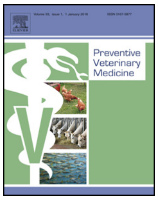
Preventive Veterinary Medicine 118 (2015) 226–237

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



From biological models to economic optimization

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

Article history:
Received 23 January 2014
Received in revised form 31 July 2014
Accepted 15 November 2014

Keywords:
Decision graphs
Markov decision processes
Dynamic programming
Value of information

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.

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

http://dx.doi.org/10.1016/j.prevetmed.2014.11.019
0167-5877/

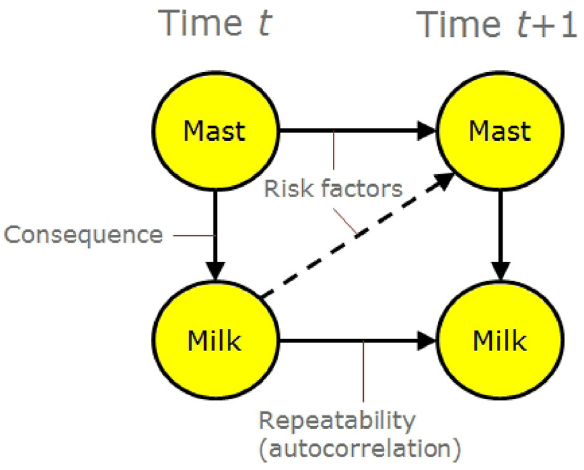


Fig. 1. A simple biological model of a dairy cow over two time stages.

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