Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE System-level Adaptive Monitoring and Control of Infrastructures: A POMDP-based Framework

PRESENTED BY Milad Memarzadeh

ACCEPTED BY THE DEPARTMENTS OF

Civil and Environmental Engineering

Matteo Pozzi — ADVISOR, MAJOR PROFESSOR — December 15, 2015 — DATE

David A. Dzombak — DEPARTMENT HEAD — December 28, 2015 — DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula — DEAN — January 4, 2016 — DATE

System-level Adaptive Monitoring and Control of Infrastructures: A POMDP-based Framework

Submitted in partial fulfillment of the requirements of

the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

Milad Memarzadeh

B.Sc., Civil Engineering, University of Tehran
M.Sc., Civil Engineering, Virginia Tech

Carnegie Mellon University
Pittsburgh, PA

December 2015

Acknowledgements

My time at Carnegie Mellon has been one of the most rewarding and inspiring time of my life, both professionally and personally. It helped me a lot in shaping my career and finding the right direction for my future research.

First of all, I would like to express my deepest gratitude to my advisor, Prof. Matteo Pozzi for his infallible guidance, immense knowledge and infinite patience. This dissertation would not have been possible without his vast experience and encouraging nature. I admire his insatiable appetite for learning and I would be honored to be an academic like him some day.

I would also like to thank Prof. J. Zico Kolter who co-advised me on projects during the early stages of my research. His unique point of view as a computer scientist helped me a lot to learn the advanced theoretical methods needed in my research and gave a new perspective to my research.

I would like to extend my gratitude to my committee members, Prof. Samer Madanat, Prof. Mitch Small, and Prof. Emma Brunskill for their insightful comments and the encouragement they have given me. I am grateful to Prof. Hae Young Noh and Prof. Chris Hendrickson for taking the time to serve as my PhD qualification committee. I would also like to thank Prof. Mario Berges not only for guiding me through graduate studies at Carnegie Mellon, but also for our great music performance in the 2015 CEE talent show.

I would like to gratefully acknowledge the funding source for supporting my PhD studies. This dissertation work is partially supported by the Dean’s Fellowship from Carnegie Institute of

Technology and the Pennsylvania Infrastructure Technology Alliance, a partnership of Carnegie Mellon, Lehigh University, and the Commonwealth of Pennsylvania's Department of Community and Economic Development (DCED), under the grant PITA YR1631571.1.9.1042204.

The Department of Civil and Environmental Engineering is fortunate to have a wonderful administrative staff in the form of Maxine Leffard, Julian Krishnamurti, Mireille Mobley, Melissa Brown, Cornelia Moore, Andrea Rooney, Jodi Russo, Daniel Joyner, and Cathy Schaefer. I am grateful for their help and assistance.

I would like to thank my other group members, Carl Malings, Pengyun Wang, Irem Velibeyoglu, and Yasamin Hashemi Tari for their valuable discussions and support during my time at Carnegie Mellon.

I am lucky to have many wonderful friends who have always taken great pride in my achievements and I am grateful for their support, especially, Prof. Mehran Tehrani, Dr. Reza Azimi, Salim Malakouti, Navid Kazem, Mohammad Ahmadpoor, and Dr. Arka Roy. We have been there for each other, shared our worries and soothed away anxieties. Pittsburgh wouldn't have seemed half as fun without them.

I would like to thank all my friends from Pittsburgh, especially my friends in Persian Student Organization, and my teammates at Civil FC which without them two CMU's intramural soccer championship would not have been possible.

Last but not least, I am tremendously grateful to my family – mom, dad and my brothers Mohsen and Meisam – for their sacrifice, love and support. I am who I am today because of them.

Abstract

Many infrastructure systems in the US such as road networks, bridges, water and wastewater pipelines, and wind farms are aging and their condition are deteriorating. Accurate risk analysis is crucial to extend the life span of these systems, and to guide decision making towards a sustainable use of resources. These systems are subjected to fatigue-induced degradation and need periodic inspections and repairs, which are usually performed through semi-annual, annual, or bi-annual scheduled maintenance. However, better maintenance can be achieved by flexible policies based on prior knowledge of the degradation process and on data collected in the field by sensors and visual inspections.

Traditional methods to model the operation and maintenance (O&M) process, such as Markov decision processes (MDP) and partially observable MDP (POMDP) have limitations that do not allow the model to properly include the knowledge available and that may result in non-optimal strategies for management of infrastructure systems. Specifically, the conditional probabilities for modeling the degradation process and the precision of the observations are usually affected by epistemic uncertainty: this cannot be captured by traditional methods.

The goal of this dissertation is to propose a computational framework for adaptive monitoring and control of infrastructures at the system-level and to connect different aspects of the management process together. The first research question we address is how to take optimal sequential decisions under model uncertainty. Second, we propose how to combine decision optimization with learning of the degradation of components and the precision of monitoring system. Specifically, we address the issue of systems made by similar components, where

transfer of knowledge across components is relevant. Finally, we propose how to assess the value of information in sequential decision making and whether it can be used as a heuristic for system-level inspection scheduling.

In this dissertation, first a novel learning and planning method is proposed, called “Planning and Learning for Uncertain dynamic Systems” (PLUS), that can learn from the environment, update the distributions of parameters, and select the optimal strategy considering the uncertainty related to the model. Validating with synthetic data, the total management cost of operating a wind farm using PLUS is shown to be significantly less than costs achieved by a fixed policy or though the POMDP framework.

Moreover, when the system is made up by similar components, data collected on one is also relevant in the management of others. This is typically the case of wind farms, which are made up by similar turbines. PLUS models the components as independent or identical and eithers learn the model for each component independently or learn a global model for all components. We extend that formulation, allowing for a weaker similarity among components. The proposed approach, called “Multiple Uncertain POMDP” (MU-POMDP), models the components as POMDPs, and assumes the corresponding model parameters as dependent random variables. By using this framework, we can calibrate specific degradation and emission models for each component while, at the same time, processing observations at the level of the entire system. We evaluate the performance of MU-POMDP compared to PLUS and discuss its potentials and computational complexity.

Lastly, operation and maintenance of an infrastructure system rely on information collected on its components, which can provide the decision maker with an accurate assessment of their condition states. However, resources to be invested in data gathering are usually limited and

observations should be collected based on their value of information (VoI). VoI is a key concept for directing explorative actions, and in the context of infrastructure operation and maintenance, it has application to decisions about inspecting and monitoring the condition states of the components. Assessing the VoI is computationally intractable for most applications involving sequential decisions, such as long-term infrastructure maintenance. The component-level VoI can be used as a heuristic for assigning priorities to system-level inspection scheduling. In this research, we propose two alternative models for integrating adaptive maintenance planning based on POMDP and inspection scheduling based on a tractable approximation of VoI: the stochastic allocation model (and its two limiting scenarios called pessimistic and optimistic) that assumes observations are collected with a given probability, and the fee-based allocation model that assumes observations are available at a given cost. We illustrate how these models can be used at component-level and for system-level inspection scheduling. Furthermore, we evaluate the quality of solution provided by pessimistic and optimistic approaches. Finally, we introduce analytical formulas based on the stochastic and fee-based allocation models to predict the impact of a monitoring system (or a piece of information) on the operation and maintenance cost of infrastructure systems.

Contents

List of Figures

List of Tables

List of Variables

s	Condition state	S^t	Sequence of states until time t
S	Set of condition states	A	Set of actions
a	Actions available to the agent	$N_{ss'}^{a}$	Number of (s, a, s') transitions
r	Reward (cost) function	N_{sz}^{a}	Number of (s, a, z) emissions
$\mathbf{T}$	Transition probability	Q	Q-value function
t	Management time steps	n_b	Number of samples in the burn-in phase of MH algorithm
γ	Discount factor	$\alpha_T, \boldsymbol{\beta}_T$	Hyper-parameters of transitions
$\bar{x}_t$	History of observed variables x until time step t	$\alpha_O, \boldsymbol{\beta}_O$	Hyper-parameters of emissions
π	Policy	$\lambda_T, \boldsymbol{\eta}_T$	Prior of hyper-parameters of transitions
V	Value	$\lambda_O, \boldsymbol{\eta}_O$	Prior of hyper-parameters of emissions
$\boldsymbol{\Theta}$	Model parameters	σ_a	Step size for proposal distribution in direction of α_T
Z, z	Observations	c_β	Concentration parameter for proposal distribution in direction of Step size for proposal distribution in direction of β_T
$\mathbf{O}$	Emission probability	$H: h$	Inspections
$\mathbf{b}$	Belief state	$\mathbf{G}$	Augmented emission probability
f	Move-forward operator	e^{I}	Emission operator of inspections
e	Emission operator	u^{I}	Updating operator of inspections

u	Updating operator	ΔV	Difference in the value
α	α-vectors representing conditional plans	VoI	Value of information
n_c	Number of conditional plans	C	Cost of inspections
$\tilde{\mathbf{b}}$	Augmented belief state	$\mathbf{E}$	Emission of additional observations
IM	Importance measure	W	System-level value
U	Value function estimation	$SVoI$	System-level value of information
$C_{M(t)}$	Minimum VoI sufficient for winning the inspection at time t	$\bar{C}_{M(t)}$	Average VoI sufficient for winning the inspection during all steps till t

Chapter 1

Introduction and Literature Review

Abstract

In this chapter, we illustrate the motivation of the research presented in this dissertation. Next, we review the literature on sequential decision making including MDPs and POMDPs, inspection scheduling, and application of them to the management of infrastructure systems. The limitations of the previous studies are also discussed. Next, we discuss the overall framework proposed in this research, and the connection among its different parts. Lastly, we list the journal and conference papers that we have submitted and published, presenting the outcome of this dissertation.

1.1 Motivation

Many infrastructure systems in the U.S. such as road networks, bridges, water and wastewater pipelines are aging and their conditions are deteriorating (ASCE 2013). Accurate risk analysis is crucial to extend the life span of these systems, and to guide decision making towards a sustainable use of resources. The degradation process of these systems can be modeled in a probabilistic framework, incorporating the effect of maintenance policy.

One of the infrastructure systems crucial for sustainability is wind farms that are playing an ever-increasing role worldwide as a renewable energy source; as a result, there will be an increasing demand for careful management of costs associated with operation and maintenance (O&M) of wind turbines. This cost on average account for approximately 25-30% of the overall energy generation costs (Marquez et al. 2012). Farms are made up by turbines of the same "typology". Their conditions degrade because of aging, fatigue load, and exposure to environmental risks. On land, there have been incidents that have showcased the risk of structural failure of wind turbines. For example, in 2009, a wind turbine in Fenner wind farm (located in NY) unexpectedly collapsed as shown in Figure 1. The cause of this structural failure was fatigue on the mast foundation. Although this did not prove to be a catastrophic event, understanding how such failures occur and how it can be avoided is both an academic and practical endeavor.

Figure 1. Collapsed wind turbine at the Fenner wind farm [Photo credit: Kevin Wigell, Everpower Wind Holdings.]

Managing a wind farm (or generally an infrastructure system) includes selecting appropriate O&M levels for the turbines (i.e. components of the system), scheduling inspections, and performing maintenance actions. A rational manager has to find a reasonable tradeoff between exploitation and exploration. Exploitation refers to the conservative maintenance policy to minimize the cost of O&M while exploration refers to learning the degradation behavior of components, the effectiveness of the maintenance actions, and the precision of the monitoring system. Thus, a robust decision making framework is needed to automatically evaluate the uncertainties related to the environment. In this context, the overall goal is to find an optimal policy that minimizes the total expected costs of the system over the management time horizon, making use of probabilistic models for predicting the degradation of the system and the effectiveness of maintenance actions.

In order to develop such framework, we have collaborated with our industry partner, Everpower Wind Holdings, to conduct the research proposed in this dissertation. They provide us with the prior knowledge on the failure of the turbine components (e.g. gearbox and yaw system) that is basis for developing the numerical examples used for validation in this research. One example of the farms they operate is Highland wind farm located in Cambria County, PA. The farm is made up by 25 Nordex N-90 turbine generators, each with a power capacity of 2.5 MW, and it is operative since August 2009. It is a very interesting application for our project, as all machines belong to the same model and have the same age. In total, Highland wind farm has the capacity to generate approximately 62.5 MW (producing enough electricity to power over 15,000 households). The turbines are instrumented with wind sensors, power sensors, accelerometers, and are periodically inspected for measuring bolt torque and symptoms of corrosion and fatigue damage.

Figure 2. Our visit to Highland Wind farm, located in Cambria County, PA.

1.2 Literature Review

1.2.1 Literature on Management of Infrastructure Systems

A fundamental framework for sequential decision making is the Markov Decision Processes (MDP). Textbooks of Sutton and Barto (1998) and Bertsekas (1996) provide comprehensive introduction of sequential decision making, optimal control and MDP. In an MDP, the environment is modeled as a finite set of states and actions that a decision maker (that from now on we refer to as "agent") can take. The goal is to choose actions that maximize the total expected reward (or minimize the total expected case in application to O&M of infrastructure systems). One of the main limitations of MDP is that it assumes that the state of the system is fully observable, which is not true in most real-world applications. Details of MDP's formulation are provided in Chapter 2.

MDP has been extensively applied to operation and maintenance of infrastructure components (Golabi et al. 1982, Guignier and Madanat 1999, Robelin and Madanat 2007), due to the computational efficiency of dynamic programming. However, an MDP assumes perfect information on the system state at any step of the decision process and, because of this, is not

suitable for investigation the impact of information gathering. Madanat (1993) proposed a methodology for optimal inspection and maintenance policies for infrastructure networks called latent MDP (LMDP) that allows the partial observability of condition state of infrastructure components. Moving beyond analysis of single component, Smilowitz and Madanat (2000) incorporated the network-level budget and condition state constrains in LMDP. Guillaumot et al. (2003) proposed an adaptive optimization method for infrastructure maintenance and inspection decisions based on LMDPs under model uncertainty. Medury and Madanat (2013a, b) have extended the state-of-the-art MDP-based methodologies in infrastructure management to integrate the two aspects of the decision making process: the financial allocation of resources for maintenance, rehabilitation and replacement policies and the operational-level implementation. In particular, they use approximate dynamic programming (Powell 2007) to model complex problems in infrastructure management.

To address the limitation of MDP, Partially Observable MDP (POMDP) generalizes MDP, where the exact state of the system cannot be observed directly but can be inferred by indirect and imperfect observations (Smallwood and Sondik 1973, Sondik 1978). Details regard POMDP formulation and implementations are provided in Chapter 2.

Extensive literature on planning inspection and maintenance for civil structures using dynamic programming and Markov processes has been reviewed by Papakonstantinou and Shinozuka (2014a). Papakonstaintou and Shinozuka (2014b, c) implemented the POMDP framework for inspection and maintenance planning of corroding reinforced concrete structure. Their method suggests inspection/monitoring and maintenance actions. Availability of different monitoring and maintenance actions, uncertain observation and action outcomes and the cost-benefit of the information are also incorporated in their formulation. Schobi and Chatzi (2015)

have used continuous POMDP for life cycle assessment and maintenance planning of infrastructure components.

1.2.2 Literature on Management of Wind Farms

In the literature, methods based on POMDP have been recently proposed for optimal management of wind farms. These methods use historical data to fix the model parameters (i.e. transition probability, describing the degradation of the system, and emission probability, describing precision of the monitoring system) and find the optimal policy based on them. Byon et al. (2010) have proposed an optimal maintenance strategy for wind turbine systems under stochastic weather conditions. They have formulated the degradation process of turbines as a POMDP, with the objective of deriving an optimal preventive maintenance policy that minimizes the expected average cost over an infinite horizon. Also, these authors have extended their proposed method to season-dependent condition-based maintenance of wind turbines to include the dynamic weather conditions, which makes the subsequent modeling of the resulting strategy season-dependent (Byon and Ding 2010). Nielsen and Sorensen (2012) have presented the use of limited information influence diagram and POMDP to assist in rational decision making for O&M of offshore turbines. McMillan and Ault (2008) have used Monte Carlo simulations to evaluate the cost effectiveness of condition based monitoring of wind turbines. Specifically they have found the effect of MDP in modeling the wind turbine deterioration and failure characteristics.

A key limit in these studies is that transition and emission probabilities (i.e. model parameters) are assumed as fixed parameters, and epistemic (and model) uncertainty is not taken into account. In those studies that have included the model uncertainty, the correct model is

being chosen among few pre-specified candidates, instead of assigning a general prior model that describes the behavior of components' parameters. Furthermore, the difference and similarities among the model parameters of different components on the system are not modeled.

1.2.3 Literature on Sequential Decision Making under Model Uncertainty

Researchers have incorporated uncertainty in the transition probabilities of the MDP framework directly in the formulations to find policies that are both optimal in terms of maximizing the total expected reward and robust to errors in the model parameters. Bagnell et al. (2001) have proposed a stochastic dynamic game to solve the problem of MDPs with uncertain transition probabilities. The proposed solution is equilibrium of the game that corresponds to that value function under the worst model. Li and Si (2007) have proposed a new optimality criterion that is a basis for development of robust policy iteration to solve this problem. Nilim and ghaoui (2005) have solved the uncertain MDP problem in the context of finite and infinite horizon using robust value iteration.

The Bayes-Adaptive POMDP (BA-POMDP) framework is a generalization of POMDP, where the transition and emission probabilities are unknown and are treated as random variables, with a prior distribution, whose distribution can be learned and updated during the process of monitoring and management (Ross et al. 2011). Details of BA-POMDP's formulation are provided in Chapter 2. Jaulmes et al. (2005a, b) have proposed an algorithm called Markovian exploration with decision based on the use of sampled model algorithm (MEDUSA) to find the optimal policy for a POMDP when the model is no known or poorly specified. Their algorithm tries to improve the POMDP incrementally using selected queries, while still optimizing the total expected reward.

1.2.4. Literature on Value of Information in Management of Infrastructure Systems

In the maintenance process, information collected by inspectors and monitoring system can provide the agent with accurate assessments and prognoses of components' condition states, which can be integrated in a probabilistic framework to model the effects of degradation and of the adopted maintenance policy. Information can reduce the uncertainty in the decision making process however, it is usually expensive to collect due to limited resources. Therefore, data collection needs to be prioritized, trading off the cost of gathering information against the potential benefits tis information might have in terms of selecting more appropriate maintenance actions. Pre-posterior analysis allows for predicting the impact of each available observation to the maintenance process, so that it can be the base for rational sequential information gathering. This is a relevant topic in a wide variety of applications from sensor scheduling (Ji et al. 2007, Shi et al. 2011, Mo et al. 2012a, Mo et al. 2012b) to scheduling for human or robot inspectors. Reference applications of the latter topic to civil infrastructure systems are provided by Straub and Faber (2004, 2005, and 2006), which proposed the so-called equidistant and threshold approaches for reliability-based inspection scheduling: a former finds an optimal inter-inspection period, whereas the latter schedules inspections when the probability of failure exceeds a threshold. Although inspection scheduling for a single component can be incorporated in the POMDP framework (Memarzadeh et al. 2015a, Papakonstantinou and Shinozuka 2014b), the system-level scheduling poses computational challenges. The concept of Value of Information (VoI) (Raiifa and Schlaifer 1961) is key to pre-posterior analysis, and can be taken as a consistent approach for ranking al available observations: VoI of an inspection is defined as the difference between the value of the management process with and without that specific observation. Introduction and application of VoI analysis to civil infrastructure systems is

provided by Pozzi and der Kiureghian (2011), Straub (2014), and Zonta et al. (2014). Application to long-term maintenance planning is shown by Straub and Faber (2006) and Konakli et al. (2015).

As mentioned, the system-level scheduling poses computational challenges, when dealing with constraints in the available resources for information gathering. It is challenging to integrate optimization of information gathering of large systems in a dynamic controlled setting where the agent is optimizing the maintenance policy as well.

In this research, we propose a framework and computational approaches that integrates learning, planning, and inspection scheduling at system-level for optimal management of infrastructure systems and addresses gaps in knowledge in the literature.

1.3 Proposed Framework and Layout of the Dissertation

Figure 3 shows three main contributions of this research and connections among them.

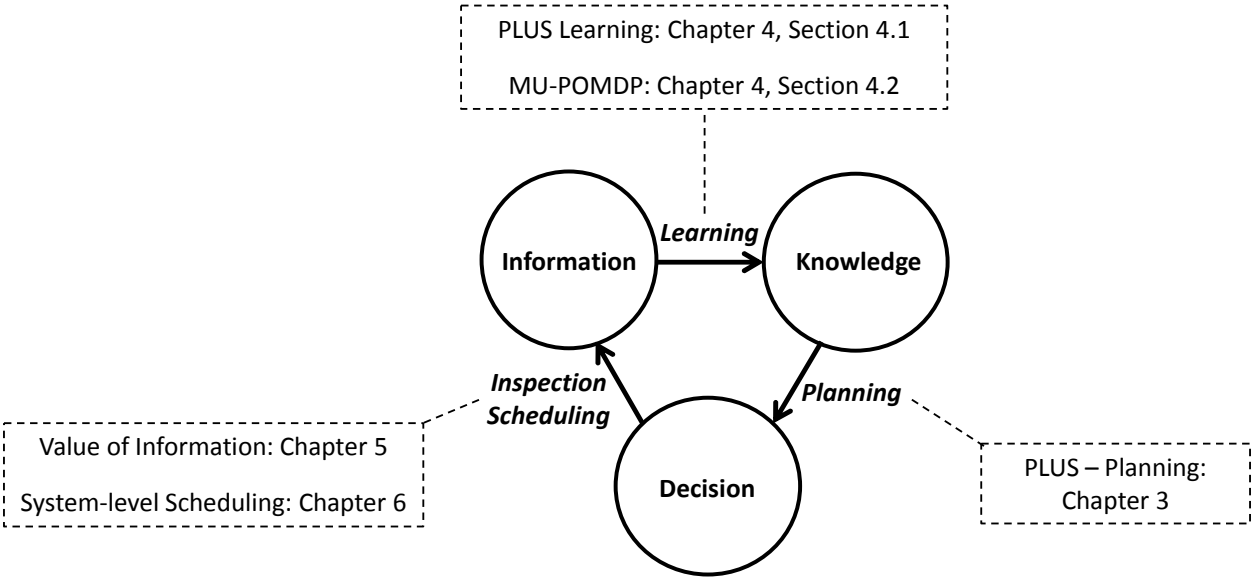

Figure 3. The proposed framework connects three important aspect of sequential decision making: learning, planning, and data collection scheduling.

First, we introduce the MDP and POMDP frameworks in full detail in Chapter 2, as we use POMDP as a baseline in the following chapters for developing the contributions of this research.

The link between *knowledge* to *decision* corresponds to the decision making under model uncertainty (i.e. planning) addressed in Chapter 3. The research question corresponds to: given the learn model parameters, how the agent can take an action for maintenance of the system considering the future consequences of her action and the uncertainty caused by dynamic environment?

The link from the *information* to *knowledge* corresponds to learning the condition state of the components as well as their degradation process (which we generally call "learning") addressed in Chapter 4. The research question is the following: given the information collected from components on the system, how an agent can learn the their degradation behavior, effectiveness of maintenance actions, and precision of the monitoring system?

Finally, the link between *decision* and *information* corresponds to the specific research question regard value of information and inspection scheduling addressed in Chapters 5 and 6. The research question is: how to predict the impact of a monitoring system as a pre-posterior analysis? Moreover, if there is restriction in availability of resources for information gathering, how an agent can prioritize this task and identify the critical components on the system at each time during the management process?

1.4 Publications Derived from this Dissertation

The first part of this dissertation was initially published in proceedings of 9th International Workshop on Structural Health Monitoring (IWSHM) (Memarzadeh et al. 2013). That paper

presents an approximate algorithm for planning and learning within the BA-POMDP framework. Later, we generalized the algorithm and proposed Planning and Learning for Uncertain dynamic Systems (PLUS) which is published in the ASCE Journal of Computing in Civil Engineering (Memarzadeh et al. 2015a). These publications are the bases for Chapter 3 and Chapter 4, section 4.1.

The second part of the research focuses on modeling systems with similar components. The preliminary results are published in the proceedings of 6th World Conference on Structural Control and Monitoring (Memarzadeh et al. 2014) and International Conference on Applications of Statistics and Probability in Civil Engineering (Memarzadeh et al. 2015f). The proposed Multiple Uncertain POMDP (MU-POMDP) framework with comprehensive validation and application on wind farm management is currently under review in Elsevier Journal of Reliability Engineering and System Safety (Memarzadeh et al. 2015b). The content of these publications is presented in Chapter 4, section 4.2.

The third part of the research is related to the computation of value of information in sequential decision making and its application for system-level inspection scheduling. We first introduced two heuristics in the Journal of Computer-Aided Civil and Infrastructure Engineering (Memarzadeh and Pozzi 2015c). Later on, we extended these heuristics to the proposed stochastic future allocation and fee-based future allocation models; one conference paper is published in proceedings of 10th IWSHM (Memarzadeh and Pozzi 2015e) and a journal paper is finalized to be submitted to Elsevier Journal of Reliability Engineering and System Safety (Memarzadeh et al. 2015d). Content of these publications are reported in Chapters 5 and 6.

Chapter 2

Markov Decision Processes: Full and Partial Observability

Abstract

In this chapter, we introduce the traditional methods for sequential decision making. The Markov decision process (MDP) is presented in detail and then the partially observable MDP (POMDP) is introduced as a generalization of MDP. We discuss details of the formulations, how to solve the optimization problem, and present an illustrative example of POMDP. In the last part, we discuss the recent advancement of POMDP to include the model uncertainty and the corresponding framework of Bayes-Adaptive POMDP (BA-POMDP).

2.1 Markov Decision Process

A fundamental model for sequential decision making is the Markov decision process (MDP). In an MDP, the environment is modeled as a finite set of states and actions that an agent can take. The goal is to choose actions that maximize the total expected reward. A typical graphical model of MDP is shown in Figure 4. Graphical models in this document follow the notation of dynamic Bayesian network and influence diagrams adopted in the textbook of Barber (2012). Circles define random variables, squares decision variables, diamonds utility variables, and arrows dependence among variables.

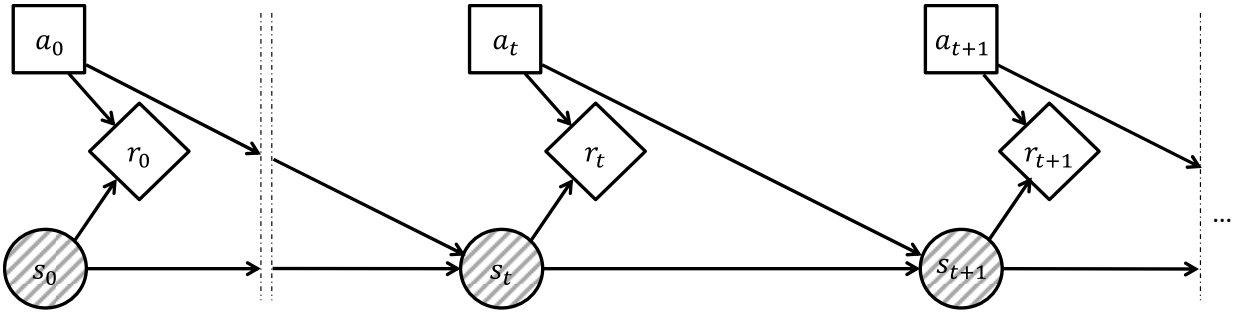

Figure 4. Graphical model of a Markov decision process

An MDP is defined by a 5-tuple $(S, A, \mathbf{T}, \mathbf{R}, \gamma)$, where $S = \{1,2, \dots, |S|\}$, $A = \{1,2, \dots, |A|\}$ are finite sets of condition states and available actions. Transition is described by a 3-dimensioona matrixes of size $|S| \times |S| \times |A|$, whose entry are defined as $T(i,j,k) = \mathbb{P}[s_{t+1} = j \mid s_t = i, a_t = k]$. In the MDP, Markov property holds: given the current state of the system and the action that an agent has taken, future states are independent of the past, so that $\mathbb{P}[s_{t+1} = j \mid \bar{a}_t, \bar{s}_t] = \mathbb{P}[s_{t+1} = j \mid s_t, a_t]$, where $\bar{s}_t = \{s_0, s_1, \dots, s_t\}$ and $\bar{a}_t = \{a_0, a_1, \dots, a_t\}$ indicate the history of states and actions, respectively. Reward (cost) matrix, of size $|S| \times |A|$, is defined as $R(i,k) = \mathbb{E}[r_t \mid s_t = i, a_t = k]$. Traditionally, letter r indicates a "reward", but in the context of infrastructure management, it refers to cost. Finally, future rewards are made equivalent to

current ones by using discount factor, γ. In the following, we summarize the parameters of a POMDP as follows: $\boldsymbol{\Theta} = \{\mathbf{T}, \mathbf{R}, \gamma\}$, since the dimension of the matrices carry information of those of relevant sets.

In the MDP, the agent starts in an initial state, s_0. At any time step t, the agent observes the current state of the system, s_t, takes an action a_t, receives a reward $R(s_t, a_t)$ (or pay the cost), and moves to the next state s_{t+1} with probability $T(s_t, a_t, s_{t+1})$. A policy, $\pi: S \rightarrow A$ is a mapping from state space to actions. The value of a policy is the corresponding expected sum of discounted costs (or rewards) when starting in some state and executing actions according to the policy. The optimal policy π^* is that achieving the minimum value (maximum value, when dealing with rewards). The optimal value for infinite time horizon is stationary and can be described by Bellman's equation (Bellman 1957):

$$V^*(s, \boldsymbol{\Theta}) = \min_{a \in A} \left\{ R(s, a) + \gamma \sum\nolimits_{s' \in S} T(s, a, s') V^*(s', \boldsymbol{\Theta}) \right\} \quad (1)$$

And the optimal policy is:

$$\pi^*(s, \boldsymbol{\Theta}) = \operatorname*{argmin}_{a \in A} \left\{ R(s, a) + \gamma \sum\nolimits_{s' \in S} T(s, a, s') V^*(s', \boldsymbol{\Theta}) \right\} \quad (2)$$

Note that if the goal of the agent is to maximize the rewards, the optimization problem needs to change to maximization over actions.

Optimal policy for MDP can be identified by two classical methods: *value iteration* and *policy iteration.* The details of these algorithms can be found in textbooks of Sutton and Barto (1998) and Russell and Norvig (2010).

One of the main limitation of MDP is that it assumes that the state of the system is fully observable, which is not a true assumption in many real-world applications.

2.2 Partially Observable Markov Decision Process

The POMDP framework shares many assumptions of MDP. At any time, the system's state s assumes one value in finite discrete set $S = \{1,2, \dots, |S|\}$, while the agent can select one action a among set $A = \{1,2, \dots, |A|\}$. Based on the current state and action, she pays cost r. Time is discretized in steps, and variable s_t, a_t, r_t indicate state, action and cost at time t respectively. Expected cost is assigned by function $R(i,k) = \mathbb{E}[r_t | s_t = i, a_t = k]$. After taking an action, the state evolves stochastically following a Markov process governed by transition probability function $T(s,a,s') = \mathbb{P}[s_{t+1} = s | s_t = s, a_t = a]$.

In MDPs, action a_t follows the observation of the full state s_t that, given the Markovian assumption, is a sufficient statistic for the process. On the contrary, POMDPs assume that at time t the agent has access only to a noisy and incomplete measure of the current state, summarized by observation z_t which can assume one value in set $Z = \{1,2, \dots, |Z|\}$. The relation between state and observation is capture by the emission probability function $O(s,a,z) = \mathbb{P}[z_t = z | s_t = s, a_{t-1} = a]$. The entire cost, transition and emission functions are listed in corresponding matrixes $\mathbf{T}, \mathbf{O}, \mathbf{R}$, of size $|S| \times |S| \times |A|$, $|S| \times |Z| \times |A|$ and $|S| \times |A|$ respectively. In summary, transition matrix $\mathbf{T}$ defines the degradation model and the effectiveness of maintenance actions, emission matrix $\mathbf{O}$ defines the accuracy of observations collected by instrumented and visual inspections, while cost matrix $\mathbf{R}$ defines the economic model.

Figure 5 shows a graphical model of a POMDP, using the classical notation of dynamic Bayesian networks and influence diagrams (Barber 2012). Only shaded variables are observed. Figure 5 allows us to follow in details the temporal process. At time t_0, the agent takes action a_0 and pay cost r_0; then time Δt passes and state evolves to s_1, that the agent observes imperfectly through z_1. Cost, new state and observation depend on the taken action. Action a_1 is selected after having analysed z_1, and the process is iterated indefinitely.

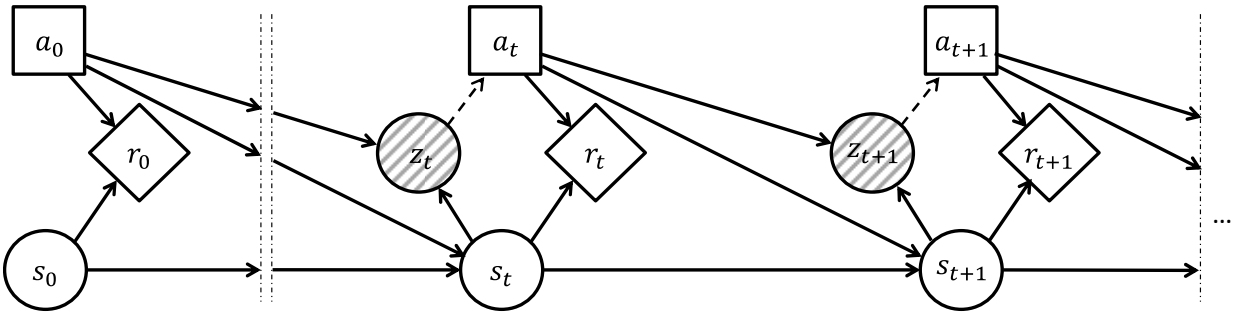

Figure 5. Graphical model of a partially observable Markov decision process.

The agent's goal is to minimize value V, defined as the expected sum of the discounted costs over an infinite time horizon: $V = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t]$, using discount factor γ. At time t, the agent's knowledge about the current state is represented by a probability distribution, or belief vector $\mathbf{b}_t$, so that the ith enty is $b_t(i) = \mathbb{P}[s_t = i|\bar{z}_t, \bar{a}_t]$, with sets $\bar{a}_t = \{a_0, \dots, a_{t-1}\}$ and $\bar{z}_t = \{z_1, \dots, z_t\}$ being the history of observations and actions up to time t, respectively. Being the belief a sufficient statistics for the process, the agent can base her decisions on that. Formally, a POMDP is defined by a 8-tuple $(S, Z, A, \mathbf{T}, \mathbf{O}, \mathbf{R}, \mathbf{b}_0, \gamma)$, where $\mathbf{b}_0$ is the initial belief. In the following, we summarize its parameters in set $\mathbf{\Theta} = \{\mathbf{T}, \mathbf{O}, \mathbf{R}, \gamma\}$, since the dimension of matrixes carry information of those of the sets S, Z and A. During the process, the agent updates her belief by iteratively processing any available observation. Transition and emission probabilities can be combined in operators that allows for predicting the state evolution and processing observations,

making use of Bayes' rule. The move-forward (f), emission (e), and updating (u) operators, or dimension $|S|$, $|Z|$ and $|S|$ respectively, are defined entry-by-entry as follows:

$$\begin{cases} f_i(\mathbf{b},k,\boldsymbol{\Theta}) = \mathbb{P}[s_{t+1}=i|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta}] & = \sum_{l=1}^{|S|} T(l,k,i)b(l) \\ e_j(\mathbf{b},k,\boldsymbol{\Theta}) = \mathbb{P}[z_{t+1}=j|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta}] & = \sum_{i=1}^{|S|} O(i,k,j)f_i(\mathbf{b},k,\boldsymbol{\Theta}) \\ u_i(\mathbf{b},k,j,\boldsymbol{\Theta}) = \mathbb{P}[s_{t+1}=i|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta},z_{t+1}=j] = \dfrac{O(i,k,j)f_i(\mathbf{b},k,\boldsymbol{\Theta})}{e_j(\mathbf{b},k,\boldsymbol{\Theta})} \end{cases} \quad (3)$$

In summary, if the agent has belief $\mathbf{b}$ at time t, takes action k and observes j at the next step, then the updated belief is $u(\mathbf{b},k,j,\boldsymbol{\Theta})$. We re-use r for indicating expected immediate cost as a function of belief $\mathbf{b}$ and action a, as $r(\mathbf{b},a,\boldsymbol{\Theta}) = \sum_{s=1}^{|S|} b(s)R(s,a)$.

The agent's behavior is defined by a policy, i.e. a map between belief and actions. When policy π is adopted, action at time $(t+1)$ is set as $a_{t+1} = \pi(\mathbf{b}_t)$. The value depends on policy π via the recursive equation:

$$V^{\pi}(\mathbf{b},\boldsymbol{\Theta}) = r(\mathbf{b},\pi(\mathbf{b}),\boldsymbol{\Theta}) + \gamma \sum_{z=1}^{|Z|} e_z(\mathbf{b},\pi(\mathbf{b}),\boldsymbol{\Theta})\, V^{\pi}[u(\mathbf{b},\pi(\mathbf{b}),z,\boldsymbol{\Theta}),\boldsymbol{\Theta}] \quad (4)$$

while the optimal value is defined by the Bellman Equation (Bellman 1957) as in Eq. (1):

$$V^{*}(\mathbf{b},\boldsymbol{\Theta}) = \min_{a\in A}\left\{ r(\mathbf{b},a,\boldsymbol{\Theta}) + \gamma \sum_{z=1}^{|Z|} e_z(\mathbf{b},a,\boldsymbol{\Theta})\, V^{*}[u(\mathbf{b},a,z,\boldsymbol{\Theta}),\boldsymbol{\Theta}] \right\} \quad (5)$$

Note that if the goal of the agent is to maximize the rewards, the optimization problem needs to change to maximization over actions.

Bellman's equation for optimal policy π^* can be formulated as in Eq. (2):

$$\pi^*(\mathbf{b}, \boldsymbol{\Theta}) = \underset{a \in A}{\operatorname{argmin}} \left\{ r(\mathbf{b}, a, \boldsymbol{\Theta}) + \gamma \sum_{z=1}^{|Z|} e_z(\mathbf{b}, a, \boldsymbol{\Theta})\, V^*[u(\mathbf{b}, a, z, \boldsymbol{\Theta}), \boldsymbol{\Theta}] \right\} \quad (6)$$

In principle, a POMDP is solved by applying the methods to solve MDPs to the belief state (Aoki 1965, Astrom 1965). However, as the belief state is a probability distribution, it is defined on an infinite space, and so exact solution for the POMDP is not generally available. In reacting to observations collected, an agent can select one conditional plan among the many available (Russell and Norvig 2010). The conditional plan can be interpreted as a policy function defined on the domain of the sequence of observations.

The number of possible conditional plans, n_c, grows exponentially with the time horizon assumed for the project. Let $\alpha_{i,\boldsymbol{\Theta}}(s)$ defined the value of executing the i-th conditional plan starting from perfect knowledge that the system is in state s for the POMDP model defined by $\boldsymbol{\Theta}$. The value of following that plan is linearly related to belief state b as $V_i(\mathbf{b}, \boldsymbol{\Theta}) = \sum_s b(s) \cdot \alpha_{i,\boldsymbol{\Theta}}(s)$. Figure 6, a graph inspired by Kaelbling et al. (1998) that refers to a simple example of a two-state POMDP. Belief is completely described by a scalar value $b(s_1)$, as $b(s_2) = 1 - b(s_1)$. Figure 6 reports the value for four conditional plans, and the bold lie indicates the optimal value, depending on the belief state.

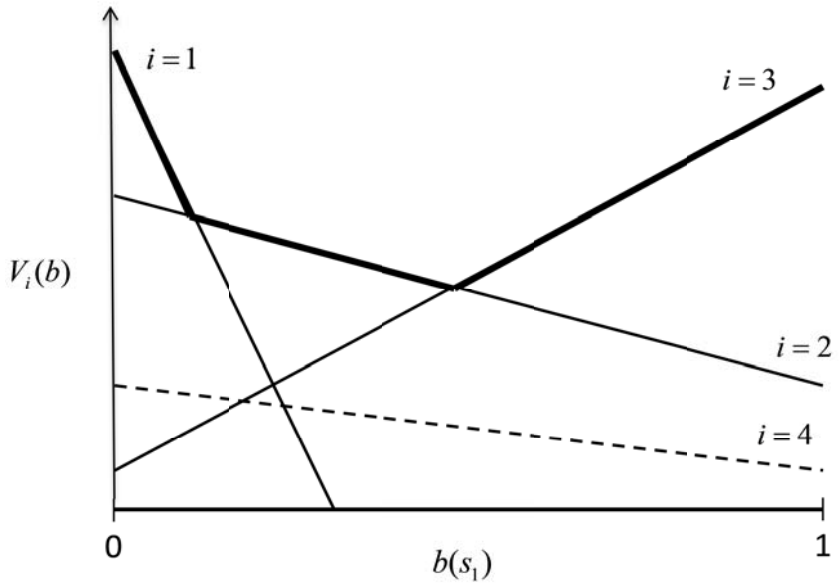

Figure 6. A simple example of value function for two state POMDP model (adapted from Kaelbling et al. 1998).

The optimal value function can be written as:

$$V^*(\mathbf{b}, \boldsymbol{\Theta}) = \min_i \sum_{s=1}^{|S|} b(s) \cdot \alpha_{i,\boldsymbol{\Theta}}(s) \tag{7}$$

Where i is defined on the domain $\{1,2,\dots,n_c\}$. The proof of Eq. (7) can be found in the work by Smallwood and Sondik (1973), which shows that the optimal value function for any finite horizon POMDP is a piecewise-linear and convex function over the domain of the belief B. Eq. (7) cannot be solved explicitly, except for very short time horizon, due to the high value of n_c. However, as it is clear in Figure 6, some conditional plans are completely dominated (e.g. plan 4) and can be neglected (Russell and Norvig 2010). It should be noted that each conditional plan begins with a specific first action, so Eq. (7) allows defining implicitly the optimal policy π^* as, for any belief state b, the optimal action is that to be executed as first one in the optimal conditional plan.

The computational complexity of solving POMDP problems and planning based on POMDP is discussed in detail by Hsu et al. (2007) and Shani et al. (2013). Exact solution of the POMDP problem can be found by the process known as *exact value iteration* (Kaelbling et al. 1998). In each iteration, the value function is updated cross the entire belief space and the size of α-vectors created in each iteration is denoted by $|V|$. The overall complexity of a single iteration is $O\left(|V| \times |A| \times |Z| \times |S|^2 + |A| \times |S| \times |V|^{|Z|}\right)$ (Shani et al. 2013). In practice, exact value iteration is only feasible for small problems as the size of the set of α-vectors grows exponentially with every iteration. As the computational cost of each iteration depends on the number of vectors in V, an exponential growth makes the algorithm prohibitively expensive.

Kaelbling et al. (1998) have proposed the so-called witness algorithm for finding the exact solution to POMDPs via value iteration. However, this algorithm is not practical when the set of states, actions, and observations are large. An alternative approach is to discretize the belief space, using either a fixed grid (Lovejoy 1991) or a variable grid (Zhou and Hansen 2001). The value of any belief is then defined by interpolation of the points on the grid. However, in general, regular grids do not scale well in problems with high dimensionality and non-regular grids suffer from expensive interpolation routines. Other point-based value iteration methods restrict search to the beliefs that can be reached starting from the initial belief state (Pineau et al. 2003). The full complexity of the point-based value iteration methods requires $O(|V| \times |A| \times |Z| \times |S|^2 + |A| \times |S| \times |Z|)$, as compared with the $O\left(|V| \times |A| \times |Z| \times |S|^2 + |A| \times |S| \times |V|^{|Z|}\right)$ of a single iteration of the exact method (Shani et al. 2013). In particular, one of the most effective point-based value iteration methods is successive approximations of the reachable space under optimal policies (SARSOP) (Kurniawati et al. 2008), which identifies the optimally reachable belief states, and approximates the optimal value function using this set. SARSOP represents the state-

of-the-art in solving POMDPs, in terms of efficiency and accuracy. As all algorithms for POMDP, SARSOP formally solves the finite horizon problem, but it can be used as an approximation to solve the infinite horizon case.

2.2.1 Illustrative Example for POMDP

In his section we illustrate how the POMDP framework operates by applying it to an illustrative example of managing a single component. Suppose an agent is managing a component whose condition state is described by only two possible states, *Intact* ($s = 1$) and *Damaged* ($s = 2$). She has access to two possible actions: *Do-Nothing* ($a = 1$, DN) and *Replace* ($a = 2$, RE). The cost of replacing a component is assumed to be \$100 and cost of damage is assumed to be \$200 and discount factor is 0.95. The transition probability of the component is given as follow:

$$\mathbf{T}_1 = \begin{bmatrix} 0.99 & 0.01 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{T}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

If the agent does nothing there is a chance of 1% for the component to become damaged in the next time step, while replacing the component improves its condition to intact with certainty. The agent also has access to noisy observations that is defined by the emission probability as follow:

$$\mathbf{O}_1 = \mathbf{O}_2 = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}$$

Where, ϵ is the probability of wrong measure. If the error is zero, then agent has perfect knowledge about the condition state of the component (which is MDP described in section 2.1), and she does nothing if the component is intact and replace otherwise. On the other hand, when

there is a probability of wrong measurement, the agent adapts a policy according to how reliable the information is. Figure 7 shows (a) the optimal value and (b) optimal policy as a function probability of damage P_{DAM} for different measurement errors, ϵ. It is clear from the figure that as the measurement error increases, the expected cost of operating this component increases as well, and the agent adapts a more conservative policy by replacing the component more often. For example in the case of $\epsilon = 0.50$, the measurements are useless and the agent cannot rely her maintenance policy on the observations, hence it adapts a very conservative policy of replacing the component as soon as the probability of damage is above 10%, while in the case of $\epsilon = 0.01$, agent has access to very reliable information about the component's condition state, hence she adapts a less conservative policy and replaces only if the probability of damage is above 35%.

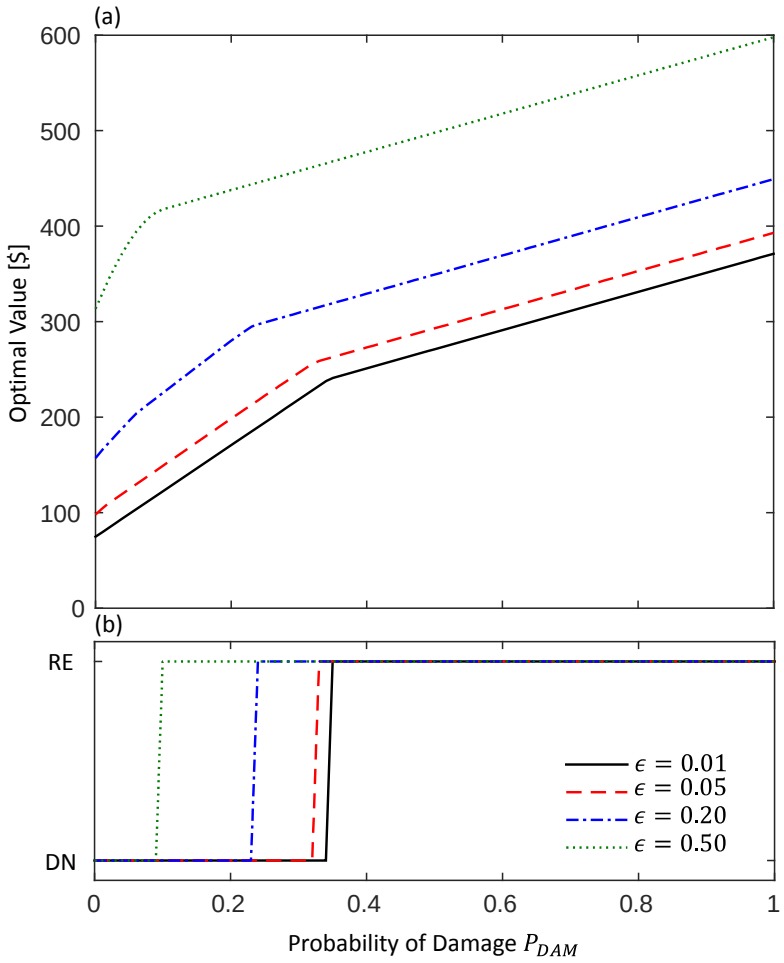

Figure 7. Optimal value (a) and optimal policy (b) as a function of P_{DAM} for different ϵ.

2.3 Bayes-Adaptive Partially Observable Markov Decision Process

Bayes-Adaptive POMDP (BA-POMDP) framework is a generalization of POMDP, where the transition and emission probabilities, $\mathbf{T}$ and $\mathbf{O}$, are unknown parameters of the model and are treated as random variables, with a prior distribution $P(\boldsymbol{\Theta})$ where $\boldsymbol{\Theta} = \{\mathbf{T}, \mathbf{O}\}$. Technically, the BA-POMDP model can be interpreted as a POMDP with a continuous state space, and with an augmented belief state that also includes $\boldsymbol{\Theta}$. The augmented belief state at time t is now defined as $\tilde{\mathbf{b}}_t = P[s_t, \boldsymbol{\Theta} \mid \bar{a}_{t-1}, \bar{z}_t]$. In principle, we can express the belief at time t as a function of that

at the previous step, as in POMDP formulation reported in Eq. (3). However, as in most cases we cannot find any closed-form representation of the posterior, in BA-POMDP it is easier to express the belief at any step by integrating the joint probability:

$$
\begin{aligned}
P(s_t, \boldsymbol{\Theta} \mid \bar{a}_{t-1}, \bar{z}_t) &\propto P(\bar{z}_t, s_t \mid \boldsymbol{\Theta}, \bar{a}_{t-1}) P(\boldsymbol{\Theta}) \\
&= P(\boldsymbol{\Theta}) \sum_{\bar{s}_{t-1} \in S^t} P(\bar{z}_t, \bar{s}_t \mid \boldsymbol{\Theta}, \bar{a}_{t-1}) \\
= P(\boldsymbol{\Theta}) \sum_{\bar{s}_{t-1} \in S^t} P(s_0) &\left[\prod_{s,a,s' \in [S \times A \times S]} \left(T(s,a,s')\right)^{N_{ss'}^a(\bar{s}_t, \bar{a}_{t-1})} \right] \times \\
&\left[\prod_{s,a,z \in [S \times A \times Z]} \left(O(s,a,z)\right)^{N_{sz}^a(\bar{s}_t, \bar{a}_{t-1}, \bar{z}_t)} \right]
\end{aligned}
\tag{8}
$$

Where S^t is the set of possible sequences of states up to time t, $N_{ss'}^a(\bar{s}_t, \bar{a}_{t-1})$ is the number of times the transition (s, a, s') appears in the process and $N_{sz}^a(\bar{s}_t, \bar{a}_{t-1}, \bar{z}_t)$ is the number of times the emission (s, a, z) appears in the process.

BA-POMDP framework can incorporate the uncertainties in the probabilities defining the transition and emission models; however its computational complexity grows exponentially with increase in the dimensionality of the problem, or longer management time horizons. In the next chapters we introduce a tractable approximate method to perform planning and learning within the BA-POMDP framework.

Chapter 3

Sequential Decision Making: Planning Under Model Uncertainty

Abstract

In this chapter, we propose an approximate method for planning under model uncertainty within the BA-POMDP framework. The proposed method includes the uncertainty in the model parameters describing the degradation behavior of components (i.e. transition probabilities) and precision of the monitoring system (i.e. emission probabilities) and it identifies the optimal action for operation and maintenance. The method is approximated, because it neglects the exploratory value of learning the model parameters. We compare the performance of the proposed method with POMDP planning on a numerical example of wind farm management.

3.1 Problem Formulation

One of the main limitations of the planning within the POMDP framework is that it assumes that the transition and emission probabilities are known with certainty. This is not a realistic assumption in many real-world management problems, because these probabilities are affected by epistemic uncertainty.

Now consider a decision making process, modeled as a POMDP, but with uncertain transition and emission probabilities, while cost function, initial belief and discount factor are fixed. The agent models her knowledge on these model parameters through a joint distribution, and she can solve the POMDP optimization problem for any model. In this setting, the problem is how to select an action: we refer to this as "planning under model uncertainty".

3.1 Proposed Method

In this section, we propose an approximate method called Planning and Learning for Uncertain dynamic Systems (PLUS) (Memarzadeh et al. 2013, 2015a). The planning method is based on two approximations. First, to neglect the exploratory value of learning variables $\mathbf{T}$, $\mathbf{O}$, i.e. the system model parameters, $\boldsymbol{\Theta} = \{\mathbf{T}, \mathbf{O}\}$. PLUS aims at identifying the optimal policy as that for transition and emission probabilities modeled by $P[\boldsymbol{\Theta} \mid \bar{a}_{t-1}, \bar{z}_t]$, neglecting the updating attributable to future observations. Consequently, according to the formulation of Durango and Madanat (2002), PLUS belongs to the "open-loop feedback control" method. They also propose the "closed-loop feedback control" method which incorporates the exploration by including the model uncertainty into the belief state. To formalize the second approximation, let us define $Q_{\boldsymbol{\Theta}}(a, \mathbf{b})$ as the quality of a belief-state-action (Q-value) for a POMDP, i.e. the value of starting

from belief $\mathbf{b}$, performing action a, and following the optimal policy after that, for a model defined by $\boldsymbol{\Theta}$, defined as follow:

$$Q_{\boldsymbol{\Theta}}(a, \mathbf{b}) = r(\mathbf{b}, a, \boldsymbol{\Theta}) + \gamma \sum_{z=1}^{|Z|} e_z(\mathbf{b}, a, \boldsymbol{\Theta}) V^*[u(\mathbf{b}, a, z, \boldsymbol{\Theta}), \boldsymbol{\Theta}] \quad (9)$$

So that:

$$V^*(\mathbf{b}, \boldsymbol{\Theta}) = \min_a Q_{\boldsymbol{\Theta}}(a, \mathbf{b}) \quad (10)$$

We can identify the optimal action a^* by the following approximate formula:

$$a^* \cong \underset{a}{\operatorname{argmin}}\, \mathbb{E}_{\boldsymbol{\Theta}}[Q_{\boldsymbol{\Theta}}(a, \mathbf{b})] \quad (11)$$

Where $\mathbb{E}_x$ indicates the statistical expectation, according to actual knowledge of variable x, and the belief state at time t is defined as in a POMDP as $\mathbf{b} = P[s_t \mid \boldsymbol{\Theta}, \bar{a}_{t-1}, \bar{z}_t]$. Eq. (11) represents an approximation, as it combines quantities related to optimal policies for different models. However, we do not use the approximation to estimate the value of the policy but only to select the current optimal action. Computationally, the advantage of Eq. (11) is that $Q_{\boldsymbol{\Theta}}(a, \mathbf{b})$ can be obtained from the results of a POMDP solver, i.e. SARSOP. Similar approaches have been used before for active learning in POMDPs with limited reinforcement using Bayes risk (Doshi-Velez et al. 2012).

The Q-value of a belief-state-action can be related to the α-vectors presented in section 2.2. For a model $\boldsymbol{\Theta}$ and belief $\mathbf{b}$, we can identify the optimal conditional plan starting with action a for each available action. We defined $\alpha^*_{a,\mathbf{b},\boldsymbol{\Theta}}(s)$ as the component referring to state s of the corresponding α-vector. The Q-value of a belief-state-action can be computed as:

$$Q_{\boldsymbol{\Theta}}(a, \mathbf{b}) = \sum_s b(s) \cdot \alpha^*_{a,\mathbf{b},\boldsymbol{\Theta}}(s) \tag{12}$$

Figure 8 presents the scheme of the planning algorithm, which is based on Eqs. (11-12). At time t, augmented belief state $\tilde{b}_t$ is represented by N samples. For each sample, we solve the corresponding POMDP problem, using SARSOP (Kurniawati et al. 2008). The outcome of SARSOP is the set of m non-dominated α-vectors. Among them, we select one optimal α-vector per each action: this is the pruning routine mentioned in the algorithm. α_j^* refers to the optimal vector for the j-th action, $Q_j^{(k)}$ to the Q-value of a belief-state-action for the k-th sampled model under the j-th action, and Q_j to the expected Q-value of a belief-state-action for the entire model space, which we compute by sample average. Action a^* is selected by identifying the maximum (minimum, when the goal is minimizing the cost) of Q_j among all possible actions.

PLUS Planning Algorithm

function PLANNING$\left(\left\{T^{(k)}, O^{(k)}, b_t^{(k)}\right\}_N^{k=1}, R, \gamma\right)$
  for $k = 1{:}\,N$ **do**
    $\{\alpha_h\}_m^{h=1} \leftarrow \text{SARSOP}(T^{(k)}, O^{(k)}, b_t^{(k)}, R, \gamma)$
    $[\alpha_1^*, \ldots, \alpha_A^*] = \text{PRUNING}(\{\alpha_h\}_m^{h=1}, b_t^{(k)})$
    for $j = 1{:}\,A$ **do**
      $Q_j^{(k)} \leftarrow \alpha_j^{*\mathrm{T}} \cdot b_t^{(k)}$
    end for
  end for
  for $j = 1{:}\,A$ **do**
    $Q_j \leftarrow \frac{1}{N}\sum_{k=1}^{N} Q_j^{(k)}$
  end for
  $a^* \leftarrow \underset{j}{\text{argmax}}\, Q_j$
  return a^*
end function

Figure 8. PLUS planning algorithm

3.2 Numerical Validation of Wind Farm Management

To validate our proposed planning approach, a numerical example of wind farm management is used through discussion with our industry collaborator, Everpower wind holdings (refer to Chapter 1 for more details). It is assumed that the condition state of each turbine can be modeled by a Markov process defined by a few states, and the observations collected can be classified within a few possible discrete values. Although PLUS can be applied to much more complicated problems, this simple setup allows us to extensively investigate the performance of the algorithm and compare it to other existing methods.

The condition state of the turbine degrades due to fatigue and aging, potentially causing a structural failure and a relevant economical loss to the agent. In turn, the agent can perform repairs to avoid failures and inspections to refine the knowledge about each condition state. In detail, we assume the farm consists of 10 turbines of the same type, so that we can refer to a unique value of transition and emission probabilities. Specifically, we assume three condition states: $s = 1$ refers to an intact structure, $s = 2$ to a damaged one, and $s = 3$ to a collapsed turbine; three actions: $a = 1$ corresponds to "Do Nothing" (DN), $a = 2$ to "Repair" (RE), and $a = 3$ to performing a "Visual Inspection" (VI). When DN is selected, the condition state evolves according to the degradation process. RE models a costly intervention which is supposed to improve the condition state, while VI models an effort providing only information on the condition state, without affecting the degradation process. Each time step is assumed to be six months, and the agent takes one action per turbine at each time step.

Observations are classified in 4 discrete outcomes: $z = 1$ is intended as a reassuring output, suggesting that the turbine is undamaged; $z = 2$ and $z = 3$ indicate two symptoms of damage; after recording $z = 4$, the agent knows that the turbine is collapsed.

We model the agent's prior knowledge is modeled on transition and emission probabilities by independent Dirichlet distributions with parameters $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ respectively. Introduction to the Dirichlet distribution can be found in the textbook of Murphy (2012). The choice of Dirichlet distribution for prior on model parameters will be clear in the next chapter (the reason is that Dirichlet is conjugate prior to multinomial distribution and transition and emission probabilities in the discrete POMDP follow a multinomial distribution, hence the posterior would be also in the form of Dirichlet). Parameter $\boldsymbol{\eta}$ can be represented by three matrices: $\boldsymbol{\eta}_{DN}$, $\boldsymbol{\eta}_{RE}$, and $\boldsymbol{\eta}_{VI}$, referring to the actions listed above:

$$\boldsymbol{\eta}_{DN} = \boldsymbol{\eta}_{VI} = \begin{bmatrix} 8 & 4 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\eta}_{RE} = \begin{bmatrix} 8 & 4 & 0 \\ 4 & 2 & 0 \\ 4 & 2 & 0 \end{bmatrix}$$

The transitions are assumed to be identical for actions DN and VI. The zeros in the matrix η_{DN} indicate that, after any of these actions, the condition state cannot improve, so that, for example, the turbine stays in a collapsed state after action DN. Generally, according to this matrix, the turbine in the intact state has a tendency to stay undamaged, but it can also become damaged or directly collapse, while that in the damaged state has a tendency to stay there, but it can also collapse. After action RE, the turbine cannot be in a collapsed state, but it can still be damaged, as the intervention is not known to be perfect and, even after a perfect repair, the turbine can transit to the damage state during the following period, considering the long time step (six months). As for any feature of the process, the effectiveness of such an intervention can be learnt by the agent during the management history. Knowledge about emissions, depending on the action, are modeled by the following values:

$$\boldsymbol{\beta}_{DN} = \boldsymbol{\beta}_{RE} = \begin{bmatrix} 8 & 4 & 2 & 0 \\ 2 & 8 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta}_{VI} = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

As can be deduced from these matrices, the agent thinks that, as a tendency, states 1 and 2 generate observations 1 and 2 respectively, under actions DN or RE. The visual inspection VI is regarded as possibly imperfect and, again, its actual effectiveness can be discovered during the management process. It is to be noted that, independently of the action, the collapse state 3 is univocally related to observation 4, so that the agent is immediately aware of any failure event.

The reward function is the sum of three components: the costs for repairing, inspecting, and down-time. The agent pays $10,000 for any repair, $500 for any visual inspection and $50,000 for any time step in which a turbine is in the collapsed state. The discount factor is assumed to be $\gamma = 0.95$.

The belief about the initial state is modeled as,

$$\mathbf{b}_0 = [0.8 \;\; 0.2 \;\; 0]$$

therefore, the agent believes that the turbines are in the “Intact” state with 80% probability and in “Damaged” state with 20% probability.

The behaviors of different turbines in the farm are assumed to be independent, and the agent refers her planning to the infinite horizon setting.

Transition and emission were fixed to a value compatible with the available knowledge, referring to this as the *true model*. The true model was assigned to each turbine in the farm, and the planning algorithm was tested for the range of all possible models representing the turbines.

In the experiments, we consider three types of agents: The *True Model* agent has perfect knowledge about the true underlying transition and emission probabilities, and adopts a POMDP model with correct value for **T** and **O**, making use of SARSOP algorithm for planning: this represents a lower bound to the performance any planning strategy under uncertainty. The

Expected Model agent derives the expected value of **T** and **O** from the prior Dirichlet distribution, and again adopts POMDP solved by SARSOP: it represents the simplest and most common approach to solve the planning problem under model uncertainty in the literature. The third agent, *PLUS*, adopts the method presented in section 3.1.

The immediate and cumulative management cost is evaluated for assessing the performance of the planning method, because they are directly related to what each agent is trying to optimize. Figure 9 reports the immediate (a) and cumulative (b) costs of O&M, for the true model, the expected model and the PLUS agents. Again, the *true model* agent represents the lower bound, leading to an immediate cost of about $2,900/6months, while the *expected model* agent achieves a cost of about $8,300/6months, and the *PLUS* agent a cost of about $7,700/6months. The difference between these latter values, i.e. $600/6months, quantifies the benefit of adapting the robust planning approach presented in this chapter. Naturally, adding the learning process as well would make PLUS perform much closer to the true model, but this experiment highlights the value of uncertainty-aware planning in and of itself. We evaluate the effect of learning in the next chapter. It should be noticed that these costs and savings are for a single turbine and the costs and savings regard the entire farm is ten times higher.

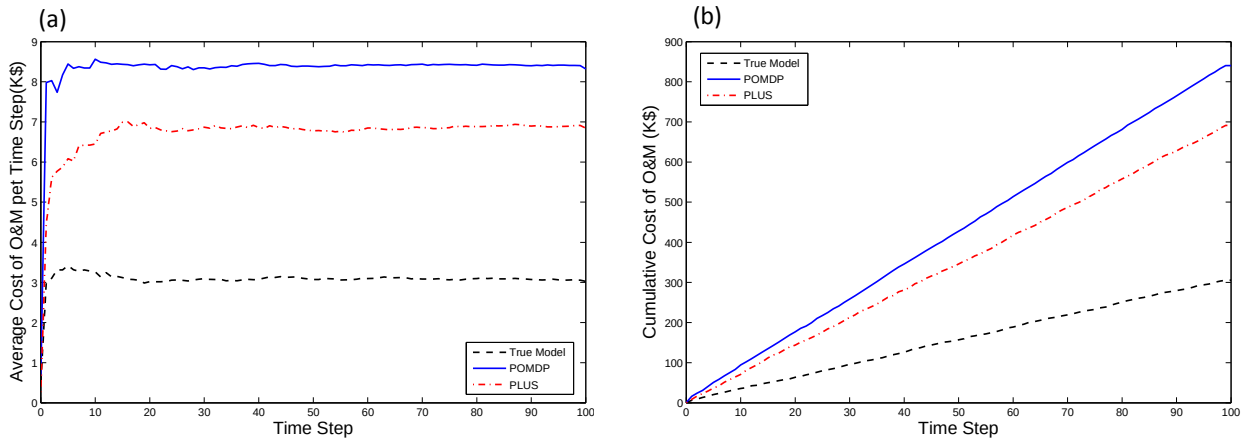

Figure 9. The planning performance of PLUS algorithm compared to POMDP and "true model" agents.

Chapter 4

Sequential Decision Making: Learning

Abstract

In this chapter, we focus on the problem of learning the degradation behavior of components (transition probabilities), as well as the precision of monitoring system (emission probabilities) by processing noisy observations. We first introduce the learning procedure in PLUS (planning and learning for uncertain dynamic systems). PLUS models the components as either independent (learning an independent model for each component) or identical (learning a global model for all components). When the system is made up by similar components, data collected on one is also relevant in the management of others. We extend the formulations of PLUS, allowing a weaker similarity among components. The proposed approach, called Multiple Uncertain POMDP (MU-POMDP), assumes the model parameters as dependent random variables among components, and allows the transfer of knowledge among them by using a set of hyper-parameters. We evaluate the performance of PLUS compared to state-of-the-art methods in reinforcement learning and then evaluate the performance of MU-POMDP compared to PLUS.

4.1 Planning and Learning for Uncertain dynamic Systems – PLUS

4.1.1 Problem Statement

PLUS introduces an approximate method for planning and learning under model uncertainty. The planning phase has been discussed in previous chapter. In this section, we focus on the learning phase. By processing noisy observations, how can an agent learn the degradation behavior and the precision of monitoring system? This is a challenging task as the condition states of the components are not observable.

4.1.2 Proposed Method

In this section, we propose an approximate method for optimally planning and learning in uncertain dynamic system (PLUS) within the BA-POMDP framework (Memarzadeh et al. 2013, 2015a). Figure 10 shows the overall PLUS method, which is organized in two main parts: learning and planning. Details regard the planning part were discussed in Chapter 3. The algorithm can be called at any stage of the process. At time t, it represents the augmented belief $\tilde{\mathbf{b}}_t$ by a set of samples, and it suggests action a^*. In the algorithm, notation $x^{(k)}$ indicates the k-th sample of variable x.

PLUS Algorithm

function PLUS $(\eta, \beta, b_0, \bar{a}_t, \bar{z}_t, R, \gamma, N)$

- **Learning**

$\left\{T^{(k)}, O^{(k)}, b_t^{(k)}\right\}_N^{k=1} \leftarrow LEARNING(\eta, \beta, b_0, \bar{a}_t, \bar{z}_t, N)$

- **Planning**

$a^* \leftarrow PLANNING\left(\left\{T^{(k)}, O^{(k)}, b_t^{(k)}\right\}_N^{k=1}, R, \gamma\right)$

return $a^*, \left\{T^{(k)}, O^{(k)}, b_t^{(k)}\right\}_N^{k=1}$

end function

Figure 10. Planning and learning for uncertain dynamic systems (PLUS) algorithm

The PLUS algorithm makes use of an approximate method based on Markov Chain Monte Carlo (MCMC) Gibbs sampling (Carter and Kohn 1994). The present approach is a slight variation of the beam sampling approach used in the context of infinite Hidden Markov Models (Van Gael et al. 2008) and infinite POMDPs (Doshi-Velez 2010). Figure 11 shows the details of the proposed algorithm for learning: the method samples N instances of **T**, **O** and belief state $\mathbf{b}_t$ from the joint posterior distribution. We start sampling **T**, **O** from the corresponding prior Dirichlet distributions, then we alternate between sampling state sequence $\bar{s}_t$, and sampling **T** and **O**. For each fixed **T** and **O**, a state sequence is drawn by forward filtering backward sampling (FFBS) (Fruhwirth-Schnatter 2006), as described below in the next section. In turn, as noted above, the posterior distribution given each sample $\bar{s}_t$ is still in the Dirichlet family. Parameter set $\boldsymbol{\eta}'$defines the updated Dirichlet distribution for the transition probabilities, depending on sampled state sequence $\bar{s}_t$, while $\boldsymbol{\beta}'$ defines that of the emission probabilities, depending on $\bar{s}_t$ and observations $\bar{z}_t$. It should be noticed that, in the limit of an infinite burn-in phase, this proposed method is selecting samples from the true posterior distribution. In Figure 11, n_b indicates the number of samples in the burn-in phase, to be discarded (Murphy, 2012), and the notation $x \sim p$ indicates that sample x is generated from distribution p.

PLUS Learning Algorithm

function LEARNING$(\eta, \beta, b_0, \bar{a}_{t-1}, \bar{z}_t, N, n_b)$

$T^{(0)} \sim \mathrm{Dir}(\eta)$

$O^{(0)} \sim \mathrm{Dir}(\beta)$

for $k = 1{:}\,(N + n_b)$ **do**

$\left(\bar{s}_t^{(k)}, b_t^{(k)}\right) \leftarrow \mathrm{FFBS}\left(T^{(k-1)}, O^{(k-1)}, b_0, \bar{a}_{t-1}, \bar{z}_t\right)$

$\eta' \leftarrow \mathrm{UpdateDirichlet}\left(\eta, \bar{s}_t^{(k)}, \bar{a}_{t-1}\right)$

$T^{(k)} \sim \mathrm{Dir}(\eta')$

$\beta' \leftarrow \mathrm{UpdateDirichlet}\left(\beta, \bar{s}_t^{(k)}, \bar{a}_{t-1}, \bar{z}_t\right)$

$O^{(k)} \sim \mathrm{Dir}(\beta')$

end for

return $\left\{T^{(k)}, O^{(k)}, b_t^{(k)}\right\}_{N+n_b}^{k=1+n_b}$

end function

Figure 11. PLUS learning algorithm

4.1.2.1 Forward Filtering Backward Sampling

FFBS is a multi-move sampling method for discrete systems (Fruhwirth-Schnatter 2006). The steps are as follow: (1) For each time step j ranging from 0 to t, we derive posterior probability $P(s_j \mid \boldsymbol{\Theta}, \bar{a}_{j-1}, \bar{z}_j)$, solving the so-called "filtering" problem; and (2) We sample state s_t' from the last distribution and s_j', from time step $j = t - 1$ backward to $j = 0$, from distribution $F(s_j) \propto P(s_j \mid \boldsymbol{\Theta}, \bar{a}_{j-1}, \bar{z}_j) P(s_{j+1}' \mid \mathbf{T}, s_j, a_j)$. The outcome of FFBS algorithm is the sequence of states $\{s_0', \ldots, s_t'\}$ sampled from distribution $P(\bar{s}_t \mid \boldsymbol{\Theta}, \bar{a}_{t-1}, \bar{z}_t)$.

4.1.3 Numerical Validation of PLUS

Details of the numerical example used for validation is provided in section 3.2 of previous chapter. We consider four types of agents: The *true model* agent has perfect knowledge about the

true underlying transition and emission probabilities, and adopts a POMDP model with correct value for **T** and **O**, making use of the SARSOP algorithm for planning: this represents a lower bound to the performance of any planning strategy under uncertainty. The *Expected Model* agent derives the expected value of **T** and **O** from the prior Dirichlet distribution and, again, adopts POMDP solved by SARSOP: it represents the simplest and most common approach to solve the planning problem under model uncertainty. The MEDUSA agent makes use of the algorithm described in Jaulmes et al. (2005a,b), while the PLUS agent adopts the method that was presented in sections 4.1.1 and 3.1.

Two different metrics are used to validate the methods. First, the immediate and cumulative management cost for assessing the performance of the planning methods is evaluated, because they are directly related to what each agent is trying to optimize. For additional validation of the learning process itself, evaluate the Kullback-Leibler (KL) divergence (Cover and Thomas 2006) between the transition (or emission) probabilities as modeled by the posterior distribution and in the true model. The KL divergence is a non-symmetric measure of the differences between two probability distributions. Specifically, the KL divergence of distributions Q from distribution P (both being distributions defined on n discrete values), denoted as $D_{KL}(P||Q)$, is a measure of information lost when Q is used to approximate P, and is defined as:

$$D_{KL}(P||Q) = \sum_{i=1}^{n} \ln\left(\frac{P(i)}{Q(i)}\right) P(i) \tag{13}$$

where ln indicates natural logarithm. In computing the KL divergence between two transition (or emission) models, the results referring to the average over all values of s_t and a_t.

In order to validate, we have fixed a model and assigned it to all turbines. This is called the *true model*, and it is defined by transition $\mathbf{T}^*$ and emission $\mathbf{O}^*$, as listed in the following:

$$\mathbf{T}^*_{DN} = \mathbf{T}^*_{VI} = \begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}^*_{RE} = \begin{bmatrix} 1 & 0 & 0 \\ 0.9 & 0.1 & 0 \\ 0.9 & 0.1 & 0 \end{bmatrix}$$

$$\mathbf{O}^*_{DN} = \mathbf{O}^*_{RE} = \begin{bmatrix} 0.8 & 0.1 & 0.1 & 0 \\ 0.05 & 0.9 & 0.05 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{O}^*_{VI} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This specific model describes a turbine that is more reliable than that defined by the expected value of the distribution reported in the previous section. These models were selected by adapting examples from the literature (Byon et al. 2010, Byon and Ding 2010), Nielsen and Sorensen 2012) after discussion with industry experts from EverPower Wind Holdings (Pittsburgh, PA). For example, the probability of a collapse in one 6-month period, for an intact turbine, is only 2%. The emissions related to the Visual Inspection models perfect information on the condition state.

For each agent, the management of the wind farm is simulated 20 times, and the average outcome is plotted in Figure 12. In each simulation, the initial state is sampled according to the distribution $\mathbf{b}_0$. Figure 12a reports the average immediate cost *vs* the time step. The black dashed line represents the true model agent, the blue line represents the expected model agent, and the red dash-dotted line represents the PLUS agent, while other colors refer to the MEDUSA algorithm, with learning rate (LR) of 0.1, 0.5 and 1.

Each agent starts with a low cost in the first steps due to the good state of the turbines, as assumed by the initial belief state. The true model and the expected model agents adopt a stationary policy, and the corresponding immediate cost converges to a constant value, which is about \$2,200/6months for the former, and \$3,500/6months for the latter agent. Fluctuations are

due to randomness in the average of the small set of simulations. Agents adopting the MEDUSA and the PLUS algorithm, on the other hand, adopt non-stationary policies because of the learning process. At each time, the knowledge about the model is affected by processing the previous observations, and the policy varies accordingly. Ideally, if sufficient information is collected, the policies (and consequently the immediate cost) of these agents should converge to that of the "true model" agent. As expected, it is apparent from the figure that the immediate cost grows in the first phase (i.e. the first 10-20 steps), and then is reduced in time, because of the effect of learning. The PLUS algorithm also performs well in the first phase because of the robust algorithm for planning. After 30 steps, the immediate cost is about $2,600/6months. In this simulation, the MEDUSA algorithm achieves a higher cost for a range of different learning rates. The benefit of the PLUS algorithm over the expected model approach can be quantified as about $1,000/6months.

Figure 12b shows the cumulative costs of O&M, computed as the integral in time of the curves plotted in Figure 12a. This representation is useful for assessing the long term benefit of adopting alternative schemes. In a 100-step period (corresponding to 50 years), the true model agent expects a cost of about $220,000, the expected model agent a cost of about $350,000, while the PLUS agent expects a cost of about $250,000. Thus, the benefit of adopting PLUS is quantifiable to about $100K for this period. It should be noticed that these costs and savings are for a single turbine and the costs and savings regard the entire farm is ten times higher.

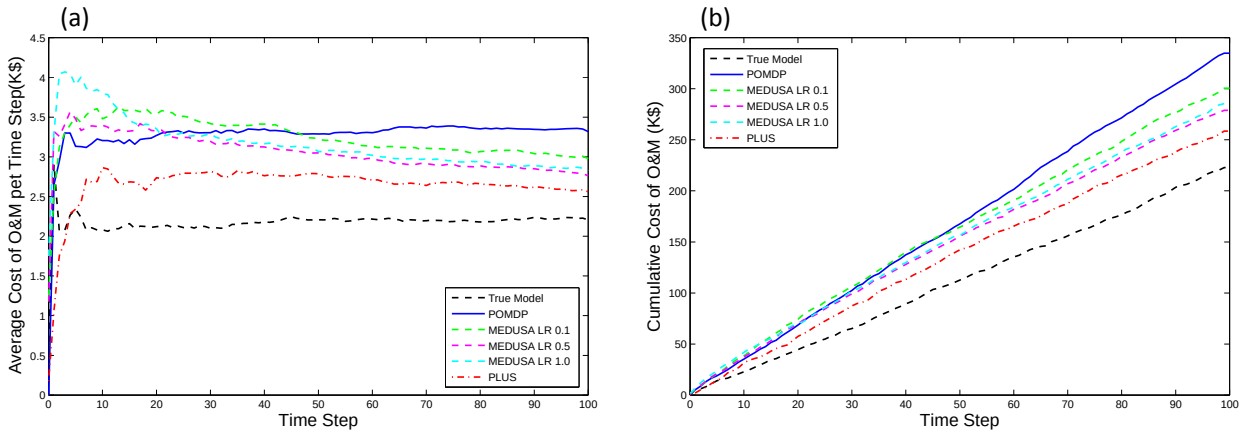

Figure 12. Costs for O&M of a wind farm versus time for six agents: (a) immediate; (b) cumulative average.

Figure 13 shows the cumulative costs for O&M of wind farm for PLUS, true model and POMDP agents including the 95% confidence intervals.

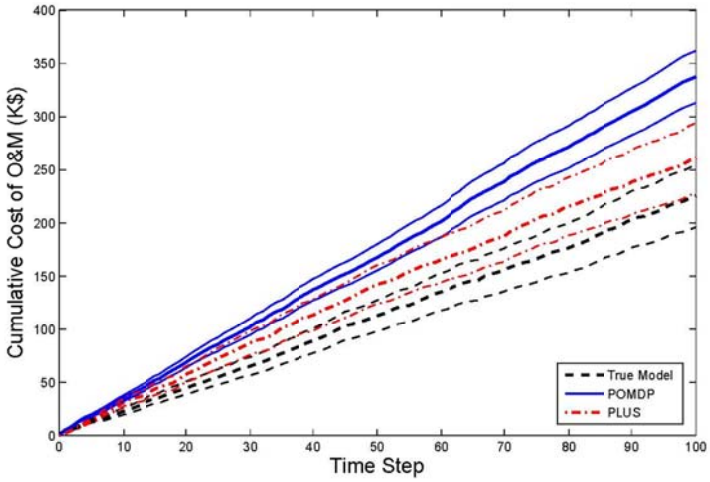

Figure 13. Cumulative costs for O&M of a wind farm vs time, for 3 agents including the 95% confidence intervals.

Figure 14 focuses on the learning process, showing the evolution of the KL divergence between the posterior distribution of the model, as formulated by each agent, and the true model. Figure 14a plots the transition probabilities and Figure 14b the emission probabilities. The expected model agent does not learn during the process and, consequently, her KL divergence is constant. The agents using MEDUSA or PLUS update their knowledge during the management

process, and we expect the KL divergence will go to zero when the information encoded in the collected observations is sufficient to identify the model. For these agents, the KL divergence is computed as the average from a set of samples generated according to the posterior distribution (as illustrated in section 4.1.1, PLUS algorithm requires to generate samples, so this further computation is straightforward). We have used 10 samples in this simulation. As shown in the figure, the learning is fast in the initial phase, but it becomes slow as more and more observations have been already collected. According to this simulation, the MEDUSA agents learn the transition probabilities well, but not the emission probabilities (Figure 14b). MEDUSA learns the emission probabilities poorly, perhaps because of their different planning approach compared with PLUS, and may need more data. However, in the long run, provided that sufficient exploration is performed, MEDUSA is conjectured to asymptotically learn the true model. Generally, MEDUSA and PLUS are different in terms of the tradeoff between computational cost and accuracy: MEDUSA is computationally cheaper and easier to scale; however, it provides less accurate solutions compared with PLUS.

Figure 14 shows that initially the KL divergence of the expected model agent is lower than that of the PLUS agent. This is a random effect owing to the selection of the true model in this simulation. The expected model agent adopts the mean transition and emission. Depending on the actual model of the turbine, it may be the case that the KL divergence can be arbitrarily small, and possibly much smaller than that of the PLUS agent. In other words, it may be the case that the model assumed by the expected model agent is actually the correct one, and therefore no learning is needed. Generally, the performance of the alternative methods depends on the specific actual model. In the next section, we perform a validation of the planning algorithm for all possible models.

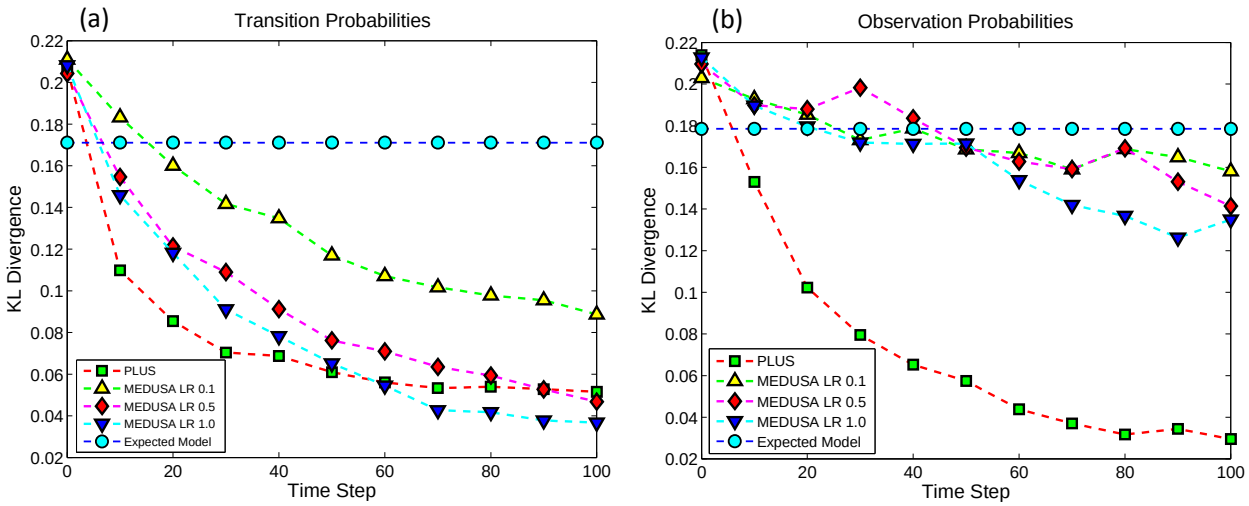

Figure 14. The performance of our proposed learning methodology (PLUS) compared to MEDUSA (with different learning rates (LR)) and POMDP (do not involve learning). The graphs show the KL divergence between each mode and the true model parameters.

Figure 15 shows the same results in figure 13a, including the 95% confidence intervals for the learning process of PLUS agent.

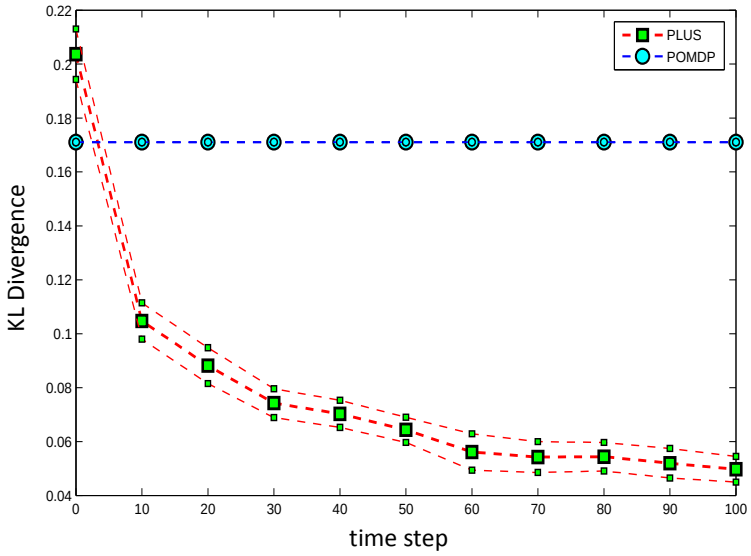

Figure 15. The performance of our proposed learning methodology (PLUS) compared to POMDP (do not involve learning) including the 95% confidence intervals. The graphs show the KL divergence between each model and the true model parameters.

4.1.4 Limitations of PLUS Learning Phase

PLUS allows for a rational treatment of data collected in-field (e.g. by sensors and visual inspections), a reliable tracking of the condition state of turbines, and robust decision making support. There are two modes of implementing the PLUS algorithm for a system made up by a set of components. The first one, that we name *Individual PLUS*, assumes that components are completely independent from each other. The second, that we name *Global PLUS*, assumes that all components are identical. In Individual PLUS the observations collected on one component are only used for updating model parameters of that specific component while in Global PLUS the observations of one component are used to update the entire system. For both implementations, PLUS allows the agent to learn, during the management process, the degradation process and the performance and reliability of the monitoring system.

4.2 Multiple Uncertain Partially Observable Markov Decision Process

4.2.1 Problem Statement

As mentioned above, depending on the implementation mode PLUS is an appropriate method to model the management of a set of components controlled by one single model (Global PLUS) or by independent models (Individual PLUS). However, a system can be composed by components controlled by similar but not identical models. This happens, e.g., when components of different typologies are exposed to the same environment, or when the components of the same typology are exposed to different environments. In this context it is appropriate to assume dependence among the models, with a degree that varies according to the application. Despite the limit cases of independent and of identical models can be solved by Individual and Global PLUS

respectively, the intermediate case poses specific computational problems, which we address in this section. Formally, the problem is defined as follow. Suppose to manage a set of components, each modeled by a POMDP. The set of parameters controlling the POMDPs are uncertain, and dependent among themselves. In this context, how can we (i) formulate a probabilistic model to capture the dependence among the parameters, (ii) develop an analytical and numerical technique to infer the variables in the problem, and (iii) define an approach to identify the optimal management policy?

4.2.2 Proposed Method

4.2.2.1 General MU-POMDP Framework

To address the first research question posed in previous section, we make use of the hierarchical Bayesian modeling approach, based on the PLUS approach (Chapter 3 and Section 4.1). Hierarchical Bayesian approach have been used before in the context of MDPs for multi-task reinforcement learning to allow transferring knowledge between different by related reinforcement learning tasks (Wilson et al. 2007). We refer to the proposed framework as Multiple Uncertain POMDP (MU-POMDP), and Figure 16 shows the corresponding probabilistic graphical model, for a system with two components. Only variables related to time steps $(t-1)$ and (t) are shown in the figure. The reader is referred to Chapter 2 for details of the classical POMDP framework which, as indicated in the figure, is used to model each component.

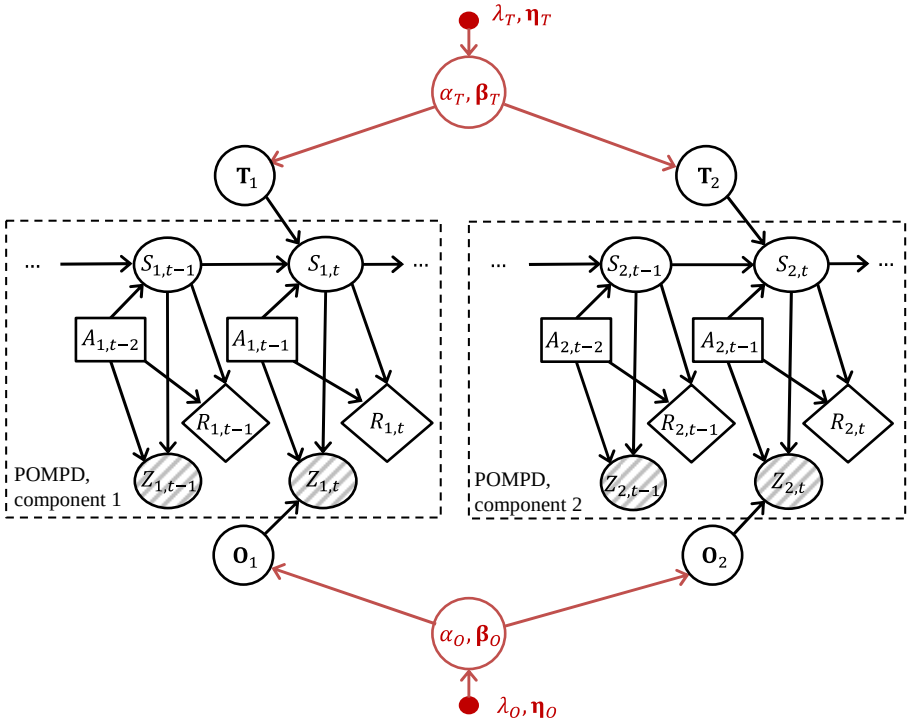

Figure 16. Graphical model of multiple uncertain POMDP (MU-POMDP) framework.

Subscript "k, t" refers the variable to component k at time t. MU-POMDP makes use of an additional layer of hyper-parameters, to model the dependence among the model parameters of different components. Hyper-parameters are marked as α_T, $\boldsymbol{\beta}_T$, α_O and $\boldsymbol{\beta}_O$ in Figure 16: the first two values define the dependence in the transitions, while the latter define that of emissions. While model parameters are different for each component, hyper-parameters are common to the entire system. Formally, matrices $\boldsymbol{\beta}_T$ and $\boldsymbol{\beta}_O$ have the same dimension of $\mathbf{T}_k$ and $\mathbf{O}_k$ respectively, while α_T and α_O are scalar variables. The role of these variables will become apparent in the following sections. Parameter matrices $\boldsymbol{\eta}_T$ and $\boldsymbol{\eta}_O$, of dimension equal to that of $\mathbf{T}_k$ and $\mathbf{O}_k$ respectively, and scalar variables λ_T and λ_O define the distribution of hyper-parameters.

The overall purpose of the inference task is to represent the posterior distribution of the variables in the problem. In this context, the posterior distribution is defined as conditional to all observations Z and actions A observed up to present time. In principles, once each conditional

distribution is analytically defined, prediction of any future variable can be performed, depending on the policy adopted. However, exact inference in not feasible in the general layout presented in Figure 16, and approximate methods needs to be adopted.

4.2.2.2 MCMC Updating Scheme

Extending the approach used in PLUS, in this section we propose to adopt a numerical scheme based on Markov Chain Monte Carlo (MCMC) (MacKay 2003). Using MCMC, the joint posterior distribution is represented by a set of samples. PLUS is based on Gibbs sampling, which is an effective implementation of MCMC. Specifically, PLUS alternates sampling the state trajectory and sampling model parameters. Using the specific distribution proposed in PLUS, the former task is accomplished by using forward filtering backward sampling (FFBS) (Fruhwirth-Schnatter 2006), once fixed the model parameters. On the other hand, once the state trajectory is assigned, the distribution of model parameters can be updated in theoretically, and a new sample can be generated. MU-POMDP is based on an extension of that method. Figure 17 reports a scheme of the inference process. In that figure, the upper bar indicates a collection of variables, from the beginning of the management process up to a specific time. For example, $\bar{S}_{k,t}$ indicates the state trajectory $\{S_{k,1}, \dots, S_{k,t}\}$ for component k. The superscript (j) refers to the j-th samples generated by the MCMC algorithm. At component level, the sampling of states and model parameters is identical to that adopted by PLUS. At system level, the hyper-parameters are sampled conditional to the sampled model parameters for all components, and we indicate with $\mathbf{T} = \{\mathbf{T}_1, \dots, \mathbf{T}_K\}$, and $\mathbf{O} = \{\mathbf{O}_1, \dots, \mathbf{O}_K\}$ the set of transition and emission respectively. This task can be accomplished by using the Metropolis-Hastings (M-H) approach (MacKay 2003). In summary, Figure 17 can be read as a recipe for generating samples from the joint posterior

distribution: model parameters and hyper-parameters are initialized at stage zero, then states and model parameters are sampled for each component, then hyper-parameters are sampled as well, and these latter steps are iterated indefinitely.

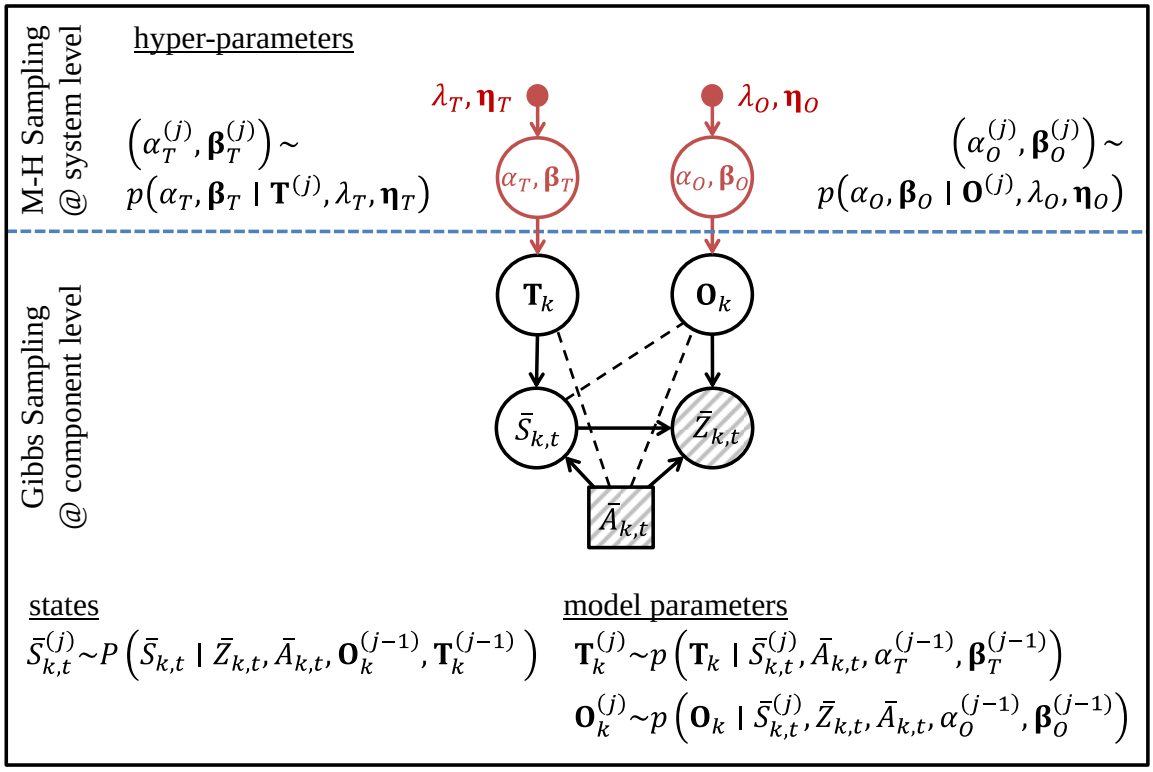

Figure 17. The proposed Markov chain Monte Carlo (MCMC) sampling approach.

4.2.2.3 Hierarchical Approach and Probabilistic Models

The graphical model in Figure 16 requires a specific assignment of marginal and conditional distributions for every random variable. In this section, we propose a probabilistic model inspired by Kemp et al, (2007), defined as follows:

$$
\begin{array}{lll}
 & \alpha_T \sim \text{Exponential}(\lambda_T) & \alpha_O \sim \text{Exponential}(\lambda_O) \\
 & \boldsymbol{\beta}_T \sim \text{Dirichlet}(\boldsymbol{\eta}_T) & \boldsymbol{\beta}_O \sim \text{Dirichlet}(\boldsymbol{\eta}_O) \\
\forall k = 1, \dots, K & \mathbf{T}_k \sim \text{Dirichlet}(\alpha_T \boldsymbol{\beta}_T) & \mathbf{O}_k \sim \text{Dirichlet}(\alpha_O \boldsymbol{\beta}_O) \\
\forall k \neq l & \mathbf{T}_k \perp \mathbf{T}_l \mid \alpha_T, \boldsymbol{\beta}_T & \mathbf{O}_k \perp \mathbf{O}_l \mid \alpha_O, \boldsymbol{\beta}_O \\
\forall t = 1,2, \dots, \infty & S_{k,t} \mid S_{k,t-1}, A_{k,t} \sim \text{Multinomial}(\mathbf{T}_k) & \\
 & Z_{k,t} \mid S_{k,t}, A_{k,t} \sim \text{Multinomial}(\mathbf{O}_k) &
\end{array}
\tag{14}
$$

where $x \sim f(y)$ indicates that variable x is distributed according to distribution f, parameterized with y, and $x \perp y \mid z$ indicates that random variables x and y are independent, given z. The reader can refer to text book of Kobayashi et al. (2012) for definition of Exponential, Dirichlet, and Multinomial distributions. Specifically, the Multinomial distribution of states and observations follows the classical assumptions of the POMDP framework. The PLUS framework can be obtained by the assumption outlined in Eq. (14), by fixing the hyper-parameters, instead of treating them as random variables. The Dirichlet distribution on model parameters is appropriate in this context, because it is conjugate prior of the multinomial distribution, and this facilitates the implementation of the Gibbs approach. As noted above, the model parameters of different components are not marginally independent, because of the common hyper-parameters parents. Consequently, observations on any component, by affecting the knowledge of the hyper-parameter, affect in turn all variables in the system. It is worth to clarify the role of hyper-parameters α and $\boldsymbol{\beta}$ in the definition of the prior distribution of model parameters. Each row in matrix $\boldsymbol{\beta}$ is normalized to one, as it follows the Dirichlet distribution. The entries in matrix $\boldsymbol{\beta}$ define the expected value of the corresponding model parameters. Scalar variable α affects the uncertainty of model parameters: intuitively a high value of α induces a low variance of the model parameters. α is modeled as an exponentially distributed random variable, and parameter λ defines the rate of this distribution. Similarly, the value of $\boldsymbol{\beta}$ describing

the system is an uncertain quantity, and $\boldsymbol{\eta}$ defines the parameters of the corresponding Dirichlet distribution.

4.2.2.4 Inference on Hyper-parameters

As outlined in Section 4.2.2.2, we propose to perform inference via the scheme reported in Figure 16. Samples of states and model parameters are generated as in PLUS, However, Eq. (14) does not allow to define in close form the conditional probability of the hyper-parameters: $p(\alpha_T, \boldsymbol{\beta}_T \mid \mathbf{T}, \lambda_T, \boldsymbol{\eta}_T)$ and $p(\alpha_O, \boldsymbol{\beta}_O \mid \mathbf{O}, \lambda_O, \boldsymbol{\eta}_O)$. As anticipated above, we propose to make use of the M-H algorithm to generate samples from these distributions. Figure 18 reports a complete algorithm to do that, for hyper-parameters α_T and $\boldsymbol{\beta}_T$ only (the corresponding procedure for α_O and $\boldsymbol{\beta}_O$ being identical, with obvious changes in the input variables). Input variables are the parameters defining the prior distribution (λ_T and $\boldsymbol{\eta}_T$), the transition probabilities for all components ($\mathbf{T}$), the step-size for the proposal distribution in the direction of α_T (σ_α), the concentration parameter for the proposal distribution in the direction of β_T (c_β), which acts as the inverse of a step-size, and the length of the Markov Chain (J). The choice of the proposal distribution is derived by the work of Kemp et al. (2007). In Figure 18, $\mathrm{Normal}(\mu, \sigma)$ indicates the normal distribution with mean μ and standard deviation σ, $\mathrm{Uniform}(a, b)$ the uniform distribution between a and b, $\mathrm{Dirichlet}(\mathbf{x}; \mathbf{y})$ the value assumed by the Dirichlet distribution with parameters $\mathbf{y}$ at $\mathbf{x}$. P indicates the un-normalized joint distribution of hyper-parameters and model parameters that, following Eq. (14), reads:

$$P(\mathbf{T}, \alpha_T, \boldsymbol{\beta}_T, \lambda_T, \boldsymbol{\eta}_T) =$$

$$\lambda_T \exp(-\lambda_T \alpha_T) \times \text{Dirichlet}(\boldsymbol{\beta}_T; \boldsymbol{\eta}_T) \times \prod_{k=1}^{K} \text{Dirichlet}(\mathbf{T}_k; \alpha_T \boldsymbol{\beta}_T) \tag{15}$$

input: $\lambda_T, \boldsymbol{\eta}_T, \mathbf{T}, \sigma_\alpha, c_\beta, J$
initialize $\alpha_T^{(1)}, \boldsymbol{\beta}_T^{(1)}$
for $j = 1, \ldots, J$ do
  sample $r \sim \text{Normal}(0,1)$
  $\alpha_T' = \exp\left[\log\left(\alpha_T^{(j)}\right) + r \times \sigma_\alpha\right]$
  sample $\boldsymbol{\beta}_T' \sim \text{Dirichlet}\left(c_\beta \boldsymbol{\beta}_T^{(j)}\right)$
  p-ratio $= P(\mathbf{T}, \alpha_T', \boldsymbol{\beta}_T', \lambda_T, \boldsymbol{\eta}_T) / P\left(\mathbf{T}, \alpha_T^{(j)}, \boldsymbol{\beta}_T^{(j)}, \lambda_T, \boldsymbol{\eta}_T\right)$
  q-ratio $= \frac{\text{Dirichlet}(\boldsymbol{\beta}_T'; c_\beta \boldsymbol{\beta}_T^{(j)})}{\text{Dirichlet}(\boldsymbol{\beta}_T^{(j)}; c_\beta \boldsymbol{\beta}_T')} \times \frac{\alpha_T'}{\alpha_T^{(j)}}$
  accept = p-ratio × q-ratio
  sample $r' \sim \text{Uniform}(0,1)$
  if accept $> r'$
    $\alpha_T^{(j+1)} = \alpha_T'$, $\boldsymbol{\beta}_T^{(j+1)} = \boldsymbol{\beta}_T'$
  else
    $\alpha_T^{(j+1)} = \alpha_T^{(j)}$, $\boldsymbol{\beta}_T^{(j+1)} = \boldsymbol{\beta}_T^{(j)}$
  end
end
output: hyper-parameters $\alpha_T^{(J+1)}$, $\boldsymbol{\beta}_T^{(J+1)}$

Figure 18. Metropolis-Hasting (MH) algorithm for sampling hyper-parameters on transition.

With this algorithm, we provide a complete recipe for a numerical implementation of the procedure outlined in Section 4.2.2.2. In that context, the value of **T** is assigned as the sample got from the Gibbs step, as reported in Figure 17.

At any state during the management process, the overall procedure provides samples of the model parameters and component state that can be used for the approximate decision optimization scheme of PLUS (Chapter 3). Following this remark, in this chapter we will not investigate the effectiveness of the policy search, and we will focus of the learning procedure

only. Specifically, we will investigate two aspects. (i) Despite it is well-known that Gibbs and M-H algorithms are consistent, we want to assess if the numerical procedure is feasible, using a reasonable number of samples; (ii) simplified procedure would derived assuming simpler dependence structure among model parameters, and we want to measure the degree of approximation induced by these assumptions. To address these questions, the next section refers to a simplified problem, which allows us an extensive numerical investigation.

4.2.3 Illustrative Example of a System with Similar Binary Components

4.2.3.1 Problem Formulation

Figure 19 shows the graphical model of a special K-component system that can modeled in the MU-POMDP framework. It is a static system, related to Figure 16 in the following way. Suppose only one action is available, so the decision variable is dummy, number of states ($|S|$) is equal 2, transition probability is so that each state is independent of the previous one (given the model parameters), and observations are perfect, meaning that the emission matrix is the identity. Given this latter remark, states and observations are identical and we can drop the state variables from the graph, relating directly model parameters to observations. In this set-up, the only model parameters for each component are the marginal probabilities assigned to the two possible observations, at any time. We name the two possible observations as *intact* and *failure*, and we assign value $Y = s_1 = 0$ to the intact state $Y = s_2 = 1$ to the failure state. We define the corresponding probabilities as $\theta_{k,i} = \mathrm{P}[Y_{k,t} = s_i]$, where $Y_{k,t}$ indicates the t-th observation from the k-th component. The two parameters describing the k-th component can be grouped in the normalized vector $\boldsymbol{\theta}_k = [\theta_{k,1} \quad \theta_{k,2}]$. Following the analogy with Section 4.2.2.1, $\boldsymbol{\theta}_k$

corresponds to $\mathbf{T}_k$ as defined before. As indicated in Figure 19 by using the plate notation (or "plate model") (Koller and Friedman 2009), we assume to have n observations from each component. Actually, the reader should think that observations are collected in time, so n corresponds to time indicator t.

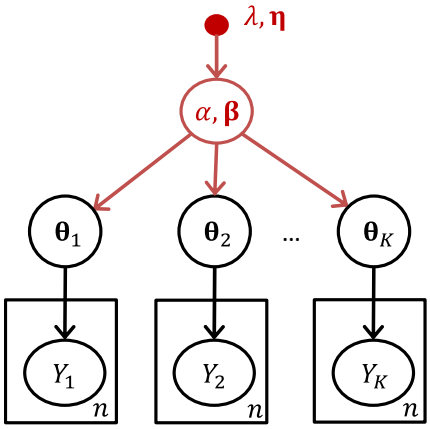

Figure 19. The MU-POMDP framework's probabilistic graphical model for the toy problem.

This formulation models the behavior of components that have a tendency to fail at any time. $\theta_{k,2}$ is the probability of failure for the k-th component, that we define as *the model parameter*. Model parameters are not known, but are assumed to be time independent and similar among components. Failures of components are perfectly observed and repaired. We developed this toy application being inspired by the textbook of Gelman et al. (2004). An analogous problem can be formulated for coins with similar biases, where $\theta_{k,2}$ represent the probability of the k-th coin to land Head, and Y indicates the outcome of the tossing. Adapting Eq. (14), distributions for variables in Figure 19 are assigned as:

$$\begin{aligned} & \alpha \sim \text{Exponential}(\lambda) && \boldsymbol{\beta} \sim \text{Dirichlet}(\boldsymbol{\eta}) \\ \forall k = 1, \dots, K \quad & \boldsymbol{\theta}_k \sim \text{Dirichlet}(\alpha\boldsymbol{\beta}) && Y_{k,l} \sim \text{Multinomial}(\boldsymbol{\theta}_k) \end{aligned} \tag{16}$$

It is to be noted that, as variables Ys are binary, the Dirichlet is equivalent to a Beta distribution, and the Multinomial is equivalent to a Binomial distribution, so the formulation belongs to the so-called hierarchical Beta-Binomial case (Gelman et al. 2004). As before, the posterior distribution cannot be evaluated in closed-form and, in the next section, we will present the approximate numerical scheme adapted from Sections 4.2.2.2-4.2.2.4.

4.2.3.2 Sampling Algorithm

As outlined in Section 4.2.2.2, the MCMC procedure we propose alternates sampling the model parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\}$ having fixed the hyper-parameters (α and $\boldsymbol{\beta}$) and sampling the hyper-parameters having fixed the model parameters. This latter step is executed following the M-H scheme reported in Figure 18, with obvious re-assignment of the random variables ($\alpha \to \alpha_T, \boldsymbol{\beta} \to \boldsymbol{\beta}_T, \boldsymbol{\theta} \to \mathbf{T}, \lambda \to \lambda_T, \boldsymbol{\eta} \to \boldsymbol{\eta}_T$). It is worth describing briefly the former step. Let us group the observations collected on the k-th component as $\mathbf{Y}_k = \{Y_{k,1}, \dots, Y_{k,n}\}$, and define counting variable $d_k = \sum_{t=1}^{n} Y_{k,t}$ and vector $\mathbf{d}_k = [n - d_k \quad d_k]$. Because of the well-known properties of the Dirichlet-Multinomial (or Beta-Binomial), for any value α and $\boldsymbol{\beta}$ of the hyper-parameters, the conditional distribution $p(\boldsymbol{\theta}_k \mid \mathbf{Y}_k, \alpha, \boldsymbol{\beta})$ is in the Dirichlet family, and it is defined by parameter $(\alpha\boldsymbol{\beta} + \mathbf{d}_k)$ (Same as PLUS learning algorithm presented in Section 4.1). Consequently, generating samples from that distribution is computationally easy. The reader should note that the PLUS method and the procedure outlined in Section 4.2.2 are based on similar properties.

4.2.4 Numerical Validation of the Illustrative Example

4.2.4.1 Alternative Processing Approaches Used for Comparison

To investigate the performance of the proposed MU-POMDP framework to the problem outlined in the previous section, we compare its performance with two alternative approaches which follow the assumptions of Global and of Individual PLUS respectively. Figure 20 reports the graphical models for these approaches, and the reader should compare these with that of MU-POMDP, as reported in Figure 18, to appreciate the differences.

The scheme for *Global PLUS* is reported in Figure 20a. This approach models all components as controlled by a single global model, that is defined by $\boldsymbol{\theta}$. Consequently, the model cannot accommodate any discrepancy among the parameters, and all observations are on the same level, for the sake of inferring $\boldsymbol{\theta}$. Figure 20b shows the *Individual PLUS* approach, which assigns an independent model to each component. Consequently, observations collected on one component are completely irrelevant for inferring the model of other components. For both approaches, we include a layer of hyper-parameters, consistently with the MU-POMDP approach. The reader should be aware that, for any practical implementation of *Global PLUS* or *Individual PLUS*, it would be easier to define directly a prior on the model paramaters, without making use of any hyper-parameter. For example, the choice of a fixed Dirichlet prior would permit to describe the posterior distribution of the model parameters in close form, without any need for sampling. However, in this section we make use of the additional layer of hyperparameters in order to achive fair comparison between MU-POMDP and the alternative approaches: making use of the same value for λ, $\boldsymbol{\eta}$ and for the conditional distributions for hyper parameters and model parameters, we get the same marginal distribution for the model parameters among all three approaches. The core of the differences across approaches is capture

by the joint distribution of models: models are marginally independent under *Invidual PLUS*, identical under *Global PLUS*, while they can be similar but not identical under MU-POMDP.

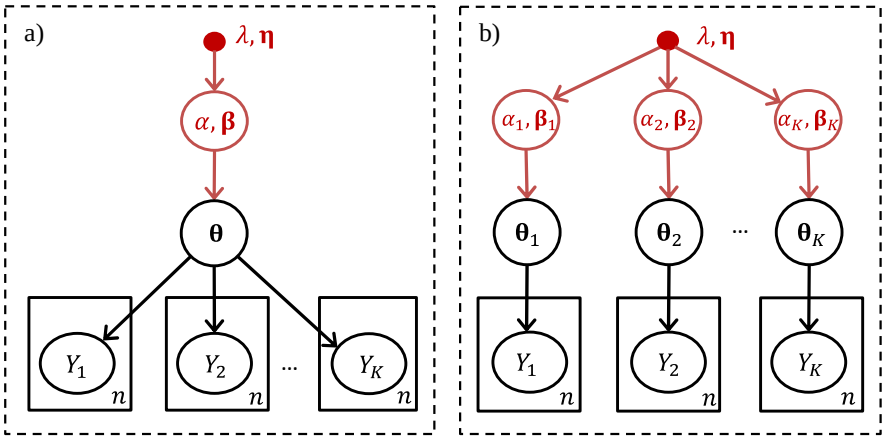

Figure 20. Graphical model for (a) Global PLUS, and (b) Individual PLUS.

4.2.4.2 Parameters for Numerical Investigation

To investigate the performance of MU-POMDP, and compare it with the alternative approaches, we consider a 5-component system and assign the following values: $\lambda = 1/1000$ and $\boldsymbol{\eta} = [\eta_1 \quad \eta_2] = [47.5 \quad 2.5]$. At this point, we can give a further insight about the relation between the choice of these values and the corresponding joint probabilities of the models. The expected value of the probability of failure $\theta_{k,2}$ is $\eta_2/(\eta_1 + \eta_2)$, that turns out to be 5% for this choice. It is hard to derive other direct relations between those parameters and features of the distribution. However, we observe that, for very high values of λ, α tends to be of high magnitude, and so the hyper-parameters of the Dirichlet distribution controlling the model $\boldsymbol{\theta}$: consequently, parameters for different components are highly correlated. On the other hand, the uncertainty in the distribution of $\boldsymbol{\beta}$ is decreasing with $(\eta_1 + \eta_2)$, so that $\boldsymbol{\beta}$ tends to be a fixed quantity when this quantity goes to infinite.

Figure 21a shows the marginal distribution for model parameter $\theta_{k,2}$ for any k-th component. The standard deviation is 3.5%. Figure 21b shows the contour plot of the joint distribution of any pair of variables $(\theta_{k,2},\theta_{l,2})$ for $k \neq l$, according to the MU-POMDP approach. As expected, random variables are dependent, and the correlation coefficient turns out to be 75%. In a nut-shell, the alternative approaches can be intended as alternative prior joint distributions. According to the *Global PLUS*, all model parameters are identical so the joint distribution collapses on the identity line $(\theta_{1,2} = \cdots = \theta_{K,2})$. According to *Individual PLUS*, the joint probability is the product of marginal distributions, as variables are independent (and therefore uncorrelated).

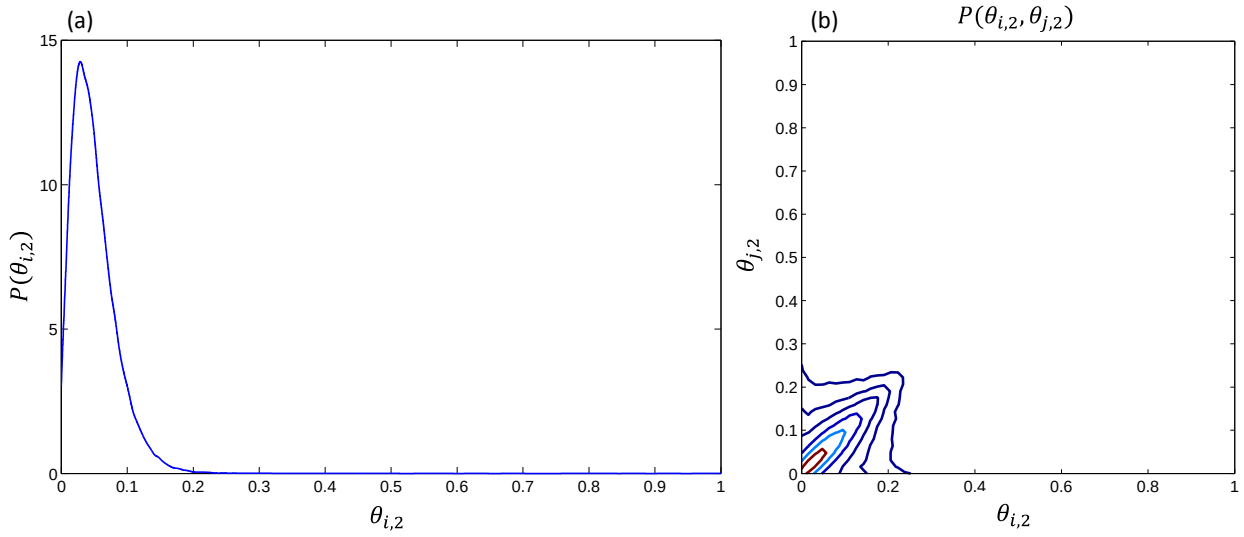

Figure 21. (a) Marginal prior density on the model parameters for each component (b) Joint prior density of the model parameters for any pair of components in the MU-POMDP framework.

4.2.4.3 Scheme for the Numerical Investigation

We assume MU-POMDP captures the correct model for all variables, and investigate (i) if the numerical procedure proposed is effective, and (ii) how approximate approaches perform. We adopt the Kullback-Leilber (KL) divergence (Cover and Thomas 2006) as a metric to assess

the performance of all approaches. The KL divergence is a non-symmetric measure of the difference between two probability distributions. Specifically the KL divergence of distribution Q from distribution P, denoted as $KL(P,Q)$, is a measure of information lost when Q is used to approximate P. In this context, suppose $\boldsymbol{\theta}_k^*$ indicates the actual model parameters for the k-th component, and $\boldsymbol{\Theta}^* = \{\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_K^*\}$ the set of corresponding values for all components in the system. An agent knowing $\boldsymbol{\Theta}^*$ exactly would predict the states (i.e. the observations) at the next time step for all components $\mathbf{y} = \{Y_1, \dots, Y_K\}$ with distribution $P(\mathbf{y}|\boldsymbol{\Theta}^*)$. Obviously, any previous observation will be irrelevant for such an agent. On the contrary, agents without perfect information on model parameters will base their prediction on inference: distribution $P(\mathbf{y}|\mathbf{Y}, \mathcal{M})$ indicates the posterior probability respect to all previous observations $\mathbf{Y}$, assuming model $\mathcal{M}$ which, depending on the agent, can be MU-POMDP, *Global* or *Individual PLUS*. The KL divergence between the distribution adopted by agents is therefore $KL[P(\mathbf{y}|\boldsymbol{\Theta}^*), P(\mathbf{y}|\mathbf{Y}, \mathcal{M})]$, and depends on $\boldsymbol{\Theta}^*$ and $\mathbf{Y}$. Treating these latter quantities as random variables, we can define an expected error e, as:

$$e(n, \mathcal{M}) = \mathbb{E}_{p(\boldsymbol{\Theta}^*)} \mathbb{E}_{P(\mathbf{Y}|\boldsymbol{\Theta}^*)} KL[P(\mathbf{y}|\boldsymbol{\Theta}^*), P(\mathbf{y}|\mathbf{Y}, \mathcal{M})] \quad (17)$$

where $\mathbb{E}_p$ indicate the statistical expectation respect to distribution p, and n indicated the number of observations collected per component (as indicated in Figs.19-20). This error measures the lack of information related to the use of model $\mathcal{M}$ in processing measures $\mathbf{Y}$, respect to observing directly $\boldsymbol{\Theta}^*$, for the sake of predicting $\mathbf{y}$. The analytical definition of KL reads, in this context:

$$KL[P(\mathbf{y}|\boldsymbol{\Theta}^*), P(\mathbf{y}|\mathbf{Y}, \mathcal{M})] = \mathbb{E}_{P(\mathbf{y}|\boldsymbol{\Theta}^*)}\left[\log \frac{P(\mathbf{y}|\boldsymbol{\Theta}^*)}{P(\mathbf{y}|\mathbf{Y}, \mathcal{M})}\right] \quad (18)$$

where the prediction using model $\mathcal{M}$ can be related to the inference on model parameters as:

$$P(\mathbf{y}|\mathbf{Y}, \mathcal{M}) = \mathbb{E}_{p(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})}[P(\mathbf{y}|\boldsymbol{\Theta})] \quad (19)$$

where $\boldsymbol{\Theta}$ indicates the set of model parameters, for all components in the system. Note that the probability of outcome $\mathbf{y}$ given parameters $\boldsymbol{\Theta}$ is:

$$P(\mathbf{y}|\boldsymbol{\Theta}) = \prod_{k=1}^{K} \theta_{k,1}^{y_k} \theta_{k,2}^{1-y_k} \quad (20)$$

Adopting a Monte Carlo approach, we can approximate any expectation with arithmetical average across samples. We start generating samples of variable $\boldsymbol{\Theta}^*$, from distribution $p(\boldsymbol{\Theta}|\mathcal{M} = \text{MU\_POMDP})$, which is represented in Figure 21. Then, in sequence, we generate samples of $\mathbf{Y}$ from distribution of $P(\mathbf{Y}|\boldsymbol{\Theta}^*)$. For each sample of $\mathbf{Y}$, we can sample $\boldsymbol{\Theta}$ from posterior distribution $p(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})$, following the inference procedure outlined in Section 4.2.3.2, for each model $\mathcal{M}$. Functions $P(\mathbf{y}|\boldsymbol{\Theta})$ and $P(\mathbf{y}|\boldsymbol{\Theta}^*)$ can be evaluated analytically using Eq. (20) on its entire domain: the number of possible outcome $\mathbf{y}$ is 2^K for a system made up by K components.

As we assume MU-POMDP to be the generative models for the validation, the agent adopting MU-POMDP is consistent. For this agent, therefore, we can drop the notion of "actual parameters" $\boldsymbol{\Theta}^*$, use $\boldsymbol{\Theta}$ in Eqs. (17) and (18), and get error e as a quantity defined internally to the graphical model reported in Figure 19, as:

$$e(n, \mathcal{M} = \text{MU\_POMDP}) = \mathbb{E}_{p(\boldsymbol{\Theta},\mathbf{y})}\{\log P(\mathbf{y}|\boldsymbol{\Theta}) - \mathbb{E}_{P(\mathbf{Y}|\boldsymbol{\Theta})}[\log \mathbb{E}_{p(\boldsymbol{\Theta}|\mathbf{Y})} P(\mathbf{y}|\boldsymbol{\Theta})]\} \quad (21)$$

In this light, error e resembles the concept of "regret", according to the definition of Raiffa and Schlaifer (1961), related to random variable $\boldsymbol{\Theta}$.

It is worth describing in details how different agents consider the collected observations, for the sake of inferring the model parameters. According to the MU-POMDP formulation, observations can be partitioned in two subsets. As shown in Figure 19, observations $\mathbf{Y}_k$, collected on component k, are particularly useful to infer model parameters $\boldsymbol{\theta}_k$, and we can call them "direct measures". On the other hand, observations $\mathbf{Y}_{l\neq k}$, collected on all components except the k-th one, are also useful for inferring $\boldsymbol{\theta}_k$, but only via the hyper-parameters α and $\boldsymbol{\beta}$, and we call them "indirect measures". In the limit for K and n going to infinite, the set of indirect measures is equivalent to a perfect observation of the hyper-parameters. This, however, would not allow to get a perfect prediction of $\boldsymbol{\theta}_k$. On the other hand, for n going to infinite the direct measures correspond to observing $\boldsymbol{\theta}_k$ directly.

As shown in Figure 20, the two PLUS approaches do not apply the distinction between direct and indirect measures. *Global PLUS* put all measures on the same level, for the sake of inferring $\boldsymbol{\theta}$. Suppose we are interested in inferring the model parameters for component k: *Global PLUS* makes use of all measures collected on the system, without giving any higher relevance to $\mathbf{Y}_k$ respect to $\mathbf{Y}_{l\neq k}$. On the other hand, *Individual PLUS* treats $\boldsymbol{\theta}_k$ and $\mathbf{Y}_{l\neq k}$ as independent variables, so observations collected on different components are discharged as irrelevant. Intuitively, indirect observations can be beneficial in many applications, especially for similar components and for a small value n of observations per component. In this context we expect *Global* to be more effective than *Individual PLUS*, because of the its capability of using all data. However,

Individual is more flexible than *Global PLUS*, as it can accommodate discrepancies among the model parameters of different components. Therefore, for high values of n we expect *Individual* to perform better than *Global PLUS*. MU-POMDP is supposed to capture the pros of both alternative methods.

4.2.4.4 Results of the Numerical Investigation

The following values have been assumed for the M-H steps (the reader is referred to Figure 18): number of steps $J = 20$, concentration $c_\beta = 600$, random step size $\sigma_\alpha = 0.1$. Furthermore, the number of cycles in the MCMC approach, as reported in Figure 16 is 1000, and the first 300 are discharged for the burn-in phase: therefore expectation in Eq. (19) is approximated by average among 700 samples. The expectation in Eq. (17) is approximation by 400 samples of "true models".

Figure 22 reports an example of outcome of the inference process, for the MU-POMDP framework. Pictures (a-b) show the joint domain of parameters ($\theta_{1,2}$,$\theta_{2,2}$) as in Figure 20b. The red star locates the value assumed as correct, and used for generating observations. The blue dots (a) and (b) show the samples generated from the posterior distribution for number of observation n equal 5 and 500 respectively. (c) and (d) report similar outcomes for the hyper-parameters (α and $\boldsymbol{\beta}$). Note that the pair (α, β_1) is sufficient to represent the entire domain, as the second component of $\boldsymbol{\beta}$ can be derived as $\beta_2 = 1 - \beta_1$. As expected, the posterior distribution becomes more skewed as more observations are processed. The figure shows an appropriate behavior of the MCMC procedure. However, as well known (MacKay 2003), the tuning of the procedure requires careful selections of its parameters, to get an appropriate rejection rate.

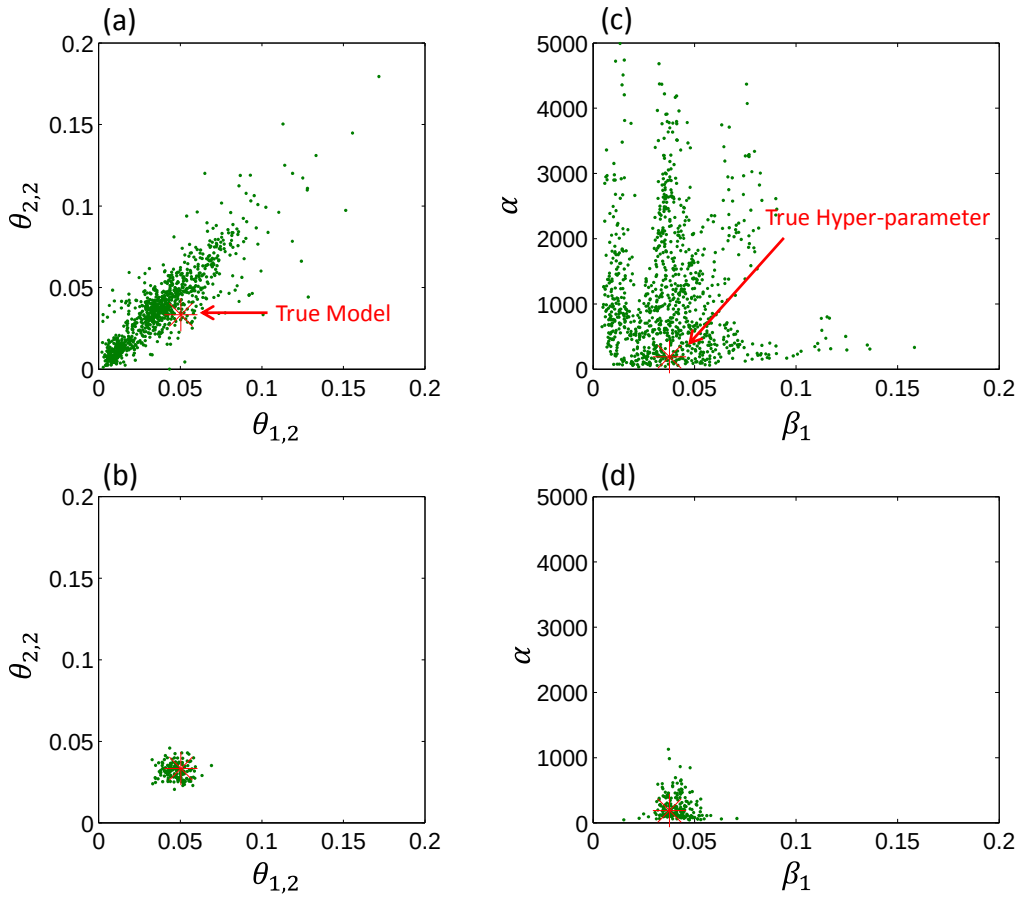

Figure 22. An example of outcome of the inference process, for the MU-POMDP framework. Samples generated from the posterior distribution of model parameters (a,b) and hyper-parameters (c,d) for (a,c) 5, and (b,d) 500 observations per component.

Figure 23 shows the outcomes of the comparison among approaches. Error e is plotted *vs* n for the three approaches. Error in the computations of e derives from approximation of expected values by samples, in Eqs. (17) and (19). We can easily estimate the confidence bound related to the approximation of Eq. (17), via computing the sample variance. Dashed lines in Figure 23 report the 95% confidence bounds. In those bounds, however, the error related to the approximation of Eq. (19) has not been included. As it can be seen in the figure, as the number of observations per components leans to infinity ($n \rightarrow \infty$), both MU-POMDP and *Individual PLUS* converge to the true model parameters while *Global PLUS* does not. This happens because true models exhibit variability among components, while *Global PLUS* assumes all models to be

identical: we except this approach to converge to the average of the components' models, and the residual errors do not vanish. The learning rate of *Individual PLUS* is lower than those of both *Global PLUS* and MU-POMDP, since it does not make use of the indirect observations. For the specific application, *Global* performs better than *Individual PLUS* up to about 200 observations per component, because of the effect of indirect measures. Up to that level, *Individual* performs better, due to its flexibility. As expected, MU-POMDP performs better than both, since it is the correct generative model for the data. As shown by the confidence bounds, the outcomes are affected by large numerical uncertainty, due to the high number of dimensions in the nested expectation defined in Eqs. (17-19).

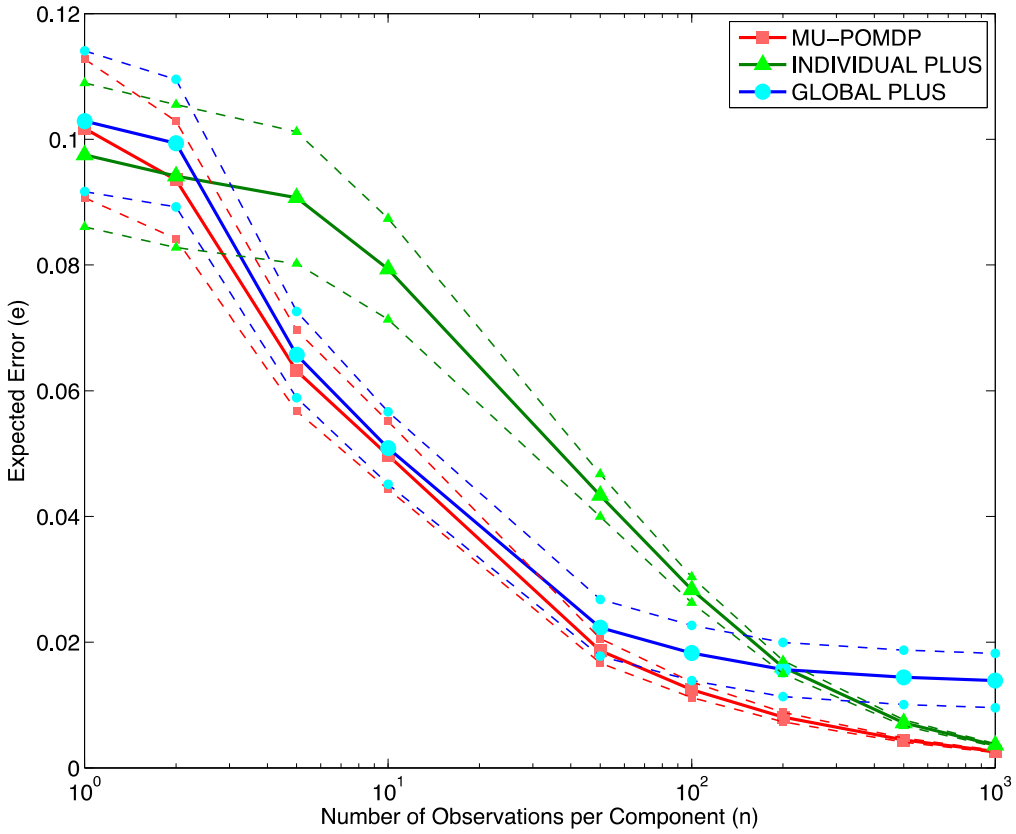

Figure 23. Comparison between MU-POMDP, Global PLUS and Individual PLUS performances in learning the model parameters.

4.2.5 Application – Wind Farm Management

In this section we evaluate the performance of proposed MU-POMDP methodology in an example of wind farm management, by adapting the setting investigated in Chapter 3 and Section 4.1. We have selected the prior parameters over the transition and emission probabilities based on literature on wind farm maintenance (Byon et al. 2010, Byon and Ding 2010, Nielesen and Sorensen 2012, Memarzadeh et al. 2013, Memarzadeh et al. 2015a) and discussion with our industry collaborator Everpower Wind Holdings.

4.2.5.1 Parameters of Numerical Investigation

For the purpose of validation, we consider a wind farm made up by 5 turbines of the same type placed in similar environmental conditions. The state condition of each turbine is discretized into three possible states where $s = 1$ refers to an intact structure, $s = 2$ to a damaged one, and $s = 3$ to the failure of the turbine; the agent receives observations from a set of four possible observations where $z = 1$ suggests that the turbine is undamaged, $z = 2$ and $z = 3$ indicate two symptoms of damage, and $z = 4$ indicates the failure of the turbine; three actions are available: Do-Nothing (DN), Repair (RE), and Visual Inspection (VI). When the agent chooses DN, the condition state of the turbine degrades owing to fatigue and aging, potentially causing a structural failure and a relevant economical loss. In turn, the agent can perform a costly intervention (i.e., RE) to avoid failure and improve the condition state of the turbines. VI better measures the condition state of the turbine (that evolves according to the degradation model, as for DN). Each time step is assumed to be six months, and the agent takes one action per turbine at each time step.

Table 1. Prior parameters over hyper-parameters for management of wind farm example.

$$\lambda_T = \lambda_O = {}^{1}/_{1000}$$

$$\boldsymbol{\eta}_{T,\mathrm{DN,VI}} = \kappa \times \begin{bmatrix} 0.57 & 0.28 & 0.15 \\ 0 & 0.67 & 0.33 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\eta}_{T,\mathrm{RE}} = \kappa \times \begin{bmatrix} 0.67 & 0.33 & 0 \\ 0.67 & 0.33 & 0 \\ 0.67 & 0.33 & 0 \end{bmatrix}$$

$$\boldsymbol{\eta}_{O,\mathrm{DN,RE}} = \kappa \times \begin{bmatrix} 0.57 & 0.28 & 0.15 & 0 \\ 0.15 & 0.57 & 0.28 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\eta}_{O,\mathrm{VI}} = \kappa \times \begin{bmatrix} 0.67 & 0.33 & 0 & 0 \\ 0.33 & 0 & 0.67 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Table 1 shows the prior parameters over hyper-parameters; subscripts report the action symbol, κ controls the skewness of the prior and has been fixed to 50, so that the corresponding average coefficient of variation of the parameters is 0.26 (the average is computed only on uncertain entries of the prior parameters resulting from Table 1). Parameter λ controls the correlation among the model parameters across components: as λ decreases, the correlation increases, and it is about 75% given the values reported above. Entries in square brackets define the expected value of transition and emission probabilities: for example, the expected value of the probability that the undamaged turbine becomes damaged under DN is 28%. The costs for repair, visual inspections and down-time due to failure are assumed to be US $25,000, $500, and $50,000, respectively. The discount factor is assumed to be $\gamma = 0.95$. The initial belief state for all turbines is defined as $\mathbf{b}_0 = [0.8 \quad 0.2 \quad 0]$, which means that the agent believes that, at the beginning of the process, the turbines are in the intact state with 80% probability and in damaged state with 20% probability.

4.2.5.2 Scheme for Numerical Investigation

To investigate the performance of MU-POMDP, we simulate the response of a system characterized by model $\boldsymbol{\Theta}^* = \{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \dots, \boldsymbol{\theta}_K^*\}$, where $\boldsymbol{\theta}_k^* = \{\mathbf{T}_k^*, \mathbf{O}_k^*\}$ defines transition and emission probabilities for component k. We evaluate the effectiveness of both learning and planning.

For learning, we evaluate the effectiveness of MU-POMDP in term of accuracy in predicting future observation for components on the system. At time step t, the probability distribution of next observation for the entire farm is defined as $P(z_{t+1}|\boldsymbol{\Theta}^*, \bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t)$, where $z_t = \{z_{1,t}, z_{2,t}, \dots, z_{K,t}\}$, $\bar{Z}_t = \{\bar{Z}_{1,t}, \bar{Z}_{2,t}, \dots, \bar{Z}_{K,t}\}$, $\bar{A}_t = \{\bar{A}_{1,t}, \bar{A}_{2,t}, \dots, \bar{A}_{K,t}\}$ and can be computed as follow (similar to Eq. (19)):

$$P(\mathbf{z}_{t+1}|\bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t) = \mathbb{E}_{p(\boldsymbol{\Theta}|\bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t)}[P(\mathbf{z}_{t+1}|\bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t, \boldsymbol{\Theta})] \tag{22}$$

The expectation in Eq. (22) can be approximated via Monte Carlo. Error in the prediction can be measured by the KL divergence:

$$\varepsilon(\boldsymbol{\Theta}^*, \bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t) = \mathrm{KL}[P(\mathbf{z}_{t+1} \mid \boldsymbol{\Theta}^*, \bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t), P(\mathbf{z}_{t+1} \mid \bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t)] \tag{23}$$

Function $\varepsilon(\boldsymbol{\Theta}^*, \bar{\mathbf{Z}}_t, \bar{\mathbf{A}}_t)$ depends on the realization of model, actions, and observations. Despite expected value can be taken, in this paper we validate the effectiveness of MU-POMDP on a specific realization. To do so, we have sampled farm model $\boldsymbol{\Theta}^*$ from the MU-POMDP priors outlined in section 4.2.5.1, and actions $\bar{\mathbf{A}}_t$ and observations $\bar{\mathbf{Z}}_t$ consequently.

4.2.5.3 Results of Numerical Investigation

We evaluate the effectiveness of learning for $t = 35, 70, 100, 500,$ and 1000 (values $t = 500$ and 1000 allow us to investigate the long term behavior or the learning process). Figure 24 reports the error in the prediction of next observation for MU-POMDP framework.

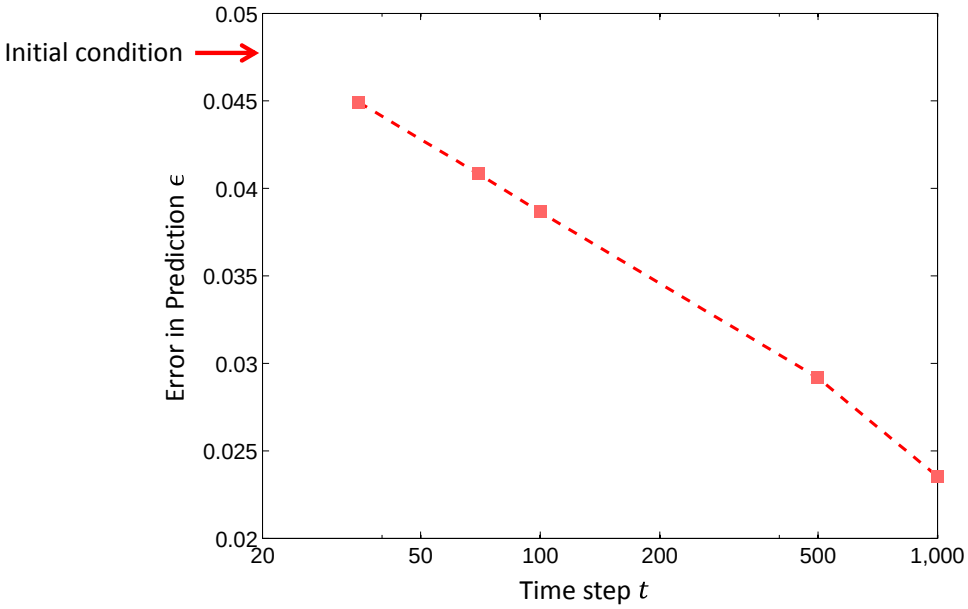

Figure 24. MUPOMDP performance in predicting the future observation as a function of number of observations received.

As shown in the figure, the error in predicting the future observation is decreased by factor 18% with only 100 data and by 50% with 1000 data. In the limit of infinite data, MU-POMDP's error in prediction of future observation should converge to zero as it learns the true model parameters accurately.

Figure 25 shows the examples of the inference process, plotting samples for one entry in the transition matrix (a-c) and emission matrix (d-f) under action DN, for components 1 and 2. The red star shows the value used for simulating the data, while the green points shows the samples generated from (a-d) the prior distribution, (b-e) MU-POMDP's posterior at $t = 70$, and (c-f)

posterior at $t = 1000$. Coefficient of variation of the posterior distribution is 0.18 (i.e., it is decreased by factor 31.1% respect to the prior distribution) after receiving 70 data (25b), and 0.14 (decreased by 47.16%) after receiving 1000 data (25c) for the specific element of transition matrix plotted in Figure 25. In the case of emission value plotted in the figure, coefficient of variation has decreased by 11.2% and 35.3% after receiving 70 (25e) and 1000 (25f) data respectively. The reader should note that only a fraction of the observations are useful for updating any specific parameter. For example, consider parameter $T(s_t = 1, a_t = DN, s_{t+1} = 1)$, i.e. the probability of next state being intact given that the current state is intact and agent performs *Do-Nothing*. First, no observation collected after any action except DN is relevant; second, only transitions starting from state 1 are relevant. Actually, given the stochastic approach for leaning, we cannot assess with certainty whether current state is 1 or not, at any time. However, we can count the occurrence of this event in each realization of state trajectory generated through FFSB. For the parameter mentioned above, we estimate that about 300 out of 1000 observations are relevant.

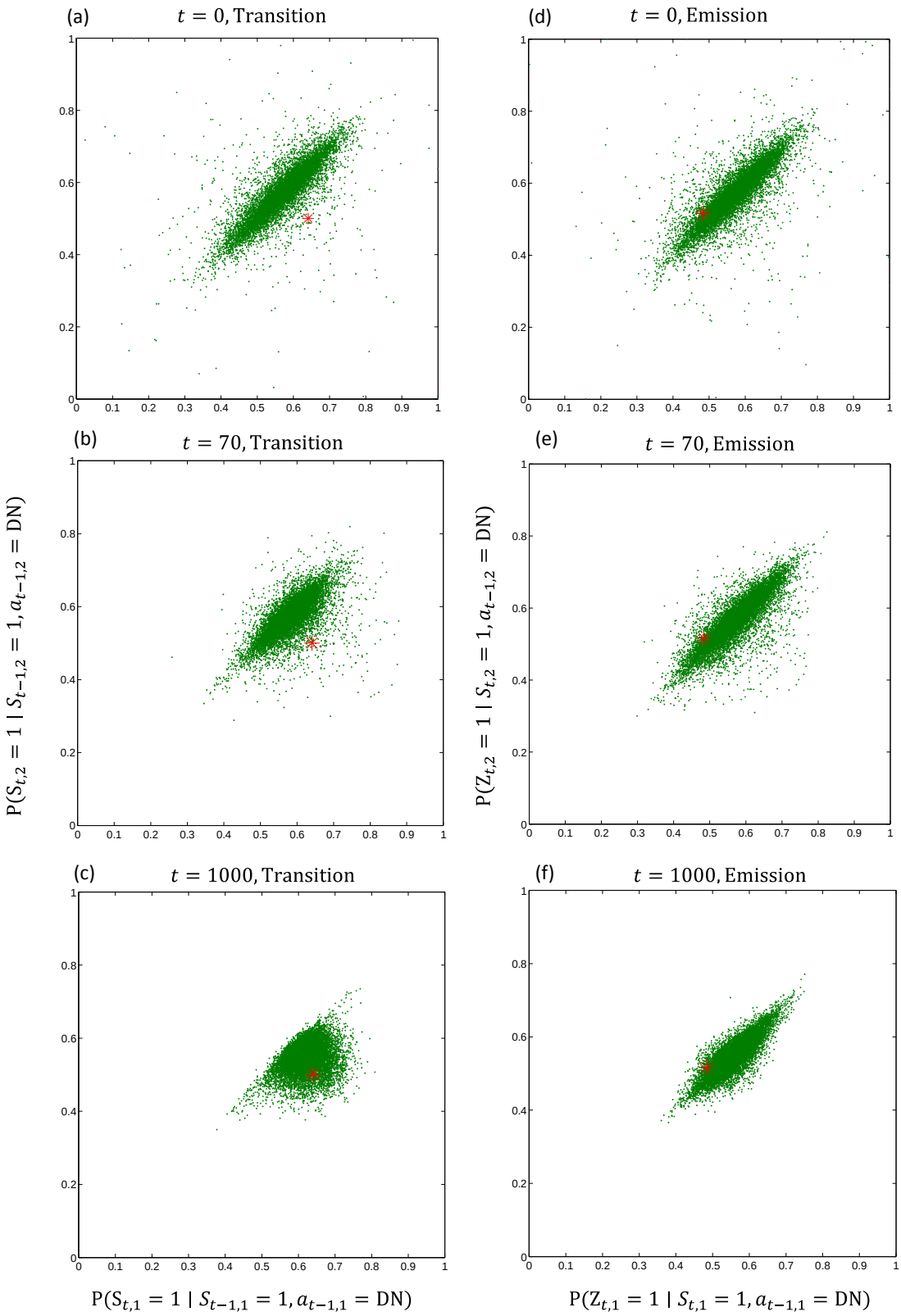

Figure 25. Examples of samples of model parameter (green dots) and exact value (red star) for MU-POMDP and PLUS.

In the final numerical campaign, we investigate the economic impact of adopting the MU-POMDP framework, showing how the more accurate learning algorithm, which accounts for discrepancies in the component models, allows for a more effective planning phase.

Figure 26 shows the cumulative cost (i.e. the negative reward) of operation and maintenance for the wind farm as a function of the time step for (1) an agent with perfect knowledge about the actual model parameters (True Model: black line), (2) an agent following MU-POMDP (MU-POMDP: red line), and (3) an agent adopting a POMDP fixed to the expectation of the prior distribution, without any learning (POMDP: blue line). Estimates are based on 100 independent simulations in the time domain, and MU-POMDP agent (red line) learns after receiving 35 and 70 data. The agent with perfect knowledge about the true model represents a lower limit for the cost (i.e. an upper bound for negative rewards). For this specific example, the value (i.e. the sum of discounted costs) for true model agent is \$56.25K per turbine, while for the MU-POMDP and POMDP agents are \$60.78K and \$63.4K per turbine respectively. We evaluate the economic benefit of using MU-POMDP framework over POMDP by computing the average reward after processing 70 observations (between $t = 70$ and $t = 100$ steps) and it is quantified as \$356.06 per time step per turbine. Of course, these values depend on the specific numerical example that have been chosen and might change with different application.

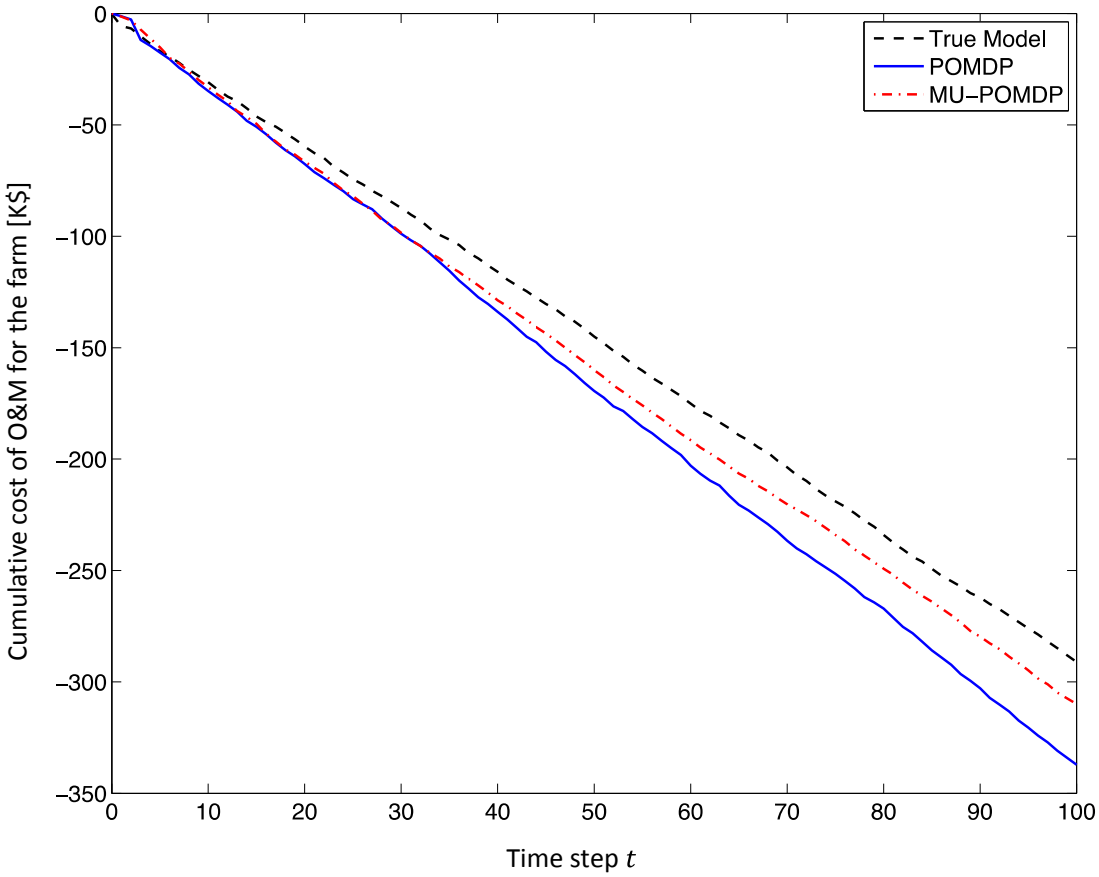

Figure 26. Cumulative O&M cost of the farm consists of five turbines for the agent knowing the true model (black), POMDP (blue), and MU-POMDP (red)

Chapter 5

Sequential Decision Making: Value of Information

Abstract

Operation and maintenance of an infrastructure system rely on information collected on its components, which can provide the decision maker with an accurate assessment of their condition states. While the methods developed in previous chapters allow for optimal information gathering, they cannot incorporate system-level constraints on resources available for this task. In this chapter, we introduce the concept of value of information (VoI), that can be used as a guide for information-gathering and, as we'll illustrate in Chapter 6, for system level inspection scheduling. In this chapter, we show how to compute the VoI in two settings: the stochastic future allocation, that assumes observations are collected with a given probability, and the fee-based future allocation that assumes observations are available at a given cost. We illustrate how these models can be used for evaluate the value of a permanent monitoring system (value of flow of information) as well as a piece of information at current time step (value of current information).

5.1 Problem Statement

The POMDP framework allows for integrating uncertain observations in the sequential decision making, including exploratory and exploitative actions. The stochasticity in the belief evolution is essentially connected to that of collected observations. In this light, each action is characterized by its expected cost and its effect on the belief evolution. Among exploitative actions, repairing can be expensive but associated with an improvement in the belief, while doing nothing may be cheaper but related to a degradation of the belief. Exploratory actions may include performing visual inspection, or collecting additional information that, while expensive, affects the belief by reducing its uncertainty. Effects of the installation of a permanent monitoring system, or of exceptional information, on the overall cost of operation and maintenance can be quantified by assessing the corresponding value of information (VoI).

5.2 Proposed Method

5.2.1 Value of Flow of Information

We assume to model the management of an infrastructure component as a POMDP. We start investigating the impact of receiving additional observations at all steps of the process, that we call a "flow of information" (Memarzadeh and Pozzi 2015d). This can happen, for example, when a monitoring system is installed, or when a component can be systematically inspected. We define h_t the additional observation of at time t, on discrete domain $H = \{1,2, \dots, |H|\}$. The relation between this observation and state s_t is modeled by emission function $E(j,i) = \mathbb{P}[h_t = j | s_t = i]$, summarized in matrix $\mathbf{E}$ or size $|S| \times |H|$. The prediction of observation h given the belief and the updating of this belief given the observation that h takes value j are

defined by emission and updating operators $\mathbf{e}^{\mathrm{I}}$ and $\mathbf{u}^{\mathrm{I}}$, of dimension $|H|$ and $|S|$ respectively, whose entries are as follows:

$$\begin{cases} e_j^{\mathrm{I}}(\mathbf{b},\mathbf{E}) = \mathbb{P}[h_t = j|\mathbf{b}_t = \mathbf{b},\mathbf{E}] = \sum_{l=1}^{|S|} E(l,j)b(l) \\ u_i^{\mathrm{I}}(\mathbf{b},j,\mathbf{E}) = \mathbb{P}[s_t = i|h_t = j,\mathbf{b}_t = \mathbf{b},\mathbf{E}] = \dfrac{E(i,j)b(i)}{\sum_{j=1}^{|H|} E(i,j)b(i)} \end{cases} \quad (24)$$

The Value of a flow of can be derived by comparing the value with different emission probabilities, as illustrated below.

5.2.1.1 Stochastic Allocation Model for Flow of Information

We generalize the problem by assuming that the availability of additional observation is uncertain (Memarzadeh and Pozzi 2015d, 2015e): at each time step, the observation is independently available only with probability P, that we call "availability". This can model possible random malfunctions of the monitoring system occurring with probability $(1-P)$. Computationally, to model this, we include an additional dummy outcome for variable h, indicating that no observation is available, with a flat emission on the state domain. The augmented emission matrix $\mathbf{E}_{(P)}^{\mathrm{S}}$, of size $|S| \times |H| + 1$ is defined as:

$$\mathbf{E}_{(P)}^{\mathrm{S}} = P[\mathbf{E} \quad \mathbf{0}_{|S|\times 1}] + (1-P)[\mathbf{0}_{|S|\times|S|} \quad \mathbf{1}_{|S|\times 1}] \quad (25)$$

where $\mathbf{0}_{s\times v}$ and $\mathbf{1}_{s\times v}$ are matrix of zeros and ones respectively, of dimension s by v. We call this the *Stochastic Allocation* model (SA), and we use superscript S to indicate it. We integrate this emission with that related to POMDP observation as follows:

$$\mathbf{O}^{\mathrm{S}}_{a,(P)} = \mathbf{O}_a \times \mathbf{E}^{S}_{(P)} \tag{26}$$

where, $\mathbf{O}_a$ is the emission matrix related to action a, and × indicates cross product matrices' columns. Each column of this matrix refer to possible pair $\{z, h\}$ including the dummy outcome for h. By grouping matrix for all action we get emission $\mathbf{O}^{\mathrm{S}}_{(P)}$, of size $|S| \times |Z|(|H|+1) \times |A|$, that can be embedded in a new POMDP with parameter set $\mathbf{\Theta}^{\mathrm{S}}_{(P)} = \{\mathbf{T}, \mathbf{O}^{\mathrm{S}}_{(P)}, \mathbf{R}, \gamma\}$.

Figure 27a illustrates the decision graph of the SA model. Usually, managers consider a monitoring effort that includes also a measure of the initial state. To take this into account, we define an optimal value U* including this initial observation, making use of Eq. (24), as follows:

$$U^*(\mathbf{b}, \mathbf{\Theta}', \mathbf{E}') = \sum_{h=1}^{|H|} e^{\mathrm{I}}_h(\mathbf{b}, \mathbf{E}')\, V^*[u^{\mathrm{I}}(\mathbf{b}, h, \mathbf{E}'), \mathbf{\Theta}'] \tag{27}$$

where $\mathbf{E}'$ is the emission of initial observation and $\mathbf{\Theta}'$ the parameter set of the POMDP following that. The difference in value $\Delta \mathrm{V_f}$ between the decision graphs in Figure 5 and 27a can be expressed as:

$$\Delta V_{\mathrm{f}}(\mathbf{b}, \mathbf{\Theta}'', \mathbf{\Theta}', \mathbf{E}') = V^*(\mathbf{b}, \mathbf{\Theta}'') - U^*(\mathbf{b}, \mathbf{\Theta}', \mathbf{E}') \tag{28}$$

where $\mathbf{\Theta}''$ indicates the parameter set of the POMDP without the initial observation. Function ΔV_{f} assesses the benefit (if positive), of the graph in Figure 27a respect to that of Figure 5. Using previous equations, we assess the VoI of this flow under the SA model ($VoI_{\mathrm{f,S}}$) as follows:

$$VoI_{\mathrm{f,S}}(\mathbf{b}, P) = \Delta V_{\mathrm{f}}\left(\mathbf{b}, \mathbf{\Theta}, \mathbf{\Theta}^{\mathrm{S}}_{(P)}, \mathbf{E}^{\mathrm{S}}_{(P)}\right) \tag{29}$$

where $\mathbf{\Theta}$ is the set of parameters without the additional observations (as in the graph of Figure 5), while $\mathbf{E}^{\mathrm{S}}_{(P)}$ and $\mathbf{\Theta}^{\mathrm{S}}_{(P)}$ are computed as in Eqs. (25-26). As expected, the VoI is a

function of multiple values: the overall setting without the flow of information (described by set **Θ**), the availability and accuracy of the flow of measures (described by P and **E** respectively), and the initial belief b. We express the VoI as a function of b and P, for convenience of notation. We also note that $VoI_{\mathrm{f,S}}(b,0)$ is zero, while $VoI_{\mathrm{f,S}}(b,1)$ express the VoI when the monitoring system is fully reliable.

This VoI can be compared with the cost for installing and operating the monitoring system, and a rational agent should adopt the system only if its cost is below its value.

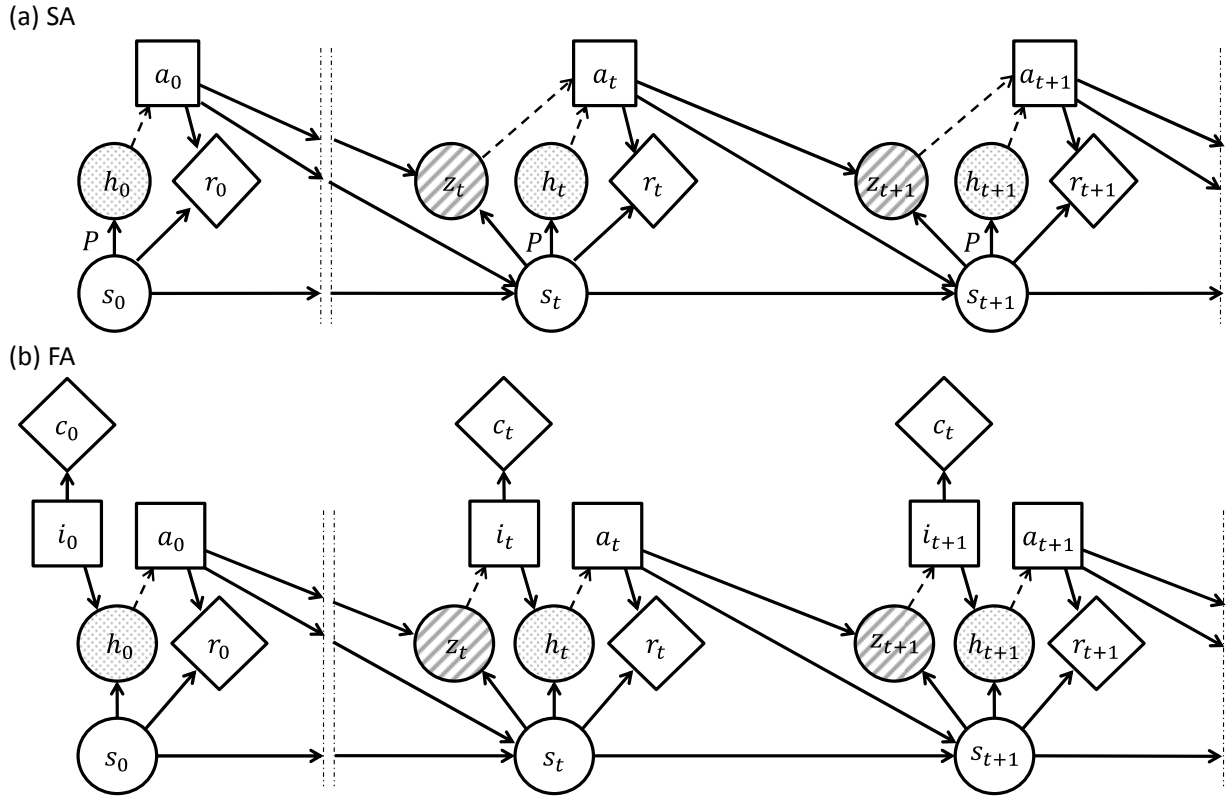

Figure 27. Decision graphs for the (a) SA model and (b) FA model.

5.2.1.2 Fee-based Allocation Model for Flow of Information

While SA model assumes that the agent has no to cover any additional cost for interrogating the monitoring system at time t. If, on the contrary, such a fee exists, the agent may choose to interrogate the system only if the belief raises concerns. The Fee-based Allocation model (FA) assumes that, at any time, variable h can be observed at non-negative cost C (Memarzadeh and Pozzi 2015d). At time t, therefore, the agent has to face two decisions in sequence: (i) in the *inspection* sub-step, binary decision $\mathrm{i_t}$ selects between *Inspect* and *Do NOT Inspect*. Only if inspection is performed, variable $\mathrm{h_t}$ is observed and cost C paid. (ii) after having processed the outcome of previous sub-step, in the *management* sub-step a management decision is taken. Figure 27b reports the decision graph according to the FA model, where cost c_t can be 0 or C depending on the decision i_t. It should be noted that, while we could easily introduce a not reliable observation, by using emission $\mathbf{E}^{\mathrm{S}}_{(P)}$, defined in Eq. (25), we prefer not to combine the FA and the SA models, for simplicity in the illustration.

To estimate the correspondent value, we have to re-formulate Eq. (5) using two sub-steps. Function $\mathrm{V}^{\mathrm{I}}_{(\mathrm{C})}$ is the value starting from the *inspection* sub-step, and function $\mathrm{V}^{\mathrm{M}}_{(\mathrm{C})}$ is that starting from the *management* sub-step, both defined on the same belief domain. Bellman Equation now reads as follows:

$$\begin{cases} V^{\mathrm{I}}_{(C)}(\mathbf{b},\boldsymbol{\Theta},\mathbf{E}) = \min\left\{ C + \sum_{h=1}^{|H|} e^{\mathrm{I}}_{h}(\mathbf{b},\mathbf{E})\, V^{\mathrm{M}}_{(C)}[\mathbf{u}^{\mathrm{I}}(\mathbf{b},h,\mathbf{E}),\boldsymbol{\Theta},\mathbf{E}]\,;\; V^{\mathrm{M}}_{(C)}(\mathbf{b},\boldsymbol{\Theta},\mathbf{E}) \right\} \\ V^{\mathrm{M}}_{(C)}(\mathbf{b},\boldsymbol{\Theta},\mathbf{E}) = \min_{a\in A}\left\{ r(\mathbf{b},a,\boldsymbol{\Theta}) + \gamma \sum_{z=1}^{|Z|} e_z(\mathbf{b},\mathrm{a},\boldsymbol{\Theta})\, V^{\mathrm{I}}_{(C)}[\mathbf{u}(\mathbf{b},a,\mathbf{z},\boldsymbol{\Theta}),\boldsymbol{\Theta},\mathbf{E}] \right\} \end{cases} \tag{30}$$

The decision among inspecting or not is defined by the minimization in the first line, where the first entry refers to inspect and the second to do not. Eq. (30) assumes that the underling

hidden state does not change between the inspection and the management sub-steps. To solve Eq. (30) requires some numerical implementation, and we illustrate in Appendix A how to re-formulate it as an equivalent single-step stationary POMDP, compatible with traditional solvers. Then VoI according to the FA model, $\mathrm{VoI}_{\mathrm{f,F}}$, is a function of fee cost C, and it can be computed as follow:

$$VoI_{\mathrm{f,F}}(\mathbf{b}, C) = V^{*}(\mathbf{b}, \boldsymbol{\Theta}) - V^{\mathrm{I}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}, \mathbf{E}) \tag{31}$$

To summarize, $\mathrm{VoI}_{\mathrm{f,F}}$ quantifies the benefit of being able to observe, at all times, variable $\mathrm{h_t}$ at cost C, before taking action $\mathrm{a_t}$, respect to not having this privilege. We note that $\mathrm{VoI}_{\mathrm{f,F}}(\mathrm{b}, \infty)$ is nil and $\mathrm{VoI}_{\mathrm{f,F}}(\mathrm{b}, 0)$ corresponds to $\mathrm{VoI}_{\mathrm{f,S}}(\mathrm{b}, 1)$, as the agent will never inspect if the fee is infinite and always is inspection is free.

5.2.2 Value of Current Information

In previous section, we address the evaluation of a flow of information, as that provided by a monitoring system. Sometimes, however, an agent has to take a decision about a current inspection (or information collection), not about a long-term monitoring effort. In this section, we investigate how to assess of VoI of short-term effort. This evaluation, however, cannot be independent of assumption about the availability of future information. The same piece of information can be relevant or redundant, depending on what other information will be available. Consider, as before, a management process modeled by a POMDP, as in Figure 5, and an inspection modeled by variable h, as in section 5.2.1 To assess the VoI related to h, we need to define: when will the component be inspected in the future? Unless we give a (probabilistic) answer to this question, the VoI is not well-defined. A possible assumption is that the component

will never be inspected again in the future: this is the pessimistic assumption mentioned in Chapter 6. An alternative one is that it will be always expected from next step: this is the optimist assumption (Chapter 6). Clearly, the VoI is different in these two cases. The application of these two cases to system-level inspection scheduling is investigated in Chapter 6. By using the SA and FA models, we can define more flexible assumptions.

5.2.2.1 Stochastic Future Allocation Model for Evaluating the Current Information

We can derive from the SA model an assumption on the future allocation of resources for information collection. Let us assume that the component will be inspected, from the next step, with probability P. To assess the corresponding VoI of current inspection, we define function ΔV_{c} as the VoI of observing h at current step, when the underlying POMDP is modeled by $\boldsymbol{\Theta}'$:

$$\Delta V_{\mathrm{c}}(\mathbf{b}, \boldsymbol{\Theta}', \mathbf{E}) = V^*(\mathbf{b}, \boldsymbol{\Theta}') - U^*(\mathbf{b}, \boldsymbol{\Theta}', \mathbf{E}) = \Delta V_{\mathrm{f}}(\mathbf{b}, \boldsymbol{\Theta}', \boldsymbol{\Theta}', \mathbf{E}) \tag{32}$$

According to the SA model, the POMDP is actually defined by $\boldsymbol{\Theta}^{\mathrm{S}}_{(P)}$, so that the corresponding VoI is:

$$VoI_{\mathrm{c,S}}(\mathbf{b}, P) = \Delta V_{\mathrm{c}}\left(\mathbf{b}, \boldsymbol{\Theta}^{\mathrm{S}}_{(P)}, \mathbf{E}\right) \tag{33}$$

where subscript c stands for “current”. This quantity can be intended as the difference between the values of two decision graphs differing one another only for the first step. Figure 28 reports the first step for the graph with and without inspection, while all future steps are modeled as in figure 27a.

In summary, Eq. (33) answers to the following question: “how much are we willing to pay for inspecting now, if future (free) inspections will be available with probability P?”

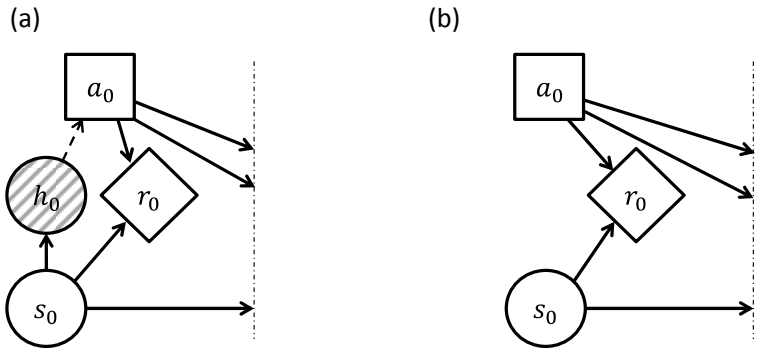

Figure 28. First step of the decision graph with (a) and without (b) inspection. The future steps are modeled as in Figure 27a.

5.2.2.2 Fee-based Future Allocation Model for Evaluating the Current Information

Similarly, we can base the VoI computation on an assumption related to the FA model. If we assume that the inspection can be repeated anytime in the future, at cost C, we can assess the VoI of current inspection as follows:

$$VoI_{\mathrm{c,F}}(\mathbf{b}, C) = V^{\mathrm{M}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}, \mathbf{E}) - V^{\mathrm{O}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}, \mathbf{E}) \tag{34}$$

where, as before, V^{M} is the value staring from a *management* sub-step, while value V^{O} is that starting by inspecting without paying any cost, obtainable as:

$$V^{\mathrm{O}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}, \mathbf{E}) = \sum_{h=1}^{|H|} e^{\mathrm{I}}_{h}(\mathbf{b}, \mathbf{E})\, V^{\mathrm{M}}_{(C)}[\mathbf{u}^{\mathrm{I}}(\mathbf{b}, h, \mathbf{E}), \boldsymbol{\Theta}, \mathbf{E}] \tag{35}$$

By confronting with Eq. (30), we can easily check that, in the setting of information flow, the agent inspects at current time only if $VoI_{\mathrm{c,F}}$ is above C.

Eq. (34) answers to the following question: “how much are we willing to pay for inspecting now, if future inspections will be available at cost C?”

5.3 Illustrative Example for Assessing VoI

We illustrate how the VoI depends on the assumed model and the decision making parameters by using a simple example. Let us consider a component whose state can assume $|S| = 3$ values, referring to Intact ($s = 1$), Damage ($s = 2$) and Failure ($s = 3$). Two maintenance actions are available to the agent: Do-nothing ($a = 1$, DN) and Replace ($a = 2$, RE). The transition probability table is reported in Table 2, where $\mathbf{T}_a$ indicated the sub-table referring to action a. If the agent does nothing, an intact component becomes damage with probability 0.5% and cannot fail directly, while a damage component cannot recover and can fail with probability 10%. If the agent replaces the component, the state becomes intact, independently on the current one. Without inspection, the only available observation discriminates between the failure and the first two states, but it does not between intact and damage: so the agent is aware of the failure as soon as it happens, but it receives no symptom of damage. Cost of repair is assumed to be \$10K and the cost of failure and downtime to be \$500K, while the discount factor is 95%.

Table 2. Matrices of the illustrative example.

$$\mathbf{T}_1 = \begin{bmatrix} 99.5\% & 0.5\% & 0 \\ 0 & 90\% & 10\% \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \mathbf{O}_{1-2} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \\ 0 & 1 \end{bmatrix}$$

In this context, the agent is evaluating a binary inspection that can detect damages according to the model reported in Table 2. The outcome of the inspection can be *alarm* or *silence*, and ϵ defines the inspector inaccuracy, both as probability of a "false alarm" (i.e. an alarm of a intact component) and of "false silence" (a silence on a damaged component), as we consider the two

probabilities to be the same for the easiness of illustration. The inspector may not be able to discriminate the failure with certainty but, as noted above, the other observation can. When ϵ is zero, the inspector is perfect and the problem becomes as MDP after current observation is taken; when ϵ is 50%, the outcome of the inspector is independent by the (not failed) state, and the inspector become useless.

Because of the emission matrix **O**, at any time the belief can assign only probability zero or one to the failure state: the agent cannot have uncertainty about the current failure of the component, and if the component is failed, the agent will replace it. Because of this, belief **b** can be completely described by the current probability of damage P_{DAM}. Figure 29a-b shows $VoI_{\mathrm{f,S}}$ as a function of belief P_{DAM}, for different values of availability P and inaccuracy ϵ.

It is to be noted that the VoI is not monotonically increasing with P_{DAM}: it is maximum where the uncertainty between doing-nothing and replacing the highest (around $P_{\mathrm{DAM}} = 3\%$), while for very high probability of damage the agent needs almost inevitably to repair and the impact of the inspections is relatively lower. Actually, VoI is the difference between two convex piecewise linear functions (Smallwood and Sondik 1973, Sondik 1978) and thus it is piecewise linear (but not necessarily convex); in Figure 29a, however, we plot it in the log-scale to highlight its behaviour for low probabilities, and this masks this feature. Also, VoI is monotonically increasing with P, as can be proven by the principle that (more) information never hurts (Heckerman et al. 1993). The relation between VoI and the accuracy of the inspector is illustrated in Figure 30a. Again, an expected, it is a monotonic relation: going from the value of a perfect inspector (for $\epsilon = 0$) to the nil value of an independent useless inspector (for $\epsilon = 50\%$). Graphs as that in Figure 30a allow us to compare accuracy and availability: the value of a perfect inspector available with probability 10% is almost equivalent to that of inspector with 30%

probability of false alarm, always available. It is not always possible to define a parameter able to capture the information accuracy (e.g. when false negative and false positive rates are different parameters). However, when this is possible (as in Pozzi and Der Kiureghian (2011), Madanat (1993)), the VoI is monotonically related to that.

Figure 29b reports the corresponding graph for the FA model (note that the graph for $C = 0$ is identical to that for $P = 1$ in Figure 29a). Again, we expected that the VoI is monotonically decreasing with fee cost C, and the monotonic relation with ϵ is illustrated in Figure 30b.

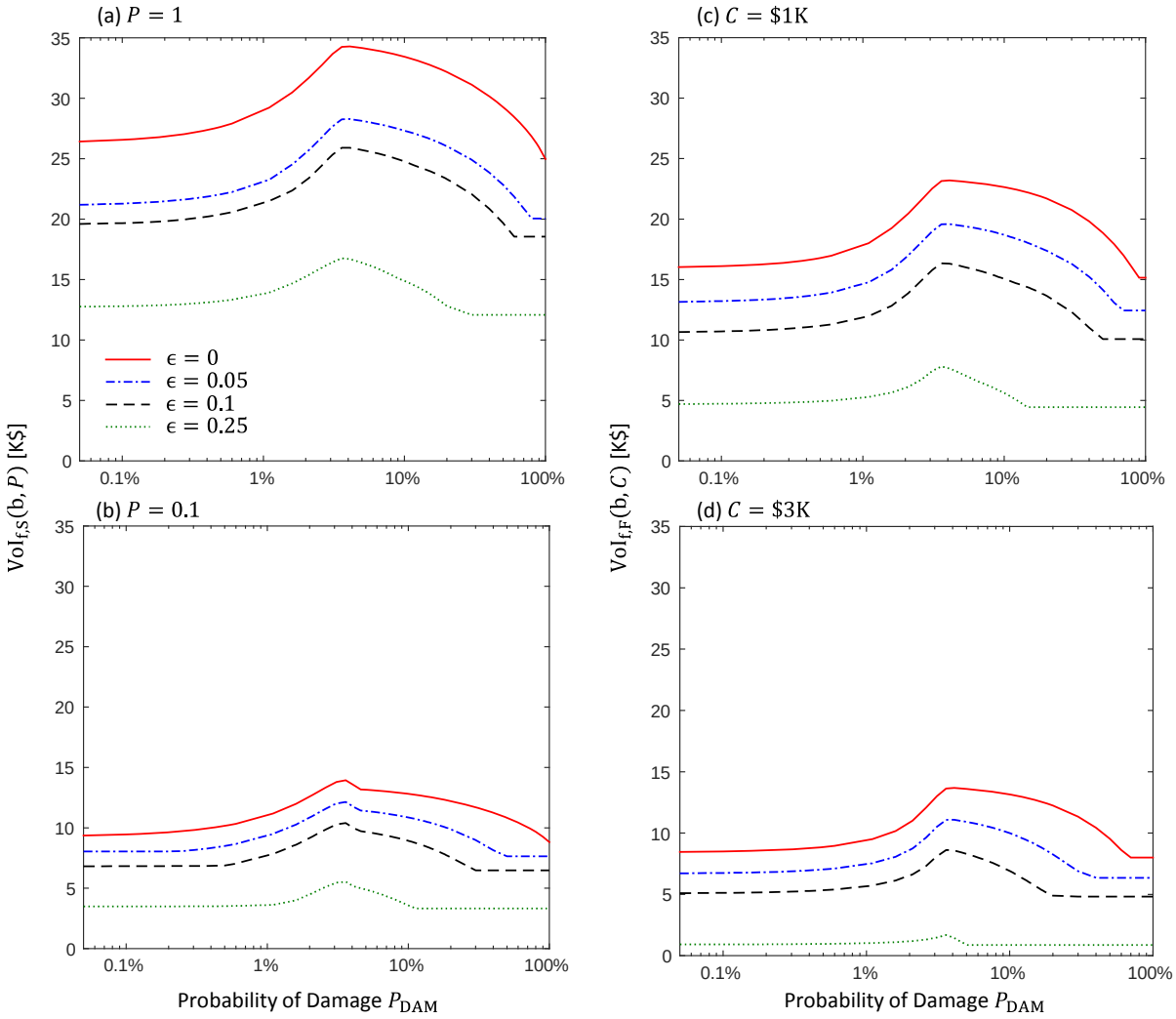

Figure 29. Value of flow of information for both SA (a,b) and FA (c,d) models as a function of probability of damage, P_{DAM} and inaccuracy ϵ.

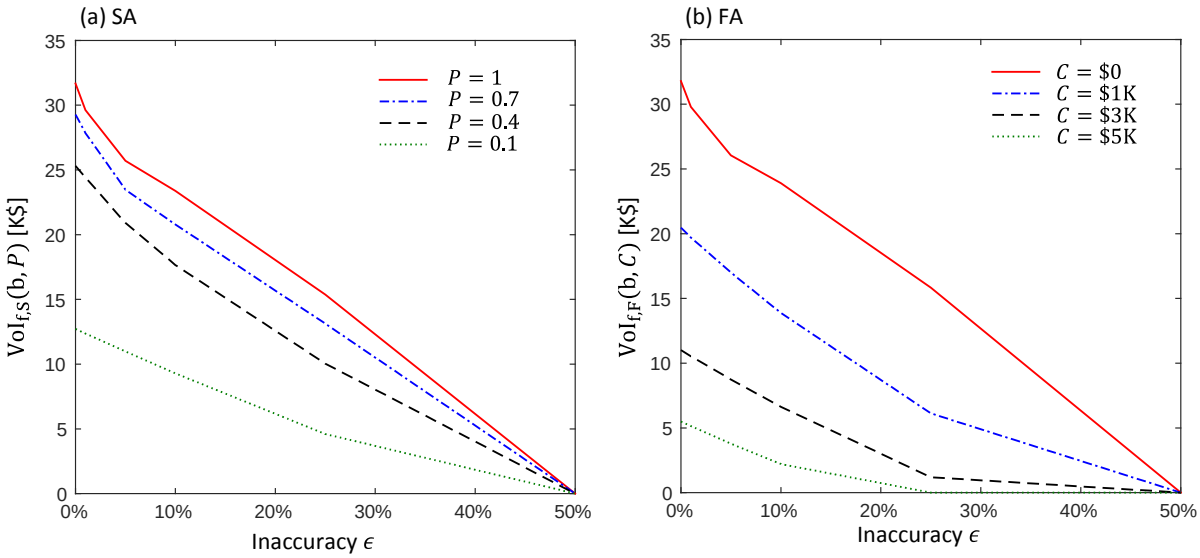

Figure 30. Value of flow of information as a function of inaccuracy, ϵ for (a) SA model with change in P and (b) FA model with change in C.

Figure 31 shows the values for the current inspection. First, we note that current VoI is zero when (as for P_{DAM} equals to zero or one) there is no uncertainty on the component's state. As for those related to the flow, VoI is monotonically decreasing with inspector inaccuracy ϵ. However, the VoI is not monotonically related to either availability P or fee C anymore. As noted in (Krause et al. 2008, Krause and Gusterin 2009, Memarzadeh and Pozzi 2015c), the VoI of one piece of information is related in a complicated way to the availability of others. If inspector is perfect and the probability of damage is 1%, the agent is ready to pay about \$1K when $C = \$1\mathrm{K}$, and up to \$2K when $C = \$3\mathrm{K}$. This can be explained as follows: when future inspections become expensive, it becomes more needed to inspect now. If, on the other hand, the probability of damage if 10% and inaccuracy is 25%, the agent is willing to pay up to about \$2K when $C = \$1\mathrm{K}$, and up to about \$500 when $C = \$3\mathrm{K}$. To explain this opposite behaviour, we reason as follows: when future inspections relatively cheap, it may be convenient to inspect at the present and future time without replacing; but when future inspections are expensive, it become pointless to inspect now, and it is more convenient to replace. Furthermore, in our framework, it is not

guarantee that the value of the flow is higher than that on the current inspection alone, even when parameters are the same. To understand why, sufficient is to note that, when availability P is zero, the flow VoI is zero (as there is nothing to evaluate), while the current VoI is not (as, in that case, zero availability means that the component will not be inspected anymore in the future). Generally, availability and fee refers only to the future, when evaluating the current inspection, and also the present, when evaluating the flow.

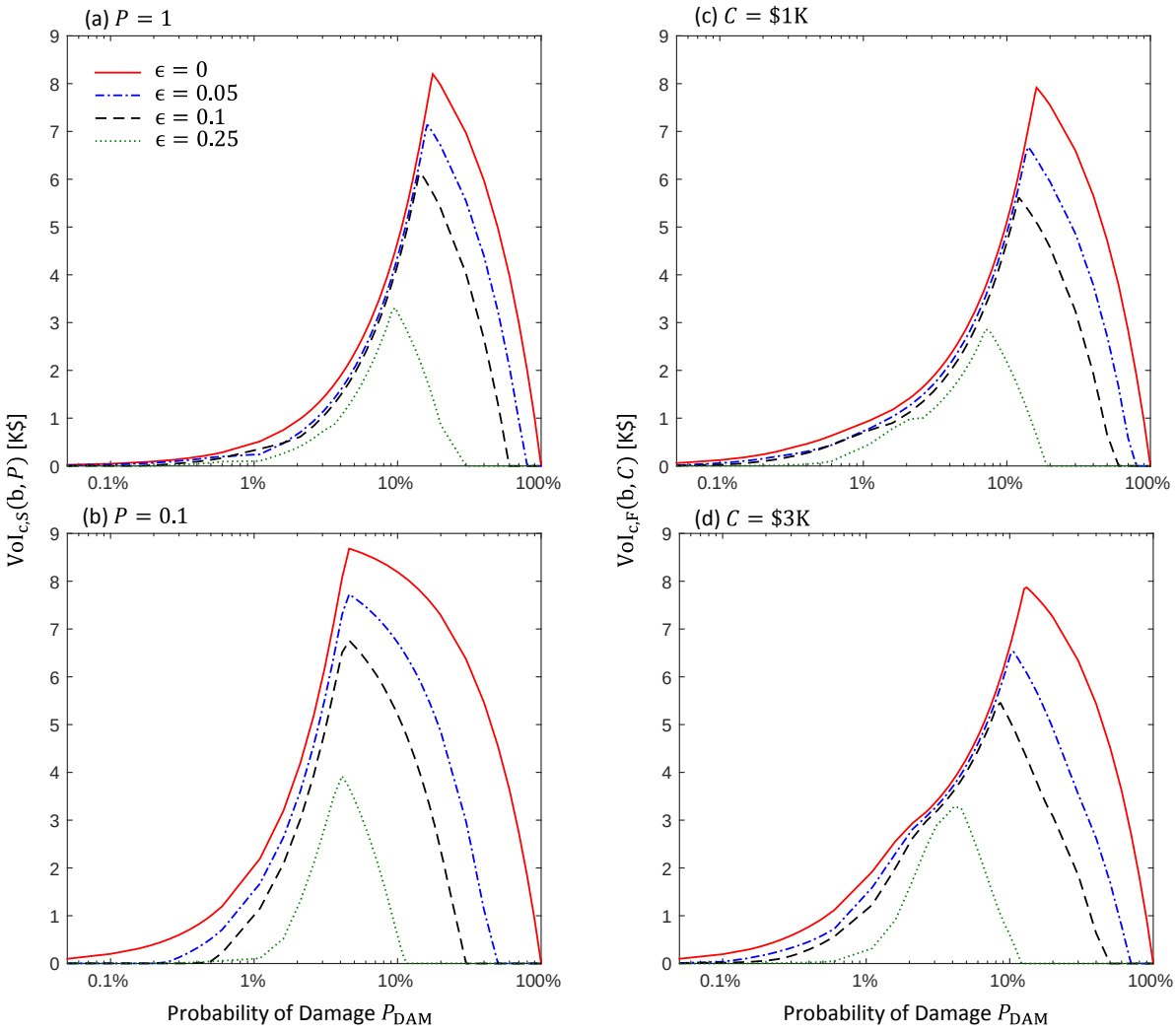

Figure 31. Value of current information for both SA (a,b) and FA (c,d) models as a function of probability of damage, P_{DAM} and inaccuracy ϵ.

The peaks of VoI in Figure 31a-b have similar values, independently on availability P, ranging from 0.1 to 1. However, this cannot be generalized to other problem settings. On the contrary, availability generally affects all aspects of current VoI. To show this, we modify the parameters of the illustrative example in Table 2 by changing the cost function and transition probability. In the first scenario, we only change the cost of repair to \$1K and cost of failure to \$2K. The results are reported in Figure 32a-b for $P = 0.1$ and $P = 1$, respectively. The current VoI now decays with increase in availability. In the second scenario, component deteriorates faster; to model this, we change the transition probability (under action *Do-Nothing*) to the following one:

$$\mathbf{T}_1 = \begin{bmatrix} 98\% & 2\% & 0 \\ 0 & 85\% & 15\% \\ 0 & 0 & 1 \end{bmatrix}$$

The results are reported in Figure 32c-d: now current VoI increases with future availability. These examples illustrate how current VoI may be highly sensitivity to future availability, and how the relation is complicate: this motivates the research of appropriate assumptions of this availability.

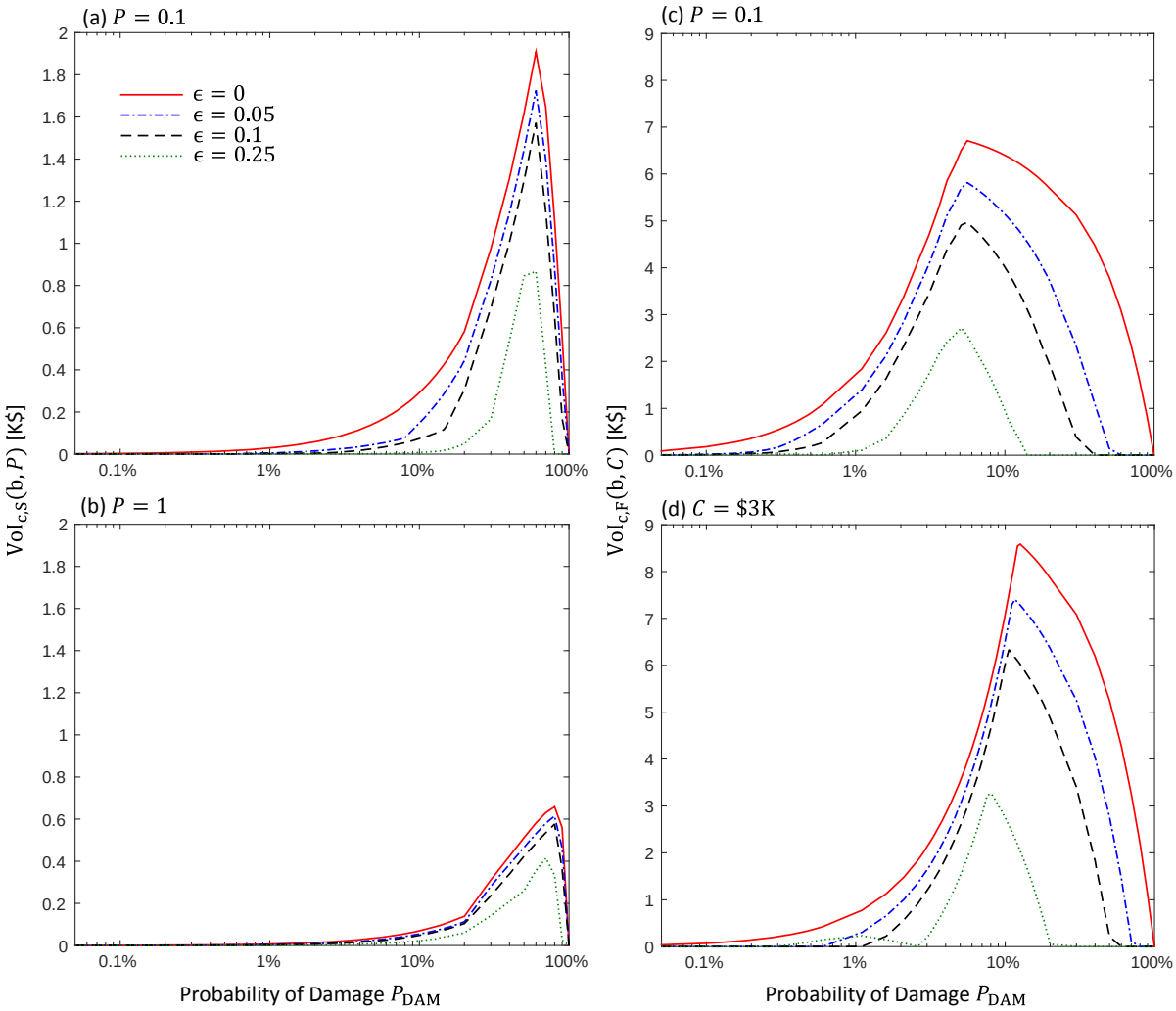

Figure 32. Value of current information for scenario 1 (a,b) and 2 (c,d) as a function of probability of damage, P_{DAM} and inaccuracy ϵ.

To better illustrate the relation between future availability of observations and management decisions, Figure 32 shows how optimal policy π^* results for the SA and FA models. For this problem, the agent will *do nothing* until the probability of damage reaches a threshold value, P^*_{DAM}, and she will *replace* after that (and if component fails), so that policy π^* is completely defined by the threshold. Figure 32a and b reports the threshold for the SA and FA models respectively, as a function of inspector inaccuracy, for different values of availability of fee. We note that the more inaccurate the inspector, the more conservative the policy. Also, the less available, of the more expensive the inspector, the more conservative the policy. An inspector

with fee cost \$5K and inaccuracy 30% is completely useless, as it will never be used: this follows consistently from both Figure 32b and 30b.

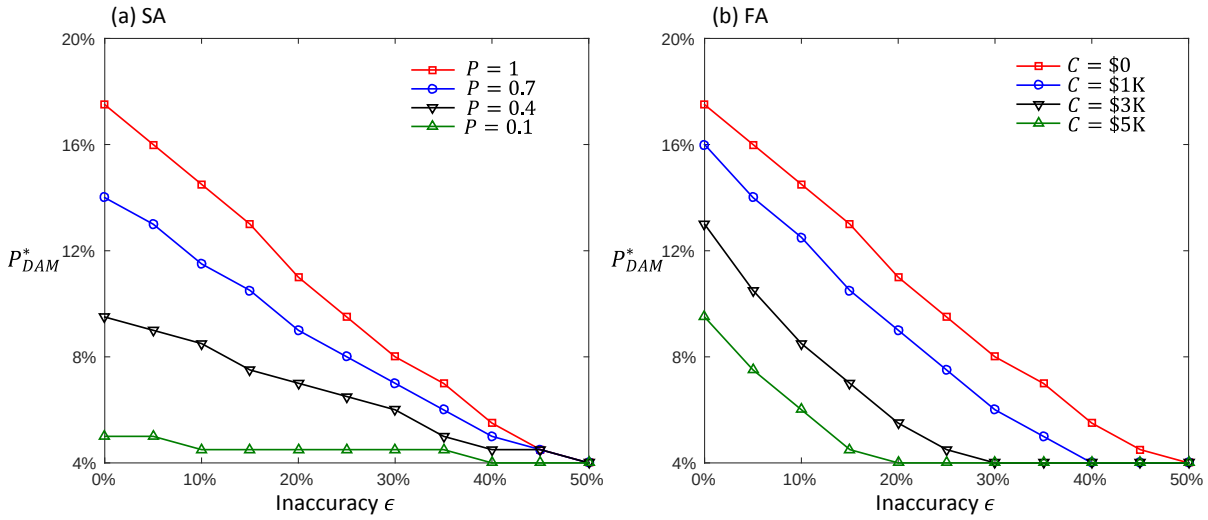

Figure 33. Threshold value, P^*_{DAM}, that the policy changes from do-nothing to repair as a function of inaccuracy, ϵ, for (a) SA model with change in P, and (b) FA model with change in C.

Chapter 6

System-level Inspection Scheduling

Abstract

Value of information (VoI) is a key concept for directing explorative actions, and in the context of infrastructure operation and maintenance, it has application to decisions about inspecting and monitoring the condition states of the components. The component-level VoI can be used as a heuristic for assigning priorities to system-level inspection scheduling, dealing with the limited resources for data collection. In this chapter, we evaluate the performance of the stochastic future allocation (and its two limiting scenarios called “pessimistic” and “optimistic”) and fee-based future allocation models for integrating adaptive maintenance planning based on POMDP and inspection scheduling based on a tractable approximation of VoI suggested by these models. We illustrate how these models can be used at system-level inspection scheduling with several numerical and analytical examples. Finally, we introduce analytical formulas based on these models to predict the impact of a monitoring system (or a piece of information).

6.1 Problem Statement

Suppose an agent is managing a system made up of several components, periodically receiving imperfect observations about their condition states. The agent also has access to inspectors that can collect additional information but, due to resource restrictions, only a limited number of components per time can be inspected. The problem we focus on is how to schedule inspections and integrate this task in the maintenance policy.

6.1.1 General Approaches for Inspection Scheduling

Priority on sequential information gathering can be based on measures of uncertainty of component condition state: the less we know about a component, the higher the need to inspect it. In this context, methods based on entropy (Cover and Thomas 2006) have been developed for sensor placement (Krause et al. 2008, Malings et al. 2013). Entropy provides a useful heuristic, but it does not guarantee an optimal ranking among inspections. Some uncertainty, in fact, is irrelevant in the maintenance process and should not be considered in the ranking as, for example, when uncertainty for a specific component is high only among states that are all acceptable and related to the same optimal maintenance action. Furthermore, that measure is not affected by the relative importance of each component.

An alternative heuristic is related to components' probability of failure, as higher probability may call for more urgent attention (Straub and Faber 2006). It should be noted that for binary-state components and low probabilities, this metric is consistent with the previous one based on entropy. While this metric is optimal in some applications, in a general context it is questionable for two reasons: (i) when the condition state is defined by many possible values, it is not

obvious how to define a unique “probability of failure”. (ii), for high probability of failure, the metric would lead to undesirable effects, assigning high priority to inspection of components that are known (even with certainty) to be in an unacceptable state, for which the prior maintenance action cannot be changed by any inspection outcome.

VoI can be understood as a combination of the previous intuitions, measuring only the uncertainty that is relevant for the maintenance process and impacts on decision-making.

6.2 Proposed Method

6.2.1 Problem Formulation

Consider a system made up of N components, each modelled by an independent POMDP with known model parameters. Optimal system-level maintenance can be found by solving independently the component-level POMDPs. But now suppose that, at each time step, the agent has access to $K \leq N$ inspectors that, depending on the setting, can provide perfect or imperfect observations on condition states: at each time-step, the agent has to decide which components to inspect. Because this decision has to be taken at system-level, the management processes of all components become dependent: if some components are inspected, some others cannot be.

Figure 33 provides a decision graph for the system level maintenance process. We add subscript i to indicate that state, observation, action and reward refers to component i. Furthermore, variable $h_{i,t}$ on domain H identifies the outcome of inspection of component i at time t, that can be observed if inspection is executed. Formally, inspection accuracy can be modelled by invariant emission probability function **E**, defined over the space of states and inspection outcomes $S \times H \rightarrow [0,1]$ as follows: $E(s,h) = P\big(\, h_{i,t} = h \mid s_{i,t} = s \,\big)$.

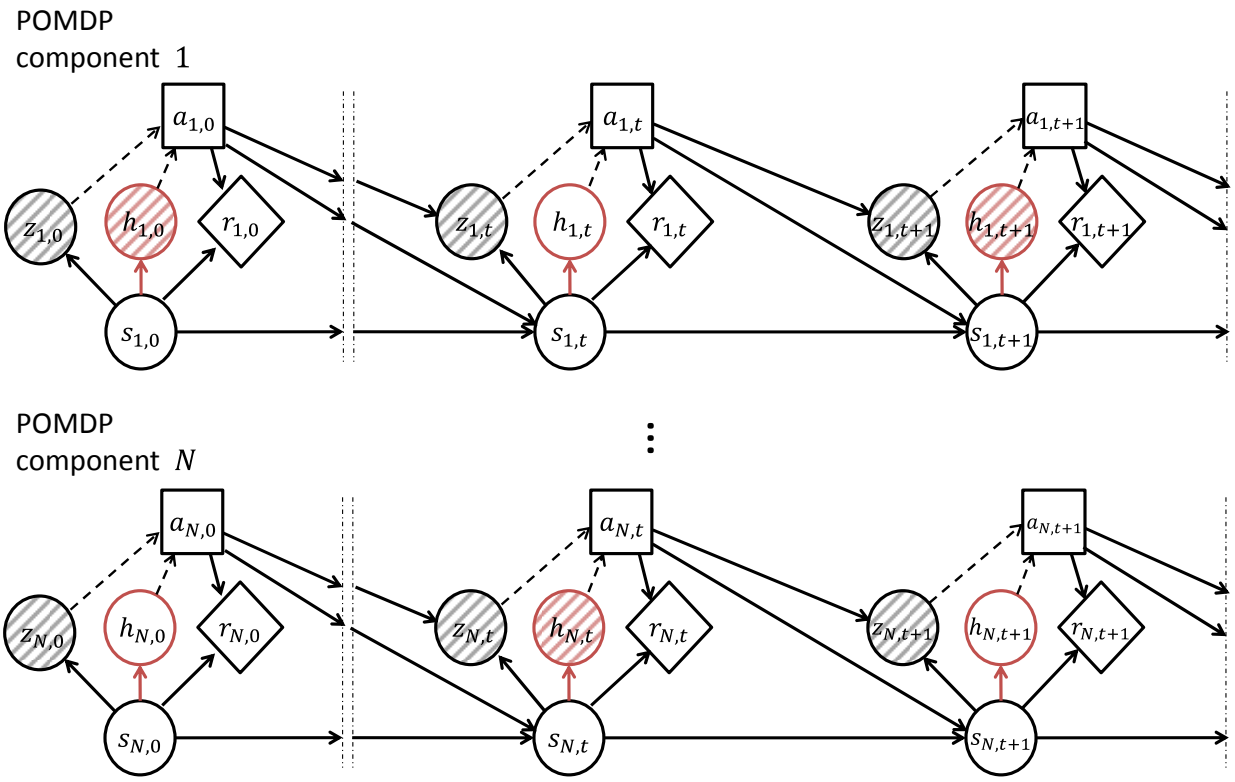

Figure 34. Decision graph for a system made up of N components modeled as POMDPs. Variables hs indicate outcomes of inspections.

In a more general setting, each element of the 8-tuple defining the POMDP can be component-dependent: for example, the number of possible condition states can be different for each component. Furthermore, function **E** could also depend on component and action. However, for ease of notation, we assume the 8-tuple and function **E** to be the same for all components, the more general setting being a trivial extension of our formulation. Also, we do not include any inspection cost and instead impose a constraint on the number of available inspectors, the extension to include inspection cost being straightforward. It should be noted that, given the maintenance actions, stochastic processes are independent for each component, so beliefs after the initial step are generally different across components, because of the randomness in the components' evolution and emission.

The problem of inspection scheduling can be described as "sequential variable selection": at time step t, while taking maintenance actions, the agent has to select only K among N variables $\{h_{1,t}, \dots, h_{N,t}\}$ to observe.

In this formulation, we assume that the agent can inspect any component at any time, without the need of planning inspections well in advance. Furthermore, inspected components at each step have not to be close in space. Removing these assumptions would induce dependence among decisions in time and space, and require a more complicated formulation.

6.2.2 Exact Solution

In principle, the problem can be formulated as a single POMDP at system-level. We can list all component states, observations and actions at current time step into an augmented state s^+, observation z^+ and action a^+, defined on domains S^+, A^+ and Z^+ of size $|S^+| = |S|^N$, $|A^+| = |A|^N$ and $|Z^+| = |Z|^N$ respectively. The maintenance process alternates two types of decisions: the agent has to select maintenance actions a^+ and to select locations Y to send inspectors to, on domain Ω_Y of size $|\Omega_Y| = \binom{N}{K} \cong N^K$. Outcomes of all inspections are listed in augmented variable h^+, on domain H^+ of size $|H^+| = |H|^K$. Details of the formulations are listed in Appendix B. In summary, the size of each augmented domain (states, observations, actions) grows exponentially with the number of components, and the problem becomes intractable except for small problems. For example, when $N = 25$, $K = 1$, $|S| = 3$, $|A| = 2$, $|Z| = 4$, and $|H| = 3$, the system-level dimension $|S^+|$ is 8.47×10^{11}, which is computationally intractable.

In the exact formulation, current inspection scheduling is coupled with future scheduling, as usually happens in sequential decision making.

6.2.3 Pessimistic and Optimistic Heuristics

In this section, we propose an approximate method to find sub-optimal solutions to the problem using the concept of VoI, by decoupling the complicated system-level optimization problem to N tractable component-level POMDP problems (Memarzadeh and Pozzi 2015c).

We introduce the idea by discussing Figure 34, which reports three schemes of observing variables h. At time t, the agent should select the set of K variables with the highest VoI. Figure 33a shows a realization of variable selection, for $K = 2$. However, VoI of any set of current variables depends on availability of observations at future time steps. For example, the VoI of inspecting component i at the current step depends on whether that component will be inspected at the next step or not. However, future inspection scheduling has not been fixed yet, so the problem has the complexity described in 6.2.2. On the other hand, it should be noted that, because of independence among component POMDPs, once future inspection scheduling is fixed, VoI can be computed independently for each variable at the current step, and optimization can be decoupled into N component-level optimization problems.

Two limit cases we investigate in this paper are shown in Figure 34b-c, and we refer to them as optimistic and pessimistic respectively. The optimistic approach assumes all components will be inspected from the next time step onwards, while the pessimistic approach assumes that no component will ever be inspected after the current time step. Although neither approach is consistent with observing K variables per step, they provide limiting scenarios for the VoI.

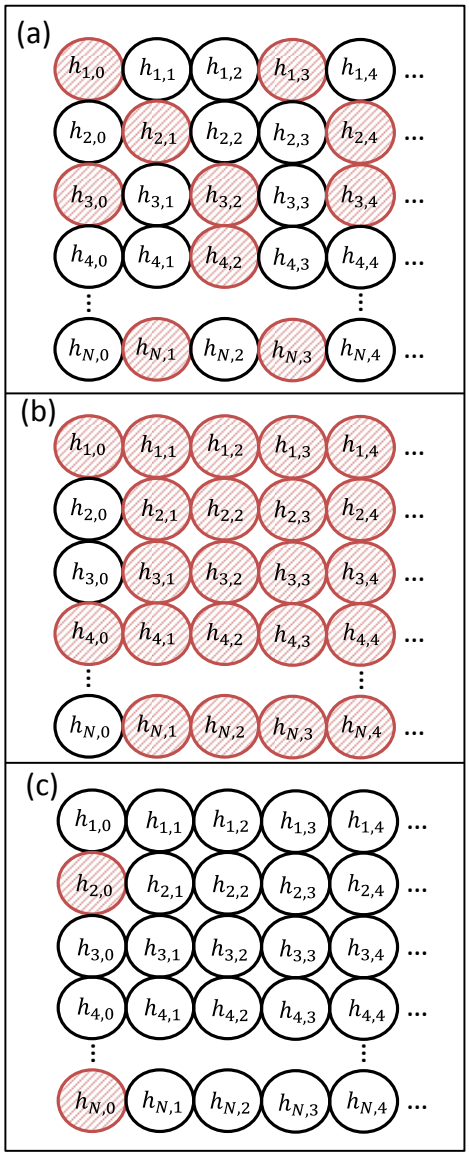

Figure 35. For $K = 2$, (a) arbitrary consistent, (b) optimistic and (c) pessimistic assumptions on inspections scheduling.

As noted above, the value related to any component can be computed independently when inspection scheduling is fixed. To compute pessimistic and optimistic VoI, we first define $V_i^{+a\to d}$ as the optimal value of managing component i, evaluated at time t, when inspections are scheduled for all times from $t+a$ to $t+d$ for that component. Specifically $V_i^{\emptyset}$, V_i^{+0}, $V_i^{+1\to\infty}$ and $V_i^{+0\to\infty}$ refer to never inspecting the component, inspecting it at the current time only, inspecting it from the next step onwards, and always inspecting the component, respectively.

6.2.3.1 Component-level Optimal Values

For component i, we define two functions related to alternative emission probabilities: $V_i^{(Z)}(b)$ and $V_i^{(Z,H)}(b)$ indicate optimal component value without and with inspections respectively at all steps, starting from belief b. Both functions can be evaluated using traditional POMDP solvers. While function $V_i^{(Z)}$ derives from emission $\mathbf{O}$, function $V_i^{(Z,H)}$ derives from emission $\mathbf{O}$ and $\mathbf{E}$, combining observations and inspections. In detail, we can concatenate observations and inspections in variable $d = \{z, h\}$, on domain $G = Z \times H$ of size $|G| = |Z||H|$. Because of conditional independence of z and h given s, we can define the emission function including inspections as $G(s, a, d) = O(s, a, z)E(s, h)$.

Value $V_i^{\emptyset}$ is equal to $V_i^{(Z)}$, while V_i^{+0} can be computed as follows from the current belief of the component b_i:

$$V_i^{+0} = \sum_{j=1}^{|H|} e_j^{\mathrm{I}}(\mathbf{b}, \mathbf{E})\, V_i^{(Z)}[u^{\mathrm{I}}(\mathbf{b}, j, \mathbf{E}), \mathbf{\Theta}] \tag{36}$$

Where e_j^{I} and u^{I} is defined in Eq. (24).

Figure 35 shows the decision tree for computing $V_i^{+1\to\infty}$. T do so, we need to first find the reachable belief states in the next time step from the current belief, b_i. If action a is taken, belief is updated to $b'_{i,a}(s_{t+1}) = \sum_{s_t} b_i(s_t)T(s_t, a, s_{t+1})$. As from the next time step, the agent always inspects, the problem can be formulated as a POMDP with emission probabilities $\mathbf{G}$ and observation d, as follow:

$$V_i^{+1\to\infty} = \min_{a\in A}\left\{r(\mathbf{b}, a, \mathbf{\Theta}) + \gamma \sum_{j=1}^{|D|} e_j(\mathbf{b}, a, \mathbf{\Theta}) V_i^{(Z,H)}[u(\mathbf{b}, a, j, \mathbf{\Theta}), \mathbf{\Theta}]\right\} \tag{37}$$

where e_j and u are defined in Eq. (3).

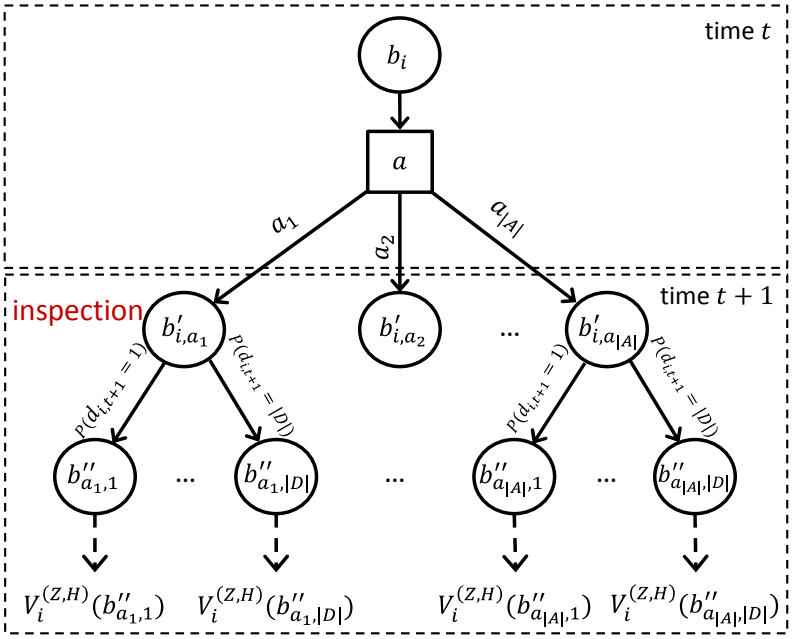

Figure 36. Decision tree for computing $V_i^{+1\to\infty}$.

6.2.3.2 VoI According to Pessimistic and Optimistic Heuristics

If inspection scheduling is decoupled, the agent allocates available inspectors to components based on an importance measure IM. As K inspectors are available, only the K components with highest importance measure can be inspected. Both the optimistic and pessimistic approaches agree that importance measure IM_i for component i is the VoI of current inspection on that component, but they disagree on how to approximate VoI.

As the pessimistic approach (P) relies on the assumption that no inspection will be available from the next time step onwards, pessimistic VoI ($VoI^{(\mathrm{P})}$) for component i can be computed as follows:

$$IM_i^{(\mathrm{P})} = VoI_i^{(\mathrm{P})} = V_i^{+0} - V_i^{\emptyset} \tag{38}$$

On the contrary, the optimistic approach (O) assumes that all components will be inspected starting next time step, and the corresponding optimistic VoI ($VoI^{(\mathrm{O})}$) is derived as:

$$IM_i^{(\mathrm{O})} = VoI_i^{(\mathrm{O})} = V_i^{+0\to\infty} - V_i^{+1\to\infty} \tag{39}$$

At each time step after having processed imperfect observations, the agent sends inspectors to components with the highest importance measures and, after receiving inspection outcomes, applies the resulting optimal maintenance policy. It is worth noting that the agents use different policies: that of the pessimistic approach is based on value function $V_i^{(Z)}$, while that of the optimistic approach is based on $V_i^{(Z,H)}$.

6.2.3.3 Bounds on the Value of the Pessimistic and Optimistic Approaches

Selecting an approach for inspection scheduling corresponds to selecting a policy, and the effectiveness of any policy should be assessed by its value. However, no closed formula is available to assess beforehand the values following the optimistic or the pessimistic policy, that we name $U^{(\mathrm{O})}$ and $U^{(\mathrm{P})}$ respectively. In this section, we define bounds for these values, which are relevant for two reasons: (i) to compare the effectiveness of different policies and (ii) to predict the overall expected discounted cost of the maintenance process. Furthermore, we can provide also bounds for the intractable optimal policy, to allow for an indirect comparison with the heuristics.

Let us define $W^{\emptyset} = \sum_{i=1}^{N} V_i^{\emptyset}$ and $W^{+1\to\infty} = \sum_{i=1}^{N} V_i^{+1\to\infty}$ as the system-level values, while never inspecting any component and always inspecting all components from the next time step onwards, respectively. As in Section 6.2.2, let us define Y as the decision variable for inspection scheduling, listing the identification of K components to be inspected. The pessimistic estimate $W_Y^{(\mathrm{P})}$ of the system-level value is as follows:

$$W_Y^{(\mathrm{P})} = W^{\emptyset} - \sum_{i\in Y} V_i^{\emptyset} + \sum_{i\in Y} V_i^{+0} = W^{\emptyset} + \sum_{i\in Y} VoI_i^{(\mathrm{P})} \tag{40}$$

It should be noted that, if no further inspection is available, $W_Y^{(\mathrm{P})}$ is the optimal value, scheduling according to Y at the current step. Similarly, the optimistic estimate $W_Y^{(\mathrm{O})}$ of the system-level value is as follows:

$$W_Y^{(\mathrm{O})} = W^{+1\to\infty} - \sum_{i\in Y} V_i^{+1\to\infty} + \sum_{i\in Y} V_i^{+0\to\infty} = W^{+1\to\infty} + \sum_{i\in Y} VoI_i^{(\mathrm{O})} \tag{41}$$

which, again, is the optimal value if inspections are available for all components from the next step onwards, scheduling according to Y at the current step.

$Y^{(\mathrm{P})}$ and $Y^{(\mathrm{O})}$ are the scheduling assigned by the pessimistic and optimistic policy respectively, following the measures defined in Eqs. (38-39). Corresponding optimal system-level estimated values are computed as follows:

$$\begin{cases} W_*^{(\mathrm{P})} \triangleq W_{Y^{(\mathrm{P})}}^{(\mathrm{P})} = \max_Y W_Y^{(\mathrm{P})} \\ W_*^{(\mathrm{O})} \triangleq W_{Y^{(\mathrm{O})}}^{(\mathrm{O})} = \max_Y W_Y^{(\mathrm{O})} \end{cases} \tag{42}$$

$W_*^{(\mathrm{P})}$ and $W_*^{(\mathrm{O})}$ are defined as the maximum of the pessimistic and optimistic estimates over all possible allocation of inspections, Y, respectively.

Bounds for $U^{(\mathrm{P})}$ and $U^{(\mathrm{O})}$ are provided below and proved in Appendix C:

$$\begin{cases} W_*^{(\mathrm{P})} \leq U^{(\mathrm{P})} \leq W_{Y^{(\mathrm{P})}}^{(\mathrm{O})} \\ U^{(\mathrm{O})} \leq W_*^{(\mathrm{O})} \end{cases} \tag{43}$$

$U^{(\mathrm{P})}$ is bounded between two values that can be computed beforehand. $U^{(\mathrm{O})}$ has an upper bound higher than the pessimistic one, but no relevant lower bound (or course, lower bounds can be found, but those computable beforehand are usually much lower than $W_*^{(\mathrm{P})}$).

Another upper bound for the value of each possible policy is that of the optimal policy (U^*) defined as in Section 6.2.2. U* cannot be computed beforehand, and it is bounded as follows:

$$W_*^{(\mathrm{P})} \leq U^* \leq W_*^{(O)} \tag{44}$$

Figure 36 summarizes the inequalities among bounds for all values. The key observation is that adopting the optimistic approach exposes the agent to low values, while the pessimistic choice has a lower bound. The next section and the appendix provide examples and clarification about these behaviours.

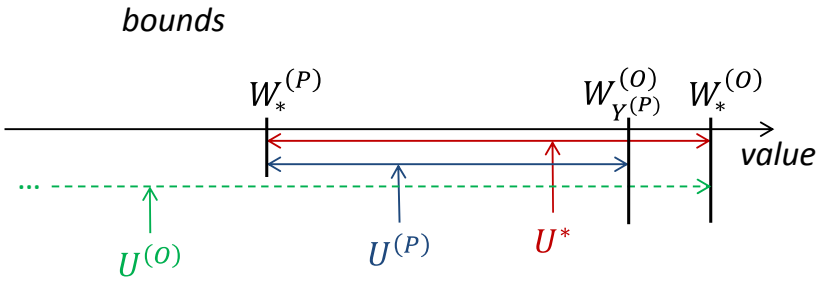

Figure 37. Bounds over values of the pessimistic, optimistic, and optimal agents.

6.2.4 Stochastic Future Allocation

As illustrated in Section 5.2.2.1, by using the SA model we can assess the value of current inspection assuming the component will inspected randomly with probability P at any future step. When all components belong to the same typology (i.e. parameter set $\boldsymbol{\Theta}$ is the same for each component), it is natural to assign availability as $P = K/N$ to everyone, for the assessment of the current VoI. When components are heterogeneous, it is more challenges to assign appropriate availability, but we recommend two rules. First, to assign future availabilities $\{P_m\}_N^{m=1}$, where P_m is that assigned to component m, so that the sum of them all is one. Doing so, the expected value of inspections at any future step would be K, as the constraint imposes. Second, if empirical data about the rate of inspection per each component, under an effective policy, are available, those data can be used for calibrating the availabilities.

The SA-based approach is approximated for two reasons: because the constraint, which must be exactly fulfilled according to the problem statement, can only be fulfilled in the expected sense; furthermore, the approach assume that future inspections are randomly scheduled, while they are actually the result of an optimization process.

6.2.5 Fee-based Future Allocation

The use of the FA-based current VoI, as shown in Section 5.2.2.2, can be intended as an effort to overcome the latter limitation of SA-based approach. The FA approach recognizes that future inspections will be allocated to components that need them. This need can be approximated by the readiness to pay fee C to receive the inspection. Therefore, even if no actual fee will ever be paid, a virtual fee C is assumed, so that future inspectors are supposed to be

allocated to those ready to pay it. By doing this, the FA-based approach mimics the output of the future optimizations for inspection scheduling. If C is set to a low value, too many components will assume the inspector available at future times, while too few will do if C is too low. We would recommend to set C so that, in the expected sense, K components will ask for future inspections. Again, if data are available about the VoI during previous steps, that C can be estimated as that necessary and sufficient to win the auction and get the inspector, in the average.

6.2.6 Predicting the Impact of Optimal Inspection Scheduling

By making use of the flow analysis, we can also assess the overall economic impact of integrating inspectors in the management process, respect to not using any inspector at all (Memarzadeh and Pozzi 2015d). According to SA model, the K inspectors will be used randomly among the N components, with corresponding availability P_m for component m. The overall system-level value of this distributed flow of information ($SVoI_{\mathrm{S}}$) is:

$$SVoI_{\mathrm{S}} = \sum_{m=1}^{N} VoI_{\mathrm{f,S},m}(\mathbf{b}_m, P_m) \tag{45}$$

where $VoI_{\mathrm{f,S,m}}$ is computed as in Eq. (29) using all parameters for component m, including the initial belief b_m. Eq. (45) tends to underestimate the system-level VoI, as it assumes that inspectors are allocated randomly.

The corresponding estimate for the FA model is as follows:

$$SVoI_{\mathrm{F}} = \left[\sum_{m=1}^{N} VoI_{\mathrm{f,F},m}(\mathbf{b}_m, C)\right] - \frac{C \cdot K}{1-\gamma} \tag{46}$$

where $VoI_{\mathrm{f,S,m}}$ derives from Eq. (31), and index m refers to the corresponding component. The FA model assumes that all components ready to pay C will get the inspector, while the second term on the right hand side of Eq. (46) takes into account that the fee is not actual.

6.3 Numerical Investigation of System-level Inspection Scheduling

6.3.1 Pessimistic and Optimistic Approaches

6.3.1.1 Example A: A System made up of 2-state Components

To investigate the performance of proposed approaches depending on the system size, we investigate how values depend on number of components in a binary setting, where each component can be in one of two possible states: Intact ($s = 1$) or Failure ($s = 2$). Two maintenance actions are available to the agent, namely: Do-Nothing (a = 1) and Replace ($a = 2$). No additional observation is available (equivalently, function $\mathbf{O}$ is flat on all states), other than the outcomes of only one perfect inspector ($K = 1$). After doing nothing, the condition state cannot recover, while replacement improves the component's state to intact. The transition probabilities for all components are reported in Table 3, where $\{T_l\}_{m,n} = T(s_t = m, a = l, s_{t+1} = n)$. The cost of replacing is assumed to be \$10K, and the cost of loss of production due to failure and down time to be \$10K, as well. The initial belief state is assumed to be 80% for intact state and 20% for failure, and discount factor is 0.95.

Table 3. Transition and emission probabilities for numerical examples in section 6.3.1.

example A

$$\mathbf{T}_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{T}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

example B

$$\mathbf{T}_1 = \begin{bmatrix} 0.90 & 0.08 & 0.02 \\ 0 & 0.90 & 0.10 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0.90 & 0.10 & 0 \\ 0.90 & 0.10 & 0 \end{bmatrix}$$

example C

$$\mathbf{T}_1 = \begin{bmatrix} 0.90 & 0.08 & 0.02 \\ 0 & 0.90 & 0.10 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}_2 = \begin{bmatrix} 0.90 & 0.10 & 0 \\ 0.80 & 0.20 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{O}_1 = \mathbf{O}_2 = \mathbf{O}_3 = \begin{bmatrix} 0.60 & 0.20 & 0.20 & 0 \\ 0.20 & 0.60 & 0.10 & 0.10 \\ 0 & 0.20 & 0.20 & 0.60 \end{bmatrix}$$

$$\mathbf{E}_{IP} = \begin{bmatrix} 0.90 & 0.08 & 0.02 \\ 0.05 & 0.90 & 0.05 \\ 0.02 & 0.08 & 0.90 \end{bmatrix} \quad \mathbf{E}_P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 37 shows the VoI based on the pessimistic and optimistic approaches, as calculated in Eq. (38) and (39), respectively, as a function of the probability of failure, $P_F = P(s_t = 2)$. VoI is always non-negative (Heckerman et al. 1993) and it is zero when the state of a component is known with certainty ($P_F = 0$ or $P_F = 1$). The maximum VoI occurs for P_F equal to 27.5% and to 53.0% for pessimistic and optimistic respectively. The difference between these two graphs indicates that priorities among components to be inspected are different depending on the approach.

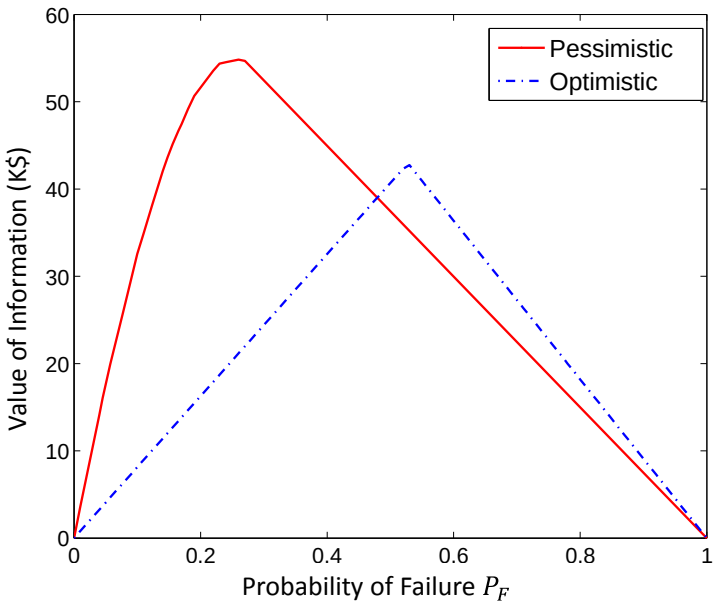

Figure 38. VoI as a function of the probability of failure (P_F) based on pessimistic and optimistic approaches.

Processing 100 forward simulations, Figure 38 compares the values of pessimistic and optimistic policies normalized with respect to number of components as N increases, reporting the 95% confidence intervals. For high N, $U^{(\mathrm{P})}$ moves towards its lower bound $W_*^{(\mathrm{P})}$: this behavior can be expected, as the assumption of not having any inspector available from next step becomes more and more realistic as ratio K/N goes to zero. On the other hand, the assumption of the optimistic agent becomes less and less accurate as that ratio vanishes, and it is to be noted that $U^{(\mathrm{O})}$ goes below even $W_*^{(\mathrm{P})}$. At a first glance, it may seem counter-intuitive that the optimistic agent receives a value lower than that achievable with no future inspections at all, but this can be explained by analyzing the optimistic planning. By relying on the availability of future inspections that actually will not be done, the optimistic agent plans poorly, e.g. postponing repairs that should be timely executed. In detail, both agents will repair after detecting a failed component and do nothing when detecting an intact one, however their behavior for un-inspected components is different: the pessimistic agent will repair when

$P_F > 27.5\%$, while the optimistic repairs only when $P_F > 53.0\%$ (it should be noted that these values correspond to the maximum VoI). While the former is the optimal policy without inspections, the latter policy is not effective when K/N is small. This also sheds light on why, as can be noted in Figure 37, $VoI^{(\mathrm{O})}$ can be higher or lower than $VoI^{(\mathrm{P})}$ for some values of P_F. Despite it seeming intuitive that an inspection has higher value if it cannot be repeated at the next step, it is well known that the value of observing one variable can be increased or decreased by the availability of other observations (Krause and Guestrin 2009).

In this example, the upper bound for pessimistic ($W_{Y^{(\mathrm{P})}}^{(\mathrm{O})}$) and optimistic ($W_*^{(\mathrm{O})}$) approaches is the same at the initial step, because beliefs for all components are the same. We note that the maximum gap between optimistic and pessimistic upper bounds can be found as follows:

$$W_*^{(O)} - W_{Y^{(P)}}^{(O)} \leq K \max_{i \in \{1,\ldots,N\}} \left\{ \max_{\mathbf{b}} VoI_i^{(O)} \right\} \tag{47}$$

where $\mathbf{b}$ is the belief, that here can be described by P_F. For this example, the maximum gap is \$4K.

When N equals 2 and 5, we can also compute the optimal value following the procedure described in section 6.2.2. The black circle in the graph shows that, for this case, both pessimistic and optimistic agents perform approximately as good as the optimal agent.

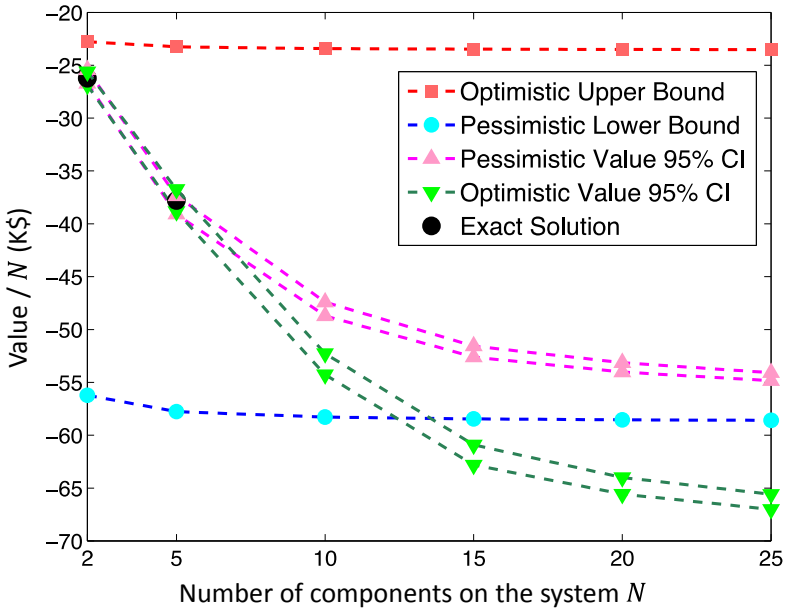

Figure 39. Value of pessimistic and optimistic agents on management of the system in example A (CI stands for confidence interval).

On this example, the performances of stochastic and fee-based future allocation models (described in Chapter 5) are similar to that of the pessimistic approach. However, we will show in Memarzadeh and Pozzi (2015c) and (2015e) that the stochastic model can outperform both pessimistic and optimistic in some settings. Moreover, we will show in section 6.3.2.3 that fee-based model can better capture optimal behavior respect to the stochastic model.

6.3.1.2 Example B: A System made up of 3-state Components

To show that the optimistic approach can outperform pessimistic when the number of inspector is high e.g. for ratio K/N close to one), we investigate a system made up by $N = 20$ components, by varying the number of inspectors. The condition state of each component is discretized into three possible states (Intact, Damaged, and Failure); again, two actions are available to the agent namely Do-Nothing ($a = 1$) and Repair ($a = 2$), and no observation is available other than the ones from perfect inspectors. The transition probabilities are reported in

Table 3. Cost of repair and failure are \$30K and \$100K respectively, while discount factor is 0.95. Figure 39 shows the outcomes from 100 simulations: when $K = 2$, pessimistic outperforms optimistic while, as K increases, optimistic performs better for $K > 10$, as her assumption is closer to reality.

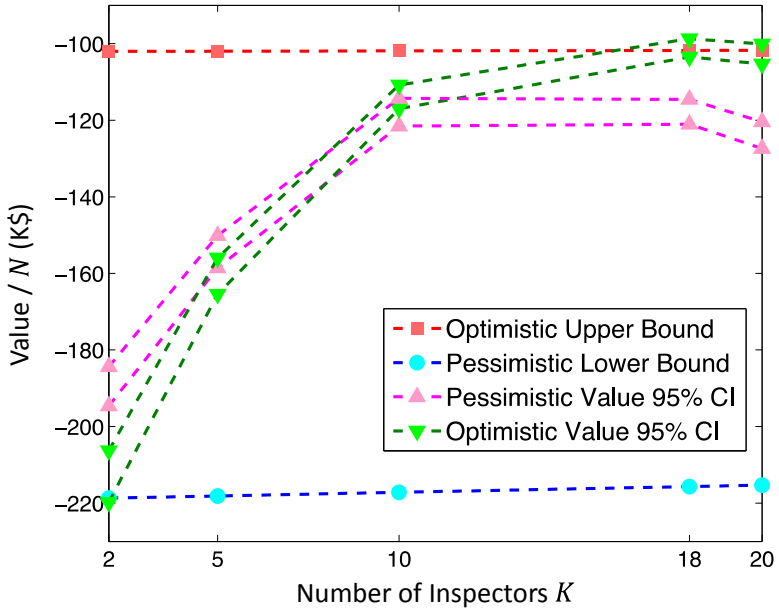

Figure 40. Value of pessimistic and optimistic agents on management of the system in example B (CI stands for confidence interval).

6.3.1.3 Example C: Wind Farm Management

By adapting the model presented in Chapter 3 and 4, we investigate the performances of the approaches for the operation and maintenance of a wind farm (the system) consisting of $N = 25$ turbines (the components). Each turbine is in one of three states: Intact (INT: $s = 1$), Damaged (DAM: $s = 2$), and Failed (FAIL: $s = 3$); four possible imperfect observations are available, along with three actions, namely: Do-Nothing (DN: $a = 1$), Repair (RE: $a = 2$), and Replace

(REP: $a = 3$). The transition and emission probabilities are reported in Table 3. We investigate the process for different numbers K of inspectors, in set $\{1, 2, 3, 5, 10, 20\}$. We also investigate two inspection accuracies: perfect and imperfect. The emission probabilities for imperfect inspection ($\mathbf{E}_{\mathrm{IP}}$) and perfect inspection ($\mathbf{E}_{\mathrm{P}}$) are reported in Table 3.

The repair, replacement and failure costs are \$10K, \$30K and \$40K respectively. The initial belief state is 80% for intact and 20% for damaged state, while the discount factor is 0.95.

Figure 40 shows the optimal policy as a function of belief without future inspections (a), and with use of imperfect (b) and perfect (c) inspections. The belief's domain is represented by an equilateral tringle, and each belief's component can be read by following the grid lines up to the corresponding side. Policy without inspections is more conservative than that of imperfect inspections, which in turn, is more conservative than that with perfect ones.

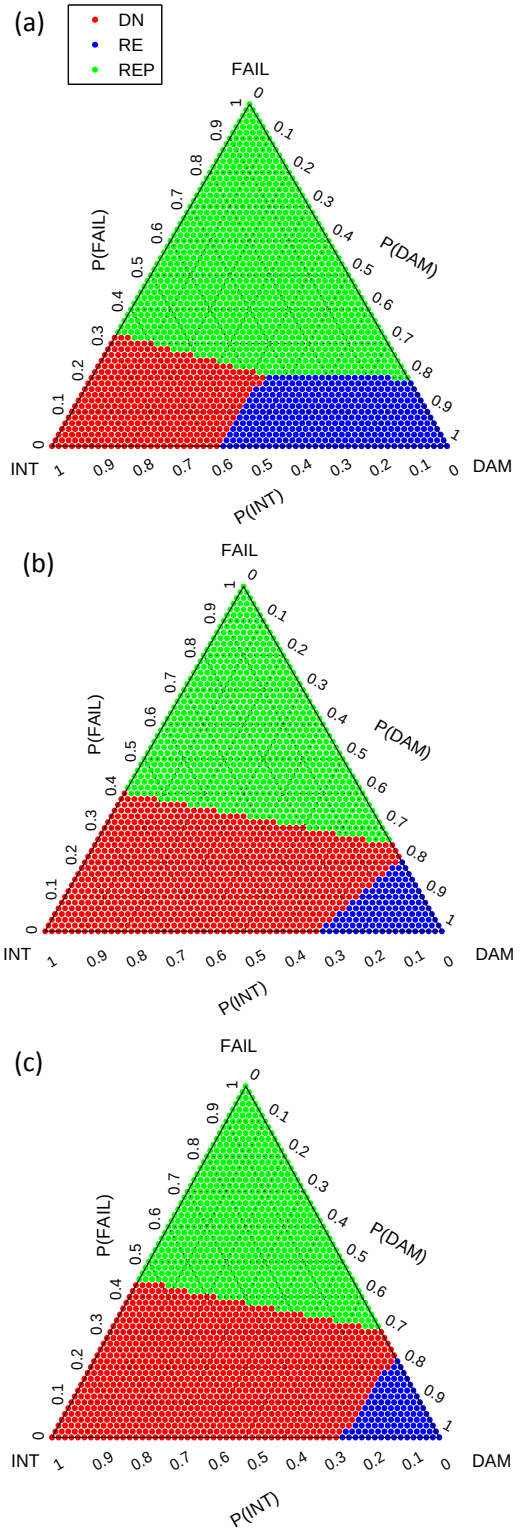

Figure 41. Policy as a function of the belief for management of the wind farm in example B (a) without any inspector, (b) with imperfect inspectors, and (c) perfect inspectors that can be available for all components at all time-steps.

Figure 41 shows the VoI as a function of belief state for the pessimistic approach with (a) perfect and (b) imperfect inspectors, and for the optimistic approach with (c) perfect and (d) imperfect inspectors. High VoI corresponds to beliefs close to the decision boundaries, where changes in action assigned by the corresponding maintenance policy occur, and it is zero when the state is known with certainty. The triangles' axes are defined as in Figure 40.

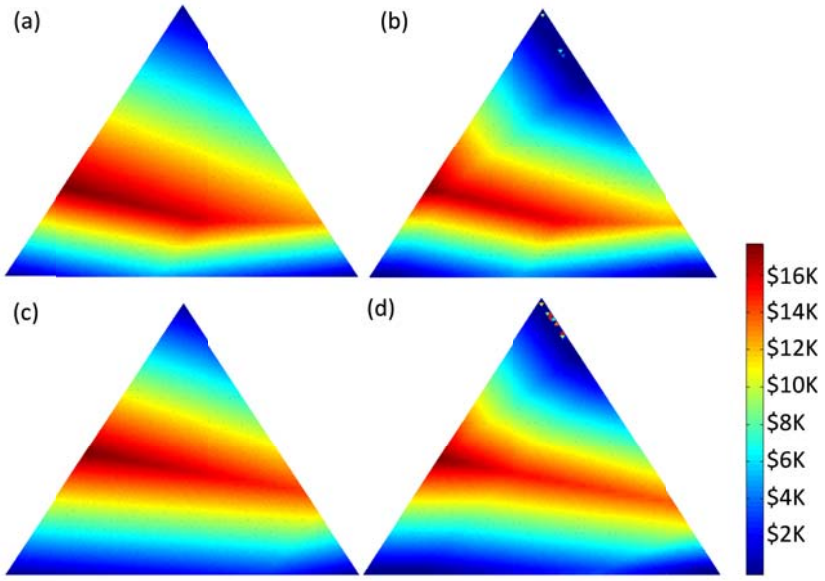

Figure 42. VoI as a function of the belief state for the pessimistic approach with (a) perfect and (b) imperfect inspectors and optimistic approach with (c) perfect and (d) imperfect inspectors.

Figure 42 illustrates the performance of optimistic and pessimistic approaches, comparing perfect and imperfect inspections. For both (a) pessimistic and (b) optimistic approaches, perfect inspectors outperform imperfect inspectors, as expected. This analysis allows for investigating the trade-off between the number of inspectors and their precision: for example, with the pessimistic approach, 10 imperfect inspectors outperform 3 perfect ones, while 5 perfect ones outperform 20 imperfect ones, as it can be seen in Figure 42a.

From these examples and other analytical examples not reported in this paper, we summarize our evaluation of the two heuristics as follows. Pessimistic and optimistic approaches are correct

if ratio K/N is equal to zero and one respectively (of course, in both settings no inspection scheduling is needed, and only planning is relevant). A reasonable conjecture that could be supported by this remark is that the optimistic value $U^{(\mathrm{O})}$ is above pessimistic value $U^{(\mathrm{P})}$ when ratio K/N is sufficiently high. However, this conjecture is incorrect in general: even one missing inspector ($\mathrm{K} = \mathrm{N} - 1$) can have a significant negative impact in the optimistic planning. On the other hand, there are cases in which $U^{(\mathrm{O})}$ is above $U^{(\mathrm{P})}$ even if the ratio is arbitrary low.

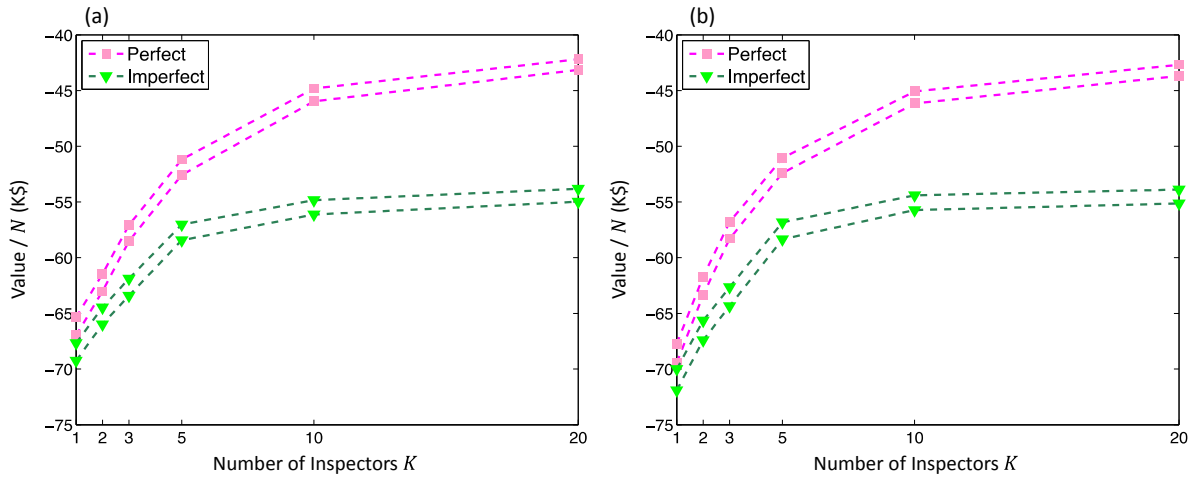

Figure 43. Comparison between the performance of the (a) pessimistic and (b) optimistic approaches with perfect and imperfect inspectors.

6.3.1.4 Discussion on Additional Examples

Pessimistic and optimistic approaches are correct if ratio K/N is equal to zero and one respectively (of course, in both settings no inspection scheduling is needed, and only planning is relevant). A reasonable conjecture that could be supported by this remark is that the optimistic value $U^{(\mathrm{O})}$ is above pessimistic value $U^{(\mathrm{P})}$ when ratio K/N is sufficiently high. However, in Appendix D, we discuss two examples, for which analytical solutions are available, disproving this conjecture. In summary, even one missing inspector ($K = N - 1$) can have a significant

negative impact in the optimistic planning. On the other hand, there are cases in which $U^{(O)}$ is above $U^{(P)}$ even if the ratio is arbitrary low.

6.3.2 Stochastic and Fee-based Approaches

6.3.2.1 Example of Pavement Management

We start with an example adapted from that reported by Guillaumot et al. (2003), about pavement management. A component indicates here a road segment. The number of possible condition states is 8, where $s = 1$ indicates Intact and $s = 1$ Failure. The agent can do-nothing ($a = 1$, DN) or replace ($a = 2,$ RE). Transition and emission probabilities are reported in Table 4: while $\mathbf{T}_1$ models the degradation process, $\mathbf{T}_1$ models an imperfect replacement. Emission $\mathbf{O}_{1-2}$ defines a binary observation that, depending on its outcome, can be a symptom of a good or of a deteriorated state. Cost of repair is \$20K and the cost of failure, to be paid when s is 8, is \$1M, while the discount factor is 95%.

Table 4. Parameters of pavement management example.

$$\mathbf{T}_1 = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.75 & 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0.35 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.55 & 0.45 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{T}_2 = \begin{bmatrix} 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \\ 0.6 & 0.4 & 0 & \dots & 0 \end{bmatrix} \qquad \mathbf{O}_{1-2} = \begin{bmatrix} 0.1 & 0.9 \\ 0.2 & 0.8 \\ 0.3 & 0.7 \\ 0.5 & 0.5 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \\ 0.9 & 0.1 \\ 1 & 0 \end{bmatrix}$$

Figure 43 shows a realization of management for an independent component, following the corresponding optimal policy. Graph (a) reports the belief, (b) the observations and taken actions, and (c) the underlying realized state evolution that, needless to say, is not accessible to the agent.

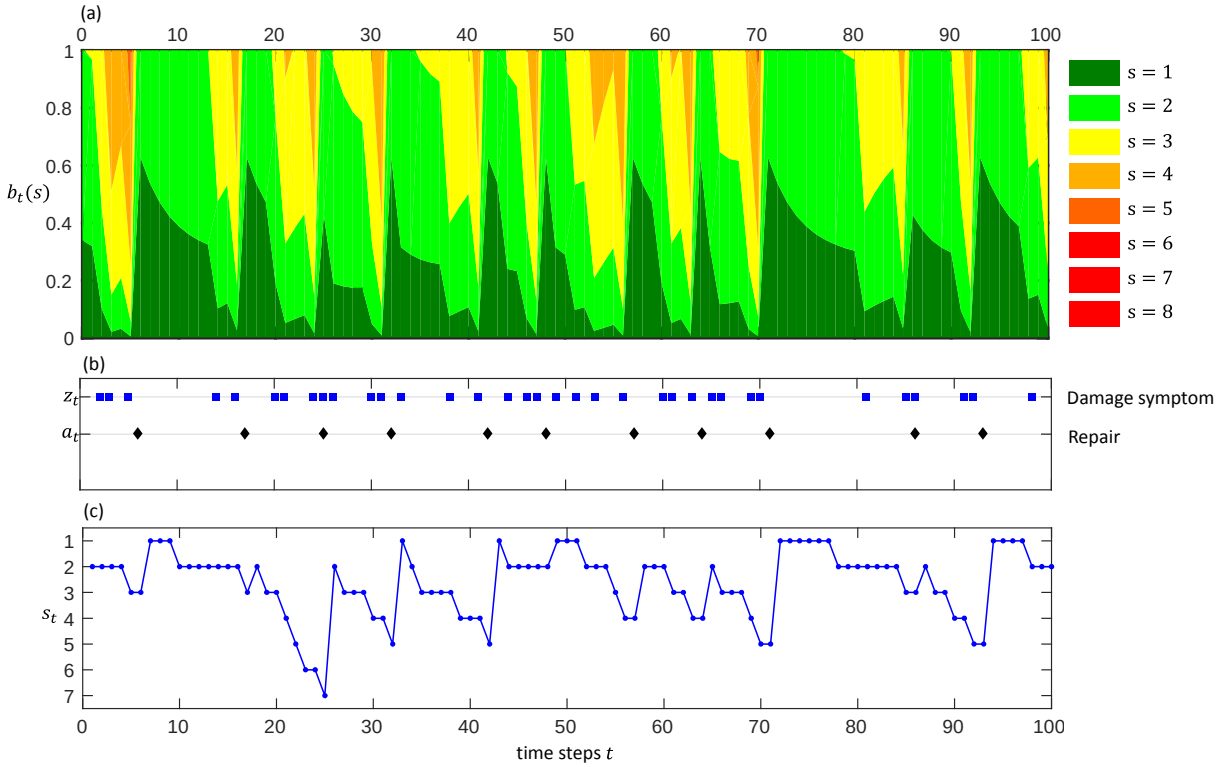

Figure 44. Realization of management for an independent component without availability of inspectors: (a) belief state, (b) damage symptoms and repairs, and (c) underlying condition states.

Figure 44 reports the realized management of a component is a system made of 20 (i.e., $N = 20$) modelled as belonging to the same typology. Now the agent has also access to two perfect inspectors (i.e., K is two and **E** the identity matrix) that needs to be allocated at each time step. The simulation is controlled by the FA model's assumption with availability $P = K/N = 10\%$. Now graph (b) also reports the realized inspections, (c) the current VoI, with continuous line, while the underlying hidden state is in (d). Comparing Figure 43b with 44b, it appears that inspecting decreases the replacement rate. Usually, VoI increases while the agent does nothing, because of the degradation, depending in the detecting symptoms, up to when it is sufficiently high the component wins the auction and gets the inspection. With the inspections, uncertainty is reduced. When the inspection detects a good state, another one is scheduled after some steps, while if it detects a bad state, a replacement is scheduled. In details, each auction involves all

components, so it may happen that the component does not get the inspector even when VoI is high, as happens at step 55: not being able to inspect, the agent prefers to replace.

Figure 44 also shows the analysis of the FA model. The corresponding current VoI is plotted in (c) with a dashed line. The assessment is so similar to that of the SA model that, in this specific example, inspection and replacement scheduling is consistent between the models.

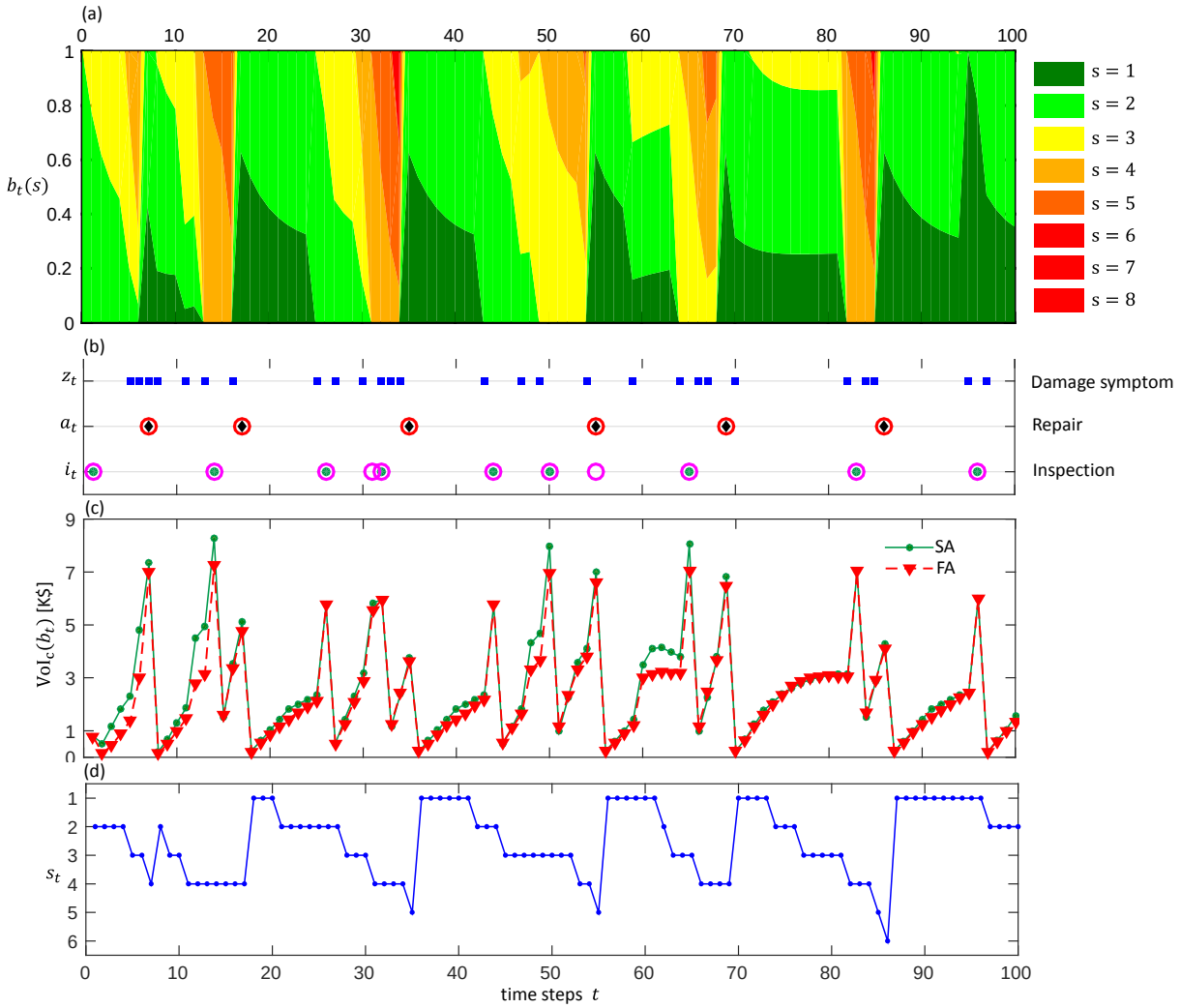

Figure 45. Realization of management for an independent component with availability of inspectors: (a) belief state, (b) damage symptoms, repairs (black diamond show the repairs based on SA model and red circles show the repairs based on FA model) and inspections (green dots show the repairs based on SA model and magnet circles show the inspections based on FA model, (c) value of current information, and (d) underlying condition states.

Assessment current VoI according the FA model depends on an assumption of fee C. Here we have adopted the following algorithm: fee is initialized at zero at step 0, and the components compete under the optimistic assumption. $C_{M(t)}$ indicates the minimum VoI sufficient for winning the auction at time t, while $\bar{C}_{M(t)}$ indicates the average during all steps up to t Both this quantities are plotted in Figure 45. At time t, we assume C equal to $\bar{C}_{M(t)}$, because this latter quantity provides a reasonable estimate of the offer needed for winning the auction and get the inspection. During the first steps, $C_{M(t)}$ assumes a low value. This is because the initial belief assumes the road components being in a good state, and no significant concern arises in the first steps. Consequently, current VoI is generally low in those steps, and so is C_M. After about step 10, components may be in high need of attention, and the auction becomes more competitive, and so C_M grows. This poses challenges to the identification of an appropriate parameter C in the FA model. At step 7, for example, basing on the data collected in so far, the agent can find appropriate to assign a low value to C, say of about \$1K; if so, each component would assume a future inspection available at that low cost anytime in the future. This assumption turns out to be incorrect, as future demand from older components is much higher, and an offer of about \$5K is necessary (in the expected sense) to get the inspector. While this effect is not significant in this example, it can be in other settings. It is hard to provide a general formula to address this issue, but we recommend that, if the future C_M is predicted to change in the future (e.g. because the condition of the component population is predicted to deteriorate), C should reflect this pattern.

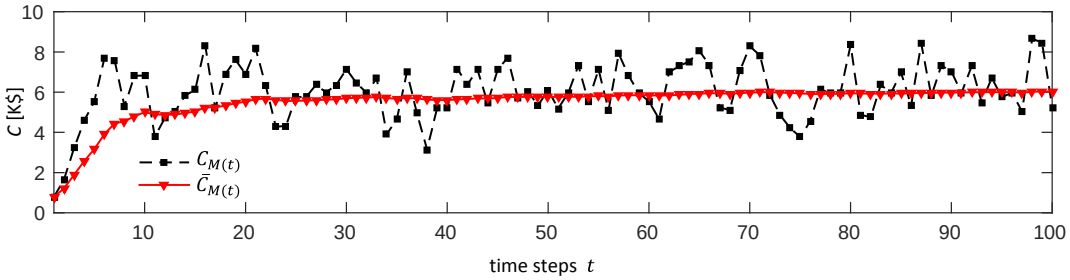

Figure 46. Realization of the minimum value of information sufficient for winning the inspection auction (black line with square markers), and the average value of information during all management time steps till time ***t*** (red line with triangle markers).

6.3.2.2 Evaluating the Impact of Inspectors

To investigate the accuracy of the predictions on Eqs. (45-46), we compare them with the results of numerical simulations. Because of the high number of simulations needed (a total of 6M steps), we use the smaller example illustrated in 5.3. We consider a system made by $N = 100$ components, and investigate the VoI using $K = 1,2,5,10,20$ and 50 perfect inspectors (i.e. $\epsilon = 0$). The actual VoI at system level is estimated from 100 forward simulations, each carried on for 100 steps. The SA model with $P = K/N$ controls the system. The value is then estimated as discounted arithmetical average of the costs, and the VoI as the difference respect to the value with no inspectors (given by the POMDP solver).

The 95% Confidence Region of the system-level VoI is reported in Figure 46 with dash-dotted lines, as a function of the number of inspectors. As expected, the VoI grows rapidly for low K and slower for high K (for $K = N$ components become MDPs). To avoid the issued about system-level synchronous aging mentioned in previous section, we assign initial beliefs to the components as those resulting from a 20-step simulation. Doing this, we assess the VoI when the system is made up by both intact and potentially damaged components. The Figure also shows the prediction of the models. For $K < N$, SA systematically underestimates the system-level VoI,

as it assumes that inspectors are randomly allocated. With the appropriate selection of the virtual fee C, on the other hand, FA is able to predict accurately the VoI, as the assumption that inspectors can be used by paying fee C is able to mimic the optimal allocation the K inspectors. This result requires an estimate of the virtual fee that, evidently, depends on the number of available inspectors.

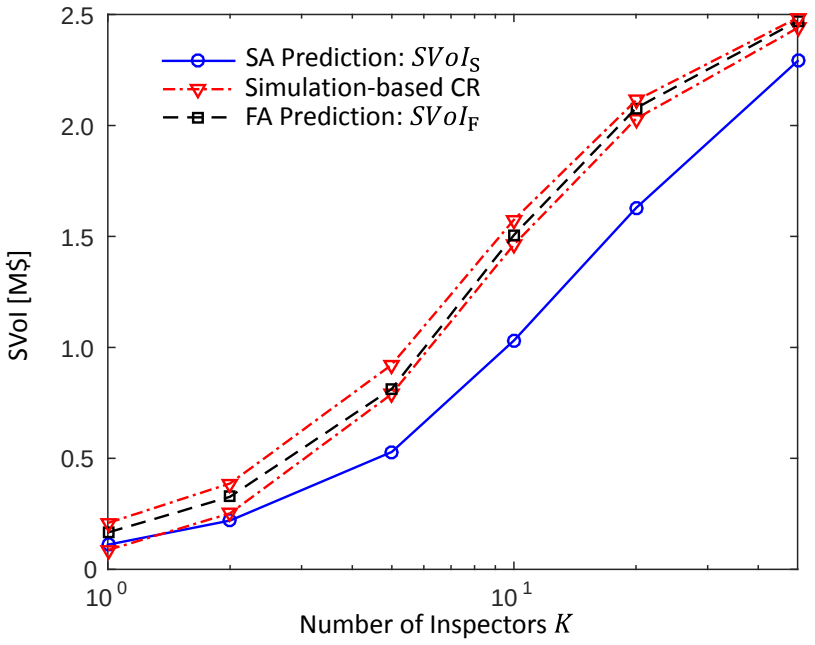

Figure 47. Confidence region of system-level value of information (red line with triangle markers) for a system of N components as a function of increase in the number of inspectors K, the SA model prediction of VoI (blue line with circle marker) and the FA model prediction of VoI (black line with square markers).

6.3.2.3 Comparing the FA and SA Models

Section 6.2.4 shows that the SA and FA models have similar performance in many settings. Section 6.2.5 shows how the prediction of impact using FA is more accurate, when an appropriate fee is used. Now we show how FA can be more effective than SA even in inspection scheduling, in specific settings. To do so, we refer to a simple two-step decision making problem with no discounting, for which analytical solutions are available. Consider a setting where

components, each made by many parts, can currently be in flawless (FL) or defective (DF) state, and this initial state is perfectly known to the agent. Because of this, inspecting is uninformative at initial step, but it is relevant at the next one. Unless protected at cost C_{PR}, a flawless component deteriorates in few of its parts; and the agent can detect all flaws only by inspecting at next step and fix them at no cost. Without inspection, the agent has to maintain a flawed component under uncertainty, at cost $C_M > C_{PR}$. A defective component, on the other hand, can be protected (say replaced with a protected one) at cost C_{RE}; if not, high deterioration will occur in some parts, and the agent cannot fix an uninspected highly deteriorate component, that will fail at cost $C_F > C_{RE}$. However, by inspecting at next step, again we assume the agent can detect and fix all issues at no cost.

An agent convinced of receiving an inspector at next step should not protect the component. On the contrary, one convinced that no inspector will be available should protect it. Specifically, if the assumed availability is lower than $P'' = 1 - C_{RE}/C_F$, the agent should protect a defective component, and if it is lower than $P' = 1 - C_{PR}/C_M$, also an intact one. If we assume $C_{PR} =$ \$1K, $C_M =$ \$5K, $C_{RE} =$ \$10K and $C_F =$ \$100K, then $P' = 90\%$ and $P'' = 80\%$.

To investigate the optimal policy, we can assess the VoI for inspecting at next step. Inspecting at next step allows for skipping current protection, so the VoI for flawless and defective components are $VoI_{\text{nx}}(\text{FL}) = C_{PR}$ and $VoI_{\text{nx}}(\text{DF}) = C_{RE}$ respectively. As $VoI_{\text{nx}}(\text{DF})$ is higher than $VoI_{\text{nx}}(\text{FL})$, the agent will give priority to inspections of unprotected defective components: only if K is higher than the number of these, unprotected flawless ones can also be inspected. Therefore, a long-sighed agent should assume a high probability of inspecting an unprotected defective component, and a low probability of inspecting an unprotected intact one. For example, if $N = 100$, $\text{K} = 20$ and the probability of a defective component is 6% then, if all

components are left unprotected, inspectors are available with probability about 1 to defective and with probability 14.9% to intact one. Consequently, the optimal policy (using component level information alone) is to protect an intact component and do not protect a defective one, as reported in Table 5, counting on the availability of inspectors for the latter. The corresponding Value, normalized for one component, is \$940. However, the SA model assumes a unique value for availability P, independent of state or belief and, because of this, cannot mimic that policy. For example, by setting $P = K/N = 10\%$, the agent will protect all components. Figure 47a shows the corresponding Value (normalized for one component), depending on the assumed availability, and the corresponding policy is reported in Table 5.

For the FA model, on the other hand, decision depends on assumed fee cost C. If C is higher than $C' = VoI_{\mathrm{nx}}(\mathrm{DF})$, the agent should protect an flawless component, and if it is higher than $C'' = VoI_{\mathrm{nx}}(\mathrm{FL})$, she should also protect a defective one. Figure 47b reports the normalized Value, depending on C and, again the corresponding policy can be read in Table 5. The optimal policy is reachable by setting C between the two VoIs.

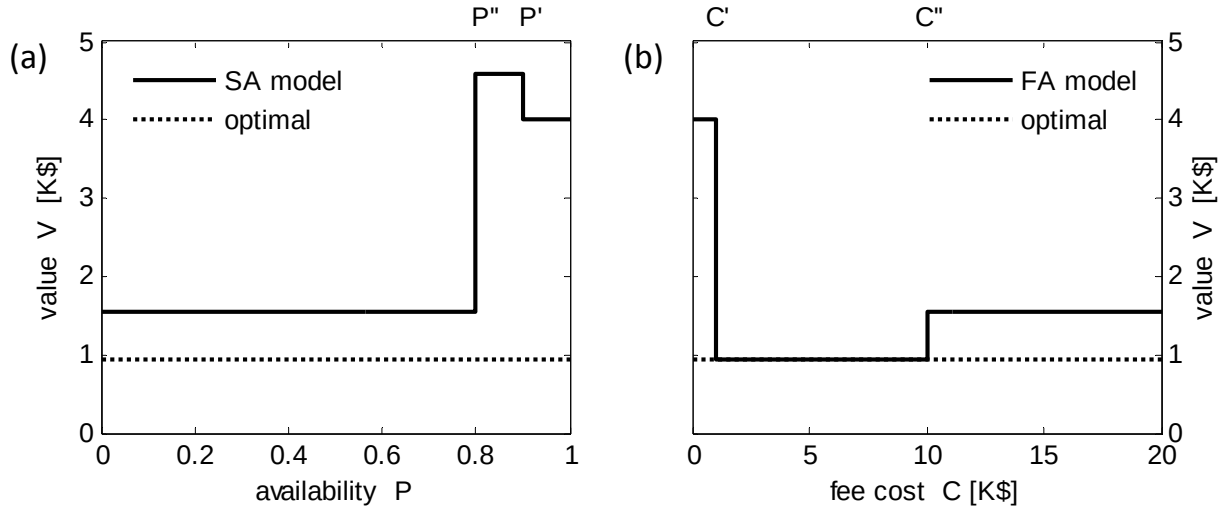

Figure 48. Value of managing a system $N = 100$ components and $K = 20$ inspectors as a function (a) availability of future inspections P and (b) the cost of inspection C. The dotted line shows the optimal solution while the weighted line shows the performance of (a) SA and (b) FA models.

Table 5. Policy depending on model and assumption, described by answering the question: is the component to be protected?

model	*assumption*	*initialstate* FL	DF
SA	$P \leq P''$	yes	yes
	$P'' < P < P'$	no	yes
	$P \geq P'$	no	no
FA	$C \leq C'$	no	no
	$C' < C < C''$	yes	no
	$C \geq C''$	yes	yes
optimal		yes	no

In summary, the example shows how the SA model is not able, in some settings, to emulate the optimal policy. Future inspection scheduling is the result of an optimization process and, consequently, availability is a function of the state. In the example above, the optimization gives priority to the unprotected defective components. But the FA model assumes that future inspections are randomly allocated, independently by the state and, doing so, it is not able to capture the optimal policy. The FA model, on the other hand, assumes that inspections will be available to those in needs (i.e. those able to cover the corresponding fee) and, by selecting an appropriate fee value, is able to get optimality.

We conclude this discussion with two remarks. First, the reader may note that a better policy is conceivable, whether information is processed at system level: if more than K components are defective, the agent can plan to inspect K of them, and protect the remaining ones. Or, if they are less, she can plan to inspect all defective ones, and also some selected flawless ones, while only the remaining ones are to be protected. While such a policy (leading to Value of $800 for component) is identifiable for this small problem, it violates the principle of using just component-level features for current decision and inspection scheduling. The complexity of system level policies for planning and inspection scheduling is illustrated in Section 6.2.2 and

suffer from the course of dimensionality. Second, the example leaves the issue of selecting an appropriate value for fee C open, as it just shows that such appropriate value exists.

Lastly, the reader is referred to Memarzadeh and Pozzi (2015c) and (2015e) for comparison between the SA model and the pessimistic and optimistic approaches.

Chapter 7

Summary and Conclusions

In this dissertation, we have developed a computational framework for system-level adaptive monitoring and control of infrastructures. Specifically, we have addressed challenges regarding: i) sequential decision making under uncertainty in the model parameters describing the degradation behavior and the precision of the monitoring system; ii) learning the model parameters by processing noisy observations, as well as modeling the dependence among the components, allowing the knowledge transfer among them; and iii) assessing the component-level value of the information and use it as a heuristic for system-level inspection scheduling, dealing with limited resources for data collection.

As a first step, in Chapter 3 and 4 (section 4.1), we proposed a method named planning and learning for uncertain dynamic systems (PLUS) for learning and planning within the BA-POMDP framework and applicable to the context of wind farm management, as well as other infrastructure systems. The BA-POMDP framework overcomes one of the primary limitations of POMDP framework by treating the transition and emission probabilities as random variables, whose distributions can be updated during the learning process. PLUS algorithm uses Markov chain Monte Carlo simulations to find an approximate solution for the BA-POMDP problem. The approach allows for a rational treatment of data collected by sensors and visual inspections, a reliable tracking of the condition states of components, and robust decision making support.

PLUS algorithm has been validated with synthetic data and is shown to outperform state-of-the-art reinforcement learning approaches, such as MEDUSA. MEDUSA was originally proposed for applications of robot navigation and it scales easier than PLUS, requiring less computational effort. However, for application to infrastructure systems, and specifically wind farms, it is believed that the computational drawback of PLUS is not a significant concern because the computational cost is low with respect to the direct costs for operation and

maintenance of a wind farm. On the contrary, in this context it is necessary to achieve a rational and robust selection of the maintenance policy, making use of the knowledge available at any state of the process. PLUS allows this; it also allows the agent to learn, during the management, the statistics of degradation process (transition probabilities) and the performance and reliability of the monitoring system (emission probabilities).

In the second step in Chapter 4 (section 4.2), we have proposed a probabilistic framework, Multiple Uncertain POMDP (MU-POMDP), for learning models in systems made up by similar components that overcomes some limitations of PLUS by explicitly defining dependence among components through a set of hyper-parameters. While Individual and Global PLUS solve the limit cases of independent or identical components, respectively, the MU-POMDP framework can potentially solve a wide range of intermediate problems. The computational complexity of MU-POMDP is higher than that of PLUS, since the former requires an extra layer of hyper-parameters. Specifically, the sampling approach makes use of the Metropolis-Hastings method, which needs careful selection of the proposal distribution to achieve effectiveness.

In many applications it is appropriate to assume some degree of similarity among components, and MU-POMDP is a consistent framework for this problem. The accuracy of approximated approaches, as Individual and Global PLUS, depends on the number of observations, as well as on the number of components and on other parameters of the application. We have measured the quality of approximation in terms of expected error, and expected economic loss. However, practical implementation of MU-POMDP requires defining an appropriate level of similarity among components, which needs careful considerations depending on the application.

Finally in Chapter 5 and 6, using POMDP framework, we have illustrated how to assess the value of information in sequential decision making. Specifically, we have proposed methods based on the Value of Information (VoI) for evaluating the impact of a monitoring system (or a set of additional information) on decision making, ranking priorities among inspections, and integrating inspections in maintenance planning. The exact method (except for approximation of the POMDP solver) is available for current and periodic observations at component-level that can be used for assigning priorities for system-level inspection scheduling. While this method is intrinsically approximate, it allows for decomposing the complex system-level optimization into component-level computation, so that the problem complexity grows just linearly with respect to the number of components on the systems. We first defined two heuristics (optimistic and pessimistic), related to different assumptions on the availability of future inspections: the optimistic approach assumes inspectors will be available for all components from next time step, whereas the pessimistic approach assumes no inspector will be available in the future. Consistent with these assumptions, heuristics are also related to alternative planning policies.

The exact solution is intractable except for small problems as discussed in Section 6.2.2. Each heuristic defines a myopic optimization algorithm whose complexity is linear with number of components N. Any formulation less myopic than those cannot rely on independence between components' values and a combinatorial explosion occurs. For example, in a two time-step search, inspections at the current step have to be coupled with those at next one, and $N \times \binom{N}{K} \cong N^{K+1}$ combinations have to be explored. Complexities of the optimistic and pessimistic approaches are similar. We have shown bounds for the value of each approach, and the pessimistic approach has better guarantees against worst case scenarios. On the other hand, the upper bound of optimistic approach is higher; however, we have found this benefit to be small in

many applications. Based on these considerations, without any additional relevant domain knowledge, we recommend adopting the pessimistic approach.

By extending the results of pessimistic and optimistic heuristics, we have introduced two models, stochastic future allocation (SA) and fee-based future allocation (FA), which differ for the assumption about the availability of future information (e.g. inspection): SA model assumes that observations are collected with a given probability (that is a generalization of the pessimistic and optimistic approaches), while FA model assumes that observations are available at a given cost. The computational complexity of FA is higher than that of SA, however it is proven to be more effective in some settings, as the assumption of a virtual fee for allocating resources on inspections can better predict the results of future optimization processes. Both models depend on the selection of internal parameters that can be chosen by expert judgment or by data analysis, if previous data are available. Specifically, the prior selection of an appropriate fee C is an open issue, and we do not provide a general formula for this task. We note that we have investigated using the solution of $C = VoI_{\mathrm{c,F}}(\mathbf{b}, C)$ for initializing C.

7.1 Future Work

PLUS planning phase is based on an assumption of neglecting the exploratory value of learning the model parameters in the planning. Basically, PLUS assumes that from next time step, the agent will receive perfect knowledge about the true model parameters, and plans optimally in this context. However, the exploratory value of learning can be incorporated into the planning feasibly by using appropriate heuristics. The challenge is that in the application to management of civil and infrastructure systems, exploration can be really costly and hence there

is a need for careful management of incorporating the exploratory value in the planning, which is part of the future work.

Moreover, one of the main assumptions in the proposed SA and FA methods is that the model parameters (i.e. transition and emission probabilities) of POMDP are assumed to be known with certainty. This assumption can be released by treating the model parameters as random variables by assigning priors to them and generalize the developed method to compute VoI under the model uncertainty, and acquire information not only with respect to their value of decreasing the uncertainty in conditions states and cost of operation and maintenance, but also with respect to their value in decreasing the uncertainty about the model parameters (i.e. transition and emission probabilities). This is a challenging task computationally, as the VoI also depends on the uncertain model parameters.

Bibliography

Aoki, M. (1965). “Optimal control of partially observable Markov systems.” J. Franklin Institute, 280(5), 367-386.

ASCE, American Society of Civil Engineers (2012), “Report Card for America’s Infrastructure”, http://www.infrastructurereportcard.org/ (accessed January 12, 2013).

Astrom, K.J. (1965). “Optimal control of Markov decision processes with incomplete state estimation.” J. Mathematical Analysis and Applications, 10, 174-205.

Bagnell, A., Ng, A., and Schneider, J. (2001). “Solving uncertain Markov decision processes.” Proc. Neural Information Processing Systems.

Barber, D. (2012). “Bayesian reasoning and machine learning”, Cambridge University Press, Cambridge, UK.

Bellman, R.E. (1957). “Dynamic programming”, Princeton University Press, Princeton, NJ.

Bertsekas, D.P. (1996). “Dynamic Programming and Optimal Control.” Athena Scientific, Volume 1.2.

Byon, E., Ntaimo, L., and Ding, Y. (2010). “Optimal maintenance strategies for wind turbine system under stochastic weather conditions.” IEEE Transactions on Reliability, 59(2), 393-404.

Byon, E., and Ding, Y. (2010). “Season-dependent condition-based maintenance for a wind turbine using a partially observed Markov decision process,” IEEE Transaction on Power Systems, 25(4), 1823-1834.

Carter, C. K., and Kohn, R. (1994). “On Gibbs sampling for state space models.” Biometrika, 81, 541-553.

Cover, T.M., and Thomas, J.A. (2006). “Elements of information theory”, John Wiley and Sons, Inc.

Doshi-Velez, F. (2010). “The infinite partially observable Markov decision process.” Proc. Neural Information Processing Systems, 22.

Doshi-Velez, F, Pineau, J., and Roy, N. (2012). “Reinforcement learning with limited reinforcement: using Bayes risk for active learning in POMDPs.” *Journal of artificial Intelligence*, 187-188, 115-132.\

Durango, P.L., and Madanat, S.M. (2002). “Optimal maintenance and repair policies in infrastructure management under uncertain facility deterioration rates: an adaptive control approach.” Transportation Research Part A, 36, 763-778.

Fruhwirth-Schnatter, S. (2006). “Finite mixture and Markov switching models.” Springer Science + Business Media, LLC.

Gelman, A., Calin, J.B., Stem, H.S., and Rubin, D.B. (2004). “Bayesian data analysis.” *Chapman & Hall/CRC*, second edition.

Golabi, K., Kulkarni, R.B., and Way, G.B. (1982). “A statewide pavement management systems”, *Interfaces*, 12, 5-21.

Guignier, F., and Madanat, S. (1999). “Optimization of infrastructure systems maintenance and improvement policies”, *Journal of Infrastructure Systems*, 5(4), 124-134.

Guillaumot, V., Durango-Cohen, P., and Madanat, S. (2003). “Adaptive optimization of infrastructure maintenance and inspection decisions under performance model uncertainty.” *Journal of Infrastructure Systems*, 9(4), 133-139.

Heckerman, D., Horvitz, E., and Middleton, B. (1993). “An approximate nonmyopic computation for value of information.” *IEEE Trans. on Patter Analysis and Machine Learning*, 15, 292-298.

Hsu, D., Lee, W., and Rong, N. “What makes some POMDP problems easy to approximate”, *Advances in Neural Information Processing Systems (NIPS)*, Vancouver and Whistler, BC, Canada.

Jaulmes, R., Pineau, J., and Precup, D. (2005a). “Active learning in partially observable Markov decision processes.” European Conference on Machine Learning.

Jaulmes, R., Pineau, J., and Precup. (2005b). “Learning in non-stationary partially observable Markov decision processes.” European Conference on Machine Learning Workshop on Reinforcement Learning in Non-Stationary Environments.

Ji, S., Parr, R., and Carin, L. (2007). “Non-myopic multi-aspect sensing with partially observable Markov decision processes”, *IEEE Transactions on Signal Processing*, 55(6), 2720-2730.

Kaelbling, L.P., Littman, M.L., and Cassnadra, A.R. (1998). “Planning and acting in partially observable stochastic domain.” J. Artificial Intelligence, 101, 99-134.

Kemp, C., Perfors, A., and Tenenbaum, J.B. (2007). “Learning overhypotheses with hierarchical Bayesian models.” *Developmental Science*, 10(3), 307-331.

Kobayashi, H., Mark, B.L., and Turin, W. (2012). “Probability, random processes and statistical analysis.” *Cambridge University Press*, Cambridge, UK.

Koller, D., and Friedman, N. (2009). “Probabilistic Graphical Models.” The MIT Press, Cambridge, Massachusetts.

Konakli, K., Sudret, B., and Faber, M. (2015). “Numerical investigation into the value of information in life cycle analysis of structural systems.” *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 10.1061/AJRUA6.0000850, B4015007.

Krause, A., Singh, A., and Gusterin C. (2008). " Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies." *Journal of Artificial Intelligence Research*, 9, 235-284.

Krause, A., and Gusterin, C. (2009). "Optimal value of information in graphical models." *Journal of Artificial Intelligence Research*, 35, 557-591.

Kurniawati, H., Hsu, D., and Lee, W. (2008). "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces." Robotics: Science and Systems.

Li, B., and Si, J. (2007). "Robust dynamic programming for discounted infinite horizon Markov decision processes with uncertain stationary transition matrices." IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning.

Lovejoy, W.S. (1991). "Computationally feasible bounds for partially observed Markov decision process." Operations Research, 39(1), 162-175.

MacKay, D.J.C. (2003). "Information theory, inference and learning algorithms." *Cambridge University Press*, Cambridge, UK.

Madanat, S. (1993). "Optimal infrastructure management decisions under uncertainty." Transportation Research Part C, 1(1), 77-88.

Malings, C., Memarzadeh, M., and Pozzi, M. (2013). "Optimal topology of sensor networks for management of infrastructure systems." *6th International Conference on Structural Health Monitoring of Intelligent Infrastructure*, Hong Kong.

Marquez, F.P.G., Tobias, A.M., Perez, J.M.P., and Papaelias, M. (2012). "Condition monitoring of wind turbines: Techniques and methods." *Renewable Energy*, 46, 169-178.

McMillan, D., and Ault, W.G. (2008). "Condition monitoring benefit for onshore wind turbines: sensitivity to operational parameters." *IET Renewable Power Generation*, 2(1), 60-72.

Medury, A., and Madanat, S. (2013a). "Incorporating network considerations into pavement management systems: A case for approximate dynamic programming." *Transportation Research Part C*, 33, 134-150.

Medury, A., and Madanat, S. (2013b). "A Simultaneous Network Optimization Approach for Pavement Management Systems." *Journal of Infrastructure Systems*, DOI: 10.1061/(ASCE)IS.1943-555X.0000149.

Memarzadeh, M., Pozzi, M., and Kolter, J.Z. (2015a). "Optimal planning and learning in uncertain environments for the management of wind farms", *Journal of Computing in Civil Engineering*, 29(5), 04014076.

Memarzadeh, M., Pozzi, M., and Kolter, J.Z. (2015b). "Multiple uncertain POMDP: hierarchical modeling of a system with similar components", *Reliability Engineering and System Safety*, under review.

Memarzadeh, M., and Pozzi, M. (2015c). “Integrated inspection scheduling and maintenance planning for infrastructure systems”, *Computer-Aided Civil and Infrastructure Engineering*, DOI: 10.1111/mice.12178.

Memarzadeh, M. and Pozzi, M. (2015d). “Value of information in sequential decision making: component inspection, permanent monitoring and system-level scheduling”, in preparation.

Memarzadeh, M. and Pozzi, M. (2015e). “System-level inspection scheduling, an approach based on stochastic future allocation.” *10th IWSHM*, Stanford, CA.

Memarzadeh, M., Pozzi, M., and Kolter, J.Z. (2015f). “Hierarchical modeling of systems with similar components.” *ICASP12,* Vancouver, Canada.

Memarzadeh, M., Pozzi, M., and Kolter, J.Z. (2014). “Managing systems made up by similar components: a probabilistic framework for the maintenance of wind farms.” *6th WCSCM*, Barcelona, Spain.

Memarzadeh, M., Pozzi, M., and Kolter, J.Z. (2013). “Probabilistic learning and planning for optimal management of wind farms.” *9th IWSHM*, Stanford, CA, 2720-2728.

Mo, Y., Garone, E., Casavola, A., and Sinopoli, B. (2012a). “Stochastic sensor scheduling in wireless sensor networks with general graph topology.” American Control Conference, Fairmont Queen Elizabeth, Montreal, Canada.

Mo, Y., Sinpoli, B., Shi, L., and Garone, E. (2012b). “Infinite-Horizon sensor scheduling for estimation over lossy networks.” IEEE Conference on Decision and Control, Maui, Hawaii, USA.

Murphy, K.P. (2012). “Machine learning: a probabilistic perspective”, The MIT Press, Cambridge, Massachusetts.

Nielsen, J.S., and Sorensen, J.D. (2012). “Maintenance optimization for offshore wind turbines using POMDP.” Proc. 16th Conference of IFIP on Reliability and Optimization of Structural Systems, Yerevan Armenia, 175-182.

Nilim, A., and Ghaoui, L.E. (2005). “Robust solutions to Markov decision problems with uncertain transition matrices.” Operations Research, 53(5).

Papakonstantinou, K.G., and Shinozuka, M. (2014a). “Planning structural inspection and maintenance policies via dynamic programing and Markov processes. Part I: Theory.” *Reliability Engineering and System Safety*, 130, 202-213.

Papakonstantinou, K.G., and Shinozuka, M. (2014b). ‘Planning structural inspection and maintenance policies via dynamic programing and Markov processes. Part II: POMDP implementation.” Reliability Engineering and System Safety, 130, 214-224.

Papakonstantinou, K.G., and Shinozuka, M. (2014c). “Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models.” Probabilistic Engineering Mechanics, 37, 93-108.

Pineau, J., Gordon, G., and Thrun, S. (2003). “Point-based value iteration: An anytime algorithm for POMDPs.” Proc. of International Joint Conference on Artificial Intelligence, Acapulco, Mexico.

Powell, W.B. (2007). “Approximate Dynamic Programming: Solving the Curses of Dimensionality.” John Wiley & Sons, Inc.

Pozzi, M., and Der Kiureghian, A. (2011). “Assessing the value of information for long-term structural health monitoring.” Proce. SPIE9784, Health Monitoring of Structural and Biological Systems, doi: 10.1117/12.881918, San Diego, California, USA.

Raiffa, H., and Schlaifer, R. (1961). “Applied Statistical Decision Theory.” John Wiley & Sons, Inc., New York, NY, USA.

Robelin, C., and Madanat, S. (2007). “History-dependent bridge deck maintenance and replacement optimization with Markov decision processes”, *Journal of Infrastructure Systems,* 13(3), 195-201.

Ross, S., Pineau, B., Chaib-draa, B., and Kreitmann, P. (2011). “A Bayesian approach for learning and planning in partially observable Markov decision process.” J. Machine Learning Research, 12, 1729-1770.

Russell, S.J., and Norvig, P. (2010). “Artificial Intelligence: A Modern Approach, Third Edition.” Prentice Hall.

Schobi, R., and Chatzi, E.N. (2015). “Maintenance planning using continuous-state partially observable Markov decision processes and non-linear action models”, *Journal of Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*, DOI: 10.1080/15732479.2015.1076485.

Shani, G., Pineau, J., and Kaplow, R. (2013). “A survey of point-based POMDP solvers.” *Autonomous Agents and Multi-Agent Systems*, 27(1), 1-51.

Shi, L., Jia, Q.S., Mo, Y., and Sinopoli, B. (2011). “Sensor scheduling over a packet-delaying network” Automatica, 47, 1089-1092.

Smallwood, R.D., Sondik, E.J. (1973). “The optimal control of partially observable Markov processes over a finite horizon.” Operations Research, 21(5), 1071-1088.

Smilowitz, K., and Madanat, S. (2000). “Optimal inspection and maintenance policies for infrastructure networks.” *Computer-Aided Civil and Infrastructure Engineering*, 15, 5-13.

Sondik, E.J. (1978). “The optimal control of partially observable Markov processes over the infinite horizon.” Operations Research, 26(2), 282-304.

Straub, D., and Faber, M.H. (2004). “System effects in generic risk based inspection planning.” Journal of Offshore Mechanics and Arctic Engineering, 126(3), 265-271.

Straub, D., and Faber, M.H. (2005). “Risk based inspection planning for structural systems.” Structural Safety, 27, 335-355.

Straub, D., and Faber, M.H. (2006). “Computational aspects of risk-based inspection planning.” Computer-Aided Civil and Infrastructure Engineering, 21, 179-192.

Straub, D. (2014). “Value of information analysis with structural reliability methods.” Structural Safety, 49, 75-85.

Sutton, R. S., and Barto, A.G. (1998). “Reinforcement Learning: an introduction.” The MIT Press, Cambridge, Massachusetts.

Van Gael, J., Saatchi, Y., The, Y.W., and Ghahramani, Z. (2008). “Beam sampling for infinite hidden Markov model.” Int. Conf. on Machine Learning, 25.

Wilson, A., Fern, A., Ray, S., and Tadepalli, P. (2007). “Multi-task reinforcement learning: A hierarchical Bayesian approach”, *Proceedings of 24th International Conference on Machine Learning*, Corvallis, Oregon, USA.

Zhou, R., and Hansen, E.A. (2001). “An improved grid-based approximation algorithm for POMDPs.” Proc. Int. Joint Conf. on Artificial Intelligence.

Zonta, D., Glisic, B., and Adriaenssens, S. (2014). “Value of information: impact of monitoring on decision-making.” Structural Control and Health Monitoring, 21, 1043-1056.

Appendix A

Formulation of Fee-based Model

In this appendix we describe how to solve Eq. (28) by reformulating the two-step process of the FA model described there as an equivalent one-step process, modeled as a standard stationary POMDP, which can be processed by a solver as SARSOP (Kurniawati et al. 2007). To do so, we group functions $V_{(C)}^{\mathrm{I}}$ and $V_{(C)}^{\mathrm{M}}$, defined on the same belief domain, in a single value function defined on the duplicate belief domain, by augmenting the state. We define $s^+ = \{s, m\}$ as an augmented belief, where m is 1 for inspection sub-steps, and 2 for management ones. We add action inspect to the original set of actions, so that the resulting set of available action (A^+) is of dimensions $|A^+| = |A| + 1$. We force the management actions to be available in the management sub-steps only (and the inspection action to be available in the other sub-steps) by imposing unbearably high costs for untimely actions. Adjust discount factor.

In the inspection sub-steps, the agent has access to two actions: do-nothing (that we assume, without loss of generality, is also an option for the management sub-steps) or inspect. Cost is C for inspecting and nil for doing-nothing. Costs for other actions are assigned to be unbearable. Transition function T_F on the augmented state is assigned so the state stays the same, but index m moves from 1 to 2: for each actions a^+, it is formally defined as $T_F(\{s, 1\}, a^+, \{s', 2\}) = \delta_{ss'}$, where δ is the Kronecker Delta. Observations are defined on the joint domain of ordinary observations and inspection outcomes. We define augmented observation z^+ on domain Z^+ of

size $|Z^+| = |Z| + |H|$, so that z^+ maps to z for the first $|Z|$ entries, and to $|Z| + h$ for the last $|H|$ entries. Emission function O_F is nil when inspection state when $z^+ \leq |Z|$. For $z^+ > |Z|$, it is defined, for each state s as:

$$\begin{cases} O_{\mathrm{F}}(\{s,1\}, a^+ \ , z^+) = 1/|H| & a^+ < |A^+| \\ O_{\mathrm{F}}(\{s,1\}, |A^+|, z^+) = E(s, z^+ - |Z|) \end{cases} \quad \text{(A1)}$$

Cost function R_F is nil for the do-nothing action ($a^+ = 1$), except, for each state $R_F(\{s,1\}, |A^+|) = C$. For other actions, the cost is unbearable.

For management sub-step and $a^+ < |A^+|$, transition and emission parameters copies those of the original POMDP, as $T_F(\{s,2\}, a^+, \{s',1\}) = T(s,a,s')$ and $O_F(\{s,2\}, a^+, z^+) = O(s,a,z)$. Cost is updated as $R(\{s,2\}, a^+) = R(s,a)/\sqrt{\gamma}$ and the overall discount factor is $\gamma^+ = \sqrt{\gamma}$. We also assign to $R(\{s,2\}, |A^+|)$ an unbearable cost, so that the other parameters for the inspection action taken from a management sub-step are irrelevant. The discount factor is updated to take into account that the first management step should not be discounted. Hence the model parameters of the FA-model POMDP can be grouped as follow: $\boldsymbol{\Theta}^{\mathrm{F}}_{(C)} = \{\mathbf{T}_{\mathrm{F}}, \mathbf{O}_{\mathrm{F}}, \mathbf{R}_{\mathrm{F}}, \gamma^+\}$. The corresponding value $G^*_{(C)}$, starting at belief $\mathbf{b}$ in the sub-inspection step, can be computed using augmented belief $\mathbf{b}^+$, defined as $b^+\{s,1\} = b(s)$, $b^+\{s,2\} = 0$. Also, value $V^{\mathrm{M}}_{(C)}$ at management sub-step can be computed from augmented belief defined as $b^+\{s,2\} = b(s)$, $b^+\{s,1\} = 0$ but, in this case, the resulting value from the POMDP value has to be multiply by factor $\sqrt{\gamma}$. In summary:

$$\begin{cases} V^{\mathrm{I}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}) = V^*\left(\mathbf{b}, \boldsymbol{\Theta}^{\mathrm{F}}_{(C)}\right) & a^+ < |A^+| \\ V^{\mathrm{M}}_{(C)}(\mathbf{b}, \boldsymbol{\Theta}) = \sqrt{\gamma}\, V^*\left(\mathbf{b}, \boldsymbol{\Theta}^{\mathrm{F}}_{(C)}\right) \end{cases} \quad \text{(A2)}$$

Appendix B

Formulating and Solving System-level POMDP

In this appendix we describe how to formulate the system-level problem of Section 6.2.2 into the POMDP framework presented in Chapter 2. The main difficulty to doing so is to convert a two-step process, alternating inspection scheduling and maintenance, into a process with uniform steps. To do so, each time step is divided into two sub-steps and, to distinguish between them, we define a binary indicator m, with possible values $m = 1$ for odd sub-steps and $m = 2$ for even ones. The intention is to use the odd sub-steps for inspections and, after receiving inspectors' outcomes, select maintenance actions at even sub-steps, receiving reward and observation. Complete state, action, and observation are defined as $s^{++} = s^{+} \cup m$, $a^{++} = a^{+} \cup Y$, and $z^{++} = z^{+} \cup h^{+}$, where h^{+} lists all inspection outcomes, on domains of size $|S^{++}| = 2|S^{+}| = 2|S|^{N}$, $|A^{++}| = \binom{N}{K} + |A|^{N}$, and $|Z^{++}| = |H|^{K} + |Z|^{N}$ respectively. In this notation, while each possible value of state s^{++} is composed of two components (s^{+} and m), actions and observations either belong to one or the other sub-domains, referring to inspection (h^{+} and Y) or to maintenance (z^{+} and a^{+}).

Complete transition probability function $T^{++}: S^{++} \times A^{++} \times S^{++} \to [0,1]$ is zero if $m_{t+1} = m_t$; $T^{++}(\{s_t^{+}, 1\}, a^{+}, \{s_{t+1}^{+}, 2\})$ is zero if $s_{t+1}^{+} \neq s_t^{+}$ and one if $s_{t+1}^{+} = s_t^{+}$; and $T^{++}(\{s_t^{+}, 2\}, a^{+}, \{s_{t+1}^{+}, 1\}) = \prod_{i=1}^{N} T(s_{i,t}, a_i, s_{i,t+1})$.

Complete emission probability function is defined as: $O^{++}: S^{++} \times A^{++} \times Z^{++} \to [0,1]$. $O^{++}(\{s_{t+1}^{+}, 2\}, a^{+}, h^{+}) = 0$, and $O^{++}(\{s_{t+1}^{+}, 2\}, a^{+}, z^{+}) = \prod_{i=1}^{N} O(s_{i,t}, a_i, z_{i,t})$; similarly $O^{++}(\{s_{t+1}^{+}, 1\}, a^{+}, z^{+}) = 0$, and $O^{++}(\{s_{t+1}^{+}, 1\}, a^{+}, h^{+}) = \prod_{i=1}^{N} G(s_{i,t}, h_{i,t})$.

The discount factor is the square root of the γ for the component-level POMDP (because we have discretized each time step into two sub-steps).

The reward function is defined as $R^{++}: S^{++} \times A^{++} \to \mathbb{R}$. $r^{++}(\{s_t^{+}, 1\}, a^{+}) = -\infty$, because at odd sub-steps we force the policy to inspect and $r^{++}(\{s_t^{+}, 1\}, Y) = 0$. Similarly, we force the policy to take maintenance actions in even sub-steps, so $r^{++}(\{s_t^{+}, 2\}, Y) = -\infty$ and $r^{++}(\{s_t^{+}, 2\}, a') = \frac{1}{\sqrt{\gamma}} \sum_{i=1}^{N} r(s_{i,t}, a_{i,t})$. The factor before the sum compounds the rewards to the first even step.

Initial belief models the knowledge that the system starts in an odd-step, so $b_0^{++}(\{s_0^{++}, 1\}) = \prod_{i=1}^{N} b_0(s_{i,0})$ and $b_0^{++}(\{s_0^{++}, 2\}) = 0$.

Once these parameters are defined, POMDP solvers can be used. As illustrated above, the computational complexity of the *system-level* POMDP problem grows exponentially with the number of components in the system and it is not tractable for most of real-world applications.

Appendix C

Proofs of Bounds

It can be proven that the VoI is always non-negative (Heckerman et al. 1993), according to the principle that Information Never Hurts (INH).

Proof that:

$$W_*^{(P)} \leq U^{(P)} \leq W_{Y^{(P)}}^{(O)} \tag{C1}$$

Let us start proving that $W_*^{(P)} \leq U^{(P)}$. Suppose no inspector is available from next time step. Then, $W_*^{(P)} = U^{(P)}$ based on the definition in Eq. (41) and the fact that the pessimistic agent acts optimally according to her assumption. During the actual management process following the pessimistic agent's policy, future observations from inspectors will be available (despite the assumption of the pessimistic agent) and based on INH we can conclude that $W_*^{(P)} \leq U^{(P)}$. Now let us prove that $U^{(P)} \leq W_{Y^{(P)}}^{(O)}$. Based on INH, $U^{(P)}$ is always less than or equal to the optimistic system-level value estimate, inspecting all components whose indices are in set $Y^{(P)}$. In other words, we can conclude that $U^{(P)} \leq W_{Y^{(P)}}^{(O)}$.

Proof that:

$$U^{(O)} \leq W_*^{(O)} \tag{C2}$$

Suppose inspectors are available for all the components of the system from the next time step. Then from Eq. (41) we have $U^{(O)} = W_*^{(O)}$. During the actual management process following the optimistic agent's policy, future observations will not be available for some components if $K \leq N$, and because of INH (i.e., lack of information never helps) we can conclude that $U^{(\mathrm{O})} \leq W_*^{(\mathrm{O})}$.

Proof that:

$$W_*^{(\mathrm{P})} \leq U^* \leq W_*^{(\mathrm{O})} \tag{C3}$$

Let us first prove the upper bound. Suppose the current set of indices of the components to be inspected, selected by the optimal agent (that solves the system-level problem and finds the exact solution) is Y^*. Then from Eq. (43) we can infer that $W_{Y^*}^{(\mathrm{O})} \leq W_*^{(\mathrm{O})}$. Next, from INH we can infer that $U^* \leq W_{Y^*}^{(\mathrm{O})}$, and hence $U^* \leq W_*^{(\mathrm{O})}$. Now let us prove the lower bound of Eq. (C3). We know that by inspecting the components in the set $Y^{(\mathrm{P})}$, and under the pessimistic assumption (that inspectors are not available from next time step) agent gets $W_*^{(\mathrm{P})}$. By the optimality of the agent's policy, the availability of future inspectors and the INH principle, we conclude that $W_*^{(\mathrm{P})} \leq U^*$.

Appendix D

Analytical Examples

We present two simple examples, for which analytical solutions are available, to illustrate how pessimistic or optimistic approaches can do better depending on the context.

Example 1: In this example we show how the optimistic approach may lead to huge losses, even when ratio K/N is arbitrary close to one. Consider a set of components with equal transition probabilities and initial belief. Five states are possible, initial state is 1, with certainty, and absorbing states 4 and 5 represent failure and disposal respectively. Four actions are available, namely *Wait* (*W*), *Dispose* (*D*), *Dispose-from-2* (*A*) and *Dispose-from-3* (*B*). The transition graph of the component is shown in Figure 48.

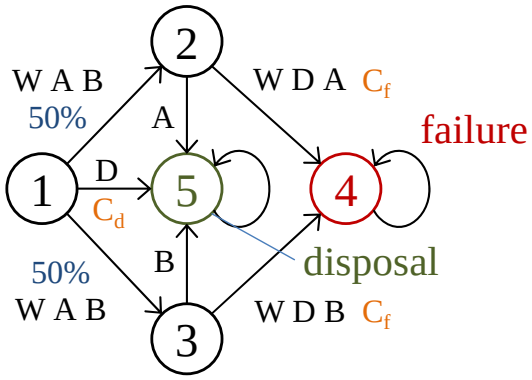

Figure 49. Transition graph for example 1 in Appendix D.

From 1, state can transit to 2 or 3 with uniform probability at no cost under any actions *W*, *A* and *B*, while action *D* takes to state 5 with disposal cost C_D. From 2 or 3, state can only move to

5, at a cost that depends on the action: from 2, failure cost $C_F \gg C_D$ has to be paid for all actions except *A*, and no cost is due for this latter action; similarly, from 3 only one action (*B*) avoids the failure cost. The only available observations are those from perfect inspectors, and discount factor is unitary. In this setting, no cost has to be paid for a component inspected at time t_1.

At time zero the pessimistic agent, assuming she will not have access to any inspectors from the next time step onwards, disposes all components, paying cost $U^{(\mathrm{P})} = W_*^{(\mathrm{P})} = NC_D$. The optimal agent knows that she has access to K inspectors at each time step, hence she disposes only $N - K$ components and her optimal value is $U^* = (N - K)C_D$.

The optimistic agent assumes inspections for all components; hence she will wait (*W*) until next time step, computing $W_*^{(\mathrm{O})} = 0$. However at time t_1 she has access to only K inspectors, and she has to select a risky action (say between *A* and *B*) for all uninspected components, so that value is $U^{(\mathrm{O})} = (N - K)C_F/2$. Note that if the cost of failure is large, the value of optimistic agent can fall well below the lower bound ($W_*^{(\mathrm{P})}$) defined for the pessimistic approach. For any values of N, K and C_D, we can find a C_F so that relative benefit, defined by the following equation:

$$U^{(\mathrm{P})} - U^{(\mathrm{O})} = NC_D - (N - K)C_F/2 \qquad \text{(D1)}$$

is positive, and the pessimistic agent does better.

The myopic planning exposes the optimistic agent to a relevant probability of failure, independent of how high failure cost C_F is. As shown by the pessimistic agent, failure events can be easily avoided by timely disposal, and no failure is foreseen in the computation of either $W_*^{(\mathrm{P})}$ or $W_*^{(\mathrm{O})}$, but the optimistic approach suggests the agent to postpose the disposal until it is

too late. It also should be noted that failures can follow optimistic planning for any N and $K < N$, and thus for arbitrary high values of the ratio K/N.

Example 2: The second example aims to show how the optimistic approach can outperform the pessimistic one even if ratio K/N is arbitrary low. We consider a system made up of two different kinds of component, with different transitions and rewards. We call one of the components "critical", while the remaining $N - 1$ are "dummy" components. The agent can use only one perfect inspector on the critical or on a dummy component. Figure 49(a) shows the transition graph for dummy components. Dummy components have binary states (state 1 and 2) independent for each time step, and the initial beliefs for all dummy components are uniform. Actions for these components are also binary (action A and B2), and can be understood as "guessing the state": if the guess is wrong, a fee $2f$ is paid, and if it is correct no fee is due. No observations are available beyond those from the inspector. Consequently, the agent has to pay expected fee f for any dummy component left uninspected. As inspecting the critical component implies not inspecting a dummy one, f can be intended as an equivalent expected fee for inspecting the critical component. The latter component is degrading up to failure related to unbearable consequences; however, failure time can be predicted with certainty, once it has been inspected. That component can be interpreted as a time-bomb, and inspecting it is equivalent to reading the timer. It can be disposed at a cost C_D, to avoid the unbearable cost of failure $C_F = \infty$. We assume discount factor $\gamma < 1$.

We anticipate the moral of the example: all agents agree that the critical component has to be inspected, but they disagree on scheduling: pessimistic prefers to inspect at time zero, as she assumes she cannot do it in the future, while optimistic can adopt the optimal action, postponing inspection until actually needed.

The transition graph of the critical component is reported in Figure 49(b).

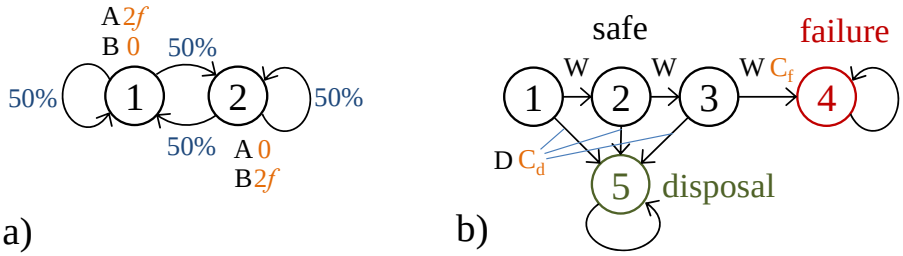

Figure 50. Transition graphs for example 2 in Appendix D: (a) dummy and (b) critical component.

The critical component's state is defined over 5 possible values: states $\{1,2,3\}$ are all safe, state 4 represents the failure, and state 5 the disposal. Two actions are available: namely *Wait* (*W*) and *Dispose* (*D*). Under *Wait*, state moves deterministically so that $s_{t+1} = s_t + 1$, up to failure. The failure and disposal states are absorbing states. Up to state 3, to *Dispose* takes component to state 5. Initial belief state is uniform between 1 and 2: this means that component is safe up to time $t = 1$, but at time $t = 2$ it may fail.

The pessimistic agent inspects the critical component at $t = 0$, and is able to dispose it when in state 3. The optimistic agent disposes it at the same time, after having waited until $t = 1$ to inspect it. This is indeed the optimal policy, and the benefit with respect to the pessimistic value is:

$$U^{(\mathrm{P})} - U^{(\mathrm{O})} = (1-\gamma)f > 0 \qquad \text{(D2)}$$

This result is invariant with respect to the number of dummy components, $N - 1$, and this proves that the optimistic assumption can provide the optimal policy, even when ratio K/N is arbitrary low.

The reader may note that, in the presented example, the ratio between the number of inspectors and critical components is actually one and, consequently, it is not a surprise that the

optimistic assumption is correct. However, our point is to disprove the conjecture that any conclusion on what agent performs better can be based on ratio K/N. It may be possible to define an adjusted ratio, identifying a sub-set of “critical” components. However, it is an open question how to define this feature in a general context.