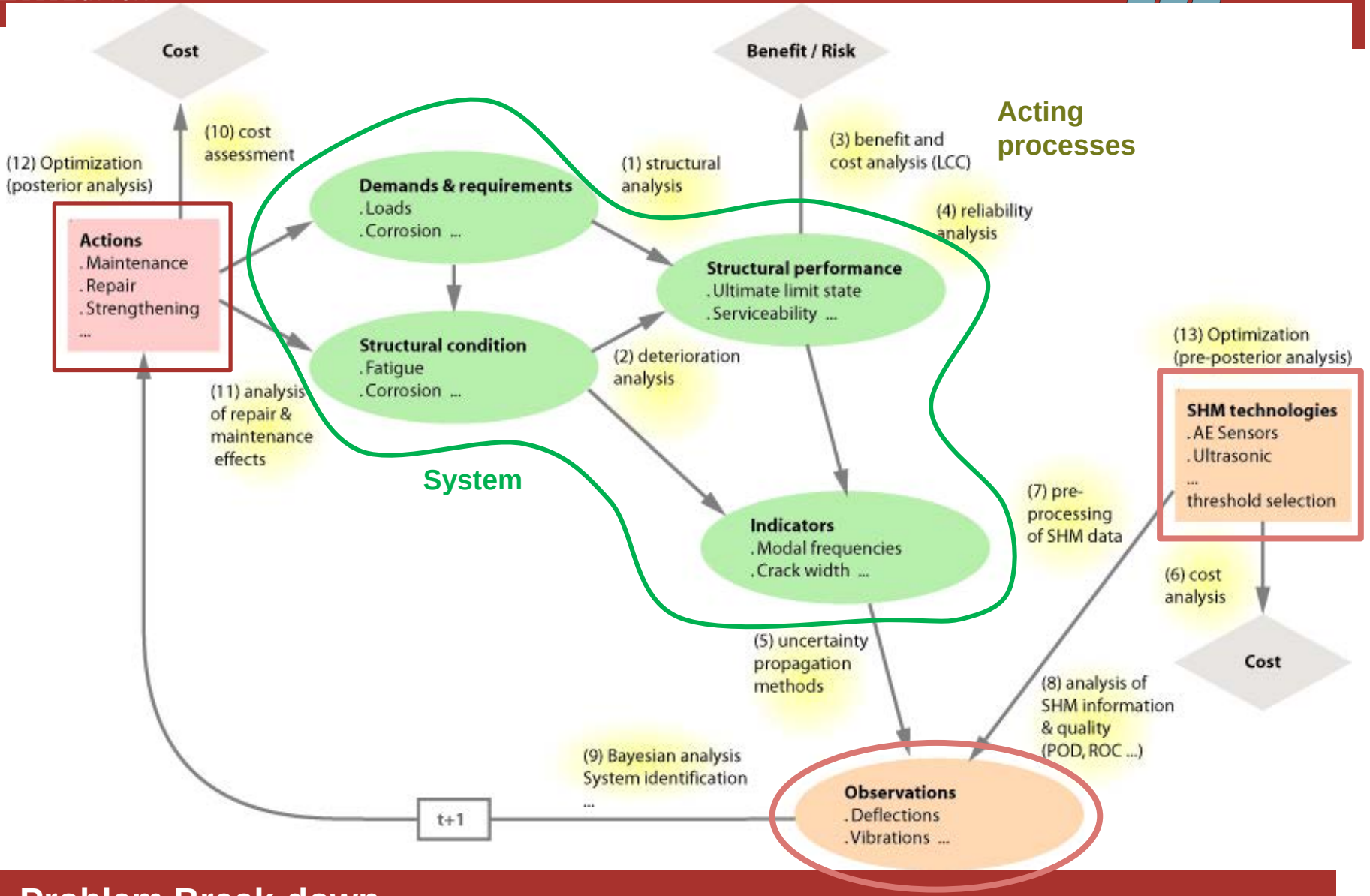
A Partially Observable Markov Decision Process approach to decision-making

Eleni Chatzi, ETH Zürich
Konstantinos Papakonstantinou, Penn State University

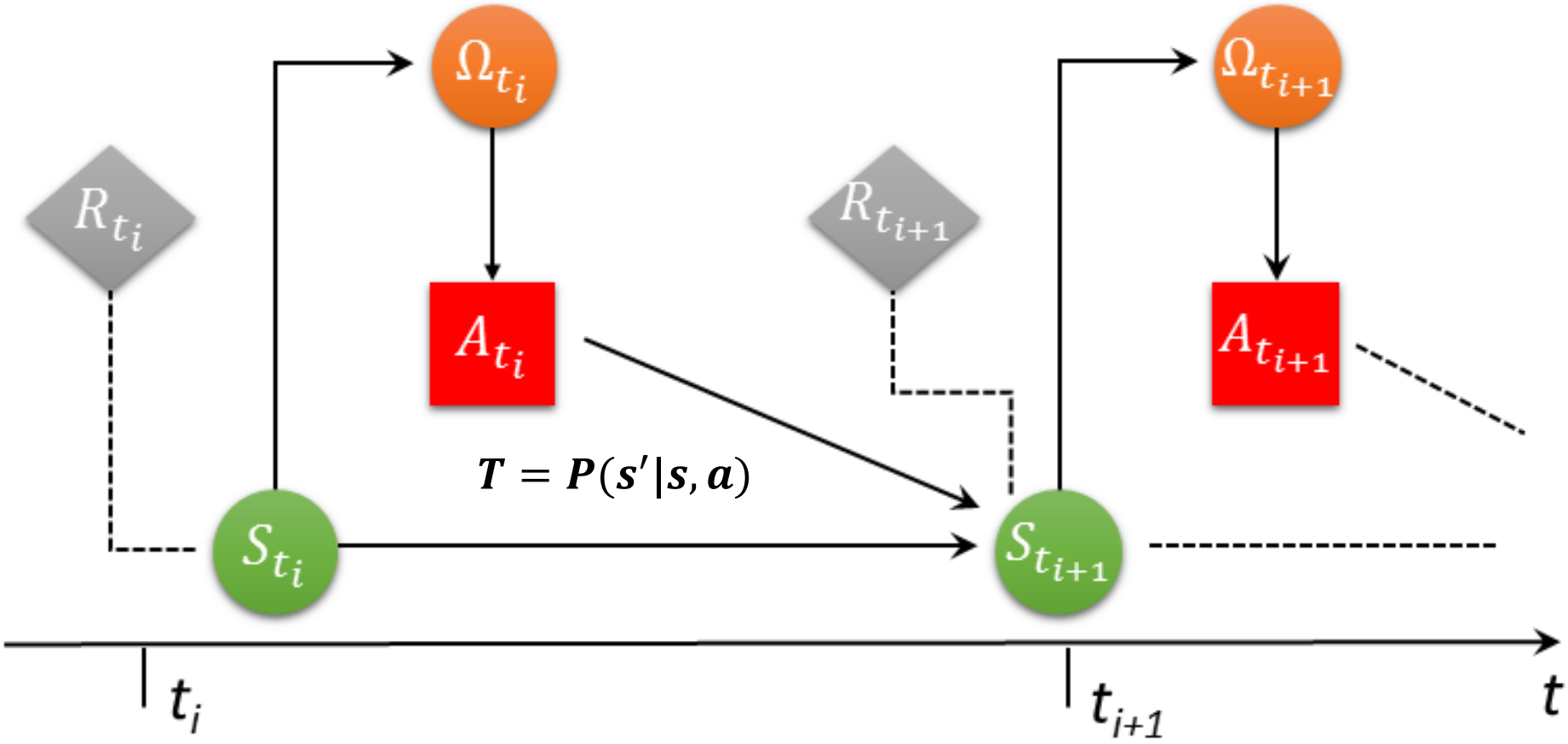
© NRC -Flickr

Optimal Inspection & Maintenance Planning



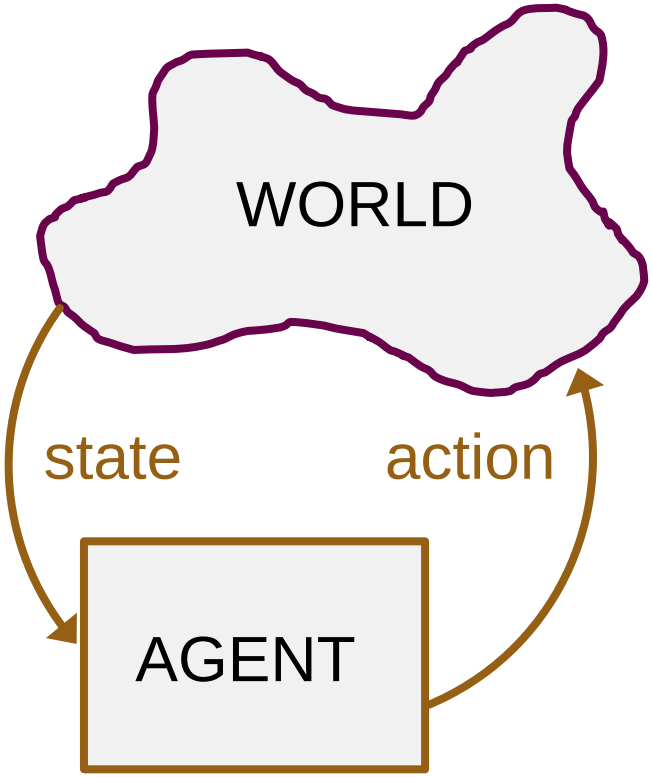
Problem Break-down

Sequential decision process with alternating actions and inspections:



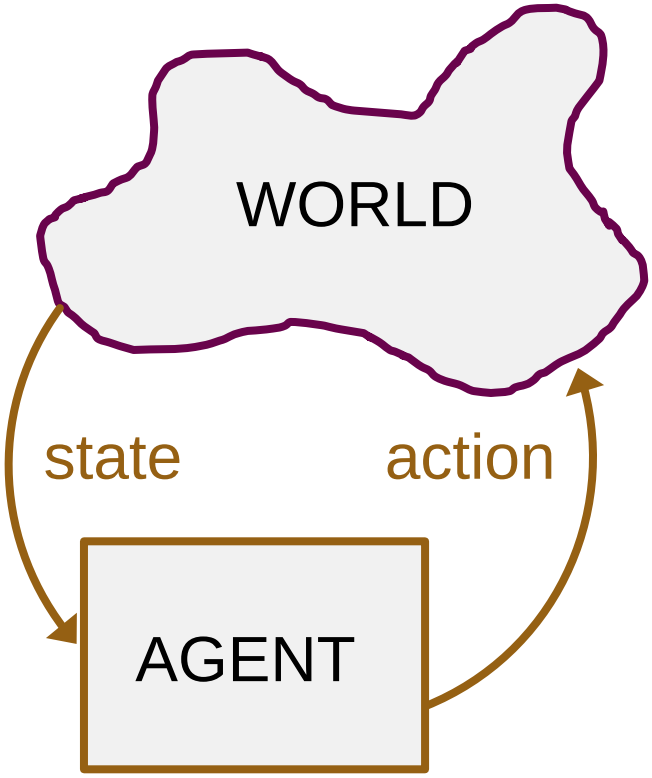
Optimal Inspection & Maintenance Planning

Fully Observable MDP

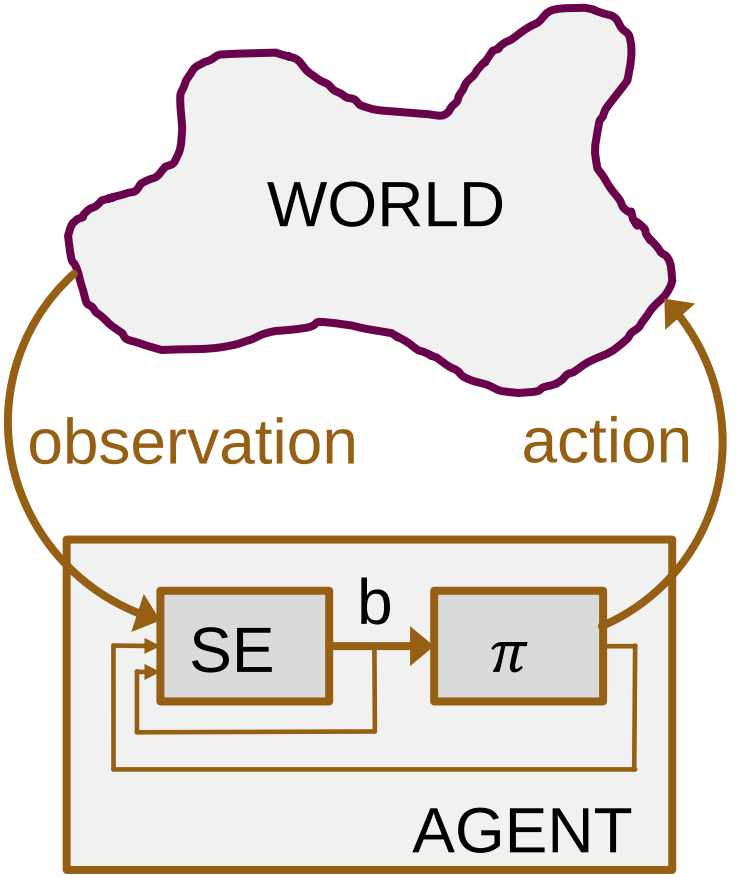


- Decision depends on current state, no history
- Initial state is known
- Action's consequences are known
- World is known
- The state is fully observable

Fully Observable MDP



Partially Observable MDP

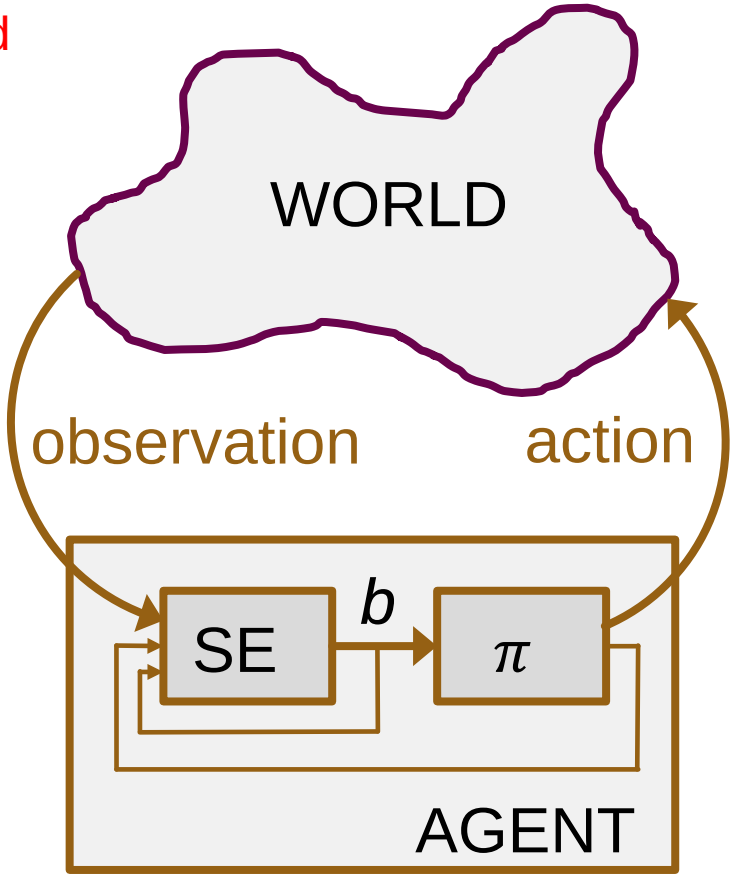


(Kaelbling, 1999)

Markov Decision Process

Partially Observable MDP ⟶ state not fully observable

- Decisions depend on current state and history
- Initial state is uncertain
- Actions are uncertain
- World is known
- Observations are uncertain
- Sequential process: action →
- observation → action . . .



SE = State Estimator
b = belief state
π = policy

(Smallwood and Sondik, 1973; Sondik, 1978)

Markov Decision Process

POMDP Applications

Industrial

Machine/structural inspection, fishing industry, quality control

Scientific

Autonomous robots, behavioral studies, machine vision

Business

Network troubleshooting, marketing, questionnaire design, corporate policy control, distributed database queries

Military

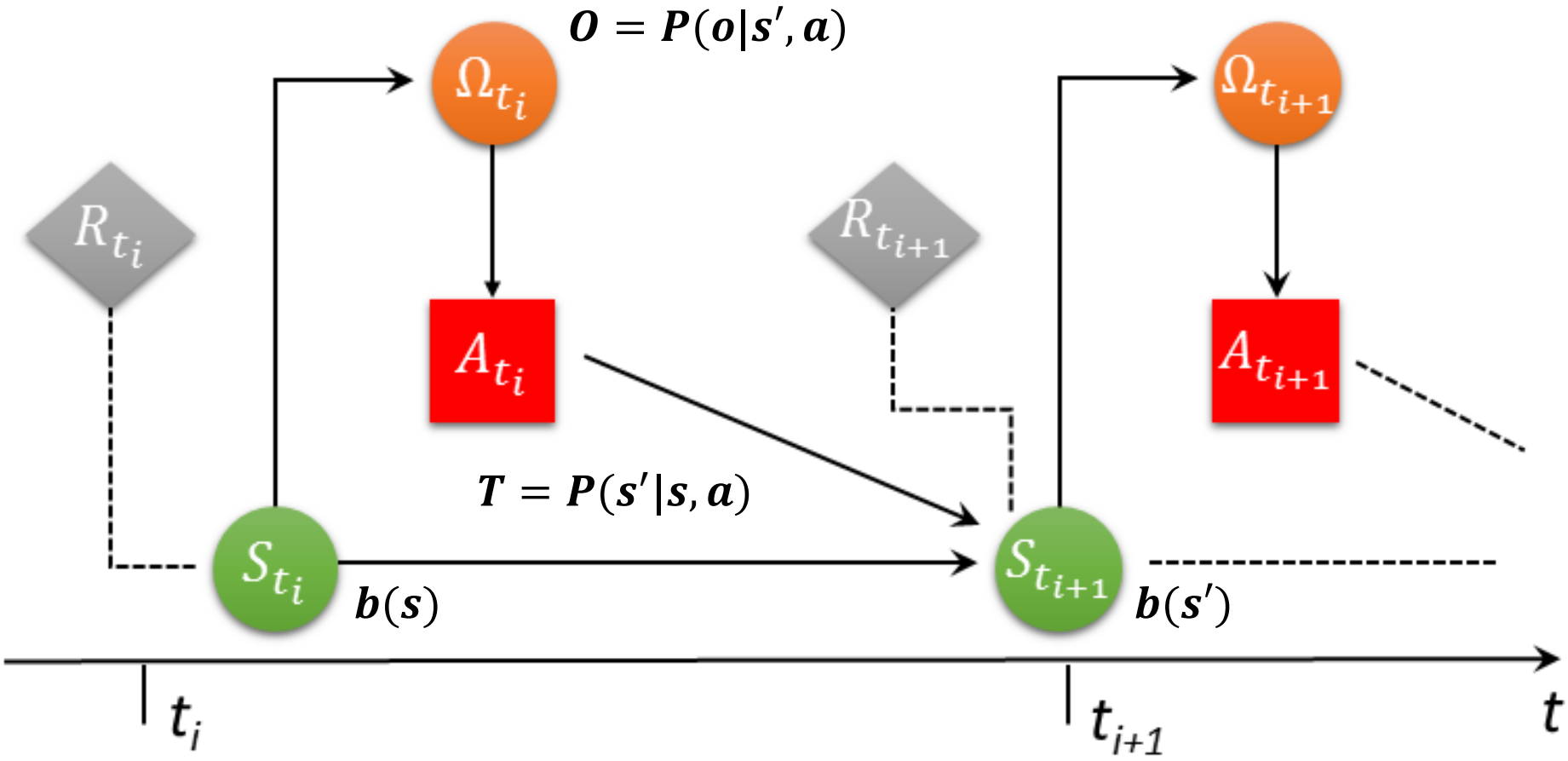
Moving target search, search and rescue, target identification, weapon allocation, finding of hidden objects

Social

Education (teaching strategies), medical diagnosis, health care policy making)

(Cassandra, 1997; Lovejoy, 1991; Monahan, 1982)

Sequential decision process with alternating actions and inspections:



The POMDP Framework

A **POMDP** framework consists of the **tuple** $\{S, A, T, \Omega, O, R\}$

S	is the set of system states
A	is the set of actions
$T: S \times A \to \Pi(S)$	is the transition model describing $p(s'\|s,a)$
Ω	is the set of discrete observations
$O: S \times A \to \Pi(\Omega)$	is the observation model describing $p(o\|s,a)$
R	is the reward function $r_a(s) \in \mathbb{R}$

The updating of a given belief state may be obtained, using Bayes' rule is (continuous states):

$$b^{a,o}(s') = \frac{p(o|s',a)}{p(o|b,a)} \int_S p\,(s'|s,a) b(s \Big)$$

Back Propagation

The **total reward** over the entire lifetime of the agent ($t = 1, \ldots, T$) is:

$$Q_{tot} = Q_{terminal} + \sum_{t=1}^{T} dQ_t \tag{2}$$

$$dQ_t = \int_s r(s,a)\, b(s) \tag{3}$$

Planning aims to maximize the expected future rewards:

$$V_n(b) = \max_a Q_n(b,a) \tag{4}$$

$$Q_n(b,a) = \int_S r_a(s)\, b(s) + \gamma \sum_o p(o|b,a)\; V_{n-1}\, b^{a,o} \tag{7}$$

where n describes the number of decision time steps left till the end of the agent's lifetime, a.k.a. **horizon**. It is then $n + t = T$.
($V_0 = Q_{\text{terminal}}$)

Continuous State POMDP

The optimal future reward is represented by a set of $a-$ functions:

$$V_n(b) = \max_{\{\alpha_{a,o}^j\}_j} \int_s \alpha_n^i(s) b(s)$$

where

$$\{\alpha_n^i(s)\}_i = r_a(s) + \gamma \sum_o \alpha_{a,o,b}(s)_{\alpha \in A}$$

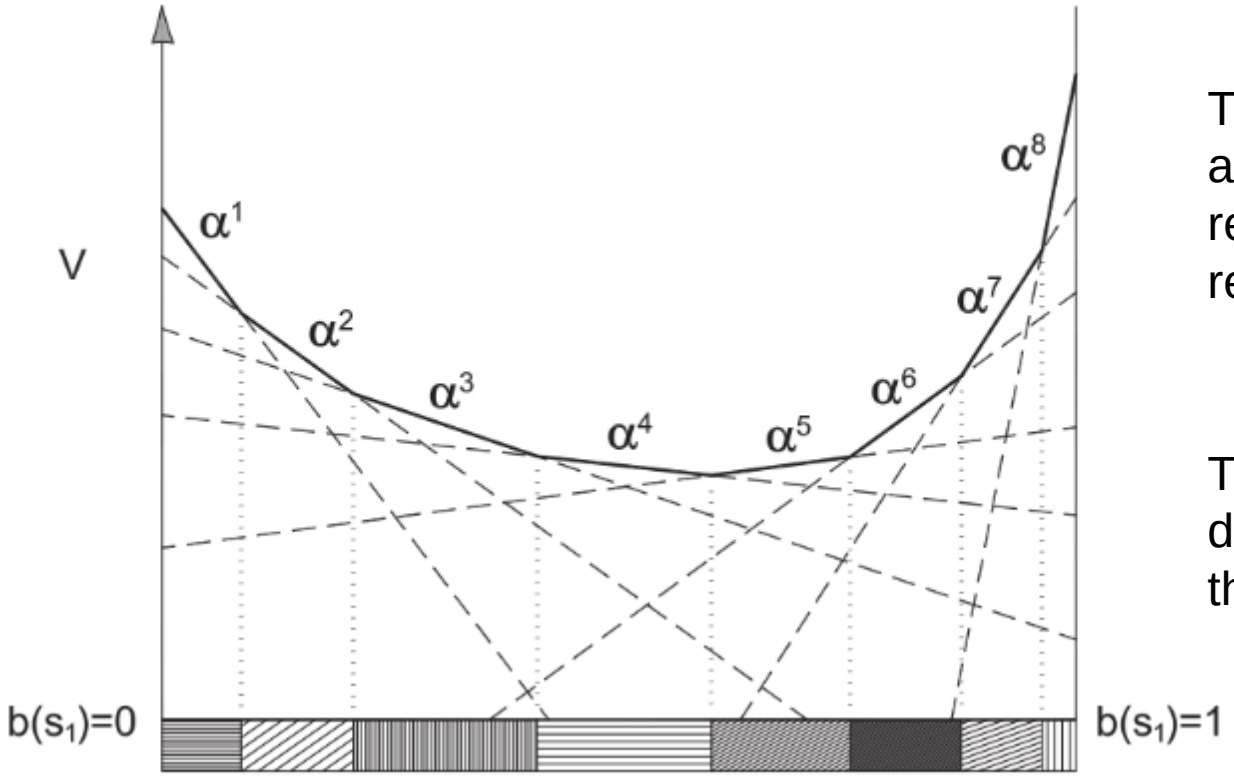
$$\alpha_{a,o,b} = \operatorname*{argmax}_{\{\alpha_{a,o}^j\}_j} \int_s \alpha_{a,o}^j(s) b(s)$$

$$\alpha_{a,o}^j = \int_{s'} \alpha_{n-1}^j(s') p(o|s') p(s'|s,a)$$

Discrete Equivalent

The optimal future reward is represented by a set of a – vectors:

$$V^*(\mathbf{b}) = \max_{\{\boldsymbol{\alpha}^i\}_i^*} \sum_{s \in S} b(s)\, \alpha^i(s),$$



Papakonstantinou & Shinozuka, 2014

The belief space is a simplex, and each vector defines a region over the simplex which represents a set of belief states.

The value function, is generally defined as the upper surface of these vectors.

POMDP Solvers

The solution of this recursive problem aims at establishing the optimal policy, i.e., planning of sequence of inspections and actions to be performed (policy).

Discrete POMDPs

- Approximations based on MDP and Q-functions
- Grid-based approximations
- Point-based value iteration methods

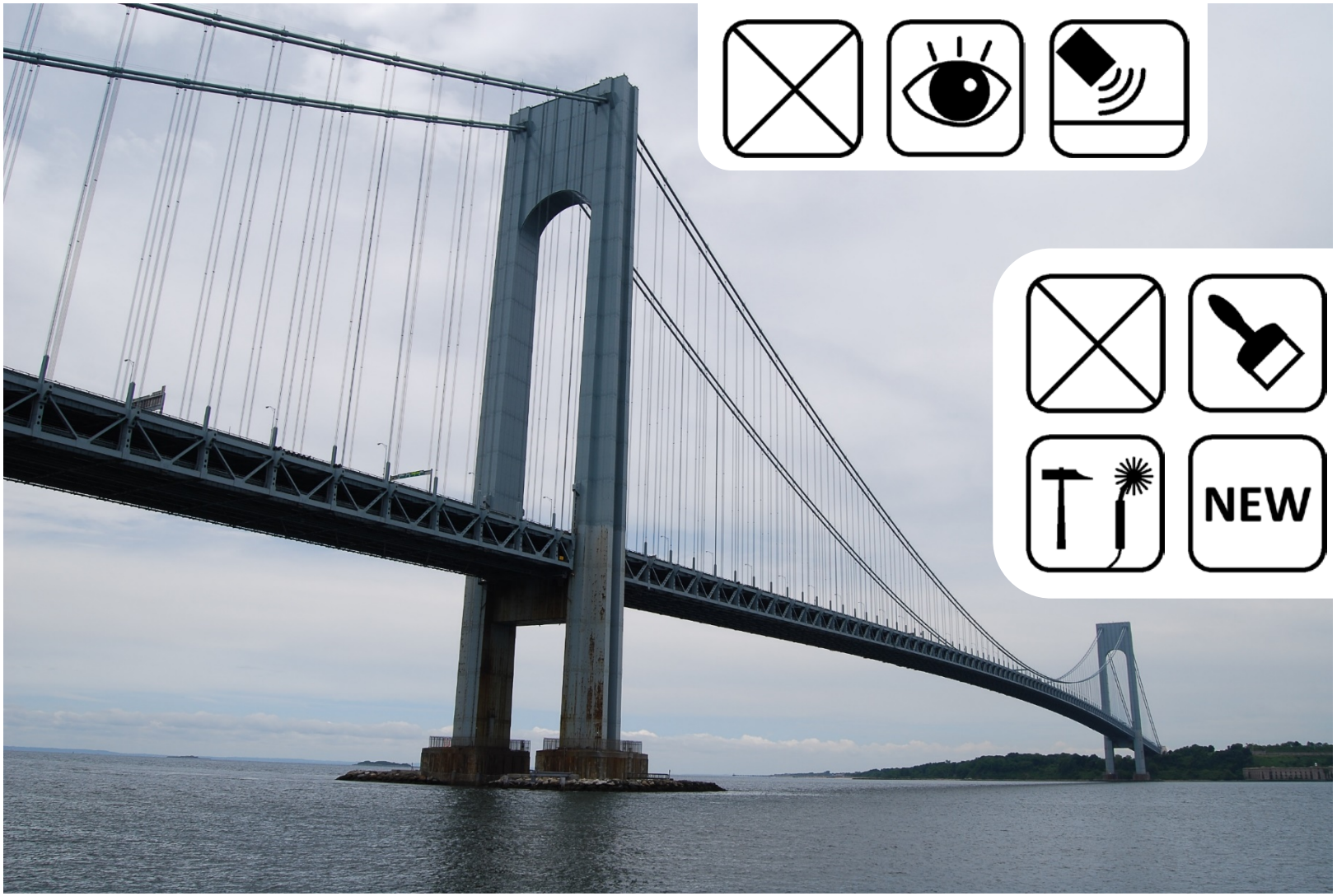
(Pineau, Gordon, & Thrun, 2003; Vlassis & Spaan 2004).

Continuous POMDPs

- Policy search methods (Aberdeen & Baxter, 2002; Baxter & Bartlett, 2001; Ng & Jordan, 2000; Williams & Singh, 1999).
- Approximate, i.e., grid- (Zhou & Hansen 2001) and point-based (Porta et al., 2005), value iteration algorithms may also be extended to fit the continuous space.

Beyond consideration of linear transition models, Schöbi & Chatzi (2016) extend the solution of continuous state POMDPs to nonlinear action models via use of Gaussian Mixtures and the Unscented Transform.

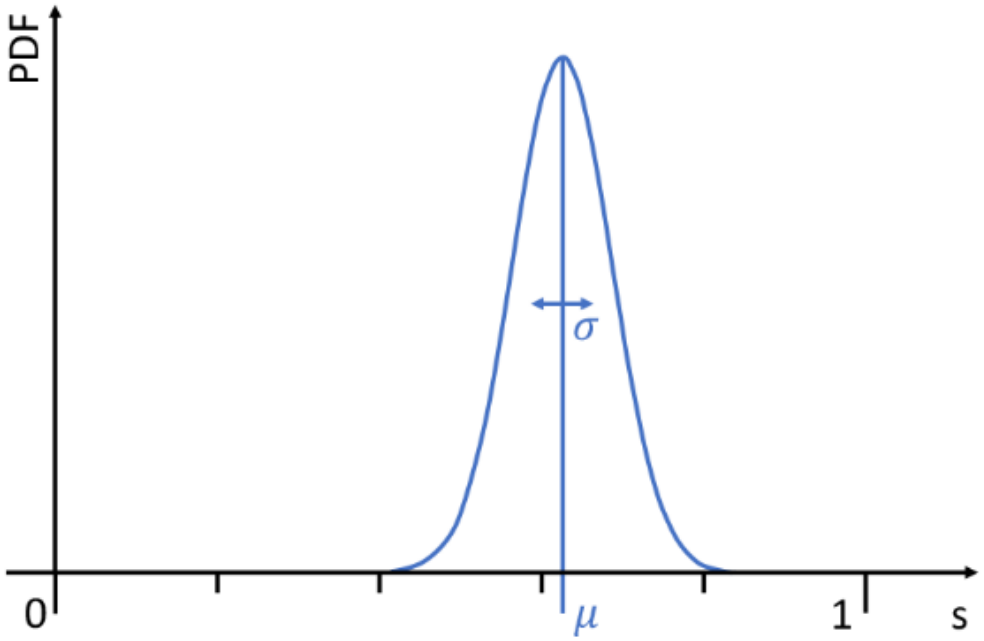
Example Application



Example Application

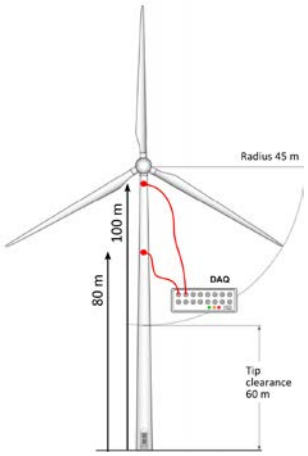
The **system state** S is 1-dimensional with a range $0 \leq s \leq 1$. (0 for failure, 1 for optimal condition of the bridge), e.g. damage index through vibrational data (natural frequencies)

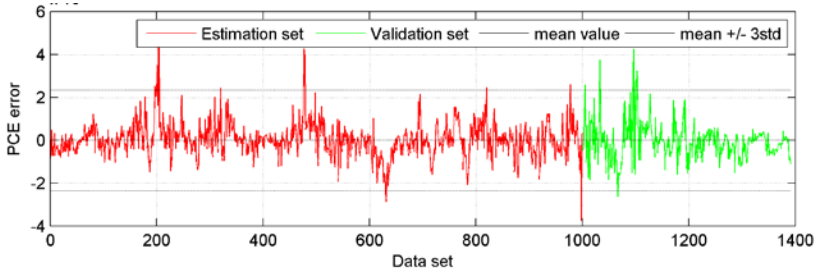
Cost for failure of the structure $C_{\text{failure}} = 1000$.



How to extract such a Condition Index?

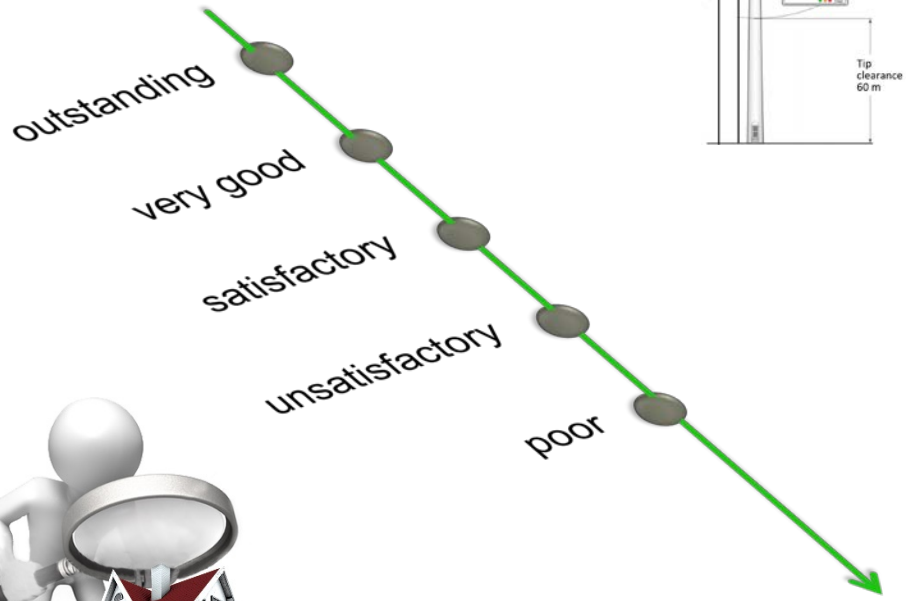
Permanent Monitoring

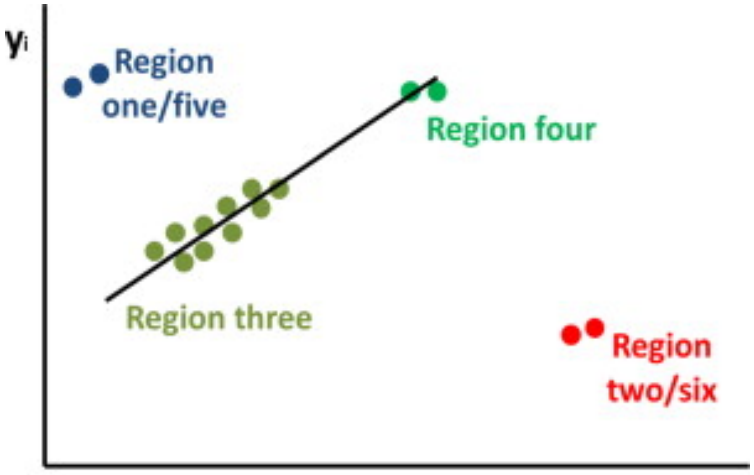




Spiridonakos & Chatzi, 2015

Inspection or NDE





Dervilis, Worden, & Cross, 2015

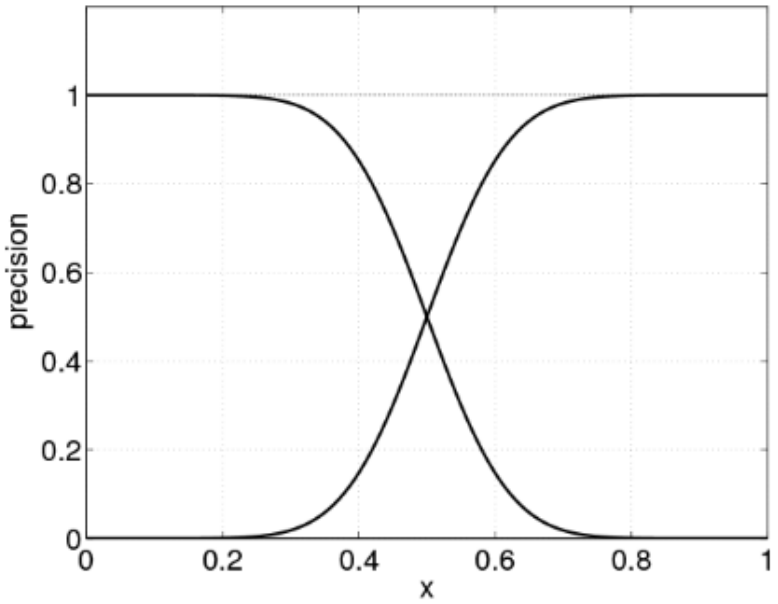
TU1406 COST ACTION

TU1402 Factsheet

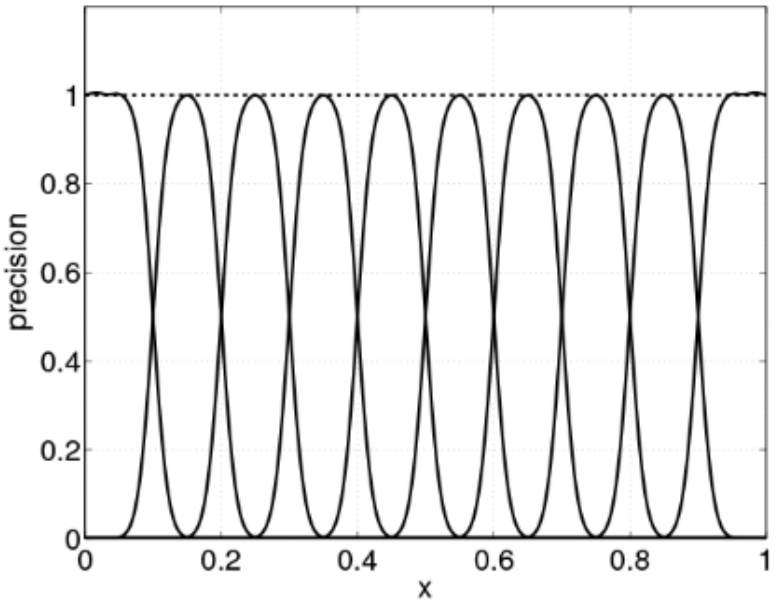
Example Application

Observations

Three possible inspection methods: "doing nothing" ($C_{\text{doing nothing}} = 0$), "visual inspection" ($C_{\text{visual}} = 1$), and "ND testing" ($C_{\text{ND}} = 5$)



(a) Visual inspection



(b) ND testing

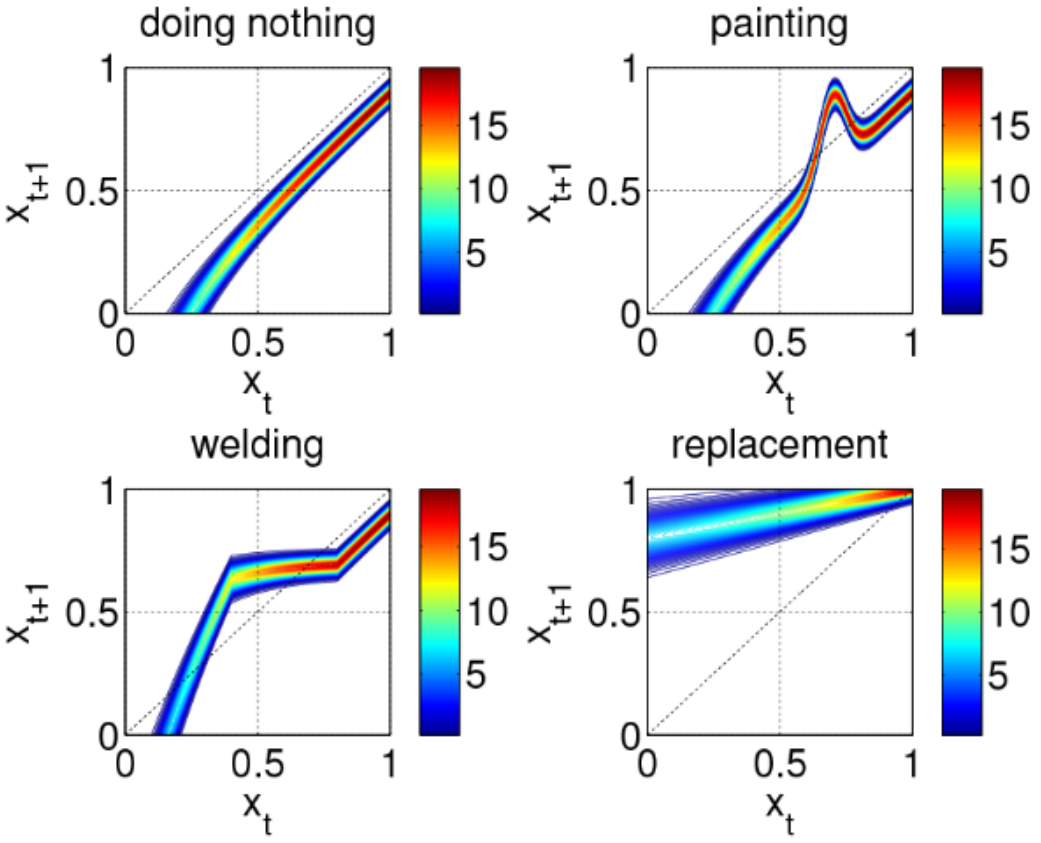
Example Application

Actions

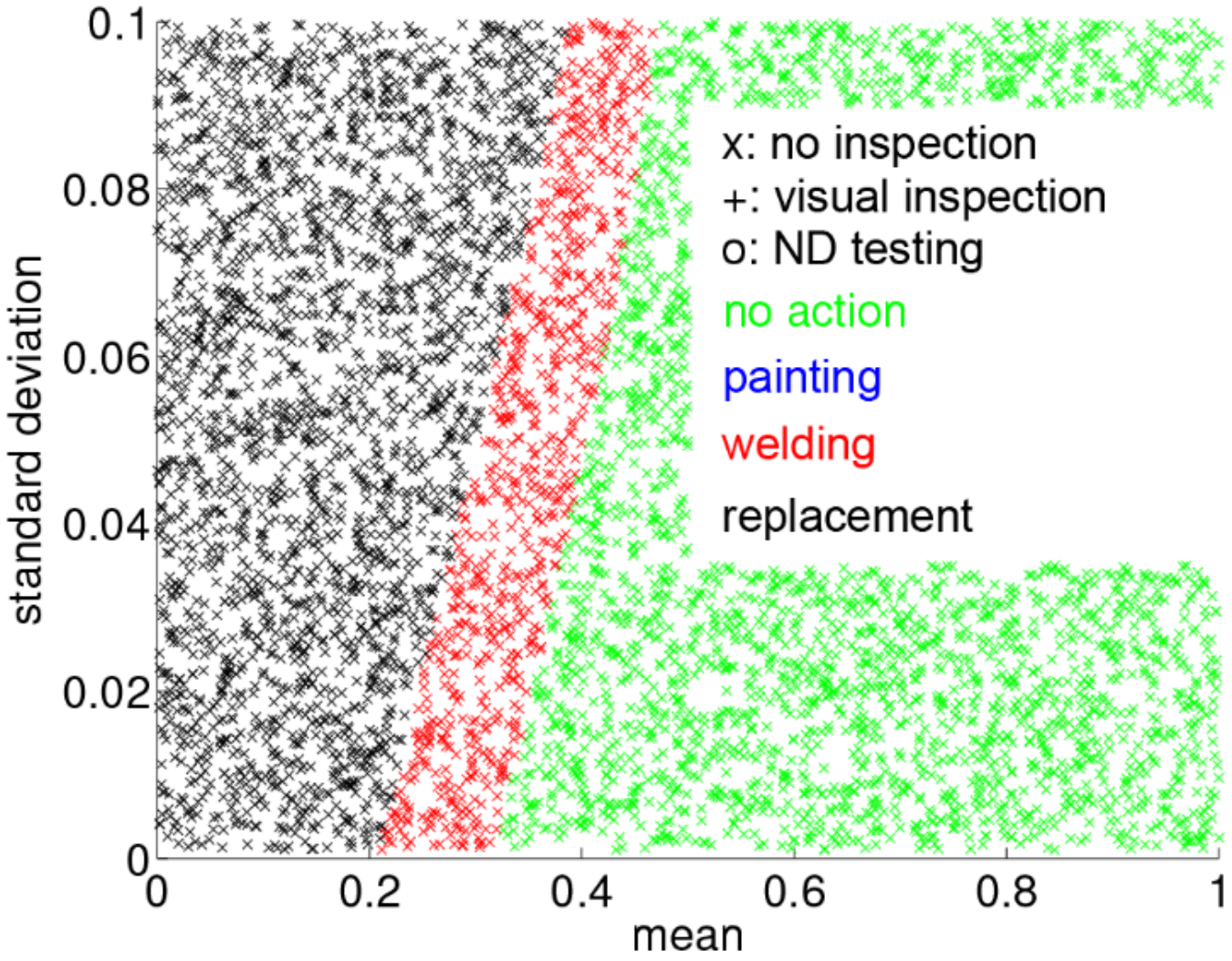
The action models are defined as the sum of a deterministic component and a stochastic component:

$$s' = f(s) = \mu_f(s) + \epsilon_f(s)$$

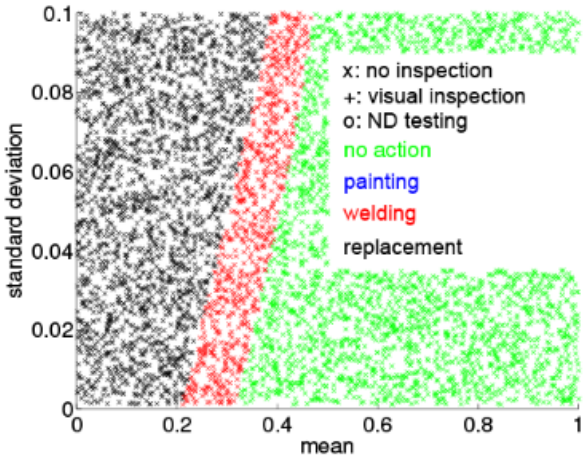
$$\begin{aligned} C_{\text{doing nothing}} &= 0 \\ C_{\text{painting}} &= 10 \\ C_{\text{welding}} &= 50 \\ C_{\text{replacement}} &= 100 \end{aligned}$$



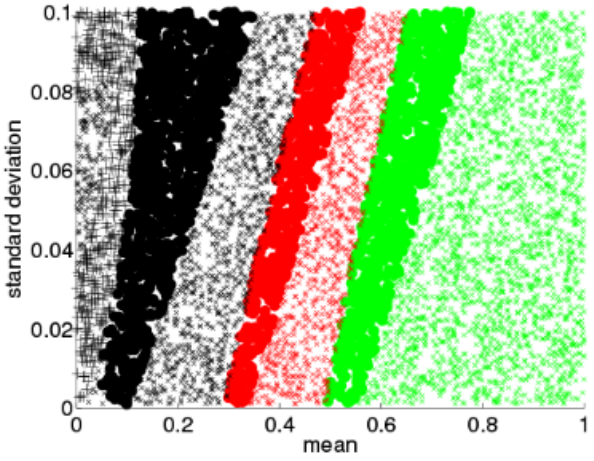
Example Application



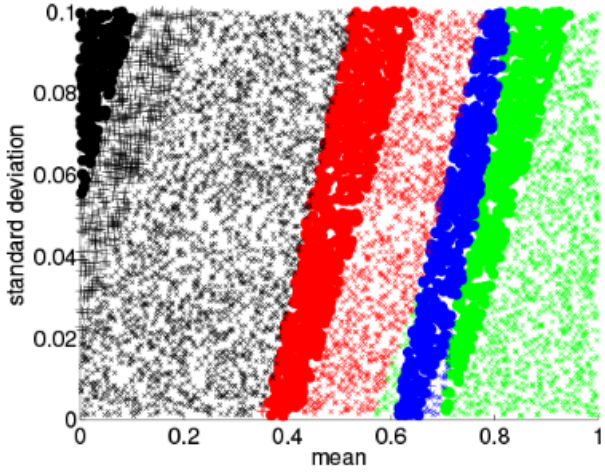
Example Application



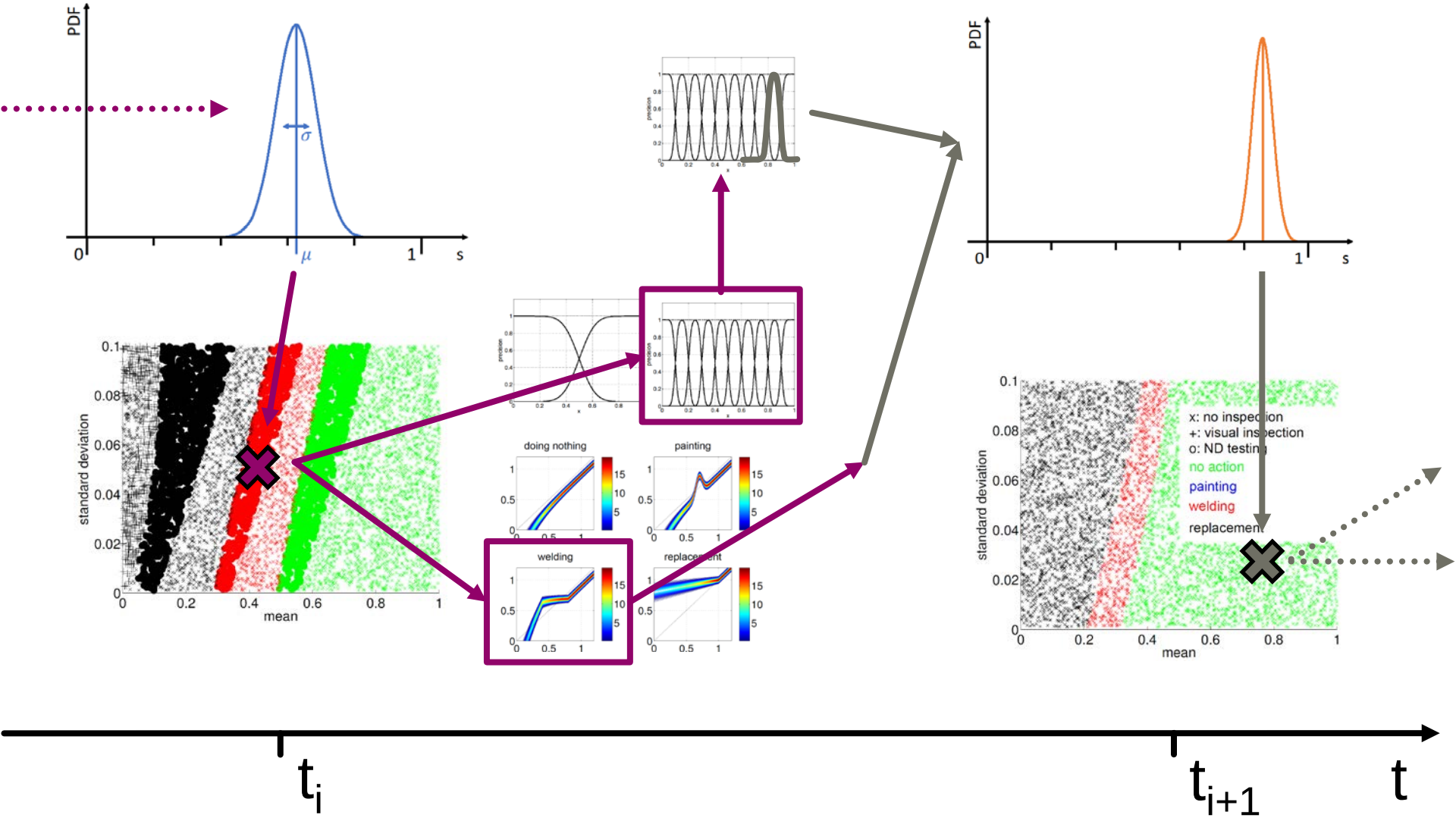
$n = 1$



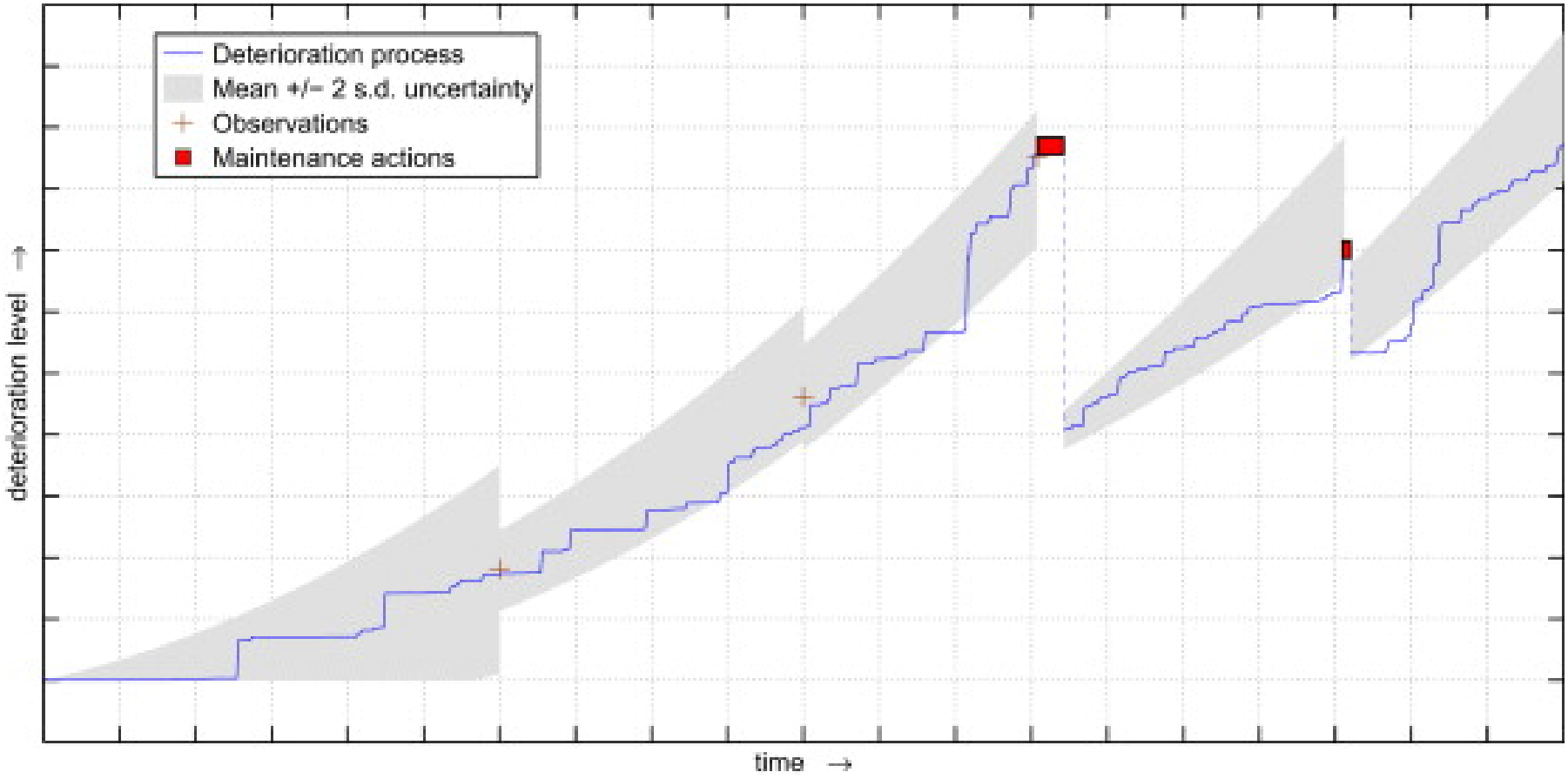
$n = 2$



$n = 3$


PDF
σ
μ
0
1
s
standard deviation
mean
doing nothing
painting
welding
replacement
x: no inspection
+: visual inspection
o: ND testing
no action
painting
welding
replacement
t_i
t_{i+1}
t

A schematic **POMDP policy** for a structural inspection and maintenance problem



Considerations

- Increase of complexity when dealing with high dimensional condition indices.
- How to properly define, and continuously update transition (action) models on the basis of inflowing information.
- How to extend the dependence to earlier points in time.
- How to properly deal with continuous monitoring systems, i.e., continuous flow of information, possibly moving closer to a real-time implementation scenario.

Literature – POMDP for Infrastructure Management

- Memarzadeh, M., Pozzi, M., and Zico Kolter, J. (2014). "Optimal Planning and Learning in Uncertain Environments for the Management of Wind Farms." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000390, 04014076.
- K.G. Papakonstantinou, M. Shinozuka, Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory, Reliability Engineering & System Safety, Volume 130, October 2014, Pages 202-213, ISSN 0951-8320
- K.G. Papakonstantinou, M. Shinozuka, Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation, Reliability Engineering & System Safety, Volume 130, October 2014, Pages 214-224
- Roland Schöbi & Eleni N. Chatzi (2016) Maintenance planning using continuous-state partially observable Markov decision processes and nonlinear action models, Structure and Infrastructure Engineering, 12:8, 977-994

ETH zürich

ibk
Institut für Baustatik und Konstruktion

We welcome questions/comments/collaboration:

chatzi@ibk.baug.ethz.ch