A time-series analysis of milk productivity changes in US dairy states

M. Li [1] K.F. Reed [2] and V.E. Cabrera[1]

[1]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison WI 53705, USA
Corresponding Author: vcabrera@wisc.edu
[2]Department of Animal Science, Cornell University, 272 Morrison Hall, Ithaca, NY 14850, USA

Abstract

It is important to characterize trends and seasonal patterns to project amounts and fluctuations in milk and milk components by states or regions. Hence, this study aimed to 1) quantify historical trends and seasonal patterns of milk and milk components production by US states; 2) classify states with similar trends and seasonal patterns into clusters; and 3) summarize the general pattern for each cluster for further applications in simulation models. Our dataset contained 9.18 million lactation records from 5.61 million Holstein cows distributed in 17 states during the period January 2006 to December 2016. Each record included a cow's total milk, fat, and protein yield during a lactation. We used time series decomposition to obtain each state's annual trend and seasonal pattern in milk productivity. Then, we classified states with agglomerative hierarchical clustering into groups according to 2 methods: (1) dynamic time warping on the original time series and (2) Euclidean distance on extracted features of trend and seasonality from the decomposition. Results showed distinguishable trends and seasonality for all states and lactation numbers. The clusters and cluster centroid pattern showed a general upward trend for all yields (ECM, milk, fat, and protein) and a steady trend for fat and protein percent for all states except Texas.

We also found a larger seasonality amplitude for all yields (ECM, milk, fat, and protein) in higher lactation numbers and a similar amplitude for fat and protein percent across lactation numbers. The results could be used for advising management decisions according to farm productivity goals. Furthermore, the trend and seasonality patterns can be used to adjust the production level in a specific state, year, and season for farm simulations to accurately project milk and milk components production.

Keywords: Time series, calving time, production.

Introduction

Dairy cow milk productivity in the US grew by 14%, from 19,895 kg to 22,761 kg, between 2006 and 2016 (USDA-NASS, 2020). Also, there is a recognized seasonal pattern of milk yield and milk components (Salfer *et al*., 2019; Ferreira *et al*., 2020). Moreover, milk yield per cow varies significantly between states and regions (USDA-ERS, 2021). Therefore, it is challenging to predict milk production by location, year, and season. Providing an accurate estimate of lactation performance would help simulation models at the farm and market level that inform management and policy actions.

Milk and milk components productivity according to calving date can be represented as time series. Time series consist of a set of variables and the time-sequential information reliant on them. Time series can be decomposed to identify trends and seasonal patterns (Hyndman and Athanasopoulos, 2018). This study aims to 1) quantify historical trends and seasonal patterns of milk and milk component production in the US; 2) cluster states with similar trends and seasonal patterns; and 3) describe the overall pattern for each cluster for future applications in simulation models.

Material and methods

Dataset

We used a large dataset provided to us by the US Council on Dairy Cattle Breeding (https://www.uscdcb.com/). The dataset included Holstein lactation records with 7 main variables: milk yield (kg/cow for a whole lactation), fat yield (kg/cow for a whole lactation), protein yield (kg/cow per lactation), lactation starting date, parity, US state, and lactation length (d) across an 11-year period (January 2006 to December 2016). We filtered lactation records so that each state had at least 100,000 lactation records, resulting in data from 17 states containing 9,184,086 million lactation records from 5,606,351 million Holsteins. We computed fat percent, protein percent, and ECM (IFCN, 2010) from the milk, fat, and protein yields. Then, we averaged the lactation-length yields for milk, fat, protein, and ECM, as well as the fat percent and protein percent by the week when lactations started.

Time series decomposition

We utilized R (R Core Team, 2020) to decompose the time series of each response using an additive model. The trend component indicated long-term change, the seasonal component reflected a cyclical process, and the residual component reflected short-term influences across the series' period (Hyndman and Athanasopoulos, 2018).

Time series clustering

To categorize all state and parity combinations, we executed agglomerative hierarchical clustering using Ward's minimal variance approach (Murtagh and Legendre, 2014) by minimizing the total variance. We determined the distance between each of the two time series in the original data using dynamic time warping distance and the trend and seasonality components using Euclidean distance. We clustered 18 time series groups into hierarchical cluster trees: 3 types of time series (original, trend, and seasonal) × 6 variables (milk yield, protein yield, fat yield, ECM yield, protein percent, and fat percent). For each one of these 18 time series groups, there were 3 parity groups (1st, 2nd, and 3rd + lactations) ×17 states: 51 time series into one hierarchical cluster tree.

Results and discussion

Each decomposed time series contained yearly trend and seasonality components. Figure 1 shows an example of the original ECM yield for New York state from 3rd + lactations data along with their decomposed components. The clustering method allowed the decomposed variable components to be grouped into clusters based on distance measurements of each pair of two time series. We identified the optimal number of clusters for each group to be 5 by tests on grouping indications. Figure 2 shows the dendrogram of clustering results of ECM yield to illustrate clusters for the original, seasonal, and trend time series.

For the clusters on original data, ECM, milk yield, fat yield, and protein yield had similar patterns for classification where the 1st lactations were mostly classified into 2 clusters that were relatively independent from 2nd and 3rd + lactations. A visual evaluation of the 1st lactation clusters indicated a lower production level and less variability. The 2nd and 3rd + lactations of the same state were commonly classified into the same group, and the centroid patterns of the two major 2nd and 3rd + lactations clusters showed a difference in production levels. The average yield of the 1st lactation and 2nd lactations were 85.4% and 98.3% of the 3rd lactations yields, respectively. The classification pattern for fat percent and protein percent was different from all yields and from each other. Regardless of lactation number, the time series of the same state were clustered together for fat percent and therefore mainly influenced by the geographic factors rather than parity effect. In general, fat percent showed greater variability and less cyclicity than protein percent, suggesting that the protein percent was less sensitive to other factors than the seasonal factors.

For clusters on trends, the classification patterns were like the original data in terms of grouping 2nd and 3rd + lactations from the same state into the same cluster for all yields (ECM, milk, fat, and protein), all lactation from the same state into the same cluster for fat percent, and 1st and 2nd lactations of the same state into the same cluster for protein percent. All yields cluster centroid patterns exhibited a rising. In contrast, the trends for fat and protein percent were relatively steady, and Texas even showed a decline trend in fat percent and protein percent. Florida had the lowest fat percent and protein percent levels, whereas California had the lowest fat percent level, but the highest protein percent for both 1st and 2nd lactations.

For clusters on the seasonal pattern, all the yield clusters had one major cluster for each of 1st, 2nd, and 3rd + lactations. Since our time in each time series was representing the week of calving, we found the valley of the centroid patterns occurred during summer calvings (mainly the last week of July for 2nd lactation, and the 2nd week of July for 3rd lactation), whereas the peak occurred in late January for 2nd lactation and mid-December for 3rd + lactations. For the 1st lactation, milk yields centroid pattern did not vary much before August. All lactations in Virginia, the 2nd lactations in Illinois, and the 2nd and 3rd + lactations in Georgia showed a distinct milk yield centroid pattern in which the yield did not decrease until May. Fat yield in some groups, particularly 3rd +

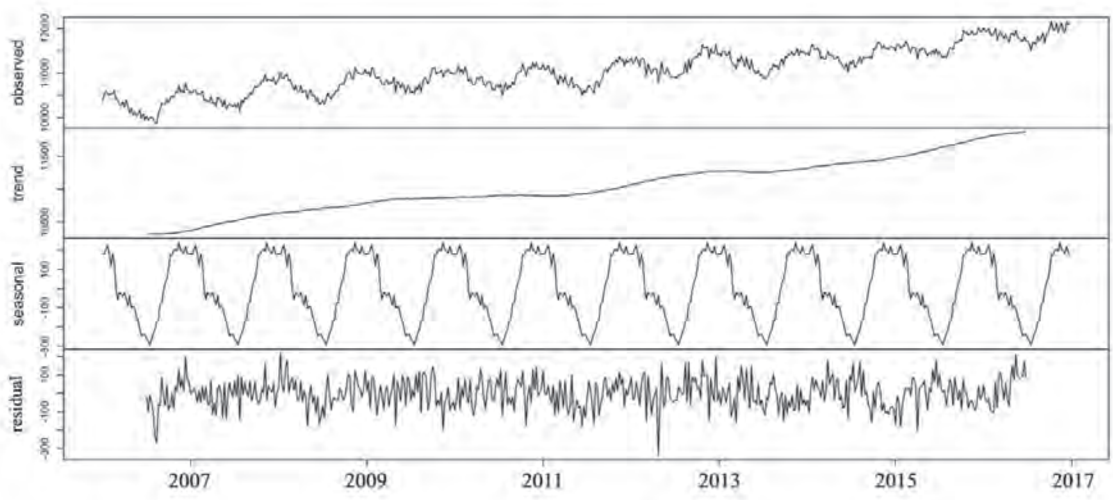

Figure 1. Time series decomposition of ECM yield for 3rd + lactations in New York . All units are in kg/cow/year.

lactations showed a valley prior to late July. Protein yield for 3rd + lactations had a valley between late June and early July. These valleys for components occurred before the milk yield, leading to an earlier valley for ECM. The fat and protein percent, where the 2nd and 3rd + lactations from the same state were clustered together, showed slopes throughout all lactations that did not appear to be different from one another. Peaks of the centroid patterns for fat percent happened during summer weeks (especially early August) and shortly after the protein percent peak. Protein percent exhibited a consistent and clear seasonal pattern across states and lactations, with peaks around late July to early August, a steady decline from November to March or April, and an increase thereafter.

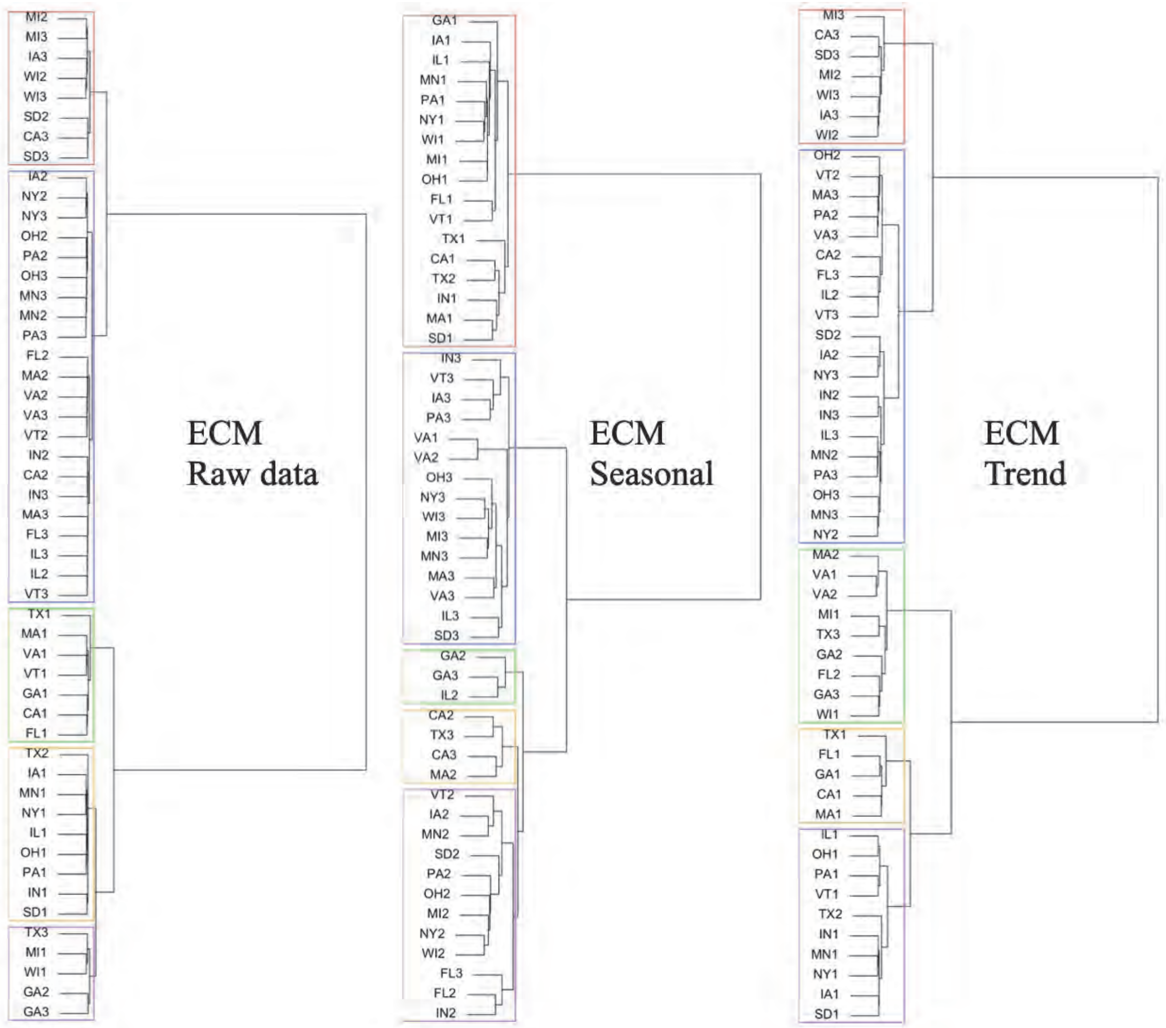

Figure 2. Dendrogram illustrating the groups formed by the cluster analysis for (a) the original data, (b) the seasonal, and (c) the trend of time series of ECM on calving date for each state and parity group. MI2 = 2nd lactation in Michigan.

Conclusion

Time series of all yields (ECM, milk, fat, and protein) were largely influenced by the long-term effect of the trend and the periodic effect of seasonality. Analysis of the long-term trend showed a general upward pattern across the years for all yields and a general flat pattern across the years for fat and protein percent. Analysis of seasonal trends showed a stronger effect of seasonality for later lactations for all yields and that seasonality of fat and protein percent are not influenced by parity. Cluster results showed distinct groups with closer long-term trend and seasonality patterns among lactations and locations. The improved classification of the original and the decomposition parameters can assist with herd and cow management decisions demonstrating the importance of seasonal patterns in production variables according to geographical location and parity.

References

Hyndman, R. J., and Athanasopoulos, G. 2018. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2

IFCN Dairy Research Center. 2010. IFCN Dairy Report 2010 – For a Better Understanding of Milk Production World-Wide. 214 pages, IFCN – International Farm Comparison Network, Kiel

Ferreira, F. C., J. S. Clay, and A. De Vries. 2020. Distribution of seasonality of calving patterns and milk production in dairy herds across the United States. J. Dairy Sci 103:8161–8173.

Murtagh, F. and P. Legendre. 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. 31(3): 274-295.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Salfer, I. J., C. D. Dechow, and K. J. Harvatine. 2019. Annual rhythms of milk and milk fat and protein production in dairy cattle in the United States. J. Dairy Sci. 102:742–753.

USDA-ERS. 2021. Milk cows and production by state and region (Annual). https://www.ers.usda.gov/data-products/dairy-data/

USDA-NASS. 2020. Agricultural Marketing Service, and Economic Research Service calculations. Accessed May 29, 2020. https://downloads.usda.library.cornell.edu/usda-esmis/files/c821gj76b/jm2159057/qv33sh51q/agpr0520.pdf