Linear Programming for large Markov Decision Processes and an application to a maintenance problem at RTE

Young Researcher Seminar

Solène Delannoy-Pavy
Supervisors: Vincent Leclère, Axel Parmentier

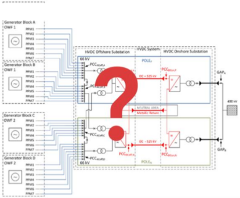
07/07/26

CERMICS
École nationale des ponts et chaussées

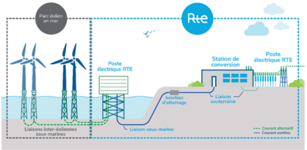
Building the French offshore electrical network

2026

design selection

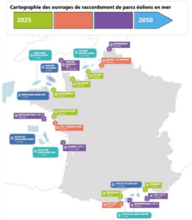


building the offshore network



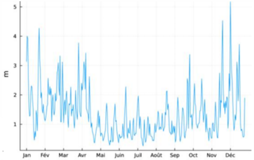
2050

45 GW capacity



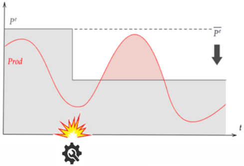
Novelties related to offshore cost estimation

Accessibility depends on the weather.

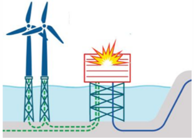


Wave height
Saint-Brieuc - 2022

Penalties outside of a contractual quota of free maintenance days.

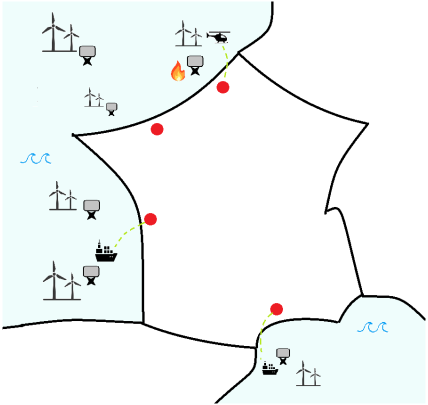


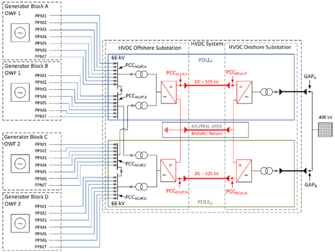
Failure



Maintenance

Optimizing maintenance of offshore electrical substations





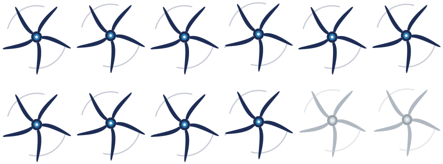
- Many components per substation
- Boats, helicopters and workers are shared among the stations

We work with very large Markov Decision Processes

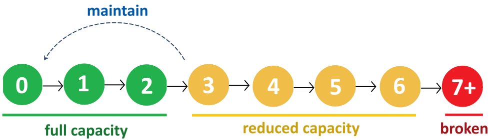
Contents

Modeling fan maintenance as a MDP

Component	**Notation**	**Fan example**
States	$x \in \mathcal{X}$	# broken fans $\in \{0, \dots, 7+\}$
Actions	$a \in \mathcal{A}$	maintain / do nothing
Costs	$c(x, a)$	capacity loss
Transitions	$p(x' \mid x, a)$	fan failure probabilities
Horizon	T	40 years



10 fans + 2 redundant



We are searching for good policies $\pi(x)$.

Linear Programming (LP) approach for MDPs

Dynamic Programming

Backward induction on $V_t(x)$:

$$V_T(x) = 0 \quad \forall\, x$$

$$V_t(x) = \min_{a \in \mathcal{A}} \Big[c(x,a) + \sum_{x'} p(x' \mid x, a) V_{t+1}(x') \Big]$$

Optimal policy: $\pi_t^*(x) = \arg\min_a [\cdots]$

Linear Programming variables: $v_x(t)$

$$\max_v \sum_x \alpha_x \, v_x(1)$$

$$\text{s.t.} \quad v_x(t) \leq c(x,a) + \sum_{x'} p(x' \mid x, a)\, v_{x'}(t+1), \quad \forall \quad x, a, t < T$$

$$v_x(T) = 0, \quad \forall x$$

$\alpha_x > 0$: initial distribution.

Under some assumptions ($\alpha_x > 0$ for all x...) any optimal LP solution satisfies the Bellman equations at equality.

The dual LP: occupancy measures

variables: $\mu_{x,a}(t) \geq 0$

Probability of visiting state x and taking action a at stage t.

$$\min_{\mu \geq 0} \quad \sum_{x,a,t} c(x,a)\, \mu_{x,a}(t)$$

Search for a stochastic process with minimum expected cost...

$$\text{s.t.} \quad \sum_a \mu_{x,a}(1) = \alpha_x \quad \forall x$$

$$\sum_a \mu_{x,a}(t) = \textstyle\sum_{x',a} p(x \mid x', a)\, \mu_{x',a}(t-1) \quad \forall x,\ t \geq 2$$

...amongst those consistent with the MDP: initial state and flow constraint.

By strong duality, the optimal values of the two LPs are equal.

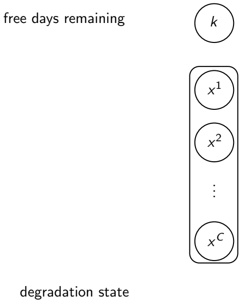
What are we doing today ?

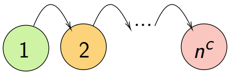
- In this talk we will focus on the dual and try to build approximations.
- Remember that the variables are probabilities.

Contents

MDP: state

The station consists of C components.



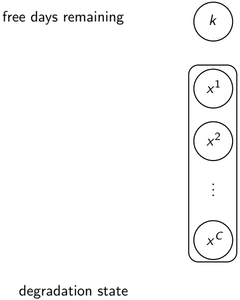


Global degradation state

$$x = (x^c)_{c \in [1:C]} \in \bigotimes_{c \in [1:C]} [1 : n^c].$$

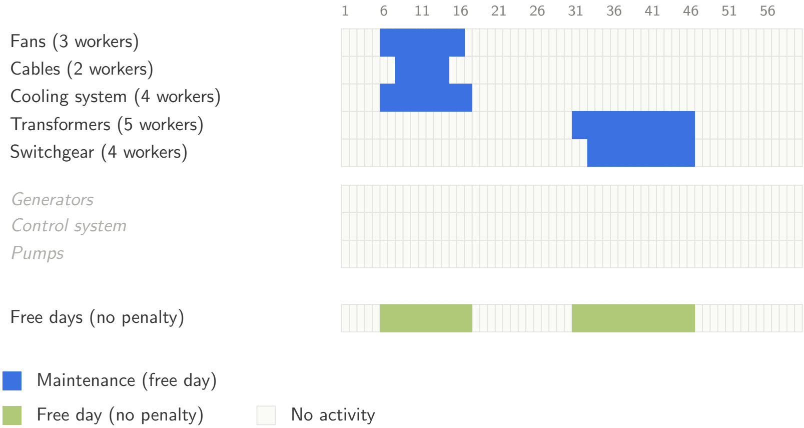
$\prod_{[1:C]} n^c$ possible degradation states.

MDP: state



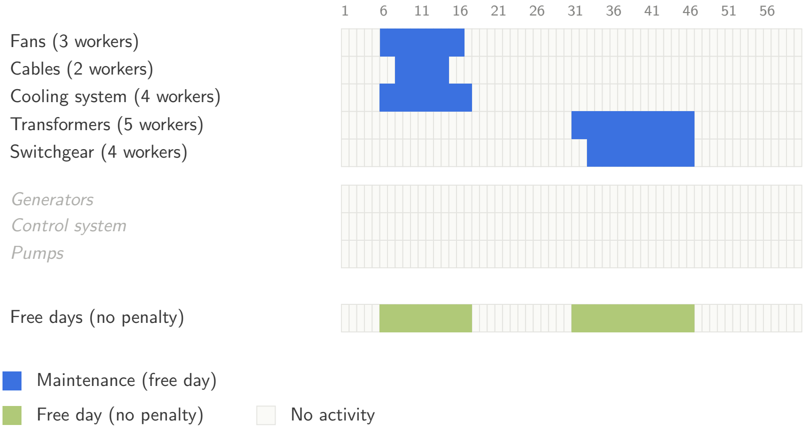
Maintenance schedules

Time step of the MDP = 2 months



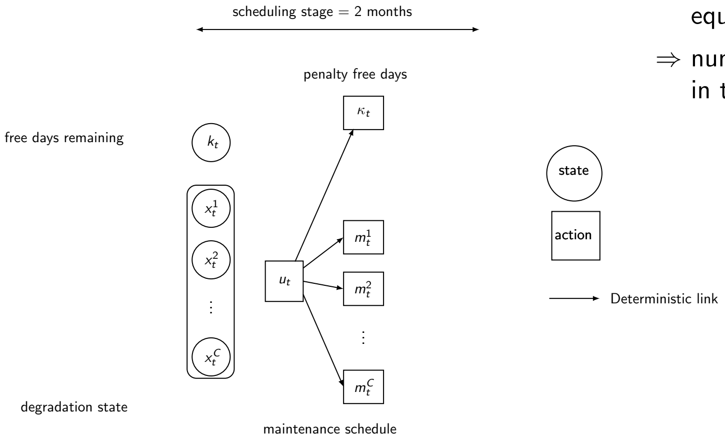
Maintenance schedules

Time step of the MDP $=$ 2 months



For now we assume we have a small catalogue of admissible maintenance schedules.

MDP: actions

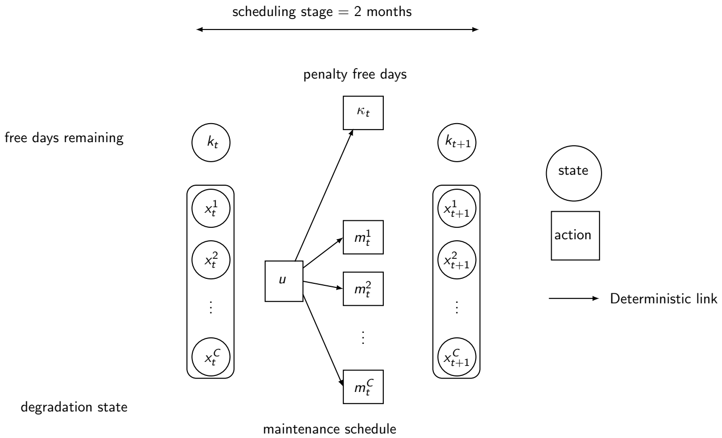


maintenance schedule u

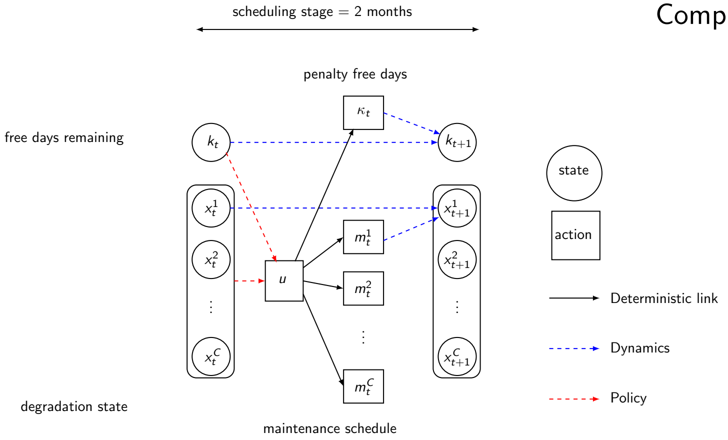
⇒ binary variable m^c for $c \in [1 : C]$, equals 1 if c is maintained.

⇒ number of penalty free days used in the scheduling stage κ.

MDP: transitions

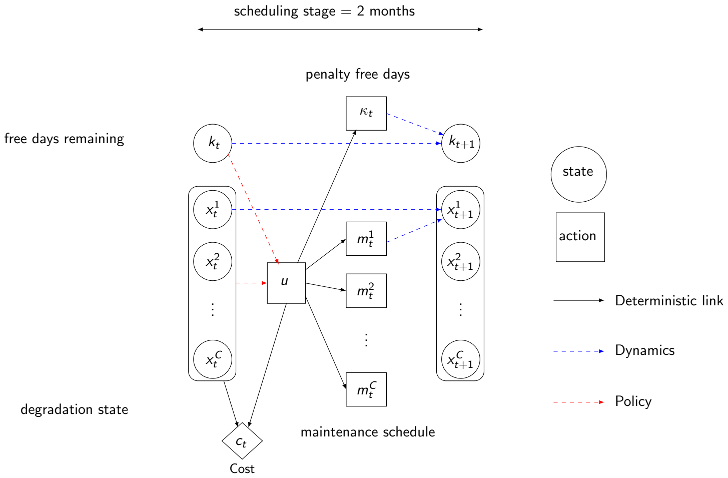


MDP: transitions

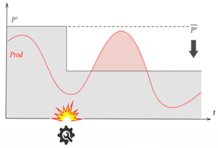


Components degrade independently.

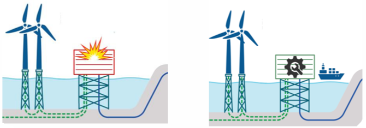
Cost



Cumulative penalties over the two months, accounting for both



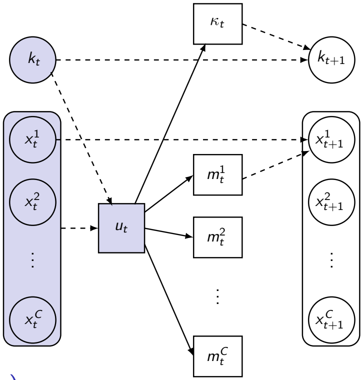
failures



and maintenance

Solving the MDP

- The size of the state space prevents solving by dynamic programming.
- The variables in the LP represent the probabilities of the values taken by the global (state, action) pair at each time step.

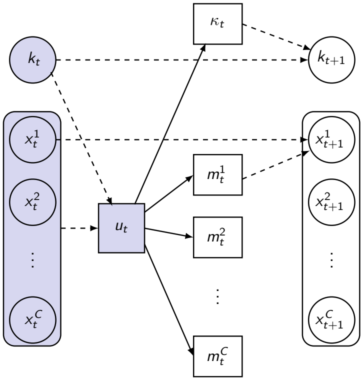


$$\begin{aligned} \underset{\text{proba}\geq 0}{\text{minimize}} \quad & \sum_t \sum_{\text{state,action}} \text{cost}_t(\text{state}, \text{action}) \times \text{proba}_t(\text{state}, \text{action}) \\ \text{s. t.} \quad & \text{Flow constraints} \end{aligned}$$

This LP is also not tractable due to the number of variables.

Solving the MDP

- The size of the state space prevents solving by dynamic programming.
- The variables in the LP represent the probabilities of the values taken by the global (state, action) pair at each time step.



$$\underset{\mu \geq 0}{\text{minimize}} \quad \sum_{t=1}^{240} \sum_{x,k,u} c_t(x,u)\mu_{xku}(t)$$

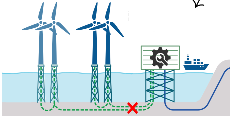
$$\text{s. t.} \quad \sum_{u} \mu_{x'k'u}(t) = \sum_{x,k,u} \mathbb{1}_{\{\phi(k,K(u))=k'\}} p(x'|x,u)\mu_{xku}(t-1)$$

This LP is also not tractable due to the number of variables.

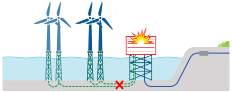
Exploiting the structure of the MDP

- The transitions decompose by component.
- The cost can be approximated as a sum of terms over a small number of subgroups of variables:

$$c_t(x,k,u) \simeq \tilde{c}_t(u) + \sum_c \tilde{c}^c(x^c, m^c) - \sum_{c,c'} \tilde{c}_t^{c,c'}(x^c, x^{c'}) + \dots$$



Maintenance



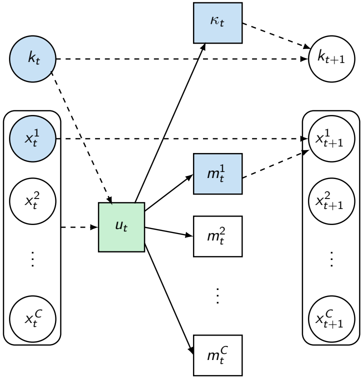
Failures

(It is a factored MDP)

Tractable LP approximation

$$\begin{aligned} \underset{\text{proba},\text{proba}\geq 0}{\text{minimize}} \quad & \sum_t \sum_{\text{action}} \text{cost}_t(\text{action}) \times \text{proba}_t(\text{action}) \\ & + \sum_c \sum_{\text{state}^c,\text{action}^c} \text{cost}_t^c(\text{state}^c, \text{action}^c) \times \text{proba}_t^c(\text{state}^c, \text{action}^c) \\ \text{s.t.} \quad & \text{Local flow constraints;} \\ & \text{Consistency between } \mu \text{ and } \nu. \end{aligned}$$

- We introduce variables representing the probabilities of the value taken by the local $(\text{state}^c, \text{action}^c)$ pair for each component c[1].



[1] Bertsimas and Mišić 2016.

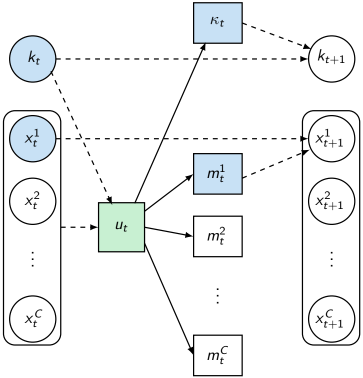
Tractable LP approximation

$$\underset{\mu,\nu\geq 0}{\text{minimize}} \quad \sum_{t=1}^{240}\sum_{u} \tilde{c}_t(u)\nu_u(t) + \sum_{c=1}^{C}\sum_{x^c,k,m^c,\kappa} \tilde{c}_t^c(x^c,m^c)\mu^c_{x^c k m^c \kappa}(t)$$

$$\text{s.t.} \quad \sum_{m^c,\kappa} \mu^c_{x'^c k' m^c \kappa}(t) = \sum_{x^c,k,m^c,\kappa} \mathbb{1}_{\{\phi(k,\kappa)=k'\}} p^c(x'^c|x^c,m^c)\mu^c_{x^c k m^c \kappa}(t-1);$$

$$\sum_{x^c,k} \mu^c_{x^c k m^c \kappa}(t) \leq \sum_{\substack{u|M^c(u)=m^c\\K(u)=\kappa}} \nu_u(t).$$

- We introduce variables representing the probabilities of the value taken by the local (state^c, action^c) pair for each component c[1].



[1] Bertsimas and Mišić 2016.

Are we happy with that ?

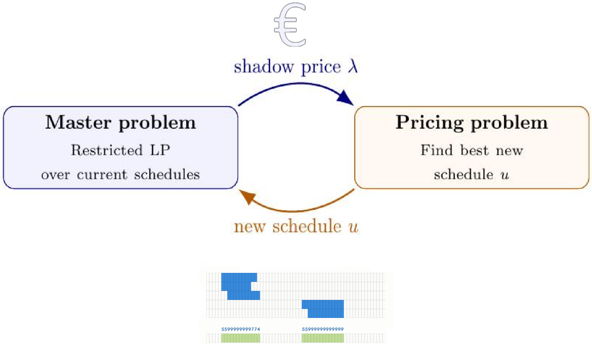
- Is it exact ? No.
- Is it good enough ? Yes.

What if the set of admissible schedules is combinatorial ?

Contents

Introducing column generation

- Iteratively add variables (columns) to the master problem.



- Simple but surprisingly effective in many applications, provided that the pricing problem can be efficiently solved.

Column generation and Rolling Horizon heuristic (RH)

Master Problem

$$\min_{\mu,\,\nu\geq 0} \quad \sum_{u\in\mathcal{U}',\,t} \tilde{c}^g_t(u)\,\nu_u(t) + \sum_{c,x^c,m^c,t} \tilde{c}^c_t(x^c,m^c)\,\mu^c_{x^c k m^c \kappa}(t)$$

s.t. flow constraint

$$\sum_{x^c,k} \mu^c_{x^c k m^c \kappa}(t) \leq \sum_{\substack{u\in\mathcal{U}':\\ M^c(u)=m^c,\,K(u)=\kappa}} \nu_u(t) \qquad [\lambda_t^{c,m^c,\kappa}].$$

Solved over restricted set $\mathcal{U}' \subseteq \mathcal{U}$

dual variables $\lambda_t^{c,m^c,\kappa}$

new column $u^\star$

$\overline{c}(u^\star,t) < 0$

Pricing Problem

$$\min_{u\in\mathcal{U}} \tilde{c}^g_t(u) + \sum_c \lambda_t^{c,\,M^c(u),\,K(u)}$$

MILP over the full combinatorial set $\mathcal{U}$

Column generation and Rolling Horizon heuristic (RH)

Master Problem

$$
\begin{aligned}
\min_{\mu,\,\nu\geq 0} \quad & \sum_{u\in\mathcal{U}',\,t} \tilde{c}_t^g(u)\,\nu_u(t) \\
& + \sum_{c,x^c,m^c,t} \tilde{c}_t^c(x^c,m^c)\,\mu^c_{x^c k m^c \kappa}(t) \\
\text{s.t.} \quad & \text{flow constraint} \\
& \sum_{x^c,k} \mu^c_{x^c k m^c \kappa}(t) \leq \sum_{\substack{u\in\mathcal{U}':\\ M^c(u)=m^c,\,K(u)=\kappa}} \nu_u(t) \\
& [\lambda_t^{c,m^c,\kappa}].
\end{aligned}
$$

Solved over restricted set $\mathcal{U}' \subseteq \mathcal{U}$

dual variables $\lambda_t^{c,m^c,\kappa}$

Pricing Problem

$$\min_{u\in\mathcal{U}} \ \tilde{c}_t^g(u) + \sum_c \lambda_t^{c,\,M^c(u),\,K(u)}$$

MILP over the full combinatorial set $\mathcal{U}$

new column $u^\star$
$\overline{c}(u^\star,t) < 0$

Rolling Horizon heuristic

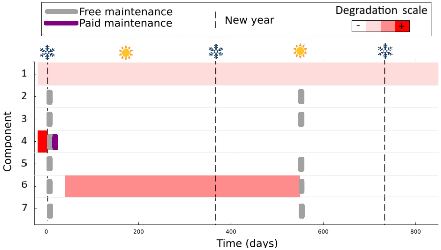
1. Fix the first-stage distribution to the current state.
2. Solve the LP over a cropped horizon.
3. Apply the highest-probability action: $a^\star = \arg\max_{\text{action}^g} \text{proba}_1(\text{action}^g)$.

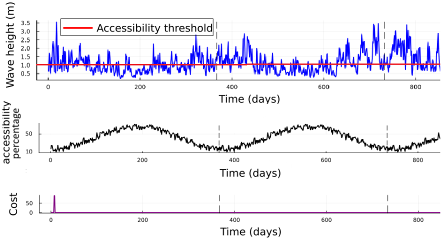
Case study

- **7** key High-Voltage Direct Current (HVDC) components.
- Weather scenarios corresponding to the upcoming **Centre Manche 1** project location, built using Copernicus data.
- Benchmark against an **operational rule** (among others).

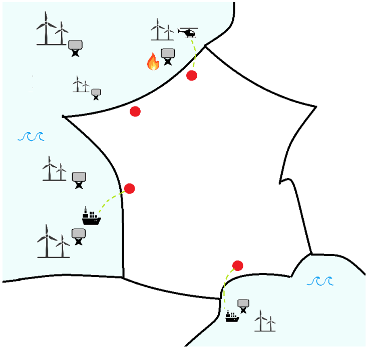
Table: Simulated cost over 40 years

Policy	Mean	$VaR_{0.05}$
RH	**126**	257
Naive	249	1671





- We solve a large MDP with combinatorial action using a tractable LP approximation exploiting the structure of the MDP.



Thank you !