AI/ML Feature Selection and Modeling of Spectral Correction Factors from FARMS-NIT

Eugeniah Arthur, Jiadi Zhang, Prabal M. Dhaubhadel, Alan J. Curran, Yuepeng Deng
First Solar, Inc. Perrysburg, OH, 43551

Introduction

Spectral correction is a metric that quantifies the deviation of the performance of a given PV module based on incident spectral deviations from the standard G-173 spectra. Current spectral correction predictive models rely on atmospheric features such as absolute airmass, precipitable water, clearness index, and clear sky index, with airmass being the most used feature.

This study aims to find additional features useful in simple predictions of spectral correction by incorporating atmospheric data like cloud type, pollutant, and turbidity data, alongside traditional atmospheric variables, into AI/ML driven variable selection for spectral correction modeling. FARMS-NIT derived spectra and metrology data sourced from the NREL opensource database (NSRDB) are used in this study. Spectral correction for TOPCon and CdTe technologies are calculated from FARMS-NIT spectra [1] to be used as modeling targets.

Objectives

- ❑ Spectral correction values (M) are calculated from FARMS-NIT Spectra
- ❑ Utilize AI/ML to discover key predictors for spectral correction (M) in CdTe and TOPCon Silicon
- ❑ Determine if spectral correction feature importance is different for TOPCon and CdTe
- ❑ Compare AI/ML model performance against existing spectral correction model functional forms, including 1-Parameter, 2-Parameter Spectral Model and Spectral 3.0

Methodology

- ➢ 104 locations across the US selected for fitting/validation
  - 2 years (2019, 2020) at each location, 1-axis orientation
- ➢ Weather features pulled from NSRDB:

  CSI, GHI, Precipitable Water, Pressure, Relative Humidity, Temperature, Solar Zenith Angle, Panel Tilt , Panel Azimuth Angle, Absolute Airmass (AM_a), Aerosol Optical Depth (AOD), Asymmetry , Dew Point, Ozone, SSA, Wind Direction, Wind Speed

Popular Spectral Correction Models

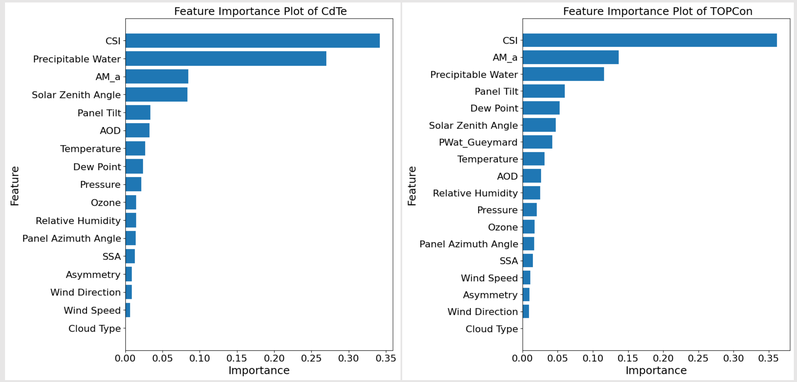
- ➢ 1-Parameter Model [2, 3]:

  $$M = A_o + A_1.AM_a + A_2.(AM_a)^2 + A_3.(AM_a)^3 + A_4.(AM_a)^4$$
- ➢ 2-Parameter Spectral Model [4]:

  $$M = a + b \cdot AM_a + c \cdot p_{wat} + d \cdot \sqrt{AM_a} + e \cdot \sqrt{p_{wat}} + f \cdot \frac{AM_a}{\sqrt{p_{wat}}}$$
- ➢ Spectral 3.0 Model [5]:

  $$M = a + b \cdot \mathrm{CSI} + c \cdot AM_a + d \cdot \sqrt{AM_a} + e \cdot p_{wat} + f \cdot \sqrt{p_{wat}}$$
  - Continuous improvement of this model has added $\sqrt{AM_a}$ and $\sqrt{p_{wat}}$ terms to this model

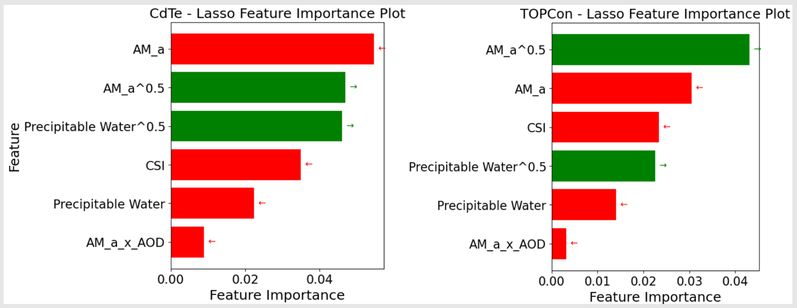
All models share the same testing and training data for an equal comparison of performance. Default coefficients of the existing models are not used.



Statistical Modeling

- ❑ XGBoost Model:
  This is a fast, scalable machine learning algorithm based on gradient boosting. It builds decision trees sequentially, with each new tree correcting the errors of the previous ones. Known for high accuracy, it handles large datasets, missing values, and includes regularization to prevent overfitting—making it a top choice in both industry and competitions.
- ❑ Feature Selection using Lasso Regression:
  Lasso Regression utilizes L1 regularization to penalize feature coefficients and shrink coefficients to 0. In this section, we use Lasso Regression to identify the next potent feature that could contribute to updated Spectral 3.0's existing model structure.
  - ➢ Steps For Lasso:
    - Compute all square root, squared and 2-degree interactions for existing features
    - Remove penalty term for existing Spectral 3.0 features
    - Fit Lasso Regression
    - Use the selected features to refit a linear regression
  - ➢ Result: The feature selected by Lasso Regression is the interaction term $AM_a \cdot AOD$.
  - ➢ Formula from Lasso Regression :

$$M = a + b \cdot \mathrm{CSI} + c \cdot AM_a + d \cdot \sqrt{AM_a} + e \cdot p_{wat} + f \cdot \sqrt{p_{wat}} + g \cdot AM_a \cdot AOD$$



Results

Metrics-CdTe	XGBoost	Lasso-Improved Spectral 3.0	Spectral 3.0	2-Parameter Spectral Model	1-Parameter Model
R2	0.79	0.75	0.68	0.41	0.12
MSE	0.00045	0.00048	0.00061	0.0011	0.0017
RMSE	0.023	0.022	0.025	0.034	0.041
MAE	0.0049	0.0072	0.0087	0.018	0.027
MBE	0.00057	0.00032	0.0012	0.0019	0.0031

Metrics-TOPCon	XGBoost	Lasso-Improved Spectral 3.0	Spectral 3.0	2-Parameter Spectral Model	1-Parameter Model
R2	0.73	0.66	0.62	0.24	0.085
MSE	0.00016	0.00020	0.00023	0.00045	0.00055
RMSE	0.0127	0.014	0.015	0.021	0.023
MAE	0.0041	0.0053	0.0053	0.011	0.014
MBE	0.00056	0.00014	0.00047	0.00087	0.0014

R2: R-Squared, MSE: Mean Square Error, RMSE: Root Mean Square Error, MAE: Mean Absolute Error, MBE: Mean Bias Error

Beyond the known important variables (CSI, precipitable water, and absolute airmass), XGBoost identifies panel tilt and solar zenith angles as predictive variables for spectral correction. Additionally, Lasso Regression demonstrates that the interaction between Air Mass and Aerosol Optical Depth is a predictive factor for spectral correction.
From the metric tables, the XGBoost model shows to be highly predictive for both technologies. Since Solar Zenith Angle, Panel tilt are among the top features for both technologies, they have the potential to improve future model performance.
Also, adding the interaction variable Air Mass * AOD (selected by Lasso Regression) successfully improved the performance of the updated spectral 3.0 in all metrics, we believe Air Mass * AOD also has the potential to improve future spectral modeling.

Conclusion

- ✓ The AI/ML models confirm current features used in building spectral models such as CSI, Precipitable Water and absolute airmass index are a step in the right direction.
- ✓ Our AI/ML models indicate that some other weather factors might be worth studying in the future.

References

1. Xie, Y., Sengupta, M., 2018. A Fast All-sky Radiation Model for Solar applications with Narrowband Irradiances on Tilted surfaces (FARMS-NIT): Part I. The clear-sky model. Sol. Energy 174, 691-702.
2. D. King, W. Boyson, and J. Kratochvill, Photovoltaic Array Performance Model, SAND2004-3535. Albuquerque, New Mexico: Sandia National Laboratories, 2004.
3. Anderson, K., Hansen, C., Holmgren, W., Jensen, A., Mikofski, M., and Driesse, A. "pvlib python: 2023 project update." Journal of Open Source Software, 8(92), 5994, (2023). DOI: 10.21105/joss.05994.
4. M. Lee and A.F. Panchula, "Spectral Correction for PV Performance Based on Air Mass and Precipitable Water," 43rd IEEE Photovoltaic Specialists Conference (2016).
5. A. J. Curran, B. Lin and Y. Deng, "Improvements on Spectral Correction Predictive Modeling for CdTe Modules," 2023 IEEE 50th Photovoltaic Specialists Conference (PVSC), San Juan, PR, USA, 2023, pp. 1-6, doi: 10.1109/PVSC48320.2023.10359979.

First Solar Proprietary & Confidential - General