Geophysical Research Abstracts
Vol. 19, EGU2017-6693, 2017
EGU General Assembly 2017

Random Forests (RFs) for Estimation, Uncertainty Prediction and Interpretation of Monthly Solar Potential

Dan Assouline (1), Nahid Mohajeri (2), and Jean-Louis Scartezzini (3)

(1) EPFL, LESO-PB, ENAC, Switzerland (dan.assouline@epfl.ch), (2) EPFL, LESO-PB, ENAC, Switzerland (nahid.mohajeri@epfl.ch), (3) EPFL, LESO-PB, ENAC, Switzerland (jean-louis.scartezzini@epfl.ch)

Solar energy is clean, widely available, and arguably the most promising renewable energy resource. Taking full advantage of solar power, however, requires a deep understanding of its patterns and dependencies in space and time. The recent advances in Machine Learning brought powerful algorithms to estimate the spatio-temporal variations of solar irradiance (the power per unit area received from the Sun, W/m^2), using local weather and terrain information. Such algorithms include Deep Learning (e.g. Artificial Neural Networks), or kernel methods (e.g. Support Vector Machines). However, most of these methods have some disadvantages, as they: (i) are complex to tune, (ii) are mainly used as a black box and offering no interpretation on the variables contributions, (iii) often do not provide uncertainty predictions (Assouline et al., 2016). To provide a reasonable solar mapping with good accuracy, these gaps would ideally need to be filled. We present here simple steps using one ensemble learning algorithm namely, Random Forests (Breiman, 2001) to (i) estimate monthly solar potential with good accuracy, (ii) provide information on the contribution of each feature in the estimation, and (iii) offer prediction intervals for each point estimate.

We have selected Switzerland as an example. Using a Digital Elevation Model (DEM) along with monthly solar irradiance time series and weather data, we build monthly solar maps for Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (GHI), and Extraterrestrial Irradiance (EI). The weather data include monthly values for temperature, precipitation, sunshine duration, and cloud cover. In order to explain the impact of each feature on the solar irradiance of each point estimate, we extend the contribution method (Kuz'min et al., 2011) to a regression setting. Contribution maps for all features can then be computed for each solar map. This provides precious information on the spatial variation of the features impact all across Switzerland maps. Finally, as RFs are based on bootstrap samples of the training data, they can produce prediction intervals by looking at the trees estimates distribution, instead of taking the mean estimate. To do so, a simple idea is to grow all trees fully so that each leaf has exactly one value, that is, a training sample value. Then, for each point estimate, we compute percentiles of the trees estimates data to build a prediction interval. Two issues arise from this process: (i) growing the trees fully is not always possible, and (ii) there is a risk of over-fitting. We show how to solve them. These steps can be used for any type of environmental mapping so as to extract useful information on uncertainty and feature impact interpretation.

References

-Assouline, D., Mohajeri, N., & Scartezzini, J. L. (2017). Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Solar Energy*, 141, 278-296.

-Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

-Kuz'min, V. E., Polishchuk, P. G., Artemenko, A. G., & Andronati, S. A. (2011). Interpretation of QSAR models based on random forest methods. *Molecular Informatics*, 30(6-7), 593-603.