DESCRIBING TEMPORAL VARIATION IN RETICULORUMINAL PH USING CONTINUOUS MONITORING DATA

Journal:	*Journal of Dairy Science*
Manuscript ID	JDS-17-12828.R2
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Denwood, Matthew; University of Copenhagen, Department of Veterinary and Animal Sciences Kleen, Joachim; CowConsult Jensen, Dan; University of Copenhagen, Department of Veterinary and Animal Sciences Jonsson, Nicholas; University of Glasgow, Institute of Biodiversity Animal Health and Comparative Medicine; Harbro Ltd
Key Words:	reticuloruminal pH, acidosis, remote sensing data, statistical model

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

skewed_sine.R

[INTERPRETIVE SUMMARY]

Monitoring the reticuloruminal pH of dairy cattle using bolus devices is becoming increasingly affordable, but making sense of the resulting mass of data is challenging. "Describing temporal variation in reticuloruminal pH using continuous monitoring data" by Denwood et al. demonstrates statistical models that summarize these data into a daily evaluation of the functional state of the rumen. These summaries are able to predict a future reduction in milk yield and feed intake in apparently healthy dairy cattle. Development of automated monitoring systems based on this work could have a large impact on milk production in industrialized settings.

[RUNNING TITLE]

TEMPORAL VARIATION IN RETICULORUMINAL PH

[TITLE AND AFFILIATIONS]

DESCRIBING TEMPORAL VARIATION IN RETICULORUMINAL PH USING CONTINUOUS MONITORING DATA

Denwood, M. J.[*1], Kleen, J. L.[†], Jensen, D.B.[*], Jonsson, N. N.[‡ #]

[*] Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

[†] CowConsult, Coldinne, Germany

[‡] Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, UK

[#] Harbro Ltd, Birkhill Mill, Lanarkshire, UK

[1] Matthew James Denwood, Grønnegårdsvej 8, 1870 Frederiksberg C, Denmark

Phone: +45 35 33 23 07, Email: *md@sund.ku.dk*

ABSTRACT

Reticuloruminal pH has been linked to subclinical disease in dairy cattle, leading to considerable interest in identifying pH observations below a given threshold. The relatively recent availability of continuously monitored data from pH boluses gives new opportunities for characterizing the normal patterns of pH over time, and distinguishing these from abnormal patterns using more sensitive and specific methods than simple thresholds. We fit a series of statistical models to continuously monitored data from 93 animals on 13 farms in order to characterize normal variation within and between animals. We used a subset of the data to relate deviations from the normal pattern to the productivity of 24 dairy cows from a single herd. Our findings show substantial variation in pH characteristics between animals, although animals within the same farm tended to show more consistent patterns. There was strong evidence for a predictable diurnal variation in all animals, and up to 70% of the observed variation in pH could be explained using a simple statistical model. For the 24 animals with available production information, there was also a strong association between productivity (as measured by both milk yield and dry matter intake) and deviations from the expected diurnal pattern of pH 2 days prior to the productivity observation. In contrast, there was no association between productivity and the occurrence of observations below a threshold pH. We conclude that statistical models can be used to account for a substantial proportion of the observed variability in pH, and that future work with continuously monitored pH data should focus on deviations from a predictable pattern rather than the frequency of observations below an arbitrary pH threshold.

Keywords: reticuloruminal pH, acidosis, remote sensing data, statistical model

INTRODUCTION

Assessment of reticuloruminal pH in cattle has been used as an indicator of an excessive intake of soluble carbohydrate or a shortage of physically effective fiber, and the consequent predisposition to a range of health problems including ruminitis, liver abscess, left displaced abomasum, diarrhea, laminitis and poor milk production (Dirksen and Smith, 1987; Nordlund and Garrett, 1997; Garrett et al., 1999; Plaizier et al., 2008). Evaluation of reticuloruminal pH is therefore of interest, with most clinicians using a single observation to define the status (Nordlund et al., 1995; Nordlund and Garrett, 1997). It is now possible to continuously monitor pH values using remote sensing data, but this generates a large volume of data that can be very challenging to interpret. As a result, researchers using continuous pH monitoring techniques to investigate diet and acidosis have most often used average values or threshold approaches for evaluating reticuloruminal pH. For example, Khafipour et al. (2011) used the overall mean pH for an individual animal over a period of time and the number of minutes that the pH was seen to be below 5.6. We aim to define and diagnose abnormal reticuloruminal pH by identifying deviations from normal pH patterns, rather than the number of observations below a given threshold. In order to achieve this goal, we first require the range of normal pH conditions to be described and quantified.

Statistical methods can be used to help quantify and explain some of the observed variability in pH over time, and may also be able to identify periods of the pH time series that are not broadly consistent with predictions. These discrepancies can be quantified based on residual variation between the fitted model and observations. There are several statistical methods that can be used to deal with time series data with temporal patterns, including recurring temporal patterns with known frequency, such as predictable daily fluctuations. A descriptive method (such as a linear decomposition or smoothed average) can be used in order to separate the predictable temporal variation from the longer-term patterns in the data. For example, Maatje et al. (1992) modeled the daily milk yield of dairy cows, which is known to follow a characteristic curve over time (Wood, 1967; Cole et al., 2009), using a simple moving average where significant deviations between the observed milk yield and the yield predicted by the moving average were considered indicative of mastitis. More recently, Ostersen et al. (2010) implemented an oestrus detection model based on the diurnal frequency of individual sows visiting a boar, by dividing the 24 hours of the day into 4 periods of 6 hours and then assigning an expected average frequency per period.

Another method is to fit a model that includes sinusoidal effects (with known frequency and varying amplitude and phase shift) in order to account for and analyze the recurring patterns. Fitting increasing numbers of these sine waves with different frequencies allows any temporally recurring pattern to be described using the sum of its harmonics, although at the cost of a potentially large number of degrees of freedom. For example, Madsen et al. (2005) modeled the daily drinking pattern of weaned pigs using between 1 and 12 harmonic waves implemented within a dynamic linear model (DLM). The same authors have also shown that significant deviations between this model and the observed drinking pattern correlate with undesirable events such as diarrhea and herd management problems (Madsen and Kristensen, 2005). Similarly, Jensen et al. (2017) modeled the drinking behavior of slaughter pigs using the sum of 3 harmonic waves implemented in a DLM, with the purpose of detecting diarrhea or pen fouling.

A third option is to fit 1 or more sine waves at frequencies corresponding to known (or suspected) biologically relevant influences on the parameter of interest. For example, a sine wave with a frequency corresponding to a daily cycle over 24 hours could be an appropriate way to deal with predictable daily variation due to a regular routine. We prefer this option as it allows the most intuitive interpretation in terms of biologically relevant factors, and allows for extrapolation to other herds and animals.

In this paper, we make use of continuously monitored reticuloruminal pH values obtained from clinically normal dairy cattle. We first quantify the variation in pH observed within and between animals and farms using standard linear models, and then more closely examine the subtle differences in pH patterns using more complex non-linear models applied to the data from each animal separately. The remaining residual variation is then used to identify periods of abnormality that can be related to decreased production traits even in these clinically normal cattle. The overall aim is to produce recommendations and statistical methods for distinguishing 'normal' from 'abnormal' when dealing with long-term continuously monitored pH data.

MATERIALS AND METHODS

Rumen Bolus Data

Remotely sensed pH data was obtained from 13 different dairy farms in Germany, with a total of 93 animals. Feeding management was consistent within each herd over the course of the data collection, and there were no reports or suspicions of either clinical disease or abnormal rumen function in any of the animals. The dairy farms were divided into those milking 3 times per day (group A: 8 farms with a total of 45 animals), and those milking twice per day (group B: 5 farms with a total of 48 animals). The 93 data series (45 from group A and 48 from group B) were first analyzed together, in order to quantify the importance of the variation between herds and animals, and then separately to allow the difference in pH patterns between individuals to be explored in more detail. An overview of the complete procedure is given in Figure 1.

Wireless, indwelling rumen pH and temperature-monitoring boluses with a measurement interval of 600 seconds were used in all animals (smaXtec animal care GmbH, Graz, Austria). The pH probes of the boluses were all calibrated in pH 4 and pH 7 buffer solutions before administration to the animals *per os*. Individual pH records were graphically assessed to remove any readings that were not compatible with rumen conditions (pH ≥ 10 or constant pH) from the beginning and end of the time series. Additionally, the first 12 hours and the last hour were removed from all data series to allow calibration time and unreliable readings from failing boluses, respectively. All data series were limited to a maximum of 50 days of observation, as per the manufacturer's guarantee.

Summary Statistics

Crude summary statistics of mean, variance, 1st percentile and 99th percentile were calculated for the pH data series from each animal. The effect of farm on the 93 recorded values of these 4 within-animal summary statistics was quantified using 4 separate linear models in R (R Core Team, 2016), with farm as a fixed effect.

Statistical Methods for Combined Data

The statistical models used to analyze the data were based on sinusoidal functions. The standard linear formulation of a sine wave is given by:

$$A\sin(\theta+\theta_0)=\alpha\sin\theta+\beta\cos\theta$$

Where θ is a known parameter reflecting the time of day that is transformed to the interval $[0, 2\pi]$, A is the amplitude, and θ_0 is the phase shift. The 2 terms α and β are linear parameters on the interval $(-\infty, \infty)$ that are typically estimated, and are related to A and θ_0 by:

$$A = \sqrt{\alpha^2 + \beta^2} \qquad \theta_0 = \operatorname{atan2}(\beta, \alpha)$$

Where atan2 is the 2-argument arc-tangent.

A fixed frequency of 1 cycle per day was used for the first sine wave to describe diurnal variation. In addition, a second sine wave with a fixed frequency corresponding to 2 or 3 cycles per day (but with equal spacing between the times of milking events rather than equal spacing throughout the day) may be justified in order to explain the influence of farm management factors associated with the frequency of daily milking. The θ parameter was therefore derived from the time of day in order to correspond to 2 different hypothesized cyclical events: diurnal variation on a 24-hour cycle, and cyclical patterns corresponding to each milking event. The total within-day variation was modeled as the sum of up to 2 separate sine waves that were simultaneously fit using periods of (a) 24 hours, and (b) 2 or 3 daily milking events (depending on the on-farm milking frequency). The daily milking events were timed to correspond to unequally spaced milking times, which were approximated for all farms as 04:30 & 15:30 h for twice daily, and 04:30, 13:00 & 20:00 h for those with 3 daily milking times.

In addition to the cyclical temporal patterns discussed above, there may be longer-term temporal trends that can affect the average pH on this longer time scale. Therefore, before fitting the linear model, a generalized additive model (GAM) was used to account for the longer-term changes in pH between days within each individual animal. A separate GAM was fitted to the pH data series from each animal using the mgcv package (Wood, 2011) for R (R Core Team, 2016), and the estimated spline functions were extracted and used to generate offset covariates for the linear model for each animal.

The complete set of observed pH data (616,945 observations from 93 animals) was then fit to the linear model with 2 sinusoidal functions, using offset covariates obtained from the GAM. Each model incorporated linear parameters of α_1 & β_1 for the daily sine wave, α_2 & β_2 for the milking-frequency sine wave, and fixed effects of farm and animal. The equation for the main effects of this model is given below:

$$pH = \left[c + m_f + n_a + \alpha_1 \sin\theta_D + \beta_1 \cos\theta_D + \alpha_2 \sin\theta_M + \beta_2 \cos\theta_M\right] + \left[s_a(t)\right]$$

Where c is the intercept, m_f is the main effect associated with the relevant farm, n_a is the main effect associated with the relevant animal, θ_D represents the time of day, θ_M represents the time elapsed since the last milking event, s_a is a smoothed spline function for animal a, and t is the time elapsed since the start of the pH series for that animal (date and time). The 2 sets of square brackets denote the elements of the model fitted using the linear model (left) and GAM (right). The models were fitted using R (R Core Team, 2016), including all 4 interactions between m_f and the α_1, β_1, α_2 & β_2 parameters, as well as all 4 interactions between n_a and the α_1, β_1, α_2 & β_2 parameters.

Statistical Methods for Individual Animal Data

A more complex modeling procedure was then used in order to further examine the nature of any differences in pH patterns between animals. These models were fitted to the data series from each animal separately, so that the estimated patterns obtained from each animal were not influenced by the data from other animals.

The standard sine wave model was extended to allow the value of the 2 rates of change to vary, so that the increase in pH from the minimum to maximum value might occur more quickly than the decrease in pH from the maximum to minimum value. Such extensions to the standard sine wave are not possible using a linear model system, but can be made using a non-linear model system. A non-linear periodic function was therefore written to extend the standard sine wave model with an additional parameter to allow the shape of the wave to be skewed so that the minimum trough points could be moved relative to their standard equidistant position between maximum peaks. This mechanism allows a skew parameter, γ, to control the rate of change for increasing and decreasing pH phases while still retaining the amplitude, periodicity and phase of the standard sine wave model. This parameter is also defined on the interval $(-\infty, \infty)$ to facilitate estimation, although it is subsequently transformed to the interval (-0.5, 0.5) using a scaled and re-centered logistic transformation. The final skewed sine wave is therefore a non-linear function of θ (which is derived from the time of day, as discussed below) and the 3 parameters α, β & γ. The R code that implements this non-linear function is given as an online supplement, and examples of the output of this function for a small number of possible values for γ are shown in Figure 2. Note that the standard sine wave is a special case of this function with skew parameter $\gamma = 0$, which was also used as a baseline model to test the improvement in model fit achieved by allowing γ to vary. The 3 parameters α, β & γ of the non-linear function were optimized simultaneously using the nlsLM function in the minpack.lm package (Elzhov et al., 2013) for R (R Core Team, 2016).

As with the simpler models described above, 2 separate sine waves with frequencies of 1 oscillation per 24 hours and either 2 or 3 oscillations per 24 hours were used to explain diurnal variation and the influence of farm management factors associated with the frequency of daily milking. A GAM was also used to account for longer-term temporal trends within each animal as before, although the GAM and non-linear components were fitted iteratively (each using the estimates from the other model as fixed covariates) for this more complex model, to allow the combined likelihood to be maximized. The complete model for each animal is represented by the equation below:

$$pH = [c + \mathrm{f}(\theta_D, \alpha_1, \beta_1, \gamma_1) + \mathrm{f}(\theta_M, \alpha_2, \beta_2, \gamma_2)] + [s(t)]$$

Where the function f is the skewed sine function, s is a smoothed spline function, θ_D represents the time of day, θ_M represents the time elapsed since the last milking event, and t is the time elapsed since the start of the pH series for that animal (date and time). The 2 sets of square brackets denote the non-linear (left) and GAM (right) elements of the model.

Model fit for the more complex models was compared in 2 distinct ways. Firstly, Akaike's Information Criterion (AIC; Akaike, 1973) was compared between the models, with the total number of degrees of freedom equal to the number of parameters in the non-linear model plus the effective number of parameters used by the GAM. The motivation for this was to assess the improvement in statistical fit obtained from increasing the complexity of the model for each animal. Secondly, the incremental reduction in residual variance was evaluated from each successively more complex model. The motivation for this was to assess the biological impact of the more complex relationships in the model in terms of the variance explained. To simplify model comparisons, the skew parameter was also included in the milking-related sine wave function if including the same parameter in the daily sine wave function improved the AIC of the model. An additional comparison was made between the 4 different possibilities for milking-related sine waves (twice and 3 times daily, each spaced equally or corresponding to approximate milking times), which had equal degrees of freedom permitting a direct comparison of the maximized likelihoods.

Identifying Abnormalities in pH

The magnitude of the residuals (difference between the pH values predicted by a fitted model and those observed from the data) can be used to identify periods during which abnormalities have occurred. Residuals can be calculated as either signed or absolute, with the former taking into account the difference between positive residuals (underestimates) and negative residuals (overestimates), and the latter removing the sign so that only the magnitude is recorded. We summarized these residuals over 24-hour periods, so that the degree of predictability or abnormality was quantified for each animal and observed day. Rather than basing the predictions on the entire data series, the model from which predictions were made was sequentially re-fit to a subset of the pH data for each observed day. This data subset excluded the 3-day period of observations from the day before to the day following that for which predictions and residuals were to be calculated, thus avoiding any potential bias from including the same data in calculating both the model fit and residuals. Each animal was modeled separately using the best-fitting model for the separate data, as described above.

Residuals from each separately fitted model were calculated for the relevant day, and the results combined into a single series of residuals for each observation per animal. These were subsequently summarized for each 24-hour time period into (1) the mean of the residuals for the day and (2) the mean of the absolute residuals for the day, which respectively represent (1) a period of unexpected sustained acidosis or alkalosis and (2) a period of unpredictability in pH (regardless of direction). Summary statistics based on the raw data were also used, representing (3) the number of pH observations for that day that were greater than an animal-specific upper threshold and (4) the number of pH observations that day that were lower than an animal-specific lower threshold. These thresholds were taken from the 1st and 99th percentile (i.e. 1% most acidic and 1% most alkaline) over the entire observation period for that animal, as previously calculated. These summary statistics were intended to be closer in concept to the traditional pH threshold for declaring acidosis, with the exception of using an animal-dependent threshold.

Relating Abnormalities to Production Characteristics

To demonstrate the ability of our statistical methods to identify abnormalities in pH series, we related the 4 daily summary statistics given above to contemporary daily measurements of dry matter intake and milk yield available for the 24 animals in herd B1. The Ali-Schaeffer curve (Ali and Schaeffer, 1987; Buttchereit et al., 2010) was used to standardize the observed daily milk yields by days since the start of the lactation, in order to simplify the comparisons. Multivariate linear mixed effects regression models were used to model the daily productivity of all 24 cows, using the cow ID as a random effect, and fixed effects relating to the 4 summary statistics of pH measurements and model predictions, as given above. Separate models were fitted to the 2 productivity parameters of dry matter intake recordings and standardized milk yields, using identical predictor variables. The data were aligned so that summary statistics for pH measurements and model predictions were used from 2 days before the corresponding productivity observations, in order to maximize the potential predictive ability of the model by using the smallest time lag that would be useful in practice. All linear mixed effects regression models were fitted using the lme4 package (Bates et al., 2014) for R (R Core Team, 2016).

RESULTS

Summary Statistics

The number and length of pH series obtained are shown in Table 1, and the mean and variance of individual animal pH series are shown in Figure 3. There is a substantial variation in the individual animal characteristics of pH, although there was also a significant effect of farm on the mean ($p < 0.001$, adjusted $R^2 = 0.32$), variance ($p < 0.001$, adjusted $R^2 = 0.44$), 1st percentile ($p=0.002$, adjusted $R^2 = 0.20$) and 99th percentile ($p < 0.001$, adjusted $R^2 = 0.37$) of the observed pH measurements for each animal.

Combined Data Results

Modeling results from the combined dataset are shown in Table 2. The mean squares estimates indicate that both daily and milking-frequency sine waves are important components of pH variation, and that mean pH varies dramatically between farms (and to a lesser extent between animals on the same farm). The results also suggest that the precise pattern of pH variation varies to some extent between farms, but appears to be more consistent between animals on the same farm. F-tests indicated that all fixed effect terms and interactions were significant ($p < 0.001$), although these p-values should be interpreted with caution due to the non-independence of residuals in the fitted model.

Individual Animal Data Results

For the purposes of illustration, the 3 individual components of the more complex statistical model corresponding to the GAM, daily-frequency sine wave and milking-frequency sine wave (as estimated from the data for 2 representative animals) can be seen in Figure 4. The proportion of the variance explained by the GAM component was between approximately 0% and 90%, depending on the animal (Figure 5a). Adding a sine wave to account for daily cyclical variability accounted for between 10% and 70% of the remaining variance, whereas adding a skew parameter γ and a second sine wave corresponding to known milking times generally accounted for less than 10% of the remaining variance (Figure 5b-d). One interesting exception is that the milking-related sine wave accounted for between 7% and 35% of the remaining variance for all 24 animals on farm B1. A comparison of model fit using AIC indicated that the addition of the skew parameter γ was supported for 82 of the 93 animals (Table 1), and the addition of the second sine wave was supported for 92 of the 93 animals. Models including the correct milking frequency were preferred for 61 of the 93 (66%) animals, and models with the second sine wave adjusted to match the interval between milking events were preferred to models with regularly spaced time intervals for 72 out of the 93 (77%) animals.

The daily pattern corresponding to the maximized likelihood of the models including a single sine wave with skew parameter (for ease of comparison between farms) is shown in Figure 6. The skew parameter (allowing different rates of increasing vs. decreasing pH) varied between data series, with 46 showing a substantially faster increasing vs. decreasing pH, 29 showing a substantially faster decreasing vs. increasing pH, and the remaining 102 data series estimated as close to a standard sine wave. The amplitude was estimated between 0.1 and 0.2 pH units for most animals. The plotted curves show a high degree of consistency for all animals in terms of the period of peak pH around 06:00 – 08:00 h, with a corresponding trough at around 18:00 h (Figure 6). The patterns appear to be more consistent within farm than between farms, for example farm A6 shows a very similar signature for all animals that is almost unique to that farm. However, farm B1 shows 2 apparent sub-groups of animals with patterns resembling those from farms B4 and A6, respectively, and there is also an individual pattern that does not closely resemble the majority at farms B3, A3, & A4. This suggests that despite the general consistency within a farm, individual-cow factors can result in unusual patterns in some animals relative to their herd mates. There is also a large amount of variation in the estimated mean pH values among animals, even within the same herd.

Relating pH Abnormalities to Production Characteristics

A total of 1,044 days of pH observations and model residuals were summarized for the 24 animals in herd B1. The patterns of mean residual and mean absolute residual indicated some similarities due to the correlation between the measurements (Supplemental Figure S1). Most animals showed little qualitative evidence of strongly 'abnormal' events, but deviations could be detected visually, for example towards the end of the series in animals 5, 6 and 11. These 3 events seem to be concurrent with abnormally high numbers of extreme pH observations (Supplemental Figure S2), although the latter plot is more difficult to interpret.

There was a total of 1,043 observations of daily dry matter intake and 970 observations of daily milk yield (after correcting for stage of lactation) that could be related to the pH observations in the 24 animals from herd B1. Results of the 2 multivariable models show that an increasing daily mean absolute residual had a significant ($p < 0.001$) and strongly negative effect on both production characteristics (Table 3). In contrast, neither the daily mean residual nor the number of extreme pH observations (either high or low pH) had a significant effect on either milk yield or dry matter intake. This indicates that deviations from a predictable daily rhythm (as represented by the mean absolute residual) are associated with decreased productivity, whereas neither changes in the daily average pH (as represented by the daily mean residual) nor extreme pH events are directly associated with productivity.

DISCUSSION

This manuscript gives an overview of the characteristics of 93 continuously monitored pH data series obtained from animals from 13 different farms. No clinical or subclinical pathology (including digestive alterations) had been reported in these animals, so the results can be considered as a baseline of normal activity over a range of animals, although the possibility of undetected sub-clinical disease in some animals cannot be excluded. The importance of the observed daily fluctuation is striking, indicating that a consistent and substantial difference in pH should be expected between samples taken in the evening vs. early morning. There were also stronger similarities between animals from the same farm than from different farms, presumably resulting from a range of dietary and management factors affecting the daily pH cycle of the rumen. However, there were also substantial differences between animals, particularly in terms of the average pH observed during the study, and the proportion of the overall variance that could be attributed to long-term predictable changes in pH and short-term cyclical behavior. There was also potentially biologically interesting variation in the peak difference in pH (amplitude) between animals, and particularly between farms. There was also substantial longer-term temporal variation in some animals, which may reflect either gradual changes in pH function (due to e.g. dietary factors) or pH drift in the sensors, and necessitated the use of a GAM.

The model-fit comparison also allows some recommendations to be made in terms of modeling these data for other purposes. The addition of a sine wave to explain short-term variation in pH was strongly justified for all animals, and was able to account for a substantial portion of the observed variation in pH, suggesting that simple daily cyclic activity should be accounted for, particularly when comparing pH values observed at different times of the day. The addition of a second sine wave to account for management factors associated with the known frequency of milking was also supported, although this had less influence than the main sine wave. Results from the simpler linear model indicate relatively little variation in the parameters of these sine waves between animals on the same farm, which could justifiably be ignored in some situations. However, we recommend that differences in temporal variation between farms and inconsistent mean pH between animals on the same farm should always be taken into account. The addition of a skew parameter to allow pH to increase and decrease at different rates was also supported for the majority of animals, but had a much smaller role in explaining the variance and requires the use of more complex non-linear models. However, further research is warranted into the potential nutritional implications of this imbalance between the rates of increase and decrease in pH. Finally, the fact that the 'correct' frequency of milking was not consistently identified on most farms suggests that the improved fit of these models is partly due to the effects of fitting an additional harmonic frequency rather than a mechanistic effect related to milking. However, farm B1 consistently showed a substantially improved fit when including the correct frequency, and for the timings linked to the estimated milking times rather than equally spaced throughout the day, suggesting that the sine wave reflects a real process on this farm. This could indicate a high degree of regularity in the relevant factors associated with the daily schedule of this farm (such as feeding times), which is far more likely to be the true cause of regular pH patterns than the milking process itself. Note that the most relevant management factors need not occur at the same time as milking, as each sine wave allows the peak/trough time to occur at any point during the cycle for each animal independently; only the frequency is fixed.

The focus of this manuscript is on providing an insight into the variation of qualitatively 'normal' pH series between dairy cattle within and between farms. However, the ultimate goal of monitoring reticuloruminal pH is not to describe normality, but to detect situations where there is evidence for an abnormal pattern. Although they are not standard threshold pH values, the 1^{st} and 99^{th} percentiles for each animal were chosen to reflect the relatively unusual observations within pH series that did not contain observations generally considered to be extreme. There was no association between productivity and the frequency of these 'extreme' pH observations. However, using the mean absolute residual between the fitted model and observed data over a daily time period, we have demonstrated that it is possible to predict changes in the milk yield and dry matter intake of the cow, even when based on data from clinically normal cows. Further work is required to refine this system in order to optimize the sensitivity to changes in pH variability, and to develop models for real-time monitoring that are able to continuously adapt to the specific pH pattern of an individual animal based only on the previously observed values. A key goal of further research will be to investigate the predictive utility of reticuloruminal pH measurements based on past measurements, perhaps using a method such as a DLM (Madsen et al., 2005; Bono et al., 2012; Jensen et al., 2017). Using residuals from a DLM was previously shown to be useful for detecting specific production-relevant events such as mastitis in dairy cows (Jensen et al., 2016) and oestrus in group-housed sows (Ostersen et al., 2010), as well as non-specific events in animal production (Madsen and Kristensen, 2005; Cornou et al., 2008). Multivariate DLM also allows data series gathered from multiple different sources to be modeled simultaneously, and the covariance between the different variables to be taken into account when forecasting expected observations (Jensen et al., 2015, 2016). It is therefore likely that using both reticuloruminal pH and feed intake data with a combined model would result in improved performance in detecting undesirable events when compared to pH data alone.

CONCLUSIONS

Continuously monitored reticuloruminal pH data show a strong and predictable short-term pattern that is well described by a simple sine wave with a frequency of 1 cycle per day. More complex models, including an additional sine wave and a skew parameter, give incrementally smaller additional benefits, but can be used to demonstrate differences between animals. Deviations from the expected daily cyclical variation were significantly associated with reduced production, but there was no apparent effect of 'extreme' pH observations on productivity. We therefore conclude that future efforts to describe continuously monitored pH data should be based on deviations from an expected predictable rhythm rather than observed measurements below some arbitrarily defined pH threshold.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Sarah Layhe (Vaetta Editing) for assistance with proofreading and formatting of the manuscript, and the two anonymous reviewers for their excellent suggestions, which have greatly improved the manuscript. Dan Jensen was financially supported from The Danish Council for Strategic Research (The PigIT project, Grant number 11-116191).

REFERENCES

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. B.N. Petrov and F. Csaki, ed. Akademiai Kiado, Budapest.

Ali, T.E., and L.R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67:637–644. doi:10.4141/cjas87-067.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://cran.r-project.org/package=lme4.

Bono, C., C. Cornou, and A.R. Kristensen. 2012. Dynamic production monitoring in pig herds I: Modeling and monitoring litter size at herd and sow level. Livest. Sci. 149:289–300.

doi:10.1016/j.livsci.2012.07.023.

Buttchereit, N., E. Stamer, W. Junge, and G. Thaller. 2010. Evaluation of five lactation curve models fitted for fat:protein ratio of milk and daily energy balance. J. Dairy Sci. 93:1702-1712. doi:10.3168/jds.2009-2198.

Cole, J.B., D.J. Null, and P.M. Vanraden. 2009. Best prediction of yields for long lactations. J. Dairy Sci. 92:1796–810. doi:10.3168/jds.2007-0976.

Cornou, C., J. Vinther, and A.R. Kristensen. 2008. Automatic detection of oestrus and health disorders using data from electronic sow feeders. Livest. Sci. 118:262–271. doi:10.1016/j.livsci.2008.02.004.

Dirksen, G., and M.C. Smith. 1987. Acquisition and analysis of bovine rumen fluid. Bov. Pract.

Elzhov, T. V, K.M. Mullen, A.-N. Spiess, and B. Bolker. 2013. minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds.

Garrett, E.F., M.N. Pereira, K. V Nordlund, L.E. Armentano, W.J. Goodger, and G.R. Oetzel. 1999. Diagnostic methods for the detection of subacute ruminal acidosis in dairy cows. J. Dairy Sci. 82:1170–1178. doi:10.3168/jds.S0022-0302(99)75340-3.

Jensen, D.B., C. Cornou, N. Toft, and A.R. Kristensen. 2015. A multi-dimensional dynamic linear model for monitoring slaughter pig production. Pages 503–512 in 7th European Conference on Precision Livestock Farming.

Jensen, D.B., H. Hogeveen, and A. De Vries. 2016. Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. J. Dairy Sci. 99:1–18. doi:10.3168/jds.2015-10060.

Jensen, D.B., N. Toft, and A.R. Kristensen. 2017. A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs. Comput. Electron. Agric. 135:51–62. doi:10.1016/j.compag.2016.12.018.

Khafipour, E., J.C. Plaizier, P.C. Aikman, and D.O. Krause. 2011. Population structure of rumen Escherichia coli associated with subacute ruminal acidosis (SARA) in dairy cattle. J. Dairy Sci. 94:351–360. doi:10.3168/jds.2010-3435.

Maatje, K., P.J.M. Huijsmans, W. Rossing, and P.H. Hogewerf. 1992. The efficacy of in-line measurement of quarter milk electrical conductivity, milk yield and milk temperature for the detection of clinical and subclinical mastitis. Livest. Prod. Sci. 30:239–249. doi:10.1016/S0301-6226(06)80013-8.

Madsen, T.N., S. Andersen, and A.R. Kristensen. 2005. Modelling the drinking patterns of young pigs using a state space model. Comput. Electron. Agric. 48:39–62. doi:10.1016/j.compag.2005.01.001.

Madsen, T.N., and A.R. Kristensen. 2005. A model for monitoring the condition of young pigs by their drinking behaviour. Comput. Electron. Agric. 48:138–154. doi:10.1016/j.compag.2005.02.014.

Nordlund, K. V, and E.F. Garrett. 1997. Rumenocentesis: A Technique for Collecting Rumen Fluid for the Diagnosis of Subacute Rumen Acidosis in Dairy Herds. Proc. Univ. Sydney Post Grad. Found. Vet. Sci. 290:195–200.

Nordlund, K. V, E.F. Garrett, and G.R. Oetzel. 1995. Herd-based rumenocentesis: a clinical approach to the diagnosis of subacute rumen acidosis. Compend. Contin. Educ. Pract. Vet. 17.

Ostersen, T., C. Cornou, and A.R. Kristensen. 2010. Detecting oestrus by monitoring sows' visits to a boar. Comput. Electron. Agric. 74:51–58. doi:10.1016/j.compag.2010.06.003.

Plaizier, J.C., D.O. Krause, G.N. Gozho, and B.W. McBride. 2008. Subacute ruminal acidosis in dairy cows: The physiological causes, incidence and consequences. Vet. J. 176:21–31. doi:10.1016/j.tvjl.2007.12.016.

R Core Team. 2016. R: A Language and Environment for Statistical Computing.

Wood, P.D. 1967. Algebraic Model of the Lactation Curve in Cattle. Nature 216:164–165.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. 73:3–36.

Table 1. The number of pH series per farm, mean (range) number of pH observations per animal, mean (range) length of time represented within the time series, and the number of time series for which AIC demonstrated the best model fit for a simple sine wave (2 degrees of freedom), sine wave with skew parameter (3 degrees of freedom), double sine wave with a second wave describing the known milking events (4 degrees of freedom), and both double sine wave and skew parameters (6 degrees of freedom)

Farm	Animals	Observations	Days	Sine	Skew	Double	Both
A1	5	7,125 (7,120 - 7,129)	49.5 (49.5 - 49.5)	0	0	2	3
A2	2	4,520 (1,912 - 7,129)	31.4 (13.3 - 49.5)	0	0	0	2
A3	6	6,812 (6,528 - 7,099)	47.5 (45.5 - 49.5)	0	0	1	5
A4	11	6,403 (4,829 - 7,129)	44.5 (33.5 - 49.5)	0	0	1	10
A5	3	7,107 (7,081 - 7,121)	49.5 (49.5 - 49.5)	0	0	1	2
A6	5	5,267 (4,743 - 5,460)	36.6 (33.0 - 37.9)	0	0	0	5
A7	6	6,929 (6,522 - 7,129)	48.1 (45.3 - 49.5)	0	0	0	6
A8	7	7,113 (7,026 - 7,129)	49.4 (48.8 - 49.5)	0	0	1	6
B1	24	6,598 (6,595 - 6,600)	45.8 (45.8 - 45.8)	0	0	0	24
B2	6	6,273 (2,046 - 7,129)	43.6 (14.3 - 49.5)	0	0	2	4
B3	4	7,044 (6,792 - 7,129)	48.9 (47.2 - 49.5)	0	0	1	3
B4	6	7,116 (7,097 - 7,129)	49.5 (49.5 - 49.5)	0	0	0	6
B5	8	6,886 (6,643 - 7,122)	48.0 (46.4 - 49.5)	1	0	1	6

Table 2: Sums of squares, degrees of freedom and mean squares obtained from a linear model fitted to 616,945 pH observations from 93 animals on 13 farms. Fixed effects of farm, animal, α_1 & β_1 for the daily sine wave, and α_2 & β_2 for the milking-frequency sine wave were fitted along with interactions, as indicated. The values for the corresponding α & β effects have been combined to give an overall estimate for the relevant model component. Effects of farm and animal represent differences in mean pH, effects of daily and milking sine waves reflect the importance of short-term temporal variation, and interactions indicate the variation in patterns between animals and farms.

Model Component	Sum of Squares	Degrees of Freedom	Mean Squares
Farm	8,422.9	12	701.9
Animal	12,508.2	80	156.4
Daily Sine Wave	3,435.8	2	1,717.9
Milking Sine Wave	567.7	2	283.9
Farm : Daily Sine (interaction)	573.3	24	23.9
Farm : Milking Sine (interaction)	539.2	24	22.5
Animal : Daily Sine (interaction)	578.6	160	3.6
Animal : Milking Sine (interaction)	325.6	160	2.0
Residuals	18,108.2	616,480	0.0

Table 3. Estimates obtained from 2 multivariable linear mixed models relating 4 different daily pH summary statistics to the daily corrected milk yield and daily dry matter intake observed 2 days later. Summary statistics are mean residual pH, mean absolute residual pH and number of pH observations below and above the most extreme 1% values for that animal

	Daily corrected milk yield		**Daily dry matter intake**	
	Estimate	**p-value**	**Estimate**	**p-value**
Mean residual	-0.137	0.911	0.019	0.988
Mean absolute residual	-8.633	< 0.001	-8.698	< 0.001
Number of obs. below lowest 1%	0.041	0.056	0.015	0.530
Number of obs. above highest 1%	0.003	0.877	0.018	0.396

FIGURE LEGENDS

Figure 1. A flow-chart representation of the complete analysis procedure as used for this manuscript, with reference to the sections given in Materials and Methods.

Figure 2. Examples of sine waves for θ in the interval $[0, 2\pi]$ corresponding to the complete range of values in e.g. time of day. The skew parameter γ is shown varying in the set 0 (a), 1.5 (b), -0.5 (c), -3 (d) with constant phase shift ($\theta_0 = 0$) and amplitude ($A = 20$) parameters. Note that the standard sine wave is a special case of $\gamma = 0$

Figure 3. The raw mean (a) and raw variance (b) in pH series obtained from 93 animals from 13 farms

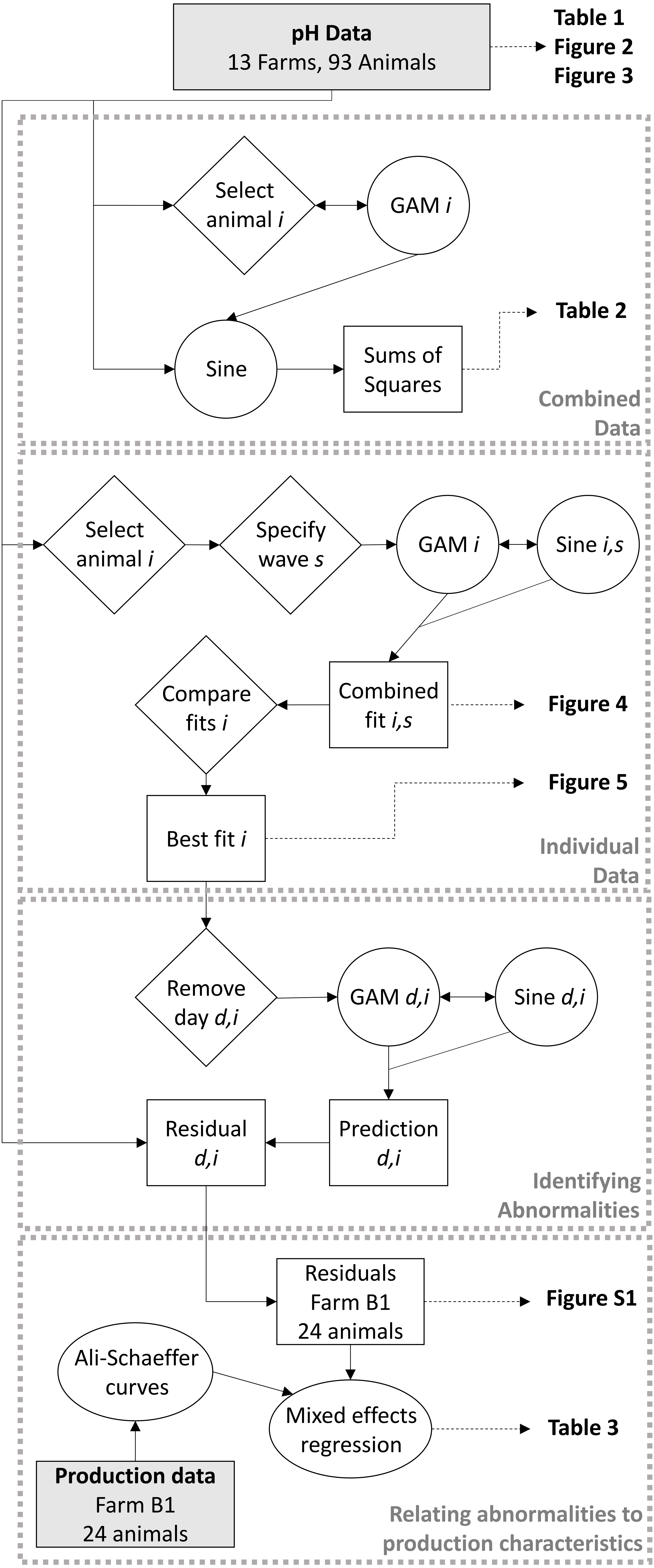
Figure 4. Illustration of the pH observations (grey dots) recorded over a 1-week period for a randomly chosen animal from each of herd B1 (left) and A1 (right), overlaid with the predictions from the GAM (top row), sine wave with daily frequency (2nd row), sine wave with milking frequency (3rd row), and combination of these 3 elements of the model (bottom row)

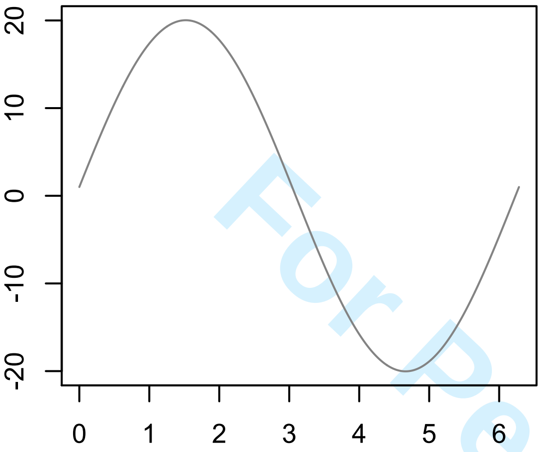
Figure 5. The proportion of the total variance explained by a GAM (a), as well as the proportion of the remaining variance explained by adding a single sine wave (b), a skew parameter (γ) to this sine wave (c), and then a second sine wave to fit the known milking times on the farm (d)

Figure 6. The fitted prediction from a single sine wave (including skew parameter) fitted to pH series obtained from 93 animals from 12 farms (longer-term variability modeled by the GAM is ignored)

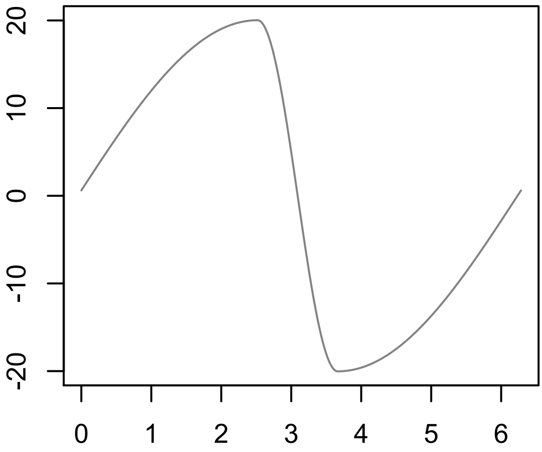
Supplemental Figure S1. The calculated daily mean residual (dashed line) and mean absolute residual (solid line) obtained from 24 animals on farm B1

Supplemental Figure S2. The daily number of pH observations above the highest 1% (crosses) and below the lowest 1% (dots) of all pH values for that animal, obtained from 24 animals on farm B1

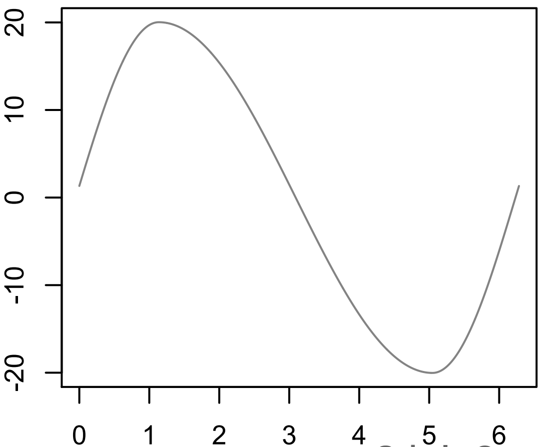

pH Data
13 Farms, 93 Animals
Table 1
Figure 2
Figure 3
Select animal i
GAM i
Sine
Sums of Squares
Table 2
Combined Data
Select animal i
Specify wave s
GAM i
Sine i,s
Combined fit i,s
Figure 4
Compare fits i
Best fit i
Figure 5
Individual Data
Remove day d,i
GAM d,i
Sine d,i
Residual d,i
Prediction d,i
Identifying Abnormalities
Residuals
Farm B1
24 animals
Figure S1
Ali-Schaeffer curves
Mixed effects regression
Table 3
Production data
Farm B1
24 animals
Relating abnormalities to production characteristics


20
10
0
-10
-20
0 1 2 3 4 5 6

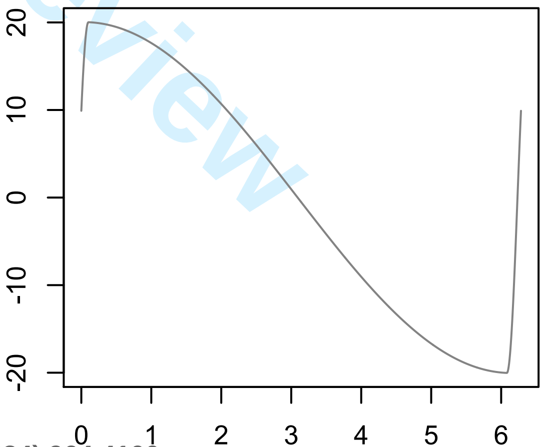
(a)


20
10
0
-10
-20
0 1 2 3 4 5 6

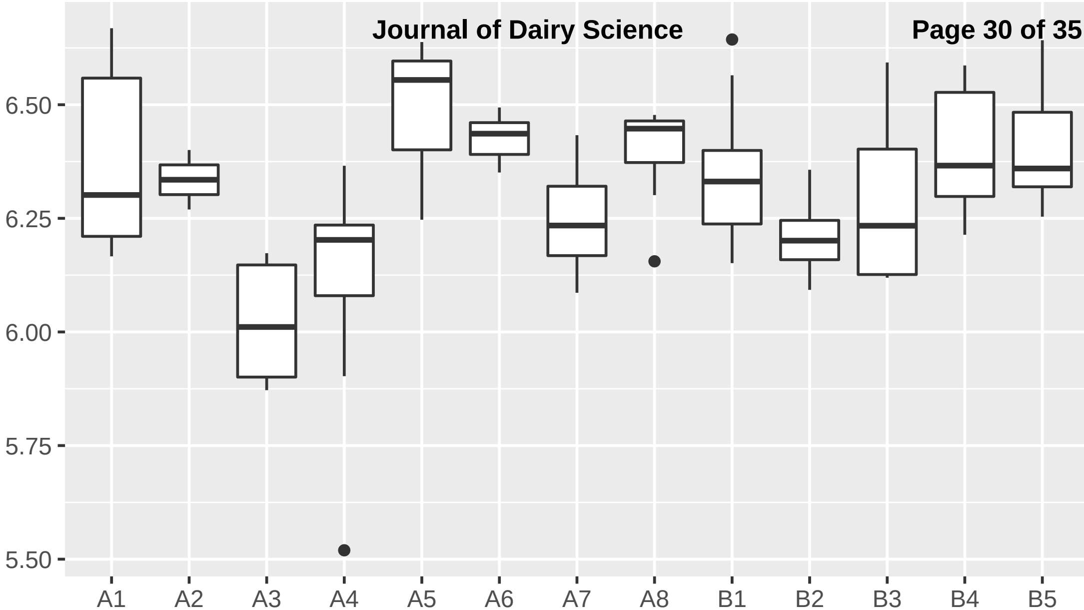
(b)


20
10
0
-10
-20
0 1 2 3 4 5 6

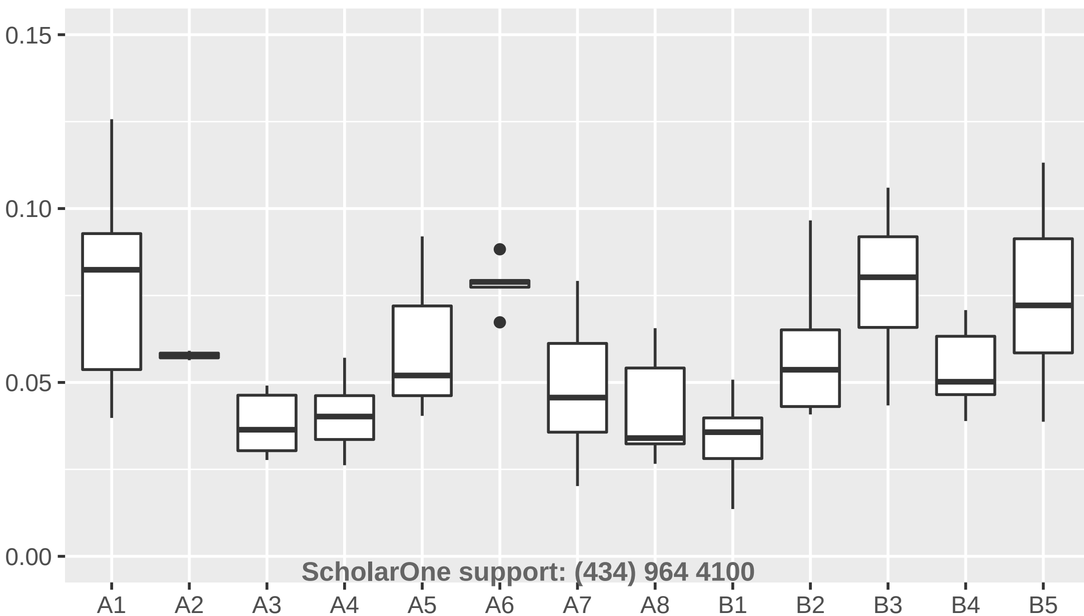
(c)


20
10
0
-10
-20
0 1 2 3 4 5 6

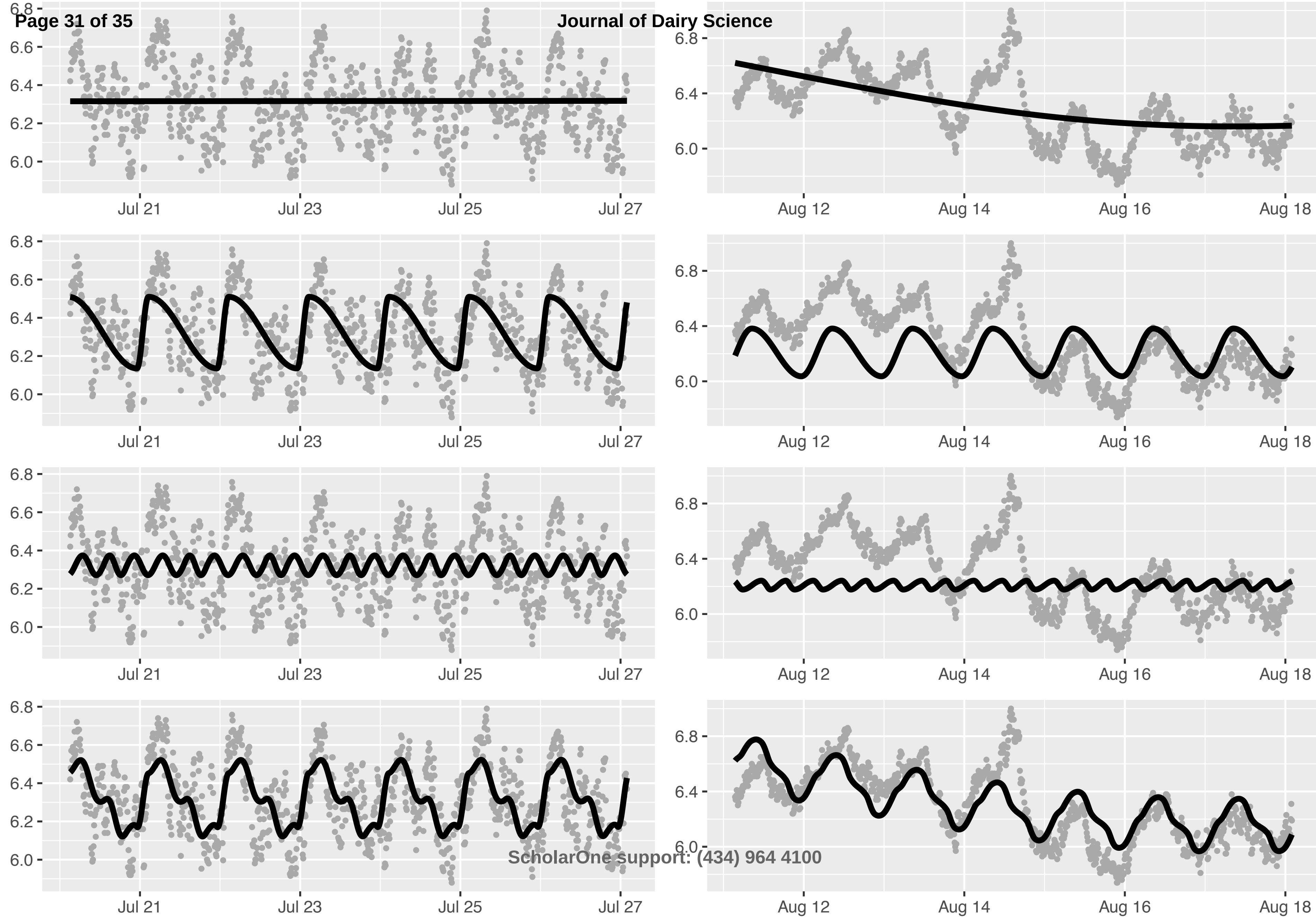
(d)

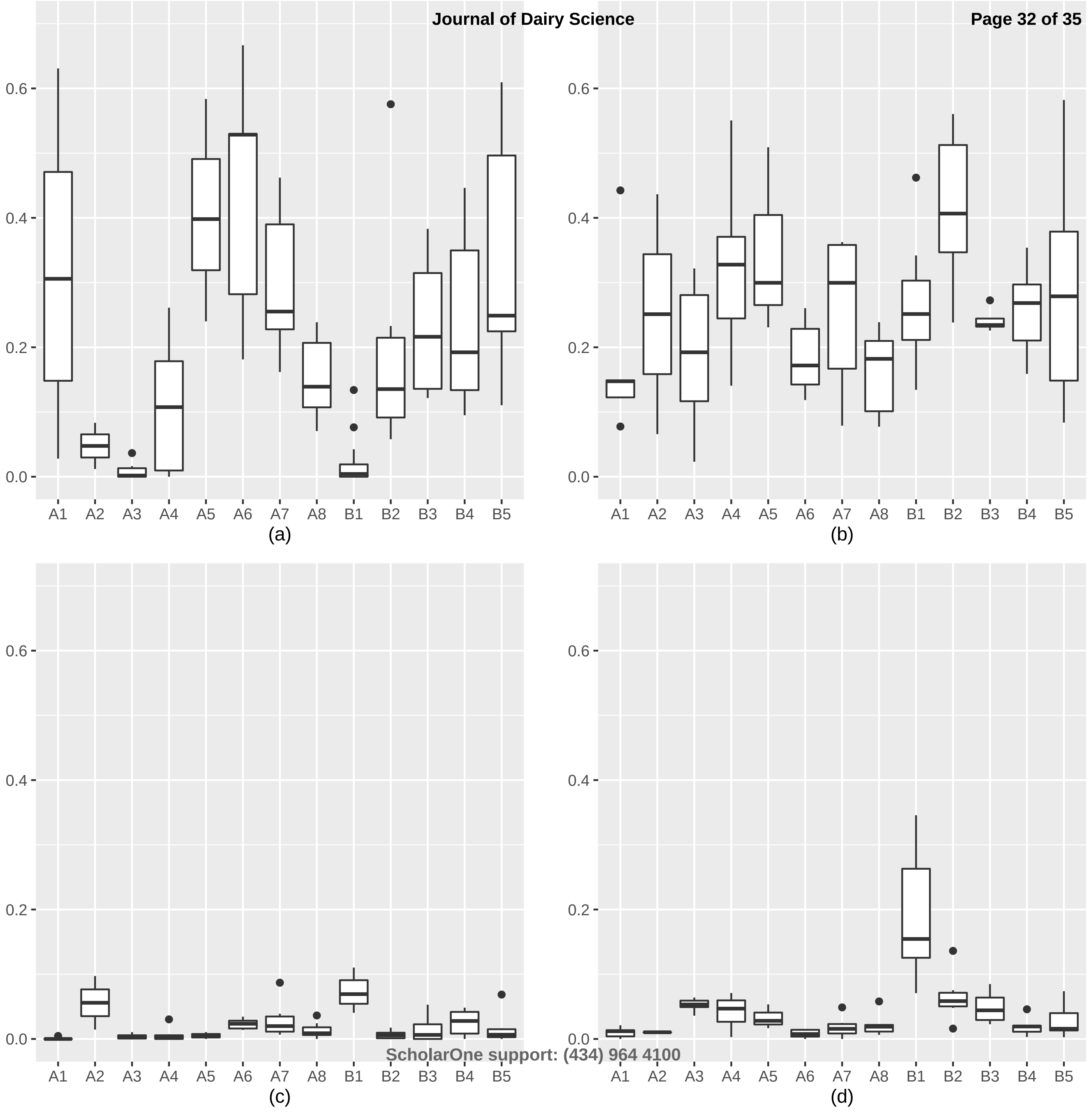

6.50
6.25
6.00
5.75
5.50
A1
A2
A3
A4
A5
A6
A7
A8
B1
B2
B3
B4
B5

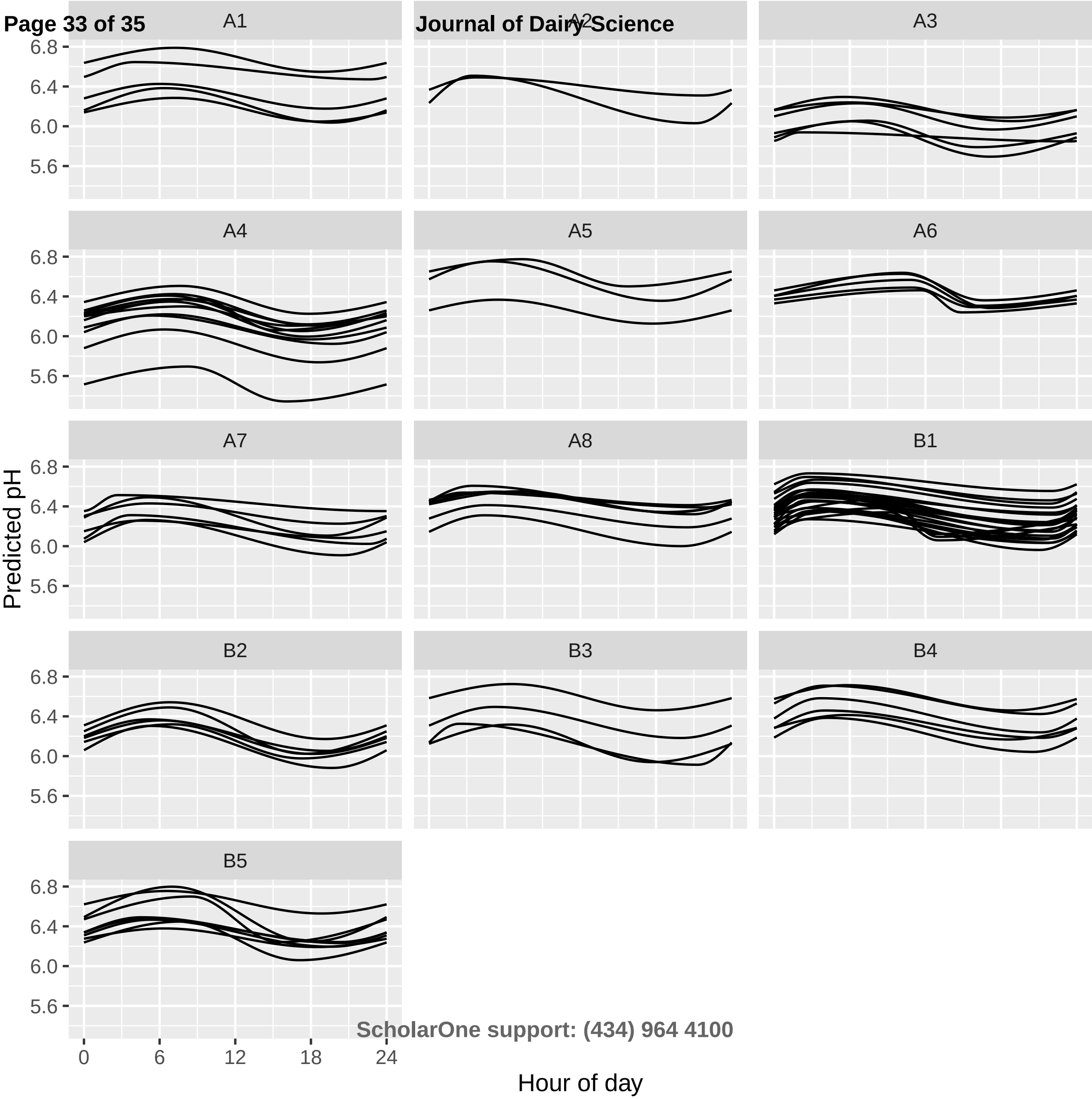
(a)

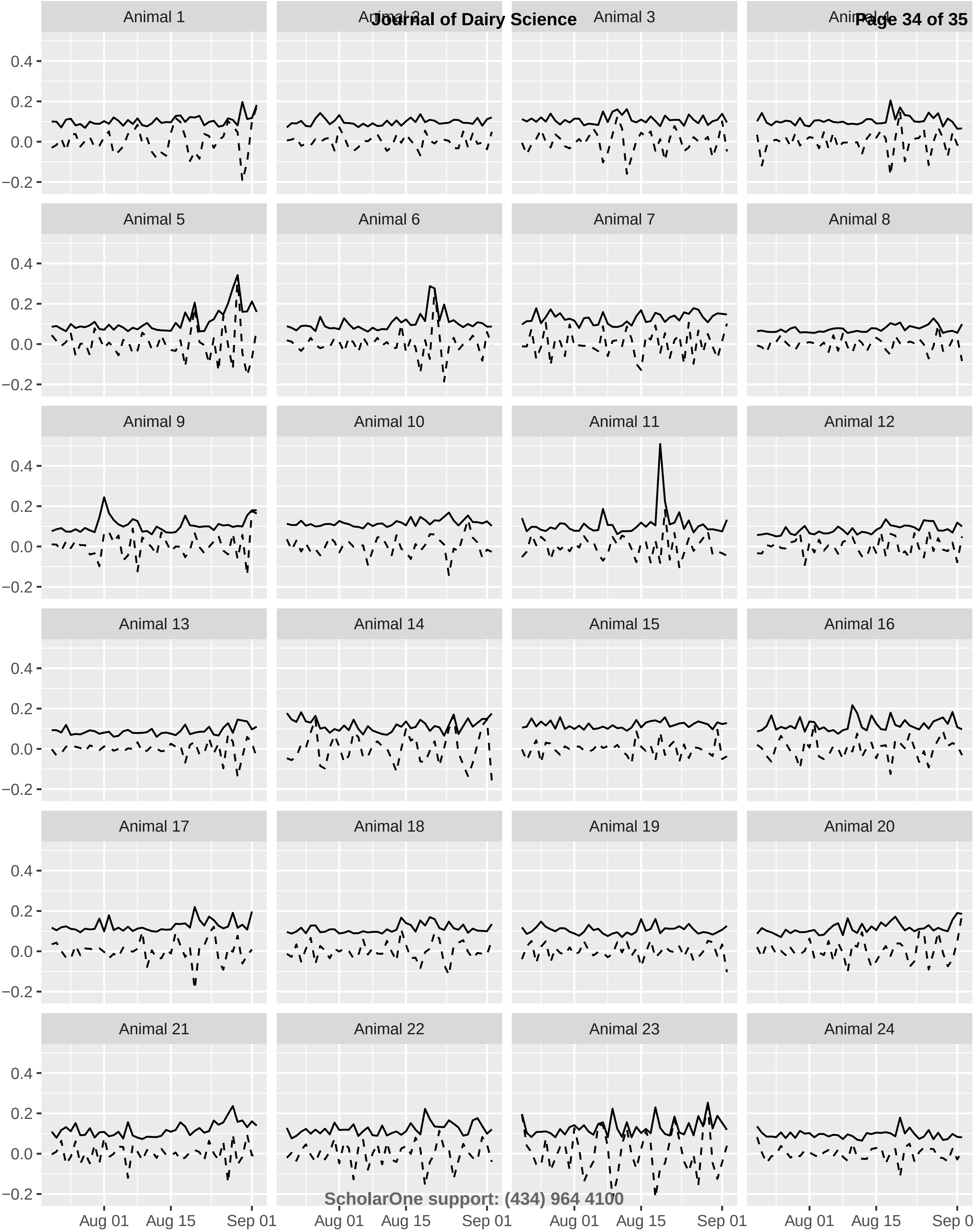

0.15
0.10
0.05
0.00
A1
A2
A3
A4
A5
A6
A7
A8
B1
B2
B3
B4
B5

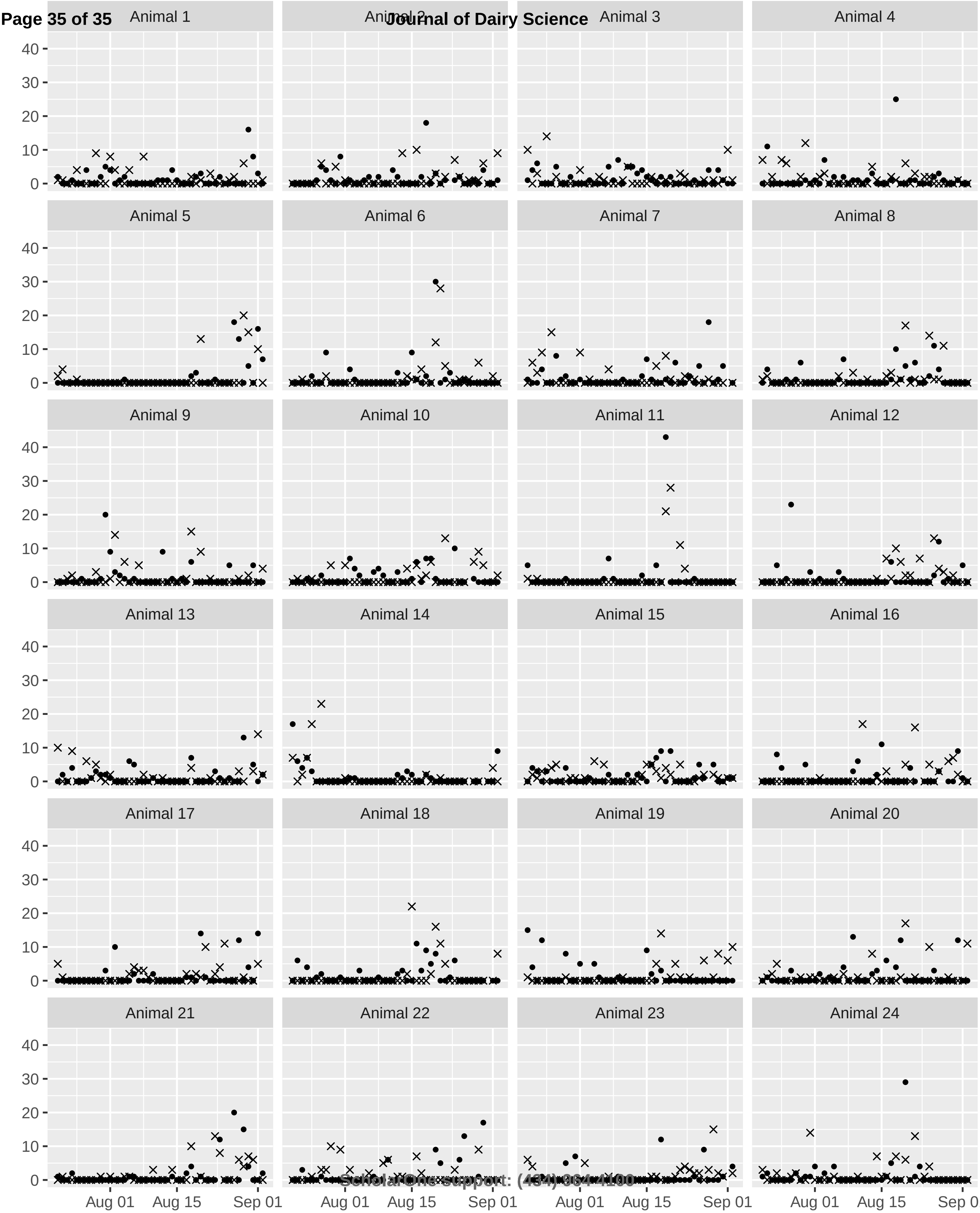
(b)


6.0
6.2
6.4
6.6
6.8
Jul 21
Jul 23
Jul 25
Jul 27
6.0
6.4
6.8
Aug 12
Aug 14
Aug 16
Aug 18


0.0
0.2
0.4
0.6
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5
(a)
(b)
(c)
(d)


A1
A2
A3
A4
A5
A6
A7
A8
B1
B2
B3
B4
B5
6.8
6.4
6.0
5.6
Predicted pH
0
6
12
18
24
Hour of day


Animal 1
Animal 2
Animal 3
Animal 4
Animal 5
Animal 6
Animal 7
Animal 8
Animal 9
Animal 10
Animal 11
Animal 12
Animal 13
Animal 14
Animal 15
Animal 16
Animal 17
Animal 18
Animal 19
Animal 20
Animal 21
Animal 22
Animal 23
Animal 24
0.4
0.2
0.0
−0.2
Aug 01
Aug 15
Sep 01


Animal 1
Animal 2
Animal 3
Animal 4
Animal 5
Animal 6
Animal 7
Animal 8
Animal 9
Animal 10
Animal 11
Animal 12
Animal 13
Animal 14
Animal 15
Animal 16
Animal 17
Animal 18
Animal 19
Animal 20
Animal 21
Animal 22
Animal 23
Animal 24
0
10
20
30
40
Aug 01
Aug 15
Sep 01