Title: Hidden-model processes for adaptive management under uncertain climate change

Authors: Matteo Pozzi, a  
Milad Memerzadeh, b  
Kelly Klima, c

Manuscript submitted to  
ASCE’s Journal of Infrastructure Systems

Version: 6/30/2016

Corresponding Author: Matteo Pozzi

Affiliation: a, Faculty, Dept. of Civil & Environmental Engineering  
Faculty Affiliate, Scott Institute for Energy Innovation  
Carnegie Mellon University  
Address: 107b Porter Hall, 5000 Forbes Av., Pittsburgh PA, 15213-3890  
Phone: (412) 268-5649  
Email: mpozzi@cmu.edu

b, Postdoctoral Scholar, Dept. of Environmental Science, Policy and Management  
University of California Berkeley  
Address: 201 Wellman Hall, UC Berkeley, Berkeley CA, 94720  
Phone: (540) 557-7087  
Email: miladm@berkeley.edu

c, Research Scientist, Dept. of Engineering and Public Policy  
Affiliate, Scott Institute for Energy Innovation  
Carnegie Mellon University  
Address: 254C Posner Hall, 5000 Forbes Av., Pittsburgh PA, 15213-3890  
Phone: (412) 268-3705  
Email: kklima@andrew.cmu.edu
ABSTRACT

Stakeholders and owners of assets exposed to extreme events have to take decisions about investments in mitigation measures. The assumption of the extent of future climate change can significantly affect the optimization of these actions. While a single climate model can be represented by a stochastic process and directly integrated into the optimization procedure, decisions under epistemic uncertainty about the climate model are computationally more challenging. Even if an agent defines a set of models with corresponding probabilities, she has still to define whether and how she will learn about the likelihood of these models. The assumed future “learning rate” about the climate model can play a significant role in the optimization procedure. For example, an agent believing, optimistically, that we will soon learn perfectly what model is right, may prefer to wait for this information before taking relevant decisions. On the other hand, an agent assuming pessimistically that no information will ever be available in the future, may prefer to immediately take an action with long term consequences.

In this paper, we propose a set of optimization procedures based on the Markov Decision Process (MDP) framework, to support decision making as a function of the assumption on future learning, thus trading-off the need of a prompt response with that for sufficient information. Specifically, we outline how an approach based on Hidden Model MDPs and point-based value iteration that can be used for a range of alternative assumption on future availability of information, and related it to simple approaches based on limit-case assumptions. We describe the complexity of these procedures, discuss their performance in different settings, and apply them to a decision making process about flood risk mitigation.

Nomenclature

We provide definitions for a subset of symbols used in the paper, listing them in order of appearance. For all of them, subscript $k$ means: “at time $t_k$”; subscript $m$: “for model $m$”. In parentheses $i$ indicates: “at state $i$”; $j$: “at next state $j$”; $a$: “for action $a$”; and $b$: “from belief $b$”; $\Phi$: “for observation $h$”.

<table>
<thead>
<tr>
<th>symbol</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{k,m}(i,a,j)$</td>
<td>Transition probability</td>
</tr>
<tr>
<td>$C_{k,m}(i,a)$</td>
<td>Immediate cost</td>
</tr>
<tr>
<td>$V_{k,m}^\Phi(i)$</td>
<td>Single-model value following policy $\Phi$</td>
</tr>
<tr>
<td>$V_{k,m}^*(i)$</td>
<td>Single-model optimal value</td>
</tr>
<tr>
<td>$\pi_{k,m}(i)$</td>
<td>Single-model optimal policy</td>
</tr>
<tr>
<td>$W_{k,\infty}^\Phi(i,b)$</td>
<td>Value following policy $\Phi$ up to the end of the process</td>
</tr>
<tr>
<td>$W_{k,\infty}^*(i,b)$</td>
<td>Optimal value under time-invariant belief</td>
</tr>
<tr>
<td>$\Phi_{k,\infty}^*(i,b)$</td>
<td>Open-loop policy</td>
</tr>
<tr>
<td>$Q_{k,m}(i,a)$</td>
<td>Action-value function</td>
</tr>
<tr>
<td>$W_{k,k+d}^\Phi(i,b)$</td>
<td>Value following policy $\Phi$ up to $t_{k+d-1}$, and single-model optimal policy after that</td>
</tr>
<tr>
<td>$W_{k,k+d}^*(i,b)$</td>
<td>Optimal value under time-invariant belief up to $t_{k+d-1}$, and perfect model knowledge after that</td>
</tr>
<tr>
<td>$\phi_{k,k+d}^*(i,b)$</td>
<td>Optimal policy with time-invariant belief up to $t_{k+d-1}$, and perfect model knowledge after that</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>$\Delta V^I_k(i, b)$</td>
<td>Incremental cost for using the near-clairvoyance policy without information</td>
</tr>
<tr>
<td>$\Delta V^{II}_k(i, b)$</td>
<td>Incremental cost for using the open-loop policy with perfect information at next step</td>
</tr>
<tr>
<td>$c_k(i, b, a)$</td>
<td>Expected immediate cost</td>
</tr>
<tr>
<td>$O_k(m, h)$</td>
<td>Emission probability</td>
</tr>
<tr>
<td>$H_k(i, a, j, b)$</td>
<td>Expected transition</td>
</tr>
<tr>
<td>$V^\Psi_k(i, b)$</td>
<td>Value following policy $\Psi$, with feedback defined by emission $O$</td>
</tr>
<tr>
<td>$V^\Psi_k(i, b)$</td>
<td>Optimal value following policy $\Psi$ with feedback defined by emission $O$</td>
</tr>
<tr>
<td>$\psi_k^\Psi(i, b)$</td>
<td>Closed-loop optimal policy, with feedback defined by emission $O$</td>
</tr>
<tr>
<td>$e_k(h, b)$</td>
<td>Marginal emission probability</td>
</tr>
<tr>
<td>$u_{k+1}(h, b)$</td>
<td>Updated belief at next step</td>
</tr>
<tr>
<td>$\Gamma_k$</td>
<td>Set of alpha vectors</td>
</tr>
<tr>
<td>$\alpha_{m, p_k(l)}$</td>
<td>Component $m$ of the alpha vector referring to conditional plan $p$</td>
</tr>
<tr>
<td>$\alpha_{m, p_{k+1}(h, j)}$</td>
<td>Component $m$ of the alpha vector referring to conditional plan $p$ after having observed that $Y_{k+1}$ is equal to $h$ and $S_{k+1}$ is equal to $j$</td>
</tr>
<tr>
<td>$\alpha_{m, i, k, m'}$</td>
<td>Component $m$ of the alpha vector referring to optimal policy under model $m'$</td>
</tr>
<tr>
<td>$\Upsilon(b, \Gamma)$</td>
<td>Lower envelope of affine functions defined by alpha vectors in set $\Gamma$</td>
</tr>
</tbody>
</table>

1. Introduction

Climate change poses specific challenges to the management of infrastructure under climate change. In many problem settings, today’s planning should consider future climate conditions, even decades in advance, which are currently unknown. However, those conditions will progressively reveal themselves, for example through observations available to climate scientists. On one hand, this may suggest postponing critical and expensive decisions until our epistemic uncertainty about climate change will evaporate. On the other hand, procrastination may expose us to high risk in the short term, and prevent us to act promptly.

In this paper, we propose and discuss optimization models for trading-off quantitatively the need of a prompt response and the need of sufficient information. Our application is protection of infrastructure assets against flood and sea-level rise. Research indicates that climate change will cause a worldwide sea level rise of possibly a meter or more by 2100 (IPCC 2007; Melillo et al., 2014), resulting in more coastal flooding events worldwide (Field et al., 2012). While decarbonizing the energy system may help reduce sea level rise (IPCC 2007), likely some amount of sea level rise will occur regardless of greenhouse gas reduction efforts. Some traditional decision analysis, as cost-benefit analysis, may not be appropriate for this setting (e.g., Morgan and Henrion, 1990). As a result, many are beginning to consider decision making under uncertain climate change. Hallegatte et al. (2012) summarizes 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. In addition, the NCA 2014 summarizes approaches, and stresses the importance of understanding the decision-makers’ preferences and application. Dittrich et al. (2016) build on this, comparing real option analysis, portfolio analysis, robust-decision making and no/low regret options and discussing when each of these may be an appropriate decision-making tool.

To model the long-term management, we use dynamic programming approaches for Markov processes. First, we posit that infrastructure planning and management under climate change can be formulated as sequential decision making problem under a non-stationary stochastic model. In
that formulation, the assumed climate model defines the specific stochastic process controlling
the system evolution. Decisions under a single model can be addressed as an optimization
problem that, if the physical state of the system is completely observable, can be formulated and
solved in the framework of Markov Decision Processes (MDPs) (Bertsekas, 1995; Sutton and
Barto, 1998). However, it is often the case that the decision maker, possibly following the expert
evaluation of community of climate scientists, cannot identify one model that completely
represents the uncertainty on the climate evolution. When the model of the dynamic evolution is
itself unknown, the appropriate planning approach should depend on the assumption of
information availability. In the two limit cases that a) perfect information about the model will be
available soon, or b) no information will ever be available, simple solutions can be found. When
only noisy or incomplete observations are available, a decision maker should use the Hidden
Model MDP (HM-MDP) framework, as introduced by Chades et al. (2012). HM-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 (Medury and Madanat, 2013; 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) reviews numerical methods for
solving POMDPs, including Perseus (Spaan and Vlasssis, 2005) and SARSOP (Kurniawati et al.,
2008).

A HM-MDP assumes that the physical state of the system is fully observability, that the time-
invariant 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) apply their
model to the management of a population of the threatened bird species endemic to Northern
Australia. Very recently, Špačková and Straub (2016), leveraging on their previous work on
flexibility in planning (Špačková and Straub, 2015), investigates a similar setting with non-
stationary models of climate change, and proposed a solution method based on Monte Carlo,
quantifying the benefit of flexibility, with example applications to waste water treatment plant
and to flood protection.

Our research specifically focuses on the role of the assumed availability of future information
in current planning. We illustrate how, depending on the assumptions on information
availability, HM-MDP can be formulated and solved. We consider methods that are exact under
limit cases assumption on information availability, and we outline an approximate point-based
value iteration method for solving finite horizon HM-MDPs with non-stationary models.

The rest of the paper is organized as follows: Section 2 reviews planning under a known
Markovian model, Section 3 defines planning under model uncertainty, Sections 4 and 5 applies
the analysis to a toy example and to a decision problem regarding flood-risk mitigation,
respectively, Section 6 defines details on the numerical methods for point-based value iteration,
while Section 7 draws conclusion.

2. Planning under a known model: MDP

2.1 Formulation

We consider an asset to be managed under climate change. We model the evolution of the asset’s
state as an MDP. Time is discretized into set \( \{t_1, t_2, t_3, ..., t_T\} \), and \( S_k \) defines the state at time \( t_k \),
on domain \( \{1, 2, ..., |S|\} \). At time \( t_k \), the manager takes action \( A_k \) on domain \( \{1, 2, ..., |A|\} \), and
pays immediate cost \( C_k \), that depends on current action and state by time-dependent function
Thus for a management process that ends at finite time horizon \( t_T \), the expected discounted cost \( V_k \) for managing the system from time \( t_k \) (that we can call the “value”) is:

\[
V_k = \sum_{i=k}^{T} \gamma^{t-k} \mathbb{E}[C_i] + \gamma^{T+1-k} V_{T+1}
\]  

(1)

where \( \gamma \) is the discount factor, \( \mathbb{E}[X] \) indicates the expectation of random variable \( X \), and \( V_{T+1} \) is the residual cost after the time horizon. While the state may be only partial observable in some applications, here we assume that the physical state is completely observable. The state stochastically evolves from state \( i \) to state \( j \) following the Markov property, according to time-variant transition function \( T_k(i,a,j) = \mathbb{P}[S_{k+1} = j | S_k = i, A_k = a] \), where \( \mathbb{P}[E] \) indicates the probability of event \( E \). As the state evolution, and consequently the future costs, depends on the action taken, the expectation in Eq.1 can be computed only when a specific policy is assigned. A decision-maker can select action \( A_k \) following policy \( \pi_k \) depending on current state \( S_k \), that represents a sufficient statistics in the MDP framework. Following time-variant policy \( \Pi = \{ \pi_1, ..., \pi_T \} \) starting from state \( S_k = i \), the agent gets value:

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

(2)

The optimal value is obtained minimizing Eq1:

\[
V_k^* (i) = \min_a \{ C_k(i, a) + \gamma \sum_{j=1}^{|S|} T_k(i, a, j) V_{k+1}^* (j) \}
\]  

(3)

Optimal policy \( \pi_k^*(i) \), at time \( t_k \), is defined by using “argmin” instead of “min” in Eq.3, and time-variant policy set \( \Pi^* \) by listing policies for all times. An agent adopting \( \Pi^* \) gets the minimum possible value that, it is worth stressing, is an expected value. Eqs.2-3 are forms of the Bellman Equation, and they can be solved iteratively, from \( k = T \) back to \( k = 1 \). Each iteration is an application of the so-called Bellman backward operator.

### 3. Planning under model uncertainty

#### 3.1 Formulation

We now extend previous formulation considering a set of \( M \) possible models, each describing an alternative process evolution. Model indicator \( M \) is one value in domain \( \{1, 2, ..., M\} \), and time-variant functions \( C_{k,m} \) and \( T_{k,m} \) define cost and transition for each model \( m \). By solving Eq.3 for each model, we could identify a set of \( M \) alternative policies \( \{ \Pi_1^*, ..., \Pi_M^* \} \), that may disagree even at initial time \( t_1 \). Given that, the agent has to consider all models jointly and take a decision accounting for the uncertainty among them. Here we assume the agent assigns belief \( \mathbb{P} \) to the models, so that \( b(m) = \mathbb{P}[M = m] \) defines the probability that model \( m \) is the right one.

#### 3.2 Robust planning via open-loop control (pessimistic policy)

The open-loop control scheme identifies an optimal time-variant policy as that performing best, in the expected sense, without any belief updating. Now \( V_{k,m}^\Phi \) indicates the value function according to model \( m \), at time \( t_k \), following policy \( \Phi = \{ \phi_1, ..., \phi_T \} \), which can be identified from Eq.2, by using functions \( C_{k,m} \) and \( T_{k,m} \) instead of \( C_k \) and \( T_k \). We refer to the corresponding value as \( W_{k,\infty}^\Phi \), to indicate persistent model uncertainty: information on the model will never be available or, equivalently, will be available infinitely far ahead. It can be defined as:
$W_{k,\infty}^\Phi(i, b) = \mathbb{E}_m V_{k,m}^\Phi(i) = \mathbb{E}_m C_{k,m}[i, \phi_k(i)] + \gamma \sum_{j=1}^{[S]} \mathbb{E}_m \{T_{k,m}(i, \phi_k(i), j) V_{k+1,m}(j)\}$ \hspace{1cm} (4)

where $\mathbb{E}_m[f_m] = \sum_{m=1}^M f_m b(m)$ indicates the expectation on $m$ using belief $b$. The minimum expected management cost achievable by open-loop control, $W_{k,\infty}^\star(i, b)$, is defined as $W_{k,\infty}^\star(i, b) = \min_\Phi W_{k,\infty}^\Phi(i, b)$. Computationally, this policy can be identified by applying iteratively the Bellman backward operator, from terminal time $t_T$, following this scheme:

$W_{k,\infty}^\Phi(i, b) = \min_\alpha \left\{ \mathbb{E}_m C_{k,m}(i, \alpha) + \gamma \sum_{j=1}^{[S]} \mathbb{E}_m \{T_{k,m}(i, \alpha, j) V_{k+1,m}^\Phi(j)\} \right\}$ \hspace{1cm} (5)

where the optimal policy $\Phi_{k,\infty}^\star = \{\phi_{k,\infty}^\star, \ldots, \phi_{T,\infty}^\star\}$ derives from using “argmin” instead of “min”. This open-loop approach is truly optimal under the assumption that no additional information about the model will ever be available, so that agent’s belief is time-invariant. Actually, this is rarely the case, as perfect observation of the state trajectory or of the actual costs inevitably contains information on the model. However, it may serve as an effective approximation in “pessimistic” scenarios, where the amount of information on the model, available during the management, process is negligible.

### 3.3 Near-clairvoyance planning via action-value function (optimistic policy)

An alternative optimization scheme, that we call “near-clairvoyance”, assumes that the model will be revealed as next step (see Memarzadeh et al., 2015). It is based on the so-called “action-value function”, for model $m$ at time $t_k$, corresponds to the content of the curly brackets in Eq.3:

$Q_{k,m}(i, a) = C_{k,m}(i, a) + \gamma \sum_{j=1}^{[S]} T_{k,m}(i, a, j) V_{k+1,m}^\star(j)$ \hspace{1cm} (6)

Using the action-value function, Eq.3 can be re-written as $V_{k,m}^\star(i) = \min_\alpha Q_{k,m}(i, a)$. The function defines the expected management cost when action $\alpha$ is taken at current time, followed, from time $t_{k+1}$, by residual part of optimal policy $\Pi_{k+1}^\star = \{\pi_{1,m}^\star, \ldots, \pi_{T,m}^\star\}$. Whether action-value functions are available for each model and action, we can compute a new value, following policy $\Phi$, as:

$W_{k,k+1}^\Phi(i, b) = \mathbb{E}_m \{Q_{k,m}[i, \phi_k(i)]\} = \sum_{m=1}^M Q_{k,m}[i, \phi_k(i)] b(m)$ \hspace{1cm} (7)

Subscript $(k, k+1)$ indicates that value is computed at time $t_k$, while policy will switch at time $t_{k+1}$. The corresponding optimal value is defined as:

$W_{k,k+1}^\star(i, b) = \min_\alpha \mathbb{E}_m \{Q_{k,m}(i, a)\} = \min_\alpha \sum_{m=1}^M Q_{k,m}(i, a) b(m)$ \hspace{1cm} (8)

Using “argmin” instead of “min” in Eq.8, we get corresponding optimal policy $\Phi_{k,k+1}^\star = \{\phi_{k,k+1}^\star, \pi_{k+1,m}^\star, \ldots, \pi_{T,m}^\star\}$, where now $m$ identifies the revealed model. This near-clairvoyance approach is optimal under the assumption that perfect information on the model will be available at the next time step: because of this, we can refer to it as an optimistic policy. Under that optimistic assumption, $W_{k,k+1}^\star$ represents the actual value the agent gets. It should be noted that the time discretization plays a key role in identifying the conditions that make the near-clairvoyance policy optimal as, depending on that, “the next time step” refers to different time moment (say: one day, or five years into the future).
3.4 Mixed planning: d-step-clairvoyance policy

By merging previous approaches, the agent can adopt policy $\Phi$ up to time $t_{k+d-1}$ and, counting on perfect information, switch to single-model optimal policy from time $t_{k+d}$. To assess the corresponding value, we can apply the single-model Bellman backward operators of Eqs. 3 from $T$ back up to time $t_{k+d}$, and the open-loop operator of Eqs.2 and 4 from time $t_{k+d-1}$ back to time $t_k$, getting value $W^\Phi_{k,k+d}(i, b)$. By using Eqs.3 and 5, on the other hand, we get the optimal value $W^*_{k,k+d}(i, b)$ and corresponding policy $\Phi^*_{k,k+d} = \{\phi^*_{k,k+d}, \phi^*_{k+d-1,k+d}, \pi^*_{k+d,m}, \ldots, \pi^*_T,m\}$ where, again, $m$ identifies the revealed model.

We call this the “d-step-clairvoyance policy”, and it is optimal when no information is available for the first $(d-1)$ steps, and perfect information is available after that. By considering parameter $d$ varying from 1 to $\infty$ (actually, up to $T + 1 - k$), policies vary from near-clairvoyance to open-loop.

3.5 Comparison of near-clairvoyance and open-loop policies

An agent should choose among the policies outlined above by considering whether and when perfect information on the model will be available. Knowing that the model will be reveal at time $t_{k+d}$, she will adopt policy $\Phi_{k,k+d}$ and pay, at time $t_k$, expected discounted cost $W^\Phi_{k,k+d}$. As “information never hurts”, having perfect information at an earlier stage is always better (strictly speaking: it is not worse), thus, if $k' < k''$ then $W^*_{k,k'} \leq W^*_{k,k''}$. 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. Following the path of Memarzadeh and Pozzi (2016a), we discuss bounds on the performance of alternative policies in different settings. On one hand, we can reasonably 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 not receiving the assumed perfect information in time.

Here we compare the performance of the open-loop and of the near-clairvoyance policies under opposite scenarios of available information. We assume that if an agent following the near-clairvoyance policy will not receive information at next step, she will assume again perfect information at the following step. Doing so, she follows policy $\Phi^*_{k+1} = \{\phi^*_{k+1,k+1}, \phi^*_{k+2,k+1}, \ldots, \phi^*_{T,k+1}\}$. By relying on the principle that “information never hurts”, and the definition of optimal policy, the following sequence of inequality holds:

$$W^*_{k,k+1}(i, b) \leq W^\Phi_{k,k+1}(i, b) \leq W^*_k(i, b) \leq W^\Phi_{k+1}(i, b)$$  \hspace{1cm} (9)

Eq.9 reads as follows: the cost of using the near-clairvoyance policy under perfect information (at next step) is less than that of using the open-loop policy under perfect information, that, in turn, is less than that of using the open-loop policy under no information that, in turn, is less than that of using the near-clairvoyance policy under no information. In details, we justify the three “less-or-equal” signs in Eq.9, from right to left, noting that they refer to “a better policy under the same information”, “more information under the same policy” and, again, “a better policy under the same information”, respectively.

To quantify how the policies perform under alternative scenarios of available information, we define $\Delta V^j_k$ and $\Delta V^{j'}_k$ as the increment of cost using the near-clairvoyance policy without actual information (respect to using the optimal appropriate open-loop policy), and that using the
open-loop policy when perfect information is available at next step (respect to using the optimal clairvoyance policy), respectively, from time $t_k$:

$$
\begin{align*}
\Delta V^I_{k}(i, b) &= W^*_{k+1}(i, b) - W^*_{k, \infty}(i, b) \\
\Delta V^H_{k}(i, b) &= W^*_{k+1}(i, b) - W^*_{k,k+1}(i, b)
\end{align*}
$$

(10)

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

Computation of these regrets may shed light on the sensitivity to the available information, depending on the adopted policy. Because of Eq.9, we can bound regret $\Delta V^H_{k}$ from above, with $(W^*_{k, \infty} - W^*_{k,k+1})$, that can be easily computed and represents the so-called Value of Information for having perfect model observation at next step, respect to not having any information. Regret $\Delta V^I_{k}$, 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 information that was supposed to be available turns out to be not so.

3.6 HM-MDP: observation on the climate model, and model updating

In previous Sections, we have assumed perfect information at one time during the process. In most cases, however, information flows more smoothly. In the general HM-MDP setting, we assume that indirect observations of the model can be available at each time. $Y_k$ indicates the observation at time $t_k$, on domain $\{1, 2, ..., |Y|\}$, and it is defined by emission function $O_k(m, h) = \mathbb{P}[Y_k = h|\mathcal{M} = m]$. It is to be noted that the set of $|Y|$ possible observations is not necessarily mapped to a time-invariant set of physical measurements: the statement $Y_k = 1$, for example, 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 on the time-invariant models can be updated sequentially following Bayes’ formula: If $Y_k$ assumes value $h$, the update at time $t_k$ reads:

$$
\begin{align*}
b_{k+1}(m) &\propto O_k(m, h)b_k(m)
\end{align*}
$$

(11)

where now $b_1$ is the initial belief while posterior belief $b_k$ at time $t_k$ is described, for $k > 1$, as $b_k(m) = \mathbb{P}[\mathcal{M} = m|Y_1, ..., Y_k]$. In principles, belief should also be a function of previous trajectory of actions and states, which contains information about the climate model. However, we consider this contribution negligible.

Figure 1 shows a decision diagram 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 decision variables, diamonds costs, continuous links define the conditional dependence structure, while dotted arrows indicate what information is available when a decision is taken. Shaded variables are observable. It is to be noted how no observation is considered as time $t_1$, as we consider all information at to be already embedded in initial belief $b_1$. Although the costs $C_k$ are also a function of the model, for clarity we do not include the corresponding links in Figure 1.
In the process outlined above, the agent should iteratively process observations and take actions depending on the updated belief. By forecasting future updating steps depending on the assumed available information, and behaving consequently, the agent adopts a “closed-loop” policy. All information for taking a decision are summarized by the augmented state \( (\mathcal{N}, \mathcal{A}) \), where ordinal \( i = S_k \) indicates the observable physical state, and vector \( \mathbf{b} = \mathbf{b}_k \) indicates belief current on the model. Now, time-variant policy \( \Psi = \{\psi_1, ..., \psi_T\} \) is defined on this augmented state. Adapting the notation of Memarzadeh and Pozzi (2016b), the value \( \mathcal{V} \), following policy \( \Psi \) is:

\[
\mathcal{V}_k(i, \mathbf{b}) = c_k[i, \mathbf{b}, \psi_k(i, \mathbf{b})] + \gamma \sum_{h=1}^{\mathcal{Y}} e_k(z, \mathbf{b}) \sum_{j=1}^{\mathcal{S}} H_k[i, \psi(i, \mathbf{b}), j, \mathbf{b}] \mathcal{V}_{k+1}[j, \mathbf{u}_{k+1}(h, \mathbf{b})]
\]

(12)

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

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

(13)

In Eq.13, the updating follows Bayes’s formula of Eq.12. Bellman equation for optimal value reads:

\[
\mathcal{V}_k^*(i, \mathbf{b}) = \min_{\mathcal{A}} \{c_k(i, \mathbf{b}, \mathbf{a}) + \gamma \sum_{h=1}^{\mathcal{Y}} e_k(h, \mathbf{b}) \sum_{j=1}^{\mathcal{S}} H_k(i, \mathbf{a}, j, \mathbf{b}) \mathcal{V}_{k+1}^*[j, \mathbf{u}_{k+1}(h, \mathbf{b})]\}
\]

(14)

As before, the optimal closed-loop policy \( \Psi^* = \{\psi_1^*, ..., \psi_T^*\} \) derives from previous equation, by using “argmin” instead of “min”. When belief is represented by standard basis vector \( \mathbf{v}_m \), made by all zeros except for a one at position \( m \), the agent knows for sure that model \( m \) is correct, and corresponding value \( \mathcal{V}_k^*(i, \mathbf{v}_m) \) is equal to single-model value \( V_{k,m}^*(i) \), as the agent cannot learn no more and policy \( \Pi_m^* \) is optimal indeed.
In summary, the closed-loop formulation allows for representation of a specific model of information availability, via emission function $O_k(m, h)$, that needs not to be as extreme those in Sections 3.2 to 3.5. The values related the open-loop and near-clairvoyance policies in those settings provide bounds for that of the closed-loop value:

$$W_{k,k+1}^*(i, b) \leq W_k^*(i, b) \leq W_{k,\infty}^*(i, b)$$  \hspace{1cm} (15)

We postpone to Section 6 details on the computational approach to solve Eq.14.

4 A toy example about the impact of the assumed available information

We start illustrating the role played by assumed available information in a simple model. An asset can be protected against extreme events, and an agent has to take a decision about its protection. We consider two stationary models: Model 1 defines an extreme event occurring with probability $P_1$ per time step, while Model 2 defines an extreme event occurring with probability $P_2 > P_1$ per time step. Failure cost $C_F$ is incurred when an extreme event occurs on an unprotected asset, independently of model or time. Investment cost $C_I$ can protect the asset indefinitely, and discount factor is $\gamma$. Belief $b$, on the two models, is of the form $[1 - \theta \quad \theta]$, where $\theta = P[m = 2]$, so $\theta$ defines the probability of the more risky model.

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

$$\theta_{nc} = \frac{r}{1 - \gamma(1 - r)}$$  \hspace{1cm} (16)

where $r = (C_I - D_1)/(D_2 - D_1)$ is the normalized investment cost. Figure 2(a) illustrates how $\theta_{nc}$ varies depending on normalized investment cost and discount factor. It is to be noted that for every problem setting (summarized by $r$) the open-loop policy is more conservative than the near-clairvoyance one. In details, except for the extreme values or $r$ equal to zero or one (for which the decision problem is trivial) or $\gamma$ equal to zero, the near-clairvoyance policy always tolerates higher values of $\theta$, before removing the risk by investing. This gap is monotonically increasing with the discount factor: for $\gamma$ approaching one, the near-clairvoyance policy prescribes to wait for the next time, without investing, except when $\theta$ is one (i.e., when Model 2 is certainly the actual one). When perfect knowledge is assumed $d$ time steps ahead, the corresponding threshold can be read in Eq.16 plugging $\gamma^d$ instead of $\gamma$: consequently, the threshold becomes closer to the open-loop one if knowledge is postponed.

In Figure 2(b), we fix $r$ to 50% and $\gamma$ to 95%, and we investigate how the closed-loop optimal policy looks like when imperfect information on the model is available. We assume that observations are binary variables, and that emission function $O$ is stationary, with $Y_k$ supposedly indicating model variable $m$. We allow for errors in the observations and, in details, we assume that $\varepsilon = P[Y_k \neq m]$ defines the probability of a wrong measure at each step, so $\varepsilon$ equal to zero indicates a perfect observation, while $\varepsilon$ equal to 50% indicates an independent irrelevant measure. For each value of $\varepsilon$, the optimal policy (identified as illustrated in Section 6) prescribes to wait until $\theta$ reaches a threshold ($\theta_{thr}$) before removing the risk. $\theta_{thr}$ starts at 50% (equal to $r$,...
i.e. equal to $\theta_{o1}$) and monotonically increases up to $\theta_{nc} = 95.2\%$ when $\varepsilon$ is zero and observation is perfect.

Overall, this toy example shows how the optimal policy depends on the assumed learning rate and how a faster leaning rate suggests a less conservative policy: consider, for example, an agent assigning 85% probability to Model 2, while normalized cost is $r = 50\%$. Despite it seems that, in this circumstances, one should invest and remove the risk, no suggestion about the optimal decision can be made without an assumption on the future information. In facts, if accurate observations are available (in details, with $\varepsilon$ less than 30%), it is optimal to wait, as can be read in Figure 2(b). Similarly, it is optimal to wait if perfect information is available within d less or equal than 3 steps.

![Figure 2: Optimal policy’s threshold for the near-clairvoyance approach, as a function of normalized investment cost and discount factor (a), threshold for the closed-loop policy, as a function of the observation precision, for 50% normalized investment cost and 95% discount factor (b).](image)

5 Illustrative application on flood-risk mitigation

5.1 Setting

Let us now consider the decision making process for protecting infrastructure near the coast. While any location near the ocean works, let us consider a location near Battery Park, NY where current and future flood risks have been estimated (e.g., Lin et al., 2016). Furthermore, while our decision-maker could range from individual to country scale, here we specifically focus on a single family house. The homeowner, even if they have a mortgage, is not currently forced to buy flood insurance through the National Flood Insurance Program (FEMA 480), or has slightly less than a 1% chance of flooding per year.

Depending on what the homeowner believes about the house’s chance of flooding now and in the future, and the assumed availability of future information, it may be optimal to invest and elevate the house, to reduce the risk. The agent can select when to elevate the house, and of how much. As in the example in Section 4, on 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 high chance of flood under some models. We investigate how the decision should depend on these assumptions.

5.2 Risk modeling

As in Špačková and Straub (2016), the agent considers three climate models, and defines the occurrence of extreme events by a varying, in time, the parameters of an extreme value
distribution. First, recognizing that many in the United States do not believe in climate change (e.g., Leiserowitz et al., 2012), we include Model 1 where “no change” occurs. Model 2 predicts “low change”, and model 3 a higher increment, that we call “high change”. This latter model derives from advanced hydrological estimates under a high emissions scenario of approximately 1m of sea level rise by 2100 as given in Lin et al., 2016 (Figure 4), while the second model is exactly in-between these two risks.

Specifically, we discretize time in year, and we indicate with variable $z$ the annual maximum flood height. For a given model, we assume annual maxima to be independent, and $p(z|m, k)$ refers to the probability density for year $k$, according to model $m$, which is a Gumbel with location parameter $\alpha_{m,k}$ and scale parameter $\beta_{m,k}$. Parameters for each model are reported in Table 1.

Table 1: parameters of the Gumbel distribution modeling flood annual maxima.

<table>
<thead>
<tr>
<th></th>
<th>$m = 1$</th>
<th>$m = 2$</th>
<th>$m = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{m,k}$ [y]</td>
<td>43.45%</td>
<td>44.55% + 0.55% $k$</td>
<td>46.33% + 1.44% $k$</td>
</tr>
<tr>
<td>$\beta_{m,k}$ [y]</td>
<td>1</td>
<td>100.1% + 0.06% $k$</td>
<td>101.2% + 0.62% $k$</td>
</tr>
</tbody>
</table>

Figure 3 plots the probability of density of $z$ for different times and models. As mentioned above, Model 1 is stationary, while the other two models assume a distribution at year 1 very close to that of model 1, but they assume an increment for the following years.

Under model $m$, the probability of a flood above level $z_e$ occurring in year $k$ is $X_{k,m} = 1 - F_{\text{Gumbel}}(z_e; \alpha_{m,k}, \beta_{m,k})$, where $F_{\text{Gumbel}}(z; \alpha, \beta)$ indicates the cumulative Gumbel distribution with position parameter $\alpha$ and shape parameter $\beta$, computed in $z$. Figure 4(a) shows this probability as a function of time $t_k$. For level $z_e$ equal to 2m, that probability is constant 1% under model 1, and it grows up to 12% under model 2 in one century, and up to 33% under model 3. The initial probability is decreased by one and two orders of magnitude for level $z_e$ equal to 3m and 4m respectively. After one century, the same levels of elevation reduce the probability to 4% and 1.25% under model 2 and to 12% and 3.5% under model 3.

It should be noted, from Table 1 and from Figure 4(a), that the three models do not agree even on the flood probability at year 1. The specific setting we use corresponds to the assumptions that the models used to agreed three years ago, and now the models no longer agree. This attempts to capture the concept that some agents have a belief that climate change has
already been occurring, but are still uncertain on the identity of the right model. The results are similar if the specific setting on the flood probability at year 1 is the same for all models.

We assume the initial level of the house is \( z_0 = 2 \text{ m} \). The decision variable \( \Delta z \) indicates the elevation is meter, so that \( (z_0 + \Delta z) \) is the house level after the decision has been made. At every year, the agent selects a value for \( \Delta z \), however we assume that it is inconvenient to elevate the house more than once during the management process, so \( \Delta z \) can be higher than zero only at one time.

We discretize the domain of \( \Delta z \) in 13 values, from 0 to 3m, equally spaced with an interval of 25cm. The “physical state” of the system is for us the house level, so the problem is described by 13 possible state, and state \( i \) corresponds to level \( z_i = z_0 + 25 \text{ cm}(i - 1) \). The cost of elevating the house includes a fixed cost for intervention and a term proportional to the elevation value: it is nil when \( \Delta z \) is zero, and it is \( C_E = $25K + $25K \cdot \frac{\Delta z}{3\text{ m}} \) for positive \( \Delta z \). The dashed-dotted line in Figure 4(b) shows \( C_E \) as a function \( \Delta z \).

Figure 4(a): probability of flood exceeding level \( z_e \), depending on time and model (a), expected discounted management cost as a function of model and initial elevation (b).

In any year when the house is flooded, we assume an incurred cost \( C_F \) of $180K that includes damages, downtime and repairs. Discount factor is 95% (corresponding to a discount rate of 5.26%). We want to model a process lasting 100 years; however, to avoid identifying an appropriate residual value at that time, we use a time-horizon of 200 years (i.e., \( T = 200 \)) and no residual value (i.e. \( V_{T+1} = 0 \)): in other words, despite we do not trust the climate modeling beyond the 100 years horizon, we assume that this assumption correctly models the process in the first 100 years.

The overall cost includes the flood-related risk and, possibly, if starting at state 1, the cost for elevation, while it is impossible to change the state when the state is higher than one. Action \( a \) takes the state to value \( a \), so transition matrix is deterministic and independent of the model: \( T_{k,m}(i,a,j) = \delta_{a,j} \), where \( \delta_{a,j} \) is the Kronecker delta. So the house level when action \( a \) is implemented is \( z_{a} \), and the cost matrix is defined as:
5.3 Single model, open loop and clairvoyance analyses

We start the analysis assuming the agent takes the action at time $t_1$, with no possibility of acting in the following years. We indicate with $V'$ the expected discounted management cost, and it is a function of initial elevation action $\Delta z$ and of model $m$, as shown in Figure 4(b). Solution derives by a simple application of Eq.2. In the same figure, circles indicate optimal actions: the agent should not elevate the house under “no change”, elevate to 1.75m under “low change” and to 2.25m under “high change”.

In general, the optimal policy, once we remove the constraints of acting at time $t_1$, may be different respect to what showed in Figure 4(b): even inside a single model, the agent may choose to wait and act at a future time, e.g. for discounting the cost of elevation. However, in the specific case, the optimal elevation time is at the beginning of the management process, and that figure also shows the optimal policy with perfect knowledge about the model. Figure 5(a) plots the optimal policy for a non-elevated house (i.e., for $S_k = 1$), depending on time. As expected, the agent never elevates under stationary Model 1. Under Model 2 or 3 if, for some reason the agent has not elevated the house in the beginning, the optimal elevation value $\Delta z^*$ is higher and higher, up to the maximum allowable value 3m. In comparison, for those belief values in that time range, the near-clairvoyance policy always prescribes not to elevate, counting on perfect information at next step.
Figure 5: Optimal action $\pi^*_{k,m}(1)$ for a non-elevated house under perfect model information, depending on time and model (a), corresponding action in the open-loop policy $\phi^*_{k,\infty}(1,b)$, for two specific belief values.

By applying Eqs.5, 8 and the procedure of Section 3.4, we can find out the open-loop, near-clairvoyance and $d$-step-clairvoyance. Figure 6 reports some of those policies, for $S_1 = 1$, time $t_1$ and for each possible belief $b$. The belief is a 3-component vector normalized to one, so its belief can be represented in a 2-D region: each point in the triangle is related to a possible belief, and vertexes refer to perfect model knowledge. In figure (a), initial open-loop policy $\phi^*_{1,\infty}$ is represented via colors: non-shaded area refers to “do-nothing” (i.e. $\Delta z^* = 0$), while each color is related to one specific elevation value. Obviously, the value at the vertexes is consistent to the initial optimal actions of individual models, as showed in Figure 5. Generally, the higher the belief in climate change, the higher the elevation. However, the specific policy is quite complicate. Before discuss it, we compare it with the $d$-step-clairvoyance policy when model is reveled at year 10, at year 5 and with the near-clairvoyance policy (when it is reveled next year). The smaller $d$, the more optimistic is the scenario about learning. Thus, for small values of $d$ the agent will prefer to wait, postponing the decision about elevation, unless she is almost certain about the model. In this latter case (i.e., in the vertexes of the belief domain), all policies are consistent.
Figure 6: open loop initial policy (a), policy with perfect information at step 10 (b), at step 5 (c) and near-clairvoyance policy (d) as a function of the belief.

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

$\Delta z^* = 2.25m$
Figure 7: optimal discounted expected management cost with no information (a), and with perfect information at next step (b).

We report in Figure 8(a) the incremental cost $\Delta V^I_1$, as the difference between the value incurred by adopting $\Phi^*_{1,+1}$ and that by adopting $\Phi^*_{1,\infty}$ with no information. As expected, the incremental cost is always positive (strictly speaking: non negative) and it is zero at the vertexes, where policies are consistent. For example, for an initial belief that assume $m$ equal to 1 or 2 with 50%/50% probability, the incremental cost is about $20K. Figure (b) reports the complementary graph: the incremental cost $\Delta V^I_1$ of adopting $\Phi^*_{1,\infty}$, instead of $\Phi^*_{1,+1}$ when perfect information is available at next step. In this case, the incremental cost for the belief cited above is about $1.25K. As noted above, $\Delta V^I_1$ tends to be higher than $\Delta V^I_1$, as the penalty for not having anticipated information is more severe than the benefit of having additional information beyond that anticipated.
We conclude this section by providing more details on the policy plotted in Figure 6(a). The strips of non-shaded area in that figure may look weird. To better present that outcome, Figure 5(a) reports the open-loop policy under two specific beliefs. When the probabilities of \( m \) equal 1 or 2 are 70% and 30% respectively, \( \Phi_{1,\infty}^* \) prescribes to elevate the house of 1.25m, while it prescribes to do nothing when those probabilities are 60% and 40% respectively. As the latter belief assigns a higher probability of climate change respect to the former, this may seems counter-intuitive. Actually, in this latter belief, policy \( \Phi_{2,\infty}^* \) just postpones the elevation of one year, but it adopts a higher value (\( \Phi_{2,\infty}^* \) corresponds to \( \Delta z^* = 1.5m \)), and it is never below the policy of the former belief after the first year.

5.4 Closed-loop analysis

We define the emission function by implicitly modeling possible evolution of the belief, following the procedure detailed in the Appendix. Instead of directly modeling possible advancement in climate analysis, the underlying idea to model the observations to the current flood probability for a non-elevated house, \( X_{k,m}(z_0) \), using a “noise-level” parameter \( \zeta_Y \) that defines the uncertainty affecting the annual observation.

Figure 9 plots \( O_k(m,h) \) for all values of \( h \) from 1 to \( |Y| \), depending on the model, for time \( t_k \) equal to 5, 10 and 20 years, and \( \zeta_Y \) equal to 50%, 1 and 2. We note that emission functions for different models are more and more separated as time passed, as consistent with Figure 3: this is because the difference among predictions of the alternative models grows with time. Furthermore, the inverse of noise-level parameter \( \zeta_Y \) can be related to the “learning rate”: when \( \zeta_Y \) is close to zero, perfect information is available at next step; on the contrary, for large \( \zeta_Y \) the emission is flat across models, and observations contain no information about the correct model.
Figure 9: Emission matrix $O_k(m, h)$, at different times and for different “noise-level” parameter.

Figure 10(a) reports 1,000 forward simulations of the belief, starting from initial belief $b_1 = [34 \ 32 \ 34]$%, that is represented by a point close to the middle of the triangle shown in Figure 8. Beliefs are simulated by sampling observations according to their probability, and process them: details on the forward simulations are reported in Appendix A. The 3 columns refer to different leaning rate: $\zeta_Y$ equal to 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. In the second and third rate, on the contrary, the learning process is faster and, for example, it is highly improbable that the belief is still similar to the initial one, after 50 years when $\zeta_Y$ is 50%. The reason is that, for that “noise-level” observations collected during 50 years are almost sufficient to reveal exactly the model. Generally, in the long-term the belief tends to converge to the right model, and therefore the simulations migrate to the domain’s vertexes. It is interesting, however, to note how they reach the vertexes. First, the simulations tend to move away from the mid-point in the left side of the triangle. That point, in facts, represents belief $b = [50 \ 0 \ 50]$%, that is uncertain between “high” and “no change”, but it excludes the possibility of “low change”: practically no sequence of observations leads to that outcome. Moreover, simulations tend to concentrate in the lower and right sides, which
represent uncertainty between two out the three models: between “high” and “low climate change”, or between “low” or “no change”. Lastly, we note that convergence to \( m \) equal to 1 or 3 is faster than to \( m \) equal to 2: the “low change” model, in facts, is placed between two other models, so both low and high observations coming from it may to be misled as coming from other models. On the contrary, the other two models are “free from one side”: for example, the “no change” model can be identified in a fast way if low-value observations are systematically collected. Figure (b) reports the corresponding closed-loop policy for the initial year. In the slow-learning scenario (i.e., for \( \zeta_Y \) equal to 2), that policy is remarkable similar to the open-loop one, reported in Figure 6(a), and it would be identical for larger \( \zeta_Y \). However, for higher leaning rates (i.e., for \( \zeta_Y \) equal to 1 or 50%), the agent prefers to do-nothing and wait, unless the probability of no climate change (Model 1) is less than about 30%. Again, when noise-level \( \zeta_Y \) goes to zero, the optimal policy converge to the near-clairvoyance one, reported in Figure 6(d).

![Figure 10: Outcomes of 1,000 forward simulations, depending on time and noise-level in the observations](image)

(a). Optimal closed-loop policy depending on the noise-level (b). Each triangle represents the belief domain, with the same scale as those in Figures 6-8.

In Figures 11 (a) and (c), we compare value and policies for the \( d \)-step-clairvoyance policy with closed-loop policy, for four alternative initial beliefs. Value \( W_{1,d}^*(1, \mathbf{b}) \) grows monotonically (strictly speaking: it grows or stays constant) with time \( t_d \), and value \( \Psi_1^*(1, \mathbf{b}) \)
grows with $\zeta_Y$. The blue crosses for $\zeta_Y$ equal to 1% and to 10 represent the value for near-clairvoyance and open-loop policies, respectively, that can be read in Figure 7. The gap between the values for two different coordinates in the horizontal axis quantifies the benefit of identifying the model earlier, or of having better observations: clearly, this depends on the belief (e.g., it would be zero if the model was already known). Figures (b) and (d) show the corresponding initial optimal action, represented in terms of elevation value $\Delta z^*$: it is zero under a threshold that depends on the belief, and consisted with the open-loop policy above that. Consequently, the value in graph (a) is flat above the threshold, as the policy is invariant respect to higher noise-level.

Figure 11: Value as a function of time when model will be revealed, for 4 selected initial beliefs $b_1$ (a), corresponding policy (b), value as a function of noise level (c) and corresponding policy (d).

5.5 Variations respect the original setting: alternative action set

To illustrate how the solution depends on the available actions, we now suppose that the only alternative respect to not elevating the house is to elevate of $\Delta z = 1$ m. Figure 12 plots the corresponding open-loop policy (a), closed-loop policy with $\zeta_Y$ equal to 50% (b), and near-clairvoyance policy (c), where shaded area refers to elevating the house. By comparison with Figures 6 and 10, it is clear that available actions strongly affect these policies. Specifically, the near-clairvoyance policy is now less conservative than that in Figure 6(d), as the agent does not need to select an appropriate (and potential expensive) intervention anymore, and not much information is needed.
Figure 12: initial action of the “open-loop” policy (a), the “closed-loop” policy with noise level equal to 50% (b), and the “clairvoyance” policy (c), when alternatives are: elevating of 1m (shaded area) or not elevating. Each triangle represents the belief domain, with the same scale as those in Figures 6-8.

5.6 Variations respect the original setting: many climate models

The example shown before included only three models, for the sake of illustration, as no belief’s domain can be shown when $M$ is higher than 3. Figure 13(a) shows the annual maxima distribution (as in Figure 3) and Figure 13(b) the probability of floods exceeding $z_e$ (as in Figure 4(a)) for a set of 10 models, by adding 7 models to the 3 used above, all intermediate between the “no change” and the “high change” models used before.

Figure 13: probability density for annual maximum at year 100 (a), probability of flood exceeding level $z_e$, depending on time and model (b), for $M = 10$.

From an initial belief assigning 19% probability to Model 1 with no change, and 9% to each of the other models, we repeat the parametric analysis that concludes Section 5.4, and plot the results in Figure 14, as in Figure 11, to show how the analysis can be performed in high-dimension.

6.1 Conditional plan and alpha vectors

We now illustrate how to solve numerically the HM-MDP formulated in Section 3.6. The hardest part in Eq.14 is to model appropriately future value \( \Psi_{k+1}^* \). 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:

\[
\Psi_k^*(i, b) \leq \min_{a \in \Gamma_{k,i}} [\alpha^T b]
\]  

(18)

where \( \Gamma_{k,i} \) is the set of so-called “alpha-vectors” for physical state \( i \), referring to time \( t_k \). Each alpha-vector is of dimension \([M \times 1]\) and it refers to a specific conditional plan (Russell and Norvig, 1995). At time \( t_k \) and state \( S_k = i \), conditional plan \( \nu_k(i) \) assigns current action \( A_k = a[\nu_k(i)] \), and an action at each future time depending on the sequence of observations. Depending on observation \( Y_{k+1} = h \) and next physical state \( S_{k+1} = j \), the plan continues into a new conditional plan \( \nu_{k+1}(h, j) \). Hence, conditional plan \( \nu_k(i) \) can be described by the initial action \( a[\nu_k(i)] \) and the set of conditional plans \( \nu_{k+1}(h, j) \), for each possible state \( j \) and observation \( h \). Consequently, from \( M_{k+1} \) conditional plans at \( t_{k+1} \), \( M_k = \sum_{j=1}^{\mid S \mid} \mid A \mid \) 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 one. If \( \Gamma_{k,i} \) contained all possible alpha-vectors, referring to all possible plans, we could use an equal sign in Eq.18.

To build an alpha-vector from the corresponding conditional plan, sufficient is to solve the problem for each model. Following Eq.12 for \( b = v_m \), component \( m \) of the alpha-vector for conditional plan \( \nu_k(i) \) can be expressed as:

\[
\alpha_{m,\nu_k(i)} = c_{k,m}[i, a[\nu_k(i)]] + \gamma \sum_{h=1}^{\mid Y \mid} O_k(m, h) \sum_{j=1}^{\mid S \mid} T_{k,m}[i, \pi_{k,m}(i, j)] \alpha_{m,\nu_{k+1}(h,j)}
\]  

(19)

where \( \alpha_{m,\nu_{k+1}(h,j)} \) is component \( m \) of the alpha vector related to conditional plan \( \nu_{k+1}(h, j) \).
Alpha vectors for each time can be built by initializing them at the end of the time horizon, and using Eq.19 as a Bellman backward operator. However, as noted before, the number of vectors grows exponentially and, after few steps back, the complete becomes intractable. Next section illustrates who to approximate the value with a limited value of alpha vectors.

6.2 Point-based value iteration

The idea of the point-based value iteration method for POMDP is to include only the alpha-vectors relevant for representing accurately the set of beliefs reachable from the initial one. Let us suppose $\Gamma_{k+1,j}$ contains a set of alpha vectors able to appropriately represent $\mathcal{V}_{k+1}(j,b)$, as in Eq.18. We can approximate Eq.14 as follows:

$$\mathcal{V}_{k}(i,b) \leq \min_a \left\{ c_k(i,b,a) + \gamma \sum_{h=1}^{|Y|} e_k(h,b) \sum_{j=1}^{|S|} H_k(i,a,j,b) \mathcal{V}[u_{k+1}(h,b),\Gamma_{k+1,j}] \right\} \tag{20}$$

where $\mathcal{V}(b,\Gamma) = \min_{a,b}[\alpha^Tb]$. By solving Eq.20 for a specific pair $(i,b)$, we can not only find out the corresponding optimal value, but also an initial action $a$ and a dominating alpha-vector for each pair $(h,j)$, corresponding to conditional plan $p_{k+1}(h,j)$:

$$b \to \forall(h,j): \alpha_{p_{k+1}(h,j)} = \arg\min_{a,e} \mathcal{V}[u_{k+1}(h,b)] \tag{21}$$

So we can derive a specific conditional plan at time $t_k$ as action $a$, followed by set of alpha-vectors $\{\alpha_{p_{k+1}(h,j)}\}$ with $(1 \leq h \leq |Y|; 1 \leq j \leq |S|)$ so that this conditional plan is optimal from belief $b_k$ and state $S_k$ defined by $b$ and $i$ respectively. Following this observation, we can select relevant alpha-vectors as those corresponding to conditional plans that are optimal for a set of relevant points, reachable from the initial belief. Despite the number of reachable points may grows exponentially in time, we can rely on a limited number of $N$ independent forward Monte Carlo simulations, from initial belief $b_1$, as described in the Appendix, 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 $\alpha_{m,l} = \mathcal{V}_{T+1}(i,m)$, where we assume residual value $V_{T+1}$ may depends 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 $k$ equal to $T$ down to 2 and for each sampled belief, we identify the optimal conditional plan using Eqs.20-21, and the corresponding alpha-vector using Eq.19, populating set $\Gamma_{k,j}$, for each state $j$. Each alpha-vector is also associated with a specific initial action.
2. By solving Eq.18 for $k$ equal 1, we identify optimal initial action $a$, and corresponding value.

Figure 16 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 represents by red points in all domains, and they are the same for all states. For illustration, we focus on one specific belief, $b$, and state equal to 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 $\mathcal{V}_{n+1}$ defines the coordinate of the updated belief, and $\mathcal{V}_{n}$ the probability of the transition in the belief. Each state at time $t_{n+1}$ is related to a set of alpha vector, so that value at each blue point can approximately computed by Eq.18. Now, expected transition $H_n(1,a,S_{n+1},b)$ can also be computed for each action $a$ and new state value. By combining transition in state and belief, and integrating immediate cost, we
can identify optimal action using Eq.20, which acts as a Bellman backward operator. The procedure has to be 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 $Y$ is invariant respect to state $S_n$.

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{figure15.png}
  \caption{Illustrative representation of the procedure for identifying optimal conditional plans.}
\end{figure}

Figure 10 illustrates the importance of identifying the reachable beliefs. While a large number of alpha-vectors may be needed for the approximating the value function with high precision in the entire domain, according to Eq.18, a limited number may be sufficient when we restrict our attention to the region in the belief domain that can be reached. For example, for $\zeta_Y$ equal to 50% and $t_k$ equal to 5 years, belief is concentrated on two sides of the triangle: therefore the value function should be well represented in that region, while the quality of the approximation outside that region is irrelevant. The method outlined above searches for vectors that are relevant for that region, as they dominate other vectors for some reachable points.

The number of alpha-vectors to be included in set $G$ depends on the target quality of approximation. The quality grows with the number of vectors, however vectors always dominated in the set of reachable beliefs can be pruned. More specifically, by neglecting vectors that provide small improvements on the value approximation, the number of vectors can be kept low. For example, when the reachable domain degenerates to a small set (as in Figure 10 with $\zeta_Y$ equal to 50% and $t_k$ equal to 50 years), the number of relevant alpha vectors drops. Likewise, at initial time $t_1$ there is just one relevant alpha vector, $\alpha_1$, related to the identified conditional plan. In any case, because of inequality in Eq.18, the policy identified by the method is guaranteed to give a value bounded from above by the approximate value $h_1^T b_1$.

Point-based value iteration methods specifically investigate reachable beliefs from an initial specific belief. To produce the graph in Figure 10(b), for the sake of illustration, we have used a set of possible initial conditions, so to represent the policy in the entire belief domain.

7. Conclusions
We have illustrated how sequential decision making under climate change can be optimized by MDPs with known or unknown dynamic models. For small state dimension, solution for MDP with known model is computationally simple to get. 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 limit cases perfect information at some step, for which closed-form are optimal. For the intermediate case of noisy observations, numerical schemes for identifying the optimal closed-loop policy, adapted from those for solving POMDPs, are illustrated in Section 6. That is an approximate method but, despite its complexity is much higher than that of extreme-case settings, they can be implemented effectively even for large class of candidate models.

Overall, we have shown how the assumption in the availability of future information about the climate model can play a key role in decision making under uncertainty. For each specific assumption, the methods outlined above allow for a consistent solution.

We conclude the analysis mentioning three issues. First, when dealing with management processing with a distant time-horizon, it may be hard to provide a complete list of possible candidate models, as new models will be probably considered by climate scientists only in the future, and we may argue those models have to be included in the current set, to assume that right one is. To address the issue, here we can only recommend the agent to include a rich set of models, able to approximately cover the entire current epistemic uncertainty.

Second, while we outline methods for specific assumed future learning rate, an open question remains about taking decision under an uncertain rate. The discussion in Section 3.5 and 5.3 may suggest that a pessimistic assumption on the rate has better guarantees, as the penalty for over-optimism tends to overpass that for over-optimism. However, we think that, to avoid wasting resources due to over-pessimism, we should identify an approximate learning scenario, and make us of it for effective planning. In the Bayesian framework, the rational path toward planning under uncertain learning rate is hierarchical modeling (where the rate may be an uncertain parameter to be learnt in time), that poses specific computational challenges in decision making (Memarzadeh et al., 2016c.

Lastly, it going without says that climate change models are not “naturally” given, but they are themselves, at least in part, the result of a decision making process, involving energy use and emission policy. Similarly, the learning rate is affected by decisions related to investment in climate studies and, locally, in risk analyses. Despite in this paper we treat those features as given, the analysis we present can be taken as a component in a more general investigation about optimizing those decisions.

Acknowledgements
The first and third authors would like to acknowledge the support of the Carnegie Mellon University’s Metro 21 Initiative, project#3505715005755, 2015 entitled “Integrated risk assessment for extreme heat events”.

References


Appendix

A.1 Policy evaluation under opposite extremes in the learning assumptions

In this section we provide details on how to evaluate sub-optimal policies, for computing Eq.10. While this may be challenging in specific setting, it turns out to be simple under the extreme condition about available information. Let us start considering perfect information is available at next step. An agent adopting the open-loop policy will update her belief, get perfect information at next step and follow the optimal single-model policy after that (as it coincides with the open-loop one, without model uncertainty). Therefore we conclude that:

$$W_{k,k+1}^\phi(i, b) = \mathbb{E}_m \{ Q_{k,m}[i, \phi_{k,\infty}^*(i, b)] \} \quad (A1)$$

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

$$W_{k,\infty}^\phi(i, b) = \mathbb{E}_m C_{k,m}[i, \phi_{k,k+1}^*(i, b)] + \gamma \sum_j |S| \mathbb{E}_m \{ T_{k,m}[i, \phi_{k,k+1}^*(i, b), j] \} V_{k+1,m}^\phi(j) \quad (A2)$$

A.2 Observation modeling for the application in Section 5

To model the relation between model and observation in the application presented in Section 5, we define measure $y_k$, at year $k$, are log-normally distributed on a continuous domain, as:

$$y_k | m \sim \mathcal{N}[\ln X_{k,m}(z_0), \zeta^2_Y] \quad (A3)$$

where $\ln \mathcal{N}(\lambda, \zeta^2)$ is the lognormal density with scale parameter $\lambda$ and scale parameter $\zeta$. So the median observation is the flood probability for a non-elevated house, $X_{k,m}(z_0)$, and $\zeta_Y$ approximates, at least when it is small, the coefficient of variation. Conditional to the model, observations are independent. We discretize the observation domain in $|Y| = 30$ possible values, by integrating the density in 30 contiguous intervals of equal length in range $[X_{k,1}(z_0)/3; X_{k,3}(z_0) \cdot 3]$. We derive matrix $O_k(m, h)$ by normalizing the probabilities, so that each model is related to an emission vector of unit sum across all possible observations.

A.3 Forward Monte Carlo sampling

To simulate reachable beliefs, as needed for the procedure in Section 6.2, illustrated in Figure 11, we make use of forward Monte Carlo simulation. To simulate beliefs from an initial one, we can simulate one model, a sequence of observations from it, and process this sequence. However, in the simulations, we do not need model consistency along time, and we can sample new beliefs, independently, for each assigned one. Let us consider belief $b$ at time $t_k$, we can sample observation $Y_{k+1} = h'$ from distribution vector $e_k$, whose component $h$ is $e_k(h, b)$, as defined in Eq.13. Updated belief derives, again, from Eq.13:

$$b' \sim e_k(b) \rightarrow u_{k+1}[h', b] \quad (A4)$$

A.4 Alpha vectors for the application in Section 5

In the application illustrated in Section 5, because of the assumption of deterministic transition depending on the action, Eq.20 is simplified to:
\[ V_k^*(i, b) = \min_a \{ EC_k(i, b, a) + \gamma \sum_{h=1}^{||Y||} e_k(h, b) \Phi \} \]

and Eq.19 to:

\[ \alpha_{m,p_k} = C_{k,m}(i, a[p_k(i)]) + \gamma \sum_{h=1}^{||Y||} O_k(m, h) \alpha_{m,p_{k+1}(h)} \]

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

\[
\begin{cases}
  i > 1: & \alpha_{m,k,i,i} = C_{k,m}(i, i) + \gamma \alpha_{m,k+1,i,i} \\
  a > 1: & \alpha_{m,k,1,a} = C_{k,m}(1, a) + \gamma \alpha_{m,k+1,a,a}
\end{cases}
\]

Finally, from state 1 and action 1, Eq.A6 reads:

\[ \alpha_{m,p_k} = C_{k,m}(1,1) + \gamma \sum_{h=1}^{||Y||} O_k(m, h) \alpha_{m,p_{k+1}(h)} \]