UNIVERSITÀ DEGLI STUDI DI BRESCIA

INFN
Istituto Nazionale di Fisica Nucleare

Long Short-Term Memory Networks and Bayesian Inference for Time-Evolving Systems: an Industrial Case

D. Pagano
UNIVERSITÀ DEGLI STUDI DI BRESCIA & INFN PAVIA

21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research

Physicists and Industry

- **About myself**: Particle physicist, Professor at the University of Brescia, Head of Laboratory of Applied Nuclear Physics (https://anplab.unibs.it/)
  - ALICE @ LHC, muography, CRNS, ...
  - statistics, machine learning (ML), metaheuristic algorithms, ...
- Over the years I collaborated (trough research contracts) with some big industries, mainly in the Brescia area
  - data analysis
  - development of ML tools
  - predictive maintenance
  - educations, ...

MAIN EU NUTS-3 REGIONS WITH STRONG SPECIALIZATION, HIGH VALUE ADDED AND LARGE EMPLOYMENT IN INDUSTRY: YEAR 2011 (Industry excluded construction. NUTS-3 Regions ranked by total value added in industry)
Source: compiled by Fondazione Edison and Confindustria Bergamo on data from Eurostat)

RANK	CODE	NUTS-3	% value added in industry	% persons employed in industry	Persons employed in industry (thousands)	Value added in industry per person employed	Total value added in industry (EUR million)
1	**ITC47**	**Brescia**	**30,7%**	**31,1%**	**167,7**	**60.268**	**10.106,9**
2	ITC46	Bergamo	34,1%	34,3%	156,3	62.254	9.730,3
3	DE913	Wolfsburg, Kreisfreie Stadt	73,2%	47,8%	55,6	155.315	8.635,5
4	ITH32	Vicenza	35,4%	38,3%	161,8	53.259	8.617,3
5	DE112	Böblingen	50,2%	34,1%	72,4	106.086	7.680,6

The industrial case

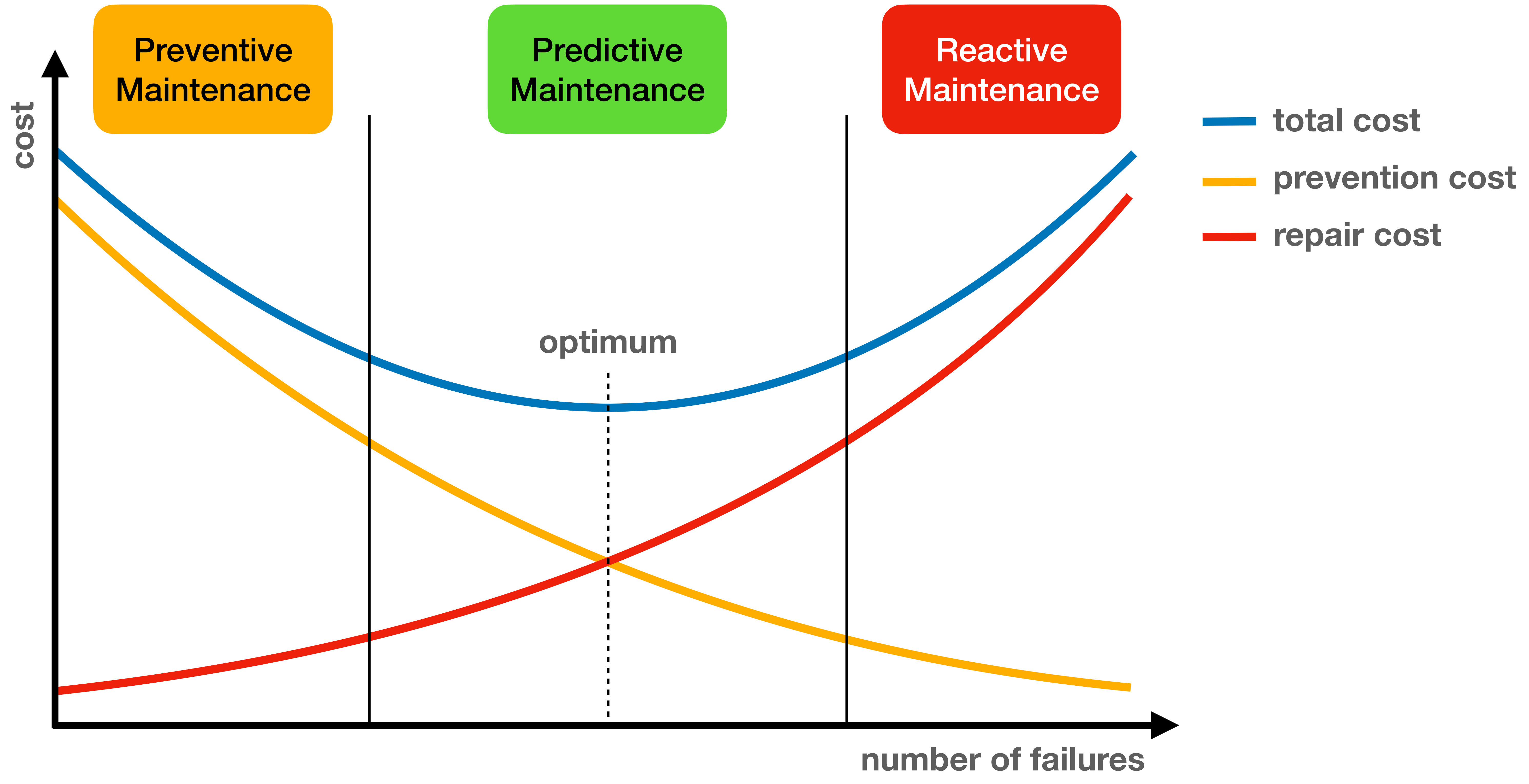
- Heavy industry factory (**Note**: because of a NDA I won't share any detail about the company and their data)

- **Goal**: development a *predictive maintenance* system for a sector of their plant, based on data from the already available sensors

Reactive Maintenance

Preventive Maintenance

Predictive Maintenance

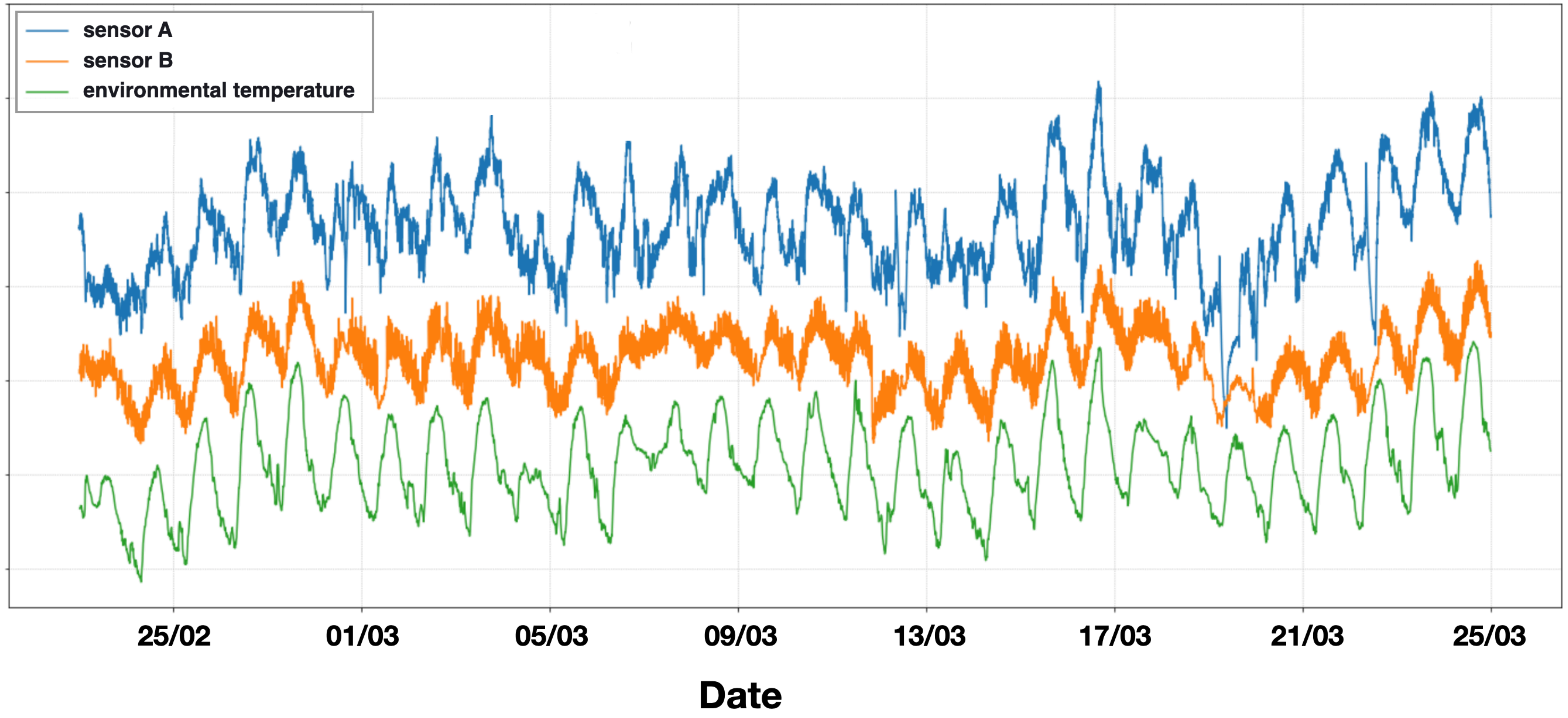


The tricky industrial data (a physicist perspective)

- Data from different sources using different protocols (MQTT, OPC UA, SQL, IBA, ...)
- Hundreds (thousands if the whole plant is considered) of **time series** sampled at different frequencies
- Data **NOT** ready to use...
  - meaning of many variables unknown
  - problems of oversampling/under-sampling
  - unreliable variable names and changes of variables names over time
  - missing values, reading errors, ...
- Data size (for this particular case): ~10 GB/month

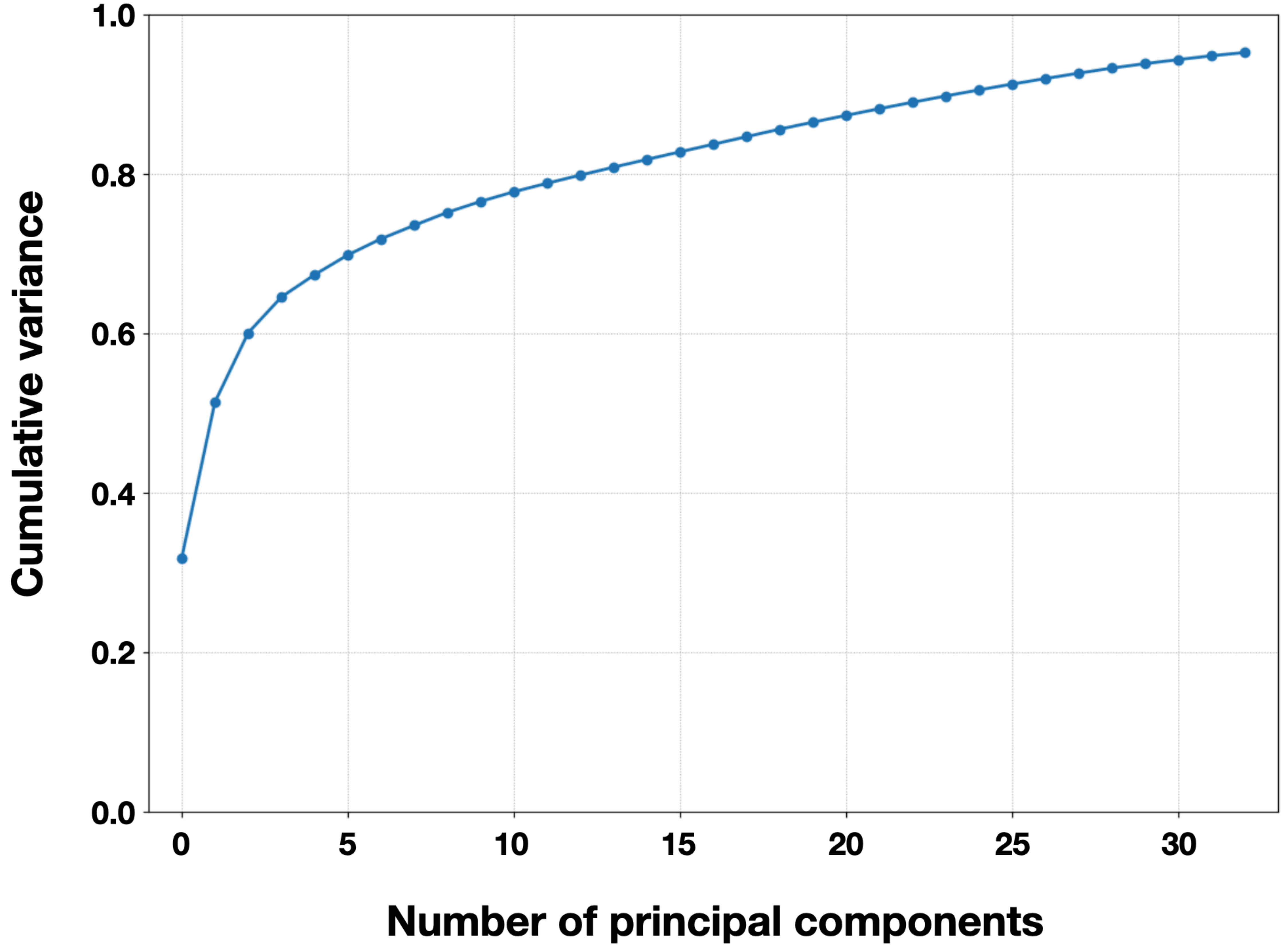
Data preparation

- For each active component of the plant (motors and other moving parts) we handled:
  - zero and near-zero variance predictors
  - fully correlated predictors
  - missing values
  - error readings
  - oversampling
  - *uninformative* sensors



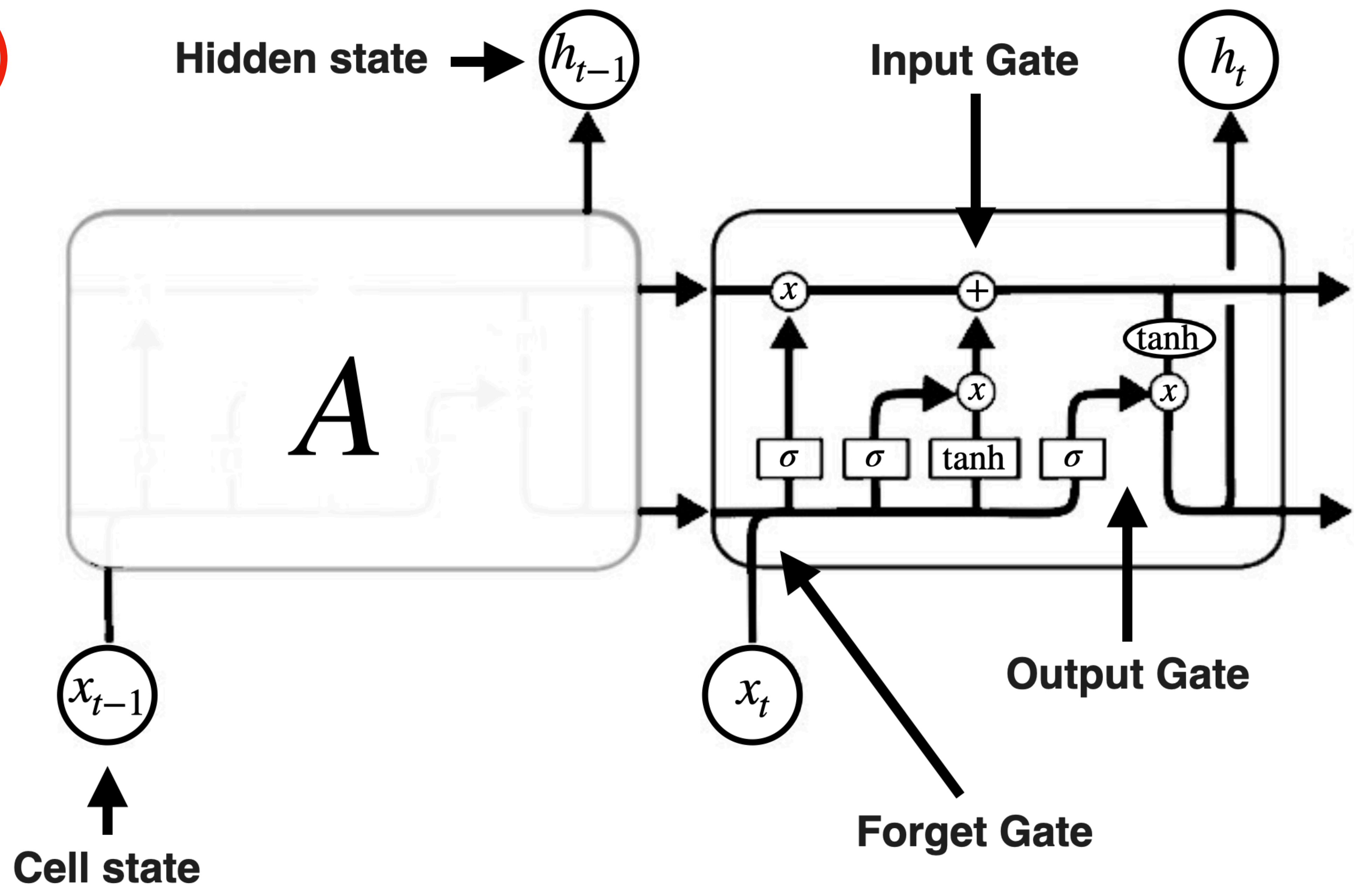
Variable selection

- For each active component of the plant a *principal components analysis* was performed
- PCA: orthogonal linear transformation of data to a new coordinate system
    - The new variable with the greatest variance is projected on the first axis, the second greatest on the second axis and so on...
- A cumulative variance, between 85% and 90%, was used to choose the number of principal components
- Up to 75% of data reduction



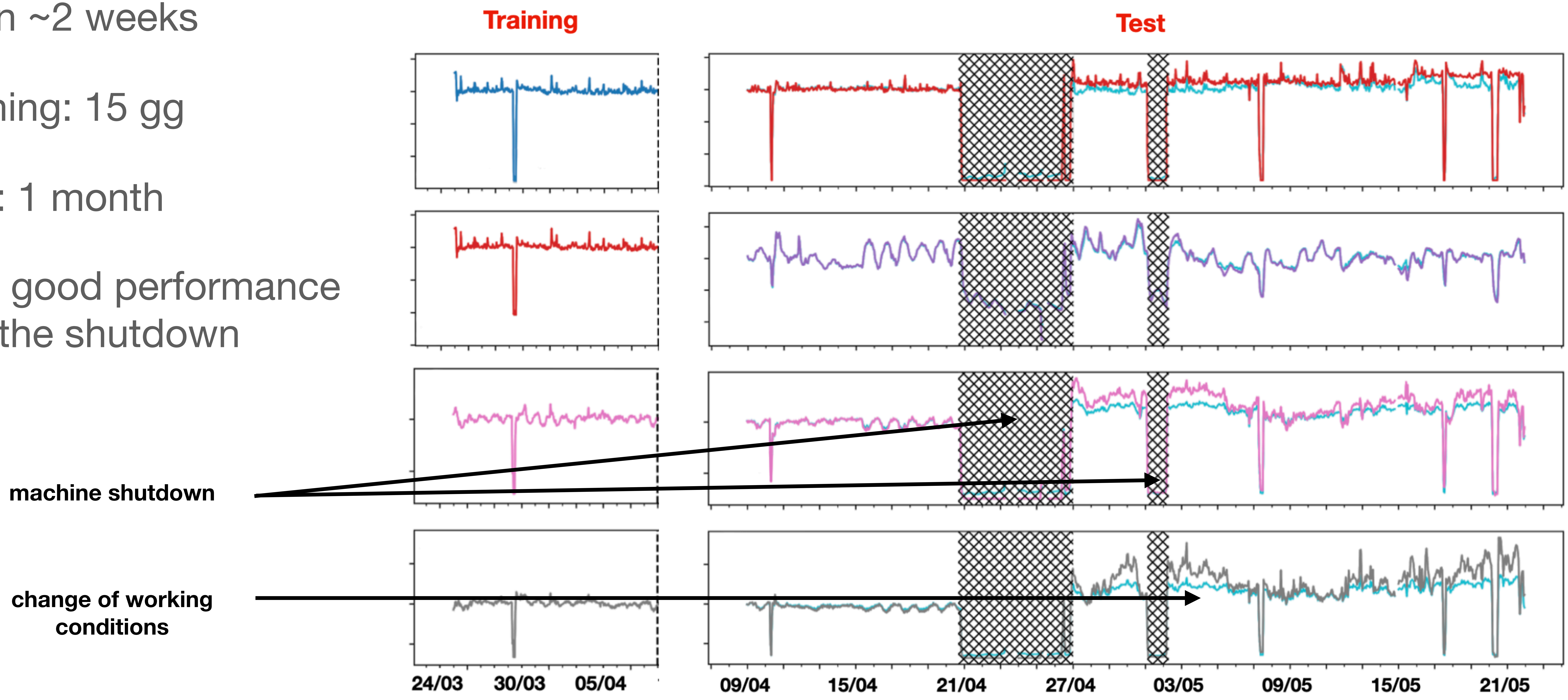
Model for data

- An industrial plant is a time-evolving system: data are time series
- Goal: **time series forecasting** with a model describing the system in a (supposedly) "good" working condition
  - Anomaly detection
  - Assess the "health" of the plant (ideally the RUL)
- We used a Long short-term memory (LSTM) neural network
  - A recurrent neural network with the capability of a long-term memory
  - Very well-suited for making predictions based on time series data



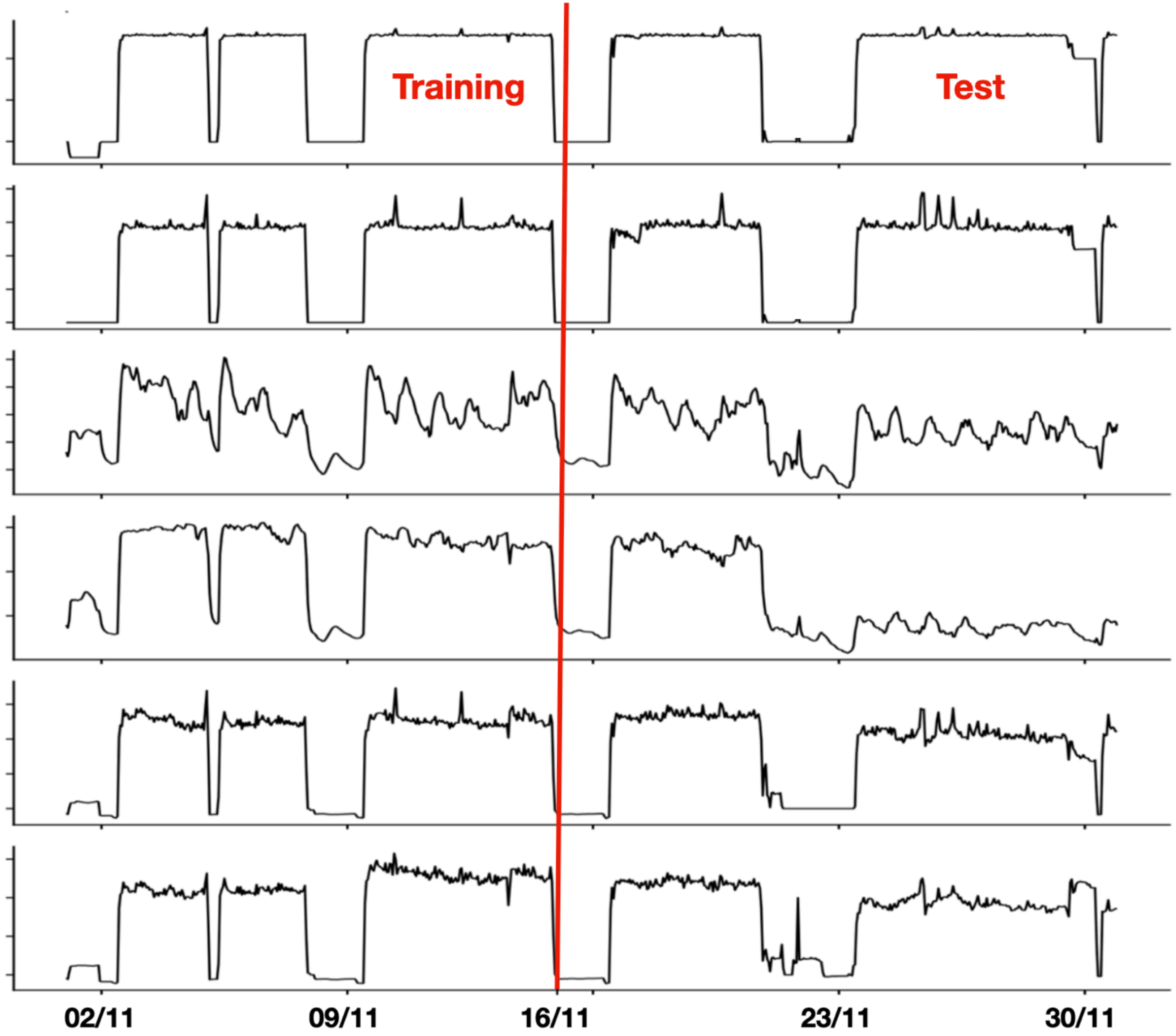
Optimization studies

- Tuning of LSTM structure to handle 20-40 variables while keeping the training window within ~2 weeks
- Training: 15 gg
- Test: 1 month
- Very good performance until the shutdown

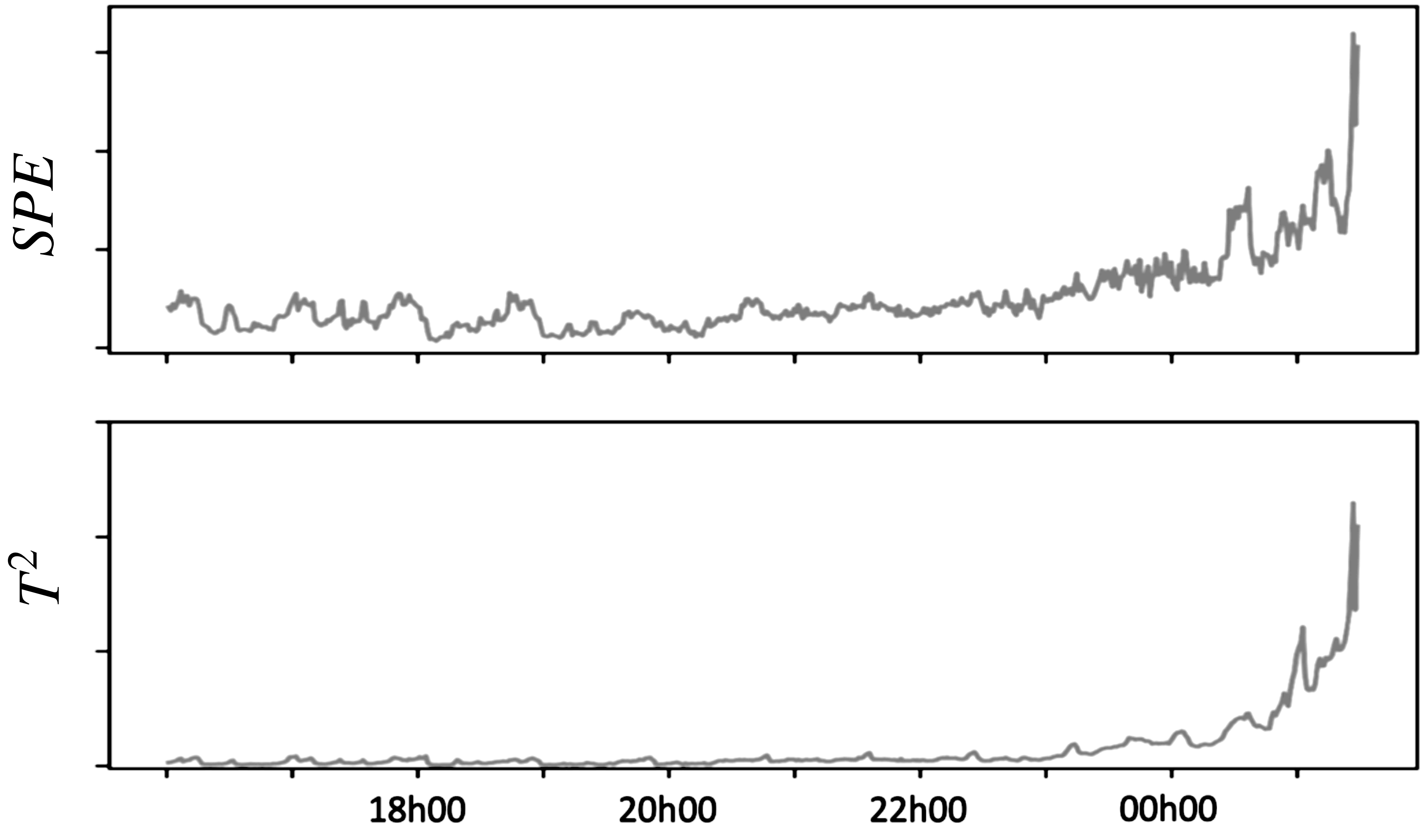


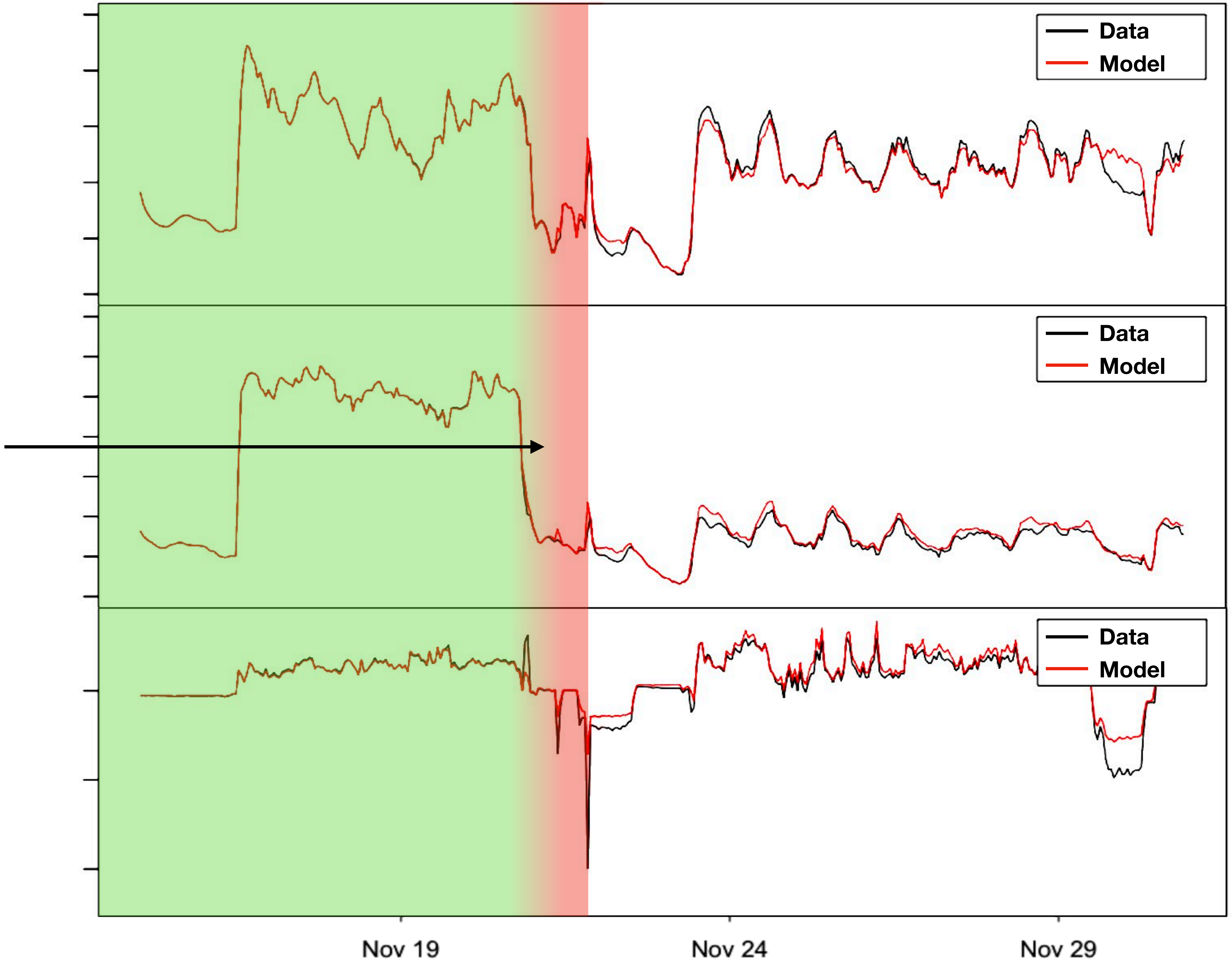
Anomaly detection

- Goal: identify previously know malfunctions
- Blind search
- Training: 15 days of data
- Agreement between model and data measured by T^2 and SPE



Anomaly detection





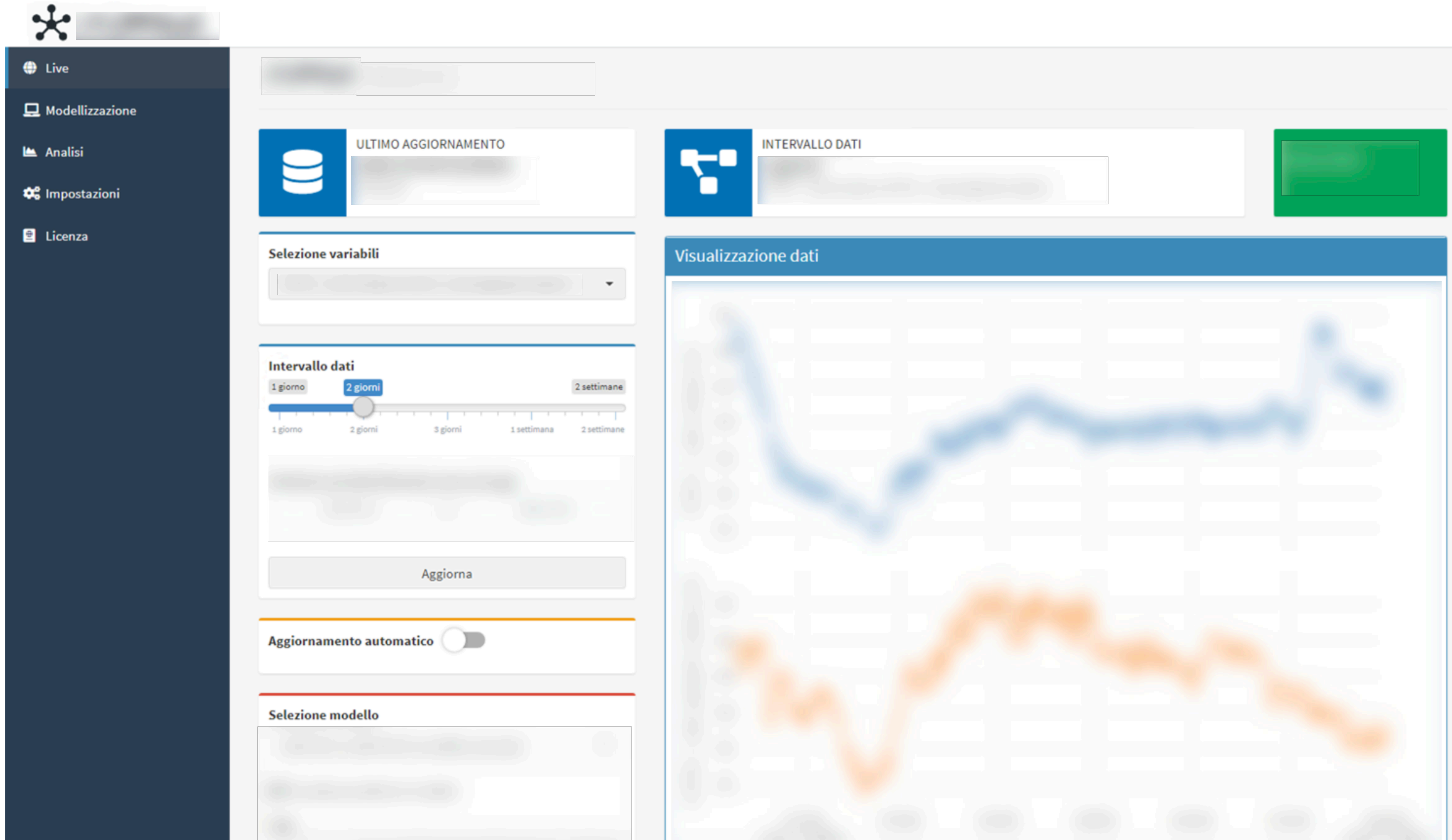
Bayesian estimation of the "health"

- Let $\boldsymbol{\omega}$ be a set of possible states describing the "health" of the plant and $\boldsymbol{x} = \{T^2, SPE\}$
- $P(\omega_j | \boldsymbol{x}) = \dfrac{p(\boldsymbol{x} | \omega_j)P(\omega_j)}{\int p(\boldsymbol{x} | \omega_j)P(\omega_j)d\omega_j}$
- The most simple approach: $\boldsymbol{\omega} = \{\omega_{good}, \omega_{bad}\}$
- The likelihoods $p(\boldsymbol{x} | \omega_j)$ can estimated from data or physical models
- In our tests $p(x_j | \omega_{good})$ and $p(x_j | \omega_{bad})$ were respectively modeled as a Normal and a Uniform distributions
    - $p(x_j | \omega_{bad})$ are tricky to model from data (they depend on the problem of the plant)

Bayesian estimation of the "health"

From $P(\omega_{good}|x)$ to maintenance

- An alarm based on the value of $P(\omega_{good}|x)$ was created
  - Note: in the long run $P(\omega_{good}|x)$ starts decreasing over time also in absence of failures (seasonality) so the model has to be regularly updated
- How does the company do it?
- All code was written in **R** (LSTM from **Keras** and **TensorFlow**)
- An interactive web app was also developed (with R Shiny)
- Running on a dedicated server of the company



Conclusions

- LSTM networks successfully used to model industrial data
- T^2 and SPE metrics used to measure the agreement between model (trained with data from a reference period) and data
- Bayesian inference to assess the "health" of the plant on the basis of $P(\omega_{good}|\boldsymbol{x})$
- This approach gave very good results in several tests
- A dashboard was also developed to allow:
  - the monitoring of the agreement between data and model
  - the training of new models
  - the raising of alarms based on $P(\omega_{good}|\boldsymbol{x})$