TRAIL

TRUSTED AI LABS

Summer Workshop 23'Nantes

Certification of AI based (sub)system

Project n°10

Team

Multitel
INNOVATION CENTRE

Emmanuel Jean
Research Leader for Trustworthy AI
AI group - Multitel

Bérangère Burgunder
R&D engineer in ML/DL
AI group - Multitel

Florian Facchin
Intern
AI group - Multitel

TRAIL

TRUSTED AI LABS

Summer Workshop 23' - Nantes

- Context and Challenges -----> EC regulation – AI Act
- Use case : Certification of an AI-based system for drone
  - Data quality
  - Uncertainty quantification

Thrustworthy AI: Context and challenges

Industrialization of AI is a crucial issue of industrial and economic competitiveness

- Increase in efficiency (i.e.: more sparing use of resources, better allocation of resources,...)
- Optimization of existing products or services
- Improvement of user experience

AI is increasingly complex and appears as black-box

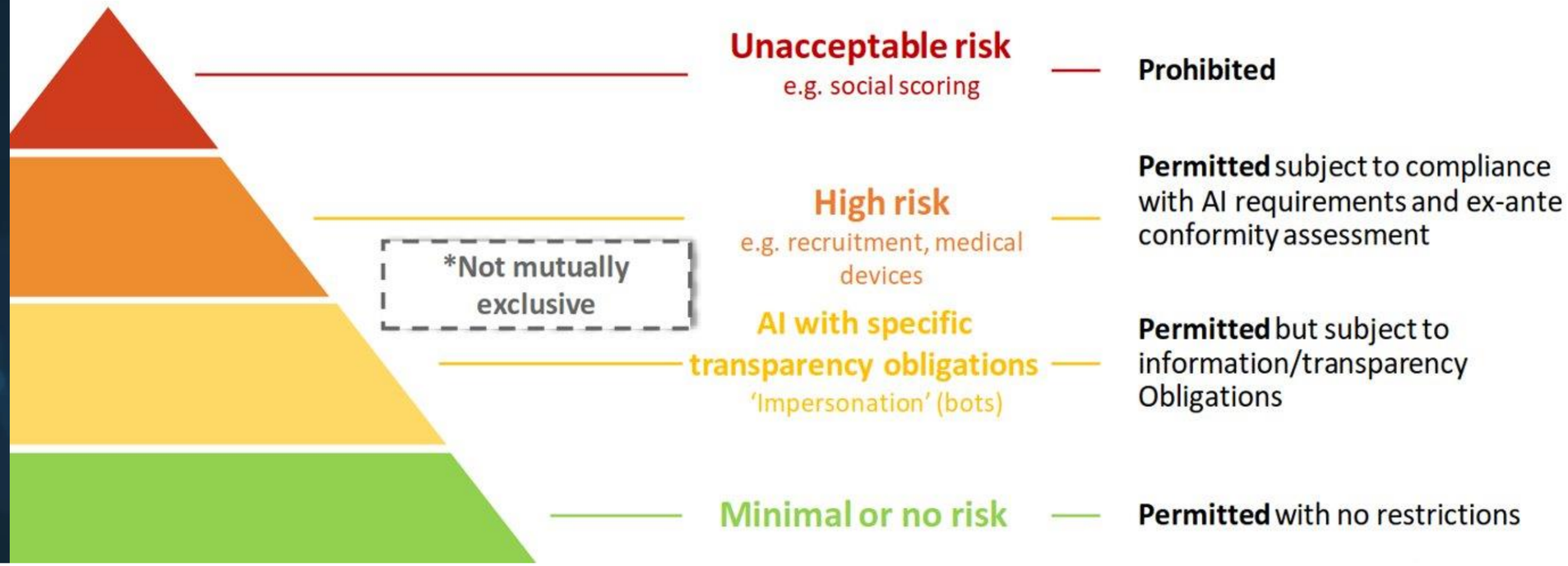
- AI creates uncertainty which restricts its acceptance by business leaders and its integration into new products
- AI creates mistrust which hinders its adoption by users

Critical systems

- What level of confidence can be given to AI techniques ?
- Are AI techniques compatible with the strict certification requirements of specific sectors (Health / Aeronautics / Space) ?

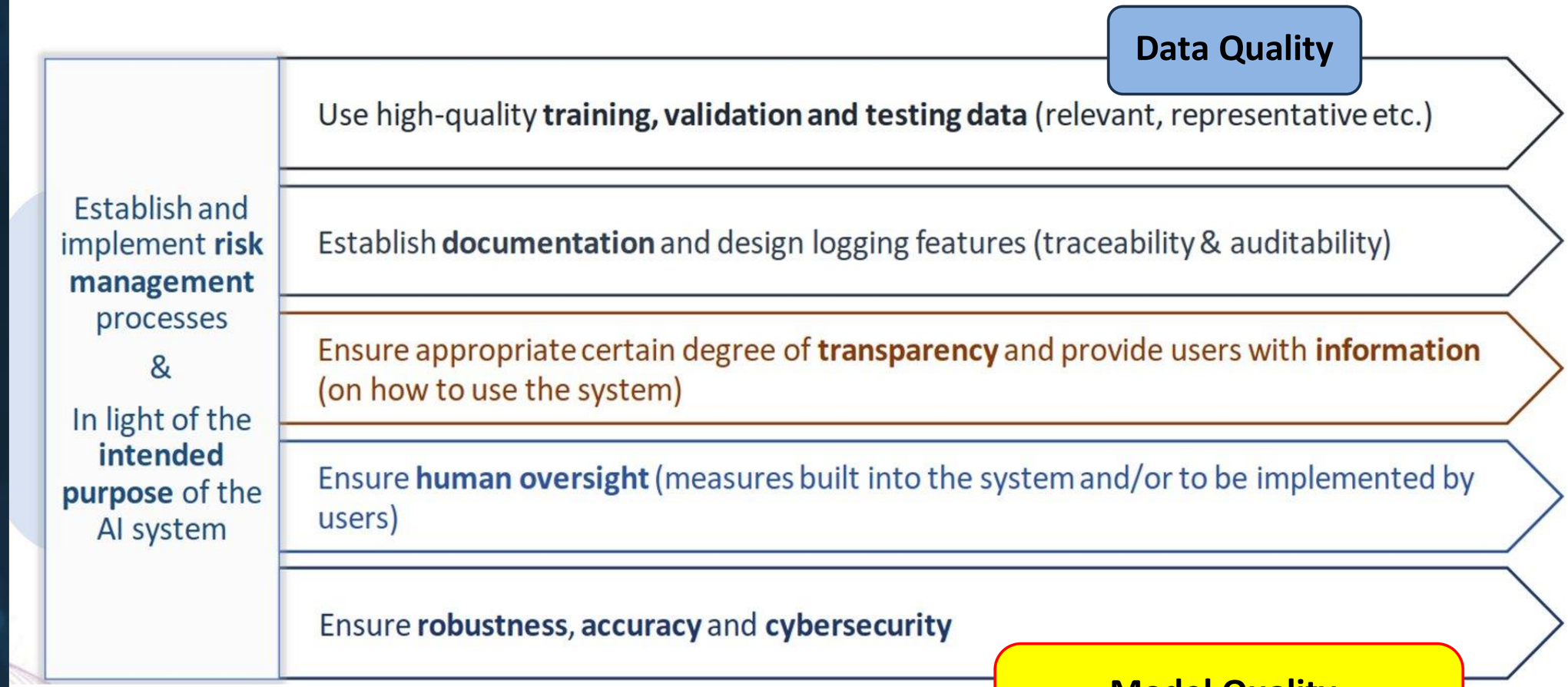
AI European Regulation

- 2019 - Guidelines based on 7 requirements
  - Put in place a whole framework for trustworthy AI in critical systems
- April 2021 -> 14 June 2023 – EU AI ACT : First Regulation on AI
  - Risk-based approach to regulate applications implementing AI components (horizontal)



https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

Requirements for high-risk AI



Data Quality

Model Quality

Uncertainty Quantification

https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

Use case: Remaining time to flight estimation system for a drone

Data:

Training data

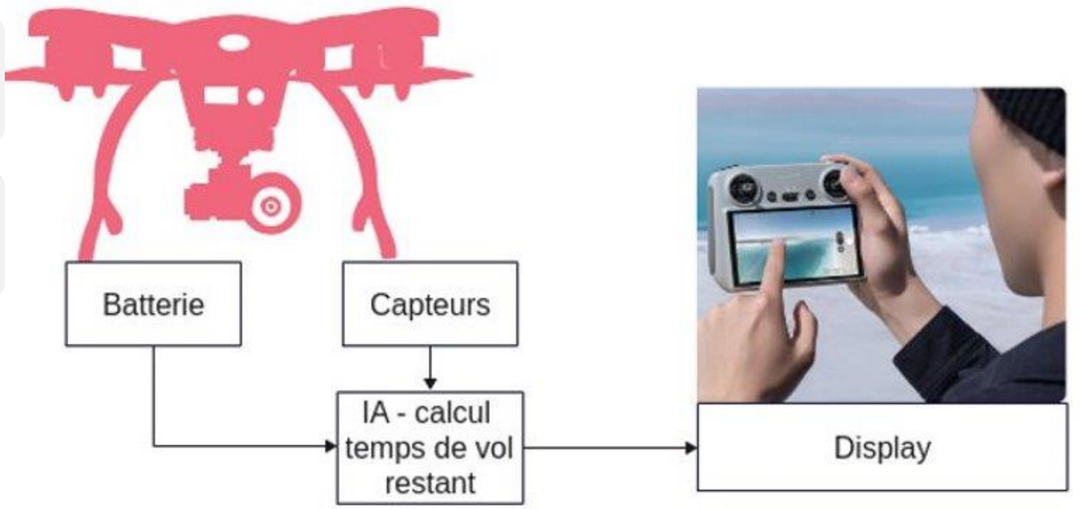
- 195 flights ~ 3 min = 10 h 45 min = 65 km travelled
- Different flight parameters (Altitude, Payload, Speed, Different routes)
- Anemometer : Wind angle and speed
- Position, orientation, speed (linear and angular) and acceleration
- Battery current and voltage (not used in training)
- Sensor box mounted on the drone to collect data

Datasets split:

- Train
- Validation
- Test

195 flights with {0, 250, 500} g

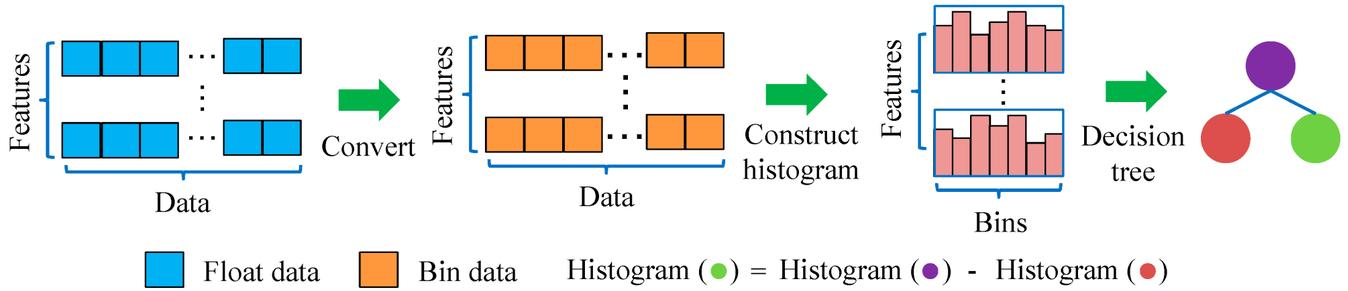
- "heavier" (OOD) → 1 flight with 750g

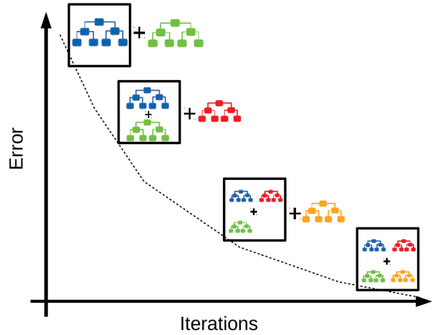


Use case: Remaining time to flight estimation system for a drone

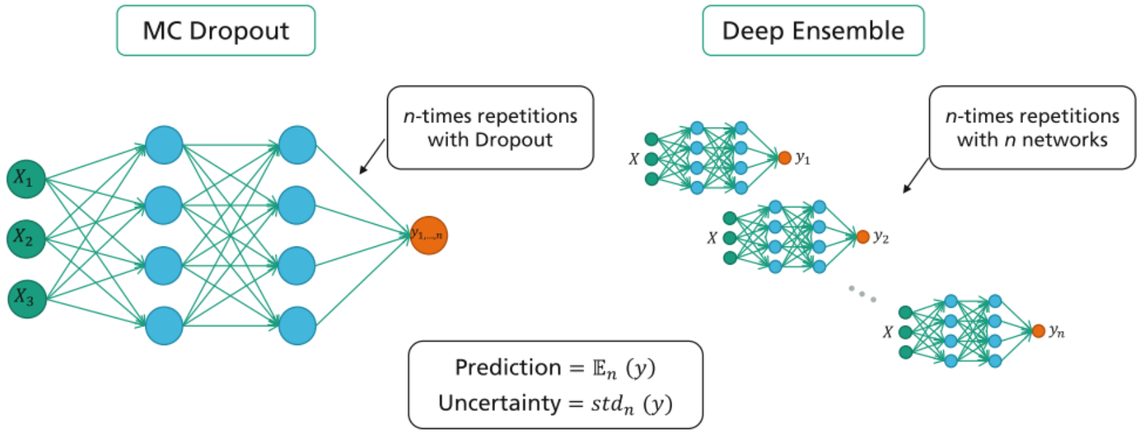
Models for the estimation of the instantaneous power

- Classical Machine Learning model : HistGradientBoostingRegressor





- Deep Learning model : Deep Ensemble of (20) MC-Dropout models (20 runs)



Sources : http://tvas.me/articles/2019/08/26/Block-Distributed-Gradient-Boosted-Trees.html

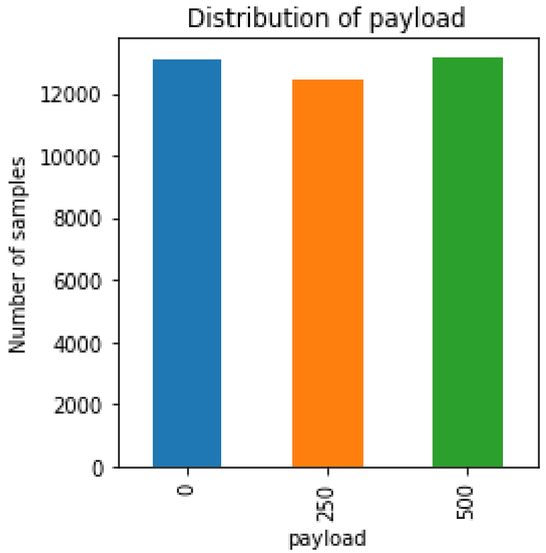
Liang, W.; Luo, S.; Zhao, G.; Wu, H. Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics* **2020**, *8*, 765. https://doi.org/10.3390/math8050765

Xinyang Wu et al., Quantification of Uncertainties in Neural Networks, In book: New Digital Work, April 2023

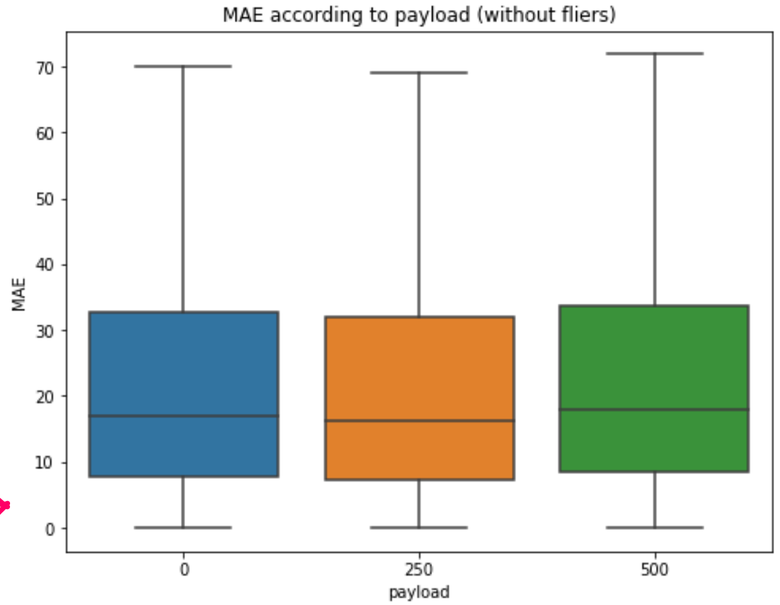
Data Quality : Fairness

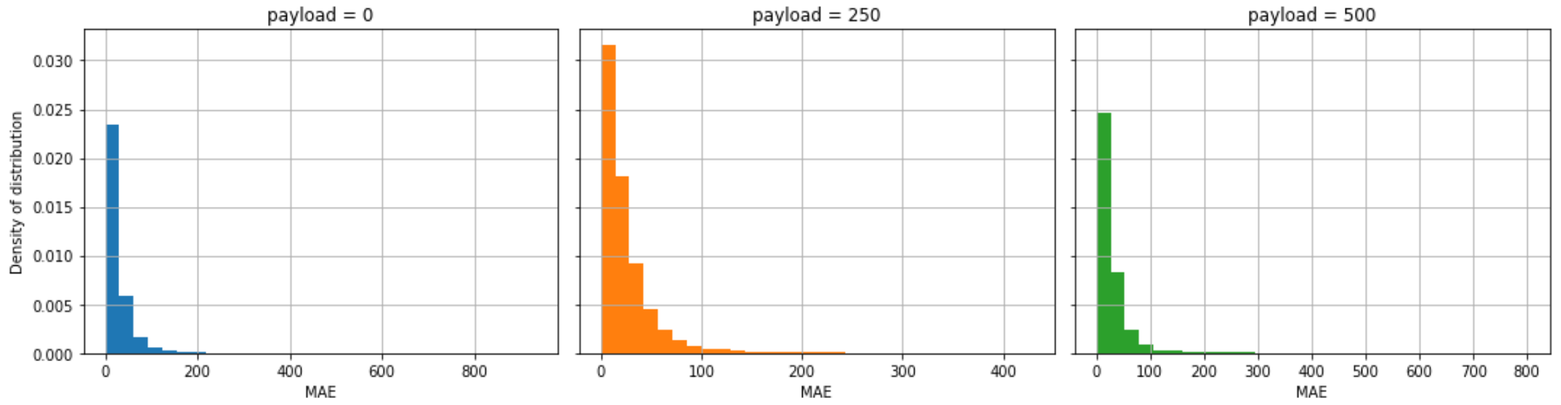
- Goal = study bias in data and models
- Example of bias in medical data : sex, age
  -> in our use case : payload

Are the subsets equally distributed ?



Does the model perform equally for every payload ?

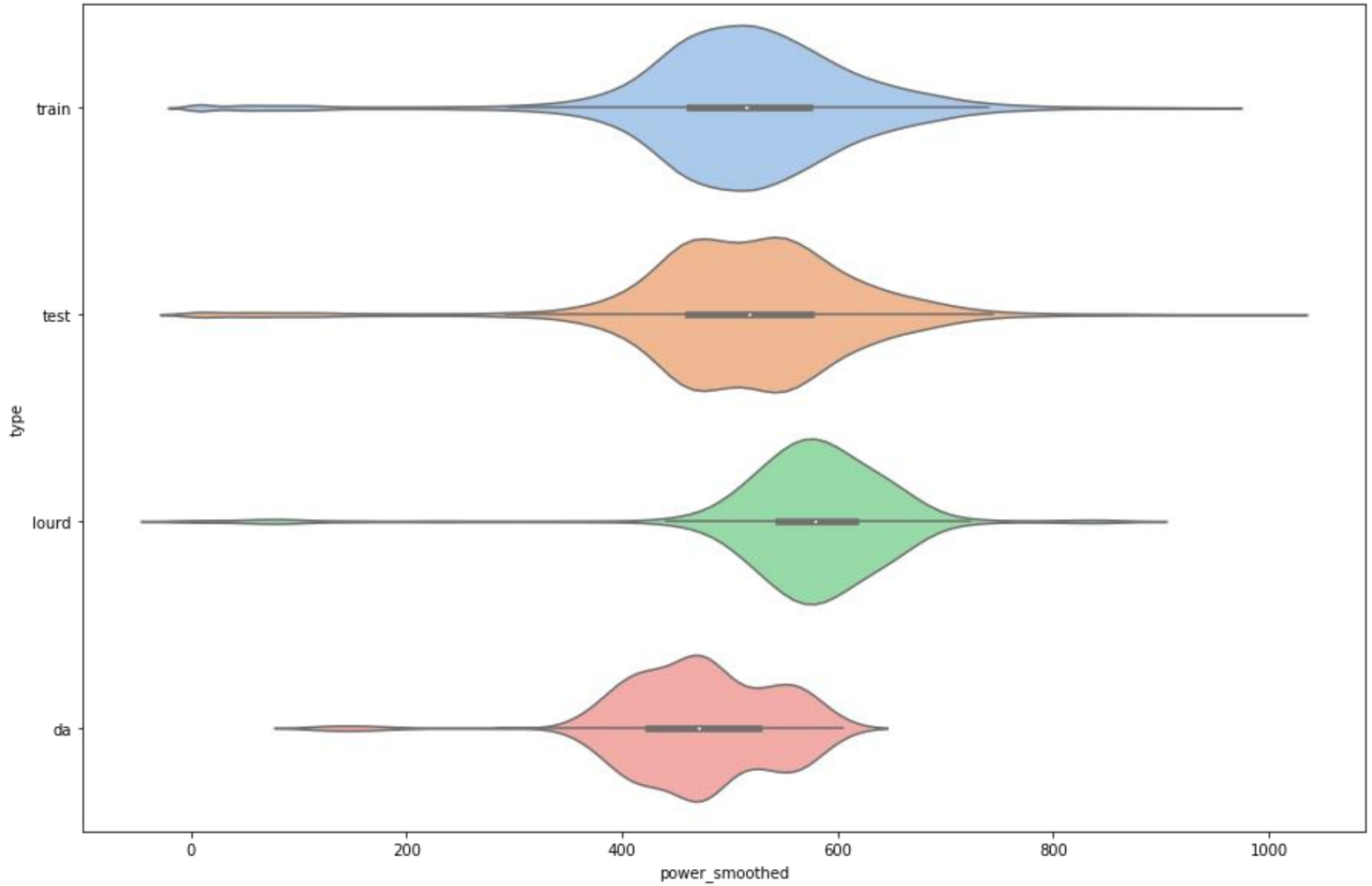




Data Quality : Out Of Distribution

The term 'out-of-distribution' (OOD) data refers to the data collected at a different time, and possibly under different conditions or environment, compared to the data collected to create the model. This data is from a 'different distribution'.
-> In our case: the 'heavier' ('lourd') dataset is OOD.

Violin plots: Display the distributions of the 'power' for multiple datasets (train, test, 'lourd', DA)



Model Quality : Prediction Intervals

Classical Machine Learning Model : HistGradientBoostingRegressor (sklearn.ensemble)

Tool : Puncc (Predictive uncertainty calibration and conformalization)

- Open-source Python library (developed by the DEEL Project)
- Collection of state-of-the-art **conformal prediction** algorithms providing prediction intervals backed by theoretical guarantees
  - Split Conformal Prediction
  - Locally Adaptive Conformal Prediction
  - Conformalized Quantile Regression
  - Ensemble Batch Prediction Intervals method
  - Locally adaptive Ensemble Batch Prediction Intervals method
  - CV + (cross-validation)
- **Evaluation**
  - Marginal Coverage (assess how often the predicted intervals contain the true values)
  - Average width of the intervals

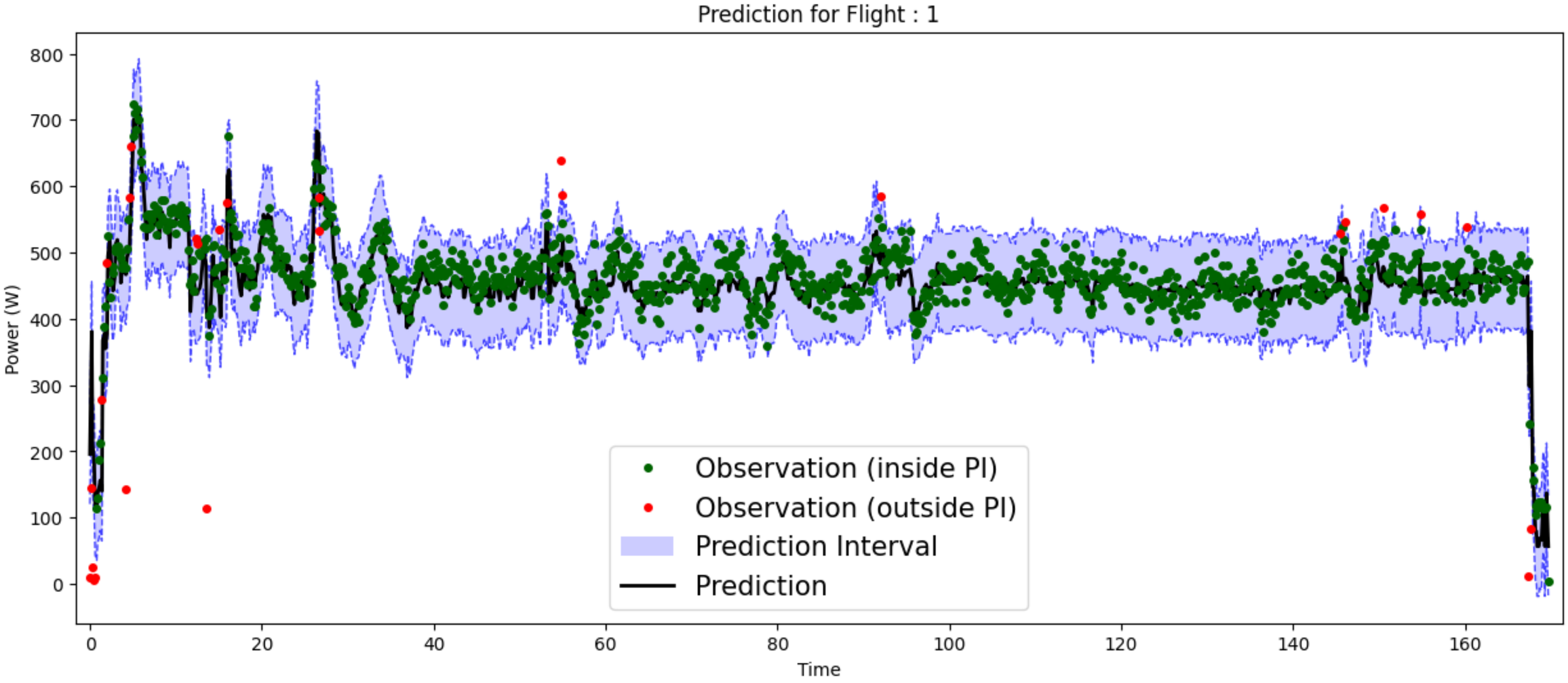
DEEL

PUNCC

A Conformal Prediction Library

https://github.com/deel-ai/puncc/

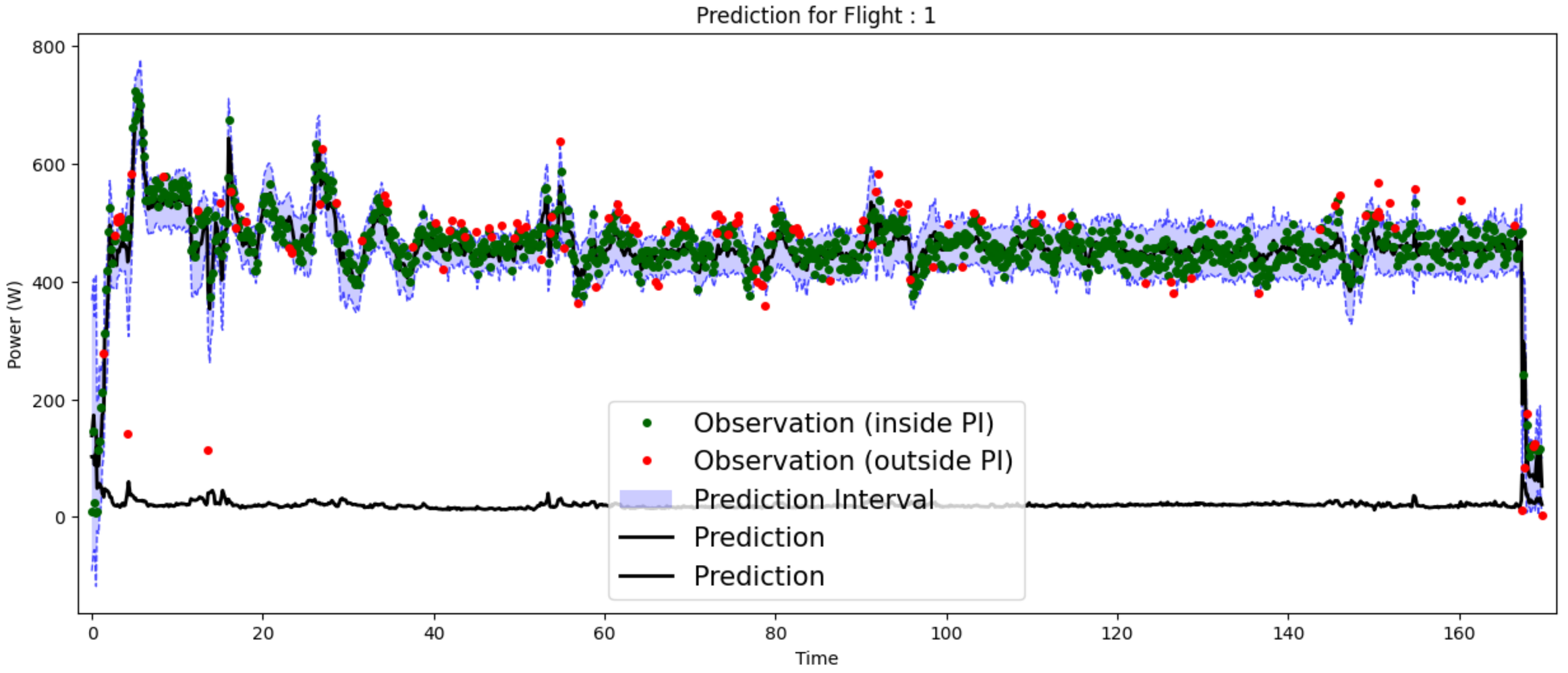
Split Conformal prediction



TEST dataset

Marginal coverage: 0.98 Average width:150.84

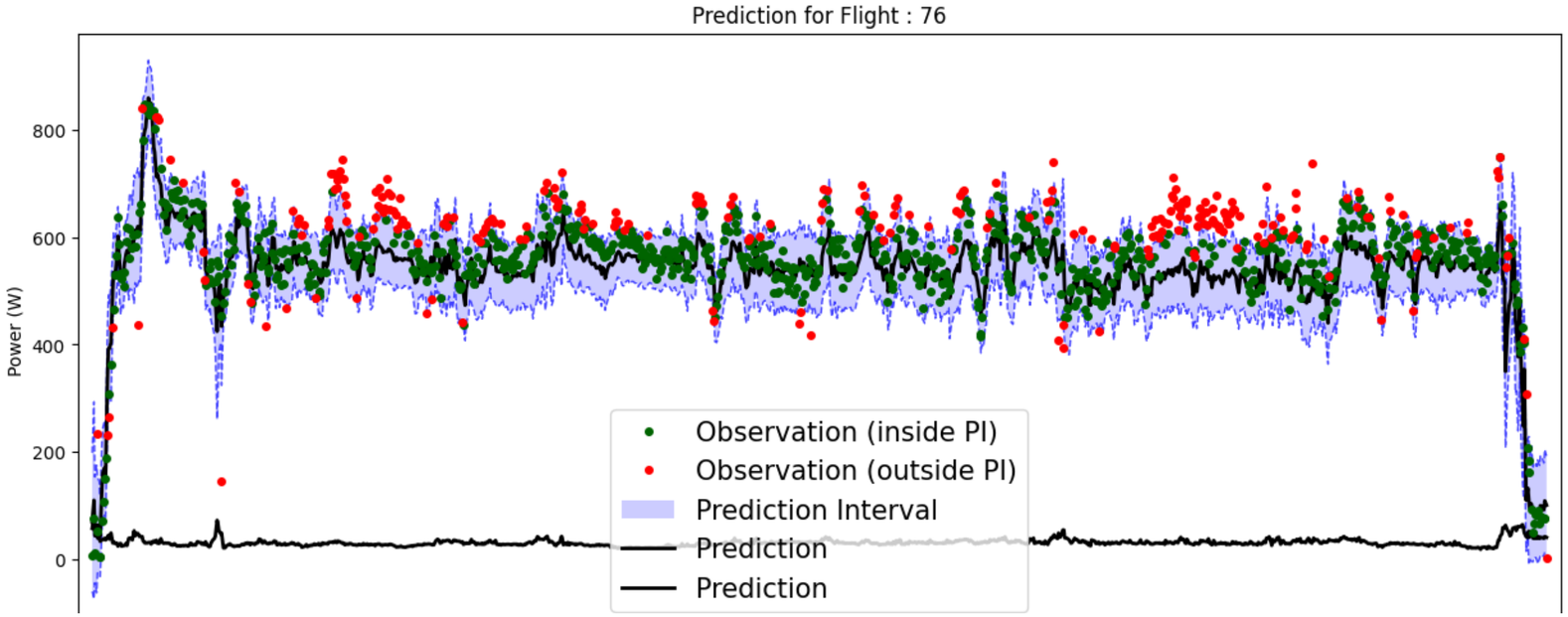
Locally adaptative Ensemble Batch Prediction Intervals



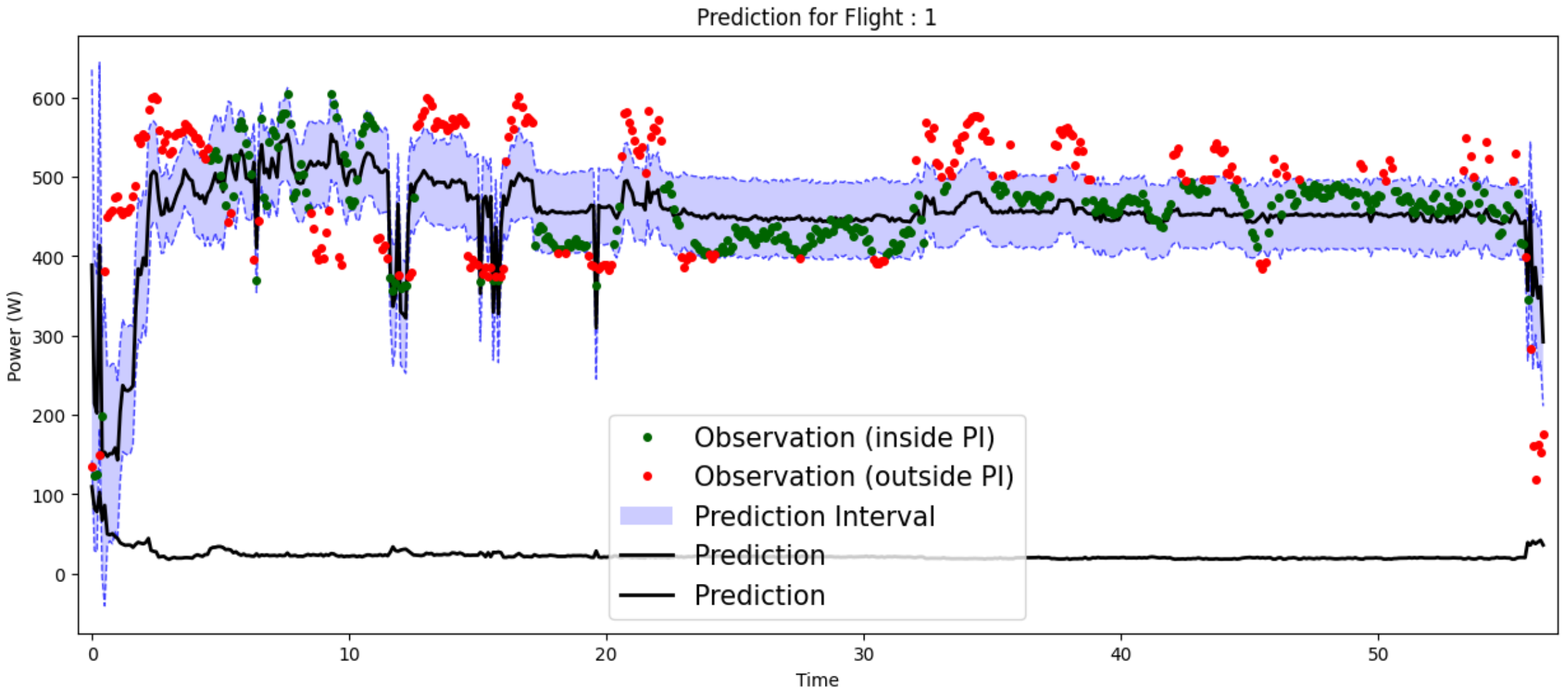
TEST dataset

Marginal coverage: 0.89 Average width:92.05

Locally adaptative Ensemble Batch Prediction Intervals



LOURD Dataset
Marginal coverage: 0.77
Average width: 139.7



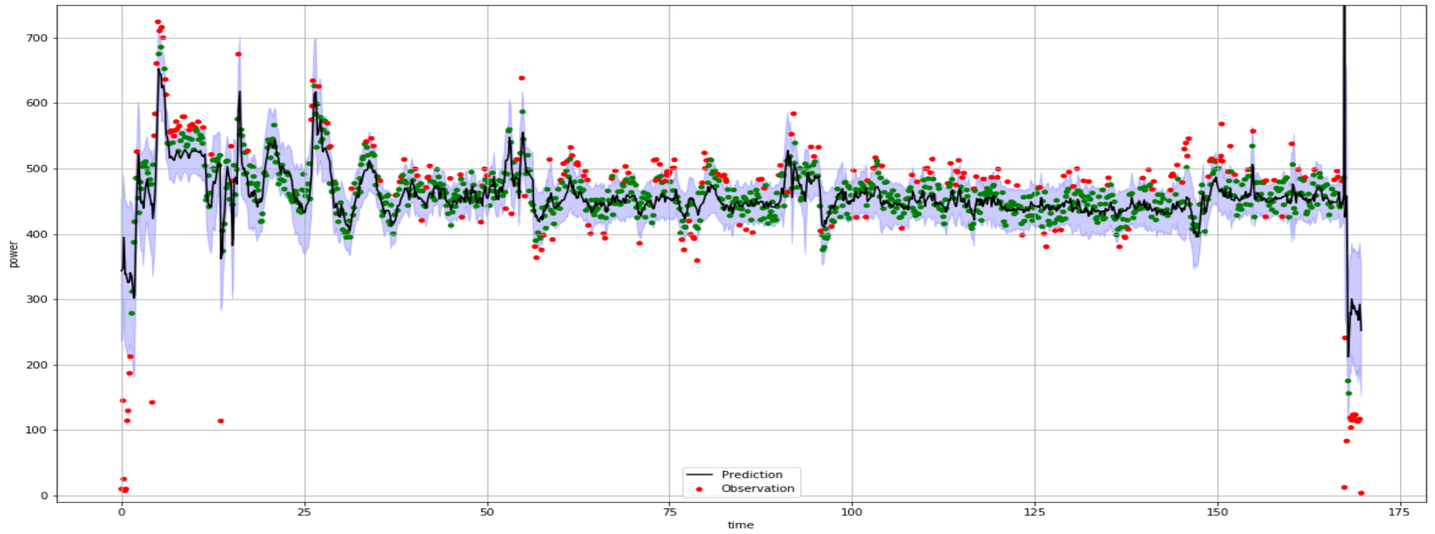
DA Dataset
Marginal coverage: 0.58
Average width: 102.61

Model Quality : Uncertainty quantification

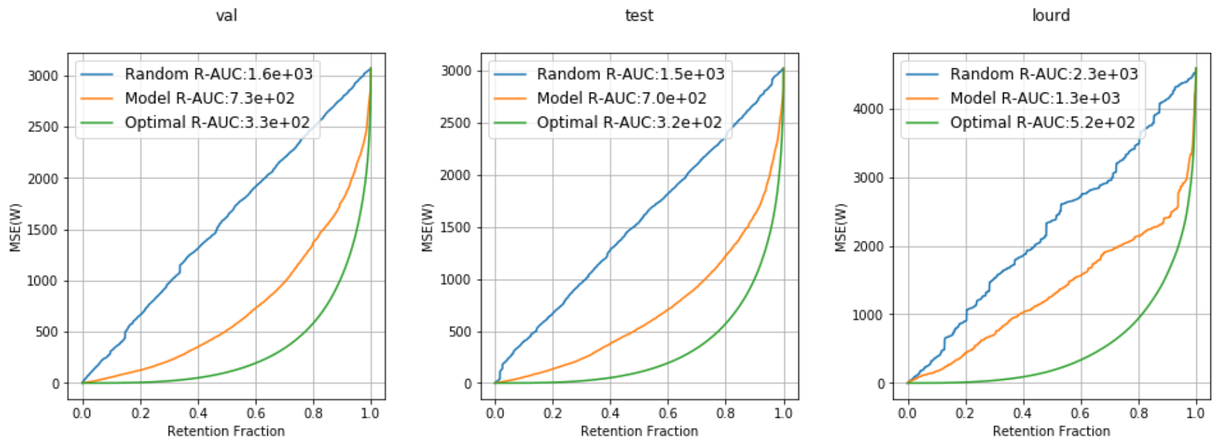
Deep Learning Model : Deep ensemble of ProbMCdropoutDNN models

From 1 input -> 400
inferences to compute mean and std

20 models in the ensemble
X 20 (sampling of by MC dropout model)



Evaluation :
the retention curve



Future works

DATA QUALITY

- **Representativeness and completeness**

MODEL QUALITY

- **Quantification of generalization bounds (operating domain definition)**
- **Certification of Robustness/stability (by formal methods)**

TRAIL

TRUSTED AI LABS

Summer Workshop 23' - Nantes

Thank you all for your attention!

Contact: jean@multitel.be