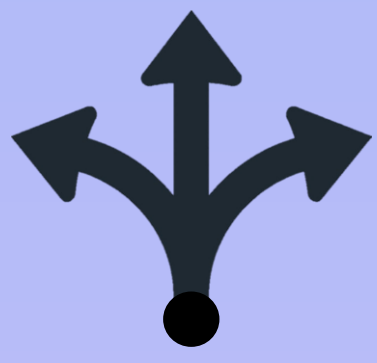


The Problem



Accurate wind power forecasts reduce uncertainty in the operation of the power system by allowing better scheduling and trading decisions¹.



Statistical forecasts that use the most recent wind speed and power measurements are commonly used for very-short-term (minutes to hours ahead) horizons².



Inaccurate assumptions and imperfect datasets mean forecast accuracy is reduced in a real life setting unless proper mitigation methods are used.

Key Results

- Missing training data can have a significant impact on results if not dealt with properly; multiple imputation is the best method to compensate (Fig. 3)
- If inputs for issuing a forecast are missing, retraining a model without these inputs shows greater performance than filling these values using a regression model with the remaining available inputs (Fig. 1 and 2)
- Spatio-temporal models including a greater number of sites are more robust to missing data (Fig. 1)
- Forecasts continue to worsen with increasing length of missing period, but the largest proportion of the loss of forecast skill comes from missing the most recent information (Fig. 2)

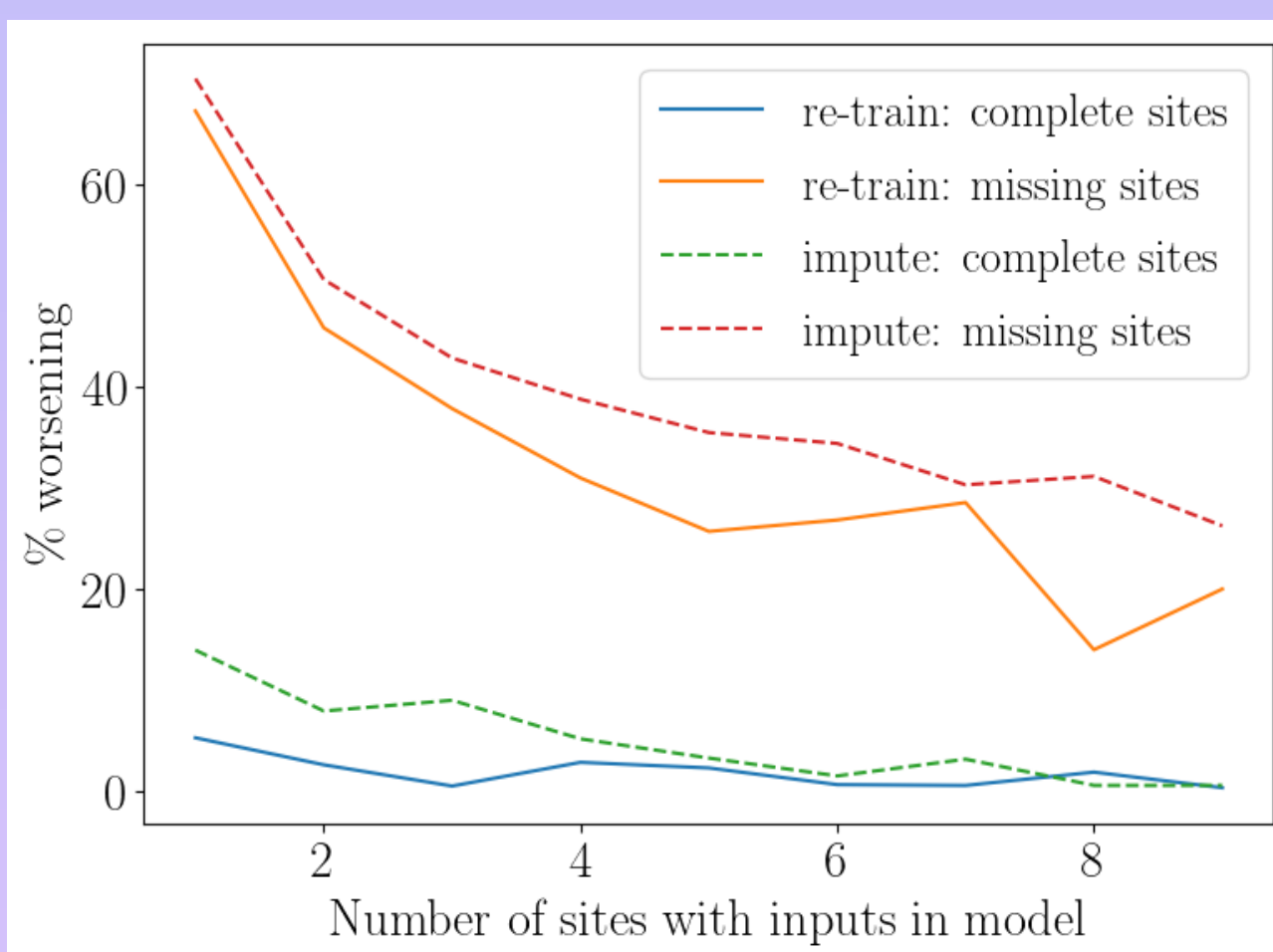


Fig. 1: Forecast error improves when more sites are included in the model, particularly for sites that are missing forecast input data.

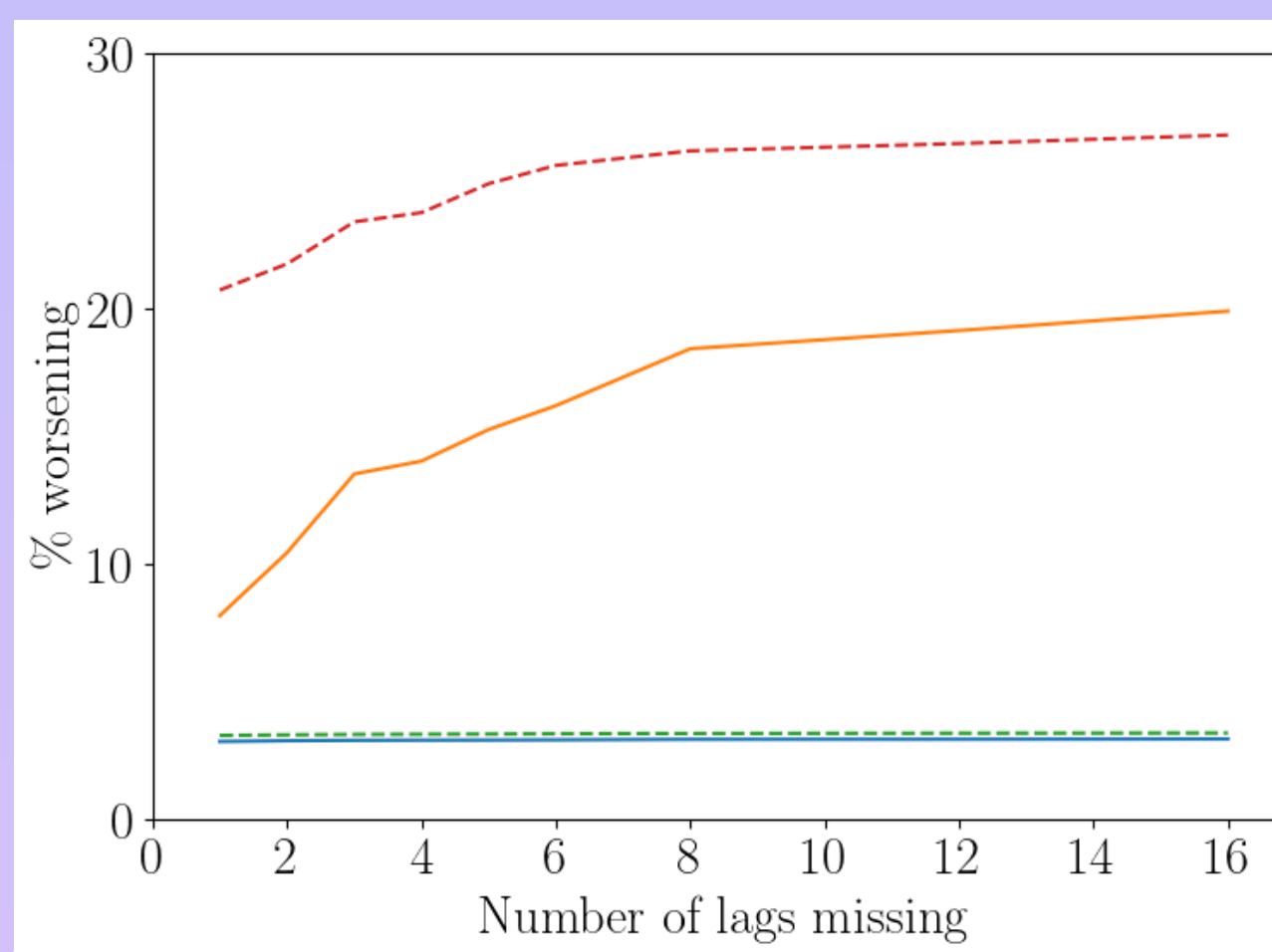


Fig. 2: The impact of cumulative lags missing shows the forecast at the missing site is worse for longer missing periods

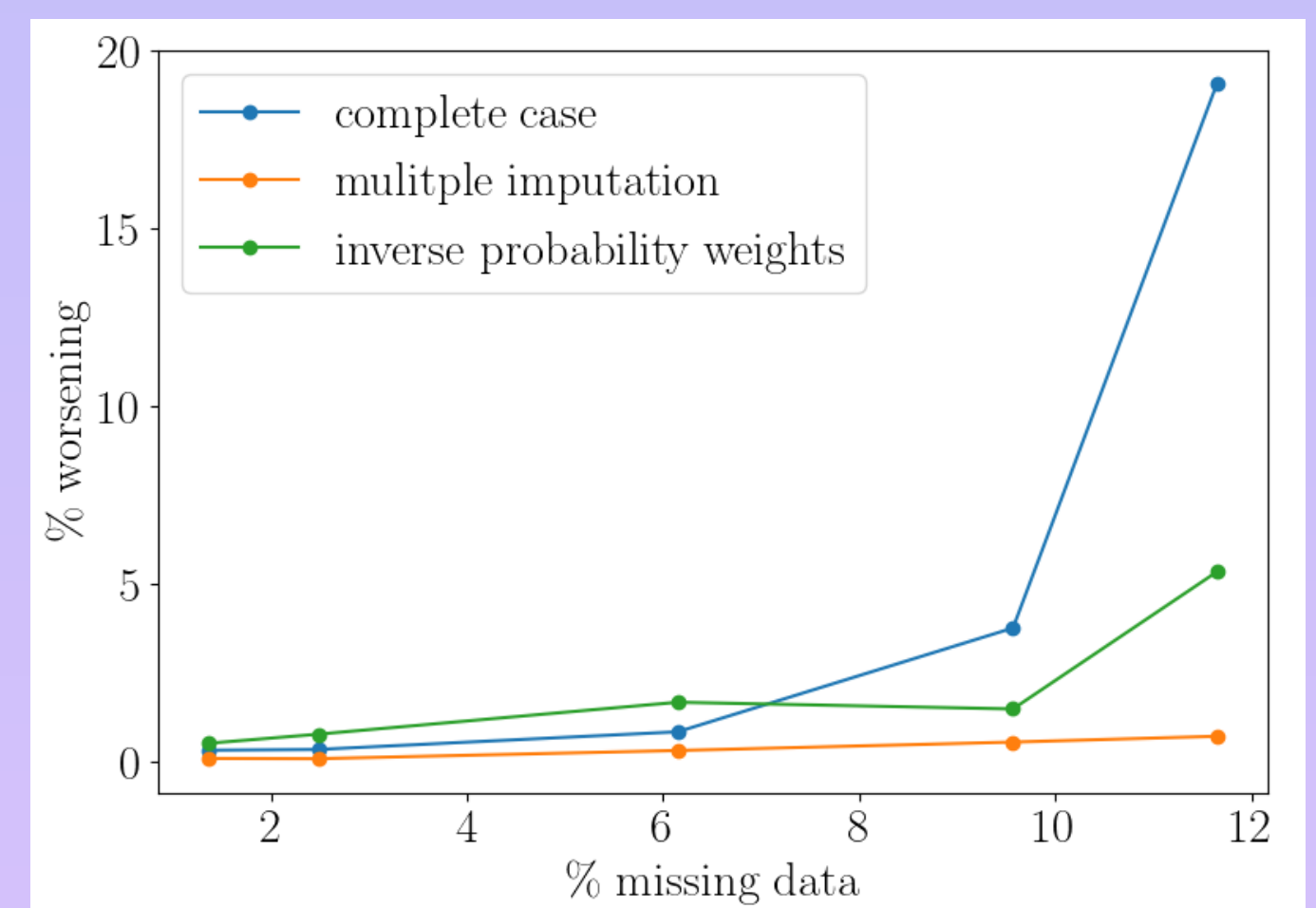
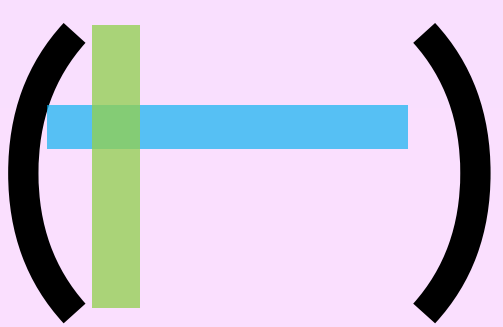


Fig. 3: Missing data in the training set. Multiple imputation shows the greatest skill in mitigating the effect of missing data

Missing Data Techniques

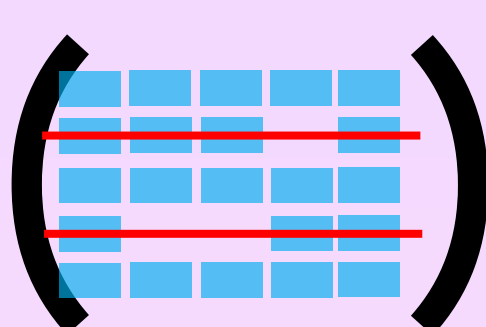
Data layout



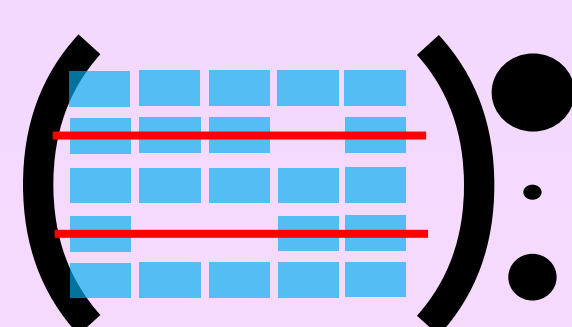
Single set of inputs
(a value for each
variable for one
time instance)

All instances of a
single variable

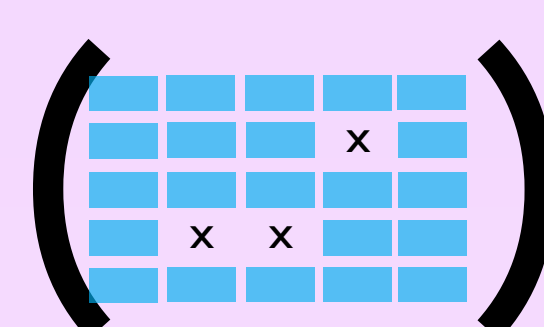
Missing data in the training set



Complete case:
delete rows with
missing data

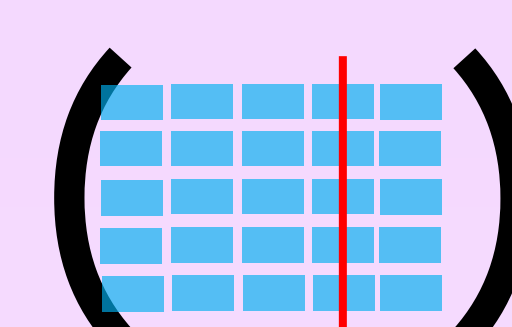


Inverse probability
weights: complete rows
weighted by
 $1/(\text{probability complete})$

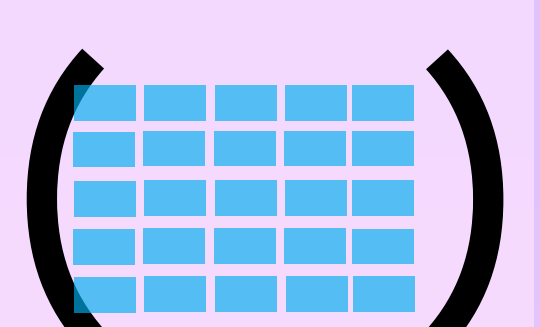


Multiple imputation:
use the remaining data
in the same row to fit
a model that fills the
gaps

Missing forecast inputs



're-train': train a new
model without
variables that are
missing as forecast
inputs



'impute': use the
remaining forecast
inputs to generate a
value for the missing
one(s)

References

- [1] J. Wang, Y. Song, F. Liu and R. Hou, 'Analysis and application of forecasting models in wind power integration', *Renewable and Sustainable Energy Reviews*, vol. 60, 2016.
- [2] G. Giebel, R. Brownsword, and G. Kariniotakis, 'The state of the art in short-term prediction of wind power: a literature overview', Tech. Rep. 2nd Edition, 2011.

