CHAPTER 3

DYNAMIC PRODUCTION MONITORING IN PIG HERDS I: MODELING AND MONITORING LITTER SIZE AT HERD AND SOW LEVEL

Claudia Bono, Cécile Cornou and Anders Ringgaard Kristensen

Published in *Livestock Science* 149 (2012) 289-300. Doi: 10.1016/j.livsci.2012.07.023

Abstract: Monitoring animal production results in real time is a challenge. Existing management information systems (MIS) in pig production are typically based on static statements of selected key figures. The objective of this paper is to develop a dynamic monitoring system for litter size at herd and sow level, with weekly updates. For this purpose, a modified litter size model, based on an existing model found in the literature, is implemented using dynamic linear models (DLMs). The variance components are pre-estimated from the individual herd database using a maximum-likelihood technique in combination with an Expectation-Maximization (EM) algorithm applied on a larger dataset with observations from 15 herds. The model includes a set of parameters describing the parity-specific mean litter sizes (herd level), a time trend describing the genetic progress (herd level), and the individual sow effects (sow level). It provides reliable forecasting with known precision, on a weekly basis, for future production. Individual sow values, useful for the culling strategy, are also computed. In a second step, statistical control tools are applied. Shewhart Control Charts and V-masks are used to give warnings in case of impaired litter size results. The model is applied on data from 15 herds, each of them including a period ranging from 150 to 800 weeks. For each herd, the litter size profile, the litter size over time, the sow

individual effect and sow economic value, are computed. Perspectives for further development of the model can take into account indices including conception rate, service rate, mortality rate etc. Such a model can be used as a basis for developing a new, dynamic, management tool.

Keywords: Monitoring; Litter size; Dynamic Linear Model; Statistical control

3.1 Introduction

In a situation with increasing herd sizes and growing competition from countries with lower production costs, the profitability of pig herds becomes increasingly sensitive to even small fluctuations in production results (Frost et al., 1997). Reliable monitoring systems therefore become of paramount importance to the farmer who, as mentioned by Frost et al. (1997), has the full responsibility for all aspects of his husbandry including monitoring of feeding, reproduction, health, growth, quality and other key indicators. Performing these tasks become increasingly important and an appropriate real-time monitoring management system could be desirable.

During the last decades the combination of sensors, databases and mathematical models has allowed the monitoring activity to be more precise, and a MIS joins human knowledge, organization's procedures and technologies, to support human decision making (Sørensen et al., 2010). Whereas current IT-based management tools available to pig farmers are well suited for storing data and providing regular to-do list, there is a great potential for improvements when it comes to systematic use of data for monitoring and forecast.

The problem of existing systems in pig production is that they are based on static statements of selected key figures calculated for a specific time interval (typically a quarter or a year). Most systems allow for comparison with the production goals of the herd, but, otherwise they don't assist the manager in any way in the interpretation of the production results. Through these systems, the advisor may help the farmer to manage his farm and do some forecasts on the future productions. Nevertheless, these systems do not allow farmer intervention in real time and this may result in decreased productivity and financial losses. Alternative systems including real time monitoring of unexpected changes in performance would allow to control the results in an acceptable range of time and to reduce economic loss (De Vries and Reneau, 2010).

Information systems for supporting the decision-making process have been applied in several fields such as financial (Humphrey, 1994), medical (Doolin, 2004), marketing (Nasir, 2005) etc. MIS is generally composed by different parts or levels (Boehlje and Eidman, 1984). In livestock production for instance, it can be applied in nutrition, animal herd health, breeding and genetics, animal products processing, animal production and management (Irvin, 2008). A potential problem in livestock industry is that the factors submitted for analysis, are biological and consequently variable and unpredictable (Frost et al., 1997). Most often the results are presented in a table as simple averages over time and animals.

This paper is the first result of a larger project with the overall aim of breaking fundamentally with those traditional principles and, instead, provide monitoring methods that are proactive and forward looking. These methods should allow the manager to make decisions based on dynamically updated forecasts, so that intervention is possible before production failures result in considerable financial losses or reduced animal health and/or welfare.

The aim of this first paper is to develop a dynamic monitoring system for litter size (total born) at herd and sow level. For accurate predictions of piglet production, the sow level is needed because individual sow productivity varies systematically with parity and genetic properties. Since these effects are herd specific, the herd level is needed as well. For this purpose a further developed version of the litter size model proposed by Toft and Jørgensen (2002) is implemented in a dynamic setting with sequential weekly updating of parameters at herd and sow level. In addition, statistical control tools (Shewhart Control Chart and V-mask) are implemented in order to give warnings in case of impaired litter size results. The system developed will be an important step towards creation of a reliable forecasting system for future production.

3.2 The litter size model proposed by Toft and Jørgensen (2002)

One of the most important measures of productivity in sow herds is the litter size. A simple average litter size is not suitable for monitoring, because it depends heavily on the age structure of the herd. A monitoring system for litter size must adjust for age distribution. In order to do this, the parity specific mean litter size must be known.

Toft and Jørgensen (2002) proposed the following litter size model for sow i at parity n

$$Y_{in} = -\phi_1 e^{-(n^2-1)\phi_2} + \phi_3 - \phi_4 n + M_i(n) + \epsilon_{in}, \quad (3.1)$$

where Y_{in} represents the litter size expressed as total born piglets for sow i in parity n ; $\phi_1, \dots, \phi_4$ are the parameters describing the curve; and where $\epsilon_{in} \sim \mathcal{N}(0, \tau^2)$ and the sow effect $M_i(n) \sim \mathcal{N}(0, \sigma^2)$ is a first order autoregressive time series described as

$$M_i(n) = e^{-\alpha} M_i(n-1) + \eta_{in}, \quad (3.2)$$

with $\eta_{in} \sim \mathcal{N}(0, (1 - e^{-2\alpha})\sigma^2)$. The parameter α represents the autocorrelation between parities of the same sow. Thus, the full litter size model has 7 parameters: $\phi_1, \phi_2, \phi_3, \phi_4, \tau, \sigma, \alpha$. Given a large sample of litter size records from a herd, these seven parameters can be estimated as described by Toft and Jørgensen (2002).

Essential for prediction of litter sizes are the four parameters $\phi_1, \dots, \phi_4$ describing the unselected herd mean (see Figure 3.1(a) for an illustration) and the sow effects $M_1(n_1), \dots, M_N(n_N)$ of the N sows in the herd. One approach could be to

(a) Model from Toft and Jørgensen (2002)

(b) Modified model (2012)

Figure 3.1: Differences between the model suggested by Toft and Jørgensen (2002) and the modified one. (a) Original model and corresponding equation. (b) Reparametrization of the previous model. The parameters ϕ_1 , ϕ_2 and ϕ_3 are replaced by the parameters μ_1 to μ_4 , allowing the use of a Dynamic Linear Model. The parameter ϕ_4 remains unchanged.

estimate the litter size profile and the variance components from the herd database as described by Toft and Jørgensen (2002), and then, given these estimates, sequentially update the estimate for the sow effect as described by Kristensen and Søllested (2004a) in the context of a replacement optimization model (i.e. not for monitoring). Such an approach has the following weaknesses for monitoring purposes:

- It cannot be applied in new herds, because no (or only very few) litter size records are available.
- The litter size profile may change over time (and is likely to do so). There can be several reasons for that, but an obvious example is the genetic trend which leads to increasing litter sizes over time.
- The estimates for the litter size profile $\phi_1, \dots, \phi_4$ and the sow effects $M_i(n)$ are mutually dependent.

It is therefore advisable to have a sequential estimation technique which is able, week by week, to simultaneously update the estimates for $\phi_1, \dots, \phi_4$ and the sow effects $M_i(n)$.

3.3 Sequential estimation of litter size profile and sow effects

This section presents the process of sequential estimation. A re-parameterized model is suggested. The reason of the re-parametrization is to make the parameters linear in order to use them in a Dynamic Linear Model. The components of the DLM are described and an example illustrates the principle behind the model design.

3.3.1 A linear litter size model

A Dynamic Linear Model as described by West and Harrison (1997) is an obvious tool for sequential estimation. In this case we face the problem that the litter size profile of the herd as suggested by Toft and Jørgensen (2002)

$$\mu_n = -\phi_1 e^{-(n^2-1)\phi_2} + \phi_3 - \phi_4 n, \quad (3.3)$$

is not linear. For “large” n , i.e. $n > 3$ we have nevertheless an approximative linear expression (cf. Figure 3.1a)

$$\mu_n \approx \phi_3 - \phi_4 n, \quad n > 3, \quad (3.4)$$

and for $n = 1$ we have

$$\mu_1 = -\phi_1 + \phi_3 - \phi_4. \quad (3.5)$$

Thus, the only mean litter sizes which are not at least approximately linear in the parameters are those for parities 2 and 3. As illustrated in Figure 3.1a, it is the parameter ϕ_2 which determines how much to subtract from the linear expression (3.4) in order to obtain μ_2 and μ_3 . We will, therefore, estimate those two mean values as

$$\mu_2 = -\phi_{22} + \phi_3 - 2\phi_4 \quad (3.6)$$

and

$$\mu_3 = -\phi_{23} + \phi_3 - 3\phi_4. \quad (3.7)$$

In other words, we use two parameters ϕ_{22} and ϕ_{23} to model the shape between μ_1 and μ_4 instead of just ϕ_2 in the original model by Toft and Jørgensen (2002). The benefit is that we obtain linear expressions for the litter size model, and the cost is the additional parameter.

When looking at the expressions (3.4 - 3.7) it seems more natural to re-parameterize the entire model, so that μ_1 , μ_2 , μ_3 and μ_4 (Figure 3.1b) are modeled and estimated directly and, for $n > 4$,

$$\mu_n = \mu_4 - (n - 4)\phi_4, \quad n > 4. \quad (3.8)$$

That will be the approach used in this study.

3.3.2 Litter size model with time trend

In the Danish breeding system high priority is given to improvement of litter size. An estimated annual genetical improvement of average litter size per parity of 0.45 live piglets (at day 5 after farrowing) has been reported (Pig Research Centre, 2010). With a sequential (and dynamic) estimation technique, the parameters of the litter size model will automatically adapt to the trend, so for pure estimation a direct modeling of the trend is less important. If, however, the model is used for forecasting, a direct modeling of the trend is important in order not to underestimate future litter sizes. In the following we shall use the notation μ_{nt} to indicate expected litter size at parity n of a sow farrowing first time at time t .

We shall, therefore, extend the model for average litter size at parity n with a linear time trend δ , so that expected litter size at parity n , μ_{nt} , of a sow with first farrowing at time t is

$$\mu_{nt} = \mu_n + \delta t, \quad (3.9)$$

where μ_n is defined as in Eq. (3.5 - 3.8). Since, however, the trend may also change over time, the recursive relation

$$\mu_{nt} = \mu_{n,t-1} + \delta_{t-1} \quad (3.10)$$

will be used to model the development of the general trend. It should be noticed that μ_{nt} for $n > 1$ does *not* refer to expected litter size of a parity n sow farrowing at time t . Instead, the index t refers to the time of *first* farrowing of a parity n sow. A trend of this kind assumes that increased litter size is the result of genetical improvement.

3.3.3 A dynamic linear model

A multivariate dynamic linear model consisting of an observation equation and a system equation as described by West and Harrison (1997) will be applied. We will use weekly observations of litter sizes of all farrowing sows to update the herd profile and the sow effects. The latent parameter vector for week t will be

$$\theta_t = (\mu_{1t}, \mu_{2t}, \mu_{3t}, \mu_{4t}, \phi_{4t}, \delta_t, M_1(n_1t), \dots, M_N(n_Nt))'. \quad (3.11)$$

Thus, the size of the parameter vector equals the number of sows, N , in the herd plus the 6 parameters describing the profile.

Observation equation

The observation vector $Y_t = (y_{f1t}, \dots, y_{fkt})'$ will consist of observed litter sizes of all k sows farrowing in week t . The observation equation linking the observations to the parameters has the general form

$$Y_t = F_t' \theta_t + v_t, \quad (3.12)$$

where F_t is called the *design matrix*, and $v_t \sim \mathcal{N}(\underline{0}, I\tau^2)$ where $\underline{0}$ is a vector of zeros and I is the identity matrix. The challenge here is to setup the design matrix.

The content of the matrix depends on the sow number and the parity. The expected litter size of a parity n sow farrowing at time t is μ_{nt_1} , where $t_1 \leq t$ is the time of first farrowing. A linear expression for μ_{nt_1} is only available if previous level estimates are remembered as long time back as the week of first farrowing of the oldest sow in the herd. In other words, we would have to include in the parameter vector also those historical values. Such a solution would, however, drastically increase the complexity of the dynamic linear model (if the oldest sow is a parity 10, the week of first farrowing will be more than 200 weeks back). We therefore rely on the following approximation:

$$\mu_{nt_1} \approx \mu_{nt} - (n - 1)d\delta_t, \quad (3.13)$$

where d is the average interval (in weeks) between two farrowings of the same sow. A value of $d = 23$ will be used. The assumptions behind the approximation is that the intervals between farrowings only vary relatively little and that the current trend δ_t is representative for the entire time interval from t_1 to t .

Now assume that $k = 5$ sows farrow in week t and they are characterized as shown in Table 3.1. The observation vector would then be $Y_t = (12, 11, 15, 16, 14)'$, and the design matrix will then look as follows:

$$F' = \begin{pmatrix} 0 & 0 & 0 & 1 & -2 & -5d & \dots & 1 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & -3d & \dots & 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & -2d & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & -d & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots \end{pmatrix} \quad (3.14)$$

Table 3.1: Example of 5 sows farrowing in week t .

Sow #	Parity	Piglets born	Sow effect
287	6	12	$M_{287}(6)$
321	4	11	$M_{321}(4)$
356	3	15	$M_{356}(3)$
410	2	16	$M_{410}(2)$
430	1	14	$M_{430}(1)$

with the corresponding parameter vector

$$\theta_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \\ \mu_{4t} \\ \phi_{4t} \\ \delta_t \\ \vdots \\ M_{287}(6t) \\ \vdots \\ M_{321}(4t) \\ \vdots \\ M_{356}(3t) \\ \vdots \\ M_{410}(2t) \\ \vdots \\ M_{430}(1t) \\ \vdots \end{pmatrix}.$$

System equation

The system equation expresses how the parameter values may change over time. The general form of the system equation is

$$\theta_t = G_t \theta_{t-1} + w_t, \quad (3.15)$$

where G_t is called the *system matrix*, and $w_t \sim \mathcal{N}(\underline{0}, W)$ where $\underline{0}$ is a vector of zeros and W is a variance-covariance matrix describing the system variance and covariance of each of the parameters.

As concerns the sow effect $M_i(n)$, the relation from parity is already given by Eq. (3.2), but in this model we need the weekly relation. Since we are only interested in the effect at farrowing we will put $\alpha = 0$, so that $e^{-\alpha} = 1$, in weeks where

no farrowing takes place for a sow and use the parity to parity value in weeks where a farrowing actually happens. Thus, the variance component in W corresponding to each sow effect is $(1 - e^{-2\alpha})\sigma^2$ with $\alpha = 0$ if no farrowing takes place (making the variance component 0 as well). The variance components corresponding to the herd level effects are estimated using the EM-algorithm described in Section 3.3.3.

Including as an example only one multi-parous farrowing sow (number 321, parity 4), one new primiparous farrowing sow (number 430), one not farrowing sow (number 415, parity 2) and a sow (number 356, parity 3) which has been culled between time $t - 1$ and time t , the parameter vectors θ_{t-1} and θ_t become

$$\theta_{t-1} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \\ \mu_{3,t-1} \\ \mu_{4,t-1} \\ \phi_{4,t-1} \\ \delta_{t-1} \\ M_{321}(3, t-1) \\ M_{356}(3, t-1) \\ M_{415}(2, t-1) \end{pmatrix}, \quad \theta_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \\ \mu_{4t} \\ \phi_{4t} \\ \delta_t \\ M_{321}(4t) \\ M_{415}(2t) \\ M_{430}(1t) \end{pmatrix}$$

and the system matrix converting θ_{t-1} to θ_t is

$$G_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{-\alpha} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.16)$$

As it is seen in the matrix G_t , the culling of Sow 356 is handled by a full column of zeros at the position of the sow in the parameter vector θ_{t-1} . Thus the sow effect $M_{356}(3, t-1)$ is ignored in the calculation of the new expected value of θ_t . Insertion of the new Sow 430 is handled by a full row of zeros at the bottom of the matrix corresponding to the index of the sow in the new parameter vector θ_t . Thus, the expected value of $M_{430}(1t)$ is zero and independent of the previous parameter vector which did not contain the sow.

For the variance-covariance matrix W_t the following structure is assumed:

$$W_t = \begin{pmatrix} W_{11} & W_{12} & W_{13} & W_{14} & 0 & 0 & 0 & 0 & 0 \\ W_{12} & W_{22} & W_{23} & W_{24} & 0 & 0 & 0 & 0 & 0 \\ W_{13} & W_{23} & W_{33} & W_{34} & 0 & 0 & 0 & 0 & 0 \\ W_{14} & W_{24} & W_{34} & W_{44} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & W_{55} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & W_{66} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (1 - e^{-2\alpha})\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}. \quad (3.17)$$

As it is seen, the changes of the level parameters $\mu_1, \dots, \mu_4$ are assumed to be mutually correlated, but independent of changes in other model elements. The sow properties are assumed to be mutually independent. Sow 321 farrows in week t , so its properties may change according to Eq. (3.2). Sow 415 does not farrow, so its properties are constant. The last sow, 430, is a new gilt farrowing, so its properties are drawn from $\mathcal{N}(0, \sigma^2)$.

Weekly update

In any week, the last week's estimate for the parameter vector $\theta_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$ is available. The vector m_{t-1} contains the estimates for the 5 litter size profile parameters, the time trend δ_{t-1} and the N_{t-1} sows present in the herd.

Based on the sows farrowing in week t , the observation vector Y_t , the design matrix F_t and the system matrix G_t will be defined as described in the previous sections and a new estimate $\theta_t \sim \mathcal{N}(m_t, C_t)$ is computed using the updating equations of the Kalman filter (filtering) as described by Kristensen et al. (2010).

As a result of the filtering, we obtain a weekly estimate of:

- The herd specific parameters $\mu_1, \mu_2, \mu_3, \mu_4, \phi_4$, where μ_n is the mean of a sow at parity n and ϕ_4 is the slope of the curve.
- The time trend δ . The trend takes into consideration the sows' genetic improvement during the years using Eq. (13) as a "correction" formula.
- The individual property of each sow in the herd, M_i .

Smoothing is a retrospective analysis applicable for those values, where fluctuations are evaluated after the events. Data are re-examined starting from the last updated, and backwards to the beginning (West and Harrison, 1997). This allows to have more precise information about the consequences of particular events, such as employment of new herdsman, disease outbreaks and so on (Kristensen et al., 2010). This retrospective analysis is essential to improve the understanding and to accelerate the decision process.

Sow values and their economical consequences

It is worth mentioning that the described DLM for monitoring complements the sow replacement model described by Kristensen and Søllested (2004a,b) in the sense that it provides a consistent framework for linking the replacement model directly to the data observed in a sow herd. Thus, the herd specific parameters $\mu_1, \mu_2, \mu_3, \mu_4, \phi_4$, are needed as input to model construction and the estimated sow value M_i is part of the state definition for individual sows.

It is therefore, by use of the replacement model, possible to provide an economical value of an individual sow in terms of the so-called future profitability originally defined by Van Arendonk (1985) for dairy cows and later used for sows by Huirne et al. (1991) who called it the *retention pay-off*. The value, which can be calculated by any replacement model based on dynamic programming, expresses the total extra profit to be expected from keeping a sow until her optimum, compared with immediate replacement.

Therefore, the sow value M_i expressing the estimated deviation (in number of piglets) of a sow compared to herd level litter size profile expressed by $\mu_2, \dots, \mu_4, \phi_4$, can by use of the replacement model by Kristensen and Søllested (2004a,b) be associated with a direct economic value in terms of the retention pay-off. The values presented in this study are computed at the time of weaning.

Initialization

In order to have a full specification of the DLM, the initial information $\theta_0 \sim \mathcal{N}(m_0, C_0)$ before anything has been observed in the herd must be defined. We will have to distinguish between two different scenarios:

New herd, no records: The subsets of m_0 and C_0 referring to the 5 litter size profile parameters $(\mu_1, \mu_2, \mu_3, \mu_4, \phi_4)$ will be defined from the population mean and the population variance-covariance matrix. For the subset of m_0 and C_0 referring to sow properties, all elements of m_0 are set equal to 0, and the corresponding diagonal elements of C_0 are set equal to σ^2 .

Existing herd, database available: From the herd database, a herd specific litter size profile with variance components τ^2, σ^2 and α , is estimated as described by Toft and Jørgensen (2002). Based on the estimated values, the sow effects of the existing sows will be calculated sequentially as described by Kristensen and Søllested (2004b). Thus, the full m_0 is defined.

EM algorithm for estimation of system variance

The EM-algorithm is an iterative algorithm used to estimate unknown parameters by Maximum Likelihood estimation (ML). A free software for statistical computing and graphics was used (R Development Core Team, 2012) for implementation of the algorithm. In ML estimation, we wish to compute the model parameter(s) for

which the observed data are the most likely. The calculation procedure is composed of iterations. The two steps of the EM-algorithm can be formulated as maximizing the conditional expectation of the log likelihood for the augmented data, given the observed data and the previous estimate (Dethlefsen, 2001).

The sow related elements of the system variance in Eq. (3.17) are known from the herd specific estimation performed by the method described by Toft and Jørgensen (2002). The herd related elements, i.e. the upper left 6×6 sub matrix of W_t , are on the other hand unknown. They are, therefore, estimated by use of the EM algorithm on a collection of herd databases from Danish sow data. Examples of EM-algorithm applications for estimation of variance components for sow data can be found in literature, e.g. in Cornou and Lundbye-Christensen (2010) and Cornou et al. (2011).

For the estimation procedure, time series from H herds are used. Each time series, $h = 1, \dots, H$, is observed over T_h weeks. Initially, a first guess of the unknown elements of W_t is used. This initial guess is successively improved over a large number of iteration steps until convergence. At each iteration step, m , the Kalman filter is applied for $t = 1, \dots, T_h$ on each time series to obtain the herd and week specific estimates m_{ht} and C_{ht} for the parameter vector θ_{ht} . Afterwards, smoothing is performed to obtain improved estimates $\tilde{m}_{ht}$ and $\tilde{C}_{ht}$ for θ_{ht} . These estimates (plain and smoothed) are then used to improve the estimate for W_t from W_t^m to W_t^{m+1} .

Inspired by the description provided by Dethlefsen (2001), an iteration step of the EM algorithm can be described as follows:

Let W^m be the current (i.e. iteration m) estimate for the upper left 6×6 sub matrix of W_t . The updated estimate, W^{m+1} , to be used as estimate at iteration $m + 1$ is found as

$$W^{m+1} = \frac{1}{\sum_{h=1}^H T_h} \sum_{h=1}^H \sum_{t=1}^{T_h} (L_{ht} + (\tilde{m}_{ht} - G_{ht}\tilde{m}_{h,t-1})(\tilde{m}_{ht} - G_{ht}\tilde{m}_{h,t-1})'), \quad (3.18)$$

where

$$L_{ht} = \tilde{C}_{ht} + G_{ht}\tilde{C}_{h,t-1}G'_{ht} - \tilde{C}_{ht}B'_{h,t-1} - B_{h,t-1}\tilde{C}'_{ht}, \quad (3.19)$$

and

$$B_{h,t-1} = C_{h,t-1}G'_{ht}(G_{ht}C_{h,t-1}G'_{ht} + W^m)^{-1}. \quad (3.20)$$

The equations (3.18), (3.19) and (3.20) are applied block-wise to the system variance-covariance matrix W and only to the herd related elements of the upper left 6×6 sub matrix corresponding to the herd level. In other words, they are applied separately to the upper left 4×4 sub matrix, to the diagonal element W_{55} and the diagonal element W_{66} .

3.4 Detection of impaired litter size results

After applying the DLM, litter sizes and individual sows' properties are monitored. Deviations of the observations as compared to the predicted values are monitored within a short- and long term period.

The short term monitoring is a weekly control of the results, where differences between observed and predicted values are monitored using a Control Chart inspired by Shewhart (Montgomery, 2005). A Control Chart is a statistical tool used to see, graphically, if observations or time series are running as they should. It is, in general, composed of three elements:

- A central line (CL) corresponding to the target value, θ' ,
- An upper control limit (UCL),
- A lower control limit (LCL).

In the current monitoring method, the target value (CL) is set to 0, since we expect that, in a normal situation, the observed values do not differ significantly from the predicted ones. The control limits are used to see whether or not the process is 'out-of-control' (Kristensen et al., 2010). They are defined as:

$$UCL_t = \theta' + as_t \quad (3.21)$$

and

$$LCL_t = \theta' - as_t \quad (3.22)$$

where a is the distance parameter and s_t is the standard deviation at week t . The value of a is set to 2, which corresponds to an approximate 5% significance level for normally distributed data (Kristensen et al., 2010).

The purpose of the long term monitoring is to detect sudden or gradual changes in level of litter size over time. It is performed over the last three years using a cumulative sum (Cusum) Control Chart combined with a V-mask as described by Montgomery (2005). The cumulative sum represents the sum of the accumulated differences between observed and predicted values. It incorporates past observations and for this reason, it is also able to detect small shifts in the process (De Vries and Reneau, 2010). If the process is in control, the Cusum should fluctuate stochastically around the 0 level. On the other hand, if data show a trend, the level can shift.

To monitor level shifts, a V-mask is applied on the Cusum Control Chart. The setting of the V-mask is defined by Barnard (1959) as the lead distance d , which reflects the model sensitivity to short term changes, and the angle Ψ , which determines how much long term drift can be accepted. In the present application of the method, the values $d = 10$ and a slope of the V-mask of 0.6 (corresponding to an opening of the V-mask, Ψ , of 31 degrees) are used. For detailed information about

(a) V-mask design

(b) V-mask applied on hypothetical Cusum

Figure 3.2: The V-mask. (a) The V-mask and scaling. The parameters d and Ψ determine the sensitivity to short term changes and (long term) drifts, respectively. (b) The V-mask applied on a hypothetical Cusum.

the choice of the settings, see Montgomery (2005). If any point of the Cusum lies outside the arms, the process is considered 'out-of-control', and the Cusum is reset to zero. Figure 3.2 illustrates the design of the V-mask and its application on a Cusum.

For both short and long term monitoring, the following components are extracted from the model: the forecast for the observation vector at time t (West and Harrison, 1997), which is:

$$f_t = F_t' G_t m_{t-1}, \quad (3.23)$$

and its variance

$$Q_t = F_t' (G_t C_t G_t + W_t) F_t + I \tau^2. \quad (3.24)$$

Let $\underline{1} = (1, \dots, 1)$ be a row vector consisting of only elements with value 1. The forecast for the total number of piglets born in week t is therefore $\underline{1} f_t$ with variance $\underline{1} Q_t \underline{1}'$ and the observed total number is $\underline{1} Y_t$. Thus the weekly forecast error, e_t , is:

$$e_t = \underline{1} Y_t - \underline{1} f_t. \quad (3.25)$$

For the Shewhart Chart, the observation in week t is e_t and the standard deviation used for control limits, is

$$s_t = \sqrt{\underline{1} Q_t \underline{1}'}. \quad (3.26)$$

Thus, the numerical value of s_t will depend heavily of the number of sows farrowing at week t .

Table 3.2: Upper left matrix of the system variance-covariance W (6 x 6). Values of the correlations are shown below the diagonal.

Matrix #	1	2	3	4	5	6
1	0.00072	0.00071	0.00069	0.00068	0	0
2	<i>0.90</i>	0.00087	0.00104	0.00121	0	0
3	<i>0.69</i>	<i>0.95</i>	0.00139	0.00176	0	0
4	<i>0.53</i>	<i>0.85</i>	<i>0.98</i>	0.00232	0	0
5	0	0	0	0	1.52e-10	0
6	0	0	0	0	0	1.48e-10

For the V-mask, the cumulative sum (Cusum) is defined as the sum of the standardized forecast errors:

$$C_t = \sum_{t=1}^t \frac{e_t}{s_t}. \quad (3.27)$$

3.5 Results

This section presents the results of the system variance estimation, of the model application and of the use of the monitoring methods to detect impaired size results. Data analysis included 15 herds.

3.5.1 System variance of herd level parameters

Convergence of the variance components for each of the six parameters was observed after 400 iterations of the EM-algorithm (Figure 3.3). Results began to converge already after 200 iterations. The inclusion of Herd 12 resulted in convergence failure; its variance was much lower than others herds so the EM-algorithm failed to converge. This herd was therefore excluded from the variance-covariance estimation procedure. The values of the upper left variance-covariance matrix are shown in Table 3.2. The variances of the model parameters ($\mu_1, \mu_2, \mu_3, \mu_4, \phi_4$) and of the time trend (δ_t), are in the diagonal. Correlations between parities were also computed and are shown below the diagonal. The highest correlation is observed between parity 3 and 4 (0.98), while parity 1 and 4 have the lowest correlation (0.53).

3.5.2 Litter size profiles

The litter size profiles for the 15 herds are shown in Figure 3.4. The profile of most of the herds appears as expected and confirms the shape of the litter size profile (Figure 3.1): a peak in production is observed around parity three and four and followed by a negative slope. At first parity the average (over herds) of the piglets born is 14.7 and the peak (around parity four) is 17.89.

Figure 3.3: Representation of convergence using the EM-algorithm, 400 iterations. (a) Convergence of the system variance in the first four parities. (b) Convergence of system correlations in the first four parities. (c) Convergence of the system variance of the slope and (d) of the trend.

Figure 3.4: Litter size profiles (number of total born piglets per parity) in 15 herds.

Exceptions are found for Herds 1, 5 and 7. In Herd 1, the number of piglets for Parity 2 increases less than expected, affecting the shape of the curve between parities 1 and 3. In Herd 5, an unexpected drop in the number of piglets is observed between parities 2 and 3. Finally, Herd 7 shows a very low productivity compared to the others herds and its slope is less steep.

The values of the model parameters for the 15 herds are shown in Table 3.3. Mean, minimum, maximum and standard deviation are also calculated. The value of trend, reflecting the genetic improvement during the three years considered, is negative for two herds (Herd 7 and 12) and the maximum value (0.0016) is observed for Herd 15.

3.5.3 Model components

A detailed analysis of the DLM components for Herd 15, over the last three years, is shown in Figure 3.5. Results of the filtering for the first four parameters of the model, corresponding to the first four parities, μ_1 to μ_4 , are shown in Figure 3.5(a), and the corresponding smoothed components in Figure 3.5(b). The number of piglets increases from parity 1 to parity 4. The smoothed values of the slope (Figure 3.5(c)) and trend (Figure 3.5(d)) are almost constant.

Table 3.3: Value of the model parameters for the 15 herds. Mean, minimum, maximum and standard deviation are shown in the bottom part.

Herd Number	μ_1	μ_2	μ_3	μ_4	ϕ_4	δ
1	15.07	16.10	17.26	17.12	0.2928	0.00162
2	15.14	17.78	18.85	19.05	0.3389	0.000959
3	13.35	15.89	16.98	17.38	0.2537	0.00151
4	14.59	17.14	18.30	18.76	0.2595	0.00129
5	16.07	17.75	17.55	17.87	0.2214	0.00207
6	14.63	16.77	17.09	17.15	0.4601	0.00125
7	11.65	13.82	14.95	15.57	0.1179	-0.000128
8	14.74	16.69	17.77	18.23	0.33	0.001042
9	14.90	17.05	17.88	18.17	0.2769	0.002202
10	15.79	18.06	19.22	19.59	0.2526	0.0023
11	14.24	16.16	16.84	17.13	0.3694	0.00173
12	15.57	17.12	17.42	17.57	0.3471	-0.000618
13	15.15	17.75	18.43	18.31	0.2452	0.00144
14	14.30	16.43	17.44	17.55	0.2439	0.000407
15	15.76	17.55	18.66	19.00	0.1899	0.00642
Mean	14.73	16.80	17.64	17.89	0.2801	0.00157
Min	11.65	13.82	14.95	15.57	0.1179	-0.000613
Max	16.07	18.06	19.22	19.59	0.4601	0.00642
SD	1.10	1.07	1.03	1.01	0.08	0.0016

Figure 3.5: Evolution of the model parameters for Herd 15 over the last three years. (a) Filtered parity components: $\mu_1, \mu_2, \mu_3, \mu_4$. (b) Smoothed parity components. (c) Smoothed slope (ϕ_4). (d) Smoothed trend (δ).

3.5.4 Sow results

The sows' properties, $M_i(nt)$, are updated on an individual basis and used for forecasting future litter sizes. These estimates are thereafter combined with an economic value (the so-called "retention payoff" or "future profitability") provided by the sow replacement model by Kristensen and Søllested (2004a,b) as described in Section 3.3.3.

Table 3.4 shows the results of eight sows from Herd 1. The sow value, $M_i(nt)$, indicates the difference (in number of piglets) between the individual sow result and the herd mean for the corresponding parity. The first two sows, 2185 and 1349, are both at parity 1. The value 1.13 for sow 2185 indicates therefore that its litter has around 1 piglet more than the general mean, resulting in a positive retention pay-off of 146.8 EUR. The value of the second sow is negative (-1.65), and this results in a negative retention pay-off implying that the sow should be replaced. The next two sows, 1820 and 1344, are both at parity 2. Even though both sow values are positive, a difference of approximately one piglet results in a large difference in term of retention pay-off. For two sows at parity 4, no economic loss is estimated even if the sow value is slightly negative. This is explained by the fact that parity 4 is just after the peak of production (around parity 3) for Herd 1 (Figure 3.4). The low retention pay-off of 2.80 EUR for Sow 2215 with $M_{2215}(4t) = -0.36$ on the other hand reflects that it can be culled with very low economical loss. From parity 5, the number of piglets per parity decreases. This has as a consequence that even if sow has a positive value (Sow 574), the corresponding retention pay-off becomes negative. The same holds for sow 688 at parity 11, of which the negative retention pay-off confirms that it is not profitable to keep a sow for so long even though the sow value is high. This is simply due to the shape of the litter size profile as seen in Figure 3.4.

Table 3.4: Estimated sow value, $M_i(nt)$, and corresponding retention pay-off of eight selected sows from Herd 1.

Sow number, i	Parity, n	Sow value, $M_i(nt)$, piglets	Retention pay-off, EUR
2185	1	1.13	146.80
1349	1	-1.65	-16.55
1820	2	0.386	32.08
1344	2	1.30	152.44
1549	4	2.25	74.54
2215	4	-0.36	2.80
574	5	0.19	-2.24
688	11	1.46	-26.94

Figure 3.6: Monitoring methods applied for Herd 15. (a) Short monitoring period (26 weeks). The central line represents the differences between observed and predicted values. The dotted lines are the upper and lower control limits. (b) Long monitoring period (156 weeks) using a V-mask applied on a Cusum.

3.5.5 Detection of alarms

Monitoring methods were applied both on short- and long term periods. Figure 3.6(a) illustrates the monitoring over a short period (26 weeks) using a Control Chart. Four alarms were observed, at weeks 161, 166, 167 and 168, where the differences between observed and predicted values (plain line) exceeded the control limits (dotted lines). Monitoring over a long period (156 weeks) is performed using a Cusum combined with a V-mask (Figure 3.6(b)). Here, the Cusum shows a level shift around Week 76, which is detected by the V-mask.

An overview of the number of alarms obtained over 156 weeks for the 15 herds is presented in Table 3.5. The total number of alarms using a V-mask is 8. Half of these alarms (4) are due to a negative trend (VM Decrease), which indicate that the observed results have been below what was predicted. For the short term monitoring, 168 alarms are observed, in which 90 of them (CC Decrease) indicate that the observed results were below what was predicted.

3.6 Discussion

Model parameters were pre-estimated using the EM-algorithm. Convergence of the system variance and correlations (Figure 3.3) was obtained when Herd 12 (which resulted in convergence failure) was left out. Since the weekly error terms are independent (cf. Eq.(3.15)), a coefficient of variation (CV) was calculated to make the interpretation of the systems variance size more understandable. As an example, Parity 1, which has a weekly variance of 0.00072 or on annual basis equal to 0.03744, has a corresponding CV, equal to 0.013. This value indicates that the herd average only varies little over time.

Concerning correlations between parities (see Table 3.2 - below the diagonal), very high values were found between Parity 1 and 2 (0.90), Parity 2 and 3 (0.95),

Table 3.5: Number of alarms in a 156 weeks period according to the detection methods: V-mask (VM) and Control Chart (CC).

Herd Number	VM Decrease	VM Increase	CC Decrease	CC Increase
1	0	0	4	2
2	0	1	8	6
3	0	0	1	2
4	0	1	3	5
5	0	0	7	3
6	0	0	4	7
7	1	0	10	4
8	1	0	9	8
9	1	0	7	3
10	0	0	4	4
11	0	0	7	4
12	0	0	2	8
13	0	1	8	9
14	0	1	4	4
15	1	0	12	9
Sum	4	4	90	78

and Parity 3 and 4 (0.98). These high correlations indicate that, for instance, if the herd average for Parity 1 increases, the average of Parity 2 will increase as well. Therefore the shape of the litter size profile will be maintained because individual parity averages will not drift independently of each other.

Filtered and smoothed data for Herd 15 were shown in Figure 3.5. The smoothing reduced the temporary random fluctuation observed in the filtered data. The smoothed data includes all information in the period of time specified, here 156 weeks, and is as such the best possible estimate. Whereas the filtered data obtained for a given week will not change later on, the corresponding smoothed value will change in the light of the later observations. Therefore, the two plots are not directly comparable. If, for instance, we decide to interrupt the analysis at Week 50, the smoothed estimate will be the same as the filtered whereas the smoothing value for that week after 150 observations, will be different. If these plots should be used in practice, the dynamic property of the smoothing should be communicated to the farmer to avoid disorientation in the interpretation of the filtered and smoothed plots.

The value of the slope also contributes to the shape of the herd profile and reflects the persistency of litter size over parities. If the numerical value of the slope is close to zero (cf. Figure 3.5(c)), the productivity of old sows will remain at a high level. If the slope is steep, it reflects a fast decrease in litter size over parities for old sows. The relative value of μ_1 , as compared to the value of μ_2 to μ_4 reflects the performance of the gilts. A low relative value may indicate problems with this

category of animals.

The prior value of the trend used in the paper (0.5 piglet per year) is inspired by the estimated genetic trend reported by Pig Research Centre (2010). This value was estimated in the breeding population and based on the number of live piglets at day five. The initial weekly value in the setting of this paper (δ_0) was therefore set to 0.00961 (0.5/52 weeks). As seen in Table 3.3, the obtained values were lower than expected. The highest value was found in Herd 15 (0.00642), while Herd 7 and 12 had negative values. This tends to indicate that these herds do not exploit the genetic progress obtained in the breeding population. Even though the breeding value may be considered as a upper limit, these results are lower than what may have been expected. On the light of these findings, breeding improvement can be suggested to the farmers.

Results were monitored both in short- and long time periods. In the short term period (26 weeks), alarms were monitored using a Shewhart Control Chart. For Herd 15, four points were outside the control limits (Figure 3.6(a)). Alarms at Weeks 166 and 168 indicate that the weekly production is lower than what was predicted. The farmer may use these alarms to investigate what happened during the insemination process, for instance. In Weeks 161 and 167, the alarms are over the UCL line. The overproduction is positive, therefore these kinds of alarms could be ignored unless the farmer wishes to be informed, in real time, of progress in terms of productivity. The addition of Warning Limits could be used to indicate when results are 'near' to be 'out-of-control'. These warning limits may be added on the Shewhart Control Chart by using a smaller value of a , for instance $a = 1.5$ (Montgomery, 2005).

Table 3.5 shows the number of alarms provided by the two monitoring methods. In the course of three years, the level of production (level shift detected by V-mask) rarely changed. There were only 4 decreasing alarms. On the other hand, the Control Chart, monitoring weekly deviations, gave a larger number of alarms. If only the decreasing alarms (90) are taken into consideration, for a total period of 2340 weeks for the 15 herds, the number of alarms per herd and per year is around 2, indicating that 3% of the monitored weeks resulted in alarm. This can be considered as an acceptable value for practical implementation of this monitoring method. Nevertheless, for practical implementation a calibration of the settings, under known production circumstances, should be performed.

An alternative monitoring method is e.g., a Tabular Cusum, for which upper and lower one-side Cusum are calculated separately. The upper Cusum accumulates the deviations above the target value, the lower, below it, and an alarm occurs when a given threshold is exceeded (Kristensen et al., 2010). This method is also suitable for the present study, since it automatically distinguishes increasing and decreasing performances. A detailed description of Tabular Cusum is found in Montgomery (2005).

The monitoring methods may also be applied on parity specific litter sizes, which can be extracted from the model. This model property is very useful if a more detailed analysis, for a given parity, is needed.

The model also estimates the sows' individual properties, which are used for reliable forecasting of litter sizes. These were in a second step used for computing retention pay-off value. The farmer can therefore combine the obtained information into a decision support model for replacement, as the one proposed by Kristensen and Søllested (2004a,b).

The present study forms the basis for the creation of a larger monitoring system. A similar analysis, at herd and sow level, should be performed using information about conception rate and mortality rate. This would allow a more precise monitoring of, for instance, the number of piglets to sell. The new information available also gaps the current bridge between the data and the decision support models, and as such, increases to the potentiality of the above mentioned decision support model for replacement.

3.7 Conclusion

A system for monitoring litter size at herd and sow level was developed. It is based on a combination of Dynamic Linear Models and methods monitoring systematic deviations, and is as such a useful tool for forecasting and monitoring litter sizes, on a weekly basis. The general profile of litter sizes appear stable and as expected, for most of the 15 herds. The amount of alarms indicating a decreased production level is, in average, two per herd and per year. This tends to indicate that the chosen settings are appropriate and that the model works properly. However, for practical implementation a calibration of the settings, under known production circumstances, is desirable. The inclusion of conception rate and mortality rate should make it a fully functional management tool to monitor and predict productions in a dynamic way.

Conflict of interest statement

The authors report that there are not conflicts of interest relevant to this publication.

Acknowledgements

The authors wish to acknowledge the 15 anonymous farmers as well as the Danish Advisory Center for providing data and the Danish Ministry of Food, Agriculture and Fisheries for financial support for this study through a grant entitled: "Development of a Management System for complete monitoring in Danish and International Pig Production".

Bibliography

- Barnard, G. A., 1959. Control charts and stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 21 (2), pp. 239–271.
- Boehlje, M. D., Eidman, V. R., 1984. *Farm Management*. John Wiley and Sons, New York.
- Cornou, C., Lundbye-Christensen, S., 2010. Classification of sows activity types from acceleration patterns using univariate and multivariate models. *Computers and Electronics in Agriculture* 72 (2), 53 – 60.
- Cornou, C., Lundbye-Christensen, S., Kristensen, A. R., 2011. Modelling and monitoring sows activity types in farrowing house using acceleration data. *Computers and Electronics in Agriculture* 76 (2), 316 – 324.
- De Vries, A., Reneau, J. K., 2010. Application of statistical process control charts to monitor changes in animal production systems. *Journal of Animal Science* 88 (13 electronic suppl), E11–E24.
- Dethlefsen, C., 2001. Space time problems and application. Ph.D. thesis, Aalborg University.
- Doolin, B., 2004. Power and resistance in the implementation of a medical management information system. *Information Systems Journal* 14 (4), 343–362.
- Frost, A., Schofield, C., Beaulah, S., Mottram, T., Lines, J., Wathes, C., 1997. A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture* 17 (2), 139 – 159.
- Huirne, R., Dijkhuizen, A., Renkema, J. A., 1991. Economic optimization of sow replacement decisions on the personal computer by method of stochastic dynamic programming. *Livestock Production Science* 28, 331–347.
- Humphrey, C., 1994. Reflecting on attempts to develop a financial management information system for the probation service in england and wales: Some observations on the relationship between the claims of accounting and its practice. *Accounting, Organizations and Society* 19 (2), 147 – 178.
- Irvin, M., 2008. Success and failure factors of management information systems in the livestock industry of developing countries. In: Tharakan, J., Trimble, J. (Eds.), *3rd International Conference on Appropriate Technology*. Kigale, Rwanda., pp. 282–288.
- Kristensen, A., Jørgensen, E., Toft, N., 2010. *Herd Management Science. II. Advanced topics*. Academic books, Copenhagen.

- Kristensen, A. R., Søllested, T. A., 2004a. A sow replacement model using bayesian updating in a three-level hierarchic markov process: I. biological model. *Livestock Production Science* 87 (1), 13 – 24.
- Kristensen, A. R., Søllested, T. A., 2004b. A sow replacement model using bayesian updating in a three-level hierarchic markov process: II. optimization model. *Livestock Production Science* 87 (1), 25 – 36.
- Montgomery, D., 2005. *Introduction to statistical quality control*, 5th Edition. Wiley, Hoboken, NJ, USA.
- Nasir, S., 2005. The development, change, and transformation of management information systems (mis): A content analysis of articles published in business and marketing journals. *International Journal of Information Management* 25, 442–457.
- Pig Research Centre, 2010. *Annual report 2010*. Tech. rep., Danish Agriculture and Food Council, Copenhagen, Denmark.
- R Development Core Team, 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Sørensen, C., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S., Basso, B., Blackmore, S., 2010. Conceptual model of a future farm management information system. *Computers and Electronics in Agriculture* 72 (1), 37 – 47.
- Toft, N., Jørgensen, E., 2002. Estimation of farm specific parameters in a longitudinal model for litter size with variance components and random dropout. *Livestock Production Science* 77 (2-3), 175 – 185.
- Van Arendonk, J. A. M., 1985. *Studies on the replacement policies in dairy cattle*. II. Optimum policy and influence of changes in production and prices. *Livestock Production Science* 13, 101–121.
- West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models*, 2nd Edition. Springer Series in Statistic. Springer, New York.