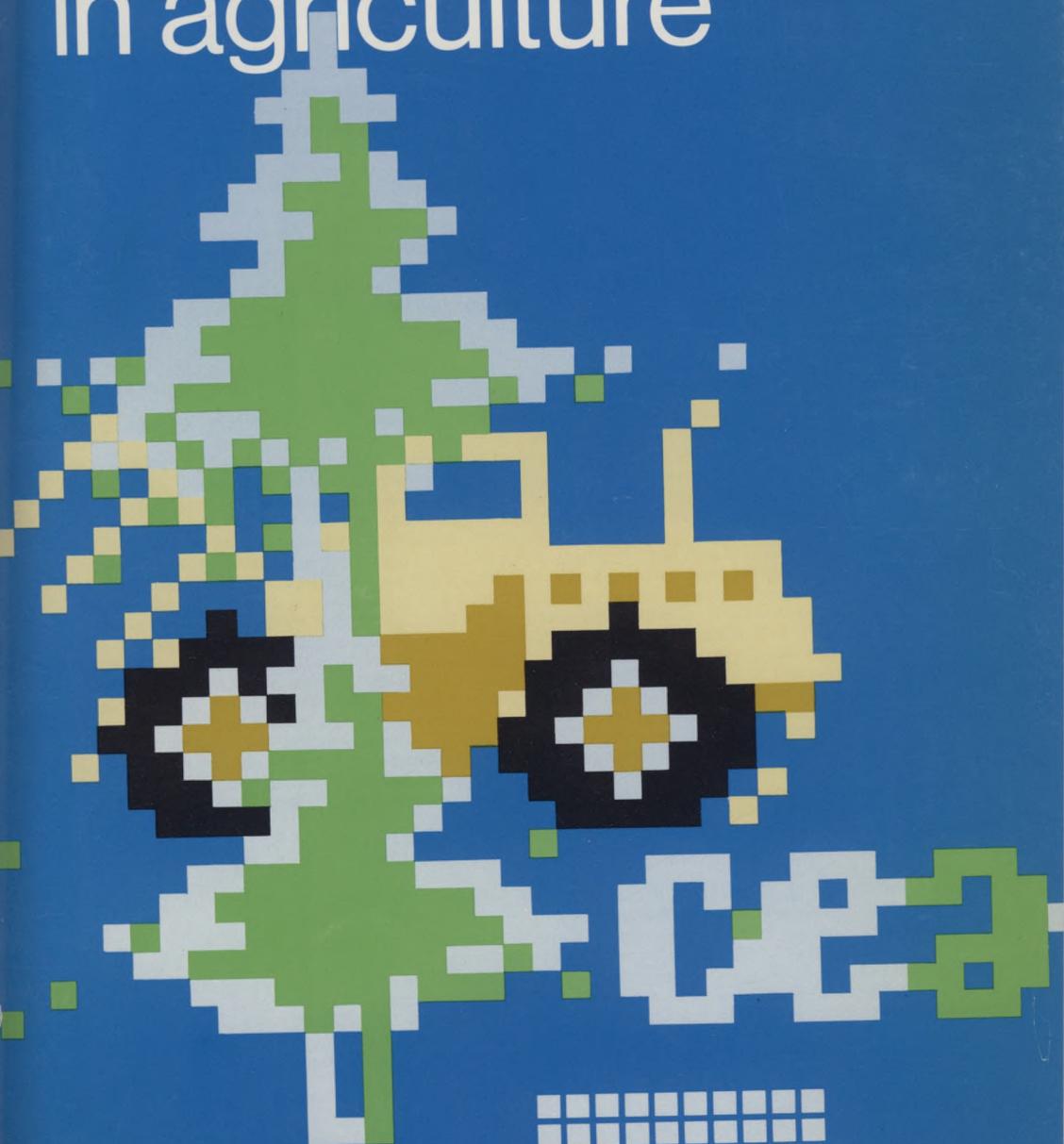
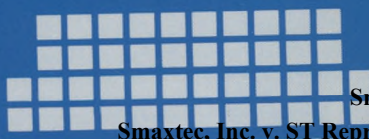


Computers and electronics in agriculture



ELSEVIER



Smaxtec Exhibit 1033
Smaxtec, Inc. v. ST Reproductive Tech, LLC

Page 1 of 21

EDITORS-IN-CHIEF

For the rest of the world

S.W.R. Cox, 18 Lister Avenue, Hitchin, Herts. SG4 9ES, UK
Fax: +44 (1462) 453871; E-mail: sidney_cox@msn.com

For the Americas

D.L. Schmoldt, c/o Brooks Forest Products Center, Virginia Tech, Blacksburg, VA, 24061-0503, USA
E-mail: schmoldt@vt.edu

BOOK REVIEW EDITOR: C.J. Studman, Department of Agricultural Engineering, Massey University, Private Bag 11222, Palmerston North, New Zealand

EDITORIAL BOARD

A.L. Abbott, Blacksburg, Virginia, USA
H. Auerhammer, Freising-Weihenstephan, Germany
G.P.A. Bot, Wageningen, The Netherlands
Y-R. Chen, Beltsville, MD, USA
N.H. Hancock, Toowoomba, Qld., Australia
S.B. Harsh, Michigan, USA
Y. Hashimoto, Matsuyama, Japan
R.B.M. Huirne, Wageningen, The Netherlands
J.W. Hummel, Urbana, IL, USA
F.J. Juste, Valencia, Spain
M. Kassler, Northbridge, NSW, Australia
E. Kline, Blacksburg, VA, USA
F. Kuhlmann, Giessen, Germany
K. Kurata, Tokyo, Japan
C.A. Madramootoo, Beaconsfield, Qué., Canada

J.A. Marchant, Silsoe, UK
J.M. McKinion, Mississippi State, MS, USA
T. Nybrant, Uppsala, Sweden
D. Ordloff, Kiel, Germany
M.J. Pitts, Pullman, WA, USA
R.E. Plant, Davis, CA, USA
W. Rossing, Wageningen, The Netherlands
H. Saarenmaa, Copenhagen, Denmark
J. Schueller, Gainesville, FL, USA
R.P. Singh, Davis, CA, USA
F.E. Sistler, Baton Rouge, LA, USA
J. Stafford, Bedford, U.K.
C.J. Studman, Palmerston North, New Zealand
A.J. Thomson, Victoria, BC, Canada
P.J. Verrier, Harpenden, UK
Mao-hua Wang, Beijing, P.R. China

GENERAL INFORMATION

Audience: All scientists, engineers, environmentalists and economists concerned with applications of computers and electronics to agriculture and related industries.

Aims and Scope: *Computers and Electronics in Agriculture* provides international coverage of advances in the application of computer hardware, software and electronic instrumentation and control systems to agriculture, forestry and related industries. The latter include horticulture (in both its food and amenity aspects), forest products, aquaculture, animal/livestock science, veterinary medicine and food processing. The journal publishes original papers, reviews, applications notes and book reviews on topics including computerized decision-support aids (e.g., expert systems and simulation models) pertaining to any aspect of the aforementioned industries; electronic monitoring or control of any aspect of livestock/crop production (e.g. soil and water, environment, growth, health, waste products) and post-harvest operations (such as drying, storage, production assessment, trimming and dissection of plant and animal material). Relevant areas of technology include artificial intelligence, sensors, machine vision, robotics and simulation modelling.

Publication information: *Computers and Electronics in Agriculture* (ISSN 0168-1699). For 1999 volumes 22, 23 and 24 are scheduled for publication. Subscription prices are available upon request from the Publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries where Air delivery via SAL mail is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, PR China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request. Claims for missing issues should be made within six months of our publication (mailing) date.

Orders, claims and product enquiries: please contact the Customer Support Department at the Regional Sales Office nearest you:

New York: Elsevier Science, P.O. Box 945, New York, NY 10159-0945, USA. Tel. (+1) 212-633-3730 [Toll free number for North American customers: 1-888-4ES-INFO (437-4636)]; Fax (+1) 212-633-3680; e-mail: usinfo-f@elsevier.com

Amsterdam: Elsevier Science, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. Tel. (+31)20-4853757; Fax (+31)20-4853432; e-mail: nlinfo-f@elsevier.nl

Tokyo: Elsevier Science, 9-15 Higashi-Azabu 1-chome, Minato-ku, Tokyo 106-0044, Japan; phone: (+81) (3) 5561 5033; fax: (+81) (3) 5561 5047; e-mail: info@elsevier.co.jp

Singapore: Elsevier Science, No. 1 Temasek Avenue, # 17-01 Millenia Tower, Singapore 039192; phone: (+65) 434 3727; fax: (+65) 337 2230; e-mail: asiainfo@elsevier.com.sg

Rio de Janeiro: Elsevier Science, Rua Sete de Setembro 111/16 Andar, 20050-002 Centro, Rio de Janeiro-RJ, Brazil; phone: (+55) (21) 509 5340; fax: (+55) (21) 507 1991; e-mail: elsevier@campus.com.br [Note (Latin America): for orders, claims and help desk information, please contact the Regional Sales Office in New York as listed above]

US mailing notice: *Computers and Electronics in Agriculture* (ISSN 0168-1699) is published monthly except in May, July and October (total 9 issues) by Elsevier Science (Molenwerf 1, P.O. Box 211, 1000 AE Amsterdam). Annual subscription price in the USA is US\$769.00 (valid in North, Central and South America), including air speed delivery. Periodicals postage paid at Rahway, NJ.

USA POSTMASTER: Send address changes to *Computers and Electronics in Agriculture* (ISSN 0168-1699) c/o Elsevier Science Regional Sales Office, Customer Support Department, 655 Avenue of the Americas, New York, NY 10010. AIRFREIGHT AND MAILING in the USA by Publications Expediting Inc., 200 Main Street, Elmont, NY 11003.

Smactec Exhibit 1033

Smactec, Inc. V. ST Reproductive Tech. LLC

Computers and electronics
in agriculture

RECEIVED

APR 24 1999

STEPHEN & LUCY CHANG

Vol. 22/1 (1999) **SCIENCE LIBRARY**

EDITORIAL BOARD

A.L. Abbott, Blacksburg, Virginia, USA
H. Auerhammer, Freising-Weihenstephan, Germany
G.P.A. Bot, Wageningen, The Netherlands
Y-R. Chen, Beltsville, MD, USA
N.H. Hancock, Toowoomba, Qld., Australia
S.B. Harsh, Michigan, USA
Y. Hashimoto, Matsuyama, Japan
R.B.M. Huirne, Wageningen, The Netherlands
J.W. Hummel, Urbana, IL, USA
F.J. Juste, Valencia, Spain
M. Kassler, Northbridge, N.S.W., Australia
E. Kline, Blacksburg, VA, USA
F. Kuhlmann, Giessen, Germany
K. Kurata, Tokyo, Japan
C.A. Madramootoo, Beaconsfield, Québec, Canada
J.A. Marchant, Silsoe, UK
J.M. McKinion, Mississippi State, MS, USA
T. Nybrant, Uppsala, Sweden
D. Ordolff, Kiel, Germany
M.J. Pitts, Pullman, WA, USA
R.E. Plant, Davis, CA, USA
W. Rossing, Wageningen, The Netherlands
H. Saarenmaa, Copenhagen, Denmark
J. Schueller, Gainesville, FL, USA
R.P. Singh, Davis, CA, USA
F.E. Sistler, Baton Rouge, LA, USA
J. Stafford, Bedford, U.K.
C.J. Studman, Palmerston North, New Zealand
A.J. Thomson, Victoria, BC, Canada
P.J. Verrier, Harpenden, UK
Mao-hua Wang, Beijing, P.R. China

Abstracts/Contents List published in: Agricola database, Agricultural Engineering Abstracts, Biological Abstracts, Computer & Control Abstracts, Computer and Information Systems Abstracts Journal, Current Contents (Agricultural Science), Electrical & Electronics Abstracts, Engineering Index, Geobase, Geographical Abstracts, Human Geography

Computers and electronics in agriculture

EDITORS-IN-CHIEF

For the rest of the world

S.W.R. Cox
Manor Close
18 Lister Avenue
Hitchin Herts. SG4 9ES
UK

For the Americas

D.L. Schmoldt
c/o Brooks Forest Products Center
Virginia Tech
Blacksburg, VA
24061-0503
U.S.A.

BOOK REVIEW EDITOR

C.J. Studman
Department of Agricultural Engineering
Massey University
Private Bag 11222
Palmerston North
New Zealand

VOLUME 22 (1999)



ELSEVIER

Amsterdam—Lausanne—New York—Oxford—Shannon—Tokyo

Smaxtec Exhibit 1033

Smaxtec, Inc. v. ST Reproductive Tech, LLC

Page 5 of 21

© 1999 Elsevier Science B.V. All rights reserved.

This journal and the individual contributions contained in it are protected under copyright by Elsevier Science B.V., and the following terms and conditions apply to their use:

Photocopying

Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier Science Rights & Permissions Department, PO Box 800, Oxford OX5 1DX, UK; phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.co.uk. You may also contact Rights & Permissions directly through Elsevier's home page (<http://www.elsevier.nl>), selecting first 'Customer Support', then 'General Information', then 'Permissions Query Form'.

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (978) 7508400, fax: (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (+44) 171 436 5931; fax: (+44) 171 436 3986. Other countries may have a local reprographic rights agency for payments.

Derivative Works

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the publisher is required for resale or distribution outside the institution.

Permission of the publisher is required for all other derivative works, including compilations and translations.

Electronic Storage or Usage

Permission of the publisher is required to store or use electronically any material contained in this journal, including any article or part of an article. Contact the publisher at the address indicated.

Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher.

Address permissions requests to: Elsevier Science Rights & Permissions Department, at the mail, fax and e-mail addresses noted above.

Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a Kalman filter

R.M. de Mol ^{a,*}, A. Keen ^{a,b}, G.H. Kroeze ^a,
J.M.F.H. Achten ^a

^a DLO Institute for Agricultural and Environmental Engineering (IMAG-DLO), P.O. Box 43,
6700 AA Wageningen, The Netherlands

^b Centre for Biometry (CPRO-DLO), Wageningen, The Netherlands

Abstract

Sensor measurements can be used in dairy farming for the detection of oestrus and diseases. A new model has been developed to process the measured variables in a combined way. It is based on time series models for milk yield, milk temperature, electrical conductivity of quarter milk and the cow's activity, and a probability distribution for the concentrates leftovers. The parameters of the time series models and the probabilities are fitted on-line for each cow after each milking by Kalman filters. This makes it possible to combine the variables and to generate cow-specific alerts. Global results on the detection of oestrus, mastitis and other diseases are given. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Dairy cattle; Oestrus detection; Health monitoring; Time series; Kalman filter; Management information systems

1. Introduction

Timely recognition of oestrus and diseases is very important in dairy farming. Oestrus detection is important because it determines the insemination time and as a derivative of this the interval between two successive calvings (calving interval). It is traditionally done by visual observations by the farmer. Cows in oestrus behave

* Corresponding author. Tel.: +31-317-476459; fax: +31-317-425670.

E-mail address: r.m.demol@imag.dlo.nl (R.M. de Mol)

differently, they are more active and stand to be mounted. Visual observation has become more difficult as the average herd size has increased. Therefore alternative methods have been developed that may be automated (De Mol et al., 1993).

Detection of diseases is also important, not only because an ill cow produces less milk but also because a disease can have harmful consequences; it may be a reason for culling animals. Mastitis (udder inflammation) is a frequently occurring disease that can lead to considerable yield reductions (Houben, 1995). Automated methods have also been developed for detection of diseases.

Several methods for automated detection of oestrus and diseases that can already be used in practice, are based on measurement of variables with sensors. Sensors are available for measuring milk yield, milk temperature, electrical conductivity of the milk, animal activity (with step counters) and concentrates intake.

The qualitative relationships between the measured variables and the occurrence of oestrus and diseases are shown in Table 1. This shows that temperature and activity are increased in case of oestrus, yield and feed intake may be decreased and conductivity is unchanged. The conductivity increases in case of mastitis.

Information from a management information system (MIS), such as the number of days in lactation, previous cases of oestrus and diseases, is useful for the interpretation of the measurements.

Models based on one variable have been developed in previous research for the detection of oestrus and diseases: for example the activity for oestrus (Eradus et al., 1992) or the conductivity for mastitis. Different variables have also been taken into account separately (Maatje et al., 1992).

It is clear from Table 1 that there is a significant potential for improvement by considering the combination of variables for the interpretation of the measurements. An increased temperature can have different causes, but when coupled with an increased activity, oestrus will be an obvious explanation; when it coincides with an increased conductivity, mastitis might be the reason. Therefore research has been carried out in which sensor measurements from the different variables and information from the MIS are processed in a combined way. This leads to a detection model for oestrus and diseases that can be a part of a MIS (De Mol et al., 1992).

Table 1
The relations between measured variables and the occurrence of oestrus and diseases^a

	Yield	Temperature	Conductivity	Activity	Feed intake
Oestrus	Neg/-	Pos	-	Pos	Neg/-
Mastitis	Neg	Pos	Pos	-	Neg/-
Other infective diseases	Neg	Pos	-	Neg	Neg/-
Metabolic diseases	Neg	-	-	Neg	Neg
Lameness	Neg	-	-	Neg	Neg

^a Neg, a significant negative influence; Pos, a significant positive influence; -, no influence (after Hogewerf et al., 1992).

2. The structure of the model

The detection model should generate alerts for oestrus and diseases (especially mastitis) based on sensor measurements and information from the MIS. These alerts are meant for the farmer to draw his attention to a cow that may be in oestrus or ill so that he can undertake appropriate action. For each cow and each milking, measurement data are available for:

- milk yield,
- milk temperature,
- electrical conductivity of the milk for each quarter of the udder, and
- activity based on the counter values of the step counter, and for each cow and each day:
- the concentrates intake and the ration.

Measurements are available from the experimental farms of IMAG-DLO in Duiven and from ID-DLO in Lelystad obtained in 1993 and 1994. The cows were milked twice a day. Maatje et al. (1992) describe the measuring methods. Reference data are available for testing: progesterone, somatic cell counts and others from laboratory analyses, veterinary treatments, and so on.

Several processing techniques are suitable for the development of such a model. In the past simple statistical techniques were used, such as the moving average. A structure that is based on more advanced statistical techniques, namely time series analysis combined with a Kalman filter, is used in this paper. Time series analysis has already been used for milk yield in Deluyker et al. (1990) where a generally applicable model has been proposed. A cow-dependent, but generally applicable, model is described here. The Kalman filter has also been used in a somewhat comparable research (Thysen, 1992), but the approach in this paper is fundamentally different.

The detection model uses underlying models that describe the 'normal' behaviour of the measured variables. These underlying models are cow-specific and estimates of parameters are updated after each milking. For each cow and each milking the following steps are taken:

1. use of the underlying models to calculate predictions for the measurements with standard errors;
2. reading of the actual new measurements;
3. comparison of the actual and the predicted values and generation of an alert if the combination of variables is aberrant; and
4. use of the new information from the measurements to update the parameter estimates in the underlying models.

In this way each cow gets her own model describing her characteristics. This makes it possible to generate cow-specific alerts in case of abnormal behaviour, possibly due to oestrus or illness. The underlying models for yield, temperature, conductivity and activity are time series models (described in Section 3), for the concentrates intake a probability distribution is used (Section 4). Kalman filters are used to update the parameters in these models (Section 5).

This approach has not been used before for the development of detection models. Similar approaches can be found in other fields, e.g. for condition monitoring in an industrial plant in Christer et al. (1997), for a river-flow forecasting model in Awwad et al. (1994) and Bidwell and Griffith (1994), for estimating dynamic tree ring climate relationships in Van Deusen (1990), for gas transport processes in Federspiel (1997) and for groundwater monitoring networks in Van Geer (1987).

3. Time series models for cow variables

Time series, like the sensor measurements, are observations of a phenomenon made sequentially in time (Chatfield, 1989). Consequently, the measurements of the cow variables are time series. A characteristic of time series is the fact that in general the successive observations are not independent. This relationship is made explicit in a time series model, which is used to forecast the measurement values for a next milking. The new measurements can then be compared with these forecasts. It is assumed that the model is valid for healthy cows that are not in oestrus. Too great deviations indicate that this assumption is no longer valid.

The usability of time series models for the measured cow variables has been examined. A model has been searched for each variable by following the standard procedure: plot the data, examine the correlograms of the autocorrelations and partial autocorrelations, select an appropriate ARIMA model (AutoRegressive Integrated Moving Average model) and fit the chosen model. This procedure has been applied for the cow variables. Appropriate time series models have been found for the cow variables milk yield, milk temperature, electrical conductivity of the milk and the cow's activity.

3.1. Yield

The measured yield is influenced by the length of the interval since the previous milking and the diurnal rhythm. A farmer is mostly concerned with the daily milk yield. An approximation of the daily milk yield based on the two latest milkings is used:

$$Y_{D,n} = (Y_{M,n} + Y_{M,n-1}) \cdot \frac{24}{24 + M_n - M_{n-2}} \quad (1)$$

with: n , latest milking; $n-1$, previous milking,...; $Y_{D,n}$, daily yield at milking n ; $Y_{M,n}$, yield at milking n ; and M_n , decimal time of milking n (between 0 and 24 h).

For the differences of successive daily yields, the following moving average (MA) model is used:

$$\nabla Y_{D,n} = Y_{D,n} - Y_{D,n-1} = Z_{Y,n} - \alpha_Y \cdot Z_{Y,n-2} \quad (2)$$

with: $\nabla Y_{D,n}$, difference of daily milk yield at milking n ; $Z_{Y,n}$, random disturbance on yield at milking n ; α_Y , parameter of yield model.

The disturbances $Z_{Y,n}$ (zero means, normally distributed) are calculated recursively, the parameter α_Y must be estimated. The last term in Eq. (2) is needed to compensate for the artificial autocorrelations introduced by Eq. (1).

3.2. Temperature

The temperature fluctuates during the day. Therefore comparison with the previous milking is not useful. An MA model for the differences of the milk temperature with two milkings ago is used:

$$\nabla T_n = T_n - T_{n-2} = Z_{T,n} - \alpha_T Z_{T,n-2} \quad (3)$$

with: ∇T_n , difference of milk temperature with lag 2 at milking n ; T_n , milk temperature at milking n ; $Z_{T,n}$, random disturbance on temperature at milking n ; α_T , parameter of temperature model.

3.3. Activity

The activity depends on the diurnal rhythm of the cow. To compensate for this diurnal effect, the hourly activity for each milking is calculated, based on the difference of the two counter values (cumulatives ranging from 0 to 999) and the interval:

$$A_{H,n} = \frac{V_n - V_{n-1}}{M_n - M_{n-1}} \quad (4)$$

with: $A_{H,n}$, hourly activity at milking n ; V_n , counter value at milking n (differences are taken modulo 1000); M_n , decimal time of activity measurement (differences are taken modulo 24.0).

For the differences in hourly activity an MA model is used:

$$\nabla A_{H,n} = A_{H,n} - A_{H,n-2} = Z_{A,n} - \alpha_A Z_{A,n-2} \quad (5)$$

with: $\nabla A_{H,n}$, difference of hourly activity with lag 2 at milking n ; $Z_{A,n}$, random disturbance on activity at milking n ; α_A , parameter of activity model.

As with the yield model, the last term in Eq. (5) is introduced to compensate for the artificial autocorrelations introduced by Eq. (4).

3.4. Conductivity

The electrical conductivity of the quarter milk depends mostly on the conductivity at the preceding milkings. Therefore an autoregressive (AR) model is used for the conductivity:

$$E_{q,n} - \mu_C = \alpha_C (E_{q,n-1} - \mu_C) + \beta_C (E_{q,n-2} - \mu_C) + Z_{Cq,n} \quad (6)$$

with: $E_{q,n}$, electrical conductivity of quarter q at milking n ; μ_C , is the average conductivity of each quarter (parameter of conductivity model); α_C and β_C , parameters of conductivity model; $Z_{Cq,n}$, random influence on conductivity of quarter q at milking n .

The same parameters, μ_C , α_C and β_C , are assumed to be valid for each quarter.

It is possible to forecast new measurement values if the values of the parameters are known. However, after fitting the models, these parameters appeared to be different for each cow and also different for successive lactations of the same cow. Therefore the parameter values should be calculated for each cow and each lactation separately. With standard techniques this is only possible at the end of a lactation, which is undesirable for practical application because results are needed during the current lactation. Application of a Kalman filter can relieve this problem (Section 5.2).

4. A stochastic model for the concentrates leftovers

The concentrates leftovers are not included in the detection model by a time series model. This variable has a different behaviour: it mostly equals zero and is sometimes higher. Therefore a different approach is used. It is assumed that successive leftovers are independent and there is a probability distribution for L , the percentage of the leftover of the concentrates ration, defined by:

$$p_{0,} = P(L = 0\%)$$

$$p_{1,} = P(0\% < L \leq 10\%)$$

$$p_{2,} = P(10\% < L \leq 30\%)$$

$$p_{3,} = P(30\% < L \leq 50\%)$$

$$p_{4,} = P(50\% < L \leq 100\%).$$

This distribution can be used to calculate the probability $P(L \geq L_n)$ of the actual leftovers L_n at a certain milking n . An alert for low concentrates intake will be given when this probability is low.

This distribution is, however cow-dependent. For some cows the leftovers are zero at most times, for other cows the leftovers are quite often greater than zero. A Kalman filter is used to fit the distribution for each individual cow (Section 5.3).

5. The Kalman filter

5.1. General description

A Kalman filter is applied because the parameters in the models for the different variables are cow-dependent and a model for the dependency between the variables is wanted. It is a method to estimate the state of a system on-line. The state is a quantity that determines the coming behaviour of the system. The estimate of the state is adjusted after each new observation by using the new information. First, a general description is given and later two applications where:

1. the state consists of the parameters in the time series models (Section 5.2);

2. the state consists of the probability distribution of the percentage of the concentrates leftover (Section 5.3).

The system must be described with state-space equations to apply the Kalman filter, consisting of an observation equation:

$$y_n = C_n \cdot x_{n-1} + v_n \quad (7)$$

and a system equation:

$$x_n = A_n \cdot x_{n-1} + w_n \quad (8)$$

In these equations x_n is the state vector, y_n the observation vector, C_n and A_n are system matrices, v_n is the random observation error and w_n is the random system error. The observation Eq. (7) describes the relationship between the measurements and the state: the state itself is not directly measurable in general. The system Eq. (8) gives the relation between the states at successive times. The distribution of v_n is $N(0, V_n)$ and of w_n is $N(0, W_n)$.

In general the estimate of the state x_n at time n using the observations $y_1, \dots, y_{n-1}$ is desired. The Kalman filter can be applied when a system is described with state equations (Harrison and Stevens, 1976; Harvey, 1989). It gives a new estimate of the state x_n after each observation and furthermore a variance-covariance matrix P_n for the state estimate.

The Kalman filter is an estimation procedure with two stages:

Stage 1 (prediction stage) is an estimate of the state based at the previous state:

$$\hat{x}_{n|n-1} = A_n \cdot \hat{x}_{n-1} \quad (9)$$

with variance-covariance matrix:

$$P_{n|n-1} = A_n \cdot P_{n-1} \cdot A_n^T + W_n \quad (10)$$

where: $\hat{x}_{n|n-1}$, estimate of state x at time n using all information up to time $n-1$; $\hat{x}_{n-1}$, estimate of state x at time $n-1$ using all information up to time $n-1$; $P_{n|n-1}$, estimate of the variance-covariance matrix P at time n using all information up to time $n-1$; P_{n-1} , estimate of the variance-covariance matrix P at time $n-1$ using all information up to time $n-1$.

Stage 2 (updating stage) updates the estimate with the observation y_n ; the estimation error is:

$$e_n = y_n - C_n \cdot \hat{x}_{n|n-1} \quad (11)$$

where: e_n , the estimation error at time n .

This gives an improved estimate of the state:

$$\hat{x}_n = \hat{x}_{n|n-1} + K_n \cdot e_n \quad (12)$$

with variance-covariance matrix:

$$P_n = P_{n|n-1} - K_n \cdot C_n \cdot P_{n|n-1} \quad (13)$$

where:

$$K_n = P_{n|n-1} \cdot C_n^T [C_n \cdot P_{n|n-1} \cdot C_n^T + V_n]^{-1} \tag{14}$$

The resulting estimates can be used in the next time step. The matrix K_n is called the Kalman gain: it gives the influence of the error at time n on the state estimate, see Eq. (12). This is also the influence of the current observation, as can be derived from Eqs. (11) and (12):

$$\hat{x}_n = K_n \cdot y_n + (I - K_n \cdot C_n) \cdot \hat{x}_{n|n-1} \tag{15}$$

where I is the identity matrix.

Harvey (1989) proves that the Kalman filter gives the minimum mean square linear estimator (MMSLE) of x_n . P_n is the unconditional variance-covariance matrix of the estimation error when estimating x_n . The variance-covariance matrix, Σ_n , of e_n is given by:

$$\Sigma_n = C_n \cdot P_{n|n-1} \cdot C_n^T + V_n \tag{16}$$

when the system matrices are fixed and known. Duncan and Horn (1972) show that even if the error vectors are not normally distributed, the Kalman filter estimator will still be the MMSLE provided the v_n and w_n are independent vectors with mean zero.

5.2. Fitting the parameters of the time series models

In standard usage of the Kalman filter the state is used to model the measured variables, e.g. the level and trend of a variable. In that case the level and trend are included in the state vector. The Kalman filter is here used to estimate the parameters of the time series models of the cow variables, therefore the state consists of these parameters. The Kalman filter gives a new estimate of the state after each milking, which means new estimates of the parameters of the time series models. With these parameters new measurement values are forecasted so that highly deviant measurements can be signalised. Also the variance-covariance matrix of the estimated state is given, this is used to relate the variables mutually. We apply the following definitions:

$$x_n = \begin{bmatrix} -\alpha_Y \\ -\alpha_T \\ -\alpha_A \\ \mu_C(1 - \alpha_C - \beta_C) \\ \alpha_C \\ \beta_C \end{bmatrix}, \quad y_n = \begin{bmatrix} \nabla Y_{D,n} \\ \nabla T_n \\ \nabla A_{H,n} \\ E_{rh,n} \\ E_{rf,n} \\ E_{lf,n} \\ E_{lh,n} \end{bmatrix}, \quad z_n = \begin{bmatrix} Z_{Y,n} \\ Z_{T,n} \\ Z_{A,n} \\ Z_{Crh,n} \\ Z_{Crf,n} \\ Z_{Clf,n} \\ Z_{Clh,n} \end{bmatrix},$$

$$C_n = \begin{bmatrix} Z_{Y,n-2} & 0 & 0 & 0 & 0 & 0 \\ 0 & Z_{T,n-2} & 0 & 0 & 0 & 0 \\ 0 & 0 & Z_{A,n-2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & E_{rh,n-1} & E_{rh,n-2} \\ 0 & 0 & 0 & 1 & E_{rf,n-1} & E_{rf,n-2} \\ 0 & 0 & 0 & 1 & E_{lf,n-1} & E_{lf,n-2} \\ 0 & 0 & 0 & 1 & E_{lh,n-1} & E_{lh,n-2} \end{bmatrix} \quad (17a)$$

$$A_n = I, \quad v_n = z_n, \quad w_n = 0 \quad (17b)$$

where the abbreviations rh (right hind), rf (right front), lf (left front) and lh (left hind) are used for the four quarters.

Using this definition, the state space equations Eqs. (7) and (8) are in fact a reformulation of the time series models as defined in Eqs. (2), (3), (5) and (6), which makes it possible to apply the Kalman filter. The matrix W_n is taken equal to the zero matrix, as we suppose that the parameters are fixed but unknown for an individual cow (Eq. (8)). The matrix V_n (the variance of the observation errors, needed in Eq. (14)) is calculated by exponential smoothing using actual values of v_n as defined in Eqs. (17a) and (17b). Elements of V_n also are taken as zero if interrelationships between measurements of variables are not plausible. For example there is no reason to suppose a relationship between measurement of yield and activity. Only influences between the measurements of the conductivity of different quarters seem possible. This means V_n has the following form:

$$V_n = \begin{bmatrix} v_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & v_{22} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & v_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & v_{44} & v_{45} & v_{46} & v_{47} \\ 0 & 0 & 0 & v_{54} & v_{55} & v_{56} & v_{57} \\ 0 & 0 & 0 & v_{64} & v_{65} & v_{66} & v_{67} \\ 0 & 0 & 0 & v_{74} & v_{75} & v_{76} & v_{77} \end{bmatrix} \quad (18)$$

5.3. Fitting the probability distribution in the concentrates leftover model

Again, a Kalman filter is applied. A description with state space equations Eq. (7) and Eq. (8) is needed for this. In this case the following definitions apply:

$$x_n = [p_0 \quad p_1 \quad p_2 \quad p_3 \quad p_4]^T, \quad y_n = [r_0 \quad r_1 \quad r_2 \quad r_3 \quad r_4]^T, \quad A_n = I, \quad C_n = I \quad (19)$$

The vector x_n is the state, here the probability distribution (Section 3), y_n is the observation with r_i defined as:

$$\begin{array}{ll}
\text{if } L_n = 0\% & r_0 = 1, r_i = 0 \text{ if } i \neq 0, \\
\text{if } 0\% > L_n \leq 10\% & r_1 = 1, r_i = 0 \text{ if } i \neq 1, \\
\text{if } 10\% > L_n \leq 30\% & r_2 = 1, r_i = 0 \text{ if } i \neq 2, \\
\text{if } 30\% > L_n \leq 50\% & r_3 = 1, r_i = 0 \text{ if } i \neq 3, \\
\text{if } 50\% > L_n \leq 100\% & r_4 = 1, r_i = 0 \text{ if } i \neq 4.
\end{array}$$

The matrices A_n and C_n are equal to the identity matrix I , $V_n = I$ and $W_n = 0.01 \cdot I$. With this definitions the estimation error is:

$$e_n = [r_0 - p_0 \quad r_1 - p_1 \quad r_2 - p_2 \quad r_3 - p_3 \quad r_4 - p_4]^T \quad (20)$$

A component of e_n is positive when $r_i = 1$ and negative when $r_i = 0$. Starting values for the probability distribution are based on observed distributions, the starting value for P_0 is taken $0.1 \cdot I$.

In this case the Kalman filter can be rewritten as:

$$\hat{x}_n = \hat{x}_{n-1} + K_n \cdot e_n, \quad P_n = (I - K_n) \cdot (P_{n-1} + 0.01 \cdot I), \quad K_n = f_n \cdot I \quad (21)$$

where the factor f_n can be calculated recursively: $f_n = \frac{f_{n-1} + 0.01}{f_{n-1} + 1.01}$, $f_0 = 0.1$

6. Detection method

The detection model is meant to draw the attention of the farmer to possible deviations in his cows; the model should generate alerts for that purpose. Alerts can be generated based on yield, temperature, conductivity and activity with the time series models for these variables. With the help of the time series models together with a Kalman filter for each milking an estimate of the observation is available following the observation equation Eq. (7). The estimate is compared with the real measurement to get the error vector e_n . The estimate of the state based on the measurements up to the preceding milking, is used for this. A normal distribution is assumed for e_n . The variance-covariance matrix of e_n can also be calculated (Eq. (16)). This matrix is used to standardise e_n . The stochastic model together with the Kalman filter gives the probability of the actual concentrates leftover.

There are two methods to generate alerts:

1. Single alerts: each component of the standardised error vector has a standard-normal distribution. Observations outside some confidence intervals result in an alert. An alert can correspond with: errors outside the 95% confidence interval, errors outside the 99% interval and errors outside the 99.9% interval. A single alert for the concentrates leftover can be given when the calculated probability is below 5, 1 or 0.1%.
2. Combined alerts: the components of the standardised error vector are mutually comparable, due to the special form chosen for V_n (Eq. (18)) This makes it possible to consider combinations of the elements. Alerts correspond with

combinations falling outside a 95%, a 99% and a 99.9% confidence region. An oestrus alert is given when the activity is rather high and the sum of standardised errors of activity, yield and temperature falls outside a certain confidence interval. A mastitis alert is given when the conductivity error is rather high and the sum of standardised errors of conductivity, yield and temperature falls outside a certain confidence interval. An illness alert is based on the sum of standardised errors of yield, temperature, activity and concentrates intake.

7. Adjustments for practical implementation

The Kalman filter as described in the previous section is adjusted in the practical implementation:

- The update of the state as defined in the updating stage, Eq. (12), is limited to prevent unwanted effects caused by start-up effects or strong deviating measurements. The updating stage is modified such that limits can be set. In practical applications the absolute change in Eq. (12) is limited to 0.1.
- The update of the matrix V_n is also limited to prevent too great changes in one step. A value of v_n outside the 99% confidence interval is replaced by the value on the border of this confidence interval.
- There are also several possibilities for applying the Kalman filter. The Kalman filter may be used only when measurements of all variables are available or used when at least one variable is measured correctly. In the latter case it is used to improve only the parameters of the ARIMA models of the right variables. Furthermore, the Kalman filter may be used only when no alerts are given or used also when there are alerts on some variables. The ARIMA model is suitable for healthy cows. A cow may be sick (or in oestrus) in case of alerts, so applying the Kalman filter may lead to wrong parameters. In both applications a Kalman filter is used when at least one variable is measured correctly and in all cases, with or without alerts. In this way all information is used as much as possible and the model adapts if the circumstances change, e.g. when the cows go to the pasture in spring.
- Measurement errors can give problems, not only for the current milking but also for following milkings, as follows from Eqs. (1), (3), (4) and (6). For example, for the calculation of the daily milk yield two successive measurements are needed (Eq. (1)). To prevent the consequences of measurement errors for the next milking the expected value of y_n is calculated and used as a substitute value for the missing measurement. These substitute values are used in cases with only one successive missing measurement. No substitute values are used when there are more missing measurements in a row.
- Measurement errors for the activity can result in false counter values and thus in great differences in Eq. (4) and wrong alerts. It is possible to neglect counter values, which are apparently coupled with a wrong cow number.

Table 2

The sensitivity and specificity for oestrus based on 537 oestrus cases and 41 803 milkings outside oestrus periods

Alerts (confidence interval, %)	Sensitivity (%)	Specificity (%)
95	94.2	94.5
99	86.5	96.9
99.9	82.5	98.1

- A combined alert is based on a combination of errors of different variables. The detection method is adapted if some variables are missing. An alert is also given when a combination of some variables minus one exceeds a similar threshold: e.g. an oestrus alert in case of increased activity and decreased yield but also a lower temperature.

8. Illustration of outcomes

The described model has been implemented and tested on two experimental farms (of IMAG-DLO and ID-DLO). Cases of oestrus or disease that were signalled by the model are true positive (TP), not signalled cases are false negative (FN). Milkings outside an oestrus or illness period are true negative (TN) if there is no alert from the model, otherwise they are false positive (FP). The model performance was expressed in the sensitivity and the specificity.

The sensitivity is the percentage of truly signalled cases: $(TP/(TP + FN)) \cdot 100\%$.

The specificity is the percentage of truly not signalled milkings outside oestrus or disease periods: $(TN/(FP + TN)) \cdot 100\%$. The specificity for mastitis is calculated by regarding cows without any mastitis case during the test period.

Some results are given in Tables 2–4; more results can be found in De Mol et al. (1997). The results depend on the chosen confidence interval. Tightening the criterion leads to a lower sensitivity and higher specificity, and vice versa. Results for oestrus are satisfying, as well as the sensitivity for mastitis. The number of false positives (as implied by the specificity) may be too high for practical implementation. The results for diseases are promising but further research is needed.

9. Discussion

The detection model is based on cow-specific time series models. Specific time series models (moving average, exponential smoothing) have been used more in detection models. Here a selection was made by a systematic search within the class of ARIMA models, resulting in a suitable MA or AR model for each variable. A general model that can be used for each cow as in Deluyker et al. (1990) was not

Table 3

The sensitivity for clinical and subclinical mastitis and the specificity for mastitis

Alerts (confidence interval, %)	Sensitivity, clinical mastitis (52 cases) (%)	Sensitivity, subclinical mastitis (21 cases) (%)	Specificity (6495 milkings) (%)
95	96	100	95.3
99	90	76	98.2
99.9	65	57	99.4

searched for here. The use of the Kalman filter makes it possible to work with cow-specific parameters for the time series models. The filter gives for each cow after each milking an estimate of the parameters that describe the normal behaviour of the cow. The model is no longer valid if new measurements widely deviate from the forecast because the cow is in oestrus or ill. An alert is given in that case.

A Kalman filter is also applied by Thysen (1992) to model the somatic cell count of milk. He has a general model for all cows and uses a 'multi-state' model in which a cow can have three possible states: normal level, an outlier or a change of level. A normal behaviour is assumed here and deviant measurements do not fit in our model. The application of time series analysis with a Kalman filter is new for a detection model. A similar approach in other fields to the use of a Kalman filter can be found in the literature as described in Section 2.

The model is developed to detect short-term changes in cow variables. After a few days the model will be adapted to the new situation. This feature will be in general advantageous because changes in grazing system, lactation stage and the like, should not result in alerts. Slow changes will be adapted by the model without generating alerts. This means that the model may not detect some diseases because the symptoms appear slowly.

A combined alert is given when the error for activity (for oestrus) or conductivity (mastitis) is high and the sum of standardised errors is outside a confidence interval. Other possibilities may lead to improved detection results: changing the threshold for the error for activity or conductivity, or changing the relative weight of variables in the sum, or excluding a variable or including new variables in the sum.

Table 4

The sensitivity for diseases (mastitis excluded) and specificity of the detection model, based on 263 cases and 40 286 milkings outside illness periods

Alerts (confidence interval, %)	Sensitivity (%)	Specificity (%)
95	99.6	86.0
99	90.5	93.5
99.9	76.8	96.7

The detection model gives each cow her own model, independent of other cows. Group effects are not taken into account. The number of FP alerts can be reduced by looking at the group effects. Oestrus might not be the reason if all cows have an increased activity, so no oestrus alerts should be given in that case.

The model is based on a milking frequency of twice a day. This is used in the time series model to include diurnal effects. Adaptation for a milking frequency of three times a day is straightforward. Adaptation to a variable milking frequency (in systems with a milking robot) is less apparent.

10. Conclusions

The existing detection models are mostly based on a moving average or exponential smoothing: these can be considered as specific time series models. The described detection model is for most variables based on time series models combined with a Kalman filter to estimate the parameters on-line and to be able to consider the mutual connection. The application of time series models gives more selection possibilities and can thus lead to better models. The distribution of the concentrates leftovers is an appropriate model for this variable. The Kalman filters make it possible to adapt the model on-line.

The results of the detection model are based on a comparison with reference data. The sensitivity is high (but depending on the chosen criterion). The specificity also seems high but may be too low for practical application, therefore additional research is directed to a reduction of the number of false positive alerts.

Further research is directed to an adaptation of the detection model for other milking frequencies (in case of an automatic milking system) and reducing the number of false positive alerts by a further processing of the standardised errors.

Acknowledgements

This paper is dedicated to co-author A. Keen, who passed away on November 7th, 1996. Bertus Keen came up with the idea of applying a Kalman filter for this purpose. This study was a co-operation among Alfa Laval Agri in Sweden, IMAG-DLO and DLO Research Institute for Animal Husbandry and Animal Health (ID-DLO) in The Netherlands.

References

- Awwad, H.M., Valdes, J.B., Restrepo, P.J., 1994. Streamflow forecasting for Han river basin. *Korean J. Water Res. Pl.-ASCE* 120, 651–673.
- Bidwell, V.J., Griffith, G.A., 1994. Adaptive flood forecasting: an application to the Waimakarini River. *J. Hydrol.-NZ* 32 (2), 1–15.
- Chatfield, C., 1989. *The Analysis of Time Series. An Introduction*, 4th ed. Chapman and Hall, London 241 pp.

- Christer, A.H., Wang, W., Sharp, J.M., 1997. A state space condition monitoring model for furnace erosion prediction and replacement. *Eur. J. Oper. Res.* 101 (1997), 1–14.
- Deluyker, H.A., Shumway, R.H., Wecker, W.E., Azari, A.S., Weaver, L.D., 1990. Modelling daily milk yield in Holstein cows using time series analysis. *J. Dairy Sci.* 73, 539–548.
- De Mol, R.M., van Zonneveld, R.T., Engel, B., Keen, A., Eradus, W.J., Kroeze, G.H., Ipema, A.H., Maatje, K., Rossing, W., 1992. A model for monitoring health and reproduction based on a combined processing of variables. In: Ipema, A.H., Lippus, A.C., Metz, J.H.M., Rossing, W. (Eds.), *Prospects for Automatic Milking. Proceedings of the International Symposium in Prospects for Automatic Milking Wageningen, Netherlands, 23–25 November 1992*. Pudoc Scientific Publishers, Wageningen, pp. 527–530 (EAAP Publication No. 65, 1992).
- De Mol, R.M., Maatje, K., Rossing, W., van Zonneveld, R.T., 1993. Tools for automated monitoring and diagnosis of reproduction and health of dairy cows. In: Annevelink, E., Oving, R.K., Vos, H.W. (Eds.), *Proceedings of the XXV CIOSTA-CIGR V Congress, Farm Planning, Labour and Labour Conditions, Computers in Agricultural Management, Wageningen Pers, Wageningen, The Netherlands*, pp. 287–294.
- De Mol, R.M., Kroeze, G.H., Achten, J.M.F.H., Maatje, K., Rossing, W., 1997. Results of a multivariate approach to automated oestrus and mastitis detection. *Livest. Prod. Sci.* 48, 219–227.
- Duncan, D.B., Horn, S.D., 1972. Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Stat. Ass.* 67, 815–821.
- Eradus, W.J., Rossing, W., Hogewerf, P.H., Benders, E., 1992. Signal processing of activity data for oestrus detection in dairy cattle. In: Ipema, A.H., Lippus, A.C., Metz, J.H.M., Rossing, W. (Eds.), *Prospects for Automatic Milking. Proceedings of the International Symposium in Prospects for Automatic Milking Wageningen, Netherlands, 23–25 November 1992*. Pudoc Scientific Publishers, Wageningen, pp. 360–369 (EAAP Publication No. 65, 1992).
- Federspiel, C.C., 1997. Estimating the inputs of gas transport processes in buildings. *IEEE Trans. Control Syst. Techn.* 5 (5).
- Harrison, P.J., Stevens, C.F., 1976. Bayesian forecasting (with discussion). *J. Roy. Stat. Soc. B Met.* 38, 205–247.
- Harvey, A.C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge UK 554 pp.
- Hogewerf, P.H., Maatje, K., Rossing, W., 1992. Computer aided system for health and reproduction control in dairy cows. In: Ipema, A.H., Lippus, A.C., Metz, J.H.M., Rossing, W. (Eds.), *Prospects for Automatic Milking. Proceedings of the International Symposium in Prospects for Automatic Milking Wageningen, Netherlands, 23–25 November 1992*. Pudoc Scientific Publishers, Wageningen, pp. 483–490 (EAAP Publication No. 65, 1992).
- Houben, E.H.P., 1995. *Economic optimization of decisions with respect to dairy cow health management*. Ph.D. Thesis, Department of Farm Management, Wageningen Agricultural University, Wageningen, The Netherlands, 146 pp.
- Maatje, K., Huijsmans, P.J.M., Rossing, W., Hogewerf, P.H., 1992. The efficacy of in-line measurement of quarter milk conductivity, milk yield and milk temperature for the detection of clinical and subclinical mastitis. *Livest. Prod. Sci.* 30, 239–249.
- Thysen, I., 1992. Monitoring bulk tank somatic cell counts by a multi-process Kalman filter. *Acta Agr. Scand. A-An.* 43, 58–64.
- Van Deusen, P.C., 1990. Evaluating time-dependent tree ring and climate relationships. *J. Environ. Qual.* 19, 481–488.
- Van Geer, F.C., 1987. *Applications of Kalman filtering in the analysis and design of groundwater monitoring networks*. Thesis, Technical University Delft, The Netherlands.