On Generalized Additive Models for Representation of Solar EUV Irradiance

Brandt, Daniel A.[1], **Vega, Erick. F.**[1], **and Ridley, Aaron J.**[2]

[1]Michigan Tech Research Institute - Michigan Technological University, Ann Arbor, MI, USA
[2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI USA

Key Points:

- Solar EUV irradiance is modeled using Generalized Additive Models parameterized using F10.7, revised Sunspot Number, and Lyman-α.
- Relative irradiance error is evaluated for solar cycle and seasonal dependence and compared to that of TIMED/SEE and SDO/EVE.
- Solar EUV irradiance hindcasts during Solar Cycle are generated to demonstrate the suitability of the GAM approach for forecasting.

Corresponding author: Daniel A. Brandt, daabrand@mtu.edu

Abstract

In the context of space weather forecasting, solar EUV irradiance specification is needed on multiple time scales, with associated uncertainty quantification for determining the accuracy of downstream parameters. Empirical models of irradiance often rely on parametric fits between irradiance in several bands and various solar indices. We build upon these empirical models by using Generalized Additive Models (GAMs) to represent solar irradiance. We apply the GAM approach in two steps: (1) A GAM is fitted between FISM2 irradiance and solar indices F10.7, Revised Sunspot Number, and the Lyman-α solar index. (2) A second GAM is fit to model the residuals of the first GAM with respect to FISM2 irradiance. We evaluate the performance of this approach during Solar Cycle 24 using GAMs driven by known solar indices as well as those forecasted 3 days ahead with an autoregressive modeling approach. We demonstrate negligible dependence of performance on solar cycle and season, and we assess the efficacy of the GAM approach across different wavelengths.

Plain Language Summary

Modeling solar irradiance at extreme ultraviolet wavelengths is very important for describing the behavior of the upper atmosphere. Many empirical models represent solar irradiance by describing it as linearly-dependent on measurements of other quantities that are very strongly correlated with it. These methods have shown great promise, but require building their models from many sources of data. We build upon these methods by showing that using only four sources of data (three solar proxies and irradiance from the FISM2 model), solar extreme ultraviolet irradiance can be modeled in different wavelengths efficiently. We use Generalized Additive Models (GAMs) for our approach, which are used to describe irradiance in terms of a sum of smooth functions of solar proxies. We show how this approach can be used to forecast solar EUV irradiance.

1 Introduction

Accurately describing space weather effects on the upper atmosphere is of critical importance for space situational awareness, satellite collision avoidance, safeguarding the electrical power grid, and protecting astronauts (Bussy-Virat et al., 2018). A key component of space weather operations involves describing the variability and effects of solar extreme ultraviolet (EUV) radiation, nearly all of which is absorbed in the thermosphere and serves as the dominant driver of energy input into the upper atmosphere during geomagnetic quiet times ((Stolarski et al., 1975; P. Richards et al., 1981)). Solar EUV irradiance additionally plays a central role in modulating the global variation of total electron content (TEC) (Hocke (2008); Lean et al. (2011)) and in driving the thermosphere response at multiple timescales (Guo et al., 2007).

Until the advent of endeavors such of SOLSTICE aboard the Upper Atmospheric Research Satellite (UARS) in 1991 (Reber et al., 1993), the SOlar Radiation and Climate Experiment (SORCE) in 2003 (Rottman, 2005), and the TIMED/SEE mission in 2001 (Woods, Bailey, et al., 2000), regular and semi-continuous measurements of solar EUV irradiance were not obtainable. As a result, solar proxies well-correlated with solar EUV irradiance which can be measured from the ground, such as F10.7 (Tapping, 2013), have seen regular usage due to their operational availability, and are routinely used as inputs for Ionosphere-Thermosphere models such as NRLMSISE 2.0 (Emmert et al., 2021) and Thermosphere Ionosphere Electrodynamics General Circulation Model (Cai et al., 2022). While these solar proxies have demonstrated applicability in downstream modeling for representing thermospheric and ionospheric climatology, they suffer from some important limitations, including:

1. Each solar index is best described as a proxy for solar processes occurring either in the photosphere, chromosphere, transition region, corona, or a combination of some of these regions, limiting their ability to capture the entire swath of variation throughout the entire EUV range (To et al., 2023).
2. The emissions most strongly correlated with each solar index are absorbed in different regions of the thermosphere and mesosphere, resulting in either increasingly complex parameterization for their ingestion into atmospheric models and non-trivial impacts on quantification of uncertainty in derived thermospheric temperatures and densities (Thayer et al., 2021).
3. The indices struggle to capture variation of solar irradiance both beyond timescales of ∼27 days, on the order of hours ((de Wit et al., 2008; Tapping & Charrois, 1994)), and do not capture the influence of solar flares inherent in EUV measurements (Pawlowski & Ridley, 2008).

These limitations have led to observed discrepancies in model performance when using different indices, especially when representing TEC (Tariku, 2019). More significantly, in the context of blackouts in solar proxies, it is possible to reconstruct those proxies for operational usage using solar radio measurements in neighboring wavelengths (Elvidge et al., 2023), which provides additional robustness for space situational awareness, but exacerbates the effects of the already inherent limitations of the index due to imperfect reconstruction.

In response to some of these concerns, empirical models of solar EUV irradiance have been developed, the outputs of which can be ingested into coupled atmospheric models such as the Global Ionosphere Thermosphere Model (GITM) (Ridley et al., 2006). The EUVAC model (P. Richards et al., 1994) is one such model that has seen regular used since its inception, and was succeeded by HEUVAC (P. G. Richards et al., 2006), which featured increased spectral resolution and flexible wavelength binning. Both models were developed based on a parameterization of F10.7, 81-day averaged F10.7 and the F74113 quiet sun reference spectrum derived from rocket measurements during the 1960s and 1970s (Heroux & Hinteregger, 1978). Other paradigms include the GOES-R EUVS model developed by the National Oceanic and Atmospheric Administration (NOAA) (Thiemann et al., 2019), the SOLAR2000 empirical solar irradiance model and forecast tool, which was implemented with the development of the E10.7 solar EUV proxy derived from the time-dependent integrated solar EUV flux at the top of the terrestrial atmosphere (Tobiska et al., 2000), and the Flare Irradiance Spectrum Model (FISM) (Chamberlin et al., 2007). The current iteration, FISM2 (Chamberlin et al., 2020), includes a daily component and a flare component, and improved empirical modeling due to the incorporation of measurements from the Solar Dynamics Observatory (SDO)/Extreme Ultraviolet Variability Experiment (EVE), SORCE/X-ray Photometer System (XPS), and SORCE/Solar Stellar Irradiance Comparison Experiment (SOLSTICE).

This work shows the suitability of FISM2 outputs for operational use, and builds upon the empirical modeling paradigm by introducing a novel framework for parameterizing solar EUV irradiance using Generalized Additive Models (GAMs), a class of linear models that relate a response variable linearly to smooth functions of predictor variables of interest (Hastie, 2017). This class of models allows for the capturing of the proportional impacts of various solar processes represented by proxies on each wavelength band, and natively support intuitive quantification of uncertainty in modeled irradiances due to the relating of the distribution of expected values of the response variable to the predictor variables through a link function. Since reliable measured solar EUV irradiance records do not extend before the early 1990s, we leverage FISM2 model results for data prior to that period, and construct a GAM to represent integrated solar EUV irradiance across all the considered wavelength bands using that data, after which we construct a second GAM to model residuals with respect to native FISM2 outputs. The results of the GAM approach are evaluated using known historical space weather drivers,

and are compared to TIMED/SEE and SDO/EVE measurements in select wavelength bands. GAM results are also evaluated using space weather drivers hindcasted 3 days into the future using an Autoregressive Modeling approach during two 30-day periods in SC24: one corresponding to low solar activity and one corresponding to high solar activity. The contributions this study include (1) the application of a rigorous mathematical modeling method for simplifying representation of solar EUV irradiance, and (2) the initial verification of this method for forecasting over time scales routinely utilized in the space weather community through hindcasts performed over low and high solar activity periods.

We wish to note that relative weakness of Solar SC24 in comparison to the preceding Solar Cycles may affect the results. While we foresee this effect being minimal since solar indices during this time show a weakness that is similar to the same degree, we acknowledge that it is of enough significance to warrant investigating the cycle-by-cycle modulations in indices and drivers in more detail in a future study. We also wish to mention that we have limited consideration of TIMED/SEE results to 16 bands between 280 and 1170 Å due to the fact that TIMED/SEE did not make spectral measurements below about 280 Å; the TIMED/SEE Level 3 irradiance below 280 Å is derived from a model, subjecting it to model error. Additionally, above 1170 Å, FISM2 is calibrated against SORCE/SOLSTICE, which is more accurate than TIMED/SEE, and disagreement between FISM2 and TIMED/SEE in this range is expected and systematic due to FISM2's higher absolute accuracy. Likewise, due its limited wavelength range, we only consider 21 wavelength bands corresponding to SDO/EVE, which suffers much less from inaccuracies due to degradation than TIMED/SEE due to regular calibrations.

The obtained GAM functions can be used for (1) operational assessment of the current state of solar irradiance, specifically when there are data outages, (2) improved thermosphere and ionosphere modeling and uncertainty quantification of the atmospheric state, and (3) forecasting future values of solar EUV irradiance across multiple wavelength bands and on multiple time scales.

The paper is organized as follows, first, the data pre-processing strategy is outlined. Then, the GAM approach for modeling irradiance is described, followed by a description of Autoregressive Models. After that, a description of the techniques used for validation of results is outlined. Thereafter, irradiance estimation results are presented and assessed by comparison against native FISM2 outputs, TIMED/SEE, and SDO/EVE. Finally, the results are discussed and conclusions drawn.

2 Methodology

2.1 Data Pre-processing

We consider F10.7, revised Sunspot Number (SSN) and Lyman-α as drivers for the GAMs. These indices were chosen primarily due to their widespread use and familiarity within the space weather community. Additionally, since indices serve as proxies primarily for the upper chromosphere/lower corona and solar photosphere (Nusinov & Katyushina, 1994; Johnson et al., 2023), constraining model parameterization to these parameters demonstrates the versatility of the GAM approach in its ability to address the problem of solar EUV modeling even with limited information. For this reason, we have not included other proxies such as the Mg-II cwr or Ca-II indices, even though their inclusion may increase accuracy by capturing the dynamics of slowly varying corona (Schonfeld et al., 2019) and better describing variation within entire chromosphere (Viereck et al., 2004). The selected indices were obtained from the NASA OMNIWeb Data Explorer (`https://omniweb.gsfc.nasa.gov/form/dx1.html`). Anomalous values were removed from the OMNIWeb solar index data, and the FISM2 Level 3 TIMED/SEE, and Level 3 SDO/EVE irradiances were obtained from the LASP Interactive Solar Irradiance Datacenter (`https://`

`lasp.colorado.edu/lisird/`) and were upsampled from their native daily resolution to the hourly resolution of the OMNIWeb solar indices. Additionally, the FISM2 and TIMED/SEE irradiances were arranged into 59 wavelength bins used by GITM for ease of comparison and for eventual ingestion into GITM (see Table 1 below). Due to the lower accuracy of TIMED/SEE compared to FISM2 outside of 280-1170 Å, only 16 non-singular wavelength bins between 280 and 1170 Å were used from TIMED/SEE for additional comparisons. In the case of SDO/EVE, the Level 3 data are bounded between wavelengths centered at 65 to 1055 Å, in 10 Å intervals; we consider 21 of these wavelength bands in our analysis. In Table 1 below, the salmon-colored bins correspond to those for which there was SDO/EVE data alone, the orchid-colored bins correspond to where there was both TIMED/SEE and SDO/EVE data, and the sky blue-colored bin corresponds to where there was TIMED/SEE data alone. FISM2 data was available in every bin.

Table 1: Solar EUV irradiance wavelength bins considered for analysis.

Lower Boundary (Å)	Upper Boundary (Å)	Bin Center (Å)
1	2	1.5
2	4	3
4	8	6
8	16	12
16	23	19.5
23	32	27.5
32	50	41
50	100	75
100	150	125
150	200	175
200	250	225
		256.3
		284.15
250	300	275
		303.31
		303.78
300	350	325
		368.07
350	400	375
400	450	425
		465.22
450	500	475
500	550	525
		554.37
		584.33
550	600	575
		609.76
		629.73
600	650	625
650	700	675
		703.31
700	750	725
		765.15
		770.41
		789.36
750	800	775
800	850	825
850	900	875
900	950	925

		977.02
950	1000	975
		1025.72
		1031.91
1000	1050	1025
1050	1100	1075
1100	1150	1125
1150	1200	1175
		1215.67
1200	1250	1225
1250	1300	1275
1300	1350	1325
1350	1400	1375
1400	1450	1425
1450	1500	1475
1500	1550	1525
1550	1600	1575
1600	1650	1625
1650	1700	1675
1700	1750	1725

Binning was performed as follows: For each spectrum at a given time, irradiance at the wavelength nearest to each of the singular wavelengths was obtained. These constitute the singular wavelengths, or 'lines'. Afterwards, the irradiance in each wavelength range (the remaining non-singular bins) was calculated by summing the irradiance in that bin. This was done simply by adding the values of irradiance corresponding to the wavelengths between the boundaries of each bin. Figure 1 shows the EUV spectrum for FISM2, TIMED/SEE, and SDO/EVE during solar maximum in SC24, obtained via our binning procedure.

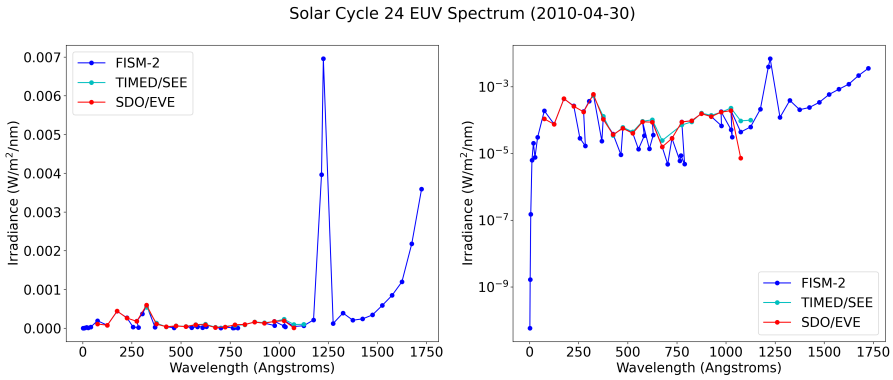

Figure 1: The Solar EUV spectrum for SC24 during April of 2010 in the ascending phase: (left) The spectrum in base units, and (right) the spectrum on a logarithmic scale. In both figures, blue represents FISM2 estimates, cyan represents TIMED/SEE measurements, and red represents SDO/EVE measurements.

2.2 Generalized Additive Models

Conventional linear models (Nelder & Wedderburn, 1972) assume a response variable y follows an exponential family distribution with mean μ, which may be a nonlinear function of $\beta_i \mathbf{X}$, where β_i are coefficients and $\mathbf{X} = [X_1, X_2, ..., X_n]$ are covariates (predictor variables):

$$g\left(\mathbb{E}\left[y\,|\mathbf{X}|\right]\right) = \beta_0 + \sum_{i=1}^{n} \beta_i \mathrm{X}_i, \tag{1}$$

where g is the link function relating the predictor variables to the expected value of the dependent variable y.

GAMs, by contrast, are linear models in which the response variable depends linearly on *unknown* smooth functions of several predictor variables. They take the following form:

$$g\left(\mathbb{E}\left[y\,|\mathbf{X}|\right]\right) = \beta_0 + \sum_{i=1}^{n} f_i\left(\mathrm{X}_i\right), \tag{2}$$

where f_i are 'feature functions' which may be constructed from various families of bases functions. In the context of this paper, the solar EUV irradiance in a given band y was regressed using the predictor variables Day-of-the-year (DOY), F10.7, SSN, and Lyman-alpha. We specifically include DOY as a model driver in order to represent seasonal variations in the solar indices. The regression was carried out using the GAM framework, which required the fitting of univariate spline functions for each those predictor variables. We assumed a normal/Gaussian error model for the responses y which naturally leads to a least-squares fitting problem. In our analysis, we proceeded with our approach as follows:

1. For SC20 through the beginning of SC23, an initial GAM Y was fit between solar EUV irradiance represented by FISM2 and the F10.7 index, revised Sunspot Number (SSN) (Clette et al., 2015), and the Lyman-α index (Woods, Tobiska, et al., 2000) using penalized B-splines and the Normal distribution.
2. A second GAMs ζ to capture the behavior of initial model residuals was fit during the remainder of SC23.
3. SC24 integrated solar EUV irradiance was modeled using the $F = Y - \zeta$ for known solar inputs (termed models F^K; likewise, Y^K refers to models Y driven by known solar inputs).
4. Relative irradiance error ε with respect to native FISM2 was compared from F^K, and assessed for solar cycle, seasonal, and solar activity dependence. This was also done for TIMED/SEE Level 3 irradiances in 16 wavelength bands between 280 and 1170 Åand SDO/EVE Level 3 irradiances in 21 bands between 65 and 1055 Å.
5. The approach was applied again for 59 different wavelength bands, to assess the behavior of the mean, standard deviation, kurtosis, and skew of relative error ε as a function of wavelength band i. These bands were selected due to their usage in global ionosphere-thermosphere models such as GITM and Aether (`https://aetherdocumentation.readthedocs.io/en/latest/index.html`). This step provided us with models $F_i = Y_i - \zeta_i, i \in [0, 59]$.
6. The above steps were applied for a 3-day ahead Autoregressive Model hindcasts of the solar indices mentioned in Step 1, which were used to drive the GAMs. This was done for two 30-day periods in SC24: one during the low solar activity in the ascending phase and one during high solar activity during the peak.

The initial fitting for Y_i was performed on solar index and FISM2 data between the beginning SC20 (October 10, 1964) and the peak of SC23 (taken as January 23, 2002). The fitting of ζ_i was done between the Y_i residuals and solar indices between the peak of SC23 and the remainder of its descending phase (up to December 1, 2008). The final models F_i were evaluated over the entirety of SC24 (December 1, 2008 to December 1, 2019). The GAMs were implemented with the aid of the recently-developed PyGAM package (Servén & Brummitt, 2018).

An example of the results of the fitting procedure in two different wavelength bands (centered at 19.5 Å and 225.0 Å is shown) is shown in Figure 2 below. The solar indices used to perform the fits also shown, along with the GAM outputs for the same period of time corresponding to the training set used for fitting.

Figure 3 shows the results of the fitting procedure for ζ_i for the wavelength bands centered at 19.5 Å and 225 Å. We observe that the aim of the inclusion of ζ_i is to account for any trends in the residuals related to season or solar activity. While we acknowledge the limitations of fitting ζ_i only during the descending Phase of SC23 to achieve that aim effectively, we elected to restrict the fitting to that period of time in order to demonstrate the effectiveness of the approach with use of limited information, and to avoid the problem of over fitting.

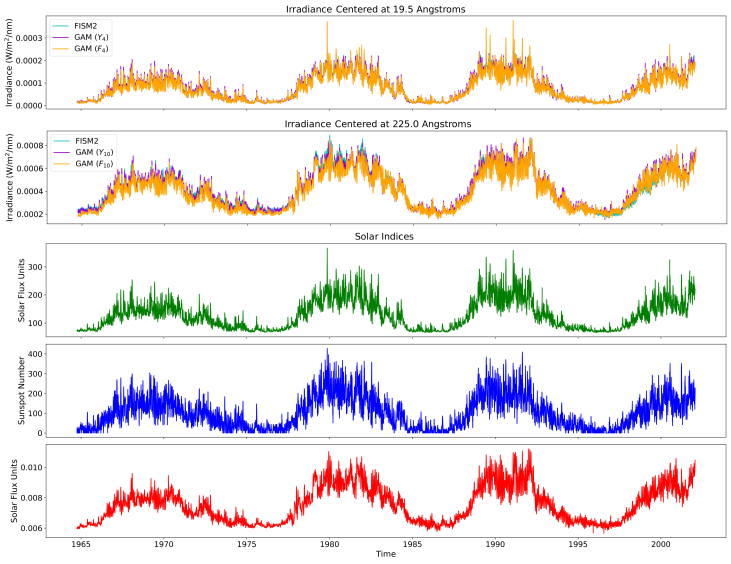

Figure 2: Time series data for solar EUV irradiance centered at 19.5 Å (top), solar EUV irradiance centered at 225.0 Å. (second from the top), F10.7 (middle), SSN (second from the bottom), and Lyman-α (bottom). The top two solar EUV plots show data for FISM2 and for the GAMs Y_i (initial fit) and F_i (initial fit - model for residuals). The model fits were performed between the beginning of SC20 and the peak of SC23.

Figure 4 shows the results of the fitting procedure in the same two bands, but viewing up close a time period during the descending phase of SC22, between March 1, 1993 and December 1, 1996. From this figure can be seen qualitatively how the inclusion of the second GAM ζ_i improves the overall correspondence between the model outputs and FISM2, while avoiding over-fitting.

For added clarification on the construction of the GAM, we show Partial Dependence Functions (PDPs) for GAMs Y_i for wavelengths centered at 19.5 Å and 225 Å in

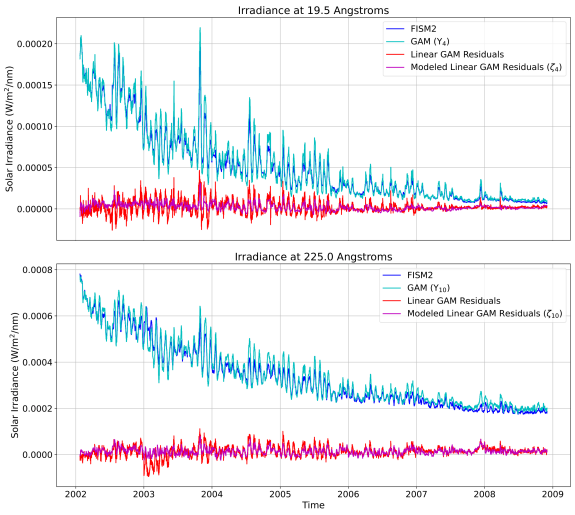

Figure 3: The results of fitting the GAM ζ_i to model residuals for solar EUV irradiance centered at 19.5 Å and at 225 Å. The time period over which this fit occurred corresponds to the descending phase of SC23.

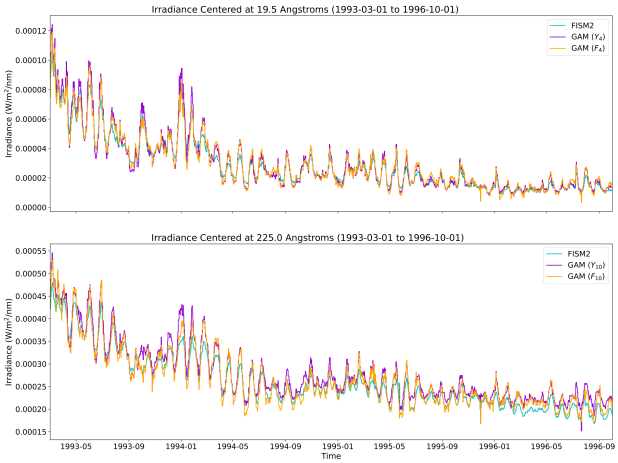

Figure 4: A closer examination (during the descending phase of SC22) of the same model results as shown in the top two plots of Figure 2.

Figures 5 and 6 below. PDPs are widely-used within the field of interpretable Machine Learning to explain the marginal effect of single features on a model prediction (Friedman, 2001). In the case of the present study, the PDPs are identical with the aforementioned feature functions used to construct the GAM. Thus, in the examples shown, they illustrate the proportional functional contributions to solar EUV irradiance in a specific band due to a specific feature represented by either DOY (a proxy for season) or a solar index. The versatility of the GAM approach is demonstrated by the varied nature of PDPs for different wavelengths. This behavior is expected, since each model driver is not expected to retain the same relationship with each wavelength band.

In particular, for the examples shown, we observe a strongly oscillatory behavior between DOY and its respective contribution to solar EUV centered at 19.5 Å that, by

inspection, is characteristic of a dominant period between 75 and 100 days in duration. This is contrasted with the same PDP for 225 Å, which shows not only similar oscillations on a shorter characteristic period of $\sim$ 50 days in duration, but an annual trend that reaches a trough just prior to the summer solstice.

Similarly, we observe unique behavior for the PDPs of the solar indices for each wavelength. For 19.5 Å, the PDP for F10.7 is nearly linear, and only fails the to meet the description of a monotonically increasing function at values beyond F10.7 = 300. This, however, is not observed with the wavelength bin centered at 225 Å which behaves nearly like a negative quadratic function until F10.7 = 300. The PDP for SSN at 19.5 Å shows an inverse relationship, while at 225 Å, this inverse relationship persists only until SSN$\sim$140, after which it reverses, terminating in a sharp, nearly-exponential relationship above SSN$\sim$ 400. Lastly, the PDP for Lyman-α is a non-monotonic function at 19.5 Å, but it is a monotonically increasing function at 225 Å.

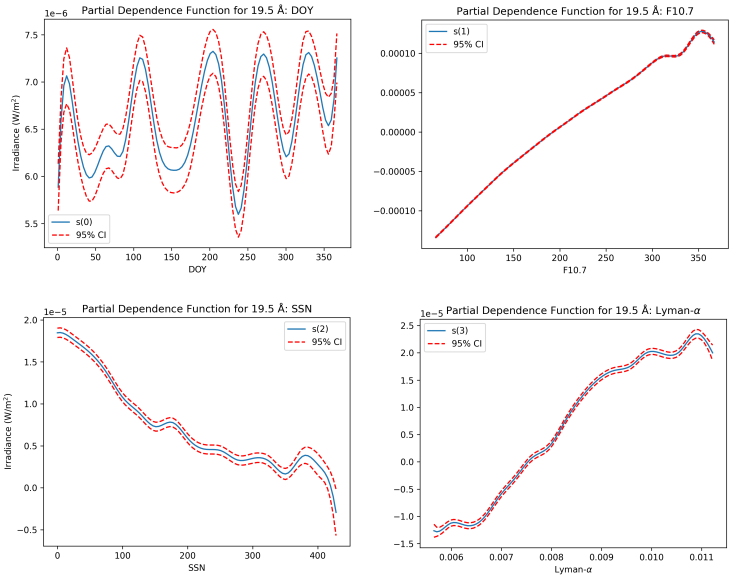

Figure 5: PDPs for the GAM Y_i used to model solar EUV irradiance centered at 19.5 Å. Clockwise from the top-left: the PDP for DOY, for F10.7, SSN, and Lyman-α. In each plot, the PDP itself is in blue, and the red dashed lines are the corresponding 95% confidence bands.

2.3 Autoregressive Models

Autoregressive (AR) models represent random processes by modeling values at future time steps as a weighted sum of values at previous time steps. Conventionally, for some AR model of order n, values of a quantity x at time t are given by

$$x_t = \sum_{i=1}^{n} \beta_i x_{t-i} + \delta_t, \tag{3}$$

where the β_i are the parameters of the model and δ_t is white noise (Box et al., 2015). AR models have been used for probabilistic forecasting of the disturbance storm time index (Chandorkar et al., 2017), predicting MeV electron fluxes in the outer radiation belt (Sakaguchi et al., 2015) and numerous AR approaches have been utilized for fore-

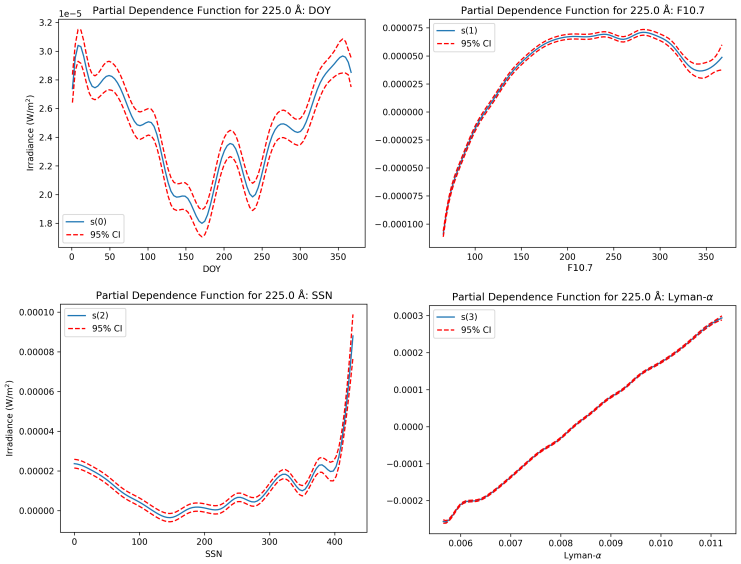

Figure 6: The same as Figure 5, but for solar EUV irradiance centered at 225 Å.

casting solar proxies (Lean et al., 2009; Si-Qing et al., 2010; Chattopadhyay & Chattopadhyay, 2012; Du, 2020). In this study, we highlight the applicability of the GAM approach for forecasting by mirroring the approach of (Du, 2020) to hindcast the solar indices F10.7, Lyman-α, and SSN 3 days into the future. These hindcasted indices are then used to drive the GAMs in SC24 over a 30-day period (just beyond the duration of a single solar rotation) during Solar Minimum (December 1, 2008 through December 29, 2008) as well as Solar Maximum (June 1, 2014 through June 29, 2014).

In our approach, we rely on an AR paradigm known as the Autoregressive Integrated Moving Average (ARIMA), which not only models the variable of interest as a function of its own prior values, but utilizes a moving average to model the regression error as a linear combination of error terms whose values occurred contemporaneously in the past. ARIMA models take advantage of non-stationarity (the mean and variance of a process vary as a function of time) and thus are suitable for forecasting solar indices, since they have been found to exhibit heteroscedasticity (Wang et al., 2018). ARIMA models have a general form given by the following:

$$x_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_p x_{t-p} + \phi_1 \delta_{t-1} + \phi_2 \delta_{t-2} + ... + \phi_q \delta_{t-q}, \tag{4}$$

where α is a constant, β_i are AR model parameters for an AR model order of p, ϕ_i are moving average model parameters up to order q, and δ_i are lagged forecast errors.

A key component of ARIMA models is that they employ differencing in order to enforce stationarity. This involves subtracting the previous value from the current value a total of d times. In the present study, during the application of our AR approach, we focused on forecasting daily solar indices, and set $p = 33$, $d = 1$, and $q = 2$, similar to the methods for short-term F10.7 AR forecasting recommended by Du (2020). An example of the solar indices hindcasted during SC24 is shown in Figures 7 and 8 below.

We utilized the Python Statsmodels package to perform model fitting and forecasting (Seabold & Perktold, 2010). We briefly consider the percentage difference between

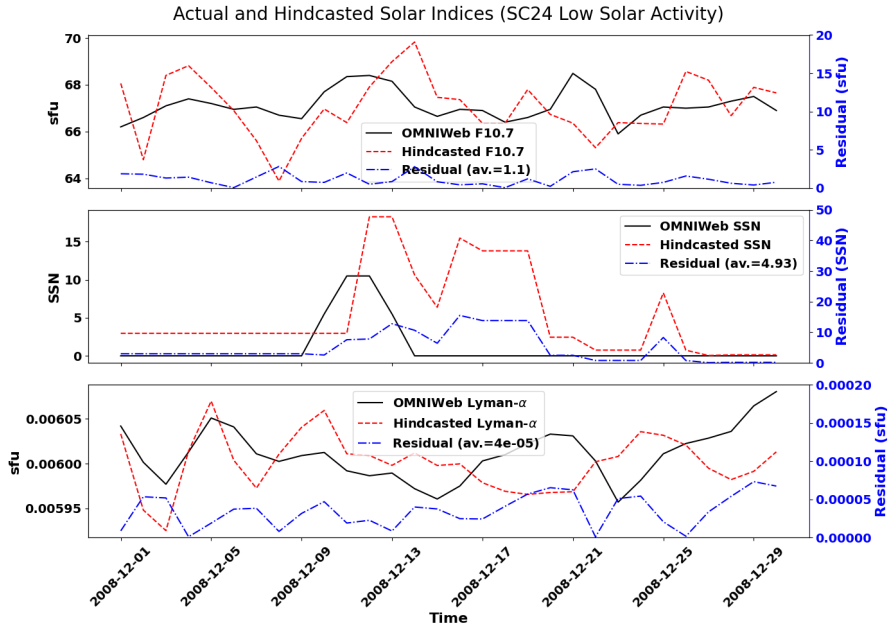

Figure 7: Solar indices hindcasted with an ARIMA approach using $(p, d, q) = (33, 1, 2)$, between Dec 1, 2008 and Dec 30, 2008 of SC24.

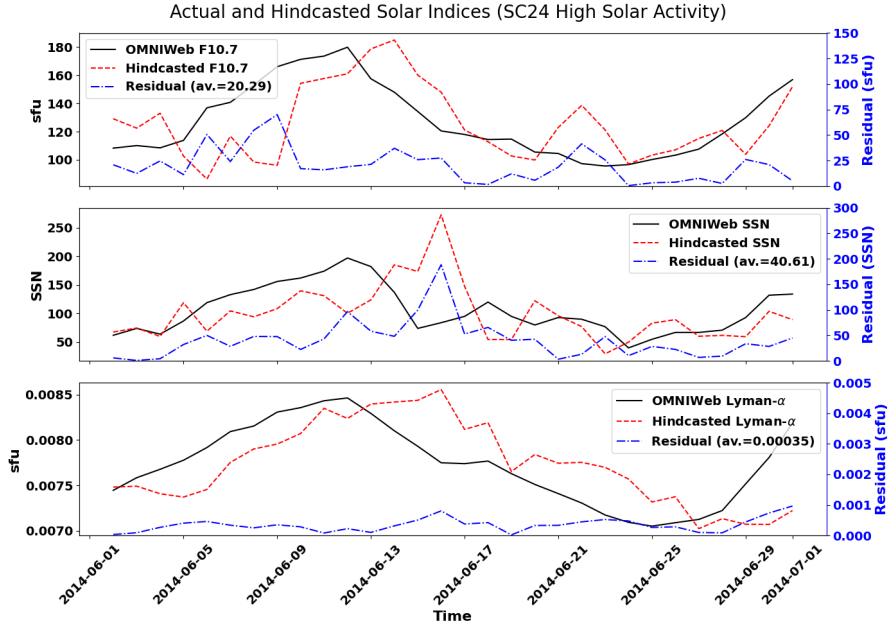

Figure 8: The same as Figure 7, but between June 1, 2014 and July 1, 2014 of SC24.

average values of the residuals for each hindcasted solar index, using the expression below:

$$P = \frac{x_i - x_f}{\left[\frac{(x_i + x_f)}{2}\right]} \times 100, \tag{5}$$

where x_i is the initial average value of the residuals (in this case corresponding to low solar activity) and x_f is the final average value of the residuals (in this case corresponding to high solar activity). Overall, we observed that the hindcasted indices exhibited lower prediction error during low solar activity, with values of the residuals increasing by a percentage difference of $P \approx 179.43\%, 156.7\%$, and 158.97% for F10.7, SSN, and Lyman-α, respectively, from low solar activity to high solar activity. We suspect this is primarily due to increased uncertainty owing to the impulsive nature and sporadic occurrence of active regions on the solar disk during high solar activity as suggested by Du (2020) and observed in F10.7 by Wilson et al. (1987), especially due to the fact that on time scales in excess of 1-2 days, magnetic structures tend to dominate in affecting fluctuations in solar irradiance (Solanki et al., 2003). Despite this, a detailed investigation and compensation for this behavior is beyond the scope of this paper. For more detail on AR models, we refer the reader to Shumway et al. (2017).

2.4 Validation

To assess the accuracy of the estimated irradiances, we followed an analogue of the analysis of (Gondelach & Linares, 2021). This involved comparing estimated irradiances with native FISM2 estimates of solar EUV irradiance. Due to non-trivial degradation issues with TIMED/SEE that have worsened in severity since late 2017, and due to the limited wavelength coverage for SDO/EVE, we avoid using both of those sources as controls for the evaluation of the GAM results. Measurements for TIMED/SEE and SDO/EVE are included in select results for qualitative comparison only. For the comparisons, we define the relative irradiance error as follows:

$$\varepsilon = \frac{\bar{I}_{EST} - \bar{I}_{FISM2}}{\bar{I}_{FISM2}} \times 100\%, \tag{6}$$

where $\bar{I}_{EST}$ and $\bar{I}_{SEE}$ indicate the daily average estimated and measured solar irradiance. For additional insight into the behavior of the relative irradiance error, we generated histograms of ε by wavelength band along with corresponding fits to a skew normal distribution, and we assessed variation of ε as a function of Day-of-the-Year (DOY) and solar activity proxied by F10.7.

We wish to note that particularly at wavelengths below 6 Å, use of Equation 6 becomes less insightful, since the characteristic irradiances are on the order of 10^{-8} W/m^2/nm, and discrepancies of two or three orders of magnitude can result in values of ε that rapidly grow beyond hundreds of percent. Since these discrepancies are expected (as it is routine for the irradiance in these lower bands to often be measured at 0, or nearly equivalent to it in numerical precision), we also make use of the Normalized Root Mean Square Error (NRMSE) when evaluating performance as a function of wavelength band. RMSE is defined as follows (Wilks, 2011):

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\bar{I}_{EST} - \bar{I}_{FISM2}\right)}, \tag{7}$$

where n represents the number of data points, and the remaining variables are defined as they are in Equation 6. We performed normalization in the same manner as Lean et al. (2009) by multiplying by 100 and dividing by the mean value of the observed values (i.e.: $NRMSE = \frac{100RMSE}{<\bar{I}_{FISM2}>}$). In the determination of this statistic, we minimized the

influence of outliers by computing the centered 24-hour rolling NRMSE with respect to FISM2, and considering the median values obtained across all considered wavelength bands.

3 Results

Results for the GAM approach are first given for integrated solar EUV irradiance across all 59 wavelength bands. We consider how well results from F^K correlate with FISM2, as well as the distribution of relative irradiance error ε over SC24. Thereafter is detailed the behavior of the variance and skew of $\bar{\varepsilon}$ as a function of band, followed by an example of forecasting 3-days ahead with the AR approach.

3.1 Integrated Solar EUV

Figure 9 shows the overall result of the GAM approach using known solar indices. With the exception of the very end of the declining phase of SC24. We observe consistent correspondence between F^K and FISM2 throughout the entirety of SC24.

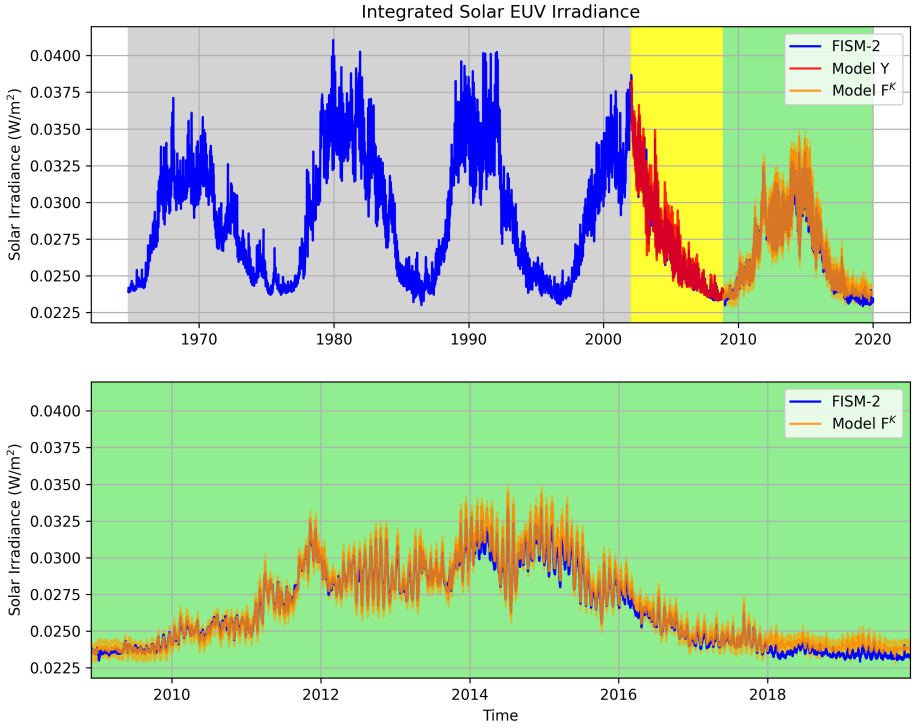

Figure 9: Results of the GAM approach applied to solar EUV irradiance integrated across all considered wavelengths, showing the entirety of the training, correction, and test sets (top) and a zoom in on the test set alone (bottom). In the top, the training region over which Y was fitted is shaded grey, the correction region over which ζ is fitted is shaded yellow, and the test region over which F is evaluated is shaded green. The pale orange shaded region in both figures corresponds to the 95% confidence interval for F^K.

For Y and ζ, we use the default condition that spline terms have a penalty on their second derivative, which encourages the feature functions to be smoother. In both cases, the regularization parameter λ, which controls the strength of this penalty, was set to 0.6 for all terms. The number of samples η used to fit each model differed due to the training set for Y being ∼5.5 times greater in length than the set used for fitting ζ. We obtained important statistics for Y and ζ, including deviance explained, scale, and McFadden's pseudo R-squared (McFadden et al., 1973). Deviance, as detailed by Wood (2017), is defined as

$$D = 2\left[l\left(\hat{\beta}_{\mathbf{max}}\right) - l\left(\hat{\beta}\right)\right]\phi, \tag{8}$$

where $l\left(\hat{\beta}_{\mathbf{max}}\right)$ is the maximized likelihood of the saturated model(a model with one parameter for each data point, so that the data are fitted exactly), $l\left(\hat{\beta}\right)$ is the maximized likelihood of the fitted model, and ϕ is the scale parameter. The scaled deviance is given by

$$D^* = D/\phi \tag{9}$$

In the case of Y and ζ, ϕ is estimated during model fitting, and represents the residual standard error squared, due to the use of the Normal distribution. The deviance explained Ξ then corresponds to the representing D^* as the proportion of total deviance explained by the current model. We also computed McFadden's adjusted pseudo R-squared ($\rho^2_{\mathbf{adj}}$) as the coefficient of determination for Y and ζ, allowing us to determine the proportion of variation of integrated solar EUV irradiance predicted by the fitted model parameters, while controlling for the number of those parameters (more detail may be found in Long and Freese (2006)). Succinctly, this statistic gives us an idea of how much variation in each of the model parameters affects changes in the irradiance. Table 2 below shows the values of η, Ξ, and $\rho^2_{\mathbf{adj}}$ for both Y and ζ.

Integrated Solar EUV Irradiance GAM Statistics			
Model	η	Ξ	$\rho^2_{\mathbf{adj}}$
Y	327,049	0.981	0.981
ζ	59,160	0.476	0.476

Table 2: Statistics for the components of F for representing integrated Solar EUV irradiance, rounded to three decimal places.

While we expect values of $\rho^2_{\mathbf{adj}}$ to run lower than conventional R^2, as shown in Figure 5.5 of McFadden et al. (1973), we note that for both Y and ζ, values of $\rho^2_{\mathbf{adj}}$ are sufficiently high (especially for the former case) to show excellent model fit for both parameterizing FISM2 irradiance and the associated residuals with respect to native FISM2 with a GAM, respectively.

Next, we consider how well the GAM correlates with native FISM2 over SC24. Figure 10 shows the correlations for F and FISM2. The correlation is significantly positive with a value of Pearson's Correlation Coefficient of 0.992. We observe that the linear fit between the outputs of F and that of native FISM2 suggests a tendency for overestimation that slightly increases as a function of solar irradiance. We attribute this primar-

ily to the relatively short period over which ζ was fit in comparison to Y, though we note that the overestimation grows only from $\sim$ 0.0003 to $\sim$ 0.001 W/m^2 throughout the entire range of irradiance values, showing a high degree of consistency.

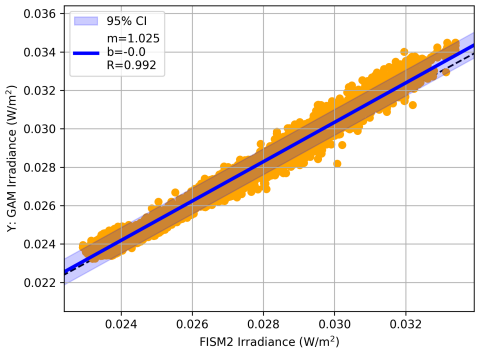

Figure 10: The linear relationship between the GAM outputs from FK and FISM2, for solar EUV irradiance integrated across all 59 wavelength bands. The blue line corresponds to the line-of-best fit, and the red shaded region corresponds to the 95% confidence interval.

We also considered the variation of relative irradiance error ε over the solar cycle in general, as shown on the left in Figure 11. To illustrate the improvement afforded by the inclusion of ζ, we show the relative error over SC24 between both Y^K and native FISM2 as well as F^K and native FISM2.

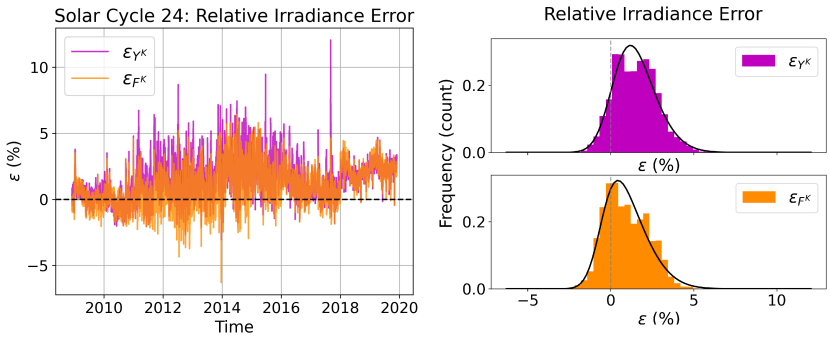

Figure 11: Left: Relative error (ε) for Y^K and F^K with respect to FISM2 integrated solar EUV irradiance throughout the entirety of SC24. These results demonstrate laudable performance at under 5% relative error for F^K throughout nearly the entirety of SC24. They additionally show that these improvements are possible in part due to the action of ζ in reducing error. Right: Relative error (ε) histograms for Y^K (top) and for F^K (bottom), with respect to FISM2 integrated solar EUV irradiance for the entirety of SC24. The action of ζ is observable in its moving the center of the distribution of errors closer to zero for F^K compared to Y^K, as well as reducing the width of the distribution.

For added clarity we generated histograms of ε for Y^K and F^K with respect to FISM2, and skew normal distributions were fit in order to elucidate the statistical behavior of the ε (on the right of Figure 11). The resulting parameters describing each distribution

can be found in Table 3. We consider in particular the shape (α), location (ξ), scale (ω), kurtosis (κ), and skewness (γ). We observe a growth in the value of α after applying ζ, which, along with a decreased value of κ, indicate that the effect of the inclusion of the second GAM is to draw mean error closer to zero and constraint the majority of errors to clustering around values in the vicinity of $\sim 2.5\%$. The negative excess kurtosis shown by the value of κ for F^K also indicates that the inclusion of ζ results in a reduction of the likelihood of errors attaining values more extreme that that corresponding to a normal distribution.

Integrated Solar EUV Irradiance ε Skew Normal Statistics					
Model	α	ξ	ω	κ	γ
Y^K	1.878	0.201	1.829	0.657	0.424
F^K	2.595	-0.551	1.971	-0.515	0.296

Table 3: Statistics for skew normal distributions of ε for Y^K and F^K, rounded to three decimal places.

3.2 Dependency of ε on Season and Solar Activity

To evaluate dependency on season and solar activity, we sorted values of ε by DOY and F10.7 (Figure 12). Linear fits to the sorted data resulted in slopes (m) and intercepts (b) suggesting a slightly negative correlation (Pearson's R≈-0.303) between season and relative error, while the converse is true regarding dependency on solar activity (Pearson's R≈0.134). We note that the clustering of values of ε plays a role in affecting the resulting linear fits, particular for dependency on solar activity. The coefficient of determination R^2 for the linear fit of ε to DOY was $\sim$ 0.092, while it was $\sim$ 0.018 for F10.7. These values indicate no statistically-significant relationship between DOY or solar activity and relative irradiance error.

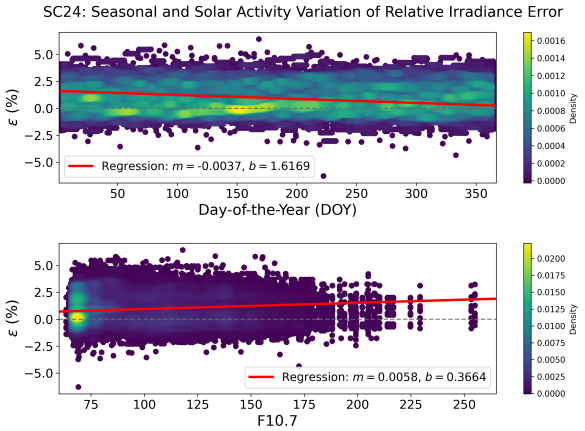

Figure 12: ε for integrated solar EUV irradiance across 59 wavelength bins as a function of season (top) and ε as a function of F10.7 (bottom). The coloring of the data in each plot relates to the density of the data points, which in each plot, have been distributed over 50 bins in both axes.

3.3 Behavior of ε as a Function of Wavelength

Next, we consider the behavior of mean (μ_{ε_i}) and standard deviation (σ_{ε_i}), kurtosis (κ_{ε_i}), and skew (γ_{ε_i}) relative irradiance error as a function wavelength band. We additionally compare the values corresponding to F_i^K with those of TIMED/SEE and SDO/EVE with respect to FISM2, in select bands. We show these results for wavelengths between 4 Å and 1750 Å, inclusive (Figure 13). For the bin centered at 1.5 Å, we observed that $\mu_{\varepsilon_0} \approx 125.18$, $\sigma_{\varepsilon_0} \approx 1546.73$, $\kappa_{\varepsilon_0} \approx 613.29$, $\gamma_{\varepsilon_0} \approx 23.30$, and for the bin centered at 3 Å, we observed that $\mu_{\varepsilon_1} \approx 355.44$, $\sigma_{\varepsilon_1} \approx 4800.59$, $\kappa_{\varepsilon_1} \approx 28.22$, $\gamma_{\varepsilon_1} \approx 955.11$. These values, particularly for μ_{ε_i} and σ_{ε_i}, differed significantly than those of the other bands, hence their suppression in Figure 13. In these two wavelength bands, FISM2 irradiances routinely were at values of zero, outside of which they would oscillate according to a pattern typical of neighboring wavelength bands and solar indices. By virtue of its construction, the GAM models quantities smoothly, and therefore would occasionally return negative values of irradiance in these bands over regions where it the prediction should be zero. In these cases, we manually zeroed the model estimate. Unfortunately in cases where FISM2 values were nonzero but exceedingly small (on the order of (10^{-11} W/m^2/nm and less) and the GAM result was greater by 1 or 2 orders of magnitude, and in the converse, we observed a dramatic increase in the estimates of relative irradiance error. In these bands, when temporal averaging is employed over a rolling 30-day period, the values of μ_{ε_i} in particular, are reduced to 87.07% for the bin centered at 1.5 Å, and to 84.55% for the bin centered at 3.0 Å (Figures 14 and 15). This is expected, as the effect of the temporal averaging is to act as a low-pass filter that excludes high frequencies corresponding to the most extreme deviations of model results from FISM2.

Figure 13: Clockwise from the top left: Mean, Standard Deviation, Kurtosis, and Skew of relative irradiance error as a function of wavelength band, for GAMs F_i^K, TIMED/SEE, and SDO/EVE, with respect to FISM2. The bands shown exclude the first two from Table 1, and are only those between 4 Å and 1750Å inclusive. The only values not shown on a symmetric log scale are σ_{ε_i}.

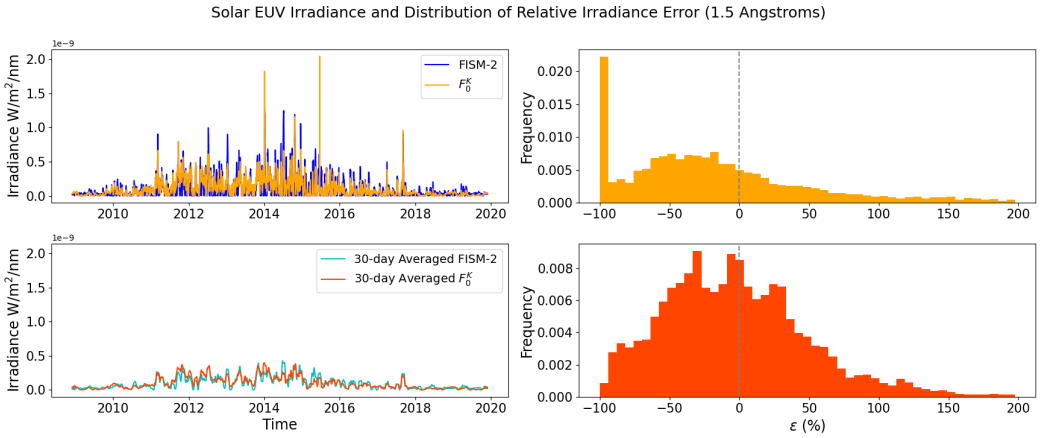

Figure 14: (Left - top and bottom) Solar EUV irradiance in the bin centered at 1.5 Åover SC24. The top shows native FISM2 outputs in blue and GAM outputs in orange. The bottom shows the a rolling 30-day average of the same. (Right - top and bottom) Histograms of the ε between F_0^K and FISM2. The top shows the original results before temporal averaging, and the bottom shows the results after a rolling centered 30-day average was applied to both FISM2 and F_0^K outputs. We note that histograms were constructed using 50 bins set within the widest range afforded by the 5th and 95th percentiles of both datasets.

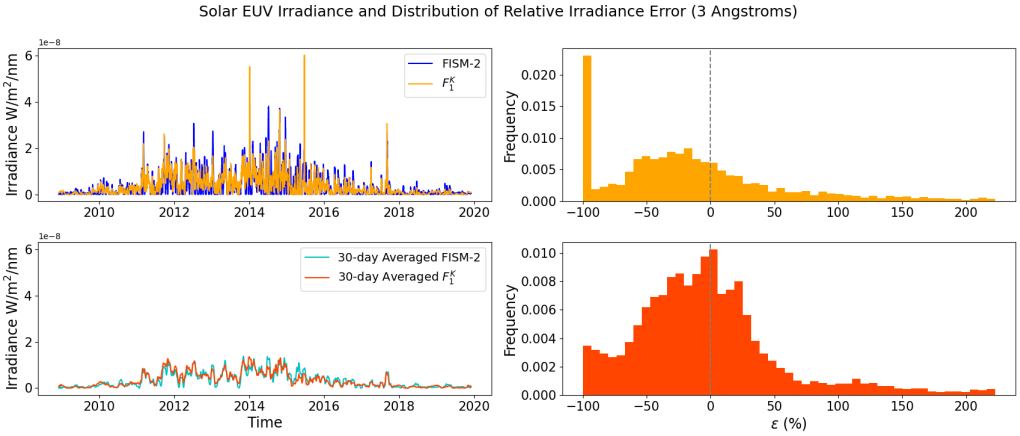

Figure 15: The same as Figure 14 but for EUV irradiance in the bin centered 3 Å.

In order to understanding the performance of the approach of the present work without recourse to temporal averaging, we also considered NRMSE values across bands. We first considered values of the rolling NRMSE computed in centered 24-hour windows (not shown). In the wavelength bands centered at 1.5 and 3.0 Å, we still observed occasional large spikes in NRMSE similar to that of ε, so we initially computed the median NRMSE value as a function of band in order to minimize the influence of outliers (Figure 16). We observe a behavior of the median values of NRMSE similar that of μ_{ε_i}, with values for the GAM approach lower than that of SDO/EVE in only wavelength band (centered at 275 Å) and lower than TIMED/SEE in all wavelength bands. Regarding the first two wavelength bands, we observe median NRMSE values of $\sim 59\%$ for both. Though these are the highest values observed for all wavelengths, we note that they remain below the

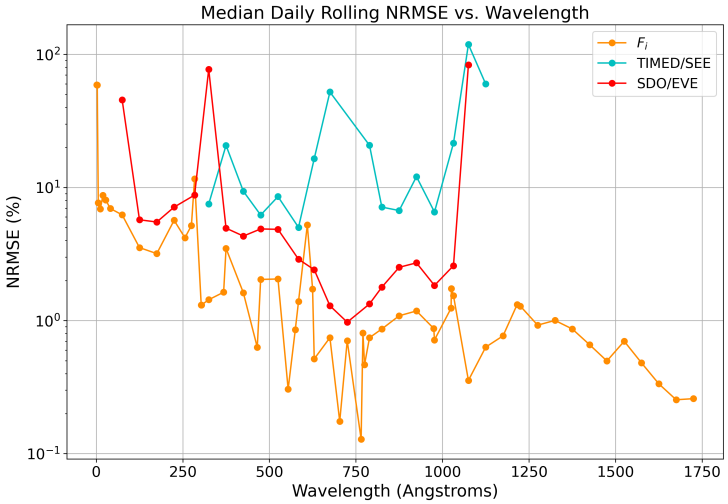

Figure 16: Median rolling daily NRMSE as a function of wavelength band, during SC24.

highest median values for SDO/EVE and TIMED/SEE, which at 1075 Å were $\sim 83\%$ and $\sim 199\%$, respectively.

Regarding μ_{ε_i}, we observe no values in excess of 10% for the GAMs F_i^K in any wavelength band. In comparison, TIMED/SEE and SDO/EVE show values of μ_{ε_i} in excess of in excess of 10% in 8 and 3 of their 16 and 21 wavelength bands, respectively. Only a single wavelength band centered at 475 Å, was μ_{ε_i} for TIMED/SEE lower than that of F_i^K, and then only by an absolute difference of $\varepsilon_i \sim 0.24$. In comparison, values of μ_{ε_i} for SDO/EVE were lower than that of F_i^K in 7 of the 21 considered bands, with the most appreciable performance exhibited by SDO/EVE at 125 Å, 175 Å, 225 Å, and 275 Å. We highlight that the values of μ_{ε_i} we observe are indicative of a systematic tendency of TIMED/SEE to overestimate FISM2, owing in part to continued degradation and reliance on calibrations corresponding to rocket measurements for which the most recent rocket flight was 2012. The daily calibrations performed for SDO/EVE partially contribute to its greater correspondence to FISM2 in several wavelength bands, such as that centered at 375 Å, as shown in Figure 17.

Results for σ_{ε_i} show that for F_i^K, there is a general downward trend as wavelength increases, and for 46 of the 59 wavelength bands considered ($\sim 78\%$) we observed values of σ_{ε_i} under 5%. These values show tight clustering of distributions of relative irradiance error for F_i^K that indicative of the most favorable performance for the GAM approach in particular above 250 Å. For TIMED/SEE, all values of σ_{ε_i} were in excess of 2%, whereas for SDO/EVE, values of σ_{ε_i} all were below those of TIMED/SEE and but in excess of F_i^K in all but 2 wavelength bands (75 Å and 1075 Å).

For κ_{ε_i}, for F_i^K, we observe a general decrease of kurtosis as wavelength increases, with distributions of ε_i having positive excess kurtosis (leptokurticity) in 34 of the 59 bands and negative excess kurtosis (platykurticity) in 25 of the 59 bands. By inspection, we observe a cutoff at $\sim$1000 Å below which leptokurticity dominates and above which we only observe platykurticity. Given that platykurtic distributions produce fewer and less extreme outliers than the normal distribution, we observe again that values of κ_{ε_i} show that the GAM approach is again more favorable as wavelength increases. For TIMED/SEE

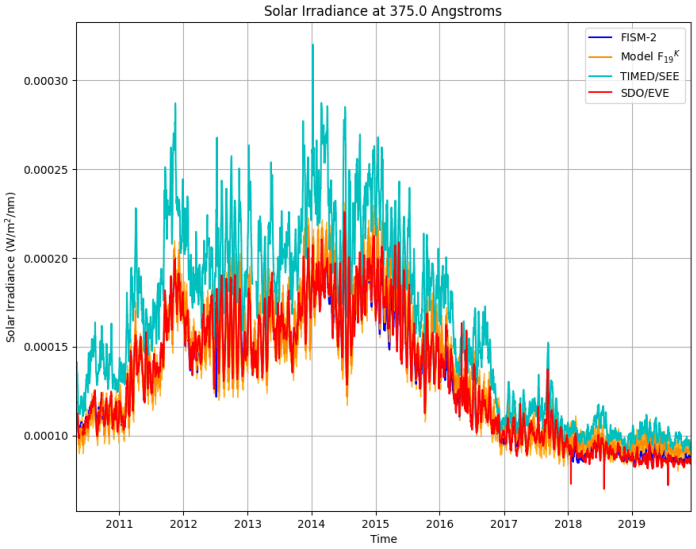

Figure 17: Time series of solar EUV irradiance centered at 375 Å, during a period of time corresponding to uninterrupted coverage provided by SDO/EVE during SC24.

and SDO/EVE, we observe platykurticity in only 1 and 3 of their respective considered bands.

Regarding skew, we observe variable skew for F_i^K that is primarily positive below $\sim$ 1050 Å, after which it is remarkably consistent in the vicinity of ~ 0.35. Positive skewness is observed in 39 of the 59 bands for F_i^K, with a trend that mirrors that of μ_{ε_i}. This indicates that over $\sim$ 66% of the solar EUV spectrum considered, relative irradiance error is most likely to deviate in the positive direction, a result indicative of a minimal but consistent tendency for the GAMs to overestimate FISM2.

3.4 Short-term Forecasting

The suitability for the GAM approach for forecasting was evaluated through hindcasts of integrated daily solar EUV irradiance. These hindcasts were performed during 30 days of low solar activity during the beginning of SC24 and 30 days high solar activity during the peak of SC24. As before, we assess this suitability first for solar EUV irradiance integrated across all 59 wavelength bands, followed by an evaluation of the behavior of ε as a function of wavelength.

In the case of integrated daily solar EUV irradiance, we observe that due to the behavior of the residuals in the hindcasted solar indices, values of ε are lower for low solar activity (average of -0.24%) than high solar activity (average of 1.68%) (Figure 18). The movement of the mean value of ε from negative to positive from low to high solar activity indicates the predilection of the GAM approach, as applied in the present work, to generally overestimate values of the solar irradiance during high solar activity when forecasting. We also observe a shift in the standard deviation σ_ε from -0.512% to 3.5%, corresponding to an increase by a percentage difference of $\sim$ 148.95%. In absolute terms, this change of P is nearly identical to that corresponding to the mean, which grew by a percentage difference of 150%. No absolute values of ε were observed to exceed 10%,

and were comparable to values resulting from successful companion techniques such as the Air Force Data Assimilative Photospheric Flux Transport (ADAPT), which relies on comprehensive estimates of the solar magnetic field distribution to derive estimates of EUV irradiance (Arge et al., 2010; Henney et al., 2015).

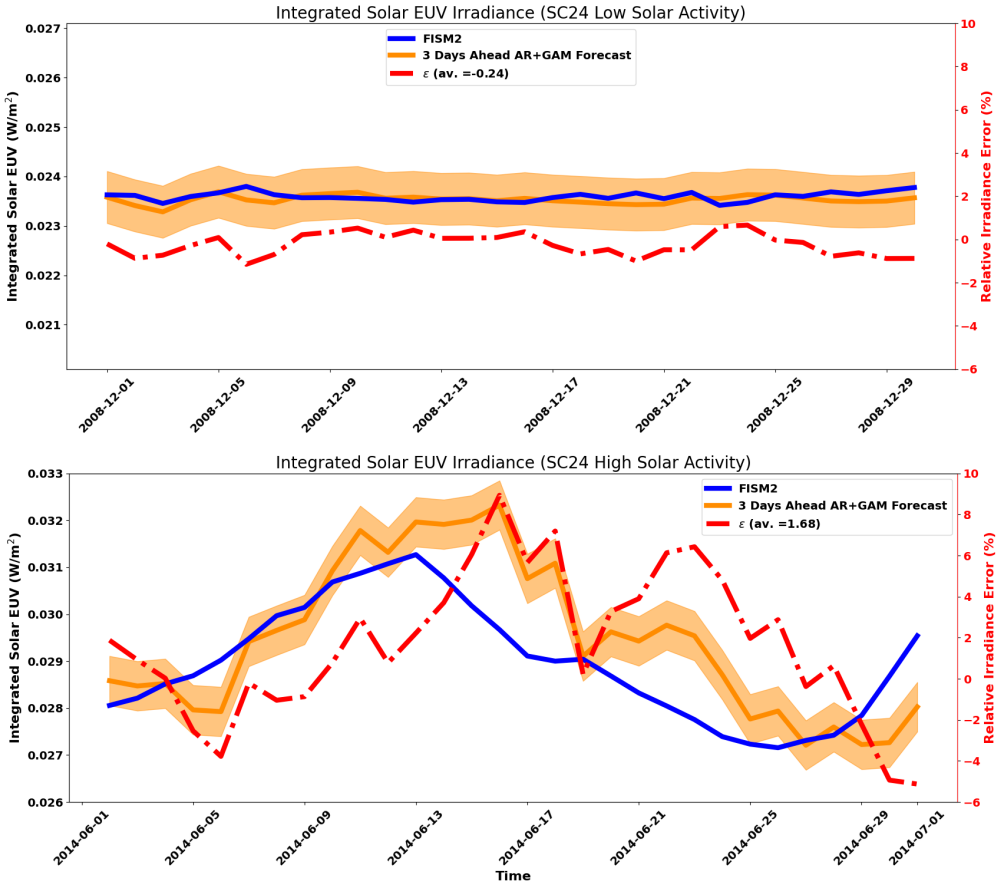

Figure 18: Time series of hindcasted integrated solar EUV irradiance during low solar activity (top) and high solar activity (bottom) during SC24. The light shaded orange region denotes the 95% Confidence Interval for the GAM results. The y-axes have been harmonized to show irradiance values that span 0.007 W/m^2 (left axis) and values of ε between -6 and 10% (right axis).

In closing, we consider values of ε as a function of wavelength band, for both low and high solar activity (Figures 19). We focus in particular on the quantities μ_{ε_i} and σ_{ε_i}, which show the most variability as wavelength increases. We note that values of μ_{ε_i} often show opposite sign from low to high solar activity. In particular, during low solar activity, values of μ_{ε_i} are most often negative below 375 Å (associated with higher values of σ_{ε_i}), they oscillate between the boundaries of $\pm 5\%$ between 375 and 1000 Å, while they remain negative but increasingly close to zero above 1000 Å. For high solar activity, values of σ_{ε_i} show the same decreasing trend as a function of increasing wavelength that is observed at low solar activity, but their baseline is notably higher, indicating a spread in errors indicative of lower accuracy during the peak of the solar cycle. Additionally, although the oscillatory behavior is again observed between 375 and 1000 Å,

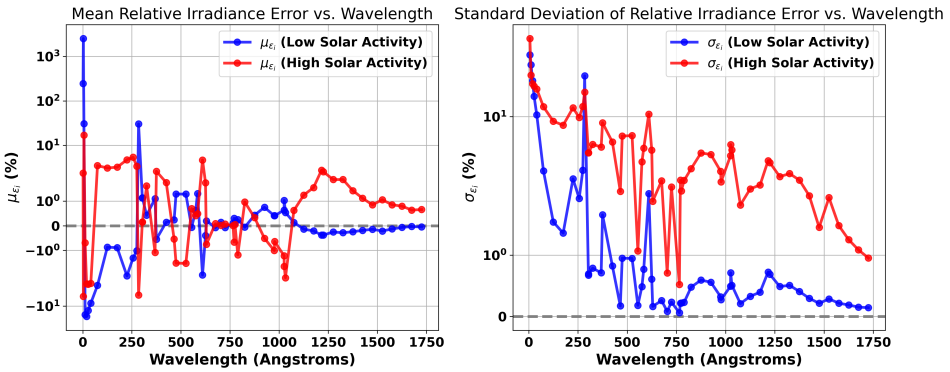

Figure 19: Mean (left) and standard deviation (right) of ε_i as a function of wavelength band, during low (blue) and high (red) solar activity during SC24, on a symmetric logarithmic scale.

the spread is slightly larger, extending up to ±10%, while values below 375 Å and above 1000 Å show systemic overestimation.

Overall, we observed the most favorable performance at low solar activity for wavelengths in excess of 1000 Å, while for wavelengths below 1000 Å, results indicate that forecasted values are on average likely to have absolute relative errors at most in the vicinity of 10% when forecasts are on the order of 3 days. This performance, however, becomes less reliable at wavelengths below 10 Å, where difficulty forecasting sharp declines in irradiance in the vicinity of zero can result in significant uncertainty. Overall, and coupled with the NRMSE results in Figure 16, we observe favorable results for the GAM approach in all wavelength bands that are comparable with and routinely exceed the accuracies of the measurements from TIMED/SEE and SDO/EVE, with forecasting errors that are comparable with companion approaches that rely on much more comprehensive information. This not only speaks to the operational suitability of the FISM2 estimates, but additionally to the fidelity of the GAM approach in retaining the statistical characteristics of the FISM2 estimates even when parameterized with a constrained set of solar drivers.

4 Conclusions and Discussion

When fitted appropriately, the GAM approach demonstrates itself as robust, statistically well-grounded, and accurate for representing solar irradiance in multiple wavelength bands. As shown for the case concerning integrated solar EUV irradiance, a robust GAM may be constructed between integrated solar EUV irradiance from FISM2 and only three solar indices with minimal sacrifice of statistical characteristics of estimated irradiance. This demonstrates the power GAMs for capturing non-linear behavior with limited drivers. This work also highlights the suitability of an approach leveraging FISM2 data for operational use, and highlights the degree to which it remains a powerful and versatile empirical paradigm for modeling of solar EUV. FISM2's capacities are inherent in its construction using three solar irradiance datasets (from SDO/EVE, SORCE/SOLSTICE, and SORCE/XPS), three solar proxies (F10.7, Mg-II cwr, and Lyman-α), and four additional solar proxies primarily from emission lines measured by SDO/EVE. When combined with an additional GAM for modeling residuals, we demonstrate appreciable performance over an entire solar cycle, as shown by absolute mean values of ε under 10% across the overwhelming majority of wavelength bands considered.

We additionally observe that combined with well-principled autoregressive model approaches for forecasting solar drivers, the GAM performs well in the context of short-term 3-day forecasts, with the resulting absolute forecast errors again regularly attaining values below 10% for both low and high solar activity, on par with companion techniques that rely on estimates of the solar magnetic field or utilize neural networks (Stevenson et al., 2022). While our approach has demonstrated suitability for nowcasting and forecasting solar EUV irradiance, it experiences some limitations, particularly in the context of heightened forecast errors during solar maximum and at wavelengths below 6 Å. We contend that these drawbacks are attributable to (1) the choice of solar drivers - improved performance may be achieved with the inclusion of other drivers such as Ca-II, Mg-II, S10, allowing for greater capturing the influence of solar chromospheric activity and solar active regions, which are particularly important during solar maximum, and (2) the fitting of ζ only during the descending phase of a single solar cycle. The latter struggles to capture variability of the residuals to a similar degree as Y capture the variability of integrated solar EUV irradiance due the usage of fewer samples. We contend that the fitting of ζ over multiple solar cycles would thus decrease kurtosis and scale of the resulting skew normal distribution of ε. Given that it has been determined that the mutual relationship between solar indices such as F10.7 and SSN show quasi-linearity that is dependent on the degree of temporal averaging (Clette, 2021), it is worth investigating how such averaging can lead to improved representation of solar EUV using the GAM approach in wavelengths below 6 Å that have shown difficult to model with a high degree of accuracy in the present work. In order to contextualize this investigation, we contend that it should be placed in the context of how the resulting irradiance estimates affect downstream ionospheric and thermospheric parameters in a coupled thermosphere-ionosphere model.

We emphasize that the principal power of this approach is its applicability for forecasting. GAMs constructed in various wavelength bands in the manner described in this paper enable forecasts of solar EUV irradiance directly from a reduced number of solar indices as drivers. With robust approaches to solar index forecasting, the GAM approach can be used to obtain much more comprehensive and accurate solar EUV forecasts for ingestion into thermospheric models, allowing for the reduction of thermospheric density errors. This is crucial especially for short-term forecasts that are needed to reduce the uncertainty of atmospheric density for satellite collision avoidance (Bussy-Virat et al., 2018).

Future work will involve the improvement of the GAM approach with the use of rigorous statistical methods such as Feature Ordering by Conditional Independence (FOCI) (Azadkia & Chatterjee, 2021), and extended application through principled medium and long-term solar driver forecasting for prediction of solar EUV irradiance on multiple timescales with quantified uncertainties.

5 Open Research

Solar indices F10.7, revised Sunspot Number, and Lyman-α were obtained via NASA OMNIWeb (`https://omniweb.gsfc.nasa.gov/form/dx1.html`). FISM2 and TIMED/SEE data were obtained via LISIRD, which is curated by the University of Colorado, Boulder (`https://lasp.colorado.edu/lisird/`). The current version of the dataset for this paper is available via Figshare at the following DOI: 10.6084/m9.figshare.24236581.v2. After publication, all of the processed data, and all code used to analyze the data will be found on Zenodo at the following DOI: 10.5281/zenodo.8250196.

Acknowledgments

This research was internally supported by Michigan Tech Research Institute (MTRI). We are grateful for the valuable statistical insight provided by others at MTRI, particularly Dr. Brian J. Thelen and Dr. Joseph W. Burns. We additionally are grateful for

insight into the FISM2 algorithm from Dr. Brandon Ponder, and we are most thankful for the valuable criticism provided by the two anonymous reviewers.

References

Arge, C. N., Henney, C. J., Koller, J., Compeau, C. R., Young, S., MacKenzie, D., ... Harvey, J. W. (2010). Air force data assimilative photospheric flux transport (adapt) model. In *Aip conference proceedings* (Vol. 1216, pp. 343–346).

Azadkia, M., & Chatterjee, S. (2021). A simple measure of conditional dependence. *The Annals of Statistics*, *49*(6), 3070–3102.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bussy-Virat, C. D., Ridley, A. J., & Getchius, J. W. (2018). Effects of uncertainties in the atmospheric density on the probability of collision between space objects. *Space Weather*, *16*(5), 519–537.

Cai, Y., Yue, X., Wang, W., Zhang, S.-R., Liu, H., Lin, D., ... others (2022). Altitude extension of the ncar-tiegcm (tiegcm-x) and evaluation. *Space Weather*, *20*(11), e2022SW003227.

Chamberlin, P. C., Eparvier, F. G., Knoer, V., Leise, H., Pankratz, A., Snow, M., ... Woods, T. N. (2020). The flare irradiance spectral model-version 2 (fism2). *Space Weather*, *18*(12), e2020SW002588.

Chamberlin, P. C., Woods, T. N., & Eparvier, F. G. (2007). Flare irradiance spectral model (fism): Daily component algorithms and results. *Space Weather*, *5*(7).

Chandorkar, M., Camporeale, E., & Wing, S. (2017). Probabilistic forecasting of the disturbance storm time index: An autoregressive gaussian process approach. *Space Weather*, *15*(8), 1004–1019.

Chattopadhyay, G., & Chattopadhyay, S. (2012). Monthly sunspot number time series analysis and its modeling through autoregressive artificial neural network. *The European Physical Journal Plus*, *127*, 1–8.

Clette, F. (2021). Is the f10. 7cm–sunspot number relation linear and stable? *Journal of Space Weather and Space Climate*, *11*, 2.

Clette, F., Cliver, E., Lefèvre, L., Svalgaard, L., & Vaquero, J. (2015). Revision of the sunspot number (s). *Space Weather*, *13*(9), 529–530.

de Wit, T. D., Kretzschmar, M., Aboudarham, J., Amblard, P.-O., Auchère, F., & Lilensten, J. (2008). Which solar euv indices are best for reconstructing the solar euv irradiance? *Advances in Space Research*, *42*(5), 903–911.

Du, Z. (2020). Forecasting the daily 10.7 cm solar radio flux using an autoregressive model. *Solar Physics*, *295*(9), 125.

Elvidge, S., Themens, D. R., Brown, M. K., & Donegan-Lawley, E. (2023). What to do when the f10. 7 goes out? *Space Weather*, *21*(4), e2022SW003392.

Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones Jr, M., Mlynczak, M., ... others (2021). Nrlmsis 2.0: A whole-atmosphere empirical model of temperature and neutral species densities. *Earth and Space Science*, *8*(3), e2020EA001321.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Gondelach, D. J., & Linares, R. (2021). Real-time thermospheric density estimation via radar and gps tracking data assimilation. *Space Weather*, *19*(4), e2020SW002620.

Guo, J., Wan, W., Forbes, J. M., Sutton, E., Nerem, R. S., Woods, T., ... Liu, L. (2007). Effects of solar variability on thermosphere density from champ accelerometer data. *Journal of Geophysical Research: Space Physics*, *112*(A10).

Hastie, T. J. (2017). Generalized additive models. In *Statistical models in s* (pp. 249–307). Routledge.

Henney, C., Hock, R., Schooley, A., Toussaint, W., White, S., & Arge, C. (2015). Forecasting solar extreme and far ultraviolet irradiance. *Space Weather*, *13*(3), 141–153.

Heroux, L., & Hinteregger, H. (1978). Aeronomical reference spectrum for solar uv below 2000 å. *Journal of Geophysical Research: Space Physics*, *83*(A11), 5305–5308.

Hocke, K. (2008). Oscillations of global mean tec. *Journal of Geophysical Research: Space Physics*, *113*(A4).

Johnson, J. R., Wing, S., O'ffill, C., & Neupane, B. (2023). Information horizon of solar active regions. *The Astrophysical Journal Letters*, *947*(1), L8.

Lean, J., Meier, R., Picone, J., & Emmert, J. (2011). Ionospheric total electron content: Global and hemispheric climatology. *Journal of Geophysical Research: Space Physics*, *116*(A10).

Lean, J., Picone, J., & Emmert, J. (2009). Quantitative forecasting of near-term solar activity and upper atmospheric density. *Journal of Geophysical Research: Space Physics*, *114*(A7).

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using stata* (Vol. 7). Stata press.

McFadden, D., et al. (1973). Conditional logit analysis of qualitative choice behavior.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Nusinov, A., & Katyushina, V. (1994). Lyman-alpha line intensity as a solar activity index in the far ultraviolet range. *Solar Physics*, *152*, 201–206.

Pawlowski, D. J., & Ridley, A. J. (2008). Modeling the thermospheric response to solar flares. *Journal of Geophysical Research: Space Physics*, *113*(A10).

Reber, C. A., Trevathan, C. E., McNeal, R. J., & Luther, M. R. (1993). The upper atmosphere research satellite (uars) mission. *Journal of Geophysical Research: Atmospheres*, *98*(D6), 10643–10647.

Richards, P., Fennelly, J., & Torr, D. (1994). Euvac: A solar euv flux model for aeronomic calculations. *Journal of Geophysical Research: Space Physics*, *99*(A5), 8981–8992.

Richards, P., Torr, M., & Torr, D. (1981). Solar euv energy budget of the thermosphere. *Advances in Space Research*, *1*(12), 53–61.

Richards, P. G., Woods, T. N., & Peterson, W. K. (2006). Heuvac: A new high resolution solar euv proxy model. *Advances in Space Research*, *37*(2), 315–322.

Ridley, A., Deng, Y., & Toth, G. (2006). The global ionosphere–thermosphere model. *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839–864.

Rottman, G. (2005). The sorce mission. *Solar Physics*, *230*, 7–25.

Sakaguchi, K., Nagatsuma, T., Reeves, G. D., & Spence, H. E. (2015). Prediction of mev electron fluxes throughout the outer radiation belt using multivariate autoregressive models. *Space Weather*, *13*(12), 853–867.

Schonfeld, S., White, S., Henney, C., Hock-Mysliwiec, R., & McAteer, R. (2019). The slowly varying corona. ii. the components of f10. 7 and their use in euv proxies. *The Astrophysical Journal*, *884*(2), 141.

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference.*

Servén, D., & Brummitt, C. (2018, mar). *pygam: Generalized additive models in python.* Retrieved from `https://doi.org/10.5281/zenodo.1208723` doi: 10.5281/zenodo.1208723

Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). Arima models. *Time series analysis and its applications: with R examples*, 75–163.

Si-Qing, L., Qiu-Zhen, Z., Jing, W., & Xian-Kang, D. (2010). Modeling research of the 27-day forecast of 10.7 cm solar radio flux (i). *Chinese Astronomy and Astrophysics*, *34*(3), 305–315.

Solanki, S., Seleznyov, A., & Krivova, N. (2003). Solar irradiance fluctuations on short timescales. In *Solar variability as an input to the earth's environment* (Vol. 535, pp. 285–288).

Stevenson, E., Rodriguez-Fernandez, V., Minisci, E., & Camacho, D. (2022). A deep learning approach to solar radio flux forecasting. *Acta Astronautica*, *193*, 595–606.

Stolarski, R., Hays, P., & Roble, R. (1975). Atmospheric heating by solar euv radiation. *Journal of Geophysical Research*, *80*(16), 2266–2276.

Tapping, K. (2013). The 10.7 cm solar radio flux (f10. 7). *Space weather*, *11*(7), 394–406.

Tapping, K., & Charrois, D. (1994). Limits to the accuracy of the 10.7 cm flux. *Solar physics*, *150*, 305–315.

Tariku, Y. A. (2019). Evaluation of the ssn and f10. 7 indices as a cause for performance variation of the iri 2016 and iri-plas 2017 models. *Advances in Space Research*, *64*(9), 1751–1763.

Thayer, J. P., Tobiska, W. K., Pilinski, M. D., & Sutton, E. K. (2021). Remaining issues in upper atmosphere satellite drag. *Space weather effects and applications*, 111–140.

Thiemann, E. M. B., Eparvier, F. A., Woodraska, D., Chamberlin, P. C., Machol, J., Eden, T., ... Woods, T. N. (2019). The goes-r euvs model for euv irradiance variability. *Journal of Space Weather and Space Climate*, *9*. doi: 10.1051/swsc/2019041

To, A. S., James, A. W., Bastian, T., van Driel-Gesztelyi, L., Long, D. M., Baker, D., ... Valori, G. (2023). Understanding the relationship between solar coronal abundances and f10. 7 cm radio emission. *The Astrophysical Journal*, *948*(2), 121.

Tobiska, W. K., Woods, T., Eparvier, F., Viereck, R., Floyd, L., Bouwer, D., ... White, O. (2000). The solar2000 empirical solar irradiance model and forecast tool. *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(14), 1233–1250.

Viereck, R. A., Floyd, L. E., Crane, P. C., Woods, T. N., Knapp, B. G., Rottman, G., ... DeLand, M. T. (2004). A composite mg ii index spanning from 1978 to 2003. *Space Weather*, *2*(10).

Wang, Z., Hu, Q., Zhong, Q., & Wang, Y. (2018). Linear multistep f10. 7 forecasting based on task correlation and heteroscedasticity. *Earth and Space Science*, *5*(12), 863–874.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Wilson, R. M., Rabin, D., & Moore, R. L. (1987). 10.7-cm solar radio flux and the magnetic complexity of active regions. *Solar physics*, *111*, 279–285.

Wood, S. N. (2017). *Generalized additive models: an introduction with r*. CRC press.

Woods, T. N., Bailey, S., Eparvier, F., Lawrence, G., Lean, J., McClintock, B., ... others (2000). Timed solar euv experiment. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, *25*(5-6), 393–396.

Woods, T. N., Tobiska, W. K., Rottman, G. J., & Worden, J. R. (2000). Improved solar lyman α irradiance modeling from 1947 through 1999 based on uars observations. *Journal of Geophysical Research: Space Physics*, *105*(A12), 27195–27215.