Safety, Reliability, Risk, Resilience and Sustainability of Structures and Infrastructure
12th Int. Conf. on Structural Safety and Reliability, Vienna, Austria, 6–10 August 2017
Christian Bucher, Bruce R. Ellingwood, Dan M. Frangopol (Editors)

A sequential decision making prospective on resilience

Matteo Pozzi[a], Milad Memarzadeh[b]

[a]Department of Civil and Environmental Engineering, Carnegie Mellon University
[b]Department of Environmental Science Policy and Management, University of California Berkeley

Abstract: We investigate how sequential decision making analysis can be used for modeling system resilience. In the aftermath of an extreme event, agents involved in the emergency management aim at an optimal recovery process, trading off the loss due to lack of system functionality with the investment needed for a fast recovery. This process can be formulated as a sequential decision-making optimization problem, where the overall loss has to be minimized by adopting an appropriate policy, and dynamic programming applied to Markov Decision Processes (MDPs) provides a rational and computationally feasible framework for a quantitative analysis. The paper investigates how trends of post-event loss and recovery can be understood in light of the sequential decision making framework. Specifically, it is well known that system's functionality is often taken to a level different from that before the event: this can be the result of budget constraints and/or economic opportunity, and the framework has the potential of integrating these considerations. But we focus on the specific case of an agent learning something new about the process, and reacting by updating the target functionality level of the system. We illustrate how this can happen in a simplified setting, by using Hidden-Model MPDs (HM-MDPs) for modelling the management of a set of components under model uncertainty. When an extreme event occurs, the agent updates the hazard model and, consequently, her response and long-term planning.

1 Introduction: resilience of controlled systems

Resilience is a key aspect in the behavior of dynamic systems, and it indicates the system's capability of recovering after disruptions as those induced by extreme events [1-2]. For civil infrastructure systems, however, the recovery process is the outcome of a complicate interplay between the physical changes, due to degradation and damages induced by the extreme event, and the decisions of the infrastructure managers, that we call "agents" in the following. The recovery process can be fast or slow depending on the priorities and decision attitude of the agents. Generally, their aim is not to restore the system' functionality as soon as possible, "at any cost": it is to find an optimal trade-off between the cost for restoration and that for the loss of functionality. We posit that the traditional metric of "long-term expected utility" or "expected cost" in rational decision making can measure resilience.

In this paper, we illustrate how resilience emerges from optimal system management, and how we can explain it in terms of optimal behavior. In Section 2 we define the basis for sequential decision making, in Section 3 we connect this to resilience analysis, in Section 4 we describe in details an application, before drawing conclusions in 5.

2 Sequential decision making

In sequential decision making, an agent acts in time, for achieving a long-term goal. As the effects of her actions cannot be decoupled, she has to identify not just an optimal current action, but a whole long-term policy. That policy strongly depends on the assumed system behaviour, which is stochastic when the agent can provide only uncertain prediction of the response to her actions, and on the reward or losses that she receives.

2.1 Markov Decision Processes

MDPs and partially observables MDPs have been extensively applied to civil systems (e.g, by [4-5]). In a Markov Decision Process (MPD), an agent interacts with a system, observing completely its current state, selecting actions and getting a loss, or reward [3]. Time is discretized into a set of instants, evenly separated by period Δt, so that discrete variables s_k and a_k indicates the state and action at time t_k, defined in domain $S = \{1,2,\dots,|S|\}$ and $A = \{1,2,\dots,|A|\}$ respectively. The immediate cost function $C(s,a)$ depends on current state s and action a, while transition probability $T(s,a,s') = \mathbb{P}[s_{k+1} = s' | s_k = s, a_k = a]$ models the Markovian dynamic of the controlled system, $\mathbb{P}[\mathrm{E}'|\mathrm{E}]$ being the conditional probability of event E' given event E. The agent selects actions relying on direct observation of the system state, with the aim of minimizing value V, i.e. the infinite-horizon expected discounted sum of costs that, at time t_0, is $V = \sum_{k=0}^{\infty} \gamma^k c_k$, where $c_k = C(s_k, a_k)$. As the system state is a sufficient statistic, the agent can base her policy to the observation of the current state. Optimal policy π^* and corresponding value V^* derives from Bellman's equation:

$$\begin{cases} V^*(s) = & \min_{a\in A}\left[C(s,a) + \gamma \sum_{s'=1}^{|S|} T(s,a,s') V^*(s')\right] \\ \pi^*(s) = \operatorname{argmin}_{a\in A} & \left[C(s,a) + \gamma \sum_{s'=1}^{|S|} T(s,a,s') V^*(s')\right] \end{cases} \qquad (1a,b)$$

where γ is the one-step discount factor. Optimal value and policy are stationary, because reward and transition matrix are so, and the agent is planning for an infinite time horizon. While recursive Eq.1(a) cannot be easily solved directly, due to its non-linearity, the value V_π following policy π can be express in a linear recursive form and in matrix-vector form as:

$$V_\pi(s) = C[s,\pi(s)] + \gamma \sum_{s'=1}^{|S|} T[s,\pi(s),s'] V_\pi(s') \qquad [\mathbf{I} - \gamma \mathbf{T}_\pi]\mathbf{v}_\pi = \mathbf{r}_\pi \qquad (2a,b)$$

where entry i of $[|S| \times 1]$ vectors $\mathbf{v}_\pi$ and $\mathbf{r}_\pi$ are $V_\pi(i)$ and $C(i,\pi(i))$ respectively, entry (i,j) of $[|S| \times |S|]$ matrix $\mathbf{T}_\pi$ is $T(i,\pi(i),j)$, and $\mathbf{I}$ is the identity matrix. We solve Eq.2 exactly or approximately. In the "policy iteration" method [3], we (i) arbitrary initialize the policy,

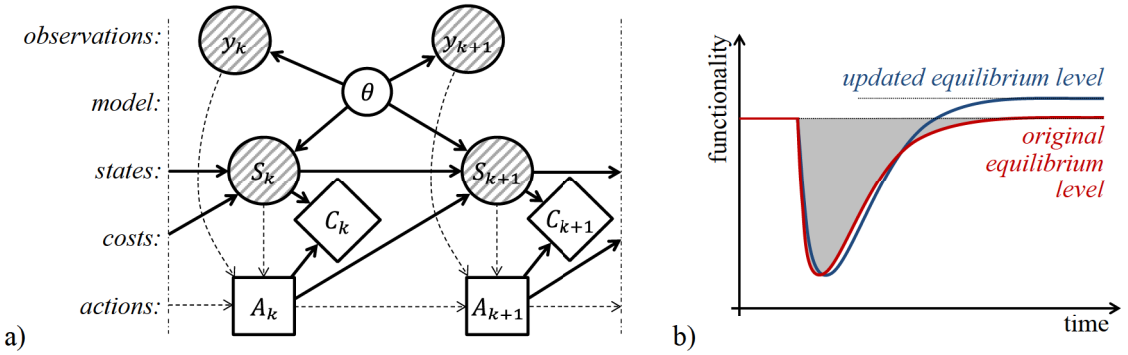

Figure 1: Graph of a HM-MDP (a), typical diagram of the system disruption in time (b).

(ii) evaluate it through Eq.2, (iii) improve it applying Eq.1(b) (using V_π instead of V^* in the right hand side), and getting a new policy, and iterate steps (ii-iii) until policy convergences.

At the end of the analysis, we get optimal policy π^*, which prescribes an action for each observed current state. Following this policy, the agent gets expected value V^*.

2.2 Bayesian model updating and MDP under model uncertainty

In previous section, we have showed how to identify the optimal policy when the dynamic model (i.e. transition function T) is known. Here we discuss Hidden-Model MDP (HM-MDP) [6]: the extended case when the model is uncertain. Model can be defined by parameter θ, that is persistent in time, but unknown to the agent. Her belief on the model is capture by probability density p_θ, so that transition T is now also a function of θ. In this setting, one task is learning and updating the probability of θ during the process: by using Bayes' rule, posterior probability at time t_k is given by $p_{\theta|Y_k} \propto p_{Y_k|\theta} p_\theta$, where Y_k includes the set of relevant information collected up to that time, including the observed trajectory of system's states. Figure 1(a) shows the influence diagram of an HM-MDP, including states, observations, costs, actions, and the model. Shaded nodes are observable in due time, and arrows describe the conditional independence structure.

The other task is to control the system under model uncertainty: depending on the assumptions and the required accuracy, a wide range of methods is available [6]. One simple method is to derive an average model, by computing the expected transition function as $T_k(i,a,j) = \mathbb{E}_{\theta|Y_k}[T(i,a,j,\theta)]$, where $\mathbb{E}_{x|z}[f(x)]$ is the expected value of f according to probability $p(x|z)$. The average model can be used in the policy iteration method, to define an optimal policy π_k^*: the agent can adopt that policy until new information comes, then she can update the distribution again, compute a new average model and a corresponding new policy. The method is inexact for two reasons. First, it does not predict that any additional information will be available in the future, and it is hence an "open-loop" method. Second, the method does not account for model persistency: policy π_k^* was optimal ended under the incorrect assumption that a new value of θ is independently generated from $p_{\theta|Y_k}$ at each future step. Nonetheless, we adopt this approach as a approximate method.

3 Describing resilience using MDPs

Figure 1(b) illustrates a typical recovery process for a system damaged by an extreme event: its functionality drops due the direct impact, and the recovery depends on the repair and maintenance policy. Resilience can be (inversely) related the shaded area in graph (b) [3]. Sometimes, the agent decides to restore the functionality to the level before the event, however she can also decide otherwise, and take the system to a new updated equilibrium. This can happen for many reasons: because of budget constraints, because of the availability of new technologies, or because of an updated system demand, or because of the "opportunity" for replacing components that the damage provides (while it was not convenient to replace the functioning components before the event).

All those features can be included in a sequential decision making framework. However, here we focus our attention to a specific reason for updating the equilibrium level: by facing the event, the agent can learn something about the hazard and/or the system vulnerability, update her knowledge and, consequently, her policy, selecting a new level. In the following Section we present in details a setting where this happens.

4 Modelling system maintenance under extreme events

To investigate how we can intend resilience from a sequential decision making prospective, we consider an agent controlling an infrastructure system before, during and after an extreme event. This agent has a prior model of the occurrence rate for these events, she has to plan in advance and can update her policy in face of new observations of these events.

4.1 Problem statement

We consider a system made up by similar components, exposed to extreme events. The system supplies a service to society, to meet its demand. We discretize time in weeks (so Δt is 7 days). Weekly demand D is unknown, and the agent models it as a set of independent random variables lognormally distributed by p_D. System state defines the number of functioning components, so that there are n_k components at time t_k. Components deteriorate, and they are prone to failure when extreme events occur. The change Δn_k in the number of functioning components from t_k to t_{k+1} is given by three contributions:

$$\Delta n_k = n_{k+1} - n_k = \Delta n_k^{(a)} - \Delta n_k^{(d)} - \Delta n_k^{(ee)} \tag{3}$$

where $\Delta n_k^{(a)}$ is the decision variable and defines the number of components to be added, $\Delta n_k^{(d)}$ of those damaged by deterioration and $\Delta n_k^{(ee)}$ of those damaged by extreme events. Binary variable e_k defines the occurrence of an extreme event in the previous time interval, and is Bernoulli distributed with rate θ, in turn modeled as $\theta \sim \text{Beta}(\alpha_\theta, \beta_\theta)$. If an event occurs, the functioning components fail independently with uncertain probability P_e, that defines the event intensity and follows distribution f_e. Hence, $\Delta n_k^{(ee)}$ is binomially distributed, while the survived components fail independently with probability P_d due to deterioration and $\Delta n_k^{(d)}$ is also binomially distributed:

$$\begin{cases} e_k \sim \text{Bernoulli}(\theta) \\ P_{e_k} \sim f_e \end{cases} ; \begin{cases} \Delta n_k^{(ee)} | [e_k = 0] = 0 \\ \Delta n_k^{(ee)} | [e_k = 1] \sim \text{Binomial}\left(n_k, P_{e_k}\right) \end{cases} ; \Delta n_k^{(d)} \sim \text{Binomial}\left(n_k - \Delta n_k^{(ee)}, P_d\right) \tag{4a,b,c}$$

Overall cost C is the sum of two contributions: C_R and C_F. Repairing cost C_R, for adding $\Delta n^{(a)}$ components is the sum of constant term c_0, if at least one component is added, and of a non linear function of the number of components added, via coefficients c_r and ν:

$$C_R\left(\Delta n^{(a)}\right) = c_0 \, I\left[\Delta n^{(a)} > 0\right] + c_r \left[\Delta n^{(a)}\right]^\nu \tag{5}$$

Expected cost for insufficient components C_F is a function of the lacking components $\Delta n_{\text{lack}} = D - n$, via parameters c_{pen} and η: function g defines this cost for one value of demand D, and C_F is its expected value:

$$g(\Delta n_{\text{lack}}) = \begin{cases} c_{\text{pen}} \left[\Delta n_{\text{lack}}\right]^\eta & \Delta n_{\text{lack}} > 0 \\ 0 & \Delta n_{\text{lack}} \le 0 \end{cases} \quad C_F(n) = \mathbb{E}_D g = \mathbb{E}_D \max\left\{c_{\text{pen}} \left[D - n\right]^\eta, 0\right\} \tag{6a,b}$$

4.2 Formulation as a MDP and a HM-MDP

To describe the problem outlined in Section 4.1 in the framework of Section 2, we define $S_k = n_k + 1$, so that the first state refers to zero active components, and $(|S| - 1)$ is the maximum number of components. Action defines the number of added components as $a_k = \Delta n_k^{(a)} + 1$, so that $a_k = 1$ if no components are added at time t_k.

We derive from Eqs.3-4 the transition matrix with no added component, and similarly those for any specified number of added components, while the complete cost function derived from Eqs.5-6.

Because of prior conjugacy, if Y_k is the observed number of extreme events in the $[t_0,t_k]$ period, the posterior probability of rate θ is:

$$\theta|Y_k \sim \text{Beta}(\alpha_\theta + Y_k, \beta_\theta + k - Y_k) \tag{7}$$

4.3 Numerical investigation

We select parameters as follows. The maximum number of components N is equal to 100. The weekly demand of functioning components is defined by probability $p_D \propto \ln\mathcal{N}(\lambda_D, \zeta_D^2)$, with $\lambda_D = \log 40$ and $\zeta_D = 10\%$, where $\ln\mathcal{N}$ is the log-normal distribution. The prior information on extreme event rate is so that the expected value and the coefficient of variation of θ are 0.2% w^{-1} (corresponding to a return period of 10 years) and 70%, respectively, so that $\alpha_\theta = 2$, $\beta_\theta = 1{,}000$. P_e is beta-distributed: $f_e = \text{Beta}(\alpha_e, \beta_e)$, with the expected value and the coefficient of variation of P_e both equal to 50%, so that $\alpha_e = \beta_e = 1.5$. P_d is 0.2%, so that the expected annual number of degraded components is 10 when $n = N$. Costs are defined by $c_0 = 4\text{K\$}$, $c_r = 10\text{K\$}$, $\nu = 2$, $c_{\text{pen}} = 1\text{K\$}$, $\eta = 2$. Weekly discount factor is $\gamma = 99.9\%$, corresponding to an annual factor of 95%.

Figure 2 summarizes inputs and outputs for the application: graph (a) reports the distribution of weekly demand D, graph (b) the corresponding expected cost C_F, obtained by Eq.6, graphs (c-d) the optimal policy and value (i.e. expected discounted cost), as derived from Eq.1. All graphs are plotted up to $n = N_M = 70$. The agent will add components below $n_{\text{opt}} = 56$, while she will prefer not to when $n \geq n_{\text{opt}}$. The less the number of components, the higher C_F, the number of components to be added, and the value. The value is 4.34M\$ when $n = N_M$ and increases of additional 6.39M\$ is the system is completely destroyed (i.e., when n is zero).

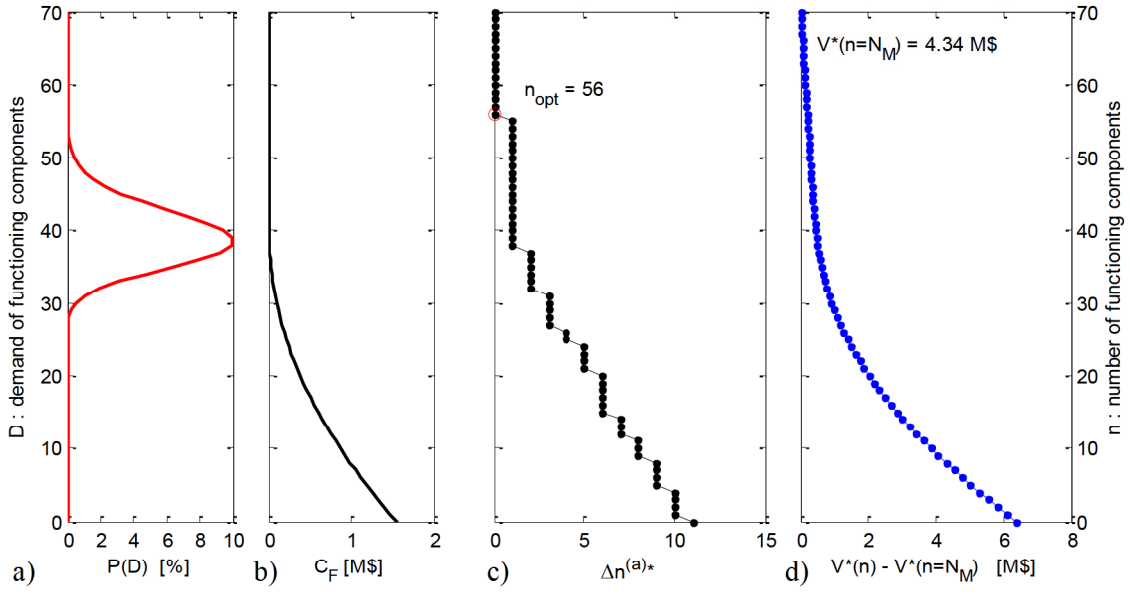

Figure 2: Demand distribution (a), expected cost for lacking components (b), optimal policy: number of components to be added (c), expected discounted cumulative cost (d).

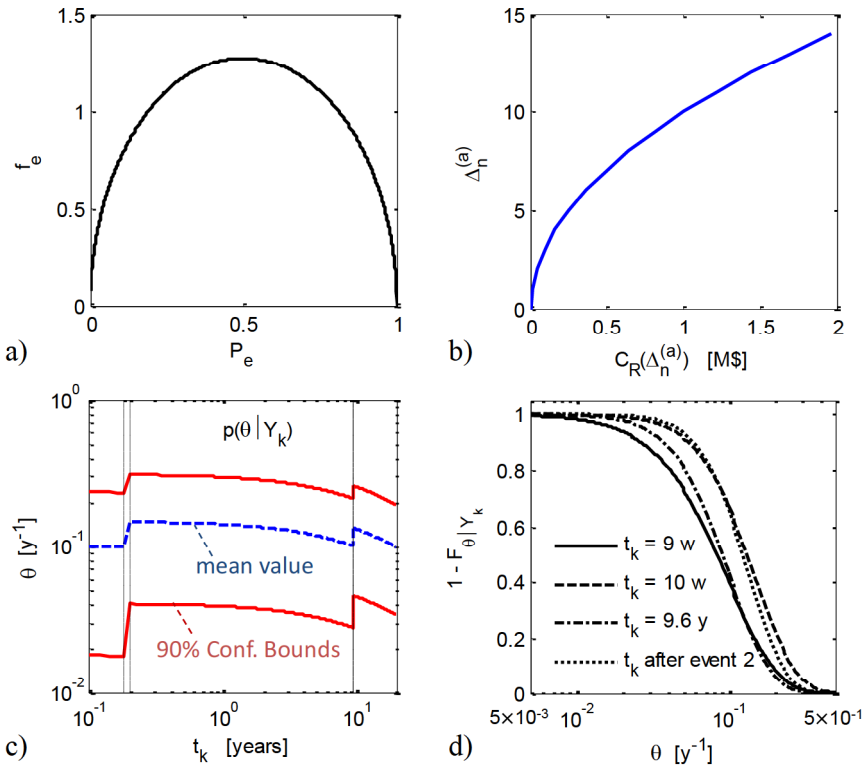

Figure 3: probability distribution of intensity (a), cost function for repairing (b), posterior distribution of extreme event rate (c) and corresponding complement CDF at given times (d).

Figure 3(a) and (b) reports the probability of intensity P_e and the cost of repairing C_R as a function of the number of added components. We assume that two extreme events occurs during the management process: at t_{10} and t_{480}, i.e. after 2 months and after about 10 years from the beginning. The corresponding posterior belief on extreme event rate θ, obtained through Eq.7, is represented in graph (c) and (d): specifically, (c) shows the 20-year time domain, and (d) focuses on 4 times, before and after the extreme events. As anticipated in Section 4.2, the agent belief shifts towards higher values of θ after an extreme event.

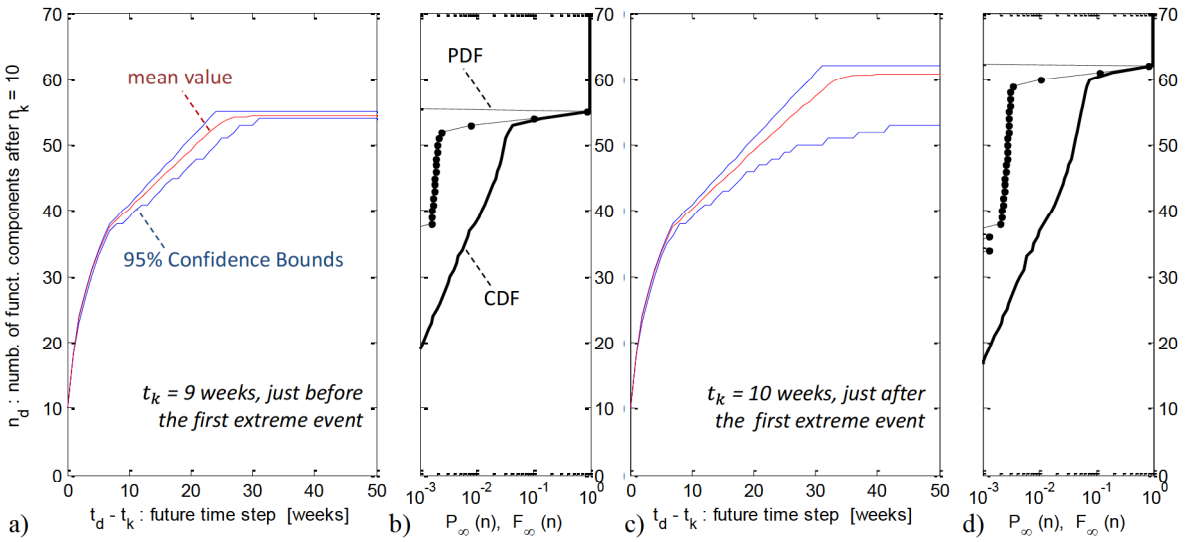

Figure 4: recovery process using the policy optimal before the first event (a), corresponding limit distribution (b), corresponding quantities for the policy after the first extreme event (c-d).

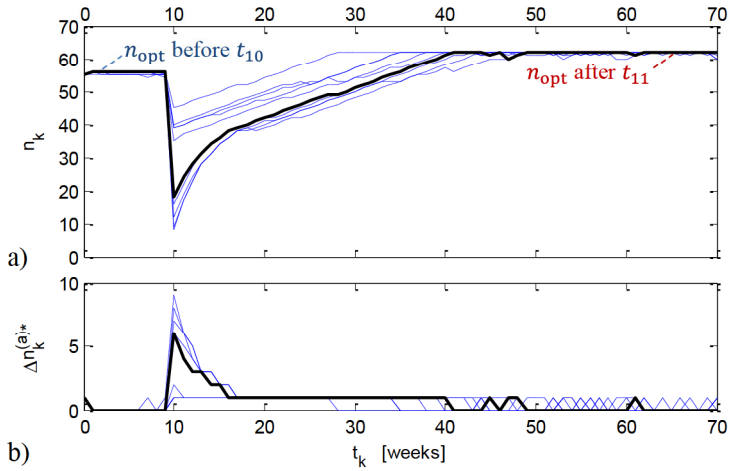

Figure 5: foward simulations of system maintenence (a), corresponding controlling actions (b).

To illustrate the recovery process at different times, we can predict the evolution of the controlled system after an extreme event. Let us assume that, at time t_k, the agent identifies optimal policy π_k^*, and adopts it from that time on. Because of randomness in the system evolution, we cannot predict it with certainty. However, having defined $T_{\pi_k^*}(i,j) = T_k(i,\pi_k^*(i),j)$ as the transition function when π_k^* is in control, $\mathbf{p}_d = \mathbf{T}_{\pi_k^*}^{d-k}\mathbf{p}_k$ defines the evolution of the probability distribution of the system state, where vector $\mathbf{p}_k$ lists probability $\mathbb{P}[n_k = i]$ for i from 0 to N. Moreover, by using a large value for d, we derive $\mathbf{p}_\infty$, the asymptotic limit distribution for the Markov process. Figure 4 shows the evolution of the controlled system after an event that takes the number of components down to 10. To do so, we initialize $\mathbf{p}_k$ to a vector of zeros except for a single 1 in position 10+1, then we can compute $\mathbf{p}_d$ for any $d > k$. Graph (a) makes use of π_9^*, the optimal policy before the first extreme event, practically identical to that at the beginning of the process, plotted in Figure 2(c). The graph reports the expected value and the 95% confidence bounds, while Figure 4(b) plots the PDF and CDF of the asymptotic limit distribution. The system never goes above n_{opt}, because the policy never prescribes to venture in that region. The randomness in the evolution is due to degradation and occurrence of extreme events.

Because of the change in the agent belief, shown in Figure 3(c-d), the assumed expected transition probability changes, and so does the optimal policy. Specifically, n_{opt} grows up to 62 components after the first extreme event, and goes down to 55 just before the second event, after which it grows again to 60. Graph (c-d) refers to time t_{10}, after the occurrence of the first extreme event. Both the short-time evolution and the limit distribution show that the agent will adopt a higher number of components, for counter-balancing the increased estimated frequency of events.

Figure 5 plots 100 time-domain forward simulations (one with a thicker line): we sample variables e_k, P_{e_k}, $\Delta n_k^{(ee)}$, $\Delta n_k^{(d)}$ for each time t_k, and we close the loop by the optimal policy. Up to $k = 9$, no extreme events have occurred and the agent follows the policy plotted in Figure 2(c) with minor adjustment due to the observation of no events. However, at t_{10}, an event occurs (i.e., $e_{10} = 1$), and the optimal policy significantly changes. Before that, the system stays at $n_{\text{opt}} = 56$, and the agent repairs all failed components. However, as n_{opt} grows up to 62 after the event, the new equilibrium point is higher than the pre-event one.

5 Conclusions

Our aim in this paper was to illustrate how to model the maintenance and control of a infrastructure system exposed to extreme events by using MDPs and HM-MDPs. Many features relevant for resilience analysis can be derived by this framework. For example, graphs as those in Figure 4 can be used for assessing resilience, as in Figure 1. The shaded area in the latter figure can be intended as a cumulative disruption, which the agent should keep low. However, we note that, in our example, the aim of the agent is more complicated. First, the disruption can be non-linearly related to the cost, according to Eq.6, while it is linearly related to the shaded area. Second, the agent can decide to have more components than those necessary to cover the demand, to increase redundancy and reduce the effects of extreme events. The solution also includes the effects on ordinary degradation and maintenance: the agent knows that to maintain many redundant components is expensive, as she has to cover their recurrent maintenance costs. Overall, the agent selects a recovery level for responding to the short-term demand, but also for preparing the response to the next shock. We also note how rich the outcome of the MDP analysis is: the value plotted in Figure 2(d), for example, can be used in the design phase, by integrating it with the construction cost to identify the optimal design in terms of number of components. That analysis is also able to formalize the updating process of the equilibrium level, and transient phase to this new target.

We have selected the simplified example presented above for the easiness of solving it (less than 10 rounds of the policy iteration algorithm are sufficient for finding the optimal solution). Despite this, we hope that this initial research can be followed by additional work to better relate sequential system control and resilience.

Acknowledgements

This research is based upon work supported by the National Science Foundation under Grant No. 1638327.

References

[1] I. Linkov, T. Bridges, F. Creutzig, J. Decker, C. Fox-Lent, W. Kröger, J.H. Lambert, A. Levermann, B. Montreuil, J. Nathwani and R. Nyer, “Changing the resilience paradigm.” Nature Climate Change, 4(6), pp.407-409. 2014

[2] G.P. Cimellaro, A.M. Reinhorn and M. Bruneau, “Framework for analytical quantification of disaster resilience” Engineering Structures, 32(11), pp.3639-3649. 2010.

[3] D.P. Bertsekas, Dynamic programming and optimal control. Belmont, MA: Athena Scientific, 1995.

[4] P.L. Durango and S.M. Madanat. Optimal maintenance and repair policies in infrastructure management under uncertain facility deterioration rates: an adaptive control approach.TranspResPart Λ; 36: 763-78. 2002

[5] M. Memarzadeh, M. Pozzi and J.Z. Kolter, “Optimal planning and learning in uncertain environments for the management of wind farms.” Journal of Computing in Civil Engineering, 29(5), p.04014076. 2014.

[6] M. Pozzi, M. Memarzadeh and K. Klima. "Hidden-model processes for adaptive management under uncertain climate change," to appear in ASCE’s Journal of Infrastructure Systems. 2017.