

**«Εξηγήσιμη Τεχνητή Νοημοσύνη για την  
ερμηνεία μοντέλων μηχανικής πρόβλεψης  
ιονοσφαιρικής δραστηριότητας»**

**Σακελλαρίου Δημήτριος**

**Διπλωματική Εργασία**

**Διεπιστημινικό Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών  
Γεωπληροφορικής**

**Σχολή Αγρονόμων και Τοπογράφων Μηχανικών – Μηχανικών  
Γεωπληροφορικής**

**Εθνικό Μετσόβιο Πολυτεχνείο**



**School of Rural, Surveying and Geoinformatics Engineering**

**National Technical University of Athens**

**-Αθήνα, 21 Ιουνίου 2024-**

**Τριμελής επιτροπή**

**Δουλάμης Νικόλαος  
Καθηγητής ΕΜΠ**

**Τσακίρη Μαρία  
Καθηγήτρια ΕΜΠ**

**Δουλάμης Αναστάσιος  
Καθηγητής ΕΜΠ**

---

---

## Ευχαριστίες

---

---

Με την συγγραφή της παρούσας διπλωματικής εργασίας, ολοκληρώνεται ένας διετής κύκλος μεταπτυχιακών σπουδών στην Σχολή Αγρονόμων και Τοπογράφων Μηχανικών - Μηχανικών Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου. Ο κύκλος αυτός, ήταν καθοριστικός για εμένα, γνώρισα και συνεργάστηκα με εξαιρετικούς καθηγητές και συμφοιτητές, ενώ δέχθηκα την στήριξη πολλών ατόμων τα οποία οφείλω να ευχαριστήσω.

Πρωτίστως, ένα μεγάλο ευχαριστώ αξίζει στον Αναπληρωτή Καθηγητή Δουλάμη Νικόλαο, για την εμπιστοσύνη του από την πρώτη στιγμή που ανέλαβα την παρούσα διπλωματική εργασία, την καθοδήγησή του και τις συμβουλές του για την συγγραφή της. Ένα μεγάλο ευχαριστώ θα πάει επίσης και στους συναδέλφους και συμφοιτητές Τσουμένη Εμμανουήλ και Σπανάκη Μάριο γιατί ήμασταν πάντα ο ένας για τον άλλο από την πρώτη ημέρα που παρουσιάστηκαμε στη ΓΥΣ, στη φοίτηση μας ως προπτυχιακοί φοιτητές στη Σχολή Αγρονόμων και Τοπογράφων Μηχανικών - Μηχανικών Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου. μέχρι και την παρουσίαση των διπλωματικών μας εργασιών ως μεταπτυχιακοί φοιτητές του ΔΠΜΣ Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου.

Ένα μεγάλο ευχαριστώ αφιερώνεται στην οικογένειά μου, σε φίλους και συγγενείς. Στη σύζυγο μου, Χρυσάνθη, και στα δύο μου παιδιά Αθανασία και Μιχαήλ, για την αδιάκοπη στήριξή τους, χάρη στους οποίους κατάφερα να ολοκληρώσω έναν πολυετή κύκλο σπουδών, αρχομένης από την επιτυχή φοίτηση στη Στρατιωτική Σχολή Ευελπίδων, τη Σχολή Πεζικού και τη Σχολή Τοπογραφίας της Γεωγραφικής Υπηρεσίας Στρατού και την αποφοίτησή μου από τη Σχολή Αγρονόμων και Τοπογράφων Μηχανικών - Μηχανικών Γεωπληροφορικής του Εθνικού Μετσόβιου Πολυτεχνείου.

Τέλος, θα ήθελα να ευχαριστήσω και τον υποψήφιο διδάκτορα Αναστάσιο Τέμμενο για την άψογη συνεργασία και τη συνεχή καθοδήγηση του, για την εκπόνηση αυτής της εργασίας.

---

---

## Περίληψη

---

---

Η αναγκαιότητα πρόβλεψης της ιονοσφαιρικής δραστηριότητας αποτελεί μείζον θέμα για τις μετρήσεις προσδιορισμού θέσης. Η καθυστέρηση της ιονόσφαιρας είναι ανάλογη με το συνολικό αριθμό ηλεκτρονίων της ιονόσφαιρας (**TEC**) κατά μήκος της διαδρομής διάδοσης, και αντιστρόφως ανάλογη του τετραγώνου της συχνότητας  $f$  του φορέα (L1 ή L2). Οι τιμές TEC (Total Electron Content) είναι απαραίτητες για τη διόρθωση της διάθλασης της ιονόσφαιρας και είναι ζωτικής σημασίας για τα συστήματα δορυφορικής πλοήγησης, προκειμένου να διασφαλιστεί η υψηλή απόδοση των δορυφορικών συστημάτων στον εντοπισμό θέσης.

Με το παρόν, προτείνουμε μια προσέγγιση βασισμένη στην εξηγήσιμη τεχνητή νοημοσύνη για μοντελοποίηση και πρόβλεψη της ιονόσφαιρας. Αυτή η μέθοδος εκμεταλλεύεται τα πλεονεκτήματα της χρήσης απλών μοντέλων που είναι εύκολο να ερμηνευτούν από τον άνθρωπο, την ανάπτυξη εργαλείων οπτικοποίησης για την αναπαράσταση των δεδομένων και των αποφάσεων του συστήματος με ευανάγνωστο τρόπο, και την ενίσχυση της διαφάνειας στον τρόπο λειτουργίας των αλγορίθμων. Στα πλαίσια της εργασίας χρησιμοποιήθηκαν εννιά (9) μοντέλα – τεχνικές εξηγήσιμης μάθησης με σκοπό να προσδιοριστεί καλύτερα οι τιμές κατακόρυφης συνολικής περιεκτικότητας ηλεκτρονίων (VTEC) και λοξής (πλάγιας) συνολικής περιεκτικότητας ηλεκτρονίων (STEC) της ιονόσφαιρας.

---

---

## Λέξεις κλειδιά

---

---

Παγκόσμιο Δορυφορικό Σύστημα Πλοήγησης, ιονοσφαιρική καθυστέρηση, μοντέλο SHAP, Συνολικός αριθμός ηλεκτρονίων ιονόσφαιρας.

---

---

## Abstract

---

---

The need to predict ionospheric activity is a major issue for positioning measurements. The ionosphere delay is proportional to the number of ionosphere electrons (TEC) along the propagation path, and inversely proportional to the square of the carrier frequency  $f$  (L1 or L2). TEC values are necessary for the correction of ionosphere refraction and are vital for satellite navigation systems in order to ensure the high performance of satellite systems at the location.

Hereby, we propose an explainable artificial intelligence-based approach for ionospheric modeling and forecasting. This method takes advantage of using simple models that are easy for humans to interpret, developing visualization tools to represent data and system decisions in an easy-to-read manner, and enhancing transparency in how algorithms work. As part of the work, nine (9) models - explanatory learning techniques were used in order to better determine the values of vertical total electron content (VTEC) and oblique (lateral) total electron content (STEC) of the ionosphere

---

---

## Keywords

---

---

GNSS, IONOSPHERIC DELAY, SHAP, STEC, VTEC



---



---

## Περιεχόμενα

---



---

Ευχαριστίες.....	1
Περίληψη.....	2
Λέξεις κλειδιά.....	2
Abstract .....	3
Keywords .....	3
Κατάλογος συντομογραφιών .....	7
Κατάλογος εικόνων .....	8
Κατάλογος πινάκων.....	9
Κεφάλαιο 1 <sup>ο</sup> Εισαγωγή.....	10
1.1 Η σημασία της πρόβλεψης της ιονοσφαιρικής δραστηριότητας .....	10
1.2 Κύριες προκλήσεις .....	11
1.2.1 Η πρόκληση της μοντελοποίησης TEC με χρήση δεδομένων GNSS .....	12
1.3 Συνεισφορά .....	13
1.4 Δομή εργασίας .....	14
Κεφάλαιο 2 <sup>ο</sup> Βιβλιογραφική ανασκόπηση.....	15
2.1 Διαστημικά Καιρικά Φαινόμενα .....	15
2.2 Νευρώνες μακράς-βραχείας διάρκειας μνήμης (Long Short-Term Memory - LSTM) ..	16
2.3 Principal component analysis (PCA) .....	16
2.4 OKSM .....	18
2.5 Τεχνικές Μοντελοποίησης DIAS .....	19
2.6 Τεχνική Support Vector Machine (SVM) .....	20
Κεφάλαιο 3 <sup>ο</sup> Σύστημα GNSS και Ατμόσφαιρα .....	21
3.1 Βασικές αρχές συστήματος GNSS .....	21
3.2 Η ατμόσφαιρα της Γης.....	22
3.2.1 Εισαγωγικές Έννοιες.....	22
3.2.2 Τα στρώματα της ατμόσφαιρας.....	22
3.3 Η Ιονόσφαιρα και τα στρώματά της.....	25
3.3.1 Εισαγωγικές Έννοιες.....	25
3.3.2 Τα στρώματα της ιονόσφαιρας και τα χαρακτηριστικά τους .....	25
3.4 Οι τιμές TEC (Total Electron Content) .....	28
3.4.1 Εισαγωγικές έννοιες.....	28
3.4.2 Βασικές αρχές.....	29
3.5 Παράμετροι επιρροής τιμών Tec .....	31

3.5.1 Εισαγωγικές έννοιες.....	31
3.5.2 Μαγνητική κ Ιονοσφαιρική Καταιγίδα.....	33
3.5.3 Ηλιακές Εκλάμψεις (Solar Flares).....	34
3.5.4 Coronal mass ejection.....	35
3.5.5 Solar Radio Flux.....	36
3.5.6 Sunspot Number.....	37
3.5.7 Ενεργητικές ροές ηλεκτρονίων (Energetic electrons fluxes).....	38
3.5.8 Γεωμαγνητικοί δείκτες K.....	39
3.5.9 Auroral Electrojet (AE).....	39
3.5.10 Disturbance storm time (Dst).....	40
Κεφάλαιο 4 <sup>ο</sup> Βασικές έννοιες Τεχνητής Νοημοσύνης.....	41
4.1 Εξηγήσιμη Τεχνητή Νοημοσύνη.....	42
4.2 Μηχανική μάθηση (Machine Learning).....	43
4.3 SHAP (SHapley Additive exPlanations).....	45
4.3.1 Αναλυτικότερα.....	45
4.3.2 SHAP Feature Importance.....	46
4.3.3 SHAP Summary Plot.....	48
4.3.4 SHAP Interaction Values.....	49
4.3.5 Πλεονεκτήματα.....	50
4.3.6 Μειονεκτήματα.....	51
4.4 Πρόβλεψη και Νευρωνικά δίκτυα (Neural Networks).....	51
4.4.1 Συνάρτηση Απώλειας (Loss Function).....	52
4.5 Linear Regrassion (Γραμμική Παλινδρόμηση).....	53
4.6 Decision Tree (Δέντρο Απόφασης).....	54
4.7 Support Vector Regression (Υποστηρικτική Διανυσματική Παλινδρόμηση).....	55
4.8 Gaussian Process Regressor (Γκαουσιανός Κατακερματισμός).....	56
4.9 XGBOOST Regressor.....	56
4.10 Multi-layer Perceptron Regressor (Πολύ-επίπεδος Επικοινωνιακός Ανιχνευτής).....	57
4.11 LightGBM Regressor.....	58
4.12 Random Forrest Regressor.....	58
Κεφάλαιο 5 <sup>ο</sup> Η γλώσσα προγραμματισμού Python.....	60
5.1 Γενικά για την γλώσσα προγραμματισμού Python.....	60
5.2 Η συμβολή της γλώσσας Python στην ανάπτυξη του δικτύου: Google Colaborratory.....	61
5.3 Σημαντικές βιβλιοθήκες Python που χρησιμοποιήθηκαν.....	61
5.3.1 Scikit Learn.....	62

5.3.2 Pandas .....	62
5.3.3 NumPy .....	63
5.3.4 Matplotlib .....	63
5.4 Υλοποίηση σε Python .....	63
Κεφάλαιο 6 <sup>ο</sup> Πειραματική διαχείριση – Περιγραφή και επεξεργασία αρχικών δεδομένων .	64
6.1 Εισαγωγικές έννοιες .....	64
6.2 Περιγραφή των δεδομένων .....	64
6.2.1 Observation Files του σταθμού COTE00ATA (Ανταρκτική) .....	64
6.2.2 Παράμετροι επιρροής τιμών Tec .....	70
6.3 Επεξεργασία των δεδομένων .....	78
Κεφάλαιο 7 <sup>ο</sup> Χρήση παραμέτρων δεδομένων TEC – Σενάρια και αξιολόγηση αυτών .....	80
7.1 Δορυφορικός Σταθμός Ανταρκτικής.....	81
7.1.1 STEC .....	81
7.1.2 VTEC.....	88
7.2 Δορυφορικός Σταθμός Αυστραλίας .....	95
7.2.1 STEC .....	95
7.3 Δορυφορικός Σταθμός Χιλής.....	101
7.3.1 STEC .....	101
7.4 Δορυφορικός Σταθμός Κένυας.....	107
7.4.1 STEC .....	108
7.5 Δορυφορικός Σταθμός Μογγολίας.....	114
7.5.1 STEC .....	114
Κεφάλαιο 8 <sup>ο</sup> Συμπεράσματα.....	122
Αναφορές .....	124
Παράρτημα Α.....	129
Πίνακας δεδομένων ανά μήνα.....	129
Παράρτημα Β.....	130
Αποτελέσματα εξηγήσιμης μάθησης για την επιρροή τιμών STEC - VTECγια τους σταθμούς Νορβηγίας και Ηνωμένων Πολιτειών Αμερικής .....	130
.....	138
.....	141
.....	144
Παράρτημα Γ .....	147
ΚΩΔΙΚΑΣ ΕΞΗΓΗΣΙΜΗΣ ΜΑΘΗΣΗΣ .....	147
STEC .....	147

---



---

## 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
MSE	Mean Square Error
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
SHAP	Shapely Additive exPLANations
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

## Κατάλογος εικόνων

Εικόνα 1. Μοντέλο PCA .....	17
Εικόνα 2. Μοντέλο OKSM.....	19
Εικόνα 3. Τεχνική SVM.....	20
Εικόνα 4. Τα στρώματα της ατμόσφαιρας .....	23
Εικόνα 5. Τα στρώματα της ιονόσφαιρας. ....	28
Εικόνα 6. Έκρηξη CME στις 31 Αυγούστου 2012 (Πηγή: NASA Goddard Space Flight Center) .....	36
Εικόνα 7. 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) .....	38
Εικόνα 8. Τεχνητή νοημοσύνη και τα υποσύνολά της. ....	42
Εικόνα 9. Συνάρτηση απώλειας Mean Absolute Error (MAE).....	52
Εικόνα 10. Συνάρτηση απώλειας Mean Square Error (MSE).....	53
Εικόνα 11. Σταθμός COTE00ATA (Πηγή: IGS).....	64
Εικόνα 12. Όλοι οι διαθέσιμοι σταθμοί του συστήματος GNSS (Πηγή: IGS).....	65
Εικόνα 13. Τυπική μορφή αρχείου RINEX.....	66
Εικόνα 14. Συμπεριφορά των τιμών TEC ανά ώρα με βάση το γεωγραφικό πλάτος .....	69
Εικόνα 15. Τιμές TEC ανά Δορυφόρο .....	69
Εικόνα 16. Η διακύμανση της τιμής TEC την 01 Ιανουαρίου 2018, ανά Δ/Φ (PRN 01, 02 κλπ) σε κάποιο εύρος υψομέτρων. ....	70
Εικόνα 17. Υπόδειγμα δεδομένων γεωμαγνητικής δραστηριότητας (K) .....	71
Εικόνα 18. 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).....	72
Εικόνα 19. Υπόδειγμα δεδομένων SME .....	72
Εικόνα 20. Υπόδειγμα δεδομένων Dst .....	73
Εικόνα 21. Υπόδειγμα δεδομένων Solar Radio Flux .....	75
Εικόνα 22. Υπόδειγμα δεδομένων Sunspot Number .....	76
Εικόνα 23. Βασιλικό Αστεροσκοπείο Βελγίου (Πηγή: wwwbis.sidc.be) .....	76
Εικόνα 24. Υπόδειγμα δεδομένων Energetic Particle Flux.....	77
Εικόνα 25. Υπόδειγμα δεδομένων Solar Flares.....	77
Εικόνα 26. Υπόδειγμα δεδομένων Coronal Mass Ejections. ....	78
Εικόνα 27. Ποιότητα σφάλματος στα σετ δεδομένων αξιολόγησης και εκπαίδευσης .....	81
Εικόνα 28. Αποτελέσματα μοντέλων εξηγήσιμης μάθησης με βάση τους δείκτες rMSE, MAE και R <sup>2</sup> .....	82
Εικόνα 29. Γραμμική παλινδρόμηση .....	83

---

---

## Κατάλογος πινάκων

---

---

Πίνακας 1. Παράμετροι επιρροής της ιονόσφαιρας. ....	31
Πίνακας 2. Excel με τα δεδομένα των παραμέτρων που χρησιμοποιήθηκαν.....	71

---

---

---

---

## Κεφάλαιο 1<sup>ο</sup> Εισαγωγή

---

---

### 1.1 Η σημασία της πρόβλεψης της ιονοσφαιρικής δραστηριότητας

---

Η μελέτη της επίδρασης διαφόρων παραμέτρων για την πρόβλεψη της ιονοσφαιρικής δραστηριότητας, αποτελεί μια διαδικασία συλλογής δεδομένων, που είναι απαραίτητη για πληθώρα εργασιών και μελετών. Το δορυφορικό σήμα GPS, καθώς διαδίδεται στην ατμόσφαιρα μέχρι τη γήινη επιφάνεια συνεχώς διαθλάται λόγω της μεταβολής του δείκτη διάθλασης της ατμόσφαιρας. Ειδικότερα το σήμα GPS, που είναι διαμορφωμένο κύμα και μεταφέρει πληροφορία, κατά τη διάδοσή του στην ιονόσφαιρα (τμήμα της ατμόσφαιρας από περίπου 50 έως 1000χλμ. ύψος από τη γήινη επιφάνεια με έντονη παρουσία ελεύθερων ηλεκτρονίων και ιόντων) εξαρτάται σημαντικά και από τις συχνότητες των συνιστωσών κυμάτων που το απαρτίζουν.

Τα σφάλματα που σχετίζονται με τη διάδοση του σήματος, όπως είναι τα ατμοσφαιρικά σφάλματα (στα οποία συμπεριλαμβάνεται και το ιονοσφαιρικό σφάλμα), αποτελούν τυχαία σφάλματα (random errors) που προφανώς είναι αναπόφευκτα κατά την εκτέλεση των μετρήσεων, και ονομάζονται και ως θόρυβος (noise, white noise). Η καθυστέρηση της ιονόσφαιρας είναι ανάλογη με το συνολικό αριθμό ηλεκτρονίων της ιονόσφαιρας (**TEC**) κατά μήκος της διαδρομής διάδοσης, και αντιστρόφως ανάλογη του τετραγώνου της συχνότητας  $f$  του φορέα (L1 ή L2). Οι τιμές TEC είναι απαραίτητες για τη διόρθωση της διάθλασης της ιονόσφαιρας και είναι ζωτικής σημασίας για τη δορυφορική πλοήγηση συστήματα, προκειμένου να διασφαλιστεί η υψηλή απόδοση των δορυφορικών συστημάτων στον εντοπισμό θέσης [1]

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 (O<sub>2</sub>, H<sub>2</sub>, He, N<sub>2</sub>) κυρίως ελεύθερων ηλεκτρονίων και θετικών ιόντων. Η ιονόσφαιρα αποτελεί μέσο διασποράς για το σήμα GPS, δηλαδή, ο δείκτης διάθλασης εξαρτάται από τη συχνότητα του σήματος και το γεγονός αυτό είναι πολύ βολικό για παρατηρήσεις GPS και στις δύο συχνότητες πχ. L1 και L2, L1 και L5. Αυτός είναι και ο βασικός σκοπός δύο τουλάχιστον συχνοτήτων L στο σήμα GPS. Με δέκτες διπλής συχνότητας το ιονοσφαιρικό σφάλμα μπορεί ουσιαστικά να απαλείφει αν χρησιμοποιήσουμε έναν κατάλληλο γραμμικό συνδυασμό πρωτογενών παρατηρήσεων των δύο φορέων, είτε ψευδοαποστάσεων από κώδικες είτε φάσεων. [2]

Περιβαλλοντικές, τοπογραφικές και γεωαδαιτικές μελέτες [3], [4], [5], [6] όλες καλούνται να απαντήσουν στην εξής απορία: Ποιος είναι ο βέλτιστος συνδυασμών παραμέτρων επιρροής της ιονοσφαιρικής δραστηριότητας, που θα οδηγήσει στην ελαχιστοποίηση του ιονοσφαιρικού σφάλματος;

Οι παραπάνω μελέτες όσο αφορά την συλλογή δεδομένων έχουν μια ειδοποιό διαφορά, τον τρόπο εύρεσης του συνδυασμού των παραμέτρων που αποσκοπεί στην ελαχιστοποίηση του ιονοσφαιρικού σφάλματος. Γενικότερα, , για τη διόρθωση της ιονόσφαιρικής καθυστέρησης χρησιμοποιούνται μοντέλα καθυστέρησης εκπομπής ιονόσφαιρας που βασίζονται σε απλά εμπειρικά μοντέλα TEC όπως τα μοντέλα Klobuchar (1987) και NeQuick (Hochegger et al. 2000, Radicella and Leitinger 2001). Καθώς όμως, ο χρόνος ανταπόκρισης και διάσωσης είναι πολύ σημαντικός, είναι αναγκαίο να υπάρξει και άλλη προσέγγιση, κατά την οποία κάνοντας χρήση μοντέλων μηχανικής μάθησης και μετρήσεων GNSS να γίνεται αυτόματα η εύρεση του βέλτιστου συνδυασμού παραμέτρων για την ελαχιστοποίηση του ιονοσφαιρικού σφάλματος.

## 1.2 Κύριες προκλήσεις

---

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable Artificial Intelligence - XAI) αναφέρεται στη δημιουργία μοντέλων μηχανικής μάθησης που είναι εύκολα ερμηνεύσιμα και κατανοητά από ανθρώπους. Για την ερμηνεία μοντέλων μηχανικής πρόβλεψης ιονοσφαιρικής δραστηριότητας, η XAI μπορεί να είναι ιδιαίτερα χρήσιμη, καθώς μπορεί να παρέχει λεπτομερείς εξηγήσεις για το πώς λειτουργούν τα μοντέλα και ποια χαρακτηριστικά των δεδομένων είναι σημαντικά για τις προβλέψεις τους.

Για παράδειγμα, ένα εξηγήσιμο μοντέλο μπορεί να αναδείξει τη σημασία διαφόρων φυσικών παραμέτρων της ιονόσφαιρας, όπως η ηλιακή δραστηριότητα, η γεωμαγνητική δραστηριότητα, η θερμοκρασία της ατμόσφαιρας και άλλες, στην πρόβλεψη της ιονοσφαιρικής δραστηριότητας. Με αυτόν τον τρόπο, οι ερευνητές και οι ειδικοί μπορούν να κατανοήσουν καλύτερα τους μηχανισμούς που διέπουν την ιονοσφαιρική δραστηριότητα και να προβλέψουν τις συνέπειές της.

Επιπλέον, η XAI μπορεί να βοηθήσει στον εντοπισμό πιθανών παραμέτρων που είναι αναξιόπιστες ή δεν είναι κατάλληλες για την πρόβλεψη της ιονοσφαιρικής δραστηριότητας, καθώς και στην ανίχνευση πιθανών παραδόξων ή ανωμαλιών στα δεδομένα.

Τέλος, η XAI μπορεί να βοηθήσει στην επικοινωνία των αποτελεσμάτων της μηχανικής πρόβλεψης ιονοσφαιρικής δραστηριότητας με τους μη ειδικούς, όπως τους πολιτικούς και τους λήπτες αποφάσεων, μέσω πιο κατανοητών και διαφανών εξηγήσεων.

Στην παρούσα διπλωματική εργασία εξετάζεται πως μέσω της εξηγήσιμης τεχνητής νοημοσύνης, υπάρχει η δυνατότητα να ερμηνευθούν μοντέλα μηχανικής πρόβλεψης της ιονόσφαιρας [8].

Το κύριο ερευνητικό πεδίο που συζητείται εδώ είναι το ιονοσφαιρικό μοντέλο TEC, που είναι ένα σημαντικό ζήτημα σε πολλές real – time εφαρμογές GNSS. Η αξιόπιστη και γρήγορη γνώση για τις ιονοσφαιρικές παραλλαγές γίνεται όλο και πιο σημαντική. Χρήστες GNSS, με δέκτες μονής συχνότητας και συστημάτων δορυφορικής πλοήγησης,



χρειάζονται ακριβείς διορθώσεις για να αφαιρέσουν την υποβάθμιση του σήματος που προκαλείται από την ιονόσφαιρα. Η ιονοσφαιρική μοντελοποίηση με κατάλληλο συνδυασμό παραμέτρων είναι το αντικείμενο συζήτησης στο παρόντα συνεισφορά.

### 1.2.1 Η πρόκληση της μοντελοποίησης TEC με χρήση δεδομένων GNSS

---

Η μεταβλητότητα των ιονοσφαιρικών παραμέτρων περιορίζει την αποτελεσματικότητα των επικοινωνιών, των ραντάρ και συστήματα πλοήγησης. Ο όρος TEC, που ορίζεται ως το ολοκλήρωμα της πυκνότητας ηλεκτρονίων σε μια διαδρομή σήματος, χρησιμοποιείται συχνά για να περιγράψει τη μεταβλητότητα της ιονόσφαιρας [9], [10]. Οι τιμές TEC είναι απαραίτητες για τη διόρθωση της διάθλασης της ιονόσφαιρας [11] και είναι ζωτικής σημασίας για τα συστήματα δορυφορικής πλοήγησης, προκειμένου να διασφαλιστεί η υψηλή απόδοση των δορυφορικών συστημάτων στον εντοπισμό θέσης.

Τα μεταδιδόμενα σήματα από τα Global Navigation Satellite Systems (GNSS) επηρεάζονται άμεσα από τις ιονοσφαιρικές διακυμάνσεις, προκαλώντας καθυστερήσεις [12]. Αυτές οι καθυστερήσεις εξαρτώνται από τη συχνότητα του σήματος και την πυκνότητα ηλεκτρονίων κατά μήκος της διαδρομής μετάδοσης. Ως εκ τούτου, εισάγεται η ιονοσφαιρική μεταβλητότητα, μια πρόσθετη πηγή σφάλματος στην τοποθέτηση GNSS [13]. Η χρήση πολλαπλών σημάτων πλοήγησης της διακριτής κεντρικής συχνότητας που μεταδίδεται από τον ίδιο δορυφόρο GNSS, επιτρέπει την άμεση εκτίμηση αυτών των ιονοσφαιρικών καθυστερήσεων. Αξιοποιώντας το γεγονός ότι διαφορετικές συχνότητες σήματος επηρεάζονται διαφορετικά από την ιονόσφαιρα, μια κατάλληλη στρατηγική επεξεργασίας σημάτων GNSS πολλαπλών συχνοτήτων, εξαλείφει το ιονοσφαιρικό σφάλμα [13].

Σε αντίθεση με τους δέκτες GNSS πολλαπλών συχνοτήτων, σε πραγματικό χρόνο (RT) μονής συχνότητας, ο εντοπισμός με χαμηλού-κόστους δέκτη, μονής συχνότητας (SF), έχει λάβει αυξανόμενη προσοχή τα τελευταία χρόνια λόγω του μεγάλο αριθμού πιθανών εφαρμογών του. Ωστόσο, σε αυτή την περίπτωση, μια σημαντική πρόκληση είναι ο αποτελεσματικός μετριασμός αυτών των ιονοσφαιρικών καθυστερήσεων [14]. Οι τεχνικές RT-SF-SPP (Standard Point Positioning)/PPP (Precise Point Positioning) χρησιμοποιούν προϊόντα ιονόσφαιρας STEC και VTEC που απελευθερώνονται από την υπηρεσία διεθνούς υπηρεσίας GNSS (IGS) σε πραγματικό χρόνο [15], για την εξάλειψη του ιονοσφαιρικού σφάλματος και την εφαρμογή διορθώσεων στο μοντέλο ως εξωτερικές παραμέτρους.

Το Κέντρο Προσδιορισμού Τροχιών στην Ευρώπη (CODE) παρέχει Παγκόσμιους Ιονοσφαιρικούς Χάρτες (GIM) σε ένα κάρναβο (2,5°×5,0°), σε χρονική διάρκεια 2 ωρών ανάλυση, οδηγώντας σε ένα χωροχρονικό αραιό μοντέλο. Αυτό σημαίνει ότι οι διορθώσεις που εφαρμόστηκαν, βελτίωσαν την ακρίβεια της παρεχόμενης λύσης για την

τοποθέτηση. Ωστόσο, οι τιμές STEC και VTEC αποτυγχάνουν να αφαιρέσουν τη συνολική ποσότητα θορύβου που προκαλείται από την ιονόσφαιρα, λόγω της συνολικής 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 προσδιορίσει όσο το δυνατόν καλύτερα το ιονοσφαιρικό μοντέλο, μέσω της εύρεσης του κατάλληλου συνδυασμού των παραμέτρων επιρροής των τιμών TEC.

Ειδικότερα, οι κύριες προκλήσεις 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 δραστηριότητας ιονόσφαιρας και σε ακανόνιστες συνθήκες.
- Η διερεύνηση της σημασίας των διαφόρων εξωτερικών παραμέτρων (π.χ. ηλιακή και γεωμαγνητική δραστηριότητα) και η εύρεση του βέλτιστου συνδυασμού αυτών, για τη βελτίωση της ακρίβειας στη μοντελοποίηση TEC.
- Η εισαγωγή αξιόπιστων δεδομένων περιφερειακής παρατήρησης για τον χαρακτηρισμό περιφερειακών ανωμαλιών που επηρεάζουν την ιονόσφαιρα σε τοπικό επίπεδο (χωρική), και επιπλέον για την εκτίμηση των αλλαγών TEC, που στοχεύουν σε χωροχρονικά μοντέλα πρόβλεψης STEC και VTEC

### 1.3 Συνεισφορά

---

Η συνολική συμβολή της παρούσας διπλωματικής εργασίας έγκειται στους ακόλουθους στόχους που καθορίστηκαν:

- να προτείνει ένα μοντέλο, για ακριβείς εκτιμήσεις TEC από δεδομένα GNSS, και για την υπέρβαση προβλημάτων που σχετίζονται με την ακανόνιστη συμπεριφορά της ιονόσφαιρας σε περιφερειακό επίπεδο
- να προτείνει ισχυρά μοντέλα εξηγήσιμης τεχνητής νοημοσύνης που εκμεταλλεύονται διάφορες εξωτερικές παραμέτρους (π.χ. σχετικές ηλιακή και γεωμαγνητική δραστηριότητα) για τη βελτίωση της πρόβλεψης και της ακρίβειας στη μοντελοποίηση TEC
- να παρουσιάσει μια ολοκληρωμένη ανάλυση για διάφορους μήνες και διάφορους σταθμούς ανά τον κόσμο, υπό διαφορετική ηλιακή και γεωμαγνητική δραστηριότητα. Τα μοντέλα ελέγχονται για διαφορετικές ημέρες για να ελεγχθεί η ανταπόκριση του μοντέλου σε περιπτώσεις υψηλής δραστηριότητας ιονόσφαιρας

## 1.4 Δομή εργασίας

---

Στα πλαίσια της παρούσας διπλωματικής εργασίας θα μελετήσουμε την επίδραση της Εξηγήσιμη Τεχνητή Νοημοσύνη για την ερμηνεία μοντέλων μηχανικής πρόβλεψης της ιονοσφαιρικής δραστηριότητας κάνοντας χρήση μοντέλων μηχανικής μάθησης και μετρήσεων GNSS. Στην ουσία μέσω της εκπαίδευσης ενός νευρωνικού δικτύου, θα μπορεί να προβλεφθεί το ιονοσφαιρικό σφάλμα, και αυτό θα έχει ως αποτέλεσμα την χρήση δέκτη μίας μόνο συχνότητας στις παρατηρήσεις GPS. Η μηχανική τους περιγραφή (μοντελοποίηση), αν και ικανοποιητική στις περισσότερες περιπτώσεις, έχει ως αποτέλεσμα την αύξηση των αγνώστων παραμέτρων στα μοντέλα συνόρθωσης και τη μείωση της ισχύος της λύσης.

Στο επόμενο κεφάλαιο θα γίνει μια βιβλιογραφική ανασκόπηση σε έρευνες που έχουν αντιμετωπίσει το πρόβλημα της πρόβλεψης του ιονοσφαιρικού σφάλματος και των τεχνικών που χρησιμοποιήσαν για να επιτευχθεί αυτό.

Στο τρίτο κεφάλαιο αναφέρονται κάποια γενικά στοιχεία για τα συστήματα GNSS, την ατμόσφαιρα της Γης και γίνεται ειδική αναφορά στις παραμέτρους που επηρεάζουν την ιονόσφαιρα.

Στο τέταρτο κεφάλαιο, γίνεται αναφορά σε βασικές έννοιες για την κατανόηση των όρων της τεχνητής νοημοσύνης, για την καλύτερη παρακολούθηση της ροής εργασίας.

Στο πέμπτο κεφάλαιο, παρουσιάζεται η γλώσσα προγραμματισμού Python στην οποία έγιναν όλα τα πειράματα, ενώ παράλληλα γίνεται αναφορά στις βιβλιοθήκες που χρησιμοποιήθηκαν και στην υλοποίηση του αλγορίθμου.

Στο έκτο κεφάλαιο περιγράφεται η πειραματική διαχείριση και αφορά την περιγραφή των δεδομένων GNSS που συλλέχθηκαν, ο τρόπος διαχείρισης τους και γίνεται αναφορά στα μοντέλα – τεχνικές εξηγήσιμης μάθησης που χρησιμοποιήθηκαν.

Στο έβδομο κεφάλαιο γίνεται αναλυτική αναφορά στα δεδομένα των δορυφορικών σταθμών που επιλέχθηκαν και στην επιλογή του βέλτιστου μοντέλου καθώς και της πιο σημαντικής παραμέτρου επιρροής τιμών TEC.

Τέλος, στο όγδοο κεφάλαιο παρατίθενται τα συμπεράσματα που προκύπτουν από όλη την προηγούμενη ανάλυση και επεξεργασία, ενώ αναγράφονται και κάποιες μελλοντικές κατευθύνσεις όσον αφορά τον συνδυασμό της μηχανικής μάθησης και της παρακολούθησης των τυχαίων σφαλμάτων που επηρεάζουν τον προσδιορισμό της θέσης ( και του χρόνου κατά περίπτωση) με παρατηρήσεις GPS.

---

---

## Κεφάλαιο 2<sup>ο</sup> Βιβλιογραφική ανασκόπηση

---

---

Στο συγκεκριμένο κεφάλαιο, γίνεται βιβλιογραφική ανασκόπηση σε εργασίες που έχουν αντιμετωπίσει με διάφορους τρόπους το πρόβλημα της πρόβλεψης τη ιονοσφαιρικής δραστηριότητας. Έχοντας ως βάση τις δυσκολίες σύμφωνα με τους τύπους δεδομένων για την υλοποίηση μιας τέτοιας διαδικασίας από το προηγούμενο κεφάλαιο, αλλά και τη σημασία αυτής της, αξίζει να μελετηθεί η μεθοδολογία αντιμετώπισης του και από άλλους ερευνητές. Οι εργασίες ταξινομούνται σύμφωνα με τον τύπο δεδομένων που χρησιμοποιήθηκαν με σκοπό την προσέγγιση στη λύση του προβλήματος.

---

### 2.1 Διαστημικά Καιρικά Φαινόμενα

---

Σύμφωνα με το [16], επισημαίνεται ότι τα διαστημικά καιρικά φαινόμενα μπορούν να προκαλέσουν σοβαρές διαταραχές της ιονοσφαιρικής συμπεριφοράς.

Εξετάζεται λοιπόν η τρέχουσα κατάσταση της ανίχνευσης, παρακολούθησης και πρόβλεψης ιονοσφαιρικών διαταραχών στην Ευρώπη με βάση το GNSS σε σχέση με τη δράση COST ES0803 «Ανάπτυξη προϊόντων και υπηρεσιών διαστημικού καιρού στην Ευρώπη».

Η έρευνα του διαστημικού καιρού και οι σχετικές μελέτες ιονόσφαιρας απαιτούν ευρεία διεθνή συνεργασία για την κοινή χρήση βάσεων δεδομένων, την ανάπτυξη λογισμικού και μοντέλων ανάλυσης και την παροχή υπηρεσιών. Γι αυτό το λόγο, αναθεωρήθηκε η ευρωπαϊκή βάση δεδομένων GNSS, συμπεριλαμβανομένων των υπηρεσιών ιονόσφαιρας που παρέχουν παράγωγα προϊόντα δεδομένων, όπως το συνολικό περιεχόμενο ηλεκτρονίων (TEC) και οι δείκτες ραδιοσπινθηρισμού. Τα θεμελιώδη φαινόμενα ιονοσφαιρικής διαταραχής που καλύπτουν αρκετά διαφορετικές κλίμακες σε χρόνο και χώρο, εξετάστηκαν υπό το πρίσμα των πρόσφατων επιτευγμάτων στην ιονοσφαιρική παρακολούθηση που βασίζεται στο GNSS. Έτσι, λαμβάνονται υπόψη διεργασίες διαταραχής μεγάλης κλίμακας που χαρακτηρίζονται από κινούμενα μέτωπα ιονισμού, ιοντίζουσες ιονόσφαιρες που μοιάζουν με κύμα και τέλος μικρής κλίμακας ανωμαλίες που προκαλούν ραδιοσπινθηρισμούς.

Ενώ οι επίγειες και διαστημικές τεχνικές παρακολούθησης GNSS είναι καλά ανεπτυγμένες, η πρόβλεψη ιονόσφαιρων διαταραχών χρειάζεται πολύ περισσότερη δουλειά για να γίνει ελκυστική για χρήστες που μπορεί να ενδιαφέρονται για συμπυκνωμένες πληροφορίες σχετικά με το βαθμό διαταραχής της ιονόσφαιρας από ισχυρούς δείκτες.

## 2.2 Νευρώνες μακράς-βραχείας διάρκειας μνήμης (Long Short-Term Memory - LSTM)

---

Σε αυτή τη μελέτη [17], όπως και η συγκεκριμένη διπλωματική εργασία, το νευρωνικό δίκτυο μακράς βραχυπρόθεσμης μνήμης (LSTM) (NN) εφαρμόζεται για την πρόβλεψη των 256 σφαιρικών αρμονικών (SH) συντελεστών που χρησιμοποιούνται παραδοσιακά για την κατασκευή παγκόσμιων ιονοσφαιρικών χαρτών (GIM). Πολλαπλά δεδομένα εισόδου, συμπεριλαμβανομένων των ιστορικών χρονοσειρών των συντελεστών SH, της ηλιακής ακραίας υπεριώδους ροής (EUV), του δείκτη χρόνου καταιγίδας διαταραχής (Dst) και της ώρας της ημέρας, χρησιμοποιούνται στο αναπτυγμένο μοντέλο LSTM NN. Διαφορετικοί συνδυασμοί των παραπάνω παραμέτρων έχουν χρησιμοποιηθεί για την κατασκευή του μοντέλου LSTM NN και βρέθηκε ότι το μοντέλο που χρησιμοποιεί και τις τέσσερις παραμέτρους έχει την καλύτερη απόδοση.

Στη συνέχεια, το μοντέλο LSTM με την καλύτερη απόδοση χρησιμοποιείται για την πρόβλεψη των συντελεστών SH και οι παγκόσμιοι ωριαίοι χάρτες TEC αναπαράγονται χρησιμοποιώντας τους 256 προβλεπόμενους συντελεστές SH. Πραγματοποιείται συνολική αξιολόγηση σε σχέση με το CODE GIM TEC.

Τα αποτελέσματα δείχνουν ότι το μέσο τετραγωνικό σφάλμα ρίζας TEC πρώτης/δεύτερης ώρας (RMSE) είναι 1,27/2,20 TECU κατά τη διάρκεια της καταιγίδας και 0,86/1,51 TECU κατά τη διάρκεια της ώρας ησυχίας, επομένως το αναπτυγμένο μοντέλο έχει καλή απόδοση τόσο σε περιόδους ησυχίας όσο και σε περιόδους καταιγίδας. Επιπλέον, τυπικές ιονοσφαιρικές δομές, όπως η ανωμαλία ιοντισμού του ισημερινού (EIA) και η πυκνότητα ενισχυμένη από καταιγίδες (SED), αναπαράγονται καλά στους προβλεπόμενους χάρτες TEC κατά τη διάρκεια της καταιγίδας. Το μοντέλο που αναπτύχθηκε δείχνει επίσης ανταγωνιστική απόδοση στην πρόβλεψη του παγκόσμιου TEC σε σύγκριση με το μοντέλο επιμονής και δύο εμπειρικά μοντέλα (IRI-2016 και NeQuick-2).

## 2.3 Principal component analysis (PCA)

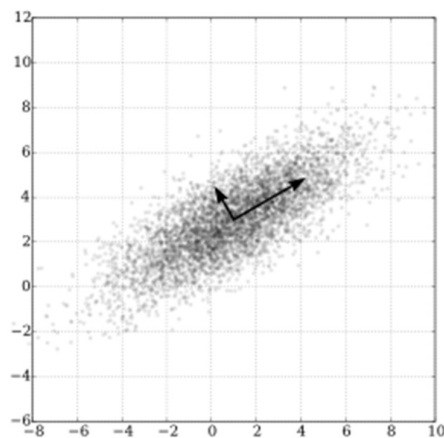
---

Αυτή η εργασία [18], 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 ηλιακός δείκτης (F10.7), 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.

Η ανάλυση κύριου συστατικού (PCA) είναι η διαδικασία υπολογισμού των κύριων στοιχείων και χρήσης τους για την πραγματοποίηση αλλαγής βάσης στα δεδομένα, μερικές φορές χρησιμοποιώντας μόνο τα πρώτα κύρια στοιχεία και αγνοώντας τα υπόλοιπα.

Το PCA χρησιμοποιείται στην διερευνητική ανάλυση δεδομένων και στη δημιουργία προγνωστικών μοντέλων. Χρησιμοποιείται συνήθως για τη μείωση των διαστάσεων προβάλλοντας κάθε σημείο δεδομένων μόνο στα πρώτα βασικά στοιχεία για να ληφθούν δεδομένα χαμηλότερης διάστασης, διατηρώντας παράλληλα όσο το δυνατόν μεγαλύτερη παραλλαγή των δεδομένων. Η πρώτη κύρια συνιστώσα μπορεί ισοδύναμα να οριστεί ως μια κατεύθυνση που μεγιστοποιεί τη διακύμανση των προβαλλόμενων δεδομένων.

Από κάθε στόχο, μπορεί να φανεί ότι τα κύρια συστατικά είναι ιδιοδιανύσματα του πίνακα συνδιακύμανσης των δεδομένων. Έτσι, τα κύρια συστατικά υπολογίζονται συχνά με ιδιοαποσύνθεση του πίνακα συνδιακύμανσης δεδομένων ή με αποσύνθεση μοναδικής τιμής του πίνακα δεδομένων. Το PCA είναι η απλούστερη από τις πραγματικές πολυμεταβλητές αναλύσεις που βασίζονται σε ιδιοδιάνυσμα και σχετίζεται στενά με την ανάλυση παραγόντων. Η παραγοντική ανάλυση συνήθως ενσωματώνει περισσότερες ειδικές υποθέσεις για την υποκείμενη δομή και επιλύει ιδιοδιανύσματα ενός ελαφρώς διαφορετικού πίνακα. Η PCA σχετίζεται επίσης με την κανονική ανάλυση συσχέτισης (CCA). Το CCA ορίζει συστήματα συντεταγμένων που περιγράφουν βέλτιστα τη διασταυρούμενη συνδιακύμανση μεταξύ δύο συνόλων δεδομένων ενώ το PCA ορίζει ένα νέο ορθογώνιο σύστημα συντεταγμένων που περιγράφει βέλτιστα τη διακύμανση σε ένα μεμονωμένο σύνολο δεδομένων.



Εικόνα 1. Μοντέλο PCA

## 2.4 OKSM

---

Αυτό το έγγραφο [19], περιγράφει τη μεθοδολογία πρόβλεψης TEC χρησιμοποιώντας το συνηθισμένο μοντέλο υποκατάστασης με βάση το Kriging (OKSM). Το OKSM αξιολογείται χρησιμοποιώντας τα δεδομένα που λαμβάνονται και συλλέγονται από τον σταθμό δέκτη IRNSS που είναι εγκατεστημένος στο ACS College of Engineering (ACSCE), στο Bengaluru (12,8913 °N, 77,4658 °E), στην Ινδία και σε άλλους σταθμούς δικτύου Διεθνούς Υπηρεσίας GNSS (IGS). Τα δεδομένα IRNSS TEC (Ιανουάριος 2018) υπολογίζονται χρησιμοποιώντας τη μέθοδο ψευδοεμβέλειας διπλής συχνότητας (L5 & S) και το TEC εξομαλύνεται με τη μέθοδο κανονικής κυβικής εξομάλυνσης spline. Το IRNSS Vertical TEC (VTEC) προβλέπεται από τις 16 Ιανουαρίου 2018 έως τις 26 Ιανουαρίου 2018 χρησιμοποιώντας προηγούμενες έξι ημέρες εκτιμώμενων τιμών VTEC. Ομοίως, το GPS VTEC για το σταθμό IGS στο IISC, Bengaluru προβλέπεται επίσης για την ίδια διάρκεια για την επικύρωση του ανεπτυγμένου OKSM.

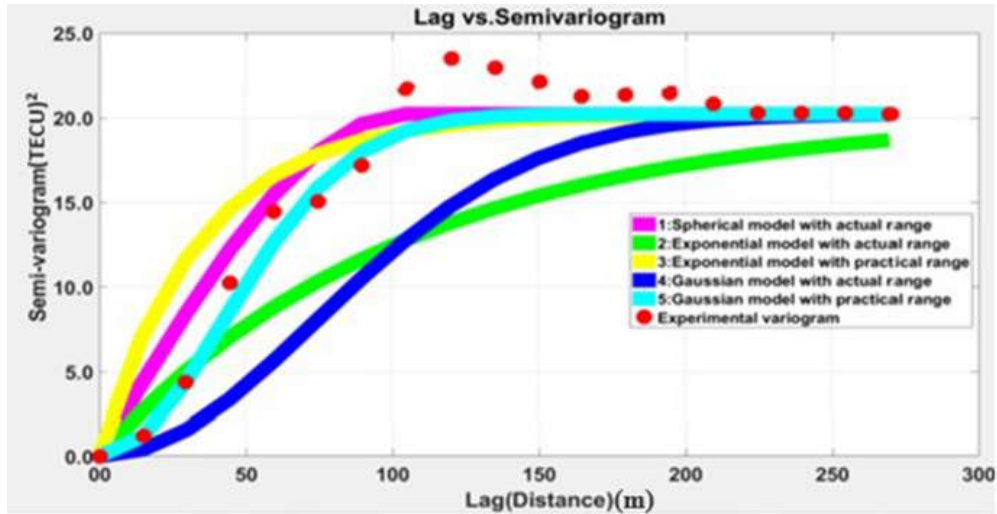
Προκειμένου να αξιολογηθεί η απόδοση του αναπτυγμένου μοντέλου πρόβλεψης για διαφορετικές γεωγραφικές τοποθεσίες, συνθήκες ηλιακής δραστηριότητας και εποχές, το VTEC προβλέπεται και αναλύεται για διαφορετικές περιοχές γεωγραφικού πλάτους, όπως σταθμός PHON χαμηλού γεωγραφικού πλάτους (6,9599 °N, 158,2101 °E), μεσαία - Γεωγραφικό πλάτος σταθμός ALGO (45,9588 °N, -78,0714 °E) και σταθμός NRIL μεγάλου πλάτους (69,3618 °N, 88,3597 °E) κατά τη διάρκεια διαφορετικών συνθηκών ηλιακής δραστηριότητας (Χαμηλή-2008, Μέση-2011 και Υψηλή ηλιακή δραστηριότητα 2013) και κατά τη διάρκεια διαφορετικών εποχών (άνοιξη, καλοκαίρι, βροχερός και χειμώνας) το έτος 2017.

Από την ανάλυση των αποτελεσμάτων πρόβλεψης OKSM, παρατηρείται ότι, το RMSE της προβλεπόμενης TEC ποικίλλει από 0,79 έως 3,6 TECU, το MAE είναι 0,4 έως 3 TECU και το MAPE είναι εντός 40% για ιονοσφαιρικές ήρεμες μέρες. Το VTEC προβλέπεται επίσης κατά τη διάρκεια των ημερών καταιγίδας (26 Οκτωβρίου 2003 έως 31 Οκτωβρίου

Για τη μελέτη της απόδοσης του μοντέλου, τα αποτελέσματα πρόβλεψης VTEC του OKSM συγκρίνονται με τα αποτελέσματα πρόβλεψης από το Standard Persistence Model (SPM) και το VTEC που προέρχονται από το μοντέλο International Reference Ionosphere (IRI-2016). Το RMSE του OKSM είναι 1,9679 TECU, το MAE είναι 1,245 TECU και το MAPE είναι 9%, ενώ για το SPM, το RMSE είναι 4,8372 TECU, το MAE είναι 3,7496 TECU και το MAPE είναι 36%. Ομοίως, για το μοντέλο IRI-2016, το RMSE είναι 7,9 TECU, το MAE είναι 7,1976 TECU και το MAPE είναι 66%. Επομένως, οι προβλέψεις TEC από την OKSM είναι καλύτερες από τις SPM και IRI-2016. Τα αποτελέσματα δείχνουν ότι το OKSM είναι χρήσιμο για εφαρμογές σε ιονοσφαιρικές μελέτες.



2003).



Εικόνα 2. Μοντέλο OKSM

## 2.5 Τεχνικές Μοντελοποίησης DIAS

Η εργασία [20], περιγράφει τη χρήση των ιονόσφαιρων μετρήσεων στις τεχνικές μοντελοποίησης DIAS για την προδιαγραφή, την πρόβλεψη και την πρόβλεψη της ιονόσφαιρας στην ευρωπαϊκή περιοχή και περιγράφει λεπτομερώς τα τελικά προϊόντα που είναι διαθέσιμα στην κοινότητα χρηστών DIAS.

Δύο είναι οι κύριοι στόχοι του έργου DIAS (European Digital Upper Atmosphere Server). Πρώτον, δημιουργεί μια πανευρωπαϊκή αποθήκη ακατέργαστων και παράγωγων ψηφιακών δεδομένων που περιγράφει την κατάσταση του ιονόσφαιρου τμήματος της ανώτερης ατμόσφαιρας, η οποία είναι ικανή να απορροφά πληροφορίες σε πραγματικό χρόνο και να διατηρεί συλλογές ιστορικών δεδομένων που παρέχονται από τους περισσότερους σταθμούς ιονόσφαιρας που λειτουργούν στην Ευρώπη.

Δεύτερον, το σύστημα DIAS παράγει και διανέμει, με βάση τη συλλογή ακατέργαστων δεδομένων, διάφορα λειτουργικά προϊόντα που απαιτούνται από διάφορες ομάδες χρηστών για σκοπούς ιονόσφαιρας τώρα μετάδοσης, πρόβλεψης και πρόβλεψης. Το έργο ολοκληρώθηκε τον Μάιο του 2006 και ο διακομιστής ΔΙΑΣ λειτουργεί έκτοτε συνεχώς.

Τα βασικά προϊόντα που παραδίδονται είναι ιονογράμματα σε πραγματικό χρόνο και ιστορικά από όλους τους ιονοσφαιρικούς σταθμούς που συνδέονται με το DIAS, γραφήματα συχνότητων και χάρτες της ιονόσφαιρας στην Ευρώπη με βάση τις παραμέτρους foF2, M(3000)F2, MUF και πυκνότητας ηλεκτρονίων, καθώς και μακροπρόθεσμη πρόβλεψη και βραχυπρόθεσμη πρόβλεψη έως και 24 ώρες μπροστά.

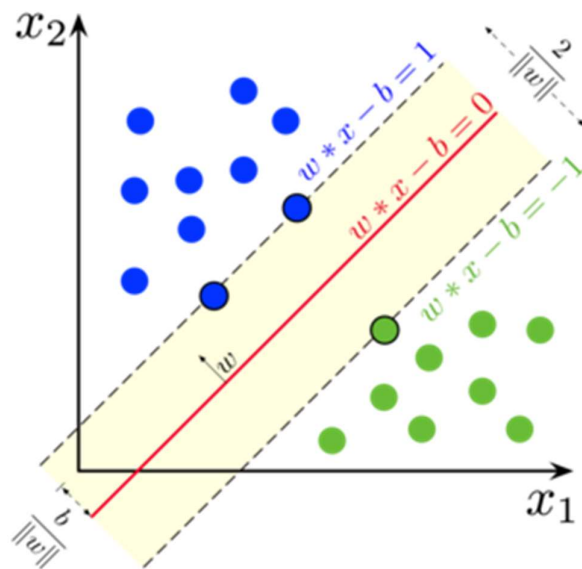


## 2.6 Τεχνική Support Vector Machine (SVM)

Αυτό το έγγραφο [21], προτείνει μια μέθοδο για την πρόβλεψη της ιονοσφαιρικής κρίσιμης συχνότητας, χρησιμοποιώντας την προσέγγιση της μηχανής διανύσματος υποστήριξης (SVM). Είναι ένας αλγόριθμος υπολογιστή που μαθαίνει με το παράδειγμα να εκχωρεί ετικέτες σε αντικείμενα.

Οι είσοδοι στο δίκτυο SVM είναι η ώρα της ημέρας, οι εποχιακές πληροφορίες, ο μέσος αριθμός ηλιακών κηλίδων 2 μηνών σε λειτουργία (R2), ο μέσος όρος λειτουργίας 3 ημερών του πλανητικού μαγνητικού δείκτη, η ηλιακή γωνία ζενίθ, η παρούσα τιμή  $f_{0F2}(t)$  και η πρώτη και η δεύτερη προσαύξησή του, η παρατήρηση του  $f_{0F2}$  στις  $t-23$  h, η μέση τιμή των 30 ημερών τη στιγμή,  $t$ ,  $f_{mF2}(t)$  και ο τρέχων μέσος όρος των προηγούμενων 30 ημερών του  $f_{0F2}$  σε  $t-23$  h  $f_{mF2}(t-23)$ . Η έξοδος είναι η προβλεπόμενη  $f_{0F2}$  1 ώρα μπροστά.

Το δίκτυο είναι εκπαιδευμένο να χρησιμοποιεί τα δεδομένα ηχογράφησης ιονόσφαιρας στους σταθμούς Haikou, Guangzhou, Chongqing, Lanzhou, Πεκίνο, Changchun και Manzhouli σε υψηλές και χαμηλές ηλιακές δραστηριότητες. Η απόδοση του μοντέλου SVM επαληθεύτηκε με δεδομένα που παρατηρήθηκαν. Δείχνεται ότι το προβλεπόμενο  $f_{0F2}$  έχει καλή συμφωνία με το παρατηρούμενο  $f_{0F2}$ . Η απόδοση του μοντέλου SVM είναι ανώτερη από αυτή των μοντέλων αυτοσυσχέτισης και επιμονής και ότι είναι συγκρίσιμη με αυτή του μοντέλου νευρωνικών δικτύων.



Εικόνα 3. Τεχνική SVM

---

---

## Κεφάλαιο 3<sup>ο</sup> Σύστημα GNSS και Ατμόσφαιρα

---

---

---

### 3.1 Βασικές αρχές συστήματος GNSS

---

Τα Global Navigation Satellite Systems (GNSS) 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.

## 3.2 Η ατμόσφαιρα της Γης

---

### 3.2.1 Εισαγωγικές Έννοιες

---

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

Η ατμόσφαιρα είναι ένα στρώμα αερίου ή στρώματα αερίων, που περιβάλλουν έναν πλανήτη και συγκρατούνται στη θέση τους από τη βαρύτητα του πλανητικού σώμα. Ένας πλανήτης διατηρεί μια ατμόσφαιρα όταν η βαρύτητα είναι μεγάλη και η θερμοκρασία της ατμόσφαιρας χαμηλή. Μια αστρική ατμόσφαιρα είναι η εξωτερική περιοχή ενός άστρου, η οποία περιλαμβάνει τα στρώματα πάνω από την αδιαφανή φωτόσφαιρα. αστέρια χαμηλής θερμοκρασίας μπορεί να έχουν εξωτερικές ατμόσφαιρες που περιέχουν μόρια ένωσης.

Η ατμόσφαιρα της Γης αποτελείται από άζωτο (78%), οξυγόνο (21%), αργό (0,9%), διοξείδιο του άνθρακα (0,04%) και ιχνοαέρια. Οι περισσότεροι οργανισμοί χρησιμοποιούν οξυγόνο για την αναπνοή. οι κεραινοί και τα βακτήρια πραγματοποιούν δέσμευση αζώτου για να παράγουν αμμωνία που χρησιμοποιείται για την παραγωγή νουκλεοτιδίων και αμινοξέων. τα φυτά, τα φύκια και τα κυανοβακτήρια χρησιμοποιούν διοξείδιο του άνθρακα για φωτοσύνθεση. Η πολυεπίπεδη σύνθεση της ατμόσφαιρας ελαχιστοποιεί τις βλαβερές επιπτώσεις του ηλιακού φωτός, της υπερϊώδους ακτινοβολίας, του ηλιακού ανέμου και των κοσμικών ακτίνων για την προστασία των οργανισμών από γενετικές βλάβες. Η σημερινή σύνθεση της ατμόσφαιρας της Γης είναι προϊόν δισεκατομμυρίων ετών βιοχημικής τροποποίησης του.

### 3.2.2 Τα στρώματα της ατμόσφαιρας

---

Η ατμόσφαιρα αποτελείται από στρώματα με βάση τη θερμοκρασία. Αυτά τα στρώματα είναι η τροπόσφαιρα, η στρατόσφαιρα, η μεσόσφαιρα και η θερμόσφαιρα. Μια περαιτέρω περιοχή σε περίπου 500 km πάνω από την επιφάνεια της Γης ονομάζεται εξώσφαιρα.

Μπορεί να χωριστεί σε στρώματα με βάση τη θερμοκρασία της, όπως φαίνεται στο παρακάτω σχήμα. Αυτά τα στρώματα είναι η τροπόσφαιρα, η στρατόσφαιρα, η μεσόσφαιρα και η θερμόσφαιρα. Μια περαιτέρω περιοχή, που ξεκινά περίπου 500 km πάνω από την επιφάνεια της Γης, ονομάζεται εξώσφαιρα.

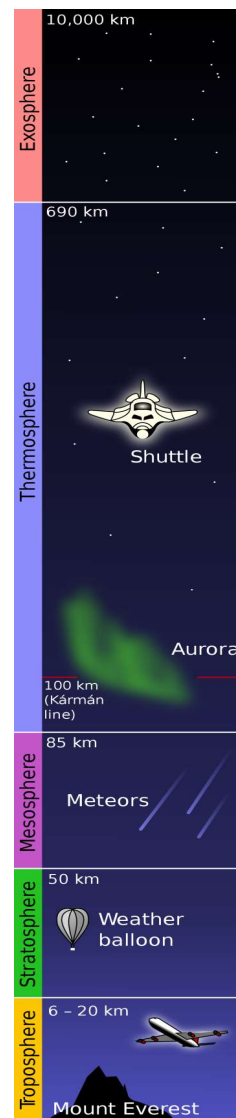
- **Η Τροπόσφαιρα**

Αυτό είναι το χαμηλότερο μέρος της ατμόσφαιρας - το μέρος στο οποίο ζούμε. Περιέχει το μεγαλύτερο μέρος του καιρού μας - σύννεφα, βροχή, χιόνι. Σε αυτό το τμήμα της ατμόσφαιρας η θερμοκρασία γίνεται πιο κρύα καθώς η απόσταση πάνω από τη γη αυξάνεται, κατά περίπου 6,5°C ανά χιλιόμετρο. Η πραγματική αλλαγή της θερμοκρασίας με το ύψος ποικίλλει από μέρα σε μέρα, ανάλογα με τον καιρό.

Η τροπόσφαιρα περιέχει περίπου το 75% του συνόλου του αέρα στην ατμόσφαιρα και σχεδόν το σύνολο των υδρατμών (που σχηματίζει σύννεφα και βροχή). Η μείωση της θερμοκρασίας με το ύψος είναι αποτέλεσμα της φθίνουσας πίεσης. Εάν ένα δέμα αέρα κινηθεί προς τα πάνω, διαστέλλεται (λόγω της χαμηλότερης πίεσης). Όταν ο αέρας διαστέλλεται, ψύχεται. Έτσι, ο αέρας ψηλότερα είναι πιο δροσερός από τον αέρα κάτω.

Το χαμηλότερο τμήμα της τροπόσφαιρας ονομάζεται οριακό στρώμα. Εδώ καθορίζεται η κίνηση του αέρα από τις ιδιότητες της επιφάνειας της Γης. Οι αναταράξεις δημιουργούνται καθώς ο άνεμος πνέει πάνω από την επιφάνεια της Γης και από θερμικές θερμοκρασίες που αναδύονται από τη γη καθώς θερμαίνεται από τον ήλιο. Αυτή η αναταραχή ανακατανέμει τη θερμότητα και την υγρασία εντός του οριακού στρώματος, καθώς και τους ρύπους και άλλα συστατικά της ατμόσφαιρας.

Η κορυφή της τροπόσφαιρας ονομάζεται τροπόπαυση. Αυτό είναι το χαμηλότερο στους πόλους, όπου είναι περίπου 7 - 10 km πάνω από την επιφάνεια της Γης. Είναι το υψηλότερο (περίπου 17 - 18 km) κοντά στον ισημερινό.



Εικόνα 4. Τα στρώματα της ατμόσφαιρας

- **Η Στρατόσφαιρα**

Αυτό εκτείνεται προς τα πάνω από την τροπόπαυση σε περίπου 50 km. Περιέχει μεγάλο μέρος του όζοντος στην ατμόσφαιρα. Η αύξηση της θερμοκρασίας με το ύψος συμβαίνει λόγω της απορρόφησης της υπεριώδους ακτινοβολίας (UV) από τον ήλιο από αυτό το όζον. Οι θερμοκρασίες στη στρατόσφαιρα είναι οι υψηλότερες στον καλοκαιρινό πόλο και οι χαμηλότερες στον χειμερινό πόλο.

Απορροφώντας την επικίνδυνη υπεριώδη ακτινοβολία, το όζον στη στρατόσφαιρα μας προστατεύει από τον καρκίνο του δέρματος και άλλες βλάβες στην υγεία. Ωστόσο, οι χημικές ουσίες (ονομαζόμενες CFC ή φρέον και halons) που χρησιμοποιούνταν κάποτε σε ψυγεία, σπρέι και πυροσβεστήρες έχουν μειώσει την ποσότητα του όζοντος στη

στρατόσφαιρα, ιδιαίτερα σε πολικά γεωγραφικά πλάτη, οδηγώντας στη λεγόμενη «τρύπα του όζοντος της Ανταρκτικής».

Τώρα οι άνθρωποι έχουν σταματήσει να παράγουν τους περισσότερους από τους επιβλαβείς CFC που αναμένουμε ότι η τρύπα του όζοντος θα ανακάμψει τελικά τον 21ο αιώνα, αλλά αυτή είναι μια αργή διαδικασία.

- **Η Μεσόσφαιρα**

Η περιοχή πάνω από τη στρατόσφαιρα ονομάζεται μεσόσφαιρα. Εδώ η θερμοκρασία μειώνεται και πάλι με το ύψος, φτάνοντας στο ελάχιστο τους  $-90^{\circ}\text{C}$  περίπου στη «μεσόπαυση».

- **Η Θερμόσφαιρα και η Ιονόσφαιρα**

Η θερμόσφαιρα βρίσκεται πάνω από τη μεσόπαυση και είναι μια περιοχή στην οποία οι θερμοκρασίες αυξάνονται και πάλι με το ύψος. Αυτή η αύξηση της θερμοκρασίας προκαλείται από την απορρόφηση της ενεργητικής υπεριώδους ακτινοβολίας και των ακτίνων X από τον ήλιο.

Στην περιοχή της ατμόσφαιρας πάνω από περίπου 80 km προκαλείται επίσης η «ιονόσφαιρα», αφού η ενεργειακή ηλιακή ακτινοβολία χτυπά τα ηλεκτρόνια από τα μόρια και τα άτομα, μετατρέποντάς τα σε «ιόντα» με θετικό φορτίο. Η θερμοκρασία της θερμόσφαιρας ποικίλλει μεταξύ νύχτας και ημέρας και μεταξύ των εποχών, όπως και ο αριθμός των ιόντων και των ηλεκτρονίων που υπάρχουν. Η ιονόσφαιρα αντανακλά και απορροφά ραδιοκύματα, επιτρέποντάς μας να λαμβάνουμε ραδιοφωνικές εκπομπές βραχέων κυμάτων στη Νέα Ζηλανδία από άλλα μέρη του κόσμου.

- **Η Εξώσφαιρα**

Η περιοχή πάνω από περίπου 500 km ονομάζεται εξώσφαιρα. Περιέχει κυρίως άτομα οξυγόνου και υδρογόνου, αλλά είναι τόσο λίγα από αυτά που σπάνια συγκρούονται - ακολουθούν «βαλλιστικές» τροχιές υπό την επίδραση της βαρύτητας και μερικά από αυτά διαφεύγουν αμέσως στο διάστημα.

- **Η Μαγνητόσφαιρα**

Η γη συμπεριφέρεται σαν ένας τεράστιος μαγνήτης. Παγιδεύει ηλεκτρόνια (αρνητικό φορτίο) και πρωτόνια (θετικά), συγκεντρώνοντάς τα σε δύο ζώνες περίπου 3.000 και 16.000 km πάνω από την υδρόγειο - τις ζώνες «ακτινοβολίας» Van Allen. Αυτή η εξωτερική περιοχή που περιβάλλει τη γη, όπου τα φορτισμένα σωματίδια κινούνται σπειροειδώς κατά μήκος των γραμμών του μαγνητικού πεδίου, ονομάζεται μαγνητόσφαιρα.

### 3.3 Η Ιονόσφαιρα και τα στρώματά της

---

#### 3.3.1 Εισαγωγικές Έννοιες

---

Η ιονόσφαιρα είναι το ιονισμένο τμήμα της ανώτερης ατμόσφαιρας της Γης, από περίπου 48 km (30 mi) έως 965 km (600 mi) υψόμετρο, μια περιοχή που περιλαμβάνει τη θερμόσφαιρα και μέρη της μεσόσφαιρας και της εξώσφαιρας. Η ιονόσφαιρα ιονίζεται από την ηλιακή ακτινοβολία. Παίζει σημαντικό ρόλο στον ατμοσφαιρικό ηλεκτρισμό και σχηματίζει το εσωτερικό άκρο της μαγνητόσφαιρας. Έχει πρακτική σημασία γιατί, μεταξύ άλλων λειτουργιών, επηρεάζει τη διάδοση του ραδιοφώνου σε μακρινά μέρη της Γης.

#### 3.3.2 Τα στρώματα της ιονόσφαιρας και τα χαρακτηριστικά τους

---

Υπάρχουν τρεις κύριες περιοχές της ιονόσφαιρας, που ονομάζονται στρώμα D, στρώμα E και στρώμα F. Αυτές οι περιοχές δεν έχουν έντονα όρια και τα υψόμετρα στα οποία εμφανίζονται ποικίλλουν κατά τη διάρκεια μιας ημέρας και από εποχή σε εποχή. Η περιοχή D είναι η χαμηλότερη, ξεκινά περίπου 60 ή 70 km (37 ή 43 μίλια) πάνω από το έδαφος και εκτείνεται προς τα πάνω σε περίπου 90 km (56 μίλια). Στη συνέχεια υψηλότερη είναι η περιοχή A, ξεκινώντας από περίπου 90 ή 100 km (56 ή 62 μίλια) και εκτείνεται στα 120 ή 150 km (75 ή 93 μίλια). Το ανώτερο τμήμα της ιονόσφαιρας, η περιοχή F, ξεκινά περίπου 150 km (93 μίλια) και εκτείνεται πολύ προς τα πάνω, μερικές φορές έως και 500 km (311 μίλια) πάνω από την επιφάνεια του πλανήτη μας.

Οι περιοχές της ιονόσφαιρας δεν θεωρούνται ξεχωριστά στρώματα, όπως η πιο οικεία τροπόσφαιρα και στρατόσφαιρα. Αντίθετα, είναι ιονισμένες περιοχές ενσωματωμένες στα τυπικά ατμοσφαιρικά στρώματα. Η περιοχή D συνήθως σχηματίζεται στο πάνω μέρος της μεσόσφαιρας, ενώ η περιοχή E εμφανίζεται συνήθως στην κάτω θερμόσφαιρα και η περιοχή F βρίσκεται στο ανώτερο τμήμα της θερμόσφαιρας.

Τη νύχτα το στρώμα F είναι το μόνο στρώμα σημαντικού ιονισμού που υπάρχει, ενώ ο ιονισμός στα στρώματα E και D είναι εξαιρετικά χαμηλός. Κατά τη διάρκεια της ημέρας, τα στρώματα D και E ιονίζονται πολύ πιο έντονα, όπως και το στρώμα F, το οποίο αναπτύσσει μια πρόσθετη, ασθενέστερη περιοχή ιονισμού γνωστή ως στρώμα F1. Το στρώμα F2 παραμένει μέρα και νύχτα και είναι η κύρια περιοχή που είναι υπεύθυνη για τη διάθλαση και την ανάκλαση των ραδιοκυμάτων.

- **D στρώμα**

Το στρώμα D είναι το πιο εσωτερικό στρώμα, 48 km (30 mi) έως 90 km (56 mi) πάνω από την επιφάνεια της Γης. Ο ιονισμός εδώ οφείλεται στην ακτινοβολία υδρογόνου σειράς άλφα Lyman σε μήκος κύματος 121,6 νανόμετρων (nm) ιονίζοντας μονοξειδίου του αζώτου (NO). Επιπλέον, η υψηλή ηλιακή δραστηριότητα μπορεί να δημιουργήσει σκληρές ακτίνες X (μήκος κύματος < 1 nm) που ιονίζουν N<sub>2</sub> και O<sub>2</sub>. Οι ρυθμοί ανασυνδυασμού είναι υψηλοί στο στρώμα D, επομένως υπάρχουν πολύ περισσότερα ουδέτερα μόρια αέρα από ιόντα.

Τα ραδιοκύματα μέσης συχνότητας (MF) και χαμηλότερης υψηλής συχνότητας (HF) εξασθενούν σημαντικά εντός του στρώματος D, καθώς τα ραδιοκύματα που περνούν προκαλούν την κίνηση των ηλεκτρονίων, τα οποία στη συνέχεια συγκρούονται με τα ουδέτερα μόρια, δίνοντας την ενέργειά τους. Οι χαμηλότερες συχνότητες έχουν μεγαλύτερη απορρόφηση επειδή μετακινούν τα ηλεκτρόνια μακρύτερα, οδηγώντας σε μεγαλύτερες πιθανότητες συγκρούσεων. Αυτός είναι ο κύριος λόγος για την απορρόφηση των ραδιοκυμάτων HF, ιδιαίτερα στα 10 MHz και κάτω, με προοδευτικά μικρότερη απορρόφηση σε υψηλότερες συχνότητες. Αυτό το φαινόμενο κορυφώνεται γύρω στο μεσημέρι και μειώνεται τη νύχτα λόγω της μείωσης του πάχους του στρώματος D. μόνο ένα μικρό μέρος παραμένει λόγω των κοσμικών ακτίνων. Ένα συνηθισμένο παράδειγμα του επιπέδου D σε δράση είναι η εξαφάνιση των απομακρυσμένων σταθμών ζώνης εκπομπής AM κατά τη διάρκεια της ημέρας.

Κατά τη διάρκεια γεγονότων ηλιακών πρωτονίων, ο ιονισμός μπορεί να φτάσει σε ασυνήθιστα υψηλά επίπεδα στην περιοχή D σε μεγάλα και πολικά γεωγραφικά πλάτη. Τέτοια πολύ σπάνια συμβάντα είναι γνωστά ως συμβάντα απορρόφησης πολικού πώματος (ή PCA), επειδή ο αυξημένος ιονισμός ενισχύει σημαντικά την απορρόφηση των ραδιοφωνικών σημάτων που διέρχονται από την περιοχή. Στην πραγματικότητα, τα επίπεδα απορρόφησης μπορούν να αυξηθούν κατά πολλές δεκάδες dB κατά τη διάρκεια έντονων γεγονότων, κάτι που είναι αρκετό για να απορροφήσει τις περισσότερες (αν όχι όλες) διαπολικές μεταδόσεις ραδιοφωνικού σήματος HF. Τέτοια συμβάντα διαρκούν συνήθως λιγότερο από 24 έως 48 ώρες.

- **E στρώμα**

Το στρώμα E είναι το μεσαίο στρώμα, 90 km (60 mi) έως 150 km (90 mi) πάνω από την επιφάνεια της Γης. Ο ιονισμός οφείλεται στον ιονισμό του μοριακού οξυγόνου (O<sub>2</sub>) με μαλακές ακτίνες X (1–10 nm) και μακρινή υπεριώδη (UV) ηλιακή ακτινοβολία. Κανονικά, σε λοξή πρόσπτωση, αυτό το στρώμα μπορεί να αντανακλά μόνο ραδιοκύματα που έχουν συχνότητες χαμηλότερες από περίπου 10 MHz και μπορεί να συμβάλει λίγο στην απορρόφηση στις παραπάνω συχνότητες. Ωστόσο, κατά τη διάρκεια έντονων σποραδικών συμβάντων E, το στρώμα Es μπορεί να αντανακλά συχνότητες έως και 50 MHz και υψηλότερες. Η κατακόρυφη δομή του στρώματος E καθορίζεται κυρίως από τα ανταγωνιστικά αποτελέσματα του ιονισμού και του ανασυνδυασμού. Τη νύχτα το στρώμα E εξασθενεί επειδή η κύρια πηγή ιονισμού δεν είναι πλέον παρούσα. Μετά τη



δύση του ηλίου, μια αύξηση στο ύψος του μέγιστου στρώματος E αυξάνει το εύρος στο οποίο τα ραδιοκύματα μπορούν να ταξιδέψουν με ανάκλαση από το στρώμα.

Αυτή η περιοχή είναι επίσης γνωστή ως στρώμα Kennelly–Heaviside ή απλά στρώμα Heaviside. Η ύπαρξή του προβλέφθηκε το 1902 ανεξάρτητα και σχεδόν ταυτόχρονα από τον Αμερικανό ηλεκτρολόγο μηχανικό Arthur Edwin Kennelly (1861–1939) και τον Βρετανό φυσικό Oliver Heaviside (1850–1925). Το 1924 ότι η ύπαρξή του εντοπίστηκε από τους Edward V. Appleton και Miles Barnett.

- **Es στρώμα**

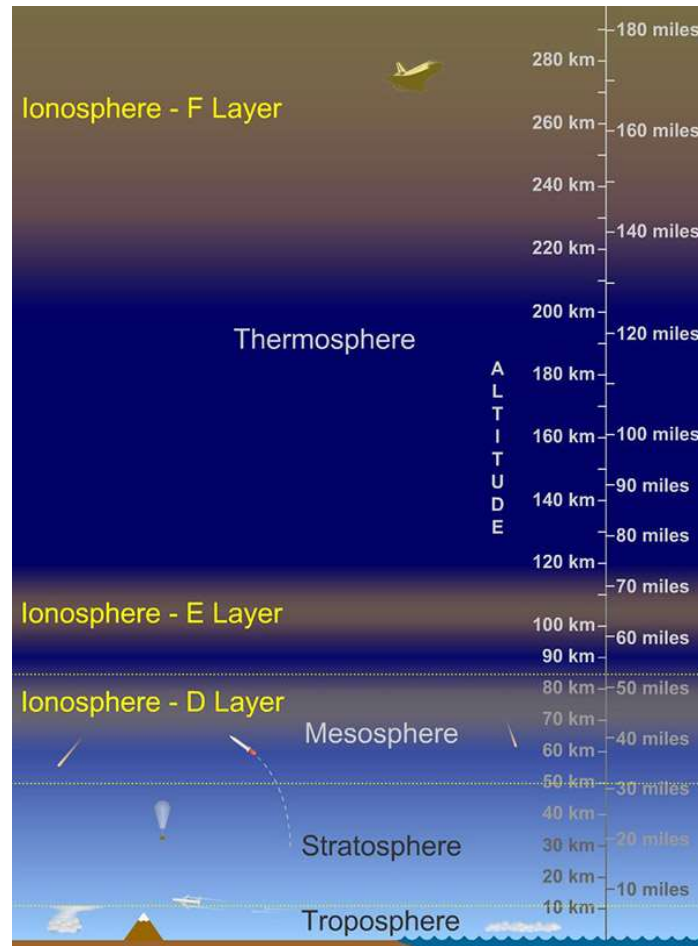
Το στρώμα Es (σποραδικό στρώμα E) χαρακτηρίζεται από μικρά, λεπτά νέφη έντονου ιονισμού, τα οποία μπορούν να υποστηρίξουν την ανάκλαση ραδιοκυμάτων, συχνά έως 50 MHz και σπάνια έως 450 MHz. Τα σποραδικά συμβάντα μπορεί να διαρκέσουν από λίγα λεπτά έως πολλές ώρες. Η σποραδική διάδοση E καθιστά τη λειτουργία VHF από ραδιοερασιτέχνες πολύ συναρπαστική όταν οι διαδρομές διάδοσης μεγάλων αποστάσεων που είναι γενικά απρόσιτες «ανοίγουν» σε αμφίδρομη επικοινωνία. Υπάρχουν πολλές αιτίες του σποραδικού-E που εξακολουθούν να αναζητούνται από τους ερευνητές. Αυτή η διάδοση συμβαίνει καθημερινά τον Ιούνιο και τον Ιούλιο στα μεσαία γεωγραφικά πλάτη του βόρειου ημισφαιρίου, όταν συχνά επιτυγχάνονται υψηλά επίπεδα σήματος. Οι αποστάσεις παράλειψης είναι γενικά περίπου 1.640 km (1.020 mi). Οι αποστάσεις για μια διάδοση λυκίσκου μπορεί να είναι οπουδήποτε από 900 km (560 mi) έως 2.500 km (1.600 mi). Η διάδοση πολλαπλών άλματος πάνω από 3.500 km (2.200 mi) είναι επίσης κοινή, μερικές φορές σε αποστάσεις 15.000 km (9.300 mi) ή περισσότερες.

- **F Στρώμα**

Το στρώμα ή η περιοχή F, επίσης γνωστή ως στρώμα Appleton–Barnett, εκτείνεται από περίπου 150 km (90 mi) σε περισσότερα από 500 km (300 mi) πάνω από την επιφάνεια της Γης. Είναι το στρώμα με την υψηλότερη πυκνότητα ηλεκτρονίων, που σημαίνει ότι τα σήματα που διεισδύουν σε αυτό το στρώμα θα διαφύγουν στο διάστημα. Η παραγωγή ηλεκτρονίων κυριαρχείται από την ακραία υπεριώδη ακτινοβολία (UV, 10–100 nm) που ionίζει ατομικό οξυγόνο. Το στρώμα F αποτελείται από ένα στρώμα (F2) τη νύχτα, αλλά κατά τη διάρκεια της ημέρας, συχνά σχηματίζεται μια δευτερεύουσα κορυφή (με την ένδειξη F1) στο προφίλ πυκνότητας ηλεκτρονίων. Επειδή το στρώμα F2 παραμένει τη μέρα και τη νύχτα, είναι υπεύθυνο για την πλειονότητα της διάδοσης των κυμάτων ουρανού των ραδιοκυμάτων και των ραδιοεπικοινωνιών υψηλής συχνότητας μεγάλων αποστάσεων (HF ή βραχέων κυμάτων).

Πάνω από το στρώμα F, ο αριθμός των ιόντων οξυγόνου μειώνεται και τα ελαφρύτερα ιόντα όπως το υδρογόνο και το ήλιο γίνονται κυρίαρχα. Αυτή η περιοχή πάνω από την κορυφή του στρώματος F και κάτω από την πλασμάσφαιρα ονομάζεται ιονόσφαιρα από την κορυφή.





Εικόνα 5. Τα στρώματα της ιονόσφαιρας.

### 3.4 Οι τιμές TEC (Total Electron Content)

#### 3.4.1 Εισαγωγικές έννοιες

Η συνθήκη ιονόσφαιρας είναι ένα από τα πιο σημαντικά χαρακτηριστικά του διαστημικού καιρού για τους χρήστες συστημάτων που βασίζονται σε ραδιόφωνα, όπως συστήματα πλοήγησης που βασίζονται στο Παγκόσμιο Δορυφορικό Σύστημα Πλοήγησης (GNSS), συστήματα επικοινωνίας υψηλής συχνότητας (HF) και τηλεπισκόπηση με βάση το διάστημα συστήματα. Τα ραδιοκύματα που διαδίδονται στην ιονόσφαιρα παρουσιάζουν καθυστέρηση στην ομαδική ταχύτητα και προχωρούν στην ταχύτητα φάσης λόγω των ηλεκτρονίων στην ιονόσφαιρα. Η καθυστέρηση της ιονόσφαιρας είναι ανάλογη με το συνολικό περιεχόμενο ηλεκτρονίων της ιονόσφαιρας (TEC) [22], κατά μήκος της διαδρομής διάδοσης, και αντιστρόφως ανάλογη του τετραγώνου της συχνότητας  $f$  του φορέα (L1 ή L2). Ο ευκολότερος τρόπος για τη διόρθωση της

ιονόσφαιρικής καθυστέρησης είναι η χρήση μοντέλων καθυστέρησης εκπομπής ιονόσφαιρας που βασίζονται σε απλά εμπειρικά μοντέλα TEC όπως τα μοντέλα Klobuchar (1987) και NeQuick (Hochegger et al. 2000, Radicella and Leitinger 2001).

### 3.4.2 Βασικές αρχές

---

Η τιμή TEC καθορίζεται από πολλούς παράγοντες, όπως η ηλιακή δραστηριότητα, η εποχή, η τοπική ώρα και η γεωμαγνητική δραστηριότητα. Υπάρχει επίσης γεωγραφική εξάρτηση στις παραλλαγές TEC [22]. Οι διακυμάνσεις του TEC που προκαλούνται από την ηλιακή δραστηριότητα, την εποχή και την τοπική ώρα μπορούν να εκτιμηθούν χρησιμοποιώντας αυτά τα απλά μοντέλα, αλλά αυτές που προκαλούνται από γεωμαγνητικές καταιγίδες και άλλα φαινόμενα δεν μπορούν να αφαιρεθούν πλήρως από αυτά τα μοντέλα. Ως εκ τούτου, οι χρήστες συστημάτων που βασίζονται σε ραδιόφωνα ενδέχεται να επηρεαστούν από θετικές ή/και αρνητικές ιονοσφαιρικές καταιγίδες. Κατά τη διάρκεια αρνητικών ιονοσφαιρικών καταιγίδων, το TEC είναι  $\geq 0$  TECU ακόμα κι αν η αρνητική καταιγίδα είναι εξαιρετικά σοβαρή. Από την άλλη πλευρά, οι ακραίες τιμές TEC κατά τη διάρκεια θετικών καταιγίδων δεν είναι άγνωστες και πρέπει να μελετηθούν. Μπορεί να μετρηθεί είτε κατακόρυφα είτε λοξά και χαρακτηρίζεται ως VTEC ή STEC αντίστοιχα.

Για το σχεδιασμό και τη λειτουργία συστημάτων που ενδέχεται να επηρεαστούν από διαστημικά καιρικά φαινόμενα, είναι σημαντικό να γνωρίζουμε την πιθανή έκταση της πρόσκρουσης και πόσο συχνά είναι πιθανό να συμβούν τέτοια γεγονότα. Επομένως, είναι σημαντικό να μελετηθούν ακραίες τιμές που σχετίζονται με διάφορα διαστημικά καιρικά φαινόμενα. Για χρήστες ραδιοϊονόσφαιρων συστημάτων, η ακραία τιμή TEC είναι μια βασική τιμή.

Έχουν μελετηθεί ακραίες τιμές ορισμένων παραμέτρων του διαστημικού καιρού. Για παράδειγμα, αυτός του δείκτη Dst διερευνήθηκε χρησιμοποιώντας μοντελοποίηση ακραίων τιμών (Tsubouchi και Omura 2007). Αυτά της ροής ακτίνων X της ηλιακής έκλαμψης, της ταχύτητας εκτόξευσης μάζας στέμματος, του δείκτη Dst και της ενέργειας πρωτονίων σε γεγονότα πρωτονίων μελετήθηκαν από τον Riley (2012) χρησιμοποιώντας συμπληρωματικές συναρτήσεις αθροιστικής κατανομής. Ωστόσο, οι ακραίες τιμές TEC μιας φοράς ανά μεγάλο χρονικό διάστημα δεν έχουν ακόμη εκτιμηθεί ποσοτικά. Αρκετές χώρες έχουν ετοιμάσει έγγραφα με κριτήρια αναφοράς καιρού για το διάστημα. Ο Λευκός Οίκος των ΗΠΑ δημοσίευσε τα «Διαστημικά καιρικά κριτήρια φάσης 1» τον Ιούνιο του 2018 (Λευκός Οίκος ΗΠΑ 2018). Αν και απαριθμεί τρεις παράγοντες που προκαλούν ιονοσφαιρικές διαταραχές, όπως οι γεωμαγνητικές καταιγίδες, δεν δόθηκαν ποσοτικά κριτήρια αναφοράς επειδή οι ιονόσφαιρες επιδράσεις των γεωμαγνητικών καταιγίδων στην ιονόσφαιρα διαφέρουν σε μεγάλο βαθμό από γεγονός σε γεγονός και ακόμη και ο μηχανισμός τους δεν είναι πλήρως κατανοητός.

Ένας άλλος λόγος για τον οποίο οι ακραίες τιμές TEC δεν έχουν μελετηθεί πλήρως είναι ότι έχουν περάσει μόνο 20 χρόνια από την έναρξη των πλήρως εξελιγμένων παρατηρήσεων TEC. Οι παρατηρήσεις TEC ξεκίνησαν με μετρήσεις της περιστροφής Faraday ή του φαινομένου Doppler πριν από πολλές. Δεδομένου ότι αυτές οι παρατηρήσεις πραγματοποιήθηκαν από λίγους πομπούς και δέκτες, είναι δύσκολο να μελετηθεί η συμπεριφορά του TEC στατιστικά. Με την εξάπλωση του GNSS και των επίγειων δεκτών του, ο αριθμός των παρατηρήσεων TEC αυξήθηκε δραματικά. Χάρη στα συστήματα παρατήρησης GNSS-TEC, μάθαμε πολλά για τη συμπεριφορά του TEC τα τελευταία 20 χρόνια. Ο σκοπός αυτής της μελέτης είναι να εκτιμήσει τις ακραίες τιμές του TEC με τα ποσοστά εμφάνισής τους. Διερευνούμε τα ποσοστά εμφάνισης ακραίων τιμών TEC στην Ιαπωνία βραχυπρόθεσμα, μεσοπρόθεσμα και μακροπρόθεσμα, τα οποία είναι μία φορά το χρόνο, 10 χρόνια και 100 χρόνια, αντίστοιχα.

Για να αξιολογηθεί το TEC που αντιστοιχεί σε ένα ποσοστό εμφάνισης μία φορά ανά 100 χρόνια, τα δεδομένα 20 ετών είναι προφανώς ανεπαρκή. Επιπλέον, η ηλιακή δραστηριότητα τα τελευταία 20 χρόνια ήταν κατά μέσο όρο μέτρια, αν και εμφανίστηκαν αρκετές έντονες γεωμαγνητικές καταιγίδες κατά τη διάρκεια του ηλιακού κύκλου 24. Σε σύγκριση με την παρατήρηση GNSS-TEC, η παρατήρηση ιονοσόνδης έχει πολύ μεγαλύτερη ιστορία. Αυτή η τεχνική αναπτύχθηκε στα τέλη της δεκαετίας του 1920 και άρχισε να εφαρμόζεται στη δεκαετία του 1940 για την παρακολούθηση της διάδοσης βραχέων κυμάτων (Gladden 1959). Στην Ιαπωνία, η παρατήρηση ιονοσόνδης ξεκίνησε το 1931. Μετά από διάφορες αλλαγές, η συνήθης παρατήρηση ιονοσόνδης ξεκίνησε από τον προκάτοχο του Εθνικού Ινστιτούτου Τεχνολογίας Πληροφοριών και Επικοινωνιών (NICT) το 1951 χρησιμοποιώντας ένα αυτόματο σύστημα. Οι ιονόσφαιρες παράμετροι που προέρχονται από τη μακροπρόθεσμη παρατήρηση ιονοσόνδης αρχειοθετούνται από το Παγκόσμιο Κέντρο Δεδομένων για την Ιονόσφαιρα στο NICT. Τα μακροπρόθεσμα δεδομένα ιονοσόνδης έχουν χρησιμοποιηθεί για διάφορες μελέτες, όπως μια μελέτη των μακροπρόθεσμων τάσεων της ιονόσφαιρας (Xu et al. 2004) και για την ανάπτυξη εμπειρικών μοντέλων (Bilitza 2018; Yue et al. 2006; Maruyama 2011). Καθώς το TEC και η μέγιστη πυκνότητα της περιοχής F που προέρχεται από την παρατήρηση ιονοσόνδης ( $NmF2$ ) είναι γνωστό ότι συσχετίζονται, το  $NmF2$  μπορεί να είναι ένας αντιπρόσωπος του TEC. Σε αυτή τη μελέτη, χρησιμοποιούνται δεδομένα περίπου 60 ετών ιονοσφαιρικών παραμέτρων που προέρχονται από τη μακροχρόνια παρατήρηση ιονοσόνδης. Αν και η περίοδος δεδομένων είναι ακόμη μικρότερη από 100 χρόνια, διερευνούμε στατιστικά χαρακτηριστικά των ακραίων τιμών TEC προκειμένου να εκτιμήσουμε την ιονόσφαιρα μία φορά ανά 100 χρόνια.

### 3.5 Παράμετροι επιρροής τιμών Tec

#### 3.5.1 Εισαγωγικές έννοιες

Οι παράμετροι που επηρεάζουν την ιονόσφαιρα παρατίθενται στον Πίνακα 1.

Πίνακας 1. Παράμετροι επιρροής της ιονόσφαιρας.

A/A	Δραστηριότητα	Παράμετρος	Υποπαράμετρος	Περίοδος καταγραφής
1	Μαγνητική δραστηριότητα	Daily geomagnetic activity average (K)	Kp	Ανά 3 ώρες
2			ap	
3			D	
4		Auroral Electrojet (AE)	SML	
5			SMU	
6		Disturbance storm time (Dst)	-	
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 Electron FluEncE >0.8 MEV	

27			GOES-15 EIElectron FluEncE >2 MEV	
28			NEutron Monitor % of bkgd	
29		Solar Flares	-	Κατά την εμφάνιση
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	

### 3.5.2 Μαγνητική κ Ιονοσφαιρική Καταιγίδα

Απότομες μεταβολές του αριθμού των ηλεκτρονίων κατά μήκος της διαδρομής του σήματος προκαλούν προβλήματα στα κυκλώματα PPL και συνεπώς πολλές απώλειες λήψης του σήματος ή παρουσία κύκλων ολίσθησης, με αποτέλεσμα τη δυσκολία του σωστού προσδιορισμού των ακέραιων ασαφειών φάσης κατά τη συνόρθωση των παρατηρήσεων. Το φαινόμενο αυτό της ιονοσφαιρικής έξαρσης (ionospheric scintillation) που συμβαίνει κυρίως στις πολικές περιοχές και σε μία ζώνη  $\pm 30^\circ$  γύρω από τον γεωμαγνητικό ισημερινό διαρκεί έως μερικά λεπτά και μπορεί να συνεχίζεται ανά διαστήματα για αρκετές ώρες. Οι παράγοντες που μπορούν να το προκαλέσουν είναι συνήθως η έντονη δραστηριότητα της ηλιακής δραστηριότητας, που προκαλεί την μαγνητική και ιονοσφαιρική καταιγίδα. Οι καταιγίδες αυτές αυξάνουν κατά πολύ τον αριθμό των ηλεκτρονίων, κατά καν'να στη διάρκεια του χειμώνα ενώ αντιθέτως μειώνουν το TEC στη διάρκεια του καλοκαιριού. Το φαινόμενο της ιονοσφαιρικής έξαρσης αποτελεί κατά κάποιον τρόπο ένα είδος πολυανάκλασης του δορυφορικού σήματος.

### 3.5.3 Ηλιακές Εκλάμψεις (Solar Flares)

---

Οι ηλιακές εκλάμψεις είναι ξαφνικές εκρήξεις ακτινοβολιών που προέρχονται από τον ήλιο, οι οποίες μεταβάλλουν σημαντικά διάφορες φυσικές και χημικές διεργασίες στην ανώτερη ατμόσφαιρα της γης. Έχει γίνει σημαντικός αριθμός μελετών στο παρελθόν, για την κατανόηση των επιπτώσεων της ηλιακής έκλαμψης στην ιονόσφαιρα (π.χ. Le et al. 2013 ) και στη θερμοσφαιρική (π.χ. Le et al. 2012, 2015· Pawlowski and Ridley 2008, 2011) υψόμετρα. Πλήθος δεδομένων από επίγειες/διαστημικές πλατφόρμες που χρησιμοποιούν κατά κύριο λόγο GPS TEC (Afraimovich 2000, Leonovich et al. 2002, 2010; Liu et al. 2006), ασυνάρτητα ραντάρ διασποράς (Mendillo and Evans 1974) σε συνδυασμό με το Themeberticallinetore al. 2008· Qian et al. 2010) έχουν παράσχει σημαντικά αποτελέσματα σχετικά με τις επιδράσεις που προκαλούνται από εκλάμψεις στην ανώτερη γήινη ατμόσφαιρα.

Πρόσφατα, τα φαινόμενα ηλιακής έκλαμψης και οι επιπτώσεις τους στην ηλεκτροδυναμική της ισημερινής ιονόσφαιρας έχουν λάβει σημαντική έλξη σε όλη την υδρόγειο (π.χ. Zhang et al. 2017· Manju 2016· Manju et al. 2012). Οι Liu et al. (2007) έδειξε ότι η ανωμαλία ιοντισμού του Ισημερινού (EIA) εξασθενούσε σημαντικά κατά τη διάρκεια της έκλαμψης λόγω του κυρίαρχου ρόλου της ηλεκτροδυναμικής παρά της φωτοχημείας. Οι μελέτες προσομοίωσης (Qian et al. 2012) έδειξαν επίσης ότι η μετατόπιση  $E \times B$  στον ισημερινό εξασθενεί κατά τη διάρκεια της έκλαμψης. Αυτό είναι σύμφωνο με τις προηγούμενες παρατηρήσεις μείωσης του ηλεκτρικού πεδίου προς τα ανατολικά στην περιοχή του δυναμό κατά τη διάρκεια της έκλαμψης (Manju και Viswanathan 2005). Ωστόσο, στο παρελθόν είχε αναφερθεί τόσο αύξηση (Qian et al. 2012) όσο και μείωση (Manju και Viswanathan 2005) στην αναλογία αγωγιμότητας Hall-to-Pedersen ενσωματωμένη στη γραμμή πεδίου σε σχέση με τον ισημερινό. Από την άλλη πλευρά, οι επιπτώσεις της ηλιακής έκλαμψης στην ουδέτερη θερμοσφαιρα, ιδιαίτερα στα γεωγραφικά πλάτη του ισημερινού, δεν έχουν γίνει σωστά κατανοητές μέχρι στιγμής, κυρίως λόγω της έλλειψης παρατηρήσεων. Αν και οι παρατηρήσεις CHALLENGING Mini Satellite Payload (CHAMP) παρείχαν αρκετές νέες γνώσεις σχετικά με αυτό (Liu et al. 2007), κανονικοποιούνται σε υψηλότερο υψόμετρο 400 km και έχουν κακή χρονική ανάλυση (~93 λεπτά) λόγω του εγγενούς χρόνου ώστε ο δορυφόρος να επανεξετάσει το ίδιο γεωγραφικό πλάτος και τοπική ώρα. Επομένως, η ανάγκη για δεδομένα υψηλού ρυθμού από την κατώτερη θερμοσφαιρα εκτιμάται καλά για την ολοκληρωμένη κατανόηση της ιονόσφαιρας-θερμοσφαιρικής σύζευξης κατά τη διάρκεια τέτοιων παροδικών γεγονότων.

Σε αυτό το πλαίσιο, οι επίγειες ουδέτερες θερμοσφαιρικές εκπομπές ημέρας λάμψης OI 630,0 nm είναι ιδανικές για τη διερεύνηση τέτοιων εφήμερων γεγονότων όπως η έκλαμψη, καθώς προέρχονται από τα υψόμετρα, όπου λαμβάνει χώρα η μέγιστη απορρόφηση EUV. Αυτές οι μετρήσεις λάμψης ημέρας στο Trivandrum, τα τελευταία χρόνια, έφεραν πολλά σημαντικά αποτελέσματα που σχετίζονται με διάφορες πτυχές των διαδικασιών σύζευξης στην ανώτερη ατμόσφαιρα του ισημερινού (π.χ., Sumod et al.

2011a, b, 2012, 2014, 2015 και αναφορές σε αυτό). Στην παρούσα εργασία, γίνεται μια προσπάθεια να διερευνηθούν οι επιπτώσεις της ηλιακής έκλαμψης X3.8, που συνέβη στις 17 Ιανουαρίου 2005, στο σύστημα ισημερινής ιονόσφαιρας-θερμόσφαιρας, χρησιμοποιώντας συνδυασμένες μετρήσεις υψηλού ρυθμού θερμοσφαιρικής λάμψης ημέρας ΟΙ 630,0 nm, με μέτρηση GPS. ιονόσφαιρα Ολικό Περιεχόμενο Ηλεκτρονίων (TEC) και η ισχύς του Ισημερινού Ηλεκτρονίου (EEJ). Η μελέτη παρέχει ίσως την «πρώτη φορά» πειραματική απόδειξη για την αποδυνάμωση της ΕΙΑ κατά τη διάρκεια της έκρηξης στον ινδικό διαμήκη τομέα.

### 3.5.4 Coronal mass ejection

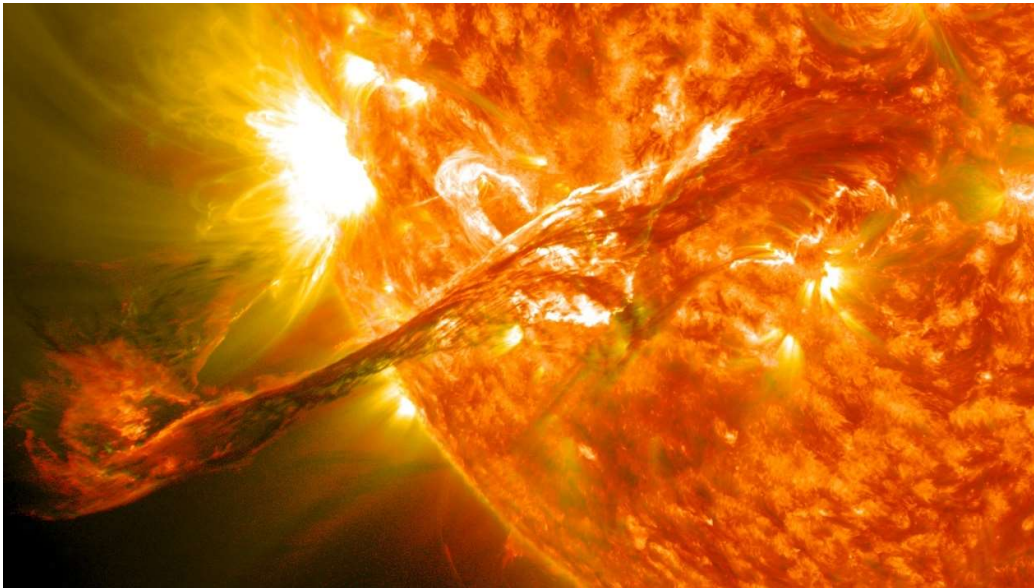
---

Αρκετές μελέτες ενδιαφέρθηκαν να διερευνήσουν τον αντίκτυπο των ηλιακών δραστηριοτήτων στο ιονόσφαιρα. Μεταξύ εκείνων των μελετών που ασχολούνται με την απόκριση της ιονόσφαιρας Η κρίσιμη συχνότητα ( $f_{oF2}$ ) στις CME, έχει γίνει υψίστης σημασίας τα τελευταία χρόνια. Στην πραγματικότητα, η ηλιακή δραστηριότητα είναι η κύρια πηγή διαταραχών και διακυμάνσεις στο περιβάλλον της Γης, ιδιαίτερα στη μαγνητόσφαιρα και στα στρώματα της ιονόσφαιρας.

Η μεταβλητότητα της ιονόσφαιρας μπορεί να αποδοθεί σε συνεισφορές από το χαμηλότερο εσωτερικό της ατμόσφαιρας διακυμάνσεις κυμάτων, γεωμαγνητικής και ηλιακής δραστηριότητας από την υψηλή ατμόσφαιρα επίσης. Στην πραγματικότητα, ηλιακός ιονιστής η ροή ποικίλλει όχι μόνο με μεγαλύτερη χρονική κλίμακα, τον ηλιακό κύκλο, αλλά και με τη μικρότερη χρονική κλίμακα, η σχεδόν 27-ημέρες περιστροφή του Ήλιου και μάλιστα σε καθημερινή βάση. Επιπλέον, προκαλείται από ηλιακή ροή Οι διακυμάνσεις της ουδέτερης θερμοκρασίας, των ανέμων και της ουδέτερης σύνθεσης εκδηλώνονται και στην ιονόσφαιρα πυκνότητες και ύψη πλάσματος.

Τα CME πιστεύεται ότι προκαλούν αύξηση της πυκνότητας ηλεκτρονίων σε μεγάλους όγκους της Γης ιονόσφαιρα. Επιπλέον, τα συμβάντα CME είναι συνήθως η πηγή έντονων γεωμαγνητικών καταιγίδων και εμφανίζονται κυρίως κατά τη διάρκεια της ηλιακής μέγιστης φάσης. Γι' αυτό θεωρούνται ότι είναι η προέλευση των διαστημικών καιρικών φαινομένων. Βασικά, τα CME είναι τεράστιες εκρήξεις πλάσματος και μαγνητικών πεδίων από το στέμμα του ήλιου. Εκτοξεύουν δισεκατομμύρια τόνους στεφανιαίου υλικού που έχει παγώσει στο μαγνητικό πεδίο που είναι μεγαλύτερα από το μαγνητικό πεδίο του ηλιακού ανέμου φόντου (IMF). Επιπλέον, είναι ταξιδεύοντας προς τα έξω από τον ήλιο με ταχύτητες που κυμαίνονται από 250 (km/s) έως και κοντά στα 3000 km/s.





Εικόνα 6. Έκρηξη CME στις 31 Αυγούστου 2012 (Πηγή: NASA Goddard Space Flight Center)

### 3.5.5 Solar Radio Flux

---

Κατά την περίοδο 2001-2010, αρκετές ισχυρές γεωμαγνητικές καταιγίδες και άμεσες παρεμβολές Solar Radio Flux επιδείνωσαν σοβαρές επιδόσεις GPS. Η εκπομπή Solar Radio Flux της ζώνης L έχει πρόσφατα θεωρηθεί ως πιθανή απειλή για τη σταθερή απόδοση του GPS και του GLONASS. Ωστόσο, η απειλή δεν έχει διερευνηθεί ή αξιολογηθεί πλήρως μέχρι στιγμής. Επιπλέον, η ανωμαλία ιονισμού σε χαμηλά γεωγραφικά πλάτη μαζί με την επίδραση των «φουσαλίδων» του ισημερινού πλάσματος αυξάνουν την πιθανότητα εξασθένησης των μεταιονοσφαιρικών σημάτων, ειδικά κατά τη διάρκεια γεωμαγνητικών καταιγίδων. Οι αστάθειες του ιονοσφαιρικού πλάσματος στα «τοιχώματα» μιας φουσαλίδας με πυκνότητα ηλεκτρονίων μικρότερη από την τιμή του υποβάθρου χαρακτηρίζονται επίσης από έντονες διαβαθμίσεις της πυκνότητας ηλεκτρονίων. Για παράδειγμα, τα τοιχώματα των φουσαλίδων μπορούν επίσης να είναι πηγή ιονοσφαιρικών σπινθηρισμών. Επομένως, ο χωρικός προσανατολισμός της φουσαλίδας πλάσματος παίζει καθοριστικό ρόλο στην ενίσχυση και την αποδυνάμωση του πλάτους και των σπινθηρισμών φάσης των σημάτων δορυφορικών οχημάτων (SV).

Αν και οι αστοχίες εμβέλειας και εντοπισμού θέσης GPS διερευνήθηκαν κάτω από τέτοιες δυσμενείς γεωφυσικές συνθήκες, η ακριβής φύση τους παραμένει ασαφής. Αυτό το κεφάλαιο είναι αφιερωμένο σε ορισμένα χαρακτηριστικά της απόδοσης GPS/GLONASS κάτω από συγκεκριμένα γεωφυσικά γεγονότα, όπως εκρήξεις ηλιακών ραδιοφωνικών εκπομπών και ισχυρή σκέδαση δορυφορικού σήματος από τις «φουσαλίδες» του ισημερινού πλάσματος. Καθώς οι ανωμαλίες επιμηκύνονται στην κατεύθυνση του μαγνητικού πεδίου, μια διαδρομή ακτινών που είναι παράλληλη με το μαγνητικό πεδίο ενσωματώνει μεγαλύτερο μήκος διαδρομής μέσω εξαντλημένων

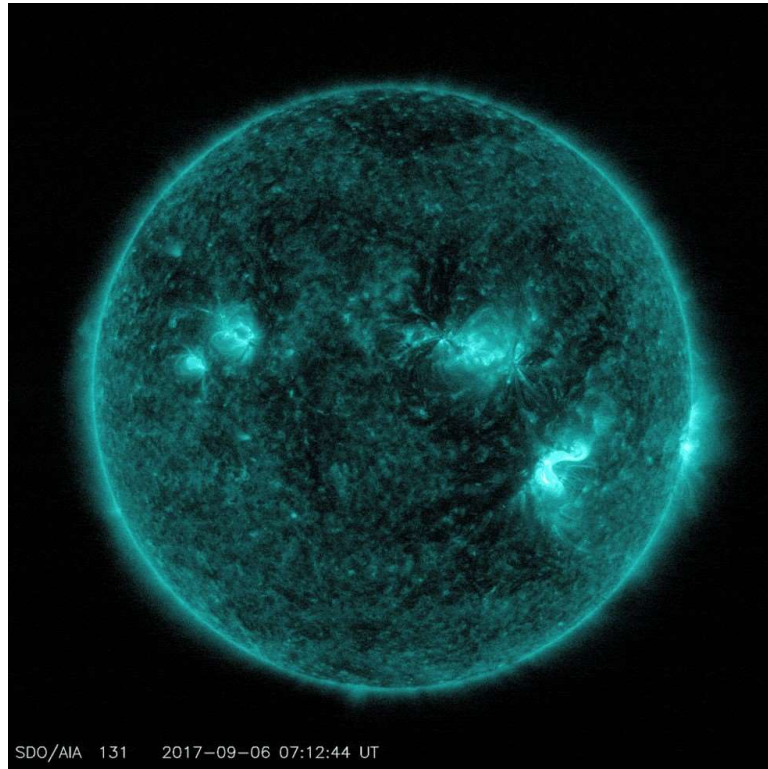
περιοχών που περιέχουν ανωμαλίες στην πυκνότητα του πλάσματος. Προτείνουμε μια μέθοδο GPS για την ανίχνευση ανωμαλιών ευθυγραμμισμένων με το μεσαίο γεωγραφικό πλάτος (FAIs) με γωνιακή σάρωση οπτικής επαφής σχετικά με το διάνυσμα τοπικού μαγνητικού πεδίου. Χρησιμοποιώντας δεδομένα GPS του ιαπωνικού δικτύου GPS (GEONET), αναλύουμε την εμφάνιση ολισθήσεων φάσης GPS και σφαλμάτων θέσης κατά τη διάρκεια της γεωμαγνητικής καταιγίδας της 12ης Φεβρουαρίου 2000.

### 3.5.6 Sunspot Number

---

Σε μια πρόσφατη μελέτη, οι ιονόσφαιρες παράμετροι μεσαίου γεωγραφικού πλάτους συγκρίθηκαν με την ηλιακή δραστηριότητα. προτάθηκε ότι η σχέση μεταξύ αυτών, που προηγουμένως θεωρούνταν σταθερή, μπορεί να αλλάζει με το χρόνο (Lastovicka, 2019). Εδώ, οι πληροφορίες επεκτείνονται σε μεγαλύτερο γεωγραφικό πλάτος (69,6°N, 19,2E) και πιο πίσω στο χρόνο. Για την ιονόσφαιρα F-περιοχή (δηλ. η κρίσιμη συχνότητα, FoF2) η ίδια συμπεριφορά παρατηρείται με ένα σημείο αλλαγής γύρω στο 1996. Για την ιονόσφαιρα E-περιοχή (δηλ. η κρίσιμη συχνότητα, foE), τα σημεία αλλαγής είναι λιγότερο εμφανή παρά στη μελέτη μεσαίου γεωγραφικού πλάτους, πιθανώς λόγω της θέσης παρατήρησης που βρίσκεται κάτω από το οβάλ σέλας.

Με μια ολοένα και πιο μεγάλη χρονική σειρά παραμέτρων από την ανώτερη ατμόσφαιρα, καθίσταται βιώσιμη η διερεύνηση της κλιματολογίας. Οι τάσεις ή οι αλλαγές στην ανώτερη ατμόσφαιρα μπορεί να αντικατοπτρίζουν αντίστοιχες στην τροπόσφαιρα (Hall et al., Reference Hall, Brekke and Cannon2007; Roble & Dickinson, Reference Roble and Dickinson1989), αλλά σε υψόμετρο που δεν επηρεάζεται άμεσα από τις τοπικές βραχυπρόθεσμες επιδράσεις. Κάποιος προσπαθεί να καθιερώσει τους μηχανισμούς που είναι υπεύθυνοι για τις παρατηρούμενες τάσεις, με έναν υποψήφιο να είναι η ηλιακή ισχύς: το επίκεντρο αυτής της μελέτης και εκείνο της Lastovicka (Αναφορά Lastovicka2019). Μόλις εξαλειφθεί αυτή η επίδραση, μπορεί να είναι δυνατός ο εντοπισμός άλλων, συμπεριλαμβανομένων των ανθρωπογενών αιτιών, εάν υπάρχουν.



*Εικόνα 7. 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)*

### 3.5.7 Ενεργητικές ροές ηλεκτρονίων (Energetic electrons fluxes)

---

Ροές ενεργειακών ηλεκτρονίων με ενέργειες λίγων δεκάδων keV έχουν παρατηρηθεί στην ιονόσφαιρα από πολυάριθμα πειράματα από την αρχή της διαστημικής εποχής (βλέπε κριτικές των Paulikas (1975) και Voss and Smith (1980)). Οι ροές ηλεκτρονίων αυξάνονται με το υψόμετρο και επίσης με τη γεωμαγνητική δραστηριότητα (Hill et al., 1970; Goldberg et al., 1974). Τρεις πληθυσμοί ηλεκτρονίων, παγιδευμένων, κατακρημνισμένων και οιονεί παγιδευμένων, τεκμηριώθηκαν επίσης (π.χ. Kohn, 1973). Αυτή η ταξινόμηση βασίζεται στη φυσική συμπεριφορά των ηλεκτρονίων με διαφορετικές τοπικές ισημερινές γωνίες βήματος (η γωνία μεταξύ της ταχύτητας ενός σωματιδίου και της γραμμής του μαγνητικού πεδίου) (π.χ. Tu et al., 2010). Συγκεκριμένα, τα σωματίδια που ταξινομούνται ως «κατακρημνιζόμενα» ή «μη παγιδευμένα» έχουν τοπικές ισημερινές γωνίες βήματος που κυμαίνονται εντός ενός κώνου απώλειας αναπήδησης, δηλαδή τα σωματίδια χάνονται μέσα σε μία περίοδο αναπήδησης λόγω διασποράς στην ατμόσφαιρα επειδή τα σημεία καθρέφτη τους βρίσκονται κάτω από 100 km. Τα σωματίδια, που κλείνουν τη διαδρομή τους γύρω από τη Γη, ονομάζονται «παγιδευμένα». Τα σωματίδια από έναν «οιονεί παγιδευμένο» πληθυσμό δεν μπορούν να κλείσουν την πλήρη μετατόπιση γύρω από τη Γη και οι γωνίες βήματος των οιονεί

παγιδευμένων σωματιδίων κυμαίνονται εντός ενός κώνου απώλειας ολίσθησης. Αυτά τα σωματίδια μπορούν να κάνουν έναν αριθμό αναπήδησεων, αλλά σε ένα συγκεκριμένο γεωγραφικό μήκος, η τοπική τους ισημερινή γωνία βήματος εμφανίζεται μέσα σε έναν κώνο απώλειας αναπήδησης και, ως εκ τούτου, τα σωματίδια κατακρημνίζονται.

### 3.5.8 Γεωμαγνητικοί δείκτες K

Ο δείκτης K είναι λογαριθμικός τοπικός δείκτης του εύρους 3 ωρών στη μαγνητική δραστηριότητα σε σχέση με μια υποτιθέμενη καμπύλη ησυχίας για μια μεμονωμένη θέση γεωμαγνητικού παρατηρητηρίου. Εισήχθη για πρώτη φορά από τον J. Bartels το 1938, αποτελείται από ένα μονοψήφιο 0 έως 9 για κάθε διάστημα 3 ωρών της καθολικής ώρας ημέρας (UT).

Ο πλανητικός δείκτης εύρους 3 ωρών K<sub>p</sub> είναι ο μέσος τυποποιημένος δείκτης K από 13 γεωμαγνητικά παρατηρητήρια μεταξύ 44 μοιρών και 60 μοιρών βόρειου ή νότιου γεωμαγνητικού γεωγραφικού πλάτους. Η κλίμακα είναι από το 0 έως το 9 που εκφράζεται σε τρίτα μιας μονάδας, π.χ. Το 5- είναι 4 2/3, το 5 είναι 5 και το 5+ είναι 5 1/3. Αυτός ο πλανητικός δείκτης έχει σχεδιαστεί για να μετράει την ηλιακή ακτινοβολία σωματιδίων με τα μαγνητικά του αποτελέσματα. Ο δείκτης 3 ωρών (ισοδύναμο εύρος) προέρχεται από τον δείκτη K<sub>p</sub> ως εξής:

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

Άλλοι πλανητικοί δείκτες περιλαμβάνουν τους Ap και AA\*. Ο δείκτης Ap ορίζεται ως η νωρίτερα εμφανιζόμενη μέγιστη τιμή 24 ωρών που λαμβάνεται με τον υπολογισμό ενός τρέχοντος μέσου όρου 8 σημείων διαδοχικών δεικτών AP 3 ωρών κατά τη διάρκεια ενός συμβάντος γεωμαγνητικής καταιγίδας και συνδέεται μοναδικά με το γεγονός καταιγίδας.

### 3.5.9 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 ζώνη του σέλαος που παράγεται από ενισχυμένα ιονοσφαιρικά ρεύματα που ρέουν κάτω και μέσα στο σέλας (Δαγκλής I., 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)).

### 3.5.10 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.

---

---

## Κεφάλαιο 4<sup>ο</sup> Βασικές έννοιες Τεχνητής Νοημοσύνης

---

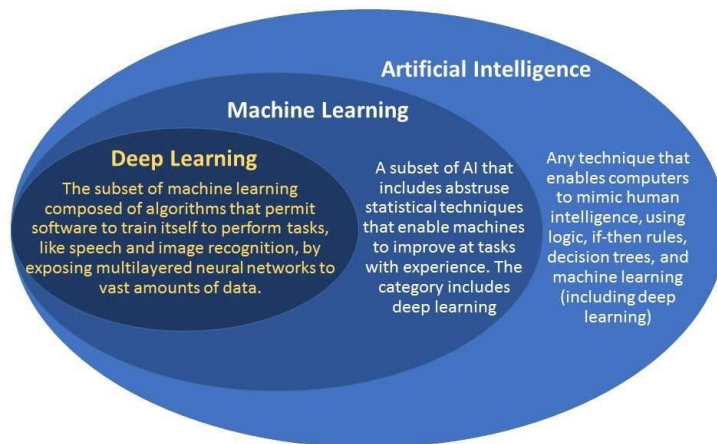
---

Η Τεχνητή Νοημοσύνη (TN) αποτελεί ένα πολύ ευρύ πεδίο μελέτης και εφαρμογής, που αναπτύσσεται με γρήγορους ρυθμούς. Εδώ είναι μερικές από τις βασικές έννοιες στον τομέα της Τεχνητής Νοημοσύνης:

- **Μηχανική Μάθηση (Machine Learning):** Η μηχανική μάθηση είναι ένα υποσύνολο της TN που εστιάζει στην ανάπτυξη αλγορίθμων και τεχνικών που επιτρέπουν στα συστήματα να μαθαίνουν από τα δεδομένα χωρίς να χρειάζεται εξαντλητική προγραμματιστική οδήγηση. Αυτό περιλαμβάνει αλγόριθμους όπως οι νευρωνικοί δίκτυα, οι δέντρα απόφασης, οι μέθοδοι συσταδοποίησης και άλλες τεχνικές.
- **Νευρωνικά Δίκτυα (Neural Networks):** Τα νευρωνικά δίκτυα αντιπροσωπεύουν ένα πολύ ισχυρό εργαλείο στην TN, εμπνευσμένο από τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Αυτά τα δίκτυα αποτελούνται από εκατομμύρια συνδεδεμένους "νευρώνες" που λειτουργούν συλλογικά για να εκτελέσουν λειτουργίες όπως η ταξινόμηση, η πρόβλεψη και η αναγνώριση προτύπων.
- **Ανάγνωση και Κατανόηση Φυσικής Γλώσσας (Natural Language Processing - NLP):** Η NLP είναι ένας τομέας της TN που αφορά την ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν σε υπολογιστές να κατανοούν και να επεξεργάζονται την ανθρώπινη γλώσσα σε όλες της τις μορφές, όπως κείμενο, φωνή και συνομιλία.
- **Εξήγηση και Διαφάνεια (Explainability and Transparency):** Η εξήγηση των αποφάσεων που παίρνουν τα συστήματα τεχνητής νοημοσύνης και η διαφάνεια των διαδικασιών είναι σημαντικές έννοιες που επιδιώκονται όλο και περισσότερο, ειδικά όταν πρόκειται για ευαίσθητες εφαρμογές, όπως η υγεία και οι οικονομικές αποφάσεις.
- **Ρομποτική (Robotics):** Η ρομποτική ασχολείται με τη σχεδίαση, την κατασκευή και την εκτέλεση ρομπότ που μπορούν να λειτουργούν αυτόνομα ή να συνεργάζονται με ανθρώπους.

Γι' αυτό και στο ακόλουθο κεφάλαιο, πρώτα θα γίνει μια αναφορά στην μηχανική μάθηση και έπειτα στην εξηγήσιμη τεχνητή νοημοσύνη, συμφωνά με την οποία υλοποιείται και η επιλογή των παραμέτρων επιρροής της ιονόσφαιρας.





Εικόνα 8. Τεχνητή νοημοσύνη και τα υποσύνολά της.

## 4.1 Εξηγήσιμη Τεχνητή Νοημοσύνη

Η "εξηγήσιμη τεχνητή νοημοσύνη" αναφέρεται στην ικανότητα ενός συστήματος τεχνητής νοημοσύνης να εξηγεί τις αποφάσεις και τις λειτουργίες του με τρόπο που είναι κατανοητός από ανθρώπους. Σε απλά λόγια, είναι η δυνατότητα να κατανοήσουμε γιατί ένα AI σύστημα πήρε μια συγκεκριμένη απόφαση ή παρήγαγε μια συγκεκριμένη έξοδο.

Η εξηγήςσιμη τεχνητή νοημοσύνη είναι σημαντική για πολλούς λόγους. Πρώτον, βοηθά στην αύξηση της εμπιστοσύνης στα συστήματα AI, καθώς οι άνθρωποι μπορούν να κατανοήσουν πώς λειτουργούν και γιατί λαμβάνουν συγκεκριμένες αποφάσεις. Δεύτερον, μπορεί να αποκαλύψει ανεπιθύμητες συμπεριφορές ή προκαταλήψεις στα συστήματα AI, επιτρέποντας έτσι στους αναλυτές και τους αναπτυσσόμενους να τις διορθώσουν ή να τις μειώσουν. Τέλος, μπορεί να βοηθήσει τους χρήστες να εκπαιδεύσουν και να βελτιώσουν τα συστήματα AI.

Οι μέθοδοι για την επίτευξη της εξηγήςσιμης τεχνητής νοημοσύνης περιλαμβάνουν τη χρήση απλών μοντέλων που είναι εύκολο να ερμηνευτούν από τον άνθρωπο, την ανάπτυξη εργαλείων οπτικοποίησης για την αναπαράσταση των δεδομένων και των αποφάσεων του συστήματος με ευανάγνωστο τρόπο, και την ενίσχυση της διαφάνειας στον τρόπο λειτουργίας των αλγορίθμων.

Η εξηγήςσιμη τεχνητή νοημοσύνη είναι ένας σημαντικός τομέας έρευνας και ανάπτυξης στον χώρο της τεχνητής νοημοσύνης, καθώς έχει ευρείες εφαρμογές σε διάφορους τομείς, συμπεριλαμβανομένων της υγείας, της ρομποτικής, των χρηματοοικονομικών και πολλών άλλων.

## 4.2 Μηχανική μάθηση (Machine Learning)

---

Νοημοσύνη ορίζεται ως η ικανότητα του ανθρώπου να επεξεργάζεται πληροφορία με σκοπό την λήψη οποιασδήποτε απόφασης. Κατ' επέκταση, Τεχνητή νοημοσύνη (Artificial Intelligence – AI) είναι η προσπάθεια του ανθρώπου να υλοποιήσει αλγορίθμους με σκοπό την επεξεργασία της πληροφορίας για την λήψη απόφασης.

Η μηχανική μάθηση είναι μέθοδος ανάλυσης δεδομένων που αυτοματοποιεί τη δόμηση αναλυτικών μοντέλων. Είναι ο κυριότερος κλάδος της τεχνητής νοημοσύνης και βασίζεται στην ιδέα ότι τα συστήματα μπορούν να μάθουν από δεδομένα, να εντοπίσουν πρότυπα και να πάρουν αποφάσεις με ελάχιστη ανθρώπινη παρέμβαση και χρησιμοποιώντας δεδομένα εκπαίδευσης. Χρησιμοποιείται κυρίως για τον εντοπισμό μίας τάσης που ακολουθούν τα δεδομένα, αλλά επικεντρώνονται και στην πρόβλεψη βάσει γνωστών πληροφοριών που μαθαίνουν από τα δεδομένα εκπαίδευσης.

Η μελέτη της μηχανικής μάθησης είναι σημαντική τόσο για την αντιμετώπιση αυτών των θεμελιωδών επιστημονικών και μηχανικών ερωτημάτων όσο και για το εξαιρετικά πρακτικό λογισμικό υπολογιστή που έχει παραγάγει και διαθέτει σε πολλές εφαρμογές. Η μηχανική μάθηση έχει προχωρήσει δραματικά τις τελευταίες δύο δεκαετίες, από την εργαστηριακή περιέργεια σε μια πρακτική τεχνολογία σε ευρεία εμπορική χρήση. Στο πλαίσιο της τεχνητής νοημοσύνης (AI), η μηχανική μάθηση έχει αναδειχθεί ως η μέθοδος επιλογής για την ανάπτυξη πρακτικού λογισμικού για όραση υπολογιστή, αναγνώριση ομιλίας, επεξεργασία φυσικής γλώσσας, έλεγχο ρομπότ και άλλες εφαρμογές.

Πολλοί προγραμματιστές συστημάτων AI αναγνωρίζουν τώρα ότι, για πολλές εφαρμογές, μπορεί να είναι πολύ πιο εύκολο να εκπαιδεύσετε ένα σύστημα δείχνοντάς του παραδείγματα επιθυμητής συμπεριφοράς εισόδου-εξόδου παρά να το προγραμματίσετε χειροκίνητα προβλέποντας την επιθυμητή απόκριση για όλες τις πιθανές εισόδους.

Στην μηχανική μάθηση, πολύ σημαντικός είναι ο επαναληπτικός παράγοντας, καθώς τα μοντέλα που κατασκευάζονται εκτίθενται σε νέα δεδομένα και μπορούν αυτόνομα να προσαρμοστούν στις ανάγκες τους. Μαθαίνουν, ουσιαστικά, από προηγούμενους υπολογισμούς να παράγουν αξιόπιστα επανειλημμένες αποφάσεις και αποτελέσματα.

Η μηχανική μάθηση διαφοροποιείται από άλλες μεθόδους, όπως η βαθιά μάθηση και η αναγνώριση προτύπων, ενώ είναι αρκετά πιο γενική μέθοδος τεχνητής νοημοσύνης σε σχέση με άλλες υπάρχουσες ως προς τους στόχους της. Οι δύο προαναφερθείσες μέθοδοι μοιάζουν μεταξύ τους, με την αναγνώριση προτύπων να ασχολείται με μεθόδους ταξινόμησης, ενώ η μηχανική μάθηση με αλγόριθμους που στοχεύουν στην εκμάθηση από τους εαυτούς τους. Η βαθιά μάθηση μπορεί να θεωρηθεί υποσύνολο της μηχανικής μάθησης, όπου το υπολογιστικό σύστημα είναι δυνατόν να εκπαιδευτεί μέσα από την αναπαράσταση των δεδομένων. Πρακτικά, ακόμη, η βαθιά μάθηση είναι ικανή να μετατρέψει ένα πρόβλημα περισσότερων διαστάσεων σε ένα αντίστοιχο με λιγότερες διαστάσεις.



Οι διάφορες κατηγορίες της μηχανικής μάθησης, δηλαδή εκπαίδευσης νευρωνικών δικτύων, παραδοσιακά χωρίζονται σε τρεις μεγάλες κατηγορίες ανάλογα με τη φύση των πληροφοριών που διατίθενται για το σύστημα εκμάθησης.

- **Επιβλεπόμενη μάθηση (Supervised Learning):** Παρέχονται από το χειριστή στον υπολογιστή παραδείγματα από δεδομένα και τα επιθυμητά τους αποτελέσματα με στόχο να εκπαιδευτεί στο γενικό κανόνα που αποτυπώνει τα δεδομένα εισόδου και εξόδου.
- **Μη επιβλεπόμενη μάθηση (Unsupervised Learning):** Δε δίνονται κατευθύνσεις στον υπολογιστή για το επιθυμητό αποτέλεσμα, επιτρέποντάς του να κατασκευάσει μόνος του ένα μοντέλο για τα δεδομένα εισόδου και να αναγνωρίσει τη δομή τους. Έχει εφαρμογή συχνά σε προβλήματα ανάλυσης-σύνδεσης και ταξινόμησης.
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Ένα υπολογιστικό πρόγραμμα αλληλοεπιδρά με ένα δυναμικό περιβάλλον εντός του οποίου πρέπει να εκτελέσει έναν συγκεκριμένο στόχο (π.χ. ο έλεγχος κίνησης ρομπότ), όπου ο αλγόριθμος εξοικειώνεται με μία στρατηγική ενεργειών. Καθώς ανακαλύπτει το χώρο του προβλήματος, ο υπολογιστής δίνει πληροφορίες για αυτό που είναι ανάλογες του κέρδους που θα έχει από την επίτευξη του στόχου, το οποίο προσπαθεί να μεγιστοποιήσει.

Η επιβλεπόμενη μάθηση χρησιμοποιείται όταν τα δεδομένα εισόδου και εξόδου ενώνονται από την σχέση:

$$Y = f(x) \quad (1.1)$$

Όπου  $f$  είναι ένα σύνολο παραμέτρων [8],

Έχουμε λοιπόν έναν ορισμό της αποστολής μας ( $T$ ): να προβλέψουμε το  $y$  από το  $x$  από τον τύπο:

$$\hat{y} = w x.$$

Ένας τρόπος μέτρησης της απόδοσης του μοντέλου είναι ο υπολογισμός του μέσου τετραγωνικού σφάλματος του μοντέλου στο δοκιμαστικό σύνολο. Αν το  $y$  δίνει τις προβλέψεις του μοντέλου στο σύνολο δοκιμής, τότε το μέσο τετράγωνο σφάλμα δίνεται από τον τύπο:

$$MSE = 1/m \sum_i (y^{\hat{}} - y)^2. \quad (5.4)$$

Για να φτιάξουμε έναν αλγόριθμο μηχανικής μάθησης, πρέπει να σχεδιάσουμε έναν αλγόριθμο που θα βελτιώσει τα βάρη  $w$  με τρόπο που να μειώνει το MSE όταν επιτρέπεται στον αλγόριθμο να αποκτήσει εμπειρία παρατηρώντας ένα σύνολο εκπαίδευσης  $(X, y)$ .

Ένας διαισθητικός τρόπος για να γίνει αυτό είναι απλώς η ελαχιστοποίηση του μέσου τετραγώνου σφάλματος στο σύνολο εκπαίδευσης, MSE [8].

### 4.3 SHAP (SHapley Additive exPlanations)

---

SHAP είναι το αρχικό για "SHapley Additive exPlanations", και αναφέρεται σε μια τεχνική που χρησιμοποιείται για την ερμηνεία και την εξήγηση προβλέψεων που παράγονται από μοντέλα μηχανικής μάθησης. Αυτή η τεχνική βασίζεται στον αλγόριθμο Shapley values, ο οποίος προέρχεται από τη θεωρία των παιγνίων και χρησιμοποιείται για να αναδείξει τη σημαντικότητα κάθε χαρακτηριστικού στην πρόβλεψη ενός μοντέλου.

Η τεχνική SHAP παρέχει εξηγήσεις για κάθε πρόβλεψη που παράγεται από ένα μοντέλο μηχανικής μάθησης, δίνοντας έναν αριθμό που αντιπροσωπεύει τη σημαντικότητα κάθε χαρακτηριστικού στην πρόβλεψη του μοντέλου. Αυτό επιτρέπει στους χρήστες να κατανοήσουν πώς και γιατί το μοντέλο κατέληξε σε μια συγκεκριμένη πρόβλεψη, καθώς και ποια χαρακτηριστικά είχαν τη μεγαλύτερη επίδραση σε αυτήν.

Η τεχνική SHAP είναι χρήσιμη για την αποσαφήνιση των αποφάσεων των μοντέλων μηχανικής μάθησης, την ανίχνευση προκαταλήψεων και τη βελτιστοποίηση της εμπιστοσύνης στα αποτελέσματά τους. Επίσης, μπορεί να χρησιμοποιηθεί για την επιλογή χαρακτηριστικών και τη βελτιστοποίηση μοντέλων. Είναι ένα ισχυρό εργαλείο για την εξήγηση της λειτουργίας μοντέλων μηχανικής μάθησης και τη βελτίωση της κατανόησης τους.

#### 4.3.1 Αναλυτικότερα

---

Ο στόχος του SHAP είναι να εξηγήσει την πρόβλεψη ενός στιγμιότυπου  $x$  υπολογίζοντας τη συμβολή κάθε χαρακτηριστικού στην πρόβλεψη. Η μέθοδος επεξήγησης SHAP υπολογίζει τις τιμές Shapley από τη θεωρία συνασπισμών παιγνίων. Οι τιμές χαρακτηριστικών μιας παρουσίας δεδομένων λειτουργούν ως παίκτες σε έναν συνασπισμό. Οι τιμές Shapley μας λένε πώς να κατανείμουμε δίκαια την «πληρωμή» (= την πρόβλεψη) μεταξύ των χαρακτηριστικών. Ένας παίκτης μπορεί να είναι μια μεμονωμένη τιμή χαρακτηριστικού, π.χ. για δεδομένα πίνακα. Ένας παίκτης μπορεί επίσης να είναι μια ομάδα τιμών χαρακτηριστικών. Για παράδειγμα, για να εξηγηθεί μια εικόνα, τα εικονοστοιχεία μπορούν να ομαδοποιηθούν σε superpixel και η πρόβλεψη να κατανεμηθεί μεταξύ τους. Μια καινοτομία που φέρνει το SHAP είναι ότι η εξήγηση της τιμής Shapley αναπαρίσταται ως μια μέθοδος απόδοσης πρόσθετων χαρακτηριστικών,

ένα γραμμικό μοντέλο. Αυτή η προβολή συνδέει τις τιμές LIME και Shapley. Το SHAP καθορίζει την εξήγηση ως εξής:

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z'_j$$

όπου  $g$  είναι το μοντέλο εξήγησης,  $z' \in \{0,1\}^M$  είναι το διάνυσμα συνασπισμού,  $M$  είναι το μέγιστο μέγεθος συνασπισμού και  $\phi_j$  η απόδοση χαρακτηριστικού για ένα χαρακτηριστικό  $j$ , οι τιμές Shapley. Αυτό που αποκαλώ "διάνυσμα συνασπισμού" ονομάζεται "απλοποιημένα χαρακτηριστικά" στο έγγραφο SHAP. Νομίζω ότι επιλέχθηκε αυτό το όνομα, γιατί π.χ. δεδομένα εικόνας, οι εικόνες δεν αναπαρίστανται σε επίπεδο pixel, αλλά συγκεντρώνονται σε superpixel. Πιστεύω ότι είναι χρήσιμο να σκεφτόμαστε ότι τα  $z$  περιγράφουν συνασπισμούς: Στο διάνυσμα συνασπισμού, μια καταχώριση 1 σημαίνει ότι η αντίστοιχη τιμή χαρακτηριστικού είναι "παρούσα" και 0 ότι είναι "απουσία". Αυτό θα πρέπει να σας ακούγεται οικείο εάν γνωρίζετε τις τιμές Shapley. Για να υπολογίσουμε τις τιμές Shapley, προσομοιάζουμε ότι παίζονται μόνο ορισμένες τιμές χαρακτηριστικών ("παρούσα") και κάποιες όχι ("απουσία"). Η αναπαράσταση ως γραμμικό μοντέλο συνασπισμών είναι ένα τέχνασμα για τον υπολογισμό των  $\phi$ . Για το  $x$ , το παράδειγμα ενδιαφέροντος, το διάνυσμα συνασπισμού  $x'$  είναι ένα διάνυσμα όλων των 1, δηλαδή όλες οι τιμές χαρακτηριστικών είναι "παρούσες". Ο τύπος απλοποιεί:

$$g(x') = \phi_0 + \sum_{j=1}^M \phi_j$$

Οι τιμές Shapley είναι η μόνη λύση που ικανοποιεί τις ιδιότητες της Απόδοσης, της Συμμετρίας, του Ομοιώματος και της Προσθετικότητας. Το SHAP ικανοποιεί επίσης αυτά, αφού υπολογίζει τις τιμές Shapley. Στο έγγραφο SHAP, θα βρείτε αποκλίσεις μεταξύ των ιδιοτήτων SHAP και των ιδιοτήτων Shapley.

Από τη παραπάνω μέθοδο προκύπτουν τα κάτωθι διαγράμματα.

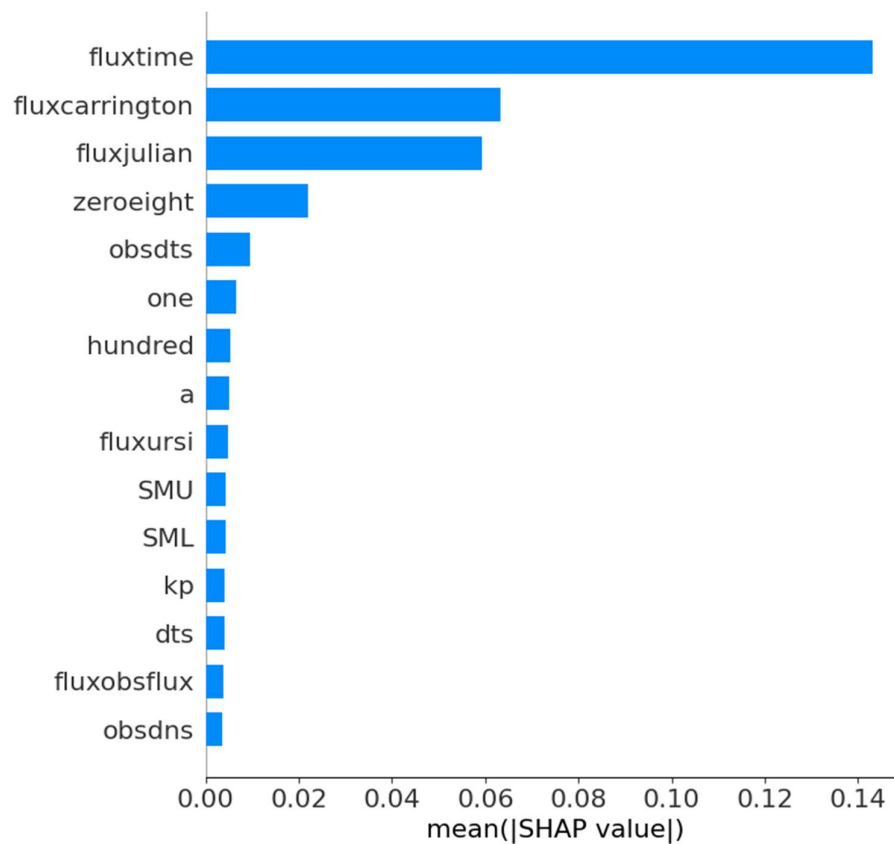
### 4.3.2 SHAP Feature Importance

---

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



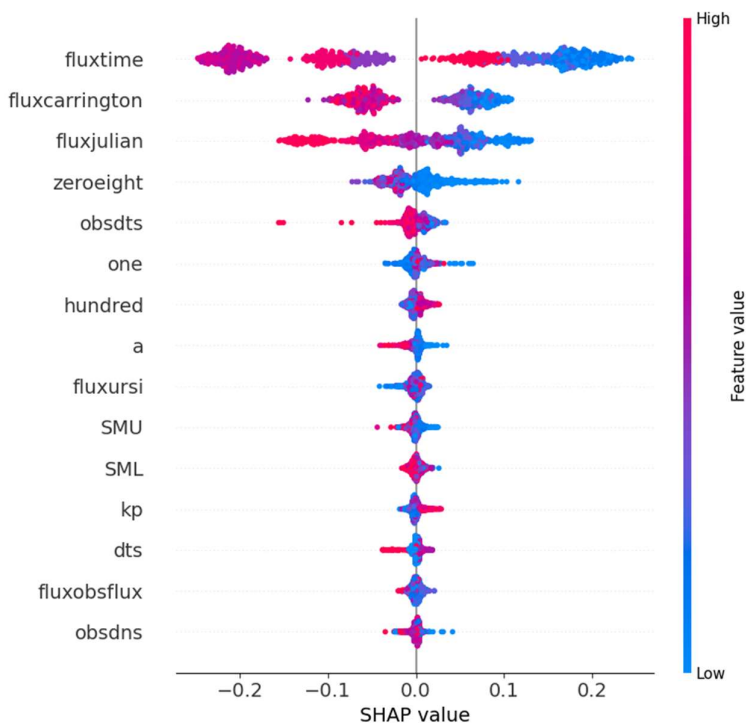
Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής STEC κατά 14 ποσοστιαίες μονάδες (0,14 στον άξονα x).

Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 4.3.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών **ανά τιμή**. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παραπάνω φαίνεται το SHAP συνοπτικό διάγραμμα. Οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν την επίδραση στις τιμές STEC, ενώ οι χαμηλές τιμές αυξάνουν την επίδραση.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

#### 4.3.4 SHAP Interaction Values

---

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

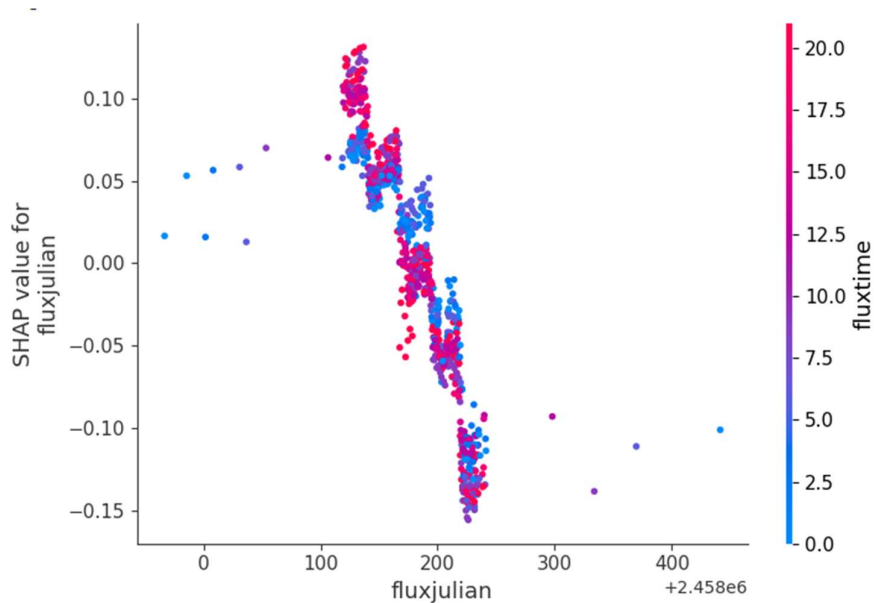
$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:



Παραπάνω φαίνεται το σχεδιάγραμμα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης. Η παράμετρος fluxjulian αλληλεπιδρά με την παράμετρο fluxtime. Σε περιπτώσεις κοντά στην τιμή 0 fluxjulian, η εμφάνιση της παραμέτρου fluxtime αυξάνει την επίδραση στο STEC. Για μεγαλύτερες τιμές παραμέτρου fluxjulian (100-300), η εμφάνιση της παραμέτρου fluxtime διατηρεί σταθερή την επίδραση στο STEC καθώς η μονοτονία του διαγράμματος είναι σταθερή και οι υψηλές καθώς και οι χαμηλές τιμές της παραμέτρου fluxtime έχουν το ίδιο βάρος επίδρασης στη παράμετρο fluxjulian..

#### 4.3.5 Πλεονεκτήματα

Εφόσον το SHAP υπολογίζει τις τιμές Shapley, ισχύουν όλα τα πλεονεκτήματα των τιμών Shapley: Το SHAP έχει μια σταθερή θεωρητική βάση στη θεωρία παιγνίων. Η πρόβλεψη κατανέμεται δίκαια μεταξύ των τιμών των χαρακτηριστικών. Λαμβάνουμε αντιθετικές εξηγήσεις που συγκρίνουν την πρόβλεψη με τη μέση πρόβλεψη.

Το SHAP συνδέει τις τιμές LIME και Shapley. Αυτό είναι πολύ χρήσιμο για την καλύτερη κατανόηση και των δύο μεθόδων. Βοηθά επίσης στην ενοποίηση του πεδίου της ερμηνεύσιμης μηχανικής μάθησης.

Το SHAP έχει μια γρήγορη υλοποίηση για μοντέλα που βασίζονται σε δέντρα. Πιστεύω ότι αυτό ήταν το κλειδί για τη δημοτικότητα του SHAP, επειδή το μεγαλύτερο εμπόδιο για την υιοθέτηση των τιμών Shapley είναι ο αργός υπολογισμός.

Ο γρήγορος υπολογισμός καθιστά δυνατό τον υπολογισμό των πολλών τιμών Shapley που απαιτούνται για τις καθολικές ερμηνείες μοντέλων. Οι καθολικές μέθοδοι ερμηνείας περιλαμβάνουν τη σημασία χαρακτηριστικών, την εξάρτηση χαρακτηριστικών, τις αλληλεπιδράσεις, την ομαδοποίηση και τα περίληψη. Με το SHAP, οι καθολικές ερμηνείες είναι συνεπείς με τις τοπικές εξηγήσεις, αφού οι τιμές Shapley είναι η «ατομική μονάδα» των καθολικών ερμηνειών. Εάν χρησιμοποιείτε LIME για τοπικές επεξηγήσεις

και γραφικά μερικής εξάρτησης συν τη σημασία της δυνατότητας μετάθεσης για καθολικές εξηγήσεις, δεν έχετε κοινή βάση.

#### 4.3.6 Μειονεκτήματα

---

Το KernelSHAP είναι αργό. Αυτό καθιστά το KernelSHAP μη πρακτικό στη χρήση όταν θέλετε να υπολογίσετε τιμές Shapley για πολλές περιπτώσεις. Επίσης, όλες οι καθολικές μέθοδοι SHAP, όπως η σημασία του χαρακτηριστικού SHAP, απαιτούν τον υπολογισμό των τιμών Shapley για πολλές περιπτώσεις.

Το KernelSHAP αγνοεί την εξάρτηση χαρακτηριστικών. Οι περισσότερες άλλες μέθοδοι ερμηνείας που βασίζονται στη μετάθεση έχουν αυτό το πρόβλημα. Αντικαθιστώντας τις τιμές χαρακτηριστικών με τιμές από τυχαίες περιπτώσεις, είναι συνήθως ευκολότερο να γίνει τυχαία δειγματοληψία από την οριακή κατανομή. Ωστόσο, εάν τα χαρακτηριστικά εξαρτώνται, π.χ. συσχετίζονται, αυτό οδηγεί σε υπερβολική βαρύτητα σε απίθανα σημεία δεδομένων. Το TreeSHAP επιλύει αυτό το πρόβλημα μοντελοποιώντας ρητά την υπό όρους αναμενόμενη πρόβλεψη.

Το TreeSHAP μπορεί να παράγει μη διαισθητικές αποδόσεις χαρακτηριστικών. Ενώ το TreeSHAP επιλύει το πρόβλημα της παρέκτασης σε απίθανα σημεία δεδομένων, το κάνει αλλάζοντας τη συνάρτηση τιμής και επομένως αλλάζει ελαφρώς το παιχνίδι. Το TreeSHAP αλλάζει τη συνάρτηση τιμής βασιζόμενος στην υπό όρους αναμενόμενη πρόβλεψη. Με την αλλαγή στη συνάρτηση τιμής, τα χαρακτηριστικά που δεν έχουν καμία επίδραση στην πρόβλεψη μπορούν να λάβουν μια τιμή TreeSHAP διαφορετική από το μηδέν.

Τα μειονεκτήματα των τιμών Shapley ισχύουν επίσης για το SHAP: Οι τιμές Shapley μπορεί να παρερμηνευθούν και απαιτείται πρόσβαση στα δεδομένα για τον υπολογισμό τους για νέα δεδομένα (εκτός από το TreeSHAP).

Είναι δυνατό να δημιουργηθούν σκόπιμα παραπλανητικές ερμηνείες με το SHAP, το οποίο μπορεί να κρύψει προκαταλήψεις, δημιουργούν παραπλανητικές εξηγήσεις). Για τους αποδέκτες μιας εξήγησης SHAP, είναι ένα μειονέκτημα: δεν μπορούν να είναι σίγουροι για την αλήθεια της εξήγησης.

## 4.4 Πρόβλεψη και Νευρωνικά δίκτυα (Neural Networks)

---

Νευρωνικό δίκτυο (Neural Network ή NN) [8], ορίζεται ως ένα κύκλωμα διασυνδεδεμένων νευρώνων. Με αναφορά τον ανθρώπινο εγκέφαλο, τα βιολογικά νευρωνικά δίκτυα (Biological Neural Networks) αποτελούνται από ένα σύνολο νευρώνων με τις διασυνδέσεις τους που έχουν ως στόχο την επεξεργασία των σημάτων που δέχεται ο



άνθρωπος καθημερινά. Τα τεχνητά νευρωνικά δίκτυα (Artificial neural networks - ANNs), αποτελούν μια μίμηση των βιολογικών καθώς δομούνται από κόμβους ή αλλιώς νευρώνες πλήρως διασυνδεδεμένους μεταξύ τους, ενώ δέχονται και επεξεργάζονται σήματα με απώτερο στόχο την μάθηση. Η υλοποίηση αυτών γίνεται μέσω της ανάπτυξης αλγορίθμων ικανών να επεξεργαστούν οποιοδήποτε σήμα.

Προκειμένου να γίνει αντιληπτή η έννοια των τεχνητών νευρωνικών δικτύων, αξίζει να αναφερθεί το πιο βασικό νευρωνικό δίκτυο που συνάμα αποτελεί και έναν νευρώνα ή αλλιώς ένα επίπεδο νευρωνικού δικτύου ή αλλιώς perceptron. Εφευρέθηκε από τον Frank Rosenblatt το 1957 και είναι η πιο απλοϊκή δομή νευρωνικού δικτύου. Εκτελεί κυρίως διεργασίες ταξινόμησης σε πρότυπα (ή κλάσεις) δεδομένων τα οποία είναι πλήρως διαχωρίσιμα μέσω μία γραμμής ή αλλιώς ευθείας σε ένα επίπεδο στο χώρο (linearly separable) [8].

#### 4.4.1 Συνάρτηση Απώλειας (Loss Function)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων.

Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο Τετραγωνικό Σφάλμα, Μέσο Απόλυτο Σφάλμα, Ρίζα Μέσου Τετραγωνικού Σφάλματος και R-Squared ή Συντελεστής προσδιορισμού χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

Στην παρούσα διπλωματική θα μας απασχολήσουν το MAE και το MSE

Το μέσο απόλυτο σφάλμα 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.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

Where,

$\hat{y}$  – predicted value of  $y$

$\bar{y}$  – mean value of  $y$

Εικόνα 9. Συνάρτηση απώλειας Mean Absolute Error (MAE)

Το μέσο τετραγωνικό σφάλμα αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Εικόνα 10. Συνάρτηση απώλειας Mean Square Error (MSE)

Οι διαφορές μεταξύ των παραπάνω είναι οι εξής:

- Το μέσο τετραγωνικό σφάλμα (MSE) και το ριζικό μέσο τετραγωνικό σφάλμα τιμωρούν τα μεγάλα σφάλματα πρόβλεψης *vi-a-vis*. Μέσο απόλυτο σφάλμα (MAE). Ωστόσο, το RMSE χρησιμοποιείται ευρέως από το MSE για την αξιολόγηση της απόδοσης του μοντέλου παλινδρόμησης με άλλα τυχαία μοντέλα, καθώς έχει τις ίδιες μονάδες με την εξαρτημένη μεταβλητή (άξονας Υ).
- Το MSE είναι μια διαφοροποιήσιμη συνάρτηση που διευκολύνει την εκτέλεση μαθηματικών πράξεων σε σύγκριση με μια μη διαφοροποιήσιμη συνάρτηση όπως η MAE. Επομένως, σε πολλά μοντέλα, το RMSE χρησιμοποιείται ως προεπιλεγμένη μέτρηση για τον υπολογισμό της Συνάρτησης Απώλειας παρόλο που είναι πιο δύσκολο να ερμηνευτεί από το MAE.
- Το MAE είναι πιο ανθεκτικό σε δεδομένα με ακραίες τιμές.
- Η χαμηλότερη τιμή των MAE, MSE και RMSE συνεπάγεται υψηλότερη ακρίβεια ενός μοντέλου παλινδρόμησης. Ωστόσο, μια υψηλότερη τιμή του τετραγώνου R θεωρείται επιθυμητή.

Επομένως, εάν συγκρίνετε την ακρίβεια πρόβλεψης μεταξύ διαφορετικών μοντέλων γραμμικής παλινδρόμησης (LR), τότε το RMSE είναι μια καλύτερη επιλογή καθώς είναι απλό στον υπολογισμό και διαφοροποιήσιμο. Ωστόσο, εάν το σύνολο δεδομένων σας έχει ακραίες τιμές, τότε επιλέξτε MAE έναντι RMSE.

Επιπλέον, ο αριθμός των μεταβλητών πρόβλεψης σε ένα μοντέλο γραμμικής παλινδρόμησης προσδιορίζεται με προσαρμοσμένο τετράγωνο R και επιλέξτε RMSE έναντι προσαρμοσμένου R τετράγωνο εάν ενδιαφέρεστε για την αξιολόγηση της ακρίβειας πρόβλεψης μεταξύ διαφορετικών μοντέλων LR. [23]

## 4.5 Linear Regression (Γραμμική Παλινδρόμηση)

Η γραμμική παλινδρόμηση (linear regression) είναι ένας από τους απλούστερους αλγορίθμους μηχανικής μάθησης και ανήκει στην κατηγορία της επιβλεπόμενης

μάθησης. Σκοπός της είναι να μοντελοποιήσει τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής (είτε συνήθως συμβολίζεται με  $y$ ) και μίας ή περισσότερων ανεξάρτητων μεταβλητών (συνήθως συμβολίζονται με  $x$ ). Η γραμμική παλινδρόμηση υποθέτει ότι η σχέση μεταξύ αυτών των μεταβλητών είναι γραμμική, δηλαδή μπορεί να περιγραφεί με έναν ευθύγραμμο τρόπο.

Το μοντέλο γραμμικής παλινδρόμησης αναπαρίσταται συνήθως από την εξίσωση:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Όπου:

- $y$  είναι η εξαρτημένη μεταβλητή που θέλουμε να προβλέψουμε.
- $x_1, x_2, \dots, x_n$  είναι οι ανεξάρτητες μεταβλητές.
- $\beta_0, \beta_1, \dots, \beta_n$  είναι οι συντελεστές παλινδρόμησης (βήτα) που υπολογίζονται από τα δεδομένα εκπαίδευσης.
- $\epsilon$  είναι ο όρος σφάλματος.

Ο στόχος είναι να βρεθούν οι βέλτιστες τιμές για τους συντελεστές παλινδρόμησης ( $\beta$ ) έτσι ώστε η πρόβλεψη που παράγει το μοντέλο να είναι όσο το δυνατόν πιο κοντά στις πραγματικές τιμές της εξαρτημένης μεταβλητής  $y$ .

Η γραμμική παλινδρόμηση είναι ένα από τα πιο βασικά εργαλεία στη στατιστική και τη μηχανική μάθηση και χρησιμοποιείται ευρέως για προβλέψεις και ανάλυση δεδομένων. Παρόλο που είναι απλό, μπορεί να είναι πολύ ισχυρό όταν χρησιμοποιείται σωστά και σε συνδυασμό με κατάλληλη προεπεξεργασία δεδομένων και επιλογή χαρακτηριστικών.

## 4.6 Decision Tree (Δέντρο Απόφασης)

---

Το Δέντρο Απόφασης (Decision Tree) είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για τόσο προβλήματα ταξινόμησης όσο και παλινδρόμησης. Στην ουσία, αναπαριστά ένα σύνολο κανόνων που μαθαίνει από τα δεδομένα ώστε να καθορίσει την κατηγορία της εισόδου.

Ένα δέντρο απόφασης αναπαρίσταται από ένα σύνολο αποφάσεων που οδηγούν σε άλλες αποφάσεις ή σε τελικά αποτελέσματα. Αρχίζει από έναν κόμβο, ο οποίος αντιπροσωπεύει την αρχική ερώτηση ή τον αρχικό διαχωρισμό στα δεδομένα, και συνεχίζει σε καθορισμένα κλαδιά, κάθε ένα από τα οποία αντιστοιχεί σε μια επόμενη ερώτηση ή διαχωρισμό, μέχρι να φτάσει σε ένα φύλλο, το οποίο αντιπροσωπεύει την τελική κατηγορία ή την τελική πρόβλεψη.

Ο αλγόριθμος εκπαιδεύει το δέντρο με χρήση δεδομένων εκπαίδευσης, προσπαθώντας να επιλέξει τις βέλτιστες ερωτήσεις ή διαχωρισμούς σε κάθε επίπεδο του δέντρου, οι οποίοι να οδηγούν σε μέγιστη απόσταση μεταξύ των κλάσεων στα φύλλα.

Οι πλεονεκτήματα των δέντρων αποφάσεων περιλαμβάνουν την ευκολία στην ερμηνεία και την απεικόνιση των αποφάσεων, καθώς και τη δυνατότητα χειρισμού και αντιμετώπισης μεγάλου αριθμού δεδομένων. Ωστόσο, μπορεί να τείνει να υπερπροσαρμοστεί στα δεδομένα εκπαίδευσης, προκαλώντας υπερ-προσαρμογή (overfitting), και μπορεί να μην είναι πάντα το καλύτερο μοντέλο για δεδομένα με σύνθετες σχέσεις.

## 4.7 Support Vector Regression (Υποστηρικτική Διανυσματική Παλινδρόμηση)

---

Η Υποστηρικτική Διανυσματική Παλινδρόμηση (Support Vector Regression - SVR) είναι μια μέθοδος παλινδρόμησης που χρησιμοποιεί τις αρχές της Υποστηρικτικής Διανυσματικής Μηχανής (Support Vector Machine - SVM) για την πρόβλεψη συνεχών μεταβλητών. Όπως και στην περίπτωση της ταξινόμησης με την υποστηρικτική διανυσματική μηχανή, η SVR αποσκοπεί στην εύρεση ενός υπερεπιπέδου (ή πολλαπλών υπερεπιπέδων σε περιπτώσεις μη γραμμικής παλινδρόμησης) που να προσεγγίζει όσο το δυνατόν καλύτερα τα δεδομένα.

Το SVR βασίζεται στην αρχή της μεγιστοποίησης του περιθωρίου (ελάχιστο περιθώριο σφάλματος) των παρατηρήσεων, δηλαδή προσπαθεί να ελαχιστοποιήσει το σφάλμα πρόβλεψης ενώ παράλληλα επιτρέπει μια ποσότητα ελευθερίας (περιθώριο) γύρω από το υπερεπίπεδο. Η μέθοδος αυτή χρησιμοποιεί ένα κριτήριο απόφασης, το οποίο είναι ορισμένο από ένα σύνολο διανυσμάτων υποστήριξης (support vectors) που βρίσκονται κοντά στο υπερεπίπεδο.

Η SVR μπορεί να χρησιμοποιηθεί είτε για γραμμική παλινδρόμηση, είτε για μη γραμμική παλινδρόμηση, χρησιμοποιώντας διάφορες συναρτήσεις πυρήνα (kernel functions) για την αναπαράσταση των μη γραμμικών σχέσεων μεταξύ των μεταβλητών.

Η υποστηρικτική διανυσματική παλινδρόμηση είναι ιδιαίτερα χρήσιμη όταν έχουμε μικρό αριθμό διαστάσεων (χαρακτηριστικών) σε σύγκριση με τον αριθμό των παρατηρήσεων (στην περίπτωση του υψηλής διαστατικότητας, η απόδοση του μοντέλου μπορεί να είναι χαμηλή), καθώς και όταν έχουμε δεδομένα που περιέχουν θόρυβο ή ακραίες τιμές. Επίσης, μπορεί να είναι αποτελεσματική σε περιπτώσεις μη γραμμικής σχέσης μεταξύ των μεταβλητών.

## 4.8 Gaussian Process Regressor (Γκαουσιανός Κατακερματισμός)

---

Ο Γκαουσιανός Κατακερματισμός (Gaussian Process Regression - GPR) είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα παλινδρόμησης. Ανήκει στην οικογένεια των μη παραμετρικών μεθόδων και βασίζεται στην αρχή των γκαουσιανών διαδικασιών, που είναι ένα ευέλικτο μοντέλο που μπορεί να μοντελοποιήσει αβέβαιες συναρτήσεις με πιθανοστικό τρόπο.

Στη βασική του μορφή, το GPR προσπαθεί να προβλέψει την τιμή μιας συνάρτησης σε ένα σημείο ή ένα σύνολο σημείων, λαμβάνοντας υπόψη την αβεβαιότητα στις προβλέψεις του. Επεκτείνοντας την έννοια της πιθανότητας από τη γκαουσιανή κατανομή, το GPR παράγει μια πρόβλεψη για την τιμή μιας συνάρτησης σε ένα σημείο, συνοδευόμενη από ένα διάστημα εμπιστοσύνης.

Η κύρια ιδέα πίσω από το GPR είναι ότι το μοντέλο προβλέπει την τιμή μιας συνάρτησης ως μια γκαουσιανή κατανομή. Αυτό σημαίνει ότι, αντί να δίνει μια μόνο τιμή πρόβλεψης, παράγει ένα πλήθος δυνατών τιμών και υπολογίζει την πιθανότητα κάθε τιμής να είναι η πραγματική τιμή. Αυτή η προσέγγιση επιτρέπει την αντιμετώπιση της αβεβαιότητας στην πρόβλεψη, προσφέροντας ταυτόχρονα έναν τρόπο για την εκτίμηση της ποιότητας των προβλέψεων.

Το GPR είναι ιδιαίτερα χρήσιμο όταν έχουμε περιορισμένο αριθμό δεδομένων και θέλουμε να προβλέψουμε τιμές συναρτήσεων με αβεβαιότητα. Μπορεί να χρησιμοποιηθεί σε πολλούς τομείς, όπως η ρομποτική, η αναγνώριση προτύπων, η πρόβλεψη χρονοσειρών και άλλες εφαρμογές που απαιτούν προβλέψεις με αβεβαιότητα.

## 4.9 XGBOOST Regressor

---

Ο XGBoost (Extreme Gradient Boosting) Regressor είναι ένας πανίσχυρος αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα παλινδρόμησης. Είναι μια βελτιωμένη υλοποίηση της μεθόδου Gradient Boosting, η οποία στοχεύει στην εκπαίδευση σειράς από αδύναμα μοντέλα μηχανικής μάθησης (συνήθως δέντρα αποφάσεων), τα οποία στο τέλος συνδυάζονται για να παράγουν ένα ισχυρό μοντέλο.

Ο αλγόριθμος XGBoost έχει πολλά πλεονεκτήματα, συμπεριλαμβανομένης της υψηλής απόδοσης, της αντοχής σε ακραίες τιμές και της ικανότητας να αντιμετωπίζει αποτελεσματικά προβλήματα μεγάλων δεδομένων. Επιπλέον, προσφέρει πολλές παραμετροποιήσιμες επιλογές που μπορούν να προσαρμοστούν στις ανάγκες του προβλήματος.

Το XGBoost χρησιμοποιεί ένα σύστημα εκπαίδευσης που βασίζεται σε Gradient Boosting, όπου κάθε νέο μοντέλο εστιάζει στα σημεία που παραμένουν από το προηγούμενο μοντέλο, καθιστώντας τον αλγόριθμο πολύ αποτελεσματικό στη μάθηση πολύπλοκων συναρτήσεων.

Το XGBoost είναι ευρέως χρησιμοποιούμενο σε πολλούς τομείς, όπως η ανάλυση δεδομένων, η αναγνώριση προτύπων, η πρόβλεψη χρονοσειρών, η ανάλυση κειμένου, η βιοπληροφορική, κ.λπ. Λόγω της απόδοσής του και της ευελιξίας του, θεωρείται ένα από τα δημοφιλέστερα εργαλεία μηχανικής μάθησης.

## 4.10 Multi-layer Perceptron Regressor (Πολύ-επίπεδος Επικοινωνιακός Ανιχνευτής)

---

Ο Πολυεπίπεδος Επικοινωνιακός Ανιχνευτής (Multi-layer Perceptron - MLP) Ανιχνευτής Επικοινωνίας είναι ένας τύπος τεχνητού νευρωνικού δικτύου που χρησιμοποιείται συχνά για προβλήματα παλινδρόμησης. Είναι ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο, που σημαίνει ότι κάθε νευρώνας σε μια στρώση συνδέεται με κάθε νευρώνα στην επόμενη στρώση.

Το MLP αποτελείται από τρία επίπεδα:

**Επίπεδο εισόδου:** Αυτό το επίπεδο λαμβάνει τις εισόδους και μεταφέρει τις τιμές τους στο δίκτυο.

**Κρυφό επίπεδο (ή περισσότερα):** Αυτά τα επίπεδα περιέχουν τους κρυφούς νευρώνες, οι οποίοι λαμβάνουν την είσοδο από το προηγούμενο επίπεδο, εκτελούν συναλλαγές και προωθούν την έξοδο στο επόμενο επίπεδο.

**Επίπεδο εξόδου:** Αυτό το επίπεδο εκφράζει την έξοδο του μοντέλου, η οποία μπορεί να είναι μια μονή τιμή (στην περίπτωση της παλινδρόμησης) ή ένα διάνυσμα πιθανοτήτων (στην περίπτωση της ταξινόμησης).

Κατά τη διάρκεια της εκπαίδευσης, το MLP εκπαιδεύεται να προβλέπει την επιθυμητή έξοδο μέσω επαναλαμβανόμενων επαναλήψεων, όπως η αντίστροφη διάδοση (backpropagation). Κατά τη διάρκεια αυτής της διαδικασίας, το δίκτυο ενημερώνεται για να μειώσει το σφάλμα μεταξύ της προβλεπόμενης εξόδου και της πραγματικής εξόδου.

Το MLP είναι ένα ισχυρό εργαλείο για προβλήματα παλινδρόμησης όταν έχουμε μεγάλο όγκο δεδομένων και πολύπλοκες σχέσεις μεταξύ των χαρακτηριστικών και της επιθυμητής εξόδου. Ωστόσο, απαιτεί συνήθως περισσότερη εκπαίδευση και

προεπεξεργασία δεδομένων σε σύγκριση με άλλους αλγορίθμους, και επίσης μπορεί να υποφέρει από υπερεκπαίδευση εάν δεν ρυθμιστούν σωστά οι υπερπαράμετροι.

#### 4.11 LightGBM Regressor

---

Ο LightGBM (Light Gradient Boosting Machine) Regressor είναι ένας από τους πιο προηγμένους αλγορίθμους μηχανικής μάθησης για προβλήματα παλινδρόμησης. Ανήκει στην κατηγορία των αλγορίθμων Boosting, οι οποίοι είναι πολύ αποτελεσματικοί στην εκπαίδευση αρκετά απλών μοντέλων με βάση το σφάλμα πρόβλεψης των προηγούμενων μοντέλων.

Ο LightGBM βασίζεται στη δομή του Gradient Boosting, αλλά είναι σχεδιασμένος για να είναι πιο γρήγορος και να καταναλώνει λιγότερη μνήμη. Χρησιμοποιεί μια τεχνική που ονομάζεται "Histogram-based Gradient Boosting", η οποία επιτρέπει την αποδοτική εκπαίδευση μεγάλων συνόλων δεδομένων.

Οι βασικές χαρακτηριστικές του LightGBM περιλαμβάνουν:

**Υψηλή απόδοση:** Ο LightGBM είναι πολύ γρήγορος και μπορεί να εκπαιδεύσει μοντέλα σε πολύ μεγάλα σύνολα δεδομένων.

**Χρήση λιγότερης μνήμης:** Ο LightGBM χρησιμοποιεί λιγότερη μνήμη σε σύγκριση με άλλους αλγορίθμους Boosting.

**Αντιμετώπιση ανισορροπιών κατηγοριών:** Ο LightGBM μπορεί να χειριστεί αποτελεσματικά δεδομένα με ανισορροπίες στις κατηγορίες.

**Ρυθμιζόμενες παράμετροι:** Οι παράμετροι του LightGBM μπορούν να ρυθμιστούν για να προσαρμοστούν στις ανάγκες του προβλήματος.

Ο LightGBM είναι ευρέως χρησιμοποιούμενος σε πολλούς τομείς, όπως η ανάλυση δεδομένων, η πρόβλεψη χρονοσειρών, η αναγνώριση προτύπων και άλλες εφαρμογές μηχανικής μάθησης. Λόγω της απόδοσής του και της ευελιξίας του, θεωρείται ένα από τα πιο ισχυρά εργαλεία για προβλήματα παλινδρόμησης.

#### 4.12 Random Forrest Regressor

---

Ο Random Forest Regressor είναι ένας από τους πιο δημοφιλείς αλγορίθμους μηχανικής μάθησης για προβλήματα παλινδρόμησης. Ανήκει στην κατηγορία των αλγορίθμων Ensemble Learning, οι οποίοι συνδυάζουν τις προβλέψεις από πολλαπλά μοντέλα για να βελτιώσουν την ακρίβεια και τη γενικευτική ικανότητα.

Το Random Forest αποτελείται από ένα σύνολο από δέντρα αποφάσεων, όπου κάθε δέντρο εκπαιδεύεται ανεξάρτητα με ένα τυχαίο υποσύνολο των δεδομένων εκπαίδευσης και των χαρακτηριστικών. Κατά τη διάρκεια της πρόβλεψης, το Random Forest συγκεντρώνει τις προβλέψεις από κάθε δέντρο και υπολογίζει το μέσο όρο των προβλέψεων ή την πιο συχνή πρόβλεψη για την τελική πρόβλεψη.

Οι βασικές χαρακτηριστικές του Random Forest περιλαμβάνουν:

**Ανθεκτικότητα στην υπερεκπαίδευση (Overfitting):** Η χρήση τυχαίων υποσυνόλων των δεδομένων για την εκπαίδευση των δέντρων βοηθά στην πρόληψη της υπερεκπαίδευσης.

**Ανθεκτικότητα στο θόρυβο:** Το Random Forest μπορεί να διαχειριστεί αποτελεσματικά δεδομένα με θόρυβο και ακραίες τιμές.

**Εύκολη παραμετροποίηση:** Οι παράμετροι του Random Forest είναι εύκολο να ρυθμιστούν και συνήθως δεν απαιτούν πολύ προσεκτική προσαρμογή.

**Επικύρωση σημασιολογικής σημασίας (Feature Importance):** Το Random Forest μπορεί να παράγει σημασιολογικές σημασίες για τα χαρακτηριστικά, που μπορούν να χρησιμοποιηθούν για την κατανόηση της σημασίας κάθε χαρακτηριστικού για την πρόβλεψη.

Ο Random Forest είναι ένα ισχυρό εργαλείο για προβλήματα παλινδρόμησης, ειδικά όταν έχουμε μεγάλα σύνολα δεδομένων με πολλά χαρακτηριστικά και πολύπλοκες σχέσεις μεταξύ των μεταβλητών.



---

---

## Κεφάλαιο 5° Η γλώσσα προγραμματισμού Python

---

---

### 5.1 Γενικά για την γλώσσα προγραμματισμού Python

---

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

Η Python είναι διερμηνευόμενη (interpreted), γενικού σκοπού (general-purpose) και υψηλού επιπέδου, γλώσσα προγραμματισμού. Ανήκει στις γλώσσες προστακτικού προγραμματισμού (Imperative programming) και υποστηρίζει τόσο το διαδικαστικό (procedural programming) όσο και το αντικειμενοστραφές (object-oriented programming) προγραμματιστικό υπόδειγμα (programming paradigm). Είναι δυναμική γλώσσα προγραμματισμού (dynamically typed) και υποστηρίζει συλλογή απορριμμάτων (garbage collection ή GC)[54], [55].

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.

Ο κύριος στόχος της είναι η αναγνωσιμότητα του κώδικά της και η ευκολία χρήσης της. Το συντακτικό της, επιτρέπει στους προγραμματιστές να εκφράσουν έννοιες σε λιγότερες γραμμές κώδικα από ότι θα ήταν δυνατόν σε γλώσσες όπως η C++ ή η Java. Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησής της. Μειονεκτεί στο 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 αποτελεί μία εύκολη στην κατανόηση της γλώσσα προγραμματισμού, κυρίως λόγω της εύκολης ανάγνωσής της. Χρησιμοποιεί αγγλικούς χαρακτήρες και δεν χρησιμοποιεί τονισμούς. Σε σχέση με άλλες γλώσσες προγραμματισμού, δεν χρησιμοποιεί αγκύλες και τα ερωτηματικά μετά τις δηλώσεις επιτρέπονται αλλά είναι σπάνια. Σύμφωνα με τον Guido Van Rossum « Το όμορφο είναι καλύτερο από το άσχημο»

Η γλώσσα αυτή είναι ανοιχτού λογισμικού και διατίθεται 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 (Γενική Άδεια Δημόσιας Χρήσης GNU) [54].

Για να εκτελεστούν διαδραστικά (interactively) μεμονωμένες εντολές ή προγράμματα είναι απαραίτητη η εγκατάσταση του διερμηνευτή της Python, ο οποίος είναι ελεύθερα διαθέσιμος για «κατέβασμα» (download) από τον επίσημο ιστότοπό της (www.python.org). Για Microsoft Windows υπάρχουν εκδόσεις των 32 ή 64 bits. Στα

Λειτουργικά συστήματα Linux και Mac OS X συνηθίζεται να είναι προεγκατεστημένη, πιθανότατα όμως να είναι μια παλαιότερη έκδοσή της[54].

Για την υλοποίηση της γλώσσας προγραμματισμού, είναι απαραίτητος ένας κειμενογράφος, είτε ένα ολοκληρωμένο περιβάλλον ανάπτυξης (Integrated Development Environment – IDE), το οποίο αποτελεί λογισμικό για την ανάπτυξη των εφαρμογών. Η Python, συνοδεύεται από ένα απλό περιβάλλον ανάπτυξης με την ονομασία IDLE (Interactive DeveLopment Environment). Το περιβάλλον αυτό είναι γραμμένο σε Python από τον Guido Van Rossum, χρησιμοποιεί τη βιβλιοθήκη γραφικών Tkinter, οπότε μπορεί να εκτελεσθεί σε περιβάλλον Linux, Windows και Mac OS X. Δίνει τη δυνατότητα να χρησιμοποιείται διαδραστικά ο διερμηνευτής της γλώσσας, να γραφτούν και να επεξεργαστούν και να εκτελεστούν προγράμματα, καθώς επίσης εκτελείται η διαδικασία της αποσφαλμάτωσης (debugging). [54]. Υπάρχει πληθώρα περιβάλλοντων ανάπτυξης που διατίθενται για κάθε τύπο λειτουργικού συστήματος (Windows, Linux, Mac OS X και άλλα), μερικά από αυτά είναι το PyCharm, Spyder, IDLE, Microsoft Visual Studio, PyDev, Eclipse και άλλα.

## 5.2 Η συμβολή της γλώσσας Python στην ανάπτυξη του δικτύου: Google Colaboratory

---

Στην παρούσα εργασία, για την εκπόνησή της, χρησιμοποιείται το διαδραστικό περιβάλλον προγραμματισμού (Interactive Python Programming Environment), Google Colaboratory (Google Colab). Το Google Colab, αποτελεί ένα προϊόν της Google Research που επιτρέπει στους προγραμματιστές να γράφουν και να εκτελούν κώδικες γλώσσας προγραμματισμού Python, μέσω του προγράμματος περιήγησής τους. Το περιβάλλον αυτό αποτελεί εξαιρετικό εργαλείο για εργασίες που απαιτούν βαθιά μάθηση. Ουσιαστικά, είναι μία εξειδικευμένη μορφή του Jupyter notebook, που δεν χρειάζεται κάποια εγκατάσταση. Το Google Colab, αποτελεί ένα κατά βάση ανοιχτό περιβάλλον προγραμματισμού, προσφέροντας υπολογιστικούς πόρους, όπως GPU και TPU, οι οποίες αποτελούν μονάδες επεξεργασίας γραφικών και μειώνουν σημαντικά τον χρόνο που απαιτείται για την εκτέλεση ενός προγράμματος. Επιπροσθέτως, το Google Colab έχει ήδη εγκατεστημένες σχεδόν όλες τις απαραίτητες βιβλιοθήκες, οι οποίες είναι διαθέσιμες ανά πάσα στιγμή. Ένα ιδιαίτερο χαρακτηριστικό του προγράμματος είναι πως πραγματοποιεί αυτόματη αποθήκευση στο Cloud και έτσι δεν υπάρχει κίνδυνος να καταστραφεί ο αλγόριθμος που έχει ήδη υλοποιηθεί. Ακόμα, το Google Colab, επιτρέπει τη συνεργασία μεταξύ των προγραμματιστών για μία κοινή εργασία. [59].

## 5.3 Σημαντικές βιβλιοθήκες Python που χρησιμοποιήθηκαν

---

Ο λόγος που χρησιμοποιήθηκε η συγκεκριμένη γλώσσα προγραμματισμού είναι, πρώτον ο χαρακτήρας της ως μια ανοιχτού κώδικα γλώσσα προγραμματισμού, που σημαίνει πως για την εγκατάστασή της και την λειτουργία της έχει μηδενικό κόστος.

### 5.3.1 Scikit Learn

---

Το "scikit-learn" είναι μια δημοφιλής βιβλιοθήκη ανοικτού κώδικα για τη μηχανική μάθηση στη γλώσσα προγραμματισμού Python. Προσφέρει ένα ευρύ φάσμα αλγορίθμων μηχανικής μάθησης για ταξινόμηση, παλινδρόμηση, συσταδοποίηση, αναγνώριση συσχέτισης, εκτίμηση πυκνότητας και πολλά άλλα.

Αυτή η βιβλιοθήκη είναι πολύ δημοφιλής λόγω της ευκολίας χρήσης, της εκτενούς τεκμηρίωσης και της ενεργής κοινότητας που την υποστηρίζει. Επιπλέον, παρέχει εργαλεία για προεπεξεργασία δεδομένων, αξιολόγηση μοντέλων, και βελτιστοποίηση υπερπαραμέτρων.

Η scikit-learn χρησιμοποιείται ευρέως στη βιομηχανία, την εκπαίδευση και την έρευνα για την ανάπτυξη μοντέλων μηχανικής μάθησης σε ποικίλες εφαρμογές, όπως η αναγνώριση προτύπων, η κατηγοριοποίηση και η πρόβλεψη.

Η βιβλιοθήκη περιλαμβάνει αλγόριθμους όπως οι SVM (Μηχανές Διανυσμάτων Υποστήριξης), οι ταξινομητές k-NN (k Πλησιέστεροι Γείτονες), οι αλγόριθμοι k-Means (K-Μέση Τιμή), η λογιστική παλινδρόμηση, οι αναδρομικοί δέντρα (Decision Trees) και πολλοί άλλοι.

Συνοπτικά, η scikit-learn αποτελεί ένα ισχυρό εργαλείο για την ανάπτυξη και την εφαρμογή αλγορίθμων μηχανικής μάθησης στο περιβάλλον της Python.

Άλλες βιβλιοθήκες που χρησιμοποιήθηκαν ήταν η Pandas, η Numpy και η Matplotlib.

### 5.3.2 Pandas

---

Η βιβλιοθήκη **Pandas** [60] 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.

### 5.3.3 NumPy

---

Η **NumPy** [56], 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.

### 5.3.4 Matplotlib

---

Τέλος η βιβλιοθήκη **Matplotlib** [64] είναι μία βιβλιοθήκη της Python που έχει σχεδιαστεί για την παραγωγή διαγραμμάτων. Παρέχει αντικειμενοστραφή API, για ενσωματωμένα γραφήματα σε εφαρμογές χρησιμοποιώντας τις εργαλειοθήκες γενικής χρήσης του GUI. Η Matplotlib δημιουργήθηκε από τον John D. Hunter και έκτοτε έχει ενεργή κοινότητα ανάπτυξης και είναι διαθέσιμη με την άδεια ανοιχτού λογισμικού BSD [64]. Στην εργασία, χρησιμοποιείται με σκοπό να αποδοθούν τα απαραίτητα γραφήματα.

## 5.4 Υλοποίηση σε Python

---

Ο κώδικας που χρησιμοποιήθηκε στη παρούσα διπλωματική για την οπτικοποίηση των αποτελεσμάτων στο σετ ελέγχου, όπου πλοτάρει σε διάγραμμα διασποράς τις πραγματικές τιμές προς τις προβλεπόμενες από το δίκτυο, έχει αναλυτικά, όπως στο Παράρτημα Γ.

---

---

## Κεφάλαιο 6<sup>ο</sup> Πειραματική διαχείριση – Περιγραφή και επεξεργασία αρχικών δεδομένων

---

---

### 6.1 Εισαγωγικές έννοιες

---

Στο ακόλουθο κεφάλαιο θα γίνει αναλυτική περιγραφή της πηγής συλλογής των δεδομένων από κάθε διαθέσιμο φορέα. Πιο συγκεκριμένα συλλέχτηκαν Observation Files από επτά (7) σταθμούς ανά τον κόσμο (και πιο συγκεκριμένα στην Ανταρκτική, Αυστραλία, Χιλή, Κένυα, Νορβηγία Ηνωμένες Πολιτείες Αμερικής και Μογγολία) και δεδομένα για τις παραμέτρους επιρροής τιμών TEC, οι οποίες αναλύθηκαν στο Κεφάλαιο 3.

### 6.2 Περιγραφή των δεδομένων

---

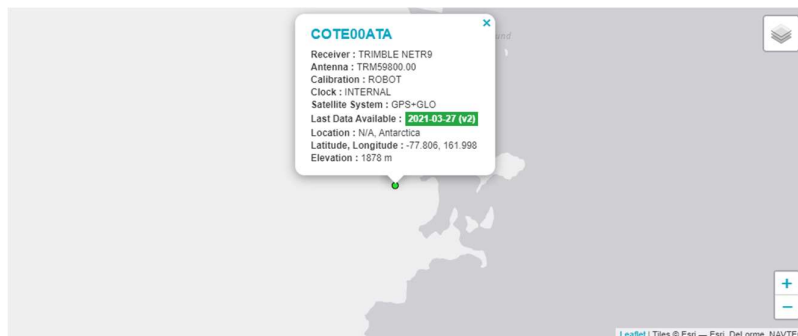
#### 6.2.1 Observation Files του σταθμού COTE00ATA (Ανταρκτική)

---

Στην παρούσα διπλωματική εργασία ακολουθήθηκαν κάποια συγκεκριμένα βήματα.

Αρχικά ένας από τους δορυφορικούς σταθμούς που επιλέχθηκε ήταν ο COTE00ATA, ο οποίος βρίσκεται στην Ανταρκτική και επιλέχθηκαν να αποθηκευτούν οι ημερησίες μετρήσεις του κατά τους μήνες Ιανουάριο, Φεβρουάριο, Μάρτιο και Απρίλιο του έτους 2018.

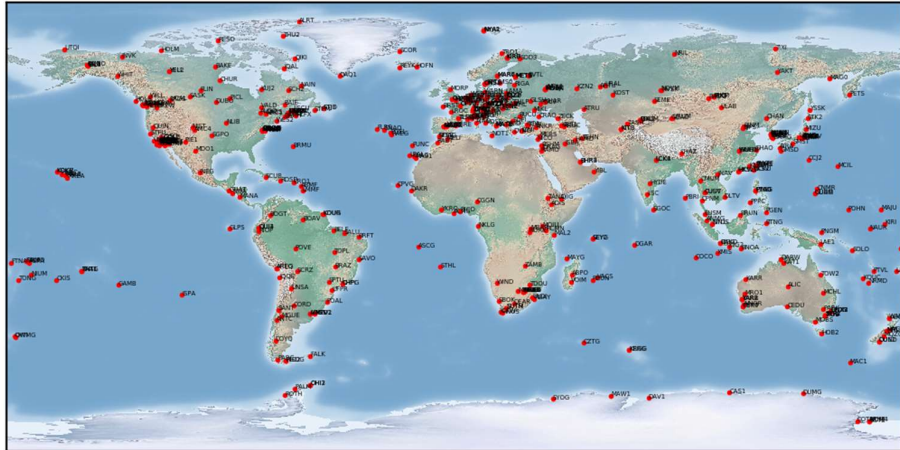
Πιο συγκεκριμένα πραγματοποιήθηκε download στα Rinex Observation files του σταθμού.



Εικόνα 11. 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 12. Όλοι οι διαθέσιμοι σταθμοί του συστήματος GNSS (Πηγή: IGS)

Για το σταθμό αυτό όπως και για τους υπόλοιπους, 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. 6.2.1.3) όπως φαίνεται και στην παρακάτω εικόνα (Εικόνα 15) 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. 6.2.1.4).





και εμπορικές εφαρμογές πράγμα που είναι ένα τεράστιο όφελος για το κοινό και ένα βασικό στοιχείο υποστήριξης για την επιστημονική πρόοδο.

- Μια 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.
- Παροχή δωρεάν και ανοιχτής πρόσβασης στα προϊόντα υψηλότερης ακρίβειας που είναι διαθέσιμα για επιστημονική πρόοδο και δημόσιο όφελος. Αυτά τα προϊόντα υποστηρίζουν μια μεγάλη ποικιλία εφαρμογών που αγγίζουν εκατομμύρια χρήστες σε σχεδόν όλα τα τμήματα της παγκόσμιας οικονομίας
- Παραγωγή προϊόντων που υποστηρίζουν την υλοποίηση του Διεθνούς Επίγειου Πλαισίου Αναφοράς παρέχοντας παράλληλα πρόσβαση σε δεδομένα παρακολούθησης από περισσότερους από 400 σταθμούς αναφοράς παγκοσμίως
- Εργασία για τη συνεχή ανάπτυξη νέων εφαρμογών και προϊόντων μέσω Ομάδων Εργασίας και Πιλοτικών Έργων
- Υποστήριξη γεωδαιτικής έρευνας και επιστημονικών δημοσιεύσεων
- Λειτουργία ως στοιχείο του Παγκόσμιου Συστήματος Γεωδαιτικής Παρατήρησης (GGOS) και μέλος του Παγκόσμιου Συστήματος Δεδομένων (WDS)

### 6.2.1.3 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.

#### 6.2.1.4 Πρόγραμμα GPS\_GOPi [29]

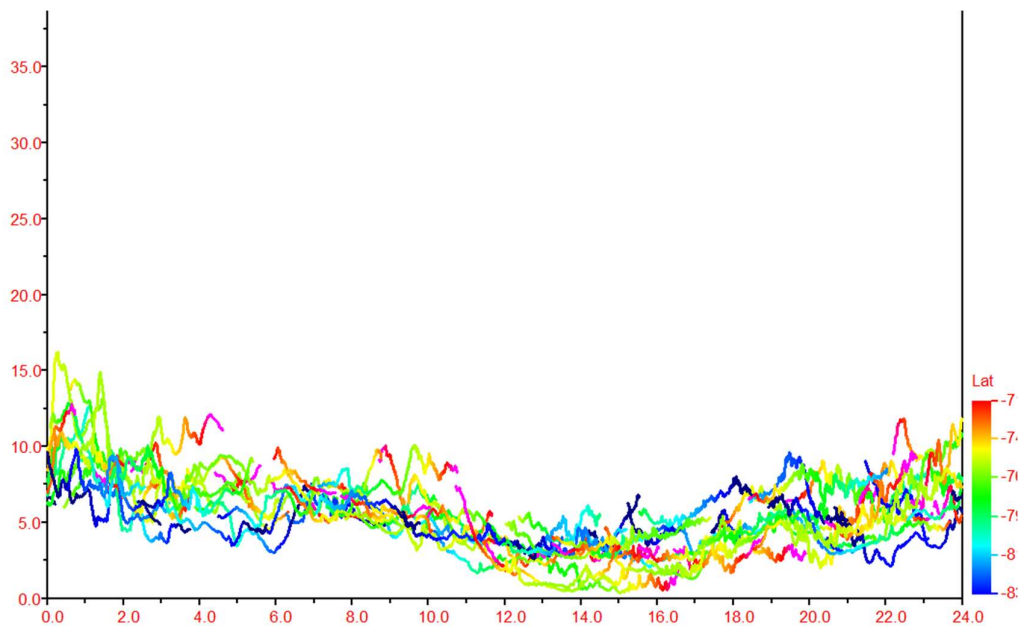
Εν συνεχεία με τη χρήση του προγράμματος GBS\_GOPi, πραγματοποιήθηκε η ανάλυση των αρχείων και η επεξεργασία τους για την εξαγωγή δεδομένων TEC.

Πιο συγκεκριμένα στο συγκεκριμένο πρόγραμμα υπάρχει η δυνατότητα ομαδικής επεξεργασίας των αρχείων εισόδου (μορφές RINEX 2 & 3 κ.λπ.) για παράδειγμα: όλα τα αρχεία του μήνα, έτος, όλοι οι σταθμοί και όλα τα αρχεία στον κατάλογο.

- Λαμβάνει ephemeris από το αρχείο πλοήγησης IGS, έχει τη δυνατότητα αυτόματης λήψης του αρχείου πλοήγησης εάν είναι συνδεδεμένο στο διαδίκτυο? εκτός αν βρει το αρχείο στον ίδιο κατάλογο με τα δεδομένα.
- Υπολογίζει το TEC από τα δεδομένα παρατήρησης του GPS Rinex, Novatel SCINDA και Leica.
- Ολισθήσεις κύκλου διεργασίας σε δεδομένα φάσης
- Ανάγνωση δορυφορικών προκαταλήψεων από αρχεία κωδικών DCB IGS, αν δεν είναι διαθέσιμα, τα υπολογίζει.
- Υπολογίζει την προκατάληψη του δέκτη
- Υπολογίζει τις διακαναλικές προκαταλήψεις για διαφορετικούς δορυφόρους στον δέκτη.
- Σχεδιάζει τις κάθετες τιμές TEC στην οθόνη και εγγράφει αρχεία εξόδου ascii (\*.CMN & \*.STD) στο ίδιο κατάλογο του αρχείου δεδομένων.

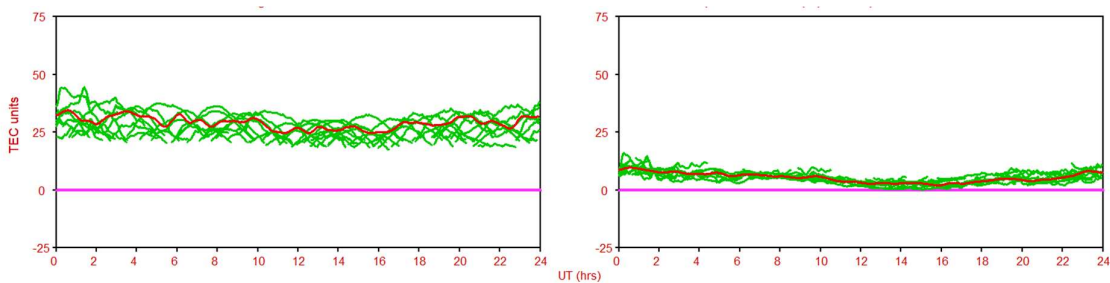
Τα δεδομένα TEC που παρήχθησαν στη συγκεκριμένη διπλωματική εργασία είναι ανά 30 sec.

Τα αποτελέσματα του συγκεκριμένου προγράμματος γίνονται στις παρακάτω εικόνες. Ενδεικτικά παρατίθενται τα αποτελέσματα του εν λόγω προγράμματος για την 01 Ιανουαρίου 2018. Πιο συγκεκριμένα στην Εικόνα 16 παρουσιάζεται η συμπεριφορά των τιμών TEC ανά ώρα με βάση το γεωγραφικό πλάτος.



Εικόνα 14. Συμπεριφορά των τιμών TEC ανά ώρα με βάση το γεωγραφικό πλάτος

Στη συνέχεια της επεξεργασίας των δεδομένων προκύπτουν τα αποτελέσματα της Εικόνας 17. Χαρακτηριστικά με την κόκκινη γραμμή απεικονίζεται ο μέσος όρος (μ.ο) της τιμής TEC για όλους τους δορυφόρους (Δ/Φ) ενώ με τις πράσινες γραμμές παρουσιάζεται η τιμή VTEC (vertical tec) για κάθε Δ/Φ ξεχωριστά. Η διαφορά των δύο διαγραμμάτων είναι πως στο δεξιά διάγραμμα έχουν αφαιρεθεί οι «επιρροές» των RX ανά Δ/Φ.

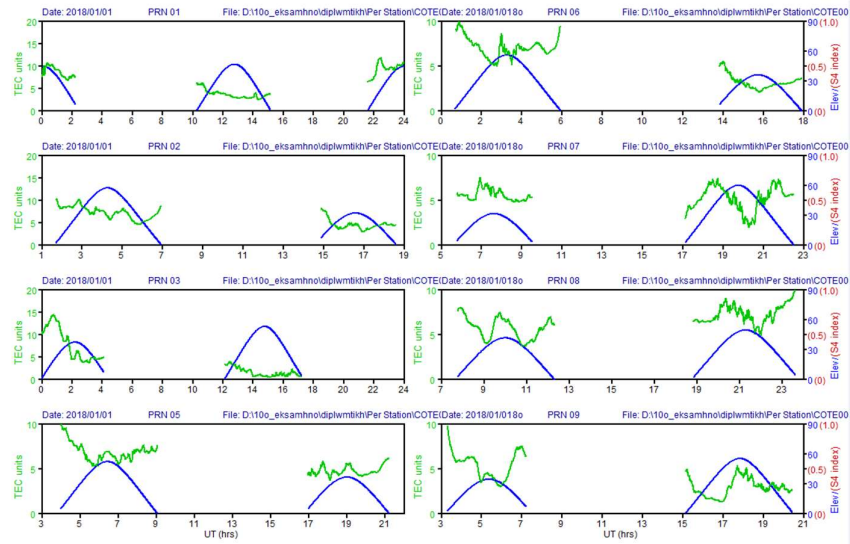


Εικόνα 15. Τιμές TEC ανά Δορυφόρο

Η επεξεργασία δεδομένων GPS για απόλυτη απόδοση τιμών TEC απαιτεί καλό υπολογισμό της «επιρροής» RX, τα οποία όμως περιλαμβάνουν ορισμένες υποθέσεις που μπορεί να θέσουν σε κίνδυνο την πραγματική ποιότητα των δεδομένων.

Γι αυτό το λόγο προτιμητέο είναι το αποτέλεσμα της διακύμανσης της τιμής TEC του πρώτου διαγράμματος, το οποίο αποδίδει την απόλυτη διακύμανση της τιμής TEC σε σχέση με την ώρα.

Τέλος η τελική οπτικοποίηση της φάσης της επεξεργασίας των δομένων του επιλεγμένου σταθμού, είναι τα διαγράμματα της Εικόνας 18. Σε αυτήν φαίνεται η διακύμανση της τιμής TEC (πράσινες γραμμές) την 01 Ιανουαρίου 2018, ανά Δ/Φ (PRN 01, 02 κλπ) σε κάποιο εύρος υψομέτρων.



Εικόνα 16. Η διακύμανση της τιμής TEC την 01 Ιανουαρίου 2018, ανά  $\Delta/\Phi$  (PRN 01, 02 κλπ) σε κάποιο εύρος υψομέτρων.

## 6.2.2 Παράμετροι επιρροής τιμών Tec

Όπως υποδηλώνει και ο τίτλος της παρούσας διπλωματικής εργασίας για την πρόβλεψη της ιονοσφαιρικής δραστηριότητας, χρησιμοποιήθηκαν ως τιμές εισόδου, τιμές των παραμέτρων επιρροής τιμών TEC, οι οποίες αναλύθηκαν στο Κεφάλαιο 3. Δεδομένα για αυτές τις παραμέτρους συλλέχθηκαν από τις δομές που θα περιγραφούν στα ακόλουθα υποκεφάλαια. Όλα τα δεδομένα που συλλέχθηκαν, καταγράφηκαν σε ένα αρχείο excel, το οποίο αποτέλεσε την κύρια πηγή δεδομένων εισόδου για το προς εκπαίδευση νευρωνικό δίκτυο. Παρακάτω παρατίθεται ενδεικτική εικόνα από το εν λόγω πρόγραμμα εργασίας. Όλα τα δεδομένα που χρησιμοποιήθηκαν σαν είσοδο στο νευρωνικό δίκτυο παρατίθενται στο Παράρτημα Α της παρούσας διπλωματικής εργασίας.

Πίνακας 2. Excel με τα δεδομένα των παραμέτρων που χρησιμοποιήθηκαν

Date	Time	STEC	VTEC	GEOMAGNETIC					Solar Radio Flux					Sunspot Number (3 hours)								
				Kp (3 hour)	Ap	D	SMI	SMU	F10.7	F10.7min	F10.7max	F10.7DF	F10.7SI	Daily Total Sunspot Number	Daily North Sunspot Number	Daily South Sunspot Number	Standard Deviation of Raw Daily Total Sunspot Data	Standard Deviation of Raw Daily Sunspot Data	Number of Observations for Daily Sunspot Number	Number of Observations for Daily Sunspot		
0	13.137	5.032396	4	2.867	12	1	-98.39444	53.3	0	*2458149.9875	2200.45	89.15	73.25	62.825	12	12	0	1.1	0.3	0	39	6
3	15.9284	6.285032	4	0.333	2	1	-50.63333	39.083333	3	*2458149.9875	2200.4275	86.675	73.125	77.1	12	12	0	1.1	0.3	0	39	6
6	12.5992	5.797062	4	0.333	2	1	-52.03667	40.488889	6	*2458149.9875	2200.405	84.2	74	75.95	12	12	0	1.1	0.3	0	39	6
9	8.99051	5.917603	4	0.333	2	1	-26.91211	45.716667	9	*2458150.1125	2200.3825	81.725	77.875	73.8	12	12	0	1.1	0.3	0	39	6
12	7.10249	2.939651	4	0.333	2	1	-26.33889	34.216667	12	*2458150.4885	2200.36	79.25	76.75	72.15	12	12	0	1.1	0.3	0	39	6
15	7.446246	3.124225	4	0	0	1	-21.97222	41.338889	15	*2458150.8615	2200.3375	76.75	75.025	70.5	12	12	0	1.1	0.3	0	39	6
18	10.1994	4.74801	4	1.333	5	1	-70.35556	44.488889	18	2458151.239	2200.205	69.3	67.3	60.5	12	12	0	1.1	0.3	0	39	6
21	12.4810	5.848979	4	1.333	5	1	-96.43889	52.048889	21	2458151.364	2200.2005	68.5	67.15	57.5	12	12	0	1.1	0.3	0	39	6
0	19.0501	6.127995	4	7	1	1	-114.7	56.794444	0	2457985.548	2194.11645	84.15	79.95	72.5	15	15	0	1.3	0.6	0	42	5
3	14.5009	6.742027	4	1	4	1	-47.8	61.133333	3	2458008.165	2194.94735	82.125	78.225	70.95	15	15	0	1.3	0.6	0	42	5
6	13.3776	6.099511	4	0.667	3	1	-43.23222	70.835556	6	2458030.782	2198.7788	80.1	78.5	69.5	15	15	0	1.3	0.6	0	42	5
9	10.2874	4.846613	4	0.667	3	1	-60.40556	47.405556	9	2458053.399	2196.60975	78.075	74.775	67.65	15	15	0	1.3	0.6	0	42	5
12	7.64487	3.548722	4	0.667	3	1	-98.63333	31.25	12	2458076.288	2197.44075	76.6	75.05	66	15	15	0	1.3	0.6	0	42	5
15	15.84751	4.605064	4	0.333	2	1	-50.47778	29.983333	15	2458098.613	2198.27125	74.025	64.35	64.35	15	15	0	1.3	0.6	0	42	5
18	10.4218	5.048906	4	0.333	2	1	-48.17222	39.488889	18	2458121.25	2199.1029	72	69.6	62.7	15	15	0	1.3	0.6	0	42	5
21	12.73279	6.137142	4	0	0	1	-17.06667	27.016667	21	2458151.364	2200.2005	69.25	67.15	60.4	15	15	0	1.3	0.6	0	42	5
0	14.8046	6.999549	4	0.333	2	1	-20.96667	26.45	0	2458156.988	2200.413	75.75	74.3	66.8	16	16	0	2.1	0	0	39	9
3	16.2381	7.13952	4	0.333	2	1	-23.86667	27.311111	3	2458156.988	2200.3905	74.75	73.25	67.7	16	16	0	2.1	0	0	39	9
6	16.2879	7.97222	3	1	4	1	-51.96667	27.4	6	2458155.738	2200.368	73.8	72.2	64.8	16	16	0	2.1	0	0	39	9
9	12.4992	5.861296	3	1	4	1	-93.96667	30.1	9	2458155.113	2200.3455	72.825	71.15	63.9	16	16	0	2.1	0	0	39	9
12	9.00606	5.987227	3	0.333	2	1	-28.16667	28.37778	12	2458154.489	2200.313	71.85	70.1	69	16	16	0	2.1	0	0	39	9
15	8.07333	3.289519	3	0.333	2	1	-49.22778	21.411111	15	2458153.864	2200.3005	70.875	69.05	62.1	16	16	0	2.1	0	0	39	9
18	11.0977	5.26291	3	1.667	6	1	-92.10556	25.22222	18	2458153.864	2200.278	69.8	68.2	61.2	16	16	0	2.1	0	0	39	9
21	14.6126	6.702707	3	1.333	5	1	-52.63333	21.666667	21	2458152.614	2200.2555	68.925	66.95	60.3	16	16	0	2.1	0	0	39	9
0	10.32241	6.942085	3	0.667	3	1	-43.13889	25.911111	0	2458157.988	2200.45	71.4	69.4	62.6	0	0	0	0	0	0	41	33
3	15.9489	7.327241	3	0	0	1	-23.97222	27.066667	3	2458157.988	2200.4275	71.25	69.25	62.45	0	0	0	0	0	0	41	33
6	15.88848	7.114541	3	0	0	1	-52.17222	24.288889	6	2458156.738	2200.405	71.1	69.1	62.3	0	0	0	0	0	0	41	33
9	11.4899	5.03209	3	0.333	2	1	-29.20556	24.488889	9	2458156.113	2200.3825	70.95	68.95	62.15	0	0	0	0	0	0	41	33
12	7.88413	3.044263	3	0.333	2	1	-16.47778	26.272222	12	2458155.489	2200.36	70.8	68.8	62	0	0	0	0	0	0	41	33
15	8.82979	3.718811	3	1	4	1	-83.13889	28.311111	15	2458154.864	2200.3775	70.65	68.65	61.85	0	0	0	0	0	0	41	33
18	9.64574	4.191919	3	2.333	9	1	-128.6111	40.944444	18	2458154.219	2200.315	70.5	68.5	61.7	0	0	0	0	0	0	41	33
21	13.9346	6.750819	3	1	4	1	-16.91667	30.244444	21	2458153.614	2200.2925	70.35	68.35	61.55	0	0	0	0	0	0	41	33
0	18.3006	5.02296	3	2.667	12	1	-118.0778	44.494444	0	2458153.864	2200.2925	70.35	68.35	61.55	0	0	0	0	0	0	41	33
3	18.8209	6.916178	3	7	7	1	-97.4	42.216667	3	2458153.864	2200.2925	70.35	68.35	61.55	0	0	0	0	0	0	41	33
6	17.839	6.877532	3	2	7	1	-136.3367	51.166667	6	2458157.738	2200.445	84.2	81.8	73.75	0	0	0	0	0	0	47	41
9	15.8034	6.100545	3	1.667	6	1	-145.7778	52.066667	9	2458157.125	2200.4275	81.725	79.4	71.825	0	0	0	0	0	0	47	41
12	7.43351	4.275282	3	15	1	1	-38.45	71.877778	12	2458156.489	2200.3985	79.25	77	69.95	0	0	0	0	0	0	47	41
15	6.80488	3.818205	3	2.333	9	1	-125.7944	51.361111	15	2458155.489	2200.3729	78.75	76	67.15	0	0	0	0	0	0	47	41
18	9.19217	5.748952	3	1	4	1	-88.22778	34.655556	18	2458155.239	2200.352	74.1	72.2	65	0	0	0	0	0	0	47	41
21	12.2953	7.084651	3	0.333	2	1	-14.13333	25.4	21	2458154.614	2200.3275	71.825	69.8	64.8	0	0	0	0	0	0	47	41

Για το σωστό έλεγχο και την αξιοπιστία του προς εκπαίδευση νευρωνικού δικτύου αποφασίστηκε η πραγματοποίηση αρκετών σεναρίων, κατά τα οποία θα προστίθενται κάθε φορά ή θα αφαιρούνταν κάποιες από τις παραμέτρους που προαναφέρθηκαν, προκειμένου να προσδιοριστεί ποιες από αυτές έχουν το μεγαλύτερο βάρος ή αλλιώς διαδραματίζουν τον πιο καθοριστικό ρόλο για την καλύτερη πρόβλεψη της ιονόσφαιρας (τιμές STEC, VTEC).

### 6.2.2.1 Geomagnetic activity (K)

Τα δεδομένα γεωμαγνητικής δραστηριότητας, παράμετρος K (δείκτες Kp, Ap, ap 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 μια 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):
#iii ii ff, ff, ff, fffff, fffff, fffff, fffff ff, ffff iiiii i
# The parameters in each line are:
#YYY MM DD hh, h hh, m_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, 000 4 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
.....
```

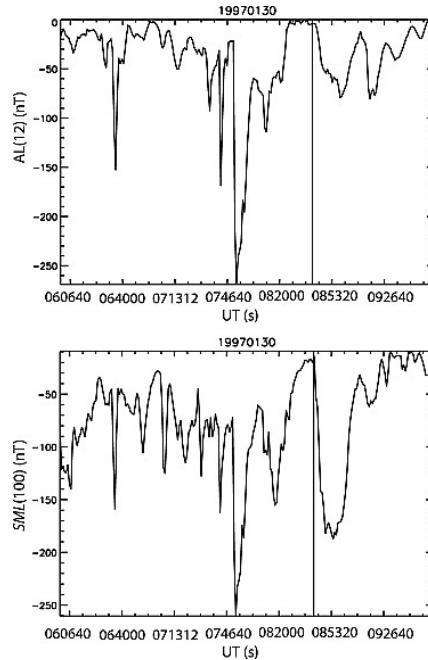
Εικόνα 17. Υπόδειγμα δεδομένων γεωμαγνητικής δραστηριότητας (K)

Το τρίωρο ισοδύναμο πλανητικό πλάτος ap προέρχεται από το Kp και το ημερήσιο ισοδύναμο πλανητικό πλάτος Ap είναι ο ημερήσιος μέσος όρος του ap. 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  $K_p$ ,  $ap$  και  $Ap$  εισήχθησαν από τον Bartels (1949, 1957). Το  $D$  υποδεικνύει εάν οι τιμές  $K_p$  και  $ap$  είναι οριστικές ( $D = 1$ ) ή προκαταρκτικές ( $D=0$ ) (Matzka et al, (2021)).

### 6.2.2.2 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).



Εικόνα 18. 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 19. Υπόδειγμα δεδομένων SME

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.

### 6.2.2.3 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

Αυτή η μορφή προσορίζεται 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 -5
1/1/2018	1:00:00 πμ		1 -7
1/1/2018	2:00:00 πμ		1 -13
1/1/2018	3:00:00 πμ		1 -9
1/1/2018	4:00:00 πμ		1 -12
1/1/2018	5:00:00 πμ		1 -18
1/1/2018	6:00:00 πμ		1 -20
1/1/2018	7:00:00 πμ		1 -20
1/1/2018	8:00:00 πμ		1 -17
1/1/2018	9:00:00 πμ		1 -18
1/1/2018	10:00:00 πμ		1 -16
1/1/2018	11:00:00 πμ		1 -12
1/1/2018	12:00:00 μμ		1 -9
1/1/2018	1:00:00 μμ		1 -9
1/1/2018	2:00:00 μμ		1 -15
1/1/2018	3:00:00 μμ		1 -17
1/1/2018	4:00:00 μμ		1 -15
1/1/2018	5:00:00 μμ		1 -13
1/1/2018	6:00:00 μμ		1 -12
1/1/2018	7:00:00 μμ		1 -10
1/1/2018	8:00:00 μμ		1 -9
1/1/2018	9:00:00 μμ		1 -9
1/1/2018	10:00:00 μμ		1 -8
1/1/2018	11:00:00 μμ		1 -7

Εικόνα 20. Υπόδειγμα δεδομένων Dst



#### 6.2.2.4 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](http://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 =  $10^{-22}$  W m<sup>-2</sup> Hz<sup>-1</sup>). Τα αρχεία είναι txt της μορφής:

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

Solar Radio Flux					
fluxtime	fluxjulian	fluxarrington	fluxobsflux	fluxadjflux	fluxursi
0	2458149.9875	2200.45	89.15	73.25	62.825
3	2458149.36275	2200.4275	86.675	73.125	77.1
6	2458149.738	2200.405	84.2	74	75.45
9	2458150.11325	2200.3825	81.725	77.875	73.8
12	2458150.4885	2200.36	79.25	76.75	72.15
15	2458150.86375	2200.3375	76.775	75.625	70.5
18	2458151.239	2200.205	69.3	67.3	60.5
21	2458151.364	2200.2095	69.15	67.15	60.4
0	2457985.548	2194.11645	84.15	79.95	72.6
3	2458008.165	2194.947525	82.125	78.225	70.95
6	2458030.782	2195.7786	80.1	76.5	69.3
9	2458053.399	2196.609675	78.075	74.775	67.65
12	2458076.016	2197.44075	76.05	73.05	66
15	2458098.633	2198.271825	74.025	71.325	64.35
18	2458121.25	2199.1029	72	69.6	62.7
21	2458151.364	2200.2095	69.15	67.15	60.4
0	2458156.988	2200.413	75.75	74.3	66.6
3	2458156.363	2200.3905	74.775	73.25	65.7
6	2458155.738	2200.368	73.8	72.2	64.8
9	2458155.113	2200.3455	72.825	71.15	63.9
12	2458154.489	2200.323	71.85	70.1	63
15	2458153.864	2200.3005	70.875	69.05	62.1
18	2458153.239	2200.278	69.9	68	61.2
21	2458152.614	2200.2555	68.925	66.95	60.3
0	2458157.988	2200.45	71.4	69.4	62.6
3	2458157.363	2200.4275	71.25	69.25	62.45
6	2458156.738	2200.405	71.1	69.1	62.3
9	2458156.113	2200.3825	70.95	68.95	62.15
12	2458155.489	2200.36	70.8	68.8	62
15	2458154.864	2200.3375	70.65	68.65	61.85
18	2458154.239	2200.315	70.5	68.5	61.7
21	2458153.614	2200.2925	70.35	68.35	61.55
0	2458156.988	2200.4915	89.15	86.6	78.05
3	2458158.363	2200.46825	86.675	84.2	75.875
6	2458157.738	2200.445	84.2	81.8	73.7
9	2458157.113	2200.42175	81.725	79.4	71.525
12	2458156.489	2200.3985	79.25	77	69.35
15	2458155.864	2200.37525	76.775	74.6	67.175

Εικόνα 21. Υπόδειγμα δεδομένων Solar Radio Flux

### 6.2.2.5 Sunspot Number

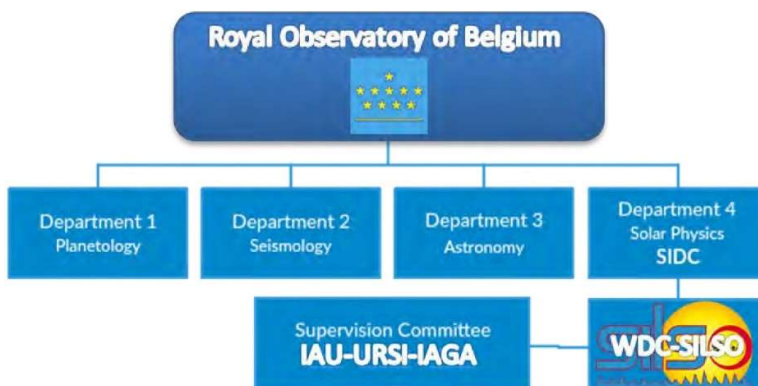
Τα δεδομένα που αφορούσαν το Sunspot Number (ηλιακή κηλίδα), ελήφθησαν από WDC-SILSO, Βασιλικό Αστεροσκοπείο του Βελγίου στις Βρυξέλλες. Ο ημερήσιος συνολικός αριθμός ηλιακών κηλίδων προκύπτει από τον τύπο:  $R = N_s + 10 * N_g$ , με  $N_s$  τον αριθμό των κηλίδων και  $N_g$  τον αριθμό των ομάδων που 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	Daily South Sunspot Number
2020-04-30	20.0	8.0	12.0
2020-04-29	24.0	12.0	12.0
2020-04-28	15.0	0.0	15.0
2020-04-27	29.0	0.0	29.0
2020-04-26	14.0	0.0	14.0
2020-04-25	6.0	0.0	6.0
2020-04-24	0.0	0.0	0.0
2020-04-23	0.0	0.0	0.0
2020-04-22	0.0	0.0	0.0
2020-04-21	0.0	0.0	0.0
2020-04-20	0.0	0.0	0.0
2020-04-19	0.0	0.0	0.0
2020-04-18	0.0	0.0	0.0
2020-04-17	0.0	0.0	0.0
2020-04-16	0.0	0.0	0.0
2020-04-15	0.0	0.0	0.0
2020-04-14	0.0	0.0	0.0
2020-04-13	0.0	0.0	0.0
2020-04-12	0.0	0.0	0.0
2020-04-11	0.0	0.0	0.0
2020-04-10	0.0	0.0	0.0
2020-04-09	0.0	0.0	0.0
2020-04-08	0.0	0.0	0.0
2020-04-07	0.0	0.0	0.0

Εικόνα 22. Υπόδειγμα δεδομένων Sunspot Number



Εικόνα 23. Βασιλικό Αστεροσκοπείο Βελγίου (Πηγή: [www.bis.sidc.be](http://www.bis.sidc.be))

### 6.2.2.6 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 ενέργειας όπως παρατηρούνται από τον κύριο δορυφόρο GOES.



### 6.2.2.8 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 28.

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 Time [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 <sup>2</sup> ]	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 02:05	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 20:24:05	191	39	241	394	782	25.0*1	5.8e+13	1.7e+28	192	C2 C3 PHTX DST Java	Poor Event

Εικόνα 26. Υπόδειγμα δεδομένων Coronal Mass Ejections.

## 6.3 Επεξεργασία των δεδομένων

Τα δεδομένα όπως περιγράφηκαν 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, δημιουργήθηκαν επτά (7) αρχεία .xlsx (ένα για κάθε δορυφορικό σταθμό) τα οποία περιέχουν το κάθε ένα μία καρτέλα για τους μήνες Ιανουάριο – Φεβρουάριο – Μάρτιο – Απρίλιο. (Παράρτημα 1).

---

---

## Κεφάλαιο 7<sup>ο</sup> Χρήση παραμέτρων δεδομένων TEC – Σενάρια και αξιολόγηση αυτών

---

---

Το αποθετήριο δεδομένων περιέχει τιμές από οκτώ (8) παραμέτρους, όπως αυτές παρουσιάστηκαν στον Πίνακα 1 του Κεφαλαίου 3. Για την πρόβλεψη της ιονοσφαιρικής δραστηριότητας χρησιμοποιήθηκαν τα κάτωθι μοντέλα μηχανικής μάθησης (τα οποία παρουσιάστηκαν αναλυτικά στο Κεφάλαιο 4):

- Linear Regression
- Decision Tree
- Support Vector Regression
- Gaussian Process Regressor
- XGBOOST Regressor
- Multi-layer Perception Regressor
- LightGBM Regressor
- Random Forest Regressor

Η κύρια τεχνική που χρησιμοποιήθηκε στην παρούσα διπλωματική για την ερμηνεία και την εξήγηση προβλέψεων που παράγονται από μοντέλα μηχανικής μάθησης είναι η SHAP το αρκτικό για "SHapley Additive exPlanations. Αυτή η τεχνική βασίζεται στον αλγόριθμο Shapley values, ο οποίος προέρχεται από τη θεωρία των παιγνίων και χρησιμοποιείται για να αναδείξει τη σημαντικότητα κάθε χαρακτηριστικού στην πρόβλεψη ενός μοντέλου.

Η τεχνική SHAP είναι χρήσιμη για την αποσαφήνιση των αποφάσεων των μοντέλων μηχανικής μάθησης (όπως αυτά αναφέρθηκαν ανωτέρω), την ανίχνευση προκαταλήψεων και τη βελτιστοποίηση της εμπιστοσύνης στα αποτελέσματά τους. Επίσης, μπορεί να χρησιμοποιηθεί για την επιλογή χαρακτηριστικών και τη βελτιστοποίηση μοντέλων. Είναι ένα ισχυρό εργαλείο για την εξήγηση της λειτουργίας μοντέλων μηχανικής μάθησης και τη βελτίωση της κατανόησης τους.

Στη παρούσα εργασία, ο στόχος του SHAP είναι να εξηγήσει την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας (βλ. Πίνακα 1 υποκεφάλαιο 3.5, πχ Fluxtime κλπ) στην πρόβλεψη. Η μέθοδος επεξήγησης SHAP υπολογίζει τις τιμές Shapley από τη θεωρία συνασπισμών παιγνίων. Μια καινοτομία που φέρνει το SHAP είναι ότι η εξήγηση της τιμής Shapley αναπαρίσταται ως μια μέθοδος απόδοσης πρόσθετων χαρακτηριστικών, ένα γραμμικό μοντέλο. Αυτή η προβολή συνδέει τις τιμές LIME και Shapley.

Για την ερμηνεία της ανωτέρω τεχνικής προκύπτουν τα γραφήματα όπως αυτά παρουσιάστηκαν στα υποκεφάλαια 4.3.2, 4.3.3, 4.3.4 και θα αναλυθούν παρακάτω.

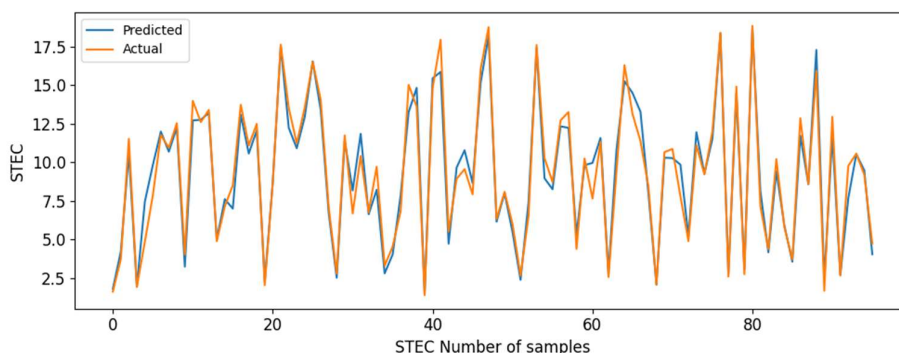
Στα επόμενα υποκεφάλαια παρατίθενται αναλυτικά η τεχνική SHAP που εφαρμόστηκε στα δεδομένα και των επτά (7) δορυφορικών σταθμών.

## 7.1 Δορυφορικός Σταθμός Ανταρκτικής

Στην ενότητα αυτή, ακολουθεί η παρουσίαση των διαγραμμάτων που προέκυψαν από την τακτική SHAP τα οποία παρουσιάζουν την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας.

### 7.1.1 STEC

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Εικόνα 27. Ποιότητα σφάλματος στα σετ δεδομένων αξιολόγησης και εκπαίδευσης

Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές STEC. Στον άξονα των x υπάρχει ο αριθμός των παρατηρήσεων μας δηλ το σετ δεδομένων ελέγχου και στον άξονα των ψ οι τιμές STEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται πολύ καλά.

#### 7.1.1.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.



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.

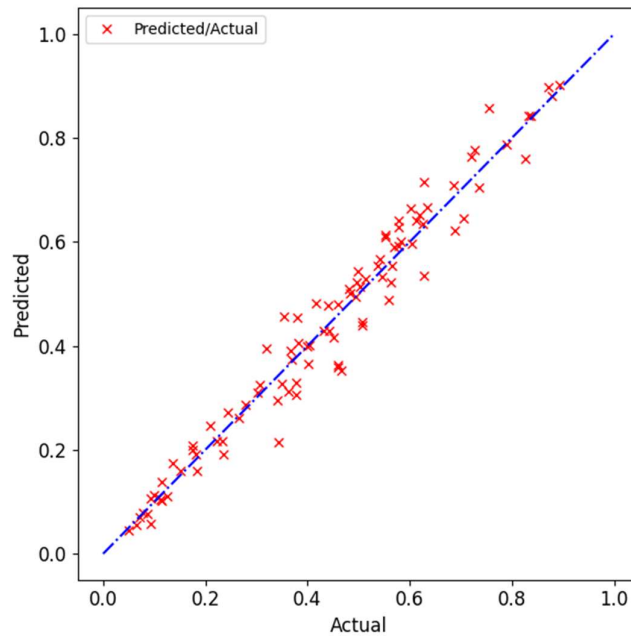
Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του XGBOOST Regressor και του Random Forrest, καθώς οι τιμές των MAE, rMSE και του  $R^2$  είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	$R^2$
Linear Regrassion	3.0724	2.5099	0.5536
Decision Tree	2.6988	2.2094	0.6556
Support Vector Regression	1.9674	1.5115	0.8170
Gaussian Process Regressor	3.0556	2.5006	0.5585
XGBOOST Regressor	0.9103	0.6856	0.9608
Multi-layer Perceptron regressor	1.2483	0.9729	0.9263
LightGBM regressor	1.0163	0.7632	0.9512
Random Forest Regressor	0.0725	0.0490	0.8973

*Εικόνα 28. Αποτελέσματα μοντέλων εξηγήσιμης μάθησης με βάση τους δείκτες rMSE, MAE και  $R^2$*

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθόσον όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική καθώς δεν υπάρχουν χωρικά ακρότατα.



Εικόνα 29. Γραμμική παλινδρόμηση

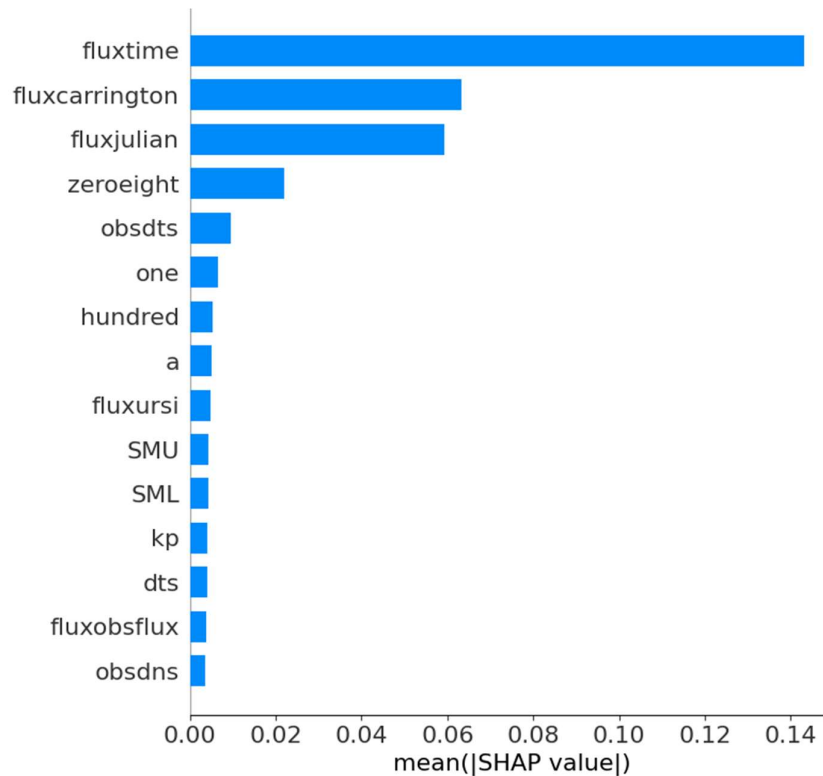
#### 7.1.1.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.





Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής STEC κατά 14 ποσοστιαίες μονάδες (0,14 στον άξονα x).

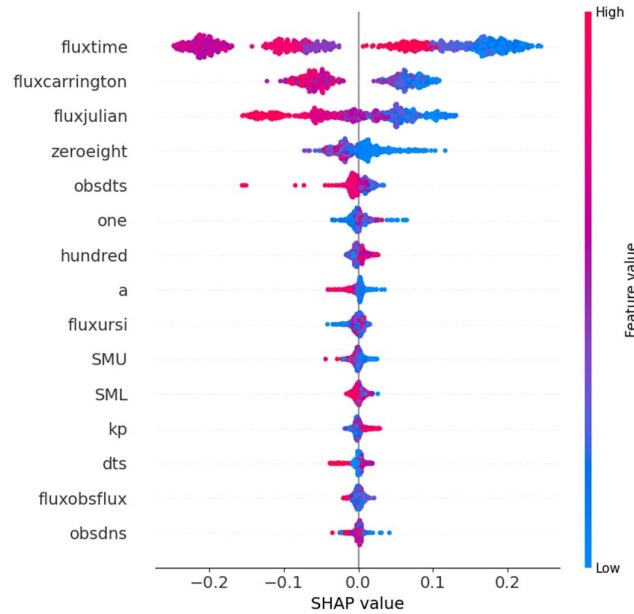
Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.1.1.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια

αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν την επίδραση στις τιμές STEC, ενώ οι χαμηλές τιμές αυξάνουν την επίδραση.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

### 7.1.1.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

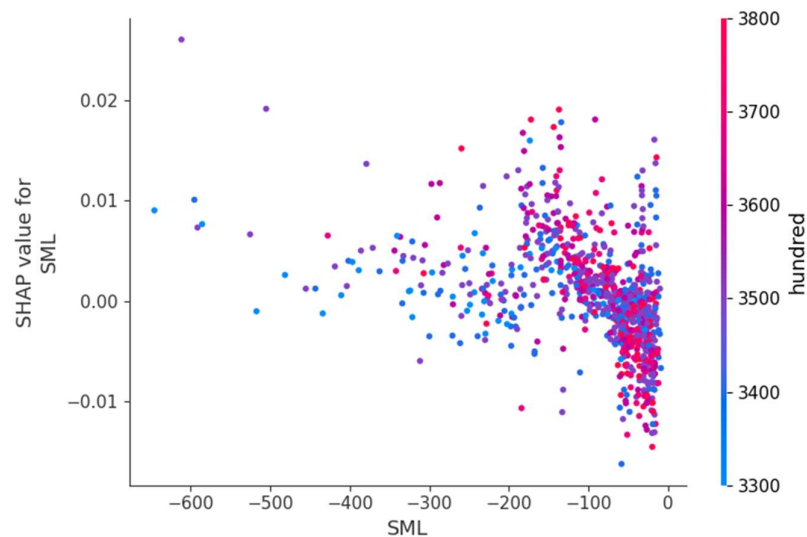
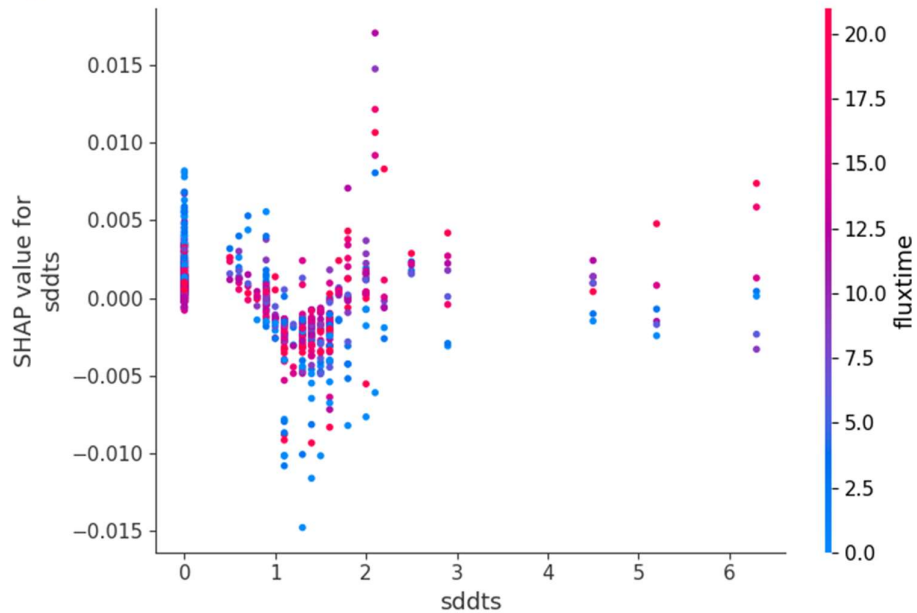
when  $i \neq j$  and:

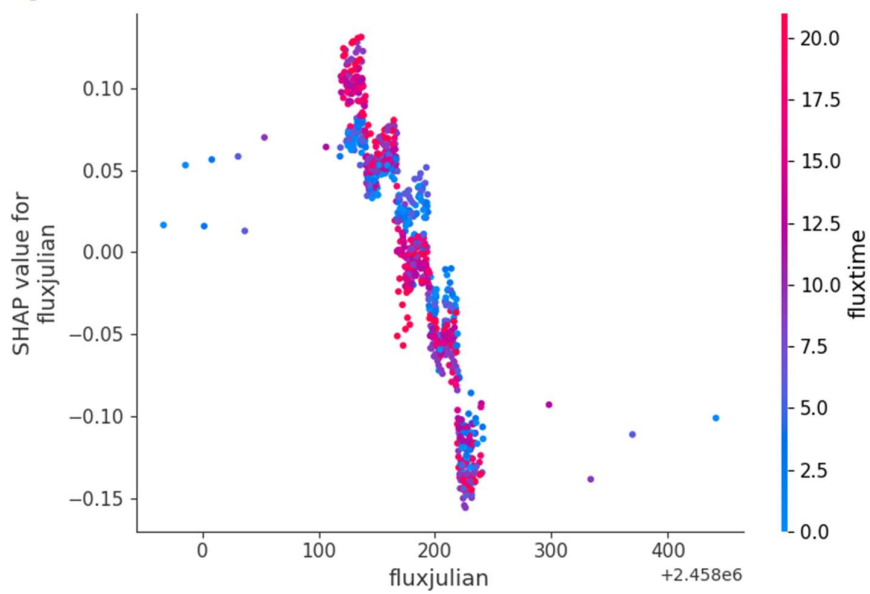
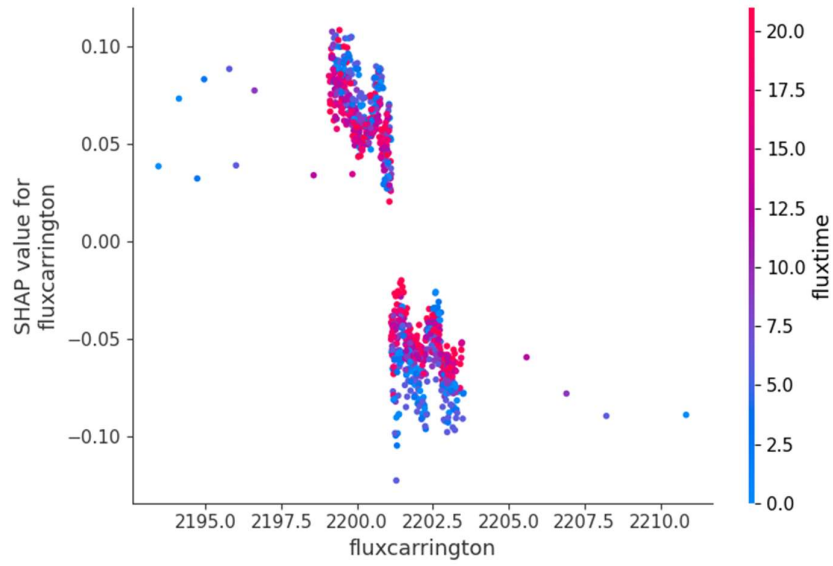
$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα

αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:





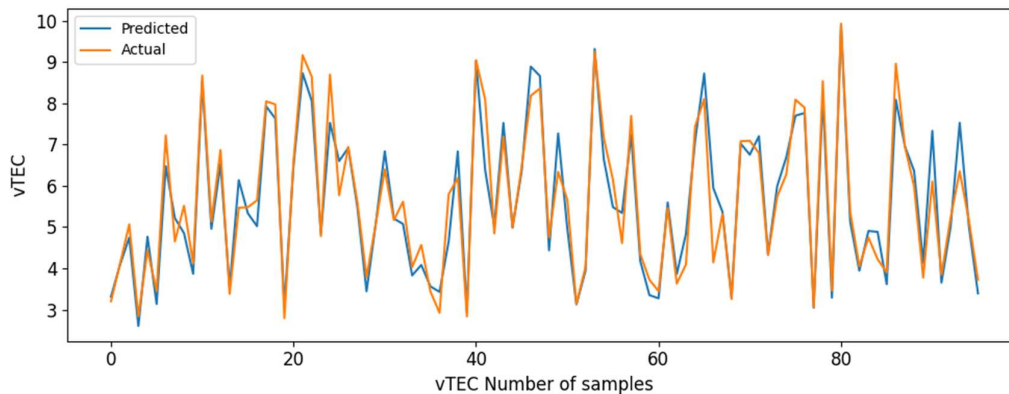
Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

Στο τελευταίο σχεδιάγραμμα η παράμετρος fluxjulian αλληλοεπιδρά με την παράμετρο fluxtime. Σε περιπτώσεις κοντά στην τιμή 0 fluxjulian, η εμφάνιση της παραμέτρου fluxtime μειώνει την επίδραση στο STEC, καθώς υπάρχουν ελάχιστα χωρικά ακρότατα. Για μεγαλύτερης τιμές παραμέτρου fluxjulian (100-300), η εμφάνιση της παραμέτρου fluxtime διατηρεί σταθερή την επίδραση στο STEC καθώς η μονοτονία του διαγράμματος είναι σταθερή και οι υψηλές καθώς και οι χαμηλές τιμές της παραμέτρου fluxtime έχουν το ίδιο βάρος επίδρασης στη παράμετρο fluxjulian.

### 7.1.2 VTEC

Αντίστοιχα για τη συμπεριφορά του μοντέλου για τις τιμές VTEC προκύπτουν τα ακόλουθα.

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές VTEC. Στον άξονα των x υπάρχει ο αριθμός των παρατηρήσεων μας δηλ το σετ δεδομένων ελέγχου και στον άξονα των ψ οι τιμές VTEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται αρκετά καλά, όχι όμως όπως οι αντίστοιχες του STEC.

7.1.2.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

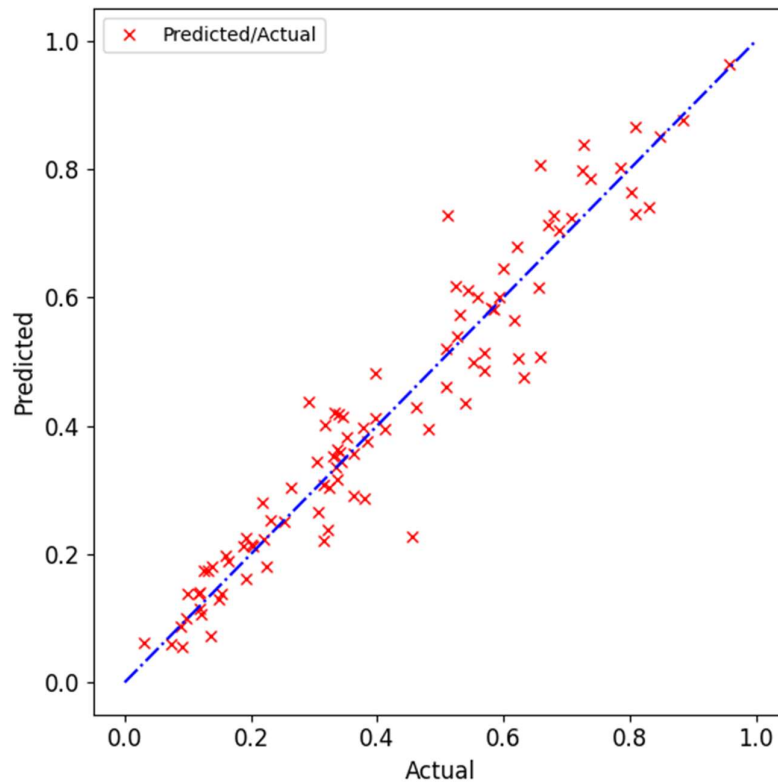
Το μέσο απόλυτο σφάλμα (mae) αντιπροσωπεύει τον μέσο όρο της απόλυτης διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τον μέσο όρο των υπολειμμάτων στο σύνολο δεδομένων.

Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του XGBOOST Regressor και του Random Forrest, καθώς οι τιμές των MAE, rMSE και του R<sup>2</sup> είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	R <sup>2</sup>
Linear Regression	1.3748	1.0877	0.4281
Decision Tree	0.9830	0.8389	0.7076
Support Vector Regression	0.9341	0.6896	0.7360
Gaussian Process Regressor	1.3638	1.0819	0.4372
XGBOOST Regressor	0.5136	0.3746	0.9202
Multi-layer Perceptron regressor	0.6051	0.4847	0.8892
LightGBM regressor	0.5348	0.3826	0.9135
Random Forest Regressor	0.0765	0.0561	0.8903

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθώς όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική. Παρατηρούνται βέβαια κάποια χωρικά ακρότατα τα οποία υποδεικνύουν ότι το μοντέλο δε συγκλίνει όπως στο STEC.

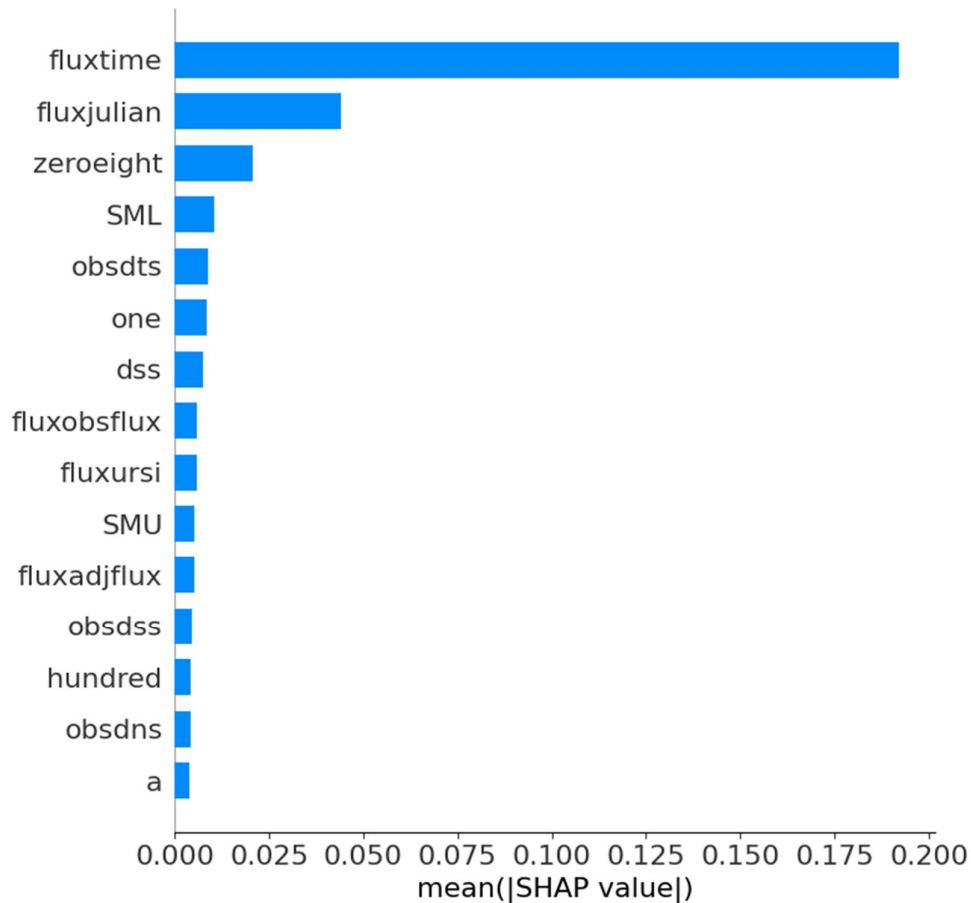


#### 7.1.2.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής VTEC κατά 20 ποσοστιαίες μονάδες (0,2 στον άξονα x).

Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

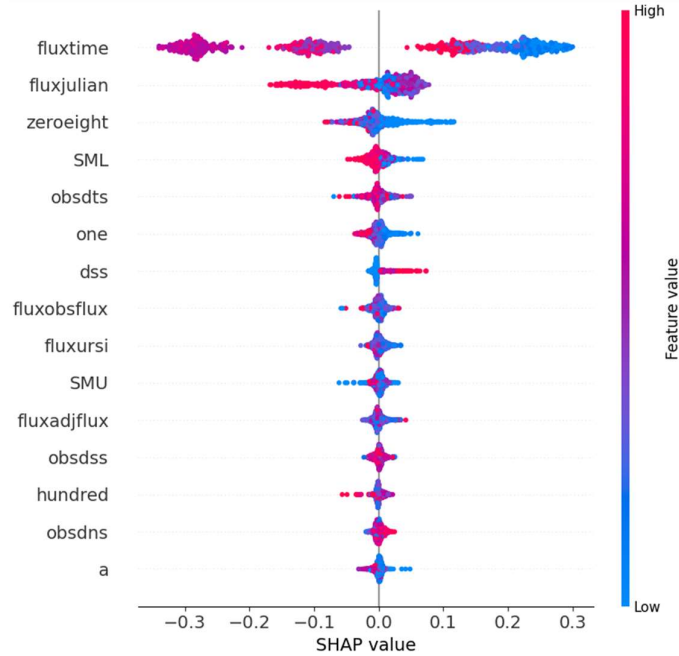
Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.1.2.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται



από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν την επίδραση στις τιμές VTEC, ενώ οι χαμηλές τιμές αυξάνουν την επίδραση.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

#### 7.1.2.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

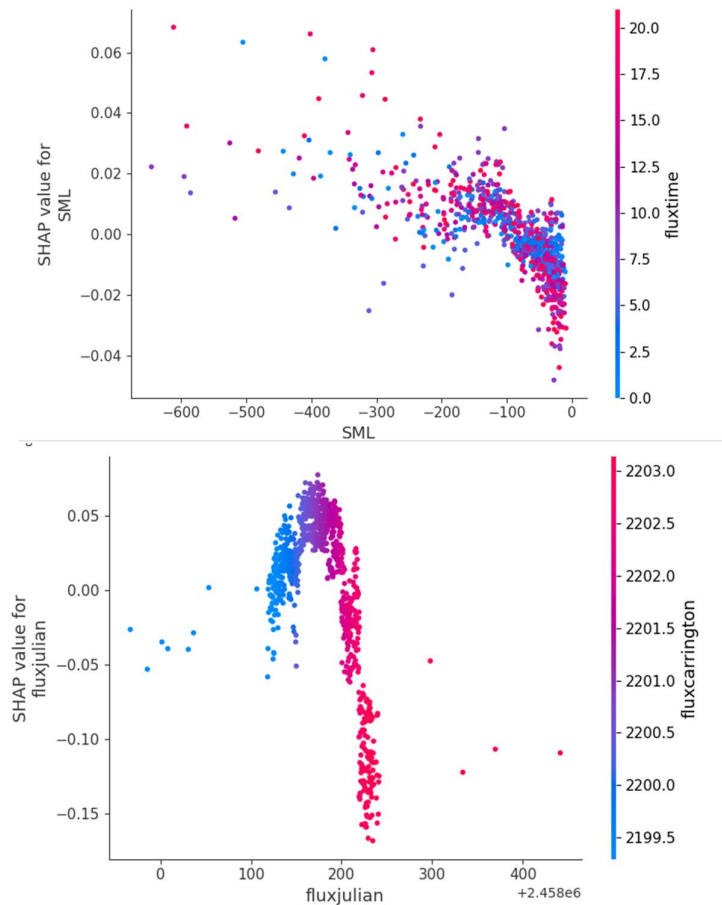
$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

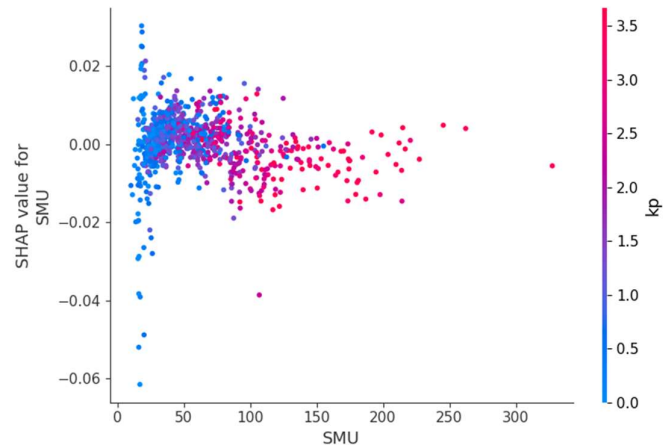
when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:





Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

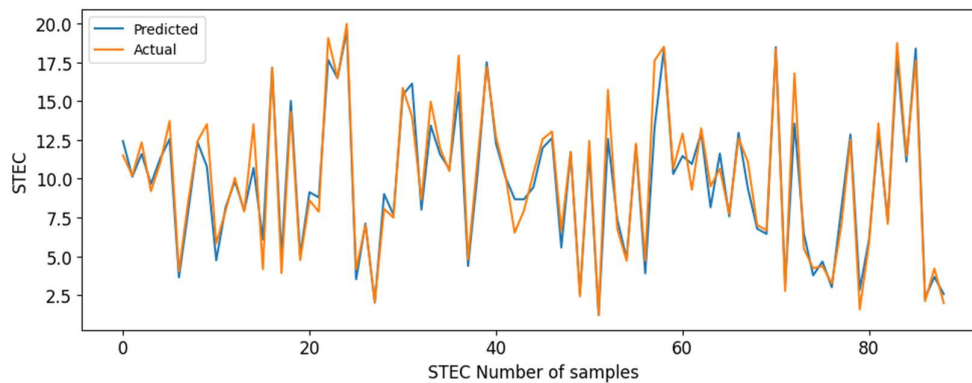
Στο πρώτο σχεδιάγραμμα η παράμετρος fluxjulian αλληλοεπιδρά με την παράμετρο fluxcarrington. Παρατηρείται ότι για τις τιμές του fluxjulian από 100-200, οι χαμηλότερες τιμές της παραμέτρου fluxcarrington επηρεάζουν θετικά το μοντέλο, ενώ για τιμές fluxjulian από 200-400 οι υψηλότερες τιμές της παραμέτρου fluxcarrington επηρεάζουν αρνητικά το μοντέλο.

## 7.2 Δορυφορικός Σταθμός Αυστραλίας

Στην ενότητα αυτή, ακολουθεί η παρουσίαση των διαγραμμάτων που προέκυψαν από την τακτική SHAP τα οποία παρουσιάζουν την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας.

### 7.2.1 STEC

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές STEC. Στον άξονα των x υπάρχει ο αριθμός των παρατηρήσεων μας δηλ το σετ δεδομένων ελέγχου και στον άξονα των ψ οι τιμές STEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται πολύ καλά.

#### 7.2.1.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

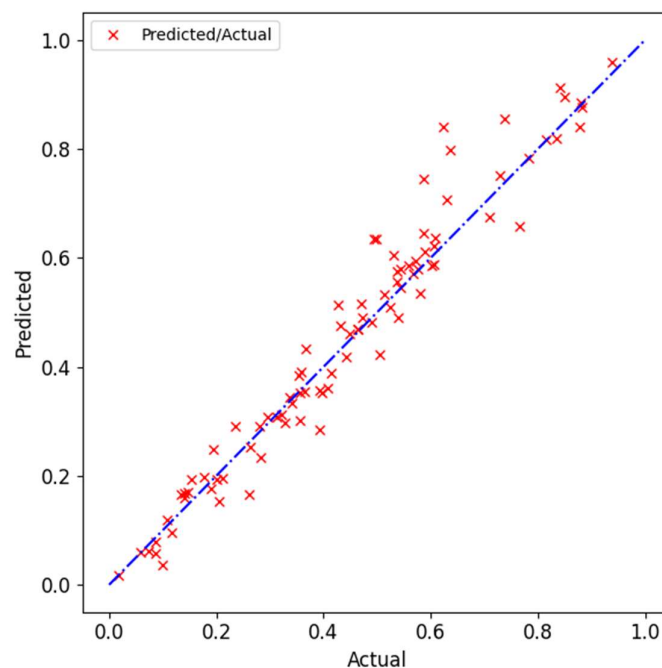
Το μέσο απόλυτο σφάλμα (mae) αντιπροσωπεύει τον μέσο όρο της απόλυτης διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τον μέσο όρο των υπολειμμάτων στο σύνολο δεδομένων.

Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του Random Forrest, καθώς οι τιμές των MAE, rMSE και του R<sup>2</sup> είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	R <sup>2</sup>
Linear Regression	3.2921	2.6615	0.5315
Decision Tree	3.4978	2.9497	0.4711
Support Vector Regression	3.1225	2.4830	0.5785
Gaussian Process Regressor	3.4538	2.6727	0.4844
XGBOOST Regressor	1.1172	0.7701	0.9460
Multi-layer Perceptron regressor	3.1751	2.4746	0.5642
LightGBM regressor	0.9120	0.6113	0.9640
Random Forest Regressor	0.0628	0.0451	0.9320

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθώς όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική καθώς δεν υπάρχουν χωρικά ακρότατα.

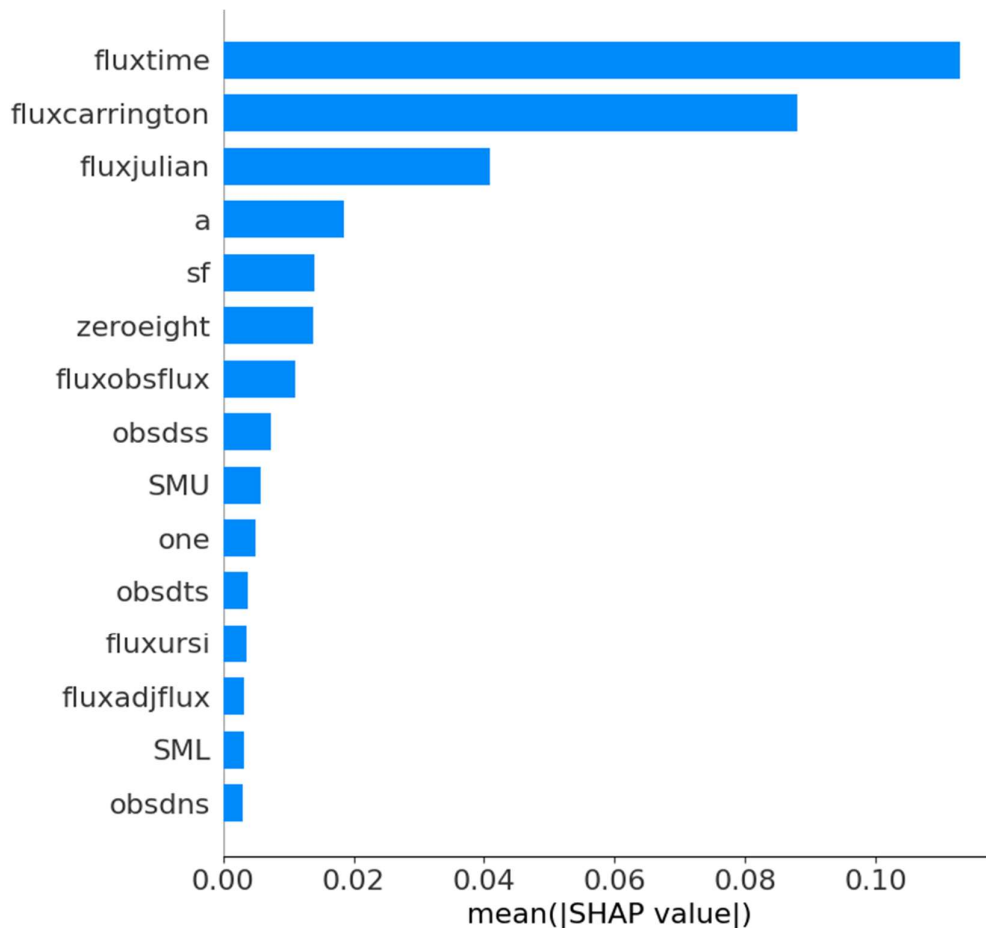


### 7.2.1.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση

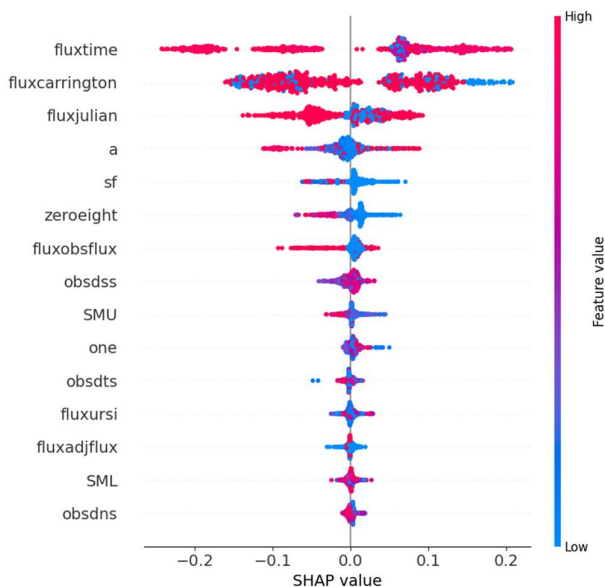
προβλεπόμενη επιρροή της τιμής STEC κατά 14 ποσοστιαίες μονάδες (0,14 στον άξονα x).

Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.2.1.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν και ταυτόχρονα αυξάνουν την επίδραση στις τιμές STEC.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

#### 7.2.1.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

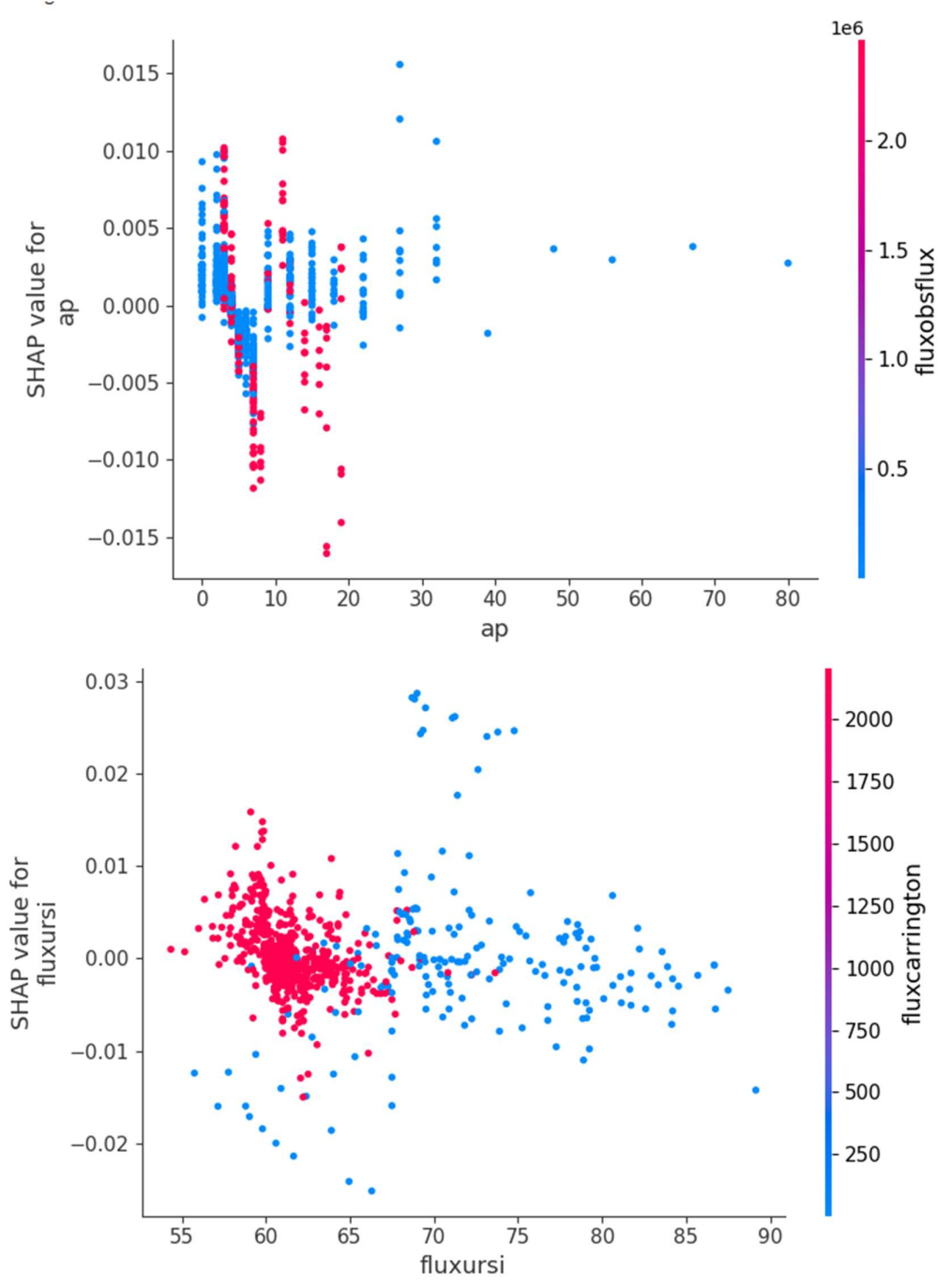
when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:





Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

Στο δεύτερο σχεδιάγραμμα η παράμετρος fluxcusi αλληλοεπιδρά με την παράμετρο fluxcarrington. Παρατηρούνται πολλά χωρικά ακρότατα και συγκεκριμένα οι υψηλές τιμές του fluxcarrington επηρεάζουν αρνητικά το μοντέλο στις χαμηλές τιμές του fluxcusi, ενώ αντίθετα οι χαμηλές τιμές του το επηρεάζουν θετικά.

Αντίστοιχα τα αποτελέσματα για τον προσδιορισμό των τιμών VTEC έχουν όπως στο Παράρτημα Β.

### 7.3 Δορυφορικός Σταθμός Χιλής

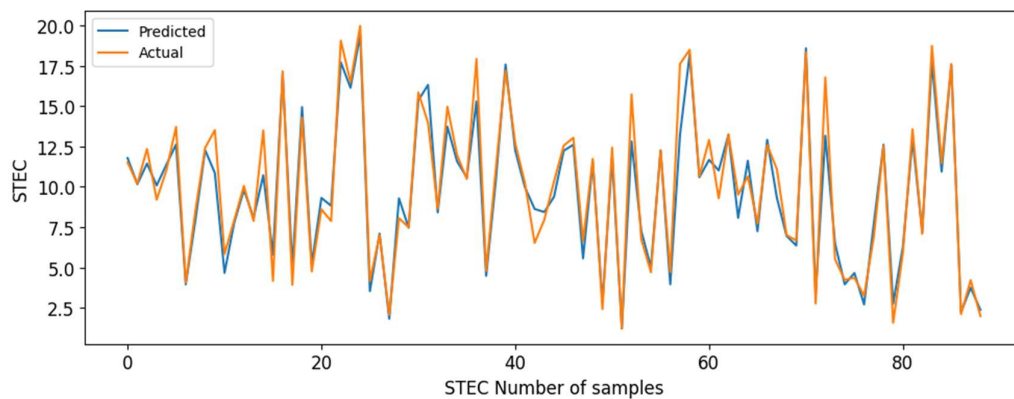
---

Στην ενότητα αυτή, ακολουθεί η παρουσίαση των διαγραμμάτων που προέκυψαν από την τακτική SHAP τα οποία παρουσιάζουν την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας.

#### 7.3.1 STEC

---

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές STEC. Στον άξονα των  $x$  υπάρχει ο αριθμός των παρατηρήσεων μας δηλ το σετ δεδομένων ελέγχου και στον άξονα των  $y$  οι τιμές STEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται πολύ καλά.

### 7.3.1.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

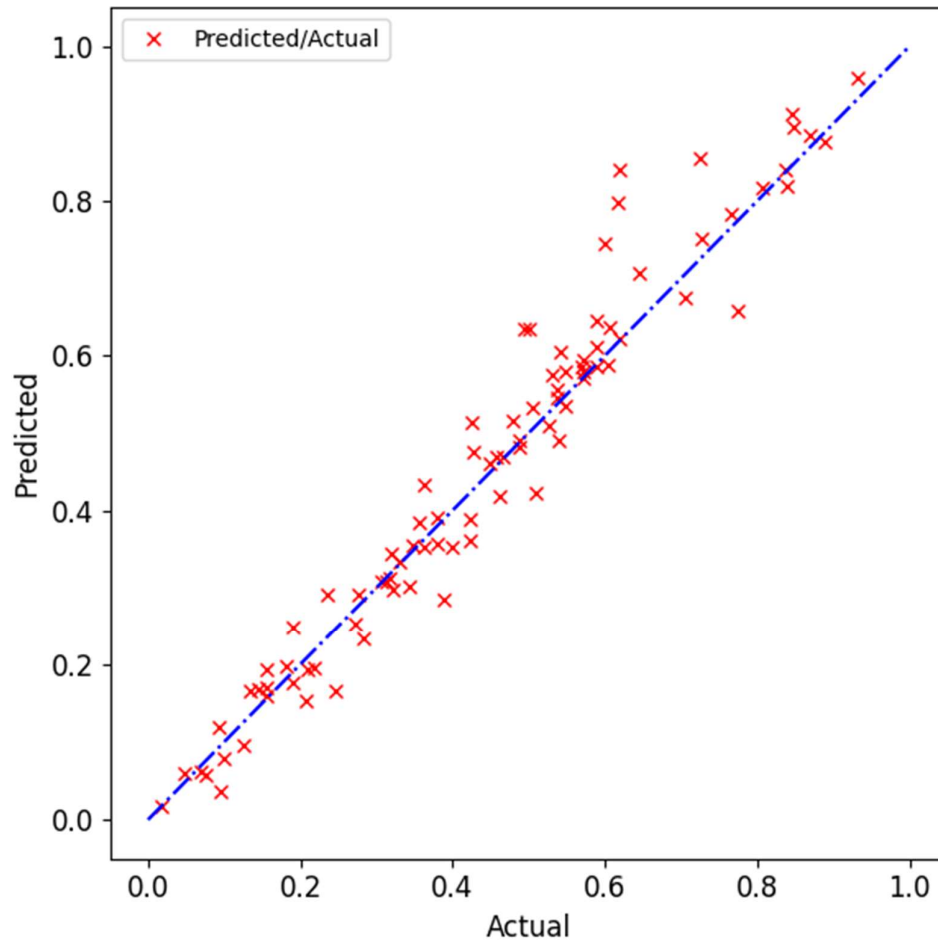
Το μέσο απόλυτο σφάλμα (mae) αντιπροσωπεύει τον μέσο όρο της απόλυτης διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τον μέσο όρο των υπολειμμάτων στο σύνολο δεδομένων.

Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του Random Forrest και του LightGBM Regressor, καθώς οι τιμές των MAE, rMSE και του  $R^2$  είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	$R^2$
Linear Regression	3.2921	2.6615	0.5315
Decision Tree	3.4978	2.9497	0.4711
Support Vector Regression	3.1855	2.5375	0.5614
Gaussian Process Regressor	3.4618	2.6897	0.4820
XGBOOST Regressor	1.1201	0.7597	0.9458
Multi-layer Perceptron regressor	3.2159	2.5458	0.5529
LightGBM regressor	0.9529	0.6391	0.9607
Random Forest Regressor	0.0626	0.0452	0.9324

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθώς όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική καθώς δεν υπάρχουν χωρικά ακρότατα.

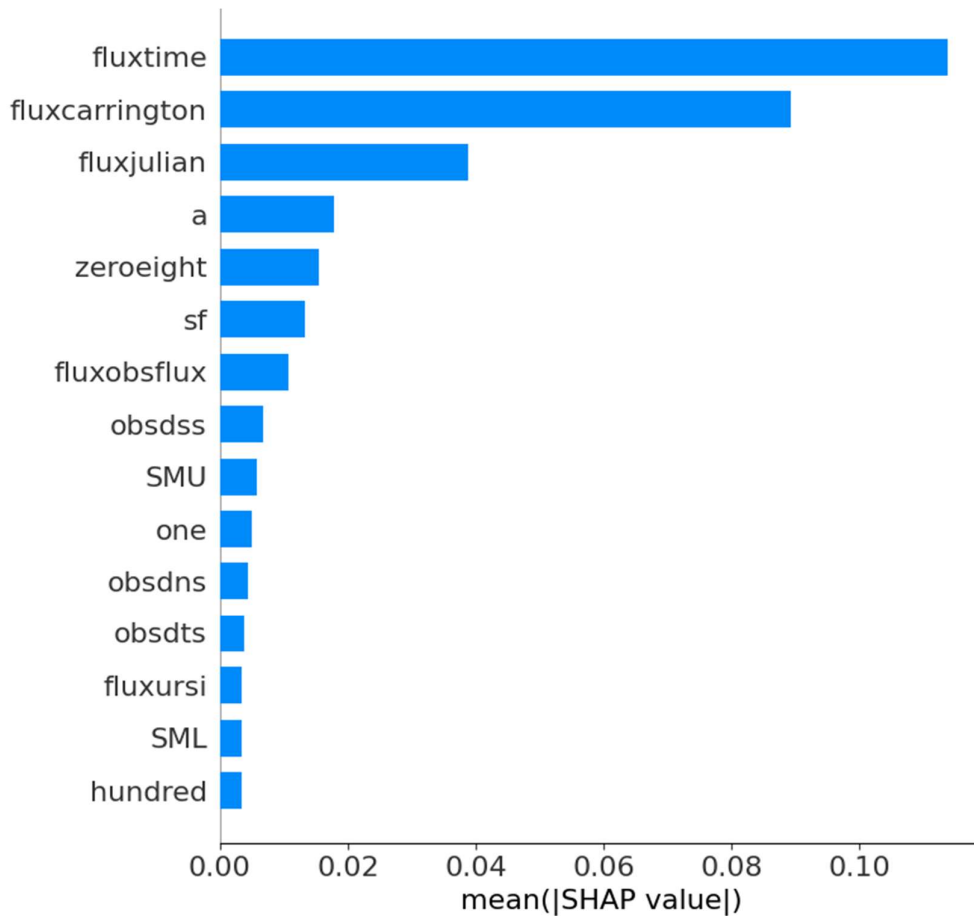


### 7.3.1.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Sharpley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Sharpley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής STEC κατά 10 ποσοστιαίες μονάδες (0,10 στον άξονα x).

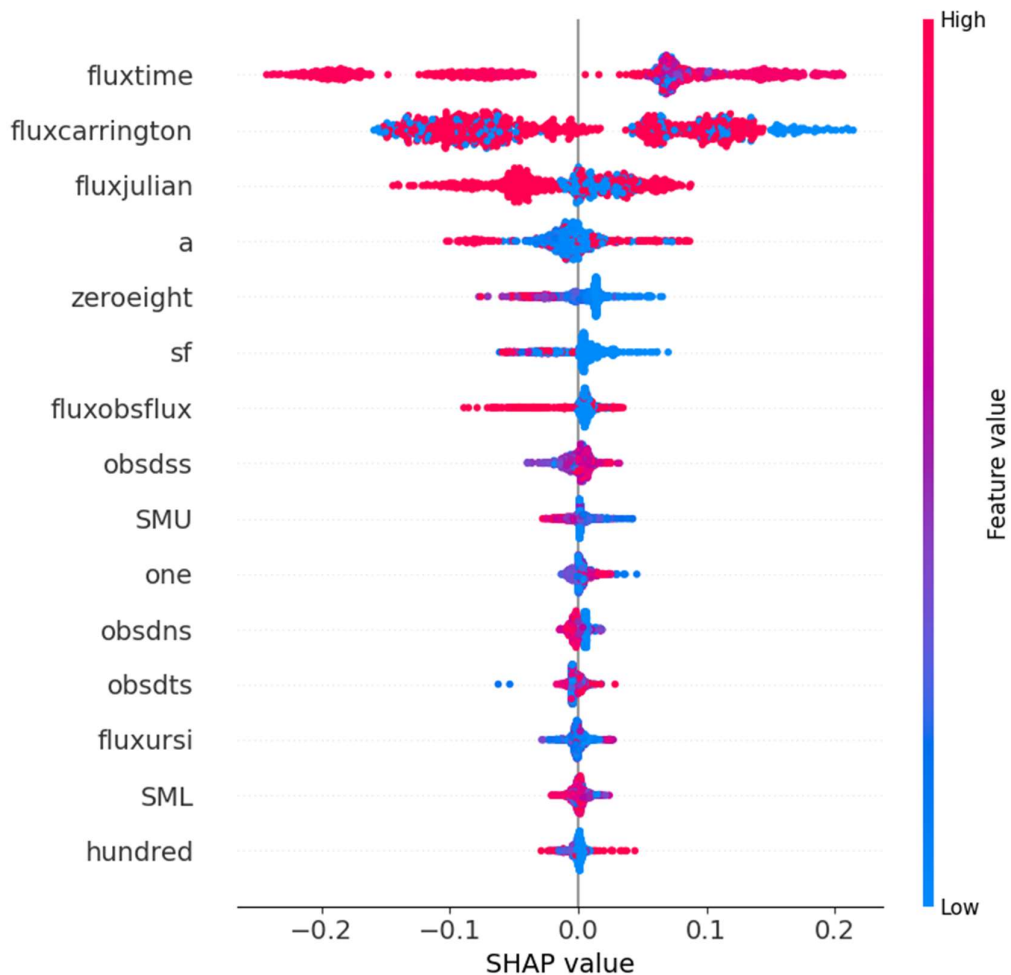
Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.3.1.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται

από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν και ταυτόχρονα αυξάνουν την επίδραση στις τιμές STEC.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

7.3.1.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

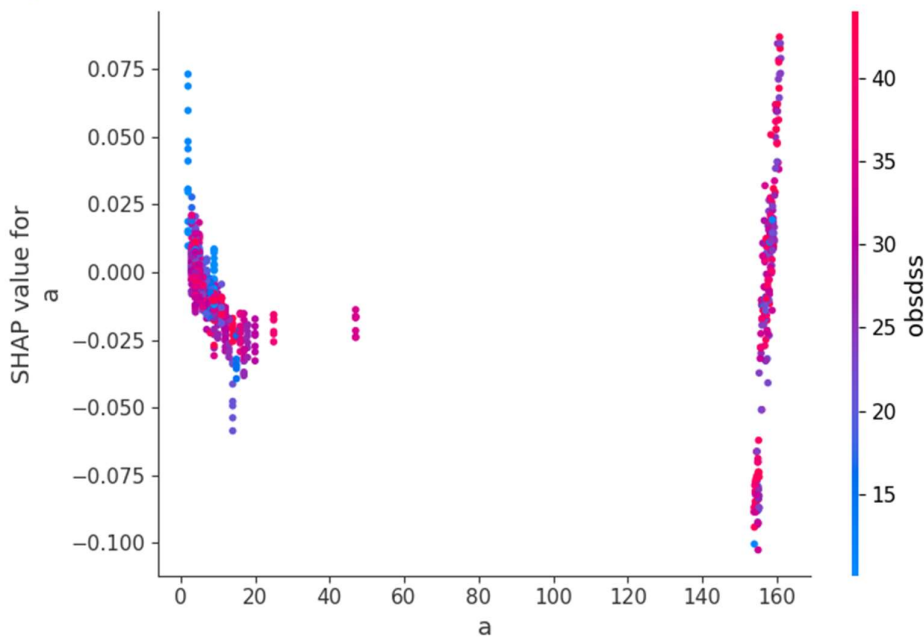
$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

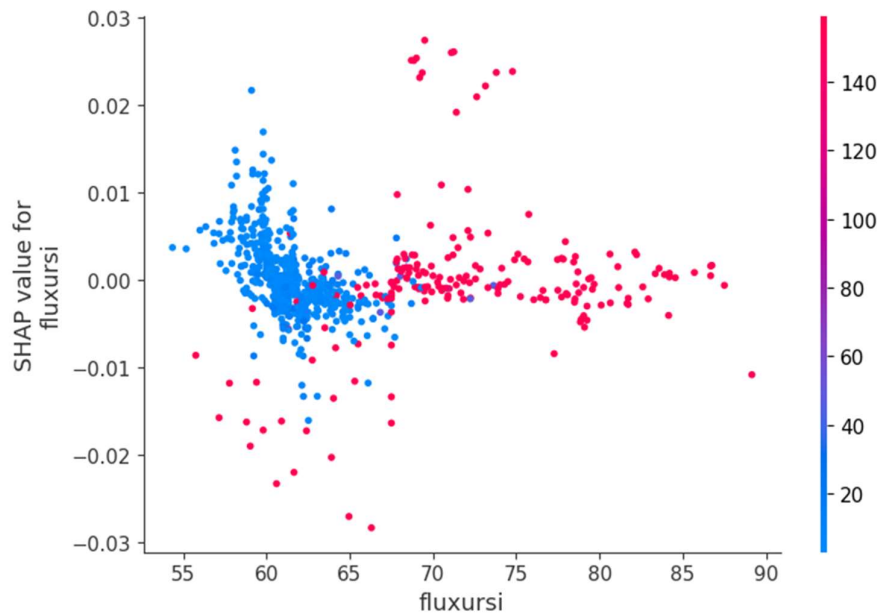
when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών S, όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις M x M, όπου M είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:





Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

Αντίστοιχα τα αποτελέσματα για τον προσδιορισμό των τιμών VTEC έχουν όπως στο Παράρτημα Β.

## 7.4 Δορυφορικός Σταθμός Κένυας

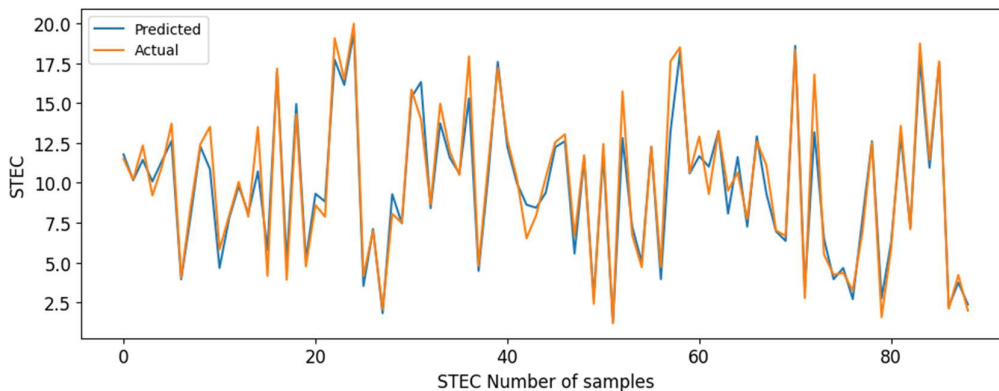
---

Στην ενότητα αυτή, ακολουθεί η παρουσίαση των διαγραμμάτων που προέκυψαν από την τακτική SHAP τα οποία παρουσιάζουν την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας.



### 7.4.1 STEC

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές STEC. Στον άξονα των x υπάρχει ο αριθμός των παρατηρήσεων μλδ το σετ δεδομένων ελέγχου και στον άξονα των ψ οι τιμές STEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται πολύ καλά.

#### 7.4.1.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

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.

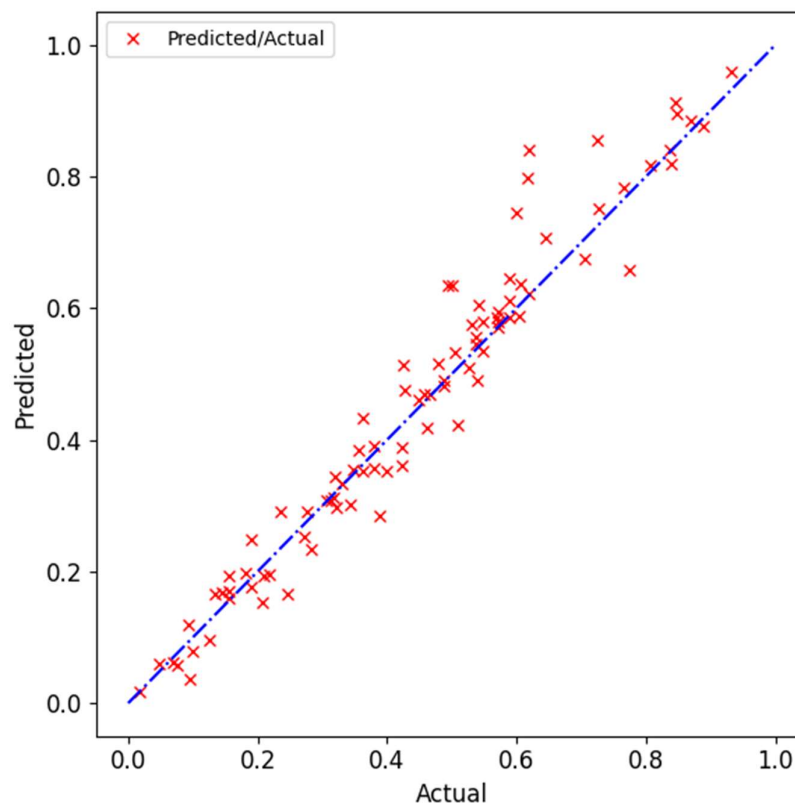
Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε

κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του Random Forrest και του LightGBM, καθώς οι τιμές των MAE, rMSE και του R<sup>2</sup> είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	R <sup>2</sup>
Linear Regression	3.2921	2.6615	0.5315
Decision Tree	3.4978	2.9497	0.4711
Support Vector Regression	3.2011	2.5485	0.5570
Gaussian Process Regressor	3.4692	2.7011	0.4797
XGBOOST Regressor	1.1201	0.7597	0.9458
Multi-layer Perceptron regressor	3.1616	2.4719	0.5679
LightGBM regressor	0.9529	0.6391	0.9607
Random Forest Regressor	0.0670	0.0478	0.9226

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθώς όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική καθώς δεν υπάρχουν χωρικά ακρότατα.

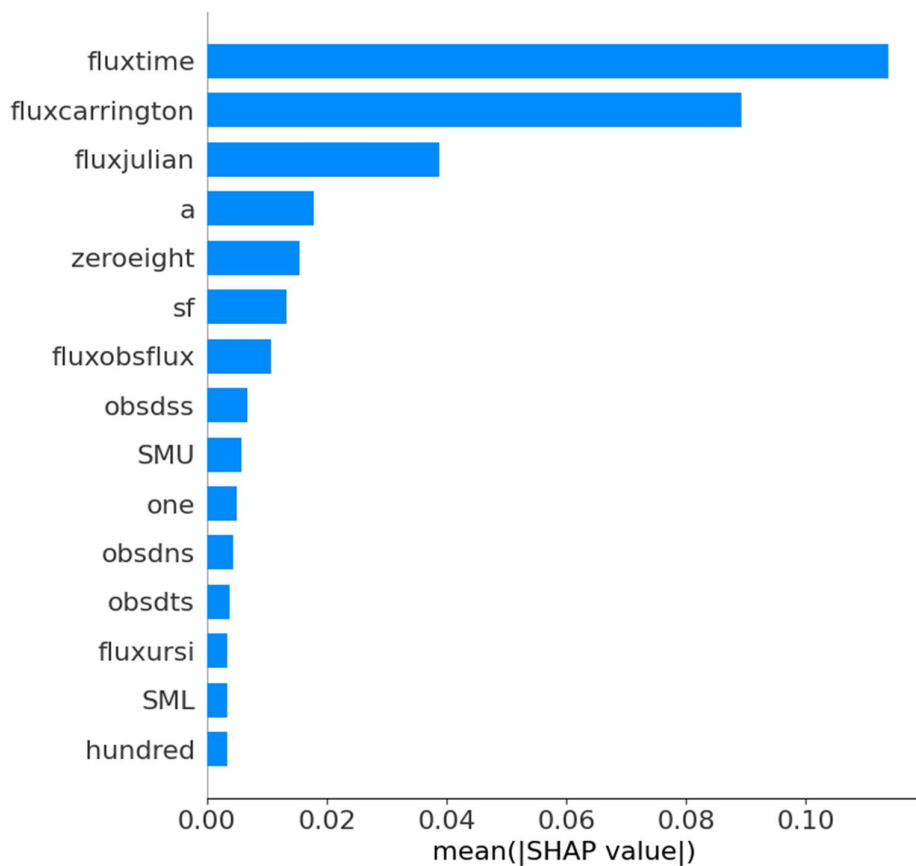


7.4.1.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



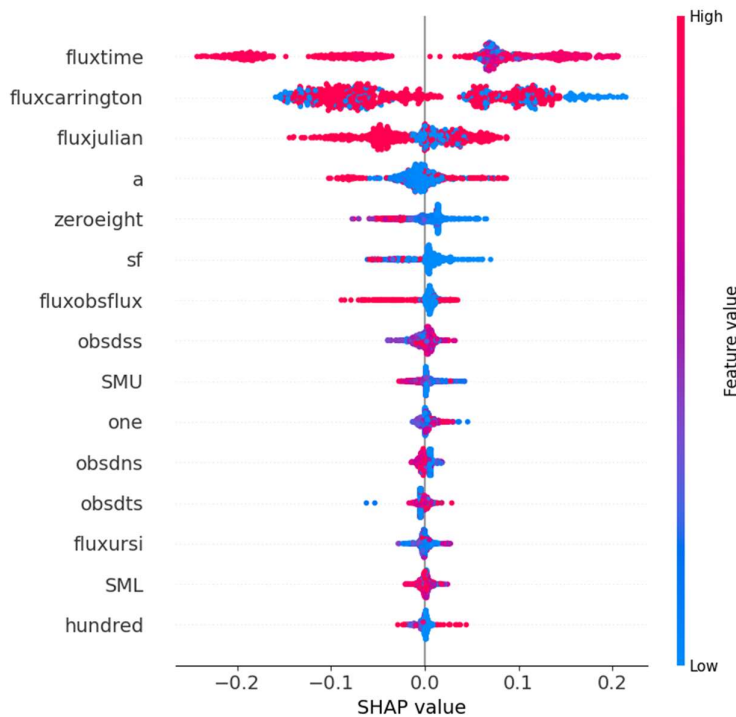
Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής STEC κατά 10 ποσοστιαίες μονάδες (0,10 στον άξονα x).

Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.4.1.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν και ταυτόχρονα αυξάνουν την επίδραση στις τιμές STEC.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

#### 7.4.1.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

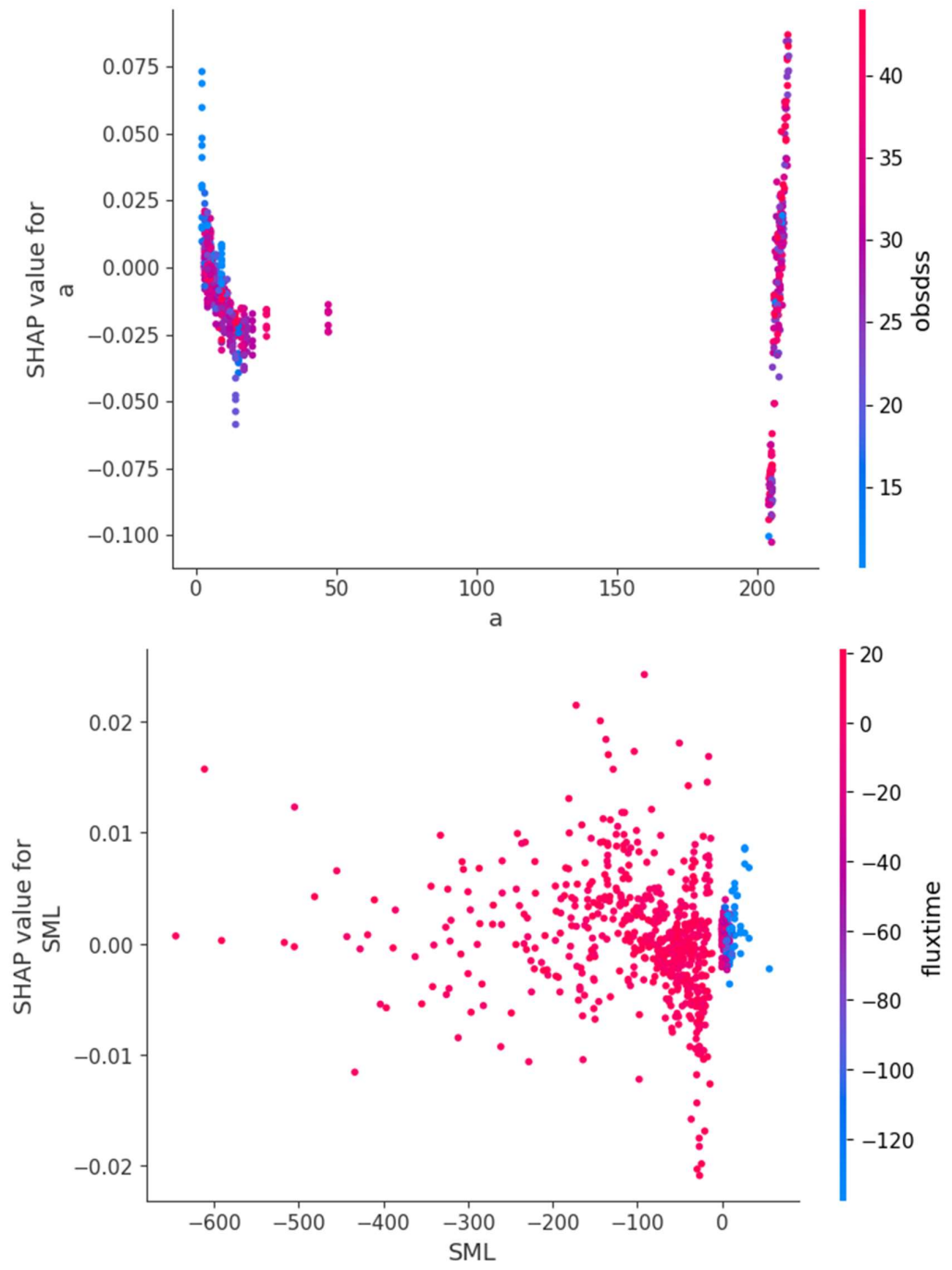
$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:



Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

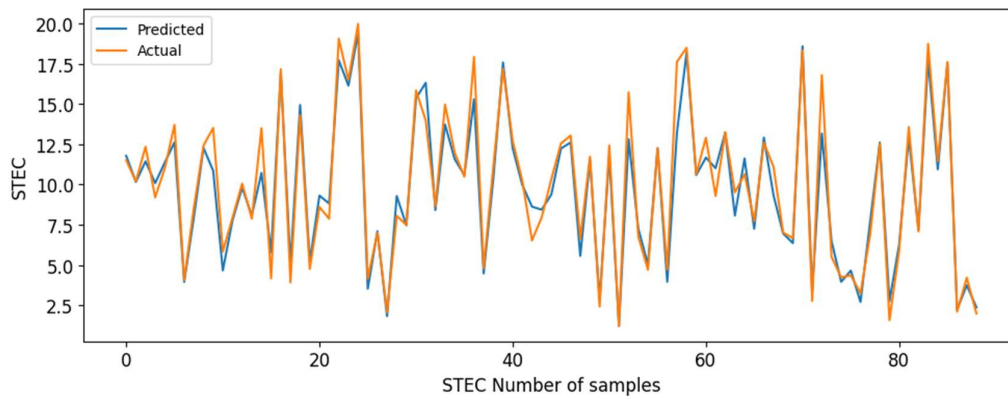
Αντίστοιχα τα αποτελέσματα για τον προσδιορισμό των τιμών VTEC έχουν όπως στο Παράρτημα Β.

## 7.5 Δορυφορικός Σταθμός Μογγολίας

Στην ενότητα αυτή, ακολουθεί η παρουσίαση των διαγραμμάτων που προέκυψαν από την τακτική SHAP τα οποία παρουσιάζουν την πρόβλεψη των τιμών STEC και VTEC, υπολογίζοντας τη συμβολή κάθε παραμέτρου επιρροής της ιονοσφαιρικής δραστηριότητας.

### 7.5.1 STEC

Αρχικά προκύπτουν τα κάτωθι διαγράμματα



Στο παραπάνω διάγραμμα παρατηρείται το πόσο καλά συγκλίνει και προσαρμόζεται το μοντέλο για τις τιμές STEC. Στον άξονα των x υπάρχει ο αριθμός των παρατηρήσεων μας δηλ το σετ δεδομένων ελέγχου και στον άξονα των ψ οι τιμές STEC. Παρατηρείται ότι και οι πραγματικές τιμές (πορτοκαλί χρώμα) αλλά και οι προβλεπόμενες τιμές (μπλε χρώμα) προσαρμόζονται πολύ καλά.

### 7.5.1.1 Mean Square error (MSE) - Mean Absolute error (MAE)

Ο στόχος της Γραμμικής Παλινδρόμησης είναι να βρει μια γραμμή που ελαχιστοποιεί το σφάλμα πρόβλεψης όλων των σημείων δεδομένων. Το βασικό βήμα σε οποιοδήποτε μοντέλο μηχανικής μάθησης είναι η αξιολόγηση της ακρίβειας του μοντέλου. Οι μετρήσεις Μέσο τετραγωνικό Σφάλμα (mse) και Μέσο Απόλυτο Σφάλμα (mae) χρησιμοποιούνται για την αξιολόγηση της απόδοσης του μοντέλου στην ανάλυση παλινδρόμησης.

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.

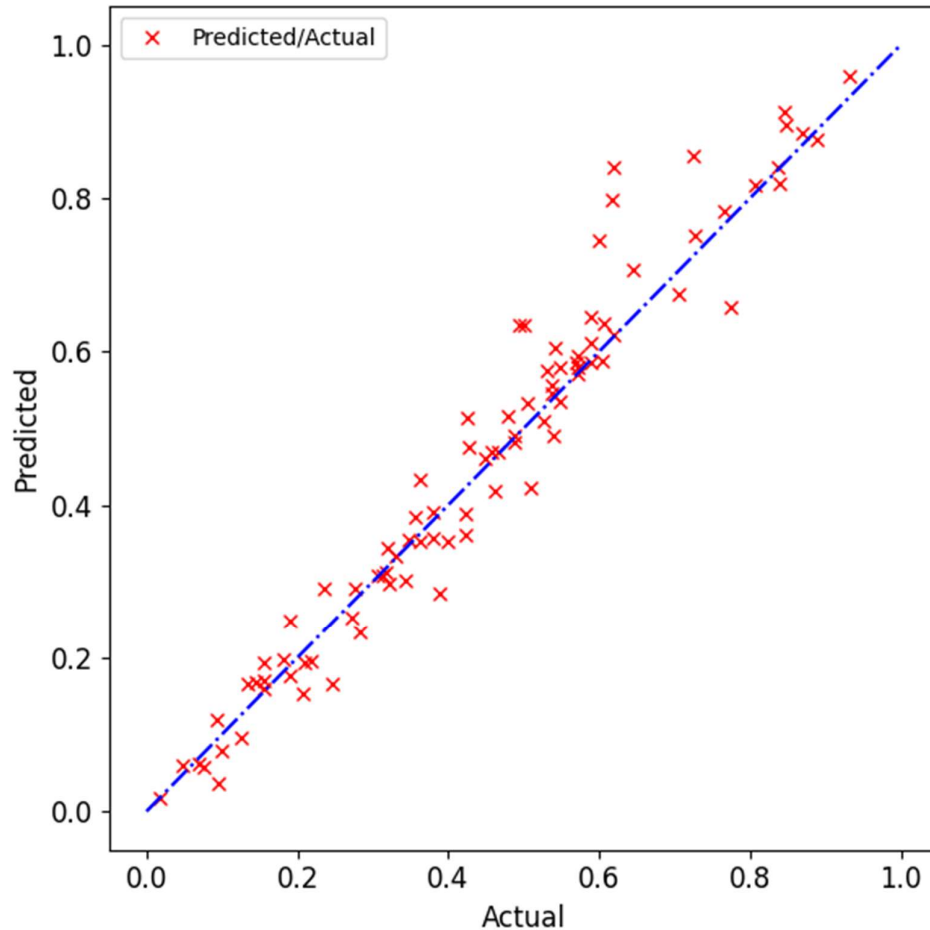
Το μέσο τετραγωνικό σφάλμα (mse) αντιπροσωπεύει τον μέσο όρο της διαφοράς στο τετράγωνο μεταξύ των αρχικών και των προβλεπόμενων τιμών στο σύνολο δεδομένων. Μετρά τη διακύμανση των υπολειμμάτων.

Όσο ο αριθμός των MAE – MSE είναι μικρός και σχεδόν ίδιος, τόσο καλύτερα προσδιορίζεται το μοντέλο και αποφεύγονται τα χωρικά ακρότατα. Στο ακόλουθο πίνακα παρατηρείται πως οι τιμές MSE, MAE είναι αρκετά μικρές αριθμητικά και σε κάποιες περιπτώσεις σχεδόν ταυτίζονται. Το σφάλμα λοιπόν έχει μικρή διασπορά και ο προσδιορισμός του μοντέλου είναι ικανοποιητικός. Διαφαίνεται λοιπόν ότι η καλύτερη τεχνική εξηγήσιμης μάθησης (μοντέλο) είναι αυτό του Random Forrest, καθώς οι τιμές των MAE, rMSE και του R<sup>2</sup> είναι πολύ χαμηλές και κοντά μεταξύ τους.

	rMSE	MAE	R <sup>2</sup>
Linear Regression	3.2921	2.6615	0.5315
Decision Tree	3.4978	2.9497	0.4711
Support Vector Regression	3.1712	2.5218	0.5653
Gaussian Process Regressor	3.4583	2.6836	0.4830
XGBOOST Regressor	1.1201	0.7597	0.9458
Multi-layer Perceptron regressor	3.1493	2.4574	0.5713
LightGBM regressor	0.9529	0.6391	0.9607
Random Forest Regressor	0.0634	0.0461	0.9308

Η παραπάνω διαπίστωση επιβεβαιώνεται και από το παρακάτω γραμμικό διάγραμμα καθώς όλα τα σημεία βρίσκονται επί της ευθείας, άρα η ακρίβεια του μοντέλου είναι ικανοποιητική καθώς δεν υπάρχουν χωρικά ακρότατα.



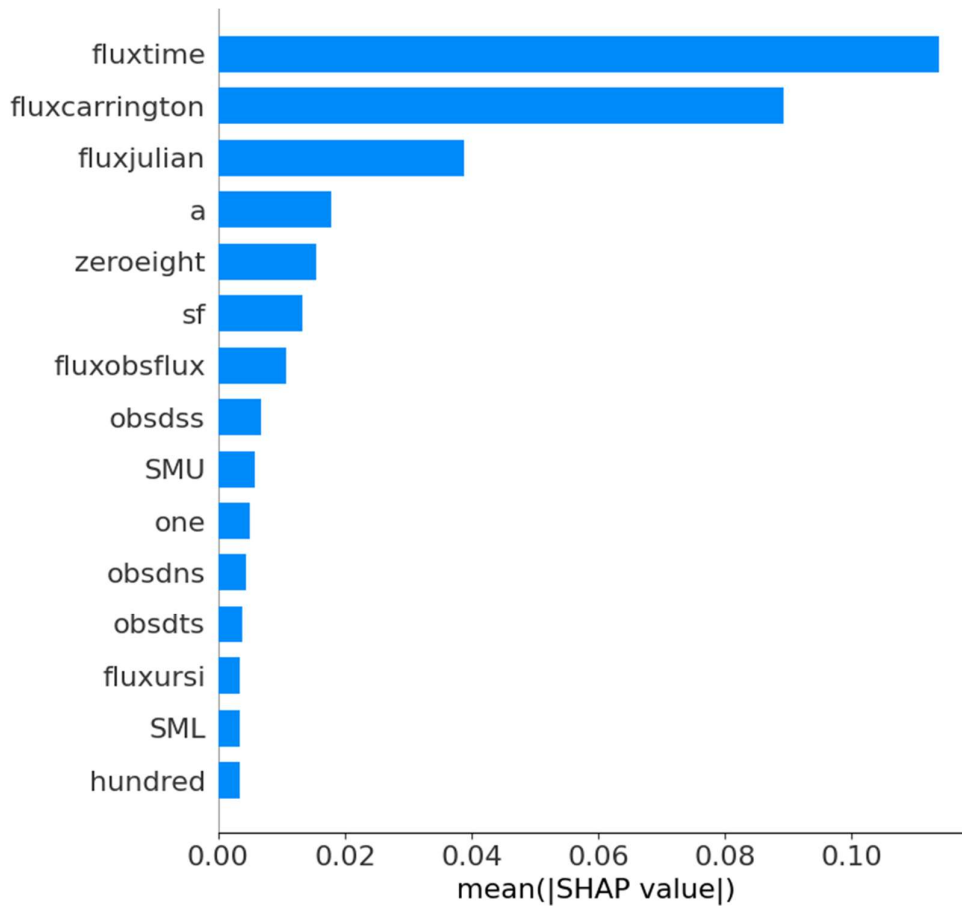


#### 7.5.1.2 SHAP Feature Importance

Η ιδέα πίσω από τη σημασία του χαρακτηριστικού SHAP είναι απλή: Τα χαρακτηριστικά με μεγάλες απόλυτες τιμές Shapley είναι σημαντικά. Εφόσον θέλουμε την παγκόσμια σημασία, υπολογίζουμε κατά μέσο όρο τις απόλυτες τιμές Shapley ανά χαρακτηριστικό στα δεδομένα:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}|$$

Στη συνέχεια, ταξινομούμε τα χαρακτηριστικά μειώνοντας τη σημασία και τα σχεδιάζουμε. Το παρακάτω σχήμα δείχνει τη σημασία του χαρακτηριστικού SHAP για το “τυχαίο δάσος” που έχει εκπαιδευτεί πριν για την πρόβλεψη της τιμής STEC.



Η σημασία του χαρακτηριστικού SHAP μετράται ως οι μέσες απόλυτες τιμές Shapley. Η παράμετρος fluxtime ήταν το πιο σημαντικό χαρακτηριστικό, αλλάζοντας την μέση προβλεπόμενη επιρροή της τιμής STEC κατά 14 ποσοστιαίες μονάδες (0,14 στον άξονα x).

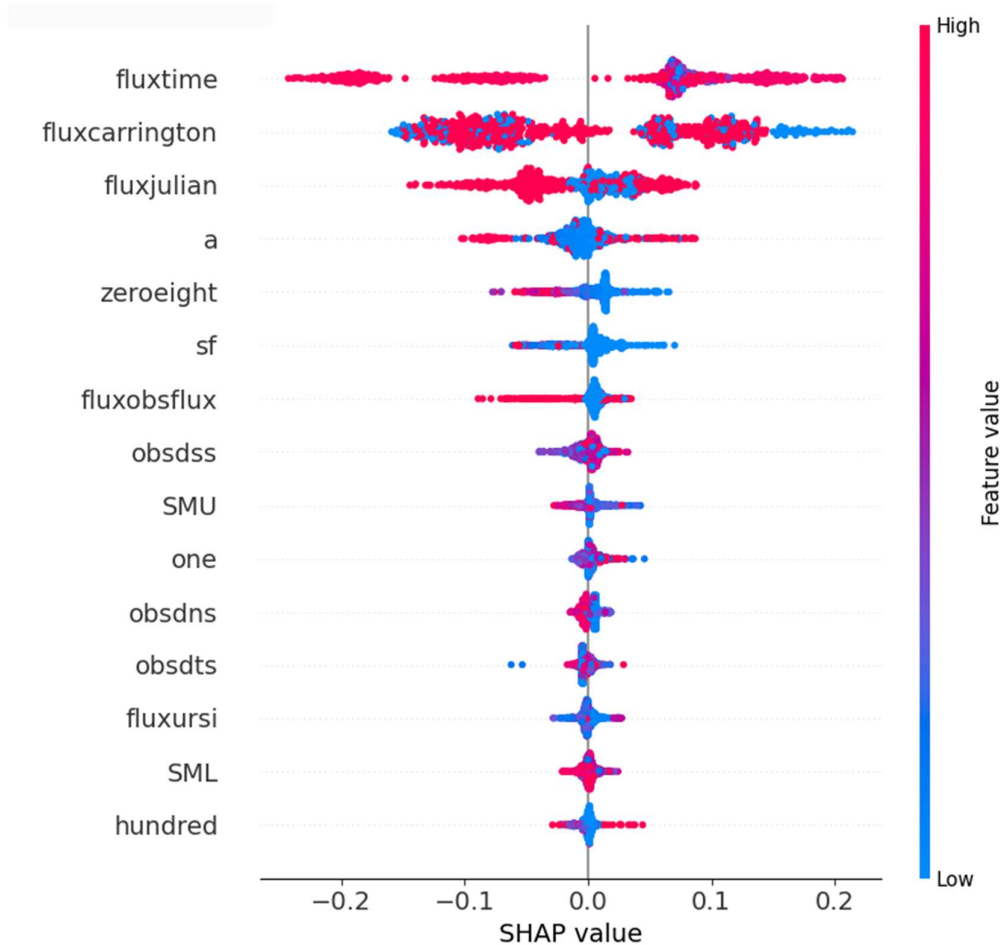
Η σημασία του χαρακτηριστικού SHAP είναι μια εναλλακτική λύση στη σημασία του χαρακτηριστικού μετάθεσης. Υπάρχει μεγάλη διαφορά μεταξύ των δύο μετρήσεων σπουδαιότητας: Η σημασία του χαρακτηριστικού μετάθεσης βασίζεται στη μείωση της απόδοσης του μοντέλου. Το SHAP βασίζεται στο μέγεθος των αποδόσεων χαρακτηριστικών.

Η γραφική παράσταση σημασίας χαρακτηριστικών είναι χρήσιμη, αλλά δεν περιέχει πληροφορίες πέρα από τις σημαντικές. Για μια πιο κατατοπιστική πλοκή, θα δούμε στη συνέχεια τη συνοπτική πλοκή.

### 7.5.1.3 SHAP Summary Plot

Η “συνοπτική πλοκή” συνδυάζει τη σημασία των χαρακτηριστικών με τα εφέ χαρακτηριστικών. Κάθε σημείο στη συνοπτική γραφική παράσταση είναι μια τιμή Shapley για ένα χαρακτηριστικό και ένα παράδειγμα. Η θέση στον άξονα y καθορίζεται

από το χαρακτηριστικό και στον άξονα x από την τιμή Shapley. Το χρώμα αντιπροσωπεύει την αξία του χαρακτηριστικού από χαμηλή σε υψηλή. Τα επικαλυπτόμενα σημεία είναι “τρελά” στην κατεύθυνση του άξονα y, οπότε έχουμε μια αίσθηση της κατανομής των τιμών Shapley ανά χαρακτηριστικό. Τα χαρακτηριστικά ταξινομούνται ανάλογα με τη σημασία τους.



Παρατηρείται ότι οι υψηλές τιμές της παραμέτρου fluxtime μειώνουν και ταυτόχρονα αυξάνουν την επίδραση στις τιμές STEC.

Στη συνοπτική γραφική παράσταση, βλέπουμε τις πρώτες ενδείξεις της σχέσης μεταξύ της αξίας ενός χαρακτηριστικού και της επίδρασης στην πρόβλεψη. Αλλά για να δούμε την ακριβή μορφή της σχέσης, πρέπει να δούμε τα σχέδια εξάρτησης SHAP.

#### 7.5.1.4 SHAP Interaction Values

Το εφέ αλληλεπίδρασης είναι το πρόσθετο εφέ συνδυασμένου χαρακτηριστικού μετά τον υπολογισμό των επιμέρους εφέ χαρακτηριστικών. Ο δείκτης αλληλεπίδρασης Shapley από τη θεωρία παιγνίων ορίζεται ως:

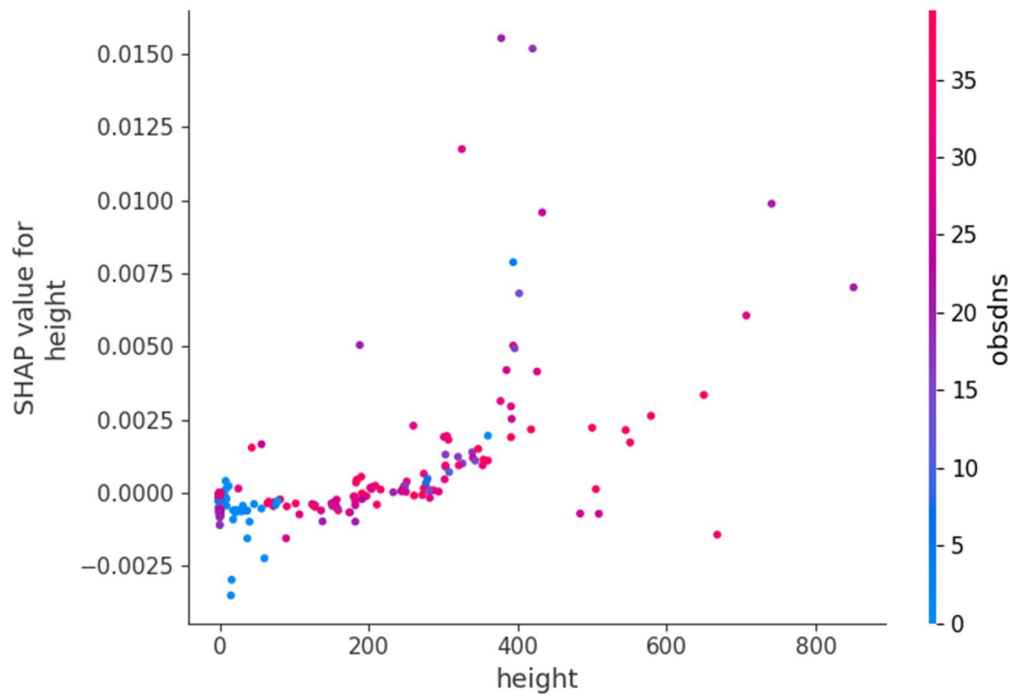
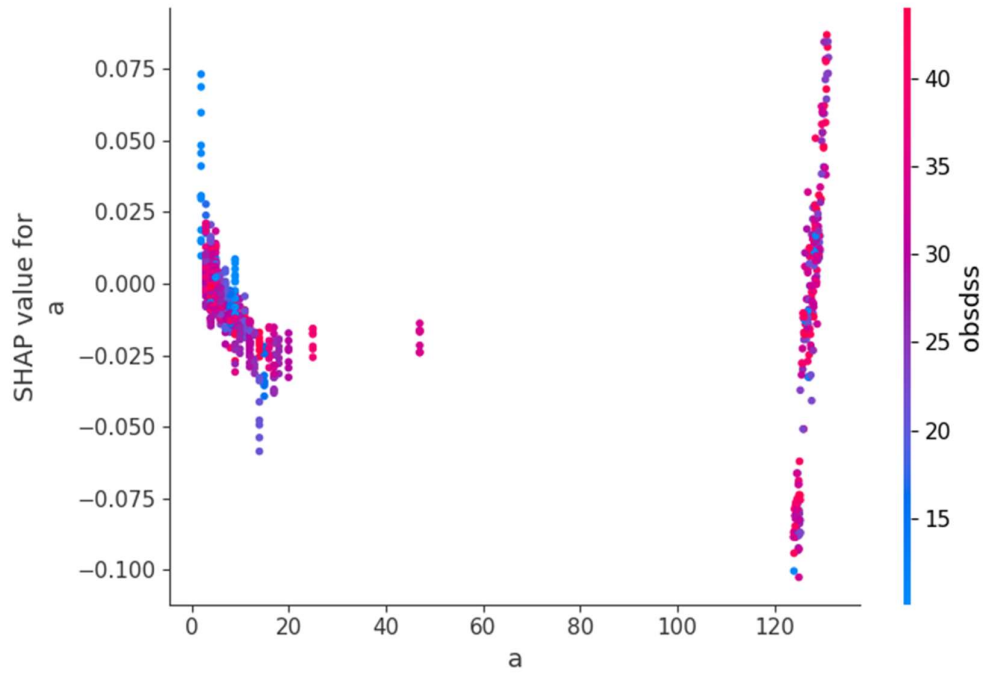
$$\phi_{i,j} = \sum_{S \subseteq \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

when  $i \neq j$  and:

$$\delta_{ij}(S) = \hat{f}_x(S \cup \{i, j\}) - \hat{f}_x(S \cup \{i\}) - \hat{f}_x(S \cup \{j\}) + \hat{f}_x(S)$$

Αυτός ο τύπος αφαιρεί το κύριο αποτέλεσμα των χαρακτηριστικών, έτσι ώστε να έχουμε το καθαρό αποτέλεσμα αλληλεπίδρασης αφού υπολογίσουμε τα μεμονωμένα αποτελέσματα. Υπολογίζουμε τον μέσο όρο των τιμών σε όλους τους πιθανούς συνασπισμούς χαρακτηριστικών  $S$ , όπως στον υπολογισμό της τιμής Shapley. Όταν υπολογίζουμε τις τιμές αλληλεπίδρασης SHAP για όλα τα χαρακτηριστικά, παίρνουμε έναν πίνακα ανά παρουσία με διαστάσεις  $M \times M$ , όπου  $M$  είναι ο αριθμός των χαρακτηριστικών.

Πώς μπορούμε να χρησιμοποιήσουμε τον δείκτη αλληλεπίδρασης; Για παράδειγμα, για να χρωματίσετε αυτόματα το διάγραμμα εξάρτησης χαρακτηριστικών SHAP με την ισχυρότερη αλληλεπίδραση:



Παραπάνω φαίνονται τα σχεδιαγράμματα εξάρτησης χαρακτηριστικών SHAP με οπτικοποίηση αλληλεπίδρασης.

Αντίστοιχα τα αποτελέσματα για τον προσδιορισμό των τιμών VTEC έχουν όπως στο Παράρτημα Β.

Επίσης τα αποτελέσματα για τους υπόλοιπους σταθμούς διαφαίνονται και αυτά στο Παράρτημα Β.

---

---

## Κεφάλαιο 8<sup>ο</sup> Συμπεράσματα

---

---

Η χρήση εξηγήσιμης τεχνητής νοημοσύνης για την ερμηνεία μοντέλων μηχανικής πρόβλεψης ιονοσφαιρικής δραστηριότητας μπορεί να οδηγήσει σε αρκετά σημαντικά συμπεράσματα:

- **Κατανόηση του Μοντέλου:** Η εξηγήσιμη τεχνητή νοημοσύνη μπορεί να βοηθήσει στην κατανόηση του τρόπου με τον οποίο λειτουργεί το μοντέλο πρόβλεψης ιονοσφαιρικής δραστηριότητας. Αναδεικνύει τις σημαντικότερες μεταβλητές και τον τρόπο με τον οποίο επηρεάζουν τα αποτελέσματα.
- **Αναγνώριση Σημαντικών Παραμέτρων:** Με τη χρήση εξηγήσιμων μεθόδων, μπορούν να αναδειχθούν οι παράμετροι που έχουν τη μεγαλύτερη επίδραση στις προβλέψεις για την ιονοσφαιρική δραστηριότητα.
- **Διαφάνεια:** Η εξηγήσιμη τεχνητή νοημοσύνη επιτρέπει στους ερευνητές και τους ενδιαφερόμενους να κατανοήσουν το μοντέλο πρόβλεψης με πιο διαφανή τρόπο, αναδεικνύοντας τις διαδικασίες λήψης αποφάσεων.
- **Αξιοπιστία και Εμπιστοσύνη:** Η δυνατότητα να κατανοήσουμε πώς λειτουργεί ένα μοντέλο πρόβλεψης οδηγεί σε μεγαλύτερη εμπιστοσύνη στα αποτελέσματά του.
- **Βελτιστοποίηση Μοντέλου:** Η κατανόηση του μοντέλου μπορεί να οδηγήσει σε βελτιώσεις του, είτε με την προσθήκη νέων χαρακτηριστικών είτε με τη βελτίωση της αρχιτεκτονικής του.
- **Κατανόηση Φαινομένων:** Μπορεί να βοηθήσει στην κατανόηση των φαινομένων που επηρεάζουν την ιονοσφαιρική δραστηριότητα και να παράσχει πιο εύστοχες προβλέψεις.

Πιο συγκεκριμένα διακρίναμε πως τα μοντέλα Random Forrest, LightGBM Regressor, ήταν τα καταλληλότερα για την πρόβλεψη της ιονοσφαιρικής δραστηριότητας ειδικά για τις τιμές STEC.

Επίσης παρατηρήθηκε ότι οι τιμές STEC, είχαν πολλή καλύτερη προσαρμογή από τις αντίστοιχες VTEC. Αυτό μπορεί να οφείλεται στο γεγονός της επιλογής των σταθμών καθόσον και των μηνών επιλογής (Ιανουάριος έως Απρίλιος)

Συνολικά, η χρήση εξηγήσιμης τεχνητής νοημοσύνης μπορεί να ενισχύσει την αξιοπιστία, την κατανοητότητα και την αποδοτικότητα των μοντέλων πρόβλεψης ιονοσφαιρικής δραστηριότητας.

Παράλληλα η συγκεκριμένη διπλωματική εργασία, απέδειξε πως μια τόσο δύσκολη και επίπονη διαδικασία, μπορεί να γίνει με ελάχιστη προεπεξεργασία, με μία απλή λήψη αρχικών δεδομένων από το διαδίκτυο, σε μια χαμηλού κόστους υπολογιστική μονάδα.

Το γεγονός πως η γλώσσα προγραμματισμού είναι η python, προσφέρει δύο μεγάλα πλεονέκτημα στην αντιμετώπιση του παραπάνω προβλήματος. Το πρώτο πλεονέκτημα είναι η φύση της γλώσσας, δηλαδή είναι ένα ανοιχτού κώδικα εργαλείο με τρομερές δυνατότητες προσφέροντας με μηδενικό κόστος εργαλεία και βιβλιοθήκες για επεξεργασία. Το δεύτερο πλεονέκτημα αφορά τα αναφερόμενα εργαλεία και οι

βιβλιοθήκες, το Tensorflow και το Keras, προσφέρουν συναρτήσεις που κάνουν ειδικά τον προγραμματισμό νευρωνικών δεικτών εφικτό σε επαγγελματίες χρήστες και μη.

Σε μελλοντικά σχέδια 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.



---

---

## Αναφορές

---

---

- [1] Siyu Tang, Zhi Huang, and Hong Yuan. Improving regional ionospheric tec mapping based on rbf interpolation. *Advances in Space Research*, 67(2):722–730, 2021.
- [2] GPS και Γεωδαιτικές Εφαρμογές, Αριστείδης Ι. Φωτίου, Χρήστος Κ. Πικριδάς
- [3] Performance evaluation of neural network TEC forecasting models over equatorial low-latitude Indian GNSS stationG. Sivavaraprasad\*, V.S. Deepika, D. SreenivasaRao, M. Ravi Kumar, M. Sridhar
- [4] *Research Article* Prediction of Ionospheric TEC Based on the NARXNeural Network Liu Guoyan ,1 Gao Wang,2 Zhang Zhengxie,2 and Zhao Qing21*Nanjing Vocational University of Industry Technology, Nanjing 210000, China* 2*South and East University, Nanjing 210000, China* Correspondence should be addressed to Liu Guoyan; 2296813027@qq.com Received 27 ugust 2021; Revised 1 October 2021; Accepted 15 October 2021; Published 31 October 2021
- [5] Predicting the Effects of Solar Storms on the Ionosphere Based on a Comparison of Real-Time Solar Wind Data with the Best-Fitting Historical Storm Event Erik Schmölder \* and Jens Berdermann
- [6] Statistical analysis of an ionospheric parameter as a base for earthquake prediction Mei Li1,2,3 and Michel Parrot1 Received 4 October 2012; revised 23 March 2013; accepted 3 May 2013; published 19 June 2013.
- [7] Pedro Lara-Benvtez, Manuel Carranza-Garcva, and Josi Riquelme. An xperimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31, 11 2020.
- [8] Deep Learning,Ian Goodfellow,Yoshua Bengio,Aaron Courville
- [9] Jeffrey M Forbes, Scott E Palo, and Xiaoli Zhang. Variability of the ionosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, 62(8):685–693, 2000.
- [10] Hu Jiang, Jingbin Liu, Zemin Wang, Jiachun An, Jikun Ou, Shulun Liu, and Ningbo Wang. Assessment of spatial and temporal tec variations derived from ionospheric models over the polar regions. *Journal of Geodesy*, 93(4):455–471, 2019.
- [11] Siyu Tang, Zhi Huang, and Hong Yuan. Improving regional ionospheric tec mapping based on rbf interpolation. *Advances in Space Research*, 67(2):722–730, 2021.
- [12] John A Klobuchar. Ionospheric time-delay algorithm for single-frequency gps users. *IEEE Transactions on aerospace and electronic systems*, (3):325–331, 1987.
- [13] Manuel Hernandez-Pajares, J. Miguel Juan, Jaume Sanz, ñngela Aragón-ñngel, Alberto Garcva-Rigo, Dagoberto Salazar, and Miquel Escudero. The ionosphere: effects, GPS modeling and the benefits for space geodetic techniques. *Journal of Geodesy*, 85(12):887–907, December 2011.

- [14] 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*, XLII- 3/W10:1175–1182, 02 2020.
- [15] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis. Context aware energy disaggregation using adaptive bidirectional lstm models. *IEEE Transactions on Smart Grid*, 11(4):3054–3067, 2020.
- [16] J. Space Weather Space Clim.2 (2012) A22 Monitoring, tracking and forecasting ionospheric perturbations using GNSS techniques
- [17] Forecasting Global Ionospheric TEC Using Deep Learning Approach, [Lei Liu](#),[Shasha Zou](#),[Yibin Yao](#),[Zihan Wang](#)
- [18] Forecasting the occurrence of ionospheric scintillation activity in the equatorial ionosphere on a day-to-day basis Author links open overlay panelDavid N.Anderson<sup>a</sup>BodoReinisch<sup>b</sup>CesarValladare<sup>c</sup>JorgeChau<sup>d</sup>OscarVeliz<sup>d</sup>
- [19] Forecasting of ionospheric TEC for different latitudes, seasons and solar activity conditions based on OKSMR. Mukesh, V. Karthikeyan, P. Soma & P. Sindhu *Astrophysics and Space Science* volume 365, Article number: 13 (2020)
- 20] Ionospheric specification and forecasting based on observations from European ionosondes participating in DIAS project Anna BELEHAKI1 , Ljiljana CANDER2 , Bruno ZOLESI3 , Juergen BREMER4 , Christer JUREN5 , Iwona STANISŁAWSKA6 , Dimitris DIALETIS7 and Michael HATZOPOULOS7
- [21] Forecasting the ionospheric  $f_0F_2$  parameter one hour ahead using a support vector machine Author links open overlay panelChunChen<sup>ab</sup>ZhensenWu<sup>a</sup>ShujiSun<sup>b</sup>PanpanBan<sup>b</sup>ZhonghuaDing<sup>b</sup>ZhengwenXu<sup>b</sup>
- [22] A Causal Long Short-Term Memory Sequence to Sequence Model for TEC Prediction Using GNSS Observations Maria Kaselimi 1,\* , Athanasios Voulodimos 2 , Nikolaos Doulamis 1 , Anastasios Doulamis 1 and Demitris Delikaraoglou 1
- [23] S. Haykin, *Neural networks and learning machines*, 3/E. Pearson Education India, 2010.
- [24] “Deep learning,” *Wikipedia*. Sep. 20, 2021. Accessed: Oct. 03, 2021. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Deep\\_learning&oldid=1045496580](https://en.wikipedia.org/w/index.php?title=Deep_learning&oldid=1045496580)
- [25] W. J. Zhang, G. Yang, Y. Lin, C. Ji, and M. M. Gupta, “On Definition of Deep Learning,” in *2018 World Automation Congress (WAC)*, Jun. 2018, pp. 1–5. doi: 10.23919/WAC.2018.8430387.
- [26] Jeffrey M Forbes, Scott E Palo, and Xiaoli Zhang. Variability of the ionosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, 62(8):685–693, 2000.
- [27] Hochreiter, S.; Jürgen, S. Long Short-Term Memory. *Neural Comput.* 1997, 8, 1735–1780.

- [28] Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. In 1999 Ninth International Conference on Artificial Neural Networks ICANN; IET: London, UK, 1999.
- [29] GPS-TEC analysis application *Developed by Dr. Gopi Krishna Seemala Indian Institute of Geomagnetism (IIG), Navi Mumbai, India Previous affiliations: Research Institute for Sustainable Humanosphere (RISH), Kyoto University, Kyoto, Japan Institute for Scientific Research, Boston College, Boston, U.S.A.*
- [30] Randa Natras, Michael Schmidt. Machine Learning Model Development for Space Weather Forecasting in the Ionosphere. 2021
- [31] 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.
- [32] George C Tiao and George EP Box. Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816, 1981.
- [33] J. 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.
- [34] 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.
- [35] 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.
- [36] Ke Su, Shuanggen Jin, and M. M. Hoque. Evaluation of ionospheric delay effects on multi-gnss positioning performance. *Remote Sensing*, 11(2), 2019.
- [37] 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.
- [38] P. T. Newell, J. W. Gjerloev. Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. 2011
- [39] Schmölter 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
- [40] 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.

- [41] Borovsky, J.E.; Denton, M.H. Differences between CME-driven storms and CIR-driven storms. *J. Geophys. Res.* 2006, *111*.
- [42] 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.
- [43] 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.
- [44] Baker, D.N.; Daly, E.; Daglis, I.; Kappenman, J.G.; Panasyuk, M. Effects of Space Weather on Technology Infrastructure. *Space Weather* 2004
- [45] 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.
- [46] Berdermann, J.; Borries, C.; Hoque, M.M.; Jakowski, N. Forecast of Total Electron Content over Europe for Disturbed Ionospheric Conditions. 2021
- [47] 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
- [48] 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
- [49] Tsagouri, I.; Belehaki, A. Ionospheric forecasts for the European region for space weather applications. *J. Space Weather Space Clim.* 201
- [50] 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.
- [51] Nicola Linty, Alex Minetto, Fabio Dovi, Luca Spogli. Effects of Phase Scintillation on the GNSS Positioning Error During the September 2017 Storm at Svalbard. 2018
- [52] 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
- [53] Schmölter, 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 Hist. 2018
- [54] G. Van Rossum, "Python Programming Language,," in USENIX annual technical conference, 2007, vol. 41, p. 36.

- [55] D. Kuhlman, A python book: Beginning python, advanced python, and python exercises. 2009.
- [56] T. E. Oliphant, A guide to NumPy, vol. 1. 2006.
- [57] G. Hackeling, Mastering Machine Learning with scikit-learn. 2017
- [58] Seemala, Gopi. (2012). GPS-TEC analysis software.
- [59] “4 Reasons Why You Should Use Google Colab for Your Next Project | by Orhan G. Yalçın | Towards Data Science.” <https://towardsdatascience.com/4-reasons-why-you-should-use-google-colab-for-your-next-project-b0c4aaad39ed> (accessed Oct. 04, 2021).
- [60] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” *ArXiv160304467 Cs*, Mar. 2016, Accessed: Oct. 04, 2021. [Online]. Available: <http://arxiv.org/abs/1603.04467>
- [61] M. Abadi *et al.*, *TensorFlow, Large-scale machine learning on heterogeneous systems*. 2015. doi: 10.5281/zenodo.4724125.
- [62] K. Team, “Keras documentation: Keras API reference.” <https://keras.io/api/> (accessed Oct. 04, 2021).
- [63] K. Team, “Keras documentation: Core layers.” [https://keras.io/api/layers/core\\_layers/](https://keras.io/api/layers/core_layers/) (accessed Oct. 04, 2021).
- [64] G. Hackeling, Mastering machine learning with scikit-learn: apply effective learning algorithms to real-world problems using scikit-learn. Birmingham: Packt Publ, 2014.
- [65] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” *ArXiv180308375 Cs Stat*, Feb. 2019, Accessed: Oct. 03, 2021. [Online]. Available: <http://arxiv.org/abs/1803.08375>

---

---

## Παράρτημα Α

---

---

### Πίνακας δεδομένων ανά μήνα

---

[https://www.dropbox.com/scl/fi/tc3yi74lfk6334bbsid41/dedomena\\_GNSS\\_stations.7z?rlkey=awtk5pdbytp9jtue9z3nwjzk&st=32e1c72z&dl=0](https://www.dropbox.com/scl/fi/tc3yi74lfk6334bbsid41/dedomena_GNSS_stations.7z?rlkey=awtk5pdbytp9jtue9z3nwjzk&st=32e1c72z&dl=0)

---

---

## Παράρτημα Β

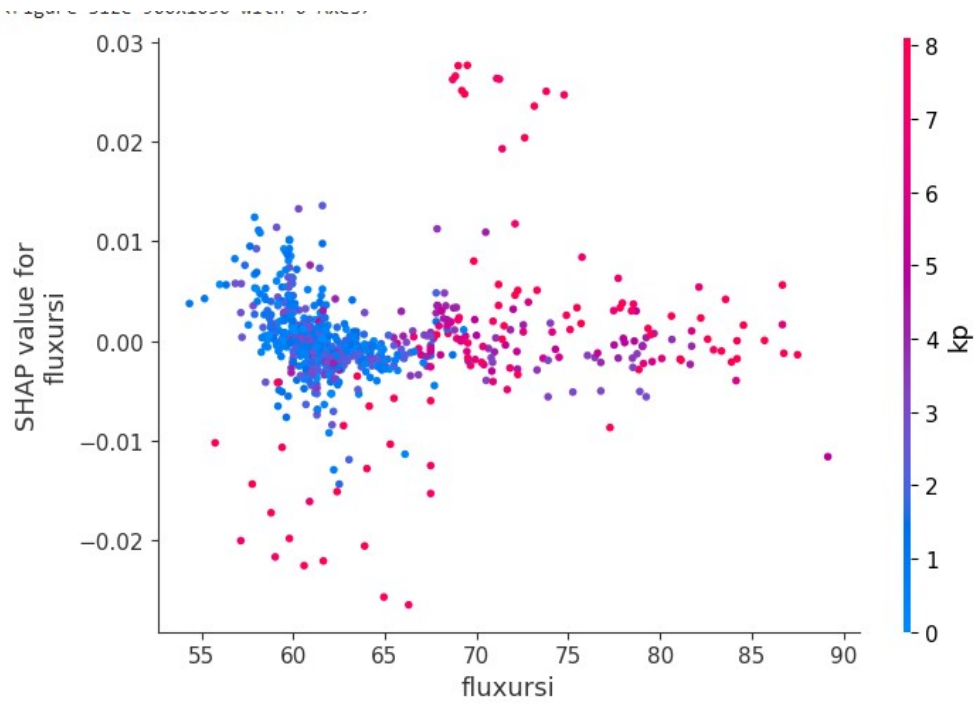
---

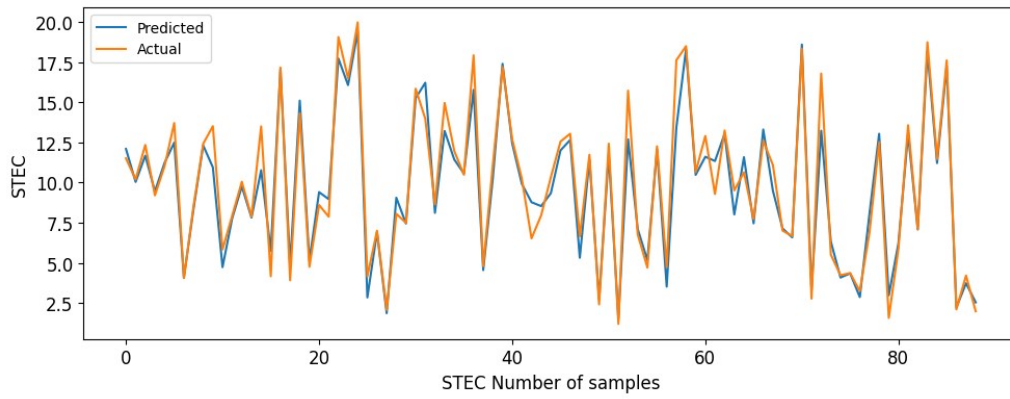
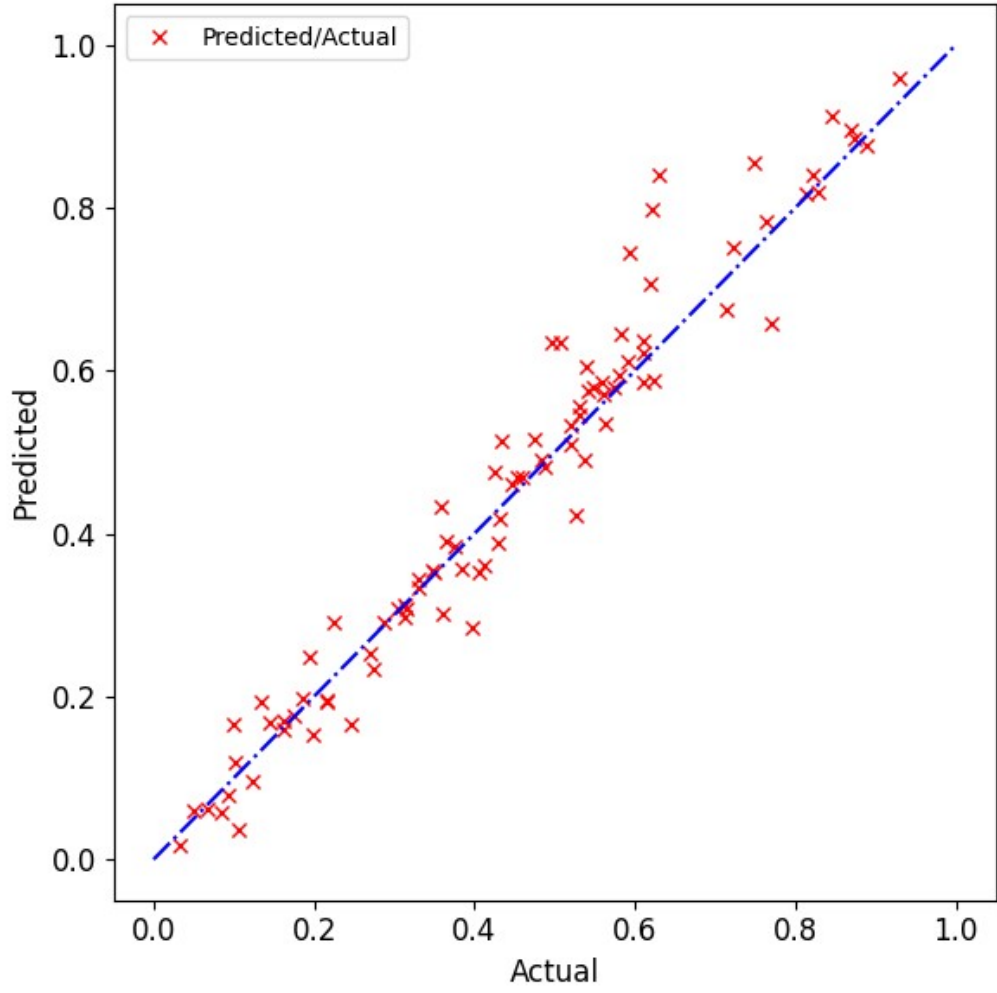
---

### Αποτελέσματα εξηγήσιμης μάθησης για την επιρροή τιμών STEC - VTEC για τους σταθμούς Νορβηγίας και Ηνωμένων Πολιτειών Αμερικής

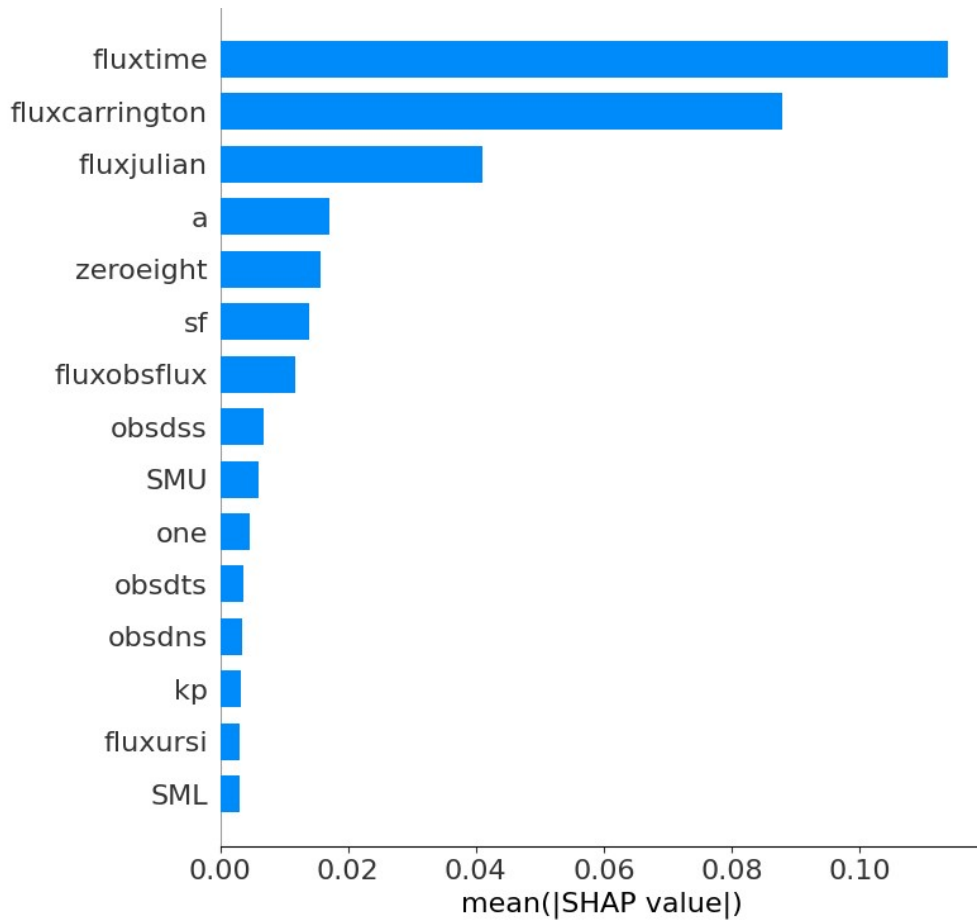
---

#### STEC NORΒΗΓΙΑΣ

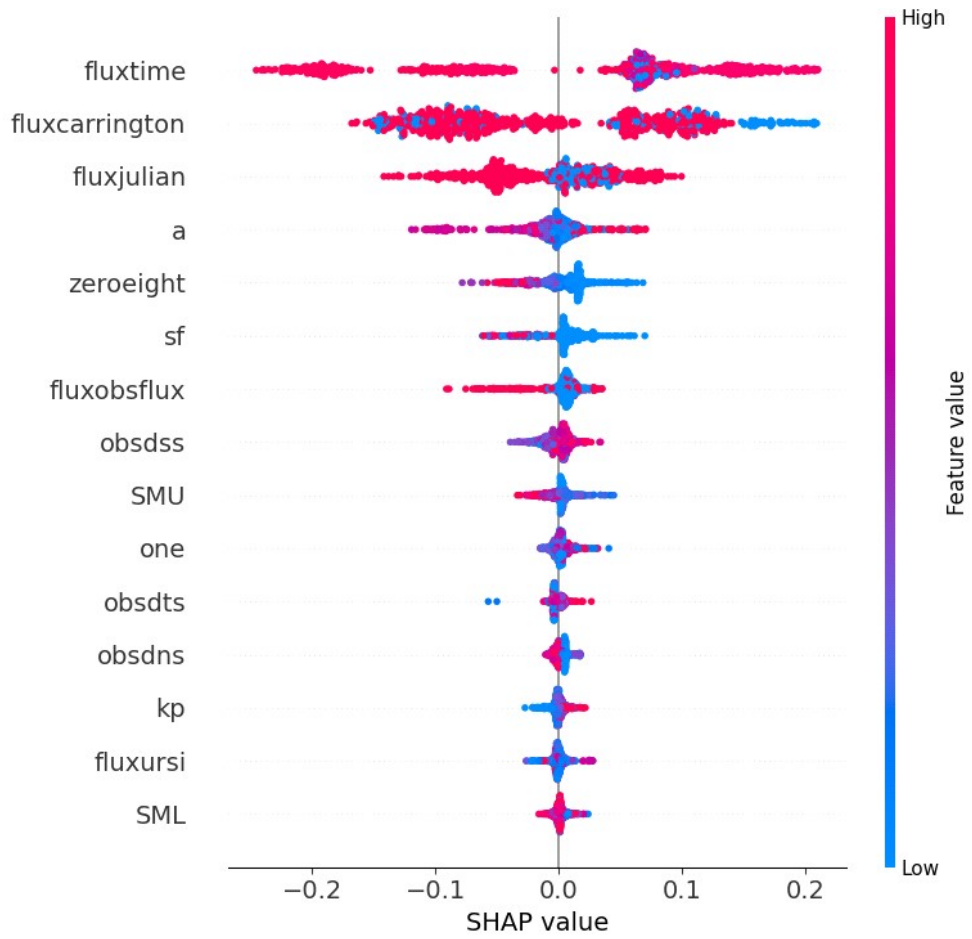




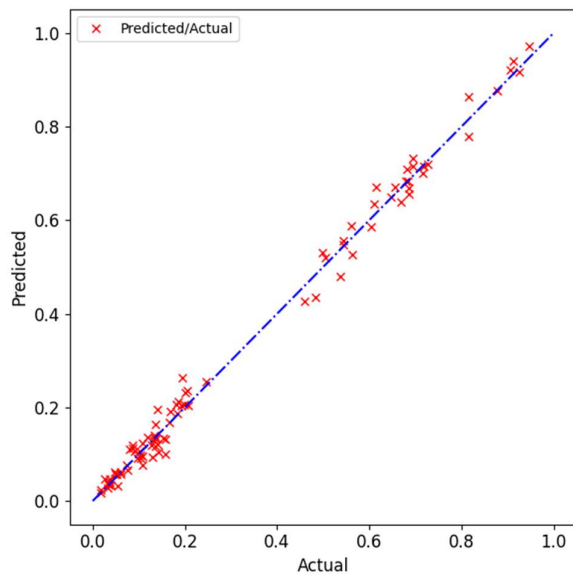


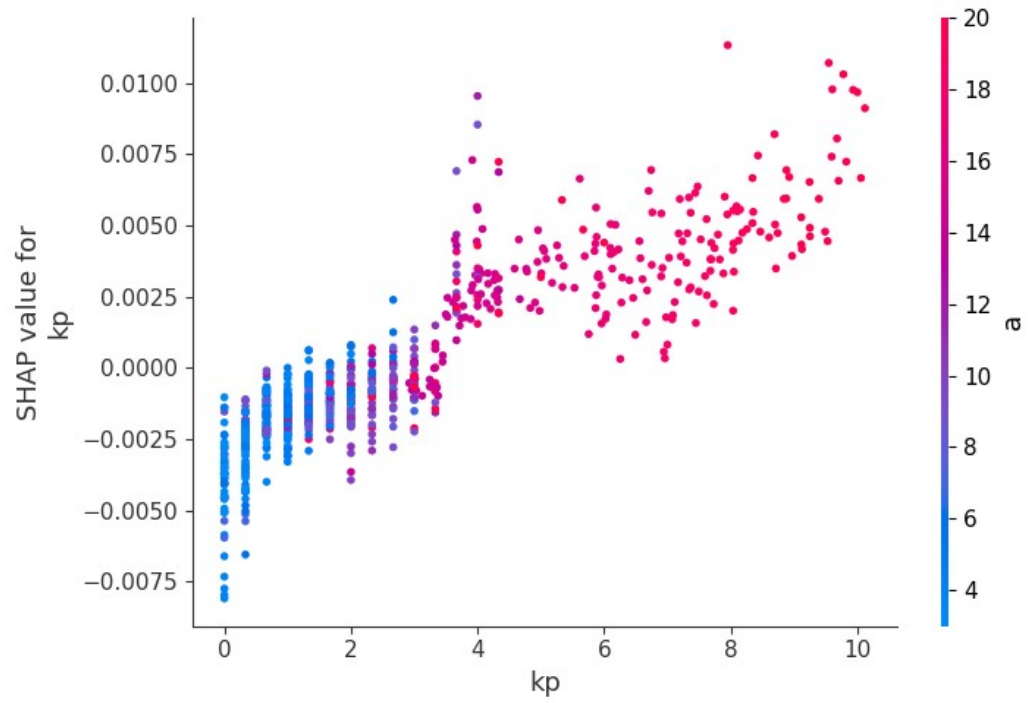
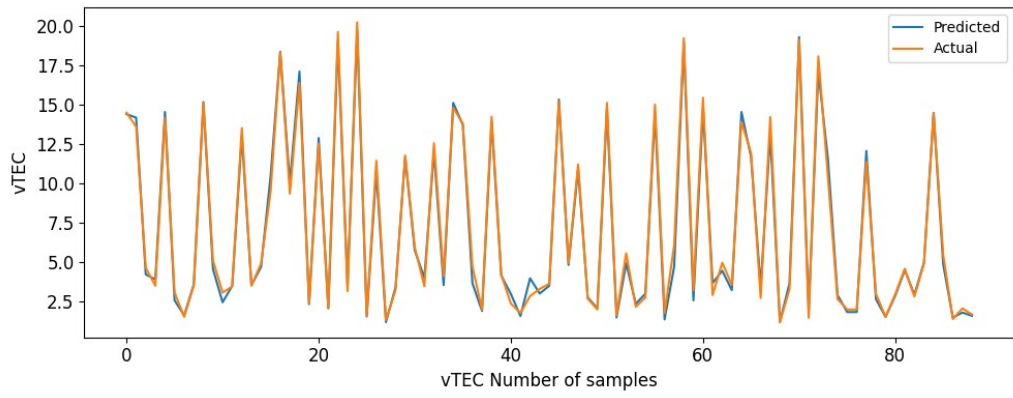


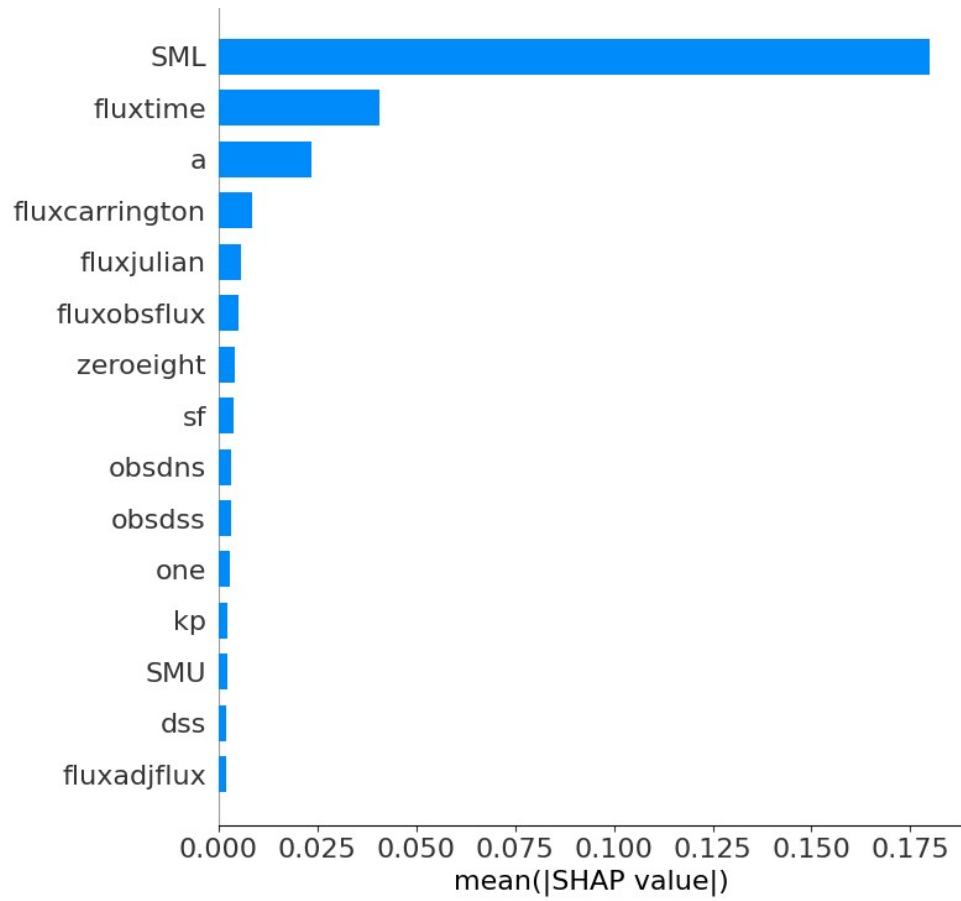
	rMSE	MAE	R <sup>2</sup>
Linear Regression	3.2921	2.6615	0.5315
Decision Tree	3.4978	2.9497	0.4711
Support Vector Regression	3.1240	2.4845	0.5781
Gaussian Process Regressor	3.4547	2.6724	0.4841
XGBOOST Regressor	1.1250	0.7744	0.9453
Multi-layer Perceptron regressor	3.1943	2.5356	0.5589
LightGBM regressor	0.9127	0.6433	0.9640
Random Forest Regressor	0.0630	0.0454	0.9317



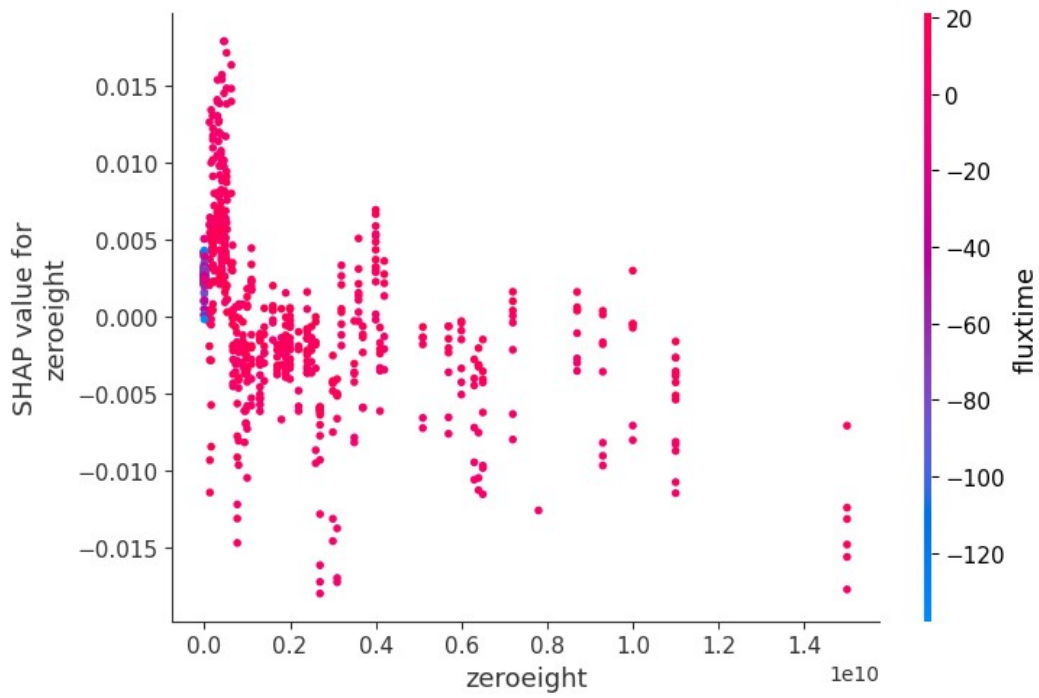
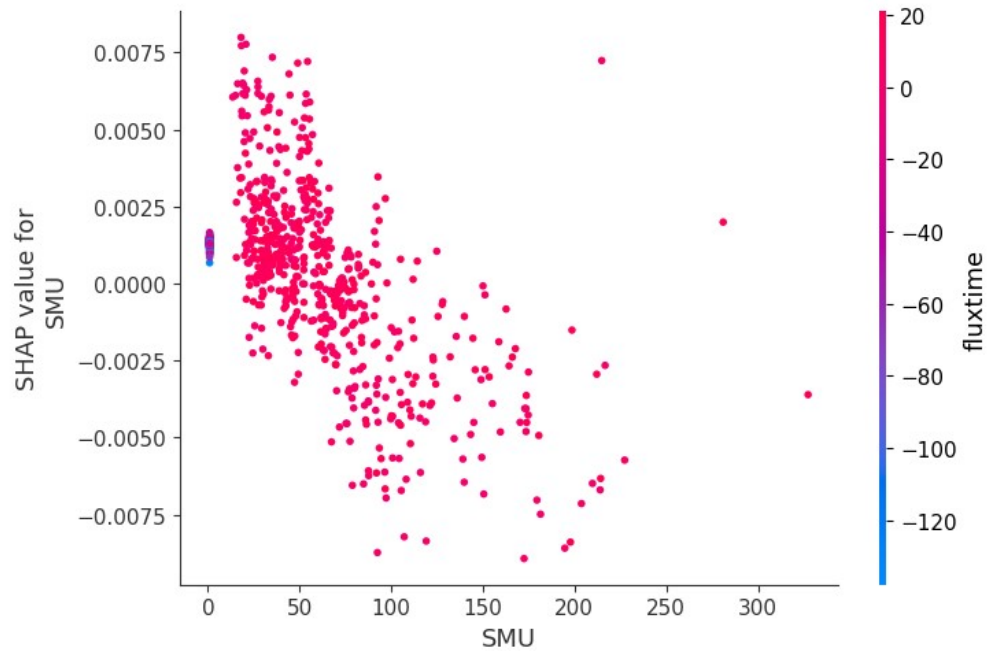
VTEC ΧΙΑΖΣ

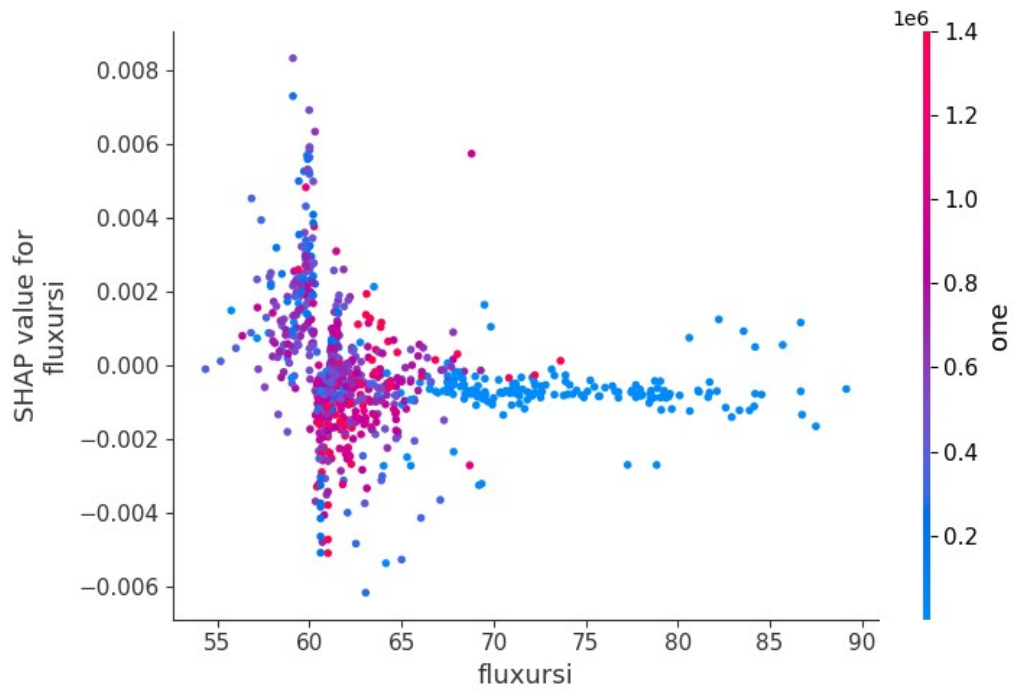
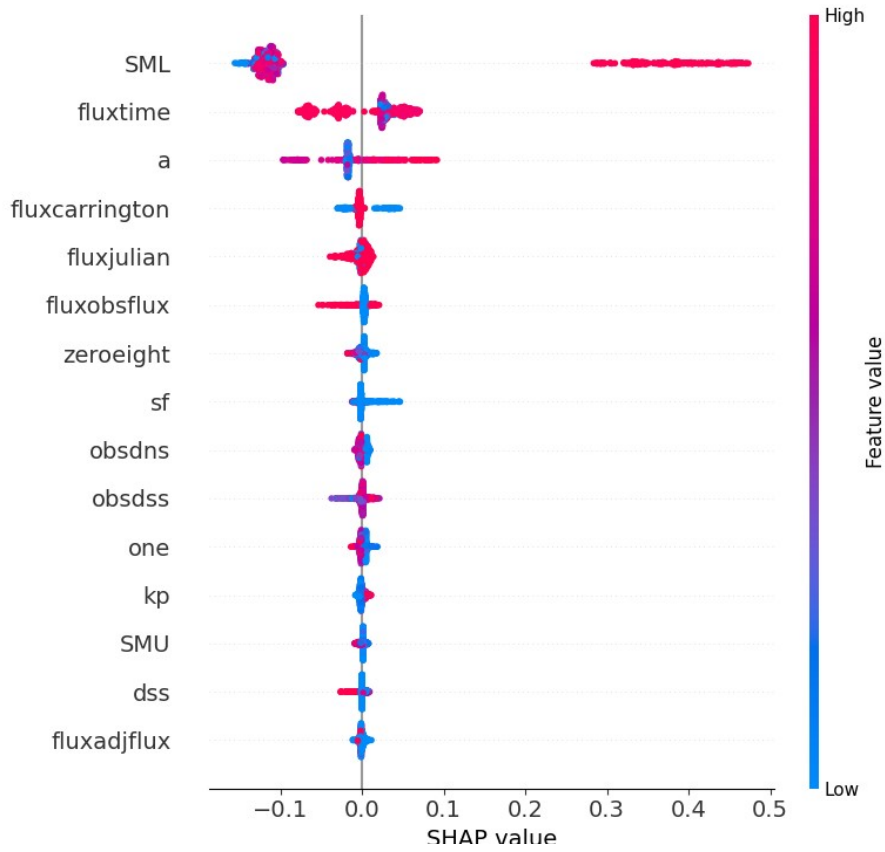




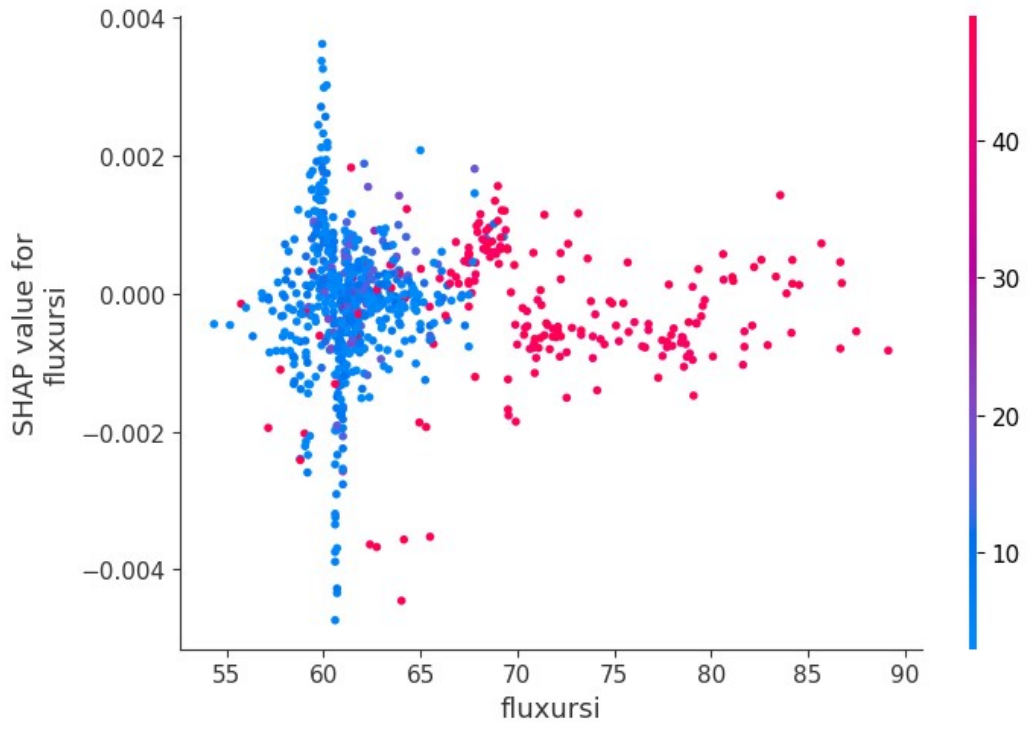


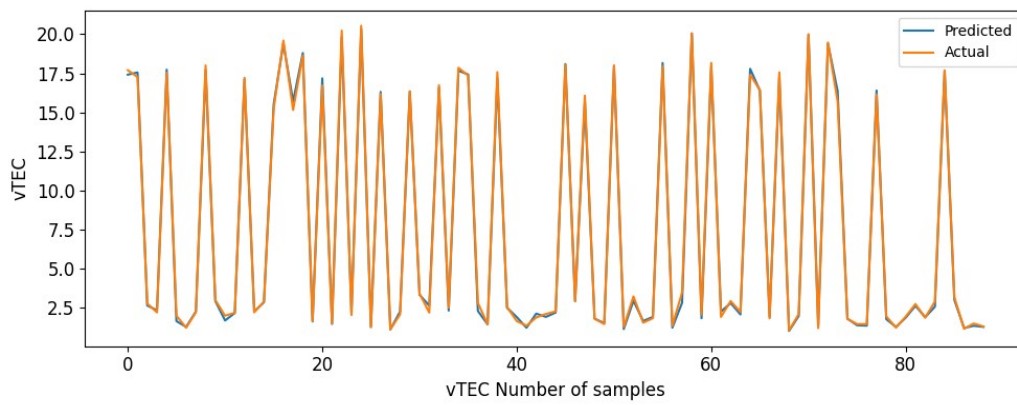
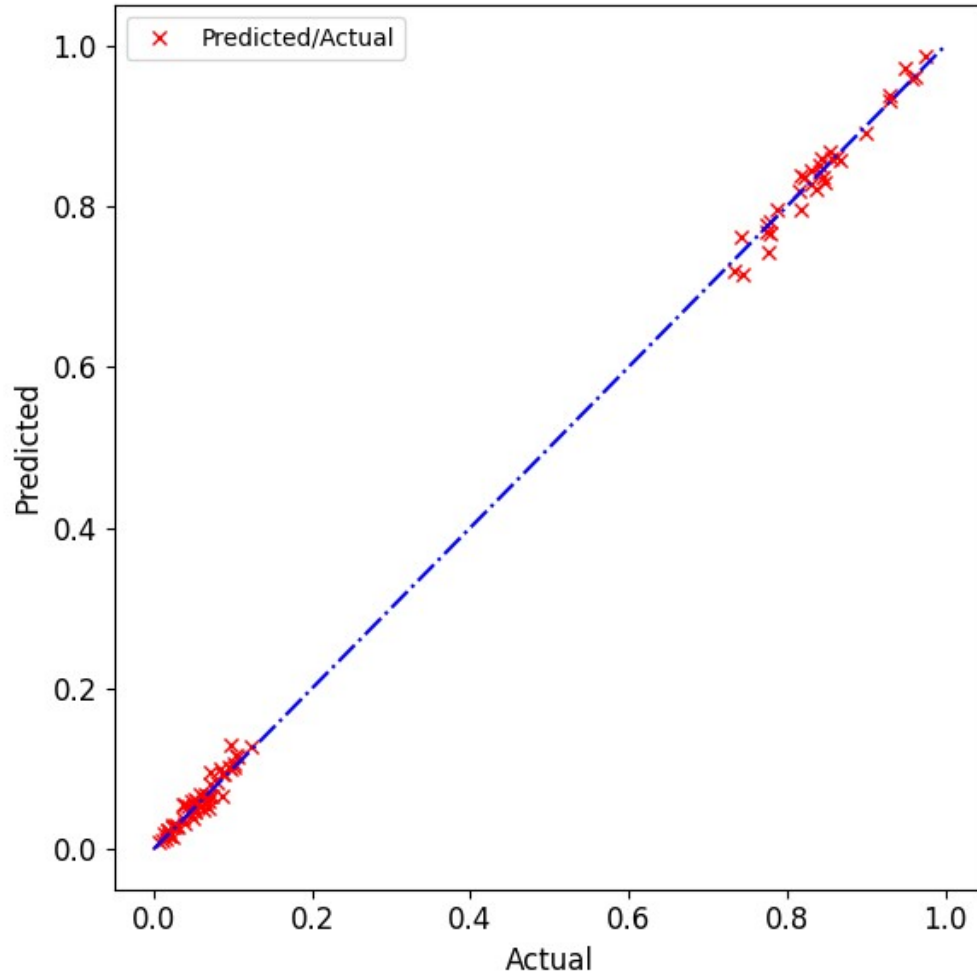
	rMSE	MAE	R <sup>2</sup>
Linear Regrassion	1.2814	1.0328	0.9505
Decision Tree	1.4146	1.1494	0.9396
Support Vector Regression	1.1811	0.9836	0.9579
Gaussian Process Regressor	1.2665	1.0193	0.9516
XGBOOST Regressor	0.4798	0.3668	0.9931
Multi-layer Perceptron regressor	1.1600	0.9370	0.9594
LightGBM regressor	0.4387	0.3254	0.9942
Random Forest Regressor	0.0292	0.0226	0.9898



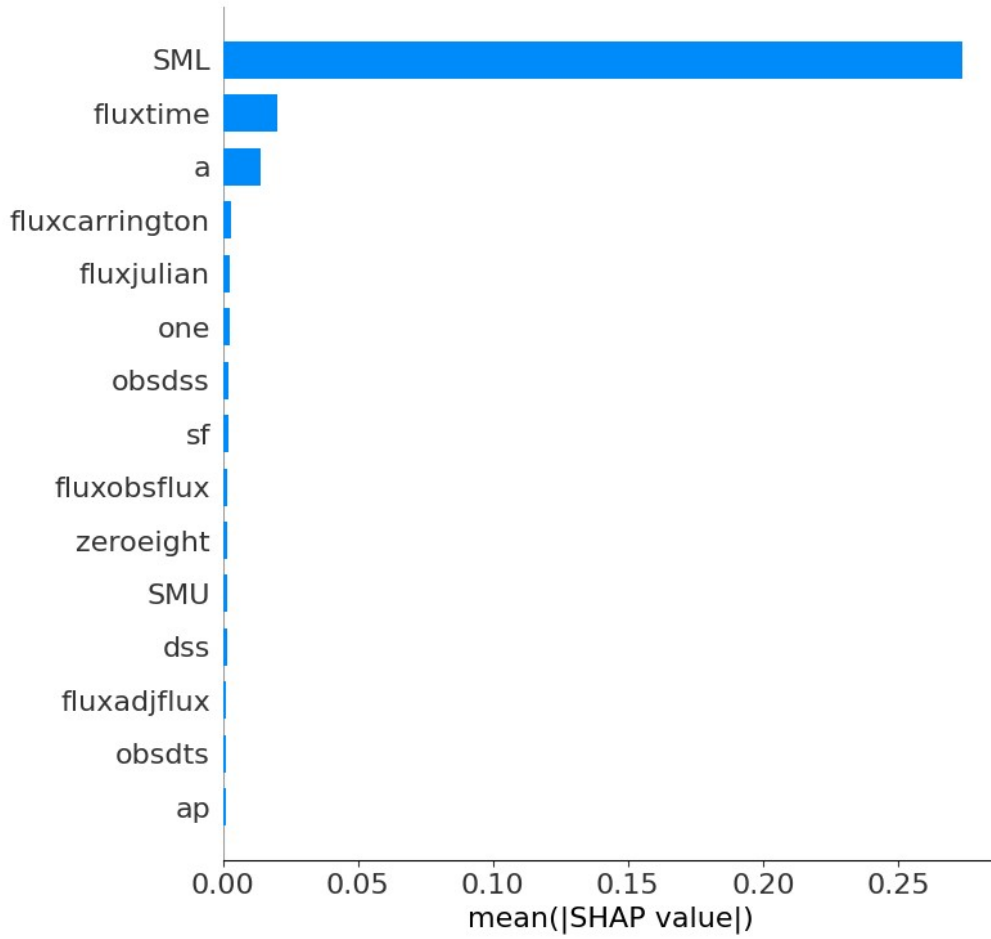


STEC ΗΠΑ

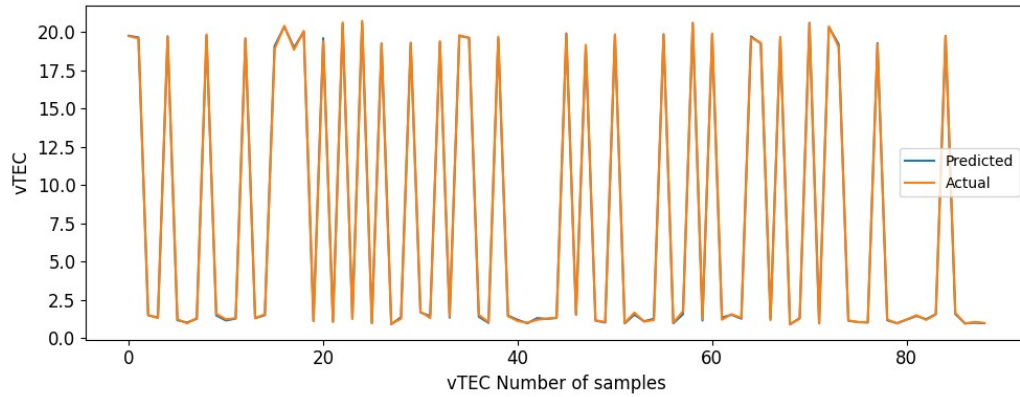






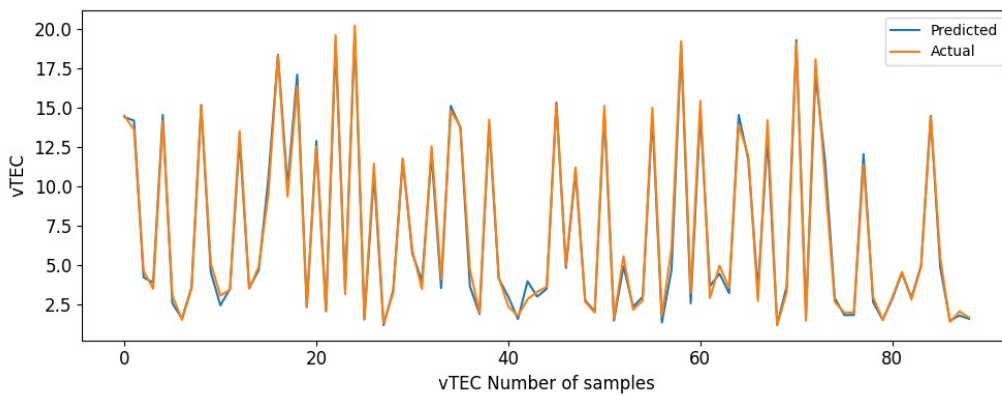
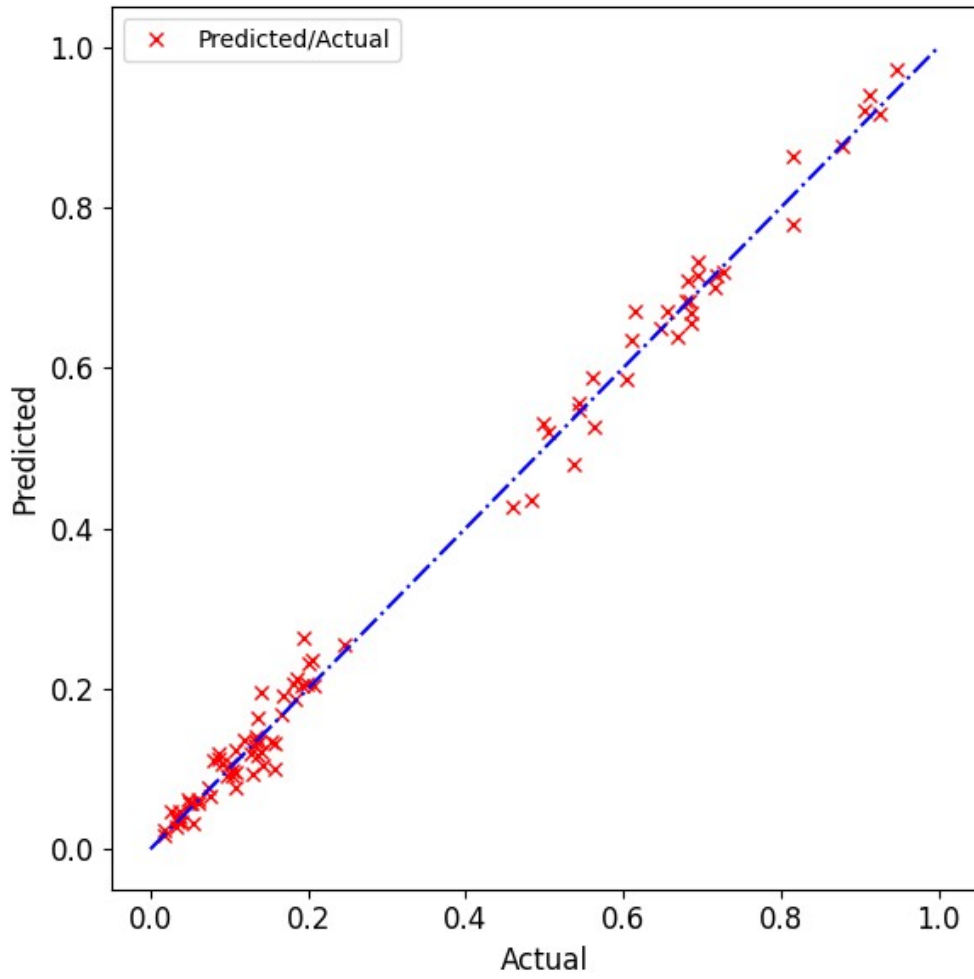


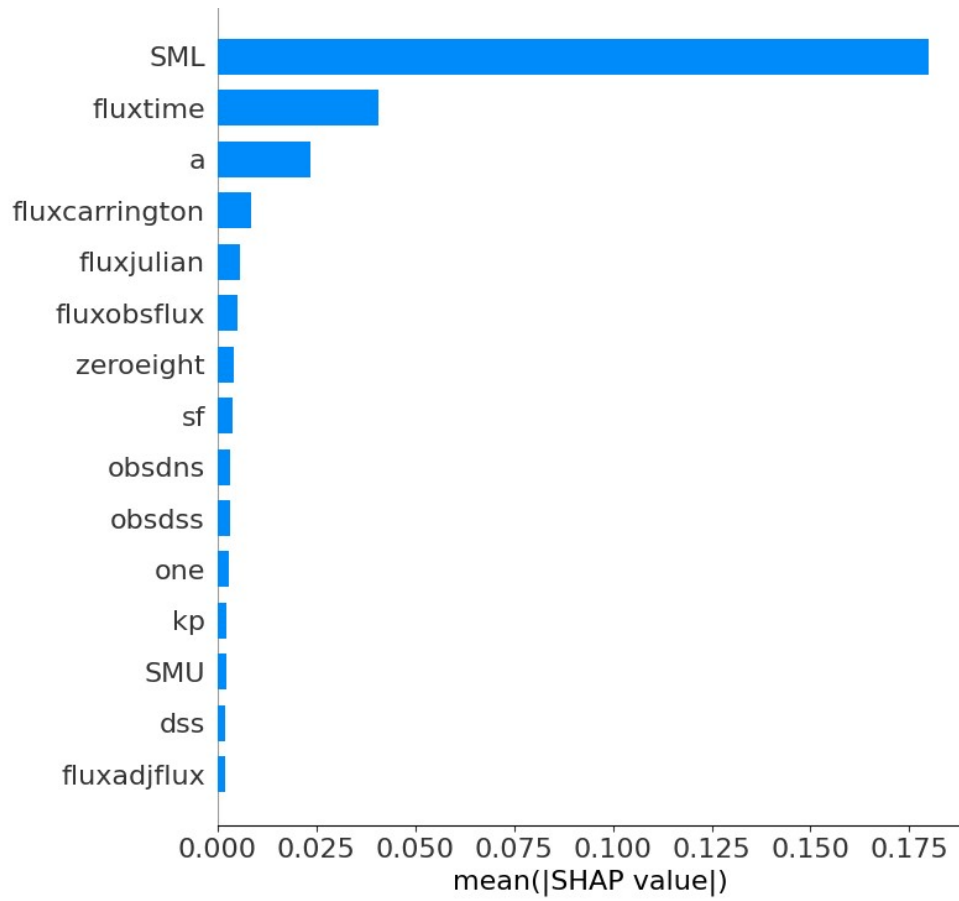
	rMSE	MAE	R <sup>2</sup>
Linear Regrassion	0.6349	0.5117	0.9930
Decision Tree	0.7009	0.5695	0.9914
Support Vector Regression	0.8914	0.7345	0.9862
Gaussian Process Regressor	0.6297	0.5059	0.9931
XGBOOST Regressor	0.2253	0.1682	0.9991
Multi-layer Perceptron regressor	0.5764	0.4753	0.9942
LightGBM regressor	0.2115	0.1611	0.9992
Random Forest Regressor	0.0148	0.0113	0.9985



	rMSE	MAE	R <sup>2</sup>
Linear Regression	0.2228	0.1796	0.9994
Decision Tree	0.2460	0.1998	0.9992
Support Vector Regression	1.0851	0.9904	0.9850
Gaussian Process Regressor	0.2226	0.1792	0.9994
XGBOOST Regressor	0.0754	0.0591	0.9999
Multi-layer Perceptron regressor	0.2417	0.1989	0.9993
LightGBM regressor	0.0747	0.0562	0.9999
Random Forest Regressor	0.0051	0.0039	0.9999

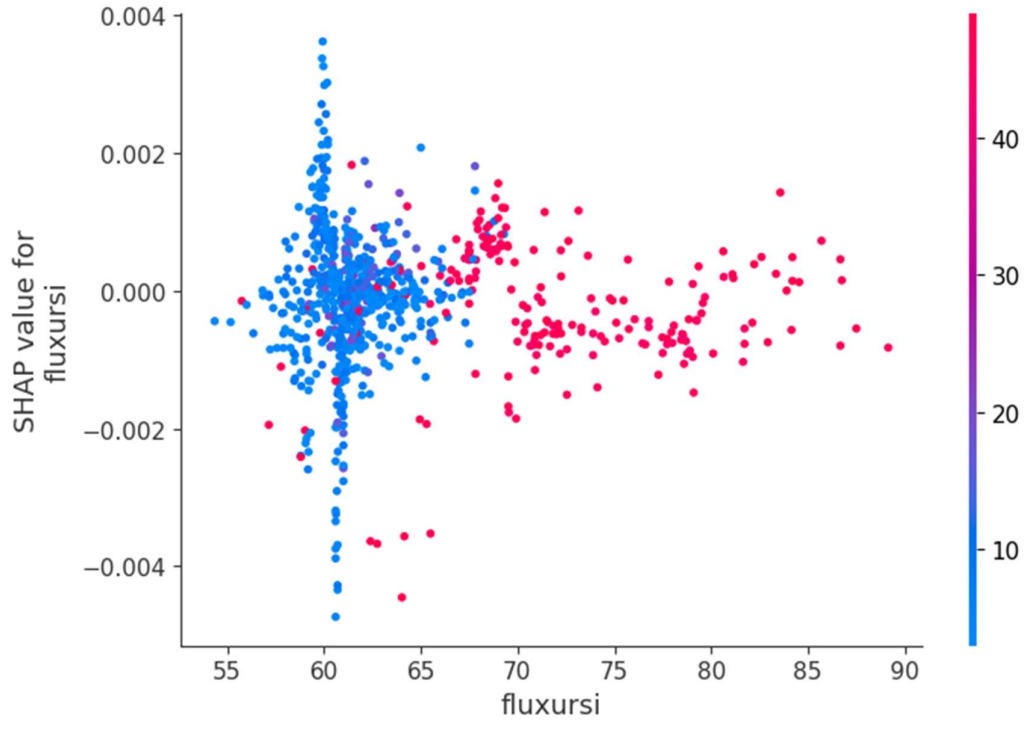
VTEC KENYAS

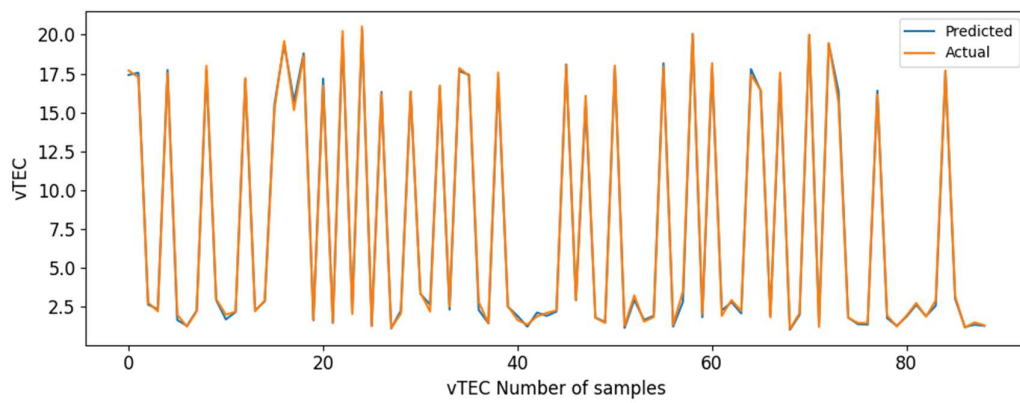
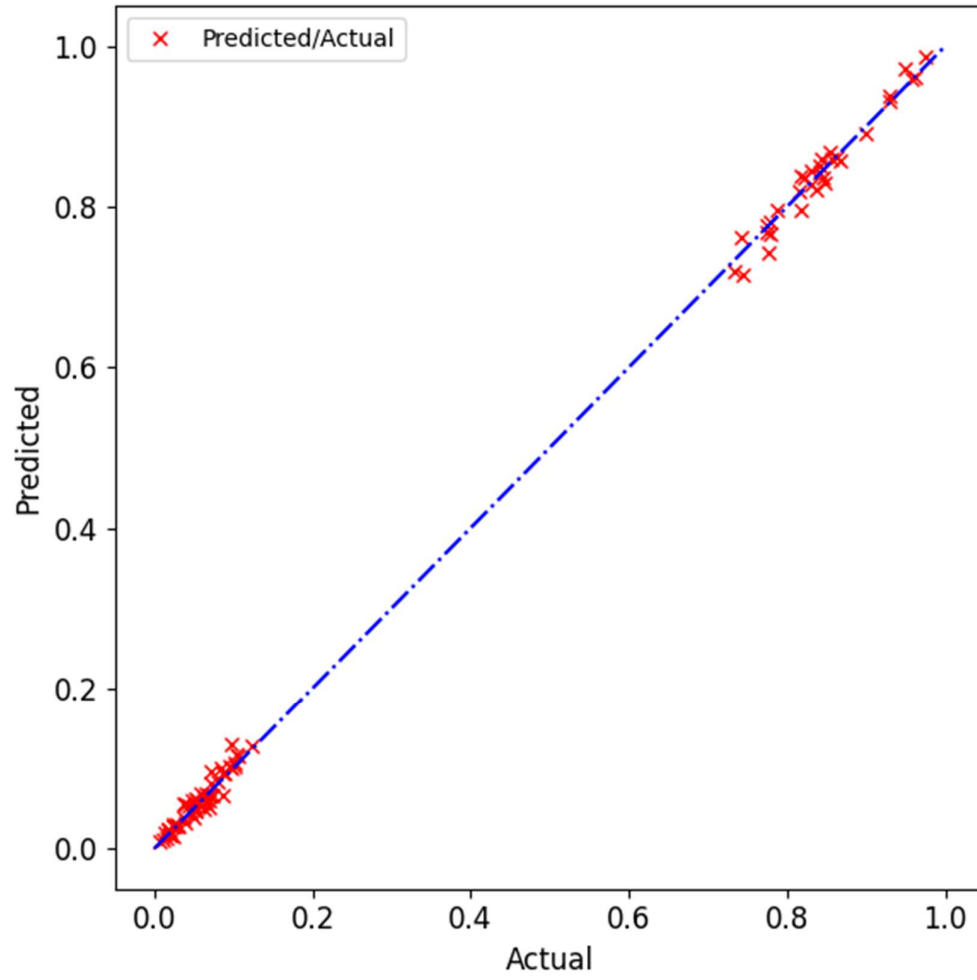


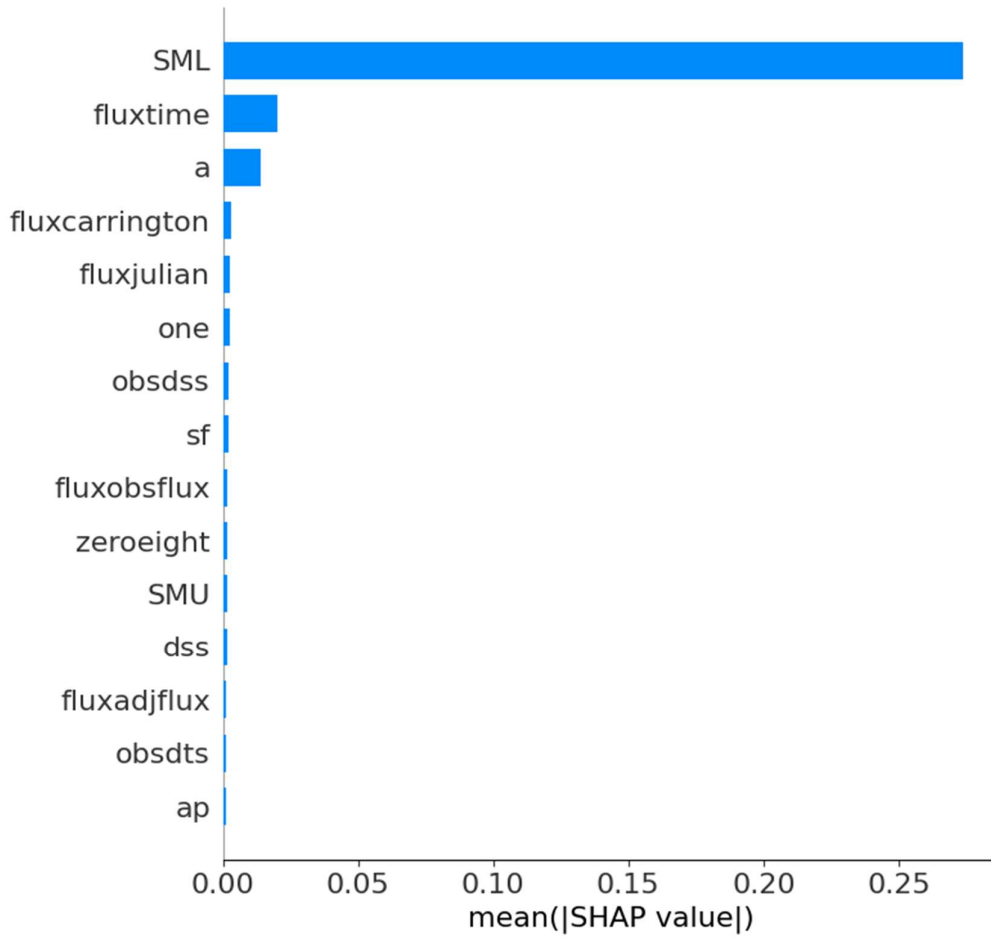


	rMSE	MAE	R <sup>2</sup>
Linear Regrassion	1.2814	1.0328	0.9505
Decision Tree	1.4146	1.1494	0.9396
Support Vector Regression	1.1811	0.9836	0.9579
Gaussian Process Regressor	1.2665	1.0193	0.9516
XGBOOST Regressor	0.4798	0.3668	0.9931
Multi-layer Perceptron regressor	1.1600	0.9370	0.9594
LightGBM regressor	0.4387	0.3254	0.9942
Random Forest Regressor	0.0292	0.0226	0.9898

VTEC ΜΟΓΓΟΛΙΑΣ







	rMSE	MAE	R <sup>2</sup>
Linear Regression	0.6349	0.5117	0.9930
Decision Tree	0.7009	0.5695	0.9914
Support Vector Regression	0.8914	0.7345	0.9862
Gaussian Process Regressor	0.6297	0.5059	0.9931
XGBOOST Regressor	0.2253	0.1682	0.9991
Multi-layer Perceptron regressor	0.5764	0.4753	0.9942
LightGBM regressor	0.2115	0.1611	0.9992
Random Forest Regressor	0.0148	0.0113	0.9985

---

---

## Παράρτημα Γ

---

---

### ΚΩΔΙΚΑΣ ΕΞΗΓΗΣΙΜΗΣ ΜΑΘΗΣΗΣ

---

#### STEC

---

```
from google.colab import drive
drive.mount('/content/drive')
```

#### Importing Libraries

```
import pandas as pd
import numpy as np
from sklearn.metrics import mean_squared_error
import sklearn
import shap
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
from sklearn.metrics import mean_absolute_error
import xgboost
import lime
import lime.lime_tabular
from sklearn.metrics import r2_score
```

#### Reading and scaling the data

```
import os

# ALLAZW TO ONOMA ARXEIOY
path = '/content/drive/MyDrive/dedomena/master/Full-
Table_Norway_FINAL_.xlsx'

path_results = '/content/drive/MyDrive/dedomena/master/results/'
```



```
ext = os.path.basename(path).split('.')[0]
print(ext)

df = pd.read_excel(path)
print(df.keys())
print(df.shape)

df.describe().to_csv(path_results+ext+"stec_stats.csv")

df2 = df[['stec', 'vtec', 'a', 'kp', 'ap', 'D', 'SML', 'SMU',
          'fluxtime', 'fluxjulian', 'fluxcarrington', 'fluxobsflux',
          'fluxadjflux', 'fluxursi', 'dts', 'dns', 'dss', 'sddts',
          'sddns',
          'sddss', 'obsdts', 'obsdns', 'obsdss', 'dpind', 'one',
          'ten', 'hundred',
          'zeroeight', 'two', 'nm', 'sf', 'pa', 'ang', 'lin',
          'height', 'speed',
          'acc', 'mpa']]

print(df2.dtypes)

col_ytrain = df2['stec']
max = col_ytrain.max()
min = col_ytrain.min()
dif = max - min
#print(type(df.max()))
#print(type(df.min()))
#print(dif)

df_for_training=(df2-df2.min())/(df2.max()-df2.min()).dropna()
# print(df_for_training.min())
# print(df_for_training.max())

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

print(trainX1.keys())
print(trainX1.min())
print(trainX1.max())
```

```
trainY1 = df_for_training[["stec"]],
"total_deaths_per_million"]
print(trainY1.keys())
print(trainY1.min())
print(trainY1.max())

#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 = 0.1, shuffle=True, random_state = 42)
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.ravel())

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

#preds_lr= preds_lr*dif + min
#y_test = y_test*dif + min

rmse_lr = np.sqrt(mean_squared_error(y_test*dif + min,
preds_lr*dif + min))
mae_lr = mean_absolute_error(y_test*dif + min, preds_lr*dif +
min)
r2_score_lr = r2_score(y_test, preds_lr)
print("Linear Regrassion: RMSE on test set:
{:.4f}".format(rmse_lr))
print("Linear Regrassion: MAE on test set:
{:.4f}".format(mae_lr))
print("Linear Regrassion: R^2 on test set:
{:.4f}".format(r2_score_lr))
print("")

# 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*dif + min,
preds_dt*dif + min))
mae_dt = mean_absolute_error(y_test*dif + min, preds_dt*dif +
min)
r2_score_DT = r2_score(y_test, preds_dt)
print("Decision Tree Regressor: RMSE on test set:
{:.4f}".format(rmse_dt))
print("Decision Tree Regressor: MAE on test set:
{:.4f}".format(mae_dt))
print("Decision Tree Regressor: R^2 on test set:
{:.4f}".format(r2_score_DT))
print("")

# 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*dif + min,
preds_svr*dif + min))
mae_svr = mean_absolute_error(y_test*dif + min, preds_svr*dif +
min)
r2_score_SVR = r2_score(y_test, preds_svr)
print("Support Vector Regression: RMSE on test set:
{:.4f}".format(rmse_svr))
print("Support Vector Regression: MAE on test set:
{:.4f}".format(mae_svr))
print("Support Vector Regression: R^2 on test set:
{:.4f}".format(r2_score_SVR))
print("")

# Lasso Regression
lassoReg = linear_model.Lasso(alpha=0.5)
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*dif + min,
preds_lasso*dif + min))
mae_lasso = mean_absolute_error(y_test*dif + min, preds_lasso*dif
+ min)
r2_score_lasso = r2_score(y_test, preds_lasso)
print("Lasso Regression: RMSE on test set:
{:.4f}".format(rmse_lasso))
print("Lasso Regression: MAE on test set:
{:.4f}".format(mae_lasso))
print("Lasso Regression: R^2 on test set:
{:.4f}".format(r2_score_lasso))
print("")

# 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*dif + min,
preds_gpr*dif + min))
mae_gpr = mean_absolute_error(y_test*dif + min, preds_gpr*dif +
min)
r2_score_gpr = r2_score(y_test, preds_gpr)
print("Gaussian Process Regressor: RMSE on test set:
{:.4f}".format(rmse_gpr))
print("Gaussian Process Regressor: MAE on test set:
{:.4f}".format(mae_gpr))
print("Gaussian Process Regressor: R^2 on test set:
{:.4f}".format(r2_score_gpr))
print("")

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

# RMSE
preds_rfg = regressor.predict(x_test)

#preds_rfg= preds_rfg*dif + min

#print("preds shape : ", preds.shape)
```

```
#print("y_test shape : ", y_test.shape)

rmse_rfg = np.sqrt(mean_squared_error(y_test, preds_rfg))
mae_rfg = mean_absolute_error(y_test, preds_rfg)
r2_score_rfg = r2_score(y_test, preds_rfg)
print("Random Forest Regression: RMSE on test set:
 {:.4f}".format(rmse_rfg))
print("Random Forest Regression: MAE on test set:
 {:.4f}".format(mae_rfg))
print("Random Forest Regression: R^2 on test set:
 {:.4f}".format(r2_score_rfg))
print("")

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

# RMSE

preds_xgb = XGBR.predict(x_test)

#preds_xgb= preds_xgb*dif + min

rmse_xgb = np.sqrt(mean_squared_error(y_test*dif + min,
preds_xgb*dif + min))
mae_xgb = mean_absolute_error(y_test*dif + min, preds_xgb*dif +
min)
r2_score_xgb = r2_score(y_test, preds_xgb)
print("XGBR : RMSE on test set: {:.4f}".format(rmse_xgb))
print("XGBR : MAE on test set: {:.4f}".format(mae_xgb))
print("XGBR : R^2 on test set: {:.4f}".format(r2_score_xgb))
print("")

from sklearn.neural_network import MLPRegressor
regr_MLP = MLPRegressor(hidden_layer_sizes=1000, random_state=1,
max_iter=5000, batch_size= 32, early_stopping= True).fit(x_train,
y_train)

preds_regr_MLP = regr_MLP.predict(x_test)

rmse_regr_MLP = np.sqrt(mean_squared_error(y_test*dif + min,
preds_regr_MLP*dif + min))
mae_regr_MLP = mean_absolute_error(y_test*dif + min,
preds_regr_MLP*dif + min)
r2_score_MLP = r2_score(y_test, preds_regr_MLP)
print("MLP: RMSE on test set: {:.4f}".format(rmse_regr_MLP))
print("MLP: MAE on test set: {:.4f}".format(mae_regr_MLP))
print("MLP: R^2 on test set: {:.4f}".format(r2_score_MLP))
print("")
```

```
import lightgbm as ltb

LGBMR = ltb.LGBMRegressor()
LGBMR.fit(x_train, y_train)

preds_LGBMR = LGBMR.predict(x_test)

rmse_LGBMR = np.sqrt(mean_squared_error(y_test*dif + min,
preds_LGBMR*dif + min))
mae_LGBMR = mean_absolute_error(y_test*dif + min, preds_LGBMR*dif
+ min)
r2_score_LGBMR = r2_score(y_test, preds_LGBMR)
print("LGBMR : RMSE on test set: {:.4f}".format(rmse_LGBMR))
print("LGBMR : MAE on test set: {:.4f}".format(mae_LGBMR))
print("LGBMR : R^2 on test set: {:.4f}".format(r2_score_LGBMR))
```

```
import matplotlib.pyplot as plt

plt.rcParams["figure.figsize"] = [9.00, 10.50]
plt.rcParams["figure.autolayout"] = True

fig, ax =plt.subplots(1,1)

data=[[rmse_lr,mae_lr,r2_score_lr],
      [rmse_dt,mae_dt, r2_score_DT],
      [rmse_svr,mae_svr, r2_score_SVR],
      [rmse_gpr,mae_gpr, r2_score_gpr],
      [rmse_xgb,mae_xgb, r2_score_xgb],
      [rmse_regr_MLP,mae_regr_MLP, r2_score_MLP],
      [rmse_LGBMR,mae_LGBMR, r2_score_LGBMR],
      [rmse_rfg,mae_rfg, r2_score_rfg]
      ]

column_labels=["rMSE", "MAE", "R^2"]

row=["Linear Regrassion ", "Decision Tree", "Support Vector
Regression", "Gaussian Process Regressor", "XGBOOST Regressor",
"Multi-layer Perceptron regressor", "LightGBM regressor", "Random
Forest 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,rowLabels=row,fo
ntsize=15,loc="center",cellLoc='center')
```

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

```
import pandas as pd
import time
import numpy as np
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure

figure(figsize=(10, 4))

start_time = time.time()
importances = regressor.feature_importances_
std = np.std([tree.feature_importances_ for tree in
regressor.estimators_], axis=0)
elapsed_time = time.time() - start_time

print(f"Elapsed time to compute the importances:
{elapsed_time:.3f} seconds")

forest_importances = pd.Series(importances, index=trainX1.keys())

fig, ax = plt.subplots(figsize=(10, 6))
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)
forest_importances.plot.bar(yerr=std, ax=ax)
ax.set_title("Feature importances using MDI", fontsize=16)
ax.set_ylabel("Mean decrease in impurity", fontsize=14)
ax.set_xlabel("Indicators", fontsize=14)
fig.tight_layout()
```

```
import statsmodels.api as sm

# #define response variable
# y = df['y']

# #define predictor variables
# x = df[['x1', 'x2']]

#add constant to predictor variables
x = sm.add_constant(trainX1)

#fit linear regression model
model = sm.OLS(trainY1, trainX1).fit()

#view model summary
```

```
print(model.summary())
```

#### Table of comparisons

```
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure

figure(figsize=(10, 4))
plt.yticks(fontsize=12)
plt.xticks(fontsize=12)

plt.ylabel('STEC', fontsize=12)
plt.xlabel('STEC Number of samples', fontsize=12)

plt.plot(preds_xgb*dif + min, label = 'Predicted')
plt.plot(y_test*dif + min, label = 'Actual')
plt.legend()
plt.show()
```

```
figure(figsize=(6, 6))
plt.yticks(fontsize=12)
plt.xticks(fontsize=12)

plt.ylabel('Predicted', fontsize=12)
plt.xlabel('Actual', fontsize=12)

plt.plot(preds_xgb, y_test, 'rx')

plt.plot([0,1], 'b-.')
plt.legend(['Predicted/Actual'])
# To show the plot
plt.show()
```

#### SHAP – RandomForestRegressor

```
# Create object that can calculate shap values
explainer = shap.TreeExplainer(XGBR)
# Calculate Shap values
shap_values = explainer.shap_values(x_train)
```

```
print(explainer.expected_value)
# SHAP Summary Plot
#shap.summary_plot(shap_values, x_train)
```



```
fig = shap.summary_plot(shap_values, x_train,
feature_names=trainX1.keys(), max_display=15, plot_type="bar",
show=False)
plt.yticks(fontsize=16)
plt.xticks(fontsize=16)
plt.xlabel('mean(|SHAP value|)', fontsize=16)
```

```
import matplotlib.pyplot as plt
f = plt.figure()
fig = shap.summary_plot(shap_values,
x_train,feature_names=trainX1.keys(), max_display=15, show=False)
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)
plt.xlabel('SHAP value', fontsize=14)
```

```
df3 = df2[['a', 'kp', 'ap', 'SML', 'SMU',
'fluxtime', 'fluxjulian', 'fluxcarrington', 'fluxobsflux',
'fluxadjflux', 'fluxursi', 'dts', 'dns', 'dss', 'sddts',
'sddns',
'sddss', 'obsdts', 'obsdns', 'obsdss', 'one', 'ten',
'hundred',
'zeroeight', 'sf', 'pa', 'ang', 'lin', 'height', 'speed',
'acc', 'mpa' ]]
```

```
shap_x_train = pd.DataFrame(x_train, columns = df3.keys())

denorm_shap_x_train = shap_x_train*(df3.max()-df3.min()) +
df3.min()
```

```
for i in range(32):
    plt.figure()
    fig = shap.dependence_plot(i,shap_values,
denorm_shap_x_train,feature_names=trainX1.keys(), show=False)
    plt.show
```