Data-Driven Survival Modeling for Predictive Maintenance

Mattias Krysander (mattias.krysander@liu.se)

12th Scandinavian Conference on SYSTEM & SOFTWARE SAFETY

Lindholmen, Göteborg, Sweden

2024-11-20

LINKÖPING UNIVERSITY

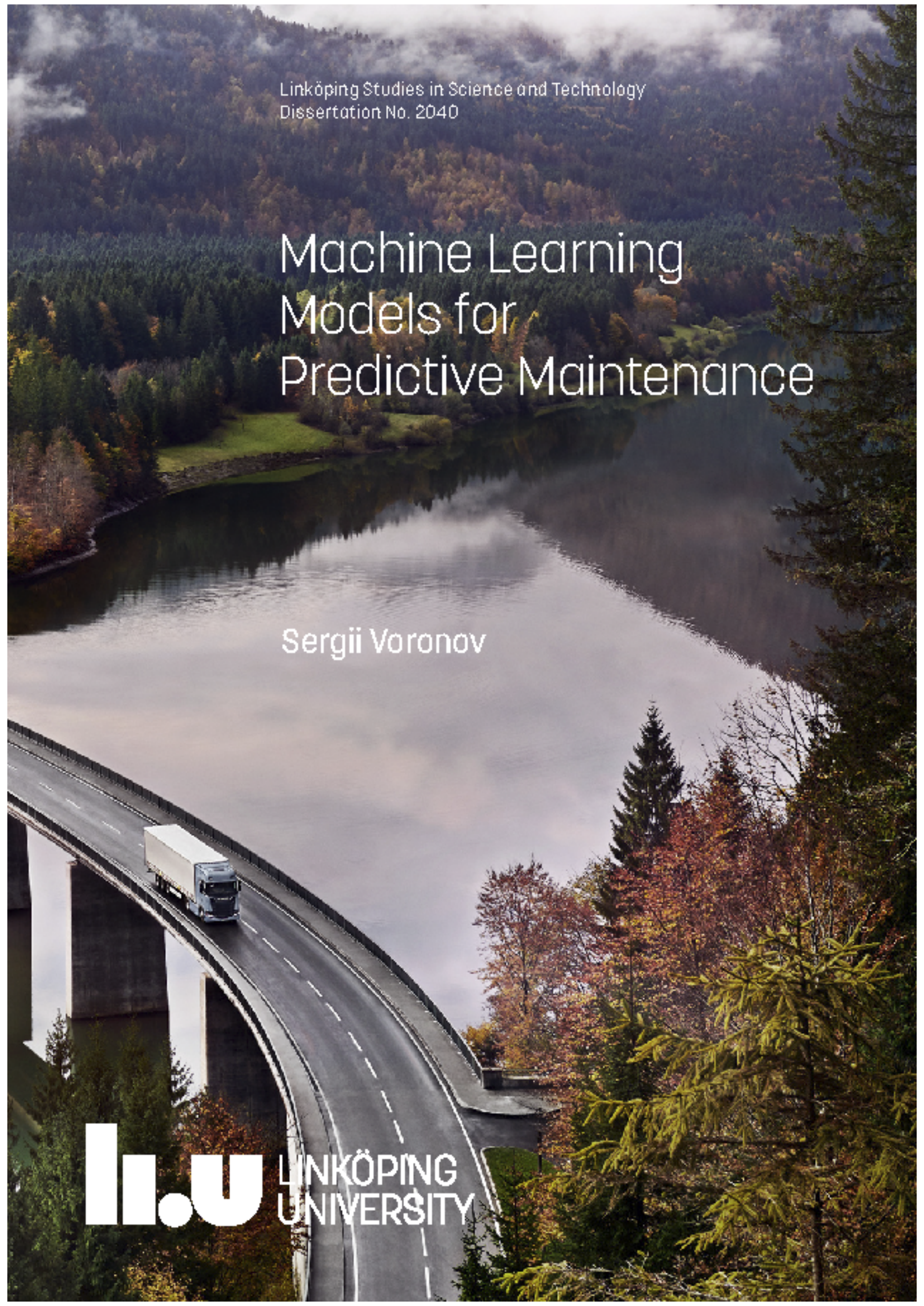
Collaborative Research Effort Since 2014

- Ongoing research started in 2014.
- Collaboration with Scania, KTH, Stockholm University, and Linköpings University

Erik Frisk

Mattias Krysander

Olov Holmer



Sergii Voronov PhD Thesis 2020

LINKÖPINGS UNIVERSITET

KTH VETENSKAP OCH KONST

Stockholms universitet

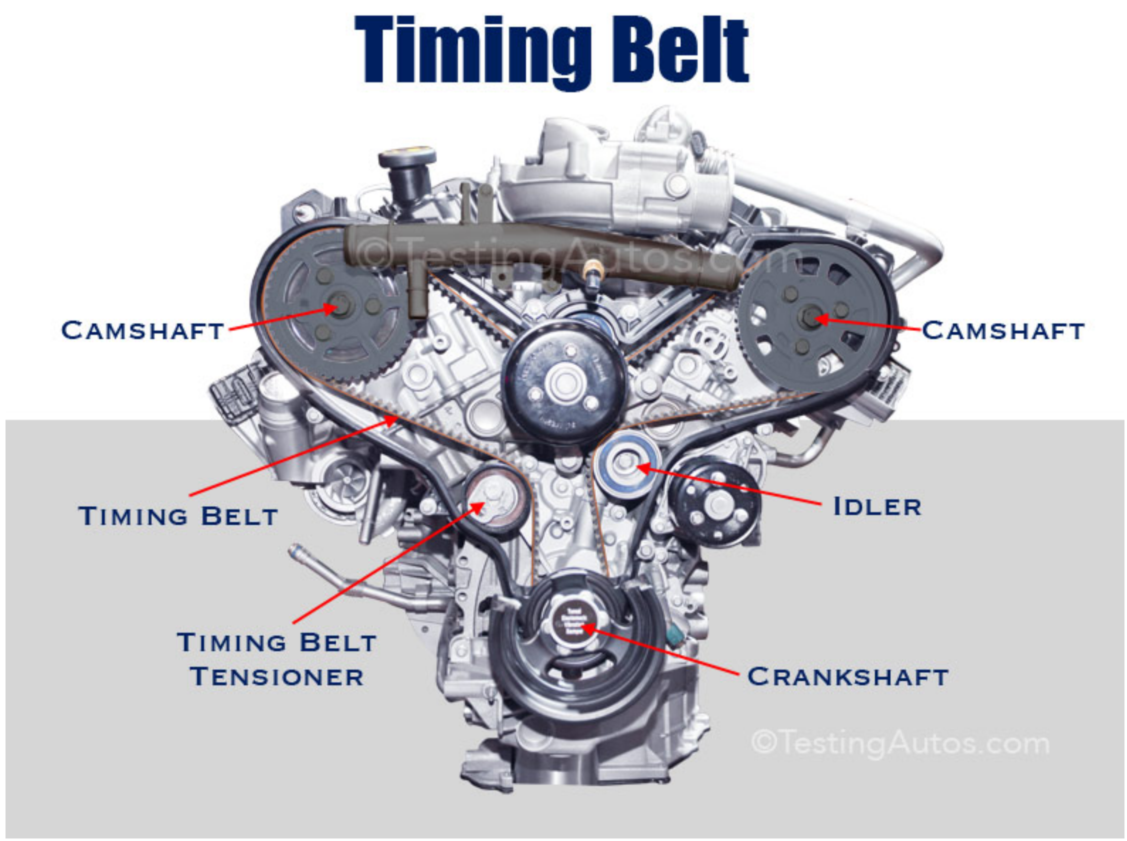
SCANIA

Maintenance Philosophies

Reactive/corrective

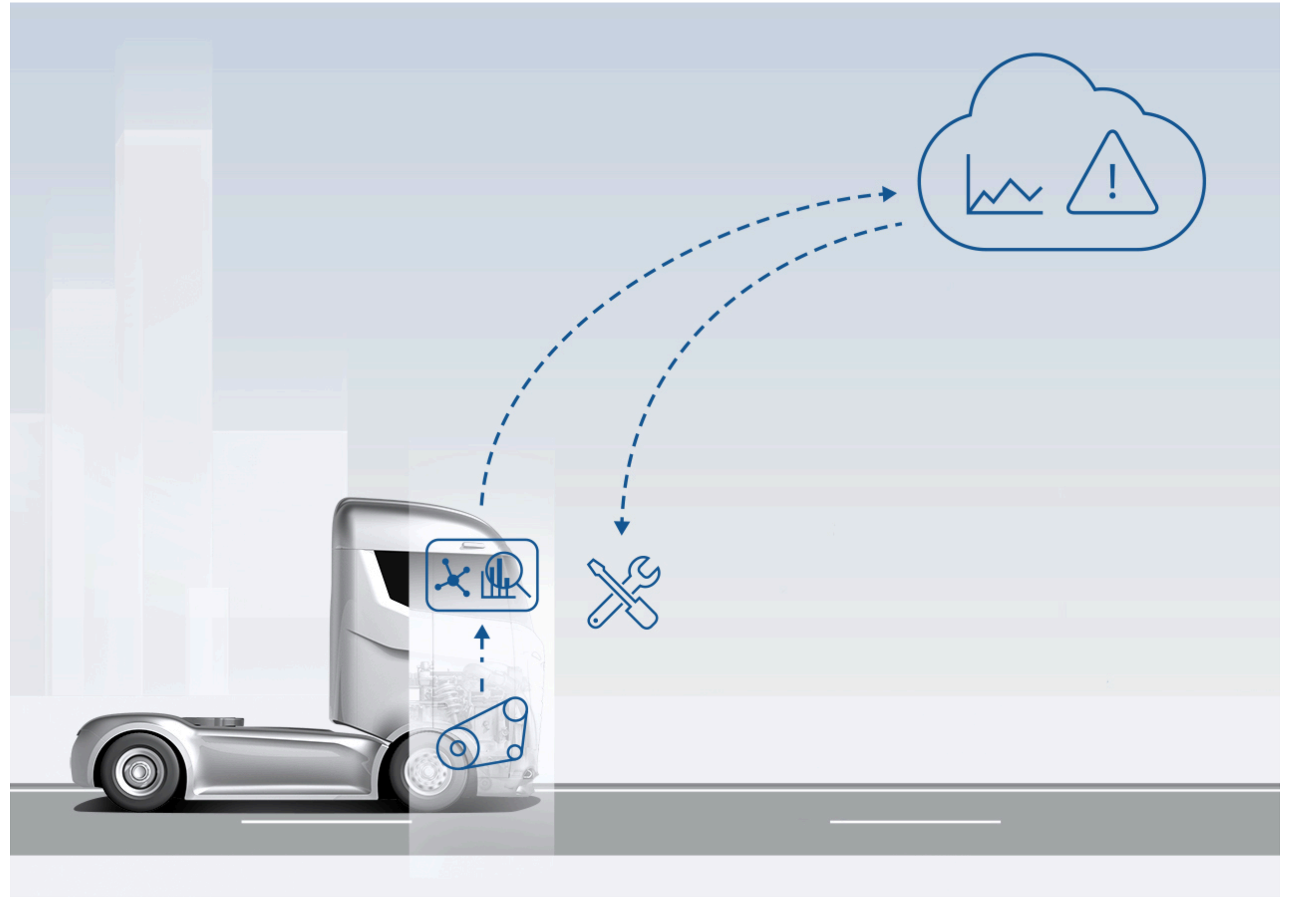
Fix it when it breaks

Preventive/scheduled



Maintain it at regular intervals to prevent breakdowns

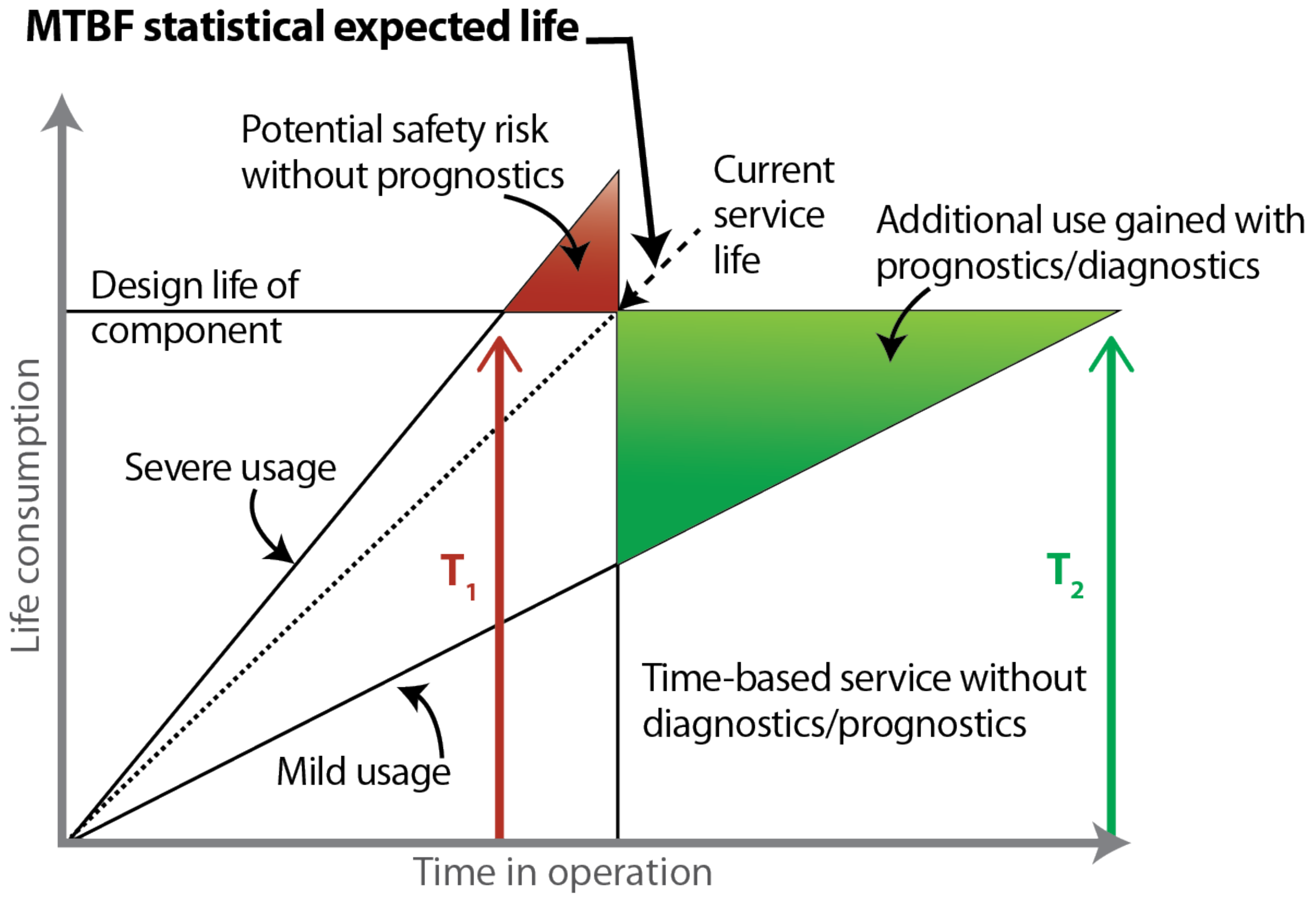
Predictive/
Condition Based Maintenance



Predict when it breaks and maintain it accordingly

Key Advantages of Predictive Maintenance

- Cost Savings:
  - Reduce unnecessary maintenance
  - Utilize component life effectively
- Operational Efficiency:
  - Minimizes unplanned stops
  - Reduce downtime
- Risk Reduction:
  - Lowers the risk of catastrophic failures by addressing issues before they escalate.



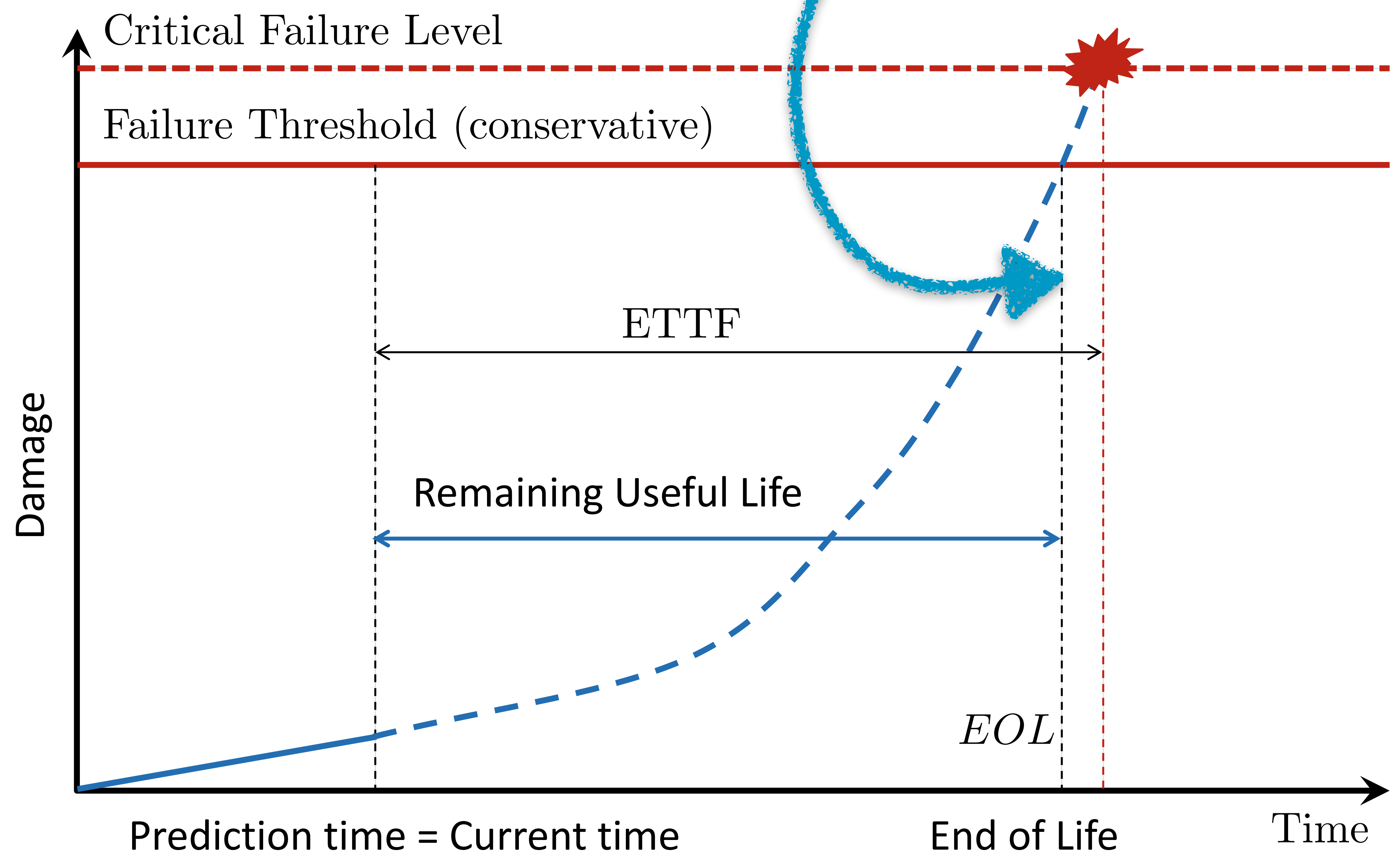
Source: Economic and Safety Benefits of Diagnostics & Prognostics (Romero et al. 1996)

Accurate end-of-life prediction of components is essential for successful predictive maintenance!

Data-driven maintenance prediction

Replacement

- To be able to predict end-of-life, track a degradation process (not just a single random event directly causing failure)
- maintenance = f(vehicle at time t)
- The health state is often not directly measurable, but sometimes possible to estimate from data
- A key difficulty: Uncertainties!



Often difficult

Physical wear and aging models

Alternative

Historical Data

Non-physical data-based models

Use Case: Heavy-duty Truck Batteries

- Trucks must maintain high availability to ensure smooth transport operations.
- Unexpected breakdowns on the road can lead to major disruptions, including:
  - Costly repairs
  - Delay deliveries with potential penalty fees
  - Cargo damage resulting in financial losses
  - Operational disruptions affecting work schedules and logistics planning

- Lead-acid starter battery issues are common causes of unplanned stops:
  - Battery's Role: Powers the starter motor for the diesel engine and auxiliary units (e.g., heating, kitchen appliances).
  - Usage Variability: Battery load varies based on usage scenarios, such as frequent stops for city trucks vs. extended operation for long-haul vehicles.

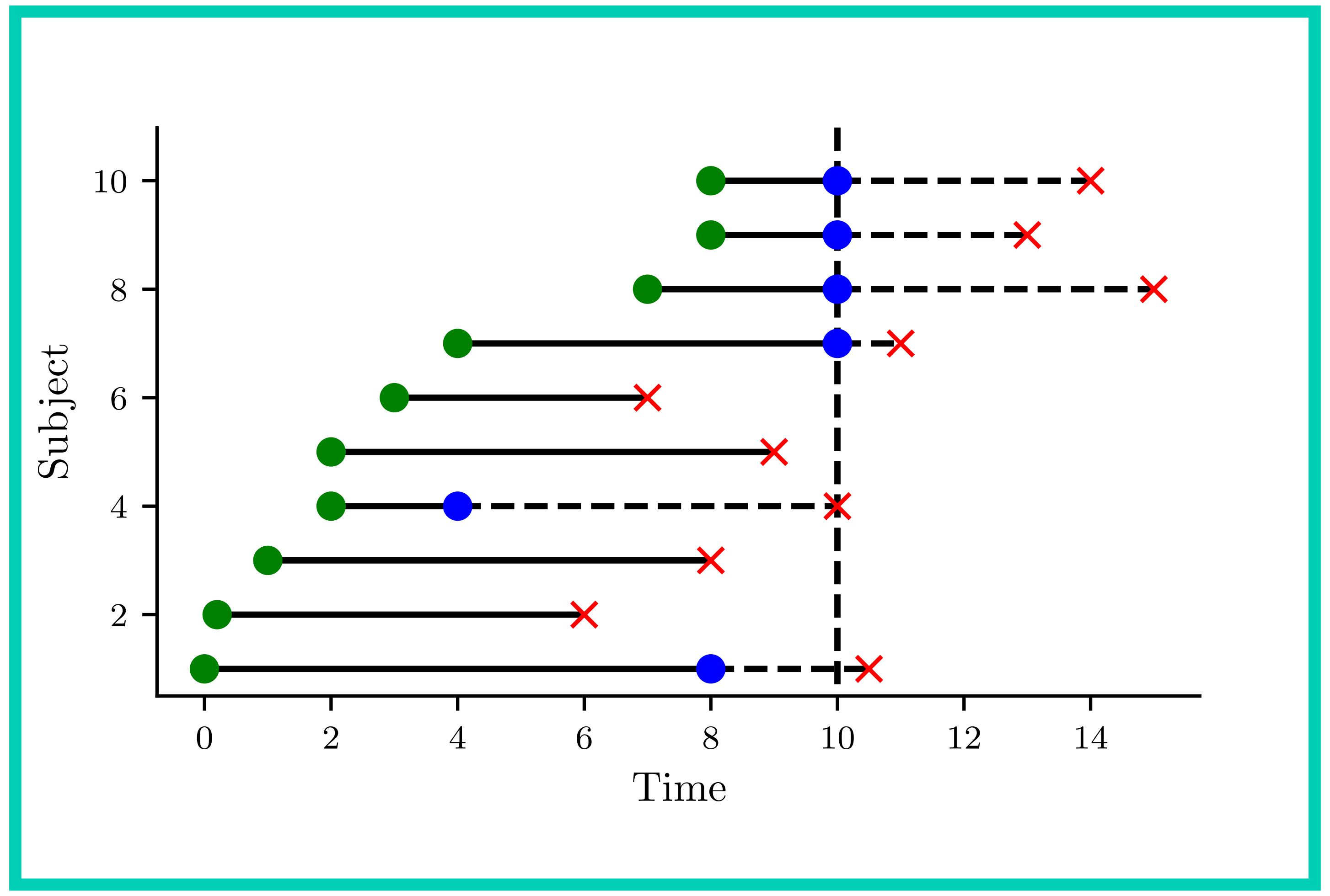
Challenges of Battery Prognostics

- Vehicle configurations: Different cabins, auxiliary systems, drivelines, etc
- Operational variations: Start-stops frequency, overnight cabin heating, …
- Environmental factors: Extreme temperatures accelerate battery wear.
- System dependencies: Degradation can result from other components, such as overcharging by the generator.

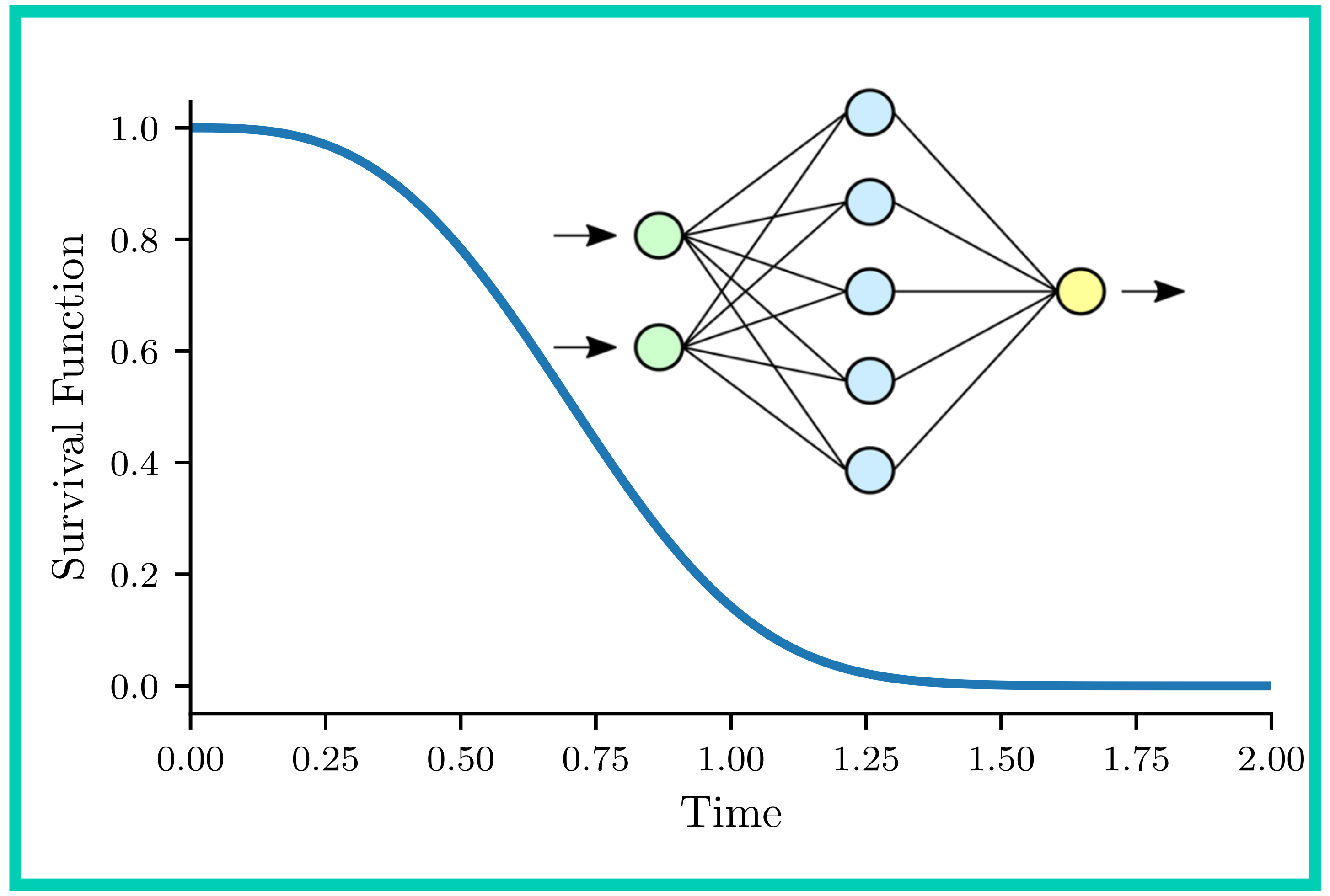
Demonstrate how data can be used to predict battery lifespan and optimize replacements

Data-Driven Survival Modeling for Predictive Maintenance

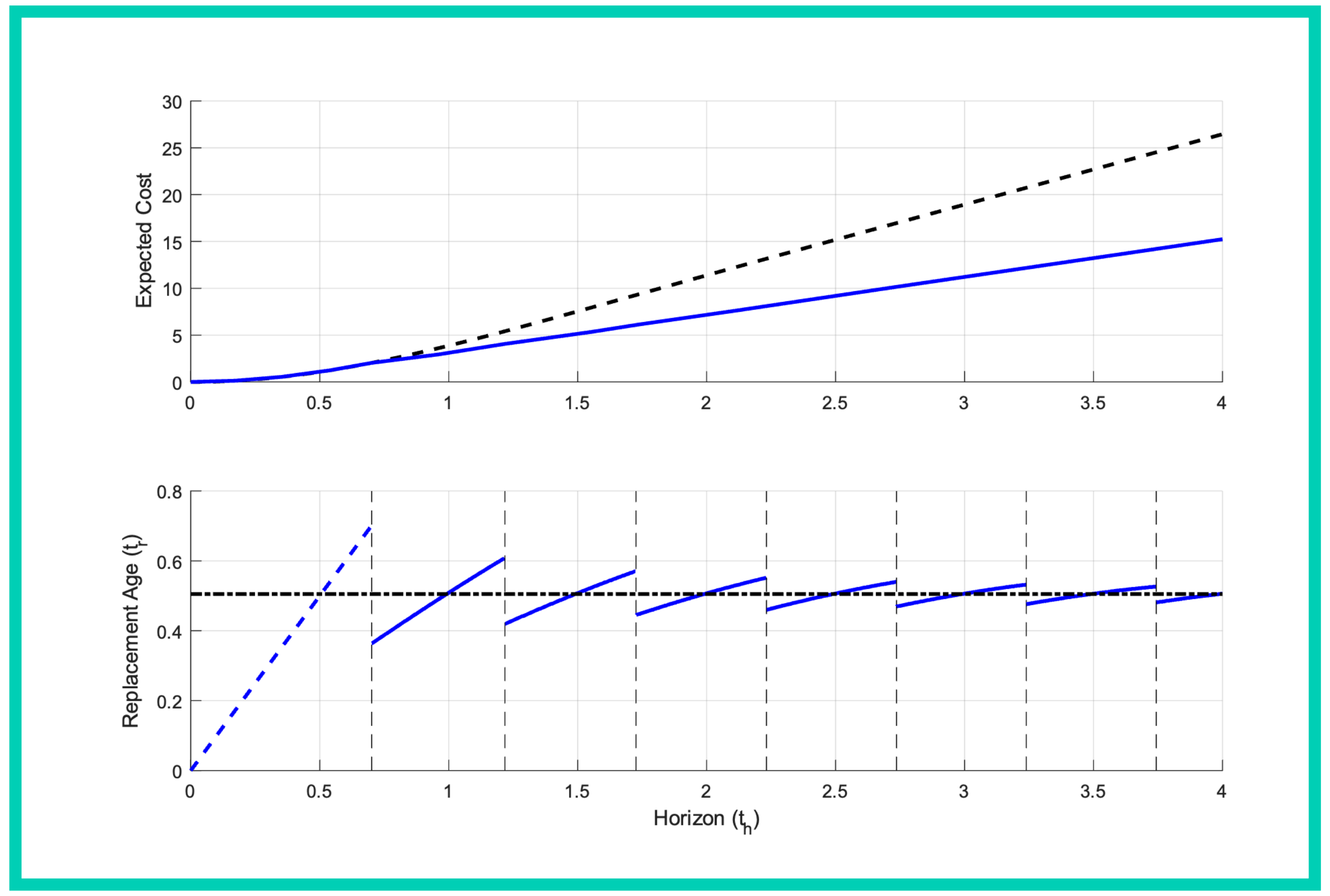
Data



Model

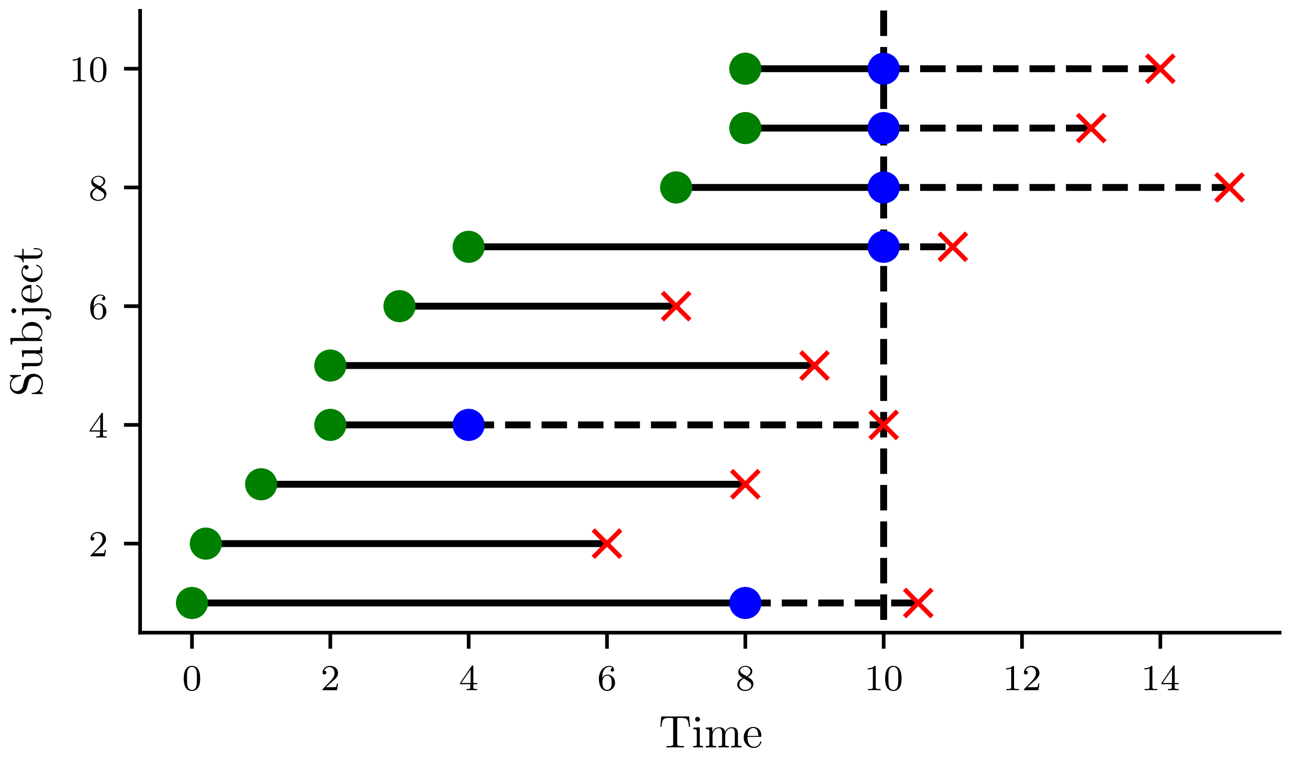


Individualized Maintenance Policy

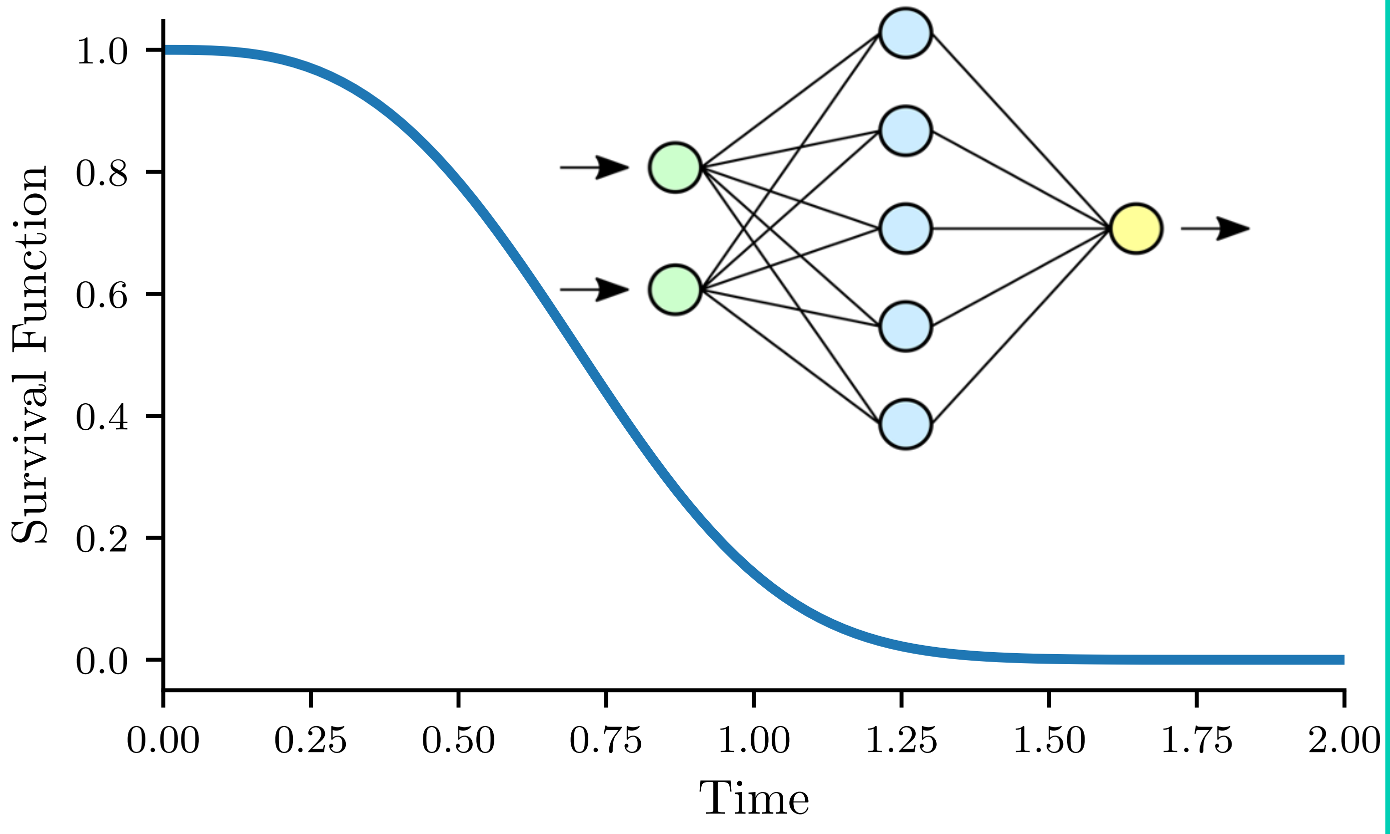


Data Description and Information Content

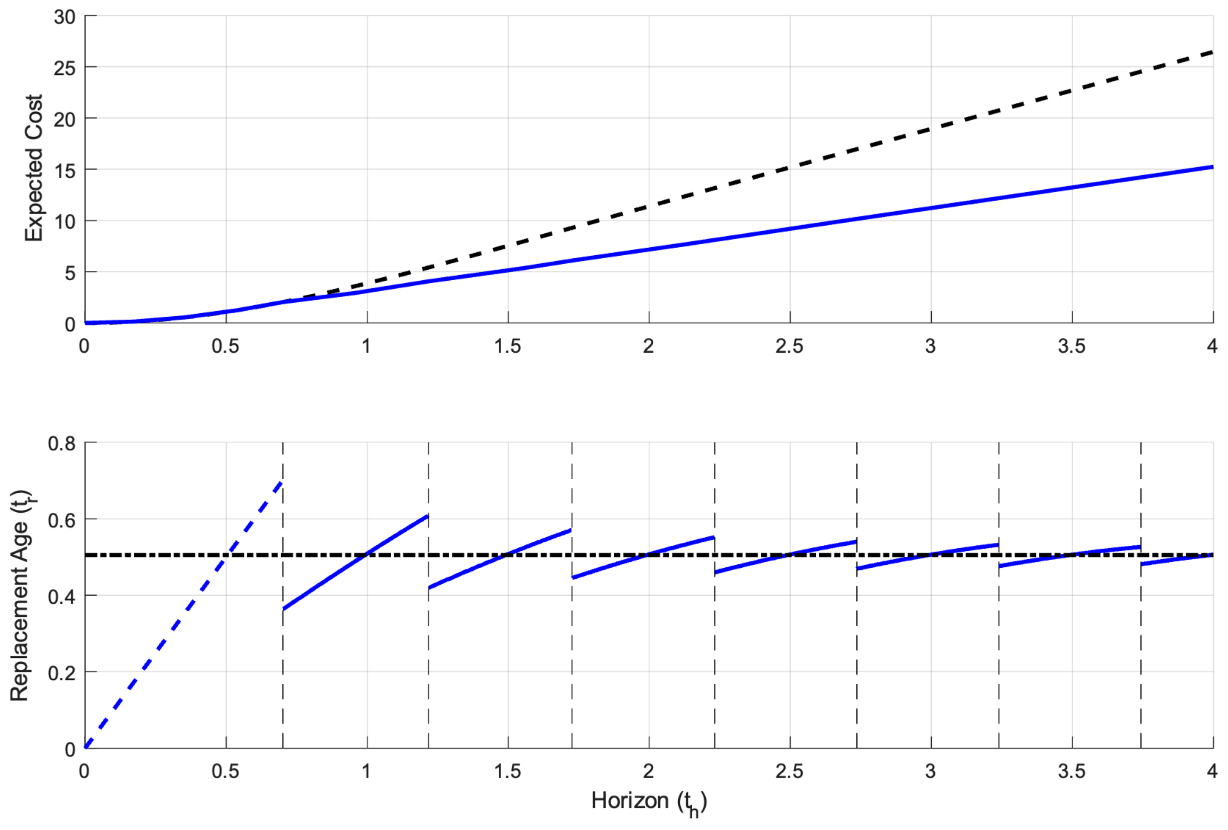
Data



Model

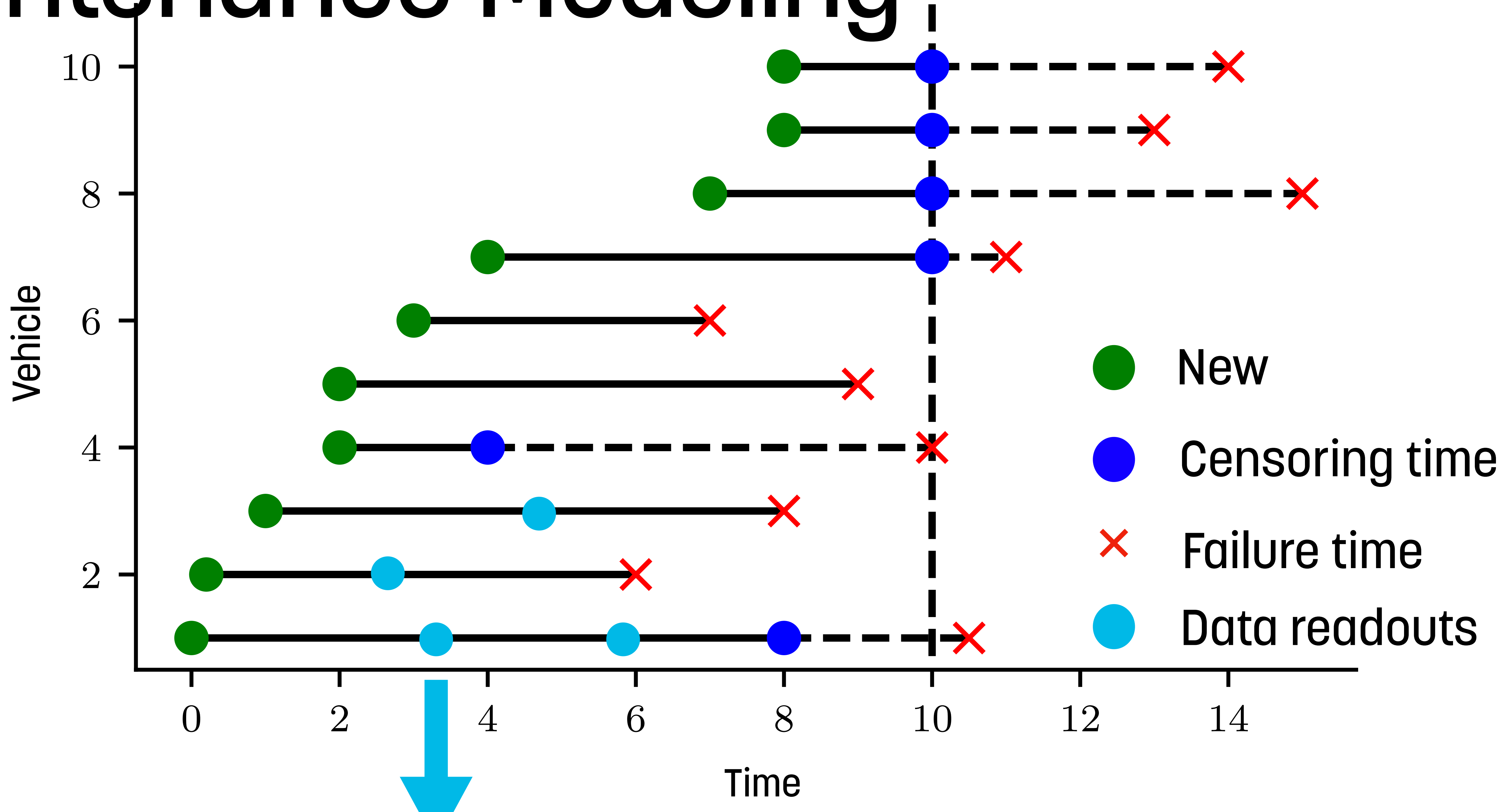


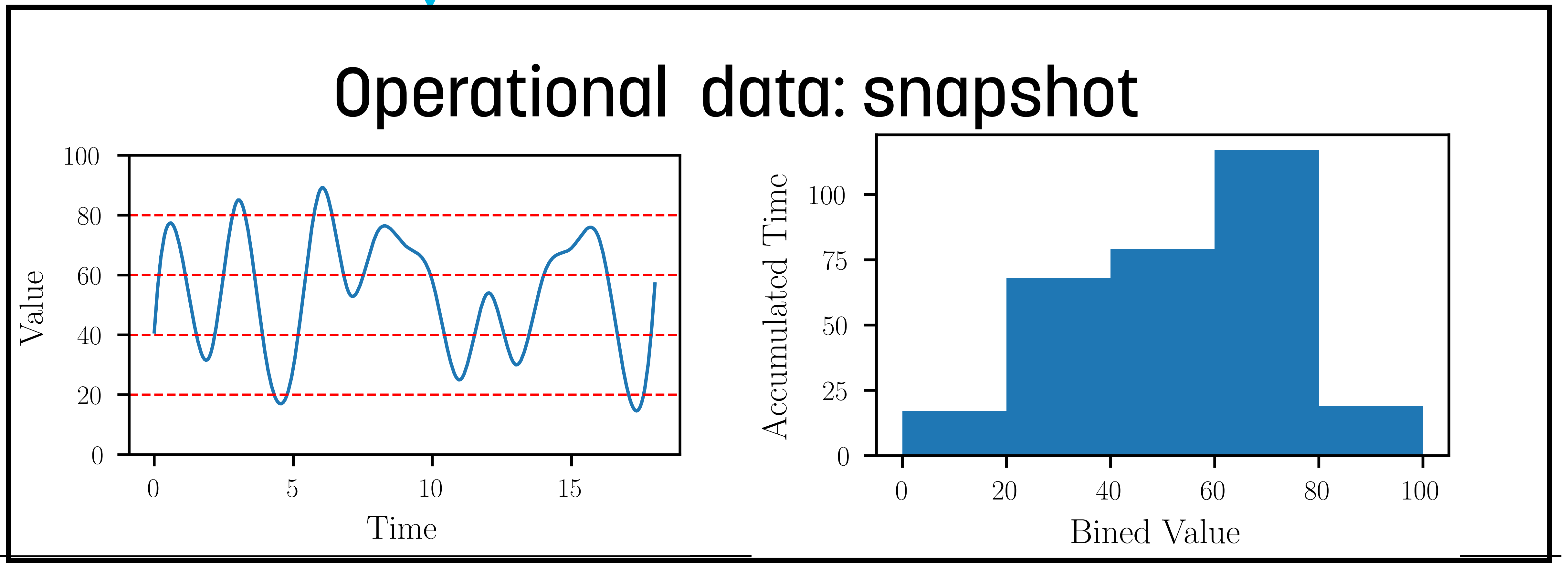
Individualized Maintenance Policy



Data Overview for Battery Maintenance Modeling

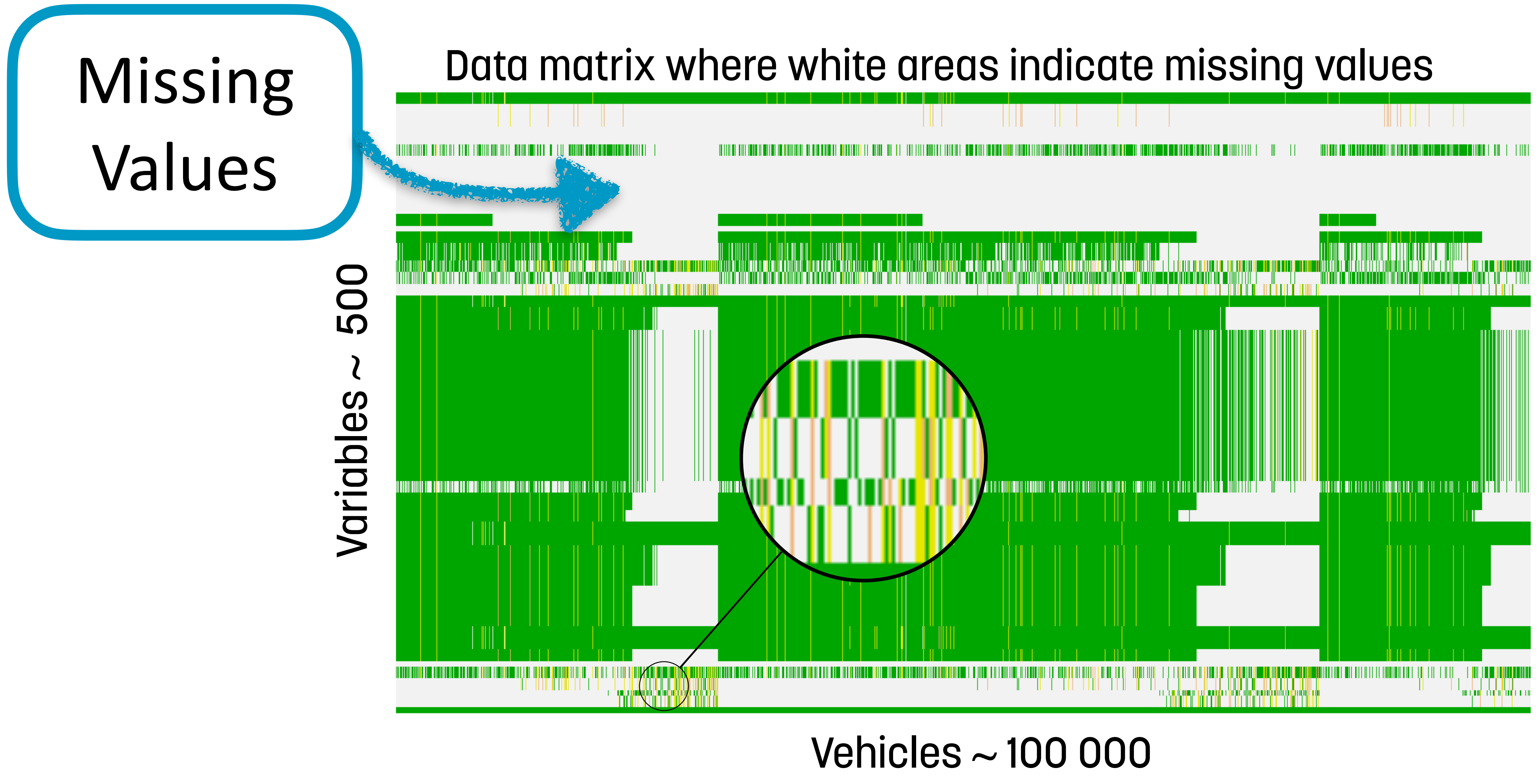
- Data sources
  - operational data
  - guarantee data
  - workshop history data
- Data characteristics
  - ≈ 100,000 vehicles with
  - High censoring rate (≈ 80-90 %)
    Good for Scania bad for modeling
  - ≈ 1,000,000 data readouts
  - Readouts now and then including aggregated usage

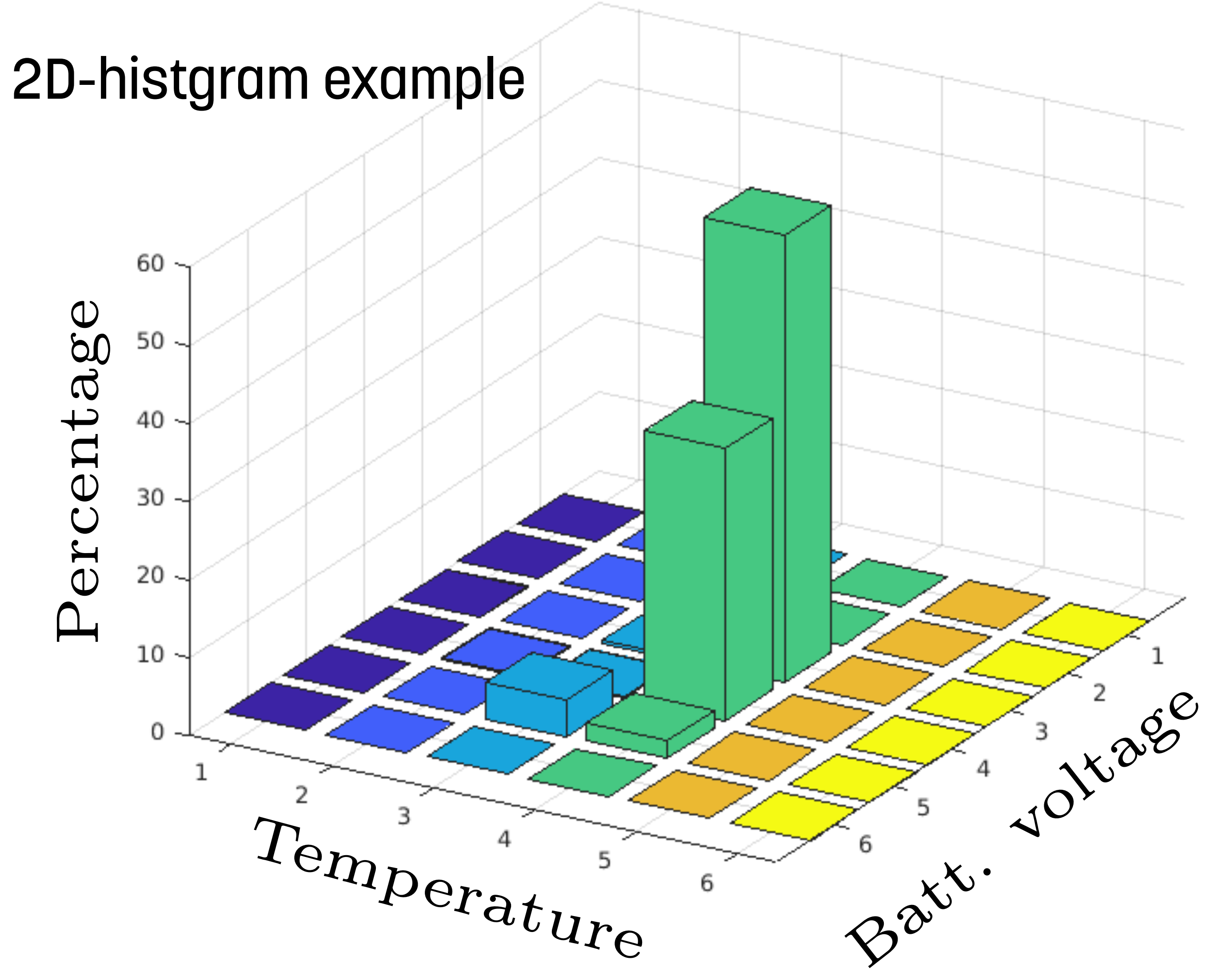




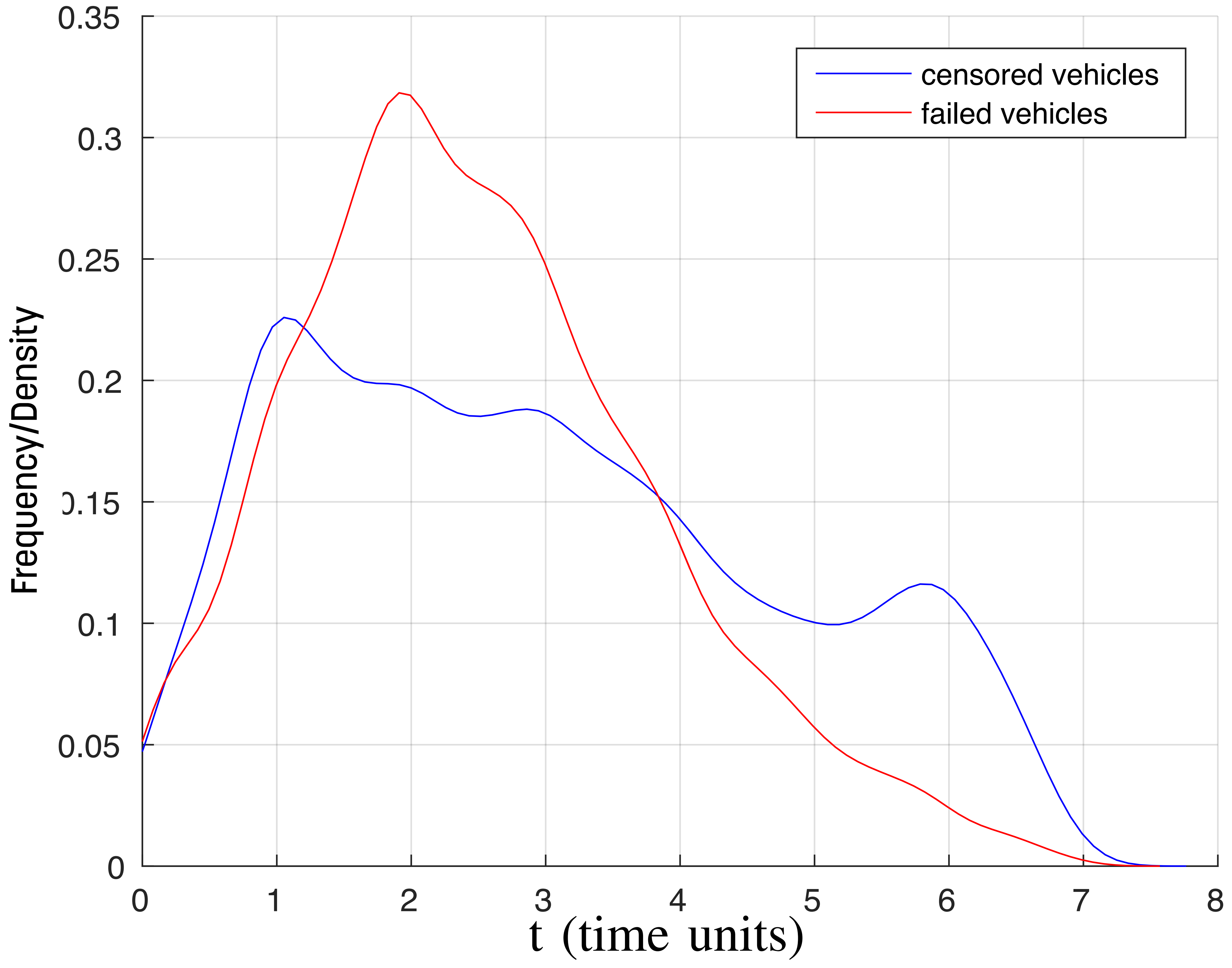
Readout Characteristics

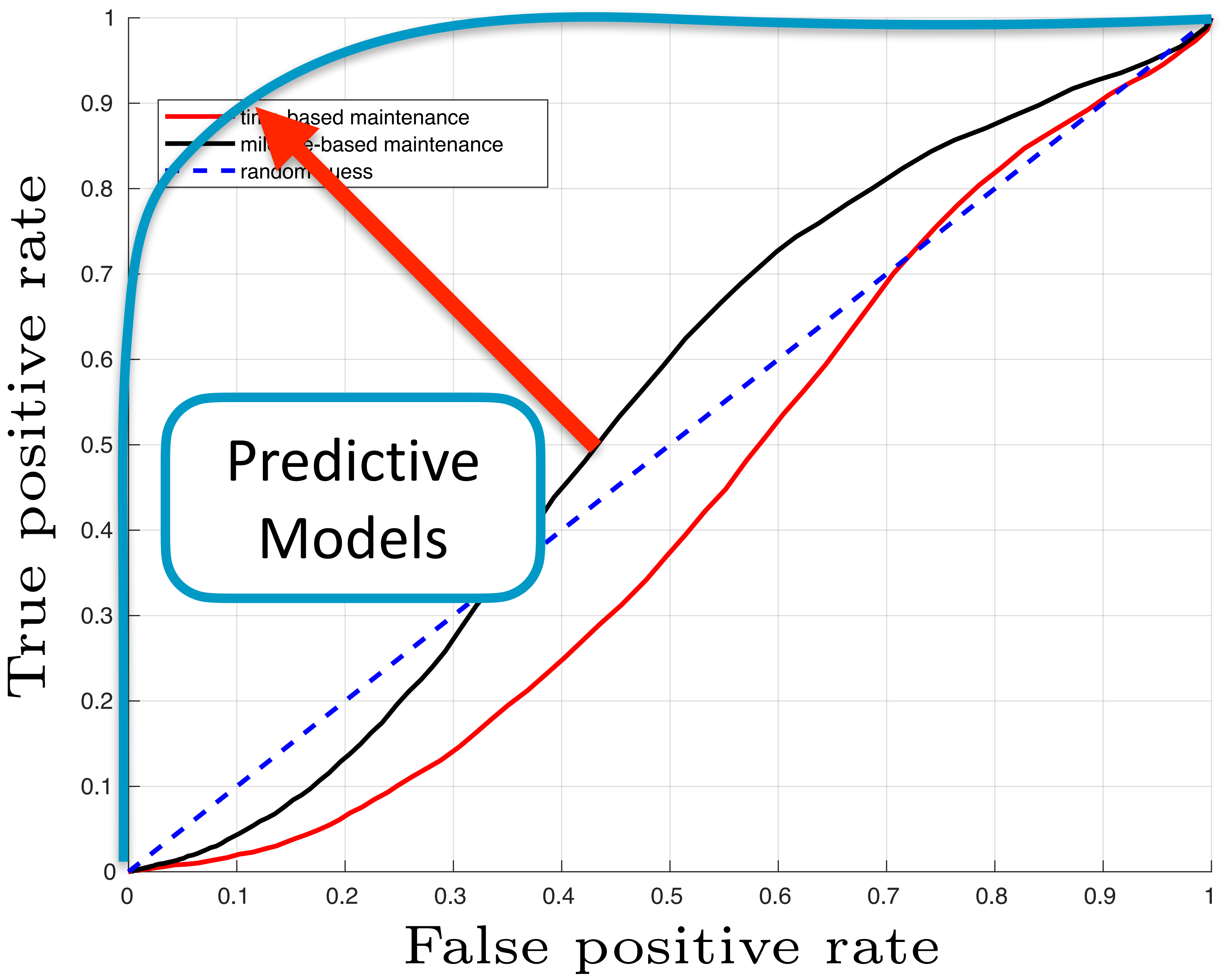
- Data readouts
  - Categorical/configuration
  - A few floating point numbers
  - 1D and 2D histograms
- About 500 variables stored
- Significant missing data rate about 40 percent
- **Important:** Not possible to estimate battery health-state from measured signals!
- Extensive data on vehicle usage and configuration, allowing for correlation with component lifespan.





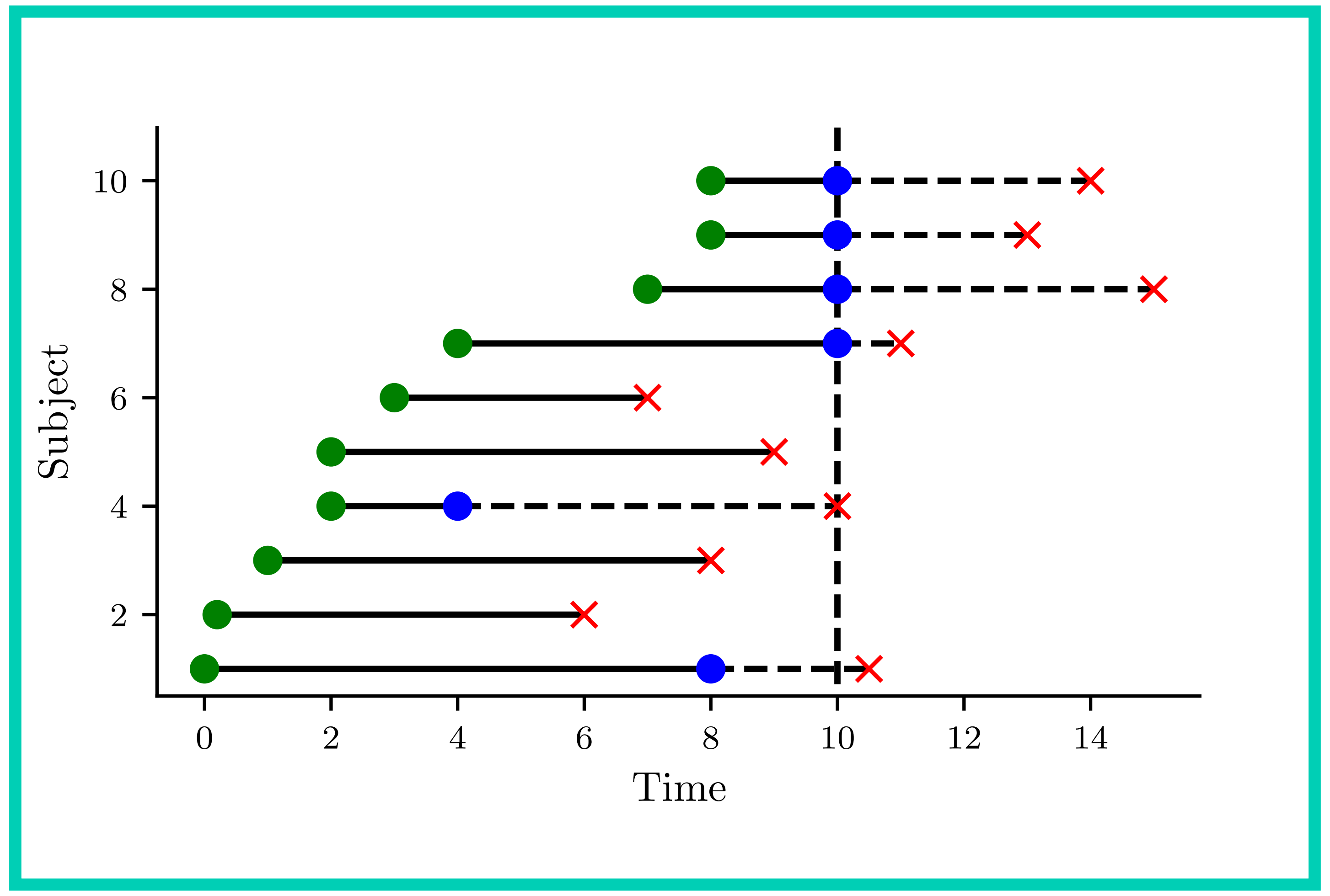
Time or mileage based maintenance plans



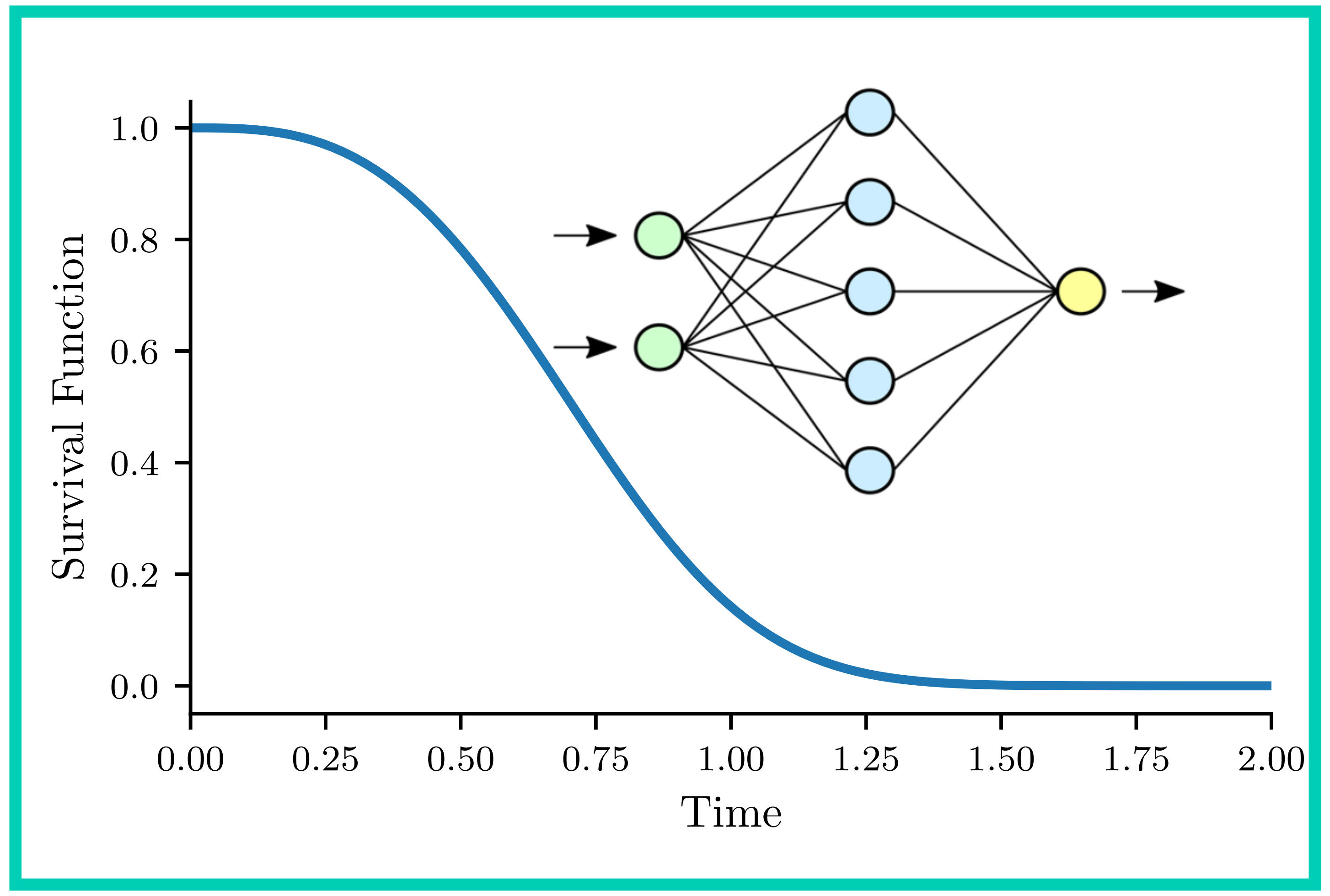


Survival Models for Predictive Maintenance

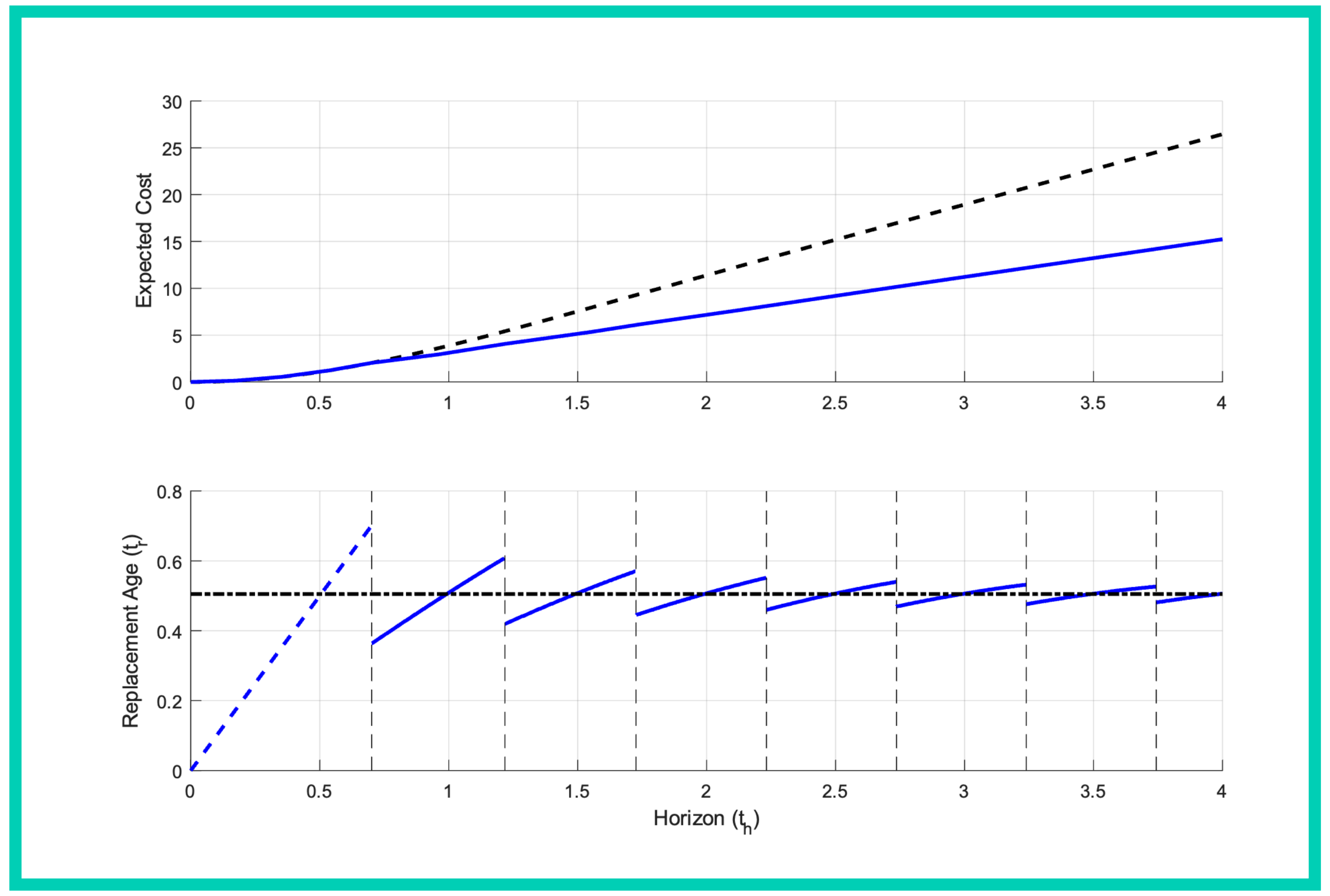
Data



Model

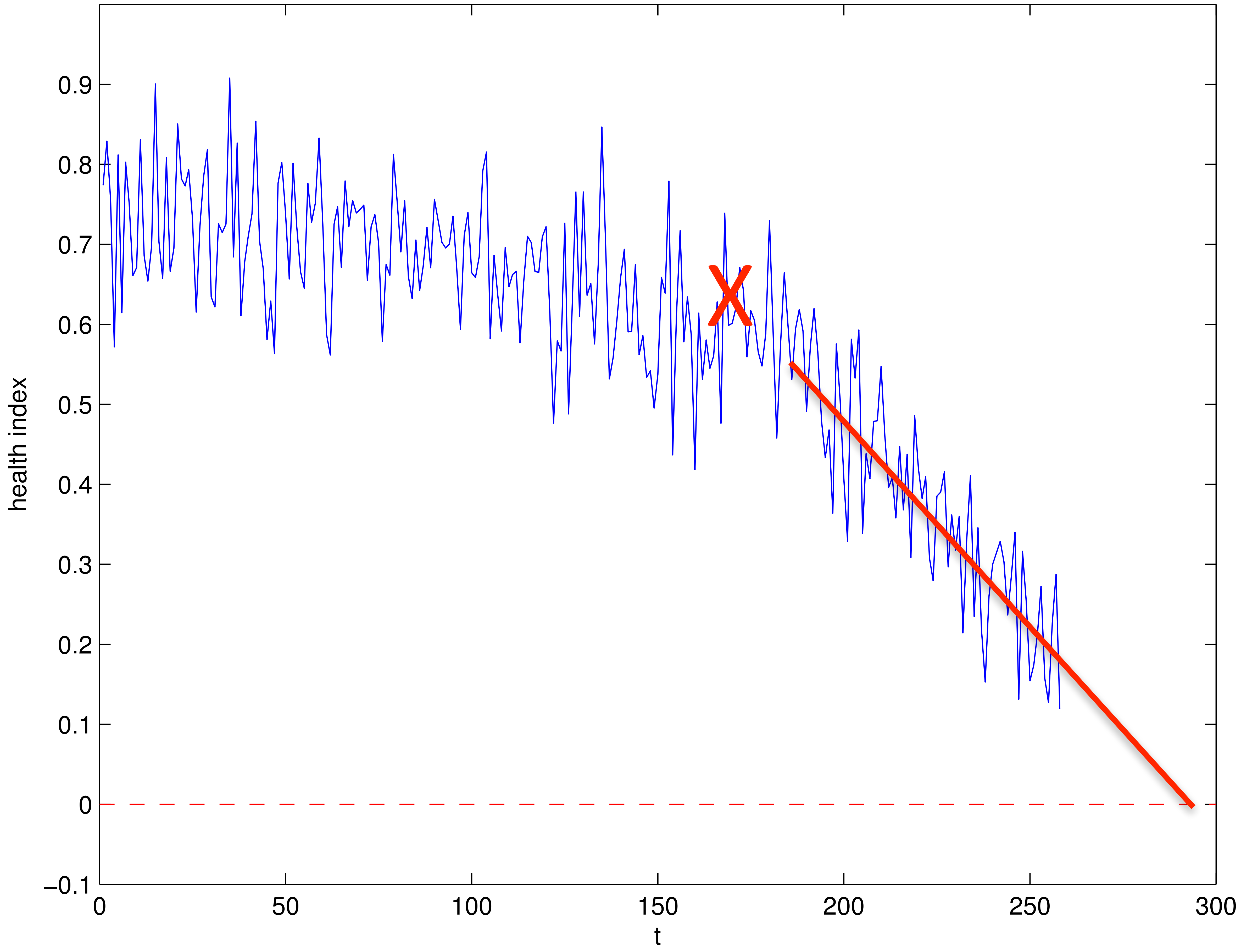


Individualized Maintenance Policy

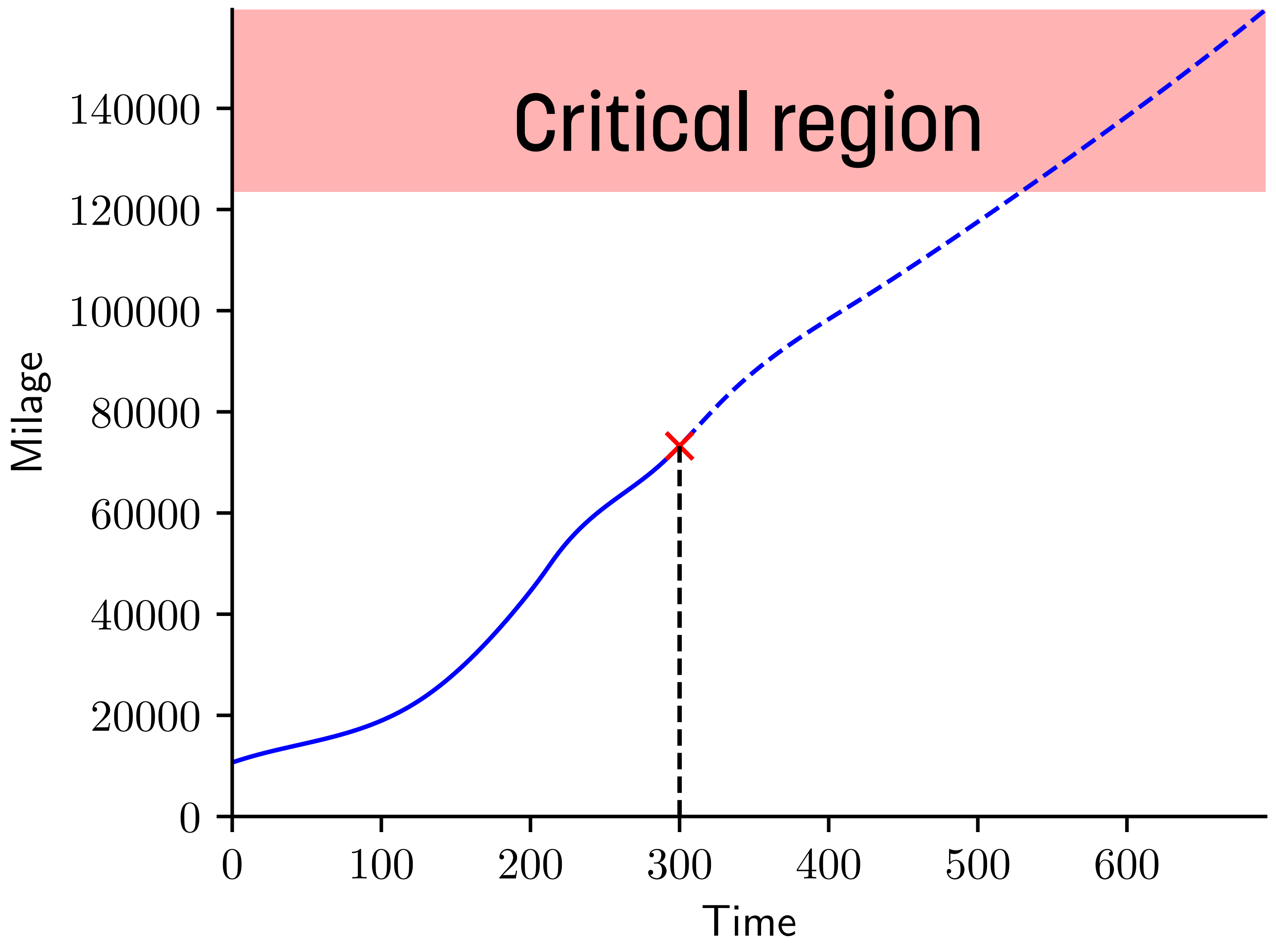


Modeling with static (or low-rate) data

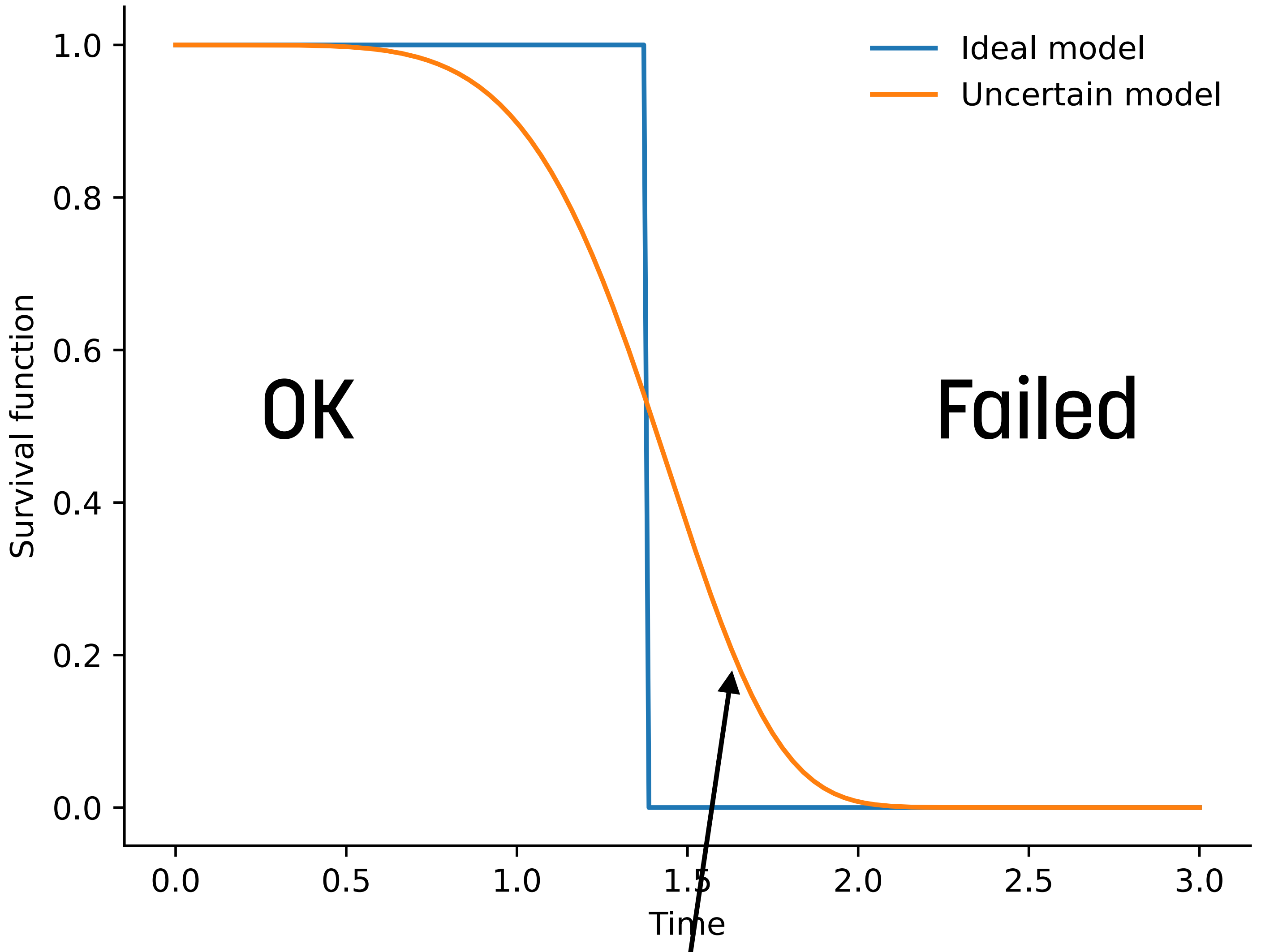
- Trad.: prognostics ~ trend analysis
- Question:
  How do you do trend analysis with few, or even single, data readouts and you can't even reliably estimate the health?
- (one) Answer:
  Look in a database and find patterns with *similar* usage and learn from their experience
- Method research: Find methods
  - How to automate "look-up"
  - How to determine "similar"

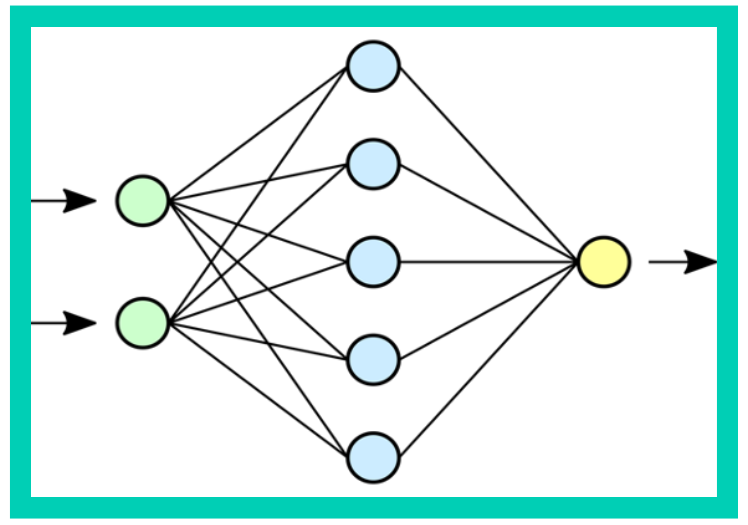


Survival models for predictive maintenance



$S(t \mid x) = P(T > t \mid x)$





Data-driven models

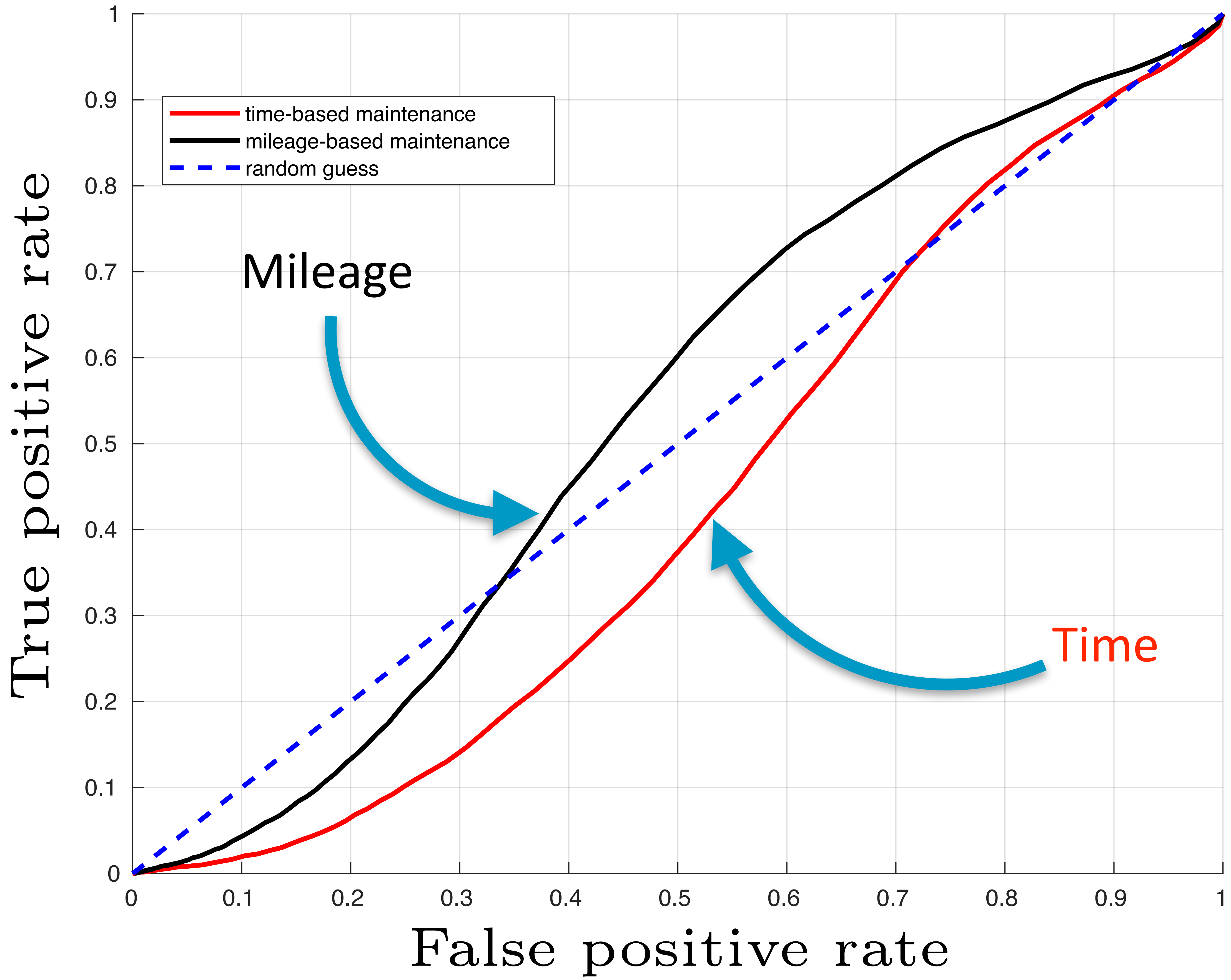
- How a system is used often has a big impact on when the system will fail
- It is often difficult/impossible to predict the exact time of failure

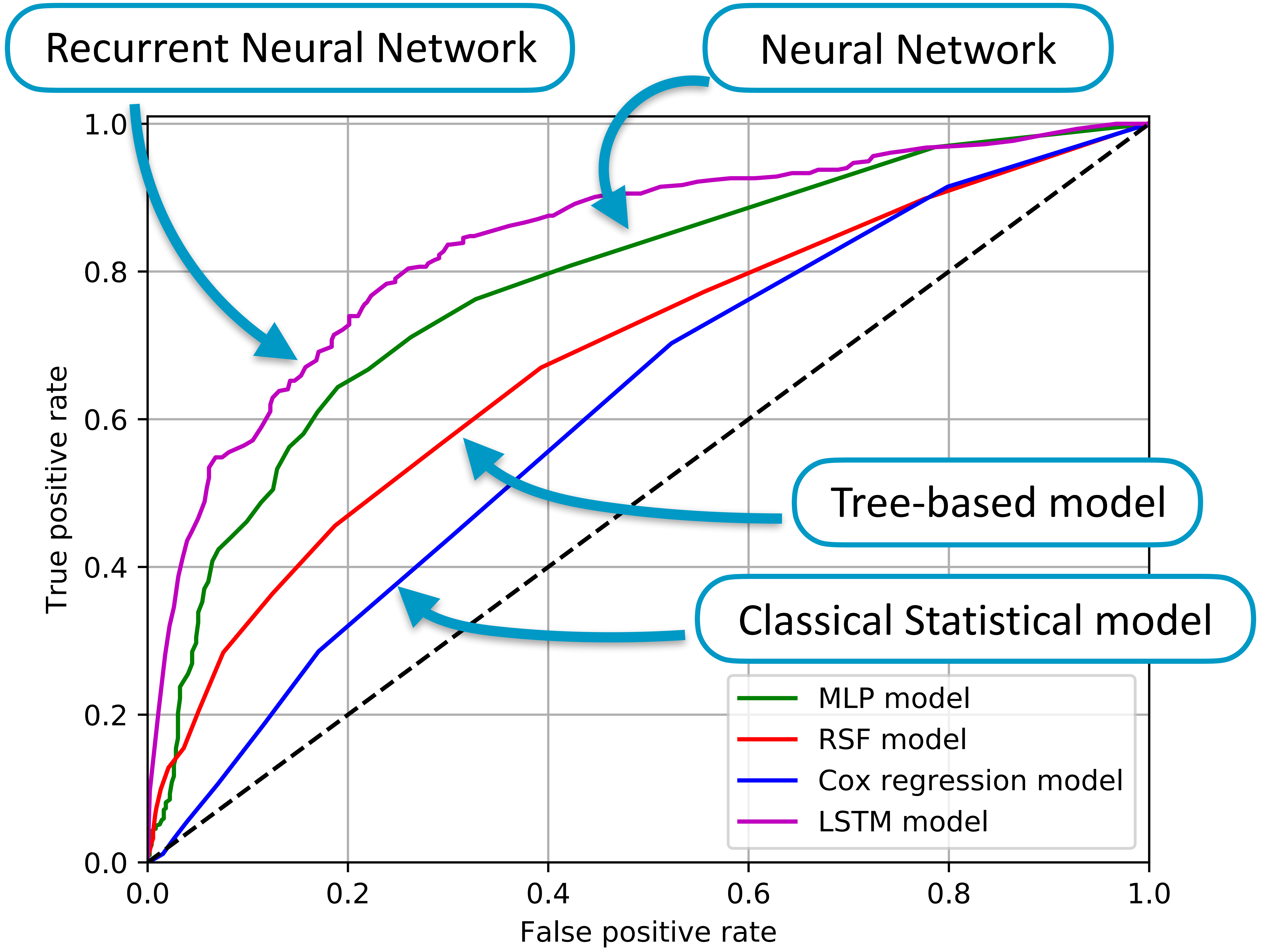
  ➡ Survival models describe the distribution of the failure time

  ➡ Problem: Estimate the survival function:

$$S(t \mid \text{data}) = P(\text{ survive until time } t \mid \text{data })$$

Model performance - a high level perspective

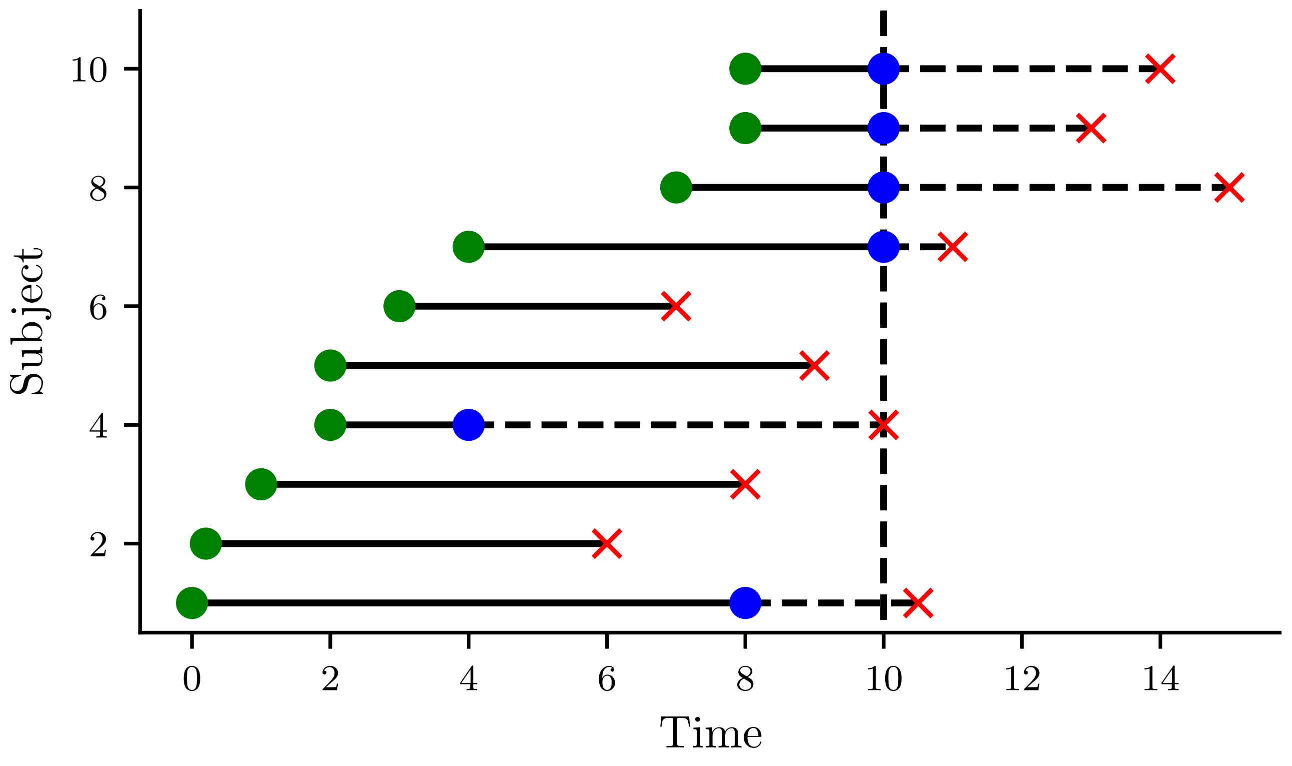




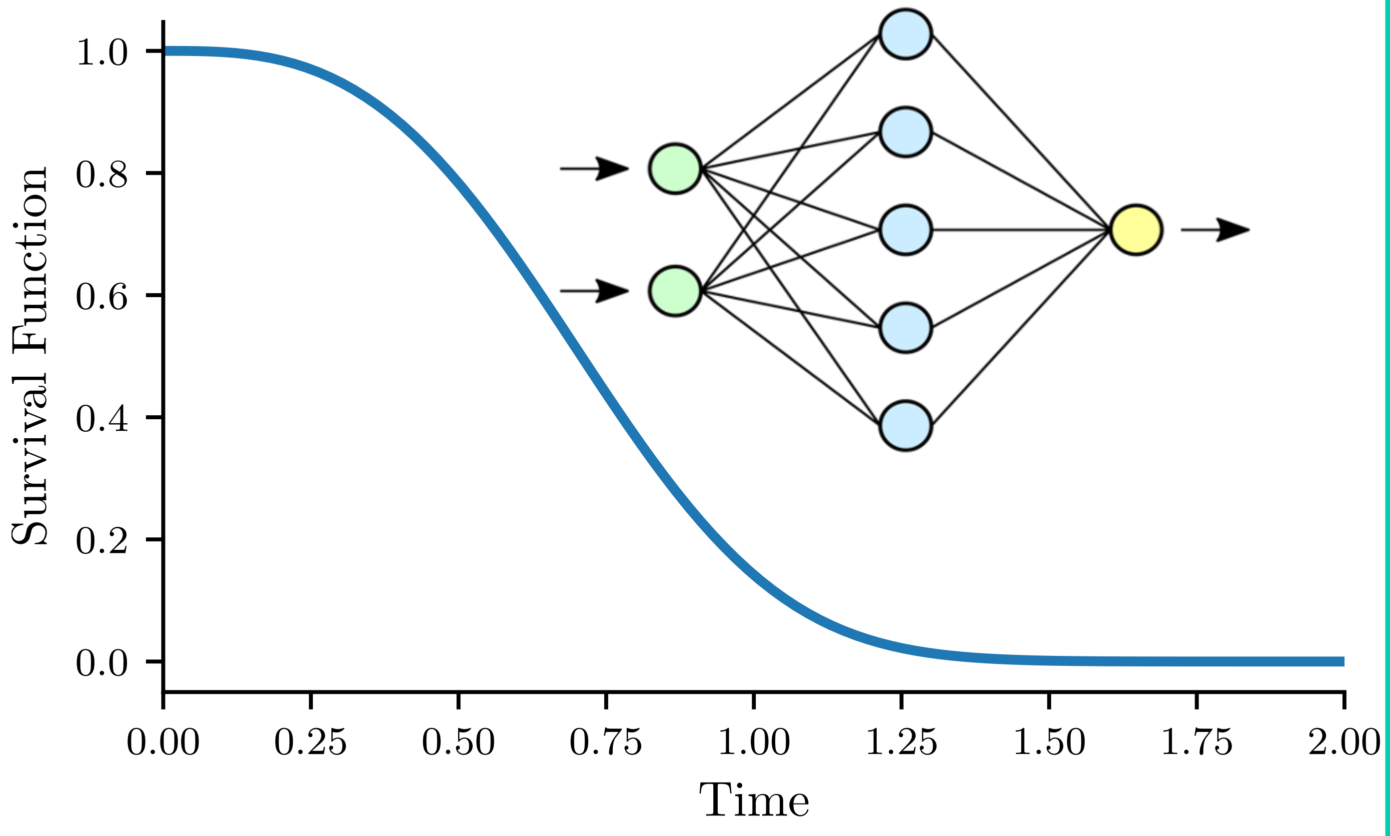
Voronov, S. "*Machine learning models for predictive maintenance*". PhD Diss. Linköping University, 2020.

Individualized Age-based Replacement

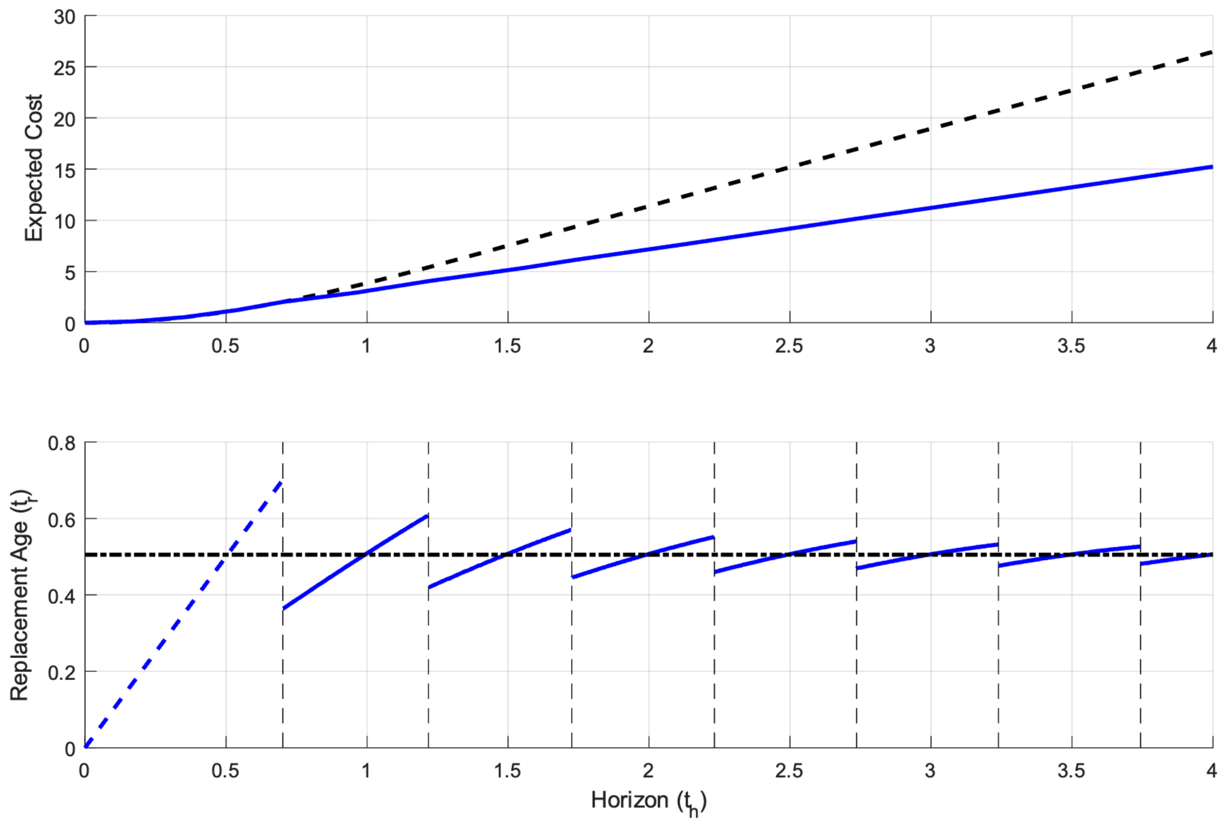
Data



Model



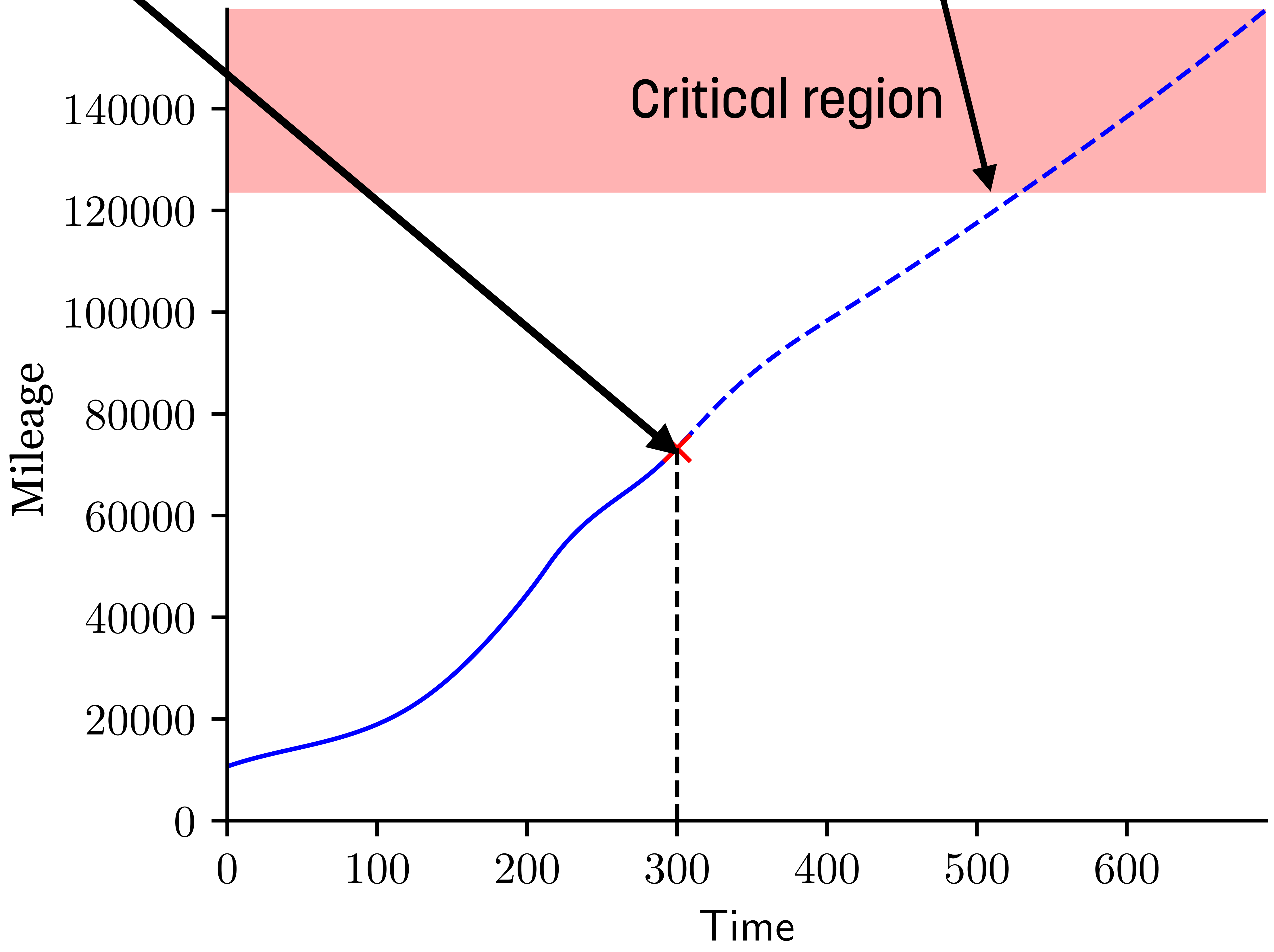
Individualized Maintenance Policy



Maintenance planning

Consider a component in a truck that tends to fail after a certain mileage.

We want to **decide as early as possible** to **repair/replace as late as possible**

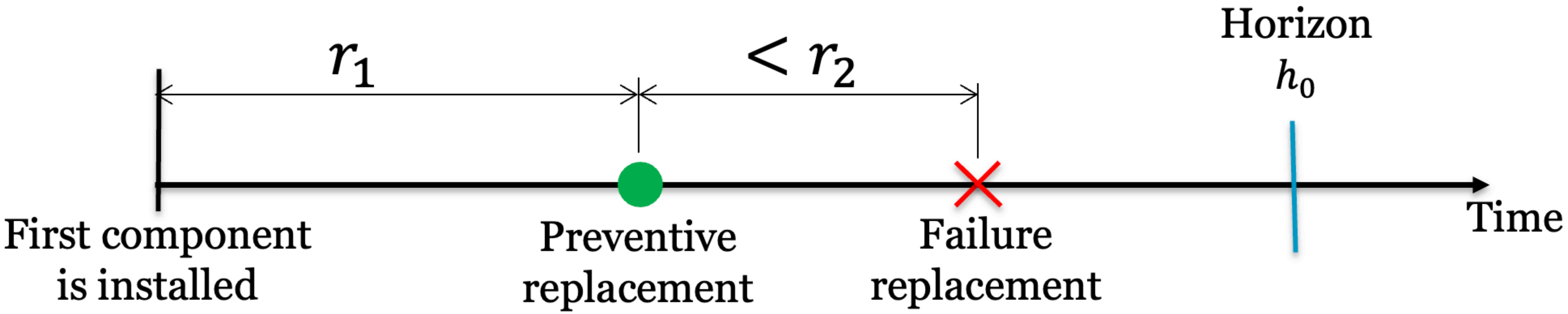


Age replacement

- The age replacement policy can be stated as follows:

  Replace the component with a new one when it reaches a specific age or when it fails, whichever comes first.

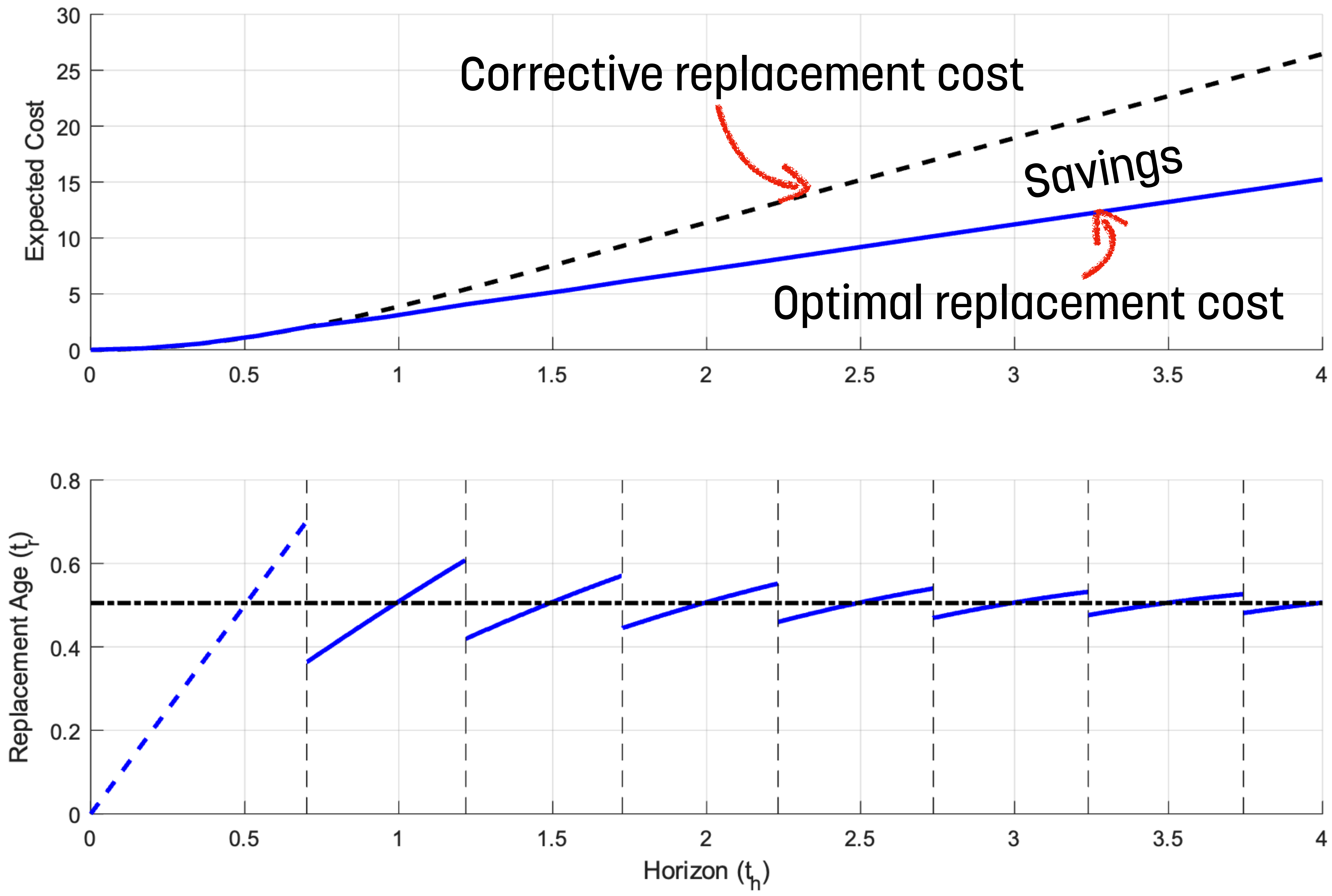
Given a statistical description (survival model) of the failure times, the optimal replacement ages can be determined



Optimal replacement policies

- One simple cost-model is
  - c_f - cost of replacing a failed component
  - c_p - cost of a preventive replacement
- Typically $c_f \gg c_p$ (otherwise, run to failure is the optimal)
- Minimize the expected maintenance cost over a finite horizon.

$$C = \mathbf{E}\left(c_f N_f + c_p N_p\right)$$



Takeaways

- Predictive maintenance: Cost-efficient, reliable operations via data-driven insights.
- Survival modeling: Tackles uncertainty and limited measurements effectively.
- Individualized age-based replacement: Reduces downtime, and optimizes resource use.
- Truck batteries: Highlight challenges of diverse configurations and incomplete data.
- SCANIA component X dataset: Real-world anonymized open dataset similar to the battery dataset.

Python package for survival modelling

SCANIA Component X Dataset