Interpretable ML Methods for forecasting the SYM-H Index

Daniel Iong

Department of Statistics
University of Michigan, Ann Arbor

Joint work with Yang Chen, Enrico Camporeale, Tamas Gombosi, Tuija Pulkkinen, Gabor Toth, Shasha Zou

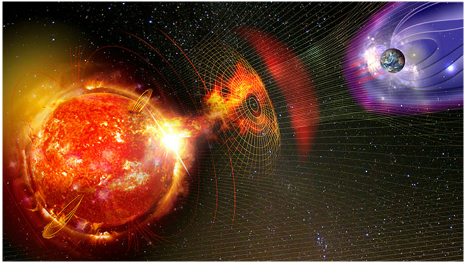
September 2, 2021

Overview

1. Introduction
   - Geomagnetic Storms
   - DST/SYM-H Index
   - Interpretable/Explainable ML
2. Methods
   - Gradient Boosting Machines
   - SHAP Values
3. Results
   - Data Preprocessing & Model Training
   - Comparison to existing methods
   - Interpreting results

Geomagnetic Storms

- **Geomagnetic storms** are brief disturbances in Earth's magnetosphere caused by energy transferred from the solar wind.
- They have the potential to cause **severe disruptions of critical infrastructure** such as satellites, power grids, oil pipelines, etc.
- The largest storms result from solar **coronal mass ejections** (CMEs).
    - Large expulsions of plasma and magnetic field from the Sun's corona

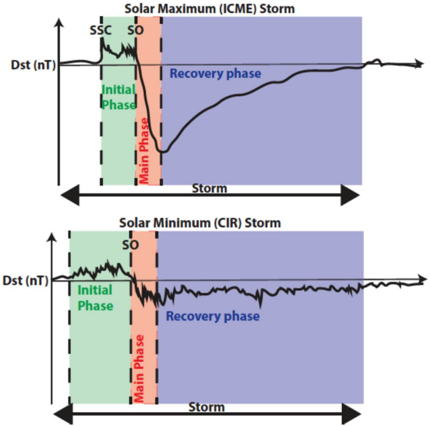


DST/SYM-H Index

- The **SYM-H index** is a 1-minute resolution version of the hourly DST index.
- The **DST index** measures the symmetric geomagnetic disturbance of the horizontal component of Earth's magnetic field near the magnetic equator on the Earth's surface
    - Mostly hovers around 0 nT during geomagnetically quiet times

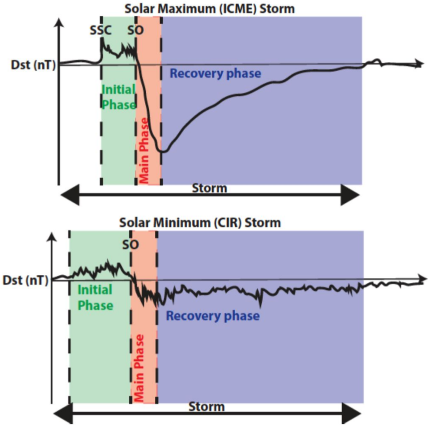
DST/SYM-H Index

- The **SYM-H index** is a 1-minute resolution version of the hourly DST index.
- The **DST index** measures the symmetric geomagnetic disturbance of the horizontal component of Earth's magnetic field near the magnetic equator on the Earth's surface
  - Mostly hovers around 0 nT during geomagnetically quiet times
- Geomagnetic storms are defined and categorized according to the DST index.
  - **Moderate storm**: -100 nT < DST < -50 nT
  - **Intense storm**: -250 nT < DST < -100 nT
  - **Super storm**: DST < -250 nT



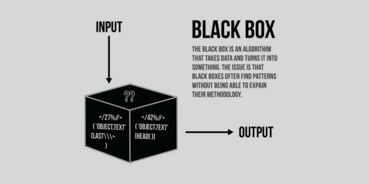
DST/SYM-H Index

- The **SYM-H index** is a 1-minute resolution version of the hourly DST index.
- The **DST index** measures the symmetric geomagnetic disturbance of the horizontal component of Earth's magnetic field near the magnetic equator on the Earth's surface
  - Mostly hovers around 0 nT during geomagnetically quiet times
- Geomagnetic storms are defined and categorized according to the DST index.
  - **Moderate storm**: -100 nT $<$ DST $<$ -50 nT
  - **Intense storm**: -250 nT $<$ DST $<$ -100 nT
  - **Super storm**: DST $<$ -250 nT
- **Our objective**: Obtain interpretable predictions of the SYM-H index 1-2 hours ahead.



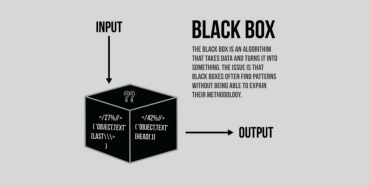
Interpretable/Explainable ML

- Most ML methods, such as neural networks, are highly accurate but are too complex for us to understand its decision making process (black box).
- Interpretability/explainability is crucial if we are to trust these methods, especially in high-risk applications like healthcare and finance.



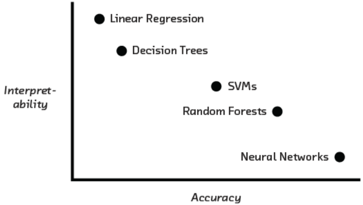
Interpretable/Explainable ML

- Most ML methods, such as neural networks, are highly accurate but are too complex for us to understand its decision making process (black box).
- Interpretability/explainability is crucial if we are to trust these methods, especially in high-risk applications like healthcare and finance.
- In the space sciences, interpretability can:
    1. help us verify if the model's predictions are consistent with physical knowledge
    2. potentially provide novel insight into underlying physics



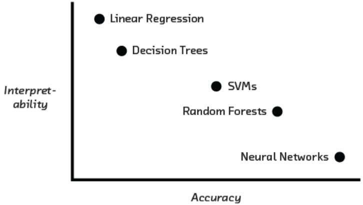
Interpretable/Explainable ML

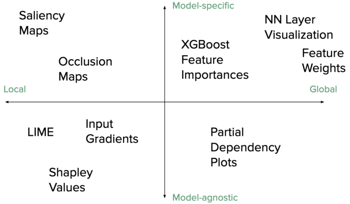
- In many cases, there exists a tradeoff between accuracy and intrinsic interpretability.



Interpretable/Explainable ML

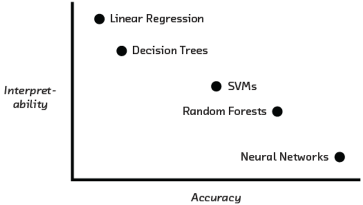
- In many cases, there exists a tradeoff between accuracy and intrinsic interpretability.
- Interpretable ML methods can be classified as:
  - **Intrinsic**: The model itself yields outputs that are interpretable
  - **Post-hoc**: Interpretable results are obtained after training.
  - **Local** vs. **global**: Explains individual predictions vs. overall model behavior
  - **Model-specific** vs. **model-agnostic**

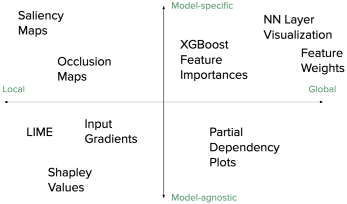




Interpretable/Explainable ML

- In many cases, there exists a tradeoff between accuracy and intrinsic interpretability.
- Interpretable ML methods can be classified as:
    - **Intrinsic**: The model itself yields outputs that are interpretable
    - **Post-hoc**: Interpretable results are obtained after training.
    - **Local** vs. **global**: Explains individual predictions vs. overall model behavior
    - **Model-specific** vs. **model-agnostic**
- In this work, we use a local model-agnostic interpretability method called **SHAP** (*Shapley Additive Explanations*) to interpret predictions from **gradient boosting machines** (*GBMs*).





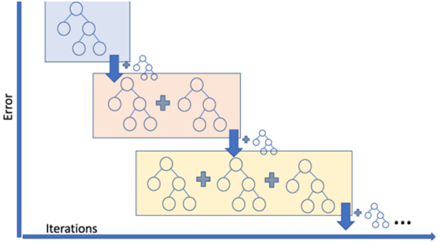
Gradient Boosting Machines (GBMs)

- GBMs, also known as *gradient boosted trees*, have been used sucessfully for prediction in a wide range of domains.
- Some have argued that they are more accurate than neural networks for prediction with tabular data.

Gradient Boosting Machines (GBMs)

- GBMs, also known as *gradient boosted trees*, have been used sucessfully for prediction in a wide range of domains.
- Some have argued that they are more accurate than neural networks for prediction with tabular data.

Intuition: Instead of a single complex model, iteratively construct simple models that improve on previously constructed models and aggregate them to make predictions.



Gradient Boosting Machines (GBMs)

$$\textbf{Model:}\ y_i = f(x_i) = \alpha + \sum_{m=1}^{M} T_m(x_i) + \epsilon_i,$$

where α is an intercept; T_m are regression trees; M is the # of iterations (trees).

- We train GBMs with XGBoost which penalizes overly-complex trees.

Gradient Boosting Machines (GBMs)

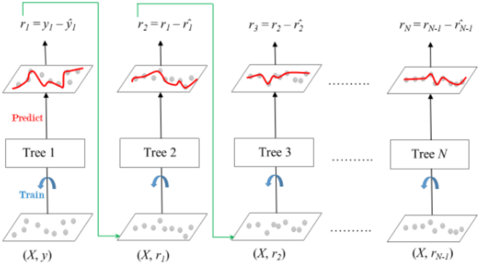
$$\textbf{Model:}\ y_i = f(x_i) = \alpha + \sum_{m=1}^{M} T_m(x_i) + \epsilon_i,$$

where α is an intercept; T_m are regression trees; M is the # of iterations (trees).

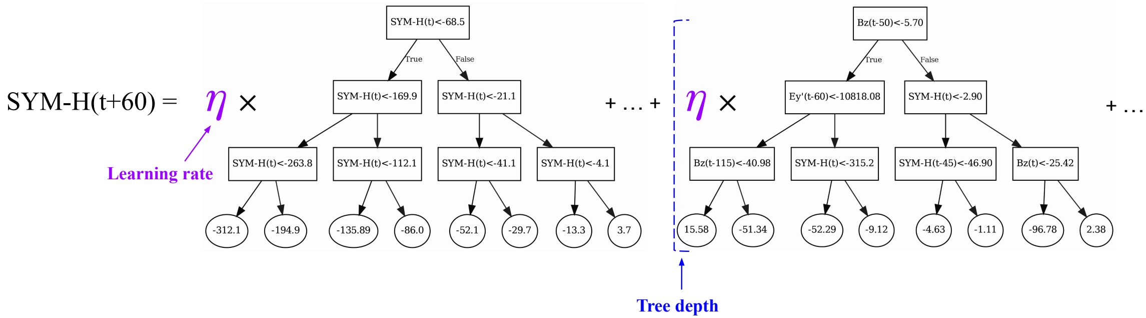
- We train GBMs with XGBoost which penalizes overly-complex trees.
- At iteration m, we construct T_m by minimizing

$$\mathcal{L}^{(m)}(f) = \sum_i \left[y_i - (\hat{y}_i^{(m-1)} + T_m(x_i))\right]^2 + \sum_{j=1}^{m} \Omega(T_j),$$

  - ℓ is the loss function
  - Ω penalizes the complexity of the regression trees
  - $\hat{y}_i^{(m-1)} = \alpha + \sum_{j=1}^{m-1} T_j(x_i)$.



Gradient Boosting Machines (GBMs)



Structure of GBM trained on past SYM-H, solar wind & IMF parameters to predict SYM-H 1 hour ahead.

- There are several hyperparameters for controlling tree complexity.
  - learning rate, max tree depth, min. child weight, feature subsampling percentage, number of trees, etc

SHAP Values

- SHAP computes **feature contributions for individual predictions** using ideas from coalition game theory.

SHAP Values

- SHAP computes **feature contributions for individual predictions** using ideas from coalition game theory.
- SHAP belongs to the class of *additive feature attribution methods* which assumes the following explanation model for an individual prediction:

$$\text{Prediction} = g(\mathbf{z}') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i'$$

  - ϕ_0 is a baseline value (e.g. mean); M is the number of input features
  - z_i' is a binary variable indicating whether feature i is present
  - ϕ_i is the feature contribution from feature i

SHAP Values

- SHAP computes **feature contributions for individual predictions** using ideas from coalition game theory.
- SHAP belongs to the class of *additive feature attribution methods* which assumes the following explanation model for an individual prediction:

$$\text{Prediction} = g(\mathbf{z}') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i'$$

  - ϕ_0 is a baseline value (e.g. mean); M is the number of input features
  - z_i' is a binary variable indicating whether feature i is present
  - ϕ_i is the feature contribution from feature i
- SHAP yields the unique solution to ϕ_i that satisfies three desirable theoretical properties (*local accuracy, missingness, consistency*)

SHAP Values

- SHAP computes **feature contributions for individual predictions** using ideas from coalition game theory.
- SHAP belongs to the class of *additive feature attribution methods* which assumes the following explanation model for an individual prediction:

$$\text{Prediction} = g(\mathbf{z}') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i$$

  - ϕ_0 is a baseline value (e.g. mean); M is the number of input features
  - z'_i is a binary variable indicating whether feature i is present
  - ϕ_i is the feature contribution from feature i
- SHAP yields the unique solution to ϕ_i that satisfies three desirable theoretical properties (*local accuracy, missingness, consistency*)
- Although SHAP is model-agnostic, it is more computationally feasible for tree-based models like GBM because of an algorithm that exploits their structure (TreeSHAP).

Data Preprocessing

- We use different combinations of the following features as inputs:
  1. Past SYM-H values
  2. IMF parameters (B_x,B_y,B_z)
  3. Solar wind parameters (density, temperature, V_x)
  4. Derived parameters: dynamic pressure, (rectified) electric field

Data Preprocessing

- We use different combinations of the following features as inputs:
  1. Past SYM-H values
  2. IMF parameters (B_x,B_y,B_z)
  3. Solar wind parameters (density, temperature, V_x)
  4. Derived parameters: dynamic pressure, (rectified) electric field
- SYM-H index data was obtained from OMNI.
- Solar wind/IMF parameters were obtained from ACE level 2 data from CDAWeb.

Data Preprocessing

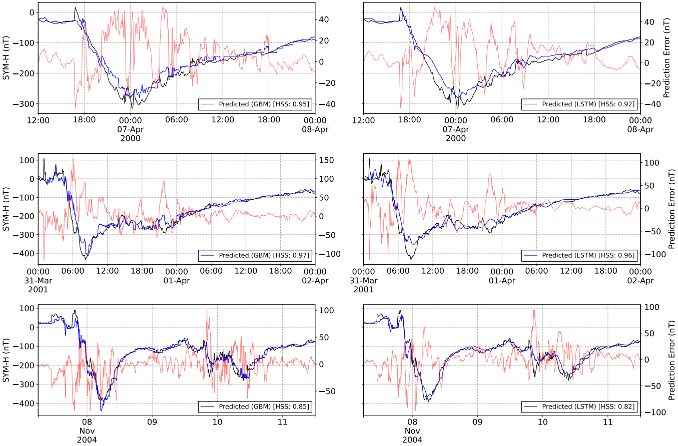
- We use different combinations of the following features as inputs:
  1. Past SYM-H values
  2. IMF parameters (B_x,B_y,B_z)
  3. Solar wind parameters (density, temperature, V_x)
  4. Derived parameters: dynamic pressure, (rectified) electric field
- SYM-H index data was obtained from OMNI.
- Solar wind/IMF parameters were obtained from ACE level 2 data from CDAWeb.
- 42 strong geomagnetic storms between 1998-2018 for training/testing.
  - SYM-H $<$ -100 nT
  - Same storms used in Siciliano et al. (2021) and Collado-Villaverde et al. (2021)

Data Preprocessing

- We use different combinations of the following features as inputs:
  1. Past SYM-H values
  2. IMF parameters (B_x,B_y,B_z)
  3. Solar wind parameters (density, temperature, V_x)
  4. Derived parameters: dynamic pressure, (rectified) electric field
- SYM-H index data was obtained from OMNI.
- Solar wind/IMF parameters were obtained from ACE level 2 data from CDAWeb.
- 42 strong geomagnetic storms between 1998-2018 for training/testing.
  - SYM-H $<$ -100 nT
  - Same storms used in Siciliano et al. (2021) and Collado-Villaverde et al. (2021)
- 2 hour history of solar wind/IMF/derived parameters & 1 hour history of past SYM-H *when past SYM-H is included*
- 30 hour history of solar wind/IMF/derived parameters *when past SYM-H is excluded*

Comparison to existing methods

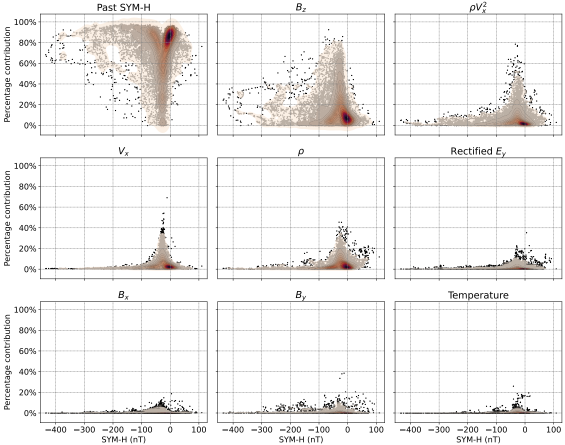
- GBMs outperform existing methods based on LSTMs for forecasting the SYM-H index 1 hour ahead *using only past SYM-H and IMF parameters*.



Storm #	GBM	LSTM1	LSTM2	Persistence
26	5.863	6.630	6.700	7.631
27	7.729	8.913	8.900	9.623
28	4.281	5.858	5.400	5.814
29	5.833	6.683	7.200	7.174
30	4.927	5.200	5.600	4.810
31	8.277	8.584	10.700	10.429
32	6.841	7.259	8.300	10.528
33	14.492	13.340	16.300	21.167
34	10.190	10.034	11.300	10.913
35	7.154	7.693	8.500	8.011
36	8.512	9.525	8.700	9.708
37	14.548	15.184	17.500	19.698
38	3.886	4.080	4.200	4.842
39	5.901	6.431	5.600	7.597
40	4.976	4.673	5.500	5.057
41	7.558	7.882	9.000	9.984
42	5.030	5.669	5.900	6.036
Mean	7.412	7.860	8.550	9.354

Interpreting results: Percentage contributions

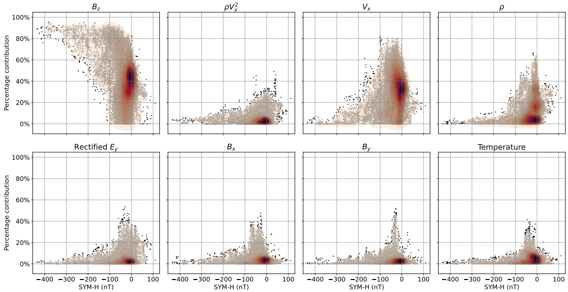
- Past SYM-H and B_z have the two largest contributions.
- Surprisingly, V_x and the rectified electric field are much less important.
- We would expect ρV_x^2 to be most important during the sudden storm commencement but the contribution is similiar to that of B_z when predicting positive SYM-H.
- ρV_x^2, V_x, ρ have marginal contributions only for predicting low negative SYM-H.



% contribution to 1-hour prediction as a function of SYM-H from GBM trained ***with*** *past SYM-H*

Interpreting results: Percentage contributions

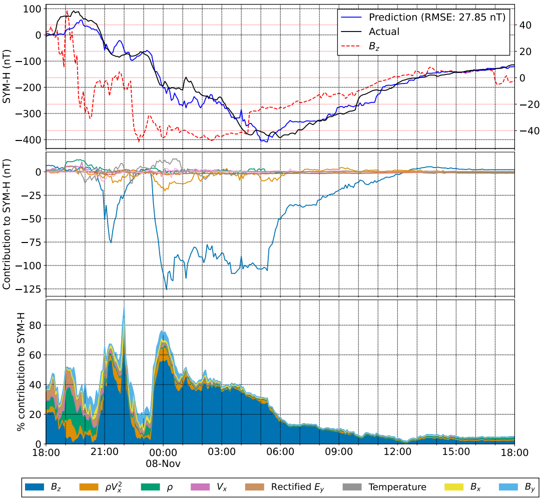
- B_z becomes the largest contributor.
- Surprisingly, the contribution from rectified E_y is still small.



% contribution to 1-hour prediction as a function of SYM-H from GBM trained ***without*** *past SYM-H*

Interpreting results: Nov. 2004 Storm

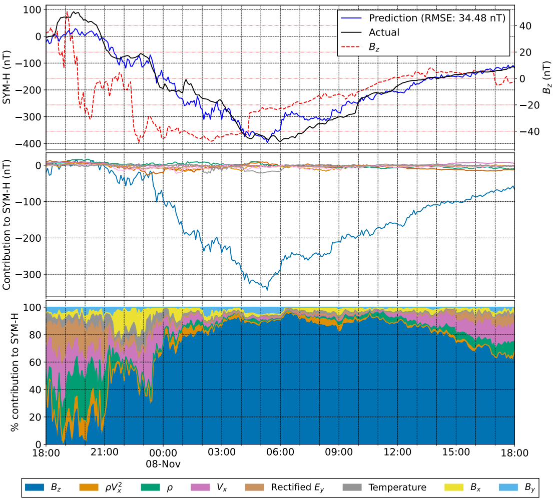
- **18:00 - 21:00 UT**: Observed SYM-H is positive
  - Density contributes $\sim 20\%$
  - Dynamic pressure should have large contribution but doesn't
- **20:45 UT Nov. 7 - 06:00 UT Nov. 8** (Main phase)
  - Relative contrib. from B_z have peaks around 21:15-22:00 UT and 24:00 UT, corresponding to local minima of B_z
- We would expect B_z to be a large contributor when it crosses 0 but this doesn't happen
- Past SYM-H dominates during recovery phase as expected.



Model trained **with** *past SYM-H*

Interpreting results: Nov. 2004 Storm

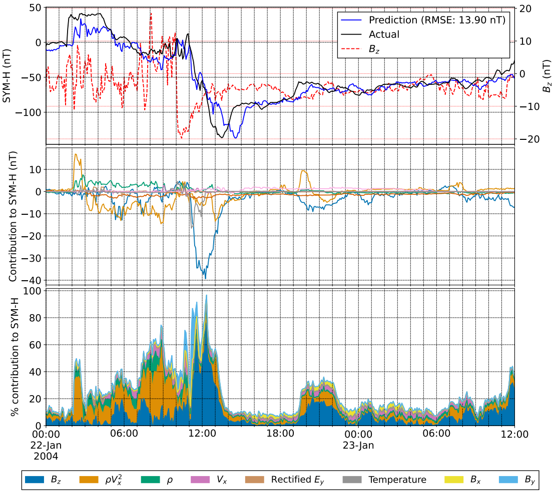
- **18:00 - 21:00 UT**: Observed SYM-H is positive
  - Main contributors are density, V_x, and rectified E_y.
- Relative contrib. from B_z gradually increases as SYM-H goes negative.
- Other features contribute more starting from 12:00 UT when B_z turns positive.



Model trained ***without*** *past SYM-H*

Interpreting results: Jan. 2004 Storm

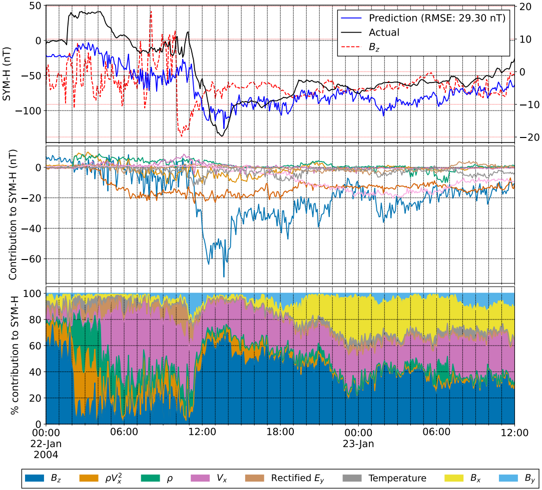
- Complicated storm due to the highly variable B_z in the CME sheath (00:00 UT - 11:00 UT Jan. 22)
- **02:00 - 11:00 UT**: ρV_x^2 is the main contributor after past SYM-H.
- **11:00 - 14:00 UT**: Relative contrib. from B_z spikes.
- **19:00 - 23:00 UT**: Relative contrib. from B_z, ρV_x^2 increases, correctly predicting the slow down of the recovery.



Model trained **with** *past SYM-H*

Interpreting results: Jan. 2004 Storm

- V_x becomes a large contributor.
- **02:00 UT**: ρV_x^2 has a large contribution when observed SYM-H turns positive.
- B_x starts to have a large contribution around 18:00 UT.



Model trained ***without*** *past SYM-H*