

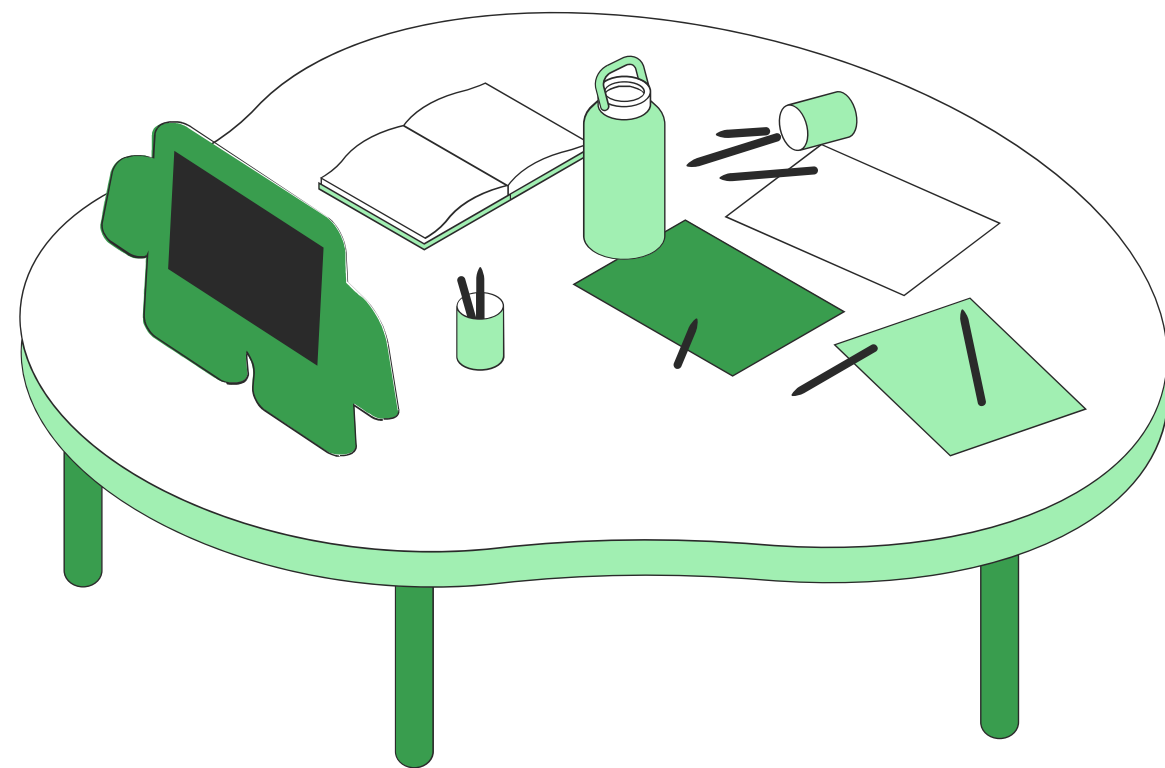
External presentation

Imputation Project team

Adrien, Albert, Jesús, Julien, Michael, Ramon
Imputation Project team



Table of content



What is the project about?

The testing method

The imputation methods

What's next?

What is the project about?

Applied Data Science and imputation

Imputation

Impute Building Management System time-series data.

Test methods

Multiple imputation methods studies and tested.

Write guidelines

Security codes and passwords could be easily accessed.

The Testing Method



Loading data

Load data without gaps



Creating gaps

Create reproducible gaps of different



Running imputations

Hot-swappable imputation methods

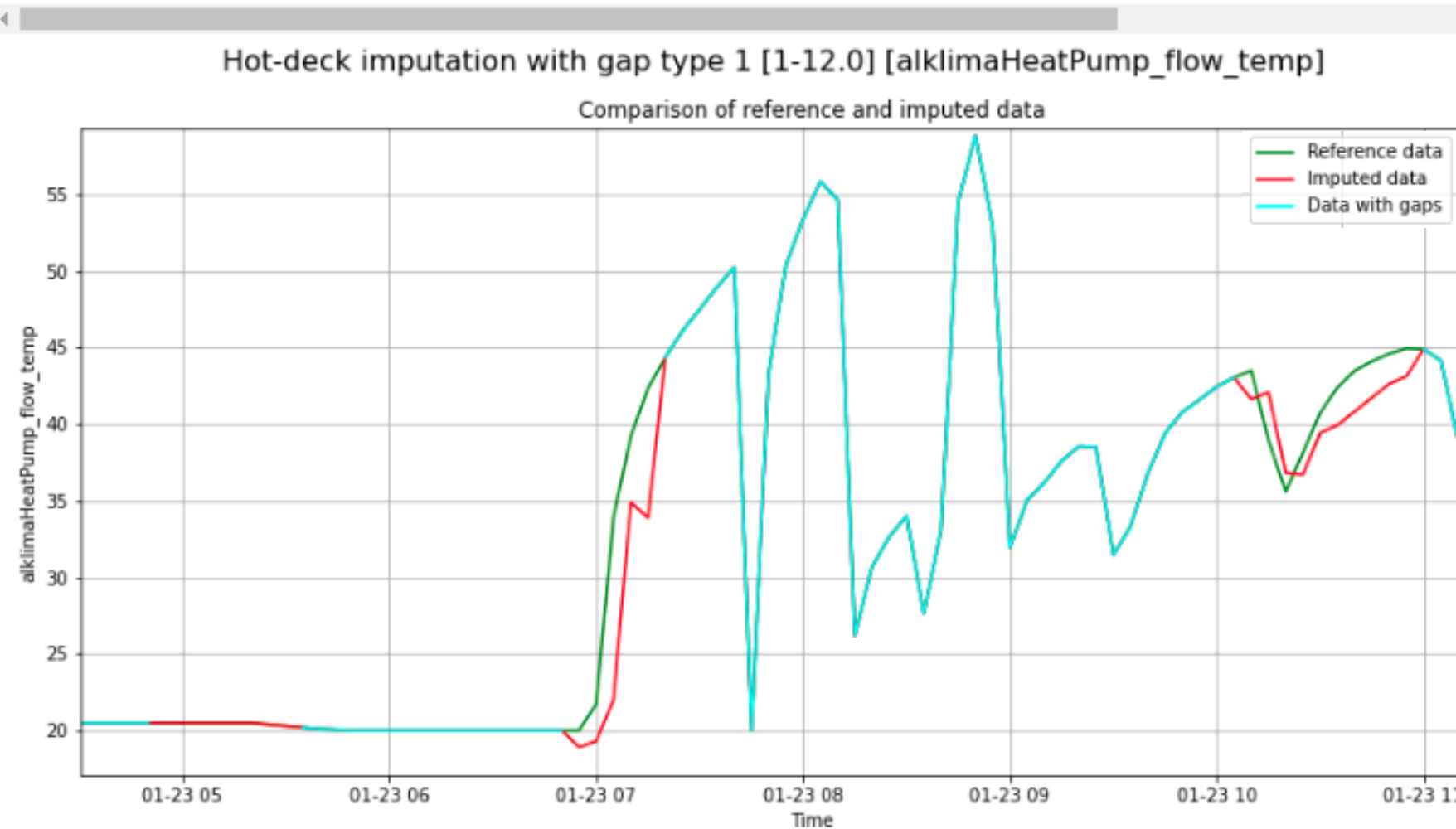


Evaluating results

Wide range of evaluation criteria

	Mean squared error	Raw bias	Absolute bias	Percent bias	Errors sum	Mean minimum error	Mean maximum error	Maximum error	Variance error	Kurtosis error	...
Hot-deck imputation - alklimaHeatPump_flow_temp - gap type 1 - 2021-12-09 15:10:00	6.032	1.291	2.427	-5.428	32165.719	0.054	1.841	44.788	6.305	-0.250	...
Hot-deck imputation - alklimaHeatPump_flow_temp - gap type 2 - 2021-12-09 15:10:00	13.125	0.866	4.105	-3.558	77597.075	-1.735	-0.130	48.260	9.134	-0.259	...
Hot-deck imputation - alklimaHeatPump_flow_temp - gap type 3 - 2021-12-09 15:10:00	23.195	0.103	7.067	-0.410	136638.115	12.110	21.905	54.067	9.484	-0.106	...
Hot-deck imputation - alklimaHeatPump_flow_temp - gap type 4 - 2021-12-09 15:10:00	25.070	1.178	7.592	-5.003	149407.675	1.632	18.724	53.057	8.732	-0.320	...
Hot-deck imputation - alklimaHeatPump_flow_temp - gap type 5 - 2021-12-09 15:10:00	33.448	1.117	6.695	-4.665	219720.519	0.403	26.540	53.578	23.054	-1.713	...

5 rows × 21 columns



Imputation Methods

Trying out as many methods as possible

Methods:

- Interpolation
- Mean / Median / Mode
- Softimpute
- Stochastic regression
- Deterministic regression

- k-NN
- Miss Forest

- Hot-deck
- Recurrent Neural Network

Recurrent Neural Network



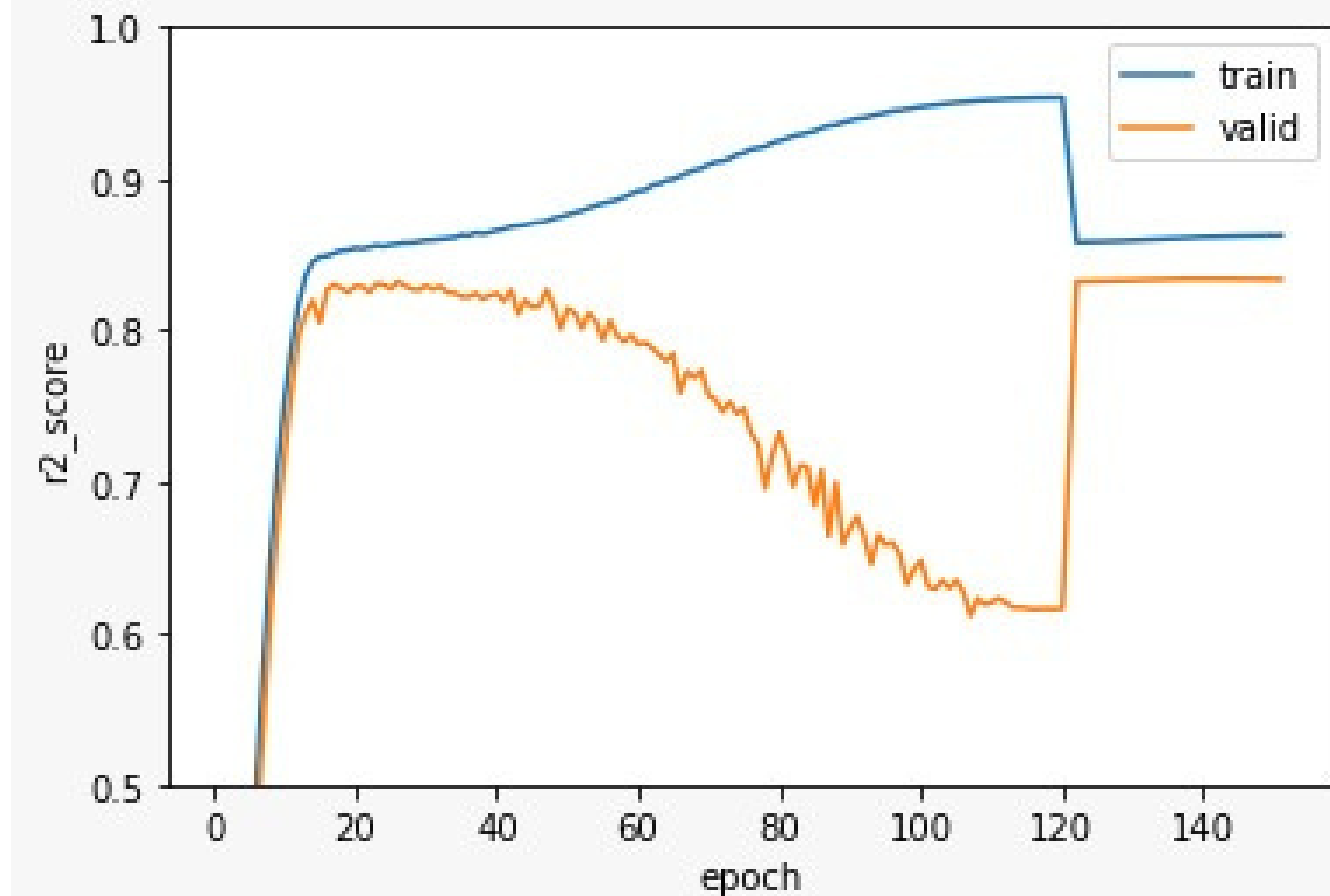
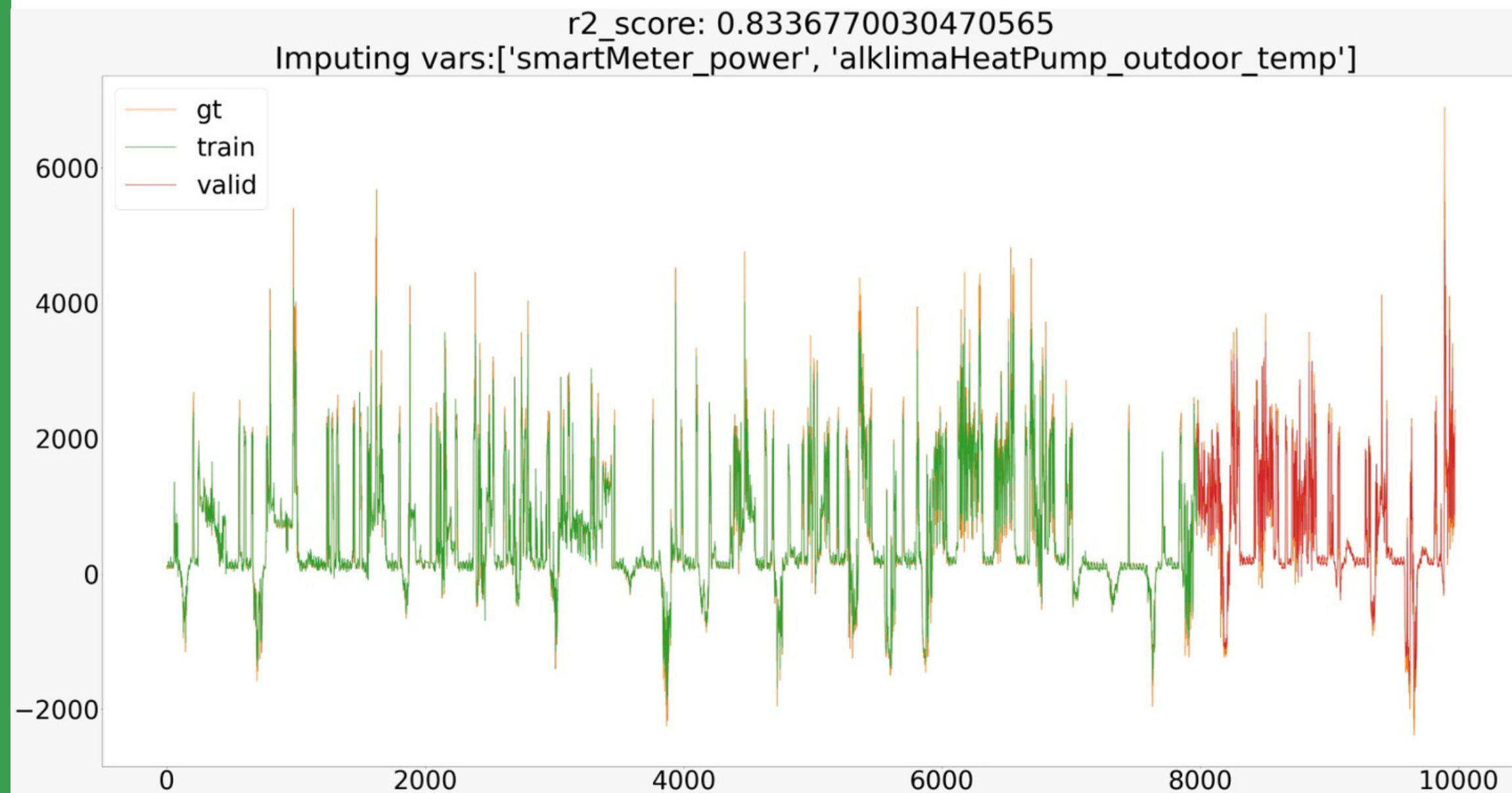
First results with RNN

Satisfactory results, still being improved



Bi-directional RNN

Use more input data for predictions



What's next?



RNN

More, and better
results with bi-
directional
imputation

Collect data

Evaluate the
imputation
methods

Fine tuning?

If necessary, fine
tune the
imputation
methods

Expose the data

Finalize the
research paper



Do you
have any
questions?