# **Machine Learning for Predicting Solar Radio Flux to Model Satellite Drag**



**A Project Report Submitted**

**By**

Deekshitha K P

Under the Supervision of

**Shri. Parthiban P**

Sci/Engr ‘SE’

Flight Dynamics Group

ISRO Bangalore

APRIL 2025 - MAY 2025

**FLIGHT DYNAMICS GROUP**

**U R RAO SATELLITE CENTRE**

**BENGALURU**

Session: **APRIL 2025 - MAY 2025**

**CERTIFICATE**

It is certified that this project report “**MACHINE LEARNING FOR PREDICTING SOLAR RADIO FLUX TO MODEL SATELLITE DRAG**” is the BONAFIDE work of “**DEEKSHITHA K P**” who carried out the project work under my supervision. The work reported herein does not form part of any other thesis or dissertation based on which a degree or award was conferred on an early occasion this or any other candidate.

**SIGNATURE**

**PARTHIBAN P**

Scientist/Engineer-SE

Flight Dynamics Group

U R Rao Satellite Centre

Bengaluru

# **ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to my supervisor and mentor, **Dr. P Parthiban** Sir, Sci/Engr-SE at U R Rao Satellite Centre, URSC-Bengaluru, for providing me with an invaluable opportunity to work on this project. His guidance, encouragement and expertise has been instrumental in shaping this work. I am thankful for his support and insights throughout this journey. His unwavering support and assistance has significantly contributed to the success of the project.

I extend my gratitude to **Dr. Kannan S**, Group Director, Flight Dynamics Group, for creating a conducive academic environment and providing ample resources for the development of this project.

I would also like to extend my gratitude to **URSC-ISRO** for providing essential resources and their support, which facilitated the smooth execution of this project.

Finally, I would like to thank my institution, **GRT Institute of Engineering and Technology** and my HOD, **Dr. Abdul Kareem D** sir, for their unwavering encouragement and understanding throughout this journey.

This project would not have been possible without the collective support, guidance and encouragement of all those mentioned above. Thank you for making this endeavour a fulfilling and enriching experience.

# **Table of Contents**

Title Page ………………………………………………………………1

Bonafide………………………………………………………………...2

Acknowledgement……………………………………………………...3

Table of Contents……………………………………………………….4

Table of Figures………………………………………………………...5

Abstract…………………………………………………………………6

1. Introduction……………………………………………………………..7
2. Forecasting using Statistical Modelling………………………………..12
3. Data…………………………………………………………………….15
   1. Data Acquisition………………………………………………...15
   2. Data Preparation………………………………………………...16
4. Machine Learning models……………………………………………..17
   1. Linear Regression……………………………………………….17
   2. SARIMA Model………………………………………………...20
   3. Long Short Term Memory(LSTM)...............................................21
   4. Artificial Neural Networks……………………………………...23
5. Failed Models………………………………………………………….25
   1. ARMA Model………………………………………………….26
   2. Decision Tree Regressor……………………………………….25
   3. XGBoost Model………………………………………………..27
6. Results and Evaluation………………………………………………..28
7. Conclusion…………………………………………………………….30
8. References……………………………………………………………..31

## **Table of Figures**

Figure 1: Scatter Plot of Observed values of Solar Radio Flux F10.7 cm……………8

Figure 2: Solar Activity from 1750s till 2025 using Monthly Sunspot Number…….10

Figure 3: Comparison Plot between NASA’s forecast and Observed values…………11

Figure 4: Monthly F10.7 and Sunspot number Forecast with McIntosh……………..12

Figure 5: 25th Cycle Prediction using McIntosh Model……………………………...14

Figure 6: Detailed Description of Dataset……………………………………………16

Figure 7: Test Train Split of SWC F10.7 Dataset…………………………………….17

Figure 8: Visualising Linear Regression Model for Solar Radio Flux Prediction……19

Figure 9: SARIMA Model Forecast………………………………………………….20

Figure 10: 25th Cycle Prediction using SARIMA and Comparison with NASA forecast………………………………………………………………………………21

Figure 11: LSTM Architecture………………………………………………………22

Figure 12: Visualising LSTM Model’s Predictions………………………………….22

Figure 13: Architecture of Artificial Neural Network……………………………….23

Figure 14: Visualising ANN Model Predictions……………………………………..24

Figure 15: Visualising Decision Tree Regressor Model Predictions…………………25

Figure 16: Visualising ARMA Model Predictions…………………………………...26

Figure 17: Visualising XGBoost Model Predictions…………………………………27

## **ABSTRACT**

In recent years (2000-2021), human-space activities have been increasing faster than ever. More than 36000 Earth-orbiting objects, all larger than 10 cm, in orbit around the Earth, are currently tracked by the European Space Agency (ESA). Around 70% of all catalogued objects are in Low-Earth Orbit (LEO). Aerodynamic drag provides one of the main sources of perturbations in this population, gradually decreasing the semi-major axis and period of the LEO satellites. Usually, an empirical atmosphere model as a function of solar radio flux and geomagnetic data is used to calculate the orbital decay and lifetimes of LEO satellites. In this respect, a good forecast for the space weather data could be a key tool to improve the model of drag. In this work, we propose using the Time Series Forecasting Model to predict the future behavior of the solar flux and to calculate the atmospheric density, to improve the analytical models and reduce the drag uncertainty. The effect of atmospheric drag on spacecraft dynamics is considered one of the predominant sources of uncertainty in [Low Earth Orbit](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/low-earth-orbit). These effects are characterised in part by the atmospheric density, a quantity highly correlated to space weather. Current atmosphere models typically account for this through proxy indices such as the F10.7, but with variations in solar radio flux forecasts leading to significant orbit differences over just a few days, the prediction of these quantities is a limiting factor in the accurate estimation of future drag conditions and, consequently, orbital prediction. In this study, we investigate the predictive performance of both traditional statistical methods and advanced machine learning models in forecasting space weather indices critical to drag modelling. Specifically, we employ the Seasonal Autoregressive Integrated Moving Average (SARIMA) model alongside a novel deep residual learning architecture designed for univariate time series prediction. Comparative analysis reveals that while SARIMA effectively captures seasonal patterns, the deep learning model exhibits superior performance in modelling non-linearities and abrupt variations in solar flux activity. We also employ Long Short-Term Memory (LSTM) networks and the ANN architecture, an advanced deep learning model designed specifically for time series forecasting. Our results indicate that while SARIMA performs well in capturing linear seasonal trends, LSTM and ANN models significantly outperform it in learning non-linear dynamics and sudden variations inherent in solar activity. Our results underscore the potential of hybrid modelling approaches to enhance the accuracy of atmospheric density forecasts and, by extension, improve the reliability of orbit propagation in the presence of space weather uncertainty.

## 

## **INTRODUCTION**

The increasing reliance on satellites for communication, navigation, Earth observation, and scientific exploration has heightened the importance of precise orbital prediction and satellite lifetime estimation, particularly for spacecraft in Low Earth Orbit (LEO). One of the dominant sources of uncertainty in LEO dynamics is atmospheric drag, a force that arises from the interaction between a spacecraft and the residual atmosphere. Unlike other perturbative forces, atmospheric drag is highly variable and depends on the density of the upper atmosphere, which in turn is influenced by solar activity and space weather phenomena.

Solar flares and related events, such as coronal mass ejections, can cause rapid and significant changes in atmospheric density. These fluctuations pose challenges in predicting satellite trajectories, potentially leading to increased fuel usage for orbit maintenance, shortened satellite lifespans, or in worst cases, loss of mission. Current atmospheric models attempt to account for these effects through empirical correlations with solar activity indices such as the F10.7 solar radio flux. However, limitations in the accuracy of solar flux forecasts hinder long-term orbit prediction, creating a critical need for more robust forecasting approaches.

Traditionally, the estimation of solar flux has relied on statistical extrapolations and empirical techniques, which are often limited in capturing abrupt changes or non-linear dynamics inherent in space weather. In recent years, advances in data-driven methods and time series forecasting have enabled new opportunities for modelling complex systems like solar activity. Statistical models like SARIMA (Seasonal Autoregressive Integrated Moving Average) offer a strong foundation for capturing seasonal trends and linear dependencies in solar flux data. However, they often fall short when faced with chaotic patterns or long-term dependencies.

To overcome these limitations, machine learning and deep learning models have been introduced for time series forecasting with encouraging results. Models such as Long Short-Term Memory (LSTM) networks are well-suited to capture temporal dependencies and have been widely adopted for various forecasting tasks. Similarly, the ANN has demonstrated state-of-the-art performance across a range of applications. By integrating these models into the solar flux prediction pipeline, we aim to reduce uncertainty in atmospheric drag estimates and improve satellite lifetime assessments.

This project explores and compares a variety of forecasting methods—from classical linear regression and SARIMA to advanced neural network architectures—on their ability to model solar activity indices relevant to spacecraft drag. The methodology includes thorough data acquisition and preprocessing, model training and evaluation, and the quantification of their impact on satellite lifetime prediction models. Additionally, the project discusses failed modelling approaches to provide insight into the limitations of certain techniques in this context.

Ultimately, this work aims to contribute to the field of space situational awareness by enhancing the fidelity of drag modelling through improved solar flux forecasting. The outcomes have the potential to support mission planning, reduce operational risks, and inform the development of more resilient satellite systems in the face of variable space weather.

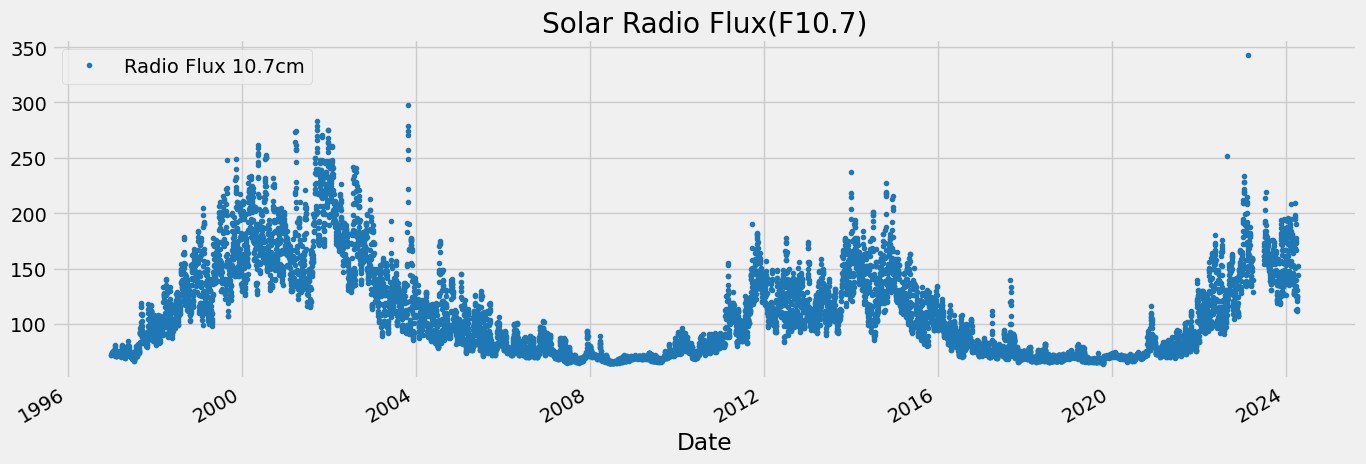


Figure 1: Scatter Plot of Observed Values of Solar Radio Flux F10.7 cm

### **LEO (Low Earth Orbit)**

Low Earth Orbit (LEO) typically refers to altitudes ranging from approximately 160 km to 2,000 km above the Earth's surface. This orbital region is home to the majority of artificial satellites, including Earth observation platforms, scientific missions, CubeSats, and a large fraction of the satellite communication infrastructure. The advantages of LEO include low latency for communications, high-resolution imaging, and reduced launch costs due to the lower energy required to reach orbit.

However, LEO is also subject to the greatest influence from Earth's atmosphere, particularly in the form of atmospheric drag. While atmospheric density at these altitudes is significantly lower than at the surface, it is still sufficient to produce measurable deceleration on orbiting objects. Over time, this drag causes satellites to lose altitude, ultimately requiring frequent orbit-raising manoeuvres or leading to uncontrolled re-entry. As such, understanding and modelling the environmental conditions in LEO, especially space weather-driven density changes, is essential for mission planning, collision avoidance, and debris management.

### **Effects of Drag in Spacecraft**

Atmospheric drag affects spacecraft in LEO by reducing their velocity and causing gradual orbital decay. The magnitude of the drag force Fd can be expressed as:

Fd​=0.5⋅Cd​⋅A⋅ρ⋅v2

where Cd​ is the drag coefficient, A is the cross-sectional area, ρ is the atmospheric density, and v is the relative velocity of the spacecraft with respect to the atmosphere. This relationship highlights that drag is sensitive not only to the spacecraft’s properties but also to the density of the surrounding atmosphere, making accurate density forecasting critically important.

Drag-induced effects include:

* **Orbital decay**: Progressive reduction in orbital altitude.
* **Increased fuel consumption**: For drag compensation using onboard propulsion systems.
* **Decreased mission lifespan**: Especially for small satellites and CubeSats with limited propulsion.
* **Tracking uncertainties**: Higher drag variability makes satellite tracking and conjunction analysis more error-prone.

### **Methods of Drag Estimation**

There are two broad categories of drag estimation: **empirical models** and **data-driven models**.

1. **Empirical Models**: These include the Jacchia models, NRLMSISE-00, and JB2008, which estimate atmospheric density based on solar and geomagnetic indices. While computationally efficient, they lack responsiveness to sudden space weather events.
2. **Data-Driven Models**: These utilise historical measurements of solar activity and atmospheric response to learn predictive patterns. They can include time series models (like SARIMA), machine learning approaches (like LSTM), or hybrid models. These models are better suited to capture nonlinear dynamics and abrupt changes in space weather parameters.

### **Atmospheric Effects of Drag**

The Earth's thermosphere responds dynamically to solar radiation, especially in the extreme ultraviolet (EUV) and X-ray bands. This radiation heats the atmosphere, causing it to expand and increase the density at satellite altitudes. The primary drivers of these effects are:

* Solar EUV and X-ray emissions
* Geomagnetic storms
* Thermospheric tides and winds

During periods of high solar activity, the atmosphere becomes denser, which amplifies the drag force experienced by satellites. Conversely, during solar minima, drag effects are milder. This variability underscores the need for accurate, real-time modelling of solar activity to forecast drag.

### **Effects of Satellite Drag using Solar Activity**

Since solar activity directly influences thermospheric density, the variability in solar indices like F10.7 and Kp leads to corresponding variability in atmospheric drag. Satellite operators use these indices as input to drag models to estimate orbit degradation rates.

Key impacts include:

* **Orbit prediction errors**: Deviations between predicted and actual positions increase with poor solar flux forecasts.
* **Increased station-keeping manoeuvres**: Leading to faster fuel depletion.
* **Lifetime predictions**: Highly sensitive to errors in density forecasts; a few per cent error in density can lead to major differences in re-entry time estimates.

Accurate modelling of solar activity is thus essential for both short-term navigation and long-term satellite mission design.

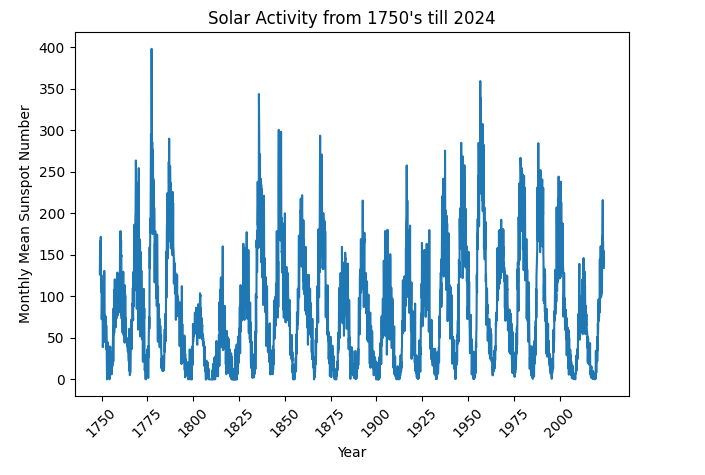


Figure 2: Solar Activity from 1750s to 2025 using Monthly mean Sunspot Number

### **Solar Flares Predicted vs Observed**

There remains a significant gap between observed solar flares and what can be accurately predicted. Current methods rely on sunspot observations, solar magnetogram data, and probabilistic models. However, the occurrence, timing, and magnitude of solar flares remain inherently unpredictable due to the complex dynamics of the Sun’s magnetic field.

This unpredictability translates into:

* Lag in atmospheric model updates after flare events.
* Mismatches between predicted and actual atmospheric conditions.
* Challenges in real-time drag compensation.

To mitigate these limitations, data-driven time series forecasting methods such as SARIMA, LSTM, and ANN can help learn from historical patterns and provide probabilistic forecasts of solar indices, particularly F10.7, which are more responsive than flare-specific forecasts.

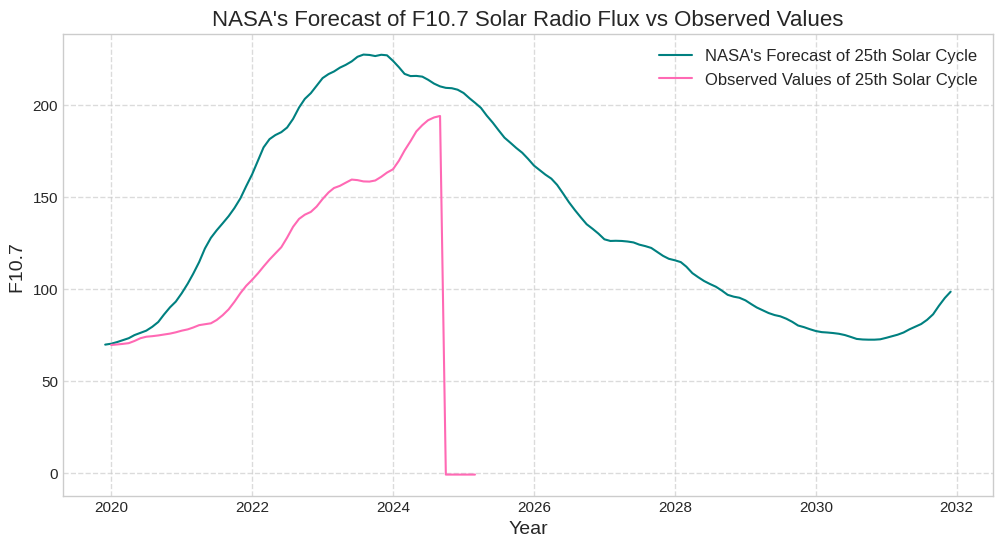


Figure 3: Comparison Plot between NASA’s forecast and Observed values

## 

## **FORECASTING USING STATISTICAL MODELLING**

Given the challenges associated with forecasting solar activity and its impact on atmospheric drag, we adopted a physically motivated, model-based statistical approach to predict the Solar Flux at 10.7 cm (SFU). This approach is grounded in empirical relationships derived from solar activity indices, specifically the **sunspot number (SSN)**, which has been widely used as a proxy for solar activity. The sunspot number is directly related to the solar radiation output, including the solar flux in the 10.7 cm wavelength band, and serves as a key input to atmospheric density models.

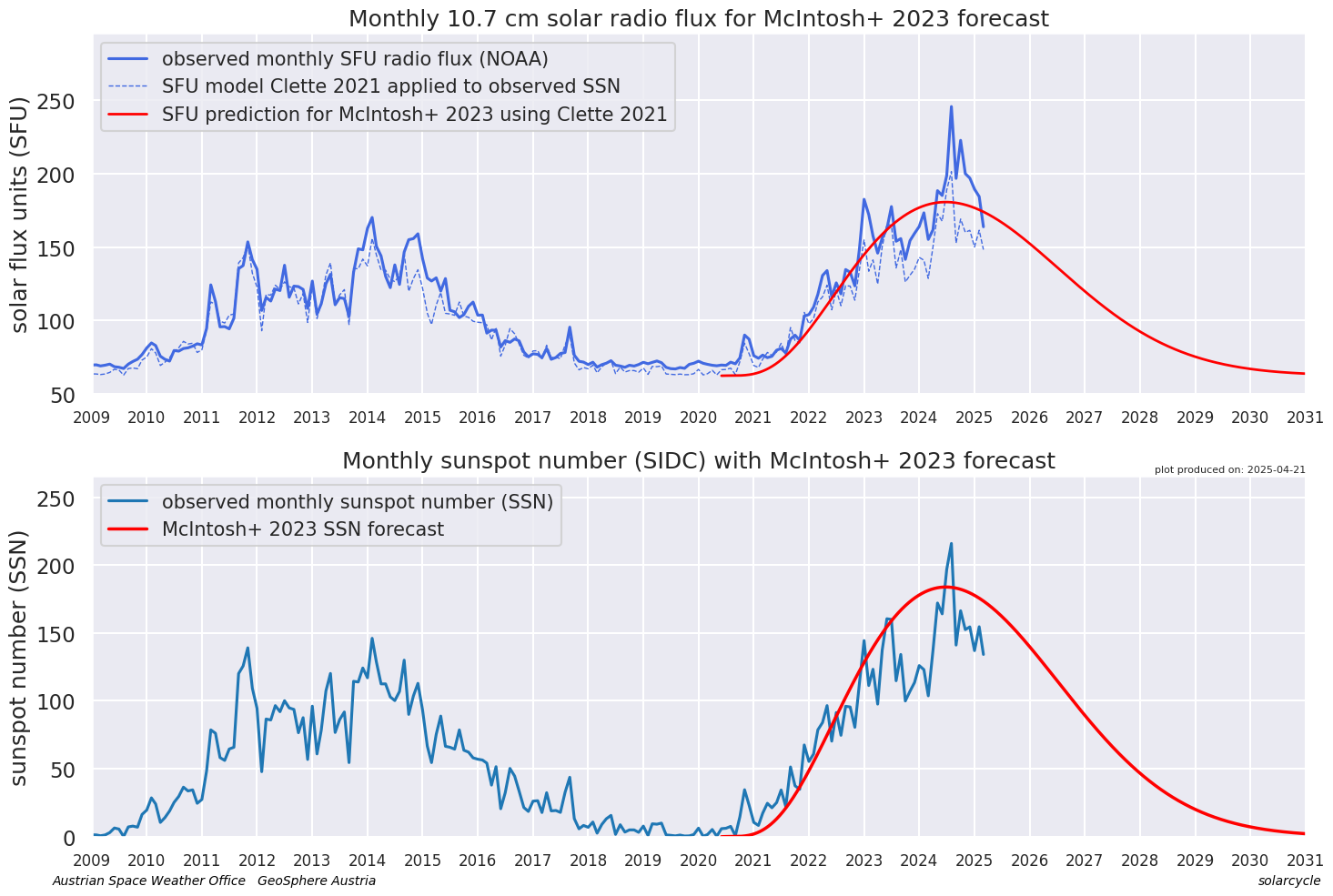


Figure 4: Monthly F10.7 and Sunspot number Forecast with McIntosh

### **Clette Model (2021)**

The **Clette 2021 model** establishes a second-order polynomial relationship between the sunspot number (SSN) and solar flux (SFU). This relationship is based on historical data and provides an improved framework for forecasting SFU values during solar cycles. The model equation is:

**SFU = 62.87 + 0.6279 × SSN + 6.141 × 10⁻⁵ × SSN²**

This second-order polynomial model takes into account both linear and non-linear effects in the relationship between SSN and SFU, providing a more nuanced forecast than simple linear models. The inclusion of a quadratic term accounts for the fact that solar flux does not increase proportionally with sunspot number, especially during the peak of the solar cycle when the intensity of solar activity is more complex.

### **Tiwari-Kumar Model (2018)**

Another model that we referenced is the **Tiwari and Kumar model (2018)**, which proposes a simplified polynomial approach to model the relationship between SSN and SFU. This model is useful for generating quick approximations of solar flux, especially when high precision is not required. The simplified equation from Tiwari and Kumar is:

**SFU = 62.51 + 0.6422 × SSN**

Although this model lacks the complexity of the second-order polynomial in the Clette model, it is computationally efficient and can be valuable for initial estimates of solar flux. It provides a linear approximation of SFU based on SSN, which is especially useful when solar flux data is missing or for operational environments that require fast predictions.

### **Hathaway Function (2015) for Solar Cycle Forecasting**

To predict future values of the sunspot number (SSN) and, consequently, solar flux, we employed the **Hathaway function (2015)**. This model provides a forecast of SSN for future solar cycles, specifically designed for Solar Cycle 25, which is currently ongoing. Hathaway’s model is based on the McIntosh+ 2023 parameters and uses a physically-based function to predict the peak and timing of the solar cycle, considering the overall solar activity level.

The Hathaway function has proven to be highly effective in forecasting solar activity patterns, allowing us to anticipate future SSN values that are essential for predicting SFU and the subsequent atmospheric drag experienced by satellites in LEO. This model helps in forecasting solar flux during both high- and low-activity phases of the solar cycle, aiding in long-term satellite mission planning.

### **Integration of Models for Solar Flux Prediction**

In our approach, we integrate the **Clette 2021 model** as the primary framework for predicting SFU using SSN. Given the complex and non-linear nature of solar activity, the second-order polynomial provides a more accurate representation of the relationship between SSN and SFU. Additionally, the Hathaway function is utilised to predict future SSN values based on the current phase of Solar Cycle 25. By combining these models, we can create a robust forecast of solar flux, accounting for both historical data and future solar activity.

The decision to use these models stems from their well-established theoretical foundations, their ability to capture both linear and non-linear dependencies in solar activity, and their demonstrated success in previous studies. Additionally, using a model-based approach with physical grounding allows us to better understand the underlying mechanisms driving solar flux and atmospheric drag, which is critical for improving the reliability of satellite drag models and predicting satellite lifetimes in LEO.

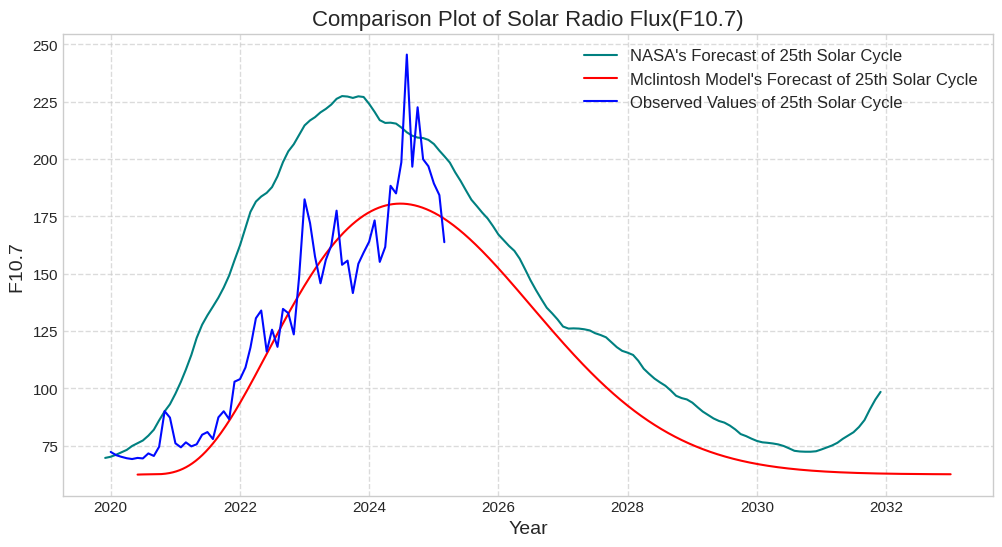


Figure 5: 25th Cycle Prediction using McIntosh Model and Comparison with NASA forecast.

## 

## 

## **DATA**

The database available in Space Weather Canada website comprises two components: measurements of the 10.7cm Flux and daily records of flux monitor output. Each measurement of the 10.7cm Solar Flux is expressed in three values: the *observed*, *adjusted* and *URSI Series D* values.

The *observed* value is the number measured by the solar radio telescope. This is modulated by two quantities: the level of solar activity and the changing distance between the Earth and the Sun. Since it is a measure of the emissions due to solar activity hitting the Earth, this is the quantity to use when terrestrial phenomena are being studied.

When the Sun is being studied, the annual modulation of the 10.7cm Solar Flux by the changing distance between the Earth and the Sun is undesirable. However, one byproduct of the ephemeris calculations needed for the solar flux monitors to properly acquire and track the Sun is the distance between the Sun and the Earth. We therefore produce an additional quantity, corrected for variations in the Earth-Sun distance, and given for the average distance. This is called the *adjusted* value.

The record for each measurement of the 10.7cm Solar Flux is as below. The quantities are separated by commas. The 10.7cm Solar Flux is given in solar flux units (an sfu = 10-22W m-2 Hz-1). For building our models we only use the observed values of past 2 solar cycles from the year 2005 till 2024.

### **Data Acquisition**

The data used in this project was obtained from **Space Weather Canada**, a reliable and authoritative source of space weather and solar activity data. The dataset includes the following key features:

* **Solar Flux (F10.7)**: The solar radio flux at 10.7 cm which is an important proxy for solar activity and plays a significant role in modelling atmospheric drag on satellites.
* **Sunspot Number (SSN)**: A measure of solar activity, representing the number of sunspots on the solar surface. Sunspot numbers are closely correlated with solar flux, and understanding this relationship is central to solar flux forecasting.
* **Geomagnetic Indices (Ap, Kp)**: These indices quantify the geomagnetic activity, which is important in understanding the response of the Earth’s atmosphere to solar wind disturbances. Geomagnetic activity can also affect atmospheric density, thus influencing satellite drag.
* **Solar Flares**: Information about solar flare events, which contribute to short-term variations in solar radiation and atmospheric drag.

The dataset spans multiple solar cycles, providing a comprehensive historical record, as well as real-time data to allow for model testing and validation. This data enables the creation of a robust forecasting framework to predict solar flux and its impact on atmospheric conditions.

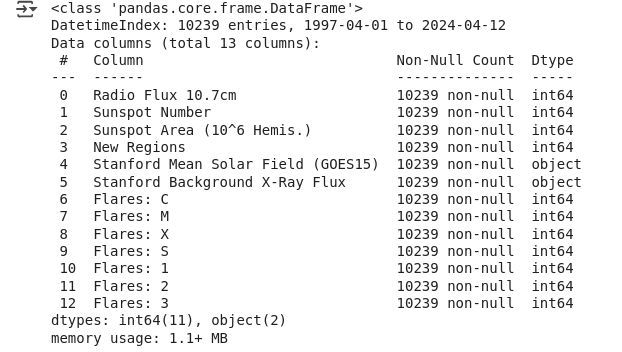


Figure 6: Detailed Description of the Dataset

### **Data Preparation**

Data preparation is a critical step in ensuring the quality and usability of the dataset for machine learning and statistical modelling. The following steps were applied to prepare the data for training and evaluation:

1. **Handling Missing Data**: Some of the solar flux and sunspot number data contained missing values due to gaps in the measurement records. To handle this, missing values were imputed using interpolation methods or forward-filling for time series data to ensure continuity in the dataset.
2. **Normalisation**: Since machine learning models often perform better when the data is normalised, the input features (such as SSN and geomagnetic indices) were scaled to a standard range (typically 0 to 1) using Min-Max normalisation or standardisation (zero mean and unit variance).
3. **Lag Feature Creation**: Time series forecasting models such as SARIMA, LSTM, and ANN require lagged values as input features. We created lag features from the solar flux and sunspot number data to allow the models to learn temporal dependencies in the dataset.
4. **Seasonality Encoding**: Solar flux data exhibits periodic patterns, especially due to the solar cycle. We added seasonality features, such as the day of the year, to capture cyclical trends in solar activity.
5. **Train-Test Split**: The dataset was divided into training (80%) and testing (20%) sets. A validation set was also used for hyperparameter tuning and model selection. The training set was used to build and tune the models, while the testing set was reserved for final evaluation.

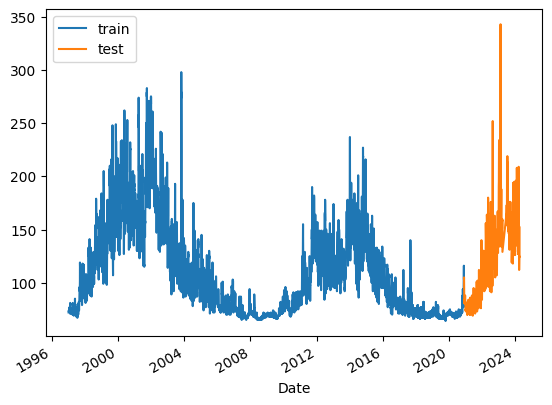


Figure 7: Test Train Split of SWC F10.7 Dataset

### 

### 

### 

### 

### 

### **MACHINE LEARNING MODELS**

Machine learning models have gained significant attention in space weather forecasting due to their ability to capture complex, non-linear relationships in large datasets. In this project, we focus on comparing several machine learning and statistical models to predict solar flux (F10.7) based on solar activity indices. These models include both traditional machine learning techniques and more advanced deep learning models. The models evaluated in this study are:

* **Linear Regression (LR)**: A baseline model for understanding the relationship between the input features (such as sunspot number, solar indices) and the target variable (solar flux).
* **SARIMA (Seasonal Autoregressive Integrated Moving Average)**: A statistical time series forecasting model used to account for both seasonal and trend components in solar flux data.
* **LSTM (Long Short-Term Memory Networks)**: A deep learning model designed to capture long-term dependencies in sequential data, especially useful for time series prediction tasks.
* **ANN (Artificial Neural Networks)**: Feedforward neural networks that can learn non-linear patterns in data through multiple hidden layers, enabling better prediction performance compared to linear models.

These models are tested and compared based on their ability to predict solar flux, which in turn affects atmospheric drag estimates for satellites in Low Earth Orbit (LEO).

## **Models**

### **Linear Regression (LR)**

**Linear Regression (LR)** is one of the simplest machine learning models used to predict a dependent variable based on one or more independent variables. In this project, Linear Regression was used to establish a baseline model for solar flux (F10.7) prediction. The model assumes a linear relationship between the independent features (such as Sunspot Number, SSN) and the dependent variable (solar flux).

#### **Approach:**

In the context of solar flux forecasting, Linear Regression models the solar flux as a weighted sum of lagged values of sunspot numbers and solar flux. The model attempts to fit the data to a linear equation where the coefficients (weights) are learned during training. The equation for the Linear Regression model is generally represented as:

SFUt ​ = β0​ + β1 ​× SSNt ​+ β2 ​× SSNt − 1​ +⋯+ ϵ

Where:

* **SFU** is the solar flux at time **t**,
* **SSN** is the Sunspot Number at time **t**,
* **β₀, β₁, β₂** are the model’s coefficients (parameters) to be learned,
* **ε** is the error term (residuals), representing the unexplained variance in the data.

The model uses **Ordinary Least Squares (OLS)** to minimise the sum of squared residuals, learning the relationship between SSN and solar flux. The coefficients that minimise this error function are used to make predictions for future solar flux values.

#### 

Figure 8: Visualising Linear Regression Model for Solar Radio Flux Prediction

In this project, Linear Regression serves as the baseline model to understand the fundamental relationship between sunspot numbers and solar flux, before transitioning to more complex models. It’s particularly useful in providing a simple and interpretable framework for predicting solar flux under steady solar conditions.

### **2. SARIMA (Seasonal ARIMA)**

**SARIMA** (Seasonal Autoregressive Integrated Moving Average) is a time series model that is an extension of ARIMA (Autoregressive Integrated Moving Average). It is specifically designed to handle time series data with seasonal patterns, such as solar flux, which exhibits periodic fluctuations due to the solar cycle.

#### **Approach:**

The SARIMA model is composed of both seasonal and non-seasonal components:

* **AR (Autoregressive)**: This term represents the relationship between the current value and previous time-step values. It captures the temporal dependencies between solar flux values over time.
* **I (Integrated)**: This term accounts for making the time series stationary by differencing the data, ensuring that trends and seasonalities are removed.
* **MA (Moving Average)**: This captures the influence of random shocks on the solar flux data by modelling the residuals of previous time steps.
* **Seasonal Components**: These components model the seasonal variation that occurs in solar flux, such as the recurring 11-year solar cycle and other periodic behaviours in solar activity.

The SARIMA model is mathematically represented as:

SFUt​ = ϕ1​ ⋅ SFUt − 1​ + θ1​⋅ϵt − 1 ​+⋯+ ϵt​

Where **ϕ** and **θ** represent the seasonal autoregressive and moving average parameters, and **ε** is the error term.

#### 

Figure 9: SARIMA Model Forecast

SARIMA is well-suited for solar flux forecasting as it can model the periodic nature of solar activity, capturing both short-term fluctuations and longer-term trends. The seasonal components of SARIMA allow it to predict solar flux across the solar cycle, which is crucial for estimating atmospheric drag variations that affect satellite trajectories in Low Earth Orbit (LEO).

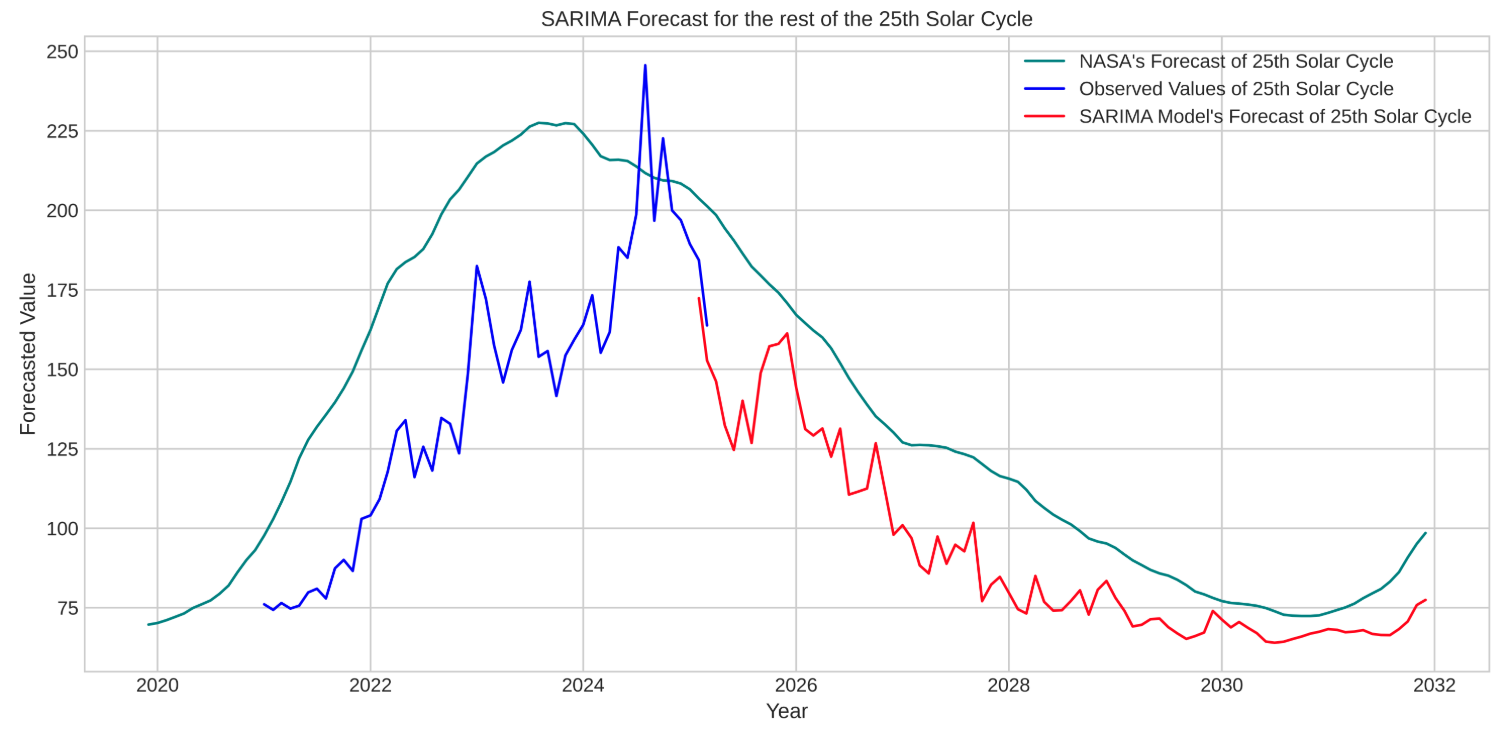


Figure 10: 25th Cycle Prediction using SARIMA and Comparison with NASA forecast.

### **3. LSTM (Long Short-Term Memory Networks)**

**LSTM** (Long Short-Term Memory) is a type of Recurrent Neural Network (RNN) designed to learn from sequential data. Unlike standard RNNs, LSTMs can capture long-term dependencies by maintaining memory of past inputs, which makes them ideal for time series forecasting tasks where past values influence future predictions.

#### **Approach:**

LSTM networks are composed of memory cells that can store information over long periods, making them particularly effective in handling time series data with complex temporal relationships, such as solar flux. The LSTM model works by passing the input sequence through multiple layers of cells that decide whether to store, update, or forget information from previous time steps.

The basic architecture of the LSTM includes:

1. **Input Layer**: The time series data (e.g., lagged solar flux) is fed as input to the model.
2. **LSTM Layers**: These layers learn temporal patterns by retaining the relevant historical information from previous time steps. LSTM units include gates like input, forget, and output gates, which regulate the flow of information.
3. **Dense Layer**: A fully connected layer at the output that generates the predicted solar flux for the next time step.
4. **Activation Function**: Usually, **ReLU** (Rectified Linear Unit) is used in the hidden layers to introduce non-linearity, while the output layer typically uses a linear activation function for regression tasks.

## **Architecture**

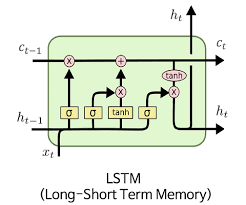


Figure 11: LSTM Architecture

LSTM is ideal for solar flux prediction because solar activity shows complex temporal dependencies, including long-term cycles (like the solar cycle) and short-term fluctuations (like solar flares). LSTM can effectively capture these dependencies and provide more accurate predictions of solar flux over time, which is critical for estimating atmospheric drag and satellite orbit changes.

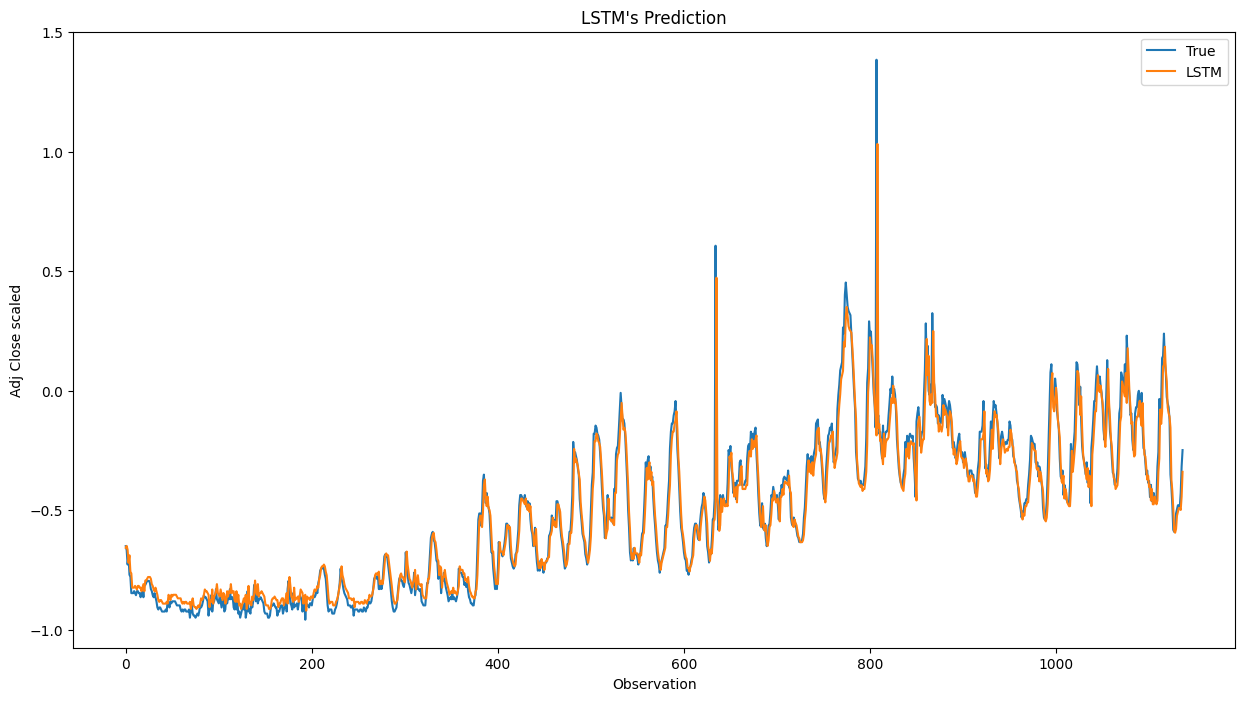


Figure 12: Visualising LSTM Model’s Predictions

### **4. ANN (Artificial Neural Networks)**

**ANN** (Artificial Neural Networks) are a class of machine learning models that consist of multiple layers of interconnected nodes or neurons. These networks are designed to learn complex, non-linear relationships in data through iterative training processes.

#### **Approach:**

The typical **Feedforward Neural Network (FNN)** architecture consists of:

1. **Input Layer**: The input layer receives time series data, such as sunspot number (SSN), lagged solar flux values, and other relevant features.
2. **Hidden Layers**: These layers contain neurons that apply weights and activation functions to learn complex patterns in the data. A common activation function is **ReLU** (Rectified Linear Unit), which helps introduce non-linearity.
3. **Output Layer**: The output layer produces a single value representing the predicted solar flux for the next time step. This layer uses a linear activation function since solar flux is a continuous variable.
4. **Backpropagation**: During training, the network adjusts its weights using **backpropagation**, minimizing the error between predicted and actual values via gradient descent.

## **Architecture**

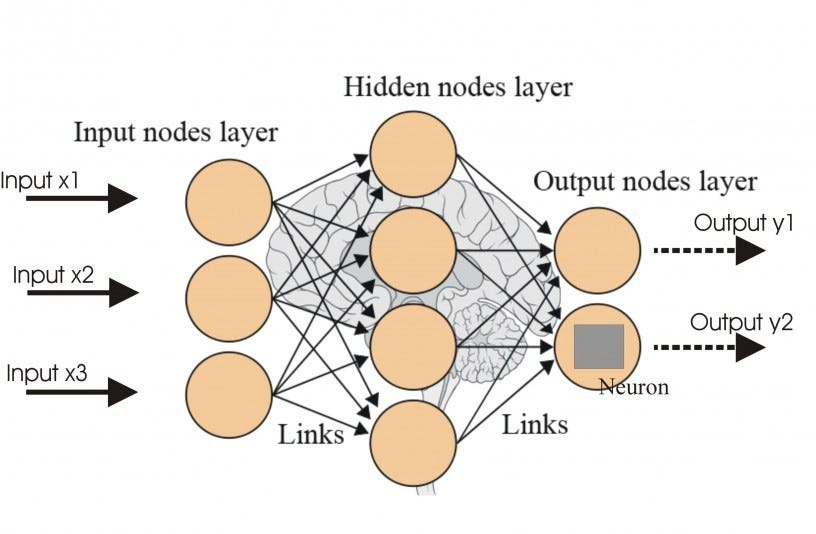


Figure 13: Architecture of an Artificial Neural Network

ANNs are particularly useful in capturing non-linear relationships that exist in solar flux data. Unlike Linear Regression, which assumes a linear relationship, ANNs can model the complex interactions between solar activity variables. However, unlike LSTM, ANNs are not inherently designed to capture long-term dependencies, which is why their performance may be slightly lower than LSTM for long-range forecasting.

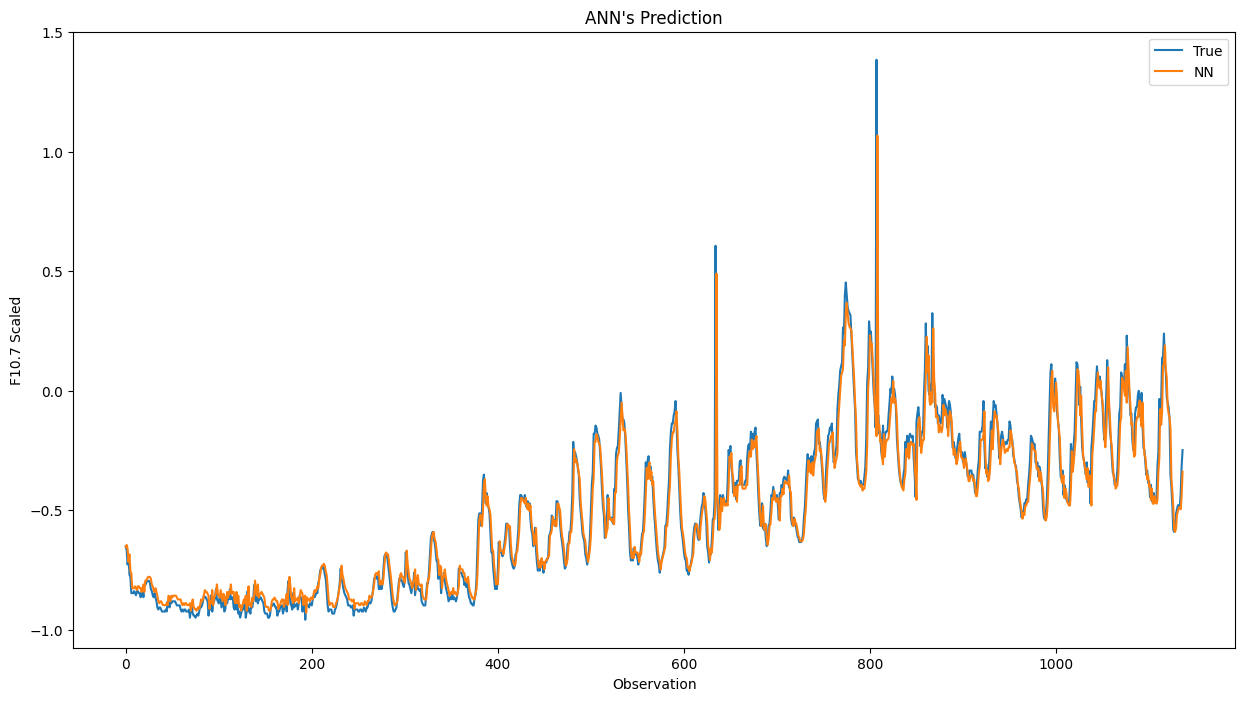


Figure 14: Visualizing ANN Model Predictions

## 

## 

## 

## 

## 

## **FAILED MODELS**

While we tested various models for predicting **F10.7 solar flux**, including **Decision Tree Regressor**, **ARMA**, and **XGBoost**, these models failed to deliver satisfactory results. Here’s a summary of why they underperformed:

### **1. Decision Tree Regressor (DTR)**

#### **Reasons for Failure:**

* **Overfitting**: Decision Trees can easily overfit, especially with complex and fluctuating data like solar flux. The model memorised the training data, leading to poor generalisation on the test set.
* **No Temporal Memory**: Decision Trees don’t capture sequential dependencies, which are crucial for time series forecasting. Without incorporating lagged values or time-based features, the model struggled to predict future values effectively.
* **Instability**: Small changes in the data caused large variations in the tree structure, resulting in unstable and inconsistent predictions.

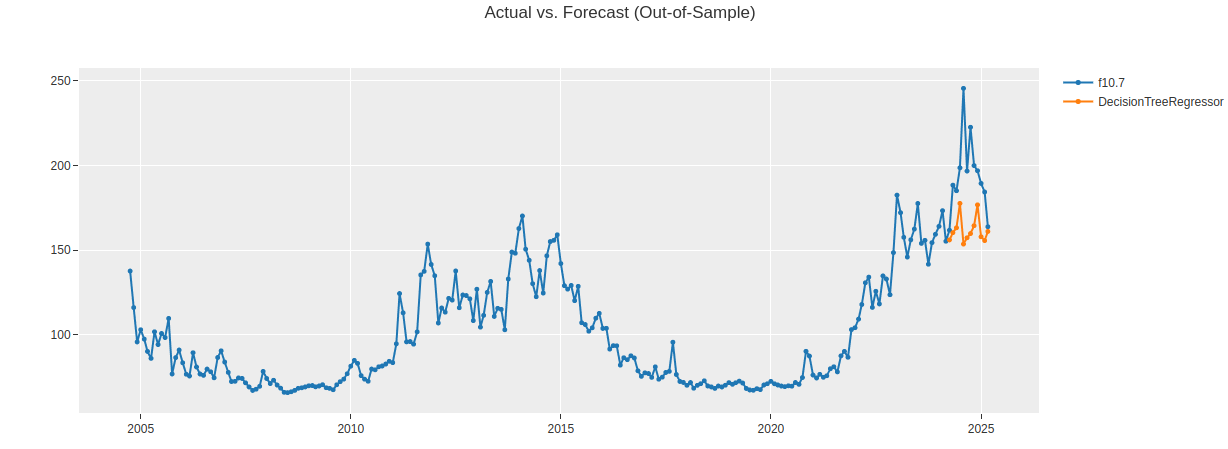


Figure 15: Visualising Decision Tree Regressor Model Predictions

#### **Conclusion:**

The model's inability to handle sequential data and its tendency to overfit made it unsuitable for this time series task.

### **2. ARMA (Autoregressive Moving Average)**

#### **Reasons for Failure:**

* **Stationarity Assumption**: ARMA models require stationary data, but solar flux exhibits non-stationary behaviour due to long-term trends like the 11-year solar cycle.
* **Limited Short-Term Dependencies**: ARMA is designed for short-term dependencies, whereas solar flux shows long-term cyclical patterns that ARMA could not capture effectively.

**Conclusion:**

* ARMA’s inability to model long-term trends and its reliance on stationarity made it unsuitable for predicting solar flux. The seasonal long-term dependency cannot be captured in the ARMA model.

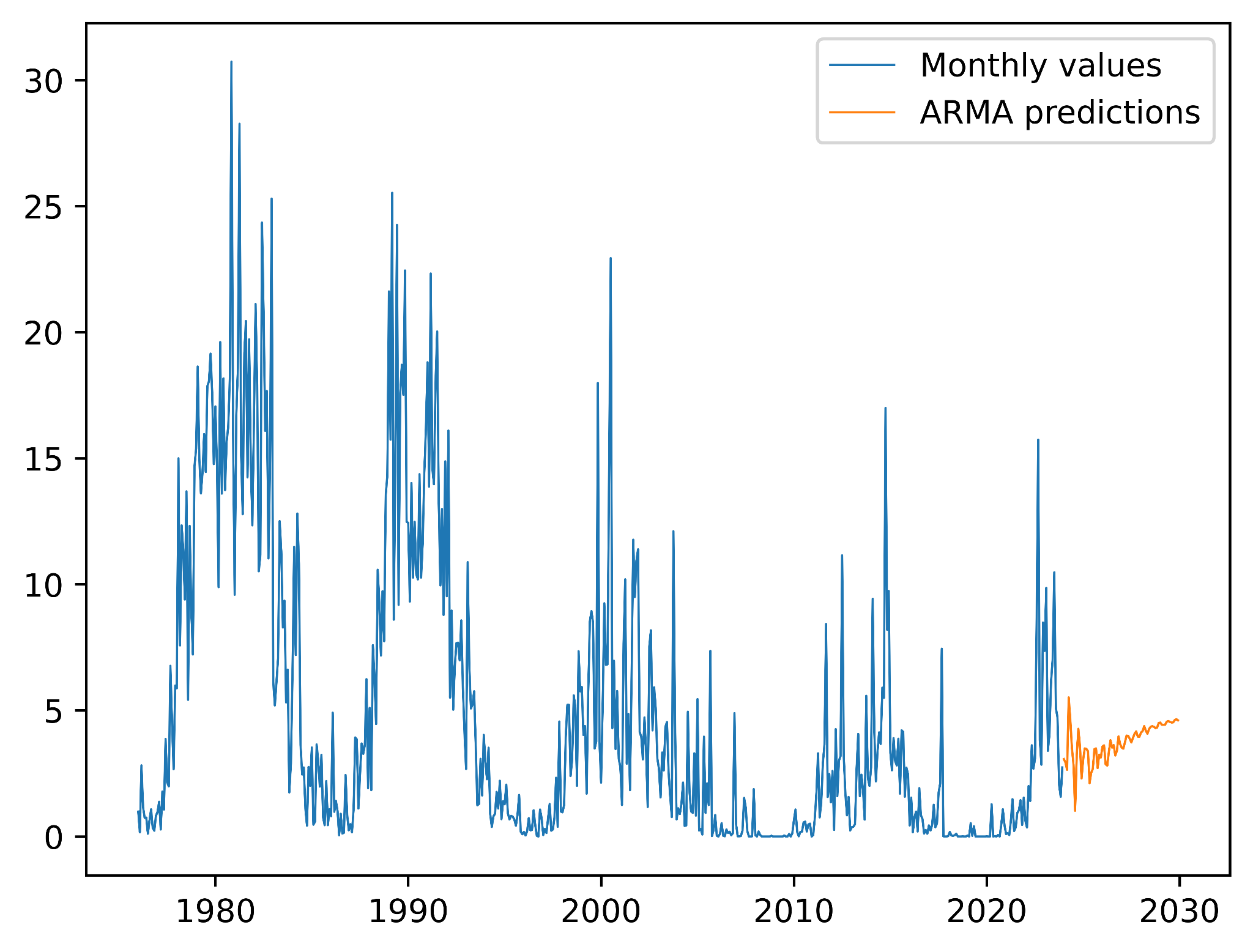


Figure 16: Visualising ARMA Model Predictions

### **3. XGBoost (Extreme Gradient Boosting)**

#### **Reasons for Failure:**

* **Overfitting**: XGBoost overfitted to the training data, especially with default hyperparameters, leading to poor performance on test data.
* **Lack of Temporal Feature Engineering**: Without explicitly engineered time-dependent features (e.g., lagged values), XGBoost struggled to learn the temporal dependencies of solar flux.
* **Inability to Capture Cyclical Patterns**: Solar flux data is influenced by periodic solar cycles, and XGBoost’s default structure couldn’t capture these long-term cyclical patterns effectively.

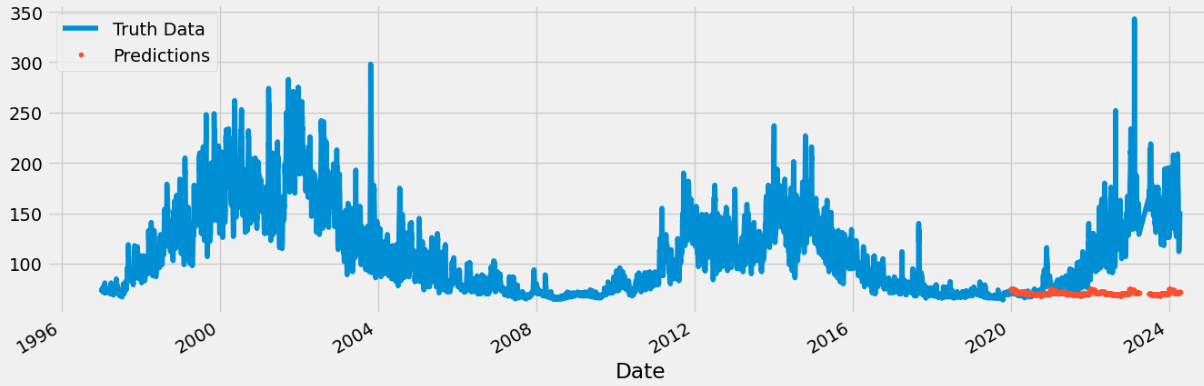


Figure 17: Visualising XGBoost Model Predictions

#### **Conclusion:**

XGBoost's lack of feature engineering and its inability to handle periodic cycles led to poor performance in forecasting solar flux. Though it performs well in single step forecasting when it comes to univariate multi-step periodic forecasting it exhibited poor performance.

### **RESULTS SUMMARY**

In this project, several models were evaluated to forecast **F10.7 solar flux**, which is crucial for predicting atmospheric drag on spacecraft in **Low Earth Orbit (LEO)**. Below is a summary of the performance of each model:

### **1. Linear Regression**

**Linear Regression** was used as a baseline model, offering decent performance in capturing the overall trend in solar flux. However, it struggled to capture the **non-linear** and **time-dependent** relationships in the data.

* **Strength**: Quick to train, simple, and interpretable, providing a general trend of solar flux.
* **Weakness**: It could not model the **non-linear patterns** or **complex dynamics** in solar activity, particularly during rapid events like solar flares.

**Conclusion**: Linear Regression was effective for baseline performance but inadequate for accurately predicting solar flux due to its linear assumptions.

### **2. SARIMA (Seasonal ARIMA)**

**SARIMA** performed well for modelling **seasonal** and **cyclical** trends, particularly capturing the 11-year solar cycle. However, it struggled with **abrupt solar events** like solar flares.

* **Strength**: Excellent for capturing long-term **seasonal** and **cyclical patterns** of solar activity.
* **Weakness**: Struggled to forecast rapid shifts in solar flux due to solar flares, as it primarily relies on past data and assumes smooth, predictable trends.

**Conclusion**: SARIMA excelled in long-term forecasting with periodic patterns but was limited in handling sudden changes in solar activity.

### **3. Artificial Neural Networks (ANN)**

**ANNs** demonstrated strong performance by capturing **non-linear relationships** and complex patterns in solar flux data. While it performed better than simpler models like Linear Regression, it still faced some challenges.

* **Strength**: Good at learning **non-linear** dynamics and complex patterns in the data, providing a more flexible approach than traditional models.
* **Weakness**: **ANNs** still require substantial preprocessing, tuning, and sufficient data for training, and they can be sensitive to overfitting, especially with smaller datasets.

**Conclusion**: ANN performed better than Linear Regression and SARIMA, effectively capturing non-linear trends, but required careful tuning and data preparation for optimal results.

### **4. LSTM (Long Short-Term Memory)**

**LSTM** models, a type of deep learning architecture, showed the best performance by effectively capturing **long-term dependencies** and **non-linear relationships** in solar flux data.

* **Strength**: Excellent at modelling both **long-term trends** and **abrupt changes** like solar flares, thanks to its ability to retain sequential information and handle complex patterns in time series data.
* **Weakness**: Computationally expensive to train and requires significant data preprocessing and hyperparameter tuning.

**Conclusion**: LSTM was the most effective model overall, offering the best prediction accuracy by handling both long-term trends and short-term fluctuations in solar flux.

## **CONCLUSION**

This project explored various models for forecasting the **F10.7 solar flux**, a key component in predicting atmospheric drag on spacecraft in **Low Earth Orbit (LEO)**. We tested multiple approaches, including **Linear Regression**, **SARIMA**, **LSTM**, **ANN**, and others like **Decision Tree Regressor**, **ARMA**, and **XGBoost**. While some models showed promise, others failed to perform effectively due to limitations in capturing the complex, cyclical, and time-dependent nature of solar flux data.

The **SARIMA** model, which incorporates seasonality and temporal dependencies, demonstrated good performance for this problem, leveraging its ability to model periodic solar cycles. Similarly, **LSTM** and **ANN** models, both deep learning approaches, performed well after appropriate preprocessing and scaling, capturing the intricate patterns of solar activity.

On the other hand, models like **Decision Tree Regressor**, **ARMA**, and **XGBoost** struggled due to overfitting, failure to account for long-term dependencies, and lack of temporal feature engineering. These models could not effectively model the cyclical nature of solar flux, highlighting the importance of selecting the right model for time series forecasting tasks.

In summary, the findings of this project suggest that **LSTM** and **SARIMA** are the most effective models for forecasting **F10.7 solar flux**, while traditional machine learning models such as **Decision Tree Regressor** and **XGBoost**, and classical statistical models like **ARMA**, are less suited for this task due to their inability to handle the complex, non-linear, and cyclical behavior inherent in solar activity data.

This work lays the foundation for future improvements in orbital prediction models, which can be enhanced by incorporating additional features and refining model architectures. Future work may also explore the use of more sophisticated deep learning models, such as **transformers** or **attention-based mechanisms**, to further improve the accuracy and reliability of solar flux predictions.

## **REFERENCES**

* **Clette 2021**: [Link to paper](https://www.swsc-journal.org/articles/swsc/pdf/2021/01/swsc200021.pdf)
* **Tiwari and Kumar 2018**: [Link to paper](https://journals.aijr.org/index.php/ias/article/view/751/172)
* A deep learning approach to Solar Radio flux Forecasting: [Link to paper](https://www.sciencedirect.com/science/article/pii/S009457652100415X)

# Machine learning to predict the solar flux and geomagnetic indices to model density and Drag in Satellites: [Link to paper](http://arxiv.org/abs/2307.05002)

* Deep Learning LSTM-based approaches for 10.7 cm solar radio flux forecasting up to 45-days: [Link to paper](https://www.sciencedirect.com/science/article/pii/S2213133724000015)
* Machine learning Mastery - [Link](https://machinelearningmastery.com/)
* Solar Cycle Progression - [Link](https://www.swpc.noaa.gov/products/solar-cycle-progression)
* Space Weather Canada - [Link](https://spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-5-en.php)
* NASA Archives - [Link](https://www.nasa.gov/solar-cycle-progression-and-forecast/archived-forecast/)