Implementing k-Nearest Neighborhood as a forecast method for Intra hour resolution with no exogenous outputs

Giuliano L. Martins[1], Rafael Antunes Campos[1], Marília Braga[1] and Ricardo Rüther[1]

[1] Universidade Federal de Santa Catarina, Florianópolis (Brazil)

Abstract

Estimating solar irradiance is a big challenge for solar power plant and grid operators and managing power inputs to the grid according to reality is key if photovoltaics (PV) are to play a bigger role in the energy mix. Therefore, forecasting techniques are an important asset, which has several methods with different singularities. In this paper, we discuss the k-Nearest Neighborhood method (kNN) in a real case scenario for a forecast horizon of 1 minute, using irradiance data from Florianópolis-SC for the year of 2018. The kNN presents itself as a simple and robust method, having accuracy results of: R^2=95% and NRMSE = 9.88% for the initial model and R^2 = 95% and NRMSE = 13.52% in real situation. Moreover, the forecasting skill was calculated for both cases, obtaining a positive result for the trained model, which was, however, not as good for the real case scenario. In addition, real case forecasting was simulated and the punctual error for each point was calculated and analyzed, thus providing subside to discuss which factors were related to the accuracy loss, concluding, afterwards that the biggest error gather was irradiance ramps, caused by cloud covers and cloud edges.

Keywords: Forecast, kNN, Kt, GHI, accuracy, Intra-Hour

1. Introduction

The implementation of photovoltaic energy in large scale generates several difficulties due to the solar resource variability. The nonlinear pattern of solar irradiance creates a challenge for the grid operator, which must predict theses variations in order to provide the right amount of energy to the grid. Therefore, methods for overcoming this issue are more and more required in order to increase renewable energy utilization.

In this paper, the k-Nearest Neighborhood (kNN) method (Chu & Coimbra, 2017) is utilized as a forecasting method due to its simple and intuitive implementation. In addition it has promising results for small and medium gaps (Pedro & Coimbra, 2012). The proposed method was used to predict global horizontal irradiance (GHI) at a site located in Florianópolis - SC (-27.43, -48.44) in a one-minute forecast timescale.

The kNN is an autoregressive method that predicts a new value based on the k nearest neighbors, where the distance among points is calculated from metrics based on available data. It is widely used to perform predictions in many solar research fields. Pedro and Coimbra (2015) predict GHI and direct normal irradiance (DNI) in intra-hour resolutions using local measured data and sky images. They concluded that the kNN has a lack of accuracy for large irradiance ramps, and that including sky images increases the accuracy by less than 5%.

Madeti and Singh (2018) used the kNN method to identify many types of faults in photovoltaic systems, such as open circuit fault and shading faults. Many other forecasting methods are available to predict solar irradiance and photovoltaic generation: autoregressive methods as ARIMA and ARMAX (Li, Su, & Shu, 2014; Scolari, Sossan, & Paolone, 2016), machine learning based as artificial neural network (ANN) (Kamadinata, Ken, & Suwa, 2019; Reno & Hansen, 2016). Yang et al. (2018) have presented a mining review of the solar forecasting literature, analyzing new methods and frequently-used terms.

2. Data and Methodology

2.1 Dataset

The irradiance data is retrieved from Fotovoltaica/UFSC solar radiation measurement station (Fig, 1), with a time resolution of one minute and is measured by a Kipp & Zonen SMP11 pyranometer. The total data period is the whole year of 2018 and the data was pre-processed using BSRN quality standards (Long & Dutton, 2010). The solar zenith angle was calculated through NREL tool (NREL, 2018) and it was used to filter the data for only day values (zenith

doi:10.18086/swc.2019.43.06 Available at http://proceedings.ises.org

< 80º).

Figure.1: Fotovoltaica – UFSC solar radiation measurement station at Florianópolis (Brazil).

3.2 The kNN method

The kNN autoregressive method predicts a new point based on the average of the k nearest neighbors points, this is, points with the minimum distances from the predicted point. The distance among points is calculated based on pre-defined parameters that are called features. Once a distance for each feature is done, the resultant distance is calculated using the Euclidian distance (d).

$$d = \sqrt{\sum_{f=1}^{n} d_f^2} \quad \text{(eq. 1)}$$

Where d_f is the distance for the feature *f*.

The features chosen were: timestamp, zenith angle, the last measured irradiance, the last kt value, moving average from the last four kts, and the variance from the last four kts. The kt used is the measured GHI value divided by the horizontal extraterrestrial irradiance, as showed in Equation 2.

$$kt = \frac{GHI}{Ext_{Hor}} \quad \text{(eq. 2)}$$

All features were normalized by their minimum and maximum values, then, the Euclidian distance is calculated based on the normalized features (Pedro and Coimbra, 2015b).

$$F_t^{norm} = \frac{F_t - \min(F)}{\max(F) - \min(F)} \quad \text{(eq. 3)}$$

Where F_t is the featured at instant *t* and F represents the entire set of F_t past values.

The whole dataset was split into three different dataframes: training dataframe, test dataframe, real case dataframe. The real case dataframe is a dataset that were not used in the context of training and testing stage of the kNN algorithm, thus, representing a forecasting real case.

3.3 The Persistence Method

The Persistence method is a widely used model to predict solar irradiance and is also used as a reference method to asses the forecast skill of other methods (Chu et al., 2015; Kaur et al., 2016; Urraca et al., 2016). It is based on the assumption that the atmospheric condition is going to be the same as the very previous condition. In irradiance terms, this means that the next (forecasted) kt is equals to the previous (measured) kt. Therefore, it is represented by Equation 4.

$$kt_{t+1} = kt_t \qquad \text{(eq. 4)}$$

Consequently, the predicted irradiance value is given by the product of this new kt and the horizontal extraterrestrial irradiation for that instant.

3.3 Accuracy assessment

To assess the accuracy of the kNN method, three main metrics were used: the R^2 of the correlation between measured and forecasted values; the Normalized Root Mean Square Error (NRMSE) and the Forecast Skill (FS) (Inman et al., 2013).

The NRMSE is the Root Mean Square Error (RMSE) normalized by the average irradiance value, and it can be calculated by Equation 5. The normalization was done in order to avoid misguidance when evaluating the error for bigger and smaller irradiance values

$$NRMSE = \frac{\sqrt{\sum_{t=1}^{T} \frac{(GHI_t - Predicted_t)^2}{T}}}{GHI_{Avg}} \qquad \text{(eq. 5)}$$

The Forecast Skill (FS) is a metric used to assess the accuracy of the method comparing it with the Persistence method and is given by Equation 6.

$$FS = 1 - \frac{NRMSE_{Predicted}}{NRMSE_{Persistenc}} \qquad \text{(eq. 6)}$$

Also, to have a better understanding of the accuracy in different conditions, a simple relative error metric calculated by Equation 7.

$$Error = \frac{GHI_t - Predicted_t}{GHI_t} \qquad \text{(eq. 7)}$$

3. Results and Discussion

Primarily, the optimal k number of neighbors was chosen through an iterative analysis where the k value was varied from 1 to 100 and it was observed that the R^2 is around 95% and 94% for the worst-case scenarios. Thus, in a trade-off between accuracy and computer efforts, the k value selected is 5.

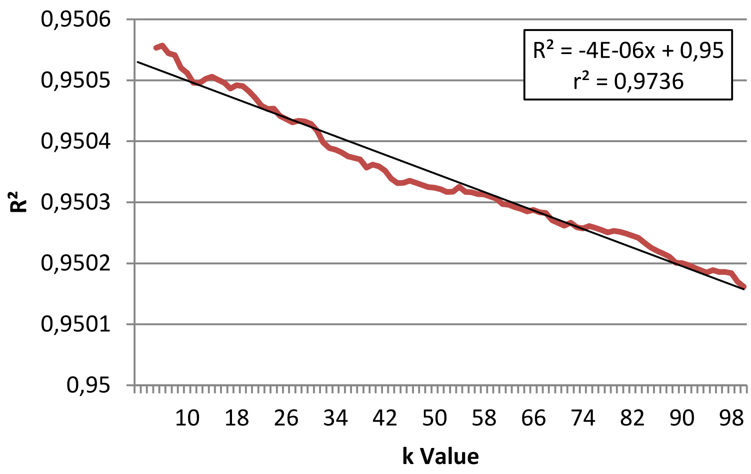

Figure.2: R² as a function of k.

Then, with the k value determined, a more complete analysis was made to evaluate the kNN accuracy in different sky conditions. The NRMSE obtained for k = 5 was 22%. In addition, to analyze the values separately we implemented a simple error metric presented in Figure 2. It is noticeable that the forecast model provides better accuracy for values between 400 and 800 W/m². For smaller irradiance values (< 200 W/m²) we have a very dispersed error, which could be explained since the kNN would perform very well for beginning and end of the day, due to its pattern behavior. On the other hand, these values are also a consequence of ramp caused by clouds, which Pedro and Coimbra (2015) already stated to be a difficult for kNN's method. Moreover, cloud edges can result in unpredictable extreme irradiance events (Rüther et al., 2017), resulting in values with very high kt's and a low accuracy for the forecast method.

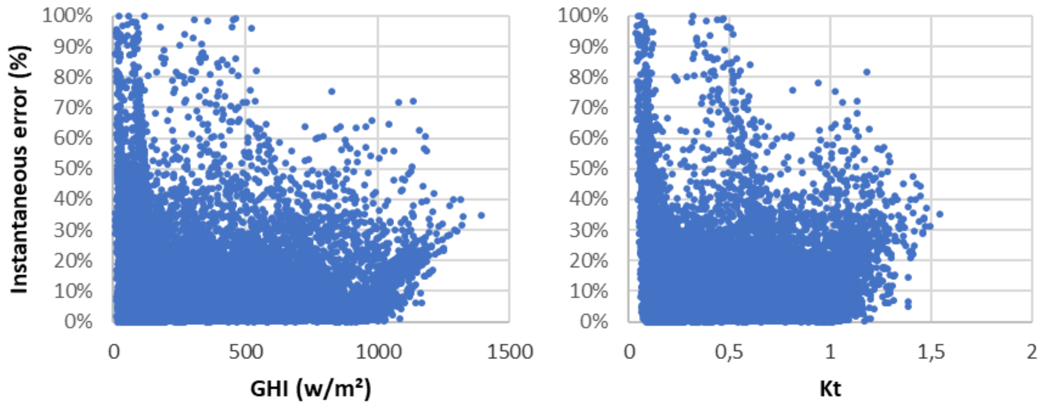

Figure.3: Correlation between GHI and Kt values and the forecast instantaneous error.

Furthermore, the persistence model was calculated, providing an R^2 =95% and NRMSE=13%. Thus, it was possible to calculate the FS. Moreover, the value obtained when comparing to the model was FS = 0.252, demonstrating a small improved if compared to the simplest possible model. However, when analyzing the real case scenario, the value obtained was FS = -0.024, a result which demonstrates no improvement in relation to the persistence model. This situation can be strongly related to the seasonal effect related to the data chosen to be used to test the model, which was the fourth quarter of the year, thus summer season in Brazil. Therefore, kNN model would be utilized in its worst conditions, with several irradiance ramps, alongside the persistence model was in a positive error condition, with more clear days than usual. According to this situation, kNN still worked as expected.

In order to better evaluate and understand the model, two real case scenarios were plot, considering two specific days, which were retrieved from the initial dataset. Thus, the first day "2018-01-04" (Figure.4 top) which represents a clear day was forecasted using the trained model. As expected, the model produced a strong representation of reality, however outputting small variations throughout the clear day, situations that can be smoothed with an optimization in the model. Furthermore, when analyzing a cloudy day "2018-08-09" (Figure.4 bottom), it is possible to observe a considerably reduced accuracy, especially related to the difficulty of kNN when forecasting irradiance ramps. Moreover, it also presents some small deviation when estimating data. Therefore, it is noticeable that the kNN model presented represents a clear day forecasting model, however generating weaker skills when related to more overcast sky days.

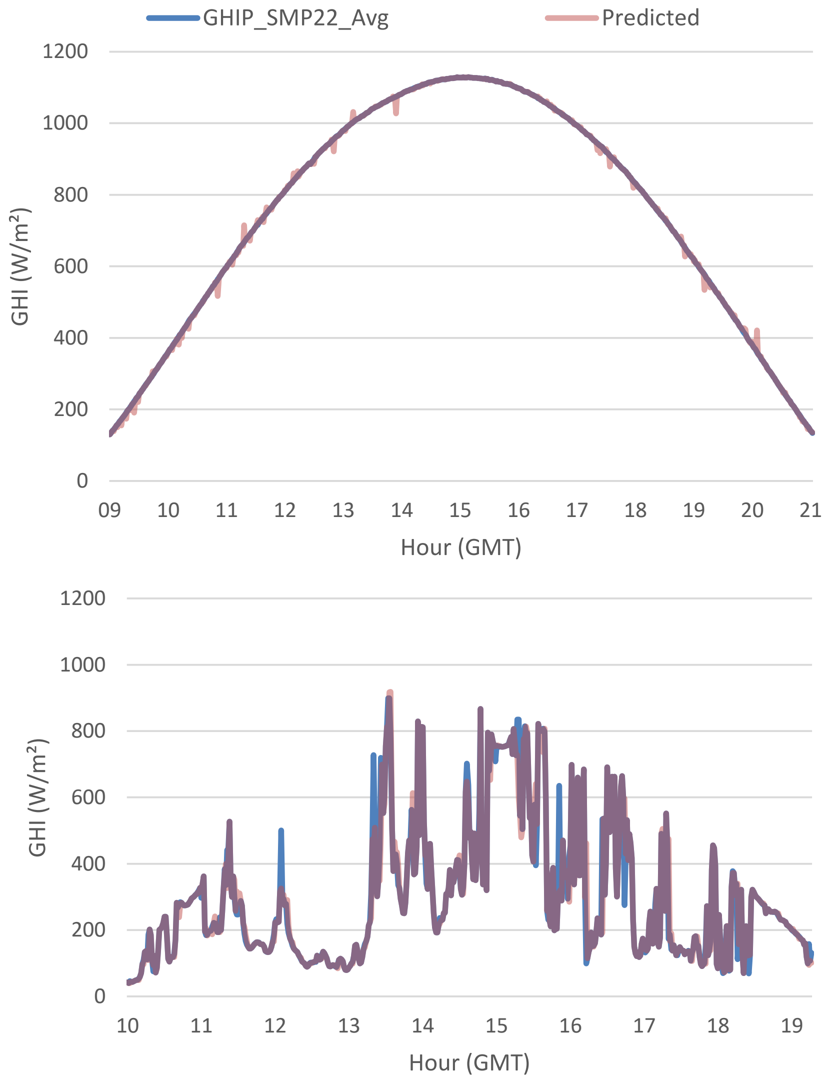

Figure.4: Comparison between forecasting in a clear sky day (top) and in a cloudy day (bottom).

4. Conclusion

Given the arguments above, we can conclude that kNN is a very consistent forecast method. Its implementation simplicity and intuitive aspects are positive factors highlighted in this method. Furthermore, presenting a R^2 = 95% and NRMSE = 22.13% demonstrated kNN as a robust method; however, this forecast tool presents difficulties regarding irradiance ramps, thus, reducing its accuracy. Moreover, future work will be carried out with sky camera analyses and other exogenous data, in order to increase kNN forecast accuracy. Also, work in progress will present results on a seasonal evaluation of kNN model to enrich the results obtained so far, complementing the work on optimization for the model presented in this paper.

5. References

Chu, Y., & Coimbra, C. F. M. (2017). Short-term probabilistic forecasts for Direct Normal Irradiance. *Renewable*

Energy, *101*, 526–536. https://doi.org/10.1016/j.renene.2016.09.012

Chu, Y., Pedro, H. T. C., Li, M., & Coimbra, C. F. M. (2015). Real-time forecasting of solar irradiance ramps with smart image processing. *Solar Energy*, *114*. https://doi.org/10.1016/j.solener.2015.01.024

Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, *39*(6), 535–576. https://doi.org/10.1016/j.pecs.2013.06.002

Kamadinata, J. O., Ken, T. L., & Suwa, T. (2019). Sky image-based solar irradiance prediction methodologies using artificial neural networks. *Renewable Energy*, *134*, 837–845. https://doi.org/10.1016/j.renene.2018.11.056

Kaur, A., Nonnenmacher, L., Pedro, H. T. C., & Coimbra, C. F. M. (2016). Benefits of solar forecasting for energy imbalance markets. *Renewable Energy*, *86*, 819–830. https://doi.org/10.1016/j.renene.2015.09.011

Li, Y., Su, Y., & Shu, L. (2014). An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, *66*, 78–89. https://doi.org/10.1016/j.renene.2013.11.067

Long, C. N., & Dutton, E. G. (2010). BSRN Global Network recommended QC tests, V2. *Journal of Climate*, *25*(24), 8542–8567. https://doi.org/10.1175/JCLI-D-11-00618.1

Madeti, S. R., & Singh, S. N. (2018). Modeling of PV system based on experimental data for fault detection using kNN method. *Solar Energy*, *173*(March), 139–151. https://doi.org/10.1016/j.solener.2018.07.038

NREL. (2018). MIDC SOLPOS Calculator. Retrieved from https://midcdmz.nrel.gov/solpos/solpos.html

Pedro, H. T. C., & Coimbra, C. F. M. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*(7), 2017–2028. https://doi.org/10.1016/j.solener.2012.04.004

Pedro, H. T. C., & Coimbra, C. F. M. (2015a). Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renewable Energy*, *80*, 770–782. https://doi.org/10.1016/j.renene.2015.02.061

Pedro, H. T. C., & Coimbra, C. F. M. (2015b). Short-term irradiance forecastability for various solar micro-climates. *Solar Energy*, *122*, 587–602. https://doi.org/10.1016/j.solener.2015.09.031

Reno, M. J., & Hansen, C. W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. *Renewable Energy*, *90*. https://doi.org/10.1016/j.renene.2015.12.031

Rüther, R., Nascimento, L. R. do, & Campos, R. A. (2017). Performance assessment issues in utility-scale photovoltaics in warm and sunny climates. *Renewable Energy and Environmental Sustainability*, *2*, 35. https://doi.org/10.1051/rees/2017035

Scolari, E., Sossan, F., & Paolone, M. (2016). Irradiance prediction intervals for PV stochastic generation in microgrid applications. *Solar Energy*, *139*, 116–129. https://doi.org/10.1016/j.solener.2016.09.030

Urraca, R., Antonanzas, J., Alia-Martinez, M., Martinez-De-Pison, F. J., & Antonanzas-Torres, F. (2016). Smart baseline models for solar irradiation forecasting. *Energy Conversion and Management*, *108*, 539–548. https://doi.org/10.1016/j.enconman.2015.11.033

Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. M. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, *168*(February), 60–101. https://doi.org/10.1016/j.solener.2017.11.023