This is to certify that the

dissertation entitled

EVALUATION OF METHODS FOR ESTIMATING 305-DAY LACTATION YIELD IN DAIRY CATTLE

presented by

PETER MALACHI SAAMA

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Animal Science

Major professor

Date January 6, 1995

EVALUATION OF METHODS FOR ESTIMATING 305-DAY LACTATION YIELD IN DAIRY CATTLE

BY

PETER MALACHI SAAMA

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Animal Science

1995

ABSTRACT

EVALUATION OF METHODS FOR ESTIMATING 305-DAY LACTATION YIELD IN DAIRY CATTLE

By

PETER MALACHI SAAMA

Sampling variations in mean of daily milk yield within parity groups were approximated using bootstrap resampling. The variance of mean milk yield was widest at the peak to mid lactation. The optimal sample size for bootstrap resampling was found to be 80% of the original sample.

Differences in the mean and variance of daily milk yield at morning and evening during a lactation in different parity and season of calving groups were investigated. Across regions, mean morning yield was higher than evening yield. The variance in daily total yield was mostly determined by variance in daily evening yield. Second lactation cows had the largest variance in daily morning and evening yield. The greatest variance occurred at peak lactation during December to February.

Mathematical functions for various shapes of mean and variance of mean yield curves were established. These functions were representative of all breeds, regions, parities, ages at calving, and seasons of calving. A data set of sample lactation curves (SLAC) was generated from these functions. Fixed effects of missing test-day, and starting days of recording were included in SLAC. The SLAC was used to compare the relative accuracy of six methods for estimating 305-d milk yield. When no yields were missing, the overall prediction bias for all methods was generally very small. Within shape of lactation curve, variability in the accuracy of the methods was evident. With missing test-day records and varying starting days of recording, some methods had smaller bias than others. The establishment of SLAC was commissioned by the International Committee of Animal Recording to examine the accuracy of current as well as future methods of estimating 305-d milk yield.

DEDICATION

To the loving memory of Sarah Kemoli

ACKNOWLEDGMENTS

The author expresses sincere thanks to Dr. I. L Mao for his companionship, guidance, and advice during the course of this study. Dr. R. LePage is cited for his resourcefulness and patience in understanding the body of this work. Dr. Gill is cited for his counsel and friendship. Utmost gratitude is extended to Dr. Ferris and Dr. Banks for their advice and guidance. The love and support of my parents, Elizabeth Okwenje, William Kelly, Judith Saurman, Sylvia Knight, Robert Wright, Alfred Stephanik, Terri Swezey, and Lee Coryell is fondly remembered. The friendship of Dr. M. C. Dong, Dr. J. Jensen, Dr. G. Jeon, Dr. G. Maria-Levrino, Ms. T. Moore, Dr. F. Ngwerume, Dr. D. Krogmeier, Ms. G. Ferreira, Mr. F. Grignola, Ms. S. Chung, Dr. Zhiwu Zhang, Mr. M. Sillanpaa, and Mr. C. Chang is greatly appreciated. Dr. H. E. Koenig, Dr. B. E. Koenig, and Dr. J. Anderson are cited for their support and collaboration in a Systems Sciences project. The author is grateful to Rick Halbert, Dr. M. Delorenzo, and Dr. L Schaeffer for their assistance in providing some of the research data. The author acknowledges Riccardo Alleandri, Todd Meinert, Anne Kjeldsen, Hans Wilmink, Nicole Bouloc, and Felipe Ruiz for their assistance in the computation of lactation totals. This research was supported by a grant from The International Committee of Animal Recording. Computing resources were provided by the Center for Genetic Improvement of Livestock ,University of Guelph, Canada.

TABLE CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Total yield in a lactation is commonly estimated from production on a few sample test days during a lactation. Milk production in a lactation has been measured as the production during the first 305 days following parturition. This standard length allows records to be compared without concern for the varying length of the production period. The term "lactation curve" refers to the graphical representation of the relationship between milk yield and length of time since calving. A frequent application of the lactation curve is in the extension of records in progress to predict yield up to a standard length. Lactation curves also can be used to monitor the cow during a lactation for health and other managerial purposes.

Numerous methods for estimating lactation totals from test-day yields have been developed. Dairy records processing centers in different regions and countries use different methods. Different methods give results of a varying degree of accuracy. For a given method, accuracy of estimation may vary due to 1) length of test interval, 2) shape of lactation curve, 3) Missing test-d yields, and 4) starting day of recording.

Furthermore, programs for genetic improvement of dairy cattle have developed rapidly over the last 40 years. This progress has been due to the availability of artificial insemination, biotechnological advances such as embryo transfer, better information processing technologies, and more accurate methods for assessing the genetic merit of individual animals. Consequently, there is an ever increasing exchange of germplasm and information between countries. Meaningful international genetic evaluations require accurate estimates of lactation totals. However, differences in methods for calculating lactation totals present an obstacle to international genetic evaluations.

The International Committee of Animal Recording (ICAR) was established 34 years ago. It is an organization which is concerned with the coordination of livestock recording worldwide. Presently, ICAR has adopted only two methods of lactation yield calculation as "official ICAR methods", namely: the centering date method (CDM) and the test interval method

(TIM). Some countries have adopted CDM and TIM. Other countries, such as USA and Italy, use different methods. In 1990, the ICAR board commissioned a working group on "lactation calculation methods" to set up guidelines and standards for lactation calculation methods and related matters. The group decided to examine the efficiency of current methods and from this information methods would be evaluated for possible use as official methods by ICAR.

This study was undertaken to examine the relative accuracy of methods for calculating lactation totals. The data from this study would then be the fundamental basis for the recommendations made by the "lactation calculation methods" working group of ICAR.

In §2 the objectives of this project are outlined. In §3, a review of concepts important to the understanding of the body of this work are discussed. Namely, sources of variation in milk yield, sampling frequency during lactation, algebraic models for the lactation curve, methods for computing lactation totals, bootstrap resampling, data smoothing, principal factor analysis, and factor rotations. The body of the research is summarized in §4, §5, §6, §7, and §8.

A necessary aspect of this project was the generation of a data set of standard lactation curves (SLAC). Different methods for estimating lactation totals would then be applied to SLAC to compare their relative accuracy. In order to create this data set, we needed to know the shapes of mean of yield from morning milking and evening milking and daily total yield and the variance associated with each of these measures, during lactation.We had to understand how the mean and variance of daily morning and evening yield affect the mean and variance of total daily yield. From a series of preliminary analyses and after incorporating information from the literature and disucssions with the working group of ICAR, a set of curves for mean and variance of daily yield was created. Mathematical functions for these curves were determined. From these functions, a Monte-Carlo method was used to generate the SLAC. The total yield for lactations in SLAC was estimated using various methods. The methods were thereafter compared.

In §4, the bootstrap resampling method is used to approximate sampling variations in mean daily yield. In §5, data sets for three regions in North America were used to investigate daily variation in total, morning, and evening yield. From §4 and §5, we had preliminary understanding of typical shapes of curves for mean and variance of daily total, morning, and evening yields.

In §6, a set of curves for mean and variance of daily, morning, and evening yields was formally presented. Mathematical functions for these curves were stipulated. A Monte-Carlo method for generating sample lactations were discussed.

In §7, a complete description of how the SLAC was created and utilized was given. In §8 methods for estimating lactation totals were outlined and compared. Concepts presented in the earlier chapters were summarized and the results of the project were discussed.

The o

specif

1) to

2) to

3) to

4) to

5) to

6) to

Th

examin

to gene

accura

2. OBJECTIVES

The overall goal of this study was to compare methods of calculating total lactation yield. The specific aims were:

1) to examine sampling variations in mean daily milk yield;
2) to examine differences in mean and variation in morning and evening yields;
3) to develop mathematical functions for the mean daily yield in a lactation;
4) to develop mathematical functions for the variance of mean daily yield in a lactation;
5) to develop a data set of standard lactation curves;
6) to establish the relative accuracy of methods for computing lactation yields.

The results from aim 1 and 2 were used to establish the shape of some of the curves examined in aims 3 and 4. The mathematical functions developed in aims 3 and 4 were used to generate the standard lactation curves data set. Six methods were evaluated for their accuracy in estimating the total yield for each of the standard lactation curves.

3. REVIEW OF LITERATURE

3.1 Sources of variation in daily milk yield

While different types of management have a marked influence on variations in yields, it is not so easily understood why this variation should exist for those herds following all the currently recommended practices. Age at calving and season of calving are main factors affecting milk production in dairy cattle (Everett and Wadell, 1970; Schultz, 1974). Milk yield increases with age at a decreasing rate and reaches a maximum at maturity (Auran, 1973; Mao et al., 1974) as is shown in Figure 1A. The effect of seasonal variation was anlalysed into its two components, seasonality of production ('spring hump seasonality') and calving month seasonality (Wood, 1969) as is depicted in Figure 1B. Daily yield was depressed during the winter months and stimulated during the spring to an extent which was independent of stage of lactation. Cows calving in winter months tended to produce more in total lactation than spring calvers (Wood, 1969).

The standard deviation of daily production among cows varied from 6 to 4.5 kg through the lactation, the magnitude being closely related to the mean (Table 1, Anderson et al., 1989).

3.1.1 Variation on test-day

Much of the variation in test-day milk yield has been attributed to the interval between milkings (Everett and Wadell, 1970; Putnam and Gilmore, 1970; Shook et al., 1980), completeness of milking (Dodd and Foot, 1948), dry matter intake (Polan et al., 1986), estrus (Hurnik, et al. 1975; Palmer, 1982), and water intake (Murphy, 1992).

Age effects which are frequently confounded with production group effects can affect test-day variation. Stanton et al. (1992) used a test-day model to study the effects of age on test-day production and concluded that age at calving would account for more of the variation in test-day production that age on test-day.

The reproductive status of a cow may contribute to test-day variation. Modest (Hurnik et al., 1975) to significant (Palmer, 1982) increases in milk yield have associated with the onset of estrus.

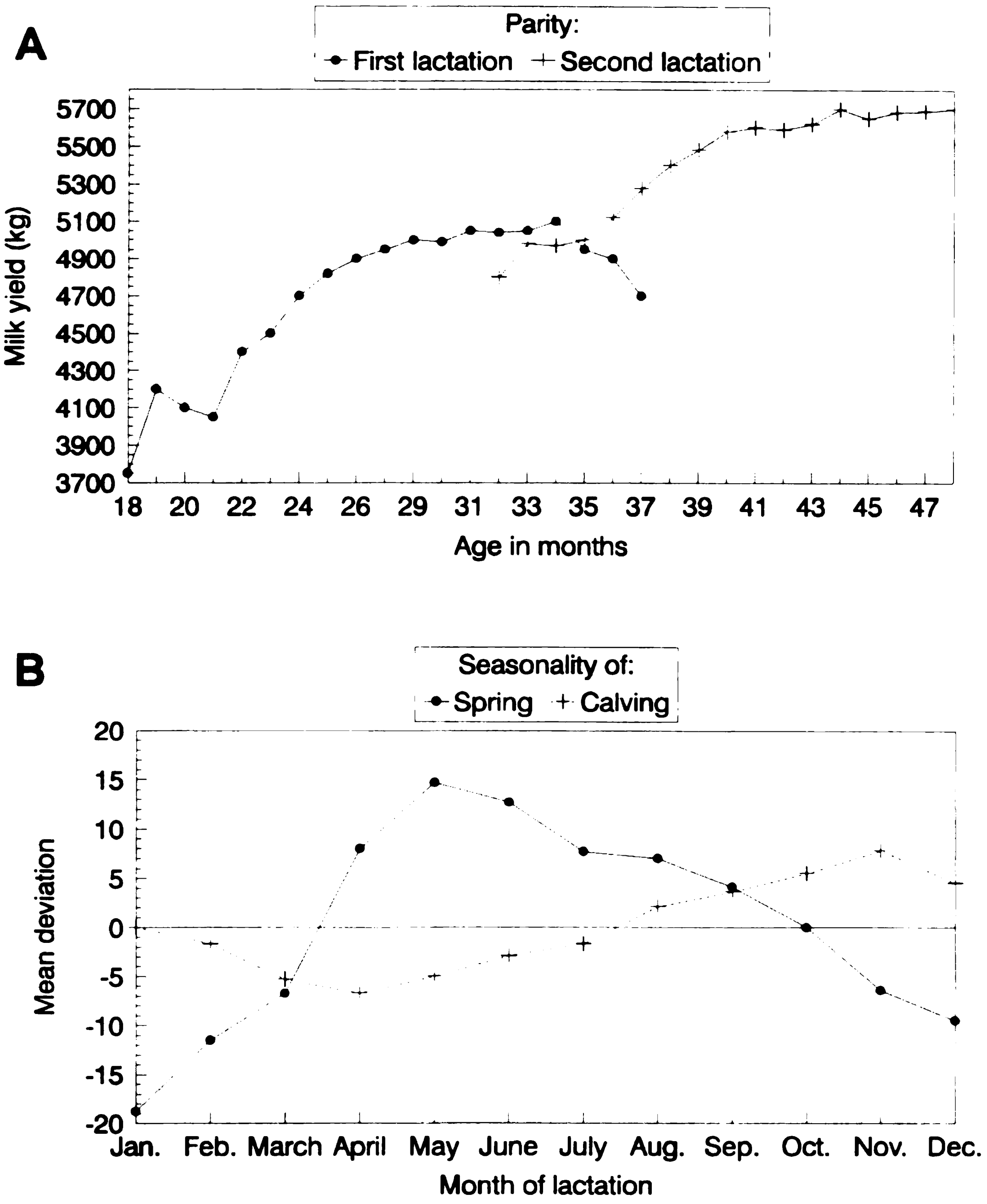


Figure 1. Mean milk yield of 305-day milk yield for first and second lactation in A (Source: Mao et al., 1974). Effect of seasonality on curve of lactation and total milk yield in B (Source: Wood, 1969).

TABLE 1. Averages and standard deviations of daily milk yield

Days in lactation	Number of cows	Average milk yield	Standard deviation	Coefficient of variation
		(kg)		
10	255	24.5	6	0.24
15	255	28.7	5.4	0.19
20	255	29.7	5.1	0.17
30	255	30.4	5.1	0.17
40	255	30.5	5.5	0.18
50	255	30.7	5.5	0.18
60	255	30.4	5.2	0.17
70	255	29.8	5.1	0.17
80	255	29.5	5.4	0.18
90	255	28.4	5.1	0.18
100	255	27.5	5	0.18
120	255	26.7	5	0.19
140	255	25.6	4.7	0.18
160	255	24.6	4.6	0.19
180	255	23.5	4.8	0.2
200	255	22	4.6	0.21
220	254	21	4.5	0.21
240	253	19.8	4.5	0.22
260	248	17.9	4.7	0.26
280	232	16.5	4.7	0.28
300	193	15.1	4.5	0.3

Source: Anderson et al. (1989)

Shook et al (1980) and Everett and Wadell (1970) concluded that differences between morning and evening milk yield were primarily a function of milking interval and the number of days in lactation.

3.1.2 Variation between test-days

The relationship between milk yield and month of calving is caused in part by the seasonal variations in feeding and care. The effect of month of calving on persistency has been observed by Sanders (1923 and 1930), Gaines (1927a), Gooch (1935), Johansson and Hansson (1940), Woodward (1945), Sikka (1950), Mahadevan (1951), Pickard (1952), Appleman (1969), Wood (1972), and Schultz (1974). These studies indicated that those cows calving in fall and winter are more persistent than those calving in spring and summer. Johansson and Hansson (1940) found that 5% of the total variation in persistency was due season of calving while Sikka (1950) reported this figure to approach 9.57%. Interactions between month of calving and stage of lactation were observed (Dannell, 1981; Miller et al., 1967). This source of variability in day to day milk yield suggested that shape of the lactation curve is dependent on month of calving.

However, Auran (1973) showed that month of calving was not as important as age at calving. Month of calving accounted for about 1.8% of the total variation in the first test-day and about 7.8% in the seventh and eight test-days. The influence of age at calving on monthly test-day yields decreased with advancing lactation, accounting for about 41% to 50% of total variation for first monthly test to about 2% to 5% for the last three days (Auran, 1973; Dannell, 1981). Thus, contrary to the age effects, the effect of month of calving is largest towards the end of lactation.

In addition, the relationship between morning and evening yields influences the shape of the lactation curve. Daily morning to evening ratios increased during later stages of the lactation (Palmer et al., 1994).

The effect of month of test-day on test-day production has been investigated (Syrstad, 1965; Everett and Wadell, 1970; Shook et al., 1980; Dannell, 1981; Ng-Kwai-Hang et al., 1984). Lindgren et al. (1980) concluded that a cow's production is less affected by month of testing immediately after calving than later in the lactation.

Gavin (1912), Brody et al. (1923), Hammond et al. (1923), Gaines et al. (1926), Johansson and Hansson 1940), Turner (1943), Louca and Legates (1968), and Chazal and Chilliard (1986) reported that the effect of pregnancy on daily milk yield is not noticeable until five months into gestation. Smith and Legates (1962) showed that production seemed to decline more rapidly 16-20 weeks following conception. Corley (1956) summarizes,

> **".. results agree in that neither persistency nor total yield is appreciably influenced by pregnancy during the first five months of gestation. However, if cows conceive early or carry a calf over 200 days of any lactation, a slight decline in total yield and a definite drop in persistency will likely occur."**

The calving interval may impact upon variation in yield during lactation. Sanders (1923), Gaines (1927a), Bonnier (1935), Johansson and Hansson (1940), Klein and Woodward (1945), Smith and Legates (1962) have shown that persistency increases with increased length of calving interval.

The effects of days open on milk yields have been studied by Wilton et al. (1967), Smith and Legates (1962), Ripley et al. (1970), Schaeffer and Henderson (1971), and Schultz (1974). Production losses due to long open periods have been reported by Louca and Legates (1968).

The effect of bovine somatotropin (bST) administration on milk yield during the lactation of cows maintained in cold environmental conditions was studied (Becker et al., 1990). Under farm conditions, bST treated cows produced 11% more milk than control-treated cows and in environmentally controlled chambers produced 17.4% more milk.

3.1.3 Variation between lactations

Ng-Kwai-Hang et al. (1984) indicated that milk production increased markedly between two and five years of age and then increased at a slower rate between five and six years of age.

Wood (1967) investigated the effect of parity on the constant a for average daily production. This constant on a log scale was 3.53, 3.72, 3.97, 3.86 for first, second, third, and forth or later parities respectively. Thus, first lactation cows were shown to be more persistent than later lactations. Results of Gowen (1920), Gaines (1927a), Turner (1927), Sanders (1923 and 1930), Gooch (1935), Dickerson et al. (1939), Johansson and Hansson (1940), Ludwick (1942), Sikka (1950), Ripley (1970), and Keown (1986) show similar findings.

Sikka (1950) noted that differences in lactation number accounted for 17.7% of the variation in persistency and 31.2% of the variation in maximum yield. Ludwick (1942) studied the records of 130 Guernsey, Holstein, and Jersey cows and reported an 8-10% drop in persistency from first to second lactation. Corley (1956) showed that cows in first lactation were 8% more persistent than cows in their second lactation. He found little difference however among subsequent lactations.

Appleman (1969) showed a significant interaction between lactation number and season of calving as did Miller et al. (1970). They found that older cows were more severely affected by summer calving than their younger counterparts.

3.1.4 Variation between cows

Cows that produce moderately with high persistency throughout lactation usually will be under less stress than cows that are less persistent have a large differential between production at peak and end of lactation. Age effects on lactation yield were demonstrated by Mao et al. (1974). Between cow differences have been attributed to herd (Auran, 1973; Everett and Wadell, 1970; Goodger et al., 1988; Mao et al., 1974), region (Mainland, 1985), additive and nonadditive genetic variance (Grossman et al., 1986), disease (Bartlett, 1991; Simerl, 1992), temperature (Becker et al., 1990; Elvinger et al., 1992), stocking rate (Baker and Leaver, 1986).

Ambient temperatures affect the performance of cows. Heat stress cows increased rectal temperatures, respiration rates, and decreased milk yield. (Elvinger et al.).

Differences between morning and evening yield may contribute towards variation between cows. Daily morning and evening ratios were plotted for selected lactations (Palmer et al., 1994). All showed large daily variations and evidence of cow differences.

3.2 *Sampling frequency during lactation*

Accepted intervals between recording yields of cows have gradually changed over the years. The 7-day interval was shown by Yapp (1915) to be a poor indicator, while Gaines (1927b) later showed that 7-day tests conducted after 60 days in lactation were more indicative as an estimate of lactation yields. M'Candlish and M'Vicar (1925) found that a 1-day test per month yielded results within 2% of actual yield, and Dick (1950) observed an average error of 2.32% from actual when cows were tested at 28d intervals. Houston (1932) found that weekly test intervals gave estimated yields approximating actual, and that to keep errors within a range of 10%, the testing interval should not exceed 30d. McDowell (1927) found that monthly and bimonthly tests varied from actual an average of 2.91 and 3.80 per cent, respectively. Clearly, there is no unanimous agreement on length of test intervals. However, a four week interval between tests is most common globally. Anderson et al. (1989) found that the four week equal interval sampling procedure gave acceptable estimates of total lactation milk yield. The crux of the matter remains that the total yield in a lactation has to be determined from periodic test-day yields. Both accuracy and precision in estimating total lactation yield increase with frequency of sampling.

3.3 Fitting lactation curves

Since the 1920's there has been considerable interest in the mathematical description and analysis of the lactation curve in dairy cattle. When a functional form is used to describe a lactation curve then:

1) the milk yield at any given stage of lactation can be predicted. Such predictions, if accurate, can be used as a basis to cull or to retain for breeding stock; 2) an individual animal's lactation and thus the average curves of groups of animals may be compared in terms of parameters of

the functional form; 3) it provides a mathematical description of average milk yield needed in any simulation model of a dairy enterprise; 4) concise summaries of patterns can be generated from which cumulative curves can be generated. The general approach has been to exploit parameters of the function in order to fit different lactation curves.

Mathematical functions for the shape (Wood, 1967; Kumar and Bhat, 1979), peak time (Sikka, 1950; Cobby and Le Du, 1978; Molina and Boschini, 1979), and declining phase (Brody, Ragsdale and Turner, 1923; Gaines, 1927a;) of the lactation curve have been proposed. Wood's equation,

$$y_n = a\ n^b \exp(-cn),$$

has been the most frequently applied. The variable n represents the length of time since calving. Coefficients a, b, c are constants determining the characteristics of curves. The equation can be estimated by ordinary least squares (OLS) in the form:

$$\log_e(y_n) = \log_e a + b\log_e n - cn\ .$$

The curve reaches a turning point at

$$n_p = -(b/c).$$

demonstrating that b is the parameter for pre-peak curvature, and c the parameter of post-peak curvature. Congleton and Everett (1980a, 1980b) examined the prediction error of Wood's equation and concluded that the function provided predictions of 305-d cumulative milk that were comparable with the estimates obtained by other DHIA techniques. Ramirez et al. (1994) found this equation performed better than the functions proposed by Brody (1923), Sikka (1950), Nelder (1966), and Colby and Le Du (1978). However, several authors (Cobby and Le Du, 1978; Danoa, 1981; Rowlands et al., 1982) have reported systematic lack of fit to lactation milk yields recorded weekly using this model.

Goodall and Sprevak (1984) retained Wood's formulation and introduced an autocorrelation function to account for systematic lack of fit,

$$\log_e \varepsilon(t) = \alpha \log_e \varepsilon(t-1) + e(t)$$

where e(t) is a random error term and α is a parameter such that $|\alpha| < 1$. Values of α greater than zero were interpreted to be an improvement over Wood (1967).

In order to account seasonal variation, Grossman et al. (1986) extended Wood's equation by adding sine and cosine terms:

$$y_n = a\, n^b \exp^{-cn} (1 + u \sin(x) + v \cos(x))$$

where a, b, c, u, and v are coefficients to be estimated, n = day of lactation; x=day of year, computed as radians. The log transformation of this model can be fitted by OLS. Batra (1986) compared this extended function to the inverse polynomial (Nelder, 1966) and found that the latter function gave a better fit than the former, based on R^2.

Grossman and Koops (1988) proposed a multiphasic function to describe the lactation curve, based on the sum of logistic functions:

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

where y_t is milk yield at t days in milk, n is the number of lactation phases, tanh is the hyperbolic tangent. Functions of parameters for each phase included initial yield computed as $a_i b_i$; contribution from each phase to 305-d yield, computed by integrating each phase from t=0 to t=305; time of peak yield for each phase, defined as c_i; and duration of each phase as days required to attain 75% of total yield and computed as $2b_i^{-1}$. DeBoer et al. (1989) fitted the multiphasic function to first through third-parity curves for milk yields for Israeli Holsteins and concluded that functions of parameters for each phase differed by parity, yield, and days open.

Lactation curve estimates also have been obtained by solving for OLS estimates of the fixed effects of days in milk (DIM) on test-d yield for which the lactation curve is partitioned into numerous DIM classes (Ngwerume, 1994; Schaeffer and Dekkers, 1994; Schaeffer et. al., 1994; Stanton et al., 1992; Trus and Buttazoni, 1990). The primary advantages of the test-d model are that it can account for 1) information from different lactations; 2) permits estimates of fixed effects to vary across herds and stages of lactation, and 3) adjust for effects of sampling date.

3.4 Methods for computing lactation totals

The traditional record of cow lactation yields is based on recording at weekly, monthly, or longer intervals. However, actual lactation yield can only be calculated by accumulating daily yields. Many procedures have been developed for computing total yield in a 305-d period. The following is a review of the methods that were compared in this study.

3.4.1 Test interval method (MSU)

The test interval method uses the time from one test-d until the following test-d as the test period (Appendix A of ICAR agreement). In general, a lactation record is calculated using three steps:

1. Estimate the sample-day yield for milk;
2. Estimate the yield from the previous sample day through the current sample day (credit for the test interval; and
3. Add the test interval credits for the lactation to determine the total lactation yield.

For most intervals, the interval yield (or credit) is calculated by multiplying the average yield between sample days by the number of days in the interval. Average daily yield in an interval is estimated as the average of the yields for the preceding and current sample days.

The credit for the first interval is calculated as

yield on sample day x day of lactation.

The test interval credit for an interval with a sample day after the first interval but prior to the last interval is calculated as

(yield on preceding sample day + yield on sample day)/2 x days in interval.

If a cow terminates her lactation before 305 d in milk, the credit for the last interval is calculated as (yield on preceding sample day + yield on sample day)/2 x days in interval

+

(yield on sample day) x (days to 305 d).

The first part of the formula gives credits for the interval immediately preceding the last one while the second part calculates credit for the days to 305 d. If the interval before 305 d is greater than 7 d, this situation is treated as resulting from an incomplete lactation and days to 305 d is set to zero. We can expect overestimation at the last interval and underestimation in the intervals spanning peak lactation.

If the yield for the current sample day is missing but the yield for the immediately preceding and subsequent sample days was recorded, an estimate of the sample day yield can be calculated as,

(yield on preceding sample day + yield on next sample day)/2.

If the yield is missing for consecutive sample days, there is no estimate for the missing yield.

For the morning/evening yield schemes, yield on a sample day is estimated as

(yield on sample day x 2)

Calculation of test interval credits is then performed in a manner similar to the one for daily yield.

The TIM method described above is illustrated in Appendix F.1

3.4.2 *Test interval method (France)*

The test interval credits from the first to the last interval are similar to those performed at MSU (Letter from Nicole Bouloc, Institut de l'elevage, Paris, France, 11/8/94). However, credits for the last interval are calculated as

(yield on preceding sample day + yield on sample day)/2 x days in interval

+

(yield on sample day) x c

where c is 14 for schemes A1, A4, A6, and AP/4 or c is 28 for scheme A8. Lactation yield is calculates as the sum of the test interval credits. There are no corrections for missing test-d yields. This method is illustrated in Appendix F.2.

3.4.3 Test interval method with adjustment factors (USA)

In the USA, most common sampling plans require weighing the milk at all milkings and collecting a composite sample during the approximate 24-h period of the sample day. Variations include AM-PM (AP) plans, for which only one milking is weighed and sampled each sample day for herds milked two times a day (2X) and only one or two milkings are weighed and one milking sampled each sample day for herds milked three times a day (3X). A complete description of this method for different sampling plans is given by Wiggans (1989) and adjustment factors are shown therein. Interval credits are computed in order to obtain 305 d yield.

The credit for the first test interval is calculated as:

$$\text{test interval credit}_{\text{first}} = \text{factor x yield on sample day x days in milk.}$$

Factors are based on breed, region, season, trait, lactation number, and stage of lactation.

The test interval credit for an interval with a preceding sample day before 40 d in milk is calculated with factors as:

$$\text{test interval credit}_{\text{peak}} = \text{factor x (yield on preceding sample day + yield on sample day)/2.}$$

The test interval credit for an interval with preceding sample day after 40 d is calculated as

$$\text{test interval credit}_{\text{post-peak}} = \text{factor x (yield on preceding sample day + yield on sample day)/2.}$$

The credit for the last interval is calculated as,

$$\text{test interval credit}_{\text{last}} = \text{factor x yield on last sample day x days to termination.}$$

The procedures for projecting lactation records of less than 305 d to 305 d is based on the number of days the cow actually milked, plus an estimate for the remainder of the 305-d lactation derived from the last available sample-day yield (Wiggans and Dickinson, 1985). For records less than 155 d in milk, the average mature-equivalent (ME) yield for cows freshening in the same herd 1 to 2 years prior to the record's last sample day also is required.

Separate factors have been developed by trait, calving season, lactation number, region of the country, and breed (Wiggans and Powell, 1980).

Records less than 305 d can be projected by

$$\hat{Y}_{305} = Y_{DIM} + (\hat{Y}_D)(305 - DIM)$$

where $\hat{Y}_{305}$ = projected 305-d yield,

Y_{DIM} = yield for the partial record, and

$\hat{Y}_D$ = estimated average daily yield for the remainder of the lactation.

For records with more than 155 d in milk, average daily yield for the remainder of the lactation can be estimated as

$$\hat{Y}_D = [\alpha_S + \beta_S(DIM)]\ (Y_S) + \alpha_S + \beta_F(DIM)$$

where α = intercept, S = sample day, β = slope, Y_S = sample-day yield, and F = factor.

For records with 155 d in milk or less, the ME herd average is included in estimating average daily yield:

$$\hat{Y}_D = [\alpha_S + \beta_S(DIM)]\ (Y_S) + \left[\frac{\alpha_H + \beta_H(DIM)}{1000}\right](Y_H)$$

where H = herd ME average and Y_H = herd-average yield.

3.4.4 Linear interpolation with standard curves (Netherlands)

This method was developed by Wilmink and Ouweltjes (1991) with the following requirements:

1 - The calculations should be independent of the recording scheme. The method should use all known test day yields as observations Cumulative yields should be estimated from these test day yields.

2 - Cumulative yield should be estimated by using corrections for the first part of lactation.

The expected shape of the lactation curve is estimated from lactations of contemporaries and is used in the calculation of cumulative yields in order to improve accuracy. Standard

lactations were estimated for separate classes of herd production level, age at calving, and season of calving (Wilmink, 1987 and 1990). These standard curves are used in the prediction of 24-h yield by interpolation and the prediction of 24-h yield before the first test day or a future yield.

Interpolation using standard lactations is performed using the equation:

$$y_i = g_i + ((y_2 - y_1) - (g_2 - g_1)) \times (x_i - x_1) / (x_2 - x_1) + (y_1 - g_1) \qquad \text{[NRS - 1]}$$

where y_i is an estimate for the uknown yield on the i^{th} day of lactation for a cow that has a recorded yield y_1 at day x_1 and yield y_2 at day x_2. Corresponding data from a pertinent standard lactation curve are yield g_1 at day x_1, yield g_2 at day x_2, and yield g_i at day x_i.

Unknown test day yields are estimated by the following prediction equation:

$$y_i = \mu_i + b_1 \times (x_p - \mu_p) + b_2 \times (y_{305} - \mu_{305}) \qquad \text{[NRS -2]}$$

where y_i = predicted yield at day i,

μ_i = expected yield at day i,

μ_p = expected yield at day p,

μ_{305} = expected yield over 305d in prior lactation,

x_p = realised yield at day p,

y_{305} = realised 305-d yield in prior lactation,

b_1, b_2 = regression factors.

All expected yields are taken from standard lactations.

To allow for varying interval lengths, the cumulative yield is computed as follows:

1 - The 24-h yield at day 0 is predicted by [NRS - 2].

2 - The 24-h yields at 30, 50, and 70 d are estimated by [NRS - 1], as long as these days are surrounded by measured test day yields. If the first test day yield is measured after 30,50, or 70d, the 24-h yield at 30, 50, and 70 d is predicted using [NRS - 2].

3 - If the lactation is completed, the 24-h yield for the last day in lactation is predicted by [NRS - 2].

4 - Using all known test day yields and calculated 24h yields (in 1, 2, and 3) the total yield is calculated by:

$$\sum_{i=2}^{n} \left[(INT_{i-1}+1)\times y_i+((INT_{i-1}-1)\times y_{i-1})\right]/2 \qquad \text{[NRS - 3]}$$

where: INT_{i-1} = number of days between y_i and y_{i-1};

y_i = the i^{th} test day yield (measured or computed)

n = total number of test day yields.

3.4.5 Centering date method (Denmark)

Denmark uses a combination of TIM and the centering date (CDM) method (Letter from O. K. Hansen, Danish Agric. Adv. Centre, Aarhus on 9/7/94). These calculations give exactly the same results when all sample milk weights are known. Each interval between sample days is divided into two parts. The interval yield is calculated as the sum of the yield in the first part of the interval and the yield in the second part of the interval.

The yield in the first interval is calculated as:

yield on sample day * day of lactation

The test interval credit for an interval with a sample day after the first interval but prior to the last interval is calculated as:

yield on preceding sample day x length of first part of the interval

\+

yield on sample day x length of the second part of the interval

The last interval is the interval from last sample to 305 d. In this calculation the credit of the last interval is calculated as:

yield on sample day x days to 305 d

Usually in Denmark, the lactation length exceeds 305 d, and routinely the yield in the last part of the 305-d period is estimated from the sample yield following the test day.

Records are only allowed for cows in the first five days after calving. In all other cases a zero yield is assumed and is used in the calculations. If there are missing records at the end of a lactation, and it is not because the cow is dry, the credit of the last interval is calculated as:

b x yield on sample day x days to 305 d

where b is a factor which is calculated from the herd average, age at 1st calving, breed, parity, season and days to 305 d. If the cow is dry a zero yield is assumed. This method is illustrated in Appendix F.3.

3.4.6 *Multiple trait projection method (Italy)*

This multiple trait projection (MTP) method was developed by Trus and Buttazzoni (1991) with the following objectives:

1 - The method should easily accommodate the range of factors affecting lactation yields

2- Partial lactation data should be treated in a statistically optimal way such that the prediction errors are minimized and projections are not sensitive to "outlier" input data points.

The MTP method has five basic elements:

1 - Divide the lactation curve into intervals corresponding to the sampling method. Observations within each interval are treated as expressions of a trait that is correlated with all other intervals.

2 - Define a fixed effects linear model which is adequate for all intervals. The same model is used for each interval.

$$Y_i = X_i\beta_i + e_i \qquad \text{for } i = 1,2,\cdots,t.$$

where $\mathbf{Y}_i$ = an observed vector observations within the i^{th} interval

$\mathbf{X}_i$ = the known design matrix for the i^{th} interval

β_i = an unknown vector of fixed effects

e_i = an unknown vector of random residuals corresponding to $\mathbf{Y}_i$

$$E\begin{bmatrix} Y_i \\ e_i \end{bmatrix} = \begin{bmatrix} X_i\beta_i \\ 0 \end{bmatrix}; \qquad Var[e_i] = R_i = Var[Y_i - X_i\beta_i]$$

The model in Italy takes into account the following fixed effects: 1) pregnancy status, no. of milkings per d, season of milking, days pregnant (DPREG) as covariate, DPREG^2 as a covariate, days in milk (DIM) since beginning of interval as a covariate, DIM^2 as a covariate, age of cow at calving (i.e. days/100) as a covariate, year of birth of cow as a covariate, and bST in a given period.

3 - Given a population, estimate model coefficients (β) and a matrix of residual (co)variances (R) among intervals:

$$R_{ii} = \left(\hat{e}'_i\hat{e}_i - \hat{e}'_i 1 1'\hat{e}_i / n_{ii}\right) / \left(n_{ii} - \text{Rank}\left[\left(X'_i X_i\right)\right]\right)$$

where n_{ii} = the number of cows with records in the i^{th} interval.

4 - For each cow on which a projection is to be made, daily yields are the sum of E(Y) and the interval residual. Residuals in intervals in which no observations occurred are predicted as weighted functions of the observed residuals. Once R and β have been estimated, there are used to calculate daily yields.

The projection of a complete lactation is achieved in three steps:

i - Calculate $\hat{e}_p = Y - X\beta$ for each observation.

ii - Calculate a residual $\hat{e}_m = R'_{pm} R_{mm}^{-1} \hat{e}_p$ for each of the remaining intervals without an observation.

iii - For each day calculate $\hat{Y} = X\hat{\beta} + \hat{e}$.

3.5 Bootstrap resampling

In ordinary usage the phrase 'resampling methods' refers to methods in which the observed data are used repeatedly, in a computer-intensive simulation analysis, to provide inferences. In simple terms, resampling does with a computer what an experimeter would do in practice, if it were possible: he or she would repeat the experiment. In resampling, the observed variable's values are randomly re-assigned to treatment groups, and the test-statistics are recomputed. These reassignments and recomputations are done thousands of times. Bootstrap resampling is only one of such methods.

Bootstrap resampling was used to examine sampling variations in mean daily milk yield.

The accuracy of a statistic describes the sampling variations of that statistic. This accuracy depends on the width of the interval spanning the statistic. In most cases, the data needed to calculate the exact half-width around a statistic never exist. Efron (1979) proposed the idea of using computer-based simulations instead of mathematical investigations to obtain the sampling properties of random variables. The bootstrap method advances the notion that by repeated sampling from the data, one can approximate the sampling variations which produced the data. Thus, the one available sample gives rise to many others. Various studies examining theoretical aspects of the bootstrap procedures have been conducted (Beran and Ducharme, 1991; LePage and Billard, 1992; Singh, 1981). The bootstrap method has been applied to practical problems that either did not have completely satisfying solutions or were resistant to statistical investigation (Westfall and Young, 1993).

The following is a brief review of the one-sample bootstrap algorithm, described more completely by Efron and Tibshirani (1986). An unknown probability distribution F produces the observed data $(x_1, x_2, \cdots, x_n) = X$ by random sampling. That is, $x_1, x_2, \cdots, x_n$ are independent and identically distributed (i.i.d) observations from F. From $x_1, x_2, \cdots, x_n$, we calculate a statistic of interest $S(x_1, x_2, \cdots, x_n)$, the numerical value of which is S^0. Let $\hat{F}$ indicate the empirical probability distribution, putting probability (1/n) on each observed value x_i $(i = 1, \cdots, n)$. A bootstrap sample $(X_1^*, \cdots, X_n^*) = X^*$ is a random sample of size n from $\hat{F}$: Each X_j^* is an independent realization of X_i with probability (1/n) (i=1, 2, ..., n). The statistic of interest S evaluated for the bootstrap data X^* is a bootstrap replication of S, the numerical value of which is S^*. We can generate as many bootstrap replications of S^* as desired. Let us call the independently generated bootstrap samples (each of which contains n bootstrap data points) X^{*1}, X^{*2}, $\cdots$, X^{*B}, where B may be at least 1000 for an approximate confidence interval. Each X^{*b} gives an independent bootstrap replication of the statistic of interest, S^b. The bootstrap replications provide bias and variance estimates for S as described by Efron and Tibshirani (1986).

The $100(1-\alpha)^{th}$ percentile of the bootstrap distribution, $S_\infty(\alpha)$, is estimated by

$$\bar{S}_B(\alpha) = 100\alpha^{th} \text{ percentile of } S^1, S^2, \cdots, S^B$$

Thus, we have an approximate confidence interval for the statistic of interest.

3.6 Smoothing of continuous functions

The need to smooth data arises in many classes of problems. Specifically, we often want to find the 'best approximation' of the function after the random noise has been removed. The problem is to find a function which has the least deviation from a given function. Much of the theory of 'best approximation' is treated in detail by Shapiro (1969). The 'moving average' method of data smoothing was used to smooth bootstrap sampling variations of mean daily milk yield in §4 and curves in §6. We formally state our problem as follows:

Given $f \in L^1(-\infty, \infty)$, can we approximate f by continuous functions in the L^1 metric? The answer is in the affirmative is and will be shown below.

Theorem: Let $f \in L^1(-\infty, \infty)$. Then, given $\varepsilon > 0$ there exists a continous function $g \in L^1(-\infty, \infty)$ such that

$$\|f-g\|_1 = \int_{-\infty}^{\infty} |f(x)-g(x)|dx < \varepsilon.$$

Proof: Let a be a positive real number. Define f_a, the moving average of f, by the formula

$$f_a(x) = \frac{1}{2a}\int_{x-a}^{x+a} f(t)dt = \frac{F(x+a)-F(x-a)}{2a} \qquad [1]$$

where

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

It is easy to check that $f_a \in L^1(-\infty, \infty)$, and that it is continuous. Hence, from integration theory,

$$\lim_{a \to 0} f_a(x) = f(x) \text{ a.e.}$$

Thus, f is the limit a.e. of a sequence of continuous functions. A detailed proof that

$f_a \in L^1(-\infty, \infty)$ is given by Shapiro (1969).

Generalizing, we rewrite [1] in the form,

$$f_a(x) = \int_{-\infty}^{\infty} f(x-t) G_a(t)\, dt$$

where $G_a(x) = \begin{cases} \frac{1}{2a} & \text{for } |x| \le a \\ 0 & \text{for } |x| > a \end{cases}$.

Whence, we define the convolution f^*g of f and g by the formula

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-t)\, g(t) dt$$

The following properties apply to the convolution product

(i) $f*g \in L^1(-\infty, \infty)$. In fact $\|f*g\|_1 \le \|f\|_1 \|g\|_1$;

(ii) $f*g = g*f$;

(iii) $f*(g*h) = (f*g)*h$

for $f, g, h \in L^1(-\infty, \infty)$.

Hence, notice that f_a is the convolution product of f and G_a. Moreover, if we set

$$K(x) = \begin{cases} \frac{1}{2}, & |x| \le 1 \\ 0, & |x| > 1 \end{cases}$$

then

$$\int_{-\infty}^{\infty} K(x) dx = 1$$

and writing for $\lambda > 0$,

$$K_\lambda(x) = \lambda K(\lambda x)$$

we have

$$G_a(x) = K_\lambda(x), \qquad \lambda = \frac{1}{a}.$$

So, the 'moving average' method of smoothing is characterized by a certain function known as the kernel, K(x). Writing $f(x; \lambda)$ for $f_a(x)$, we have

$$f(x;\lambda) = (f * K_\lambda)(x) = \int_{-\infty}^{\infty} f(x-t)\,\lambda\,K(\lambda t)dt$$
$$= \int_{-\infty}^{\infty} f\left(x - \frac{t}{\lambda}\right) K(t)dt.$$

3.7 Principal factor analysis

Factor analysis was used to examine differences between methods for computing total lactation yield. Factor analysis entails the statistical analysis of multivariate data from a mixture of finitely many populations. The task, at hand, is to find fundamental and meaningful dimensions of a multivariate domain by examining the intercorrelations between a set of traits of interest.

In the initial step, a composite score measuring what these traits have in common is generated. This score must explain the maximum variance among the variables. The principal axis, or component, defines the factor or basic dimension the variables are measuring in common. This procedure is called principal components analysis (Morrison, 1976). The resulting principal factors are used as a set of reference axes for determining the most easily interpretable set of factors for the domain in question. This whole process, which Harman (1960) calls multiple-factor analysis, is reviewed within the context of the sample space model by Cooley and Lohnes (1971).

Subjective procedures are proposed (Harman, 1960) for developing the transformation from some initial solution to the multiple-factor form of the solution. The methods consist of the build up of a series of rotations in a plane using simplification of the rows or columns of the factor matrix.

After the initial factor extraction, the common factors are uncorrelated with each other. If the factors are rotated by an orthogonal transformation, the rotated factors are also uncorrelated. If the factors are rotated by an oblique transformation, the rotated factors become more correlated. However, oblique rotations often produce more useful patterns than do orthogonal rotations.

3.7.1 *Orthogonal varimax factor rotation*

The varimax criterion involves simplification of the columns of the factor matrix and has become the most widely accepted and employed standard for orthogonal rotation of factors since its development by Kaiser (1958). He defines the simplicity of the factor as the variance of its squared loadings (A factor loading is a correlation between the underlying factor and the observed trait in question):

$$V_k = \frac{\left\{ p\sum_{j=1}^{p}\left(s_{jk}^2\right)^2 - \left(\sum_{j=1}^{p} s_{jk}^2\right)^2 \right\}}{p^2}$$

where s_{jk}^2 is the new factor loading for variable j on factor k; j=1, 2, ..., p, and k = 1, 2, ..., n. Then for the entire factor matrix the varimax criterion is:

$$V = \sum_{k=1}^{n} V_k = \sum_{k=1}^{n} \left\{ \frac{p\sum_{j=1}^{p}\left(s_{jk}^2\right)^2 - \left(\sum_{j=1}^{p} s_{jk}^2\right)^2}{p^2} \right\}\Bigg|_{max}$$

To eliminate some slight bias associated with the column sums $\sum_{j=1}^{p} s_{jk}^2$, Kaiser redefined the criterion by 'normalizing' the loadings,

$$V = \sum_{k=1}^{n} \left\{ \frac{p\sum_{j=1}^{p}\left(s_{jk}^2/h_j^2\right)^2 - \left(\sum_{j=1}^{p} s_{jk}^2/h_j^2\right)^2}{p^2} \right\}\Bigg|_{max}$$

where h_j^2 is the communality of the j^{th} trait (A communality is the proportion of the variance of the j^{th} trait that is explained by all n factors). Kaiser (1958) delineates the method fully. The criterion V is maximized by the iterative application of trigonometric functions.

3.7.2 *Oblique Procrustean factor rotation*

First, the matrix of factor loadings is rotated to orthogonal simple structure using the varimax criterion. Then these orthogonal results are rotated to a least squares fit to give the ideal oblique solution. Hendrickson and White (1964) define a matrix $P=(p_{kj})$ such that:

$$p_{kj}=\left|s_{kj}^{u+1}\right|/s_{kj}$$

with u > 1. Thus, each element of this matrix is the uth power of the corresponding element in the row-column normalized orthogonal matrix. Then find the OLS fit of the orthogonal matrix of factor loadings to the pattern matrix, P:

$$L=(E'E)^{-1}E'P$$

where L is the unnormalized transformation matrix of the reference vector and E is the orthogonal rotated matrix. This is the 'Procrustes' equation described by Hurley and Cattell (1962). The columns of L are normalized such that their sums of squares are equal to unity. This provides the transformation matrix from the orthogonal factors to the oblique reference vectors.

3.7.3 Oblique Harris-Kaiser factor rotation

A derived oblique solution which employs only positive definite diagonal matrices (D matrices) and orthonormal matrices (T matrices) is presented by Harris and Kaiser (1964). This feature then permits translating the problem of the developing an oblique solution into the problem of orthogonal rotation of a matrix that differs in certain ways from the initial orthogonal solution, F. The preliminaries are as follows:

1) R^*, the correlation or covariance matrix of the traits of interest.

2) $R^*=QM^2Q', \quad Q'Q=I, \quad QQ'\neq I$

where M^2 is positive definite and diagonal and Q consists of m columns of normalized eigenvectors corresponding to the nonzero eigenvalues of R^*. The tautological expression is:

$$R^*=QM^2Q'=(QMT_2D_2T_1D_1)$$
$$\cdot(D_1^{-1}T'_1D_2^{-1}T'_2M^{-1}M^2M^{-1}T_2D_2^{-1}T_1D_1^{-1})(D_1T'_1D_2T'_2MQ')$$

in which all T matrices are orthonormal $(T'T=TT'=I)$. An oblique solution is obtained by setting $T_2=I$ and $D_3=I$ (with $T_1\neq I, D_2\neq I, L\neq I$). Then, for an independent cluster solution, define

$$A = Q\,T_1\,D_1,$$
$$L = D_1^{-1}\,T'_1\,M^2\,T_1\,D_1^{-1}, \qquad \text{(Case II; Harris and Kaiser, 1964)}$$
$$B = Q\,M^2\,T_1\,D_1^{1}.$$

Here A is regarded as a pattern matrix and B as a structure matrix in the sense of Harman (1964), and the matrix L designates the intercorrelations of the factors.

3.7.3.1 Harris-Kaiser rotation with Cureton-Mulaik weights

Kaiser's iterative algorithm for the varimax rotation fails when a) there is a substantial cluster of variables near the middle of each bounding hyperplane, and/or b) there are appreciably more than m traits whose loadings on one of initial F-matrix, usually the first, are near zero. Cureton and Mulaik (1975) proposed an approach for overcoming these difficulties by weighting the factors, giving maximum weights to those likely to be near the primary axes, intermediate weights to those likely to be near hyperplanes but not near primary axes, and near-zero weights to those almost collinear with or almost orthogonal to the first initial F-axis. For a solution, normalize the rows of the initial F-matrix and call the result G, with elements $g_{jk} = f_{jk} / h_j$, where h is the square root of the communality of trait j. Isolate all rows of G whose first-factor loadings are negative, and call the result A. The desirable weighting function is:

$$w_k = \cos^2\left(\frac{\cos^{-1}\sqrt{(1/m)} - \cos^{-1} a_{k1}}{\cos^{-1}\sqrt{(1/m)}} \times 90^\circ \right) + .001 \quad \text{if } a_{k1} \geq \sqrt{(1/m}$$

$$w_k = \cos^2\left(\frac{\cos^{-1}\sqrt{(1/m)} - \cos^{-1} a_{k1}}{90^\circ - \cos^{-1}\sqrt{(1/m)}} \times 90^\circ \right) + .001 \quad \text{if } a_{k1} < \sqrt{(1/m}$$

Let W be a diagonal matrix of the m weights for the m traits. Then the weighted varimax orthogonal approximation to simple structure is:

$$V_v = F\,\Lambda_w,$$

where Λ_w is the transformation matrix of $V_w = WA$. Applications to the Procrustean rotation are discussed by Cureton and Mulaik (1975). An extension to case II of the Harris-Kaiser oblique procedure exists if their T_1 is replaced with the weighted varimax transformation matrix Λ_w.

4. BOOTSTRAP ASSESSMENT OF THE SAMPLING VARIATIONS IN MEAN DAILY MILK YIELD

4.1 Abstract

The notion behind bootstrap is that by sampling repeatedly from data, one can approximate the sampling variations which produced that data. The objective was to estimate sampling variations in mean of daily milk yield (DMY), throughout a lactation, within subclasses of parity and season of calving. We used 4 yr daily milk records of 340 Holstein cows from a Michigan herd. Parity groups were 1st, 2nd, 3rd, 4th or higher. Season of calving groups were April through October and November through March. Bootstrap resampling was done within each of the eight parity-season groups: A random sample comprising of a fixed percentage of the total number of lactations was obtained to form a hypothetical random sample from the population. This sample was duplicated, or cloned, to form a proxy for the population. A random sample of the fixed percentage, the bootstrap sample, was drawn without replacement from the cloned population. Different percentages were studied to determine optimum size for resampling. This resampling was repeated 5000 times. The mean of DMY for each of the bootstrap samples was deviated from those calculated from the hypothetical sample to give an approximation to the sampling variations in the mean of DMY.

4.2 Introduction

The availability of electronic identification and decreased cost of electronic data acquisition have made feasible the daily monitoring of milk yield for individual cows. The physiological state of cows can be associated with abnormal fluctuations in daily milk yield.

The sources of variation in milk yield have been examined (Everett and Wadell, 1970) and reviewed (Palmer et al., 1994). The gross standard deviation and coefficient of variation of total daily was 4.5 to 6 kg and 17 to 30% within 300-d lactations (Anderson et al., 1989).

Because variance estimates differ based on the sample, there is a need to have a better understanding of the sampling variations around daily milk yields.

However, the data needed to compute the true variance of daily milk yields are too expensive to acquire. The bootstrap (BS) method, introduced by Efron (1989), can provide a good approximation to the true variance given a relatively small sample that is representative of the population. It is a computer-intensive method that achieves this approximation by repeated sampling from the original sample. In simple terms, resampling does with a computer what an experimeter would do in practice, if it were possible: he or she would repeat the experiment. In resampling, the observed variable's values are randomly re-assigned to treatment groups, and the test-statistics are recomputed. These reassignments and recomputations are done thousands of times.

The objectives of this study were: 1) to approximate sampling variations in mean daily milk yield; 2) to determine the optimum proportion of the sample data for bootstrap resampling; 3) to use bootstrap confidence intervals for generating lactation records.

4.3 Materials and methods

The data were daily milk yield for 340 primiparous Holstein cows from a low somatic cell count herd in Michigan. There were 89 cows in the first lactation and 251 in the second and later lactations. 67% of the first lactation cows and 88% of the second and later lactation cows had > 305 days in milk. Parity subclasses were defined as 1) first parity and 2) second and later parities. To be consistent with the literature, all records were truncated at 305d.

Sampling variations in mean DMY were approximated by bootstrap resampling. To determine the optimum proportion of the original data for resampling, two sampling proportions (p) were defined; $p = .2$ for light and $p = .8$ for heavy sampling. Within parity the BS method was implemented by the following steps:

a. Generate a proxy for the population. Note that each cow generates 305 data points.

i) Obtain random sample of n cows, where n = p*N (e.g. N=89 for 1st parity).

ii) Compute sample mean.

iii) To reduce noise, smooth the sample mean by convolution:

Define a suitable kernel estimator,

$$k(y-x) = \exp^{-200\left(\frac{i}{305}\right)^2} ;\; i = 1, \cdots, 305$$

Then let f(y) = a(ii). Then

$$f(x) = \int f(y)\, k(y-x) dy$$

and

convolution = InverseFourier[Fourier[f(x)] Fourier[k(y-x)]]

iv) Clone the sample (N / n) times.

b. Generate the bootstrap sample.

i) Obtain a random sample of size n without replacement from the clone. This is known as the BS sample.

ii) Compute mean of the BS sample.

iii) To reduce noise, smooth the mean of the BS sample by convolution.

iv) Compute difference a(iii) and b(iii).

c. Repeat b 5000 times to give 5000 x 305 matrix of smoothed BS differences.

d. To obtain naive 100(1 - α)% confidence interval (CI),

i) Sort smoothed BS differences in c

ii) The half width for mean DMY is the i^{th}, j^{th} element of e where,

$$i = 5000(1 - \alpha)$$

iii) CI = b(iii) ± d(ii). Due differences between a(iii) and b(iii), we can expect some degree of under or over-estimation by the BS approach.

e. A lactation curve can be generated by random sampling of a real value between the upper and lower bounds described by d(iii) on each day in lactation.

4.4 Results and discussion

The BS resampling analyses were performed on an Intel 486/66MHz IBM PC compatible computer with 20MB of RAM. The average CPU time was 58 sec per iteration. Figure 2 shows the 100(1-α)% BS confidence intervals of mean DMY for first lactation cows after light sampling (p=.2) of the original data. The mean of the original sample, which represents the population mean, is shown on each plot. These data were consistent with theory; the 99% CI > 95% CI > 80% CI. However, the 80% CI for mean DMY did not cover the pseudo-population mean. The half-width was widest in the interval spanning peak lactation and narrowest at the beginning and end of 305-d lactation. The figure shows that the "sample mean" lead to an overestimation during the middle part of the lactation. This overestimation was due to "sampling error". The BS estimates of confidence intervals were considered to be unbiased and were the primary focus of these results.

Contrary to the data reported by Anderson (1989), the variance of daily yield varied during lactation. However, our data are in agreement with those of Palmer et al. (1994).

The CI's after heavy sampling of the original data are shown in Figure 3. The half-width following heavy sampling was narrower than that for light sampling. Furthermore, the 80% CI covered the pseudo-population mean. Because the BS samples contained more information about the population, the coverage was much better. Light sampling led to wider BS estimates of the half-width. Results for the second and later lactation cows (not shown) were similar.

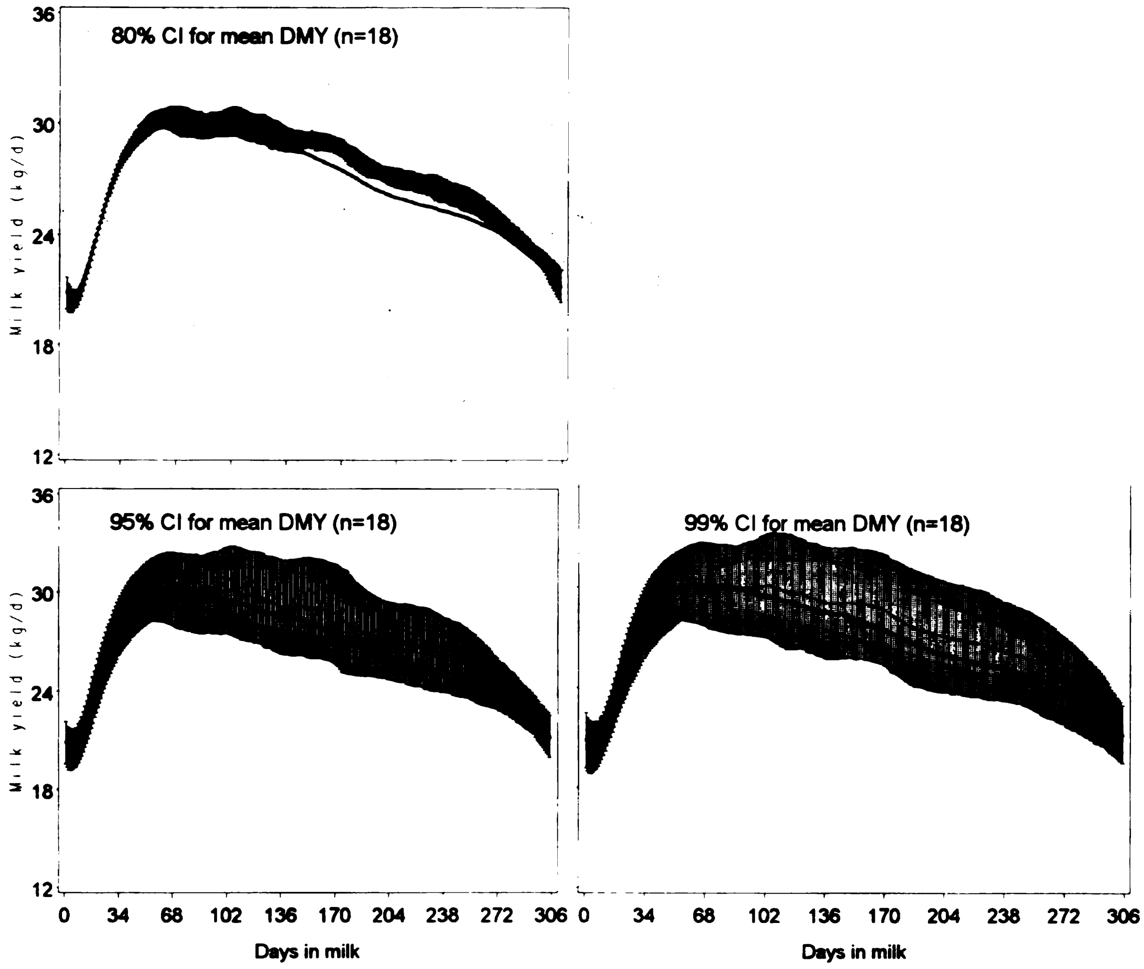


Figure 2. Bootstrap confidence intervals for daily milk yield. Data are for first lactation cows following sampling a light proportion (p=.2) of the original data. The pseudo-population mean is also shown on each plot.

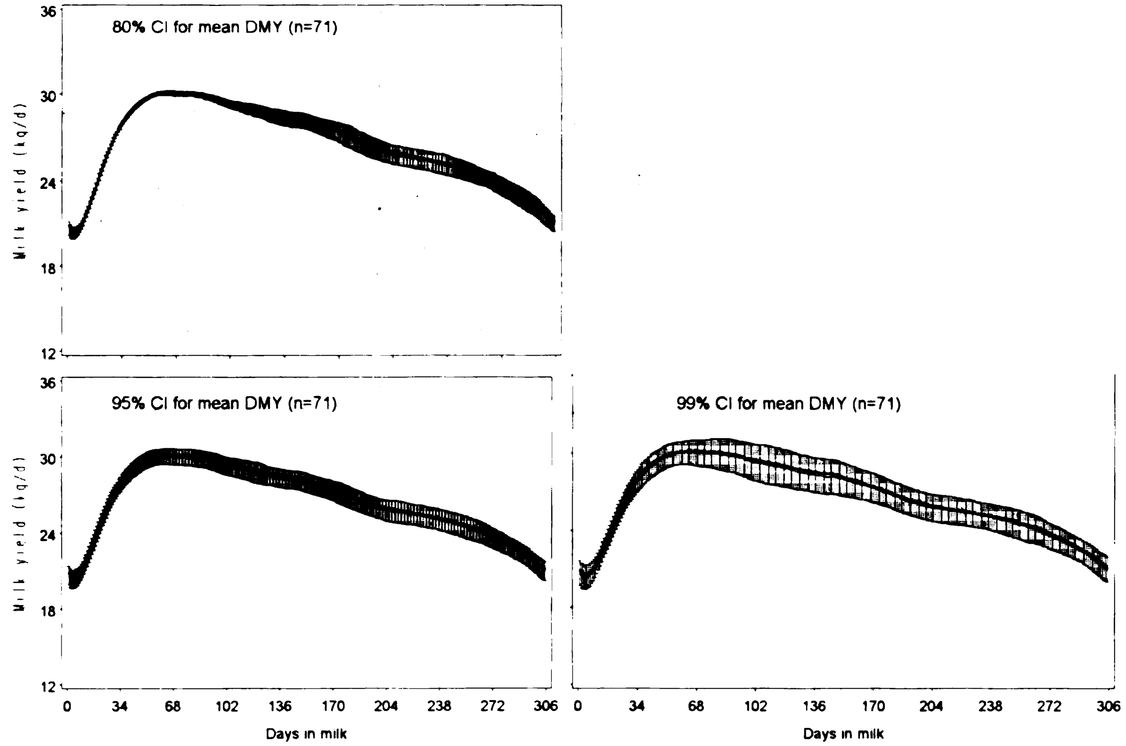


Figure 3. Bootstrap confidence intervals for daily milk yield. Data are for first lactation cows following sampling a heavy proportion (p=.8) of the original data. The pseudo-population mean is also shown on each plot.

A sample lactation curve generated using the 99% CI is shown in Figure 4. The fluctuations in milk yield depicted by this curve are comparable to those in the observed data.

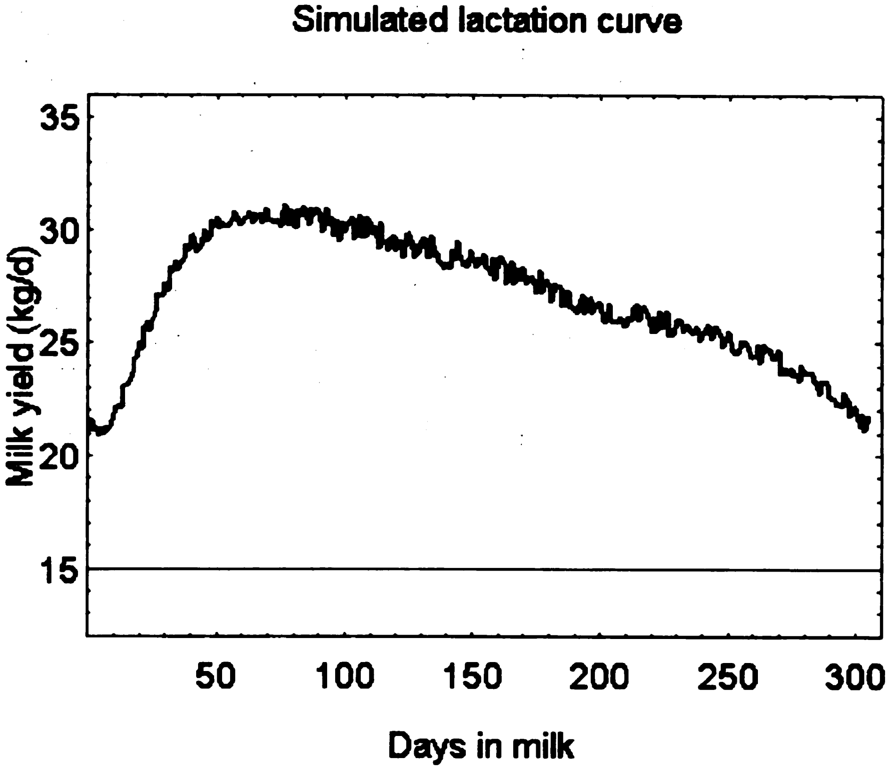


Figure 4. A sample lactation curve simulated by random sampling from 99% bootstrap confidence intervals.

4.5 Conclusions

The bootstrap method was used to approximate the sampling variations of mean daily milk yield. The bootstrap 100(1 - α)% CI of mean daily milk yield were accurate. However accuracy of the sampling variations was sensitive to the size of the bootstrap sample. The coverage of the confidence intervals was much better when the size of the bootstrap sample was equal to 80% of the original sample size. When the bootstrap sample size was only 20% of the original sample size, the confidence intervals appeared to be over estimated. The data show that confidence interval of mean daily milk yield is widest at peak to mid lactation. The confidence intervals obtained are useful in generating biologically consistent lactation records. The method is relatively easy to implement. The main dis-advantage of bootstrap resampling is the heavy computation involved and availability of computer software to perform analyses.

Most applications require individually crafted programs. In addition, once software is in place, the thousands of simulations required can be uncomfortably time-consuming, particularly for those with limited computing facilities. The use of bootstrap resampling in this study was not only descriptive but well suited to the problem of approximated the distributional characteristics of mean daily yield. The approach was suitable because the mean for the original sample was unusual compared to the resampling distribution. The resampling method was a convenient and asymptotically valid way of incorporating the unknown dependence structures inherent in the data. In the absence of the requisite large sample sizes, the method can be used to verify analytic results from standard analyses.

5. VARIATION IN MORNING AND EVENING YIELD DURING LACTATION OF HOLSTEIN COWS

5.1 Abstract

Differences in mean and variance of daily milk yield at morning and evening during a lactation in different parity and season of calving were investigated. Data on a total of 956,680 lactations of 3295 cows distributed in 26 herds from Dairy Herd Improvement Centers of Alberta and Ontario, and Florida Agricultural Experiment Station were used. Size of herds and number of cows with records had an impact on the magnitude of the variances within subclasses of parity and season of calving. Across regions, mean morning yield was higher than evening yield. The shape of the mean curve for total yield followed that for evening yield closely. Rank correlations between daily morning, evening, and total yield were as high as .998 ($P < .0001$). Variance in daily total yield was mostly determined by variance in daily evening yield. Second lactation cows had the largest variance in daily morning and evening yield. Parity differences were least in the Florida data. The highest variation occurred at 50 d postpartum during the lactation of cows calving in December to February. Fluctuations in daily yield were highest for cows freshening in June through August.

5.2 Introduction

Recent technological advances in identification and automated data capture have made feasible the daily recording of milk yields for individual cows. The accumulated data can be utilized in monitoring the health status and in facilitating the management of lactating cows.

Variation in milk yield due to different intervals between morning and evening milkings has been reported (Anderson et al., 1989, Everett and Wadell, 1970; Palmer et al., 1994; Schmidt, 1960). Gilbert et al. (1973) investigated diurnal variations of milk yield and found that milk production at morning (AM) and evening (PM) were not equal even though the milking interval was 12h. Hyde et al. (1981) observed that variation in milk yield was not constant throughout the entire lactation and varied between first and later parities.

Adjustment factors for the difference between AM and PM milk yields have been developed. These adjustment factors accommodate differences among herds, cows, seasons, month of lactation milking interval, month of lactation, breeds, and age at calving (Palmer et al., 1994; Putnam and Gilmore, 1970; Shook et al., 1980). Palmer et al. (1994) calculated a range of 2 to 6kg in gross standard deviation of individual milk yield within 305-d lactations. However, the literature is lacking in studies that show the actual magnitude of variation in the AM and PM milk yields throughout lactation. The objective of this study was to examine differences in mean and variation in AM and PM yields and their relationship to daily total yield. The effects of parity, season, and region on AM and PM yield both in mean and variation were also examined.

5.3 Materials and methods

Data of morning and evening milk yield for Holsteins cows from Florida, Alberta, and Ontario were used. Data from the two provinces were collected in an automated milking systems project conducted by the Canadian Dairy Herd Improvement Program. To be included in analyses, all lactation records had to have calving dates and parity number specified and were truncated at 305 d postpartum (pp). The resulting data sets included whole and part-lactation records with either a AM or PM milk weight on each day of lactation. The intervals between morning and evening milkings varied. However the exact length of the interval could not be determined because milking times for the sampled and previous milkings were not available.

The extent of the data is indicated in Table 2. The Ontario data set had the most records, lactations, cows, and herds. Six of the Ontario herds had less than 90 cows while all 5 Alberta herds had more than 160 cows. In contrast to the Florida data, the number of records increased with parity for the Canadian herds. Alberta herds tended to have more late Fall and Winter calvings whereas the Ontario data consisted of higher number of calvings during the Spring and Summer. Most of the Florida lactation records were initiated in late Fall.

TABLE 2. Test day data for Holstein cows utilized for analyses by location.

Category	Data source Alberta	 Ontario	 Florida
No. of lactation days	287325	451423	217932
No. of lactations	1288	2715	587
No. of cows	960	1924	411
No. of herd	5	20	1
Minimum herd size (cows)	161	28	___
Maximum herd size (cows)	437	490	___
1st lactation records	409	787	272
2nd lactation records	300	691	146
3rd or later lactation records	579	1237	169
Lactations started Dec. - Feb.	325	657	129
Lactations started Mar. - May	295	700	141
Lactations started Jun. - Aug.	308	723	130
Lactations started Sep. - Nov.	360	635	187

Figure 5 shows the number of cows with lactations records by location. All locations had few cows with records from 1 to 7 d pp. Thereafter, the Alberta data showed a drop in number of cows with records (NCR) after 280 d pp. A steady decline in NCR occurred in the Ontario data after 100 d pp. The NCR remained the same after 7 d pp for the Florida data. Whereas, NCR for Alberta and Florida was similar at a.m. and p.m., Ontario had more NCR for AM after 100 d pp. Most of the cows with records at a.m. had a record at p.m.

Within each location, parity, and season of calving, the mean, standard error of mean, 95% confidence interval of mean and variance of mean AM, PM, and daily total yield were computed. Product-moment correlations between mean AM, PM and daily total yield were calculated. A test for significant differences between correlations was conducted. All analyses were accomplished using analytical procedures of SAS (1992).

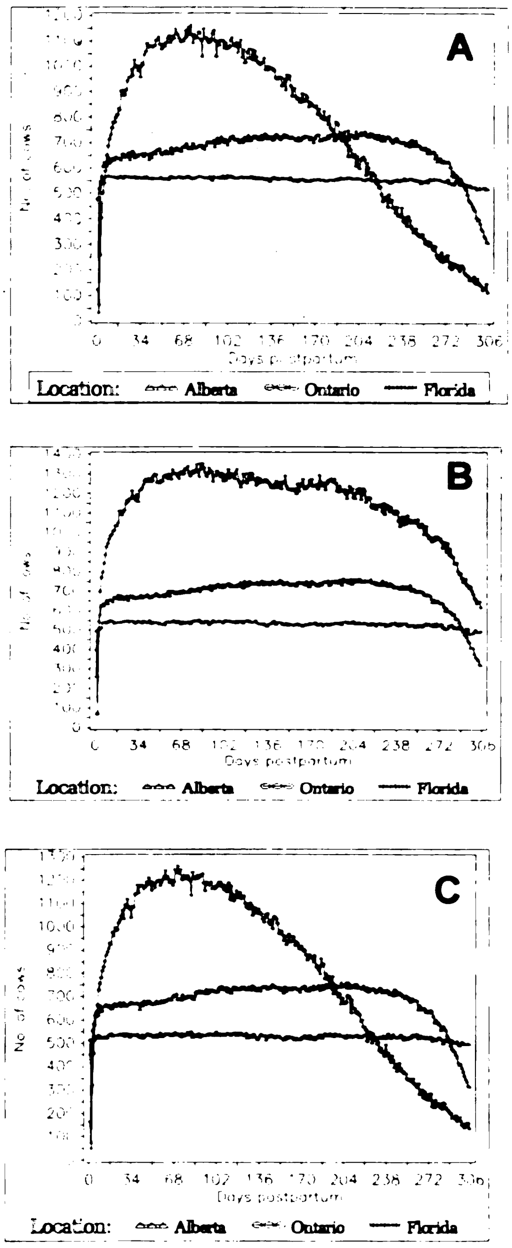


Figure 5. Number of cows with lactation records by location: A = total milk yield, B = morning milk yield, and C = evening yield.

5.4 Results and discussion

Table 3 shows the within location product-moment correlations between mean total, AM, and PM yields. Across locations, correlation coefficients were significantly high ($P < .0001$) but they were not significantly different from each other ($P > .1$). The magnitude of these correlations was not regarded as ample evidence for predictive ability because the sources of variation in total (Everett and Wadell, 1970; Murphy, 1992), AM, and PM yields are numerous (Amos et al., 1985; Becker et al., 1990; Elvinger et al., 1992; Lewis and Newman, 1984).

TABLE 3. Product moment correlations between mean morning, evening, and total yields.

	Alberta			Ontario			Florida		
	Total	AM[1]	PM[2]	Total	AM	PM	Total	AM	PM
Total	..	.994	.996	..	.970	.992	..	.993	.993
AM		..	.980		..	.99	..	..	.982

[1]AM = daily morning yield.

[2]PM =daily evening yield.

All correlation coefficients significantly different from zero ($P < .0001$) but not significantly different from each other.

5.4.1 Overall variance of daily milk yield

Figure 6 shows 95% confidence intervals for mean total, AM, and PM yield for Alberta, Ontario, and Florida. In all cases, mean AM yield was higher than mean PM yield. This relationship between AM and PM yields was similar to that observed by Putnam and Gilmore (1970). However, the general shape of mean curve for total yield appeared to follow that for PM. For both Alberta and Ontario, the variance of AM yield rose sharply, peaked at 20 d pp, declined rapidly then started to rise in late lactation. Ontario had much smaller herds. In small herds, time of AM milking tends to be consistent but PM milking time tends to be less so due to field work and social life. Thus, more variation in milk yield would be associated with cows in small herds. On the contrary, in the Florida data, the variance of AM peaked at 45 d and then declined steadily throughout lactation. In all cases, the variance of total yield followed that for PM. In this and subsequent illustrations, the apparently higher variance at the beginning and end of the lactation could be attributed to more missing observations caused by incomplete lactations.

5.4.2. Within parity variance of daily milk yield

95% confidence intervals and mean total, AM, and PM yields for Alberta, Ontario, and, Florida are shown in Figures 7, 8, and 9 respectively. Mean yield increased with parity. The variance of daily morning, evening, and total yield was highest for 2nd lactation cows. This difference in the variance peaks was associated with a possible region by parity interaction. This variance peaked at 7 d pp in the Alberta, and Ontario data but at 50 d in Florida. In addition, differences between 1st and 2nd lactation curves were much less for Florida. Because the Florida data were from a single herd, this lack of differences between 1st and 2nd parities was attributed to management within that herd.

5.4.3 Within season variance of daily milk yield

5.4.3.1 Alberta.

There were seasonal differences in shapes of the curves for mean milk yield (Figure 10). The largest peak for mean and variance of AM and PM was at 50 d pp, for cows freshening in December through February. The data for mean daily yield are in agreement with those of Keown and Van Vleck (1973) who reported that cows calving in January through April had the highest peak production. The least variance during lactation was for cows giving birth in March to May. The variance in AM and PM peaked at 7 d pp in June through August.

5.4.3.2 Ontario.

The magnitude of the variance for AM, PM, and total yield was highest at 45 d pp for cows calving in March to May (Figure 11). As few as 10 cows with records at the beginning and end of the lactation caused variances to be high. Fall freshening cows were more persistent and had the highest variance during lactation. The persistency in the Fall provides biological evidence for the Spring stimulus to production discussed by Wood (1969).

5.4.3.3 Florida.

The largest variance in AM, PM, and total yield occurred at 50 d pp for cows calving in December through February (Figure 12) and agreed with Florida. Mean milk yield was more depressed and had greater fluctuations in June to August but the variance declined gradually following the 50 d pp peak. Seasonal effects on mean daily yield differ from effects reported by Keown and Van Vleck (1973), who found cows freshening in May through August had the highest average production.

5.5 Conclusions

Mean morning yield was higher than evening yield. The data indicated that the shape of the mean curve for total daily yield followed that for evening yield. In these data, the variance of evening yield appeared to determine the variance of total daily yield. Regional differences in the magnitude of the variance of daily total, morning, and evening yield were observed. These differences could be associated, in part, with stage of lactation, parity, seasonal effects. The data did not contain complete morning and evening yields for each cow during lactation. Hence, it was not possible to quantify the observed region by season, region by parity, and parity by season interactions. Future studies should investigate these interactions.

ALBERTA

ONTARIO

FLORIDA

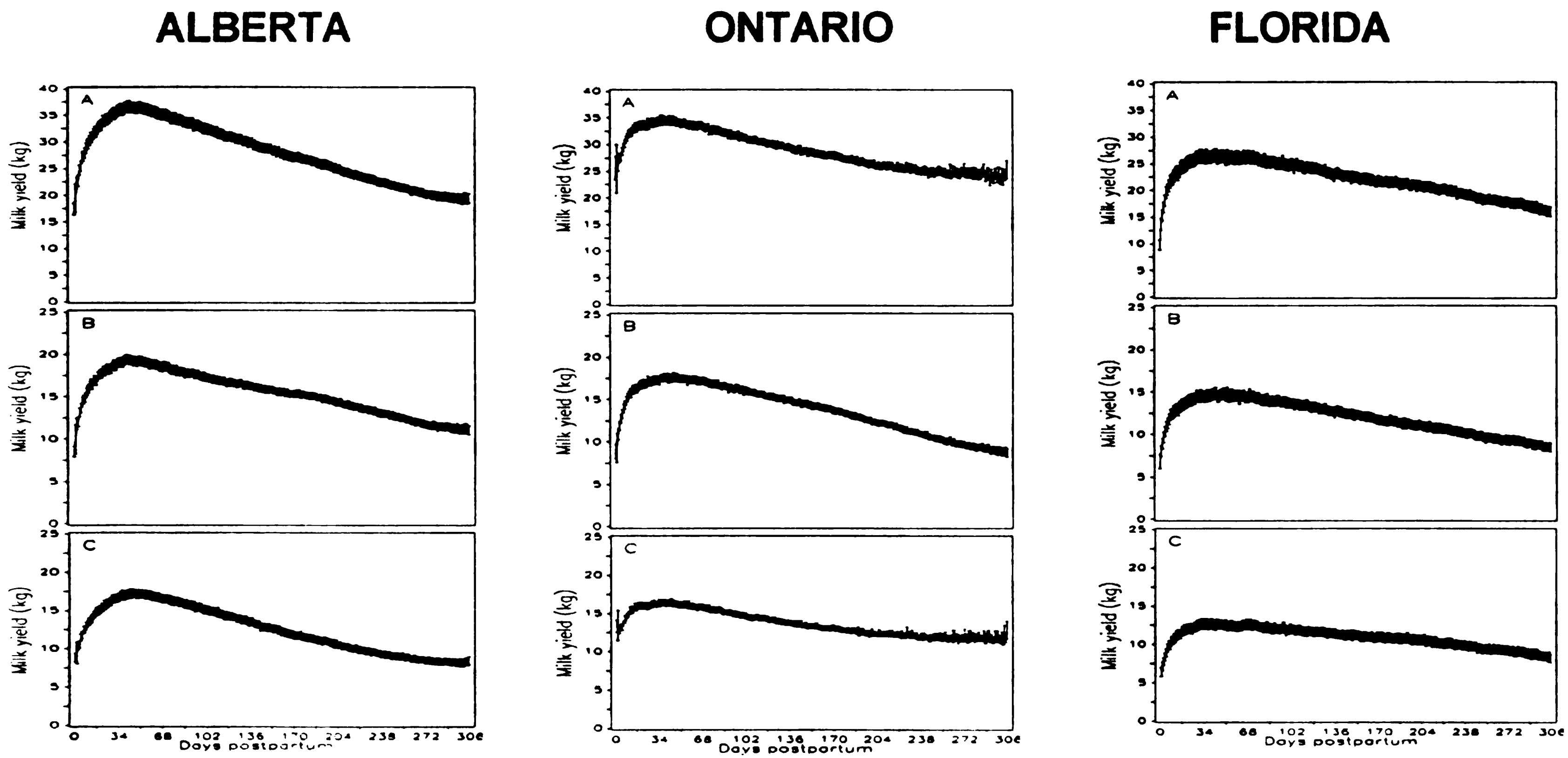


Figure 6. Estimated 95% confidence interval of mean daily milk yield for Alberta, Ontario, and Florida: A= total yield, B = morning yield, and C = evening yield.

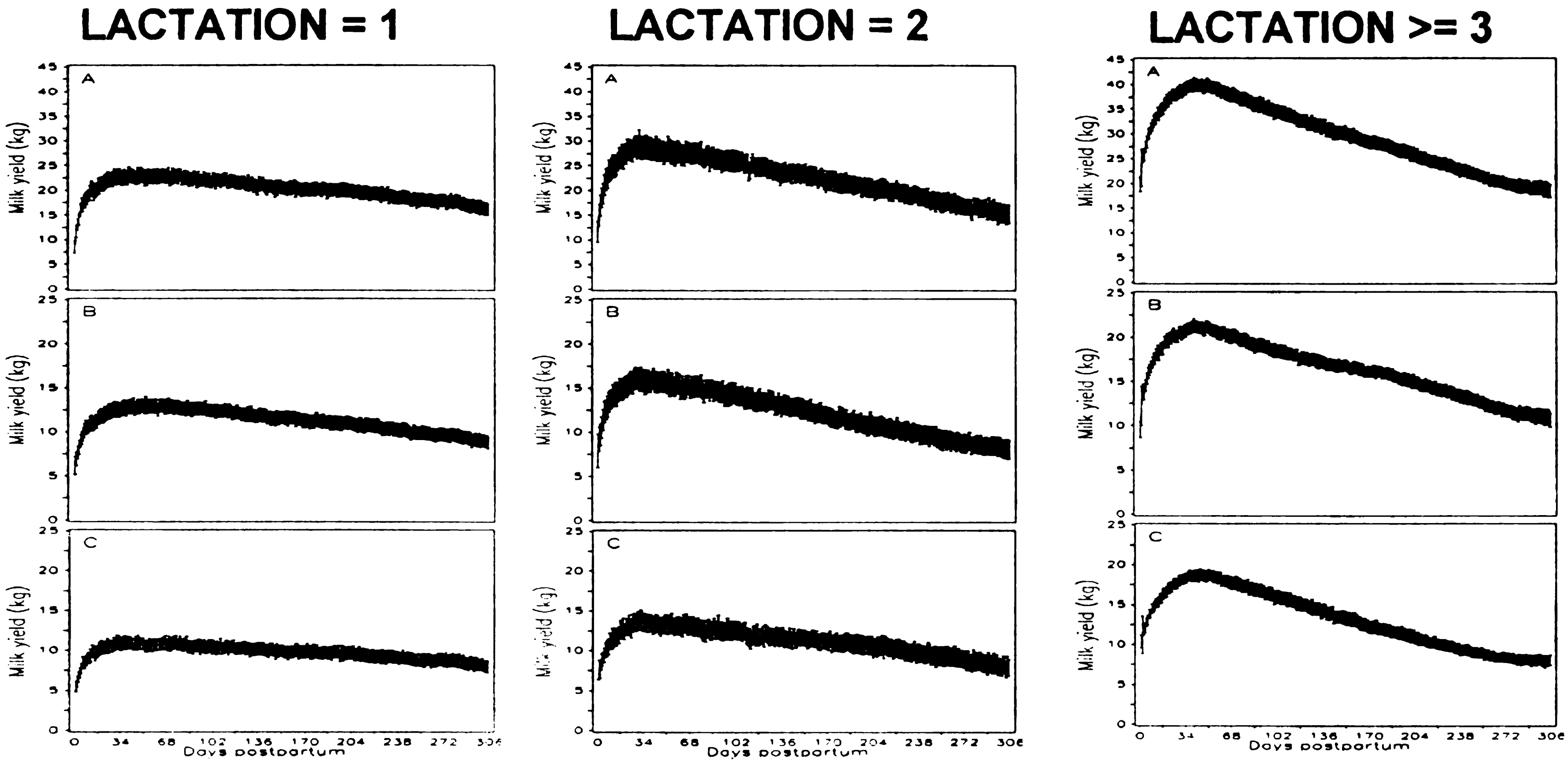


Figure 7. Estimated 95% confidence interval of mean daily milk yield within parity for Alberta, Canada: A = total yield, B = morning yield, and C = evening yield.

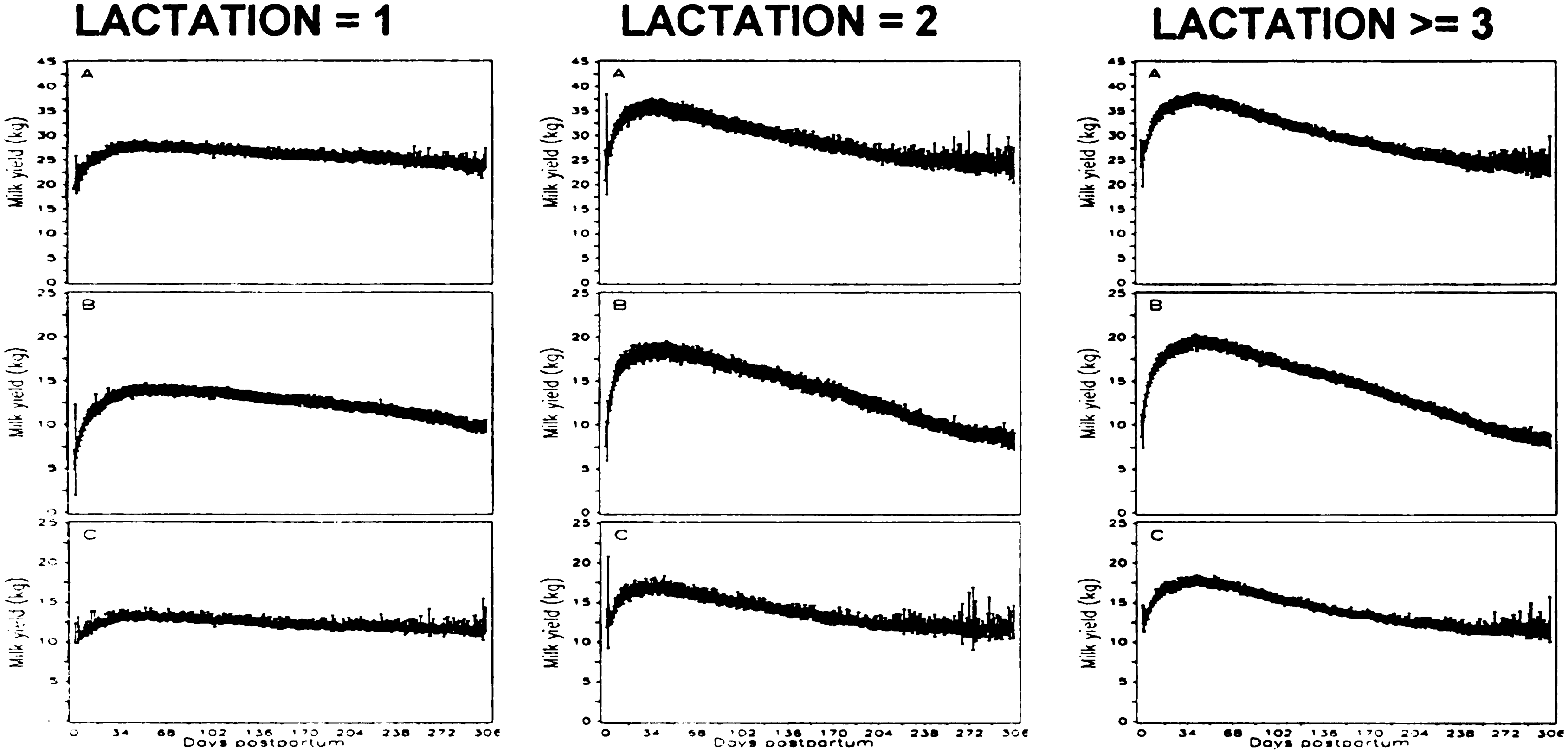


Figure 8. Estimated 95% confidence interval of mean daily milk yield within parity for Ontario, Canada: A = total yield, B = morning yield, and C = evening yield.

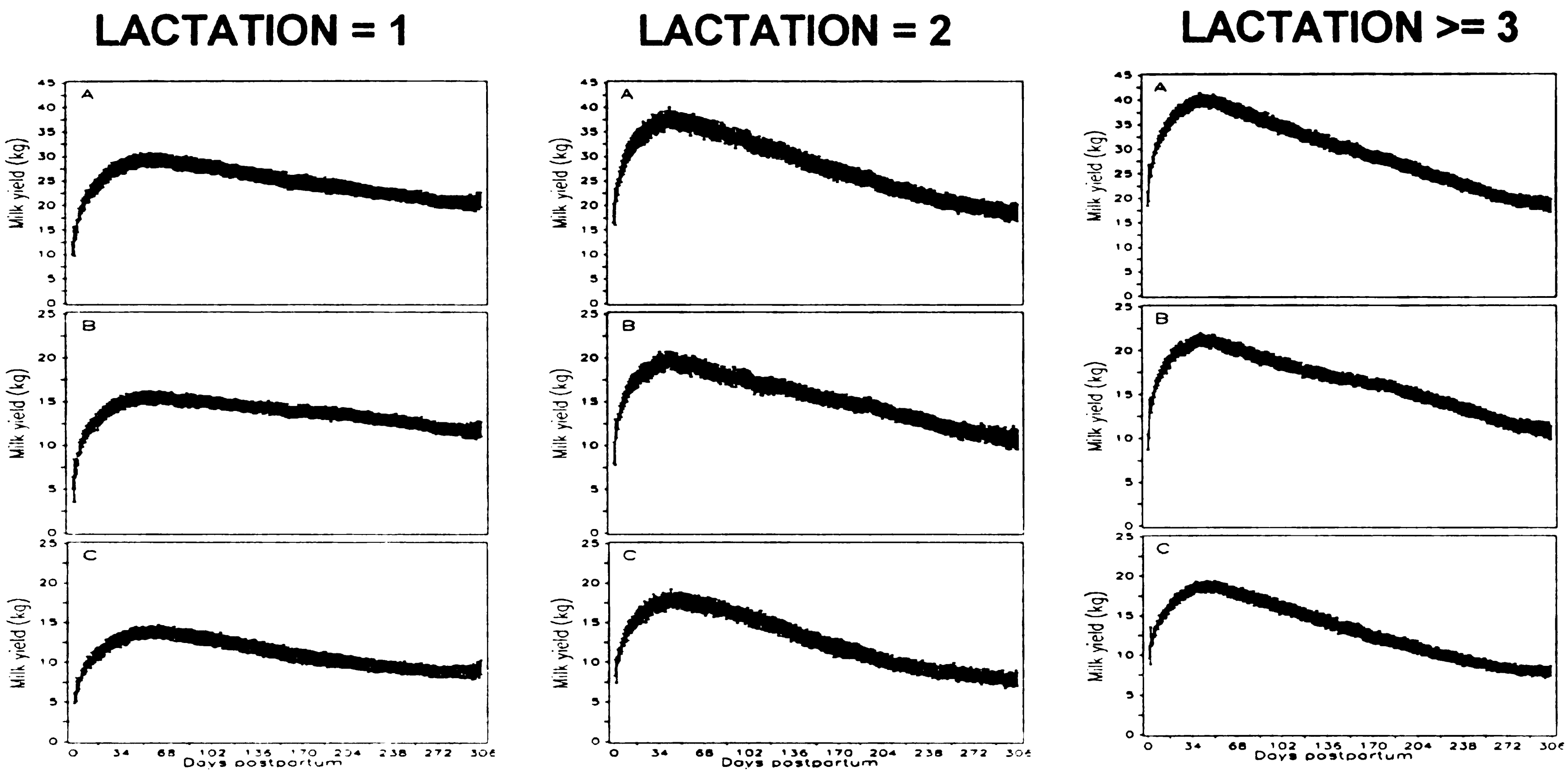


Figure 9. Estimated 95% confidence interval of mean daily milk yield within parity for Florida, USA: A = total yield, B = morning yield, and C = evening yield.

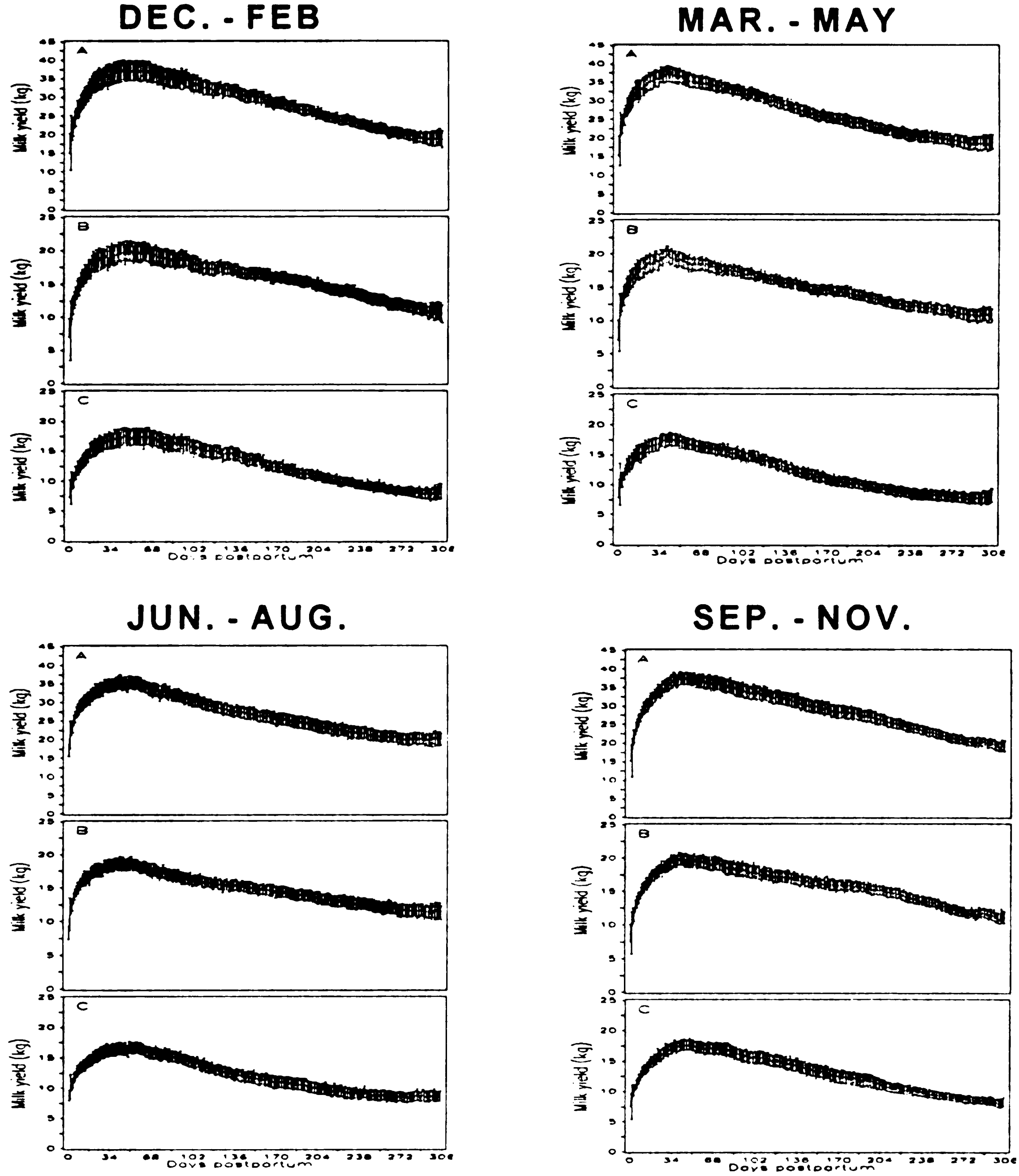


Figure 10. Estimated 95% confidence interval of mean daily milk yield within calving season for Alberta, Canada: A = total yield, B = morning yield, and C = evening yield.

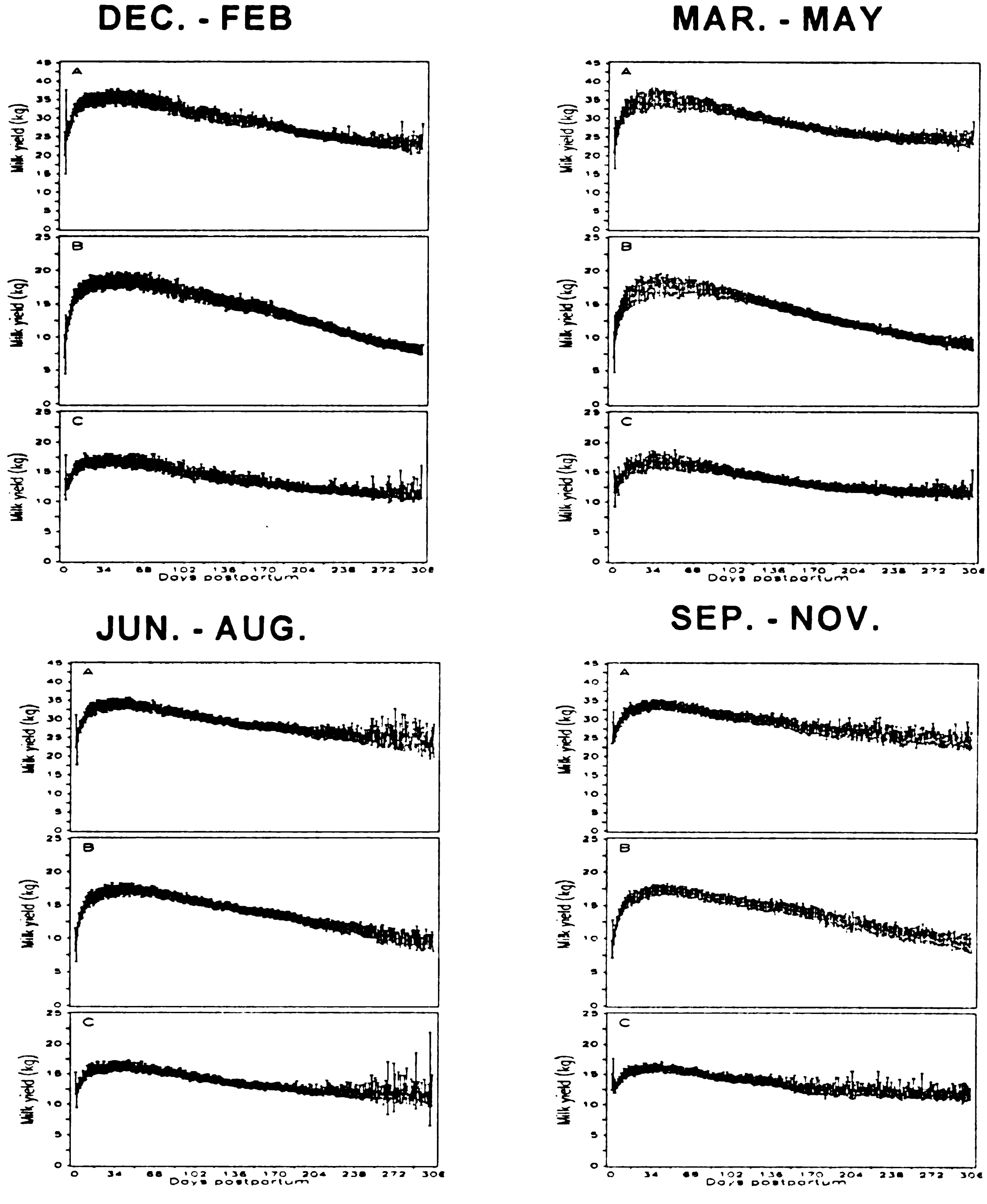


Figure 11. Estimated 95% confidence interval of mean daily milk yield within calving season for Ontario, Canada: A = total yield, B = morning yield, and C = evening yield.

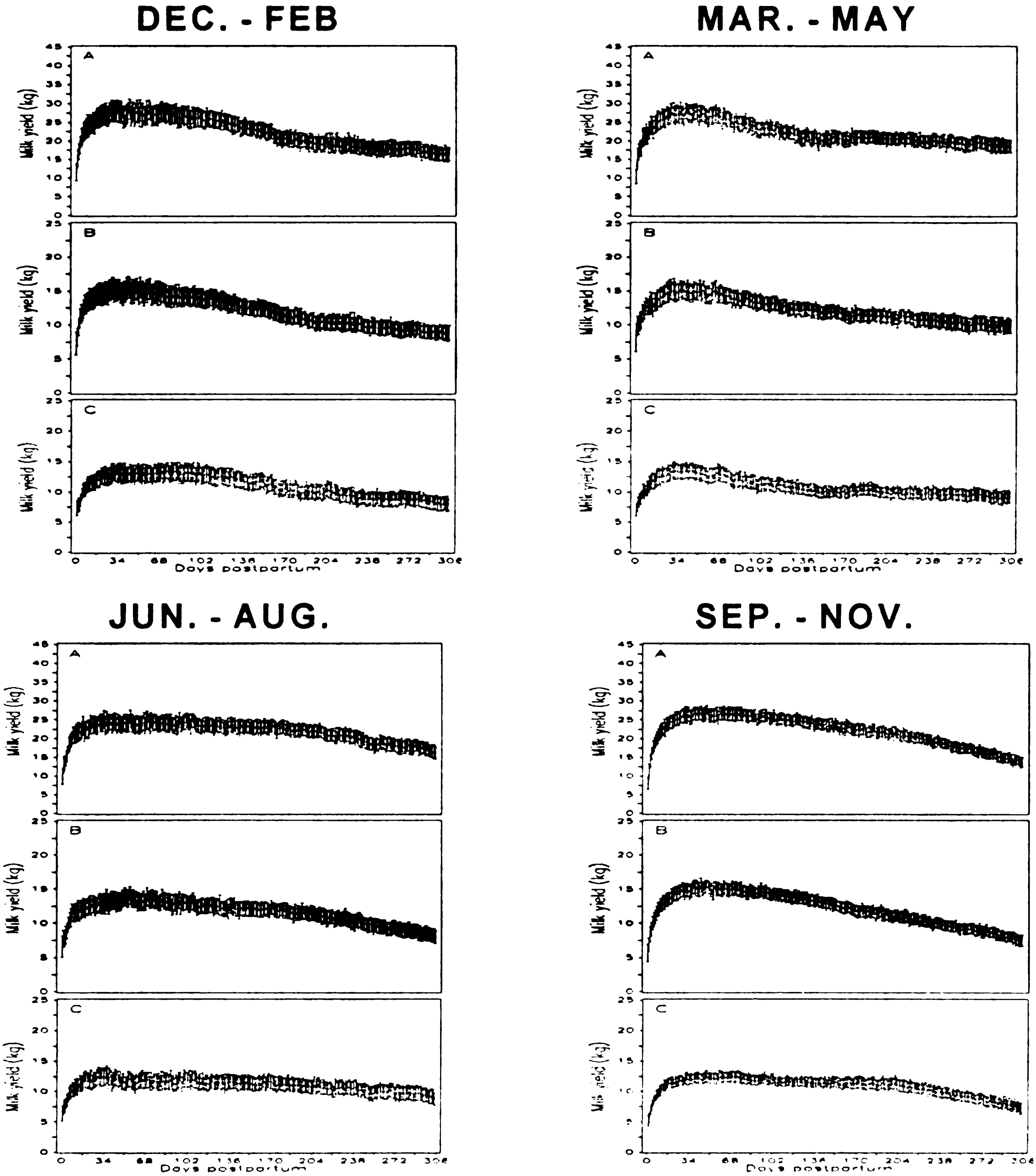


Figure 12. Estimated 95% confidence interval of mean daily milk yield within calving season for Florida, USA: A = total yield, B = morning yield, and C = evening yield.

6. MATHEMATICAL DESCRIPTION OF LACTATION CURVES IN DAIRY CATTLE.

6.1 Abstract

Tools for fitting mathematical functions to various shapes of lactation curves of dairy cows were established. A method for generating sample lactation curves from the functions was described. This method transcends any method that approximates the shape and variance of the lactation curve from a single mathematical model. The method was demonstrated by fitting ten curves for mean yield and four variance curves for mean yield during lactation.

6.2 Introduction

Frequent applications of the lactation curve include extension of records in progress to predict yield up to a standard length, calculation of lactation totals, and generation of yields for simulation studies. Lactation curves also can be used on a routine basis to monitor the health and progress of individual animals. Therefore, a clear, concise, and accurate mathematical description of the lactation curve is imperative. Such a description must be representative of different breeds, regions, seasons of calving, age at calving, days open, days dry, stage of pregnancy, stage of lactation, BST treatment, and other environmental factors.

An early attempt to develop a model which would describe the lactation curve was made by Gaines (1927). He proposed the formula $Y_t = A\exp(-Kt)$ in which Y_t is the yield in month t, A is the starting yield (when t = 0) and K is the rate of decline per month of lactation. This expression however makes no attempt to explain the initial rise in production, a portion of the curve which is of extreme importance.

Vujicic and Bacic (1961) suggested a modification of the formula of Gaines. They proposed the expression $Y_t = tm^{-a}\exp(-mt)$, where Y_t is the yield in period t while a and m are parameters. Unlike Gianes' formula, the expression proposed by Vujicic and Bacic varies both direc

early

N

$Y_t =$

wh

yie

in

$n^{}$

$=$

i

T

c

directly as exponentially with time. Thus, it accounts for the initial rise in production during early lactation.

Nelder (1966) described a family of inverse polynomial curves of the form, $Y_x = (b_0/x + b_1 + b_2 x)^{-1}$ where Y_x is the yield in week x and $b_0, b_1, \text{and}\, b_2$ are constants which could be estimated by Ordinary Least Squares (OLS). Under this model, maximum yield occurs when $x = (b_0 / b_2)$. The maximum yield is equal to $(2/(b_0 b_2) + b_1)^{-1}$.

Using a similar approach, Wood (1967, 1969, 1972) described the lactation curve as the incomplete gamma (IG) function $Y_n = an^b \exp(-cn)$ where Y_n is the average daily yield in the n^th week of lactation and a, b, and c are constants. The equation reaches a maximum when n = b/c. The expected maximum yield is $a(b/c)^b \exp(-b)$. The exponential form accounted for 95.4% of the variation in month yield as opposed to 84.4% for the inverse polynomial. The Wood's IG function has been the most widely used because of the flexibility of fitting curves of different shapes by estimating the three curve parameters.

Grossman et al. (1986) extended the IG equation by including sine and cosine terms to account for seasonal variations other than season of calving, $Y_n = an^b \exp(-cn)(1 + u\sin(x) + v\cos(x))$ where x is the day of the year in radians and all other terms are as defined previously. This extended function gave a better OLS fit than the inverse polynomial function (Batra, 1986). In order to account the different stages of lactation, Grossman and Koops (1988) proposed a multiphasic function,

$Y_t = \sum_{i=1}^{n} a_i b_i \{1 - \tanh^2(b_i(t - c_i))\}$, where Y_t is milk yield at time t (t = days in milk); n is the number of lactation phases; tanh is the hyperbolic tangent; a_i, b_i, c_i are parameters for the i^th phase. A diphasic function was found to be sufficient to describe the lactation curve. This function accounts for smaller and more random residuals and provides easily interpretable parameters that have biological importance. In addition, linear (Schaeffer and Dekkers, 1994; Rowlands et al., 1982; Stanton et al., 1992), non-linear (Freeze and Richards, 1992; Schaeffer et al., 1977) models to include shape, genetic and environmental effects have been proposed.

A

mate

becau

dim

curv

cur

err

6

c

(

All of the above mathematical expressions and models require raw data in order to estimate the parameters of the lactation curve and they diverge from typical lactation curves because shapes of lactation curves differ for over a wide range of ages, parities, regions, climates, and breeds. For empirical studies, the hitherto approach of relying on estimates of curve parameters in order to accommodate factors which affect the shape of the lactation curve is not only unsuitable but prohibitive. The objective of this study was to propose mathematical functions which describe typical shapes of the lactation curve in dairy cattle.

6.3 Materials and methods

Ten curves that were representative of typical lactation curves were identified. These curves as shown in Figure 13. Two curves for morning and evening yield also were identified (Figure 14). Observed from the literature were mean curve 1 (Grossman and Koops, 1988), curve 2 (Mainland, 1985), and curve 3 (Congleton and Everett, 1980). Mean curves 4, 5, 6, 7, 8, 9, 10, 11, and 12 were perceived from a preliminary analysis of data from a herd in Michigan.

Four curves that were representative of the variation in mean daily yield during lactation were identified. These variance curves, shown in Figure 15, were observed from preliminary analysis of data from a herd in Michigan. All the mean and variance curves were endorsed by the "Lactation Computation Working Group" of the International Committee of Animal Recording as representative of all breeds, regions, parities, ages at calving, and seasons of calving, and other grouping effects.

A collection of approximately 40 basic mathematical functions from the literature (Abramowitz and Stegun, 1965; Papajcsik and Bodero, 1988) were used to handcraft the functions for the curves in Figure 13, 14, and 15. To reduce random noise, the 'moving average' approach was used to smooth mean curves 1, 2, 3, 4, and 5; and variance curve 1. This was achieved by defining the kernel estimator,

$$k(y-x) = \exp^{-200\left(\frac{i}{305}\right)^2}; \qquad i,\cdots,305$$

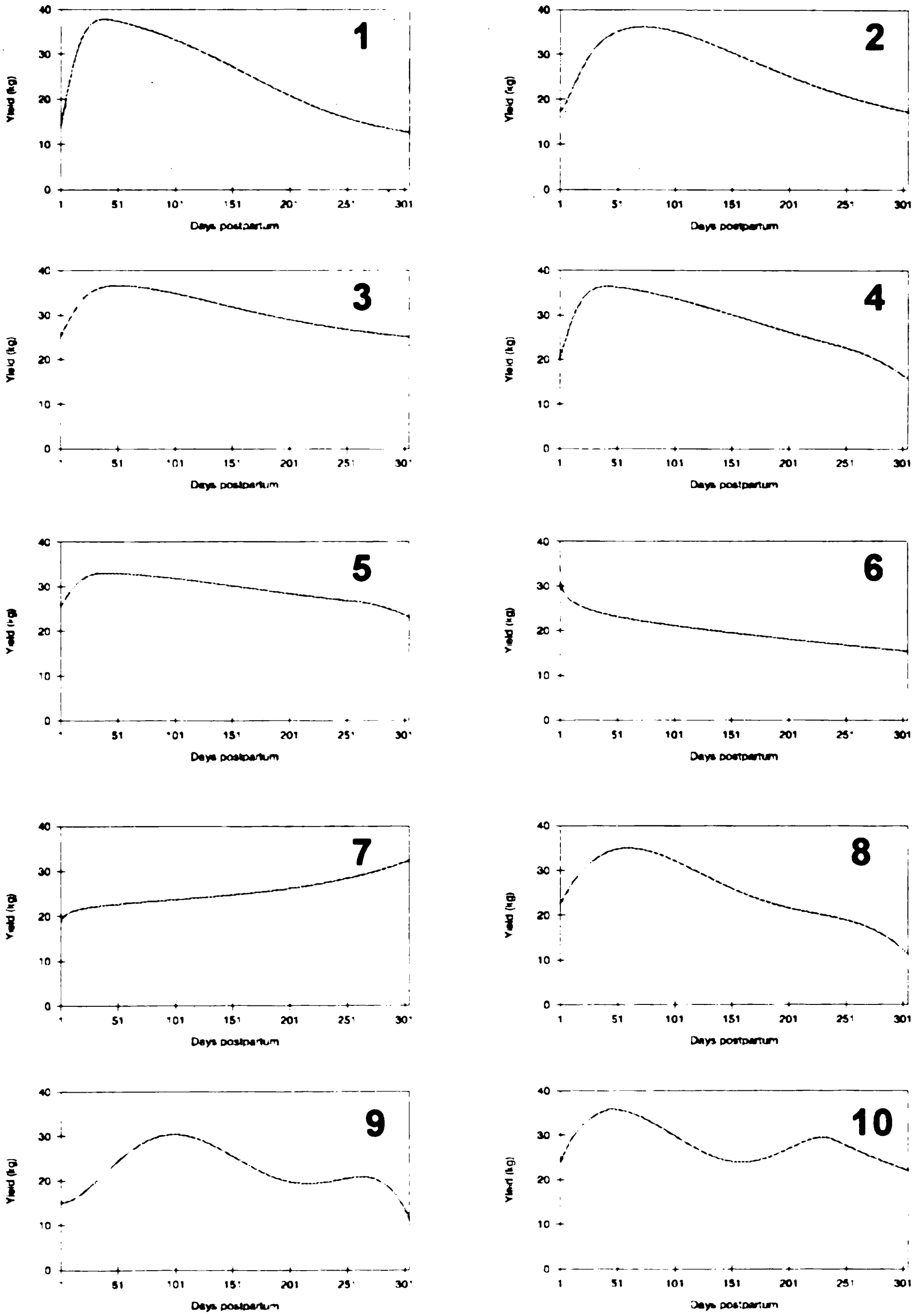


Figure 13. Lactation curves for daily yield.

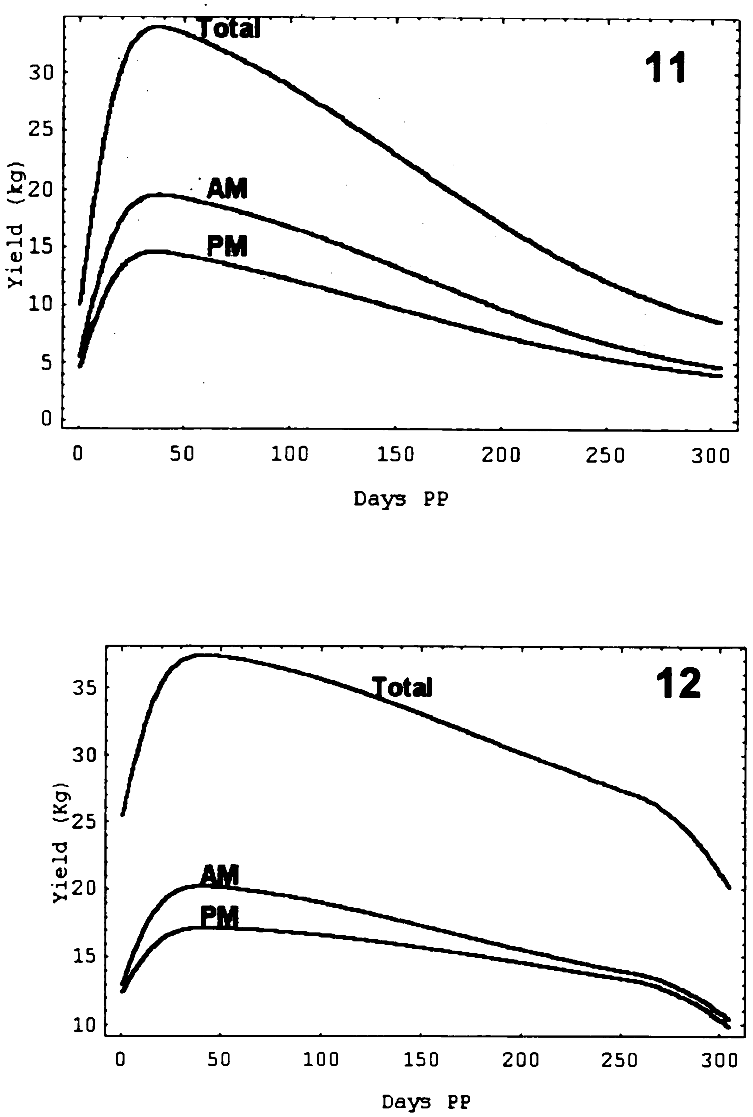


Figure 14. Lactation curves for morning (AM) and evening (PM) yield.


Variation (kg)
110
100
90
80
70
60
Variation (kg)

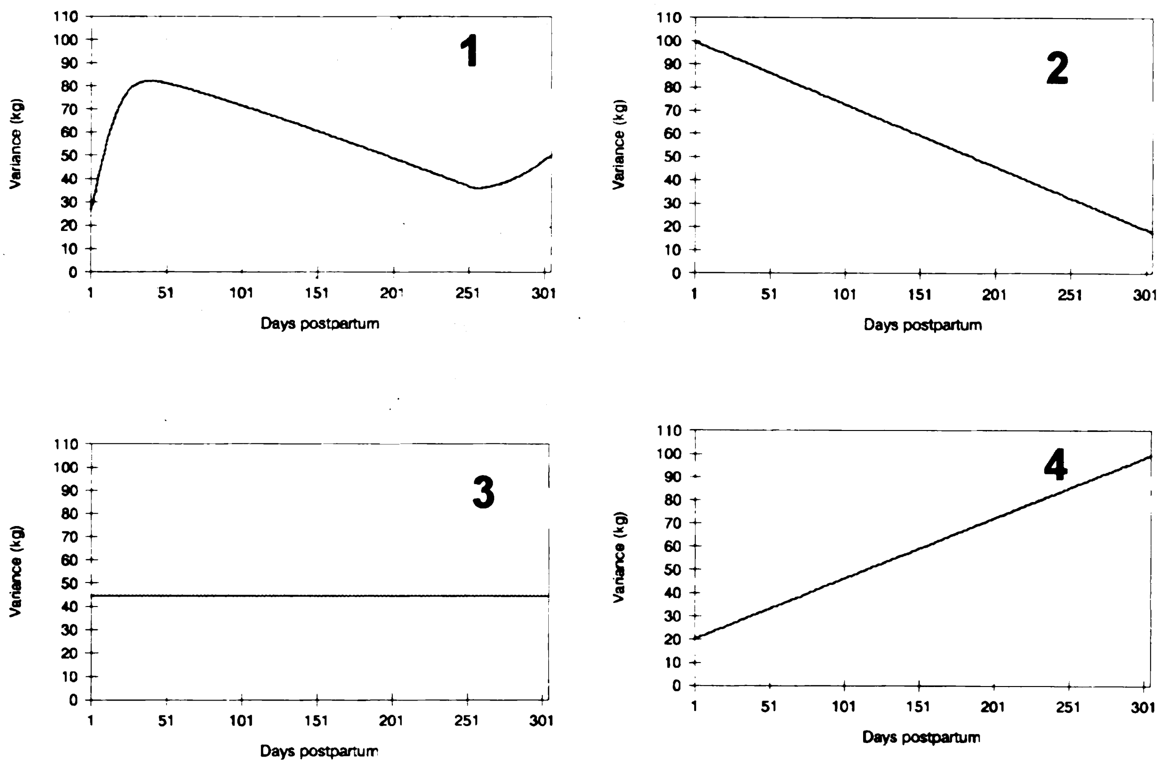


Figure 15. Variance curves for mean daily yield during lactation.

The unifying integral for the convolution was

$$f(x) = \int f(y)\,k(y-x)dy$$

and

convolution = InverseFourier[Fourier[**f(x)**] Fourier[**k(y-x)**]]

where $f(y)$ is the raw data for the curve. All analyses were performed using Mathematica (Wolfram, 1988).

6.4 Results and discussion.

The mathematical functions for the mean and variance curves were as follows:

Mean curve 1:

$$f(x) = 14.7819\,(1 - \mathrm{Tanh}[.008471\,(x - 150.656)]); \quad x, 1, \ldots, 305;$$

Mean curve 2:

$$f(x)=10x^{252}\,\mathrm{Sech}[.009x]; \quad x=1,\ldots,305;$$

Mean curve 3:

$$f(x)=10x^{102}\,\mathrm{Sech}[.009x]; \quad x=1,\ldots,305;$$

Mean curve 4:

$$f(x)=\begin{cases} y=10+24\,i^{008}\,\mathrm{Sech}[.006i]; & i=1,\ldots,250,\ 1\leq x\leq 250 \\ y=23-.06i-.001\,i^{2}; & i=6,\ldots,60,\ 250<x\leq 305; \end{cases}$$

Mean curve 5:

$$f(x)=\begin{cases} y=20+11\,i^{.005}\,\mathrm{Sech}[.006i]; & i=1,\ldots,250,\ 1\leq x\leq 250 \\ y=26.8-.001\,i^{2}-.01\,\mathrm{Log}[i]; & i=6,\ldots,60,\ 250<x\leq 305; \end{cases}.$$

Mean curve 6:

$$f(x)=30+\mathrm{Exp}[-.07x]-.02x-1.5\,\mathrm{Log}[x]; \quad x=1,\ldots,305;$$

Mean curve 7:

$$f(x)=20+.3\,\mathrm{Exp}[.011x]+.007x^{.14}+.9\,\mathrm{Log}[.2x]; \quad x=1,\ldots,305;$$

Mean curve 8:

$$f(x)=22.1539+.500083x-.00620379x^{2}+.000025815\,6x^{3}+3.68262*10^{-8}x^{4}; \quad x=1,\ldots,305\cdot$$

Mean curve 9:

$$f(x)=15+.00019x+.00731x^{2}-.00009x^{3}+3.724*10^{-7}x^{4}-5.1379*10^{-10}x^{5}; \quad x=1,\ldots,305\cdot$$

Mean curve 10:

$$f(x)=\begin{cases} y_1=5+10.5x^3\,\mathrm{Sech}[.009i]; & i=7,\ldots,50 \quad x=1,\ldots,43 \\ y_2=263+.0002i+.0059i^2-.0001i^3+3.62*10^{-7}i^4-5.228*10^{-10}i^5; & i=85,\ldots,283 \quad x=44,\ldots,243 \\ y_3=10.08+13i^3\,\mathrm{Sech}[.009i]; & i=244,\ldots,305 \quad x=244,\ldots,305 \end{cases}$$

Mean curve 11:

$$y_{am}=2.5+9.30139\,(1-\mathrm{Tanh}[0.007471\,(-150.656+x)]); \quad x=1,\ldots,305$$
$$y_{pm}=2+7.61007\,(1-\mathrm{Tanh}[0.006162\,(-130.546+x)]); \quad x=1,\ldots,305$$

Mean curve 12:

$$y_{am}=\begin{cases} y_1=8.8+11i^{.005}\,\mathrm{Sech}[.006i],\ i=1,\ldots,250; & 1\leq x\leq 250 \\ y_2=y_{1,(250)}-.001i^2-.01\,\mathrm{Log}[i],\ i=6,\ldots,60; & 250<x\leq 305 \end{cases}$$

$$y_{pm}=\begin{cases} y_1=5+11i^{.005}\,\mathrm{Sech}[.004i],\ i=1,\ldots,250; & 1\leq x\leq 250 \\ y_2=y_{1,(250)}-.001i^2-.01\,\mathrm{Log}[i],\ i=6,\ldots,60; & 250<x\leq 305 \end{cases}$$

Variance curve 1:

$$f(x)=\begin{cases} y=35-.2i+.4\,\mathrm{Log}[i];\ i=1,\ldots,255, & 1\leq x\leq 255 \\ y=8.5+\dfrac{.001i^2}{\mathrm{Exp}[-1.9i]};\ i=5,\ldots,54, & 255<x\leq 305 \end{cases}$$

Variance curve 2:

$$f(x)=100-.27x+.002\,\mathrm{Log}[.02x]; \quad x=1,\ldots,305$$

Variance curve 3:

$$f(x)=44.6652; \quad x=1,\ldots,305$$

Variance curve 4:

$$f(x)=20+.26x; \quad x=1,\ldots,305.$$

Alternatively, the piecewise functions could have been fitted using: 1) polynomial cubic splines with several knots (Mathews, 1987); 2) fourier series and trigonometric polynomials (Mathews, 1987); 3) the bounded Riemann Zeta-function (Karatsuba and Voronin, 1992); 4) simple differential equations (O'Neil, 1983) such as

$$dX = \frac{a_x}{X} dt + \sigma_x^2 dw$$

where a and σ^2 are parameters which could be modified as required.

Lactation curves for empirical studies can be obtained from these curves by the Monte-Carlo approach of applying the variance curves to each mean curve. Figure 16A illustrates this for mean curve 8 and variance curve 1. Daily yields can be generated by random sampling of real values between the limits determined by the variance. Figure 16B shows 5 sample lactation curves generated by this approach. An individual lactation curve is shown in Figure 16C. As expected, the curves are sensitive to the shape of the variance. Increasing the sample size guarantees full representation of the variation in yield. Because these curves are comprised of linear and nonlinear trajectories, no single function or mathematical model can be used to generate all of them.

6.5 Conclusions

Mathematical functions for 12 mean curves and 4 variance curves for lactation yield in dairy cattle were determined. These curves were representative of curves that are observable in the various dairy cattle populations globally. The vastly different shapes can best be fitted by mathematical functions such as those described. A method for generating a sample lactation from these functions was established. This method generated lactation curves that were more representative than those that could be generated from a single mathematical model that relies on estimation of curve parameters such as those in Wood's (1967) equation. These functions also can be used as tools for instruction in dairy production courses.

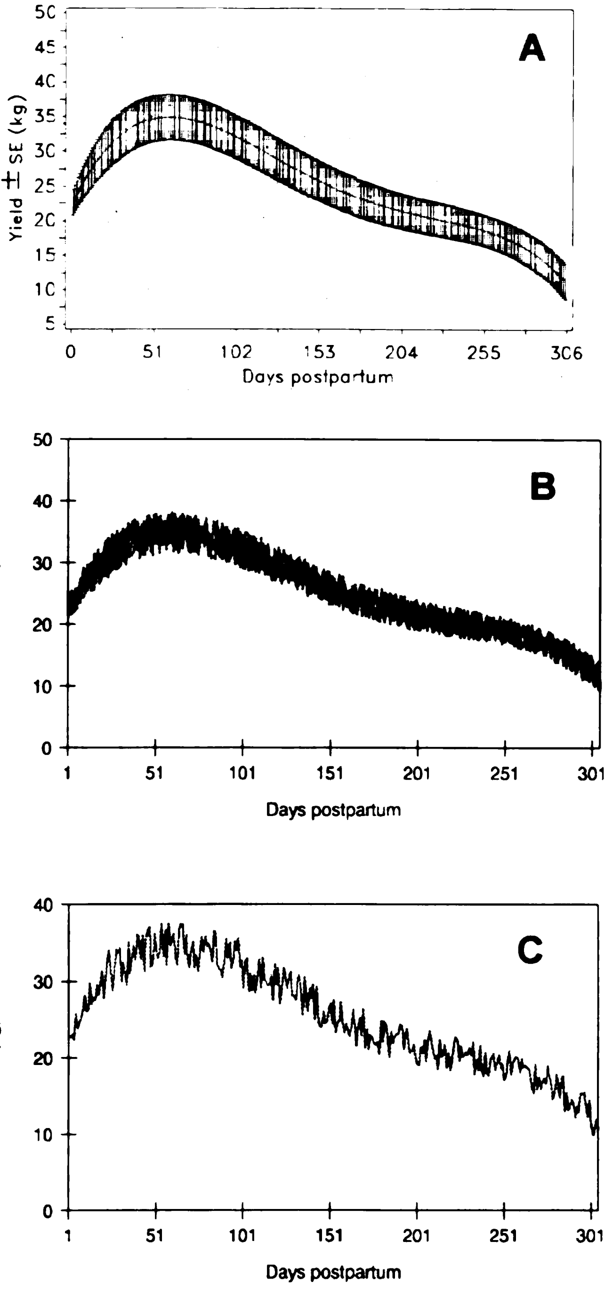


Figure 16. Mean ± standard error for lactation yield (A), 5 sample lactation curves (B), and 1 lactation curve (C).

7. DESCRIPTION OF THE STANDARD LACTATION CURVES (SLAC) PROJECT

7.1 Summary

The data files of standard lactation curves (SLAC) were generated. The SLAC consists of test-day yields for a total of 1,126,080 lactation records representing five testing schemes, six patterns of missing test-day yields, varying starting days of recording postpartum from 3600 sample lactations of various shapes of lactation curves. Collaborating Data Processing Centers (DPC) were asked to calculate the lactation totals of these lactations using their operating procedures. The calculated lactation totals were sent back to Michigan State University (MSU) for analyses and summarization.

A total of 40 parameter curves were designed by combining ten curves of different shapes for mean daily yield in a lactation with four curves for variance of daily yield. For each of the 40 parameter curves, 90 sample curves were simulated to generate a total of 3,600 sample curves. True lactation total yields of the sample curves were calculated and kept at MSU.

Four test schemes of A1, A4, A6, and A8 were applied to each of the sample curves. Only one starting test-day postpartum was considered for the A1 scheme, but four starting days were considered for the other three schemes. For each of the simulated curves under a specific test scheme with a specified starting day postpartum, up to five patterns of missing test-day records were established. The patterns in which missing records occurred at the end of a lactation were intended to simulate incomplete lactations which will need to be "extended" in order to obtain 305-d yields. After application of different test schemes with different starting days and patterns of missing test-days on the 3600 sample curves, there are test records on a total of 136,800 lactations.

Mean daily AM yield curves and the corresponding mean PM yield curves were designed for each of two mean daily total yield curves, and two variance curves. From a composite of simulated AM and PM data, 90 sample curves on daily total yields were generated for each of the four parameter curves to give a total of 360 sample curves. The AP/4 test scheme was applied to these curves by taking monthly test records alternately from the AM and PM test results. After the application of the four starting test-days postpartum and the patterns of missing test day records, there were AP/4 records on a total of 3,960 lactations. The total number of lactations was now 140,760.

Each simulated lactation was coded for one trait, one breed, one of two parities, and one of four seasons of calving. Each cooperating DPC had to designate the codes to fit categories of its choice before processing the SLAC data for total lactation yields. The DPC was asked to inform MSU of its designations. For SLAC, we assumed that all codes (1x1x2x4=8 combinations) would be designated, thus, each of the 140,760 lactations were repeated 8 times to give a total of 1,126,080 lactations.

The SLAC data file on 1,126,080 lactations was partitioned into five parts by test schemes and was stored in a fixed format. Column positions and codes used for the files are described and a printout of the first three records of each file are shown in Appendix C. The original 103 MB data file was packed into a size of 5 MB. Instructions for retrieving SLAC by DPC with FTP service were provided. Transfer with IBM PC compatible floppy disks was a popular alternative as most DPCs did not have FTP services.

Collaborating DPCs were asked to send the yield totals calculated from lactations in SLAC back to MSU. The format and a sample printout of the first three records of each of the five return files are shown in Appendix D. Alternative means for transfer/retrieval were also suggested.

7.2 Design of lactation curves for daily total yield

A total of ten curves of different shapes for mean daily yield in a lactation and four curves for variance of daily yield were designed. These curves are shown in Figure 13 and 15, respectively.

a) Combining each of ten mean curves with each of four variance curves gives a set of 10 x 4 = 40 curves (see illustration in Figure 16).

b) Mathematical functions were established for each of the 10 mean curves. Then for each of the mean curves, 90 sample curves were simulated according to the assigned variance curve. A total of 3,600 curves was generated. For each of the curves, true lactation yield was calculated by accumulating daily yields. The true lactation yield data was kept at MSU for analysis purposes. Note that the optimal size required to compute a significant difference between mean biases at the 99% level was found to be 90.

c) For each of the simulated curves, four test schemes with different lengths of interval (A1, A4, A6, and A8) were applied. A1 stands for a 1 week interval between milk recording, A4 stands for a 4 week interval recording, A6, stands for a 6 week interval between recording, and A8 stands for an 8 week interval between recording.

Starting day of recording was 7d postpartum for A1 scheme, while there were four (4) different starting days, 7d, 14d, 21d, and 28d postpartum, applied to A1, A4, A6 schemes. The number of test-day records for each of the combinations of test schemes by starting day is shown in Table 4.

TABLE 4. Number of records in a complete lactation by scheme and starting day for milk recording.

	Starting day			
Scheme	7d pp	14d pp	21d pp	28d pp
A1	43	—	—	—
A4	11	11	11	10
A6	8	7	7	7
A8	6	6	6	5

d) For each of the simulated curves under a specific test scheme and starting day, up to five (5) patterns of missing test-day records were established. They are shown in Table 5 below and a graphical representation is in Appendix B.

The number of lactation curves before (not in parentheses) and after (in parentheses) applying schemes, starting days, and missing patterns is shown in Table 6.

TABLE 5. Weeks in which data were missing by scheme and starting day of milk recording.

	Starting day			
Scheme	7d pp	14d pp	21d pp	28d pp
A1	1, ..., 3 6, ..., 9 10, ..., 13 26, ..., 43 36, ..., 43	—	—	—
A4	9, ...,13	6, ..., 10	3 7 11, ..., 15 27, ..., 43 35, ..., 43	—
A6	13	—	3 9 15 27, ..., 39 33, ..., 39	—
A8	9	—	3 11 19 27, ..., 43 35, ..., 43	—

TABLE 6. Number of curves by scheme, starting day, and missing pattern

	Starting day				
Scheme	7d pp	14d pp	21d pp	28d pp	Total
A1	3600+(3600x5)=21600				21600
A4	3600+(3600)=7200	3600+(3600)=7200	3600+(3600x5)=21600	3600	39600
A6	3600+(3600)=7200	3600	3600+(3600x5)=21600	3600	36000
A8	3600+(3600)=7200	3600	3600+(3600x5)=21600	3600+3600=7200	39600
Total	43200	14400	64800	14400	136800

7.3 Design of lactation curves for AM and PM yield

Mean daily AM yield curves and the corresponding mean daily PM yield curves were designed for each of the two mean daily total yield curves. Two shapes of variance curves were chosen. The two mean curves are shown in Figure 14 and the two variance curves are shown in Figure 15 as curves 3 and 4.

a) Combining each of the yield curves with each of the two variance curves gave a set of 2 x 2 = 4 curves.

b) For each of 4 curves, a total of 90 sample curves were simulated for a total of 360 curves. For each of the simulated curves, true daily yield was calculated by summing the corresponding AM and PM daily yields, and true lactation yield was calculated by summing the true daily yields.

c) For each of the simulated AM/PM curves, the AP/4 test scheme was applied by taking the yield on AM curve at week 1, 9, 17, 25, 33, and 41 and the yield on the PM curve at week 5, 13, 21, 29, and 37 (see illustration in Appendix A).

Also applied were four starting days of recording 7d, 14d, 21d, and 28d postpartum.

For each simulated curve, patterns of missing records were established in a manner similar to that for A4. The total number of curves for AP/4 is shown in Table 7.:

TABLE 7. Number of curves for AP/4 by starting day and missing pattern.

	Starting day				
Scheme	7d pp	14d pp	21d pp	28d pp	Total
AP/4	360+360=720	360+360=720	360+(360x5)=2160	360	3960

The total number of AP/4 lactation curves was 3,960 and the combined total number of lactation curves for both daily total yield and AP/4 was (136,800+3,960)=140,760.

7.4 Designation of trait/breed/parity/calving season categories

Each simulated lactation included a trait designation, a breed designation, and designations for parity and calving season categories. Each of the Data Processing Centers needs to fit their designation codes to the codes described below and, in the process, to choose the specific categories in order to capture the most popular categories. In every case, the Data Processing Center was required to inform MSU of the category definition of the chosen designation codes.

Designation codes were required for the following categories:

(1) Trait: Only "1" for milk yield;

(2) Breed: Only "1" for one cattle breed. We suggested that the breed be Holstein/ Friesian, which was the case for Italy and Canada for example. For Norway, the breed could be Norwegian cattle. For Switzerland, the breed could be either Simmental, Brown or Holstein;

(3) Parity: Up to two categories. Code "1" denoted either first parity or all parities, while

code "2" denoted either second or second plus later parities. For U.S.A., for example, "1" was for the first parity and "2" for second plus later parities. If the algorithm of a Data Processing Center did not distinguish between parities, the DPC used "1" to denote all parities.

(4) Season: For season of calving, up to four categories. For US, "1" denoted December through February, "2" was for March through May, "3" June through August, and "4" represented September through November. If a Data Processing Center considered only two calving seasons, for example, the designation would be either "1" or "2" only.

The purposes of such designations were:

(1) To ensure that the record format contains basic information necessary for inputting the data into the algorithm/program for calculating total yields at a Data Processing Center, and

(2) To suggest specific categories and a specific number of categories in order for a Data Processing Center to choose specific algorithm/programs for those specific categories.

(3) To achieve a reasonable degree of standardization, because of the great variety of category designations among Data Processing Centers. The number of categories for each designation was kept at a minimum in order to keep the data size manageable.

It was assumed that the maximum number of the above categories (1,2,3, and 4) would be chosen. Thus each of the 140,760 curves was repeated 8 (1x1x2x4=8) times. The total number of curves was therefore 1,126,080. The combinations of categories are shown in Table 8.

TABLE 8. Combinations of fixed effects for calculating lactation totals.

Combination code (1,...,8)	Lactation (1,2)	Breed (1)	Trait (1)	Season (1,...,4)
1	1	1	1	1
2	1	1	1	2
3	1	1	1	3
4	1	1	1	4
5	2	1	1	1
6	2	1	1	2
7	2	1	1	3
8	2	1	1	4

7.5 Data files of standard lactation curves (SLAC)

The 1,126,080 curves made up the SLAC. The data were stored in a fixed format. The record format in SLAC was:

Mean curve No.; Variance curve No.; Replicate No.; Scheme; Starting day; Missing pattern; Combination code for fixed effects; Days PP; Test-day yield.

The SLAC was partitioned into five (5) parts by test schemes. An exact description of column positions and codes used for the files and a printout of the first 3 records of each file are shown in Appendix C.

This SLAC was made available in one of the following two ways to the Data Processing Center (DPC) of each of the ICAR member which agreed to collaborate:

Choice 1: DPC's with FTP service could retrieve SLAC via Guest FTP to 35.8.124.45 (Guest login password was: *slac*). Files could be transferred from the sub-directory *\pub\slac\outgoing*. All files were stored in an archived format. Instructions for transfer/

retrieval of the archived files were contained in the file *\pub\slac\outgoing\readme.1st* and are included herein as Appendix E.2. Problems with file transfer were addressed to saama@ msu.edu.

Choice 2: DPC's without FTP services could receive SLAC on IBM PC compatible 1.44MB 3½" or 1.2MB 5¼" disk media with installation instructions included herein as Appendix E.1.

7.6 Data files of calculated lactation totals to be returned to MSU

These calculated lactation totals by the DPC will be sent to Michigan State University in the corresponding five (5) parts by test scheme with the format below:

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (1, ..., 10)
Numeric	1	3	Variance curve No. (1, ...,4)
Numeric	3	4-6	Replicate No. (1, ...,120)
Numeric	1	7	Scheme (1=A1, 2=A4, 3=A6, 4=A8, 5= AP/4)
Numeric	1	8-9	Starting day (7=7d PP, 14=14d PP, 21=21d PP, 28=28d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5=tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	5	12-16	Total yield (kg) in Integer format.

The data could be returned to MSU using only one of the following methods:

Choice 1: Guest FTP to 35.8.124.45 (Guest login password was: *slac*). Files could be placed in the sub-directory *\pub\slac\incoming*. A description of the files transmitted was required. After transmission, DPC's were asked to send an E-mail message to saama@msu.edu. The E-mail message had to include data source and a list of the files transmitted to MSU.

Choice 2: IBM PC compatible 1.44MB 3½" or 1.2MB 5¼" disk (Provide disk catalog).

Choice 3: ASCII or EBCDIC tape (Provide tape catalog).

A sample printout of the first 3 records of the files for the corresponding 5 parts is shown in Appendix F. Note that the total yield shown for each record in the sample printouts is solely for the purpose of illustration.

7.7 Participating data records processing centers

A request for participation in the SLAC project was solicited by the Secretary Generalof ICAR. A respose was received at MSU from Australia, Italy, the Netherlands, New Zealand, USA, Denmark, Jersey (England & Wales), France, Switzerland, Germany, Austria, and Mexico. Subsequently, the SLAC was sent to each of these countries. Results for the test-interval method were computed at MSU.

7.8 Data files for computed lactation totals returned to MSU

Data files of computed lactation totals were returned from USA, Netherlands, Denmark, Italy, and France. As a result of in-house changes in the method for computing lactation totals, Germany, New Zealand, and Australia decided against computing or returning data files. These radical changes were not atypical of the DPC's but some may have been motivated by preliminary SLAC results presented to the working group on "Lactation computation and related matters" at the 29th session of the ICAR General Assemby meetings in Ottawa,

8. COMPARISONS OF DIFFERENT METHODS OF ESTIMATING 305-D MILK YIELD

8.1 Abstract

In order to evaluate the relative accuracy of methods for calculating total yield in a lactation, a data set of standard lactation curves was generated. Standard lactation curves were designed. Replicates by simulation, test schemes, missing data patterns and starting day of recording were empirically imposed on each sample standard curve. Separately, morning and evening lactation curves were designed for an alternating testing scheme. The total number of sample test-day records was 1,126,080. A total of six methods for calculating lactation totals from test-d yields were compared. They were methods of centering date ignoring missing test-day data, test interval without adjustment factors but correction for missing data, test interval without adjustment factors and no correction for missing data, test interval with adjustment factors, interpolation with standard curves, and multiple-trait projection. The differences between actual and calculated lactation yields were analyzed within each method by fixed classification models. Factor analysis using the squared multiple correlations of the methods as priors was conducted. Main effects of method, scheme, patterns of missing yields, and starting day of recording postpartum and two way interactions between scheme and shape of lactation curve, interactions between scheme and starting day of recording and, interactions between scheme and pattern of missing test-day records were significant ($P < .0001$). When no yields were missing, the overall prediction bias for all methods was generally small. Within shape of lactation curve, variability in the accuracy of the methods was evident. With missing test-day records and varying starting days of recording, some methods had smaller bias than others.

8.2 Introduction

Test-day records in a lactation are the basic unit of information on a cow's production for management decisions and genetic evaluation. Lactation totals are calculated from information collected on test-days. The test-day yields are most commonly recorded at monthly intervals, which are then used to estimate a cow's lactation yield over a standard lactation period, and the convention is to use 305 days. Some of the officially approved testing schemes by the International Committee of Animal Recording (ICAR) are A4, A6, and A8 where the A denotes testing schemes in which data collection, handling, and transferring is done by authorized technicians at 4, 6, and 8 wk intervals.

To calculate 305-d yield, typically, a continuous lactation curve is simulated by linear interpolation between test-d records. This approach is weakened by non-linearity in the shape of the lactation curve at the beginning, peak, and end of lactation. To account for this weakness, various methods have been developed (Wood, 1972; Schaeffer et al., 1977; Grossman et al., 1986; Shook et al., 1980; Wilmink, 1987).

The projection of incomplete lactations to 305 d also provides useful information for both management and genetic evaluation. Factors have been estimated that can be used to extend partial lactation records (Batra and Lee, 1985; Keown et al., 1986; Wiggans and Van Vleck, 1979; Wiggans, 1981; Wiggans, 1986). Most of these factors were estimated from empirical relationships between cumulative and last test-d yields and lactations from animals with complete lactation records (Shook et al., 1980, Palmer et al., 1994), but some used factors calculated from lactation curves (Schaeffer et al., 1977; Wilmink, 1987).

Different methods for computing lactation totals are being used in different regions and countries. These methods are at various levels of sophistication and produce estimates of lactation totals with different degrees of precision and accuracy. With the increased need for pooling data bases and the tremendous growth in the exchange of germplasm between countries, a set of standard methods is needed. Such methods must be flexible enough to suit different shapes of lactation curves due to breed and management differences, different intervals of sampling and different rules of screening records in different schemes, different patterns of missing test-day records, different starting days of milk recording, and different production traits. To evaluate, summarize, and compare the precision and accuracy of current methods, a "Lactation Computation" working group, consisting of 12 member countries, was established by ICAR. This study was commissioned by ICAR to establish the relative precision and accuracy of methods for computing 305-d lactation yields. Twelve countries participated in the computation of lactation totals.

8.3 Materials and methods

A test-day data set on standard lactation curves (SLAC) was generated.

8.3.1 Design of SLAC

A total of 40 parameter curves were designed for 10 curves of different shapes for mean daily yield in a lactation by 4 curves describing variance of daily yield throughout a lactation. For each of the parameter curves, 90 sample curves were simulated to generate a total of 3,600 sample curves. True lactation yield of the sample curves were calculated. Four test schemes of A1, A4, A6, and A8 were applied to each of the sample curves. Only one starting day postpartum (7 d) was considered for scheme A1, but four starting days (7, 14, 21, and 28 d) were considered for the other three schemes. For each of the simulated curves under a specific test scheme with a specified starting day postpartum, up to six patterns of missing test-d yields (none, early, early peak, late peak, late, and tail-end of lactation) were established. After application of different test schemes with different starting days and patterns of missing test-d yields on the 3600 sample curves, there were test-d records on a total of 136,800 lactations.

Mean daily morning and the corresponding evening yield curves were designed for each of two mean daily total yield curves, for each of which two variance curves were designed. Again, 90 sample curves were generated for each of the four parameter curves to give a total of 360 sample curves. An AP/4 test scheme was then applied to these curves by taking alternate morning and evening test-day yields from one month to the next. After the application of the four starting test-days postpartum and patterns of missing test-day records, there were AP/4 records on a total of 3,960 lactations. The combined total number of lactations was now 140,760.

Each simulated lactation was coded for one trait, one breed, one of two parities, and one of four seasons of calving to accommodate the application of different methods for calculating lactation totals. Thus, each of 140,760 lactations was duplicated 8 times to give a total of 1,126,080 lactations. The SLAC curves were simulated using Mathematica (Wolfram, 1988).

8.3.2 Methods for calculating 305-d lactation yield

Six methods were compared:

1) Test interval method (TIM-A1: Wiggans, 1985; Wiggans and Dickinson, 1985) with prediction of missing test-day yields and 305-d projections using adjustment factors for parity, and season effects. No lactation totals were computed for lactations in scheme A6 and A8, and missing pattern of 'late lactation';

2) Linear interpolation using standard curves (ISC: Wilmink, 1987; Mimeo from Wilmink and Ouweltjes, March 1991, NAS-report, 91-0355/WO/CA, Netherlands Royal Syndicate, AL Arhem) with prediction of missing test-day yields and 305-d projections using standard yield curves;

3) Centering date method (CDM: Letter from O. K. Hansen, Danish Agric. Adv. Centre, Aarhus on 9/7/94) with no prediction of missing test-day yields but 305-d projections using adjustment factors for parity, and season effects. No lactation totals were computed for scheme AP/4;

4) Test interval method (TIM-U1: Appendix A of Int. Comm. of Anim. Recording Agreement) with an estimate for a single missing test-d yield and 305-d projections for lactations that were at least 300d but with no adjustments for parity or season effects. Interval credits for scheme AP/4 were calculated after doubling the sample test-d yields;

5) Test interval method (TIM-U2: Letter from Nicole Bouloc, Institut de l'elevage, Paris, France, on 11/8/94) with no adjustments for parity and season effects and no corrections for missing test-d yield. Interval credits for the last interval computed as,

$$((\text{Test-d yield} + \text{previous yield})/2 \times \text{interval length}) + (\text{test-d yield} \times c)$$

where c = 14 for schemes A1, A4, AP/4, and A6 or c = 28 for scheme A8. No lactation totals were computed for incomplete lactations;

6) Multiple trait projection (MTP: Trus and Buttazzoni, 1991; Mimeo from D. Trus, Ottawa, Canada, on 8/15/94) method with prediction of missing test-day yields and 305-d projections using expectations of parity and season sub-class means, and error estimates from the residual covariances among intervals (15). Lactation totals were computed for only schemes A4 and A6.

8.3.3 Statistical analyses

Bias in calculated lactation total yield from each of the above methods was expressed as

bias = (actual yield - estimated yield).

Thus, a positive mean bias was synonymous with overestimation while negative bias was indicative of underestimation.

For purposes of analysis, biases were stratified into three data sets in a manner that was consistent with the SLAC design:

1) Data set A - For general inferences about the methods, biases from lactation records with starting d of 7 d postpartum and no missing test-d yields were used.

Using data set A, crude means for bias were computed within each test scheme. Within shape of mean curve, sources of variation in bias were analyzed by the following linear models:

$$y_{ijklmn} = \mu + S_i + M_j + V_k + P_{l(j)} + T_{m(j)} + MV_{jk} + SM_{ij} + SV_{ik} + SP_{il(j)} + ST_{im(j)} + \varepsilon_{ijklmn} \qquad [1a]$$

where y_{ijklmn} is bias (kg); μ is a constant common to all observations; S_i is the fixed effect of scheme i with i=1, 2, ..., 4, with 1 = A1, 2 = A4, 3 = A6, 4 = A8; M_j is fixed effect of method j with j = 1, 2, ..., 6; V_k is fixed effect of variance curve k with k = 1, 2, 3, 4; P_l is the fixed effect of parity l with l = 1, ..., 4 with 1 = first, 2 = second, 3 = second and later, 4 = not used; T_m is the fixed effect of season of calving m with m = 1, ..., 7 with 1 = December to February, 2 = December to January, 3 = March to May, 4 = April to May, 5 = June to August, 6 = September to November, 7 = not used; $P_{l(j)}$ and $T_{m(j)}$ are nested effects and

MV_{jk}, SM_{ij}, SV_{ik}, $SP_{il(j)}$, $ST_{im(j)}$ are interaction effects between the respective factors; ε_{ijklmn} is the random residual error distributed as $N(0, I\sigma_e^2)$, where σ_e^2 was assumed to be homogeneous across all groups and zero covariances between groups. All other interactions were assumed to be negligible.

$$y_{jklmn} = \mu + M_j + V_k + P_{l(j)} + T_{m(j)} + MV_{jk} + \varepsilon_{jklmn} \qquad [2a]$$

where y_{jklmn} is bias (kg) for scheme AP/4; all terms are as defined in [1a].

2) Data set B - For inferences regarding missing test-d yields, biases from all lactation records for scheme A1, but only lactations with postpartum starting d of 21d for schemes A4, A6, A8, and AP/4. Within each shape of mean curve, sources of variation in bias were analyzed by the following linear models:

$$\begin{aligned} y_{ijklmqr} = {} & \mu + S_i + M_j + V_k + P_{l(j)} + T_{m(j)} + F_q + MV_{jk} + SM_{ij} + SP_{il(j)} + ST_{im(j)} \\ & + SF_{iq} + \varepsilon_{ijklmqr} \end{aligned} \qquad [1b]$$

where $y_{ijklmqr}$ is bias (kg); F_q is fixed effect of pattern of missing test-d yield, q = 1, 2, ..., 6, with 1 = None, 2 = early lactation, 3 = early peak, 4 = late peak, 5 = late lactation, 6 = tail end; SF_{iq} is an interaction effect between the respective factors. All other terms are as defined in [1a].

$$y_{jklmqr} = \mu + M_j + V_k + P_{l(j)} + T_{m(j)} + F_q + MV_{jk} + \varepsilon_{jklmqr} \qquad [2b]$$

where y_{jklmqr} is bias (kg) for scheme AP/4. All other terms are as defined in [2a].

3) Data set C - For inferences about starting day postpartum of recording, biases from lactation records for schemes A4, A6, A8, and AP/4 with no missing test-d yields were used. Within each method, sources of variation in bias were analyzed by the following linear models:

$$\begin{aligned} y_{ijklmno} = {} & \mu + S_i + M_j + V_k + P_{l(j)} + T_{m(j)} + D_n + MV_{jk} + SM_{ij} + SP_{il(j)} + ST_{im(j)} \\ & + SD_{in} + \varepsilon_{ijklmno} \end{aligned} \qquad [1c]$$

where $y_{ijklmno}$ is bias (kg); D_n is fixed effect of starting day postpartum, n=1, 2, 3, 4 with 1 = 7d, 2 = 14d, 3 = 21d, 4 = 28d postpartum; SD_{in} is an interaction effect between the respective factors. All other terms are as defined in [1a].

$$y_{jklmno} = \mu + M_j + V_k + P_{l(j)} + T_{m(j)} + D_n + MV_{jk} + \varepsilon_{jklmno} \qquad [2c]$$

where y_{jklmno} is bias (kg) for scheme AP/4. All other terms as defined in [3a].

To further delineate between the methods, principal factor analysis (Harman, 1960) using squared multiple correlations for the prior communality estimates was conducted across and within the schemes. Kaiser's measure of sampling adequacy (MSA) is a summary, for each variable and for all variables, of how much smaller the partial correlations are than the simple correlations (Cerny and Kaiser, 1977). Kaiser's MSA was computed (Cerny and Kaiser, 1977); values of .8 and .9 were considered very good, while MSA below .5 suggested sampling inadequacy. Orthogonal varimax prerotation (Harman, 1960) followed by oblique Procrustean rotation (Hendrickson and White, 1964) of the initial factor matrix was carried out. Harris-Kaiser's rotation (Harris and Kaiser, 1964) with Cureton-Mulaik weights (Cureton and Mulaik, 1975) was used to obtain an independent cluster oblique solution to the simple structure of the factors. All statistical analyses were conducted using SAS (SAS, 1992).

8.4 Results and discussion

8.4.1 Biases of different methods associated with test schemes on daily yield

The crude mean, standard deviation, standard error, mean : standard deviation, and coefficient of variation of bias by method of estimating lactation totals and schemes A1 and A4 are shown in Table 9 while data for schemes A6 and A8 are summarized in Table 10. For all methods, the mean bias did not increase with decreased sampling frequency. However, the standard deviation and standard error was higher under scheme A8 when compared with scheme A1 ($P < .0001$). The magnitude of mean : standard deviation followed the sampling frequency and was a reliable measure for ranking methods within each test scheme. The mean for TIM-A1 under schemes A4 and A8 was similar but larger than that under scheme A1; the mean under scheme A6 was smallest ($P < .0001$). The mean bias for ISC and TIM-U2 increased with decreased sampling frequency. However, the mean bias for TIM-U2 was smaller under scheme A8. The MTP method also showed an increase in the mean under scheme A6 when compared with scheme A4 ($P < .0001$). The mean for TIM-U1 under schemes A1 and A4 were close but increased under schemes A6 and A8. This reduced accuracy was attributed to the increase in the interval between recording.

TABLE 9. Mean, standard deviation, standard error, coefficient of variation, and mean : standard deviation of bias by method of estimating lactation totals and schemes A1 and A4.

	Scheme									
	A1					A4				
Method	$\bar{X}$	SD	$\bar{X}$/SD	SE	CV	$\bar{X}$	SD	$\bar{X}$/SD	SE	CV
	(kg)			(kg)	%	(kg)			(kg)	%
TIM-A1	53.4	29.2	1.8	.2	54.7	120.5	61.6	2.0	.4	51.1
ISC	-8.8	28.1	-.3	.1	-319.5	11.5	61.9	.2	.3	535.8
CDM	-10.1	28.2	-.4	.2	-279.4	8.7	61.4	.1	.4	706.2
TIM-U1	-10.1	28.2	-.4	.2	-279.4	8.7	61.4	.1	.4	706.2
MTP						43.6	120.8	.4	.7	277.2
TIM-U2	-29.3	67.9	-.4	.2	-231.4	30.9	140.9	.2	.3	455.3

CDM = Centering date method, ISC = interpolation with standard curves, MTP = multiple trait projection, TIM-A1 = test interval method, TIM-U1 = TIM without adjustments for fixed effects, and TIM-U1 = TIM without adjustments for fixed effects and no corrections for missing test-d yield.

TABLE 10. Mean, standard deviation, standard error, coefficient of variation, and mean : standard deviation of bias by method of estimating lactation totals and schemes A6 and A8.

	Scheme									
	A6					A8				
Method	$\bar{X}$	SD	$\bar{X}$/SD	SE	CV	$\bar{X}$	SD	$\bar{X}$/SD	SE	CV
	(kg)			(kg)	%	(kg)			(kg)	%
TIM	32.5	83.9	.4	.5	258.2	61.4	112.5	.5	.7	183.1
ISC	63.9	87.8	.7	.4	137.6	89.2	120.9	.7	.5	135.5
CDM	70.9	87.2	.8	.5	123.1	106.4	120.6	.9	.7	113.4
TIM-U1	70.9	87.2	.8	.5	123.1	106.4	120.6	.9	.7	113.4
MTP	63.8	115.0	.6	.7	180.0					
TIM-U2	259.0	206.4	1.3	.4	79.7	178.1	282.2	.6	.6	158.5

CDM = Centering date method, ISC = interpolation with standard curves, MTP = multiple trait projection, TIM-A1 = test interval method, TIM-U1 = TIM without adjustments for fixed effects, and TIM-U1 = TIM without adjustments for fixed effects and no corrections for missing test-d yield.

The CDM and TIM-U1 gave equivalent results because, within scheme, the interval between recording was fixed. Agreement between the centering date and test interval methods was also observed by Sargent et al. (1968). The increase in standard error with length of testing interval is comparable with the results of Erb et al. (1952) and Anderson et al. (1989). The considerably high coefficient of variation signified that the prediction error was due to numerous factors and that further partitioning of the mean bias was necessary.

8.4.1.1 All test-day records available

Within method, from model [1a], main effects and interactions between scheme and shape of mean curve were significant ($P < .0001$). In this and subsequent analyses, the main effect of variance curve was not significant ($P > .1$). The fixed effects of parity and season of calving were assigned to lactation curves in order to: 1) accomodate computing algorithms for the various methods; 2) facilitate computation of lactation totals. The bias introduced by these effects was removed by consideration of these effects in the linear models.

The least squares mean bias by method, scheme and shape of mean curve is shown in Figure 17. From this and other models, standard errors of least squares means ranged from .1 to .85 kg. Hence, least squares means that were not within 2 standard errors of each other were considered to be significantly different ($P < .0001$). Observe that, in this and all charts showing adjusted mean bias, scales for plots vary from scheme to scheme. Under scheme A1, all methods, with the exception of TIM-A1, tended to overestimate lactation yields for all shapes of the lactation curve with the exception of shapes 6 and 7. Estimation of lactation yields by TIM-A1 had the lowest bias for shape 1. In comparison with other methods, TIM-A1 had the largest absolute bias for all shapes except shape 1. The CDM, TIM-U1, and TIM-U2 performed best for shape 6 and 7 while the ISC approach gave the lowest bias for shape 7.

The results for scheme A4 indicate a tendency towards underestimation by all methods for most of the shapes except the MTP method which overestimated yields for shapes 2, 8, and 9. The ISC, CDM, and TIM-U1 tended to overestimate yields for shape 9. The MTP, TIM-A1, and TIM-U2 gave equivalent results for shapes 3 and 10. The MTP approach did best for shapes 1 and 4 but was not accurate for shape 7. The ISC, CDM, and TIM-U1 were similar and gave consistently lower biases across all shapes.

A remarkable improvement in TIM-A1, relative to other methods, was observed from scheme A4 to A6. The bias for TIM-A1 was relatively low as compared to the other methods. All methods tended to underestimate yields for all shapes except shapes 6 and 7. The ISC, CDM, TIM-U1, and TIM-U2 were similar. Similar relationships between the methods were observed for scheme A8 which was in agreement with the crude statistics in Table 10 where the biases were much larger for scheme A8. The TIM-A1 had relatively lower absolute bias than all other methods.

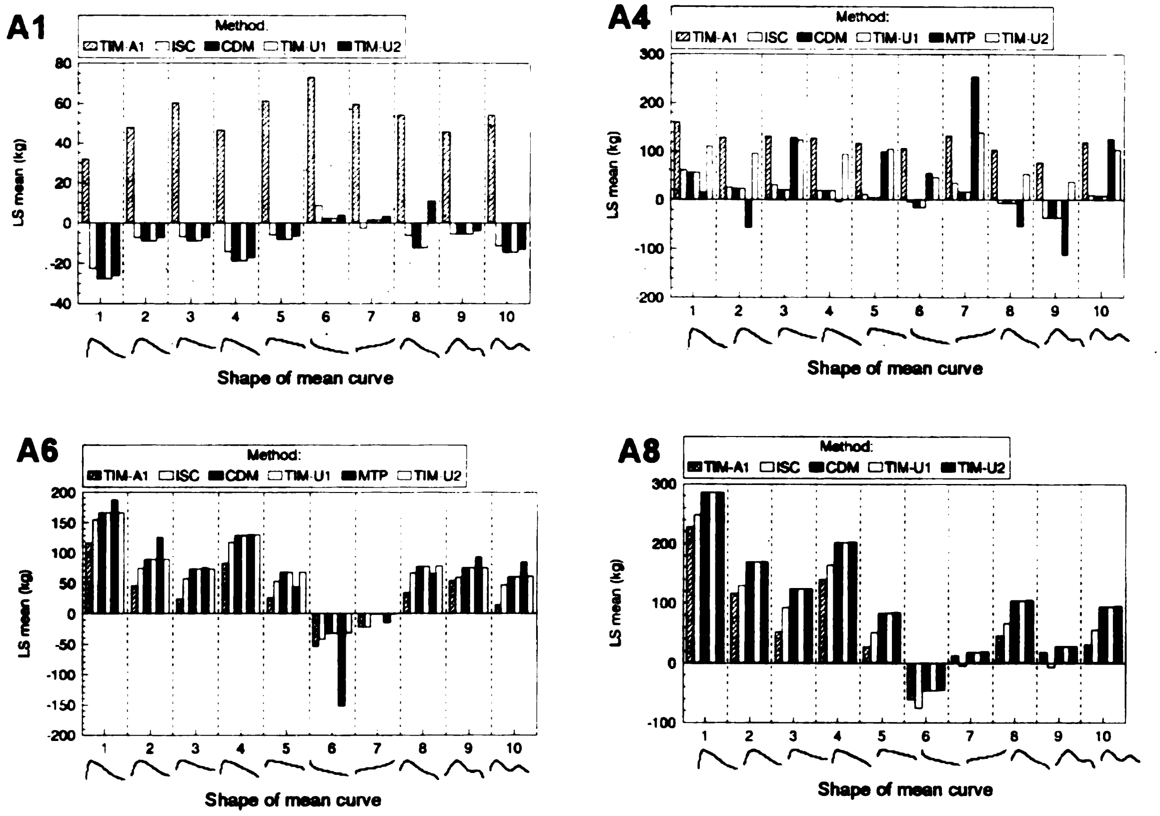


Figure 17. Least squares mean bias by method, shape of lactation curve and test scheme A1, A4, A6, and A8.

The total variation in the TIM, ISC, CDM, TIM-U1, and TIM-U2 was explained by at least two factors. Plots of the reference structure of the predominant factors for the methods within test schemes A1, A4, A6, and A8 are shown in Figure 18. A distance of 5mm was used as a measure of close proximity between methods. Methods within 3mm of each other were considered to have formed a cluster. The ISC and TIM-A1 formed a cluster which was close to CDM, TIM-U1, and TIM-U2 under scheme A1. However, ISC, TIM-A1, CDM, and TIM-U1 formed a cluster which was close to TIM-U2 but distant from MTP under scheme A4. For schemes A6, CDM, ISC, TIM-U1 and TIM-U2 clustered together and were close to TIM-A1. The MTP and TIM-A1 were close. Therefore estimation of lactation yields by CDM, ISC, TIM-U1, and TIM-U2 was expected to give results that were similar and close to those given by TIM-A1 for scheme A6. The data for scheme A8 showed a similar relationship between the methods. However, TIM-A1 was isolated from the other methods. The MSA's ranged between .67 and .91 implying that the data were suitable for the factor analysis. Results from factor analysis were in agreement with those from the linear model analysis. Although, the data were inadequate for factor analysis within scheme and shape of mean curves, this analysis added strength to interpretations by considering the (co)variance structure of methods.

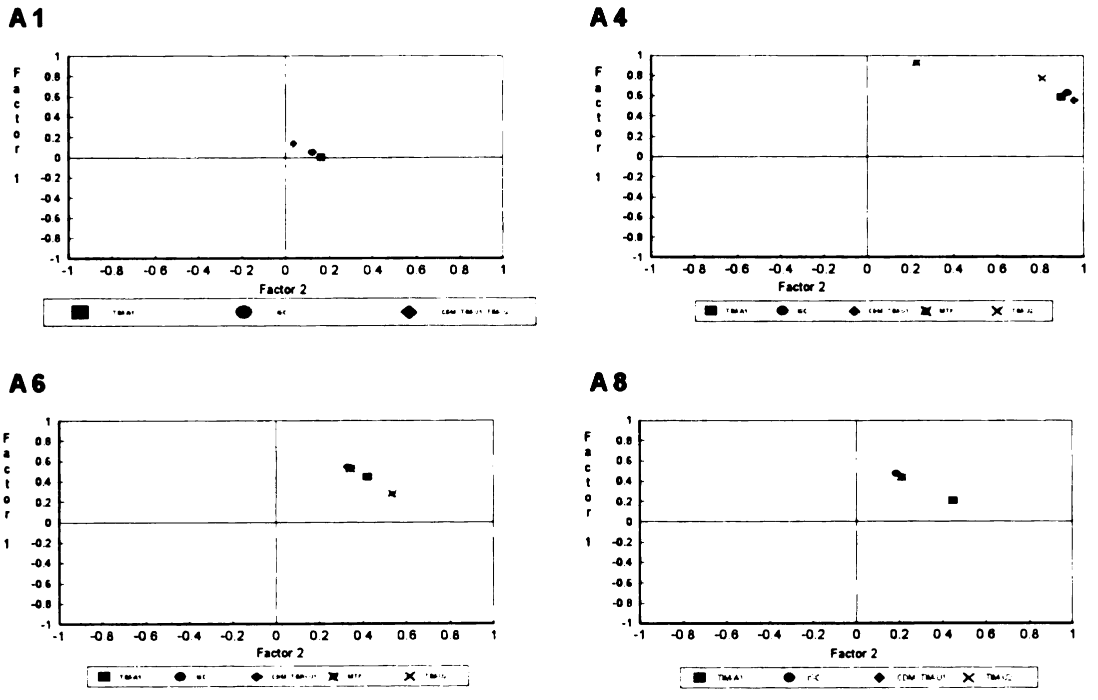


Figure 18. Reference structure for weighted Harris-Kaiser rotation of factors for test scheme A4 and test scheme A6. Procrustean rotation of factors for scheme A8.

8.4.1.2 Missing test-day records

Within method, from model [1b], the main effects and interactions between schemes and patterns of missing test-d records were significant ($P < .0001$). The least squares mean bias by method, scheme, and pattern of missing test-d records is shown in Figure 19. When all records were available, all methods were accurate under the different recording schemes except TIM-A1, TIM-U1 and TIM-U2 in scheme A6. This reduced accuracy was attributed to a possible starting day by scheme interaction. For scheme A1, the accuracy of the ISC method was consistently high for all patterns of missing data. The data for TIM-A1 were comparable to ISC expect when records were missing towards the later part of lactation. Therefore, the ISC method was more accurate than TIM-A1 for the incomplete lactations. Biases for CDM were large when records were missing in all but the late lactation and tail end stages of lactation. The TIM-U1 was not accurate in all patterns and grossly underestimated lactation yields for incomplete lactations. The TIM-U2 was comparable to ISC when records were missing during early and peak lactation.

As the sampling frequency decreased from A1 to A8, biases increased but the relative accuracy between methods remained essentially the same. Biases in CDM and TIM-U1 became very exaggerated. The MTP method was relative more accurate than TIM-A1 when test-d records were missing during lactation but was comparable to ISC for incomplete lactations in schemes A4 and A6. The TIM-U2 had lower biases than TIM-U1. Biases for TIM-U1 were low when test-d records were missing during early peak under scheme A4 and A8, and during late peak lactation under scheme A8. This implied that estimating missing yields using the average of the previous and next test-day yield improved the accuracy of TIM-U1 in schemes A4 and A8.

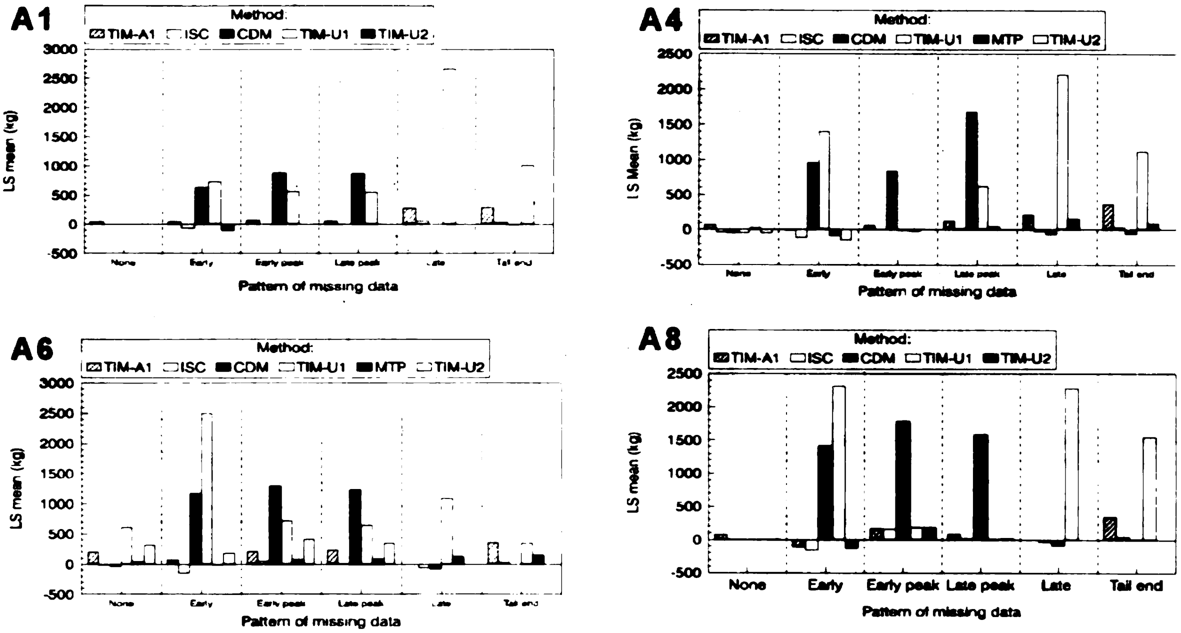


Figure 19. Least squares mean bias by method, pattern of missing test-d records and test scheme A1, scheme A4, scheme A6, and scheme A8.

Plots of the reference structure of the important factors for methods by scheme, for lactations with missing test-d records, are shown in Figure 20. Under scheme A1, all the methods, except ISC and TIM-U2 were apart from each other. The furthest distance was between CDM and TIM-A1. The plot for scheme A4 showed that TIM-A1 was close to both MTP and ISC; ISC was close to TIM-A1 and TIM-U2; TIM-U2 was distant from MTP. For scheme A6, the TIM-A1, MTP, and TIM-U2 methods formed a cluster that was close in proximity to ISC. The CDM and TIM-U1 were distant from this cluster. Therefore, these methods were not expected to yield equivalent results under schemes A6 and A8. The MSA's ranged between .6 for scheme A8 and .9 for schemes A4 and A6. Hence, the data for scheme A8 were suitable for the factor analysis.

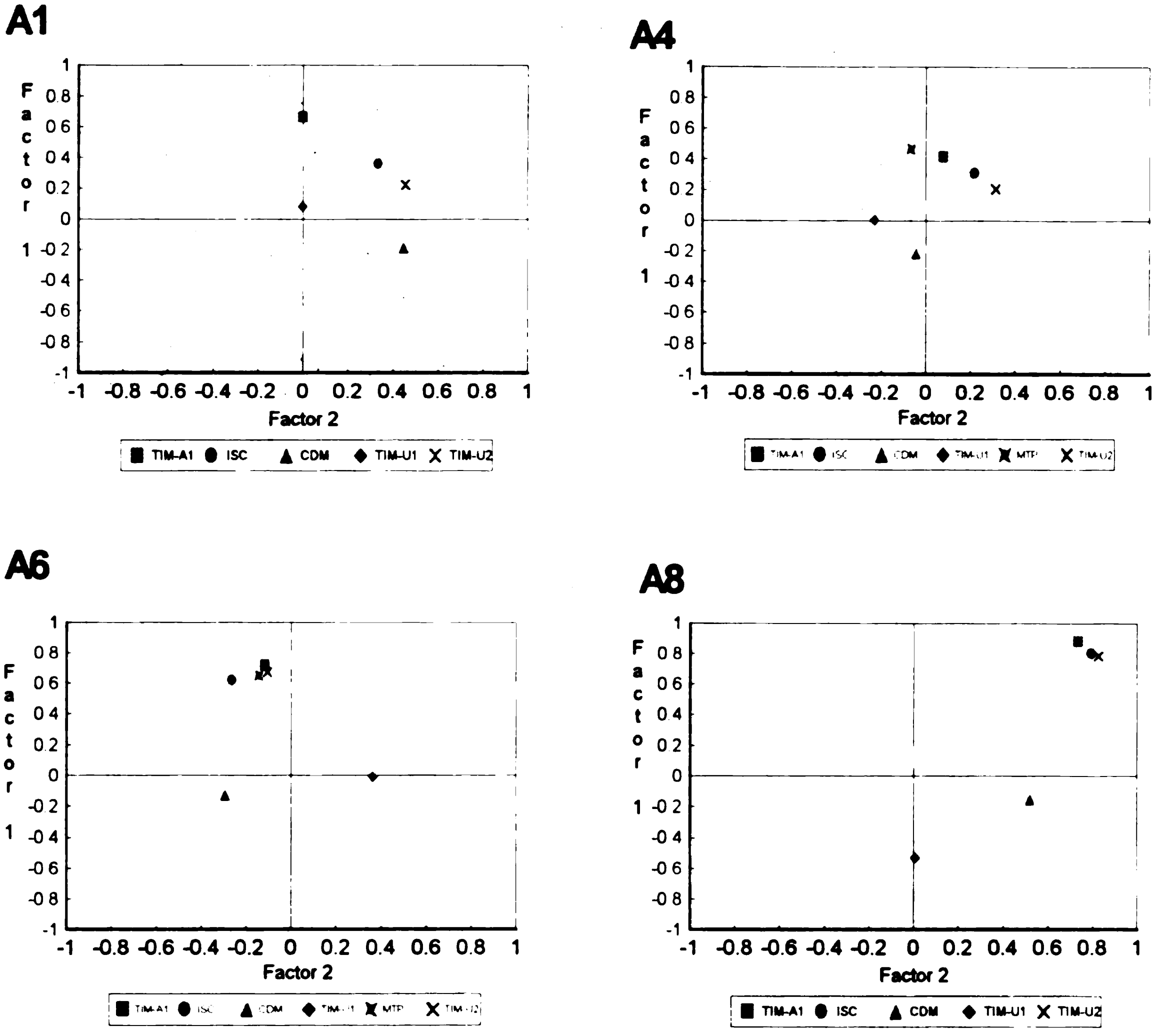


Figure 20. Reference structure for weighted Harris-Kaiser rotation of factors for test scheme A1, scheme A4, scheme A6, and scheme A8.

8.4.1.3 Starting day of recording postpartum

Within method, from model [1c], the main effects and interactions between schemes and starting day of recording were significant ($P < .0001$). The least squares mean bias by method, scheme and starting day of recording is shown in Figure 21. For scheme A4, all methods performed well. The ISC, CDM, and TIM-U2 showed an increase in overestimation as the duration to the starting day for recording was increased from 7 d to 28 d postpartum. Biases for CDM and TIM-U1 had opposite signs and also were largest when the starting day of recording was 28 d postpartum. Erb et al. (1952) also found that the day of the initial supervised test influenced the accuracy of lactation records.

The data for scheme A6 show similar trends for ISC and CDM. Biases for TIM-A1 and TIM-U1 were large for all starting days except for 7; biases for TIM-A1 were much smaller. The accuracy of TIM-U2 improved as the duration to the initial test-d increased from 14 d to 28 d. The bias for MTP was small and relatively stable but decreased when the starting day of recording was 28 d postpartum.

For scheme A8, all methods performed well for all starting days except for 28 d postpartum where biases for TIM-A1, TIM-U1, and TIM-U2 increased with the latter grossly underestimating lactation yields.

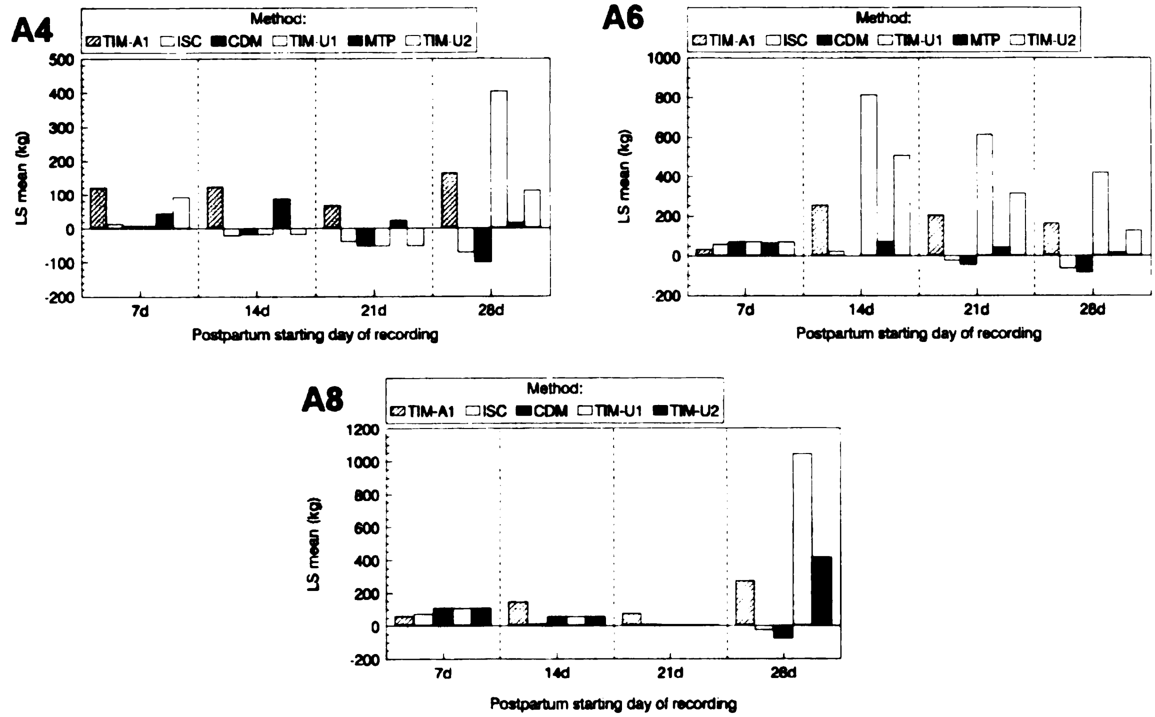


Figure 21. Least squares mean bias by method, starting day of recording and scheme A4, scheme A6, and scheme A8.

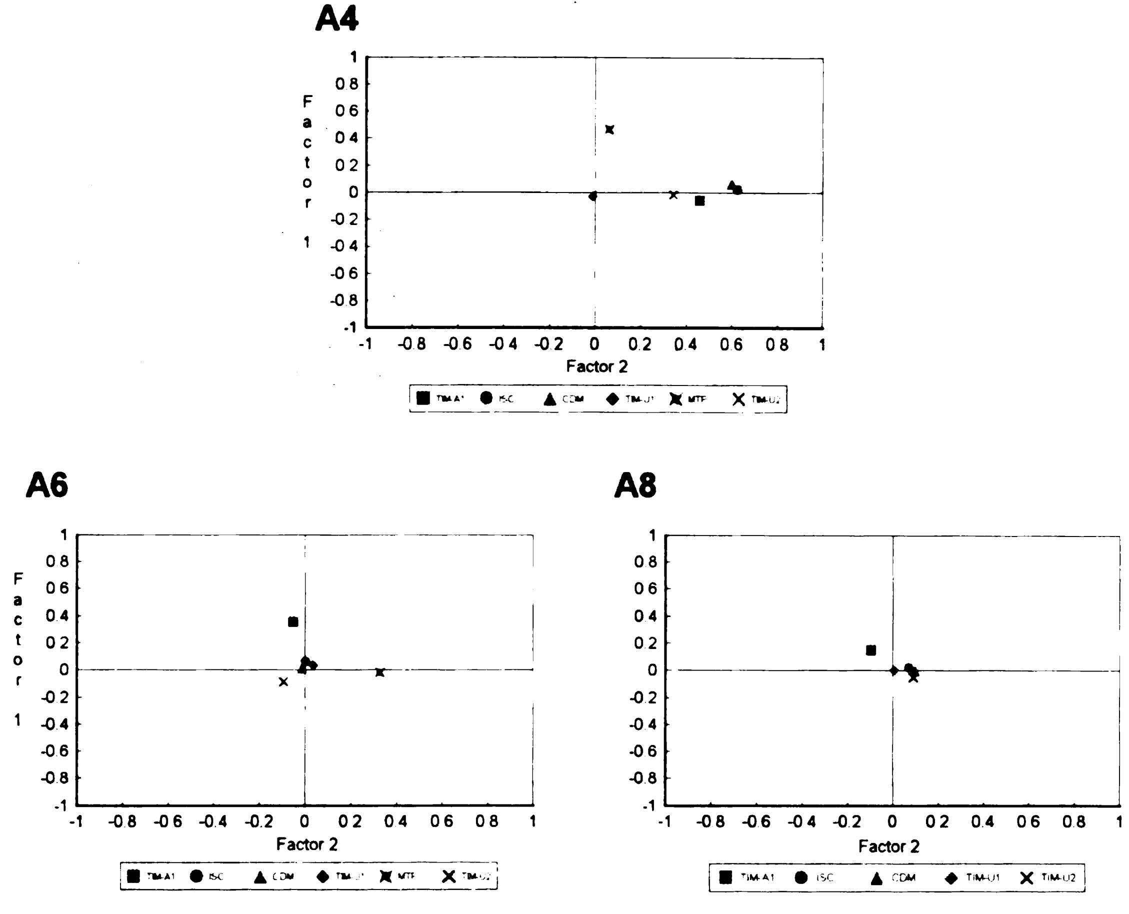


Figure 22. Reference structure for weighted Harris-Kaiser rotation of factors for test scheme A4, scheme A6, scheme A8.

Plots of the reference structure of the important factors for methods by scheme, when starting days of recording were allowed to vary, are shown in Figure 22. Under scheme A4, the CDM and ISC formed a cluster that was close to TIM-A1; TIM-A1 was close to TIM-U2. All other methods were isolated. The plot for scheme A6 showed a cluster between CDM, ISC, TIM-U1, and TIM-U2 which was distant from MTP and TIM-A1. For scheme A8, the cluster was between CDM, ISC, TIM-U1, and TIM-U2; TIM-A1 was close TIM-U1. The MSA's ranged between .5 for scheme A8, and .8 for schemes A4 and A6. Therefore, the data for scheme A8 were marginally suitable for the factor analysis.

8.4.2 Biases of different methods associated with the AP/4 test scheme

The crude mean, standard deviation, standard error, coefficient of variation, and mean : standard deviation of bias by method of estimating lactation totals for scheme AP/4 are shown in Table 11. The ISC method had the smallest bias while TIM-A1 had the largest bias but all biases and mean : standard deviation were much larger than those for scheme A4 shown in Table 9. However, the coefficient of variation for TIM-A1 was relatively low. The variance for TIM-U1 was small as compared to other methods. The mean : standard deviation were smallest for ISC and TIM-U2. Standard errors were similar but larger than those for other testing schemes shown in Table 9 and 10.

TABLE 11. Mean, standard deviation, standard error, coefficient of variation, mean : standard deviation of bias by method of estimating lactation totals and schemes AP/4.

Method	$\overline{X}$	SD	$\overline{X}$/SD	SE	CV
	(kg)			(kg)	%
TIM	222.1	87.1	2.5	1.6	39.2
ISC	25.0	86.1	.3	1.6	344.4
TIM-U1	34.9	9.5	3.7	1.7	256.3
TIM-U2	-42.4	196.3	.2	1.2	-463.2

ISC = Interpolation with standard curves, TIM-A1 = test interval method, TIM-U1 = TIM without adjustments for fixed effects, and TIM-U1 = TIM without adjustments for fixed effects and no corrections for missing test-d yield.

8.4.2.1 All test-day records available

From model [2a], the effect of shape of mean curve was significant ($P < .0001$). Least squares mean bias by method and shape of mean curves is shown in Figure 23A. The biases for TIM-A1, ISC, and TIM-U2 were invariant to the shapes of the underlying lactation curves. However, TIM-U1 had a larger bias under shape 11. This difference was attributed to more variability around peak and postpeak yields for shape 11.

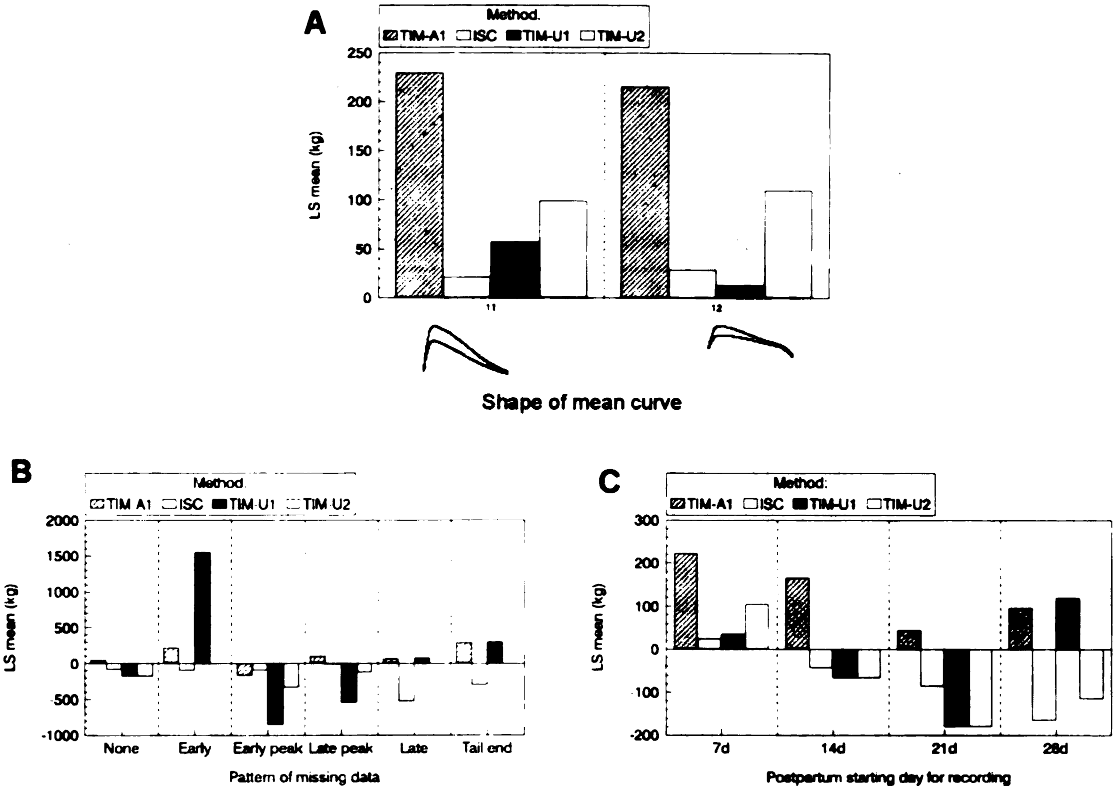


Figure 23. Least squares mean bias by scheme AP/4, method, and shape of lactation curve (A), pattern of missing test-d records (B), and starting day of recording (C).

A plot for the reference structure of the first two factors is shown in Figure 24A. The ISC clustered with TIM-U2 while TIM-A1 was close to TIM-U1. The overall MSA was .8 which suggested that the data were suitable for factor analysis.

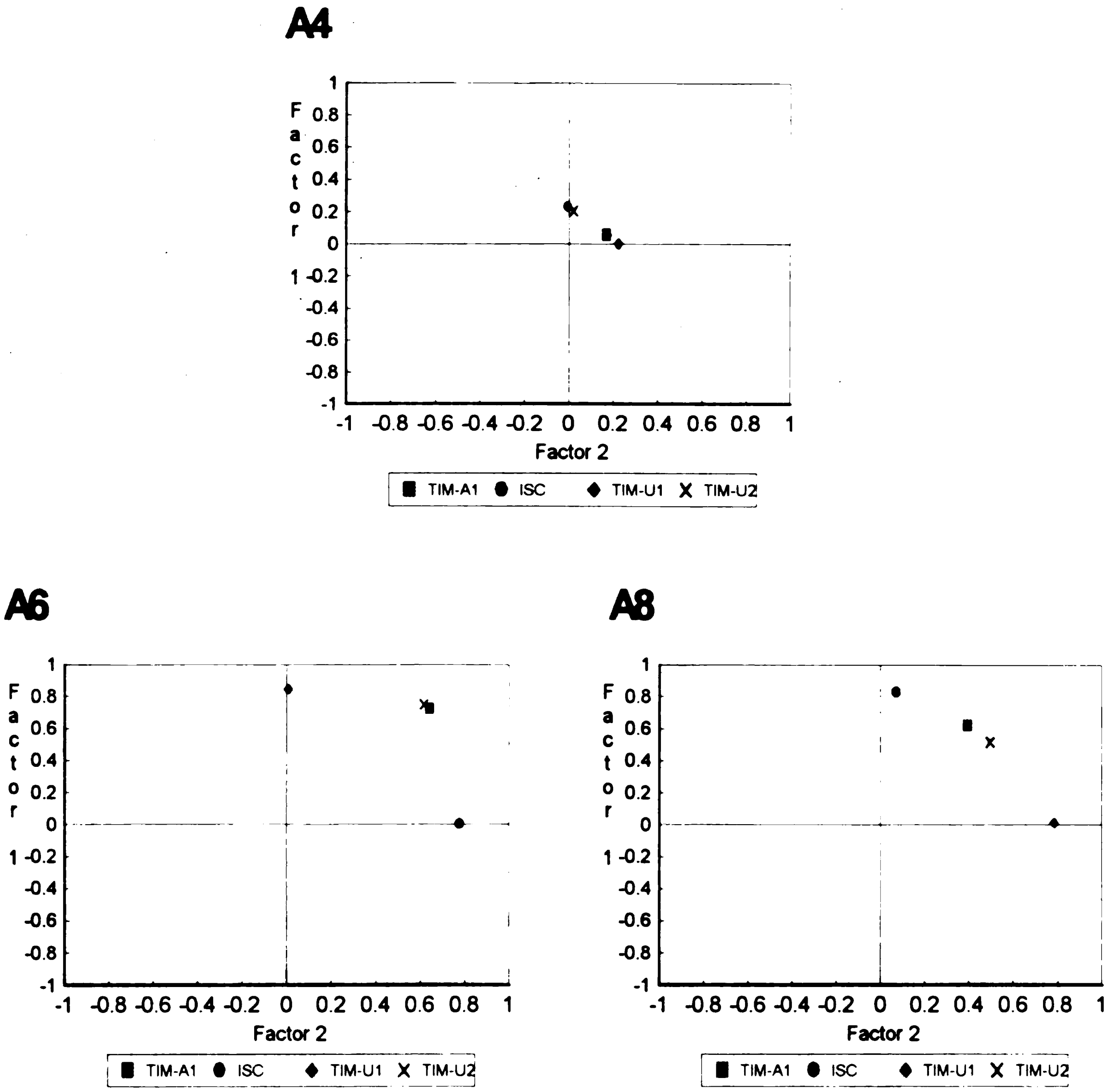


Figure 24. Reference structure for weighted Harris-Kaiser rotation of factors for scheme AP/ 4 and method for lactations with no missing data(A), lactations with missing data (B), and lactations with varying starting day of recording (C).

8.4.2.2 Missing test-day records

From model [2b], the effect of missing test-d records was significant ($P < .0001$). Least squares mean bias by method and pattern of missing test-d records is shown in Figure 23B.

for SLAC.

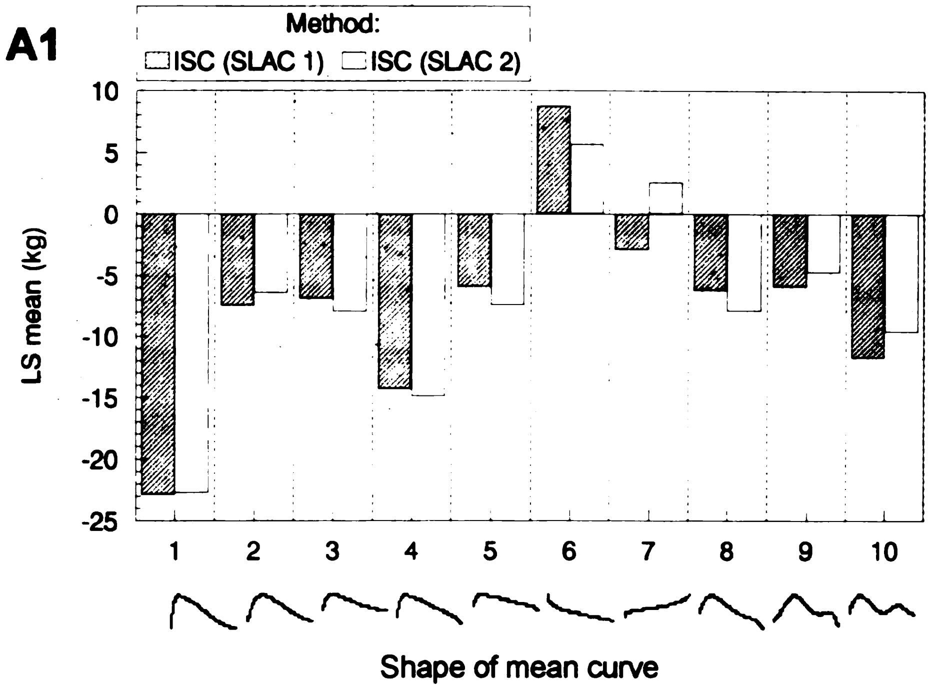


Figure 25. Least squares mean bias by method, data set, and shape of mean curve for scheme A1.

Contrary to the data for scheme A4, the ISC method had considerably large biases for incomplete lactations. The bias for TIM-A1 was largest when test-d yields were missing at the tail end of lactations. The largest biases for TIM-U1 occurred when data were missing early and peak periods of a lactation. During those periods, the performance of TIM-U1 was inferior to TIM-A1 and ISC. For incomplete lactations, the performance of TIM-U1 was comparable to TIM. The accuracy of TIM-U2 was highest when test-d yields were missing in early lactation but was lowest when data were missing during early peak lactation.

A plot of the reference structure for the important factors is shown in Figure 24B. This plot depicted close proximity between ISC and TIM-U2. All other methods were distant. The overall MSA was .65 which implied that the data were well-suited for factor analysis.

8.4.2.3 Starting day of recording postpartum

From model [2c], the effect of starting day of recording was significant ($P < .0001$). Least squares mean bias by method and starting day of recording is shown in Figure 23C. Biases for TIM-A1 were large when the starting day of recording was 7d but appeared to decline as the starting day occurred later in lactation.

A plot of the reference structure for the important factors is shown in Figure 24C. The TIM-A1 and TIM-U2 were close to each other. All other methods were distant from each other indicating a lack of equivalence between them. The overall MSA of .7 implied that these data were marginally favorable for the factor analysis.

8.4.3 Validation of sampling design

To examine if the performance of methods was caused by the sampling design of SLAC, a different data set (SLAC 2) was created as defined in §8.3.1. However, values for the variance curves used in creating SLAC 2 were twice those shown in Figure 15. Thus, test-day yields for lactations had more day to day variation. The ISC method was used to compute lactation totals for SLAC 2. Scheme A1 results for SLAC 1 and SLAC 2 are shown in Figure 25. The biases for both data sets were very similar. This consistency established that the performance of methods was not due to sampling design

8.5 Conclusions

Generation of SLAC enabled the examination of the relative accuracy of different methods for calculating lactation totals. The shapes of mean curves were designed to cover factors that affect milk yield in a lactation. The patterns of missing test-d records, starting days of recording, and testing schemes were representative of typical situations. The relative accuracy was examined under different shapes of mean curve, patterns of missing test-d records, and starting days of recording. Methods for computing lactation yields are many and some are currently being modified. The methods compared in this study were considered to be fairly representative of different approaches. The sampling frequency did not affect the mean bias but increased the standard errors. The accuracy of methods for scheme AP/4 was lower when compared with scheme A4. When all test-d records are available, and when the duration to the initial test was 7d, all the methods performed well. The accuracy of these methods differed according to underlying shapes of the lactation curve. In general the interpolation with standard curves, centering date, multiple trait projection were more accurate than the other methods. While the adjustment factors used with the test interval method appeared to underestimate lactation yields in scheme A1, the method produced small biases for lactations with missing test-d records. The centering date and test interval methods were considered suitable for recording schemes that exclude lactations with missing test-d yields. When the starting day of recording was delayed to 28 d postpartum, the accuracy of the test interval methods without adjustment for fixed factors was reduced. The shapes of the lactation curves in this data set were known apriori. Thus, implementation of the interpolation with standard curves methods was easy. In most situations, information about the underlying shapes is not readily available. Therefore, the benefit of using this method should be considered after a careful review of its merit and the availability of such data. In order for a data processing center to adopt a given method, shapes of lactation curves, patterns of missing test-d records, and starting days of recording postpartum inherent in the data need to be considered. The SLAC enabled the International Committee of Animal Recording to evaluate the accuracy of any method for computing lactation totals. From these evaluations, standards for recognizing methods as official were established.

9. SUMMARY

Sampling variations in mean daily milk yield were investigated using bootstrap resampling. The bootstrap confidence intervals of mean daily yield were accurate and consistent with the theory. The variation in yields was highest at the peak and postpeak stages of lactation. It was established that for a given mean and variance curve, sample lactation curves could be generated.

Variation in morning and evening yields of Holstein cows was studied. Mean morning yield was higher than evening yield. The variance of evening yield appeared to determine the variance of total daily yield. The variance was highest at peak lactation of morning and evening yields. Regional differences in the magnitude of variance of daily total, morning, and evening yield were observed. These results provided a preliminary understanding of the shapes of curves for mean milk yield and its variance.

A set of curves for mean and variance of total, morning, and evening yields that was representative of the global cattle population was established. Mathematical functions were fitted to these curves. A data set of standard lactation curves (SLAC) was generated using these functions. Thus lactation records of SLAC were biologically consistent and accomodated different missing test-day records, starting days of recording, parities, breeds, seasons of calving, ages at calving, regions, and other grouping effects.

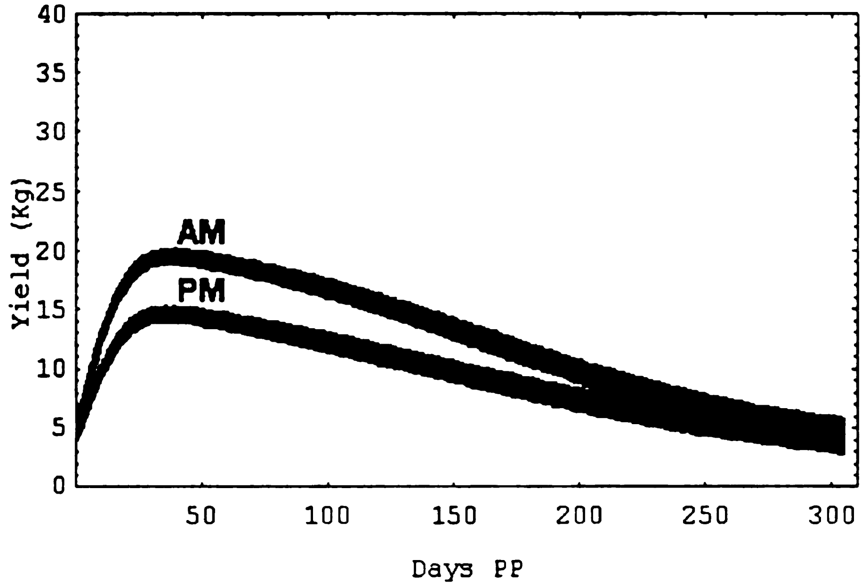
The SLAC was sent to six participating data processing centers. Each center calculated and returned lactation totals for SLAC. A total of six methods for calculating lactation totals were compared. When all test-d records were available, all methods performed well. The sampling frequency did not affect the mean bias but increased the standard errors. Within shape of lactation curves, the accuracy of the methods varied. With missing test-d yields, some methods gave a lower bias than others. The SLAC was commissioned by ICAR to identify the sensitivity of a given method to the various shapes of the lactation curve, patterns of missing test-d yields, and starting days of recording.

10. APPENDICES

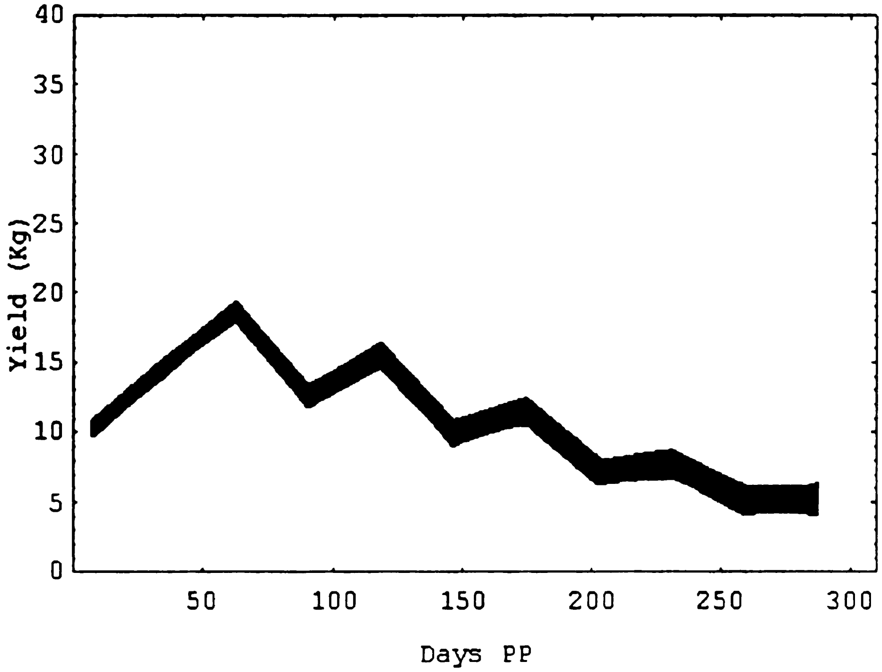
APPENDIX A:

Illustration of parameter lactation curves for daily AM and PM yield using mean daily total yield curve 11 and variance curve 4

AM/PM MEAN CURVE 11 - VARIANCE CURVE 4:



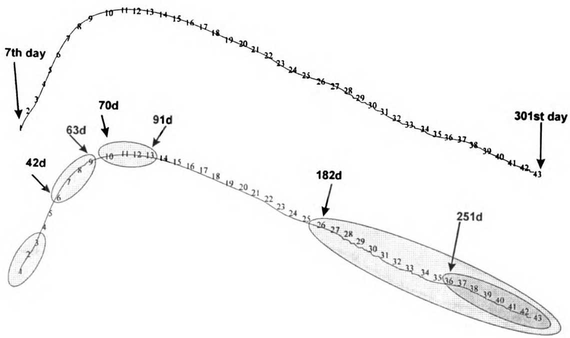
AP/4 CURVES (AM/PM MEAN CURVE 11 - VARIANCE CURVE 4):



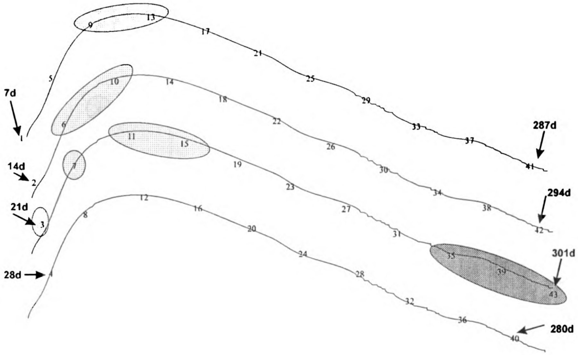
APPENDIX B:

Graphic illustration of patterns of missing test-day records

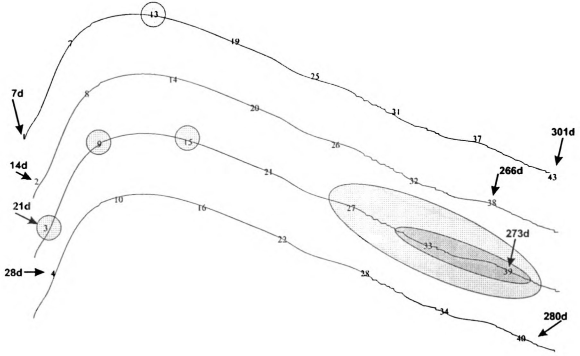
SCHEME: A1



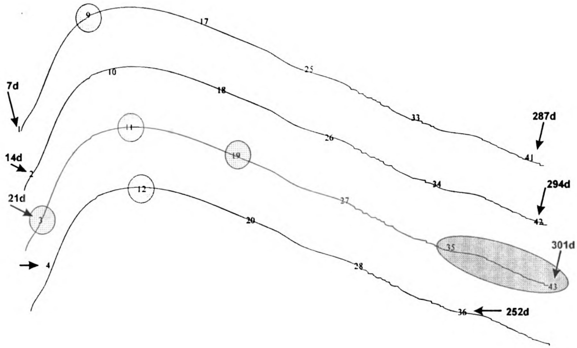
SCHEMES: A4 and AP/4



SCHEME: A6


7d
14d
21d
28d
301d
266d
273d
280d
1
7
13
19
25
31
37
43
2
8
14
20
26
32
38
3
9
15
21
27
33
39
4
10
16
22
28
34
40

SCHEME: A8


7d
14d
21d
287d
294d
301d
252d
1
2
3
4
9
10
11
12
17
18
19
20
25
26
27
28
33
34
35
36
41
42
43

APPENDIX C:
Format for SLAC data files and sample printout of data

(1) FILE 1 (slac1td.dat): Contains test-day data from test scheme A1

Number of records: 172,800

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (1, ..., 10)
Numeric	1	3	Variance curve No. (1, ...,4)
Numeric	3	4-6	Replicate No. (1, ...,120)
Numeric	1	7	Scheme (1=A1)
Numeric	1	8-9	Starting day (7=7d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5=tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	1	12	Days PP at 1st recording
Numeric	2	13-14	Days PP at 2nd recording
"	"	...	
"	"	37-38	Days PP at 14th recording
Numeric	3	39-41	Days PP at 15th recording
"	"	...	
Numeric	3	123-125	Days PP at 43rd recording
Numeric	3	126-128	Test-day yield at 1st recording (divide by 10 to get yield, 0=missing)
Numeric	3	129-131	Test-day yield at 2nd recording (")
"	"	...	
Numeric	3	252-254	Test-day yield at 43rd recording (")

ILLUSTRATION OF SLAC FILE 1 (slac1td.dat) with 1st and last two days PP and corresponding test day yields shown

First 3 records on-file (Unformatted):

```
 11  11 701 714...294301228289...135118
 11  11 702 714...294301228289...135118
 11  11 703 714...294301228289...135118
```

First 3 records after reading (Formatted):

```
1 1 1 1 7 0 1 7 14 ... 294 301 22.8 28.9 ... 13.5 11.8
1 1 1 1 7 0 2 7 14 ... 294 301 22.8 28.9 ... 13.5 11.8
1 1 1 1 7 0 3 7 14 ... 294 301 22.8 28.9 ... 13.5 11.8
```

(2) FILE 2 (slac2td.dat): Contains test-day data from test scheme A4

Number of records: 316,800

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (1, ..., 10)
Numeric	1	3	Variance curve No. (1, ..., 4)
Numeric	3	4-6	Replicate No. (1, ..., 120)
Numeric	1	7	Scheme (2=A4)
Numeric	2	8-9	Starting date (7=7d PP, 14=14d PP, 21=21d PP, 28=28d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5= tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	2	12-13	Days PP at 1st recording
Numeric	2	14-15	Days PP at 2nd recording
Numeric	2	16-17	Days PP at 3rd recording
Numeric	3	18-20	Days PP at 4th recording
"	"	...	
Numeric	3	39-41	Days PP at 11th recording (999 = Not applicable for scheme/starting date of recording)
Numeric	3	42-44	Test-day yield at 1st recording (divide by 10 to get yield, 0=missing)
Numeric	3	45-47	Test-day yield at 2nd recording (")
"	"	...	
Numeric	3	72-74	Test-day yield at 11th recording (divide by 10 to get yield, 0=missing, 999 = Not applicable for scheme/starting date of recording)

ILLUSTRATION OF SLAC FILE 2 (slac2td.dat) with 1st and last two days PP and corresponding test day yields shown

First 3 records on-file (Unformatted):

```
 11  12 701 735...259287228380...153131
 11  12 702 735...259287228380...153131
 11  12 703 735...259287228380...153131
```

First 3 records after reading (Formatted):

```
1 1 1 2 7 0 1 7 35 ... 259 287 22.8 38.0 ... 15.3 13.1
1 1 1 2 7 0 2 7 35 ... 259 287 22.8 38.0 ... 15.3 13.1
1 1 1 2 7 0 3 7 35 ... 259 287 22.8 38.0 ... 15.3 13.1
```

(3) FILE 3 (slac3td.dat): Contains test-day data from test scheme A6

Number of records: 288,000

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (1, ..., 10)
Numeric	1	3	Variance curve No. (1, ..., 4)
Numeric	3	4-6	Replicate No. (1, ..., 120)
Numeric	1	7	Scheme (3=A6)
Numeric	2	8-9	Starting date (7=7d PP, 14=14d PP, 21=21d PP, 28=28d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5= tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	2	12-13	Days PP at 1st recording
Numeric	2	14-15	Days PP at 2nd recording
Numeric	3	16-18	Days PP at 3rd recording
"	"	...	
Numeric	3	31-33	Days PP at 8th recording (999= Not applicable for scheme/starting date of recording)
Numeric	3	34-36	Test-day yield at 1st recording (divide by 10 to get yield, 0=missing)
Numeric	3	37-39	Test-day yield at 2nd recording (")
"	"	...	
Numeric	3	55-57	Test-day yield at 8th recording (divide by 10 to get yield, 0=missing, 999 = Not applicable for scheme/starting date of recording)

ILLUSTRATION OF SLAC FILE 3 (slac3td.dat) with 1st and last two days PP and corresponding test day yields shown

First 3 records on-file (Unformatted):

```
 11  13 701 749...259301228379...153118
 11  13 702 749...259301228379...153118
 11  13 703 749...259301228379...153118
```

First 3 records after reading (Formatted):

```
1 1 1 3 7 0 1 7 49 ... 259 301 22.8 37.9 ... 15.3 11.8
1 1 1 3 7 0 2 7 49 ... 259 301 22.8 37.9 ... 15.3 11.8
1 1 1 3 7 0 3 7 49 ... 259 301 22.8 37.9 ... 15.3 11.8
```

(4) FILE 4 (slac4td.dat): Contains test-day data from test scheme A8

Number of records: 316,800

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (1, ..., 10)
Numeric	1	3	Variance curve No. (1, ..., 4)
Numeric	3	4-6	Replicate No. (1, ..., 120)
Numeric	1	7	Scheme (4=A8)
Numeric	2	8-9	Starting date (7=7d PP, 14=14d PP, 21=21d PP, 28=28d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5= tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	2	12-13	Days PP at 1st recording
Numeric	2	14-15	Days PP at 2nd recording
Numeric	3	16-18	Days PP at 3rd recording
"	"	...	
Numeric	3	25-27	Days PP at 6th recording (999 = Not applicable for scheme/starting date of recording)
Numeric	3	28-30	Test-day yield at 1st recording (divide by 10 to get yield, 0=missing)
Numeric	3	31-33	Test-day yield at 2nd recording (")
"	"	...	
Numeric	3	43-45	Test-day yield at 6th recording (divide by 10 to get yield, 0=missing, 999 = Not applicable for scheme/starting date of recording)

ILLUSTRATION OF SLAC FILE 4 (slac4td.dat) with 1st and last two days PP and corresponding test day yields shown

First 3 records on-file (Unformatted):

```
 11  14 701 763...231287228368...182131
 11  14 702 763...231287228368...182131
 11  14 703 763...231287228368...182131
```

First 3 records after reading (Formatted):

```
1 1 1 4 7 0 1 7 63 ... 231 287 22.8 36.8 ... 18.2 13.1
1 1 1 4 7 0 2 7 63 ... 231 287 22.8 36.8 ... 18.2 13.1
1 1 1 4 7 0 3 7 63 ... 231 287 22.8 36.8 ... 18.2 13.1
```

(5) FILE 5 (slac5td.dat): Contains test-day data from test scheme AP/4

Number of records: 31,680

Type	Length	Position	Description
Numeric	2	1-2	Mean curve No. (11 = AM/PM curve 1, 12 = AM/PM curve 2)
Numeric	1	3	Variance curve No. (5 = AM/PM variance curve 1, 6 = AM/PM variance curve 2)
Numeric	3	4-6	Replicate No. (1, ..., 120)
Numeric	1	7	Scheme (5=AP/4)
Numeric	2	8-9	Starting date (7=7d PP, 14=14d PP, 21=21d PP, 28=28d PP)
Numeric	1	10	Missing pattern (0=none, 1=early lactation, 2=early peak, 3=late peak, 4=late, 5= tail end)
Numeric	1	11	Combination code for levels of fixed effects (1 => Lactation=1, breed=1, trait=1, season=1; 2 => Lactation=1, breed=1, trait=1, season=2; 3 => Lactation=1, breed=1, trait=1, season=3; 4 => Lactation=1, breed=1, trait=1, season=4; 5 => Lactation=2, breed=1, trait=1, season=1; 6 => Lactation=2, breed=1, trait=1, season=2; 7 => Lactation=2, breed=1, trait=1, season=3; 8 => Lactation=2, breed=1, trait=1, season=4)
Numeric	2	12-13	Days PP at 1st recording
Numeric	2	14-15	Days PP at 2nd recording
Numeric	2	16-17	Days PP at 3rd recording
Numeric	3	18-20	Days PP at 4th recording
"	"	...	
Numeric	3	39-41	Days PP at 11th recording (999 = Not applicable for scheme/starting date of recording)
Numeric	3	42-44	Test-day yield at 1st recording (divide by 10 to get yield, 0=missing)
Numeric	3	45-47	Test-day yield at 2nd recording (divide by 10 to get yield, 0=missing)
"	"		...
Numeric	3	72-74	Test-day yield at 11th recording (divide by 10 to get yield, 0=missing, 999 = Not applicable for scheme/starting date of recording)

ILLUSTRATION OF SLAC FILE 5 (slac5td.dat) with 1st and last two days PP and corresponding test day yields shown

First 3 records on file (Unformatted):

```
115  15 701 735...259287 96148... 46 50
115  25 701 735...259287106147... 55 49
115  35 701 735...259287104141... 51 45
```

First 3 records after reading (Formatted):

```
11 5 1 5 7 0 1 7 35 ... 259 287 9.6  14.8 ... 4.6 5.0
11 5 2 5 7 0 1 7 35 ... 259 287 10.6 14.7 ... 5.5 4.9
11 5 3 5 7 0 1 7 35 ... 259 287 10.4 14.1 ... 5.1 4.5
```

APPENDIX D:
Illustration of files containing computed lactation total yield

FILE 1 (slac1sum.dat) for SCHEME A1

First 3 records on-file:

```
 11  11 70110111
 11  11 70210112
 11  11 70310113
```

FILE 2 (slac2sum.dat) for SCHEME A4

First 3 records on-file:

```
 11  12 70110111
 11  12 70210112
 11  12 70310113
```

FILE 3 (slac3sum.dat) for SCHEME A6

First 3 records on-file:

```
 11  13 70110111
 11  13 70210112
 11  13 70310113
```

FILE 4 (slac4sum.dat) for SCHEME A8

First 3 records on-file:

```
 11  14 70110111
 11  14 70210112
 11  14 70310113
```

FILE 5 (slac5sum.dat) for SCHEME AP/4

First 3 records on-file:

```
115  15 70110111
115  25 70110112
115  35 70110113
```

APPENDIX E.1

Installation instructions for the SLAC dataset

A. Distribution disks

INSTALL/DATA DISK #1

```
 Filename    Type Length   Date     Description
=============================================
autoinst.bat  A     3452 02-16-94  DOS batch file (alternative to install.exe)
readme.1st    A     2495 02-01-94  Instructions for data retrieval via FTP
slac.fmt      A    13814 02-01-94  File formats for SLAC
install.exe   B   175132 02-16-94  SLAC installation software
*.b_          A                    Internal files (used by install.exe)
*.t_          A                    Internal files (used by install.exe)
*.x_          B                    Internal files (used by install.exe)
slac1.1       A   797559 02-14-94  Part I of archive for File 1 of SLAC
```

INSTALL/DATA DISK #2

```
 Filename    Type Length   Date     Description
=============================================
slac1.2       A   797559 02-14-94  Part II of archive for File 1 of SLAC
```

INSTALL/DATA DISK #3

```
 Filename    Type Length   Date     Description
=============================================
slac2.zoo     B  1287447 01-15-94  Archive for File 2 of SLAC (slac2td.dat)
```

INSTALL/DATA DISK #4

```
 Filename    Type Length   Date     Description
=============================================
slac3.zoo     B   940150 01-15-94  Archive for File 3 of SLAC (slac3td.dat)
```

INSTALL/DATA DISK #5

```
 Filename    Type Length   Date     Description
=============================================
slac4.zoo     B   934519 01-15-94  Archive for File 4 of SLAC (slac4td.dat)
slac5.zoo     B   119312 01-31-94  Archive for File 5 of SLAC (slac5td.dat)
```

The SLAC is partitioned into 5 parts according to milk recording scheme. A detailed description of each part is given in the file *slac.fmt* on INSTALL/DATA DISK 1. The uncompressed SLAC files should occupy 103 MB disk space.

B. Data Installation

To install SLAC:

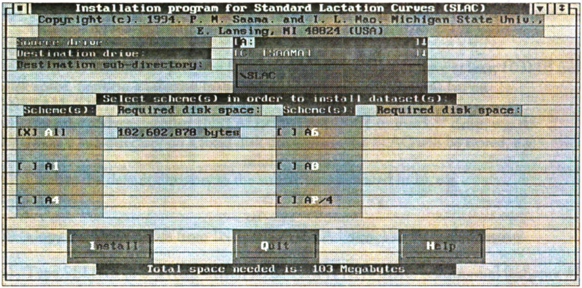
Step 1

Place the diskette labelled INSTALL/DATA DISK 1 in the 3.5" drive, e.g. A:

Step 2

Type A:INSTALL and press <Enter> (Executes SLAC installation program).

You should see a screen that looks like this:



Installation program for Standard Lactation Curves (SLAC)

Copyright (c). 1994. P. M. Saama. and I. L. Mao. Michigan State Univ.,
E. Lansing, MI 48824 (USA)

Source drive: [A:]
Destination drive: [C: [SAAMA]]
Destination sub-directory: \SLAC

Select scheme(s) in order to install dataset(s):

Scheme(s):	Required disk space:	Scheme(s):	Required disk space:
[X] All	102,602,878 bytes	[ ] A6	
[ ] A1		[ ] A8	
[ ] A4		[ ] AP/4	

Total space needed is: 103 Megabytes

In the event that this screen does not come up, we recommend that you stream-line your system start-up files. If problems persist, skip to step 4.

Step 3

The following are required as input:

a) The data source drive:

To choose from a list of all available drives on your system, use [Up] and [Down] arrow keys or mouse. This should be the 3.5" drive. The default is A:

b) The Destination drive:

This is the drive on which the data will be decompresseed. Use the [Tab] key move forward from input field to input field. [SHIFT-Tab] moves backwards between input fields.

c) The destination sub-directory:

This is the location on the destination drive where the data will be stored.

d) Scheme:

Selective installation of data for any desired scheme is made possible by pressing the spacebar. The default choice installs data for all schemes. The [Up] and [Down] arrow keys can be used to move between the check boxes. Pressing [Alt] plus the [highlighted letter] or [number] selects or invokes the respective feature. For example Alt-1 selects scheme A1 while [Alt-H] invokes the on-line help. The disk space needed is shown at the bottom of the screen. Press [Alt-I] to start the data installation. The program will prompt you for the appropriate disks.

Step 4

If the installation program fails to work, perform a full installation using the batch file *autoins.bat* on INSTALL/DATA DISK 1.

a) Place the diskette labelled INSTALL/DATA DISK 1 in the 3.5" drive, e.g. A:
b) Type A:AUTOINS [source drive:] [destination drive:\path] and press <Enter>
This should install SLAC on the [destination drive:\path].

If the installation program fails to work and you do not have at least 103MB free disk space, perform a selective installation using the batch file *selins.bat* on INSTALL/DATA DISK 1.

a) Place the diskette labelled INSTALL/DATA DISK 1 in the 3.5" drive, e.g. A:
b) Make a sub-directory on a fixed disk drive where the SLAC will be stored, e.g.

```
mkdir c:\slac
```

c) Change the default directory to the sub-directory created in b), e.g.

```
c: <Enter>
cd\slac <Enter>
```

d) Copy the file *selins.bat* from INSTALL/DATA DISK 1 to this sub-directroy.

e) Type SELINS [SOURCE drive:] [DESTINATION drive:\pathname] [scheme #] and press <Enter> then follow the instructions on the screen.

Valid values for [scheme #] are:

1	for scheme A1
2	for scheme A4
3	for scheme A6
4	for scheme A8
5	for scheme AP/4

For example:

```
SELINS A: C:\SLAC 1 ----- Installs SCHEME A1 data from A: to C:\SLAC
```

This should perform a selective installation of SLAC.

APPENDIX E.2

Installation instructions for electronic retrieval of SLAC via FTP

Files were stored in the directory pub\slac\incoming\ (guest ftp 35.8.124.45, pwd: slac)

Filename	Type	Length	Date	Description
readme.1st	A	2495	02-01-94	This file (Instructions for data retreival)
slac.fmt	A	3814	02-01-94	File formats for SLAC
slac1.zoo	B	1595118	01-15-94	Archive for File 1 of SLAC (slac1td.dat)
slac2.zoo	B	1287447	01-15-94	Archive for File 2 of SLAC (slac2td.dat)
slac3.zoo	B	940150	01-15-94	Archive for File 3 of SLAC (slac3td.dat)
slac4.zoo	B	934519	01-15-94	Archive for File 4 of SLAC (slac4td.dat)
slac5.zoo	B	119312	01-31-94	Archive for File 5 of SLAC (slac5td.dat)
zoo210.exe	B	55721	01-17-94	Self extracting - makes/extracts/views ZOO archives

The SLAC is partitioned into 5 parts according to milk recording scheme. A detailed description of each part is given in the file \pub\slac\incoming\slac.fmt. Due to the large size of the files each one was compressed to enable transfer from MSU. Five (5) MB disk space is required for transferring SLAC. The uncompressed SLAC files take

up 103 MB disk space. The following sequence of commands is suggested (rootdir is used to represent the destination directory):

	Command:	Purpose:
1.	lcd \rootdir	(Changes your default directory)
1.	bi	(Sets file transfer mode to binary)
2.	get filename.ext	(Transfers *filename.ext* to *\rootdir*, repeat as *desired)*
	or	
	mget *.*	(Transfers all files to \rootdir)
3.	quit	(Closes the FTP connection).

Note: The file compression\decompression utility is Dhesi's ZOO version 2.10 for MSDOS 3.x or higher and should now be in your \rootdir. The compressed files should be compatible with the UNIX version of ZOO.

4.	zoo210	Executes zoo210.exe and extracts the files ZOO.EXE & ZOO.MAN
5.	zoo x arc *.*	Extracts the contents of \rootdir\arc.zoo (where arc is the filename for the archive (e.g. slac1 for \rootdir\slac1.zoo). See ZOO.MAN for further instructions.

APPENDIX F.1: Illustration of the test interval method (MSU)

Example 1: *All testday yields available (Scheme A4)*
SLAC record (Mean curve/ Variance curve/ Replicate No./ Scheme/ Starting date/ Missing pattern/ Code for fixed effects) = 10/4/015/2/07/0/8

Day of lact.	Testday yield	Interval days	Interval yield	Cum. Yield
7	26.9	7	188.3	188.3
35	35.1	28	868.0	1056.3
63	35.1	28	982.8	2039.1
91	31.9	28	938.0	2977.1
119	27.3	28	828.8	3805.9
147	24.2	28	721.0	4526.9
175	23.2	28	663.6	5190.5
203	27.5	28	709.8	5900.3
231	29.2	28	793.8	6694.1
259	26.6	28	781.2	7475.3
287	22.9	46	1105.2	8580.5

Notes:
- length of last interval = (287 - 259) + (305 - 287) = 46 days credited for.

Example 2: *Not all testday yields available (SCHEME A4)*
SLAC record (Mean curve/ Variance curve/ Scheme/ Replicate No./ Starting date/ Missing pattern/ Code for fixed effects) = 10/4/010/2/21/2/8

Day of lact.	Testday yield	Interval days	Interval yield	Cum. Yield
21	32.6	21	684.6	684.6
49	____	28	914.9	1599.5
77	32.9	28	919.1	2518.6
105	28.5	28	859.6	3378.2
133	24.7	28	744.8	4123.0
161	23.6	28	676.2	4799.2
189	26.6	28	702.8	5502.0
217	27.6	28	758.8	6260.8
245	28.1	28	779.8	7040.6
273	26.3	28	761.6	7802.2
301	21.7	32	758.8	8561.0

Notes:
- length of last interval = (301 - 273) + (305 - 301) = 32 days credited for.
- estimate for missing yield on day of lactation = 49 was calculated as (32.6 + 32.9) /2 = 32.8 kg

APPENDIX F.2 Illustration of the test interval method (France).

Example: *All testday yields available (*SCHEME A8*)*
Mean curve/ Variance curve/ Replicate/ Scheme/ Starting day/ Missing pattern/ = 7/ 4/ 1/ 4/ 21/ 0

Day of lact.	Testday yield	Interval days	Interval yield	Cum. Yield
21	21.4	21	449.4	449.4
77	23.4	56	1254.4	1703.8
133	25.1	56	1358.0	3061.8
189	25.6	56	1419.6	4481.4
245	27.8	56	1495.2	5976.6
301	31.1	56	1649.2	7625.8
305	31.1	4	124.4	7750.2

APPENDIX F.3: Illustration of the centering date method (Denmark)

Example 1: *All testday yields available (Scheme A4)*
SLAC record (Mean curve/ Variance curve/ Replicate No./ Scheme/ Starting date/ Missing pattern/ Code for fixed effects) = 10/4/015/2/07/0/8

Day of lact.	Testday yield	Interval days	Interval yield	Cum. Yield
7	26.9	7	188.3	188.3
35	35.1	28	868.0	1056.3
63	35.1	28	982.8	2039.1
91	31.9	28	938.0	2977.1□
119	27.3	28	828.8	3805.9
147	24.2	28	721.0	4526.9
175	23.2	28	663.6	5190.5
203	27.5	28	709.8	5900.3
231	29.2	28	793.8	6694.1
259	26.6	28	781.2	7475.3
287	22.9	46	1105.2	8580.5

Example 2: *Not all testday yields available (SCHEME A4)*
SLAC record (Mean curve/ Variance curve/ Scheme/ Replicate No./ Starting date/ Missing pattern/ Code for fixed effects) = 10/4/010/2/21/2/8

Day of lact.	Testday yield	Interval days	Interval yield	Cum. Yield
21	32.6	21	684.6	684.6
49	0.0	28	456.4	1141.0
77	32.9	28	460.6	1601.6
105	28.5	28	859.6	2461.2
133	24.7	28	744.8	3206.0
161	23.6	28	676.2	3882.2
189	26.6	28	702.8	4585.0
217	27.6	28	758.8	5343.8
245	28.1	28	779.8	6123.6
273	26.3	28	761.6	6885.2
301	21.7	32	758.8	7644.0

11. REFERENCES

Abramowitz, M., and I. A. Stegun. 1972. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Eds. U.S. Govt. Print. Off., Washington, D.C.

Amos, H. E., T. Kiser, and M. Loewenstein. 1985. Influence of milking frequency on productive and reproductive efficiencies of dairy cows. J. Dairy Sci. 68:732.

Anderson, S. M., I. L. Mao, and J. L. Gill. 1989. Effect of frequency and spacing of sampling on accuracy and precision of estimating total lactation milk yield and characteristics of the lactation curve. J. Dairy Sci. 72:2387.

Appleman, R. D., S. D. Musgrave, and R. D. Morrison. 1969. Extending incomplete lactation records of Holstein cows with varying levels of production. J. Dairy Sci. 52:360.

Auran, T. 1973. Studies on monthly and cumulative monthly milk yield records. I. Effect of age, month of calving, and length of first test period. Acta Scand. 23:189.

Baker, A. M. C. and J. D. Leaver. 1986. Effect of stocking rate in early season on dairy cow performance and sward characteristics. Grass Forage Sci. 41:333.

Bartlett, P. C., J. Van Wijk, D. J. Johnston, and C. D. Green. 1991. Temporal patterns of lost milk production following clinical mastitis in a large Michigan Holstein herd. J. Dairy Sci. 74:1561.

Batra, T. R. 1986. Comparison of two mathematical models in fitting lactation curves for pureline and crossbred dairy cows. Can. J. Anim. Sci. 66:405.

Batra, T. R., and A. J. Lee. 1985. Comparison of three methods of predicting 305-day milk and fat production in dairy cows. Can. J. Anim. Sci. 65:341.

Becker, B. A., H. D. Johnston, R. Li, and r. J. Collier. 1990. Effect of farm and simulated laboratory cold environmental conditions on the performance and physiological responses of lactating dairy cows supplemented with bovine somatotropin (BST). Int. J. Biometeorol. 34:151.

Beran, R. and G. R., Ducharme. 1991. Asymptotic theory for bootstrap methods in statistics. Les publications CRM. Université de Montréal, Québec, Canada.

Bonnier, G. 1935. Is the shape of the lactation curve genetically determined? Hereditas, 20:199.

Brody, S., A. C. Ragsdale, and C. W. Turner. 1923. The rate of decline of milk secretion with the advance of the period of lactation. J. Gen. Physiol. 5:441.

Cerny, B. A., and H. F. Kaiser. 1977. A measure of sampling adequacy for factor-analytic correlation matrices. Multiv. Behav. Res. 12:43.

Chazal, M. P. , and Y. Chilliard. 1986. Effect of stage of lactation, milk yield and herd management on seasonal variation in spontaneous lypolysis in bovine milk. J. Dairy Res. 53:529.

Cobby, J. M., and Y. L. P. Le Du. 1978. On fitting curves to lactation data. Anim. Prod. 26:127.

Congleton, W. R, Jr. and R. W. Everett. 1980a. Error and bias in using the incomplete gamma function to describe lactation curves. J. Dairy Sci. 63:101.

Congleton, W. R, Jr. and R. W. Everett. 1980b. Application of the incomplete gamma function to predict cumulative milk production. J. Dairy Sci. 63:109.

Cooley, W. W., and P. R. Lohnes. 1971. Multivariate data analysis. John Wiley & Sons, New York, NY.

Corley, E. L. 1956. A study of persistency of lactation in dairy catttle. Ph.D. Thesis, University of Wisconsin, Madison, WI.

Cureton, E. E., and S. A. Mulaik. 1975. The weighted varimax rotation and the promax rotation. Psychometrika 40:183.

Danoa, M. S. 1981. A note on an alternative form of the lactation model of Wood. Anim. Prod. 32:349.

Dannell, B. 1981. Evaluation of sires on first lactation yield of Swedish dairy cattle. Ph. D. Thesis. Swedish Univ. of Agric. Sci., Uppsala.

De Boer, J. A., J. I. Weller, T. A. Gipson, and M. Grossman. 1989. Multiphasic analysis of milk and fat yield curves of Israeli Holsteins. J. Dairy Sci. 72:2143.

Dick, I. D. 1950. Some results on the accuracy of milk sampling over a complete lactation and daily variability in milk yields. N. Z. Sci. Technol 32(A):25.

Dickerson, G. E. 1939. The effect of age and dry period on production at different levels of producing ability. Proc. Am. Soc. Anim. Prod. Page. 73.

Dodd, F. H., and A. S. Foot. 1948. Residual milk. Agric. 55:238

Efron, B. 1979. Bootstrap methods: another look at the jacknife. Ann. Statist. 7:1.

Efron, B., and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1:54.

Elvinger, F., R. P. Natzke, and P. J. Hansen. 1992. Interactions of heat stress and bovine somatotropin affecting physiology and immunology of lactating cows. J. Dairy Sci. 75:449.

Erb, R. E., M. M. Goodwin, R. A. Morrison, and A. O. Shaw. 1952. Lactation studies. III. Variations in dialy yields of dairy cows. J. Dairy Sci. 12:967.

Everett, R. W., H. W. Carter, and J. D. Burke. 1951. Evaluation of the dairy herd improvement association record system. J. Dairy Sci. 51:153.

Everett, R. W., and L. H. Wadell. 1970. Sources of variation affecting the difference between morning and evening daily milk production. J. Dairy Sci. 53:1425.

Fieller, E. C. 1954. Some problems in interval estimation. J. Roy. Stat. Soc. 16:175.

Gaines, W. L. 1926. Interpretation of the lactation curve. J. Gen Physiol. 10:27.

Gaines, W. L. 1927a. Persistency of lactation in dairy cows. Ill. Agric. Exp. Stn. Bull. No. 288:355

Gaines, W. L. 1927b. Deferred short-time test as a measure of the performance of dairy cows. J. Agr. Res. 35:237.

Gavin, W. 1912. The interpretation of milk records. J. Royal Agric. Soc. of Engl. 73:153.

Gilbert, G. R., G. L. Hargrove, and M. Kroger. 1973. Diurnal variations in milk yield, fat yield, milk fat percentage, and milk protein percentage of Holstein-Friesian cows. J. Dairy Sci. 56:409.

Goodall, E. A., and D. Sprevak. 1984. A note on a stochastic model to describe the milk yield of a dairy cow. Anim. Prod. 38:1 3.

Goodger, W. J., J. C. Galland, and V. E. Christiansen. 1988. Survey of milking management on large dairies and their relationship to udder health and production variables. J. Dairy Res. 71:2535.

Gooch, M. 1935. An analysis of the time change in milk production in individual lactatations. J. Agri. Sci. 25:71.

Gowen, J. W. 1920. Studies in milk secretion. 5. On the variations and correlations of milk secretion with age. Genetics 5:111.

Grossman, M, and W. J. Koops. 1988. Multiphasic analysis of lactation curves in dairy cattle. J. Dairy Sci. 71:1598.

Grossman, M., A. L. Kuck, and H. W. Norton. 1986. Lactation curves of purebred and crossbred dairy cattle. J. Dairy Sci. 69:195.

Hammond, J., and H. G. Sanders. 1923. Some factors affecting milk yield. J. Agri. Sci. 13:74.

Harman, H. H. 1960. Modern factor analysis. Univ. of Chicago Press, Chicago.

Harris, C. W., and H. F. Kaiser. 1964. Oblique factor analytic solutions by orthogonal transformations. Psychometrika 29:347

Hendrickson, A. E., and P. O. White. 1964. Promax: A quick method for rotation to oblique simple structure. British J. Stat. Psych. 17:65.

Houston, J. 1932. The errors involved in certain methods of estimating lactation yield of milk and butterfat. J. Dairy Res. 4:37.

Hurnik, J. F., G. J. King, and H. A. Robertson. 1975. Estrous and related behaviour in postpartum Holstein cows. Appl. Anim. Ethol. 2:55.

Hurley, J. R., and R. B. Cattell. 1962. The Procrustes program: producing direct rotation to test a hypothesized factor structure. Behav. Sci. 7:258.

Hyde, G. M., H. B. Puckett, E. F. Olver, and K. E. Harshbarger. 1981. A step toward dairy herd management by exception. Trans. Am. Soc. Agric. Eng. Paper 24:202.

Johansson, I., and A. Hansson. 1940. Causes of variation in milk and butterfat yield of dairy cows. Kungl. Lantbruksakadmiens Tidskrift. 6:127.

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23:187.

Karatsuba, A. A., and S. M. Voronin. 1992. The Riemann Zeta-function. Walter de Gruyter, NY.

Kendrick, J. F. 1940. The cow tester's manual. USDA Misc. Pub. 359.

Keown, J. F., and L. D. Van Vleck. 1973. Extending lactation records in progress to 305-day equivalent. J. Dairy Sci. 60:1308.

Keown, J. F., R. W. Everett, N. B. Empet, and L. H. Wadell. 1986. Lactation curves. J. Dairy Sci. 69:769.

Klein, J. W., and T. E. Woodward. 1943. Influcnce of length of dry period upon the quantity of milk produced in the subsequent lactation. J. Dairy Sci. 26:705.

Kumar, R. and P. N. Bhat. 1979. Lactation curve in Indian buffaloes. Indian J. of Dairy Sci. 32:156.

LePage, R. D., and L. Billard. 1992. Exploring the limits of bootstrap. John Wiley and Sons, Inc., New York, New York.

Lewis, G. S. and S. K. Newman. 1984. Changes throughout estrous cycles of variables that might indicate estrus in dairy cows. J. Dairy Sci. 67:146.

Lindgren, N., J. Philipsson, and A. Elofson-Bernstedt. 1980. Studies on monthly protein records of individual cows. Acta. Agric. Scand. 30:437.

Louca, A., and J. E. Legates. 1968. Production loesses in dairy cattle due to days open. J. Dairy Sci. 51:573.

Ludwick, T. M. 1942. Inheritance of persistency in lactation in dairy cattle. Ph.D. Thesis. University of Minnesota, St. Paul, MN.

Mahadevan, P. 1951. The effect of environment and heredity on lactation. II. Persistency of lactation. J. Agric. Sci. 41: 89.

Mainland, D. D. 1985. A note on lactation curves of dairy cows in Scotland. Anim. Prod. 41:413.

Mathews, J. H. 1987. Numerical methods for computer science, engineering, and mathematics. Prentice-Hall, Inc., Englewood Cliffs, NJ. Pages 240-263

Mao, I. L, J. W. Wilton, and E. B. Burnside. 1974. Parity in age adjustment for milk and fat yield. J. Dairy Sci. 57:100.

M'Candlish, A. C., and A. M'Vicar. 1925. Are milk record association results accurate!. Scot. J. Agric. 8:201.

McDowell, J. C. 1927. Testing cows for production every other month. Holstein-Friesian world 24:1970.

Miller, R. H., B. T. McDaniel, and Corley, E. L. 1967. Variation in ratio factors for age-adjusting part-lactation records. J. Dairy Sci. 50:1819.

Miller, P. D., W. E. Lentz, and C. R. Henderson. 1970. Joint influence of month and age of calving on milk yield of Holstein cows in the north-eastern United States. J. Dairy Sci. 53:351.

Molina, J. R. and C. Boschini. 1979. Adjustment of the dairy curve of a Holstein herd with a linear modal model. Agronomia Costarricense 3:167.

Morant, S. V., and A. Gnanasakthy. 1989. A new approach to the mathematical formulation of lactation curves. Anim. Prod. 49:151.

Morrison, D. F. 1976. Multivariate statistical methods. McGraw-Hill Book Co., New York, NY.

Murphy, M. R. 1992. Water metabolism of dairy cattle. J. Dairy Sci. 75:326

Nelder, J. A. 1966. Inverse polynomials, a useful group of multi-factor response functions. Biometrics 22:128.

Noreen, E. W. 1989. Computer intensive methods for testing hypotheses. John Wiley & Sons, Inc. New York, NY.

Ngwerume, F. 1994. Application of a multi-trait animal model to predict next test-day milk production. PhD. Diss., Michigan State Univ., East Lansing.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1984. Variability of test-day milk production and composition and relation of somatic cell counts with yield and compositional changes of bovine milk. J. Dairy Sci. 67:361.

O'Neil, P. V. 1983. Advance engineering mathematics. Wadsworth Publ. Co. Belmont, CA. Pages 618-627.

Palmer, R. W., 1982. Effects of estrus and disease states on daily milk production and their application to computer assisted management of dairy cows. PhD. Diss., Univ. Wisconsin, Madison.

Palmer, R. W., E. L. Jensen, and A. R. Hardie. 1994. Removal of within-cow differences between morning and evening milk yields. J. Dairy Sci. 77:2663.

Papajcsik and Bodero. 1988. Modelling lactation curves of friesian cows in a subtropical climate. Anim. Prod. 47:201.

Pickard, G. C., W. J. Tyler, I. D. Poterfield, and H. O. Henderson. 1952. Persistency in dairy cattle. M.S. Thesis. West Virginia Univ.

Polan, C. E., R. E. Blaser, C. N. Miller, and D. D. Wolf. 1986. Utilization of pasture by lactating cows. J. Dairy Sci. 6:1604.

Putnam, D. N., and H. C. Gilmore. 1970. Factors to adjust milk production to a 24-hour basis when milking intervals are not equal. J. Dairy Sci. 53:685.

Ramirez R., G. Ramirez, nd R. Nunez-Dominguez. 1994. Comparison of equations and estimation procedures of lactation curves in Angus, Brown Swiss, and their crosses. Proc. 5th World Cong. Genet. Appl. Livest. Prod., Guelph, Ontario, Canada. 17:136.

Ripley, R. L., W. L. Tucker, and H. H. Voelker. 1970. Effect of days open on lactation production. J. Dairy Sci. 53:654.

Rowlands, G. J. S. Lucey, and A. M. Russell. 1982. A comparison of different models of the lactation curve in dairy cattle. Anim. Prod. 35:135.

Sanders, H. G. 1923. The shape of the lactation curve. J. Agr. Sci. 13:169.

Sanders, H. G. 1930. The analysis of the lactation curve into maximum yield and persistency. J. Agri. Sci. 20: 145.

Sargent, F. D., V.H. Lyton, and O. G. Wall, Jr. 1968. Test interval method of calculating Dairy Hard Improvement Association records. J. Dairy Sci. 51:170.

SAS®/STAT User's guide, Release 6.08 Edition. 1992. SAS Inst., Inc., Cary, NC.

Schaeffer, L. R., and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Cong. Genet. Appl. Livest. Prod., Guelph, Ontario, Canada. 18:443.

Schaeffer, L. R., and C. R. Henderson. 1971. Effects of days dry and days open non Holstein milk production. J. Dairy Sci. 55:107

Schaeffer, L. R., Jamrozik, J., and A. A. Aarcher. 1994. Multiple trait prediction of 305-day milk, fat, and protein production of dairy cows on automated milking systems. Proc. 29th Session Int. Comm. Anim. Recording, Ottawa, Ontario, Canada.

Schaeffer, L. R., C. E. Minder, I. McMillan, and E. B. Burnside. 1977. Non-linear techniques for predicting 305-day lactation production of Holstein and Jerseys. J. Dairy Sci. 60:1636.

Schmidt, G. H. 1960. Effect of milking intervals on the rate of milk and fat secretion. J. Dairy Sci. 43:213.

Shapiro, H. S. 1969. Smoothing and approximation of functions. Van Nostrand Reinhold Co., New York, NY.

Shook, G. E., E. L Jensen, and F. N. Dickinson. 1980. Factors for estimating sample-day yield in a.m.-p.m. sampling plans. Dairy Herd Improvement Lett. 56(4).

Schultz, A. A. 1974. Factors affecting the shape of the lactation curve and its mathematical description. M. S. Thesis, Univ. Wisconsin, Madison.

Sikka, L C. 1950. A study of lactation as affected by heredity and environment. J. Dairy Res. 17:231.

Simerl, N. A., C. J. Wilcox, and w. W. Thatcher. 1992. Postpartum performance of dairy heifers freshening at young ages. J. Dairy Sci. 75:590.

Singh, 1981. On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. 9:1187.

Smith, J.W., and J. E. Legates. 1962. Relation of days open and days dry to lactation milk and fat yields. J. Dairy Sci. 45:1192.

Stanton, T. L, L R. Jones, R. W. Everett, and S. D. Kachman. 1992. Estimating milk, fat, and protein lactation curves with a test day model. J. Dairy Sci. 75:1691.

Syrstad, O. 1965. Studies on dairy herd records. II. Effect of age and season of calving. Acta. Agric. Scand. 20:31.

Turner, C. W. 1927. Persistency of fat secretion during the lactation period as affected by age. J. Dairy Sci. 10:95.

Turner, C. W. 1943. Factors influencing the persistency of lactation. Guernsey Breeders Journal 64:377.

Trus, D, and L Buttazzoni. 1991. Multiple trait approach to modelling the lactation curve. Proc. 4th World Congr. Genet. Appl. Livest. Prod. Edinburgh, Scotland XIII:492.

Vujicic, I, and B. Bacic. 1961. New equation of the lactation curve. Novi Sad. Ann. Sci. Agric. No. 5.

Westfall, P. H. and S. S. Young. 1993. Resampling-based multiple testing. John Wiley & Sons, Inc. New York, NY.

Wiggans, G. R. 1981. Estimation of lactation projection factors as joined straight lines. J. Dairy Sci. 64:703.

Wiggans, G. R. 1985. Procedures for calculating lactation records. USDA Dairy Herd Improvement Fact Sheet G-1.

Wiggans, G. R. 1986. Projection of protein yield to 305 days. J. Dairy Sci. 69:2908.

Wiggans, G. R, and F. N. Dickinson. 1985. Standardization of NCDHIP dairy cattle lactation records. USDA Dairy Herd Improvement Fact Sheet G-2.

Wiggans, G. R., and Powell. 1980. Projection factors for milk and fat lactation records. USDA Dairy Herd Improvement Lett. 56(1):1.

Wiggans, G. R., and L. D. Van Vleck. 1979. Extending partial lactation milk and fat records with a function of last-sample production. J. Dairy Sci. 62:316.

Wilmink, J. B. M. 1987. Comparison of different methods of predicting 305-d milk yield using means calculated from within-herd lactation curves. Livestock Prod. Sci. 17:1

Wilmink, J. B. M. 1990. Het voorspelllen van een toekomstige produktie bij de berkening van de laktatiewaarde en het NAS-kwotumplan. NAS-report, 90-0700/HW/MD, Netherlands Royal Syndicate, AL Arnhem, The Netherlands.

Wilmink, J. B. M., and W. Ouweltjes. 1991. Computation of lactation yields in the Netherlands. NAS-report, 91-0355/WO/CA, Netherlands Royal Syndicate, AL Arhem, The Netherlands.

Wilton, J. W., E. B. Burnside, and J. Rennie. 1967. The effects of days dry and days open on the milk and butterfat production of Holstein-Friesian cattle. Can. J. Anim. Sci. 47:85.

Wolfram, S. 1988. Mathematica: A System for Doing Mathematics by Computer. Addison-Wesley Publ., Inc., Champaign, Ill.

Wood, P. D. P. 1967. Algebraic model of he lactation curve in cattle. Nature 218:894.

Wood, P. D. P. 1969. Factors affecting the shape of the lactation curve in cattle. Anim. Prod. 11:307.

Wood, P. D. P. 1972. A note on seasonal fluctuations in milk production. Anim. Prod. 15:89.

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.

Woodward, T. E. 1945. Some studies of lactation records. J. Dairy Sci. 28:209.

Yapp, W. W. 1915. Relative reliability of official tests for dairy cows. Ill. Agr. Exp. Sta. Bull. 215., Urbana, Ill.