Develop the Oracle of DSCOVR

ReDSCOVR

Re-discover the oracle of DSCOVR by recurrent neural network

Chin-Yang Lin

Wei-Hsiang Chiu

Yu-Cheng Lin

Rong-Zhu Lin

Chen-Hsun Ma

Background picture: https://svs.gsfc.nasa.gov/11298

FFT analysis

DSCOVR and GFZ time series data

↓

Fast Fourier transform

↓

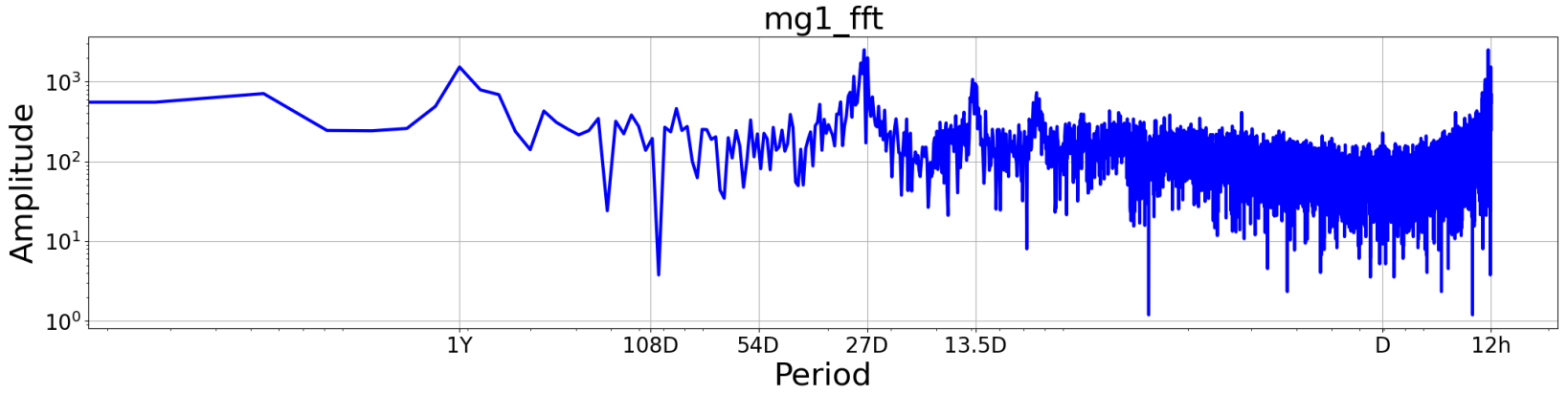
Frequency (period) spectrum

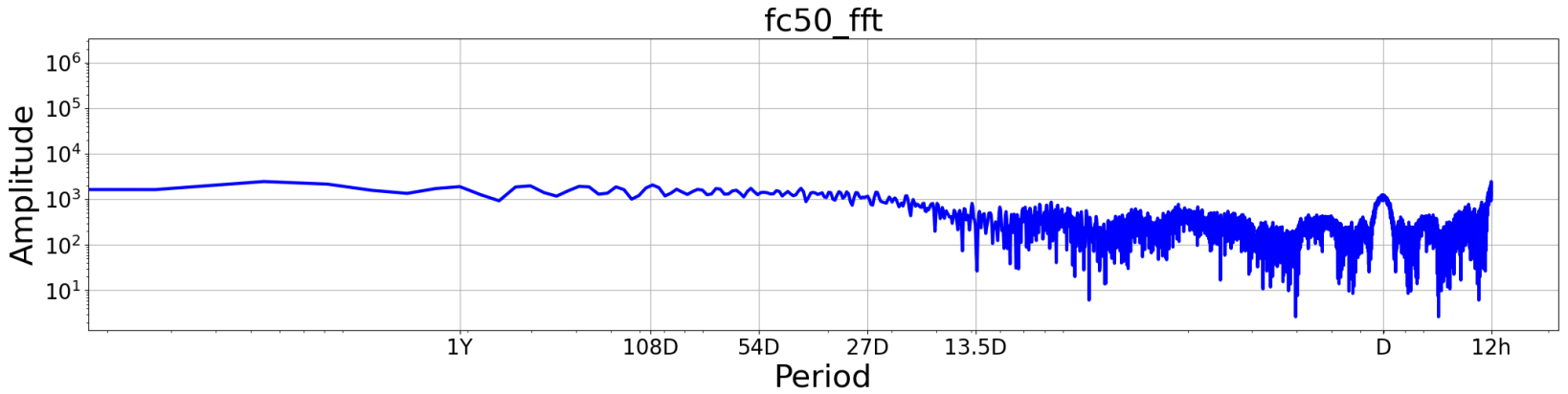
↓

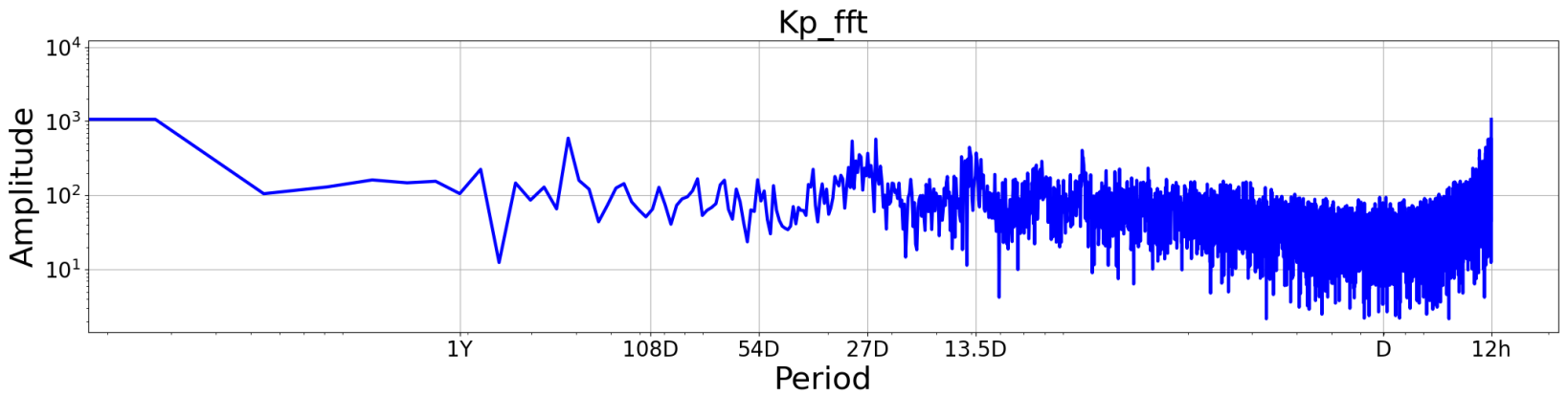
Observe the periods with the multiple of 1-year, 27-days, 1-day, and even longer of 11-years

↓

Include the sin and cos of these periods into data to capture the periodic correlation





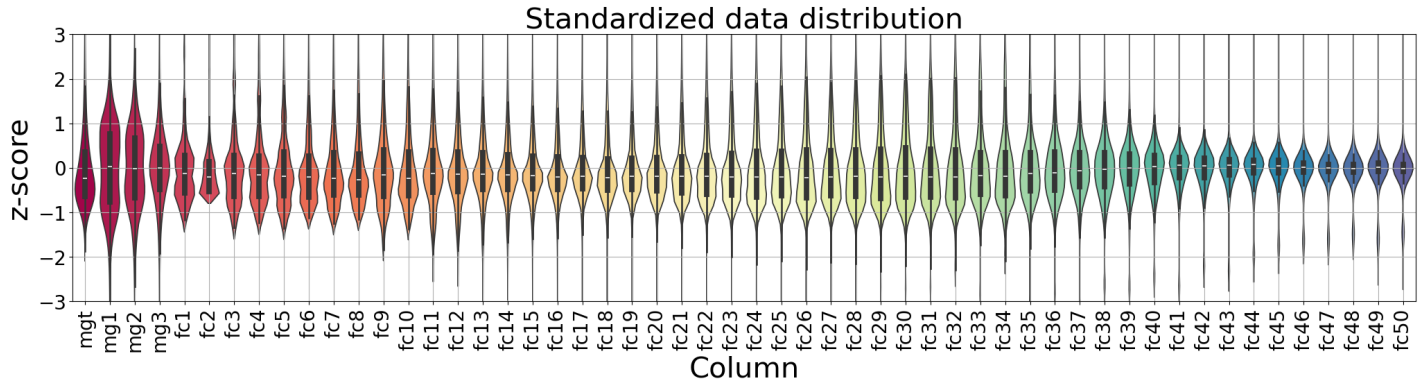


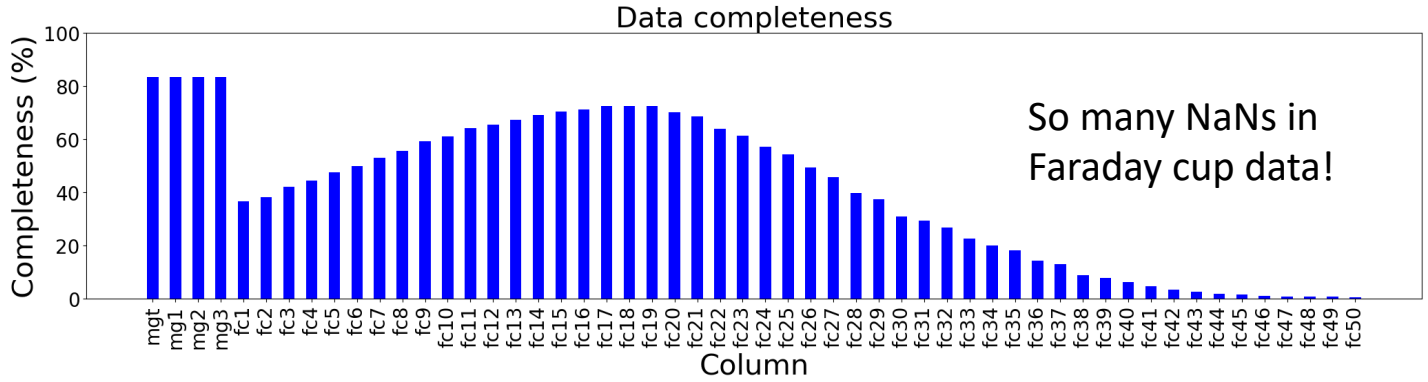
Inspect and preprocess data

Data type	Description	Number of columns
Time period	sin and cos with periods of 11yr, 1yr, 27day, 1day, 1hr(11Ysin, 11Ycos, …)	10
DSCOVR PlasMag	magnetic field vector and its magnitude(mg1-3, mgt), faraday cup data (fc1-50)	4(mg)+50(fc)=54
GFZ metric	sunspot number, F10.7obs, F10.7adj, Ap, ap, Kp	6
Not-nan checking	boolean to indicate whether the above data is not NaN(ck)	(10+54+6)x2=70

Data preprocessing:

1. Resample time resolution to 1hr
2. Split to training and testing data
3. Standardize data with the mean and std of the training data (no access to the testing data!)
4. Fill NaN with zero





Predict Kp-index with LSTM model

LSTM model

Past data (not limited to 24 hrs!)

↓

LSTM layer

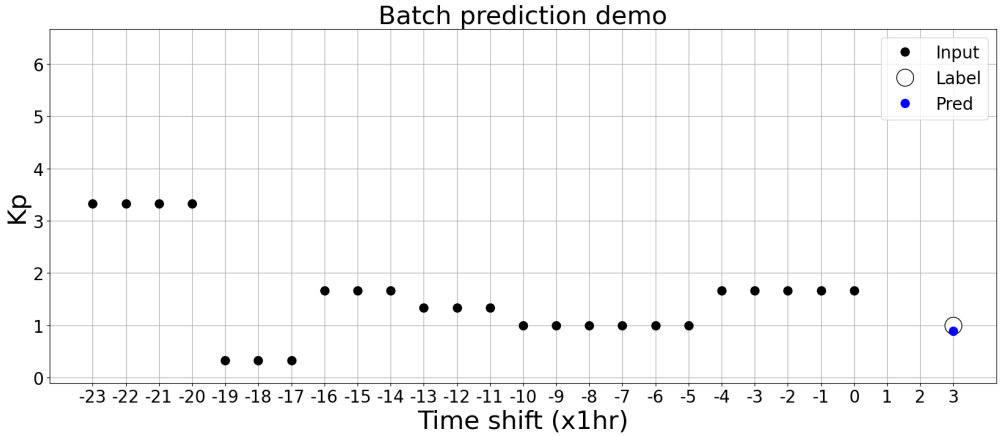
↓

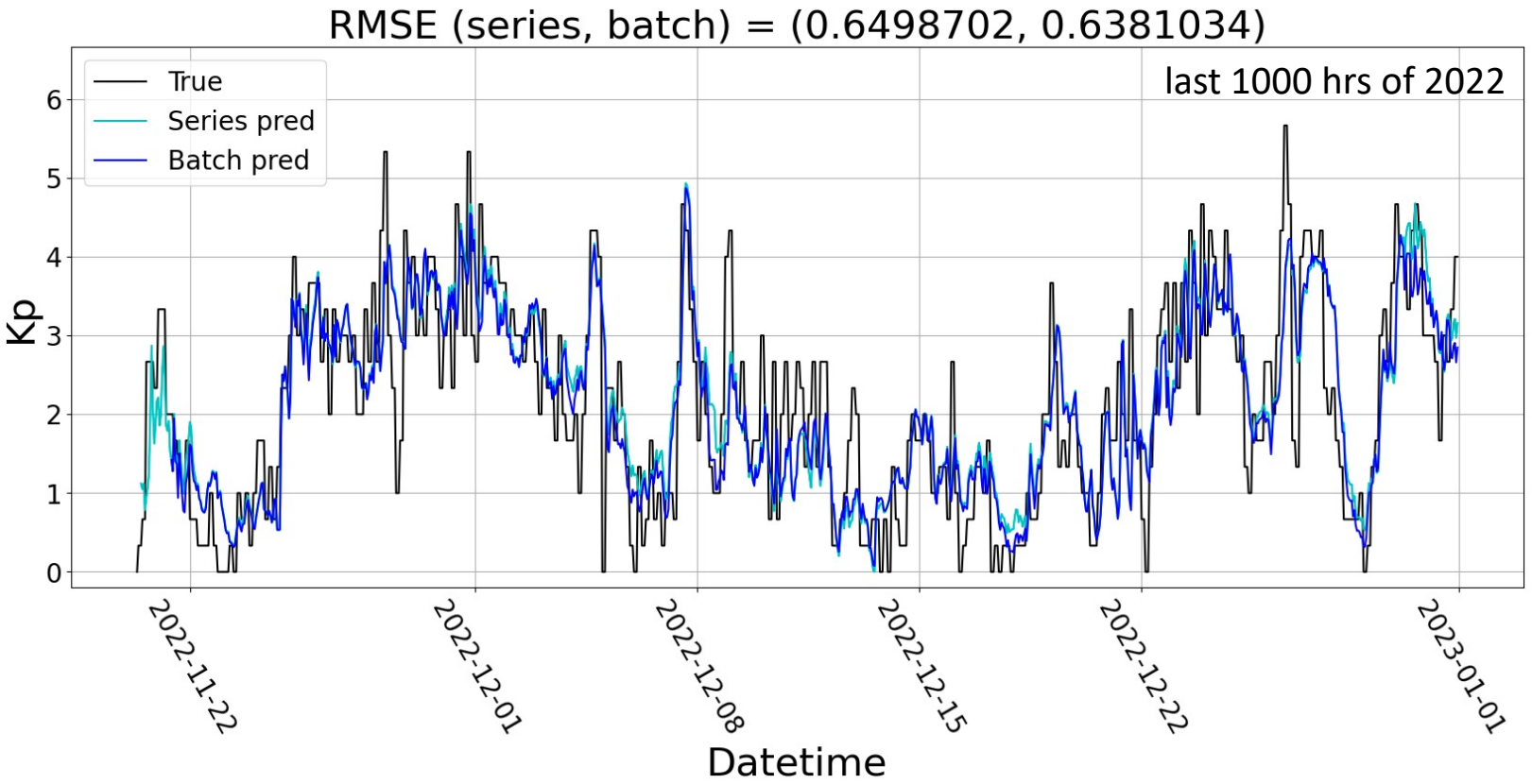
Hidden layer (size=128)

↓

3hrs in advance Kp

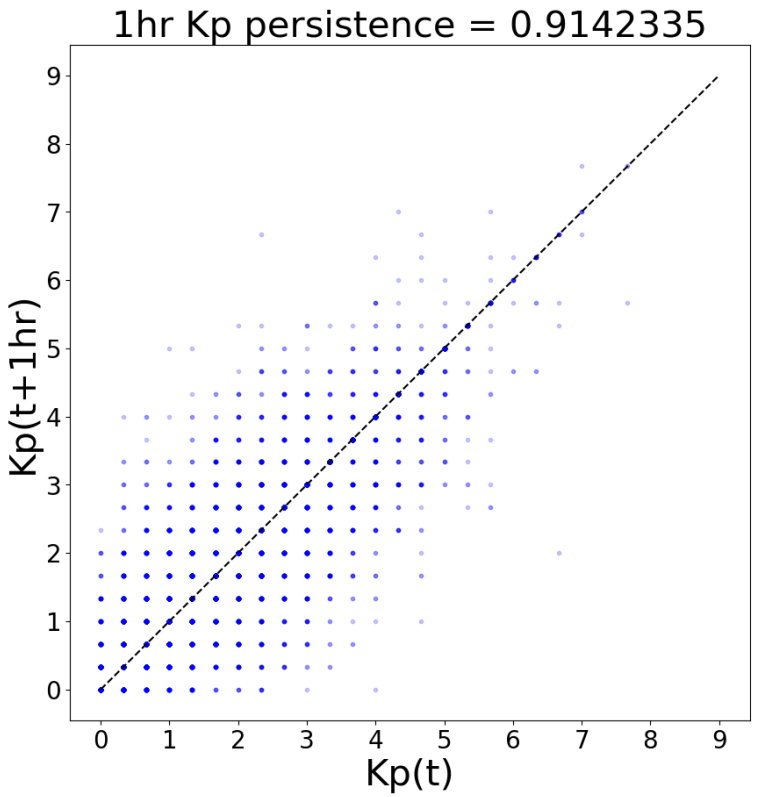
Use past 24 hours data to train the model that predict 3 hours in advance Kp



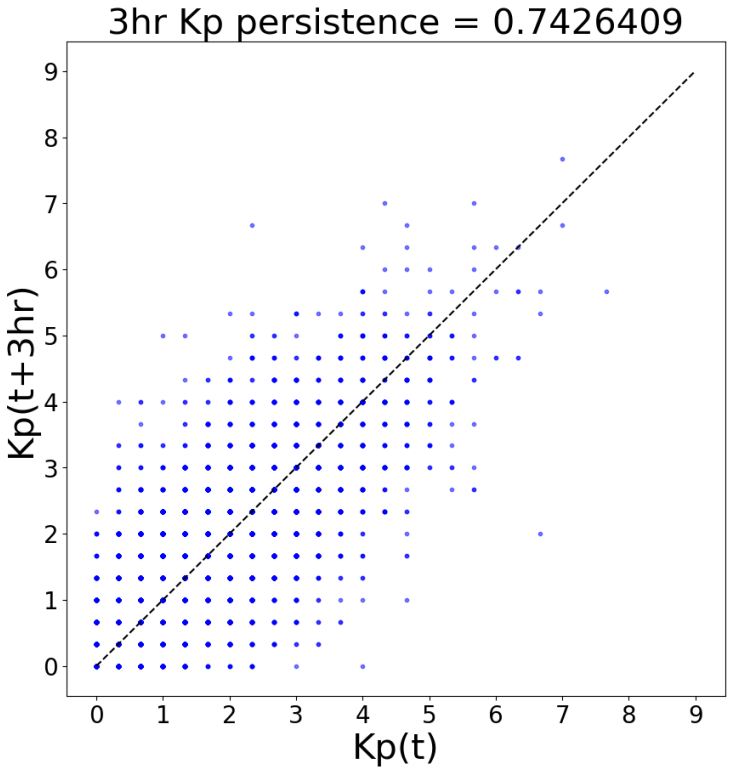


Kp persistence and Kp correlation

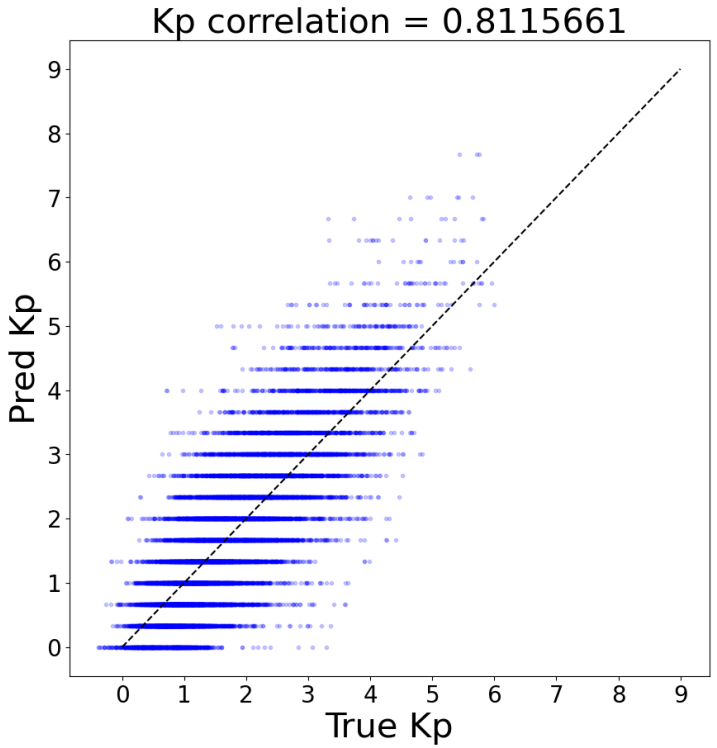
It is easy to predict 1hr Kp due to its high 1hr persistence, also DSCOVR can provide 15min-60min ahead warning.



In 3hr, exceeding the DSCOVR 60min warning, Kp persistence is lower.



Our model gives the higher correlation between pred Kp and true Kp.



pred-Kp error = 0.41

Gradient and set-to-nan analysis

Gradient analysis:

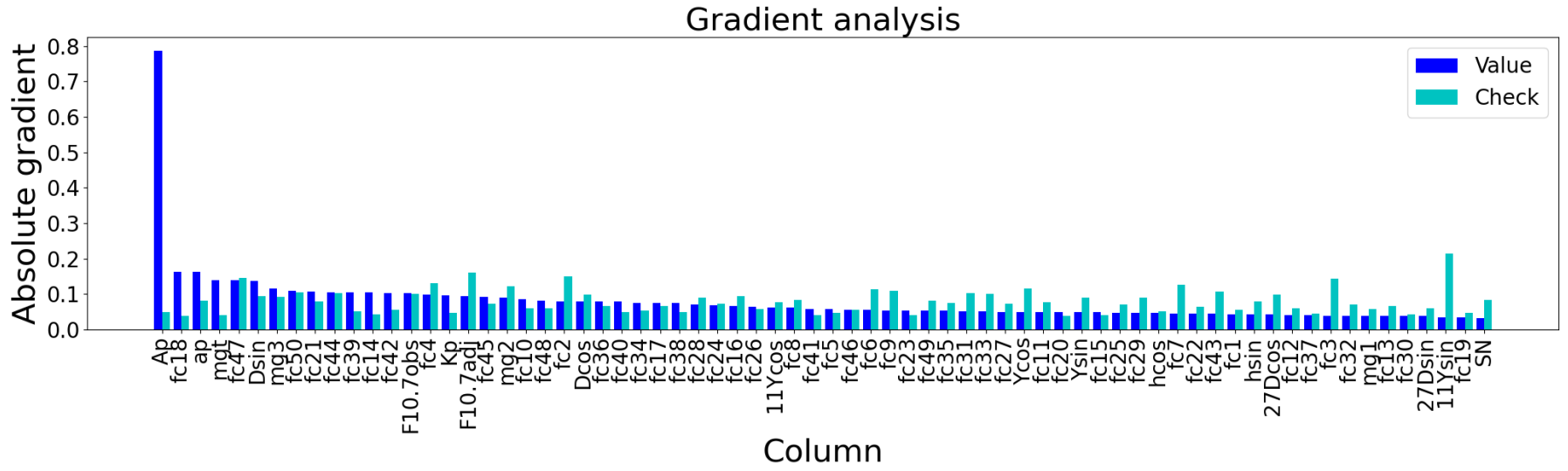
- Take grad of pred-Kp w.r.t. each input
- higher grad = Pred-Kp is more sensitive to this input.

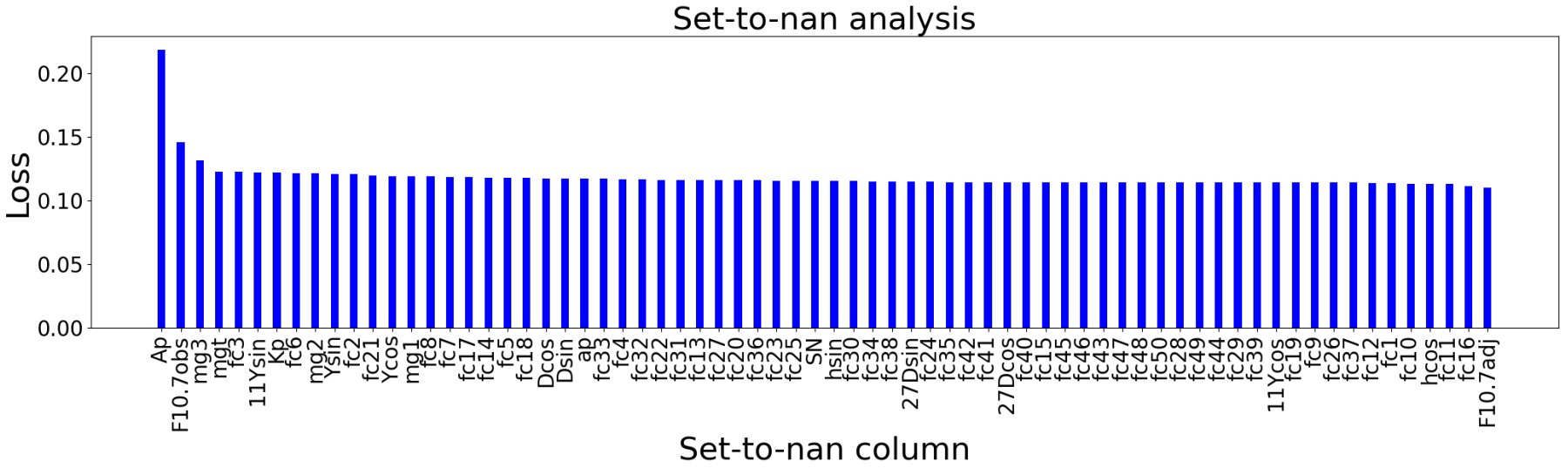
Set-to-nan analysis

- Set one type of input to be NaN
- higher loss = It is more important to feed this input

Insights

- Not surprising, Ap is important.
- In DSCOVR data, mgt, mg3 and some of fc data are important.





Insights and summary

- We train a LSTM model to predict 3hr in advance Kp only with the small error of 0.41, before DSCOVR 15min-60min warning.
- By FFT and set-to-nan analysis, the model rely on the 1 day and 11 years period to predict Kp.
- The gradient and set-to-nan analysis show that for DSCOVR data, mgt, mg3, fc3 and fc18 are important.
- The data completeness of Faraday cup data is low, which is the biggest anomalies of DSCOVR. However, the two analysis show that they are important.
- Though DSCOVR data contains large number of missing values, our model predicts very well.