A Wavelet Based Approach for Solar Irradiance Forecasting

Tom Blau
Irena Koprinska
School of Information Technologies
Faculty of Engineering & Information Technologies

Problem

Solar power is useful but unpredictable. A power provider needs to know the output of a PV plant in advance so that they can plan accordingly and keep customers from experiencing power outages.

Contribution

We improve on standard time-series forecasting techniques by predicting individually on the components of a wavelet-decomposed signal. We further improve accuracy with a staged model that uses the outputs of certain predictors as inputs for others.

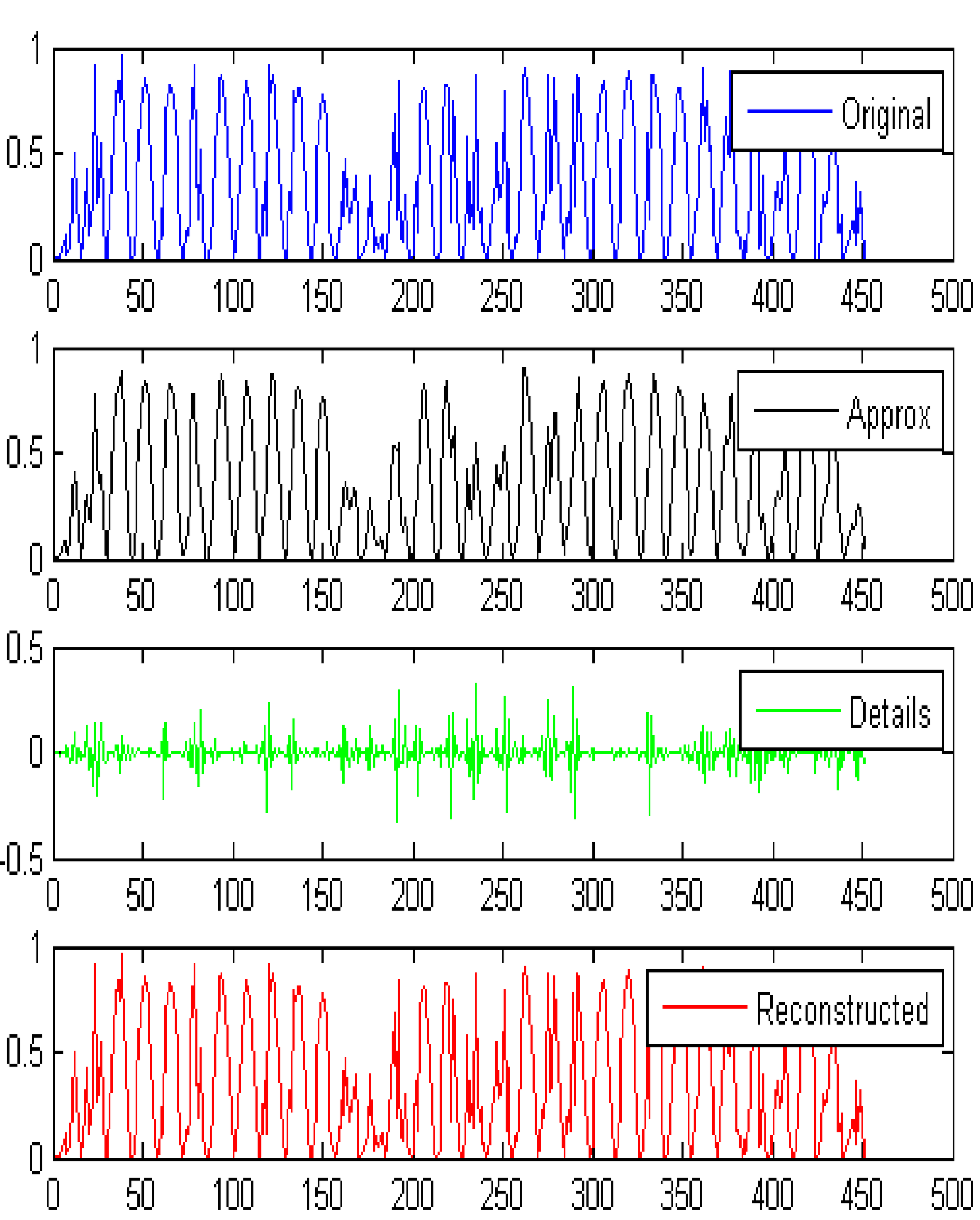
Background

Time series forecasting with machine learning

Predicting the next entry of a time series is a well studied problem. Machine learning approaches such as linear regression, support vector regression and neural networks have been effective in tackling it.
However, classical ML techniques still have difficulty when the input signal is noisy.

Stationary Wavelet Transform

Wavelet functions [1] allow us to decompose a time series into two components- a high-frequency component describing local changes (details) and a low-frequency component approximating the original signal (approximation)



Dataset

The data for our experiments comes from the University of Queensland's St. Lucia campus. It includes hourly measurements of solar irradiance for the year 2013, as well as various atmospheric measurements that assist in prediction.

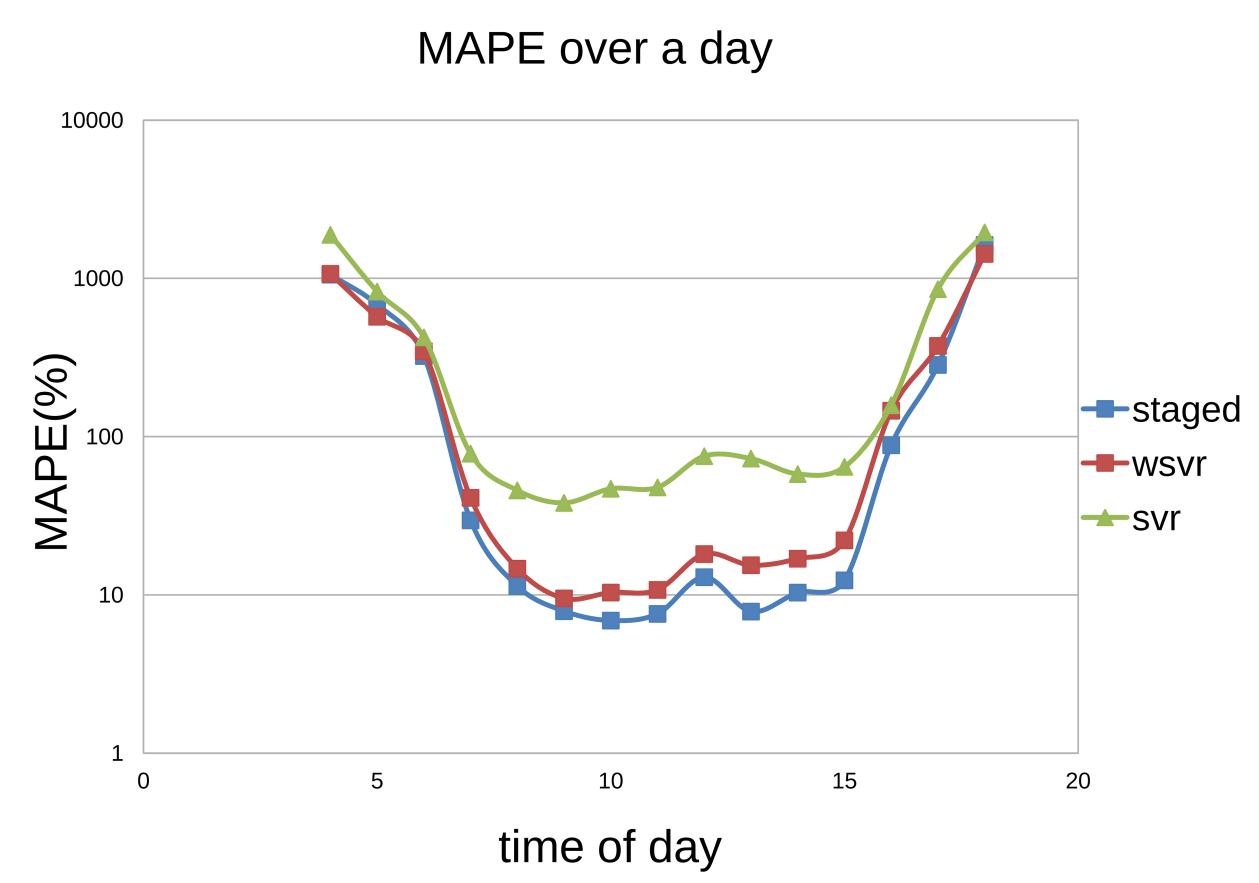
Approach

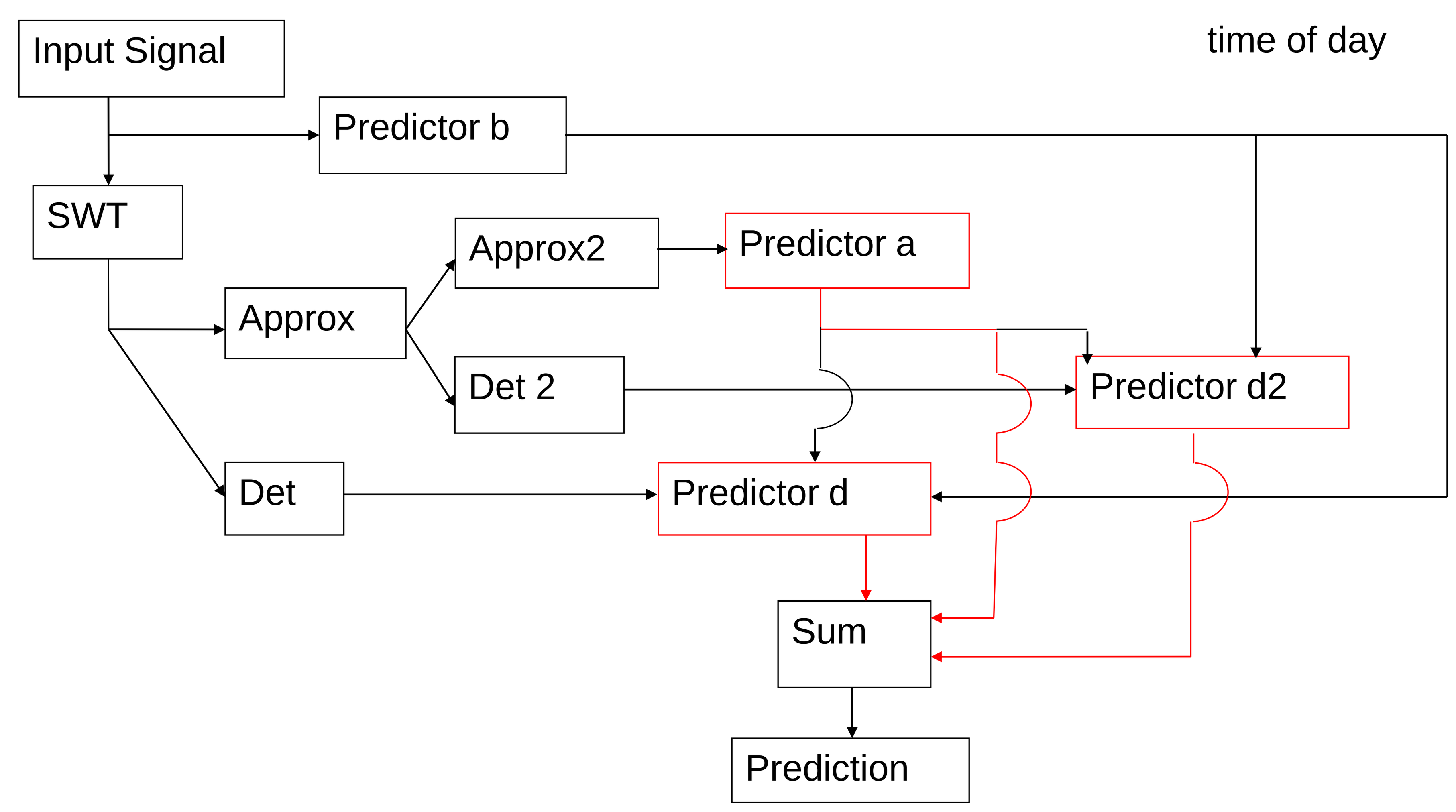
Waveleted component prediction

We decompose the input signal using the Daubechies wavelet, then decompose the approximation signal, to get 3 signals in total. We then train and predict on the individual components, and sum the results to get a prediction of the original signal.

Staged model

To improve prediction on the noisy details signals, we give their predictors additional input- predictions on the approximation signal and the original signal.





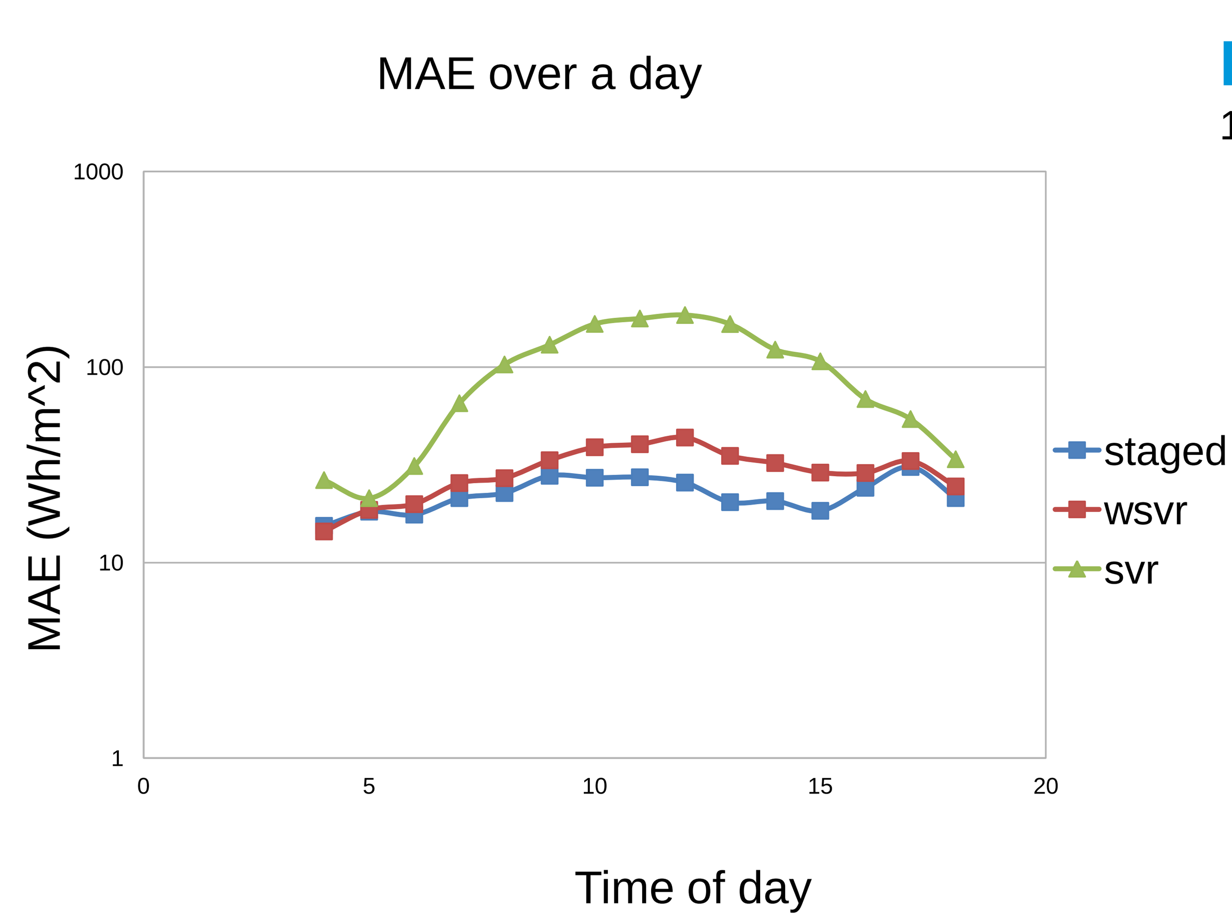
Evaluation

We assess the accuracy of prediction using mean absolute error (MAE) and mean absolute percentage error (MAPE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|p_i - a_i\right|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - a_i}{a_i}\right|$$

Where p is the prediction, a is the actual value, and n is the number of instances.
The charts show a comparison of three models:
• A support vector regressor, using the unprocessed irradiance signal, serves as our baseline.
• A wavelet support vector regressor, predicting each wavelet decomposed component and summing the predictions.
• A staged model, as represented in the above figure.



Conclusions

• Wavelet decomposition yields a big improvement, and the staged model takes this even further. T-tests confirm this improvement is statistically significant ($p < 0.05$).
• Even though the details signal looks like noise, we can actually predict it surprisingly well.
• The wavelet-based and staged models improve on the baseline significantly in midday hours, but less so at morning and evening.

Future Work

• Finding ways to improve prediction on the non-waveleted signal can improve the capability of the staged model.
• Different configurations of the staging, or combining different techniques for the different component predictors (e.g. svm for one, nn for another).
• Training separate predictors for different times of day or year, using fuzzy logic, to smooth the error curve.

References

1. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, 2nd ed. Wellesley-Cambridge Press, 1997