Total consumption prediction of Renewable Energy through Random Forest Regression

Alexandre Cruz Sanchetta, Pablo de Freitas Hernandez, Laís Castro de Carvalho, Emilson Pereira Leite. UNICAMP

1. Summary

Policymakers worldwide have sought to accomplish the climate goals of reducing carbon emission set forth in the Paris Agreement and the sustainable development goals of the 2030 agenda. The present work analyzes the renewable sources of energy by forecasting temporal series using Random Forest Regression. Random Forest application is justified by its prediction power and adaptability.

The data preparation for this machine learning application is achieved by considering the time series as a column of a data matrix and adding new columns with a shifted version of the original time series, which creates a new framing of the time series from the desired length of the input and output sequences. This procedure allows to explore different framings of a time series problem with machine learning algorithms to find the best performing models. A time series can't be evaluated using the standard k-fold method. Alternatively, using the walk-forward validation technique.

We then separated the last 12 years for the testing set, with the exception of the last year while the rest of the dataset was used for training. The last year measured was used for an out-of-sample forecast (OSF) prediction, that is, to make an extrapolation and predict data that is not found in the training/test set. Using this database separation and a Random Forest Regressor, we predicted the total final consumption of renewable resources in the United States of America measured in Kilotons of Oil Equivalent (KTOE).

The quality of the prediction is measured by the mean absolute error (MAE) and by Pearson correlation coefficient (PCORR) in the training stage and by comparing values in the test stage we compare only a single predicted value and a single observed value. The optimal modeling was achieved by shifting the original prediction based on the MAE (Prediction + MAE). We reapplied the MAE over the optimal model generating the final mean absolute error (FMAE).

2. Results and Conclusions

In the training phase, after the MAE correction, the internal validation of the last 12 years recovered the following KTOE results (Observed, Predicted): 1) (61554.2, 63142.7); 2) (63894.1, 66435.0); 3) (66660.9, 65803.6); 4) (66327.7, 65214.1); 5) (70155.2, 70228.7); 6) (73488.6, 74749.7); 7) (76326.9, 76706.2); 8) (83421.2, 78623.7); 9) (84686.8, 80960.1); 10) (81546.9, 81943.8); 11) (82062.8, 83676.4); 12) (82231.5, 84872.2).

The results above show the best scenario with a FMAE of 1749.2 KTOE and represents only about 2% of error. The data average is 74363.1 KTOE. The PCORR between observed and predicted data is 0.962.

After obtaining satisfactory validation results in the training phase, we used the optimal parameters to make an OSF prediction. The predicted value for the last year was 91555.8 while the observed value was 89897.5, resulting in an error of less than 2%, which indicates that this methodology is robust and can apply to problems of the same nature.