EPFL

Machine Learning for Predictive Maintenance Applications:
Fleet approaches / Prescriptive maintenance/operations concepts

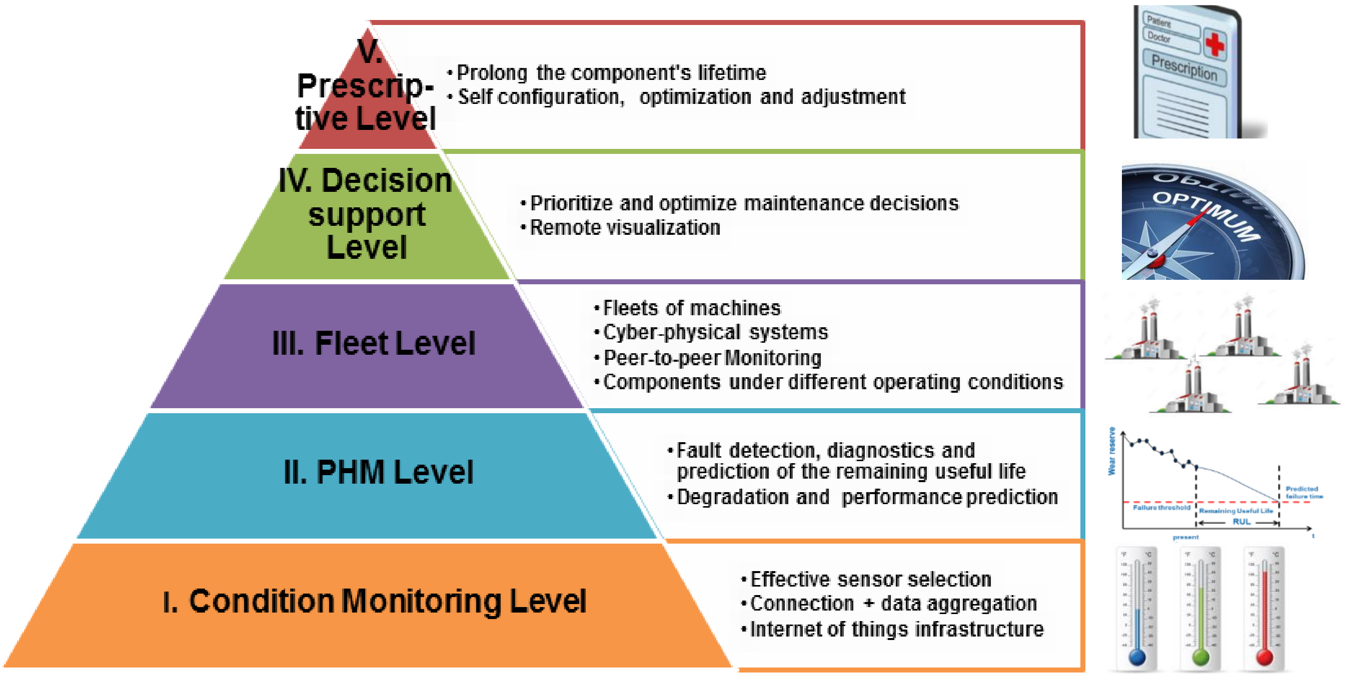
Prof. Dr. Olga Fink

December 2025

École polytechnique fédérale de Lausanne

Fleet approaches

Five levels of condition-based and predictive maintenance



Challenges in Predictive Maintenance

Varying and evolving operating conditions → Even healthy system conditions are not always representative due to limited observation time period
→ Representative operating conditions (and features) required

Algorithms also for systems required that are newly taken into operation

What do we start with?

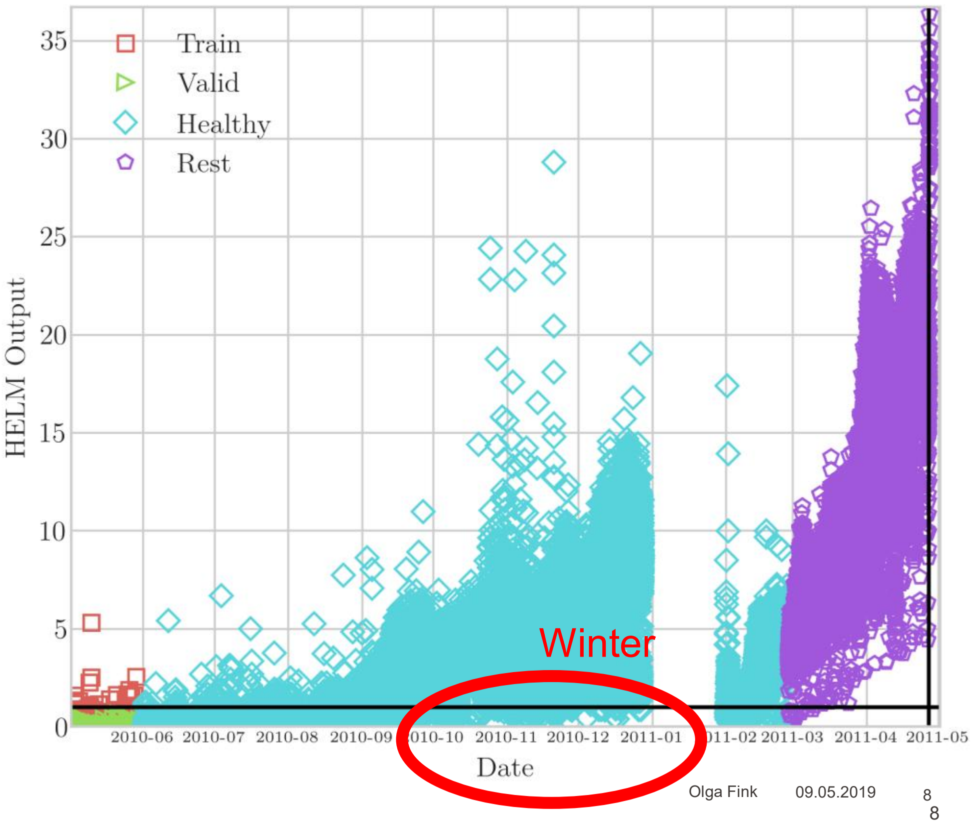
- Limited number of faults (labels)
- Large variety of condition monitoring data under different operating conditions
- Several units of the same fleet (but units have variability in their configurations and operating conditions)
- Heterogenous operating conditions and configurations of the fleet units
- Limited observation time periods
- **Limited representativeness** of the collected data for the expected operating conditions

What are we trying to achieve?

- Compile representative training datasets that are valid for the specific units under the specific operating conditions (homogeneous datasets)
- Using labeled and unlabeled data as efficiently as possible at the level of an entire fleet
- Develop also algorithms for new units
- Transferring knowledge (on operating conditions and faults) between the single units of a fleet
- Learn robust features that are invariant to different operating conditions

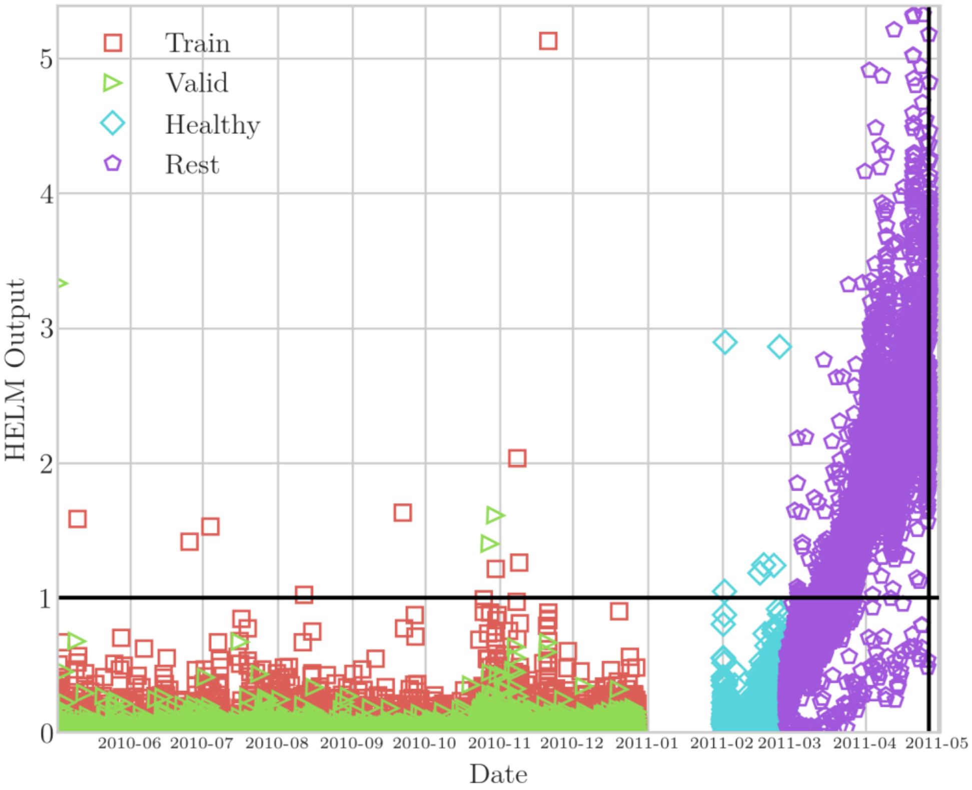
Example Gas Turbines

Only a short observation period



Example Gas Turbines

Extend the observation period



Different options for fleet learning (1/4)

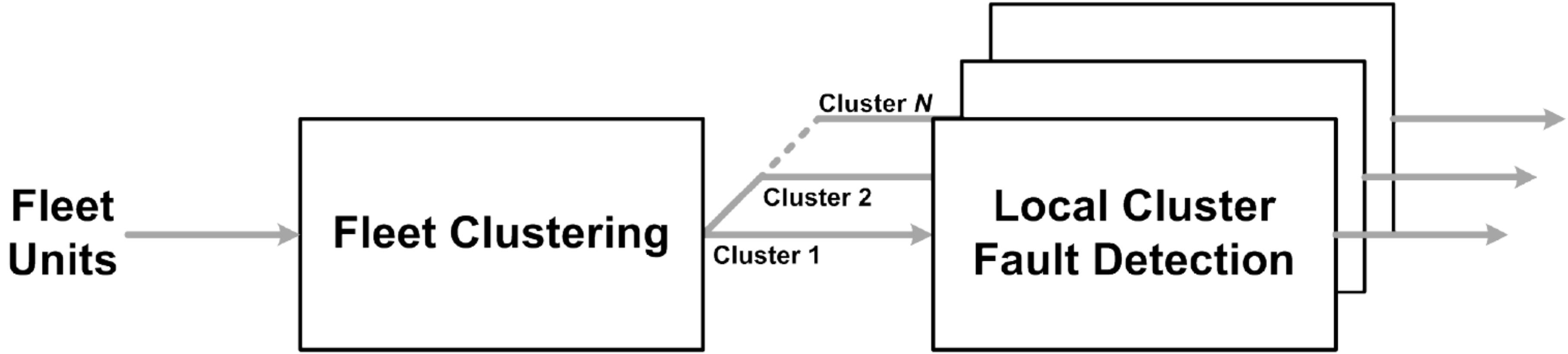
Goal: identify similar units of a fleet that could form a homogenous sub-fleet

Problem: What does similar mean? How could similarity be defined?

1. Identify some relevant operating or design parameters of the units (e.g. average operating regimes)
2. Find sub-fleets (possibly with clustering) defined by similar characteristics based on the selected parameters;
3. Use the subsets of condition monitoring data of each of the sub-fleets to train the algorithms
4. Apply “specialized” models for the PHM tasks on all the units within the sub-fleet

Challenge: Aggregated parameters used for comparison may not cover all the relevant conditions or the aggregated parameters may not be representative of the unit specificities

Basic principle of identifying similar units of a fleet

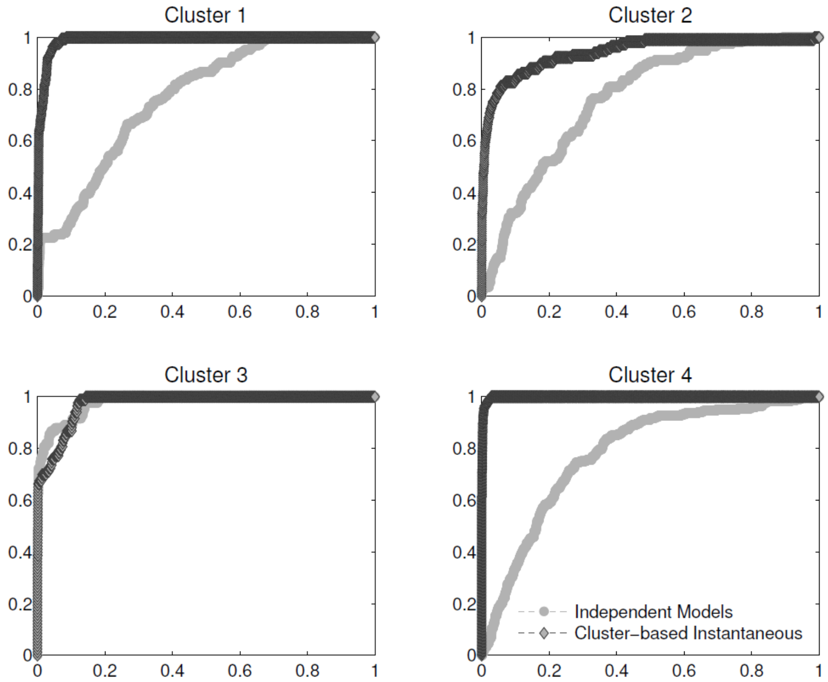


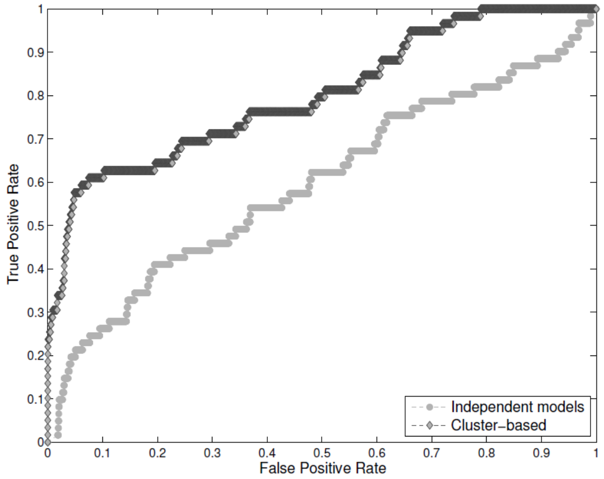
Lapira, Edzel R. *Fault detection in a network of similar machines using clustering approach*. Diss. University of Cincinnati, 2012.

Olga Fink

EPFL

Receiver operating characteristic curves for servo-gun (left) and wind turbines (right) fault detection





Lapira, Edzel R. *Fault detection in a network of similar machines using clustering approach*. Diss. University of Cincinnati, 2012.

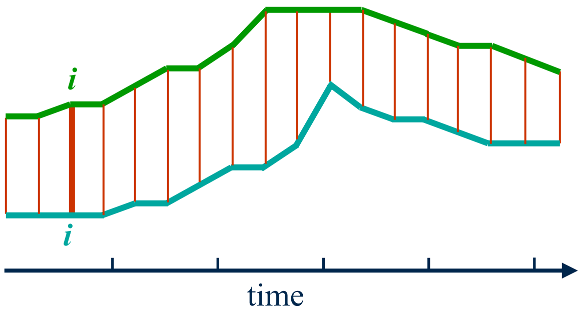
Different options for fleet learning (2/4)

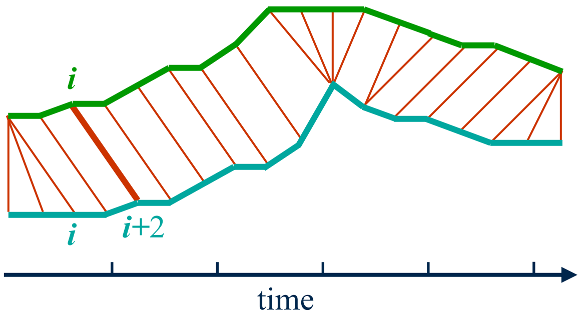
1. Use the entire time series of condition monitoring signals to identify similar sub-fleets
2. Perform time-series clustering to find sub-fleets
3. Use the subsets of condition monitoring data of each of the sub-fleets to train the algorithms
4. Apply “specialized” models for the PHM tasks on all the units within the sub-fleet

Challenges:

1) Comparing the distances between time series is affected by the curse of dimensionality.
2) Time series cluster analysis becomes even more challenging when operating conditions evolve over time.

Why Dynamic Time Warping?





Any distance (Euclidean, Manhattan, …) which aligns the *i*-th point on one time series with the *i*-th point on the other will produce a poor similarity score.

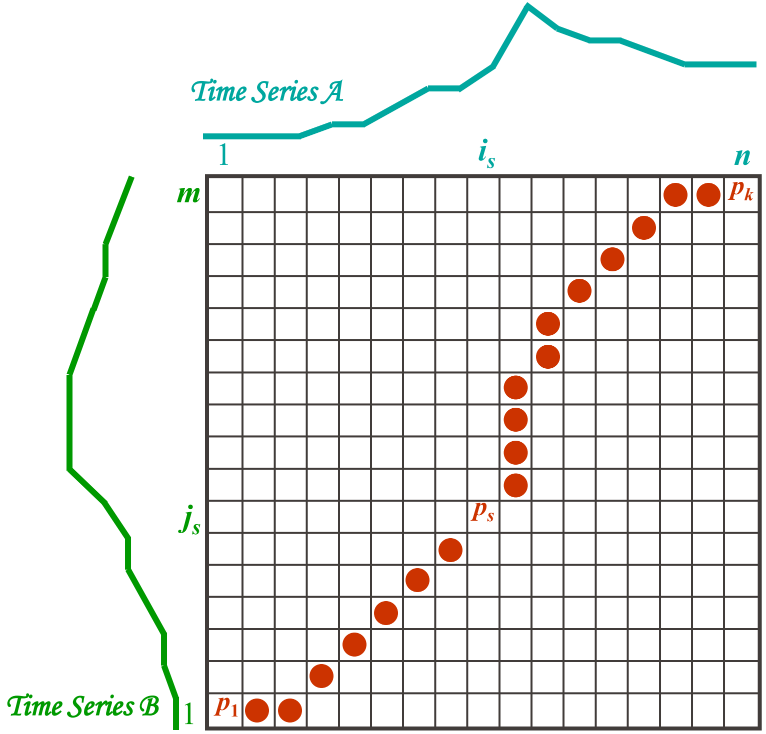
A non-linear (elastic) alignment produces a more intuitive similarity measure, allowing similar shapes to match even if they are out of phase in the time axis.

Source: Elena Tsiporkova

Dynamic Time Warping

- Dynamic Time Warping (DTW) is a technique used to measure the similarity between two temporal sequences, which may vary in speed or length.
- It aligns sequences by warping them non-linearly in the time dimension to minimize the distance between corresponding points.
- DTW is commonly applied in time-series analysis, speech recognition, and gesture recognition.
- It computes an optimal alignment path between sequences using **dynamic programming** to find the minimal cumulative distance.
- The method is robust to shifts, scaling, and distortions in time, making it suitable for comparing sequences with varying patterns.
- DTW can be computationally intensive for long sequences.

Warping Function



To find the *best alignment* between $\mathcal{A}$ and $\mathcal{B}$ one needs to find the path through the grid

$P = p_1, \dots, p_s, \dots, p_k$

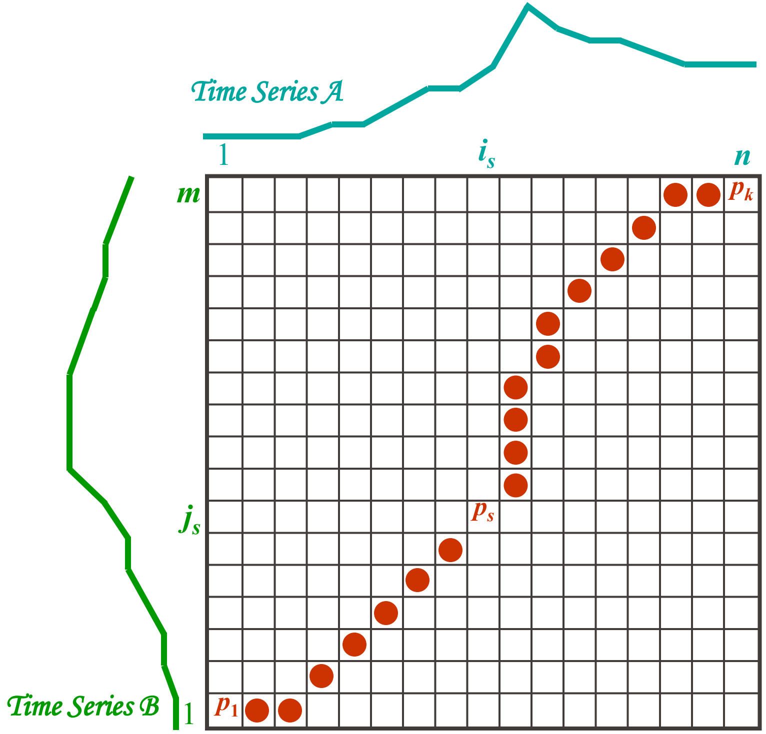
$p_s = (i_s, j_s)$

which *minimizes* the total distance between them.

P is called a *warping function*.

Source: Elena Tsiporkova

Time-Normalized Distance Measure



<u>*Time-normalized distance*</u> between $\mathcal{A}$ and $\mathcal{B}$:

$$D(\mathcal{A}, \mathcal{B}) = \left[\frac{\sum_{s=1}^{k} d(p_s) \cdot w_s}{\sum_{s=1}^{k} w_s} \right]$$

$d(p_s)$: distance between i_s and j_s

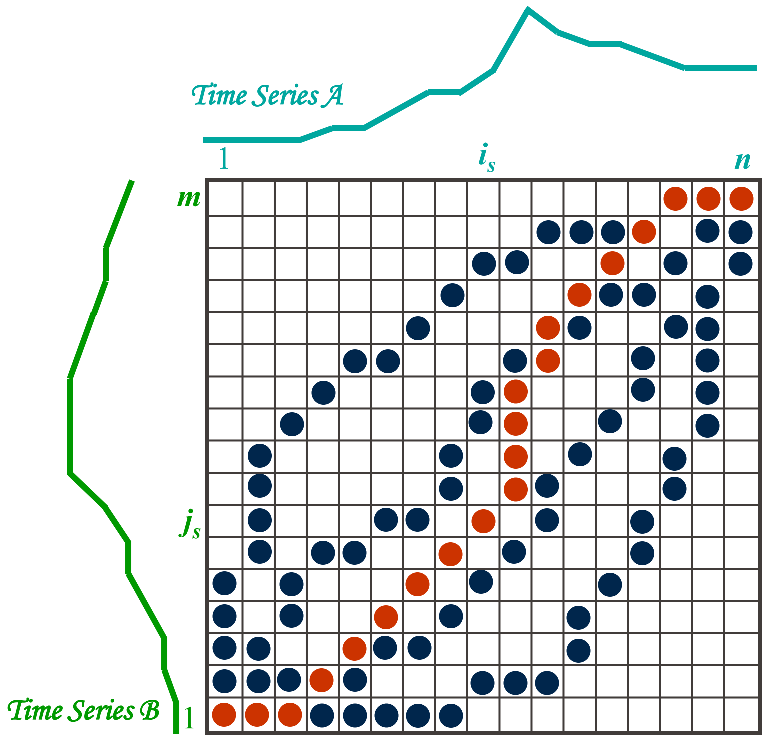
$w_s > 0$: weighting coefficient.

<u>*Best alignment path*</u> between $\mathcal{A}$ and $\mathcal{B}$:

$$P_0 = \arg\min_{P} (D(\mathcal{A}, \mathcal{B}))$$

Source: Elena Tsiporkova

Optimisations to the DTW Algorithm



The number of possible warping paths through the grid is exponentially explosive!

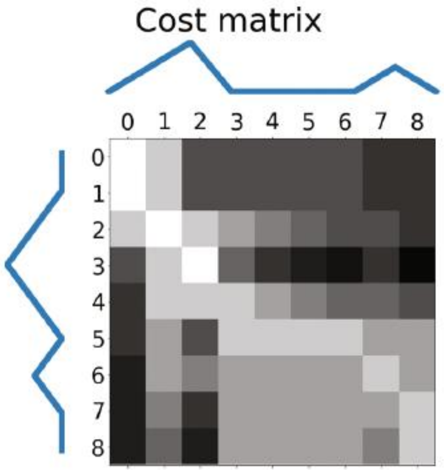
reduction of the search space

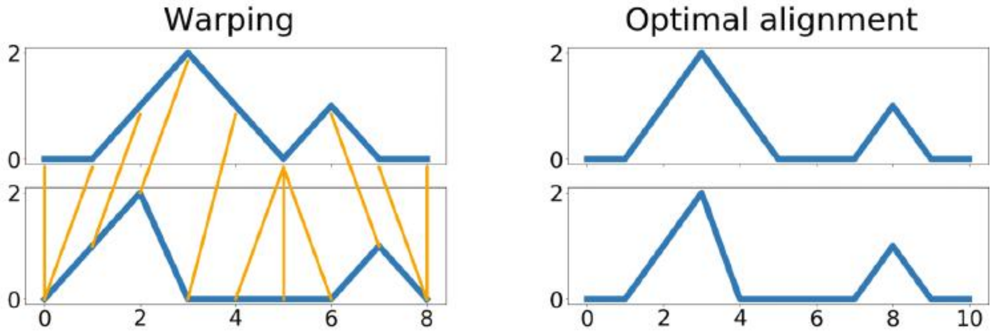
Restrictions on the warping function:

- monotonicity
- continuity
- boundary conditions
- warping window
- slope constraint.

Source: Elena Tsiporkova

Example

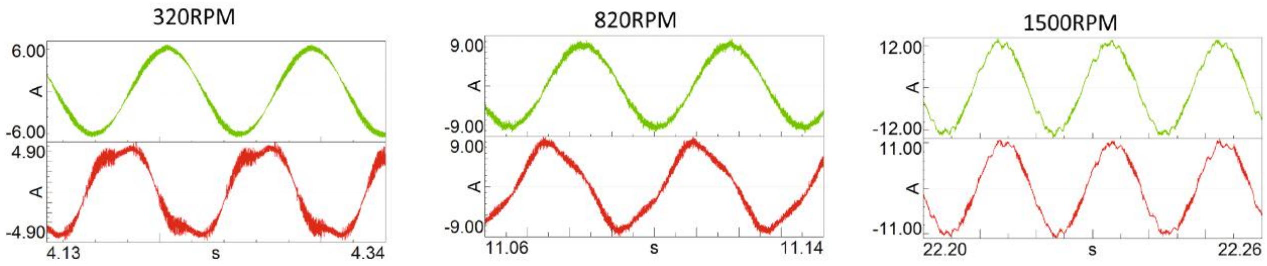




Hendrickx, Kilian, et al. "A fleet-wide approach for condition monitoring of similar machines using time-series clustering." 2019.

Case study

- Drivetrains 1–5: Squirrel cage induction motor (SCIM) at the driving side and DC motor at the load side.
- Drivetrains 6–10: SCIM at the driving side and Wound Rotor Synchronous Motor (WRSM) at the load side.
- A phase unbalance is introduced at one drivetrain for each of these groups



Hendrickx, Kilian, et al. "A fleet-wide approach for condition monitoring of similar machines using time-series clustering." 2019.

Example of alignment

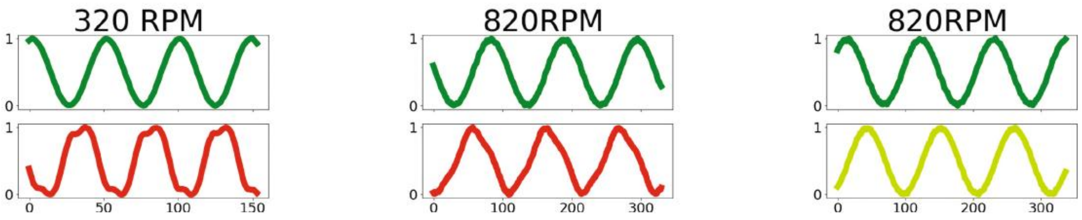


Fig. 5. Raw signals showing a faulty (red) and two healthy (green) drivetrains

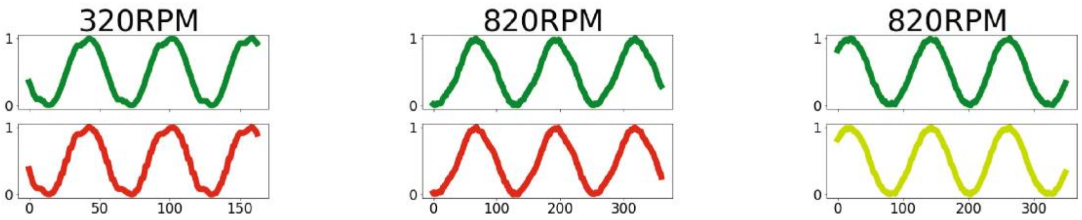


Fig. 6. Warped signals for a faulty (red) and two healthy (green) waveforms

Hendrickx, Kilian, et al. "A fleet-wide approach for condition monitoring of similar machines using time-series clustering." 2019.

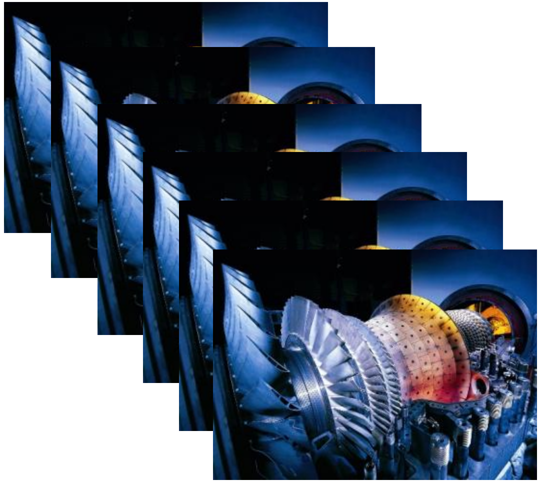
EPFL

Different options for fleet learning (3/4)

1. Develop models for the functional behaviour of the units
2. Apply the models of the functional bahaviour to other units (train on one unit → apply to other units)
3. Define thresholds for the similar functional behavior between the units
4. Use the subsets of condition monitoring data of each of the sub-fleets to train the algorithms
5. Apply “specialized” models for the PHM tasks on all the units within the sub-fleet

Challenge: one of the underlying requirements is that the units experience a sufficient similarity in their operating regimes. If the units are operated in a dissimilar way, large fleets may be required to find units with a sufficient similarity.

Solution: Using the fleet experience! → Transfer the experience



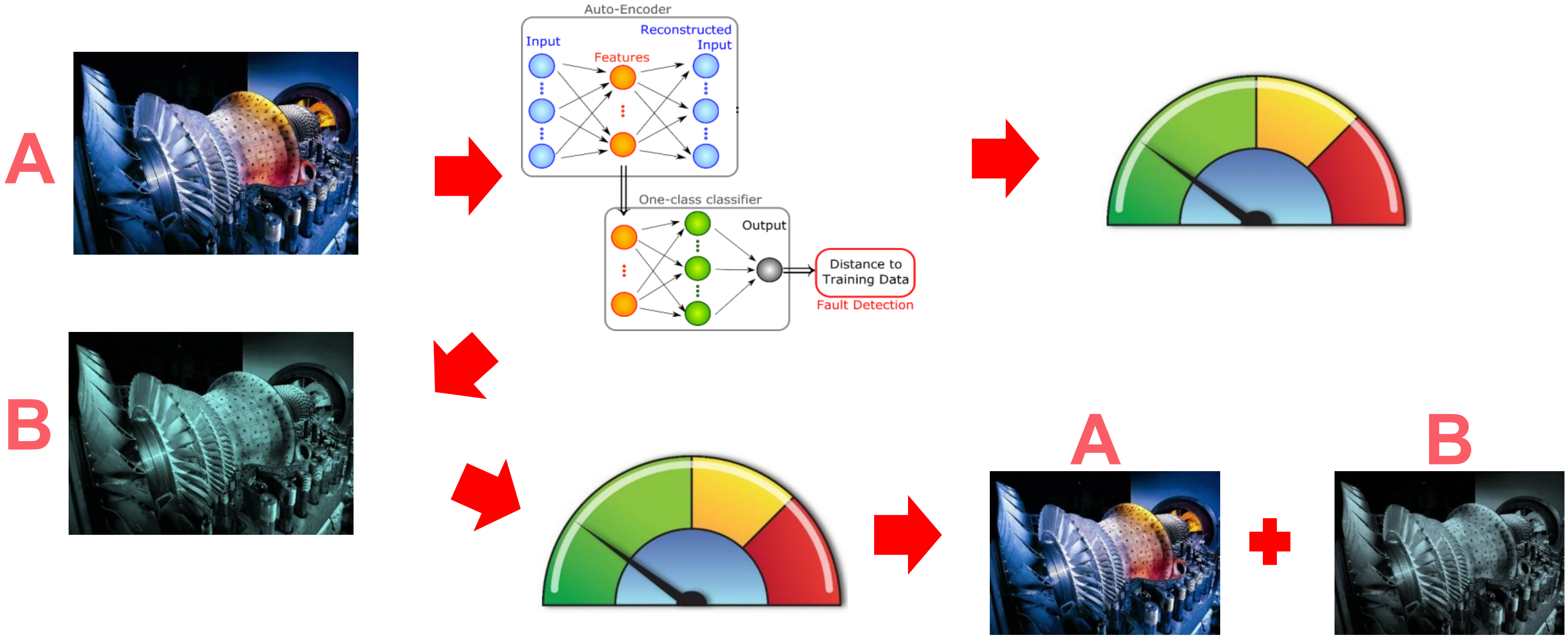
Transfer of experience with respect to the healthy operating conditions→ Enlarge the set of representative «healthy data»

Transfer the experience with respect to faulty system conditions

But: single units are operated differently, have different configurations and environmental conditions

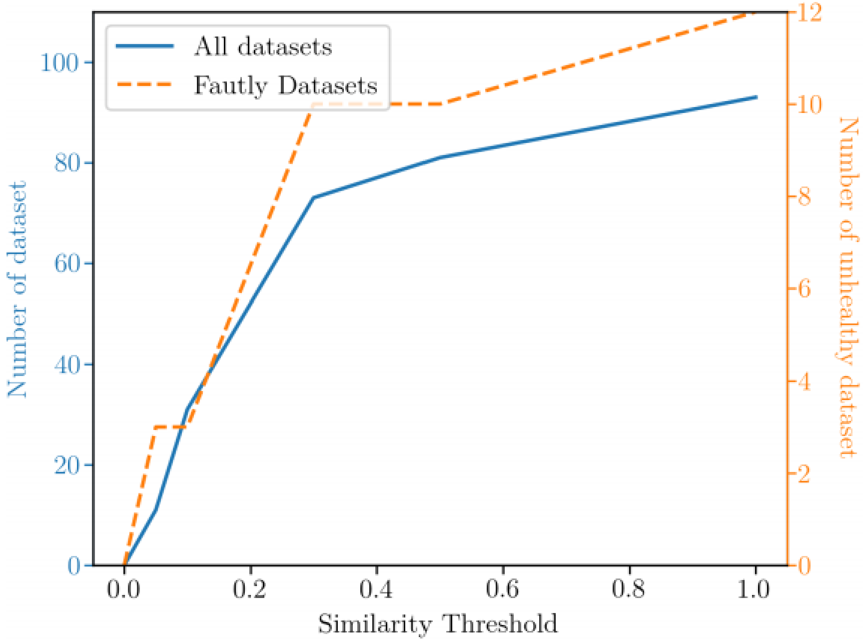
Challenge: If fleet units too similar→ no additional experience added
If fleet units too different→ faulty system conditions recognized as healthy

Using the fleet experience



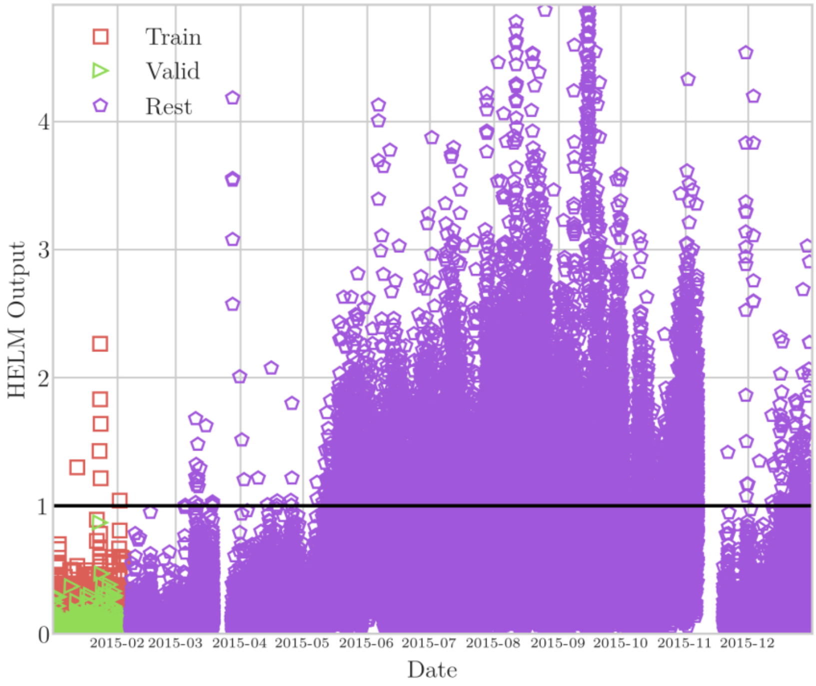
Michau, Gabriel, Thomas Palmé, and Olga Fink. 2018. “Fleet PHM for Critical Systems: Bi-Level Deep Learning Approach for Fault Detection.” In European Prognognostics and Health Management Conference. Utrecht.

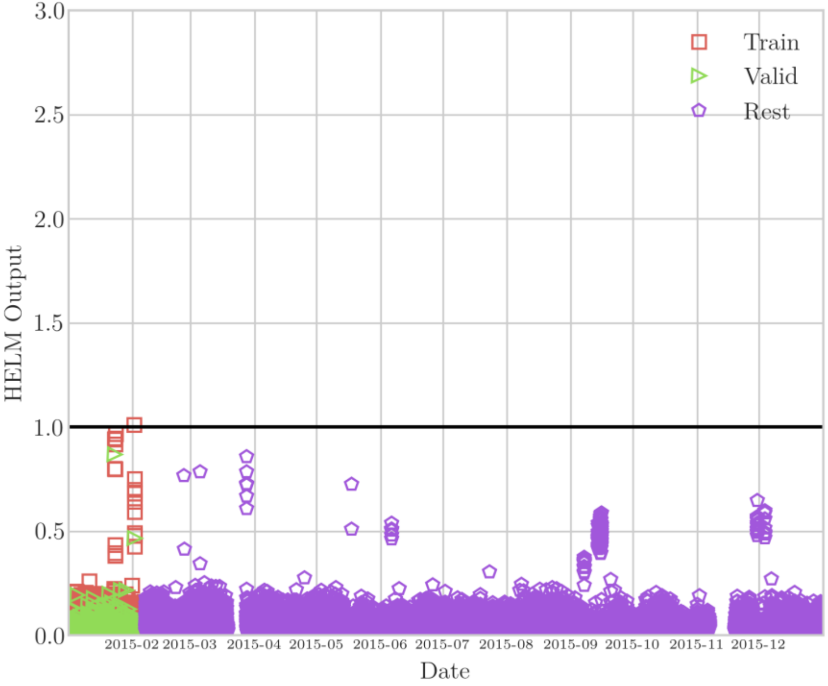
Dataset dissimilarities



Olga Fink

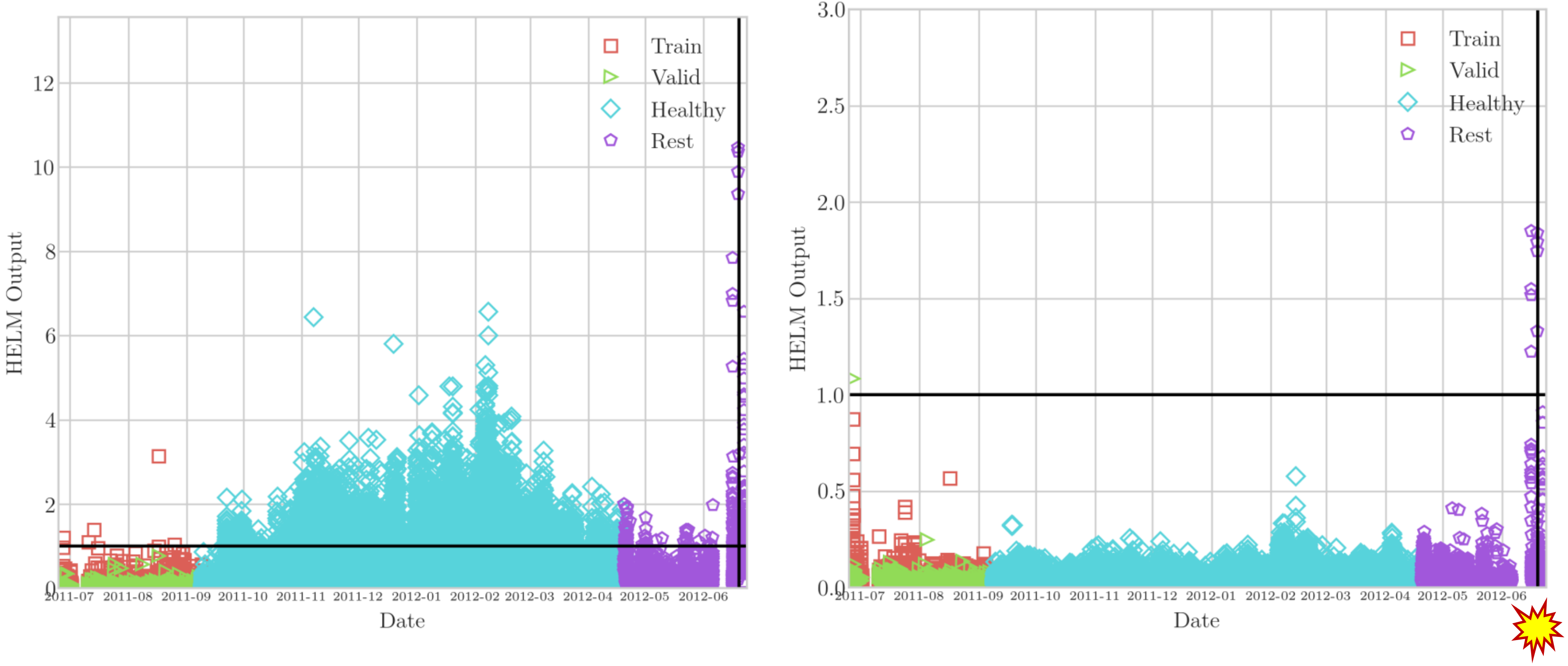
Fleet of Gas Turbines Plant 1: Healthy





Michau, Gabriel, Thomas Palmé, and Olga Fink. 2018. “Fleet PHM for Critical Systems: Bi-Level Deep Learning Approach for Fault Detection.” In European Prognognostics and Health Management Conference. Utrecht.

Fleet of Gas Turbines Plant 2: Fault



Michau, Gabriel, Thomas Palmé, and Olga Fink. 2018. “Fleet PHM for Critical Systems: Bi-Level Deep Learning Approach for Fault Detection.” In European Prognognostics and Health Management Conference. Utrecht.

Different options for fleet learning (4/4)

1. Perform domain alignment in the feature space of the different units to compensate for the distribution shift between different units of a fleet.
   → typically, pairwise transfer of models with source and target units
2. Apply the trained models to the target units.

Challenge:

- alignment is performed in an unsupervised way

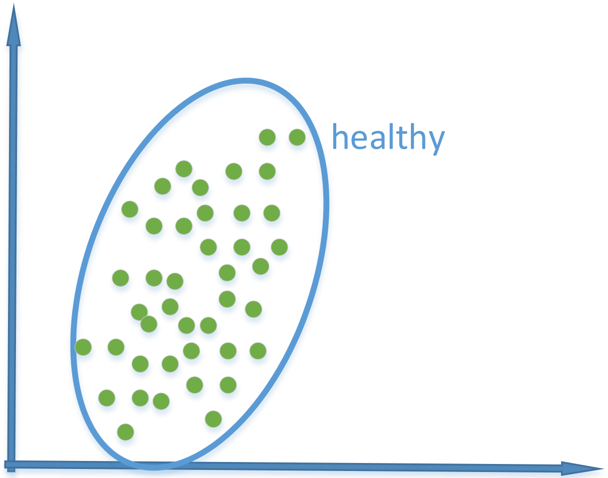
→ performance depends on the assumption that the future operating conditions of the unit of interest will be representative to the aligned operating conditions

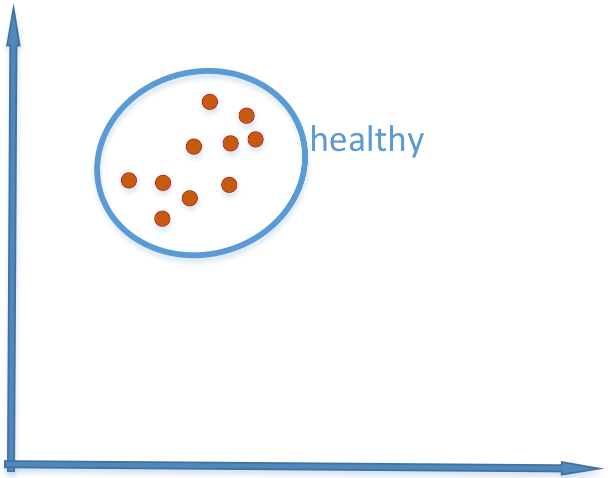
→ no guarantees can be made that the system of interest will be behaving in a similar way in the future

- (However, this limitation is in fact true for all the fleet PHM approaches since the past experience of other fleet units is transferred to the unit of interest.)

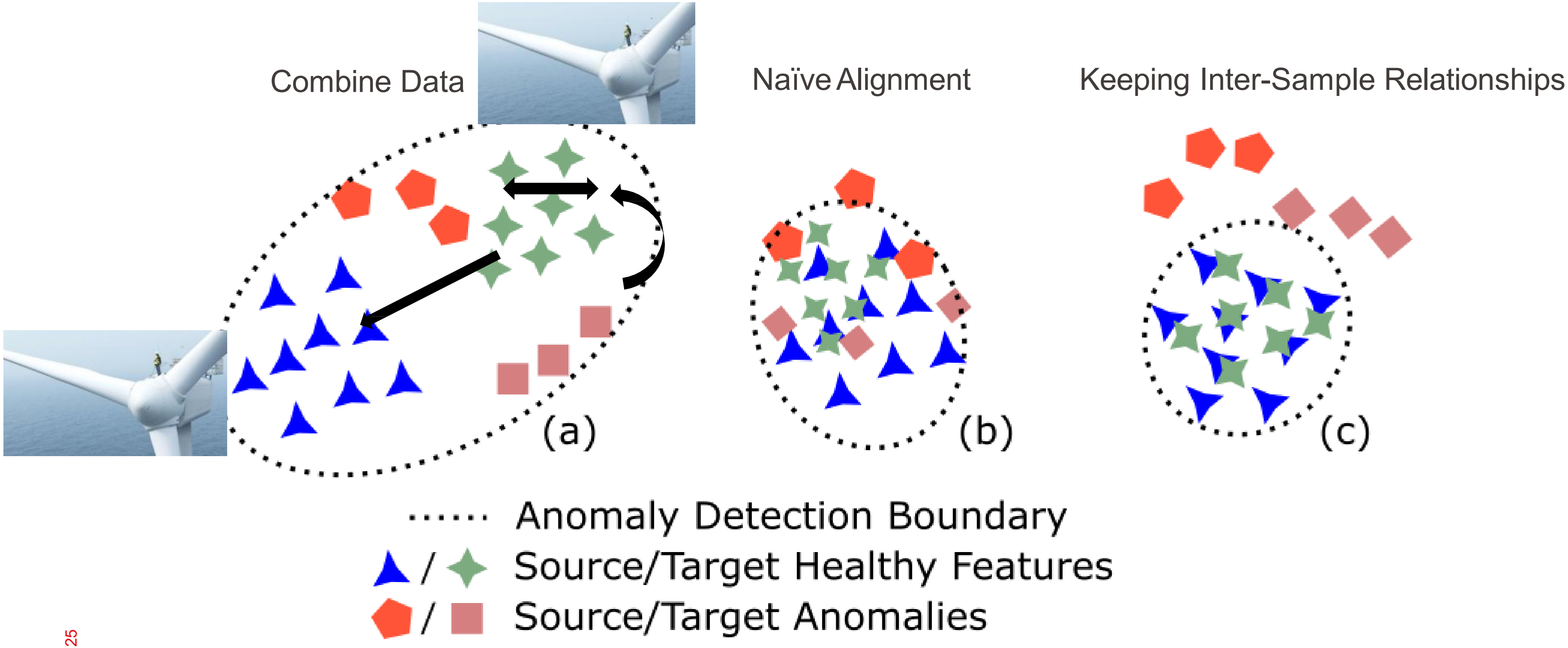
Only healthy data for source and target available

Source **Target**



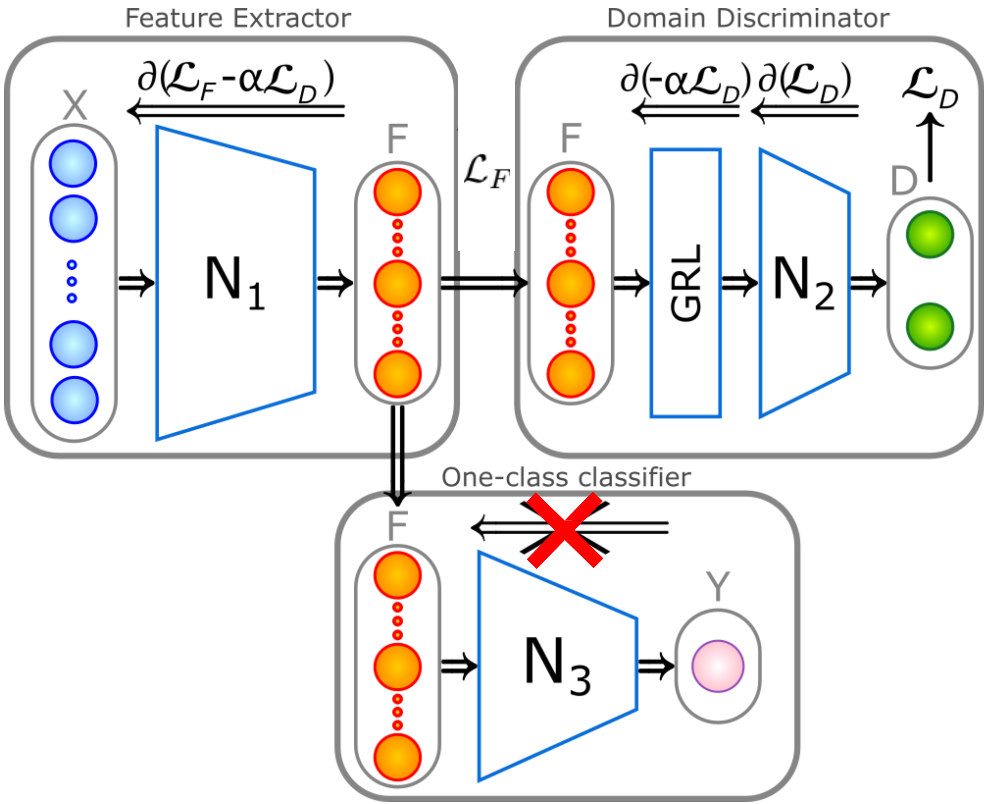


Transfer operational experience between units



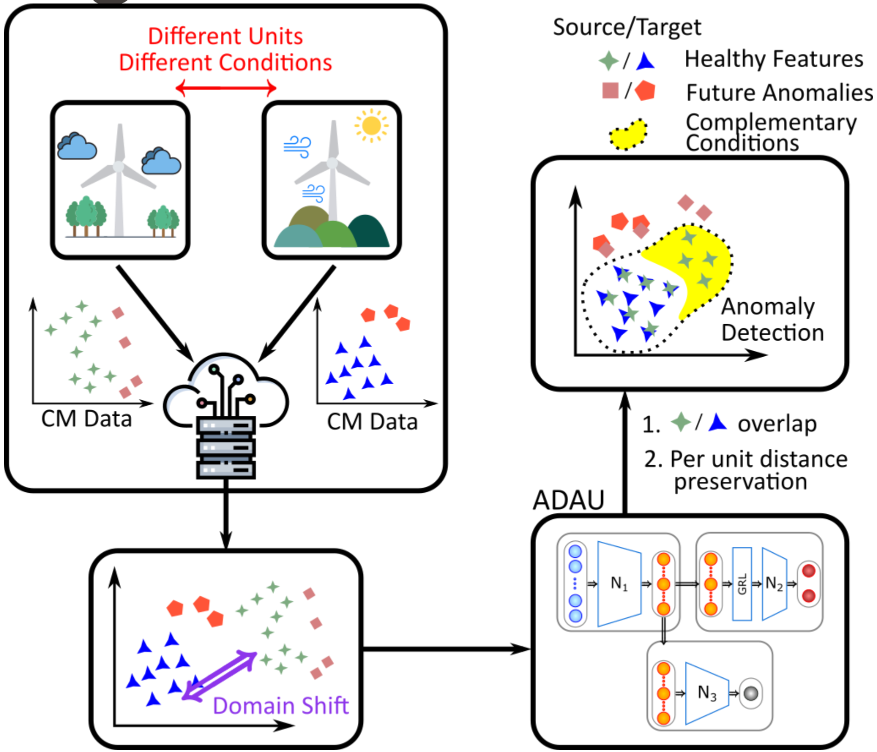
Michau, G. & O. Fink (2021): Unsupervised Transfer Learning for Anomaly Detection: Application to Complementary Operating Condition Transfer, Knowledge-Based Systems, 216, 106816

Domain Adaptation



Michau, G. & O. Fink (2021): Unsupervised Transfer Learning for Anomaly Detection: Application to Complementary Operating Condition Transfer, Knowledge-Based Systems, 216, 106816

Proposed Framework for Unsupervised Transfer Learning



Michau, G. & O. Fink (2021): Unsupervised Transfer Learning for Anomaly Detection: Application to Complementary Operating Condition Transfer, Knowledge-Based Systems, 216, 106816

Source to Target - Suggested Scaled Domain Adaptation

- Enforce “scaled” projection between input and latent space

$$\mathcal{L}_F = \sum_{S \in \left\{ \substack{\text{Source} \\ \text{Target}} \right\}} \frac{1}{|S|} \sum_{(i,j) \in S} \left\| \|X_i - X_j\|_2 - \eta \|F_i - F_j\|_2 \right\|_2 \qquad \eta = \underset{\tilde{\eta}}{\text{Argmin}}\, \mathcal{L}_F(\tilde{\eta})$$

Michau, G. & O. Fink (2021): Unsupervised Transfer Learning for Anomaly Detection: Application to Complementary Operating Condition Transfer, Knowledge-Based Systems, 216, 106816

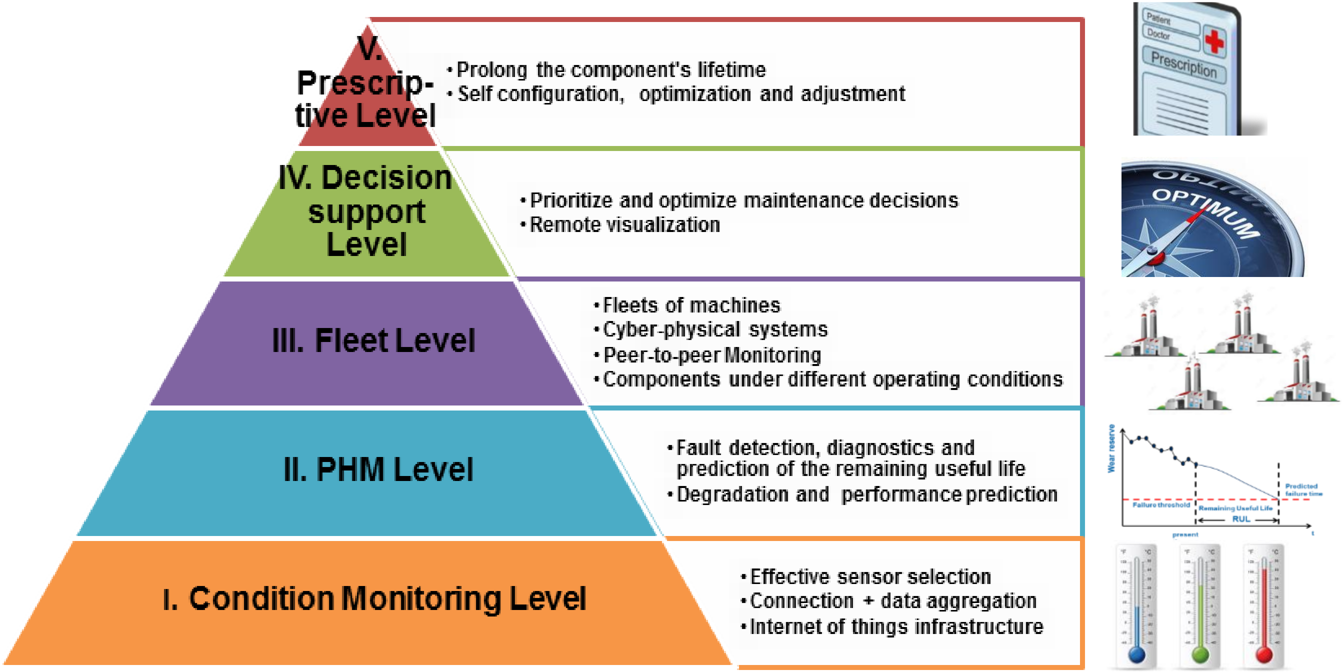
Number of matched units

Unit	HELM	β-VAE	β-VAEs	β-VAEw	HFA	AFAs	AFAw	HAFAs	HAFAw
1	11	74	65	74	83	83	85	86	79
2	0	5	20	13	10	13	24	5	12
3	10	28	22	22	21	23	30	32	34
4	17	30	21	32	54	55	54	52	49
5	94	68	47	67	90	59	63	80	85
6	92	51	68	63	85	77	79	92	93
7	0	13	29	24	29	45	31	34	26
8	95	40	42	43	67	61	63	65	58
9	2	19	19	18	26	28	32	22	39
10	1	18	15	8	21	28	24	34	29
11	2	20	35	47	59	63	51	60	51
12	0	3	3	4	2	2	1	1	3
R% (5%)	27.3	31.1	32.5	34.9	46.0	45.2	45.2	47.4	47.0
R% (1%)	13.5	20.6	22.0	25.8	30.5	27.0	25.5	32.8	30.1

G.Michau & O. Fink, 2019 "Domain Adaptation for One-Class Classification: Monitoring the Health of Critical Systems Under Limited Information." International Journal of Prognostics and Health Management

Prescriptive Maintenance/ Operation (health-aware control)

Five levels of condition-based and predictive maintenance



Predictive vs. Prescriptive Maintenance /Presrciptive Operation

Predict the remaining useful life

Anticipate the failure

Reduce the impact of the failure

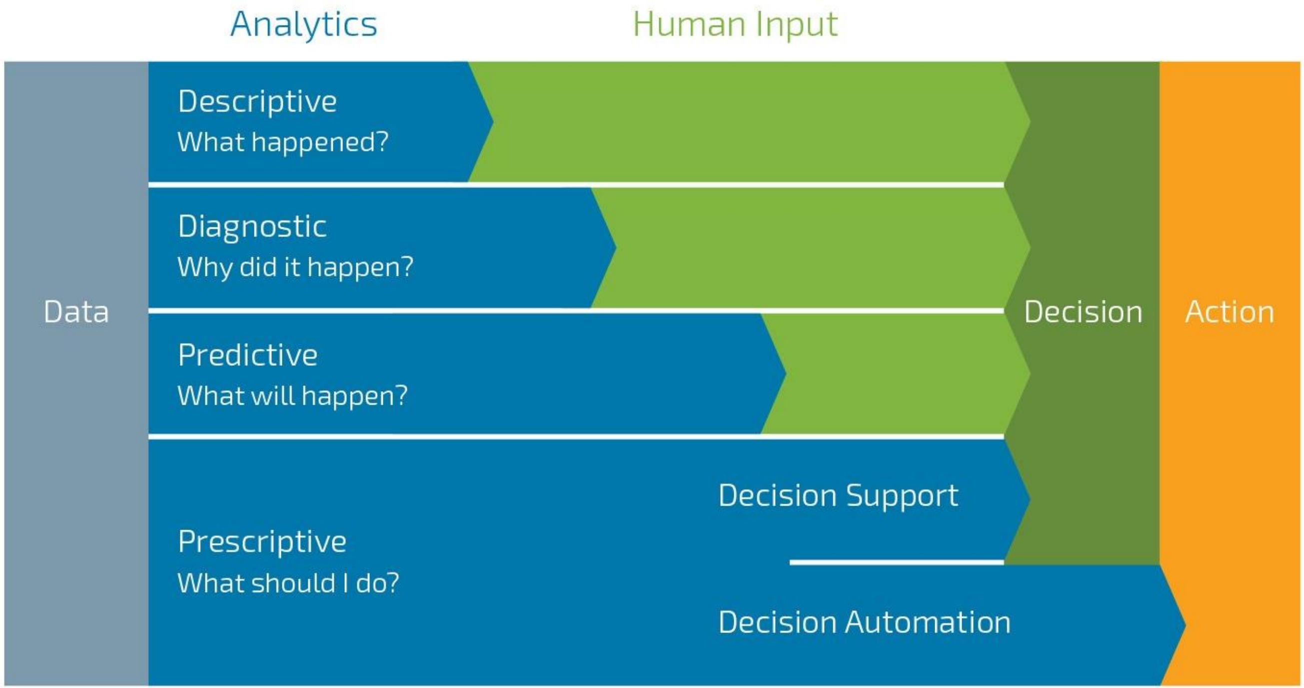
Determine the optimal point in time for maintenance intervention

What can we do to prolong the remaining useful life?

How can we proactively adjust the operating conditions?

How can we control the process parameters?

Concept of prescriptive analytics

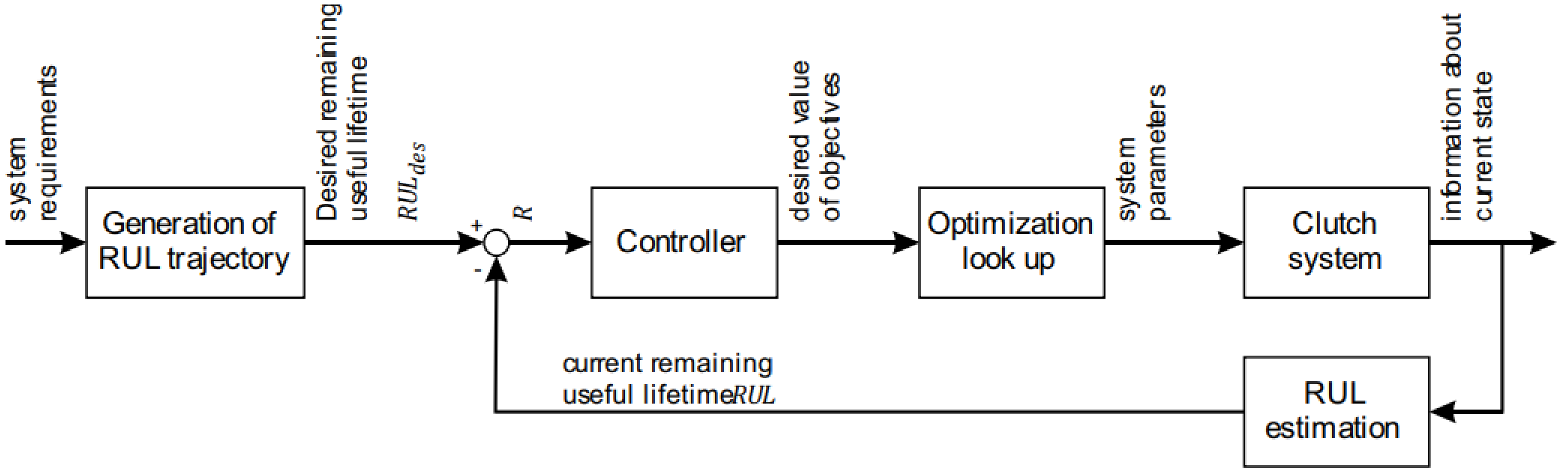


Source Gartner

Prescriptive Maintenance / Operation

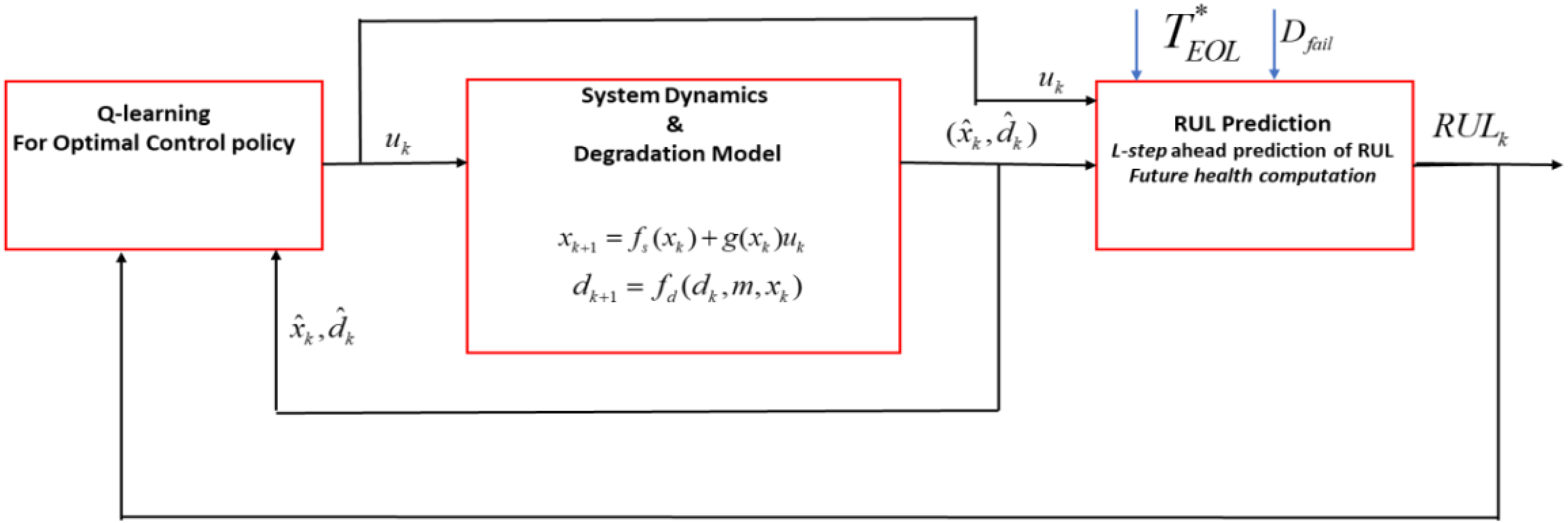
- Prescribe **sequentially** optimal operational parameters that have an impact on the remaining useful lifetime in order to achieve a certain goal (e.g. finish a mission, finish a production goal, extend the time interval to the planned maintenance intervention etc.)
- The set of the operational parameters are optimized based on their **impact on the consumption of the remaining useful lifetime**
- A good prognostics model required
- A good understanding required of what influences the remaining useful lifetime

Controlling the Remaining Useful Lifetime using Self-Optimization



Meyer, Tobias, et al. "Controlling the remaining useful lifetime using self-optimization." *Chemical Engineering Transactions* 33 (2013).

A Reinforcement Learning Approach to Health Aware Control Strategy

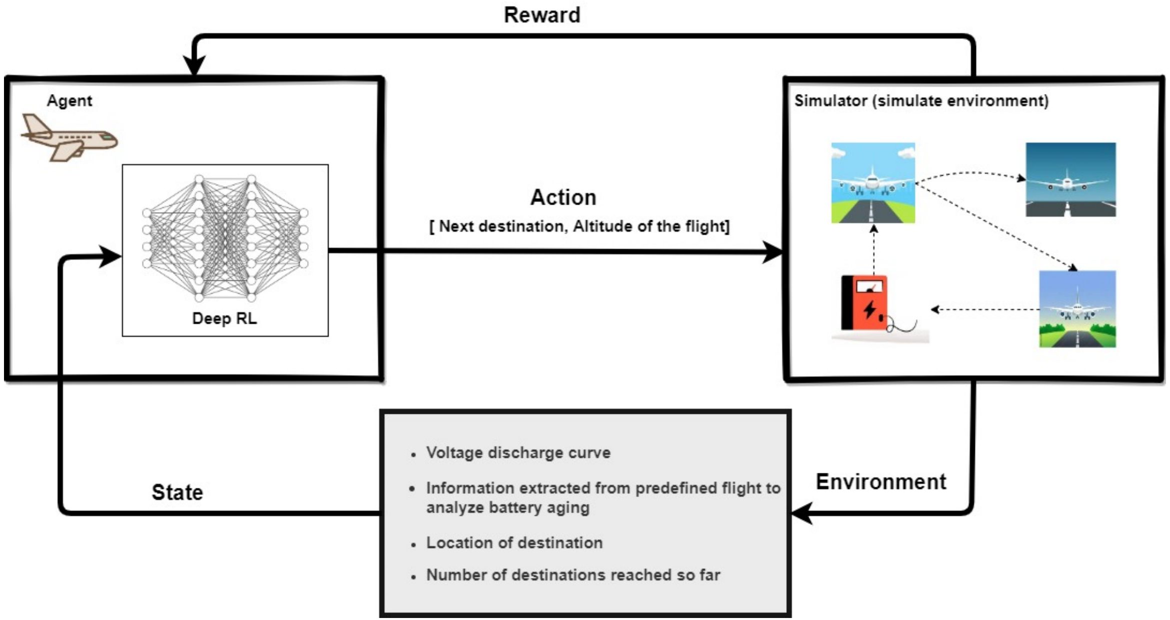


Jha, Mayank S., et al. "A reinforcement learning approach to health aware control strategy." *2019 27th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2019.

Prescriptive operation

How can adapt the operating parameters of a drone based on the health state of the battery, the current environmental conditions+ planned mission proactively → prescribe optimal parameters!

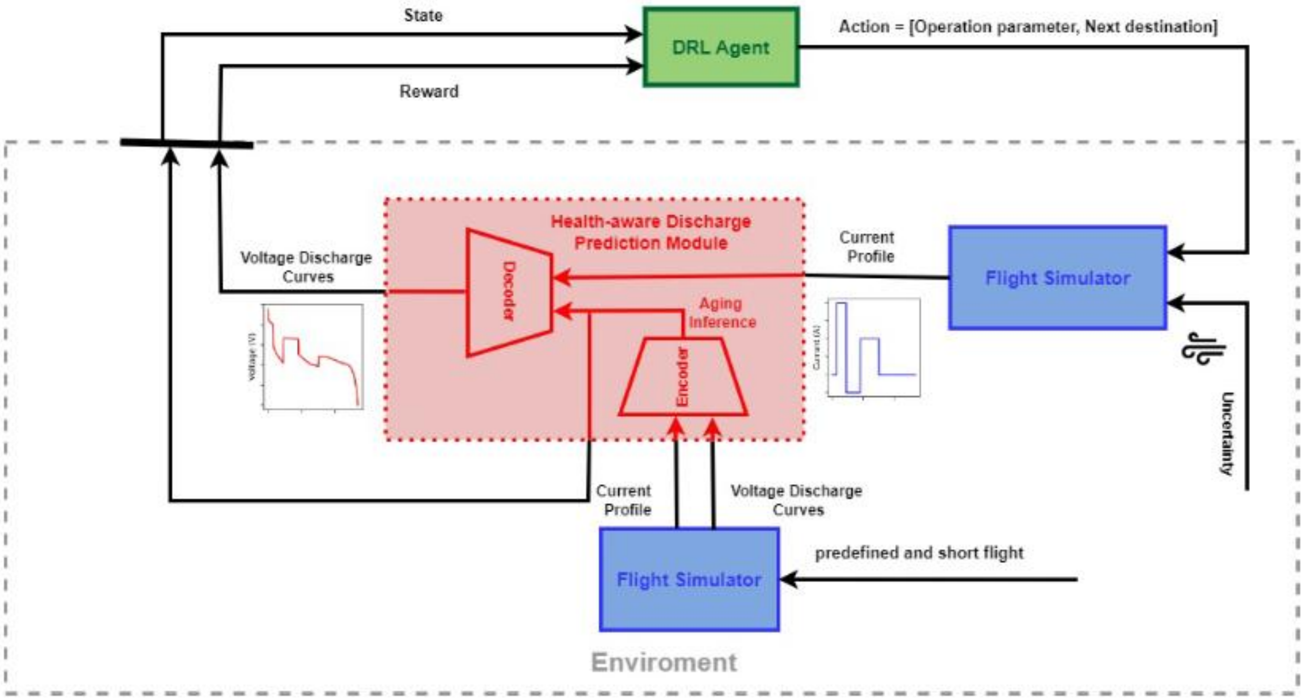
Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning



- Voltage discharge curve
- Information extracted from predefined flight to analyze battery aging
- Location of destination
- Number of destinations reached so far

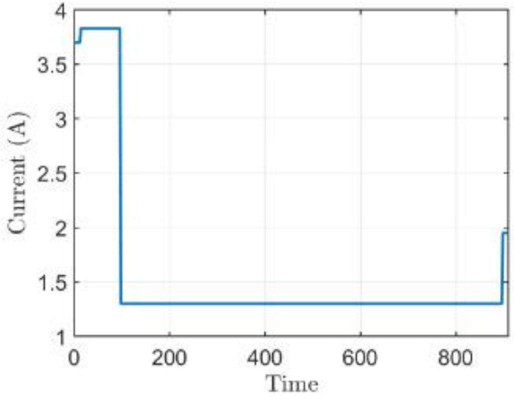
Montazeri, M., Kulkarni, C., & Fink, O.: Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning, in preparation

Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning

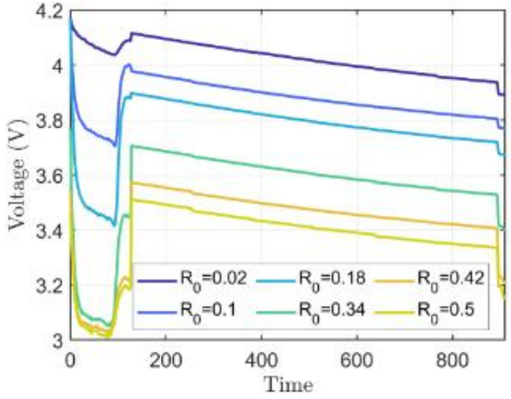


Montazeri, M., Kulkarni, C., & Fink, O.: Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning, in preparation

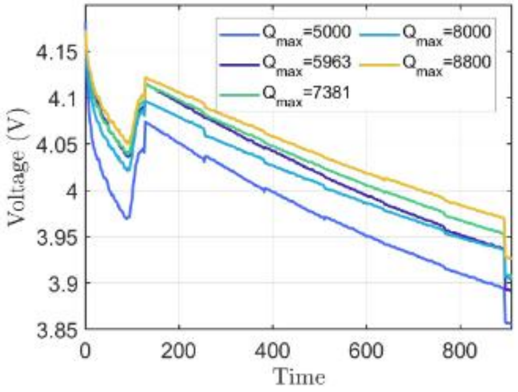
Effect of varying degradation parameters on the voltage discharge curves



(a) Current profile for specific flight

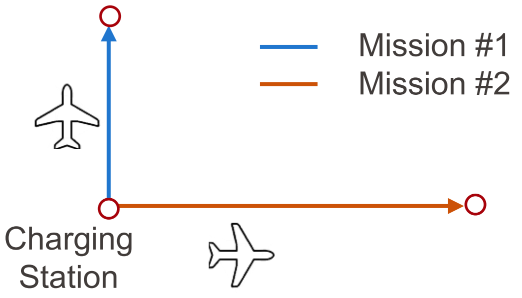


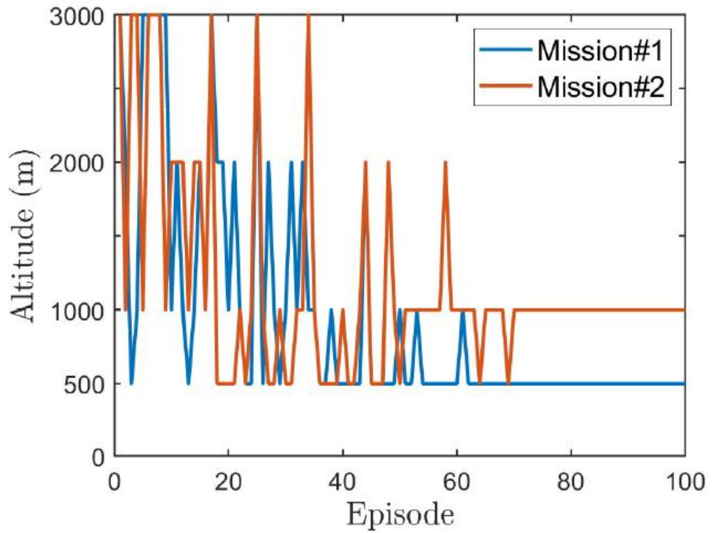
(b) Varying R_0 and keeping q_{max} fixed



(c) Varying q_{max} and keeping R_0 fixed

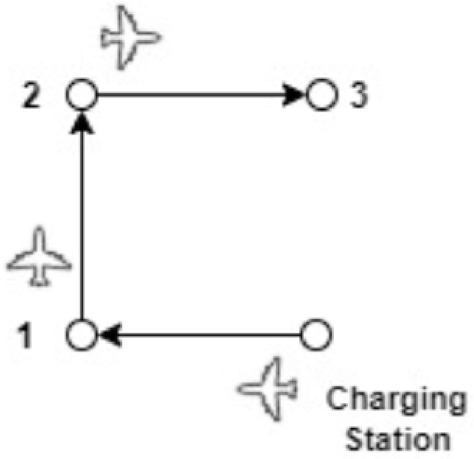
Single mission → altitude optimization



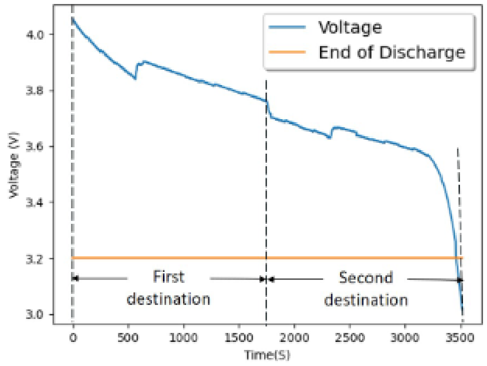


Montazeri, M., Kulkarni, C., & Fink, O.: Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning, under review

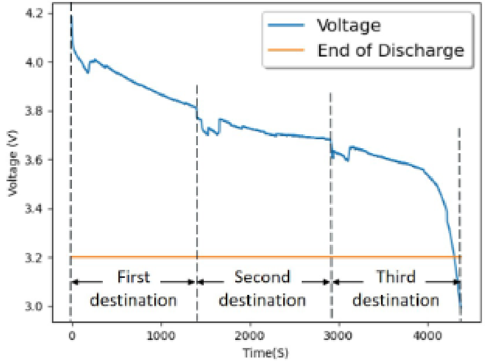
Impact of flight altitude on voltage discharge curve and the number of reached destinations (before EOD)



(a) Map of an example mission



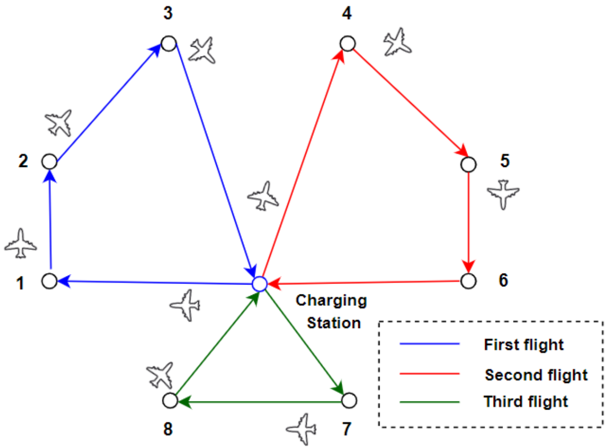
(b) Voltage discharge curve for cruise altitude=3000 m

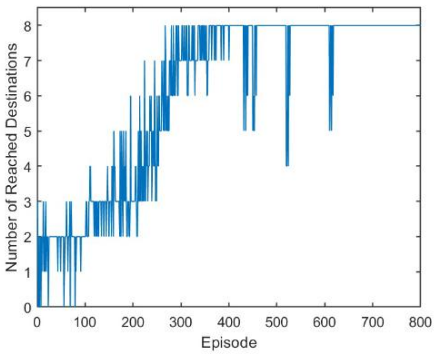


(c) Voltage discharge curve for cruise altitude=1000 m

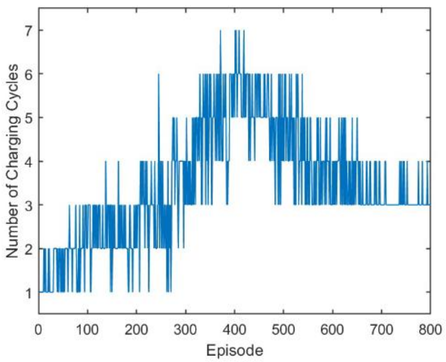
Montazeri, M., Kulkarni, C., & Fink, O.: Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning, under review

Number of reached destinations and charging cycles





(a) Number of reached destinations in each episode



(b) Number of charging cycles in each episode

Montazeri, M., Kulkarni, C., & Fink, O.: Prescribing Optimal Operation for Urban Air Mobility Using Deep Reinforcement Learning, in preparation