Filling Gaps in Complex Terrain Using LightGBM

Lucas Jurema

Intro

This study introduces a new Measure Correlate Predict (MCP) method for data reconstruction using the Machine Learning (ML) model LightGBM [1], as an alternative for filling gaps in complex terrain, addressing its impact on wind resource patterns.

Methods

1. The validation set comprised 3 clusters representing regions with different terrain complexities: non-complex (NCT), complex I (CT I - shear inversion), and complex II (CT II - Multimodal Distribution). The clusters contained 9, 9, and 4 metmasts, respectively, each with 4 types of artificial gaps (1, 10, 100, 1000 hours) per metmast.
2. Benchmarking using the MTS method results from the Windographer software.
3. Single parameterization for all sets.
4. Metrics: Mean Absolute Error (MAE), Pearson's chi-squared test statistic as the Distribution Error (DE) [2] and Energetic Error (EE) calculated using the NREL Onshore Wind Turbine Power Curves [3].

Results

- LightGBM achieved better results in time series affected by complex terrain.
- LightGBM delivered competitive results while maintaining a good trade-off between accuracy and distribution error.

Discussion

- The fact that MTS performed better in Non-Complex Terrain may indicate that the LightGBM model could extract and use more relevant information in Complex Terrain.

Patterns in Time Series Associated with Complex Terrain

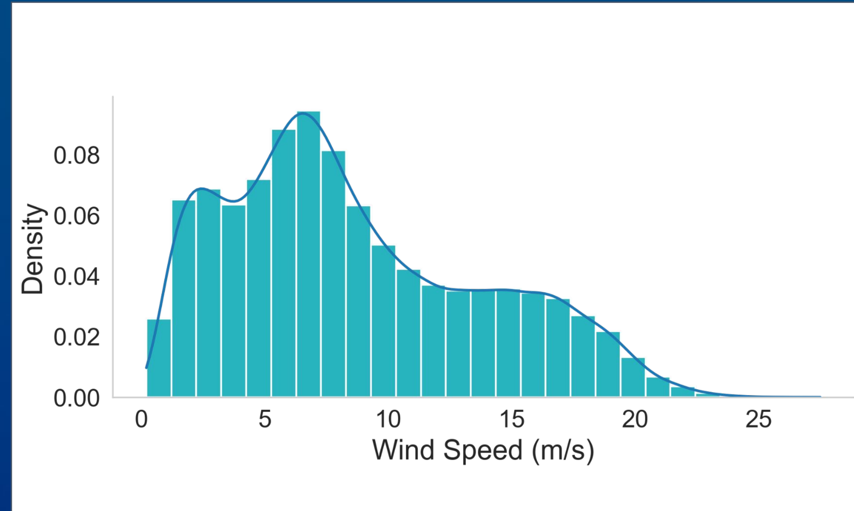

Figure 1: Multimodal wind speed distribution

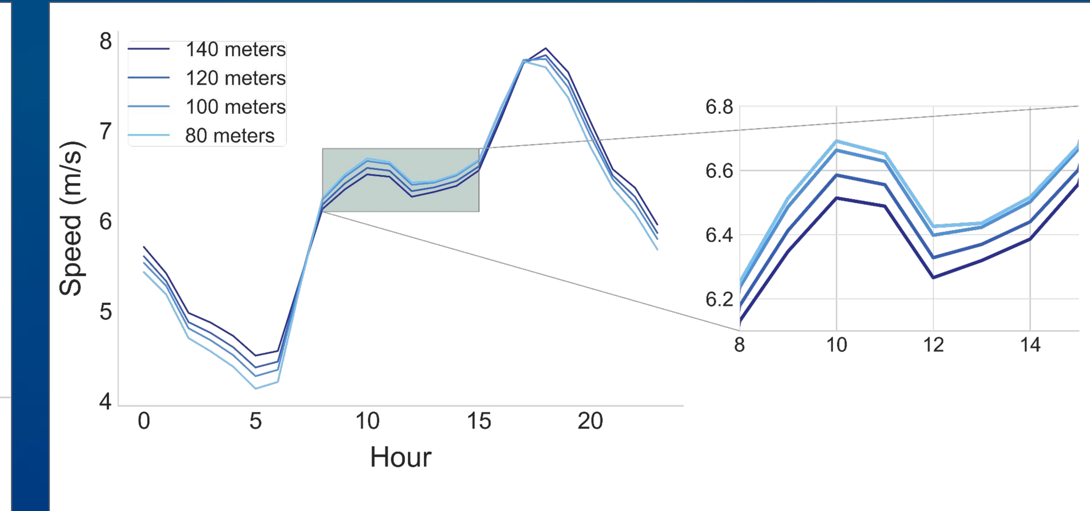

Figure 2: Shear Inversion in direction bin 75° (31% of the time series)

Filling GAPS

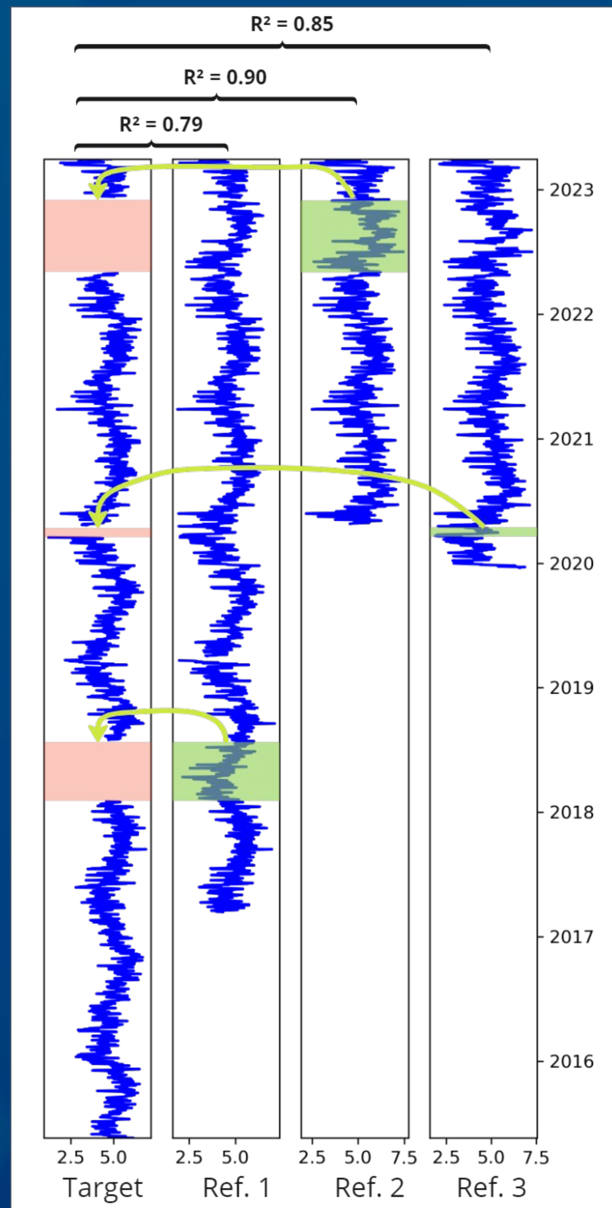

Figure 8: Example of a procedure for MCP Filling Gaps

Accuracy vs. Distribution Error: **A Trade-off**

Model Tuning

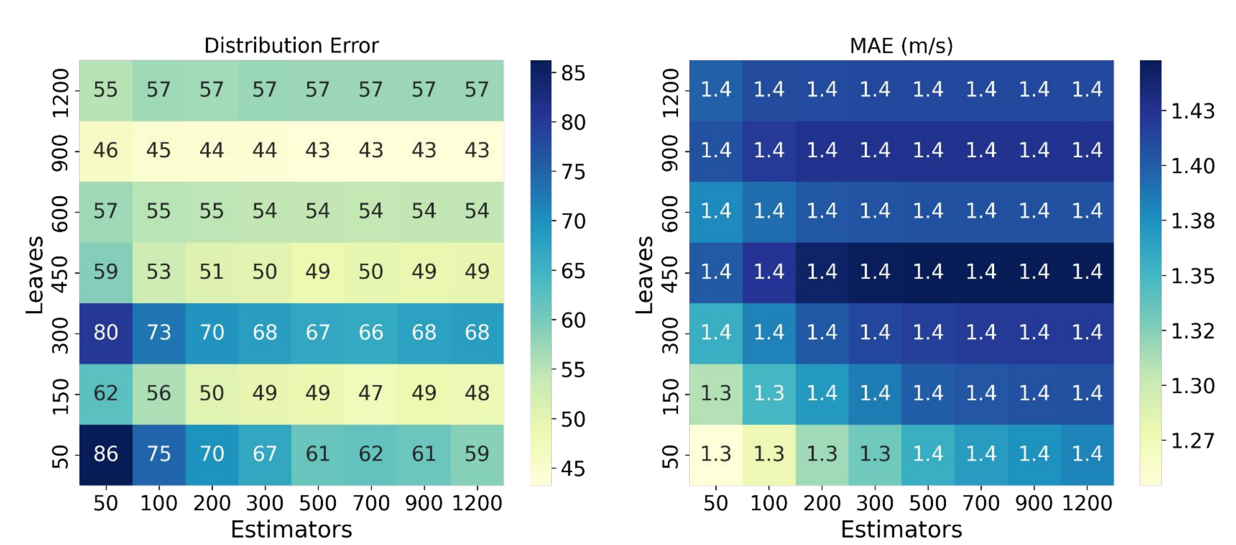

Figure 3: Model complexity effect over the DE and accuracy. Complexity by Leaves and Estimators. The learning rate was set to 0.5. As a reference, the results using the MTS method for MAE and DE are 1.52 (m/s) and 30.01, respectively.

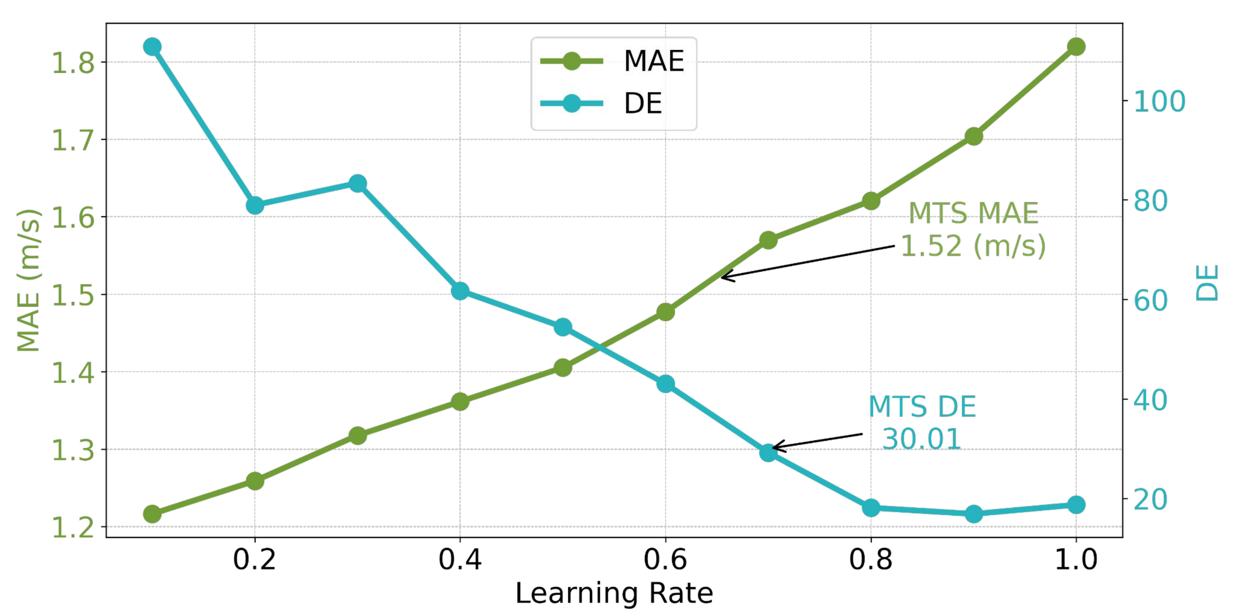

Figure 4: Trade-off between DE and Accuracy by Learning Rate. Number of leaves = 600 and number of estimators = 300.

Results

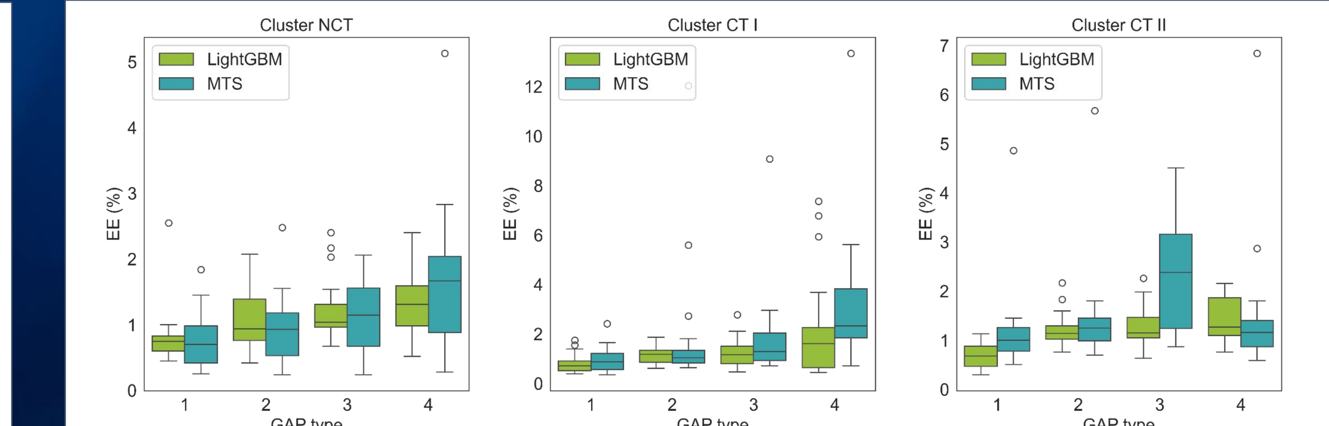

Figure 5: Distribution of Energetic Error (EE) by cluster and gap type.

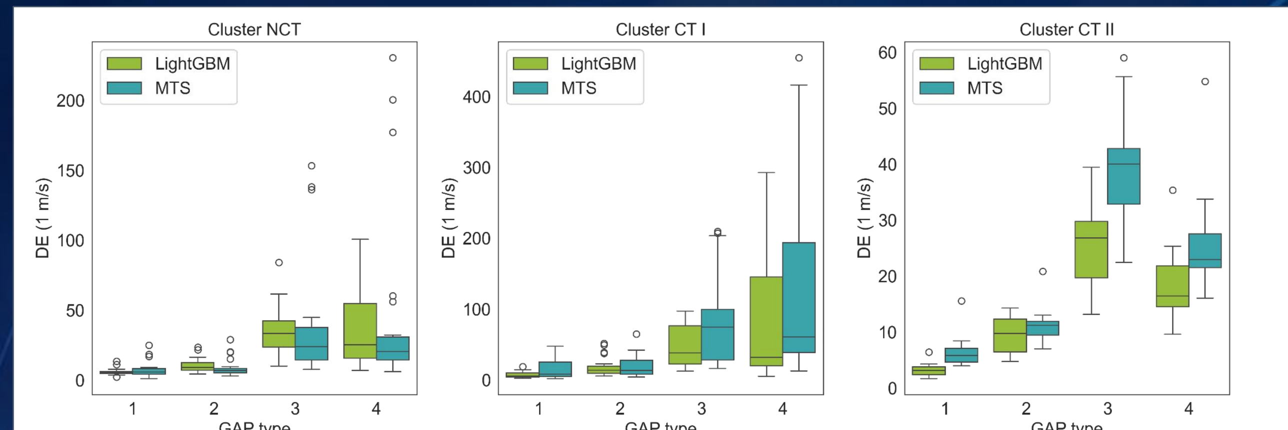

Figure 6: Distribution of Distribution Error (DE) by cluster and gap type.

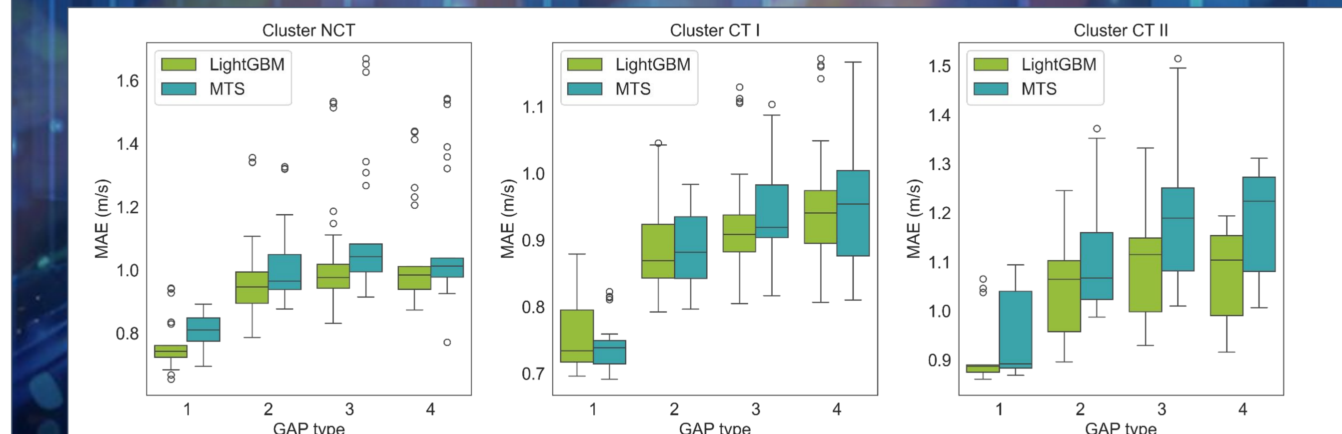

Figure 7: Distribution of MAE by cluster and gap type.

RESOURCE &TECH

AMERICAN CLEAN POWER

Hyperparameters/Features

Hyperparameters

task : 'train'
boosting_type : ['gbdt']
objective : 'regression'
metric : ['l1' , 'l2']
learning_rate : 0.5
lambda_l1 : 0.0
lambda_l2 : 0.0
min_child_samples : 1
feature_fraction : 1
bagging_fraction : 1
verbose : 0
num_leaves : 600
n_estimators : 300

Features

Hour of the Day

Reference Wind Speed:
lags:
[1, 2, 3, 6, 18]
leads:
[1, 2, 3, 6, 18]
Mean, Max, Min:
[hour, day, week, Month]
Percentiles (.9, .8, .7, .3, .2, .1) :
[hour, day, week, Month]
Skewness and Kurtosis:
[day, week, Month]

Summarized Results

Table 1. Energetic Error Mean (Abs. %)

Cluster	GAP type	LightGBM	MTS
CT I	1	0.79	0.94
CT I	2	1.15	1.66
CT I	3	1.21	1.75
CT I	4	1.98	2.99
CT II	1	0.69	1.27
CT II	2	1.24	1.56
CT II	3	1.35	2.35
CT II	4	1.42	1.69
NCT	1	0.80	0.77
NCT	2	1.05	0.95
NCT	3	1.19	1.15
NCT	4	1.31	1.68

Table 2. Mean Absolute Error (m/s)

Cluster	GAP type	LightGBM	MTS
CT I	1	0.76	0.74
CT I	2	0.89	0.88
CT I	3	0.93	0.94
CT I	4	0.95	0.96
CT II	1	0.92	0.95
CT II	2	1.05	1.12
CT II	3	1.10	1.22
CT II	4	1.07	1.18
NCT	1	0.76	0.81
NCT	2	0.99	1.02
NCT	3	1.04	1.12
NCT	4	1.05	1.09

References

1. LightGBM's documentation. 2024. Available from: https://lightgbm.readthedocs.io/en/latest/index.html
2. UL SOLUTIONS. Windographer: Wind data analytics and visualization solution. Available from https://store.ul-renewables.com/products/windographer
3. NREL Wind Turbine Power Curve Archive Repository. 2020. Available from https://github.com/NREL/turbine-models/tree/master.

Contact Info

L. JUREMA, Data Specialist, CDV Desenvolvimento
lucas.cordeiro@casadosventos.com.br

cdv desenvolvimento