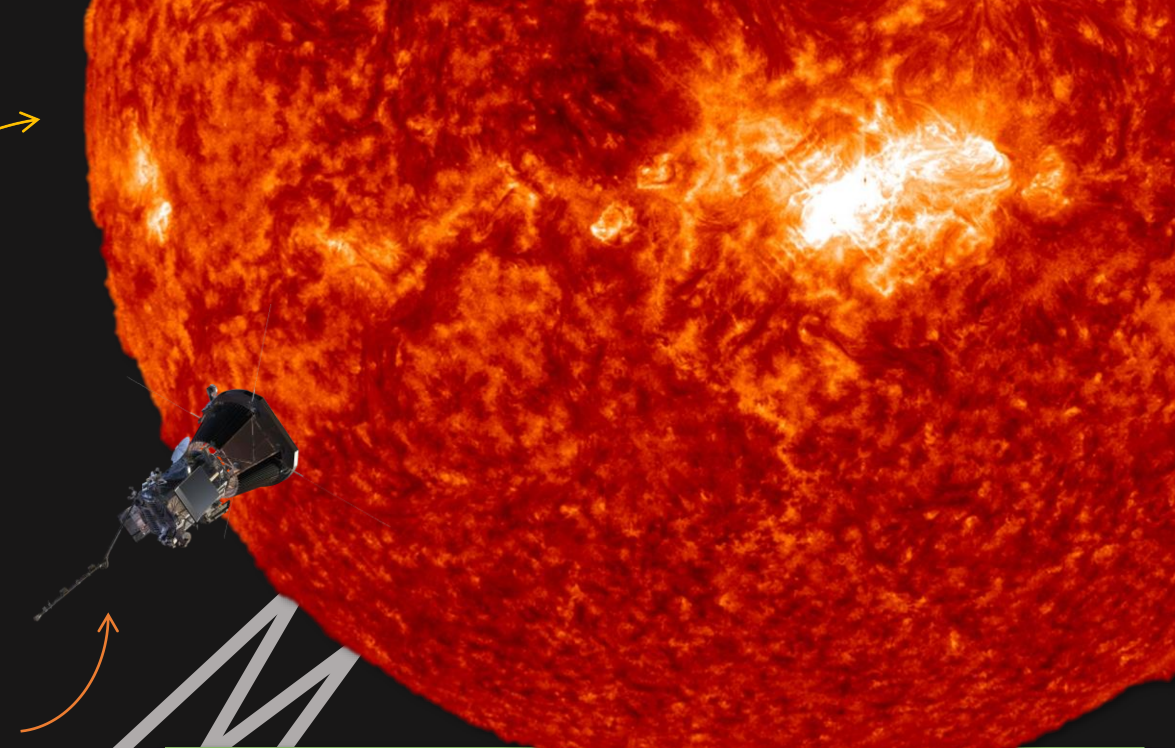
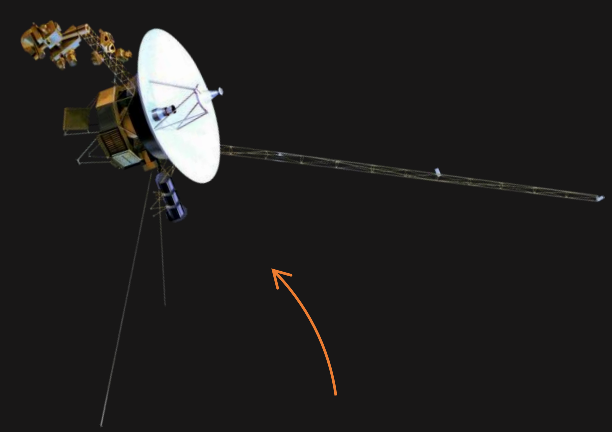
FILLING THE GAPS

Neural Nets for Space Stats





THE DATA

The Sun is constantly producing a stream of charged particles that race through the solar system. This "solar wind" is a key part of space weather, and can endanger our space technologies during periods of intense solar activity. Understanding this complicated system is crucial to protecting our satellites, power grids, and astronauts from harm. A host of satellites provide us with measurements of various aspects of space weather. Here we work with the magnetic fields.

THE STATS

A time series of the magnetic field strength of the solar wind often appears chaotic and unpredictable. In order to make sense of this data, we calculate statistics that summarise how the field is changing. In particular, we can examine the scales on which the field fluctuates. Here we focus on the structure function, a function of distance between two points that is central to turbulence models. These, in turn, are critical to improve our space weather models.

MAGNETIC FIELD OVER TIME

$$\langle |B_j(t) - B_j(t+\tau)|^p \rangle$$

Mathematical equation

Neural network?

STRUCTURE FUNCTION

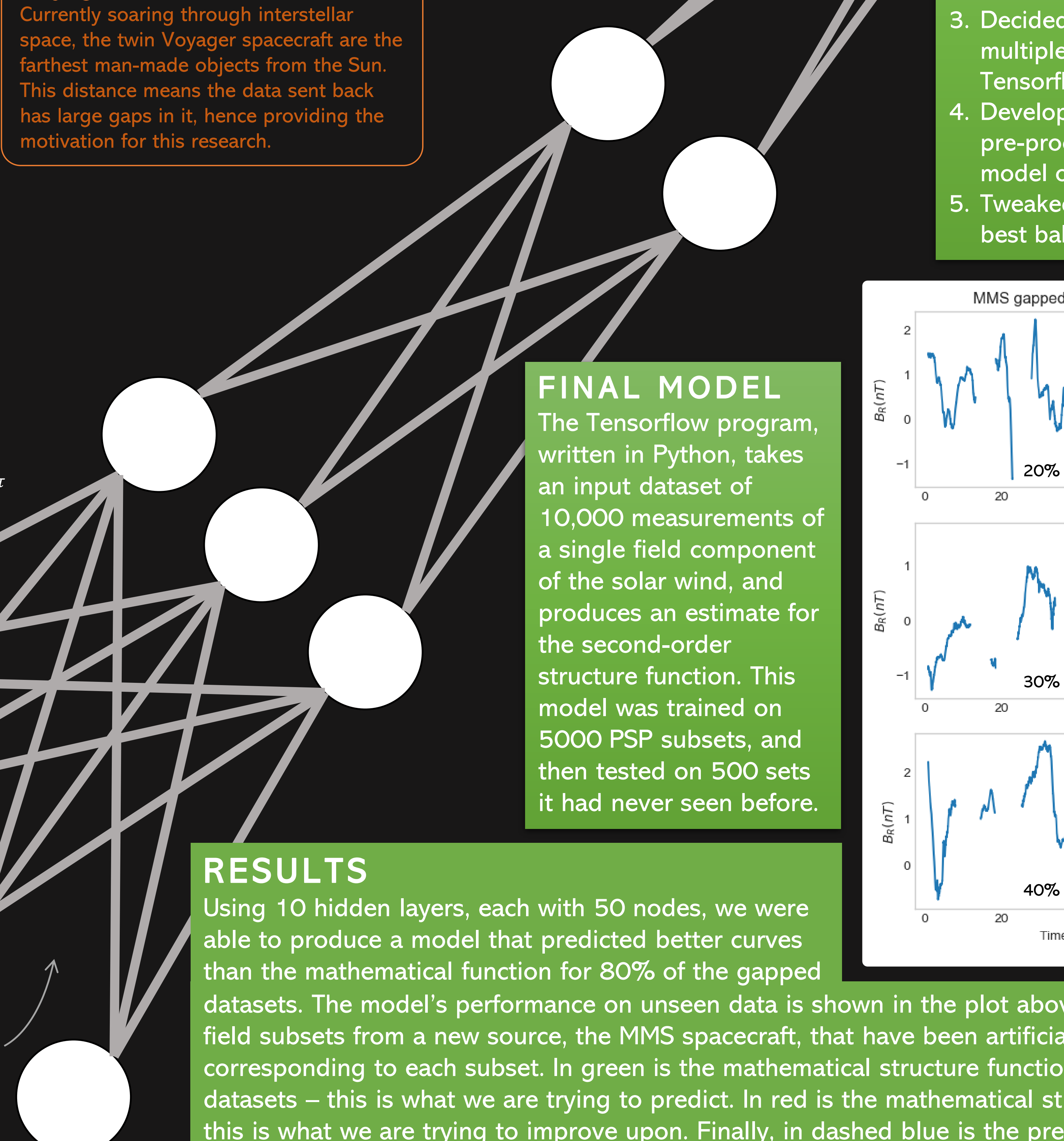
THE PROBLEM

Telemetry, instrumentation and calibration problems onboard spacecraft mean the data they send back to Earth often has large gaps in it. This means the statistics become noisy and unhelpful for extracting information from the data. Previous attempts[1][2] to rectify this have focussed on frequency-domain interpolation, or trying to estimate what the missing values should be.

[1] Gallana *et al.* (2015), *J. Geophys Res.*
[2] Fraternale *et al.* (2019), *ApJ*

A NEW SOLUTION?

Instead of finding a way to "fill in" the missing values, and then compute the structure function from this imputed dataset, a more direct, data-driven approach was proposed to solve this problem. **Neural networks** are a method of supervised machine learning, in which a computer "learns" how to associate an input with its expected output, in a similar way to how neurons fire in the brain.

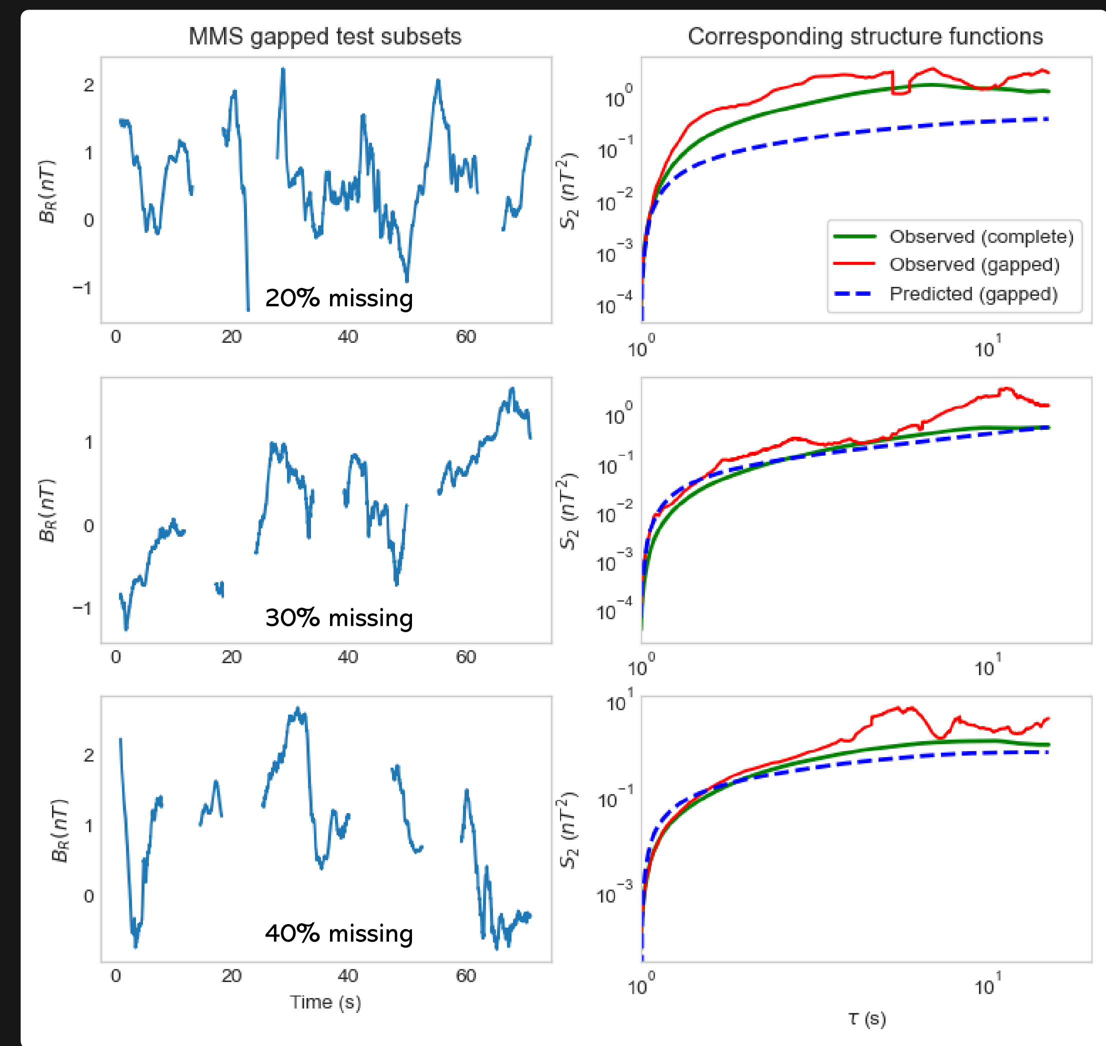


An artificial neural network.

In an attempt to model the workings of the human brain, this diagram shows the process whereby outputs from one layer of nodes become inputs to the next layer. The best weight parameters at each layer are iteratively selected in a learning technique known as "back-propagation".

METHOD

1. Became familiar with the data, statistics, and physics surrounding turbulence in space plasma.
2. Investigated potential methods of dealing with missing data, including traditional time series analysis methods such as ARIMA models.
3. Decided to pursue neural networks and studied multiple regression modelling and programming with Tensorflow.
4. Developed a machine learning pipeline to import and pre-process the data and finally train and test the model on a high-performance computing cluster.
5. Tweaked the model parameters in order to find the best balance between underfitting and overfitting.



FINAL MODEL

The Tensorflow program, written in Python, takes an input dataset of 10,000 measurements of a single field component of the solar wind, and produces an estimate for the second-order structure function. This model was trained on 5000 PSP subsets, and then tested on 500 sets it had never seen before.

RESULTS

Using 10 hidden layers, each with 50 nodes, we were able to produce a model that predicted better curves than the mathematical function for 80% of the gapped datasets. The model's performance on unseen data is shown in the plot above. In the left column are three magnetic field subsets from a new source, the MMS spacecraft, that have been artificially gapped, and on the right are the curves corresponding to each subset. In green is the mathematical structure function calculated from the original, ungapped datasets – this is what we are trying to predict. In red is the mathematical structure function from the gapped datasets – this is what we are trying to improve upon. Finally, in dashed blue is the prediction from our neural network. In the bottom two rows we can see the model is proving useful, as the blue line is closer to the green line than the red line is.

CONCLUSION

While this has shown the viability of using neural networks for estimating curves from missing data, the true test of the utility of this model will be how well it generalises to data from other spacecraft at different scales. Ideally, we would be able to apply it to Voyager datasets, which can have up to 70% of their data missing. Further research could involve extending such models to other statistics and other physical regimes, thereby greatly increasing our understanding of space weather in the face of incomplete data.

Research and poster by Daniel Wrench, under the supervision of Dr. Tulasi Parashar and Dr. Ramesh Rayudu
School of Chemical and Physical Sciences, Faculty of Science, Victoria University of Wellington