



HAL
open science

Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows

Marius C. Codrea, Soren Hojsgaard, Nicolas N.C. Friggens

► To cite this version:

Marius C. Codrea, Soren Hojsgaard, Nicolas N.C. Friggens. Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows. *Journal of Animal Science*, 2011, 89, pp.3089-3098. 10.2527/jas.2010-3753 . hal-01186805

HAL Id: hal-01186805

<https://hal.science/hal-01186805>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows

M. C. Codrea, S. Højsgaard and N. C. Friggens

J ANIM SCI 2011, 89:3089-3098.

doi: 10.2527/jas.2010-3753 originally published online April 8, 2011

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org/content/89/10/3089>



American Society of Animal Science

www.asas.org

Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows¹

M. C. Codrea,*² S. Højsgaard,† and N. C. Friggens‡§

*Faculty of Science and Technology, Department of Animal Health and Bioscience, and †Faculty of Science and Technology, Department of Genetics and Biotechnology, Aarhus University, Blichers Allé 20, PO Box 50, Tjele, 8830, Denmark; ‡INRA, UMR 791 Modélisation Systémique Appliquée aux Ruminants, 16 rue Claude Bernard, 75005 Paris, France; and §AgroParisTech, UMR 791 Modélisation Systémique Appliquée aux Ruminants, 16 rue Claude Bernard, 75005 Paris, France

ABSTRACT: Recent advances in on-farm technology now provide us with multiple time-series of reliably measured indicators of animal performance and status at the level of the individual. This paper presents a smoothing approach for extracting biologically meaningful features from such time series using bovine milk yield data as an example. The main goal of this study was to illustrate how the method can be used to detect production deviations, extract quantifiable features of the deviation profiles, and thus provide means to examine hypotheses concerning the nature of the deviations. The effectiveness of the method was assessed with complete lactation curves from 47 Holstein cows. Within their lactations, the cows were each subjected to 1 nutritional challenge for a period of 4 d (their standard diet: a maize silage-based total mixed ration was diluted with 60% wheat straw), which provoked a decline in the milk yield in all cows. The challenge was imposed between the same calendar days for all cows. Thus, the cows were at different stages of lactation: early ($n = 14$), mid ($n = 15$), and late ($n = 18$). Each

milk-yield curve was decomposed into components that capture the short-term deviations of the cow such as the response to the nutritional challenge and describe the phenotypic potential yield function of that cow throughout its lactation. The difference between the 2 components gives a measure of the milk loss. In all, 480 deviations were detected from the complete lactations of 47 cows. The milk loss provoked by the feeding challenge ($n = 47$) was significantly related to the milk yield immediately before the challenge ($r = 0.86$, $P < 0.01$). The correlation between the rate of recovery and milk loss was ($r = 0.94$, $P < 0.01$). Further, there was no significant slope ($P > 0.1$) to the relationship between the ratio (rate of recovery/milk loss) and days from calving, indicating that the force of recovery was unaffected by stage of lactation. These results suggest that differential smoothing can be a useful tool for quantifying biological disturbances in animal performance and for extracting features that relate to the potential and robustness of an animal.

Key words: B-spline, feature extraction, phenotype, smoothing

©2011 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2011. 89:3089–3098
doi:10.2527/jas.2010-3753

INTRODUCTION

Traditionally, in agricultural research, smoothing has been used largely to reduce the variability or noise in

data so as to make it easier to detect differences between treatments (e.g., feeds or other systematic factors). In such situations, the aim of smoothing is to remove short-term disturbances from the data. Rather than using smoothing to eliminate disturbances, an alternative approach is to use smoothing to capture them. Disease episodes often cause short-term disturbances (e.g., a decrease in milk yield followed by a recovery; Fourichon et al., 1999). Similarly, if an animal goes “off-feed,” this will disturb production traits, although in this case, the disturbance can be of longer duration (Bareille et al., 2003). Being able to identify and quantify the extent of these disturbances has been shown to be of value for

¹This study, which was part of the Biosens project, used data from the “On-line data to monitor health status of cows” project. The work was supported by the Danish Ministry of Food, Agriculture and Fisheries, the Danish Cattle Industry via Finance Committee Cattle (Copenhagen, Denmark), and the Faculty of Agricultural Sciences, Aarhus University (Aarhus, Denmark).

²Corresponding author: marius.codrea@agrsci.dk

Received December 2, 2010.

Accepted April 3, 2011.

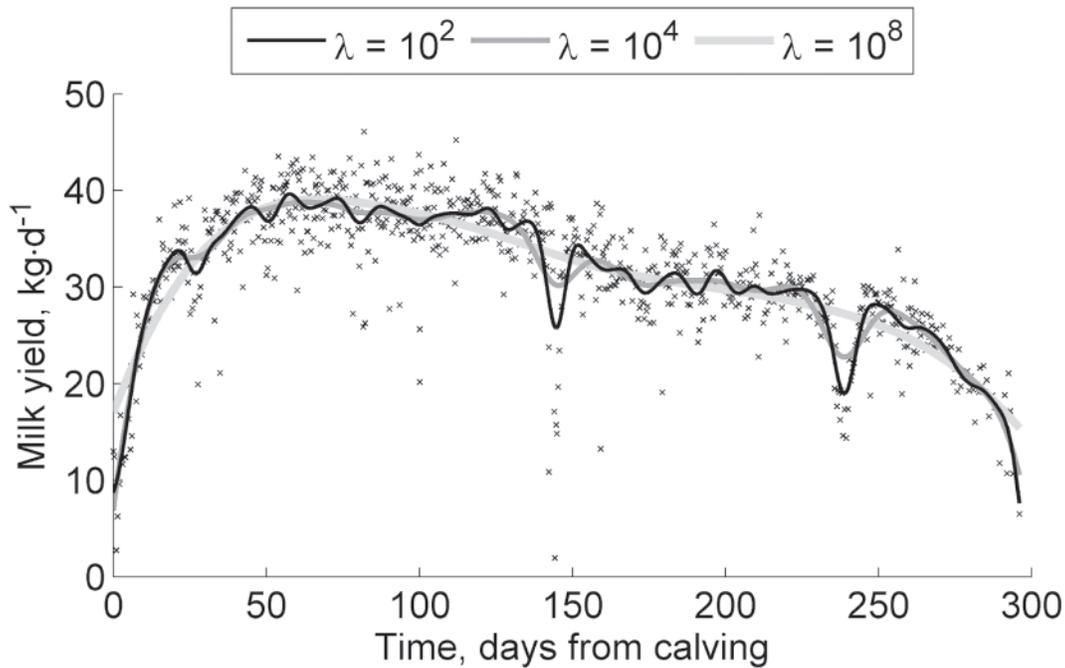


Figure 1. An example of different degrees of smoothing applied to a lactation curve of milk-yield records. The different degrees of smoothing are obtained by imposing different weights on the roughness penalty term λ (see Eq. 1). Short-term disturbances in milk yield can be detected ($\lambda = 100$), as well as the general long-term trend ($\lambda = 10^8$; i.e., the natural changes in milk yield related to stage of lactation).

health management (Friggens et al., 2007; Kamphuis et al., 2008a).

The deviation and severity of an animal response to a disturbance (e.g., disease challenge, nutritional shortage) reflects, to some degree, the capacity of that animal to respond. As such, these responses contain biologically meaningful information. Thus, a method for quantifying the features of such data may provide quantitative information about the phenotype of the animal at a functional level. Using smoothing to identify disturbances implies decomposing the time series in question into the underlying long-term trend and deviations from it. The aims of this study are to examine the issues involved in decomposing time-series data to capture biologically meaningful features and to evaluate the ability of this approach to detect (known) disturbances in the time series of milk production. To this end, we have chosen a smoothing method that is an example of a functional form, which facilitates the extraction of features without compromising flexibility to fit different curve shapes. It is important to note that no assumption on the parameterization of the curve is needed.

MATERIALS AND METHODS

Experimental procedures involving animals were approved by the Danish Animal Experiments Inspectorate and complied with the Danish Ministry of Justice Law No. 382 (June 10, 1987) and Acts 739 (December 6, 1988) and 333 (May 19, 1990) concerning animal experimentation and care of experimental animals.

Animals and Milking

In this study we analyzed the lactation curves from 47 Holstein cows from the resident herd at the Danish Cattle Research Centre (Tjele, Denmark). The cows were housed in a loose-housing environment with free stalls containing mats of hard rubber material and sawdust as bedding. Cows had free access to water and to an automatic milking system [Voluntary Milking System (VMS), DeLaval, Tumba, Sweden]. Because the cows visit the VMS on a voluntary basis, the milking intervals (the time interval between successive milking events) are variable. The minimum milking interval permitted by the VMS was 5 h. Cows that had not presented to the VMS in the previous 18 h were brought to the VMS for milking.

In all, 39,962 milkings were recorded from all lactations of the 47 cows with an average milking interval of 0.41 d (SD = 0.14 d). The milk yield per visit at the automatic milking system was converted to milk yield per day by dividing the measured milk yield by the milking interval, in days. Very short milking intervals, usually caused by premature removal of the milking set (i.e., those associated with incomplete preceding milkings), give rise to erroneously large yield values. Therefore, we adjusted the milking intervals shorter than 0.2 d to a value of 0.2 d. Similarly, very long milking intervals, usually caused by a cow being removed from the group for management reasons, can give rise to implausibly low yields, and we consequently set the maximum interval to 0.8 d. These adjustments affected 203 short- and 937 long-milking intervals.

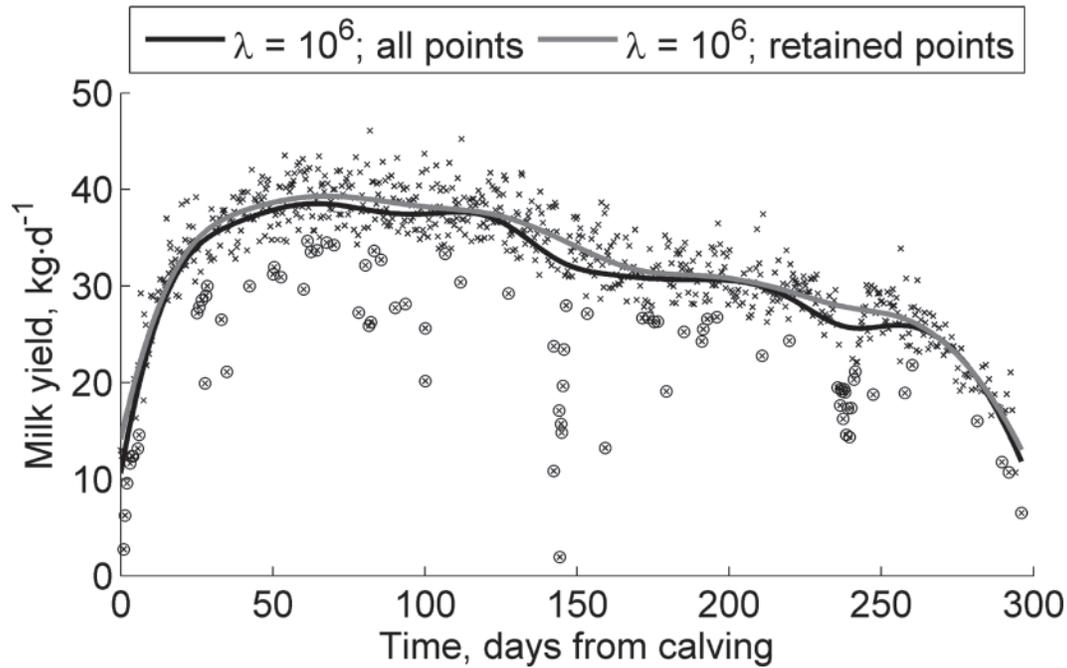


Figure 2. The use of residual filtering to offset bias in the yield records. An initial smoothing can indicate the data points that give large negative residuals (encircled points), which can be discarded from a second smoothing. A slightly increased and smoother curve is found with the retained points.

Within their lactations, the cows were each subjected to 1 nutritional challenge for a period of 4 d. During this period, their standard diet (a maize silage-based total mixed ration) was diluted with 60% wheat straw. The challenge was imposed between the same calendar days for all cows, and thus the challenge occurred at

different stages of lactation: early ($n = 14$), mid ($n = 15$), and late ($n = 18$). All cows were healthy before and during the challenge trial. The complete lactations were used in this study and the effectiveness of the proposed method was particularly evaluated around the known challenge period. Other disease episodes,

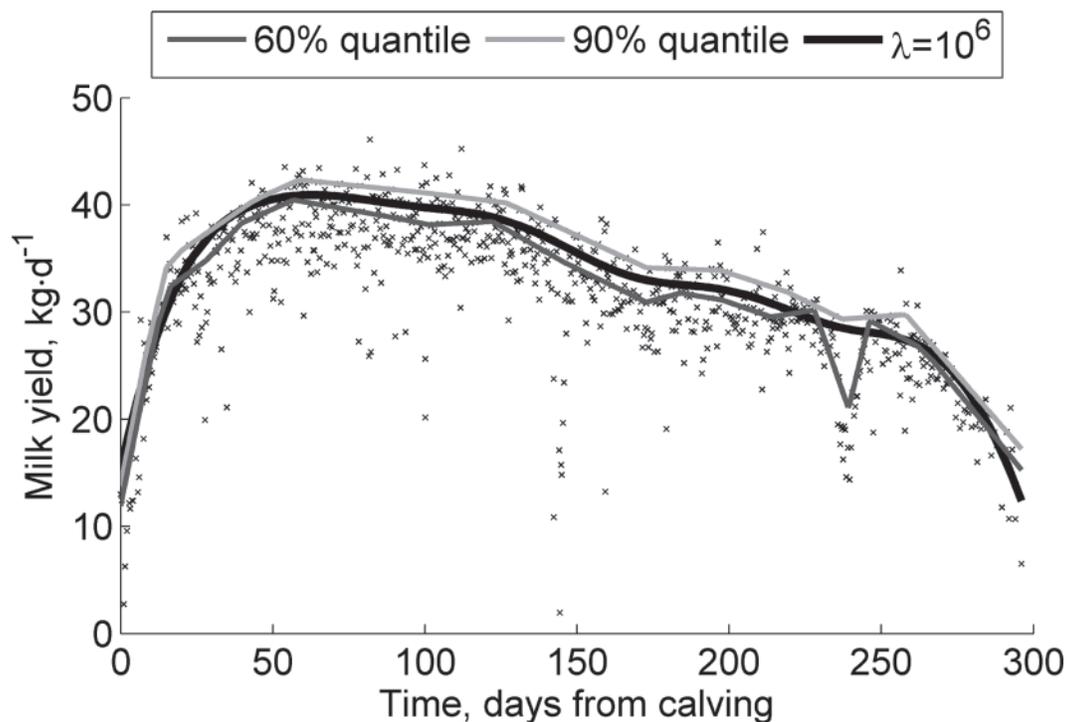


Figure 3. The use of quantile regression to offset bias in the yield records. The resulting curve ($\lambda = 10^6$) represents undisturbed phenotypic potential when smoothing using data points within 60 and 90% quantiles.

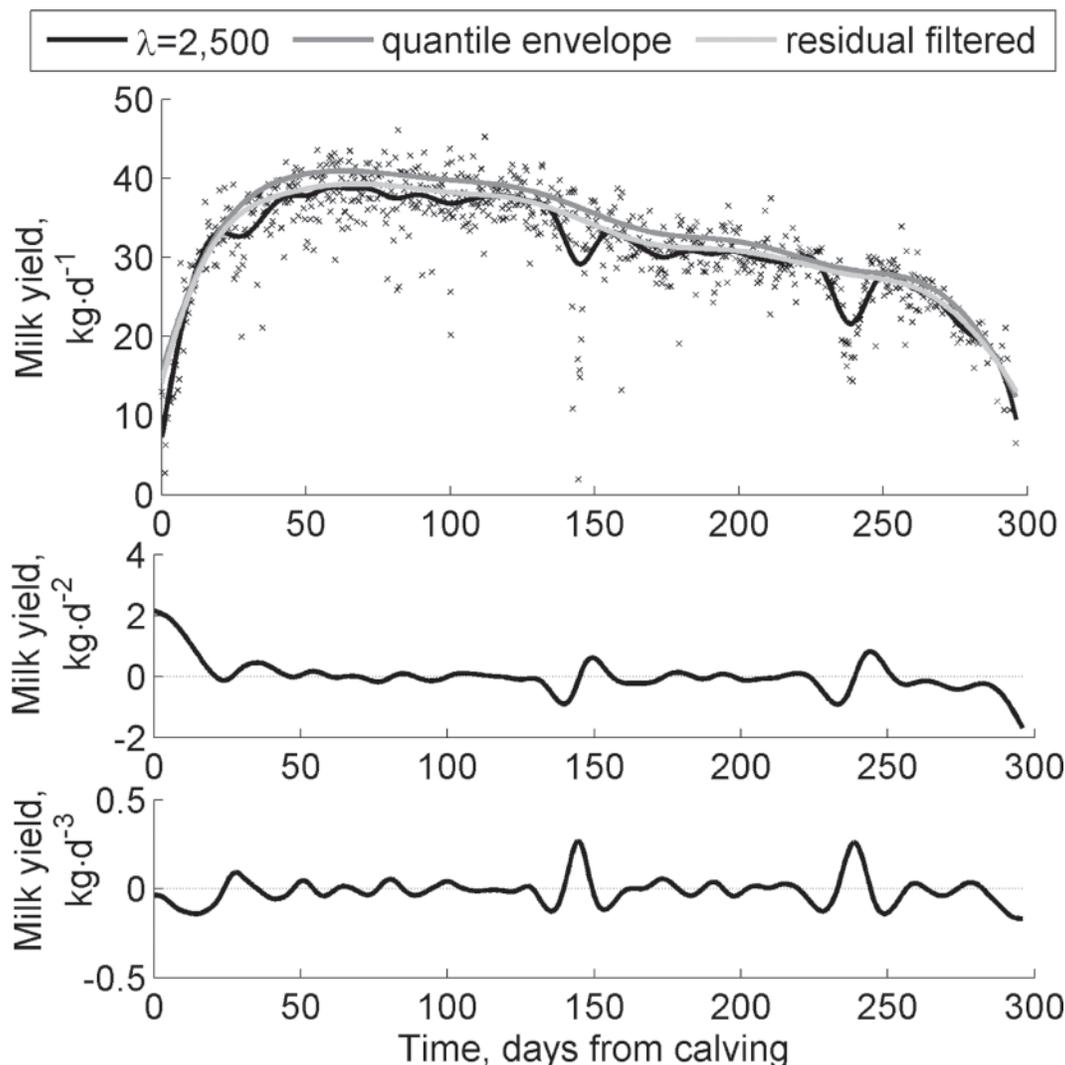


Figure 4. Top panel: an example of differential smoothing to obtain an offset curve representing undisturbed phenotypic potential (quantile envelope) and the much rougher curve ($\lambda = 2,500$), which captures biological disturbances. First and second derivatives of the rough curve can be estimated at any time point along the lactation curve. These are shown in the middle and bottom panels, respectively.

whether known or unknown, were not thoroughly investigated. The undisturbed phenotypic potential of each cow (i.e., the estimated milk yield if free from challenges) was the self-reference baseline for quantifying the deviations in yield.

Smoothing Method

We adopted the smoothing approach described, and termed functional data analysis, by Ramsay and Silverman (2005).

Formally, the objective function $[f_{MY}(c)]$ being minimized is

$$f_{MY}(c) = \sum_{j=1}^N [y_j - x(t_j)]^2 + \lambda \int_0^L [\partial^4 x(t)]^2 dt \quad [1]$$

$$x(t) = c\varphi(t),$$

where y_j are N observed raw values in the time-series interval $[0, L]$, and $x(t)$ is the smooth function constructed from the set of basis functions $\varphi(t)$, which are linearly combined with coefficients c . The scalar λ is a weighting term whose role is discussed in the following. We used B-spline basis functions, which can construct piece-wise polynomial functions joined at the knots (the points y_j used in the fitting procedure). The roughness penalty, in this case the curvature of the second derivatives, is controlled by the scalar λ . When $\lambda = 0$, $x(t)$ fits the data as well as is possible in terms of the squared error. As λ increases, more weight is put on the penalty term and thus the second derivative converges to a straight line (zero curvature). Therefore, different values for λ give different degrees of smoothing in both the function itself, as well as in its derivatives.

It is important to note that no a priori assumptions constrain the shape of the curve or the distance between knots. That is, the raw data need not be equidistantly sampled. Actually, it may be desirable to have

more knots at regions with sharp changes, and only a few knots are required where the curve is rather flat, and thus, focus the smoothing to capture finer details in specific regions of interest. In this study, to imitate real-life situations, we did not use the information about the known challenge period for choosing knots. With the data described in functional form (basis functions, coefficients, and knots), the smooth curve can be estimated at any interior time point within the range of the data. This means that we can sample the curve with arbitrarily small resolution.

Having access to the derivatives of the function opens several possibilities: i) as they reflect rates of changes, they have intrinsic value in characterizing the dynamic behavior of the system under investigation; ii) deriva-

tives provide means to find points of interest (landmarks) in the data (e.g., at peaks and troughs in the data, the first derivative is zero); and iii) landmarks can also be used to align and compare individual time profiles. Hence, several features can be extracted from the derivatives themselves, as well as from the timing and the amplitude of landmarks.

RESULTS

Differential Smoothing

By imposing different values of λ , which weights the roughness penalty, different degrees of smoothing can be readily obtained (Figure 1). Thus, it is possible to

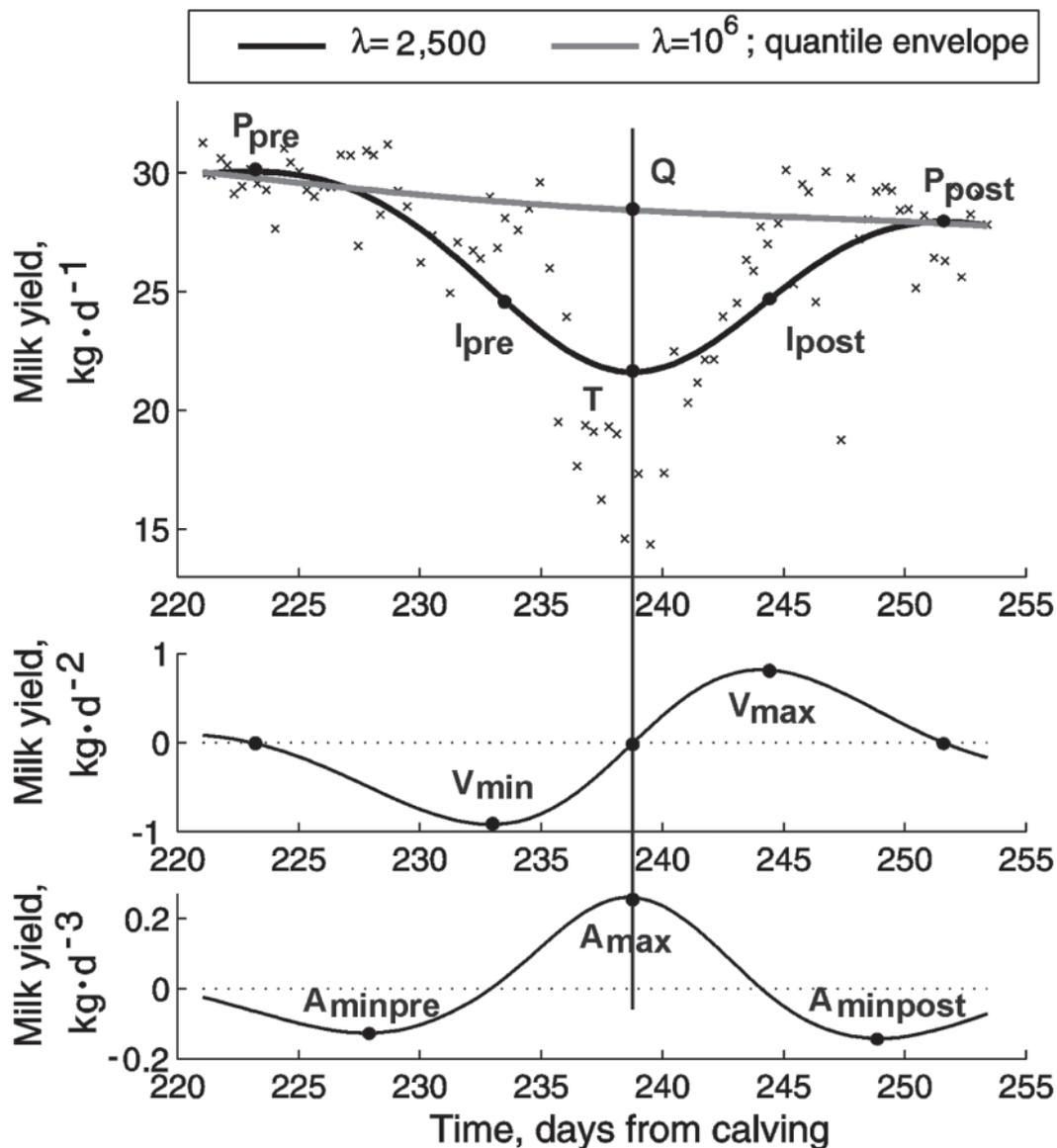


Figure 5. Possible features for detecting and quantifying deviations in milk yield. Each deviation profile has a trough (T), which splits it into pre- and post-trough periods that correspond to response and recovery periods in milk yield. The trough occurs where the first derivative crosses 0 from negative to positive. Two peaks (P_{pre} and P_{post}) can be located on either side of T where the first derivative crosses 0 from positive to negative. The pre-trough inflection point (I_{pre}) corresponds to the minimum value of the first derivative (V_{min}), and the post-trough inflection point (I_{post}) corresponds to the maximum of the first derivative (V_{max}). The extreme values of the second derivative of the smooth function are denoted by A_{max} , A_{minpre} , and $A_{minpost}$, respectively. The potential level (Q) from the quantile envelope is an estimate of what the milk yield would have been in the absence of the challenge.

Table 1. Summary of the profile features of deviations in milk yield provoked by the nutritional challenge in cows ($n = 47$)

Item ¹	Mean	SD
First derivative features, $\text{kg}\cdot\text{d}^{-2}$		
V_{\min}	0.98	0.45
V_{\max}	0.84	0.33
Second derivative features, $\text{kg}\cdot\text{d}^{-3}$		
$A_{\min\text{pre}}$	0.17	0.08
A_{\max}	0.28	0.10
$A_{\min\text{post}}$	0.13	0.05
Timing features, d		
$\Delta t(I_{\text{pre}}, P_{\text{pre}})$	8.48	3.03
$\Delta t(T, P_{\text{pre}})$	15.73	7.49
$\Delta t(T, I_{\text{pre}})$	7.25	7.44
$\Delta t(I_{\text{post}}, T)$	5.62	1.33
$\Delta t(P_{\text{post}}, I_{\text{post}})$	11.80	7.43
$\Delta t(P_{\text{post}}, T)$	17.43	7.84
Milk yield loss, ² $\text{kg}\cdot\text{d}^{-1}$ ($Q - T$)	7.52	2.38

¹Each deviation profile has a trough (T), which splits it into pre- and post-trough periods that correspond to response and recovery periods in milk yield. Two peaks (P_{pre} and P_{post}) and 2 inflection points (I_{pre} and I_{post}) can be detected on either side of T. The extreme values of the first and second derivatives of the smooth curve are denoted by (V_{\min} and V_{\max}) and (A_{\max} , $A_{\min\text{pre}}$, and $A_{\min\text{post}}$), respectively. The timing features (Δt) are absolute time differences among the points of interest P_{pre} , I_{pre} , T, I_{post} , and P_{post} .

²Milk yield loss ($Q - T$) = the difference between the milk level at trough (T) and the corresponding potential level (Q) on the quantile envelope, which is a positively offset-smoothed curve that represents the undisturbed phenotypic potential yield.

capture different aspects of the milk-yield curve. Using a small value of λ captures the finer details of the milk-yield curve (e.g., $\lambda = 100$; Figure 1), which is useful for describing the short-term responses of the cow to, for example, infection or nutritional challenges. Using a larger value of λ provides a more stringent smoothing (e.g., $\lambda = 10^8$) that can reveal the long-term component of the yield curve (i.e., the natural changes in milk yield related to stage of lactation). The case with $\lambda = 10^4$ illustrates an intermediate profile.

To quantify the short-term disturbances (in this case, milk loss) the sudden decline in milk yield needs to be related to another quantity (e.g., the long-term curve). However, it is desirable to relate the loss with respect to what the cow would have produced when free from biological and technical disturbances. Such a baseline, termed here the “phenotypic potential yield curve,” would provide more accurate measures of the biological response and requires that the general long-term smoothed curve is adjusted for these disturbances. The rationale in constructing such a function is as follows: i) there is a likelihood of a systematic bias in random error (e.g., milking machine errors are mostly low readings, mainly milk yields resulting from interrupted milkings), and ii) the majority of biological disturbances decrease milk yield. These 2 effects will create some degree of negative bias and an underestimate of the phenotypic potential curve, if ignored. Therefore, we need to positively offset the smoothed curve to obtain

a curve that is representative of undisturbed phenotypic potential. A simple way of achieving this is by setting up asymmetric bounds for acceptable residuals, and then either exclude extreme values or down-weight their contribution to the sum of squares. This can be done in several ways, and here we illustrate 2 such alternatives. An initial smoothing can indicate the data points that give large negative residuals (>1 SD of all the residuals across lactation), which can be discarded from a second smoothing. Such an offset curve, called here “residual filtered,” is illustrated in Figure 2. A second example, using quantile regression (Koenker, 2005) with confidence limits of 90 and 60%, is shown in Figure 3. We call the resulting curve a quantile envelope. For the actual calculation of the quantiles we used Koenker’s `quantreg` R package (<http://cran.r-project.org/web/packages/quantreg/>). We chose a nonparametric smoothing method with a penalty factor on the gradient, which gives piecewise linear fitted quantile functions. Thus, the quantile envelope is derived from piecewise estimates of the residual variation (data points above the 90% and below the 60% quantiles are excluded from the smoothing fit). In both offsetting strategies, the adjustment accounts for the inherent variability in the original measures.

Examples of offset curves representing undisturbed phenotypic potential and the much rougher curve that captures biological disturbances are shown in Figure 4 (top panel). In the subsequent analyses and results, we have chosen to use the quantile envelope method for deriving the undisturbed phenotypic potential because it is more internally consistent with the time-series approach as it accommodates the possibility of differences in residual variation at different times from calving.

For the purposes of the present paper, the choice of which method is used for off-setting the smooth fit to derive the undisturbed phenotypic potential is not important. What is important is to be clear that the choices of values for parameters such as λ , the number of knots in the fit, and the limits for the quantile regression were not based on a rigorous statistical optimization of parameters. Because the objective of this approach is to provide a biologically meaningful differential smoothing of the data, these parameters were chosen to best capture (i.e., fit by eye) the biological phenomena we wish to isolate. This “bio-logic” approach would be problematic if these parameters were adjusted for each individual curve being fitted because this would render invalid any comparison across cows. Thus, the same parameter values were used across all cows and, within the coefficients chosen, fitting was based on minimizing least squares and taking into account the distributions of residuals.

Feature Extraction: Quantifying Deviations and Deviation Profiles

In Figure 4, the middle and bottom panels are the first and the second derivatives of the rough curve ($\lambda =$

2,500). The derivatives can be useful features in their own right and also provide means to find relevant landmarks of the smooth curve. At peaks and troughs in the data, the first derivative is 0, and at inflection points (where the function changes concavity), the second derivative is 0. Figure 5 presents these features for 1 deviation profile (caused by the feed restriction imposed between 235 and 239 d from calving) in the smoothed milk yield curve shown in Figure 4. By detecting peaks and troughs, the deviation profile can be split into pre- and post-trough periods that correspond to response and recovery periods in milk yield. A panel of features that relate to aspects of the deviation profile of the cow can then be extracted. Examples of such features to quantify the loss, the timing of the response, and other aspects of the deviation profile are shown in Figure 5. The trough (T) is located where the first derivative values cross 0 from negative to positive. The pre-trough peak (P_{pre}) and the post-trough peak (P_{post}) are located where the first derivative values

cross 0 from positive to negative. The pre-trough inflection point (I_{pre}) corresponds to the minimum value of the first derivative (V_{min}) and the post-trough inflection point (I_{post}) corresponds to the maximum of the first derivative (V_{max}). The extreme values of the second derivative are denoted by (A_{max} , A_{minpre} , and $A_{minpost}$), respectively. The milk yield loss ($Q - T$) is the difference between the milk level (T) and the corresponding potential level (Q) on the quantile envelope, which is a positively off-set smoothed curve that represents the undisturbed phenotypic potential yield.

In addition to the absolute levels of the features, their relative timing can also be powerful profile descriptors. Thus, we can construct timing features (Δt) as absolute time differences among the points of interest P_{pre} , I_{pre} , T , I_{post} , and P_{post} . For example, the response time $\Delta t(P_{pre}, T)$ can be calculated as the difference between the absolute time(P_{pre}) - time(T). Similarly, the recovery time can be quantified by $\Delta t(T, P_{post})$. Another example could be $\Delta t(T, I_{post})$, which measures how fast

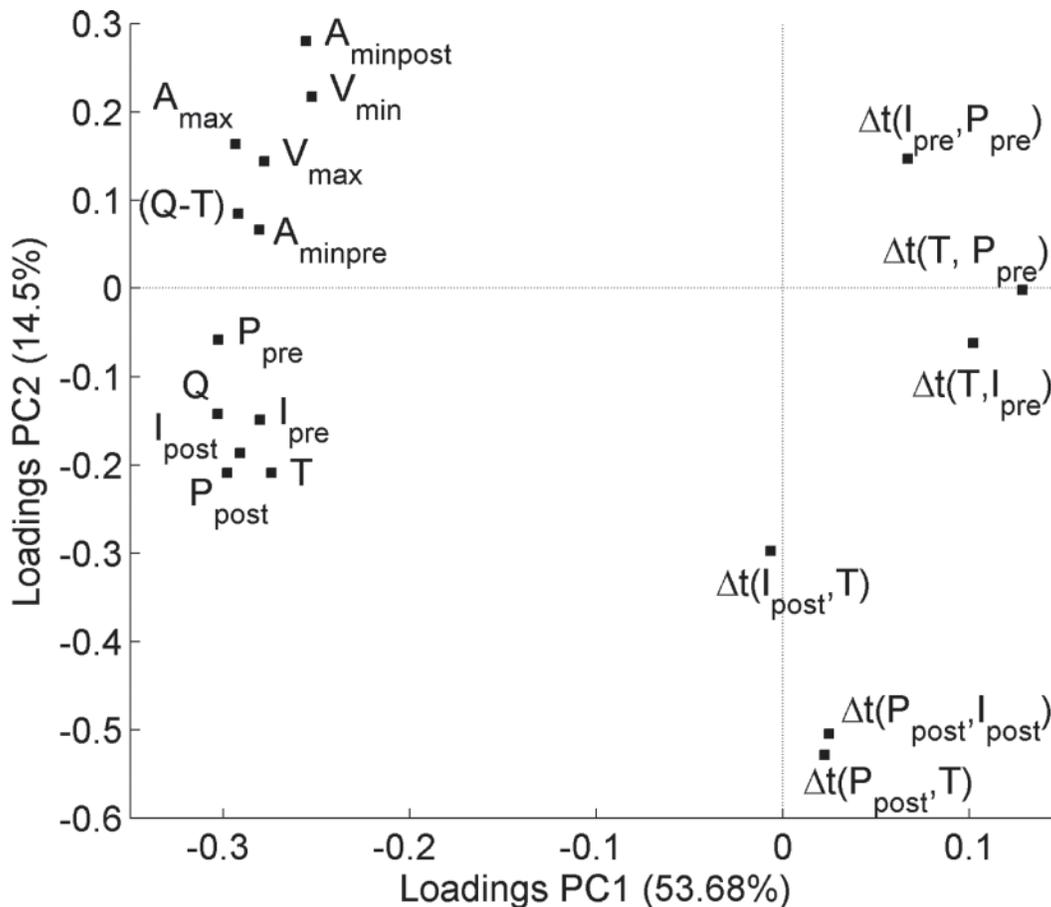


Figure 6. The relationships between the different deviation profile features (Figure 5) in a principal components analysis. Each deviation profile has a trough (T), which splits it into pre- and post-trough periods that correspond to response and recovery periods in milk yield. Two peaks (P_{pre} and P_{post}) and 2 inflection points (I_{pre} and I_{post}) can be detected on either side of T . The extreme values of the first and second derivatives of the smooth curve are denoted by (V_{min} and V_{max}) and (A_{max} , A_{minpre} , and $A_{minpost}$), respectively. The milk yield loss ($Q - T$) is the difference between the milk level (T) and the corresponding potential level (Q) on the quantile envelope, which is a positively offset-smoothed curve that represents the undisturbed phenotypic potential yield. A clear grouping can be observed along the first principal component (PC1): the timing features (Δt), which are absolute time differences among the points of interest P_{pre} , I_{pre} , T , I_{post} , and P_{post} and are well separated from the absolute milk levels (P_{pre} , I_{pre} , T , Q , I_{post} , and P_{post}) and from the derivative features (V_{min} , V_{max} , A_{max} , A_{minpre} , and $A_{minpost}$). The later 2 groups are well distinguished in the second principal component (PC2).

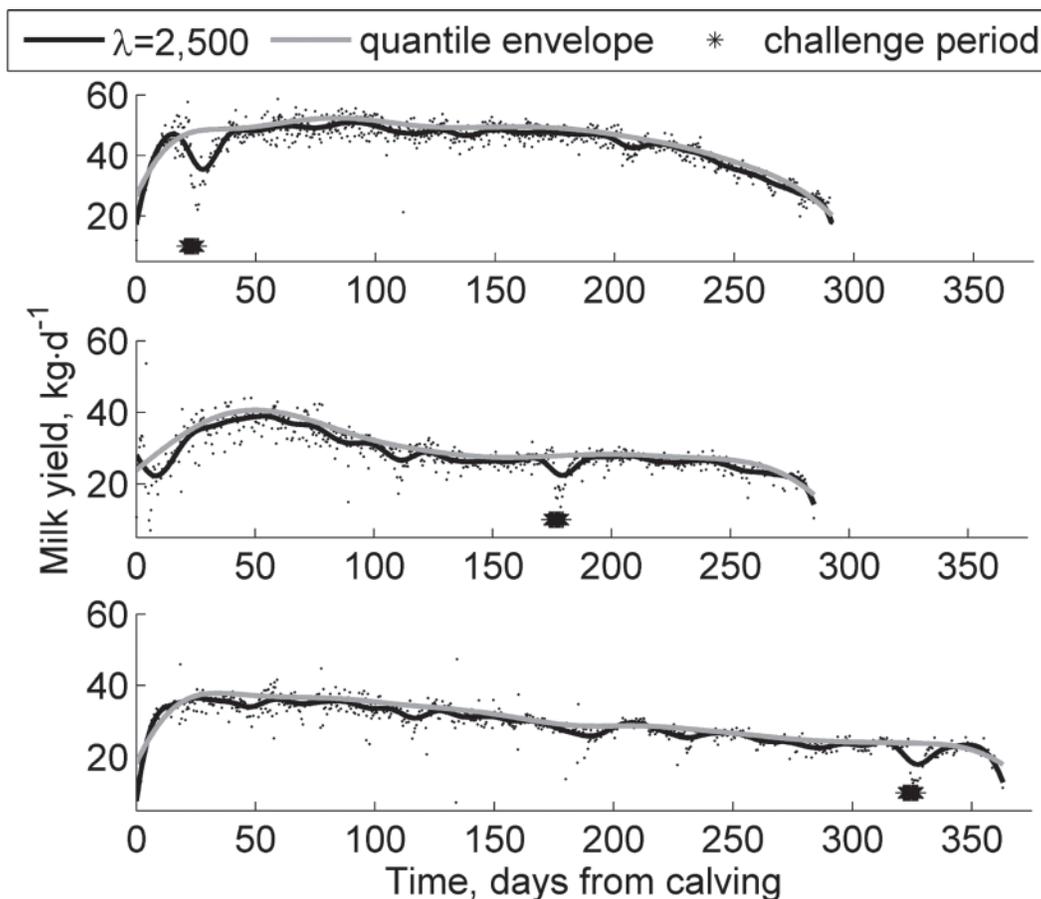


Figure 7. Examples of milk yield curves from 3 cows with the off-feed challenge at different stages of lactation. The roughly smoothed curves clearly capture the decline in milk yield during the challenge period for each cow. Other deviations in milk yield can be observed throughout lactations.

a cow can reach the maximum recovery rate (because V_{\max} is at I_{post}). The second derivative can also be used in a similar way to extract features from the rates of response (Figure 5). Table 1 gives a summary of the profile features of deviations in milk yield provoked by the nutritional challenge in the 47 cows. The absolute values of the rates of changes are shown (e.g., although V_{\min} is numerically negative, its absolute value denotes the rate of decrease, just like V_{\max} denotes the rate of increase).

Such a set of features can be seen as a panel of candidates for quantifying aspects of the ability of the cow to respond (i.e., her robustness; Friggens et al., 2010). Considering the post-trough recovery period, one can use these features to ask questions about how different animals recover. For example, the ability of the animal to respond could be modeled using the analogy of a spring or elastic band where the larger the deviation, the stronger the force of recovery to the original state. This model would suppose a positive correlation between the size of the decline in milk yield ($Q - T$), and the rate of recovery, V_{\max} . In contrast, it has been argued that excessive rates of change in milk yield may be a physiological stress for the animal, and thus, animals may seek a constant rate of recovery independent of the

size of the milk-yield loss. In this study, the correlation between V_{\max} and $(Q - T)$ was $r = 0.94$ ($P < 0.01$). Further, there was no significant slope ($P > 0.1$) to the relationship between $V_{\max}/(Q - T)$ and days from calving, indicating that the force of recovery was unaffected by stage of lactation. Thus, extracting response features provides a means to examine hypotheses concerning the nature of responses.

This approach can be further extended using multivariate statistical exploration of data to generate new hypotheses. For example, Figure 6 shows clear groups of features in a principal components analysis. Along the first principal component (**PC1**), the timing features Δt , which largely capture the shape of the deviation profiles, are clearly distinct from the absolute milk levels (P_{pre} , I_{pre} , T , Q , I_{post} , and P_{post}) and from the derivative features (V_{\min} , V_{\max} , A_{\max} , $A_{\min\text{pre}}$, and $A_{\min\text{post}}$). The latter 2 groups are well distinguished in the second principal component (**PC2**), and not surprisingly, $(Q - T)$ is closer to the derivative than the absolute level features. By construction, $(Q - T)$ is the instantaneous difference between the produced and the phenotypic potential yield curve. Therefore, we can conclude that the proposed estimation of the phenotypic potential yield is a suitable baseline for measuring disturbances.

Effectiveness of Quantifying Deviations

One test of the differential smoothing approach, and the choice of parameters values, is to examine the extent to which the deviations in milk yield provoked by the feeding challenge can be distinguished from other deviations in the roughly smoothed ($\lambda = 2,500$) milk yield curve. Figure 7 illustrates such deviations in the milk yield curves from 3 different cows with the off-feed challenge at different stages of lactation. The roughly smoothed curves clearly indicate the decline in milk yield during the challenge period for each cow, and also exhibit other deviations with unknown cause or unrelated to the off-feed experiment. In all, 480 deviations were detected from the complete lactations of the 47 cows. As can be seen in Figure 8, the size of the unplanned deviations was largely independent of the P_{pre} . The average milk yield loss of these unplanned deviations was $2.73 \text{ kg}\cdot\text{d}^{-1}$ and the $\text{SD} = 1.54 \text{ kg}\cdot\text{d}^{-1}$. In contrast, the size of the deviations provoked by the feeding challenge was significantly related to P_{pre} ($r = 0.86$, $P < 0.01$). The residual SD of these feeding challenge-provoked deviations was $1.24 \text{ kg}\cdot\text{d}^{-1}$. Thus, with the parameter values indicated above, the differential smoothing method was clearly able to detect the feeding challenge-provoked deviations.

DISCUSSION

Smoothing has largely been used to reduce the variability or noise in data to make it easier to detect differences between treatments or other systematic factors by removing short-term disturbances from the data. Taking weekly averages of daily measures or using group averages are simple forms of smoothing, but far more complex methods, such as using random components in regressions on time (e.g., Ingvarlsen and Friggens, 2005) or using Kalman filters (Kalman, 1960), are becoming increasingly common.

In this study, we used smoothing to identify disturbances rather than eliminating them. We illustrated the approach on milk production records from a controlled “off-feed” experiment, and the results indicate that differential smoothing can be a useful tool for quantifying biological disturbances in animal performance. Further, biologically meaningful features can be extracted from the smoothed curves without the need for a priori assumptions about the shape of the lactation curve. As presented here, the method is particularly suited for off-line usage (when complete deviation profiles are available). It provides a means to quantify the consequences of disturbing factors, such as disease and feeding disturbances, as well as allowing hypotheses about the biological nature of the response to be explored. If the primary goal is to develop on-line monitoring systems to detect when the physiological balance of an animal is being disrupted and intervention by the farm manager may be needed, additional capabilities such as dynamic update of the time series and forecasting

are required. Forecasting can be simple (e.g., based on the difference between the latest measure and a rolling baseline; Chagunda et al., 2006), or more complex (e.g., using nonparametric smoothing with latent factor models; Shen, 2009), or using multivariate measures to quantify degree of infection (Højsgaard and Friggens, 2010).

The amount of time and money that a farmer should invest in treating or preventing a disease will depend, in part, on the economic losses that disease causes (i.e., the size and duration of the production disturbance; Fourichon et al., 2000). Further, if a method for quantifying the disturbance is developed using clinical disease cases, then it may be extended to identify subclinical cases, which are by definition more difficult to identify from classical symptoms alone (Sloth et al., 2003). In this context, capturing disturbances (i.e., deviations from the underlying long-term trends) has been shown to provide early identification of onset of diseases, such as ketosis (Nielsen et al., 2005) and mastitis (Chagunda et al., 2006; Kamphuis et al., 2008b).

The approach of using the biological information in short-term disturbances has been greatly facilitated by the increasing availability of reliable, frequently measured, time-series data. It is now common for milk yield to be measured at every milking. Another example is the use of weigh platforms in milking systems that provide BW records 2 or more times a day. With this type

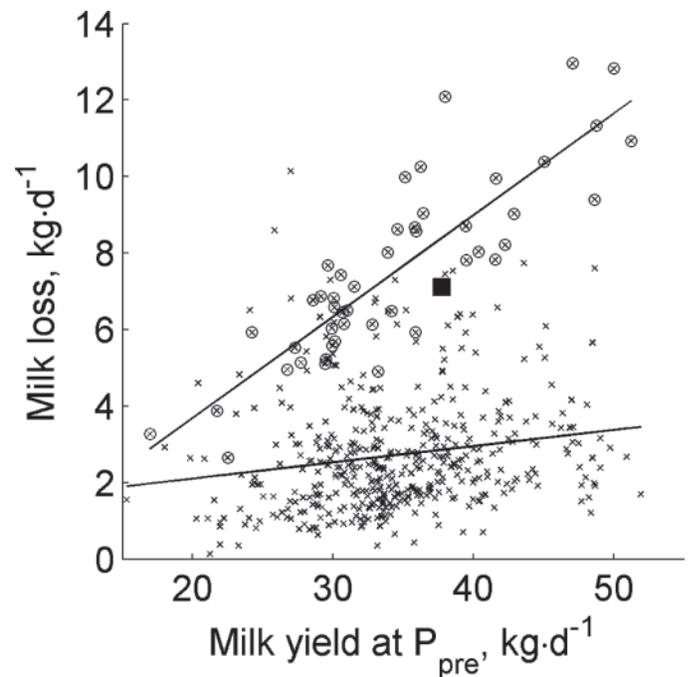


Figure 8. Milk loss [defined as the difference between the milk level at the trough (T) of the deviation profile and the corresponding potential level (Q) on the quantile envelope, which is a positively offset-smoothed curve that represents the undisturbed phenotypic potential yield] vs. the milk yield at the preceding peak (P_{pre} in Figure 5). Encircled cases are from the off-feed challenge experiment. The ■ is the first deviation in Figure 4 (at ~ 140 d from calving), which corresponds to a confirmed case of mastitis.

of technology and high measurement frequency, it becomes possible to distinguish biologically related disturbances from random measurement error.

Although demonstrated here using milk-yield data, the proposed approach is, in principle, generalizable to other biological time-series data. The increasing development of in-line measurement technology on the farm both affords the opportunity to develop such tools and presents the challenge of extracting the key information from such large quantities of data.

LITERATURE CITED

- Bareille, N., F. Beaudeau, S. Billon, A. Robert, and P. Faverdin. 2003. Effects of health disorders on feed intake and milk production in dairy cows. *Livest. Prod. Sci.* 83:53–62.
- Chagunda, M. G. G., N. C. Friggens, T. Larsen, and M. D. Rasmussen. 2006. A biological model for detecting individual cow mastitis risk based on lactate dehydrogenase: Model description. *J. Dairy Sci.* 89:2980–2998.
- Fourichon, C., H. Seegers, N. Bareille, and F. Beaudeau. 1999. Effects of disease on milk production in the dairy cow: A review. *Prev. Vet. Med.* 41:1–35.
- Fourichon, C., H. Seegers, and X. Malher. 2000. Effect of disease on reproduction in the dairy cow: A meta-analysis. *Theriogenology* 53:1729–1759.
- Friggens, N. C., M. G. G. Chagunda, M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen. 2007. Estimating degree of mastitis from time-series measurements in milk: A test of a model based on lactate dehydrogenase measurements. *J. Dairy Sci.* 90:5415–5427.
- Friggens, N. C., D. Sauvant, and O. Martin. 2010. Vers des définitions opérationnelles de la robustesse s'appuyant sur des faits biologiques: l'exemple de la nutrition. *INRA Prod. Anim.* 23:43–52.
- Højsgaard, S., and N. C. Friggens. 2010. Quantifying degree of mastitis from common trends in a panel of indicators for mastitis in dairy cows. *J. Dairy Sci.* 93:582–592.
- Ingvartsen, K. L., and N. C. Friggens. 2005. To what extent do variabilities in hormones, metabolites and energy intake explain variation in milk yield? *Domest. Anim. Endocrinol.* 29:294–304.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME. J. Basic Eng.* 82:35–45.
- Kamphuis, C., D. Pietersma, R. van der Tol, M. Wiedemann, and H. Hogeveen. 2008a. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying mastitis and abnormal milk. *Comp. Elec. Agric.* 62:169–181.
- Kamphuis, C., R. Sherlock, J. Jago, G. Mein, and H. Hogeveen. 2008b. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *J. Dairy Sci.* 91:4560–4570.
- Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, New York, NY.
- Nielsen, N. I., N. C. Friggens, M. G. G. Chagunda, and K. L. Ingvartsen. 2005. Predicting the risk of ketosis in dairy cows using on-line measurements of beta-hydroxybutyrate in milk: A biological model. *J. Dairy Sci.* 88:2441–2453.
- Ramsay, J. O., and B. W. Silverman. 2005. *Functional Data Analysis*. 1st ed. Springer, New York, NY.
- Shen, H. 2009. On modeling and forecasting time series of smooth curves. *Technometrics* 51:227–238.
- Sloth, K. H. M. N., N. C. Friggens, P. L. Løvendahl, P. H. Andersen, J. Jensen, and K. L. Ingvartsen. 2003. Potential for improving description of bovine udder health status by combined analysis of milk parameters. *J. Dairy Sci.* 86:1221–1232.

References

This article cites 14 articles, 0 of which you can access for free at:
<http://jas.fass.org/content/89/10/3089#BIBL>