Optimising maintenance operations in photovoltaic solar plants using data analysis for predictive maintenance

Gøran Sildnes Gedde-Dahl
Msc. Data Science

Overview

- Background
- Theoretical background
- Analysis
- Conclusion & future work

Motivation

Importance of reliability

Consequence of unexpected failure

Trade off: avoiding unexpected failure vs. exploiting full potential lifetime

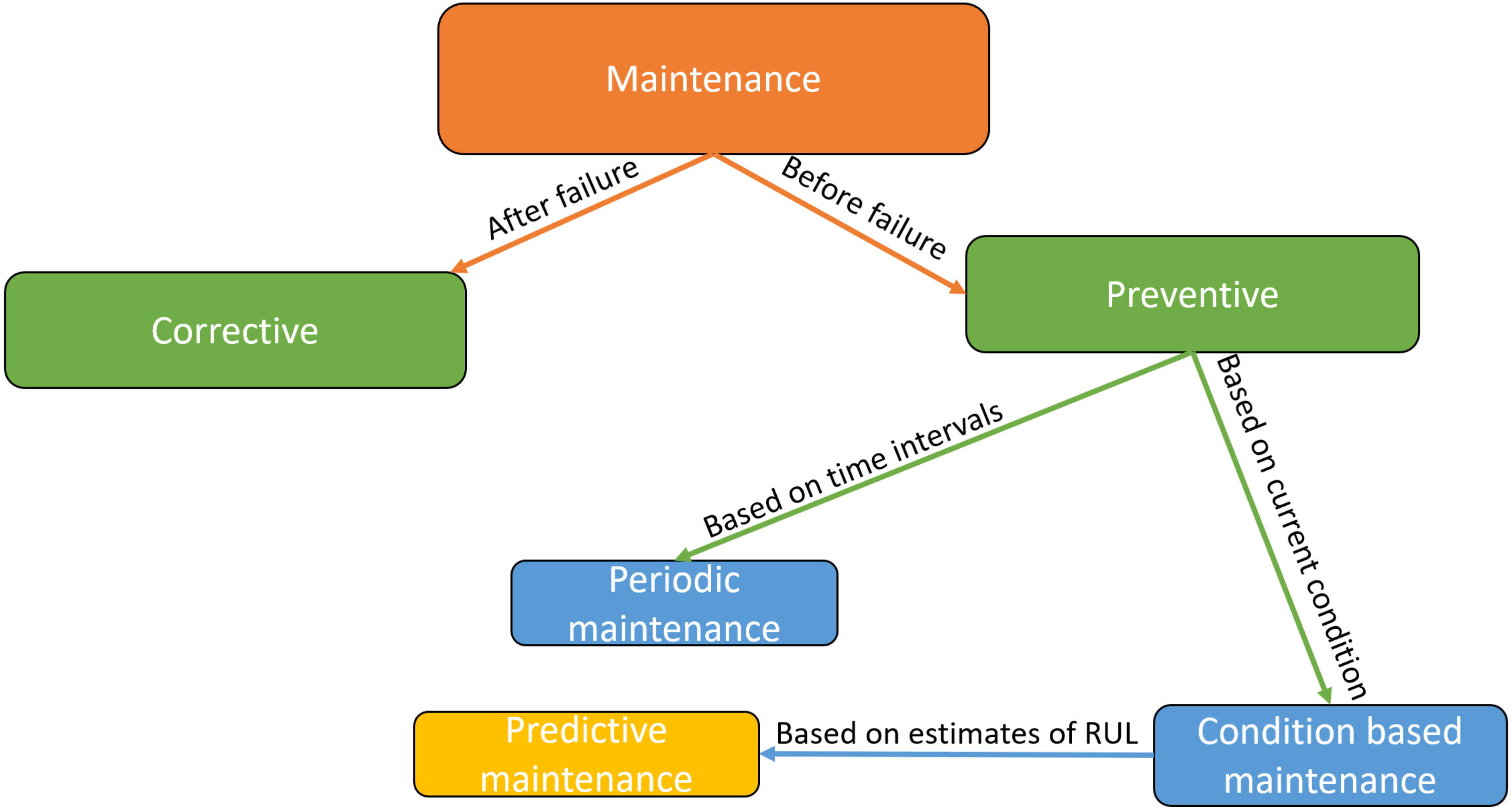
Goals

Conduct a proof of concept for predicting RUL (Remaining Useful Lifetime) using data science

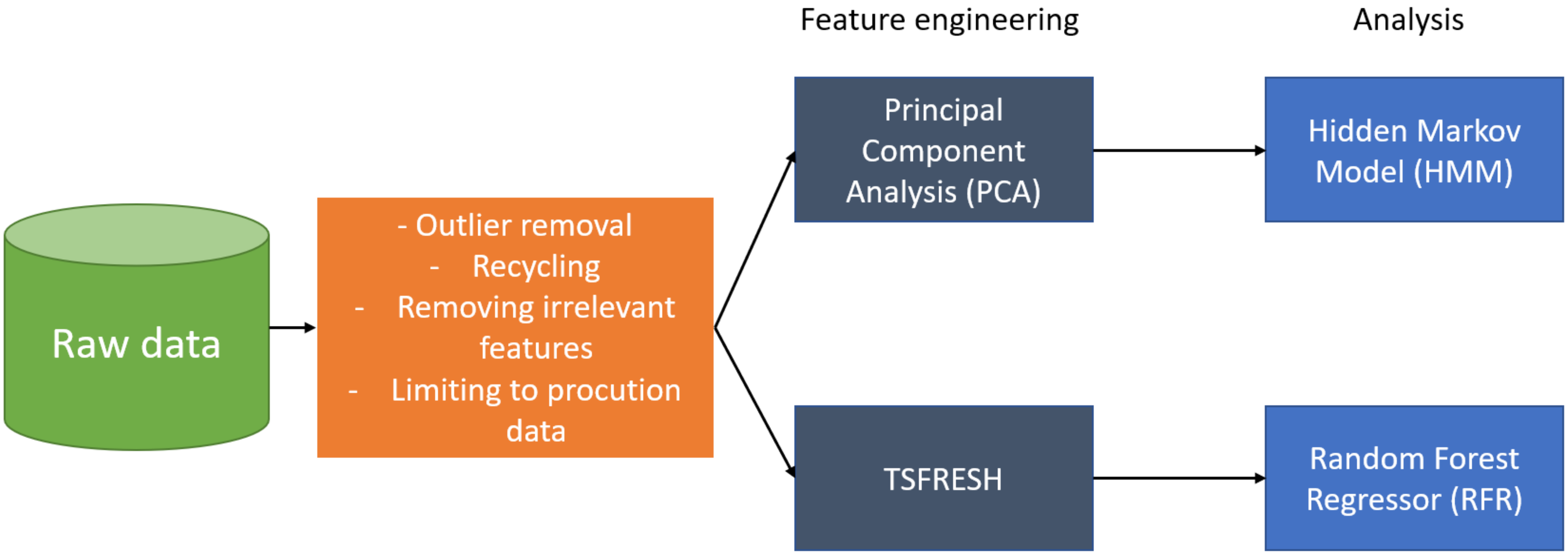
Make use of statistical and machine learning models

Evaluate and compare the competing approaches

Predictive maintenance



Proposed pipelines



Methodology

Principal Component Analysis (PCA)

Working method

- Dimensionality reduction
- Creates new features from input features– principal components

User choice

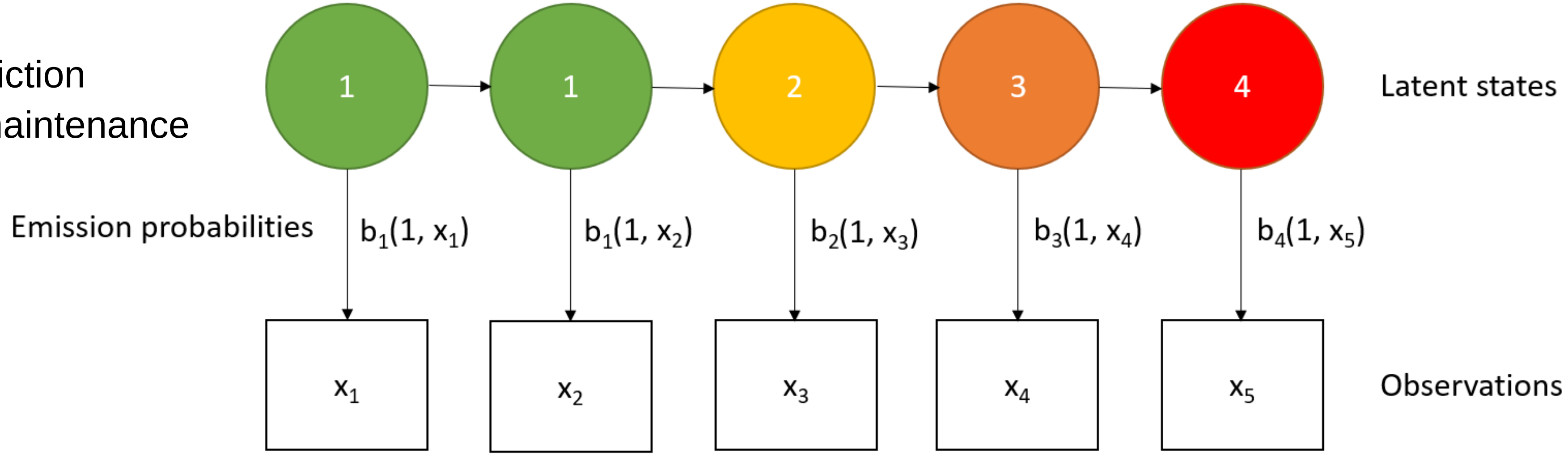
- Trade-off: number of features in output vs. explained variance

Hidden Markov Model

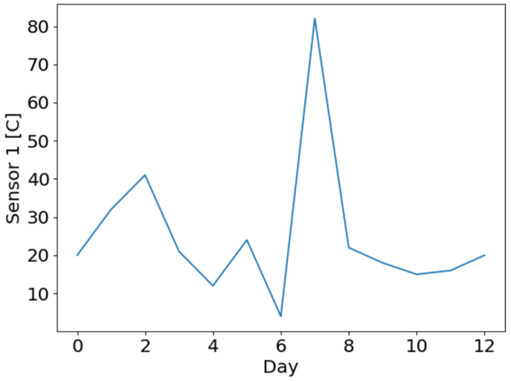
- A statistical method, unsupervised, probability based
  - Assumptions

- Parameters
  - Number of states
  - Start probability
  - Transition probability matrix
  - Covariance type

- Endpoint predicted state
  - Possibilities for RUL prediction
  - Use state for predictive maintenance



TSFRESH – Time Series Feature Extraction using Scalable Hypothesis test[1]



- 16 data points
- 1 Attribute
- 1 Period
- 1 ID

TSFRESH
feature extraction

S1.Peaks	S1.Mean	S1.Median	...	S1.max
3	25,15	20.0	...	82

- 1 data point per period per ID
- 789 new attributes per attribute

Automated feature extraction method

- Temporally invariant and variant information

Summarizing time series into one row

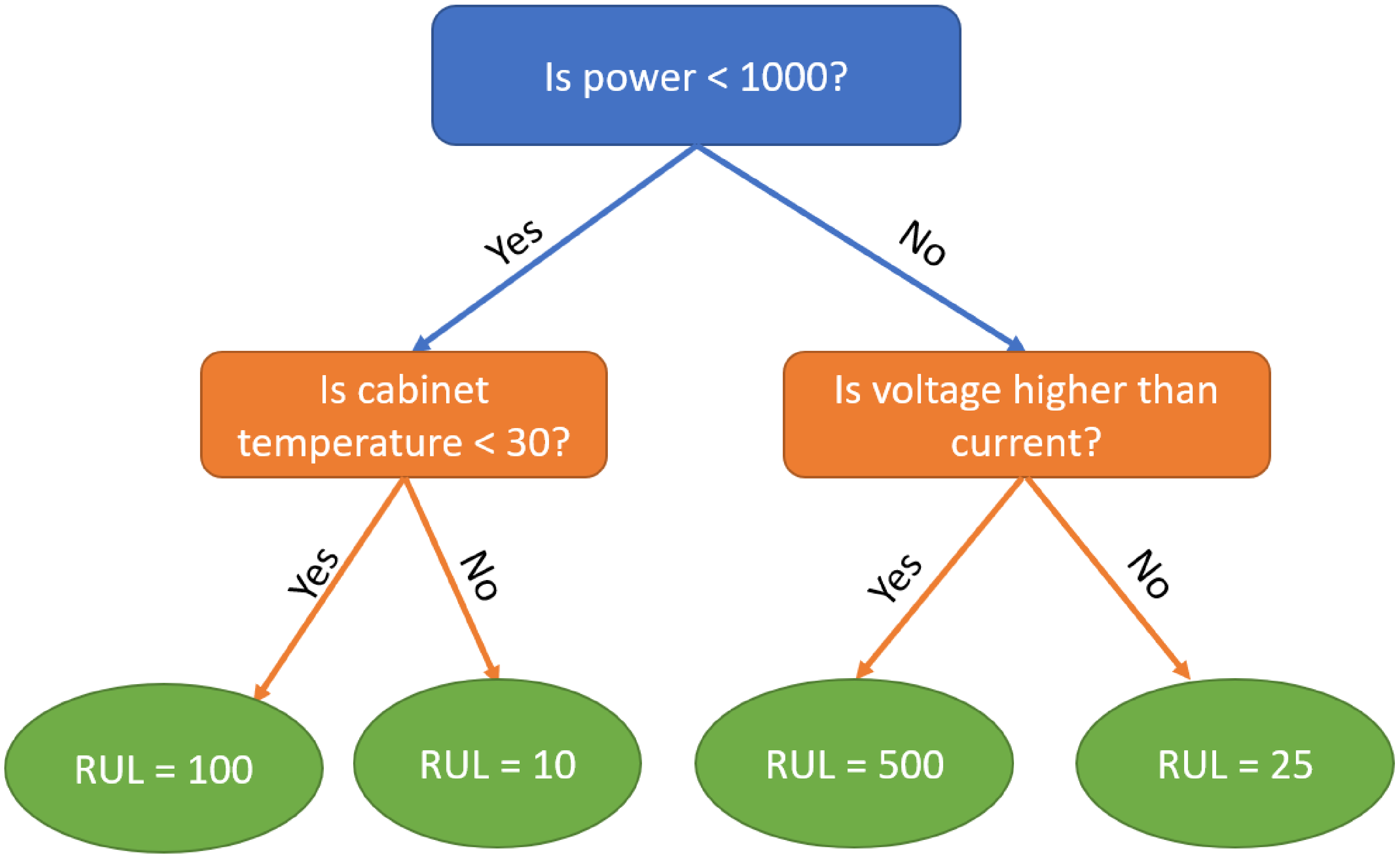
- 789 new columns per feature in the input data
- 77 functions with different parameter settings

Feature selection

- Scalable Hypothesis test
- Benjmini-Yeuketeli – eliminating false discovery

1 - https://tsfresh.readthedocs.io/en/latest/

Random Forest Regressor



Experiments

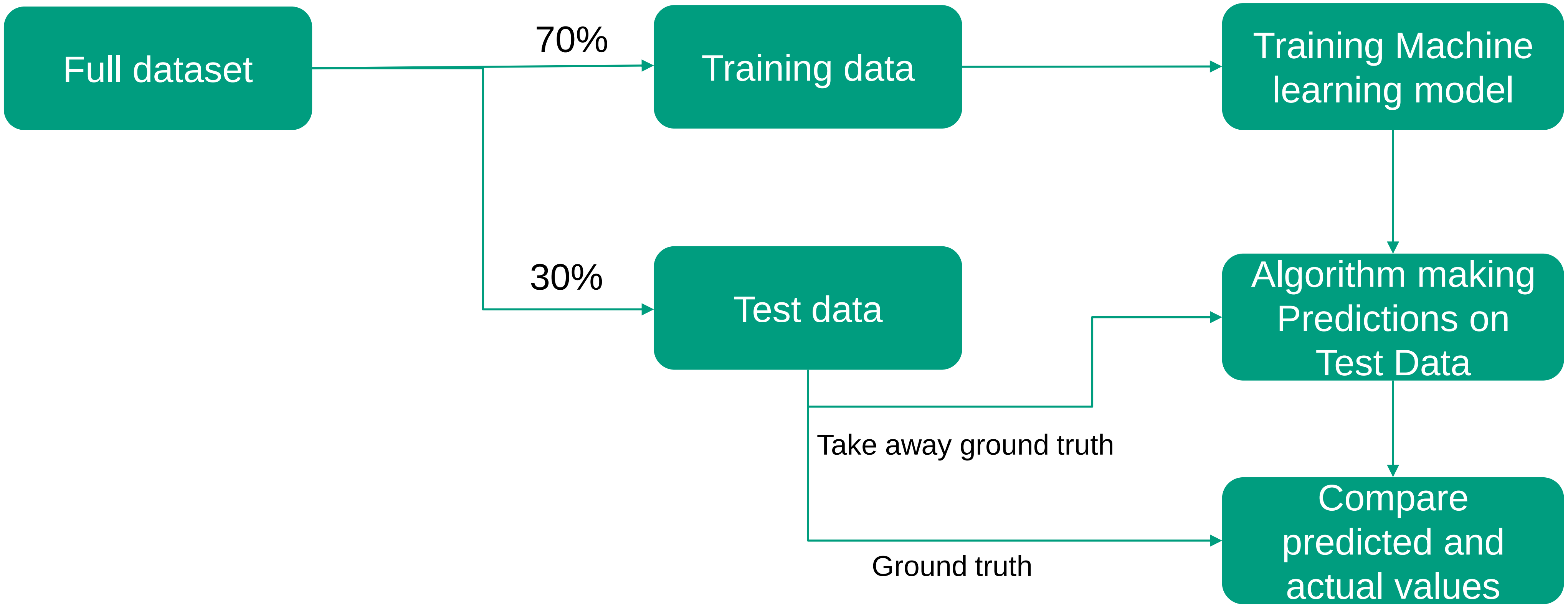
Data background

- All Inverters and weather stations in all Egypt plants
  - Sampled every minute from June 2019 to 2022
  - Resampled to be every 30 minutes by mean value aggregation

- One certain failure type determined for prediction

Inverter	Timestamp	s1 DC power	s1 DC voltage	s1 DC current	s1 Phase 1 module temperature	...	Status	Irradiation horizontal	Irradiation incline
01	06.05.2019 08:00	855,19	1193,29	718,75	83,5	...	164	998,4	1037,0
01	06.05.2019 08:30	1031,68	1143,38	910,16	93,1	...	164	1025,4	1037,4
01	06.05.2019 09:00	1226,49	1095,22	1123,15	105,3	...	164	1035,9	1039,4

Experimental setup

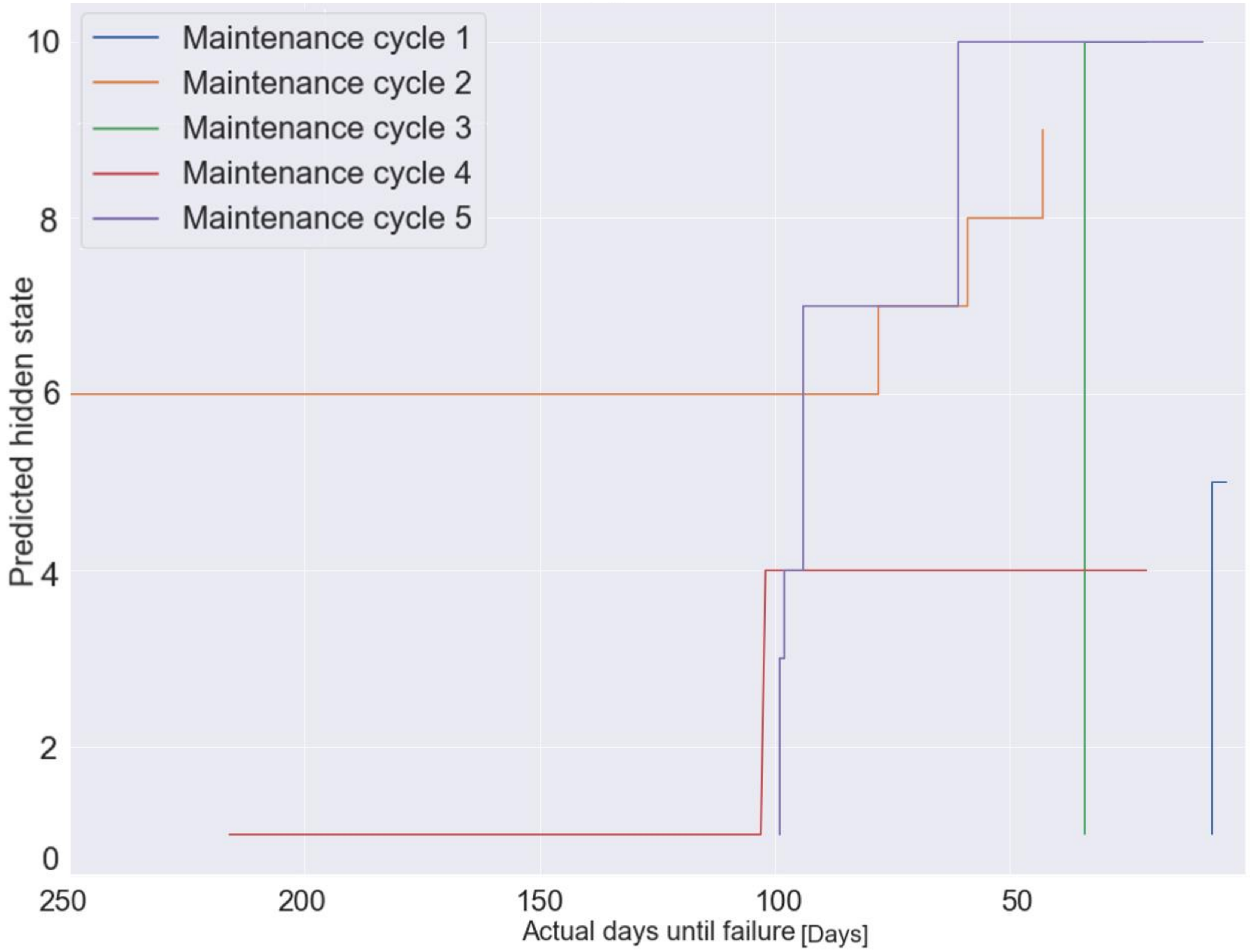


Approach 1

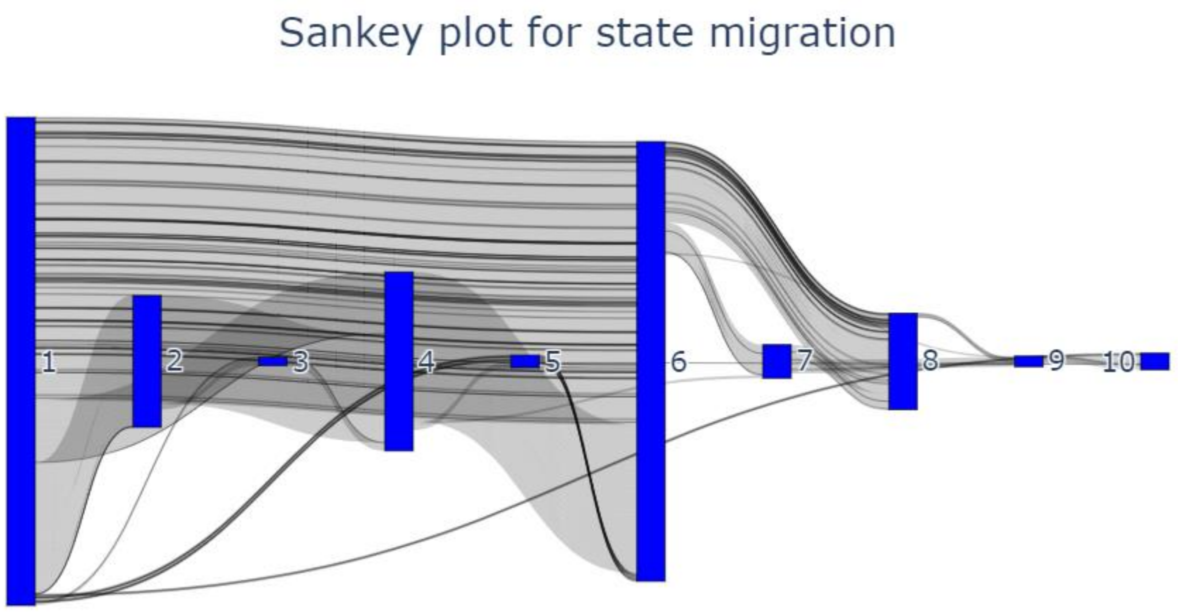
Degree of failure risk using a statistical Hidden Markov Model

Results

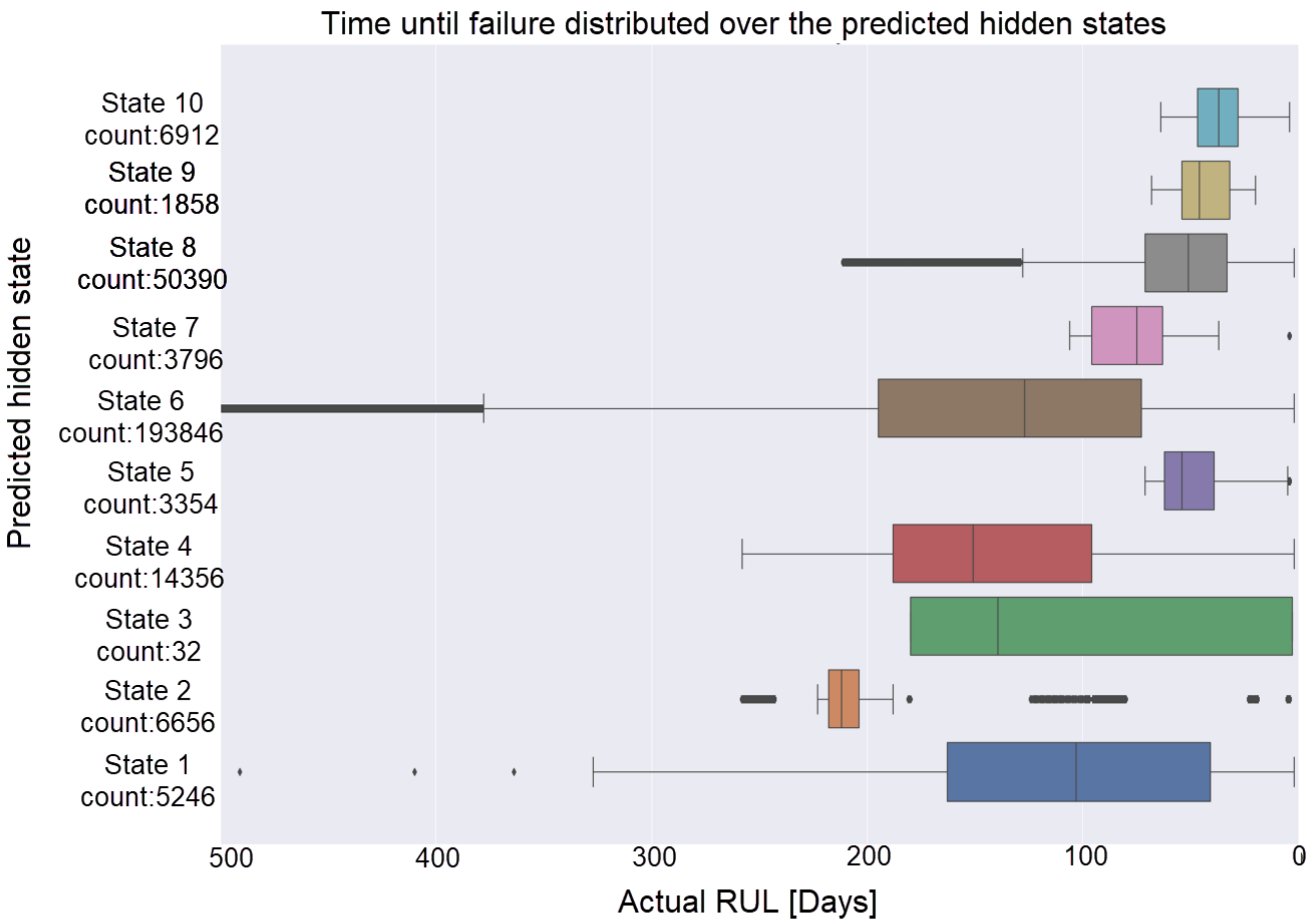
HMM state prediction for example maintenance cycles



Results



Sankey plot for state migration
1
2
3
4
5
6
7
8
9
10

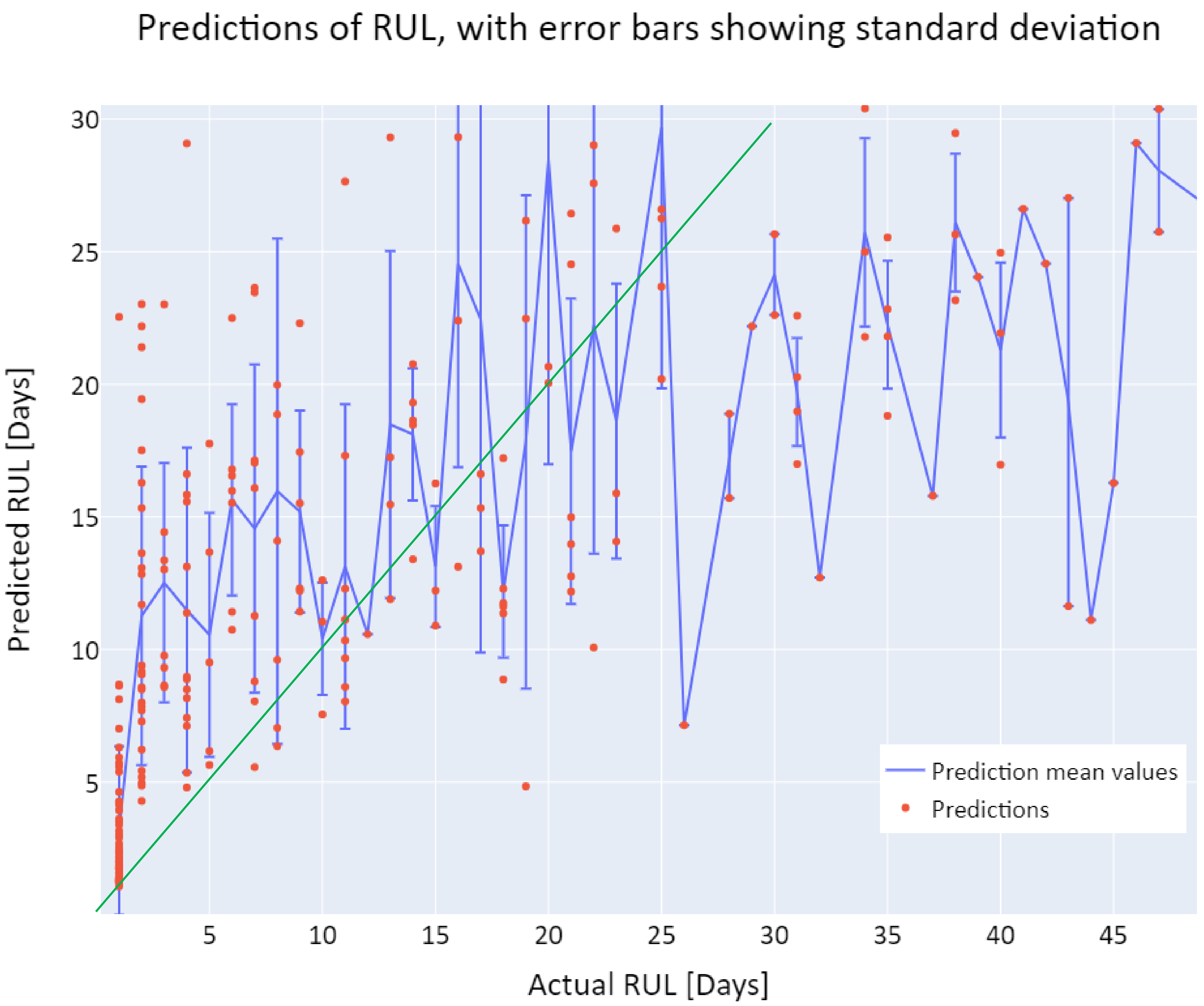


Time until failure distributed over the predicted hidden states
Predicted hidden state
State 10
count:6912
State 9
count:1858
State 8
count:50390
State 7
count:3796
State 6
count:193846
State 5
count:3354
State 4
count:14356
State 3
count:32
State 2
count:6656
State 1
count:5246
500
400
300
200
100
0
Actual RUL [Days]

Approach 2

Machine learning algorithm estimating the time until failure at a given point in time

Results



Predictions of RUL, with error bars showing standard deviation
Predicted RUL [Days]
Actual RUL [Days]
30
25
20
15
10
5
5
10
15
20
25
30
35
40
45
Prediction mean values
Predictions

Results

- Feature importances
- Randomness
- Relevance – feedback from experts

Feature name	Importance
Total_Capacitive_Reactive_Energy_in_the_inverter_symmetry_looking_r_0.3	0.37
Section_2_Status_symmetry_looking_r_0.1	0.14
Liquid_Cooling_flow_variance_larger_than_standard_deviation	0.13
Section_2_DC_Power_Measurement_has_duplicate	0.08
Section_1_Phase_3_power_module_temperature_has_duplicate	0.07
Line_Voltage_Measurement_of_Phases_2_and_3_large_standard_deviation_r_0.1	0.07
Section_2_Phase_2_power_module_temperature_large_standard_deviation_r_0.25	0.06

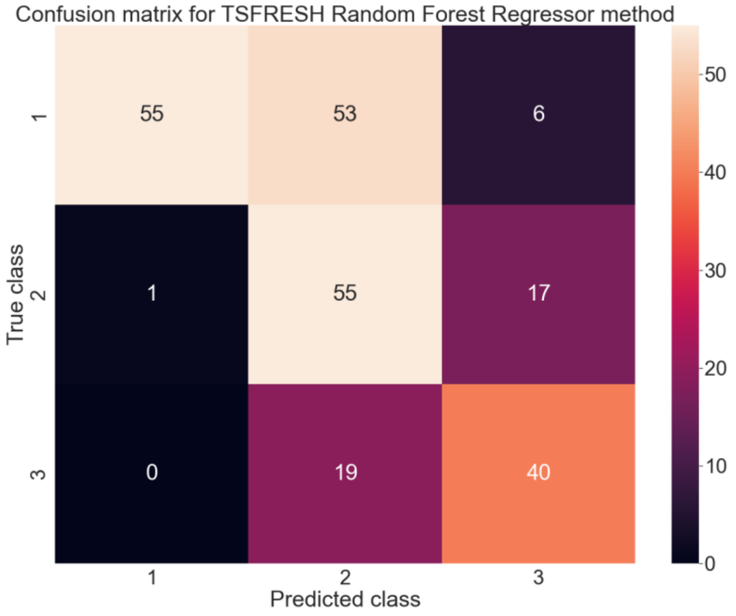
Comparing the methods

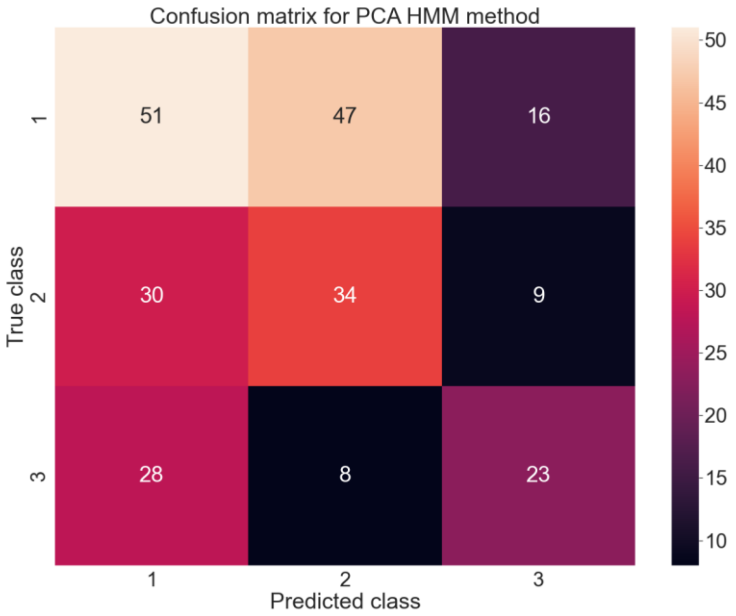
- Common benchmark process
  - Prediction transformed to risk class

Prediction	RFR	HMM
Risk 1	$[0, 5]$	$[8, 10]$
Risk 2	$(5, 20]$	$[5, 7]$
Risk 3	$(20, \infty)$	$[1, 4]$

- Confusion matrices

- Interpretation





Conclusion and Future Work

What do these results show?

There is a potential of predictive maintenance using data currently sampled

How can it be improved, what are the next steps?

Remove errors not leading to downtime

Testing on various geographical locations

Implementing an infrastructure for data analysis and displaying results

Testing upsampling of data with large RUL

How do the two approaches compare?

RFR seems to be more stable in the confusion matrix

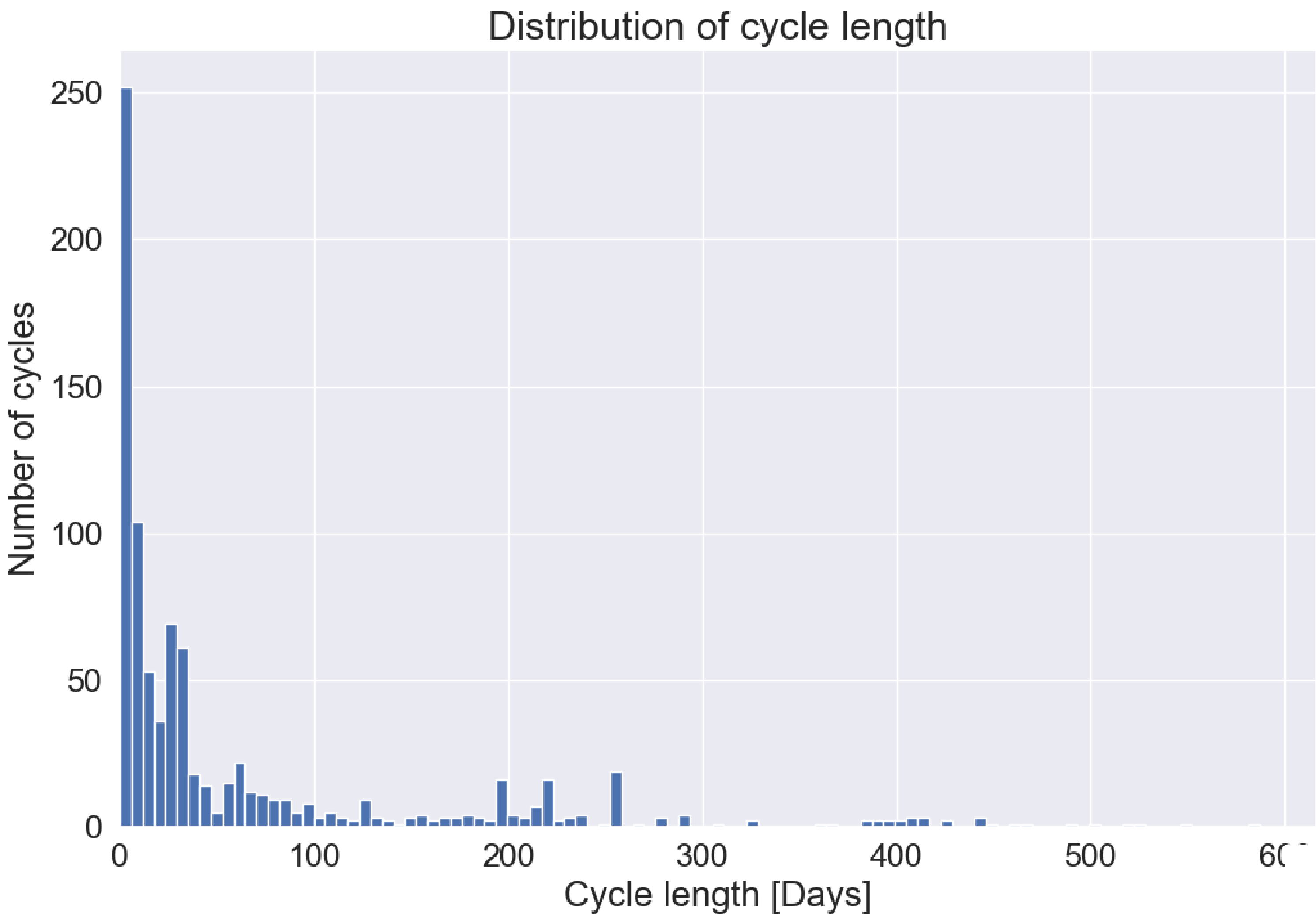
Differences in outputs and predictions

Thank you for your attention!

- Questions?

- Feel free to contact me or further questions or comments
  - E-mail: goran.sildnes@icloud.com – goran.sildnes@prevas.no
  - LinkedIn: Gøran Sildnes Gedde-Dahl

Exploratory data analysis



Preprocessing

Filtering

- Non-production data
- Data from periods of failure
- Non-physical values
- Removing irrelevant columns (i.e. Error columns)

Outlier removal

- Extreme values indicating sensor error

Re-sampling maintenance cycles

- Random ending point before failure, maximum 50 days

Experimental setup: Re-sampling of maintenance cycles