Solar Flare Prediction with LSTM and Polarity Inversion Line Detection

Hu Sun

Department of Statistics
University of Michigan, Ann Arbor

September 27, 2019

Solar Flare Prediction with LSTM

Data source:

- Space-Weather HMI-Active Region Patch (SHARP) features, with 12 min cadence.
- GOES dataset from 2010-2018

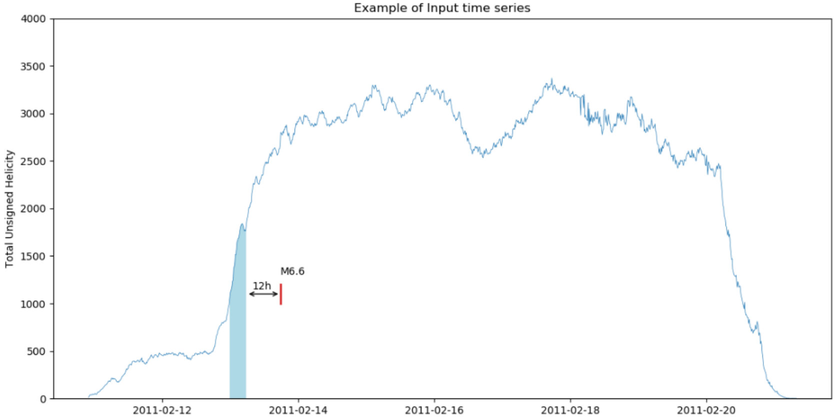
Solar Flare Prediction with LSTM

Parameter	Description
TOTUSJH:	Total unsigned current helicity
TOTUSJZ:	Total unsigned vertical current
SAVNCPP:	Sum of the modulus of the net current per polarity
USFLUX:	Total unsigned flux
ABSNJZH:	Absolute value of the net current helicity
TOTPOT:	Proxy for total photospheric magnetic free energy density
SIZE ACR:	De-projected area of active pixels (B_z magnitude larger than noise threshold) on image in micro-hemisphere (defined as one millionth of half the surface of the Sun)
NACR:	The number of strong LoS magnetic-field pixels in the patch
MEANPOT:	Proxy for mean photospheric excess magnetic energy density
SIZE:	Projected area of the image in micro-hemispheres
MEANJZH:	Current helicity (B_z contribution)
SHRGT45:	Fraction of area with shear > 45°
MEANSHR:	Mean shear angle
MEANJZD:	Vertical current density
MEANALP:	Characteristic twist parameter, α
MEANGBT:	Horizontal gradient of total field
MEANGAM:	Mean angle of field from radial
MEANGBZ:	Horizontal gradient of vertical field
MEANGBH:	Horizontal gradient of horizontal field

Input Time Series

There are two parameters that can configure the input time series. **Data length** and the **Prediction time**.

Input Time Series



Input Time Series

We have tried data length to be 1, 6, 12, 24 hours and prediction time to be 1, 6, 12, 24, 48, 72 hours.

Machine Learning Task

Task: With all SHARP physical quantities' time series of 1/6/12/24 hours length as input, classify whether the flare 1/6/12/24/48/72 hours later is an M/X class flare or a B-class flare.

Machine Learning Task

Hours Before an Event	1 hour				6 hours				12 hours			
Hours of Data for Training	1	6	12	24	1	6	12	24	1	6	12	24
Num. Strong Flares	585	579	565	543	579	565	559	529	565	559	546	510
Num. Weak Flares	851	838	814	768	838	817	794	749	814	794	769	726
Num. ARs	632	628	618	606	628	619	612	601	618	612	608	588
Hours Before an Event	24 hours				48 hours				72 hours			
Hours of Data for Training	1	6	12	24	1	6	12	24	1	6	12	24
Num. Strong Flares	543	529	510	480	475	463	453	423	422	412	403	382
Num. Weak Flares	768	749	726	669	660	631	609	564	560	545	524	476
Num. ARs	606	601	588	567	563	552	542	520	518	512	504	485

Machine Learning Task

Model variations include:

- M/X vs Quite Time
- First M/X flare vs Quiet Time
- First Flare (X/M/C/B) vs Quiet Time
- X vs M vs C vs B intensity regression (w/o Quiet Time)
- Predict the intensity of the strongest flare in the next 24/48 hours.

Machine Learning Task

Model variations include:

- M/X vs Quite Time
- First M/X flare vs Quiet Time
- First Flare (X/M/C/B) vs Quiet Time
- X vs M vs C vs B intensity regression (w/o Quiet Time)
- Predict the intensity of the strongest flare in the next 24/48 hours.

Machine Learning Task

Model variations include:

- M/X vs Quite Time
- First M/X flare vs Quiet Time
- First Flare (X/M/C/B) vs Quiet Time
- X vs M vs C vs B intensity regression (w/o Quiet Time)
- Predict the intensity of the strongest flare in the next 24/48 hours.

Machine Learning Task

Model variations include:

- M/X vs Quite Time
- First M/X flare vs Quiet Time
- First Flare (X/M/C/B) vs Quiet Time
- X vs M vs C vs B intensity regression (w/o Quiet Time)
- Predict the intensity of the strongest flare in the next 24/48 hours.

Machine Learning Task

Model variations include:

- M/X vs Quite Time
- First M/X flare vs Quiet Time
- First Flare (X/M/C/B) vs Quiet Time
- X vs M vs C vs B intensity regression (w/o Quiet Time)
- Predict the intensity of the strongest flare in the next 24/48 hours.

Train-Test Set Splitting

Three types of train-test set split, all with roughly two-thirds of the samples in the training set and the rest in testing set:

- Random Split
- Split by Active Region
- Split by Year

Data standardization is done only on the training set. When validating the test set, we still use mean and standard deviation from the training set.

Train-Test Set Splitting

Three types of train-test set split, all with roughly two-thirds of the samples in the training set and the rest in testing set:

- Random Split
- Split by Active Region
- Split by Year

Data standardization is done only on the training set. When validating the test set, we still use mean and standard deviation from the training set.

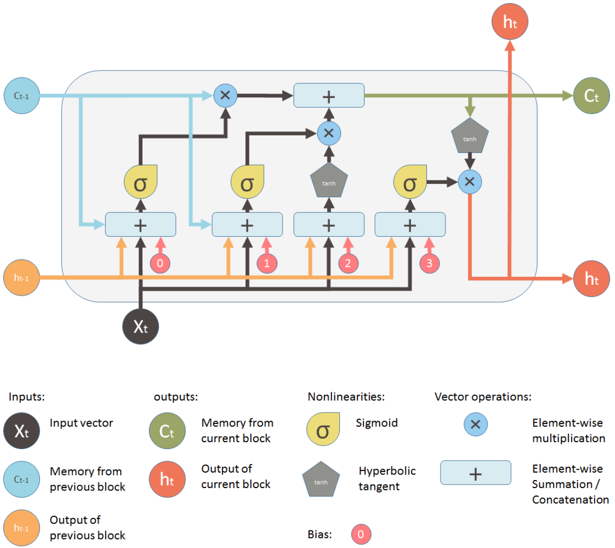
Train-Test Set Splitting

Three types of train-test set split, all with roughly two-thirds of the samples in the training set and the rest in testing set:

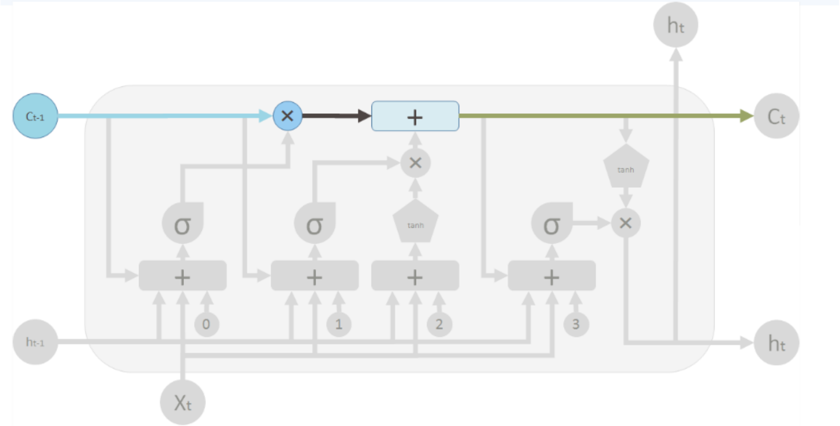
- Random Split
- Split by Active Region
- Split by Year

Data standardization is done only on the training set. When validating the test set, we still use mean and standard deviation from the training set.

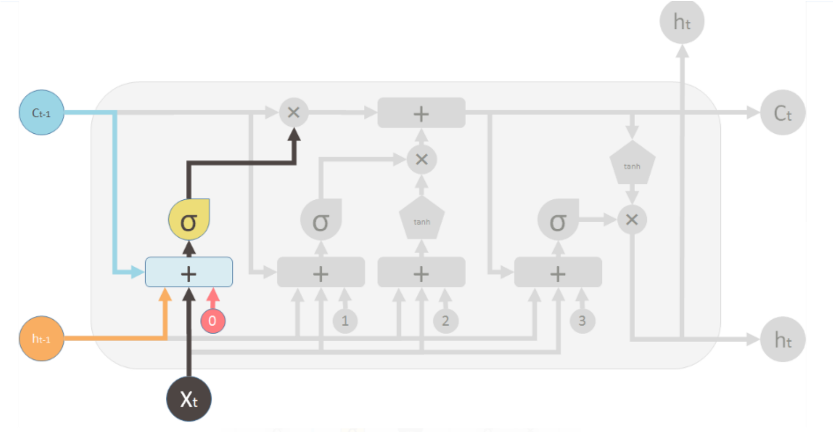
About LSTM Model



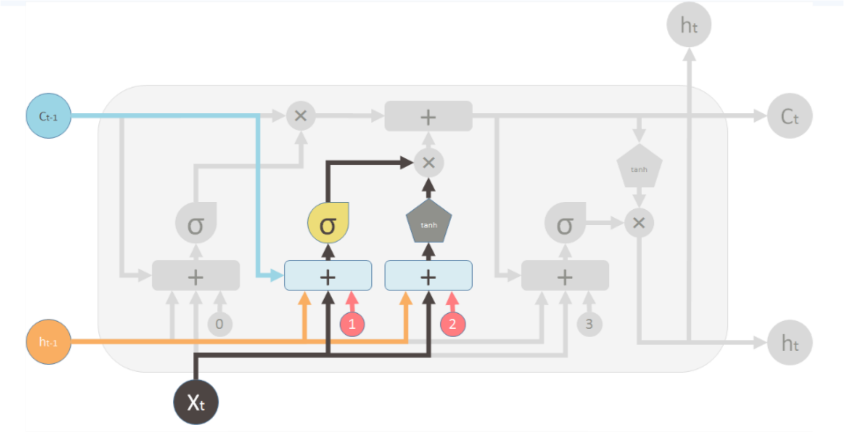
About LSTM Model



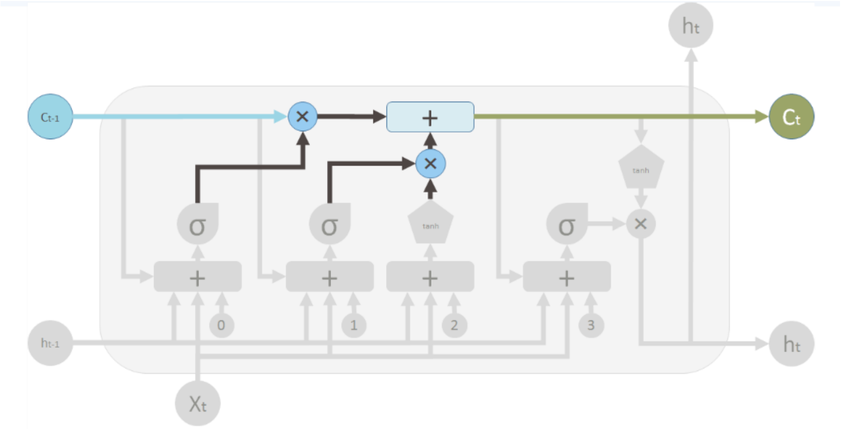
About LSTM Model



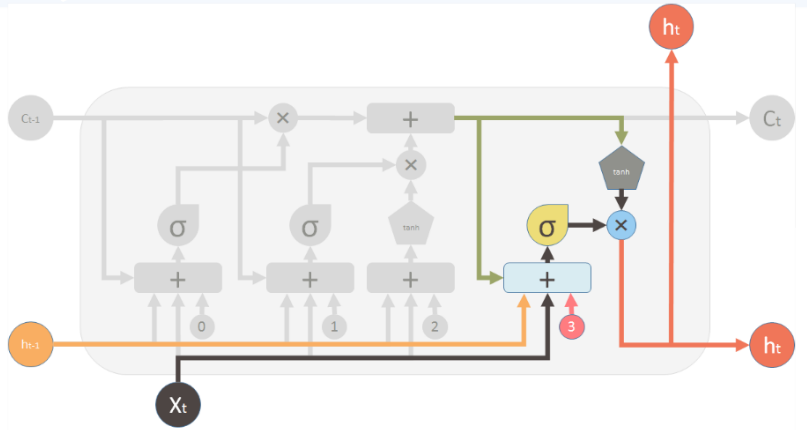
About LSTM Model



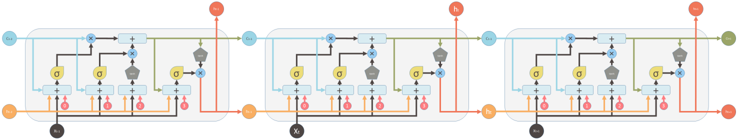
About LSTM Model



About LSTM Model



About LSTM Model



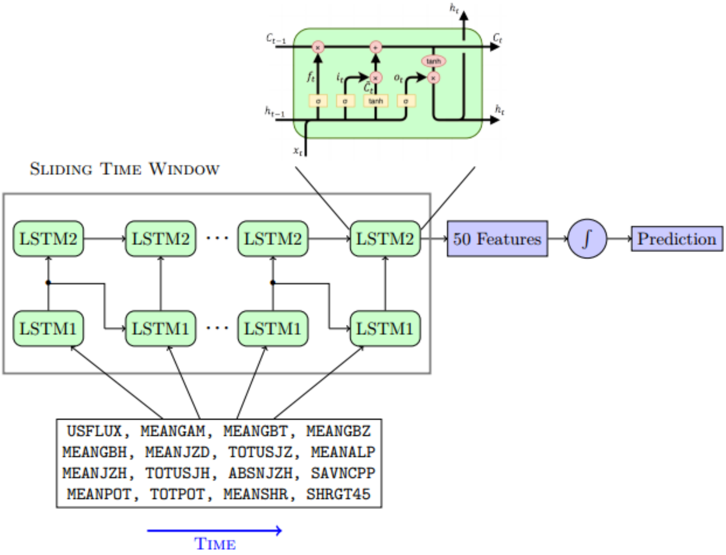
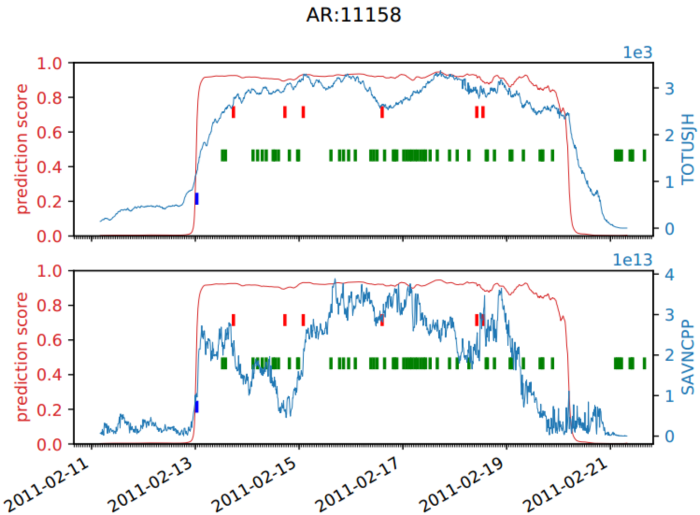


Figure: LSTM Model

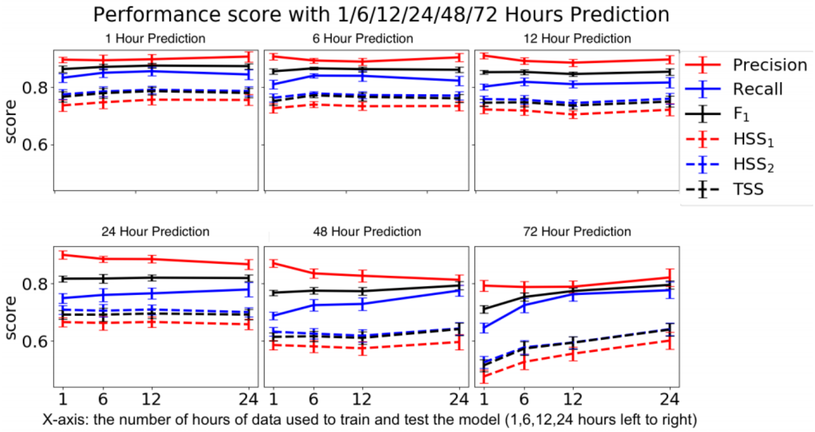
Prediction Score example

To do prediction, we can simply plug in any time series with the same data length as the input data, and get a prediction score which indicates the probability of seeing an M/X class flare at the prediction time.

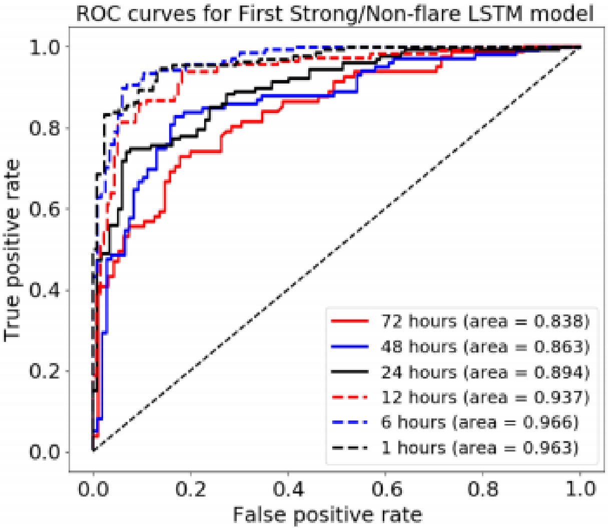
Prediction Score example

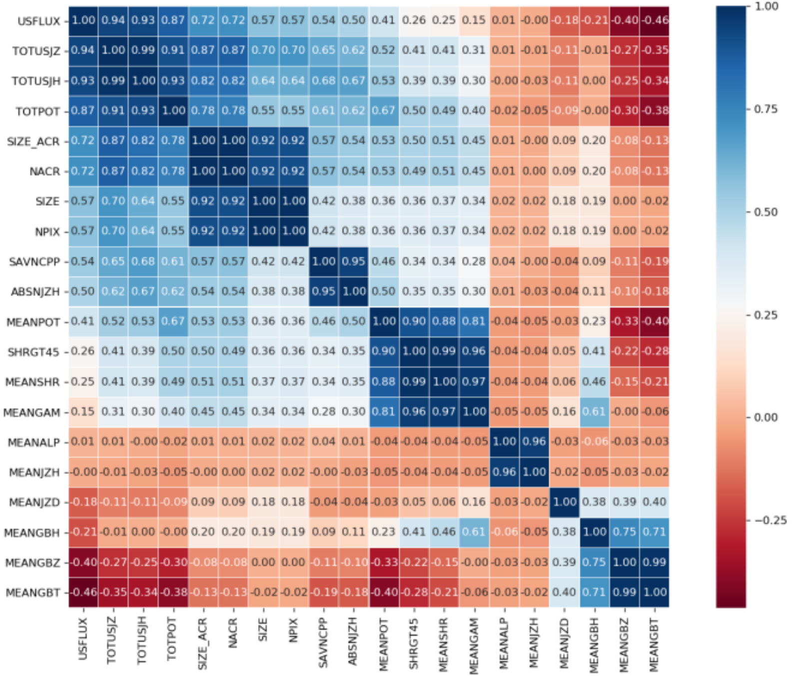


The Result Metrics



The Result Metrics

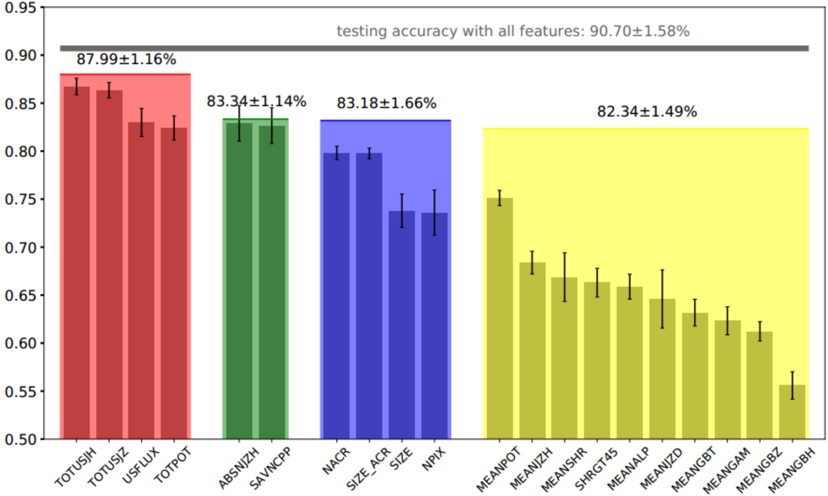



USFLUX 1.00 0.94 0.93 0.87 0.72 0.72 0.57 0.57 0.54 0.50 0.41 0.26 0.25 0.15 0.01 -0.00 -0.18 -0.21 -0.40 -0.46
TOTUSJZ 0.94 1.00 0.99 0.91 0.87 0.87 0.70 0.70 0.65 0.62 0.52 0.41 0.41 0.31 0.01 -0.01 -0.11 -0.01 -0.27 -0.35
TOTUSJH 0.93 0.99 1.00 0.93 0.82 0.82 0.64 0.64 0.68 0.67 0.53 0.39 0.39 0.30 -0.00 -0.03 -0.11 0.00 -0.25 -0.34
TOTPOT 0.87 0.91 0.93 1.00 0.78 0.78 0.55 0.55 0.61 0.62 0.67 0.50 0.49 0.40 -0.02 -0.05 -0.09 -0.00 -0.30 -0.38
SIZE_ACR 0.72 0.87 0.82 0.78 1.00 1.00 0.92 0.92 0.57 0.54 0.53 0.50 0.51 0.45 0.01 -0.00 0.09 0.20 -0.08 -0.13
NACR 0.72 0.87 0.82 0.78 1.00 1.00 0.92 0.92 0.57 0.54 0.53 0.49 0.51 0.45 0.01 0.00 0.09 0.20 -0.08 -0.13
SIZE 0.57 0.70 0.64 0.55 0.92 0.92 1.00 1.00 0.42 0.38 0.36 0.36 0.37 0.34 0.02 0.02 0.18 0.19 0.00 -0.02
NPIX 0.57 0.70 0.64 0.55 0.92 0.92 1.00 1.00 0.42 0.38 0.36 0.36 0.37 0.34 0.02 0.02 0.18 0.19 0.00 -0.02
SAVNCPP 0.54 0.65 0.68 0.61 0.57 0.57 0.42 0.42 1.00 0.95 0.46 0.34 0.34 0.28 0.04 -0.00 -0.04 0.09 -0.11 -0.19
ABSNJZH 0.50 0.62 0.67 0.62 0.54 0.54 0.38 0.38 0.95 1.00 0.50 0.35 0.35 0.30 0.01 -0.03 -0.04 0.11 -0.10 -0.18
MEANPOT 0.41 0.52 0.53 0.67 0.53 0.53 0.36 0.36 0.46 0.50 1.00 0.90 0.88 0.81 -0.04 -0.05 -0.03 0.23 -0.33 -0.40
SHRGT45 0.26 0.41 0.39 0.50 0.50 0.49 0.36 0.36 0.34 0.35 0.90 1.00 0.99 0.96 -0.04 -0.04 0.05 0.41 -0.22 -0.28
MEANSHR 0.25 0.41 0.39 0.49 0.51 0.51 0.37 0.37 0.34 0.35 0.88 0.99 1.00 0.97 -0.04 -0.04 0.06 0.46 -0.15 -0.21
MEANGAM 0.15 0.31 0.30 0.40 0.45 0.45 0.34 0.34 0.28 0.30 0.81 0.96 0.97 1.00 -0.05 -0.05 0.16 0.61 -0.00 -0.06
MEANALP 0.01 0.01 -0.00 -0.02 0.01 0.01 0.02 0.02 0.04 0.01 -0.04 -0.04 -0.04 -0.05 1.00 0.96 -0.03 -0.06 -0.03 -0.03
MEANJZH -0.00 -0.01 -0.03 -0.05 -0.00 0.00 0.02 0.02 -0.00 -0.03 -0.05 -0.04 -0.04 -0.05 0.96 1.00 -0.02 -0.05 -0.03 -0.02
MEANJZD -0.18 -0.11 -0.11 -0.09 0.09 0.09 0.18 0.18 -0.04 -0.04 -0.03 0.05 0.06 0.16 -0.03 -0.02 1.00 0.38 0.39 0.40
MEANGBH -0.21 -0.01 0.00 -0.00 0.20 0.20 0.19 0.19 0.09 0.11 0.23 0.41 0.46 0.61 -0.06 -0.05 0.38 1.00 0.75 0.71
MEANGBZ -0.40 -0.27 -0.25 -0.30 -0.08 -0.08 0.00 0.00 -0.11 -0.10 -0.33 -0.22 -0.15 -0.00 -0.03 -0.03 0.39 0.75 1.00 0.99
MEANGBT -0.46 -0.35 -0.34 -0.38 -0.13 -0.13 -0.02 -0.02 -0.19 -0.18 -0.40 -0.28 -0.21 -0.06 -0.03 -0.02 0.40 0.71 0.99 1.00
USFLUX TOTUSJZ TOTUSJH TOTPOT SIZE_ACR NACR SIZE NPIX SAVNCPP ABSNJZH MEANPOT SHRGT45 MEANSHR MEANGAM MEANALP MEANJZH MEANJZD MEANGBH MEANGBZ MEANGBT
1.00
0.75
0.50
0.25
0.00
−0.25

LSTM Model Variable Importance

To understand the driving force for the prediction score, we used each of the 20 SHARP features as a single predictor for M/X class flare, and see how the variable importance would drop.

LSTM Model Variable Importance

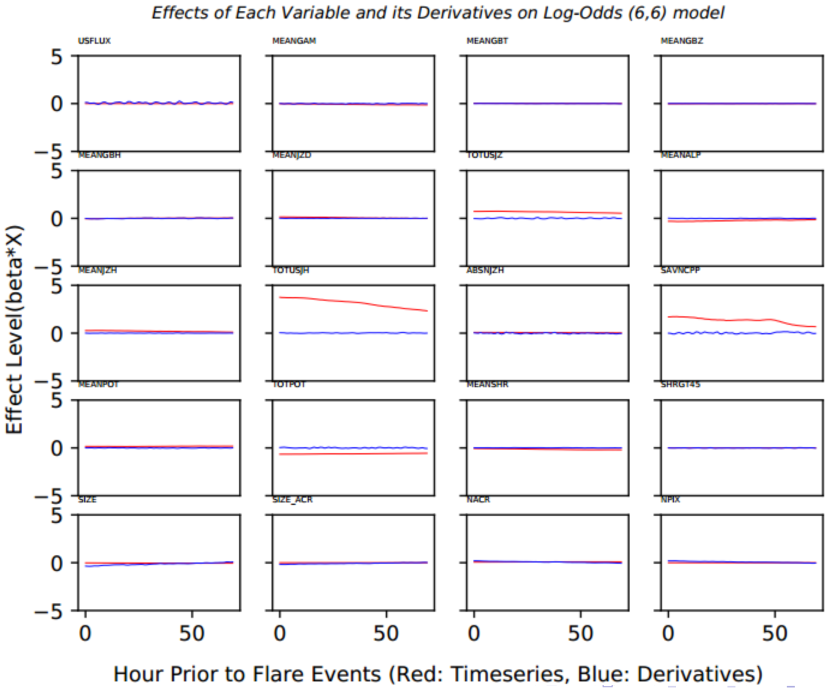


Variable Importance Selection with LASSO regression

We also tried a simple logistic regression for the classification task, **with LASSO penalty**.

$$\min_{\beta_0,\beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 \right\} \text{ subject to } \sum_{j=1}^{p} |\beta_j| \le t.$$

LASSO penalty is a kind of penalized regression tool. It will tend to push some of the variables to have zero effects on the classification and only retain those variables that are important. It can also select the most important variable from a block of highly correlated variables.



Effects of Each Variable and its Derivatives on Log-Odds (6,6) model
USFLUX
MEANGAM
MEANGBT
MEANGBZ
MEANGBH
MEANJZD
TOTUSJZ
MEANALP
MEANJZH
TOTUSJH
ABSNJZH
SAVNCPP
MEANPOT
TOTPOT
MEANSHR
SHRGT45
SIZE
SIZE_ACR
NACR
NPIX
Effect Level(beta*X)
Hour Prior to Flare Events (Red: Timeseries, Blue: Derivatives)

Sharp Transitions of Prediction Score

We define that the prediction curve of an active region before its first M/X flare goes through a **Sharp Transition** if:

- there is a time when prediction score is above 0.7, and persists at least 36 minutes (after transition time)
- starting from the "after transition time", any time when prediction score is below 0.3 and persists at least 36 minutes is called "before transition time"

Sharp Transitions of Prediction Score

We define that the prediction curve of an active region before its first M/X flare goes through a **Sharp Transition** if:

- there is a time when prediction score is above 0.7, and persists at least 36 minutes (after transition time)
- starting from the "after transition time", any time when prediction score is below 0.3 and persists at least 36 minutes is called "before transition time"

Sharp Transitions of Prediction Score

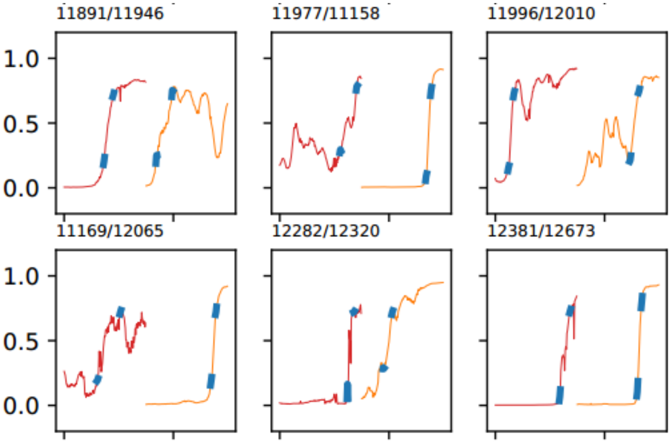


Figure: A few selected sharp transitions

Sharp Transitions of Prediction Score

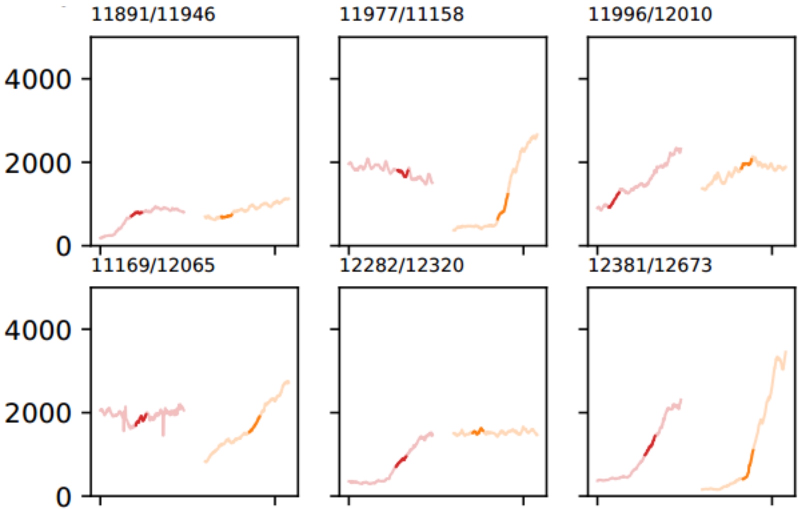
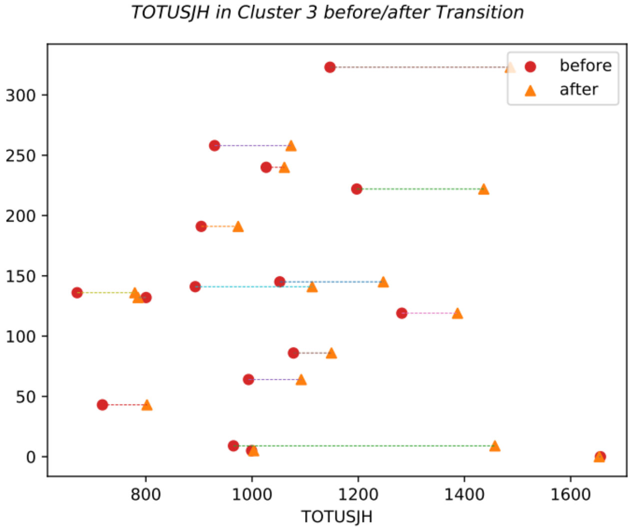


Figure: TOTUSJH time series during transition time

Sharp Transitions of Prediction Score



Conclusion

- LSTM model gives high prediction accuracy for many different classification tasks
- TOTUSJH and SAVNCPP are selected to be the most significant contributors to the LSTM prediction
- 23 active regions with sharp transitions detected
- Driving force for sharp transition is not uniform, but more evident within sub-group of active regions

Conclusion

- LSTM model gives high prediction accuracy for many different classification tasks
- TOTUSJH and SAVNCPP are selected to be the most significant contributors to the LSTM prediction
- 23 active regions with sharp transitions detected
- Driving force for sharp transition is not uniform, but more evident within sub-group of active regions

Conclusion

- LSTM model gives high prediction accuracy for many different classification tasks
- TOTUSJH and SAVNCPP are selected to be the most significant contributors to the LSTM prediction
- 23 active regions with sharp transitions detected
- Driving force for sharp transition is not uniform, but more evident within sub-group of active regions

Conclusion

- LSTM model gives high prediction accuracy for many different classification tasks
- TOTUSJH and SAVNCPP are selected to be the most significant contributors to the LSTM prediction
- 23 active regions with sharp transitions detected
- Driving force for sharp transition is not uniform, but more evident within sub-group of active regions

Step 1: Select PIL candidate pixels

Data Type: Magnetogram with 200,000+ pixels.

Step 1: Select PIL candidate pixels

Data Type: Magnetogram with 200,000+ pixels.

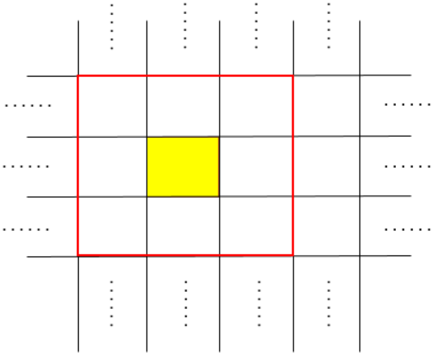


Figure: A pixel's neighborhood, 3×3 window

Step 1: Select PIL candidate pixels

Data Type: Magnetogram with 200,000+ pixels.

In every pixel's neighborhood, we check if there is both a strong positive polar and a strong negative polar. (default threshold ± 100Gauss)

Step 1: Select PIL candidate pixels

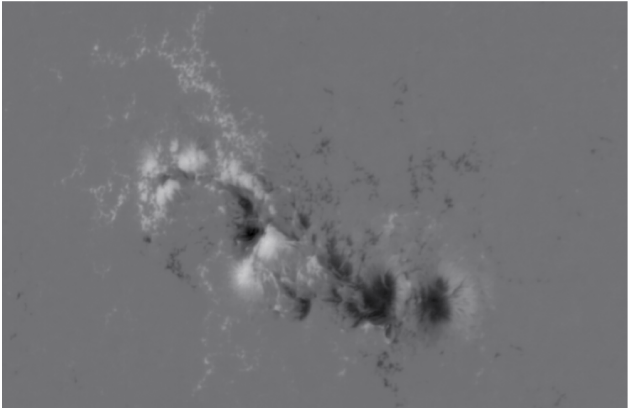


Figure: B-z Component, HARP377, 2011-02-17 14:00:00TAI

Step 1: Select PIL candidate pixels

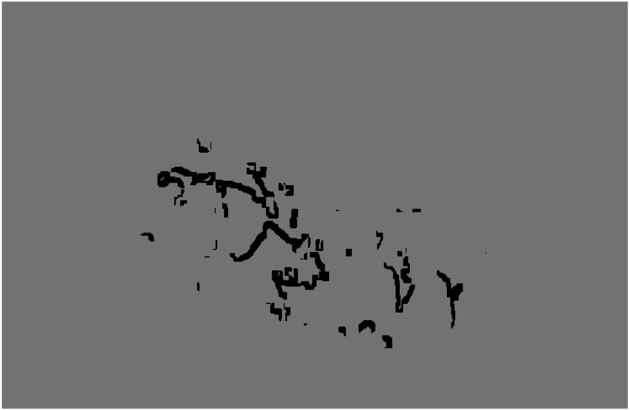


Figure: Pixels remained after neighborhood check (5×5 window)

Step 1: Select PIL candidate pixels

Property of the pixels selected:

- Having shape of multiple PIL segments
- Having miscellaneous clusters of pixels not of our interest
- Pixels on the same PIL are not linked well
- The PILs seem to be too thick.

Step 1: Select PIL candidate pixels

Property of the pixels selected:

- Having shape of multiple PIL segments
- Having miscellaneous clusters of pixels not of our interest
- Pixels on the same PIL are not linked well
- The PILs seem to be too thick.

Step 1: Select PIL candidate pixels

Property of the pixels selected:

- Having shape of multiple PIL segments
- Having miscellaneous clusters of pixels not of our interest
- Pixels on the same PIL are not linked well
- The PILs seem to be too thick.

Step 1: Select PIL candidate pixels

Property of the pixels selected:

- Having shape of multiple PIL segments
- Having miscellaneous clusters of pixels not of our interest
- Pixels on the same PIL are not linked well
- The PILs seem to be too thick.

Step 2: Non-maxima Suppression

Use Prewitt filter to calculate local B-z gradient for each of the pixel:

$$\mathbf{G_x} = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} * \mathbf{A} \quad \text{and} \quad \mathbf{G_y} = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * \mathbf{A}$$

Step 2: Non-maxima Suppression

With both G_x and G_y, we could calculate the direction of the gradient by $\arctan(\frac{G_y}{G_x})$.

And with the direction of the gradient, we could define the adjacent pixels along the gradient direction of any pixel with non-zero gradient.

Step 2: Non-maxima Suppression

With both G_x and G_y, we could calculate the direction of the gradient by $\arctan(\frac{G_y}{G_x})$.

And with the direction of the gradient, we could define the adjacent pixels along the gradient direction of any pixel with non-zero gradient.

Step 2: Non-maxima Suppression

With both G_x and G_y, we could calculate the direction of the gradient by $\arctan(\frac{G_y}{G_x})$.

And with the direction of the gradient, we could define the adjacent pixels along the gradient direction of any pixel with non-zero gradient.

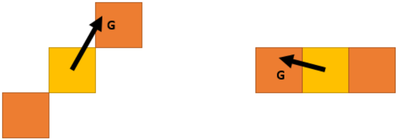


Figure: Adjacent pixels along gradient direction

Step 2: Non-maxima Suppression

With both G_x and G_y, we could calculate the direction of the gradient by $\arctan(\frac{G_y}{G_x})$.

And with the direction of the gradient, we could define the adjacent pixels along the gradient direction of any pixel with non-zero gradient.

Now we only retain the pixels whose gradient norm is local maxima along it gradient direction.

Step 2: Non-maxima Suppression

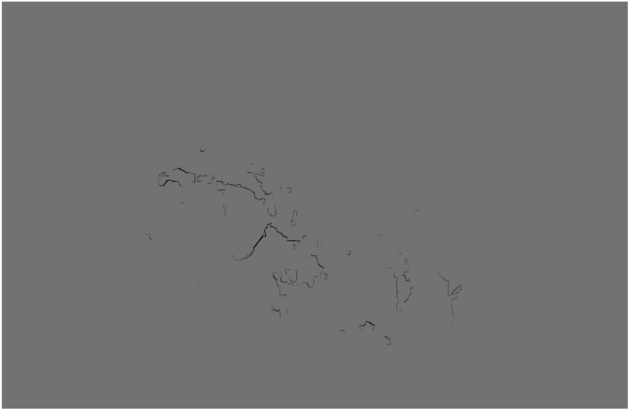


Figure: Thin Edges after Non-maxima Suppression

Step 3: PIL extension

In order to make the PIL drawn to be more connected, we apply a simple continuity extension to pixels on the PIL.

Step 3: PIL extension

In order to make the PIL drawn to be more connected, we apply a simple continuity extension to pixels on the PIL.

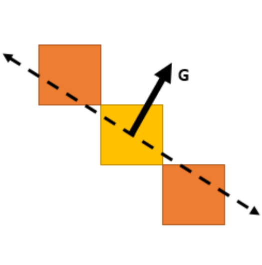


Figure: PIL extension direction

Step 3: PIL extension

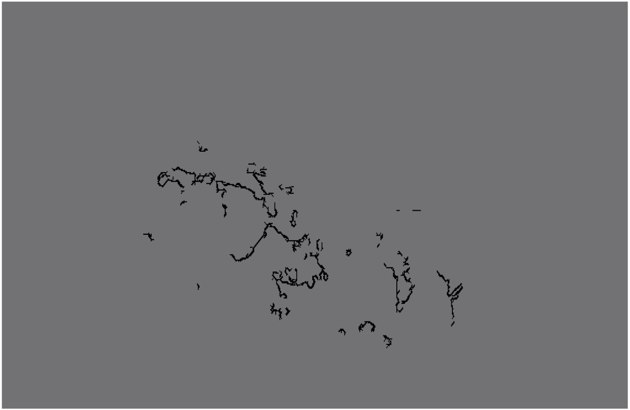


Figure: More connected PILs

Step 3: PIL extension

Such an extension would add more connectivity to each PIL, but will add many pixels with small B-z gradients into the PIL.

Optionally, one could only extend on points with large gradients, and this extension can be conducted **recursively**.

Step 4: Density-Based Clustering

- Generally, the longer the PIL is, the more complex the local magenetic field structure there is.
- We need to retain the longest PILs and erase the miscellaneous PILs.

Step 4: Density-Based Clustering

- Generally, the longer the PIL is, the more complex the local magenetic field structure there is.
- We need to retain the longest PILs and erase the miscellaneous PILs.

Step 4: Density-Based Clustering

How to classify PIL pixels into several clusters?

- Connected-component analysis
- Density-based clustering algorithm

Step 4: Density-Based Clustering

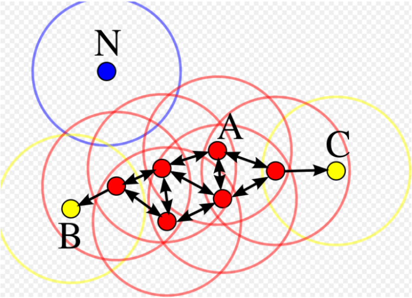
Density-based spatial clustering of applications with noise **(DBSCAN)** :

- Locate some "core" points as the seed for a cluster
- All points within a certain range of core points are in the same cluster as the core points

Step 4: Density-Based Clustering

Density-based spatial clustering of applications with noise **(DBSCAN)** :

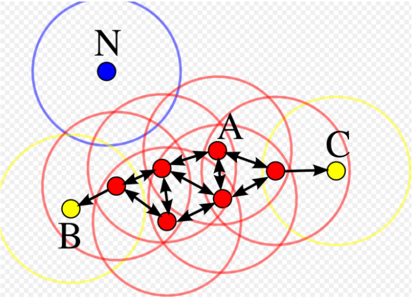
- Locate some "core" points as the seed for a cluster
- All points within a certain range of core points are in the same cluster as the core points



Step 4: Density-Based Clustering

Density-based spatial clustering of applications with noise **(DBSCAN)** :

- Locate some "core" points as the seed for a cluster
- All points within a certain range of core points are in the same cluster as the core points



Step 4: Density-Based Clustering

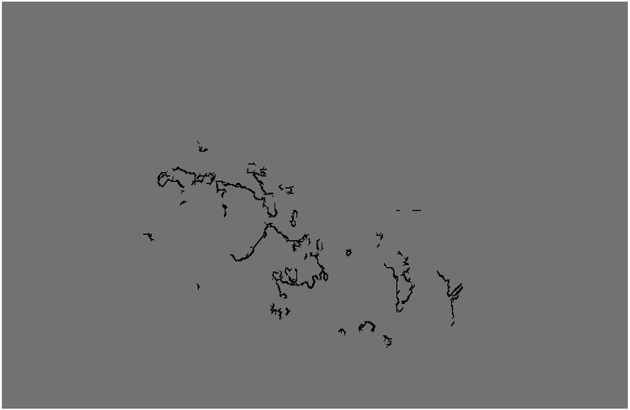


Figure: More connected PILs

Step 4: Density-Based Clustering

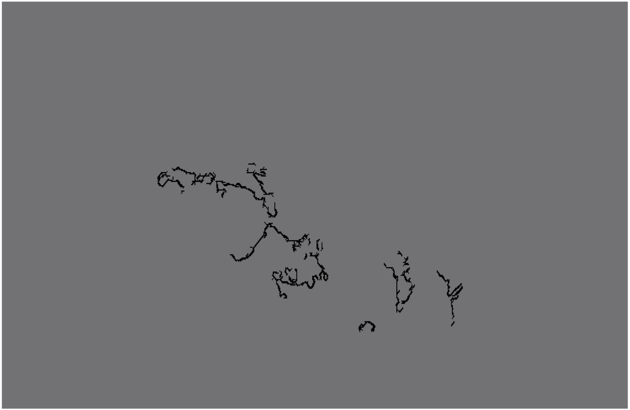
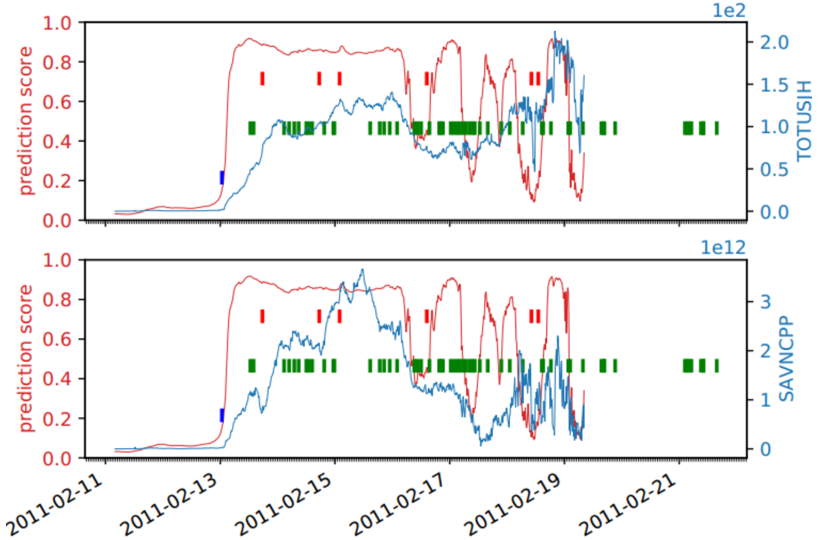


Figure: PILs left after deleting small clusters

AR:11158



1e2
1.0
0.8
0.6
0.4
0.2
0.0
prediction score
2.0
1.5
1.0
0.5
0.0
TOTUSJH
1e12
3
2
1
0
SAVNCPP
2011-02-11
2011-02-13
2011-02-15
2011-02-17
2011-02-19
2011-02-21

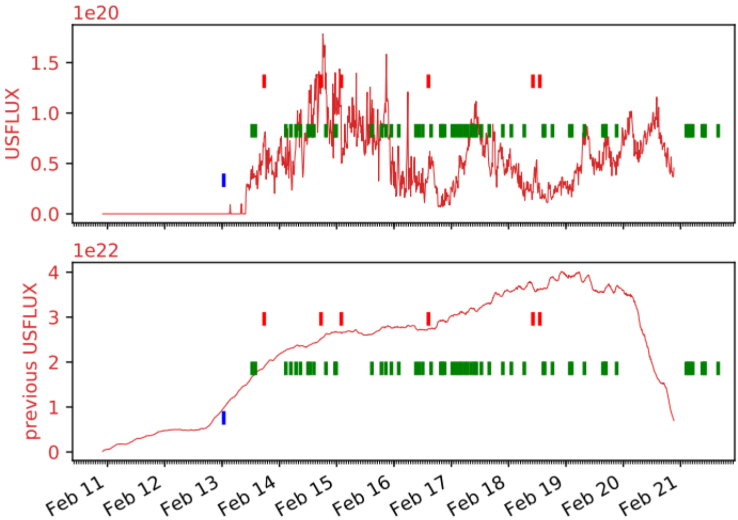


Figure: Recalculate USFLUX for HARP377