A Model to Determine the Optimal Sampling Schedule of Diet Components¹

N. R. St-Pierre² and B. Cobanov

Department of Animal Sciences, The Ohio State University, 2029 Fyffe Rd., Columbus 43210

ABSTRACT

Various recommendations have been issued regarding the sampling schedule of diet components, especially forages. Their basis is unclear and none are justified from an economic standpoint. The objective of this research was to derive a general method for determining the optimal sampling design for forages. The process of forage removal from storage can be conceptualized as a quality control issue that can be monitored using a Shewhart X-bar chart. This procedure requires 3 control parameters: number of samples (n), sampling interval (h), and control limits (L). A quality cost function made of 4 parts is proposed: cost per cycle while the process is in-control (I); cost per cycle while the process is out-of-control (O); cost per cycle for sampling and analyses (A); and expected duration of a cycle (D). Thirteen inputs enter the cost function: the mean time that process is in control, the number of animals in the herd, the unit price of milk, the milk production loss due to white noise, the milk production loss from an abrupt change in forage composition, the time to sample and analyze one item, the expected time to discover the assignable cause, the expected time to fix the diet, the cost per false alarm, the cost to fix the diet, the fixed cost of sampling at each sampling time, the cost for each unit sampled, and the number of standard deviation slips when forage changes. The total quality cost per day $C = (I + O + A)/D$. The C function can be optimized with respect to n , h , and L to yield an optimal sampling schedule. Because n and h are discrete variables in a highly nonlinear function, parametric optimization algorithms cannot be used to optimize the function. A genetic algorithm was used for the minimization of C . Results showed that the optimal sampling designs are close to current practices in small herds of 50 cows, but very different in large

herds of 1,000 cows, resulting in reduced total quality costs of \$250/d. Total sensitivity of C was greatest for the number of cows in the herd, the shift in milk production when forage changes, the mean time that the process is in control, the price of milk, and the time to sample and analyze one item. Total sensitivities of n , h , and L were greatest for the mean time that the process is in control, the extent of the change in composition when there is a change in forage composition, the number of cows in the herd, the shift in milk production when forage changes, and the cost per unit sampled.

Key words: genetic algorithm, optimal sampling schedule, forage composition, sensitivity analysis

INTRODUCTION

In dairy production, as in most farm production systems, there are many processes that must be monitored to achieve maximum economic efficiency. Monitoring must be done because a given process should be producing expected outcomes while in-control, but it can shift at any point in time to an out-of-control state (Ishikawa, 1976). Feed production, including ingredient sourcing and purchasing as well as mixing and delivery to the animals, is a process that requires monitoring and intervention because feed costs are the major expense class in most US livestock production (St-Pierre et al., 2003) and because this process does go out-of-control periodically on most farms. Ingredient variation can be accounted for during diet formulation in instances in which the distribution of nutrients within feedstuffs is known at the time of formulation (Tozer, 2000). Alternatively, statistical process control methods could be implemented to monitor feed composition. In dairy production, forage represents generally over 50% of the DM consumed by lactating cows. Forage composition varies for most nutrients of economic importance (Table 15-1 of NRC, 2001; Sauvant et al., 2004). Variation in forage composition has been associated with significant potential production losses in dairy production (Fadel et al., 2006; St-Pierre and Weiss, 2006).

Received November 5, 2006.

Accepted August 31, 2007.

¹Salaries and research support were provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University. Manuscript No. 12-07AS.

²Corresponding author: st-pierre.8@osu.edu

Changes in forage composition on a given farm are essentially of 3 types. The first type is primarily associated with changes in measurements due to laboratory and method variations (analytical variance) as well as sampling and subsampling variances (St-Pierre and Weiss, 2006). These variations are of no consequence to animals because they reflect changes in apparent nutritional composition but not true composition. The second type reflects small changes in composition, essentially white noise, caused by a multitude of factors, each having a small effect. This is the variation that one would expect, for example, in the true composition of forage harvested from a single, uniform field, from a single forage variety or hybrid, on the same harvesting day (St-Pierre and Weiss, 2006). Little can be economically done to reduce these sources of variation. The last type of change is characterized by relatively large and sudden threshold changes in composition, such as could be expected when sourcing forages from different fields, varieties, species, harvesting dates, and so on (St-Pierre and Weiss, 2006). These changes can be managed by separation of inventories or by sampling for laboratory determination of chemical composition. In most instances, large storage structures such as those used for the storage of grass, legume, and grain silages do not allow for meaningful inventory separation, leaving periodic sampling and laboratory determinations as the predominant means of controlling the effect of forage variation in the diets.

The determination of an optimal feed sampling schedule in animal production is an unresolved problem and is receiving more attention because of its potential impact on animal performance and farm profitability. Control charts have long been used to monitor commercial production processes to detect special causes that produce changes in the output (Shewhart, 1931). They are slowly gaining acceptance in animal agriculture as ways to incorporate statistical process control to various animal production systems (DeVries et al., 1997; Lukas et al., 2005). Many charts have been proposed, such as the Shewhart (**X-bar**), exponentially weighed moving average, and cumulative sum charts. The basic X-bar control chart consists of sampling from a process over time and charting mean measurements at each sampling time (Drain, 1997). As long as the mean of measurements at each sampling time falls inside predetermined control limits (L), the process is considered in-control. Shewhart X-bar charts with “3-sigma” control limits [i.e., $L = 3$ standard deviations (**SD**) of the mean] are often used (Reynolds and Stoumbos, 2005). In most agricultural systems, it is customary to set the control limits at 2 SD of the mean of a set of measurements taken at the same time, traditionally called a subgroup. This is in

contrast to industrial production, in which control limits are generally set at 3 SD (DeVries et al., 1997; Drain, 1997). This practice of using 2 SD, however, does not consider the respective size of type I (i.e., concluding that the process is out-of-control when it is in-control; the false alarms) and type II errors (i.e., concluding that the process is in-control when in fact it has shifted to an out-of-control state), as well as all costs incurred. The objectives of this research were to derive a general methodology to determine optimum sample size (n), sampling interval (h), and location of L for an X-bar chart used to monitor forage or feed composition and to perform a sensitivity analysis to parameter values to identify those parameters requiring precise estimates.

MATERIALS AND METHODS

Model Derivation

The monitoring of forage quality can be viewed as a stochastic process; the question becomes which stochastic model seems most reasonable. There are 3 basic types of models for stochastic processes: Gaussian processes and time series, Poisson and renewal processes, and Markov processes (Parzen, 1999). Renewal reward models are of the second type and are characterized by memoryless and nonself recovery attributes. In the case of quality monitoring of forage, the first characteristic implies that the forage does not “remember” its last change in composition. This characteristic appears reasonable in our case. The second characteristic implies that a diet unbalanced by a permanent change in forage composition will not readjust itself to a state of balance; human intervention is required. Forage composition is never permanently changed. But the interval between major compositional changes is generally much greater than the time to sample, analyze, and chart the results (i.e., the minimum time to intervention). As long as the potential frequency of sampling remains far greater than the frequency of major abrupt changes, considering the process as a renewal reward process approximates closely the true minimal quality costs (Knappenberger and Grandage, 1969). Thus, the renewal process should be a good approximation to the true but unknown stochastic process underlying forage quality. There are instances, however, in which the frequency of change in composition approaches that of minimum sampling interval, such as DM composition in horizontal silage structures. In those instances, the process does not conform to a renewal reward process and other means of control need to be investigated. The approach used in our derivation is based on the general approach used by Lorenzen and Vance (1986) of a total quality cost function

Figure 1. Graphical representation of a quality cycle: $1/\lambda$ = mean time process is in-control, τ = expected time of occurrence of abrupt change after last sampling while in control, T_c = expected time between the occurrence of abrupt change and the next sampling time, h = sampling interval, T_e = expected time until an out-of-control signal occurs, T_1 = expected time to investigate the cause of the change, T_2 = expected time to reformulate and implement new diets, and T_f = expected time to investigate the cause of the change, reformulate, and implement new diets.

for renewal reward processes. To optimize the quality control procedure, and thus the sampling schedule, we must derive the total cost of quality as a function of the number of samples taken at each sampling time, the sampling interval (i.e., the frequency of sampling), and the control limits. To do so, a quality cycle is defined as the time between the start of successive in-control periods (Figure 1). In our application, an in-control period is defined as when the forage has a mean nutritional composition equal to that assumed during diet formulation. The costs incurred during the in-control period are 1) costs from sampling the process, 2) costs due to white-noise variation in forage composition, and 3) costs of false alarms, which are the costs due to a type I error (i.e., when the control chart erroneously signals an out-of-control state). We assume that when the process goes out-of-control, it shifts abruptly to a different state, meaning that the change in forage composition is sudden. The importance of this assumption will be assessed in the companion paper (St-Pierre and Cobanov, 2007). As discussed previously, we also assume that the process cannot return to an in-control state without intervention. The costs incurred during the out-of-control period; that is, when forage composition has changed due to a major assignable cause but the diet has not yet been rebalanced are 1) costs due to sampling, 2) costs due to lost milk production if forage quality dropped, or a lost opportunity to reduce

feed costs if forage quality improved, and 3) costs due to searching for the cause, repairing the system (diet reformulation, changes in purchased feeds, etc.), and possible downtime in the case of a general renewal reward process. This last item is not relevant in dairy feeding systems (i.e., animals are still being fed while diets are being reformulated) but is included here for general application to any renewal reward processes. The costs due to lost milk production and increased feed costs are the direct result of a type II error (i.e., when the control chart erroneously signals an in-control state when the forage composition has, in fact, changed). The total quality cost per day is calculated as the total cost per cycle divided by the cycle length calculated in days.

Cycle Length

First, let us consider the total length of a cycle. It is the sum of the following factors.

Time Until Occurrence of an Abrupt Change in Forage Composition. Given that we assumed a memoryless process subject to abrupt changes, the in-control period is distributed as a negative exponential random variable with mean $1/\lambda$ (Ryan, 2000; Yashin, 1993). If production continues during searches for an assignable cause, the average time for occurrence of the assignable cause is simply $1/\lambda$. Production always

continues during searches in the context of monitoring forage quality for milk production. But production could cease in other future applications of quality costs minimization of a renewal reward process. In such instances, the function becomes more complicated and is derived here for future applications. Let T_0 be the expected search time (d) for a false alarm. Then, the expected time spent searching during false alarms is T_0 times the expected number of false alarms:

$$T_a = T_0 \frac{s}{ARL_0} \quad [1]$$

where T_a = time spent searching during false alarms (d), T_0 = expected search time for a false alarm (d), s = expected number of samples taken while in-control, and ARL_0 = average run length while in-control (d).

A few steps of algebraic derivations based on a renewal reward Poisson process (Parzen, 1999), and hence a negative exponential distribution of ARL_0 , show that:

$$s = \frac{e^{-\lambda h}}{1 - e^{-\lambda h}} \quad [2]$$

where e = the exponential function, λ = reciprocal of the mean length of the in-control period (1/d), and h = sampling interval; that is, the time interval between 2 successive sampling occurrences (d).

In [2], λ is a characteristic of the harvesting and storage system, independent of any control chart.

Also:

$$ARL_0 = \frac{1}{\alpha} \quad [3]$$

where $\alpha = P(\text{out-of-control signal} \mid \text{process is in-control})$; that is, the probability of a type I error with P symbolizing “the probability of,” and $\mid$ symbolizing “given.” Lorenzen and Vance (1986) suggested a way of combining both conditions. Let $\delta_1 = 1$ if production continues during searches, and $\delta_1 = 0$ if production ceases during searches. Then,

$$T_b = \frac{1}{\lambda} + \frac{sT_0(1 - \delta_1)}{ARL_0} \quad [4]$$

where T_b = expected time until the change in composition occurs (d), and δ_1 = dummy binary variable depending on production during searches. T_b is equal to $1/\lambda$ in the case of forage composition monitoring because production continues even when diets are being reformulated.

Time Until the Next Sample is Taken. Let τ be the expected time (d) of occurrence of the abrupt change in forage composition, given that it occurs between the i th and $(i + 1)$ th samples; that is, the time between the last sample while in-control and the abrupt change in forage composition. By derivation, we get:

$$\tau = \frac{1 - (1 - \lambda h)e^{-\lambda h}}{\lambda(1 - e^{-\lambda h})} \quad [5]$$

with all symbols as previously defined. The expected time (d) between the occurrences of the abrupt change in forage composition and the next sampling time (T_c) equals

$$T_c = h - \tau. \quad [6]$$

Time to Analyze the Sample and Chart the Result. Let E be the expected time (d) to sample, analyze, and chart the results of one sampling time. Note that E can be considered equal to 0 if the measurements are made on the farm as in the determination of silage DM concentration (the measurements are quite fast, and usually performed on site) but can be significant if samples are sent to an outside laboratory for extensive chemical analyses. For a sample of n items, the time (d) to sample, analyze the sample and chart the result (T_d) is given by:

$$T_d = nE \quad [7]$$

if T_d is proportional to the number of samples taken, and

$$T_d = E \quad [8]$$

if T_d is a constant, independent of the number of samples. In general, [8] is the norm in forage analyses and will be retained for the remainder of this paper, remembering that [7] could be substituted in the derivation when applicable.

Time Until the Chart Gives an Out-of-Control Signal. The expected time (d) until an out-of-control signal occurs (T_e) is given by

$$T_e = h(ARL_1 - 1) \quad [9]$$

where ARL_1 = the average run length when the process has shifted to an out-of-control state (d).

If the sample statistics are independent (i.e., samples taken at a given sampling time are true replicates and not subsamples), then

$$ARL_1 = \frac{1}{1 - \beta} \quad [10]$$

where $\beta = P(\text{in-control signal} \mid \text{process is out-of-control})$; that is, the probability of a type II error.

Note that ARL_1 depends on the underlying distribution, the control limit L , the sample size n , and the extent of the change in composition (Δ) when there is an abrupt change in forage composition. It is important to understand that when forage quality has changed, it is not necessarily detected (signaled) the next time samples are taken, leading to type II errors.

Time to Find the Cause and Repair the Process.

Let T_1 be the expected time (d) to investigate the cause of the change, and T_2 be the expected time (d) to reformulate and implement new diets (equivalent of repairing the machinery or process in industrial production). Then,

$$T_f = T_1 + T_2 \quad [11]$$

where T_f = the expected time to investigate the cause of a composition change and reformulate and implement new diets (d).

Total Cycle Length. The total cycle length T_{CL} is simply calculated as the sum of its elements:

$$T_{CL} = T_b + T_c + T_d + T_e + T_f \quad [12]$$

When fully expanded, [12] yields

$$T_{CL} = \frac{1}{\lambda} + \frac{sT_0(1 - \delta_1)}{ARL_0} - \tau + E + hARL_1 + T_1 + T_2. \quad [13]$$

Costs per Cycle

The costs per cycle are incurred from loss of production while in-control due to white noise in forage composition, as well as loss of production when forage quality declines and increased feed costs when forage quality increases while out-of-control (i.e., when forage composition has changed abruptly), for false alarms, for location and repair of the assignable cause (i.e., identifying remedies, and reformulating and implementing new diets), and for sampling and analyzing forages.

Cost per Cycle Due To Reduced Production. Let C_0 and C_1 be the costs per day due to lost production from white noise in forage composition while the process is in-control, and to lost production or increased feed costs while the process is out-of-control, respec-

tively. Then, the expected cost per cycle due to reduced production (C_a) equals

$$C_a = \frac{C_0}{\lambda} + C_1(E - \tau + hARL_1 + \delta_1 T_1 + \delta_2 T_2) \quad [14]$$

where C_a = cost per cycle due to reduced production (\$), C_0 = cost per day from reduced production while in-control (\$/d), C_1 = cost per day from reduced production while out-of-control (\$/d), and δ_2 = dummy binary variable that depends on state of production during repair ($\delta_2 = 1$ if production continues during repair, and $\delta_2 = 0$ if production ceases).

In dairy production,

$$C_0 = \Delta_{M0} P_m N_c \quad [15]$$

where Δ_{M0} = milk production loss per cow and per day due to white noise in forage composition when the process is in-control (kg/d per cow), P_m = unit price of milk (\$/kg), and N_c = number of lactating cows in the herd.

Also,

$$C_1 = \Delta_{M1} P_m N_c \quad [16]$$

where Δ_{M1} = milk production loss per cow and per day due to an abrupt change in forage composition when the process is out-of-control (kg/d per cow).

Cost per Cycle Due To False Alarms, Identifying the Cause, and Repairing. Let Y be the cost per false alarm. This includes the cost of searching and testing for the cause of the potential forage change when applicable, plus the cost of downtime if production ceases during the search. Also, let W be the cost to investigate the cause of the change and to reformulate and implement new diets when an abrupt change in forage composition occurs. Then, the expected cost for false alarms and for identifying the cause of the change and repairing the system when out-of-control (new diet formulation and implementation) is:

$$C_b = \frac{sY}{ARL_0} + W \quad [17]$$

Cost per Cycle for Sampling and Analyses. Let a be the fixed cost per sampling time (depending on the conditions, this could be the cost of mailing samples or the labor to set up the sampling and sample preparation) and b the cost per unit sampled (e.g., cost of laboratory analyses). The sampling and analyses costs at each sampling time (C_c) are then

$$C_c = a + bn. \quad [18]$$

The expected cost for sampling and analyses of samples is given by $(a + bn) \times (\text{days of production in a cycle}) / h$. Thus, using [13], the expected cost per cycle for sampling (C_d) equals

$$C_d = \frac{(a + bn) \left(\frac{1}{\lambda} - \tau + E + hARL_1 + \delta_1 T_1 + \delta_2 T_2 \right)}{h} \quad [19]$$

Total Quality Cost per Cycle. Combining [14], [17], and [19], we get the total quality cost per cycle (C_{CL}) as

$$C_{CL} = C_a + C_b + C_d. \quad [20]$$

When expanded, [20] yields:

$$C_{CL} = \frac{C_0}{\lambda} + C_1(E - \tau + hARL_1 + \delta_1 T_1 + \delta_2 T_2) + \frac{sY}{ARL_0} + W + \frac{(a + bn) \left(\frac{1}{\lambda} - \tau + E + hARL_1 + \delta_1 T_1 + \delta_2 T_2 \right)}{h} \quad [21]$$

Total Quality Cost per Day. Because cycle lengths are variable and are functions of h , n , and L , we must express the cost function per unit of time (days), not per cycle. This is accomplished by dividing equation [21] by equation [13] to yield the total quality cost per day, C :

$$C = \frac{\frac{C_0}{\lambda} + C_1(E - \tau + hARL_1 + \delta_1 T_1 + \delta_2 T_2) + \frac{sY}{ARL_0} + W}{\frac{1}{\lambda} + \frac{sT_0(1 - \delta_1)}{ARL_0} - \tau + E + hARL_1 + T_1 + T_2} + \frac{(a + bn) \left(\frac{1}{\lambda} - \tau + E + hARL_1 + \delta_1 T_1 + \delta_2 T_2 \right)}{\left(\frac{1}{\lambda} + \frac{sT_0(1 - \delta_1)}{ARL_0} - \tau + E + hARL_1 + T_1 + T_2 \right) h} \quad [22]$$

All we are left with is the derivation of ARL_0 and ARL_1 in equation [22]. For an X-bar chart, it is assumed that the observations are independent and identically distributed from a normal distribution with mean equal to the center line (CL) of the chart and standard deviation σ . The independence assumption requires that the samples taken at a given sampling time are true replicates and not subsamples. Thus,

from the mean (X) of a set of measurements taken at the same time we have:

$$\alpha = P\left(X - CL > \frac{L\sigma}{\sqrt{n}}\right) + P\left(X - CL < \frac{-L\sigma}{\sqrt{n}}\right) = 2\phi(-L) \quad [23]$$

where ϕ = is the cumulative distribution function for the standard normal distribution.

When the process is out-of-control, the observations are independent and identically distributed from a normal distribution with mean $CL + \Delta\sigma$, so that

$$\beta = P\left(\frac{-L\sigma}{\sqrt{n}} < X - CL < \frac{L\sigma}{\sqrt{n}}\right) \quad [24]$$

which equates to:

$$\beta = \phi(L - \Delta\sqrt{n}) - \phi(-L - \Delta\sqrt{n}).$$

Substitution of $1/\alpha$ for ARL_0 and $1/(1 - \beta)$ for ARL_1 in [22] in combination with [2], [5], [15], and [16] gives the quality cost per day for any process to be monitored by an X-bar chart.

Minimizing the Total Quality Costs per Day Function

The total quality cost per day function is given by [2], [5], [15], [16], [22], [23], and [24] and is a function of 3 quality control variables: the sample size (n), the sampling interval (h), and the control limit (L). Function [22] with $1/\alpha$ substituting for ARL_0 and $1/(1 - \beta)$ for ARL_1 as per [23] and [24] is highly nonlinear in each of the 3 parameters. Therefore, nonlinear optimization techniques must be used to identify the optimum n , h , and L leading to the minimum cost. In our application the number of samples n must clearly be considered an integer variable. Additionally, the sampling interval h should likely be considered an integer variable in field applications (i.e., samples can only be shipped once a day to an outside laboratory). The resulting mixed-integer nonlinear optimization problem was solved using the genetic algorithm, a nonparametric optimization algorithm (Goldberg, 1989). The genetic algorithm (Haupt and Haupt, 1998) was programmed using the Borland Delphi 2005 development environment with the following structure: initial population size = 4,000, working population size = 1,000,

Table 1. Input variables to the total quality cost model

Input variable	Description	Central values	Sensitivity analysis		
			Distribution	Lower bound	Upper bound
$1/\lambda$	Mean time process is in control (d)	30	Uniform	7	60
Δ	Number of standard deviations slip when process moves out-of-control (abrupt change in forage composition)	1.5	Uniform	0.75	3.00
T_0	Expected search time for false alarm (d)	2	Uniform	0.50	7.00
T_2	Expected time to repair the process (i.e., to rebalance and implement new rations) (d)	2	Uniform	0.50	7.00
E	Time to sample and analyze all items (i.e., time awaiting for lab results) (d)	4	Uniform	0.50	14.00
N_c	Number of cows in the herd	100	Uniform	50	1,000
Δ_{m0}	Milk production loss due to white noise change in forage composition while the process is in-control (kg/cow per d)	0.45	Uniform	0.22	0.91
Δ_{m1}	Milk production loss when process goes out-of-control (abrupt change in forage composition) (kg/cow per d)	2.27	Uniform	0.91	4.55
P_m	Price of milk (\$/kg)	0.286	Uniform	0.22	0.374
Y	Cost per false alarm (\$)	200	Uniform	50	1,000
W	Cost to locate and repair the assignable cause (i.e., to rebalance and implement new rations) (\$)	200	Uniform	50	1,000
a	Fixed cost per sampling time (preparation and mailing of samples) (\$)	15	Uniform	5	25
b	Cost per unit sampled (\$)	25	Uniform	10	125

mutation rate = 12%, elite member rate = 0.5%, and number of iterations = 60.

Sensitivity Analysis

Equation [22] contains 16 input variables and 3 optimization variables (n , h , and L). Although we could not find specific recommended sampling designs for forages in the scientific literature, it is common on many US commercial dairy farms to take one sample of each forage once a month and to use control limits of 2 SD or 0 SD (i.e., always adjust the diet based on the most recent forage analysis). Thus, an industry standard was used for comparison with $n^- = 1$, $h^- = 30$ d, and $L^- = 2$ SD (the $\sim$ symbol signifies that these are the values of n , h , and L for the industry standard). Of the 16 input variables to the total quality cost function in [22] (an equation that is applicable to any renewal reward process), 13 of them could vary in our specific application, as two of the input variables (δ_1 and δ_2) have a fixed value of 1, and a third (T_1) has a fixed value of 0. A set of default central values for all the resulting 13 input variables were selected as a starting basis for the sensitivity analyses (Table 1). This set of default values will be referred to as the set of central values. These were essentially based on expert opinion. Sensitivity analyses should indicate which input variables need to be estimated precisely in model application. Three types of sensitivity analyses were performed: a univariate sensitivity analysis, a global linear sensitivity analysis, and a global nonlinear sensitivity analysis (Saltelli et al., 2004). For all

3 types, sensitivity analyses were done for total quality cost per day (C), and for each of the 3 variables forming the sampling design (n , h , and L).

Univariate Sensitivity Analysis. A conventional sensitivity analysis was performed by iteratively varying each input variable from its minimum to its maximum value (Table 1), while keeping all other input variables at their central values. The calculated univariate sensitivity indices (S_L) express the average percentage change in C , n , h , or L from a 1% change in each of the input variables while keeping all others constant at their central values.

Global Linear Sensitivity Analysis. A global linear sensitivity analysis based on standardized regression coefficients (**SRC**) was performed according to Saltelli et al. (2004) using SimLab V1.1 (Giglioli and Saltelli, 2006). The SRC were fitted by least squares using the GLM procedure of SAS (SAS Institute, 2004).

If the true model is linear, the SRC coefficients determine the fraction of the total variation caused by each of the parameters. In such an instance, the total variance is completely decomposed through SRC coefficients and the sum of the SRC squared equals 1. If the true model is not linear, the SRC coefficients decompose the variance for the linear effect of each parameter, but ignore the contribution from the nonlinear and interaction effects. The variance decomposition attributed to the nonlinear part of the model remains unknown. The validity of the SRC as a measure of sensitivity is conditional on the degree to which the regression model fits the data; that is, to the coefficient of model determination R_y^2 (Saltelli et al., 2004).

Large values of R_y^2 indicate near linearity of the model and the calculated SRC can be used to assess the contribution of the individual input variables to the uncertainty in model prediction. Conversely, small values of R_y^2 indicate strong nonlinear sensitivity behavior, and consequently higher order methods must be used for sensitivity analyses. Whereas the univariate sensitivity analysis assesses sensitivity in only one dimension, SRC offers a measure of sensitivity that is multidimensionally averaged, meaning that the SRC coefficients provide a measure of variance contribution for an input variable that is averaged over all possible values of the remaining input variables.

Global Nonlinear Sensitivity Analysis. The extended Fourier amplitude sensitivity analysis (e-FAST) is a variance-based technique that estimates the total fractional contribution of each input factor X_i to the total variance of the model output Y (Saltelli et al., 1999; Giglioli and Saltelli, 2006). Furthermore, e-FAST allows the simultaneous evaluation of the first (S_i) and total effect (S_{Ti}) indices. The estimation of the pair (S_i , S_{Ti}) is important to appreciate the difference between the impact of factor X_i alone on Y (that is, S_i) and the overall impact of factor X_i through interactions with the others on Y (provided by S_{Ti}). All e-FAST calculations were made using SimLab V1.1 (Giglioli and Saltelli, 2006).

RESULTS AND DISCUSSION

Optimal Sampling Schedule with the Set of Central Input Variables

The optimum design when all input variables are set at their central values is: $n^* = 2$, $h^* = 15$ d, and $L^* = 1.098$ SD (the * symbol signifies that these are the optimum values of n , h , and L). This is very different from the industry standard design schedule of taking 1 sample every 30 d, and to consider the process out of control if the result deviates by 2 SD from the mean; that is, $n^- = 1$, $h^- = 30$, and $L^- = 2$. The cost implication of these differences are discussed in the next section in conjunction with the effect of herd size.

Univariate Sensitivity Analysis

Figure 2 shows the univariate sensitivity of the sampling design parameters and C to herd size (N_c). The optimal number of samples to be taken at each sampling time is equal to 2 except for the smallest herd size investigated ($N_c = 50$). The optimal sampling interval follows an exponential decline as herd size increases, reaching an asymptote of 4 d when N_c exceeds 670 cows. For all but the smallest herd size, the optimal

control limit L exceeds 1 and fluctuates between 1.1 and 1.4. In Figure 2, total quality cost C is shown both for the optimal sampling design as well as for the industry standard design. Both costs are nearly identical at a herd size of 50 cows, but the optimal design reduces total quality cost by nearly \$250/d at a herd size of 1,000 cows. At an average daily production of 30 kg/d, this difference equates to net savings of \$0.83/100 kg of milk. This figure corresponds to 19 to 37% of the average long-term net farm income per cow per day in the United States (St-Pierre et al., 2000). It is of interest to note that the industry standard design ($n^- = 1$, $h^- = 30$, and $L^- = 2$) is relatively close to the optimum design for a herd of 50 cows, except for the control limits ($n^* = 1$, $h^* = 27$, and $L^* = 0.6$). Thus, the practice of taking a forage sample once a month probably served the industry well when herd sizes were smaller, but its use negatively impacts large herds which are becoming the norm in many parts of the United States.

Table 2 reports the results of the univariate sensitivity analyses for C , L , n , and h for all input variables. The linear sensitivity indices $S_L(C)$, $S_L(L)$, $S_L(n)$, and $S_L(h)$ represent the average percentage changes in C , L , n , and h from a 1% change in each of the 13 input variables while maintaining all others constant. Using this criterion, total quality costs are most sensitive to N_c , Δ_{m1} , and $1/\lambda$ than to all other input variables to the model. For these 3 input variables, a 1% change results in a 0.5 to 1.39% change in P . For all other input variables, C changes by less than 0.33% for each 1% change in the variable. The control limit L is most sensitive to $1/\lambda$, Δ , b , Δ_{m1} , and Y . In all instances, a 1% change in these input variables results in a greater than 1% change in L . The number of samples per sampling time n is most sensitive to Y , b , $1/\lambda$, Δ , N_c , Δ_{m1} , and W . A 1% change in these input variables results in a 0.5 to 1.0% change in n . Finally, h is most sensitive to Δ_{m1} , N_c , b , Δ , and W , with a 1% change in their input values resulting in a 0.53 to 1.26% change in h . These results can only be accurate at the central values of the input variables. This is a severe limitation to the traditional univariate sensitivity analysis. In addition, they do not account for nonlinear effects as well as interaction effects between input variables. In the presence of such effects, the S_L can be grossly inaccurate in assessing sensitivity to inputs (Saltelli et al., 2004).

Global Linear Sensitivity Analysis

Table 3 reports the SRC coefficients and descriptive statistics for all parameters of the model describing sensitivity of C to input variables. Based on the SRC

Figure 2. Effect of number of cows in the herd (N_c) on optimal sampling design and total quality cost per day; — = the sampling interval (h); —△— = number of standard deviations used as control limits of the X-bar chart (L); —◇— = number of samples taken (n); —■— = the total daily quality cost (C); and —○— = the total daily quality cost from current sampling design in the dairy industry ($h = 30$ d, $n = 1$, $L = 2$ SD).

decomposition, C is most sensitive to $1/\lambda$, Δ , N_c , and Δ_{m1} . The R_y^2 of 0.75 indicates that the contribution of most variables to the total variance is probably reasonably well approximated solely by its linear effect. In contrast, the R_y^2 for the SRC decomposition of the sensitivity of L , h , and n to input variables were all relatively small ($R_y^2 < 0.5$, results not shown) indicating the necessity of factoring nonlinear and interaction effects in the analysis.

Global Nonlinear Sensitivity Analysis

Results from the e-FAST first-order sensitivity indices were generally in line with the results obtained with the SRC: sensitivity was predominantly the result of N_c , Δ_{m1} , and $1/\lambda$ (results not shown). Recall that the difference between S_{Ti} and S_i represents the overall impact of input variable i through interactions with other input variables. A useful way of expressing the importance of the interaction effects is to measure the proportion of S_{Ti} that is accounted by S_i (Table 4). For C , the sensitivity to input variables is predominantly from a direct effect of each input variable and not from their interactions with each other. These results confirm those obtained using SRC. For L , n , and h , sensitivity is predominantly from interaction effects. For these 3 design variables, it is thus important that total sensitivity to inputs be measured using total sensitivity indices calculated by a global nonlinear

method and that comparisons across input variables be on that basis. Therefore, univariate S_L indices and global linear SRC indices are not providing an accurate assessment of the true sensitivity of L , n , and h to input variables. Table 5 expresses the proportion of the sum of all e-FAST total sensitivity indices that is accounted by for each individual total sensitivity index. This greatly facilitates the identification of input variables for which a design parameter has large total sensitivity. For C , it is clear that nearly all of the total sensitivity is the result of N_c , Δ_{m1} , $1/\lambda$, P_m , and E . These input variables account for more than 98% of the total sensitivity of C to inputs. For L , 57.4% of the total sensitivity is the result of N_c , Δ_{m1} , Δ , $1/\lambda$, and W . For n , 76.3% of the total sensitivity is the result of b , Δ , Y , $1/\lambda$, and Δ_{m1} . Finally for h , 66.9% of the total sensitivity is the result of N_c , Δ_{m1} , $1/\lambda$, W , and Δ_{m0} . These results indicate that the variables with a large influence on total quality costs are not the same as those with a large influence on the parameters of the sampling design. The 3 design parameters and C are all sensitive to $1/\lambda$ and Δ_{m1} , and none of them are particularly sensitive to T_0 , T_2 , and a . Total quality costs C is very sensitive, whereas L and h are somewhat sensitive to N_c . Both L and n are sensitive to Δ , but not C and h . Finally, E , Δ_{m0} , P_m , Y , W , and b result in important total sensitivity for only 1 of the 4 variables. Therefore, precise and accurate estimates of N_c , Δ_{m1} , $1/\lambda$, P_m , and E are required for accurate

Table 2. Linear sensitivity analysis for total quality costs per day (C), control limits (L), number of samples (n), and sampling interval (h)

Input variable ¹	Min/max values	C (\$/d)	L (SD units)	n	H (d)	$S_L(C)^2$	$S_L(L)$	$S_L(n)$	$S_L(h)$
$1/\lambda$	7	61.70	0.01	1	19	0.50	1.65	0.56	0.09
	60	33.11	1.22	2	17				
Δ	0.75	45.37	0.00	2	23	0.15	1.66	0.56	0.66
	3	37.91	1.81	1	10				
T_0	0.5	42.09	1.10	2	15	0.00	0.00	0.00	0.00
	7	42.09	1.10	2	15				
T_2	0.5	41.18	1.14	2	14	0.07	0.08	0.00	0.16
	7	44.67	1.04	2	17				
E	0.5	39.85	1.12	2	14	0.13	0.12	0.00	0.25
	14	46.72	0.98	2	19				
N_c	50	25.21	0.62	1	26	1.39	0.50	0.56	1.22
	1,000	278.77	1.17	2	4				
Δ_{m0}	0.22	37.88	1.10	2	14	0.23	0.01	0.00	0.21
	0.91	49.98	1.08	2	18				
Δ_{m1}	0.91	24.66	0.08	1	64	0.72	1.46	0.56	1.26
	4.55	62.11	1.15	2					
P_m	0.220	34.61	1.08	2	18	0.32	0.06	0.00	0.33
	0.374	51.28	1.16	2	12				
Y	50	39.16	0.24	1	12	0.11	1.29	1.00	0.38
	1,000	44.88	1.91	4	19				
W	50	38.86	1.11	2	14	0.29	0.49	0.56	0.53
	1,000	58.06	0.60	1	27				
a	5	41.38	1.21	2	13	0.03	0.12	0.00	0.17
	25	42.73	1.05	2	16				
b	10	39.25	1.56	3	11	0.16	1.56	0.83	0.73
	125	47.27	0.05	1	28				

¹ $1/\lambda$ = mean time process is in control (d); Δ = number of standard deviations slip when process moves out-of-control (abrupt change in forage composition); T_0 = expected search time for false alarm (d); T_2 = expected time to repair the process (i.e., to rebalance and implement new rations) (d); E = time to sample and analyze all items (e.g., time awaiting for lab results) (d); N_c = number of cows in the herd; Δ_{m0} = milk production loss due to white noise change in forage composition while the process is in-control (kg/cow per d); Δ_{m1} = milk production loss when process goes out-of-control (abrupt change in forage composition) (kg/cow per d); P_m = price of milk (\$/kg); Y = cost per false alarm (\$); W = cost to locate and repair the assignable cause (i.e., to rebalance and implement new rations) (\$); a = fixed cost per sampling time (preparation and mailing of samples) (\$); b = cost per unit sampled (\$).

² $S_L(C)$ = average percentage change in C from a 1% change in the input variable; $S_L(L)$ = average percentage change in L from a 1% change in the input variable; $S_L(n)$ = average percentage change in n from a 1% change in the input variable, and $S_L(h)$ = average percentage change in h from a 1% change in the input variable.

estimation of total quality costs, whereas precise and accurate estimates of $1/\lambda$, Δ , N_c , Δ_{m1} , and b are required for a reasonably accurate identification of the optimum sampling design. Under field conditions, reasonably precise estimates of N_c , P_m , and E should be relatively easy to obtain. Precise estimate of Δ would require prior monitoring records of forage quality on individual farms. This would prove to be more difficult in many instances. A precise estimate of Δ_{m1} is problematic. One could possibly use a nutritional model such as NRC (2001) to estimate expected production changes from shifts in forage composition for each of the nutrients being monitored by their individual control chart. This would naively transform a nutritional requirement system into a nutrient response system, which has been shown to be inaccurate (St-Pierre and Thraen, 1999). In addition, the best strategy to moni-

tor multiple nutrients is unclear. Each nutrient could have its own control chart and its own sampling design but this may require a global sampling schedule, which is not acceptable in practice. Clearly, additional work is needed in this area.

CONCLUSIONS

A model with 16 input variables and 3 design parameters can be used to calculate the total quality cost per unit of time of any renewal reward process based on X-bar quality charts. When applied to the control of forage variation, the model reduces to 13 input variables and 3 design parameters. The current practice of taking one forage sample per month and intervening when the results differ by more than 2 SD from the expected value appears to be close to the optimal sam-

Table 3. Global linear sensitivity analysis based on the standard regression coefficients (SRC) method for total quality costs C^1

Input variable ²	Estimates	SE	t-value	$P > t $	Low bound	High bound	SD of input variables	SRC
$1/\lambda$	-3.153	0.0177	-178.21	<0.0001	7	60	15.30	-0.327
Δ	-25.17	0.4095	-61.46	<0.0001	0.75	3	0.65	-0.111
T_0	-4.150	0.1448	-28.65	<0.0001	0.5	7	1.88	-0.053
T_2	0.474	0.1448	3.27	0.0011	0.5	7	1.88	0.006
E	2.695	0.0699	38.55	<0.0001	0.5	14	3.90	0.071
N_c	0.380	0.0010	382.45	<0.0001	50	1,000	274.24	0.704
Δ_{m0}	21.52	0.6142	35.05	<0.0001	0.22	0.91	0.43	0.063
Δ_{m1}	22.09	0.1161	190.32	<0.0001	0.91	4.55	2.31	0.345
P_M	424.2	10.4681	40.53	<0.0001	0.22	0.37	0.02	0.058
Y	-0.0096	0.0010	-9.70	<0.0001	50	1,000	274.24	-0.018
W	-0.0008	0.0010	-0.77	0.4424	50	1,000	274.24	-0.001
a	-1.417	0.0464	-30.52	<0.0001	5	25	5.77	-0.055
b	-0.0240	0.0082	-2.94	0.0033	10	125	33.20	-0.005

¹ $R_y^2 = 0.75$ under a sampling scheme that resulted in a standard deviation for C of \$147.77/d.

² $1/\lambda$ = mean time process is in control (d); Δ = number of standard deviations slip when process moves out-of-control (abrupt change in forage composition); T_0 = expected search time for false alarm (d); T_2 = expected time to repair the process (i.e., to rebalance and implement new rations) (d); E = time to sample and analyze all items (e.g., time awaiting for lab results) (d); N_c = number of cows in the herd; Δ_{m0} = milk production loss due to white noise change in forage composition while the process is in-control (kg/cow per d); Δ_{m1} = milk production loss when process goes out-of-control (abrupt change in forage composition) (kg/cow per d); P_M = price of milk (\$/kg); Y = cost per false alarm (\$); W = cost to locate and repair the assignable cause (i.e., to rebalance and implement new rations) (\$); a = fixed cost per sampling time (preparation and mailing of samples) (\$); b = cost per unit sampled (\$).

Table 4. Global nonlinear sensitivity analysis (e-FAST¹): proportions (expressed as percentages) of first-order sensitivity indices (S_i) to total sensitivity indices (S_{Ti}) for total quality costs per day (C), control limits (L), number of samples (n), and sampling interval (h)

Input variable ²	C	L	n	h
$1/\lambda$	62.4	0.7	24.8	7.4
Δ	56.9	62.4	28.1	0.3
T_0	0.0	0.4	0.1	0.4
T_2	42.3	1.0	1.8	0.0
E	47.4	2.8	7.1	0.1
N_c	85.0	4.3	13.0	20.2
Δ_{m0}	68.4	1.0	1.5	4.0
Δ_{m1}	70.0	1.5	13.4	15.4
P_M	70.3	0.2	1.5	0.9
Y	36.1	45.4	34.6	0.1
W	60.4	2.1	0.5	4.4
a	3.1	0.6	0.6	0.0
b	54.2	49.2	53.3	0.5

¹e-FAST = extended Fourier amplitude sensitivity test.

² $1/\lambda$ = mean time process is in control (d); Δ = number of standard deviations slip when process moves out-of-control (abrupt change in forage composition); T_0 = expected search time for false alarm (d); T_2 = expected time to repair the process (i.e., to rebalance and implement new rations) (d); E = time to sample and analyze all items (e.g., time awaiting for lab results) (d); N_c = number of cows in the herd; Δ_{m0} = milk production loss due to white noise change in forage composition while the process is in-control (kg/cow per d); Δ_{m1} = milk production loss when process goes out-of-control (abrupt change in forage composition) (kg/cow per d); P_M = price of milk (\$/kg); Y = cost per false alarm (\$); W = cost to locate and repair the assignable cause (i.e., to rebalance and implement new rations) (\$); a = fixed cost per sampling time (preparation and mailing of samples) (\$); b = cost per unit sampled (\$).

Table 5. Global nonlinear sensitivity analysis (e-FAST¹): proportions (expressed as percentages) of individual total sensitivity indices (S_{Ti}) to the sum of the total sensitivity indices (ΣS_{Ti}) for total quality costs per day (C), control limits (L), number of samples (n), and sampling interval (h)

Input variable ²	C	L	n	h
$1/\lambda$	11.8	10.9	8.1	11.6
Δ	0.4	12.3	18.3	3.9
T_0	0.1	3.4	2.3	4.0
T_2	0.6	4.7	2.6	3.7
E	2.5	3.2	2.8	2.9
N_c	55.9	12.6	4.7	19.1
Δ_{m0}	2.2		3.0	9.3
Δ_{m1}	20.1	12.5	6.0	17.4
P_M	5.6	4.2	2.7	4.5
Y	0.2	6.4	14.4	3.6
W	0.2	9.1	2.6	9.5
a	0.1	4.4	3.1	5.8
b	0.3	7.8	29.6	4.6

¹e-FAST = extended Fourier amplitude sensitivity test.

² $1/\lambda$ = mean time process is in control (d); Δ = number of standard deviations slip when process moves out-of-control (abrupt change in forage composition); T_0 = expected search time for false alarm (d); T_2 = expected time to repair the process (i.e., to rebalance and implement new rations) (d); E = time to sample and analyze all items (e.g., time awaiting for lab results) (d); N_c = number of cows in the herd; Δ_{m0} = milk production loss due to white noise change in forage composition while the process is in-control (kg/cow per d); Δ_{m1} = milk production loss when process goes out-of-control (abrupt change in forage composition) (kg/cow per d); P_M = price of milk (\$/kg); Y = cost per false alarm (\$); W = cost to locate and repair the assignable cause (i.e., to rebalance and implement new rations) (\$); a = fixed cost per sampling time (preparation and mailing of samples) (\$); b = cost per unit sampled (\$).

pling design in small herds (50 cows). However, it appears to be incorrect in large herds, in which the optimal design requires taking 2 samples every 4 d and intervening if the average of the 2 samples differ by more than 1.2 SD from expectation. This optimal design reduces daily costs by about \$250.

Sensitivity of total quality costs is not affected much by the interaction of input variables. Total sensitivity of C is primarily for N_c , Δ_{m1} , $1/\lambda$, P_m , and E , whereas total sensitivities of the design parameters n , h , and L are primarily for $1/\lambda$, Δ , N_c , Δ_{m1} , and b . Sensitivities of design parameters have large interaction and high-order components, indicating that the traditional sensitivity analysis of varying one parameter at a time yields incorrect sensitivity assessments.

ACKNOWLEDGMENTS

This project was a component of the NC-1119 Regional Research Project. The authors thank Joanne Knapp, Maurice Eastridge, and 2 anonymous reviewers for helpful comments and suggestions on an earlier version of this manuscript.

REFERENCES

- DeVries, A., B. J. Conlin, W. E. Marsh, and J. K. Reneau. 1997. Monitoring daily milk weights with statistical process control techniques. *J. Dairy Sci.* 80(Suppl. 1):272. (Abstr.)
- Drain, D. 1997. *Statistical Methods for Industrial Process Control*. Chapman & Hall, New York, NY.
- Fadel, J. G., H. A. Johnson, and P. H. Robinson. 2006. Impact of variation in diet nutrient inputs on model outputs. *J. Dairy Sci.* 89(Suppl. 1):383. (Abstr.)
- Giglioli, N., and A. Saltelli. 2006. SimLab 1.1. Software for Sensitivity and Uncertainty Analysis. Joint Research Centre, European Commission, Institute for Systems Informatics and Safety. Ispra, Italy. <http://www.citebase.org/cgi-bin/citations?identifier=oai:arXiv.org:cs/0011031> Accessed July 29, 2006.
- Goldberg, E. D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Professional, Boston, MA.
- Haupt, R. L., and S. E. Haupt. 1998. *Practical Genetic Algorithms*. John Wiley and Sons, New York, NY.
- Ishikawa, K. 1976. *Guide to Quality Control*. Asian Productivity Organization, Tokyo, Japan.
- Knappenberger, H. A., and A. H. Grandage. 1969. Minimum cost quality control tests. *AIIE Trans.* 1:24–32.
- Lorenzen, T. J., and L. C. Vance. 1986. The economic design of control charts: A unified approach. *Technometrics* 28:3–10.
- Lukas, J. M., D. M. Hawkins, M. L. Kinsel, and J. K. Reneau. 2005. Bulk tank somatic cell counts analyzed by statistical process control tools to identify and monitor subclinical mastitis incidence. *J. Dairy Sci.* 88:3944–3952.
- NRC. 2001. *Nutrient Requirements of Dairy Cattle*. 7th rev. ed. Natl. Acad. Sci. Washington, DC.
- Parzen, E. P. 1999. *Stochastic Processes*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Reynolds, M. R., and Z. G. Stoumbos. 2005. Should exponentially weighted moving average and cumulative sum charts be used with Shewhart limits? *Technometrics* 47:409–424.
- Ryan, T. P. 2000. *Statistical Methods for Quality Improvement*. 2nd ed. Wiley, New York, NY.
- Saltelli, A., K. Chan, and M. Scott. 2000. *Sensitivity Analysis*. Wiley, New York, NY.
- Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, New York, NY.
- Saltelli, A., S. Tarantola, and K. S. Chan. 1999. Quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41:39–56.
- Sauvant, D., J.-M. Perez, and G. Tran. 2004. *Tables de Composition et de Valeur Nutritive des Matières Premières Destinées aux Animaux d'Élevage*. INRA Editions, Paris, France.
- Shewhart, W. A. 1931. *Economic Control of Quality of Manufactured Products*. Van Nostrand, New York, NY.
- SAS Institute. 2004. *SAS/Stat 9.1 User's Guide*. Vol. 4. SAS Inst., Inc. Cary, NC.
- St-Pierre, N. R., and B. Cobanov. 2007. Optimal sampling schedule of diet components: Model robustness to departure from assumptions. *J. Dairy Sci.* 90:5395–5404.
- St-Pierre, N. R., B. Cobanov, and G. Schnitkey. 2003. Economic losses from heat stress by U.S. livestock industries. *J. Dairy Sci.* 86(E. Suppl.):E52–E77.
- St-Pierre, N. R., D. Shoemaker, and L. R. Jones. 2000. The next \$120,000: A case study to illustrate analysis of alternative farm investments in fixed assets. *J. Dairy Sci.* 83:1159–1169.
- St-Pierre, N. R., and C. S. Thraen. 1999. Animal grouping strategies, sources of variation, and economic factors affecting nutrient balance on dairy farms. *J. Anim. Sci.* 77(Suppl. 2):72–83.
- St-Pierre, N. R., and W. P. Weiss. 2006. Managing feedstuff variation in nutritional practice. *J. Dairy Sci.* 89(Suppl. 1):383. (Abstr.)
- Tozer, P. R. 2000. Least-cost ration formulations for Holstein dairy heifers by using linear and stochastic programming. *J. Dairy Sci.* 83:443–451.
- Yashin, E. 1993. Statistical control schemes: Methods, applications and generalizations. *Int. Stat. Rev.* 61:41–66.