**From biological models to economic optimization**

Kristensen, Anders Ringgaard

# From biological models to economic optimization

CrossMark

Anders Ringgaard Kristensen*

HERD - Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark

## ARTICLE INFO

## ABSTRACT

This article addresses the additional challenges being faced when biological models are used as a basis for decision support in livestock herds. The challenges include dealing with uncertain information, observation costs, herd dynamics and methodological issues in relation to the computational methods applied particularly in the dynamic case. The desired key property of information included in models is that it can be used as the basis for unbiased prediction of the future performance of the animals. Often there will be a tradeoff between uncertainty and costs in the sense that the level of uncertainty can be reduced (for instance through additional tests) at some cost. Thus, the decision about which (and how many) tests to perform can be seen as an optimization problem in itself. Another way of expressing the tradeoff is to talk about the value of information which can sometimes be assessed by modeling different approaches and levels of detail in data collection. Various optimization methods of relevance to herd health management are discussed with the main emphasis on decision graphs in the static case and Markov decision processes (dynamic programming) in a dynamic context.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

In their textbook Kristensen et al. (2010a, Chapter 2) discuss the role of models in a herd context. They emphasize that the reason for building models is to become able to provide better decisions than human experts. They mention that the main advantages of model based decision support are, in the ideal situation, the ability to take herd and animal specific conditions into account, provision of a concise framework for combination of information from different sources, direct representation of uncertainty and efficient search algorithms for determination of optimal decisions. Models may further contribute with extensive sensitivity analyses concerning optimal decisions, deviating conditions and parameter values.

The above perspective is from a decision making point of view, but as discussed by Kristensen et al. (2010b, Chapter 14), another purpose of particularly simulation models is to improve the understanding of a system by combining research results from different areas in order to obtain a comprehensive description of the system. If the system being modeled is a biological system (for instance a herd or a larger population of animals) it is natural to refer to the model as a biological model.

Biological models in the context of preventive veterinary medicine are typically defined either at animal level (for instance dairy cow replacement models), herd level (for instance herd simulation models) or population level (disease spread models) even though some models comprise two levels (herd and animal or population and herd).

A biological model is first of all a *model*. The adjective *biological* only indicates that the model describes biological phenomena. A biological model does not represent the "biological truth", but it describes *our understanding* of a

* Tel.: +45 35 33 30 91.
  E-mail address: ark@sund.ku.dk
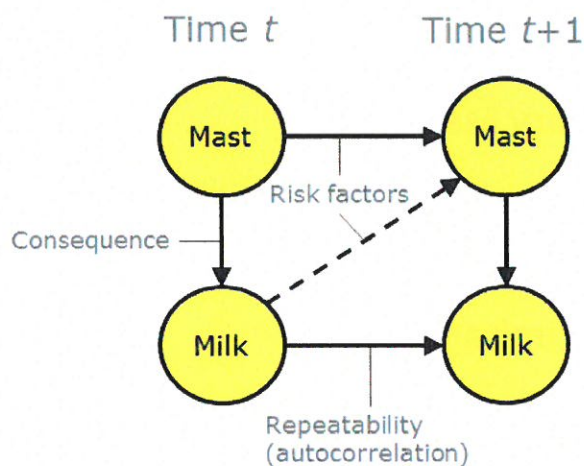
## Time $t$      Time $t+1$



**Fig. 1.** A simple biological model of a dairy cow over two time stages.

biological system. It is important to emphasize that this understanding can be more or less detailed, precise, biased or wrong. It is therefore desirable that the model interacts with actual observations (data) so that we can test the validity of the model.

Having built a biological model with the initial purpose of system comprehension, a natural next step is to consider whether the model could also be used for decision support. The aim of this article is to address the additional challenges being faced when biological models are extended and used for decision support in a livestock herd. The challenges include dealing with uncertain information, observation costs, herd dynamics and methodological issues in relation to the computational methods applied particularly in the dynamic case.

This article is organized as follows: after a short definition of the concept of a biological model a graphical language for describing such models is introduced and later extended to include utility and decisions. The graphical language is used throughout the article initially for the static case where decision graphs are applied and later for the dynamic case where the optimization technique is based on Markov decision processes.

## 2. A graphical language for models

### 2.1. Biological models

Usually, the description of a model is in mathematical and statistical terms, but in this article we will use a graphical modeling language known as *directed acyclic graphs* (DAGs). It is the graphical language used for Bayesian networks and decision graph. For a more thorough introduction to the concepts, reference is made to Jensen and Nielsen (2007) or Cowell et al. (1999).

The first step in using the graphical language for biological models is to identify the variables. Assuming an animal level model, the variables represent different traits of the animal (for instance the milk yield or the disease state). In the graphical language, the variables are simply shown as circles as in Fig. 1. The next step is to identify

the relations between the variables. Relations are shown as directed edges between pairs of variables. The direction of an edge is given by causality. In other words, $A \rightarrow B$ indicates that the value of $A$ influences the value of $B$. The causal influence may be in terms of consequences of certain states, risk factors or repeatability over time.

Fig. 1 shows an example of a very simple biological model with two time stages. As it is seen, two variables are considered at each stage: mastitis ("Mast") and milk yield ("Milk"). The edge from "Mast" to "Milk" at both stages suggests that the mastitis state influences the milk yield. The edge from "Mast" at time $t$ to "Mast" at time $t+1$ expresses that a previous mastitis case is a risk factor for a new case. The edge from "Milk" at time $t$ to "Milk" at time $t+1$ has a similar meaning in the sense that it corresponds to the repeatability of milk yield in a dairy cow.

Edges give important information about the model and the biological assumptions behind it, but it is important to understand that the absence of an edge gives just as important information. In the figure, a dashed edged has been added from "Milk" at time $t$ to "Mast" at time $t+1$. The purpose of the dashes is only to discuss whether or not the edge should be there. If it is omitted, the biological understanding is that "Milk" at time $t$ is *conditionally independent* of "Mast" at time $t+1$ given "Mast" at time $t$. For an in-depth discussion of the concept of conditional independence, reference is made to Jensen and Nielsen (2007). If, on the other hand, the edge is added it implies, that *even* if the value of "Mast" at time $t$ is known, additional information of the milk yield at time $t$ will alter the probability distribution of "Mast" at time $t+1$.

In Fig. 1 there is no edge from "Mast" at time $t$ to "Milk" at time $t+1$. The biological assumption behind the missing edge is that mastitis only indirectly influences the *future* milk yield through the increased risk of new mastitis cases. If, on the other hand, the biological understanding is that mastitis directly influences also the future milk yield, an edge should be added.

Thus, the DAG of Fig. 1 provides a consistent framework for logical testing of the model against the biological understanding of the system being modeled. It is natural to refer to the DAG as the *structure* of the model. Before it can be used it is, of course, also necessary to define the numerical properties of the model. In the example, it would be natural to let "Mast" be a categorical variable with states {"Yes", "No"}, or if more details are needed the states could be {"Gram Pos", "Gram Neg", "Other", "No"} or similar. The numerical specification for time $t$ would simply be a probability distribution over the state space.

The natural understanding of the "Milk" variable is that it is continuous, but in model implementations it will often be handled as a discrete variable with a number of states each corresponding to an interval of milk yield. The number of discrete intervals to use depends on the desired accuracy of the model, and often the number is a trade-off between accuracy and computational complexity. For a discussion of principles for discretization of continuous variables, reference is made to Nielsen et al. (2010). The numerical specification at time $t$ would be a conditional probability distribution over the (continuous or) discrete state space. If "Mast" is modeled as just "Yes" or "No", we
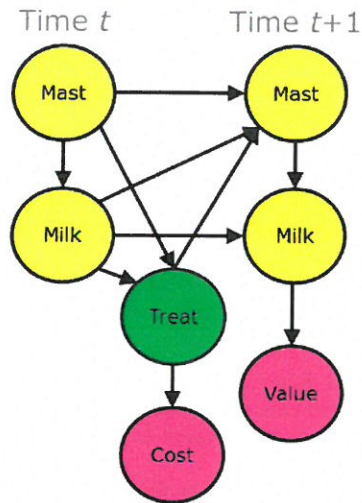
**Fig. 2.** The biological model from Fig. 1 extended with a decision variable and two utility variables. The model is now a decision graph.

might for instance use a normal distribution for the "Yes" case and another normal distribution for the "No" mastitis case.

The numerical specification of the variables at time $t + 1$ is slightly more complicated. For the "Mast" variable we need a set of conditional probability distributions for each combination of the *parent variables* "Mast" and "Milk" at time $t$, and for the "Milk" variable we need conditional probability distributions given previous milk yield and present mastitis state. All conditional (and unconditional) probability distributions should be estimated from data.

If the milk yield variables are modeled as discrete and the probability distributions have actually been estimated, we have a full working model which is an example of a *Bayesian Network* as described by Jensen and Nielsen (2007). The model would, for instance, be able to forecast the variables at time $t + 1$ given observations at time $t$. Alternatively, it might also estimate the probability that there had been a mastitis case at time $t$ given observations at time $t + 1$.

### 2.2. Optimization models

An obvious decision to consider in relation to the model shown in Fig. 1 is treatment in case of mastitis. It would therefore be natural to add a *decision variable* to the DAG. Treatment is not free, so in addition to the decision variable there should be a *utility variable* capturing the costs. The benefit of a treatment at time $t$ would be that it decreases the probability of mastitis at time $t + 1$. This may potentially lead to increased milk yield at time $t + 1$. We therefore need another utility variable to capture the value of the milk at time $t + 1$. The decision to treat will be based on a tradeoff between the cost of treatment and the expected higher milk yield if treated. Thus, we have an economic optimization problem.

Adding a decision variable "Treat", a utility variable "Cost" and a utility variable "Value" result in the DAG shown in Fig. 2 where decision and utility variables are

shown in colors deviating from the biological variables. The DAG is now referred to as a *decision graph* and, if implemented numerically, algorithms are available for optimization of decisions. One of them is known as single policy updating and was originally described by Lauritzen and Nilsson (2001). They also used the term *limited memory influence diagram* (LIMID) for decision graphs being solved by single policy updating. The algorithm has been implemented in the Esthauge LIMID Software System.[1]

The interpretation of edges into decision and utility variables deviates a little from edges into biological variables. Thus, an edge into a decision variable means that the value of the parent is known when the decision is made. In other words, if $V_1$ and $V_2$ are parents of the decision $D$, it means that an optimal decision strategy for $D$ will depend on $V_1$ and $V_2$ so that for each combination of values of $V_1$ and $V_2$ there will be an optimal decision. Therefore, edges into decision variables are also called *information edges*. An edge into a utility variable simply means that the utility value depends on the value of the parent. Edges out of a utility variable are not allowed.

### 3. The static case

#### 3.1. Static versus dynamic

The term "static" implies that time does not play a major role in the model. There may, as in Figs. 1 and 2 be two (or more) time stages involved, but they are modeled explicitly with limited time horizon. In the static case, a decision graph is an obvious tool for economic optimization. In the examples of this section the Esthauge Limid Software System will be used for optimization in the decision graphs.

The opposite of "static" is "dynamic" where time plays an important role, and where we often face a long and even undefined ("infinite") time horizon. The optimization techniques applied to dynamic models are different and will be dealt with in a later section.

#### 3.2. The two-cow problem

For the static case, a "two-cow problem" will be used and extended step by step throughout the section. It is heavily inspired by the "two-sow problem" described by Kristensen et al. (2010b, Example 6.1).

First, a verbal description of the decision problem is given:

Problem description: A smallholder dairy farmer has only housing capacity for one cow. Recently his cow died. He has no heifer, so he wants to buy a new cow. His neighbor has two pregnant 1st lactation cows for sale: Cow A and Cow B. Both of them will calve two weeks from now. The price is the same.

Decision problem: Should the smallholder buy Cow A or Cow B from his neighbor?

---

[1] http://www.esthauge.dk

**Table 1**
Optimal decision strategies and expected utilities for the four observation strategies.

| Cows obs. | Optimal strategy | Exp. util | Value of inf. |
|---|---|---|---|
| None | Choose Cow A or Cow B | 8407 | 0 |
| Cow A | Choose Cow A if $Y_{1A} > 7500$[a] | 8546 | 139 |
| Cow B | Choose Cow A if $Y_{1B} \leq 7500$[a] | 8546 | 139 |
| Both | Choose Cow A if $Y_{1A} > Y_{1B}$[a] | 8600 | 193 |

[a] Otherwise choose Cow B.

The biological model behind this problem is a model of milk yield in a dairy cow. Let $Y_{nA}$ and $Y_{nB}$ be the nth lactation milk yield of Cow A and Cow B, respectively. We assume that the mean milk yield of a first parity cow is 7500 kg and 8400 kg for a second lactation cow. The standard deviation (between cows, within herd) is for both parities assumed to be 10% of the mean. Finally, the repeatability of milk yield from first to second lactation is assumed to be 0.45.

Before continuing with a solution to the decision problem it is worth noticing that according to usual partial budgeting principles (see Dijkhuizen et al., 1995, for a discussion) everything that does not depend on the decision can be ignored in the optimization. In this case we can therefore ignore the price of the new cow and the value of the calf. The only relevant variable to consider is therefore the value of the milk. A decision graph implementation of the two-cow problem in the Esthauge LIMID Software System is shown in Fig. 3(a).

### 3.3. Observation scenarios and value of information

In Fig. 3 four alternative observation scenarios are shown. The four scenarios are

1. Observe nothing (Fig. 3(a)).
2. Observe first lactation milk yield of Cow A (Fig. 3(b)).
3. Observe first lactation milk yield of Cow B (Fig. 3(c)).
4. Observe first lactation milk yield of Cow A and Cow B (Fig. 3(d)).

In the software system, the different scenarios are simply defined by adding the relevant information edges as seen in the figure. In each case, the Single Policy Updating algorithm is able to provide:

1. The optimal decision strategy.
2. The expected utility under the optimal strategy.

Running an optimization for each of the four scenarios produces the results shown in Table 1.

If neither of the first lactation milk yields are observed, it does not matter whether the smallholder dairy farmer buys Cow A or Cow B. In both cases the expected utility (measured as amount of milk) is simply 8407 corresponding to the mean milk yield of an arbitrarily selected second parity cow (it is because of the inaccuracies introduced by the discretization of the milk yield variables that the result is not exactly 8400).

If first lactation milk yield of one of the cows is observed, it does not matter whether it is Cow A or Cow B. In both

cases the expected utility increases to 8546 or 139 more than the no observation scenario. In both cases the strategy is to choose the observed cow if its first lactation milk yield is higher than average for a first lactation cow. If it is lower than average the other cow is chosen.

Finally, if both first lactation milk yields are observed, the optimal strategy is, of course, to choose the cow having the highest first lactation milk yield. In this scenario the expected utility is 8600 or 193 higher than the no observation scenario.

The results of Table 1 illustrate several important characteristics of observations and decisions:

- Observations have a value because they enable the decision maker to make better (i.e. more informed) intervening decisions. This value is known as the *value of information*.
- The value of information is found as the expected utility *with* information minus the expected utility *without* information.
- The value of information shows diminishing returns to scale. Thus, the value of observing both cows is less than twice the value of observing both cows. This result is generally true in cases like this where the amount of information is measured as the number of conditionally independent observations (Kihlstrom, 1974). Under other conditions it is not necessarily true (Chade and Schlee, 2002).

### 3.4. Observation costs and observation strategy

In the previous subsection it was implicitly assumed that there were no observation costs, and in that case the optimal observation strategy would be to observe both first lactation milk yields. The assumption mentioned in Section 3.2 was, however, that the two cows had their first lactations in the herd of the neighbor. Thus, the smallholder farmer does not have direct access to those observations. In principle, they are available in the neighbor's herd, but perhaps the neighbor will not supply this information for free.

Assume that the price for getting the information is $x$ per cow. Depending on $x$ it may be optimal to completely abstain from observations, to buy information about one cow at the price of $x$ or to buy the information for both cows at the price of $2x$. Jensen and Nielsen (2007, p312) describe how to include the decision to observe directly in a decision graph. The trick in this example is to include two additional variables corresponding to the observed milk yields of first lactation as seen in Fig. 4. If it is decided, for instance, to observe Cow A, the value of the observation variable $O_{1A}$ will be equal to $Y_{1A}$. If, on the other hand, it is decided not to observe, the value of $O_{1A}$ will simply be "Unobserved". For the correctness of the model it is important to remember, that the second lactation milk yield of the selected cow depends on the actual first lactation milk yield (not the observed).

While Fig. 4 provides a general framework for optimization of the observation strategy, the example in this case is so simple that it is easily seen (without using the decision

(a) Observe nothing

(b) Observe Cow A

(c) Observe Cow B

(d) Observe both cows
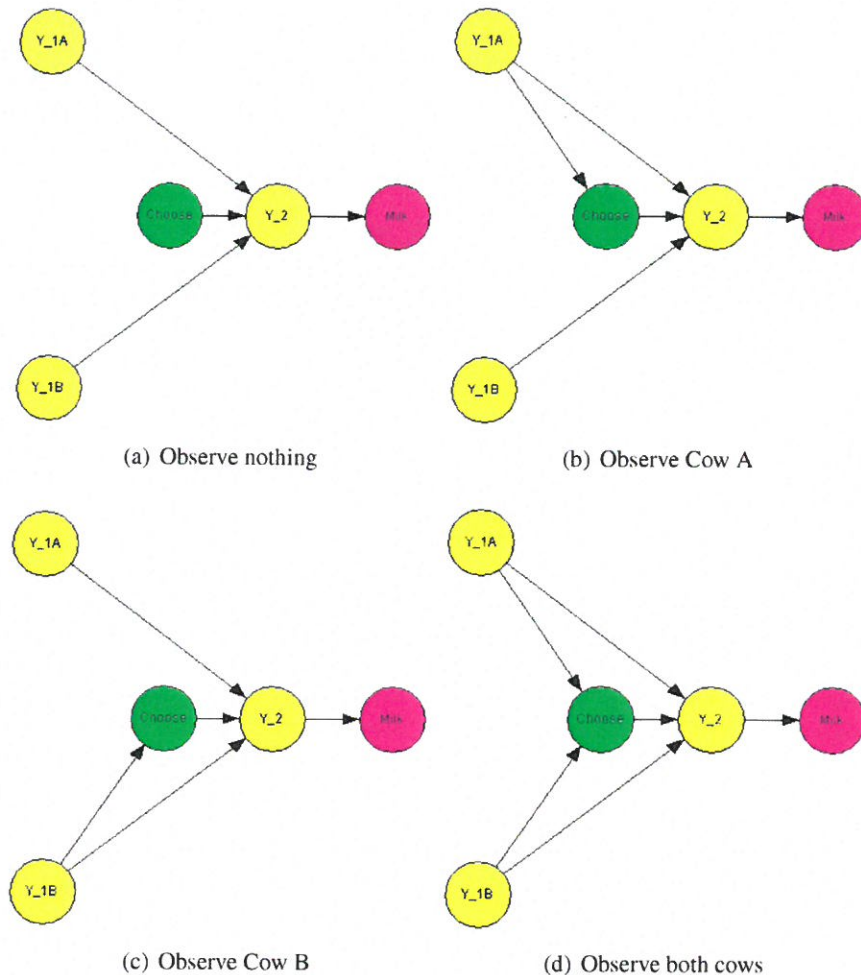
**Fig. 3.** The two-cow problem with four different observation scenarios implemented in the Esthauge LIMID Software System. The variables Y_1A and Y_1B are first lactation milk yield of Cow A and Cow B, respectively, and Y_2 is the second lactation milk yield of the chosen cow. "Choose" is the decision (either Cow A or Cow B), and "Milk" is the value of the milk.

graph) from Table 1 that the optimal observation strategy depends on the cost $x$ as follows:

- Observe both cows if $2x \leq 193$.
- Observe one cow, if $2x > 193$, but $x \leq 139$.
- Abstain from observation if $x > 139$.

From this example we conclude that for a given observation cost there is an optimal observation strategy. Thus, the decision to observe should also be based on economic optimization.

### 3.5. Observation precision

Very often in practise observations are not precise. Diagnostic tests, for instance, are characterized by their sensitivity and specificity, and many measurements of continuous variables are associated with a measurement error. In both cases, the solution in terms of modeling is to distinguish the observed value from the true (unobservable) value in the same way as for milk yield in Fig. 4. Now,

however, if the variable is observed, $O_{1A}$ will not hold the same value as $Y_{1A}$, but instead we might have a relation like

$$O_{1A} = Y_{1A} + e, \quad e \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where $e$ is the observation error. The assumption behind Eq. (1) is that the observation is unbiased with the *precision* $1/\sigma^2$.

The value of this, less precise, information will of course be less than in the case with perfect information. Just like in Section 3.3 there are four observation scenarios, and one of them (observing Cow A) is shown in Fig. 5.

Running an optimization (Single Policy Updating) for all four scenarios assuming very imprecise measurement of milk yield with $\sigma = 300$ reveals that the value of observing one cow is 128 and the value of observing both cows is 179. Comparing with Table 1 we see that the value is less than under perfect information, but actually only slightly. Therefore, if the imprecise information is cheaper than the exact, it may very well be profitable only to buy the less precise information.
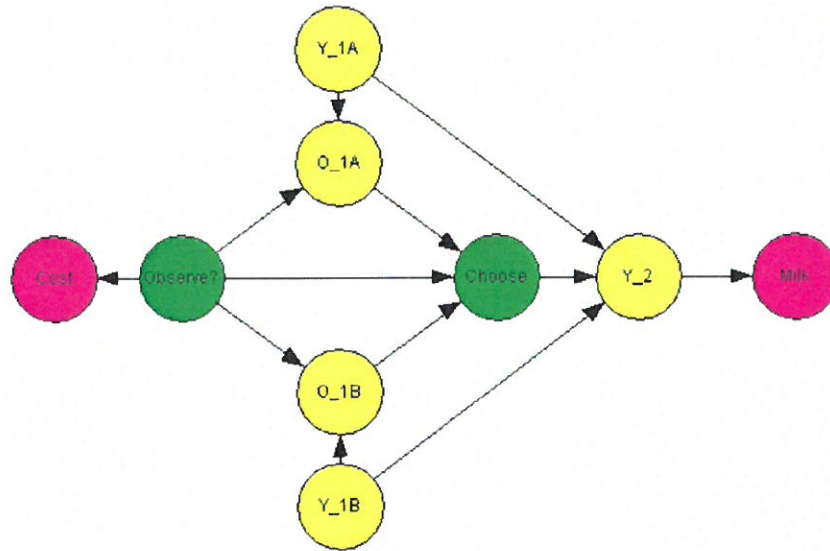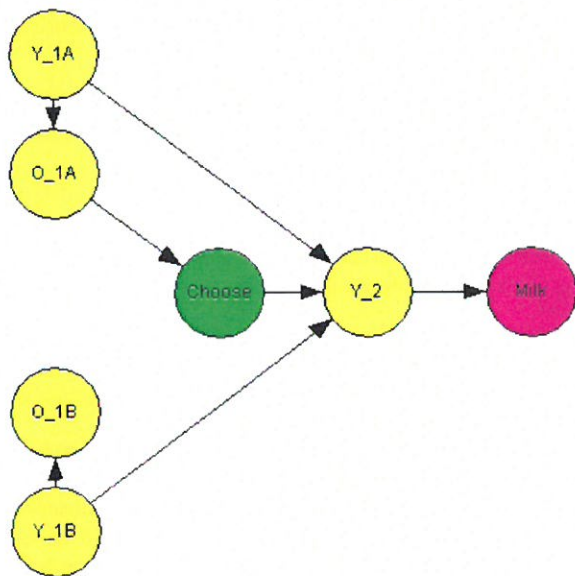
**Fig. 4.** The two-cow problem with observation as a decision involving a cost. The variables O_1A and O_1B are the *observed* first lactation milk yields of Cow A and Cow B, respectively, "Observe?" is the observation decision and "Cost" is the observation costs. Other variables are explained in Fig. 3. If a cow is not observed, the corresponding observation variable will be in state "Unobserved". Otherwise it takes the same numerical value as the actual first lactation milk yield.

Thus, we can conclude that the value of information depends on the precision, and if high precision also has a higher cost it may be beneficial only to aim at the lower precision.

### 3.6. Decision graphs for economic optimization

In the static case, a decision graph is an almost perfect tool for economic optimization. The technique enables us to handle unobservable traits and thus distinguish observed

values from true underlying latent values. Typically, however, software implementations require that continuous variables are transformed to discrete variables with a finite number of states with some loss of accuracy as a consequence.

## 4. The dynamic case

### 4.1. The combinatorial explosion

When moving from the static case to the dynamic case, the first challenge faced is the combinatorial explosion as illustrated in Fig. 6. The present cow is a Parity five cow. If it is kept, it will turn into a Parity six cow. If replaced, the next lactation will be a first parity cow. Continuing only three
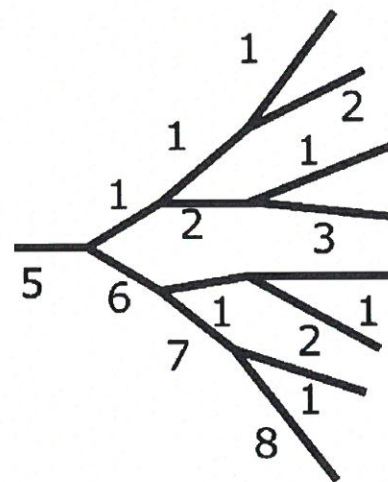


**Fig. 5.** The two-cow problem with observation errors. The variables are explained in Figs. 3 and 4. The observed milk yields are influenced by an unbiased observation error. Here, it is assumed that Cow A is observed.



**Fig. 6.** The combinatorial explosion of the replacement problem in dairy cows. The numbers refer to parities.

lactations forward, there will be eight possible scenarios. If also other properties were considered, there would soon be prohibitively many scenarios.

A very efficient method for handling the combinatorial explosion is a *Markov decision process* (also known as *dynamic programming*). Numerous applications of the method in livestock production have been published. Some recent examples (Nielsen et al., 2010; Demeter et al., 2011) focus on production and reproduction traits whereas other (Bar et al., 2008a,b; Cha et al., 2010, 2011, 2014a) focus on health traits. All the models mentioned have been implemented by use of the MLHMP software system[2] developed by Kristensen (2003). Recently, Nielsen and Kristensen (2014) gave a comprehensive review of the applications of the technique.

Even though decision graphs are not able to solve Markov decision processes with undefined time horizon, the DAG language will also be used in this section to visualize the models.

### 4.2. Markov decision processes

An example of a Markov decision process with focus on mastitis is shown in Fig. 7. The example is heavily inspired by models presented by Bar et al. (2008b,a) and Cha et al. (2011, 2014a). The model shown considers three properties (age, milk yield and mastitis) of a dairy cow at regular time intervals called *stages*. The stage may be a lactation, a month or a week depending on the level of detail. Each of the three properties is shown as a separate *state variable* at each time stage.

The age is typically measured as parity (often supplemented by month of lactation), the milk yield as a number of discrete levels, and the mastitis variable may simply have the states "Yes" and "No" like in Bar et al. (2008a), or it may be modeled at a more detailed level. Thus, Cha et al. (2011) considered the Gram type of the mastitis infection and Cha et al. (2014a) even considered the bacteria causing the infection. The combined values of all three state variables is referred to as the *state* of the Markov decision process.

Based on the current state a *decision* is made as shown in the figure. In the example the decision options could for instance be "Treat", "Inseminate" or "Replace" (and perhaps combinations of these). Having observed the state and made a decision, a *reward* is received. Typically, the reward would be the gross margin for a cow given state and decision. The rewards are shown as utility variables in Fig. 7.

Unlike a decision graph, a Markov decision process does not necessarily have a well-defined planning horizon. In Fig. 7 this is illustrated by the dots at the right-hand side. Often the time horizon is considered as infinite. If, on the other hand, a Markov decision process actually has a well defined planning horizon with a finite number of stages, it is just a special case of a decision graph and, accordingly, Single Policy Updating as described by Lauritzen and

Nilsson (2001) can be used for optimization even though it is seldom done.

Dynamic programming as a method for dealing with sequential decision problems was described by Bellman (1957), but later the term Markov decision process was introduced by Howard (1960) who also described the classical optimization techniques known as *policy iteration* for the infinite stage case and *value iteration* for the finite stage case. Because of its simplicity, value iteration has often been used as an approximate optimization technique also for the infinite case.

The first example known to the author of application of Markov decision processes in dairy cows was presented by Jenkins and Halter (1963), but a far more comprehensive early application was published by Giaever (1966).

### 4.3. Challenges when modeling with Markov decision processes

Even though Markov decision processes have been applied in numerous dairy cow replacement, insemination and treatment models in literature, the modeler faces some challenges when using the technique. In this section three major (and mutually related) challenges are discussed.

#### 4.3.1. The Markov property

The Markov property is essential for the correctness of the model and this is probably where the biggest pitfalls are for the modeler. The Markov property states that the present state and decision contain all relevant information for predicting the future. What has happened before (i.e. previous states and decisions) is of no relevance for the future.

A very simple illustration of the Markov property, what it means and how to compensate in case of violations is given in Fig. 8 where (for simplicity) only one property, "Milk yield" is considered over four stages. The graphical modeling language in the form of DAGs is very well suited for such an illustration.

In Fig. 8(a) the model says that future milk yields only depends on present milk yield. Historical observations of milk yield are of no relevance given that the present milk yield is known. If that is true (also from a biological point of view), the Markov property is satisfied and the model will be valid.

If, on the other hand, biological knowledge implies that future milk yield is influenced not only by present milk yield, but also by previous milk yield as illustrated in Fig. 8(b), then the Markov property is violated. Even though the model shown makes sense it is not a Markov process.

If what is seen in Fig. 8(b) corresponds to our biological understanding of the system (a dairy cow), we face a dilemma. A straightforward modeling as in Fig. 8(a) will be directly erroneous, but the model shown in Fig. 8(b) is not a Markov process and cannot be solved as such. The traditional approach for compensation for such a violation of the Markov property is to include one or more *memory variables* in the state space as shown in Fig. 8(c) where the state space has been redefined to include a state variable simply remembering the milk yield of the previous stage. Thus

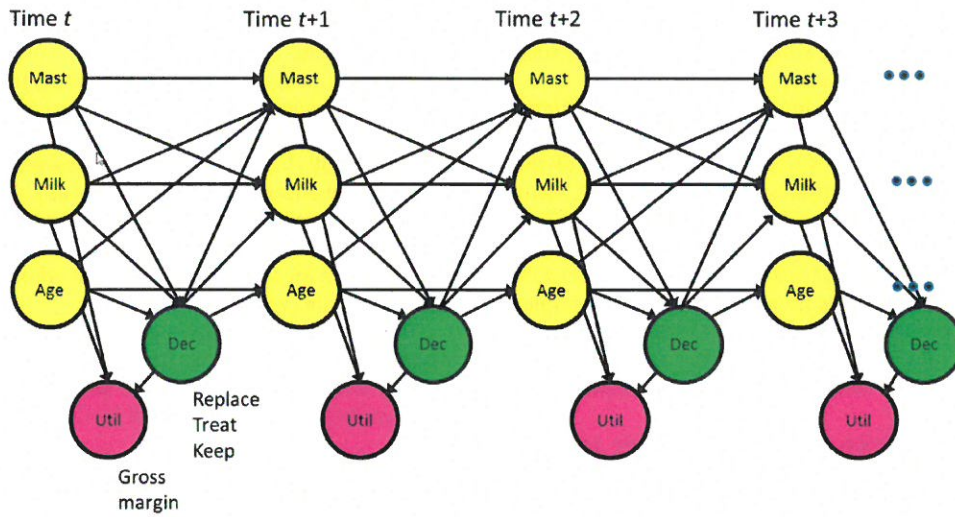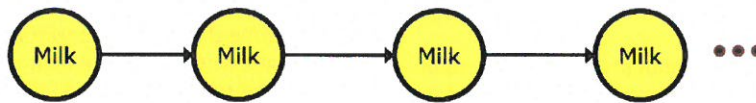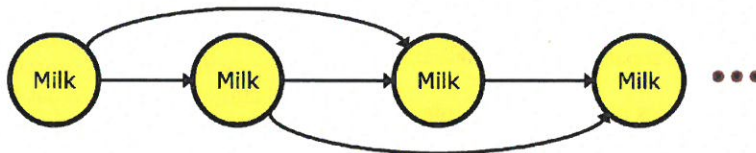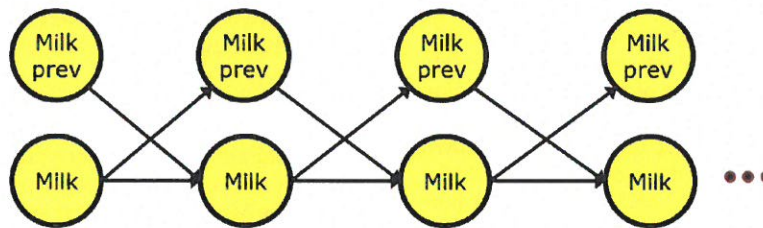---

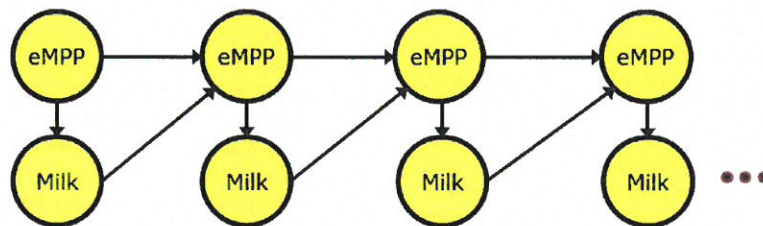**Fig. 7.** A Markov decision process with three state variables.



(a) Markov property satisfied



(b) Violation of Markov property



(c) Compensation by memory variable



(d) Compensation by Bayesian updating

**Fig. 8.** Illustration of the Markov property and how to compensate for violations.

the state is defined as the combined values of present and previous milk yield, and the Markov property is satisfied.

The approach with memory variables has been used in numerous models published in literature. For milk yield is has for instance been done exactly as described in models by van Arendonk (1985), van Arendonk and Dijkhuizen (1985) and Kristensen (1987, 1989). A similar approach for number of mastitis cases was used by Houben et al. (1994). For a more complete overview, reference is made to Nielsen and Kristensen (2014).

A weakness of the approach with memory variables is that it is not always obvious how many to include. Is it, for instance, sufficient to include state variables for present and previous milk yield or should also memory variables for even older milk yield observations be included? A more general approach is, therefore, to use Bayesian updating as illustrated in Fig. 8(d). The idea, which was simultaneously (but independently) introduced by Kennedy and Stott (1993) and Kristensen (1993), is to see the observations of milk yield over time as influenced by a latent milk production potential (MPP) of the cow. This potential can then be estimated sequentially as observations are done so that we at each stage have an estimated milk production potential (eMPP) as illustrated in the figure. The technique was further generalized by Nielsen et al. (2011) and it has already been applied in several models for milk yield modeling (e.g. Nielsen et al., 2010; Demeter et al., 2011).

If the same principles were to be applied on disease data then actual disease cases should be seen as results of an underlying latent disease risk which is re-estimated by Bayesian updating each time a new case is seen or not seen. Only very few attempts have been done to implement such a framework even though Ge et al. (2010, 2014) implemented a population level model for foot and mouth disease control with a latent disease risk sequentially estimated by Bayesian updating.

### 4.3.2. Full observability of state space

In the static case imperfect observations were discussed in Section 3.5. It was assumed that only an imperfect observation of milk yield was available. The scenario was illustrated in Fig. 5 where it is seen that the decision is based on the imperfect observation, whereas it is the true first lactation milk yield that influences the second lactation milk yield.

In a decision graph this is not really a problem, but in a sequential decision problem solved by Markov decision programming it is not possible to handle such a case correctly. It is a precondition for the method that the state is fully observable, because otherwise the future milk yield is influenced by something that is not observed. This is a violation of the Markov property because the decision is not based on all relevant information.

In the model shown in Fig. 7 mastitis is one of the properties considered. As mentioned, mastitis can be handled at different levels of detail. Cha et al. (2014a) modeled the disease at a very detailed pathogen specific level. Such an information level allows for pathogen specific effects on milk yield, conception, mortality, treatment cost, and risk of new cases, but those advantages must be weighed against increased culture costs to determine the pathogen.
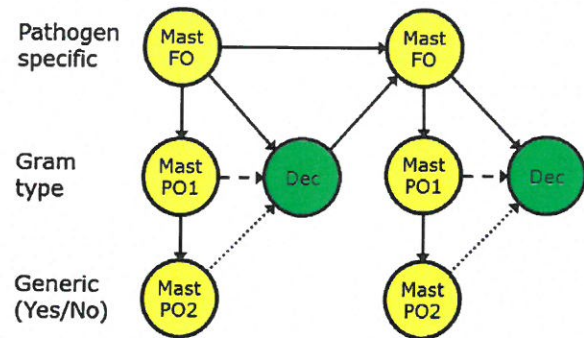


**Fig. 9.** Three possible observation scenarios for mastitis. FO: full observation at pathogen level (solid information edge); PO1: partial observation at Gram type level (dashed information edge); PO2: partial observation at generic level (dotted information edge). Only full observation satisfies the Markov property.

Instead of observing mastitis cases at pathogen level it is worth considering whether a cheaper test method only determining the Gram type should be applied instead as assumed in Cha et al. (2011) or even completely without testing the type of mastitis as assumed by Bar et al. (2008a).

Based on these considerations three observation scenarios can be defined as illustrated in Fig. 9:

1. Observe mastitis at pathogen level (full observation, FO) illustrated by the solid information edge.
2. Observe mastitis at Gram type level (partial observation, PO1) illustrated by the dashed information edge.
3. Observe mastitis at generic level (partial observation, PO2) illustrated by the dotted information edge.

It should be noted that the edges from "FO" to "PO1" and further to "PO2" correspond to going from detailed to less detailed knowledge. A certain pathogen in "FO" will, for instance, always be in one of the classes defined for "PO1" and, accordingly, the conditional probability distributions behind the edges are trivial.

The problem in Fig. 9 is that if the pathogen specific model reflects our biological understanding of the disease dynamics of a dairy cow, the Markov property will only be satisfied if the decision is based on full observation. If instead only a partial observation is done ("PO1" or "PO2") it means that the future depends on aspects that are not known when the decision is made. Thus, one of the fundamental assumptions in a Markov decision process is violated.

Nevertheless, it is also in the dynamic case relevant to ask about the value of information. Full observation will increase the possibilities for targeting the treatment to the pathogen, but the culture costs are higher. Unlike in the static case, the question cannot be answered exactly, so approximate methods must be used.

One option used in an ongoing study by Cha et al. (2014b) is to combine optimization of the model with full observability with simulation of simplified strategies where the same decision is made for all pathogens in for instance the group of Gram positive bacteria. In that way an approximate value of information can be

estimated. Another option, which has not been tested in practice, would be to use a technique based on combination of Markov decision processes and decision graphs as described by Jørgensen et al. (2014).

### 4.3.3. The curse of dimensionality

When several state variables (cow traits) are considered at a realistic number of levels, the state space grows to prohibitive dimensions so dairy cow replacement models published in literature often have millions of state combinations. Therefore, a direct optimization is not always possible.

A solution is to decompose the state space according to time and build a hierarchical model as originally proposed by Kristensen (1988) and later further extended by Kristensen and Jørgensen (2000). Such a decomposition has a tremendous effect on computational performance and even models with millions of state combinations can be solved. The technique has been used in numerous dairy cow replacement models as well as for other decision problems. For an overview of applications reference is made to Nielsen and Kristensen (2014). A special software system (MLHMP) is available for implementation of hierarchical models (Kristensen, 2003).

Directly inspired by a dairy cow replacement model developed by Demeter et al. (2011) a hierarchical Markov decision process with three levels is illustrated in Fig. 10. The idea of the technique is to decompose the state space according to time horizon so moving from the *founder level* to the *child level* and the *grandchild level* the time horizon is decreased and the stage lengths are accordingly shorter. Conversely, the level of detail increases when moving from founder level to child and grandchild levels.

In the example of Fig. 10 the founder level is shown at the top, the child level in the middle and the grandchild level at the bottom. Each level corresponds to separate Markov decision processes with their own stages, states and decisions as follows:

Founder level: The process at the founder level has an infinite horizon and models a chain of cows successively replacing each other.

Stage: A stage is defined as the lifetime of the present cow in the herd.
State: Only one state variable, the breeding index at first calving, is defined.
Decision: No decisions are made at this level.

Child level: The process at the child level has a limited time horizon corresponding to the maximum number of lactations considered for a cow.

Stage: A stage is defined as the duration of a full lactation cycle (including dry period).
State: Only one state variable, the estimated permanent milk yield potential of the cow, is considered.
Decision: No decisions are made at this level.

Grandchild level: The process has a limited time horizon corresponding to the maximum duration of a lactation cycle.

Stage: A stage is defined as a fixed time interval (e.g. a month, a week or a day). Demeter et al. (2011) assumed monthly stages.
State: Two state variables are considered: the estimated permanent milk yield potential of the cow and the estimated temporary potential – see Demeter et al. (2011) for details.
Decision: "Keep and inseminate", "Keep", "Replace".

It should be noticed that, unlike the ordinary Markov decision process in Fig. 10, age is not included as a state variable in the hierarchical model described above. The age of the cow will simply be known from the stage numbers at the child and grandchild level.

Due to the different time horizons (infinite in the founder process and finite in the child and grandchild processes), an optimization technique combining policy iteration in the founder process and value iteration in the child and grandchild processes is applied as described by Kristensen and Jørgensen (2000).

## 5. Discussion

In this article the optimization problem in relation to biological modeling has been discussed for the static case as well as for the dynamic case. A common graphical language in the form of DAGs has been used to visualize the models and important characteristics and challenges have been described and discussed. Even though all examples have referred to dairy cows, the principles discussed apply to modeling of livestock production in general.

The desired key property of information included in models is that it can be used as the basis for unbiased prediction of the future performance of the animals. Straightforward representation of observed data is seldom desirable because it often leads to computationally intractable models (the use of memory variables in Markov decision processes is an obvious example of this). Instead emphasis should be put on sequential Bayesian updating of latent variables given data.

In the static case, Bayesian updating is straightforward, and the decision graph framework is very well suited for such an approach with latent, unobservable, properties. In the dynamic case, this is more complicated because of the underlying assumption of full observability of the state space. Nevertheless, the use of Bayesian updating in a dynamic context is now quite well described and understood when it comes to normally distributed data, where Nielsen et al. (2011) presented a framework. When it comes to categorical variables, however, the concept has not been described very well even though Ge et al. (2010, 2014) applies a Bayesian updating technique to a variable following a Poisson distribution. Bayesian updating for categorical variables in Markov decision processes is certainly an area where more research is desirable.

All the examples presented in this article have, for simplicity, been at animal level. A few examples in literature
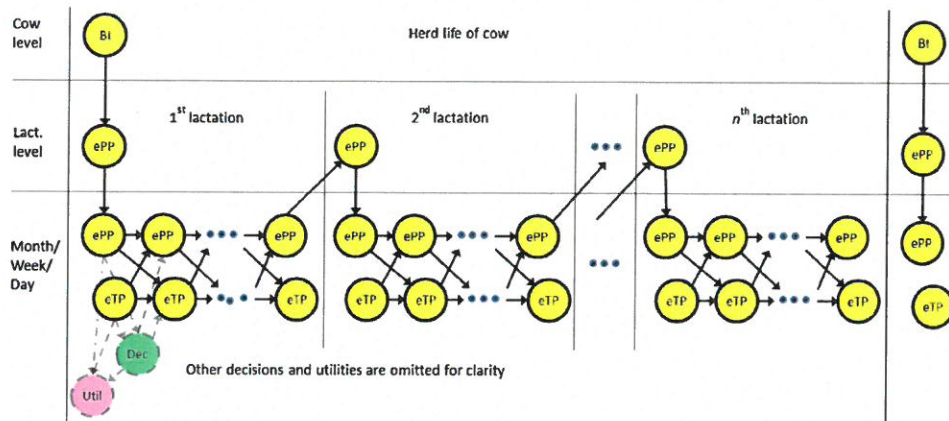
**Fig. 10.** A hierarchical Markov decision process with three levels: cow level, lactation level and lactation stage level. The four variables to the right (after the vertical line) correspond to the next cow. The variables are BI: breeding index; ePP: estimated permanent potential; eTP: estimated temporary potential.

have applied Markov decision processes also at population level (e.g. Ge et al., 2010, 2014; Viet et al., 2012) for disease control decisions, but the main application area remains the animal level.

A problem, which has not been discussed in this article, is the interaction between herd and animal level in the dynamic case. It is obvious that such interactions are very relevant if contagious diseases are modeled, but also herd constraints of various kinds (limited supply of heifers, milk quota, etc.) will create interactions between animals. No good optimization techniques are available to link herd and animal level, even though a few attempts are reported in literature (Ben-Ari and Gal, 1986; Kristensen, 1992). This is also an area where more research is needed.

## 6. Conclusions

Based on the previous sections the following conclusions are made for observations and their value:

- Observations have a value because they enable the decision maker to make better (i.e. more informed) intervening decisions. This value is known as the *value of information*.
- The value of information is found as the expected utility *with* information minus the expected utility *without* information.
- The value of information shows diminishing returns to scale.
- The decision to observe should be based on economic optimization.
- The value of observation depends on the precision.
- The value of information is often more difficult to estimate in a dynamic framework.

Methodologically it is concluded that decision graphs are very well suited for static decision problems whereas Markov decision processes is the preferred tool for dynamic decision problems.

## References

Bar, D., Tauer, L.W., Bennett, G., González, R.N., Hertl, J.A., Schulte, H.F., Schukken, Y.H., Welcome, F.L., Gröhn, Y.T., 2008a. Use of a dynamic programming model to estimate the value of clinical mastitis treatment and prevention options utilized by dairy producers. Agric. Syst. 99 (1), 6–12.

Bar, D., Tauer, L.W., Bennett, G., Gonzalez, R.N., Hertl, J.A., Schukken, Y.H., Schulte, H.F., Welcome, F.L., Gröhn, Y.T., 2008b. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J. Dairy Sci. 91 (6), 2205–2214.

Bellman, R.E., 1957. Dynamic Programming. Princeton University Press, Princeton.

Ben-Ari, Y., Gal, S., 1986. Optimal replacement policy for multicomponent systems: an application to a dairy herd. Eur. J. Oper. Res. 23, 213–221.

Cha, E., Bar, D., Hertl, J., Tauer, L., Bennett, G., González, R., Schukken, Y., Welcome, F., Gröhn, Y., 2011. The cost and management of different types of clinical mastitis in dairy cows estimated by dynamic programming. J. Dairy Sci. 94 (9), 4476–4487.

Cha, E., Hertl, J., Bar, D., Groehn, Y., 2010. The cost of different types of lameness in dairy cows calculated by dynamic programming. Prev. Vet. Med. 97 (1), 1–8.

Cha, E., Kristensen, A.R., Hertl, J.A., Schukken, Y.H., Tauer, L., Welcome, F.L., Gröhn, Y.T., 2014a. Optimal insemination and replacement decisions to minimize the cost of pathogen specific clinical mastitis in dairy cows. J. Dairy Sci. 97, 2101–2117.

Cha, E., Smith, R.L., Kristensen, A.R., Hertl, J.A., Schukken, Y.H., Tauer, L.W., Welcome, F.L., Gröhn, Y.T., 2014b. The Value of Pathogen Information in Treating Mastitis Estimated by An Economic Optimization Model Based on Dynamic Programming (in preparation).

Chade, H., Schlee, E.E., 2002. Another look at the Ragner–Stiglitz nonconcavity in the value of information. J. Econ. Theory 107, 421–452.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L., Spiegelhalter, D.J., 1999. Probabilistic Networks and Expert Systems. Statistics for Engineering and Information Science. Springer-Verlag, New York, Berlin, Heidelberg.

Demeter, R.M., Kristensen, A.R., Dijkstra, J., Lansink, A.G.J.M.O., Meuwissen, M.P.M., van Arendonk, J.A.M., 2011. A multi-level hierarchic Markov process with Bayesian updating for herd optimization and simulation in dairy cattle. J. Dairy Sci. 94 (12), 5938–5962.

Dijkhuizen, A.A., Huirne, R.B.M., Jalvingh, A.W., 1995. Economic analysis of animal diseases and their control. Prev. Vet. Med. 25, 135–149.

Ge, L., Kristensen, A.R., Mourits, M.C., Huirne, R.B., 2014. A new decision support framework for managing foot-and-mouth disease epidemics. Ann. Oper. Res. 219, 49–62.

Ge, L., Mourits, M.C., Kristensen, A.R., Huirne, R.B., 2010. A modelling approach to support dynamic decision-making in the control of FMD epidemics. Prev. Vet. Med. 95 (3–4), 167–174.

Giaever, H.B., 1966. Optimal dairy cow replacement policies. University of California, Berkeley/University Microfilms, Ann Arbor, MI, Ph.D. Thesis.

Houben, E.H.P., Huirne, R.B.M., Dijkhuizen, A.A., Kristensen, A.R., 1994. Optimal replacement of mastitis cows determined by a hierarchic Markov process. J. Dairy Sci. 77, 2975–2993.

Howard, R., 1960. Dynamic Programming and Markov Processes. The MIT Press, Cambridge, MA.

Jenkins, K.B., Halter, A.N., 1963. A Multistage Stochastic Decision Model: Application to Replacement of Dairy Cows. Technical Bulletin 67, Agricultural Experimental Station. Oregon State University.

Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs. Information Science and Statistics, 2nd ed. Springer, New York.

Jørgensen, E., Kristensen, A.R., Nilsson, D., 2014. Markov LIMID processes for representing and solving renewal problems. Ann. Oper. Res. 219, 63–84.

Kennedy, J.O.S., Stott, A.W., 1993. An adaptive decision-making aid for dairy cow replacement. Agric. Syst. 42, 25–39.

Kihlstrom, R., 1974. A Bayesian model of demand for information about product quality. Int. Econ. Rev. 15 (1), 99–118.

Kristensen, A.R., 1987. Optimal replacement and ranking of dairy cows determined by a hierarchic Markov process. Livest. Prod. Sci. 16, 131–144.

Kristensen, A.R., 1988. Hierarchic Markov processes and their applications in replacement models. Eur. J. Oper. Res. 35, 207–215.

Kristensen, A.R., 1989. Optimal replacement and ranking of dairy cows under milk quotas. Acta Agric. Scand. 39, 311–318.

Kristensen, A.R., 1992. Optimal replacement in the dairy herd: a multi-component system. Agric. Syst. 39, 1–24.

Kristensen, A.R., 1993. Bayesian updating in hierarchic Markov processes applied to the animal replacement problem. Eur. Rev. Agric. Econ. 20, 223–239.

Kristensen, A.R., 2003. A general software system for Markov decision processes in herd management applications. Comput. Electron. Agric. 38, 199–215.

Kristensen, A.R., Jørgensen, E., 2000. Multi-level hierarchic Markov processes as a framework for herd management support. Ann. Oper. Res. 94, 69–89.

Kristensen, A.R., Jørgensen, E., Toft, N., 2010a. Herd Management Science: I. Basic Concepts. Academic Books, Copenhagen.

Kristensen, A.R., Jørgensen, E., Toft, N., 2010b. Herd Management Science: II. Advanced Topics. Academic Books, Copenhagen.

Lauritzen, S.L., Nilsson, D., 2001. Representing and solving decision problems with limited information. Manag. Sci. 47, 1235–1251.

Nielsen, L.R., Jørgensen, E., Højsgaard, S., 2011. Embedding a state space model into a Markov decision process. Ann. Oper. Res. 190, 289–309.

Nielsen, L.R., Jørgensen, E., Kristensen, A.R., Østergaard, S., 2010. Optimal replacement policies for dairy cows based on daily yield measurements. J. Dairy Sci. 93 (1), 75–92.

Nielsen, L.R., Kristensen, A.R., 2014. Markov decision processes to model livestock systems. In: Plà-Aragonés, L.M. (Ed.), Handbook of Operations Research in Agriculture and the Agri-Food Industry. Springer.

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. Livest. Prod. Sci. 13, 101–121.

van Arendonk, J.A.M., Dijkhuizen, A.A., 1985. Studies on the replacement policies in dairy cattle: III. Optimum policy and influence of changes in production and prices. Livest. Prod. Sci. 13, 333–349.

Viet, A.-F., Jeanpierre, L., Bouzid, M., Mouaddib, A.-I., 2012. Using Markov decision processes to define an adaptive strategy to control the spread of an animal disease. Comput. Electron. Agric. 80 (0), 71–79.