SGRE Data Ingestion Results with ModelarDB

ESR 2.3: Model-based for time series

March 8, 2023

Abduvoris Abduvakhobov

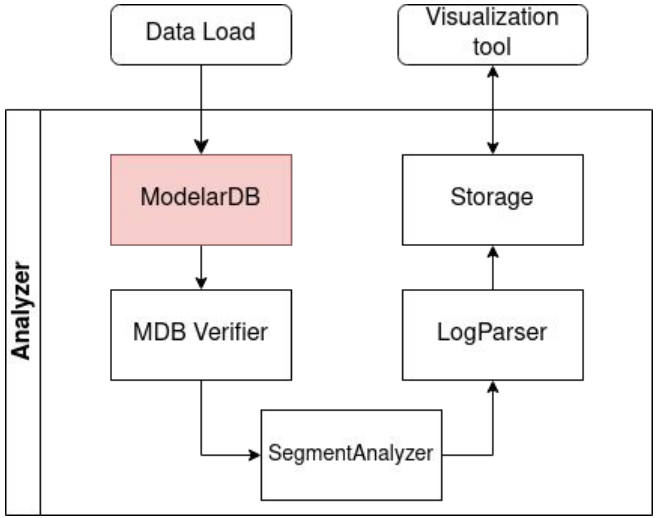
Current status

Tentative Title of Paper 1: A tool for analysis of the efficiency of model-based compression in ModelarDB

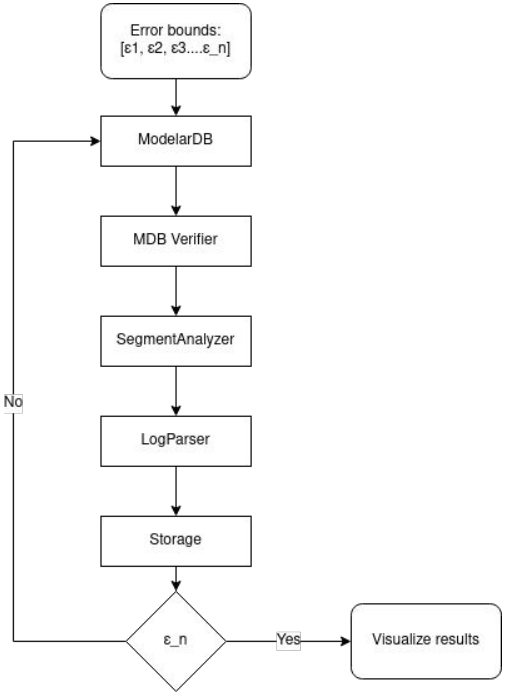
Tasks performed:

- Active collaboration with SGRE
- Python “Analyzer” tool
- Technical report with SGRE
- Scheduled visit to SGRE next week

Analyzer



Components of Analyzer

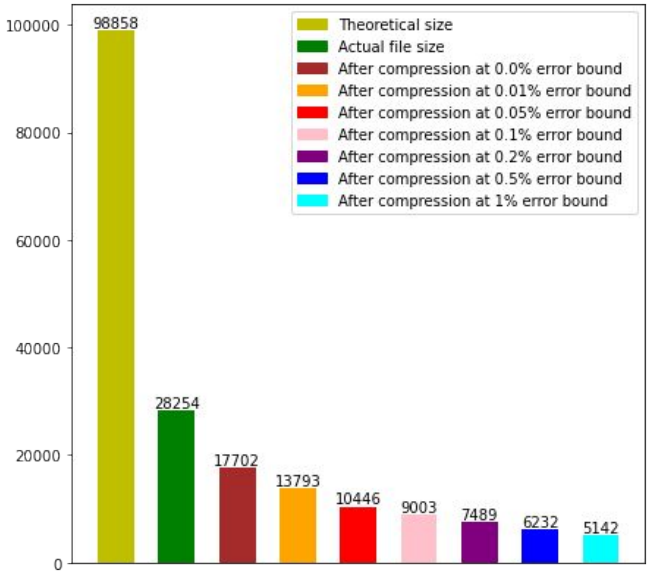


Analyzer running process

SOME RESULTS

Ingestion of the Full Dataset

- **2.5 years** of data from **Power Controller**
- **~480mln** rows, **46** columns (multivariate time series)
- **ORC file format**
- **150 ms** sampling interval (although not perfectly regular)
- **Error bounds used**: 0%, 0.01%, 0.05%, 0.1%, 0.2%, 0.5% and 1%
- ***Theoretical size*** *= Rows x (Columns x Value + Timestamp)*

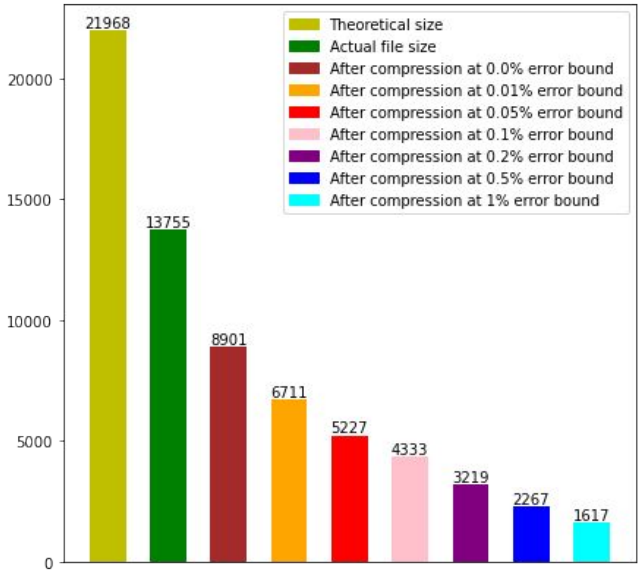


In MB

Ingestion of the Analytics Dataset

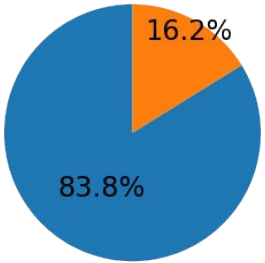
Used columns:

1. TimeStamp
2. ActivePower
3. ActivePower60
4. ActivePower600
5. AvailablePower
6. Frequency
7. PowerError
8. PowerLowerLimit
9. PowerUpperLimit
10. RawPower
11. RawReactivePower

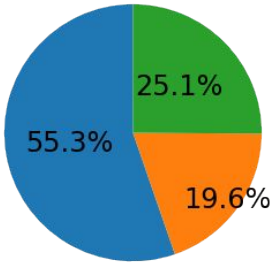


In MB

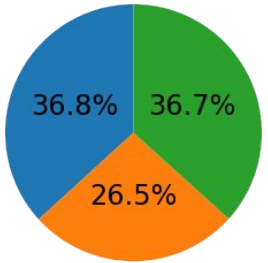
General model performance by data points (Analytics Dataset)

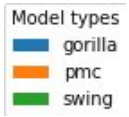


Error bound 0.0%

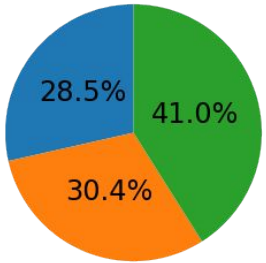


Error bound 0.01%

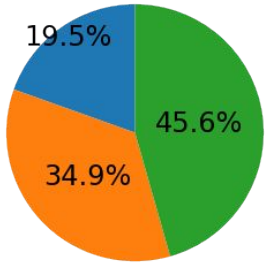




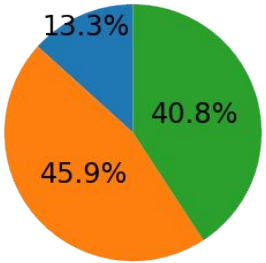
Error bound 0.05%



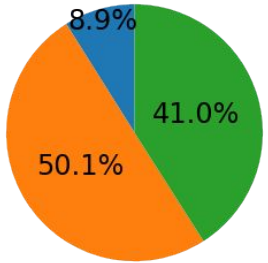
Error bound 0.1%



Error bound 0.2%

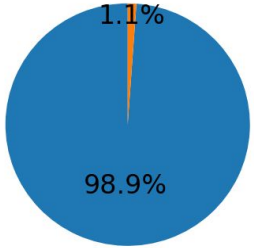


Error bound 0.5%

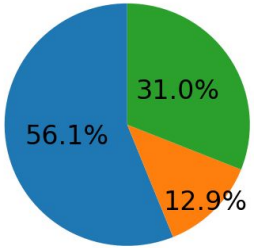


Error bound 1.0%

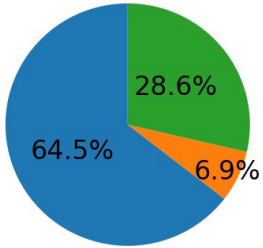
General model performance by segments (Analytics dataset)



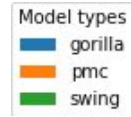
Error bound 0%

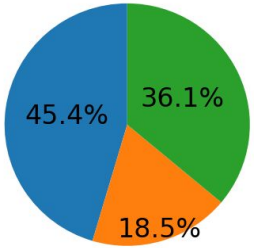


Error bound 0.05%

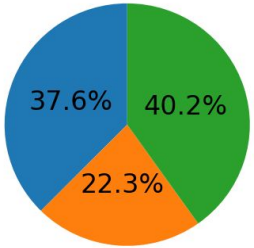


Error bound 0.01%

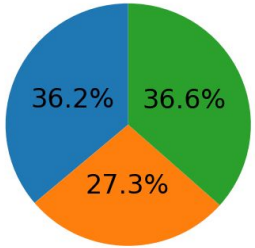




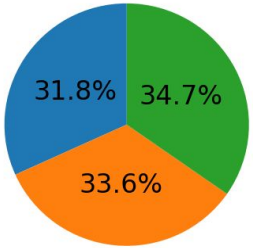
Error bound 0.1%



Error bound 0.2%

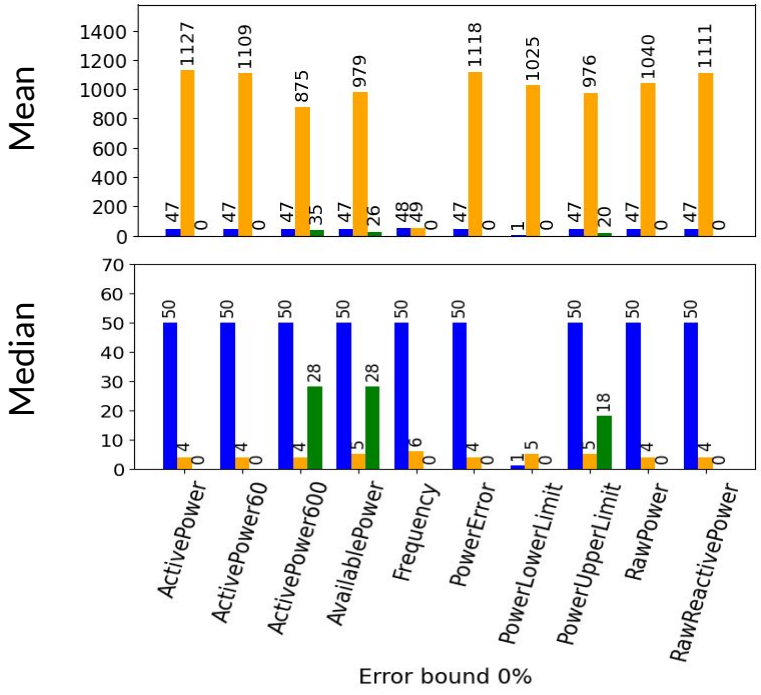


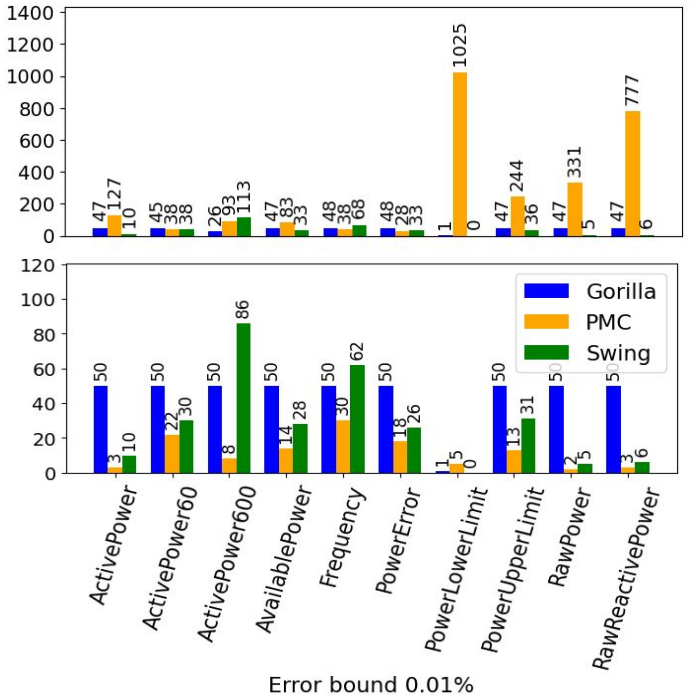
Error bound 0.5%



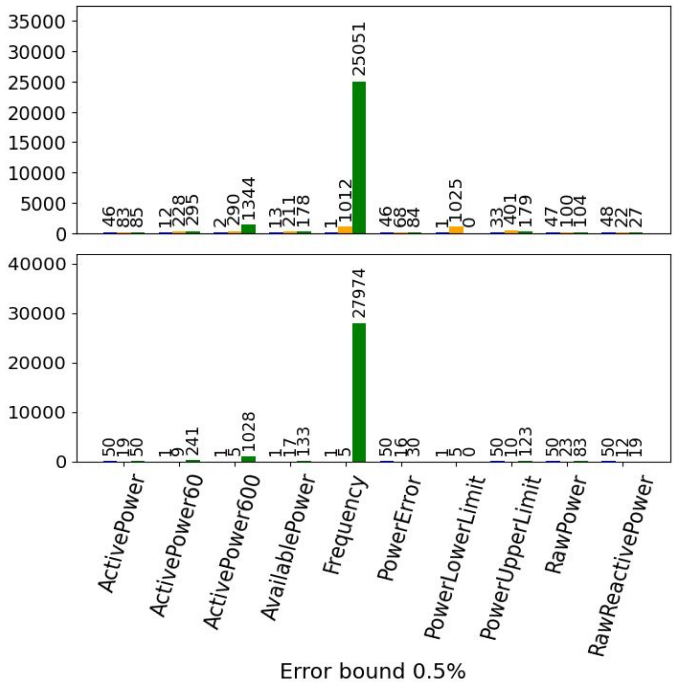
Error bound 1%

Length of segments

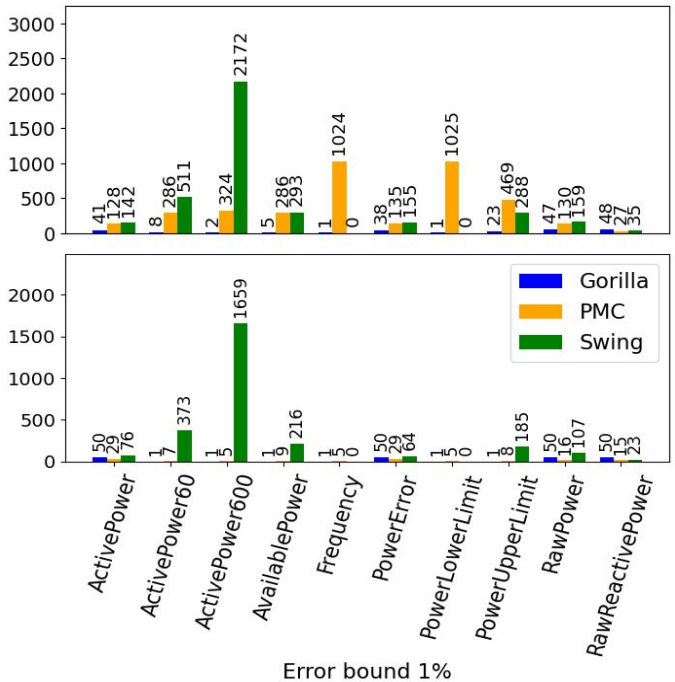




Length of segments

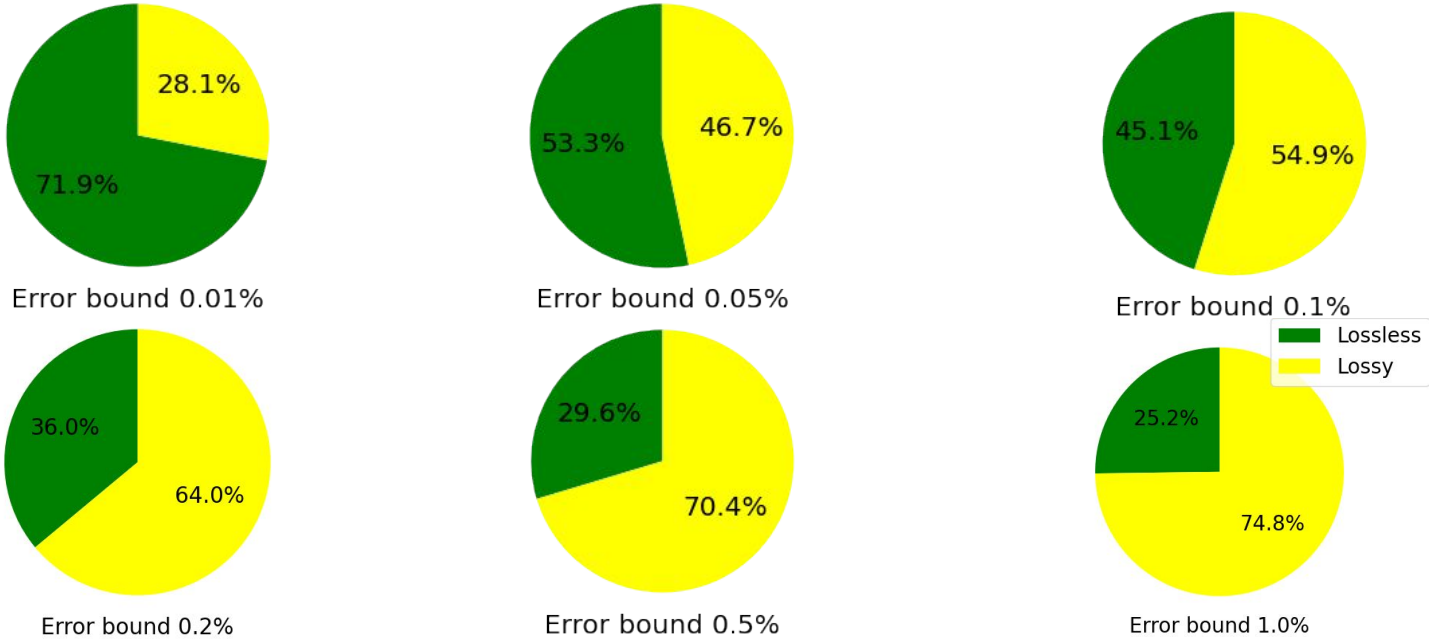


Error bound 0.5%

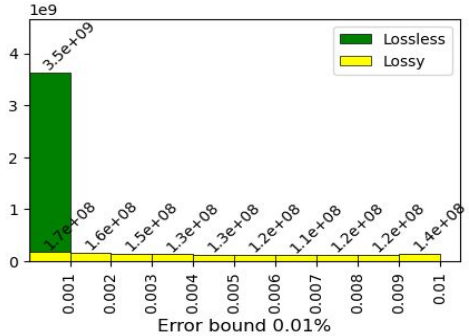


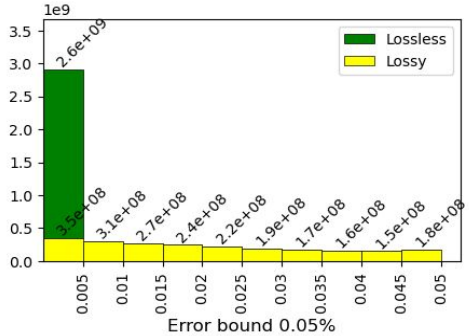
Error bound 1%

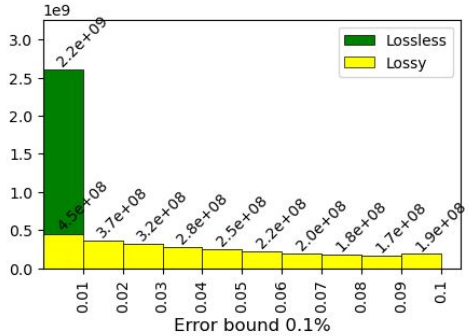
Losslessly compressed data points (Analytics Dataset)

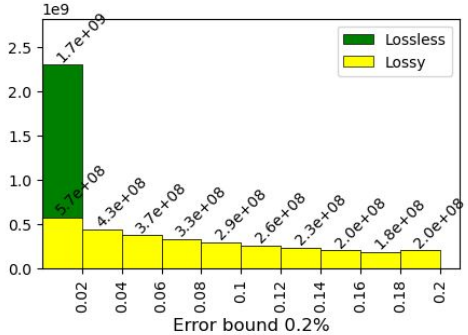


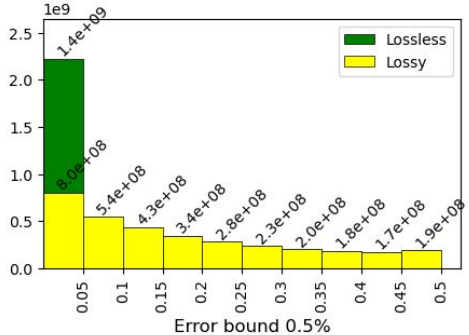
Distribution of actual errors (Analytics dataset)

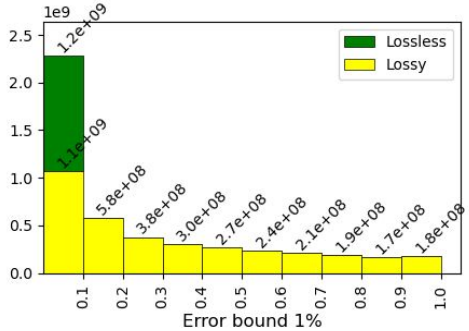






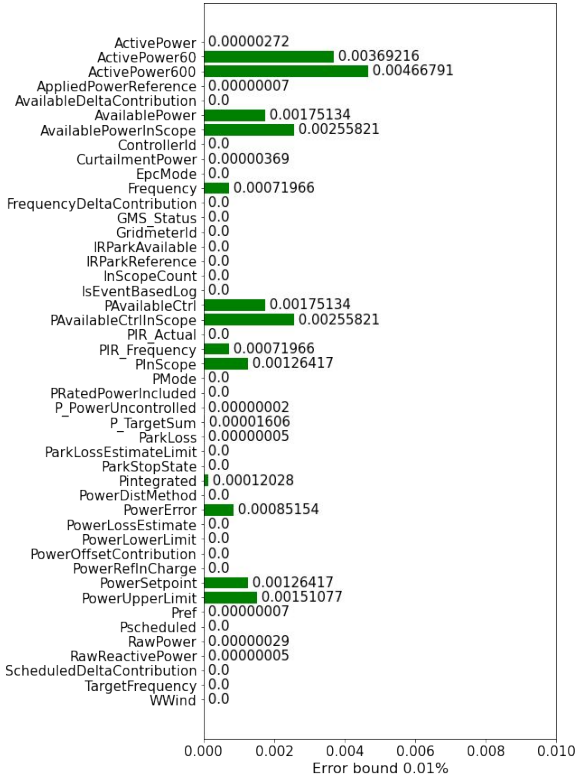




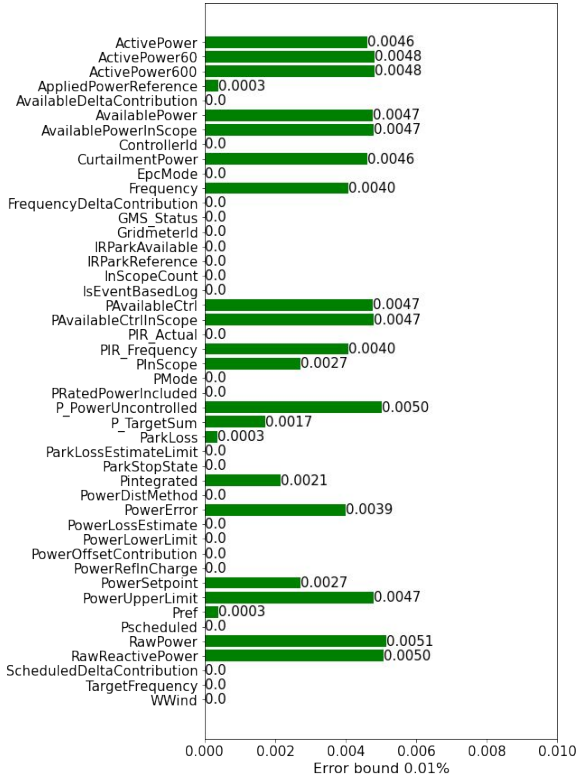


Average actual error (0.01% error bound)

MAPE (Mean Average Percentage Error)

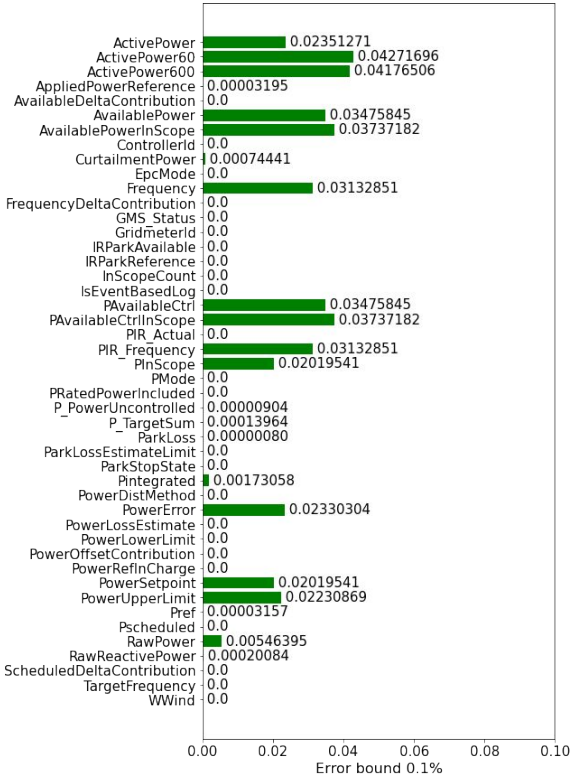


MAPE, |r| > 0

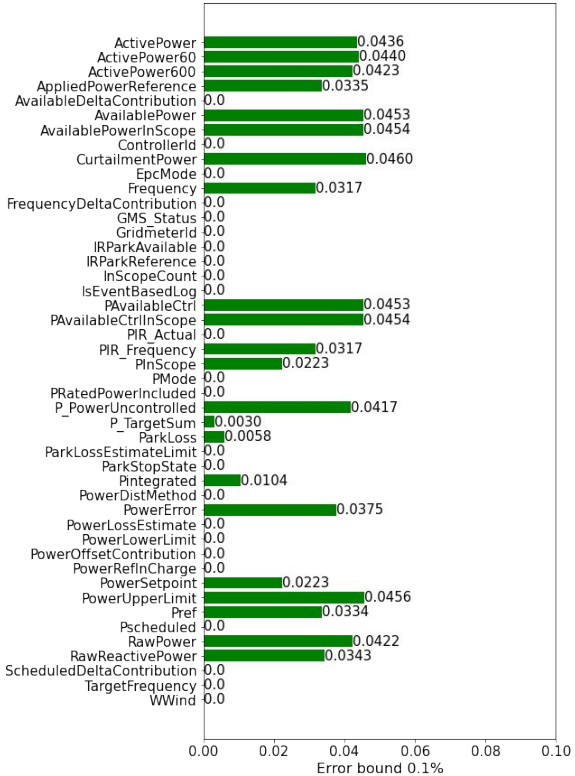


Average actual error (0.1% error bound)

MAPE

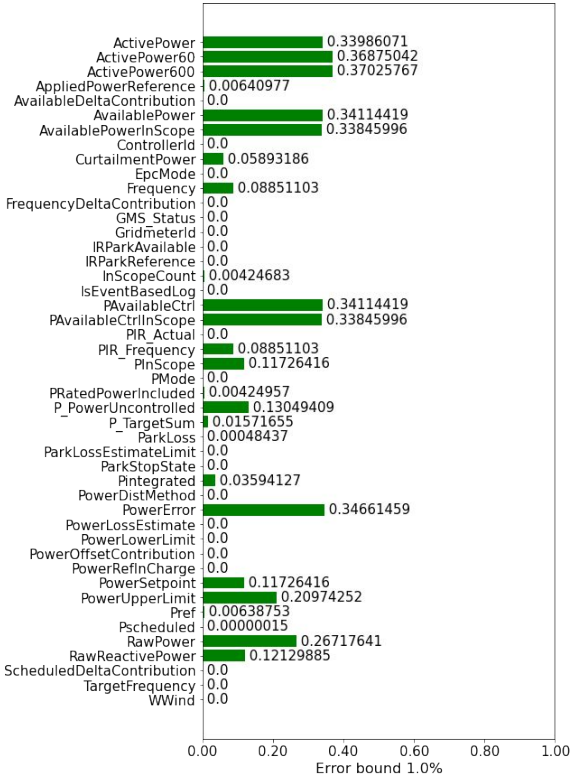


MAPE, |r| > 0

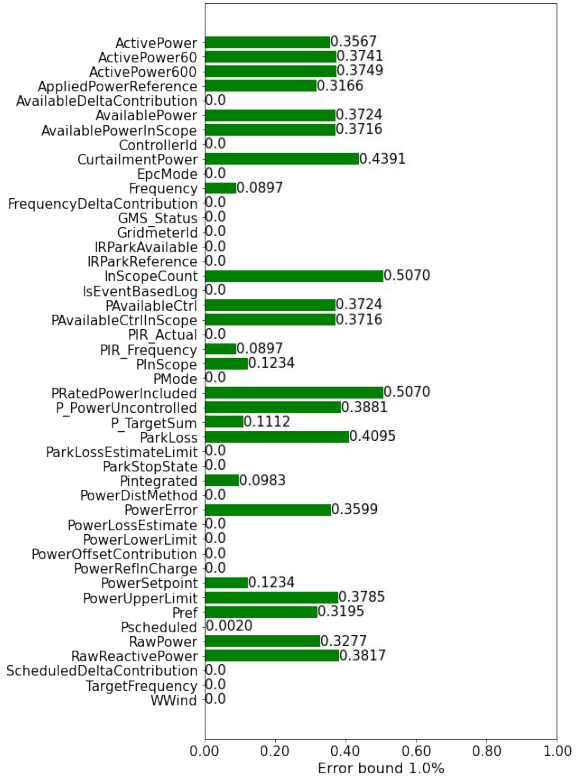


Average actual error (1% error bound)

MAPE



MAPE, |r| > 0



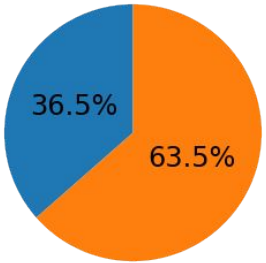
Next steps

- Technical report with SGRE
- Convert it into a paper
- New model types
- Optimization of existing ones
- Constant performance evaluation using Analyzer

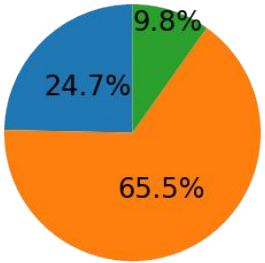
Feedbacks from audience

1. Mention a vision of this paper. What is the goal, targets you set, contribution you make
2. What could be improved in models you use?
3. For actual average error calculation, think about including percentiles to convey a more extensive message

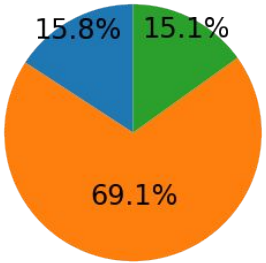
General model performance by data points (Full Dataset)



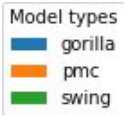
Error bound 0.0%

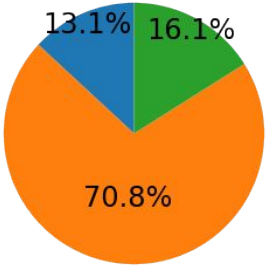


Error bound 0.01%

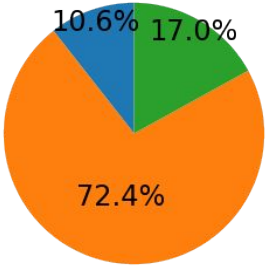


Error bound 0.05%

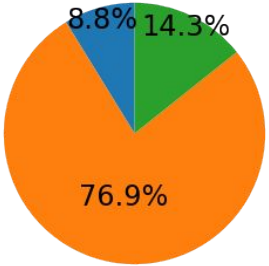




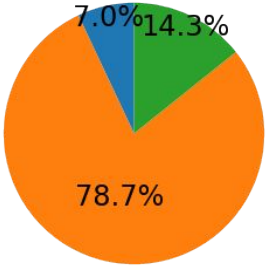
Error bound 0.1%



Error bound 0.2%

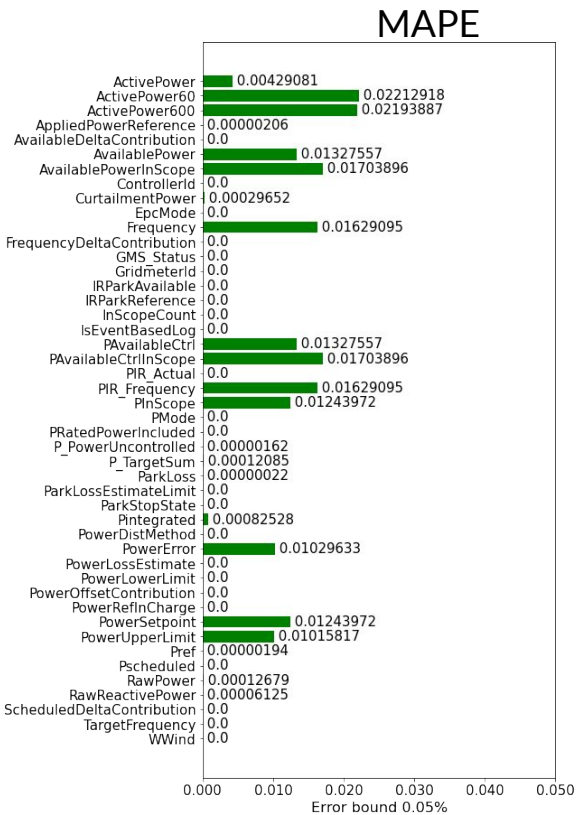


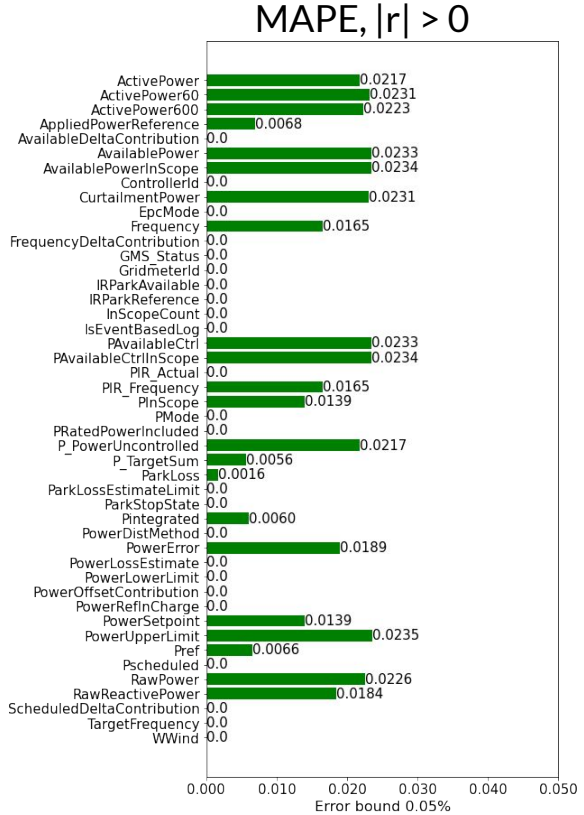
Error bound 0.5%



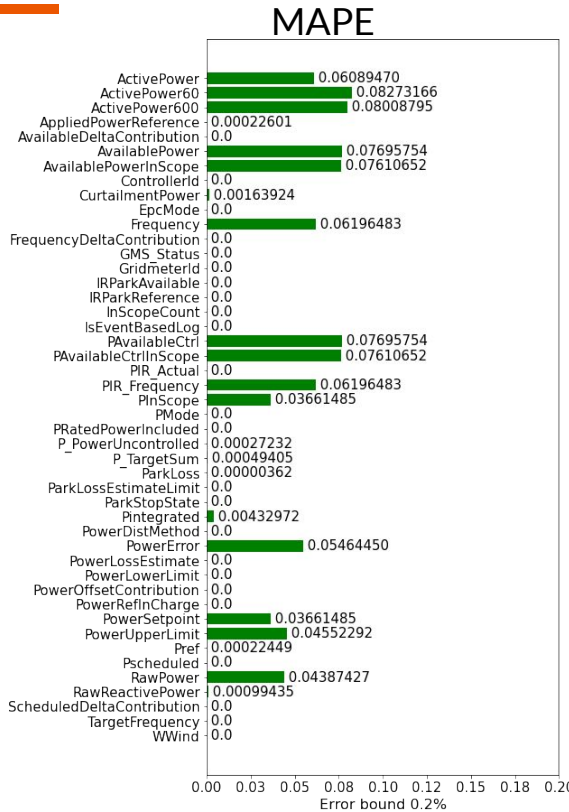
Error bound 1.0%

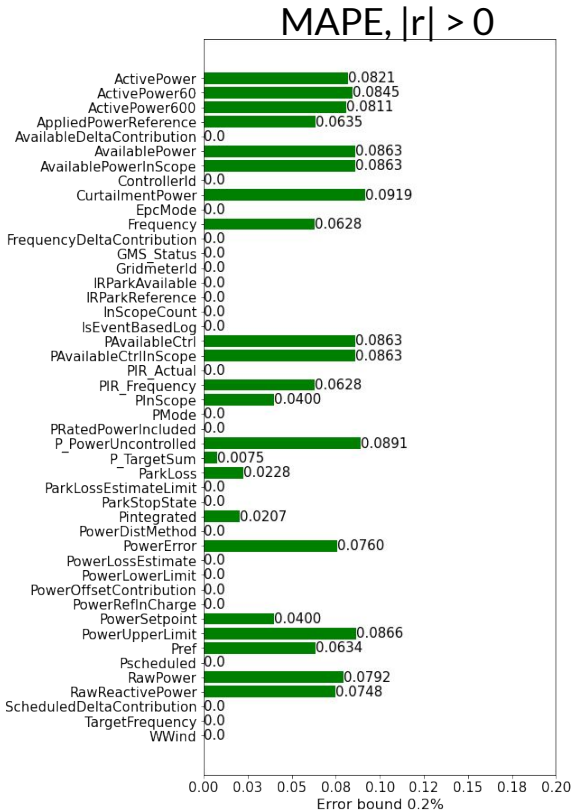
Average actual error (0.05% error bound)





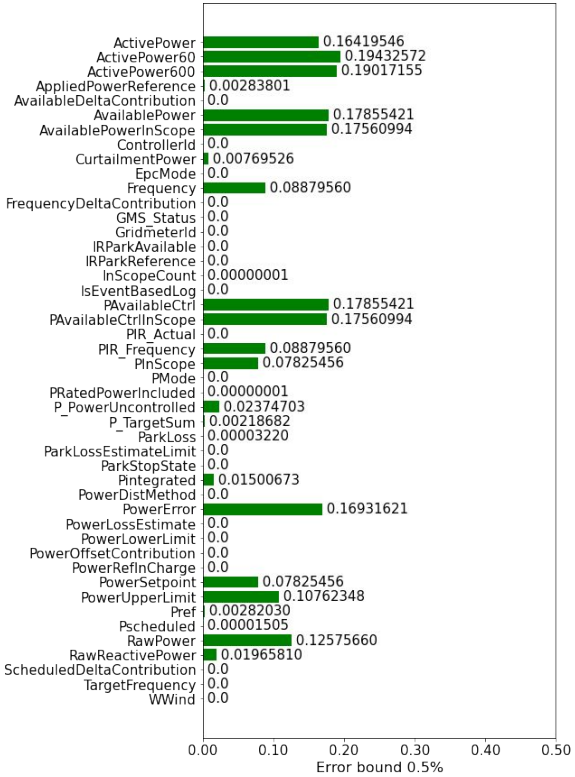
Average actual error (0.2% error bound)



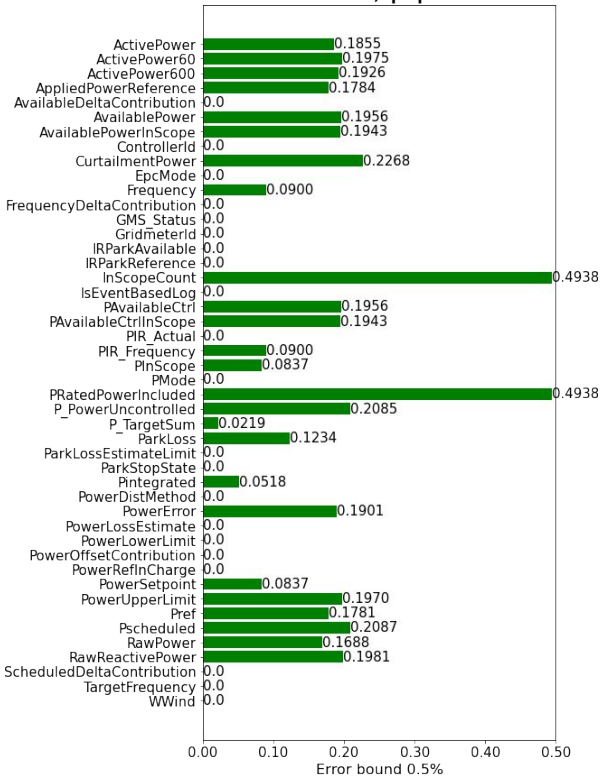


Average actual error (0.5% error bound)

MAPE



MAPE, |r| > 0



Length of segments