

Article

# Clustering and Characterization of the Lactation Curves of Dairy Cows Using *K*-Medoids Clustering Algorithm

Mingyung Lee <sup>1,†</sup> , Seonghun Lee <sup>2,†</sup> , Jaehwa Park <sup>2,\*</sup>  and Seongwon Seo <sup>1,\*</sup> 

<sup>1</sup> Division of Animal and Dairy Sciences, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea; mingyung1203@cnu.kr

<sup>2</sup> Department of Computer Science and Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea; gnstjdok@cau.ac.kr

\* Correspondence: jaehwa@cau.ac.kr (J.P.); swseo@cnu.kr (S.S.)

† These authors contributed equally to this work as first author.

Received: 13 July 2020; Accepted: 30 July 2020; Published: 4 August 2020



**Simple Summary:** A lactation curve (LC) provides valuable insights in planning appropriate management strategies related to health, nutrition, and breeding in dairy cows. A clustering based approach on LC patterns analysis is presented. The *k*-medoids algorithm is adopted for the clustering. This approach generates several clusters which have similar milking characteristics of total milk yield, peak milk yield, and days in milk at peak yield. The LCs of some groups represent characteristics of atypical milking patterns which are not considered much in previous approaches, whereas LCs of the other groups show the typical LC patterns similar to the results of previous methods. This approach could be used as a tool to manage an abnormal herd of cows.

**Abstract:** The aim of the study was to group the lactation curve (LC) of Holstein cows in several clusters based on their milking characteristics and to investigate physiological differences among the clusters. Milking data of 330 lactations which have a milk yield per day during entire lactation period were used. The data were obtained by refinement from 1332 lactations from 724 cows collected from commercial farms. Based on the similarity measures, clustering was performed using the *k*-medoids algorithm; the number of clusters was determined to be six, following the elbow method. Significant differences on parity, peak milk yield, DIM at peak milk yield, and average and total milk yield ( $p < 0.01$ ) were observed among the clusters. Four clusters, which include 82% of data, show typical LC patterns. The other two clusters represent atypical patterns. Comparing to the LCs generated from the previous models, Wood, Wilmink and Dijkstra, it is observed that the prediction errors in the atypical patterns of the two clusters are much larger than those of the other four cases of typical patterns. The presented model can be used as a tool to refine characterization on the typical LC patterns, excluding atypical patterns as exceptional cases.

**Keywords:** dairy cow; lactation curve; *k*-medoids clustering; milking characteristics; model fitting

## 1. Introduction

Lactation curves (LC) provide invaluable information that could be used to evaluate the genetic and management status of lactating dairy cows. An LC is a graphical representation of change in milk yield during the lactation period following calving [1]. It typically increases rapidly until the peak yield is achieved and then decreases slowly until the drying period. The shape of an LC, described based on peak yield, peak time, and persistency, is used to determine the milking potential of a dairy cow and is an index that facilitates feed and health management, in addition to breeding decision [2,3]. For a large

group of cows, the mean LC follows a typical pattern, which can be described mathematically using a simple equation [4]. Various equations have been proposed [5–8], and they fit the mean LC of a large group of cows relatively well [2,9].

The shapes of individual LC, however, are quite diverse because LC shape varies based on various factors, including biological factors such as breed, genetic effects, physiological conditions, pregnancy, parity, and age, and environmental factors such as feed, herd management, and calving season [5,10–13]. Some studies have reported atypical shapes, such as the absence of peak yield, in approximately 20~30% of cases [9,14,15]. Despite their considerable proportion, such cases have been considered outliers or disregarded in analyses, as they are difficult to explain using previously developed equations [16–18]. To facilitate precision dairy farming, however, it is useful to identify and analyze LC shapes that can be influenced by individual variation, feed management, or disease. In previous studies, LCs have been analyzed after grouping dairy cows based on production levels, parity differences, or presence of disease [3,19–21]. But the limitations remain because those studies tried to derive LCs for predefined groups instead of extracting physiological characteristics from various LCs.

Clustering analysis facilitates the defining of groups objectively based on phenotypic observations. Particularly, the unsupervised and non-hierarchical clustering methods (i.e., *k*-means and *k*-medoids) are extensively used for the analysis of large datasets because they are powerful and have lower computational cost requirements than hierarchical clustering methods [22,23]. Several studies in the field of dairy science have applied the *k*-means method to classify dairy cows according to genetic status and health status based on peak milk yield, milk components, and blood properties [24–27].

The objective of the present paper, therefore, was to cluster lactation data based on the shapes of LCs using the *k*-medoids, a variant of *k*-means, clustering algorithm. We obtained three years of milking records from commercial Korean dairy farms and used them for the analyses. After clustering, the representative LCs of each cluster were fitted using several conventional LC models. We also investigated differences in physiological and milking characteristics among the clustered groups.

## 2. Backgrounds

An LC can be thought as a time-series of daily milk production for an individual dairy cattle. Therefore, grouping LCs can be handled by time-series clustering. Time-series clustering and its application have been widely studied in various fields [28–32]. Liao T. W. [28] provided guidance on how to apply clustering algorithms to time-series data. Aghabozorgi et al. [31] also classified the time-series clustering into three categories: whole time-series clustering, subsequence time-series clustering and time point clustering. In the study, the authors only focused to whole time-series clustering because subsequence time-series clustering was considered meaningless by Keogh and Lin [33] and time point clustering is similar to time-series segmentation [31].

There are two most extensively used time-series clustering algorithms: the connectivity-based hierarchical clustering and the centroid-based *k*-means clustering. Hierarchical clustering repeatedly merges to form a larger cluster (i.e., agglomerative clustering) or divides a larger cluster into smaller clusters until the cluster is not partitioned (i.e., divisive clustering). This clustering algorithm is often used as a primary prescription for time-series clustering, because it requires no prior understanding of the datasets and clusters [31]. However, due to its high computational complexity, the hierarchical clustering method is not recommended for large datasets [28,32]. Conversely, *k*-means algorithm is a non-trivial algorithm, which implies that it relies heavily on the initial center of the cluster [31,34]. In addition, *k*-means algorithm has a slow initial convergence rate [34]. To address the challenge, a weighted probability distribution can be adopted to initialize the center of the cluster, which is introduced in the *k*-means++ algorithm [34]. In the method, the initial center is selected with probability  $P(x) = D(x)^2 / \sum_{x' \in X} D(x')^2$ , where  $D(x)$  is the shortest distance from the data point to the nearest center already selected. Therefore, the method increases the rate of convergence, and evenly distributes the initial centers.

Although the  $k$ -means algorithm is the most extensively applied, the  $k$ -medoids algorithm has promising applications because it is more robust against noise and outliers [23,35]. For example, Sauder et al. [36] compared the performance of hierarchical clustering and two non-hierarchical clustering algorithms (i.e.,  $k$ -means and  $k$ -medoids) in the classification of Holstein dairy cows based on their growth curves, and reported that the  $k$ -medoids method was the most appropriate for grouping cows. They also observed significant differences in milking performance between the clustered groups.

$K$ -medoids algorithm is a  $k$ -means variant that complements the noise vulnerabilities of  $k$ -means algorithm.  $K$ -means algorithm adopts the mean value as the center of the cluster. Therefore, it naturally suffers from noise [32]. Conversely,  $k$ -medoids algorithm is less sensitive to outliers because it uses the median for the selection of center values, which makes the computation procedure in  $k$ -medoids more complex than the procedure in  $k$ -means algorithm; however,  $k$ -medoids is still a more powerful tool for clustering large datasets than hierarchical clustering [31]. In addition, unlike  $k$ -means clustering,  $k$ -medoids does not require additional calculation for the inter-LC distances whenever the center is updated.

### 3. Materials and Methods

#### 3.1. Dataset

The datasets were collected from January 2016 to December 2018 from four commercial farms in Chungcheong Province, South Korea. An automatic milking system (*Astronaut A4*; Lely Industries NV, Maassluis, the Netherlands) was installed in three farms, while a conventional milking parlor system (DeLaval international AB, Tumba, Sweden) was used in one farm. Milk yield data were stored individually using radio-frequency identification (RFID) tags in both milking systems. A total of 1332 lactations (i.e., records of milk production events during a single lactation period) were recorded from 724 cows (Table 1).

**Table 1.** Descriptive statistics of the raw data of milking records.

	Conventional Milking System		Automatic Milking System		Total	
	Mean	SE	Mean	SE	Mean	SE
Total milking days (day)	103,086		203,444		306,530	
Milking days (day/lactation)	207.00	5.99	243.94	5.20	230.13	3.98
Number of animals	287		437		724	
Number of total lactations	498		834		1332	
- Number of primiparous cows	191		269		460	
- Number of multiparous cows	307		565		872	
Parity	2.24	0.06	2.58	0.06	2.45	0.04
Daily milk yield (L/day/lactation)	30.75	0.52	32.82	0.39	32.06	0.31
Total milk yield (L/lactation)	7056	224	8349	189	7876	146

SE: standard error. SE calculated as the standard deviation divided by the square root of the number of lactations. The values of total milking days and the number of animals and lactations represent the total number, and the others describe the means and standard errors.

The milking records from 10 to 280 days-in-milk (DIM; calving day was considered as 0 DIM) were used. This was because the data before 10 DIM and after 280 DIM were error-prone due to differences in management practices among the four farms. In addition, an ideal LC is formed using a concave function and the peak is often achieved within 10 weeks [37]. Because the peak information is critical source of information in LCs, we set the non-interpolation period within the 10 through 70 DIM. Lactation records were preprocessed according to three criteria for incomplete data (Figure 1).

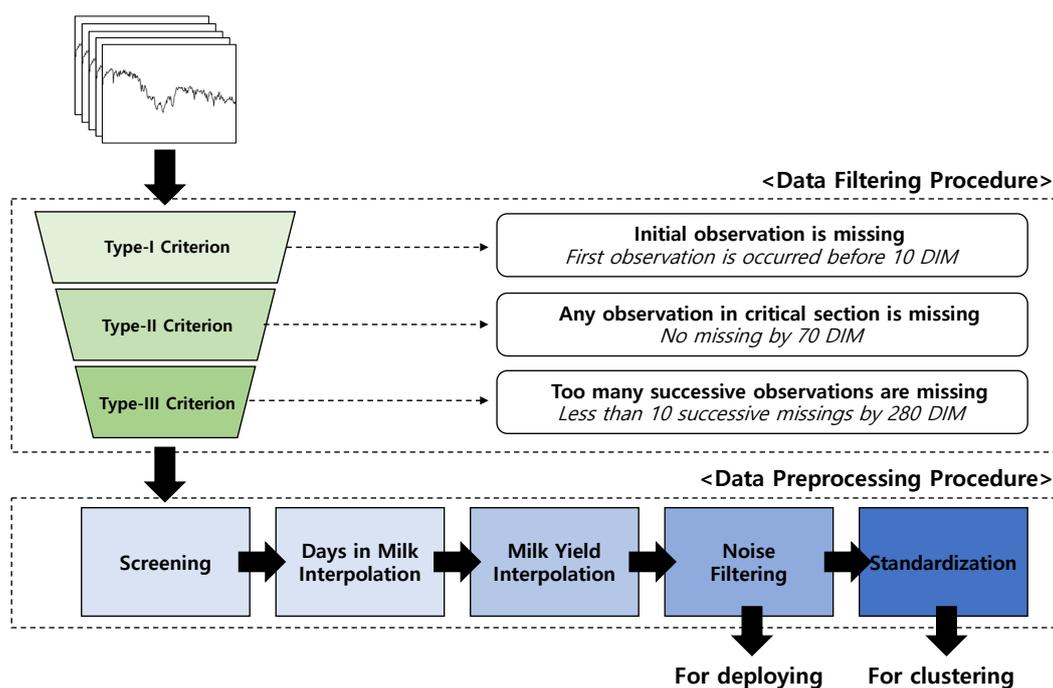


Figure 1. Data filtering and preprocessing procedure.

First of all, data recorded prior to 10 DIM were filtered out (Type-I). Datasets which had any missing records between 10 to 70 DIM were also discarded (Type-II). Finally, lactations that contained missing data for more than 10 consecutive days between 70 to 280 DIM were also discarded dataset (Type-III).

Out of the total 1332 lactation data, 228 (22.8%), 42 (4.2%), and 732 (73.1%) records were excluded via the Type-I, II and III criteria, respectively. Consequently, a total of 330 (24.74%) lactation records remained in the final dataset (Table 2). One-hundred and one (33.6%) records were from primiparous cows (i.e., cows at first lactation), with an average parity of 2.3, which is close to the Korean national average parity [38].

Data preprocessing was performed on the final dataset to fill gaps and minimize noise in the data. Missing values in lactation records were compensated for via the linear interpolation method [39]. Table 2 presents lactation performance values within the dataset. The milking days, average daily milk yield per head, and total milk yield in a lactation increased with an increase in parity number. After the compensation step, noise filtering was also performed. The moving average was used based on a 10-day window.

Before the clustering, the datasets were normalized using Z-score transformation. The normalization process minimizes clustering divergence caused by differences in dynamic range and emphasizes inter-cluster phenotypic homogeneity.

Table 2. Lactation statistics for the filtered data according to each parity.

Parity	Number of Lactations (Cows)	Milking Days		Daily MY		Total MY	
		Mean	SD	Mean	SD	Mean	SD
1	111	358.95	73.55	31.65	4.86	11,303	2649
2	101	377.30	78.56	36.00	5.41	13,545	3403
3	61	375.21	69.06	36.89	6.12	13,786	3310
≥4	57	406.58	81.59	37.09	6.60	14,981	3601

MY: milk yield; SD: standard deviation. All missing values are interpolated.

### 3.2. Clustering

A metric for similarity measures is required for clustering [28,32]. Root mean square (RMS) of Euclidean distance denoted as 'd' is used to measure the similarity between two records as,

$$d(A, B) = \sqrt{\frac{\sum_{i=1}^N (m_{A,i} - m_{B,i})^2}{N}} \quad (1)$$

where:

$d(A, B)$ : Distance between lactation dataset  $A$  and  $B$   
 $m_{A,i}$ : Amount of milk yield on  $i^{th}$  day in lactation dataset  $A$   
 $m_{B,i}$ : Amount of milk yield on  $i^{th}$  day in lactation dataset  $B$   
 $N$ : The total number of milking days

K-medoids clustering, a center-based clustering method, was used. It is advantageous for large datasets, although a pre-defined number of clusters is necessary [28,31,32]. Unlike the  $k$ -means algorithm, the  $k$ -medoids algorithm considers the median as the center of a cluster. Using a median instead of mean enhances robustness against outliers [40]. In the  $k$ -medoids algorithm, a cluster  $S$  is given as:

$$S = \arg \min_S \sum_{A=1}^k \sum_{X \in S} d(X, \mu_A), \quad (2)$$

where:

$\mu_A$ : A median dataset of a cluster  $A$   
 $d(X, \mu_A)$ : Euclidean distance between lactation datasets  $X$  and  $\mu_A$

The elbow method was adopted for the selection of the pre-defined number of clusters. The elbow method is a heuristic approach for determining the point where the local optimum is observed [41]. Six was the most appropriate cluster size with a low root mean square error (RMSE) in our experiments.

### 3.3. Characterization and Comparison of the Clusters

After clustering, differences in lactation and physiological characteristics (i.e., parity, peak milk yield, DIM at peak milk yield, total milk yield from 10 to 280 DIM, the day at 70 DIM in a lactation) were compared among the six clusters. The mean LCs of each cluster and the entire dataset were generated. Subsequently, the mean LCs were regressed to the three of the most popular LC models to investigate differences in estimated parameters and lactation characteristics (i.e., peak milk yield, DIM at peak milk yield, and persistency of milk production) among clusters. The models included the Wood model, an incomplete gamma function proposed by Wood [5], the Wilmink model, a model that combines exponential and linear decline function presented by Wilmink [6], and the Dijkstra model, a mechanical model presented by Dijkstra et al. [7], as shown in Table 3. The residuals of the models did not follow the normal distribution, rather shaped as bell curves.

An HPE Proliant Gen10 DL380 server equipped with 16 core CPUs (2.10 GHz) and 32 GB ECC DDR4 RAM was used for the computation. It took several minutes to group and train data. The least-square method was applied for the regression. Clustering algorithm was implemented in Python (v3.6). For the regression, the "SciPy" package (v1.3.0) was used.

### 3.4. Statistical Analysis

Analysis of variance was performed to evaluate the differences in milking characteristics among clusters. Differences in means were tested using Fisher's Least Significant Difference. Statistical significance was set at  $p < 0.05$ , and  $0.05 \leq p < 0.1$  were considered trends.

Two fitting errors,  $\epsilon_f$  and  $\epsilon_c$  were measured to test fitting performance.  $\epsilon_f$  is the difference between the regression curve and a representative LC of a cluster measured in RMSE. The mean value

of the cluster is chosen as the representative LC.  $\epsilon_c$  is the difference measured in RMSE between the regression curve and all the LCs in a cluster. When calculating  $\epsilon_c$ , all curves are each Z-normalized to compare shapes.

**Table 3.** Traditional regression models.

Model	Lactation Curve
Wood (1967)	$y = a \cdot t^b \cdot e^{-c \cdot t}$ [5]
Wilmink (1987)	$y = a + b \cdot e^{-k \cdot t} + c \cdot t$ [6]
Dijkstra (1997)	$y = a \cdot e^{b \cdot \frac{1-e^{-c \cdot t}}{c} - d \cdot t}$ [7]

$y$ : daily milk yield;  $t$ : time from parturition. All parameters  $a$ ,  $b$ ,  $c$  and  $d$  define the scale and shape of the LC. Wood model [5]: ' $a$ ' represents the general production level, ' $b$ ' represents the increasing period of the lactation before the peak yield, and ' $c$ ' represents the decreasing period of the lactation after the peak yield. Wilmink [6] ' $a$ ' represents the general production level, ' $b$ ' represents the decreasing period of the lactation after the peak yield, and ' $c$ ' represents the increasing period of the lactation before the peak yield, and ' $d$ ' represents the peak day. Dijkstra [7] ' $a$ ' represents the cell population at parturition (theoretical initial milk production), ' $b$ ' represents the cell proliferation rate at birth, ' $c$ ' represents the rate of decrease in cell proliferation during lactation, and ' $d$ ' represents the cell death rate during lactation. All parameters are non-zero positive except  $b$  and  $c$  from Wilmink model.  $b$  and  $c$  in Wilmink model are non-zero negative.

#### 4. Results

The  $k$ -medoids clustering algorithm ( $k = 6$ ) grouped 330 datasets into six clusters of 119, 64, 50, 47, 38, and 12 datasets. The clusters were denoted (a)–(f) (Table 4). The largest cluster (a), had 36% of the datasets, while the smallest cluster (f) had 4% of the datasets, which represented the smallest dataset. Excluding the 70th DIM in a lactation, there were significant differences in parity, average daily milk yield, total milk yield from 10 to 280 DIM, peak DIM, and peak milk yield ( $p < 0.01$ ) among the clusters. Clusters (a) and (d), which had high proportions of multiparous cows, had a parity of 2.6, which was significantly higher than those of other clusters ( $p < 0.05$ ). In contrast, clusters (e) and (f) contained more primiparous cows than multiparous cows, and had lower average parities when compared with the other clusters ( $p < 0.05$ ). The daily average milk yield and total milk yield were the highest in cluster (a) (39.5 L and 10,713 L) and the lowest in cluster (e) (35 L and 9509 L) when compared to other clusters ( $p < 0.05$ ). Peak milk yield was the highest in cluster (d) (54 L), with an 11.6-L difference from cluster (e), which had the lowest peak yield ( $p < 0.05$ ). Clusters (a), (b), and (d) had peak yields at the early lactation period, while clusters (c), (e), and (f) had peak yields at the mid-lactation period. In cluster (e), which had the lowest peak yield, DIM at peak yield was 144 days, which was the latest among the six clusters ( $p < 0.05$ ).

The regression graphs and model parameters of the three conventional models for the average lactation data in the clusters are presented in Figure 2 and Table 5, respectively. Clusters (a) and (b) curves displayed typical shape similar to that of the average LC of the 330 lactations, while clusters (d) and (f) curves had a different shape (Figure 2). Clusters (c) and (e) curves were gentle with lower peak milk yield than clusters (a) and (b). Cluster (d) curves exhibited rapid declines in milk production in the mid-lactation period after the high peak yields, and cluster (f) curves displayed an abnormal shape rather than a general LC. Cluster (f) curve did not fit properly in all three model. Distorted parameter estimates were also derived from the Wilmink model (Table 5). When the  $\epsilon_f$ s of the three models for the clusters were compared, the model with the lowest  $\epsilon_f$  varied across the clusters (Wood model-cluster [e], Wilmink model-clusters [a] and [c], Dijkstra model-cluster [b]). The clusters (d) and (f), which exhibited abnormal shapes, had high  $\epsilon_f$ s (2.53 and 1.96, respectively) on average for all models, unlike other clusters. The  $\epsilon_c$  for the lactation data of the samples within the cluster was the lowest in cluster (a) with the ideal curve shape, and the  $\epsilon_c$  of cluster (f) was twice as high as that of cluster (a).

**Table 4.** Descriptive statistics of each cluster and the entire dataset.

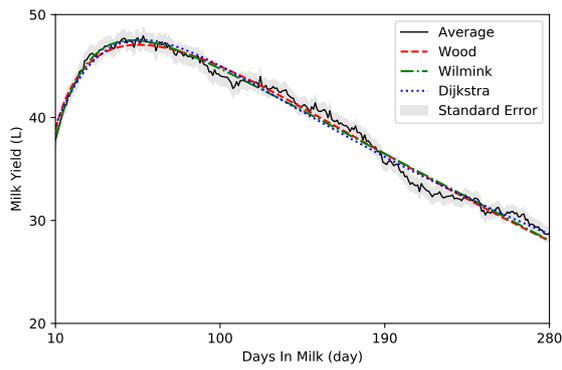
	Cluster						Total	p-Value
	(a)	(b)	(c)	(d)	(e)	(f)		
No. of lactations	119	64	50	47	38	12	330	
Primiparous	17	28	18	10	28	10	111	
Multiparous	102	36	32	37	10	2	219	
Parity	2.61 <sup>a</sup> (0.12)	2.16 <sup>c</sup> (0.17)	2.30 <sup>b</sup> (0.19)	2.57 <sup>a</sup> (0.20)	1.63 <sup>d</sup> (0.20)	1.25 <sup>e</sup> (0.18)	2.31 (0.07)	<0.001 (0.16)
70 DIM in a lactation (days)	Aug. 1. (21.79)	Sep. 29. (24.92)	Oct. 10. (26.28)	Aug. 20. (37.73)	Sep. 30. (32.88)	Jun. 30. (90.25)	Aug. 31. (12.47)	0.272 (28.93)
MY (L/day)	39.53 <sup>a</sup> (0.61)	36.64 <sup>d</sup> (0.96)	38.20 <sup>c</sup> (0.93)	38.76 <sup>b</sup> (1.02)	35.09 <sup>e</sup> (1.13)	36.18 <sup>d</sup> (1.69)	38.03 (0.39)	0.006 (0.88)
MY (L)	10,713 <sup>a</sup> (164.73)	9930 <sup>d</sup> (261.38)	10,351 <sup>c</sup> (252.30)	10,504 <sup>b</sup> (274.96)	9509 <sup>e</sup> (305.46)	9806 <sup>d</sup> (459.28)	10,305 (104.65)	0.006 (238.49)
Peak DIM (days)	59.92 <sup>d</sup> (2.73)	86.25 <sup>c</sup> (4.61)	119.68 <sup>b</sup> (8.42)	54.94 <sup>e</sup> (6.11)	144.00 <sup>a</sup> (9.67)	119.92 <sup>b</sup> (25.73)	85.24 (3.01)	<0.001 (6.01)
Peak MY (L)	53.08 <sup>b</sup> (0.83)	46.39 <sup>d</sup> (1.24)	47.14 <sup>c</sup> (1.08)	54.25 <sup>a</sup> (1.34)	42.70 <sup>e</sup> (1.33)	45.82 <sup>d</sup> (2.26)	49.59 (0.54)	<0.001 (1.13)

MY: milk yield; DIM: days in milk. The number in parentheses represents a standard error, and particularly, the rightmost values mean the pooled standard error. Superscripts that are not common indicate significant differences within the cluster ( $p < 0.05$ ).

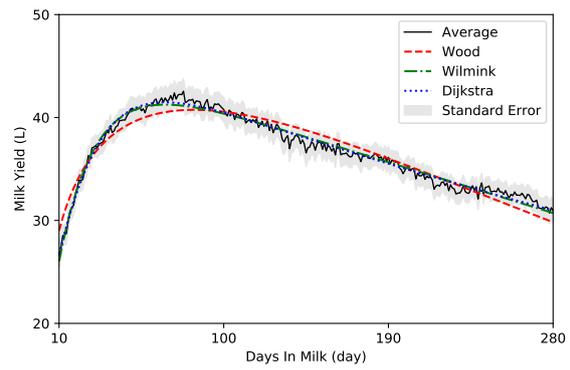
**Table 5.** The regression parameters and the RMSE results per each cluster.

Model	Parameter	Cluster						Total
		(a)	(b)	(c)	(d)	(e)	(f)	
Wood	a	24.6645 (0.3161)	15.9437 (0.3522)	12.5155 (0.3401)	44.8198 (2.1754)	12.2953 (0.1697)	30.2838 (1.3543)	22.1761 (0.2843)
	b	0.2142 (0.0037)	0.2738 (0.0063)	0.3406 (0.0076)	0.0170 (0.0143)	0.2743 (0.0038)	0.0374 (0.0127)	0.2000 (0.0037)
	c	0.0039 (≈0)	0.0033 (0.0001)	0.0035 (0.0001)	0.0016 (0.0001)	0.0018 (≈0)	≈0 (0.0001)	0.0029 (≈0)
	$\epsilon_f$ (L)	0.6680	0.9529	1.1474	2.6746	0.5068 *	1.9778	0.6090
	$\epsilon_c$	0.6125	0.8121	0.8298	0.8827	0.8343 *	1.2126	0.9087
	Wilmink	a	54.2924 (0.1707)	46.0936 (0.1337)	65.3834 (2.2023)	47.6263 (0.3927)	42.0509 (0.4718)	6604.6668 (0.1981)
b		-24.9865 (0.7756)	-31.9822 (0.6105)	-38.4509 (1.9260)	-58.5113 (35.6232)	-21.6106 (0.3697)	-6570.6721 (0.1981)	-25.8178 (0.3245)
c		-0.0937 (0.0009)	-0.0549 (0.0007)	-0.1200 (0.0072)	-0.0594 (0.0023)	-0.0247 (0.0019)	≈0 (9.1156)	-0.0582 (0.0003)
k		0.0494 (0.0020)	0.0495 (0.0012)	0.0124 (0.0009)	0.1715 (0.0496)	0.0195 (0.0008)	≈0 (0.0014)	0.0543 (0.0008)
$\epsilon_f$ (L)		0.6676 *	0.5240	1.0063 *	2.5458	0.5509	1.8807 *	0.2436 *
$\epsilon_c$		0.6119 *	0.7866	0.8230 *	0.8682	0.8359	1.1974 *	0.9006 *
Dijkstra	a	34.0962 (0.5020)	18.8021 (0.3568)	28.2765 (0.4192)	13.6532 (8.0760)	21.3504 (0.2697)	12.0427 (7.7063)	25.1318 (0.2315)
	b	0.0197 (0.0011)	0.0467 (0.0018)	0.0126 (0.0005)	0.1950 (0.1342)	0.0152 (0.0007)	0.1686 (0.1449)	0.0346 (0.0009)
	c	0.0357 (0.0015)	0.0506 (0.0011)	0.0124 (0.0010)	0.1525 (0.0356)	0.0230 (0.0009)	0.1525 (0.0443)	0.0520 (0.0008)
	d	0.0026 (≈0)	0.0015 (≈0)	0.0032 (0.0002)	0.0016 (0.0001)	0.0006 (≈0)	≈0 (0.0001)	0.0016 (≈0)
	$\epsilon_f$ (L)	0.6933	0.4644 *	1.1214	2.3680 *	0.5784	2.0179	0.2479
	$\epsilon_c$	0.6131	0.7852 *	0.8316	0.8510 *	0.8371	1.2012	0.9010

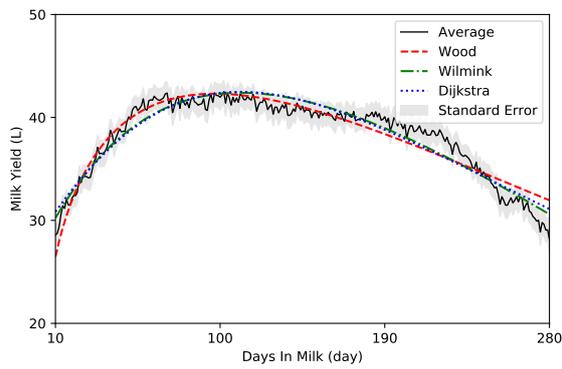
RMSE: root means square error. Cluster means the RMSE calculation is based on each cluster, whereas Total is based on entire data. Models are fit using the averages from the records for each group. The RMSE is calculated between the average fitted model and each record in the group. The numbers in parentheses represent standard error.  $\epsilon_f$  is calculated from based on the cluster average and the model.  $\epsilon_c$  is the average RMSE between each Z-transformed sample in a cluster and a model; \* The lowest error in each cluster.



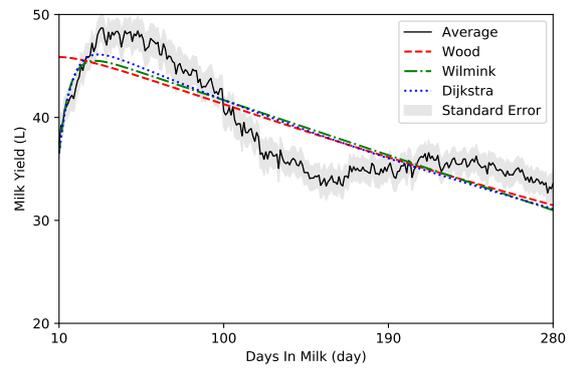
(a) 119 samples



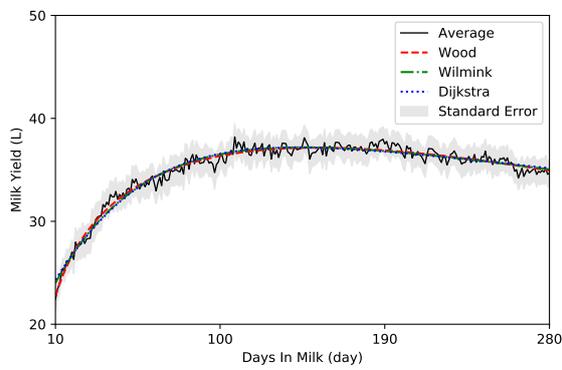
(b) 64 samples



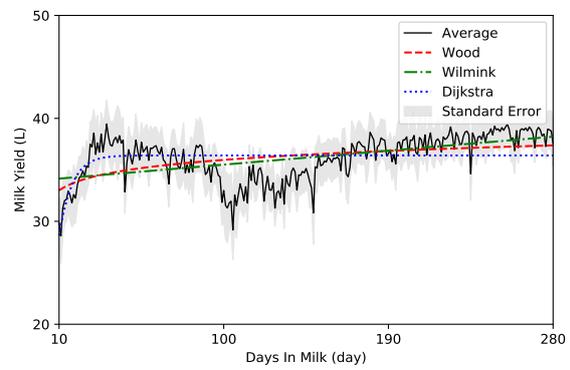
(c) 50 samples



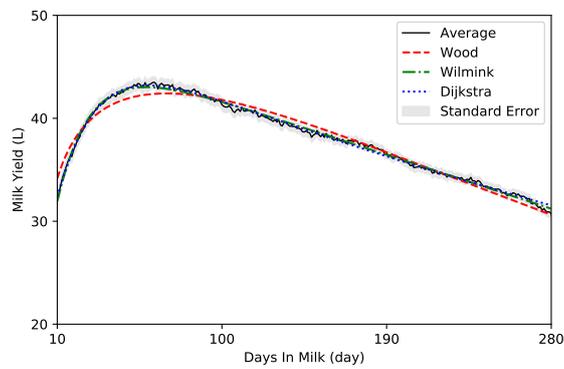
(d) 47 samples



(e) 38 samples



(f) 12 samples



(Total) 330 samples

Figure 2. The averages of each cluster from 330 samples.

The estimated values of the parameters are listed in Table 6, except Wilmink model. Predicted values of peak milk yield were close between the Wood model and the Dijkstra model, but were underestimated by 6 L when compared to the actual lactation data. The greatest difference was 8 L, which was observed in cluster (d). Excluding in cluster (a), the calculated peak DIM value varied between the two models, and there was a gap of 15 days on average. The greatest difference between the models was 20 days, which was observed in cluster (d). Compared to the actual lactation value, the difference in the predicted peak DIM value from two models was relatively low, at 3 days in cluster (a) and (e), but high in cluster (d), at 34 days. Persistency, the relative declining rate at the half point between peak milk yield and the end of lactation, was the lowest in cluster (a) and the highest in cluster (e).

**Table 6.** Peak yield, peak DIM, and persistency calculated using the obtained regression parameters.

	Cluster						Total
	(a)	(b)	(c)	(d)	(e)	(f)	
Wood model <sup>1</sup>							
Peak yield (L)	46.96	40.65	42.34	45.87	37.10	N/A	42.34
Peak DIM (days)	54.92	82.97	97.31	10.63	152.39	N/A	68.97
Persistency	−0.0026	−0.0017	−0.0016	−0.0015	−0.0005	N/A	−0.0017
Dijkstra model <sup>2</sup>							
Peak yield (L)	47.50	41.48	42.37	46.14	37.02	N/A	43.13
Peak DIM (days)	56.73	67.95	110.53	31.50	140.53	N/A	59.11
Persistency	−0.0025	−0.0015	−0.0020	−0.0016	−0.0005	N/A	−0.0016

Peak yield: peak milk yield; Peak DIM: days in milk at the peak milk yield; Persistency: relative rate of decline at the point halfway between peak milk yield and end of lactation [42]; N/A: not applicable; <sup>1</sup> The features of LC were calculated by the equations as follows [5,42]: Peak yield (L),  $a * (b/c)^b * e^{-b}$ ; Peak DIM (days),  $b/c$ ; Persistency,  $b/t_h - c$ ; <sup>2</sup> The features of LC were calculated using the following equations [7,42]: Peak yield (L),  $a * (d/b)^{d/c} * \exp[(b - d)/c]$ ; Peak DIM (days),  $c^{-1} \ln(b/d)$ ; Persistency,  $b * \exp(-c * t_h) - d$ ; \*  $t_h = (t_m + t_f)/2$ ;  $t_m$ , Peak DIM;  $t_f$ , length of lactation.

## 5. Discussion

Clustering of LC shapes yielded clusters (a), (b), (c), and (e), which accounted for 82% of the total individuals, and had typical LC shapes (Table 4). In particular, clusters (a) and (e) showed typical LC shapes of multiparous and primiparous cows, respectively. Primiparous cows generally have a flatter LC and have relatively high persistency, whereas multiparous cows exhibit rapid increases in daily milk yield from calving to the peak milk yield, followed by a significant decline [5,43]. Our study revealed significant differences in milking characteristics such as peak milk yield, peak time, and persistency ( $p < 0.01$ ). The results are consistent with the findings of previous studies [44–49], which reported that total milk yield in a 305-d lactation increases with an increase in the parity of Holstein cows. Particularly, some studies reported that peak milk yield increased dramatically from parity 1 to parity 2, and peak milk yield was observed at later periods in primiparous cows and earlier in multiparous cows [47,48]. Other studies have reported that milk yield is generally higher in multiparous cows than in primiparous cows, while persistency is often greater in the primiparous cows with less developed mammary glands [50–52]. In addition, most primiparous cows are not physically mature [53]. Since primiparous cow require nutrients for their own growth, their metabolic status is different from that of multiparous cows [54]. Such results support the observation that the clustering method used in the present study can discriminate milking characteristics such as total milk yield, peak milk yield, and DIM at peak yield, which vary depending on parity.

Clusters (d) and (f), which accounted for approximately 18% of the cows (Table 4), exhibited abnormal LC shapes (Figure 2). Cluster (f), which consisted of only 12 individuals (4%), had an LC shape with no peaks and no significant changes in milk production. The LC of cluster (d), which had high and rapid peaks, had an undulating shape due to a sharp decrease in milk yield during the

mid-lactation period. Previous studies have reported that LCs with atypical shapes, such as no peak in LC, account for approximately 20~30% individuals in datasets [2,9,13–15], which is consistent with the observations of the present study. Some studies have also reported very high peaks in the early lactation stages are accompanied by sharp decreases in milk yield during subsequent lactation periods, which could be due to metabolic stress and negative energy balance in some instances [55,56]. In addition, previous studies have reported that cows with high milk production and relatively early peak milk yields in early-lactation periods exhibit slow rates of recovery of body condition score, increased physiological stress, and have high risks of udder disease during mid- and late lactation periods [57–59]. The results suggest that the undulating shape of cluster (d) indicated energy unbalance or metabolic disorder. However, such atypical shapes (e.g., non-peak or undulating) of LCs could be caused by individual characteristics [2,9,60]. Arnal, et al. [61] classified LCs of dairy goats using principal component analysis and reported that LCs that were undulating in the mid lactation (120 DIM) period after peak yield were not associated with health problems or environmental factors. To facilitate precision dairy farming, it is critical to determine when such atypical LC shapes emerge, and whether they are attributable to variations, or health problems. Therefore, future studies should develop algorithms that can distinguish such variations or statuses.

The best fitting models are not the same in each clusters as shown in Table 5. The results are consistent with the those of previous studies [3,19]. Previous studies performed goodness of fit tests of the LC model for lactation data by grouping the data according to the period in which the milking data was obtained, the level of milk production, and parity, with the aim of finding a model that optimally describes lactation. Direct comparison on fitting accuracy of presented method to the previous studies is somewhat difficult to perform, since data collection duration, number of cows breed, metric of average milk yield and the data grouping method were not unified among the previous studies. The LC models in previous studies [3,19] were established from the selected groups considering ability of milk production and parity from a large breeding group. Abnormal populations were not considered for the LC model and excluded as exceptional cases. The previous approaches are not appropriate for studying the atypical cases of LC. The LC characteristics of atypical cases would be taken into account using the clustering method.

The three conventional LC models applied in the present study had high fitting errors for clusters with atypical shapes (Table 5). Cluster (d), which had the undulating shape, had the highest fitting error among the clusters, and cluster (f), which exhibited an abnormal shape without peaks, had distorted parameter estimates. Consequently, it was not possible to calculate peak yield, DIM, and persistence in cluster (f) using parameters derived from the model (Table 6). In addition, in cluster (d), the peak milk yield and DIM calculated using model parameters were significantly different from the actual milking data when compared with other clusters. Cluster (f) individuals accounted for only 4% of the individuals in the entire dataset and were less similar between individual LCs in the group. Therefore, the validation of individual data should be performed before further analyses (Table 5;  $\epsilon_c$ ). However, cluster (d) individuals accounted for a considerable proportion of the total data (47 out of 330 cows). In addition, since the similarity between individual LCs in the group was comparable to the similarity observed within other clusters (Table 5;  $\epsilon_c$ ), it can be considered as one of the LC types that can emerge in the commercial farms. In previous studies, LCs with atypical shapes have been omitted from initial datasets or their influence on the overall dataset has been diluted using mean values [16–18]. As mentioned above, such LC types are attributable to diverse factors ranging from simple individual differences to health problems [2,9,55,56,60], so more detailed investigations are required to determine their specific origins. However, in the present study, the conventional model had a high fitting error for LCs with atypical shapes, which is consistent with the findings of previous studies [9,15,62]. Therefore, it is necessary to develop a high-resolution model that can distinguish detailed features with high goodness of fit for diverse LC shapes.

In summary, clusters delineated from LC shape have different milking characteristics due to the varying proportions of multiparous and primiparous cows in each cluster. In addition, 18% of the

individuals had atypical LC shapes. One of the clusters had an undulating LC shape and accounted for 14% of the individuals. This observation is presumably due to metabolic problems caused by rapid peaks and high production at the early stages of lactation. The fit of the model varied for each cluster following the fitting the three popular LC models to average lactation data clusters. We confirmed, however, that the conventional model is not suitable for clusters with atypical shapes due to high fitting errors. The method proposed in the present study may facilitate the identification and management of cows that require attention in herds. One of the practical applications of this study is to predict the daily milk production of individual cows using the cluster information. If the LC of currently milking cow follows a typically shaped cluster, then the regression models of the cluster can be thought as a predictor. The present study, however, was carried out using a limited amount of lactation data, so more commercial farm data should be collected to validate the various LC shapes. In addition, the correlations between LC shapes and the factors influencing such correlations should be investigated further based on biological and environmental data from individuals linked to lactation data. It is also necessary to investigate how many clusters of lactation data are appropriate for application in herd management. Finally, further studies should be conducted to develop a model that can mathematically explain LCs with various shapes.

## 6. Conclusions

This study described the development of a pattern-based clustering method for LCs using *k*-medoids algorithms. This methodology was able to successfully group the individual lactation data with similar milking characteristics (i.e., total milk yield, peak milk yield, and days in milk at peak yield). These results suggest that our methodology can facilitate the identification and management of cows that require attention in the herd.

**Author Contributions:** Writing—original draft, M.L.; formal analysis, S.L.; writing—review and editing, J.P.; supervision, S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Advanced Production Technology Development Program (Project No. 318005-04-2-HD030), Ministry of Agriculture, Food and Rural Affairs, Korea.

**Acknowledgments:** The authors specially thank to Younghwa Ham at AgriRoboTech Co., Ltd. and Jae June Cho at Korea Dairy Committee for providing the data and cooperation for this research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Papajcsik, I.A.; Boderó, J. Modelling lactation curves of Friesian cows in a subtropical climate. *Anim. Sci.* **1988**, *47*, 201–207. [[CrossRef](#)]
2. Olori, V.; Brotherstone, S.; Hill, W.; McGuirk, B. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* **1999**, *58*, 55–63. [[CrossRef](#)]
3. Sherchand, L.; Mcnew, R.W.; Kellogg, D.W.; Johnson, Z.B. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *J. Dairy Sci.* **1995**, *78*, 2507–2513. [[CrossRef](#)]
4. Brody, S.; Ragsdale, A.C.; Turner, C.W. The rate of decline of milk secretion with the advance of the period of lactation. *J. Gen. Physiol.* **1923**, *5*, 441–444. [[CrossRef](#)]
5. Wood, P.D.P. Algebraic model of the lactation curve in cattle. *Nature* **1967**, *216*, 164–165. [[CrossRef](#)]
6. Wilmink, J. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* **1987**, *16*, 335–348. [[CrossRef](#)]
7. Dijkstra, J.; France, J.; Dhanoa, M.; Maas, J.; Hanigan, M.; Rook, A.; Beever, D. A model to describe growth patterns of the mammary gland during pregnancy and lactation. *J. Dairy Sci.* **1997**, *80*, 2340–2354. [[CrossRef](#)]
8. Pollott, G. A biological approach to lactation curve analysis for milk yield. *J. Dairy Sci.* **2000**, *83*, 2448–2458. [[CrossRef](#)]
9. Macciotta, N.; Vicario, D.; Cappio-Borlino, A. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *J. Dairy Sci.* **2005**, *88*, 1178–1191. [[CrossRef](#)]

10. Ferris, T.; Mao, I.; Anderson, C. Selecting for lactation curve and milk yield in dairy cattle. *J. Dairy Sci.* **1985**, *68*, 1438–1448. [[CrossRef](#)]
11. Tassell, C.V.; Jones, L.; Eicker, S. Production evaluation techniques based on lactation curves. *J. Dairy Sci.* **1995**, *78*, 457–465. [[CrossRef](#)]
12. Pérochon, L.; Coulon, J.B.; Lescourret, F. Modelling lactation curves of dairy cows with emphasis on individual variability. *Anim. Sci.* **1996**, *63*, 189–200. [[CrossRef](#)]
13. Tekerli, M.; Akinci, Z.; Dogan, I.; Akcan, A. Factors affecting the shape of lactation curves of Holstein cows from the balikesir province of Turkey. *J. Dairy Sci.* **2000**, *83*, 1381–1386. [[CrossRef](#)]
14. Congleton, W.R.J.; Everett, R.W. Error and bias in using the incomplete gamma function to describe lactation curves. *J. Dairy Sci.* **1980**, *63*, 101–108. [[CrossRef](#)]
15. Rekik, B.; Gara, A. Factors affecting the occurrence of atypical lactations for Holstein–Friesian cows. *Livest. Prod. Sci.* **2004**, *87*, 245–250. [[CrossRef](#)]
16. Daniel, J.B.; Friggens, N.C.; van Laar, H.; Ingvarsten, K.L.; Sauvant, D. Modeling homeorhetic trajectories of milk component yields, body composition and dry–matter intake in dairy cows: Influence of parity, milk production potential and breed. *Animal* **2018**, *12*, 1182–1195. [[CrossRef](#)]
17. Jeretina, J.; Babnik, D.; Škorjanc, D. Modeling lactation curve standards for test–day milk yield in Holstein, Brown Swiss and Simmental cows. *J. Anim. Plant Sci.* **2013**, *23*, 754–762.
18. Pietersma, D.; Lacroix, R.; Lefebvre, D.; Block, E.; Wade, K. A case–acquisition and decision–support system for the analysis of group–average lactation curves. *J. Dairy Sci.* **2001**, *84*, 730–739. [[CrossRef](#)]
19. Cunha, D.D.N.F.V.D.; Pereira, J.C.; Silva, F.F.e.; Silva, O.F.d.; Campos, J.L.; Braga, J.A.M. Selection of models of lactation curves to use in milk production simulation systems. *Rev. Bras. De Zootec.* **2010**, *39*, 891–902. [[CrossRef](#)]
20. Andersen, F.; Østerås, O.; Reksen, O.; Gröhn, Y.T. Mastitis and the shape of the lactation curve in Norwegian dairy cows. *J. Dairy Res.* **2011**, *78*, 23–31. [[CrossRef](#)]
21. Hostens, M.; Ehrlich, J.; Ranst, B.V.; Opsomer, G. On–Farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. *J. Dairy Sci.* **2012**, *95*, 2988–3007. [[CrossRef](#)] [[PubMed](#)]
22. Huang, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min. Knowl. Discov.* **1998**, *2*, 283–304. [[CrossRef](#)]
23. Park, H.S.; Jun, C.H. A simple and fast algorithm for k-medoids clustering. *Expert Syst. Appl.* **2009**, *36*, 3336–3341. [[CrossRef](#)]
24. Huquet, B.; Leclerc, H.; Ducrocq, V. Characterization of French dairy farm environments from herd–test–day profiles. *J. Dairy Sci.* **2012**, *95*, 4085–4098. [[CrossRef](#)]
25. Savegnago, R.P.; do Nascimento, G.B.; de Magalhães Rosa, G.J.; de Carneiro, R.L.R.; Sesana, R.C.; Faro, L.E.; Munari, D.P. Cluster analyses to explore the genetic curve pattern for milk yield of Holstein. *Livest. Sci.* **2016**, *183*, 28–32. [[CrossRef](#)]
26. Koster, J.D.; Salavati, M.; Grelet, C.; Crowe, M.; Matthews, E.; O’Flaherty, R.; Opsomer, G.; Foldager, L.; Hostens, M. Prediction of metabolic clusters in early–lactation dairy cows using models based on milk biomarkers. *J. Dairy Sci.* **2019**, *102*, 2631–2644. [[CrossRef](#)]
27. Grelet, C.; Vanlierde, A.; Hostens, M.; Foldager, L.; Salavati, M.; Ingvarsten, K.L.; Crowe, M.; Sorensen, M.T.; Froidmont, E.; Ferris, C.P.; et al. Potential of milk mid–IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. *Animal* **2019**, *13*, 649–658. [[CrossRef](#)]
28. Liao, T.W. Clustering of time series data—A survey. *Pattern Recognit.* **2005**, *38*, 1857–1874. [[CrossRef](#)]
29. chung Fu, T. A review on time series data mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181. [[CrossRef](#)]
30. Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. On clustering validation techniques. *J. Intell. Inf. Syst.* **2001**, *17*, 107–145. [[CrossRef](#)]
31. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
32. Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [[CrossRef](#)]
33. Keogh, E.; Lin, J. Clustering of time-series subsequences is meaningless: Implications for previous and future research. *Knowl. Inf. Syst.* **2005**, *8*, 154–177. [[CrossRef](#)]

34. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms; Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 7–9 January 2007; pp. 1027–1035.
35. Madhulatha, T.S. Comparison between k-means and k-medoids clustering algorithms. In *Advances in Computing and Information Technology*; Wyld, D.C., Wozniak, M., Chaki, N., Meghanathan, N., Nagamalai, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 472–481.
36. Sauder, C.; Cardot, H.; Disenhaus, C.; Le Cozler, Y. Non-parametric approaches to the impact of Holstein heifer growth from birth to insemination on their dairy performance at lactation one. *J. Agric. Sci.* **2013**, *151*, 578–589. [[CrossRef](#)]
37. National Research Council. *Nutrient Requirements of Dairy Cattle: Seventh Revised Edition, 2001*; The National Academies Press: Washington, DC, USA, 2001. [[CrossRef](#)]
38. Ministry of Agriculture, Food and Rural Affairs. *2018 DHI Annual Report in Korea*; Dairy Cattle Improvement Center, National Agricultural Cooperative Federation: Gyeonggi-do, Goyang-si, Korea, 2018.
39. Meijering, E. A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proc. IEEE* **2002**, *90*, 319–342. [[CrossRef](#)]
40. Velmurugan, T.; Santhanam, T. Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points. *J. Comput. Sci.* **2010**, *6*. [[CrossRef](#)]
41. Ketchen, D.J.; Shook, C.L. The Application of cluster analysis in strategic management research: An analysis and critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [[CrossRef](#)]
42. Thornley, J.; France, J. *Mathematical Models in Agriculture. Quantitative Methods for the Plant, Animal and Ecological Sciences*, 2nd ed.; CABI: Wallingford, CT, USA, 2007.
43. Wood, P. Factors affecting the shape of the lactation curve in cattle. *Anim. Sci.* **1969**, *11*, 307–316. [[CrossRef](#)]
44. Arbel, R.; Bigun, Y.; Ezra, E.; Sturman, H.; Hojman, D. The effect of extended calving intervals in high lactating cows on milk production and profitability. *J. Dairy Sci.* **2001**, *84*, 600–608. [[CrossRef](#)]
45. Dematawewa, C.; Berger, P. Genetic and phenotypic parameters for 305-day yield, fertility, and survival in Holsteins. *J. Dairy Sci.* **1998**, *81*, 2700–2709. [[CrossRef](#)]
46. Lee, J.Y.; Kim, I.H. Advancing parity is associated with high milk production at the cost of body condition and increased periparturient disorders in dairy herds. *J. Vet. Sci.* **2006**, *7*, 161–166. [[CrossRef](#)] [[PubMed](#)]
47. López, S.; France, J.; Odongo, N.; McBride, R.; Kebreab, E.; AlZahal, O.; McBride, B.; Dijkstra, J. On the analysis of canadian holstein dairy cow lactation curves using standard growth functions. *J. Dairy Sci.* **2015**, *98*, 2701–2712. [[CrossRef](#)] [[PubMed](#)]
48. Weller, J.; Ezra, E. Genetic and phenotypic analysis of daily Israeli Holstein milk, fat, and protein production as determined by a real-time milk analyzer. *J. Dairy Sci.* **2016**, *99*, 9782–9795. [[CrossRef](#)] [[PubMed](#)]
49. Ray, D.; Halbach, T.; Armstrong, D. Season and lactation number effects on milk production and reproduction of dairy cattle in arizona. *J. Dairy Sci.* **1992**, *75*, 2976–2983. [[CrossRef](#)]
50. Weber, M.; Purup, S.; Vestergaard, M.; Akers, R.; Sejrsen, K. Regulation of local synthesis of insulin-like growth factor-I and binding proteins in mammary tissue. *J. Dairy Sci.* **2000**, *83*, 30–37. [[CrossRef](#)]
51. Sorensen, A.; Knight, C. Endocrine profiles of cows undergoing extended lactation in relation to the control of lactation persistency. *Domest. Anim. Endocrinol.* **2002**, *23*, 111–123. [[CrossRef](#)]
52. Miller, N.; Delbecchi, L.; Petittlerc, D.; Wagner, G.; Talbot, B.; Lacasse, P. Effect of stage of lactation and parity on mammary gland cell renewal. *J. Dairy Sci.* **2006**, *89*, 4669–4677. [[CrossRef](#)]
53. Coffey, M.; Hickey, J.; Brotherstone, S. Genetic aspects of growth of Holstein-Friesian dairy cows from birth to maturity. *J. Dairy Sci.* **2006**, *89*, 322–329. [[CrossRef](#)]
54. Wathes, D.; Cheng, Z.; Bourne, N.; Taylor, V.; Coffey, M.; Brotherstone, S. Differences between primiparous and multiparous dairy cows in the inter-relationships between metabolic traits, milk yield and body condition score in the periparturient period. *Domest. Anim. Endocrinol.* **2007**, *33*, 203–225. [[CrossRef](#)]
55. Swalve, H.H. Theoretical basis and computational methods for different test—Day genetic evaluation methods. *J. Dairy Sci.* **2000**, *83*, 1115–1124. [[CrossRef](#)]
56. Jakobsen, J.; Rekaya, R.; Jensen, J.; Sorensen, D.; Madsen, P.; Gianola, D.; Christensen, L.; Pedersen, J. Bayesian estimates of covariance components between lactation curve parameters and disease liability in Danish Holstein cows. *J. Dairy Sci.* **2003**, *86*, 3000–3007. [[CrossRef](#)]

57. Lassen, J.; Hansen, M.; Sørensen, M.K.; Aamand, G.P.; Christensen, L.G.; Madsen, P. Genetic relationship between body condition score, dairy character, mastitis, and diseases other than mastitis in first—Parity Danish Holstein cows. *J. Dairy Sci.* **2003**, *86*, 3730–3735. [[CrossRef](#)]
58. Hansen, J.; Friggens, N.; Højsgaard, S. The influence of breed and parity on milk yield, and milk yield acceleration curves. *Livest. Sci.* **2006**, *104*, 53–62. [[CrossRef](#)]
59. Yamazaki, T.; Takeda, H.; Nishira, A.; Togashi, K. Relationship between the lactation curve and udder disease incidence in different lactation stages in first-lactation Holstein cows. *Anim. Sci. J.* **2009**, *80*, 636–643. [[CrossRef](#)]
60. Silvestre, A.; Petim-Batista, F.; Colaço, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *J. Dairy Sci.* **2006**, *89*, 1813–1821. [[CrossRef](#)]
61. Arnal, M.; Robert-Granié, C.; Larroque, H. Diversity of dairy goat lactation curves in France. *J. Dairy Sci.* **2018**, *101*, 11040–11051. [[CrossRef](#)]
62. Silvestre, A.; Martins, A.; Santos, V.; Ginja, M.; Colaço, J. Lactation curves for milk, fat and protein in dairy cows: A full approach. *Livest. Sci.* **2009**, *122*, 308–313. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).