained as before. To illustrate, the expected fat content curves for each of the four lactation cycles of animal  $i = 135$  graphed in black in Figure 6.18 were obtained. In this figure the observed data are again plotted in green, while the red lines represent the least squares estimate of the fat content curve obtained using the 4parameter Morant model. The 90% HPD intervals and 90% prediction interval for the expected fat content curves obtained as explained before are represented by magenta and blue broken lines respectively. The Bayesian approach again eliminates the atypical form of fat content curves obtained through least squares estimation in lactation cycles 1 through to 3.



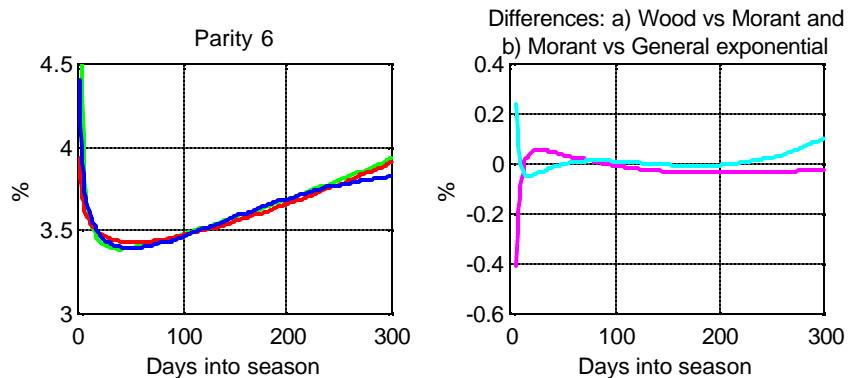
**Figure 6.18:** For each of the four lactation cycle of Animal 135 the expected fat content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

Figure 6.19 shows the expected fat content curves with 95% HPD intervals for parities 1 and 7 or greater when all other cofactors have been averaged out using this method. Very little difference between the results of the two parity groups is evident, although the 90% HPD interval for parity 7 or greater is wider towards the end of the lactation cycle.



**Figure 6.19:** Expected fat content curves with 95% HPD intervals resulting from the 4-parameter Morant model for parities 1 and 7 or greater when other cofactors have been averaged out.

Figure 6.20 compares the expected fat content curves of parity 6 that resulted from the 4-parameter Morant model, the Wood model and the general exponential model. The graph on the left shows that very little differences in both shape and scale occurred among the three models. The graph on the right provides the actual differences between expected fat content curves of the Wood and 4-parameter Morant models (in magenta) and the 4-parameter Morant and general exponential models (in turquoise). The greatest differences are observed at beginning and end of the lactation cycle, although these differences are relatively small when compared to the scale on which fat content is measured.



**Figure 6.20: (1) On the left the expected fat content curves for parity 6 using the Morant model (—), the general exponential model (---) and the Wood model (---) when other cofactors have been averaged out.**

**(2) On the right the differences in fat content between the expected curves for parity 6 the plotted over time for (a) Wood vs Morant model (---) and (b) Morant vs general exponential model (---).**

The results with respect to the expected fat content curves of all seven parities for the 4-parameter Morant model are summarised in Table 6.9. On average the expected fat content of the parities on the first day of the lactation cycle for the Morant model is 7,1% higher than for the Wood model and 6,6% higher than for the general exponential model. These high starting values resulting from the 4-parameter Morant model are however unrealistically high. On average the minimum fat content of the parities for the Morant model are 0,055% lower and from 5 to 18 days earlier than for the Wood model, but 0,009% lower and from 9 to 12 days earlier than for the general exponential model. On day 300 the expected fat content of the Morant model is on average 0,087% higher than that of the general exponential model, but when compared to the Wood model there is no consistent difference on day 300.

**Table 6.9: Summary of the expected fat content for the seven parities using the 4-parameter Morant model.**

Parity	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1	9,7567	3,3025	33	3,8703
2	10,8073	3,3800	36	3,9555
3	12,6422	3,3733	37	3,9143
4	11,0194	3,3908	39	3,9341
5	10,8564	3,3938	38	3,9063
6	10,6673	3,3929	40	3,9324
7 or >	11,3618	3,3838	40	3,9403

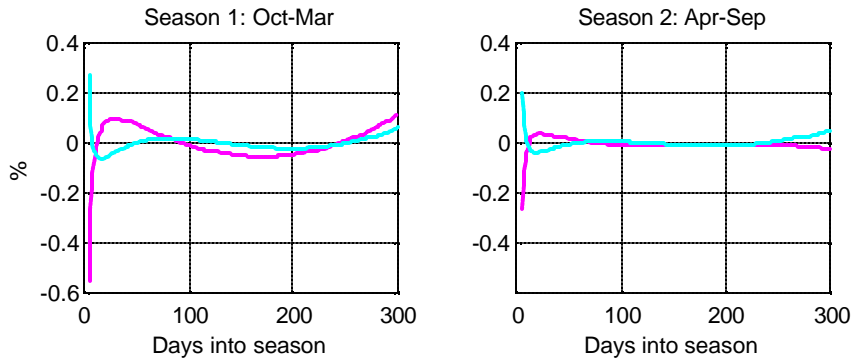
Table 6.10 provides a summary of the expected fat content curves for the four years. When comparing these results of the expected fat content for the four years from the 4-parameter Morant model to that of the Wood and general exponential models, the results are similar to that of the parities above, with the exception of the day of minimum fat content that occurs 10 to 17 days earlier than in the case of the Wood model and 9 to 11 days earlier than for the general exponential model.

**Table 6.10: Summary of the expected fat content for the four years using the 4-parameter Morant model.**

Year	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1995	13,2231	3,4251	38	3,9014
1996	10,1939	3,3661	37	3,9150
1997	9,9220	3,4207	36	3,9079
1998	10,8413	3,2861	39	3,8824

In Figure 6.21 the expected fat content curve of obtained from the 4-parameter Morant model for season 2 differs less from both that of the Wood and general exponential models than is the case for season 1.

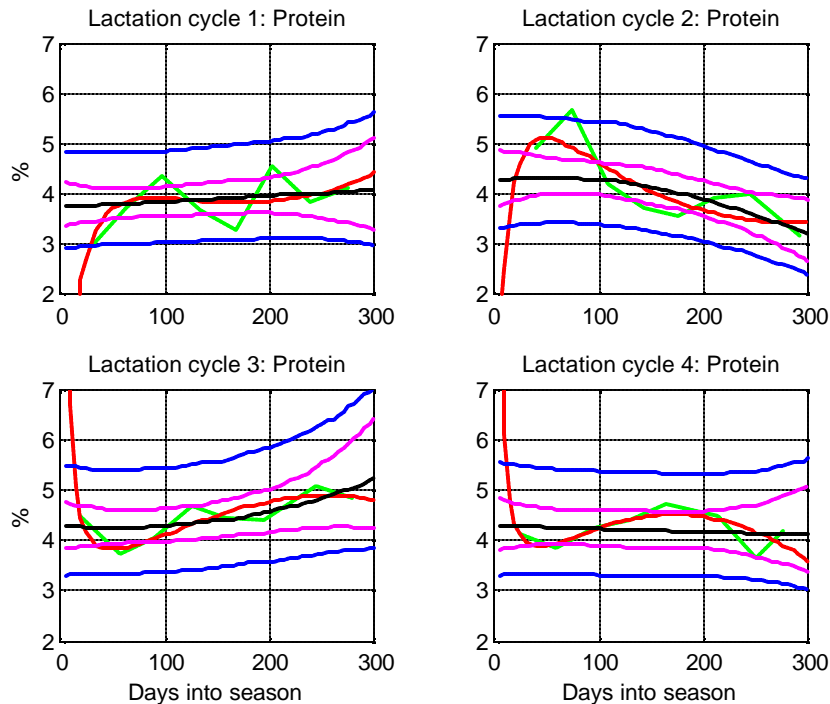




**Figure 6.21:** For seasons 1 and 2 respectively, the differences in expected fat content curves plotted over time for (a) Wood vs Morant model (—) and (b) Morant vs general exponential model (—).

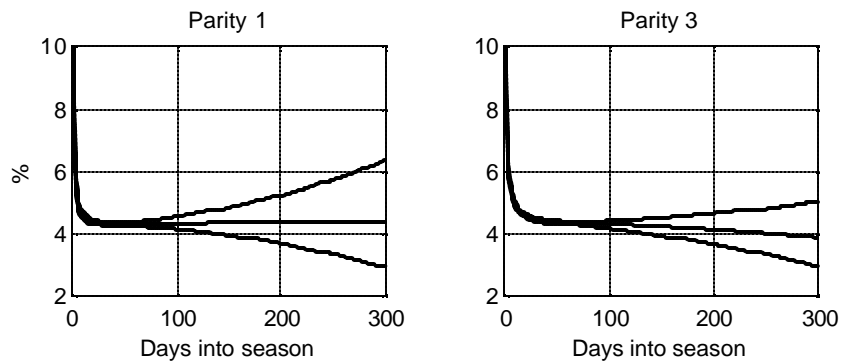
### 6.3.3 The lactation trait protein content of milk

To illustrate the result obtained with respect to the expected protein content for an individual animal using the 4-parameter Morant model, Figure 6.22 provides the expected protein content curves for each of the four lactation cycles of animal  $i = 135$  in black.



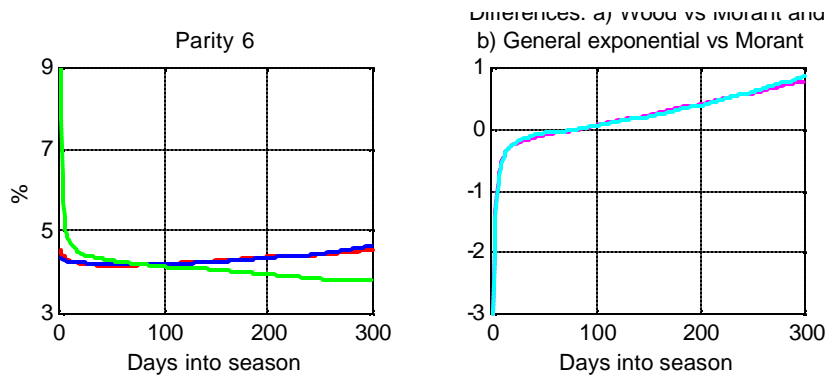
**Figure 6.22:** For each of the four lactation cycle of Animal 135 the expected protein content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the protein curve is given by —. The observed data for Animal 135 is given by —.

Expected protein content curves for the herd with respect to one of the levels of a cofactor of interest was again obtained by applying the same procedure as before. The expected protein content curves for parities 1 and 3 when all other cofactors have been average out, together with the 95% HPD intervals of these expected curves are shown in Figure 6.23. The expected protein curves of these two parities are very similar during the initial stages of lactation but differ in the final stages of lactation.



**Figure 6.23:** Expected protein content curves with 95% HPD intervals resulting from the 4-parameter Morant model for parities 1 and 3 when the other cofactors have been averaged out.

The graph on the left in Figure 6.24 compares the expected protein content curves of parity 6 obtained from the 4-parameter Morant model in green, the general exponential model in blue and the Wood model in red, while in the graph on the right the difference over time between the expected protein content curves of parity 6 for the Wood model against the 4-parameter Morant model is plotted in magenta and that for the 4-parameter Morant model against the general exponential model is plotted in turquoise. From these graphs it is clear that the expected protein content curves of the Wood and general exponential models are similar in shape and scale, but that that these models differ slightly from the 4-parameter Morant model with respect to expected protein content curve of parity 6. In the initial stages of the 4 parameter Morant model estimated the expected protein content to be at a higher level than the Wood and general exponential models. After approximately day 80 until the end of the lactation cycle the 4parameter Morant model estimates the expected protein content to be lower than that of the Wood and general exponential models. On the final day of the lactation cycle the 4-parameter Morant model estimates the expected protein content to be lower than the other two models. This result also applies to the expected lactation curves of all other parities, calving years and calving seasons. The results with respect to the expected protein content of all seven parities are summarised in Table 6.11 below.



**Figure 6.24:** (1) On the left the expected protein content curves for parity 6 using the Morant model (—), the general exponential model (---) and the Wood model (---) when other cofactors have been averaged out. (2) On the right the differences in protein content between the expected curves for parity 6 the plotted over time for (a) Wood vs Morant model (---) and (b) Morant vs general exponential model (—).

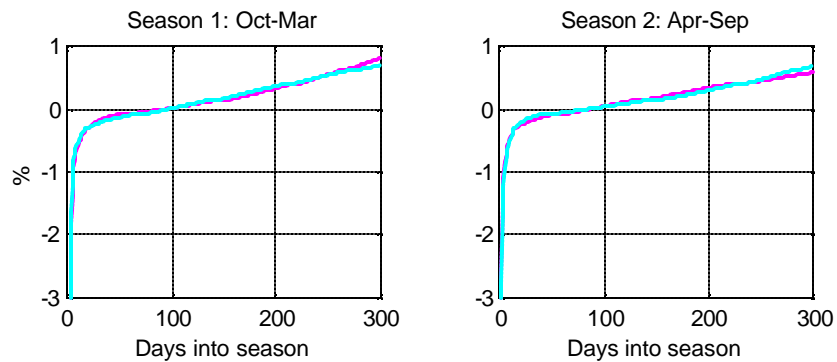
**Table 6.11:** Summary of the expected protein content for the seven parities 4-parameter Morant model.

Parity	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1	13,0102	4,3136	49	4,4249
2	14,6219	4,1895	300	4,1895
3	17,2458	3,9159	300	3,9159
4	14,7471	3,9800	300	3,9800
5	14,3695	3,8659	300	3,8659
6	14,2351	3,7816	300	3,7816
7 or >	14,6661	4,1122	280	4,1137

When comparing the results from Table 6.11 for the expected protein content of the seven parities using the 4-parameter Morant model to the same results generated for the Wood and general exponential models, we see that the minimum expected protein content for parities 2 to 6 from the 4-parameter Morant model is lower than those from the Wood model, while the minimum expected protein content for parities 2 to 7 or > are lower than those from general exponential models. It should however also be noted that for both the Wood and general exponential model the minimum expected protein content is attained early in the lactation cycle (no later than day 74), while for the 4-parameter Morant model this minimum is on day 49 for parity 1, on day 300 for parities 2 to 6 and on day 280 for parity 7 or greater. Results similar to parities 2 to 6 also apply to the expected protein content curves of the four calving years summarised in Table 6.12 below.

**Table 6.12: Summary of the expected protein content for the four years 4-parameter Morant model.**

Year	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1995	17,0806	3,8262	300	3,8262
1996	13,5405	3,8748	300	3,8748
1997	13,1187	4,1921	300	4,1921
1998	15,1139	4,1676	300	4,1676

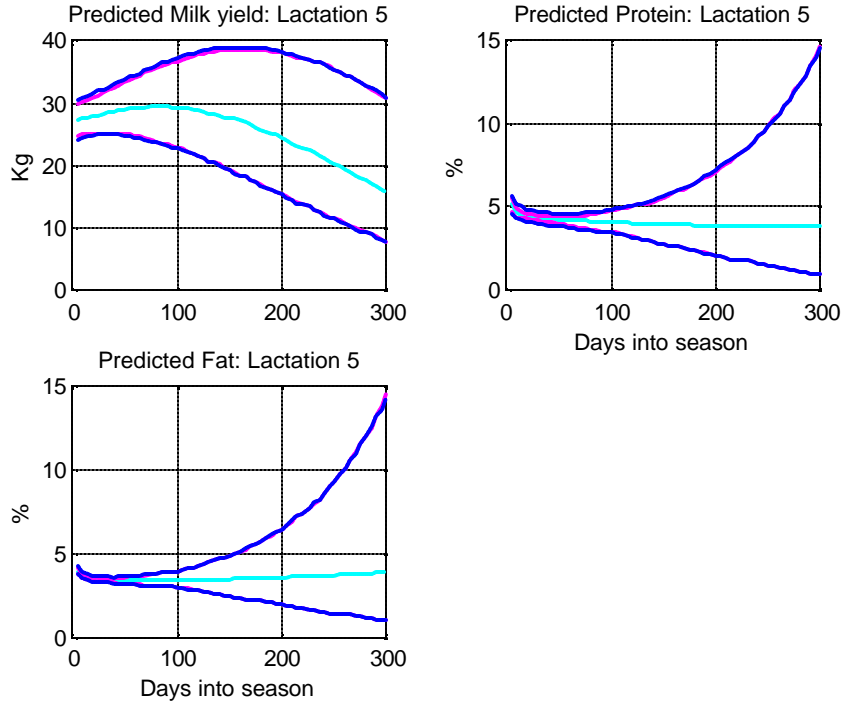


**Figure 6.25: For seasons 1 and 2 respectively, the differences in expected protein content curves plotted over time for (a) Wood vs Morant model (---) and (b) Morant vs general exponential model (—).**

Figure 6.25 confirms that the expected protein content curves for the two seasons using the Morant model differ from that of the Wood and general exponential model in the same manner as explained above.

### 6.3.4 Predictions based on the 4-parameter Morant model

To predict milk yield, fat content and protein content curves for the fifth lactation cycle of animal  $i = 135$  given the performance in the preceding four lactation cycles modelled by means of the 4-parameter Morant model, the procedure applied in section 5.7 of Chapter 5 is again applied here. In Figure 6.26 the predicted curves are represented by turquoise lines, while the 90% HPD and 90% prediction intervals are represented by magenta and blue broken lines respectively.



**Figure 6.26:** For the fifth lactation cycle of Animal 135 the predicted milk yield, fat content and protein content curves is given by —, their 90% HPD interval by - - and 90% prediction interval by - - .

## 6.4 THE ADAPTED WILMINK MODEL

The Wilmink II model was introduced by Wilmink in 1987 to model lactation in Dutch Friesian cows. This model assumes that the expected milk yield (in kg/day) of an animal at time  $t$  can be represented over the lactation period by

$$E(W_t) = a + bt + c \exp(-wt) \quad (6.16)$$

where  $w = 0,05$ . Taking into account the discussion on assumptions made about the errors in section 3.5 of Chapter 3, the Wilmink II model is unsuitable to fit using the Bayesian approach suggested in Chapter 4. To be suitable the lactation model should be such that, when the assumption of multiplicative errors is applied and a log transformation of the model is performed, the result is a model of linear form to which additive errors applies. As a result the Wilmink II model had to be adapted in such a way that, with the assumption of multiplicative errors and after a natural log transformation, it would be of linear form.

The adapted Wilmink model suggested in this study, assumes that the expected milk yield (in kg/day) of an animal at time  $t$  can be represented over the lactation period by

$$E(W_t) = \exp(a + bt + c \exp(-wt)). \quad (6.17)$$

where  $-\infty < a < \infty$ ,  $b < 0$ ,  $c < 0$  and  $0 < w < 1$  instead of fixing the value of  $w$  at 0,05 or 0,61 as previously done by researchers. The parameters  $a$ ,  $b$  and  $c$  are unknown and may differ from one animal to another, while the parameter  $w$  remains the same throughout the herd.

The adapted Wilmink model also has the ability to take on a convex form and may therefore be applied to model the milk composition traits fat and protein content. The parameters of the functions of the composition traits fat and protein content over time are required to be  $-\infty < a < \infty$ ,  $b > 0$ ,  $c > 0$  and  $0 < w < 1$  in order for the functions to be convex in nature.

If multiplicative errors are assumed for the adapted Wilmink model in (6.17), then by performing a natural log transformation on this model, the observation model for animal  $i$  is written as:

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) = a_{ijs} + b_{ijs} t_{ijp} + c_{ijs} \exp(-wt_{ijp}) + e_{ijps} \quad (6.18)$$

where  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ , and for the Jersey data specifically  $i = 1, \dots, 1411$ ,  $j = 1, \dots, 4$ ,  $s = 1, 2, 3$ ,  $p = 1, \dots, n_{ij}$  and  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ , where  $8 \leq n_{ij} \leq 10$ . The parameter  $w$  applies to the herd of which animal  $i$  forms part and not to a lactation trait of a specific lactation cycle of animal  $i$ .

The generalised linear model form of the model for animal  $i$  during lactation cycle  $j$  is as described in (4.3) for the general case:

$$Y_{ij} = X_{ij} \mathbf{M}_{ij} + E_{ij} \quad (6.19)$$

where  $\text{vec } E'_{ij}(3n_{ij} \times 1) \sim N(\text{vec } 0, F \otimes \mathbf{I}_{n_{ij}})$ . In the case of the adapted Wilmink model there are  $v = 3$  regression coefficients in the generalised linear form of the lactation model for the  $i^{\text{th}}$  animal during its  $j^{\text{th}}$  lactation cycle for the lactation trait indicated by  $s = 1, 2, 3$ . These are  $a_{ijs}$ ,  $b_{ijs}$  and  $c_{ijs}$ , so that

$$\mathbf{M}_{ij}(3 \times 3) = \begin{bmatrix} a_{ij1} & a_{ij2} & a_{ij3} \\ b_{ij1} & b_{ij2} & b_{ij3} \\ c_{ij1} & c_{ij2} & c_{ij3} \end{bmatrix}. \quad (6.20)$$

The  $r$ -th row of design matrix  $\mathbf{X}_{ij}$  ( $n_{ij} \times 3$ ) contains the elements that would return the adapted Wilmlink model in generalised linear form when multiplied with the matrix of regression coefficients  $\mathbf{M}_{ij}$  in (6.20) above, i.e.

$$\mathbf{X}_{ij}^{(r)} = [1 \quad t_{ijr} \quad \exp(-wt_{ijr})]. \quad (6.21)$$

The model extended for animal  $i$  to all  $q = 4$  lactation cycles then is,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_{i1} \\ \mathbf{Y}_{i2} \\ \mathbf{Y}_{i3} \\ \mathbf{Y}_{i4} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (6.22)$$

where  $\text{vec } \mathbf{E}'_i (3n_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^4 n_{ij}$ ,  $\mathbf{X}_i (n_i \times 12) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i (12 \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \\ \mathbf{M}_{i3} \\ \mathbf{M}_{i4} \end{bmatrix}. \quad (6.23)$$

The matrix of covariances between the model parameters of the same animal in successive lactation cycles,  $\mathbf{R}$  remains of same structural form as denoted in (6.8) for the Jersey data. The additional information on the factors parity number, region, calving year and calving season available in the Jersey data has to again be included in the model. This is done as before by constructing a covariate matrix  $\mathbf{Z}_i (17 \times 4)$  for each animal  $i$  as explained in section 5.3 of Chapter 5. This covariate matrix is then used as described in section 4.1 of Chapter 4 for the generalised Bayesian model.

The adapted Wilmlink model has one additional parameter  $w$ , that is *not* included in the model as one of the regression coefficients of the generalised linear form of the lactation model for the  $i^{\text{th}}$  animal during its  $j^{\text{th}}$  lactation cycle for the lactation trait denoted by  $s = 1, 2, 3$ , or  $\mathbf{M}_{ij}$ . This requires the following prior be specified for this parameter:

$$\pi(w) \propto \frac{1}{w}, \quad (6.24)$$

The prior distributions of  $\mathbf{m}_i$ ,  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  remain as specified in (4.12) to (4.16).

The conditional distributions of the model parameters  $\mathbf{m}_i, \mathbf{B}, \mathbf{F}^{-1}, \mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  still are as given in equations (4.17) to (4.21), while conditional distribution of  $w$ , contained in row  $r$  of  $\mathbf{X}_{ij}(n_{ij} \times 3)$  as  $\mathbf{X}_{ij}^{(r)} = [1 \quad t_{ijr} \quad \exp(-wt_{ijr})]$ , is derived as follows:

$$\begin{aligned}
& f(w \mid \mathbf{y}_i, \mathbf{m}_i, \mathbf{F}) \\
& \propto f(\mathbf{y}_i \mid \mathbf{m}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R}) \cdot f(w) \\
& \propto \prod_{i=1}^k \exp\left[-\frac{1}{2}(\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i)\mathbf{m}_i)'(\mathbf{F} \otimes \mathbf{I}_{n_i})^{-1}(\mathbf{y}_i - (\mathbf{I}_u \otimes \mathbf{X}_i)\mathbf{m}_i)\right] \cdot \frac{1}{w} \\
& \propto \text{etr}\left[-\frac{1}{2}\sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i\mathbf{M}_i)\mathbf{F}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\mathbf{M}_i)'\right] \cdot \frac{1}{w} \\
& \propto w^{-1} \text{etr}\left[\mathbf{F}^{-1}\left\{-\frac{1}{2}\sum_{i=1}^k (\mathbf{Y}_i - \mathbf{X}_i\mathbf{M}_i)'(\mathbf{Y}_i - \mathbf{X}_i\mathbf{M}_i)\right\}\right] \tag{6.25}
\end{aligned}$$

Next the Gibbs sampler as described in section 4.2 of Chapter 4 with the full conditional distributions of all the model parameters of interest as set out in equations (4.17) to (4.21) and the Metropolis-Hastings algorithm using equation (6.25) above were applied to obtain the marginal posterior distributions conditional on the observed data only. The moments estimators from the sampling distributions of  $\mathbf{S}$  and  $\mathbf{R}$  were again used as the hyperparameters  $\mathbf{G}$  and  $\mathbf{P}$  required to generate marginal distributions for  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , while the degrees of freedom were again kept as small as possible,  $\mathbf{d} = 4$  and  $g = 10$ .

MCMC procedures applied were put through a ‘‘burn-in’’ period of 2000 simulation iterations, after which 10 000 sets of parameters were generated and kept using equations (4.17) to (4.21), together with (6.25) for  $w$ . The distributions of  $\mathbf{M}_i$  and  $\mathbf{B}$  were as before obtained by sampling from two normal distributions, while the distributions of  $\mathbf{F}^{-1}, \mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$  were again generated from their respective Wishart distributions as explained at the end of section 5.2 in Chapter 5. However, to obtain the distribution of  $w$  required a somewhat more complicated sampling scheme in the form of the Metropolis-Hastings algorithm discussed in section 4.5 of Chapter 4.

To obtain the marginal posterior distribution of  $w$  from (6.25) the Metropolis-Hastings algorithm is employed with the restriction that  $0 < w < 1$  placed on the value of this parameter. A candidate value  $w^c$  of the following form is obtained:



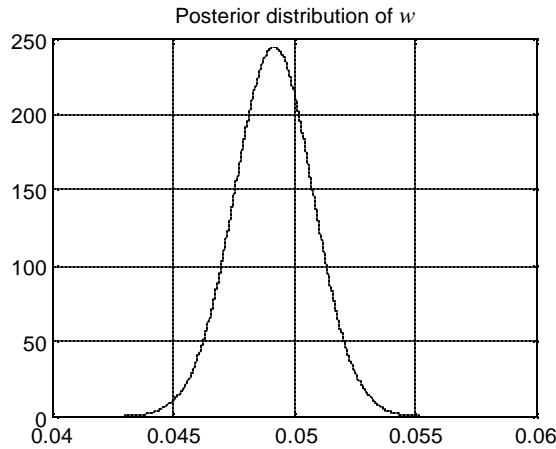
$$w^c = w \exp(z) \quad (6.26)$$

where  $z$  is randomly sampled from a uniform distribution in the interval  $-0,5$  to  $0,5$ . This candidate value is then accepted with probability

$$\min \left( 1, \frac{f(w^c | \mathbf{Y}, \mathbf{M}, \mathbf{F})}{f(w | \mathbf{Y}, \mathbf{M}, \mathbf{F})} \right) \quad (6.27)$$

As starting point value to  $w$  a randomly selected value in the interval 0 to 1 was used. As mentioned before, an acceptable acceptance rate for such proposal values is considered to usually be between 40% and 70%. In the case of  $w$  the acceptance rate was 47%.

Only the 10 000 sets of parameters generated for  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$ ,  $\mathbf{S}^{-1}$ ,  $\mathbf{R}^{-1}$  and  $w$  were stored due to both speed and storage constraints in Matlab. The posterior distribution of the parameter  $w$  that resulted through MCMC simulation is depicted in Figure 6.27 below. This posterior distribution has a mean of 0,0491.



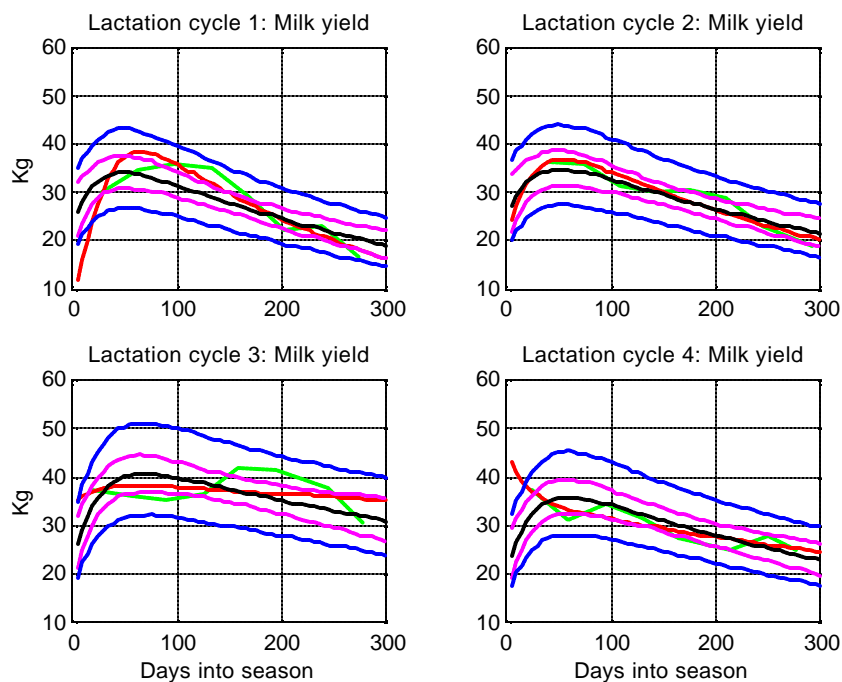
**Figure 6.27:** The posterior distribution of the parameter  $w$  contained in the adapted Wilmink model.

At 90% HPD level the intervals for the elements of the regression matrix  $\mathbf{B}$  ( $9 \times 17$ ), constructed using the 10 000  $\mathbf{B}$  matrices resulting from the Gibbs sampler, indicated 94 of the 153 elements of this matrix to significantly affect the parameters of the three lactation trait curves, milk yield, percentage of fat and percentage of protein in milk composition. All nine base level effects are significant, while the effects of all covariates on the nine trait curve parameters are significant for at least 2, but in some cases all 9 parameters. As a result all cofactors were kept in the model.

Once the Wilmlink model had been fitted to the Jersey data and the marginal distributions of the parameters obtained through Gibbs sampling, further investigation of the results for Jersey data based on this lactation model was performed.

### 6.4.1 The lactation trait milk yield

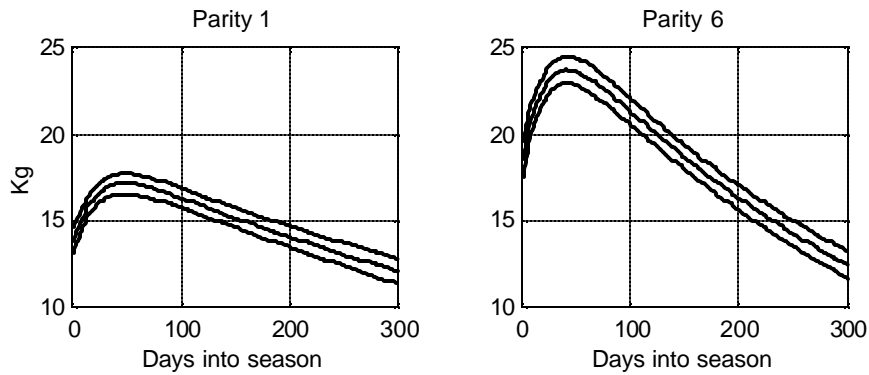
By following the usual procedure, the expected lactation curve for each of the four lactation cycles of animal  $i = 135$ , where lactation is modelled using the adapted Wilmlink model, were determined. These expected lactation curves are represented by the black solid lines in Figure 6.28, together with their 90% HPD and 90% prediction intervals represented by magenta and blue broken lines respectively. The observed data are represented by green lines and the least squares estimate of the adapted Wilmlink model for milk yield by red lines.



**Figure 6.28:** For each of the four lactation cycle of Animal 135 the expected lactation curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

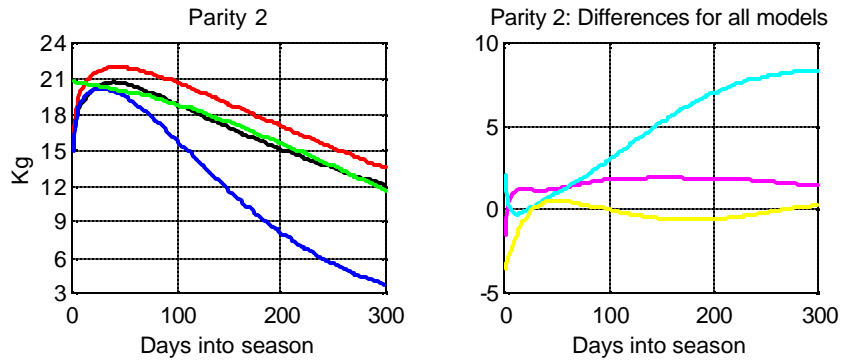
Again using the procedure described in section 5.5 of Chapter 5, the expected lactation or milk yield curve of the herd with respect to one of the levels of a cofactor, which also provides the Bayesian equivalent of the Standard lactation curve (SLAC) with respect to this

cofactor level, may be obtained. Figure 6.29 gives the expected lactation curves with 95% HPD intervals for parities 1 and 6 when all other cofactors have been averaged out.



**Figure 6.29:** Expected lactation curves with 95% HPD intervals resulting from the adapted Wilmink model for parities 1 and 6 when other cofactors have been averaged out.

Figure 6.30 compares the expected lactation curves of parity 2 obtained for all four lactation models. From the graph on the left of this figure where the expected lactation curves of parity 2 obtained for the adapted Wilmink model is given in black, for the 4-parameter Morant model in green, for the general exponential model in blue and for the Wood model in red, it follows that the result of the adapted Wilmink model most similar to that of the 4-parameter Morant model with respect to scale, followed by the Wood model and then only the general exponential model. This is confirmed by the graph on the right of this figure in which the difference over time between the expected lactation curves of parity 2 for the Wood against the adapted Wilmink model is plotted in magenta, for the adapted Wilmink against the general exponential model in turquoise and the adapted Wilmink against the 4-parameter Morant model in yellow. From this is clear that the resulting expected lactation curves of the adapted Wilmink and 4-parameter Morant models are more similar in scale than the expected lactation curves of the adapted Wilmink model either the Wood model or the general exponential model. The shape of the adapted Wilmink model is however closer to that of the Wood model. This result also applies to the expected lactation curves of all other parities.



**Figure 6.30:** (1) On the left the expected lactation curves for parity 2 using the adapted Wilmink model (—), the Morant model (---), the general exponential model (- · -) and the Wood model (····) when all other cofactors have been averaged out. (2) On the right the differences in milk yield between the expected curves for parity 2 the plotted over time for (a) Wood vs adapted Wilmink (---), (b) adapted Wilmink vs general exponential (- · -) and (c) adapted Wilmink vs Morant (—) models.

The full result of the expected lactation curves of all seven parity groups are summarised in Table 6.13. The lowest level of expected peak milk yield occurs in first parity, a sharp rise in expected peak level occurs during the second parity followed by a steady increase in expected peak milk yield up to parity 6. In parity 7 or > expected peak yield again drops off slightly. This is similar to the result obtained for the Wood model in section 5.5 of Chapter 5. For the adapted Wilmink time of expected peak yield is latest in parity 1 and occurs on the same day as that of the Wood model.

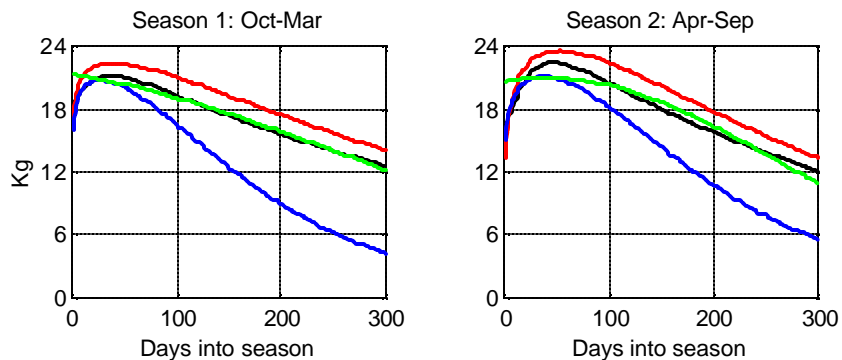
**Table 6.13:** Summary of the expected milk yield for the seven parities using the adapted Wilmink model.

Parity	Expected kg's milk on day 1	Maximum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1	13,8344	17,1406	48	12,0914
2	17,2717	20,6966	40	12,0967
3	18,2434	22,4932	41	12,4924
4	18,2026	23,1510	42	12,5651
5	17,7031	23,2800	44	12,5751
6	18,6048	23,6892	42	12,4617
7 or >	17,3409	22,9125	44	12,1696

In Table 6.14 the results with respect to the expected lactation curves of the four years are summarised. The expected lactation curves commence at a slightly lower level than that of the Wood or 4-parameter Morant models, but at a higher level than the general exponential model. Time of expected peak yield for the adapted Wilmink model is closest to that of the Wood model.

**Table 6.14: Summary of the expected milk yield for the four years using the adapted Wilmink model.**

Year	Expected kg's milk on day 1	Minimum expected kg's milk	Day of maximum expected kg's milk	Expected kg's milk on day 300
1995	17,1853	21,7217	43	12,1161
1996	16,8395	21,0316	43	12,2487
1997	17,3553	22,0936	44	12,6990
1998	17,5939	22,3162	43	12,3312

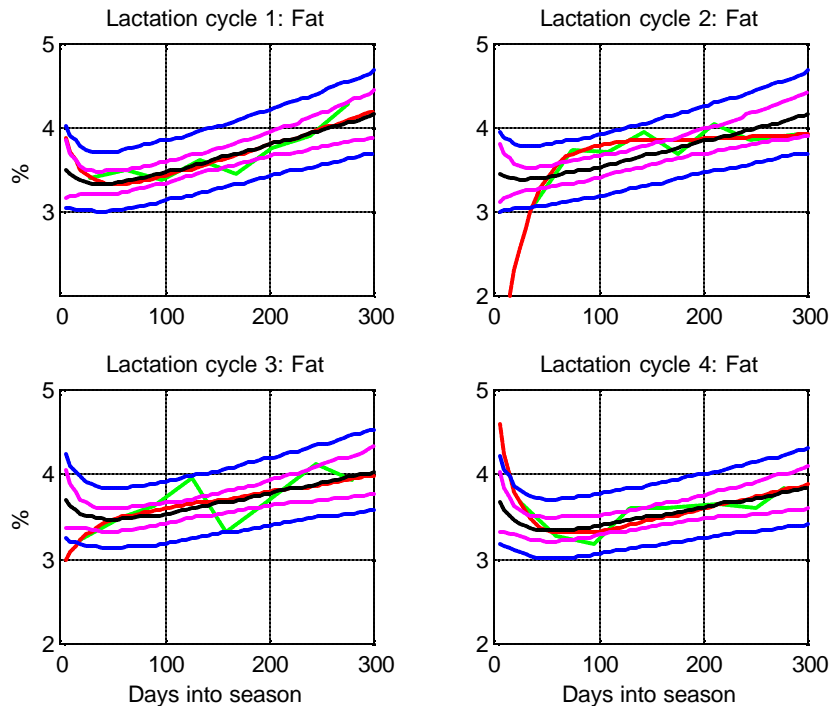


**Figure 6.31: The expected lactation curves for seasons 1 and 2 using the adapted Wilmink model (—), the 4-parameter Morant model (---), the general exponential model (- · -) and the Wood model (·····) when the other cofactors have been averaged out.**

In Figure 6.31 the expected lactation curves for the two seasons resulting from the adapted Wilmink model, the Wood model, the 4-parameter Morant model and the general exponential model are compared. The result of the expected lactation curve of the adapted Wilmink model again is closest in scale to that of the 4-parameter Morant model, while the shape of the curve for the adapted Wilmink model more closely resembles that of the Wood model.

## 6.4.2 The lactation trait fat content of milk

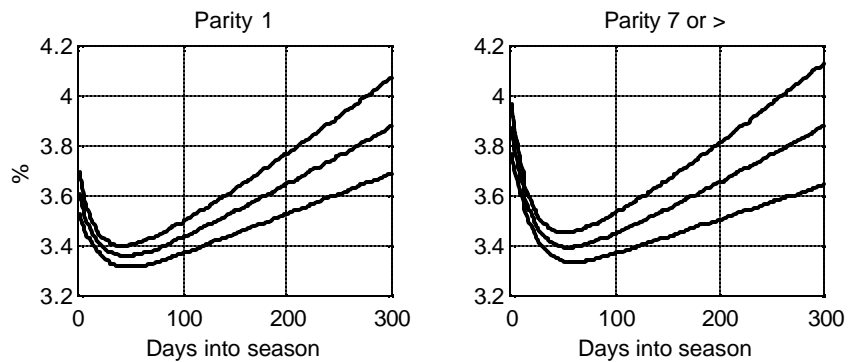
The expected fat content curves for each of the four lactation cycles of animal  $i = 135$ , together with the 90% HPD and 90% prediction intervals of these expected fat content curves are given in Figure 6.32. The observed data values, as well as the least squares estimate of the fat content curve using the adapted Wilmink model are also provided for each lactation cycle.



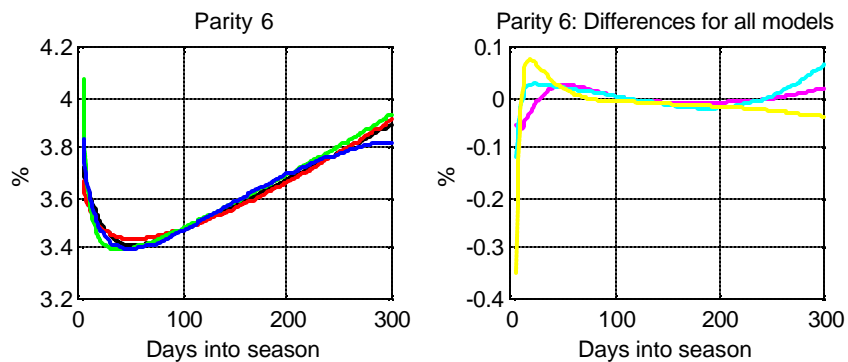
**Figure 6.32:** For each of the four lactation cycle of Animal 135 the expected fat content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 135 is given by —.

For both lactation cycles 2 and 3 the expected fat content curves following from the Bayesian approach are not atypical, while the least squares estimates of the fat content curves for these lactation cycles are. The Bayesian approach therefore again succeeds in preventing the occurrence of an atypical result.

In Figure 6.33 the expected fat content curves and their 95% HPD intervals for parities 1 and 7 or > when all other cofactors in the model have been averaged out, are given. Both the shape and scale of the expected fat content curves for these parities are more similar to the equivalent results from the Wood model than to the other model discussed in this chapter.



**Figure 6.33:** Expected fat content curves with 95% HPD intervals resulting from the adapted Wilmink model for parities 1 and 7 or greater when other cofactors have been averaged out.



**Figure 6.34:** (1) On the left the expected fat content curves for parity 6 using the adapted Wilmink model (—), the Morant model (---), the general exponential model (- · -) and the Wood model (·····) when all other cofactors have been averaged out. (2) On the right the differences in fat content between the expected curves for parity 6 the plotted over time for (a) Wood vs adapted Wilmink (---), (b) adapted Wilmink vs general exponential (- · -) and (c) adapted Wilmink vs Morant (—) models.

Figure 6.34 compares the expected fat content curves of parity 6 obtained using the adapted Wilmink, the Wood, the 4-parameter Morant and the general exponential models. In the graph on the left it follows that there are very little differences in these models with respect to scale and shape of the expected fat content curve for this parity. This is confirmed by the graph on the right of this figure, which also indicates that the largest of these small differences occur at the start and end of the lactation cycle. The results with respect to the expected fat content curves of all seven parities are summarised in Table 6.15. On average the minimum expected fat content of the parities is 0,021% lower for the adapted Wilmink model than for the Wood model, 0,034% and 0,026% lower for the Morant and general exponential models than for the adapted Wilmink. Time of minimum expected fat content for the adapted Wilmink model is closest to that of the Wood model.

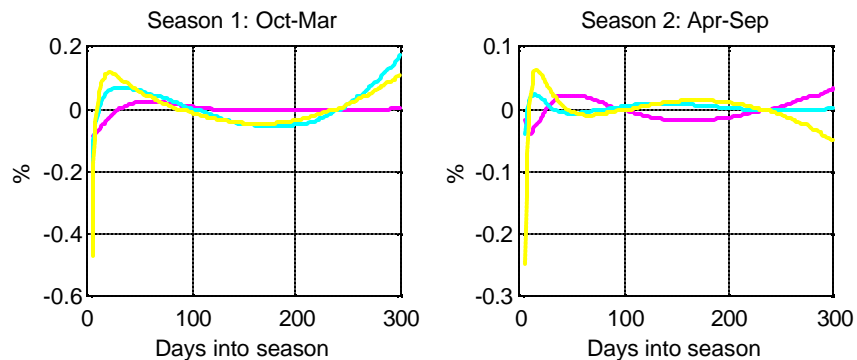
**Table 6.15: Summary of the expected fat content for the seven parities using the adapted Wilmink model.**

Parity	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1	3,6103	3,3617	46	3,8796
2	3,7833	3,4227	50	3,9689
3	3,8079	3,4210	51	3,9582
4	3,8508	3,4182	53	3,9486
5	3,8475	3,4206	54	3,9146
6	3,8180	3,4154	53	3,8929
7 or >	3,8701	3,3982	55	3,8825

The results of the expected fat content curves for the four calving years are summarised in Table 6.16. When the expected fat content curves for the four years using the adapted Wilmink model are compared to those obtained using the Wood, 4parameter Morant and general exponential models, the results are similar to that of the parities mentioned above.

**Table 6.16: Summary of the expected fat content for the four years using the adapted Wilmink model.**

Year	Expected fat % on day 1	Minimum expected fat %	Day of minimum expected fat %	Expected fat % on day 300
1995	3,7972	3,4673	51	3,9306
1996	3,7596	3,3976	50	3,9244
1997	3,8600	3,4572	53	3,9430
1998	3,7733	3,3142	53	3,8833



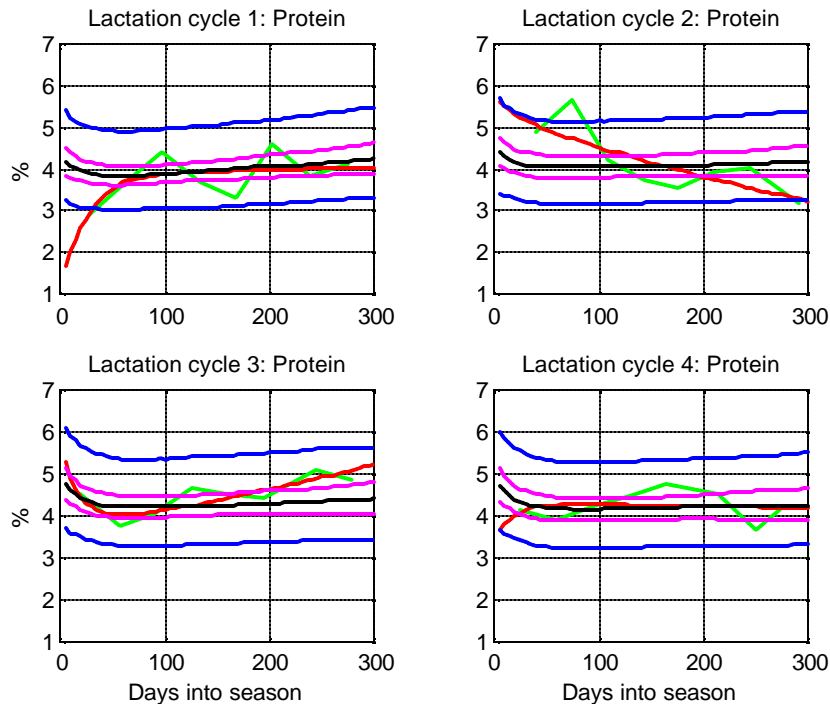
**Figure 6.35: For seasons 1 and 2 respectively, the differences in expected fat content curves over time for (a) Wood model vs adapted Wilmink model (---), (b) adapted Wilmink model vs general exponential model (---) and (c) adapted Wilmink model vs Morant model (—).**



From Figure 6.35 it follows that in season 1 the differences between the expected fat content curves of the adapted Wilmlink model compared to the Wood model are smallest, while in season 2 these differences are smallest for the adapted Wilmlink model compared to the general exponential model.

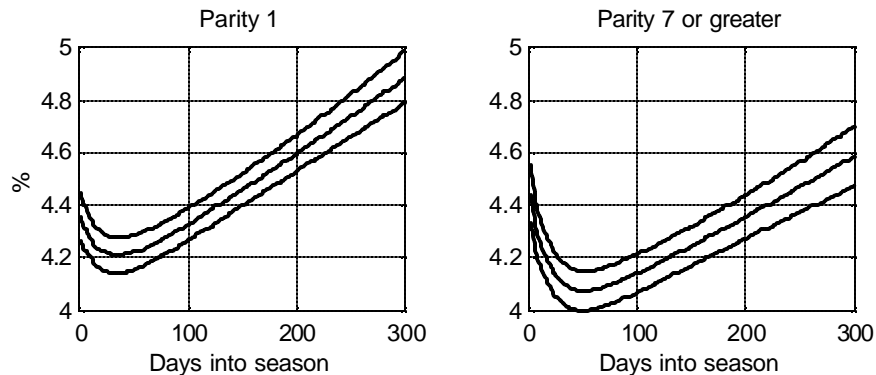
### 6.4.3 The lactation trait protein content of milk

The expected protein content curve for each of the four lactation cycles of animal  $i = 135$ , the 90% HPD and prediction intervals of the expected protein content curves, the observed data and the least squares estimate of the protein curve for this animal using the adapted Wilmlink model are shown in Figure 6.36. The atypical fit of the least squares estimate of the adapted Wilmlink model occurring in lactation cycles 1 and 4 are again successfully eliminated using the Bayesian approach.



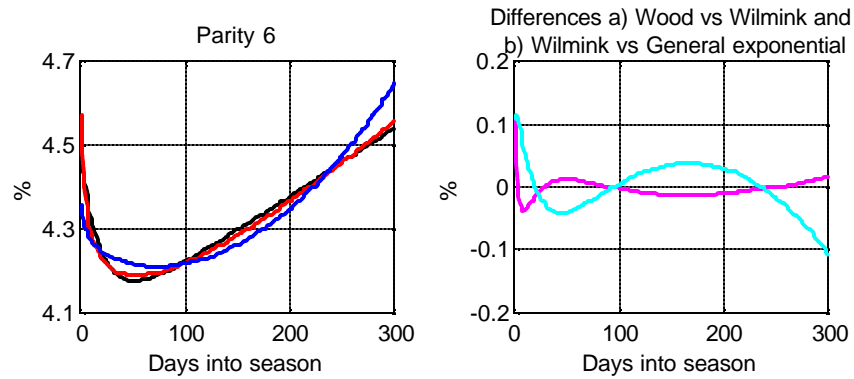
**Figure 6.36** For each of the four lactation cycle of Animal 135 the expected protein content curve is given by —, its 90% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the protein curve is given by —. The observed data for Animal 135 is given by —.

By applying the same procedure as explained before, the expected protein content curve for the herd with respect to one of the levels of a cofactor of interest may again be obtained, for all cofactor levels. To illustrate such a result, Figure 6.37 gives the expected protein content curves together with 95% HPD intervals for parities 1 and 7 or > when the effects of all other cofactors have been eliminated through averaging.



**Figure 6.37:** Expected protein content curves with 95% HPD intervals resulting from the adapted Wilmink model for parities 1 and 7 or greater when the other cofactors have been averaged out.

In Figure 6.38 the expected protein content curves of parity 6 obtained using the adapted Wilmink model, the Wood model and the general exponential model are compared. The expected protein content curve of this parity obtained from the 4parameter Morant model was dropped from this comparison after it was discovered that in the initial stages of lactation it differed considerably from the results of the other three models resulting in differences among these models being difficult to investigate. From the graph on the left of this figure where the expected lactation curves of parity 6 obtained for the adapted Wilmink model is given in black, for the general exponential model in blue and for the Wood model in red, it follows that the result of the adapted Wilmink model is most similar to that of the Wood model with respect to both shape and scale. This is confirmed by the graph on the right of this figure in which the differences over time between the expected lactation curves of parity 6 for the Wood model against the adapted Wilmink model is plotted in magenta and for the adapted Wilmink model against the general exponential model in turquoise. This result also applies to the expected lactation curves of all other parities.



**Figure 6.38:** (1) On the left the expected protein content curves for parity 6 using the adapted Wilmink model (—), the general exponential model (- - -) and the Wood model (.....) when all other cofactors have been averaged out. (2) On the right the differences in protein content between the expected curves for parity 6 the plotted over time for (a) Wood model vs adapted Wilmink model (- - -) and (b) adapted Wilmink model vs general exponential model (- · ·).

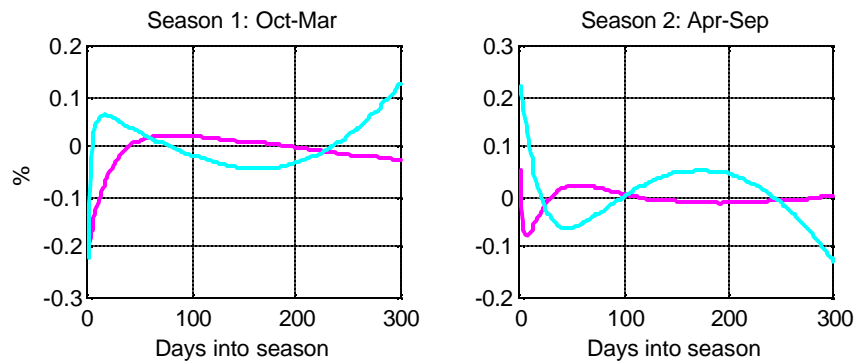
The results with respect to the expected protein content curves of all seven parities are summarised in Table 6.17 below. The greatest variation in differences between the results of the adapted Wilmink model and any one of the other three models occur with respect to the expected protein content curves. The same is true for the expected protein content curves for the four calving years summarised in Table 6.18.

**Table 6.17:** Summary of the expected protein content for the seven parities using the adapted Wilmink model.

Parity	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1	4,3543	4,2107	35	4,8862
2	4,4453	4,2976	38	4,8435
3	4,4851	4,2614	45	4,7595
4	4,4900	4,2227	48	4,6881
5	4,4968	4,1639	51	4,6367
6	4,4638	4,1758	52	4,5412
7 or >	4,4435	4,0775	51	4,5870

**Table 6.18: Summary of the expected protein content for the four years using the adapted Wilmink model.**

Year	Expected protein % on day 1	Minimum expected protein %	Day of minimum expected protein %	Expected protein % on day 300
1995	4,3678	4,1087	45	4,6513
1996	4,3936	4,1577	47	4,6039
1997	4,5257	4,2701	45	4,8335
1998	4,5307	4,2767	47	4,7322



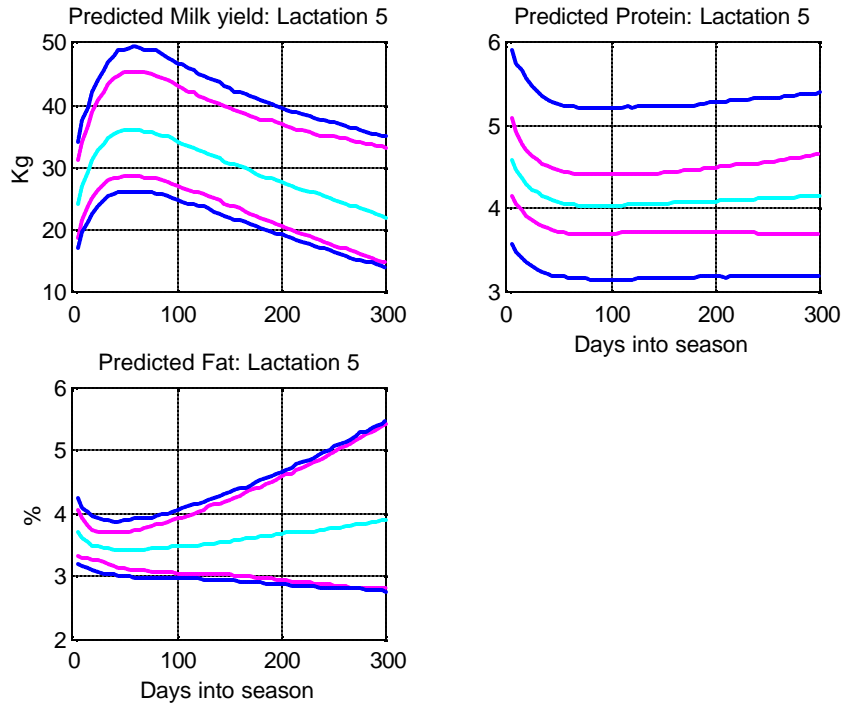
**Figure 6.39: For seasons 1 and 2 respectively, the differences in expected protein content curves plotted over time for (a) Wood vs adapted Wilmink model (---) and (b) adapted Wilmink vs general exponential model (- - -).**

Figure 6.39 also shows that the differences between the expected protein content curves of the adapted Wilmink and Wood models are smaller than those between the curves of the adapted Wilmink and general exponential models. The greatest differences occur between the expected protein content curves of the adapted Wilmink and 4parameter Morant models not shown in the illustration.

#### 6.4.4 Predictions based on the adapted Wilmink model

Again using the prediction procedure applied in section 5.7 of Chapter 5, the milk yield, fat content and protein content curves for fifth lactation cycle of animal  $i = 135$  based on the results from the preceding four lactation cycles modelled by means of the adapted Wilmink model, were predicted. The results are provided in Figure 6.40 below, where the solid turquoise line represents the predicted curve, while the 90% HPD intervals are given as magenta coloured broken lines and the 90% prediction intervals as blue broken lines. It is

noted that for the adapted Wilmlink model these intervals are not as wide for predictions of the lactation traits in the fifth lactation cycle as they were for the general exponential and 4 parameter Morant models discussed earlier in this chapter.



**Figure 6.40:** For the fifth lactation cycle of Animal 135 the predicted milk yield, fat content and protein content curves is given by —, their 90% HPD interval by - - and 90% prediction interval by - - .

## 6.5 ALTERNATIVE LACTATION MODELS IN SUMMARY

When comparing the expected milk yield or lactation curves for the levels of the cofactor of interest, when the effects of all other cofactors were averaged out, obtained using the general exponential, 4parameter Morant, adapted Wilmlink and Wood models, it follows that the shape of the Wood and adapted Wilmlink models are very similar, although the scale of the adapted Wilmlink model with respect to expected milk yield corresponds more to that of the 4-parameter Morant model. The general exponential model differs most with respect to both shape and scale of the expected milk yield curves from that of the other models and possibly result in too low levels of expected milk yield towards the end of the lactation cycle, making its shape less desirable. When considering the expected milk yield curves, the 4-parameter

Morant model is criticised for the fact that it often indicates peak milk yield to occur when the lactation cycle commences which is considered not to be true for lactation in Jersey cows. Therefore, of the three models discussed in this chapter the adapted Wilmink model best describes expected milk yield.

With respect to the expected fat content curves of the levels of a cofactor of interest, when the effects of the other cofactors have been eliminated through averaging, similar results in both shape and scale were obtained from all four lactation models considered in this study. The 4-parameter Morant model is however worth taking special note of with respect to the expected fat content curves, as this is the only lactation trait it seems to model effectively when considering the shapes of lactation trait curves that are considered realistic with respect to lactation in Jersey cows.

Considering the results of the expected protein content curves for the levels of a cofactor of interest, when the effects of the other cofactors have been average out, the general exponential model, adapted Wilmink model and the Wood model are very similar in shape and scale, although the resulting expected protein content curves of the adapted Wilmink model are closest to that of the Wood model. The 4-parameter Morant model, however, takes on a less realistic form with respect to the expected protein content curves. This same result was also obtained when predicting the protein content curve of the fifth lactation cycle based on the results for the preceding four cycles using the 4-parameter Morant model.

With respect to the 90% HPD and prediction intervals obtained for prediction of the fifth lactation cycle based on the results from the preceding four cycles, the adapted Wilmink model displayed the most consistent interval widths through out the lactation cycle in all three lactation traits.

Considering all of the above, the adapted Wilmink model fares best of the three models considered in this chapter at describing the expected lactation curves of a level of the cofactor of interest when the effects of all other cofactors have been eliminated. To further investigate the appropriateness of the four models considered in this study in modelling lactation for the Jersey data under consideration, a diagnostic comparison of these lactation models will be performed in the next chapter.

# CHAPTER 7

## BAYES FACTORS AND MODEL COMPARISON

### 7.1 BAYES FACTORS

One of the first and probably most significant contributions to hypothesis testing in a Bayesian context was that of Jeffreys (1935, 1961). His groundbreaking work allowed for the comparison of predictions resulting from two competing scientific theories through the introduction of statistical models that represent the probability of the data according to each theory. The posterior probability that one of the theories is correct is then calculated by means of Bayes' theorem. This has since developed into formulating two the competing statistical models as two hypotheses and calculating, as the so-called Bayes factor, the posterior probability of the model under one of the hypotheses when it is assumed that the prior probabilities of the two hypotheses are equal.

Kass and Raftery (1995) explain the approach used to determine Bayes factors as follows. It is assumed that the data  $\mathbf{D}$  resulted under one of two models specified in the hypothesis  $H_1$  and  $H_2$  respectively, according to probability density  $p(\mathbf{D}|H_1)$  or  $p(\mathbf{D}|H_2)$ . Using the prior probabilities  $p(H_1)$  and  $p(H_2) = 1 - p(H_1)$ , the data produce the posterior probabilities  $p(H_1|\mathbf{D})$  and  $p(H_2|\mathbf{D}) = 1 - p(H_1|\mathbf{D})$ . From this result it follows that prior information is transformed through consideration of the data to posterior information, and therefore the transformation itself represents the evidence provided by the data.

Next consider the odds ratio scale,  $\text{Odds} = \frac{\text{Probability}}{1 - \text{Probability}}$ , together with Bayes' theorem:

$$p(H_k|\mathbf{D}) = \frac{p(\mathbf{D}|H_k)p(H_k)}{p(\mathbf{D}|H_1)p(H_1) + p(\mathbf{D}|H_2)p(H_2)} \quad k = 1, 2. \quad (7.1)$$

To determine the posterior odds of hypothesis  $H_1$ , the odds ratio scale results in

$$\frac{p(H_1|\mathbf{D})}{1 - p(H_1|\mathbf{D})} = \frac{p(H_1|\mathbf{D})}{p(H_2|\mathbf{D})}. \quad (7.2)$$

Determining the odds ratio in (7.2) in terms of Bayes' theorem given in (7.1) leads to the transformation of prior to posterior taking on a simple form:

$$\frac{p(H_1|\mathbf{D})}{p(H_2|\mathbf{D})} = \frac{p(\mathbf{D}|H_1)}{p(\mathbf{D}|H_2)} \frac{p(H_1)}{p(H_2)}. \quad (7.3)$$

This implies that the transformation for prior odds to posterior odds is multiplication by

$$B_{12} = \frac{p(\mathbf{D}|H_1)}{p(\mathbf{D}|H_2)} \quad (7.4)$$

or the Bayes factor. This means that the posterior odds of the hypothesis  $H_1$  is equal to the product of the Bayes factor and the prior odds of the hypothesis  $H_1$ . The Bayes factor therefore represents the ratio of the posterior odds of the hypothesis  $H_1$  to its prior odds, irrespective of the value of the prior odds. If the prior probabilities of the two hypotheses under consideration may be assumed equal, the Bayes factor reduces to the posterior odds of  $H_1$ . This may, however, not always be the case.

When the two hypotheses under consideration contain no free parameters, the Bayes factor would simply be the likelihood ratio. However, if there are unknown parameters under either or both hypotheses, the Bayes factor is still calculated as in (7.4), but the probability densities  $p(\mathbf{D}|H_k)$  ( $k = 1, 2$ ) then requires integration over the parameter space so that

$$p(\mathbf{D}|H_k) = \int p(\mathbf{D}|\theta_k, H_k) \pi(\theta_k|H_k) d\theta_k \quad (7.5)$$

where under hypothesis  $H_k$  the parameter is  $\theta_k$ , with prior density  $\pi(\theta_k|H_k)$  and likelihood function  $p(\mathbf{D}|\theta_k, H_k)$ . Note that  $\theta_k$  may also be a vector and that if this is the case  $p(\mathbf{D}|H_k)$  is the marginal probability or marginal likelihood of the data. Note that none of the constants obtained in determining  $p(\mathbf{D}|H_k)$  may be discarded when calculating Bayes factors.

It often happens that the marginal likelihood as denoted by equation (7.5) is intractable and therefore requires determination by computational methods. This has proved to sometimes be extremely challenging. Chib (1995) has however found a way out in this respect, by determining the marginal likelihood from the output of the Gibbs sampler as discussed in the next section.

Using Bayes factors to evaluate the evidence in favour of a model has the benefit that it allows for all other information used in the modelling approach to be included in the assessment. To summarise the Bayes factor, denoted by  $B_{jk}$  and interpreted as the Bayes factor for  $H_j$  against  $H_k$ , is a measure of the evidence provided by the data in favour of the model under hypothesis  $H_j$  as opposed that the model under hypothesis  $H_k$ .



## 7.2 MARGINAL LIKELIHOOD FROM GIBBS OUTPUT

Chib (1995) introduced a simple approach to determine the marginal density of the sample data given parameter draws from the posterior distribution. From this result the Bayes factors required for model comparison may easily be obtained as part of the output of the Gibbs sampler. The approach, as described in detail below, is based on the fact that the marginal density of the data can be expressed as the prior times the likelihood function over the posterior density. Note that it now follows that model  $M_1$  is the model specified under hypothesis  $H_1$ , while model  $M_2$  is the model under hypothesis  $H_2$ .

The Bayes factor in favour of model  $M_1$  when compared to model  $M_2$  then is

$$BF_{12} = \frac{m(\mathbf{y} | M_1)}{m(\mathbf{y} | M_2)} \quad (7.6)$$

where  $m(\mathbf{y} | M_i) = \int p(\mathbf{y} | \boldsymbol{\theta}, M_i) \pi(\boldsymbol{\theta} | M_i) d\boldsymbol{\theta}$  is the marginal likelihood under model  $M_i$ . As mentioned before the problem is to find  $m(\mathbf{y} | M_i)$  if the integral is intractable as often happens.

The joint posterior of the parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$  is

$$\pi(\boldsymbol{\theta} | \mathbf{y}, M_i) = \frac{p(\mathbf{y} | \boldsymbol{\theta}, M_i) \pi(\boldsymbol{\theta} | M_i)}{m(\mathbf{y} | M_i)} \quad (7.7)$$

so that,

$$m(\mathbf{y} | M_i) = \frac{p(\mathbf{y} | \boldsymbol{\theta}^*, M_i) \pi(\boldsymbol{\theta}^* | M_i)}{\pi(\boldsymbol{\theta}^* | \mathbf{y}, M_i)} \quad (7.8)$$

which hold for any  $\boldsymbol{\theta}$ , and therefore also for the specified  $\boldsymbol{\theta}^*$ . The numerator of  $m(\mathbf{y} | M_i)$  in (7.8) can be evaluated directly, since the functions are known. The problem, however, is in the denominator of  $m(\mathbf{y} | M_i)$  in (7.8), because the value of the joint posterior at  $\boldsymbol{\theta}^*$  is unknown. The Gibbs sampler does, however, provide samples from the marginal posterior distributions. But for specified  $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_p^*)$

$$\pi(\boldsymbol{\theta}^* | \mathbf{y}, M_i) = \pi(\theta_1^* | \mathbf{y}, M_i) \pi(\theta_2^* | \theta_1^*, \mathbf{y}, M_i) \dots \pi(\theta_p^* | \theta_1^*, \dots, \theta_{p-1}^*, \mathbf{y}, M_i) \quad (7.9)$$

The last term of (7.9) is the full conditional of  $\theta_p^*$ , which is known and can be evaluated exactly. The first term is the marginal posterior of  $\theta_1$  at  $\theta_1^*$ , and can therefore be obtained from the original Gibbs sampling:

$$\pi\left(\theta_1^* \mid \mathbf{y}, M_i\right) = \frac{1}{K} \sum_{j=1}^K \pi\left(\theta_1^* \mid \theta_{(-1)}^{(j)}, \mathbf{y}, M_i\right) \quad (7.10)$$

where  $K$  is the number of simulation iteration used in Gibbs sampling. For each of the rest of the terms, additional Gibbs sampling is required. For example, to find  $\pi\left(\theta_2^* \mid \theta_1^*, \mathbf{y}, M_i\right)$ , draw successively from

$$\begin{aligned} & \pi\left(\theta_2 \mid \theta_1^*, \theta_3, \dots, \theta_p, \mathbf{y}, M_i\right), \\ & \pi\left(\theta_3 \mid \theta_1^*, \theta_2, \theta_4, \dots, \theta_p, \mathbf{y}, M_i\right), \\ & \dots\dots\dots \\ & \pi\left(\theta_p \mid \theta_1^*, \theta_2, \dots, \theta_{p-1}, \mathbf{y}, M_i\right). \end{aligned} \quad (7.11)$$

Then

$$\pi\left(\theta_2^* \mid \theta_1^*, \mathbf{y}, M_i\right) = \frac{1}{K} \sum_{j=1}^K \pi\left(\theta_2^* \mid \theta_1^*, \theta_3^{(j)}, \dots, \theta_p^{(j)}, \mathbf{y}, M_i\right). \quad (7.12)$$

Similarly for  $\pi\left(\theta_3^* \mid \theta_1^*, \theta_2^*, \mathbf{y}, M_i\right)$ , and so forth. For the best results,  $\theta^*$  should be a high density point, such as the mode or the mean.

Furthermore, the approach described here to obtain the marginal likelihood from the output of the Gibbs sampling algorithm is fully automatic and stable and requires no input beyond the draws from the Gibbs simulation. This means that no additional tuning functions are required when determining the marginal likelihood.

### 7.3 BAYES FACTORS IN APPLICATION

The above results are now applied to the general case as described for the Bayesian approach in Chapter 4 in order to find the marginal likelihood or marginal density of the data. From this, Bayes factors are calculated in order to compare the evidence provided by the data in

favour of one model when compared to another for each of the four models that were fitted to the Jersey data in Chapters 5 and 6. In general, to determine the marginal likelihood, let  $\mathbf{?} = \{\mathbf{M}(vqk \times u), \mathbf{F}(\text{diag}(u)), \mathbf{B}(uv \times m), \mathbf{R}(q \times q), \mathbf{S}(uv \times uv)\}$ , it therefore follows that

$$\begin{aligned} \pi(\mathbf{?}^* | \mathbf{Y}) &= \pi(\mathbf{M}^* | \mathbf{Y}) \cdot \pi(\mathbf{B}^* | \mathbf{M}^*, \mathbf{Y}) \cdot \pi(\mathbf{S}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{Y}) \cdot \\ &\quad \pi(\mathbf{R}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{S}^*, \mathbf{Y}) \cdot \pi(\mathbf{F}^{*-1} | \mathbf{M}^*, \mathbf{Y}) \end{aligned} \quad (7.13)$$

The five terms on the right hand side of (7.13) can be evaluated as follows: Determine a  $\mathbf{?}^*$  from the marginal posteriors, including for all elements of  $\mathbf{M}_{ij}^*$  ( $v \times u$ ) for all  $i, j$ . We start out with:

$$1) \quad \pi(\mathbf{M}^* | \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^K \pi(\mathbf{M}^* | \mathbf{B}^{(j)}, \mathbf{F}^{(j)}, \mathbf{R}^{(j)}, \mathbf{S}^{(j)}, \mathbf{Y})$$

This is the product of equation (4.17) in Chapter 4 for all  $i = 1, \dots, k$ , evaluated at  $\mathbf{M}^*$  for each set of the original number of  $K$  Gibbs simulations for the other parameters.

$$2) \quad \pi(\mathbf{B}^* | \mathbf{M}^*, \mathbf{Y}) \text{ requires a cycle of three distributions. First, draw } \mathbf{B} \text{ from } \pi(\mathbf{B} | \mathbf{M}^*, \mathbf{R}, \mathbf{S}, \mathbf{Y}), \text{ i.e. (4.18) in Chapter 4 with fixed } \mathbf{M}^*. \text{ The starting values for } \mathbf{R} \text{ and } \mathbf{S} \text{ remain as before. Second, draw } \mathbf{R} \text{ from } \pi(\mathbf{R}^{-1} | \mathbf{M}^*, \mathbf{B}, \mathbf{S}, \mathbf{Y}) \text{ in (4.21). Third, draw } \mathbf{S} \text{ from } \pi(\mathbf{S}^{-1} | \mathbf{M}^*, \mathbf{B}, \mathbf{R}, \mathbf{Y}) \text{ in (4.20). The output for } \mathbf{R} \text{ and } \mathbf{S} \text{ is then used to obtain } \pi(\mathbf{B}^* | \mathbf{M}^*, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^K \pi(\mathbf{B}^* | \mathbf{M}^*, \mathbf{R}^{(j)}, \mathbf{S}^{(j)}, \mathbf{Y}).$$

$$3) \quad \pi(\mathbf{S}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{Y}) \text{ requires a cycle with two distributions, the first } \pi(\mathbf{S}^{-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{R}, \mathbf{Y}) \text{ in (4.20) and thereafter } \pi(\mathbf{R}^{-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{S}, \mathbf{Y}) \text{ in (4.21) from which } \pi(\mathbf{S}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{Y}) = \frac{1}{K} \sum_{j=1}^K \pi(\mathbf{S}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{R}^{(j)}, \mathbf{Y}) \text{ is obtained.}$$

$$4) \quad \pi(\mathbf{R}^{*-1} | \mathbf{M}^*, \mathbf{B}^*, \mathbf{S}^*, \mathbf{Y}) \text{ is the full conditional distribution and is evaluated exactly from (4.21).}$$

5)  $\pi\left(\mathbf{F}^{*-1} \mid \mathbf{M}^*, \mathbf{Y}\right)$  is also evaluated exactly from (4.19) since  $\mathbf{F}$  only depends on  $\mathbf{M}$ .

The product of these five terms then provides the denominator of the marginal likelihood in

(7.8). The joint distribution of all quantities is  $p\left(\mathbf{Y} \mid ?^*\right) \pi\left(?^*\right)$ , which is the product:

$$p\left(\mathbf{Y} \mid \mathbf{M}^*, \mathbf{F}^*\right) \pi\left(\mathbf{M}^* \mid \mathbf{B}^*, \mathbf{S}^*, \mathbf{R}^*\right) \pi\left(\mathbf{F}^{*-1}\right) \pi\left(\mathbf{B}^*\right) \pi\left(\mathbf{S}^{*-1}\right) \pi\left(\mathbf{R}^{*-1}\right).$$

and forms the numerator of the marginal likelihood:

$$m(\mathbf{Y}) = \frac{p\left(\mathbf{Y} \mid ?^*\right) \pi\left(?^*\right)}{\pi\left(?^* \mid \mathbf{Y}\right)}. \quad (7.14)$$

The same procedure must now be followed for another model. Once the marginal likelihood for two competing models have been obtained as explained above, the evidence in favour of one model may be determined through calculation of the Bayes factor in favour of that model using:

$$BF_{12} = \frac{m(\mathbf{Y} \mid M_1)}{m(\mathbf{Y} \mid M_2)}. \quad (7.15)$$

At this point it is worth noting that it is sometimes computationally more convenient when calculating the marginal likelihood to do so on natural logarithm scale, so that

$$\ln(m(\mathbf{Y})) = \ln\left(p\left(\mathbf{Y} \mid ?^*\right)\right) + \ln\left(\pi\left(?^*\right)\right) - \ln\left(\pi\left(?^* \mid \mathbf{Y}\right)\right). \quad (7.16)$$

The natural logarithms of the marginal likelihoods are then used to obtain the natural logarithm of the Bayes factor:

$$\ln(BF_{12}) = \ln(m(\mathbf{Y} \mid M_1)) - \ln(m(\mathbf{Y} \mid M_2)). \quad (7.17)$$

Good (1985) referred to the *ln* Bayes factor as the “weight of the evidence”. We therefore consider the *ln* Bayes factor as the weight of the evidence in favour of model  $M_1$ .

## 7.4 MODEL COMPARISON USING BAYES FACTORS

For computational convenience all four the models fitted to the Jersey data in Chapters 5 and 6 required determination of the marginal likelihoods on the natural logarithm scale through the use the method suggested by Chib (1995) as discussed in sections 7.2 and 7.3 above. The

results for the natural logarithm of the marginal likelihood of each of the four models were as follows.

**Table 7.1: The natural logarithm of the marginal likelihoods for the four lactation models fitted to the Jersey data using the Chib procedure**

Model	$\ln(m(Y))$
Wood : $E(W_i) = t^b \exp(a + ct)$	$-1.338887914040548 \times 10^{11}$
General Exponential: $E(W_i) = \exp(a + bt + c \exp(wt))$	$-8.626831563671578 \times 10^{13}$
4-parameter Morant: $E(W_i) = \exp(a + bt + ct^2 + d/t)$	$-5.179540981122800 \times 10^{15}$
Adapted Wilmink : $E(W_i) = \exp(a + bt + c \exp(wt))$	$-1.282449628353239 \times 10^{11}$

If the Bayes factors of these models are now evaluated, starting with the weight of evidence in favour of the model with the largest  $\ln(m(Y))$ , the results are as follows:

**Table 7.2: The  $\ln$  Bayes factors measuring the weight of evidence in favour of the adapted Wilmink model compared to each of the other three lactation models fitted.**

<i>Adapted Wilmink model vs</i>	$\ln(BF)$
Wood model	$5.643828568730896 \times 10^9$
General exponential model	$8.614007067388045 \times 10^{13}$
4-parameter Morant model	$5.179412736159965 \times 10^{15}$

From the results of Table 7.2 is evident that when the adapted Wilmink model suggested by this study is compared to any of the other three models fitted to the Jersey data, it is the most appropriate model to use in modelling lactation for these cows.

**Table 7.3: The  $\ln$  Bayes factors measuring the weight of evidence in favour of the Wood model compared to the general exponential and 4-parameter Morant models fitted.**

<i>Wood model vs</i>	$\ln(BF)$
General exponential model	$8.613442684531172 \times 10^{13}$
4-parameter model Morant	$5.179407092331396 \times 10^{15}$

From the results of Table 7.3 in which the  $\ln$  Bayes factors comparing the Wood model to the general exponential and 4-parameter Morant models respectively are considered, the weight of evidence in favour of the Wood model is in both cases overwhelming.

When comparing the general exponential model to the 4-parameter Morant model using the natural logarithm of the Bayes factor, the value  $5.093272665486084 \times 10^{15}$  that represents the weight of evidence in favour of the general exponential model, indicates it to be preferred over the 4-parameter Morant model.

From the above results it follows that the adapted Wilmink model should be the model of preference when modelling the Jersey data, followed by the Wood and general exponential models, while the 4-parameter Morant model seem to be the worst choice of lactation model of the four to fit to the data. These results support the suspicion that arose through visual inspection of the seemingly good results that were obtained for the adapted Wilmink model when the expected lactation trait curves of the levels of a cofactor of interest after the effects of all other cofactors have been averaged out are considered, while the 4-parameter Morant model did not really seem to provide realistic expected curves for two of the three lactation traits.

It is interesting to note the two models (Wood and adapted Wilmink) containing three parameters per lactation trait in the matrix of regression coefficients, or  $\mathbf{M}_{ij}$ , perform better than the two models with four parameters per lactation trait in  $\mathbf{M}_{ij}$ . The four-parameter models could possibly be over-parameterisations of the lactation process.

# CHAPTER 8

## FURTHER PREDICTION BASED ON MODELS FITTED

### 8.1 INTRODUCTION

So far, with respect to prediction of lactation trait curves for individual animals, only prediction of the fifth lactation cycle for a cow contained in the Jersey data, i.e. where the records of the preceding four lactation cycles are known and were used in developing the lactation model, were considered. Two further aspects of importance with respect to prediction of lactation trait curves of individual animals will be considered in this chapter. The first is how the lactation model fitted to the Jersey data may be used to predict lactation trait curves for an individual animal that do not form part of the Jersey data used to develop the model, on condition that it may be assumed that such an individual animal is similar to those used in the model development. The second is how the availability of lactation records and lactation trait curves fitted to a group of animals in a preceding lactation cycle or cycles, aid the prediction of lactation trait curves for such a group of animals in coming lactation cycles if these animals did not form part of the Jersey data set used to develop the lactation model, but may be assumed similar to the animals in the Jersey data for which the lactation model was developed.

To perform this part of the study, lactation records for 10 Jersey cows that originate from the same larger data set as the so-called Jersey data were obtained. These 10 Jersey cows were all observed during each of the four lactation cycles in the calving years from 1995 to 1998, but were excluded from the originally constructed Jersey data because seven test day records occurred in at least one of the observed four lactation cycles. For inclusion in the originally constructed Jersey data the condition was set that an animal had to have at least eight test day records per lactation cycle. In the data on these 10 Jersey cows all of the cofactors, parity number, region, calving year and calving season, present in the original Jersey data were also included. The data on these 10 Jersey cows will from this point onwards be referred to as the Further Jersey data. The only difference in characteristics between the Jersey data set and the Further Jersey data is that in the latter the number of test day records per lactation cycle over

all four lactation cycles for all ten animals ranged from 7 to 9, with the majority having 8 observed test day records.

## 8.2 PREDICTION FOR FURTHER DATA: THE GENERAL CASE

Suppose that for a cow in the Further Jersey data the lactation trait curves of milk yield, fat content and protein content of the four lactation cycles from 1995 to 1998 have to be predicted before any test day records for these lactation cycles become available. On condition that this cow is similar to those used in the development of the lactation model considered, such a model developed for the original Jersey data may be used to do so.

It is reasonable to assume that for cows similar to those in the original Jersey data certain information will be readily available. One could expect to at least know most, if not all the information pertaining to the cofactors included in a lactation model through the use of the covariate matrix,  $\mathbf{Z}_i$ , discussed in section 5.3 of Chapter 5. These cofactors are the parity of the cow, region in which it occurred and season it is expected to calve in. The cofactor year of calving is related to the lactation cycle to be predicted and is therefore anyway considered known. If any of these cofactors are not known for the animal of which the lactation trait curves are to be predicted, the effect of such cofactor(s) may be averaged out as explained in section 5.5 in Chapter 5, by taking as the values of the covariates of the cofactor the value one divide by the number of levels in the cofactor to be averaged out.

Before the test day records of any of the lactation cycles of an individual animal  $i$  for which the lactation traits are to be predicted are known, the predictive density follows from the fact that

$$\mathbf{Y}_i | \mathbf{M}_i, \mathbf{F} \sim N(\mathbf{X}_i \mathbf{M}_i, \mathbf{F} \otimes \mathbf{I}_{n_i}), \quad (8.1)$$

where  $\mathbf{M}_i (uv \times q)$  follows from the product  $\mathbf{BZ}_i$ . The matrix  $\mathbf{B} (uv \times m)$  refers to the regression matrix obtained for the original Jersey data during modelling. The covariate matrix  $\mathbf{Z}_i$  is constructed from the known cofactor information of animal  $i$  for which the lactation trait curves are to be predicted or by averaging out the effects of unknown cofactors as briefly explained above.

Next it is assumed that the test day records of the first of the  $q$  lactation cycle under consideration become available, but that the test day data for remaining  $(q - 1)$  lactation cycles are still unknown. Thereafter it is assumed that the test day records of the first two



lactation cycles are known, but that the remaining  $(q - 2)$  lactation cycles have to be predicted, and so forth until all  $q$  lactation cycles under consideration have been observed. In section 4.1 of Chapter 4 the procedure for the prediction of one or more lactation cycle of an animal  $i$  based on the results from one or more the preceding lactation cycles is given for the general case when the lactation model have been developed specifically for these animals. Lactation trait curves for all animals that did not form part of the data for which the model was developed, but that may be assumed similar to the animals for which the model was developed may then also be predicted using this procedure.

Assuming that  $h$  lactation cycles were observed, it is possible to predict the results of  $(q - h)$  lactation cycles for an animal similar to those for which the model was developed. This means that if the test day records of the first lactation cycle of an animal are available, the method allows for the prediction of the remaining  $(q - 1)$  lactation cycles. Using Gibbs simulation and

$$\mathbf{m}_i | \mathbf{y}_i, \mathbf{B}, \mathbf{F}, \mathbf{S}, \mathbf{R} \sim N \left\{ \left[ \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \right) \mathbf{y}_i + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \mathbf{B} \right) \mathbf{z}_i \right] \right], \right. \\ \left. \left[ \left[ \left( \mathbf{F}^{-1} \otimes \mathbf{X}_i' \mathbf{X}_i \right) + \left( \mathbf{R}^{-1} \otimes \mathbf{S}^{-1} \right) \right]^{-1} \right] \right\}, \quad i = 1, \dots, k \quad (8.2)$$

the 10 000 set of parameters of the model for the first lactation cycle,  $\mathbf{m}_i^{(1)*}$ , is drawn where  $\mathbf{B}$ ,  $\mathbf{F}$ ,  $\mathbf{S}$  and  $\mathbf{R}$  have the values obtained from Gibbs sampling for the original Jersey data and  $\mathbf{z}_i$  is constructed for the observed lactation cycle. Next, based on the values of  $\mathbf{m}_i^{(1)*}$ ,  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\mathbf{R}$  above, 10 000 sets of  $\mathbf{m}_{if}^{(2)*}$  are drawn from

$$\mathbf{m}_{if}^{(2)} | \mathbf{m}_i^{(1)} = \mathbf{m}_i^{(1)*}, \mathbf{B}, \mathbf{S}, \mathbf{R} \sim N \left( \left( \mathbf{I}_{q-h} \otimes \mathbf{B} \right) \mathbf{z}_i^{(2)} + \left( \mathbf{R}_{21}^{(h)} \mathbf{R}_{11}^{(h)-1} \otimes \mathbf{I}_{uv} \right) \mathbf{m}_i^{(1)*} - \left( \mathbf{I}_h \otimes \mathbf{B} \right) \mathbf{z}_i^{(1)}, \right. \\ \left. \left( \mathbf{R}_{22}^{(h)} \otimes \mathbf{S} \right) - \left( \mathbf{R}_{21}^{(h)} \mathbf{R}_{11}^{(h)-1} \mathbf{R}_{12}^{(h)} \otimes \mathbf{S} \right) \right) \quad (8.3)$$

using Gibbs sampling. Finally the predictive density, which again follows from (8.1), for the  $(q - h)$  future lactation cycles becomes:

$$\mathbf{y}_{if}^{(2)} | \mathbf{m}_{if}^{(2)} = \mathbf{m}_{if}^{(2)*}, \mathbf{F} \sim N \left( \left( \mathbf{I}_u \otimes \mathbf{X}_{if}^{(2)} \right) \mathbf{m}_{if}^{(2)*}, \mathbf{F} \otimes \mathbf{I}_{n_{ij}(q-h)} \right). \quad (8.4)$$

The mean of the 10 000 expected values of this distribution provides the predicted lactation trait curves for the  $(q - h)$  lactation cycles to be predicted.

No efficient method exists to measure the performance of the model prediction of a future lactation cycle of an animal  $i$  if the data of that lactation cycle of animal  $i$  are not available at

some point. However, if the data are or become available the efficiency of prediction can be measured. This is done by obtaining the sum of squared errors (SSE) for animal  $i$ , which is the sum of the squared differences between each observed data point and the value of the predicted lactation trait curve at the same point in time as the observed data point over all observed data points. If this is done for all previously predicted lactation cycles as data become available, it is possible to investigate whether the inclusion of data improved the efficiency of prediction of future lactation cycles.

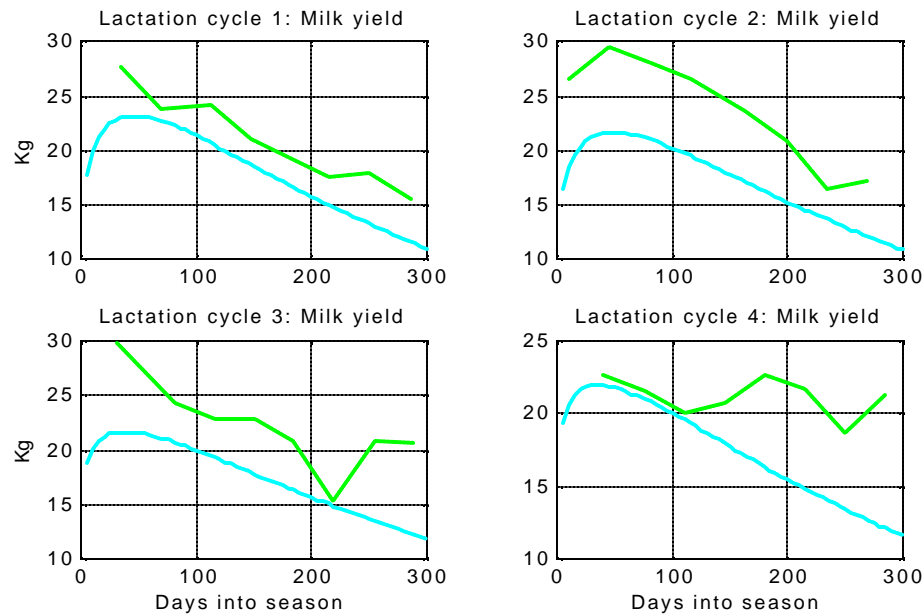
To explain, suppose four lactation cycles are predicted with respect to milk yield. This prediction process may be broken down into four phases. In phase 1, initially no data will be present and all four cycles will as a consequence have to be predicted. Then the milk yield records in the first of the four lactation cycles are observed and a milk yield curve fitted to it. In phase 2 only the remaining three unknown lactation cycles have to be predicted, now however with knowledge of the fitted milk yield curve on the first included in the prediction. Once the milk yield records of the second cycle are also observed and a milk yield curve fitted to it, then in phase 3 the prediction of the last two cycles will be done by incorporating the knowledge of the curves fitted to the first and second lactation cycles in the prediction. Once the third lactation cycle has also been observed and a milk yield curve fitted to it, then in phase 4 the fourth lactation cycle will be predicted through the inclusion of the knowledge of the fitted milk yield curves from all three preceding lactation cycles. Once the data on all four the lactation cycles have been observed, the improvement in prediction of the unknown milk yield curves in every phase may be evaluated by calculating the SSE for each of the predicted curves in every phase. By comparing the SSE of the same lactation cycle over all phases for which the cycle was predicted, the efficiency of the inclusion of data in the improvement of the prediction may be evaluated. An improvement is simply based on a reduction in to the sum of the squared errors for that lactation cycle from one phase to the next.

In the application that follows, prediction as explained through the four phases above is applied to the 10 animals in the Further Jersey data. Based on the result, the improvement in prediction is assessed by means of the SSE values. This is done for the group of 10 animals by obtaining the sum of the SSE values for the individual animals for each lactation cycle in each prediction phase, as well as by considering for how many animals an improvement through the addition of data occurred for each lactation cycle predicted in each phase.

Finally, once the test day records on all four lactation cycles have been obtained, the efficiency of the lactation model in finding the expected lactation trait curves for animals similar to, but not included in the development of the model may be compared for the four different lactation models fitted using this Bayesian method. To do this the SSE values for the expected lactation trait curve for all four lactation cycles of the ten animals are obtained for each of the four lactation models fitted in this study.

### 8.3 PREDICTION AND THE WOOD MODEL

Using the results of the Wood model fitted to the Jersey data by means of the Bayesian method proposed in this study, the four lactation cycles of each of the 10 animals in the Further Jersey data were predicted through the 4phase prediction process explained above. To illustrate, the results for the prediction of the lactation or milk yield curves of each phase of the 4phase prediction process for the four lactation cycles of animal 3 are given in Figures 8.1 to 8.4 below.

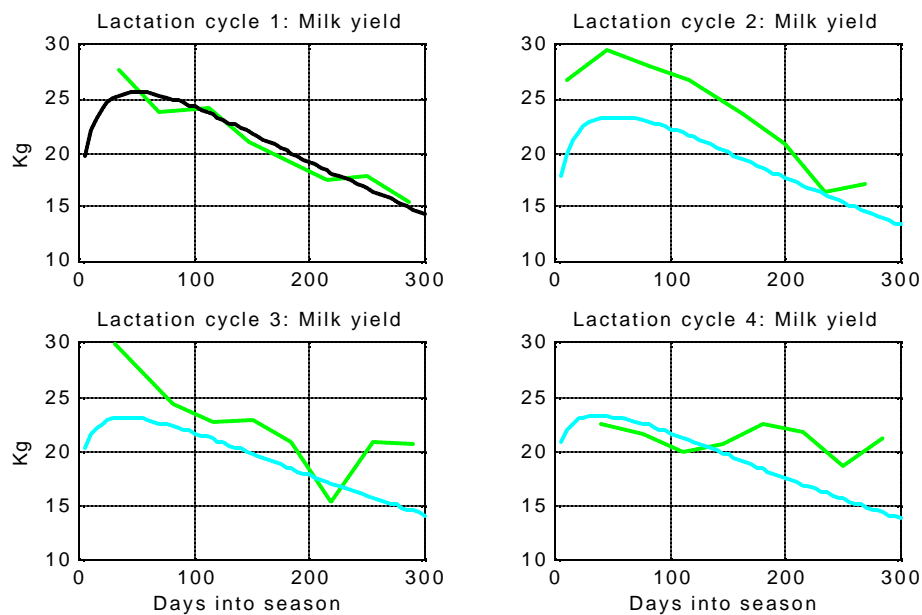


**Figure 8.1: Prediction phase 1. Predicting milk yield curves for all 4 lactation cycles of animal 3, where milk yield test day records are represented by — and predicted milk yield curves are represented by — .**

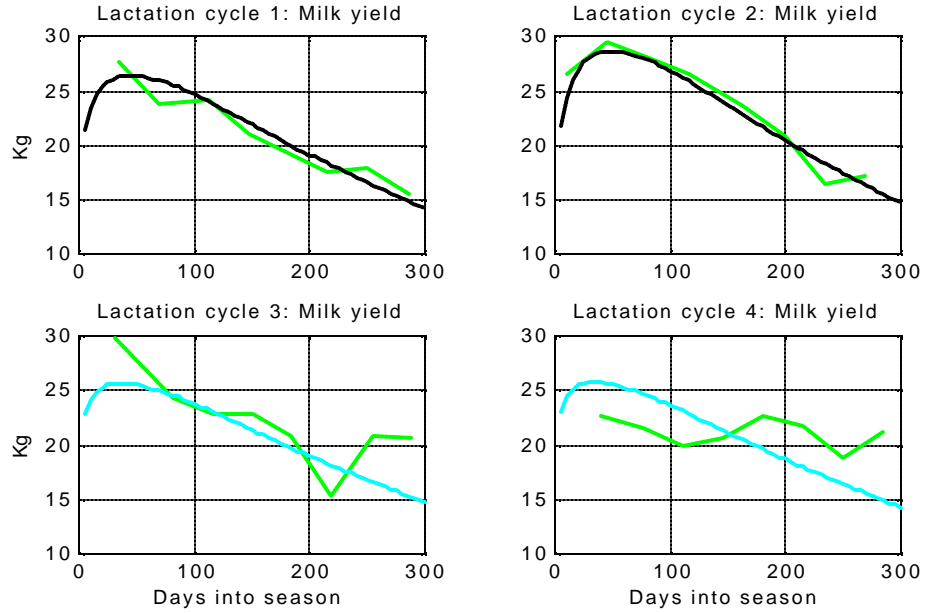
In Figure 8.1 where the predicted milk yield curves that result from phase 1 in the prediction process are represented by the turquoise lines and the data on the milk yield test day records of each lactation cycle that would supposedly only be known later are represented by green

lines (consecutive data points were connected by straight line segments). The predictions of the milk yield curves of animal 3 in this phase include no information on milk yield test day records of this animal. These predictions are based solely on the information on milk yield test day records from the Jersey data. The only information from animal 3 included in these prediction are the information supplied by the cofactors parity number, region, calving year and calving season. When calculating the SSE for each of lactation cycles 1, 2, 3 and 4 of animal 3 in this phase of prediction the results are 83,122; 333,382; 263,775 and 207,3042 respectively.

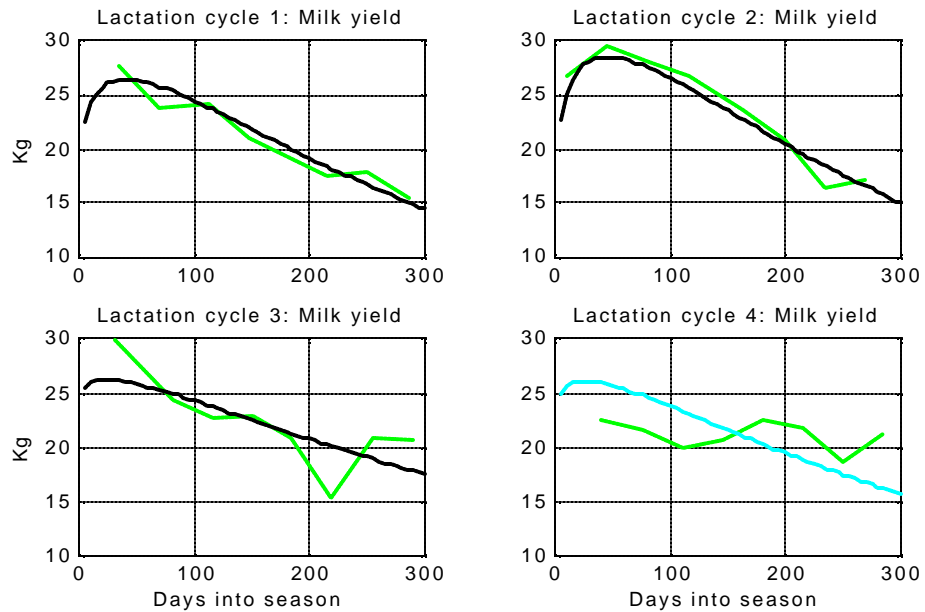
In Figure 8.2 the result of phase 2 of the prediction process is illustrated, where the black line now represents the fitted milk yield curve in lactation cycle 1 of animal 3 after the milk yield test day records of this cycle were observed. Lactation cycles 2, 3 and 4 are now again predicted, but this time incorporating the additional information that became available through the observed milk yield test day records of the first lactation cycle. When considering the SSE values for lactation cycles 2, 3 and 4 these values are 168,988; 133,580 and 103,804 respectively. When comparing these values to the SSE calculated for the corresponding lactation cycles in the first phase of prediction a reduction is seen for all lactation cycles.



**Figure 8.2: Prediction phase 2. Predicting milk yield curves for lactation cycles 2, 3 and 4 of animal 3 after the milk yield curve of lactation cycle 1 has been fitted. Milk yield test day records are represented by — , predicted milk yield curves are represented by — , and the fitted milk yield curve by — .**



**Figure 8.3: Prediction phase 3. Predicting milk yield curves for lactation cycles 3 and 4 of animal 3 after milk yield curves of lactation cycles 1 and 2 have been fitted. Milk yield test day records are represented by — (green), predicted milk yield curves are represented by — (cyan), and fitted milk yield curves by — (black).**



**Figure 8.4: Prediction phase 4. Predicting the milk yield curve for lactation cycle 4 of animal 3 after milk yield curves of lactation cycles 1, 2 and 3 have been fitted. Milk yield test day records are represented by — (green), the predicted milk yield curve is represented by — (cyan) and fitted milk yield curves by — (black).**

Figure 8.3 now shows the result of phase 3 of the prediction process, where only the milk yield curves of lactation cycles 3 and 4 of animal 3 are predicted. The resulting predicted milk yield curves of lactation cycles 3 and 4 are now much closer to the data that would later be observed. The SSE values of lactation cycles 3 and 4 now are 77,472 and 99,516 respectively, which is again a reduction in SSE values for these two cycles when compared to their result from phase 2 of the prediction process.

Finally in Figure 8.4 illustrating phase 4 of the prediction process, only the milk yield curve of the fourth lactation cycle has to be predicted this time based on the results of the Jersey data and the information from the preceding three lactation cycles of animal 3. When the SSE value for the prediction in the lactation cycle 4 is calculated the resulting value is 77,243 and is down from an SSE value of 99,516 for this lactation cycle in the phase 3 of the prediction process.

When all four the figures representing the results of prediction for the 4phase prediction process, together with the SSE values of each phase for animal 3 are considered, it is clear that as data from preceding lactation cycles become available the prediction ability of the method applied improves. The results obtained for the remaining 9 animals in the Further Jersey data are similar, although a reduction in SSE of a predicted lactation cycle is not always obtained from one prediction phase to the next as data on test day records of preceding lactation cycles become available. A summary of the results now follows.

To assess the improvement over all animals in the Further Jersey data, the sum of the SSE values for each lactation trait curve over all ten animals for each lactation cycle in each phase of the prediction process was obtained. The results are given in Tables 8.1, 8.3 and 8.5 below for the lactation traits milk yield, fat content and protein content respectively. In Tables 8.2, 8.4 and 8.6 the number of animals for which the predicted milk yield curve, predicted protein content curve and predicted fat content curve improved in fit as data on preceding lactation cycles became available, are provided. An improvement is based on a reduction in the sum of the squared errors for that lactation cycle when compared to actual data only available when the lactation cycle is eventually observed.

**Table 8.1: Sum of squared errors over all animals for the predicted milk yield curve fitted using the Wood model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	3977,761	2541,331	3592,989	1767,084
1 only (Phase 2)	◆	1276,762	1591,863	1059,939
1 and 2 (Phase 3)	◆	◆	1296,605	844,827
1, 2 and 3 (Phase 4)	◆	◆	◆	993,519

From Table 8.1 it is evident that for the whole of the Further Jersey data set the prediction ability of the method improved with respect to predicted milk yield curves during each of the first three phases of the prediction process. Only in the fourth phase did the inclusion of the test day records of all animals on the preceding three lactation cycles not improve the result with respect to predicting the milk yield curves.

**Table 8.2: The number of animals for which, according to the SSE, the predicted milk yield curve from the Wood model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	7	7	8
1 and 2 (Phase 3)	◆	8	7
1, 2 and 3 (Phase 4)	◆	◆	6

The results in Table 8.2 show that in every phase of the prediction process, except the final phase, there was an improvement in the prediction of milk yield curves for all lactation cycle in at least seven of the ten animals in the Further Jersey data. In the final prediction phase the inclusion of milk yield test day data on the third lactation cycle only improved the prediction of the milk yield curve of the fourth lactation cycle in six animals. This is decline in the number of animals for which the SSE reduced in the fourth phase of prediction is probably why the corresponding SSE in Table 8.1 was not smaller than the SSE for lactation cycle 4 in the phase 3 of the prediction process when sum of the SSE's over all animals are obtained for each cycle during each prediction phase.

**Table 8.3: Sum of squared errors over all animals for the predicted fat content curve fitted using the Wood model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	7,8121	6,2111	7,6299	12,8373
1 only (Phase 2)	◆	3,9946	4,7064	10,8117
1 and 2 (Phase 3)	◆	◆	3,6902	9,8881
1, 2 and 3 (Phase 4)	◆	◆	◆	9,7885

Table 8.3 shows that for the whole of the Further Jersey data set the prediction ability of the method improved in all phases of the prediction process with respect to predicted fat content. This is surprising when considering the result of Table 8.4, where in two lactation cycles the prediction of fat content curve improved in less than 6 animals. However, in all cases where no improvement resulted the difference between the SSE's for the same lactation cycle from one phase to the next is small compared to difference in cases where improvements did occur.

**Table 8.4: The number of animals for which, according to the SSE, the predicted fat content curve from the Wood model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	5	6	6
1 and 2 (Phase 3)	◆	8	8
1, 2 and 3 (Phase 4)	◆	◆	3

From the overall SSE results provided in Table 8.5 it was only during the third phase of prediction for protein content that no improvement occurred. When considering the number of animals in each lactation cycle during each phase of prediction for which the SSE values reduced as provided in Table 8.6, it follows that phase 3 of the prediction process also improved prediction for protein content in the smallest number of animals (4 only).



**Table 8.5: Sum of squared errors over all animals for the predicted protein content curve fitted using the Wood model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	55,1751	42,5692	28,7319	35,2074
1 only (Phase 2)	◆	39,9202	22,0870	34,2604
1 and 2 (Phase 3)	◆	◆	24,7528	37,4509
1, 2 and 3 (Phase 4)	◆	◆	◆	35,3853

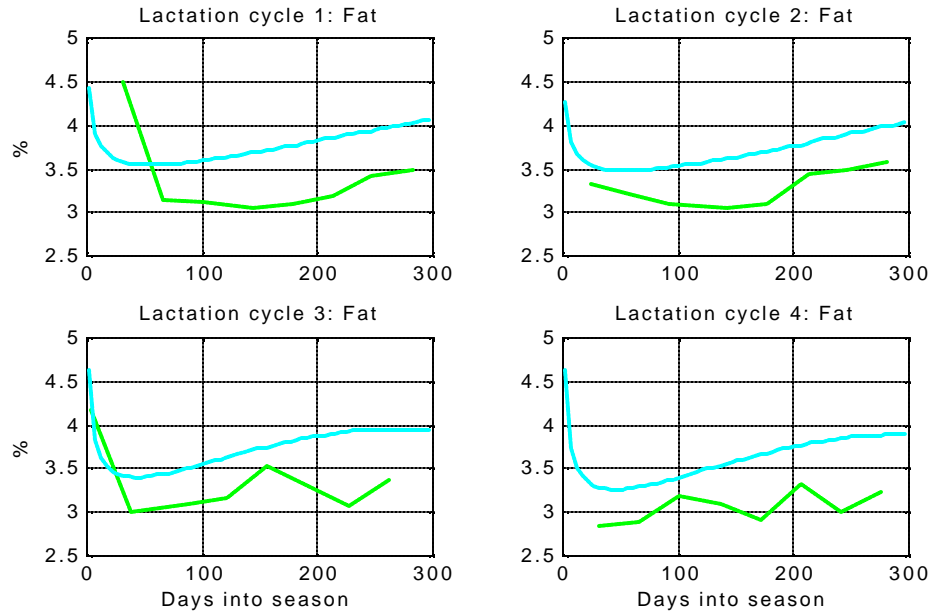
**Table 8.6: The number of animals for which, according to the SSE, the predicted protein content curve from the Wood model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	6	9	6
1 and 2 (Phase 3)	◆	4	4
1, 2 and 3 (Phase 4)	◆	◆	6

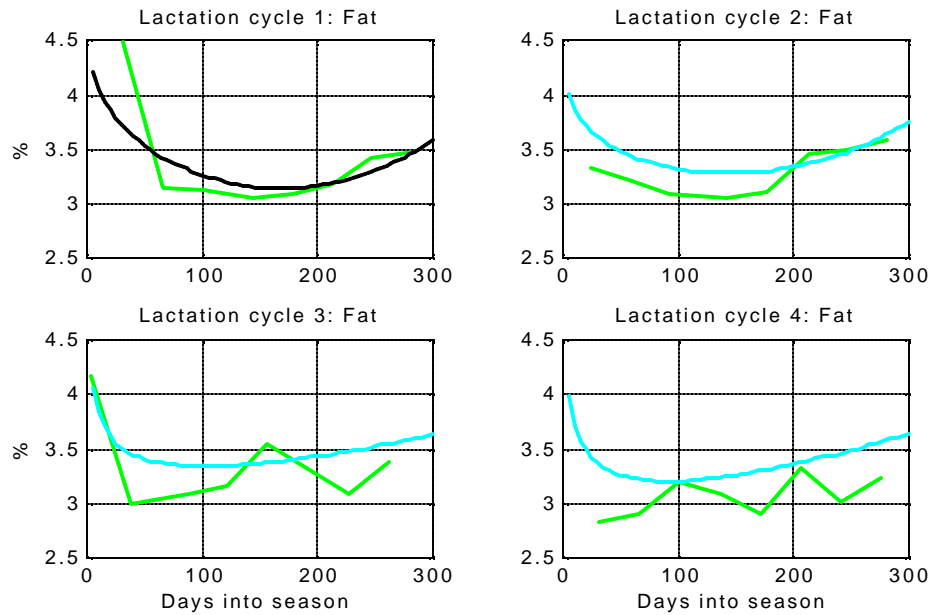
## 8.4 PREDICTION AND THE GENERAL EXPONENTIAL MODEL

For the general exponential model fitted to the Jersey data by means of the Bayesian method, the four lactation cycles for each of the 10 animals in the Further Jersey data were also predicted through the 4-phase process explained earlier. This time the results of the prediction of the fat content curve for each phase in the 4phase prediction process for the four lactation cycles of animal 10 are given in Figures 8.5 to 8.8.

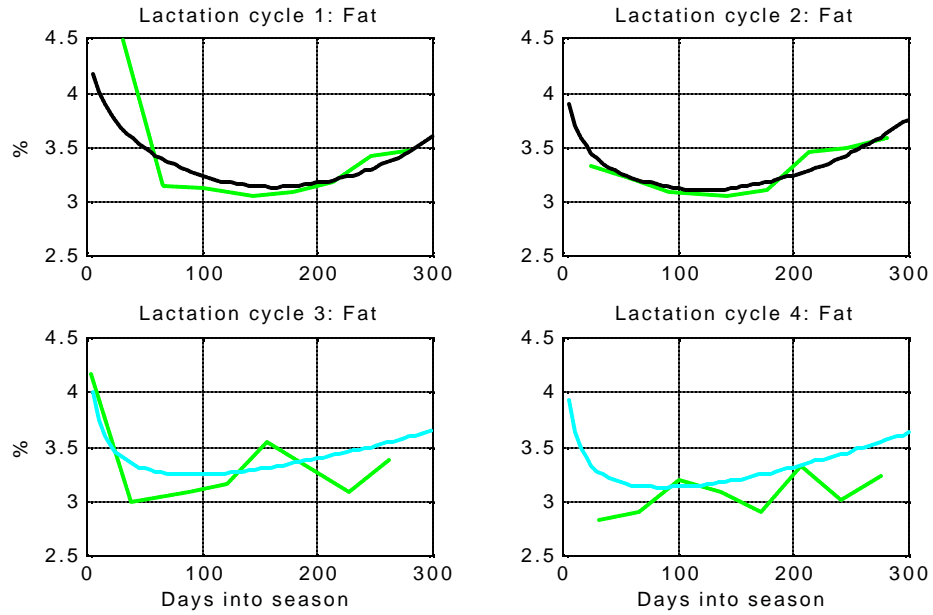
Through visual inspection of the graphs in Figures 8.5 to 8.8 it seems as if in each phase of the prediction process the predicted fat content curve of the lactation cycles moved closer to the actual data that was later observed. When, for animal 10, the SSE values of the lactation cycles in a prediction phase is compared to that of the next, these values confirm this results. For example, the SSE values of lactation cycle 4 in prediction phase 1 through to 4 are 2,502; 0,824; 0,649 and 0,426 respectively.



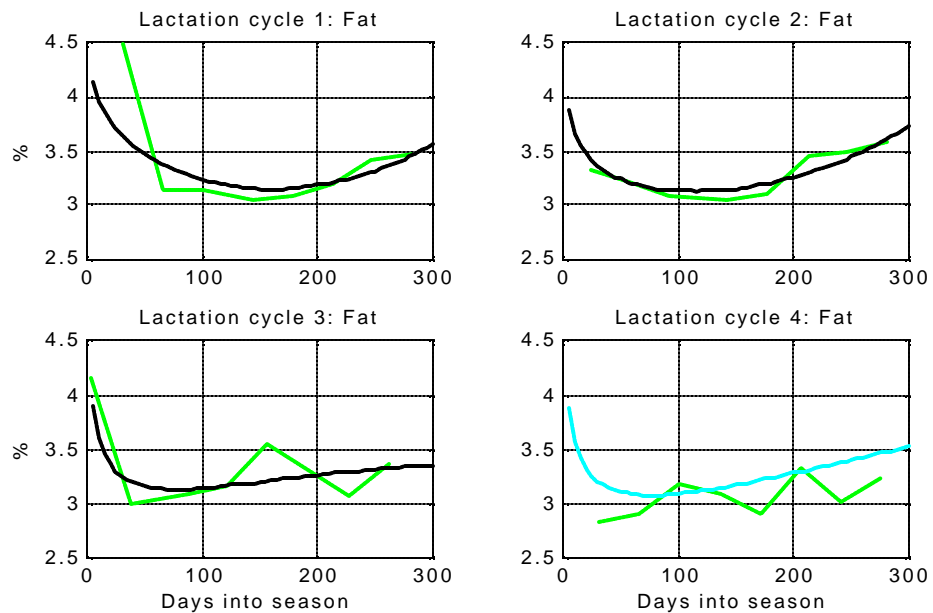
**Figure 8.5: Prediction phase 1. Predicting fat content curves for all 4 lactation cycles of animal 10, where fat content test day records are represented by — and predicted milk yield curves are represented by —.**



**Figure 8.6: Prediction phase 2. Predicting fat content curves for lactation cycles 2, 3 and 4 of animal 10 after the fat content curve of lactation cycle 1 has been fitted. Fat content test day records are represented by —, predicted milk yield curves are represented by — and the fitted fat content curve by —.**



**Figure 8.7: Prediction phase 3. Predicting fat content curves for lactation cycles 3 and 4 of animal 10 after fat content curves of lactation cycles 1 and 2 have been fitted. Fat content test day records are represented by — , predicted fat content curves are represented by — , and fitted fat content curves by — .**



**Figure 8.8: Prediction phase 4. Predicting the fat content curve for lactation cycle 4 of animal 10 after fat content curves of lactation cycles 1, 2 and 3 have been fitted. Fat content test day records are represented by — , predicted fat content curves are represented by — and fitted fat content curves by — .**

From the results of the sum of the SSE values of the predicted milk yield curves over all 10 animals in the data provided Table 8.7 it follow that for the whole of the Further Jersey data set the prediction ability of the method improved with respect to predicted milk yield curves during all of the phases of the prediction process. When considering the number of individual animals in the Further Jersey data for which this is true, Table 8.8 shows that an improvement in the prediction of milk yield as data from preceding cycles are obtained occur for at least 5 animals in a lactation cycle in any phase of prediction.

**Table 8.7: Sum of squared errors over all animals for the predicted milk yield curve fitted using the general exponential model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	4145,351	2601,262	3570,845	1774,791
1 only (Phase 2)	◆	1673,640	1744,586	1554,898
1 and 2 (Phase 3)	◆	◆	1210,771	1307,678
1, 2 and 3 (Phase 4)	◆	◆	◆	1273,962

**Table 8.8: The number of animals for which, according to the SSE, the predicted milk yield curve from the general exponential model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	8	5	7
1 and 2 (Phase 3)	◆	7	8
1, 2 and 3 (Phase 4)	◆	◆	5

In Table 8.9 the results of the sum of squared errors over all animals in the case of the predicted fat content curves indicates an improvement in prediction in all but the last phase of prediction, where there is only a slight increase in the SSE value for lactation cycle 4 when compare to the SSE value of this lactation cycle in phase 3 of prediction. It also follows from Table 8.10 that for the fourth phase of prediction an improvement in the prediction of the fat content curve for lactation cycle 4 only occurred in 3 animals.

**Table 8.9: Sum of squared errors over all animals for the predicted fat content curve fitted using the general exponential model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	7,3844	6,3534	7,3795	12,3854
1 only (Phase 2)	◆	4,1259	4,5116	10,2579
1 and 2 (Phase 3)	◆	◆	3,6451	9,2508
1, 2 and 3 (Phase 4)	◆	◆	◆	9,4466

**Table 8.10: The number of animals for which, according to the SSE, the predicted fat content curve from the general exponential model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	5	6	6
1 and 2 (Phase 3)	◆	8	7
1, 2 and 3 (Phase 4)	◆	◆	3

To consider the improvement in the prediction ability with respect to the predicted protein content curves over all animals in the data set, the sum of the SSE's over all animals in the data for this lactation trait are given in Table 8.11. The only SSE values of that indicate improvements in the prediction of the protein content curves are those of lactation cycles 2 and 3 in phase 2, and that of lactation cycle 4 in phase 4. Table 8.12 provides the numbers of animals in each lactation cycle of each phase for which such an improvement in predicted protein content curve occurred.

**Table 8.11: Sum of squared errors over all animals for the predicted protein content curve fitted using the general exponential model before and as data on preceding lactation cycles became available.**

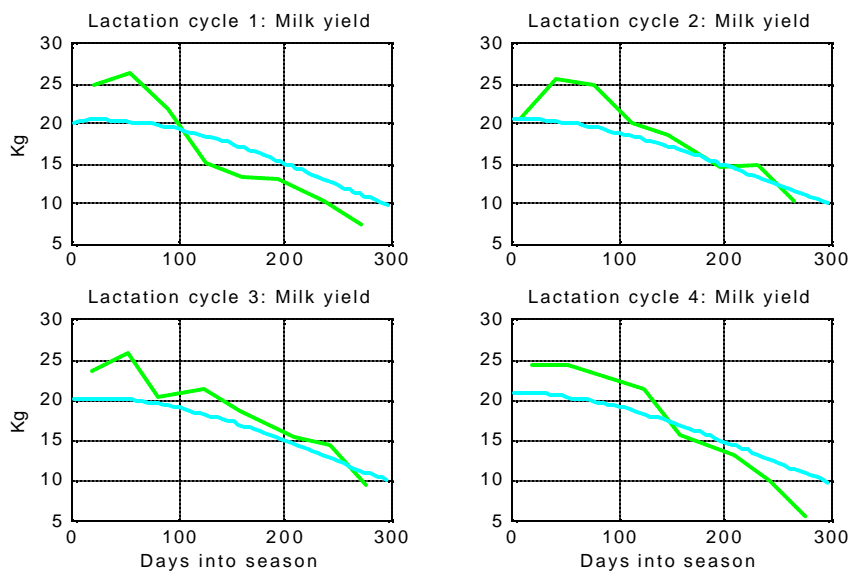
Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	54,7181	42,8749	27,4354	34,1457
1 only (Phase 2)	◆	41,3787	21,8013	35,2503
1 and 2 (Phase 3)	◆	◆	23,9290	38,6943
1, 2 and 3 (Phase 4)	◆	◆	◆	35,3198

**Table 8.12:** The number of animals for which, according to the SSE, the predicted protein content curve from the general exponential model improved in fit as data on previous lactation cycle became available.

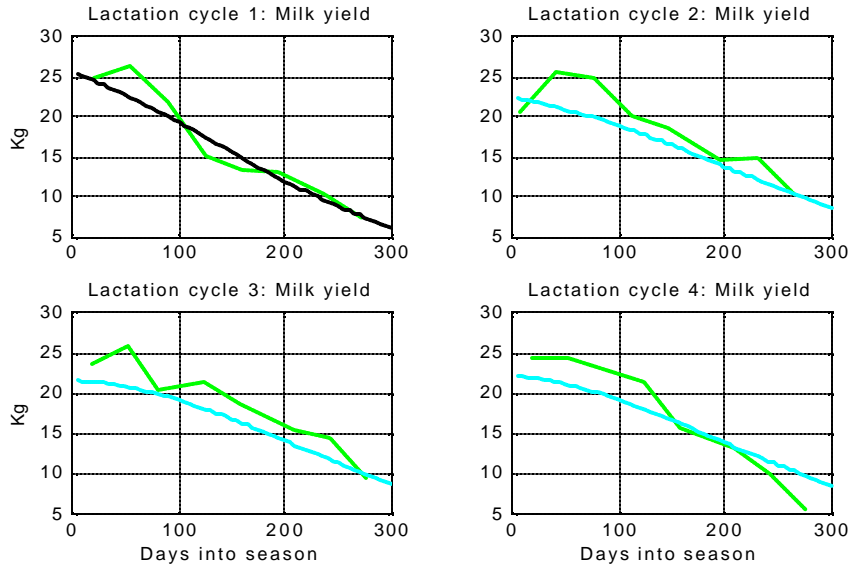
Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	5	7	5
1 and 2 (Phase 3)	♦	3	4
1, 2 and 3 (Phase 4)	♦	♦	8

## 8.5 PREDICTION AND THE 4-PARAMETER MORANT MODEL

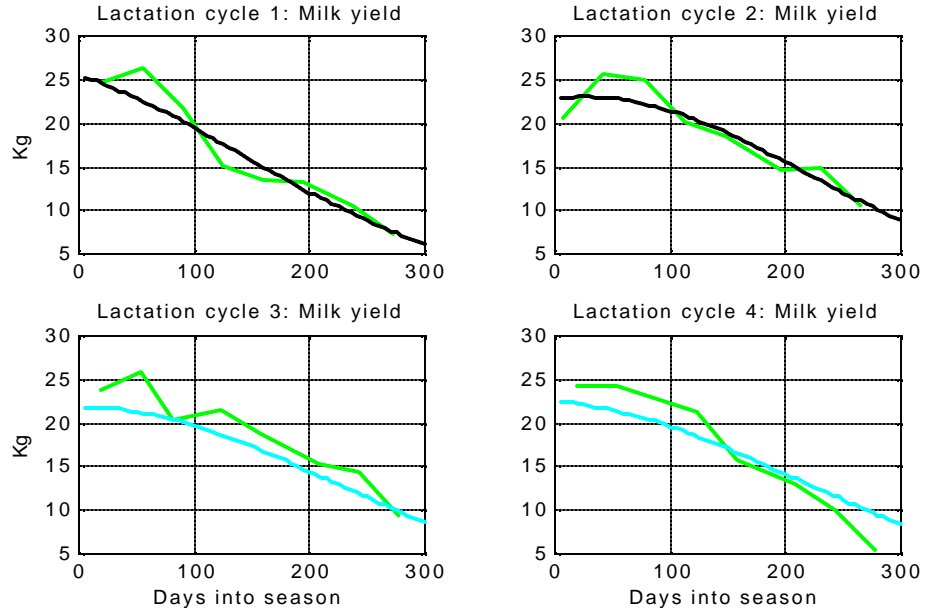
Figures 8.9 to 8.12 provide the results of the prediction of the milk yield curve for each of the 4 phases of the prediction process for the 4 lactation cycles of animal 5 in the Further Jersey data when the 4-parameter Morant model fitted to the original Jersey data by means of the Bayesian method is used in prediction. When considering these graphs, the improvement is not very clear from a visual perspective, but when studying the SSE values for the predicted milk yield curves of each lactation cycle in the four phases of prediction, the SSE values of each of the lactation cycle are reduced in every consecutive phase of prediction and therefore indicted an improvement in the prediction ability of the predicted milk yield curves for the lactation cycles as test day records in preceding lactation cycles become available.



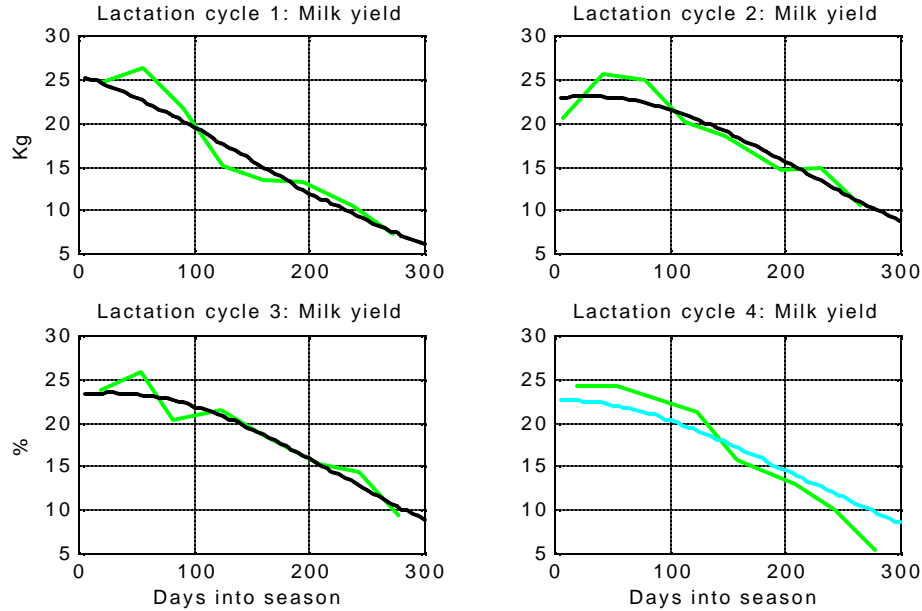
**Figure 8.9:** Prediction phase 1. Predicting milk yield curves for all 4 lactation cycles of animal 5, where milk yield test day records are represented by — and predicted milk yield curves are represented by — .



**Figure 8.10: Prediction phase 2. Predicting milk yield curves for lactation cycles 2, 3 and 4 of animal 5 after the milk yield curve of lactation cycle 1 has been fitted. Milk yield test day records are represented by — , predicted milk yield curves are represented by — and the fitted milk yield curve by — .**



**Figure 8.11: Prediction phase 3. Predicting milk yield curves for lactation cycles 3 and 4 of animal 5 after milk yield curves of lactation cycles 1 and 2 have been fitted. Milk yield test day records are represented by — , predicted milk yield curves are represented by — and fitted milk yield curves by — .**



**Figure 8.12: Prediction phase 4. Predicting the milk yield curve for lactation cycle 4 of animal 5 after milk yield curves of lactation cycles 1, 2 and 3 have been fitted. Milk yield test day records are represented by — , the predicted milk yield curve is represented by — , fitted milk yield curves by — .**

From the sum of the SSE values over all 10 animals in the Further Jersey data for predicted milk yield curves given in Table 8.13 it follows that overall prediction of milk yield curves continued to improve for all lactation cycles from one phase to the next. From Table 8.14 it follows that when the predicted milk yield curves of individual animals are considered, this improvement occurs in at least 7 animals for the lactation cycles of any phase of prediction. For lactation cycles in the second phase of prediction the number of individual animals to which an improvement applies is exceptionally high.

**Table 8.13: Sum of squared errors over all animals for the predicted milk yield curve fitted using the 4-parameter Morant model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	4179,922	2701,934	3759,006	1793,776
1 only (Phase 2)	♦	1724,578	2611,169	1170,907
1 and 2 (Phase 3)	♦	♦	2078,987	969,9002
1, 2 and 3 (Phase 4)	♦	♦	♦	844,2139



**Table 8.14:** The number of animals for which, according to the SSE, the predicted milk yield curve from the 4-parameter Morant model improved in fit as data on previous lactation cycle became available.

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	10	9	10
1 and 2 (Phase 3)	♦	7	8
1, 2 and 3 (Phase 4)	♦	♦	7

A result similar to the above also follows from Table 8.15 with respect to the SSE values for the predicted fat content curves. In Table 8.16 it, however, follows that numbers of individual animals to which such an improvement in the prediction of fat content curve for lactation cycles of each prediction phase apply, are lower.

**Table 8.15:** Sum of squared errors over all animals for the predicted fat content curve fitted using the 4-parameter Morant model before and as data on preceding lactation cycles became available.

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	7,5603	6,5498	7,5923	12,6259
1 only (Phase 2)	♦	5,4710	6,3302	10,9788
1 and 2 (Phase 3)	♦	♦	5,2154	10,5165
1, 2 and 3 (Phase 4)	♦	♦	♦	9,7103

**Table 8.16:** The number of animals for which, according to the SSE, the predicted fat content curve from the 4-parameter Morant model improved in fit as data on previous lactation cycle became available.

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	5	8	6
1 and 2 (Phase 3)	♦	6	6
1, 2 and 3 (Phase 4)	♦	♦	6

When compared to how the inclusion of data from preceding lactation cycles benefited the prediction of curves for the previous two lactation traits using the 4-parameter Morant model, the model fared worst at predicting protein content curves using the 4-phase prediction process. When considering Table 8.17 and 8.18, this is especially true for the second phase of prediction using the 4-parameter Morant model. When considering the sum of the SSE values over all 10 animals in the Further Jersey data the prediction of the protein content curve did not improve for any of the lactation cycles in the second phase. For the individual animals there was only an improvement in the prediction for one animal during all the lactation cycles in the second phase of prediction. This poor performance of the 4-parameter Morant model should however have been expected when taking into account the less realistic shape produced by this model for the expected protein content curves of a level of the cofactor of interest when eliminating the effects of all other cofactors as discussed in section 6.3.3 of Chapter 6.

**Table 8.17: Sum of squared errors over all animals for the predicted protein content curve fitted using the 4-parameter Morant model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	63,4016	40,2124	28,1340	35,4038
1 only (Phase 2)	◆	4030,256	7211,573	277,4784
1 and 2 (Phase 3)	◆	◆	3349,729	504,6070
1, 2 and 3 (Phase 4)	◆	◆	◆	301,1018

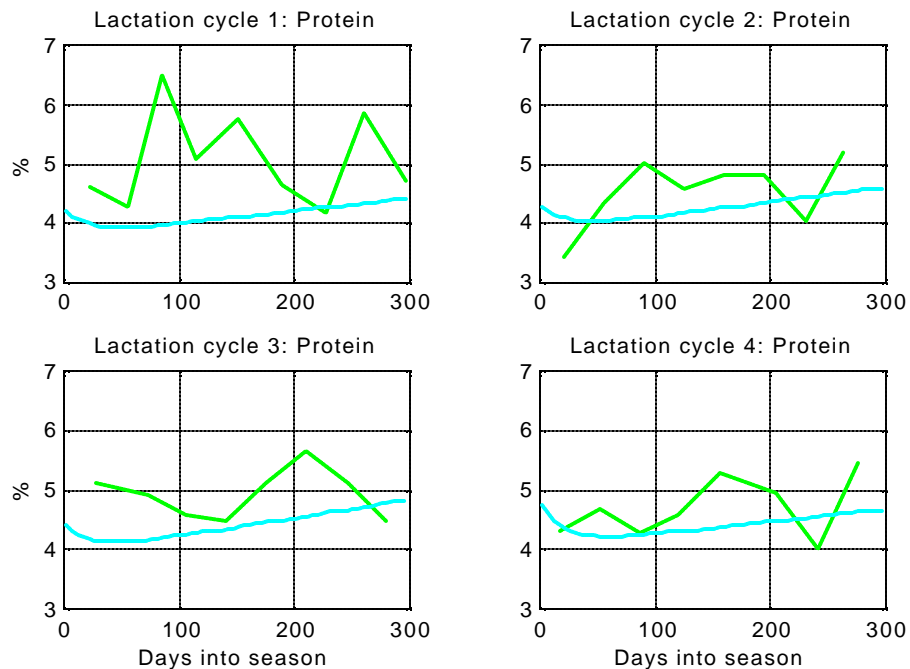
**Table 8.18: The number of animals for which, according to the SSE, the predicted protein content curve from the 4-parameter Morant model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	0	0	1
1 and 2 (Phase 3)	◆	6	7
1, 2 and 3 (Phase 4)	◆	◆	3

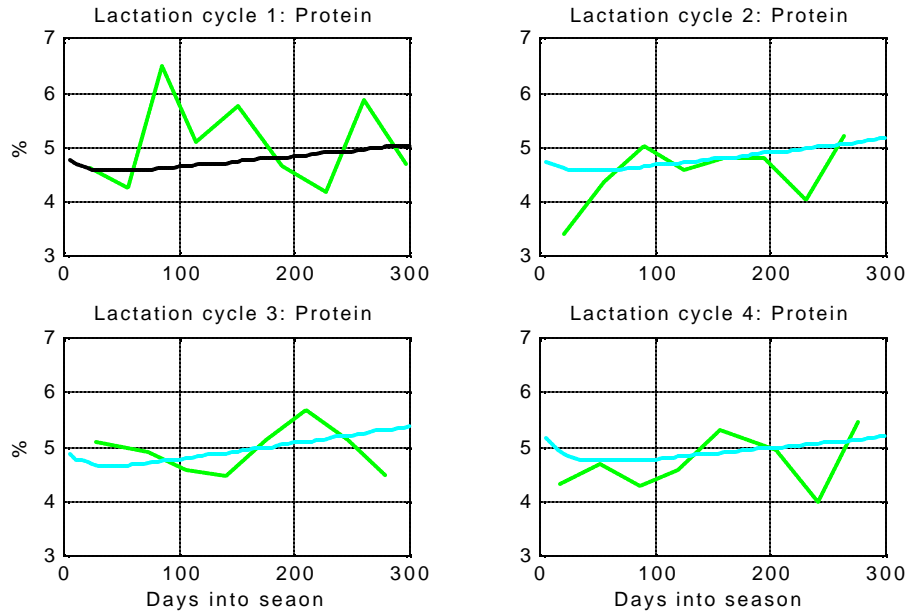
## 8.6 PREDICTION AND THE ADAPTED WILMINK MODEL

From the results of the adapted Wilmink model fitted to the Jersey data by using the Bayesian method, the four lactation cycles of each of the 10 animals in the Further Jersey data were predicted by means of the 4-phase prediction process. The results for the prediction of the protein content curves of each phase of the 4-phase prediction process for the four lactation cycles of animal 2 follow in Figures 8.13 to 8.16 below.

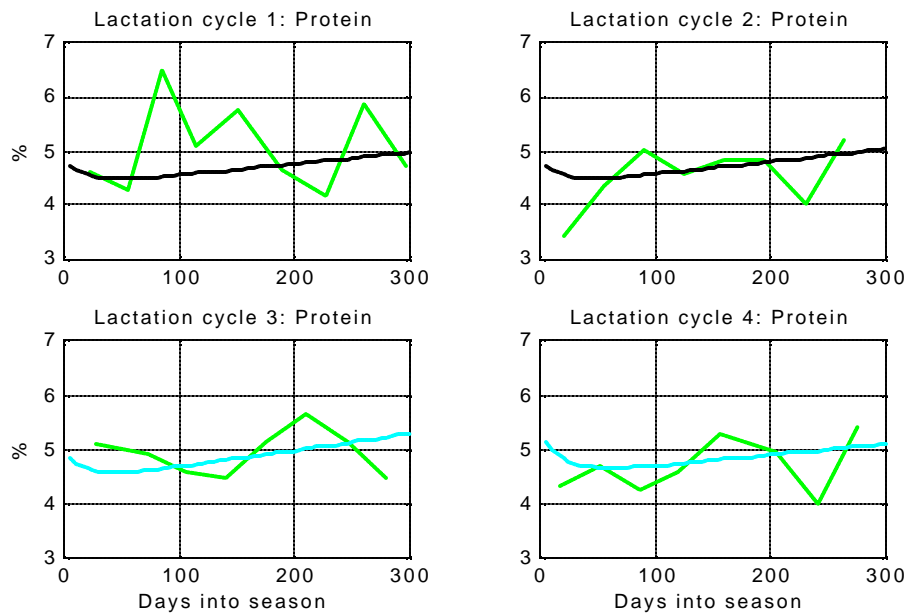
By considering the SSE values of the predicted protein content curves of the lactation cycles in each phase of the prediction process together with these graphs, it is clear that an improvement in prediction of the protein content occurs in all predictions up the third phase of prediction. No reduction in the value of SSE for the prediction of the protein content curve of lactation cycle 4 however occurs from phase 3 to phase 4.



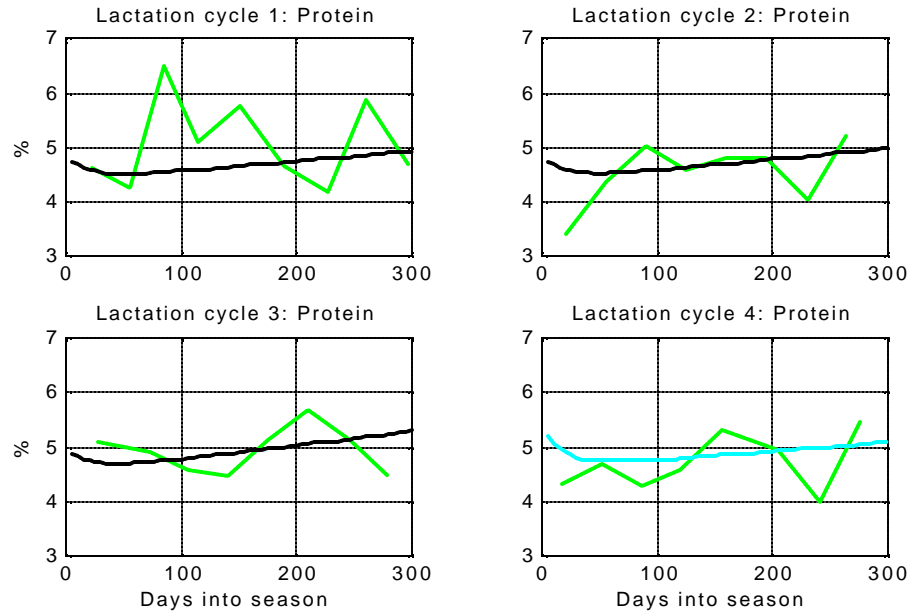
**Figure 8.13: Prediction phase 1. Predicting protein content curves for all 4 lactation cycles of animal 2, where protein content test day records are represented by — and predicted milk yield curves are represented by — .**



**Figure 8.14: Prediction phase 2. Predicting protein content curves for lactation cycles 2, 3 and 4 of animal 2 after the protein content curve of lactation cycle 1 has been fitted. Protein content test day records are represented by —, predicted protein content curves by — and the fitted protein content curve by —.**



**Figure 8.15: Prediction phase 3. Predicting protein content curves for lactation cycles 3 and 4 of animal 2 after protein content curves of lactation cycles 1 and 2 have been fitted. Protein content test day records are represented by —, predicted protein content curves by — and fitted protein curves by —.**



**Figure 8.16: Prediction phase 4. Predicting the protein content curve for lactation cycle 4 of animal 2 after protein content curves of lactation cycles 1, 2 and 3 have been fitted. Protein content test day records are represented by —, the predicted protein content curves by — and the fitted protein content curves by —.**

From Tables 8.19 and 8.21 it follows that for both the lactation traits milk yield and fat content the sum of the SSE values over all 10 animals in the Further Jersey data indicate an improvement in each prediction phase, except the fourth, for all lactation cycles for which trait curves are predicted. In Tables 8.20 and 8.22 when considering the numbers of animals in the predicted lactation cycles of each phase for which there is an improvement in prediction, these however differ for the two traits.

**Table 8.19: Sum of squared errors over all animals for the predicted milk yield curve fitted using the adapted Wilmlink model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	3862,066	2468,901	3521,751	1763,617
1 only (Phase 2)	◆	1311,834	1437,660	1236,904
1 and 2 (Phase 3)	◆	◆	1163,581	908,764
1, 2 and 3 (Phase 4)	◆	◆	◆	1104,218

**Table 8.20:** The number of animals for which, according to the SSE, the predicted milk yield curve from the adapted Wilmink model improved in fit as data on previous lactation cycle became available.

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	7	7	8
1 and 2 (Phase 3)	♦	8	6
1, 2 and 3 (Phase 4)	♦	♦	6

**Table 8.21:** Sum of squared errors over all animals for the predicted fat content curve fitted using the adapted Wilmink model before and as data on preceding lactation cycles became available.

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	7,6878	6,2944	7,6691	12,8178
1 only (Phase 2)	♦	4,2416	4,4920	10,7805
1 and 2 (Phase 3)	♦	♦	3,4783	9,6613
1, 2 and 3 (Phase 4)	♦	♦	♦	9,8153

**Table 8.22:** The number of animals for which, according to the SSE, the predicted fat content curve from the adapted Wilmink model improved in fit as data on previous lactation cycle became available.

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	5	6	6
1 and 2 (Phase 3)	♦	9	7
1, 2 and 3 (Phase 4)	♦	♦	4

Table 8.23 provides the sum of the SSE values over all ten animals in the Further Jersey data when the protein content curves are predicted using the 4-phase prediction process. The total SSE values over all animals are not reduced in lactation cycle 4 of the second phase of prediction or in any of the lactation cycles of the third phase of prediction. From Table 8.24 it follows that these lactation cycles just mentioned also have low numbers of animals for which no improvement in prediction resulted when data on more preceding lactation records became available.

**Table 8.23: Sum of squared errors over all animals for the predicted protein content curve fitted using the adapted Wilmink model before and as data on preceding lactation cycles became available.**

Data available on lactation cycle(s)	Sum of squared errors summed over all ten animals for			
	Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
No cycle (Phase 1)	55,5602	43,1743	29,4088	35,4958
1 only (Phase 2)	♦	38,9655	23,2168	36,0907
1 and 2 (Phase 3)	♦	♦	26,3870	39,0673
1, 2 and 3 (Phase 4)	♦	♦	♦	36,2666

**Table 8.24: The number of animals for which, according to the SSE, the predicted protein content curve from the adapted Wilmink model improved in fit as data on previous lactation cycle became available.**

Data available on lactation cycle(s)	The number of animals for which the prediction improved in		
	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4
1 only (Phase 2)	6	9	5
1 and 2 (Phase 3)	♦	6	4
1, 2 and 3 (Phase 4)	♦	♦	7

## 8.7 PREDICTION OF FURTHER DATA IN SUMMARY

The number of improvements in the predictions for each trait curve over all 60 lactation cycles where improvement for the individual animals in the Further Jersey data is possible (i.e. phases 2, 3 and 4 of the prediction process) with respect to the four lactation models fitted follow in Table 8.25.

**Table 8.25: The number of lactation cycles for which prediction improved as data on preceding lactation cycles became available for the three lactation models applied. Improvement is possible in a total of 60 lactation cycles.**

Predicted curves for	Wood	General Exponential	4-parameter Morant	Adapted Wilmink
Milk yield	43	40	51	42
Fat content	36	35	37	37
Protein content	35	32	17	37

On condition that the data for which such prediction have to made is similar to that for which a lactation model was developed through the Bayesian method in this study, this method is useful in predicting outside of the data set for which modelling initially took place. It is also clear that the Bayesian method is effective in improving the prediction ability of a model as data on preceding lactation cycles become available. It is however important to consider such an improvement in the correct context, i.e. even if the prediction ability of the model improves using this method it does not imply that the model fitted to the data is necessarily the best possible model for the data.

When considering the sum of the SSE values over all 10 animals in the Further Jersey data for each of the expected lactation trait curves in each of the four lactation cycles, the results for each lactation model used to model the original Jersey data are as follow in Table 8.26. Take note that expected lactation trait curves refer to the resulting lactation trait curves obtained when using the results from the original Jersey data and the known test day records of each of the four the lactation cycles.

**Table 8.26: The sum of the SSE values over all 10 animals in the Further Jersey data for each of the expected lactation trait curves in each of the four lactation cycles and the total of the sum of the SSE values over all four lactation cycles for the lactation models**

Model	Lactation trait	Sum of SSE values over all 10 animals for:				Total of sum of SSE's over lactation cycles
		Lactation cycle 1	Lactation cycle 2	Lactation cycle 3	Lactation cycle 4	
Wood	Milk yield	270,1961	262,4482	264,3009	295,6840	1092,6293
	Fat content	2,9999	1,6861	2,3723	4,1843	11,2426
	Protein content	35,1771	28,8601	21,0316	28,2097	113,2786
General Exponential	Milk yield	273,0856	216,5720	224,7424	255,6289	934,0289
	Fat content	2,0573	1,4295	1,6902	2,7165	7,8936
	Protein content	26,9857	23,4978	18,3482	23,5944	92,4261
4-parameter Morant	Milk yield	274,9743	282,2374	294,5502	230,0762	1081,8381
	Fat content	2,0670	1,4505	1,6470	2,9170	8,0816
	Protein content	25,2571	22,5437	17,4763	21,9867	87,2638
Adapted Wilmink	Milk yield	272,8605	271,2695	283,7072	331,1656	1159,0028
	Fat content	2,9133	1,7233	2,1744	3,7227	10,5337
	Protein content	39,0549	30,7830	22,2458	30,2439	122,3277



In the above table the sum of SSE values highlighted in red indicates the lowest values for each lactation cycle, as well as for the total of the sum of the SSE values over the lactation cycles with respect to the expected milk yield curves. Similar lowest SSE values with respect to the expected fat content curves are highlighted in blue and those for the expected protein content curve in orange. From the total of the sum of the SSE values over all ten animals obtained over all four lactation cycles in the Further Jersey data as provided in the last column of Table 8.26 for each of the lactation traits, it follows that for expected milk yield curves in the Further Jersey data the best result over all animals and all lactation cycles is obtained using the General exponential model fitted to the original Jersey data. This result is followed in second place by the expected milk yield curves obtained by means of the 4-parameter Morant model, but only because for lactation cycle 4 in case of milk yield the Morant model resulted in a very low sum of SSE values compared to the sum of SSE values for all lactation cycles in the other models. In the other three lactation cycles the Morant model actually always fared worst.

When considering the expected fat content curves, the lowest overall SSE value for the expected curves of this trait in the last column of the table is also that of General exponential model. This is also true for three of the four lactation cycles with respect to expected fat content curves. It is only in lactation cycle 3 that the 4-parameter Morant model fares slightly better. Overall with respect to expected fat content curves the Morant model is the second best model when considering the sum of the SSE values over all animals.

With respect to the expected protein content curves the 4-parameter Morant model outperformed all other models with respect to the fit obtained. When considering the shape of the expected protein content curve for a level of the cofactor of interest when the effect of all other cofactors have been eliminated in section 6.3.3 of Chapter 6, the question arises as to whether the observed protein content values in the Further Jersey data are indeed such that the more traditionally shaped convex protein content curve should be fitted to it.

Even though through the use of Bayes factors the adapted Wilmink model was found to be best suited to model lactation in the original Jersey data, for the Further Jersey data the best choice of model with respect to obtaining the expected lactation curves of all traits simultaneously would be the General exponential model fitted to the Jersey data, when the sum of SSE values are studied. The 4-parameter Morant model follows in second place, with the Wood and adapted Wilmink models in third and fourth place respectively. It is important

to remember that the expected trait curves for the Further Jersey data are based the observed test day records of the Further Jersey data, together with the results from the relevant model fitted to the original Jersey data.

When considering all of the above, the comment by Wood (1974) that “It is a truism that a model fits best the data from which it was constructed, ...” comes to mind.

# **CHAPTER 9**

## **THE WOOD MODEL AND THE INCOMPLETE SAANEN DATA**

### **9.1 THE SAANEN DATA**

The Saanen dairy goats are the most popular dairy goats in the world and have their origin in the Saanen Valley of Switzerland. Saanen does are exceptionally high yielding milk producers when compared to other goat species and milk usually has a fat content of three to four percent. (Department of Animal Science, Oklahoma State University, 1996.) When compared to cow's milk, milk from Saanen does are very similar in fat and protein content to that of Friesland cows, although it may be a little sweeter in taste (The South African Stud Book and Livestock Improvement Association, 2001). The Saanen is, however, a relatively new breed of goat in South Africa, where the population developed from 1923 onward only, when animals were imported from Switzerland, England and Germany. Goat's milk has two main uses in South Africa. It is often used as baby feed for babies allergic to cow's milk, or as a substitute for cow's milk by people with digestive disorders. An important benefit of goats as milk producers over cows is that they do not contract tuberculosis, which saves on the cost of inoculating animals and testing of milk (SA Milch Goat Breeders' Society, 2001).

The data used in this study was obtained from the South African Stud Book and Livestock Improvement Association in Bloemfontein. The original data set contained information on 1 057 lactation cycles that included a total of 8 941 test day records, collected from 1990 to 1992 for 713 lactating Saanen does kept by a number of registered Saanen goat breeders in South Africa. Information such as farmer member number, animal identification number, kidding date, birth date of doe, parity number, test day date, milk weight in kilograms as measured during each of either two or three milking sessions on each test day, as well as the percentages of fat, protein and lactose in milk composition were included in the data. Of the 1 057 lactation cycles, 90 contained less than 5 test day records, while only 3 lactation cycles were collected for 1992. Removing these lactation cycles together with all lactation cycles for which either parity number or year of kidding or both were unknown, as well as all lactation cycles containing obvious errors, left a data set referred to from this point onwards as the Saanen data.

The Saanen data consist of 755 lactation cycles, collected during 1990 and 1991, for 493 lactating Saanen does. Of the 493 does, 117 were recorded in 1990 only, 114 were recorded in 1991 only, and the remaining 262 were recorded during both years. In total 6 842 observations of test day records for does in various stages of lactations and different parities, with a minimum of 5 and a maximum of 11 test day records for all lactation cycles are contained in the Saanen data. Parity numbers ranged from first to fifth, with 151 lactation cycles in the first parity, 211 in the second parity, 254 in the third parity, 131 in the fourth parity and only eight in fifth parity. It was decided, however, that only four parities should be considered in order to obtain a better balance among the parity groups. For this reason the fourth and fifth parities were combined into a single group referred to as does in parity four or greater.

The number of days from kidding to test day had to be calculated using the kidding date and the test day date for each test day record. Once the number of days from kidding to test day had been established, this value had to be corrected by subtracting 3 days in order to make provision for the initial secretion of colostrum in the first three days after kidding as it is assumed that the secretion of true milk only commences on the fourth day after parturition. In the case of South African Saanen goats the kidding season starts at the end of July and lasts approximately four months, with kidding usually occurring at intervals of at least twelve months. For this reason no attention will be paid in this study to the season of kidding, but rather to the point in time during the season at which kidding occurred. The lactation cycle of a Saanen dairy goat is considered to be 300 days and commences on day four once the secretion of colostrum have ended.

A summarised version of the results in Chapter 9 was published in the *Journal of Agricultural, Biological and Environmental Statistics* (Groenewald and Viljoen, 2003).

## **9.2 THE WOOD MODEL FITTED TO INCOMPLETE DATA**

In the discussion above it was pointed out that, in the case of the Saanen data, lactation cycles were observed over a two-year period. For some of the does in the data two consecutive lactation cycles were observed, one in each year. For other does in this data set only one lactation cycle, which could possibly have occurred in either the first or the second year of

this two-year period, was observed. The resulting data therefore has the same structure as that of incomplete data discussed in section 4.4 of Chapter 4.

The Wood model and the Bayesian approach, with the inclusion of the amendments required in order to handle incomplete data, was used to model lactation in these Saanen does. The Wood model assumes that the expected milk yield (in kg/day) of an animal at time  $t$  is represented over the lactation period by

$$E(W_t) = t^b \exp(a + ct) \quad (9.1)$$

where  $-\infty < a < \infty$ ,  $b > 0$  and  $c < 0$ . The parameters  $a$ ,  $b$  and  $c$  are unknown and may differ from one animal to another.

As discussed in section 5.2 of Chapter 5, the Wood model may be used to estimate the milk composition traits fat and protein content when measured as percentages of composition. However, based on the finding by Sakul and Boylan (1992) that the Wood model was unable to describe the percentage of lactose contained in sheep milk, it was decided not to include the lactose contained in the Saanen milk as part of the modelling process. As a result of the convex nature of the functions of percentages of fat and protein in milk composition over time, model parameters  $b$  and  $c$  for these traits should again be  $b < 0$  and  $c > 0$ .

With the assumption of multiplicative errors in (9.1), and after a log transformation the observation model of the  $i^{\text{th}}$  animal is written as

$$Y_{ijs}(t_{ijp}) = \ln(W_{ijs}(t_{ijp})) = a_{ijs} + b_{ijs} \ln(t_{ijp}) + c_{ijs} t_{ijp} + e_{ijps} \quad (9.2)$$

where  $e_{ijps} \sim i.i.d.N(0, \sigma_s^2)$ ,  $i = 1, \dots, k$  as before. Now, however,  $j$  serves to index the year in which the lactation cycle occurs (year 1 or 2), while  $q_i$  is the number of lactation cycles observed for animal  $i$ , which for this analysis is  $q_i = 1$  or  $2$ . As before  $s = 1, \dots, u$  where  $u$  is the number of lactation traits observed, and  $p = 1, \dots, n_{ij}$  where  $n_{ij}$  is the number of test days for animal  $i$  during lactation cycle  $j$ , with  $8 \leq n_{ij} \leq 10$  for all values of  $i$  and  $j$ . For the Saanen data the number of does are  $k = 493$ , while the  $u = 3$  lactation traits milk yield, percentage of fat and percentage of protein in milk composition are modelled here.

The generalised linear model form of the model for animal  $i$  during lactation cycle  $j$  as described in (4.3) for the general case is:

$$Y_{ij} = X_{ij} M_{ij} + E_{ij} \quad (9.3)$$

where  $\text{vec } \mathbf{E}'_{ij}(3n_{ij} \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_{ij}})$ .

For the Wood model the  $v = 3$  regression coefficients of the generalised linear form of the lactation model for the  $i^{\text{th}}$  animal during its  $j^{\text{th}}$  lactation cycle for the lactation trait indicated by  $s = 1, 2, 3$  are  $a_{ijs}$ ,  $b_{ijs}$ , and  $c_{ijs}$ , so that as before

$$\mathbf{M}_{ij}(3 \times 3) = \begin{bmatrix} a_{ij1} & a_{ij2} & a_{ij3} \\ b_{ij1} & b_{ij2} & b_{ij3} \\ c_{ij1} & c_{ij2} & c_{ij3} \end{bmatrix}. \quad (9.4)$$

The  $r$ -th row of design matrix  $\mathbf{X}_{ij}(n_{ij} \times 3)$  contains the elements that would return the Wood model in generalised linear form when multiplied with the matrix of coefficients  $\mathbf{M}_{ij}$ ,

therefore 
$$\mathbf{X}_{ij}^{(r)} = [1 \quad \ln t_{ijr} \quad t_{ijr}]. \quad (9.5)$$

For animal  $i$  over all  $q_i = 1$  or 2 lactation cycles observed,

$$\mathbf{Y}_i(n_i \times 3) = \begin{bmatrix} \mathbf{Y}_{i1} \\ \vdots \\ \mathbf{Y}_{iq_i} \end{bmatrix} = \mathbf{X}_i \mathbf{M}_i + \mathbf{E}_i \quad (9.6)$$

where  $\text{vec } \mathbf{E}'_i(3n_i \times 1) \sim N(\text{vec } 0, \mathbf{F} \otimes \mathbf{I}_{n_i})$ ,  $n_i = \sum_{j=1}^{q_i} n_{ij}$ ,  $\mathbf{X}_i(n_i \times 3q_i) = \text{diag}\{\mathbf{X}_{ij}\}$  and

$$\mathbf{M}_i(3q_i \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \vdots \\ \mathbf{M}_{iq_i} \end{bmatrix}. \quad (9.7)$$

The dimensions of both  $\mathbf{Y}_i$  and  $\mathbf{M}_i$  therefore depends on the number of lactation cycles  $q_i$  observed for animal  $i$ . For example, if  $q_i = 1$  then  $\mathbf{M}_i(3 \times 3) = [\mathbf{M}_{i1}]$ , but if  $q_i = 2$  then

$$\mathbf{M}_i(6 \times 3) = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \end{bmatrix}.$$

From the earlier discussion on factors that could possibly have a significant influence on lactation, it follows that the additional information on year of kidding, parity number and kidding date available in the Saanen data has to also be included in the model. This is done by means of the covariate matrix or vector  $\mathbf{Z}_i(6 \times q_i)$  for the  $i^{\text{th}}$  animal, which is then used as described in Chapter 4. A full description of how this covariate matrix  $\mathbf{Z}_i$  is obtained for each animal follows in section 9.3 below.

Because  $q_i = 1$  or 2 lactation cycles, the matrix  $\mathbf{R}(q_i \times q_i)$  only contains the covariances between the model parameters of the same animal in successive lactation cycles when  $q_i > 1$ .

In this application where  $\max(q_i) = 2$ ,  $\mathbf{R}_i$  is a  $(2 \times 2)$  matrix if an animal has been recorded over both the lactation cycles observed for 1990 and 1991,

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{12} & \mathbf{r}_{22} \end{bmatrix} \quad (9.8)$$

otherwise  $\mathbf{R}_i$  is a scalar. If an animal has been recorded in one lactation cycle only, depending on the year this lactation cycle was observed,  $\mathbf{R}_i = \mathbf{r}_{11}$  if the does has been recorded in 1990 and  $\mathbf{R}_i = \mathbf{r}_{22}$  if it has been recorded in 1991.

Through the use of the Gibbs sampler as described in section 4.2 of Chapter 4 and the full conditional distributions of the model parameters  $\mathbf{m}_i$ ,  $\mathbf{B}$ ,  $\mathbf{F}^{-1}$  and  $\mathbf{S}^{-1}$  as set out in equations (4.28) to (4.31), and the joint conditional distribution on the elements of  $\mathbf{R}$  given in equation (4.33) together with the Metropolis-Hastings algorithm with restrictions as described in section 4.5 of Chapter 4, the marginal posterior distributions conditional on the observed data only was obtained. The hyperparameters  $\mathbf{P}$  and  $\mathbf{G}$  required to generate marginal distributions for  $\mathbf{S}^{-1}$  and  $\mathbf{R}^{-1}$ , were taken as the moments estimators from the sampling distributions of  $\mathbf{S}$  and  $\mathbf{R}$  with degrees of freedom as small as possible,  $d = 2$  and  $g = 10$ . Applying the approach suggested by MacEachern and Berliner (1994), the sampling process using the Gibbs sampler and Metropolis-Hastings algorithm was put through a “burn-in” period of 2000 simulation iterations, after which 12 000 sets of parameters were generated and kept using equations (4.28) to (4.31) and (4.33). This required simulating from normal distributions to obtain the distributions of  $\mathbf{M}_i$  and  $\mathbf{B}$ , from Wishart distributions for  $\mathbf{F}^{-1}$  and  $\mathbf{S}^{-1}$ , and using the Metropolis-Hastings algorithm with restrictions as set out in section 4.5 of Chapter 4 to obtain the elements of  $\mathbf{R}^{-1}$  from their joint distributions.

### 9.3 SETTING THE COVARIATES CONTAINED $\mathbf{Z}_i$

Additional information available in the Saanen data on cofactors that could possibly have a significant influence on lactation is again included in the model using covariates. The cofactors for the Saanen data are year of kidding, parity number and kidding date. Does kidding in two years, 1990 and 1991, were included in the data, while parity numbers range from 1 to 4 or greater. Kidding date was provided as the actual date of parturition. To include these cofactors in the model, a covariate vector  $\mathbf{z}_i$  or matrix  $\mathbf{Z}_i$  has to be constructed for each animal  $i$ , where  $i = 1, \dots, 493$ .

For does recorded during one lactation cycle only ( $q_i = 1$ ), the cofactors were translated into a vector of covariates  $z_i(6 \times 1)$ , while for animal recorded during both of lactation cycles ( $q_i = 2$ ) a matrix of covariates  $Z_i(6 \times 2) = [z_{i1} z_{i2}]$  was constructed, where the column vectors  $z_{i1}(6 \times 1)$  and  $z_{i2}(6 \times 1)$  are the covariate vectors of the first and second of the two lactation cycles of animal  $i$  respectively. A total of  $m = 6$  covariates were required for each lactation cycle.

In a covariate vector  $z_i$ , for example, the first element or  $z_{i1}$  is always a constant, 1. The second element of this vector is used to identify the year of kidding, with  $z_{i2}$  equal to zero if the doe kidded in 1990 and one if the doe kidded in 1991. Elements three, four and five of such a covariate vector are used to identify the parity number. For an animal in first parity all three these elements are set at zero, for an animal in second parity  $z_{i3}$  is equal to 1 and  $z_{i4}$  and  $z_{i5}$  are set at zero, for an animal in third parity  $z_{i4}$  is equal to one, but  $z_{i3}$  and  $z_{i5}$  are set at zero and so forth. The final element of the covariate vector  $z_i$  is used to indicate the kidding date. However, instead of setting  $z_{i6}$  equal to zero or one, a positive numerical value is assigned to this position in the covariate vector. This value is determined by using the earliest kidding date in the Saanen data for that particular year as reference point and assigning it the time code 1. Other kidding dates for that particular year are then coded as days after earliest kidding date. The kidding date time code of animal  $i$  is then used as its  $z_{i6}$  covariate value. For 1990 the latest kidding date was coded as day 124 after the start of the kidding season, while for 1991 the last kidding date was day 115.

By setting covariates one and six of the covariate vector  $z_i$  equal to one and covariates two to five equal to zero, the so-called base levels of the cofactors are obtained. The base levels of these cofactors are 1990 in the case of year of kidding, parity 1 for parity number and time code zero for kidding date. A typical example of such a covariate vector would be:

$$z_i(6 \times 1) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 34 \end{bmatrix} \quad (9.9)$$

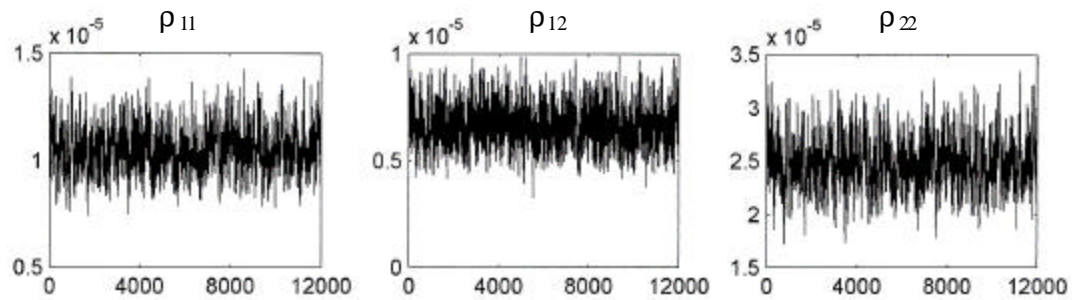
from which it is evident that this doe, was observed during 1991 in its third parity, when it kidded on the 34<sup>th</sup> day since the start of the kidding season if it is assumed that the first doe to kid in that year indicates the start of the kidding season.



A covariate matrix  $\mathbf{Z}_i(6 \times 2) = [z_{i1} z_{i2}]$  simply consist of two such vectors, one for each of the lactation cycles on which lactation records for doe  $i$  are available.

## 9.4 THE RESULTS OF THE GIBBS SAMPLER

After convergence, like in the case of the Jersey data described in section 5.4 of Chapter 5, the results of the parameter sample values obtained through Gibbs sampling for  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\mathbf{F}$  are stable. To illustrate the stability of the results, after convergence, of the three elements of the matrix  $\mathbf{R}$  when 12 000 sets of parameters have been generated by means of the Metropolis-Hastings algorithm, the scans of these elements of  $\mathbf{R}$  are considered in Figure 9.1 below.



**Figure 9.1: Scans of the three elements of the covariance matrix  $\mathbf{R}$  for the 12 000 simulations retained.**

The correlation between the model parameters of the same animal in successive years is estimated from the elements of  $\mathbf{R}$  as 0,412 with 95% HPD interval [0,343 ; 0,477].

The regression matrix  $\mathbf{B}(9 \times 6)$  represents the effects of the covariates on the parameters of the three lactation trait curves, milk yield, percentage of fat and percentage of protein in milk composition, while the covariate vector  $z_i$  or matrix  $\mathbf{Z}_i$  is responsible for the identification of the levels of the cofactors relevant to animal  $i$ . The first column of the matrix  $\mathbf{B}$  contains the base level effects on the three lactation trait curve parameters, while the elements in columns two through to six represent the additional effect on the parameters as a result of the cofactors for which the “levels” are indicated by the covariates. The product of  $\mathbf{B}$  and the covariate vector or matrix of animal  $i$  is used to determine the sum of the base level effects and the additional effects as a result of the settings of the covariates of animal  $i$  on the nine parameters of the three traits curves for that animal. From the 90% HPD intervals of the marginal posterior distributions of the elements of the matrix  $\mathbf{B}$ , it was found that 45 of the 54 elements

significantly affect the parameters of the three lactation trait curves milk yield, percentage of fat and percentage of protein. All nine base level effects contained in  $\mathbf{B}$  are significant, while the effects of all other covariates on the parameters of the trait curves are significant for at least five parameters. The matrix  $\mathbf{B}$  that follows contains the mean of the 12 000 simulated matrices resulting from the Gibbs sampler, with all values indicated in red representing the mean elements of this matrix for which the 90% HPD intervals were found to be significant with respect to their affect on the parameters of the lactation trait curves. All covariates included in the model were kept as a consequence of the results obtained from the 90% HPD intervals.

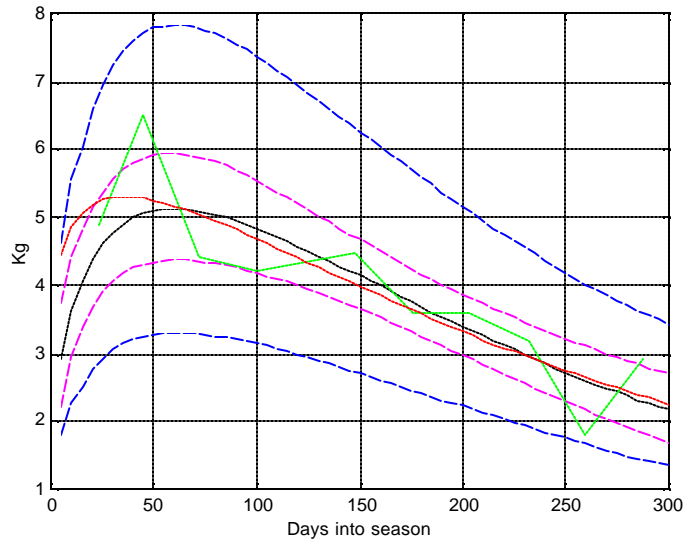
$$\mathbf{B} = \begin{bmatrix} 0,29637 & 0,21546 & 0,24473 & 0,13795 & 0,09278 & -0,00454 \\ 0,27593 & -0,11905 & 0,02769 & 0,07598 & 0,09794 & 0,00143 \\ -0,00423 & 0,00186 & -0,00152 & -0,00182 & -0,00243 & -0,00002 \\ 2,97572 & -0,80337 & -0,23708 & -0,54179 & -0,90765 & -0,01452 \\ -0,52978 & 0,22953 & 0,07117 & 0,14541 & 0,24751 & 0,00342 \\ 0,00422 & -0,00165 & -0,00075 & -0,00118 & -0,00211 & -0,00001 \\ 1,69588 & -0,08635 & -0,10075 & -0,06122 & -0,12391 & -0,00052 \\ -0,16655 & 0,02324 & 0,02570 & 0,00570 & 0,01935 & -0,00013 \\ 0,00136 & -0,00006 & -0,00020 & 0,00004 & -0,00011 & 0,00001 \end{bmatrix}$$

## 9.5 THE LACTATION TRAIT MILK YIELD

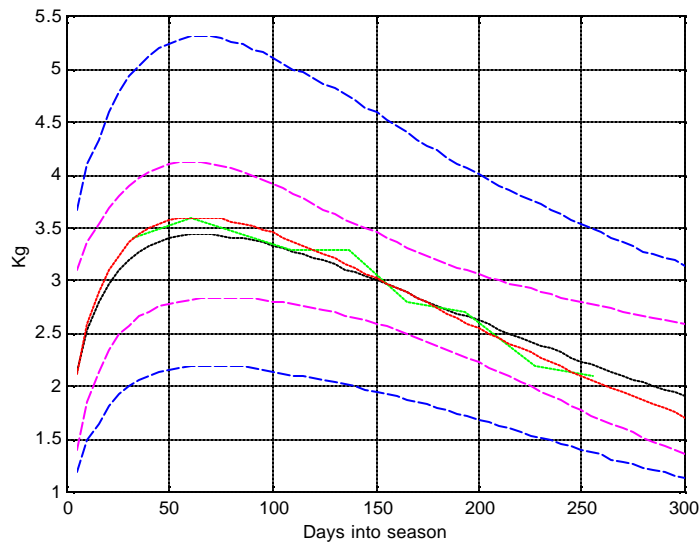
Once the 12 000 simulated  $M_i$  matrices for each animal  $i$  over the lactation cycle or cycles relevant to it have been obtained as a result of Gibbs sampling, 12 000 lactation or milk yield curves may be generated from this result for the lactation cycle or cycles of animal  $i$ . The mean of the 12 000 lactation curves of a lactation cycle of animal  $i$  then gives the expected lactation curve for that animal in the particular lactation cycle under consideration.

To illustrate, this was done for three does in the Saanen data. The first, doe 35 was recoded in 1990 only, when it was in its third parity and kidded on the 14<sup>th</sup> day of the kidding season. The second, doe 162 was recorded in 1991 only, when it was in its first parity and kidded on the 52<sup>nd</sup> day of the kidding season. The expected lactation curve for these does are represented by black solid lines in Figure 9.2 and 9.3 respectively, while the green lines represent the observed milk yield values connected by straight-line segments and the red lines represent the least squares estimate of the lactation curves fitted to the data. The 95% HPD intervals for the expected lactation curves are represented by the magenta coloured broken

lines, while the 95% prediction intervals are displayed as blue broken lines. The same method used to obtain the prediction interval as 95% HPD interval of the predictive density applied to the Jersey data in section 5.5 of Chapter 5 was again used here.

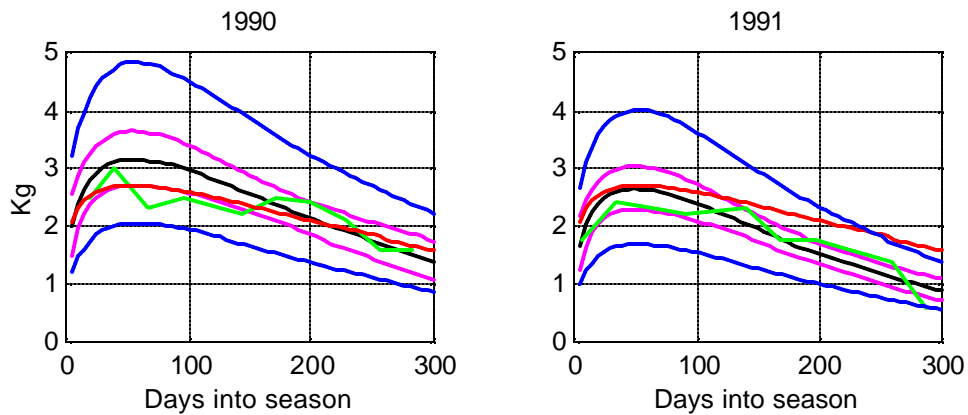


**Figure 9.2:** The expected lactation curve of Animal 35 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 35 is given by —.



**Figure 9.3:** The expected lactation curve of Animal 162 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 162 is given by —.

The third doe or doe 232 was recoded in both 1990 and 1991. In 1990 it was in its fourth parity and kidded on the 18<sup>th</sup> day of the kidding season, while in 1991 it was in its fifth party and kidded on the 21<sup>st</sup> day of the kidding season. For both these years this doe was grouped into the parity group “parity 4 or greater”. The expected lactation curves for each of the two lactation cycles of this doe are represented as a black solid line Figure 9.4, while as before the green line represents the observed milk yield data of each lactation cycle, the red line the least squares estimate of the lactation curve, the magenta coloured broken lines the 95% HPD intervals of the expected lactation curve and the blue broken lines the 95% prediction intervals for the expected lactation curves of the two cycles.



**Figure 9.4:** For each of the two lactation cycle of Animal 232 the expected lactation curve is given by—, its 95% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by — . The observed data for Animal 232 is given by — .

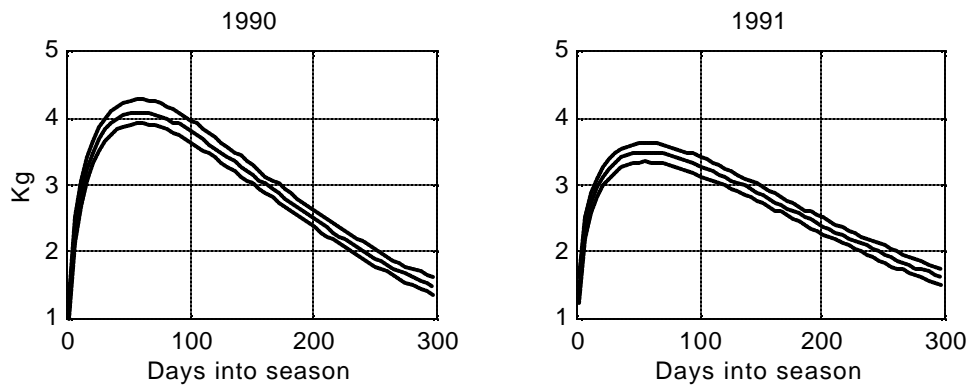
It is also possible to obtain the expected lactation curve of a herd or breed with respect to one of the levels of the cofactors year of kidding and parity number. First 12 000 vectors,

$$m^c = Bz^*, \quad (9.10)$$

for each of the 12 000 simulated  $B$  matrices are calculated, where  $z^*$  is a vector that identifies through its covariates the cofactor level of interest and eliminates the other cofactors by taking the averages over the levels of these cofactors as the values of their covariate(s) in this vector  $z^*$ . For the cofactor “year of kidding”, which only has 2 levels, the average effect of each of its cofactor levels would be  $\frac{1}{2}$ . To average out the effect of the cofactor year of kidding, the value  $\frac{1}{2}$  is assigned to the one covariate used to indicate its level. Similarly for the cofactor “parity number” which has 4 levels identified by 3 covariates, the average effect of  $\frac{1}{4}$  has to be assigned to all 3 these covariates to eliminate its effect through averaging. The only other cofactor present in the model is kidding date, which is in its covariate coded as the

number of days since the start of the kidding season. The time codes since the start of the kidding season is modelled as a Gamma variate, with parameters estimated for each different parity number.

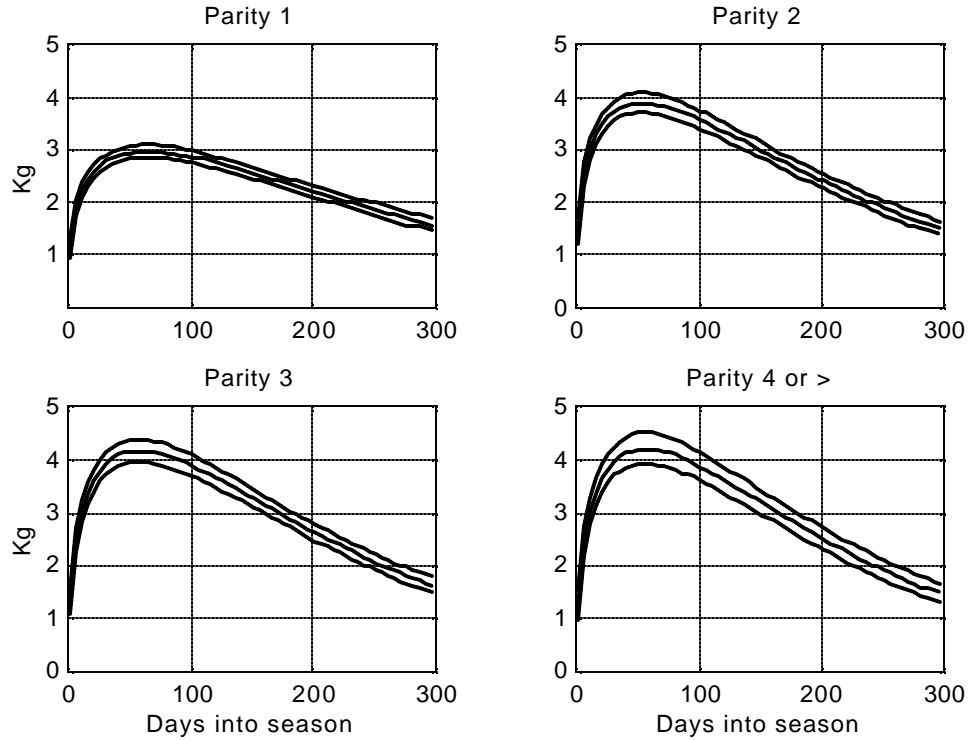
The parameter values contained in  $m^c$  related to milk yield are then used to construct 12 000 lactation curves for the cofactor level of interest and the mean of these curves then gives the expected lactation curve for that cofactor level only. Figure 9.5 below provides the expected lactation curves, together with 95% HPD intervals, for the two years of kidding after the cofactors parity number and kidding date were averaged out as explained above.



**Figure 9.5: Expected lactation curves with 95% HPD intervals for 1990 and 1991 after all other cofactors have been averaged out.**

From Figure 9.5 it is seen that peak yield in 1990 is at a much higher level than in 1990, although time of peak yield seems to be very similar for the two years. It is expected that the total yield in 1990 would be higher than that in 1991. The extent to which the above is true for the mentioned characteristics of milk yield was also investigated using the functions suggested by Wood for these lactation characteristics. The results of this investigation follow in the next section.

In Figure 9.6 expected lactation curves with 95% HPD intervals for the four parity numbers are given after the effects of year of kidding and kidding date were averaged out. After careful consideration of these expected lactation curves for the four parities, it could be commented that parity 1 has the lowest and latest occurring peak milk yield, the flattest curve and, as a result, probably the greatest persistency. Peak milk yields increase with parity number in parities 2 and 3, but in parity 4 or > there again is a slight reduction in peak level. Time of peak milk yield is earliest in parity 2, followed by parity 4 or >, parity 3 and then parity 1.



**Figure 9.6: Expected lactation curves with 95% HPD intervals for the four parities after all other cofactors have been averaged out.**

## 9.6 THE WOOD LACTATION CURVE CHARACTERISTICS

Posterior distributions of the following characteristics as defined by Wood (1967) may again be obtained:

Total milk yield: 
$$f = \int_0^{300} t^b \exp(a + ct) dt, \quad (9.11)$$

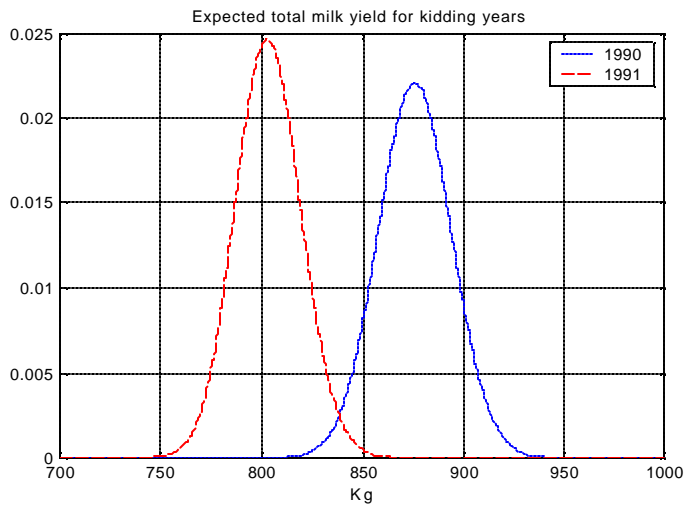
Peak milk yield: 
$$h = \left(-\frac{b}{c}\right)^b \exp(a - b), \quad (9.12)$$

Time of peak yield: 
$$t = -\frac{b}{c} \quad \text{and} \quad (9.13)$$

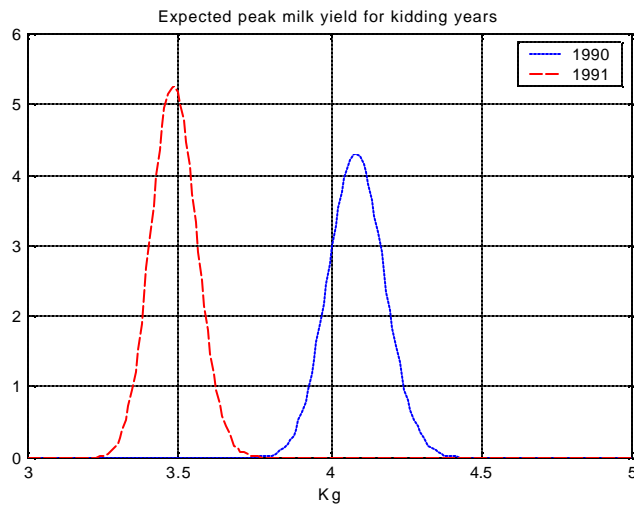
Persistency of lactation: 
$$y = -(b + 1) \ln(-c). \quad (9.14)$$

Once 12 000 function values have been obtained for animal  $i$  using the parameters associated with milk yield in the simulated  $M_i$  matrices for that animal, a histogram is constructed for these values and a Pearson curve fitted to the histogram to find the marginal posterior density of the function.

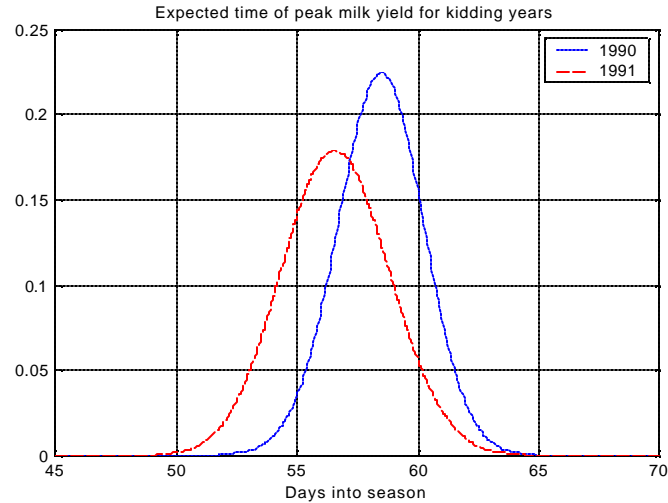
The distributions of these functions with respect to one of the levels of a specified cofactor, rather than for an individual animal, may also be determined by using the same approach applied in sections 5.5.1 to 5.5.4 of Chapter 5. When considering the marginal posterior distributions of these functions per cofactor, the results are as given in Figure 9.7 to 9.14 below. Note that it was not possible to efficiently assess persistency of lactation from Figures 9.5 and 9.6. The marginal posterior distributions of this characteristic for the cofactors kidding year and parity number, however, now enables us to do so.



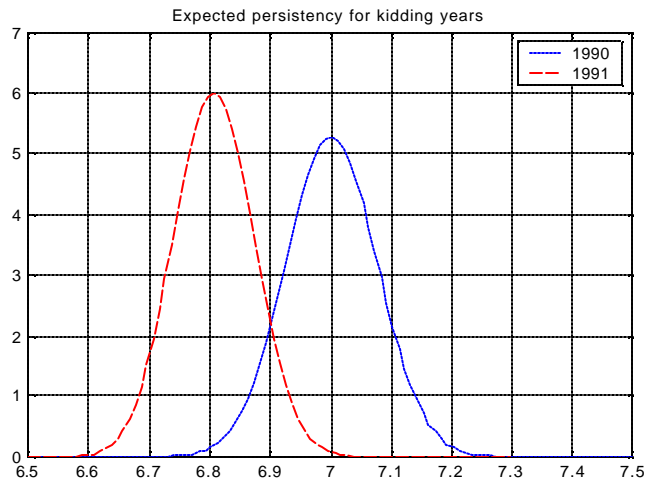
**Figure 9.7: Expected total milk yields for two kidding years after all other cofactors have been averaged out.**



**Figure 9.8: Expected peak milk yields for two kidding years after all other cofactors have been averaged out.**



**Figure 9.9: Expected time of peak milk yields for two kidding years after all other cofactors have been averaged out.**

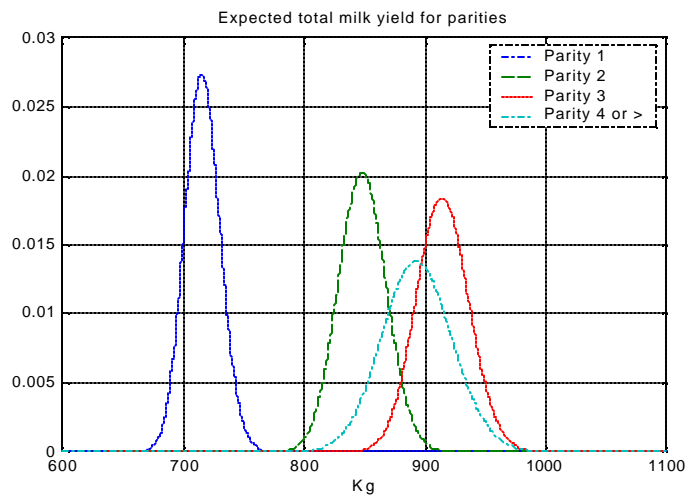


**Figure 9.10: Expected persistency of lactation for two kidding years after all other cofactors have been averaged out.**

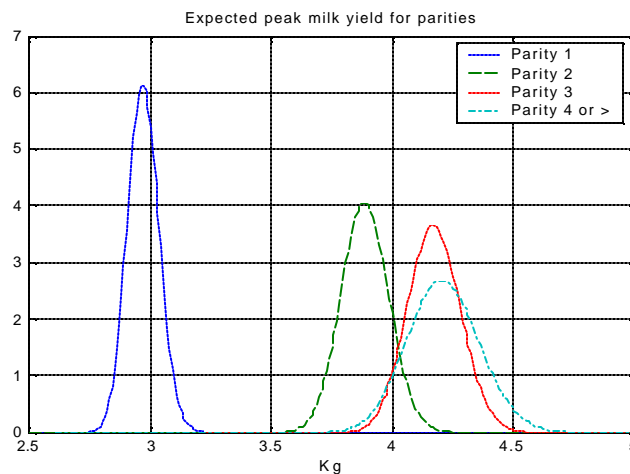
From the results in Figures 9.7 to 9.10 above it is confirmed that expected total milk yield, peak milk yield and persistency of lactation is higher in 1990 than in 1991. Expected time of peak milk yield is also later in 1990 than in 1991. When considering 95% HPD intervals for the differences in these characteristics between 1990 and 1991, only the differences in expected total milk yield and expected peak milk yield between the kidding years were found to be significant. When considering the expected time of peak milk yield in Figure 9.9, the mean of 1990 is only approximately 2 days later than that of 1991. The kidding year 1990, however, out-performed 1991 in all respects as far milk yield in the Saanen goats is concerned. These results corresponds to that of other studies on lactation of dairy goats from



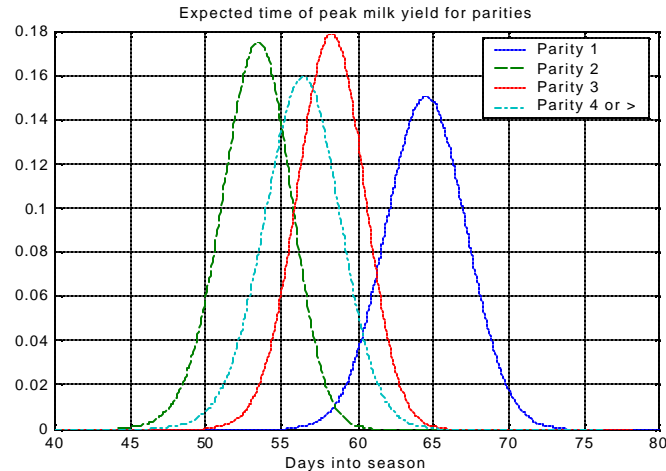
which it was found that the year of kidding significantly affects milk yield. Kominakis et al. (2000) found that the effect of production year was statistically significant in Skopelos dairy goats. Mavrogenis et al. (1984) found that year of kidding had a significant effect on milk yield of Damascus goats in Cyprus. Year of kidding also accounted for variation of milk yield in two Indian goat breeds (Kala and Prakash, 1990) and the production year had a significant effect on the milk yield of Spanish Verata goats (Rabasco et al., 1993). Note that Gipson and Grossman (1990) stated that differences in goat breed are usually not important in accounting for variation in shape of the lactation curve, but that it does play a role in the in the scale of the curve. The shape of a lactation curve usually remains unaffected by breed, but the curve shifts upward for higher producing breeds and downward for breeds with lower production levels.



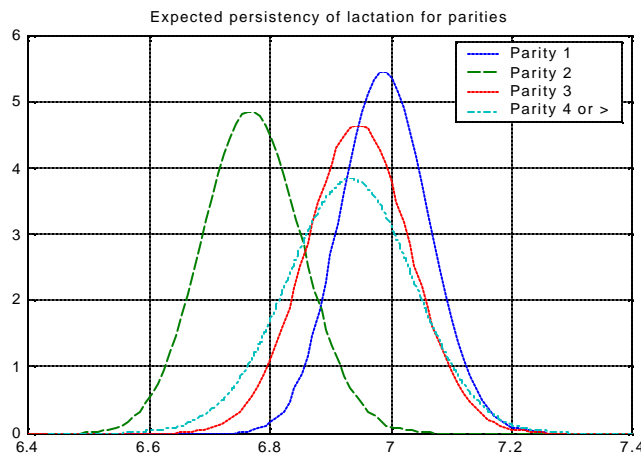
**Figure 9.11** Expected total milk yields for four parities after all other cofactors have been averaged out.



**Figure 9.12** Expected peak milk yields for four parities after all other cofactors have been averaged out.



**Figure 9.13** Expected time of peak milk yields for four parities after all other cofactors have been averaged out.



**Figure 9.14** Expected persistency of lactation for four parities after all other cofactors have been averaged out.

Figure 9.11 to 9.14 show parity 1 to have the lowest expected total and expected peak milk yields. Expected time of peak yield is, however, latest for parity 1, while persistency is greatest in parity 1. For parities 2 and 3 expected total and expect peak milk yield increase with parity number, but for parity 4 or > there is again a slight drop in expected total milk yield to just less than parity 3 level. Expected peak yield is, however, greatest in parity 4 or >. Expected time of peak milk yield is significantly later for parity 1 than for the other parities and, as a result of the combined effect of all the afore mentioned characteristics, parity 1 has the greatest persistency followed by parity 3, 4 or >, and then parity 2 with the lowest persistency. For all four these lactation characteristics parity 4 or > had the greatest variance.

The above results are consistent with the results of other studies on the lactation curves of dairy goats as discussed in Gipson and Grossman (1990) and the numerous references therein. In these studies it was found that the initial and peak yields for first-parity does were usually lower than for later parity does and that time of peak yield was generally later in the lactation cycle of first-parity does than for later-parity does. Because of the flatter lactation curve resulting in the first parity, persistency was found to be greatest in first parity does and decreased with increasing parity. Ruvuna et al. (1995) also found that older does, i.e. those in later parities, have a higher peak yield and lower persistency than younger does or those in earlier parities. Note that there is a negative relationship between the level of production (or total yield) and persistency, as the level of production increases the persistency decreases.

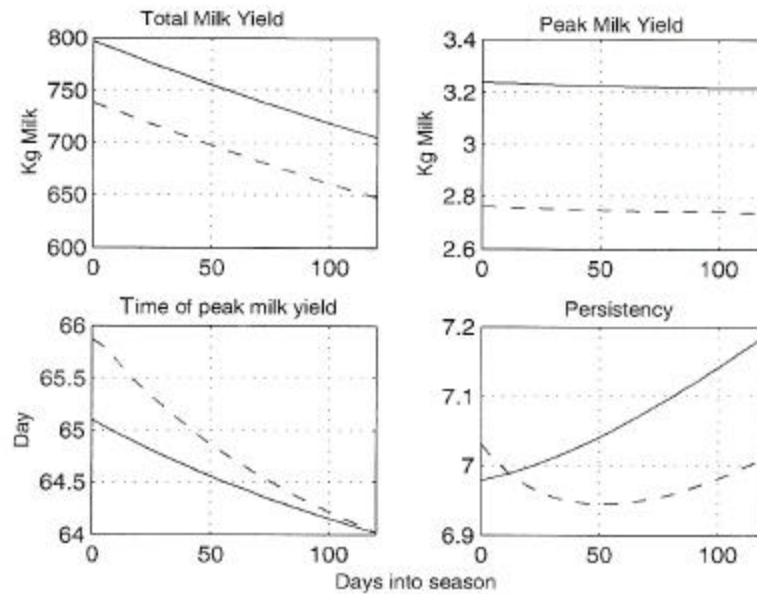
Eight possible combinations of kidding year and parity number are possible for these two cofactors in the data set and any doe in the Saanen data will belong to one of these eight combinations. The only cofactor that could potentially be different for the does in the Saanen data is the time code of kidding date. It was therefore decided to also determine the mean values together with 90% HPD intervals for the marginal posterior distributions of the lactation curve characteristics in equations (9.11) to (9.14) for the eight combinations of the cofactors kidding year and parity number. This was done as explained earlier by obtaining the functions of the characteristics of lactation from the parameter values of milk yield in the 12 000 vectors  $m^c = Bz^*$ , where  $z^*$  was compiled using each of the eight combinations of cofactor levels for the cofactors “kidding year” and “parity number”, while the average kidding date time code of 50 days since the start of the kidding season was used as value of the cofactor kidding date. The result is given in Table 9.1 below.

**Table 9.1: Mean value of four characteristics of the lactation curve, 90% HPD intervals in parenthesis.**

Characteristic	Year	Parity			
		First	Second	Third	Fourth and >
Total yield (kg milk)	1990	755 (732-780)	896 (868-926)	967 (935-999)	945 (895-997)
	1991	698 (678-720)	823 (792-854)	887 (854-920)	865 (828-904)
Peak yield (kg milk)	1990	3.22 (3.11-3.33)	4.21 (4.06-4.36)	4.53 (4.36-4.69)	4.57 (4.30-4.83)
	1991	2.75 (2.63-2.97)	3.60 (3.46-3.74)	3.86 (3.70-4.01)	3.89 (3.70-4.07)
Time of peak yield (days)	1990	64.6 (60.6-68.2)	54.4 (51.3-57.3)	58.8 (55.8-51.7)	57.0 (53.1-60.8)
	1991	64.9 (57.7-71.8)	50.8 (46.4-55.0)	57.0 (53.0-60.9)	54.8 (50.9-58.7)
Persistency	1990	7.04 (6.91-7.16)	6.84 (6.72-6.96)	7.02 (6.89-7.15)	7.01 (6.83-7.18)
	1991	6.94 (6.80-7.08)	6.65 (6.53-6.76)	6.83 (6.77-6.95)	6.80 (6.67-6.93)

The results in table 9.1 correspond in most part to that obtained from the characteristics of the lactation curves for the cofactors year of kidding and parity number in Figures 9.7 to 9.14.

The third cofactor, or kidding date, in the Saanen data had as many as 124 level settings in 1990, making it difficult to meaningfully display its results. Probably the best way of considering the cofactor kidding date, which was coded as days from the start of the kidding season, is to graph the expected posterior characteristics of the lactation curve for the parities as a function of time since the start of the kidding season, i.e. over the levels of the cofactor kidding date, for each or the two kidding years.

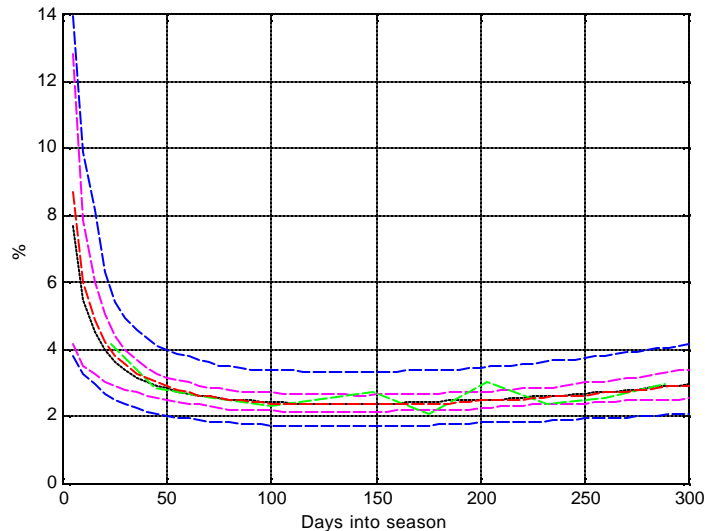


**Figure 9.15:** Expected posterior characteristics of the lactation curve of does in parity 1 for 1990 (—) and 1991(- -) as a function of kidding date coded as days since start of the kidding season.

From Figure 9.15 above it is evident that for first parity does the kidding year 1990 most definitely out-performed 1991. However, the only characteristic that is significantly affected is total milk yield, where the mean yield is highest for animals lactating early during the season, and dropping by more than 90kg for animals lactating at the end of the season. It would also be possible to use the same form of graphical representation to display the effects of the cofactor year of kidding on the characteristics when the effect of the cofactor parity number is averaged out, or the effects of the cofactor parity number when the effect of the cofactor kidding year is averaged out.

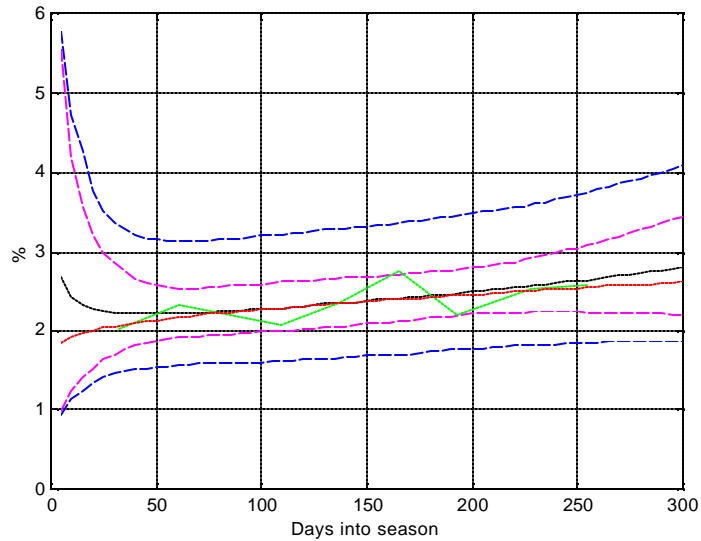
## 9.7 THE LACTATION COMPOSITION TRAITS

The Wood model was used to simultaneously model not only milk yield, but also the milk composition traits fat content and protein content both measured as percentages in the composition of milk. The mean of the 12 000 curves generated for a composition trait of animal  $i$  provides the expected composition trait (fat or protein) curve for that animal in the particular lactation cycle under consideration. This procedure was again followed to obtain the expected trait curves for does 35, 162 and 232 with respect to fat and protein content in milk yield. The results are given in Figures 9.17 to 9.21. In all cases the expected curve is represented by a black solid line, the green line is the observed data, the red line represents the least squares estimate of the curve fitted to the data, the 95% HPD interval of the expected curve is represented by magenta broken lines, and the 95% prediction interval is displayed as blue broken lines. The prediction intervals are again obtained as the 95% HPD intervals of the predictive density in equation (5.10) in section 5.5 of Chapter 5 as mentioned earlier.

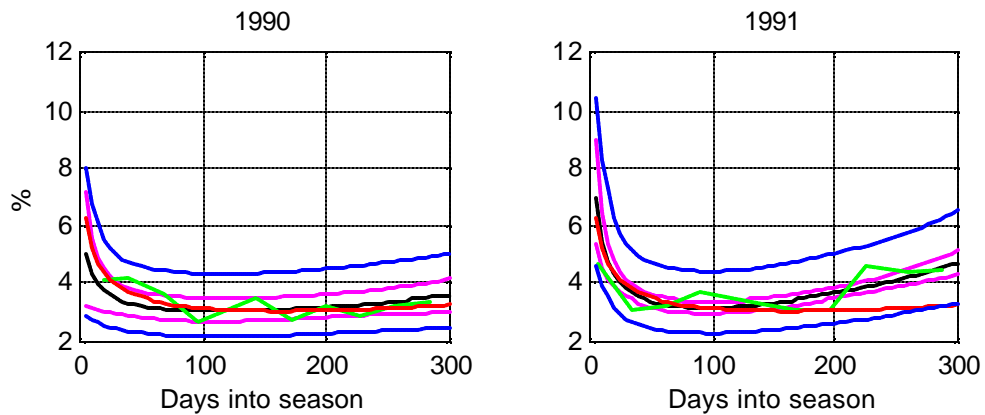


**Figure 9.16:** The expected fat content curve for Animal 35 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 35 is given by —.

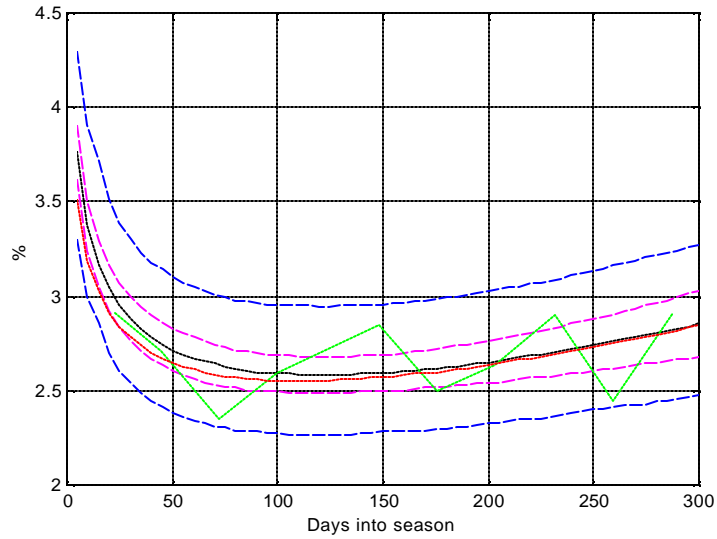
In Figure 9.17 the ability of this Bayesian method in reducing the number of atypical curves fitted to individual animals in the data as result of the influence of information from the whole herd on the expected curve of an individual animal is again demonstrated. The least squares estimate of the fat content curve of animal 162 is atypical in nature, but the expected lactation curve fitted through this Bayesian method is *not* atypical.



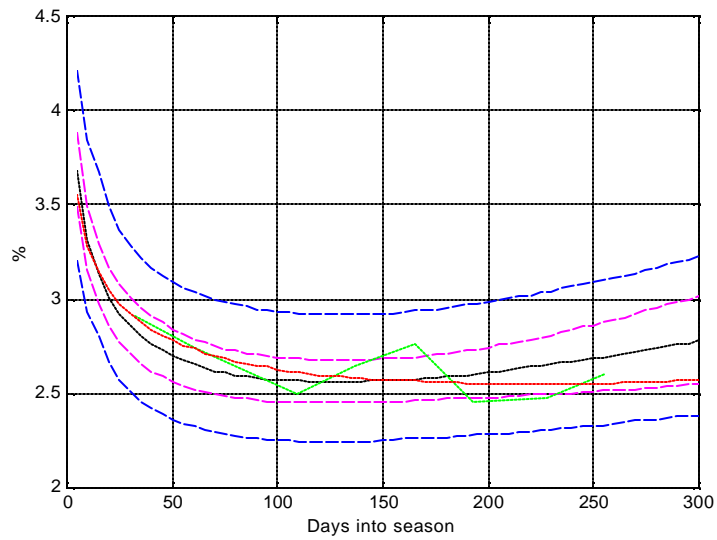
**Figure 9.17:** The expected fat content curve for Animal 162 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 162 is given by —.



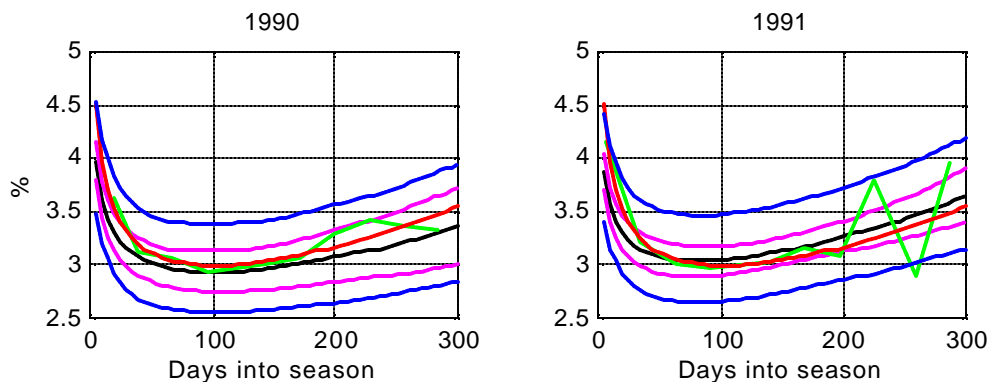
**Figure 9.18:** For each of the two lactation cycle of Animal 232 the expected fat content curve is given by —, its 95% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 232 is given by —.



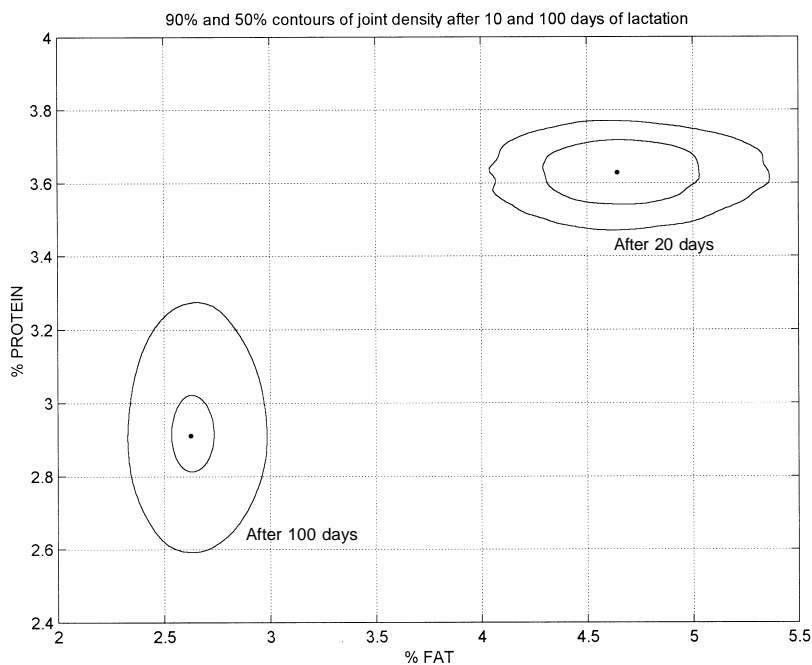
**Figure 9.19:** The expected protein content curve for Animal 35 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 35 is given by —.



**Figure 9.20:** The expected protein content curve for Animal 162 is given by —, its 95% HPD interval by - - and its 95% prediction interval by - - -. The least squares estimate of the lactation curve is given by —. The observed data for Animal 162 is given by —.



**Figure 9.21:** For each of the two lactation cycle of Animal 232 the expected protein content curve is given by —, its 95% HPD interval by - - and its 90% prediction interval by - - . The least squares estimate of the lactation curve is given by —. The observed data for Animal 232 is given by —.



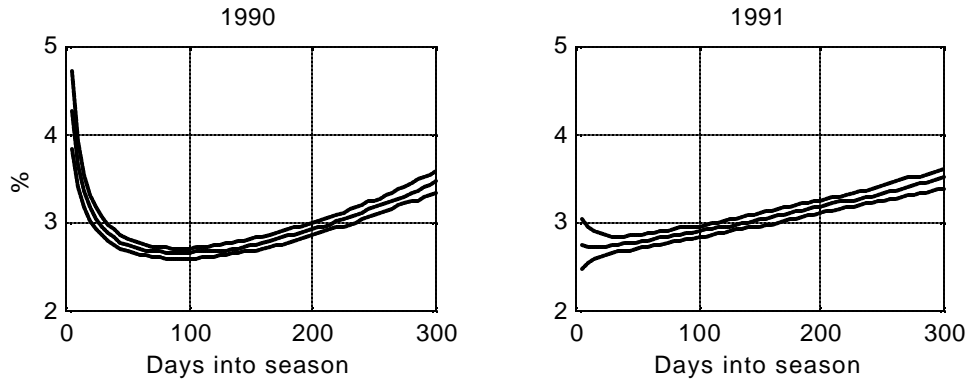
**Figure 9.22:** The mean, 50% and 95% contours of the joint posterior distribution of fat and protein content after 20 and 100 days since the start of lactation for 1990 and parity 1.

Figure 9.22 shows the joint posterior distribution of fat and protein percentage in the composition of milk after 20 days and also after 100 days of from the start of lactation by plotting the mean, 50% and 95% contours of the joint posterior distribution for the cofactor levels 1990 and parity 1, when the effect of the cofactor kidding date is averaged out. This clearly shows that the higher content in both composition traits occur early in the lactation cycle of the Saanen does. Although not much work has been done in the past on the

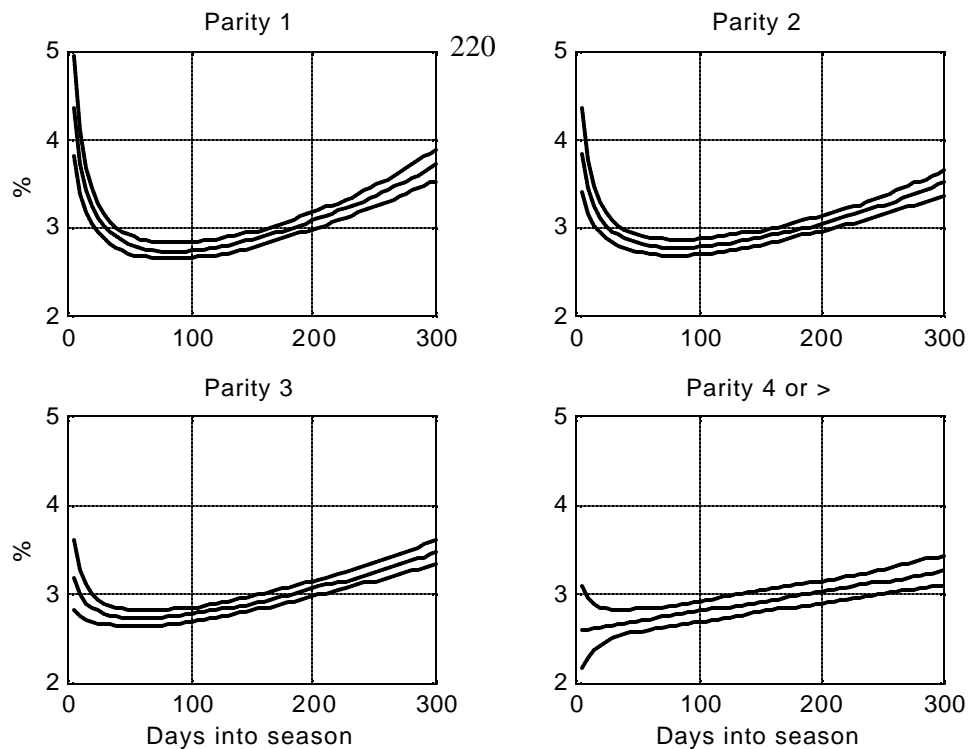


composition of goat milk with respect to fat and protein content, Kominakis et al. (2000) and other references therein draw attention to the negative genetic correlation between milk yield and fat content, and milk yield and protein content.

The expected fat and protein content curves for the herd with respect to one of the levels of the cofactors year of kidding and parity number may be obtained as before. Figures 9.23 to 9.26 provide these expected fat and protein content curves together with 95% HPD intervals.

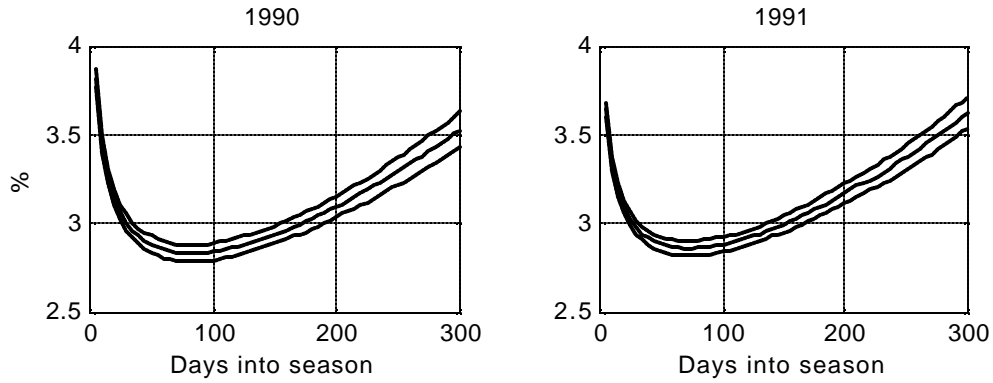


**Figure 9.23:** Expected fat content curves with 95% HPD intervals for 1990 and 1991 after all other cofactors have been averaged out.

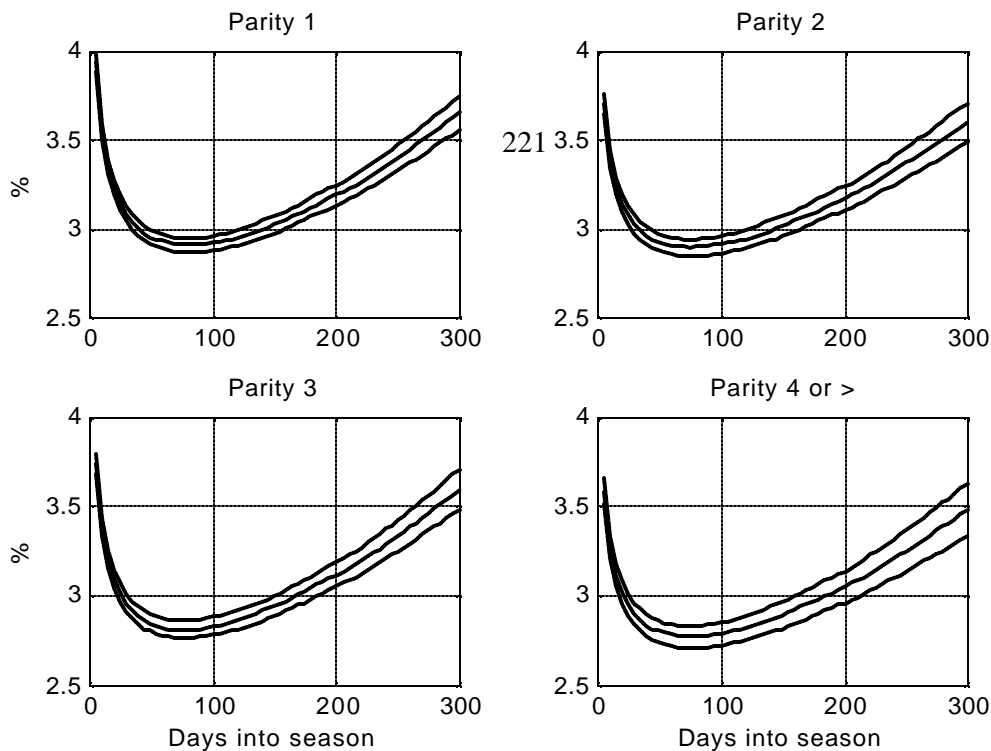


**Figure 9.24:** Expected fat content curves with 95% HPD intervals for the four parities after all other cofactors have been averaged out.

Figure 9.24 shows that parities 1 and 2 are very similar with respect to the expected fat content of milk, except that parity 1 commences with and ends in slightly higher expected percentages. Parity 4 or greater commences with and ends in the lowest expected fat content when compared to all other parities.



**Figure 9.25:** Expected protein content curves with 95% HPD intervals for 1990 and 1991 after all other cofactors have been averaged out.



**Figure 9.26:** Expected protein content curves with 95% HPD intervals for the four parities after all other cofactors have been averaged out.

Figure 9.26 shows that the expected protein content of milk in parity 1 commences and ends with slightly higher percentages than the other parities. Parity 4 or greater seems to have the

lowest expected protein content throughout most of the lactation cycle when compared to all other parities.

## 9.8 PREDICTING MISSING LACTATION RECORDS

Lactation records in the Saanen data were recorded over two kidding years or lactation cycles, 1990 and 1991. Of the total of 493 does in the data, only 262 were recorded during both lactation cycles. For 117 does recorded in 1990 only, it is possible to predict the missing lactation record for 1991.

In section 4.1 of Chapter 4 the procedure for the general case where, of a possible number of  $q$  consecutive lactation cycles for animal  $i$  of which the lactation records of  $h < q$  cycles are known and  $(q - h)$  cycles have to be predicted, is explained. For the Saanen data each doe has a total of  $q = 2$  possible lactation cycles and for all does recorded in 1990 only this implies that  $h = 1$  lactation cycle is known and that one further lactation cycle has to be predicted to complete missing lactation records for those does.

To do so the model parameters of the lactation cycle to be predicted or  $\mathbf{m}_{if}^{(2)}$  have to be determined if the model parameters of the known lactation cycle is given by  $\mathbf{m}_i^{(1)} = \mathbf{m}_i^{(1)*}$ . This is done through MCMC simulation using the conditional distribution of  $\mathbf{m}_{if}^{(2)}$  specified in equation (4.22). For the Saanen data, however, where  $q = 2$  and  $h = 1$ , this conditional distribution for predicting the second year's model parameters for an individual or  $\mathbf{m}_{if}^{(2)}$ , given the first year's performance reduces to:

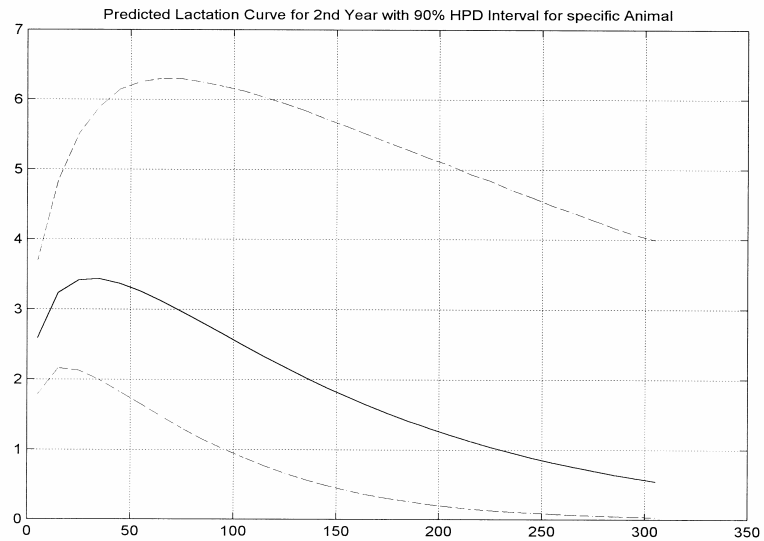
$$\mathbf{m}_{if}^{(2)} | \mathbf{m}_i^{(1)}, \mathbf{B}, \mathbf{R}, \mathbf{S} \sim N \left( \mathbf{Bz}_{i2} + \frac{\mathbf{r}_{12}}{\mathbf{r}_{11}} (\mathbf{m}_i^{(1)} - \mathbf{Bz}_{i1}), \frac{1}{\mathbf{r}_{11}} | \mathbf{R} | \mathbf{S} \right). \quad (9.15)$$

For predicting the second years performance,  $\tilde{\mathbf{Y}}_{if}$ , we then use

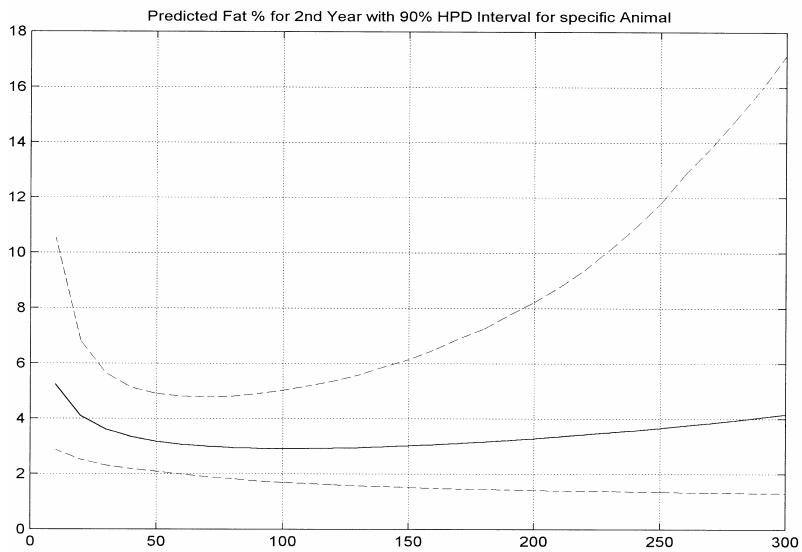
$$\mathbf{y}_{if}^{(2)} | \mathbf{m}_{if}^{(2)} = \mathbf{m}_{if}^{(2)*}, \mathbf{F} \sim N \left( (\mathbf{I}_3 \otimes \mathbf{X}_{if}^{(2)}) \mathbf{m}_{if}^{(2)}, \mathbf{F} \otimes \mathbf{I}_{n_{if}} \right). \quad (9.16)$$

From the result of equation (9.16) the lactation or milk yield curve, as well as the fat and protein content curves for the second year given first year's performance may then be

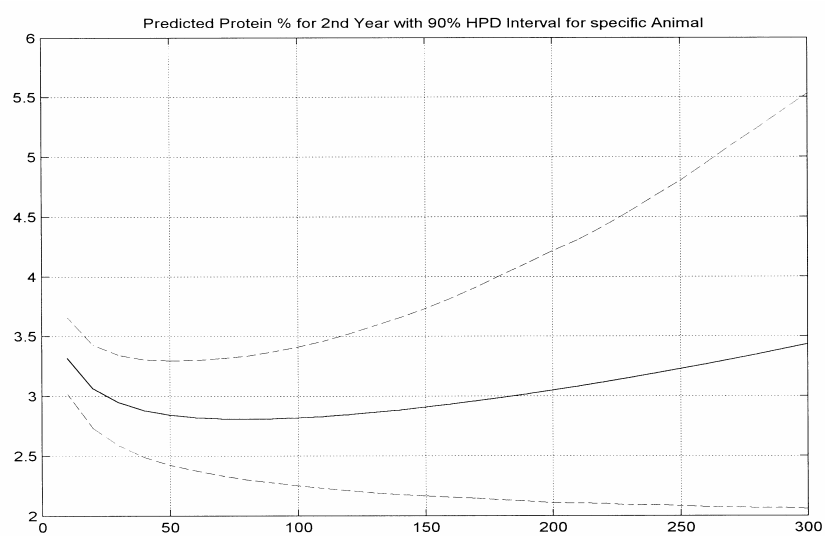
graphed. If this is done for doe 35 in the data for which only the lactation record of 1990 was observed, the results are as follows.



**Figure 9.27: Predicted lactation curve with 90% HPD intervals for doe 35 in 1991.**



**Figure 9.28: Predicted fat content curve with 90% HPD intervals for doe 35 in 1991.**



**Figure 9.29:** Predicted protein content curve with 90% HPD intervals for doe 35 in 1991.

## 9.9 SAANEN DATA RESULTS: A SUMMARY

In general, the season of lactation, 1990 or 1991, significantly affected total yield and peak yield, with 1991 resulting in lower values of both these characteristics in all four parities. Total milk yield was in turn affected by the time of kidding. Total milk yield was higher for does that started lactation early during the season and declined for does that started lactation later during the season.

Parity had a large effect on the characteristics of the lactation curve in dairy goats. It was found that peak yield increased with increasing parity up to about the third or fourth parity, while time of peak yield is later for first-parity does than for later parity does. The effect of parity on total milk yield was again similar to the effect on peak yield, with total milk yield increasing up to the third parity, where after the total milk yield starts to decline. Persistency was influenced by parity in the same way as total milk yield. The effect of parity on total milk yield and persistency could play an important role in management of dairy animals with respect to culling.

# CHAPTER 10

## EPILOGUE

### 10.1 THE RESULTS IN SUMMARY

The Wood, general exponential, 4-parameter Morant and adapted Wilmlink models were used to simultaneously model the lactation traits milk yield, percentage of fat and percentage of protein in the milk composition from test day records observed for Jersey cows by means of a Bayesian approach using MCMC methods. The results obtained from the Wood model, when applied to lactation of dairy cows, in the main corresponds to the results and conclusions obtained in other studies with respect to the effects of cofactors, such as parity, region and season of calving, on the lactation curve characteristics total milk yield, peak milk yield, time of peak milk yield, and persistency, as discussed in Chapter 5. Although not much research on modelling the fat and protein content of milk by means of the Wood model have previously been done, the results and conclusions obtained in this research also mainly corresponds to those of other studies.

When considering the expected lactation curves with respect to the levels of a cofactor of interest obtained from the general exponential, 4parameter Morant and adapted Wilmlink models, the adapted Wilmlink model fared best at describing milk yield. The 4-parameter Morant model often resulted in curves that peak when lactation commenced which is considered an unrealistic scenario with respect to lactation in Jersey cows, while the general exponential model tends to under estimate milk yield in the final stages of lactation. Considering the expected fat curves with respect to the levels of a cofactor of interest very similar results were obtained for the four models. Fat content was also the only lactation trait in the original Jersey data that the 4parameter Morant model seemed to model effectively. With respect to expected protein content of the levels of a cofactor of interest the 4-parameter Morant model seemed unable to take on the shape expected for such curves, while the results from the Wood, adapted Wilmlink and general exponential models were similar in scale and shape.

When comparing the ability of all four models in modelling the three lactation traits of the Jersey data using Bayes factors, the adapted Wilmlink model is the model of preference,

followed by the Wood and general exponential models, while the 4parameter Morant model fared worst.

The results from the Bayesian approach for these models obtained from the original Jersey data were also used to predict and model lactation with respect to the Further Jersey data that contains animals similar to those in the original Jersey data but that do not form part of the original Jersey data. Here it was found that the general exponential model resulted in the lowest total sum of SSE values over all lactation cycles of all 10 animals in the Further Jersey data with respect to the lactation traits milk yield and fat content, while for protein content the 4-parameter Morant model provided the lowest total sum of SSE values over all lactation cycles of all 10 animals. This was however the case when using results generated for the models fitted to the original Jersey data and *not* by modelling the Further Jersey data using these models and the Bayesian approach.

To illustrate the use of the Bayesian approach in modelling incomplete lactation data, the Wood model was used to model the Saanen data. The results obtained with respect to the effects of cofactors such as year of kidding, parity and kidding date on the lactation curve characteristics total milk yield, peak milk yield, time of peak milk yield and persistency also on the whole corresponds to those of other studies in which the Wood model was used to model lactation in goats. This correspondence was also found with respect to the results of the fat and protein content of the milk of goats modelled by means of the Wood model.

It was also illustrated how the Bayesian approach also allows for the prediction of lactation trait curves in future lactation cycles of animals contained in the Jersey using the Wood, general exponential, 4parameter Morant and adapted Wilmink model, as well as for animals in the Saanen data using the Wood model.

The ability of the Bayesian approach in reducing the occurrence of atypical curves fitted to milk yield, fat content and protein content is also very evident throughout this study.

## **10.2 CONTRIBUTION TO LACTATION RESEACH IN PERSPECTIVE**

In this research a hierarchical Bayesian approach was applied in modelling lactation data with the intention of investigating lactation traits and not for the purpose of genetic evaluation and

selection as was done before. For the first time consideration is given to the effect that the assumption about errors has on fitting the various forms of lactation models to the data and the problems encountered with respect to some of the models based on the error assumption.

A generalised hierarchical Bayesian approach to modelling lactation using any lactation model, together with the assumption of multiplicative errors, that may then be re-written in linear form by means of a logarithmic transformation so that the errors then become additive, was developed.

The effects of cofactors were included into the model through the use of covariates in order to enable the study of the effects of the various levels of such cofactors on the lactation trait curves.

The method that averages out the effects of certain cofactors in order to obtain the lactation trait curve applicable to a certain level of the specific cofactor of interest was introduced. This provides a Bayesian equivalent to Standard Lactation Curves (SLAC's) obtained by means of IRIS-Dairy.

This modelling approach has the ability to handle both complete and incomplete data. Complete data referring to the case where the same lactation cycles were recorded for all animals in the data, while in the case of incomplete data not all animals were recorded during all the considered lactation cycles. From the results of such a model it is possible to predict missing lactation cycles for animals based on preceding lactation cycles for which test day records were observed and included in modelling. Future lactation cycles of the animals in the data following those already included in modelling may also be predicted.

It is also possible to predict lactation trait curves for animals that did not form part of the data set for which modelling took place, on condition that it may be assumed that such animals are similar to those for which modelling was done.

An adaptation of the Wilmink model II was performed in order to enable its use as a model that, when multiplicative errors are assumed, may be transformed using a logarithmic transformation so that the model becomes linear with additive errors. In addition to this  $w$  in the adapted Wilmink model is used as a parameter applicable to the herd under investigation instead of as a constant as was previously done in the Wilmink II model. The resulting



adapted Wilmink model has proved to be a model that is worthy of further consideration in all the studies of lactation in dairy animals.

In summary, the main advantages of the Bayesian approach, combined with MCMC methods, is, firstly, the ease with which it can handle the estimation of parameters in complex multi-level hierarchical models, and, secondly, the ability to generate accurate credibility intervals for these parameters. Also, credibility intervals for non-linear combinations of these parameters, such as the lactation characteristics, and prediction intervals follow directly. This is not possible when applying the classical methods, except with the use of approximations.

## REFERENCES

- [1] Ali, T.E. and Schaeffer, L.R. (1987). Accounting for covariance among test day milk yields in dairy cows. *Can. J. Anim. Sci.*, 67:637-644.
- [2] Apka, G.N., Asiribo, E.O., Oni, O.O. and Alawa, J.P. (2001). The influence of non-genetic factors on the shape of lactation curves in Red Sokoto goats. *Anim. Sci.*, 72:233-239.
- [3] Arman, P., Kay, R.N.B., Goodall, E.D. and Sharman, G.A.M. (1974). The composition and yield of milk from captive red deer (*Cervus elaphus L.*). *J. Reprod. Fertil.*, 37:67-84.
- [4] Batra, T.R. (1986). Comparison of two Mathematical Models in fitting Lactation curves for Pureline and Crossline Dairy Cows. *Can. J. Anim. Sci.*, 66:405-414.
- [5] Brody, S., Ragsdale A.C. and Turner C.W. (1923). The rate of decline of milk secretion with the advance of the period of lactation. *J. Gen. Physiol.*, 5:441-444.
- [6] Brody, S., Turner C.W. and Ragsdale A.C. (1924). The relation between the initial rise and the subsequent decline of milk secretion following parturition. *J. Gen. Physiol.*, 6:541.
- [7] Cappio-Borlino, A., Pulina, G. and Rossi, G. (1995). A non-linear modification of Wood's equation fitted to lactation curves of Sardinian dairy ewes. *Small Ruminant Res.*, 18:75-79.
- [8] Casella, G. and George, E.I. (1992). Explaining the Gibbs Sampler. *Am. Stat.*, 46:167-174.
- [9] Chang Y-M., Rekaya, R., Gianola, D. and Thomas, D.L. (2001). Genetic variation of lactation curves in diary sheep: a Bayesian analysis of Wood's function. *Livest. Prod. Sci.*, 71:241-251.
- [10] Chang Y-M., Rekaya, R., Gianola, D. and Thomas, D.L. (2002). Genetic variation of lactation curves in dairy sheep: quadratic function versus Wood's model. *Livest. Prod. Sci.*, 77:147-153.

- [11] Chib, S. (1995). Marginal Likelihood From the Gibbs Output. *J. Am. Stat. Assoc.*, 90:1313-1321.
- [12] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *Am. Stat.*, 49: 327-335.
- [13] Cobby, J.M. and Le Du, Y.L.P. (1978). On fitting curves to lactation data. *Anim. Prod.*, 26:127-133.
- [14] Dave, B.K. (1971). First lactation curve of Indian water buffalo. *JNKVV Res J.*, 5:93.
- [15] De Boer, J.A., Weller, J.I., Gipson, T.A. and Grossman, M. (1989). Multiphasic Analysis of Milk and Fat Yield Curves of Israeli Holsteins. *J. Dairy Sci.*, 72:2143-2152.
- [16] Department of Animal Science, Oklahoma State University (1996). Breeds of Livestock: Saanen. [Online]. Available: <http://www.ansi.okstate.edu/breeds/goats/>, [Assessed: 8 December 2005].
- [17] Department of Animal Science, Oklahoma State University (1997). Breeds of Livestock: Jersey. [Online]. Available: <http://www.ansi.okstate.edu/breeds/cattle/>, [Assessed: 7 October 2005].
- [18] Dhanoa, M.S. (1981). A note on an alternative form of the lactation model of Wood. *Anim. Prod.*, 32:349-351.
- [19] Ferris, T.A., Mao, I.L. and Anderson, C.R. (1985). Selecting for Lactation Curve and Milk Yield in Dairy Cattle. *J. Dairy Sci.*, 68:1438-1448.
- [20] Fischer, A. (1958). Research with Württemberg Spotted Mountain cows on the shape of the lactation curve and how it may be influenced by non-genetic factors. *Züchtungskunde*, 30:296-304.
- [21] Franci, O., Pugliese, C., Acciaioli, A., Parisi, G. and Lucifero, M. (1999). Application of two models to the lactation curve of Massese ewes. *Small Ruminant Res.*, 31:91-96.
- [22] Freeze, B.S. and Richards, T.J. (1992). Lactation Curve Estimation for use in Economic Optimization Models in the Dairy Industry. *J. Dairy Sci.*, 75:2984-2989.

- [23] Fuller, W.A. (1969). Grafted polynomials as approximating functions. *Aust. J. Agric. Econ.*, 13:35-46.
- [24] Garcia, A., Landete-Castillejos, T., Molina, A. Albiñana, B., Fernández, C., Garde, J. and Gallego, L. (1999). Lactation curves in captive Iberian red deer (*Cervus elaphus hispanicus*). *J. Anim. Sci.*, 77:3150-3155.
- [25] Gelfand, A.E., Hills, S.E, Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *J. Am. Stat. Assoc.*, 85:972-985.
- [26] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *J. Am. Stat. Assoc.*, 85:398-409.
- [27] Gelman, A. (1992). Iterative and Non-Iterative Simulation Algorithms. *Computer Science and Statistics: Proceedings of the 24<sup>th</sup> Symposium on the Interface.*, 24:433-438.
- [28] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin D. B. (1995). Markov chain simulation, in *Bayesian Data Analysis*, London: Chapman & Hall, pp. 322-326.
- [29] Geman, S. and Geman, D (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721-741.
- [30] Geyer, C.J. (1991). Markov Chain Monte Carlo Maximum Likelihood. *Computer Science and Statistics: Proceedings of the 23<sup>rd</sup> Symposium on the Interface.*, ed. E.M. Keramidas, Fairfax Station, 23:156-163.
- [31] Gipson, T.A. and Grossman, M. (1989). Diphasic Analysis of Lactation Curves in Dairy Goats. *J. Dairy Sci.*, 72:1035-1044.
- [32] Gipson, T.A. and Grossman, M. (1990). Lactation curves in dairy goats: a review. *Small Ruminant Res.*, 3:383-396.
- [33] Good, I.J. (1985). Weight of Evidence: A Brief Survey, in *Bayesian Statistics 2*, eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, New York: Elsevier, pp 249-269.

- [34] Goodall, E.A. (1983). An analysis of seasonality of milk production. *Anim. Prod.*, 36:69-72.
- [35] Goodall, E.A. (1986). Prediction of Milk and Milk Solids Production. *Agric. Syst.*, 21:189-200.
- [36] Goodall, E.A. and Sprevak, D. (1984). A note on a stochastic model to describe the milk yield of a dairy cow. *Anim. Prod.*, 38:133-136.
- [37] Goodall, E.A. and Sprevak, D. (1985). A Bayesian estimation of the lactation curve of a dairy cow. *Anim. Prod.*, 40:189-193.
- [38] Groenewald, P.C.N., Ferreira, A.V., Van der Merwe, H.J. and Slippers, S.C. (1995). A mathematical model for describing and predicting the lactation curve of Merino ewes. *Anim. Sci.*, 61:95-101.
- [39] Groenewald, P.C.N., Ferreira, A.V., Van der Merwe, H.J. and Slippers, S.C. (1996). Application of Bayesian inference in the comparison of lactation curves of Merino ewes. *Anim. Sci.*, 62:63-69.
- [40] Groenewald, P.C.N. and Viljoen, C.S. (2003). A Bayesian Model for the Analysis of Lactation Curves of Dairy Goats. *J. Agric. Biol. Environ. Stat.*, 8:75-83.
- [41] Grossman, M., Hartz, S.M. and Koops, W.J. (1999). Persistency of Lactation Yield: A Novel Approach. *J. Dairy Sci.*, 82:2192-2197.
- [42] Grossman, M. and Koops, W.J. (1988). Multiphasic Analysis of Lactation Curves in Dairy Cattle. *J. Dairy Sci.*, 71:1598-1608.
- [43] Guo, Z. and Swalve, H.H. (1995). Modelling of the lactation curve as a sub-model in the evaluation of test day records. *Paper presented at the INTERBULL open meeting, 7-8 September 1995, Prague, Czech Republic.*
- [44] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57:97-109.

- [45] Hurley, W.L., (2003). Lactation Biology. [Online]. Available: <http://classes.aces.uiuc.edu/AnSci308/intro.html>, [Accessed 24 September 2003], University of Illinois, Urbana-Champaign.
- [46] Jamrozik, J. and Schaeffer, L.R. (1997). Estimates of Genetic Parameters for a Test Day Model with Random Regressions for Yield Traits of First Lactation Holsteins. *J. Dairy Sci.*, 80:762-770.
- [47] Jeffreys, H. (1935). Some Tests of Significance, Treated by the Theory of Probability. *Proceedings of the Cambridge Philosophy Society*, 31:203-222.
- [48] Jeffreys, H. (1961). *Theory of Probability*, 3<sup>d</sup> Ed., Oxford: Oxford University Press.
- [49] Jenkins, T.G. and Ferrell, C.L. (1984). A note on lactation curves of crossbred cows. *Anim. Prod.*, 39:479-482.
- [50] Kala, S.N. and Prakash, B. (1990). Genetic and Phenotypic Parameters of Milk Yield and Milk Composition in two Indian Goat Breeds. *Small Ruminant Res.*, 3:475-484.
- [51] Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *J. Am. Stat. Assoc.*, 90:773-795.
- [52] Keown, J.F., Everett, R.W., Empet, N.B. and Wadell L.H. (1986). Lactation Curves. *J. Dairy Sci.*, 69:769-781.
- [53] Kim, N.S., Ju, J.C., Won, Y.S., Choi, S.B. and Choi, Y.L. (1998). A Comparison of Different Equations of the Lactation Curve in Hanwoo. *Proceedings: 6<sup>th</sup> World Congress on Genetics applied to Livestock Production*, 24:90-93.
- [54] Kominakis, A., Rogdakis, E., Vasiloudis, Ch. and Liaskos, O. (2000). Genetic and Environmental Sources of Variation in Milk Yield of Skopelos Dairy Goats. *Small Ruminant Res.*, 36:1-5.
- [55] Landete-Castillejos, T. and Gallego, L. (2000). Technical note: The ability of mathematical models to describe the shape of lactation curves. *J. Anim. Sci.*, 78:3010-3013.
- [56] Lee, P.C., Majluf, P. and Gordon, I.J. (1991). Growth, weaning and maternal investment from a comparative perspective. *J. Zool. (Lond.)*, 225:99-114.

- [57] Lennox, S.D., Goodall, E.A. and Mayne, C.S. (1992). A mathematical model of the lactation curve of the dairy cow to incorporate metabolizable energy intake. *The Statistician*, 41:285-293.
- [58] Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 34:1-41.
- [59] Loudon, A.S.I., McNeilly, A.S. and Milne, J.A. (1983). Nutrition and lactational control of fertility in red deer. *Nature*, 302:145-147.
- [60] MacEachern, S.N. and Berliner, L.M. (1994). Subsampling the Gibbs Sampler. *Am. Stat.*, 48:188-190.
- [61] Madalena, F.E., Martinez, M.L. and Freitas, A.F. (1979). Lactation curves of Holstein-Friesian and Holstein-Friesian  $\times$  Gir cows. *Anim. Prod.*, 29:101-107.
- [62] Marquardt, D.W. (1963). An algorithm for Least Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.*, 11:431-441.
- [63] Mavrogenis, A.P., Constantinou, A. and Louca, A. (1984). Environmental and Genetic Causes of Variation in Production Traits of Damascus Goats. 2. Goat Productivity. *Anim. Prod.*, 38:99-104.
- [64] Mepham, B. (1976). *The Secretion of Milk*, London: Edward Arnold (Publishers) Limited.
- [65] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087-1092.
- [66] Molina, J.R. and Boschini, C. (1979). Adjustment of the dairy curve of a Holstein herd with a linear modal model. *Agron. Costarric.*, 3:167-172.
- [67] Morant, S.V. and Gnanasakthy, A. (1989). A New Approach to the Mathematical Formulation of Lactation Curves. *Anim. Prod.*, 49:151-162.

- [68] Mostert, B.E., Theron, H.E. and Kanfer, F.H.J. (2001). The effect of calving season and age at calving on production traits of South African dairy cattle. *South Afr. J. Anim. Sci.*, 31:205-214.
- [69] Nelder, J.A. (1966). Inverse polynomials, a useful group of multi-factor response functions. *Biometrics*, 22:128-141.
- [70] Olori, V.E., Brotherstone, S., Hill, W.G. and McGuirk B.J. (1999). Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.*, 58:55-63.
- [71] Olori, V.E. and Galesloot, P.J.B. (1999). Projection of records in progress and calculation of 305-day yields for dairy cattle in the Republic of Ireland. *CR Delta - NRS Report*, 19 October 1999.
- [72] Papajcsik, I.A. and Boderó, J. (1988). Modelling lactation curves of Friesian cows in a subtropical climate. *Anim. Prod.*, 47:201-205.
- [73] Portolano, B., Spatafora, F., Bono, G., Margiotta, S., Todaro, M., Ortoleva, V. and Leto, G. (1996). Application of the Wood model to lactation curves of Comisana sheep. *Small Ruminant Res.*, 24:7-13.
- [74] Rabasco, A., Serradilla, J.M., Padilla, J.A. and Serrano, A. (1993). Genetic and Non-genetic Sources of Variation in Yield and Composition of Milk in Verata Goats. *Small Ruminant Res.*, 11:151-161.
- [75] Rekaya, R., Carabaño, M.J. and Toro, M.A. (2000). Bayesian Analysis of Lactation Curves of Holstein-Friesian Cattle Using a Nonlinear Model. *J. Dairy Sci.*, 83:2691-2701.
- [76] Roberts, G.O. and Polson, N.G. (1994). On the Geometric Convergence of the Gibbs Sampler. *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 56:377-384.
- [77] Rook, A.J., France, J. and Dhanoa, M.S. (1993). On the mathematical description of lactation curves. *J. Agric. Sci.*, 121:97-102.
- [78] Rowlands, G.J., Lucey, S. and Russell, A.M. (1982). A comparison of different models of the lactation curve in dairy cattle. *Anim. Prod.*, 35:135-144.



- [79] Ruvuna, F., Kogi, J.K., Taylor, J.F. and Mkuu, S.M. (1995). Lactation curves among crosses of Galla and East African with Toggenburg and Anglo Nubian goats. *Small Ruminant Res.*, 16:1-6.
- [80] SA Milch Goat Breeders' Society (2001). SA Milch Goat Breeders' Society. [Online]. Available: <http://www.sa-breeders.co.za/org/milch-goats/>, [Assessed: 8 December 2005].
- [81] Sakul, H. and Boylan, W.J. (1992). Lactation curves for several US sheep breeds. *Anim. Prod.*, 54:229-233.
- [82] Schneeberger, M. (1981). Inheritance of Lactation Curve in Swiss Brown Cattle. *J. Dairy Sci.*, 64:475-483.
- [83] Scott, T.A., Yandell, B., Zepeda, L., Shaver R.D. and Smith, T.R. (1996). Use of Lactation Curves for Analysis of Milk Production Data. *J. Dairy Sci.*, 79:1885-1894.
- [84] Serrano, M., Ugarte, E., Jurado, J.J., Pérez-Guzmán, M.D. and Legarra, A. (2001). Test day models and genetic parameters in Latxa and Manchega dairy ewes. *Livest. Prod. Sci.*, 67:253-264.
- [85] Shanks, R.D., Berger, P.J., Freeman, A.E. and Dickinson, F.N. (1981). Genetic Aspects of Lactation Curves. *J. Dairy Sci.*, 64:1852-1860.
- [86] Sherchand, L., McNew, R.W., Kellog, D.W. and Johnson, Z.B. (1995). Selection of Mathematical Model to Generate Lactation Curves Using Daily Milk Yields of Holstein Cows. *J. Dairy Sci.*, 78:2507-2513.
- [87] Sikka, L.C. (1950). A study of lactation as affected by heredity and environment. *J. Dairy Res.*, 17:231.
- [88] Singh, R.P. and Gopal, R. (1982). Lactation curve analysis of buffaloes maintained under village conditions. *Indian J. Anim. Sci.*, 52:1157-1163.
- [89] Sölkner, J. and Fuchs, W. (1987). A Comparison of Different Measures of Persistency with Special Respect to Variation of Test-day Milk Yields. *Livest. Prod. Sci.*, 16:305-319.

- [90] Stanton, T.L., Jones, L.R., Everett, R.W. and Kachman, S.D. (1992). Estimating Milk, Fat, and Protein Lactation Curves with a Test Day Model. *J. Dairy Sci.*, 75:1691-1700.
- [91] Tekerli, M., Akinci Z., Dogan, J. and Akcan A. (2000). Factors Affecting the Shape of Lactation Curves of Holstein Cows from the Balikesir Province of Turkey. *J. Dairy Sci.*, 83:1381-1386.
- [92] The South African Stud Book and Livestock Improvement Association (2001). Goat's milk: Questions and Answers. [Online]. Available: <http://studbook.co.za/Society/Milch/milk.htm>, [Assessed: 8 December 2005].
- [93] Torres-Hernandez, G. and Hohenboken, W.D. (1980). Biometric Properties of Lactations in Ewes Raising Single or Twin Lambs. *Anim. Prod.*, 30:431-436.
- [94] Tozer, P.R. and Huffaker R.G. (1999). Mathematical equations to describe lactation curves for Holstein-Friesian cows in New South Wales. *Aust. J. Agric. Res.*, 50:431-440.
- [95] Vargas, B., Koops, W.J., Herrero, M. and Van Arendonk, J.A.M. (2000). Modeling Extended Lactations of Dairy Cows. *J. Dairy Sci.*, 83:1371-1380.
- [96] Varona, L., Moreno, C., Garcia Cortes, L.A. and Altarriba, J. (1998). Bayesian Analysis of Wood's Lactation Curve for Spanish Dairy Cows. *J. Dairy Sci.*, 81:1469-1478.
- [97] Vujcic, I. and Bacic, B. (1961). [New equation of the lactation curve.] *Novi Sad. Ann. Sci. Agric.*, No. 5.
- [98] Wang, C.S., Rutledge, J.J. and Gianola, D. (1993). Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.*, 25:41-62.
- [99] Weigel, K.A., Craig, B.A., Bidwell, T.R. and Bates, D.M. (1992). Comparison of alternative diphasic lactation curve models under bovine somatotropin administration. *J. Dairy Sci.*, 75:580-589.
- [100] Williams, J.C. (1993). An empirical model for the lactation curve of white British dairy goats. *Anim. Prod.*, 57:91-97.

- [101] Wilmink, J.B.M. (1987). Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season and Stage of Lactation. *Livest. Prod. Sci.*, 16:335-348.
- [102] Whittemore, C.T. (1980). *Lactation of the dairy cow*, London: Longman Group Limited.
- [103] Wood, P.D.P. (1967). Algebraic model of the lactation curve in cattle. *Nature*, 216:164-165.
- [104] Wood, P.D.P. (1968). Factors affecting Persistency of Lactation in Cattle. *Nature*, 218:894.
- [105] Wood, P.D.P. (1969). Factors affecting the shape of the lactation curve in cattle. *Anim. Prod.*, 11:307-316.
- [106] Wood, P.D.P. (1970). A note on the repeatability of parameters of the lactation curve in cattle. *Anim. Prod.*, 12:535-538.
- [107] Wood, P.D.P. (1972). A note on seasonal fluctuation in milk production. *Anim. Prod.*, 15:89-92.
- [108] Wood, P.D.P. (1974). A note on the estimation of total lactation yield from production on a single day. *Anim. Prod.*, 19:393-396.
- [109] Wood, P.D.P. (1976). Algebraic models of the lactation curves for milk, fat and protein production, with estimates of seasonal variation. *Anim. Prod.*, 22:35-40.
- [110] Wood, P.D.P. (1977). The biometry of lactation. *J. Agric. Sci.*, 88:333-339.
- [111] Zellner, A. and Min, C. (1995). Gibbs Sampler Convergence Criteria. *J. Am. Stat. Assoc.*, 90:921-927.

# APPENDIX

## A1: An example of MATLAB<sup>®</sup> code: Finding marginal posterior distribution through Gibbs sampling in the case of complete data

The MATLAB program code below illustrates how the marginal distributions of the parameters of the three lactation traits using the adapted Wilmink model were obtained by means of the Gibbs sampler when applied to complete data, i.e. where all animals in the data were observed during all of the considered lactation cycles. In the case of the Jersey data this means that all animals in the data were observed during each of four lactation cycles. Similar code, with the necessary adjustments relevant to the lactation model applied, were also used to obtain the marginal distributions of the parameters of the three lactation traits using the Wood, general exponential and 4-parameter Morant models.

```
%%ADAPTED WILMINK: Jersey data
%(Gibbs sampling: Equal numbers of lactation cycles for all
animals)
clear
jersey
format compact
k=length(ni);      p=3;      Nij=[0;cumsum(nij)];
Y=log(D);          t=X-3;
clear D X
n=sum(nij);        Z=Z';

%STARTING VALUES
%-----
randn('seed',sum(100*clock));
load startwml7k    % File with starting values loaded
R=Rst; Sig=Sigst; B=Bst; s=sst; Phi=diag(s); w=wst;
S=[]; TMc=[]; BB=[]; RR=[]; SIG=[]; Kor=[]; W=[]; TEL=[];

for jj=1:10000    %Perform 10 000 iterations of Gibbs sampler

%SIMULATING Mic
%-----
MB1=0; MB2=0; E=0; GG=0; Mc=[]; M=[]; Mi=[]; ZZ=[]; Mw=[];
for i=1:4:4*k-1
    y1=Y(Nij(i)+1:Nij(i+1),:);
    y2=Y(Nij(i+1)+1:Nij(i+2),:);
    y3=Y(Nij(i+2)+1:Nij(i+3),:);
    y4=Y(Nij(i+3)+1:Nij(i+4),:);
    y=[y1;y2;y3;y4];
    yr=y';
    yc=yr(:);
```

```

    ti1=[t(Nij(i)+1:Nij(i+1))];
    ti2=[t(Nij(i+1)+1:Nij(i+2))];
    ti3=[t(Nij(i+2)+1:Nij(i+3))];
    ti4=[t(Nij(i+3)+1:Nij(i+4))];
    Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
    Zr=[Z(:,i) Z(:,i+1) Z(:,i+2) Z(:,i+3)];
    xi1=[ones(length(ti1),1) (ti1) exp(-w*ti1)];
    xi2=[ones(length(ti2),1) (ti2) exp(-w*ti2)];
    xi3=[ones(length(ti3),1) (ti3) exp(-w*ti3)];
    xi4=[ones(length(ti4),1) (ti4) exp(-w*ti4)];
    xi=[xi1 zeros(length(xi1),9)
        zeros(length(xi2),3) xi2 zeros(length(xi2),6)
        zeros(length(xi3),6) xi3 zeros(length(xi3),3)
        zeros(length(xi4),9) xi4];
    vmc=inv(kron(xi'*xi,inv(Phi))+kron(inv(R),inv(Sig)));
    mmc=vmc*(kron(xi',inv(Phi))*yc + kron(inv(R),inv(Sig)*B)*Zc);
    mc=sqrtm(vmc)*randn(size(mmc)) + mmc;
    mi=[mc(1:9) mc(10:18) mc(19:27) mc(28:36)];

m=[mc(1:3)';mc(4:6)';mc(7:9)';mc(10:12)';mc(13:15)';mc(16:18)';mc(19
:21)';mc(22:24)';mc(25:27)';mc(28:30)';mc(31:33)'; mc(34:36)'];

    mB1=mi*inv(R)*Zr';
    mB2=Zr*inv(R)*Zr';
    MB1=mB1+mB1;
    MB2=mB2+mB2;
    e=(y-xi*m)'*(y-xi*m);
    E=E+e;
    Mw=[Mw m];
    M=[M mi];
    Mi=[Mi;mi];
    ZZ=[ZZ;Zr];
end

%SIMULATING B
%-----
iZRZ=inv(MB2);
mB=(MB1)*iZRZ;
Bg=sqrtm(kron(iZRZ,Sig))*randn(size(mB(:)))+mB(:);
B=[Bg(1:9) Bg(10:18) Bg(19:27) Bg(28:36) Bg(37:45) Bg(46:54)
Bg(55:63) Bg(64:72) Bg(73:81) Bg(82:90) Bg(91:99) Bg(100:108)
Bg(109:117) Bg(118:126) Bg(127:135) Bg(136:144) Bg(145:153)];

%SIMULATING Phi
%-----
C=diag(E);
s=C'./sum((randn(n,3).^2));
Phi=diag(s)

%SIMULATING Sig
%-----
q=4*k;
g=3*p+1;
G=diag([0.1414822
        0.0000029
        0.3448342

```

```

0.0512994
0.0000011
0.2515051
0.0161449
0.0000003
0.0665536]);
ke=kron(speye(k),inv(R));
F=(M-B*Z)*full(ke)*(M-B*Z)';
H=inv(g*G+F);
r1=randn(3*p,q+g);
r2=sqrtm(H)*r1;
Sig=inv(r2*r2');

%SIMULATING R
%-----
G2=0;   wr=4;   %wr=delta and Wr=hyperparameter P
Wr=[1.1751  1.1243  1.1635  1.1617
     1.1243  1.2290  1.1907  1.1907
     1.1635  1.1907  1.3001  1.2395
     1.1617  1.1907  1.2395  1.3259];

for j=1:k
    g2=M(:,1+4*(j-1):4*j)-B*Z(:,1+4*(j-1):4*j);
    GG=g2'*inv(Sig)*g2;
    G2=G2+GG;
end
L=inv(wr*Wr+G2);
r3=randn(4,wr+3*p*k);
r4=sqrtm(L)*r3;
R=inv(r4*r4')

%SIMULATING w
%-----
ym=0;
for i=1:4:4*k-1
    ii=(i+3)/4;
    y1=Y(Nij(i)+1:Nij(i+1),:);
    y2=Y(Nij(i+1)+1:Nij(i+2),:);
    y3=Y(Nij(i+2)+1:Nij(i+3),:);
    y4=Y(Nij(i+3)+1:Nij(i+4),:);
    y=[y1;y2;y3;y4];
    yr=y';
    yc=yr(:);
    ti1=[t(Nij(i)+1:Nij(i+1))];
    ti2=[t(Nij(i+1)+1:Nij(i+2))];
    ti3=[t(Nij(i+2)+1:Nij(i+3))];
    ti4=[t(Nij(i+3)+1:Nij(i+4))];
    Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
    Zr=[Z(:,i) Z(:,i+1) Z(:,i+2) Z(:,i+3)];
    xi1=[ones(length(ti1),1) (ti1) exp(-w*ti1)];
    xi2=[ones(length(ti2),1) (ti2) exp(-w*ti2)];
    xi3=[ones(length(ti3),1) (ti3) exp(-w*ti3)];
    xi4=[ones(length(ti4),1) (ti4) exp(-w*ti4)];
    xi=[xi1 zeros(length(xi1),9)
        zeros(length(xi2),3) xi2 zeros(length(xi2),6)
        zeros(length(xi3),6) xi3 zeros(length(xi3),3)

```

```

        zeros(length(xi4),9) xi4];
        m=Mw(:,(3*(ii-1)+1):(3*(ii-1)+3));
        yma=(-.5*trace((y-xi*m)*inv(Phi)*(y-xi*m)'));
        ym=ym+yma;
    end
    lnfw=ym;
    a1=-20; zw=rand(1,1);
    tel=0;
    while zw>=exp(a1)
        tel=tel+1;
        w0=w*(0.75+(rand(1,1)*.5))
        ym0=0;
    for i=1:4:4*k-1
        ii=(i+3)/4;
        y1=Y(Nij(i)+1:Nij(i+1),:);
        y2=Y(Nij(i+1)+1:Nij(i+2),:);
        y3=Y(Nij(i+2)+1:Nij(i+3),:);
        y4=Y(Nij(i+3)+1:Nij(i+4),:);
        y=[y1;y2;y3;y4];
        yr=y';
        yc=yr(:);
        ti1=[t(Nij(i)+1:Nij(i+1))];
        ti2=[t(Nij(i+1)+1:Nij(i+2))];
        ti3=[t(Nij(i+2)+1:Nij(i+3))];
        ti4=[t(Nij(i+3)+1:Nij(i+4))];
        Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
        Zr=[Z(:,i) Z(:,i+1) Z(:,i+2) Z(:,i+3)];
        xi10=[ones(length(ti1),1) (ti1) exp(-w0*ti1)];
        xi20=[ones(length(ti2),1) (ti2) exp(-w0*ti2)];
        xi30=[ones(length(ti3),1) (ti3) exp(-w0*ti3)];
        xi40=[ones(length(ti4),1) (ti4) exp(-w0*ti4)];
        xi0=[xi10 zeros(length(xi10),9)
            zeros(length(xi20),3) xi20 zeros(length(xi20),6)
            zeros(length(xi30),6) xi30 zeros(length(xi30),3)
            zeros(length(xi40),9) xi40];
        m=Mw(:,(3*(ii-1)+1):(3*(ii-1)+3));
        yma=(-.5*trace((y-xi0*m)*inv(Phi)*(y-xi0*m)'));
        ym0=ym0+yma;
    end
    lnfw0=ym0;%-log(w0);
    a1=min([0 lnfw0-lnfw]);
    eal=exp(a1);
    zw=rand(1,1);
    end
    w=w0
    TEL=[TEL;tel];

    S=[S;s];
    BB=[BB;B];
    RR=[RR;R];
    SIG=[SIG;Sig];
    W=[W;w];

End

save sim17wm S BB RR SIG W jj TEL

```

## A2: An example of MATLAB<sup>®</sup> code: Determining the marginal likelihood required for calculation of Bayes factors

The MATLAB program code below illustrates how the marginal likelihood required in the calculation of Bayes factors is determined using the procedure suggested by Chib (1995) when the general exponential model is used to model lactation data for the three lactation traits. With only minor adjustments to make provision for the difference in generalised linear form, the marginal likelihood for the Wood, 4-parameter Morant or adapted Wilmink models may also be determined.

```
%For Marginal likelihood required in Bayes factors - Part 1: (Chib)
%Finding means for Mi's in general exponential model
clear
load siml7ge          %Load data from 10 000 original Gibbs simulations
B=0; R=0; Sig=0;
for ii=1:jj
    b=BB((12*(ii-1)+1):(12*(ii-1)+12),:);
    B=B+b;
    r=RR((4*(ii-1)+1):(4*(ii-1)+4),:);
    R=R+r;
    sig=SIG((12*(ii-1)+1):(12*(ii-1)+12),:);
    Sig=Sig+sig;
end

%%Means from original simulations%%
mB=B/jj; mR=R/jj; mS=mean(S); mSig=Sig/jj; mPhi=diag(mS)

jersey
k=length(ni); Nij=[0;cumsum(nij)];
Y=log(D); t=X-3;
clear D X
n=sum(nij); Z=Z';
MMi=[];
for i=1:4:4*k-1
    y1=Y(Nij(i)+1:Nij(i+1),:);
    y2=Y(Nij(i+1)+1:Nij(i+2),:);
    y3=Y(Nij(i+2)+1:Nij(i+3),:);
    y4=Y(Nij(i+3)+1:Nij(i+4),:);
    y=[y1;y2;y3;y4];
    yr=y'; yc=yr(:);
    ti1=[t(Nij(i)+1:Nij(i+1))];
    ti2=[t(Nij(i+1)+1:Nij(i+2))];
    ti3=[t(Nij(i+2)+1:Nij(i+3))];
    ti4=[t(Nij(i+3)+1:Nij(i+4))];
    xi1=[ones(length(ti1),1) log(ti1) ti1 ti1.^2];
    xi2=[ones(length(ti2),1) log(ti2) ti2 ti2.^2];
    xi3=[ones(length(ti3),1) log(ti3) ti3 ti3.^2];
    xi4=[ones(length(ti4),1) log(ti4) ti4 ti4.^2];
    xi=[xi1 zeros(length(xi1),12)
```



```

        zeros(length(xi2),4) xi2 zeros(length(xi2),8)
        zeros(length(xi3),8) xi3 zeros(length(xi3),4)
        zeros(length(xi4),12) xi4];
    Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
    Zr=[Z(:,i) Z(:,i+1) Z(:,i+2) Z(:,i+3)];
    Mvar=inv(kron(xi'*xi,inv(mPhi))+kron(inv(mR),inv(mSig)));
    Mmi=Mvar*(kron(xi',inv(mPhi))*yc +
kron(inv(mR),inv(mSig)*mB)*Zc);
    MMi=[MMi Mmi];           %orde: 48 x 1141
    mm=[Mmi(1:12) Mmi(13:24) Mmi(25:36) Mmi(37:48)];
    MM=[MM mm];           %order: 12 x 4564
end
save bf1.mat MMi MM mB mR mSig mS mPhi

```

```

% For Marginal likelihood required in Bayes factors:
% General exponential model - Part 2: (Chib)
clear
load sim17gex
load bf1
jersey
k=length(ni); Nij=[0;cumsum(nij)];
Y=log(D); t=X-3;
qi=4;           %number of consecutive cycle per animal
p=4;           %number of parameters in model
kof=17;        %number of cofactors in model
s=3;           %number of traits investigated (Prod, Pt & BV)
clear D X
n=sum(nij);
Z=Z';

randn('seed',sum(100*clock));

LBSTER=[];    LSigSTER=[];
BBt=[]; RRb=[]; SIGb=[]; SIGt=[]; RRsigs=[];
R=mR; Sig=mSig;

jj=10000      %number of Gibbs iterations

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% B* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
logMPBst=0;

for j=1:jj

%Draw B: mean of R and Sig from simulations as start & fixed M*%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
MB1=0; MB2=0;
for i=1:k
    Mi=MMi(:,i);           %fixed Mi*
    mi=[Mi(1:12) Mi(13:24) Mi(25:36) Mi(37:48)];
    ii=4*i;
    Zr=[Z(:,ii-3) Z(:,ii-2) Z(:,ii-1) Z(:,ii)];
    mB1=mi*inv(R)*Zr';
    mB2=Zr*inv(R)*Zr';
    MB1=MB1+mB1;

```

```

        MB2=MB2+mB2;
    end
    iZRZ=inv(MB2);
    gmB=(MB1)*iZRZ;
    Bg=real(sqrtm(kron(iZRZ,Sig)))*randn(size(gmB(:)))+gmB(:);
    B=[Bg(1:12) Bg(13:24) Bg(25:36) Bg(37:48) Bg(49:60) Bg(61:72)
    Bg(73:84) Bg(85:96) Bg(97:108) Bg(109:120) Bg(121:132) Bg(133:144)
    Bg(145:156) Bg(157:168) Bg(169:180) Bg(181:192) Bg(193:204)];
    BBT=[BBt;B];

    %%Draw R: mean(Sig) from simulations as start, B as above & fixed M*%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    G2=0; w=4;
    W=[1.183312777    0.789010185    0.826199702    0.806821427
        0.789010185    1.246930017    0.832558042    0.800748753
        0.826199702    0.832558043    1.204569336    0.854740445
        0.806821427    0.800748753    0.854740445    1.239184366];
    for i=1:k
        Mi=MMi(:,i); %fixed Mi*
        mi=[Mi(1:12) Mi(13:24) Mi(25:36) Mi(37:48)];
        g2=mi-B*Z(:,1+4*(i-1):4*i);
        GG=g2'*inv(Sig)*g2;
        G2=G2+GG;
    end
    L=inv(w*W+G2);
    r3=randn(4,w+3*p*k);
    r4=sqrtm(L)*r3;
    R=inv(r4*r4');
    RRb=[RRb;R];

    %%Draw Sig with B & R drawn and fixed M*%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    q=4*k; g=3*p+1;
    G=diag([ 1.88732364997719
            0.21803441972297
            0.00010366578937
            0.0000047229
            2.19366342439675
            0.25209117888476
            0.00010727024472
            0.0000043419
            0.42040640140500
            0.04872026647853
            0.00002081770497
            0.0000008403]);
    ke=kron(speye(k),inv(R));
    F=(MM-B*Z)*full(ke)*(MM-B*Z)';
    H=inv(g*G+F);
    r1=randn(3*p,q+g);
    r2=sqrtm(H)*r1;
    SigB=inv(r2*r2');
    SIGb=[SIGb;SigB];

    MB1=0; MB2=0;
    for i=1:k
        Mi=MMi(:,i); %fixed Mi*

```

```

mi=[Mi(1:12) Mi(13:24) Mi(25:36) Mi(37:48)];
ii=4*i;
Zr=[Z(:,ii-3) Z(:,ii-2) Z(:,ii-1) Z(:,ii)];
mB1=mi*inv(R)*Zr';
mB2=Zr*inv(R)*Zr';
MB1=MB1+mB1;
MB2=MB2+mB2;
end

%%Determine marginal posterior of B* from simulation results.%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
iZRZ=inv(MB2);
gmB=(MB1)*iZRZ;
cons=(-.5*(3*p)*kof)*log(2*pi);
logMPb=cons-.5*(3*p)*log(det(iZRZ))- .5*kof*log(det(SigB))+((trace(-
.5*inv(SigB)*(mB-gmB)*inv(iZRZ)*(mB-gmB)'));
logMPBst=logMPBst+logMPb;
end

logBSTER=logMPBst/jj;
LBSTER=[LBSTER;logBSTER];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Sig* %%%%%%%%%
R=mR;

logMPSigst=0;
q=4*k;
g=3*p+1;
Gamln=[];
for ii=1:(3*p)
    gamln=gammaln(.5*(q+g+1-ii));
    Gamln=[Gamln;gamln];
end
consj=((.25*(3*p)*(3*p-1))*log(pi)+sum(Gamln));

for j=1:jj

%%Draw Sig with mean(R) from simulations as start and fixed M* &
B*%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
G=diag([1.88732364997719
        0.21803441972297
        0.00010366578937
        0.0000047229
        2.19366342439675
        0.25209117888476
        0.00010727024472
        0.0000043419
        0.42040640140500
        0.04872026647853
        0.00002081770497
        0.0000008403]);
ke=kron(speye(k),inv(R));
F=(MM-mB*Z)*full(ke)*(MM-mB*Z)';          %mB=fixed B*; MM=fixed M*
H=inv(g*G+F);

```

```

r1=randn(3*p,q+g);
r2=sqrtm(H)*r1;
Sig=inv(r2*r2');
SIGt=[SIGt;Sig];

%%Draw R with Sig drawn and fixed M* & B*%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
G2=0; w=4;
W=[1.183312777    0.789010185    0.826199702    0.806821427
    0.789010185    1.246930017    0.832558042    0.800748753
    0.826199702    0.832558043    1.204569336    0.854740445
    0.806821427    0.800748753    0.854740445    1.239184366];
for i=1:k
    Mi=MMi(:,i);
    mi=[Mi(1:12) Mi(13:24) Mi(25:36) Mi(37:48)];
    g2=mi-mB*Z(:,1+4*(i-1):4*i);
    GG=g2'*inv(Sig)*g2;
    G2=G2+GG;
end
L=inv(w*W+G2);
r3=randn(4,w+3*p*k);
r4=sqrtm(L)*r3;
R=inv(r4*r4');
RRsig=[RRsig;R];

%%Determine marginal posterior of Sig* from simulation results.%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
ke=kron(speye(k),inv(R));
F=(MM-mB*Z)*full(ke)*(MM-mB*Z)';          %mB=fixed B*;  MM=fixed M*
H=inv(g*G+F);
logMPSigi=-consj+(-.5*(q+g))*log(det(2*H))+trace(-
    .5*inv(H)*inv(mSig))+(.5*((q+g)-(3*p)-1))*log(det(inv(mSig)));
logMPSigst=logMPSigst+logMPSigi;
end

logSigSTER=logMPSigst/jj;
LSigSTER=[LSigSTER;logSigSTER];

    save resbf2 LSigSTER LBSTER Bbt SIGt RRb SIGb RRsig jj

clear
load resbf2
p=4;
Rb=RRb((jj*4)-4+1:jj*4,:);
Sigb=SIGb((jj*(3*p))-(3*p)+1:jj*(3*p),:);
Rsig=RRsig((jj*4)-4+1:jj*4,:);
save bayesfst Rb Sigb Rsig

%%For Marginal likelihood required in Bayes factors:
%%General exponential model - Part 3 (Chib)
clear
load sim17gex
load bf1

```

```

load bayesfst
jersey
k=length(ni); Nij=[0;cumsum(nij)];
Y=log(D); t=X-3;
qi=4; p=4; kof=17; s=3;
clear D X
n=sum(nij); Z=Z';

randn('seed',sum(100*clock));

LBSTER=[]; LSigSTER=[]; Bbt=[]; RRb=[]; SIGb=[]; SIGt=[]; RRsigt=[];
R=mR; Sig=mSig;

jj=10000 %number of simulations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% M*
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Find marginal posterior of Mi* from simulation results (log vorm)
%%Use as good estimate of Mi* the value from the mean of the normal
%%distribution from which it originated, calculate the mean values
of
%%B, R, Phi and Sig from the 10000 simulations.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
logMPmst=0;

for j=1:jj
    j
    phi=diag(S(j,:));
    B=BB(((3*p)*(j-1)+1):((3*p)*(j-1)+(3*p)),:);
    R=RR((4*(j-1)+1):(4*(j-1)+4),:);
    Sig=SIG(((3*p)*(j-1)+1):((3*p)*(j-1)+(3*p)),:);
    Em=0; Ps=0;
    for i=1:k
        Mi=MMi(:,i);
        y1=Y(Nij(((i-1)*4)+1)+1:Nij((((i-1)*4)+1)+1),:);
        y2=Y(Nij(((i-1)*4)+2)+1:Nij((((i-1)*4)+1)+2),:);
        y3=Y(Nij(((i-1)*4)+3)+1:Nij((((i-1)*4)+1)+3),:);
        y4=Y(Nij(((i-1)*4)+4)+1:Nij((((i-1)*4)+1)+4),:);
        y=[y1;y2;y3;y4];
        yr=y'; yc=yr(:);
        ti1=[t(Nij(((i-1)*4)+1)+1:Nij((((i-1)*4)+1)+1))];
        ti2=[t(Nij(((i-1)*4)+2)+1:Nij((((i-1)*4)+1)+2))];
        ti3=[t(Nij(((i-1)*4)+3)+1:Nij((((i-1)*4)+1)+3))];
        ti4=[t(Nij(((i-1)*4)+4)+1:Nij((((i-1)*4)+1)+4))];
        xi1=[ones(length(ti1),1) log(ti1) ti1 ti1.^2];
        xi2=[ones(length(ti2),1) log(ti2) ti2 ti2.^2];
        xi3=[ones(length(ti3),1) log(ti3) ti3 ti3.^2];
        xi4=[ones(length(ti4),1) log(ti4) ti4 ti4.^2];
        xi=[xi1 zeros(length(xi1),12)
            zeros(length(xi2),4) xi2 zeros(length(xi2),8)
            zeros(length(xi3),8) xi3 zeros(length(xi3),4)
            zeros(length(xi4),12) xi4];
        Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
        Mvar=inv(kron(xi'*xi,inv(phi))+kron(inv(R),inv(Sig)));
        Mmu=Mvar*(kron(xi',inv(phi))*yc + kron(inv(R),inv(Sig)*B)*Zc);
        em=(Mi-Mmu) '*inv(Mvar)*(Mi-Mmu);
        Em=Em+em;
    end
end

```

```

        ps=(-1/2)*(log(det(Mvar)));
        Ps=Ps+ps;
    end
    con=(-(3*p)*qi*k/2)*log(2*pi);
    logMPmsti=con+(trace((-1/2)*Em))+Ps;
    logMPmst=logMPmst+logMPmsti;
end
logMSTER=logMPmst/jj;

save bf3 logMSTER

```

```

% For Marginal likelihood required in Bayes factors - Part 4:
(Chib) % General exponential model

```

```

clear
load sim17gex
load bf1
jersey
k=length(ni); Nij=[0;cumsum(nij)];
Y=log(D); t=X-3;
qi=4; p=4; kof=17; s=3; (Prod, Pt & BV)
clear D X
n=sum(nij); Z=Z';

```

```

randn('seed',sum(100*clock));

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Y|M*,Phi*
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Determin p(y|Mi*,Phi*) (in log form)!!!

```

```

Ey=0; Ps=0;
for i=1:k
    Mi=MMi(:,i);
    y1=Y(Nij(((i-1)*4)+1)+1:Nij(((i-1)*4)+1)+1,:);
    y2=Y(Nij(((i-1)*4)+2)+1:Nij(((i-1)*4)+1)+2,:);
    y3=Y(Nij(((i-1)*4)+3)+1:Nij(((i-1)*4)+1)+3,:);
    y4=Y(Nij(((i-1)*4)+4)+1:Nij(((i-1)*4)+1)+4,:);
    y=[y1;y2;y3;y4];
    yr=y'; yc=yr(:);
    ti1=[t(Nij(((i-1)*4)+1)+1:Nij(((i-1)*4)+1)+1)];
    ti2=[t(Nij(((i-1)*4)+2)+1:Nij(((i-1)*4)+1)+2)];
    ti3=[t(Nij(((i-1)*4)+3)+1:Nij(((i-1)*4)+1)+3)];
    ti4=[t(Nij(((i-1)*4)+4)+1:Nij(((i-1)*4)+1)+4)];
    xi1=[ones(length(ti1),1) log(ti1) ti1 ti1.^2];
    xi2=[ones(length(ti2),1) log(ti2) ti2 ti2.^2];
    xi3=[ones(length(ti3),1) log(ti3) ti3 ti3.^2];
    xi4=[ones(length(ti4),1) log(ti4) ti4 ti4.^2];
    xi=[xi1 zeros(length(xi1),12)
        zeros(length(xi2),4) xi2 zeros(length(xi2),8)
        zeros(length(xi3),8) xi3 zeros(length(xi3),4)
        zeros(length(xi4),12) xi4];
    Yvar=(kron(mPhi,eye(ni(i))));
    Ymu=(kron(eye(3),xi))*Mi;
    ey=(yc-Ymu) '*inv(Yvar)*(yc-Ymu);
    Ey=Ey+ey;
end

```

```

        ps=(-3*ni(i)/2)*log(2*pi)+(-ni(i)/2)*log(det(mPhi));
        Ps=Ps+ps;
    end
    logpY=Ps+(trace((-1/2)*Ey))
save bf4 logpY
%General exponential priors evaluated at *-values {for Bayesfactors}
clear
load bf1
jersey
k=length(ni); Nij=[0;cumsum(nij)];
Y=log(D); t=X-3;
qi=4; p=4; kof=17; s=3;
clear D X
n=sum(nij); Z=Z';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Prior of M* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Em=0;
for i=1:k
    Mi=MMi(:,i);
    Zc=[Z(:,i);Z(:,i+1);Z(:,i+2);Z(:,i+3)];
    Mvar=kron(mR,mSig);
    Mmu=kron(eye(qi),mB)*Zc;
    em=(Mi-Mmu)'*inv(Mvar)*(Mi-Mmu);
    Em=Em+em;
end
con=(-p*s*qi*k/2)*log(2*pi)+(-qi*k/2)*log(det(mSig))+(-
p*s*k/2)*log(det(mR));
logPM=con+(trace((-1/2)*Em));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Prior of Phi* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
logPPhi=log(inv(prod(diag(mPhi))));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Prior of B* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
logPB=log(1);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Prior of R* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
w=4;
W=[1.183312777    0.789010185    0.826199702    0.806821427
    0.789010185    1.246930017    0.832558042    0.800748753
    0.826199702    0.832558043    1.204569336    0.854740445
    0.806821427    0.800748753    0.854740445    1.239184366];

Gamlnr=[];
for ii=1:w
    gamlnr=gammaln(.5*(w+1-ii));
    Gamlnr=[Gamlnr;gamlnr];
end
consjr=(.25*qi*(qi-1))*log(pi)+sum(Gamlnr);

logPR=-consjr+(-.5*w)*log(det(2*inv(w*W)))+(trace(-
.5*(w*W)*inv(mR)))+(0.5*(w-qi-1))*log(det(inv(mR)))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Prior of Sig* %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
g=3*p+1;

```

```

G=diag([ 1.88732364997719
        0.21803441972297
        0.00010366578937
        0.0000047229
        2.19366342439675
        0.25209117888476
        0.00010727024472
        0.0000043419
        0.42040640140500
        0.04872026647853
        0.00002081770497
        0.0000008403]);

GamlnS=[];
for ii=1:(3*p)
    gamlnS=gammaln(.5*(g+1-ii));
    GamlnS=[GamlnS;gamlnS];
end
consjs=((.25*(3*p)*(3*p-1))*log(pi)+sum(GamlnS));

logPS=-consjs-0.5*g*log(det(2*inv(g*G)))+(trace(-
.5*(g*G)*inv(mSig))+0.5*(g-(3*p)-1)*log(det(inv(mSig))))

save prior logPM logPPhi logPB logPR logPS

%% Determine Marginal likelihood for general exponential model %%
clear
load ster          %contains all *-matrices generated
load bfl
load prior
jersey
k=length(ni); p=4; s=3; Z=Z'; ZiRZ=0;
for i=1:k
    ii=4*i;
    Zr=[Z(:,ii-3) Z(:,ii-2) Z(:,ii-1) Z(:,ii)];
    ziRz=Zr*inv(mR)*Zr';
    ZiRZ=ZiRZ+ziRz;
end
m=17;                %number of cofactors
cons=(-p*s*m/2)*log(2*pi)+(-m/2)*log(det(mSig))+(-
p*s/2)*log(det(ZiRZ))
num=logpY+logPM+logPPhi+logPB+logPS+logPR+cons
den=logMSTER+logPhiSTER+LBSTER+LSigSTER+logRSTER
ml=num-den
mlgexp=ml

save mlg mlgexp

```



### **A3: An example of MATLAB® code: Finding marginal posterior distribution through Gibbs sampling in the case of incomplete data**

The MATLAB program code below illustrates how the marginal distributions of the parameters for the three lactation traits using the Wood model were obtained by means of the Gibbs sampler when applied to incomplete data, i.e. where *not* all animals in the data were observed during all of the considered lactation cycles. For the Saanen data modelled here, a number of does were observed during both of the possible two lactation cycles, while others were observed only once which could either have been during the first or the second of the two lactation cycles.

```
%% WOOD MODEL: Saanen data

%%(Gibbs sampling: Unequal numbers of lactation cycles for animals)

clear
melkbok
format compact
k1=length(ni1); k2=length(ni2); k=k1+k2; p=3;
af=sum(nij1); bf=sum(nij2);
nij=[nij1;nij2];
Nij1=[0;cumsum(nij1)]; Nij2=[0;cumsum(nij2)]; Nij=[0;cumsum(nij)];
Y1=log(D1); Y2=log(D2);
X1=X1-3; X2=X2-3;
t1=X1; t2=X2;
clear D1 D2 X1 X2
n1=sum(nij1); n2=sum(nij2); n=n1+n2;
%Y=[Y1;Y2];
Z1(:,6)=Z1(:,6); Z2(:,6)=Z2(:,6);
Z1=Z1'; Z2=Z2';
II1=find(Z1(2,:)==0); %Finds doe with only a cycle in year 1
II2=find(Z1(2,:)==1); %Finds doe with only a cycle in year 2

%STARTING VALUES
%-----
randn('seed',sum(100*clock));
load start2
s=sst(1:p); R=Rst; Sig=Sigst; B=Bst;
Phi=diag(s);

S=[]; TM1c=[]; TM2c=[]; BB=[]; RO=[]; SIG=[]; Kor=[];

for jj=1:12000

%SIMULATING Mic
%-----
MB11=0; MB12=0; E1=0; M1c=[]; M1=[]; Rp=[];
```

```

for i=1:k1
    if Z1(2,i)==0
        Rr=R(1,1);
    else
        Rr=R(2,2);
    end
    Rp=[Rp;Rr];
    y1=Y1(Nij1(i)+1:Nij1(i+1),:);
    ylc=y1(:);
    til=[t1(Nij1(i)+1:Nij1(i+1))];
    xil=[ones(length(til),1) log(til) til];
    vmlc=inv(kron(inv(Phi),xil'*xil)+inv(Rr*Sig));
    mmlc=vmlc*(kron(inv(Phi),xil')*ylc + inv(Rr*Sig)*B*Z1(:,i));
    m1c=sqrtm(vmlc)*randn(size(mmlc)) + mmlc;
    m1=[m1c(1:3) m1c(4:6) m1c(7:9)];

    mB11=m1c*Z1(:,i)'/Rr; mB12=Z1(:,i)*Z1(:,i)'/Rr;
    MB11=MB11+mB11; MB12=MB12+mB12;

    e1=(y1-xil*m1)'*(y1-xil*m1);
    E1=E1+e1;

    M1c=[M1c;m1c]; M1=[M1 m1c];
end
I1=2; I2=35; I3=62; I4=76; I5=162; I6=194;
M1cc=[M1c(9*(I1-1)+1:9*I1);M1c(9*(I2-1)+1:9*I2);M1c(9*(I3-
1)+1:9*I3);M1c(9*(I4-1)+1:9*I4);M1c(9*(I5-1)+1:9*I5);M1c(9*(I6-
1)+1:9*I6)];
MB21=0; MB22=0; E2=0; G2=0; M2c=[]; M2=[]; Mi2=[]; Z222=[];
for i=1:2:2*k2-1
    y21=Y2(Nij2(i)+1:Nij2(i+1),:);
    y22=Y2(Nij2(i+1)+1:Nij2(i+2),:);
    y2=[y21;y22];
    y2c=[y21(:);y22(:)];
    ti21=[t2(Nij2(i)+1:Nij2(i+1))];
    ti22=[t2(Nij2(i+1)+1:Nij2(i+2))];
    Z2c=[Z2(:,i);Z2(:,i+1)];
    Z22=[Z2(:,i) Z2(:,i+1)];
    xi21=[ones(length(ti21),1) log(ti21) ti21];
    xi22=[ones(length(ti22),1) log(ti22) ti22];
    xi2=[xi21 zeros(length(xi21),3)
        zeros(length(xi22),3) xi22];
    vm2c=inv(kron(inv(Phi),xi2'*xi2)+kron(inv(R),inv(Sig)));
    mm2c=vm2c*(kron(inv(Phi),xi2')*y2c +
kron(inv(R),inv(Sig)*B)*Z2c);
    m2c=sqrtm(vm2c)*randn(size(mm2c)) + mm2c;
    mi2=[m2c(1:9) m2c(10:18)];
    m2=[m2c(1:3) m2c(4:6) m2c(7:9)
        m2c(10:12) m2c(13:15) m2c(16:18)];

    mB21=mi2*inv(R)*Z22'; mB22=Z22*inv(R)*Z22';
    MB21=MB21+mB21; MB22=MB22+mB22;

    e2=(y2-xi2*m2)'*(y2-xi2*m2);
    E2=E2+e2;

```

```

    g2=(mi2-B*Z22)'*inv(Sig)*(mi2-B*Z22);
    G2=G2+g2;
    M2c=[M2c;m2c]; M2=[M2 mi2]; Mi2=[Mi2;mi2]; Z222=[Z222;Z22];
end
J1=246; J2=1; J3=70; J4=150; J5=185; J6=144;
M2cc=[M2c(18*(J1-1)+1:18*J1);M2c(18*(J2-1)+1:18*J2);M2c(18*(J3-
1)+1:18*J3);M2c(18*(J4-1)+1:18*J4);M2c(18*(J5-
1)+1:18*J5);M2c(18*(J6-1)+1:18*J6)];

%SIMULATING B
%-----
iZRZ=inv(MB12+MB22);
mB=(MB11+MB21)*iZRZ;
%vB=kron(iZRZ,Sig);
B=sqrtm(Sig)*randn(size(mB))*sqrtm(iZRZ)+mB;

%SIMULATING Phi
%-----
E=E1+E2;
C=diag(E);
s=C'./sum((randn(n,p).^2));
Phi=diag(s);

%SIMULATING Sig
%-----
iRp=Rp.^(-1);
q=k1+2*k2; g=3*p+1; G=(eye(3*p))/10;
F1=(M1-B*Z1)*diag(iRp)*(M1-B*Z1)';
F2=(M2-B*Z2)*kron(eye(k2),inv(R))*(M2-B*Z2)';
F=F1+F2;
H=inv(g*G + F);
r1=randn(3*p,q+g);
r2=sqrtm(H)*r1;
Sig=inv(r2*r2');
[jj;R(1,1);Sig(1,1);rank(iZRZ)]

%SIMULATING R - Metropolis-Hastings%%
%-----
format compact
M11=M1(:,II1); M12=M1(:,II2);
n1r=length(M11); n2r=length(M12); n12r=0.5*length(M2);
r11=R(1,1); r12=R(1,2); r22=R(2,2);
d=4; P=[0.002 0;0 0.003]/10;
R11=[]; R12=[]; R22=[];
mb11=M11-B*Z1(:,II1); x11=0;
for j=1:n1r
    fx11=mb11(:,j)'\*inv(Sig)*mb11(:,j);
    x11=x11+fx11;
end
mb12=M12-B*Z1(:,II2); x12=0;
for j=1:n2r
    fx12=mb12(:,j)'\*inv(Sig)*mb12(:,j);
    x12=x12+fx12;
end
mb2=M2-B*Z2; x2=zeros(2,2);
for j=1:2:2*n12r-1

```

```

        fx2=mb2(:,j:j+1)'*inv(Sig)*mb2(:,j:j+1);
        x2=x2+fx2;
    end
    for i=1:10
        R=[r11 r12;r12 r22];
        lnfr111=-((9*n12r+d-3)/2)*log(det(R))-0.5*trace(inv(R)*(d*P + x2));
        lnfr112=-9*n1r/2*log(r11)-0.5*x11/r11;
        lnfr11=lnfr111+lnfr112;
        l11=(R(1,2)^2)/R(2,2);
        a1=-100; z2=rand(1,1);
        while z2>=exp(a1);
            z1=rand(1,1)-0.5;
            r110=(r11-l11)*exp(z1)+l11;
            R1=[r110 r12;r12 r22];
            lnfr1110=-((9*n12r+d-3)/2)*log(det(R1))-0.5*trace(inv(R1)*(d*P +
            x2));
            lnfr1120=-9*n1r/2*log(r110)-0.5*x11/r110;
            lnfr110=lnfr1110+lnfr1120;
            a1=min([0 lnfr110-lnfr11]);
            z2=rand(1,1);
        end
        r11=r110;
        %R11=[R11;r11];

        R=[r11 r12;r12 r22];
        lnfr221=-((9*n12r+d-3)/2)*log(det(R))-0.5*trace(inv(R)*(d*P + x2));
        lnfr222=-9*n2r/2*log(r22)-0.5*x12/r22;
        lnfr22=lnfr221+lnfr222;
        l22=(R(1,2)^2)/R(1,1);
        a3=-100; z5=rand(1,1);
        while z5>=exp(a3);
            z4=rand(1,1)-0.5;
            r220=(r22-l22)*exp(z4)+l22;
            R3=[r11 r12;r12 r220];
            lnfr2210=-((9*n12r+d-3)/2)*log(det(R3))-0.5*trace(inv(R3)*(d*P +
            x2));
            lnfr2220=-9*n2r/2*log(r220)-0.5*x12/r220;
            lnfr220=lnfr2210+lnfr2220;
            a3=min([0 lnfr220-lnfr22]);
            z5=rand(1,1);
        end
        r22=r220;

        R=[r11 r12;r12 r22];
        lnfr12=-((9*n12r+d-3)/2)*log(det(R))-0.5*trace(inv(R)*(d*P + x2));
        l12=sqrt(r11*r22);
        a2=-100; z3=rand(1,1);
        while z3>=exp(a2);
            r120=l12*randn(1,1)/3+r12;
            while r120<=-l12 | r120>=l12
                r120=l12*randn(1,1)/3+r12;
            end
            R2=[r11 r120;r120 r22];
            lnfr120=-((9*n12r+d-3)/2)*log(det(R2))-0.5*trace(inv(R2)*(d*P +
            x2));
            a2=min([0 lnfr120-lnfr12]);

```

```

z3=rand(1,1);
end
r12=r120;
%R12=[R12;r12];
end

R=[r11 r12;r12 r22];

Ro=[R(1,:) R(2,2)]; kor=R(1,2)/sqrt(R(1,1)*R(2,2));
S=[S;s];TM1c=[TM1c M1cc]; TM2c=[TM2c M2cc];
BB=[BB;B]; RO=[RO;Ro]; Kor=[Kor;kor];
SIG=[SIG;Sig];
save final Kor S TM1c TM2c BB RO SIG Y1 Y2 t1 t2 nij1 nij2 I1 I2 I3
I4 I5 I6 J1 J2 J3 J4 J5 J6 jj

```

# SUMMARY

This thesis was written with the aim of modelling the lactation process in dairy cows and goats by applying a hierarchical Bayesian approach. Information on cofactors that could possibly affect lactation is included in the model through a novel approach using covariates. Posterior distributions of quantities of interest are obtained by means of the Markov chain Monte Carlo methods. Prediction of future lactation cycle(s) is also performed.

In chapter one lactation is defined, its characteristics considered, the factors that could possibly influence lactation mentioned, and the reasons for modelling lactation explained.

Chapter two provides a historical perspective to lactation models, considers typical lactation curve shapes and curves fitted to the lactation composition traits fat and protein of milk. Attention is also paid to persistency of lactation.

Chapter three considers alternative methods of obtaining total yield and producing Standard Lactation Curves (SLAC's). Attention is paid to methods used in fitting lactation curves and the assumptions about the errors.

In chapter four the generalised Bayesian model approach used to simultaneously model more than one lactation trait, while also incorporating information on cofactors that could possibly influence lactation, is developed. Special attention is paid not only to the model for complete data, but also how modelling is adjusted to make provision for cases where not all lactation cycles have been observed for all animals, also referred to as incomplete data. The use of the Gibbs sampler and the Metropolis-Hastings algorithm in determining marginal posterior distributions of model parameters and quantities that are functions of such parameters are also discussed. Prediction of future lactation cycles using the model is also considered.

In chapter five the Bayesian approach together with the Wood model, applied to 4564 lactation cycles of 1141 Jersey cows, is used to illustrate the approach to modelling and prediction of milk yield, percentage of fat and percentage of protein in milk composition in the case of complete data. The incorporation of cofactor information through the use of the covariate matrix is also considered in greater detail. The results from the Gibbs sampler are evaluated and convergence there-of investigated. Attention is also paid to the expected lactation curve characteristics as defined by Wood, as well as obtaining the expected lactation

curve of one of the levels of a cofactor when the influence of the other cofactors on the lactation curve has been eliminated.

Chapter six considers the use of the Bayesian approach together with the general exponential and 4-parameter Morant model, as well as an adaptation of a model suggested by Wilmink, in modelling and predicting milk yield, fat content and protein content of milk for the Jersey data.

In chapter seven a diagnostic comparison by means of Bayes factors of the results from the four models in the preceding two chapters, when used together with the Bayesian approach, is performed. As a result the adapted form of the Wilmink model fared best of the models considered!

Chapter eight illustrates the use of the Bayesian approach, together with the four lactation models considered in this study, to predict the lactation traits for animals similar to, but not contained in the data used to develop the respective models.

In chapter nine the Bayesian approach together with the Wood model, applied to 755 lactation cycles of 493 Saanen does collected during either or both of two consecutive years, is used to illustrate the approach to modelling and predicting milk yield, percentage of fat and percentage of protein in milk in the case of incomplete data.

Chapter ten provides a summary of the results and a perspective of the contribution of this research to lactation modelling.

**KEYWORDS:** 4-parameter Morant model, adapted Wilmink model, Bayes factors, Bayesian modelling, covariate, general exponential model, lactation curves, Markov chain Monte Carlo, Standard Lactation Curves, Wood model

## OPSOMMING

Hierdie tesis is geskryf met die doel om die laktasieproses in suiwelkoeie en suiwelbokke te modelleer deur 'n hierargiese Bayesbenadering toe te pas. Inligting aangaande kofaktore wat maandelik laktasie kan beïnvloed, is in die model ingesluit deur middel van 'n unieke benadering wat van koveranderlikes gebruik maak. Posteriorverdelings van hoeveelhede van belang word deur middel van die Markovketting Monte Carlo metodes verkry. Voorspelling van toekomstige laktasiesiklus(se) is ook uitgevoer.

In hoofstuk een word laktasie gedefinieer, die eienskappe daarvan beskou, die faktore wat maandelik laktasie mag beïnvloed genoem, en die redes vir die modellering van laktasie verduidelik.

Hoofstuk twee lewer 'n historiese perspektief tot laktasiemodelle, beskou tipiese laktasiekurwe vorms, asook kurwes gepas aan die laktasiesamestellingskenmerke vet en proteïen van melk. Aandag word ook aan die volhoubaarheid van laktasie geskenk.

Hoofstuk drie beskou alternatiewe metodes om totale opbrengs te verkry en Standaard Laktasiekurwes (SLAC's) voort te bring. Aandag word geskenk aan metodes wat gebruik word in die passing van laktasiekurwes en die aannames aangaande die foute.

In hoofstuk vier word die veralgemeende Bayesmodelbenadering ontwikkel om meer as een laktasiekenmerk gelyktydig te modelleer, terwyl inligting aangaande kofaktore wat maandelik laktasie kan beïnvloed ook ingesluit word. Spesiale aandag word nie net aan die model vir volledige data geskenk nie, maar ook aan hoe modellering aangepas moet word om voorsiening te maak vir gevalle waar nie al die laktasiesiklusse vir alle diere waargeneem is nie, wat ook na verwys word as onvolledige data. Die gebruik van Gibbssteekproefneming en die Metropolis-Hastings algoritme in die bepaling van posterior randverdelings van die model parameters en hoeveelhede wat funksies van sulke parameters is, word ook bespreek. Voorspelling van toekomstige laktasiesiklusse deur die model te gebruik word ook beskou.

In hoofstuk vyf word die Bayesbenadering saam met die Woodmodel, toegepas op 4564 laktasiesiklusse van 1141 Jerseykoeie, ter illustrasie van die benadering tot modellering en voorspelling van melkopbrengs, persentasie vet en persentasie proteïen in die samestelling van melk in die geval van volledige data gebruik. Die insluiting van kofaktorinligting deur



die gebruik van die matriks van koveranderlikes word ook in meer besonderhede beskou. Die resultaat vanaf Gibbssteekproefneming word evalueer en die konvergensie daarvan ondersoek. Aandag word ook geskenk aan die verwagte laktasiekurwe eienskappe soos gedefinieer deur Wood, asook die bepaling van verwagte laktasiekurwes vir een van die vlakke van 'n kofaktor indien die invloed van die ander kofaktore op die laktasiekurwe uitgeskakel word.

Hoofstuk ses beskou die gebruik van die Bayesbenadering saam met die veralgemeende eksponensiaal en 4parameter Morant model, asook 'n aanpassing van 'n model wat deur Wilmink voorgestel is, in die modellering en voorspelling van melkopbrengs, asook die samestellingskenmerke vet en proteïene in melk vir die Jerseydata.

In hoofstuk sewe word 'n diagnostiese vergelyking deur middel van Bayesfaktore uitgevoer op die resultate van die vier modelle in die voorafgaande twee hoofstukke wanneer dit saam met die Bayesbenadering gebruik word. As resultaat hiervan het die aangepaste vorm van die Wilmink model die beste van die modelle wat oorweeg is, gevaar!

Hoofstuk agt illustreer die gebruik van die Bayesbenadering, saam met die vier laktasiemodelle onder beskouing in hierdie studie, om die laktasiekenmerke te voorspel van diere soortgelyk aan, maar nie ingesluit in die data wat gebruik is in die ontwikkeling van die onderskeie modelle nie.

In hoofstuk nege word die Bayesbenadering saam met die Wood model toegepas op 755 laktasiesiklusse van 493 Saanenooie om die benadering tot modellering en voorspelling van melkopbrengs, persentasie vet en persentasie proteïene in die samestelling van melk in die geval van onvolledige data te illustreer.

Hoofstuk tien lewer 'n opsomming van die resultate en 'n perspektief van die bydrae van hierdie navorsing tot laktasiemodellering.