ICIC Express Letters
Part B: Applications
Volume **14**, Number **1**, January **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **73–79**

PREDICTING CALVING TIME OF DAIRY COWS BY AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (*ARIMA*) MODEL AND EXPONENTIAL SMOOTHING MODEL

Tunn Cho Lwin, Thi Thi Zin and Pyke Tin

Graduate School of Engineering
University of Miyazaki
1-1, Gakuen Kibanadai-Nishi, Miyazaki 889-2192, Japan
z322t02@student.miyazaki-u.ac.jp; thithi@cc.miyazaki-u.ac.jp; pyketin11@gmail.com

Received May 2022; accepted July 2022

Abstract. *Calving time prediction is an important factor in dairy farming. The careful monitoring of cows can help to decrease the loss of calf rates during the calving time; moreover, to know the exact time of birth is crucial to make sure timely assistance. However, direct visual observation is time-wasting for observers, and the continuous presence of observers during calving time may disturb cows. In this study, the recording from video cameras and counting the number of standing-to-lying and lying-to-standing transitions of 25 cows from a large farm at Oita prefecture, Japan, before 72 hours of calving time are applied. To be specific, we model the number of changes in behaviors of standing and lying as a time series in hourly basis. The time series approaches, namely the exponential distribution probability, autoregressive integrated moving average (ARIMA) model, and double exponential smoothing (DES) model, are applied to predicting the calving time and the root mean square error (RMSE) is used to check the accuracy and error value of the experiment. By investigating the changes in behavior patterns a few days before the calving events, the proposed method can predict accurately the time of occurrence of calving events by the developed ARIMA (2,0,1) model. Therefore, the developed model can be used to estimate the calving time which has significantly positive impact for livestock specialists.*
Keywords: *ARIMA* model, *DES* model, Calving time prediction, Transition changes, Exponential distribution, Time series analysis

1. **Introduction.** Among agricultural sectors, dairy farming is one of the most important segments all over the world as dairy products are necessary nutrients of humans, and it can also be expressed as vital earning of farmers who are majorities in most of the countries. For dairy farming, most relevant and efficient forms of livestock are dairy cows, which can also be called as dairy cattle. In the cattle business, calf losses are regarded as a major concern since it has significant impacts on profitability and reproduction. According to numerous researches, it is indicated that calf losses can be averted if intervention in the process of parturition is fulfilled properly.

Most of the calf loss cases can be assumed with the cause of perinatal mortality, which is known as death of perinate within 48 hours of calving, and at least 260 days of gestation periods are also included. Therefore, difficulties of calving had become a concern of all cattlemen as it can cause calf deaths and cause rebreeding failures of the second time in reducing calf crop rates. That is, the highest possibility of calving difficulties can have lower fertility at rebreeding significantly [1]. For calving, 75% of calf died within an hour of calving, 10% were in the pre-calving stage and post-partum death was 15% [2]. Therefore, in order to be safe for both cattle and new born calf, predicting calving time plays an important factor.

DOI: 10.24507/icicelb.14.01.73

Therefore, to prevent perinatal mortality, researchers tried to predict calving stages with different approaches. Automated devices were used to predict calving time by measuring tail raising, lying time, uterine contractions, measuring vaginal temperature [3]. Before the calving period, the number of standing positions between lying rest moments is increased with discomforts. During 72 hours before calving time, numerous numbers of clinical signs can be observed which are hormonal changes. Furthermore, behavioral changes can be observed directly which can be seen several days before parturition, sometimes, those can be observed 4 to 10 days before calving and sometimes, only on the actual day of calving [4].

Accurate prediction of calving time is patently helpful to cattle ranchers and livestock specialist, as it is considered as a key factor in preparation of preventive measures to minimize perinatal mortality rate and avoid cases like dystocia. Nonetheless, constant observation is inadvisable, by reason of causing potential disturbance to cows, and time-wasting for observers [5]. For these reasons, in this paper, based on hourly data three days before calving time, wherein standing-to-lying and lying-to-standing transition is used, we will apply the time series analysis that includes autoregressive integrated moving average ($ARIMA$) model and survival probability with exponential distribution, which is described as a scientifically proven method and one of the most efficient methodologies in data modeling of time and forecasting, to predicting calving time.

2. **Literature Review.** The continuous monitoring systems can test the behavioral changes occurring on the day of calving, most of them being accentuated within a previous couple of hours before delivery, standing/lying transitions, tail raising, and feeding time. Applying these behavioral changes has the potential to improve the management of calving. Prediction is the operation of constructing an assumption about the future values of studied variables. A time series is nothing but observations per the chronological order of time [6]. By time series analysis, the forecasting accuracies rely on the characteristic of time series. If the transition curves show stability and periodicity, we will reach high forecasting accuracies, whereas we cannot hope high accuracies if the curves contain highly irregular patterns [7].

There are several researches that have been trying to forecast calving time with varying results. According to the findings of [8], predicting the calving time of dairy cows by using the actual collected transition data with Geometric and Poisson distribution, the nature geometric distribution is suitable for determining the calving process namely calving or not calving so that it is useful to forecast the waiting time to the success or the occurrence of the calving event. Since the use of the data are transition of one posture to another, the Poisson distribution is suitable and the two proposed models are good enough for the calving time prediction.

In a machine-learning-based calving prediction system [9], the researchers collected data from 20 primiparous and 33 of multiparous dairy cattle. In addition, they used the HR tag, which collects neck activity and rumination data in 2-hour increments, and IceQube which collects the number of steps, lying, and standing times. The machine-learning method was effective in performing retrospective calving prediction. By monitoring the activities and standing-lying behaviors, it could accurately predict the calving day 8 hours before calving time.

3. **Proposed Methods.** In this study, the exponential distribution probability, $ARIMA$ modeling, DES model and the root mean square error ($RMSE$) are applied because the observed transition data is distributing with exponential distribution and both of the forecasting models are principally adopting univariate time series analysis.

Definition 3.1. *The exponential distribution probability is a type of continuous probability distribution that anticipates the time interval between successive events. It is one of the widely used continuous distributions to model the time elapsed between events.*

The probability density function [10] of exponentially distribution is given by Equation (1),

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Equivalently, its distribution function is given by Equation (2),

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where λ is the parameter of exponential distribution and x is continuous nonnegative random variable.

The detailed workflow of *ARIMA* model is expressed in Figure 1. To summarize that, there are three main steps [11]. They are (i) model identification – the use of the data, and of any information on how the series was generated by stationarity, (ii) parameter estimation – use of the data to make inferences about the parameters conditional on the adequacy of the model, and (iii) goodness-of-fit – checking the fitted model in its relation to the data with intent to reveal model inadequacies.

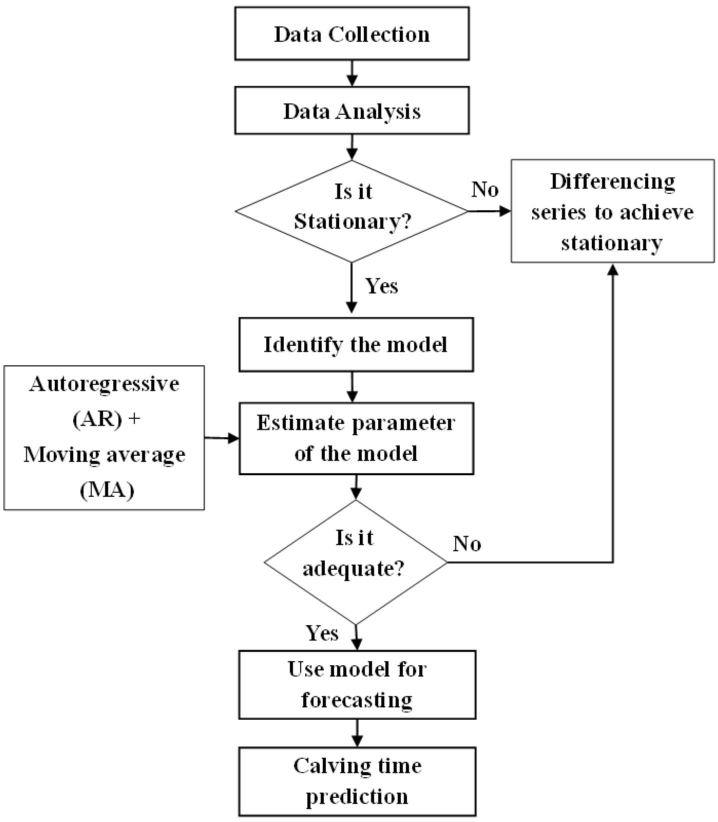

Figure 1. Flowchart of *ARIMA* modeling

Definition 3.2. *In autoregressive integrated moving average, ARIMA (p,d,q) model, we are able to see the order of AR(p) and MA(q) only. Integrating or differencing term, d, is for non-stationary series. In the model, the predicted values of the process are described as the sum of finite linear combination of previous values and random error (white noise). The sample equation of ARIMA (p,d,q) model is expressed as follows:*

$$\hat{Y}_t = \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

where $\hat{Y}_t$ is the predicted value at time t; α is the constant term; $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are the observed values at time $t-1, t-2, \ldots, t-p$, respectively; ϕ_j, $j = 1, 2, \ldots, p$ are the coefficients of AR terms to be determined; θ_k, $k = 1, 2, \ldots, q$ are the coefficients of MA terms to be determined; ε_t is the error term (white noise).

Definition 3.3. *The time series $\{Y_t, t \in Z\}$, with index set $Z = \{0, \pm 1, \pm 2, \ldots\}$ is said to be stationary if (i) $\mu_{x_t}^2 < \infty$ for all $t \in Z$, (ii) $\mu_{x_t} = m$, (m = constant) for all $t \in Z$, and (iii) the autocovariance function $\gamma_x(r, s) = \gamma_x(r + t, s + t)$, for all $r, s, t \in Z$.*

Remark 3.1. *If $\{Y_t, t \in Z\}$ is stationary, then $\gamma_x(r, s) = \gamma_x(r - s, 0)$ for all $r, s \in Z$. It is, therefore, convenient to redefine the autocovariance function of a stationary process as the function of just one variable as follows:*

$$\gamma_x(h) \equiv \gamma_x(h, 0) = Cov(Y_{t+h}, Y_t) \text{ for all } t, h \in Z \tag{4}$$

Example 3.1. [12] *Let $Y_t = A\cos(\theta t) + B\sin(\theta t)$ where A and B are two uncorrelated random variables with zero means and unit variance with $\theta \in [-\pi, \pi]$. Then*

$$\begin{aligned}
\gamma_x(t+h, t) &= Cov(Y_{t+h}, Y_t) \\
&= Cov(A\cos(\theta(t+h)) + B\sin(\theta(t+h)), A\cos(\theta t) + B\sin(\theta t)) \\
&= \cos(\theta(t+h))\cos(\theta t)Cov(A, A) + \cos(\theta(t+h))\sin(\theta t)Cov(A, B) \\
&\quad + \sin(\theta(t+h))\cos(\theta t)Cov(B, A) + \sin(\theta(t+h))\sin(\theta t)Cov(B, B) \\
&= \cos(\theta(t+h))\cos(\theta t)\,Var(A) + \sin(\theta(t+h))\sin(\theta t)\,Var(B) \\
&= \cos(\theta(t+h))\cos(\theta t) + \sin(\theta(t+h))\sin(\theta t) \\
&= \cos(\theta h),
\end{aligned}$$

which is independent of t. Therefore, this time series is stationary.
Since A and B are two uncorrelated random variables, then $Cov(A, B) = 0$.

Definition 3.4. *The double exponential smoothing method is also known as Holt's linear exponential method [13], and has two smoothing constants α and γ, and three equations:*

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{5}$$

$$b_t = \gamma(l_t - l_{t-1}) + (1 - \gamma)b_{t-1} \tag{6}$$

$$\hat{y}_{t+n} = l_t + nb_t \tag{7}$$

where l_t and b_t are the estimate of level and growth of the observed data series at time t, y_t is the observed data, respectively, α and γ are smoothing constants.

Definition 3.5. *The root mean square error (RMSE) is the square root of mean square error to measure the goodness of fit of the regression model. It measures the differences between values predicted by the created mathematical model and observed values. It is calculated as Equations (8) and (9).*

$$MSE = \sum_{t=1}^{n} \frac{\left(Y_t - \overline{Y}_t\right)^2}{n} \tag{8}$$

$$RMSE = \sqrt{MSE} \tag{9}$$

where Y_t and $\overline{Y}_t$ are the actual values and forecasted values, respectively.

4. **Results and Discussion.** This study examines the effectiveness of calving time forecasting in dairy cattle. The data in this study are hourly time series of the number of transitions from standing-to-lying and lying-to-standing over the period of 72 hours before calving time. These data are from total 25 cows of Oita branch and collected by using the videos taken by camera 360. The collected data are as shown in Table 1 with 4 hours interval.

Table 1. Transition from standing-to-lying and lying-to-standing of cows' data

Before calving time (h)	**Number of standing-to-lying and lying-to-standing transitions**							
	Cow 1	**Cow 2**	**Cow 3**	**Cow 4**	⋯	**Cow 23**	**Cow 24**	**Cow 25**
−72	15	1	7	2		13	5	4
−68	4	3	6	2		3	5	1
−64	6	6	5	2		5	3	5
−60	7	2	1	0		7	1	1
−56	1	5	7	2		0	2	3
−52	5	3	0	0		4	5	3
−48	11	2	5	1		11	4	3
−44	4	4	4	1		3	3	3
−40	6	5	5	0		7	4	7
−36	8	1	2	0	⋯	7	4	2
−32	1	1	4	0		2	0	3
−28	9	5	0	2		7	6	3
−24	12	2	2	0		12	4	4
−20	8	4	2	0		5	4	1
−16	4	10	2	1		2	1	6
−12	4	2	0	1		8	3	2
−8	12	7	0	2		11	4	8
−4	22	13	7	3		21	20	14

The exponential probability is useful when the observed data have significant growth per hour. Clearly, the closer the time is, the higher the chance of calving happens as shown in Table 1. This signifies that the cow will be in labor at the next time interval. Moreover, this probability method is beneficial when the observed data with uncertain time span prior parturition is received. On that occasion, the threshold parameter values from training data must be used to predict about calving time.

In building the *ARIMA* model, according to the Box-Jenkin principle, after verification of the stationary of the series as shown in Example 3.1, identifying the order of *AR* terms and *MA* terms is carried out. In determining the values of p and q, *ARIMA* (p,d,q) model gives minimum values error of Akiake information criterion (*AICc*). The result is presented in Table 3. Autocorrelation function (*ACF*) and partial autocorrelation function (*PACF*) were used to identify the model. In this stage, the best model is as simple as possible and minimizes the certain criteria *AICc* [14].

Moreover, Table 3 summarizes the values of the different models' parameters calculated by the method of moments in *R* and proves the selection of the model on which we will base our predictions. The chosen model is that of *ARIMA* (2,0,1). For the other models, one of the values of the minimization criteria is bigger than that obtained for the *ARIMA* model (2,0,1) with the constant value.

Furthermore, *ARIMA* (2,0,1) model without constant has larger *AICc* values than with constant value. The chosen parameters are grouped in Table 2. The developed model is

Table 2. Developed *ARIMA* model's parameter

Transitions from standing-to-lying and lying-to-standing of cow series			**Estimate**	**Standard error**
Constant			0.7	0.024
AR	Lag 2	ϕ_1	0.2	0.047
		ϕ_2	−0.1	0.0125
MA	Lag 1	θ_1	0.2	0.144

TABLE 3. Coefficient of different *ARIMA* model

Characteristics \ Models		***ARIMA* (1,0,0)**	***ARIMA* (0,0,1)**	***ARIMA* (1,0,1)**	***ARIMA* (2,0,1)**	***ARIMA* (2,0,2)**	***ARIMA* (2,0,1) without constant**
***AR* (1)**	ϕ_1	0.1322	–	−0.4465	0.2027	−1.0826	0.2749
***MA* (1)**	θ_1	–	0.2146	0.6646	0.2122	1.4496	0.2547
***AR* (2)**	ϕ_2	–	–	–	−0.1214	−0.8088	0.0419
Constant	α	1.548	1.1159	0.8985	0.7021	0.1909	–
AICc		172.53	171.89	172.13	**169.25**	178.2	173.54

given by Equation (10),

$$\overline{Y}_t = 0.7 + 0.2Y_{t-1} - 0.1Y_{t-2} + \varepsilon_t - 0.2\varepsilon_{t-1} \tag{10}$$

where $\overline{Y}_t$ is the forecasted standing/lying data at period t, Y_{t-1} and Y_{t-2} are the observed standing-to-lying and lying-to-standing transitions of periods $t-1$, and $t-2$, respectively. ε_t and ε_{t-1} are the residuals of periods t and $t-1$, respectively.

In addition, double exponential smoothing (*DES*) model is also used because the observed data of standing/lying transition is significantly in the shape of trend pattern. For this case, firstly, we will calculate initial value of level and trend line, l_0 and b_0 by fitting least-square regression in EXCEL to half of observed data. And substituting those values in Equations (5) and (6), continue the estimation. There are two parameters γ and α that are also determined by minimizing the *RMSE* values.

Moreover, whether the cow gives birth or not, the sum of mean standard deviation lines would be able to forecast the estimated time interval. If the forecast of the data passes through that line, the cow will be calving at the upcoming time interval. That will help us to take the right decisions related to the calving time of dairy cows. The comparison of estimation results of *ARIMA* (2,0,1) model and *DES* model is provided in Figure 2. It shows that the survival probability of expected calving time is near to give birth when the *DES* model passes through the second standard deviation line; however,

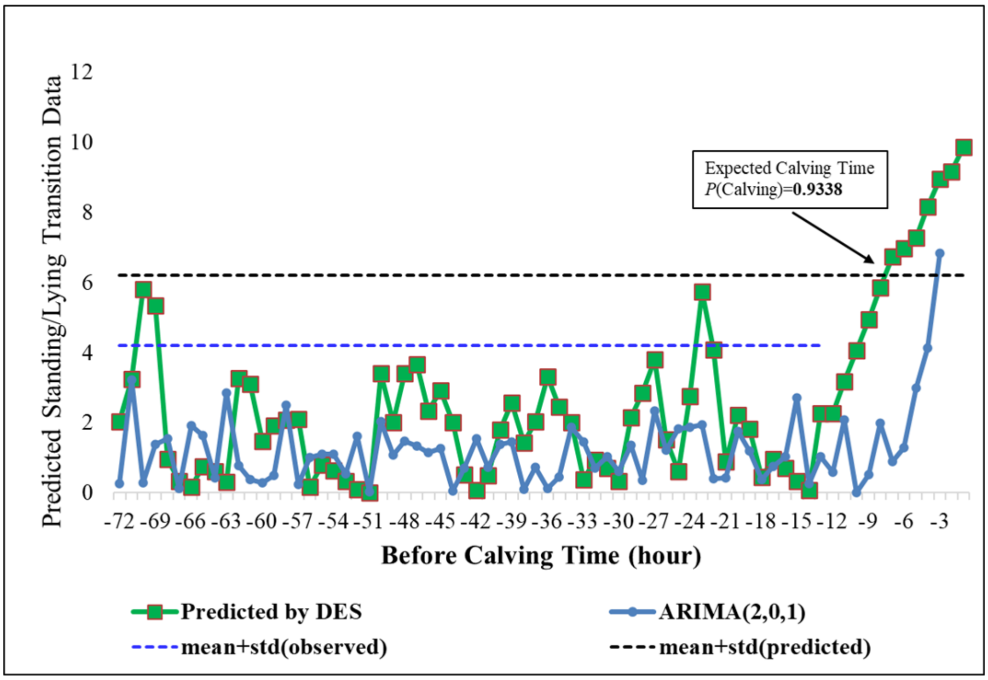

FIGURE 2. Predicted values by *ARIMA* (2,0,1) model and *DES* model

Table 4. Accuracy and error results of *ARIMA* model and *DES* model

Methods	**Testing accuracy**	**RMSE**
ARIMA (2,0,1)	86%	1.01
Double exponential smoothing (*DES*)	86%	1.14

it is 6 hours ahead by real calving time. In this case, *ARIMA* (2,0,1) model's prediction result is nearer with real calving time. On the other hand, the developed *ARIMA* model gives least *RMSE* error values of forecast than *DES* model as indicated in Table 4.

5. **Conclusions.** The *ARIMA* model that is selected and which minimizes the previous criteria is *ARIMA* (2,0,1). The analysis results might be summarized as follows: (i) we can estimate the calving time with the exact time interval by using the exponential distribution probability, and (ii) in *ARIMA* and *DES* modeling, the developed models can be used for modeling and forecasting the calving time of dairy cows. During the last 24-hour interval, the hourly trend of standing-lying transitions appeared to be increasing steadily. Furthermore, we can also hope to know the cows' behavioral changes are whether normal or not. Moreover, by using more dataset and finding seasonal pattern, the model can be improved by adding seasonal component. Therefore, this system can help to protect calf loss rate during calving and also it is effective in predicting calving time of dairy cows.

REFERENCES

[1] L. R. Sprott, *Recognizing and Handling Calving Problems*, Texas A&M AgriLife Extension, 2022.

[2] J. F. Mee, Newborn dairy calf management, *Veterinary Clinic of North America, Food Animal Practice*, vol.24, pp.1-17, 2008.

[3] M. Saint-Dizier and S. Chastant-Maillard, Methods and on farm devices to predict calving time in cattle, *Vet. Journal*, vol.205, pp.349-356, 2015.

[4] M. A. G. Von Keyserlingk and D. M. Weary, Maternal behavior in cattle, *Hormones and Behavior*, vol.52, no.1, pp.106-113, 2007.

[5] T. C. Lwin, T. T. Zin and P. Tin, Predicting calving time of dairy cows by exponential smoothing model, *Proc. of the 2020 IEEE 9th Global Conference on Consumer Electronics*, Kobe, Japan, pp.360-361, 2020.

[6] C. B. Bozarth and R. B. Handfield, *Introduction to Operations and Supply Chain Management*, 4th Edition, Pearson, Raleign, 2016.

[7] M. Mitsutaka and I. Akira, Examination of demand forecasting by time series analysis for auto parts remanufacturing, *J. Remanuf.*, vol.5, DOI: 10.1186/s13243-015-0010-y, 2015.

[8] T. T. Zin, K. Sumi, P. T. Seint, P. Tin, I. Kobayashi and Y. Horii, A stochastic modeling procedure for predicting the time of calving in cattle, *ICIC Express Letters, Part B: Applications*, vol.13, no.1, pp.49-56, 2022.

[9] M. R. Borchers et al., Machine-learning-based calving prediction from activity, lying and rumination behaviors in dairy cattle, *Journal of Dairy Science*, vol.100, no.7, pp.5664-5674, 2017.

[10] A. Gupta, W.-B. Zeng and Y. Wu, *Probability and Statistical Model: Foundation for Problems in Reality and Financial Mathematics*, DOI: 10.1007/978-0-8176-4987-6, 2010.

[11] S. Saengwong, C. Jatuporn and S. W. Roan, An analysis of Taiwanese livestock prices: Empirical time series approaches, *Journal of Animal and Veterinary Advances*, vol.11, no.23, pp.4340-4346, 2012.

[12] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd Edition, Springer, 1991.

[13] R. J. Hyndman, A. B. Koehler, J. K. Ord and R. D. Snyder, *Forecasting with Exponential Smoothing. The State Space Approach*, Springer, 2008.

[14] G. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*, 3rd Edition, Holden-Day, San Francisco, 1994.