Embedding a State Space Model Into a Markov Decision Process

Lars Relund Nielsen*, Erik Jørgensen and Søren Højsgaard

Research Unit of Bioinformatics, Genetics and Statistics, Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, P.O. Box 50, DK-8830 Tjele, Denmark

January 28, 2010

Abstract: In agriculture Markov decision processes (MDPs) with finite state and action space are often used to model sequential decision making over time. For instance, states in the process represent possible levels of traits of the animal and transition probabilities are based on biological models estimated from data collected from the animal or herd.
State space models (SSMs) are a general tool for modeling repeated measurements over time where the model parameters can evolve dynamically.
In this paper we consider methods for embedding an SSM into an MDP with finite state and action space. Different ways of discretizing an SSM are discussed and methods for reducing the state space of the MDP are presented. An example from dairy production is given.

Keywords: state space model, Markov decision process, sequential decision making, stochastic dynamic programming

1 Introduction

Many decision problems are dynamic in nature and must hence be re-evaluated over time based on the state of some crucial underlying variables, e.g. animal weight, number of treatments, milk yield etc. Examples of sequential decisions in animal production include replacement of animals, insemination and medical treatment. Often these problems can be modeled using *Markov decision processes* (*MDPs*) which have been widely used to model stochastic environments due to their analytical tractability. Some examples of MDPs applied to agriculture over the last decades are Stott, Jones, Humphry, and Gunn [30], Yalcin and Stott [35], Kristensen and Søllested [20], Houben, Huirne, Dijkhuizen, and Kristensen [11], Lien, Kristensen, Hegrenes, and Hardaker [22].

*Corresponding author, lars@relund.dk. Present address: Centre for Operations Research Applications in Logistics (CORAL) Department of Business Studies, Aarhus School of Business, University of Aarhus.

Time-series of on-line monitoring are often available today due to bio-sensors. *State space models* (*SSMs*) provide statistical methods modeling patterns of interest from large datasets obtained using on-line sensors. However, as shown in the following, only a few examples exist where data from on-line monitoring is used in combination with a formal framework for economic optimization of sequential decisions, such as MDPs.

The objective of this study is to present a general framework for embedding an SSM into an MDP with finite state and action space. Different ways of discretizing the SSM is discussed and methods for reducing the state space of the MDP are presented. An example from dairy production is given.

The paper is organized as follows. A short overview over existing literature on sequential decision making and online monitoring in animal production is given in the remaining of this section. In Section 2 we provide preliminaries of MDPs and SSMs. Section 3 present the framework for embedding an SSM into an MDP. How decisions and new information will affect the SSM is discussed in Section 4 and in Section 5 we consider techniques for reducing the state space of the MDP. Finally, in Section 6 discussion and conclusions are given.

1.1 Sequential decision making

Sequential decision-making problems can be modeled with MDPs. At a specified point in time, a decision maker observes the state of a system and chooses an action. The action choice and the state produce two results: the decision maker receives an immediate reward (or incurs an immediate cost), and the system evolves probabilistically to a different state at a subsequent discrete point in time. At this subsequent point in time, the decision maker faces a similar problem, but the system may be in a new state. The goal is to find a policy of actions (dependent on the observation of the state) which maximizes the reward after a certain time. Finding an optimal policy of an MDP is a well studied topic. For an overview see Puterman [29].

The definition of the state in the MDP must comprise all relevant information about the system being observed. In real-world models, the system is usually described by several discrete state variables, each representing a relevant trait. The overall state space is then defined as the cartesian product of the individual value spaces of the state variables resulting in that the number of states grow exponentially in the number of state variables. This explosion of the state space is referred to as the “curse of dimensionality” in literature and has historically been a major problem in the application of the method.

Hierarchical MDPs (*HMDPs*) were introduced by Kristensen [17] and Kristensen and Jørgensen [16]. The basic idea is that if a state variable is constant over a number of stages, the state variable should be represented by a so-called child process, which again may expand stages further to new child processes leading to multiple levels. Each child process has a finite time-horizon and the process at the top level has an infinite time-horizon. As a result an HMDP may be considered as an MDP with infinite time-horizon and parameters defined in a special way, but nevertheless in accordance with all usual rules and conditions relating to such processes.

HMDPs contribute to the reduction of the “curse of dimensionality” problem, because the number of states in the infinite time-horizon process can be reduced. Hence the solution methods for finding the optimal policy may become computationally tractable. HMDPs are especially suited to cyclic production systems, as often encountered within agriculture. A standard software system for solving both MDPs and HMDPs has been developed by Kristensen [19].

1.2 Online monitoring in agriculture

In agriculture the use of on-line sensors for monitoring production, detection of oestrus and disease in dairy herds has been suggested for well over two decades. Pedometric activity sensors, milk temperature, milk yield, and milk composition combined with regular hormonal measures have all been suggested as candidates for this monitoring see e.g. Firk, Stamer, Junge, and Krieter [7] and Cornou [3]. Often the amount of data collected using on-line monitoring is large since data is collected regularly over a relatively long period, e.g. a lactation period of the animal or the lifetime of the animal, or within plant production, the crop production during a year on a field or the entire sequence of crops on a field during 10 years.

State space models (SSMs) provide statistical methods modeling patterns of interest from large datasets obtained using on-line sensors. SSMs consists of a set of latent variables and a set of observed variables at each time-point. At a specified point in time the conditional distribution of the observed variables is a function of the latent variables specified via the observation equations. The latent variables changes over time, as described via the system equations. The observations are conditionally independent given the latent variables. Thus the value of the latent variables at the time point may be considered as the state of the system, and the SSM framework allows to predict the latent variables/state of the system via the observed variables, both the current state and the future development in the state variables. Examples of SSMs applied to agricultural problems are: Thysen [31], who monitored somatic cell count of dairy cows at herd level, de Mol, Keen, Kroeze, and Achten [4] which use a multivariate SSM for automated oestrus and mastitis detection, Madsen, Andersen, and Kristensen [23] who use an SSM for predicting drinking patterns of young pigs, and Van Bebber, Reinsch, Junge, and Kalm [33] which formulate an SSM for monitoring daily milk yield.

One way of applying SSM models in decision support systems is to trigger an alarm if the predicted value exceed some predetermined bound. The manager can then react to the alarm. The goal is to assist the manager in reaching his main objective, namely, to maximize his economical revenue under constraints imposed by health, regulation and management considerations. However, SSMs have only rarely been used in economic sequential decision models calculating optimal decisions.

1.3 Combining sequential decision making and online monitoring

The use of SSMs in sequential decision models are rare, but examples may be found. Kennedy and Stott [14] (although not explicitly stated) use an SSM to predict yield potential for a lactation based on previous lactation yields and embed the yield potential in the MDP. In Kristensen and Søllested [21] and Jørgensen [13] a sow replacement model is formulated. An SSM for the litter size of the sow is formulated and embedded into the MDP. Lien et al. [22] present a model investigating the optimal economic life cycle of grass leys with winter damage problems in northern Norway. The model embeds an extended SSM for updating the knowledge of yield potential and damage level.

The limited use of SSMs in sequential decision models may be due to that a framework for embedding an SSM into an MDP has not been explicitly pointed out in any papers. The goal of this paper is to provide such a framework.

2 Preliminaries

In this section we present some preliminaries of Markov decision processes and state space models.

2.1 Markov decision processes

Markov decision processes are models for sequential decision making when outcomes are uncertain. We consider a multi-level hierarchical MDP (*HMDP*), which is a series of MDPs with finite time-horizon (called child processes) built together into one MDP with infinite time-horizon called the founder process. In all processes a finite state and action space is assumed.

2.1.1 Ordinary MDPs with finite time-horizon

To have a frame of reference we first introduce the notation of an ordinary MDP having a finite time-horizon $\{1,...,T\}$ and $T-1$ stages. That is, decision number t is made at the beginning of stage t which corresponds to the time interval from decision number t to decision number $t+1$ (not including this time point).

At stage t the system occupies a *state*. We denote the finite set of system states $\mathscr{S}_t$. Given the decision maker observes state $s \in \mathscr{S}_t$ at stage t, he may choose an *action* a from the set of finite allowable actions $\mathscr{A}_{s,t}$ generating *reward* $r_t(s,a)$ (a cost if negative). Moreover, we let $p_t(\cdot \mid s,a)$ denote the *probability distribution* or *transition probabilities* of obtaining states $s' \in \mathscr{S}_{t+1}$ at stage $t+1$.

Since no decision is made at the end of stage $T-1$, the reward at this point of time is a function of the state $s \in \mathscr{S}_T$ denoted $r_T(s,a_T)$ where a_T denotes a dummy action. Moreover, at stage one we assume without loss of generality that we only have a single state and action (can be obtained by inserting a dummy stage).

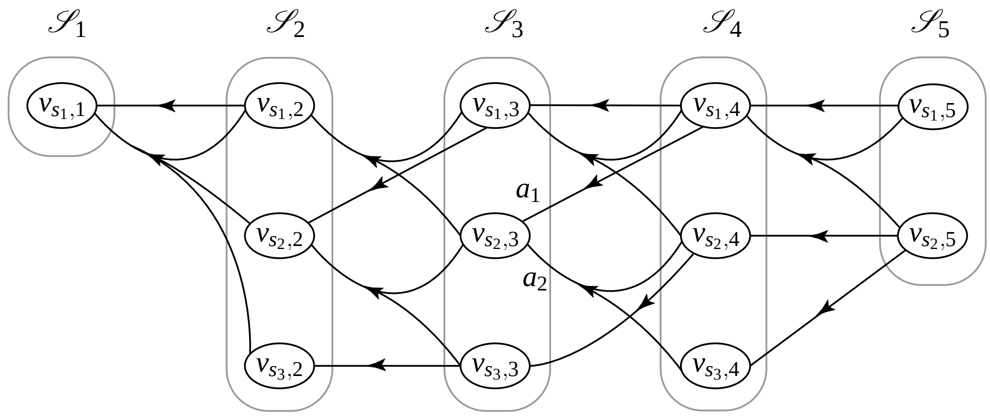


Figure 1: A state-expanded hypergraph for an MDP with time horizon $T = 5$. At stage t each node $v_{s_j,t}$ corresponds to a state in $\mathscr{S}_t$. The hyperarcs correspond to actions, e.g. if the system at stage 3 is in state s_2 then there are two possible actions. Action a_1 results in a deterministic transition to state s_1 (because there is only one tail) at stage 4 and a_2 results in a transition to either state s_2 or s_3 with a certain probability. For further details see Example 1.

A *decision rule* at stage t is a function $\delta_t : \mathscr{S}_t \to \mathscr{A}_{s,t}$ which specifies the action choice given state s at stage t. We let $\mathscr{D}_t$ denote the set of possible decision rules at stage t.

A *policy* or *strategy* specifies the decision rules to be used at all stages and provides the decision maker with a plan of which action to take given stage and state. That is, a policy δ is a sequence of decision rules, $\delta = (\delta_1, ..., \delta_T)$ with $\delta_t \in \mathscr{D}_t$ for $t = 1, ..., T$.

Given a criterion of optimality, e.g. the expected total discounted reward criterion the policy δ which maximizes the expected total discounted reward can be found by analyzing a sequence of simpler inductively defined single-stage problems. This is often referred to as *value iteration* or *dynamic programming* [29, chap. 4].

2.1.2 Illustrating an MDP using a directed hypergraph

A *directed hypergraph* is a pair $\mathscr{H} = (\mathscr{V}, \mathscr{E})$ where $\mathscr{V} = (v_1, ..., v_{|\mathscr{V}|})$ is the set of *nodes* and $\mathscr{E} = (e_1, ..., e_{|\mathscr{E}|})$ is the set of *hyperarcs*. A hyperarc $e \in \mathscr{E}$ is a pair $e = (T(e), h(e))$ where $T(e) \subset \mathscr{V}$ denotes the set of *tail* nodes and $h(e) \in \mathscr{V} \setminus T(e)$ denotes the *head* node. Note that a hyperarc has exactly one node in the head, and possibly several nodes in the tail. For an illustration see Figure 1.

Directed hypergraphs represent a general modeling and algorithmic tool, which has been used in many different research areas such as artificial intelligence, database systems, fuzzy systems, propositional logic, and transportation networks [8, 25, 26]. For a general overview on directed hypergraphs see [1].

As pointed out in Nielsen and Kristensen [24] a finite-horizon MDP can be represented using a *state-expanded directed hypergraph* as is illustrated in Example 1.

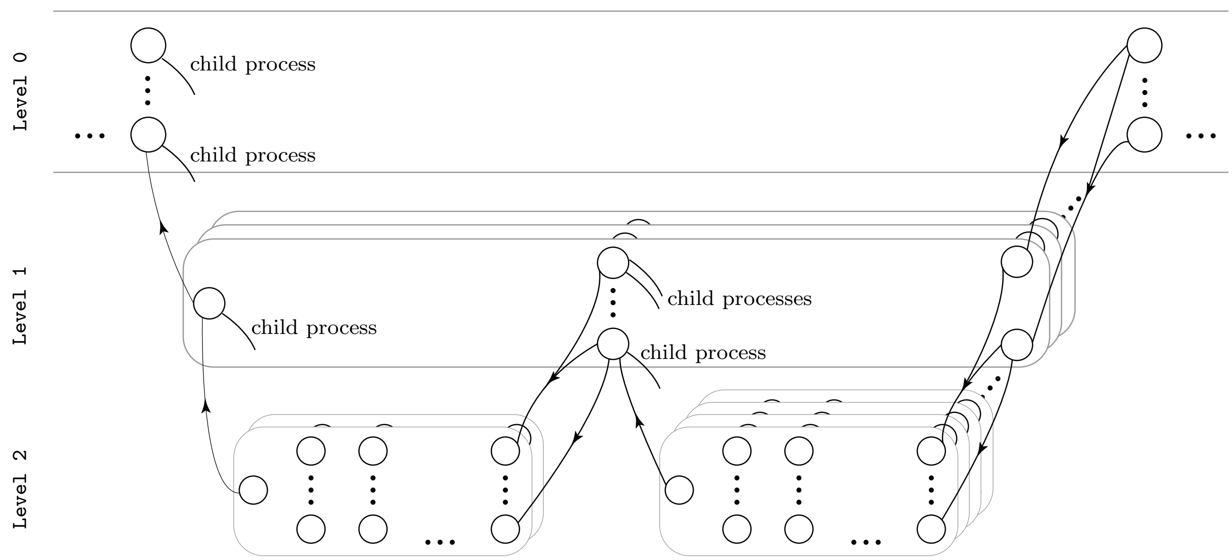


Figure 2: A hypergraph representation of a stage in an HMDP. Level 0 indicate the founder level, and the nodes indicates states at the different levels. A child process (oval box) is represented using its state-expanded hypergraph (hyperarcs not shown) and is uniquely defined by a given state and action of its parent process.

Example 1 A hypergraph $\mathscr{H}$ representing an MDP with time-horizon $T = 5$ is shown in Figure 1. Each node corresponds to a specific state in the MDP and a hyperarc is defined for each possible action. For instance, node $v_{s_2,3}$ corresponds to a state $s_2 \in \mathscr{S}_3$. The two hyperarcs with head in node $v_{s_2,3}$ show that two actions are possible given state s_2. Action a_1 corresponds to a deterministic transition to state s_1 and action a_2 corresponds to a transition to state s_2 or s_3 at stage 4 with a certain probability greater than zero. $\square$

The hypergraph representation provides us with a way to visualizing the MDP which will be used though out the paper. Furthermore, by assigning relevant weights to the hyperarcs the optimal policy of the MDP can be found by processing the nodes backward in time (the reason that the hyperarcs points opposite time). This is beyond the scope for this paper. For further details see [24].

2.1.3 Hierarchical MDPs

A *hierarchical MDP* (*HMDP*) is an infinite stage Markov decision process with parameters defined in a special way, but nevertheless in accordance with all usual rules and conditions relating to such processes. The basic idea of the hierarchical structure is that stages of the process can be expanded to a so-called *child process*, which again may expand stages further to new child processes leading to multiple levels.

To illustrate consider a stage of an HMDP shown in Figure 2. The process has three levels. At `Level 2` we have a set of ordinary MDPs with finite time-horizon (one for each oval box) which all can be represented using a state-expanded hypergraph (hyperarcs not shown, only hyperarcs connecting processes are shown). An MDP at `Level 2` is uniquely defined by a given state s and action a of its *parent process* at `Level 1` (illustrated by the arcs with head and tail node at `Level 1` and `Level 2`, respectively). Moreover, when a child process at `Level 2` terminates a transition from a state $s \in \mathscr{S}_T$ of the child process to a state at the next stage of the parent process occur (illustrated by the (hyper)arcs having head and tail at `Level 2` and `Level 1`, respectively). Note that information in the state of the parent process defining the child process may be used throughout the child process to define e.g. transition probabilities of the child process.

Similarly, at `Level 1` we have a set of MDPs with finite time-horizon. However, the rewards and transition probabilities, at a stage of a process at `Level 1` are equal to the total expected rewards and transition probabilities of the corresponding child process under the policy chosen for the child.

At `Level 0` only a single process exists. We refer to that process as the *founder process.* Note that the founder process has an infinite time-horizon (only one stage is shown in Figure 2). Since the founder process is running over an infinite number of stages, we assume stationary state and action spaces, i.e. the states, action and transition probabilities of the child processes are the same when we consider different stages of the founder process.

Figure 2 shows that a directed hypergraph can be used to represent a stage of the founder process. For each child process a dummy arc is inserted to represent the transition from the parent process to the child process. Furthermore, hyperarcs are inserted to represent the transition from the child process back to the parent process.

Since the founder process has an infinite time-horizon the optimal policy maximizing the expected total discounted reward can be found using a modified policy iteration algorithm [16].

2.2 Linear normal state space models

State space models are models of phenomena evolving in time e.g. blood pressure and milk yield.

A set of q-dimensional latent continuous variables $\theta_{\{t=0,1,\dots\}}$ evolves as a first order Markov process using *system equation*

$$\theta_t = G_t \theta_{t-1} + \omega_t, \tag{1}$$

with $\omega_t \sim N(0, W_t)$. We assume that $\theta_0 \sim N(m_0, C_0)$. Moreover, we have a set of r-dimensional observable continuous variables $Y_{\{t=1,2,\dots\}}$ which are dependent on the latent variable using *observation equation*

$$Y_t = F_t' \theta_t + \nu_t, \tag{2}$$

with $\nu_t \sim N(0, V_t)$. The error sequences ω_t and ν_t are internally and mutually independent. Hence given θ_t we have that Y_t is independent of all other observations and in general the past and the future are independent given the present.

Example 2 Throughout the paper we will use the model from Nielsen, Jørgensen, Kristensen, and Østergaard [27] to exemplify. This model is intended for use in dairy herds, and the core of the model is an SSM describing the milk yield of the cows. The milk yield of cows vary both due to permanent factors such as breeding value and due to short term influences e.g. disease, change of feed etc. An SSM model is well suited for this purpose. The specification of the SSM is summarized below.

Consider a cow in a dairy herd. Let $Y_{t,j}$ be the *residual milk yield* at day t (measured as days from calving) in lactation j. To keep the notation simple we do not consider indices for the cow and herd. We assume the following model:

$$Y_{t,j} = M_{t,j} - \mu_{t,j} = A_j + X_{t,j} + \nu_{t,j}. \tag{3}$$

where the variables correspond to:

$M_{t,j}$ The milk yield of the cow.

$\mu_{t,j}$ Mean milk yield curve for the herd at lactation j.

A_j Production potential of the specific cow in lactation j. It is assumed that $A_j \sim N(0, \sigma_A^2)$.

$X_{t,j}$ An autoregressive process of order one with mean zero and autocorrelation $\rho(t) = \rho^t$, $|\rho| < 1$, i.e.

$$X_{t,j} = \rho X_{t-1,j} + \varepsilon_{t,j}.$$

Letting $\mathrm{var}(X_{t,j}) = \sigma_X^2$ the auto-covariance becomes $\mathrm{cov}\left(X_{t,j}, X_{t+t',j}\right) = \sigma_X^2 \rho\left(t'\right)$. $X_{t,j}$ may be considered as short term environmental influence on milk yield.

$\nu_{t,j}$ Mutually independent random variables describing the measurement error. It is assumed that $\nu_{t,j} \sim N(0, \sigma_\nu^2)$.

To illustrate, the daily milk yield $M_{t,3}$ for three cows at lactation 3 in a Danish herd is shown in Figure 3(a). The mean milk yield curve for the herd $\mu_{t,3}$ (estimated using a mixed model) is given by the solid black line. Note that cow 1 is below average, cow 2 around average and cow 3 above average. The corresponding residual milk yield $Y_{t,3}$ is shown in Figure 3(b). A value of $Y_{t,3}$ above zero denotes that the yield is above average.

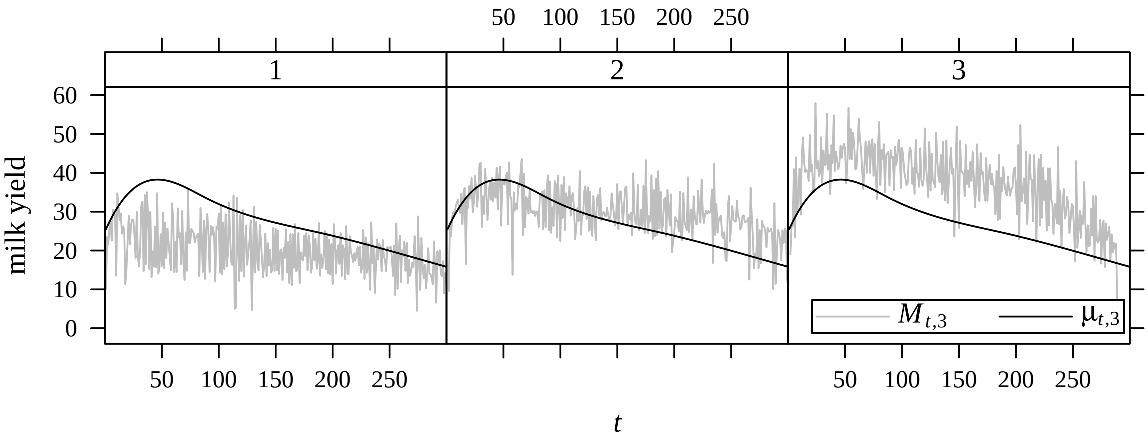
An SSM of (3) is specified for each lactation j using the following measurement equation and system equation

$$Y_{t,j} = F'\theta_{t,j} + \nu_{t,j} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} A_j \\ X_{t,j} \end{pmatrix} + \nu_{t,j}, \tag{4a}$$

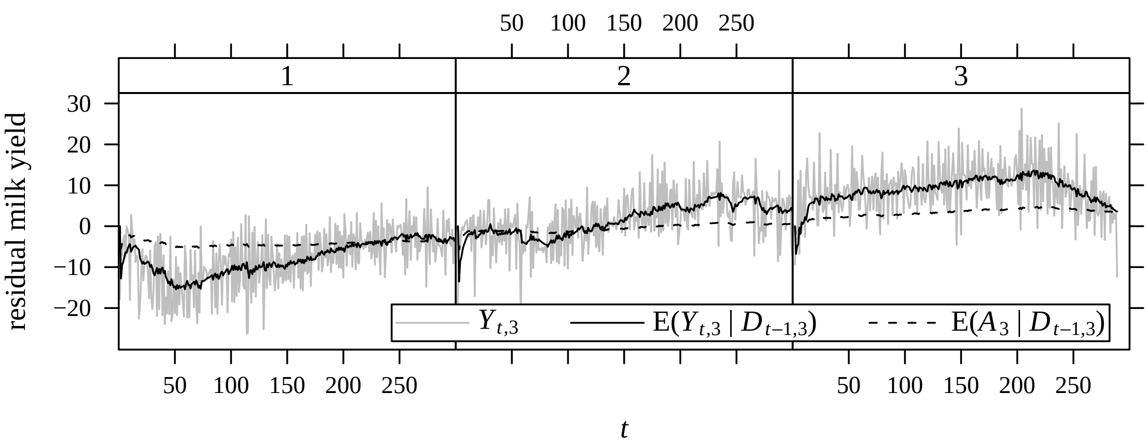
$$\theta_{t,j} = G\theta_{t-1} + \omega_{t,j} = \begin{pmatrix} A_j \\ X_{t,j} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} A_j \\ X_{t-1,j} \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_{t,j} \end{pmatrix}, \tag{4b}$$

where $\omega_{t,j} \sim N\left(0, W\right)$ with

$$W = \begin{pmatrix} 0 & 0 \\ 0 & \left(1-\rho^2\right)\sigma_X^2 \end{pmatrix}.$$



(a) Daily milk yield and herd mean curve for three different cows.



(b) Residual milk yield and forecasted values for three different cows.

Figure 3: Milk yield and Residual milk yield for 3 different cows (lactation 3).

Equation (4a) is the model (3) written using matrix notation and (4b) describe the relations between the latent variables from time $t-1$ to time t.

The mean and covariance matrix of the prior $\theta_{0,j} \sim N\left(m_{0,j}, C_{0,j}\right)$ is

$$m_{0,j} = \begin{pmatrix} \hat{A}_j \\ 0 \end{pmatrix}, \quad C_{0,j} = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix}. \tag{5}$$

We take $\hat{A}_1 = 0$ and $\hat{A}_j = \hat{A}_{j-1},\ j > 1$.

The variance structure is described by the hyper-parameters

$$\psi = (\sigma_A, \sigma_v, \sigma_X, \rho). \tag{6}$$

To estimate the hyper-parameters we need several time-series of the daily yield from different cows and lactations in the herd. Using for instance a mixed model on the previous time-series from the herd, we can now simultaneously estimate (6) and $\mu_{t,j}$. The hyper-parameters describe the "between cows" variances in the herd, i.e. they are the same for the SSMs monitoring different cows in the herd. □

As illustrated in the example above the SSM is formulated such that it is valid for each cow in a given lactation. That is, we have an instance of the SSM for each cow and lactation where the hyper-parameters $(m_0, C_0, F_t, G_t, V_t, W_t)$ of the SSM is the same. Since the hyper-parameters are estimated within a population we will call these estimates *population based estimates*.

We will not go into detail how to find the population based estimates. An example focusing on the present context is presented in Toft and Jørgensen [32]. Similar methods are used routinely for evaluation of breeding values.

In the decision model we will assume that the hyper parameters are known. Given these estimates interference about the latent variable can be made. Let $D_{t-1} = (Y_1, \dots, Y_{t-1}, m_0, C_0)$ denote the information available up to time $t-1$. Given the posterior of the latent variable at time $t-1$ we can use the Kalman filter to update the distributions at time t [34, Thm 4.1]. Suppose that at time $t-1$ we have

$$(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1}), \quad \text{(posterior at time } t-1\text{)}.$$

then

$$\begin{aligned} (\theta_t \mid D_{t-1}) &\sim N(b_t, R_t), && \text{(prior at time } t\text{)} \\ (Y_t \mid D_{t-1}) &\sim N(f_t, Q_t), && \text{(one-step forecast at time } t-1\text{)} \\ (\theta_t \mid D_t) &\sim N(m_t, C_t), && \text{(posterior at time } t\text{)} \end{aligned}$$

where

$$\begin{aligned} b_t &= G_t m_{t-1}, & R_t &= G_t C_{t-1} G_t' + W_t \\ f_t &= F_t' b_t, & Q_t &= F_t' R_t F_t + V_t \\ e_t &= Y_t - f_t, & B_t &= R_t F_t Q_t^{-1} \\ m_t &= b_t + B_t e_t, & C_t &= R_t - B_t Q_t B_t'. \end{aligned} \tag{7}$$

In a Kalman filter iteration the *predict step* corresponds to calculating b_t and R_t (f_t and Q_t) while the *correction step* corresponds to finding m_t and C_t.

Note that the one-step forecast mean f_t only depends on m_{t-1}, i.e. we only need to keep the most recent conditional mean of θ_{t-1} to forecast the next value. Hence when making prediction based on D_{t-1}, we need only to store m_{t-1}. Similar the variance Q_t only depends on the number of observations made, i.e. we can calculate a sequence $Q_1, \dots, Q_t$ without knowing the observations $Y_1, \dots, Y_t$.

Example 2 (continued) If we apply the Kalman filter on a subset of the cows in the data set using the predetermined estimates of (6) (found using statistical analysis on historical data from the same population), we get a result as shown in Figure 3(b). The forecasted daily milk yield $f_{t,3} = \mathrm{E}(Y_{t,3} \mid D_{t-1,3})$ is shown with solid black line. As mentioned this expectation can be calculated based on the two-dimensional vector $m_{t-1,3}$, and Q_t that only depends on t. The prior estimate of A_3 at $t = 0$ is set to zero and the estimate of A_3 slowly adapts to the level of the cow.

An estimate of A_3 above zero corresponds to a cow we predict to yield better than an average cow and an estimate of A_3 below zero to a cow we predict yielding worse than average. The estimate of A_3 is important since in a decision model a high estimate will correspond to a high yielding cow that we want to keep longer than a low yielding cow. □

3 Embedding an SSM into a single policy HMDP

An SSM is in general a process following a specific policy, e.g. the policy δ to do nothing. Decisions deviating from this policy can affect the SSM in different ways, e.g. the process may stop due to dropout or the mean change. In this section we consider the SSM under policy δ.

MDPs requires all states to be discrete and observable. Hence two problems arise: 1) how can the latent variable given the current information be embedded into the MDP and 2) how do we discretize the continuous variables efficiently. We consider these problems in the following two sections.

3.1 Embedding the latent variable

The latent variable cannot be observed but the distribution of θ_t given all previous information can be calculated. The benefit of modeling the problem by the system equation (1) and observation equation (2) is that the distribution of θ_t given all previous information only depends on m_t and C_t which can be found using the previous observation and the prior of the latent variable (see (7)). As a result we do not have to store all previous information in the set of states of the MDP, this information can be represented by storing m_t and C_t instead without any loss of information at all.

Notice that if we assume that the information stored in the states of the MDP have no influence on matrices G_t, F_t, W_t and V_t then the variance of the latent variable only depends on the number of observations made, i.e. time-instance t. Hence C_t does not have to be stored in the MDP since it is independent of the state at time t. This is assumed throughout the rest of the paper.

3.2 Discretizing the state space

To embed the SSM in an MDP, observations Y_t and values of the conditional mean m_t has to be discretized. As a result the MDP will be an approximation of the SSM. In general the continuous space is partitioned into discrete regions which are estimated using a specific point in the region.

Definition 1 Let an n-dimensional *hypercube* Π be defined as $\Pi = \times_{i=1,\ldots,n} I_i$, where I_i is an open or closed interval of $\mathbb{R}$

A *rectangular partition* of $\mathbb{R}^n$ is a finite set of hypercubes $\Omega = \{\Pi_1, \ldots, \Pi_k\}$ such that $\bigcup_{i=1,\ldots,k} \Pi_i = \mathbb{R}^n$ and $\Pi_j \cap \Pi_i = \emptyset$, $i \neq j$.

Given hypercube Π, let *center point* π denote a specific point in Π defined as the center of the hypercube, or as the mean point in the hypercube with respect to some distribution.

Let $\Omega^m = \{\Pi_1^m, \ldots, \Pi_h^m\}$ denote a rectangular partition of $\mathbb{R}^q$ in which m_t takes values (we use the same partition for all t). Similar let $\Omega^Y = \{\Pi_1^Y, \ldots, \Pi_l^Y\}$ denote a partition of Y_t.

Consider the prior $(\theta_t \mid D_t) \sim N(m_t, C_t)$ at time t and assume that $m_t \in \Pi_i^m$. We approximate the distribution of $(\theta_t \mid D_t)$ by replacing m_t with the center point of Π_i^m, i.e.

$$(\theta_t \mid D_t) \sim_A N(\pi_i^m, C_t)$$

which implies that the one-step forecast is

$$(Y_{t+1} \mid D_t) \sim_A N(F' G \pi_i^m, Q_{t+1}). \tag{8}$$

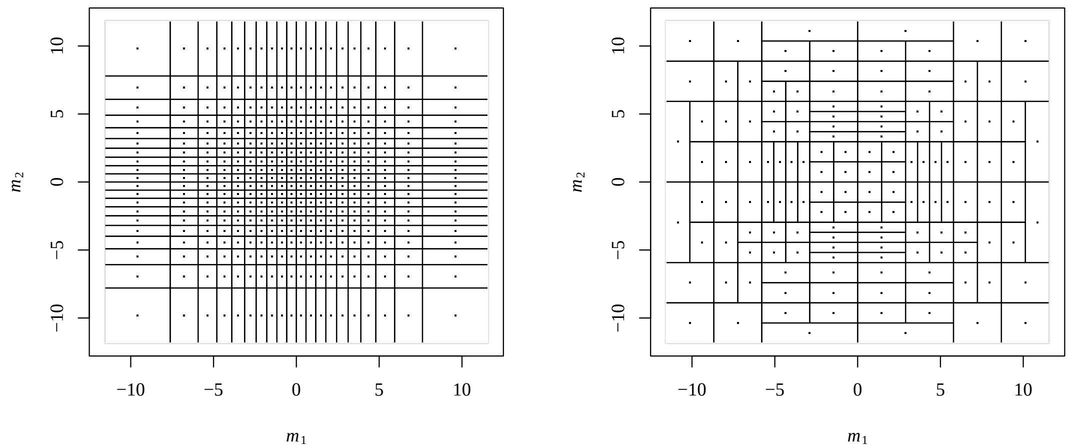
The rectangular partition Ω^Y is used to discretize the density of $(Y_{t+1} \mid D_t)$ as follows

$$\varphi_{t+1}(\Pi_k^Y \mid \Pi_i^m) = P(Y_{t+1} \in \Pi_k^Y \mid \pi_i^m, Q_{t+1}) = \int_{\Pi_k^Y} p(y) dy, \tag{9}$$

where $p(\cdot)$ denote the (multivariate) normal density of (8).

In general Ω^m and Ω^Y should be defined such that they represent a good approximation of the distributions of the variables. To measure the quality or error introduced by the discretization we use the *Kullback-Liebler (KL) distance* between two probability density functions [15]. Small KL corresponds to a good approximation and if KL is zero the two probability density functions are equal.

Traditionally, in the literature [21, 22] a rectangular partition Ω representing the discrete approximation of a multivariate normal distributed variable Z is obtained using an *univariate discretization*. Each variable of Z is discretized separately and the rectangular partition is obtained by taking the cross product of the intervals obtained for each variable. However, a rectangular partition of the same quality containing fewer hypercubes can be obtained using an *multivariate discretization*, which discretizes multidimensional domains as a whole, rather than discretizing each variable separately. The algorithm works by maintaining a priority queue of hypercubes and the hypercube with largest KL distance is selected and divided into a two smaller hypercubes which are inserted into the priority queue. The step is repeated until the specified quality of the approximation is obtained. For more details see Kozlov and Koller [15].



(a) Univariate. (b) Multivariate.

Figure 4: Rectangular partition of the mean $m_{t,3}$ using univariate and multivariate discretization with a KL distance approximate equal to 0.2. Using a multivariate discretization reduce the number of states by 64%.

Example 2 (continued) Assume that we base the discretization of $m_{t,3}$ on the distribution of the prior $\theta_0 \sim N((0,0)', C_0)$ with covariance matrix estimated to

$$C_{0,3} = \begin{pmatrix} 21.44 & 0 \\ 0 & 22.46 \end{pmatrix}$$

Using an univariate discretization where we divide each variable into 20 intervals each having a probability of 0.05 we get a rectangular partition representing 400 states as illustrated in Figure 4(a). Note that some hypercubes are unbounded such that the rectangular partition satisfies Definition 1. The center points for each hypercube (illustrated with a dot) are found by taking the mean over each hypercube.

Using a multivariate discretization we get a rectangular partition representing 143 states as illustrated in Figure 4(b).

Both approximations of the distribution have a KL distance approximately equal to 0.2. Note that using a multivariate discretization results in that the number of states is reduced by 64%. □

Note, there is a tradeoff between making the approximation error small and the number of states in the HMDP. Thus a poor approximation may result in that the optimal values and policies of the HMDP are not correct with respect to the problem under consideration. On the other hand, an approximation with a small error may make the HMDP untractable due to the "curse of dimensionality". The suggested discretization using the KL distance ensures the closest approximation to the probability distribution. However, the overall ap-

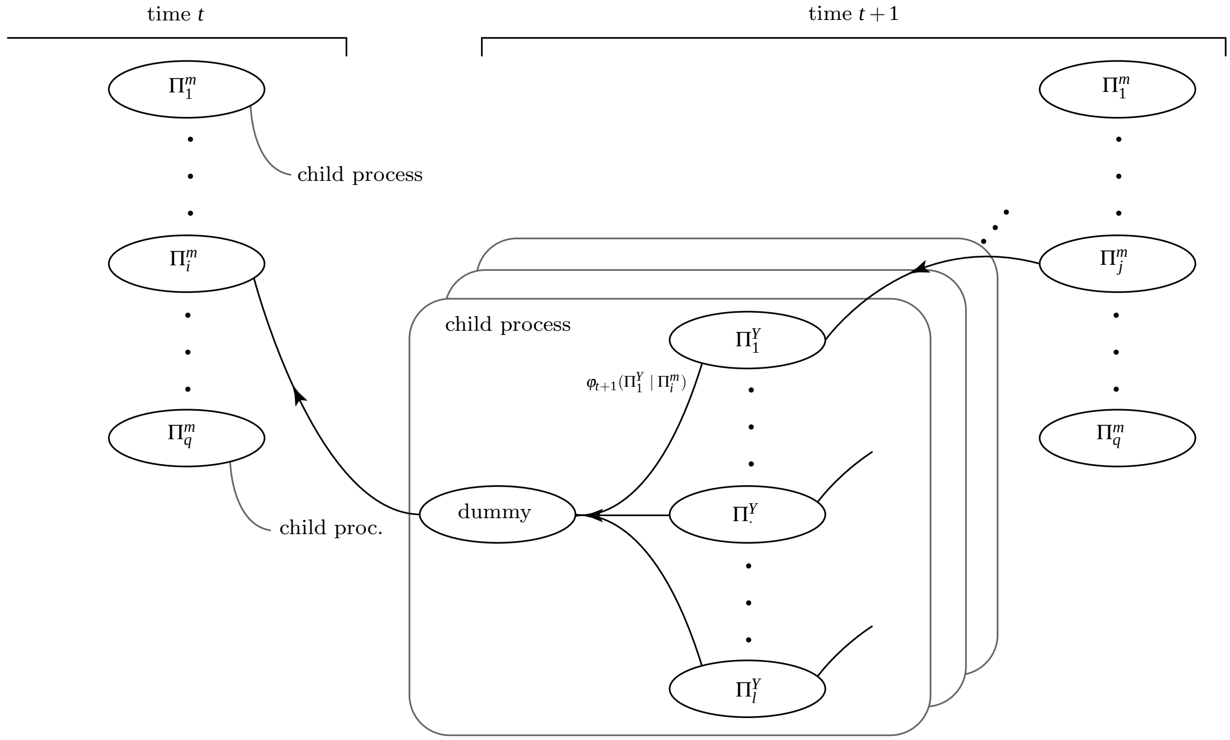


Figure 5: The SSM embedded into an HMDP from time t to $t+1$ (founder level not shown). At level 1 the means of the latent variables are stored. Node Π_i^m indicates that we consider state i where $m_t \in \Pi_i^m$. At the observations Y_{t+1} are kept. The `dummy` node signifies the start of the child process. The different child processes is indicated by the slices.

proximation error does also depend on the rewards. Thus, in practice an iterative approach may also be used by decreasing the approximation error and each time solve the HMDP until the difference in the optimal values are tolerable.

3.3 HMDP formulation

The SSM can be embedded in an HMDP. A stage at the founder level describes the lifetime of an individual within a population, e.g. the lifetime of a cow. The time-horizon is infinite. At the next stage of the founder process the individual is replaced with a new individual and hence the time-series used by the SSM is terminated and a new time-series is started using the population based estimates.

At child levels the SSM is embedded as illustrated in Figure 5 (founder level not shown). At the first level the means of the latent variable are stored. The set of states at time t is $\mathscr{S}_t = \Omega^m$. For each state $s \in \mathscr{S}_t$ a dummy action (the action of doing nothing) is used to define a child process. The child process is a one stage MDP starting with a dummy state (representing the start of the process) and at the next decision epoch the set of states is $\mathbf{\Omega}^Y$. The transition probability of the transition from the dummy state to state Π_k^Y is defined using (9)

$$P(\Pi_k^Y \mid \texttt{dummy}) = \varphi_{t+1}(\Pi_k^Y \mid \Pi_i^m). \tag{10}$$

Given state Π_i^m (defining the child process) and state Π_k^Y (the state in the child process) we can update the mean of the latent variable at time $t+1$ using the center points of Π_i^m and Π_k^Y. Using the formulas in (7), let

$$\check{m}_{t+1} = G\pi_i^m + B_t(\pi_k^Y - F'G\pi_i^m)$$

and let the transition probability from state Π_k^Y back to the parent process at time $t+1$ be

$$P(\Pi_j^m \mid \Pi_k^Y) = \begin{cases} 1 & \text{if } \check{m}_{t+1} \in \Pi_j^m \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

as illustrated by the arc from the child process to the parent process in Figure 5.

The rewards of the process for each state and action is calculated as a function of the states in the process, e.g. in the dairy cow example the reward would be calculated based on the observed milk yield.

Example 2 (continued) Using the multivariate discretization in Figure 4(b) we have $\mathbf{\Omega}^m$ consists of 143 states. Similar if we base the discretization of Y_t on the distribution of the prior $Y_0 \sim N(0, 7^2)$, we have that $\mathbf{\Omega}^Y$ consists of 11 states when the KL distance is equal to 0.2. □

4 Extending the model with interventions and decisions

In the previous section we have embedded the SSM in the MDP with only one (dummy) decision. In this section we will show how to implement decisions in the model. Many of the relevant interventions such as culling can be seen as premature termination of the process, but we will also present other types of intervention.

However, the first step is to implement process termination that takes place independently of the state variables in the model or without the control of the decision maker.

4.1 Random termination of the child process

In the literature on animal replacement models, see e.g. [35, 20, 11], this part of the model is often called *involuntary culling*, i.e. culling due to sudden death, illness etc. As an example, a large part of the so-called involuntary culling in sow herds is due to reproductive failure, which typically is a measure of poor productivity and not an indication of failure to produce. However, from a modeling point of view a better term is *random termination*

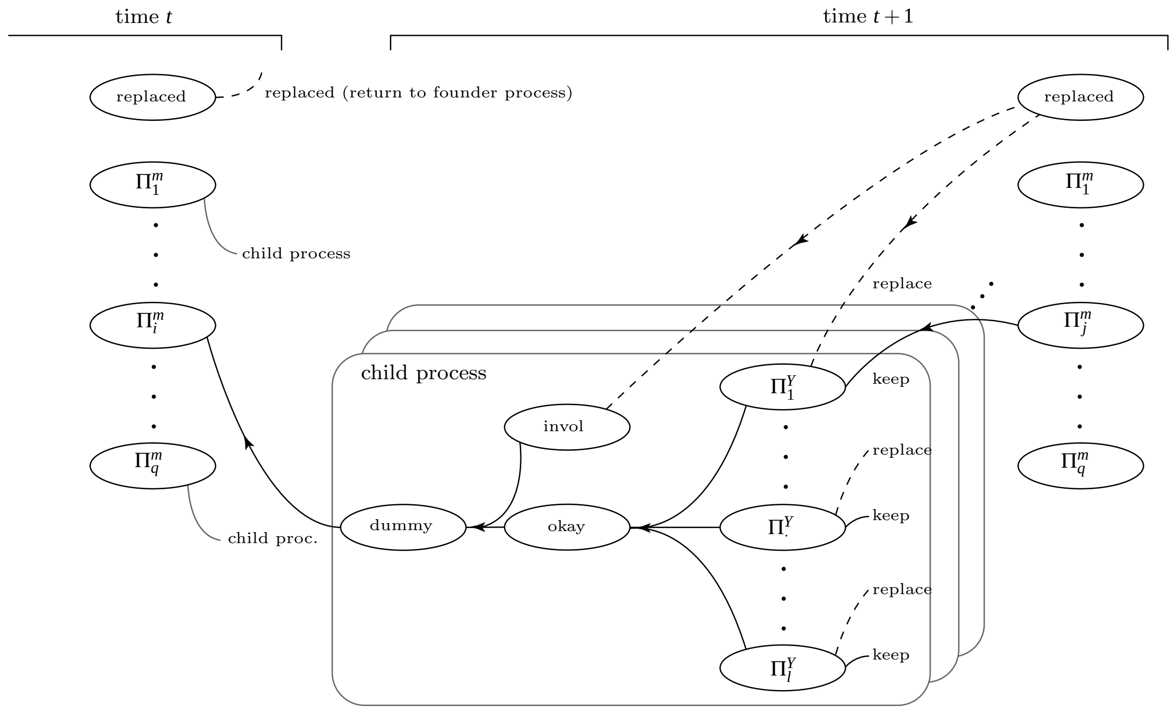


Figure 6: The HMDP for Example 2 (time t to $t+1$). Corresponds to Figure 5 but with additional nodes added. Dummy state `replaced`, represent that the cow has been replaced. The nodes `invol/okay` represent involuntary culling. Two actions `keep/replace` are possible. Dashed lines indicate replacement. For details see text.

(RT) since the time-series of the SSM is terminated randomly which corresponds to that the child process embedding the SSM is terminated randomly, i.e. not based on a decision.

Usually, the time-series used for the estimation of RT will be censored [5]. Observations used by the SSM may terminate prematurely, e.g. because of death or culling. This reflects both RT and the current decision policy. Thus estimating RT may be difficult. However, models can be formulated where it is possible to obtain an estimate of RT [20].

Example 2 (continued)

To represent involuntary culling in the HMDP the state space has to be expanded as illustrated in Figure 6.

A stage at the founder level (not shown) represent the lifetime of a cow, i.e. at each stage of the founder level we start monitoring a new cow inserted when the current cow has been replaced. Since the time-horizon of the founder process is infinite the founder process represent the current cow and all its successors. The process at the first level in Figure 6 has a new dummy state `replaced` added, representing that the cow has been replaced. If the system enters state `replaced` it makes a transition to the founder state with probability one and a new stage of the founder process is started, i.e. a new cow is inserted. Because replaced is a dummy state the time for a transition is zero.

At the second level two nodes are inserted to take involuntary culling into consideration. The process enters the involuntary culled state invol with probability $\hat{p}$ which is the rate of RT (dependent on lactation, days from calving, and pregnancy state). In this case the cow is replaced, i.e. the process returns to state replaced in the parent process. This is illustrated with the dotted arrow from replaced to invol. Otherwise the cow is in state okay with probability $(1-\hat{p})$ and the milk yield is observed at time $t+1$. □

4.2 Termination of the child process due to a decision

Often decisions are taken such that the time-series of the SSM is ended, e.g. we may end the life of a cow, or similarly, we may decide to establish a new crop on the field.

This is simple to handle within the framework, and is similar to the handling of RT, except that it is included as a decision variable as illustrated in the following example.

Example 2 (continued) The decisions whether to keep or replace the cow is represented in the HMDP after the milk yield, e.g. $Y_{t+1} \in \Pi_1^Y$, is observed. Replacing the cow makes the process return to state replaced in the parent process with probability one. Keeping the cow corresponds to letting the time-series of the SSM continue, i.e. we make a transition to a state Π_j^m in the parent process (see Figure 6). □

4.3 Non-terminating decisions

In some cases, we can give a realistic formulation of how an intervention will affect the states of the SSM [34, Chap. 11] even though that it is not terminating the time-series. If such interventions occur regularly they may be included in the HMDP as decisions. A decision representing an intervention could be that the animal is treated for an illness, that we start measuring the online variables at different time-intensities or that we change the feeding strategy etc.

Such decisions can affect the SSM in different ways: it may change the predicted mean and variance of the latent or observed variables or it may change the hyper-parameters of the model in a predictable manner. Suppose that the intervention information of the chosen decision is given by $I_t = \{h_t, H_t\}$, where h_t is the mean and H_t is the covariance matrix of the random variable $\xi_t \sim N(h_t, H_t)$. If the intervention is effected by adding ξ_t to (1) then

$$\theta_t = G_t \theta_{t-1} + \omega_t + \xi_t,$$

and hence (7) becomes

$$b_t^* = b_t + h_t \qquad R_t^* = R_t + H_t.$$

That is, the transition probabilities of the HMDP will be different compared to if the decision not was taken.

In many of these cases, the Kalman-filter technique will no longer give an exact representation of the HMDP. For instance if $H_t \neq 0$ then the variance at time t will now depend on the number of decisions made in the past. Hence the covariance matrix (or subsets of it) must be represented in the states of the HMDP, i.e. the states Ω^m in Figure 5 will be replaced with the cross product of Ω^m and an (approximation) set of possible covariance matrices.

4.4 Other types of information

The previous decisions and information could all be seen as a part of the model framework. However, other information may indicate, that the system has changed so much, that the population parameters of the model are no longer representative. Strictly speaking, new information may indicate that the historical data for estimating the population parameters in the SSM and in other parts of the model are no longer valid. If such relevant information suddenly becomes available then this information should be incorporated into the HMDP and afterwards the HMDP should be reoptimized to find a new optimal policy. Unexpected relevant information could be: a new illness suddenly present in the herd, high increase in market prices, new environmental conditions etc. Such new information may change the variance components of the SSM, make sudden shifts in the mean m_t of the latent variable ([34, Chap. 11]) and change the reward defined for the MDP.

5 Techniques for reducing the state space

Even though that classic dynamic programming algorithms solve MDPs in polynomial time in the size of the state space, the size of the state space may be very large in practice. If for instance the latent variable θ_t is of dimension four and we use a univariate discretization approach (see Section 3.2) which divide each variable into 10 intervals the number of states in Ω^m becomes 10^4. For each state in Ω^m a child process is defined containing the discretization of the observation variables (see Figure 5). If for instance Y_t is of dimension five and we use a univariate discretization approach with 10 intervals for each variable the number of states in Ω^Y becomes 10^5 and the total number of states for each time instance is 10^9. This is one example on the "curse of dimensionality" problem and techniques to overcome it are an important topic of research.

One approach is to represent the number of states implicit in the model using factored MDPs [2] which is beyond the scope of this paper. In this framework, a state in the MDP is implicitly described using a dynamic Bayesian network. This representation often allows an exponential reduction in the representation size of structured MDPs, but the complexity of exact solution algorithms for such MDPs can grow exponentially in the representation size. Techniques to avoid this growth are considered in [9]. Models with continuous variables are considered in [10, 6].

In this paper we consider classical dynamic programming methods which require an explicit state space enumeration. As shown in Example 2 on page 13 using a multivariate

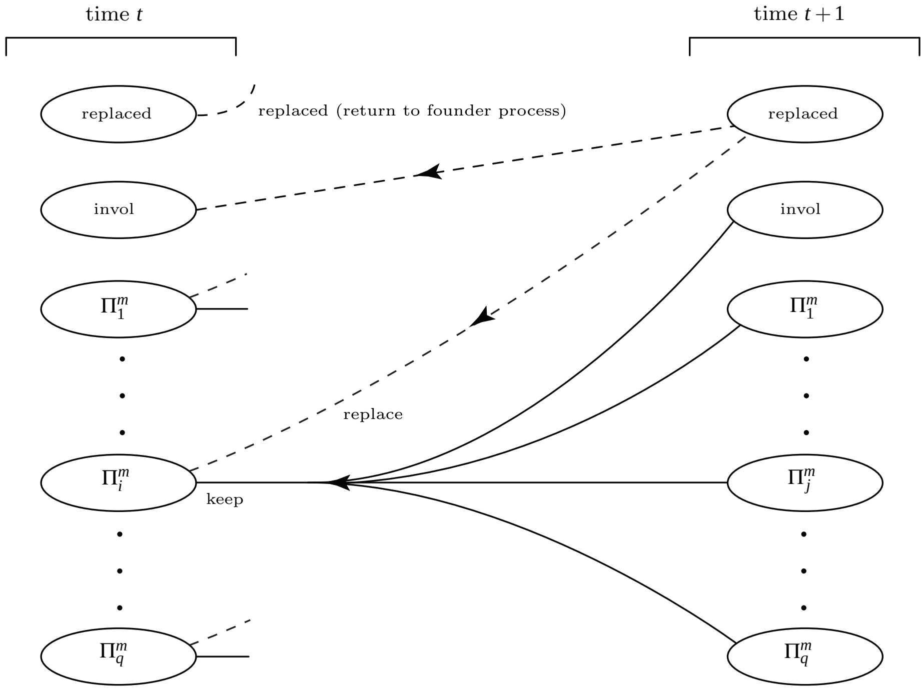


Figure 7: The HMDP for Example 2 (time t to $t+1$). The child process storing the observations has been removed. The state `invol` has been moved from the child process to the parent process and direct transitions probabilities have been specified.

discretization approach will reduce the number of states in $\boldsymbol{\Omega}^m$. The same approach can be used to discretize $\boldsymbol{\Omega}^Y$ on the child level. However, the child process defined for each state in $\boldsymbol{\Omega}^m$ can be removed since the mean m_t represent all information available up to time instance t.

Example 2 (continued)
Consider the HMDP in Figure 6. Here decisions are defined such that they are taken when an observation of the milk yield is given. However, the decision might as well be taken after we have calculated the new updated mean m_{t+1} in the parent process. As a result the child process only contains dummy actions and describe the transition from state Π_i^m at time t to a state in $\boldsymbol{\Omega}^m$ at time $t+1$. Furthermore, the state `invol` may be removed from the child process and placed in the parent process. Hence the child process can be removed from the MDP model and direct transitions in the parent process can be specified as illustrated in Figure 7. Here given state Π_i^m at time t we enter state `invol` with probability $\hat{p}$ and the transition probability from Π_i^m to Π_j^m at time $t+1$ is based on the transition probabilities defined in the child process containing only dummy actions.

By removing the child process the number of states used at stage t in the HMDP is reduced from $143 \cdot 15 + 1$ to $143 + 2$, i.e. 93%. □

Note that the child process representing the observation Y_t is a discrete approximation of the continuous observation variable Y_t. Since the child process can be removed from the HMDP process there is no need to discretize the density of the observation variable, in fact the density of $(m_{t+1} \mid m_t)$ can be calculated using standard formulas for an linear transformation of the multivariate normal distribution. From Equation (7) we have that

$$m_{t+1} = b_{t+1} + B_{t+1}(Y_{t+1} - f_{t+1}) = B_{t+1}Y_{t+1} + (b_{t+1} - B_{t+1}f_{t+1}) \tag{12}$$

and since $Y_{t+1} \sim N(f_{t+1}, Q_{t+1})$ is the only random variable in (12) at time t, we have that $(m_{t+1} \mid m_t) \sim N(\mu, \Sigma)$ with mean and covariance matrix

$$\begin{aligned} \mu &= B_{t+1}f_{t+1} + (b_{t+1} - B_{t+1}f_{t+1}) = G_{t+1}m_t \\ \Sigma &= B_{t+1}Q_{t+1}B'_{t+1} = R_{t+1} - (R_{t+1} + B_{t+1}Q_{t+1}B'_{t+1}) = R_{t+1} - C_{t+1} \end{aligned} \tag{13}$$

As a result we can calculate the transition probabilities using an discrete approximation of (13). Assume that $m_t \in \Pi_i^m$ at time t. We approximate the distribution of $(m_{t+1} \mid m_t)$ by replacing m_t with the center point of Π_i^m, i.e.

$$(m_{t+1} \mid m_t) \sim_A N(G_{t+1}\pi_i^m, R_{t+1} - C_{t+1}) \tag{14}$$

which implies that the transition probabilities are defined as

$$\psi_t(\Pi_j^m \mid \Pi_i^m) = P(m_{t+1} \in \Pi_j^m \mid m_t \in \Pi_i^m) = \int_{\Pi_j^m} p(m)dm \tag{15}$$

where $p(\cdot)$ denote the (multivariate) normal density of (14). Note that by discretizing (14) we obtain a better approximation compared to the approach described in Section 3.2 where approximations were introduced at both parent and child level.

Example 2 (continued) The SSM and HMDP described in the example throughout this paper are taken from Nielsen et al. [27], which implement the methodology described in the present paper. A 3-level HMDP is built with 3011884 states. We will end the running example by describing some of the results of the model.

A stage in the founder process represent the lifetime of a cow and the duration and reward of the stage are defined by a child process at level 1 which is a finite time-horizon MDP with 10 stages. Each stage corresponds to a possible lactation which again is expanded to a child process at level 2. Here the lactation is divided into daily stages where the SSM based on daily yield measurements is embedded as described in Example 2 on page 19. Transition probabilities are calculated using equation (15).

The HMDP is solved maximizing the expected total discounted reward using a yearly interest rate of 2.5%. At a specific time instance the value $m_{t,j}$ together with lactation number and pregnancy status can be used to identify the state in the HMDP and the optimal policy of the HMDP specify which decision to take.

The optimal policy is illustrated in Figure 8 for the three cows presented in Figure 3. Here the x-values are days from calving in lactation 3 and the y-values the *retention payoff*

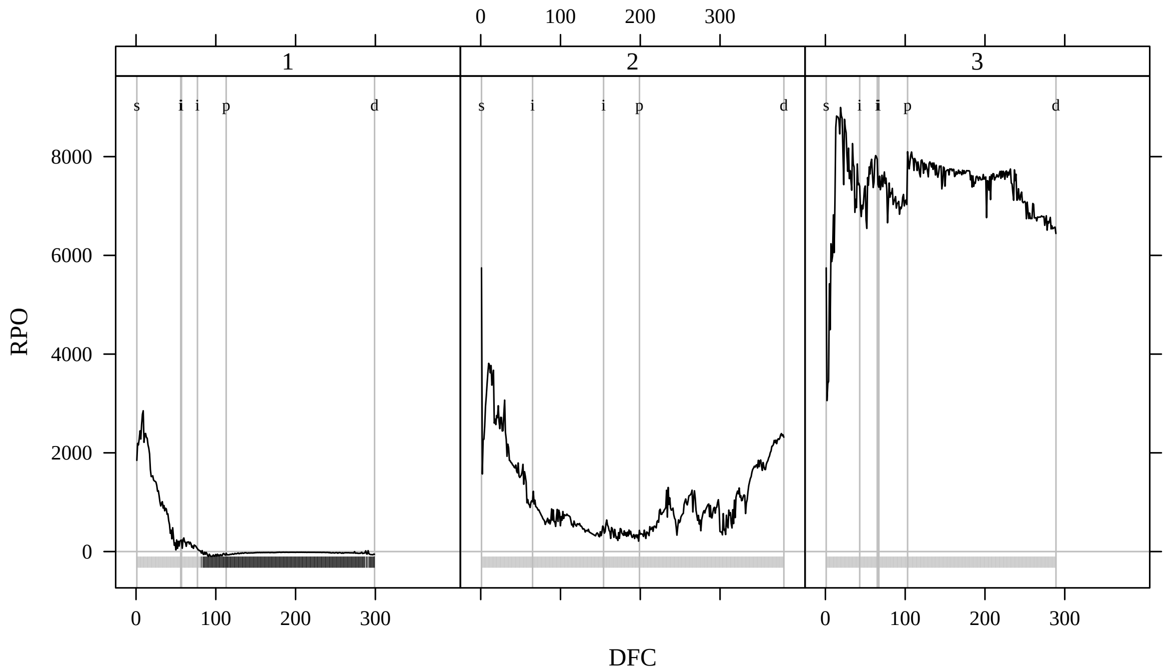


Figure 8: RPO for the cows in Figure 3. The horizontal line in the bottom of each plot indicate whether the optimal decision is to keep or replace (black = replace, gray = keep). Vertical lines correspond to s: start of lactation, i: inseminated, p: positive pregnancy test, d: dry.

(RPO) which denote the difference between the discounted reward under decision `keep` and the discounted reward under decision `replace` for the specific state in the HMDP, i.e. if RPO is positive the optimal decision is to keep the cow and if RPO is negative the optimal decision is to replace the cow. The RPO may be considered as the total extra (discounted) reward of keeping the cow until her optimal replacement time. The horizontal line in the bottom of each plot indicate whether the optimal decision is to keep or replace (black = replace, gray = keep). Vertical lines correspond to s: start of lactation, i: inseminated, p: positive pregnancy test, d: dry.

It can be seen that cow 1 is optimal to replace 82 days from calving while cow 2 and 3 are recommended kept under the whole lactation. For further details see Nielsen et al. [27]. □

6 Discussion and conclusion

The technological possibilities for detailed monitoring of production processes within agriculture give new opportunities for more detailed decision support. State space models for describing these data series is a flexible framework for building such decision support systems that rely on Markov Decision processes.

As illustrated in this paper the transformation to an MDP can be described in general terms, and need not rely on the specific formulation of the SSM. The methods we present embed the SSM into an HMDP. In Section 3 we embedded the SSM using three levels in the HMDP where a stage on the founder level represents the production process, e.g. the lifetime of a cow. However, additional levels in the hierarchy may be a better choice in other contexts. At the next level the states represent information about the variability of the latent variables as estimated from historical observations from the population. In the SSM the latent variables is defined as unobserved. Thus, the latent variable of the SSM cannot be modeled directly in the MDP. However, in the case of SSMs with Gaussian distributions, we can store the distributions implicitly using the conditional mean of the latent variable given previous observations of the process, i.e. the so-called Bayesian updating used in e.g. Kristensen [18].

Note that even though this holds true for the large class of models based on the Gaussian distribution, it is because the covariance matrix is independent of the current state in the MDP and can be calculated based on the time/stage instance. If decisions concerning intervention are modeled in the MDP, the resulting covariance matrix will have to be represented in the states of the MDP too, since the covariance matrix can no longer be calculated based only on the stage/time instance.

The conditional mean of the latent variable in the SSM cannot be embedded directly into the HMDP since it must be discrete. We have followed the standard approach within HMDP and discretized the space of the conditional mean. Increasing the precision of the approximation results in more states in the MDP. If the dimension of the conditional mean is high, this may become intractable to solve. As illustrated in section 3.2, this effect can be reduced using a multivariate discretization and a measure of how well the discretized state space fits the continuous distribution.

However, it may be more important to reduce the states in the HMDP before performing the optimization as pointed out in Section 5. By removing the child process representing the observations we reduce the amount of states for a given time instance dramatically provided that the number of multivariate observation variables are high which is often the case in modern herd management. Here we have many biosensors to predict few traits of the animal. In general, such reductions should be exploited as far as possible.

Discrete random variables which are not a part of the SSM are often represented in the states of the HMDP. For instance, in the dairy cow replacement problem, the calving interval is usually represented as a state variable which is used to measure when the cow will end the lactation. We have omitted these additional state variables from this paper, because in most cases these variables may be considered independent of the variables in the SSM. Hence, the combined transition probabilities can be calculated by multiplying the probabilities from the discrete variables and the variables in the SSM.

The use of SSM for modelling the biological production function may seem unnecessary complicated. Other ways of handling time-series from on-line monitoring in an MDP are possible. For instance a subset of the previous observations could be stored (e.g. in our example the last 5 yield observations or as used by [12] the litter size in previous 2 parities). However, notice that the number of states at each stage becomes the cross product of the intervals used for each of the previous observations which often is larger than if an SSM is used. Moreover, the subset of previous observations only represent an approximation of the previous information provided by the time-series. In some cases the parameters in the time-series may be estimated using observed transition probabilities between observations with discrete levels [28]. However, in the general case, the observations will be much to sparse to include observations of all transition, and such methods can not handle informative censoring of the data, leading to bias in the parameters. The parametric formulation used in the SSM avoids these complications.

Finally, the parameters of the model rely on historical information. However, circumstances may change so much that the historical information is irrelevant, as discussed in section 4.4. Because we rely on a statistical model for the observed data, it should be possible to monitor the validity of the model on a daily base. Again such monitoring could improve the use of decision support system, but how to incorporate it is a topic for future study.

Acknowledgement

The authors would like to thank Professor Anders Ringgaard Kristensen for useful comments on an earlier draft of this paper. This study was funded by a grant financed by the Directorate for Food, Fisheries and Agri Business, Lattec I/S, the Danish Cattle Association and the Faculty of Agricultural Sciences.

References

[1] G. Ausiello, P.G. Franciosa, and D. Frigioni. Directed hypergraphs: Problems, algorithmic results, and a novel decremental approach. In *Theoretical Computer Science : 7th Italian Conference, ICTCS 2001*, volume 2202 of *Lecture Notes in Computer Science*, pages 312–328, Torino, Italy, October 4-6 2001.

[2] Craig Boutilier, Richard Dearden, and Moisés Goldszmidt. Stochastic dynamic programming with factored representations. *Artificial Intelligence*, 121(1-2):49–107, 2000. doi:10.1016/S0004-3702(00)00033-3.

[3] C. Cornou. Automated oestrus detection methods in group housed sows: review of the current methods and perspectives for development. *Livestock Science*, 105:1–11, 2006. doi:10.1016/j.livsci.2006.05.023.

[4] R.M. de Mol, A. Keen, G.H. Kroeze, and J.M.F.H. Achten. Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a Kalman filter. *Computers and Electronics in Agriculture*, 22(2-3):171–185, 1999. doi:10.1016/S0168-1699(99)00016-2.

[5] P. Diggle and M.G. Kenward. Informative dropout in longitudinal data analysis. *Appl. Statist.*, 43:49–93, 1994.

[6] Zhengzhu Feng, Richard Dearden, Nicolas Meuleau, and Richard Washington. Dynamic programming for structured continuous Markov decision problems. In *AUAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 154–161, Arlington, Virginia, United States, 2004.

[7] R. Firk, E. Stamer, W. Junge, and J. Krieter. Automation of oestrus detection in dairy cows: a review. *Livestock Production Science*, 75:219–232, 2002. doi:10.1016/S0301-6226(01)00323-2.

[8] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen. Directed hypergraphs and applications. *Discrete Applied Mathematics*, 42(2-3):177–201, 1993. doi:10.1016/0166-218X(93)90045-P.

[9] C. Guestrin, D. Koller, R. Parr, and S. Venkataraman. Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research*, 19:399–468, 2003.

[10] C. Guestrin, M. Hauskrecht, and B. Kveton. Solving factored MDPs with continuous and discrete variables. In *AUAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 235–242, Arlington, Virginia, United States, 2004.

[11] E.H.P. Houben, R.B.M. Huirne, A.A. Dijkhuizen, and A.R. Kristensen. Optimal replacement of mastitis cows determined by a hierarchic Markov process. *Journal of Dairy Science*, 77:2975–2993, 1994.

[12] R.B.M. Huirne, A.A. Dijkhuizen, and J.A. Renkema. Economic optimization of sow replacement decisions on the personal computer by method of stochastic dynamic programming. *Livestock Production Science*, 28:331–347, 1991. doi:10.1016/0301-6226(91)90014-H.

[13] E. Jørgensen. Sow replacement: Reduction of state space in dynamic programming model and evaluation of benefit from using the model. Dina Research Report 6, National Institute of Animal Science, 1992.

[14] John O. S. Kennedy and Alistair W. Stott. An adaptive decision-making aid for dairy cow replacement. *Agricultural Systems*, 42(1-2):25–39, 1993. doi:10.1016/0308-521X(93)90066-B.

[15] A.V. Kozlov and D. Koller. Nonuniform dynamic discretization in hybrid networks. In *The Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 314–325, August 1-3 1997.

[16] A. R. Kristensen and E. Jørgensen. Multi-level hierarchic Markov processes as a framework for herd management support. *Annals of Operations Research*, 94:69–89, 2000. doi:10.1023/A:1018921201113.

[17] A.R. Kristensen. Hierarchic Markov processes and their applications in replacement models. *European Journal of Operational Research*, 35(2):207–215, 1988. doi:10.1016/0377-2217(88)90031-8.

[18] A.R. Kristensen. Bayesian updating in hierarchical Markov-processes applied to the animal replacement-problem. *European Review of Agricultural Economics*, 20(2):223–239, 1993.

[19] A.R. Kristensen. A general software system for Markov decision processes in herd management applications. *Computers and Electronics in Agriculture*, 38(3):199–215, 2003. doi:10.1016/S0168-1699(02)00183-7.

[20] A.R. Kristensen and T.A. Søllested. A sow replacement model using Bayesian updating in a three-level hierarchic Markov process I. Biological model. *Livestock Production Science*, 87(1):13–24, 2004. doi:10.1016/j.livprodsci.2003.07.004.

[21] A.R. Kristensen and T.A. Søllested. A sow replacement model using Bayesian updating in a three-level hierarchic Markov process, II. Optimization model. *Livestock Production Science*, 87(1):25–36, 2004. doi:10.1016/j.livprodsci.2003.07.005.

[22] G. Lien, A.R. Kristensen, A. Hegrenes, and J.B. Hardaker. Optimal length of leys in an area with winter damage problems. *Grass and Forage Science*, 58(2):168–177, 2003. doi:10.1046/j.1365-2494.2003.00367.x.

[23] T.N. Madsen, S. Andersen, and A.R. Kristensen. Modelling the drinking patterns of young pigs using a state space model. *Computers and Electronics in Agriculture*, 48 (1):39–61, 2005. doi:10.1016/j.compag.2005.01.001.

[24] L.R. Nielsen and A.R. Kristensen. Finding the K best policies in a finite-horizon Markov decision process. *European Journal of Operational Research*, 175(2):1164–1179, 2006. doi:10.1016/j.ejor.2005.06.011.

[25] L.R. Nielsen, K.A. Andersen, and D. Pretolani. Finding the K shortest hyperpaths. *Computers and Operations Research*, 32(6):1477–1497, 2005. doi:10.1016/j.cor.2003.11.014.

[26] L.R. Nielsen, D. Pretolani, and K.A. Andersen. *Multiobjective Programming and Goal Programming*, volume 618 of *Lecture Notes in Economics and Mathematical Systems*, chapter Bicriterion shortest paths in stochastic time-dependent networks, pages 57–67. Springer Berlin Heidelberg, 2009. doi:10.1007/978-3-540-85646-7_6.

[27] L.R. Nielsen, E. Jørgensen, A.R. Kristensen, and S. Østergaard. Optimal replacement policies for dairy cows based on daily yield measurements. *Journal of Dairy Science*, 93(1):75–92, 2010. doi:10.3168/jds.2009-2209.

[28] L.M. Pla, C. Pomar, and J. Pomar. A Markov decision sow model representing the productive lifespan of herd sows. *Agricultural Systems*, 76(1):253–272, April 2003. doi:10.1016/S0308-521X(02)00102-6.

[29] M.L. Puterman. *Markov Decision Processes*. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, 1994.

[30] A.W. Stott, G.M. Jones, R.W. Humphry, and G.J. Gunn. Financial incentive to control paratuberculosis (Johne's disease) on dairy farms in the United Kingdom. *Veterinary Record*, 156(26):825–831, 2005.

[31] I. Thysen. Monitoring bulk tank somatic cell counts by a multi-process Kalman filter. *Acta Agric. Scand. Section A - Animal Science*, 43:58–64, 1993.

[32] N. Toft and E. Jørgensen. Estimation of farm specific parameters in a longitudinal model for litter size with variance components and random dropout. *Livestock Production Science*, 77(2-3):175–185, November 2002. doi:10.1016/S0301-6226(02)00061-1.

[33] J. Van Bebber, N. Reinsch, W. Junge, and E. Kalm. Monitoring daily milk yields with a recursive test day repeatability model (Kalman filter). *Journal of Dairy Science*, 82 (11):2421, 1999.

[34] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer-Verlag, New York, 2 edition, 1997.

[35] C. Yalcin and A.W. Stott. Dynamic programming to investigate financial impacts of mastitis control decisions in milk production systems. *Journal of Dairy Research*, 67 (4):515–528, 2000. doi:10.1017/S0022029900004453.