Modeling the Variability of the Sun's Total Solar Irradiance through Supervised Machine Learning Techniques

Ashley Lieber [1,2], Laura Sandoval [1], Joshua P. Elliott [1], Dr. Wendy Carande [1]
[1]Laboratory for Atmospheric and Space Physics (LASP), [2]University of Arkansas

Abstract

With the Sun being Earth's main source of energy and heat, it is critical that we understand how and why it varies. The Sun's irradiance changes over its 11-year solar cycle driven by variabilities in the magnetic field. The primary objective of this project was to create a machine learning regression model that predicts the total solar irradiance (TSI) and to analyze the contributions of solar features in driving change of the Sun's TSI. This improves upon the previous model by incorporating machine learning techniques on a wider array of data.

This model uses data of solar features visible in intensitygrams and line-of-sight magnetograms from the Helioseismic and Magnetic Imager (HMI) instrument aboard the Solar Dynamics Observatory (SDO). This data includes HARP data which gives information on active regions (AR) on the Sun's surface and SHARP data which provides space weather parameters. The TSI values were measured by Total Irradiance Monitors (TIM) on two separate satellites. In order to make the data compatible between the two satellites, we utilized TSI composite which minimizes the systemic differences between the datasets and adjusts the data accordingly without introducing more error.

Initially, we utilized the machine learning technique of multiple linear regression to have the model predict an irradiance value for a given day based on given solar features. To ascertain the sunspot area, we performed image processing on the intensitygrams from the SDO HMI. Additionally, seventeen space weather parameters from the SHARP data were added to the model. Due to the number of features, a correlation matrix was created to eliminate features that were highly correlated and determine how features correlated with TSI. Preliminary results show that training the model on these features yields a root mean squared error of 0.052 showing that the model is performing as expected in this early stage. A residual plot was also created to showcase the relation between the actual TSI values and the predicted TSI values.

Objectives

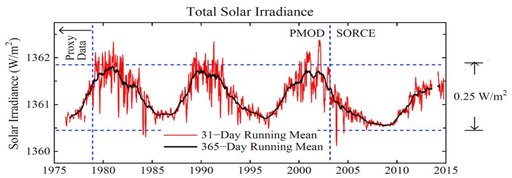


Figure 1: Solar Cycle depicted with Total Solar Irradiance over time. Increase magnetic activity occurs with increased number of sunspots. Credit: Wikipedia Commons

The total solar irradiance of the Sun follows a fairly regular 11-year pattern called the Solar Cycle shown in black in Figure 1. As this pattern progresses there are minute variations along the way often attributed to the complex magnetic field of the Sun. The objectives of the project are as follows:

1. To create a model using machine learning to predict total solar irradiance (TSI).
2. To analyze the contributions of solar features in driving such variabilities. The previous model NRL TSI2 from the Naval Research Laboratory also employed such machine learning techniques but we aim to incorporate more data on differing features in hopes of a more efficient and accurate model.

Motivation

- With the Sun as our main source of energy and heat, understanding how and why it changes is critical to understanding the Sun's impact on Earth. Even small changes on the Sun can cause profound effects on Earth.
- Moreover, determining which feature cause these changes in irradiance can help us better understand the mechanisms of change in the Sun and even potentially direct our research effects.
- A model that can predict irradiance values can help us to fill in gaps in the irradiance record when satellites that measure TSI are offline due to maintenance or space weather events.

Methods

In order to create a machine learning regression model, we had to be selective in which features to input in our model. Pre-processing our data is a major component of our project as outlined below.

Machine Learning

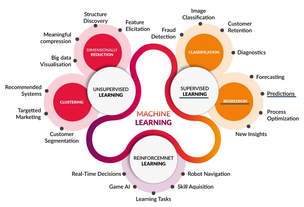


Figure 2: Machine Learning Breakdown
Credit: http://www.cognub.com/index.php/cognitive-platform/

- Why Machine Learning?
  - It's a great technique to tackle large sets of data
  - Allows us to create a model that can predict irradiance by training
- For our project, the branch of machine learning that we utilized is as follow:
  - Supervised Machine Learning
  - Multiple Linear Regression
  - Predictive Model

Data

For our target variable, Total Solar Irradiance (TSI) we gathered data from four satellites that each have an instrument aboard to measure total solar irradiance.

1. Solar Radiation & Climate Experiment (SORCE) 2003-2020
2. Total and Spectral Irradiance Sensor 1 (TSIS 1) aboard the International Space Station (2017-Present)
3. Total Solar Irradiance Calibration Transfer (TCTE) 2013-2019
4. Solar and Heliospheric Observatory (SOHO) 1995-Present

With data coming from multiple instruments, we utilized the TSI Composite Data Product from the Laboratory for Atmospheric and Space Physics (LASP) in order to make the data compatible across all four instruments.

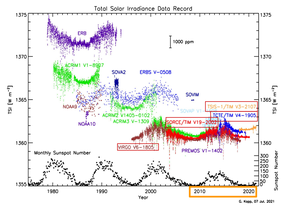


Figure 3: Total Solar Irradiance Record

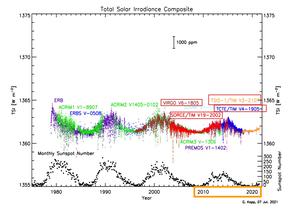


Figure 4: Total Solar Irradiance Composite

We gathered data on our features from the Solar Dynamics Observatory (SDO) which achieved first light in 2010 and is still currently taking data. The Helioseismic and Magnetic Imager (HMI) aboard provides a multitude of information on the photospheric magnetic activity and space weather features. We utilized line-of-sight magnetograms to learn about the active regions (AR) present on the Sun.

Feature Selection

Features used in this model:

- **Total Sunspot Area:**
  - Total sunspot area ∝ irradiance
  - For each snapshot of the Sun, we summed the total sunspot area for the image. See Feature Processing for how area was obtained.

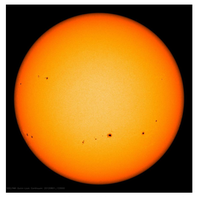


Figure 5: Image of Sun from SDO showcasing a sampling of sunspots. Credit: NASA SDO

- **Total Unsigned Flux:**
  - Total unsigned flux ∝ irradiance
  - Unsigned flux is related to the area of the plage regions near faculae on the Sun's surface (Photosphere)

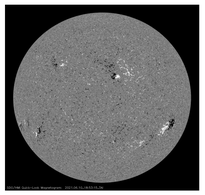


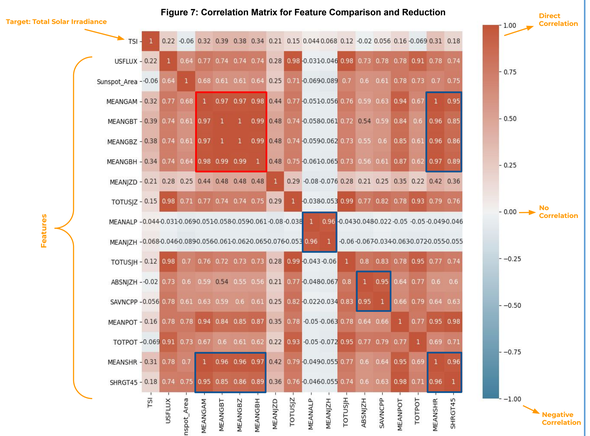
Figure 6: Example Magnetogram from SDO HMI Satellite Credit: NASA/SDO/Goddard

- **Spaceweather Parameters (SHARPs):**
  - A SHARP contains various spaceweather quantities calculated from the photospheric vector magnetogram data.

Feature Reduction

Incorporating too many features into our model can actually be counterproductive and cause problems in the model. In order to carefully select our features, we use a correlation matrix shown below to visually see which features are highly correlated and can thus be reduced. We removed six features using this tool based on the regions enclosed by boxes that show high correlations across multiple features.

Figure 7: Correlation Matrix for Feature Comparison and Reduction



Feature Processing

Most features can be put into our data frame directly from the data released by the satellites.

To obtain total sunspot area, we performed image processing using Python. The code isolated the area of the umbra and penumbra as shown in Figure 8 to the right. The areas were then summed for the total of each snapshot of the Sun and added to our dataframe.

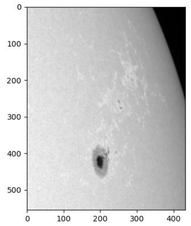


Figure 8: Sample Image from SDO showing a sunspot. The area of the sunspot (both umbra and penumbra) can be determined by image processing. Credit: NASA SDO

Results

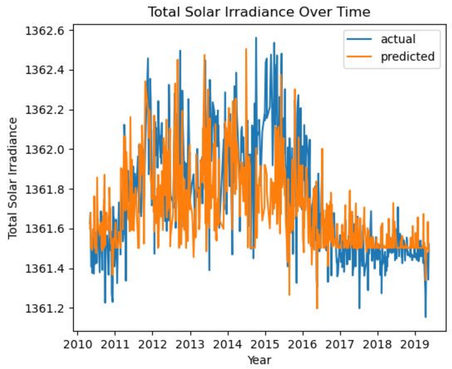


Figure 10: Plot of the actual TSI data overlain with our model's predictions

The plot above shows the TSI composite data product (actual TSI measurements) overlain with our model's TSI predictions. You will notice that the model does a decent job of predicting irradiance, but there is much room for improvement.

It is worth noting that prior iterations of this model were greatly skewed due to the presence of outliers. To correct for this, we used Random Sampling Consensus (RANSAC) which helps us achieve robust linear regression to mitigate the effects of these outliers.

Evaluation of the Model

Residual = Actual Data - Predicted Value

Residual	Meaning
Positive Value	Predicted too low
Zero	Prediction spot on
Negative Value	Predicted too high

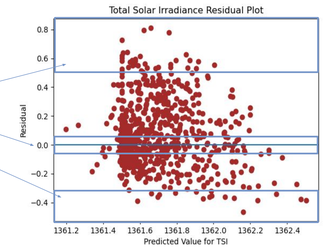


Figure 10: Plot of the residuals which show how far the predicted data is from the actual datapoints.

Root Mean Squared Error (RMSE) for our Model: 0.0491 4.91%	Mean Squared Error (MSE) (Quadratic Loss) for our Model: 0.0491 4.91%	Mean Absolute Error (MAE) for our Model: 0.1600 16.00%
$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$	$\mathrm{MSE} = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}$	$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \lvert y_i - x_i \rvert}{n}$
$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values	$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values	$y_1, y_2, \ldots, y_n$ are actual values
$y_1, y_2, \ldots, y_n$ are observed values	$y_1, y_2, \ldots, y_n$ are observed values	$y_1, y_2, \ldots, y_n$ are observed values
n is the number of observations	n is the number of observations	n is the number of observations

Analysis of Feature Importance

Helps us to understand what features were most important in predicting irradiance in our model.

Features:

- USFLUX: Total unsigned flux in Maxwells
- Sunspot_Area: Sum of the areas of the penumbra and umbra
- MEANGAM: Mean inclination angle, gamma, in degrees
- MEANJZD: Mean vertical current density, in mA/m2
- TOTUSJZ: Total unsigned vertical current, in Amperes
- MEANALP: Total twist parameter, alpha, in 1/Mm
- MEANJZH: Mean current helicity in G2/m
- SAVNCPP: Sum of the Absolute Value of the Net Currents Per Polarity
- MEANPOT: Mean photospheric excess magnetic energy density in ergs per cubic centimeter
- TOTPOT: Total photospheric magnetic energy density in ergs per cubic centimeter
- MEANSHR: Mean shear angle (measured using Btotal)

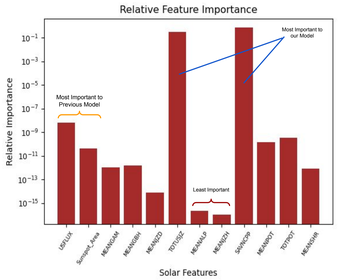


Figure 11: Plot of the relative feature importance between the features

Future Plans

In the future, we hope to work on refining and understanding the model.

- Testing out other machine learning models such a Random Forest Regression and Support Vector Regression
- Add additional features such as the location of the sunspots on the face of the Sun and number of sunspots
- Build a better understanding of the physical reasonings behind why certain features were deemed more important in the model than others. In hopes of better understanding the mechanisms behind the variations in irradiance.

Acknowledgements

This research was supported by the National Science Foundation REU Program, award #1659878

Special thanks to my REU mentors Laura Sandoval and Josh Elliott for their incredible support over the ten weeks of research!

Also to Caroline Mather and Megan Smith for their previous work which was utilized in this project and Dr. Greg Kopp who worked on the TSI Composite Data Product which was also utilized.

Contact Information: Please feel free to contact me at aelieber@uark.edu or connect with me on LinkedIn for any questions or comments.

References

- Coddington, O., Lean, J., Pilewskie, P., et al. 2019, Earth Space Sci., 6, 2525
- Bobra, M.G., Sun, X., Hoeksema, J.T., et al., "The Helioseismic and Magentic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs - Space Weather HMI Active Region Patches," Solar Physics, 2014, doi: 10.1007/s11207-014-0529-3
- Kopp, G., Krivova, N., Lean, J., and Wu, C.J., "The Impact of the Revised Sunspot Record on Solar Irradiance Reconstructions," Solar Physics, 2016, doi: 10.1007/s11207-016-0853-x
- Wu, C.-J., Krivova, N. A., Solanki, S. K., Usoskin, I. G., A&A, 620, A120, 12, https://doi.org/10.1051/0004-6361/201832956

RESEARCH POSTER PRESENTATION DESIGN © 2019
www.PosterPresentations.com