Deep Horizon

Modeling of Earth's Radiation Environment. Final Presentation

Team: Florian Peste, Jens Huthmacher, Marcus Münzer, Tanveer Hannan, Ziyang Zhou

Supervisor: Max Berrendorf, Evgeniy Faerman, Dr. Elena Kronberg

Supervising Professor: Prof. Dr. Matthias Schubert

Introduction

Introduction
Background

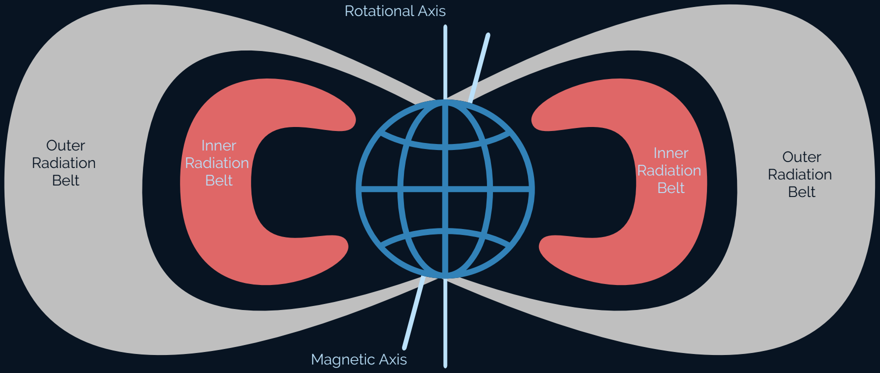
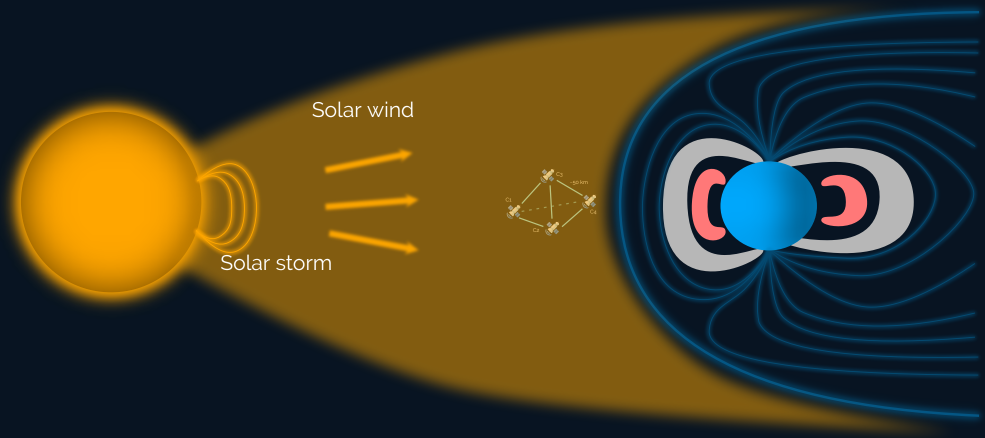


Figure: https://en.wikipedia.org/wiki/Van_Allen_radiation_belt

Introduction
Background



Introduction
Motivation

Energetic particles

- ... are hazard for modern spacecrafts (e.g., satellites)
- ... contaminate data in space observation (e.g. XMM)

Such contamination

- ... could cause severe data loss
- ... is highly dynamic
- ... is not yet well understood by the physicists

Introduction
Data

Cluster

- **Proton Intensities on 7 Energy Channels** (p1 – p7)
- **Positions** (Coordinates, Distance from Earth)

OMNI

- **Geomagnetic Activity** (AE Index, SYM-H Index)
- **Solar Activity** (Solar Radio Flux)
- **Solar Wind Activity** (Speed, Density, Temperature, Dynamic Pressure)

Foot Type: Connection to the Magnetic Field Line to the Earth

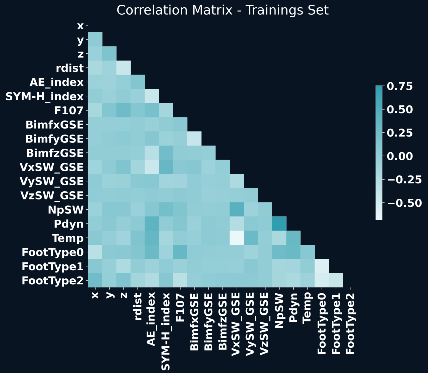
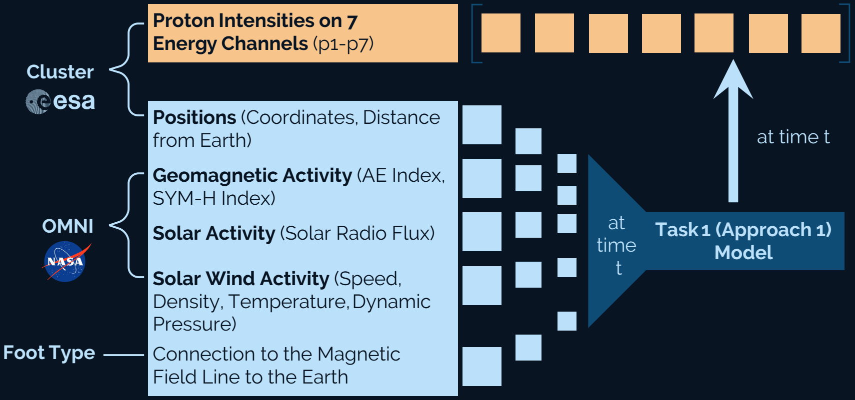


Figure: Correlation Matrix (Trainings Set)

Our Approaches
Approach 1



Our Approaches
Approach 2

Cluster esa

Proton Intensities on 7 Energy Channels (p1-p7)

Positions (Coordinates, Distance from Earth)

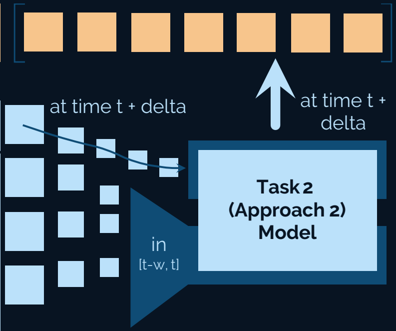
OMNI NASA

Geomagnetic Activity (AE Index, SYM-H Index)

Solar Activity (Solar Radio Flux)

Solar Wind Activity (Speed, Density, Temperature, Dynamic Pressure)

Foot Type — Connection to the Magnetic Field Line to the Earth

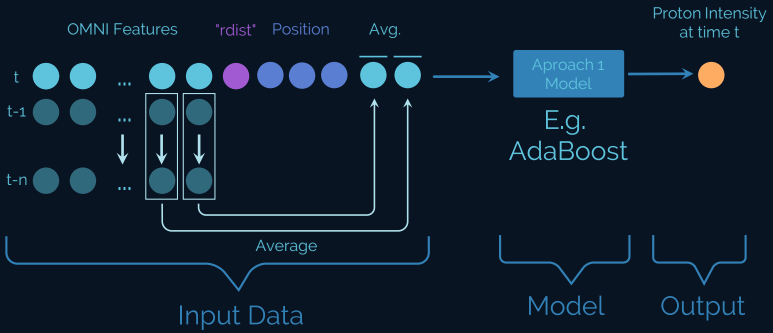


Approach 1

Approach 1
Architecture

Basic architecture for each channel 1-7:

- $Input + OMNI_t \rightarrow$ Prediction of proton intensities at timestamp t
- **Additional features**: Average of the last 1, 2, 4, 8, 16 ... hours of data
- **Output**: Proton intensities at timestamp t for channel p



Approach 1
Models and Techniques

Tree-Based

- Averaging:
  - → Averages independent estimators
  - → Low variance
  - → **ExtraTrees**, **RandomForest**
- Boosting:
  - → Sequentially combined weak estimators
  - → Robust, works well with non-linear decision boundaries
  - → **AdaBoost**, **HistGradientBoosting**, **GradientBoosting**, **LightGBM**
- Decision Trees (Regression Trees):
  - → Simple decision rules
  - → Low costs

Approach 1
Models and Techniques

Non Tree-Based

- Linear:
  - $\rightarrow$ Linear combination of features
  - $\rightarrow$ **LarsRegression**, **RidgeRegression**
- **kNeighborsRegression**

Baseline:

- Mean: Predicts the mean of a channel
- Historical Binning: Predicts mean over an according spatial bin

Approach 2 (DeepHorizon)

Approach 2 (DeepHorizon)
Architecture

Proton intensities:

- For a point p in 3D space (i.e. (x, y, z))
- At time $t + \Delta$

Parameters:

- Sequence: 3 hours
- Forecast: 5 minutes

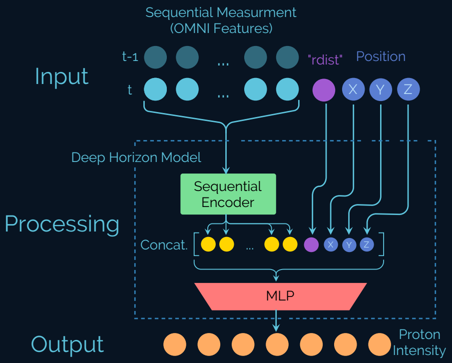


Figure: Deep Horizon Architecture (Approach 2)

Approach 2 (DeepHorizon)
Model Classes

Model Class	Model(s)	Inductive Bias
RNN	LSTM / GRU	Sequentiality
CNN	(Causal) CNN (TCN)	Locality
Differential Equation	Neural ODE / Neural CDE	Continuity

Approach 2 (DeepHorizon)
NeuralODE/CDE

Idea:

- Learn the **rate of change**

Recipe:

- Learn **derivative** $f_\theta = \frac{\partial F_\theta}{\partial x}$
- Integrate over derivative $F_\theta = \int f_\theta$
- Solve differential equation by adding initial condition $z_0 + \int f_\theta$

Example:

$$y = x^2 + 5$$
$$y' = 2x$$

Perfect case:

$$y_{learned} = 5 + \int 2x \, dx$$
$$y_{learned}(5) = 5 + 25 = 30 = y(5)$$

Final Model:

$y \approx \ell_\theta^1(z_T)$, where $z_t = z_0 + \int_0^t f_\theta(z_s)ds$ and $z_0 = \ell_\theta^2(x)$

Downside:

- Initial-value-problem $\rightarrow$ **Neural CDE**

Results

Results
Evaluation of the Outcome

❞ We were able to reach very good results in predicting/forecasting proton intensity!

How did we evaluate our models?

- Initially we calculated 5 different metrics for evaluating our models
  → MSE, MAE, Pearson Correlation, Spearman Correlation, R2 Score / Prediction Efficiency
- Finally, we just use the **Spearman Correlation**

Explanation Spearman-Correlation

- Statistical measurement of the strength of a **monotonic** relationship
- The closer SC is to ± 1 the stronger the monotonic relation

$$SC = \frac{covariance(rank(X), rank(Y))}{std(rank(X)) \cdot (rank(Y))} \qquad with,\ SC \in [-1, 1]$$

Results
Approach 1

🏆 Overall best result **0.63** on channel 1 from **ExtraTree**!

Channel	Model	SC
1	**ExtraTree**	**0.63**
2	GradientBoostingReg	0.5639
3	AdaBoostReg	0.5923
4	AdaBoostReg	0.5637
5	HistGradientBoostingReg	0.4579
6	GradientBoostingReg	0.2939
7	GradientBoostingReg	0.2926

Simple Regression and all leading models are **tree-based**!

- General known: Tree-Based Models are quite good!
- Self-learning $\rightarrow$ less dependent on paramter than e.g. SVM

Results
Approach 1 - Plots (Best Model)

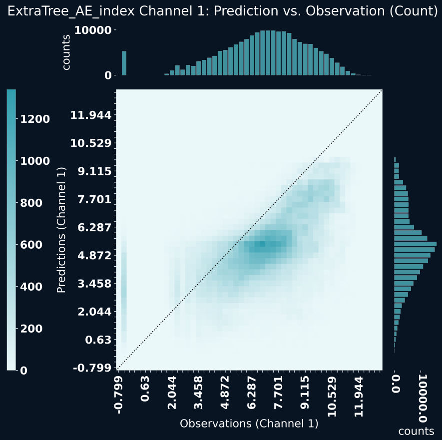
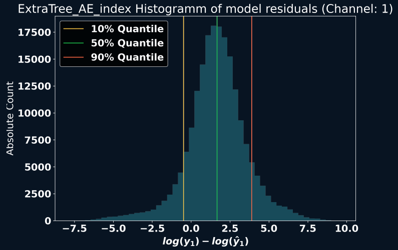


Figure: ExtraTree (Channel 1): Prediction vs. Observation



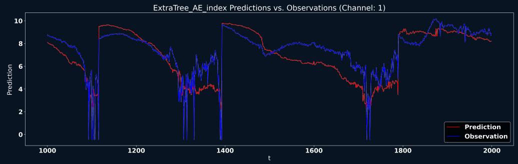


Figure: ExtraTree (Channel 1): Residuals and Predictions vs. Observation over time

Results
Approach 1 - Feature Importances

Feature importances for second best task 1 model (**AdaBoost**) on channel 1.

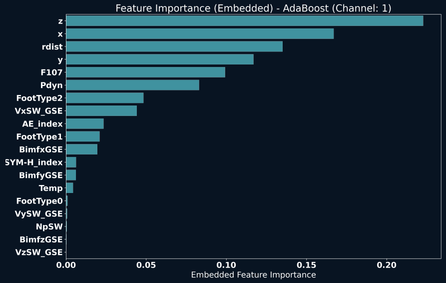


Figure: Embedded Feature Importance

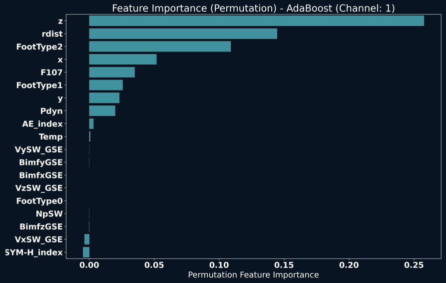


Figure: Permutation Feature Importance

Results
Approach 2

Best task 2 result **0.6537** on channel 1 from **GRU**!

Channel	Model	SC
1	**GRU**	**0.6537**
2	GRU	0.5687
3	GRU	0.604
4	GRU	0.5607
5	GRU	0.3988
6	NCDE	0.2022
7	NCDE	0.1918

We beat the task 1 models on nearly every channel!

Results
Approach 2 - Plots (Best Model)

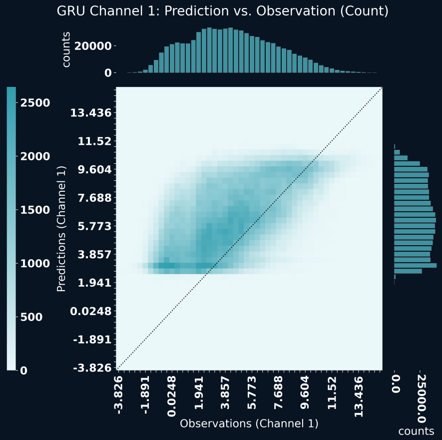
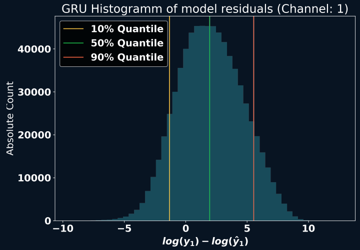


Figure: GRU (Channel 1): Prediction vs. Observation (Observations > 1)



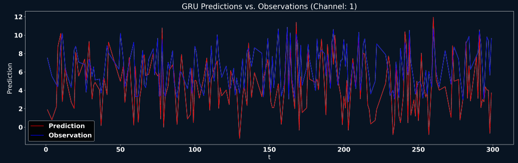


Figure: GRU (Channel 1): Residuals and Predictions vs. Observation over time

Results
Approach 1 vs. Approach 2

🏆 Overall best result **0.6537** on channel 1 from **GRU**!

Channel	SC (Task 1)	SC (Task 2)
1	**0.63**	**0.6537**
2	0.5639	0.5687
3	0.5923	0.604
4	0.5637	0.5607
5	0.4579	0.3988
6	0.2939	0.2022
7	0.2020	0.1918

What can we derive from those results:

- Task 2 models achieved slightly better performance
- Task 2 differences that could relate to the output:
  → ... more complex models to capture highly dynamic relations
  → ... sequential input
  → ... ability to capture trends
- Channel 6/7 in general quite hard to prediction
  → ... a lot of missing data!
  → ... unpredictable particle behavior in hight channels (high dynamic)

Results

Some remarks on statistical interpretability

2-sided hypothesis test on spearman correlation:

- Null hypothesis: Predictions are uncorrelated to Observations
- Result: $p = 0$ for all channels $\rightarrow$ Null hypothesis is rejected
- Interpretation: **Our model learned the trend of proton intensities!**

Paired t-test on residuals:

- Null hypothesis: Mean of residuals $= 0$
- Result: $p = 0$ for all channels $\rightarrow$ Null hypothesis is rejected
- Interpretation: Predictions has systematic error

Results
XMM Use Case

About XMM:

- The XMM-Newton X-Ray telescope is the biggest scientific satellite ever built in Europe
- It has very sensitive cameras that can see much more than any previous X-ray satellite

Use Case

- We have data about the contamination level of the XMM telescope
- Idea:

  $\rightarrow$ Take the x, y, z coordinates from the XMM + OMNI data
  $\rightarrow$ Use best model for predicting the 7 energy levels
  $\rightarrow$ Goal: See which channel correlates best with the "contamination level"

Result: We were not able to see a correlation between our predictions and the XMM contamination level.

Results
Conclusion and Future Work

"We achieved very promising results and we were able to make a good contribution to predicting proton intensites by implementing ...

- ... a "simple" **regression approach** and
- ... a **sequential model**

Future Work:

Fusion of temporal (OMNI) and non-temporal (spatial) features: $encode(OMNI_{t-\Delta:t}|POS_t)$

Modeling systematic error (we are in a highly dynamic set up)

Results
Practical Usage

Finally, what can we do with the work we did? Here we have two example use cases.

Hydroelectric Powerplants

- Geomagnetic storms can damage hydroelectrostations

🏆 Forecasting a high proton intensity could help react on geomagentic storms beforehand (e.g. shut down systems)!

Space Measurments

- High proton intensity are hazardous for satellites (e.g. harm components)
- High proton intensity affect and disturb the measurments

🏆 Forecasting a high proton intensity at a specific position could help to react on the situation, e.b. change trajectory of satellite.

In general: The models and their predictions could help to understand contamination and influencing factors in the outer space.

Src.: https://www.nasa.gov/press-release/nasa-astronauts-launch-from-america-in-historic-test-flight-of-spacex-crew-dragon

A SpaceX Falcon 9 rocket carrying the company's Crew Dragon spacecraft is launched from Launch Complex 39A on NASA's SpaceX Demo-2 mission to the ISS with NASA Saturday, May 30, 2020, at NASA's Kennedy Space Center in Florida

Thank you!
Questions?