

A Comparative Study of Geoeffectiveness Prediction Models using Synthetic L1 Data

Senne DOUMEN

Supervisor: Prof. Dr. Stefaan Poedts
KU Leuven, Department of Mathematics

Master thesis submitted in fulfillment
of the requirements for the degree in
Master of Science in Mathematics, KU Leuven

Academic year 2023-2024

© Copyright by KU Leuven

Without written permission of the promoters and the authors it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to KU Leuven, Faculteit Wetenschappen, Geel Huis, Kasteelpark Arenberg 11 bus 2100, 3001 Leuven (Heverlee), Telephone +32 16 32 14 01. A written permission of the promoter is also required to use the methods, products, schematics and programs described in this work for industrial or commercial use, and for submitting this publication in scientific contests.

Leuven, 22 August 2024

Promotor

Author

Prof. Dr. Stefaan Poedts

Senne Doumen

Preface

Aim of the Work

Space weather research aims to understand and accurately predict the behavior of space weather and geomagnetic storms. In this research field, many models have been proposed to predict geomagnetic indices (which quantify how strong a geomagnetic storm is) from interplanetary conditions, e.g. the particle density, proton pressure and particle velocity in space.

Since the 1970s, coronal mass ejections (CMEs) have been known to be the main cause of non-recurrent geomagnetic storms. In recent years, much progress has been made in simulating the propagation of these CMEs to predict interplanetary conditions solely from solar observation.

This thesis aims to test the performance of prediction models for geomagnetic indices when they are given synthetic L1 data (simulated interplanetary conditions) instead of observational data. The usage of synthetic L1 data as input for the prediction models would allow the forecast to be made one or two days in advance, as opposed to only a few hours as is currently the case with the usage of real-time L1 data [Maharana et al., 2024]. This would greatly enhance the preparation time governments would have to prepare their satellites and power grids for the harmful effects of the geomagnetic storm.

Acknowledgements

This thesis could not have been completed without the irreplaceable contributions of several individuals, to whom I am deeply grateful.

- Firstly, thanks are given to Prof. Dr. Stefaan Poedts and drs. Anwesha Maharana, whose support and guidance were key to completing this thesis.
- To Alexey Isavnin and Andrey Kochanov for their help and quick responses to the many technical issues I encountered.
- My family for always being there for me, giving me a place to call home and encouraging me to achieve my dreams.
- My friends, without whom these five years in Leuven would have been much less fun, funny, educational, amusing, and enjoyable.

- My girlfriend for her daily support and reassurance, which were essential in the final stages of the completion of this thesis.

Contribution Statement

The resources and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

Prof. Dr. Stefaan Poedts and ESA provided access to the Virtual Space Weather Modeling Center, which enables access to EUHFORIA, the Newell Kp model, the O'Brien Dst model, and the SNGI models. Dr. Peter Wintoft provided access to the feed-forward neural network Kp prediction model. Maria Tahtouh, Guillaume Bernoux, and Antoine Brunet from the French Aerospace Lab ONERA provided access to the recurrent neural network model SWING.

Popularizing Summary

Space weather research can be seen as the weather forecast for space. Just like a person wants to know whether a rain cloud passes over their house, a space weather researcher wants to know whether a coronal mass ejection or CME hits the Earth on its trajectory. A coronal mass ejection is a giant blob of plasma ejected from the Sun. CMEs are the main drivers of geomagnetic storms, just as clouds are the main drivers of a rainstorm. So, in this comparison, a CME can be seen as a space cloud.

It is crucial to accurately predict these geomagnetic storms, as they can interfere with radio waves, damage satellites, disrupt GPS, and completely shut down power grids. It has been estimated that a powerful geomagnetic storm would cost one trillion USD and take four to ten years in repairs to recover fully. To measure how strong a geomagnetic storm is, specific indices have been defined to quantify their strength. This thesis investigates two of these indices: the Kp and Dst index. They can be compared to the expected rainfall associated with a cloud.

Just as not every cloud is a rain cloud, not every CME will trigger a geomagnetic storm. Specific parameters of a CME determine whether it will or will not trigger a storm. Currently, a satellite between the Sun and the Earth measures these parameters. Many prediction models have been devices for predicting how strong the Kp and Dst index will be from these measurements. In this thesis, seven prediction models are investigated: four for the Kp index and three for the Dst index.

Predictions based on satellite measurements have the disadvantage that they can only predict these indices a few hours in advance. Currently, researchers are investing a lot of effort in simulating how a CME ejected by the Sun propagates in our solar system. One of these simulation models is EUHFORIA, which was developed at KU Leuven by Pomoell and Poedts [2018]. EUHFORIA uses a synoptic magnetogram (the scientific term for a picture of the Sun that shows how strong the magnetic field is at each position) and observationally determined parameters of the CME (e.g., its velocity and propagation angle) to simulate how this CME will propagate in the three-dimensional domain, predict whether it will hit Earth, and determine its relevant parameter values.

This thesis investigates how well the prediction models match the observed values when the EUHFORIA simulations are used as input instead of the satellite measurements. With this procedure, the forecast can already be made at the beginning of the CME's trajectory, not just when it hits the satellite. This allows the forecast to be made one or two days in advance instead of only a few hours [Maharana et al., 2024]. This would significantly enhance the preparation time governments would have to prepare their satellites

and power grids for the harmful effects of the geomagnetic storm.

Five statistical measures are calculated to quantify the performance of these models using simulated input values, and they are compared to each other and to a reference model. It does this for two test cases: the CME events of 12 July 2012 and 4-6 September 2017. Finally, this thesis also investigates this for sixteen virtual satellites placed around the Earth to evaluate how sensitive all these models are to spatial variations and small deviations in the input values.

The analysis shows that all four Kp prediction models consistently outperform the reference model in the five statistical measures. The same conclusion can be drawn for two out of three Dst prediction models. The results of this research also suggest that the discussed models are not oversensitive to spatial derivations and small derivations in the input values.

Scientific Summary

In this thesis the performance is assessed of four Kp prediction models and three Dst prediction models with synthetic L1 data as input. The use of synthetic L1 data allows for a forecast to be made one or two days in advance, as opposed to only a few hours as is currently the case with real-time L1 data [Maharana et al., 2024]. This would greatly enhance the preparation time governments would have to prepare their satellites and power grids for the harmful effects of the geomagnetic storm.

The four Kp prediction models discussed in this thesis are the coupling function devised by Newell et al. [2008], a feed-forward neural network model trained by Wintoft et al. [2017], a recurrent neural network called SWING, and the SNGI Kp prediction model which is based on the NARMAX methodology as described by Ayala Solares et al. [2016]. The three Dst prediction models discussed in this thesis are the AK2 model derived by O'Brien and McPherron [2000a], a recurrent neural network called SWING, and the SNGI Dst prediction model which is based on the NARMAX methodology described by Ayala Solares et al. [2016].

In this study, two test cases were investigated, these being the CME events of 12 July 2012 and 4-6 September 2017. The synthetic L1 data for these test cases is obtained by performing a EUHFORIA simulation. EUHFORIA is a three-dimensional time-dependent magnetohydrodynamic (MHD) model designed by Pomoell and Poedts [2018] for space weather research and operational purposes.

For each EUHFORIA simulation the plasma parameters are extracted at the virtual location of the Earth and of 16 virtual satellites situated on what is in this thesis referred to as the 5 ° and 10 ° lattice. These are included to investigate how sensitive the models are to spatial variations and small deviations in the input values. Each of these plasma parameter sets is then used as the input for the aforementioned four Kp prediction models and three Dst prediction models. To each of the predictions using synthetic L1 input data five metrics are applied to quantify how well they resemble the observed Kp and Dst values for these events. The metrics are also applied to the reference models $Kp = 0$ and $Dst = 0$. These five metrics are the bias, the root mean square error, the Pearson correlation coefficient, the coefficient of determination and the self similarity factor obtained using the dynamic time warping methodology.

The analysis performed shows that all four Kp prediction models consistently outperform the reference model in each of the five statistical measures. The same conclusion can be drawn for two out of three Dst prediction models. From the results in this research it can also be concluded that the discussed models are indeed not oversensitive to spatial derivations and small derivations in the input values.

List of Acronyms and Symbols

L1	The first Lagrangian point
CME	Coronal Mass Ejection
VSC	Vlaams Supercomputer Centrum/ Flemish Supercomputer Center
FWO	Fonds Wetenschappelijk Onderzoek/ Research Foundation - Flanders
EUHFORIA	European Heliospheric Forecasting Information Asset
Kp	Planetarische Kennziffer/ planetary index
Dst	Storm-time variation of the disturbance field
SNGI	Sheffield NARMAX geomagnetic indices models
ESA	European Space Agency
SWING	Recurrent neural network model developed at ONERA to predict geomagnetic indices
ONERA	Office national d'études et de recherches aérospatiales/ French national aerospace research centre
NARMAX	Nonlinear AutoRegressive Moving Average eXogenous models
AK2	ACE-Kyoto 2 model, internal nomenclature referring to the Dst prediction model derived by O'Brien and McPherron [2000a]
MHD	Magnetohydrodynamics
C -index	Daily international character figure
K -index	Kennziffer
AE -index	Auroral electrojet index
am indices	Mean amplitude indices
D	Disturbance field
H	Horizontal component of the geomagnetic field in the southward direction

λ	Magnetic dipole latitude of observatory
H_{base}	Baseline value of the geomagnetic main field
Sq	Quiet-day solar variation
UT	Universal time
Ks	Standardized K values
AK1	ACE-Kyoto 1 model, internal nomenclature referring to the Dst prediction model derived by Burton et al. [1975]
Dst^*	Corrected Dst
b	Empirically determined constant
P_{dyn}	Solar wind dynamic pressure
nPa	Nanopascal
c	Empirically determined constant
Q	Function proportional to the rate of energy injection into the ring current
τ	Decay time
VBs	Dawn-to-dusk component of interplanetary electric field
VB_z	z-component of interplanetary electric field
B_z	z-component of interplanetary magnetic field
OMNI	NASA database for solar wind magnetic field and plasma data, energetic proton fluxes (> 1 to > 60 MeV), and geomagnetic and solar activity indices
ODI	Open Data Interface, a common backend and database system for space environment data processing systems at ESA
t	Time
Δt	Timestep
n	Particle density (cm^{-3})
cm	Centimeter
v	Velocity (km/s)
km/s	Kilometer per second
B_T	Interplanetary field magnitude (nT)
nT	Nanotesla

θ_c	IMF clock angle
IMF	Interplanetary magnetic field
B_y	y-component of the interplanetary magnetic field
$\frac{d\Phi_{MP}}{dt}$	Coupling function defined by Equation 2.6
r	Correlation coefficient
FNN	Feed-forward neural network
a	Modeling scalar
$\vec{v}$	Row matrix of modeling parameters
$\vec{b}$	Column matrix of modeling parameters
W	$m \times n$ matrix of modeling parameters
$\vec{x}$	Column matrix of input variables
g	Transfer function tanh
ACE	Advanced Composition Explorer
T_{UT}	UT hour
D_{UT}	UT day of year
RNN	Recurrent neural network
GRU	Gated recurrent unit
LSTM	Long-short-term memory
GP	Gaussian process
y_t	Output at time t
b_o	Output layer bias
J	Number of hidden weights
v_j	Hidden weights
F	Nonlinear activation function
a_j	Hidden layer biases
w_{ij}	Weights
u_{jk}	Recurrent weights
z_t^j	Update gate at time t

$\tilde{h}_t^j$	Candidate activation at time t
σ	Sigmoid function
W_z	Weight matrix that operates on the input vector
U_z	Weight matrix that operates on the past activation
x_t	Input vector at time t
h_{t-1}	Past activation at time t
W_h	Weight matrix
U_h	Weight matrix
r_t^j	Reset gate at time t
W_r	Weight matrix
U_r	Weight matrix
SNRB	NARMAX model to predict electron fluxes at geostationary orbit covering the GOES 15 energy channels $\geq 800\text{keV}$ and $\geq 2\text{MeV}$
GOES	Geostationary Operational Environmental Satellite
keV	Kiloelectronvolt
MeV	Megaelectronvolt
NARX	Nonlinear AutoRegressive with eXogenous input
$f(\cdot)$	Function to be determined from data
$u(k)$	System input
$y(k)$	System output
$e(k)$	System noise
n_u	Maximum lag for the input signal
n_y	Maximum lag for the output signal
l	Order of polynomial
OFR	Orthogonal forward regression
ERR	Error Reduction Ratio
MPO	Model-prediction output
V	Solar wind speed (km/s)
B _s	Southward interplanetary magnetic field (nT)

VBs	$V \cdot Bs/1000$
p	Solar wind pressure (nPa)
h	Modeling horizon
$Kp(k)$	Measured Kp index at time k
$\widehat{Kp}(k)$	Measured Kp index at time k
NASA	National Aeronautics and Space Administration
AU	Astronomical unit
PFSS	Potential field source surface model
SCS	Schatten Current Sheet model
f	Flux tube expansion factor
d	Distance from the footpoint of the flux tube to the nearest coronal hole boundary
v_{sw}	Solar wind speed
LFSS	Linear force-free spheromak model
vr	Radial velocity
$vlon$	Longitudinal velocity
$vclt$	Co-latitudinal velocity
$vtot$	Total velocity
Br	Radial magnetic field strength
$Blon$	Longitudinal magnetic field strength
$Bclt$	Co-latitudinal magnetic field strength
$Btot$	Total magnetic field strength
N	Number of data points
$\bar{y}$	Predicted value
y	Observed value
RMSE	Root mean square error
CORR	Pearson linear correlation coefficient
$\langle \cdot \rangle$	Mean value
R^2	Coefficient of determination

DTW	Dynamic time warping
$D(i, j)$	Cumulative DTW cost
$\delta(s, q)$	Euclidian distance between the points s and q
SSF	Sequence similarity factor
O	Observed data
M	Modeled data
$\bar{O}$	Averaged observed data

List of Figures

5.2.1 Results of EUHFORIA simulations for the 2012 test case.	22
5.2.2 Results of EUHFORIA simulations for the 2017 test case.	23
7.1.1 Kp values for the 2012 case using the Newell formula.	30
7.1.2 Kp values for the 2017 case using the Newell formula.	30
7.1.3 Dst values for the 2012 case using the O'Brien formula.	31
7.1.4 Dst values for the 2017 case using the O'Brien formula.	31
7.2.1 Kp values for the 2012 case using the FNN model.	32
7.2.2 Kp values for the 2017 case using the FNN model.	32
7.2.3 Kp values for the 2012 case using the SWING model.	33
7.2.4 Kp values for the 2017 case using the SWING model.	33
7.2.5 Dst values for the 2012 case using the SWING model.	34
7.2.6 Dst values for the 2017 case using the SWING model.	34
7.3.1 Kp values for the 2012 case using the SNGI model.	35
7.3.2 Kp values for the 2017 case using the SNGI model.	35
7.3.3 Dst values for the 2012 case using the SNGI model.	36
7.3.4 Dst values for the 2017 case using the SNGI model.	36
7.4.1 Kp values for the EUHFORIA simulation at Earth for the 2012 test case. .	37
7.4.2 Kp values for the EUHFORIA simulation at Earth for the 2017 test case. .	38
7.4.3 Dst values for the EUHFORIA simulation at Earth for the 2012 test case.	39
7.4.4 Dst values for the EUHFORIA simulation at Earth for the 2017 test case.	39
8.1.1 Bias of the Kp models for the 2012 test case.	42
8.1.2 Bias of the Kp models for the 2017 test case.	43
8.1.3 Bias of the Dst models for the 2012 test case.	44
8.1.4 Bias of the Dst models for the 2017 test case.	44
8.2.1 Root mean square error of the Kp models for the 2012 test case.	45

8.2.2 Root mean square error of the Kp models for the 2017 test case.	46
8.2.3 Root mean square error of the Dst models for the 2012 test case.	47
8.2.4 Root mean square error of the Dst models for the 2017 test case.	47
8.3.1 Pearson correlation coefficient of the Kp models for the 2012 test case. . .	48
8.3.2 Pearson correlation coefficient of the Kp models for the 2017 test case. . .	49
8.3.3 Pearson correlation coefficient of the Dst models for the 2012 test case. . .	50
8.3.4 Pearson correlation coefficient of the Dst models for the 2017 test case. . .	50
8.4.1 Coefficient of determination of the Kp models for the 2012 test case. . . .	51
8.4.2 Coefficient of determination of the Kp models for the 2017 test case. . . .	52
8.4.3 Coefficient of determination of the Dst models for the 2012 test case. . . .	53
8.4.4 Coefficient of determination of the Dst models for the 2017 test case. . . .	53
8.5.1 SSF scores of the Kp models for the 2012 test case.	54
8.5.2 SSF scores of the Kp models for the 2017 test case.	55
8.5.3 SSF scores of the Dst models for the 2012 test case.	56
8.5.4 SSF scores of the Dst models for the 2017 test case.	57

List of Tables

1.1	K index range limits at the Niemegk observatory	6
1.2	Kp measurement network	7
4.1	Data Set Variables	16
8.1	Metric values for the Kp prediction models using the EUHFORIA synthetic data at Earth for the 2012 test case.	58
8.2	Metric values for the Kp prediction models using the EUHFORIA synthetic data at Earth for the 2017 test case.	59
8.3	Metric values for the Dst prediction models using the EUHFORIA synthetic data at Earth for the 2012 test case.	60
8.4	Metric values for the Dst prediction models using the EUHFORIA synthetic data at Earth for the 2017 test case.	60

Contents

Preface	i
Popularizing Summary	iii
Scientific Summary	v
List of Acronyms and Symbols	vii
List of Figures	xiii
List of Tables	xv
Table of contents	xvii
Introduction	1
1 Geomagnetic Indices	5
1.1 <i>Dst</i> Index	5
1.2 <i>Kp</i> Index	6
2 Models based on an Empirical Relation	9
2.1 AK2 Dst Model by O’Brien and McPherron [2000a]	9
2.2 Kp Model by Newell et al. [2008]	10
3 Neural Network Models	11
3.1 Feed-forward Neural Networks	11
3.2 Recurrent Neural Networks	12
4 NARMAX Models	15
4.1 SNGI	15

5	Providing L1 Data	19
5.1	Observational Data	19
5.2	Synthetic Data	19
5.2.1	EUHFORIA	19
5.2.2	Results	20
6	Metrics	25
6.1	Bias	25
6.2	Root Mean Square Error	25
6.3	Pearson Correlation Coefficient	25
6.4	Coefficient of Determination	26
6.5	Dynamic Time Warping	26
7	Results	29
7.1	Empirical Models	29
7.1.1	Newell Kp Model	29
7.1.2	O'Brien Dst Model	31
7.2	Neural Network Models	32
7.2.1	FNN Kp Model	32
7.2.2	SWING Kp Model	33
7.2.3	SWING Dst Model	34
7.3	NARMAX Models	35
7.3.1	SNGI Kp Model	35
7.3.2	SNGI Dst Model	36
7.4	Comparison	37
7.4.1	Kp Models	37
7.4.2	Dst Models	38
8	Analysis	41
8.1	Bias	41
8.1.1	Kp Models	41
8.1.2	Dst Models	43
8.2	Root Mean Square Error	45
8.2.1	Kp Models	45
8.2.2	Dst Models	46

8.3	Pearson Correlation Coefficient	48
8.3.1	<i>Kp</i> Models	48
8.3.2	<i>Dst</i> Models	49
8.4	Coefficient of Determination	51
8.4.1	<i>Kp</i> Models	51
8.4.2	<i>Dst</i> Models	52
8.5	SSF Score	54
8.5.1	<i>Kp</i> Models	54
8.5.2	<i>Dst</i> Models	55
8.6	Comparison	58
8.6.1	<i>Kp</i> Models	58
8.6.2	<i>Dst</i> Models	59
9	Conclusion	61
	Bibliography	63

Introduction

Holy Helios who rises in the east, good mariner, trustworthy leader of the sun's rays, reliable witness, who of old didst establish the mighty wheel of the heavens, holy orderer, ruler of the axis of the heaven, Lord, Brilliant Leader, King, Soldier. I, Senne, student of KU Leuven, present my dissertation before you, that you will appear to me without causing me fear, and you will be revealed to me without causing me terror, and you will conceal nothing from me and will tell me truthfully all that I desire.

- Adapted from the "Prayer to the sun god Helios" from the SEIPHER HAZIM, as translated by Morgan [2022].

Humanity has always had a special connection with our Sun. Throughout history, many nations and cultures have woven myths, legends, traditions, and superstitions about the Sun [Olcott, 1914]. This is no surprise, as the Sun gives us life, light, and warmth. In light of modern science, myths and legends have made place for knowledge and facts. It is not his loyal steeds that pull Helios' chariot from east to west, but the rotation of the earth makes our Sun rise and set. It is not his shining aureole that provides us with light, but the process of nuclear fusion that takes place in the Sun's core leads to the creation of photons. Although they cannot be attributed to his erratic nature, for which all Olympian gods are so well known, we must still be aware of the dangers the Sun might pose to us and our society.

The impact of space weather on technology has been known since the mid-19th century, e.g., through the effects of geomagnetically induced currents on telegraph systems [Hapgood, 2011]. Over the past decades, our civilization has greatly increased its sensitivity to space weather as technological systems have become more critical to the functioning of economies and societies worldwide.

Ever since the invention and deployment of the first electrical communication system, the electrical telegraph [Bothmer and Daglis, 2007], it has been observed that geomagnetic storms can interfere with man-made systems on Earth: they can interfere with radio waves, damage satellites, disrupt GPS and completely shut down power grids [Gruet et al., 2018]. Moreover, Council [2008] has estimated that at present, the occurrence of an event such as the one that took place in September 1859 (which is referred to as the Carrington event) would cost around one trillion USD and would take four to ten years to recover from.

Luckily, our scientific understanding of the processes that cause these impacts has dramatically increased in the past decades. Observational studies of the solar atmosphere in

the early 1970s revealed that the Sun sporadically ejects vast quantities of matter into the Solar System and the constant outflow known as the solar wind. These sporadic ejections are known today as coronal mass ejections (CMEs) [Green and Baker, 2015]. CMEs are common events that occur with a frequency that follows the (on average) 11-year solar cycle. At cycle minimum, approximately one CME is formed per day, and at cycle maximum, as many as five are formed per day. They can be understood as a mechanism by which the Sun ejects an atmospheric magnetic field that harbours magnetic energy and a complex morphology. When CMEs collide with the Earth's magnetic field (the magnetosphere), a magnetic reconnection between the CME and the magnetosphere can occur if the magnetic field of the CME has the correct orientation. Magnetic reconnection is a fundamental physical process in which plasma and magnetic field become momentarily decoupled. This reconfiguration of the field converts magnetic energy into other forms of energy. The magnetic field at the subsolar point of the magnetosphere (directly between the Sun and Earth) is strongly northward, so a strong southward-directed magnetic field within the CME could lead to magnetic reconnection and a joining of the CME's magnetic field with the magnetosphere. This drives the magnetosphere into a disturbed state, producing geomagnetic substorms and storms that lead to many space weather effects [Green and Baker, 2015]. The discovery of these coronal mass ejections in the 1970s and the recognition that these, rather than eruptive flares, are the cause of non-recurrent geomagnetic storms was a major turning point in our understanding of space weather [Weather et al., 2008].

As early as 1885, a first attempt was made to characterize geomagnetic activity. It was aimed at estimating the geomagnetic disturbances on a daily base. The first defined geomagnetic index was the so-called "C" index, to which an observer assigned a numerical value of 0, 1, or 2 by judging the relative degree of disturbance of the magnetogram. These values correspond to a quiet, moderately disturbed, or disturbed state. However, the crudeness of this index and the observer's subjectivity in its determination eventually led to the introduction of new indices that allow objective monitoring of irregular variations.

Bartels et al. [1939] first introduced the K index, for which its measurement in a network of stations was then combined to form the Kp index. Later, also the Dst (ring current behavior, [Sugiura, 1963]), AE (maximum of the auroral electrojet intensity, [Davis and Sugiura, 1966]) and am indices (K -derived planetary activity index, [Mayaud, 1968]) were proposed. In this work the Kp and Dst indices are regarded. Their descriptions can be found in Chapter 1.

Many efforts have been made to predict these indices from observational solar wind data using several methodologies. Firstly, an empirical analysis can be performed to derive a relationship or coupling function between the interplanetary conditions and the geomagnetic index. For the Dst index, this work refers to the prediction model derived by O'Brien and McPherron [2000a], which is a correction to the model derived initially by Burton et al. [1975]. For the Kp index, this work regards the coupling function derived by Newell et al. [2008]. The description of these models can be found in Chapter 2.

Secondly, a neural network can be trained to predict the geomagnetic indices values given as input parameters the interplanetary conditions. This work considers both feed-forward neural networks (only for the Kp index) and recurrent neural networks. The description of these models can be found in Chapter 3.

Lastly, the lesser-known NARMAX methodology is also investigated, which is a system identification modeling approach. The description of these models can be found in Chapter 4.

Whereas these models have all been derived or trained using observational solar wind data, this work aims to assess their performance using synthetic solar wind data. Using synthetic L1 data would allow a forecast to be made one or two days in advance, as opposed to only a few hours, as is currently the case with real-time L1 data [Maharana et al., 2024]. This would significantly enhance the preparation time governments would have to prepare their satellites and power grids for the harmful effects of the geomagnetic storm. The synthetic data used in this thesis is obtained by simulations using EUHFORIA, a three-dimensional time-dependent magnetohydrodynamic (MHD) model designed by Pomoell and Poedts [2018]. A further description of EUHFORIA can be found in Chapter 5.

The performance of the models using synthetic data is assessed using several metrics. The explanation of these metrics can be found in Chapter 6.

Finally, the results of these prediction models with synthetic input data can be found in Chapter 7, and the resulting statistical analysis is discussed in Chapter 8.

Chapter 1

Geomagnetic Indices

The following chapter provides a description of the two geomagnetic indices discussed in this thesis. These two are the *Dst* index and the *Kp* index. The following explanations roughly follow the ones provided by Menvielle et al. [2010].

1.1 *Dst* Index

The Disturbance storm time index, or *Dst* index, was developed by Sugiura [1963] and measures the strength of the ring current. The ring current is generally believed to be the primary measure of the intensity of a geomagnetic storm. *Dst* stands for the Storm-Time variation of the disturbance field *D*. This *D*-field is the part of the Earth's magnetic field generated by external currents and their associated induced Earth currents [Akasofu and Chapman, 1961]. More specifically, the *Dst* index is a measure of the variation in the horizontal component of the geomagnetic field in the southward direction parallel to the geomagnetic dipole axis. This horizontal component is denoted by *H*, and the variation of disturbance in this component is denoted by *D*. These variations are measured at four magnetic observatories in Hermanus (South Africa), Kakioka (Japan), Honolulu (Hawaii, USA) and San Juan (Puerto Rico).

The following description is based on Sugiura [1991]. The reader is referred to this article for further technical details. The *Dst* index at time *t* is defined as the average of these variations *D_i*, divided by the average of the cosines of the dipole latitudes at the observatories:

$$Dst(t) = \sum_{i=1}^4 D_i(t) / \sum_{i=1}^4 \cos(\lambda_i), \quad (1.1)$$

where λ_i is the magnetic dipole latitude at the *i*-th observatory. Note that the normalization is by the average of the four cosines and not by for each *D_i* by $\cos(\lambda_i)$. According to Sugiura [1991], this normalization procedure has been found to minimize the undesirable effects of missing hourly values. Each *D_i*(*t*) is calculated from the observed *H*(*t*) by subtracting the baseline value of the geomagnetic main field *H_{base}*(*t*) and the quiet-day solar variations *Sq*(*t*). For more information on how *H_{base}*(*t*) and *Sq*(*t*) are calculated, the reader is again referred to Sugiura [1991].

Hourly Dst values are provided by the World Data Center for Geomagnetism in Kyoto. The data can be retrieved online from their website ¹. There are three classes of Dst index services: Real-time (Quicklook) Dst , provisional Dst and final Dst . The classes are defined according to the data processing stage provided by the observatories, which differ mainly in determining the baseline $H_{base}(t)$.

1.2 Kp Index

The Kp index is one of the most commonly used parameters to quantify geomagnetic activity. It stands for planetarische Kennziffer, which is German for planetary index [Ayala Solares et al., 2016]. It is the weighted sum of 13 K indices measured all around the globe. The K index was introduced by Bartels et al. [1939] and is based on the irregular variations of the two horizontal geomagnetic components after eliminating the non- K variations. The vertical component is not considered.

The difficulty in scaling the K indices lies in determining the non- K variations. According to Mayaud [1967], the non- K variations are defined as the simplest and least speculative smooth curve that corresponds to a possible solar regular variation. The K indices are determined on a 3-hour basis and are integers ranging from 0–9. These integers represent ten classes of ranges. The exact range limits differ for each observatory to account for the variations of the observed geomagnetic perturbations with geomagnetic latitude. The range limits for the Niemegk Observatory in Germany can be seen in Table 1.1.

Range (nT)	0 - 5	5 - 10	10 - 20	20 - 40	40 - 70	70 - 120	120 - 200	200 - 330	330 - 500	> 500
K value	0	1	2	3	4	5	6	7	8	9

Table 1.1: K index range limits at the Niemegk observatory

An individual K index is thus an integer in the range 0 to 9 corresponding to a class that contains the largest range of geomagnetic disturbances in either of the two horizontal components during a 3 h UT interval.

It is generally accepted that $K \approx 0 - 2$ corresponds to periods of magnetic quietness, $K \approx 3 - 5$ to periods of moderate activity, and $K \approx 6 - 9$ to periods of intense geomagnetic activity.

These K values are then standardized using the conversion tables defined by Bartels [1951]. These standardized K values are referred to as Ks . The purpose of this standardization is to eliminate time- and seasonal-dependent features. Where there were ten options for the values of K , there are 28 options for the values of Ks . These are scaled as $0o, 0+, 1-, 1o, 1+, 2-, \dots$ to $9o$. These are also sometimes expressed as $3Ks$ ranging from 0 to 27.

Finally, the Kp index is defined as the average of Ks values from 13 measuring stations. This network of stations can be seen in table 1.2.

The Ks data for the two stations, Brorfelde and Uppsala (6) and Eyrewell and Canberra (11), are combined so that their averages enter into the calculation of Kp .

¹<https://wdc.kugi.kyoto-u.ac.jp/dst/dir/>

(1)	Meannook	Canada
(2)	Sitka	USA
(3)	Lerwick	UK
(4)	Ottowa	Canada
(5)	Eskdalemuir	UK
(6)	Brorfelde	Denmark
	Uppsala	Sweden
(7)	Fredericksburg	USA
(8)	Wingst	Germany
(9)	Niemegk	Germany
(10)	Hartland	UK
(11)	Eyrewell	New-Zealand
	Canberra	Australia

Table 1.2: Kp measurement network

The Kp index has the same scaling as the Ks indices.

Chapter 2

Models based on an Empirical Relation

The following chapter provides a description of the investigated Dst and Kp prediction models based on empirical relations.

2.1 AK2 Dst Model by O'Brien and McPherron [2000a]

The investigated model is the AK2 model derived by O'Brien and McPherron [2000a]. It uses a dynamical equation nearly identical to the one derived by Burton et al. [1975] with only a slight correction. The model derived by Burton et al. [1975] is also referred to as the AK1 model [O'Brien and McPherron, 2000b]. It was shown by O'Brien and McPherron [2000b] that the AK2 model provides better results than the AK1 model. Both equations remove the contribution of the magnetopause current system when studying Dst . This correction is given by:

$$Dst^* = Dst - b\sqrt{P_{dyn}} + c, \quad (2.1)$$

where P_{dyn} is the solar wind dynamic pressure and b and c are empirically determined constants. The Burton equation is a simple differential equation for the evolution of the corrected Dst^* and is given by:

$$\frac{dDst^*}{dt} = Q(t) - \frac{Dst^*(t)}{\tau}, \quad (2.2)$$

where Q is proportional to the energy injection rate into the ring current and τ is the decay time. Q is generally given in terms of VBs , the dawn-to-dusk component of the interplanetary electric field, defined as:

$$VBs = \begin{cases} |VB_z|, & \text{if } B_z < 0, \\ 0, & \text{if } B_z \geq 0. \end{cases} \quad (2.3)$$

The AK2 model modifies the AK1 model to allow the decay time τ to vary with VBs instead of using a constant value. This is a valuable correction, as O'Brien and McPherron

[2000b] showed that the AK2 model outperforms the AK1 model. The constant values used in the AK2 model were determined using the OMNI database of hourly solar wind and Dst measurements, which used data from 1964-1996. The AK2 model is then given by equations (2.1) and (2.2) where $b = 7.26 \text{ nT}/\sqrt{nPa}$, $c = 11 \text{ nT}$,

$$Q(nT/h) = -4.4(VBs - 0.5) \quad \text{and} \quad \tau(h) = 2.4 \exp\left(\frac{9.74}{4.69 + VBs}\right), \quad (2.4)$$

where VBs is given in mV/m and P_{dyn} in nPa.

In this implementation of the AK2 model, the initial Dst value is taken during the model start-up by querying the ODI database. The model then uses a simple forward difference to integrate the Dst index from the initial value:

$$Dst^*(t + \Delta t) = Dst^*(t) + \left(Q(t) - \frac{Dst^*(t)}{\tau}\right) \Delta t. \quad (2.5)$$

2.2 Kp Model by Newell et al. [2008]

The empirical Kp model used here is based on the empirical equation derived by Newell et al. [2008]. Newell et al. [2008] not only investigates the values of Kp but considers ten different magnetospheric state variables, among which both the Kp and Dst indices. It investigates 32 coupling functions, of which twenty are viscously related, and twelve are merging related solar wind-magnetosphere coupling functions. The term coupling function is used because these functions couple the Kp index to the interplanetary conditions. Their study considers the 496 unique binary combinations to find the pair that best predicts these state variables.

For all of these candidates, the correlation coefficients were calculated and compared. It was found that the best performing pair consisted of $d\Phi_{MP}/dt$ combined with $n^{1/2}v^2$ where v is the velocity in km/s and n the particle density in cm^{-3} . The coupling function $d\Phi_{MP}/dt$ was introduced by Newell et al. [2007] and is defined as:

$$\frac{d\Phi_{MP}}{dt} = v^{4/3} B_T^{2/3} \sin^{8/3}(\theta_c/2), \quad (2.6)$$

where v is the velocity in km/s, B_T is interplanetary field magnitude in nT and θ_c is the IMF clock angle which is defined as $\theta_c = \arctan(B_y/B_z)$. The article also provides the explicit equation for the least variance linear prediction of Kp :

$$Kp = 0.05 + 2.244 \times 10^{-4} d\Phi_{MP}/dt + 2.844 \times 10^{-6} n^{1/2} v^2, \quad (2.7)$$

which has a correlation of $r = 0.866$.

Chapter 3

Neural Network Models

The following chapter describes the investigated *Dst* and *Kp* prediction models, which are derived using neural networks. This is opposed to Chapter 2, where empirical methods were used.

3.1 Feed-forward Neural Networks

The investigated FNN model is as presented in Wintoft et al. [2017]. This article only discusses the *Kp* prediction model, but a similar *Dst* model is introduced in Wintoft and Wik [2018]. This thesis discusses only the *Kp* prediction model, as the author was not granted access to the *Dst* model during this study. In what follows, a short overview of these models is given.

The algorithm used is a feed-forward neural network with one hidden layer. The function gives the mapping from input to output:

$$y(\vec{x}) = a + \vec{v}g(\vec{b} + \mathbf{W}\vec{x}) = a + \sum_{i=1}^m v_i g\left(b_i + \sum_{j=1}^n w_{i,j}x_j\right), \quad (3.1)$$

where a is a scalar, $\vec{v}$ is the row matrix $[v_1, \dots, v_m]$, $\vec{b}$ is the column matrix $[b_1, \dots, b_m]^T$, $\mathbf{W}$ is the $m \times n$ matrix

$$\begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{1,m} & \cdots & w_{n,m} \end{pmatrix} \quad (3.2)$$

and $\vec{x}$ is the input column matrix $[x_1, \dots, x_n]^T$. The transfer function g is the bounded nonlinear function $\tanh$. All free parameters $(a, \vec{v}, \vec{b}, \mathbf{W})$ have been determined through a training and validation procedure using ACE data and the final *Kp/Dst*.

The *Kp* model was constructed using the ACE solar wind data set from 1998 to 2015. The complete data set is divided into three independent data sets: the training set, the validation set, and the test set. The three sets are composed of data from the years (1998, 1999, 2002-2005, 2007-2010, 2013-2015), (2000, 2006, 2012), and (2001, 2011),

respectively. The solar wind variables are normalized to lie in the range $[+1, -1]$, where the normalization constants are determined from the training set.

The *Dst* model used OMNI solar wind data and the final *Dst*, with data from 1981, 1996, 2001 and 2008 as test data. The test set is completely removed from the training-validation loop to provide an unbiased estimate of the performance of the models.

The model also incorporates local time and day of year dependencies, such as sine and cosine of $2\pi T_{UT}/24$ and $2\pi D_{UT}/365$ where T_{UT} and D_{UT} are UT hours and day of the year.

As the network in equation 3.1 has no internal memory, the dynamics are instead coded into the inputs using past solar wind values. After some testing, it was found that two previous 3-hour segments were sufficient for the *Kp* model. Over the past 64 hours, the inputs for the *Dst* model have been extended to capture the dynamics of the recovery phase.

3.2 Recurrent Neural Networks

Another popular type of neural network is a recurrent neural network. A RNN models the dynamic behavior through internal states so that the output depends on both the input and the internal state. Thus, temporally correlated structures can be modeled without explicitly parameterizing the temporal dependence. Wintoft and Wik [2021] also developed three RNN prediction models, all based on different methodologies. The first is an Elman neural network model, the second is a gated recurrent unit (GRU) model, and the third is a long-short-term memory (LSTM) model.

However, in this thesis, the investigation of the models by Wintoft and Wik [2021] is not considered since the author was not given access to these models during this study. Fortunately, the author was granted access to another RNN prediction model called SWING. SWING is an LSTM neural network model that ONERA developed. It is very similar to the model presented by Gruet et al. [2018], also developed at ONERA. The model by Gruet et al. [2018] combines an LSTM recurrent neural network with a Gaussian process (GP) model to bring together the sequence modeling capabilities of a recurrent neural network with the error bars and confidence bounds provided by a GP. However, this thesis does not explore this GP feature, and the SWING model is regarded as a single-point prediction.

To explain the principles behind the LSTM neural network, an explanation following that given by Wintoft and Wik [2021] is presented here. This also includes the explanation of the Elman and GRU networks, which are not considered in this thesis. Nonetheless, it is the opinion of the author that the inclusion of their explanations here provides an intermediate step in the conceptual understanding of the LSTM neural network and thus allows for easier comprehension of the concepts involved.

The Elman network [Elman, 1990] was first developed for *Dst* prediction by Wu and Lundstedt [1997] and later implemented for real-time operation by Lundstedt et al. [2002].

The Elman network at time t is described by

$$y_t = b_o + \sum_{j=1}^J v_j h_t^j = \mathbf{V} \mathbf{h}_t, \quad (3.3)$$

$$h_t^j = F\left(a_j + \sum_{i=1}^I w_{ij} x_t^i + \sum_{k=1}^J u_{jk} h_{t-1}^k\right) = F(\mathbf{W} \mathbf{x}_t + \mathbf{U} \mathbf{h}_{t-1})^j, \quad (3.4)$$

where b_o is the output layer bias, v_j are J hidden weights, F is a nonlinear activation function, a_j are J hidden layer biases, w_{ij} are $J \times I$ weights and u_{jk} are $J \times J$ recurrent weights. The article uses a minimalistic Elman network, which only uses one input and one linear unit in the hidden layer. This translates to $b_o = 0$ and $v_1 = 1$, leading to

$$y_t = y_{t-1} + a_1 + w_{11} x_t^1 - (1 - u_{11}) y_{t-1}. \quad (3.5)$$

The gated recurrent unit (GRU) network [Cho et al., 2014] has a more complex architecture. The output is again given by equation 3.3. The GRU layer output at unit j is

$$h_t^j = z_t^j h_{t-1}^j + (1 - z_t^j) \tilde{h}_t^j, \quad (3.6)$$

where z_t^j is the update gate and $\tilde{h}_t^j$ is the candidate activation. The update gate is defined as

$$z_t^j = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1})^j, \quad (3.7)$$

where σ is the sigmoid function with output range zero to one, $\mathbf{W}_z$ is the weight matrix that operates on the input vector $\mathbf{x}_t$ and $\mathbf{U}_z$ is the matrix that operates on the past activation $\mathbf{h}_{t-1}$. The candidate activation $\tilde{h}_t^j$ is defined as

$$\tilde{h}_t^j = F(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h (\mathbf{r} \odot \mathbf{h}_{t-1}))^j, \quad (3.8)$$

where F is a nonlinear function with two additional weight matrices $\mathbf{W}_h$ and $\mathbf{U}_h$. The $\mathbf{U}_h$ matrix operates on the past activation weighted by the reset gate

$$r_t^j = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1})^j, \quad (3.9)$$

which again has a set of weight matrices $\mathbf{W}_r$ and $\mathbf{U}_r$. The GRU network is more complex than the Elman Network and has approximately three times more weight than the Elman network regarding the same number of units.

The long short-term memory (LSTM) network [Hochreiter and Schmidhuber, 1997] was introduced before the GRU network and has even more complexity with approximately four times more weights than the Elman network given the same type of units. Instead of GRU's two gating functions, it has three that control the information: the output gate, the forget gate, and the input gate. The equations are not repeated in the thesis; instead, the reader is referred to Chung et al. [2014].

The SWING Kp and Dst models were trained using high-resolution data, with a cadence of 10 minutes, from the OMNI database and used as input the plasma parameters that

correspond to the EUHFORIA outputs as described in Subsection 5.2.1. The training period spans from 1997 to 2021 (inclusive), with the exclusion of the testing years 2012 and 2017, which correspond to the years of the CME events that are used as test cases in this thesis, as described in Subsection 5.2.2. The validation data set covers the years 1995 and 1996. During its development, the model was never trained, validated, or tested on data from 2012 and 2017 to ensure the absence of data leakage.

Chapter 4

NARMAX Models

The following chapter describes the investigated *Dst* and *Kp* prediction models that follow the NARMAX methodology.

4.1 SNGI

The Sheffield NARMAX geomagnetic indices (SNGI) models are two forecast models developed at the University of Sheffield (UK) [Poedts, 2022]. These models were derived using a system identification technique based on Nonlinear AutoRegressive Moving Average eXogenous models (NARMAX). They are two of the four Sheffield NARMAX (SN) models; the other two are the SNRB models that simulate the electron fluxes in geostationary orbit covering the GOES 15 energy channels > 800 keV and > 2 MeV [Boynton et al., 2015]. However, these two are not relevant to the results of this thesis and are therefore not discussed here. In what follows, a short overview of the SNGI model for the *Kp* index is given, as was presented in Ayala Solares et al. [2016]. This article discusses using NARX models (Nonlinear AutoRegressive with eXogenous input), but these are simply a special case of the general NARMAX model [Billings et al., 1989].

The first step in dynamic system identification is to detect an appropriate model structure. This challenging task has been studied extensively, and much information is available in the literature. A popular approach is the NARX methodology, which has been used successfully to identify non-linear systems [Billings, 2013]. A NARX model can be written as

$$y(k) = f(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u)) + e(k), \quad (4.1)$$

where $f(\cdot)$ is a function to be determined from data, $u(k)$ is the system input, $y(k)$ the system output and $e(k)$ is the system noise (with $k = 1, 2, \dots, N$). The maximum lags for the input and output signals are n_u and n_y , respectively. Most approaches assume that the function $f(\cdot)$ can be approximated by a linear combination of a predefined set of past outputs and input functions. A popular choice for these functions is polynomials, which give rise to the polynomial NARX representation. A NARX model of order l means that the maximal order of each polynomial term in the model cannot exceed l . The most popular algorithm for NARX modeling is the Orthogonal Forward Regression (OFR) algorithm [Billings, 2013]. In the case of a polynomial NARX representation, the

OFR algorithm performs a stepwise regression procedure to identify the most significant model terms. To achieve this, it uses the Error Reduction Ratio (ERR) index to measure the significance of each candidate model term [Billings, 2013].

The second step is parameter estimation, which is typically performed using the traditional least square method, gradient descent, and the Metropolis-Hastings algorithm. The final step is model validation, for which several authors have developed different approaches. Because the NARX model depends on past outputs, a reliable way to check the model's validity is through the model-predicted output (MPO), which uses past predicted outputs to estimate future ones and provide details about the stability and predictability range of the models.

The article presents two approaches: a recursive sliding window approach and a direct approach. Although both models obtained good forecasts, the article concludes that the direct approach provides better forecasts than the sliding-window approach. Hence, only the methodology of the direct approach is discussed in this overview. The model's input parameters are presented in Table 4.1.

Symbol	Description
V	Solar wind speed (km/s)
B_s	Southward interplanetary magnetic field (nT)
VB_s	$V \cdot B_s / 1000$
p	Solar wind pressure (nPa)
$\sqrt{p}$	Square root of solar wind pressure

Table 4.1: Data Set Variables

Instead of training a model many times and using it recursively to calculate forecasts, the direct approach obtains a separate model for a horizon h of interest. In such a case, equation 4.1 becomes

$$y(k) = f(y(k-h), y(k-h-1), \dots, y(k-h-n_y), u(k-1), \dots, u(k-n_u)) + e(k). \quad (4.2)$$

The model is used to make predictions 3, 6, 12, and 24 hours in advance, so in this case, the direct approach produces four models. The training process uses the adaptive orthogonal search algorithm described in Billings and Wei [2008]. Numerical experimental results suggested $n_y = 4$ and $n_u = 2$ were an appropriate choice. The NARX model structure is then given by:

$$\begin{aligned} \widehat{Kp}(k) = f & \left(Kp(k-h), \dots, Kp(k-h-3), \right. \\ & V(k-1), V(k-2), Bs(k-1), Bs(k-2), \\ & VB_s(k-1), VB_s(k-2), p(k-1), p(k-2), \\ & \left. \sqrt{p(k-1)}, \sqrt{p(k-2)} \right), \end{aligned} \quad (4.3)$$

where $f(\cdot)$ is chosen to be a polynomial of degree $l = 2$, $Kp(k)$ is the measured Kp index at time k , and $\widehat{Kp}(k)$ is the predicted Kp index at time k . The article uses the first six months of 2000 for training, while the second half of the year is used for validation.

Explicit expressions of the identified models for this data set and the acquired results can be found in Section 5 of the article by Ayala Solares et al. [2016].

These results show a bias for low and high magnetic disturbances. Furthermore, the error bars become odd and difficult to interpret for high values of Kp ($Kp > 8$). This is because there are very few occurrences of high-value Kp indexes, so few predictions are made in such cases, and hence, they tend to be underpredicted. In general, this approach tends to slightly underpredict the Kp index on average. Furthermore, the forecasts are less accurate as the number of hours to predict ahead increases.

The last thing to note about the direct approach is that the explicit expression of the found models can be used to identify a set of significant model terms shared by all models, regardless of the time horizon of interest. This produces further information showing the relative contributions of these different drivers to changes in the Kp index.

Chapter 5

Providing L1 Data

5.1 Observational Data

This thesis's observational data were taken from the NASA OMNIWeb Service database, which can be accessed through <https://omniweb.gsfc.nasa.gov/>.

5.2 Synthetic Data

The synthetic data in this thesis were obtained by performing an EUHFORIA simulation for both test cases. A description of EUHFORIA can be found in Subsection 5.2.1 and a description of the simulation results in Subsection 5.2.2

5.2.1 EUHFORIA

The "European Heliospheric Forecasting Information Asset" (EUHFORIA) is a three-dimensional time-dependent magnetohydrodynamic (MHD) model designed by Pomoell and Poedts [2018] for space weather research and operational purposes. It consists of two parts: a coronal model and a heliospheric model. This division is possible because, at 0.1 AU and beyond, the solar wind plasma is supersonic and super-Alfvénic, meaning no information travels towards the Sun from there. As a result, the heliospheric model is agnostic of the coronal model, which makes the coupling between the two domains a one-way process.

The coronal model takes synoptic magnetograms as input and provides plasma conditions at 0.1 AU. First, the model computes the lower corona's global three-dimensional coronal magnetic field using a Potential Field Source Surface (PFSS) model [Altschuler and Newkirk, 1969]. Then, the Schatten Current Sheet (SCS) model extends the computed magnetic field to 0.1 AU [Schatten et al., 1969]. The solar wind speed at 0.1 AU is then found using an empirical formula $v_{sw} = v(f, d)$ that takes two of the magnetic field properties as input. These two properties are the flux tube expansion factor f and the distance d from the flux tube's foot point to the nearest coronal hole boundary. Finally,

the other required solar wind parameters can be derived from the solar wind speed v_{sw} using the formulas presented by Pomoell and Poedts [2018].

These parameters are then used as boundary conditions in the heliospheric model. This is a three-dimensional time-dependent MHD simulation that self-consistently evolves the background solar wind and the propagation of CMEs. The equations solved are those of ideal MHD augmented with gravity. The computational domain extends from 0.1 to 2 AU with resolution 512 and spans from -60° to 60° in latitude and 360° in longitude, both with a resolution of 2° . There are several options for modeling CMEs in EUHFORIA. In this work, the linear force-free spheromak (LFFS) model is used that was implemented in EUHFORIA by Verbeke et al. [2019]. It models the CME as a flux-rope structure, which it constructs as a force-free magnetic field configuration characterized by a global spherical shape. It was shown by Scolini et al. [2019] that this CME model gives better magnetic field predictions at L1 than the popular cone model. The output parameters of EUHFORIA are the particle density n , the thermal pressure P , the radial, longitudinal and co-latitudinal components of the velocity and the radial, longitudinal and co-latitudinal components of the interplanetary magnetic field strength, as can be seen in Figures 5.2.1 and 5.2.2. In these figures, the total velocity and interplanetary magnetic field strength are also displayed. These are not the model's direct output parameters but can be easily calculated from their components.

5.2.2 Results

This report uses two test cases to investigate the different models. These two cases are the CME events that occurred on 12 July 2012 and 4 – 6 September 2017. The 2012 CME event was a single CME, while the 2017 event consisted of three CMEs that rapidly followed each other. EUHFORIA provides plasma and magnetic field properties in three dimensions. The output is extracted not only at Earth, but also at multiple virtual satellites. For each case, the Dst and Kp index are modeled for the EUHFORIA output at Earth and sixteen other virtual satellites. These can be grouped into two sets of eight. From now on, one set is referred to as the 5° lattice and the other as the 10° lattice. The 5° lattice is given by the angular coordinates (0, 5), (5, 5), (5, 0), (0, -5), (-5, -5), (-5, 0), (5, -5) and (-5, 5) in the heliographic coordinate system. The 10° lattice is structured analogously.

The plasma parameters obtained using an EUHFORIA simulation for the 2012 case can be seen in Figure 5.2.1. The obtained parameter values are the particle density n , the plasma pressure P , the radial component of the velocity vr , the co-latitudinal component of the velocity $vclt$, the longitudinal component of the velocity $vlon$, the radial component of the interplanetary magnetic field strength Br , the co-latitudinal component of the interplanetary magnetic field strength $Bclt$ and the longitudinal component of the interplanetary magnetic field strength $Blon$. The total velocity $vtot$ and total interplanetary magnetic field strength $Btot$ are also calculated using the relationships $vtot = \sqrt{vr^2 + vclt^2 + vlon^2}$ and $Btot = \sqrt{Br^2 + Bclt^2 + Blon^2}$.

Of particular interest is the variable $Blon$, as this component of the interplanetary magnetic field is often linked to magnetic reconnection and the occurrence of geomagnetic storms.

In Figure 5.2.1, the blue line indicates the simulation results at Earth, while the green shaded area indicates how much the virtual satellites differ from these. The observed plasma parameters are also displayed in orange for reference. The data gaps in the observations were solved using linear interpolation and are shown in red. The results of the EUHFORIA simulation of the 2017 test case are plotted analogously and can be seen in Figure 5.2.2.

2012 Test Case

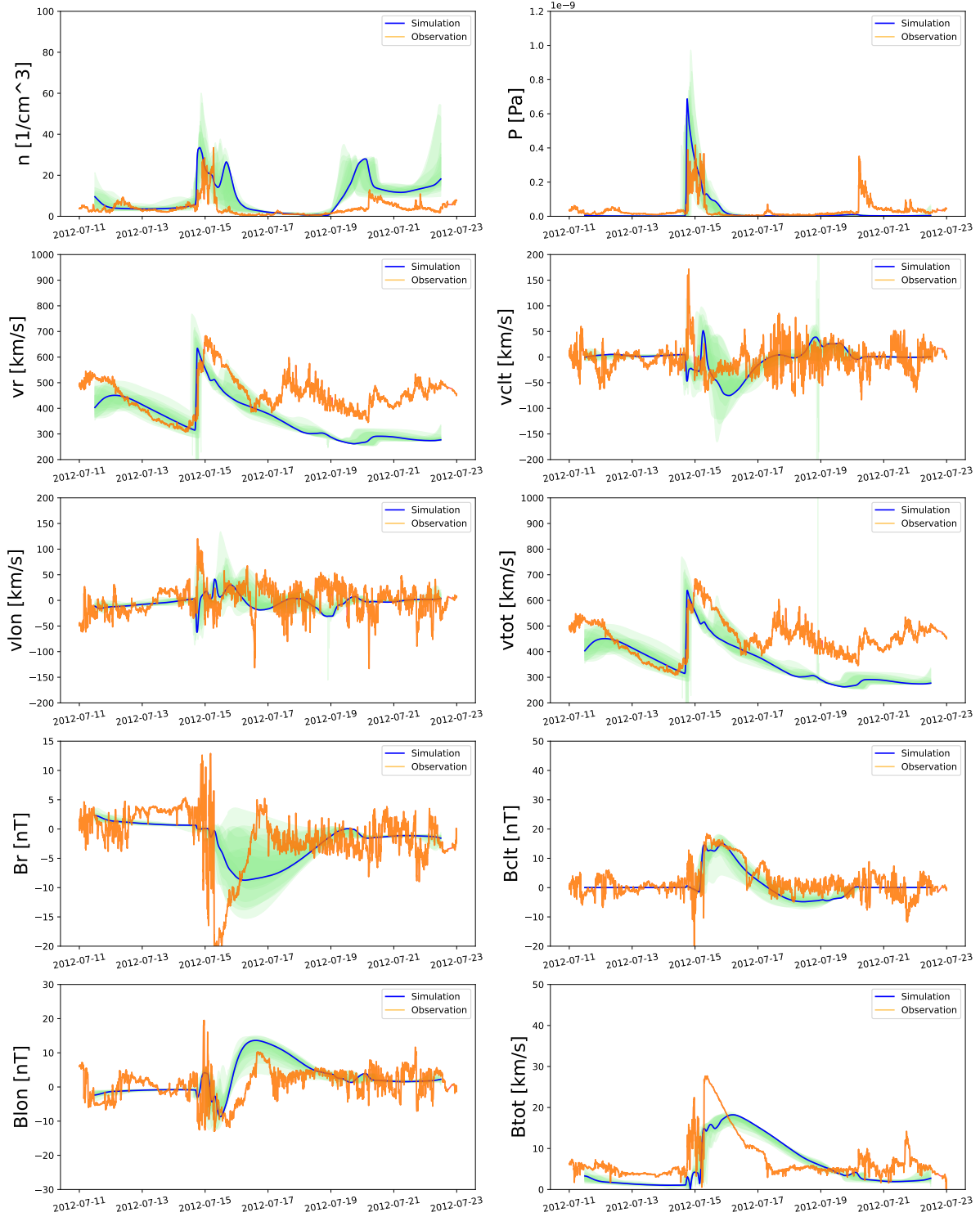


Figure 5.2.1: Results of EUHFORIA simulations for the 2012 test case.

2017 Test Case



Figure 5.2.2: Results of EUHFORIA simulations for the 2017 test case.

Chapter 6

Metrics

The following chapter provides a description of the different metrics used to assess the performance of the various models.

6.1 Bias

The bias is defined as

$$\text{BIAS} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i - y_i, \quad (6.1)$$

where N is the number of data points, $\bar{y}$ is the predicted Kp or Dst value and y is the observed Kp or Dst value [Wintoft and Wik, 2018]. For a perfect prediction, $\text{BIAS} = 0$.

6.2 Root Mean Square Error

The root mean square error is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - y_i)^2}, \quad (6.2)$$

where N is again the number of data points, $\bar{y}$ is the predicted Kp or Dst value and y is the observed Kp or Dst value [Wintoft and Wik, 2018]. For a perfect prediction, $\text{RMSE} = 0$.

6.3 Pearson Correlation Coefficient

The Pearson correlation coefficient is defined as

$$\text{CORR} = \frac{\sum_{i=1}^N (\bar{y}_i - \langle \bar{y} \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^N (\bar{y}_i - \langle \bar{y} \rangle)^2} \sqrt{\sum_{i=1}^N (y_i - \langle y \rangle)^2}}, \quad (6.3)$$

where N is again the number of data points, $\bar{y}$ is the predicted Kp or Dst value, y is the observed Kp or Dst value and $\langle \cdot \rangle$ indicates the mean value [Wintoft and Wik, 2018]. In this thesis, the mean value is replaced by the reference value $Kp = 0$ or $Dst = 0$.

The definition then becomes:

$$\text{CORR} = \frac{\sum_{i=1}^N \bar{y}_i y_i}{\sqrt{\sum_{i=1}^N \bar{y}_i^2} \sqrt{\sum_{i=1}^N y_i^2}}, \quad (6.4)$$

For a perfect forecast, $\text{CORR} = 1$.

6.4 Coefficient of Determination

The coefficient of determination, also known as the R^2 statistic, is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^N (\bar{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \langle y \rangle)^2}, \quad (6.5)$$

where N is again the number of data points, $\bar{y}$ is the predicted Kp or Dst value, y is the observed Kp or Dst value and $\langle \cdot \rangle$ indicates the mean value [Wintoft and Wik, 2018]. Again, the mean value is replaced by the reference value $Kp = 0$ or $Dst = 0$. The definition then becomes:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\bar{y}_i - y_i)^2}{\sum_{i=1}^N y_i^2}, \quad (6.6)$$

For a perfect forecast, $R^2 = 1$. An reference model which always predicts $\bar{y} = 0$ gives $R^2 = 0$.

6.5 Dynamic Time Warping

The statistics mentioned above are all so-called point-by-point metrics. All of them are widely used to compare the amplitudes between observed and predicted time series in a point-by-point manner. The most significant deficiency of this type of statistic is found in their weakness in quantifying time uncertainties [Samara et al., 2022]. A time window metric is also included in the analysis to solve this problem. The chosen time window metric is Dynamic Time Warping (DTW). DTW is a well-known technique for comparing two time series with similar patterns but differing in time. Although it was initially developed for speech recognition purposes, it has been widely applied in other scientific fields over the years. It was shown by Samara et al. [2022] that this technique could be applied successfully to compare solar wind characteristics. It is this procedure that is employed in the analysis of this thesis. A short explanation of how DTW is employed here is provided based on Samara et al. [2022].

The method determines how similar two time series are by optimizing their temporal alignment. DTW is a distance measure but is not a metric, as it violates the triangular inequality. Nevertheless, it can be used as one by obeying the following three principles:

1. The first and last points of both sequences should be matched. (However, these matches are not required to be unique.)
2. The mapping of elements should be monotonically increasing in time.
3. There can be no data gaps; every point must be matched with at least one other point. This is called the continuity rule.

It then optimizes the alignment by finding the path through the DTW cost matrix that minimizes the total cumulative cost. This path is called the warping path. The DTW cost matrix is filled using the following equation:

$$D(i, j) = \delta(s_i, q_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\}, \quad (6.7)$$

where $D(i, j)$ is the cumulative DTW cost and $\delta(s_i, q_j) = |s_i - q_j|$ corresponds to the Euclidean distance between the point s_i from the first time series and the point q_j from the other time series. The first element of the matrix $D(0, 0)$ is equal to $\delta(s_0, q_0)$. After filling the DTW cost matrix, the warping path can be found efficiently.

The last (bottom right) cell in the DTW cost matrix indicates the DTW score, representing the minimum effort the technique puts in to align the time series. To give this score an actual meaning, it must be compared with the ideal scenario (that is, the DTW score = 0) and a non-ideal reference prediction. This reference prediction is defined as the scenario where no geomagnetic activity occurs. This translates to $Dst = 0$ and $Kp = 0$. One advantage of this reference prediction is that it has no variation in time. The sequence similarity factor (SSF) is then defined as:

$$SSF = \frac{DTW_{score}(O, M)}{DTW_{score}(O, \bar{O})}, \quad SSF \in [0, \infty), \quad (6.8)$$

where O , M and $\bar{O}$ stand for Observed, Modeled and Averaged Observed data, respectively. The SSF is equal to zero when a perfect forecast is achieved and equal to one when our forecast is as bad as an average straight-line prediction.

Chapter 7

Results

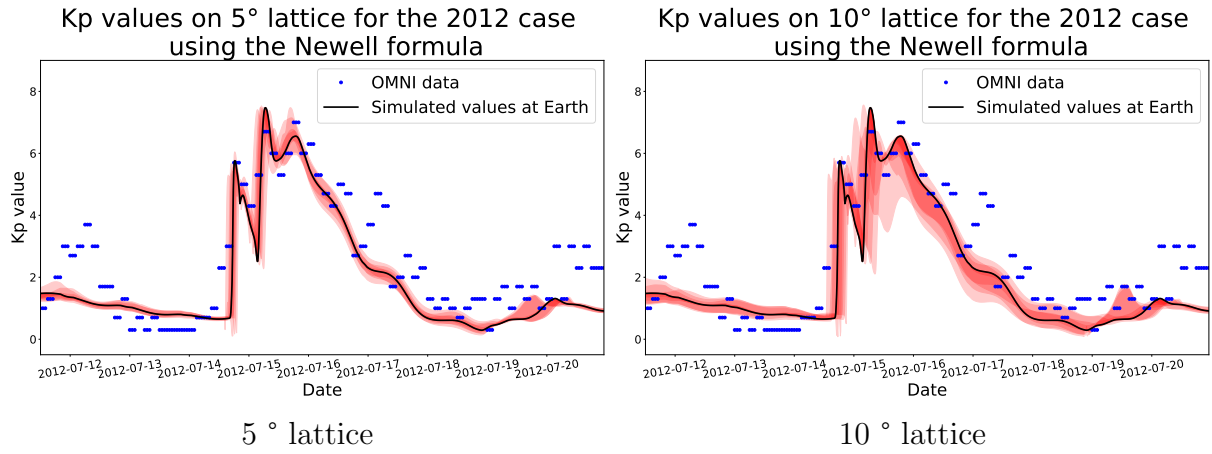
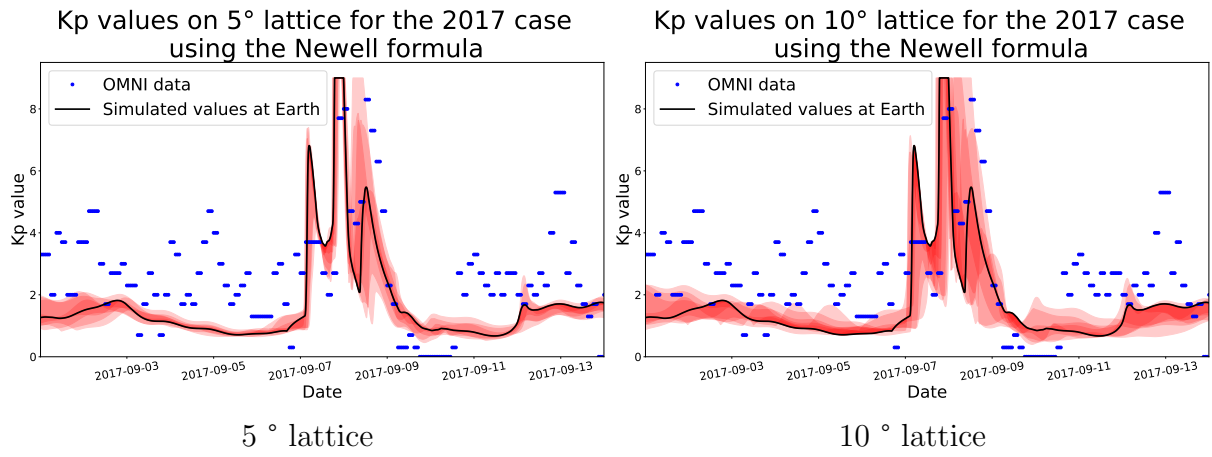
In this chapter, the results of the different prediction models are displayed. It was initially intended to run all models with both the observational and synthetic L1 data. However, only the synthetic data is included in this thesis. Further research is needed to analyze the results obtained from using observational data in prediction models. The synthetic data at Earth and the sixteen virtual satellites that make up the 5 ° and 10 ° lattice were used as input for each model.

7.1 Empirical Models

7.1.1 Newell Kp Model

The Newell Kp model results can be seen in Figures 7.1.1 and 7.1.2. On the left side, for example, in Figure 7.1.1, the results of the virtual satellites on the 5 ° lattice are displayed. The black line shows the simulation results at Earth. The red-shaded region indicates how much the virtual satellites differ from these. The blue dots are the observed Kp values. On the right side, for example, again in Figure 7.1.1, the results of the virtual satellites on the 10 ° lattice are displayed. All other figures in this chapter also follow this framework.

Looking at Figure 7.1.1, one can already notice that the virtual satellites on the 10 ° lattice, in general, deviate more from the simulation at Earth than the 5 ° lattice. The same can be observed in Figure 7.1.2. This is in line with expectations, as one would indeed suspect smaller deviations for virtual satellites closer to Earth and larger deviations for those further away. On Figure 7.1.2 one can also recognise that the results obtained with the EUFHORIA input showcases three peaks, in accordance with this being a thee CME event. Although it may appear in this figure that the peak estimation is not entirely shown, the reader is reminded of the fact that the Kp index can only take values in the range 0 – 9, as was mentioned in Section 1.2, and thus these peak values are capped at the value $Kp = 9$.

Figure 7.1.1: Kp values for the 2012 case using the Newell formula.Figure 7.1.2: Kp values for the 2017 case using the Newell formula.

7.1.2 O'Brien Dst Model

Figures 7.1.3 and 7.1.4 show the results of the O'Brien Dst model. In both figures, it can be observed that this model overestimates the Dst values before the occurrence of the CME shock wave. The same overestimation is again observed after the recovery phase. Where one would expect values around $Dst = 0$ here, this model implementation by O'Brien and McPherron [2000a] predicts a value around $Dst = 50nT$. In 7.1.4, it can be seen that the peak strength for the 2017 test case is overestimated, which may stem from the overestimation of the plasma parameter $Blon$ as can be seen in Figure 5.2.2.

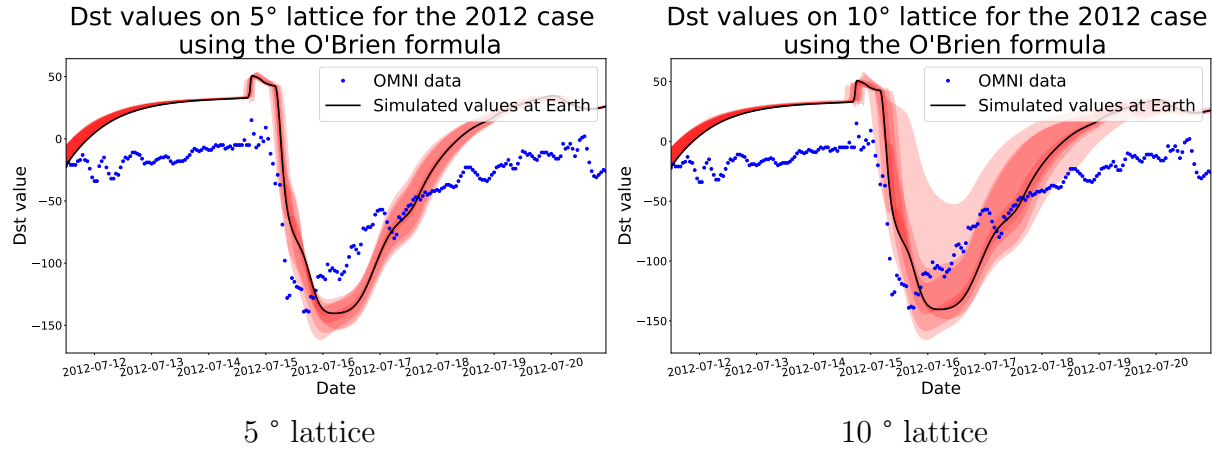


Figure 7.1.3: Dst values for the 2012 case using the O'Brien formula.

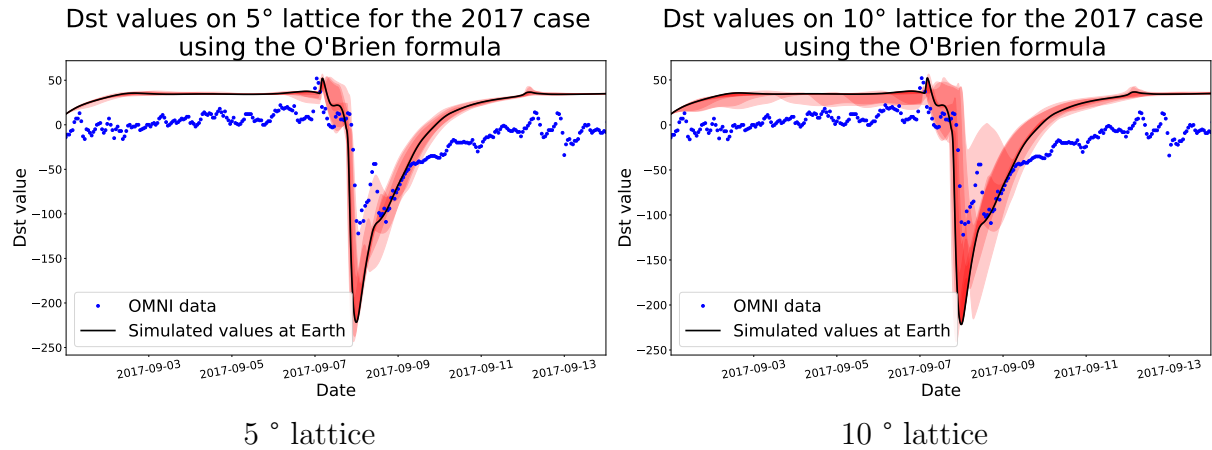


Figure 7.1.4: Dst values for the 2017 case using the O'Brien formula.

7.2 Neural Network Models

7.2.1 FNN Kp Model

The results for the feed-forward Neural Network are displayed in Figures 7.2.1 and 7.2.2. Note that only the results for Kp are present for this model type, as the Dst model was not made available for this study. The Kp predictions in Figures 7.2.1 and 7.2.2 show similar features to those discussed in Subsection 7.1.1. The comparison of these is discussed further in Section 7.4.

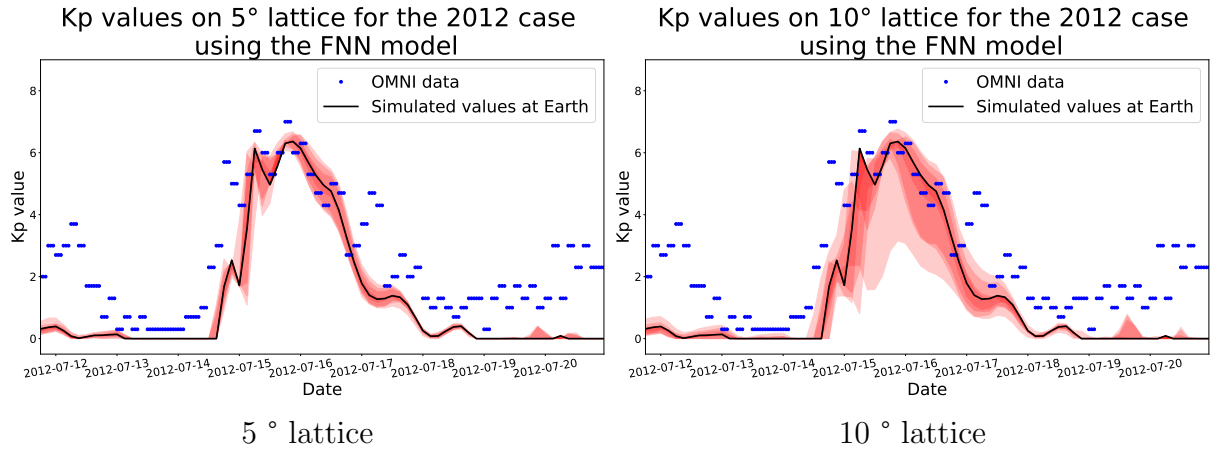


Figure 7.2.1: Kp values for the 2012 case using the FNN model.

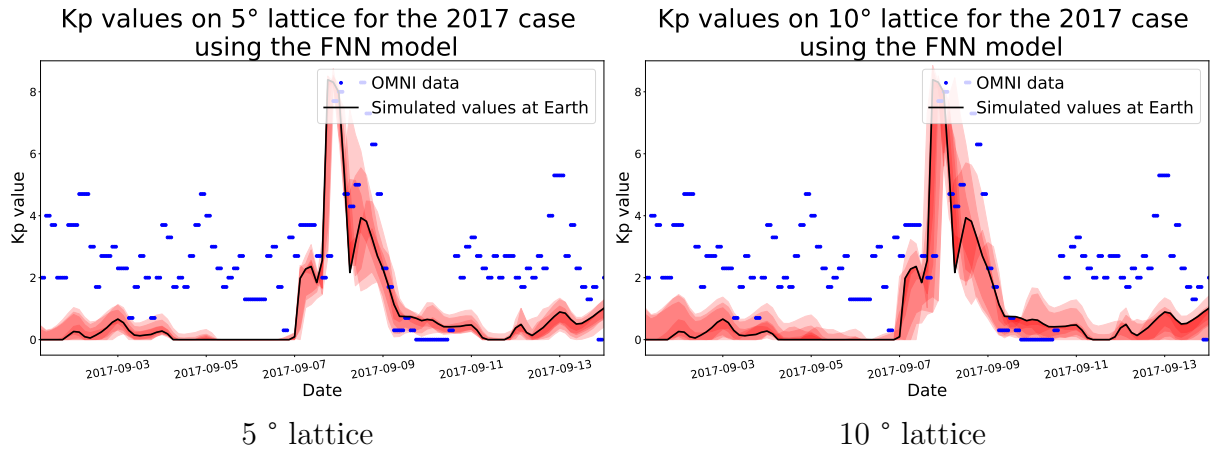


Figure 7.2.2: Kp values for the 2017 case using the FNN model.

7.2.2 SWING Kp Model

The results of the SWING model predictions for the Kp index are shown in Figures 7.2.3 and 7.2.4. Apart from the secondary peak that can be observed in Figure 7.2.3 for a number of virtual satellites on the 10° lattice, these results again show similar features to those of the Newell model discussed in Subsection 7.1.1.

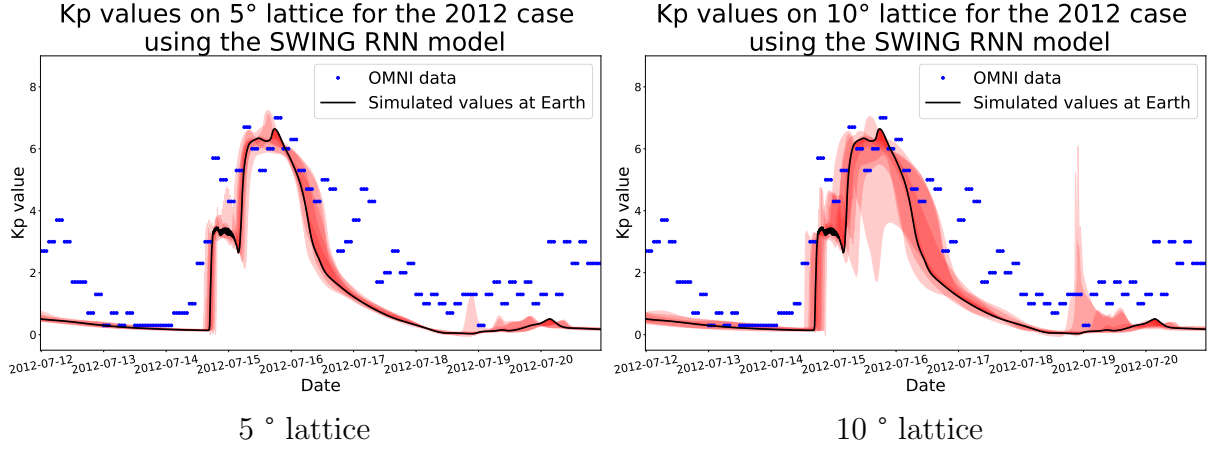


Figure 7.2.3: Kp values for the 2012 case using the SWING model.

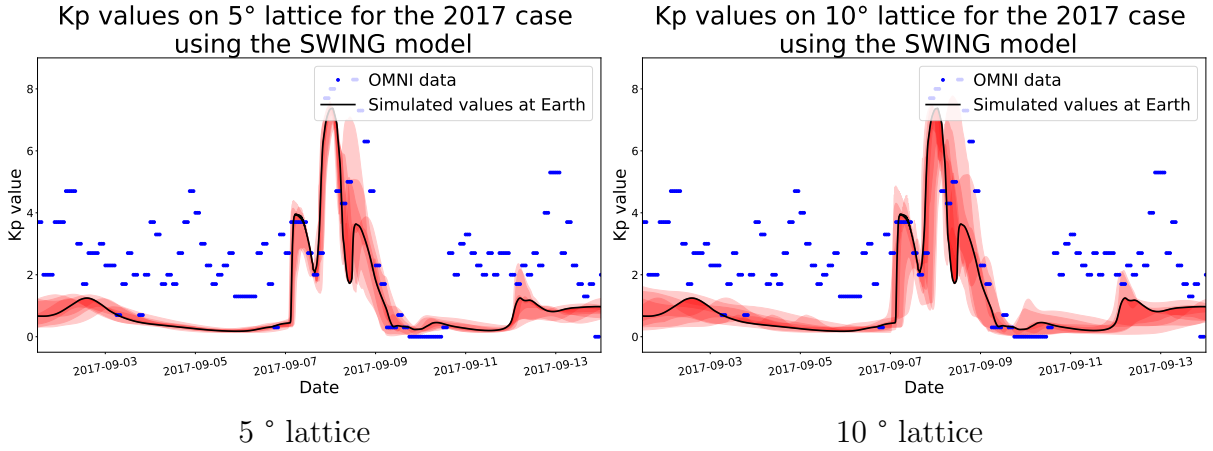


Figure 7.2.4: Kp values for the 2017 case using the SWING model.

7.2.3 SWING Dst Model

The results of the SWING models for the Dst index are shown in Figures 7.2.5 and 7.2.6. In contrast to what was observed for the O'Brien model in Subsection 7.1.2, the Dst predictions made by SWING do not overestimate the values before the shock arrival and after the recovery phase. In Figure 7.2.6, the peak strength is also not as heavily overestimated as in Figure 7.1.4. Similar to what was observed on the 10° lattice for the SWING Kp prediction model in the 2012 test case, a secondary peak can be seen in Figure 7.2.5 for several virtual satellites on the 10° lattice.

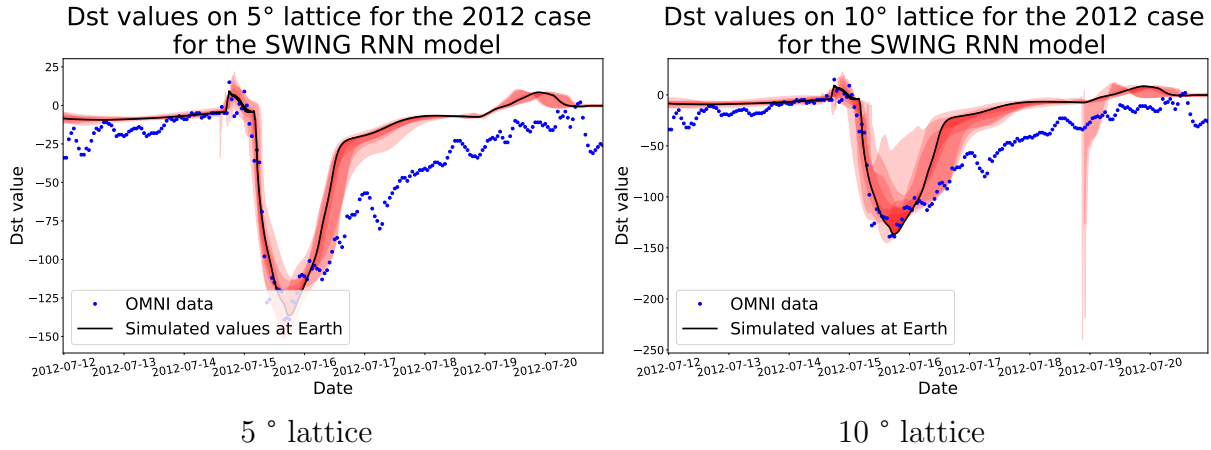


Figure 7.2.5: Dst values for the 2012 case using the SWING model.

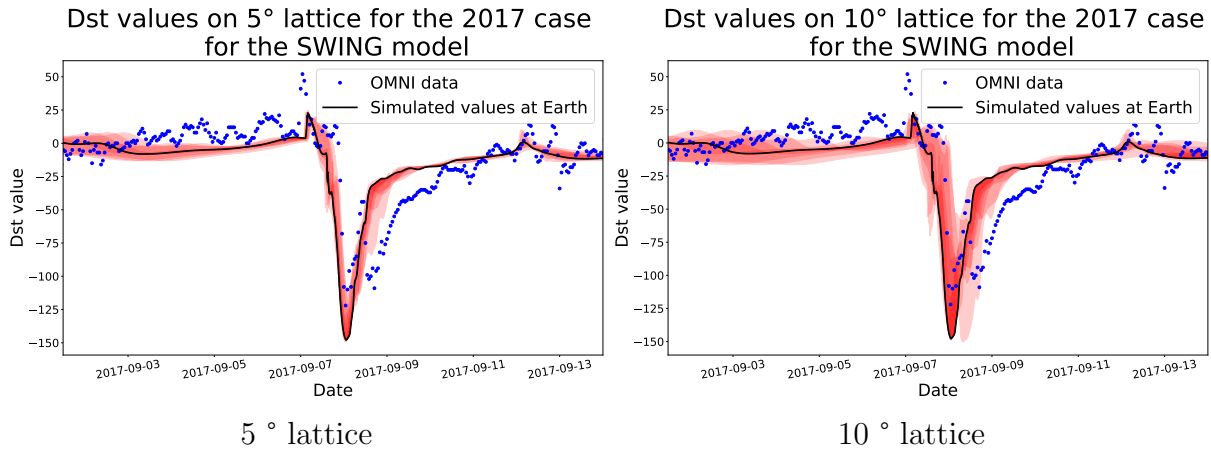


Figure 7.2.6: Dst values for the 2017 case using the SWING model.

7.3 NARMAX Models

7.3.1 SNGI Kp Model

The results of the SNGI model predictions for the Kp index are shown in Figures 7.3.1 and 7.3.2. These results do not display any features that were not yet discussed previously. Their comparison with the results of the other models is left for Section 7.4.

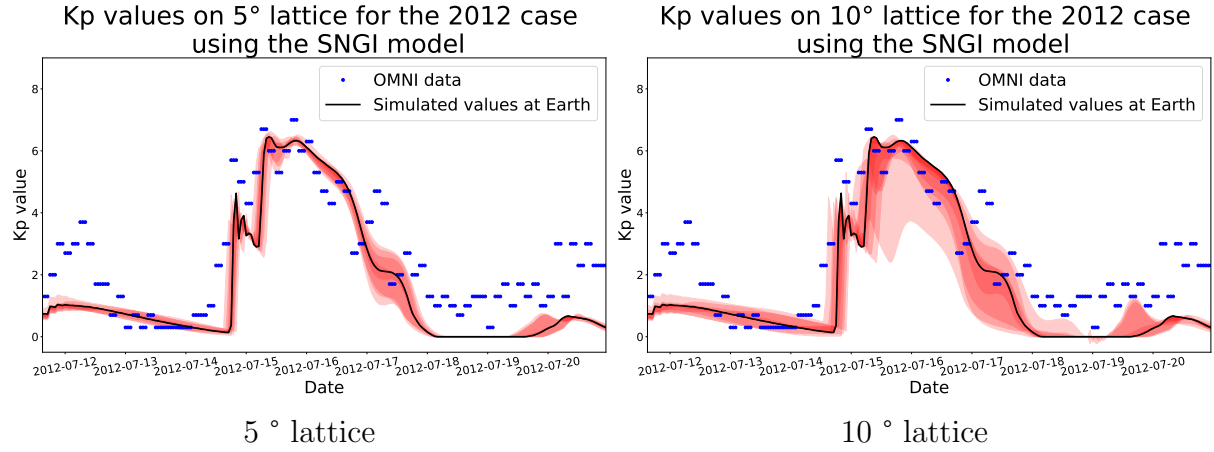


Figure 7.3.1: Kp values for the 2012 case using the SNGI model.

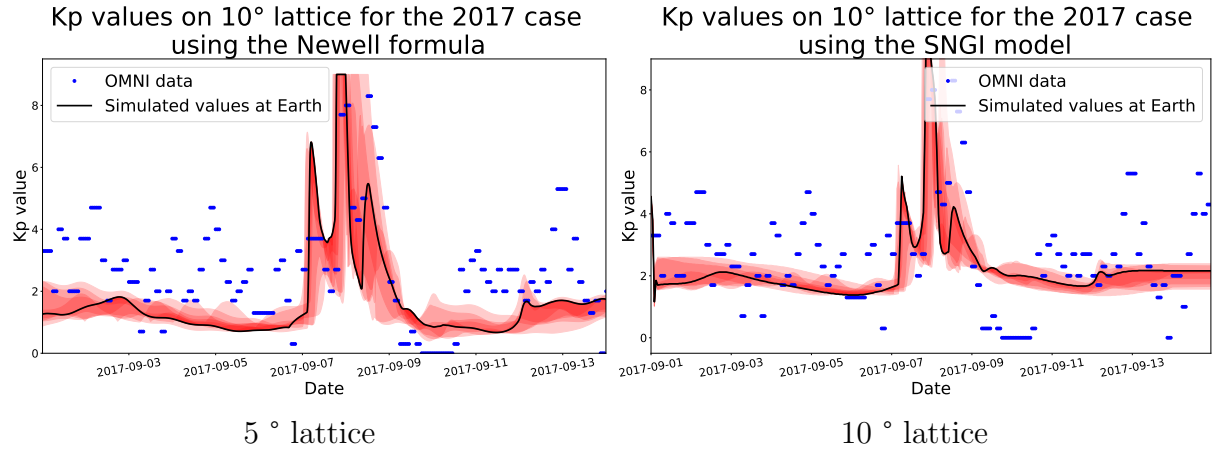


Figure 7.3.2: Kp values for the 2017 case using the SNGI model.

7.3.2 SNGI Dst Model

The results of the SNGI model predictions for the Dst index are shown in Figures 7.3.3 and 7.3.4. Similar to the results of the model by O'Brien and McPherron [2000a], the results in Figure 7.3.3 show an overestimation of the Dst values before the CME arrival and after the recovery time, although not as gravely as in Figure 7.1.3. The results shown in Figure 7.3.4 also show an overestimation of the peak strength similar to that in Figure 7.1.4.

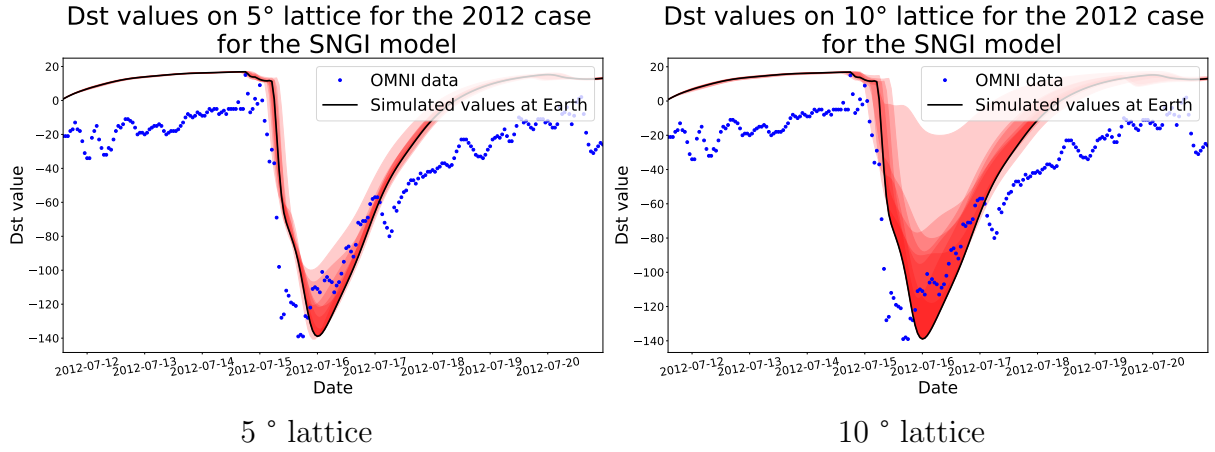


Figure 7.3.3: Dst values for the 2012 case using the SNGI model.

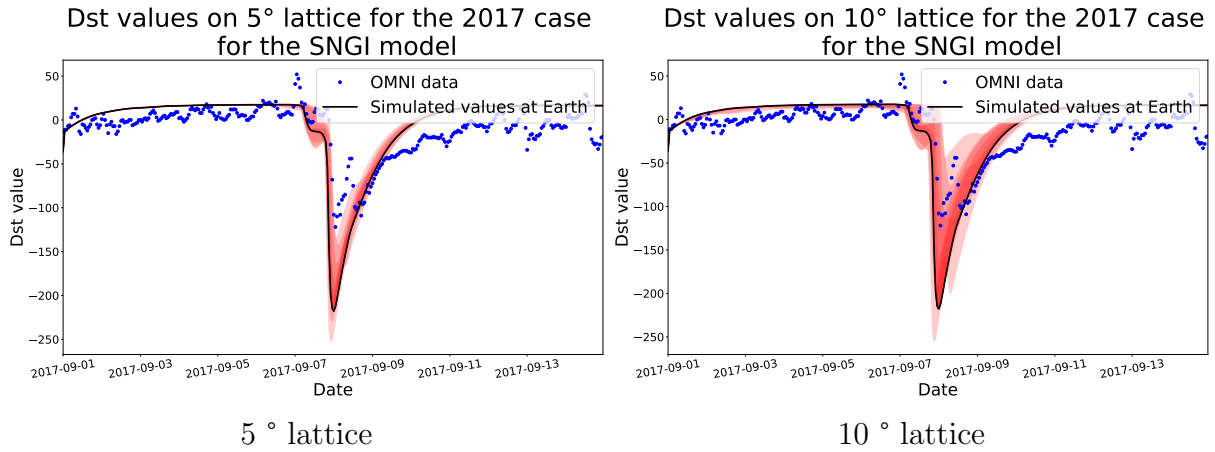


Figure 7.3.4: Dst values for the 2017 case using the SNGI model.

7.4 Comparison

7.4.1 Kp Models

To allow for easier comparison between the different models, the Kp results for the EUHFORIA simulation at Earth are summarized together with the observed Kp values in Figure 7.4.1 for the 2012 test case and in Figure 7.4.2 for the 2017 test case.

Looking at Figure 7.4.1, one can see that all four models have very similar predictions of the peak intensity for the 2012 test case. Figure 7.4.2 shows more variation in the prediction in peak strength for the 2017 test case, although all models nicely showcase the three peaks associated with this being a three CME event. Here, one can also see a clear difference between the models in predicting the Kp values before and after the geomagnetic storm associated with this CME.

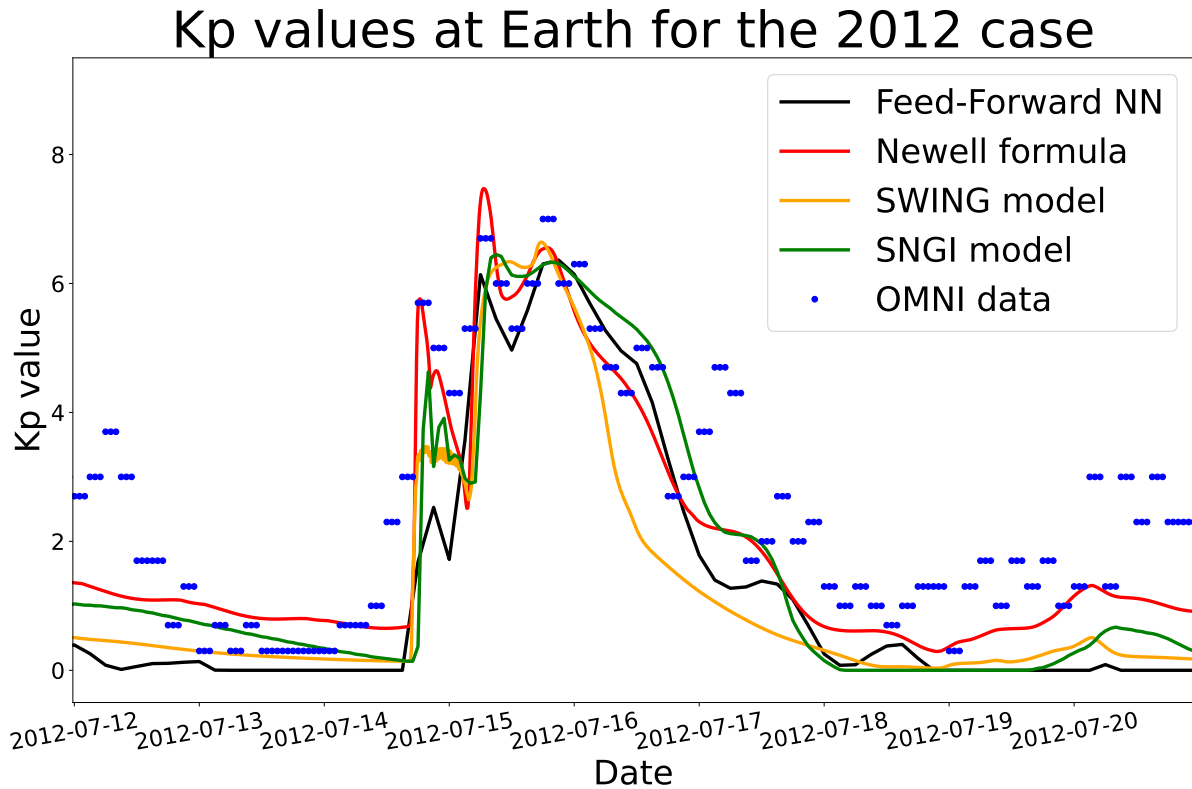


Figure 7.4.1: Kp values for the EUHFORIA simulation at Earth for the 2012 test case.

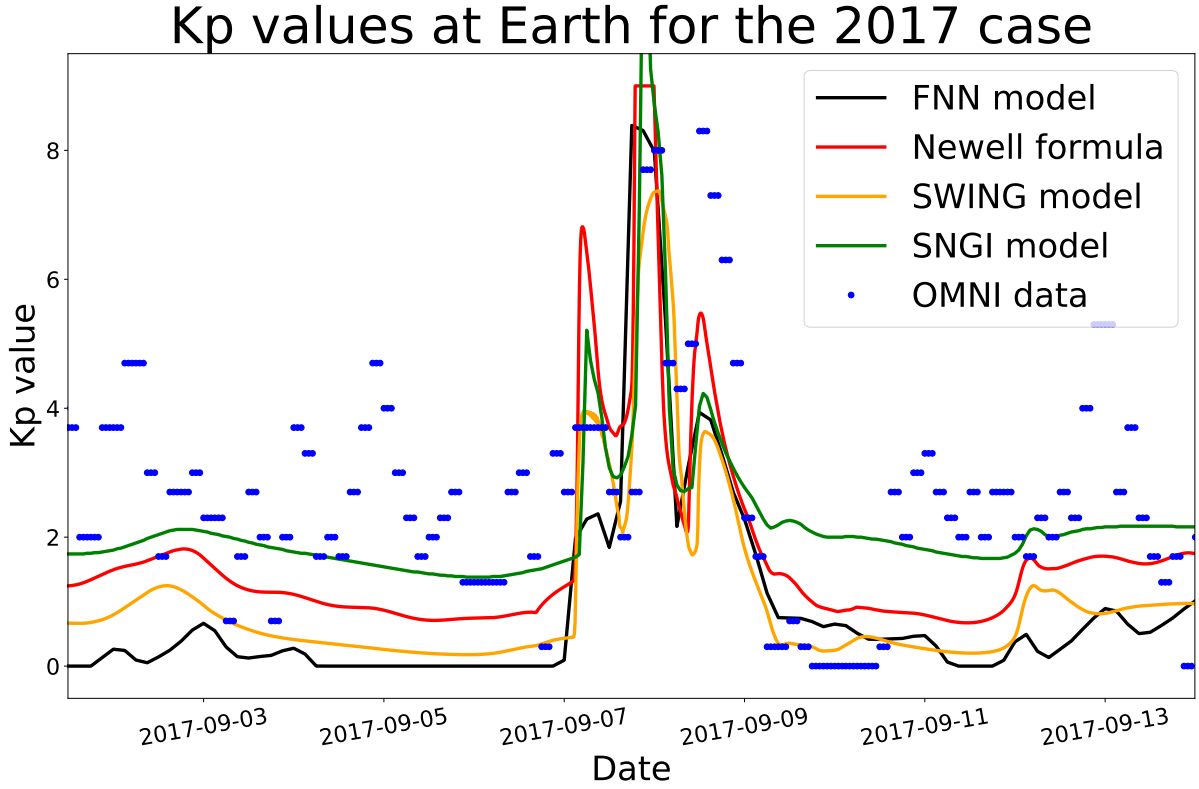


Figure 7.4.2: K_p values for the EUHFORIA simulation at Earth for the 2017 test case.

7.4.2 Dst Models

The Dst results for the EUHFORIA simulation at Earth are summarized together with the observed Dst values in Figure 7.4.3 for the 2012 test case and in Figure 7.4.4 for the 2017 test case.

In Figure 7.4.3, one can clearly observe the difference between the different models of the predicted Dst values before the occurrence of the CME arrival and after the recovery phase. The model by O'Brien and McPherron [2000a] strongly overestimates these values, the SNGI model also overestimates these, and the SWING model makes the most accurate prediction. Figure 7.4.3 also shows that the three models predict the peak strength quite well.

On the contrary, Figure 7.4.4 shows that both the O'Brien and the SNGI models overestimate the peak strength, while the SWING model performs better. Here, one can also observe the same prediction behavior for the Dst values before the occurrence of the CME arrival and after the recovery phase, as was already discussed for the 2012 test case.

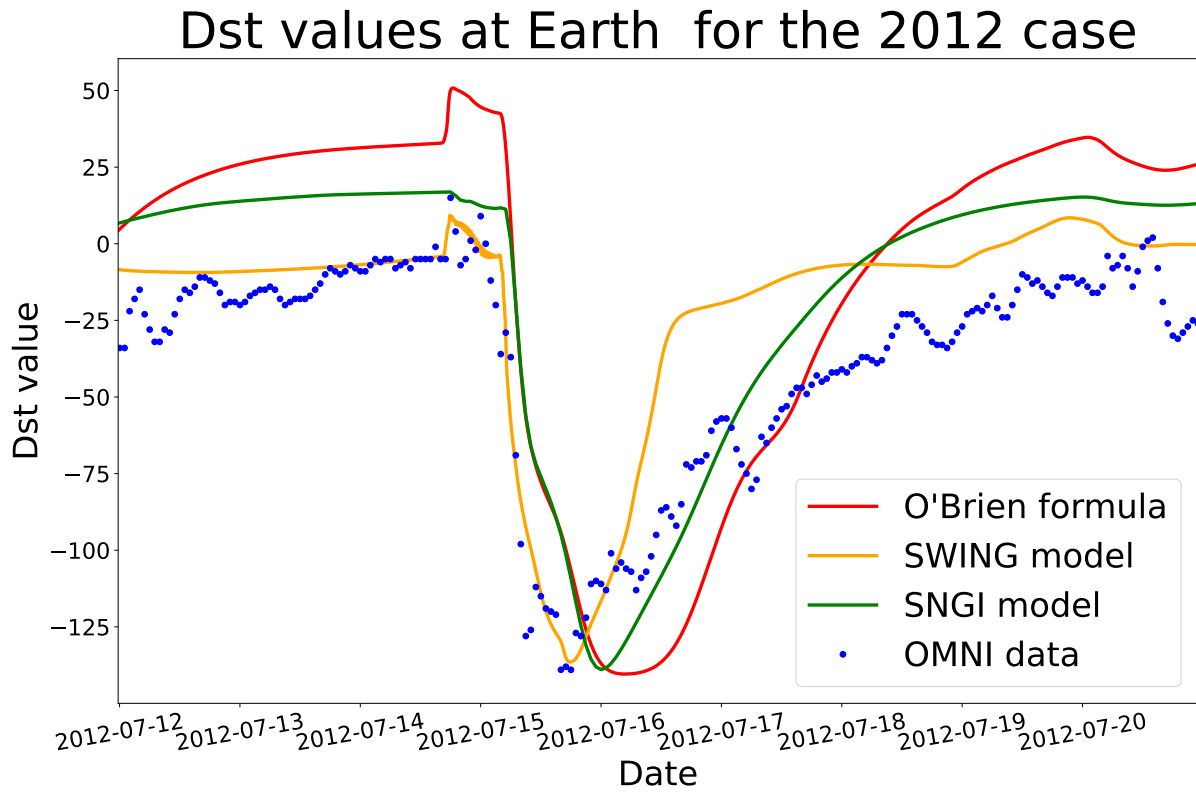


Figure 7.4.3: *Dst* values for the EUHFORIA simulation at Earth for the 2012 test case.

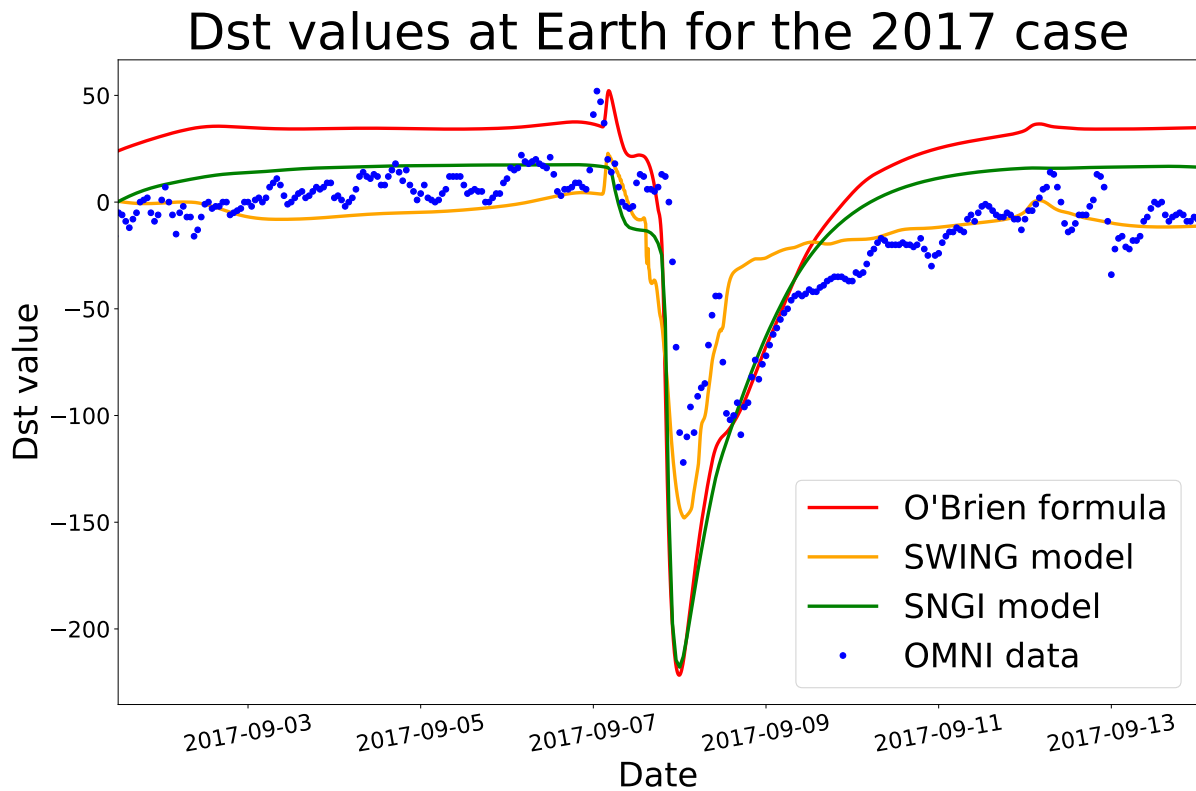


Figure 7.4.4: *Dst* values for the EUHFORIA simulation at Earth for the 2017 test case.

Chapter 8

Analysis

In this chapter, the metrics described in Chapter 6 are applied to the results obtained in Chapter 7. The results are summarized in Figures 8.1.1 to 8.5.4. These figures display the obtained results for the EUHFORIA simulation at Earth using coloured, dashed lines. The box plots show the spread of the values obtained for the virtual satellites, subdivided into one box plot showing the results on the 5 ° lattice and another box plot showing the results on the 10 ° lattice.

For each metric, the results for a reference model were also calculated. The reference model for the Kp values is defined as $Kp = 0$ and for the Dst values as $Dst = 0$. The results for these reference models are displayed using the black dash-dotted line. The Pearson correlation coefficient is an exception to this reference model definition. As can be seen in Equation 6.4, the reference models $Kp = 0$ and $Dst = 0$ give rise to the indeterminate form $\frac{0}{0}$. The reference Pearson correlation coefficient is therefore defined as $CORR = 0$.

To allow for easier comparison of the different models, the obtained results for the EUHFORIA simulation at Earth, which in the figures are displayed using dashed lines, are also summarized in Tables 8.1 to 8.4.

8.1 Bias

In this section, the bias, as defined in Equation 6.1, is calculated for the predictions made by the model. The bias indicates whether a model's predictions tend to be over- or underestimated and by what amount. A perfect forecast has a bias equal to zero.

8.1.1 Kp Models

In Figure 8.1.1, it can be seen that for the 2012 test case, all four models obtain a bias in the range $[-2, 0]$. Since the reference model obtains a bias of about 2.5, it can be concluded that all four models perform reasonably well compared to the reference model. In this test case, the Newell model generally has the highest bias (lowest absolute value, as they are all negative), followed by the SNGI model. The simulations of the feed-forward

neural network and SWING have very similar values at Earth. In 8.1, it can be seen that the feed-forward neural network performs a little better here. However, it is also a prediction made by the forward neural network for a virtual satellite simulation on the 10° lattice that has the lowest bias (highest in absolute value).

In case of the 2017 test case, Figure 8.1.2 shows that all four prediction models obtain a bias in the range $[0, 2.1]$ while the reference model obtains a bias of about 2.8. Thus, it can also be concluded that all four models perform reasonably well compared to the reference model in this test case. Figure 8.1.2 also shows that the SNGI model has the highest bias (lowest in absolute value), followed by the Newell model. The feed-forward neural network and the SWING model again obtain a similar bias, and Table 8.2 shows that for the 2017 test case, SWING obtains a higher bias (lower in absolute value) for the EUHFORIA simulation at Earth.

The resulting biases seen in Figures 8.1.1 and 8.1.2 remind of the Kp predictions before and after the geomagnetic storm associated with this CME as was seen in Figures 7.4.1 and 7.4.2. From this observation, it could be derived that the bias might be an appropriate metric to measure the performance of the prediction models in quiet periods. However, a more detailed analysis is required before such a conclusion can be made.

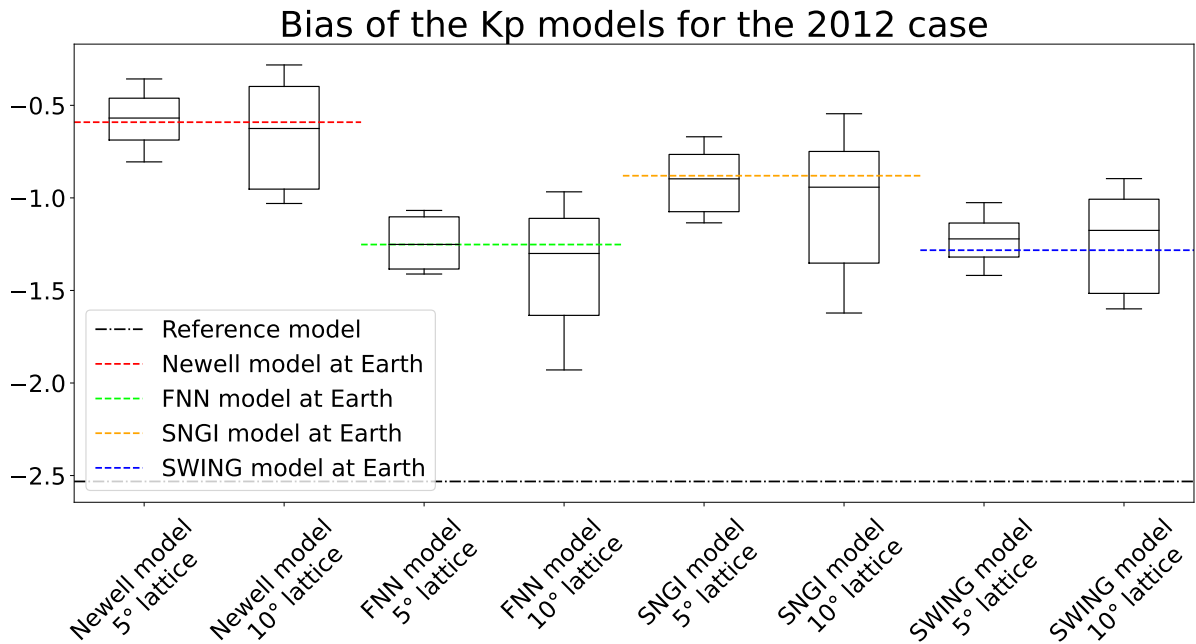


Figure 8.1.1: Bias of the Kp models for the 2012 test case.

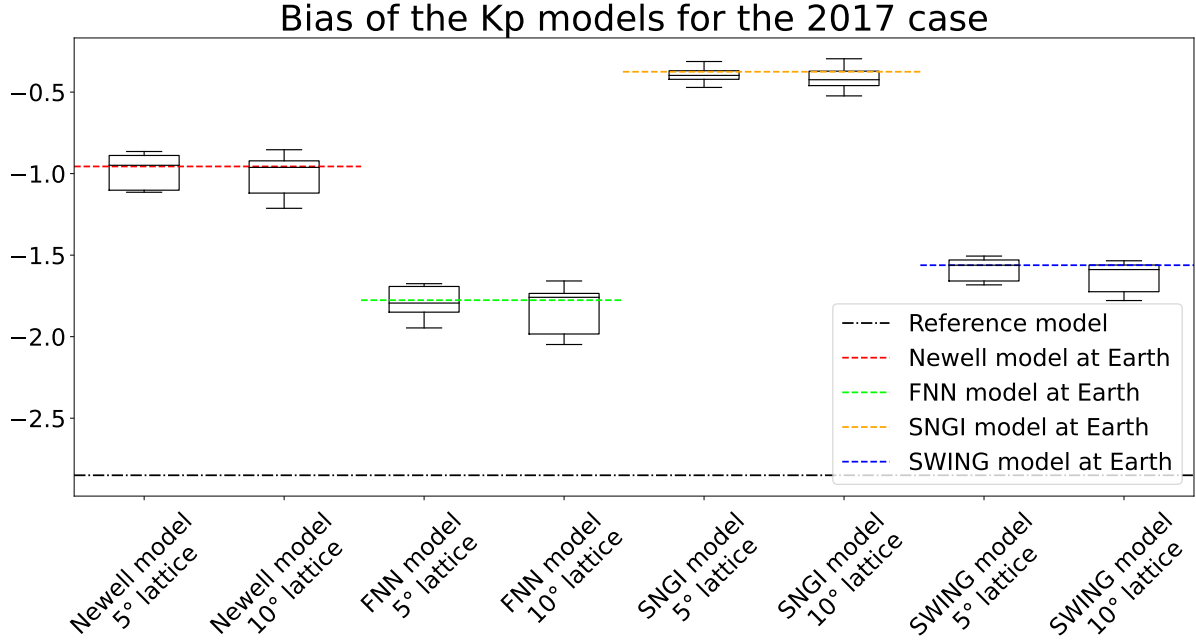
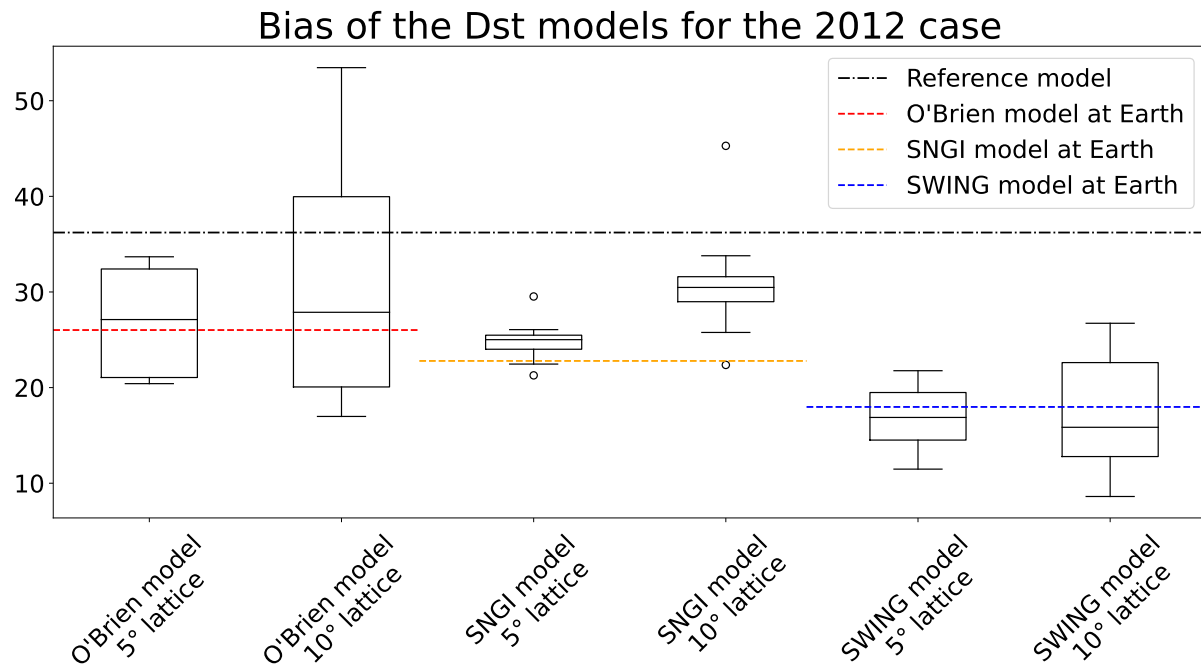
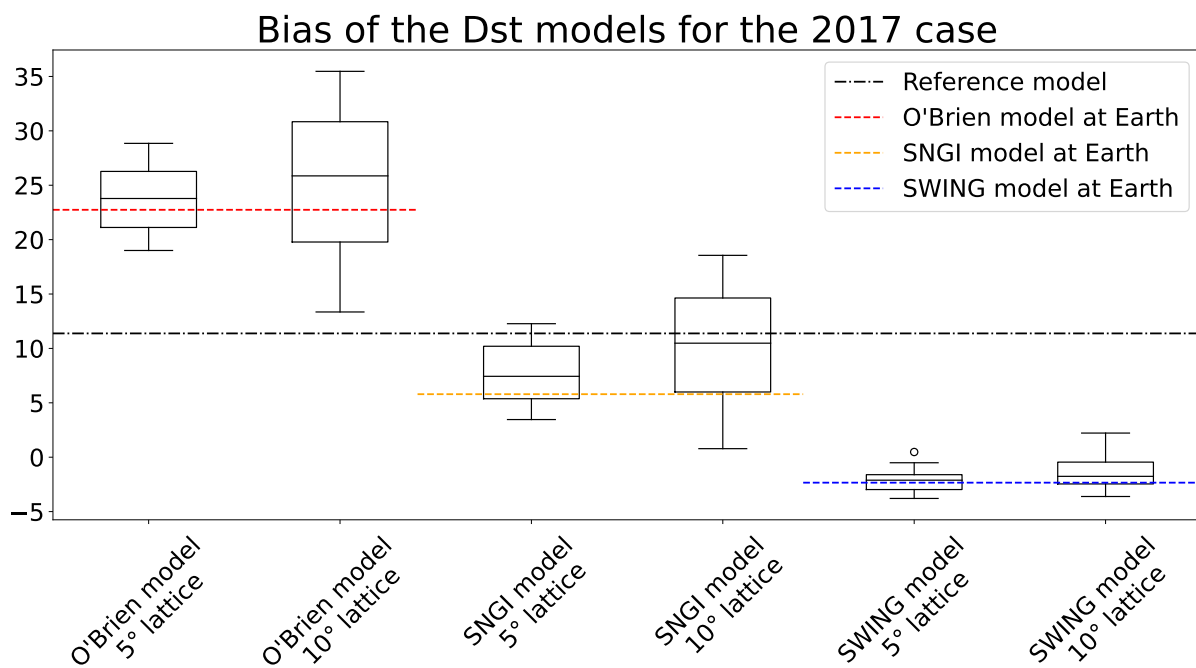


Figure 8.1.2: Bias of the Kp models for the 2017 test case.

8.1.2 Dst Models

Figures 8.1.3 and 8.1.4 show that in both test cases, only the SWING model can consistently perform better than the reference model. The SNGI model can perform better than the reference model when the EUHFORIA simulation at Earth is used as input but performs worse for one (for the 2012 test case) or more (for the 2017 test case) of the investigated virtual satellites. The O'Brien model performs the worst. As can be seen in Figure 8.1.4, not a single prediction made using this model for the 2017 test case can score better than the reference model.

The results of the bias for the Dst prediction models again remind us of the observations that were made about Figures 7.4.3 and 7.4.4 about the predicted values before the occurrence of the CME arrival and after the recovery phase.

Figure 8.1.3: Bias of the *Dst* models for the 2012 test case.Figure 8.1.4: Bias of the *Dst* models for the 2017 test case.

8.2 Root Mean Square Error

In this section, the root mean square error, as was defined in Equation 6.2, is calculated for the predictions made by the model. It is a frequently used and perhaps the most famous method to compare predicted and observed time series. A perfect forecast has a root mean square error equal to zero.

8.2.1 Kp Models

Looking at Figures 8.2.1 and 8.2.2 reveals that all Kp prediction models perform better than the reference model for every performed simulation. In general, the Newell model and SNGI model obtain a lower root mean square error than the FNN model and SWING model.

A distinction between the Newell model and the SNGI model can not be made, as the Newell model obtains a lower root mean square error in the 2012 case, while the SNGI model has the lowest root mean square error in the 2012 case. The feed-forward neural network and the SWING model obtain an almost identical root mean square error in the 2012 case for the EUHFORIA simulation at Earth, as seen in Table 8.1. In the 2017 test case, the SWING model obtains a lower root mean square error than the feed-forward neural network.

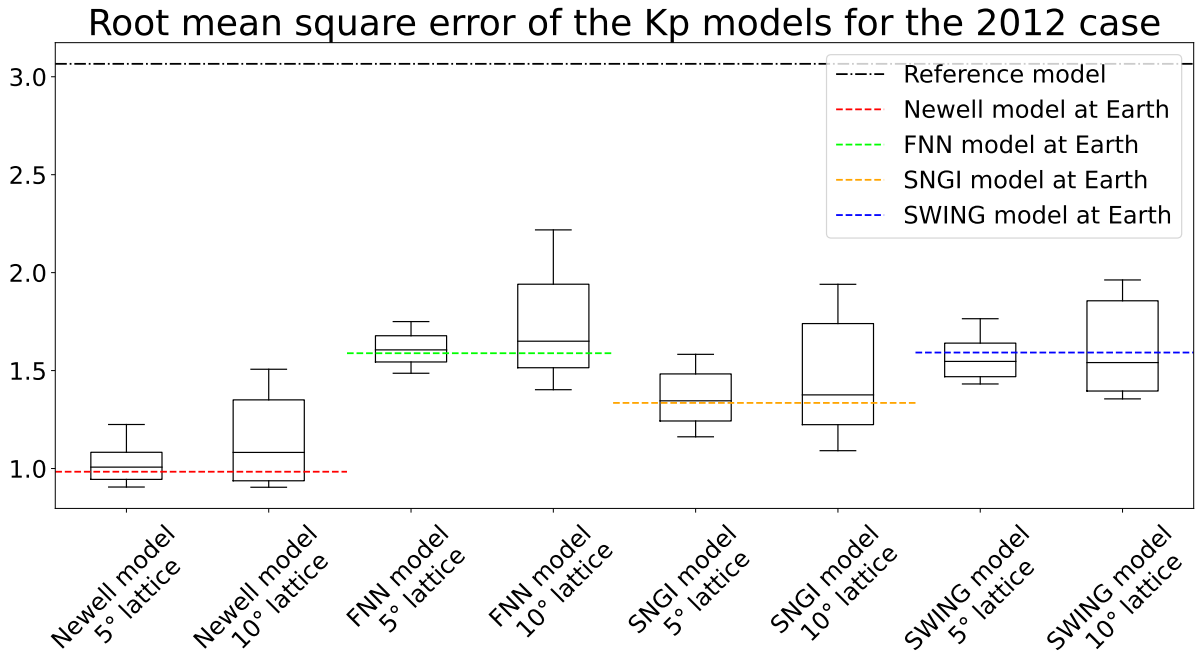


Figure 8.2.1: Root mean square error of the Kp models for the 2012 test case.

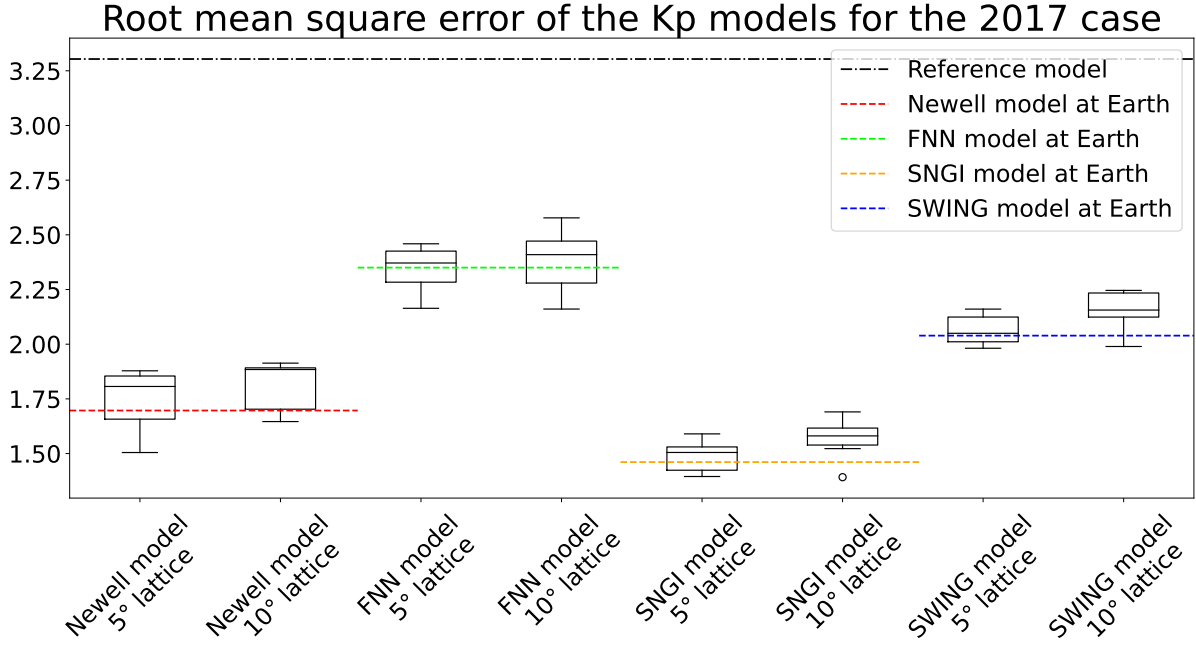


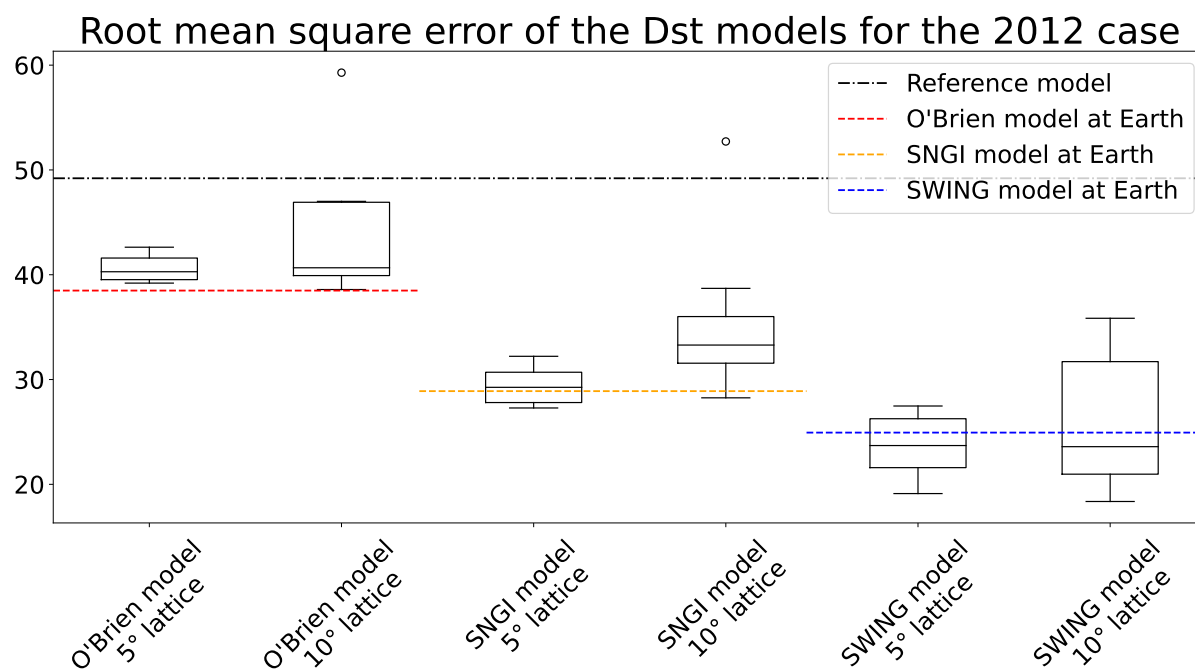
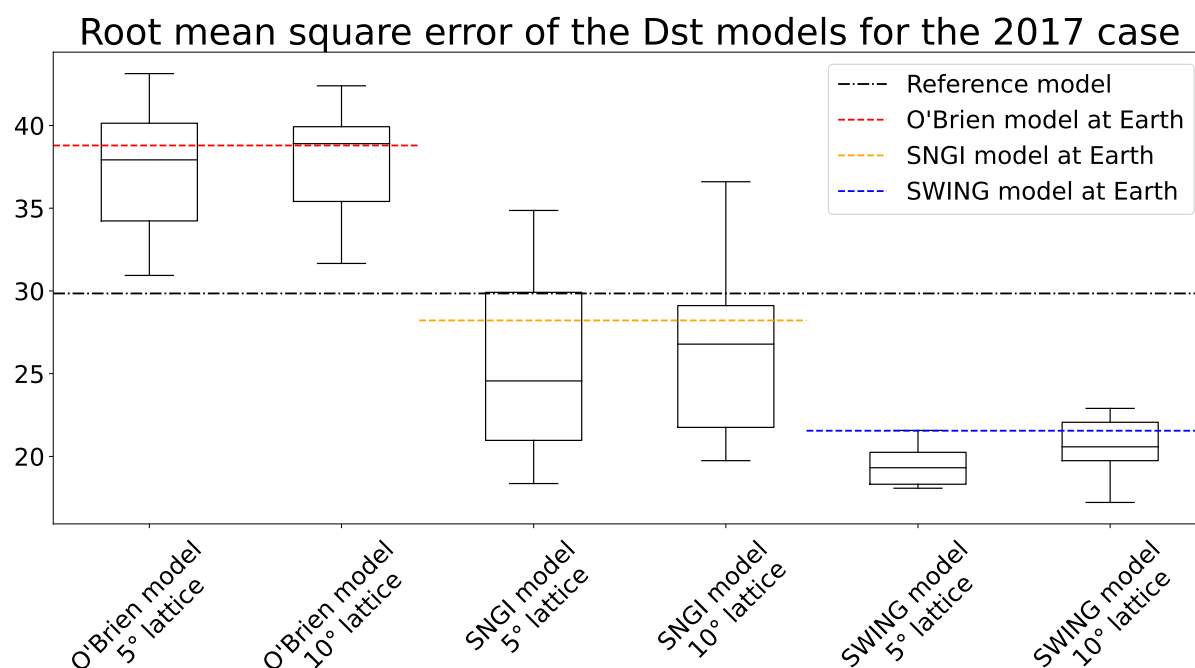
Figure 8.2.2: Root mean square error of the Kp models for the 2017 test case.

8.2.2 Dst Models

Looking at Figures 8.2.3 and 8.2.4 reveals a clear distinction in the root mean square error obtained by the models. It is clear that only the SWING model consistently obtains a lower root mean square error than the reference model. Generally, it also scores lower than the SNGI and the O'Brien models. These figures also show that in general, the SNGI model is also able to obtain a lower root mean square error than the O'Brien model.

For the 2012 test case, both the O'Brien model and the SNGI model can obtain a lower root mean square error than the reference model, except for a single outlier on the 10 ° lattice for both models.

The same does not hold for the 2017 test case, as about a quarter of the virtual satellites on both the 5 ° and 10 ° obtain higher root mean square error than the reference model when looking at the SNGI model. The O'Brien model performs even worse for the 2017 test case, as not a single simulation has a lower root mean square error than the reference model.

Figure 8.2.3: Root mean square error of the *Dst* models for the 2012 test case.Figure 8.2.4: Root mean square error of the *Dst* models for the 2017 test case.

8.3 Pearson Correlation Coefficient

In this section, the Pearson correlation coefficient, as was defined in Equation 6.4, is calculated for the predictions made by the model. The Pearson correlation coefficient is, in essence, a normalized measure of the covariance between two time series. It can thus only obtain a result between -1 and 1 . A perfect forecast has a Pearson correlation coefficient equal to one.

8.3.1 Kp Models

Looking at Figure 8.3.1 it is clear that for the 2012 case all models have Pearson correlation coefficient close to each other and close to unity. This can also be seen in Table 8.1, which shows that The Pearson correlation coefficients for the EUHFORIA simulation at Earth lie in the range $[0.86, 0.96]$. Figure 8.3.2 reveals that the Pearson correlation coefficients are generally lower for the 2017 test case. They still, however, perform better than the reference model.

Looking at both Figure 8.3.1 and 8.3.2 shows that in general, the Newell model and SNGI model perform slightly better than the SWING model and the feed-forward neural network model and that the SWING model generally performs better than the feed-forward neural network model. A distinction can not be made between the Newell and SNGI model since the Newell model obtains higher Pearson correlation coefficient in the 2012 test case while the SNGI model obtains a higher Pearson correlation coefficient in the 2017 test case.

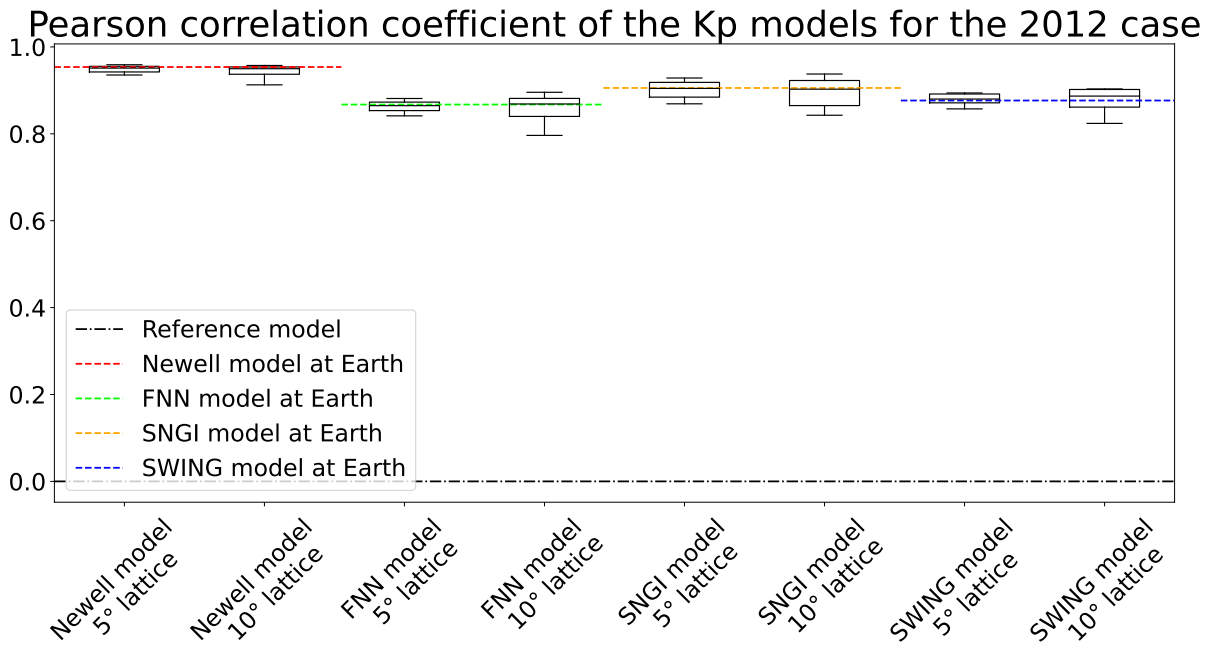


Figure 8.3.1: Pearson correlation coefficient of the Kp models for the 2012 test case.

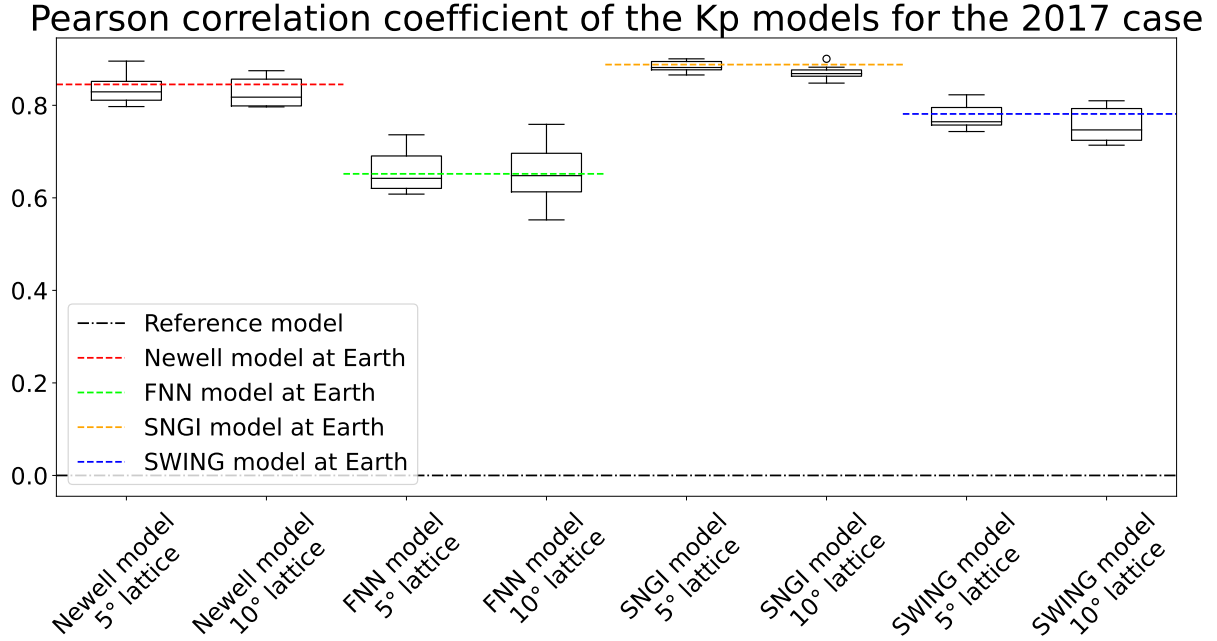
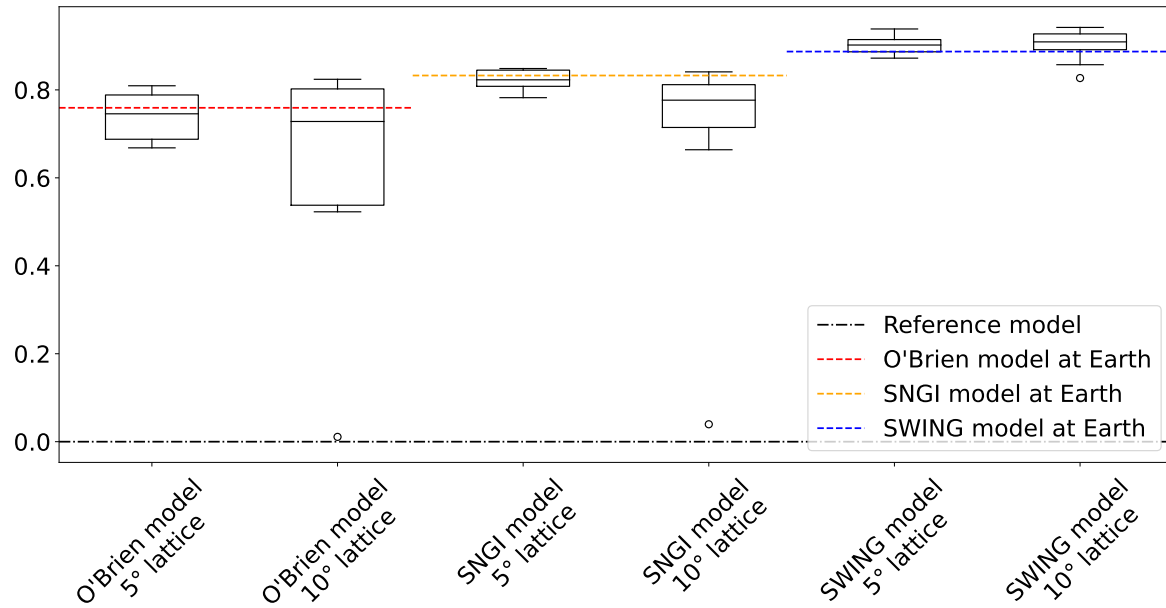
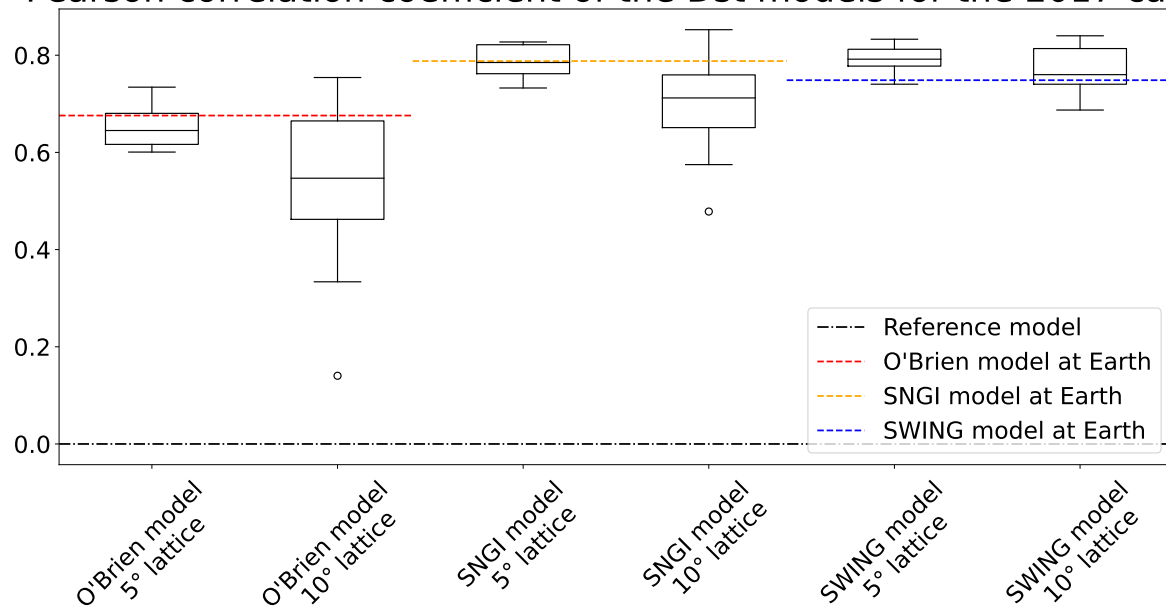


Figure 8.3.2: Pearson correlation coefficient of the Kp models for the 2017 test case.

8.3.2 Dst Models

Looking at Figure 8.3.3, one can see that all models obtain high Pearson correlation coefficients compared to the reference model, apart from one outlier on the 10 ° lattice for the O'Brien model and one outlier on the 10 ° lattice for the SNGI model. The SWING model does not display such an outlier and performs better than the O'Brien and SNGI models. The results of the O'Brien and SNGI model overlap quite a bit, but the O'Brien model appears to be a bit lower in general and shows a more extensive spread downwards than the SNGI model.

Figure 8.3.4 shows that for the 2017 case, the SWING model does not outperform the other two but that, this time, the SNGI model shows comparative results. The SNGI model obtains a higher Pearson correlation coefficient for the EUHFORIA simulation at Earth. Still, the results for the virtual satellites appear to be more spread and generally a bit lower than for the SWING model. The results of the O'Brien model for the 2017 test case appears to perform worse again, but it can still obtain quite decent results compared to the reference model.

Pearson correlation coefficient of the *Dst* models for the 2012 caseFigure 8.3.3: Pearson correlation coefficient of the *Dst* models for the 2012 test case.Pearson correlation coefficient of the *Dst* models for the 2017 caseFigure 8.3.4: Pearson correlation coefficient of the *Dst* models for the 2017 test case.

8.4 Coefficient of Determination

In this section, the coefficient of determination, as was defined in Equation 6.6, is calculated for the predictions made by the model. The coefficient of determination measures the sum of the squared residuals (just like the root mean square error) concerning the sum of the squared residuals obtained for a simple average estimate. A perfect forecast has a coefficient of determination equal to one. An average model, which always predicts the reference value of the observations, has a coefficient of determination equal to zero.

8.4.1 Kp Models

Looking at Figures 8.4.1 and 8.4.2 again, the same pattern arises as was observed for the previous metrics: all models obtain higher coefficients of determination than the reference model for both test cases. In general, the feed-forward neural network obtains the lowest values, followed by the SWING model. The Newell model performs best for the 2012 test case, while the SNGI model performs best for the 2017 test case.

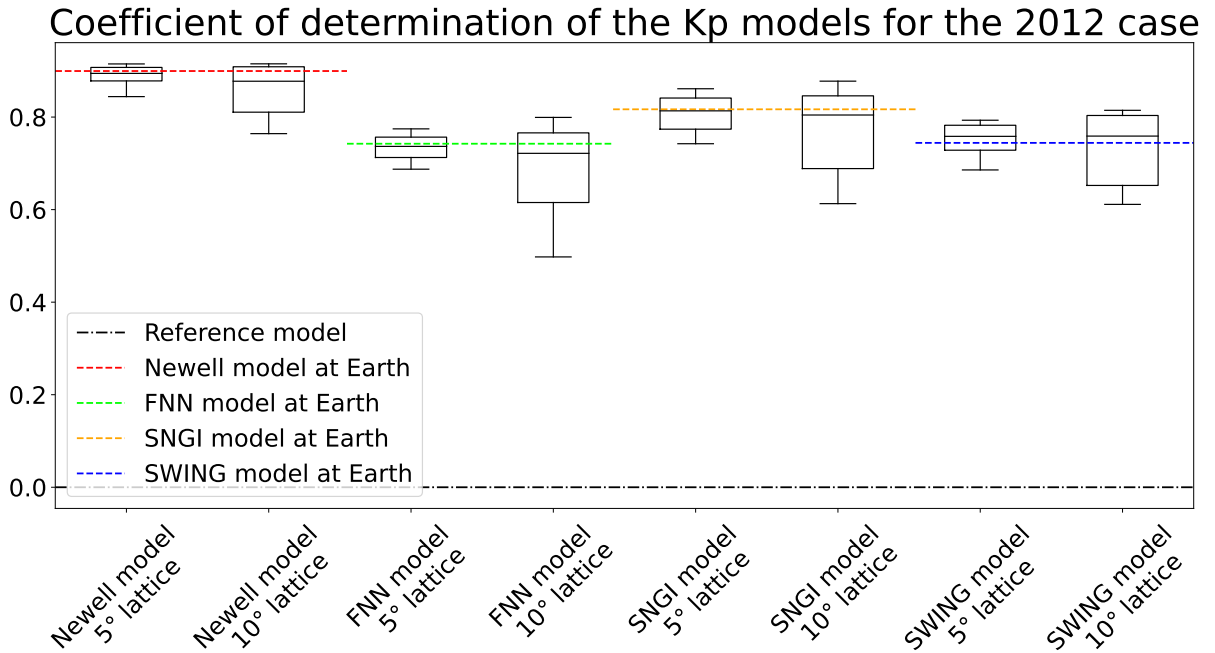


Figure 8.4.1: Coefficient of determination of the Kp models for the 2012 test case.

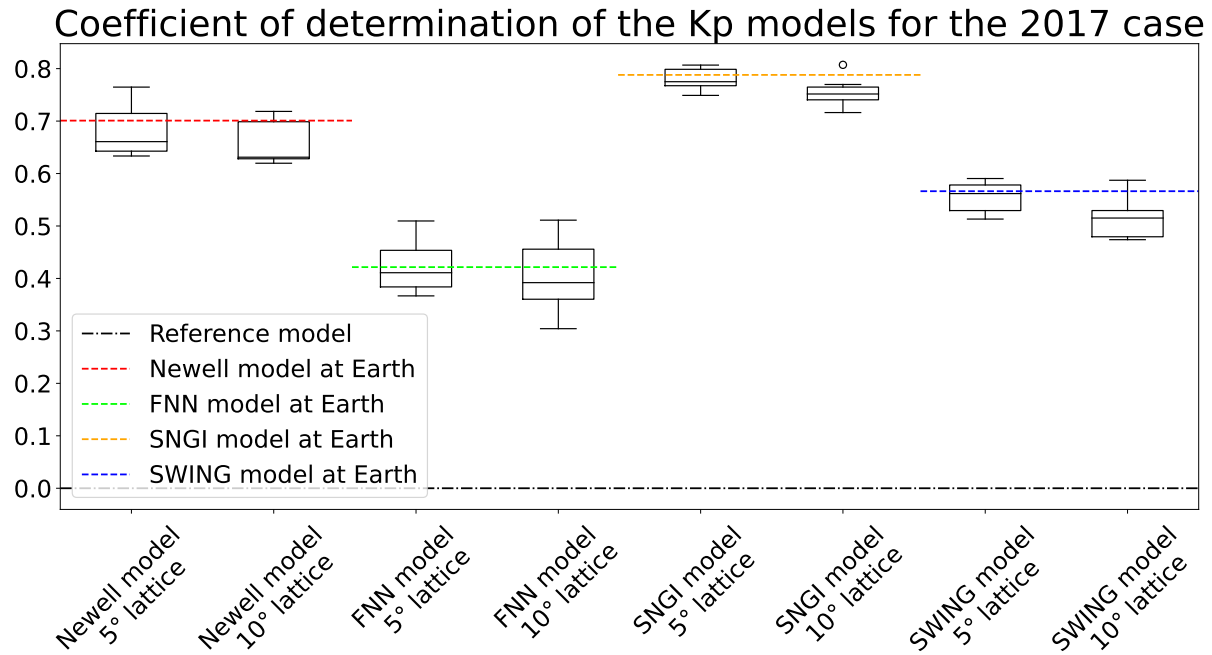
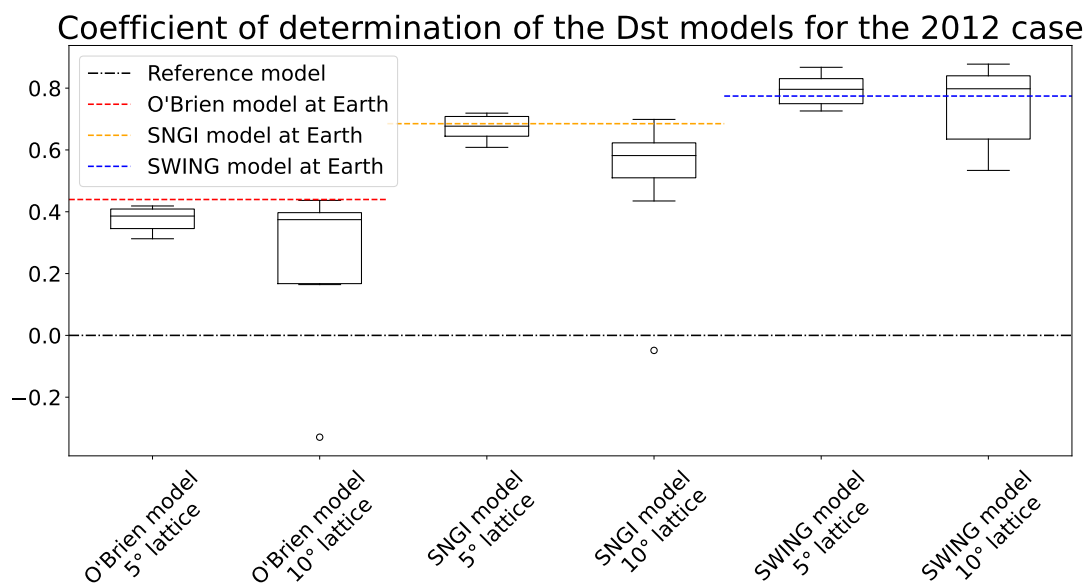
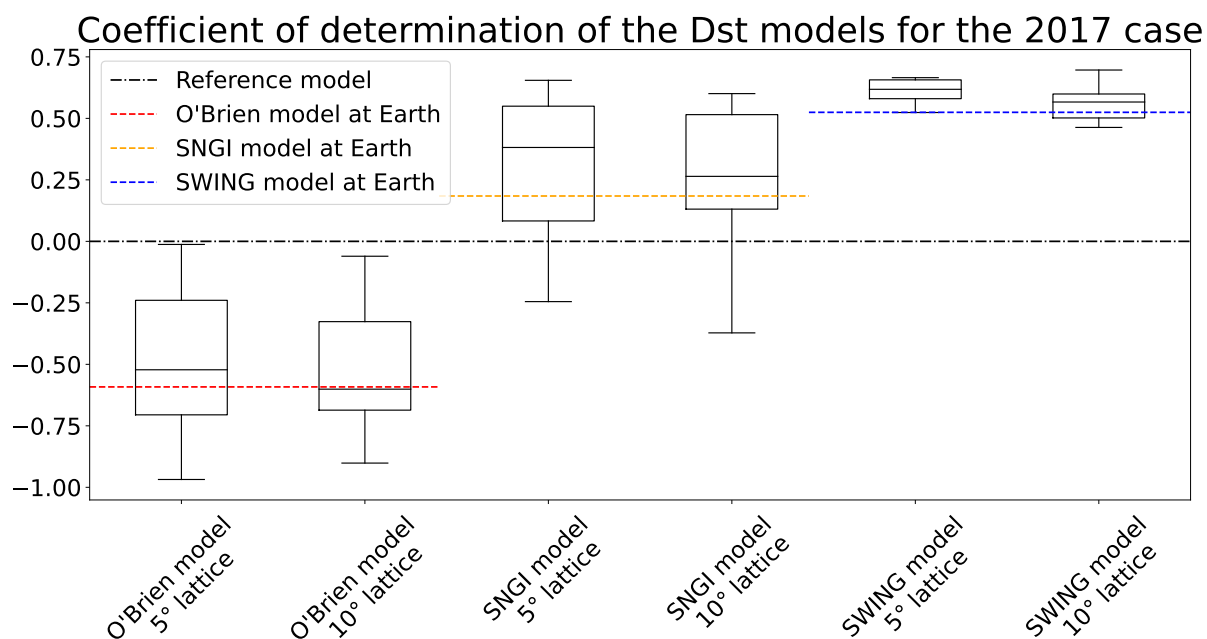


Figure 8.4.2: Coefficient of determination of the Kp models for the 2017 test case.

8.4.2 Dst Models

Looking at Figures 8.4.3 and 8.4.4 again, the same distinction can be made based on the coefficients of determination: the SWING model obtains, in general, the highest coefficient, followed by the SNGI model. The O'Brien model performs the worst, again not being able to score better than the reference model in the 2017 test case for any of its simulations. For the 2012 case, one can again observe one outlier on the 10° latitude for the O'Brien model and one outlier on the 10° latitude for the SNGI model that obtains a lower coefficient of determination than the reference model.

Figure 8.4.3: Coefficient of determination of the *Dst* models for the 2012 test case.Figure 8.4.4: Coefficient of determination of the *Dst* models for the 2017 test case.

8.5 SSF Score

In this section, the sequence similarity factor, as was defined in Equation 6.8, is calculated for the predictions made by the models. In contrast with the other metrics, which are all point-by-point metrics, the SSF is a time window metric. For a perfect forecast, the sequence similarity factor is equal to zero. The SSF is normalized by the reference models $Kp = 0$ or $Dst = 0$, so the self-similarity factor is equal to one for these reference models.

8.5.1 Kp Models

Figure 8.5.1 shows that the Newell model performs best for the 2012 test case. The SNGI and SWING models obtain similar SSF scores, and the feed-forward neural network receives the highest SSF score of all the models. As is shown in Figure 8.5.2, the same is not valid for the 2017 test case. Here, the SNGI model obtains the lowest SSF score, followed by the Newell model. The feed-forward neural network, in general, obtains the highest SSF score, just as it did in the 2012 test case. All models can obtain SSF scores below the reference model in both test cases.

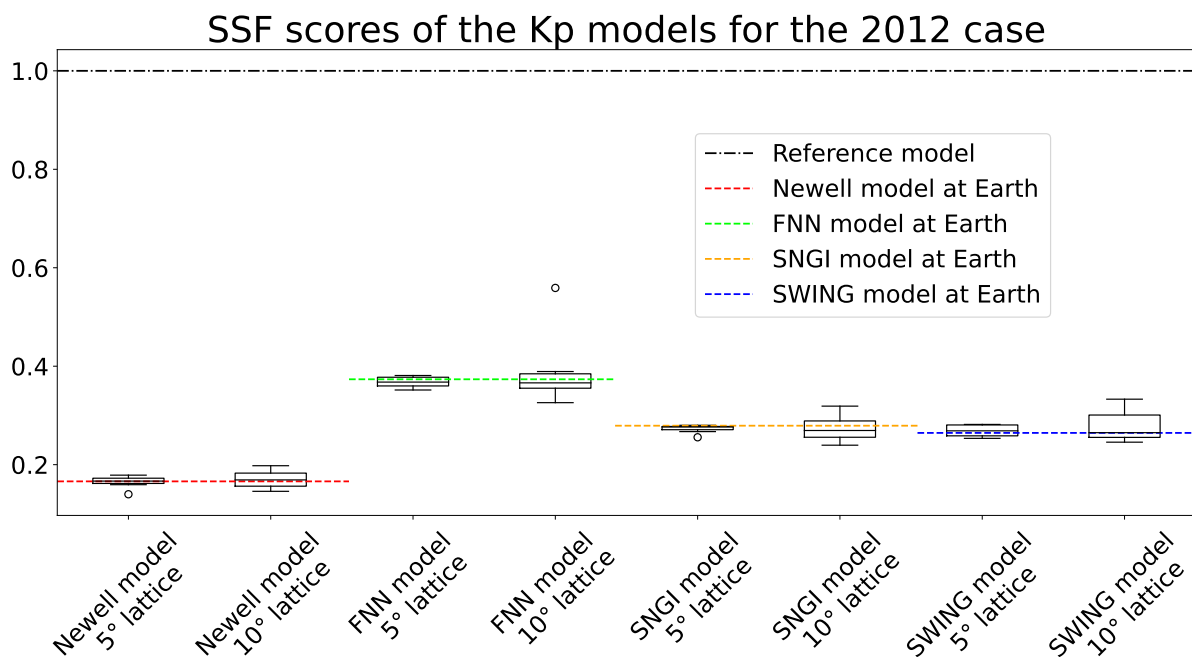


Figure 8.5.1: SSF scores of the Kp models for the 2012 test case.

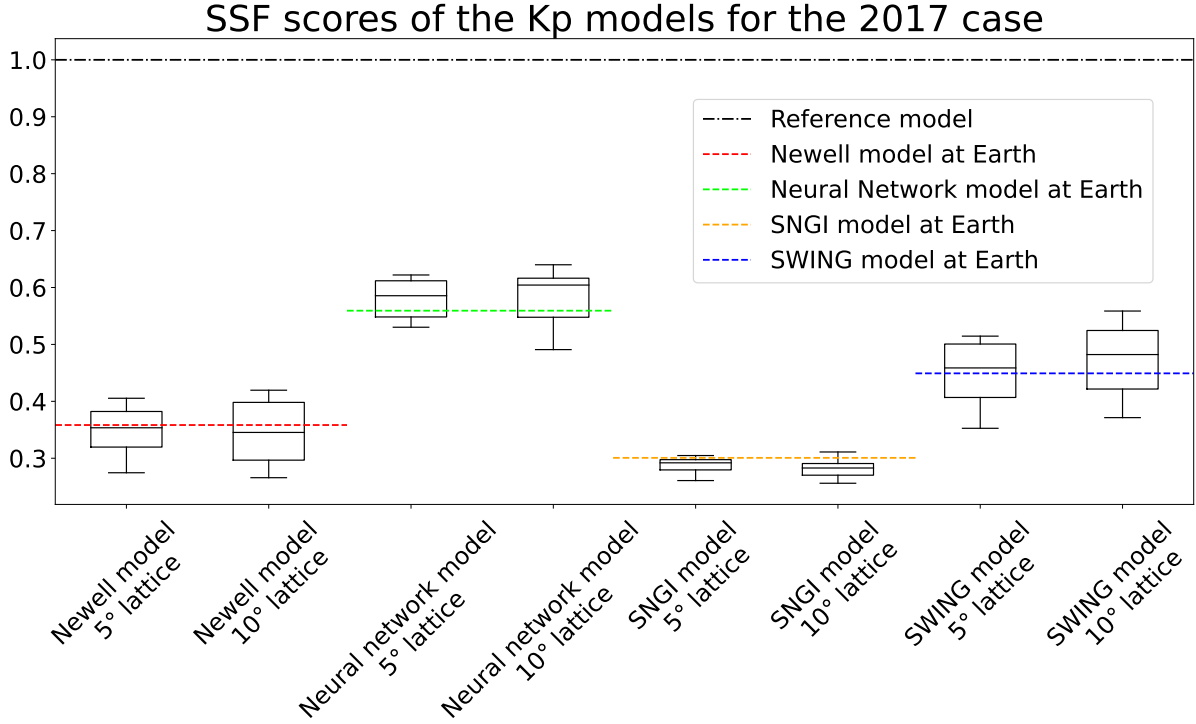
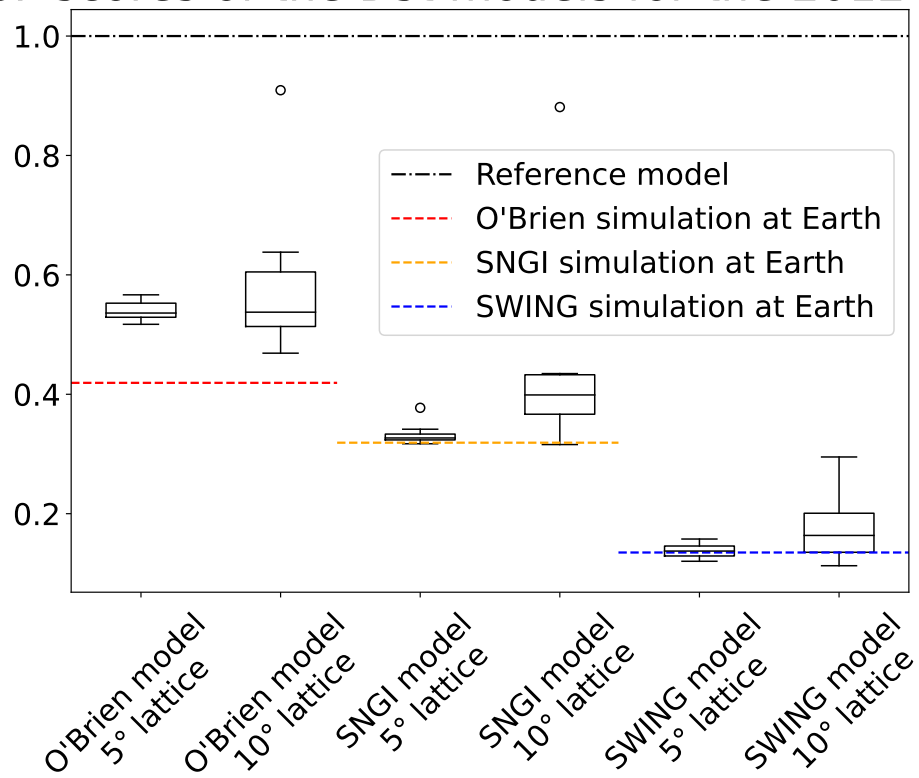


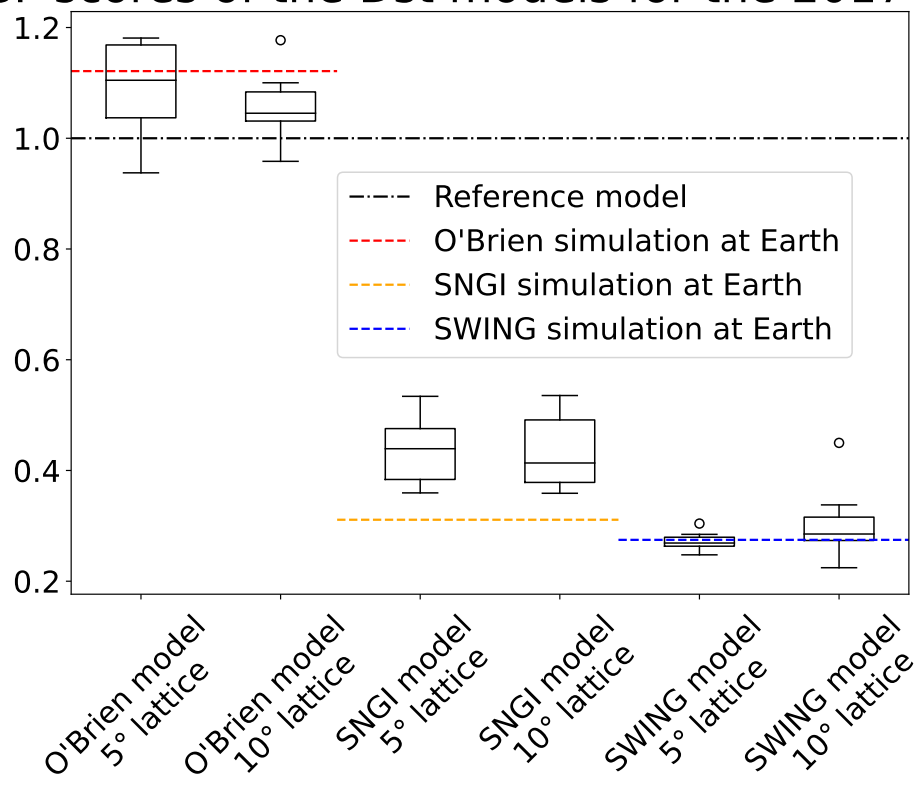
Figure 8.5.2: SSF scores of the Kp models for the 2017 test case.

8.5.2 Dst Models

Figures 8.5.3 and 8.5.4 show a clear distinction between the performance of the different models. In general, the SWING model obtains the lowest SSF score. The SNGI model scores a bit higher but is generally able to outperform the reference model easily. An exception is an outlier on the 10 ° lattice for the 2012 test case, which scores an SSF value of about 0.9. The O'Brien model performs the worst. Generally, it obtains a higher SSF score than the other two models. In the 2012 test case, it can still outperform the reference model. Still, for the 2017 case, this is generally not the case, and only a handful of virtual satellite simulations can attain an SSF score lower than one, and only just.

SSF scores of the *Dst* models for the 2012 caseFigure 8.5.3: SSF scores of the *Dst* models for the 2012 test case.

SSF scores of the Dst models for the 2017 case

Figure 8.5.4: SSF scores of the *Dst* models for the 2017 test case.

8.6 Comparison

8.6.1 Kp Models

The results indicate that all models perform well using synthetic L1 data at Earth compared to the reference model. The four Kp prediction models outperform the reference model for each of the five applied metrics, as seen in Tables 8.1 and 8.2. In both Tables, the models have smaller bias and root mean square error than the reference model. The 2012 test case has Pearson correlation coefficients close to unity, and the 2017 test case also has relatively high Pearson correlation coefficients. The 2012 case also has quite good R^2 values, while for the 2017 case the SNGI and Newell models clearly outperform the other two models. However, all models still easily outperform the reference model. Both test cases have SSF scores smaller than unity, effortlessly outperforming the reference values.

As discussed in this chapter's previous sections, the same applies to all virtual satellites. The obtained values for the virtual satellites are all similar to the values obtained for the EUHFORIA simulation at Earth and do not directly display a great spread around this value. This shows that all four Kp models are robust against small deviations in input values and are also not oversensitive to small spatial errors introduced due to the precision of the L1 monitor.

Across all calculated metrics, the Newell model was the best-performing model for the 2012 case, while the SNGI model was the best-performing model for the 2017 case. Because of this phenomenon, crowning a single model as the best-performing model across all test cases and metrics is impossible. The SWING model was the third best-performing model in both test cases. In both test cases, the feed-forward neural network was generally the worst-performing model, although it still easily outperformed the reference model across all metrics.

Model	Bias	RMSE	CORR	R^2	SSF
Reference	-2.5318	3.0659	0	0	1
Newell	-0.59106	0.98367	0.95373	0.89951	0.16605
FNN	-1.2520	1.5887	0.86731	0.74240	0.37343
SNGI	-0.88027	1.3350	0.90549	0.81678	0.27914
SWING	-1.2826	1.5921	0.87661	0.74423	0.26450

Table 8.1: Metric values for the Kp prediction models using the EUHFORIA synthetic data at Earth for the 2012 test case.

Model	Bias	RMSE	CORR	R^2	SSF
Reference	-2.8500	3.3038	0	0	1
Newell	-0.95605	1.6968	0.84515	0.70089	0.35836
FNN	-1.7762	2.3501	0.65186	0.42159	0.55926
SNGI	-0.37524	1.4610	0.88801	0.78804	0.30068
SWING	-1.5617	2.0391	0.78137	0.56632	0.44916

Table 8.2: Metric values for the Kp prediction models using the EUHFORIA synthetic data at Earth for the 2017 test case.

8.6.2 Dst Models

A distinction has to be made for the performance of the Dst models. Both the SNGI model and the SWING model show good performance in both test cases. Except for a few outliers in the virtual satellite simulations, they consistently outperform the reference and the O’Brien models. In Tables 8.3 and 8.4, it can be seen that they have lower bias, lower root mean square error, relatively high Pearson correlation coefficients, positive R^2 values and low SSF scores.

The O’Brien model obtains good Pearson correlation coefficients but performs worse on all other metrics than the SNGI and SWING models. In the 2012 case, it still managed to outperform the reference model, but this is generally not the case for the 2017 test case. It has a higher bias and root mean square error than the reference model, a negative R^2 value, and even obtains an SSF score greater than unity. Figure 7.1.4 reveals that the O’Brien model greatly overestimates the Dst values before and after the shock. Where one would expect Dst values around zero here, the model predicts Dst values close to fifty. This effect can also be easily observed in Figures 7.4.3 and 7.4.4. Strangely enough, this effect is not present in the paper written by O’Brien and McPherron [2000a] where this model is presented. The origin of this modelling error in this implementation requires further investigation.

As discussed in the previous sections of this chapter, the results of the virtual satellite simulations are consistent with those of the simulations at Earth’s location, except for a few outliers. One could suggest that the spread of the results here is more significant than for the Kp models. Still, one has to take into account that the Kp values are restricted to the range 0 – 9, whereas the Dst index can take any value, which could be a possible explanation for the larger spread. Thus, it can be concluded that the SNGI and SWING Dst models are quite robust against small deviations from the input values and are also not oversensitive to small spatial errors introduced due to L1 monitor precision. Further research is required on the modelling errors observed in this implementation of the O’Brien model before any conclusions can be drawn about this model.

The results in 8.3 and 8.4 show that the SWING model performs the best across the different metrics. Except for the Pearson correlation coefficient in the 2017 test case, it performs better on all metrics in both test cases. Looking back at Figure 8.3.4, it can be seen that the difference in the Pearson correlation coefficient in the 2017 test case is minimal between the SWING and SNGI models and that, in general, the SWING model obtains higher Pearson correlation coefficients for the virtual satellite here than the SNGI

model. The SNGI model is the second best-performing model, while the O'Brien model is the worst-performing model, often even worse than the reference model in the 2017 test case.

Model	Bias	RMSE	CORR	R^2	SSF
Reference	36.216	49.205	0	0	1
O'Brien	26.020	38.489	0.75926	0.43965	0.41915
SNGI	22.788	28.893	0.83270	0.68499	0.31889
SWING	17.977	24.936	0.88714	0.77437	0.13450

Table 8.3: Metric values for the *Dst* prediction models using the EUHFORIA synthetic data at Earth for the 2012 test case.

Model	Bias	RMSE	CORR	R^2	SSF
Reference	11.380	29.849	0	0	1
O'Brien	22.742	38.791	0.67585	-0.59181	1.1213
SNGI	5.7931	28.218	0.78797	0.18439	0.31101
SWING	-2.3407	21.553	0.74859	0.52457	0.27470

Table 8.4: Metric values for the *Dst* prediction models using the EUHFORIA synthetic data at Earth for the 2017 test case.

Chapter 9

Conclusion

This thesis assesses the performance of four Kp prediction models and three Dst prediction models with synthetic L1 data as input. The four Kp prediction models discussed in this thesis are the coupling function devised by Newell et al. [2008], a feed-forward neural network model trained by Wintoft et al. [2017], a recurrent neural network called SWING, and the SNGI Kp -model which is based on the NARMAX methodology as described by Ayala Solares et al. [2016]. The three Dst prediction models discussed in this thesis are the AK2 model derived by O'Brien and McPherron [2000a], a recurrent neural network called SWING, and the SNGI Dst model which is based on the NARMAX methodology described by Ayala Solares et al. [2016].

In this assessment, two test cases were investigated: the CME events of 12 July 2012 and 4-6 September 2017. This synthetic L1 data was obtained using an EUHFORIA simulation for both test cases. From each simulation, the plasma parameters were extracted at the virtual locations of the Earth and 16 virtual satellites situated on what is referred to in this thesis as the 5 ° and 10 ° lattice. Each of these plasma and magnetic parameter sets was used as input for four Kp prediction models and three Dst prediction models, as mentioned above. These predictions were shown in Chapter 7. Using observational plasma parameters as input in the models mentioned above was left for further research.

Five metrics were applied to each prediction using synthetic L1 input data to quantify how well it resembles the observed Kp and Dst values for these events. These five metrics are the bias, the root mean square error, the Pearson correlation coefficient, the coefficient of determination and the self-similarity factor obtained using the dynamic time-warping methodology. The results of the application of these metrics were discussed in Chapter 8.

This chapter also concluded that all four Kp prediction models using synthetic L1 data perform well compared to the reference model $Kp = 0$, show robustness to minor deviations, and are not particularly sensitive to small spatial errors. The same was concluded for two of three Dst models, the SNGI and the SWING models. The observed modeling errors observed for this implementation of the O'Brien model require further research. It was found that the best-performing Kp prediction models were the Newell model and the SNGI model, but a distinction between the two could not be made. The SWING model was the third-best-performing Kp prediction model. The feed-forward neural network model performed the worst, although it still performed better than the reference model. The best-performing Dst model was found to be the SWING model. The SNGI model

was found to be the second-best-performing model, while the O'Brien model was found to be the worst-performing model. In the 2017 test case, the O'Brien model performed worse than the reference model for four of the five metrics applied.

Bibliography

- Syun-Ichi Akasofu and Sydney Chapman. A study of magnetic storms and auroras. 1961.
- Martin D Altschuler and Gordon Newkirk. Magnetic fields and the structure of the solar corona: I: Methods of calculating coronal fields. *Solar Physics*, 9:131–149, 1969.
- Jose Roberto Ayala Solares, Hua-Liang Wei, Richard J Boynton, Simon N Walker, and Stephen A Billings. Modeling and prediction of global magnetic disturbance in near-earth space: A case study for kp index using narx models. *Space Weather*, 14(10): 899–916, 2016.
- J Bartels. An attempt to standardize the daily international magnetic character figure. *IATME Bull. 12e*, 109, 1951.
- J Bartels, NH Heck, and HF Johnston. The three-hour-range index measuring geomagnetic activity. *Terrestrial Magnetism and Atmospheric Electricity*, 44(4):411–454, 1939.
- SA Billings, S Chen, and MJ Korenberg. Identification of mimo non-linear systems using a forward-regression orthogonal estimator. *International journal of control*, 49:2157–2189, 1989.
- Stephen A Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.
- Stephen A Billings and Hua-Liang Wei. An adaptive orthogonal search algorithm for model subset selection and non-linear system identification. *International Journal of Control*, 81(5):714–724, 2008.
- Volker Bothmer and Ioannis A Daglis. *Space weather: physics and effects*. Springer Science & Business Media, 2007.
- RJ Boynton, MA Balikhin, and SA Billings. Online narmax model for electron fluxes at geo. In *Annales Geophysicae*, volume 33, pages 405–411. Copernicus GmbH Göttingen, Germany, 2015.
- Rande K Burton, RL McPherron, and CT Russell. An empirical relationship between interplanetary conditions and dst. *Journal of geophysical research*, 80(31):4204–4214, 1975.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

- Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- National Research Council. *Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report*. The National Academies Press, Washington, DC, 2008. ISBN 978-0-309-12769-1. doi: 10.17226/12507. URL <https://nap.nationalacademies.org/catalog/12507/severe-space-weather-events-understanding-societal-and-economic-impacts-a>.
- T Neil Davis and Masahisa Sugiura. Auroral electrojet activity index ae and its universal time variations. *Journal of Geophysical Research*, 71(3):785–801, 1966.
- Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- Lucie Green and Deb Baker. Coronal mass ejections: a driver of severe space weather. *Weather*, 70(1):31–35, 2015.
- Marina A Gruet, M Chandorkar, Angélica Sicard, and Enrico Camporeale. Multiple-hour-ahead forecast of the dst index using a combination of long short-term memory neural network and gaussian process. *Space Weather*, 16(11):1882–1896, 2018.
- MA Hapgood. Towards a scientific understanding of the risk from extreme space weather. *Advances in Space Research*, 47(12):2059–2072, 2011.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- H Lundstedt, H Gleisner, and P Wintoft. Operational forecasts of the geomagnetic dst index. *Geophysical Research Letters*, 29(24):34–1, 2002.
- Anwesha Maharana, W Douglas Cramer, Evangelia Samara, Camilla Scolini, Joachim Raeder, and Stefaan Poedts. Employing the coupled euhforia-opengcm model to predict cme geoeffectiveness. *Space Weather*, 22(5):e2023SW003715, 2024.
- Pierre-Noël Mayaud. *Atlas of indices K*. IUGG Publication Office, 1967.
- Pierre-Noël Mayaud. *Indices K_n , K_s et K_m* . Éd. du Centre national de la recherche scient., 1968.
- Michel Menvielle, Toshihiko Iyemori, Aurélie Marchaudon, and Masahito Nosé. Geomagnetic indices. In *Geomagnetic observations and models*, pages 183–228. Springer, 2010.
- Michael A Morgan. *Sepher Ha-Razim: The Book of Mysteries*, volume 25. SBL Press, 2022.
- Patrick T Newell, Thomas Sotirelis, Kan Liou, and FJ Rich. Pairs of solar wind-magnetosphere coupling functions: Combining a merging term with a viscous term works best. *Journal of Geophysical Research: Space Physics*, 113(A4), 2008.

- PT Newell, T Sotirelis, K Liou, C-I Meng, and FJ Rich. A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research: Space Physics*, 112(A1), 2007.
- T Paul O'Brien and Robert L McPherron. An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay. *Journal of Geophysical Research: Space Physics*, 105(A4):7707–7719, 2000a.
- T Paul O'Brien and Robert L McPherron. Forecasting the ring current index dst in real time. *Journal of Atmospheric and Solar-Terrestrial Physics*, 62(14):1295–1299, 2000b.
- William Tyler Olcott. *Sun lore of all ages: a collection of myths and legends concerning the sun and its worship*. GP Putnam's sons, 1914.
- Stefaan Poedts. The esa virtual space weather modelling centre. In *EGU General Assembly Conference Abstracts*, pages EGU22–2862, 2022.
- Jens Pomoell and Stefaan Poedts. Euhforia: European heliospheric forecasting information asset. *Journal of Space Weather and Space Climate*, 8:A35, 2018.
- Evangelia Samara, Brecht Laperre, Rungployphan Kieokaew, Manuela Temmer, Christine Verbeke, Luciano Rodriguez, J Magdalenic, and Stefaan Poedts. Dynamic time warping as a means of assessing solar wind time series. *The Astrophysical Journal*, 927(2):187, 2022.
- Kenneth H Schatten, John M Wilcox, and Norman F Ness. A model of interplanetary and coronal magnetic fields. *Solar Physics*, 6:442–455, 1969.
- Camilla Scolini, Luciano Rodriguez, Marilena Mierla, Jens Pomoell, and Stefaan Poedts. Observation-based modelling of magnetised coronal mass ejections with euhforia. *Astronomy & Astrophysics*, 626:A122, 2019.
- Masahisa Sugiura. Hourly values of equatorial dst for the igy. Technical report, 1963.
- Masahisa Sugiura. Equatorial d_{st} index 1957-1986. *IAGA bulletin*, 40:17–38, 1991.
- C Verbeke, J Pomoell, and Stefaan Poedts. The evolution of coronal mass ejections in the inner heliosphere: implementing the spheromak model with euhforia. *Astronomy & Astrophysics*, 627:A111, 2019.
- Economic Impacts Of Severe Space Weather et al. Severe space weather events—understanding societal and economic impacts: A workshop report. In *Severe Space Weather Events: Understanding Societal and Economic Impacts Workshop Report*, 2008.
- P Wintoft and M Wik. Evaluation of kp and dst predictions using ace and dscovr solar wind data. *Space Weather*, 16(12):1972–1983, 2018.
- Peter Wintoft and Magnus Wik. Exploring three recurrent neural network architectures for geomagnetic predictions. *Frontiers in Astronomy and Space Sciences*, 8:664483, 2021.

- Peter Wintoft, Magnus Wik, Jürgen Matzka, and Yuri Shprits. Forecasting kp from solar wind data: input parameter study using 3-hour averages and 3-hour range values. *Journal of Space Weather and Space Climate*, 7:A29, 2017.
- Jian-Guo Wu and Henrik Lundstedt. Neural network modeling of solar wind-magnetosphere interaction. *Journal of Geophysical Research: Space Physics*, 102(A7): 14457–14466, 1997.

