

Δημιουργία μοντέλων μηχανικής μάθησης για τη μελέτη της απόκρισης της ιονόσφαιρας σε έντονα ηλιακά φαινόμενα με χρήση παρατηρήσεων GNSS

Τσουμένης Εμμανουήλ

Διπλωματική Εργασία

Σχολή Αγρονόμων και Τοπογράφων Μηχανικών – Μηχανικών Γεωπληροφορικής
Εθνικό Μετσόβιο Πολυτεχνείο



School of Rural, Surveying and Geoinformatics Engineering
National Technical University of Athens

Φεβρουάριος 2022

Επιβλέπων: Δουλάμης Νικόλαος
Αναπληρωτής Καθηγητής ΕΜΠ

Τριμελής Επιτροπή

Δουλάμης Νικόλαος
Αναπληρωτής Καθηγητής

Πανταζής Γεώργιος
Καθηγητής

Γκίκας Βασίλειος
Καθηγητής

Περίληψη

Οι μετρήσεις προσδιορισμού θέσης με χρήση συστημάτων GNSS αποτελούν ένα τεράστιο τμήμα της καθημερινότητας μας σε διάφορους τομείς. Όμως, η μετρήσεις αυτές επηρεάζονται σε μεγάλο βαθμό μεταξύ άλλων και από την ατμόσφαιρα της Γης και, πιο συγκεκριμένα, από την ιονόσφαιρα. Ως εκ τούτου, ένα μεγάλο πρόβλημα των ημερών μας, είναι η ανάγκη πρόβλεψης της ιονοσφαιρικής δραστηριότητας. Η απόκριση της ιονόσφαιρας είναι ανάλογη του συνολικού αριθμού ηλεκτρονίων (TEC) κατά μήκος της διαδρομής διάδοσης, και αντιστρόφως ανάλογη του τετραγώνου της συχνότητας f του φορέα (L1 ή L2). Μέσω των δεδομένων TEC γίνεται η διόρθωση της διάθλασης στην ιονόσφαιρα και είναι ζωτικής σημασίας για τα συστήματα δορυφορικής πλοήγησης, προκειμένου να διασφαλιστεί η υψηλή απόδοση τους στον εντοπισμό θέσης.

Με το παρόν, έγινε μία προσπάθεια δημιουργίας μοντέλων μηχανικής μάθησης για τη μελέτη της απόκρισης της ιονόσφαιρας σε έντονα ηλιακά φαινόμενα με χρήση παρατηρήσεων GNSS. Αυτή η μέθοδος εκμεταλλεύεται τα πλεονεκτήματα της μηχανικής μάθησης και συγκεκριμένα του παλινδρομητή XGBoost για την πρόβλεψη των τιμών TEC (VTEC και STEC), χρησιμοποιώντας ως δεδομένα εισόδου αυτά που προκύπτουν από τις 8 γεωμαγνητικές και ηλιακές παραμέτρους που επηρεάζουν την ιονόσφαιρα. Ο σκοπός της προτεινόμενης μεθόδου είναι να αξιολογήσει την εκτίμηση της ιονοσφαιρικής καθυστέρησης σε ημέρες με έντονα ηλιακά φαινόμενα όπως solar flares ή CMEs) και να ελέγξει τη διαφορά της απόκρισης της κάτω από αυτά τα φαινόμενα. Αυτό θα οδηγήσει στην καλύτερη κατανόηση των προβλημάτων που δημιουργούν τα φαινόμενα αυτά και των παραμέτρων που επηρεάζουν και, ιδανικά, θα βοηθήσει στη πρόβλεψη τέτοιων φαινομένων με σκοπό τον προσδιορισμό της ιονόσφαιρας και τη δημιουργία καλύτερων τοπικών μοντέλων πρόβλεψης (χάρτες TEC).

Για τον προσδιορισμό του καλύτερου παλινδρομητή χρησιμοποιήθηκαν οι συναρτήσεις Ρίζας Μέσου Τετραγωνικού Σφάλματος (rMSE) και Μέσου Απόλυτου Σφάλματος (MAE).

Λέξεις Κλειδιά

GNSS, TEC, STEC, VTEC, rMSE, MAE, ηλιακά φαινόμενα, ηλιακές εκλάμψεις, ιονόσφαιρα, XGBoost, Random Forest, Python, ηλιακές εκρήξεις πλάσματος

Abstract

Positioning using GNSS systems are a huge part of our daily lives in various fields (scientific or not). However, these measurements are greatly influenced, among other things, by the Earth's atmosphere and, more specifically, by the ionosphere. Therefore, a major problem today is the need to predict ionospheric activity. The ionospheric response is proportional to the total number of electrons (TEC) along the propagation path, and inversely proportional to the square of the carrier frequency f (L1 or L2). TEC data corrects ionosphere refraction and is vital for satellite navigation systems to ensure their high performance in positioning.

This thesis is an attempt to develop machine learning models for the study of the ionospheric response to intense solar phenomena using GNSS observations. This method exploits the advantages of machine learning and in particular the XGBoost regressor for predicting TEC values (VTEC and STEC), using as input data those resulting from the 8 geomagnetic and solar indices that affect the ionosphere. The purpose of the proposed method is to evaluate the estimate of ionospheric delay on days with intense solar phenomena (solar flares or CMEs) and to check the difference in response under those phenomena. This will lead to a better understanding of the problems created by those phenomena and the parameters that affect them and, ideally, will help in predicting such phenomena in order to determine the ionosphere and create better local prediction models (TEC maps).

The square root average error (rMSE) and absolute mean error (MAE) functions were used to determine the best regressor.

Keywords

GNSS, TEC, STEC, VTEC, rMSE, MAE, solar phenomena, solar flares, ionosphere, XGBoost, Random Forest, Python, CMEs

Ευχαριστίες

Με την συγγραφή αυτής της διπλωματικής εργασίας, ολοκληρώθηκε ακόμα ένας κύκλος σπουδών. Μετά την επιτυχή φοίτηση στη Στρατιωτική Σχολή Ευελπίδων, τη Σχολή Μηχανικού και τη Σχολή Τοπογραφίας της Γεωγραφικής Υπηρεσίας Στρατού, η αποφοίτησή μου από τη Σχολή Αγρονόμων και Τοπογράφων Μηχανικών – Μηχανικών Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου έρχεται να κλείσει ένα μεγάλο κύκλο σπουδών και να ανοίξει την πόρτα σε ένα κύκλο επαγγελματικής και ακαδημαϊκής εξέλιξης σύμφωνα με τα εφόδια που έλαβα όλα αυτά χρόνια.

Θα ήθελα μέσα από την εργασία αυτή να πω ένα μεγάλο ευχαριστώ στην οικογένεια μου που με στήριξε όλα αυτά τα χρόνια και ήταν δίπλα μου σε όλες τις αντιξοότητες που αντιμετώπισα.

Ένα μεγάλο ευχαριστώ θα πάει επίσης και στους συναδέλφους και συμφοιτητές Σακελλαρίου Δημήτριο και Σπανάκη Μάριο γιατί ήμασταν πάντα ο ένας για τον άλλο από την πρώτη ημέρα που παρουσιαστήκαμε στη ΓΥΣ μέχρι και την παρουσίαση των διπλωματικών μας εργασιών.

Επιπλέον θα ήθελα να ευχαριστήσω δεόντως τον Αναπληρωτή Καθηγητή Δουλάμη Νικόλαο γιατί από την πρώτη ημέρα που τον συναντήσαμε στη Σχολή Τοπογραφίας της ΓΥΣ μας άνοιξε τα μάτια και μας έδωσε το έναυσμα για να ξεκινήσει αυτό το ταξίδι στον τομέα της Τοπογραφίας και της Γεωπληροφορικής. Ήταν δίπλα μας από την πρώτη στιγμή και μας έδειξε απίστευτη εμπιστοσύνη και κατά την ανάληψη της παρούσας διπλωματικής. Τέλος, θα ήθελα επίσης να ευχαριστήσω και τη Δρ. Κασελίμη Μαρία για την άψογη συνεργασία και τη συνεχή καθοδήγηση στην εκπόνηση αυτής της εργασίας.

Περιεχόμενα

Περίληψη	3
Λέξεις Κλειδιά	3
Abstract	4
Keywords.....	4
Ευχαριστίες	5
Περιεχόμενα	6
Κατάλογος Πινάκων	8
Κατάλογος Εικόνων.....	9
Κατάλογος Συντομογραφιών	10
1. Εισαγωγή	12
1.1. Η σημασία της μελέτης της ιονόσφαιρας.....	12
1.2. Κύριες προκλήσεις	12
1.3. Συνεισφορά	13
1.4. Δομή εργασίας	14
2. Βιβλιογραφική Ανασκόπηση	15
2.1. Απόκριση ιονόσφαιρας σε τρία ακραία γεγονότα που συμβαίνουν κοντά στην εαρινή ισημερία το 2012, το 2013 και το 2015, που παρατηρήθηκαν από το περιφερειακό μοντέλο GNSS-TEC.....	15
2.2. Μελέτη της απόκρισης της ιονόσφαιρας στην X9.3 Flare στις 6 Σεπτεμβρίου 2017 και οι επιπτώσεις της για τις υπηρεσίες πλοήγησης στην Ευρώπη	16
2.3. Επιδράσεις του σπινθηρισμού φάσης στο σφάλμα θέσης GNSS κατά τη διάρκεια της καταιγίδας του Σεπτεμβρίου 2017 στο Svalbard	17
2.4. Πρόβλεψη των επιπτώσεων των ηλιακών καταιγίδων στην ιονόσφαιρα με βάση μια σύγκριση δεδομένων ηλιακού ανέμου σε πραγματικό χρόνο με το καταλληλότερο ιστορικό γεγονός καταιγίδας	18
2.5. Ανάπτυξη μοντέλου μηχανικής μάθησης για πρόγνωση διαστημικού καιρού στην Ιονόσφαιρα.....	18
2.6. Εφαρμογή αλγορίθμου πρόβλεψης υβριδικών ιονοσφαιρικών δεδομένων TEC με χρήση της μεθόδου PCA-NN	19
3. Το σύστημα GNSS.....	21

4.	Η ατμόσφαιρα της Γης.....	22
4.1.	Εισαγωγικές Έννοιες.....	22
4.2.	Η Ατμόσφαιρα της Γης	22
4.3.	Στρώματα	22
5.	Ιονόσφαιρα.....	24
5.1.	Εισαγωγικές Έννοιες.....	24
5.2.	Δομή	24
5.3.	Tec	28
5.4.	Παράμετροι.....	29
5.4.1.	Geomagnetic activity (K).....	31
5.4.2.	Auroral Electrojet (AE)	32
5.4.3.	Disturbance storm time (Dst)	33
5.4.4.	Solar Radio Flux.....	34
5.4.5.	Sunspot Number	34
5.4.6.	Energetic Particle Flux.....	34
5.4.7.	Solar Flares.....	35
5.4.8.	Coronal Mass Ejections	36
6.	Μηχανική Μάθηση.....	37
6.1.	Εισαγωγικές Έννοιες.....	37
7.	Αλγόριθμοι Παλινδρόμησης.....	39
7.1.	Αλγόριθμος Random Forest.....	39
7.2.	Αλγόριθμος XGBoost.....	40
8.	Προγραμματισμός επίλυσης σε Python.....	42
8.1.	Η γλώσσα προγραμματισμού Python.....	42
8.2.	Google Colab	43
8.3.	Βιβλιοθήκες Python	43
8.4.	Υλοποίηση σε Python.....	45
9.	Πειραματική Διαχείριση	47
9.1.	Περιγραφή των δεδομένων	47

9.1.1.	Observation Files του σταθμού COTE00ATA	47
9.1.2.	Geomagnetic activity (K).....	52
9.1.3.	Auroral Electrojet (AE)	53
9.1.4.	Disturbance storm time (Dst)	54
9.1.5.	Solar Radio Flux.....	55
9.1.6.	Sunspot Number	57
9.1.7.	Energetic Particle Flux.....	58
9.1.8.	Solar Flares.....	59
9.1.9.	Coronal Mass Ejections	61
9.2.	Επεξεργασία δεδομένων.....	62
9.3.	Εκτέλεση προγράμματος – Εξαγωγή αποτελεσμάτων – Αξιολόγηση	62
9.3.1.	Σύγκριση παλινδρομητών.....	63
9.3.2.	Αποτελέσματα	66
9.3.3.	Αξιολόγηση	68
10.	Συμπεράσματα	69
	Βιβλιογραφία	70
	Παραρτήματα	75
1.	Πίνακας Δεδομένων ανά μήνα	75
2.	Κώδικας RF_Solar.....	75
2.1.	STEC	75
2.2.	VTEC.....	82

Κατάλογος Πινάκων

Πίνακας 1.	Παράμετροι επηρεασμού της ιονόσφαιρας.	29
Πίνακας 2.	Αντιστοιχία του δείκτη Kp με τον δείκτη ap.....	32
Πίνακας 3.	Πίνακας σύγκρισης παλινδρομητών Ιανουαρίου για STEC και VTEC αντίστοιχα.....	63
Πίνακας 4.	Πίνακας σύγκρισης παλινδρομητών Φεβρουαρίου για STEC και VTEC αντίστοιχα.....	63

Πίνακας 5. Πίνακας σύγκρισης παλινδρομητών Μαρτίου για STEC και VTEC αντίστοιχα	64
Πίνακας 6. Πίνακας σύγκρισης παλινδρομητών Απριλίου για STEC και VTEC αντίστοιχα	64

Κατάλογος Εικόνων

Εικόνα 1. STEC/ VTEC (Πηγή: Shuanggen Jin, Estel Cardellach, Feiqin Xie, Ground GNSS Ionosphere Sounding, 2013)	29
Εικόνα 2. Δίκτυο σταθμών ΑΕ (Μπλε χρώμα) και SME (Κόκκινο χρώμα)	33
Εικόνα 3. Οι Χ2.2 και Χ9.3 Solar Flares που εξέπεμψε ο Ήλιος στις 6 Σεπτεμβρίου 2017. Οι εικόνες καταγράφηκαν από το Solar Dynamics Observatory της NASA και δείχνουν φως στο μήκος κύματος των 131 angstrom. (Πηγή: NASA/Goddard/SDO)	35
Εικόνα 4. Έκρηξη CME στις 31 Αυγούστου 2012 (Πηγή: NASA Goddard Space Flight Center)	36
Εικόνα 6. Αναπαράσταση του αλγορίθμου Random Forest (Πηγή: levelup.gitconnected.com).	40
Εικόνα 7. Η εξέλιξη του αλγορίθμου XGBoost (Πηγή: towardsdatascience.com)	41
Εικόνα 8. Σταθμός COTE00ATA (Πηγή: IGS)	47
Εικόνα 9. Όλοι οι διαθέσιμοι σταθμοί του συστήματος GNSS (Πηγή: IGS)	48
Εικόνα 10. Τυπική μορφή αρχείου RINEX.	48
Εικόνα 11. Παραγόμενα διαγράμματα από το λογισμικό GPS Gori	52
Εικόνα 12. Υπόδειγμα δεδομένων γεωμαγνητικής δραστηριότητας (Κ)	52
Εικόνα 13. Ανταπόκριση παραμέτρων ΑΕ και SME σε γεγονός την 30 Ιανουαρίου 1997 (Πηγή: Newell and Gjerloev, Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power, 2011)	53
Εικόνα 14. Υπόδειγμα δεδομένων SME	54
Εικόνα 15. Υπόδειγμα δεδομένων Dst	55
Εικόνα 16. Υπόδειγμα δεδομένων Solar Radio Flux	56
Εικόνα 17. Υπόδειγμα δεδομένων Sunspot Number	57
Εικόνα 18. Βασιλικό Αστεροσκοπείο Βελγίου (Πηγή: wwwbis.sidc.be)	58
Εικόνα 19. Υπόδειγμα δεδομένων Energetic Particle Flux	59
Εικόνα 20. Υπόδειγμα δεδομένων Solar Flares.	60

Εικόνα 21. Υπόδειγμα δεδομένων Coronal Mass Ejections.	62
Εικόνα 22. Γραφήματα σύγκρισης αποτελεσμάτων Ιανουαρίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα	66
Εικόνα 23. Γραφήματα σύγκρισης αποτελεσμάτων Φεβρουαρίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα	67
Εικόνα 24. Γραφήματα σύγκρισης αποτελεσμάτων Μαρτίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα	67
Εικόνα 25. Γραφήματα σύγκρισης αποτελεσμάτων Απριλίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα	68

Κατάλογος Συντομογραφιών

AE	Auroral Electrojet
CDDIS	Crustal Dynamics Data Information System
CDP	Crustal Dynamics Project
CME	Coronal Mass Ejection
CODE	Center for Orbit Determination in Europe
DORIS	Doppler Orbitography and Radio-positioning Integrated by Satellite
Dst	Disturbance storm time
EGNOS	European Geostationary Navigation Overlay Service
GIM	Global Ionosphere Maps
IAGA	International Association of Geomagnetism and Aeronomy
IDS	International DORIS Service
IGS	International GNSS Service
ILRS	International Laser Ranging Service
IVS	International VLBI Service for Geodesy and Astrometry
GLONASS	GLOBAL Navigation Satellite System
GNSS	Global Navigation Satellite System
GOES-15	Geostationary Operational Environmental Satellite
GPS	Global Positioning System

GTEC	Global Total Electron Content
IRNSS	Indian Regional Navigational Satellite System
LASO	Large Angle and Spectrometric Coronagraph
MAE	Mean Absolute Error
MEV	Mega ElectronVolt
NASA	National Aeronautics and Space Administration
NAVSTAR	NAVigation Satellite Timing And Ranging
NOAA	National Oceanic and Atmospheric Administration
PCA-NN	Principal Component Analysis Neural Networks
QZSS	Quasi-Zenith Satellite System
RHESSI	Ramaty High Energy Solar Spectroscopic Imager
RINEX	Receiver Independent Exchange Format
RMSE	Rooted mean squared error
SBAS	Satellite-Based Augmentation System
SLR	Satellite Laser Ranging
SME	SuperMAG Electrojet
SML	SuperMAG Upper
SMU	SuperMAG Lower
STEC	Slant Total Electron Content
SWPC	Space Weather Prediction Center
TEC	Total Electron Content
TECU	Total Electron Content Units
VLBI	Very Long Baseline Interferometry
VTEC	Vertical Total Electron Content
WAAS	Wide Area Augmentation System
XGBoost	eXtreme Gradient Boosting

1. Εισαγωγή

1.1. Η σημασία της μελέτης της ιονόσφαιρας

Η ιονόσφαιρα είναι το τμήμα της ατμόσφαιρας που επεκτείνεται από περίπου 50 έως 1000 χλμ. πάνω από τη γήινη επιφάνεια και χαρακτηρίζεται από την ύπαρξη ιονιζόντων αερίων κυρίως ελεύθερων ηλεκτρονίων και θετικών ιόντων. Η ιονόσφαιρα αποτελεί το μέσο διάδοσης και διασποράς για το σήμα GNNS. Όμως, το δορυφορικό σήμα GNNS, καθώς διαδίδεται στην ατμόσφαιρα μέχρι τη γήινη επιφάνεια συνεχώς διαθλάται λόγω της μεταβολής του δείκτη διάθλασης της ατμόσφαιρας. Ειδικότερα το σήμα GNNS, που είναι διαμορφωμένο κύμα και μεταφέρει πληροφορία, κατά τη διάδοσή του στην ιονόσφαιρα εξαρτάται σημαντικά και από τις συχνότητες των συνιστωσών κυμάτων που το απαρτίζουν και, άρα, οι διάφορες συνιστώσες του ταξιδεύουν με διαφορετική ταχύτητα στην ιονόσφαιρα. Αντίθετα, στην τροπόσφαιρα δεν υπάρχει εξάρτηση από τη συχνότητα και όλες οι συνιστώσες του του σήματος έχουν ίδια ταχύτητα διάδοσης (Φωτίου, Πικριδάς, 2012). Το ιονοσφαιρικό σφάλμα, αποτελεί τυχαίο σφάλμα (random error) που προφανώς είναι αναπόφευκτο κατά την εκτέλεση των μετρήσεων, και ονομάζεται θόρυβος (noise, white noise).

Για την εκτίμηση των σφαλμάτων αυτών, στα υποπλάισια 4 και 5 του μηνύματος πλοήγησης που εκπέμπει ο κάθε δορυφόρος υπάρχουν πληροφορίες για τις παραμέτρους της ιονοσφαιρικής επίδρασης. Οι παράμετροι αυτοί είναι οκτώ, αφορούν μαγνητική και ηλιακή δραστηριότητα και χρησιμοποιούνται για μία προσεγγιστική εκτίμηση της ιονοσφαιρικής καθυστέρησης του σήματος. Ειδικότερα, για τις παραμέτρους της ηλιακής δραστηριότητας, έχει παρατηρηθεί ότι κατά τη διάρκεια έντονης ηλιακής δραστηριότητας, αυξάνεται κατά πολύ ο αριθμός των ηλεκτρονίων (TEC) στη διάρκεια του χειμώνα και αδυνατίζει στη διάρκεια του καλοκαιριού. Το φαινόμενο της ιονοσφαιρικής έξαρσης (ionospheric scintillation) συμβαίνει κυρίως στις πολικές περιοχές και σε μία ζώνη $\pm 30^\circ$ γύρω από το γεωμαγνητικό ισημερινό και διαρκεί έως μερικά λεπτά.

Στα πλαίσια της παρούσας διπλωματικής εργασίας θα μελετηθεί η απόκριση της ιονοσφαιρικής δραστηριότητας κάτω από έντονα ηλιακά φαινόμενα όπως εμφανίζονται στις παραμέτρους που αφορούν την ηλιακή δραστηριότητα. Για την μελέτη αυτή θα γίνει χρήση μοντέλων μηχανικής μάθησης και μετρήσεων GNSS. Στην ουσία μέσω της εκπαίδευσης ενός νευρωνικού δικτύου, θα γίνει προσπάθεια να προβλεφθεί το ιονοσφαιρικό σφάλμα σε στιγμές έντονης ηλιακής δραστηριότητας, και θα μελετηθεί αν και κατά πόσο επηρεάζει τις μετρήσεις GNSS.

1.2. Κύριες προκλήσεις

Το κύριο ερευνητικό πεδίο που συζητάται εδώ είναι το ιονοσφαιρικό μοντέλο TEC, που είναι ένα σημαντικό ζήτημα σε πολλές εφαρμογές GNSS πραγματικού χρόνου. Η αξιόπιστη γνώση των ιονοσφαιρικών μεταβολών γίνεται όλο και πιο σημαντική. Χρήστες GNSS, με δέκτες μονής συχνότητας χρειάζονται ακριβείς διορθώσεις για να αφαιρέσουν την υποβάθμιση του σήματος που προκαλείται από την ιονόσφαιρα.

Παράλληλα, πρέπει να είναι σε θέση να προβλέψουν αν και κατά πόσο επηρεάζεται το σήμα από την ηλιακή δραστηριότητα.

Τα νευρωνικά δίκτυα βαθιάς μάθησης (Deep Learning Neural Networks) μαθαίνουν πολύπλοκες αντιστοιχίσεις μεταξύ δεδομένων εισόδου και εξόδου και υποστηρίζουν πολλαπλές εισόδους και εξόδους. Έχουν αποδειχθεί αποτελεσματική λύση δεδομένης της ικανότητάς τους να αυτοματοποιούνται. Ωστόσο, επιλέγοντας το πιο βολικό τύπο νευρωνικού δικτύου βαθιάς μάθησης και η παραμετροποίησή του, είναι ένα σύνθετο έργο που απαιτεί σημαντική εξειδίκευση.

Το Ευρωπαϊκό Κέντρο Προσδιορισμού Τροχιών (CODE) παρέχει Παγκόσμιους Ιονοσφαιρικούς Χάρτες (GIM) σε ένα κάναβο ($2,5^{\circ} \times 5,0^{\circ}$), με ανάλυση 2 ωρών. Ωστόσο, οι τιμές STEC και VTEC αποτυγχάνουν να αφαιρέσουν τη συνολική ποσότητα θορύβου που προκαλείται από την ιονόσφαιρα, λόγω της ευρείας τους κάλυψης και την αδυναμίας τους να προσαρμοστούν με ακρίβεια σε περιφερειακό (και περισσότερο σε τοπικό) επίπεδο.

Ο στόχος αυτής της διπλωματικής εργασίας είναι να μελετήσει την απόκριση της ιονόσφαιρας σε έντονα ηλιακά φαινόμενα χρησιμοποιώντας μεθόδους βαθιάς μάθησης και να ξεπεραστούν τα προβλήματα που προκύπτουν χρησιμοποιώντας τις παραδοσιακές μεθόδους για τη μοντελοποίηση TEC. Έτσι, χρήστες κοντά στην περιοχή αναφοράς αυτών των περιφερειακών μοντέλων βαθιάς μάθησης, θα μπορούσαν να επωφεληθούν από αυτό το μοντέλο αντί για το παγκόσμιο.

Ειδικότερα, οι ακόλουθοι ερευνητικοί στόχοι που πηγάζουν από το στόχο της παρούσας διπλωματικής εργασίας είναι :

- Η βελτίωση της ακρίβειας των υφιστάμενων περιφερειακών μοντέλων TEC και η διερεύνηση της απόκρισης του μοντέλου σε περιπτώσεις υψηλής ηλιακής δραστηριότητας.
- Η διερεύνηση της σημασίας των διαφόρων υποπαραμέτρων ηλιακής δραστηριότητας και η εύρεση των σημαντικότερων εξ αυτών, για τη βελτίωση της ακρίβειας στη μελέτη της απόκρισης της ιονόσφαιρας.

1.3. Συνεισφορά

Η συνολική συνεισφορά της παρούσας διπλωματικής εργασίας παρουσιάζεται στους ακόλουθους στόχους:

- να ελεγχθεί η ανταπόκριση του μοντέλου σε περιπτώσεις υψηλής ηλιακής δραστηριότητας.
- να παρουσιάσει μια ολοκληρωμένη ανάλυση για διάφορους μήνες, υπό διαφορετική ηλιακή δραστηριότητα. Τα μοντέλα ελέγχονται για

διαφορετικές ημέρες για να ελεγχθεί η ανταπόκριση του μοντέλου σε κάθε περίπτωση ηλιακής δραστηριότητας.

1.4. Δομή εργασίας

Στο δεύτερο κεφάλαιο γίνεται μία βιβλιογραφική ανασκόπηση σε αντίστοιχες εργασίες που έχουν μελετήσει σε κάποιο βαθμό την απόκριση της ιονόσφαιρας σε έντονα ηλιακά φαινόμενα ή έχουν κάνει χρήση σχετικών δεδομένων προκειμένου να βρεθούν τρόποι αντιμετώπισης του θέματος.

Στο τρίτο κεφάλαιο αναφέρονται κάποια γενικά στοιχεία το παγκόσμιο σύστημα προσδιορισμού θέσης.

Στο τέταρτο κεφάλαιο αναφέρονται κάποια γενικά στοιχεία για την ατμόσφαιρα της Γης και γίνεται ειδική αναφορά στον τρόπο που τα ατμοσφαιρικά σφάλματα επηρεάζουν τον προσδιορισμό της θέσης.

Στο πέμπτο κεφάλαιο αναλύεται η ιονόσφαιρα, το ιονοσφαιρικό σφάλμα και πως αυτό επηρεάζει το δορυφορικό σήμα GNSS. Επίσης αναλύονται οι παράμετροι που επηρεάζουν την ιονοσφαιρική δραστηριότητα και αναλύεται η έννοια των δεδομένων TEC.

Στο έκτο κεφάλαιο πραγματοποιείται επιγραμματική ανάλυση της μηχανικής μάθησης.

Στο έβδομο κεφάλαιο σχολιάζονται κάποιοι αλγόριθμοι παλινδρόμησης που χρησιμοποιήθηκαν για την εκπόνηση του πειραματικού σταδίου της εργασίας.

Στο όγδοο κεφάλαιο γίνεται μνεία στη γλώσσα προγραμματισμού Python και αναλύεται λεπτομερώς ο κώδικας του λογισμικού που συντάχθηκε για την επίλυση του προβλήματος

Στο ένατο κεφάλαιο παρατίθενται τα πειραματικά δεδομένα που χρησιμοποιήθηκαν για την επίλυση του θέματος και αναλύεται η επεξεργασία που απαιτήθηκε να γίνει σε αυτά. Επίσης, παρουσιάζονται τα αποτελέσματα και γίνεται η αξιολόγησή τους.

Τέλος, στα παραρτήματα παρατίθενται το σύνολο των δεδομένων καθώς και ο κώδικας του προγράμματος που συντάχθηκε.

2. Βιβλιογραφική Ανασκόπηση

Στο συγκεκριμένο κεφάλαιο, γίνεται βιβλιογραφική ανασκόπηση σε εργασίες που έχουν μελετήσει με διάφορους τρόπους την απόκριση της ιονοσφαιρικής δραστηριότητας σε έντονα ηλιακά φαινόμενα. Έχοντας ως βάση τις δυσκολίες σύμφωνα με τους τύπους δεδομένων για την υλοποίηση μιας τέτοιας διαδικασίας και τη σημασία αυτής της, αξίζει να μελετηθεί η μεθοδολογία αντιμετώπισής του θέματος και από άλλους ερευνητές. Οι εργασίες ταξινομούνται σύμφωνα με τα δεδομένα και τη λύση που προτείνεται.

2.1. Απόκριση ιονόσφαιρας σε τρία ακραία γεγονότα που συμβαίνουν κοντά στην εαρινή ισημερία το 2012, το 2013 και το 2015, που παρατηρήθηκαν από το περιφερειακό μοντέλο GNSS-TEC

Το άρθρο αυτό [48] αναφέρει ότι ο 24^{ος} ηλιακός κύκλος χαρακτηρίζεται από την εμφάνιση τριών ισχυρών διαταραχών κοντά σε μια ισημερία. Σε αυτή τη δημοσίευση, αναλύθηκαν οι αποκρίσεις της ιονόσφαιρας στις καταιγίδες της ισημερίας το 2012, το 2013 και το 2015 που οδηγήθηκαν από εκτινάξεις μάζας στεμματών. Λόγω της δυναμικής φύσης των αλλαγών στην ιονόσφαιρα, η ακρίβεια και η ανάλυση των υφιστάμενων παγκόσμιων μοντέλων ιονόσφαιρας είναι συχνά ανεπαρκείς για να αντικατοπτρίζουν λεπτομερώς τις επιπτώσεις της καταιγίδας. Έχοντας αυτό υπόψη, εφαρμόστηκε ένα νέο περιφερειακό μοντέλο ιονόσφαιρας υψηλής ακρίβειας και υψηλής ανάλυσης για τη μελέτη της απόκρισης αυτού του στρώματος σε σοβαρές γεωμαγνητικές καταιγίδες στην Ευρώπη. Οι νέοι περιφερειακοί χάρτες TEC προήλθαν αποκλειστικά από ακριβή δεδομένα φάσης φέροντος παγκόσμιων δορυφορικών συστημάτων πλοήγησης (GNSS). Αν και αυτή η μελέτη βασίζεται σε διαστημική γεωδαιτική τεχνική, πραγματοποιήθηκε επίσης σε σχέση με τις παρατηρήσεις που παρέχονται από ευρωπαϊκούς ιονοσόντες (παλμικά ραντάρ, των οποίων η συχνότητα μεταβάλλεται εντός συγκεκριμένου εύρους). Επιπλέον, οι περιφερειακοί χάρτες συγκρίθηκαν με τους τελικούς IGS Global Ionosphere Maps, όπου η νέα λύση έδειξε καλύτερο επίπεδο λεπτομέρειας. Τα αποτελέσματα των χρονικών αλλαγών TEC χρόνου καταιγίδας που παρέχονται από τα δεδομένα GNSS επιβεβαιώθηκαν από αλλαγές NmF2 που προέρχονται από ιονοσόνδρες. Και οι δύο πηγές δεδομένων επιβεβαίωσαν την υψηλή συμβατότητά τους για τη μελέτη της διαταραγμένης ιονόσφαιρας. Οι μαγνητικές καταιγίδες που σημειώθηκαν στις 7, 9, 12 και 15 Μαρτίου 2012 ήταν διαφορετικής φύσης. Η μεγαλύτερη αλλαγή στο συνολικό περιεχόμενο ηλεκτρονίων παρατηρήθηκε κατά τη διάρκεια της καταιγίδας της 9ης Μαρτίου. Αυτή η καταιγίδα συνδέθηκε με μια διαπλανητική στεφανιαία εκτίναξη μάζας στις 7 Μαρτίου που έφτασε στη Γη 2 ημέρες αργότερα. Τα άλλα γεγονότα που αναλύθηκαν το 2013 και το 2015 συνέβησαν την ίδια ημέρα του έτους—17 Μαρτίου. Πυροδοτήθηκαν από εκτοξεύσεις μάζας στεμματών, οι οποίες έπληξαν επίσης τη

γήινη μαγνητόσφαιρα την ίδια ώρα της ημέρας. Ωστόσο, και πάλι η παρατηρούμενη απόκριση της ιονόσφαιρας σε αυτά τα γεγονότα ήταν διαφορετική.

2.2. Μελέτη της απόκρισης της ιονόσφαιρας στην X9.3 Flare στις 6 Σεπτεμβρίου 2017 και οι επιπτώσεις της για τις υπηρεσίες πλοήγησης στην Ευρώπη

Το άρθρο αυτό [10] ασχολείται με την απόκριση της ιονόσφαιρας σε μία συγκεκριμένη ηλιακή έκλαμψη. Στις 6 Σεπτεμβρίου 2017, μια έκλαμψη κατηγορίας X μεγέθους 9,3 σημειώθηκε γύρω στο μεσημέρι UT, που είναι το ισχυρότερο συμβάν έκλαμψης της τελευταίας δεκαετίας. Η ίδια η έκλαμψη ήταν το αποκορύφωμα μιας αρκετά ενδιαφέρουσας φάσης αλληλεπιδράσεων ηλιακής-γήινης που προκλήθηκε από την ενεργό περιοχή γνωστή ως ομάδα ηλιακών κηλίδων της Κατάνια 46 ή ενεργή περιοχή με αριθμό 2673 στον κατάλογο NOAA. Από τις 3 έως τις 13 Σεπτεμβρίου 2017 σημειώθηκαν ισχυρές δραστηριότητες εκλάμψεων, συνοδευόμενες από έναν αριθμό ραδιοεκρήξεων και εκτινάξεις μάζας στέμματος κατευθυνόμενες προς τη γη. Οι επιρροές του ηλιακού ανέμου στη Γη ήταν μέτριες κατά τη δραστηριότητα των εκλάμψεων και περιορίστηκαν στις πολικές περιοχές (Linty et al., 2018,). Όμως, η ίδια η ισχυρή έκλαμψη X9.3 είχε επιπτώσεις στην ιονόσφαιρα της ημέρας προκαλώντας ορισμένα προβλήματα στις υπηρεσίες πλοήγησης όπως παρουσιάζουμε σε αυτό το έγγραφο. Τα δεδομένα συμβάντων που παρατηρήθηκαν και αναλύθηκαν μας δίνουν την ευκαιρία να βελτιώσουμε την κατανόησή μας για τα ισχυρά και ακραία καιρικά φαινόμενα στο διάστημα και μας επιτρέπουν να διακρίνουμε την επιρροή των διαφορετικών κατηγοριών συμβάντων στην τεχνολογική μας υποδομή σε περιόδους έντονης ηλιακής δραστηριότητας.

Την τελευταία δεκαετία η συνειδητοποίηση της κοινωνίας μας στο γεγονός ότι ο διαστημικός καιρός αποτελεί σημαντικό κίνδυνο για την τεχνολογική μας υποδομή έχει αυξηθεί συνεχώς. Σε πολλές δημοσιεύσεις, (π.χ. Borries et al., 2009, 2015; Förster & Jakowski, 2000; Lekshmi et al., 2011) η επίδραση των καιρικών φαινομένων στο διάστημα έχει αναλυθεί εστιάζοντας στην απόκριση της ιονόσφαιρας σε σχέση με τα CMEs ως κυρίαρχα απειλή. Πρόσφατα, αρκετοί συγγραφείς εστιάζουν ιδιαίτερα στην απόδοση τοποθέτησης κατά τη διάρκεια ιονοσφαιρικών καταιγίδων (π.χ. Jacobsen & Andalsvik, 2016; Rodriguez-Bilbao et al., 2015). Σε αυτήν την εργασία αναλύθηκαν οι επιπτώσεις της ισχυρής έκλαμψης X9.3 στις 6 Σεπτεμβρίου 2017, μετά τον αντίκτυπό της στην ιονόσφαιρα και τα σοβαρά προβλήματα που προκύπτουν για την ακριβή τοποθέτηση και τις υπηρεσίες υποστήριξης πλοήγησης GNSS. Οι ηλιακές εκλάμψεις μπορούν να επηρεάσουν τις υπηρεσίες πλοήγησης για έως και ώρες προκαλώντας κρίσιμες καταστάσεις σε πολλές εφαρμογές πλοήγησης. Ειδικότερα, οι ηλιακές εκλάμψεις με ισχυρό στοιχείο EUV περίπου 30 nm μπορούν να επηρεάσουν σοβαρά τις υπηρεσίες εντοπισμού θέσης GNSS που χρησιμοποιούνται στην αεροπορία και τη

θαλάσσια ναυσιπλοΐα. Παρέχονται αποτελέσματα από PPP κατά τη διάρκεια της ισχυρής ηλιακής έκλαμψης σε σύγκριση με σχεδόν ήσυχες συνθήκες της προηγούμενης ημέρας για να τονίσουμε την επίδραση των γεγονότων ηλιακής έκλαμψης στην απόδοση τοποθέτησης GNSS. Υποδεικνύεται ότι εκτός από τη διαδικασία PPP (Precise Point Positioning) μπορεί να επηρεαστεί και το σύστημα EGNOS, το οποίο παρέχει υπηρεσίες πλοήγησης για την ασφάλεια της ζωής σε χρήστες αεροπορίας, ναυτιλίας και χερσαίας βάσης. Οι ηλιακές εκλάμψεις μπορεί να οδηγήσουν σε απώλεια σημάτων στον δέκτη GNSS με σημαντική επίδραση στην ακρίβεια και την ακεραιότητα της απόδοσης πλοήγησης. Το AIS, ένα σύστημα που εισήχθη για τη βελτίωση της ασφάλειας στη θάλασσα, μπορεί να είναι μια άλλη πιθανή υπηρεσία που επηρεάζεται κατά τη διάρκεια καιρικών φαινομένων στο διάστημα. Ο αντίκτυπος των διαστημικών καιρικών φαινομένων δεν επηρεάζει μόνο τα υπάρχοντα συστήματα όπως δείξαμε, αλλά πιθανότατα αποτελούν επίσης απειλή για μελλοντικές τεχνολογίες όπως η αυτόνομη πλοήγηση αυτοκινήτων. Εκτός από την προετοιμασία για μελλοντικά σοβαρά ή ακραία διαστημικά καιρικά φαινόμενα, είναι σημαντικό να αντιμετωπιστεί ολόκληρη η αλυσίδα αλληλεπίδρασης του διαστημικού καιρού από τον Ήλιο στη Γη με έμφαση στον αντίκτυπο σε συστήματα και υπηρεσίες. Αυτό θα βοηθήσει στην απόκτηση της απαραίτητης γνώσης για την ευπάθεια της τεχνολογικής μας υποδομής σε διαστημικά καιρικά φαινόμενα.

2.3. Επιδράσεις του σπινθηρισμού φάσης στο σφάλμα θέσης GNSS κατά τη διάρκεια της καταιγίδας του Σεπτεμβρίου 2017 στο Svalbard

Το άρθρο αυτό εξετάζει την ηλιακή καταιγίδα που επηρέασε το Svalbard το 2017 [47]. Στις αρχές Σεπτεμβρίου 2017, αρκετά διαστημικά καιρικά φαινόμενα προκάλεσαν διαταραγμένες συνθήκες του διαστήματος κοντά στη Γη. Ο συνδυασμός δύο αφίξεων εκτίναξης στεφανιαίας μάζας, που σχετίζονται με μια έκλαμψη κατηγορίας X, προκάλεσε μια ισχυρή γεωμαγνητική καταιγίδα στις 7 και 8 Σεπτεμβρίου, προκαλώντας έτσι διάχυτους σπινθηρισμούς ιονόσφαιρας φάσης στα σήματα του Παγκόσμιου Δορυφορικού Συστήματος Πλοήγησης (GNSS). Αυτή η εργασία αναλύει τα αποτελέσματα και τον πραγματικό αντίκτυπο τέτοιων σπινθηρισμών φάσης στα σήματα του μεταιονοσφαιρικού συστήματος παγκοσμίου εντοπισμού θέσης (GPS) και στη σχετική ακρίβεια εντοπισμού θέσης. Η έρευνα επικεντρώνεται ιδιαίτερα σε δεδομένα GPS L1 μεγάλου πλάτους, που καταγράφηκαν κατά τη διάρκεια δοκιμαστικής εκστρατείας στο Svalbard της Νορβηγίας. Η κοινή επίδραση των δορυφόρων σε χαμηλό υψόμετρο και η έκθεση της ιονόσφαιρας στη γεωδιαστημική επιβάρυνση καθιστούν την πλοήγηση ένα κρίσιμο έργο για ένα τόσο απαιτητικό περιβάλλον. Η ανάλυση δεδομένων δείχνει ότι η απόδοση των αλγορίθμων εξομάλυνσης φορέων επηρεάστηκε από την παρουσία μέτριου και ισχυρού σπινθηρισμού φάσης. Σε αυτή τη μελέτη φαίνεται ότι τα σφάλματα εντοπισμού θέσης διπλασιάζονται όταν χρησιμοποιούνται σήματα GPS που επηρεάζονται από

σπινθηρισμό. Αυτή η εργασία δείχνει ότι οι σπινθηρισμοί προκαλούν ένα σημαντικό αποτέλεσμα ομαδοποίησης στις λειασμένες λύσεις τοποθέτησης. Επομένως, προτείνεται μια μεθοδολογία αυτόματης και αυτόνομης ανίχνευσης των ορίων του γεγονότος σπινθηρισμού σύμφωνα με ένα τέτοιο φαινόμενο υψηλού επιπέδου. Η χρήση ραδιοφωνικών δεκτών που καθορίζονται από λογισμικό για αυτόματη λήψη και επεξεργασία δεδομένων GNSS που επηρεάζονται από τον σπινθηρισμό αποτελεί προστιθέμενη αξία στην ανάλυση, καθώς προσφέρει τη δυνατότητα εφαρμογής προηγμένων τεχνικών επεξεργασίας σήματος και βαθύτερη παρατήρηση της επίδρασης των σπινθηρισμών στα σήματα.

2.4. Πρόβλεψη των επιπτώσεων των ηλιακών καταιγίδων στην ιονόσφαιρα με βάση μια σύγκριση δεδομένων ηλιακού ανέμου σε πραγματικό χρόνο με το καταλληλότερο ιστορικό γεγονός καταιγίδας

Αυτή η μελέτη [35] παρουσιάζει μια νέα προσέγγιση μοντελοποίησης που στοχεύει σε μακροχρόνιες προβλέψεις (περισσότερο από 12 ώρες) ιονοσφαιρικών διαταραχών που προκαλούνται από γεγονότα ηλιακών καταιγίδων. Το προτεινόμενο μοντέλο θα λειτουργεί σε ένα λειτουργικό πλαίσιο για να παρέχει γρήγορες και ακριβείς τοπικές προειδοποιήσεις για αυτές τις διαταραχές στο μέλλον. Η προσέγγιση βάσει δεδομένων ηλιακού ανέμου χρησιμοποιεί μια βάση δεδομένων ιστορικών επιπτώσεων ηλιακής καταιγίδας που καλύπτουν δύο ηλιακούς κύκλους για την ανακατασκευή μελλοντικών γεγονότων και που προκύπτουν ιονοσφαιρικές διαταραχές. Τα βασικά στοιχεία του μοντέλου παρουσιάζονται και συζητούνται σε αυτή τη μελέτη και τα δυνατά σημεία της ανακατασκευής που βασίζεται σε ιστορικά γεγονότα παρουσιάζονται δείχνοντας τις καλές συσχετίσεις για την προβλεπόμενη και παρατηρούμενη γεωμαγνητική δραστηριότητα. Τα αρχικά αποτελέσματα σχετικά με την απόκριση της ιονόσφαιρας συζητούνται για όλα τα ιστορικά γεγονότα χρησιμοποιώντας παγκόσμιο συνολικό περιεχόμενο ηλεκτρονίων (GTEC) και λεπτομερέστερα χρησιμοποιώντας χάρτες ολικού περιεχομένου ηλεκτρονίων (TEC) για δύο συγκεκριμένες μελέτες περιπτώσεων (συμπεριλαμβανομένης της γεωμαγνητικής καταιγίδας της Ημέρας του Αγίου Πατρικίου κατά τη διάρκεια της 17ης Μαρτίου 2015). Οι μέσες τιμές ρίζας μέσου τετραγωνικού σφάλματος (RMSE) 3,90 και 5,21 TECU υπολογίζονται για αυτές τις περιπτώσεις επιβεβαιώνοντας καλά αποτελέσματα για την τρέχουσα διαμόρφωση του μοντέλου. Πιθανές μελλοντικές βελτιώσεις των επιμέρους εξαρτημάτων του μοντέλου, καθώς και οι προγραμματισμένες επεκτάσεις και εφαρμογές συζητούνται λεπτομερώς.

2.5. Ανάπτυξη μοντέλου μηχανικής μάθησης για πρόγνωση διαστημικού καιρού στην Ιονόσφαιρα

Στη μελέτη αυτή [12] γίνεται ανάπτυξη μοντέλου πρόγνωσης διαστημικού καιρού στην ιονόσφαιρα. Πολλά πολύπλοκα φυσικά προβλήματα, όπως η πρόβλεψη καιρού

στο διάστημα, έχουν προκλήσεις που πρέπει να προβλεφθούν με ακρίβεια και μακροπρόθεσμα. Μία από τις μεγάλες προκλήσεις για τον διαστημικό καιρό είναι οι μεταβλητές δυναμικές διεργασίες μεταξύ του Ήλιου και της Γης για τις οποίες δεν υπάρχει ακόμα επαρκής κατανόηση όλων των υποκείμενων διεργασιών και σχέσεων. Ακριβή συστήματα πρόβλεψης και έγκαιρης προειδοποίησης χρειάζονται επείγοντως για τη σημερινή κοινωνία, η οποία βασίζεται σε διαστημικές και επίγειες τεχνολογικές υποδομές, οι οποίες είναι ιδιαίτερα ευάλωτες σε ακραία διαστημικά καιρικά φαινόμενα και δεν προστατεύονται επαρκώς. Η προσέγγιση ML έχει τη δυνατότητα να βρει μη γραμμικές συναρτήσεις για να προσεγγίσει τις διαστημικές καιρικές διεργασίες και να προβλέψει την εκδήλωσή τους στο μαγνητικό πεδίο της Γης και στην ιονόσφαιρα. Αυτό το άρθρο εισάγει τη ροή εργασίας της ανάπτυξης μοντέλων ML για την πρόβλεψη του διαστημικού καιρού στην ιονόσφαιρα. Η μελέτη βρίσκεται ακόμη σε εξέλιξη. Τα μοντέλα βασίζονται σε δεδομένα, αποκτώντας γνώση από δεδομένα, τα οποία περιγράφουν την ηλιακή δραστηριότητα, την ταχύτητα του ηλιακού ανέμου, το γεωμαγνητικό πεδίο και την ιονόσφαιρα. Για να αξιοποιήσετε στο έπακρο το ML, απαιτείται να βρείτε και να προετοιμάσετε σχετικά δεδομένα και χρήσιμες λειτουργίες που μπορούν να βελτιώσουν τη μάθηση, ειδικά κατά τη διάρκεια καιρικών φαινομένων στο διάστημα. Όταν αντιμετωπίζετε ένα σύνολο δεδομένων ανισορροπίας, πρέπει να αναπτυχθεί ένα έξυπνο σύστημα που να είναι σε θέση να ξεπεράσει μια τέτοια προκατάληψη. Επιπλέον, θα πρέπει να παρέχεται η αβεβαιότητα της πρόγνωσης του καιρού. Είναι επίσης σημαντικό να κατανοήσουμε και να μπορούμε να εξηγήσουμε τι έχει μάθει το μοντέλο και εάν είναι μεροληπτικό με οποιονδήποτε τρόπο. Η επιθεώρηση του τι έχει μάθει το μοντέλο ML μπορεί επίσης να μας βοηθήσει να ανακαλύψουμε κρυφά μοτίβα και να κατανοήσουμε καλύτερα το πρόβλημα.

2.6. Εφαρμογή αλγορίθμου πρόβλεψης υβριδικών ιονοσφαιρικών δεδομένων TEC με χρήση της μεθόδου PCA-NN

Η πρόβλεψη του καιρού στο διάστημα της ιονόσφαιρας είναι ζωτικής σημασίας για τη βελτίωση της ακρίβειας των παγκόσμιων δορυφορικών συστημάτων πλοήγησης (GNSS). Ωστόσο, η κατανόηση της μη ομοιογενούς ιονοσφαιρικής μεταβλητότητας κάτω από περιβαλλοντικές συνθήκες της διαστημικής γης είναι μια σημαντική πρόκληση, και το ίδιο συμβαίνει και με την ανάπτυξη ενός ακριβούς μοντέλου πρόβλεψης ιονόσφαιρας. Οι πολύπλοκες χωρικές και χρονικές παραλλαγές στην ιονόσφαιρα είναι τα αποτελέσματα των ηλιακών και διαπλανητικών δραστηριοτήτων, εκτός από τις διεργασίες της μαγνητόσφαιρας, της μεσόσφαιρας, της θερμόσφαιρας, της στρατόσφαιρας, της τροπόσφαιρας και της λιθόσφαιρας. Επομένως, αυτό απαιτεί επείγουσα ανάγκη να αναπτυχθεί ένας κατάλληλος αλγόριθμος πρόβλεψης ιονόσφαιρας για την καταγραφή των ιονοσφαιρικών διαταραχών. Το συνολικό περιεχόμενο ηλεκτρονίων (TEC) είναι η βασική παράμετρος

που προέρχεται από δέκτες GNSS για να αντιπροσωπεύει την κατάσταση της ιονόσφαιρας. Αυτή η εργασία [46] εισάγει έναν νέο αλγόριθμο πρόβλεψης ιονόσφαιρας που βασίζεται στη σύντηξη μεθόδων ανάλυσης κύριων συστατικών και τεχνητών νευρωνικών δικτύων (PCA-NN) για την πρόβλεψη των τιμών TEC της ιονόσφαιρας. Για την εφαρμογή μεθοδολογιών τεχνητής νοημοσύνης χρησιμοποιήθηκαν ο ηλιακός δείκτης (F10.7), ο γεωμαγνητικός δείκτης (δείκτης Ap) και δεδομένα TEC 20 ετών (1997–2016) σε σημείο πλέγματος της Ιαπωνίας (34,95 °N και 134,05 °E). Τα πειραματικά αποτελέσματα υπογραμμίζουν την αξιοπιστία του προτεινόμενου αλγορίθμου στην πρόβλεψη των επιπτώσεων χρονοκαθυστέρησης της ιονόσφαιρας.

3. Το σύστημα GNSS

Τα Global Navigation Satellite Systems (GNSS) [49] είναι παγκόσμια δορυφορικά συστήματα πλοήγησης και αναφέρονται σε αστερισμούς δορυφόρων που παρέχουν σήματα από το διάστημα που μεταδίδουν δεδομένα θέσης και χρονισμού σε δέκτες GNSS. Στη συνέχεια, οι δέκτες χρησιμοποιούν αυτά τα δεδομένα για να προσδιορίσουν τη θέση. Εξ ορισμού, το GNSS παρέχει παγκόσμια κάλυψη. Παραδείγματα GNSS περιλαμβάνουν το Galileo της Ευρώπης, το NAVigation Satellite Timing And Ranging Global Positioning System (NAVSTAR GPS) των ΗΠΑ, το GLObal Navigation Satellite System (GLONASS) της Ρωσίας, το δορυφορικό σύστημα πλοήγησης COMPASS της Κίνας, το Quasi-Zenith Satellite System (QZSS) της Ιαπωνίας και το Indian Regional Navigational Satellite System (IRNSS) καθώς και συμπληρωματικά συστήματα SBAS (EGNOS, WAAS) και συνεργασίες με διεθνείς οργανισμούς και υπηρεσίες Τηλεπικοινωνιών, Αεροπλοΐας κα.

Η απόδοση του GNSS αξιολογείται χρησιμοποιώντας τέσσερα κριτήρια:

- Ακρίβεια: η διαφορά μεταξύ της μέτρησης και της πραγματικής θέσης, της ταχύτητας ή του χρόνου ενός δέκτη
- Ακεραιότητα: η ικανότητα ενός συστήματος να παρέχει ένα κατώφλι εμπιστοσύνης και, σε περίπτωση ανωμαλίας στα δεδομένα τοποθέτησης, ένας συναγερμός.
- Συνέχεια: η ικανότητα ενός συστήματος να λειτουργεί χωρίς διακοπή.
- Διαθεσιμότητα: το ποσοστό χρόνου που ένα σήμα πληροί τα παραπάνω κριτήρια ακρίβειας, ακεραιότητας και συνέχειας.

Αυτή η απόδοση μπορεί να βελτιωθεί μέσω περιφερειακών συστημάτων αύξησης μέσω δορυφόρου (SBAS), όπως η Ευρωπαϊκή Υπηρεσία Επικάλυψης Πλοήγησης Γεωστατικών (EGNOS). Το EGNOS βελτιώνει την ακρίβεια και την αξιοπιστία των πληροφοριών GPS διορθώνοντας τα σφάλματα μέτρησης σήματος και παρέχοντας πληροφορίες σχετικά με την ακεραιότητα των σημάτων του.

4. Η ατμόσφαιρα της Γης

4.1. Εισαγωγικές Έννοιες

Η ατμόσφαιρα [50][51][52] είναι ένα στρώμα αερίων που περιβάλλει έναν πλανήτη ή άλλο υλικό σώμα και συγκρατείται από τη βαρύτητα του σώματος.

4.2. Η Ατμόσφαιρα της Γης

Με τον όρο ατμόσφαιρα της Γης εννοούμε το αέριο στρώμα που περιβάλλει τον πλανήτη Γη και συγκρατείται λόγω της βαρύτητάς της. Το όριο ανάμεσα στην ατμόσφαιρα και το διάστημα δεν είναι αυστηρά καθορισμένο. Όσο μεγαλώνει η απόσταση της από τη Γη η ατμόσφαιρα σταδιακά εξασθενεί.

«Ατμόσφαιρα είναι το αεριώδες περίβλημα της Γης, που συγκρατούμενο από τη δύναμη της βαρύτητας αποτελεί ένα σώμα με τη Γη και συμμετέχει σε όλες τις κινήσεις της.»

Αστεριάδης, 1983

Η ατμόσφαιρα της Γης αποτελείται κατά 78,08 % από άζωτο, 20,95% από οξυγόνο, 0,93% από αργό, 0,0395% από διοξείδιο του άνθρακα και από ίχνη άλλων αερίων όπως, ευγενή αέρια, ίχνη υδρογόνου, όζοντος κτλ. Στην ατμόσφαιρα επίσης αιωρούνται σχεδόν πάντα και μόρια κονιορτού, καπνού, άλατος κτλ., καθώς και υδρατμοί από την εξάτμιση θαλασσών, λιμνών, ποταμών κτλ.

Το όριο ανάμεσα στη γήινη και την ηλιακή ατμόσφαιρα δεν είναι ορισμένο με ακρίβεια: Το εξωτερικό σύνορο της ατμόσφαιρας αντιστοιχεί στην απόσταση στην οποία τα μόρια των ατμοσφαιρικών αερίων δεν υπόκεινται πλέον στην έλξη της γης και την επίδραση του μαγνητικού της πεδίου. Οι συνθήκες αυτές ισχύουν σε ένα υψόμετρο που ποικίλει ανάλογα με το γεωγραφικό μήκος και συνήθως είναι στα 60 χλμ. πάνω από τους πόλους και στα 30 χλμ. πάνω από τον ισημερινό. Αυτά τα νούμερα είναι απλά ενδεικτικά: το γήινο μαγνητικό πεδίο παραμορφώνεται διαρκώς από τον ηλιακό άνεμο. Η πυκνότητα της ατμόσφαιρας, επίσης, παρουσιάζει σημαντικές αυξομειώσεις.

4.3. Στρώματα

Η χημική σύνθεση της ατμόσφαιρας μέχρι το ύψος των 80-100 χλμ. είναι σχεδόν αμετάβλητη. Ανάλογα όμως της μεταβολής της θερμοκρασίας διακρίνονται σε αυτή τα ακόλουθα στρώματα:

- Τροπόσφαιρα, από ύψος 0 μέχρι 9-18 χλμ. όπου και η τροπόπαυση. Χαρακτηρίζεται από σχεδόν ομοιόμορφη ελάττωση της θερμοκρασίας με το ύψος, αύξηση της ταχύτητας του ανέμου με το ύψος, σημαντική περιεκτικότητα σε υδρατμούς, ιδιαίτερα στα κατώτερα στρώματα,

- κατακόρυφες κινήσεις του αέρα και συνεχή εναλλαγή ατμοσφαιρικών φαινομένων. Το ύψος της τροπόπαυσης παρουσιάζει εποχική μεταβολή και είναι μεγαλύτερο κατά το τέλος του καλοκαιριού και στις αρχές του φθινοπώρου και μικρότερο στο τέλος του χειμώνα και στις αρχές της άνοιξης.
- Στρατόσφαιρα, από την τροπόπαυση μέχρι τα 50 χλμ. όπου και η στρατόπαυση. Από την τροπόπαυση μέχρι τη στάθμη των 35 χιλιομέτρων, η θερμοκρασία δεν παρουσιάζει ουσιαστική μεταβολή με το ύψος. Χαρακτηρίζεται από μεγάλη ξηρασία και ευστάθεια. Οι κατακόρυφες κινήσεις των αέριων μαζών είναι σημαντικά περιορισμένες ενώ αντίθετα υπάρχει αξιόλογη οριζόντια κυκλοφορία με ισχυρούς ανέμους. Από τα 35 μέχρι τα 50-55 χιλιόμετρα, η θερμοκρασία αυξάνει με το ύψος και φτάνει τους 15°C περίπου. Η αύξηση της θερμοκρασίας στο τμήμα αυτό οφείλεται στην απορρόφηση μεγάλου μέρους της ηλιακής υπεριώδους ακτινοβολίας από το όζον του οποίου η μεγαλύτερη συγκέντρωση εμφανίζεται στα 20-40 χιλιόμετρα.
 - Μεσόσφαιρα, από την στρατόπαυση μέχρι τα 80 χλμ. όπου και η μεσόπαυση. Είναι το στρώμα στο οποίο η θερμοκρασία ελαττώνεται μέχρι το ύψος των 80-85 χιλιομέτρων και φτάνει τους -90°C. Η ελάττωση της θερμοκρασίας οφείλεται εκτός άλλων και στην απουσία όζοντος. Μεσόπαυση, που είναι η διαχωριστική ζώνη ανάμεσα στη μεσόσφαιρα και το επόμενο στρώμα και η θερμοκρασία σε αυτή κατεβαίνει μέχρι -150°C, δηλαδή είναι η ψυχρότερη περιοχή της ατμόσφαιρας.
 - Θερμόσφαιρα ή Ιονόσφαιρα, από την μεσόπαυση μέχρι 800 χλμ. όπου η θερμόπαυση. Χαρακτηρίζεται από συνεχή αύξηση της θερμοκρασίας που πολλές φορές φτάνει τους 1500°C. Οι θερμοκρασίες που σημειώνονται στη θερμόσφαιρα επηρεάζονται από το πλάτος, την ώρα της ημέρας και την ηλιακή δραστηριότητα. Κατά το ελάχιστο των ηλιακών κηλίδων, στα μέσα πλάτη, η μέση μέγιστη θερμοκρασία είναι περίπου 750°C και κατά το μέγιστο των κηλίδων είναι περίπου 1300°C. Όταν σημειώνονται εκλάμψεις (flares) στον Ήλιο η θερμοκρασία φτάνει τους 1700°C.
 - Εξώσφαιρα, από θερμόπαυση μέχρι 3.500 χλμ.

Ο συνδυασμός των φωτεινών και λοιπών ακτινοβολιών του Ήλιου, η διαφορετική πυκνότητα των στρωμάτων της ατμόσφαιρας και τα διαφορετικά φυσικά χαρακτηριστικά (όπως η θερμοκρασία, η πίεση και η υγρασία), η εκάστοτε μορφή των υδρατμών (αέρια, υγρά και στερεά) στην ατμόσφαιρα όπως και άλλοι παράγοντες π.χ. ιόντα αερίων, προκαλούν σειρά από διάφορα οπτικά, ακουστικά και ηλεκτρικά φαινόμενα.

5. Ιονόσφαιρα

5.1. Εισαγωγικές Έννοιες

Η ιονόσφαιρα [53][54] είναι τμήμα της ανώτερης ατμόσφαιρας που εκτείνεται από περίπου τα 50 έως τα 1000 χιλιόμετρα πάνω από την επιφάνεια της Γης και αποτελείται από ηλεκτρόνια και ιόντα. Λειτουργεί τόσο σαν μέσο διάδοσης του ηλεκτρισμού όσο και σαν μερικώς ιονισμένο αέριο που, ως τέτοιο, υπόκεινται στους νόμους της φυσικής του πλάσματος. Το ιονοσφαιρικό πλάσμα λόγω των ηλεκτρομαγνητικών του ιδιοτήτων αλληλεπιδρά με το μαγνητικό πεδίο της Γης και ως ένα βαθμό καθορίζεται από αυτό. Είναι σημαντικό να αναφερθεί ότι ακόμα και στο τμήμα της ιονόσφαιρας που εμφανίζει τη μεγαλύτερη συγκέντρωση ιονισμένων σωματιδίων, που τοποθετείται μεταξύ 250 και 400 χιλιομέτρων, αυτή είναι περίπου τρεις τάξεις μεγέθους μικρότερη από τη συγκέντρωση των ουδέτερων σωματιδίων. Αυτό σημαίνει ότι και αυτά παίζουν πολύ σημαντικό ρόλο στον σχηματισμό της.

Οι φυσικές και χημικές διεργασίες που οδηγούν στο σχηματισμό της ιονόσφαιρας είναι αρκετά διαφορετικές όσο μεταβάλλεται το ύψος και για το λόγο αυτό χωρίζεται σε περιοχές. Βασικός παράγοντας ιονισμού της ιονόσφαιρας είναι η υπεριώδης ηλιακή ακτινοβολία με μήκος κύματος 102.7 nm η οποία απορροφάται από τα μόρια και άτομα της ατμόσφαιρας. Υπάρχουν και άλλοι παράγοντες ιονισμού όπως οι ακτίνες X, οι κοσμικές ακτίνες και η ρέουσα σωματιδιακή ακτινοβολία, αλλά έχουν σποραδικό χαρακτήρα που συνδέεται με την ηλιακή δραστηριότητα (Pulinets & Boyarchuk, 2004).

5.2. Δομή

Η Ιονόσφαιρα αποτελείται από επιμέρους στρώματα:

- Περιοχή D Η περιοχή D βρίσκεται σε ύψος μεταξύ 50 και 90 χιλιομέτρων, συχνά το τμήμα μεταξύ 50 και 70 χιλιομέτρων καλείται περιοχή C και σχηματίζεται κυρίως από την επίδραση κοσμικών ακτίνων. Τα ουδέτερα μόρια της περιοχής αυτής είναι τα N₂, O₂, Ar, CO₂, He καθώς και έντονα μεταβαλλόμενες ποσότητες O₃ και H₂O (Zolesi & Cander, 2014). Κυριότεροι παράγοντες ιονισμού είναι οι ακτίνες Lyman-α με λ=121.6 nm που μπορούν να ιονίσουν τις μικρές ποσότητες NO που υπάρχουν στην ατμόσφαιρα και να παράγουν ιόντα NO⁺ στα ύψη μεταξύ 60 και 90 χιλιομέτρων. Επίσης ηλιακές και γαλακτικές ακτίνες X (λ<λ<λ<λ<λ και ρέοντα φορτισμένα σωματίδια μεταβάλλουν ενεργειακά τα μόρια O₂ και N₂ και στη συνέχεια η υπεριώδης ακτινοβολία με 102.8<λ<111.8 ιονίζει τα ήδη διεγερμένα μόρια O₂(1Δg). Χαρακτηρίζεται από το γεγονός ότι εντός της δημιουργούνται τα πιο σύνθετα ιόντα. Περιέχει μεγάλα συμπλέγματα ιόντων όπως H⁺(H₂O)_n, NO⁺(H₂O)_n, NO⁺(CO₂)_n και άλλα. Λόγω της δημιουργίας ανιόντων, είναι η μόνη περιοχή

της ιονόσφαιρας όπου η συγκέντρωση κατιόντων δεν ισούται με τη συγκέντρωση των ηλεκτρονίων. Σε ύψη χαμηλότερα των 75 χιλιομέτρων ο αριθμός των ανιόντων είναι πολύ μεγαλύτερος από εκείνον των ηλεκτρονίων. Κατά τη διάρκεια του χειμώνα η περιοχή D εμφανίζει μεγαλύτερη συγκέντρωση ηλεκτρονίων σε σχέση με το καλοκαίρι και αυτό οφείλεται στο γεγονός ότι το χειμώνα δεν δημιουργούνται ιόντα NO^+ . Η ημερήσια μεταβολή της συγκέντρωσης των ηλεκτρονίων μπορεί να κυμαίνεται από 1.5 τάξη μεγέθους σε περιόδους ήπιας ηλιακής δραστηριότητας μέχρι 2.5 τάξεις μεγέθους σε περιόδους έντονης ηλιακής δραστηριότητας. Η περιοχή D σύμφωνα με το διαχωρισμό της ατμόσφαιρας που περιγράφεται σε προηγούμενο κεφάλαιο αντιστοιχεί στη μεσόσφαιρα, όπου απαντάται αντιστροφή της θερμοκρασίας από τους -13°C στους -93°C για ύψη από 50 έως 85 χιλιόμετρα, μετά η θερμοκρασία αυξάνει σταθερά (Pulinets & Boyarchuk, 2004).

- Περιοχή E Εντοπίζεται σε ύψη 90 έως 140 χιλιομέτρων. Βασικός παράγοντας ιονισμού της περιοχής E είναι η υπεριώδης ακτινοβολία με μήκη κύματος $80 < \lambda < 102.8 \text{ nm}$. Κυριότερες είναι οι ακτίνες $\text{L}\beta = 102.6 \text{ nm}$ και $\text{CIII} = 97.7 \text{ nm}$ που ιονίζουν τα μόρια O_2 και οι ακτίνες X με $1 < \lambda < 10 \text{ nm}$. Εδώ επικρατούν πρωτίστως τα ιόντα O_2^+ και NO^+ και δευτερευόντως τα O^+ και N_2^+ . Βασικός παράγοντας απώλειας φορτισμένων σωματιδίων στην περιοχή αυτή είναι ο διασπαστικός ανασυνδυασμός (Dissociative Recombination, DR) μοριακών ιόντων με τα ηλεκτρόνια. Η συμπεριφορά της περιοχής E είναι κανονική και εξαρτάται από τη ζενίθια γωνία του Ήλιου, με το μέγιστο να εντοπίζεται κατά τη μεσουράνηση του Ήλιου σε κάθε τόπο, από την εποχή, με το μέγιστο να εντοπίζεται το καλοκαίρι και από τον ενδεκαετή ηλιακό κύκλο (Zolesi & Cander, 2014). Η δημιουργία και απώλεια ιόντων βασίζεται σε φωτοχημικές διεργασίες. Ωστόσο κατά την ώρα που δύνει ο Ήλιος και τη νύχτα, παράγοντες όπως η φωτεινή πλευρά της εξόχως απόκεντρης περιοχής της γήινης ατμόσφαιρας (geocorona) και οι ηλιακές ακτίνες των γραμμών υδρογόνου με $\lambda = 121.6 \text{ nm}$ & $\lambda = 102.6 \text{ nm}$ και ηλίου με $\lambda = 58.4 \text{ nm}$ & $\lambda = 30.4 \text{ nm}$ περιορίζουν τις πηγές ιονισμού της περιοχής αυτής. Εντός της περιοχής E μπορεί να δημιουργηθεί ένα λεπτό στρώμα, που ονομάζεται σποραδικό στρώμα Es, λόγω της σύγκλισης της κάθετης ροής των μεταλλικών ιόντων. Αυτό μπορεί να περιλαμβάνει εντός του διάφορα ανομοιογενή στρώματα. Πειραματικά καταγράφηκαν στο Es τα παρακάτω ιόντα $\text{Na}^+(23^+)$, $\text{Mg}^+(24^+)$, $\text{Mg}^+(26^+)$, $\text{Mg}^+(27^+)$, $\text{Al}^+(27^+)$, $\text{Si}^+(29^+)$, $\text{K}^+(39^+)$, $\text{Ca}^+(40^+)$, $\text{K}^+(41^+)$, $\text{Ca}^+(44^+)$, $\text{Ti}^+(48^+)$, $\text{Fe}^+(54^+)$, $\text{Mn}^+(55^+)$, $\text{Fe}^+(56^+)$, $\text{Fe}^+(57^+)$, $\text{Fe}^+(58^+)$, $\text{Ni}^+(58^+)$, $\text{Co}^+(59^+)$, $\text{Ni}^+(60^+)$, $\text{Ni}^+(61^+)$, $\text{Ni}^+(62^+)$, $\text{Ni}^+(64^+)$, $\text{Xe}^+(66^+)$, οξείδια και υδρίτες μεταλλικών ιόντων AlO^+ , SiO^+ , MgO^+ , SiOH^+ , $\text{Na}^+(\text{H}_2\text{O})$, $\text{Na}^+(\text{H}_2\text{O}_2)$, NaO^+ , NaOH^+ . Τα ιόντα αυτά

προέρχονται κατά κύριο λόγο από μετεωρίτες, όμως υπάρχει πιθανότητα να προέρχονται και από την επιφάνεια της Γης. Η μεταφορά των ιόντων για το σχηματισμό του σποραδικού στρώματος Es γίνεται κατά κύριο λόγο μέσω ανέμων, παλιρροιών, βαρυτικών κυμάτων και ηλεκτρικών πεδίων. Το πάχος του σποραδικού στρώματος ποικίλει από εκατοντάδες μέτρα μέχρι λίγα χιλιόμετρα και παρουσιάζει μεγαλύτερη πυκνότητα στα άκρα, μίας ή δύο τάξεων μεγέθους υψηλότερη, από ότι αναμένεται σε ένα στρώμα της ιονόσφαιρας. Η δομή του είναι ακανόνιστη τόσο κατά την οριζόντια, λόγω της αστάθειας των κυμάτων που συμμετέχουν στο σχηματισμό του, όσο και κατά την κατακόρυφη διεύθυνση, λόγω της ύπαρξης εντός του επιμέρους διαφορετικών στρωμάτων. Η πυκνότητα των ηλεκτρονίων στο στρώμα αυτό μπορεί να είναι πολύ υψηλή και μερικές φορές είναι υψηλότερη από το αντίστοιχο μέγιστο της περιοχής F (Pulinets & Boyarchuk, 2004).

- Περιοχή F Η περιοχή F είναι το πιο σημαντικό τμήμα της ιονόσφαιρας λόγω της επίδρασής της στη διάδοση των ραδιοκυμάτων και αποτελείται από δύο στρώματα τα F1 και F2.
 - Το στρώμα F1 εμφανίζεται ως κοίτασμα ή ως σημείο καμψής της πυκνότητας των ηλεκτρονίων κατά την κατακόρυφη διεύθυνση, μεταξύ των στρωμάτων E και F2 σε ύψη 160-200 χιλιομέτρων. Το ύψος του εξαρτάται από την ηλιακή και γεωμαγνητική δραστηριότητα, καθώς και την εποχή και σπάνια φτάνει πάνω από τα 180 χιλιόμετρα υψόμετρο. Ο σχηματισμός του οφείλεται στην επίδραση της υπεριώδους ακτινοβολίας με μήκη κύματος $10 < \lambda < 90$ nm, που ιονίζει τα N₂, O, και σε μικρότερο βαθμό το O₂. Τα ιόντα που επικρατούν στο στρώμα αυτό είναι τα ίδια με της περιοχής E, δηλαδή NO⁺ και O₂⁺. Ο αριθμός των ιόντων που προκύπτει από φωτοχημικές διεργασίες είναι μεγαλύτερος σε σχέση με αυτά που προέρχονται από μεταφορά, τουλάχιστον κατά τη διάρκεια της ημέρας. Χαρακτηρίζεται από ανισοτροπία της θερμοκρασίας ιόντων ηλεκτρονίων. Εμφανίζεται συνήθως το καλοκαίρι κατά τη διάρκεια της ημέρας, ενώ κατά τη διάρκεια της νύχτας αποτελεί μία κοιλότητα μεταξύ των στρωμάτων E και F2 που στη φυσική της ιονόσφαιρας ονομάζεται κοιλάδα.
 - Το στρώμα F2 είναι το πιο δυναμικό και πυκνό, από την άποψη της συγκέντρωσης φορτισμένων σωματιδίων, στρώμα της ιονόσφαιρας. Το μέγιστο της συγκέντρωσης, αναλόγως των γεωφυσικών συνθηκών, εντοπίζεται σε ύψη μεταξύ 250 και 500 χιλιομέτρων. Με βάση αυτό το μέγιστο χωρίζεται σε δύο τμήματα. Το κατώτερο που έχει μικρό πάχος και εκτείνεται από τα 200 χιλιόμετρα μέχρι το μέγιστο της συγκέντρωσης και καθορίζεται από φωτοχημικές διεργασίες και το

ανώτερο που φτάνει μέχρι τα 1000 ή και τα 2000 χιλιόμετρα και καθορίζεται από τη διαδικασία της διάχυσης. Επηρεάζει καθοριστικά την διάδοση των ραδιοκυμάτων με μήκος κύματος δέκατα του εκατοστού και είναι η κύρια πηγή των σφαλμάτων που εντοπίζονται στις μετρήσεις των συστημάτων πλοήγησης. Χαρακτηρίζεται από έντονες μεταβολές που σε χρονική κλίμακα κυμαίνονται από τα ένδεκα χρόνια του ηλιακού κύκλου ή και περισσότερο μέχρι μερικά δευτερόλεπτα λόγω της αλληλεπίδρασης με την πλασμόσφαιρα σε υψόμετρα μεγαλύτερα των 1000 χιλιομέτρων. Η αναπλήρωση σε φορτισμένα σωματίδια που προέρχονται από την πλασμόσφαιρα παίζει πολύ σημαντικό ρόλο στη διατήρηση της περιοχής F κατά τη διάρκεια της νύχτας. Βασικός παράγοντας ιονισμού, όπως και στο στρώμα F1, είναι η υπεριώδης ηλιακή ακτινοβολία με $10 < \lambda < 90$ nm. Κυρίως ιονίζονται άτομα οξυγόνου O. Κυριαρχούν τα ιόντα O^+ , ενώ σε ύψη μεγαλύτερα των 700 χιλιομέτρων αρχίζουν να επικρατούν τα ελαφρύτερα ιόντα ηλίου He^+ και υδρογόνου H^+ . Ο σχηματισμός του στρώματος F2 δεν εξαρτάται από το ύψος του ήλιου ούτε από το γεωγραφικό πλάτος. Ωστόσο μετά την ανατολή του ηλίου η συγκέντρωση των ηλεκτρονίων αυξάνει με γοργούς ρυθμούς φθάνοντας τη μέγιστη τιμή της κάποια στιγμή μέσα στη διάρκεια της ημέρας. Η σύνθετη συμπεριφορά που παρουσιάζει η περιοχή αυτή οφείλεται σε χημικές μεταβολές αφού η συγκέντρωση των ηλεκτρονίων εξαρτάται από την αναλογία $[O]/[N_2]$ και την αναλογία $[O]/[O_2]$, στη διαφορά της θερμοκρασίας μεταξύ ημέρας και νύχτας, στους ανέμους σε ύψη γύρω στα 300 χιλιόμετρα με ταχύτητες από δεκάδες μέχρι εκατοντάδες m/s, που σε μέσα πλάτη κατευθύνονται προς τους πόλους κατά τη διάρκεια της μέρας και προς τον ισημερινό κατά τη διάρκεια της νύχτας και στα ηλεκτρικά πεδία που αναπτύσσονται στα ανώτερα στρώματα της ατμόσφαιρας και μεταβάλλουν την κατακόρυφη κατανομή των φορτισμένων σωματιδίων. Η επίδραση των ηλεκτρικών πεδίων είναι σημαντικότερη κοντά στον ισημερινό, ενώ η επίδραση των ανέμων είναι σημαντικότερη στα μέσα γεωγραφικά πλάτη. Σημαντικό ρόλο στην απώλεια ηλεκτρονίων λόγω της διαδικασία του διασπαστικού ανασυνδυασμού (DR) παίζουν τα μόρια N_2 και O_2 , ενώ στη δημιουργία τους τα άτομα O. Τα ηλεκτρόνια που δημιουργούνται από τη διαδικασία ιονισμού είναι θερμότερα από τα ουδέτερα μόρια από τα οποία προήλθαν. Αυτή η παραπάνω ενέργεια μεταβιβάζεται σταδιακά στα θετικά ιόντα όχι όμως και στα ουδέτερα μόρια. Έτσι η

Θερμοκρασία του πλάσματος είναι μεγαλύτερη από εκείνη του ουδέτερου αέρα. Κατά τη διάρκεια της ημέρας η θερμοκρασία των ηλεκτρονίων είναι μεγαλύτερη από εκείνη των ιόντων, ενώ κατά τη διάρκεια της νύχτας οι θερμοκρασίες περίπου εξισώνονται. Όταν το πλάσμα είναι θερμό ανεβαίνει σε μεγαλύτερα υψόμετρα που η διαδικασία του ανασυνδυασμού επιβραδύνεται και επομένως διατηρεί την πυκνότητα του. Όταν η θερμοκρασία του πέφτει κατά τη διάρκεια της νύχτας, κατεβαίνει πάλι χαμηλότερα και συμβάλει στη διατήρηση της περιοχής F παρά τον ανασυνδυασμό που λαμβάνει χώρα στα χαμηλότερα υψόμετρα (Zolesi & Cander, 2014).

5.3. Tec

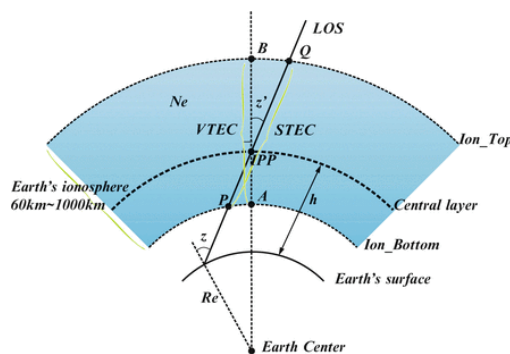
Η συνολική περιεκτικότητα ηλεκτρονίων (ή TEC) [55] είναι μια σημαντική περιγραφική ποσότητα για την ιονόσφαιρα της Γης. Το TEC είναι ο συνολικός αριθμός ηλεκτρονίων ανά m^2 που περιέχονται μεταξύ δύο σημείων, κατά μήκος ενός σωλήνα κυλινδρικής διατομής $1 m^2$. Συχνά μετριέται σε μονάδες TEC, δηλαδή σε $TECU=10^{16}$ electrons/ m^2 .

Το TEC είναι σημαντικό για τον προσδιορισμό των εξάρσεων και των καθυστερήσεων φάσης ενός ραδιοκύματος μέσω ενός μέσου. Τα δεδομένα TEC χαρακτηρίζονται από την παρατήρηση καθυστερήσεων φάσης των φορέων των λαμβανόμενων ραδιοσημάτων που μεταδίδονται από δορυφόρους που βρίσκονται πάνω από την ιονόσφαιρα. Το TEC επηρεάζεται έντονα από την ηλιακή δραστηριότητα.

Εξαρτάται, εξ ορισμού, από τη διαδρομή και μπορεί να υπολογιστεί ενσωματώνοντας την κατά μήκος διαδρομή d_s μέσω της ιονόσφαιρας με την εξαρτώμενη από τη θέση πυκνότητα ηλεκτρονίων $n_e(s)$:

$$TEC = \int n_e(s) ds$$

Μπορεί να μετρηθεί είτε κατακόρυφα είτε λοξά και χαρακτηρίζεται ως VTEC ή STEC αντίστοιχα.



Εικόνα 1. STEC/ VTEC (Πηγή: Shuanggen Jin, Estel Cardellach, Feiqin Xie, Ground GNSS Ionosphere Sounding, 2013)

Η ιονοσφαιρική καθυστέρηση είναι ανάλογη της πυκνότητας των ηλεκτρονίων και αντιστρόφως ανάλογη του τετραγώνου της συχνότητας f του φορέα.

$$\tau_p^{\text{iono}} = -\kappa \frac{\text{TEC}}{f^2}$$

όπου

$$\kappa = q^2 / (8\pi^2 m_e \epsilon_0) = c^2 r_e / (2\pi)$$

όπου:

- q : το φορτίο των ηλεκτρονίων
- m_e : η μάζα
- r_e : η ακτίνα
- c : η ταχύτητα του φωτός
- ϵ_0 : είναι η διαπερατότητα του κενού

Η τιμή της σταθεράς είναι $\kappa \approx 40,308193 \text{ m}\cdot\text{s}^{-2}\text{TECU}^{-1}$.

5.4. Παράμετροι

Όπως αναφέρθηκε και παραπάνω οι παράμετροι που επηρεάζουν την ιονόσφαιρα είναι οκτώ και αφορούν γεωμαγνητική και ηλιακή δραστηριότητα. Αναλυτικότερα:

Πίνακας 1. Παράμετροι επηρεασμού της ιονόσφαιρας.

α/α	Δραστηριότητα	Παράμετρος	Δείκτης	Περίοδος καταγραφής
1	Μαγνητική δραστηριότητα	Geomagnetic activity (K)	Kp	Ανά 3 ώρες
2			ap	
3			D	
4		SML		

5		Auroral Electrojet (AE) /SME	SMU	
6		Disturbance storm time (Dst)	-	Ωριαία
7	Ηλιακή δραστηριότητα	Solar Radio Flux	Fluxtime	Ανά 3 ώρες
8			Fluxjulian	
9			Fluxcarrington	
10			Fluxobsflux	
11			fluxadjflux	
12			fluxursi	
13		Sunspot Number	Daily Total Sunspot Number	
14			Daily North Sunspot Number	
15			Daily South Sunspot Number	
16			Standard Deviation of Raw Daily Total Sunspot Data	
17			Standard Deviation of Raw Daily North Sunspot Data	
18			Standard Deviation of Raw Daily South Sunspot Data	
19			Number of Observations for Daily Total Sunspot Number	
20	Number of Observations for Daily North Sunspot Number			

21			Number of Observations for Daily South Sunspot Number	
22			Definitive/Provisional Indicator	
23		Energetic Particle Flux	GOES-15 Proton FluEncE >1 MEV	Ημερήσια
24			GOES-15 Proton FluEncE >10 MEV	
25			GOES-15 Proton FluEncE >100 MEV	
26			GOES-15 EIEctron FluEncE >0.8 MEV	
27			GOES-15 EIEctron FluEncE >2 MEV	
28			NEutron Monitor % of bkgd	
29			Solar Flares	
30		Coronal Mass Ejections	Central PA	
31			Angular	
32			Linear	
33			2nd-order Speed at final height	
34			2nd-order Speed at 20 Rs	
35			Acceleration	
36			Mass	
37			Kinetic Energy	
38			MPA	

5.4.1. Geomagnetic activity (K)

Η παράμετρος γεωμαγνητικής δραστηριότητας (K) αντιπροσωπεύει τη γεωμαγνητική δραστηριότητα που οφείλεται στην είσοδο σωματιδίων ηλιακής ακτινοβολίας στην

μαγνητόσφαιρα. Η έκφραση του δείκτη αυτού γίνεται με την ανάθεση ενός ακέραιου αριθμού από το 0 έως το 9 σε κάθε διάστημα 3 ωρών του παγκόσμιου χρόνου σε μία ημέρα. Ο δείκτης Kp αφορά και αυτός διαστήματα 3 ωρών και είναι ο κανονικοποιημένος δείκτης K από δεκατρία γεωμαγνητικά παρατηρητήρια τα οποία έχουν επιλεγθεί ειδικά να βρίσκονται σε γεωγραφικά πλάτη ανάμεσα στις 44 και 60° στο Βόρειο και το Νότιο ημισφαίριο. Στην περίπτωση του δείκτη Kp, η κλίμακα από 0 έως 9 του δείκτη K έχει υποδιαιρέσεις του ενός τρίτου. Για παράδειγμα, ο δείκτης 5- αντιπροσωπεύει το 4 και 2/3, ο δείκτης 5o σημαίνει 5 και ο δείκτης 5+ είναι 5 και 1/3. Εφόσον οι δείκτες K και Kp βασίζονται σε λογαριθμική κλίμακα, έχει εισαχθεί και ο αντίστοιχος γραμμικός δείκτης ap, ο οποίος προκύπτει από τον Kp με τον ακόλουθο τρόπο:

Πίνακας 2. Αντιστοιχία του δείκτη Kp με τον δείκτη ap

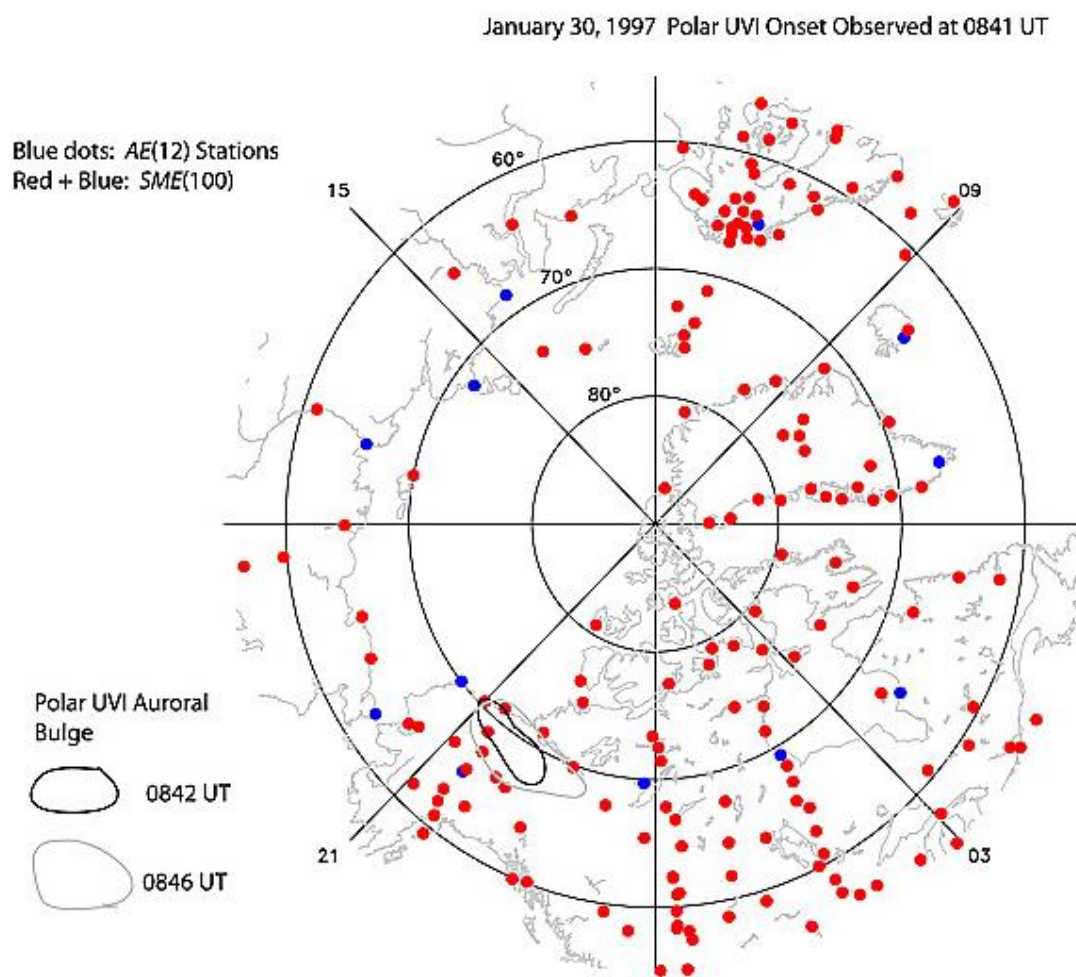
Kp	0o	0+	1-	1o	1+	2-	2o	2+	3-	3o	3+	4-	4o	4+
ap	0	2	3	4	5	6	7	9	12	15	18	22	27	32
Kp	5-	5o	5+	6-	6o	6+	7-	7o	7+	8-	8o	8+	9-	9o
ap	39	48	56	67	80	94	111	132	154	179	207	236	300	400

Ο μέσος όρος των οκτώ δεικτών ap για μία ημέρα μας δίνει τον ημερήσιο δείκτη Ap. Ο γεωμαγνητικός δείκτης Ap λοιπόν, παρέχει ένα μέσο όρο της γεωμαγνητικής δραστηριότητας ημερησίως, δηλαδή τις ημέρες με εντονότερη γεωμαγνητική δραστηριότητα και οι τιμές του Ap είναι μεγαλύτερες, και θα τον μελετήσουμε στην εργασία αυτή καθώς είναι πολύ χρήσιμος για τη μελέτη της διαμόρφωσης της κοσμικής ακτινοβολίας. Τέλος, ο δείκτης D υποδεικνύει εάν οι τιμές Kp και ap είναι οριστικές ($D = 1$) ή προκαταρκτικές ($D=0$)

5.4.2. Auroral Electrojet (AE)

Ο δείκτης Auroral Electrojet (σελαϊκών ηλεκτροχειμάρρων) εισήχθη αρχικά από τους Davis και Sugiura το 1966 ως μέτρο της παγκόσμιας δραστηριότητας των ηλεκτροχειμάρρων στη σελαϊκή ζώνη. Παρέχει μια ποσοτική μέτρηση της μαγνητικής δραστηριότητας στη ζώνης του σέλαος που παράγεται από ενισχυμένα ιονοσφαιρικά ρεύματα που ρέουν κάτω και μέσα στο σέλας (Δαγκλής Ι., Σημειώσεις Διαστημικής Φυσικής, 2018). Προέρχεται από διακυμάνσεις του γεωμαγνητικού πεδίου στη συνιστώσα H που μετράται σε 12 σταθμούς κατά μήκος της ζώνης του σέλαος στο βόρειο ημισφαίριο. Ο δείκτης Auroral Electrojet $AE(12) = AU(12) - AL(12)$, όπου AU και AL είναι τα ανώτερα και κατώτερα συστατικά του AE (οι μεγαλύτερες και μικρότερες τιμές της συνιστώσας H μεταξύ των μαγνητικών σταθμών που χρησιμοποιούνται).

Ως εξέλιξη του δείκτη AE που χρησιμοποιείται παραδοσιακά, ο δείκτης SME που εισήχθη από τους συνεργάτες του SuperMAG διαφέρει από τον παραδοσιακό δείκτη AE κυρίως από τον αριθμό των σταθμών που συμμετείχαν (100). Όλα τα δεδομένα μετατρέπονται πρώτα σε συντεταγμένες με τη συνιστώσα H να δείχνει προς τον τοπικό μαγνητικό βορρά και η γραμμή βάσης αφαιρείται. Τότε ο σταθμός με τη μεγαλύτερη τιμή H συνεισφέρει στην τιμή SMU, ενώ ο σταθμός με τη μικρότερη τιμή συνεισφέρει στην τιμή SML. Ο δείκτης SME είναι $SME = SMU - SML$ (Newell and Gjerloev (2012)).



Εικόνα 2. Δίκτυο σταθμών AE (Μπλε χρώμα) και SME (Κόκκινο χρώμα)

5.4.3. Disturbance storm time (Dst)

Ο δείκτης χρόνου διαταραχής καταιγίδας (Dst) χρησιμοποιείται για ανάλυση της ισχύος και της διάρκειας των γεωμαγνητικών καταιγίδων. Το Dst είναι ένα μέτρο της μείωσης της οριζόντιας συνιστώσας του μαγνητικού πεδίου της Γης κοντά στον μαγνητικό ισημερινό λόγω της αύξησης του ρεύματος του μαγνητοσφαιρικού δακτυλίου (δείτε τις Συχνές ερωτήσεις παρακάτω). Τιμές μικρότερες από -50 nanotesla (nT) υποδηλώνουν υψηλή γεωμαγνητική δραστηριότητα.

Ο αρχικός δείκτης Dst παρέχεται από το Παγκόσμιο Κέντρο Δεδομένων για τον Γεωμαγνητισμό, Κιότο, Ιαπωνία. Ο δείκτης παράγεται σε ωριαίο ρυθμό χρησιμοποιώντας χρονική και χωρική μέση τιμή από τέσσερις επίγειους σταθμούς παρατήρησης χαμηλού γεωγραφικού πλάτους.

5.4.4. Solar Radio Flux

Οι διάφορες εκδηλώσεις της ηλιακής δραστηριότητας καθοδηγούνται από τη συνολική ποσότητα μαγνητικής ροής που αναδύεται μέσω της φωτόσφαιρας στη χρωμόσφαιρα και το στέμμα και τη χρονική και χωρική κατανομή της. Για λόγους που δεν είναι σαφώς κατανοητοί, η ηλιακή δραστηριότητα υποχωρεί και ρέει σε έναν κύκλο περίπου 11 ετών. Ο δείκτης Solar Radio Flux είναι μια μέτρηση της ενσωματωμένης εκπομπής σε μήκος κύματος 10,7 cm (2800 MHz) και είναι ένας εξαιρετικός δείκτης της ηλιακής δραστηριότητας. Μπορεί εύκολα να μετρηθεί αξιόπιστα σε καθημερινή βάση από την επιφάνεια της Γης, σε όλους τους τύπους καιρού. Αναφέρεται σε «μονάδες ηλιακής ροής», (s.f.u.), το F10.7, όπως συχνά αναφέρεται, μπορεί να ποικίλλει από κάτω από 50 s.f.u. έως πάνω από 300 s.f.u., κατά τη διάρκεια ενός ηλιακού κύκλου.

5.4.5. Sunspot Number

Η αφθονία των ηλιακών κηλίδων στον Ήλιο ποικίλλει σε χρονικές κλίμακες από μερικές ώρες έως πολλά χρόνια. Ιστορικά, ο δείκτης Sunspot Number έχει χρησιμοποιηθεί για να ποσοτικοποιήσει την αφθονία αυτή. Αυτός ο δείκτης εξακολουθεί να χρησιμοποιείται ευρέως σήμερα, αν και για ορισμένους σκοπούς έχει αντικατασταθεί από πιο εύκολα και με συνέπεια μετρούμενους δείκτες όπως το F10.7. Το κύριο πλεονέκτημα του αριθμού ηλιακών κηλίδων είναι ότι είναι ο μόνος δείκτης για τον οποίο έχουμε μια μακρά και λεπτομερή ιστορική καταγραφή. Ο Sunspot Number (εδώ συμβολίζεται με R) ορίζεται ως:

$$R = K * (10 * G + I)$$

Όπου:

- G είναι ο αριθμός των ομάδων ηλιακών κηλίδων που είναι ορατές στον Ήλιο.
- I είναι ο συνολικός αριθμός των ορατών μεμονωμένων σημείων.
- K είναι ένας καθοριστικός παράγοντας για να ληφθούν υπόψη οι διαφορές μεταξύ παρατηρητών και παρατηρητηρίων.

5.4.6. Energetic Particle Flux

Η μαγνητόσφαιρα της Γης βομβαρδίζεται αδιάλειπτα από μια σχεδόν ισοτροπική ροή ενεργητικών φορτισμένων σωματιδίων που ονομάζεται κοσμική ακτινοβολία. Τα ισχυρά αυτά ηλιακά γεγονότα διαρκούν αρκετές ημέρες και απελευθερώνουν πρωτόνια και βαρύτερα ιόντα. Από άποψη ενέργειας τα σωματίδια αυτά είναι της τάξης των μερικών εκατοντάδων MeV και κυρίως επηρεάζουν συστήματα που

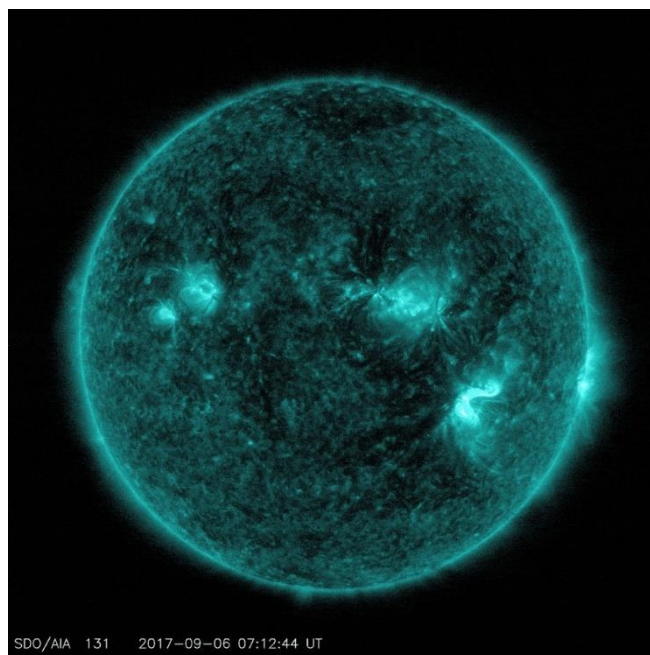
βρίσκονται σε μεγαλύτερο υψόμετρο. Περιστασιακά καταγράφονται ηλιακά γεγονότα που παράγουν σωματίδια με ενέργεια της τάξης GeV. Τα σωματίδια αυτά είναι εξαιρετικά ισχυρά και μπορούν να φτάσουν μέχρι και τα ισημερινά πλάτη (Dyer C. and Rodgers D., 1998). Για τη μελέτη της ηλιακής δραστηριότητας στο συγκεκριμένο τομέα, καταγράφονται συμβάντα πρωτονίων, ηλεκτρονίων και νετρονίων σε συγκεκριμένα κατώφλια σύμφωνα με τον NOAA.

5.4.7. Solar Flares

Τα Solar Flares ή ηλιακές εκλάμψεις είναι μια έντονη έκρηξη ηλεκτρομαγνητικής ακτινοβολίας στην ατμόσφαιρα του Ήλιου. Ιστορικά παρακολουθούνταν καλύτερα στο μήκος κύματος Η-άλφα και εμφανίζονται στη χρωμόσφαιρα (η κατώτερη περιοχή της ατμόσφαιρας του Ήλιου), αν και περιστασιακά παρατηρούνται εκλάμψεις λευκού φωτός στη φωτόσφαιρα (η ορατή «επιφάνεια» του Ήλιου). Τώρα, τα μήκη κύματος των ηλιακών ακτίνων Χ παρακολουθούνται μέσω δορυφόρου για ηλιακές εκλάμψεις.

Τα Solar Flares χαρακτηρίζονται από χρόνο ανόδου της τάξης των λεπτών και αποσύνθεση της τάξης των δεκάδων λεπτών. Η συνολική ενέργεια που δαπανάται σε μια τυπική έκρηξη είναι περίπου 10^{30} ergs. το μαγνητικό πεδίο είναι εξαιρετικά υψηλό, φτάνοντας τιμές από 100 έως 10.000 gauss.

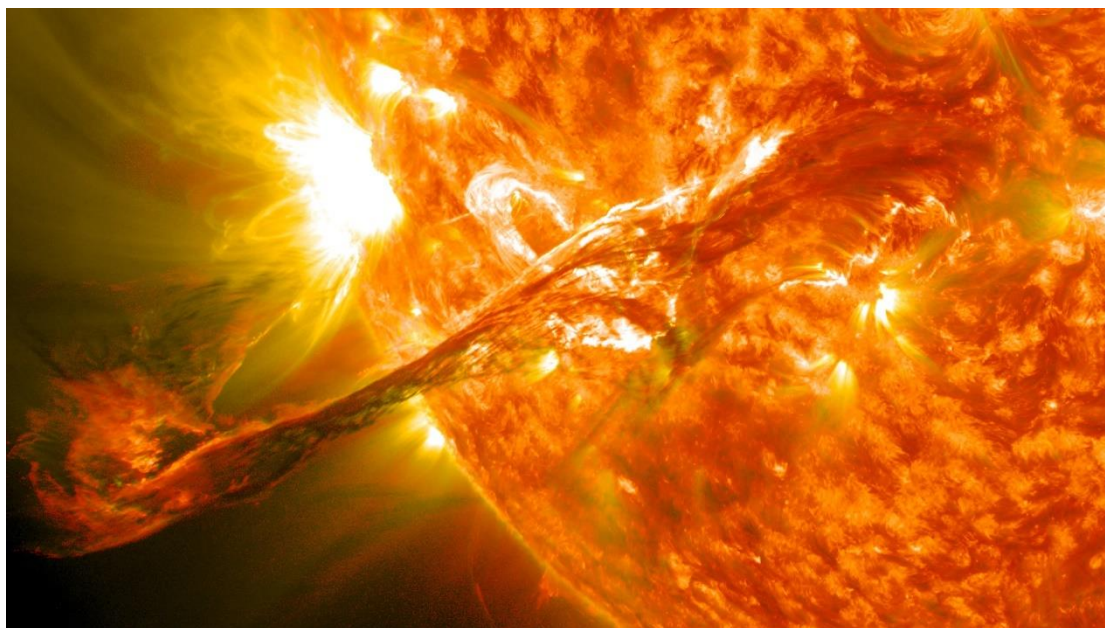
Οι εκλάμψεις συμβαίνουν σε ενεργές περιοχές και συχνά, αλλά όχι πάντα, συνοδεύονται από coronal mass ejections, συμβάντα ηλιακών σωματιδίων και άλλα ηλιακά φαινόμενα.



Εικόνα 3. Οι X2.2 και X9.3 Solar Flares που εξέπεμψε ο Ήλιος στις 6 Σεπτεμβρίου 2017. Οι εικόνες καταγράφηκαν από το Solar Dynamics Observatory της NASA και δείχνουν φως στο μήκος κύματος των 131 angstrom. (Πηγή: NASA/Goddard/SDO)

5.4.8. Coronal Mass Ejections

Οι Coronal Mass Ejections (CMEs) είναι μεγάλες αποβολές πλάσματος και μαγνητικού πεδίου από το στέμμα του Ήλιου. Μπορούν να εκτινάξουν δισεκατομμύρια τόνους στεφανιαίου υλικού και να φέρουν ένα ενσωματωμένο μαγνητικό πεδίο που είναι ισχυρότερο από την ισχύ του διαπλανητικού μαγνητικού πεδίου του ηλιακού ανέμου (IMF). Τα CME ταξιδεύουν προς τα έξω από τον Ήλιο με ταχύτητες που κυμαίνονται από μικρότερες από 250 km/s έως και 3000 km/s. Τα ταχύτερα CME που κατευθύνονται στη Γη μπορούν να φτάσουν στον πλανήτη μας σε μόλις 15-18 ώρες. Τα πιο αργά CME μπορεί να χρειαστούν αρκετές ημέρες για να φτάσουν. Επεκτείνονται σε μέγεθος καθώς διαδίδονται μακριά από τον Ήλιο και τα μεγαλύτερα CME μπορούν να φτάσουν σε μέγεθος που να περιλαμβάνει σχεδόν το ένα τέταρτο του χώρου μεταξύ της Γης και του Ήλιου μέχρι να φτάσει στον πλανήτη μας.



Εικόνα 4. Έκρηξη CME στις 31 Αυγούστου 2012 (Πηγή: NASA Goddard Space Flight Center)

6. Μηχανική Μάθηση

6.1. Εισαγωγικές Έννοιες

Νοημοσύνη είναι η ικανότητα επεξεργασίας της πληροφορίας με σκοπό τη λήψη απόφασης. Τεχνητή Νοημοσύνη (Artificial Intelligence – AI) είναι η προσπάθεια υλοποίησης αλγορίθμων με σκοπό την επεξεργασία της πληροφορίας για την λήψη απόφασης.

Η μηχανική μάθηση (Machine Learning - ML) είναι η μελέτη αλγορίθμων υπολογιστών που μπορούν να βελτιωθούν αυτόματα μέσω της εμπειρίας και με τη χρήση δεδομένων [57][58]. Θεωρείται υποσύνολο της τεχνητής νοημοσύνης. Ο Tom M. Mitchell διευκρινίζει τον όρο μάθηση με την ακόλουθη πρόταση: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E » [57].

Οι αλγόριθμοι μηχανικής μάθησης χτίζουν ένα μοντέλο που βασίζεται σε δείγματα δεδομένων, γνωστά ως δεδομένα εκπαίδευσης, προκειμένου να κάνουν προβλέψεις ή να πάρουν αποφάσεις χωρίς να είναι ρητά προγραμματισμένοι να το κάνουν. Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται σε μια μεγάλη ποικιλία εφαρμογών, όπως στην ιατρική, την αναγνώριση ομιλίας και την όραση υπολογιστή, όπου είναι δύσκολο ή ανέφικτο να αναπτυχθούν συμβατικοί αλγόριθμοι για την εκτέλεση των απαιτούμενων εργασιών.

Αναλυτικότερα, για ένα σύνολο δεδομένων εισόδου, ένας αλγόριθμος μηχανικής μάθησης μπορεί να εκτελέσει τις ακόλουθες βασικές εργασίες:

- **Ταξινόμηση (Classification):** Ο αλγόριθμος μηχανικής μάθησης καλείται να προσδιορίσει την κατηγορία στην οποία ανήκουν τα δεδομένα εισόδου.
- **Παλινδρόμηση (Regression):** Ο αλγόριθμος μηχανικής μάθησης προσπαθεί να γραμμικοποιήσει τις τιμές εισόδου, δηλαδή να μεταφράσει τη συμπεριφορά τους μέσω γραμμικής εξίσωσης.
- **Συσταδοποίηση (Clustering):** Ο αλγόριθμος μηχανικής μάθησης ζητείται να ομαδοποιήσει σε έναν πεπερασμένο αριθμό κλάσεων τα δεδομένα εισόδου.

Είναι προφανές πως οι αλγόριθμοι μηχανικής μάθησης εκτελούν πολλές παραπάνω εργασίες όπως π.χ. Transcription, Machine translation, Anomaly detection, Denoising, κ.λ.π.

Ένας βασικός στόχος ενός εκπαιδευόμενου είναι να γενικεύσει από την εμπειρία του. Η γενίκευση σε αυτό το πλαίσιο είναι η ικανότητα μιας μαθησιακής μηχανής να αποδίδει με ακρίβεια σε νέα, αόρατα παραδείγματα/εργασίες αφού έχει βιώσει ένα σύνολο δεδομένων μάθησης. Τα παραδείγματα εκπαίδευσης προέρχονται από

κάποια γενικά άγνωστη κατανομή πιθανοτήτων (που θεωρείται αντιπροσωπευτική του χώρου των περιστατικών) και ο εκπαιδευόμενος πρέπει να δημιουργήσει ένα γενικό μοντέλο για αυτόν τον χώρο που του επιτρέπει να παράγει επαρκώς ακριβείς προβλέψεις σε νέες περιπτώσεις.

Η υπολογιστική ανάλυση των αλγορίθμων μηχανικής μάθησης και η απόδοσή τους είναι ένας κλάδος της θεωρητικής επιστήμης των υπολογιστών που είναι γνωστός ως υπολογιστική θεωρία μάθησης. Επειδή τα εκπαιδευτικά σύνολα είναι πεπερασμένα και το μέλλον είναι αβέβαιο, η θεωρία μάθησης συνήθως δεν παρέχει εγγυήσεις για την απόδοση των αλγορίθμων. Αντίθετα, τα πιθανοτικά όρια για την απόδοση είναι αρκετά κοινά. Η αποσύνθεση μεροληψίας-διακύμανσης είναι ένας τρόπος για να ποσοτικοποιηθεί το σφάλμα γενίκευσης.

Για την καλύτερη απόδοση στο πλαίσιο της γενίκευσης, η πολυπλοκότητα της υπόθεσης θα πρέπει να ταιριάζει με την πολυπλοκότητα της συνάρτησης που βρίσκεται κάτω από τα δεδομένα. Εάν η υπόθεση είναι λιγότερο σύνθετη από τη συνάρτηση, τότε το μοντέλο έχει υπολογίσει τα δεδομένα. Εάν η πολυπλοκότητα του μοντέλου αυξάνεται σε απόκριση, τότε το σφάλμα εκπαίδευσης μειώνεται. Αλλά εάν η υπόθεση είναι πολύ περίπλοκη, τότε το μοντέλο υπόκειται σε υπερβολική προσαρμογή και η γενίκευση θα είναι φτωχότερη.

Τέλος, οι προσεγγίσεις μηχανικής μάθησης χωρίζονται παραδοσιακά σε τρεις μεγάλες κατηγορίες [56], ανάλογα με τη φύση του «σήματος» ή της «ανάδρασης» που διατίθεται στο σύστημα εκμάθησης:

- **Εποπτευόμενη εκμάθηση:** Ο υπολογιστής παρουσιάζεται με παραδείγματα εισόδων και τις επιθυμητές εξόδους τους, που δίνονται από έναν «δάσκαλο» και ο στόχος είναι να μάθουμε έναν γενικό κανόνα που αντιστοιχίζει τις εισόδους στις εξόδους.
- **Εκμάθηση χωρίς επίβλεψη:** Δεν δίνονται ετικέτες στον αλγόριθμο εκμάθησης, αφήνοντάς τον μόνο του να βρει δομή στην εισαγωγή του. Η μάθηση χωρίς επίβλεψη μπορεί να είναι ένας στόχος από μόνος του (ανακάλυψη κρυφών μοτίβων στα δεδομένα) ή ένα μέσο για την επίτευξη ενός σκοπού (εκμάθηση χαρακτηριστικών).
- **Ενισχυτική εκμάθηση:** Ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να εκτελέσει έναν συγκεκριμένο στόχο (όπως η οδήγηση ενός οχήματος ή το παιχνίδι ενός παιχνιδιού εναντίον ενός αντιπάλου). Καθώς περιηγείται στον χώρο προβλημάτων του, το πρόγραμμα λαμβάνει ανατροφοδότηση που είναι ανάλογη με τις ανταμοιβές, τις οποίες προσπαθεί να μεγιστοποιήσει.

7. Αλγόριθμοι Παλινδρόμησης

Στη στατιστική μοντελοποίηση, η ανάλυση παλινδρόμησης είναι ένα σύνολο στατιστικών διαδικασιών για την εκτίμηση των σχέσεων μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η πιο κοινή μορφή ανάλυσης παλινδρόμησης είναι η γραμμική παλινδρόμηση, στην οποία βρίσκουμε τη γραμμή (ή έναν πιο σύνθετο γραμμικό συνδυασμό) που ταιριάζει περισσότερο στα δεδομένα σύμφωνα με ένα συγκεκριμένο μαθηματικό κριτήριο. Για παράδειγμα, η μέθοδος των ελαχίστων τετραγώνων υπολογίζει τη μοναδική γραμμή (ή υπερεπίπεδο) που ελαχιστοποιεί το άθροισμα των τετραγώνων διαφορών μεταξύ των αληθινών δεδομένων και αυτής της γραμμής (ή υπερεπίπεδου). Για συγκεκριμένους μαθηματικούς λόγους, αυτό επιτρέπει στον ερευνητή να εκτιμήσει την υπό όρους προσδοκία (ή μέση τιμή πληθυσμού) της εξαρτημένης μεταβλητής όταν οι ανεξάρτητες μεταβλητές λαμβάνουν ένα δεδομένο σύνολο τιμών. Οι λιγότερο κοινές μορφές παλινδρόμησης χρησιμοποιούν ελαφρώς διαφορετικές διαδικασίες για να εκτιμήσουν εναλλακτικούς παραμέτρους τοποθεσίας ή να εκτιμήσουν την υπό όρους προσδοκία σε μια ευρύτερη συλλογή μη γραμμικών μοντέλων (π.χ. μη παραμετρική παλινδρόμηση). Η ανάλυση παλινδρόμησης χρησιμοποιείται κυρίως για δύο εννοιολογικά διακριτούς σκοπούς.

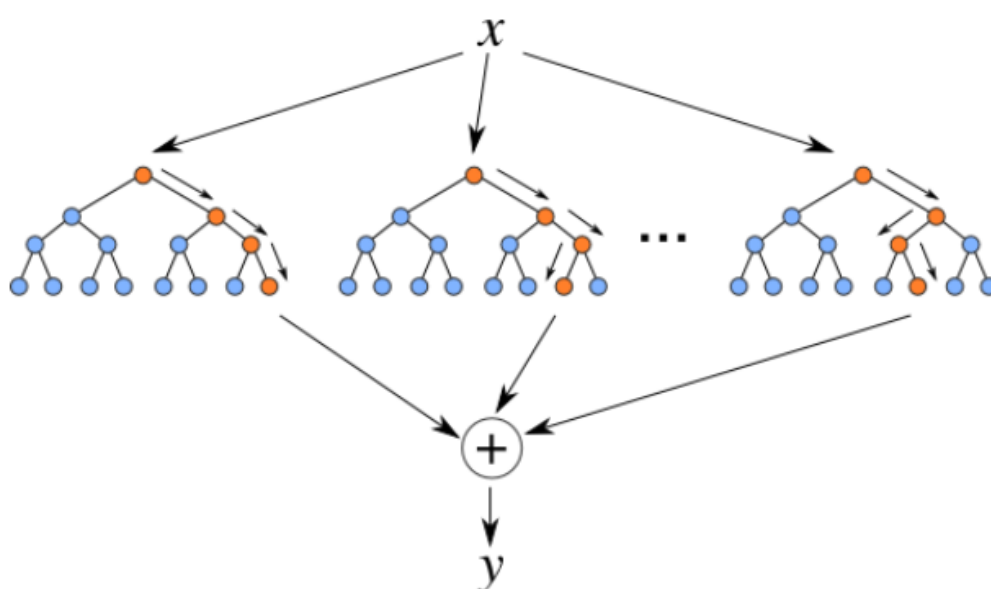
- Πρώτον, η ανάλυση παλινδρόμησης χρησιμοποιείται ευρέως για την πρόβλεψη, όπου η χρήση της έχει ουσιαστική επικάλυψη με το πεδίο της μηχανικής μάθησης.
- Δεύτερον, σε ορισμένες περιπτώσεις, η ανάλυση παλινδρόμησης μπορεί να χρησιμοποιηθεί για να συμπεράνει αιτιώδεις σχέσεις μεταξύ των ανεξάρτητων και των εξαρτημένων μεταβλητών. Είναι σημαντικό ότι οι παλινδρομήσεις από μόνες τους αποκαλύπτουν μόνο σχέσεις μεταξύ μιας εξαρτημένης μεταβλητής και μιας συλλογής ανεξάρτητων μεταβλητών σε ένα σταθερό σύνολο δεδομένων [61][62].

Για τους σκοπούς της συγκεκριμένης διπλωματικής εργασίας, χρησιμοποιήθηκαν ποικίλοι αλγόριθμοι για το πρόβλημα της παλινδρόμησης. Έγινε χρήση απλής γραμμικής παλινδρόμησης, SVM παλινδρόμησης, Gaussian Process, δέντρων απόφασης, Random Forest αλλά και πιο σύγχρονων αλγορίθμων μηχανικής μάθησης όπως ο XGBoost. Από όλους αυτούς τους αλγορίθμους, αυτοί που είχαν τα καλύτερα αποτελέσματα και θα αναλυθούν περαιτέρω είναι ο αλγόριθμος Random Forest και ο αλγόριθμος XGBoost.

7.1. Αλγόριθμος Random Forest

Ο ταξινομητής Random Forest (RF) [6][63], είναι ένας από τους πιο γνωστούς ομοιογενής, συνδυαστικούς ταξινομητές, ο οποίος χρησιμοποιεί πολλά, ασυσχέτιστα μεταξύ τους δέντρα αποφάσεων για να πραγματοποιήσει μία εκτίμηση (Belgiu M.

and Dragut L. (2016)). Η βασική ιδέα πίσω από το μοντέλο Random Forest, είναι η μείωση της συσχέτισης μεταξύ των ταξινομητών που το απαρτίζουν και του φαινομένου overfitting που προκαλεί την προσαρμογή του μοντέλου στον θόρυβο που υπάρχει στα δεδομένα. Κάθε δέντρο αποφάσεων δίνει μια κατηγοριοποίηση, προτεινόμενη τιμή της εξαρτημένης μεταβλητής. Από το σύνολο των δέντρων αποφάσεων κάθε πιθανή τιμή της εξαρτημένης μεταβλητής, έχει προταθεί συγκεκριμένες φορές από το σύνολο των δέντρων αποφάσεων, που αποτελεί και τη βάση ταξινόμησης με βάση τον αλγόριθμο random forest (Breiman, 2001). Βασικό πλεονέκτημα του ταξινομητή Random Forest είναι η ικανότητά του να διαχειρίζεται αποτελεσματικά μεγάλο αριθμό ανεξάρτητων μεταβλητών και ο μικρός απαιτούμενος χρόνος εκτέλεσής του (Breiman, 2001).



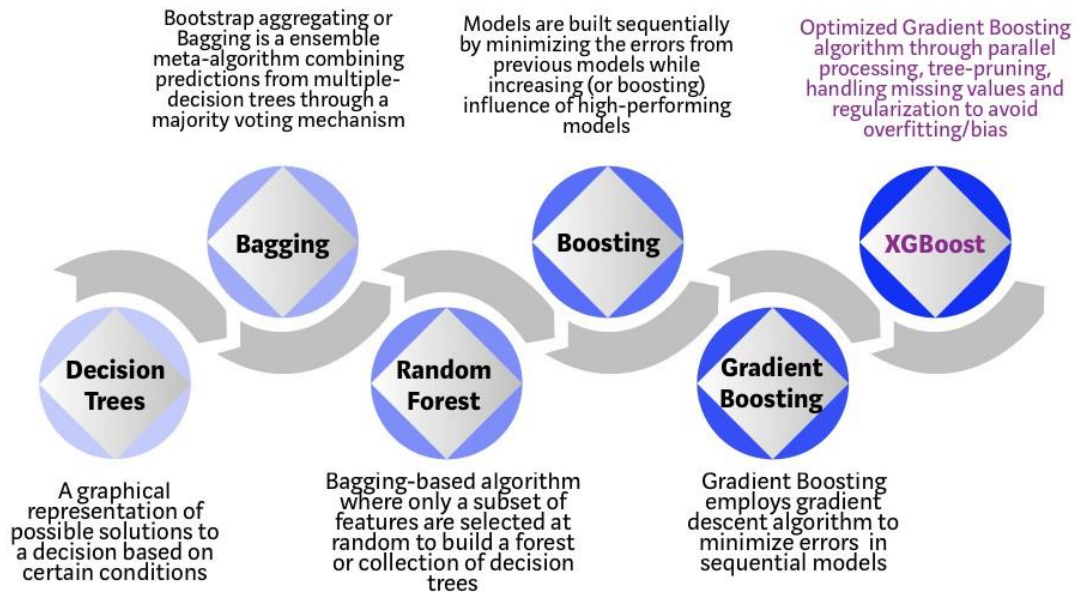
Εικόνα 5. Αναπαράσταση του αλγορίθμου Random Forest (Πηγή: levelup.gitconnected.com).

7.2. Αλγόριθμος XGBoost

Ο αλγόριθμος XGBoost είναι συντομογραφία του «eXtreme Gradient Boosting». Είναι ένας αλγόριθμος μηχανικής μάθησης που βασίζεται σε δέντρα αποφάσεων και χρησιμοποιεί ένα framework σταδιακής ενίσχυσης. Σε προβλήματα πρόβλεψης που αφορούν μη δομημένα δεδομένα (εικόνες, κείμενο, κ.λπ.), τα τεχνητά νευρωνικά δίκτυα τείνουν να έχουν καλύτερη απόδοση από όλους τους άλλους αλγόριθμους ή framework. Ωστόσο, όταν πρόκειται για μικρά έως μεσαία δομημένα δεδομένα, οι αλγόριθμοι που βασίζονται σε δέντρα αποφάσεων θεωρούνται οι καλύτεροι στην κατηγορία αυτή τη στιγμή.

Η βασική ιδέα του Boosting είναι ο συνδυασμός των αποτελεσμάτων πολλών «αδύναμων» αλγορίθμων μάθησης. Δηλαδή, ο συνδυασμός αλγορίθμων σφαλμάτων (error rate) είναι ελαφρώς καλύτερος από την τυχαία επιλογή. Με την τεχνική του boosting μειώνεται τόσο η διακύμανση (variance) όσο και η μεροληψία (bias). Το Gradient Boosting είναι μια τεχνική όπου προστίθενται νέα μοντέλα για τη διόρθωση

των υπολειπόμενων σφαλμάτων που δημιουργήθηκαν από τα προηγούμενα μοντέλα. Τα μοντέλα προστίθενται διαδοχικά έως ότου δεν γίνονται περαιτέρω βελτιώσεις. Χρησιμοποιείται σε ένα ευρύ φάσμα εφαρμογών: Μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων παλινδρόμησης, ταξινόμησης, κατάταξης και πρόβλεψης που καθορίζονται από τον χρήστη.



Εικόνα 6. Η εξέλιξη του αλγορίθμου XGBoost (Πηγή: towardsdatascience.com)

8. Προγραμματισμός επίλυσης σε Python

8.1. Η γλώσσα προγραμματισμού Python

Για την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python και συγκεκριμένα Python 3.6.

Η Python είναι μια διερμηνευόμενη (interpreted) γλώσσα προγραμματισμού γενικού σκοπού υψηλού επιπέδου. Η σχεδιαστική του φιλοσοφία δίνει έμφαση στην αναγνωσιμότητα του κώδικα με τη χρήση σημαντικών εσοχών. Οι γλωσσικές κατασκευές του καθώς και η αντικειμενοστραφής προσέγγισή του στοχεύουν να βοηθήσουν τους προγραμματιστές να γράφουν σαφή, λογικό κώδικα για έργα μικρής και μεγάλης κλίμακας. Η Python είναι δυναμική γλώσσα προγραμματισμού (dynamically typed) και υποστηρίζει συλλογή απορριμμάτων (garbage collection ή GC). Υποστηρίζει πολλαπλά παραδείγματα προγραμματισμού, συμπεριλαμβανομένου δομημένου, αντικειμενοστραφούς και λειτουργικού προγραμματισμού [1][64][65].

Δημιουργήθηκε από τον Ολλανδό Γκίντο βαν Ρόσσουμ (Guido van Rossum) στο ερευνητικό κέντρο Centrum Wiskunde & Informatica (CWI) το 1989 και κυκλοφόρησε για πρώτη φορά το 1991.

Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησής της. Μειονεκτεί όμως, στο γεγονός ότι επειδή είναι διερμηνευόμενη είναι πιο αργή από τις μεταγλωττιζόμενες (compiled) γλώσσες, όπως η C και η C++. Για αυτόν τον λόγο δεν είναι κατάλληλη για γραφή λειτουργικών συστημάτων.

Η Python αναπτύσσεται ως ανοιχτό λογισμικό (open source) και η διαχείρισή της γίνεται από τον μη κερδοσκοπικό οργανισμό Python Software Foundation. Ο κώδικας διανέμεται με την άδεια Python Software Foundation License η οποία είναι συμβατή με την GPL

Για τη συγγραφή προγραμμάτων είναι απαραίτητος ένας κειμενογράφος ή ακόμα καλύτερα ένα ολοκληρωμένο περιβάλλον ανάπτυξης (Integrated Development Environment - IDE), το οποίο είναι ένα ειδικό λογισμικό για την ανάπτυξη εφαρμογών. Για την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας, αρχικά, χρησιμοποιήθηκε ως ολοκληρωμένο περιβάλλον ανάπτυξης το PyCharm, με δυνατότητες επεξεργασίας, συγγραφής, αποσφαλμάτωσης (debugging) κ.τ.λ.. Στη συνέχεια, και μέχρι την ολοκλήρωση της εργασίας, η ανάπτυξη του κώδικα και η εξαγωγή των αποτελεσμάτων έγινε στο Google Colab. Ο λόγος της μετάβασης σε Colab ήταν η φορητότητα και η μεγαλύτερη υπολογιστική ισχύς.

8.2. Google Colab

Το Colaboratory, ή "Colab" για συντομία, είναι ένα προϊόν από την Google Research. Το Colab επιτρέπει σε οποιονδήποτε να γράψει και να εκτελέσει αυθαίρετο κώδικα python μέσω του προγράμματος περιήγησης και είναι ιδιαίτερα κατάλληλο για μηχανική μάθηση, ανάλυση δεδομένων και εκπαίδευση. Πιο τεχνικά, το Colab είναι μια φιλοξενούμενη υπηρεσία Jupyter (διαδικτυακή εφαρμογή που προσφέρει διαδραστικής ανάπτυξης σημειωματάρων, κώδικα και δεδομένων και είναι ιδανικό για εφαρμογές μηχανικής μάθησης και data science) που δεν απαιτεί εγκατάσταση για χρήση, ενώ παρέχει δωρεάν πρόσβαση σε υπολογιστικούς πόρους, συμπεριλαμβανομένων των GPU. Είναι ιδανικό για:

- Γραφή και εκτέλεση κώδικα σε Python
- Τεκμηρίωση του κώδικα που υποστηρίζει μαθηματικές εξισώσεις
- Δημιουργία/Μεταφόρτωση/Κοινή χρήση σημειωματάρων
- Εισαγωγή/Αποθήκευση σημειωματάρων από/στο Google Drive
- Εισαγωγή/Δημοσίευση σημειωματάρων από το GitHub
- Εισαγωγή εξωτερικών συνόλων δεδομένων π.χ. από το Kaggle
- Ενσωμάτωση PyTorch, TensorFlow, Keras, OpenCV
- Υπηρεσία Cloud με δωρεάν GPU

8.3. Βιβλιοθήκες Python

Η συγκεκριμένη γλώσσα προγραμματισμού χρησιμοποιήθηκε, πρώτον, διότι είναι μία γλώσσα ανοιχτού κώδικα και για την εγκατάστασή της και την λειτουργία της έχει μηδενικό κόστος. Δεύτερον, διότι διαθέτει πάρα πολλές βιβλιοθήκες για την επίλυση προβλημάτων συμπεριλαμβανομένου και μηχανικής μάθησης. Στη συγκεκριμένη εργασία χρησιμοποιήθηκαν οι κάτωθι βιβλιοθήκες:

- Pandas: αποτελεί μια βιβλιοθήκη της Python η οποία προσφέρει δομές δεδομένων και λειτουργίες για το χειρισμό αριθμητικών πινάκων και χρονοσειρών. Είναι δωρεάν λογισμικό που κυκλοφορεί με την άδεια BSD τριών ρητρών. Το όνομα προέρχεται από τον όρο "Panel Data", ένας όρος οικονομετρίας για σύνολα δεδομένων που περιλαμβάνουν παρατηρήσεις σε πολλαπλές χρονικές περιόδους για τα ίδια άτομα. Υπάρχει η δυνατότητα εισαγωγής δεδομένων από διάφορες μορφές αρχείων, όπως CSV, JSON, πίνακες ή ερωτήματα βάσεων δεδομένων SQL και Microsoft Excel. Επιτρέπει διάφορες λειτουργίες χειρισμού δεδομένων, όπως συγχώνευση, αναμόρφωση, επιλογή, καθώς και λειτουργίες καθαρισμού δεδομένων και σύγχυσης δεδομένων.
- Numpy [66]: αποτελεί μια βιβλιοθήκη της python η οποία υλοποιήθηκε με στόχο την επεξεργασία και την διαχείριση μαθηματικών προβλημάτων μέσω

της δομής πινάκων. Η NumPy βρίσκει πληθώρα εφαρμογών σε προβλήματα γραμμικής άλγεβρας, μετασχηματισμού Fourier και υλοποίηση πράξεων μεταξύ πινάκων. Η NumPy δημιουργήθηκε το 2005 από τον Travis Oliphant και είναι ένα εργαλείο ανοιχτού κώδικα που μπορεί να χρησιμοποιηθεί ελεύθερα. Ο λόγος που χρησιμοποιείται η συγκεκριμένη βιβλιοθήκη είναι γιατί στην Python υπάρχουν οι λίστες που εξυπηρετούν το σκοπό των πινάκων, αλλά είναι αργές στην επεξεργασία τους. Η NumPy στοχεύει στην παροχή ενός πίνακα που για οποιαδήποτε πράξη είναι αρκετά πιο γρήγορος από μια παραδοσιακή λίστα.

- Sklearn (scikit-learn) [67]: αποτελεί μια βιβλιοθήκη της Python η οποία υλοποιήθηκε με στόχο την ανάπτυξη αλγορίθμων μηχανικής μάθησης. Η βιβλιοθήκη sklearn περιέχει πολλά αποτελεσματικά εργαλεία για μηχανική μάθηση και στατιστική μοντελοποίηση, όπως ταξινόμηση, παλινδρόμηση, ομαδοποίηση και μείωση διαστάσεων. Η sklearn διαθέτει μερικές από τις ακόλουθες δυνατότητες:
 - Supervised learning algorithms
 - Cross-validation
 - Unsupervised learning algorithms
 - Various toy datasets
 - Feature extraction
- SHAP (SHapley Additive exPlanation): αποτελεί μια βιβλιοθήκη της Python η οποία είναι μια θεωρητική προσέγγιση παιγνίων που εξηγεί τα αποτελέσματα οποιουδήποτε μοντέλου μηχανικής μάθησης. Συνδέει τη βέλτιστη κατανομή πιστώσεων (credit allocation) με τοπικές επεξηγήσεις χρησιμοποιώντας τις κλασικές τιμές Shapley από τη θεωρία παιγνίων και τις σχετικές επεκτάσεις τους αντιπροσωπεύοντας τη μόνη δυνατή, συνεπή και τοπικά ακριβή μέθοδο απόδοσης πρόσθετων χαρακτηριστικών με βάση τις προσδοκίες.
- Statistics: αποτελεί μια βιβλιοθήκη της Python η οποία παρέχει λειτουργίες για τον υπολογισμό μαθηματικών στατιστικών αριθμητικών δεδομένων.
- Matplotlib: αποτελεί μια βιβλιοθήκη της Python η οποία επιτρέπει τη σχεδίαση γραφημάτων καθώς και στατικών, κινούμενων και διαδραστικών απεικονίσεων.
- Xgboost: αποτελεί μια βιβλιοθήκη της Python η οποία εφαρμόζει αλγόριθμους μηχανικής μάθησης στο framework του Gradient Boosting. Το XGBoost παρέχει μια παράλληλη ενίσχυση δέντρων (tree boosting: επίσης γνωστή ως GBDT, GBM) που λύνει πολλά προβλήματα data science με γρήγορο και ακριβή τρόπο.
- LIME (Local Interpretable Model-agnostic Explanations): αποτελεί μια βιβλιοθήκη της Python η οποία δημιουργεί αραιά γραμμικά μοντέλα γύρω

από κάθε πρόβλεψη για να εξηγήσει πώς λειτουργεί ένα κλειστό μοντέλο (μαύρου κουτιού) σε αυτήν την τοπική περιοχή.

- Ορενρυχι: αποτελεί μια βιβλιοθήκη της Python για την ανάγνωση/εγγραφή αρχείων Excel 2010.

8.4. Υλοποίηση σε Python

Αφού ολοκληρώθηκε η επεξεργασία των πειραματικών δεδομένων, αποφασίστηκε η γλώσσα προγραμματισμού και βρέθηκαν οι κατάλληλες βιβλιοθήκες, ξεκίνησε η διαδικασία υλοποίησης του κώδικα. Αναλυτικά:

Εγκαταστάθηκαν και εισήχθησαν οι παραπάνω βιβλιοθήκες.

- `path = '/content/drive/MyDrive/Colab/jan.xlsx'`: Ορίστηκε η διαδρομή ανάγνωσης του αρχείου δεδομένων για τον κάθε μήνα.
- `df = pd.read_excel(path)`: Έγινε ανάγνωση του αρχείου xlsx και μετατροπή του σε pandas dataframe.
- `dates = df['date']`, `timestep = df['timestep']`, `stec = df['stec']` & `col_stec = df['stec']`: Ορίστηκαν τα αντίστοιχα dataframes.
- `stec_std = stec.std()` & `std = df_no_date.std()`: Υπολογίστηκε η τυπική απόκλιση του dataframe.
- `stec_mean = stec.mean()` & `mean = df_no_date.mean()`: Υπολογίστηκε ο μέσος όρος του dataframe.
- `df_no_date = df.drop(['timestep', 'date', 'D', 'dpind'], axis=1)`: Έγινε απόρριψη στηλών που δεν χρειαζόντουσαν σε αυτό το dataframe.
- `max = col_stec.max()`: Βρέθηκε ο μέγιστος όρος του dataframe.
- `min = col_stec.min()`: Βρέθηκε ο μέγιστος όρος του dataframe.
- `dif = max - min`: Υπολογίστηκε η διαφορά μεταξύ max και min.
- `df_for_training = (df_no_date - mean) / std`: Έγινε κανονικοποίηση των όλων των δεδομένων του dataframe.
- `trainX1 = df_for_training[['time', 'a', 'kp', 'ap', 'SML', 'SMU', 'fluxobsflux', 'fluxadjflux', 'fluxursi', 'dts', 'dns', 'dss', 'sddts', 'sddns', 'sddss', 'obsdts', 'obsdns', 'obsdss', 'one', 'ten', 'hundred', 'zeroeight', 'two', 'nm', 'sf', 'pa', 'ang', 'lin', 'height', 'speed', 'acc', 'mass', 'ke', 'mpa', 'msd', 'ked']]` & `trainY1 = df_no_date[["stec"]]`: Ορίστηκαν τα δεδομένα εισόδου και εξόδου για την εκπαίδευση.
- `trainX, trainY = np.array(trainX1), np.array(trainY1)`: Έγινε μετασχηματισμός των δεδομένων σε numpy array.
- `x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(trainX, trainY, test_size = 0.4, shuffle=True)`: Έγινε χωρισμός των δεδομένων σε δεδομένα ελέγχου και εκπαίδευσης.

- Στα παραπάνω δεδομένα εφαρμόστηκαν οι προαναφερθέντες αλγόριθμοι παλινδρόμησης και υπολογίστηκαν τα αντίστοιχα MAE και rMSE.
- Ακολουθώντας, έγινε οπτικοποίηση των MAE και rMSE σε πίνακα. Από τα αποτελέσματα έγινε η αλλαγή της αρχικής επιλογής παλινδρομητή από Random Forest σε XGBoost.
- `test_x = pd.DataFrame(x_test, columns = trainX1.keys())` & `test_y = pd.DataFrame(y_test, columns = trainY1.keys())`: Δημιουργήθηκαν τα αντίστοιχα pandas dataframe από τα παραπάνω δεδομένα.
- `test = pd.concat([test_x, test_y], axis=1)`: Συνδέθηκαν τα ανωτέρω dataframe.
- `test_with_dates = pd.merge(test, df_for_training, how="inner", on=['time', 'a', 'kp', 'ap', 'SML', 'SMU', 'fluxobsflux', 'fluxadjflux', 'fluxursi', 'dts', 'dns', 'dss', 'sd dts', 'sddns', 'sddss', 'obsdts', 'obsdns', 'obsdss', 'one', 'ten', 'hundred', 'zeroeight', 'two', 'nm', 'sf', 'pa', 'ang', 'lin', 'height', 'speed', 'acc', 'mass', 'ke', 'mpa', 'msd', 'ked'])`: Ενώθηκαν τα αναφερόμενα dataframe.
- `test_with_dates.drop(['vtec', 'stec_y'], inplace=True, axis=1)`: Αφαιρέθηκαν οι αναφερόμενες στήλες.
- `test_with_dates = test_with_dates.sort_values(by=['timestep'])`: Έγινε ταξινόμηση των δεδομένων με βάση το χρονικό βήμα (ωριαίο)
- `XTest, YTest = np.array(XTest), np.array(YTest)`: Έγινε μετασχηματισμός των δεδομένων σε numpy array.
- `preds = regressor.predict(XTest)`: Υπολογίστηκαν οι προβλέψεις για τις τιμές `stev` ή `vtec` αντίστοιχα.
- `preds_denorm = preds*stec_std + stec_mean` & `YTest_denorm = YTest*stec_std + stec_mean`: Έγινε αποκανονικοποίηση των δεδομένων.
- Τέλος, έγινε οπτικοποίηση των αποτελεσμάτων σε γραφήματα ανά μήνα.

Το project ονομάστηκε RF_Solar εκ των αρχικών των λέξεων του αλγορίθμου Random Forest καθώς, αρχικά, είχε θεωρηθεί ότι θα πρόσφερε τη βέλτιστη λύση.

9. Πειραματική Διαχείριση

Για την ολοκλήρωση της μελέτης, απαιτήθηκε η συλλογή των απαραίτητων δεδομένων από κάθε διαθέσιμο και πιστοποιημένο φορέα.

9.1. Περιγραφή των δεδομένων

Τα δεδομένα που συλλέχθηκαν είναι:

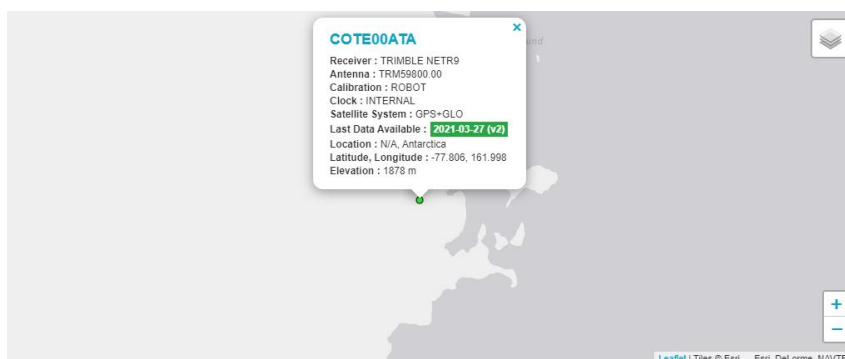
- Observation Files από το σταθμό COTE00ATA.
- Δεδομένα για την παράμετρο Geomagnetic activity (K).
- Δεδομένα για την παράμετρο Auroral Electrojet (AE).
- Δεδομένα για την παράμετρο Disturbance storm time (Dst).
- Δεδομένα για την παράμετρο Solar Radio Flux.
- Δεδομένα για την παράμετρο Sunspot Number.
- Δεδομένα για την παράμετρο Energetic Particle Flux.
- Δεδομένα για την παράμετρο Solar Flares.
- Δεδομένα για την παράμετρο Coronal Mass Ejections.

Τα παραπάνω δεδομένα χορηγούνται σε διάφορες μορφές και απαιτήθηκε εκτενής και πολύωρη επεξεργασία ώστε να έρθουν σε μορφή διαχειρίσιμη συνολικά από ένα χρήστη. Αναλυτικά:

9.1.1. Observation Files του σταθμού COTE00ATA

Τα observation files όλων των σταθμών καθώς και διάφορα άλλα σχετικά δεδομένα που αφορούν GNSS, SLR, VLBI και DORIS δίνονται δωρεάν για ερευνητικούς σκοπούς από τη NASA μέσω του CDDIS.

Από όλους τους διαθέσιμους σταθμούς επιλέχθηκε ο σταθμός COTE00ATA, ο οποίος βρίσκεται στην Ανταρκτική (Εικόνα 8).

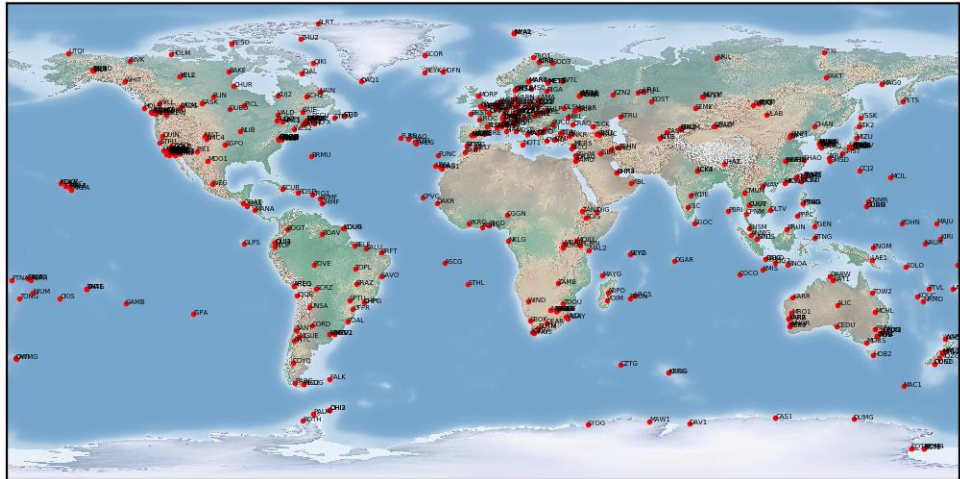


Εικόνα 7. Σταθμός COTE00ATA (Πηγή: IGS)

Η επιλογή του συγκεκριμένου σταθμού έγινε για τους εξής λόγους:

- Το φαινόμενο της ιονοσφαιρικής έξαρσης (ionospheric scintillation) συμβαίνει κυρίως στις πολικές περιοχές και σε μία ζώνη $\pm 30^\circ$ γύρω από το γεωμαγνητικό ισημερινό και διαρκεί έως μερικά λεπτά.

- Ήταν από τους λίγους σταθμούς που είχε συνεχή δεδομένα
- Θεωρήθηκε ότι δεν επηρεάζεται από κανενός είδους ανθρώπινες παρεμβάσεις (κατασκευές, δίκτυα τηλεφωνίας κτλ)



Εικόνα 8. Όλοι οι διαθέσιμοι σταθμοί του συστήματος GNSS (Πηγή: IGS)

Για το σταθμό αυτό ελήφθησαν τα observation files για τους μήνες Ιανουάριο – Απρίλιο του έτους 2018. Η επιλογή των συγκεκριμένων μηνών έγινε εξαιτίας της ύπαρξης ημερών με έντονη ηλιακή δραστηριότητα.

Τα αρχεία ήταν τύπου RINEX (βλ. 9.1.1.3).όπως φαίνεται και στην παρακάτω εικόνα (Εικόνα 10) και για την επίλυση τους και την εξαγωγή των δεδομένων TEC που απαιτούνταν για την εκτέλεση της παρούσας διπλωματικής έγινε χρήση του λογισμικού GPS GOPI (βλ. 9.1.1.4).

```

2.11      OBSERVATION DATA      M (MIXED)      RINEX VERSION / TYPE
teqc 2018Jan30 UNAVCO Archive Ops 20180201 01:02:11UTC PGM / RUN BY / DATE
Solaris x86_5.10 [AMD64] cc SCS.8 -xarch=amd64 [+]+
0.000      (antenna height)      COMMENT
-77.80590314 (latitude)      COMMENT
+161.99781078 (longitude)      COMMENT
1878.374      (elevation)      COMMENT
BIT 2 OF LLI FLAGS DATA COLLECTED UNDER A/S CONDITION      COMMENT
COTE      MARKER NAME
56090M001      MARKER NUMBER
Terry Wilson      Ohio State University      OBSERVER / AGENCY
5034K69672      TRIMBLE NETR9 5.10      REC # / TYPE / VERS
4852A62823      TRMS9800.00 SCS      ANT # / TYPE
-1265757.4165      417822.4915 -6214230.8474      APPROX POSITION XYZ
0.00000      0.00000      0.00000      ANTENNA: DELTA H/E/N
1 1      WAVELENGTH FACT L1/2
7 L1 L2 C1 P2 P1 S1 S2      # / TYPES OF OBSERV
30.0000      INTERVAL
18      LEAP SECONDS
input file: cote201801310000a.tgd      COMMENT
Forced Modulo Decimation to 30 seconds      COMMENT
RINEX file created by UNAVCO GPS Archive.      COMMENT
%monument ID: 21831      COMMENT
JUNAVCO 4-char name: COTE      COMMENT
4-char name from Log or data file: COTE      COMMENT
%monument location: -77.8059 161.9978 1878.1      COMMENT
%visit ID: 117469      COMMENT
POLENET-NSF/OSU/UNAVCO      COMMENT
doi:10.7283/TSGT5KGN      COMMENT
End of DB comments      COMMENT
SNR is mapped to RINEX snr flag value [0-9]      COMMENT
L1 & L2: min(max(int(snr_dBHz/6), 0), 9)      COMMENT
2018 1 31 0 0 0 0.000000      GPS      TIME OF FIRST OBS
END OF HEADER
18 1 31 0 0 0.000000 0 23R17R06R00R24G22G32R01G17R07G31G14R15
G25R23G01G19G12R16G02G06R14G24G03
112379164.792 8 87406084.335 7 21000737.836 21000740.309 21000736.313
48.500 46.000
127147974.451 5 98892907.495 4 23827481.977 23827484.516 23827480.402
32.400 27.400
114161736.087 7 88792544.948 7 21318910.219 21318913.691 21318909.824
45.600 45.900
104291009.446 7 81115254.573 6 19502946.250 19502946.688 19502944.441
47.200 38.000
119202007.791 7 92884700.20245 22683386.516 22683388.270
46.900 33.000

```

Εικόνα 9. Τυπική μορφή αρχείου RINEX.

9.1.1.1. CDDIS

Το Crustal Dynamics Data Information System (CDDIS) αναπτύχθηκε αρχικά για να παρέχει μια κεντρική τράπεζα δεδομένων για το Crustal Dynamics Project (CDP) της NASA. Το σύστημα συνεχίζει να υποστηρίζει την κοινότητα διαστημικής γεωδαισίας και γεωδυναμικής μέσω του Έργου Διαστημικής Γεωδαισίας της NASA καθώς και της Επιχείρησης Earth Science Enterprise της NASA. Το CDDIS ιδρύθηκε το 1982 ως ειδική τράπεζα δεδομένων για την αρχειοθέτηση και τη διανομή συνόλων δεδομένων που σχετίζονται με τη γεωδαισία του διαστήματος. Σήμερα, το CDDIS αρχειοθετεί και διανέμει κυρίως δεδομένα GNSS (GNSS, επί του παρόντος Παγκόσμιο Σύστημα Εντοπισμού Θέσης GPS και Παγκόσμιο Δορυφορικό Σύστημα Πλοήγησης GLONASS), SLR (τόσο σε τεχνητούς δορυφόρους, SLR, και σε σεληνιακό, LLR), Very Long Baseline Interferometry (VLBI), και Doppler Orbitography and Radio-positioning Integrated by Satellite (DORIS) για μία συνεχώς αυξανόμενη κοινότητα χρηστών γεωφυσικών.

Το CDDIS έχει χρησιμεύσει ως παγκόσμιο κέντρο δεδομένων για τη Διεθνή Υπηρεσία GNSS (IGS) από το 1992. Υποστηρίζει επίσης ενεργά τη International Laser Ranging Service (ILRS), τη International VLBI Service for Geodesy and Astrometry (IVS), τη Διεθνή Υπηρεσία DORIS (IDS) και τη Διεθνή Υπηρεσία Συστημάτων Περιστροφής και Αναφοράς Γης (IERS) ως παγκόσμιο κέντρο δεδομένων.

9.1.1.2. IGS

Η International GNSS Service (IGS) είναι μία εθελοντική ομοσπονδία περισσότερων από 200 αυτοχρηματοδοτούμενων φορέων, πανεπιστημίων και ερευνητικών ιδρυμάτων σε περισσότερες από 100 χώρες/περιοχές. συνεργάζονται για να παρέχουν τις δορυφορικές τροχιές GPS υψηλότερης ακρίβειας στον κόσμο.

Η IGS, ως μέρος του Global Geodetic Observing System, λειτουργεί ένα παγκόσμιο δίκτυο επίγειων σταθμών GNSS, κέντρων δεδομένων και κέντρων ανάλυσης δεδομένων για την παροχή δεδομένων και προϊόντων που είναι απαραίτητα για την έρευνα της επιστήμης της Γης. Πολυεπιστημονικές εφαρμογές εντοπισμού θέσης, πλοήγησης και χρονισμού (PNT) και εκπαίδευση. Ο Reference Frame Coordinator της IGS καθορίζει τις συντεταγμένες των σταθμών παρακολούθησης και τις ταχύτητες στο Διεθνές Επίγειο Σύστημα Αναφοράς (ITRF) και οργανώνει τη συνεισφορά του IGS στο ITRF. Μάθε περισσότερα

Τα προϊόντα IGS περιλαμβάνουν:

- Εφημερίδες δορυφόρων GNSS.
- Παράμετροι περιστροφής της γης.
- Συντεταγμένες και ταχύτητες παγκόσμιων σταθμών παρακολούθησης.
- Πληροφορίες ρολογιού δορυφόρων και σταθμών παρακολούθησης.
- Εκτιμήσεις καθυστέρησης ζενίθιας τροποσφαιρικής διαδρομής.

- Παγκόσμιοι χάρτες ιονόσφαιρας

Αυτά τα προϊόντα υποστηρίζουν τις επιστημονικές αναλύσεις της Γης και άλλες προσπάθειες, όπως π.χ:

- Βελτίωση και επέκταση του Διεθνούς Επίγειου Συστήματος Αναφοράς (ITRF) που διατηρεί η Διεθνής Υπηρεσία Συστημάτων Περιστροφής Γης και Αναφοράς (IERS).
- Παρακολούθηση παραμόρφωσης της Γης.
- Παρακολούθηση περιστροφής της γης.
- Παρακολούθηση της τροπόσφαιρας και της ιονόσφαιρας.
- Προσδιορισμός τροχιών επιστημονικών δορυφόρων και άλλες ποικίλες εφαρμογές.

9.1.1.3. RINEX

Στον τομέα της γεωδαισίας, το Receiver Independent Exchange Format (RINEX) είναι μια μορφή ανταλλαγής δεδομένων για ακατέργαστα δεδομένα συστήματος δορυφορικής πλοήγησης. Αυτό επιτρέπει στον χρήστη να επεξεργάζεται εκ των υστέρων τα δεδομένα που λαμβάνονται για να παράγει ένα πιο ακριβές αποτέλεσμα — συνήθως με άλλα δεδομένα άγνωστα στον αρχικό δέκτη, όπως καλύτερα μοντέλα των ατμοσφαιρικών συνθηκών τη στιγμή της μέτρησης.

Η τελική έξοδος ενός δέκτη πλοήγησης είναι συνήθως η θέση, η ταχύτητά του ή άλλα σχετικά φυσικά μεγέθη. Ωστόσο, ο υπολογισμός αυτών των ποσοτήτων βασίζεται σε μια σειρά μετρήσεων από έναν ή περισσότερους δορυφορικούς αστερισμούς. Αν και οι δέκτες υπολογίζουν τις θέσεις σε πραγματικό χρόνο, σε πολλές περιπτώσεις είναι ενδιαφέρον να αποθηκεύονται ενδιάμεσα μέτρα για μελλοντική χρήση. Το RINEX είναι η τυπική μορφή που επιτρέπει τη διαχείριση και τη διάθεση των μέτρων που παράγονται από έναν δέκτη, καθώς και την off-line επεξεργασία τους από πληθώρα εφαρμογών, ανεξάρτητα από τον κατασκευαστή του δέκτη και της εφαρμογής υπολογιστή.

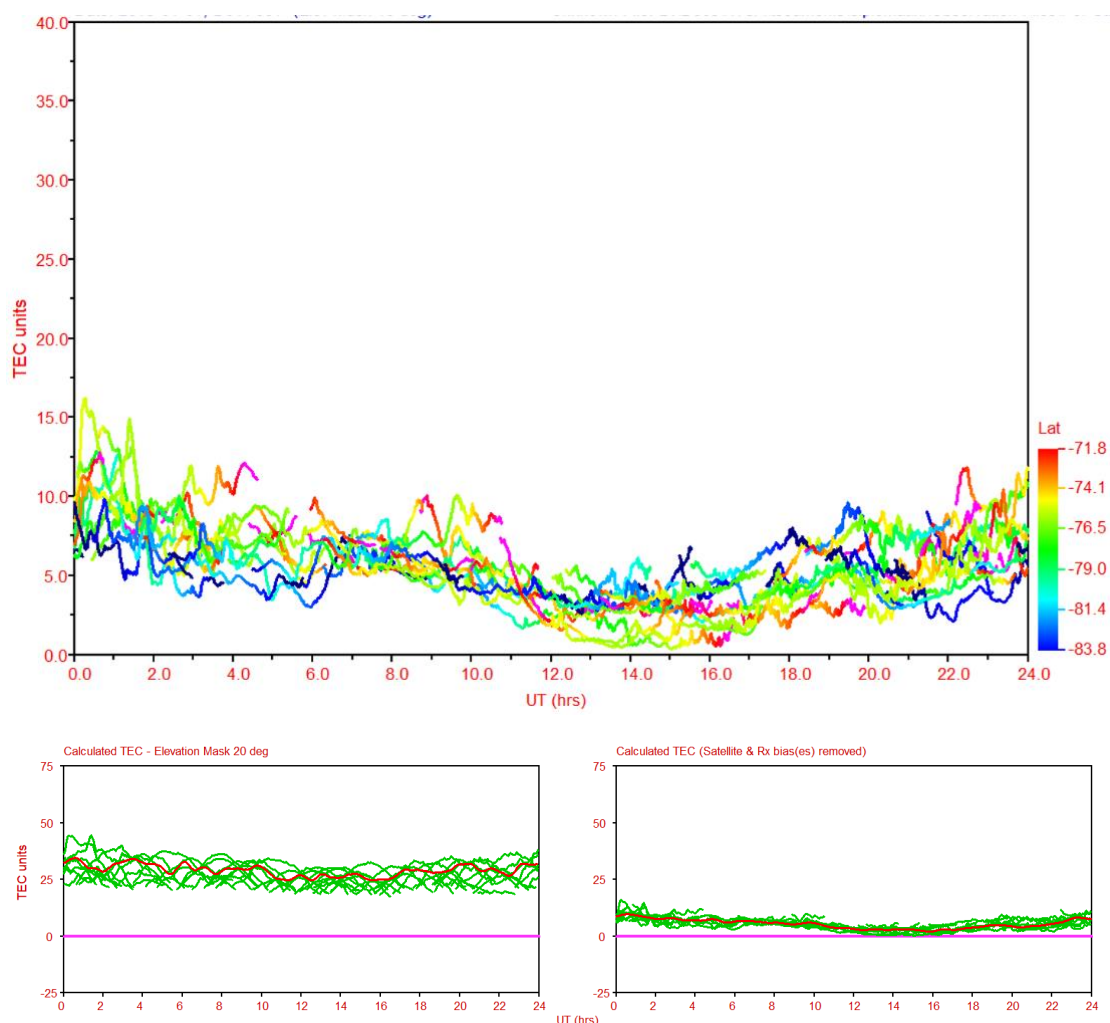
Η μορφή RINEX έχει σχεδιαστεί για να εξελίσσεται με την πάροδο του χρόνου, προσαρμοζόμενη σε νέους τύπους μετρήσεων και νέα συστήματα δορυφορικής πλοήγησης. Η πρώτη έκδοση RINEX αναπτύχθηκε από τον W. Gurtner το 1989 και δημοσιεύτηκε από τους W. Gurtner και G. Mader στο CSTG GPS Bulletin του Σεπτεμβρίου/Οκτωβρίου 1990. Από το 1993 το RINEX 2 είναι διαθέσιμο, το οποίο έχει αναθεωρηθεί και υιοθετηθεί πολλές φορές. Το RINEX επιτρέπει την αποθήκευση των μετρήσεων ψευδοαπόστασης, φάσης φορέα, Doppler και σήματος προς θόρυβο από GPS (συμπεριλαμβανομένων σημάτων εκσυγχρονισμού GPS, π.χ. L5 και L2C), GLONASS, Galileo, Beidou, μαζί με δεδομένα από δορυφόρους EGNOS και WAAS συστήματα αύξησης (SBAS), QZSS, ταυτόχρονα. Η έκδοση 3.02 του RINEX υποβλήθηκε

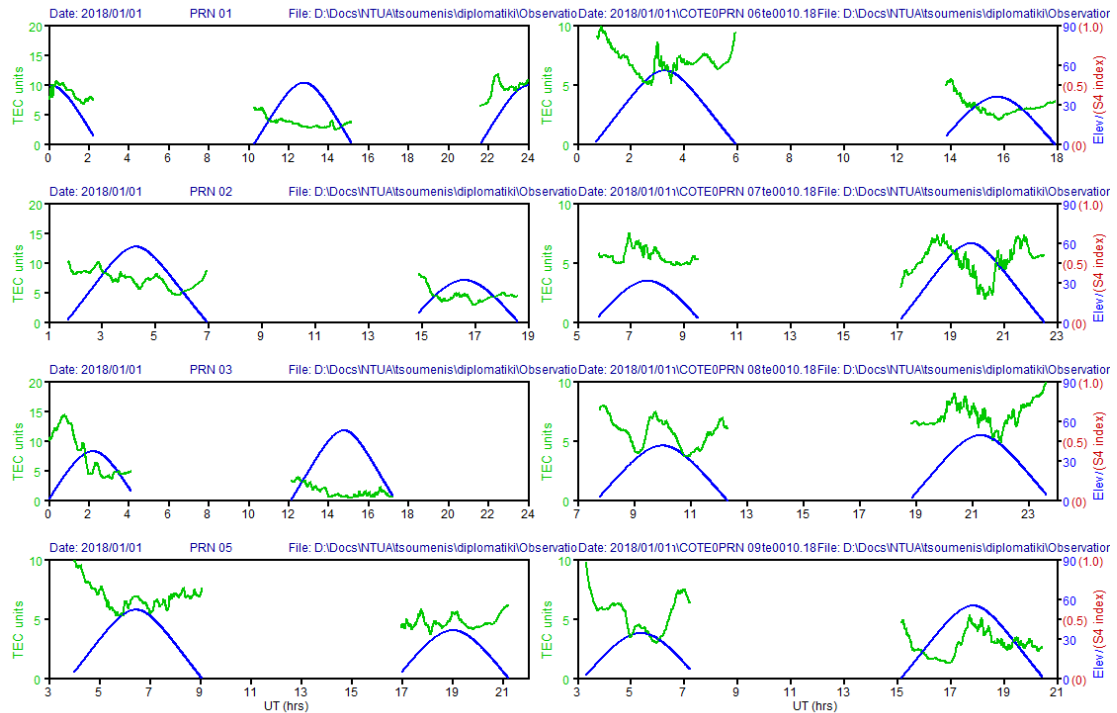
τον Απρίλιο του 2013 και περιέχει νέους κωδικούς παρατήρησης από συστήματα GPS ή Galileo. Η πιο πρόσφατη έκδοση είναι το RINEX 4.00 από τον Δεκέμβριο του 2021.

Αν και δεν αποτελεί μέρος της μορφής RINEX, το σχήμα συμπίεσης Hatanaka χρησιμοποιείται συνήθως για τη μείωση του μεγέθους των αρχείων RINEX, με αποτέλεσμα μια μορφή CompactRINEX ή CRINEX που βασίζεται σε ASCII. Χρησιμοποιεί χρονικές διαφορές υψηλότερης τάξης για να μειώσει τον αριθμό των χαρακτήρων που απαιτούνται για την αποθήκευση δεδομένων χρόνου.

9.1.1.4. GPS GOPI

Το λογισμικό GPS Gopi αναπτύχθηκε από τον Dr. Gopi Krishna Seemala και χρησιμοποιείται για την ανάλυση δεδομένων πλοήγησης [68]. Έχει δυνατότητα επεξεργασίας δεδομένων RINEX (ημερήσια ή μεγάλων περιόδων) και, όντας συνδεδεμένο στο διαδίκτυο μπορεί να λαμβάνει της εφημερίδα των δορυφόρων από το αρχείο πλοήγησης του IGS. Επίσης, μπορεί να υπολογίζει το TEC (STEC και VTEC) από τα δεδομένα παρατήρησης των GPS Rinex, Novatel, SCINDA και Leica. Τέλος, υπολογίζει την πόλωση του δέκτη και τις διακαναλικές πολώσεις για διαφορετικούς δορυφόρους στον ίδιο δέκτη. Το λογισμικό παράγει τα κάτωθι γραφήματα:





Εικόνα 10. Παραγόμενα διαγράμματα από το λογισμικό GPS Gori

9.1.2. Geomagnetic activity (K)

Τα δεδομένα γεωμαγνητικής δραστηριότητας, παράμετρος K (δείκτες Kp, Ap, ar και D), ελήφθησαν από το Γεωμαγνητικό Παρατηρητήριο Niemegk, GFZ Γερμανικό Ερευνητικό Κέντρο Γεωεπιστημών (Geomagnetic Observatory Niemegk, GFZ German Research Centre for Geosciences). Τα δεδομένα είναι μία δημοσίευση των Matzka J., Stolle C., Yamazaki Y., Bronkalla O. και Morschhauser A..

```
# The format for each line is (i stands for integer, f for float):
#iiii ii ff,f ffff,ffff fffff,ffff ff,fff iiii i
# The parameters in each line are:
#YYY MM DD hh,h hh,_m days days_m Kp ap D
2018 01 01 00,0 01,50 31412,00000 31412,06250 3,333 18 1
2018 01 01 03,0 04,50 31412,12500 31412,18750 3,667 22 1
2018 01 01 06,0 07,50 31412,25000 31412,31250 2,333 9 1
2018 01 01 09,0 10,50 31412,37500 31412,43750 2,333 9 1
2018 01 01 12,0 13,50 31412,50000 31412,56250 2,667 12 1
2018 01 01 15,0 16,50 31412,62500 31412,68750 1,000 4 1
2018 01 01 18,0 19,50 31412,75000 31412,81250 1,000 4 1
2018 01 01 21,0 22,50 31412,87500 31412,93750 1,333 5 1
2018 01 02 00,0 01,50 31413,00000 31413,06250 1,667 6 1
2018 01 02 03,0 04,50 31413,12500 31413,18750 0,667 3 1
2018 01 02 06,0 07,50 31413,25000 31413,31250 0,333 2 1
2018 01 02 09,0 10,50 31413,37500 31413,43750 0,667 3 1
2018 01 02 12,0 13,50 31413,50000 31413,56250 1,667 6 1
2018 01 02 15,0 16,50 31413,62500 31413,68750 1,000 4 1
2018 01 02 18,0 19,50 31413,75000 31413,81250 2,000 7 1
2018 01 02 21,0 22,50 31413,87500 31413,93750 1,000 4 1
2018 01 03 00,0 01,50 31414,00000 31414,06250 0,000 0 1
2018 01 03 03,0 04,50 31414,12500 31414,18750 0,667 3 1
2018 01 03 06,0 07,50 31414,25000 31414,31250 1,000 4 1
2018 01 03 09,0 10,50 31414,37500 31414,43750 0,667 3 1
2018 01 03 12,0 13,50 31414,50000 31414,56250 0,333 2 1
2018 01 03 15,0 16,50 31414,62500 31414,68750 1,000 4 1
2018 01 03 18,0 19,50 31414,75000 31414,81250 0,333 2 1
2018 01 03 21,0 22,50 31414,87500 31414,93750 0,000 0 1
2018 01 04 00,0 01,50 31415,00000 31415,06250 0,667 3 1
2018 01 04 03,0 04,50 31415,12500 31415,18750 0,000 0 1
```

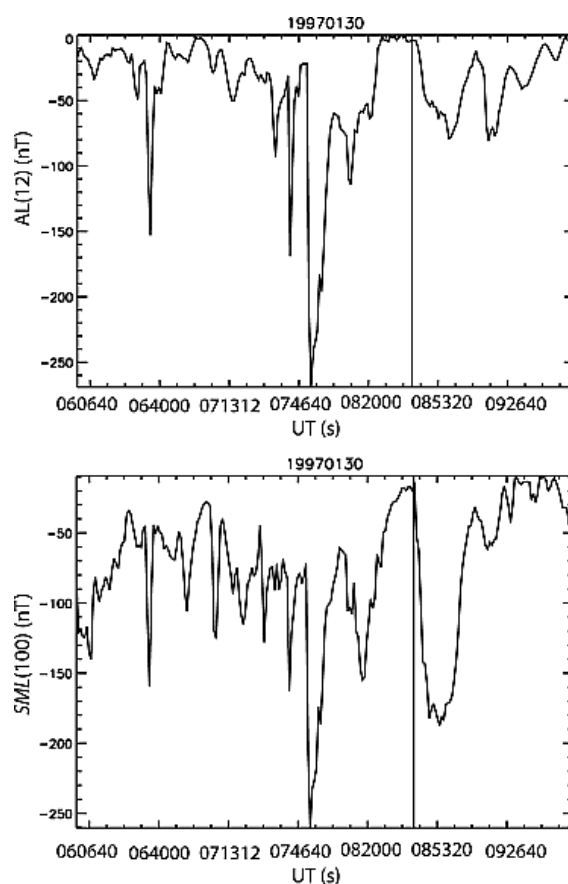
Εικόνα 11. Υπόδειγμα δεδομένων γεωμαγνητικής δραστηριότητας (K)

Το τρίωρο ισοδύναμο πλανητικό πλάτος ar προέρχεται από το Kp και το ημερήσιο ισοδύναμο πλανητικό πλάτος Ap είναι ο ημερήσιος μέσος όρος του ar. Το Kp είναι χωρίς μονάδες, το Ap και το ar είναι χωρίς μονάδες και μπορούν να

πολλαπλασιαστούν με 2 nT για να προκύψει η μέση γεωμαγνητική διαταραχή σε γεωμαγνητικό γεωγραφικό πλάτος 50 μοιρών. Τα Kp, ap και Ap εισήχθησαν από τον Bartels (1949, 1957). Το D υποδεικνύει εάν οι τιμές Kp και ap είναι οριστικές (D = 1) ή προκαταρκτικές (D=0) (Matzka et al, (2021)).

9.1.3. Auroral Electrojet (AE)

Για τα δεδομένα που αφορούσαν την παράμετρο AE, αποφασίστηκε η λήψη δεδομένων SME ως ποιο ακριβή λόγω του πλήθους των σταθμών από τους οποίους προέρχονται και τις καλύτερης ανταπόκρισης που έχουν στις γεωμαγνητικές καταιγίδες (Εικόνα 13) [34].



Εικόνα 12. Ανταπόκριση παραμέτρων AE και SME σε γεγονός την 30 Ιανουαρίου 1997 (Πηγή: Newell and Gjerloev, Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power, 2011)

Τα δεδομένα SME χορηγήθηκαν από τον οργανισμό SuperMAG για τους μήνες Ιανουάριο – Απρίλιο του έτους 2018. Η επιλογή των συγκεκριμένων μηνών έγινε εξαιτίας της ύπαρξης ημερών με έντονη ηλιακή δραστηριότητα. Τα αρχεία ήταν τύπου txt με μορφή όπως φαίνεται στην εικόνα 14.

```

=====
<year> <month> <day> <hour> <min> <sec> <SML (nT)> <SMU (nT)>
2018 01 01 00 00 00 -187 81
2018 01 01 00 01 00 -188 84
2018 01 01 00 02 00 -185 83
2018 01 01 00 03 00 -187 85
2018 01 01 00 04 00 -187 83
2018 01 01 00 05 00 -181 82
2018 01 01 00 06 00 -181 82
2018 01 01 00 07 00 -180 85
2018 01 01 00 08 00 -174 91
2018 01 01 00 09 00 -178 96
2018 01 01 00 10 00 -183 99
2018 01 01 00 11 00 -189 99
2018 01 01 00 12 00 -198 96
2018 01 01 00 13 00 -207 94
2018 01 01 00 14 00 -210 93
2018 01 01 00 15 00 -210 95
2018 01 01 00 16 00 -197 103
2018 01 01 00 17 00 -191 102
2018 01 01 00 18 00 -194 98
2018 01 01 00 19 00 -198 95
    
```

Εικόνα 13. Υπόδειγμα δεδομένων SME

9.1.3.1. SuperMAG

Το SuperMAG είναι μια παγκόσμια συνεργασία οργανισμών και εθνικών φορέων που σήμερα λειτουργούν πάνω από 300 επίγεια μαγνητόμετρα. Παρέχει εύκολη πρόσβαση σε επικυρωμένες διαταραχές μαγνητικού πεδίου στο ίδιο σύστημα συντεταγμένων, ταυτόσημη χρονική ανάλυση και με μια κοινή προσέγγιση αφαίρεσης γραμμής βάσης. Χρησιμοποιεί τρισδιάστατες διανυσματικές μετρήσεις του μαγνητικού πεδίου που λαμβάνονται από μαγνητόμετρα βάσεων εδάφους. Συμπεριλαμβάνονται σταθμοί που παρέχουν απόλυτες μετρήσεις (π.χ. Διαμαγνητικά Παρατηρητήρια) καθώς και σταθμοί που παρέχουν σχετικές μετρήσεις. Το SuperMAG εστιάζει στις διακυμάνσεις που προκαλούνται από τα ηλεκτρικά ρεύματα που ρέουν στην ιονόσφαιρα και τη μαγνητόσφαιρα και ως εκ τούτου αφαιρεί το κυρίαρχο και αργά μεταβαλλόμενο κύριο πεδίο της Γης.

9.1.4. Disturbance storm time (Dst)

Για τα δεδομένα που αφορούσαν την παράμετρο Dst, έγινε λήψη τους από World Data Center for Geomagnetism του Kyoto. Ο δείκτης Dst υπολογίζεται από δεδομένα γεωμαγνητικού πεδίου που εξετάστηκαν οπτικά για τεχνητούς θορύβους (M. Nose, T. Iyemori, M. Sugiura, T. Kamei (2015)). Τα δεδομένα ήταν της μορφής IAGA-2002

9.1.4.1. IAGA-2002

Αυτή η μορφή προορίζεται ως μορφή ανταλλαγής δεδομένων για γεωμαγνητικά δεδομένα (δείγματα και μέσους) από παρατηρητήρια και σταθμούς σε χρονικά διαστήματα από χιλιοστά του δευτερολέπτου έως και μηνιαίων μέσων. Η μορφή περιλαμβάνει:

- Δώδεκα (12) υποχρεωτικές και μία (1) προαιρετική εγγραφή κεφαλίδας αρχείου
- Απεριόριστες προαιρετικές εγγραφές σχολίων
- Μία (1) υποχρεωτική εγγραφή κεφαλίδας δεδομένων
- Μια σειρά από αρχεία δεδομένων.
- Κάθε εγγραφή έχει μήκος 70 χαρακτήρες συν την επιστροφή / τροφοδοσία γραμμής που εξαρτάται από το μηχάνημα.

DATE	TIME	DOY	DST
1/1/2018	12:00:00 πμ		1
1/1/2018	1:00:00 πμ		1
1/1/2018	2:00:00 πμ		1
1/1/2018	3:00:00 πμ		1
1/1/2018	4:00:00 πμ		1
1/1/2018	5:00:00 πμ		1
1/1/2018	6:00:00 πμ		1
1/1/2018	7:00:00 πμ		1
1/1/2018	8:00:00 πμ		1
1/1/2018	9:00:00 πμ		1
1/1/2018	10:00:00 πμ		1
1/1/2018	11:00:00 πμ		1
1/1/2018	12:00:00 μμ		1
1/1/2018	1:00:00 μμ		1
1/1/2018	2:00:00 μμ		1
1/1/2018	3:00:00 μμ		1
1/1/2018	4:00:00 μμ		1
1/1/2018	5:00:00 μμ		1
1/1/2018	6:00:00 μμ		1
1/1/2018	7:00:00 μμ		1
1/1/2018	8:00:00 μμ		1
1/1/2018	9:00:00 μμ		1
1/1/2018	10:00:00 μμ		1
1/1/2018	11:00:00 μμ		1

Εικόνα 14. Υπόδειγμα δεδομένων Dst

9.1.5. Solar Radio Flux

Τα δεδομένα που αφορούν το Solar Radio Flux (ηλιακή ροή), ήταν διαθέσιμα από την κυβέρνηση του Καναδά, στο site www.spaceweather.gc.ca.

Η βάση δεδομένων που είναι διαθέσιμη εδώ περιλαμβάνει δύο στοιχεία: μετρήσεις του Solar Radio Flux των 10,7 cm και ημερήσιες καταγραφές του Solar Radio Flux. Κάθε μέτρηση της ηλιακής ροής 10,7 cm εκφράζεται σε τρεις τιμές: τις τιμές που παρατηρήθηκαν, τις προσαρμοσμένες και τιμές URSI σειράς D.

Η παρατηρούμενη τιμή είναι ο αριθμός που μετράται από το ηλιακό ραδιοτηλεσκόπιο. Αυτό διαμορφώνεται από δύο μεγέθη: το επίπεδο της ηλιακής δραστηριότητας και τη μεταβαλλόμενη απόσταση μεταξύ της Γης και του Ήλιου. Δεδομένου ότι είναι ένα μέτρο των εκπομπών που οφείλονται στην ηλιακή δραστηριότητα που χτυπά τη Γη, αυτή είναι η ποσότητα που πρέπει να χρησιμοποιείται όταν μελετώνται επίγεια φαινόμενα.

Όταν μελετάται ο Ήλιος, η ετήσια διαμόρφωση της ηλιακής ροής των 10,7 cm από τη μεταβαλλόμενη απόσταση μεταξύ της Γης και του Ήλιου είναι ανεπιθύμητη. Ωστόσο, ένα υποπροϊόν των υπολογισμών που απαιτούνται από τους δέκτες ώστε να αποκτήσουν και να παρακολουθήσουν σωστά τον Ήλιο είναι η απόσταση μεταξύ του Ήλιου και της Γης. Επομένως, παράγεται μια πρόσθετη ποσότητα, διορθωμένη από διακυμάνσεις στην απόσταση Γης-Ήλιου και δίνεται για τη μέση απόσταση. Αυτό ονομάζεται προσαρμοσμένη τιμή. Οι απόλυτες μετρήσεις της πυκνότητας της ροής είναι αρκετά δύσκολες, και στα πρώτα χρόνια της ηλιακής ραδιοαστρονομίας, καταβλήθηκε σημαντική προσπάθεια σε όλο τον κόσμο για να γίνουν απόλυτες μετρήσεις της πυκνότητας της ηλιακής ροής σε πολλές διαφορετικές συχνότητες. Έγινε τότε μια προσπάθεια να χωρέσουν όλα αυτά τα διάφορα δεδομένα σε ένα φάσμα. Σε κάθε σύνολο μετρήσεων δόθηκε στη συνέχεια ένας παράγοντας κλιμάκωσης που θα τις μετακινούσε απευθείας στο προσαρμοσμένο φάσμα. Για την ηλιακή ροή 10,7 cm υπολογίστηκε ένας συντελεστής κλιμάκωσης 0,9. Έτσι, δίνεται επίσης στη βάση δεδομένων το Series D Flux, που είναι η προσαρμοσμένη τιμή πολλαπλασιαζόμενη επί 0,9.

Τρεις προσδιορισμοί ροής γίνονται κάθε μέρα. Μεταξύ Μαρτίου και Οκτωβρίου οι μετρήσεις γίνονται στις 17.00, 20.00 και 23.00 UT. Από τον Νοέμβριο έως τον Φεβρουάριο, οι χρόνοι προσδιορισμού της ροής αλλάζουν σε 1800, 2000 και 2200, έτσι ώστε ο Ήλιος να βρίσκεται αρκετά ψηλά πάνω από τον ορίζοντα για να γίνει μια καλή μέτρηση.

Η ηλιακή ροή 10,7 cm δίνεται σε μονάδες ηλιακής ροής (ένα sfu = 10-22W m⁻² Hz⁻¹). Τα αρχεία είναι txt της μορφής:

- Ημερομηνία καταγραφής
- Ώρα
- Ιουλιανή ημερομηνία
- Αριθμός περιστροφής Carrington
- παρατηρούμενη ροή
- προσαρμοσμένη ροή
- ροή της σειράς D.

fluxdate	fluxtime	fluxjulian	fluxcarrington	fluxobsflux	fluxadjflux	fluxursi
20180101	180000	2458120.239	2199.069	70	67.7	60.9
20180101	200000	2458120.322	2199.072	69.1	66.8	60.1
20180101	220000	2458120.416	2199.0723	72.6	70.1	63.1
20180102	180000	2458121.25	2199.1029	72	69.6	62.7
20180102	200000	2458121.322	2199.108	69.5	67.2	60.5
20180102	220000	2458121.406	2199.111	70	67.7	60.9
20180103	180000	2458122.239	2199.142	71.4	69	62.1
20180103	200000	2458122.322	2199.145	70.7	68.3	61.5
20180103	220000	2458122.406	2199.148	70.5	68.2	61.4
20180104	180000	2458123.239	2199.179	70.1	67.8	61
20180104	200000	2458123.322	2199.182	69.5	67.2	60.5
20180104	220000	2458123.406	2199.185	69.6	67.3	60.5
20180105	180000	2458124.239	2199.215	69.1	66.8	60.1
20180105	200000	2458124.322	2199.218	69.3	67	60.3
20180105	220000	2458124.406	2199.221	68.6	66.4	59.7
20180106	180000	2458125.239	2199.252	69.6	67.3	60.6
20180106	200000	2458125.322	2199.255	69.4	67.1	60.4
20180106	220000	2458125.406	2199.258	69.3	67	60.3

Εικόνα 15. Υπόδειγμα δεδομένων Solar Radio Flux

9.1.6. Sunspot Number

Τα δεδομένα που αφορούσαν το Sunspot Number (ηλιακή κηλίδα), ελήφθησαν από WDC-SILSO, Βασιλικό Αστεροσκοπείο του Βελγίου στις Βρυξέλλες. Ο ημερήσιος συνολικός αριθμός ηλιακών κηλίδων προκύπτει από τον τύπο: $R = N_s + 10 * N_g$, με N_s τον αριθμό των κηλίδων και N_g τον αριθμό των ομάδων που υπολογίζονται σε ολόκληρο τον ηλιακό δίσκο. Το αρχείο είναι μορφής txt και περιλάμβανε όλα τα δεδομένα από το 1992 έως το 2020 ως ακολούθως:

- Ημερομηνία
- Ημερήσιος συνολικός αριθμός ηλιακών κηλίδων
- Ημερήσιος αριθμός βόρειων ηλιακών κηλίδων
- Ημερήσιος αριθμός νότιων ηλιακών κηλίδων
- Τυπική απόκλιση των ακατέργαστων ημερήσιων συνολικών δεδομένων ηλιακών κηλίδων
- Τυπική απόκλιση ακατέργαστων ημερήσιων δεδομένων βόρειων ηλιακών κηλίδων
- Τυπική απόκλιση ακατέργαστων ημερήσιων δεδομένων νότιων ηλιακών κηλίδων
- Αριθμός παρατηρήσεων για ημερήσιο συνολικό αριθμό ηλιακών κηλίδων
- Αριθμός παρατηρήσεων για ημερήσιο αριθμό βόρειων ηλιακών κηλίδων
- Αριθμός παρατηρήσεων για τον ημερήσιο αριθμό νότιων ηλιακών κηλίδων
- Οριστικός / Προσωρινός Δείκτης

```
Date,Daily Total Sunspot Number,Daily North Sunspot Number,  
2020-04-30,20.0,8.0,12.0,1.2,0.6,1.0,20.0,19.0,31.0,0.0  
2020-04-29,24.0,12.0,12.0,1.6,0.0,0.5,22.0,5.0,2.0,0.0  
2020-04-28,15.0,0.0,15.0,2.6,0.0,0.8,18.0,24.0,13.0,0.0  
2020-04-27,29.0,0.0,29.0,2.0,0.0,1.1,20.0,35.0,2.0,0.0  
2020-04-26,14.0,0.0,14.0,1.0,0.0,0.0,23.0,33.0,3.0,0.0  
2020-04-25,6.0,0.0,6.0,0.4,0.0,0.5,22.0,32.0,19.0,0.0  
2020-04-24,0.0,0.0,0.0,0.0,0.0,0.0,42.0,33.0,33.0,0.0  
2020-04-23,0.0,0.0,0.0,0.0,0.0,0.0,43.0,35.0,35.0,0.0  
2020-04-22,0.0,0.0,0.0,0.0,0.0,0.0,43.0,34.0,34.0,0.0  
2020-04-21,0.0,0.0,0.0,0.0,0.0,0.0,37.0,29.0,29.0,0.0  
2020-04-20,0.0,0.0,0.0,0.0,0.0,0.0,38.0,28.0,28.0,0.0  
2020-04-19,0.0,0.0,0.0,0.0,0.0,0.0,37.0,28.0,28.0,0.0  
2020-04-18,0.0,0.0,0.0,0.0,0.0,0.0,37.0,28.0,28.0,0.0  
2020-04-17,0.0,0.0,0.0,0.0,0.0,0.0,38.0,29.0,29.0,0.0  
2020-04-16,0.0,0.0,0.0,0.0,0.0,0.0,42.0,33.0,33.0,0.0  
2020-04-15,0.0,0.0,0.0,0.0,0.0,0.0,44.0,35.0,35.0,0.0  
2020-04-14,0.0,0.0,0.0,0.0,0.0,0.0,41.0,35.0,35.0,0.0  
2020-04-13,0.0,0.0,0.0,0.0,0.0,0.0,38.0,30.0,30.0,0.0  
2020-04-12,0.0,0.0,0.0,0.0,0.0,0.0,38.0,29.0,29.0,0.0  
2020-04-11,0.0,0.0,0.0,0.0,0.0,0.0,46.0,36.0,36.0,0.0  
2020-04-10,0.0,0.0,0.0,0.0,0.0,0.0,45.0,35.0,35.0,0.0  
2020-04-09,0.0,0.0,0.0,0.0,0.0,0.0,47.0,36.0,36.0,0.0  
2020-04-08,0.0,0.0,0.0,0.0,0.0,0.0,46.0,35.0,36.0,0.0  
2020-04-07,0.0,0.0,0.0,0.0,0.0,0.0,43.0,34.0,34.0,0.0
```

Εικόνα 16. Υπόδειγμα δεδομένων Sunspot Number

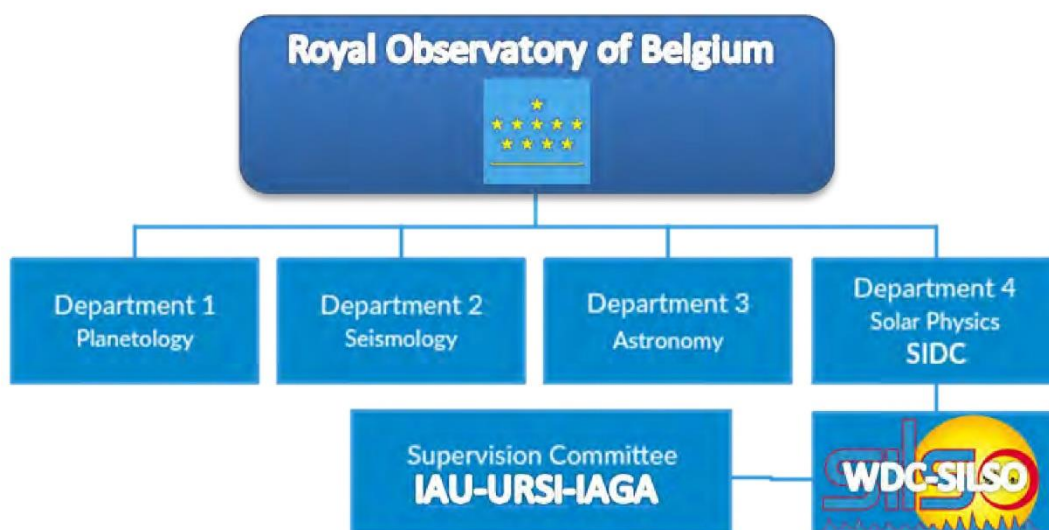
9.1.6.1. WDC-SILSO, Βασιλικό Αστεροσκοπείο του Βελγίου

Το Βασιλικό Αστεροσκοπείο του Βελγίου, βρίσκεται στον δήμο Uccle των Βρυξελλών από το 1890. Ιδρύθηκε στο Saint-Josse-ten-Noode το 1826 από τον William I υπό την

παρόρμηση του Adolphe Quetelet. Ήταν το σπίτι ενός ανακλαστήρα Zeiss με διάμετρο 100 cm το πρώτο μισό του 20ου αιώνα, ένα από τα μεγαλύτερα τηλεσκόπια στον κόσμο εκείνη την εποχή. Κατέχει μια ποικιλία άλλων αστρονομικών οργάνων, όπως αστρογράφους, καθώς και μια σειρά από εξοπλισμό σειсмоγράφου (για την ανίχνευση σεισμών). Οι κύριες δραστηριότητές του είναι:

- Συστήματα αναφοράς και γεωδυναμική.
- Αστρομετρία και δυναμική ουράνιων σωμάτων.
- Αστροφυσική.
- Ηλιακή φυσική.

Το WDC-SILSO (World Data Center for the production, preservation and dissemination of the international sunspot number) είναι το Παγκόσμιο Κέντρο Δεδομένων για την παραγωγή, τη διατήρηση και τη διάδοση του διεθνούς αριθμού ηλιακών κηλίδων. Είναι μια δραστηριότητα της Επιχειρησιακής Διεύθυνσης «Solar Physics and Space Weather» επίσης γνωστή διεθνώς ως Κέντρο Ανάλυσης Δεδομένων Ηλιακών Επιρροών (SIDC). Το SIDC είναι τμήμα του Βασιλικού Αστεροσκοπείου του Βελγίου. Αποστολή του είναι να διατηρήσει, να αναπτύξει και να διαδώσει τη γνώση των μακροπρόθεσμων παραλλαγών της ηλιακής δραστηριότητας, ως εισροή αναφοράς στις μελέτες του μηχανισμού του ηλιακού κύκλου και της ηλιακής ώθησης στο κλίμα της Γης.



Εικόνα 17. Βασιλικό Αστεροσκοπείο Βελγίου (Πηγή: wwwbis.sidc.be)

9.1.7. Energetic Particle Flux

Τα δεδομένα που αφορούσαν το Energetic Particle Flux (Ροή ενεργών σωματιδίων), ελήφθησαν από το SWPC-NOAA. Το αρχείο είναι σε μορφή txt και περιλαμβάνει

δεδομένα πρωτονίων, ηλεκτρονίων και νετρονίων σε διάφορες τιμές ενέργειες όπως παρατηρούνται από τον κύριο δορυφόρο GOES.

GOES-15 Proton FluEncE			GOES-15 ElElectron FluEncE		NEutron
Protons/cm2-day-sr			ElElectrons/cm2-day-sr		Monitor
>1 MEV	>10 MEV	>100 MEV	>0.8 MEV	>2 MEV	% of bkgd
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
560000	16000	3400	500000000	1600000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
470000	15000	3600	1100000000	2300000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99
400000	16000	3500	970000000	2000000	-999.99

Εικόνα 18. Υπόδειγμα δεδομένων Energetic Particle Flux.

9.1.7.1. SWPC-NOAA

Η National Oceanic and Atmospheric Administration είναι μια αμερικανική επιστημονική και ρυθμιστική υπηρεσία στο Υπουργείο Εμπορίου των Ηνωμένων Πολιτειών που προβλέπει καιρικές συνθήκες, παρακολουθεί τις ωκεάνιες και ατμοσφαιρικές συνθήκες, χαρτογραφεί τις θάλασσες, διεξάγει εξερεύνηση βαθέων υδάτων και διαχειρίζεται την αλιεία και την προστασία θαλάσσιων θηλαστικών και απειλούμενων ειδών στην αποκλειστική οικονομική ζώνη των ΗΠΑ.

Το Space Weather Prediction Center (SWPC) είναι ένα εργαστήριο και κέντρο υπηρεσιών της Εθνικής Μετεωρολογικής Υπηρεσίας των ΗΠΑ (NWS), μέρος της Εθνικής Υπηρεσίας Ωκεανών και Ατμόσφαιρας (NOAA), που βρίσκεται στο Boulder του Κολοράντο. Το SWPC παρακολουθεί και προβλέπει συνεχώς το διαστημικό περιβάλλον της Γης, παρέχοντας ηλιακές-γήινες πληροφορίες. Το SWPC είναι η επίσημη πηγή ειδοποιήσεων και προειδοποιήσεων διαστημικού καιρού για τις Ηνωμένες Πολιτείες.

9.1.8. Solar Flares

Τα δεδομένα που αφορούσαν τα Solar Flares (Ηλιακές εκλάμψεις), ελήφθησαν από τη NASA μέσω του προγράμματος RHESSI. Το αρχείο είναι σε μορφή txt και περιλαμβάνει τις καταγραφές των εκλάμψεων για την περίοδο μελέτης όπως φαίνεται στην εικόνα 20.

HESSI Flare List (generated 7-Oct-2020 13:36)

Total # flares: 121206 Time range: 12-Feb-2002 21:29:56.000 - 3-Mar-2018 04:12:20.000

Flare	Start time	Peak	End	Dur s	Peak c/s	Total Counts	Energy keV	X asec	Pos asec	Y asec	Radial asec	AR θ	A0 D
18010601	6-Jan-2018	16:31:44	16:32:26	16:32:56	72	128	4152	3-6	0	0	0	0	A0 D
18011801	18-Jan-2018	07:29:24	07:34:58	07:42:56	812	256	86808	25-50	88	-121	150	2696	A0 P
18011802	18-Jan-2018	08:11:08	08:12:54	08:16:16	308	40	13446	6-12	110	-129	169	2696	A0 D
18012001	20-Jan-2018	04:50:48	04:51:58	04:53:28	160	20	4200	6-12	501	-130	517	2696	A0 P
18012002	20-Jan-2018	04:53:28	04:54:30	04:57:28	240	16	5608	6-12	499	-115	512	2696	A0 P
18012202	22-Jan-2018	02:43:28	02:53:22	03:15:04	1896	40	82840	6-12	817	-157	832	2696	A0 D
18020401	4-Feb-2018	21:14:28	21:15:06	21:19:12	284	22	4696	6-12	-960	-115	966	2699	A0 D
18020402	4-Feb-2018	21:19:12	21:22:18	21:28:12	540	14	6264	6-12	-960	-115	966	2699	A0 D
18020601	6-Feb-2018	18:48:04	18:55:38	19:04:20	976	36	39272	6-12	-821	113	828	2699	A0 D
18020602	6-Feb-2018	19:04:20	19:05:30	19:09:08	288	13	5544	6-12	-845	-35	845	2699	A0 D
18020603	6-Feb-2018	19:09:08	19:12:30	19:17:24	496	13	9024	6-12	-829	-58	831	2699	A0 D
18020701	7-Feb-2018	13:34:44	13:44:46	13:45:00	616	288	87832	25-50	-689	-45	690	2699	A0 D
18020702	7-Feb-2018	14:16:52	14:17:02	14:19:04	132	6	1160	6-12	-729	-38	730	2699	A0 D
18020703	7-Feb-2018	14:19:04	14:20:22	14:24:40	336	6	2328	6-12	-729	-38	730	2699	A0 D
18020801	8-Feb-2018	17:00:04	17:00:14	17:11:24	680	15	9808	6-12	-508	-49	510	2699	A0 D
18020802	8-Feb-2018	18:42:44	18:44:38	18:46:12	208	24	4648	6-12	-499	-38	500	2699	A0 D
18020803	8-Feb-2018	19:52:56	19:53:54	19:56:56	240	25	7646	6-12	-515	-16	515	2699	A0 D
18020901	9-Feb-2018	14:53:32	14:55:22	14:58:32	300	13	3008	6-12	-345	-38	347	2699	A0 D
18020902	9-Feb-2018	15:45:28	15:46:54	15:52:36	428	28	9088	6-12	0	0	0	2699	A0 D
18020903	9-Feb-2018	16:41:28	16:42:54	16:43:32	124	18	2888	6-12	-345	-38	347	2699	A0 D
18020904	9-Feb-2018	18:15:56	18:17:26	18:17:40	104	16	1656	6-12	-268	-38	271	2699	A0 D
18021001	10-Feb-2018	13:04:36	13:06:46	13:07:04	148	15	2224	6-12	-115	-38	121	2699	A0 D
18022601	26-Feb-2018	15:49:56	15:51:18	15:53:52	236	16	3312	6-12	115	192	223	2700	A0 D
18030301	3-Mar-2018	04:04:08	04:06:22	04:12:20	492	42	15816	6-12	926	128	935	2700	A0 P

Εικόνα 19. Υπόδειγμα δεδομένων Solar Flares.

9.1.8.1. NASA

Η Εθνική Υπηρεσία Αεροναυτικής και Διαστήματος (NASA) είναι μια ανεξάρτητη υπηρεσία της ομοσπονδιακής κυβέρνησης των ΗΠΑ που είναι υπεύθυνη για το μη στρατιωτικό διαστημικό πρόγραμμα, καθώς και για την αεροναυτική και διαστημική έρευνα.

Η NASA ιδρύθηκε το 1958, διαδεχόμενη την Εθνική Συμβουλευτική Επιτροπή για την Αεροναυτική (NACA). Ο νέος οργανισμός επρόκειτο να έχει έναν σαφώς πολιτικό προσανατολισμό, ενθαρρύνοντας ειρηνικές εφαρμογές στη διαστημική επιστήμη. Από την ίδρυσή της, οι περισσότερες προσπάθειες εξερεύνησης του διαστήματος των ΗΠΑ καθοδηγούνται από τη NASA, συμπεριλαμβανομένων των αποστολών προσγείωσης του Apollo Moon, του διαστημικού σταθμού Skylab και αργότερα του Διαστημικού Λεωφορείου. Η NASA υποστηρίζει τον Διεθνή Διαστημικό Σταθμό και επιβλέπει την ανάπτυξη του διαστημικού σκάφους Orion, του Space Launch System, των οχημάτων Commercial Crew και του σχεδιαζόμενου διαστημικού σταθμού Lunar Gateway. Ο οργανισμός είναι επίσης υπεύθυνος για το Πρόγραμμα Υπηρεσιών Εκτόξευσης, το οποίο παρέχει επίβλεψη των λειτουργιών εκτόξευσης και διαχείριση αντίστροφης μέτρησης για εκτοξεύσεις της NASA χωρίς πλήρωμα.

Η επιστήμη της NASA επικεντρώνεται στην καλύτερη κατανόηση της Γης μέσω του Συστήματος Παρατήρησης της Γης· στην προώθηση της ηλιοφυσικής μέσω των προσπαθειών του Ερευνητικού Προγράμματος Ηλιοφυσικής της Διεύθυνσης Επιστημονικής Αποστολής στην εξερεύνηση σωμάτων σε όλο το Ηλιακό Σύστημα με προηγμένα ρομποτικά διαστημόπλοια όπως το New Horizons και έρευνα θεμάτων αστροφυσικής, όπως το Big Bang, μέσω των Μεγάλων Αστεροσκοπειών και των σχετικών προγραμμάτων. Το Space Weather Prediction Center (SWPC) είναι ένα εργαστήριο και κέντρο υπηρεσιών της Εθνικής Μετεωρολογικής Υπηρεσίας των ΗΠΑ (NWS), μέρος της Εθνικής Υπηρεσίας Ωκεανών και Ατμόσφαιρας (NOAA), που

βρίσκεται στο Boulder του Κολοράντο. Το SWPC παρακολουθεί και προβλέπει συνεχώς το διαστημικό περιβάλλον της Γης, παρέχοντας ηλιακές-γήινες πληροφορίες. Το SWPC είναι η επίσημη πηγή ειδοποιήσεων και προειδοποιήσεων διαστημικού καιρού για τις Ηνωμένες Πολιτείες.

9.1.8.2. *RHESSI*

Το Ramaty High Energy Solar Spectroscopic Imager (RHESSI, αρχικά High Energy Solar Spectroscopic Imager ή HESSI ή Explorer 81) ήταν ένα παρατηρητήριο ηλιακών εκλάμψεων της NASA. Ήταν η έκτη αποστολή στο πρόγραμμα Small Explorer (SMEX), που επιλέχθηκε τον Οκτώβριο του 1997 και ξεκίνησε στις 5 Φεβρουαρίου 2002, στις 20:58:12 UTC. Λόγω δυσκολιών που προέκυψαν στο σύστημα επικοινωνιών, σταμάτησε τη λειτουργία στις 11 Απριλίου 2018. Η κύρια αποστολή του ήταν να εξερευνήσει τη φυσική της επιτάχυνσης των σωματιδίων και την απελευθέρωση ενέργειας στις ηλιακές εκλάμψεις.

9.1.9. Coronal Mass Ejections

Οι Coronal Mass Ejections (CMEs) είναι τεράστιες εκρήξεις πλάσματος από τον ήλιο. Τα δεδομένα που αφορούσαν τα CMEs, ελήφθησαν από τη NASA μέσω του προγράμματος SOHO-LASCO. Το αρχείο είναι σε μορφή txt και περιλαμβάνει τις καταγραφές των εκλάμψεων για την περίοδο μελέτης όπως φαίνεται στην εικόνα 19.

Ένας από τους επιστημονικούς στόχους του SOHO-LASCO (Large Angle and Spectrometric Coronagraph) είναι να κατανοήσει γιατί συμβαίνουν αυτά τα συμβάντα. Πιστεύεται ότι προκαλούνται από αστάθειες στο ηλιακό μαγνητικό πεδίο, το οποίο εξελίσσεται συνεχώς. Ένας άλλος επιστημονικός στόχος είναι να κατανοηθούν οι επιδράσεις των CME στον διαπλανητικό χώρο και πολύ σημαντικό τι επιπτώσεις έχουν οι CME όταν συναντούν το περιβάλλον της γης.

Ο διαθέσιμος κατάλογος CME δημιουργείται και διατηρείται στο CDAW Data Center από τη NASA και το Καθολικό Πανεπιστήμιο της Αμερικής σε συνεργασία με το Ναυτικό Ερευνητικό Εργαστήριο. Το SOHO (NASA Solar and Heliospheric Observatory) είναι ένα έργο διεθνούς συνεργασίας μεταξύ της ESA και της NASA.

First C2 Appearance Date [UT]	Central PA [deg]	Angular Width [deg]	Linear Speed [km/s]	2nd-order Speed at final height [km/s]	2nd-order Speed at 20 Rs [km/s]	Accel [m/s ²]	Mass [gram]	Kinetic Energy [erg]	MPA [deg]	s, plots, & links	Remarks
2018/01/01 05:48:05	315	28	100	182	369	5.8*1	2.4e+15	1.2e+29	301	C2 C3 PHTX DST Java	Very Poor Event
2018/01/02 12:05	48	12	377	307	0	-67.5*1	----	----	51	C2 C3 PHTX DST Java	Very Poor Event; Only C2
2018/01/03 12:05	104	9	244	301	355	4.0*1	----	----	111	C2 C3 PHTX DST Java	Poor Event
2018/01/04 03:48:05	278	26	128	175	402	6.4*1	4.6e+13	3.7e+27	276	C2 C3 PHTX DST Java	Very Poor Event; Only C2
2018/01/04 20:05	191	39	241	394	782	25.0*1	5.8e+13	1.7e+28	192	C2 C3 PHTX DST Java	Poor Event

Εικόνα 20. Υπόδειγμα δεδομένων Coronal Mass Ejections.

9.2. Επεξεργασία δεδομένων

Τα δεδομένα όπως περιεγράφηκαν παραπάνω βρισκόντουσαν σε διάφορες μορφές αρχείων και ήταν καταγεγραμμένα σε διαφορετικά χρονικά διαστήματα όπως φαίνεται και στον πίνακα 1. Αφού επεξεργάστηκαν όλα τα δεδομένα και ήρθαν σε ίδια μορφή, αποφασίστηκε να ταξινομηθούν σε διαστήματα 3 ωρών (00.00-03.00-06.00-09.00-12.00-15.00-18.00-21.00) όπως φαίνεται στο Παράρτημα 1 και να εισαχθούν σε ένα αρχείο xlsx ώστε να είναι πιο διαχειρίσιμα και εύληπτα για οποιονδήποτε χρήστη. Για την ταξινόμησή τους σε διαστήματα 3 ωρών έπρεπε να γίνουν κάποιες διαδικασίες. Αρχικά, όσα δεδομένα ήταν σε διαστήματα μικρότερα των 3 ωρών (πχ τα δεδομένα TEC) λαμβάνοντας τους μέσους όρους για τα διαστήματα των 3 ωρών, προέκυψε η τελική τους τιμή. Για δεδομένα με διαστήματα μεγαλύτερα των 3 ωρών όπως τα Sunspot Number (Ημερήσια), τοποθετήθηκε η καταγεγραμμένη τιμή σε όλα τα τρίωρα διαστήματα της εκάστοτε μέρας. Για τα δεδομένα που είχα καταγραφές συγκεκριμένων ωρών όπως το Solar Radio Flux, συμπληρώθηκε ο πίνακας κάθε μέρας με χρήση γραμμικής παρεμβολής. Τέλος, για τα δεδομένα τυχαίας εμφάνισης όπως τα CMEs, τοποθετήθηκαν στην αντίστοιχη μέρα και ώρα εμφάνισής τους και όλες οι υπόλοιπες ώρες συμπληρώθηκαν με 0. Ολοκληρώνοντας, αφού αντικαταστάθηκαν όσα κελιά που δεν είχαν δεδομένα λόγω κάποιου προβλήματος στο δέκτη με 0, συμπληρώθηκε ο πίνακας για κάθε μήνα σε ξεχωριστές καρτέλες (Παράρτημα 1). Τέλος, παρήχθησαν 4 αρχεία xlsx για κάθε μήνα ώστε να μπορεί να χρησιμοποιηθεί από το πρόγραμμα RF_Solar.

9.3. Εκτέλεση προγράμματος – Εξαγωγή αποτελεσμάτων – Αξιολόγηση

Αφού ολοκληρώθηκε η επεξεργασία των δεδομένων και δημιουργήθηκαν τα αρχεία για κάθε μήνα, ξεκίνησε η εκτέλεση του προγράμματος όπως αναφέρεται στο

προηγούμενο κεφάλαιο. Ακολουθήθηκαν τα παρακάτω βήματα για την εξαγωγή των αποτελεσμάτων:

- Εισήχθη το αρχείο δεδομένων του μήνα μελέτης.
- Επιλέχθηκε η αντίστοιχη επίλυση για STEC ή VTEC.
- Επιλέχθηκε test size 40% ως το ελάχιστο με ορθά αποτελέσματα.
- Έγινε έλεγχος για την εύρεση του βέλτιστου παλινδρομητή σύμφωνα με τα δεδομένα του μήνα μελέτης.
- Επιλέχθηκες ο βέλτιστος παλινδρομητής.
- Εξήχθησαν τα αποτελέσματα και παράχθηκαν τα γραφήματα για κάθε μήνα.

9.3.1. Σύγκριση παλινδρομητών

Η σύγκριση των παλινδρομητών που αναφέρθηκαν και στο κεφάλαιο 7 (απλής γραμμικής παλινδρόμησης, SVM παλινδρόμησης, Gaussian Process, δέντρων απόφασης, Random Forest και XGBoost) έγινε με χρήση rMSE και MAE. Για κάθε μήνα τα αποτελέσματα της σύγκρισης εκτυπώθηκαν σε πίνακα όπως φαίνεται και παρακάτω και βάσει αυτού επιλέχθηκε η βέλτιστη λύση.

Πίνακας 3. Πίνακας σύγκρισης παλινδρομητών Ιανουαρίου για STEC και VTEC αντίστοιχα

	rMSE	MAE
Linear Regression	12.9865	10.5401
Decision Tree	10.6696	8.4739
Support Vector Regression	12.9832	10.7950
Gaussian Process Regressor	12.9505	10.8157
Random Forest Regressor	7.1382	5.5434
XGBOOST Regressor	5.2099	4.0715

	rMSE	MAE
Linear Regression	9.5424	3.4367
Decision Tree	1.9869	1.6163
Support Vector Regression	2.7397	2.3539
Gaussian Process Regressor	2.7690	2.2628
Random Forest Regressor	1.6610	1.2899
XGBOOST Regressor	1.5059	1.1479

Πίνακας 4. Πίνακας σύγκρισης παλινδρομητών Φεβρουαρίου για STEC και VTEC αντίστοιχα

	rMSE	MAE
Linear Regression	1,936.8111	265.4322
Decision Tree	10.0302	7.9035
Support Vector Regression	13.3100	11.2984
Gaussian Process Regressor	13.4311	11.3024
Random Forest Regressor	6.2555	4.9465
XGBOOST Regressor	5.0705	4.2008

	rMSE	MAE
Linear Regression	587.7603	71.0805
Decision Tree	2.1142	1.8109
Support Vector Regression	3.4594	2.8827
Gaussian Process Regressor	4.3563	3.1991
Random Forest Regressor	1.2494	0.9808
XGBOOST Regressor	1.2895	1.0335

Πίνακας 5. Πίνακας σύγκρισης παλινδρομητών Μαρτίου για STEC και VTEC αντίστοιχα

	rMSE	MAE
Linear Regression	50.4985	16.7076
Decision Tree	8.3981	7.0983
Support Vector Regression	11.5203	9.7260
Gaussian Process Regressor	11.9241	10.0540
Random Forest Regressor	4.5382	3.5558
XGBOOST Regressor	3.8422	2.9757

	rMSE	MAE
Linear Regression	3.7089	2.9620
Decision Tree	1.9655	1.5427
Support Vector Regression	3.0418	2.5679
Gaussian Process Regressor	3.0755	2.5429
Random Forest Regressor	1.1954	0.9487
XGBOOST Regressor	1.0919	0.8508

Πίνακας 6. Πίνακας σύγκρισης παλινδρομητών Απριλίου για STEC και VTEC αντίστοιχα

	rMSE	MAE
Linear Regression	10.9658	8.7339
Decision Tree	5.9504	4.6787
Support Vector Regression	10.0178	7.9809
Gaussian Process Regressor	10.3219	8.3171
Random Forest Regressor	3.8886	3.0445
XGBOOST Regressor	3.4691	2.5869

	rMSE	MAE
Linear Regression	2.9962	2.2921
Decision Tree	1.5753	1.2331
Support Vector Regression	2.6657	2.2130
Gaussian Process Regressor	2.6803	2.1523
Random Forest Regressor	1.1622	0.8806
XGBOOST Regressor	1.0450	0.7966

Όπως παρουσιάζεται σε όλους τους παραπάνω πίνακες, ο αλγόριθμος XGBoost έχει καλύτερη προσαρμογή σε όλα τα δεδομένα και επιλέχθηκε η χρήση του σε όλους τους μήνες.

9.3.1.1. rMSE

Η απόκλιση ρίζας μέσου τετραγώνου (rMSD) ή σφάλμα ρίζας μέσου τετραγώνου (rMSE) είναι ένα συχνά χρησιμοποιούμενο μέτρο των διαφορών μεταξύ των τιμών (τιμές δείγματος ή πληθυσμού) που προβλέπονται από ένα μοντέλο ή έναν εκτιμητή και των τιμών που παρατηρούνται. Το RMSE αντιπροσωπεύει την τετραγωνική ρίζα της δεύτερης ροπής δείγματος των διαφορών μεταξύ των προβλεπόμενων τιμών και των παρατηρούμενων τιμών ή τον τετραγωνικό μέσο όρο αυτών των διαφορών. Αυτές οι αποκλίσεις ονομάζονται υπολείμματα όταν οι υπολογισμοί εκτελούνται στο δείγμα δεδομένων που χρησιμοποιήθηκε για εκτίμηση και ονομάζονται σφάλματα (ή σφάλματα πρόβλεψης) όταν υπολογίζονται εκτός δείγματος. Το RMSE χρησιμεύει για τη συγκέντρωση των μεγεθών των σφαλμάτων στις προβλέψεις για διάφορα σημεία δεδομένων σε ένα μόνο μέτρο προγνωστικής ισχύος. Το RMSE είναι ένα μέτρο ακρίβειας, για τη σύγκριση σφαλμάτων πρόβλεψης διαφορετικών μοντέλων για ένα συγκεκριμένο σύνολο δεδομένων και όχι μεταξύ συνόλων δεδομένων, καθώς εξαρτάται από την κλίμακα.

Το RMSE είναι πάντα μη αρνητικό και μια τιμή 0 (σχεδόν ποτέ δεν επιτυγχάνεται στην πράξη) θα υποδηλώνει τέλεια προσαρμογή στα δεδομένα. Γενικά, ένα χαμηλότερο RMSE είναι καλύτερο από ένα υψηλότερο. Ωστόσο, οι συγκρίσεις μεταξύ διαφορετικών τύπων δεδομένων δεν θα ήταν έγκυρες επειδή το μέτρο εξαρτάται από την κλίμακα των αριθμών που χρησιμοποιούνται.

Το RMSE είναι η τετραγωνική ρίζα του μέσου όρου των τετραγωνικών σφαλμάτων. Η επίδραση κάθε σφάλματος στο RMSE είναι ανάλογη με το μέγεθος του τετραγώνου του σφάλματος. Έτσι τα μεγαλύτερα σφάλματα έχουν δυσανάλογα μεγάλη επίδραση στο RMSE. Κατά συνέπεια, το RMSE είναι ευαίσθητο σε ακραίες τιμές. Υπολογίζεται ως εξής:

$$\text{RMSE}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} = \sqrt{\text{E}((\hat{\theta} - \theta)^2)}.$$

Όπου

- θ' η αληθής τιμή,
- θ η τιμή της πρόβλεψης

9.3.1.2. MAE

Το μέσο απόλυτο σφάλμα (MAE) αντιπροσωπεύει τον μέσο όρο της απόλυτης διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τον μέσο όρο των υπολειμμάτων στο σύνολο δεδομένων. Υπολογίζεται ως εξής:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}.$$

Όπου

- x η αληθής τιμή,
- y η τιμή της πρόβλεψης και
- n ο αριθμός των παρατηρήσεων

Είναι ένα μέτρο των σφαλμάτων μεταξύ των παρατηρήσεων που εκφράζουν το ίδιο φαινόμενο. Το μέσο απόλυτο σφάλμα χρησιμοποιεί την ίδια κλίμακα με τα δεδομένα που μετρώνται. Αυτό είναι γνωστό ως μέτρο ακρίβειας που εξαρτάται από την κλίμακα και επομένως δεν μπορεί να χρησιμοποιηθεί για να γίνουν συγκρίσεις μεταξύ σειρών χρησιμοποιώντας διαφορετικές κλίμακες. Το μέσο απόλυτο σφάλμα είναι ένα κοινό μέτρο του σφάλματος πρόβλεψης στην ανάλυση χρονοσειρών.

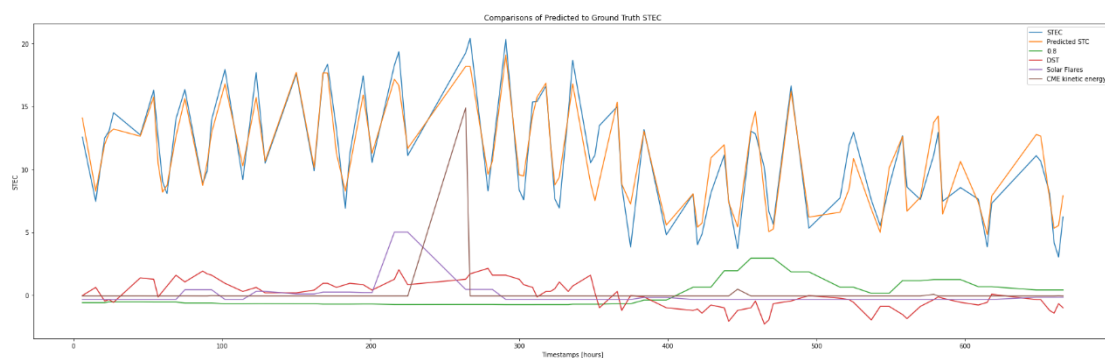
9.3.2. Αποτελέσματα

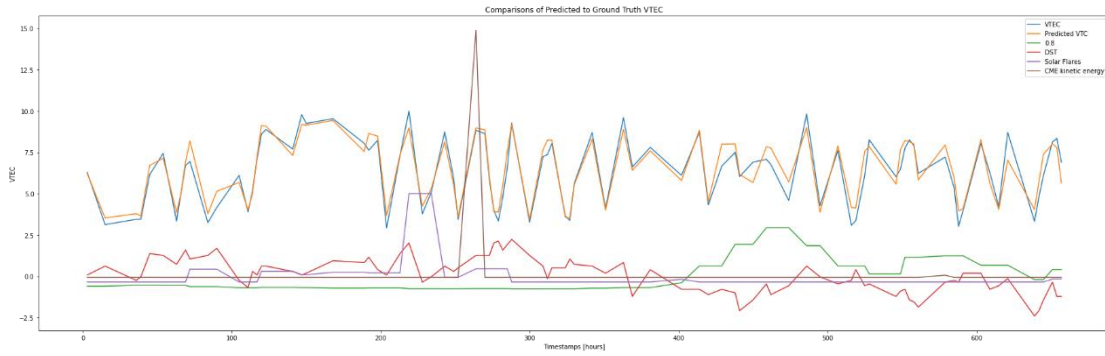
Κατά την παραγωγή των αποτελεσμάτων επιλέχθηκε η παρουσίαση τους σε γραφήματα για καλύτερη κατανόηση των αποτελεσμάτων και ευκολότερη εξαγωγή συμπερασμάτων. Στα γραφήματα όπως φαίνεται στις εικόνες 22-25 παρουσιάζονται: με μπλε χρώμα οι πραγματικές τιμές STEC και VTEC αντίστοιχα:

- με πορτοκαλί χρώμα οι προβλέψεις του παλινδρομητή XGBoost
- με πράσινο χρώμα οι τιμές GOES-15 Electron Fluence >0.8 MEV (Energetic Particle Flux)
- με κόκκινο χρώμα οι τιμές Dst
- με μωβ χρώμα οι τιμές των Solar Flares
- με καφέ χρώμα οι τιμές CMEs (κινητική ενέργεια)

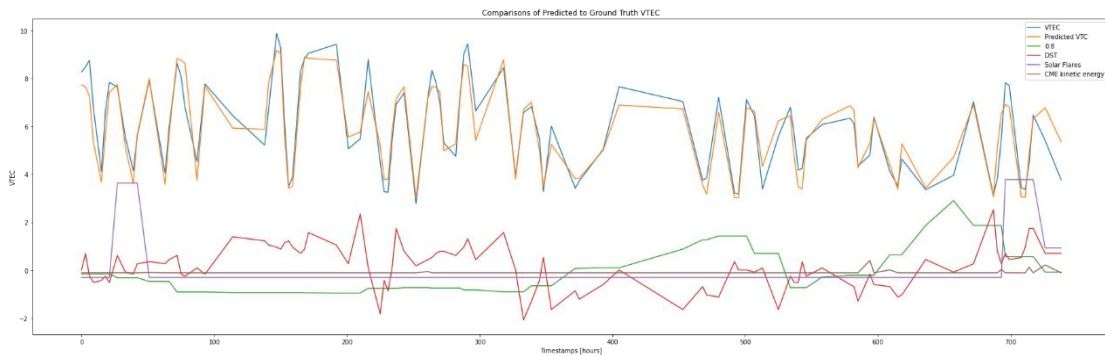
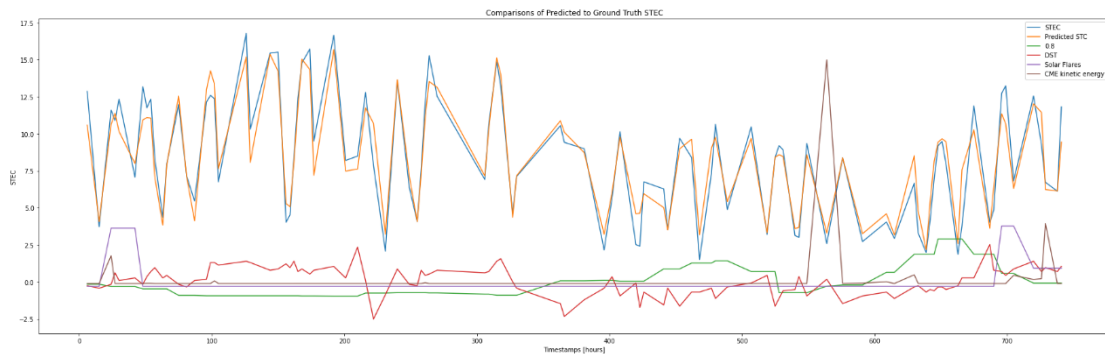


Εικόνα 21. Γραφήματα σύγκρισης αποτελεσμάτων Ιανουαρίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα

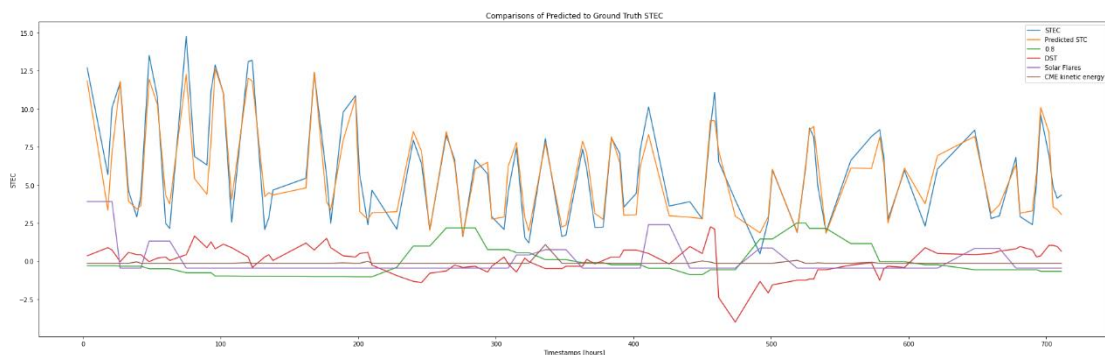


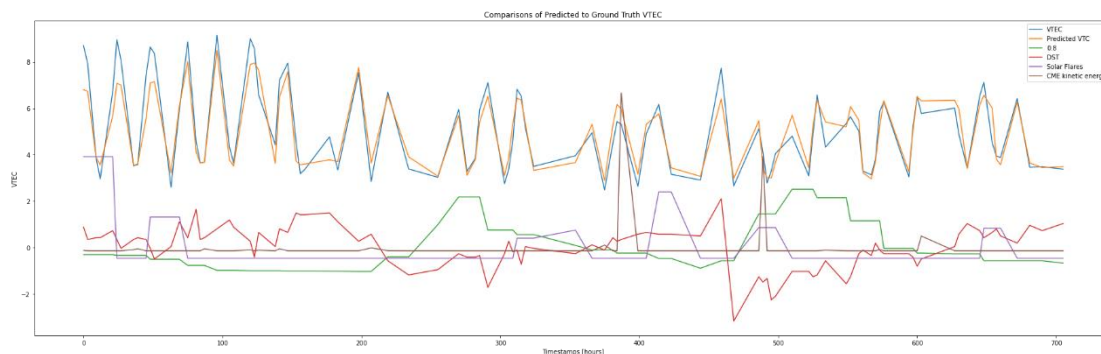


Εικόνα 22. Γραφήματα σύγκρισης αποτελεσμάτων Φεβρουαρίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα



Εικόνα 23. Γραφήματα σύγκρισης αποτελεσμάτων Μαρτίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα





Εικόνα 24. Γραφήματα σύγκρισης αποτελεσμάτων Απριλίου παλινδρομητή XGBoost για STEC και VTEC αντίστοιχα

9.3.3. Αξιολόγηση

Από τα παραπάνω γραφήματα και ύστερα από εκτενή μελέτη και της βιβλιογραφίας εξήχθησαν τα κάτωθι αποτελέσματα:

- Ο παλινδρομητής XGBoost είχε πολύ καλή προσαρμογή στα δεδομένα και οι προβλέψεις του ακολουθούσαν την πραγματικότητα εξαιρετικά.
- Αρχικά παρήχθησαν αποτελέσματα μόνο με τη χρήση των Solar Flares και των CMEs αλλά τα STEC και VTEC δεν είχαν σταθερές συμπεριφορές. Ως εκ τούτου και ύστερα από μελέτη άλλων εργασιών που έχουν εκπονηθεί από επιστήμονες ανά τον κόσμο [10][12][24][26][28][35][39][47][48] και υπό την επίβλεψη κέντρων που ασχολούνται ενδελεχώς με το αντικείμενο της μελέτης του διαστημικού καιρού, επιλέχθηκε η εμφάνιση και των τιμών των GOES-15 Electron Fluence >0.8 MEV και των Dst. Αυτό οφείλεται κυρίως στο γεγονός ότι δεν απαραίτητο ότι ένα Solar Flare ή ένα CME που θα παρατηρηθεί, να φτάσει στη Γη. Επίσης, δεν φτάνουν όλα στη Γη στον ίδιο χρόνο και ούτε έχουν ίδιες επιπτώσεις. Από την άλλη, οι τιμές των GOES-15 Electron Fluence >0.8 MEV και των Dst αφορούν εξ ολοκλήρου το γήινο μαγνητικό πεδίο.
- Από τις τιμές των GOES-15 Electron Fluence >0.8 MEV και των Dst φάνηκε επακριβώς πώς επηρεάζονται τα δεδομένα STEC και VTEC. Από μία απλή παρατήρηση, φαίνεται ότι τα δεδομένα STEC και VTEC ακολουθούν πλήρως τη συμπεριφορά της γραμμής των δεδομένων Dst. Τα δεδομένα Dst από την άλλη δείχνουν να επηρεάζονται πολύ από τα δεδομένα των τιμών των GOES-15 Electron Fluence >0.8 MEV και μία αύξηση τους προκαλεί μείωση των Dst και αντιστρόφως.

10. Συμπεράσματα

Με την εκπόνηση της παραπάνω διπλωματικής εργασίας, αντιμετωπίστηκε το πρόβλημα της δημιουργίας μοντέλων μηχανικής μάθησης για τη μελέτη της απόκρισης της ιονόσφαιρας σε έντονα ηλιακά φαινόμενα με χρήση παρατηρήσεων GNS. Το αποτέλεσμα της κρίνεται ως θετικό και σίγουρα θα μπορούσε να βοηθήσει στην πρόβλεψη έντονων ηλιακών φαινομένων. Παράλληλα φάνηκε σε όλα τα αποτελέσματα η επίδραση των έντονων ηλιακών φαινομένων στην ιονόσφαιρα.

Η παραπάνω διπλωματική εργασία απέδειξε πως είναι εφικτό για έναν αλγόριθμο να αξιολογεί «έξυπνα» διάφορες παραμέτρους, με την βοήθεια της βαθιάς μάθησης. Ο βέλτιστος αλγόριθμος με βάσει αυτούς που εξετάστηκαν ήταν ο XGBoost και αυτός χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων.

Ως γλώσσα προγραμματισμού για τη δημιουργία του προγράμματος RF_Solar επιλέχθηκε η Python καθώς προσφέρει δύο μεγάλα πλεονέκτημα στην αντιμετώπιση του παραπάνω προβλήματος. Το πρώτο πλεονέκτημα ότι είναι ένα ανοιχτού κώδικα γλώσσα με τρομερές δυνατότητες που διαθέτει χωρίς κόστος άπειρα εργαλεία και βιβλιοθήκες. Το δεύτερο πλεονέκτημα αφορά τα αναφερόμενα εργαλεία και οι βιβλιοθήκες pandas, shap και sklearn, προσφέρουν συναρτήσεις που κάνουν ειδικά τον προγραμματισμό δικτύων μηχανικής μάθησης εφικτό σε επαγγελματίες χρήστες και μη.

Τέλος, από την ολοκλήρωση της εργασίας, κρίνεται απαραίτητη η μελέτη ακόμα περισσότερων δεδομένων, πολύ περισσότερων ετών για την εξαγωγή αξιόπιστων αποτελεσμάτων ακόμα και σε ακραία φαινόμενα και τη ενδελεχή μελέτη της απόκρισης της ιονόσφαιρας σε αυτά.

Βιβλιογραφία

- [1] G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.
- [2] J. A. Anderson, An Introduction to Neural Networks, MIT Press, Cambridge, 1995.
- [3] S. Haykin, Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall, Upper Saddle Point, 1999.
- [4] N. K. Bose and P. Liang, Neural Networks Fundamentals with Graphs, Algorithms and Applications, McGraw-Hill, New York, 1996.
- [5] R. Lippmann, An Introduction to Computing with Neural Networks, IEEE ASSP Magazine, 4-22, 1987.
- [6] L. Breiman, —Random forests, Machine Learning, 45(1), pp. 5-32, 2001.
- [7] Gardner, M. ., & Dorling, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment, 32(14-15), 2627–2636. 1998
- [8] J. Friedman. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 2001.
- [9] Nishio M, Nishizawa M, Sugiyama O, Kojima R, Yakami M, Kuroda T, et al. Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization. 2018
- [10] J. Berdermann, M. Kriegel, D. Banyś, F. Heymann, M. M. Hoque, V. Wilken, C. Borries, A. Heßelbarth, N. Jakowski. Ionospheric Response to the X9.3 Flare on 6 September 2017 and Its Implication for Navigation Services Over Europe. 2018
- [11] Kaselimi M. Machine learning methods for modelling and analysis of time series signals in geoinformatics. 2021
- [12] Randa Natras, Michael Schmidt. Machine Learning Model Development for Space Weather Forecasting in the Ionosphere. 2021
- [13] Changqing Cheng, Akkarapol Sa-Ngasoongsong, Omer Beyca, Trung Le, Hui Yang, Zhenyu Kong, and Satish Bukkapatnam. Time series forecasting for nonlinear and nonstationary processes: A review and comparative study. IIE Transactions, 01 2015.
- [14] George C Tiao and George EP Box. Modeling multiple time series with applications. journal of the American Statistical Association, 76(376):802–816, 1981.
- [15] Lawrence C Hamilton, Cliff Brown, and Barry D Keim. Ski areas, weather and climate: time series models for new england case studies. International Journal of Climatology: A Journal of the Royal Meteorological Society, 27(15):2113–2124, 2007.
- [16] Giovanni Sparacino, Francesca Zanderigo, Stefano Corazza, Alberto Maran, Andrea Facchinetti, and Claudio Cobelli. Glucose concentration can be

- predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Transactions on biomedical engineering*, 54(5):931–937, 2007.
- [17] Nikolay Laptev, Saeed Amizadeh, and Ian Flint. Generic and scalable framework for automated time-series anomaly detection. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1939–1947, 2015.
- [18] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. Machine learning strategies for time series forecasting. In *European business intelligence summer school*, pages 62–77. Springer, 2012.
- [19] O. Mandrikova, Nadezhda Fetisova, Riad Al-Kasasbeh, D. Klionskiy, Vladimir Geppener, and Ilyash Maksim. Ionospheric parameter modelling and anomaly discovery by combining the wavelet transform with autoregressive models. *Annals of Geophysics*, 58, 11 2015.
- [20] D Venkata Ratnam, Nagireddy Vanga, P S Brahmanadam, B Sada Siva Rao, K Sarat Kumar, et al. Auto regressive ionospheric prediction model for gps applications. *International Journal of Computer Applications*, 48(4):7–9, 2012.
- [21] Yulong Kong, Hongzhou Chai, JinYao Li, ZongPeng Pan, and Yang Chong. A modified forecast method of ionosphere vtec series based on arma model. In *2017 Forum on Cooperative Positioning and Service (CPGPS)*, pages 90–95. IEEE, 2017.
- [22] D. Venkata Ratnam, Yuichi Otsuka, G. Sivavaraprasad, and J.R.K. Kumar Dabbakuti. Development of multivariate ionospheric tec forecasting algorithm using linear time series model and arma over low-latitude GNSS station. *Advances in Space Research*, 63(9):2848–2856, 2019. *Multi-GNSS: Methods, Benefits, Challenges, and Geosciences Applications*.
- [23]] John Bosco Habarulema, Lee-Anne McKinnell, Pierre J. Cilliers, and Ben D.L. Opperman. Application of neural networks to south african gps tec modelling. *Advances in Space Research*, 43(11):1711–1720, 2009. *Ionosphere - Modelling, Forecasting, and Telecommunications I*.
- [24] ATİLA Yılmaz, KE Akdogan, and M Gurun. Regional tec mapping using neural networks. *Radio Science*, 44(3), 2009.
- [25] Z. Huang and H. Yuan. Ionospheric single-station tec short-term forecast using rbf neural network. *Radio Science*, 49(4):283–292, 2014.
- [26] S. Wang, L. Liu, Liangke Huang, Y. Yang, and H. Peng. Performance evaluation of ionospheric tec forecasting models using gps observations at different latitudes. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII3/W10:1175–1182, 02 2020.
- [27] John A Klobuchar. Ionospheric time-delay algorithm for single-frequency gps users. *IEEE Transactions on aerospace and electronic systems*, (3):325–331, 1987.
- [28] Ersin Tulunay, Erdem Turker Senalp, Sandro Maria Radicella, and Yurdanur Tulunay. Forecasting total electron content maps by neural network technique. *Radio Science*, 41(04):1– 12, 2006.

- [29] Jyothi Ravi Kiran Kumar Dabbakuti and Bhavya Lahari G. Application of singular spectrum analysis using artificial neural networks in tec predictions for ionospheric space weather. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12):5101– 5107, 2019.
- [30] Ke Su, Shuanggen Jin, and M. M. Hoque. Evaluation of ionospheric delay effects on multi-GNSS positioning performance. *Remote Sensing*, 11(2), 2019.
- [31] Maria Kaselimi, Nikolaos Doulamis, and Demitris Delikaraoglou. Ionospheric modeling in GNSS positioning using deep learning models, 04 2019.
- [32] Rongxin Tang, Fantao Zeng, Zhou Chen, Jing-Song Wang, Chun-Ming Huang, and Zhiping Wu. The comparison of predicting storm-time ionospheric tec by three methods: Arima, LSTM, and seq2seq. *Atmosphere*, 11(4):316, 2020.
- [33] Maria Kaselimi, Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Demitris Delikaraoglou. Deep recurrent neural networks for ionospheric variations estimation using GNSS measurements. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–15, 2021.
- [34] P. T. Newell, J. W. Gjerloev. Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. 2011
- [35] Schmölder E., Berdermann J. Predicting the Effects of Solar Storms on the Ionosphere Based on a Comparison of Real-Time Solar Wind Data with the Best-Fitting. 2021
- [36] Milan, S.E.; Clausen, L.B.N.; Coxon, J.C.; Carter, J.A.; Walach, M.T.; Laundal, K.; Østgaard, N.; Tenfjord, P.; Reistad, J.; Snekvik, K.; et al. Overview of Solar Wind-Magnetosphere-Ionosphere-Atmosphere Coupling and the Generation of Magnetospheric Currents. *Space Sci. Rev.* **2017**, *206*, 547–573.
- [37] Borovsky, J.E.; Denton, M.H. Differences between CME-driven storms and CIR-driven storms. *J. Geophys. Res.* **2006**, *111*.
- [38] Yang, N.; Le, H.; Liu, L. Statistical analysis of the mid-latitude trough position during different categories of magnetic storms and different storm intensities. *Earth Planets Space* **2016**, *68*, 171–181.
- [39] Dugassa, T.; Habarulema, J.B.; Nigussie, M. Equatorial and low-latitude ionospheric TEC response to CIR-driven geomagnetic storms at different longitude sectors. *Adv. Space Res.* **2020**, *66*, 1947–1966.
- [40] Baker, D.N.; Daly, E.; Daglis, I.; Kappenman, J.G.; Panasyuk, M. Effects of Space Weather on Technology Infrastructure. *Space Weather* **2004**
- [41] Balan, N.; Skoug, R.; Ram, S.T.; Rajesh, P.K.; Shiokawa, K.; Otsuka, Y.; Batista, I.S.; Ebihara, Y.; Nakamura, T. CME front and severe space weather. *J. Geophys. Res. Space Phys.* **2014**, *119*, 10041–10058.
- [42] Berdermann, J.; Borries, C.; Hoque, M.M.; Jakowski, N. Forecast of Total Electron Content over Europe for Disturbed Ionospheric Conditions. 2021
- [43] Borries, C.; Berdermann, J.; Jakowski, N.; Hoque, M.; Bothmer, V. Preparation of an advanced TEC forecast based on the statistical analysis of

- historical ionospheric storms. In Proceedings of the 4th International Galileo Science Colloquium, Prague, Czech Republic, 4–6 December 2013
- [44] Schrijver, C.J.; Kauristie, K.; Aylward, A.D.; Denardini, C.M.; Gibson, S.E.; Glover, A.; Gopalswamy, N.; Grande, M.; Hapgood, M.; Heynderickx, D.; et al. Understanding space weather to shield society: A global road map for 2015–2025 commissioned by COSPAR and ILWS. *Adv. Space Res.* **2015**
- [45] Tsagouri, I.; Belehaki, A. Ionospheric forecasts for the European region for space weather applications. *J. Space Weather Space Clim.* **201**
- [46] Mallika, I. Lakshmi, Devanaboyina Venkata Ratnam, Yuichi Ostuka, Gampala Sivavaraprasad and Saravanan Raman. "Implementation of Hybrid Ionospheric TEC Forecasting Algorithm Using PCA-NN Method." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (2019): 371-381.
- [47] Nicola Linty, Alex Minetto, Fabio Dovis, Luca Spogli. Effects of Phase Scintillation on the GNSS Positioning Error During the September 2017 Storm at Svalbard. 2018
- [48] Anna Krypiak-Gregorczyk. Ionosphere response to three extreme events occurring near spring equinox in 2012, 2013 and 2015, observed by regional GNSS-TEC model. 2018
- [49] C. J. Hegarty and E. Chatre, "Evolution of the Global Navigation Satellite System (GNSS)," in *Proceedings of the IEEE*, vol. 96, no. 12, pp. 1902-1917, Dec. 2008, doi: 10.1109/JPROC.2008.2006090.
- [50] Calvert, J. G.. "Glossary of atmospheric chemistry terms (Recommendations 1990)" *Pure and Applied Chemistry*, vol. 62, no. 11, 1990, pp. 2167-2219
- [51] National Geographic Society. Atmosphere
- [52] I. I. Borzenkova and I. Ye. Turchinovich. Department of Climatology, State Hydrological Institute, Russia. History of Atmospheric Composition.
- [53] Schunk, R. W. The Terrestrial Ionosphere. 1983
- [54] H. Rishbeth. The ionospheric E-layer and F-layer dynamos — a tutorial review. *Journal of Atmospheric and Solar-Terrestrial Physics*. 1997
- [55] ScienceDirect. Total Electron Content. 2020
- [56] Bishop C. M.. Pattern Recognition and Machine Learning. 2006
- [57] T. M. Mitchell, "Machine learning," 1997.
- [58] Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet. Foundations of Machine Learning. 2012.
- [59] P. A. Schirmer, I. Mporas and M. Paraskevas, "Evaluation of Regression Algorithms and Features on the Energy Disaggregation Task," 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019
- [60] T. Doan and J. Kalita, "Selecting Machine Learning Algorithms Using Regression Models," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015
- [61] David A. Freedman. Statistical Models: Theory and Practice. 2009

- [62] R. Dennis Cook, Sanford Weisberg. Criticism and Influence Analysis in Regression, *Sociological Methodology*, Vol. 13. 1982
- [63] Ho TK. The Random Subspace Method for Constructing Decision Forests. 1998
- [64] G. Van Rossum, "Python Programming Language.," in USENIX annual technical conference, 2007, vol. 41, p. 36.
- [65] D. Kuhlman, A python book: Beginning python, advanced python, and python exercises. 2009.
- [66] T. E. Oliphant, A guide to NumPy, vol. 1. 2006.
- [67] G. Hackeling, Mastering Machine Learning with scikit-learn. 2017
- [68] Seemala, Gopi. (2012). GPS-TEC analysis software.

Παραρτήματα

1. Πίνακας Δεδομένων ανά μήνα

<https://www.dropbox.com/s/xxio754m78y9fzz/Indices.txt?dl=0>

2. Κώδικας RF_Solar

2.1. STEC

```
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
```

Pip install SHAP

```
!pip install shap
!pip install lime
!pip install --upgrade openpyxl
```

Importing Libraries

```
import pandas as pd
import numpy as np
import openpyxl
from sklearn.metrics import mean_squared_error
import sklearn
import shap
import statistics
from sklearn import ensemble
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.svm import SVR
from sklearn import linear_model
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import DotProduct, WhiteKernel
nel
from sklearn.metrics import mean_absolute_error
import xgboost
import lime
import lime.lime_tabular
import matplotlib.pyplot as plt
```

Reading and scaling the data

```
#reading the data
```

```
path = '/content/drive/MyDrive/Colab/jan.xlsx'

df = pd.read_excel(path)
print(df.head())
dates = df['date']
timestep = df['timestep']

stec = df['stec']
stec_std = stec.std()
stec_mean = stec.mean()

df_no_date = df.drop(['timestep', 'date', 'D', 'dpind'], axis=1)

print(df_no_date.head())
print(df_no_date.keys())
col_stec = df['stec']
max = col_stec.max()
min = col_stec.min()
std = df_no_date.std()
mean = df_no_date.mean()

dif = max - min

df_for_training = (df_no_date - mean) / std

df_for_training['dates'] = dates
df_for_training['timestep'] = timestep

print('df_for_training : ', df_for_training.head())

#trainX1 = df_no_date
trainX1 = df_for_training[[
    'time', 'a',
    'kp', 'ap', 'SML',
    'SMU',
    'fluxjulian', 'fluxcarrington', 'fluxobsflux',
    'fluxadjflux', 'fluxursi', 'dts',
    'dns',
    'dss', 'sddts', 'sddns',
    'sddss',
    'obsdts', 'obsdns', 'obsdss',
    'one',
    'ten', 'hundred', 'zeroeight',
    'two',
    'nm', 'sf', 'pa',
    'ang',
    'lin', 'height', 'speed',
    'acc',
```

```
        'mass',      'ke',      'mpa',
        'msd',      'ked']]
print('trainX1')
print(trainX1.head())

trainY1 = df_no_date[["stec"]]
print(trainY1.head())

#convert to np array
trainX, trainY = np.array(trainX1), np.array(trainY1)
print('trainX shape == {}'.format(trainX.shape))
print('trainY shape == {}'.format(trainY.shape))

#splitting the data
x_train, x_test, y_train, y_test = sklearn.model_selection.train_
test_split(trainX, trainY, test_size = 0.4, shuffle=True)
print('x_train shape : ', x_train.shape)
print('y_train shape : ', y_train.shape)

print('x_test shape : ', x_test.shape)
print('y_test shape : ', y_test.shape)
```

Other regressor approaches

```
# Plain Linear Regrassion
lr = LinearRegression().fit(x_train, y_train)

# RMSE - MAE : LR
preds_lr = lr.predict(x_test)

rmse_lr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
preds_lr*stec_std + stec_mean))
mae_lr = mean_absolute_error(y_test*stec_std + stec_mean, preds_l
r*stec_std + stec_mean)
print("Linear Regrassion: The root mean squared error (rMSE) on t
est set: {:.4f}".format(rmse_lr))
print("Linear Regrassion: The root mean squared error (rMSE) on t
est set: {:.4f}".format(mae_lr))

# Decision Tree
dt = DecisionTreeRegressor(max_depth=2)
dt.fit(x_train, y_train.ravel())

# RMSE - MAE : DT
preds_dt = dt.predict(x_test)

#preds_dt= preds_dt*dif + min
```

```
rmse_dt = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
    preds_dt*stec_std + stec_mean))
mae_dt = mean_absolute_error(y_test*stec_std + stec_mean, preds_d
t*stec_std + stec_mean)
print("Decision Tree Regressor: The root mean squared error (rMSE
) on test set: {:.4f}".format(rmse_dt))
print("Decision Tree Regressor: The root mean squared error (rMSE
) on test set: {:.4f}".format(mae_dt))

# Support Vector Regression
svr = SVR().fit(x_train, y_train.ravel())

# RMSE - SVR
preds_svr = svr.predict(x_test)

#preds_svr= preds_svr*dif + min

rmse_svr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_svr*stec_std + stec_mean))
mae_svr = mean_absolute_error(y_test*stec_std + stec_mean, preds_
svr*stec_std + stec_mean)
print("Support Vector Regression: The root mean squared error (rM
SE) on test set: {:.4f}".format(rmse_svr))
print("Support Vector Regression: The root mean squared error (rM
SE) on test set: {:.4f}".format(mae_svr))

# Lasso Regression
lassoReg = linear_model.Lasso(alpha=0.1)
lassoReg.fit(x_train, y_train.ravel())

# RMSE - Lasso
preds_lasso = lassoReg.predict(x_test)

#preds_lasso= preds_lasso*dif + min

rmse_lasso = np.sqrt(mean_squared_error(y_test*stec_std + stec_me
an, preds_lasso*stec_std + stec_mean))
mae_lasso = mean_absolute_error(y_test*stec_std + stec_mean, pred
s_lasso*stec_std + stec_mean)
print("Lasso Regression: The root mean squared error (rMSE) on te
st set: {:.4f}".format(rmse_lasso))
print("Lasso Regression: The root mean squared error (rMSE) on te
st set: {:.4f}".format(mae_lasso))

# Gaussian Process Regressor
kernel = DotProduct() + WhiteKernel()
gpr = GaussianProcessRegressor(kernel=kernel,random_state=0).fit(
x_train, y_train.ravel())
```

```
# RMSE - Gaussian Process Regressor
preds_gpr = gpr.predict(x_test)

#preds_gpr= preds_gpr*dif + min

rmse_gpr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_gpr*stec_std + stec_mean))
mae_gpr = mean_absolute_error(y_test*stec_std + stec_mean, preds_
gpr*stec_std + stec_mean)
print("Gaussian Process Regressor: The root mean squared error (r
MSE) on test set: {:.4f}".format(rmse_gpr))
print("Gaussian Process Regressor: The root mean squared error (r
MSE) on test set: {:.4f}".format(mae_gpr))

# Random Forest Regressor:
rfregressor = ensemble.RandomForestRegressor()
rfregressor.fit(x_train, y_train.ravel())

# RMSE - Random Forest Regressor
preds_rfg = rfregressor.predict(x_test)

rmse_rfg = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_rfg*stec_std + stec_mean))
mae_rfg = mean_absolute_error(y_test*stec_std + stec_mean, preds_
rfg*stec_std + stec_mean)
print("Random Forest Regressor:The root mean squared error (rMSE)
on test set: {:.4f}".format(rmse_rfg))
print("Random Forest Regressor:The root mean squared error (rMSE)
on test set: {:.4f}".format(mae_rfg))

XGBR = xgboost.XGBRegressor().fit(x_train, y_train)

# RMSE XGBoost Regressor

preds_xgb = XGBR.predict(x_test)

rmse_xgb = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_xgb*stec_std + stec_mean))
mae_xgb = mean_absolute_error(y_test*stec_std + stec_mean, preds_
xgb*stec_std + stec_mean)
print("XGBoost Regressor: The root mean squared error (rMSE) on t
est set: {:.4f}".format(rmse_xgb))
print("XGBoost Regressor: The root mean squared error (rMSE) on t
est set: {:.4f}".format(mae_xgb))

#min-max dif
yg = y_test*stec_std + stec_mean
```

```
yh = preds_xgb*stec_std + stec_mean  
diff = np.absolute(yg - yh)
```

```
MIN = np.amin(diff)  
MAX = np.amax(diff)  
print( "RF MIN : ", MIN)  
print( "RF MAX : ", MAX)
```

Table of comparisons

```
plt.rcParams["figure.figsize"] = [10.00, 5.50]  
plt.rcParams["figure.autolayout"] = True  
fig, ax =plt.subplots(1,1)  
data=[[rmse_lr,mae_lr],  
      [rmse_dt,mae_dt],  
      [rmse_lasso,mae_lasso],  
      [rmse_gpr,mae_gpr],  
      [rmse_rfg,mae_rfg],  
      [rmse_xgb,mae_xgb]  
      ]  
column_labels=["rMSE", "MAE"]  
row=["Linear Regression ", "Decision Tree", "Support Vector Regress  
ion", "Gaussian Process Regressor", "Random Forest Regressor", "XGB  
OOST Regressor"]  
df=pd.DataFrame(data,columns=column_labels)  
df.update(df.applymap('{:,.4f}'.format))  
ax.axis('tight')  
ax.axis('off')  
the_table = ax.table(cellText=df.values,colLabels=df.columns,rowL  
abels=row,fontsize=15,loc="center",cellLoc='center')  
the_table.auto_set_font_size(False)  
the_table.set_fontsize(14)  
plt.show()
```

Creating Graphs

```
regressor = ensemble.RandomForestRegressor()  
regressor.fit(x_train, y_train.ravel())  
  
test_x = pd.DataFrame(x_test, columns = trainX1.keys())  
test_y = pd.DataFrame(y_test, columns = trainY1.keys())  
test = pd.concat([test_x, test_y], axis=1)  
  
test_with_dates = pd.merge(test, df_for_training, how="inner", on  
=['time','a', 'kp', 'ap', 'SML', 'SMU',  
  'fluxjulian', 'fluxcarrington', 'fluxobsflux', 'fluxadjflu  
x',
```



```
        'fluxursi', 'dts', 'dns', 'dss', 'sddts', 'sddns', 'sdds'
, 'obsdts',
        'obsdns', 'obsdss', 'one', 'ten', 'hundred', 'zeroeight',
'two', 'nm',
        'sf', 'pa', 'ang', 'lin', 'height', 'speed', 'acc', 'mass'
, 'ke', 'mpa', 'msd', 'ked'])

test_with_dates.drop(['vtec', 'stec_y'], inplace=True, axis=1)

test_with_dates = test_with_dates.sort_values(by=['timestep'])

print(test_with_dates.head())
print(test_with_dates.keys())

XTest = test_with_dates[[
        'time', 'a',
        'kp', 'ap', 'SML',
        'SMU',
        'fluxjulian', 'fluxcarrington', 'fluxobsflux'
,
        'fluxadjflux', 'fluxursi', 'dts',
        'dns',
        'dss', 'sddts', 'sddns',
        'sdds',
        'obsdts', 'obsdns', 'obsdss',
        'one',
        'ten', 'hundred', 'zeroeight',
        'two',
        'nm', 'sf', 'pa',
        'ang',
        'lin', 'height', 'speed',
        'acc', 'mass', 'ke', 'mpa', 'msd', 'ked']]

YTest = test_with_dates[['stec_x']]

#convert to np array
XTest, YTest = np.array(XTest), np.array(YTest)
print('XTest shape == {}'.format(XTest.shape))
print('YTest shape == {}'.format(YTest.shape))

#RMSE
preds = regressor.predict(XTest)
preds_denorm = preds*stec_std + stec_mean
YTest_denorm = YTest*stec_std + stec_mean

#Create Graphs
plt.figure(figsize=(25, 8))
plt.plot(test_with_dates['timestep'], YTest, label = 'STEC')
```

```
plt.plot(test_with_dates['timestep'],preds, label = 'Predicted ST  
C')  
plt.plot(test_with_dates['timestep'],test_with_dates['zeroeight']  
, label = '0.8')  
plt.plot(test_with_dates['timestep'],test_with_dates['a'], label  
= 'DST')  
plt.plot(test_with_dates['timestep'],test_with_dates['sf'], label  
= 'Solar Flares')  
plt.plot(test_with_dates['timestep'],test_with_dates['ke'], label  
= 'CME kinetic energy')  
plt.legend(loc='upper right')  
plt.title('Comparisons of Predicted to Ground Truth STEC')  
plt.xlabel('Timestamps [hours]')  
plt.ylabel('STEC')  
plt.show()
```

2.2. VTEC

```
#reading the data  
path = '/content/drive/MyDrive/Colab/jan.xlsx'  
  
df = pd.read_excel(path)  
print(df.head())  
dates = df['date']  
timestep = df['timestep']  
  
stec = df['vtec']  
stec_std = stec.std()  
stec_mean = stec.mean()  
  
df_no_date = df.drop(['timestep','date','D','dpind'], axis=1)  
  
print(df_no_date.head())  
print(df_no_date.keys())  
col_stec = df['vtec']  
max = col_stec.max()  
min = col_stec.min()  
std = df_no_date.std()  
mean = df_no_date.mean()  
  
dif = max - min  
  
df_for_training = (df_no_date-mean)/std  
  
df_for_training['dates'] = dates  
df_for_training['timestep'] = timestep
```

```
print('df_for_training : ', df_for_training.head())

#trainX1 = df_no_date
trainX1 = df_for_training[[
    'time', 'a',
    'kp', 'ap', 'SML',
    'SMU',
    'fluxjulian', 'fluxcarrington', 'fluxobsflux',
    'fluxadjflux', 'fluxursi', 'dts',
    'dns',
    'dss', 'sddts', 'sddns',
    'sddss',
    'obsdts', 'obsdns', 'obsdss',
    'one',
    'ten', 'hundred', 'zeroeight',
    'two',
    'nm', 'sf', 'pa',
    'ang',
    'lin', 'height', 'speed',
    'acc',
    'mass', 'ke', 'mpa',
    'msd', 'ked']]
print('trainX1')
print(trainX1.head())

trainY1 = df_no_date[["vtec"]]
print('trainY1.head()')
print(trainY1.head())

#convert to np array
trainX, trainY = np.array(trainX1), np.array(trainY1)
print('trainX shape == {}'.format(trainX.shape))
print('trainY shape == {}'.format(trainY.shape))

#splitting the data
x_train, x_test, y_train, y_test = sklearn.model_selection.train_
test_split(trainX, trainY, test_size = 0.4, shuffle=True)
print('x_train shape : ', x_train.shape)
print('y_train shape : ', y_train.shape)

print('x_test shape : ', x_test.shape)
print('y_test shape : ', y_test.shape)

# Plain Linear Regrassion
lr = LinearRegression().fit(x_train, y_train)

# RMSE - MAE : LR
preds_lr = lr.predict(x_test)
```

```
rmse_lr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
    preds_lr*stec_std + stec_mean))
mae_lr = mean_absolute_error(y_test*stec_std + stec_mean, preds_l
r*stec_std + stec_mean)
print("Linear Regrassion: The root mean squared error (rMSE) on t
est set: {:.4f}".format(rmse_lr))
print("Linear Regrassion: The root mean squared error (rMSE) on t
est set: {:.4f}".format(mae_lr))

# Decision Tree
dt = DecisionTreeRegressor(max_depth=2)
dt.fit(x_train, y_train.ravel())

# RMSE - MAE : DT
preds_dt = dt.predict(x_test)

#preds_dt= preds_dt*dif + min

rmse_dt = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
    preds_dt*stec_std + stec_mean))
mae_dt = mean_absolute_error(y_test*stec_std + stec_mean, preds_d
t*stec_std + stec_mean)
print("Decision Tree Regressor: The root mean squared error (rMSE
) on test set: {:.4f}".format(rmse_dt))
print("Decision Tree Regressor: The root mean squared error (rMSE
) on test set: {:.4f}".format(mae_dt))

# Support Vector Regression
svr = SVR().fit(x_train, y_train.ravel())

# RMSE - SVR
preds_svr = svr.predict(x_test)

#preds_svr= preds_svr*dif + min

rmse_svr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_svr*stec_std + stec_mean))
mae_svr = mean_absolute_error(y_test*stec_std + stec_mean, preds_
svr*stec_std + stec_mean)
print("Support Vector Regression: The root mean squared error (rM
SE) on test set: {:.4f}".format(rmse_svr))
print("Support Vector Regression: The root mean squared error (rM
SE) on test set: {:.4f}".format(mae_svr))

# Lasso Regression
lassoReg = linear_model.Lasso(alpha=0.1)
lassoReg.fit(x_train, y_train.ravel())
```

```
# RMSE - Lasso
preds_lasso = lassoReg.predict(x_test)

#preds_lasso= preds_lasso*dif + min

rmse_lasso = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
preds_lasso*stec_std + stec_mean))
mae_lasso = mean_absolute_error(y_test*stec_std + stec_mean, preds_lasso*stec_std + stec_mean)
print("Lasso Regression: The root mean squared error (rMSE) on test set: {:.4f}".format(rmse_lasso))
print("Lasso Regression: The root mean squared error (rMSE) on test set: {:.4f}".format(mae_lasso))

# Gaussian Process Regressor
kernel = DotProduct() + WhiteKernel()
gpr = GaussianProcessRegressor(kernel=kernel,random_state=0).fit(x_train, y_train.ravel())

# RMSE - Gaussian Process Regressor
preds_gpr = gpr.predict(x_test)

#preds_gpr= preds_gpr*dif + min

rmse_gpr = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
preds_gpr*stec_std + stec_mean))
mae_gpr = mean_absolute_error(y_test*stec_std + stec_mean, preds_gpr*stec_std + stec_mean)
print("Gaussian Process Regressor: The root mean squared error (rMSE) on test set: {:.4f}".format(rmse_gpr))
print("Gaussian Process Regressor: The root mean squared error (rMSE) on test set: {:.4f}".format(mae_gpr))

# MODEL
regressor = ensemble.RandomForestRegressor()
regressor.fit(x_train, y_train.ravel())

# RMSE
preds_rfg = regressor.predict(x_test)

rmse_rfg = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean,
preds_rfg*stec_std + stec_mean))
mae_rfg = mean_absolute_error(y_test*stec_std + stec_mean, preds_rfg*stec_std + stec_mean)
print("The root mean squared error (rMSE) on test set: {:.4f}".format(rmse_rfg))
```

```
print("The root mean squared error (rMSE) on test set: {:.4f}".fo
rmat(mae_rfg))

XGBR = xgboost.XGBRegressor().fit(x_train, y_train)

# RMSE

preds_xgb = XGBR.predict(x_test)

rmse_xgb = np.sqrt(mean_squared_error(y_test*stec_std + stec_mean
, preds_xgb*stec_std + stec_mean))
mae_xgb = mean_absolute_error(y_test*stec_std + stec_mean, preds_
xgb*stec_std + stec_mean)
print("The root mean squared error (rMSE) on test set: {:.4f}".fo
rmat(rmse_xgb))
print("The root mean squared error (rMSE) on test set: {:.4f}".fo
rmat(mae_xgb))

#min-max dif
yg = y_test*stec_std + stec_mean
yh = preds_xgb*stec_std + stec_mean
diff = np.absolute(yg - yh)

MIN = np.amin(diff)
MAX = np.amax(diff)
print("RF MIN : ", MIN)
print("RF MAX : ", MAX)

plt.rcParams["figure.figsize"] = [10.00, 5.50]
plt.rcParams["figure.autolayout"] = True
fig, ax =plt.subplots(1,1)
data=[[rmse_lr,mae_lr],
      [rmse_dt,mae_dt],
      [rmse_lasso,mae_lasso],
      [rmse_gpr,mae_gpr],
      [rmse_rfg,mae_rfg],
      [rmse_xgb,mae_xgb]
      ]
column_labels=["rMSE", "MAE"]
row=["Linear Regression ", "Decision Tree", "Support Vector Regress
ion", "Gaussian Process Regressor", "Random Forest Regressor", "XGB
OOST Regressor"]
df=pd.DataFrame(data,columns=column_labels)
df.update(df.applymap('{:,.4f}'.format))
ax.axis('tight')
ax.axis('off')
the_table = ax.table(cellText=df.values,colLabels=df.columns,rowL
abels=row,fontsize=15,loc="center",cellLoc='center')
```

```
the_table.auto_set_font_size(False)
the_table.set_fontsize(14)
plt.show()

test_x = pd.DataFrame(x_test, columns = trainX1.keys())
test_y = pd.DataFrame(y_test, columns = trainY1.keys())
test = pd.concat([test_x, test_y], axis=1)

test_with_dates = pd.merge(test, df_for_training, how="inner", on
=['time', 'a', 'kp', 'ap', 'SML', 'SMU',
  'fluxjulian', 'fluxcarrington', 'fluxobsflux', 'fluxadjflux',
  'fluxursi', 'dts', 'dns', 'dss', 'sddts', 'sddns', 'sddss',
  'obsdts',
  'obsdns', 'obsdss', 'one', 'ten', 'hundred', 'zeroeight',
  'two', 'nm',
  'sf', 'pa', 'ang', 'lin', 'height', 'speed', 'acc', 'mass',
  'ke', 'mpa', 'msd', 'ked'])

test_with_dates.drop(['stec', 'vtec_y'], inplace=True, axis=1)

test_with_dates = test_with_dates.sort_values(by=['timestep'])

print(test_with_dates.head())
print(test_with_dates.keys())

XTest = test_with_dates[['time', 'a',
  'kp', 'ap', 'SML',
  'SMU',
  'fluxjulian', 'fluxcarrington', 'fluxobsflux',
  'fluxadjflux', 'fluxursi', 'dts',
  'dns',
  'dss', 'sddts', 'sddns',
  'sddss',
  'obsdts', 'obsdns', 'obsdss',
  'one',
  'ten', 'hundred', 'zeroeight',
  'two',
  'nm', 'sf', 'pa',
  'ang',
  'lin', 'height', 'speed',
  'acc', 'mass', 'ke', 'mpa', 'msd', 'ked']]

YTest = test_with_dates[['vtec_x']]

#convert to np array
XTest, YTest = np.array(XTest), np.array(YTest)
```

```
print('XTest shape == {}'.format(XTest.shape))
print('YTest shape == {}'.format(YTest.shape))

#RMSE
preds = regressor.predict(XTest)
preds_denorm = preds*stec_std + stec_mean
YTest_denorm = YTest*stec_std + stec_mean

#Create Graphs
plt.figure(figsize=(25, 8))
plt.plot(test_with_dates['timestep'],YTest, label = 'VTEC')
plt.plot(test_with_dates['timestep'],preds, label = 'Predicted VT
C')
plt.plot(test_with_dates['timestep'],test_with_dates['zeroeight']
, label = '0.8')
plt.plot(test_with_dates['timestep'],test_with_dates['a'], label
= 'DST')
plt.plot(test_with_dates['timestep'],test_with_dates['sf'], label
= 'Solar Flares')
plt.plot(test_with_dates['timestep'],test_with_dates['ke'], label
= 'CME kinetic energy')
plt.legend(loc='upper right')
plt.title('Comparisons of Predicted to Ground Truth VTEC')
plt.xlabel('Timestamps [hours]')
plt.ylabel('VTEC')
plt.show()
```