KiK|MPi
Kennis- en innovatiecentrum Maintenance Procesindustrie

BEMAS

Interreg
EUROPESE UNIE
Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling

Masterclass "Voorspellen levensduur van regelkleppen"

Interreg project
Circulair Onderhoud
22 november 2022

Circulair onderhoud

Workshop

data analysis for predictive maintenance

Gert Jacobusse, Mischa Beckers

HZ UAS – november 2022

Timeline

2020

Requirements for analysis products

2021

Analysis of available datasets

2022

Data simulation and analysis results

Requirements for analysis products

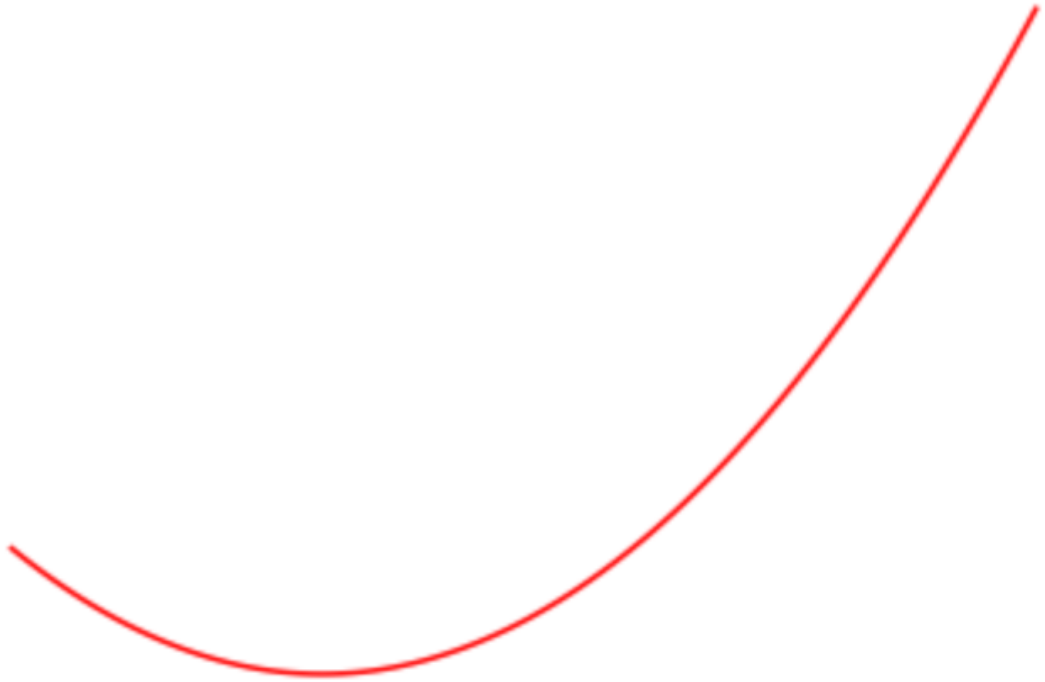
Targets	Features
The failures that need to be predicted: replacements, unplanned maintenance, notifications, type of problem	The parameters that may provide early warnings for a failure (or even cause failure): temperature, flow, pressure, …
• SAP • “switched on since” in positioner data? • Other data sources?	• Selected info from positioner data • Process data (sensors) • Context information

Requirements: survival curve estimation

For survival curve estimation, **only targets** are required

Based on the lifetime (time between installation and failure) of a large number of devices, the probability of failure can be modeled as a function of time.

Without features, probabilities slowly change and provide limited information



Requirements: outlier detection

For outlier detection, **only features** are required

Based on a model of “normal behavior” a signal can be sent when the deviation exceeds a certain threshold

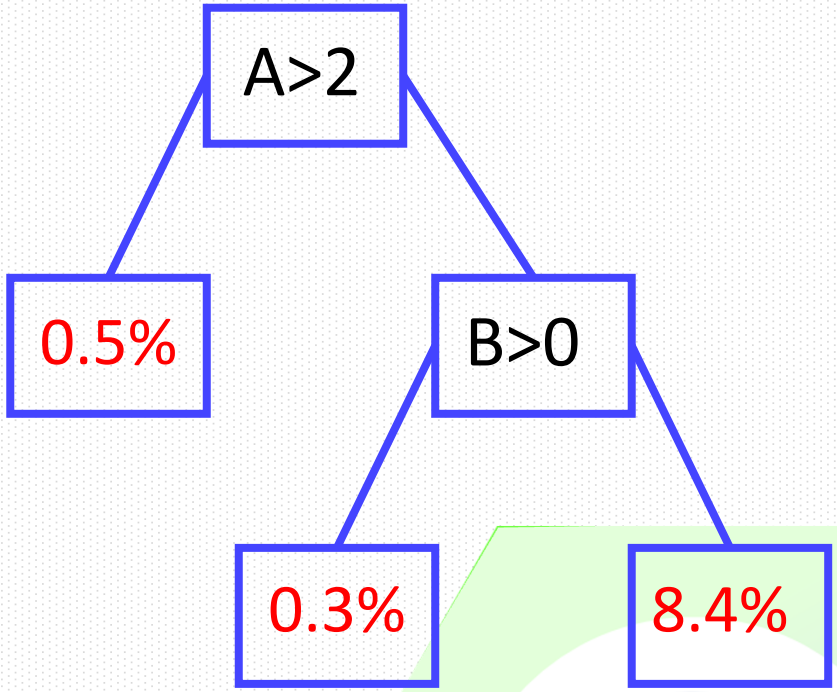
Without knowledge about targets, it is hard to choose (or weight) the features and set the right threshold

Requirements: failure probability prediction

For failure probability prediction, **both features and targets** are required

Relations between features and targets can be exploited to “learn” under what conditions the probability of failure is most increased.

To learn more complex models, a larger number of failure examples is needed.



Discussion

- What data does your organization have?
  - Targets: maintenance, replacement
  - Features: process data, sensors, …

Discussion

- What data does your organization have?
  - Targets: maintenance, replacement
  - Features: process data, sensors, …
- What type of analysis do you currently use to support maintenance?

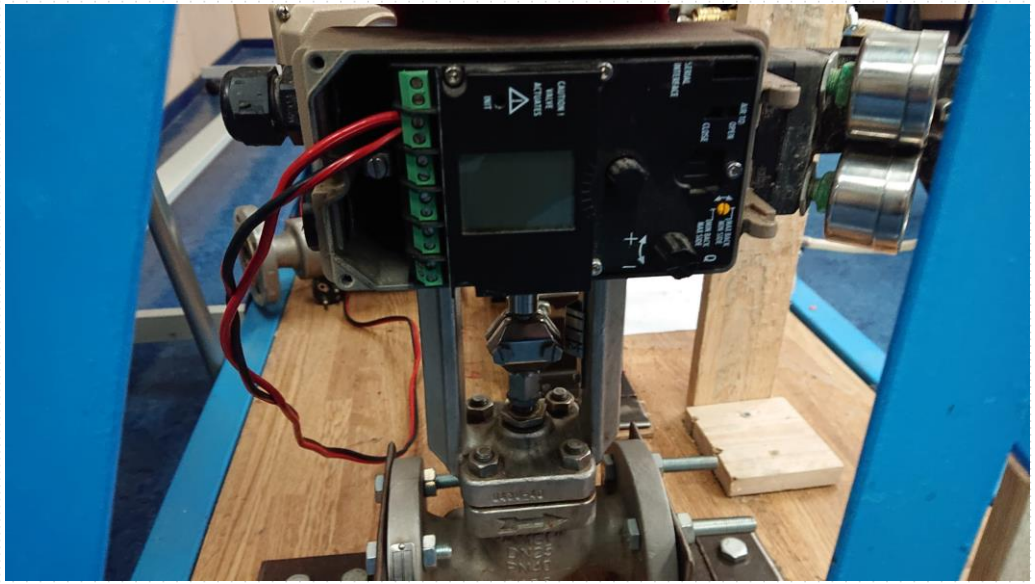
Discussion

- What data does your organization have?
  - Targets: maintenance, replacement
  - Features: process data, sensors, …
- What type of analysis do you currently use to support maintenance?
- What type of information would be most useful for you?
  - Survival curve estimation
  - Outlier detection
  - Failure probability prediction

UNIVERSITY OF APPLIED SCIENCES

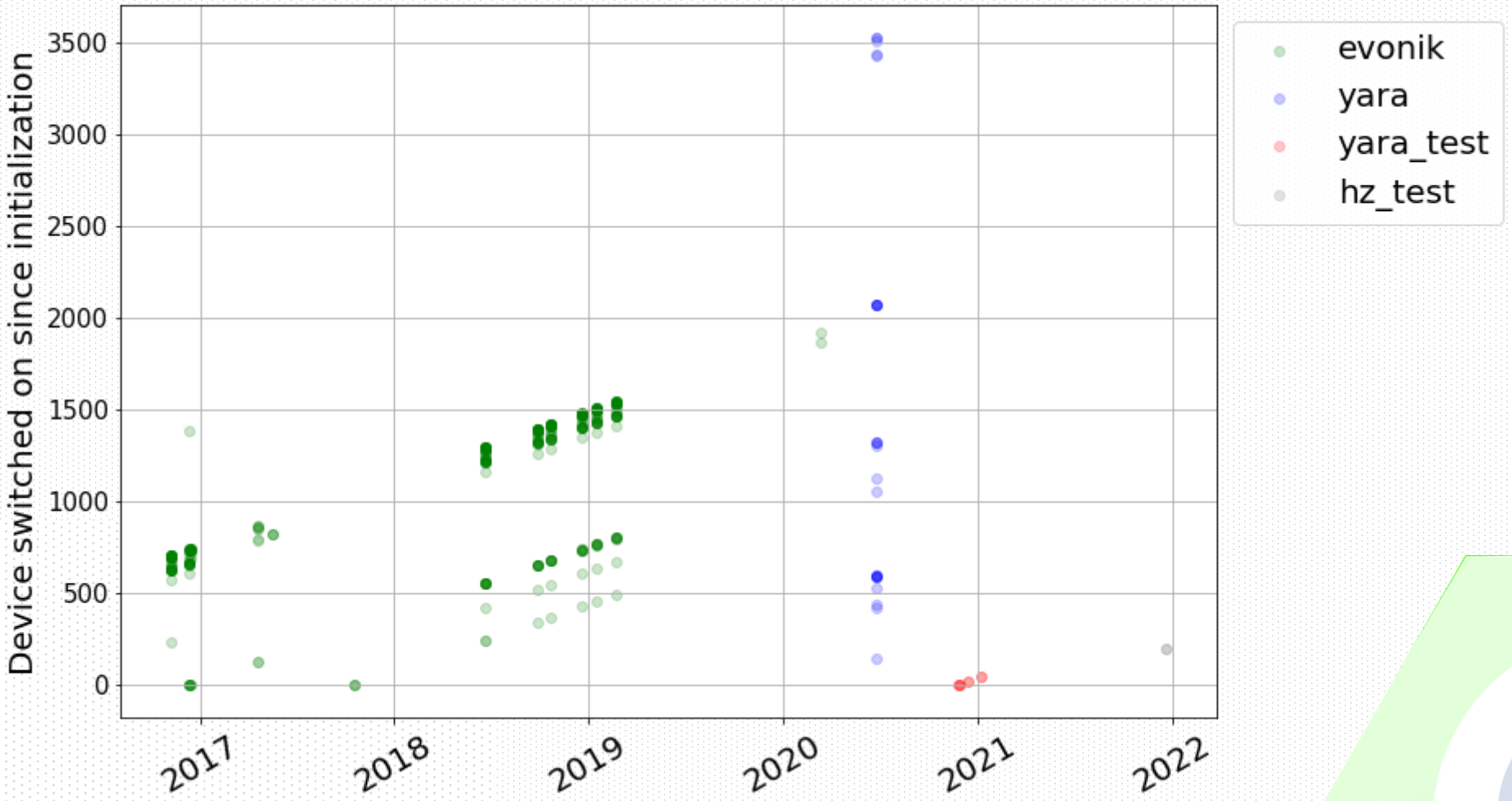
Analysis of available datasets

- Samson positioner data
  - Yara, Evonik, student experiment
  - Binary file that needs manual export
  - Semi structured XML format
  - Missing information on timing
  - Small number of available snapshots

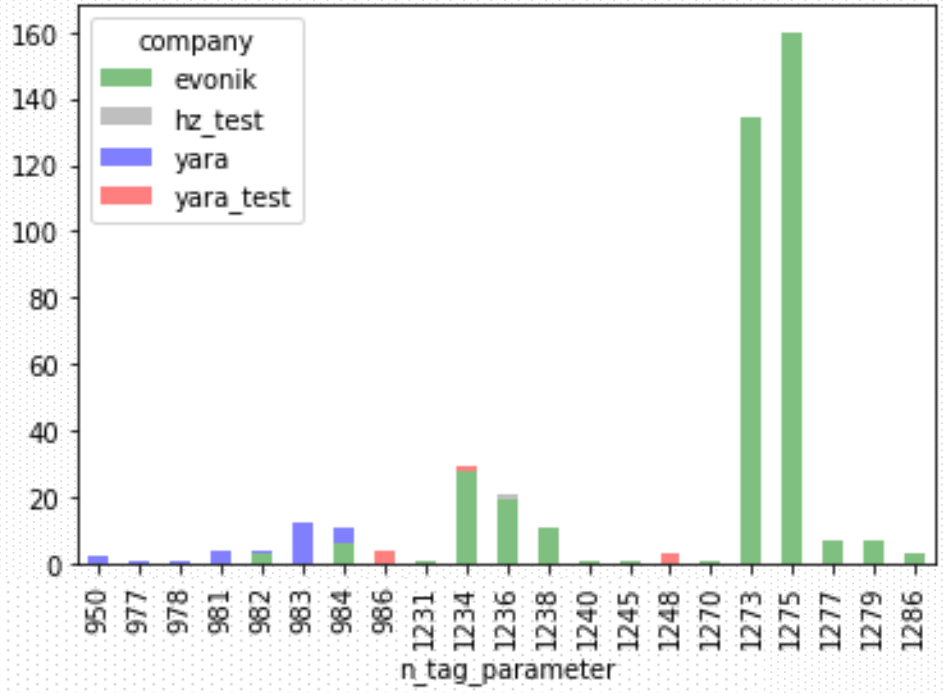


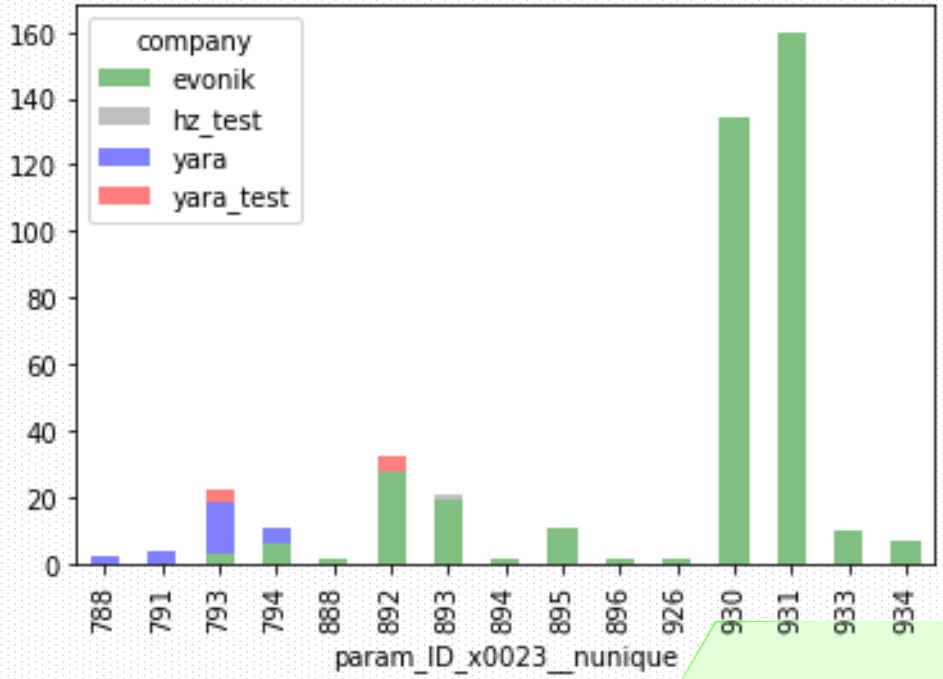
- Yara historian
  - Process data/ trendminer tool
  - Not available for analysis

Samson positioner data: timing



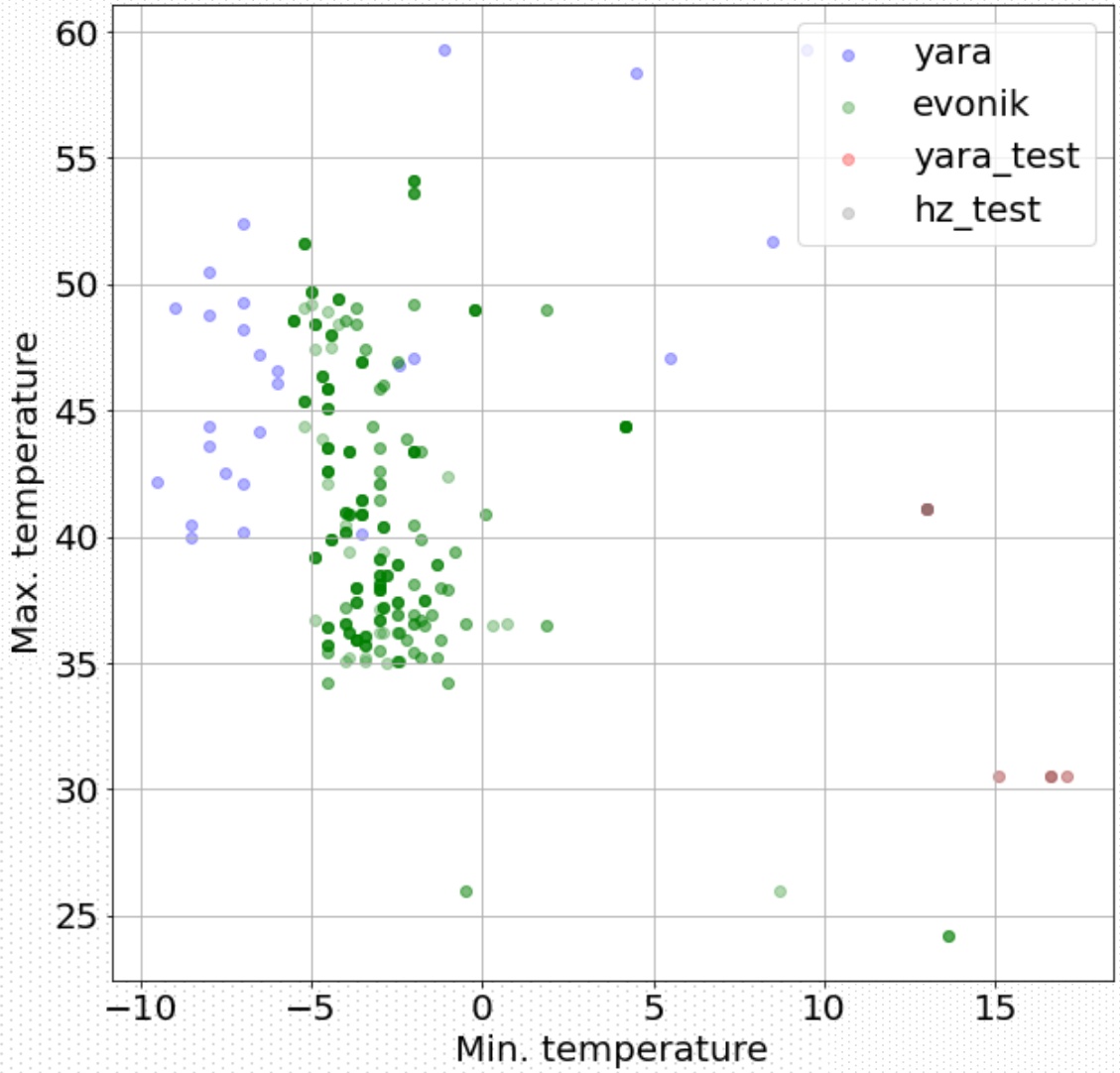
Samson positioner data: parameters

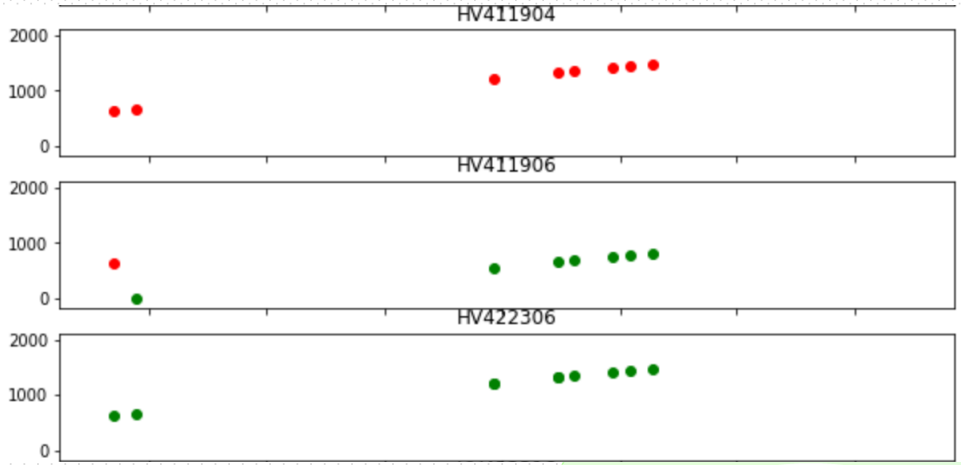




focus on parameters that occur **once** in **every** file

Samson positioner data: explorative analysis





Maintenance required

Analysis of available datasets: conclusion

It is hard to have the right data available for analysis

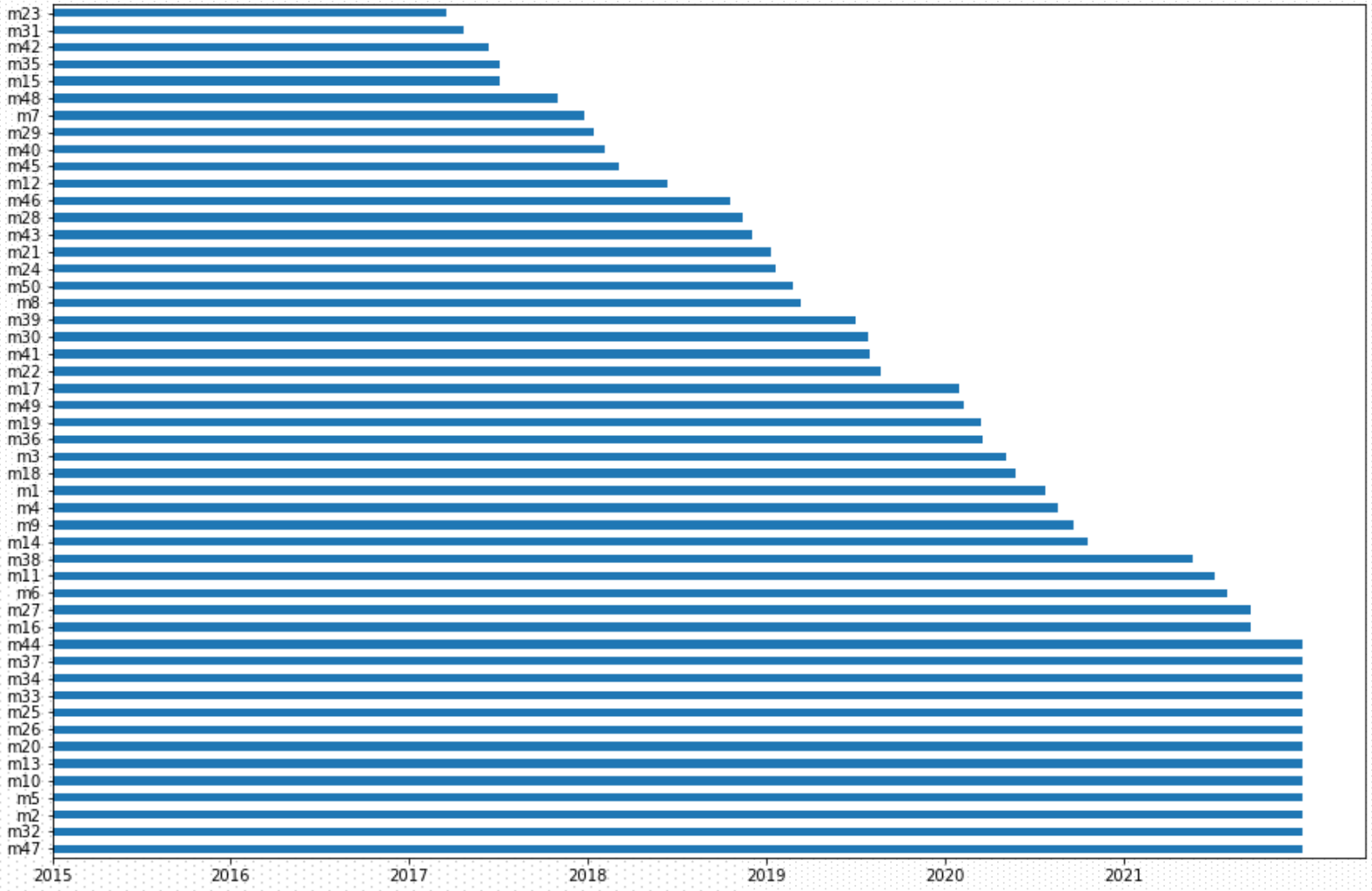
- Information on timing is crucial for analysis
- Data on positioners are gathered with a different goal, by a third party
- Security considerations limit:
  - Automated/ wireless data gathering
  - Sharing of data
- Failures (important as target) are rare
- Data about maintenance and failures are not gathered in a standardized way

To illustrate the analysis opportunities in a workshop, data simulation is necessary

Data simulation and analysis results only targets → survival curve estimation

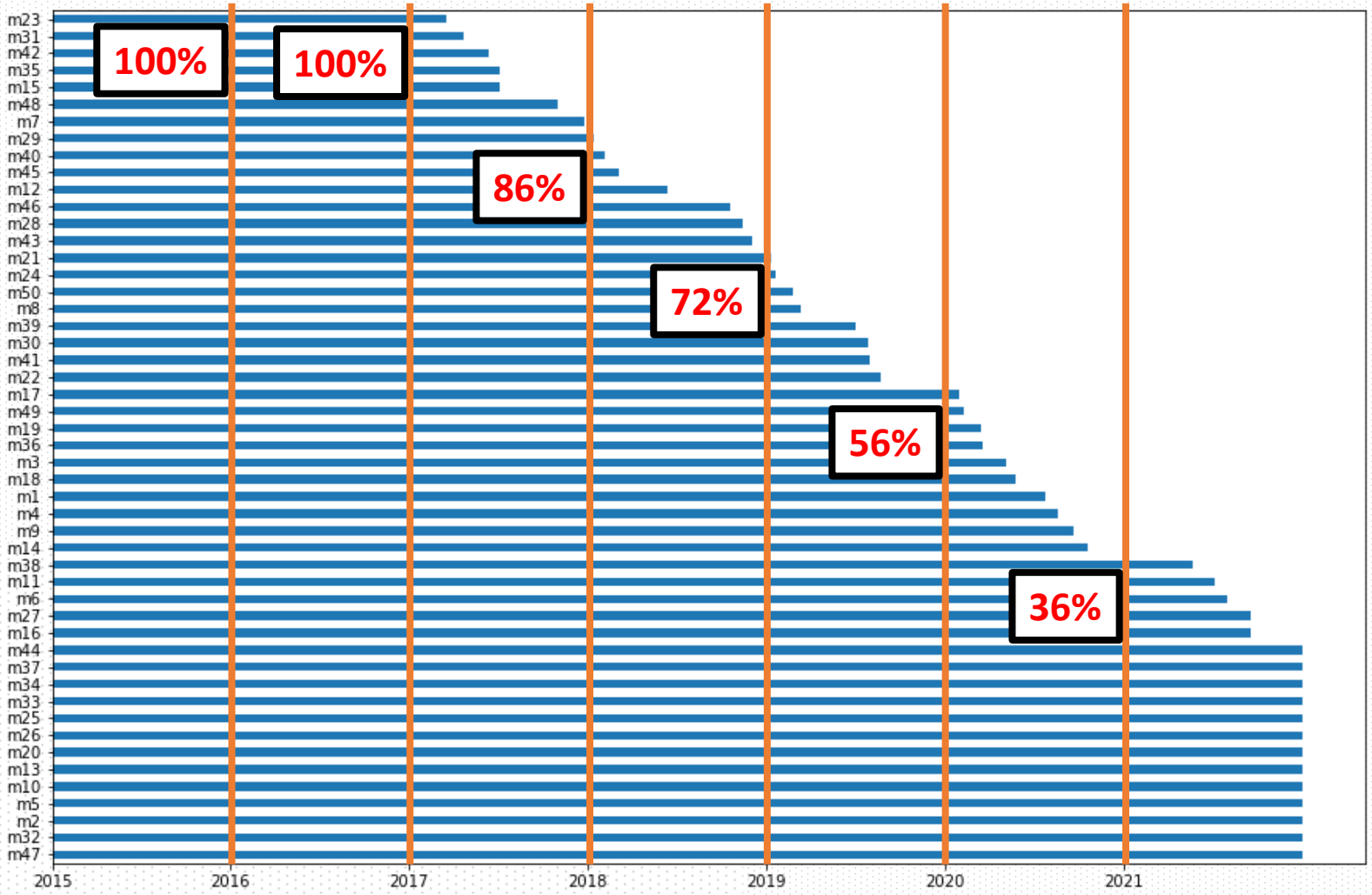
- 50 machines, 7 year follow up
- question: how long will an asset function without problems, from the moment it is installed

Data simulation and analysis results only targets → survival curve estimation



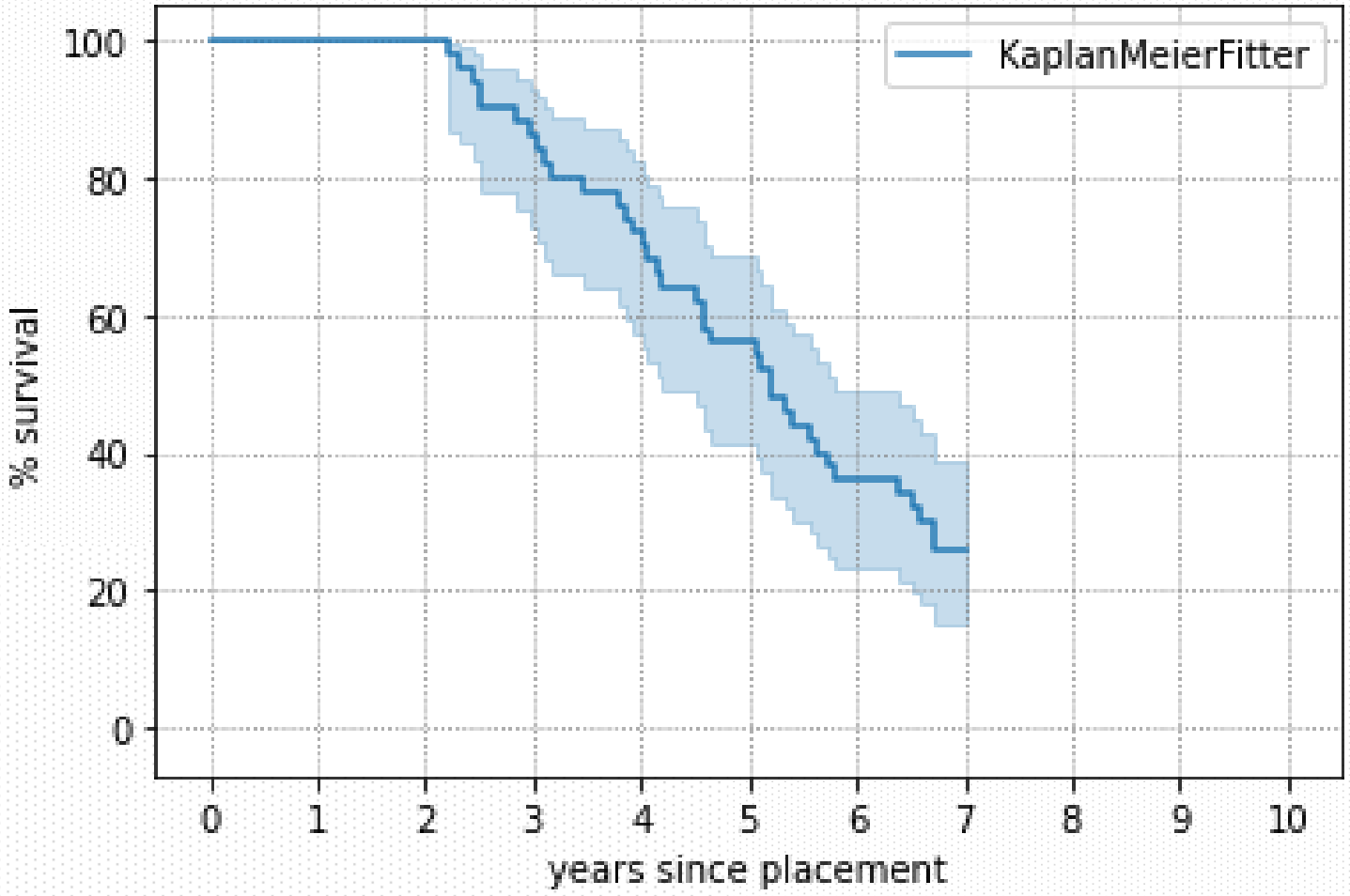
- sort by duration of functioning without problems

Data simulation and analysis results only targets → survival curve estimation



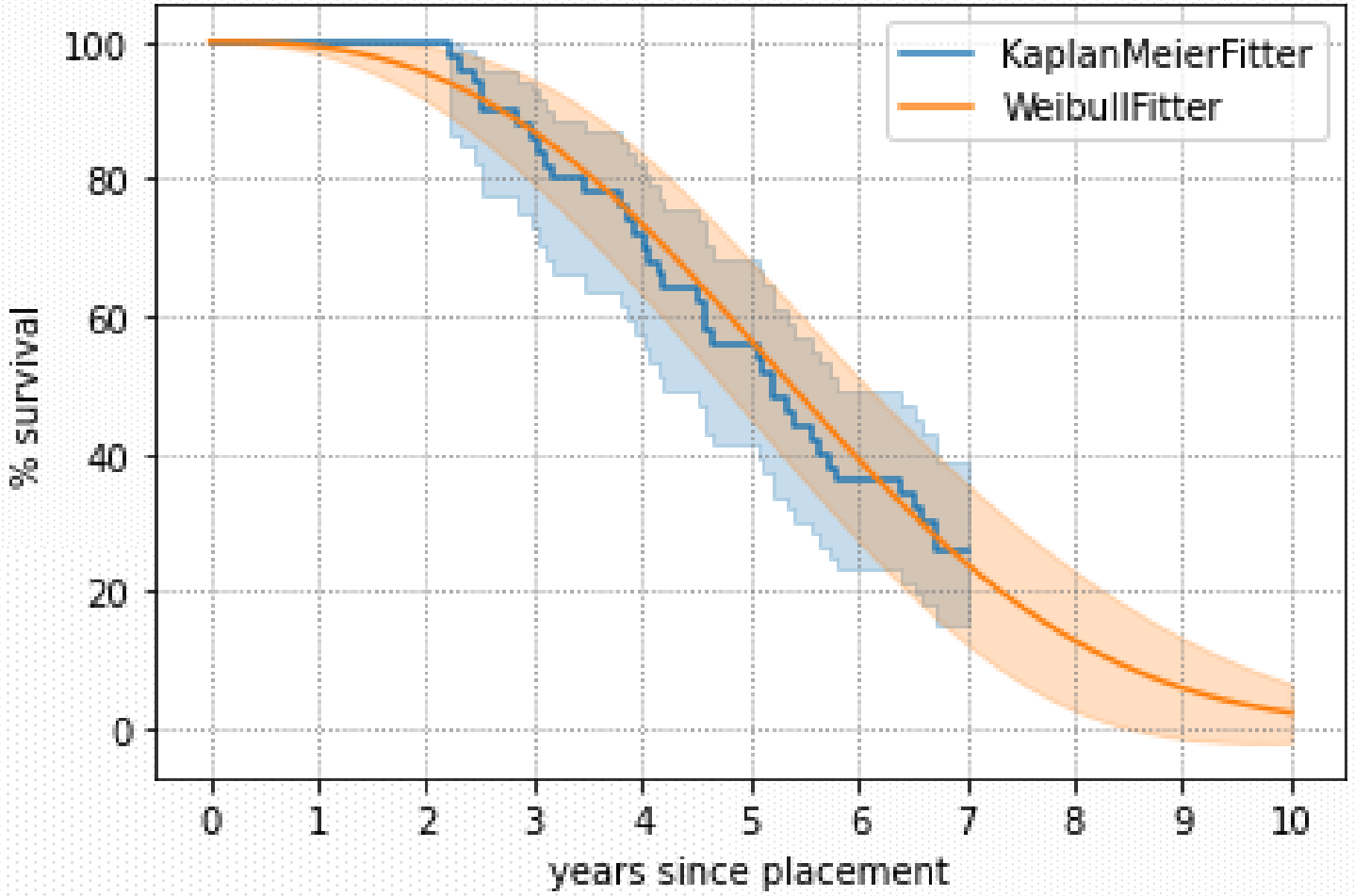
- calculate survival after every year

Data simulation and analysis results only targets → survival curve estimation



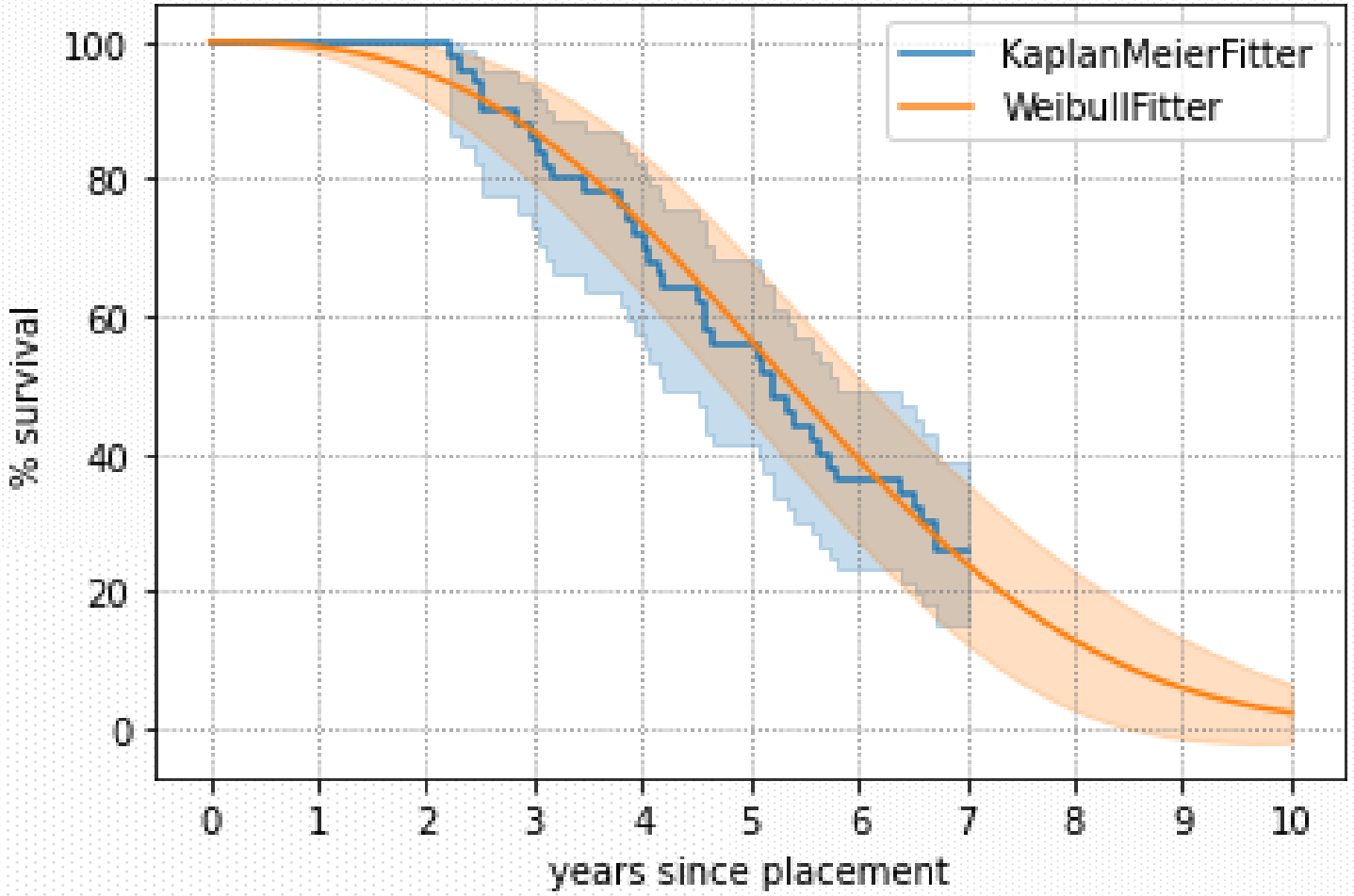
- survival curve, including uncertainty of estimates

Data simulation and analysis results only targets → survival curve estimation



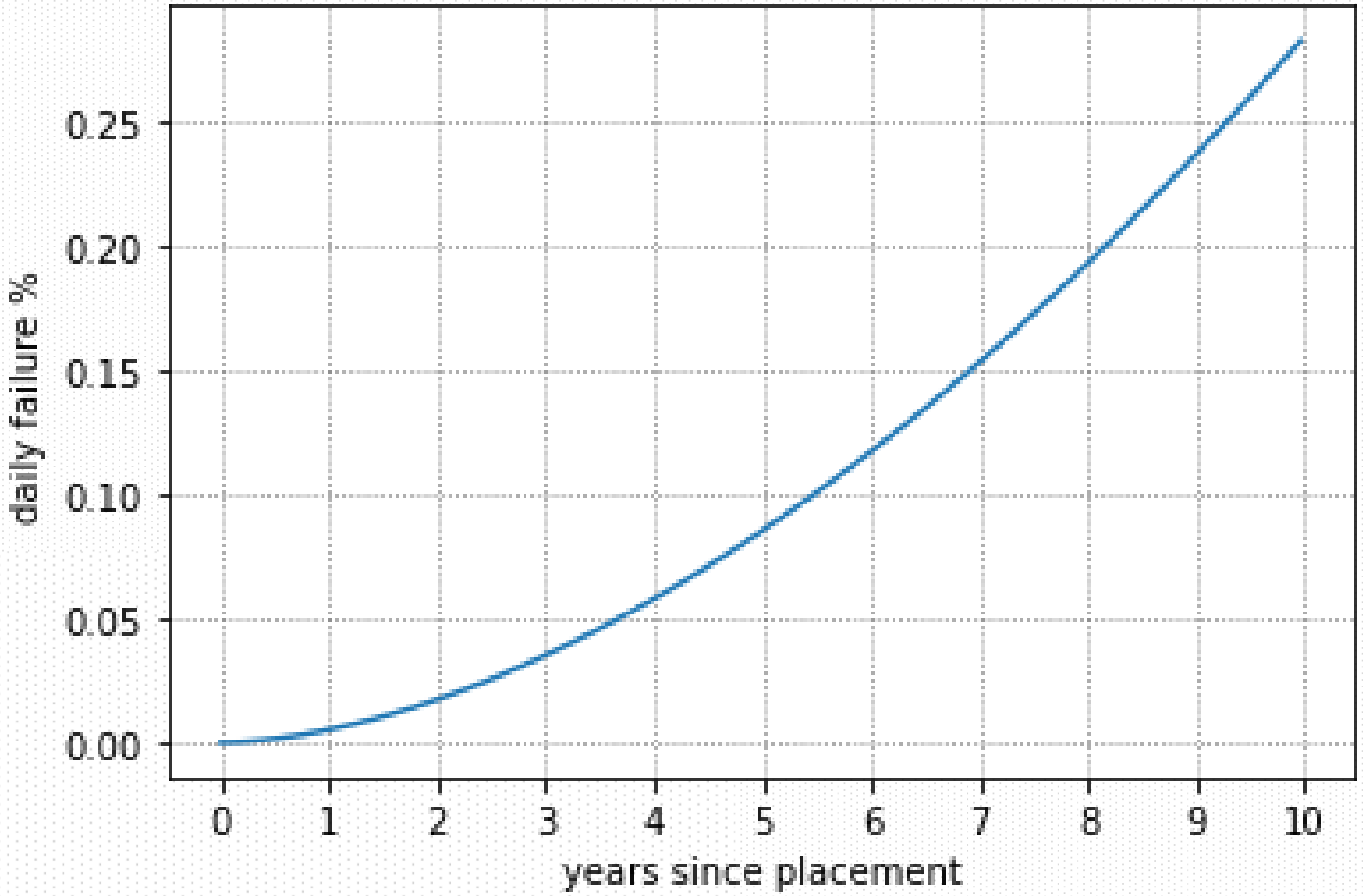
- model to extrapolate into the future
- how can this result be applied in practice?

Data simulation and analysis results only targets → survival curve estimation



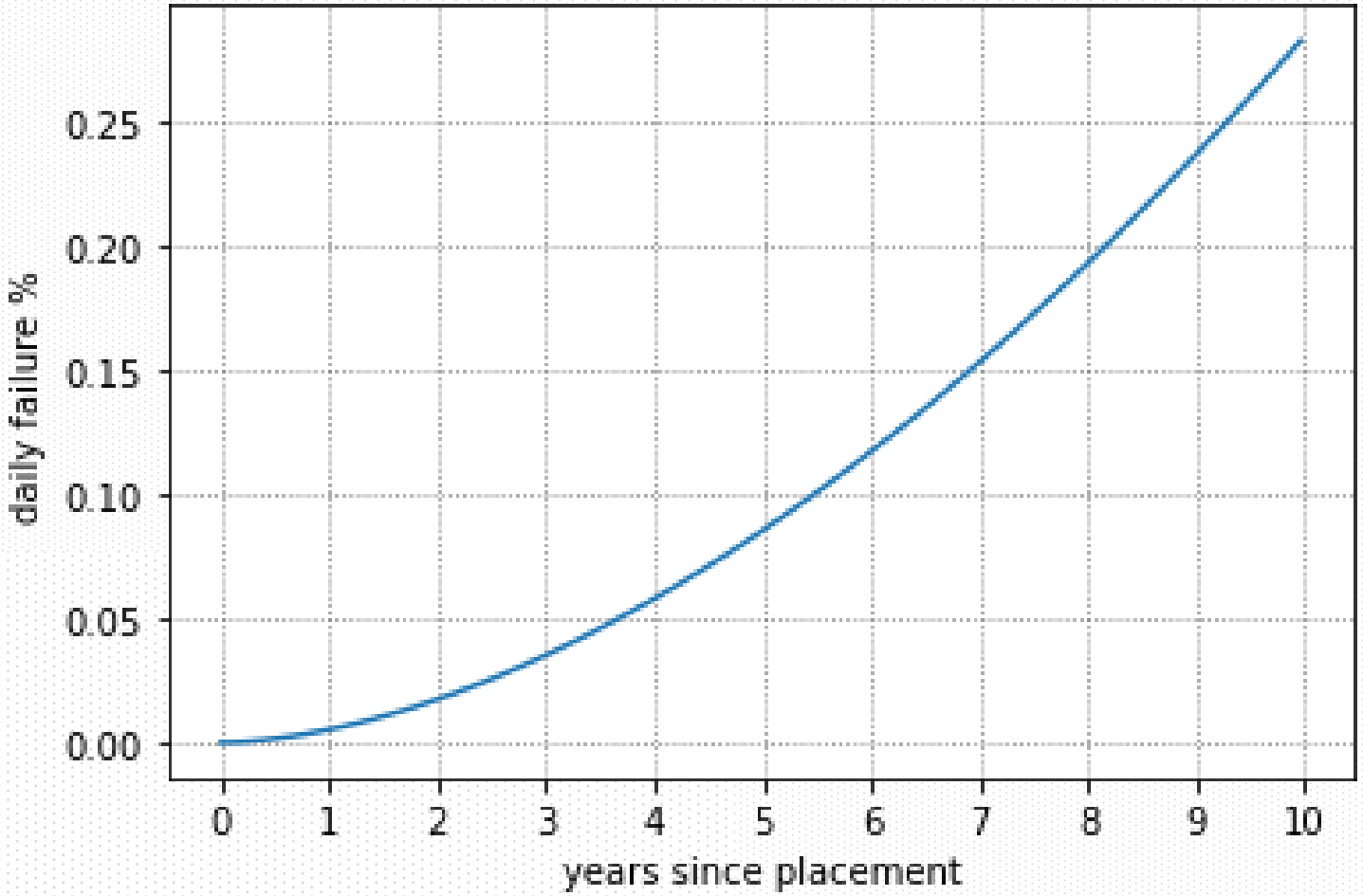
- model to extrapolate into the future
- → predict cost of reactive replacement strategy, based on age composition

Data simulation and analysis results only targets → survival curve estimation



- daily failure % increases with time
- how can this result be applied in practice?

Data simulation and analysis results only targets → survival curve estimation



- daily failure % increases with time
- → prescribe proactive replacement strategy, based on acceptable failure percentage

Discussion: challenges in practice

Required: time until (first) failure for all assets of the same type

- Not all assets have the same type – when time passes, new models become available
  - Information on new models is most useful, but hardest to obtain

Discussion: challenges in practice

Required: time until (first) failure for all assets of the same type

- Not all assets have the same type – when time passes, new models become available
  - Information on new models is most useful, but hardest to obtain
- Usage (instead of age) might better explain wear

UNIVERSITY OF APPLIED SCIENCES

Discussion: challenges in practice

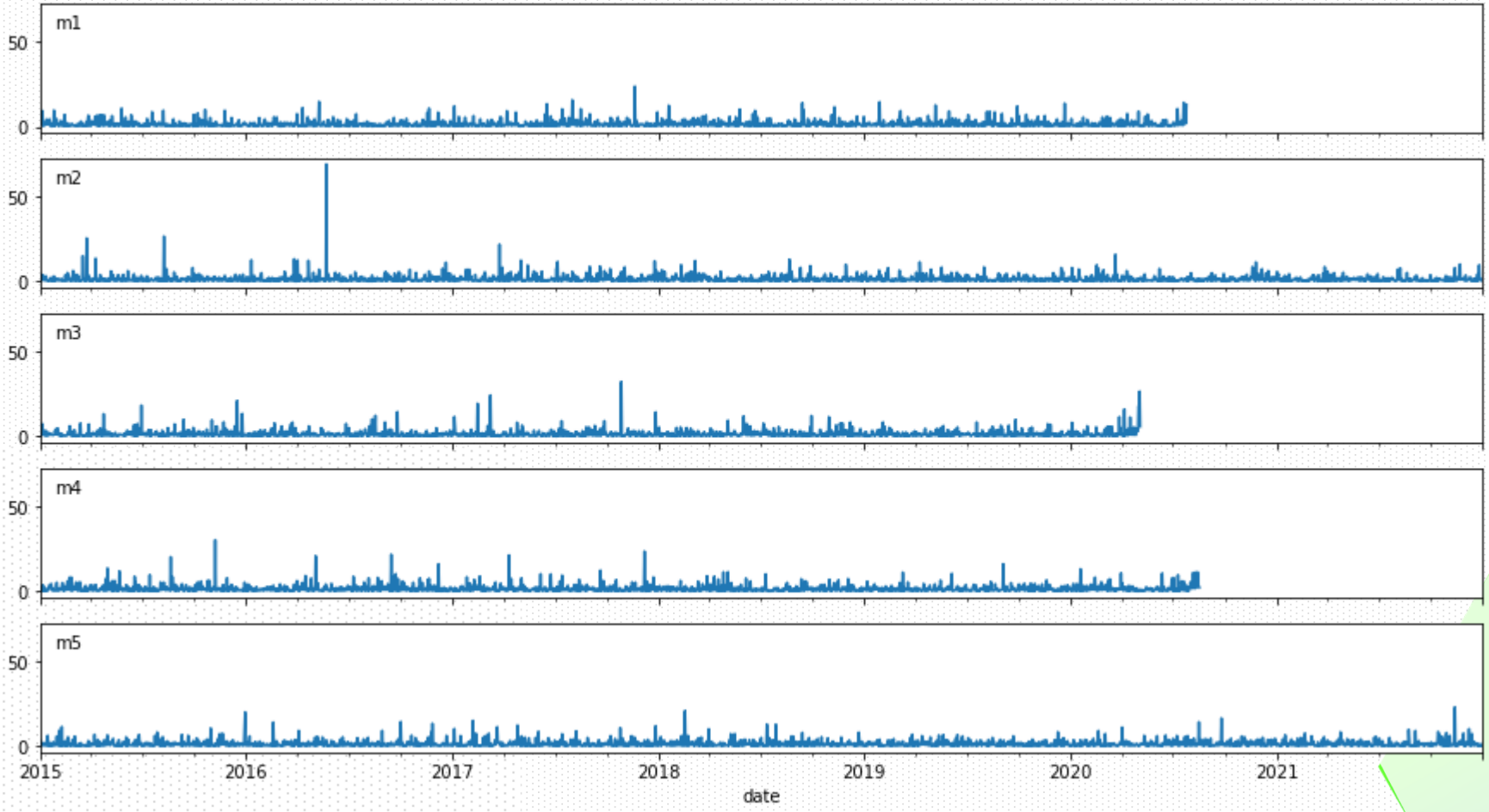
Required: time until (first) failure for all assets of the same type

- Not all assets have the same type – when time passes, new models become available
  - Information on new models is most useful, but hardest to obtain
- Usage (instead of age) might better explain wear
- Different types of failure may need separate analysis
  - infant mortality vs wear out (bathtub curve)
  - treatment (repair/ replace) and cost
  - missing data: failure type/ cause, timing

UNIVERSITY OF APPLIED SCIENCES

Data simulation and analysis results only features → outlier detection

Vibration data (only m1 to m5 shown, of 50)

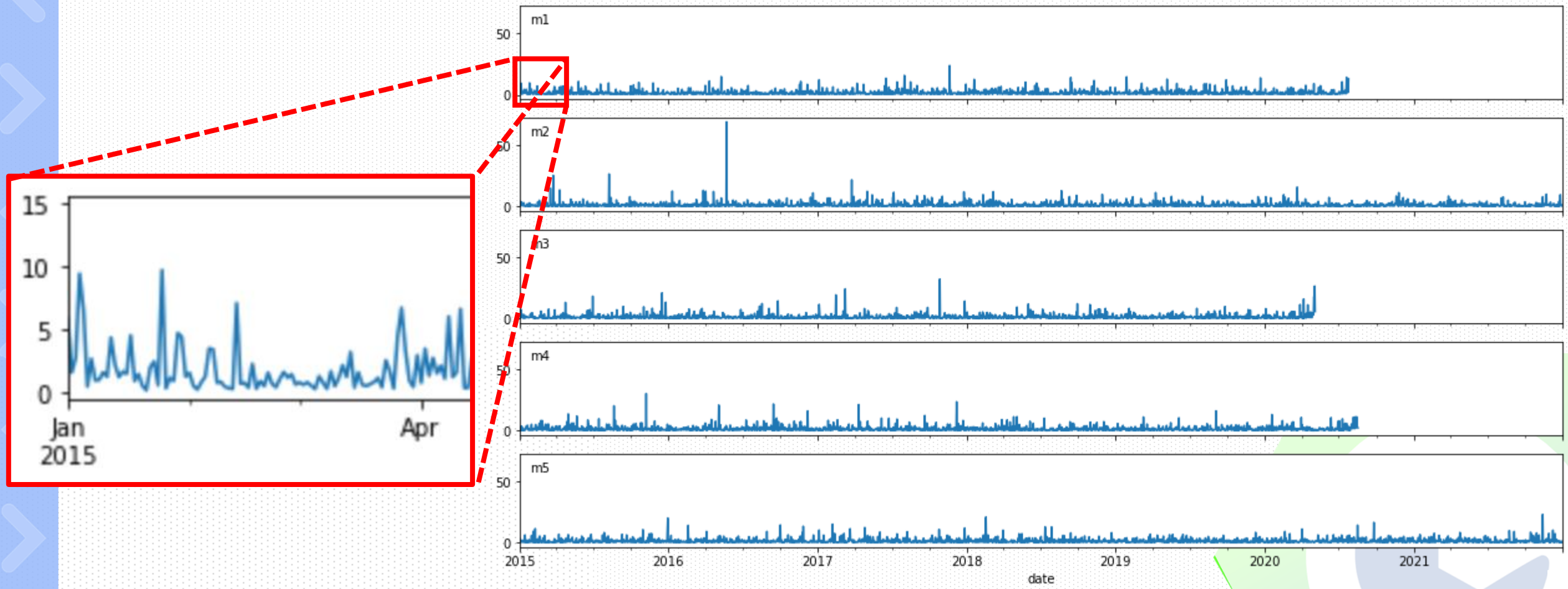


Optional data preparation

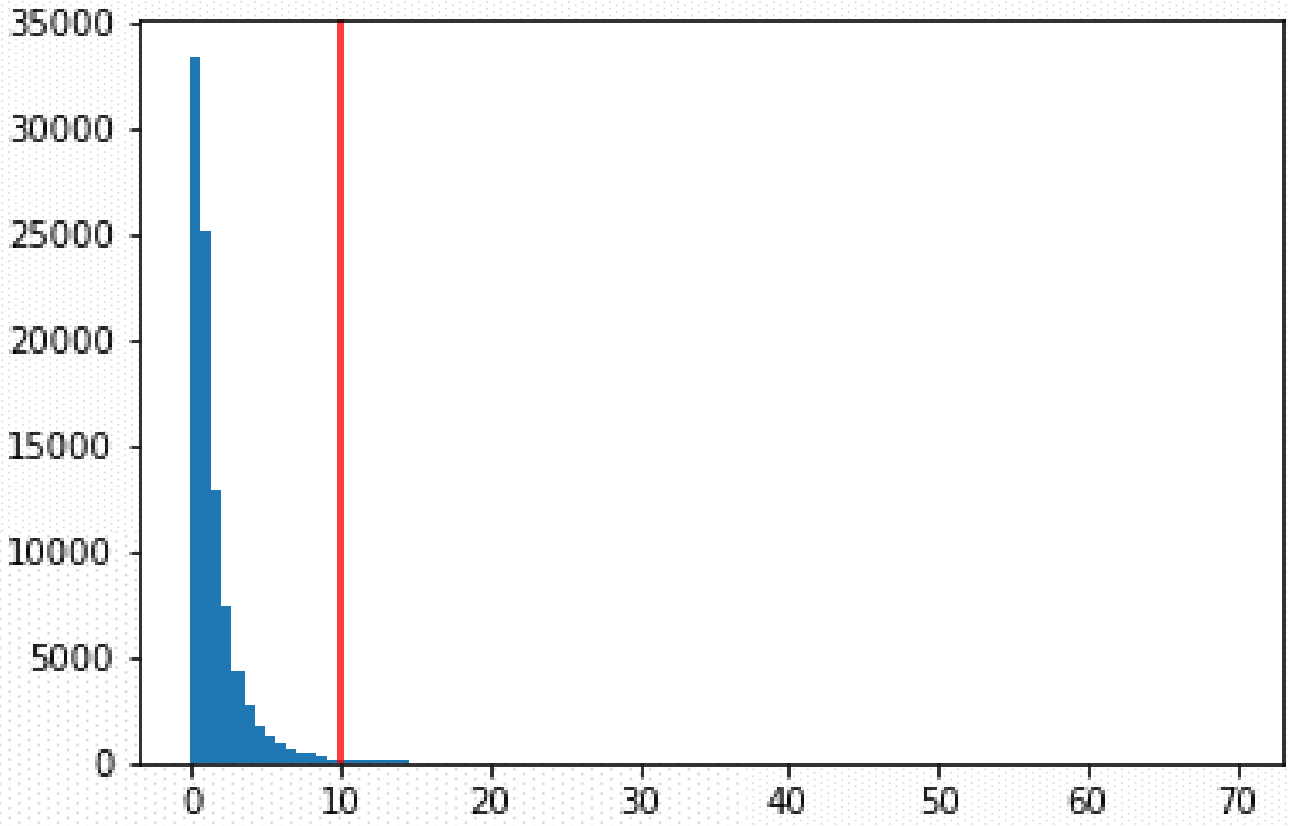
- derivatives
- Fourier transform

Data simulation and analysis results only features → outlier detection

Vibration data (only m1 to m5 shown, of 50)



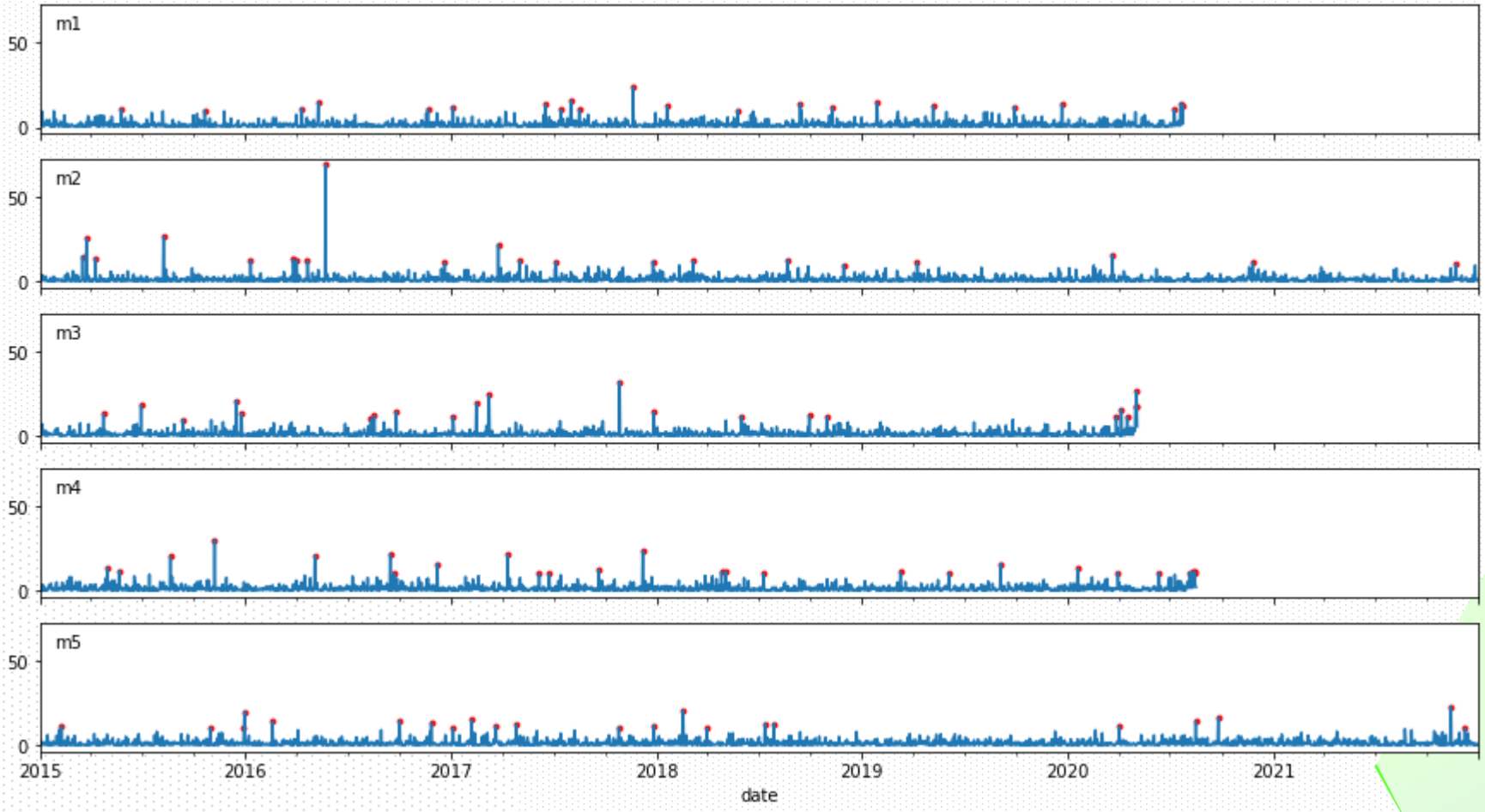
Data simulation and analysis results only features → outlier detection



- Histogram to identify a threshold for extreme values
- Can be done without targets

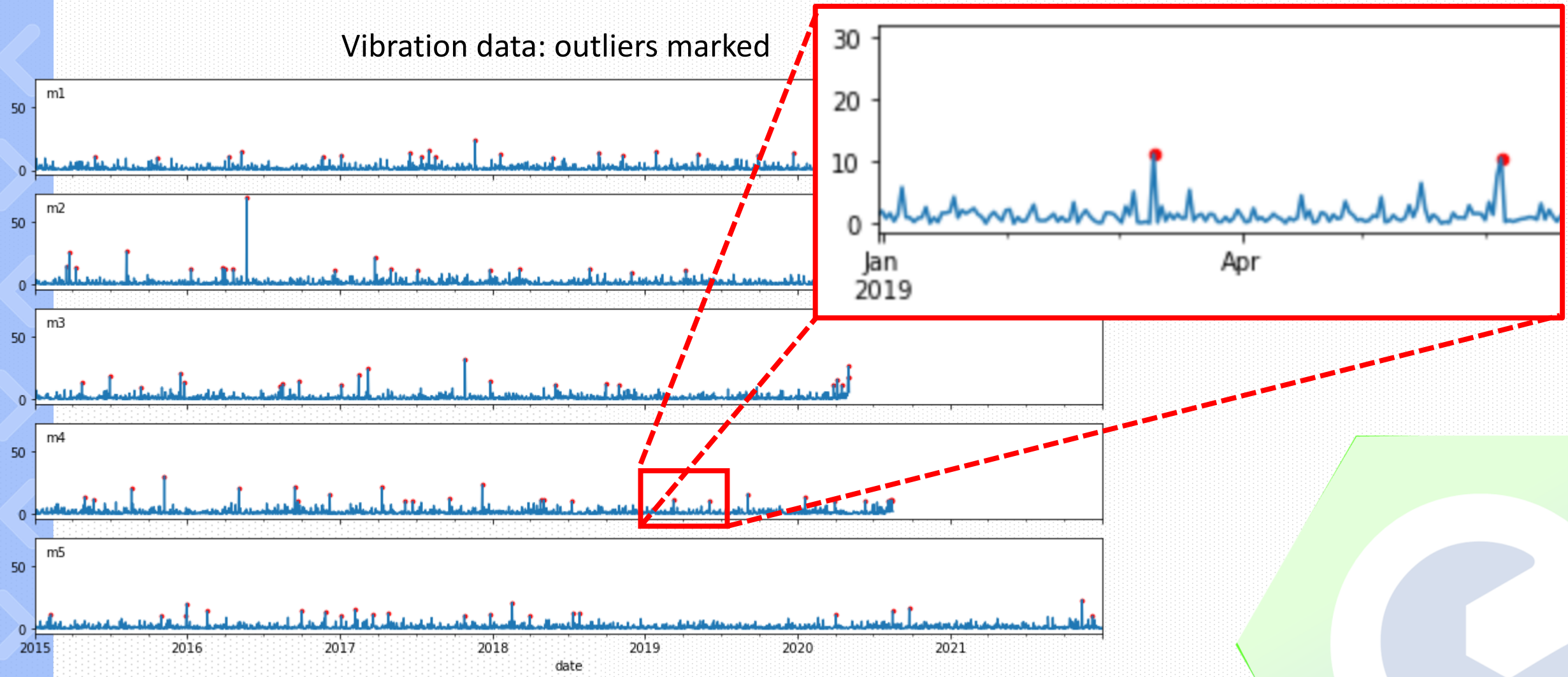
Data simulation and analysis results only features → outlier detection

Vibration data: outliers marked

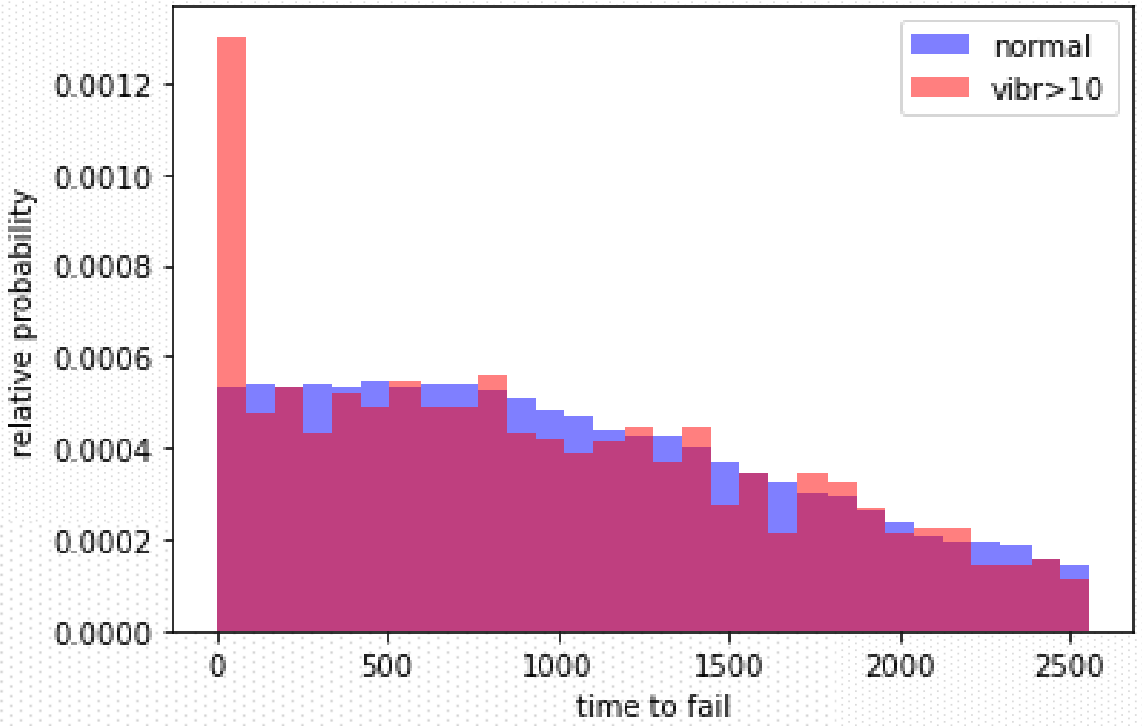


Data simulation and analysis results only features → outlier detection

Vibration data: outliers marked



Data simulation and analysis results only features → outlier detection

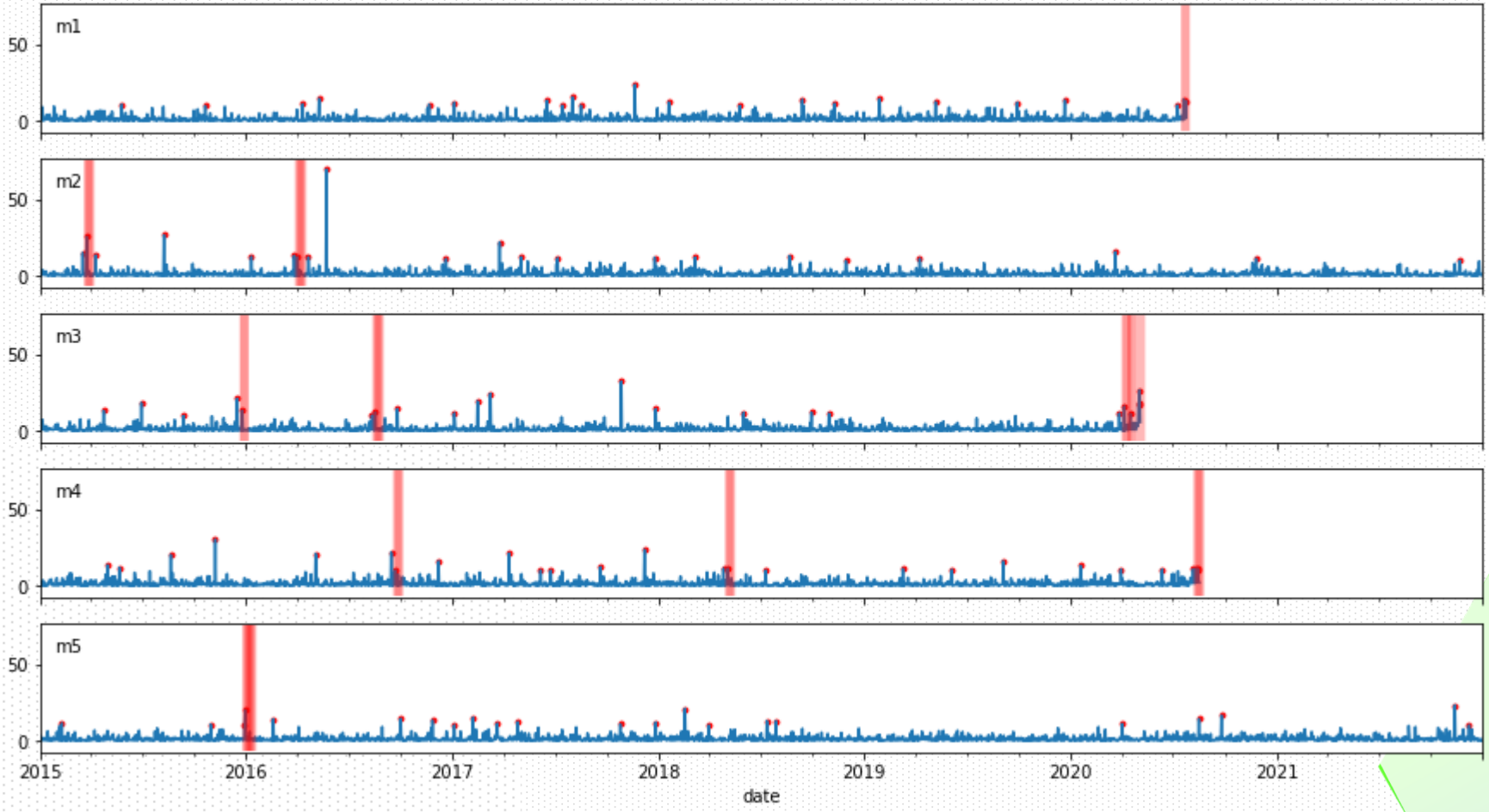


- validation by looking at available targets

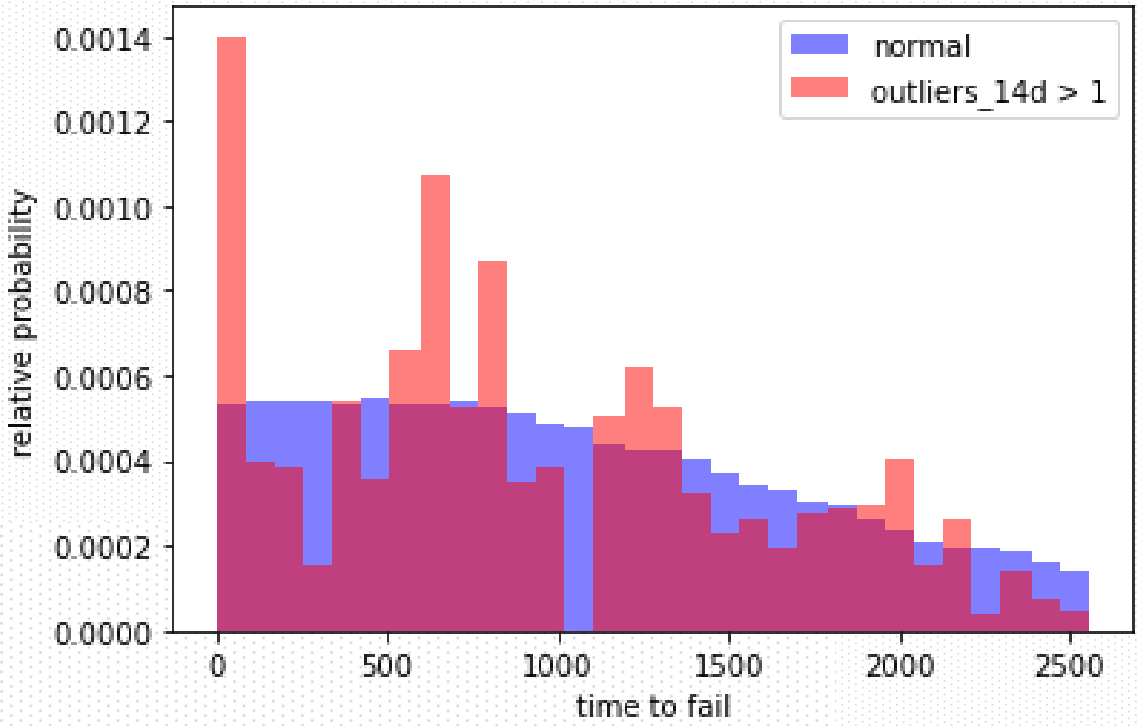
outlier	time_to_fail	ttf<7	ttf<30	count
False	1024.272	0.003	0.015	92012
True	957.572	0.059	0.086	1055

Data simulation and analysis results only features → outlier detection

Vibration data: 14-day-periods with multiple outliers highlighted



Data simulation and analysis results only features → outlier detection



outliers_14d	time_to_fail	ttf<7	ttf<30	count
0.0	1025.667	0.002	0.013	80090
1.0	1016.379	0.009	0.028	11996
2.0	963.071	0.050	0.070	927
3.0	534.630	0.435	0.435	46
4.0	1.143	1.000	1.000	7
5.0	0.000	1.000	1.000	1

→ set threshold, based on validation

Discussion: challenges in practice

- Cost of sensors (installation, maintenance)
  - collection and storage
  - data and metadata

Discussion: challenges in practice

- Cost of sensors (installation, maintenance)
  - collection and storage
  - data and metadata

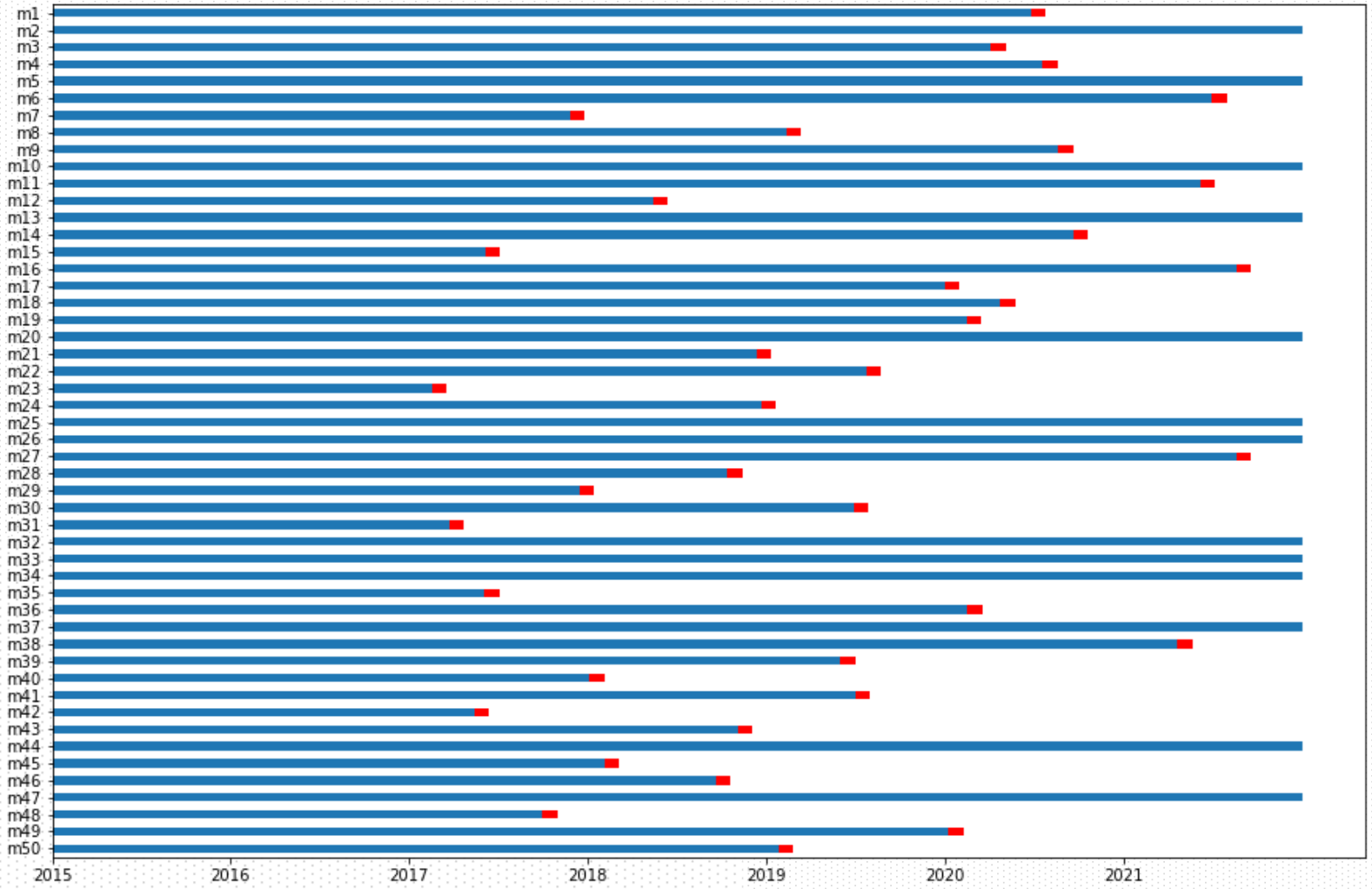
- To gather vibration data, with influences from the environment
  - weather: wind, rain, thunder, temperature
  - condition of other components
  - activity of other assets
  - (non random) failure of sensors

Discussion: challenges in practice

- Cost of sensors (installation, maintenance)
  - collection and storage
  - data and metadata
- To gather vibration data, with influences from the environment
  - weather: wind, rain, thunder, temperature
  - condition of other components
  - activity of other assets
  - (non random) failure of sensors
- To set a threshold for outliers when no targets are available
  - assumed relation between feature and targets
  - threshold cannot be validated

UNIVERSITY OF APPLIED SCIENCES

Data simulation and analysis results targets and features → failure probability

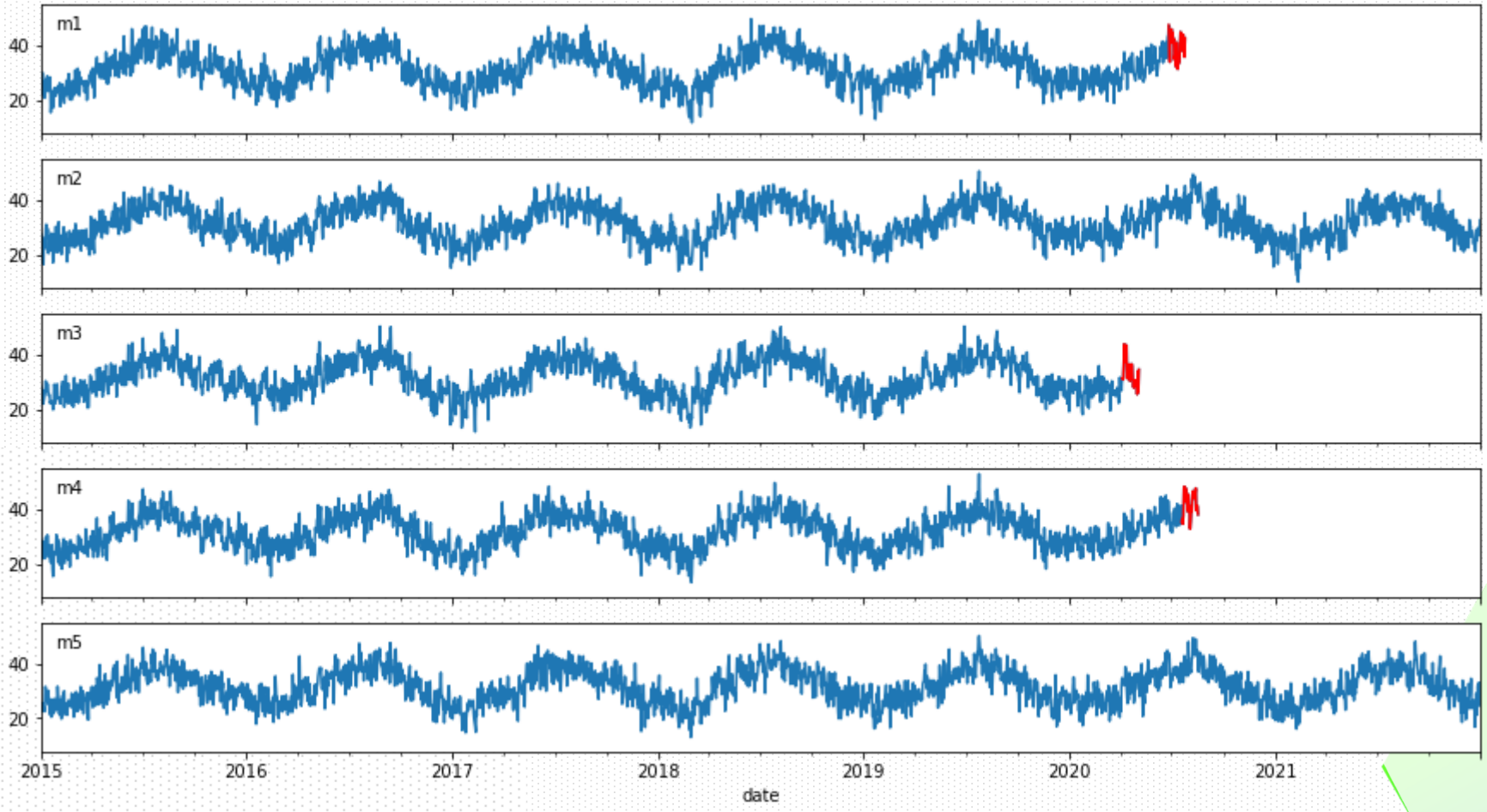


- Focus: how to recognize the moment when a failure will happen within 30 days (red areas)

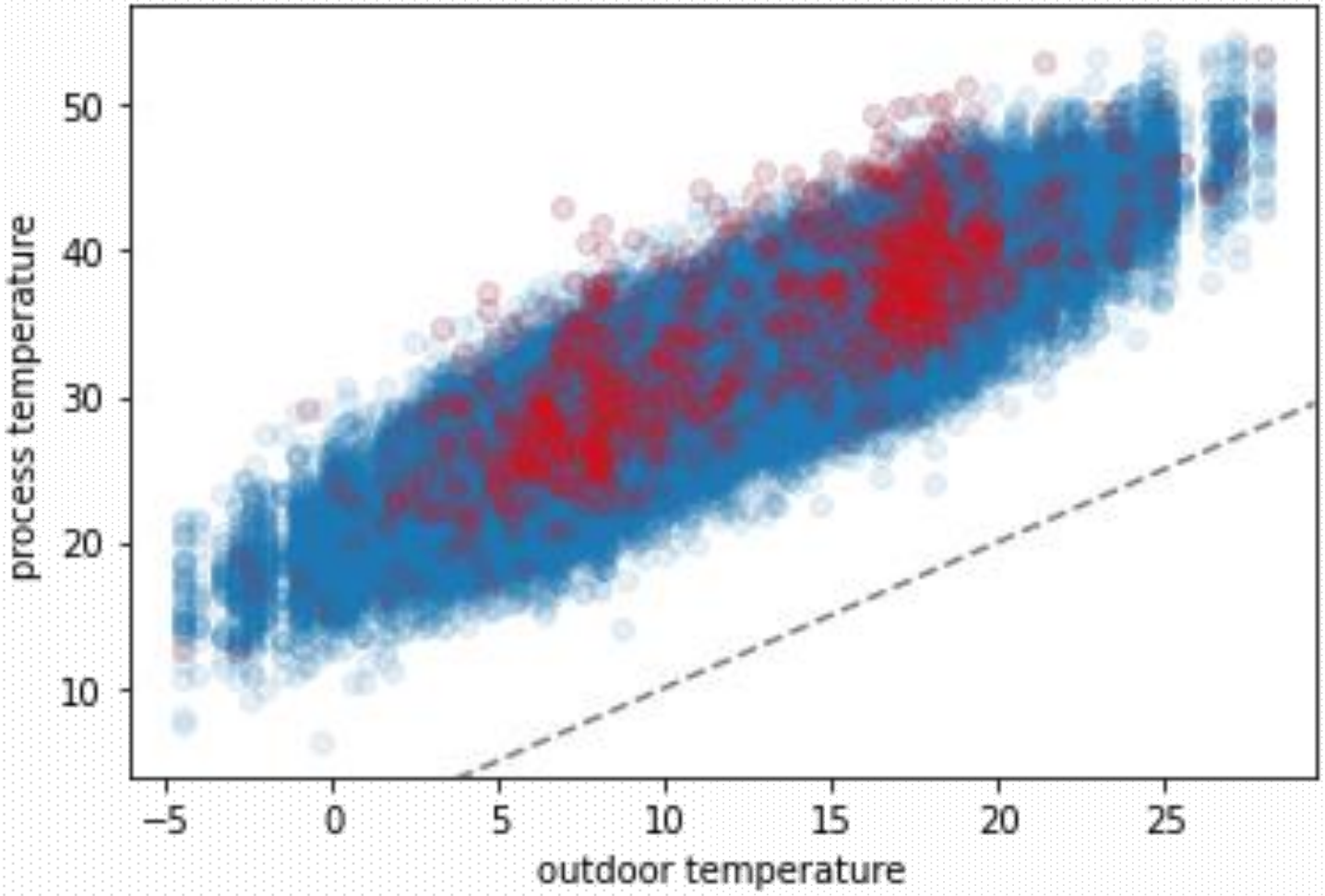
Data simulation and analysis results targets and features → failure probability

Interreg Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling

Process temperature



Data simulation and analysis results targets and features → failure probability



- Classification model needs to distinguish between red (failure within 30 days) and blue

Data simulation and analysis results targets and features → failure probability

no failure within 30 days → 91957

1110 ← # failure within 30 days

?

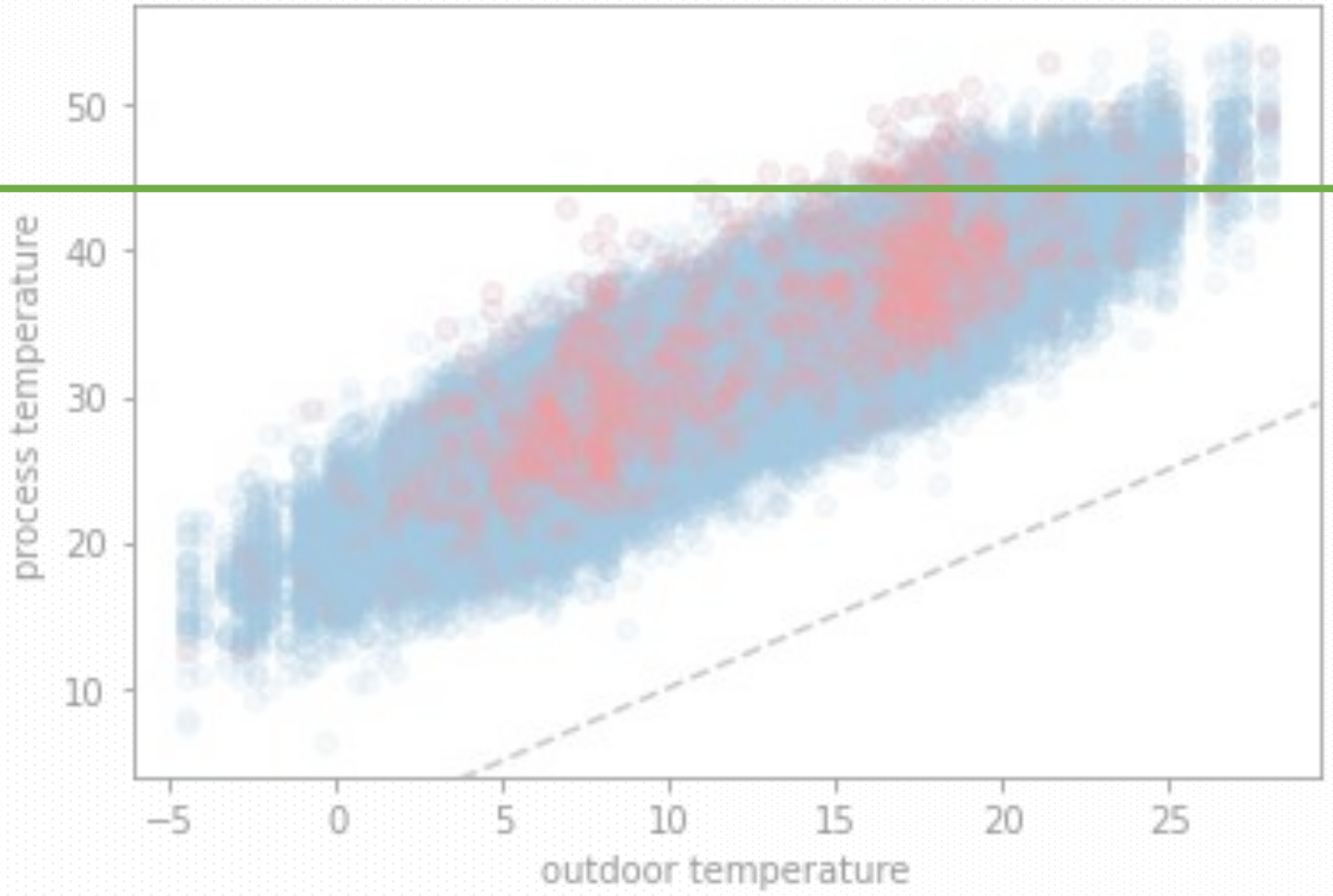
Decision Rule

Meant to split the dataset in parts that have a different probability of failure within 30 days

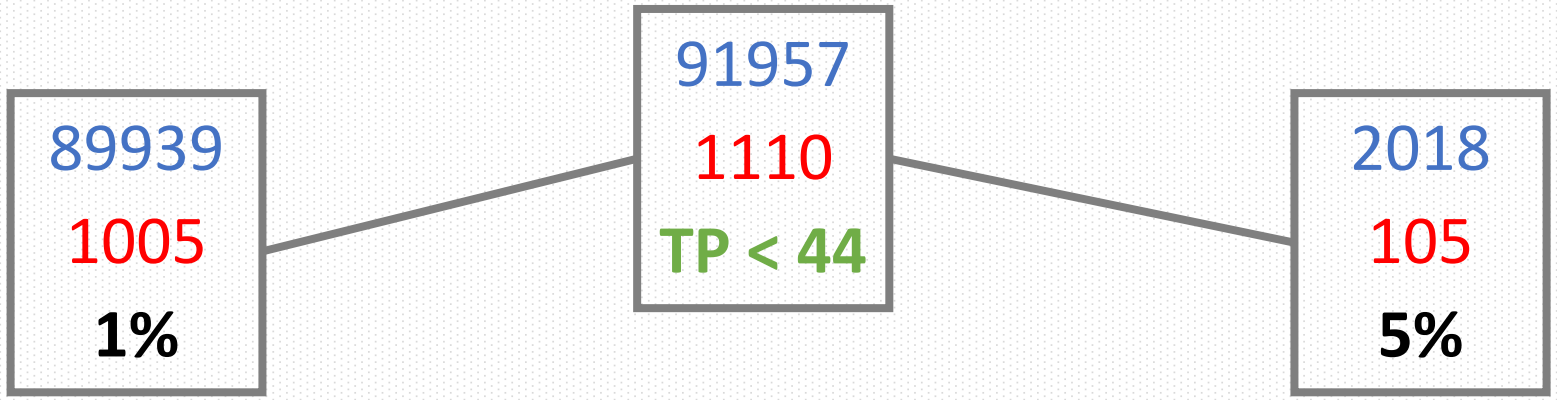
Many models

- linear
- neighbors
- neural nets

Data simulation and analysis results targets and features → failure probability

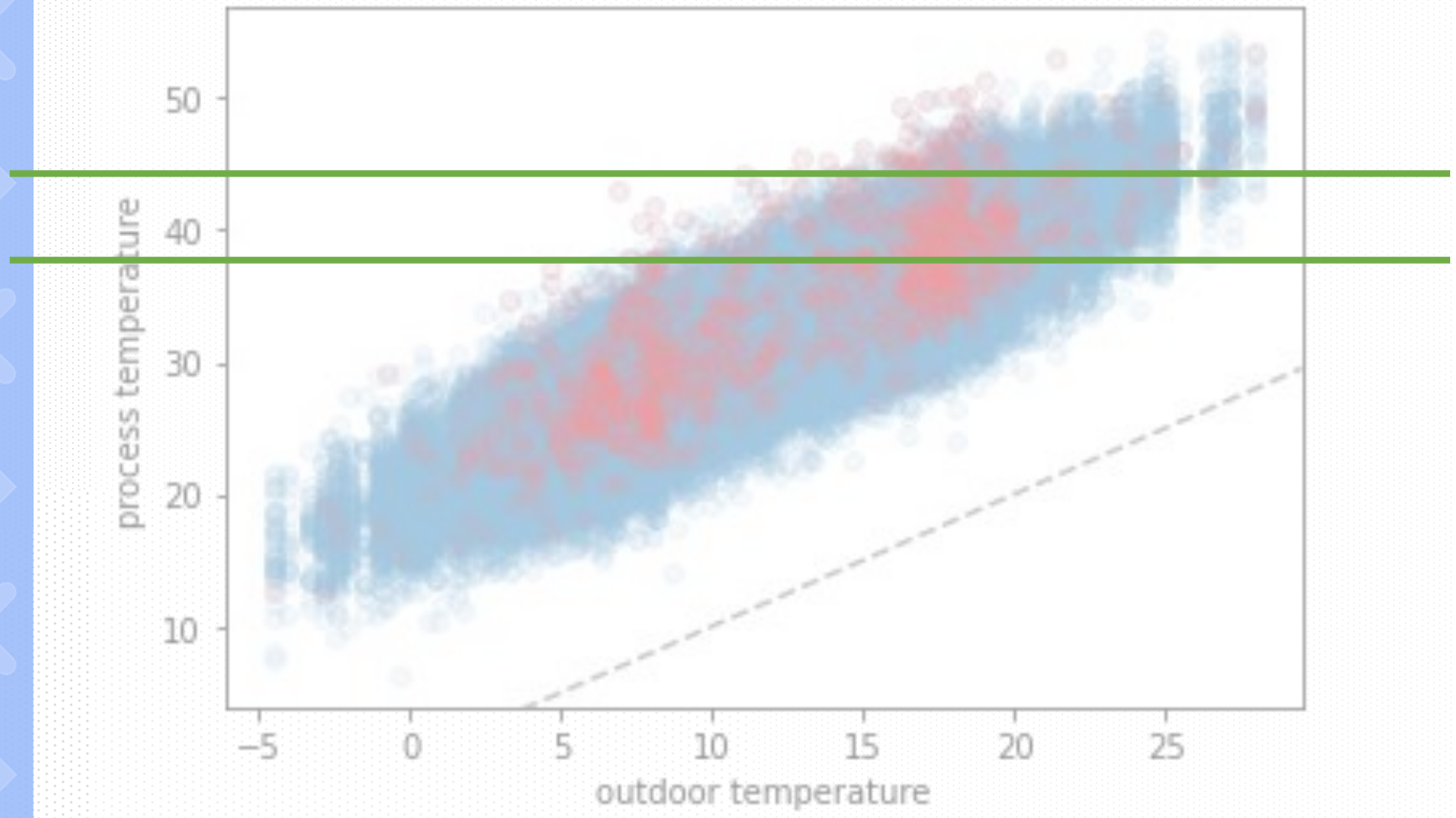


Data simulation and analysis results targets and features → failure probability

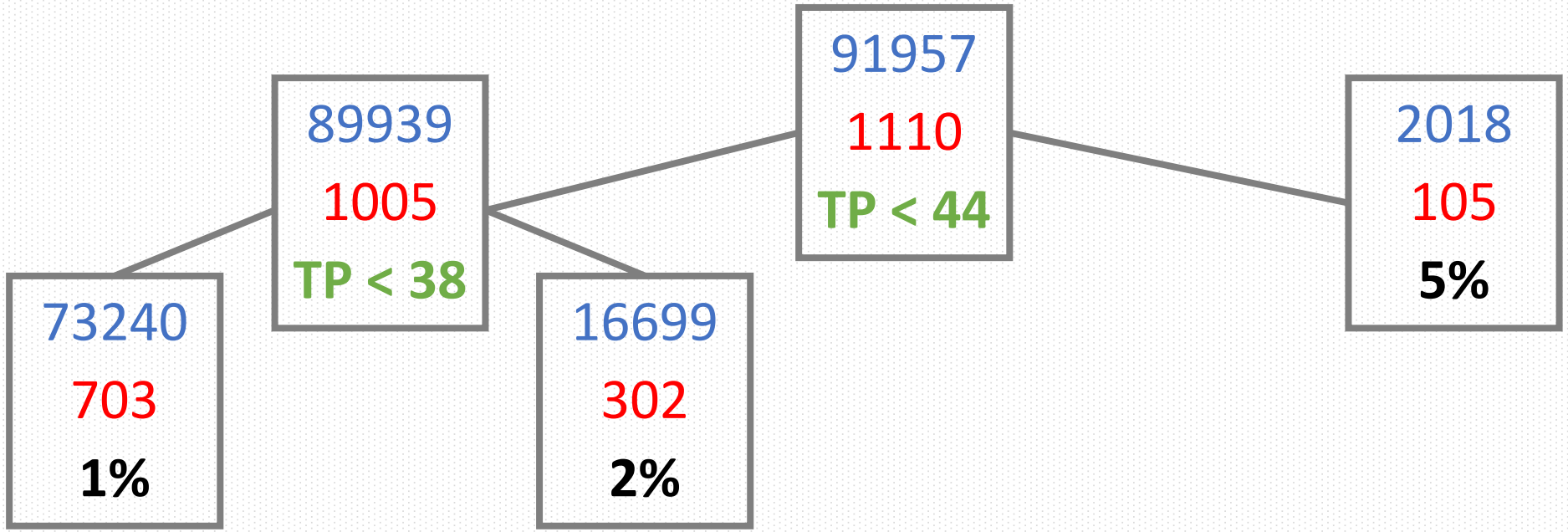


TP = Process Temperature

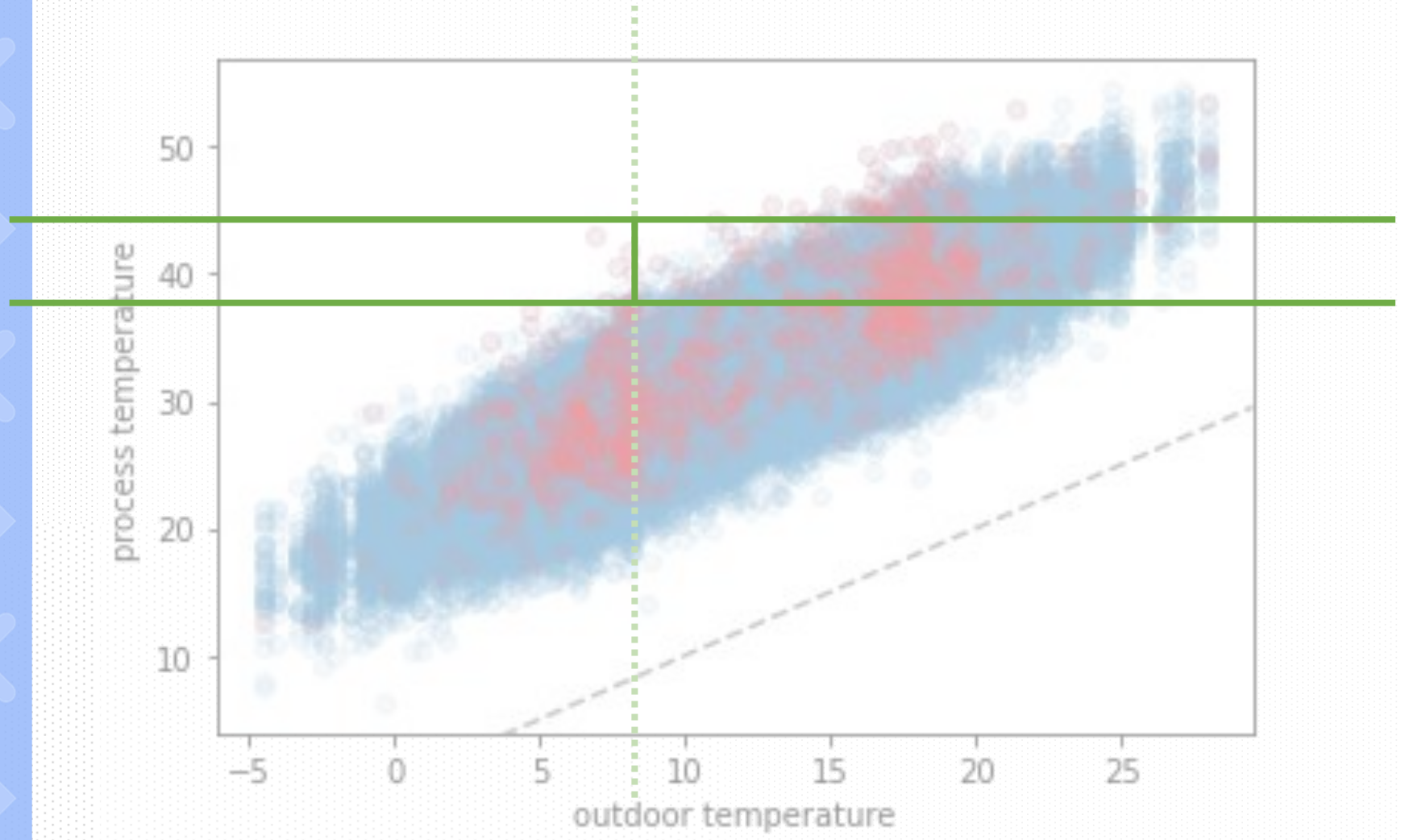
Data simulation and analysis results targets and features → failure probability



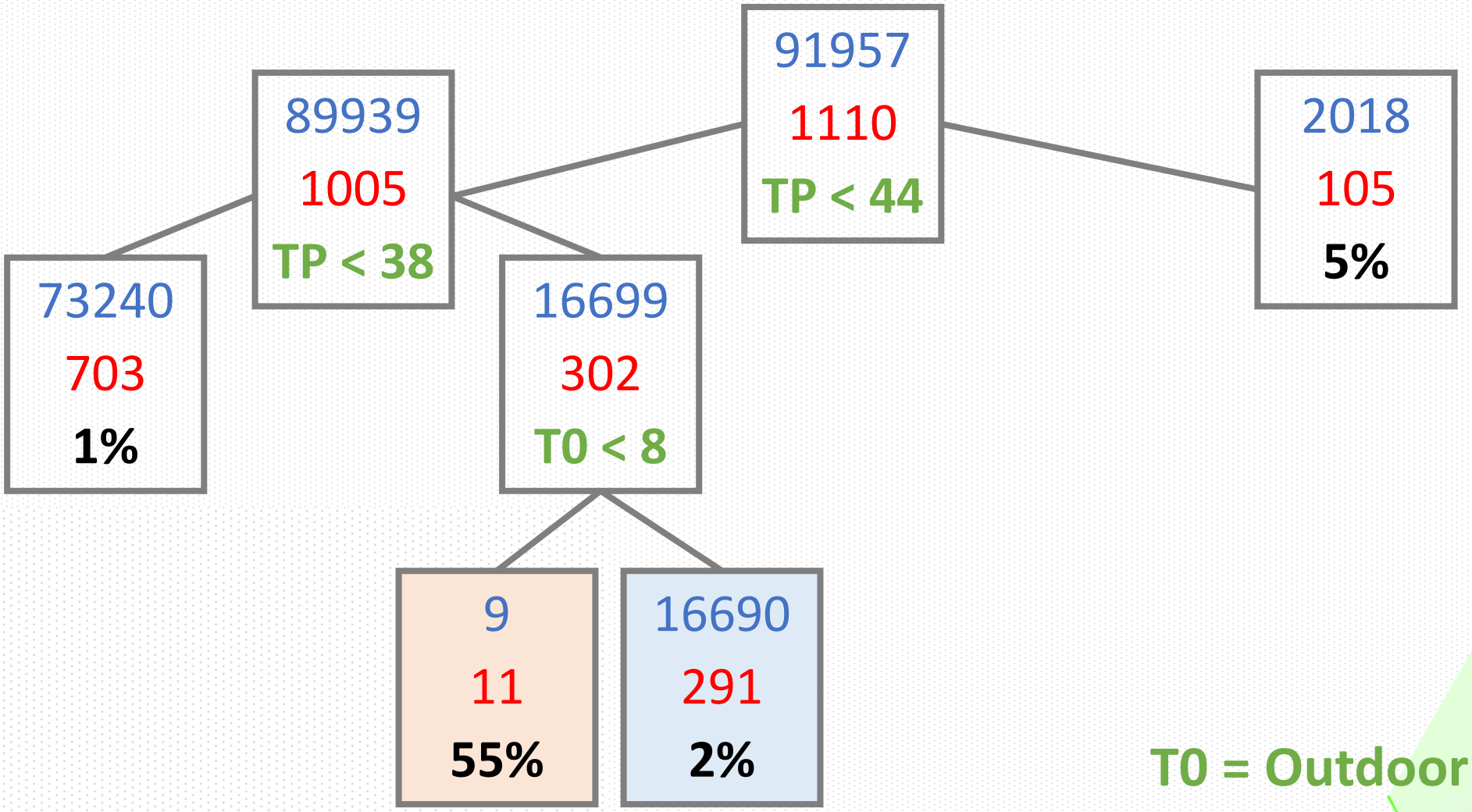
Data simulation and analysis results targets and features → failure probability



Data simulation and analysis results targets and features → failure probability

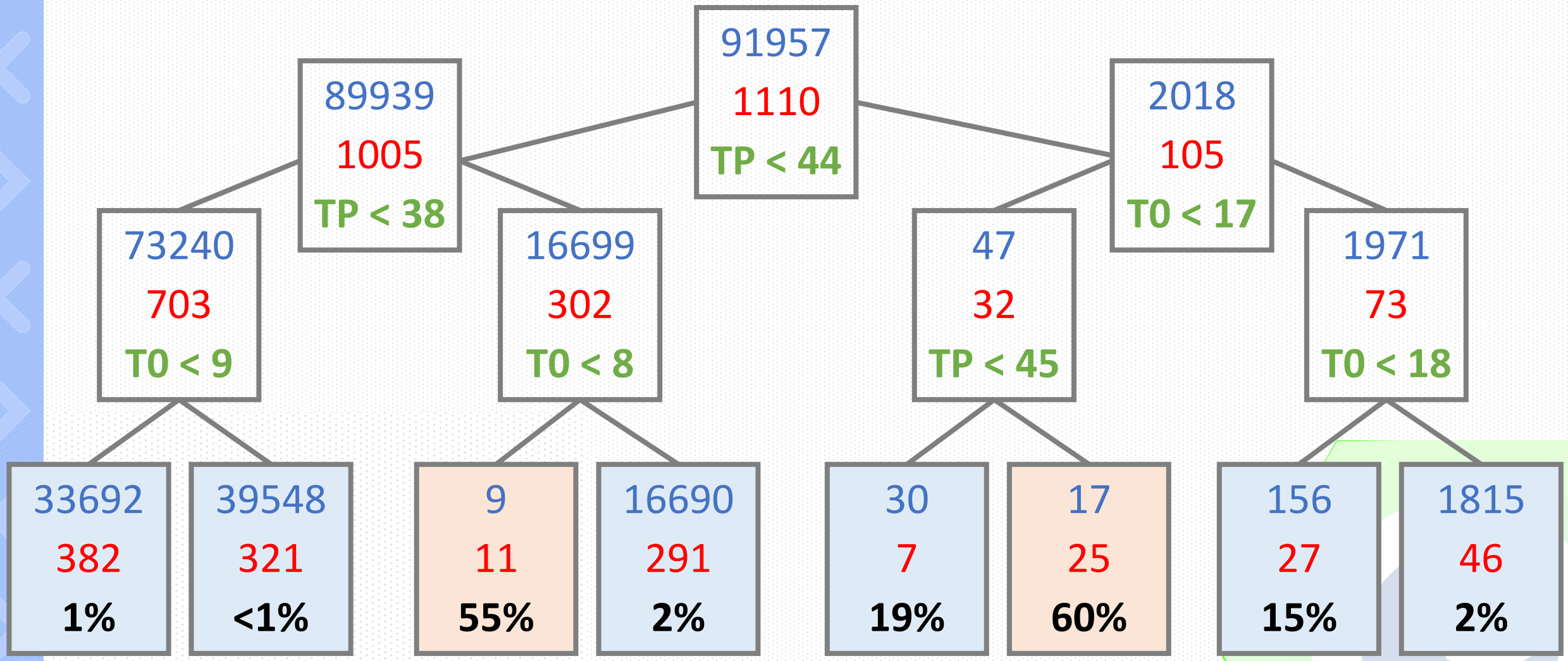


Data simulation and analysis results targets and features → failure probability

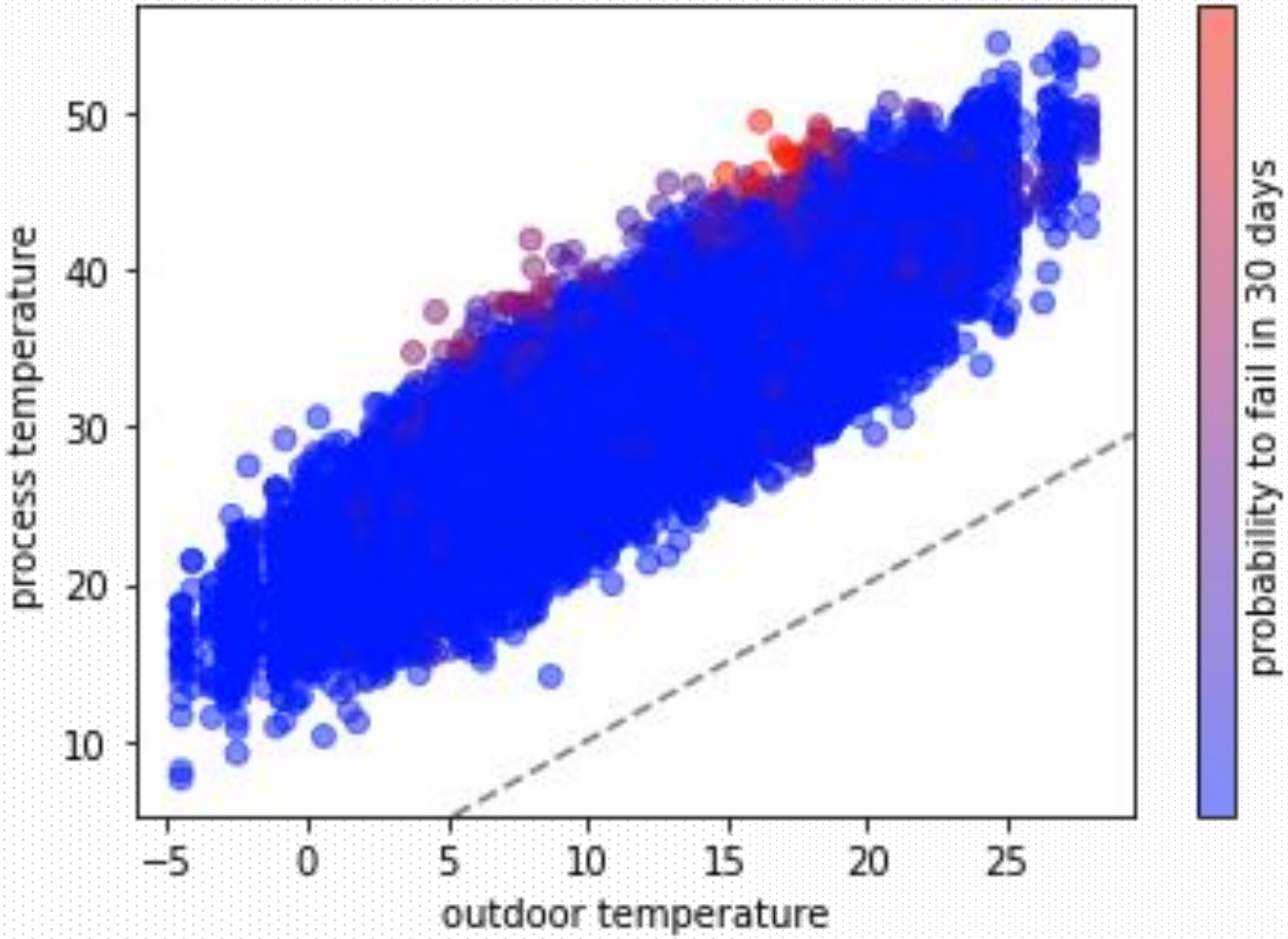


T0 = Outdoor Temperature

Data simulation and analysis results targets and features → failure probability

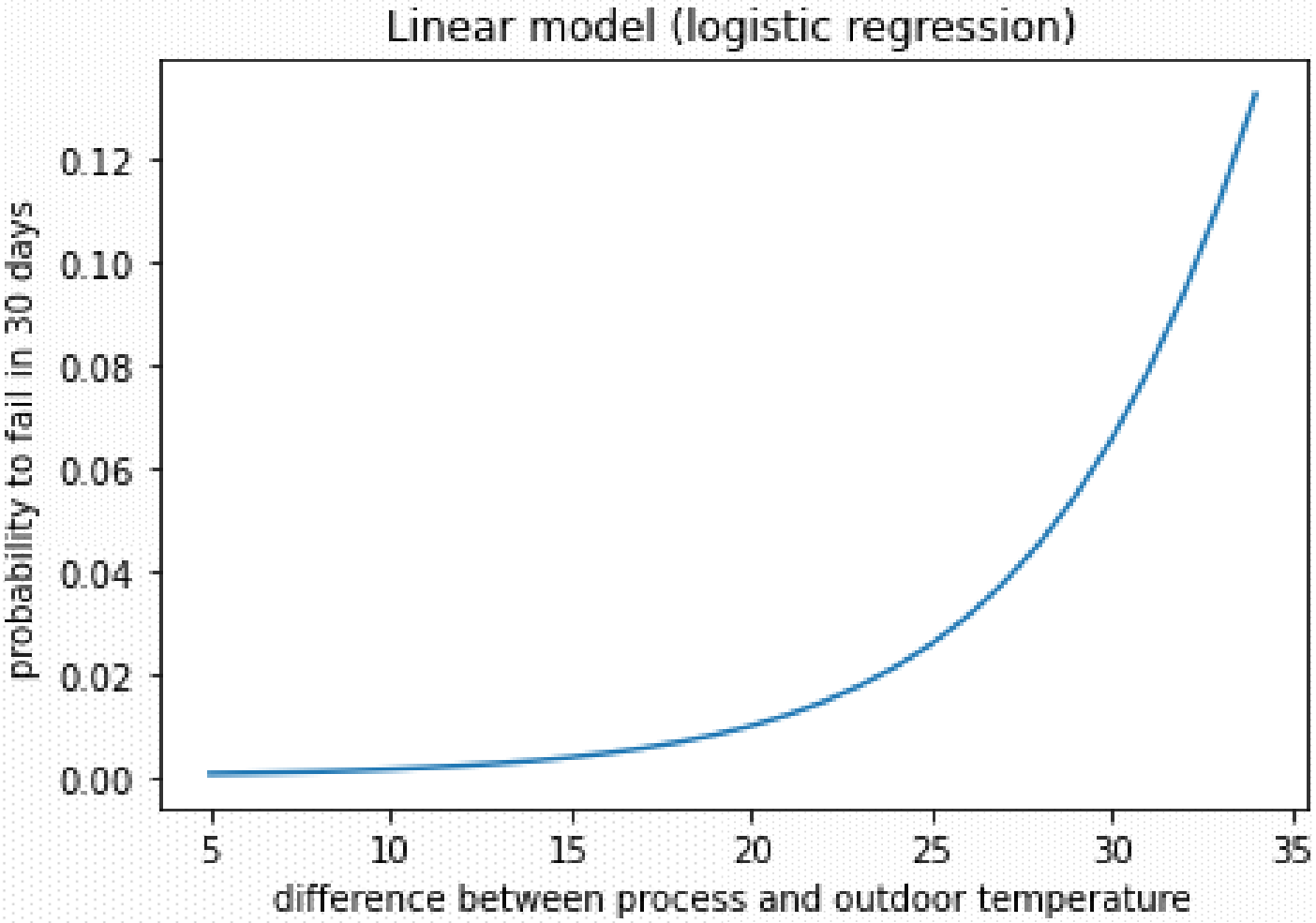


Data simulation and analysis results targets and features → failure probability



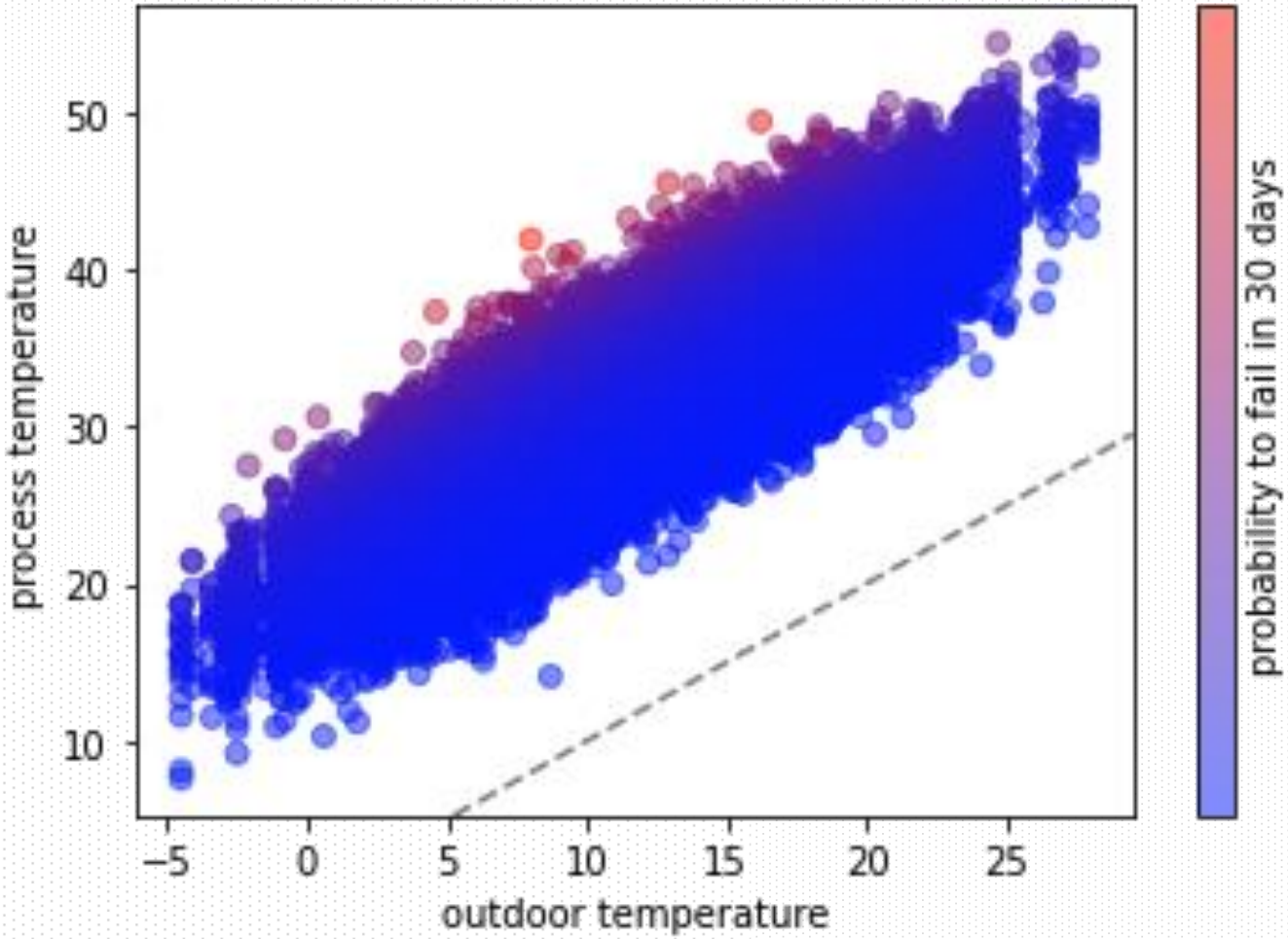
- These probabilities are output of the decision tree model

Data simulation and analysis results targets and features → failure probability



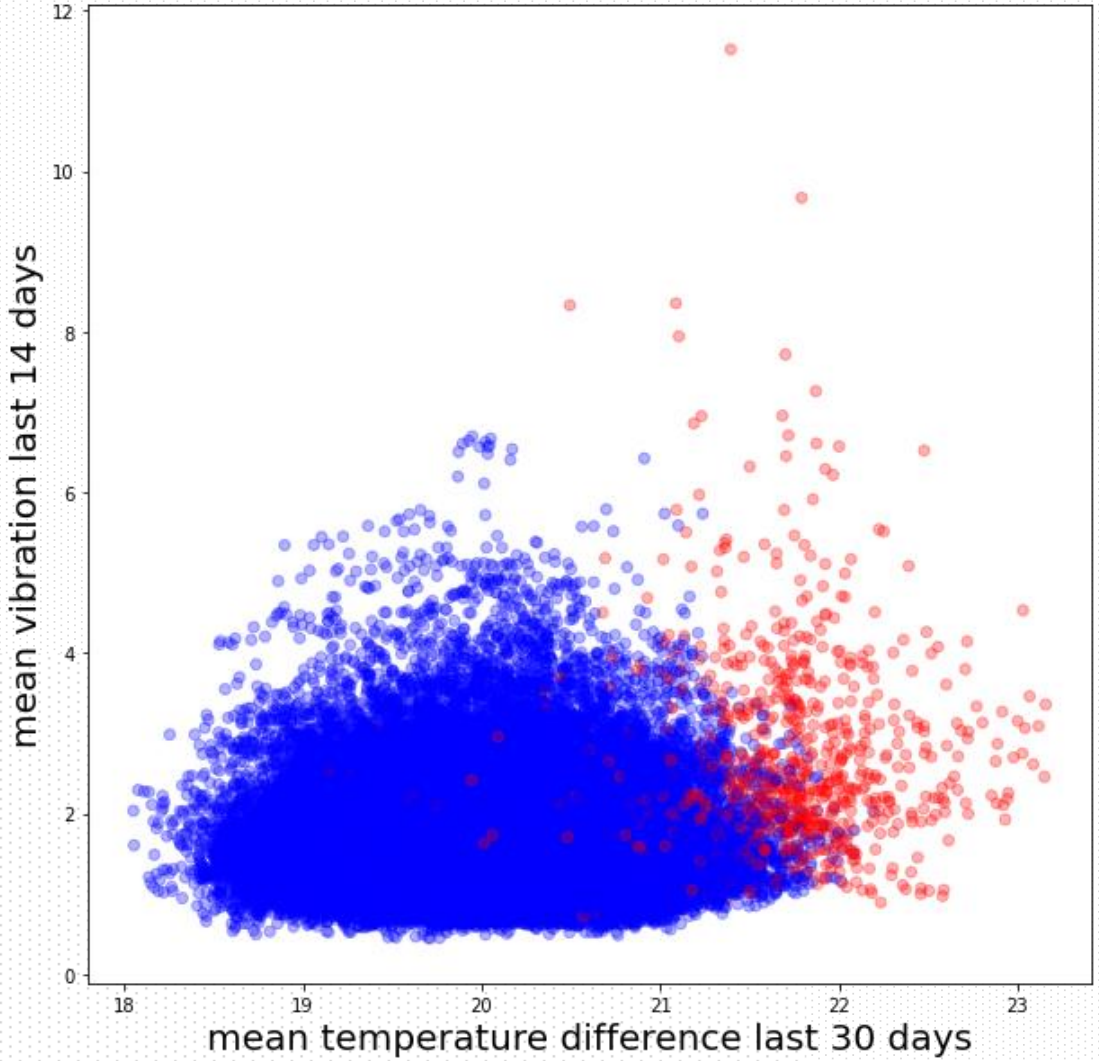
- The difference between process- and outdoor temperature is added as an extra feature to help a **linear model** recognize this relation

Data simulation and analysis results targets and features → failure probability



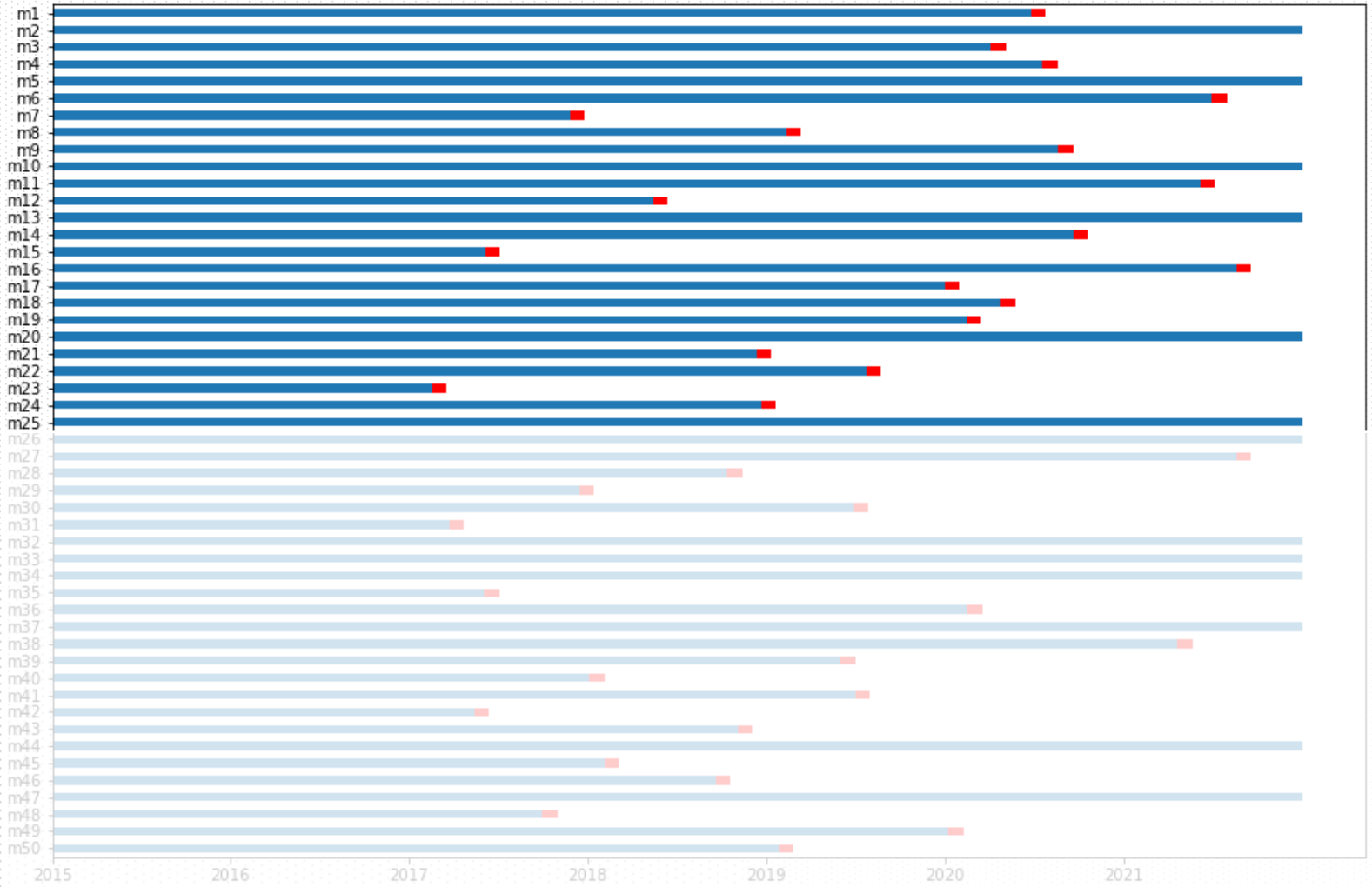
- These probabilities are output of the linear model

Data simulation and analysis results targets and features → failure probability



- Now the mean temperature difference (between process and outdoor) over the last 30 days is added
- Together with the mean vibration last 14 days as a second feature

Data simulation and analysis results targets and features → failure probability



- Train the model on one half of the dataset, to learn the relation between features and target
- Test the model on the other half of the dataset, to see how well it works on unseen data

Data simulation and analysis results targets and features → failure probability

	predicted not functioning	predicted functioning
not functioning	191	349
functioning	6	42996

René: “in practice, predictions should be correct **more than 50%** of the time”

Is this result good enough?

Data simulation and analysis results targets and features → failure probability

	predicted not functioning	predicted functioning
not functioning	191	349
functioning	6	42996

Accuracy: 99%

Data simulation and analysis results targets and features → failure probability

	predicted not functioning	predicted functioning
not functioning	191	349
functioning	6	42996

Precision: 97%

Data simulation and analysis results targets and features → failure probability

	predicted not functioning	predicted functioning
not functioning	191	349
functioning	6	42996

Recall: 35%

Data simulation and analysis results targets and features → failure probability

	predicted not functioning	predicted functioning
not functioning	191	349
functioning	6	42996

Recall: 35% ?

- Not functioning entails last 30 days of 18 machines in the test set
- Of these 18 machines, 17 have at least one (max 21) of last 30 days predicted not functioning

Final discussion and questions

- How long in advance would you need to be warned of a failure?

contact: gert.jacobusse@hz.nl

UNIVERSITY OF APPLIED SCIENCES

Final discussion and questions

- How long in advance would you need to be warned of a failure?
- What precision and recall would you need…
  - or would something like identifying the worst 10% also be useful, to prioritize maintenance or replacement?

contact: gert.jacobusse@hz.nl

UNIVERSITY OF APPLIED SCIENCES

Final discussion and questions

- How long in advance would you need to be warned of a failure?
- What precision and recall would you need…
  - or would something like identifying the worst 10% also be useful, to prioritize maintenance or replacement?
- What is more useful/ likely, features that:
  - provide early warnings, like vibration or power consumption
  - cause failure, like temperature, pressure or acidity

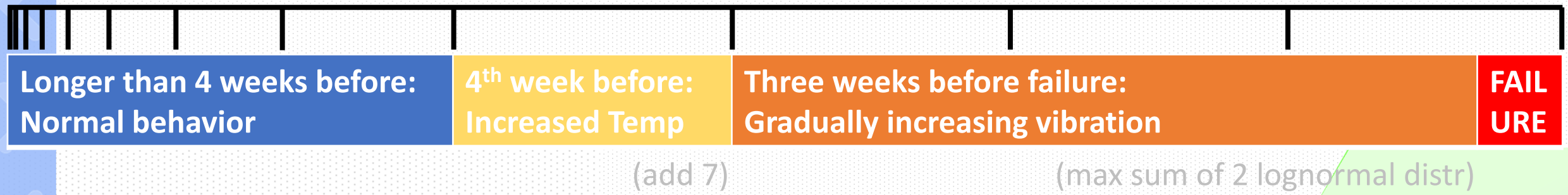
contact: gert.jacobusse@hz.nl

UNIVERSITY OF APPLIED SCIENCES

Behind the scenes: simulation design

Normal behavior

- Vibration: random lognormal distribution
- Temperature: outdoor temperature + 20 + random noise (stdev 3)



Failure probability

- increases proportional to age$^{1.5}$ (max 0.002)