Hidden-Model Processes for Adaptive Management under Uncertain Climate Change

Matteo Pozzi, A.M.ASCE[1]; Milad Memarzadeh[2]; and Kelly Klima[3]

Abstract: Predictions of climate change can significantly affect the optimization of measures reducing the long-term risk for assets exposed to extreme events. Although a single climate model can be represented by a Markov stochastic process and directly integrated into the sequential decision-making procedure, optimization under epistemic uncertainty about the model is computationally more challenging. Decision makers have to define not only a set of models with corresponding probabilities, but also whether and how they will learn more about the likelihood of these models during the asset-management process. Different assumed *learning rates* about the climate can suggest opposite behaviors. For example, an agent believing, optimistically, that the correct model will soon be identified may prefer to wait for this information before making relevant decisions; on the other hand, an agent predicting, pessimistically, that no further information will ever be available may prefer to immediately take actions with long-term consequences. This paper proposes a set of optimization procedures based on the Markov decision process (MDP) framework to support decision making depending on the assumed learning rate, thus trading off the need for a prompt response with that for reducing uncertainty before deciding. Specifically, it outlines how approaches based on the MDP and hidden-mode MDPs, dynamic programming, and point-based value iteration can be used, depending on the assumptions on future learning. The paper describes the complexity of these procedures, discusses their performance in different settings, and applies them to flood risk mitigation. **DOI: 10.1061/(ASCE)IS.1943-555X.0000376.**

Introduction

Climate change poses specific challenges to long-term asset and infrastructure management. Much of today's planning should consider future climate conditions, even decades in advance, which are currently unknown and will progressively reveal themselves. On the one hand, this may suggest postponing critical and expensive decisions until epistemic uncertainty about climate change evaporates. On the other hand, procrastination may expose decision makers to high risk in the short term and prevent them from acting promptly.

This paper proposes and discusses optimization models for trading off quantitatively the need for a prompt response and the need for collecting information to reduce uncertainty before making decisions. The motivating application is protection of infrastructure assets against flooding and sea-level rise. Research indicates that climate change may cause a worldwide sea level rise of one meter or more by 2100 (IPCC 2014; Melillo et al. 2014), resulting in more coastal flooding events worldwide (Field et al. 2012). This prediction is affected by significant uncertainty; Hallegatte et al. (2012) summarized methods that might be appropriate for decision making under deep uncertainty, including cost-benefit analysis under uncertainty, cost-benefit analysis with real options, robust decision making, and climate informed decision analysis. The National Climate Assessment report (Melillo et al. 2014) also discussed approaches and stressed the importance of understanding the decision makers' preferences and application. Dittrich et al. (2016) built on this, comparing real option analysis, portfolio analysis, robust decision making, and no-regret/low-regret options and discussing when each of these may be an appropriate decision-making tool.

Although those approaches are suitable for identifying robust policies for decision processes of one or a few stages, this paper considers an agent aiming at optimal long-term planning, sequentially adapting their policy depending on the available information. This is a risk-neutral agent ready to assign a distribution to an exhaustive set of climate evolutions, including model and scenario uncertainty as defined by Hawkins and Sutton (2009); models and distribution in turn depend on assumed emissions scenarios that the agent cannot control. The problem setting belongs to classic decision theory as defined by von Neumann and Morgenstern (1944), and excludes *deep uncertainty* that cannot be explicitly modeled by any distribution.

Long-term asset management under climate change can be formulated as a sequential decision-making problem with nonstationary stochastic models governing the system evolution. The Markov decision processes (MDPs) framework (Bertsekas 1995; Sutton and Barto 1998) can model the planning process under a single climate model when the state of the system is completely observable. However, it is often the case that the decision maker, possibly following the expert evaluation of the scientific community, cannot identify one model that completely represents the uncertainty of the climate evolution. When the model is itself unknown, the appropriate planning approach should depend on the assumption of information availability. In the two limit cases in which either perfect information about the model will be available soon or no information will ever be available, simple solutions can be found. When only noisy

[1]Faculty, Dept. of Civil and Environmental Engineering, Faculty Affiliate, Scott Institute for Energy Innovation, Carnegie Mellon Univ., 107b Porter Hall, 5000 Forbes Ave., Pittsburgh, PA 15213-3890 (corresponding author). E-mail: mpozzi@cmu.edu

[2]Postdoctoral Scholar, Dept. of Environmental Science, Policy and Management, Univ. of California Berkeley, 201 Wellman Hall, Berkeley, CA 94720. E-mail: miladm@berkeley.edu

[3]Adjunct Assistant Professor, Research Scientist, Dept. of Engineering and Public Policy, Scott Institute for Energy Innovation, Carnegie Mellon Univ., Baker 129, 5000 Forbes Ave., Pittsburgh, PA 15213-3890. E-mail: kelly.klima@gmail.com

Note. This manuscript was submitted on July 12, 2016; approved on March 7, 2017; published online on June 23, 2017. Discussion period open until November 23, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

or incomplete observations of the model are progressively available in time, a decision maker should use the hidden-mode MDP (HM-MDP) framework, as introduced by Chades et al. (2012). Hidden-mode MDPs are a special case of the partially observable Markov decision processes (POMDPs) (Smallwood and Sondik 1973; Sondik 1978); POMDPs have been extensively used for infrastructure management under state uncertainty (Madanat 1993; Papakonstantinou and Shinozuka 2014; Memarzadeh and Pozzi 2016a, b) and also under state and model uncertainty (Memarzadeh et al. 2015, 2016), mostly with stationary models. Shani et al. (2013) reviewed efficient numerical methods for solving POMDPs, including Perseus (Spaan and Vlassis 2005) and SARSOP (Kurniawati et al. 2008).

A HM-MDP assumes that the state of the system is fully observable, that the persistent hidden dynamic model is only partially observable, and that the model and the available information are not affected by the agent's actions. Chades et al. (2012) applied their approach to the management of a population of the threatened bird species endemic to Northern Australia. Špačková and Straub (2017), leveraging their previous work on flexibility in planning (Špačková and Straub 2015), investigated a similar setting with nonstationary models of climate change and proposed a solution method based on Monte Carlo methods, quantifying the benefit of flexibility, with example applications to a wastewater treatment plant and to flood protection.

This paper specifically focuses on the role of the assumed availability of future information in current planning, illustrating how, depending on the *learning rate* (i.e., the assumed information availability), HM-MDP can be formulated and solved. It considers exact methods under limit-case assumptions and approximate methods for intermediate cases.

Planning under Known Model: MDP

Consider an asset to be managed under climate change. Time is discretized into set $\{t_1, t_2, t_3, \ldots, t_T\}$, and S_k defines the asset's state at time t_k on domain $\{1,2, \ldots, |S|\}$. At time t_k, the manager takes action A_k on domain $\{1,2, \ldots, |A|\}$, and pays immediate cost C_k that depends on the current action and state by time-dependent function $C_k(S_k, A_k)$. Thus the expected discounted cost V_k (the *value*) for managing the system from time t_k to the time horizon t_T is

$$V_k = \sum_{t=k}^{T} \gamma^{t-k} \mathbb{E}[C_t] + \gamma^{T+1-k} V_{T+1} \tag{1}$$

where γ = one-step discount factor; $\mathbb{E}[X]$ indicates the expectation of random variable X; and V_{T+1} = residual cost one step after the time horizon. Although the state may be only partial observable in some applications, in the MDP framework it is completely observable. The state stochastically evolves from value i to j following the Markov property, according to time-varying transition function $T_k(i, a, j) = \mathbb{P}[S_{k+1} = j | S_k = i, A_k = a]$, where $\mathbb{P}[E|F]$ indicates the probability of event E conditional to event F. Because the state evolution and consequently the future costs depend on the action taken, the expectation in Eq. (1) can be computed only when a specific policy is assigned. In an MDP, current state S_k represents a sufficient statistic, and the decision maker can adopt a policy $A_k = \pi_k(S_k)$ depending on that state only. Following time-varying policy $\Pi = \{\pi_1, \ldots, \pi_T\}$ starting from state $S_k = i$, the agent obtains value

$$V_k^{\Pi}(i) = C_k[i, \pi_k(i)] + \gamma \sum_{j=1}^{|S|} T_k[i, \pi_k(i), j] V_{k+1}^{\Pi}(j) \tag{2}$$

The optimal value is obtained by minimizing Eq. (2)

$$V_k^*(i) = \min_a \left\{ C_k(i, a) + \gamma \sum_{j=1}^{|S|} T_k(i, a, j) V_{k+1}^*(j) \right\} \tag{3}$$

Optimal policy $\pi_k^*(i)$, at time t_k, is defined by using argmin instead of min in Eq. (3), and time-varying policy set Π^* is defined by listing policies for all times. An agent adopting Π^* obtains the minimum possible value (it is worth stressing that this expresses an expected quantity). Eqs. (2) and (3) are forms of the Bellman equation, and they can be solved iteratively, from $k = T$ back to $k = 1$, each iteration being an application of the so-called Bellman operator.

Planning under Model Uncertainty

Extend the previous formulation considering a set of M possible models, each describing a persistent stochastic behavior. Model indicator $\mathcal{M}$ assumes one value in domain $\{1,2, \ldots, M\}$, and time-varying functions $C_{k,m}$ and $T_{k,m}$ define the cost and transition, respectively, for model m. By solving Eq. (3) for each model, one can identify a set of M policies $\{\Pi_1^*, \ldots, \Pi_M^*\}$, that might disagree even at initial time t_1. Thus the agent has to consider all models jointly and make a decision accounting for the uncertainty among them. Assume the agent assigns belief $\mathbf{b}$ to the models, so that $\mathbf{b}(m) = \mathbb{P}[\mathcal{M} = m]$ defines the probability that model m is the correct one. One approximate planning approach is to derive a single expected model by averaging transition probabilities and immediate costs over the belief, and then to apply Eq. (3) to that model. The following section introduces an alternative approach that accounts for the persistency of the model.

Robust Planning via Open-Loop Control (Pessimistic Policy)

The open-loop control scheme identifies an optimal time-varying policy as that which performs best, in the expected sense, with no belief updating during the process. Now $V_{k,m}^{\Phi}$ indicates the value function according to model m, at time t_k, following policy $\Phi = \{\phi_1, \ldots, \phi_T\}$, which can be identified by Eq. (2), using functions $C_{k,m}$ and $T_{k,m}$ instead of C_k and T_k. The corresponding value under model uncertainty, $W_{k,\infty}^{\Phi}$ can be defined as

$$\begin{aligned} W_{k,\infty}^{\Phi}(i, \mathbf{b}) &= \mathbb{E}_m V_{k,m}^{\Phi}(i) \\ &= \mathbb{E}_m C_{k,m}[i, \phi_k(i)] \\ &\quad + \gamma \sum_{j=1}^{|S|} \mathbb{E}_m \{T_{k,m}[i, \phi_k(i), j] V_{k+1,m}^{\Phi}(j)\} \end{aligned} \tag{4}$$

where $\mathbb{E}_m[f_m] = \sum_{m=1}^{M} f_m b(m)$ indicates the expectation of m using belief $\mathbf{b}$, and subscript (k, ∞) indicates that the value is computed at time t_k and policy is prescribed up to an infinite horizon. The minimum expected cost achievable by open-loop control, $W_{k,\infty}^*$, is defined as $W_{k,\infty}^*(i, \mathbf{b}) = \min_{\Phi} W_{k,\infty}^{\Phi}(i, \mathbf{b})$ that is optimal when information on the model will never be available or, equivalently, will be available infinitely far ahead. Computationally, this value can be identified by iteratively applying the Bellman operator, from terminal time t_T, following the scheme

$$\begin{aligned} W_{k,\infty}^*(i, \mathbf{b}) = \min_a \Big\{ & \mathbb{E}_m C_{k,m}(i, a) \\ & + \gamma \sum_{j=1}^{|S|} \mathbb{E}_m [T_{k,m}(i, a, j) V_{k+1,m}^{\Phi_{k,\infty}^*}(j)] \Big\} \end{aligned} \tag{5}$$

where the optimal policy $\Phi^*_{k,\infty} = \{\phi^*_{k,\infty}, \ldots, \phi^*_{T,\infty}\}$ derives from using argmin instead of min. This open-loop approach is truly optimal when no information on the model is available, and the agent's belief is time-invariant. Actually, this is rarely the case, because perfect observation of the state trajectory or of the actual costs inevitably contains information about the model. However, it may serve as an effective approximation in *pessimistic* scenarios, where information about the model is negligible.

Near-Clairvoyance Planning via Action-Value Function (Optimistic Policy)

The alternative *near-clairvoyance* optimization scheme assumes that the model will be revealed at next step (Memarzadeh et al. 2015). It is based on the so-called *action-value function* (Sutton and Barto 1998) for model m at time t_k that corresponds to the content of the curly brackets in Eq. (3)

$$Q_{k,m}(i,a) = C_{k,m}(i,a) + \gamma \sum_{j=1}^{|S|} T_{k,m}(i,a,j) V^*_{k+1,m}(j) \quad (6)$$

so that Eq. (3) can be rewritten as $V^*_{k,m}(i) = \min_a Q_{k,m}(i,a)$. The action-value function defines the expected management cost when action a is taken at the current time, followed, from time t_{k+1} onward, by the residual part of optimal policy $\Pi^*_m = \{\pi^*_{1,m}, \ldots, \pi^*_{T,m}\}$. Whether or not functions are available for each model and action, one can compute a new value following policy Φ for one step, as

$$W^{\Phi}_{k,k+1}(i,\mathbf{b}) = \mathbb{E}_m\{Q_{k,m}[i,\phi_k(i)]\} = \sum_{m=1}^{M} Q_{k,m}[i,\phi_k(i)] b(m) \quad (7)$$

where subscript $(k, k+1)$ indicates that value is computed at time t_k, and the policy will switch at time t_{k+1}. The corresponding optimal value is defined as

$$W^*_{k,k+1}(i,\mathbf{b}) = \min_a \mathbb{E}_m\{Q_{k,m}(i,a)\} = \min_a \sum_{m=1}^{M} Q_{k,m}(i,a) b(m) \quad (8)$$

Using argmin instead of min in Eq. (8) gives the corresponding optimal policy $\Phi^*_{k,k+1} = \{\phi^*_{k,k+1}, \pi^*_{k+1,m}, \ldots, \pi^*_{T,m}\}$, where m now identifies the revealed model. The near-clairvoyance approach is optimal under the assumption that perfect information on the model will be available at the next time step, and therefore it can be considered as an *optimistic* policy. Under that optimistic assumption, $W^*_{k,k+1}$ represents the actual value the agent obtains. It should be noted that the time discretization plays a key role in identifying the conditions that make the near-clairvoyance policy optimal because, depending on that, *the next time step* refers to different times (e.g., 1 day, or 5 years in the future).

Mixed Planning: d-Step-Ahead Clairvoyance Policy

By merging previous approaches, the agent can adopt policy Φ up to time t_{k+d-1} and switch to a single-model optimal policy from time t_{k+d}, counting on perfect information at that time. To assess the corresponding value, one can apply the single-model Bellman operators of Eq. (3) from T back to time t_{k+d}, and the open-loop operator of Eqs. (2) and (4) from time t_{k+d-1} back to time t_k, obtaining value $W^{\Phi}_{k,k+d}(i,\mathbf{b})$. Combining Eqs. (3) and (5) gives the optimal value $W^*_{k,k+d}(i,\mathbf{b})$ and corresponding policy $\Phi^*_{k,k+d} = \{\phi^*_{k,k+d}, \ldots, \phi^*_{k+d-1,k+d}, \pi^*_{k+d,m}, \ldots, \pi^*_{T,m}\}$, where, again, m identifies the revealed model.

This *d-step-ahead clairvoyance policy* is optimal when perfect information is revealed after d steps and no information is available before that. By considering parameter d varying from 1 to ∞ (actually, up to $T+1-k$), these policies vary from near-clairvoyance to open-loop.

Comparison of Near-Clairvoyance and Open-Loop Policies

Agents should choose among the policies outlined previously by considering whether and when perfect information on the model will be available. Knowing that the model will be revealed at time t_{k+d}, they will adopt, from time t_k, policy $\Phi^*_{k,k+d}$ for expected discounted cost $W^*_{k,k+d}$. Because "information never hurts" (Krause 2008), having perfect information at an earlier stage is always better (strictly speaking, it is not worse), thus if $k' < k''$ then $W^*_{k,k'} \le W^*_{k,k''}$. However, optimistic policy $\Phi^*_{k,k'}$ may perform better or worse than more pessimistic policy $\Phi^*_{k,k''}$, depending on the actual availability of information. The path followed by Memarzadeh and Pozzi (2016a) allows for discussing bounds on the performance of alternative policies in different settings and, specifically, for comparing the open-loop and the near-clairvoyance policies under opposite scenarios. On the one hand, it is reasonable to assume that an agent receiving perfect information, even before the predicted time, will switch to the corresponding single-model optimal policy. On the other, it is less clear how one reacts to not receiving the assumed perfect information at the predicted time. Assume that if an agent following the near-clairvoyance policy does not receive information at the next step, she or he will predict receiving perfect information at the following step. Doing so, the agent follows policy $\Phi'_{k,+1} = \{\phi^*_{k,k+1}, \phi^*_{k+1,k+2}, \ldots, \phi^*_{T,T+1}\}$. By relying on the principle that information never hurts, and the definition of optimal policy, the following sequence of inequality holds:

$$W^*_{k,k+1}(i,\mathbf{b}) \le W^{\Phi^*_{k,\infty}}_{k,k+1}(i,\mathbf{b}) \le W^*_{k,\infty}(i,\mathbf{b}) \le W^{\Phi'_{k,+1}}_{k,\infty}(i,\mathbf{b}) \quad (9)$$

Eq. (9) means that the cost of using the near-clairvoyance policy under perfect information (at the next step) is less than that of using the open-loop policy under perfect information, which is less than that of using the open-loop policy under no information, which is less than that of using the near-clairvoyance policy under no information. The three less-than-or-equal-to signs in Eq. (9), from right to left, can be justified by noting that they refer to *a better policy under the same information*, *more future information under the same initial policy* and, again, *a better policy under the same information*, respectively.

From time t_k, the increment of cost using the near-clairvoyance policy without actual information (with respect to using the appropriate open-loop policy), ΔV^{I}_k, and that of using the open-loop policy when perfect information is available at the next step (with respect to using the appropriate clairvoyance policy), ΔV^{II}_k, are obtained as

$$\Delta V^{\mathrm{I}}_k(i,\mathbf{b}) = W^{\Phi'_{k,+1}}_{k,\infty}(i,\mathbf{b}) - W^*_{k,\infty}(i,\mathbf{b})$$
$$\Delta V^{\mathrm{II}}_k(i,\mathbf{b}) = W^{\Phi^*_{k,\infty}}_{k,k+1}(i,\mathbf{b}) - W^*_{k,k+1}(i,\mathbf{b}) \quad (10)$$

where functions $W^{\Phi'_{k,+1}}_{k,\infty}$ and $W^{\Phi^*_{k,\infty}}_{k,k+1}$ can be evaluated as illustrated in Appendix 1.

Computation of these cost increments (or so-called *regrets*) may shed light on the sensitivity to the available information depending on the adopted policy. Because of Eq. (9), regret ΔV^{II}_k can be bounded from above by $W^*_{k,\infty} - W^*_{k,k+1}$, which is easily computed and represents the so-called value of information for having perfect

model observation at the next step, with respect to not having any information. Regret ΔV^{I}, on the other hand, cannot be easily bounded from above, as shown in Eq. (9); this is related to the discussion in Memarzadeh and Pozzi (2016a) about the potential high loss occurring when anticipated information does not become available.

HM-MDP: Model Observation and Updating

Previous sections assumed perfect information on the model at one time during the process. In most cases, however, information flows more smoothly, and the general HM-MDP setting assumes that indirect observations of the model can be available at each time. Observation Y_k at time t_k on domain $\{1,2, \ldots, |Y|\}$ is defined by observation probability $O_k(m,h) = \mathbb{P}[Y_k = h|\mathcal{M} = m]$. The set of $|Y|$ possible observations is not necessarily mapped to a time-invariant set of physical measurements; for example, the statement $Y_k = 1$ may refer to different measurement values depending on k. As for Chades et al. (2012), actions do not affect the model, nor the flow of information.

Beliefs about the model can be updated sequentially following Bayes' formula. If Y_k assumes value h, the update at time t_k is

$$b_{k+1}(m) \propto O_k(m,h) b_k(m) \tag{11}$$

where now $\mathbf{b}_1$ = initial belief; and posterior belief $\mathbf{b}_k$ at time t_k is described for $k > 1$ as $b_k(m) = \mathbb{P}[\mathcal{M} = m|Y_2, \ldots, Y_k]$. In principle, belief should also be a function of the previous trajectory of actions and states, which contain information about the climate model. However, this contribution is considered to be negligible.

Fig. 1 shows a decision graph of the management process, following the corresponding HM-MDP framework, in which time flows from left to right. As in the traditional notation of probabilistic graphical models, circles represent random variables, squares represent decision variables, diamonds represent costs, continuous links define the conditional dependence structure, and dotted arrows indicate what information is available before a decision is made. Shaded variables are observable. No observation is considered at time t_1 because all relevant information collected at that time can be embedded in initial belief $\mathbf{b}_1$. Although the immediate costs are also a function of the model, for readability of the graph the corresponding links are not included in Fig. 1.

Closed-Loop Control for HM-MDPs

In the process outlined in the previous section, the agent should iteratively process observations and take actions depending on the updated belief. By forecasting future model updating steps, depending on the assumed available information, and behaving consequently, the agent adopts a closed-loop policy. All information for making decision A_k at time t_k are summarized by the augmented state $(i, \mathbf{b})$, where ordinal $i = S_k$ indicates the observable physical state and vector $\mathbf{b} = \mathbf{b}_k$ indicates current model belief. Time-varying policy $\Psi = \{\psi_1, \ldots, \psi_T\}$ is defined on this augmented state $(i, \mathbf{b})$. Adapting the notation of Memarzadeh and Pozzi (2016b), the value $\mathbb{V}$, following policy Ψ, is

$$\mathbb{V}_k^{\Psi}(i,\mathbf{b}) = c_k[i,\mathbf{b},\psi_k(i,\mathbf{b})] + \gamma \sum_{h=1}^{|Y|} e_k(z,\mathbf{b}) \times \sum_{j=1}^{|S|} H_k[i,\psi_k(i,\mathbf{b}),j,\mathbf{b}]\mathbb{V}_{k+1}^{\Psi}[j,\mathbf{u}_{k+1}(h,\mathbf{b})] \tag{12}$$

where immediate cost c_k, observation operator e_k, and entry m in updated belief $\mathbf{u}_{k+1}$, and expected transition H_k are defined as

$$\begin{aligned}
c_k(i,\mathbf{b},a) &= \mathbb{E}[C_k|\mathbf{b}_k = \mathbf{b}, S_k = i, A_k = a] \\
&= \sum_{m=1}^{M} C_{k,m}(i,a) b(m) \\
e_k(h,\mathbf{b}) &= \mathbb{P}[Y_k = h|\mathbf{b}_k = \mathbf{b}] \\
&= \sum_{m=1}^{M} O_k(m,h) b(m) \\
u_{k+1,m}(h,\mathbf{b}) &= \mathbb{P}[\mathcal{M} = m|\mathbf{b}_k = \mathbf{b}, Y_{k+1} = h] \\
&= \frac{O_k(m,h) b(m)}{e_k(h,\mathbf{b})} \\
H_k(i,a,j,\mathbf{b}) &= \mathbb{P}[S_{k+1} = j|S_k = i, A_k = a, \mathbf{b}_k = \mathbf{b}] \\
&= \sum_{m=1}^{M} T_{k,m}(i,a,j) b(m)
\end{aligned} \tag{13}$$

In Eq. (13), the updating follows Bayes' formula of Eq. (12). Bellman's equation for optimal value is

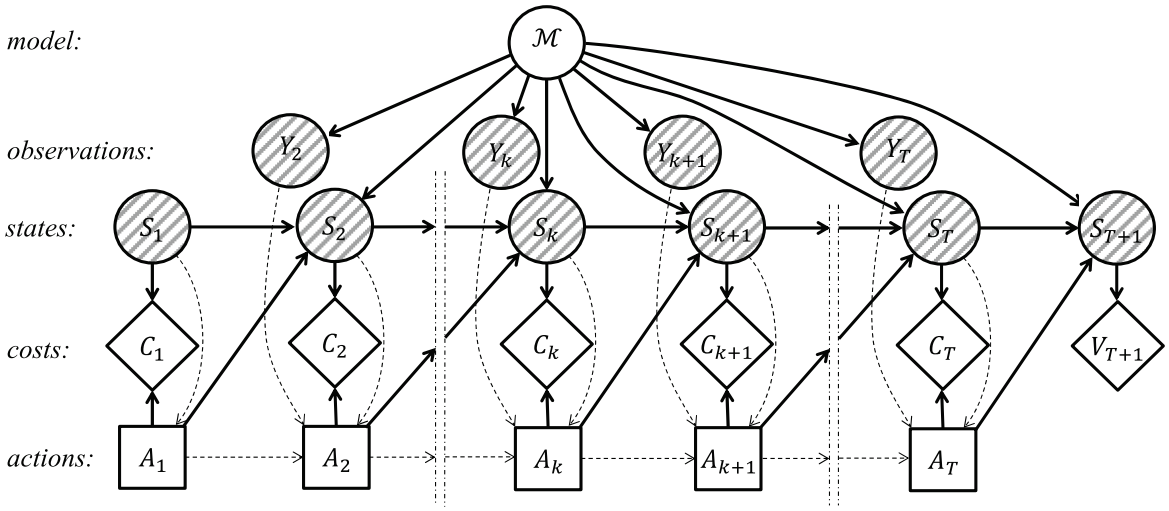

Fig. 1. Decision graph of the HM-MDP management process

$$\mathbb{V}_k^*(i,\mathbf{b}) = \min_a \Big\{ c_k(i,\mathbf{b},a) + \gamma \sum_{h=1}^{|Y|} e_k(h,\mathbf{b}) \sum_{j=1}^{|S|} H_k(i,a,j,\mathbf{b}) \mathbb{V}_{k+1}^*[j,\mathbf{u}_{k+1}(h,\mathbf{b})] \Big\} \tag{14}$$

The optimal closed-loop policy $\Psi^* = \{\psi_1^*, \ldots, \psi_T^*\}$ derives from the previous equation by using argmin instead of min. When belief is represented by standard basis vector $\mathbf{v}_m$, made of all zeros except for a 1 at position m, the agent is certain that model m is correct, and corresponding value $\mathbb{V}_k^*(i,\mathbf{v}_m)$ is equal to single-model value $V_{k,m}^*(i)$, because the agent cannot learn anything more and policy Π_m^* is indeed optimal.

In summary, via observation probability $O_k(m,h)$ the closed-loop formulation allows for representing assumptions about information availability that are more general than those for the d-step-ahead clairvoyance setting can handle. The values related to the open-loop and near-clairvoyance policies in those settings provide bounds for that of the closed-loop value

$$W_{k,k+1}^*(i,\mathbf{b}) \le \mathbb{V}_k^*(i,\mathbf{b}) \le W_{k,\infty}^*(i,\mathbf{b}) \tag{15}$$

Appendix 1 details the computational approach to solve Eq. (14) and identify the optimal closed-loop policy. Although any POMDP solver (Shani et al. 2013) can be applied to this task, for the HM-MDP setting, where the belief evolution does not depend on the adopted policy, simpler approaches, such as that illustrated in Appendix 1, are effective. The complexity of those approaches scales well with the number of models and actions available.

Simple Example of Impact of Assumed Available Information

To investigate the role played by assumed available information, consider an agent who must decide about the protection of an asset against extreme events over an infinite time horizon. Two stationary models are possible: Model 1 assumes extreme events occurring with probability P_1 per time step, whereas Model 2 assumes probability $P_2 > P_1$ per time step. Failure cost C_F is incurred when an extreme event occurs to an unprotected asset, independently of model or time. Investment cost C_I can protect the asset indefinitely, and the discount factor is γ. Belief $\mathbf{b}$ is of the form $[1-\theta \quad \theta]$, where $\theta = \mathbb{P}[m=2]$ defines the probability of the riskier model.

Clearly, optimal policies related to each single model in isolation are stationary. If doing nothing is optimal at time t_1 under model m, it will always be so; the corresponding annual risk is $R_m = P_m C_F$, and the cumulative risk doing nothing is $D_m = R_m/(1-\gamma)$. Now assume C_I between D_1 and D_2, so that a rational agent with perfect knowledge should take the risk and do nothing under Model 1, but invest to remove the risk under Model 2. Under model uncertainty, the open-loop approach prescribes doing nothing if belief parameter θ is less than the normalized investment cost $r = (C_I - D_1)/(D_2 - D_1)$, whereas the near-clairvoyance approach allows for doing nothing up to level θ_{nc}, defined as

$$\theta_{\text{nc}} = \frac{r}{1-\gamma(1-r)} \tag{16}$$

Fig. 2(a) llustrates how θ_{nc} varies depending on the normalized investment cost and discount factor. For every setting (summarized by r), the open-loop policy is more conservative than the near-clairvoyance policy—with the exception of the extreme values of $r=0$ or $r=1$ (for which the decision is trivial) or for $\gamma = 0$, the near-clairvoyance policy always tolerates higher values of θ before removing the risk by investing. This gap monotonically increases with the discount factor: for γ approaching 1, the near-clairvoyance policy prescribes waiting for the next time step, without investing, except when $\theta = 1$ (i.e., when Model 2 is certain). When perfect knowledge is assumed d time steps ahead, the corresponding threshold can be read in Eq. (16) using γ^d instead of γ; consequently, the d-step-ahead clairvoyance threshold becomes closer to the open-loop threshold if knowledge is postponed. Fig. 2(b) fixes r at 50% and γ at 95%, and plots with a dashed-dotted line the upper threshold (θ_{thr}) as a function of d. To investigate how the closed-loop optimal policy looks like when imperfect information on the model is available, assume that observations are binary variables, and that observation probability O is stationary, with Y_k supposedly indicating model variable m. *Inaccuracy* $\varepsilon = \mathbb{P}[Y_k \neq m]$ defines the probability of a wrong measure at each step; $\varepsilon = 0$ indicates a perfect observation, whereas $\varepsilon = 50\%$ indicates irrelevant measures. Fig. 2(b) shows the upper threshold of the belief as a function of ε. It starts at 50% (equal to r) and monotonically increases up to $\theta_{\text{nc}} = 95.2\%$ when $\varepsilon = 0$ and observation is

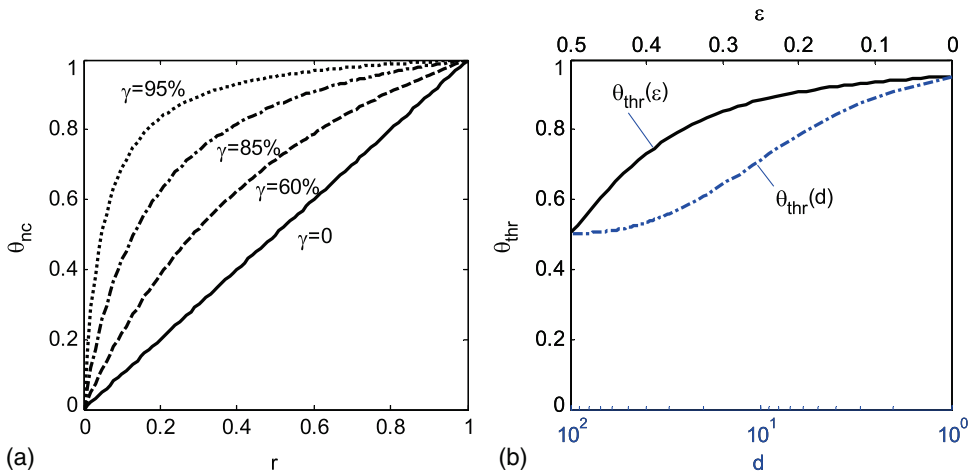

Fig. 2. (a) Optimal policy's threshold for the near-clairvoyance approach, as a function of normalized investment cost and discount factor; (b) threshold for the closed-loop policy, as a function of the delay for getting perfect information or of observation inaccuracy, for 50% normalized investment cost and 95% discount factor

perfect. This is consistent with the behavior of the threshold as a function of delay d.

Overall, this simple example shows how the optimal policy depends on the assumed learning rate and how a faster learning rate suggests a less conservative policy. Consider, for example, an agent assigning 85% probability to Model 2, while normalized cost is $r = 50\%$. Although it seems that in these circumstances one should immediately invest and remove the risk, no suggestion about the optimal decision can be made without an assumption about the future information. In fact, if sufficiently accurate observations are available (in detail, with $\varepsilon < 30\%$), it is optimal to wait, as can be seen in Fig. 2(b). Similarly, it is optimal to wait if perfect information is available within d equals three or fewer steps.

Application to Flood-Risk Reduction

Consider the protection of assets near the coast, in a location near Battery Park, New York, where current and future flood risks have been estimated (e.g., Lin et al. 2012). Although decision making can range from an individual scale to a national scale, this example specifically focuses on a single asset, such as a single-family house. The asset has slightly less than a 1% per year chance of flooding, and thus even if the owner has a mortgage, she or he is not currently forced to buy flood insurance through the National Flood Insurance Program (FEMA 480). The decision maker can select when to elevate the asset and by how much. As in the example of the previous section, on the one hand, under model uncertainty, it may be convenient to wait until further evidence about the model is available; on the other, to wait may be risky, because of a high chance of a flood under some models. The analysis investigates how the decision should depend on the learning assumptions.

As in Špačková and Straub (2017), the agent considers three climate models and defines the occurrence of extreme events by an extreme value distribution with time-varying parameters. First, recognizing that many people in the United States do not believe in climate change (Leiserowitz et al. 2012), Model 1 assumes that *no change* occurs. Next, Model 2 predicts *low climate change*, and Model 3 predicts a relatively *high climate change*. Model 3 derives from Lin et al. (2012) (Fig. 4), who used advanced hydrological estimates under an ensemble of climate models for a high-emissions scenario resulting in approximately 1 m of sea level rise by 2100. The second model is exactly between these two predictions. The analysis is extended to more models subsequently, but here it is restricted to these three models for the sake of illustration.

Table 1. Parameters of Weibull Distributions for Flood Annual Maxima

Parameters	$m = 1$	$m = 2$	$m = 3$
$\lambda_{m,k}$ (m)	43.45%	$44.55 + 0.55\%k$	$46.33 + 1.44\%k$
$\beta_{m,k}$	1	$100.1 + 0.06\%k$	$101.2 + 0.62\%k$

Time is discretized in years, and variable z indicates the annual maximum flood height. Annual maximums are assumed to be independent, and $p(z|m,k)$ refers to the probability density for year k, according to model m, which is a Weibull distribution with scale parameter $\lambda_{m,k}$ and shape parameter $\beta_{m,k}$. Parameters for each model are time-varying and are reported in Table 1.

Fig. 3 plots the probability density of z for different years and models. As mentioned, Model 1 is stationary, whereas the other two models assume a distribution at Year 1 close to that of Model 1 but an increment for the following years.

Under model m, the probability of a flood above level z_e occurring in year k is $X_{k,m} = 1 - F_{\text{Weibull}}(z_e; \lambda_{m,k}, \beta_{m,k})$, where $F_{\text{Weibull}}(z; \lambda, \beta)$ indicates the cumulative Weibull distribution with scale parameter λ and shape parameter β, computed at z. Fig. 4(a) shows this probability as a function of time t_k: for level $z_e = 2$ m, that probability is a constant 1% under Model 1, and over 100 years it increases to 12% under Model 2 and to 33% under Model 3. The initial probability is decreased by one and two orders of magnitude for level $z_e = 3$ and 4 m, respectively. After 100 years, the same levels of elevation reduce the probability to 12 and 3.5% under Model 2 and to 4 and 1.25% under Model 3.

It should be noted from Table 1 and Fig. 4(a) that the models do not exactly agree on the flood probability at Year 1. The models agreed some time ago (specifically, 3 years ago), but their current assessment is already slightly different, and the agent cannot identify which model is correct. Results are similar if the specific setting of the flood probability at Year 1 is the same for all models.

The initial level of the asset is $z_0 = 2$ m, the decision variable Δz indicates the elevation, so that $z_0 + \Delta z$ is the level after the decision has been made. Every year, the agent selects a value for Δz; however, assume here that it is inconvenient to elevate the asset more than once during the management process, so Δz can be greater than zero only once during the process. The setting would be computationally just slightly more complicated if one allowed for the possibility of re-elevating the asset.

The domain of Δz is discretized in 13 values from 0 to 3 m, equally spaced with an interval of 25 cm. The *physical state*

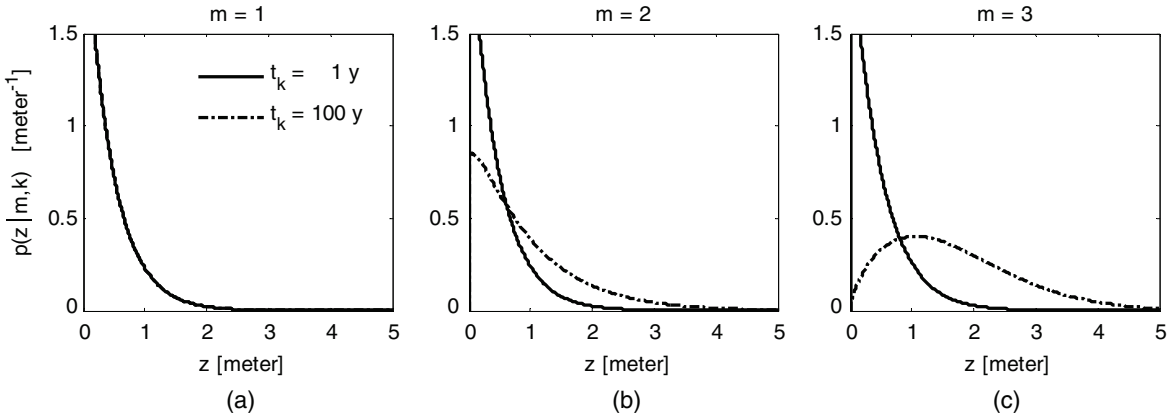

Fig. 3. Probability density for annual maximum at Year 1 and Year 100: (a) Model 1; (b) Model 2; (c) Model 3

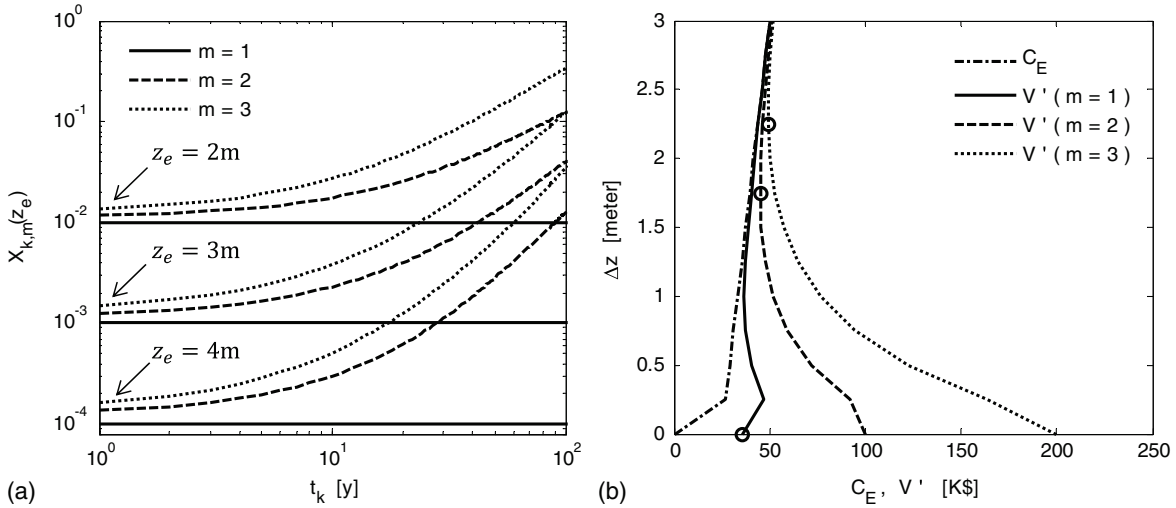

Fig. 4. (a) Probability of flood exceeding level z_e, depending on time and model; (b) discounted cost as a function of model and initial elevation

of the system is the asset level, so the problem is described by 13 possible states, and state i corresponds to level $z_i = z_0 + 25\ \text{cm}(i-1)$. The cost of elevating C_E includes a fixed cost for intervention and a term proportional to the elevation value; it is zero when Δz is zero and it is $C_E = \$25\text{K} + \$25\text{K} \cdot \Delta z/3\ \text{m}$ for positive Δz. The dashed-dotted line in Fig. 4(b) shows C_E as a function Δz.

In any year when the asset is flooded, the agent incurs cost C_F of \$180K that includes damages, downtime, and repairs. The annual discount factor is 95% (corresponding to a discount rate of 5.26%). The management process lasts 100 years; however, to avoid the need for identifying an appropriate residual value at the final time, the time-horizon is 200 years (i.e., $T = 200$), without residual value (i.e., $V_{T+1} = 0$). In other words, despite not trusting the climate modeling beyond the 100-year horizon, it is assumed that it correctly models the overall value for the first years.

The overall cost includes the flood-related risk and, possibly, if starting at State 1, the cost for elevation, although it is impossible to change the state when it is higher than one. Action a takes the state to value a, so the transition matrix is deterministic and independent of the model $T_{k,m}(i,a,j) = \delta_{a,j}$, where $\delta_{\cdot,\cdot}$ is the Kronecker delta. In this form, the asset level after action a is z_a, and the immediate cost matrix is defined as

$$C_{k,m}(i,a) = \begin{cases} C_F X_{k,m}(z_a) & a = i \\ C_F X_{k,m}(z_a) + C_E(a) & a > i = 1 \\ \infty & a \neq i > 1 \end{cases} \quad (17)$$

Single Model, Open-Loop, and Near-Clairvoyance Analyses

Start by assuming that the agent acts at time t_1, with no further possibility of revising her or his decision. Fig. 4(b) shows the expected discounted management cost V' as a function of initial elevation action Δz and model m. The solution derives from a simple application of Eq. (2). In the same figure, circles indicate optimal actions: the agents should not elevate the asset under *no change*, they should elevate the asset to 1.75 m under *low climate change* and they should elevate the asset to 2.25 m under *high climate change*.

Once the constraint of acting at time t_1 is removed, the optimal policy may differ from that shown in Fig. 4(b). Even under a single model, the agent may choose to wait and elevate at a future time, e.g., to discount the corresponding cost. However, in this specific case, the optimal elevation time is at the beginning of the management process, and that figure shows the optimal policy. Fig. 5(a) plots the optimal policy for a nonelevated asset (i.e., for $S_k = 1$) depending on time. As expected, the agent never elevates under stationary Model 1. Under Models 2 or 3 if, for some reason, the agent did not elevate the asset at the beginning, the optimal elevation value Δz^* increases during the process, up to the maximum allowable value of 3 m. Fig. 5(b) shows the open-policy for two examples of belief. The near-clairvoyance policy always prescribes not elevating, counting on perfect information at next step.

The open-loop, near-clairvoyance, and d-step-ahead clairvoyance policies derive from Eqs. (5) and (8). Fig. 6 reports some of those outcomes for $S_1 = 1$, time t_1, and for each possible belief

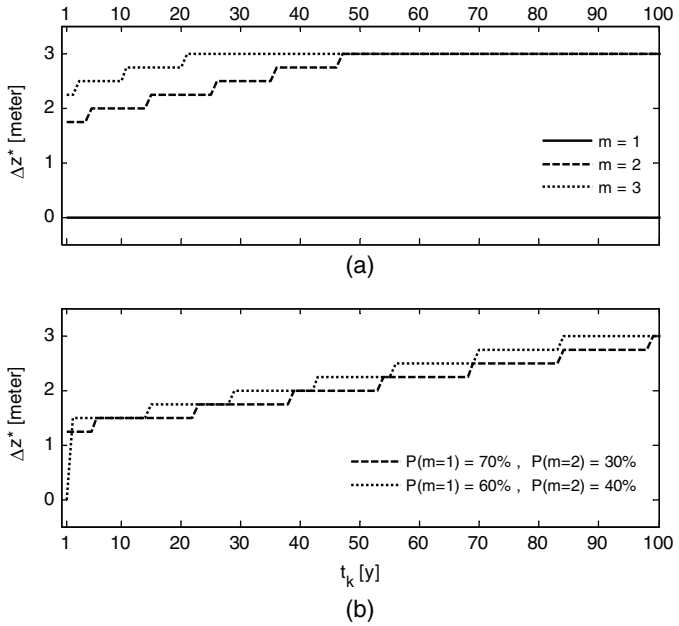

Fig. 5. (a) Optimal action $\pi^*_{k,m}(1)$ for a nonelevated asset under perfect model information, depending on time and model; (b) corresponding action in the open-loop policy $\phi^*_{k,\infty}(1,\mathbf{b})$, for two specific belief values

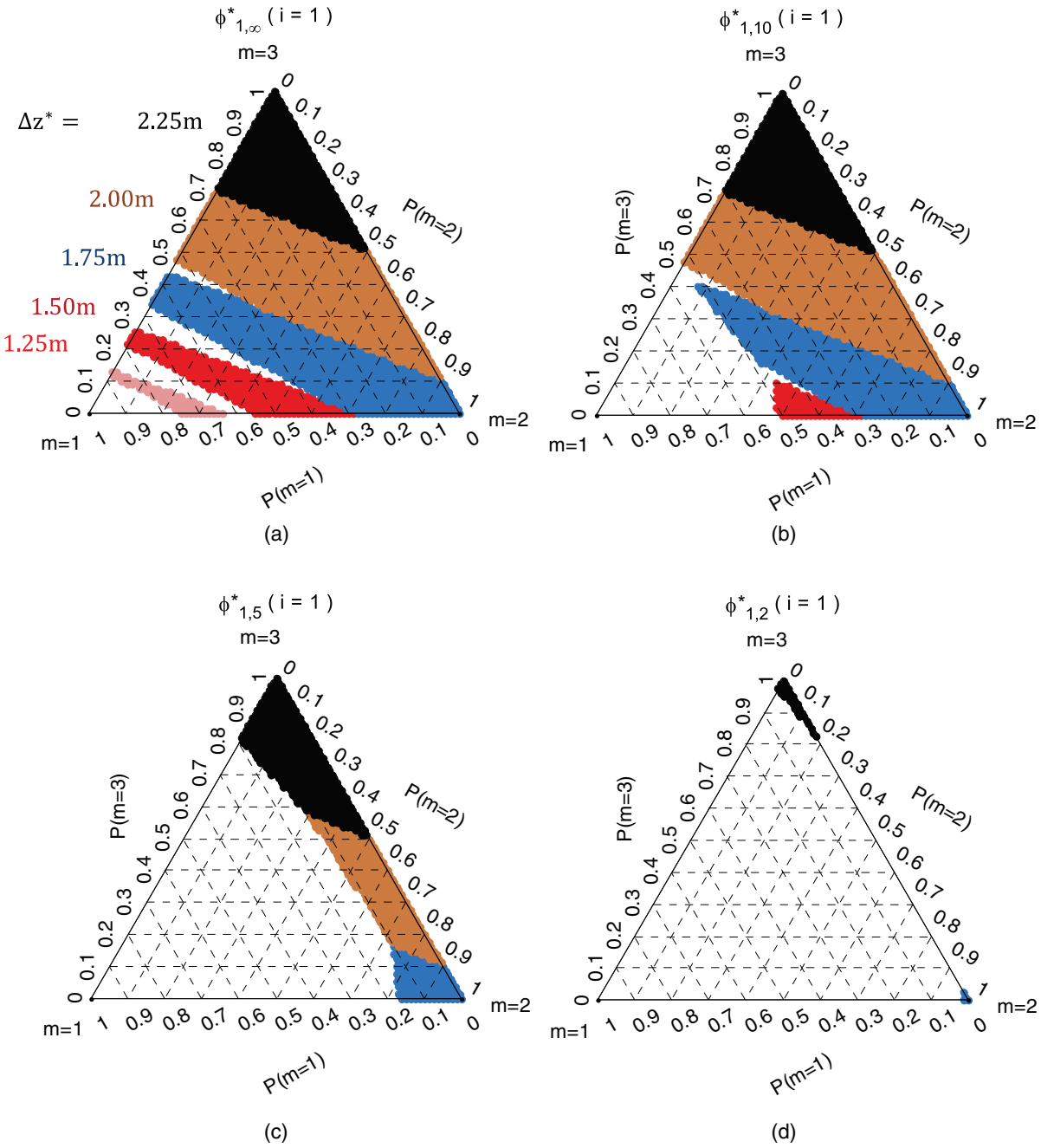

Fig. 6. Policies as a function of the belief: (a) open-loop initial policy; (b) policy if perfect information will be available at Year 10; (c) policy if perfect information will be available at Year 5; (d) near-clairvoyance policy

b. The belief is a three-component vector normalized to 1, so it can be represented in a two-dimensional (2D) region, in which each point in the triangle represents a possible belief and the vertexes refer to perfect model knowledge. Fig. 6(a) represents the initial open-loop policy $\phi^*_{1,\infty}$ via colors: the nonshaded area refers to *do nothing* (i.e., $\Delta z^* = 0$), and each color is related to one specific elevation value. Generally, the higher the belief in climate change, the higher the elevation. However, the specific policy is quite complicated, and should be compared with the d-step-ahead clairvoyance policy when the model is revealed at Year 10 and at Year 5 and with the near-clairvoyance policy (when it is revealed next year). The smaller the value of d, the more optimistic is the learning scenario. For small values of d the agent will prefer to wait, postponing the decision about elevation unless she or he is almost certain about the model. In this latter case (i.e., in the vertexes of the belief domain), all policies are consistent among themselves and with the initial optimal actions under single models, as shown in Fig. 5.

The values at initial time corresponding to open-loop ($W^*_{1,\infty}$) and near-clairvoyance ($W^*_{1,2}$) policies are reported in Figs. 7(a and b), respectively. Although, again, the values in the vertexes are the same, $W^*_{1,\infty}$ is always higher (strictly speaking, it is non-less) than $W^*_{1,2}$, as predicted by Eq. (9).

Fig. 8(a) reports the incremental cost ΔV^I_1, as defined in Eq. (10). As expected, the incremental cost is always positive (strictly speaking, it is non-negative) and it is zero at the vertexes,

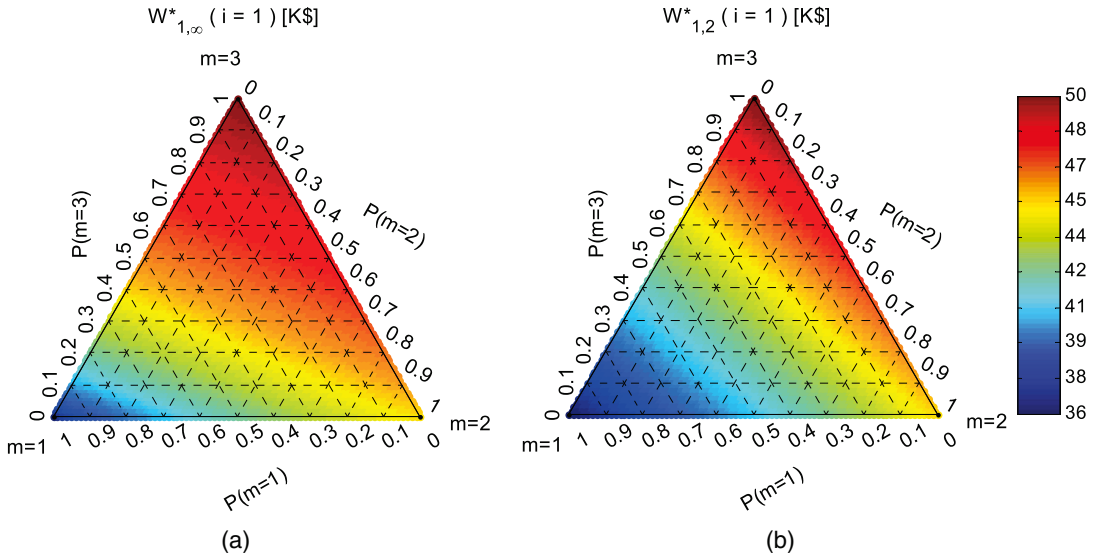

Fig. 7. Optimal discounted expected cost: (a) with no information at next step; (b) with perfect information at next step

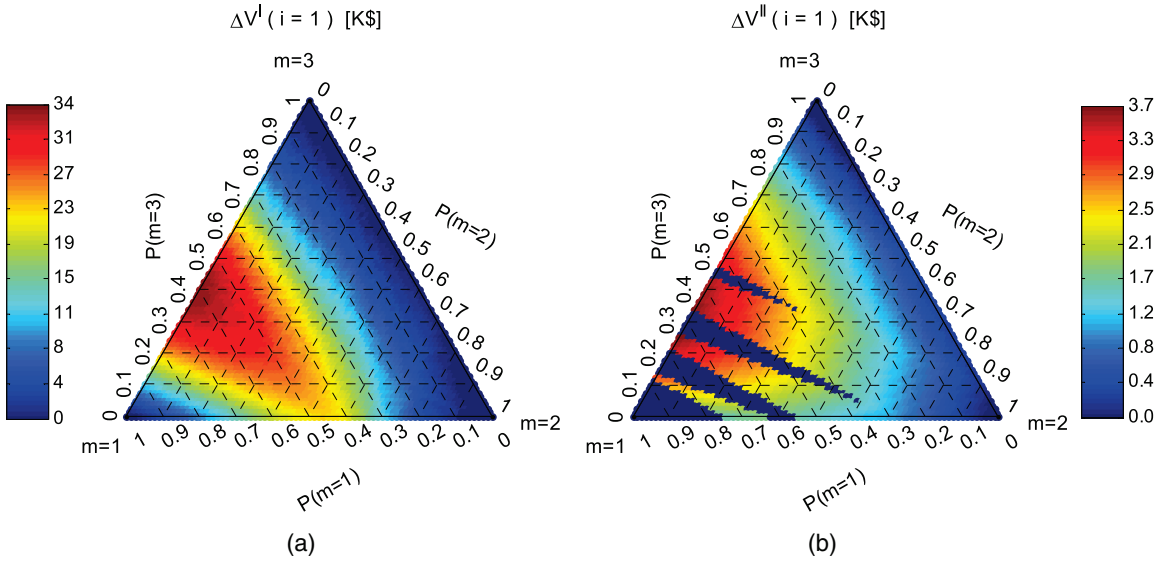

Fig. 8. (a) Incremental cost of adopting the near-clairvoyance instead of open-loop policy in the no-learning scenario; (b) incremental cost of using open-loop instead of near-clairvoyance policy when perfect information is available at next step

where policies are consistent. For example, for an initial belief assuming $m = 1$ or 2 with 50%/50% probability, the incremental cost is approximately \$20K. Fig. 8(b) reports incremental cost ΔV_1^{II}, that is approximately \$1.25K for that belief. As noted previously, ΔV_1^{I} tends to be higher than ΔV_1^{II} because the penalty for not having anticipated information is more severe than the benefit of having additional information beyond that anticipated.

In Fig. 6(a), the strips of nonshaded area may look counterintuitive. To better understand that policy, examine Fig. 5(a): when the probabilities of $m = 1$ or 2 are 70 and 30%, respectively, the open-loop policy prescribes elevating the asset to 1.25 m, whereas it prescribes doing nothing when those probabilities are 60 and 40% respectively, even if the latter belief assigns a higher probability of climate change than does the former. Actually, in this latter belief the policy postpones the elevation by one year, but it adopts a higher elevation value ($\Delta z^* = 1.5$ m), and it is never below the policy of the former belief after the first year.

Closed-Loop Analysis

Now consider scenarios of possible evolutions of the belief about climate models by introducing an appropriate observation probability. Instead of directly defining possible advancement in climate analysis, this model uses indirect measures of the current floods in regions close to the asset. A noise-level parameter ζ_Y defines the uncertainty in those annual observations according to the formula reported in Appendix 1.

Fig. 9 plots $O_k(m, h)$ for all values of h from 1 to $|Y|$, depending on the model, for time $t_k = 5$, 10, and 20 years, and $\zeta_Y = 50\%$, 1, and 2. Observation probabilities for different models are increasingly separated as time passes, consistent with Fig. 3. This is because the difference among predictions of models grows with time. Furthermore, the inverse of noise-level parameter ζ_Y can be related to the learning rate. When ζ_Y is close to zero, perfect information is available at next step; in contrast, for large ζ_Y the observation

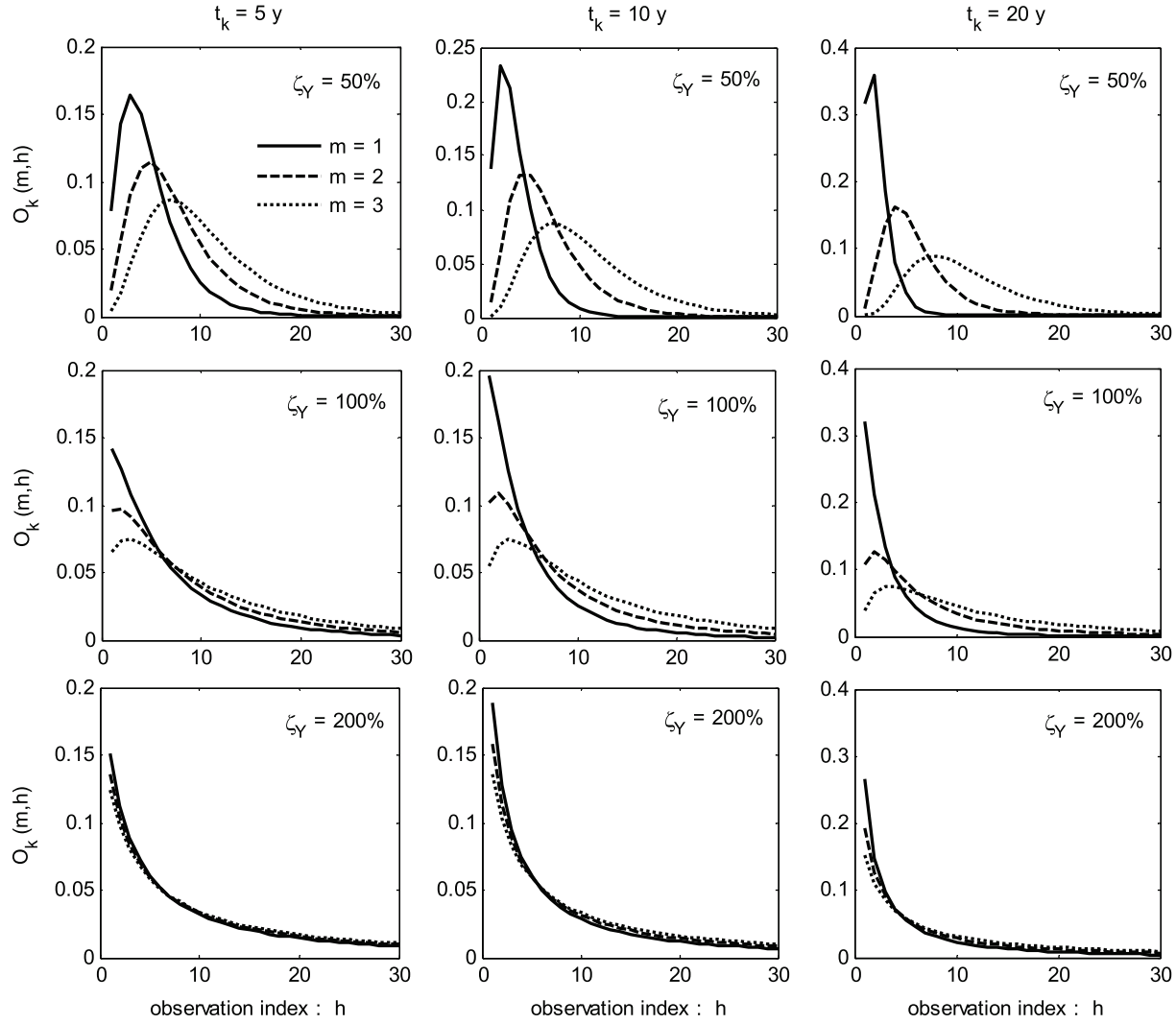

Fig. 9. Observation probability matrix $O_k(m,h)$, at different times and for different noise levels

probability is flat across models, and observations contain no information about the correct model.

Fig. 10(a) reports 1,000 forward simulations of the belief evolution, starting from initial belief $\mathbf{b}_1 = [34 \quad 32 \quad 34]\%$, which is represented by a point close to the middle of the triangle shown in Fig. 8. Beliefs are simulated by sampling observations according to their probability and processing them. Details on the forward simulations are reported in Appendix 1. The three columns refer to different leaning rates for $\zeta_Y = 2$, 1, and 50%, respectively. In the first case, learning is slow: after 5 years the belief tends to be close the initial one, and after 50 years the likelihood of being close to that point is still high. At the second and third rates the learning process is faster; for example, when ζ_Y is 50% it is highly improbable that the belief is still similar to the initial belief after 50 years. The reason is that, for that noise level, observations collected during 50 years are almost sufficient to reveal exactly the model for all cases. In the long term the belief tends to converge to the correct model, and therefore the simulations migrate to the domain's vertexes. It is interesting to note, however, how they reach the vertexes. First, the simulations tend to move away from the midpoint of the left side of the triangle. That point represents belief $\mathbf{b} = [50 \quad 0 \quad 50]\%$, i.e., it is uncertain between high climate change and no change, but it excludes the possibility of low climate change: practically no sequence of observations leads to that outcome. Moreover, simulations tend to approach the lower and right sides, which represent uncertainty between two out the three models—between high climate change and low climate change, or between low climate change and no change. Lastly, note that convergence to $m = 1$ or 3 is faster than convergence to $m = 2$ (i.e., the low climate change model). This happens because Model 2 is intermediate between other two, so both low and high observations from Model 2 may be mistaken as coming from other models. In contrast, Models 1 and 3 are *free from one side*; for example, the no change model can be quickly identified if low-value observations are systematically collected. Fig. 10(b) reports the corresponding closed-loop policy for the initial year. In the slow-learning scenario (i.e., for $\zeta_Y = 2$), that policy is similar to the open-loop policy reported in Fig. 6(a), and it would become identical for larger ζ_Y. However, for faster learning rates (i.e., for $\zeta_Y = 1$ or 50%), the agent prefers to do nothing and wait, unless the probability of no climate change (i.e., of Model 1) is less than approximately 30%. Again, when noise level ζ_Y goes to zero, the optimal policy converges to the near-clairvoyance policy reported in Fig. 6(d).

Figs. 11(a and c) compare values and policies for the d-step-ahead clairvoyance approach and the closed-loop approach, for four initial beliefs. Value $W^*_{1,d}(1,\mathbf{b})$ grows monotonically (strictly speaking, it grows or stays constant) with time t_d, and value

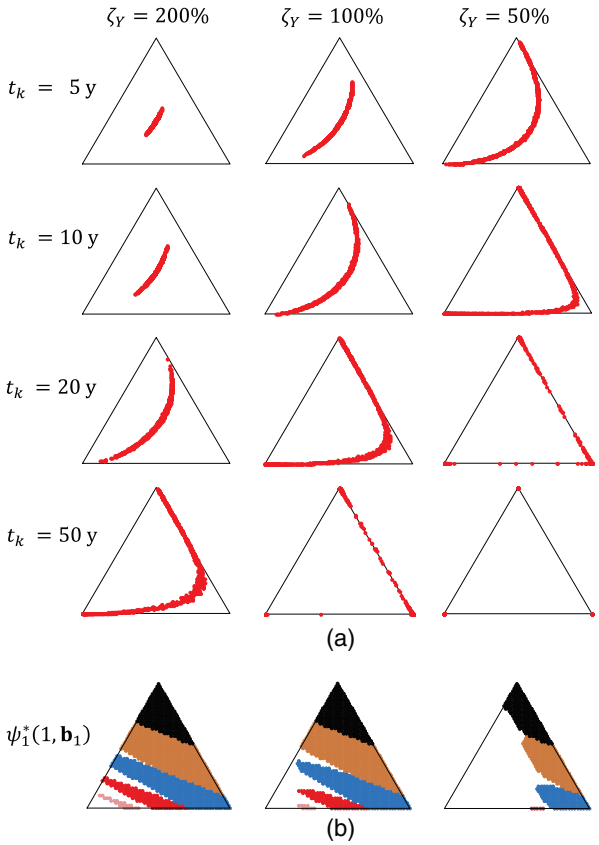

Fig. 10. (a) Outcomes of 1,000 forward simulations, depending on time and noise level of observations: (b) closed-loop policy depending on the noise level; each triangle represents the belief domain, with the same scale as those in Figs. 6–8

$\mathbb{V}_1^*(1, \mathbf{b})$ grows with ζ_Y. The blue crosses reported at $\zeta_Y = 1\%$ and $\zeta_Y = 10$ represent the values for the near-clairvoyance and open-loop policies, respectively, that can be read in Fig. 7. The gap between the values for two different coordinates in the horizontal axis quantifies the value of information, that is, the benefit of identifying the model earlier or of getting better observations. Clearly, this depends on the belief (e.g., it would be zero if the model was already known). Figs. 11(b and d) show the corresponding initial optimal action, represented in terms of elevation value Δz^*; it is zero under a threshold that depends on the belief, and consistent with the open-loop policy above that threshold. Consequently, the value in Fig. 11(a) is flat above the threshold, because the policy is invariant with respect to higher noise level.

Variations of Original Setting: Alternative Action Set

To illustrate how the solution depends on the available actions, suppose that the only alternative with respect to not elevating the asset is to elevate it to $\Delta z = 1$ m. Fig. 12 plots the corresponding open-loop policy [Fig. 12(a)], closed-loop policy with ζ_Y equal to 50% [Fig. 12(b)], and near-clairvoyance policy [Fig. 12(c)], where shaded area refers to elevating the asset. By comparison with Figs. 6 and 10, it is clear that available actions strongly affect these policies. Specifically, the near-clairvoyance policy is now less prone to procrastination than that in Fig. 6(d) because the agent no longer needs to select an appropriate (and potentially expensive) intervention, and not much information is needed.

Variations of Original Setting: Many Climate Models

The example described previously included only three models for the sake of illustration, because no belief's domain can be shown when M is greater than 3. Fig. 13(a) shows distribution of the

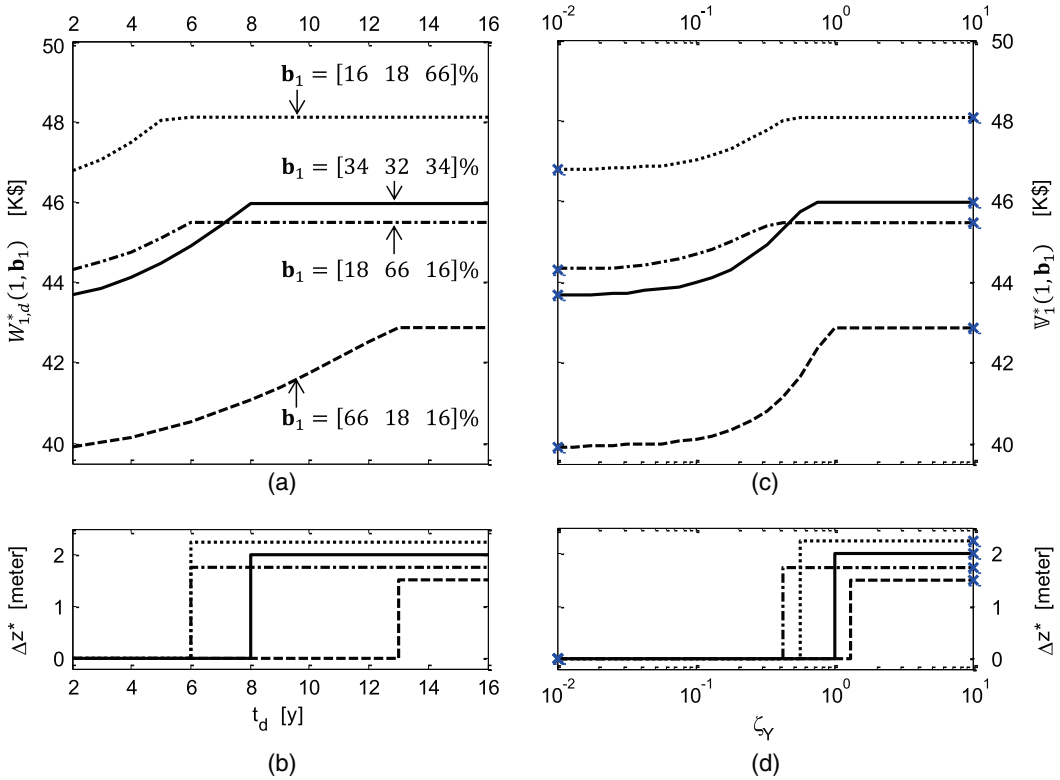

Fig. 11. (a) Optimal discounted expected cost depending on the time of perfect information, for four selected initial beliefs $\mathbf{b}_1$; (b) corresponding policy; (c) cost for the closed-loop approach, depending on noise level; (d) corresponding policy

J. Infrastruct. Syst., 2017, 23(4): 04017022

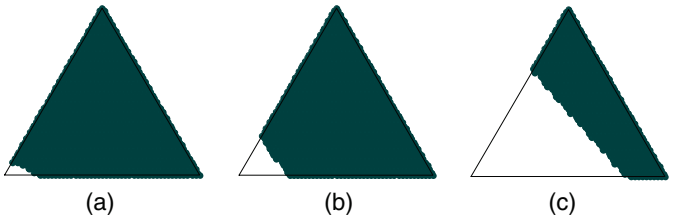

Fig. 12. Initial action of (a) the open-loop policy; (b) the closed-loop policy with noise level equal to 50%; and (c) the near-clairvoyance policy, when alternatives are elevating by 1 m (shaded area) or not elevating; each triangle represents the belief domain, with the same scale as those in Figs. 6–8

annual maximum (as in Fig. 3) and Fig. 13(b) shows the probability of floods exceeding z_e [as in Fig. 4(a)] for a set of 10 models by adding 7 models to the 3 used above, all intermediate between the no change model and the high climate change model.

The analysis can be performed in this high-dimension domain, and Fig. 14 reports the outcomes of a parametric analysis (similarly to Fig. 11) from an initial belief assigning 19% probability to Model 1 with no change, and 9% to each of the other models.

Conclusions

The framework and the examples in this paper illustrate how sequential decision making under climate change, with known or unknown dynamic models, can be optimized. For small dimensions (i.e., for small $|S|$ and $|A|$) and under a known model, solving MDPs is computationally simple. When, as in the HM-MDP framework, the model is within a set of M possible candidates, exact solution is still computationally efficient (with complexity growing linearly with M) in the special case of having perfect information available at some step. For the intermediate case of imperfect observations, numerical schemes for identifying the optimal closed-loop policy, adapted from those for solving

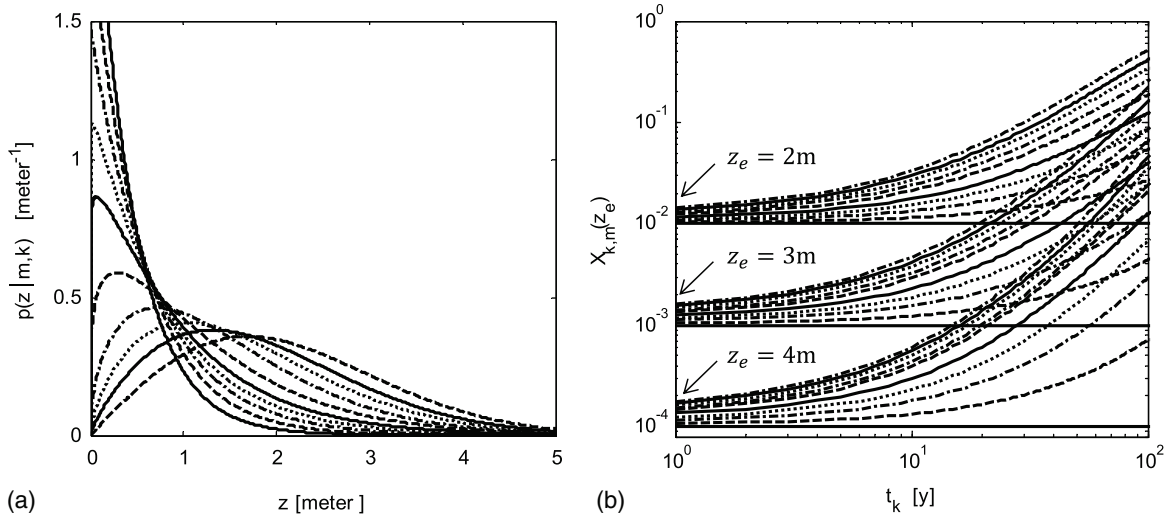

Fig. 13. For a set of 10 models: (a) probability density for annual maximum at Year 100; (b) probability of flood exceeding level z_e, depending on time and model

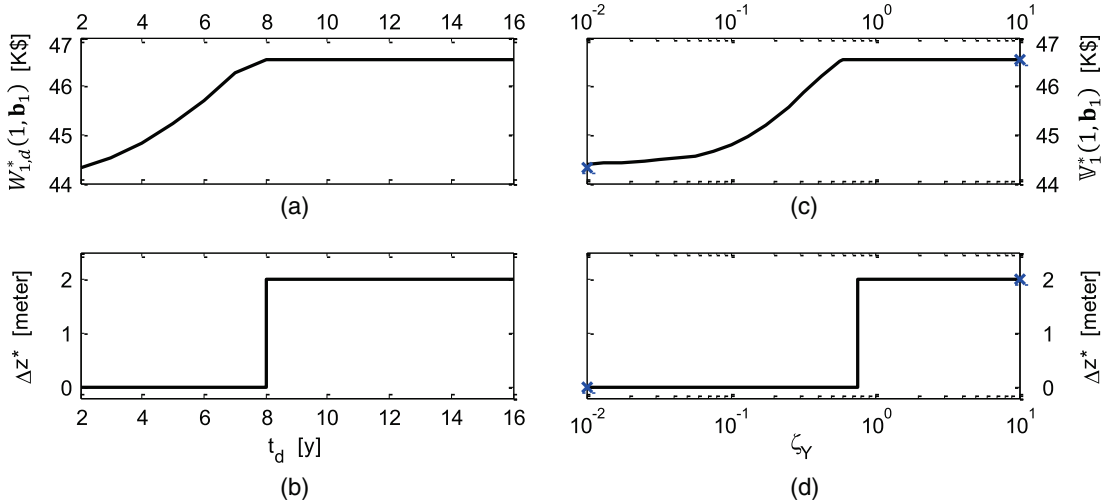

Fig. 14. (a) Optimal discounted expected cost depending on the time of perfect information; (b) corresponding policy; (c) cost for the closed-loop approach, depending on the noise level; (d) corresponding policy

POMDPs, are also available, as illustrated in Appendix 1. Those are approximate methods, but despite their complexity being much higher than that of the limit-case settings cited previously, they can be implemented effectively even for a large set of candidate models.

Overall, the analysis showed that the assumption regarding the availability of future information about the climate model can play a key role in current decision making under uncertainty.

Three final issues are worth mentioning. First, when dealing with long-term management processes, it may be hard to provide a complete list of possible evolutions, because additional models will be developed by climate scientists in the future, and one may argue that those yet-to-be developed models must be included in the current set. To address this issue, the authors can only recommend including a sufficiently rich set of models able to approximately cover the entire current epistemic uncertainty.

Second, although this paper outlined methods that assume a specific future learning rate, open questions remain about making decisions under an uncertain rate. Previous discussion may suggest that a pessimistic assumption on the rate has better guarantees, because the penalty for overoptimism tends to surpass that for overpessimism. However, the authors posit that in order to also avoid wasting resources due to overpessimism, one should identify and adopt an approximate learning scenario. In the Bayesian framework, the rational approach for planning under an uncertain rate is hierarchical modeling (where the rate is a random variable to be learnt during the process); that approach poses specific computational challenges (Memarzadeh et al. 2016).

Finally, climate change models are not *naturally* given, but they are at least partially the result of a decision-making process involving energy use and emissions policy. Similarly, the learning rate is affected by decisions related to investment in climate studies and, locally, in risk analyses. Although this paper treats those features as given, the analysis can be taken as a component in a more general investigation about optimizing those decisions.

Appendix 1. Numerical Methods for HM-MDPs

Point-Based Value Iteration

This appendix illustrates how to solve numerically HM-MDPs. Given immediate cost function, discount factor, transition function, observation probability, and initial belief, Eqs. (13) and (14) allow for identifying optimal policy and corresponding value. In doing so, the hardest part is to model appropriately the value at next step, $\mathbb{V}^*_{k+1}$, in Eq. (14). However, it is well known that (for any POMDP) the convex value function can be approximated from above, at any time, by the envelope of a set of affine functions, on the belief's domain

$$\mathbb{V}^*_k(i,\mathbf{b}) \le \min_{\boldsymbol{\alpha}\in\boldsymbol{\Gamma}_{k,i}}[\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{b}] \tag{18}$$

where $\boldsymbol{\Gamma}_{k,i}$ = set of so-called *alpha vectors* for state i and time t_k. Each alpha vector is of size $[M \times 1]$ and refers to a specific conditional plan (Russell and Norvig 1995). At time t_k and state $S_k = i$, conditional plan $\wp_k(i)$ assigns current action $A_k = a[\wp_k(i)]$ and an action at each future time depending on the sequence of collected observations. Depending on observation $Y_{k+1} = h$ and next state $S_{k+1} = j$, the plan continues into a new conditional plan $\wp_{k+1}(h,j)$. Hence conditional plan $\wp_k(i)$ can be described by the initial action $a[\wp_k(i)]$ and the set of conditional plans $\wp_{k+1}(h,j)$, for each possible state j and observation h. Consequently, from n_{k+1} conditional plans at t_{k+1}, $n_k = n_{k+1}^{|Y||S|}|A|$ distinct possible plans can be defined at t_k. However, most of them can be neglected as dominated by other plans, at least in the set of beliefs that are reachable from the initial plan. If $\boldsymbol{\Gamma}_{k,i}$ contained all possible alpha vectors, referring to all possible plans, Eq. (18) could have been written with an equal sign.

To build an alpha vector from the corresponding conditional plan, it is sufficient to assess the value under each single model. Following Eq. (12) for $\mathbf{b} = \mathbf{v}_m$, component m of the alpha vector for conditional plan $\wp_k(i)$ can be expressed as

$$\alpha_{m,\wp_k(i)} = C_{k,m}\{i, a[\wp_k(i)]\} + \gamma \sum_{h=1}^{|Y|} O_k(m,h) \sum_{j=1}^{|S|} T_{k,m}\{i, \pi^*_{k,m}(i), j\}\alpha_{m,\wp_{k+1}(h,j)} \tag{19}$$

where $\alpha_{m,\wp_{k+1}(h,j)}$= component m of the alpha vector related to conditional plan $\wp_{k+1}(h,j)$.

Alpha vectors for each time can be built by initializing a set of them at the end of the time horizon, and using Eq. (19) as a Bellman operator. However, as noted previously, the number of vectors grows exponentially and, after few steps backward, the complete set becomes intractable. A point-based value iteration method, such as those for solving POMDPs, can approximate the value function, in the set of beliefs reachable from the initial belief, with a limited number of relevant alpha vectors. Suppose $\boldsymbol{\Gamma}_{k+1,j}$ contains a set of alpha vectors able to appropriately represent $\mathbb{V}^*_{k+1}(j,\mathbf{b})$, as in Eq. (18). Eq. (14) can be approximated as follows:

$$\mathbb{V}^*_k(i,\mathbf{b}) \le \min_a \Big\{ c_k(i,\mathbf{b},a) + \gamma \sum_{h=1}^{|Y|} e_k(h,\mathbf{b}) \sum_{j=1}^{|S|} H_k(i,a,j,\mathbf{b})\Upsilon[\mathbf{u}_{k+1}(h,\mathbf{b}), \boldsymbol{\Gamma}_{k+1,j}]\Big\} \tag{20}$$

where $\Upsilon(\mathbf{b},\boldsymbol{\Gamma}) = \min_{\boldsymbol{\alpha}\in\boldsymbol{\Gamma}}[\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{b}]$. By solving Eq. (20) for a specific pair $(i,\mathbf{b})$, one obtains not only the corresponding optimal value but also the corresponding action a and a dominating alpha vector for each pair (h,j), corresponding to conditional plan $\wp_{k+1}(h,j)$

$$\mathbf{b} \to \forall(h,j): \boldsymbol{\alpha}_{\wp_{k+1}(h,j)} = \operatorname{argmin}_{\boldsymbol{\alpha}\in\boldsymbol{\Gamma}_{k+1,j}}[\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}_{k+1}(h,\mathbf{b})] \tag{21}$$

The specific conditional plan at time t_k is composed of action a followed by set of alpha vectors $\{\boldsymbol{\alpha}_{\wp_{k+1}(h,j)}\}$ with $(1 \le h \le |Y|; 1 \le j \le |S|)$ so that this conditional plan is optimal from belief $\mathbf{b}_k$ and state S_k defined by $\mathbf{b}$ and i, respectively. Following this observation, one can select as relevant alpha vectors those corresponding to conditional plans that are optimal for a set of relevant points, reachable from the initial belief. Although the number of reachable points may grow exponentially over time, the approximation can rely on a limited number (N) of independent forward Monte Carlo simulations, from initial belief $\mathbf{b}_1$, as described subsequently, to get N sampled beliefs at each time during the decision process.

At time t_{T+1}, the process is over, and the only alpha vector is defined as $\alpha_{m,i} = V_{T+1}(i,m)$, assuming that residual value V_{T+1} may depend on model and state. Relying on this initialization and on the set of samples, the complete procedure for identifying the optimal initial action is as follows:

1. From $k = T$ down to 2 and for each sampled belief, identify the optimal conditional plan using Eqs. (20) and (21) and the corresponding alpha vector using Eq. (19), populating set $\boldsymbol{\Gamma}_{k,j}$, for each state j. Each alpha vector is also associated with a specific action.

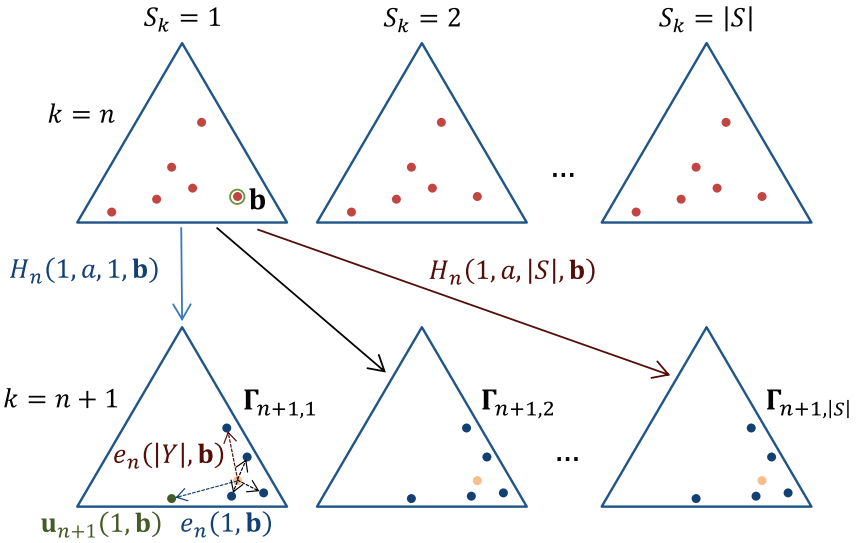

Fig. 15. Illustrative representation of the procedure for identifying optimal conditional plans and value

2. By solving Eq. (20) for $k = 1$, identify optimal initial action a, and the corresponding value.

Fig. 15 provides an illustrative scheme of the method. The upper row of triangles represents belief domains at time t_n for different states. The sampled beliefs are represented by red points in all domains, and they are the same for all states. For illustration, the figure focuses on one specific belief, $\mathbf{b}$, and state = 1. This belief is copied in orange, for the sake of illustration, in the domains of the second row, referring to time t_{n+1}. Blue points represent reachable beliefs, one for each possible value of observation variable h, so that $\mathbf{u}_{n+1}$ defines the coordinate of the updated belief, and e_n is the probability of the belief transition. Each state at time t_{n+1} is related to a set of alpha vectors, so that value at each blue point can be approximately computed by Eq. (18). Expected transition $H_n(1, a, S_{n+1}, \mathbf{b})$ also can be computed for each action a and next state value S_{n+1}. By combining the transitions in state and belief, and integrating immediate cost, one can identify the optimal action using Eq. (20), which acts as a Bellman operator. The procedure is repeated for each state and sampled belief, but it should be noted that many operations, e.g., the computation of the updated beliefs and future value Y, are invariant respect to the state value S_n.

Fig. 10 illustrates the importance of identifying the set of reachable beliefs. Although a large number of alpha-vectors may be needed to approximate the value function with high precision in the entire domain, according to Eq. (18), a limited number may be sufficient for approximating it in the region that can be reached. For example, for $\zeta_Y = 50\%$ and $t_k = 20$ years, belief is concentrated on two sides of the triangle. Therefore the value function should be well represented in that region, whereas the quality of the approximation outside that region may be irrelevant. The method outlined here searches for vectors that are relevant, because they dominate all other vectors, for points in that reachable region.

The number of alpha vectors to be included in set $\mathbf{\Gamma}$ depends on the target quality of the approximation. The quality grows with the number of vectors; however, vectors that are always dominated in the set of reachable beliefs can be pruned. More specifically, by neglecting vectors that provide small improvements to the value approximation, the number of vectors can be kept low. For example, when the reachable domain shrinks to a small set, the number of relevant alpha vectors drops. Likewise, at initial time t_1 there is just one relevant alpha vector, α_1, related to the identified conditional plan. Because of Eq. (18), the policy identified by the method is guaranteed to give a value bounded from above by the approximate value $\alpha_1^{\mathrm{T}}\mathbf{b}_1$.

Point-based value iteration methods, such as that illustrated here, specifically investigate reachable beliefs from a single initial belief. For the sake of illustration, Fig. 10(b) makes use of a set of possible initial conditions, in order to represent the policy in the entire belief domain.

Forward Monte Carlo Sampling

Forward Monte Carlo simulations allow for sampling reachable beliefs, as is needed for the procedure in previous section. To simulate beliefs following an initial belief, one can sample one model, simulate a sequence of observations from that model, and process this sequence. However, beliefs can also be propagated independently. To do so, consider belief $\mathbf{b}$ at time t_k. One can sample observation $Y_{k+1} = h'$ from distribution vector $\mathbf{e}_k$, whose component h is $e_k(h, \mathbf{b})$, as defined in Eq. (13). Updated belief derives, again, from Eq. (13)

$$h' \sim \mathbf{e}_k(\mathbf{b}) \rightarrow \mathbf{u}_{k+1}[h', \mathbf{b}] \tag{22}$$

Policy Evaluation under Limit Learning Assumptions

This section provides details on how to evaluate suboptimal policies, for computing Eq. (10). Although this may be challenging in specific settings, it turns out to be simple under the extreme conditions about available information. Start by considering that perfect information is available at the next step. An agent adopting the open-loop policy will update her or his belief, obtain perfect information at the next step, and follow the optimal single-model policy after that (because it coincides with the open-loop policy, without model uncertainty). Therefore one concludes that

$$W^{\Phi^*_{k,\infty}}_{k,k+1}(i, \mathbf{b}) = \mathbb{E}_m\{Q_{k,m}[i, \phi^*_{k,\infty}(i, \mathbf{b})]\} \tag{23}$$

For the opposite case, consider the scenario without information. In that case, the belief is time-invariant, and one can evaluate the near-clairvoyance policy in a similar way as in Eq. (5)

$$W^{\Phi'_{k,+1}}_{k,\infty}(i,\mathbf{b}) = \mathbb{E}_m C_{k,m}[i,\phi^*_{k,k+1}(i,\mathbf{b})] + \gamma \sum_j^{|S|} \mathbb{E}_m\{T_{k,m}[i,\phi^*_{k,k+1}(i,\mathbf{b}),j] V^{\Phi'_{k+1,m}}_{k+1,m}(j)\} \quad (24)$$

Appendix 2. Details of the Application to Flood Protection

Observation Modeling

The example of flood protection assumes that available measures are probabilistically related to the model as follows. Measure y_k, at year k, is lognormally distributed on a continuous domain, as

$$y_k|m \sim \ln\mathcal{N}[\ln X_{k,m}(z_0), \zeta_Y^2] \quad (25)$$

where $\ln\mathcal{N}(\lambda,\zeta^2)$= lognormal density with position parameter λ and scale parameter ζ^2. Therefore the median observation is the flood probability for a nonelevated asset, $X_{k,m}(z_0)$, and ζ_Y approximates, at least when it is small, the coefficient of variation. Conditional to the model, observations are independent. The observation domain is discretized in $|Y| = 30$ possible values by integrating the density in 30 contiguous intervals of equal length in range $[X_{k,1}(z_0)/3; X_{k,3}(z_0)\cdot 3]$. Matrix $O_k(m,h)$ derives from the normalizing of these probabilities, so that each model is related to an observation probability vector of unit sum across all possible observations.

Alpha Vectors for the Specific Application

In the application to flood protection, because of the assumption of deterministic transition depending on the action, Eq. (20) is simplified to

$$\mathbb{V}^*_k(i,\mathbf{b}) \le \min_a \left\{ EC_k(i,\mathbf{b},a) + \gamma \sum_{h=1}^{|Y|} e_k(h,\mathbf{b}) \Upsilon[\mathbf{u}_{k+1}(h,\mathbf{b}), \mathbf{\Gamma}_{k+1,a}] \right\} \quad (26)$$

and Eq. (19) is simplified to

$$\alpha_{m,\not{p}_k(i)} = C_{k,m}\{i, a[\not{p}_k(i)]\} + \gamma \sum_{h=1}^{|Y|} O_k(m,h) \alpha_{m,\not{p}_{k+1}\{h,a[\not{p}_k(i)]\}} \quad (27)$$

Moreover, for a state i higher than 1 (i.e., when the asset has already been elevated), there is just one possible conditional plan (because the agent cannot make other decisions), which is defined by a single alpha vector that can be called $\alpha_{k,i,i}$ without ambiguity. For State 1 and action a higher than 1, again there is a unique conditional plan and vector, called $\alpha_{k,i,a}$. Component m for these two vectors is

$$\begin{aligned} i > 1{:}\alpha_{m,k,i,i} &= C_{k,m}(i,i) + \gamma\alpha_{m,k+1,i,i} \\ a > 1{:}\alpha_{m,k,1,a} &= C_{k,m}(1,a) + \gamma\alpha_{m,k+1,a,a} \end{aligned} \quad (28)$$

Finally, from State 1 and Action 1, Eq. (19) reads

$$\alpha_{m,\not{p}_k} = C_{k,m}(1,1) + \gamma \sum_{h=1}^{|Y|} O_k(m,h) \alpha_{m,\not{p}_{k+1}(h)} \quad (29)$$

Acknowledgments

The first author acknowledges the support of NSF project CMMI #1638327, titled "CRISP Type 1/Collaborative Research: A Computational Approach for Integrated Network Resilience Analysis under Extreme Events for Financial and Physical Infrastructures." The authors thank the Center for Engineering and Resilience for Climate Adaptation (CERCA) of the CEE/EPP departments at Carnegie Mellon University for inspiring this research.

Notation

The following symbols are used in this paper:

$C_{k,m}(i,a)$ = immediate cost;
$c_k(i,b,a)$ = expected immediate cost;
$e_k(h,b)$ = marginal observation probability;
$H_k(i,a,j,b)$ = expected transition;
$O_k(m,h)$ = observation probability; for this and following symbols, h means *for observation* h;
$Q_{k,m}(i,a)$ = action-value function;
$T_{k,m}(i,a,j)$ = transition probability, $(i,\ j,\ a)$ indicates *for state* i, *next state* j, and *action* a, respectively;
$u_{k+1}(h,\mathbf{b})$ = updated belief at next step;
$\mathbb{V}^*_k(i,\mathbf{b})$ = optimal value with feedback defined by observation probability O;
$\mathbb{V}^\Psi_k(i,\mathbf{b})$ = value following policy Ψ, with feedback defined by observation probability O;
$V^*_{k,m}(i)$ = single-model optimal value;
$V^\Phi_{k,m}(i)$ = single-model value following policy Φ;
$W^*_{k,\infty}(i,\mathbf{b})$ = optimal value under time-invariant belief;
$W^\Phi_{k,\infty}(i,\mathbf{b})$ = value following policy Φ up to the end of the process; for this and following symbols, $\mathbf{b}$ means "from belief $\mathbf{b}$";
$W^*_{k,k+d}(i,\mathbf{b})$ = optimal value under time-invariant belief up to i, and single-model optimal policy after that;
$W^\Phi_{k,k+d}(i,\mathbf{b})$ = value following policy Φ up to t_{k+d-1}, and single-model optimal policy after that;
$\Upsilon(\mathbf{b},\Gamma)$ = lower envelope of affine functions defined by alpha vectors in set $\mathbf{\Gamma}$;
$\alpha_{m,i,k,m'}$ = component m of the alpha vector referring to optimal policy under model m';
$\alpha_{m,\not{p}_k(i)}$ = component m of the alpha vector referring to conditional plan $\not{p}$;
$\alpha_{m,\not{p}_{k+1}(h,j)}$ = component m of the alpha vector referring to conditional plan $\not{p}$ after having observed that Y_{k+1} is equal to h and S_{k+1} is equal to j;
$\Gamma_{k,i}$ = set of alpha vectors;
$\pi^*_{k,m}(i)$ = single-model optimal policy;
$\phi^*_{k,\infty}(i,\mathbf{b})$ = open-loop policy;
$\phi^*_{k,k+d}(i,\mathbf{b})$ = optimal policy with time-invariant belief up to t_{k+d-1}, and perfect model knowledge after that;
$\psi^*_k(i,\mathbf{b})$ = closed-loop optimal policy, with feedback defined by observation probability O;
$\Delta V^I_k(i,\mathbf{b})$ = incremental cost for using the near-clairvoyance policy without information; and
$\Delta V^{II}_k(i,\mathbf{b})$ = incremental cost for using the open-loop policy with perfect information at the next step.

Subscripts

k = at time t_k; and
m = for model m.

References

Bertsekas, D. P. (1995). *Dynamic programming and optimal control*, Vol. 1, Athena Scientific, Belmont, MA.

Chades, I., Carwardine, J., Martin, T., Nicol, S., Sabbadin, R., and Buffet, O. (2012). "MOMDPs: A solution for modeling adaptive management problems." ⟨https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4990⟩ (May 15, 2017).

Dittrich, R., Wreford, A., and Moran, D. (2016). "A survey of decision-making approaches for climate change adaptation: Are robust methods the way forward?" *Ecol. Econ.*, 122, 79–89.

Field, C. B., ed. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change*, Cambridge University Press, Cambridge, U.K.

Hallegatte, S., Shah, A., Lempert, R., Brown, C., and Gill, S. (2012). *Investment decision making under deep uncertainty: Application to climate change*, World Bank, Washington, DC.

Hawkins, E., and Sutton, R. (2009). "The potential to narrow uncertainty in regional climate predictions." *Bull. Am. Meteorol. Soc.*, 90(8), 1095–1107.

IPCC (Intergovernmental Panel on Climate Change). (2014). "Climate change 2014: Impacts, adaptation, and vulnerability." *Working Group II Contribution to the Fifth Assessment Rep. of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, U.K.

Krause, A. (2008). "Optimizing sensing: Theory and applications." Ph.D. dissertation, School of Computer Science, Carnegie Mellon Univ., Pittsburgh.

Kurniawati, H., Hsu, D., and Lee, W. (2008). "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces." *Robotics: Science and Systems IV*, Eidgenössische Technische Hochschule Zürich (ETHZ), Zurich, Switzerland.

Leiserowitz, A., Maibach, E., Roser-Renouf, C. (2012). "Climate change in the American mind: Americans' global warming beliefs and attitudes in March 2012." *Rep. from Yale Project on Climate Change Communication*, Yale Univ. and George Mason Univ., New Haven, CT.

Lin, N., Emanuel, K., Oppenheimer, M., and Vanmarcke, E. (2012). "Physically based assessment of hurricane surge threat under climate change." *Nat. Clim. Change*, 2(6), 462–467.

Madanat, S. (1993). "Optimal infrastructure management decision under uncertainty." *Transp. Res. Part C*, 1(1), 77–88.

Melillo, J. M., Richmond, T. C., and Yohe, G. W., eds. (2014). *Climate change impacts in the United States: The third national climate assessment*, U.S. Global Change Research Program, Washington, DC, 418–440.

Memarzadeh, M., and Pozzi, M. (2016a). "Integrated inspection scheduling and maintenance planning for infrastructure systems." *Comput. Aided Civil Infrastruct. Eng.*, 31(6), 403–415.

Memarzadeh, M., and Pozzi, M. (2016b). "Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling." *Reliab. Eng. Syst. Saf.*, 154, 137–151.

Memarzadeh, M., Pozzi, M., and Kolter, J. Z. (2015). "Optimal planning and learning in uncertain environments for the management of wind farm." *J. Comput. Civil Eng.*, 10.1061/(ASCE)CP.1943-5487.0000390, 04014076.

Memarzadeh, M., Pozzi, M., and Kolter, J. Z. (2016). "Hierarchical modeling of systems with similar components: A framework for adaptive monitoring and control." *Reliab. Eng. Syst. Saf.*, 153, 159–169.

Papakonstantinou, K., and Shinozuka, M. (2014). "Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation." *Reliab. Eng. Syst. Saf.*, 130, 214–224.

Russell, S., and Norvig, P. (1995). *Artificial intelligence: A modern approach*, Pearson Education, Inc., Boston.

Shani, G., Pineau, J., and Kaplow, R. (2013). "A survey of point-based POMDP solvers." *Auton. Agents Multi-Agent Syst.*, 27(1), 1–51.

Smallwood, R. D., and Sondik, E. J. (1973). "The optimal control of partially observable Markov processes over a finite horizon." *Oper. Res.*, 21(5), 1071–1088.

Sondik, E. J. (1978). "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs." *Oper. Res.*, 26(2), 282–304.

Spaan, M., and Vlassis, N. (2005). "Perseus: Randomized point-based value iteration for POMDPs." *J. Artif. Intell. Res.*, 24, 195–220.

Špačková, O., and Straub, D. (2015). "Cost-benefit analysis for optimization of risk protection under budget constraints." *Risk Anal.*, 35(5), 941–959.

Špačková, O., and Straub, D. (2017). "Long-term adaption decisions via fully and partially observable Markov decision processes." *Sustainable Resilient Infrastruct.*, 2(1), 37–58.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement learning: An introduction*, MIT Press, Cambridge, MA.

von Neumann, J., and Morgenstern, O. (1944). *Theory of games and economic behavior*, Vol. 60, Princeton University Press, Princeton, NJ.