Near-Term Solar Proxy Forecasting for High-Precision Orbit Propagation Using Online Machine Learning

Fabio Chiappina
Aerospace Software Engineer, a.i. solutions, Inc
Dr. Katie Kosak
Radiation Engineer, a.i. solutions, Inc
John Smigo III
Aerospace Software Engineer, a.i. solutions, Inc

ABSTRACT

Solar activity strongly influences the Earth's atmosphere. Solar storms and associated phenomena, including unexpected solar flares and coronal mass ejections, can interact with the Earth's magnetic field and atmosphere, increasing the atmospheric density and therefore the drag experienced by spacecraft for several days. Accurately forecasting future solar activity is critical for safe and efficient space operations in that it minimizes uncertainty and potentially unnecessary fuel expenditure. This investigation presents an in-depth analysis of near-term forecasting models for solar proxies – observable metrics that vary according to the 11-year solar cycle and that serve as inputs to atmospheric density models used in orbit propagation and other applications. A variety of traditional and novel forecasting approaches are evaluated across thousands of forecasts spanning a complete solar cycle, thus allowing for a complete characterization of forecaster performance (i.e., error mean and variance) across a 60-day forecast horizon, with a particular emphasis on performance in the first 1-5 days. A software architecture for daily retrieval of new solar proxy observational data, re-training of forecasting models, and forecaster evaluation and selection is described. The performance improvement of machine learning based forecasters over traditional forecasting methods is highlighted. The daily ingestion of newly available data into such forecasters allows satellite operators to incorporate the latest solar activity into their mission planning, leading to safer and more efficient operations in space.

1. INTRODUCTION

1.1 The Solar Cycle

Solar activity varies according to an 11-year solar cycle due to the reversal of the global magnetic field of the Sun, which in turn is due to the solar dynamo. The theory describing the underlying mechanisms of the solar dynamo is an area of ongoing research [1]. The solar cycle is not symmetric, as evidenced by the variations seen in several observable metrics that vary according to the solar cycle, known collectively as solar proxies [2].

Solar proxies grant insight into the transient and long-term behavior of the Sun and are thus used in a variety of downstream applications where an understanding of solar activity is necessary. Though multiple solar proxies follow the solar cycle, not every solar proxy does so linearly, and the shape of their variations over the 11-year cycle period may differ in timing and amplitude; see Fig. 1. below for an example of these variations across two solar proxies, both of which are discussed in greater detail below. An understanding of the behavior of these solar proxies, the correlations between them, and the physical mechanisms underlying their variability is an important aspect of modern solar physics.

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

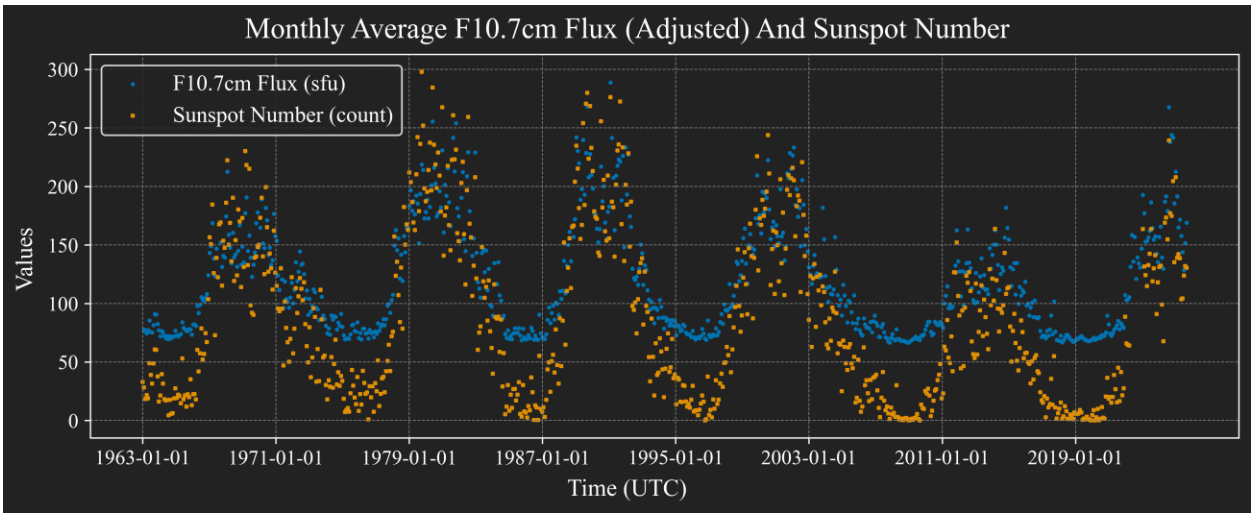


Fig. 1. Monthly averages for two solar proxies – $F_{10.7}$ and sunspot number – from 1963-present.

Some of the most commonly forecasted solar proxies include $F_{10.7}$, sunspot number, and Ap geomagnetic index. The $F_{10.7}$ index is a disc-integrated emission from the Sun high in the chromosphere and low in the corona at radio wavelength of 10.7 cm (2800 MHz) [3]. $F_{10.7}$ has been recorded daily by the Canadian Solar Radio Monitoring Programme since 1946. $F_{10.7}$ has a baseline of 67 solar flux units (sfu), where 1 sfu = $10^{-22} Wm^{-2}Hz^{-1}$. $F_{10.7}$ is a critical input to many atmospheric density models due to its strong (linear) relationship with atmospheric temperature. Bachmann and White showed that $F_{10.7}$ lags by one month behind another widely used solar proxy: sunspot number (SSN) [4]. Sunspots are defined as regions of the Sun with intense magnetic field strength [5]. The final solar proxy considered in this analysis is the Ap geomagnetic index, which has measured since 1932 the geomagnetic storm activity level with eight daily observations on a three-hour cadence.

In addition to the solar cycle, other periodic behaviors in solar activity exist. Table 1 lists the most common modes that are frequently referenced, but there are more modes beyond the scope of this discussion [5]. The 27-day period of the Sun is often observed in time-series data and associated with Active Regions moving along the Sun's equator. Starting with the SSN data from 1850-2013, and removing the 27-day rotation and solar cycle variation from the SSN, the signal can be smoothed with a Gaussian shape filter with Full Width at Half Maximum (FWHM) of 54 days, reducing all signals lower than 54 days. A second FWHM Gaussian filter of 24 rotations is applied to produce the remaining solar activity signal, showing quasi-periodic behavior. Rieger et al. noticed the 154-day periodicity of solar activity from observations on board Solar Maximum Mission (SMM) with the Gamma Ray Spectrometer (GRS) with gamma ray flare activity from 1980-1983 [6]. Bai and Cliver found the same periodicity in the same epoch, as well as from 1958-1971 in proton flares [7]. A two-year periodicity also occurs in the solar cycle [8]. At solar maximum, the two-year periodicity is seen as a double-peak feature in solar proxies and solar magnetic field strength [9].

Table 1. Common solar activity modes.

Mode	**Occurrence**
Solar Cycle	11 years
Quasi-Biennial Oscillation (QBO)	1-6 years
27 Day	If an Active Region rotates, 27 days.
P-Modes	5-7 minutes
G-modes	Exist, but not detected.

1.2 Motivation For Solar Proxy Forecasting

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

Towards solar maximum, solar activity, including the occurrence of solar flares and coronal mass ejections (CMEs), increases, as do EUV and X-ray solar emissions. Predicting the increase and decrease of solar activity with the solar cycle is important to space weather forecasting, spacecraft mission planning, and many other areas of the space industry. This is especially true for the more energetic events, such as solar flares, due to the damage they can cause. Studying the solar cycle allows for better forecasting of Solar Energetic Particles (SEPs), which is important to spacecraft. SEPs are particles emitted by the Sun that have near-relativistic energies and are associated with flare sites and shock waves from the Sun. The energy ranges from a few keV to GeV. One notable example of SEP impact on a spacecraft was from a X17.2 flare and a halo CME, when 5 SEPs penetrated SOHO, causing a malfunction to its CCD cameras. SEPs cause degradation to electronic and optical components, spacecraft surface charging, spacecraft surface erosion, and Single Event Effects (SEE). Examples of SEEs are data corruption, image sensor noise, and electronic faults, which can cause spacecraft system shutdowns.

Earth's atmosphere consists of layers of gases surrounding Earth, held by Earth's gravitational field. The Earth's atmosphere is directly influenced by the Sun's activity. Increased solar activity during a solar maximum increases atmospheric heating, causing increased atmospheric temperature and density in low orbits [10]. Decreased solar activity during a solar minimum has the opposite effect: the atmosphere cools and contracts. Since the 1950s, many models of Earth's atmosphere have been developed: the exponential atmospheric model, Jacchia 1977, Jacchia-Bowman 2008, etc. [11, 12]. Jacchia 1977, Jacchia-Bowman 2008, and other Jacchia variants are among the most common models used in the satellite industry because they are non-proprietary atmospheric models that consider both the solar cycle and geomagnetic effects.

While rudimentary atmospheric density models may approximate density with time-independent exponential decay, modern atmospheric models, including the widely used Jacchia models, incorporate solar proxies like $F_{10.7}$ as critical inputs. Improved forecasting of $F_{10.7}$ thus leads to improved atmospheric density modeling and, consequently, improved estimates of atmospheric drag force and trajectory state and covariance. Particularly in low earth orbit (LEO), the accuracy of state and covariance predictions is essential for applications like mission planning, radio frequency (RF) analysis, re-entry analysis, and conjunction assessment; indeed, in short-term orbit propagation for conjunction assessment, atmospheric density (ρ) uncertainty is the dominating factor of drag given by $a_{drag} = -\frac{1}{2}(C_D A/M)\,\rho v_r^2 e_v$ [13]. Even minor solar activity forecasting inaccuracies, and the resulting error in atmospheric drag models, can change the risk associated with a given conjunction, potentially causing concerning events to be mischaracterized, unnecessary and/or risky maneuvers to be performed, or inaction even in the presence of legitimate risk [13]. Given that the consequences of mischaracterizing even a single high-risk event can be catastrophic – i.e., a collision and the resulting pollution of the space environment with an innumerable influx of dangerous debris fragments – any improvements that can be made to solar activity forecasting accuracy could be immeasurably important to keeping space mission assets and the space environment as a whole safe.

This analysis focuses on $F_{10.7}$ forecasting due to its direct relationship with atmospheric density modeling. Additional forecasts are generated for sunspot number and Ap geomagnetic index but are not discussed in detail for the sake of brevity. The processes and forecasting models described are designed to be agnostic to any specific dataset; an extension of this analysis to any number of timeseries, forecasted individually or as a group, and on any time scale, is therefore possible.

1.3 Existing Solar Proxy Forecasting Models

The NOAA/SWPC (Space Weather Prediction Center) and Dr. Kenneth Schatten's solar cycle prediction models represent two approaches to forecast the $F_{10.7}$ Solar Cycle. The NOAA/SWPC model is primarily empirical, relying on statistical trends, historical sunspot numbers, and inputs from subject matter experts. It often assumes a predefined solar cycle shape, such as a modified Gaussian, and scales this shape based on observed data. The result is a conservative and operationally focused forecast that is updated regularly to reflect new observations. This approach is ideal for near-term, practical applications like space weather forecasting and satellite operations on a shorter timescale planning.

In contrast, Schatten's model is based off of solar physics, specifically using a surface flux transport (SFT) model and dynamo theory. It simulates the evolution of the Sun's surface magnetic fields, particularly the polar fields, which are strong predictors of the next solar cycle's amplitude [14]. This model then feeds those magnetic field

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

conditions into a dynamo model to predict the $F_{10.7}$.'s maximum. Because it relies on dynamo theory and the magnetic field, the predictions are able to predict on a timescale to decades.

Fig. 2. illustrates both the long-term periodic nature of $F_{10.7}$ variations over a solar cycle and Schatten model forecasts approximating this long-term behavior. Schatten forecasts require periodic (i.e., roughly every 6 months) re-fitting by a subject matter expert to the latest observational data; times when such re-fits occurred are indicated in the figure below by vertical dashed lines. Transitions from one forecast to another near the time of re-fit are slightly smoothed by Gaussian kernel smoothing with a standard deviation of 20 days.

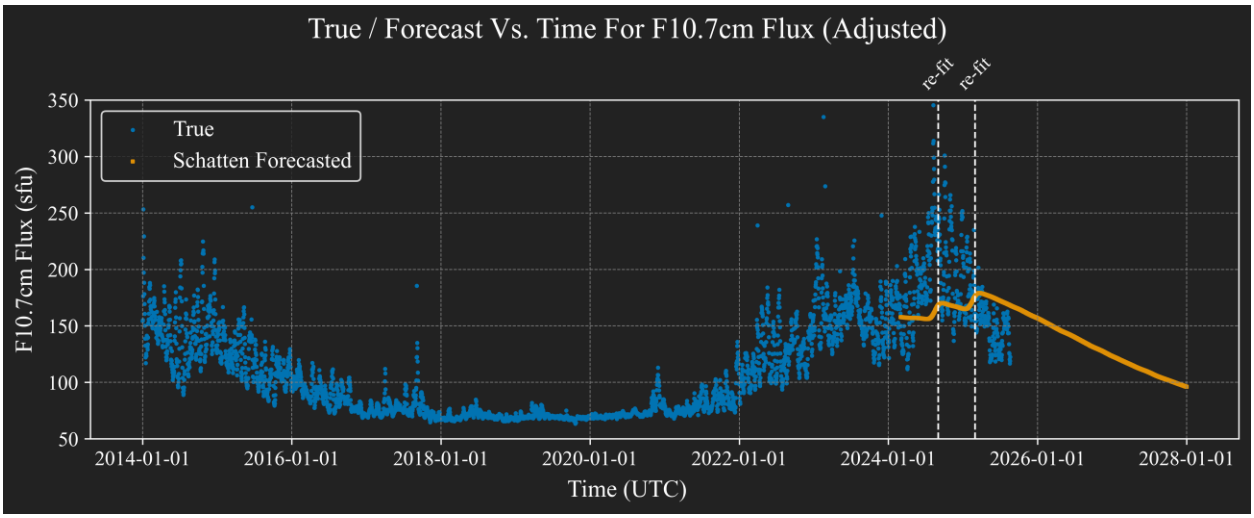


Fig. 2. Long-term periodic variations in $F_{10.7}$ over an 11-year period beginning in January of 2014 until present (August of 2025), along with Schatten $F_{10.7}$ forecasts beginning in March of 2024 and extending through 2027. Vertical dashed lines indicate transitions between Schatten forecasts that occur when the model's tuning parameters are manually re-fit by a subject matter expert.

Recent advancements in machine learning have facilitated new developments in solar proxy timeseries forecasting. Daniell and Mehta showed the potential of machine learning models to significantly improve upon linear auto-regressive baselines over 1-6 day forecast horizons and conducted an evaluation of a multitude of configuration hyperparameters and their impact on performance [15]. This investigation expands upon their analysis by characterizing forecaster performance over longer forecast horizons (60 days) and introducing additional flexibility in the selection of model features and hyperparameters.

2. MACHINE LEARNING FOR TIMESERIES FORECASTING

2.1 Regression

Regression analysis is the approximation of the relationship between one or more independent variables, often referred to as inputs or features, and one or more dependent variables, or outputs. This approximation problem can be formulated as follows: In general, outputs are continuous, real-valued variables, though predicting a discrete or categorical output variable – a process known as classification – can be achieved by similar methods.

In one common mathematical formulation of the regression problem, the relationship between dependent variables (**Y**) and independent variables (**X**) can be described as the sum of a function transformation on the independent variables (f(**X**)) and noise or other unmodeled behavior (**E**):

$$Y = f(X) + E \tag{1}$$

The goal of regression analysis is to determine a function f that most appropriately approximates the aforementioned relationship. Some regression models, like linear or polynomial regression, begin with an assumption that the

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

relationship between **X** and **Y** is governed by (or at least, can be modeled by) an underlying mathematical function; thus, these regression models can rewrite equation (1) such that *f* is parameterized by a set of unknown parameters that the regression problem attempts to estimate. Other regression models, like artificial neural networks, make no such assumptions, allowing the function *f* to take any arbitrary shape.

One application of regression is timeseries forecasting. A timeseries is a sequence of time-ordered data points and their associated times, and it may be univariate, where each data point contains only a single value, or multivariate, where each data point contains more than one value. Timeseries forecasting is the process of predicting future values in a timeseries using values observed in the past. A regression model can be utilized to extract meaningful patterns from historical timeseries values and then apply those learned patterns to predict future behavior. Using the mathematical formulation of regression presented in equation (1), timeseries forecasting is an application of regression that considers past timeseries values as the independent variables (**X**) and future timeseries values as the dependent variables (**Y**). Each solar proxy forecasted in this investigation is a timeseries – some univariate and others multivariate.

2.2 Linear Regression

Linear regression is a specific regression model that assumes that the relationship between **X** and **Y** is linear. Specifically, the dependent variable(s) can be written as the sum of the inner product of a weight matrix or vector with the independent variable(s), plus noise. The function *f* in equation (1) can be expressed for this case as follows, where **B** is a coefficient matrix (or, if the dependent variables can be expressed as a single column vector **y**, then a coefficient vector **β** is used instead):

$$f(X) = XB \tag{2}$$

Linear regression can be performed on univariate or multivariate timeseries. Extensions of this procedure involve extending the features such that weights on their nonlinear transformations (e.g., the square or cube of a single feature, the products between features, etc.) are also learned. While linear regression can be effective for certain regression tasks, it relies on the assumption of the linearity of the relationship between the independent and dependent variables. Since solar proxies are governed by complex and likely highly nonlinear processes occurring within the Sun, linear regression models are not expected to make effective solar proxy forecasters but still serve as a useful baseline against which to compare more powerful models.

2.3 Artificial Neural Networks

An artificial neural network is a regression model comprised of a set of interconnected units, or neurons. Each neuron receives inputs from other units, transforms those inputs through a function defined by learnable parameters, and computes an output through an additional, possibly nonlinear, function transformation. Neurons are generally organized into layers such that the output of one layer is dependent on inputs from any number of previous layers and can serve as the input to one or more later layers. A layer can thus be expressed as follows, where $\mathbf{h}_i$ is the output of layer *i, g* is a function specific to the layer type, and $\boldsymbol{\theta}_i$ is a set of learnable parameters for layer *i*.

$$h_i = g(h_0, h_1, \ \ldots h_{i-1}; \theta_i) \tag{3}$$

The output of the artificial neural network is thus the output of the final layer, which is itself a function of the outputs of each layer before it; thus, the function *f* in equation (1) can be expressed as follows:

$$f(X) = h_N = g_N(h_0, h_1, \ \ldots h_{N-1}; \theta_N) \tag{4}$$

Among the first implementations of an artificial neural network was Rosenblatt's perceptron, which established the foundation that would eventually evolve into the architectures of the modern day [16]. Rumelhart, Hinton, and Williams later introduced backpropagation, an algorithm for artificial neural network parameter adjustment that enables the learning of complex features [17]. Today, many variants of artificial neural networks exist. One of the

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

simplest and most widely used is the multi-layer perceptron (MLP), in which each layer's function g is a simple linear combination of the outputs from the previous layer, followed by a simple nonlinear activation function. First introduced by Hochreiter and Schmidhuber in 1997, a more sophisticated variant is the long short-term memory (LSTM) network, which is better optimized for learning patterns in sequential data by maintaining a hidden state that remembers long-term dependencies [18].

Artificial neural networks can learn extremely complex relationships that may not be obvious to a human eye, and they do not require any assumptions about the relationship between the independent and dependent variables. However, artificial neural networks do require examples – often called training data – of input and output pairs from which the relationship between the independent and dependent variables can be learned. In general, a larger network – that is, one with more learnable parameters – can learn more complex patterns than a smaller network but also tends to require more training data to learn those patterns. Large networks can also be prone to overfitting – a phenomenon where the network learns a function that fits the training data well but generalizes poorly to unseen data due to its excessive complexity. Since a network that is too small may be unable to learn complex enough patterns to effectively capture the relationship between the independent and dependent variables, the selection of network size and architecture can be a challenging balancing act that requires careful consideration and evaluation to ensure the resulting model will be effective when analyzing new inputs.

3. DATA ACQUISITION AND PROCESSING

3.1 Data Sources

Historical data for $F_{10.7}$, sunspot number, and Ap geomagnetic index are obtained from the OMNIWeb database maintained by NASA's Goddard Space Flight Center Space Physics Data Facility [19]. For daily updating datasets, the daily solar flux report from the Dominion Radio Astrophysical Observatory (DRAO) is utilized for $F_{10.7}$, and a daily updating webpage maintained by Jan Alvestad is utilized for sunspot number and Ap geomagnetic index [20, 21]. Data for all three proxies begins in January of 1963 and continues through present day. For simplicity, all proxies are sampled at a once-daily cadence; only the measurements at 20:00UTC are considered for $F_{10.7}$, and the 3-hourly Ap measurements are replaced with three daily features: minimum, maximum, and average Ap within each 24-hour period.

3.2 Data Segmentation for Training and Evaluation

The training dataset for the artificial neural network models considered in this analysis contains the majority, but not entirety, of all available historical data. Of the remaining data, a subset – often referred to as the validation data – is reserved to allow experimentation with model configurations and hyperparameters, where the models that perform best on this validation data are selected. A separate subset – often referred to as the test data – is reserved for final model evaluation; models do not learn from any data in this subset, and thus the performance of a model on this subset is the best possible characterization of how that model will perform when it attempts to make predictions on new, unseen, potentially not yet existing data. Training datasets often comprise 60-70 percent of the complete dataset, with validation and test datasets each accounting for 15-20 percent – though the exact split may vary according to the amount of available data and/or other characteristics of the specific task at hand; in this investigation, roughly 64% of the data is used for training, with roughly 18% reserved for each of validation and testing, as tabulated in Table 2. Although some baseline models have no learnable parameters and thus would not require this segmentation between training, validation, and test datasets if evaluated in isolation, they are still evaluated on an identical test dataset such that the performance of all models can be easily compared.

In many cases, each sample from the complete set of available data is randomly assigned to either the training, validation, or test dataset. However, in timeseries forecasting problems, this random assignment is generally avoided, since the temporal dependence of adjacent samples could result in the training dataset containing samples that are highly similar to samples in both the validation and test datasets, which can lead to artificially improved performance on those datasets and a failure to replicate similar performance on unseen data. Thus, for this analysis, three contiguous timespans are selected as the training, validation, and test datasets. The validation and test datasets each span 11 years so that the evaluation of our models captures their variance across all phases of the solar cycle.

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

The training dataset is the first of the three contiguous timespans so that model evaluations beginning at the start of the validation and test timespans can leverage past data as context; however, in an effort to capture any new trends in solar activity that may differ from the behavior of cycles long past, the model is additionally trained on what is referred to in this analysis as the online learning timespan, which contains data beginning in January of 2024.

Table 2. Training, validation, and test timespans used to train machine learning models.

Dataset Name	Start Date	End Date	Timespan (Days)
Training	01 Jan 1963	31 Dec 2001	14,245
Validation	01 Jan 2002	31 Dec 2012	4,018
Testing	01 Jan 2013	31 Dec 2023	4,017
Online Learning	01 Jan 2024	-	600+

3.3 Data Processing

A process known as imputation, where missing samples are filled with synthetic values to create a complete dataset, is performed on data from the three proxies considered in this analysis. The imputation process affects only days where data is missing or invalid; existing, nominal data is not smoothed or otherwise altered. Missing or invalid samples are replaced by the evaluation of a Gaussian kernel, where the values from the *k*-nearest samples are weighted according to their proximity in time to the missing sample. The number of nearest neighbors to utilize is computed as one half of the square root of the number of total samples, and the standard deviation of the Gaussian kernel is equal to the greatest distance in time from the missing sample from among the nearest neighbor samples. The resulting datasets each have exactly one sample per day.

4. FORECASTER ARCHITECTURES

4.1 Persistence Forecaster

Perhaps the simplest forecaster, all values predicted by the persistence forecaster predictions are equivalent to the latest observation, with that single, constant value “persisting” indefinitely through any forecast horizon. This forecaster is used as the primary baseline for this analysis. Depending on the variance of the timeseries under forecast, persistence forecasting can often perform well in the short-term. For timeseries like solar proxies with distinct long-term behavior of a greater amplitude than short-term variations, persistence forecasting performance is expected to diminish for longer forecast horizons. Persistence forecasters often benefit from making, on average, nearly unbiased forecasts, since their predictions for roughly periodic timeseries are generally just as likely to overpredict as they are to underpredict. The variance of these forecasters, on the other hand, is generally higher than more sophisticated methods and can increase substantially with forecast horizon.

4.2 Linear Regression Forecaster

A linear regression forecaster is also considered as an additional baseline in this analysis. For any individual forecast, the linear regression forecaster fits its coefficient matrix **B** (equation 2) using the *k* most recent observations before the forecast start time. For simplicity, only a small number of selections for *k* are analyzed in detail; the values 5, 27, and 54 are selected, where the former focuses on trends in the immediate two, and the latter two are selected to include observations from the previous one/two Carrington rotation(s). A separate linear regressor is fit on each solar proxy feature; cross-feature interactions and other nonlinear terms are not considered.

4.3 Schatten $F_{10.7}$ Forecaster

Though designed to forecast $F_{10.7}$ in the long term (i.e., on the order of years, or entire solar cycles), this analysis aims to characterize the performance of the Schatten $F_{10.7}$ forecaster in the short term, where output values are resampled from their original monthly cadence to a daily cadence to match the sampling rate of other forecasters. The Schatten model is manually and periodically re-fit to observational data by a subject matter expert. While the

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

model has historically been successful at predicting long-term solar activity, due to the relatively infrequent nature of the model updates (~2-4 times per year), this forecaster is not well-equipped to capture short-term variations; as such, all other forecasters considered in this analysis, including the other baselines, are generally expected to outperform the Schatten model. This forecaster is also unique among those in this investigation due to its specificity to solar physics data; while all other models considered can be applied to the generic timeseries forecasting problem, the Schatten forecaster is specifically fit to the shape of $F_{10.7}$, and thus its methodology cannot be generalized to other applications.

4.4 Multi-Timeseries to Multi-Timeseries LSTM Forecaster

The most flexible and powerful forecaster architecture considered in this analysis is an artificial neural network forecaster, the core advantage of which is its use of LSTM cells to learn both long- and short-term temporal dependencies. The bulk of this analysis consists of the performance evaluation of a multitude of combinations of hyperparameters in this forecaster architecture. The selection of lookback indices (see below), in particular, is influenced by consideration of the Carrington rotation and other known periodic solar activity behaviors (see section 1.2).

The LSTM forecaster can make predictions on any number of timeseries with any number of total features, and it can use as context past observational data both from the same timeseries it forecasts, and from additional timeseries that it does not directly forecast (e.g., synthetic features – see section 5.2). The network architecture is comprised of an input layer of shape equal to that of the model input matrices (discussed below), followed by one or more hidden LSTM layers, followed by one or more hidden dense layers, and finally followed by an output layer of shape equal to that of the model output matrices. The number of hidden LSTM and/or dense layers, as well as their size (i.e., number of neurons), are among the many tunable hyperparameters available to this model architecture. Other hyperparameters include the activation function between layers, the inclusion of layer normalization after each LSTM layer, the inclusion of batch normalization after each dense layer, the inclusion of dropout layers after each LSTM or dense layer, and recurrent and non-recurrent L1 and L2 regularization coefficients. For brevity, a detailed discussion of each of these parameters is omitted, but the final selections for the best performing configuration are presented in Table 3.

To effectively train an artificial neural network of this variety on the short-term solar proxy forecasting problem, the network must be presented with example input/output pairs from which to learn the temporal relationships it attempts to model. At each observation time, one such pair is assembled, comprised of an input sequence of one or more observations taken in the past or present relative to the observation time, as well as an output sequence of one or more future observations. Each sequence is of a predetermined and fixed length, though input sequences need not share their length with output sequences, and consecutive elements in the input and output sequences need not be consecutive in time.

Two key hyperparameters are referred to in this analysis as the lookback and lookahead indices. Lookback indices, $\boldsymbol{l_b}$, are a list of one or more nonpositive integers, always including 0, defining the indices relative to a given input/output pair of past observations to assemble into an input matrix. If the timeseries histories from all of the model's M constituent solar proxies, $\boldsymbol{p}_i$, each of which has an identical number of samples following the resampling process described in section 3.3, are concatenated into a matrix $\mathbf{P}$, then the following equations describe how that matrix of raw data is converted into a single input sequence at time t_i for use by the model in training:

$$P = [p_1\ p_2\ \dots\ p_M] \tag{5}$$

$$l_b = (b_1,\ b_2,\ \dots\ b_B) \tag{6}$$

$$I_i = P[i + l_b] \in \mathbb{R}^{B \times M} \tag{7}$$

Similarly, lookahead indices, $\boldsymbol{l_a}$, are a list of one or more positive integers defining the indices relative to a given input/output pair of past observations to assemble into an output matrix. Since these timeseries models need not output all proxies that they receive as input, instead forecasting on Q output proxies $\boldsymbol{p}_{oi}$ comprising a subset of the

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

input matrix **P**, to compute a single output sequence at time t_i, that subset matrix $\mathbf{P_o}$ is indexed at time i plus each of the positive integers in the lookahead indices:

$$P_o = \left[p_{o1}\, p_{o2} \cdots p_{oQ}\right] \subseteq P \tag{8}$$

$$l_a = (a_1,\, a_2,\, \ldots\, a_A) \tag{9}$$

$$O_i = P[i + l_a] \in \mathbb{R}^{AxQ} \tag{10}$$

With lookback and lookahead indices selected, the model then learns to map a particular set of past observations to a corresponding set of predicted future observations. However, depending on the length of the forecast desired, a single inferencing step in which the machine learning model is invoked with a set of sequence inputs may not be sufficient to produce predicted output values at each of the requested forecast times; that is, if the maximum lookahead index is H, but the forecast horizon is $H+1$ timesteps, then one pass through the model does fully generate the desired forecast. To remediate this, the outputs at the first lookahead index from one inferencing step can then be concatenated with the original input sequences, creating a combined sequence of with one additional row and containing both true and predicted values; this combined sequence can then be fed back into the model as a new input sequence, and the model can make a new prediction, advancing the forecast one additional step forward in time. By repeating this process of chaining output predictions together with true input data and performing additional inferencing steps, a forecast of arbitrary length can ultimately be generated.

Two methods are considered for combining the outputs of multiple inferencing steps together to produce a final forecast result. In the first of these, regardless of whether the lookahead indices have more than a single element, only predictions at the first lookahead index are represented in the output forecast; predictions at later lookahead indices are discarded entirely. In the second of these, the value at each timestep is computed as the average of the values from all inferencing steps that made a prediction at that timestep; in this way, models with more than a single lookahead index use predictions at later lookahead indices to attempt to smooth their output forecasts.

5. FORECASTER FEATURE SELECTION

5.1 Cross-Proxy Correlation

Since each of the solar proxies considered in this investigation are measurements of activity governed by the same or related underlying physical processes occurring within the Sun, distinct solar proxies are expected to be correlated with one another. These correlations suggest that a solar proxy forecasting model could potentially leverage relationships between solar proxies to improve its predictive accuracy. Forcing the model to make predictions that consider accuracy across multiple timeseries may assist in preventing the model from overfitting to noise within a particular timeseries. In addition, if two solar proxies have varying levels of responsiveness to a particular solar phenomenon, data within one proxy may provide new insight into that phenomenon that could not have been gleaned from the other proxy. However, if additional proxies are extremely noisy, prone to outliers, or poorly correlated with the other proxies alongside which they are predicted, their inclusion could result in reduced performance. Thus, this investigation considers both forecasters that predict a single solar proxy and forecasters that predict multiple solar proxies simultaneously, with the goal of characterizing the performance differences between the two. For simplicity, this analysis focuses on the effect of cross-proxy correlations provided by Ap and sunspot number on $F_{10.7}$ forecasting performance, though the methodology described can be applied to analyze any arbitrary timeseries and the effect on performance that results from the consideration of its correlations with other timeseries.

5.2 Synthetic Feature Augmentation

Though raw solar proxy observational data may well be sufficient for a forecasting model to extract meaningful patterns and make predictions with reasonable accuracy, additional known information about the problem space can supplement raw data to potentially improve predictive performance. The following timeseries – referred to hereafter as *synthetic* due to their ability to be readily computed for any past or future timestep without requiring observation – are considered:

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

5.2.1 Solar Cycle Phase

Due to the relatively well-understood period of the solar cycle, a potentially valuable feature with which to augment raw solar proxy measurements is the proximity of the current observation time to solar min/max. For maximal effectiveness, a feature encoding this "position" within the solar cycle should be both smooth and continuous; such properties ensure that two observations that are close together in time will map to close together feature values. This is most easily accomplished by generating a single timeseries with two features: the *sin* and *cos* of the solar cycle phase, computed at each measurement time t and defined in this analysis as θ below:

$$\theta(t) = \frac{2\pi\left(t - t_{previous}\right)}{t_{next} - t_{previous}} \tag{11}$$

In equation (11), $t_{previous}$ is defined as the observed time of the previous solar cycle minimum, and t_{next} is defined as the observed time of the next solar cycle minimum. For times in the current solar cycle, the next solar cycle minimum is predicted by a simple linear regression method that considers as input features the historical time deltas between two consecutive solar minima. Taking the *sin* and *cos* of this solar cycle phase quantity ensures that feature values are bounded on [0, 1] and that the aforementioned smoothness and continuity properties are satisfied. The inclusion of this feature is intended to supply models with context encoding a priori known information about the long-term periodicity of the solar cycle – context that may be both valuable in improving predictive accuracy and difficult to learn from raw observational data alone.

5.2.2 Schatten $F_{10.7}$ Forecasts

Though the Schatten $F_{10.7}$ forecaster is evaluated as its own, standalone forecaster in this investigation, its outputs double as additional input features that attempt to improve the predictive accuracy of short-term forecasting models. Since the Schatten forecaster is empirically validated and has baked into its calculations the expertise of the solar physicists responsible for its tuning, and since the model has historically been successful at predicting long-term solar cycle behavior, it stands to reason that the forecasted values it generates for $F_{10.7}$ approximate the true future behavior to a significant degree. These outputs are offset by an arbitrarily select thirty days into the future – that is, at any given observation time, an associated synthetic feature sample is generated, equal to the forecasted outputs from the Schatten model thirty days after that observation time. Each of these synthetic feature samples is a twelve-dimensional vector. Nine of those features represent forecasted $F_{10.7}$ at each combination of solar cycle timings and strengths from among early/nominal/late timing and low/mean/high strength. The remaining three features represent forecasted Ap geomagnetic index values at early/nominal/late solar cycle timings.

6. FORECASTER EVALUATION METHODOLOGY

6.1 Multi-Forecast Evaluation Framework

A primary objective of this investigation is to fully characterize the performance of forecasters throughout the entirety of a solar cycle, ensuring both the performance near solar min and solar max are accounted for. Understanding the advantages and limitations of different forecaster methodologies, including their performance as a function of forecast horizon, enables analysts to make better informed decisions (e.g., switching from a short-term optimized model to a long-term optimized model once the performance of the former begins to drop below that of the latter).

This evaluation is accomplished by initiating a new forecast starting at each day within the validation timespan, which spans one complete solar cycle. Each of these 4018 forecasts is executed for a forecast horizon of a maximum of 60 days (or fewer, if the forecast start date is fewer than 60 days from the validation timespan end date). The error of each forecast is then computed as the difference between the observed values and the forecasted values. All 4018 forecasts are then aggregated together into a single array, where, regardless of each forecast start time, all forecasts are aligned at the first array index and span a maximum of 60 indices. The mean and variance of the 4018 error arrays is then computed per-index, resulting in arrays for each solar proxy describing the mean and variance of forecast residuals at each forecast horizon timestep (i.e., at 1 day in the future, 2 days in the future, up to 60 days in the future). In this way, comparisons between error means/variances versus forecast horizon across multiple

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

forecaster architectures reveal both how well those forecasters track the true values they attempt to predict and how steeply their performances drop the further into the future they forecast.

In addition to the aggregate error mean and variance computation, a similar computation is performed with symmetric mean absolute percentage error (SMAPE), where SMAPE for a given timestep is computed for a pair of true ($\mathbf{y}_{Ti}$) and forecast ($\mathbf{y}_{Fi}$) values using a small epsilon (ε) as follows:

$$SMAPE_i = \frac{|y_F - y_T|}{|y_F| + |y_T| + \varepsilon} \tag{12}$$

SMAPE is then transformed into an alternative quantity, referred to here as *forecast accuracy*, as follows:

$$\alpha_i = 100(1 - SMAPE_i) \tag{13}$$

It should be noted that some formulations of equation (12) include additional scalar terms, and that the quantity in (13) does not refer to the concept of accuracy in the traditional sense, where it is most closely associated with classification rather than regression. Instead, this accuracy quantity assigns a score bounded by (0,100] to a forecast value regardless of the absolute scale of that value, making it a useful and interpretable quantity for easy comparison between forecaster architectures. Means and variances of accuracy versus forecast horizon are computed as they are for error.

After performing this evaluation on a multitude of forecaster configuration options using the validation timespan, this process is repeated on a subset of the best performing forecasters using the test timespan. Hyperparameters are not further tuned given examination of the results in the test timespan.

6.2 Per-Timeseries Forecaster Scoring

In an effort to assign a single numerical score to each forecaster, allowing for simplistic comparisons and a final designation of a single forecaster as the superior choice, a scoring metric that reduces the results of the multi-forecast evaluation process described in 6.1 is defined. The aforementioned SMAPE-based accuracy is chosen as the representative metric with which to compute overall forecaster scores due to its uniform bounding by (0,100], allowing scores across different timeseries types to remain easily comparable.

The reduction of a curve of accuracy versus forecast horizon to a single scalar value can be accomplished via an exponentially decaying weighted sum, where values at shorter forecast horizons are prioritized over values at longer forecast horizons such that performance over the first few days contributes most heavily to the overall score. This exponential weighting can be expressed as follows, where s is the accuracy score, α_i is the accuracy at forecast horizon timestep i, and r is a scalar in (0,1], fixed to 0.9 in this analysis:

$$s = \frac{\sum_{i=1}^{60} \alpha_i \, r^{i-1}}{\sum_{i=1}^{60} r^{i-1}} \tag{14}$$

A unique score is computed for each solar proxy feature. Scores are computed on the validation timespan for most models and on the test timespan for a smaller subset of the best performing among all validated models.

7. ONLINE MACHINE LEARNING SOFTWARE DESIGN

A software architecture is developed to facilitate the creation, evaluation, and selection of forecaster models that continue to learn from the latest solar proxy observational data as it becomes available. This architecture enables the following:

1. The automatic retrieval and processing of data for each solar proxy in consideration;

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

2. The organization of solar proxies into one or more groups, where each group is configured by a subject matter expert and represents solar proxies that are expected to be correlated, and thus should be forecasted together;
3. The initialization of a multitude of new models from a set of configured hyperparameter selections;
4. The training of forecaster models on training and online learning timespans, ensuring the incorporation of the latest available data;
5. The filtering of poorly performing forecasters by evaluation of their loss functions on the validation timespan;
6. The multi-forecast evaluation of the best performing forecasters, as described in section 6.1;
7. The scoring of forecasters according to the metric defined in section 6.2, with which a single best forecaster is selected; and
8. The execution of a new forecast from the current date until 14 days into the future, utilizing the best scoring forecaster.

Through the exploration of multiple forecaster architectures and configuration parameters, and the ingestion of the latest available solar proxy data, this software system ensures that the forecasts it ultimately produces are as accurate and informed of recent solar activity trends as possible.

It should be noted that in machine learning, online learning generally refers to incrementally updating a model as new data becomes available, avoiding retraining the complete model from scratch as in batch/offline learning. To avoid the model forgetting trends learned from the original dataset, several techniques exist, including subsampling a portion of the original dataset and combining with newly received data, then retraining on that combination. However, due to the relatively small size of solar proxy datasets used in this analysis, fully retraining the models on the entire dataset whenever new data becomes available is simpler and has a lesser chance of falling victim to catastrophic forgetting.

8. RESULTS

8.1 Multi-Forecast Evaluation Results

The multi-forecast evaluation and scoring process described in section 6.1 aims to (1) identify a forecaster that significantly improves upon the performance of baseline models and (2) illustrate the effects of cross-proxy correlations and synthetic feature augmentation, as described in section 5, on forecaster performance.

8.1.1 Comparison Across Forecaster Type

By experimenting with forecaster configuration hyperparameters and comparing their multi-forecast evaluation results against other models of the same type, each model with hyperparameters – in this analysis, the linear regression and multi-timeseries to multi-timeseries LSTM forecasters – is compared with other models of the same type in search of the best performing configuration. The tuning and selection of forecasters is performed using the validation timespan, and the selected forecasters are then evaluated on the test timespan. A comparison is then performed and illustrated in Fig. 3-6 below between the best performing model of each type, where LSTM (A) and Linear Regression Forecaster (A) are the representative forecasters selected from among all LSTMs and all linear regressors, respectively (see Table 3 for configuration hyperparameters).

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

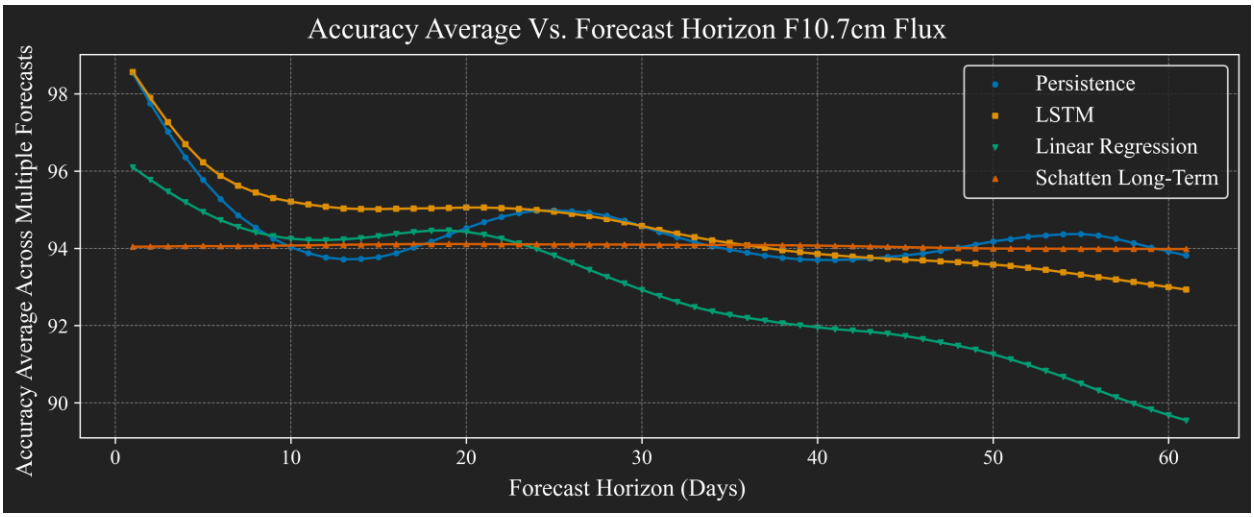


Fig. 3. Comparison of mean $F_{10.7}$ SMAPE accuracy score versus forecast horizon over 4017 forecasts in the test timespan for the best performing forecaster from each forecaster type.

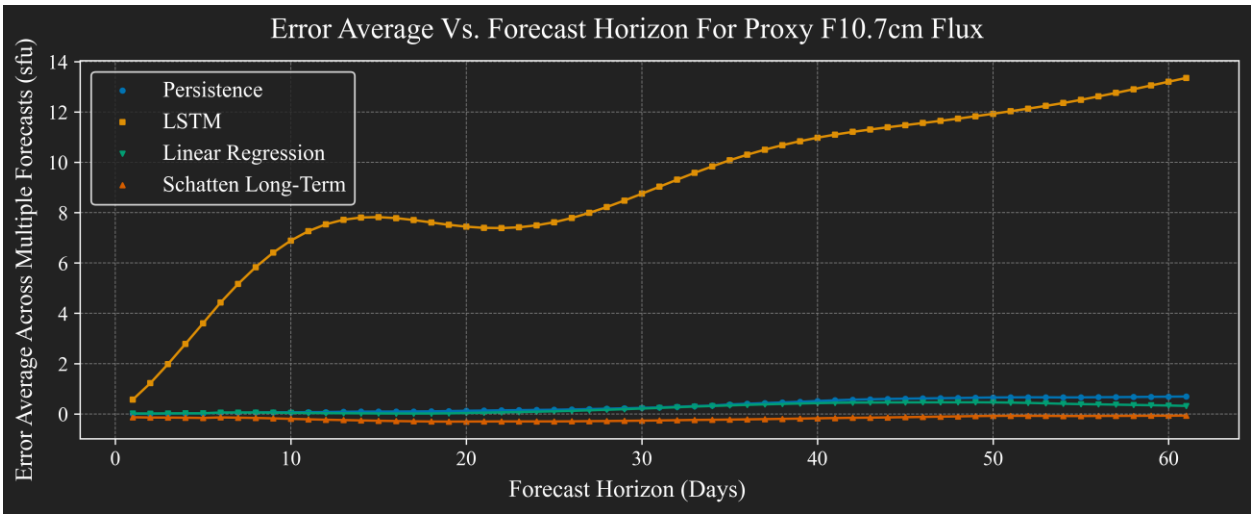
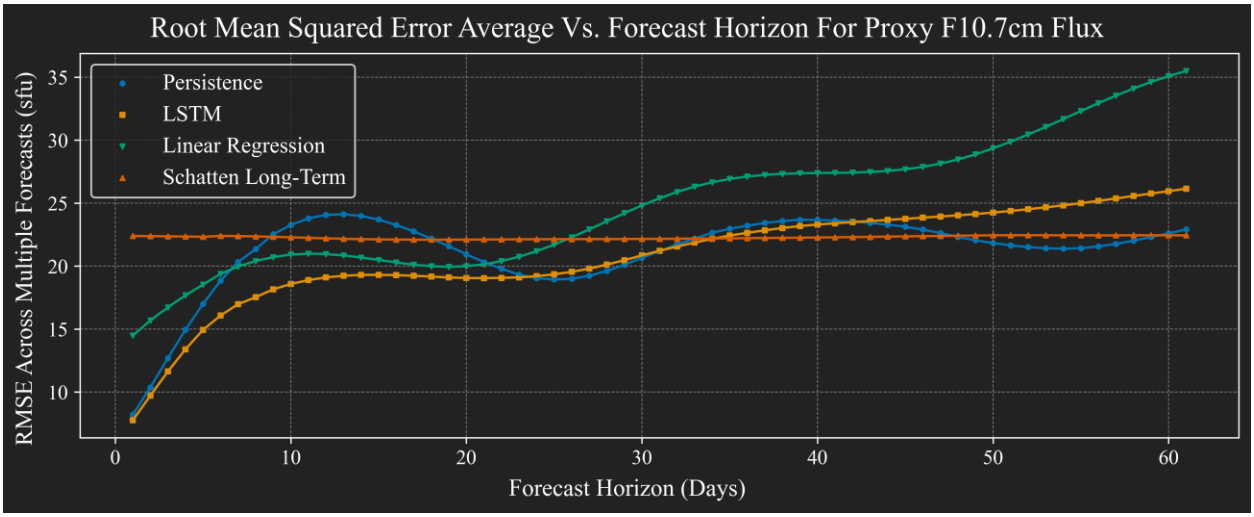


Fig. 4. Comparison of mean $F_{10.7}$ forecast error (solar flux units) versus forecast horizon over 4017 forecasts in the test timespan for the best performing forecaster from each forecaster type.



Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

Fig. 5. Comparison of $F_{10.7}$ forecast root mean squared error (solar flux units) versus forecast horizon over 4017 forecasts in the test timespan for the best performing forecaster from each forecaster type.

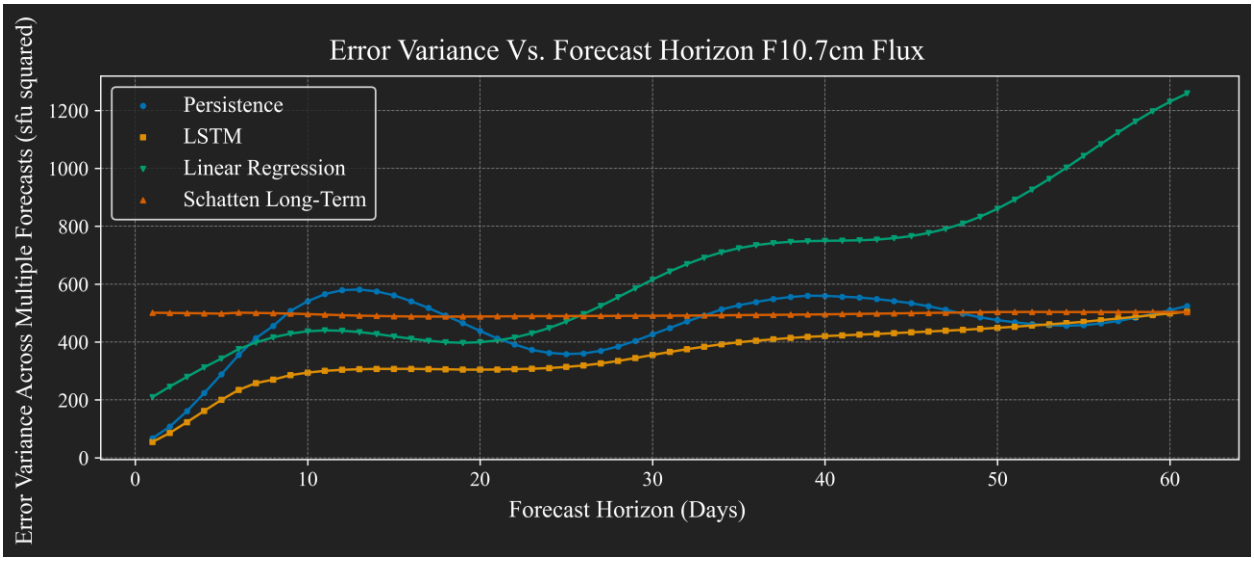


Fig. 6. Comparison of variance of $F_{10.7}$ forecast error (solar flux units squared) versus forecast horizon over 4017 forecasts in the test timespan for the best performing forecaster from each forecaster type.

Fig. 3. reveals several interesting phenomena. The LSTM forecaster (LSTM (A) in Table 3 below) demonstrates superior accuracy scores through a 35-day forecast horizon, after which the accuracy of the Schatten long-term forecaster, and later the persistence forecaster, surpass it. The persistence forecaster shows local maxima in performance near 27 and 54 days, revealing the effects of the short-to-medium-term periodic and quasi-periodic solar activity behaviors discussed in section 1.1. Despite the increased sophistication of the linear regression forecaster over the persistence forecaster, it only performs better over forecast horizons of 10-19 days, after which its performance begins to drop steeply compared to the other considered forecasters. Imbued with the historical shape of the solar cycle and the subject matter expertise of the engineers that tune it, the Schatten long-term forecaster retains nearly constant performance across the complete 60-day forecast horizon, but due to its inability to respond to transient changes or automatically update in the presence of new data, its performance in the immediate term is significantly lower than that of the other models.

Fig. 4. underscores what is in some ways an advantage of the persistence, linear regression, and Schatten long-term forecasters over the LSTM forecaster; while the latter is highly responsive to short-term variations, changing its output predictions significantly based on trends in the latest available data, the other three models issue forecasts that are much more static. In the relatively short-term, where the 11-year periodic effects of the solar cycle are not strongly pronounced, such nearly constant forecasts tend to overpredict as frequently as they underpredict, unlike the forecasts of the LSTM model, which is optimized over a complex objective function in a way that does not guarantee no resulting systemic bias.

Though the bias revealed in Fig. 4. appears to be a disadvantage of the LSTM model compared to traditional approaches, Fig. 5. illustrates that the metrics in the former figure do not paint a complete picture; while the average error remains near zero for the baseline models, their root mean squared errors (RMSEs) – obtained by squaring the error of each forecast, averaging across all forecasts in the test timespan, and taking the element-wise square root per forecast horizon – exceed that of the LSTM model for forecast horizons up to 22 days. At a horizon of 1 day, the LSTM forecaster boasts a 5.42% decrease (8.195 to 7.751 sfu) in RMSE over the next best (persistence) forecaster; at a horizon of 10 days, the LSTM forecaster has a 11.18% lower RMSE (20.911 to 18.574 sfu) than the next best (linear regression) forecaster. Fig. 6. additionally highlights that the error variance of the LSTM forecaster is the least among all considered forecaster across essentially the entire 60-day forecast horizon. At horizons of 1 and 10 days, respectively, the LSTM forecaster has lower error variances than the next-best baseline (persistence at 1-day, linear regression at 10) by 11.08% and 32.37%. Interestingly, the Schatten forecaster's error variance – just as all of

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

its other evaluation metrics – remains nearly flat over the complete considered forecast horizon. While the LSTM RMSE surpasses that of the Schatten forecaster at a 34-day forecast horizon, the LSTM variance does not surpass that of the Schatten forecaster throughout the complete 60-day forecast horizon.

8.1.2 Impact of Feature Selection and Hyperparameters

As discussed in sections 4 and 5, a multitude of forecaster architectures and feature enrichments are considered. The performance of a plethora of models is evaluated according to the process described in section 6, and a small subset of the best performing forecasters and their associated feature and hyperparameter selections is presented in Table 3 below.

Table 3. Evaluation results on the validation and test timespans of a variety of forecaster architectures with multiple combinations of feature augmentations/selections. Scores are SMAPE-based accuracy scores computed according to equation (14) and represent an evaluation solely of the forecasting performance on $F_{10.7}$.

Forecaster Name/Type	**Validation Score**	**Test Score**	**Feature Selections**	**Hyperparameter Selections**
LSTM (A)	96.23	96.14	$F_{10.7}$	1 hidden LSTM layer of size 64; 8 hidden dense layers of size 8; Lookback: [-155:-150, -145:5:-55, -54:-50, -45:5:-30, -27:-25, -20:0]; Lookahead: [1]
LSTM (B)	95.86	96.04	$F_{10.7}$, solar cycle phase	Identical to LSTM (A)
LSTM (C)	95.71	95.77	$F_{10.7}$, solar cycle phase	2 hidden LSTM layers of size 64; 3 hidden dense layers of size 64; LSTM layer normalization; Dropout, L1 and L2 regularization, and recurrent L1 and L2 regularization of 0.0001; Lookback: [-54, -53, -27, -26, -7:0]; Lookahead: [1, 2, 3]; Lookahead averaging
LSTM (D)	95.68	95.84	$F_{10.7}$	1 hidden LSTM layer of size 64; 3 hidden dense layers of size 32; LSTM layer normalization; L1 and L2 regularization, and recurrent L1 and L2 regularization of 0.0001; Lookback: [-100:5:-15, -14:0]; Lookahead: [1, 27]; Lookahead averaging
LSTM (E)	95.65	95.62	$F_{10.7}$, Schatten forecasts	Identical to LSTM (C)
LSTM (F)	95.54	95.87	$F_{10.7}$, solar cycle phase	Identical to LSTM (D)
Persistence	95.44	95.53	$F_{10.7}$	-
LSTM (G)	95.06	95.32	$F_{10.7}$, Ap, sunspot number	2 hidden LSTM layers of size 64; 3 hidden dense layers of size 48; LSTM layer normalization; Dropout of 0.001; L1 and L2 regularization, and recurrent L1 and L2 regularization of 0.0001; Lookback: [-81, -80, -54, -53, -27, -26, -5:0]; Lookahead: [1:14]; Lookahead averaging

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

Linear Regression (A)	94.57	94.73	$F_{10.7}$	k-nearest observations = 54
Linear Regression (B)	92.45	92.94	$F_{10.7}$	k-nearest observations = 27
Linear Regression (C)	89.71	89.54	$F_{10.7}$	k-nearest observations = 5
Schatten Long-Term	81.09	94.07	$F_{10.7}$, Ap	Nominal timing; Mean strength

The results in Table 3 reveal the interesting phenomenon that the inclusion of additional features – whether true observational data like measurements from additional related solar proxies, or synthetically generated features like solar cycle phase or Schatten forecasts – appears to bear relatively little weight on overall forecaster performance. In fact, the very best performing model (if ranked by score on the validation timespan) in this analysis included no additional features whatsoever – only raw observational data from the single solar proxy it attempted to forecast. Further research is necessary to understand why these additional features failed to improve the performance of the forecasting models, and whether an improved selection of features exist that could better correlate with $F_{10.7}$ variations and thus potentially succeed where these features failed.

Though a large number of forecaster architectures are evaluated, most of which are not present in Table 3 due to their inferior performance, several of the best-performing architectures appear to retain their success across different feature selections. LSTMs (A) and (B) in Table 3 share an architecture despite their differing feature selections; the same is true for LSTMs (C) and (E), and for LSTMs (D) and (F). The success of these architectures/hyperparameter selections across differing feature selections over dozens of alternatives suggests that the architecture selection may be more important for this analysis than the feature selection.

Note that the poor validation performance of the Schatten forecaster is likely attributable to having fewer re-fitted models available during that timespan than during the test timespan. For this reason, the score on the test timespan is the more useful metric for that model.

8.2 Single-Forecast Analysis

An example forecast executed within the test timespan is displayed in Fig. 7. and illustrates the ability of the best performing LSTM forecaster to capture short-term trends and produce an accurate forecast in the immediate term. Sudden changes preceded by little indication in the raw observational data, however, like the dramatic slope change that occurs on December 23rd, remain a limitation of this forecaster. Additional feature context from an external source (i.e., another correlated solar proxy dataset, or an additional carefully chosen synthetic feature) may be necessary to anticipate such large and seemingly random shifts.

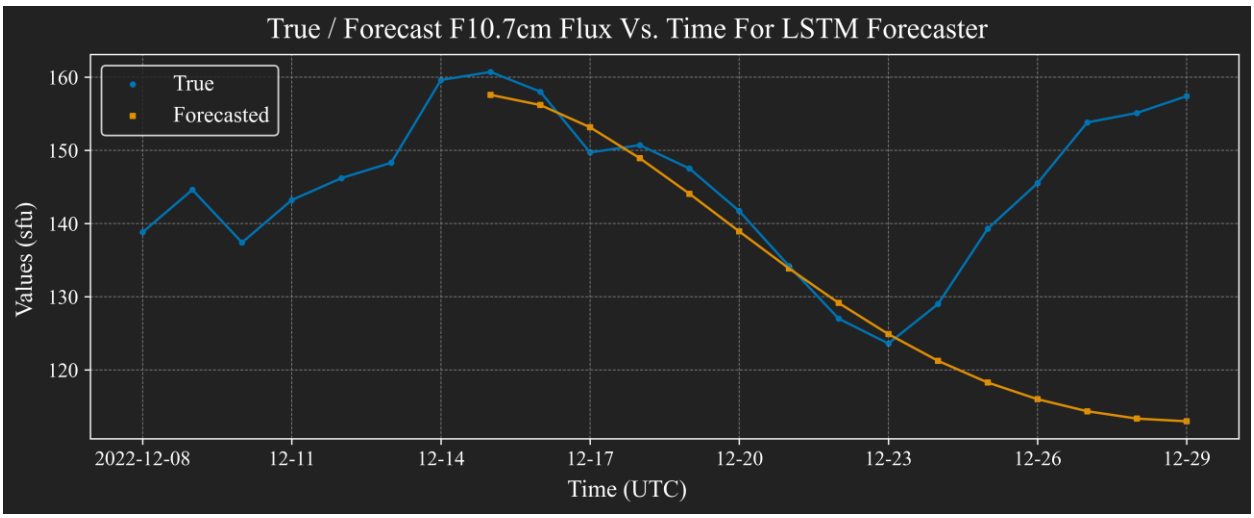


Fig. 7. True and forecasted values of $F_{10.7}$, where forecasted values begin on December 15th, 2025. The forecast is generated using model LSTM (A) (see Table 3).

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

However, regularly re-training the model on the newest available data and re-generating forecasts helps mitigate this limitation, as below in Fig. 8. When new solar proxy observations are received, the software architecture described in section 7 enables the rapid updating of the best performing forecasting model, allowing new predictions to incorporate the latest data and improve in accuracy. As Fig. 8. highlights, the forecaster is regularly re-trained on new observations, and forecasts are regenerated as observations arrive; for simplicity, though a new forecast could be generated each day (or at whatever cadence new data is received), only three forecasts are displayed. Though the second forecast yields lower accuracy forecasts than the first, the third and final forecast corrects significantly for the large upwards trend over the previous four days, underpredicting by a significantly lesser margin than the previous two forecasts. While the baseline models, like the persistence or linear regression forecaster, can also change their forecasts as new data becomes available, none does so as accurately or as consistently, and none takes into consideration the trends across sequences of observations. The LSTM forecaster thus demonstrates superior responsiveness over the baseline models to transient trends in the latest observational data – a significant advantage for operationally focused applications.

A similar forecast evolution analysis is performed in Fig. 9., where forecasts are generated on May 6th, May 9th, and May 12th of 2024, during the Gannon Storm – a period of significant solar and geomagnetic activity. Though the first forecast significantly underpredicts what would be a dramatic spike between May 6th –9th, the two subsequent forecasts offer substantial improvements.

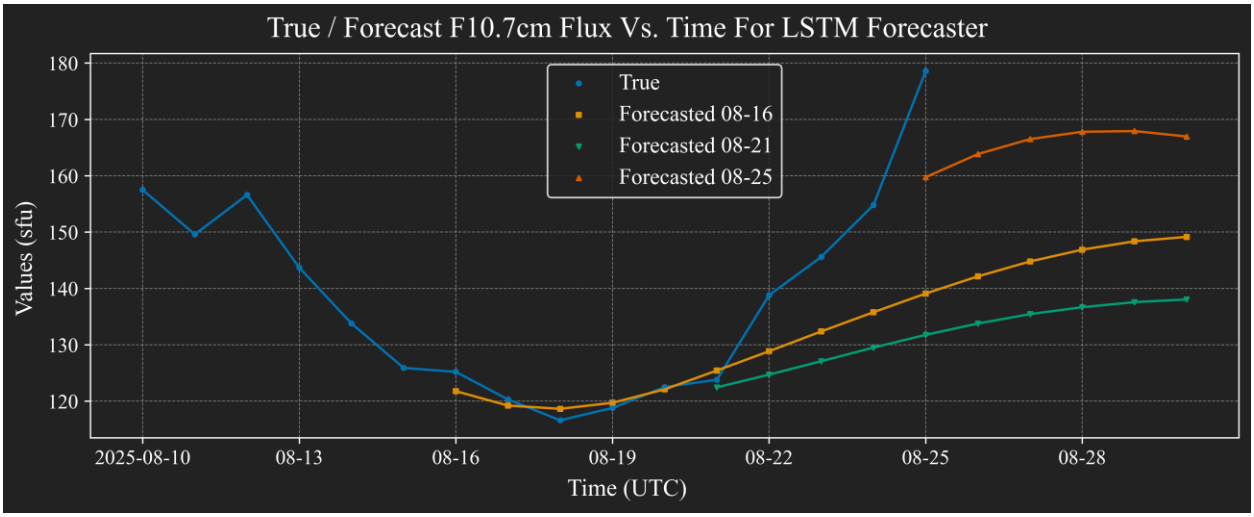


Fig. 8. True and forecasted values of $F_{10.7}$. A forecast is initially generated on August 16th, 2025, where the model is trained only on data available up until that date. The model is then re-trained on all available data up until August 21st, at which point a new forecast is generated. A final forecast is generated with a re-trained model on August 25th. All forecasts are generated using model LSTM (A) (see Table 3).

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

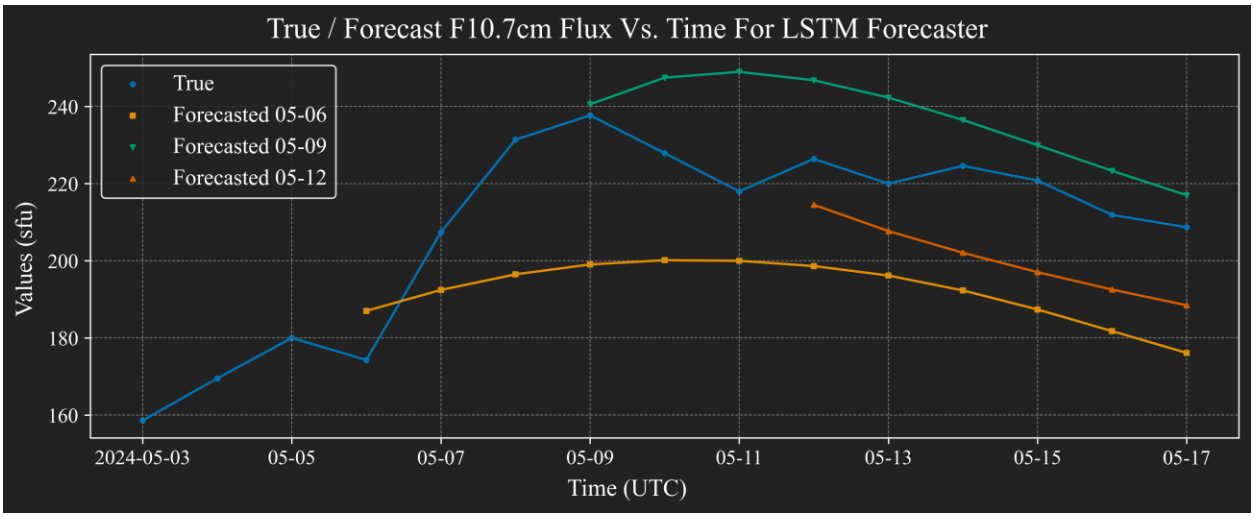


Fig. 9. True and forecasted values of $F_{10.7}$ during the Gannon Storm. A forecast is initially generated on May 6th, 2024, where the model is trained only on data available up until that date. The model is then re-trained on all available data up until May 9th, at which point a new forecast is generated. A final forecast is generated with a re-trained model on May 12th. All forecasts are generated using model LSTM (A) (see Table 3).

9. CONCLUSION

An analysis of various architectures for forecasting $F_{10.7}$ reveals that machine learning (LSTM-based) models improve upon baseline models in the near-term according to SMAPE-based accuracy scores on the validation timespan, and similarly high performance is retained on the test timespan, indicating the models generalize relatively well to previously unseen data. These near-term performance improvements are particularly relevant for conjunction assessment, where a clearer picture of solar activity over the next several days can help to reduce operational risks, and other applications with similar timescales and dependence on solar and atmospheric conditions. The benefits of making daily updates to forecasts generated by these models, facilitated by a software architecture for automatic data retrieval, processing, and retraining, are highlighted. Additional synthetic features and multivariate timeseries forecasting using additional correlated solar proxies failed to significantly alter forecaster performance. A complete characterization over the validation and test datasets of the forecast performance versus forecast horizon is presented, setting expectations for what level of certainty can be anticipated from these forecasters' predictions.

10. ACKNOWLEDGEMENTS

The authors would like to thank aiLabs, a.i. solutions' Internal Research & Development Program, for sponsoring this investigation. The authors also appreciate the continued feedback and invaluable insight received from Mark Miesch of NOAA SWPC and Anthony DeStefano and Jenna Martin of NASA Marshall.

11. REFERENCES

[1] P. Charbonneau. Dynamo Models of the Solar Cycle, *Living Reviews in Solar Physics*, 7(1):3, 2010. https://doi.org/10.12942/lrsp-2010-3
[2] M. Waldmeier. Neue Eigenschaften der Sonnenfleckenkurve, *Astronomische Mitteilungen der Eidgenössischen Sternwarte Zurich*, 14:105–136, 1935.
[3] K.F. Tapping and D.P. Charrois. Limits to the Accuracy of the 10.7-CENTIMETER Flux, Solar Physics, 150(1-2):305–315, 1994. https://doi.org/10.1007/BF00712892
[4] K.T. Bachmann and O.R. White. Observations of Hysteresis in Solar Cycle Variations among Seven Solar Activity Indicators. Solar Physics, 150(1-2):347–355, 1994. https://doi.org/10.1007/BF00712896
[5] K. Kosak. Seismology of Pseudomodes and Slow Magnetoacoustic Oscillations [Doctoral thesis, University of Warwick], 2023.

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com

[6] E. Rieger, G.H. Share, D.J. Forrest, G. Kanbach, C. Reppin, and E.L. Chupp. A 154-day periodicity in the occurrence of hard solar flares? *Nature*, 312:623-625, 1984. https://doi.org/10.1038/312623a0
[7] T. Bai and E.W. Cliver. A 154 Day Periodicity in the Occurrence Rate of Proton Flares. *The Astrophysical Journal*, 363:299, 1990. https://doi.org/10.1086/169342
[8] A. Johnston, B. Roberts, and A.N. Wright. High Frequency Waves and Chromospheric Magnetism. *Astronomical Society of the Pacific Conference Series,* 76:264, 1995.
[9] Y.M. Wang. On the Fluctuating Component of the Sun's Large-Scale Magnetic Field, *The Astrophysical Journal,* 590(2):1111, 2008. https:/doi.org/10.1086/375026
[10] T. Pulkkinen. Space Weather: Terrestrial Perspective, *Living Reviews in Solar Physics*, 4(1):1, 2007. https:/doi.org/10.12942/lrsp-2007-1
[11] L. G. Jacchia. Thermospheric temperature, density, and composition: New models, *Smithsonian Institution Astrophysical Observatory Special Report No. 375,* 1977.
[12] B. Bowman, F. Marcos, C. Huang, C. Lin, and W. Burke. A new empirical thermospheric density model jb2008 using new solar and geomagnetic indices. *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, 2008. https:/doi.org/10.2514/6.2008-6438.
[13] M.D. Hejduk and D.E. Snow. The effect of neutral density estimation errors on satellite conjunction serious event rates, *Space Weather*, 16, 849–869, 2018. https://doi.org/10.1029/2017SW001720
[14] K.H. Schatten and W.D. Pesnell. An Early Solar Dynamo Prediction: Cycle 23 ∼ Cycle 22, *Geophysical Research Letters*, 20(20), [2275–2278?], 1993. https://doi.org/10.1029/93GL02431
[15] J.D. Daniell and P.M. Mehta. Probabilistic Solar Proxy Forecasting with Neural Network Ensembles, *Space Weather,* 21(9), 2023. https://doi.org/10.1029/2023SW003675
[16] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington, D.C., 1962.
[17] D.E. Rumelart, G.E. Hinton, and R.J. WIlliams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. https://doi.org/10.1038/323533a0
[18] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735
[19] N.E. Papitashvili and J.H. King. OMNI Hourly Data [Data set], NASA Space Physics Data Facility, 2020. https://doi.org/10.48322/1shr-ht18. Accessed August 28, 2025.
[20] Natural Resources Canada. Space Weather Canada, Government of Canada. https://www.spaceweather.gc.ca/solar_flux_data/daily_flux_values/fluxtable.txt. Accessed August 28, 2025.
[21] J. Alvestad. Solar and geomagnetic data, Solen. https://solen.info/solar/indices.html. Accessed August 28, 2025.

Copyright © 2025 Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) – www.amostech.com