Forecasting the Remaining Duration of an Ongoing Solar Flare

Jeffrey W Reep[1] and Will T Barnes[2]

[1]US Naval Research Laboratory
[2]National Research Council Postdoctoral Associate residing at US Naval Research Laboratory

November 24, 2022

Abstract

The solar X-ray irradiance is significantly heightened during the course of a solar flare, which can cause radio blackouts due to ionization of the atoms in the ionosphere. As the duration of a solar flare is not related to the size of that flare, it is not directly clear how long those blackouts can persist. Using a random forest regression model trained on data taken from X-ray light curves, we have developed a direct forecasting method that predicts how long the event will remain above background levels. We test this on a large collection of flares observed with GOES-15, and show that it generally outperforms simple linear regression. This forecast is computationally light enough to be performed in real time, allowing for the prediction to be made during the course of a flare.

Forecasting the Remaining Duration of an Ongoing Solar Flare

Jeffrey W. Reep[1]

Will T. Barnes[2]

[1]Space Science Division, US Naval Research Laboratory
[2]National Research Council Research Associate Residing at the Naval Research Laboratory

Key Points:

- A random forest regression model has been trained on GOES-15 XRS data to predict the amount of time for an ongoing flare to return to background flux levels
- The random forest prediction out-performs simple linear regression
- This forecast can be run in real-time using only GOES/XRS data in order to forecast how long radio communications will be impacted

Corresponding author: Jeffrey Reep, `jeffrey.reep@nrl.navy.mil`

Abstract

The solar X-ray irradiance is significantly heightened during the course of a solar flare, which can cause radio blackouts due to ionization of the atoms in the ionosphere. As the duration of a solar flare is not related to the size of that flare, it is not directly clear how long those blackouts can persist. Using a random forest regression model trained on data taken from X-ray light curves, we have developed a direct forecasting method that predicts how long the event will remain above background levels. We test this on a large collection of flares observed with GOES-15, and show that it generally outperforms simple linear regression. This forecast is computationally light enough to be performed in real time, allowing for the prediction to be made during the course of a flare.

Plain Language Summary

The X-ray emissions from a solar flare can impact the Earth's ionosphere, ionizing the atoms and thereby increasing the total electron content. This is turn reduces the range of radio communications, effectively causing a blackout. Unfortunately, it is not clear how long any given flare might last, and therefore how long communications could be adversely impacted. The duration of a solar flare is not related to the size of that flare, so it is not straightforward to predict how long it might last from simple observations. In this work, we develop a method that allows us to forecast that duration in real time using a machine learning algorithm. This would then allow prediction of how long radio communications will be impacted.

1 Introduction

Solar flare emissions are measured and classified by their soft X-ray emissions as measured by the X-ray Sensors (XRS) on board the Geostationary Operational Environmental Satellites (GOES). These satellites measure the X-ray emission in the 1–8 Å and 0.5–4 Å wavelength bands, where the peak level in the former is used as a flare classification with X-class corresponding to peak emission above 10^{-4} W m^{-2}, and M, C, B, and A classes in decreasing orders of magnitude. The frequency distribution of flare sizes is a power law distribution with slope of approximately -2 (*e.g.* Hudson 1991), so the largest flares are significantly rarer than smaller ones.

Solar flare durations vary widely, and are not correlated with the GOES class of the flare. In Reep & Knizhnik (2019), for example, it was shown that the GOES class is not correlated with the full-width-at-half-maximum (FWHM) in either of the 1–8 Å or 0.5–4 Å XRS wavelength bands, and furthermore that the distributions of FWHM are consistent with log-normal distributions, approximately ranging from tens of seconds to a few hours. There were also no relations found between the GOES FWHM and the temperature, emission measure (EM), thermal energy, ribbon area, or active region area, though there may be a relation in large flares between the magnetic flux and duration (Reep & Knizhnik, 2019). Because the durations do not correlate with the flare class, one cannot estimate the duration from the peak flux of that flare (or similarly from other simple parameters). Toriumi et al. (2017) did find a linear relation between the separation of the flare ribbon centroids and the duration in a set of large flares (above M5), but this cannot be easily measured in real time, nor done with flares occurring at the limb, and would need to be extended to smaller flares to be of general use.

It is not clear that the duration of a solar flare in X-rays can be predicted from simple scaling laws. The duration may be related to some parameters that can be calculated *post facto*, but this means that a real time forecast would be impossible. However, given that there are thousands of flares observed by GOES of widely varying sizes and durations, and at various levels of solar activity, the statistical data should contain information to estimate likelihood of the duration. This is the goal of this paper.

Therefore, we wish to develop a real-time forecasting tool that predicts how long a solar flare will take to cool from its current flux level to a defined background level. For example, suppose that an X-class flare like the one in Figure 1 were occurring at the present moment. The two GOES/XRS channels are shown (top) in red (1–8 Å) and blue (0.5–4 Å), as well as the derivative of the long wavelength channel (bottom). Given the light curves up to the present and their current flux values (solid lines), could we estimate how long they would take to return to the background level? What is the likely duration and what is the range of potential values? In other words, how well can we forecast the true light curves (dotted lines)? To do so, we use quantities measured from the light curves at five times t_i, denoted by vertical pink lines, defined by the derivatives, for the 1–8 Å channel. These are fully explained in Section 2. We attempt to predict the timing of the final vertical line, which we call t_4 in this paper.

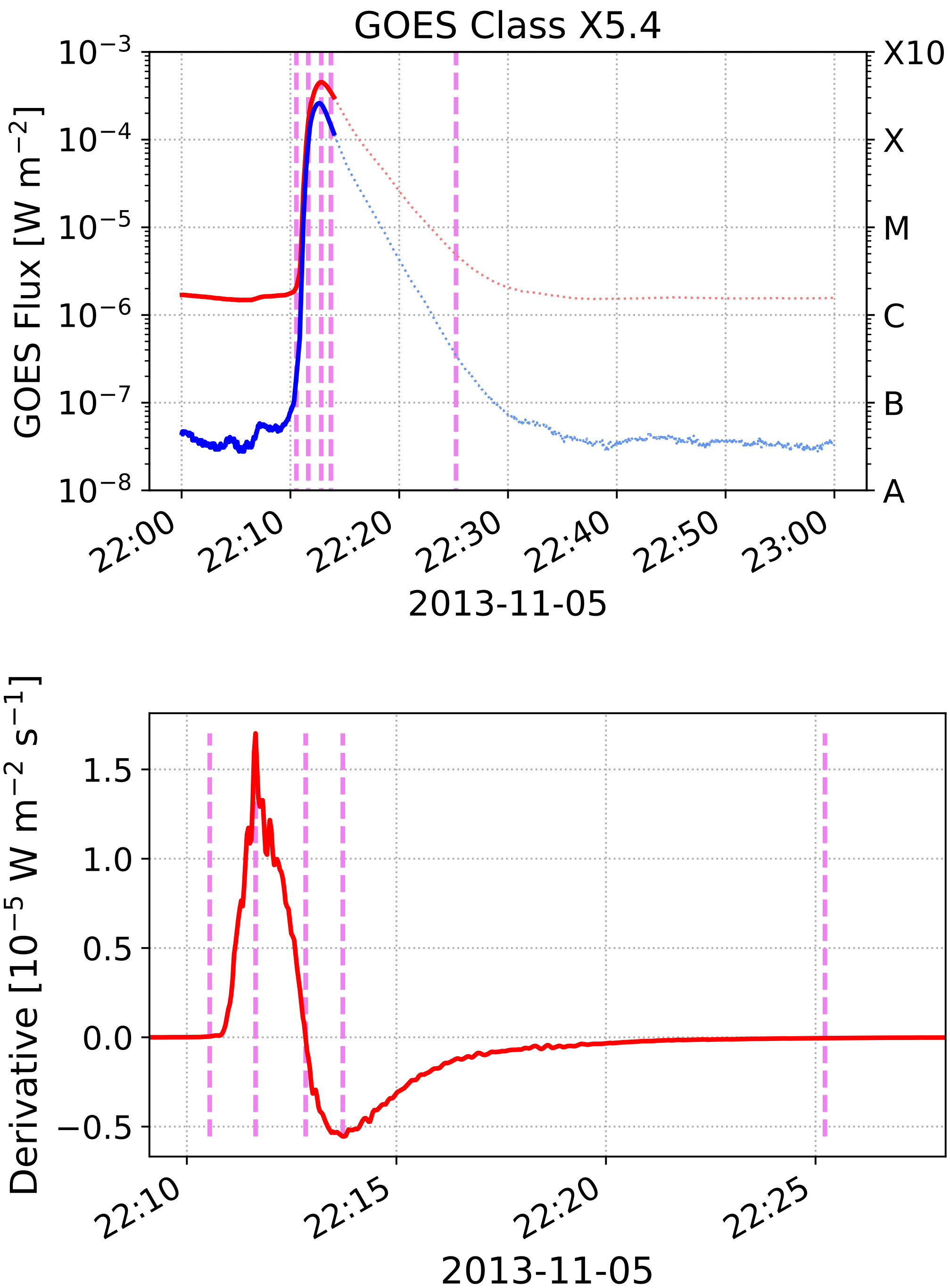


Figure 1. A statement of the problem. If this flare was currently ongoing (at about 22:14), how long might we estimate it to take to cool from its current flux back down to the background level? How much uncertainty is there in the estimate of duration? At top, the figure shows the GOES 1–8 Å light curve (red) and 0.5–4 Å light curve (blue) up to about UT 22:14 for an X5.4 flare on 5 November 2013 (solid lines), while the dotted lines show the true light curves. The bottom plot shows the time derivative of the 1–8 Å channel, and the five dashed lines mark the values of each t_i (see Section 2).

In this work, we use a random forest regressor trained only with parameters measured from observed GOES/XRS data to estimate the total duration of the flare in XRS and calculate a likelihood interval of the duration. We show that this model improves its prediction over the course of a flare, and that it outperforms a linear regression model.

2 Data Preparation

We first prepare data taken from XRS observations of flares with measurements made by the GOES-15 satellite. We examine all flares in NOAA's GOES event catalogue from the launch of GOES-15 through the deactivation of the satellite on 2 March 2020. While we have chosen to only focus on GOES-15 data, the model could be trained with data from any GOES satellite. Because the instrumental response varies slightly for each satellite, it is possible that predictions trained with one satellite's data may not work well for another satellite, which should be tested in the future.

We expand the time range of the events in the GOES catalogue to ten minutes earlier and sixty minutes later than the listed start and end times to better include pre-flare rise and late phase decay. We do not background subtract the data, since we do not know *a priori* whether the background activity impacts flare duration. We remove events from the data set if any of the following apply to an event: (1) there are any data gaps in either GOES/XRS channel; (2) the GOES catalogue is missing any data for an event; (3) the signal-to-noise ratio in either GOES/XRS channel is less than 2; (4) the event is not sufficiently isolated in time from other events, which we explain below based on the timings.

For each flare, we measure 18 parameters for each event from the light curves for both the 1–8 Å and 0.5–4 Å channels. We first calculate the derivatives in both channels using a 32-point Savitzky-Golay smoothing filter (Savitzky & Golay, 1964). The filter is particularly important for small flares where the signal-to-noise ratio is not as good as in larger flares. We then define five times in terms of the derivative with respect to time of the light curve: t_0 the onset of the flare as the derivative begins to rise, t_1 when the derivative is maximized, t_2 when the flux peaks, t_3 when the derivative is minimized, and t_4 when the derivative approximately returns to 0. More specifically, we define t_0 as the first time when the derivative exceeds 10^{-4} times the peak flux; for example, for an X1 flare, we define t_0 as the first time when the derivative exceeds 10^{-8} W m^{-2} s^{-1}.

Table 1. Definitions of the flare timings that we use in this work. The values are all implicitly defined in terms of the first time derivative of the light curves $F(t)$ of each XRS channel. For all flares in our data set, we require $t_i < t_{i+1}$ to be certain that each event is isolated from other events in time.

Time	Meaning	Definition
t_0	Approximate Start Time	$\frac{dF}{dt}\|_{t_0} = (10^{-4}\ \mathrm{s}^{-1}) \times F(t_2)$
t_1	Maximum of $\frac{dF}{dt}$	$\frac{dF}{dt}\|_{t_1} = \max\left(\frac{dF}{dt}\right)$
t_2	Peak of Flare	$\frac{dF}{dt}\|_{t_2} = 0$
t_3	Minimum of $\frac{dF}{dt}$	$\frac{dF}{dt}\|_{t_3} = \min\left(\frac{dF}{dt}\right)$
t_4	Approximate End Time	$\frac{dF}{dt}\|_{t_4} = -(10^{-4}\ \mathrm{s}^{-1}) \times F(t_2)$

We similarly define t_4 as the first time after t_3 when the derivative is above -10^{-4} times the peak flux; for an X1 flare, t_4 is the first time after t_3 when the derivative increases above -10^{-8} W m^{-2} s^{-1}. We list the definitions in Table 1.

At each time t_i, we measure the flux level $F_i = F(t_i)$, the derivative of the flux with time, $\frac{dF}{dt}|_{t_i}$, and the definite integral of the flux $A_i = \int_{t_0}^{t_i} F(t)\, dt$. By definition, $A_0 = 0$ and $\frac{dF}{dt}|_{t_0} \approx 0$, so these are neglected. This gives 36 total parameters, 18 for each channel. We train the random forest regressor twice: once using the values through time t_2 in order to forecast t_3 and t_4 for each channel, and once using the values through time t_3 in order to forecast t_4. Since we do not know *a priori* what parameters act as strong predictors, we consider all features when training the random forest predictor.

Because flares often occur within short time periods of one another, it can be difficult to measure some of the parameters with which we might wish to train a machine learning algorithm. For example, if a second flare occurs before an earlier flare finishes and the X-ray emission has not yet returned to background levels, we cannot measure the true end time of the first flare. This means that we cannot include it in our training data set. Specifically, we enforce the condition that each t_i occurs in sequential order, and discard events where this condition does not hold. This happens, for example, when there are flares of similar size occurring in close succession.

After pruning the data set, we are left with 7055 events, ranging in GOES class from A2.2 to X13. The FWHM of the light curves range between 24 and 16068 s in the 1–8 Å

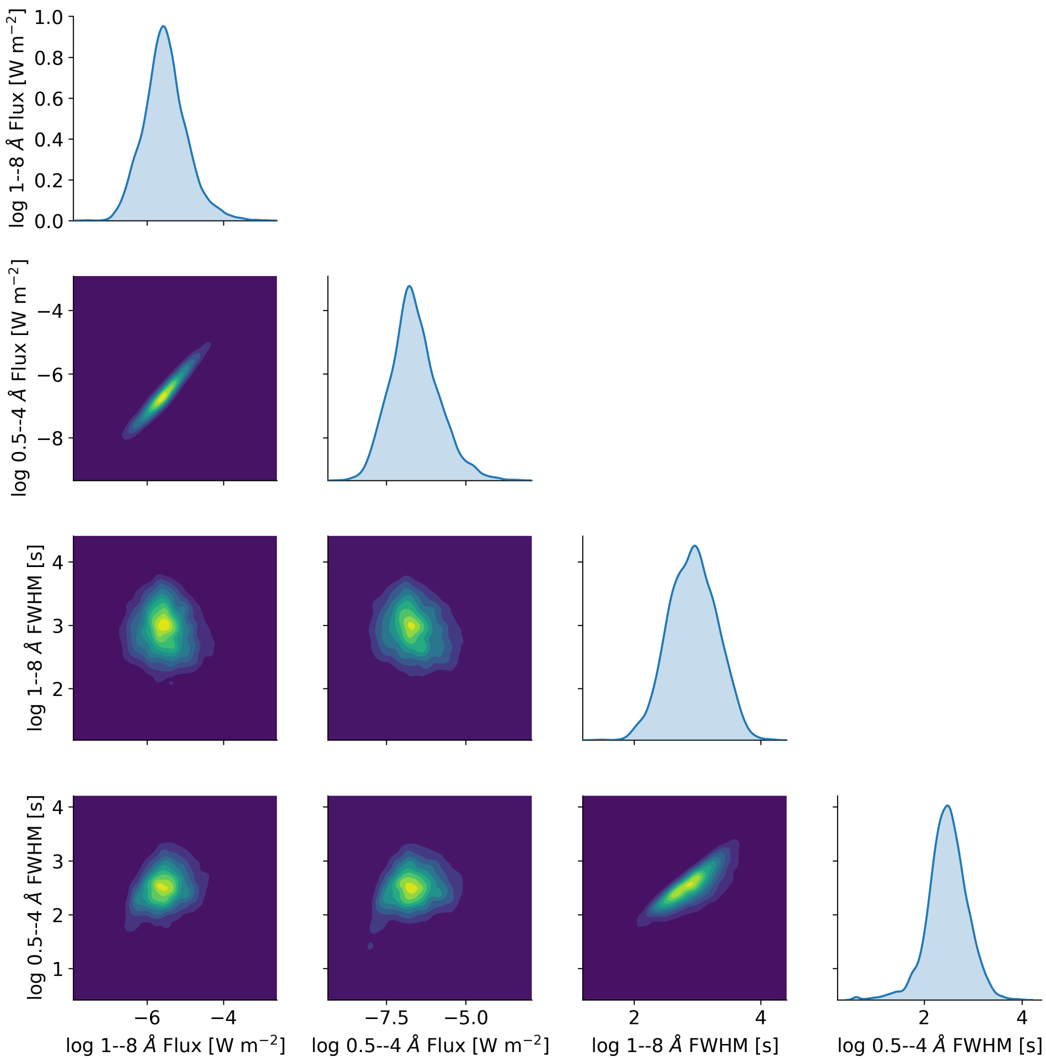


Figure 2. Plots of the distributions of the peak fluxes and FWHM in the two XRS channels. The diagonal elements show the distributions themselves, while off-diagonal plots show heat maps comparing each pair of variables. The flare sizes and durations are uncorrelated, and each distribution is consistent with log-normal (see Reep & Knizhnik 2019).

channel and between 4 and 10307 s in the 0.5–4 Å channel. The flares occurred between 4 September 2010 and 24 January 2020 across a wide range of solar activity levels. In Figure 2 we show the distributions of fluxes and durations in both XRS channels, as well as heat maps showing the relations between each variable. The fluxes and durations are uncorrelated, while each individual distribution is consistent with log-normal, demonstrated with Kolmogorov-Smirnov tests in Reep & Knizhnik (2019).

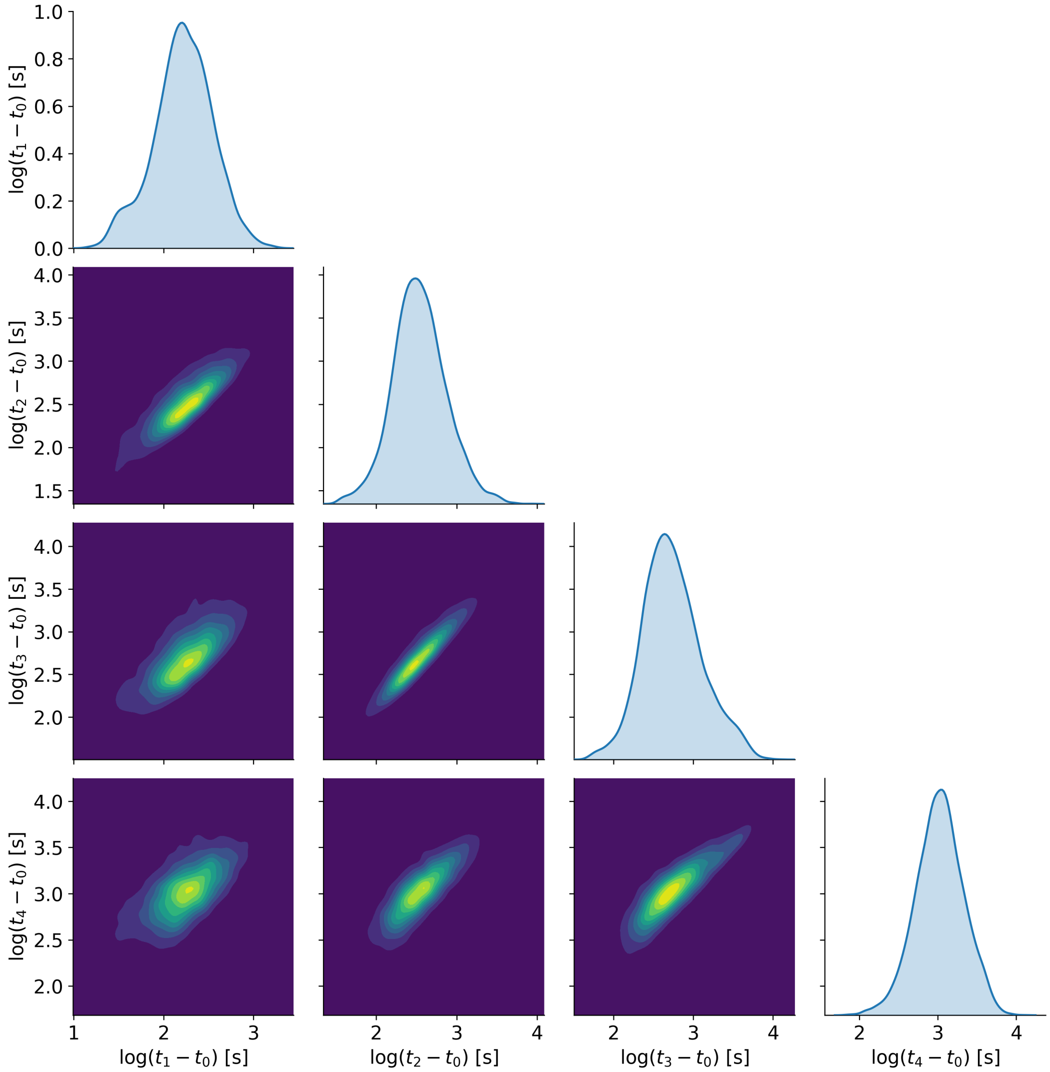


Figure 3. A pair plot showing the relationships between the timings $t_i - t_0$, for each $i > 0$, in the 1–8 Å channel. While there is a general positive and monotonic correlation, there is also noticeable scatter.

Similarly, Figure 3 shows the relationships between the timings in the events in the 1–8 Å channel. We show each $t_i - t_0$, $i > 0$, which in all cases show a positive and monotonic correlation. However, the scatter is noticeable, particularly when going from early timings to late ones. It is clear that a linear regression could then give a prediction for t_3 and t_4, but would not capture the scatter. Examining the last row, for example, it is clear that t_1 would not give a good prediction for t_4, but t_2 would give a better prediction, and t_3 better still. We show in the next section that a random forest regressor outperforms a simple linear case.

3 Duration Prediction

We use the random forest regressor implemented in the `scikit-learn` package (Pedregosa et al., 2011) to calculate this prediction. Random forest methods (Breiman, 2001) create an ensemble of decision tree predictors, which when averaged give a robust prediction. The chief advantage of a random forest predictor is that it does not assume any functional form for the data being fit (whether through classification or regression), so that the model can map complex relationships between the inputs and outputs. Breiman (2001) notes that the other main advantages are that random forests are fast, robust against noise, and can be used to quantify errors and correlations easily.

Decision trees (Breiman et al., 1984) work by recursively splitting a data set, based on values of training features, into nodes. The trees split a certain number of times (the depth), which can be fixed or randomized. While an individual decision tree is not robust in predictions, by combining multiple trees with differing nodes and depths, the prediction is improved. The random forest predictor randomizes the elements of the set that are used in a decision tree, the features used in the splitting at each node, the values by which they are split, and the depth of the trees. When combined, these many varying decision tree predictors significantly reduce the variance of the model.

3.1 Predictions at time t_2

We separate the full data set randomly into a training group (67% of events) and test group (33% of events), using the 36 features of the light curves defined above to train the data. We train the random forest regressor with the former group. We then test the efficacy the model by predicting t_3 and t_4 for the remaining 33% of the events and compare them to the actual, observed t_3 and t_4 values.

These predictions are presented in Figure 4 for the test group. Using the parameters through time t_2 to train the random forest, we show the estimated values of t_3 (top row) and t_4 (bottom row) for the 1–8 Å (left column) and 0.5–4 Å (right column) XRS passbands as compared to their true values. We perform a linear regression with a non-parametric Theil-Sen estimator (Sen, 1968; Theil, 1992) to estimate the slope and intercept for these scalings. Theil-Sen estimators are robust and insensitive to outliers in the data, unlike ordinary least squares linear regression. We would expect a slope of 1 for a perfect prediction, such that the predicted time grows at the same rate. The pre-

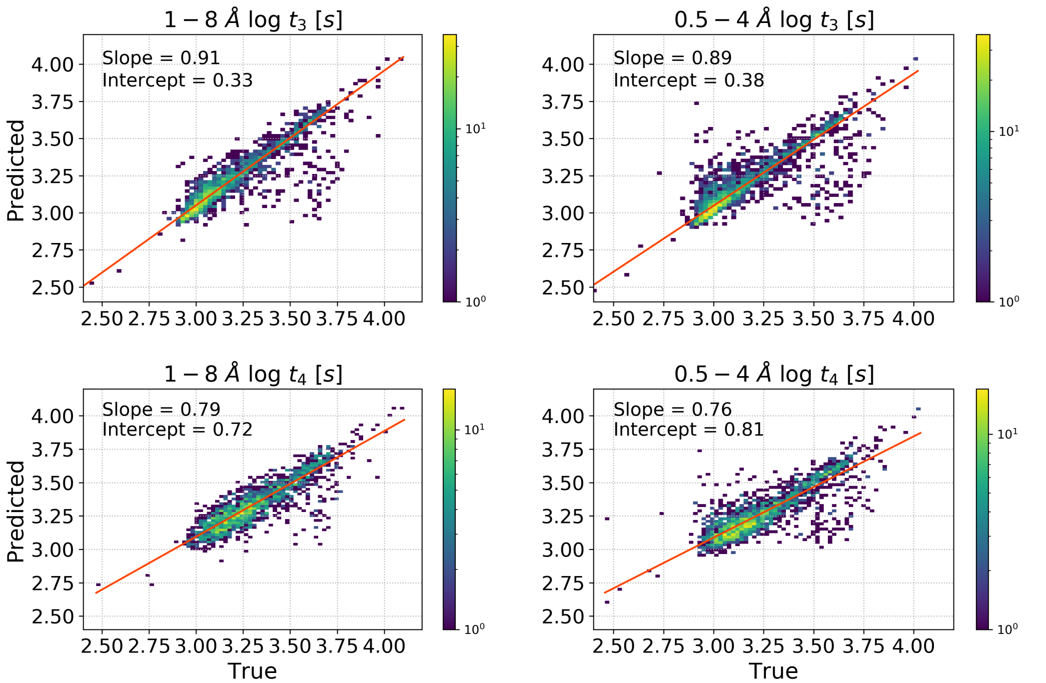


Figure 4. The random forest regressor prediction of the remaining durations, using a random forest trained with the parameters through time t_2. The plots show 2D histograms comparing the predicted and true values of t_3 (top row) and t_4 (bottom row) for the XRS 1–8 Å (left column) and 0.5–4 Å (right column) channels for the 33% of events placed in the test group. The predictions were performed using only the features through time t_2. The orange lines are Theil-Sen linear regression fits to the data, with the slope and intercepts indicated. The slopes are less than 1, indicating that the durations are underestimated in general.

dictions for t_3 scale approximately with a slope of 0.9, while the predictions for t_4 scale more slowly, suggesting that the forecast will generally underestimate the duration for longer events.

To demonstrate that these predictions are reasonable, we show six example flares in Figure 5 with both relatively short (left) and long (right) durations, ranging in class from C-class (top), M-class (middle), to X-class (bottom). We estimate the remaining duration from the peak of each channel at time t_2. On each plot, we show the predicted value of t_4 with a dashed line that terminates with an X mark, where the red case shows the prediction for 1–8 Å and blue for 0.5–4 Å. These lines are calculated with a spline fit connecting the true t_2 values with the predicted values of t_4, and are meant simply to guide the eye. The outer dashed lines similarly show the $\pm 1\sigma$ confidence intervals for each prediction of t_4. In general, the predictions of t_4 for short duration flares are close to the true values, while those for long duration flares are underestimated. This agrees with Figure 4, which shows that the predictions for t_4 scale more slowly than the true values (slope < 1).

3.2 Predictions at time t_3

We repeat the analysis, this time training the random forest using the parameters through time t_3 (the minima of the first derivatives) in order to forecast t_4. We use the same train-test split of the data as before to ensure a valid comparison. Figure 6 shows the results of the new predictions of t_4 for the two XRS channels. In this case, the slopes are approximately 0.9, improving upon the estimates in Figure 4. The scatter is similarly reduced. In general, the method still slightly underestimates the durations of events.

In Figure 7, we show the predictions for the same six flares presented previously, including C, M, and X class events that are both short and long durations. We see that the predictions are now slightly improved when compared to their true values, although the estimates of t_4, particularly for the longer events, are still somewhat short. For example, the long duration X-flare, at bottom right, is still underestimated. This suggests that as the flare proceeds from times t_2 (the peak) to t_3 (the cooling phase), we can improve our estimate of the end of the flare t_4.

3.3 Prediction Skill

`scikit-learn` also calculates the feature importance, or which features are most important in determining the predictions, as a product of its predictions. For regression problems, this is quantified by the reduction in variance for each feature, and then the features are ranked accordingly. For predictions of t_4 made at time t_2, the most important features are, respectively, t_2, t_0, A_2, A_1, t_1. The flux levels and derivative magnitudes are, perhaps surprisingly, almost negligible in their impact on the predictions.

How does this prediction fare? We calculate a skill score by comparing to a simple linear regression model. Since t_2 is the most important feature in predicting t_4, we use it alone to give a simple prediction. In Figure 8, we show scatter plots of t_2 versus

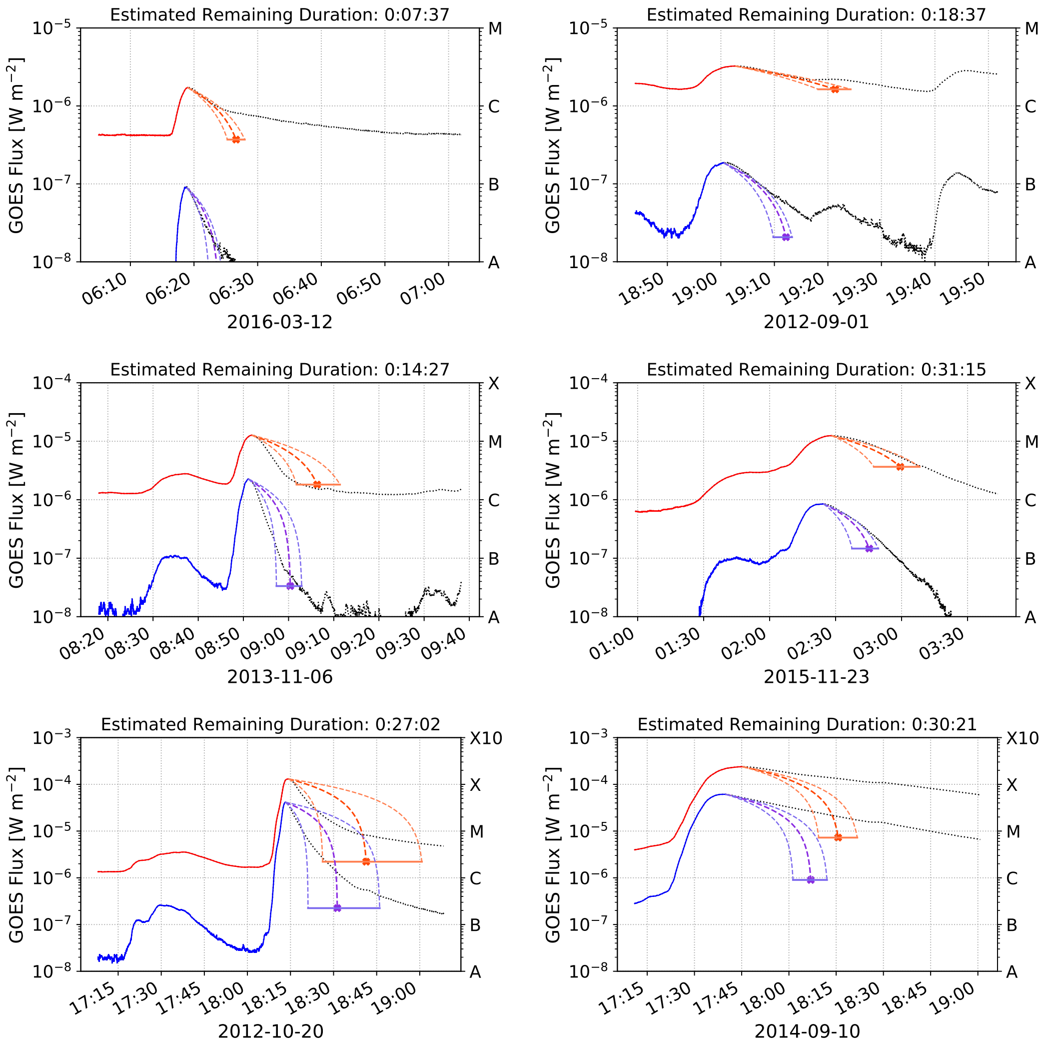


Figure 5. Forecasts of (t_4 - t_2) calculated at the peak for six example flares, with relatively short (left column) and long (right column) total durations. The top row shows two C-class flares, the middle row M-class, and the bottom row X-class. The red lines are for the 1–8 Å channel, and blue for the 0.5–4 Å channel. The middle dashed line that terminates with an X mark is the predicted duration, while the outer lines mark the $\pm 1\sigma$ confidence intervals. The black dotted lines show the true evolution of the light curves. The method tends to predict short duration flares accurately, while it tends to underestimate the duration of longer duration flares.

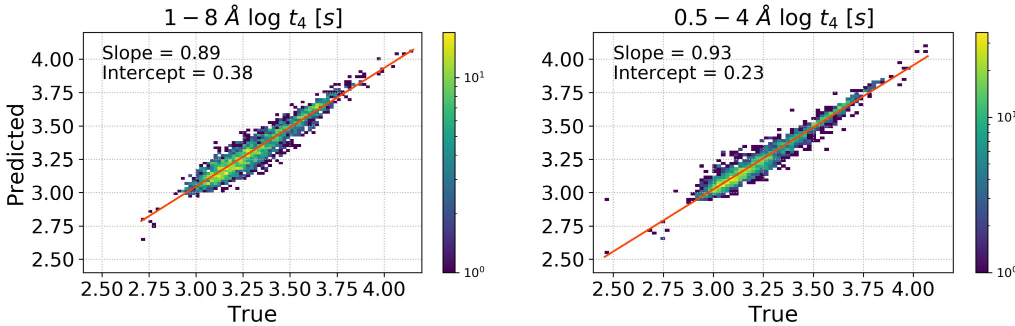


Figure 6. Similar to Figure 4, except that the random forest has now been trained using features through time t_3 to predict the values of t_4. The slopes are closer to 1 in this case, indicating a better fit, though still slightly underestimating the duration in general.

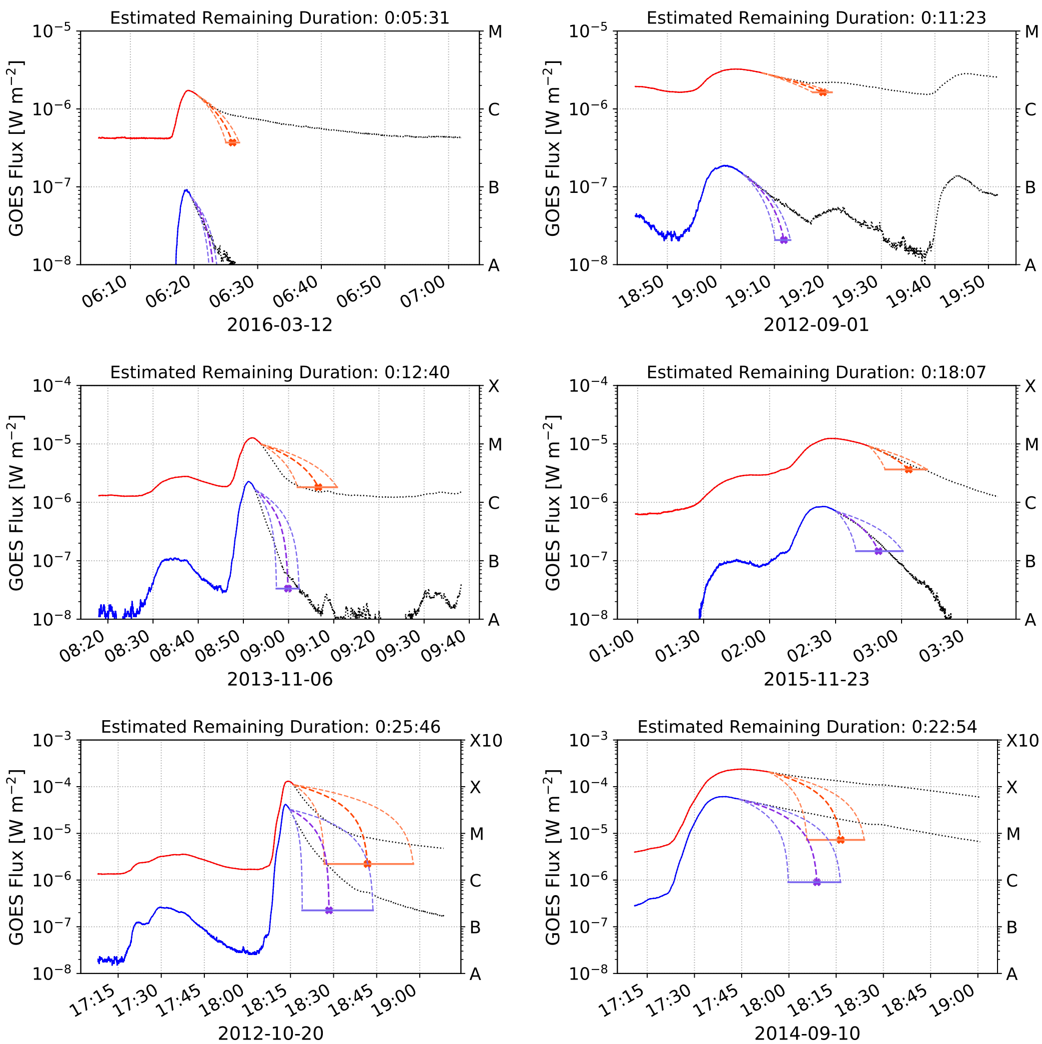


Figure 7. Forecasts of (t_4 - t_3), similar to Figure 5, showing the predictions beginning at time t_3 (the minimum of the first derivative) for the same six events.

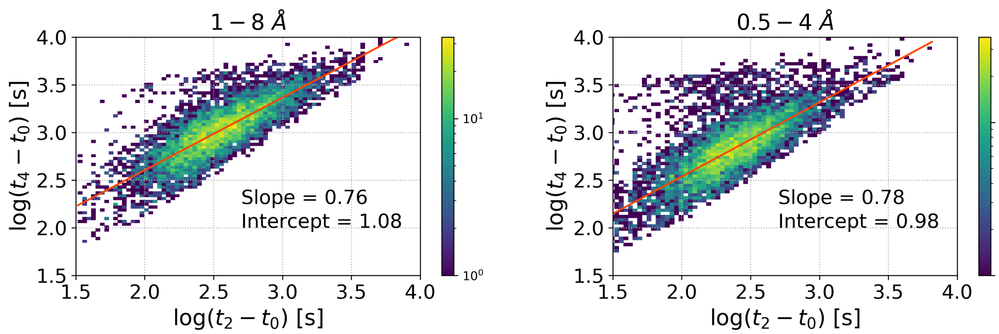


Figure 8. 2D histograms comparing the true values of t_2 and t_4 in the 1–8 Å(left) and 0.5–4 Å(right) channels for the entire data set. The distributions have been fit with a Theil-Sen estimator (orange lines), for which the slope and y-intercepts are indicated. We use this linear regression model to calculate a reference forecast which we show is outperformed by the random forest method.

t_4 for the whole data set for each GOES channel. We also show a Theil-Sen linear regression to the data. With these fits, we then calculate a crude estimate of the final time $t_{4,\mathrm{lin}} = m\,t_2 + b$, for m the slope and b the y-intercept. We then define a skill score as:

$$\mathrm{Skill\,score} = 1 - \frac{MSE_{rf}}{MSE_{lin}} \tag{1}$$

where MSE_{rf} is the mean squared error from the random forest prediction for t_4 and MSE_{lin} is the mean squared error from the linear regression prediction. A skill score of exactly 1 corresponds to a perfect forecast for all events, while a skill score of 0 or less indicates that the random forest regressor was less accurate in predicting flare duration than the linear regressor. We use only the test data to calculate the MSE to avoid biasing the skill score. For the forecasts through time t_2, we find a skill score of 0.056 in the 1–8 Å channel and 0.215 in the 0.5–4 Å channel. For the forecasts through time t_3, these skill scores improve to 0.693 and 0.883, respectively. That is, the random forest slightly outperforms a linear regression when the forecast occurs at time t_2 (the peak flux in each channel), and strongly outperforms the linear regression when forecasted from time t_3 (the minima of the first derivatives). The random forest model provides a reasonable, though imperfect, forecast that statistically tends to underestimate the durations.

4 Conclusions

This model is simple, lightweight, and accurate. It can be run in Python to make direct predictions of the remaining duration on an ongoing solar flare. We have trained the data set using only parameters that are easily measured from GOES/XRS light curves, and which can be easily calculated in real-time. The model generally performs well, with a skill score that outperforms a simpler linear regression model. The most important features for forecasting the remaining duration are the timings of the start and peak as well as the integrals of the flux to the peak.

We should note that, in principle, quasi-periodic pulsations (QPPs; Nakariakov & Melnikov 2009) could be measured from GOES/XRS light curves and that the period of QPPs are correlated with flare duration (Hayes et al., 2020). As noted by Hayes et al. (2020), however, the period changes during the course of a longer flare. Furthermore, the signal-to-noise ratio needs to be relatively large to measure this period, which would exclude forecasting for smaller flares. Finally, it is not clear that all flares exhibit QPPs: a sample of X-class flares detected QPPs in only about 80% of the events (Simões et al., 2015). In future work, it would be worthwhile to build QPP measurements into the model, particularly with newer GOES satellites which have better cadence.

While this model has been built with simplicity in mind, the addition of other data sets would likely improve the prediction. The two XRS channels are generally sensitive only to plasma exceeding approximately 10 MK, and their ratio can be used to estimate temperature (Garcia, 1994). However, they do not actually measure the distribution of plasma at various temperatures, and the combination of this data with data from other instruments might strongly improve forecasts. For example, the Extreme Ultraviolet Variability Experiment (EVE; Woods et al. 2012) onboard the Solar Dynamics Observatory (SDO; Pesnell et al. 2012) provides irradiance measurements of a wide range of spectral lines that form at different temperatures and heights in the solar atmosphere, which could potentially prove useful for these forecasts.

Acknowledgments

J.W.R. was supported by a NASA Living With a Star Grant. W.T.B. was supported by NASA's Hinode project through the National Research Council program. The authors thank Douglas Drob and Dennis Socker for helpful discussions in the development of this work, and what parameters to examine. We thank Laura Hayes for help with SunPy usage, in particular GOES/XRS data in SunPy. The authors also thank the SunPy collaboration for their efforts at making accessible tools for the solar community.

This work has made use of the following packages: `numpy` (Oliphant, 2006), `matplotlib` (Hunter, 2007), `scipy` (Virtanen et al., 2020), `pandas` (Wes McKinney, 2010), `scikit-learn` (Pedregosa et al., 2011), `seaborn` (Waskom et al., 2018), and `sunpy` (The SunPy Community et al., 2020; Mumford et al., 2020).

Data Availability Statement.

The data and routines used to produce the results in this paper are available at `https://github.com/USNavalResearchLaboratory/flare_duration_forecasting` (using v1.0.0, DOI: 10.5281/zenodo.4592403), archived at `https://doi.org/10.5281/zenodo.4592403`. GOES/XRS data are provided by NOAA and were accessed through the SunPy package.

References

Breiman, L. (2001, October). Random forests. *Mach. Learn.*, *45*(1), 532. Retrieved from `https://doi.org/10.1023/A:1010933404324` doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and regression trees*. Taylor & Francis.

Garcia, H. A. (1994, October). Temperature and Emission Measure from Goes Soft X-Ray Measurements. *Solar Physics*, *154*(2), 275-308. doi: 10.1007/BF00681100

Hayes, L. A., Inglis, A. R., Christe, S., Dennis, B., & Gallagher, P. T. (2020, May). Statistical Study of GOES X-Ray Quasi-periodic Pulsations in Solar Flares. *The Astrophysical Journal*, *895*(1), 50. doi: 10.3847/1538-4357/ab8d40

Hudson, H. S. (1991, June). Solar flares, microflares, nanoflares, and coronal heating. *Solar Physics*, *133*(2), 357-369. doi: 10.1007/BF00149894

Hunter, J. D. (2007, May). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Mumford, S. J., Freij, N., Christe, S., Ireland, J., Mayer, F., Hughitt, V. K., … Murray, S. A. (2020, July). *sunpy v2.0.1.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.3940415` doi: 10.5281/zenodo.3940415

Nakariakov, V. M., & Melnikov, V. F. (2009, December). Quasi-Periodic Pulsations in Solar Flares. *Space Science Reviews*, *149*(1-4), 119-151. doi: 10.1007/s11214-009-9536-3

Oliphant, T. (2006). *A Guide to Numpy.* USA: Trelgol Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. (2012, January). The Solar Dynamics Observatory (SDO). *Solar Physics*, *275*(1-2), 3-15. doi: 10.1007/s11207-011-9841-3

Reep, J. W., & Knizhnik, K. J. (2019, April). What Determines the X-Ray Intensity and Duration of a Solar Flare? *Astrophysical Journal*, *874*(2), 157. doi: 10.3847/1538-4357/ab0ae7

Savitzky, A., & Golay, M. J. E. (1964, January). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*, 1627-1639.

Sen, P. K. (1968). Estimates of the regression coefficient based on kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. Retrieved from `http://www.jstor.org/stable/2285891`

Simões, P. J. A., Hudson, H. S., & Fletcher, L. (2015, December). Soft X-Ray Pulsations in Solar Flares. *Solar Physics*, *290*(12), 3625-3639. doi: 10.1007/s11207-015-0691-2

The SunPy Community, Barnes, W. T., Bobra, M. G., Christe, S. D., Freij, N., Hayes, L. A., … Dang, T. K. (2020). The sunpy project: Open source development and status of the version 1.0 core package. *The Astrophysical Journal*, *890*, 68-. Retrieved from `https://iopscience.iop.org/article/10.3847/1538-4357/ab4f7a` doi: 10.3847/1538-4357/ab4f7a

Theil, H. (1992). A rank-invariant method of linear and polynomial regression analysis. In B. Raj & J. Koerts (Eds.), *Henri theil's contributions to economics and econometrics: Econometric theory and methodology* (pp. 345–381). Dordrecht: Springer Netherlands. Retrieved from `https://doi.org/10.1007/978-94-011-2546-8_20` doi: 10.1007/978-94-011-2546-8_20

Toriumi, S., Schrijver, C. J., Harra, L. K., Hudson, H., & Nagashima, K. (2017, Jan). Magnetic Properties of Solar Active Regions That Govern Large Solar

Flares and Eruptions. *Astrophysical Journal*, *834*(1), 56. doi: 10.3847/1538-4357/834/1/56

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... van Mulbregt, P. (2020, February). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 1–12. doi: 10.1038/s41592-019-0686-2

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., ... Qalieh, A. (2018, July). *mwaskom/seaborn v0.9.0.* Retrieved from `https://doi.org/10.5281/zenodo.1313201` doi: 10.5281/zenodo.1313201

Wes McKinney. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (p. 56 - 61). doi: 10.25080/Majora-92bf1922-00a

Woods, T. N., Eparvier, F. G., Hock, R., Jones, A. R., Woodraska, D., Judge, D., ... Viereck, R. (2012, January). Extreme Ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO): Overview of Science Objectives, Instrument Design, Data Products, and Model Developments. *Solar Physics*, *275*(1-2), 115-143. doi: 10.1007/s11207-009-9487-6

Figure 1a.

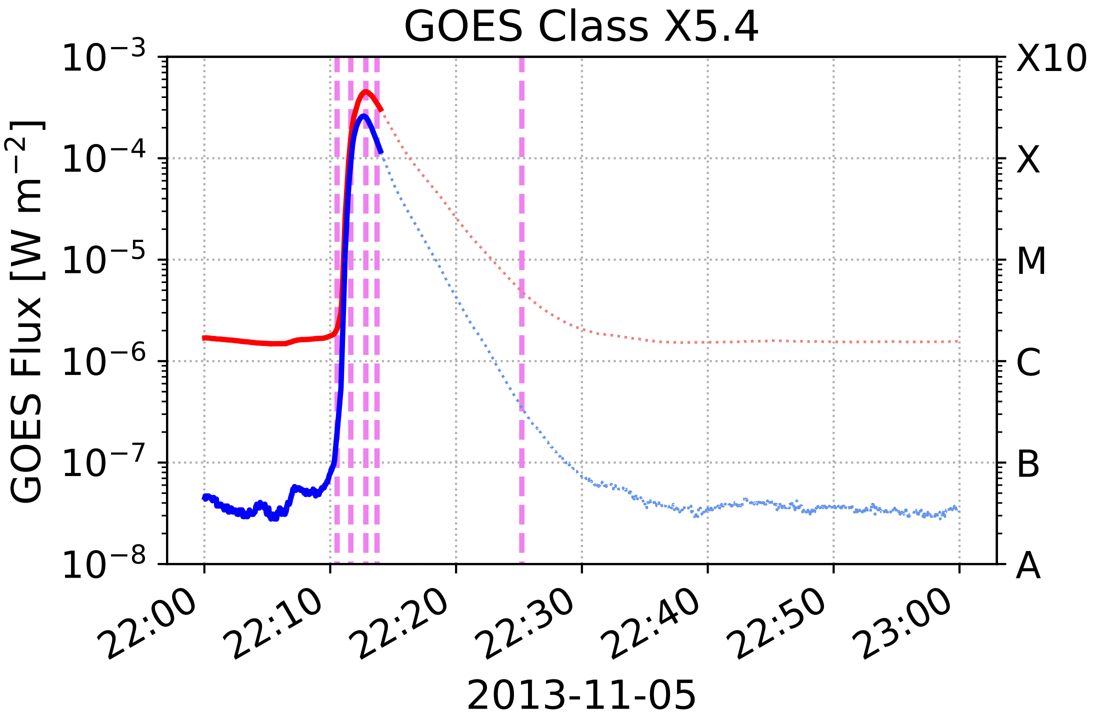
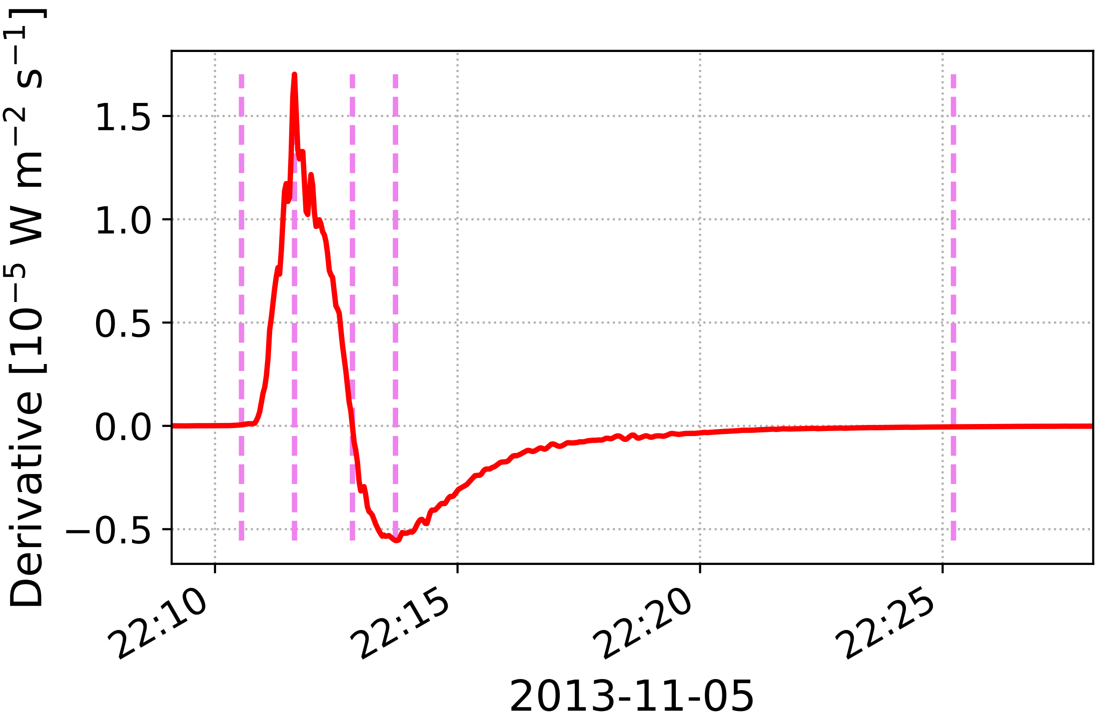
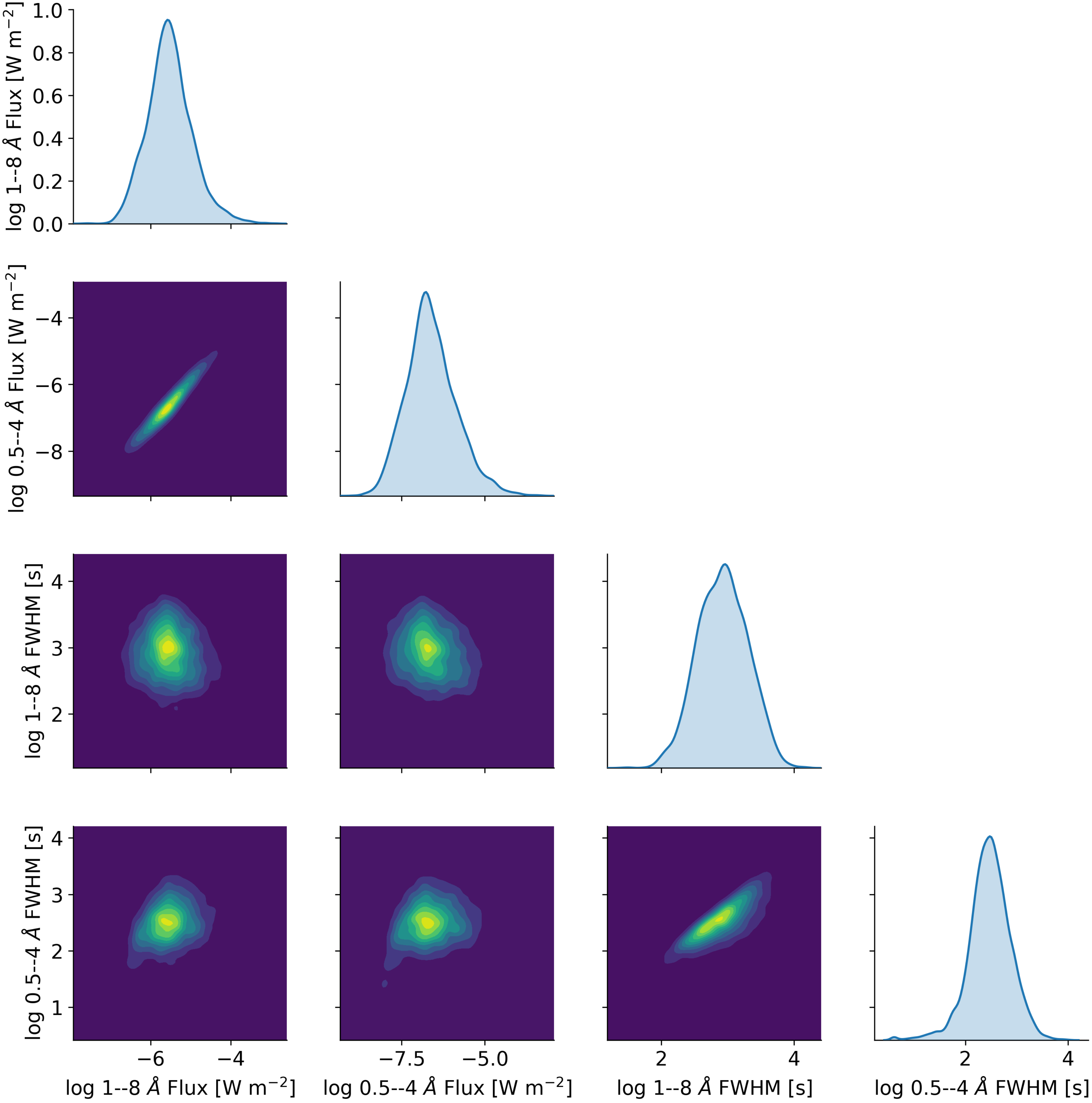

GOES Class X5.4
GOES Flux [W m^{-2}]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X10
X
M
C
B
A
22:00
22:10
22:20
22:30
22:40
22:50
23:00
2013-11-05

Figure 1b.



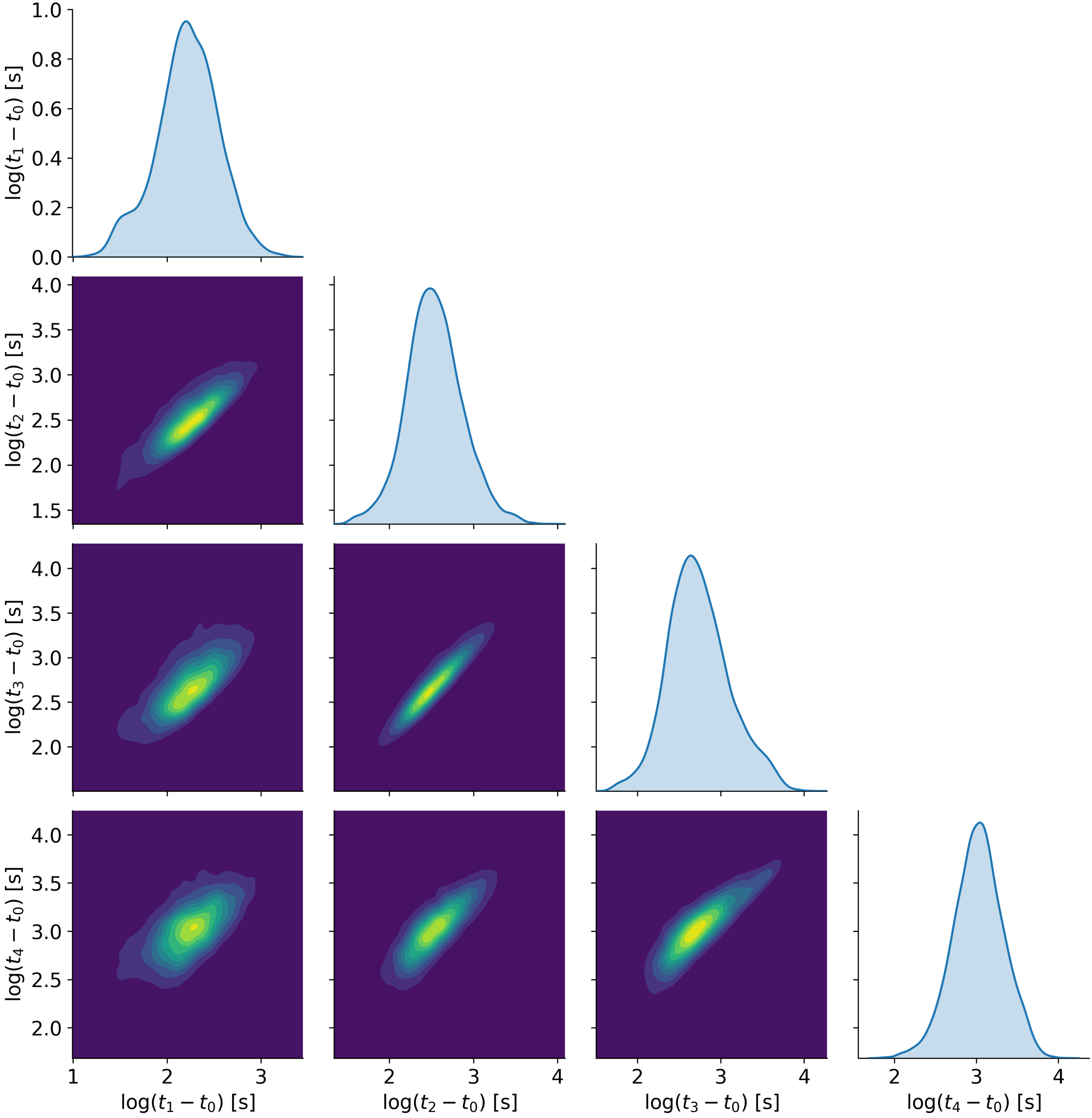
Derivative [10^{-5} W m^{-2} s^{-1}]
1.5
1.0
0.5
0.0
−0.5
22:10
22:15
22:20
22:25
2013-11-05

Figure 2.



log 1--8 Å Flux [W m^{-2}]
log 0.5--4 Å Flux [W m^{-2}]
log 1--8 Å FWHM [s]
log 0.5--4 Å FWHM [s]

Figure 3.



log(t_1 − t_0) [s]
log(t_2 − t_0) [s]
log(t_3 − t_0) [s]
log(t_4 − t_0) [s]
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1
2
3
4

Figure 4a.

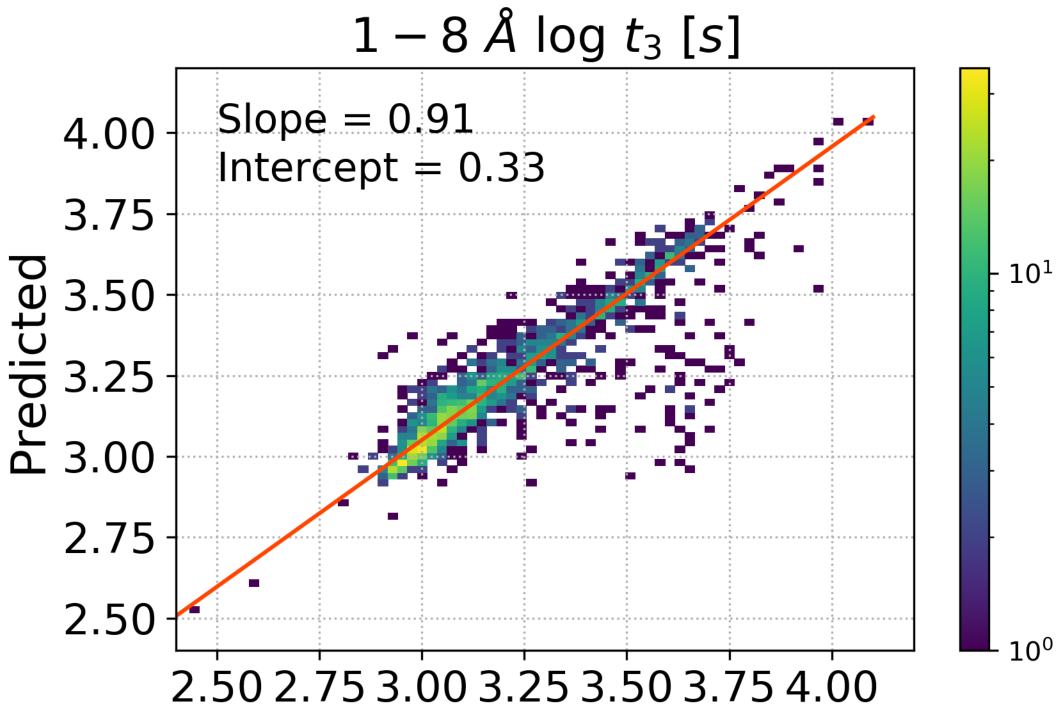
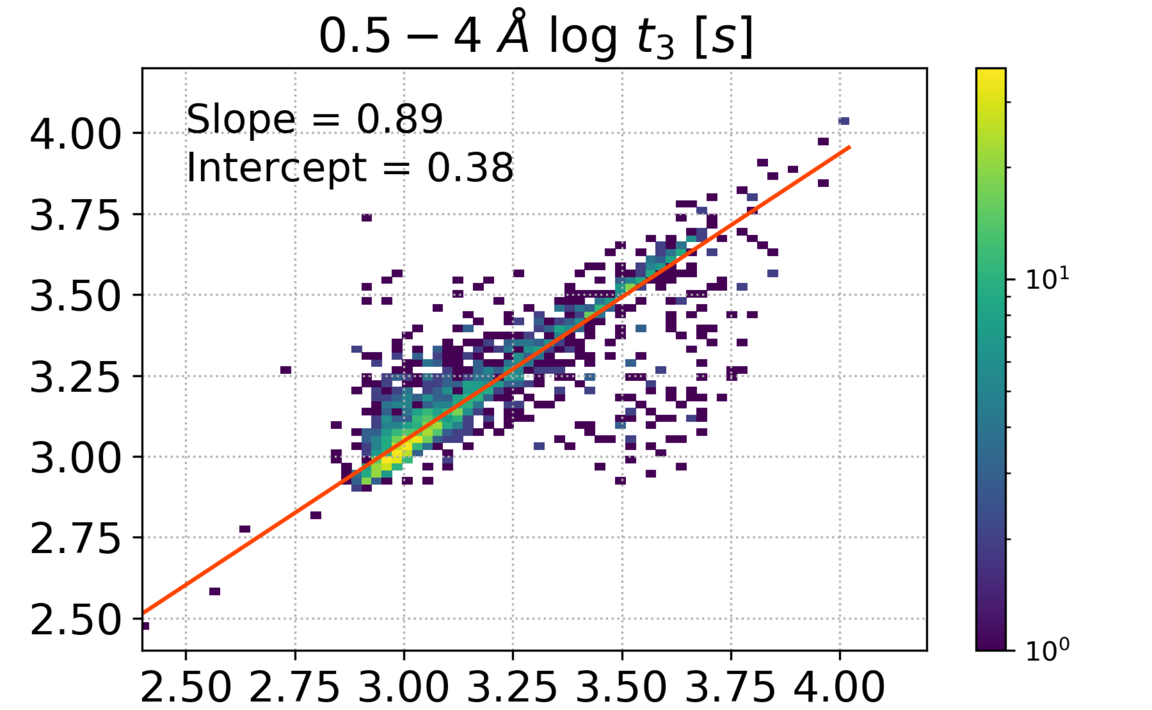
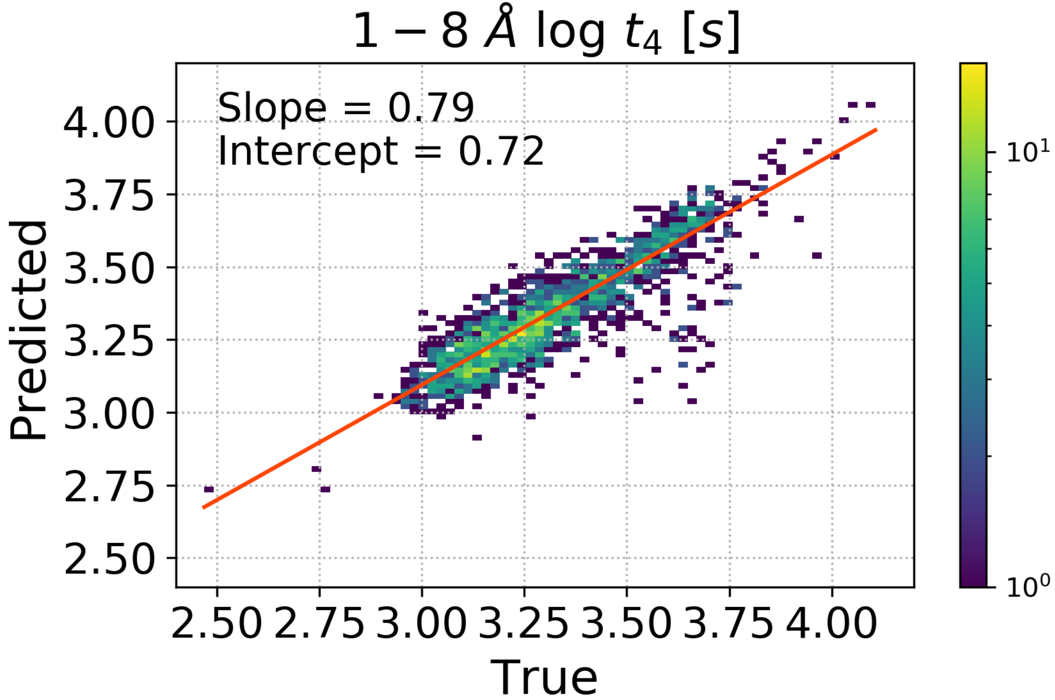

$1-8\ \AA\ \log t_3\ [s]$
Slope = 0.91
Intercept = 0.33
Predicted
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.50 2.75 3.00 3.25 3.50 3.75 4.00
10^1
10^0

Figure 4b.



$0.5 - 4$ Å $\log t_3$ $[s]$
Slope = 0.89
Intercept = 0.38
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.50 2.75 3.00 3.25 3.50 3.75 4.00
10^1
10^0

Figure 4c.



$1 - 8\ \AA\ \log t_4\ [s]$
Slope = 0.79
Intercept = 0.72
Predicted
True
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.50 2.75 3.00 3.25 3.50 3.75 4.00
10^1
10^0

Figure 4d.

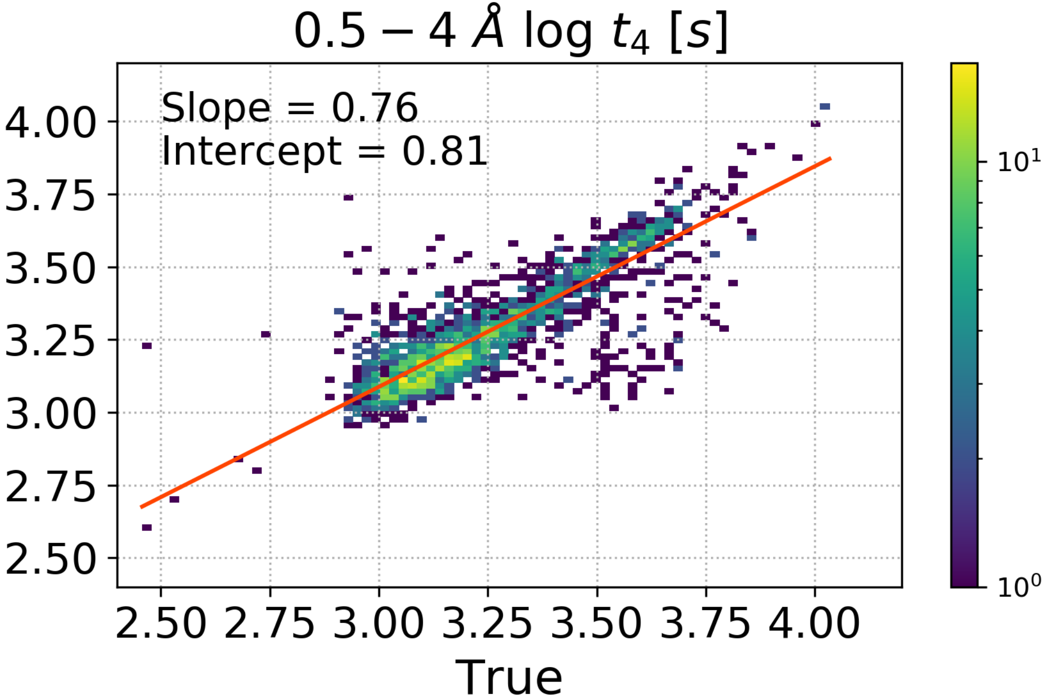
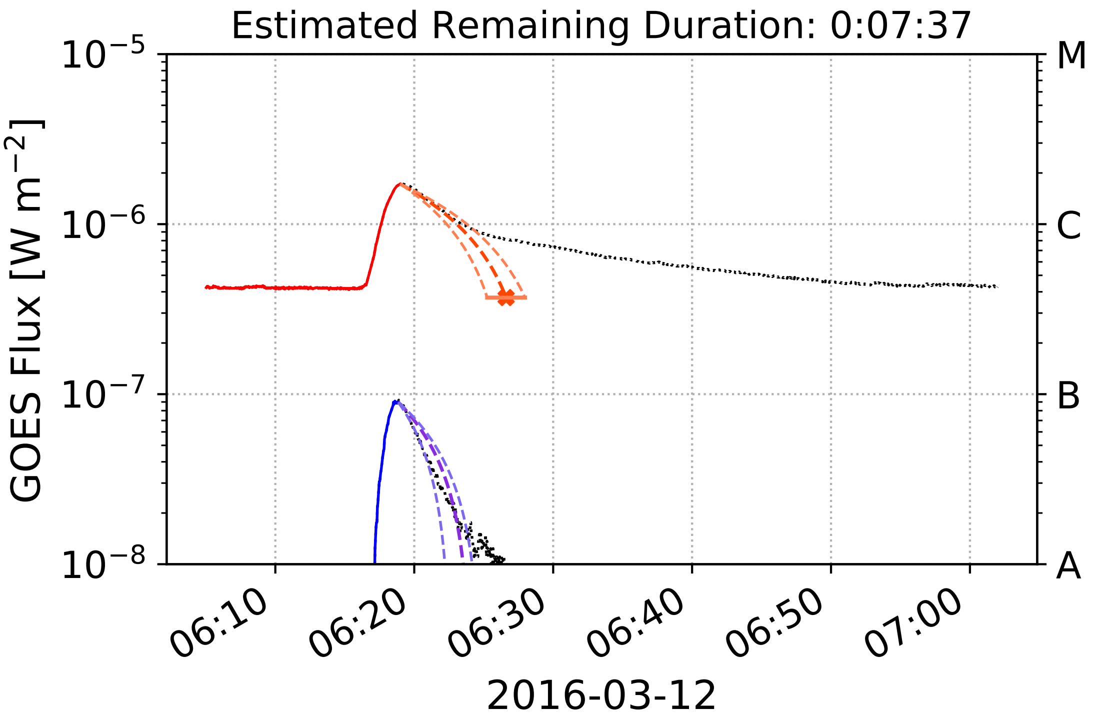
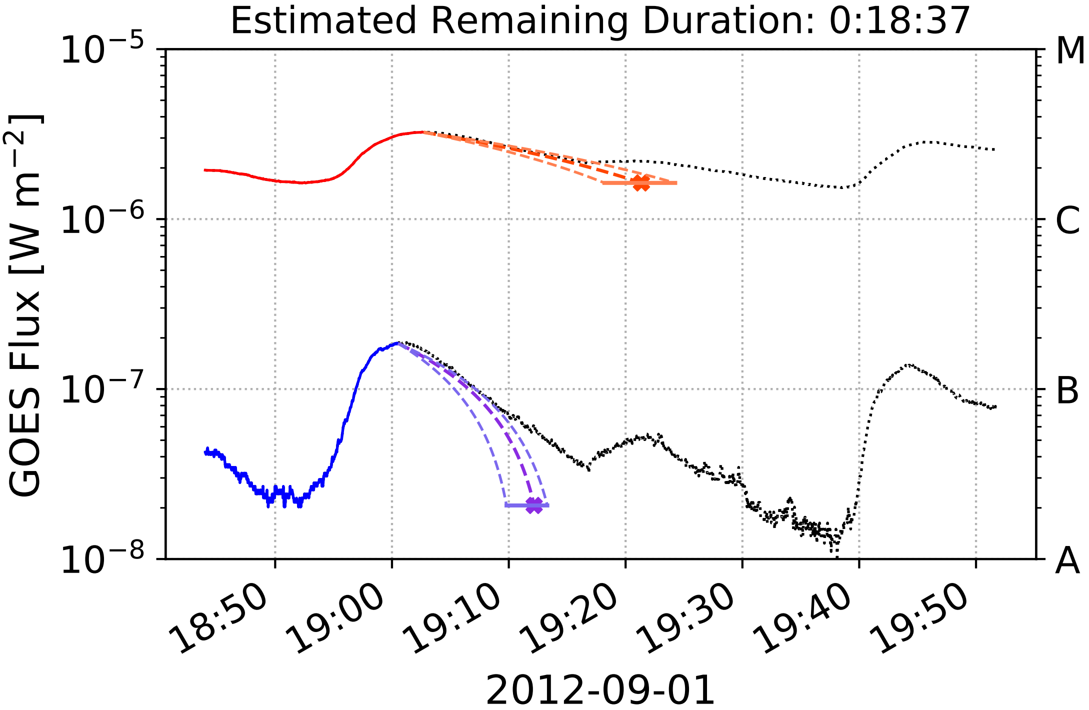

$0.5 - 4$ Å $\log t_4$ $[s]$
Slope = 0.76
Intercept = 0.81
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.50
2.75
3.00
3.25
3.50
3.75
4.00
True
10^1
10^0

Figure 5a.



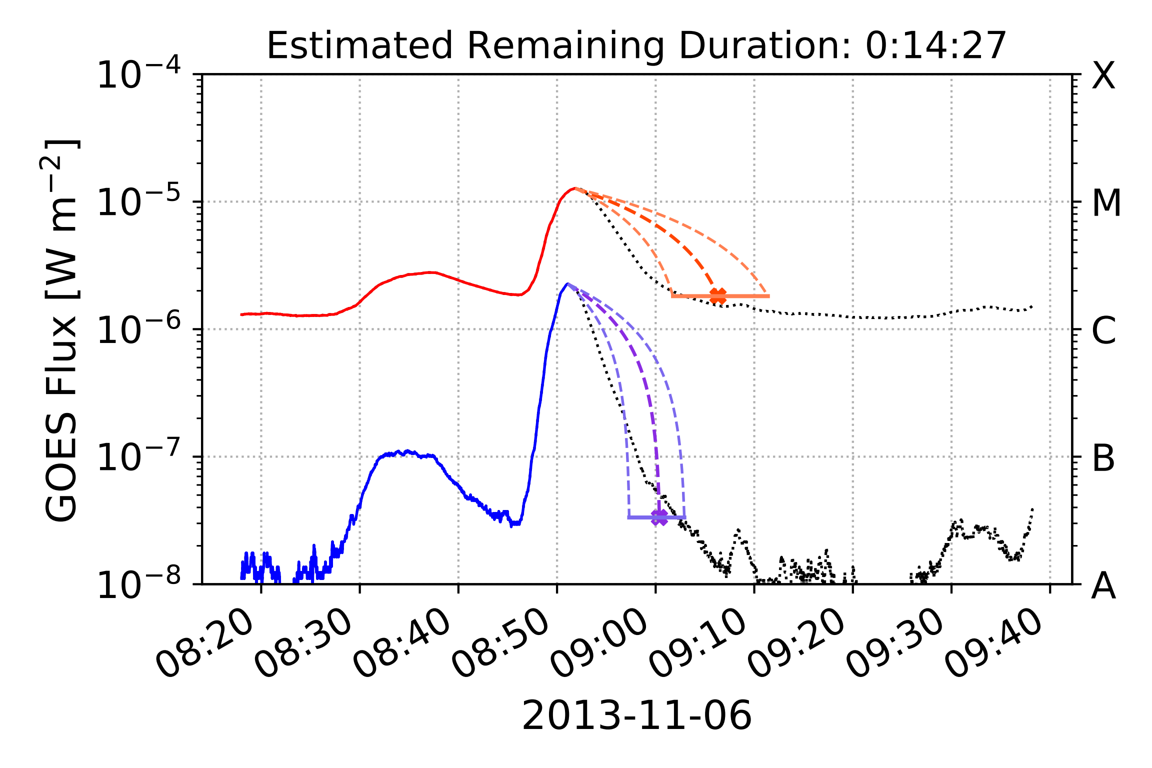
Estimated Remaining Duration: 0:07:37
GOES Flux [W m^{-2}]
10^{-5}
10^{-6}
10^{-7}
10^{-8}
M
C
B
A
06:10
06:20
06:30
06:40
06:50
07:00
2016-03-12

Figure 5b.



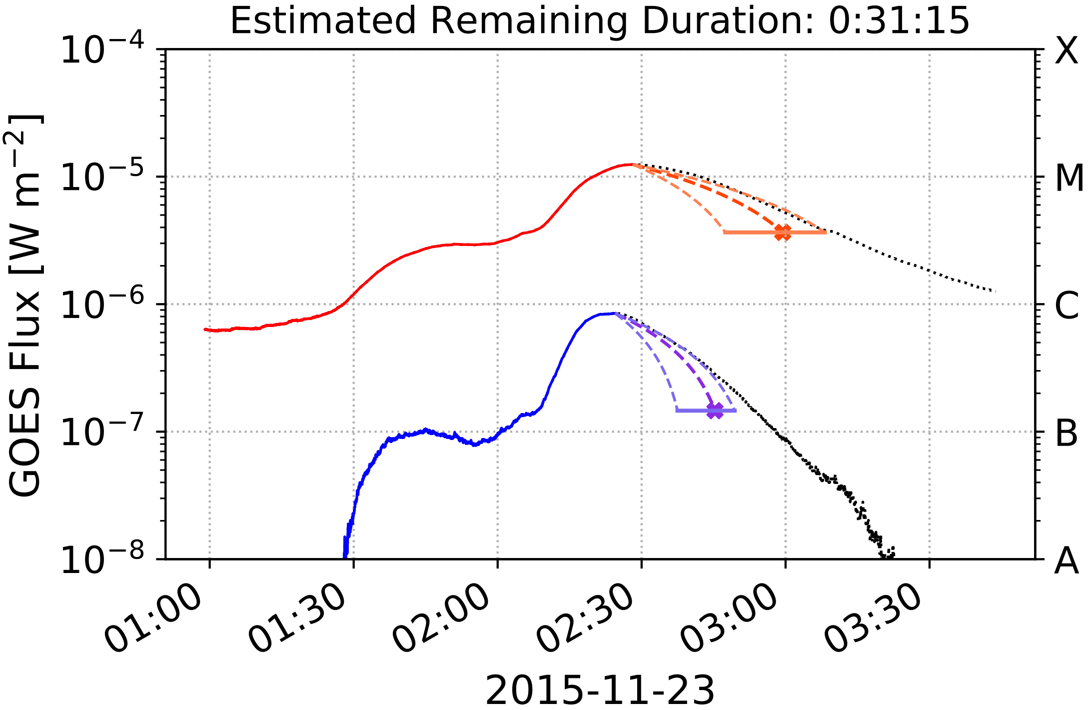
Estimated Remaining Duration: 0:18:37
GOES Flux [W m^{-2}]
10^{-5}
10^{-6}
10^{-7}
10^{-8}
M
C
B
A
18:50
19:00
19:10
19:20
19:30
19:40
19:50
2012-09-01

Figure 5c.



Estimated Remaining Duration: 0:14:27
GOES Flux [W m^{-2}]
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X
M
C
B
A
08:20
08:30
08:40
08:50
09:00
09:10
09:20
09:30
09:40
2013-11-06

Figure 5d.



Estimated Remaining Duration: 0:31:15
GOES Flux [W m^{-2}]
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X
M
C
B
A
01:00
01:30
02:00
02:30
03:00
03:30
2015-11-23

Figure 5e.

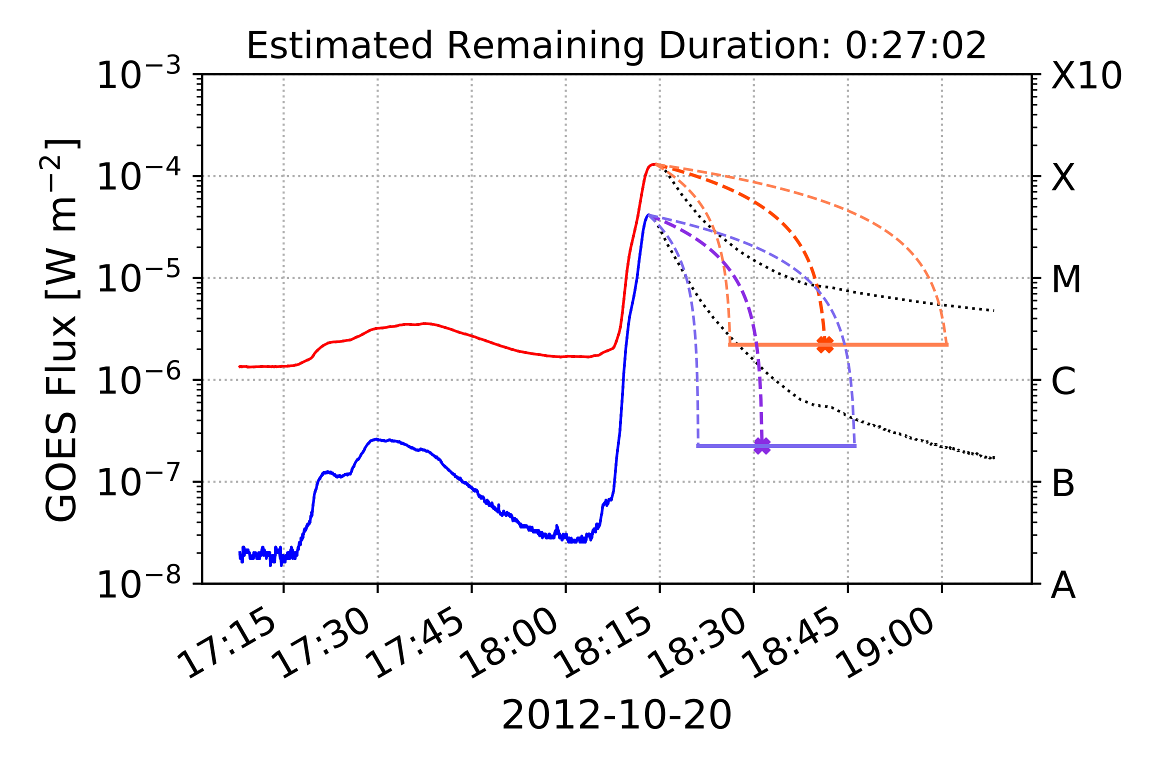
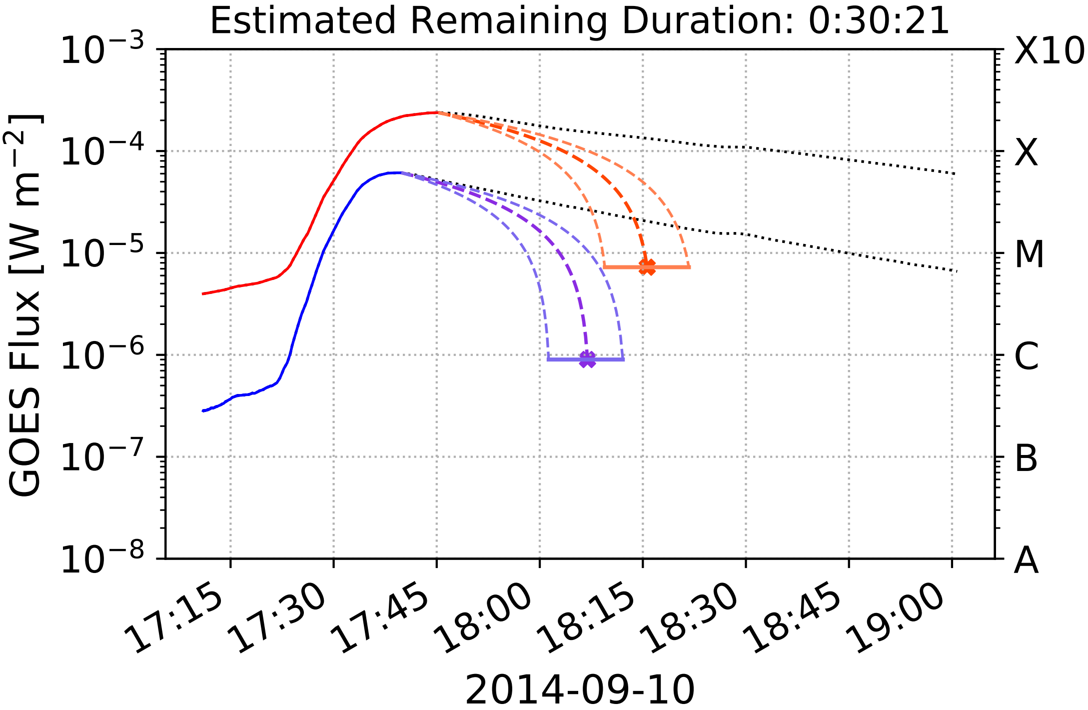
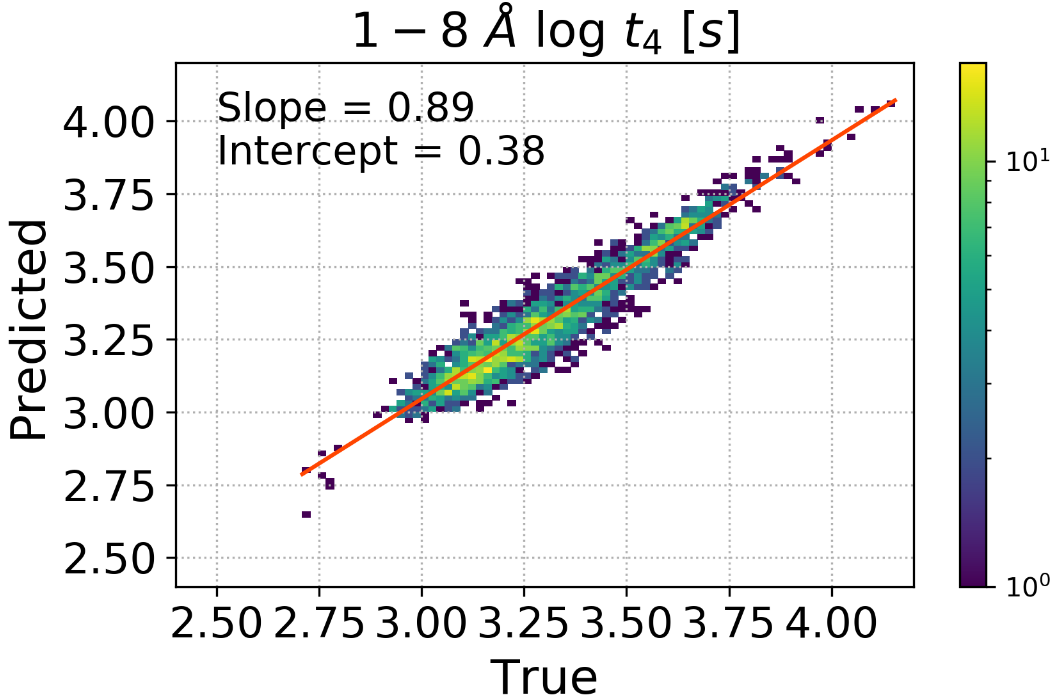

Estimated Remaining Duration: 0:27:02
GOES Flux [W m^{-2}]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X10
X
M
C
B
A
17:15
17:30
17:45
18:00
18:15
18:30
18:45
19:00
2012-10-20

Figure 5f.



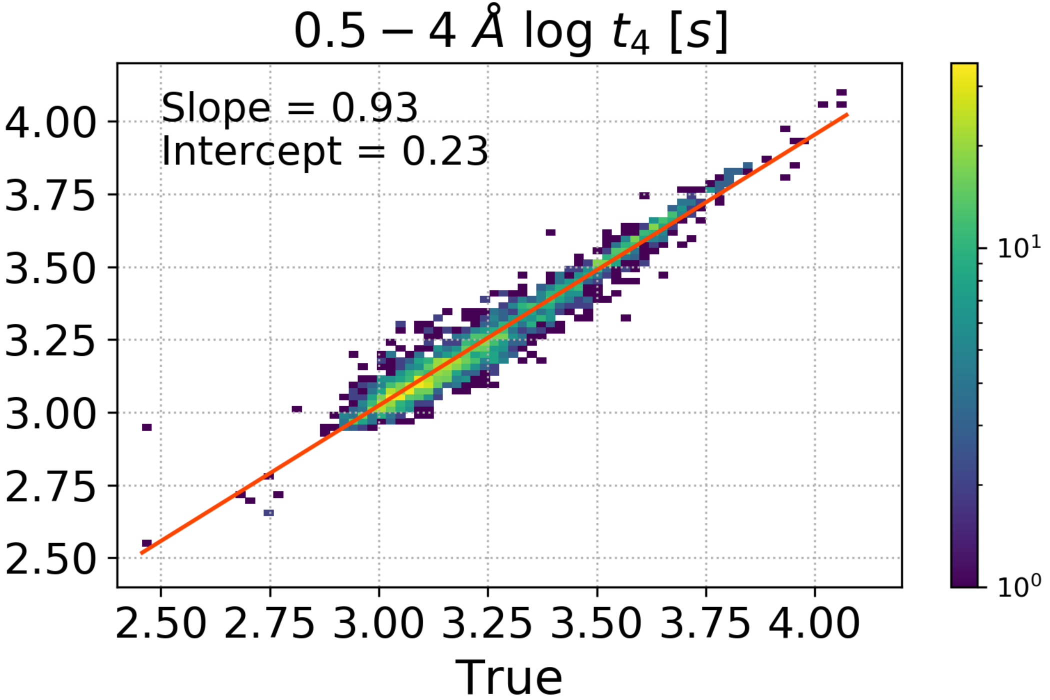
Estimated Remaining Duration: 0:30:21
GOES Flux [W m^{-2}]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X10
X
M
C
B
A
17:15
17:30
17:45
18:00
18:15
18:30
18:45
19:00
2014-09-10

Figure 6a.



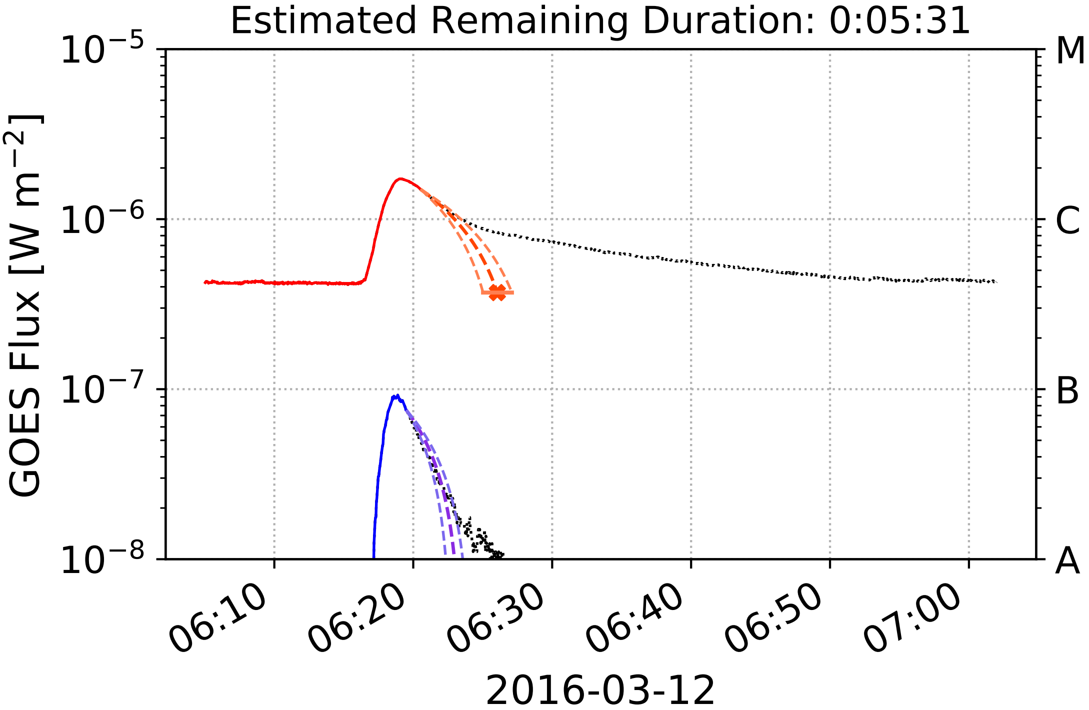
$1-8\ \AA\ \log t_4\ [s]$
Slope = 0.89
Intercept = 0.38
Predicted
True
2.50
2.75
3.00
3.25
3.50
3.75
4.00
10^1
10^0

Figure 6b.



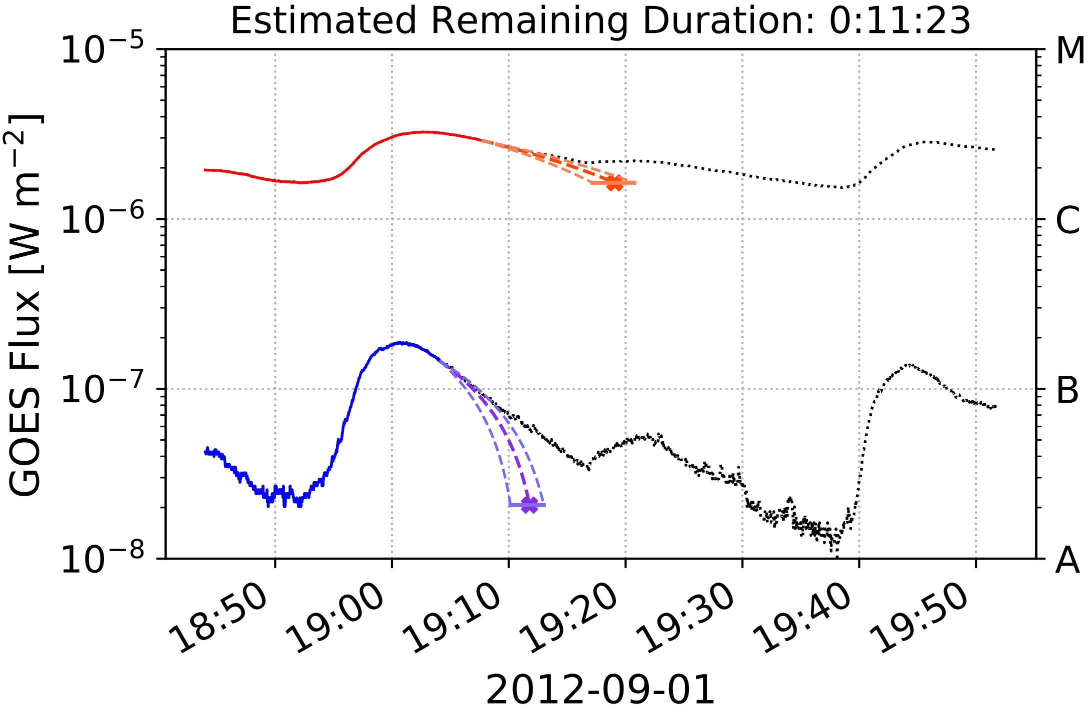
$0.5 - 4\ \AA\ \log\ t_4\ [s]$
Slope = 0.93
Intercept = 0.23
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.50 2.75 3.00 3.25 3.50 3.75 4.00
True
10^1
10^0

Figure 7a.



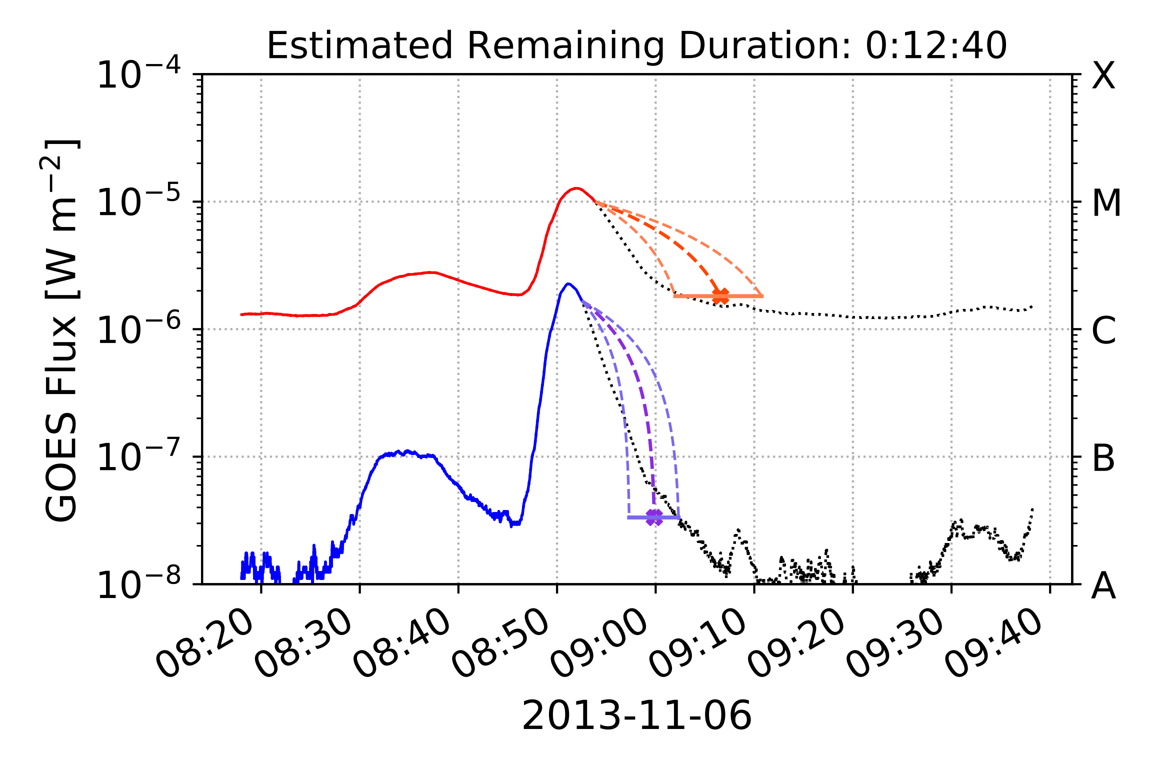
Estimated Remaining Duration: 0:05:31
GOES Flux [W m^{-2}]
10^{-5}
10^{-6}
10^{-7}
10^{-8}
M
C
B
A
06:10
06:20
06:30
06:40
06:50
07:00
2016-03-12

Figure 7b.



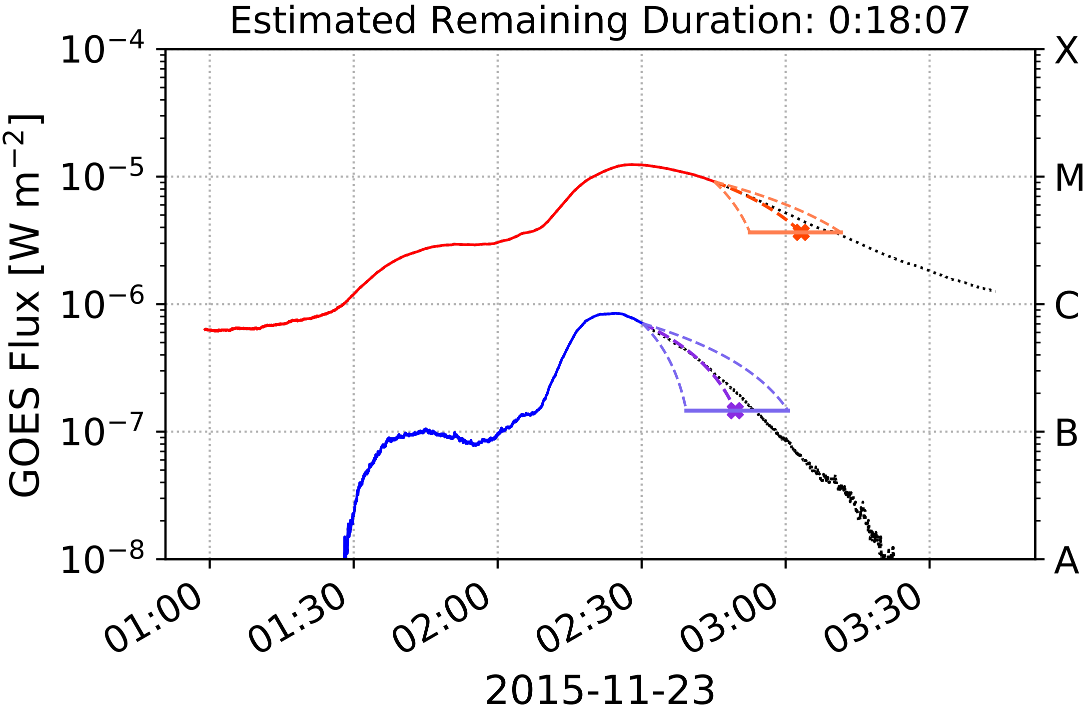
Estimated Remaining Duration: 0:11:23
GOES Flux [W m^{-2}]
10^{-5}
10^{-6}
10^{-7}
10^{-8}
M
C
B
A
18:50
19:00
19:10
19:20
19:30
19:40
19:50
2012-09-01

Figure 7c.



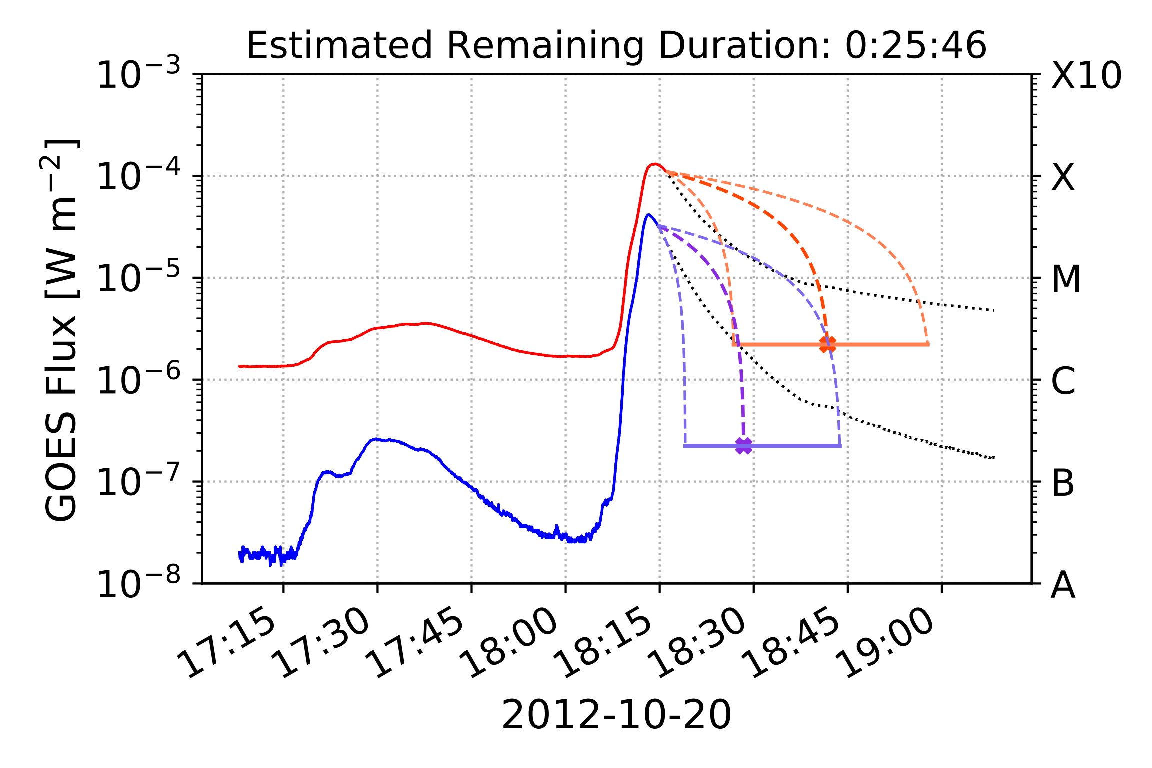
Estimated Remaining Duration: 0:12:40
GOES Flux [W m^{-2}]
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X
M
C
B
A
08:20
08:30
08:40
08:50
09:00
09:10
09:20
09:30
09:40
2013-11-06

Figure 7d.



Estimated Remaining Duration: 0:18:07
GOES Flux [W m^{-2}]
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X
M
C
B
A
01:00
01:30
02:00
02:30
03:00
03:30
2015-11-23

Figure 7e.



Estimated Remaining Duration: 0:25:46
GOES Flux [W m^{-2}]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X10
X
M
C
B
A
17:15
17:30
17:45
18:00
18:15
18:30
18:45
19:00
2012-10-20

Figure 7f.

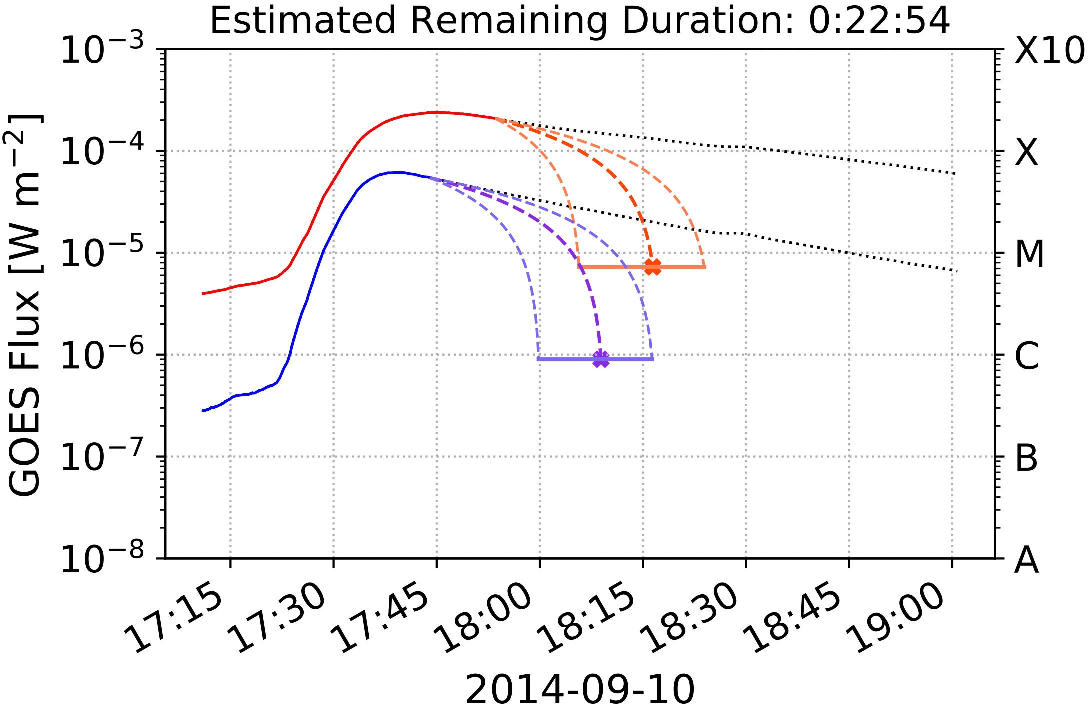
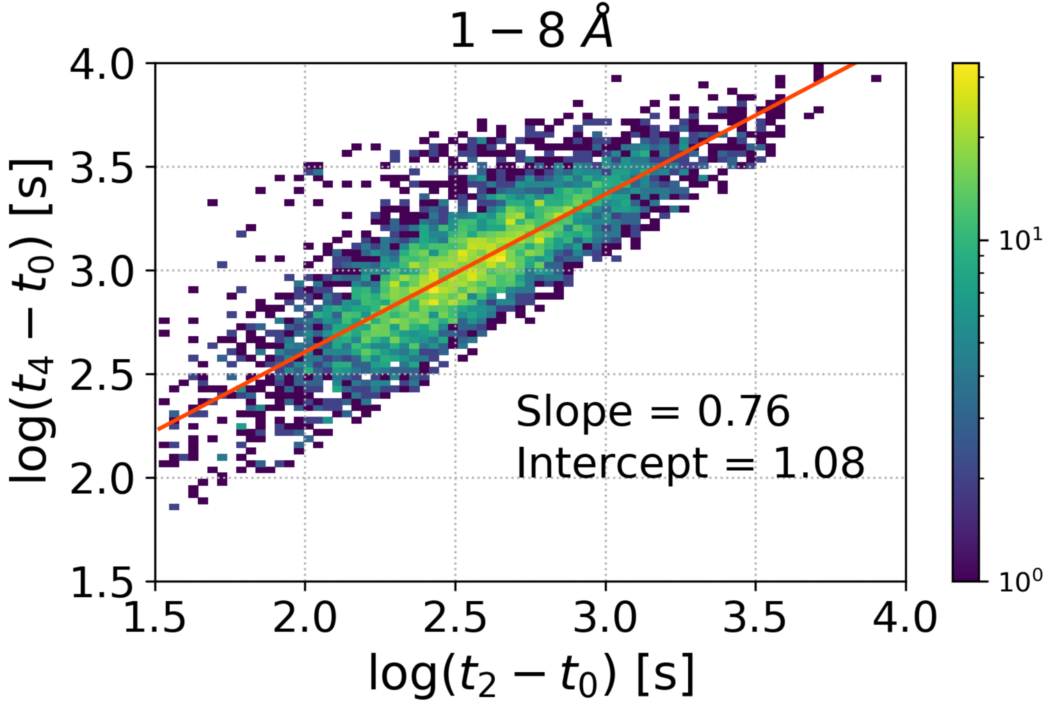
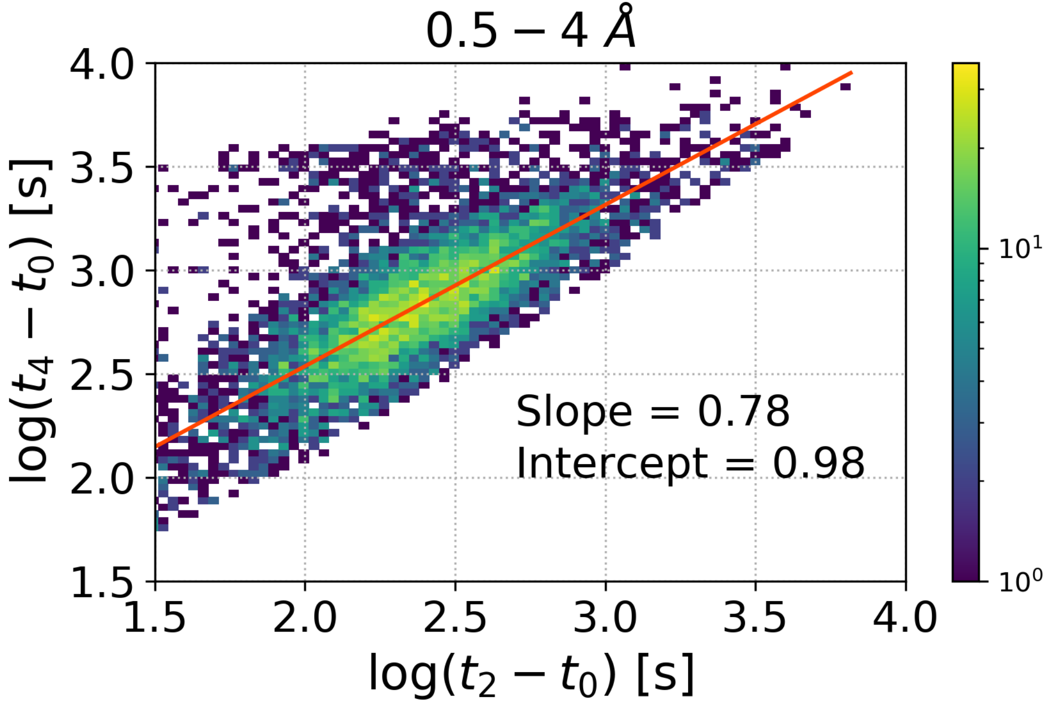

Estimated Remaining Duration: 0:22:54
GOES Flux [W m^{-2}]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
X10
X
M
C
B
A
17:15
17:30
17:45
18:00
18:15
18:30
18:45
19:00
2014-09-10

Figure 8a.



$1-8\ \AA$
Slope = 0.76
Intercept = 1.08
$\log(t_4 - t_0)$ [s]
$\log(t_2 - t_0)$ [s]
4.0
3.5
3.0
2.5
2.0
1.5
1.5
2.0
2.5
3.0
3.5
4.0
10^1
10^0

Figure 8b.



$0.5 - 4\ \AA$
Slope = 0.78
Intercept = 0.98
$\log(t_2 - t_0)$ [s]
$\log(t_4 - t_0)$ [s]
1.5
2.0
2.5
3.0
3.5
4.0
10^0
10^1