AGU PUBLICATIONS

Space Weather

RESEARCH ARTICLE
10.1002/2017SW001637

Special Section:
Low Earth Orbit Satellite Drag: Science and Operational Impact

Linear forecasting of the $F_{10.7}$ proxy for solar activity

Harry P. Warren[1], **John T. Emmert**[1], **and Nicholas A. Crump**[1]

[1] Space Science Division, Naval Research Laboratory, Washington, District of Columbia, USA

Abstract The ability to accurately forecast variations in the solar extreme ultraviolet irradiance is important to many aspects of operational space weather. For example, variations in the Sun's radiative output at these wavelengths drive changes in thermospheric density, which perturbs the trajectories of objects in low Earth orbit. Thus, predicting the conjunction of an operational satellite with orbital debris requires accurate forecasts of solar activity. In this paper we present a simple linear forecasting model for the 10.7 cm radio flux ($F_{10.7}$), a commonly used proxy for solar activity. Comparisons with simple reference models indicate that this linear model has positive skill for all forecast days that we have considered. We also examine the impact of the $F_{10.7}$ forecast skill on empirical model predictions of thermospheric density and ionospheric total electron content.

Key Points:
- A new forecasting model for the $F_{10.7}$ cm radio flux has been developed and tested
- Accurate forecasts of thermospheric conditions will require additional information

Correspondence to:
H. P. Warren,
harry.warren@nrl.navy.mil

Citation:
Warren, H. P., J. T. Emmert, and N. A. Crump (2017), Linear forecasting of the $F_{10.7}$ proxy for solar activity, *Space Weather*, *15*, 1039–1051, doi:10.1002/2017SW001637.

Received 31 MAR 2017
Accepted 30 JUN 2017
Accepted article online 7 JUL 2017
Published online 24 AUG 2017

1. Introduction

Forecasting is an essential aspect of space weather. There are many applications where conditions in the heliosphere need to be estimated in advance. Predicting conjunctions between operational satellites and orbital debris is one such example. Spacecraft debris avoidance maneuvers are not routine, and these predictions must be made several days in advance. For objects in low Earth orbit (below about 2000 km altitude) atmospheric drag produces a significant orbital perturbation. Atmospheric density at these heights is strongly modulated by the Sun's radiative output at extreme ultraviolet wavelengths (EUV, 5–120 nm), which can be highly variable. Further, *Emmert et al.* [2017a] have shown that uncertainties in the EUV irradiance lead to satellite orbit errors whose variance grow with time as t^5. Thus, predicting conjunctions requires very accurate forecasts of solar activity.

In this paper we present a forecasting model for the $F_{10.7}$ cm radio flux [e.g., *Tapping*, 1987, 2013], a proxy for the solar EUV irradiance that is commonly used as an input to models of the density in the Earth's thermosphere and exosphere, which corresponds to altitudes above about 90 km. On the time scales relevant to space weather, variations in the solar EUV irradiance are generally dominated by the 27 day rotation of individual active regions across the solar disk. We exploit this quasi-periodic nature of solar activity and use linear combinations of the previous 81 days of observations to forecast activity over the next 45 days. Comparisons between this model and two simpler reference models (persistence and climatology) yield positive skill scores for all forecast days, although the skill for forecasts longer than about 10 days is relatively small. The maximum skill is for 6 day forecasts, suggesting that this model may be able to contribute to improved conjunction predictions.

The forecasting of proxies for solar activity has been considered previously by *Lean et al.* [2009]. That study used a linear autoregressive method, where 1 day forecasts are advanced recursively. That is, the previous n days of data are used to forecast the proxy for the next day, which is then incorporated into the forecast for the next day, and so on. We take a different approach and compute independent models optimized for each forecast day. We find that this generally produces smaller forecast errors for $F_{10.7}$ than the autoregressive approach. *Tobiska et al.* [2008] considered forecasts of 1 to 4 days using an approach similar to ours and appears to achieve similar performance. The United States Air Force also makes daily 45 day forecasts of the $F_{10.7}$ proxy for solar activity.

More recently, the forecasting of $F_{10.7}$ using a surface magnetic flux transport simulation was studied by *Henney et al.* [2012]. Variations in the radiance of individual emission lines are well correlated with the total unsigned magnetic flux [e.g., *Schrijver*, 1987; *Pevtsov et al.*, 2003], so it is not surprising that the magnetic flux can be used as a proxy for solar activity. Furthermore, the evolution of surface magnetic fields on the Sun

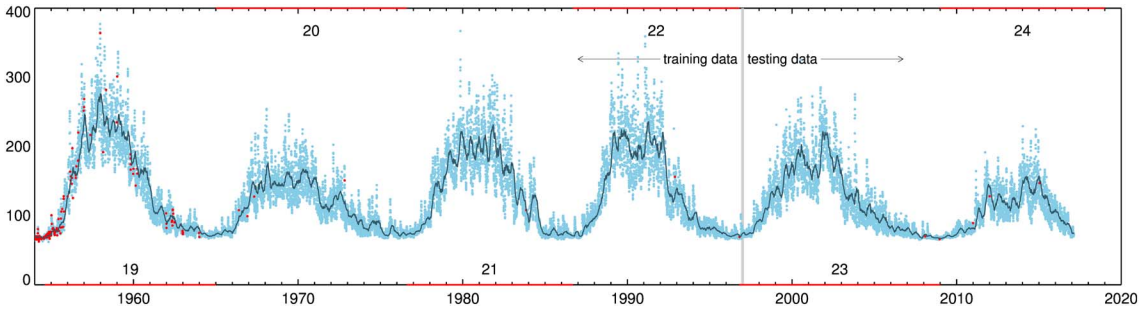

Figure 1. The $F_{10.7}$ cm solar radio flux over the past six solar cycles. Daily values are shown in light blue, and the 81 day running mean is shown in black. Interpolated values are indicated in red. Data before 1 January 1997 are used in the training set; data after this date are used for testing and evaluation.

has been studied extensively and is generally well understood [e.g., *Wang et al.*, 1989]. We find, however, that the linear forecasting model developed in this paper slightly outperforms the flux transport model used by *Henney et al.* [2012].

Longer-term, one potential advantage to flux transport modeling is its ability to incorporate information on active regions that have emerged on the far side of the Sun or are predicted to emerge on the near side. Presently, magnetic fields are observed only from instruments on or near the Earth. This means that half of all regions emerge unseen on the far side of the Sun and cannot be accounted for in any model that uses retrospective data. *Ugarte-Urra et al.* [2015] have used EUV images from the Solar Terrestrial Relations Observatory mission [*Howard et al.*, 2008], which has two spacecraft in orbit around the Sun, to infer the magnetic properties of newly emerged regions and shown that their subsequent evolution can be accurately modeled with flux transport. Similar results on detecting far side solar activity may be possible using ground-based helioseismology [e.g., *Lindsey and Braun*, 2000]. Helioseismology may also be able to predict the emergence of large solar active regions on the near side of the Sun in advance [e.g., *Ilonidis et al.*, 2011; *Barnes et al.*, 2014].

In this paper we also briefly examine the effect that an $F_{10.7}$ forecast with improved skill has on model predictions of thermospheric density and ionospheric total electron content. We find that other sources of error, including unmodeled atmospheric meteorology and errors in $F_{10.7}$'s representation of thermospheric UV heating, contribute significantly to the modeling of the terrestrial upper atmosphere, similar to the conclusions of *Lean et al.* [2009]. Nonetheless, improved $F_{10.7}$ forecasts can substantially improve thermospheric density and ionospheric total electron content (TEC) forecasts, especially when the terrestrial model predictions are adjusted to account for temporally local differences between modeled and observed values. We estimate this adjustment with the same forecast techniques used for the $F_{10.7}$ forecasts.

2. The Linear Forecasting Model

To develop this forecasting model, we consider observations of the $F_{10.7}$ cm radio flux from 1 January 1954 to 16 February 2017, which are shown in Figure 1. $F_{10.7}$ observations actually begin in 1947, but there are frequent data dropouts for these early years. Solar cycle 19 is one of the most active on record, and the data availability is generally good beginning in 1954, so we chose to begin at that time.

$F_{10.7}$ observations prior to 2017 were downloaded from NOAA, while data after that date were downloaded from Natural Resources Canada. Of the 23,058 days in this interval there are 169 days without an observation. To avoid the complexity of dealing with missing data, we perform a cubic spline interpolation to estimate values for these dates. The amount of missing data is small relative to the actual data values (0.7%), and it seems unlikely that this interpolation would play any role in the formulation or evaluation of the model.

To forecast $F_{10.7}$ at future times, we use a simple linear combination of previous observations:

$$f_{+n} = c^n + a_0^n f_0 + a_{-1}^n f_{-1} + a_{-2}^n f_{-2} + \cdots, \tag{1}$$

where f_0 is the most recent available observation, f_{-1} is the observation before that, and so on up to 81 days. Here n is the number of days in the future for the forecast (herein called the "forecast day"). We consider forecasts of 1 to 45 days. The coefficients are optimized by considering a large number of sets of "past" and "future" observations. Note that we compute an independent set of coefficients for each forecast day.

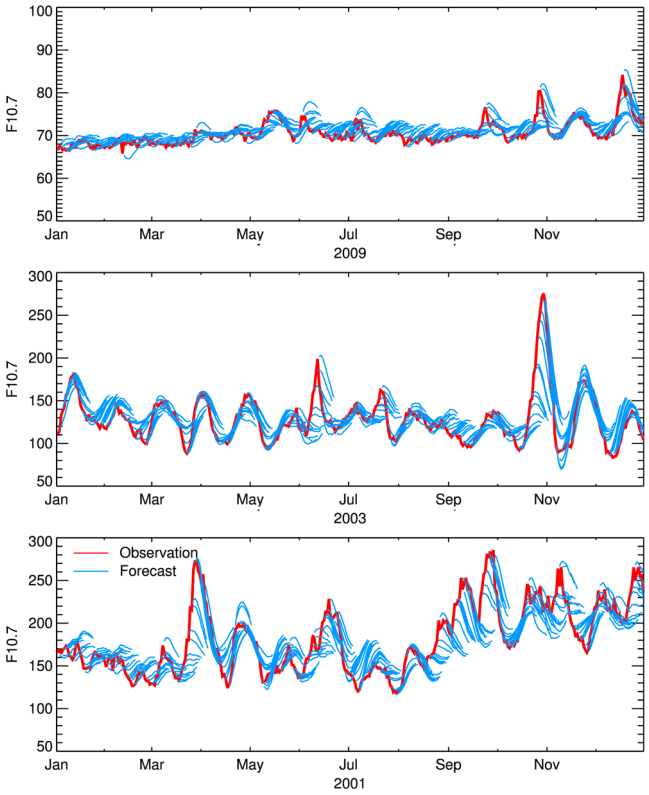

Figure 2. Examples of 1 to 10 day forecasts for three levels of solar activity: low (2009), moderate (2003), and high (2001). For each plot the observations for the given year are shown (red lines) along with the 10 day forecast made on each day (blue lines). Forecasts made during periods of low and moderate activity are generally more accurate than forecasts made during periods of high activity, when new active regions routinely emerge during the forecast period. Here low, moderate, and high refer to general levels of solar activity. Later in the paper we will quantify the model performance for periods where $F_{10.7}$ is above and below 100. In this, and all subsequent plots unless otherwise noted, the data displayed are from the testing period and have not been used to optimize the model parameters.

Thus, this method differs from autoregression, which recursively computes 1 day forecasts to arrive at forecasts further in the future.

To compute the model coefficients, only observations between 1 January 1954 and 31 December 1996 are used. To evaluate the model performance (e.g., computing RMS errors, correlation coefficients, or forecast skill) only data after 1 January 1997 are used. The segregation of the data into independent training (68%) and testing (32%) sets prevents overfitting the data. Making this split between solar cycles instead of simply randomizing all of the data further reduces the correlation between the training and test data sets. For example, a 10 day forecast for a given day is likely to be very similar to the 10 day forecast for the previous day. Note that there is actually a small amount of overlap in the training and testing data sets. These dates correspond to the date on which the forecast is being made. Forecasts near the end of the training period require data points from the testing period to evaluate equation (1). Also note that since we have a finite amount of data there are some differences in the number of forecasts available for each forecast day. To simplify the subsequent analysis, we consider only those days for which all 45 forecast days are available.

Finally, we note that we experimented with using different amounts of past data in the forecasting model. We found that models with 27 to 81 days of past data preformed similarly.

Figure 2 shows some of the forecasts made using the model. Here we display the actual observations as a function of time for three representative levels of solar activity. Also shown are the 1 to 10 day forecasts for each date. As one would expect, high levels of activity, such as were observed in 2001 or late 2003, are difficult to forecast with retrospective observations, since active regions are constantly emerging. Inspection of these plots suggests that the forecasts are relatively more accurate during periods of moderate or low activity. These plots also suggest that the declining phase of a solar rotation is generally well reproduced by the linear model.

Figure 3 systematically illustrates the differences between the observations and the model for several forecast horizons for 2003, a period showing both moderate and high levels of solar activity. Here we define the residual for a given pair of forecast and observation as $100 \times (\textit{forecast} - \textit{observation})/\textit{observation}$. As one would expect, the differences between the forecast and the observations tend to grow with increasing forecast time. Differences of greater than 10% are very rare for 1 day forecasts, for example, but become relatively common for forecasts of 7 days or longer. The period at the end of 2003, when several very large active regions emerged at about the same time, is an example of a particularly poor forecast with errors in excess of 50% at 10 days or more. This is essentially an upper bound to the EUV forecast errors at this time horizon.

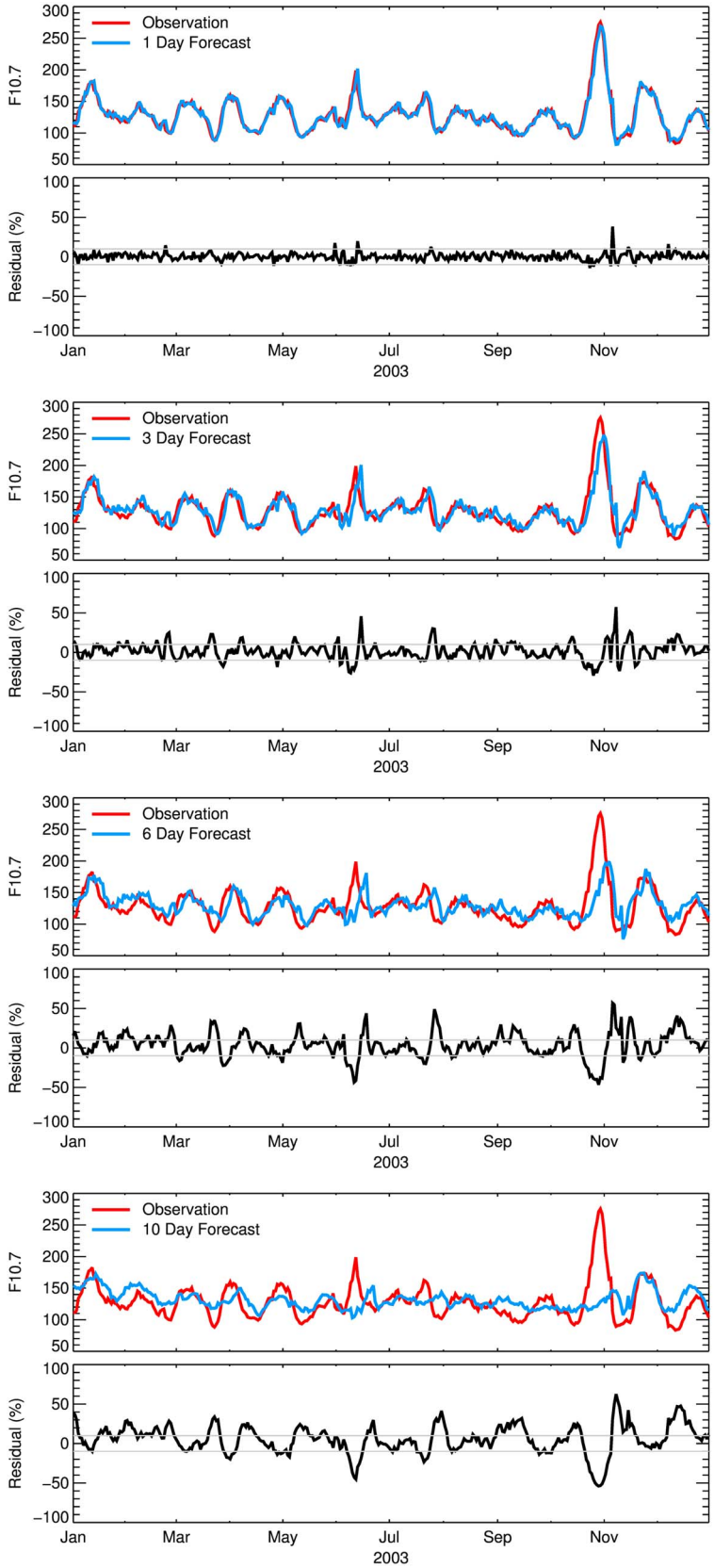

Figure 3. Plots of 1, 3, 6, and 10 day forecasts for 2003. In each plot the observed value for that day and the value *n* days ago for that day are shown. Also shown are the residual between the observation and the forecast for each day.

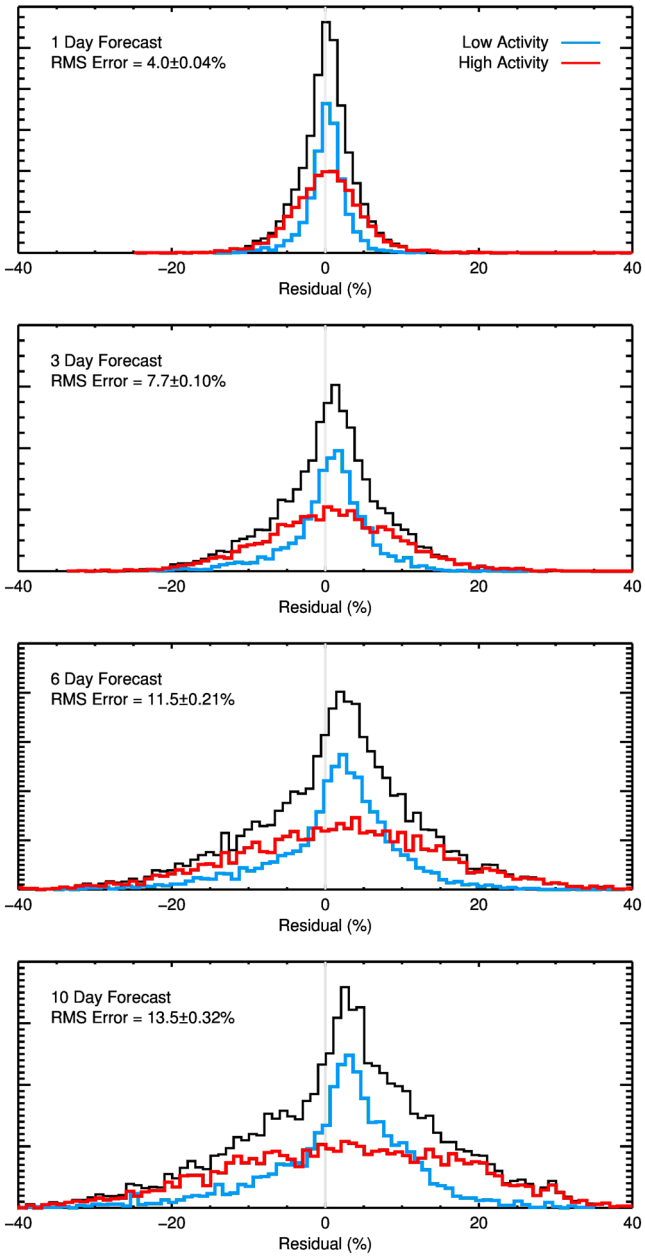

Figure 4. Histograms of the residuals for all 1, 3, 6, and 10 day forecasts from the testing period. Also shown are the residuals for periods of moderate and high solar activity ($F_{10.7} > 100$) and low solar activity ($F_{10.7} < 100$). The RMS error indicated in the figure is for all of the testing data.

To study the errors in the forecasts more systematically, we have computed histograms of the residuals for each forecast day using the testing data. Several of these histograms are shown in Figure 4. These calculations confirm the trends evident in the forecasts displayed in Figure 3 and the residual time series displayed in Figure 2. Errors in the forecasts grow with time and periods of low solar activity are easier to forecast than periods of moderate or high solar activity.

The widths of these histograms give information on the uncertainty in the model forecasts. However, since our sample is finite, these widths also have uncertainty associated with them. One complication in estimating the uncertainty of the root-mean-square (RMS) errors is the fact that successive values of forecast errors (for a given forecast day) are not independent (a typical assumption in statistical inference) but are temporally correlated. To obtain a fairly rigorous estimate of the uncertainty of the RMS error values, we conducted a 1000-sample Monte Carlo calculation, in which each simulated sample was constrained to have the same

autocorrelation as the sample autocorrelation (maximum lag 40) of the 20 year forecast error time series [cf. *Emmert and Picone*, 2010] for each forecast day. Some resulting estimates are indicated in Figure 4, and, as one would expect for a sample containing over 7000 observations, they are very small relative to the RMS error. The estimated uncertainty in the RMS error, for example, increases with increasing forecast day to a maximum of 0.47% at 45 days, which is still very small relative to the RMS error of 15.9% for that forecast.

These histograms also show that the distributions of the residuals are not normally distributed but have broad wings. Further, we also see that the peak of the histograms is skewed toward somewhat positive values (the peak in the distribution rises with forecast day is at approximately 5% at 9 days). The origin of this skew is unclear. Inspection of the residuals suggests that the model tends to overpredict $F_{10.7}$ during the decay phase of a solar rotation.

It is useful to quantify the uncertainty associated with each forecast. In principle, histograms such as those shown in Figure 4 can be used for this purpose. For a given forecast day and level of solar activity we could collect similar forecasts made during the testing period and compute the width of the resulting distribution of residuals. To simplify the error analysis, we have developed a simple parametric representation for the error as a function of forecast day and level of solar activity. To compute these parametric error models, we first binned all of the residuals for each forecast day by the $F_{10.7}$ flux level on the day that the forecast was made. For example, for the 1 day forecasts we consider forecasts for $F_{10.7}$ in the ranges 60–70, 70–80, 80–90, and so on. We then compute the standard deviation of the residuals in each bin. The resulting error surface is shown in Figure 5. Note that in this context, where we have binned the data into relatively small intervals, the uncertainty in the RMS error becomes appreciable. This is indicated by the error bars in Figure 5. Finally, for each forecast day we fit the standard deviations using

$$\text{Error(\%)} = p_0^n \tanh\left[\left(\frac{F_{10.7}}{p_1^n}\right)^3\right], \tag{2}$$

where the superscripts on the parameters (p_0^n, p_1^n) indicate that there is a different set of parameters for each forecast day. Example fits and the resulting modeled error surface are shown in Figure 5. These computed errors give the range of possible values that we expect for a given forecast. For a 6 day forecast with a current $F_{10.7}$ flux of 100, for example, we expect that about 68% of the time the observation will be within about 11.5% of the forecast, and about 95% of the time the observation will be with about 23.0%. It is important to remember that they are only estimates of the error. The distributions of residuals are only approximately Gaussian, and as mentioned previously, there is also some bias to the forecasts. Still, these estimated uncertainties are preferable to no uncertainties at all.

We now turn to evaluating model performance systematically. Perhaps, the most commonly used metric in forecasting is the root-mean-square error (RMSE), which, for this work, we take to be the square root of the mean of the residuals squared. For our distributions of residuals, whose means are close to zero, the RMS error is closely related to the standard deviation of the residuals. The RMSE for each forecast day is displayed in Figure 6. We see that the RMSE increases rapidly with forecast day, until about day 10, and then plateaus. From the 1 day to 10 day forecast the RMSE rises from 4.0% to 13.5%. Between the 10 and 45 day forecast the RMSE rises only to 15.9%. Here we also show the RMSE for the training data, which clearly follows the trends in the testing data very closely and demonstrates that we have not overfit the data.

Some studies use the Pearson correlation coefficient to quantify model performance. In Figure 6 we show the correlation coefficient between the forecast and observed values as a function of forecast day. For 1 day forecasts the correlation is very high (0.99) and the correlation declines to about 0.85 for 45 day forecasts. As noted by *Murphy* [1988], there are potential problems with using the correlation coefficient for evaluating forecasts. For example, the correlation coefficient between two data sets can be high even if there is a constant offset between them.

Low RMS errors and high correlation coefficients can obscure problems with the forecasting model. For example, the correlation between the model and the observations is high (0.99) for 1 day forecasts, but the correlation between observations on successive days is also very high and it could be that more accurate near-term forecasts could be obtained by simply using the last observed value instead of the model. Similarly, the RMS error of 15.9% and correlation of 0.85 for 45 day forecasts is more a reflection of the limited variability of the $F_{10.7}$ over that time scale rather than some insight into solar variability embedded in the model. In this

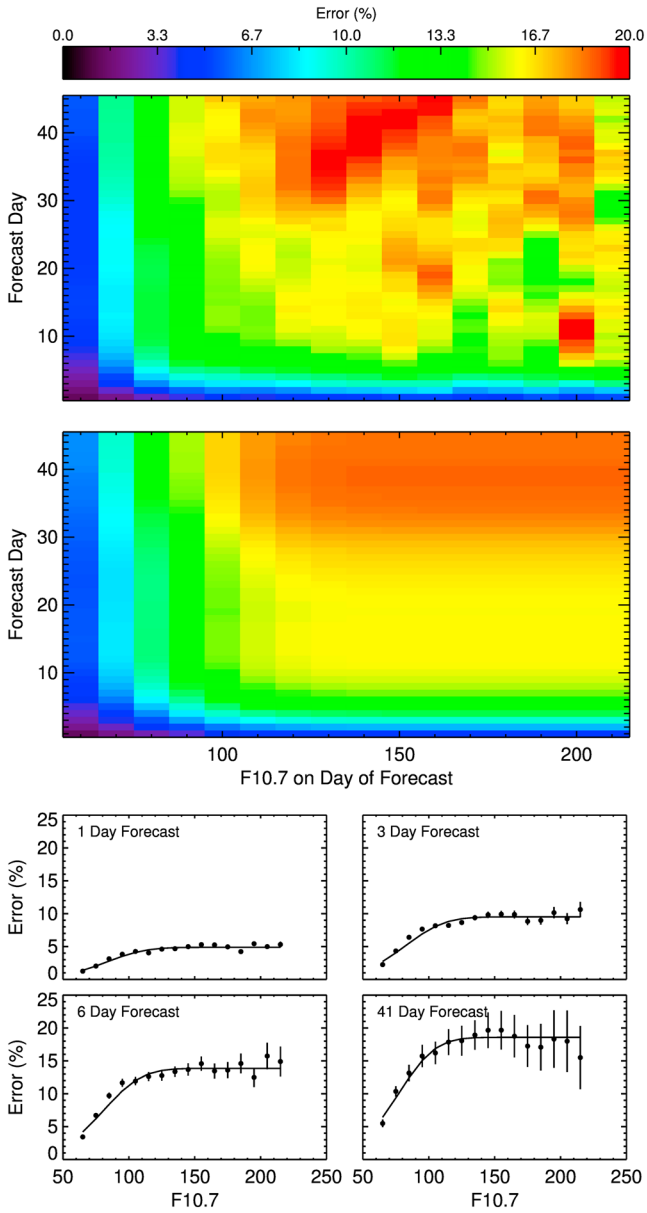

Figure 5. Forecast error as a function of level of solar activity and forecast day. (top) The 1σ relative error derived from binning the residuals from each forecast by level of solar activity. (middle) The 1σ relative errors derived from a parametric fit to the observed residuals for each forecast day. These parametric fits can be used to easily compute uncertainties for a given forecast. (bottom) Examples of observed (dots) and fit (solid lines) errors as a function of solar activity for four different forecast days. The error bars denote the estimated 1σ uncertainty of the RMSE (see text for details).

case more accurate forecasts might be obtained by simply using the mean of recent data for the long-term forecast.

To quantify the skill of the model, we compute a formal skill score [e.g., *Murphy*, 1988] based on the mean square error of the linear model and a simple reference forecasting method. The skill score is defined as

$$\text{Skill} = 1 - \frac{\text{MSE(model)}}{\text{MSE(reference)}}, \tag{3}$$

where "model" refers to the linear model presented in this paper and "reference" corresponds to a the simple model with which we are comparing. We use two simple reference forecast models: "persistence"—predicting that the last measured value will be observed for the next 45 days and "climatology"—predicting that the

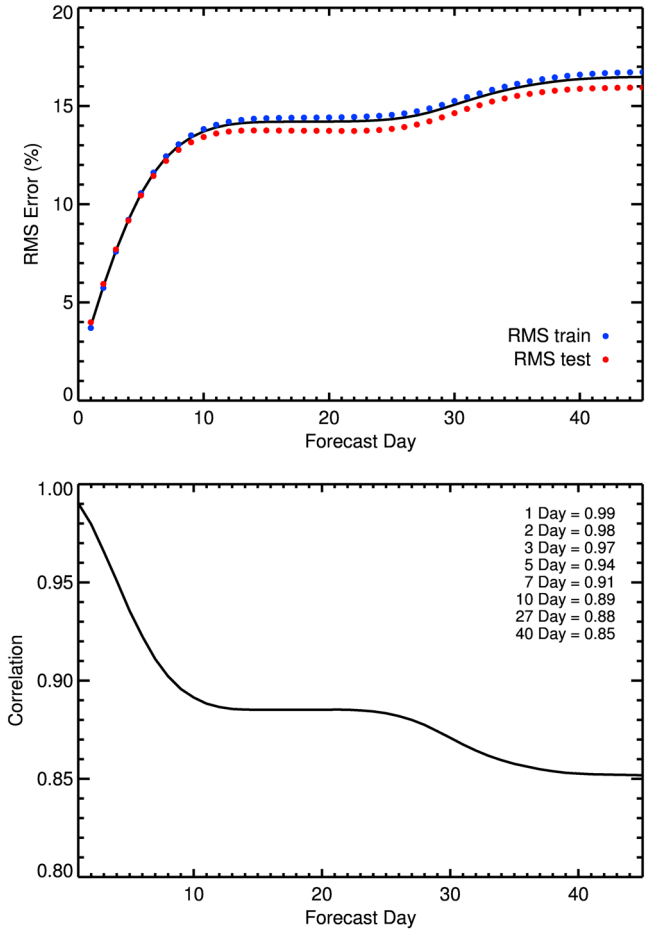

Figure 6. (top) RMS error as a function of forecast day for the linear model. Here the results from both the training and test data sets are shown. They are comparable for all forecasts, suggesting the model is not overfitting the data. (bottom) The Pearson correlation coefficient between the model and the observation as a function of forecast day.

average value from the previous 81 days will be observed over the next 45 days. Note that a perfect forecast model would have a skill score of 1, while a model that fails to outperform a simple reference model would have a negative skill score.

Figure 7 shows that the RMS error for the linear model is smaller than that of the persistence and climatology models for all forecast days. Thus, the linear model has positive skill for all forecast days. For very short and very long forecasts, however, the skill scores are small, reflecting the fact that it is relatively easy to forecast activity on these time scales with very simple models. At intermediate time scales we see the largest differences between the linear model and the reference models peaks at about 6 days.

We note that this calculation implicitly tests another possible reference model: "recurrence"—using the observed value from 26 days ago or one solar rotation, to make a 1 day forecast. This is equivalent to using the persistence model to make a 26 day forecast. Figure 7 shows a clear dip in the RMS error in the persistence model for this forecast date, but it does not outperform the climatology model.

Our analysis of the forecast residuals indicates that the errors depend on the level of solar activity. We have recomputed the RMS errors and skill scores independently for periods of moderate and high solar activity ($F_{10.7} > 100$) and low solar activity ($F_{10.7} < 100$). At the higher levels of activity the RMS errors for the linear model and the reference models are systematically larger by almost identical amounts and the skill scores are very similar to what is shown in Figure 7. At low levels of activity the reference models are closer in performance to the linear model and the skill is reduced significantly. This is illustrated in Figure 8.

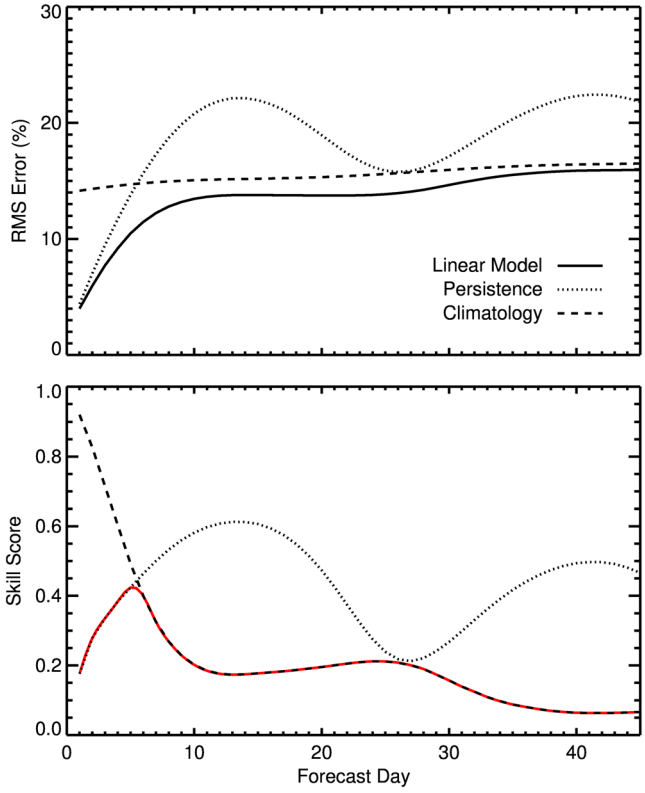

Figure 7. (top) RMS error for the linear model as well as two simple reference models: persistence, using the last observed value for all forecasts, and climatology,' using the mean of the last 81 observations for all forecasts. (bottom) The skill score as a function of forecast day computed with persistence and climatology. The red line represents the minimum skill derived from the two reference models, which peaks at 6 days.

3. Model Comparisons

As mentioned in section 1, forecasting of the $F_{10.7}$ cm radio flux has been considered by several previous authors. Unfortunately, the performance metrics reported in these studies have not been consistent and it is not possible to perform comprehensive comparisons for all metrics and all time scales. One of the motivations for the present work was to evaluate forecast performance using a number of different metrics and a full range of relevant time horizons.

Lean et al. [2009] considered 1–10 day forecasts and corresponding observations for the period 1980–2005. They report RMS errors of 5.2%, 10.5%, 16.4%, and 19.6% for 1, 3, 6, and 10 day forecasts. The linear model presented here has RMS errors of 4.1%, 8.2%, 12.6%, and 15.0% for the same forecast horizons and time period. Note that *Lean et al.* [2009] compute their RMS errors slightly differently than we have. Instead of computing the error of the forecast relative to the corresponding observation, they compute it relative to the running 81 day mean. We have recalculated our RMS errors using this formulation and find only small differences. For example, 4.1% and 4.4% for 1 day forecasts and 15.0% and 15.9% for the 10 day forecasts.

Tobiska et al. [2008] present correlation coefficients for observed and forecast $F_{10.7}$ values for two brief periods in 2001 and 2005 of about 180 days each. Their correlation coefficients for 1–4 day forecasts for these times are (0.98, 0.95, 0.90, and 0.68) and (0.98, 0.95, 0.90, and 0.60). For these periods the linear model from this paper produces correlation coefficients of (0.96, 0.91, 0.84, and 0.76) and (0.95, 0.90, 0.82, and 0.73). Based on these comparisons, the *Tobiska et al.* [2008] model outperforms the linear model for days 1–3 but underperforms the linear model significantly for the 4 day forecast. These periods, however, are so short relative to the entire data set that it is difficult to know how to interpret these comparisons. Also, *Tobiska et al.* [2008] do not discuss partitioning into training and test periods so it is possible that the good performance for forecast days 1–3 is the result of overfitting the data. Clearly, more extensive comparisons are needed.

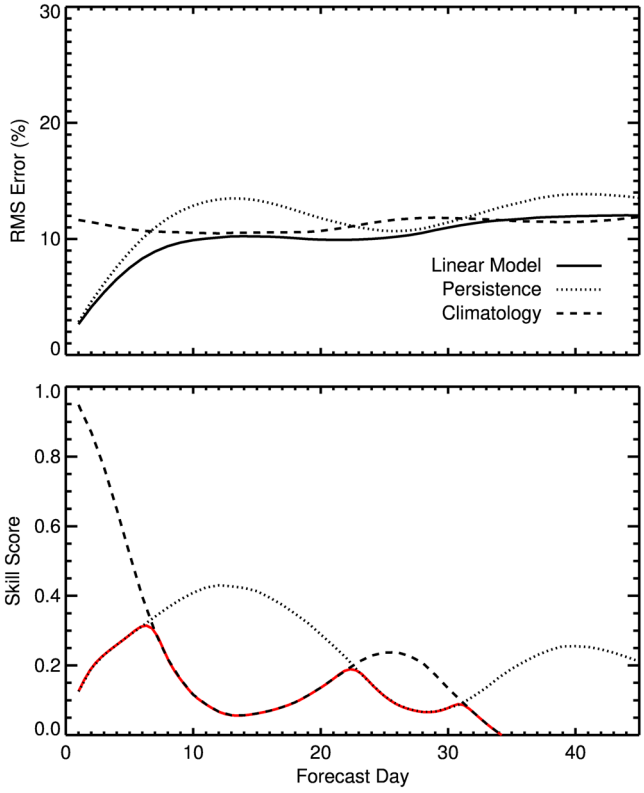

Figure 8. RMS error and skill for low levels of solar activity ($F_{10.7} < 100$). The format is the same as Figure 7. Note that for forecasts of 35 days and greater climatology has a lower RMS error and the linear model has negative skill relative to this simple reference. The skill scores for high levels of solar activity ($F_{10.7} > 100$) are very similar to those shown in Figure 7 for all data.

Henney et al. [2012] present results from forecasting $F_{10.7}$ using surface magnetic flux transport simulations for the period 1993–2010. They give Pearson correlation coefficients for 1.13, 3.13, and 7.13 day forecasts. Note that the small difference between the forecast time and the more usual 1, 3, and 7 day targets is due to differences in the times when the $F_{10.7}$ and magnetogram data are taken. They report correlation coefficients of (0.96, 0.95, and 0.91), while we obtain comparable values of (0.99, 0.97, and 0.93) for this same period. We have not attempted to adjust for the offsets in forecast times in these comparisons.

Finally, we note that the National Oceanic and Atmospheric Administration's (NOAA) Space Weather Prediction Center also publishes forecasts for the $F_{10.7}$ cm radio flux for various time horizons. The methodology and performance of these forecasts, however, does not appear to have been published in the peer-reviewed literature. *Lean et al.* [2009] analyzed the archived 1 and 3 day forecasts and found RMS errors of 5.6% and 10.7% for the period 1996–2005. The linear model has RMS errors of 4.4% and 8.4% over this same period.

4. Application to Thermospheric Density and Ionospheric TEC Forecasts

Empirical and first-principles models of the thermosphere and ionosphere often use $F_{10.7}$ as a proxy for EUV irradiance, which is the dominant driver of variations in the upper atmosphere. Therefore, $F_{10.7}$ forecasts are needed to predict thermosphere and ionosphere properties such as mass density (used in orbit prediction) and column electron density (total electron content, or TEC, which affects radiowave propagation). To assess the skill of the $F_{10.7}$ linear forecast model in these applications, we input the $F_{10.7}$ predictions into empirical models of global average mass density at 400 km altitude derived from orbital drag [*Emmert*, 2015] and global average TEC derived from GPS receiver data [*Emmert et al.*, 2017b].

The density and TEC empirical models depend on the daily $F_{10.7}$ averaged over several days (0–4 days prior to day of the density prediction and 5 days prior to 1 day after the day of the TEC prediction). The models also depend on the 81 day average $F_{10.7}$ (centered on the day of the prediction) and the *Kp* geomagnetic

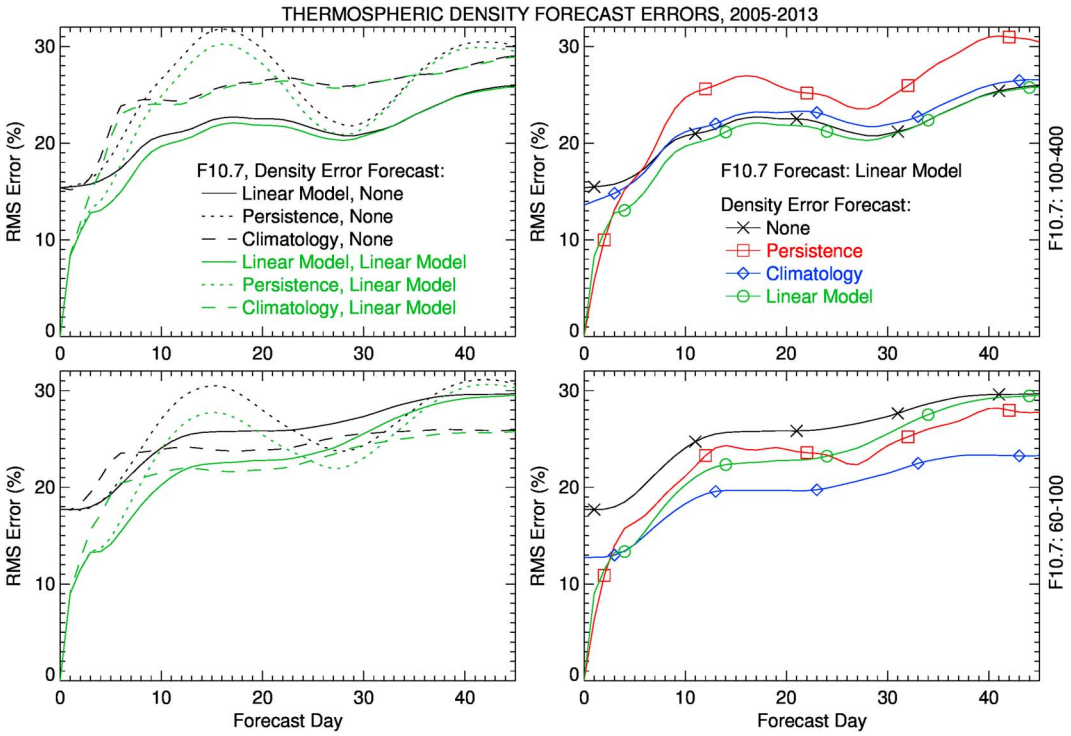

Figure 9. Thermospheric density RMSE for selected combinations of $F_{10.7}$ and density error forecast models. (left column) Thermospheric density forecast RMS error as a function of forecast day, using $F_{10.7}$ linear model (solid lines), persistence (dotted), and climatology (dashed) forecasts together with a density error linear model forecast (green lines) or with no density error forecast (black lines). (right column) Same as Figure 9 (left column) but showing results using the $F_{10.7}$ linear model forecast with different density error forecasts: none (black crosses), persistence (red squares), climatology (blue diamonds), and a linear model (green circles). The RMS errors were calculated using 2005–2013 orbit-derived density data and forecasts, under (bottom row) low solar activity conditions ($F_{10.7} > 100$) and (top row) moderate to high solar activity ($F_{10.7} > 100$).

activity index; for those arguments, we use observed values rather than forecast values, in order to isolate the contribution of the daily $F_{10.7}$ forecasts to the thermosphere-ionosphere error budget.

Even when the empirical model arguments are known with zero error, there are other sources of error in their predictions: $F_{10.7}$ and *Kp* are imperfect indices of the energy input into the thermosphere-ionosphere system, and lower atmospheric meteorology (neglected in the models) drives additional variability. The standard deviation of the data-minus-model residuals, using observed (versus forecasted) model arguments, is ~13% for density at 400 km and ~8% for TEC [*Emmert et al.*, 2014, 2017b]. These values represent a lower limit for forecast uncertainty unless a temporally local adjustment to the model predictions is applied to bring the model into better agreement with recent observed values. Since this model adjustment must also be extrapolated into the future for prediction purposes, we denote it as a density (or TEC) error forecast. Such forecasting is used in operational orbit prediction to adjust model densities [*Storz et al.*, 2005] and is possible because the errors have some temporal coherency [e.g., *Emmert and Picone*, 2011].

Here we consider density error forecasts using persistence, climatology, a linear model (employing the same methodology used for the $F_{10.7}$ forecasts), and the additional option of no forecast, in which the empirical model prediction is unadjusted. Thus, there are 12 possible combinations of $F_{10.7}$ (persistence, climatology, and linear model) and density error forecast techniques. The density error linear forecast model is trained with 1971–2005 data, which is the same period used to train the empirical model itself.

Figure 9 (left column) show the RMS error in thermospheric density for all three types of $F_{10.7}$ forecasts and the "none" and "linear model" types of density error forecasts (i.e., 6 of the 12 possible combinations). The RMS error is calculated using 2005–2013 data, so that the statistic is independent of the data used to train the $F_{10.7}$ linear model, the empirical density model, and the density error linear model. In the moderate-to-high solar activity case (top left), the $F_{10.7}$ linear forecast model reduces the density forecast RMS error by 0–10%,

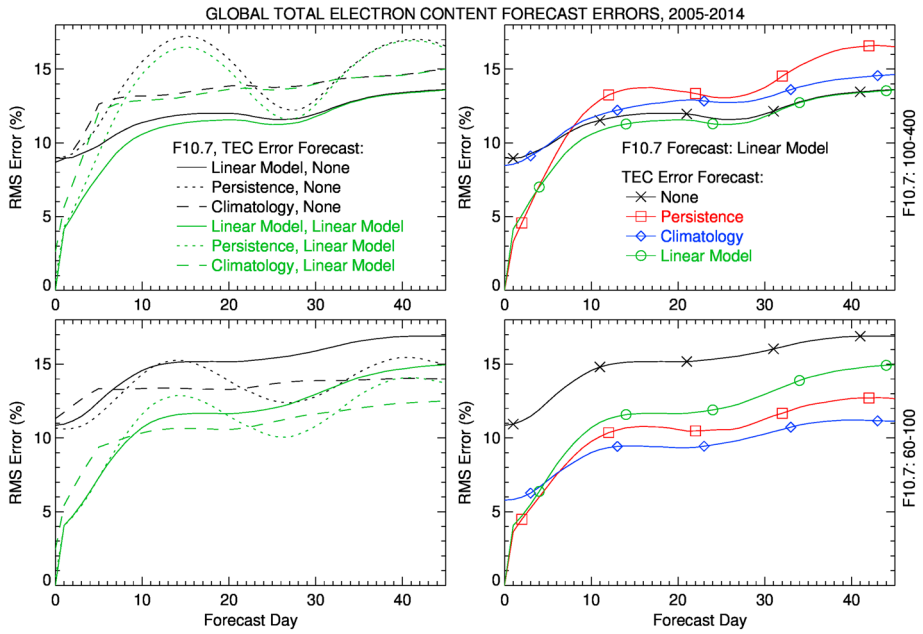

Figure 10. Same as Figure 9 but showing 2005–2014 global average TEC forecast RMS errors.

compared to the persistence and climatology $F_{10.7}$ forecasts. However, in the low solar activity case, the $F_{10.7}$ climatology forecast produces lower RMS density error than the $F_{10.7}$ linear model after forecast day 10, as does the $F_{10.7}$ persistence forecast around forecast day 27.

This counterintuitive result that a more accurate $F_{10.7}$ forecast can produce a less accurate density forecast is due to the anomalously low density that occurred from 2006 to 2009. The observed density was consistently up to ~20% lower than predicted by the empirical model [*Emmert et al.*, 2014], and this bias contributes strongly to the RMS error metric, so that the accuracy of the $F_{10.7}$ forecast becomes less relevant. We note that the RMS error metric does not remove the sample mean (which is unknown in a forecast application), unlike the root variance and standard deviation statistics. During the 2006–2009 solar minimum, it appears that the less accurate (and therefore essentially more random) $F_{10.7}$ persistence and climatology forecasts were more likely than the $F_{10.7}$ linear forecast to produce low $F_{10.7}$ values that would compensate for the density bias. Because the empirical density model is nonlinear with respect to $F_{10.7}$, we suspect that the positive part of the $F_{10.7}$ forecast error distribution is less influential in this case, but further study is needed to confirm this.

The application of a density error forecast linear model substantially reduces the RMS error of the density forecasts, compared to using the unadjusted empirical density model (green versus black lines in Figure 9). However, the density error linear model is not necessarily optimal. As shown in Figure 9 (bottom right), in the solar minimum case a density error climatology forecast (i.e., adjustment of the empirical model by the average error over the past 81 days) produces the lowest RMS error after forecast day 6. This indicates that the density errors during the anomalous 2006–2009 solar minimum are better characterized by a simple bias than by the stochastic process represented by the linear model (which was trained using 1971–2005 data).

Similar results are obtained for global TEC forecast errors, as shown in Figure 10. We conclude that to get the maximum benefit from improved solar EUV irradiance forecasts, other sources of error in thermospheric density and ionospheric TEC predictions should be considered. *Lean et al.* [2009] reached a similar conclusion with respect to thermospheric density. We will examine the error budget of thermospheric density in a future paper.

5. Summary and Discussion

We have presented the formulation and evaluation of a simple forecasting model for the $F_{10.7}$ cm radio flux, a widely used proxy for solar activity. This model uses linear combinations of the previous 81 days of observations to produce daily forecasts from 1 to 45 days. Skill for the linear model relative to the persistence and climatology forecasts peaks at about 6 days.

Comparisons with other models show that the linear model performs comparably, although these comparisons are somewhat limited in scope. This work suggests that future forecasting efforts should include a larger number of performance metrics, focusing on RMS errors and skill scores relative to simple reference models. Considering extended periods of high and low activity is also essential.

We conjecture that it will be difficult to create an $F_{10.7}$ forecasting model based solely on retrospective observations that significantly outperforms the linear model presented here. We have experimented with artificial neural networks with various architectures using Google's TensorFlow library. In each case the performance of the neural network matched that of the linear model almost exactly. Clearly, much more work is required before a definitive conclusion can be reached, but these initial experiments were not encouraging.

The magnetic flux transport modeling considered by *Henney et al.* [2012] appears to be a much more promising approach to more accurate forecasting models. While flux transport currently performs comparably to the linear model, it will undoubtedly improve as we learn more about how magnetic flux emerges and evolves on the Sun. Flux transport simulations are also a useful framework for incorporating information on far side activity or predicted near side flux emergence.

Acknowledgments
This research was sponsored by the Chief of Naval Research. Observations of the $F_{10.7}$ cm radio flux were downloaded from ftp://ftp.ngdc.noaa.gov/stp/geomagnetic_data/indices/kp_ap and ftp://ftp.geolab.nrcan.gc.ca/data-/solar_flux/daily_flux_values/fluxtable.txt.

References

Barnes, G., A. C. Birch, K. D. Leka, and D. C. Braun (2014), Helioseismology of pre-emerging active regions. III. Statistical analysis, *Astrophys. J.*, *786*, 19, doi:10.1088/0004-637X/786/1/19.

Emmert, J. T. (2015), Altitude and solar activity dependence of 1967–2005 thermospheric density trends derived from orbital drag, *J. Geophys. Res. Space Physics*, *120*, 2940–2950, doi:10.1002/2015JA021047.

Emmert, J. T., and J. M Picone (2010), Climatology of globally averaged thermospheric mass density, *J. Geophys. Res.*, *115*, A09326, doi:10.1029/2010JA015298.

Emmert, J. T., and J. M. Picone (2011), Statistical uncertainty of 1967–2005 thermospheric density trends derived from orbital drag, *J. Geophys. Res.*, *116*, A00H09, doi:10.1029/2010JA016382.

Emmert, J. T., S. E. McDonald, D. P. Drob, R. R. Meier, J. L. Lean, and J. M. Picone (2014), Attribution of interminima changes in the global thermosphere and ionosphere, *J. Geophys. Res. Space Physics*, *119*, 6657–6688, doi:10.1002/2013JA019484.

Emmert, J. T., H. P. Warren, A. M. Segerman, J. M. Byers, and J. M. Picone (2017a), Propagation of atmospheric density errors to satellite orbits, *Adv. Space Res.*, *59*, 147–165, doi:10.1016/j.asr.2016.07.036.

Emmert, J. T., A. J. Mannucci, S. E. McDonald, and P. Vergados (2017b), Attribution of interminimum changes in global and hemispheric total electron content, *J. Geophys. Res. Space Physics*, *122*, 2424–2439, doi:10.1002/2016JA023680.

Henney, C. J., W. A. Toussaint, S. M. White, and C. N. Arge (2012), Forecasting $F_{10.7}$ with solar magnetic flux transport modeling, *Space Weather*, *10*, S02011, doi:10.1029/2011SW000748.

Howard., R. A., et al. (2008), Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI), *Space Sci. Rev.*, *136*, 67–115, doi:10.1007/s11214-008-9341-4.

Ilonidis, S., J. Zhao, and A. Kosovichev (2011), Detection of emerging sunspot regions in the solar interior, *Science*, *333*, 993–996, doi:10.1126/science.1206253.

Lean, J. L., J. M. Picone, and J. T. Emmert (2009), Quantitative forecasting of near-term solar activity and upper atmospheric density, *J. Geophys. Res.*, *114*, A07301, doi:10.1029/2009JA014285.

Lindsey, C., and D. C. Braun (2000), Seismic images of the far side of the Sun, *Science*, *287*, 1799–1801, doi:10.1126/science.287.5459.1799.

Murphy, A. H. (1988), Skill scores based on the mean square error and their relationships to the correlation coefficient, *Mon. Weather Rev.*, *116*(12), 2417–2424, doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Pevtsov, A. A., G. H. Fisher, L. W. Acton, D. W. Longcope, C. M. Johns-Krull, C. C. Kankelborg, and T. R. Metcalf (2003), The relationship between X-ray radiance and magnetic flux, *Astrophys. J.*, *598*, 1387–1391, doi:10.1086/378944.

Schrijver, C. J. (1987), Solar active regions—Radiative intensities and large-scale parameters of the magnetic field, *Astron. Astrophys.*, *180*, 241–252.

Storz, M. F., B. R. Bowman, M. J. I. Branson, S. J. Casali, and W. K. Tobiska (2005), High Accuracy Satellite Drag Model (HASDM), *Adv. Space Res.*, *36*, 2497–2505, doi:10.1016/j.asr.2004.02.020.

Tapping, K. F. (1987), Recent solar radio astronomy at centimeter wavelengths—The temporal variability of the 10.7-cm flux, *J. Geophys. Res.*, *92*, 829–838, doi:10.1029/JD092iD01p00829.

Tapping, K. F. (2013), The 10.7 cm solar radio flux ($F_{10.7}$), *Space Weather*, *11*, 394–406, doi:10.1002/swe.20064.

Tobiska, W. K., S. D. Bouwer, and B. R. Bowman (2008), The development of new solar indices for use in thermospheric density modeling, *J. Atmos. Sol. Terr. Phys.*, *70*, 803–819, doi:10.1016/j.jastp.2007.11.001.

Ugarte-Urra, I., L. Upton, H. P. Warren, and D. H. Hathaway (2015), Magnetic flux transport and the long-term evolution of solar active regions, *Astrophys. J.*, *815*, 90, doi:10.1088/0004-637X/815/2/90.

Wang, Y.-M., A. G. Nash, and N. R. Sheeley Jr. (1989), Magnetic flux transport on the Sun, *Science*, *245*, 712–718, doi:10.1126/science.245.4919.712.