Forecasting the Solar Energetic Protons Integral Flux using the Bi-Directional Long Short-Term Memory Neural Network

Mohamed Nedal*[1], Kamen Kozarev[1], Nestor Arsenov[1], Peijin Zhang[1], and Yordan Darakchiev[2]

[1] Institute of Astronomy with National Astronomical Observatory, Bulgarian Academy of Sciences (IANAO-BAS), Sofia, Bulgaria
[2] Department of Astronomy, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

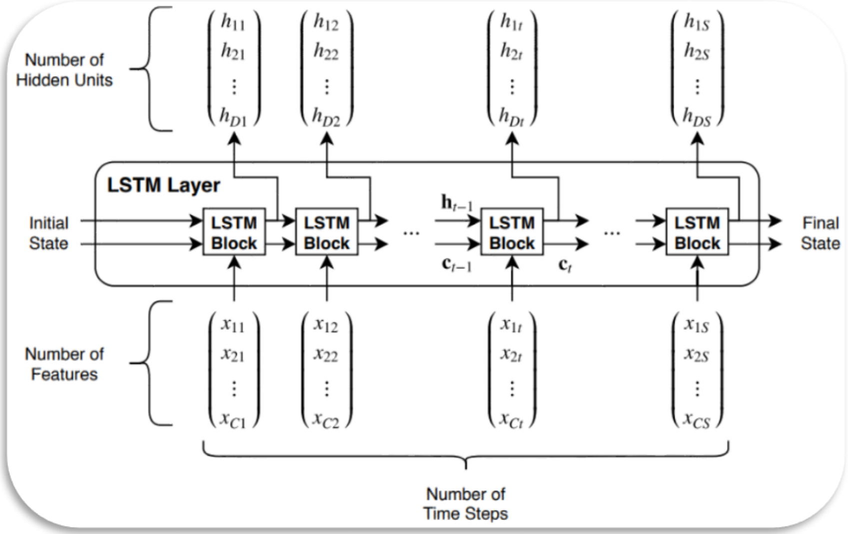
Machine Learning in Heliophysics
24th March, 2022

Introduction

- Solar energetic particles (SEP) are high-energy particles usually originating from solar eruptions, and they are composed of protons, electrons, and some heavy ions with energy ranging from a few tens of keV to several GeV
- All man-made space instruments are continuously exposed to SEPs that permeate the space environment
- That continuous exposure to solar particle radiation leads to charges to build up on the satellites and the spacesuits of astronauts, which cause electrical discharges, malfunctioning, and radiation sickness and increased risk of cancer
- Additionally, the highest energetic protons (>100 MeV) can enhance neutron count rates at ground levels through secondary radiation effects, which are known as Ground Level Enhancements (GLEs) which in turn cause problems on Earth
- Therefore, it is essential to study the SEPs and forecast their flux at the near-Earth orbit
- In this work, we implement a short-term and long-term forecasting models to predict the solar protons integral flux at 1 AU using the Bi-directional Long Short-Term Memory (LSTM) neural network technique

Bi-LSTM Neural Network Model

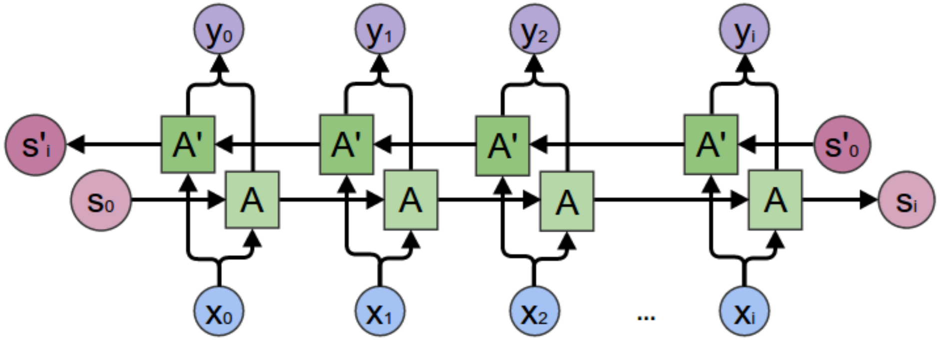
- The LSTM networks are a subset of the Recurrent Neural Network (RNN) used in deep learning to address contextual information by integrating a loop that allows information to flow from one time step to the next
- This is managed by learning when to remember and when to forget, through their forget gate weights
- In the case of the Bi-directional LSTM network, the input flows in two directions to preserve the future and the past information, which gives better results in our case compared to the regular LSTM model
- This diagram shows the flow of a time series *X* with *C* features of length *S* through a regular LSTM layer — Here, ***h_t* and *c_t* denote the output (*also known as the hidden state*) and the cell state at time step *t***, respectively



Source: MathWorks

Bi-LSTM Neural Network Model – cont.

- The Bi-LSTM layer is made up of 2 layers – the forward layer and the backward layer
- The forward layer process the information from the past to the future (+ time direction), and the backward layer does the opposite (- time direction)
- Here, *A* and *A'* are the activation cells in the forward and backward layers, respectively
- *Xi* and *Yi* are the input and output sequences, respectively
- *S* and *S'* are the hidden states in the forward and backward layers, respectively

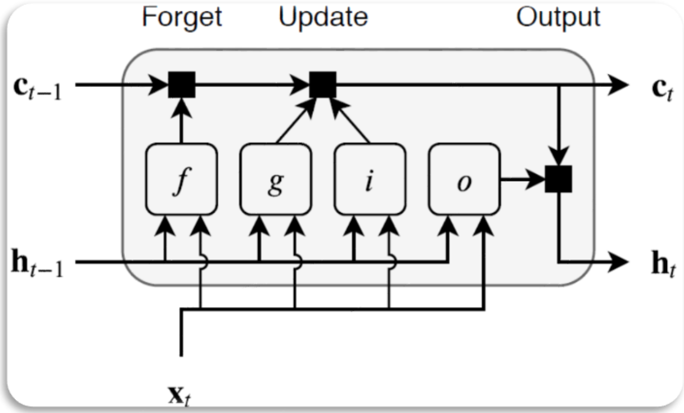


Source: colah's blog

Bi-LSTM Neural Network Model – cont.

The LSTM unit is composed of a cell, an input gate, an output gate and a forget gate — The cell remembers values over the time steps and the 3 former gates adjust the information stream into and out of the cell

This diagram shows the flow of data at time step *t* and highlights how the gates forget, update, and output the cell and hidden states — Here, ***f, g, i, o* are the forget gate, the cell candidate, the input gate, and the output gate**, respectively



Source: MathWorks

Bi-LSTM Neural Network Model – cont.

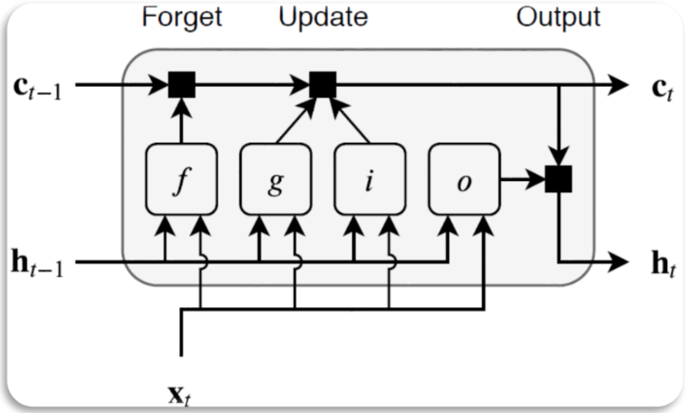
We used **9** features –

the **sunspot number**; obtained from the World Data Center for the production, preservation, and dissemination of the international sunspot number;

the **solar radio flux density**, the **IMF**, the **SW speed**, and the **proton integral flux** at *3* energy channels (*>10 MeV, >30 MeV, and >60 MeV*) obtained from OMNI database;

the *long*- and *short*-wavelength bands of **X-ray flux**; obtained from GOES database

The reason for choosing those features is because the dynamics of the solar activity influence the protons flux since they travel within the inner heliosphere — By doing correlation analysis, we selected the **top 6** correlated features with the proton flux for each energy channel



Source: MathWorks

Bi-LSTM Neural Network Model – cont.

The data is split into 75% (from 1976 to 2008) for training the model and 25% (from 2008 to 2019) for validating the performance

A multivariate multi-step Bi-LSTM NN model is implemented, based on the *Multiple Output Strategy*, to forecast the integral protons flux in 3 energy channels (>10 MeV, >30 MeV, and >60 MeV) throughout the following 6 hour, 12 hours, and 24 hours for the short-term mode – and throughout the following 3 days, 5 days, and 7 days for the long-term mode

N_layer: 2 Bi-LSTM layers + 1 Dense layer

N_cells: 32
Batch size: 512, Epochs: 70

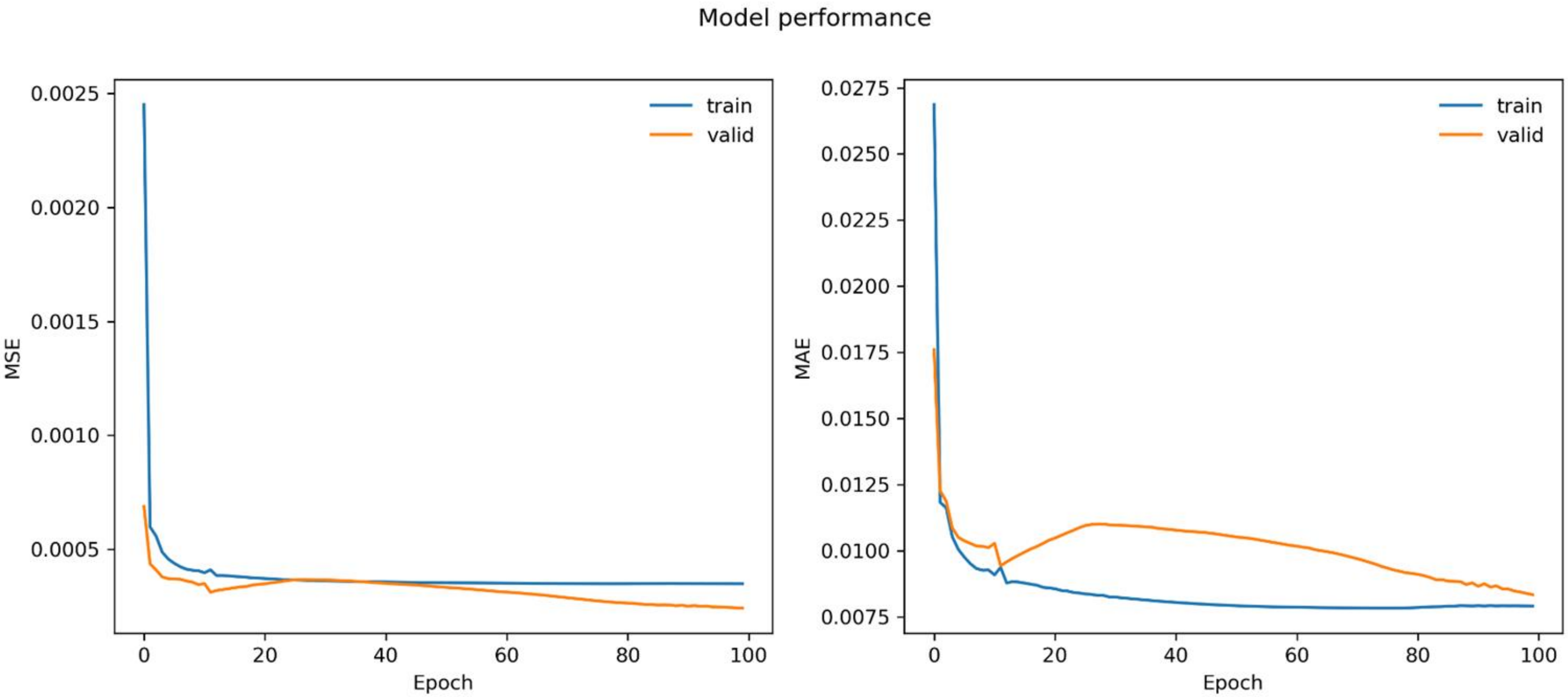
N_input timesteps: 730 (~24 days) → daily data

N_input timesteps: 720 (~30 hours) → hourly data

Example of Hourly Data

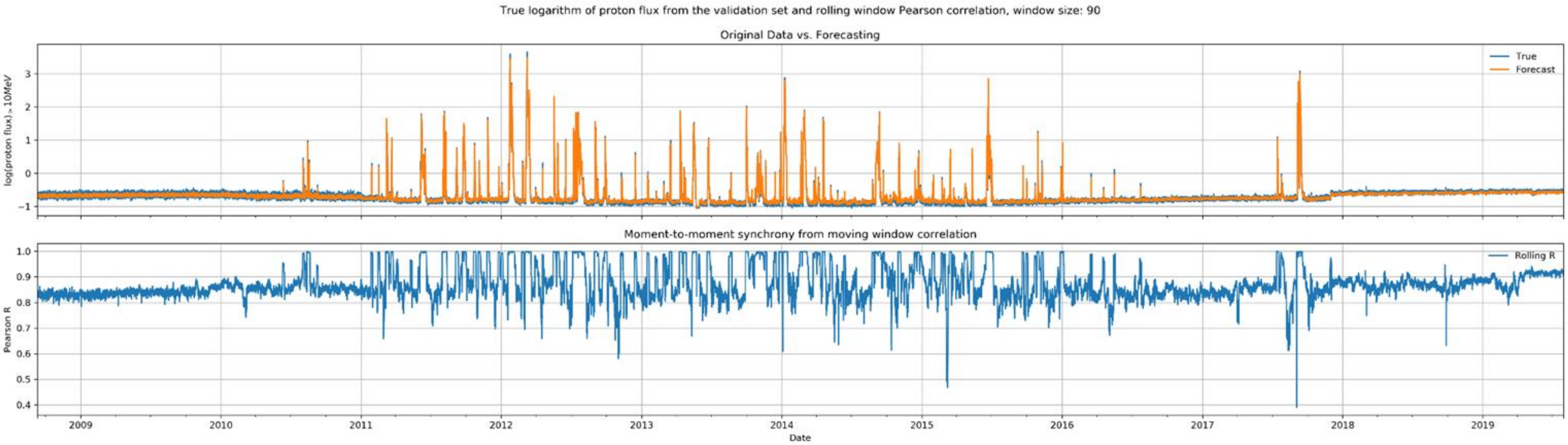
6-hr pf >10 MeV

Results: 6-hr forecasting of proton flux >10 MeV



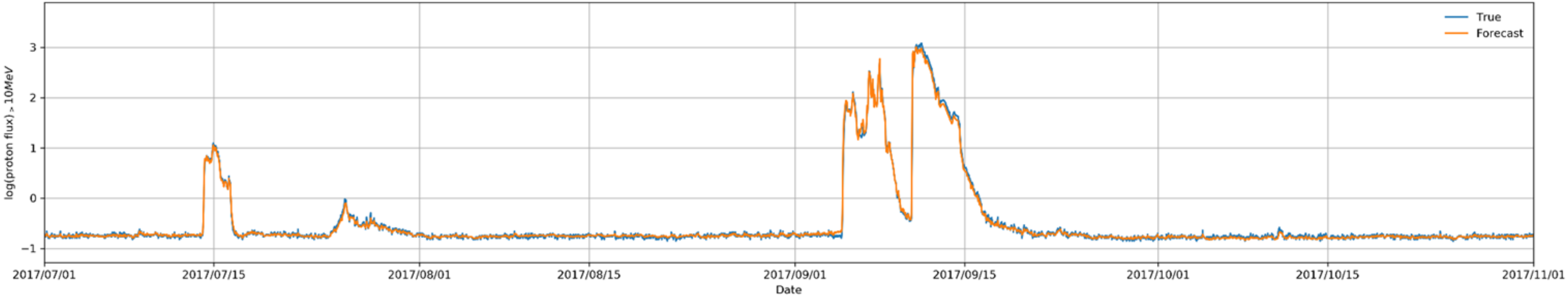
Loss curve to check the performance using MSE and MAE as the error metrics

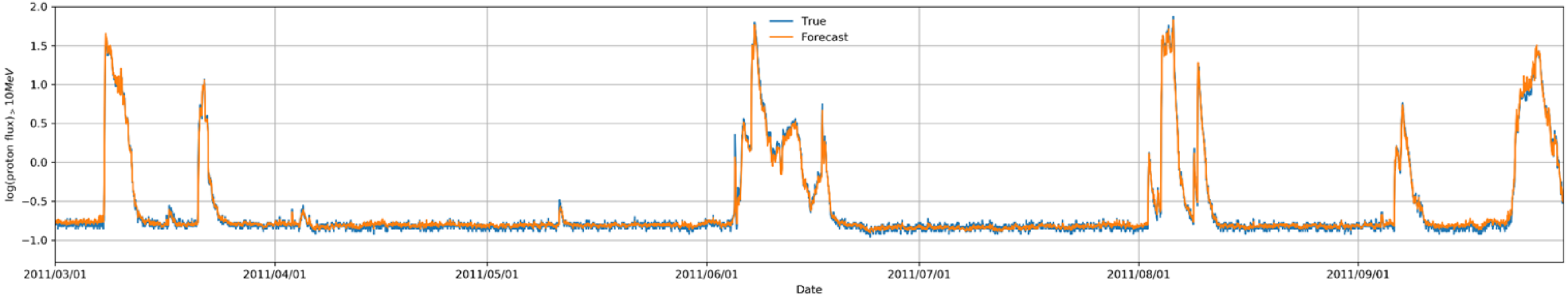
Results: 6-hr forecasting of proton flux >10 MeV – cont.



Comparison between the model's output and the validation set, with the rolling-window Pearson corr. of window size = 90

Results: 6-hr forecasting of proton flux >10 MeV – cont.



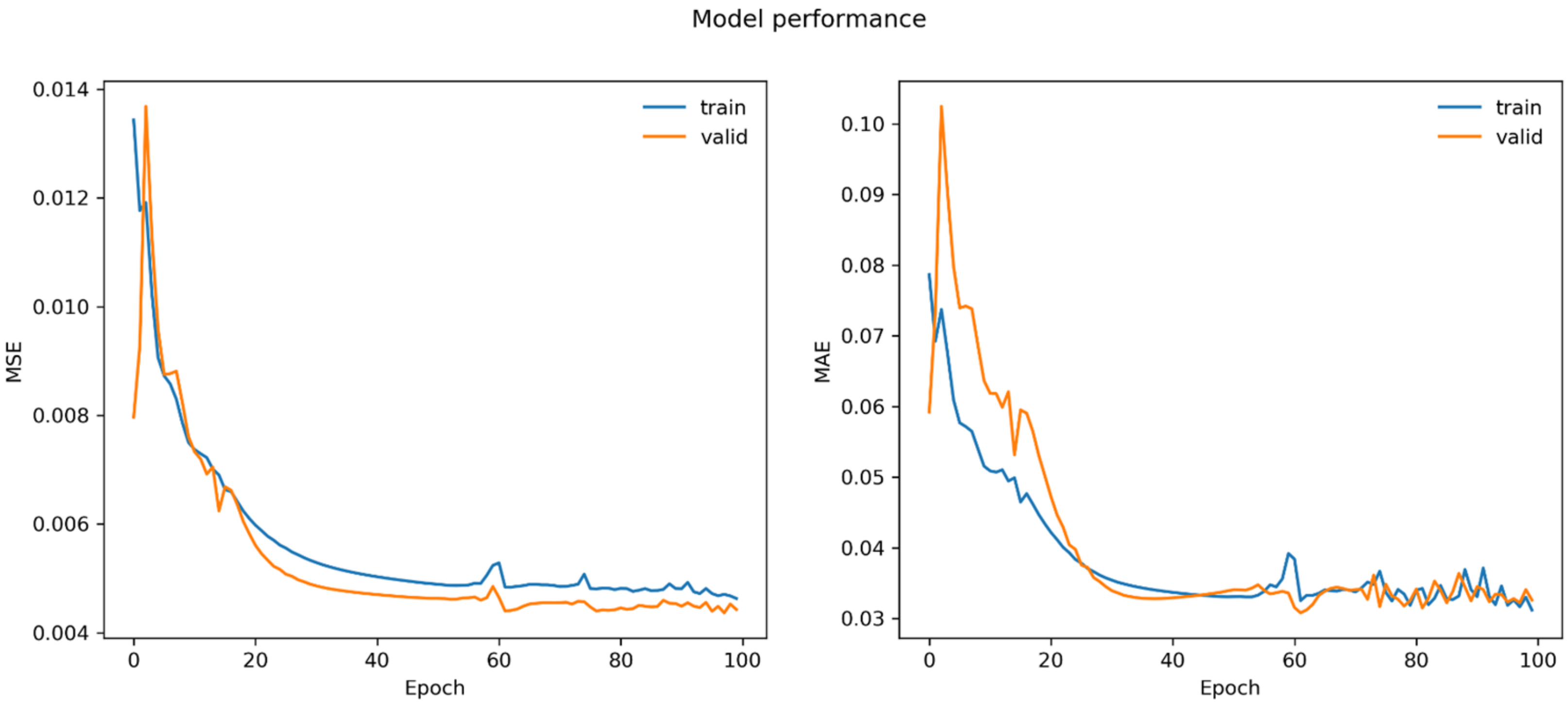


Comparison between the rea and modelled data – parts of the validation set

Example of Daily Data

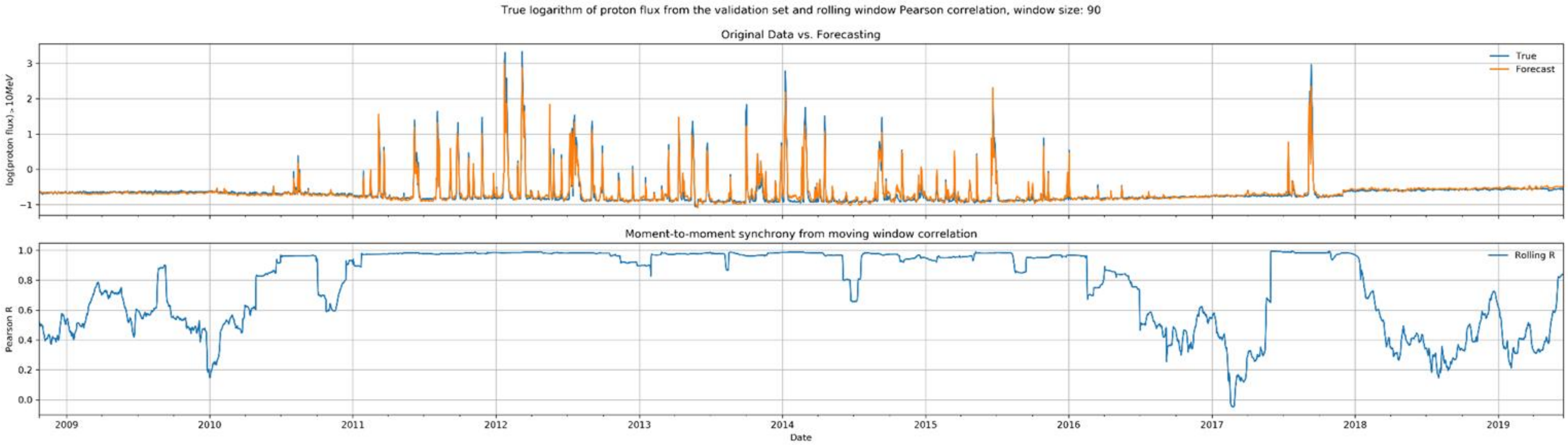
3-day pf >10 MeV

Results: 3-day forecasting of proton flux >10 MeV



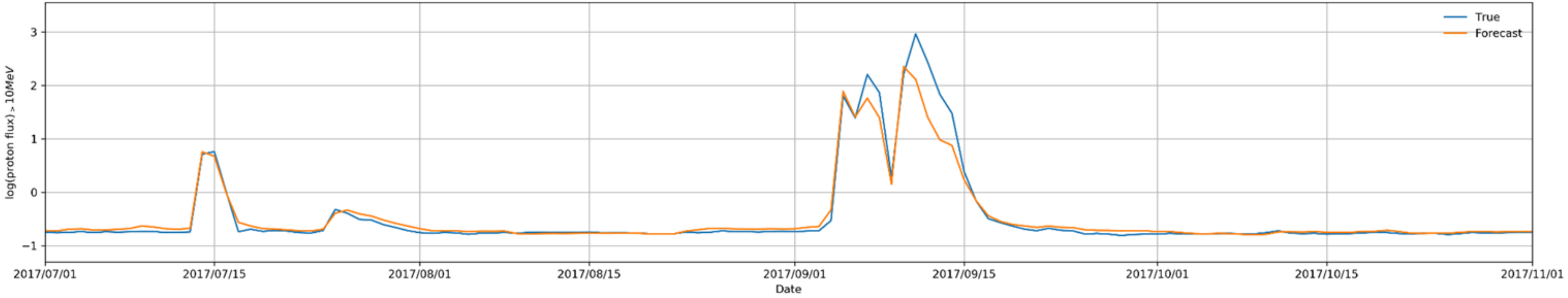
Loss curve to check the performance using MSE and MAE as the error metrics

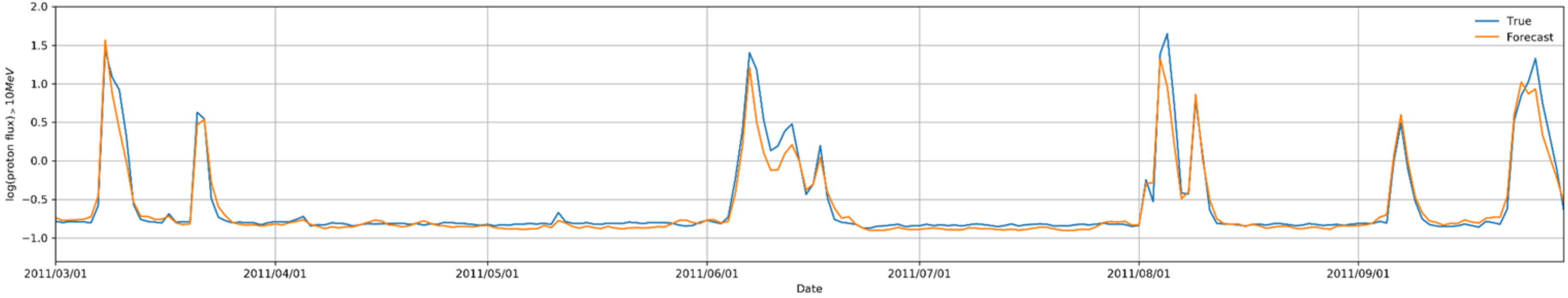
Results: 3-day forecasting of proton flux >10 MeV – cont.



Comparison between the model's output and the validation set, with the rolling-window Pearson corr. of window size = 90

Results: 3-day forecasting of proton flux >10 MeV – cont.





Comparison between the rea and modelled data – parts of the validation set

Prediction Errors for the Validation Set

3-day

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.009 MAE = 0.054	MSE = 0.382 MAE = 0.505	MSE = 0.334 MAE = 0.476

5-day

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.012 MAE = 0.065	MSE = 0.370 MAE = 0.498	MSE = 0.294 MAE = 0.446

7-day

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.012 MAE = 0.070	MSE = 0.354 MAE = 0.487	MSE = 0.290 MAE = 0.447

6-hr

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.001 MAE = 0.027	MSE = 0.618 MAE = 0.676	MSE = 0.442 MAE = 0.571

12-hr

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.173 MAE = 0.305	MSE = 0.536 MAE = 0.632	MSE = 0.442 MAE = 0.571

24-hr

pf >10 MeV	pf >30 MeV	pf >60 MeV
MSE = 0.173 MAE = 0.305	MSE = 0.536 MAE = 0.632	MSE = 0.442 MAE = 0.571

Conclusion

- We implemented forecasting models to do short-term and long-term forecasting for the integral protons flux in 3 energy channels based on the data of the previous 4 solar cycles and by using 7 input features that reflect the solar activity state
- The MSE of prediction of the PF>30 MeV channel is generally the highest in both the long-term and short-term forecasting, while the MSE for the PF>10 MeV is the lowest
- For the long-term forecasting, the MSE increases at larger future horizons, as expected, except for the PF>30 MeV and PF>60 MeV — The MSE for the PF>10 MeV is similar up to 3 digits for the 5-day and 7-day forecasting
- The MSE for the PF>10 MeV is similar up to 3 digits for the 12-hr and 24-hr forecasting. The same applies for the PF>30 MeV
- The MSE for the PF>60 MeV is similar up to 3 digits for the 3 forecasting windows, which means that changing the future horizon has very little impact on the model performance
- The model still needs more fine-tuning and the performance can potentially be greatly improved

Thank You!