



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ  
ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ  
Δ.Π.Μ.Σ. «ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ»

**Σύγκριση Στατιστικών Μεθόδων και Μεθόδων Υπολογιστικής  
Νοημοσύνης για Βραχυπρόθεσμη Πρόβλεψη Μονομεταβλητών  
Κυκλοφοριακών Χρονοσειρών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΚΑΜΠΙΤΑΚΗ Φ. ΕΜΜΑΝΟΥΗΛ**

**Επιβλέπουσα :** Βλαχογιάννη Ελένη  
Αναπληρώτρια Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2020





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ  
Δ.Π.Μ.Σ. «ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ»

## **Σύγκριση Στατιστικών Μεθόδων και Μεθόδων Υπολογιστικής Νοημοσύνης για Βραχυπρόθεσμη Πρόβλεψη Κυκλοφοριακών Χρονοσειρών**

Η παρούσα Διπλωματική Εργασία  
εκπονήθηκε στα πλαίσια των σπουδών  
για την απόκτηση  
**του Μεταπτυχιακού Διπλώματος Ειδίκευσης**  
**«Εφαρμοσμένες Μαθηματικές Επιστήμες»**  
που απονέμει η  
**Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών**  
**του**  
**Εθνικού Μετσόβιου Πολυτεχνείου**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή:

Βλαχογιάννη Ελένη  
Αναπληρώτρια  
Καθηγήτρια Ε.Μ.Π.

Γιαννής Γεώργιος  
Καθηγητής Ε.Μ.Π.

Καρόνη Χρυσίς  
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2020

**ΚΑΜΠΙΤΑΚΗΣ Φ. ΕΜΜΑΝΟΥΗΛ**

Διπλωματούχος Πολιτικός Μηχανικός Ε.Μ.Π.

© 2020 – All rights reserved

*Στους γονείς μου*



*«Όλα τα μοντέλα είναι λάθος,  
αλλά κάποια είναι χρήσιμα.»*

**George E. P. Box**





# **Σύγκριση Στατιστικών Μεθόδων και Μεθόδων Υπολογιστικής Νοημοσύνης για Βραχυπρόθεσμη Πρόβλεψη Μονομεταβλητών Κυκλοφοριακών Χρονοσειρών**

**Καμπιτάκης Φ. Εμμανουήλ**

**Επιβλέπουσα: Βλαχογιάννη Ι. Ελένη**

## **Περίληψη**

Η βραχυπρόθεσμη πρόβλεψη κυκλοφορίας αποτελεί μία από τις σημαντικότερες προκλήσεις των Ευφυών Μεταφορικών Συστημάτων καθώς η κυκλοφοριακή συμφόρηση και τα επακόλουθα που δημιουργεί είναι από τα σημαντικότερα προβλήματα των σύγχρονων μεγάλων πόλεων. Αν και μελετάται εκτενώς στην βιβλιογραφία για πάνω από 5 δεκαετίες, η ραγδαία εξέλιξη των τελευταίων χρόνων στον κλάδο της μηχανικής μάθησης και της βαθιάς μηχανικής μάθησης καθώς και η εμφάνιση νέων πηγών πληροφοριών, οι οποίες μας προσφέρουν πλέον περισσότερα και ποιοτικότερα δεδομένα, έχουν ανοίξει νέους ορίζοντες για μία αποτελεσματικότερη και πιο ακριβής βραχυπρόθεσμη πρόβλεψη κυκλοφοριακών μεγεθών. Παρόλα αυτά, λίγα έχουν γίνει σε επίπεδο συγκριτικής ανάλυσης με σκοπό την πραγματική κατανόηση των νέων μεθόδων και πως εκείνες θα μπορούσαν να βοηθήσουν ουσιαστικά σε εφαρμογές πραγματικού χρόνου διαχείρισης κυκλοφορίας.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η διεξαγωγή συγκριτικής ανάλυσης 15 διαφορετικών μεθόδων πρόβλεψης, από τις οποίες οι 10 αφορούν σε μοντέλα Βαθιάς Μηχανικής Μάθησης (MLP, LSTM, GRU, CNN, Dilated CNN, CNN-LSTM, Conv-LSTM, Encoder-Decoder LSTM, Encoder-Decoder CNN-LSTM και Encoder-Decoder Conv-LSTM) ενώ οι υπόλοιπες 5 αφορούν σε γνωστές στατιστικές μεθόδους χρονοσειρών (STL αποσύνθεση και ARIMA, ARIMAX, STL αποσύνθεση και ETS, TBATS και Theta) για βραχυπρόθεσμη πρόβλεψη μονομεταβλητών κυκλοφοριακών χρονοσειρών. Οι τεχνικές πρόβλεψης εφαρμόζονται σε χρονοσειρές μέσης ταχύτητα διαδρομής (ανά 5 λεπτά) σε μία αρτηρία την πόλης Xi'an στην Κίνα. Οι χρονοσειρές έχουν προκύψει από ανεπεξέργαστα GPS δεδομένα των ταξί της εταιρίας DiDi. Απώτερος σκοπός είναι η ανάδειξη δυσκολιών που σχετίζονται με την εκπαίδευση τέτοιων μοντέλων (Βαθιάς Μηχανικής Μάθησης) και πόσο εύκολο είναι να ξεπεράσουν σε προβλεπτική

ικανότητα τις παραδοσιακές στατιστικές μεθόδους. Εξετάστηκαν επίσης, δύο μέθοδοι προεπεξεργασίας δεδομένων, η αφαίρεση της εποχικής και της συνιστώσας τάσης, δανεισμένες από την κλασική ανάλυση χρονοσειρών, και πως αυτές επηρεάζουν την επίδοση μοντέλων Βαθιάς Μηχανικής Μάθησης. Για τον υπολογισμό της 1<sup>ης</sup> συνιστώσας έγινε με την μέθοδο “LOESS”, ενώ της 2<sup>ης</sup> με την μέθοδο των κυβικών σπλινών. Ως μέθοδος αξιολόγησης της προβλεπτικής ικανότητας των μοντέλων χρησιμοποιήθηκε η μέθοδος κυλιόμενης αρχής χωρίς επανεκπαίδευση με σταθερή τιμή του ορίζοντα πρόβλεψης, για 5 διαφορετικές μετρικές σφάλματος. Ο ορίζοντας πρόβλεψης ορίστηκε από 5 λεπτά έως 60 λεπτά (1-12 βήματα). Το πρόβλημα της πρόβλεψης διατυπώθηκε ως πρόβλημα αυτοπαλινδρόμησης, AR(p), χωρίς εξωγενείς μεταβλητές. Υπολογίστηκε η υπολογιστική πολυπλοκότητα για κάθε μέθοδο η οποία ορίζεται ως η επιπλέον «προσπάθεια» (πολλαπλασιαστικά), σε σύγκριση με την μέθοδο αναφοράς, που απαιτείται για την αντίστοιχη μείωση του σφάλματος πρόβλεψης. Επιπροσθέτως, υπολογίστηκε η ποσοστιαία προσαρμογή κάθε μεθόδου, κανονικοποιημένη από το τετράγωνο της μέσης τιμής της χρονοσειράς, και εξετάστηκε γραφικά η σχέση της με την προβλεπτική ικανότητα. Τέλος, διερευνήθηκε και πάλι γραφικά η ευαισθησία των μη παραμετρικών μοντέλων στον χώρο των παραμέτρων που εξετάστηκαν.

Τα αποτελέσματα δείχνουν ότι όλες οι μέθοδοι έχουν καλύτερη προβλεπτική ικανότητα από την μέθοδο αναφοράς, ενώ φαίνεται οι μέθοδοι Βαθιάς Μηχανικής Μάθησης να υπερτερούν των στατιστικών μεθόδων (εκτός της μεθόδου MLP), χωρίς η επιλογή της μετρικής σφάλματος να επηρεάζει την επιλογή της καλύτερης μεθόδου. Καλύτερη μέθοδος από τις στατιστικές μεθόδους για τις μέσες τιμές των σφαλμάτων και για όλες τις μετρικές είναι η μέθοδος ARIMA με εποχική αποσύνθεση, ενώ η αντίστοιχα καλύτερη μέθοδος για τις μεθόδους βαθιάς μηχανικής μάθησης είναι η μέθοδος Encoder-Decoder LSTM, η οποία είναι και η καλύτερη σε προβλεπτική ικανότητα από συνολικά όλες που εξετάστηκαν. Ακόμα, προκύπτει ότι και οι 2 στατιστικές μέθοδοι προεπεξεργασίας που εφαρμόστηκαν είναι στατιστικά σημαντικές για την βελτίωση της απόδοσης των μεθόδων βαθιάς μηχανικής μάθησης. Μάλιστα, η αφαίρεση της εποχικής συνιστώσας βρέθηκε να έχει σημαντική μειωτική επιρροή στο σφάλμα για όλες τις μεθόδους που εξετάστηκαν. Η υπολογιστική πολυπλοκότητα των μεθόδων βαθιάς μηχανικής μάθησης βρέθηκε ότι είναι τρεις τάξεις μεγέθους μεγαλύτερη από την αντίστοιχη των στατιστικών μεθόδων, και τέλος, αναδείχθηκε η ανάγκη αυτοματοποίησης της εκπαίδευσης των πρώτων καθώς και περαιτέρω διερεύνησης σημαντικών παραμέτρων για το πρόβλημα της πρόβλεψης κυκλοφορίας.

# **Comparison of Statistical and Computational Intelligence Methods for Univariate Short-term Traffic Timeseries Forecasting**

**Kampitakis Ph. Emmanouil**

**Supervisor: Vlahogianni I. Eleni**

## **Abstract**

Short-term traffic prediction is one of the most important challenges in ITS systems as traffic congestion and its aftermaths are of paramount importance in metropolitan areas. In the last five decades this subject has been thoroughly studied in dozens of research papers, but the rapid development in machine learning and deep learning, and the emergence of new data sources makes it possible to examine and predict traffic conditions in big cities more accurately than ever. Despite this, there is still not a clear view about the various requirements involved in training these new methods and how to effectively use them in traffic forecasting in order to be essentially helpful in real time traffic management applications.

The aim of this master thesis, is the conduction of a comparative analysis in 15 different forecasting methods, 10 of them are from the area of deep learning (MLP, LSTM, GRU, CNN, Dilated CNN, CNN-LSTM, Conv-LSTM, Encoder-Decoder LSTM, Encoder-Decoder CNN-LSTM and Encoder-Decoder Conv-LSTM) and the rest of them are from the area of statistical time series (STL decomposition and ARIMA, ARIMAX, STL decomposition and ETS, TBATS and Theta). The data are the average speed (1 observation per 5 minutes) from October and November in 2016 for an arterial in the city of Xi'an in China which were extracted from raw GPS data collected from DiDi's taxi drivers. The ultimate goal is to highlight the difficulties associated with the training of such models (Deep Learning) and how easy it is to overcome traditional statistical methods in predictive capability. Two data preprocessing methods were examined, deseasonalization and detrending, borrowed from statistical time series in order to investigate how these two methods affect the performance of deep learning models. For the calculation of the seasonal component the "LOESS" method was used, while for the trend component the cubic splines method was used. The predictive capability of each model was tested using the rolling

origin without retraining and constant horizon method. Also, five different error measures were used. The horizon was set to be from 5 minutes to 60 minutes (1-12 steps). The time series forecasting problem was formalized as an auto-regressive task, i.e., based on an AR(p) model, with no exogenous variables. Computational Complexity (CC) is used to determine the time needed to train a given model and use it for forecasting. Thus, CC can be simply defined as the computational time required by the model to predict a time series, divided by the corresponding time needed by the baseline method to achieve the same task. In this regard, we end up with a relative metric indicating the additional, proportional time required for obtaining the forecasts from the more complex methods. Additionally, the accuracy of how well a model fits the historical data is calculated which is actually the Mean Squared Error (MSE) of the in-sample forecasts, normalized by the squared mean value of the time series being examined. Finally, we examined graphically the sensitivity of each model's average MAE to hyperparameter changes within the bounds searched.

The results indicate that all methods have better predictive performance than the baseline method (Seasonal Naïve method), while the deep learning methods seem to outperform the statistical methods (except the MLP method), without the choice of the error measure affecting the choice of the best method. The best method from the statistical ones in terms of predictive performance for all mean error measures is the seasonal decomposition with ARIMA method, while the corresponding one for deep learning methods is the Encoder-Decoder LSTM method, which is the superior one to all other methods that we examined. Both of the statistical preprocessing methods mentioned above are statistically significant for improving the performance of deep learning methods. Moreover, the seasonal adjustment was found to have a significant reductive effect on error for all deep learning models. Finally, the computational complexity of deep learning methods was found to be three orders of magnitude greater than that of statistical methods. We conclude by highlighting the need to automate the training procedure of deep learning methods and to further examine important parameters that we believe that highly affect the traffic forecasting problem.

## Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω θερμά την, επιβλέπουσα κα Ελένη Βλαχογιάννη, Αναπληρώτρια Καθηγήτρια της Σχολής Πολιτικών Μηχανικών, για την ανάθεση της παρούσας Διπλωματικής Εργασίας, για την υποστήριξη και την πολύτιμη καθοδήγησή της σε όλο το διάστημα της εκπόνησής της. Η διάθεση για συζήτηση, οποιουδήποτε προβληματισμού και απορίας και οι κατευθύνσεις που πήρα, συνέβαλλαν ουσιαστικά για την ολοκλήρωση της διπλωματικής μου εργασίας.

Θα ήθελα ακολούθως να ευχαριστήσω τους καθηγητές της εξεταστικής επιτροπής για τον χρόνο που διέθεσαν και για τις συμβουλές τους οι οποίες ήταν καίριες για την ολοκλήρωση της διπλωματικής εργασίας.

Ένα ευχαριστώ στα ακαδημαϊκό προσωπικό του μεταπτυχιακού προγράμματος για την ευγένεια τους, τον χρόνο που διέθεσαν και την σκληρή δουλειά που κατέβαλαν για να μεταλαμπαδεύσουν στους φοιτητές τους την απαιτούμενη γνώση για ένα καλύτερο αύριο.

Τέλος, θα ήθελα να εκφράσω θερμές ευχαριστίες στους φίλους μου για την στήριξη τους και στους γονείς μου για την αμέριστη συμπαράστασή τους καθ' όλη την διάρκεια των σπουδών μου, καθώς και στην Μυρτώ που με στηρίζει καθημερινά στον αγώνα του ακαδημαϊκού στίβου.



# Παράρτημα

<b>Λίστα Διαγραμμάτων.....</b>	<b>xi</b>
<b>Λίστα Εικόνων .....</b>	<b>xix</b>
<b>Λίστα Πινάκων.....</b>	<b>xxi</b>
<b>1 Εισαγωγή.....</b>	<b>- 1 -</b>
1.1 Γενικά.....	- 1 -
1.2 Πρόβλεψη κυκλοφορίας.....	- 3 -
1.3 Κατηγοριοποίηση μοντέλων βραχυπρόθεσμης πρόβλεψης κυκλοφορίας.....	- 4 -
1.4 Βραχυπρόθεσμη Πρόβλεψη Κυκλοφορίας .....	- 6 -
1.5 Η Επέλαση της Βαθιάς Μηχανικής Μάθησης.....	- 9 -
1.6 Αντικείμενο της διπλωματικής εργασίας .....	- 11 -
1.6.1 Συνεισφορά.....	- 11 -
1.6.2 Οργάνωση κειμένου.....	- 12 -
<b>2 Σύντομη Βιβλιογραφική Ανασκόπηση .....</b>	<b>- 13 -</b>
<b>3 Προετοιμασία και προεπεξεργασία των Δεδομένων .....</b>	<b>- 17 -</b>
3.1 Μέθοδοι Συλλογής Κυκλοφοριακών Δεδομένων.....	- 17 -
3.2 Η Πηγή του Ανεπεξέργαστου Συνόλου Δεδομένων .....	- 18 -
3.3 Δημιουργία του Συνόλου Δεδομένων .....	- 20 -
3.4 Οπτικοποίηση των δεδομένων.....	- 22 -
3.4.1 Οι Συνιστώσες μίας Χρονοσειράς – Ανάλυση Αυτοσυσχέτισης και Μερικής Αυτοσυσχέτισης .....	- 22 -
3.4.2 Αναλύοντας σε βάθος το σύνολο δεδομένων.....	- 27 -
3.5 Στασιμότητα .....	- 36 -
3.6 Προεπεξεργασία των δεδομένων .....	- 38 -
3.6.1 Τι είναι η προεπεξεργασία δεδομένων; .....	- 38 -
3.6.2 Ελλείπουσες τιμές και τρόποι αντιμετώπισης.....	- 38 -
3.6.3 Εντοπισμός ακραίων τιμών και τρόποι αντιμετώπισης.....	- 46 -
3.7 Ανάλυση σε Συνιστώσες – Εποχική Αποσύνθεση.....	- 53 -
3.8 Χαρακτηριστικά του τελικού συνόλου δεδομένων και των εποχικά προσαρμοσμένων δεδομένων .....	- 56 -
3.9 Προβλεψιμότητα του συνόλου δεδομένων .....	- 69 -

<b>4</b>	<b>Θεωρητικό Υπόβαθρο Μεθόδων Πρόβλεψης .....</b>	<b>- 71 -</b>
4.1	Μέθοδοι βραχυπρόθεσμης πρόβλεψης κυκλοφορίας .....	- 71 -
4.2	Μέθοδος Αναφοράς .....	- 72 -
4.2.1	Μέθοδος <i>Seasonal Naive</i> .....	- 72 -
4.3	Στατιστικές Μέθοδοι.....	- 73 -
4.3.1	Αυτοπαλίνδρομες διαδικασίες .....	- 73 -
4.3.2	Διαδικασίες Κινητού Μέσου.....	- 73 -
4.3.3	Διαφόριση .....	- 75 -
4.3.4	Μη-εποχικά μοντέλα <i>ARIMA</i> .....	- 76 -
4.3.5	Η μέθοδος <i>Arima</i> με Εποχική αποσύνθεση ( <i>STL+Arima</i> ) .....	- 77 -
4.3.6	Δυναμική Αρμονική Παλινδρόμηση με σφάλματα <i>ARIMA</i> .....	- 77 -
4.3.7	Εισαγωγή στις μεθόδους <i>Exponential Smoothing</i> .....	- 79 -
4.3.8	Η μέθοδος <i>ETS</i> με Εποχική αποσύνθεση ( <i>STL+ETS</i> ) .....	- 83 -
4.3.9	Μέθοδος <i>TBATS</i> .....	- 83 -
4.3.10	Μέθοδος <i>Theta</i> .....	- 86 -
4.4	Μέθοδοι Βαθιάς Μηχανικής Μάθησης.....	- 88 -
4.4.1	Η Μέθοδος <i>Multi-layer Perceptron (MLP)</i> .....	- 88 -
4.4.2	Τα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα.....	- 94 -
4.4.3	Το κελί <i>LSTM</i> .....	- 98 -
4.4.4	Το κελί <i>GRU</i> .....	- 100 -
4.4.5	Τα Συνελκτικά Νευρωνικά Δίκτυα.....	- 102 -
4.4.6	Τα Διεσταλμένα Αιτιώδη Συνελκτικά Νευρωνικά Δίκτυα.....	- 106 -
4.4.7	Η Μέθοδος <i>CNN-LSTM</i> .....	- 106 -
4.4.8	Η Μέθοδος των Συνελκτικών <i>LSTM</i> δικτύων.....	- 107 -
4.4.9	Μέθοδοι Κωδικοποιητή-Αποκωδικοποιητή .....	- 108 -
<b>5</b>	<b>Μεθοδολογική Προσέγγιση .....</b>	<b>- 111 -</b>
5.1	Περιγραφή του προβλήματος .....	- 111 -
5.2	Δημιουργία συνόλου εκπαίδευσης, αξιολόγησης και ελέγχου.....	- 112 -
5.3	Μέθοδος αξιολόγησης της προβλεπτικής ικανότητας .....	- 114 -
5.4	Μετρικές Αξιολόγησης .....	- 117 -
5.4.1	Μετρικές Σφάλματος.....	- 117 -
5.4.2	Υπολογιστική Πολυπλοκότητα.....	- 119 -



5.4.3	<i>Προσαρμογή Μοντέλου</i> .....	- 119 -
5.5	Μέθοδοι προεπεξεργασίας για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.....	- 120 -
5.6	Βελτιστοποίηση των υπερπαραμέτρων και κανονικοποίηση.....	- 122 -
5.6.1	<i>Κανονικοποίηση</i> .....	- 123 -
5.6.2	<i>Πρόωρο Σταμάτημα</i> .....	- 123 -
5.6.3	<i>Dropout</i> .....	- 124 -
5.6.4	<i>Κανονικοποίηση Παρτίδων</i> .....	- 125 -
5.6.5	<i>L2 ποινικοποίηση</i> .....	- 126 -
<b>6</b>	<b>Αποτελέσματα</b> .....	<b>- 129 -</b>
6.1	Αποτελέσματα Μεθόδου Αναφοράς.....	- 129 -
6.2	Αποτελέσματα Στατιστικών Μεθόδων .....	- 133 -
6.2.1	<i>Μέθοδος ARIMA με εποχική αποσύνθεση (STL+ARIMA)</i> .....	- 133 -
6.2.2	<i>Δυναμική Αρμονική Παλινδρόμηση με σφάλματα ARIMA</i> .....	- 138 -
6.2.3	<i>Μέθοδος ETS με εποχική αποσύνθεση (STL+ETS)</i> .....	- 143 -
6.2.4	<i>Μέθοδος TBATS</i> .....	- 148 -
6.2.5	<i>Μέθοδος Theta</i> .....	- 153 -
6.3	Αποτελέσματα Μεθόδων Βαθιάς Μηχανικής Μάθησης .....	- 157 -
6.3.1	<i>Μέθοδος Multi-layer Perceptron (MLP)</i> .....	- 157 -
6.3.2	<i>Μέθοδος Long-Short Term Memory (LSTM)</i> .....	- 165 -
6.3.3	<i>Μέθοδος Gated Recurrent Unit (GRU)</i> .....	- 172 -
6.3.4	<i>Μέθοδος Convolutional Neural Networks (CNN)</i> .....	- 179 -
6.3.5	<i>Μέθοδος Dilated Convolutional Neural Networks (Dilated CNN)</i> .....	- 186 -
6.3.6	<i>Μέθοδος CNN-LSTM</i> .....	- 193 -
6.3.7	<i>Μέθοδος Convolutional LSTM (Conv-LSTM)</i> .....	- 200 -
6.3.8	<i>Μέθοδος Encoder-Decoder LSTM</i> .....	- 207 -
6.3.9	<i>Μέθοδος Encoder-Decoder CNN-LSTM</i> .....	- 214 -
6.3.10	<i>Μέθοδος Encoder-Decoder Conv-LSTM</i> .....	- 221 -
6.4	Συγκριτική ανάλυση των αποτελεσμάτων .....	- 228 -
<b>7</b>	<b>Συμπεράσματα</b> .....	<b>- 251 -</b>
7.1	Γενικά.....	- 251 -
7.2	Μεθοδολογία και Βασικά Συμπεράσματα .....	- 251 -
7.3	Προτάσεις για περαιτέρω έρευνα .....	- 254 -



## ΛΙΣΤΑ ΔΙΑΓΡΑΜΜΑΤΩΝ

---

<b>Διάγραμμα 1</b> Το σύνολο των πραγματικών δεδομένων (ανά 5 λεπτά).....	- 22 -
<b>Διάγραμμα 2</b> Το Διάγραμμα αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 1440.....	- 25 -
<b>Διάγραμμα 3</b> Το διάγραμμα αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 50.....	- 25 -
<b>Διάγραμμα 4</b> Το Διάγραμμα μερικής αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 288. ....	- 26 -
<b>Διάγραμμα 5</b> Το Διάγραμμα μερικής αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 50. ....	- 27 -
<b>Διάγραμμα 6</b> Μέση ταχύτητα ανά ημέρα με ομαλοποίηση κινητού μέσου.....	- 27 -
<b>Διάγραμμα 7</b> Μέση ταχύτητα για καθημερινές και σαββατοκύριακο (μέσοι όροι και ομαλοποιημένες τιμές κινητού μέσου).....	- 28 -
<b>Διάγραμμα 8</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Δευτέρες του συνόλου δεδομένων. ....	- 29 -
<b>Διάγραμμα 9</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Δευτέρες του συνόλου δεδομένων. ....	- 29 -
<b>Διάγραμμα 10</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Τρίτες του συνόλου δεδομένων. ....	- 30 -
<b>Διάγραμμα 11</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Τρίτες του συνόλου δεδομένων. ....	- 30 -
<b>Διάγραμμα 12</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Τετάρτες του συνόλου δεδομένων. ....	- 31 -
<b>Διάγραμμα 13</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Τετάρτες του συνόλου δεδομένων. ....	- 31 -
<b>Διάγραμμα 14</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Πέμπτες του συνόλου δεδομένων. ....	- 32 -
<b>Διάγραμμα 15</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Πέμπτες του συνόλου δεδομένων. ....	- 32 -
<b>Διάγραμμα 16</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Παρασκευές του συνόλου δεδομένων.....	- 33 -
<b>Διάγραμμα 17</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Παρασκευές του συνόλου δεδομένων. ....	- 33 -
<b>Διάγραμμα 18</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλα τα Σάββατα του συνόλου δεδομένων. ....	- 34 -
<b>Διάγραμμα 19</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τα Σάββατα του συνόλου δεδομένων. ....	- 34 -
<b>Διάγραμμα 20</b> Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλα τις Κυριακές του συνόλου δεδομένων. ....	- 35 -
<b>Διάγραμμα 21</b> Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Κυριακές του συνόλου δεδομένων. ....	- 35 -

<b>Διάγραμμα 22</b>	Το σύνολο δεδομένων ανά ημέρα (ομαλοποιημένες τιμές κινητού μέσου)....	- 40 -
<b>Διάγραμμα 23</b>	Κατανομή ελλειπουσών τιμών στο σύνολο δεδομένων.....	- 41 -
<b>Διάγραμμα 24</b>	Ποσοστό ελλειπουσών τιμών ανά ημέρα. ....	- 41 -
<b>Διάγραμμα 25</b>	Οι συχνότερες εμφανίσεις ακολουθιών ελλειπουσών τιμών στο σύνολο δεδομένων.....	- 42 -
<b>Διάγραμμα 26</b>	Το συμπληρωμένο σύνολο δεδομένων. ....	- 43 -
<b>Διάγραμμα 27</b>	Το συμπληρωμένο και το αρχικό σύνολο δεδομένων. ....	- 43 -
<b>Διάγραμμα 28</b>	Το συμπληρωμένο σύνολο δεδομένων για την 2 <sup>η</sup> εβδομάδα. ....	- 44 -
<b>Διάγραμμα 29</b>	Το συμπληρωμένο σύνολο δεδομένων για την 8 <sup>η</sup> ημέρα.....	- 44 -
<b>Διάγραμμα 30</b>	Το Διάγραμμα αυτοσυσχέτισης για το συμπληρωμένο σύνολο δεδομένων. ..	- 45 -
<b>Διάγραμμα 31</b>	Το Διάγραμμα μερικής αυτοσυσχέτισης για το συμπληρωμένο σύνολο δεδομένων.....	- 45 -
<b>Διάγραμμα 32</b>	Εντοπισμός των ακραίων παρατηρήσεων του συνόλου δεδομένων με την μέθοδο «gesd».....	- 51 -
<b>Διάγραμμα 33</b>	Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με τις ακραίες παρατηρήσεις να έχουν αντικατασταθεί.....	- 52 -
<b>Διάγραμμα 34</b>	Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με αντικατεστημένες τις ακραίες παρατηρήσεις για την 2 <sup>η</sup> εβδομάδα. ....	- 52 -
<b>Διάγραμμα 35</b>	Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με αντικατεστημένες τις ακραίες παρατηρήσεις για την 8 <sup>η</sup> ημέρα. ....	- 53 -
<b>Διάγραμμα 36</b>	Τελικό σύνολο δεδομένων και ομαλοποιημένα δεδομένα με κινητό μέσο (τιμή υστέρησης = 288).....	- 55 -
<b>Διάγραμμα 37</b>	Ανάλυση συνιστωσών (τάσης, εποχική, υπόλοιπο) του τελικού συνόλου δεδομένων με χρήση της μεθόδου "LOESS". ....	- 55 -
<b>Διάγραμμα 38</b>	Το τελικό σύνολο δεδομένων, τα εποχιακά προσαρμοσμένα δεδομένα και η συνιστώσα τάσης για την 2 <sup>η</sup> εβδομάδα.....	- 56 -
<b>Διάγραμμα 39</b>	Το τελικό σύνολο δεδομένων μετασχηματισμένο κατά Box-Cox.....	- 60 -
<b>Διάγραμμα 40</b>	Τα εποχιακά προσαρμοσμένα δεδομένα μετασχηματισμένα κατά Box-Cox. ....	- 61 -
<b>Διάγραμμα 41</b>	Το ιστόγραμμα του τελικού συνόλου δεδομένων.....	- 61 -
<b>Διάγραμμα 42</b>	Το ιστόγραμμα του τελικού συνόλου δεδομένων μετασχηματισμένο κατά Box-Cox. ....	- 62 -
<b>Διάγραμμα 43</b>	Το ιστόγραμμα των εποχιακά προσαρμοσμένων δεδομένων.....	- 62 -
<b>Διάγραμμα 44</b>	Το ιστόγραμμα των εποχιακά προσαρμοσμένων δεδομένων μετασχηματισμένο κατά Box-Cox. ....	- 63 -
<b>Διάγραμμα 45</b>	Το Διάγραμμα αυτοσυσχέτισης για τα εποχιακά προσαρμοσμένα δεδομένα. ....	- 64 -
<b>Διάγραμμα 46</b>	Το Διάγραμμα μερικής αυτοσυσχέτισης για τα εποχιακά προσαρμοσμένα δεδομένα. ....	- 64 -
<b>Διάγραμμα 47</b>	Το Διάγραμμα αυτοσυσχέτισης για τα μετασχηματισμένα εποχιακά προσαρμοσμένα δεδομένα.....	- 65 -
<b>Διάγραμμα 48</b>	Το Διάγραμμα μερικής αυτοσυσχέτισης για τα μετασχηματισμένα εποχιακά προσαρμοσμένα δεδομένα.....	- 65 -
<b>Διάγραμμα 49</b>	Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το τελικό σύνολο δεδομένων.....	- 66 -

<b>Διάγραμμα 50</b> Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το τελικό σύνολο δεδομένων. ....	- 67 -
<b>Διάγραμμα 51</b> Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων δεδομένων. ....	- 67 -
<b>Διάγραμμα 52</b> Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων δεδομένων. ....	- 68 -
<b>Διάγραμμα 53</b> Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων και μετασχηματισμένων δεδομένων.....	- 68 -
<b>Διάγραμμα 54</b> Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων και μετασχηματισμένων δεδομένων.....	- 69 -
<b>Διάγραμμα 55</b> Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.....	- 131 -
<b>Διάγραμμα 56</b> Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.....	- 131 -
<b>Διάγραμμα 57</b> Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 132 -
<b>Διάγραμμα 58</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Seasonal Naive.....	- 132 -
<b>Διάγραμμα 59</b> Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.....	- 135 -
<b>Διάγραμμα 60</b> Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.....	- 136 -
<b>Διάγραμμα 61</b> Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 136 -
<b>Διάγραμμα 62</b> Η εποχική συνιστώσα που υπολογίστηκε για την μέθοδο STL+ARIMA με μέθοδο LOESS.....	- 137 -
<b>Διάγραμμα 63</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου STL+ARIMA.....	- 137 -
<b>Διάγραμμα 64</b> Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.....	- 141 -
<b>Διάγραμμα 65</b> Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.....	- 142 -
<b>Διάγραμμα 66</b> Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 142 -
<b>Διάγραμμα 67</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA. ....	- 143 -
<b>Διάγραμμα 68</b> Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.....	- 146 -
<b>Διάγραμμα 69</b> Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.....	- 146 -
<b>Διάγραμμα 70</b> Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 147 -
<b>Διάγραμμα 71</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου STL+ETS.....	- 147 -
<b>Διάγραμμα 72</b> Προβλέψεις της μεθόδου Tbats για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.....	- 151 -

<b>Διάγραμμα 73</b> Προβλέψεις της μεθόδου Tbats για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 151 -
<b>Διάγραμμα 74</b> Προβλέψεις της μεθόδου Tbats για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 152 -
<b>Διάγραμμα 75</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Tbats. ....	- 152 -
<b>Διάγραμμα 76</b> Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 155 -
<b>Διάγραμμα 77</b> Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 155 -
<b>Διάγραμμα 78</b> Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 156 -
<b>Διάγραμμα 79</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Theta. ....	- 156 -
<b>Διάγραμμα 80</b> Ευαισθησία της μεθόδου Multi-layer Perceptron (MLP). ....	- 158 -
<b>Διάγραμμα 81</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Multi-layer Perceptron (MLP). ...	- 159 -
<b>Διάγραμμα 82</b> Ευαισθησία της μεθόδου Multi-layer Perceptron (MLP) ανά υπερπαραμέτρο. ....	- 159 -
<b>Διάγραμμα 83</b> Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 163 -
<b>Διάγραμμα 84</b> Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 163 -
<b>Διάγραμμα 85</b> Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 164 -
<b>Διάγραμμα 86</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Multi-layer Perceptron (MLP). ..	- 164 -
<b>Διάγραμμα 87</b> Ευαισθησία της μεθόδου LSTM. ....	- 165 -
<b>Διάγραμμα 88</b> Χρόνος ανά παραμετροποίηση για την μέθοδο LSTM. ....	- 166 -
<b>Διάγραμμα 89</b> Ευαισθησία της μεθόδου LSTM ανά υπερπαραμέτρο. ....	- 166 -
<b>Διάγραμμα 90</b> Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 170 -
<b>Διάγραμμα 91</b> Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 170 -
<b>Διάγραμμα 92</b> Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 171 -
<b>Διάγραμμα 93</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου LSTM. ....	- 171 -
<b>Διάγραμμα 94</b> Ευαισθησία της μεθόδου GRU. ....	- 172 -
<b>Διάγραμμα 95</b> Χρόνος ανά παραμετροποίηση για την μέθοδο GRU. ....	- 173 -
<b>Διάγραμμα 96</b> Ευαισθησία της μεθόδου GRU ανά υπερπαραμέτρο. ....	- 173 -
<b>Διάγραμμα 97</b> Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 177 -
<b>Διάγραμμα 98</b> Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 177 -

<b>Διάγραμμα 99</b> Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 178 -
<b>Διάγραμμα 100</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου GRU. ....	- 178 -
<b>Διάγραμμα 101</b> Ευαισθησία της μεθόδου CNN. ....	- 179 -
<b>Διάγραμμα 102</b> Χρόνος ανά παραμετροποίηση για την μέθοδο CNN. ....	- 180 -
<b>Διάγραμμα 103</b> Ευαισθησία της μεθόδου CNN ανά υπερπαράμετρο. ....	- 180 -
<b>Διάγραμμα 104</b> Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 184 -
<b>Διάγραμμα 105</b> Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 184 -
<b>Διάγραμμα 106</b> Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 185 -
<b>Διάγραμμα 107</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου CNN. ....	- 185 -
<b>Διάγραμμα 108</b> Ευαισθησία της μεθόδου Dilated CNN. ....	- 186 -
<b>Διάγραμμα 109</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Dilated CNN. ....	- 187 -
<b>Διάγραμμα 110</b> Ευαισθησία της μεθόδου Dilated CNN ανά υπερπαράμετρο. ....	- 187 -
<b>Διάγραμμα 111</b> Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 191 -
<b>Διάγραμμα 112</b> Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 191 -
<b>Διάγραμμα 113</b> Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 192 -
<b>Διάγραμμα 114</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Dilated CNN. ....	- 192 -
<b>Διάγραμμα 115</b> Ευαισθησία της μεθόδου CNN-LSTM. ....	- 193 -
<b>Διάγραμμα 116</b> Χρόνος ανά παραμετροποίηση για την μέθοδο CNN-LSTM. ....	- 194 -
<b>Διάγραμμα 117</b> Ευαισθησία της μεθόδου CNN-LSTM ανά υπερπαράμετρο. ....	- 194 -
<b>Διάγραμμα 118</b> Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 198 -
<b>Διάγραμμα 119</b> Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 198 -
<b>Διάγραμμα 120</b> Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 199 -
<b>Διάγραμμα 121</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου CNN-LSTM. ....	- 199 -
<b>Διάγραμμα 122</b> Ευαισθησία της μεθόδου Conv-LSTM. ....	- 200 -
<b>Διάγραμμα 123</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Conv-LSTM. ....	- 201 -
<b>Διάγραμμα 124</b> Ευαισθησία της μεθόδου Conv-LSTM ανά υπερπαράμετρο. ....	- 201 -
<b>Διάγραμμα 125</b> Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 205 -
<b>Διάγραμμα 126</b> Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 205 -
<b>Διάγραμμα 127</b> Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 206 -
<b>Διάγραμμα 128</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Conv-LSTM. ....	- 206 -

<b>Διάγραμμα 129</b> Ευαισθησία της μεθόδου Encoder Decoder LSTM. ....	- 207 -
<b>Διάγραμμα 130</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder LSTM..	- 208 -
<b>Διάγραμμα 131</b> Ευαισθησία της μεθόδου Encoder-Decoder LSTM ανά υπερπαραμέτρο...	- 208 -
<b>Διάγραμμα 132</b> Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 212 -
<b>Διάγραμμα 133</b> Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 212 -
<b>Διάγραμμα 134</b> Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12. ....	- 213 -
<b>Διάγραμμα 135</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder LSTM. -	213
<b>Διάγραμμα 136</b> Ευαισθησία της μεθόδου Encoder-Decoder CNN-LSTM. ....	- 214 -
<b>Διάγραμμα 137</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder CNN-LSTM. -	215 -
<b>Διάγραμμα 138</b> Ευαισθησία της μεθόδου Encoder-Decoder CNN-LSTM ανά υπερπαραμέτρο. . -	215 -
<b>Διάγραμμα 139</b> Προβλέψεις της μεθόδου Encoder-Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 219 -
<b>Διάγραμμα 140</b> Προβλέψεις της μεθόδου Encoder Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 219 -
<b>Διάγραμμα 141</b> Προβλέψεις της μεθόδου Encoder Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 220 -
<b>Διάγραμμα 142</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder CNN-LSTM. ....	- 220 -
<b>Διάγραμμα 143</b> Ευαισθησία της μεθόδου Encoder Decoder Conv-LSTM.....	- 221 -
<b>Διάγραμμα 144</b> Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder Conv-LSTM. -	222 -
<b>Διάγραμμα 145</b> Ευαισθησία της μεθόδου Encoder Decoder Conv-LSTM ανά υπερπαραμέτρο... -	222 -
<b>Διάγραμμα 146</b> Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1. ....	- 226 -
<b>Διάγραμμα 147</b> Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6. ....	- 226 -
<b>Διάγραμμα 148</b> Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.....	- 227 -
<b>Διάγραμμα 149</b> Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder Conv-LSTM. ....	- 227 -
<b>Διάγραμμα 150</b> Μετρική MAE ανά ορίζοντα για όλες τις μεθόδους.....	- 228 -
<b>Διάγραμμα 151</b> Μετρική MAE ανά ορίζοντα για τις στατιστικές μεθόδους. ....	- 229 -
<b>Διάγραμμα 152</b> Συνοπτικά μέτρα για την μετρική σφάλματος MAE για τις στατιστικές μεθόδους.....	- 229 -
<b>Διάγραμμα 153</b> Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.....	- 230 -



<b>Διάγραμμα 154</b> Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 πρώτες τιμές του ορίζοντα πρόβλεψης. ....	- 231 -
<b>Διάγραμμα 155</b> Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 μεσαίες τιμές του ορίζοντα πρόβλεψης. ....	- 231 -
<b>Διάγραμμα 156</b> Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 τελευταίες τιμές του ορίζοντα πρόβλεψης. ....	- 232 -
<b>Διάγραμμα 157</b> Συνοπτικά μέτρα για την μετρική σφάλματος MAE για τις μεθόδους Βαθιάς Μηχανικής Μάθησης. ....	- 232 -
<b>Διάγραμμα 158</b> Μετρική MAE για βραχυπρόθεσμο μέσο ορίζοντα ανά μέθοδο. ....	- 233 -
<b>Διάγραμμα 159</b> Μετρική MAE για μεσοπρόθεσμο μέσο ορίζοντα ανά μέθοδο. ....	- 234 -
<b>Διάγραμμα 160</b> Μετρική MAE για μακροπρόθεσμο μέσο ορίζοντα ανά μέθοδο. ....	- 234 -
<b>Διάγραμμα 161</b> Μέσο MAE ανά μέθοδο. ....	- 235 -
<b>Διάγραμμα 162</b> Σύγκριση της μεθόδου STL+ARIMA με τις μεθόδους Βαθιάς Μηχανικής Μάθησης για τα συνοπτικά μέτρα της μετρικής σφάλματος MAE. ....	- 235 -
<b>Διάγραμμα 163</b> Σύγκριση της μεθόδου STL+ARIMA με την μέθοδο Encoder-Decoder LSTM ανά ορίζοντα και για τα συνοπτικά μέτρα της μετρικής σφάλματος MAE. ....	- 237 -
<b>Διάγραμμα 164</b> Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για βραχυπρόθεσμο ορίζοντα πρόβλεψης. ....	- 237 -
<b>Διάγραμμα 165</b> Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης. ....	- 238 -
<b>Διάγραμμα 166</b> Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης. ....	- 238 -
<b>Διάγραμμα 167</b> Το μέσο σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος. ..	- 239 -
<b>Διάγραμμα 168</b> Το σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος για βραχυπρόθεσμο ορίζοντα πρόβλεψης. ....	- 240 -
<b>Διάγραμμα 169</b> Το σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης. ....	- 240 -
<b>Διάγραμμα 170</b> Το σφάλμα για τις Βαθιάς Μηχανικής Μάθησης μεθόδους ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης. ....	- 241 -
<b>Διάγραμμα 171</b> Το μέσο σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος. ....	- 242 -
<b>Διάγραμμα 172</b> Ευαισθησία των μεθόδων Βαθιάς Μηχανικής Μάθησης στο χώρο παραμέτρων που εξετάστηκε κάθε μία μέθοδος. ....	- 243 -
<b>Διάγραμμα 173</b> Επιρροή της αποεποχικοποίησης των δεδομένων στο μέσο RMSE. ....	- 244 -
<b>Διάγραμμα 174</b> Προσαρμογή κάθε μεθόδου σε σχέση με το μέσο MAE. ....	- 244 -
<b>Διάγραμμα 175</b> Γραμμική παλινδρόμηση για την προσαρμογή μοντέλου ως προς το μέσο MAE. ....	- 245 -
<b>Διάγραμμα 176</b> Υπολογιστική πολυπλοκότητα όλων των μεθόδων σε συνάρτηση με το μέσο MAE. ....	- 246 -
<b>Διάγραμμα 177</b> Υπολογιστική πολυπλοκότητα ανά μοντέλο σε μορφή barplot. ....	- 246 -
<b>Διάγραμμα 178</b> Χρόνοι βελτιστοποίησης υπερπαραμέτρων, εκπαίδευσης και πρόβλεψης για όλες τις μεθόδους. ....	- 247 -

<b>Διάγραμμα 179</b> Χρόνοι βελτιστοποίησης υπερπαραμέτρων, εκπαίδευσης και πρόβλεψης για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.....	- 248 -
<b>Διάγραμμα 180</b> Προβλέψεις της μεθόδου Encoder-Decoder LSTM για οριζόντα πρόβλεψης ίσο με 1.....	- 249 -
<b>Διάγραμμα 181</b> Προβλέψεις της μεθόδου Encoder-Decoder LSTM για οριζόντα πρόβλεψης ίσο με 12.....	- 249 -

## ΛΙΣΤΑ ΕΙΚΟΝΩΝ

---

<b>Εικόνα 1</b> Αριθμός εργασιών ανά κατηγορία μοντέλων (μέχρι το 2018) [9] .....	- 5 -
<b>Εικόνα 2</b> Ο 2 <sup>ος</sup> δακτύλιος της πόλης Xi'an (Κίνα) και το χωρικό εύρος των μετρήσεων (Πηγή: <a href="https://gaia.didichuxing.com">https://gaia.didichuxing.com</a> ) .....	- 18 -
<b>Εικόνα 3</b> Ο 2 <sup>ος</sup> δακτύλιος της πόλης Xi'an (Κίνα) (Πηγή: <a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a> ) -	19 -
<b>Εικόνα 4</b> Δείγμα από το σύνολο των ακατέργαστων δεδομένων (Πηγή: <a href="https://gaia.didichuxing.com">https://gaia.didichuxing.com</a> ) .....	- 19 -
<b>Εικόνα 5</b> Η αρχιτεκτονική του MLP με δύο εισόδους, ένα κρυφό επίπεδο, και τρεις νευρώνες εξόδου [88]. .....	- 90 -
<b>Εικόνα 6</b> Διορθωμένη γραμμική μονάδα ( <i>ReLU</i> ) .....	- 93 -
<b>Εικόνα 7</b> Ένας ανατροφοδοτούμενος νευρώνας (αριστερά) "ξεδιπλωμένος" στο χρόνο (δεξιά). -	95 -
<b>Εικόνα 8</b> Ένα επίπεδο ανατροφοδοτούμενων νευρώνων (αριστερά) ξεδιπλωμένο στον χρόνο (δεξιά) [88]. .....	- 95 -
<b>Εικόνα 9</b> Ακολουθία-σε-ακολουθία (αριστερά) και ακολουθία-σε-διάνυσμα (δεξιά) δίκτυο [88].-	97 -
<b>Εικόνα 10</b> Δίκτυο RNN με βάθος (αριστερά) ξεδιπλωμένο στον χρόνο (αριστερά) [88]. .....	- 97 -
<b>Εικόνα 11</b> Το κελί LSTM [88]. .....	- 98 -
<b>Εικόνα 12</b> Το κελί GRU [88]. .....	- 101 -
<b>Εικόνα 13</b> Τα 3 στάδια ενός τυπικού συνελκτικού επιπέδου. .....	- 104 -
<b>Εικόνα 14</b> Παράδειγμα μονοδιάστατης συνέλιξης για μονομεταβλητές χρονοσειρές [92]. -	104 -
<b>Εικόνα 15</b> Look-ahead μεροληψία στην κλασσική συνέλιξη [94]. .....	- 105 -
<b>Εικόνα 16</b> Αιτιώδης συνέλιξη [94]. .....	- 105 -
<b>Εικόνα 17</b> Αιτιώδης διεσταλμένη συνέλιξη [94]. .....	- 106 -
<b>Εικόνα 18</b> Δίκτυο CNN-LSTM. ....	- 107 -
<b>Εικόνα 19</b> Το κελί Conv-LSTM. ....	- 108 -
<b>Εικόνα 20</b> Αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή [88]. .....	- 109 -
<b>Εικόνα 21</b> Αξιολόγηση με κυλιόμενη αρχή σταθερού ορίζοντα χωρίς επανεκπαίδευση. ....	- 116 -
<b>Εικόνα 22</b> Υπολογισμός τάσης των δεδομένων εκπαίδευσης με χρήση κυβικών σπλινών. -	121 -
<b>Εικόνα 23</b> Πρόβλεψη τάσης για ορίζοντα 12 βημάτων στο σύνολο ελέγχου. ....	- 121 -
<b>Εικόνα 24</b> Η μέθοδος του πρόωρου σταματήματος. ....	- 124 -
<b>Εικόνα 25</b> Νευρωνικό δίκτυο πριν (αριστερά) και μετά την εφαρμογή του dropout (δεξιά).. -	124 -
<b>Εικόνα 26</b> Το επίπεδο batch normalization σχηματικά. ....	- 125 -
<b>Εικόνα 27</b> Η μέθοδος της L2 ποινικοποίησης. ....	- 126 -



## ΛΙΣΤΑ ΠΙΝΑΚΩΝ

---

<b>Πίνακας 1</b> Βασικότερα κριτήρια για το πρόβλημα πρόβλεψης κυκλοφορίας σύμφωνα με το [11]. .....	- 8 -
<b>Πίνακας 2</b> Παραμετροποιήσεις και δομή κανονικοποίησης για τις μεθόδους βαθιάς μηχανικής μάθησης. ....	- 127 -
<b>Πίνακας 3</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Seasonal Naive. ....	- 130 -
<b>Πίνακας 4</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Seasonal Naive. ....	- 130 -
<b>Πίνακας 5</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Seasonal Naive. ....	- 130 -
<b>Πίνακας 6</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο STL+ARIMA. ....	- 133 -
<b>Πίνακας 7</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο STL+ARIMA. ....	- 134 -
<b>Πίνακας 8</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο STL+ARIMA ως προς την μέθοδο αναφοράς. ....	- 134 -
<b>Πίνακας 9</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο STL+ARIMA ως προς την μέθοδο αναφοράς. ....	- 135 -
<b>Πίνακας 10</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο STL+ARIMA. ....	- 135 -
<b>Πίνακας 11</b> Συντελεστές των όρων fourier της δυναμικής παλινδρόμησης. ....	- 138 -
<b>Πίνακας 12</b> Συντελεστές της διαδικασίας ARIMA που ακολουθούν τα σφάλματα της δυναμικής παλινδρόμησης. ....	- 138 -
<b>Πίνακας 13</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA. ....	- 139 -
<b>Πίνακας 14</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA. ....	- 139 -
<b>Πίνακας 15</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA. ....	- 140 -
<b>Πίνακας 16</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA ως προς την μέθοδο αναφοράς. ....	- 140 -
<b>Πίνακας 17</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA ως προς την μέθοδο αναφοράς. ....	- 140 -
<b>Πίνακας 18</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο STL+ETS. ....	- 144 -
<b>Πίνακας 19</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο STL+ETS. ....	- 144 -
<b>Πίνακας 20</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο STL+ETS. ....	- 145 -
<b>Πίνακας 21</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο STL+ETS ως προς την μέθοδο αναφοράς. ....	- 145 -
<b>Πίνακας 22</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο STL+ETS ως προς την μέθοδο αναφοράς. ....	- 145 -
<b>Πίνακας 23</b> Οι συνιστώσες τον ARMA σφαλμάτων της TBATS μεθόδου. ....	- 148 -
<b>Πίνακας 24</b> Αρχικές τιμές των παραμέτρων της εποχικότητας για την μέθοδο TBATS. ....	- 149 -

<b>Πίνακας 25</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Tbats.....	- 149 -
<b>Πίνακας 26</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Tbats.....	- 149 -
<b>Πίνακας 27</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Tbats.....	- 150 -
<b>Πίνακας 28</b> Ποσοστιαία μείωση για όλες τις μετρικές ράσφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Tbats ως προς την μέθοδο αναφοράς.....	- 150 -
<b>Πίνακας 29</b> Συνοπτικά μέτρα σφάλματος για όλους τους ορίζοντες για την μέθοδο Tbats. ....	- 150 -
<b>Πίνακας 30</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Theta.....	- 153 -
<b>Πίνακας 31</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Theta. ....	- 153 -
<b>Πίνακας 32</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Theta.....	- 153 -
<b>Πίνακας 33</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Theta ως προς την μέθοδο αναφοράς. ....	- 154 -
<b>Πίνακας 34</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Theta ως προς την μέθοδο αναφοράς. ....	- 154 -
<b>Πίνακας 35</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Multi-layer Perceptron (MLP).-	160 -
<b>Πίνακας 36</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Multi-layer Perceptron (MLP)..-	161 -
<b>Πίνακας 37</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Multi-layer Perceptron (MLP)...	- 161 -
<b>Πίνακας 38</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Multi-layer Perceptron (MLP).....	- 161 -
<b>Πίνακας 39</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Multi-layer Perceptron (MLP) ως προς την μέθοδο αναφοράς.....	- 162 -
<b>Πίνακας 40</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Multi-layer Perceptron (MLP) ως προς την μέθοδο αναφοράς. ....	- 162 -
<b>Πίνακας 41</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο LSTM. ....	- 167 -
<b>Πίνακας 42</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο LSTM. ....	- 168 -
<b>Πίνακας 43</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο LSTM.....	- 168 -
<b>Πίνακας 44</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο LSTM. ....	- 168 -
<b>Πίνακας 45</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο LSTM ως προς την μέθοδο αναφοράς.....	- 169 -
<b>Πίνακας 46</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο LSTM ως προς την μέθοδο αναφοράς. ....	- 169 -
<b>Πίνακας 47</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο GRU. ....	- 174 -
<b>Πίνακας 48</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο GRU. ....	- 175 -
<b>Πίνακας 49</b> Συνοπτικά μέτρα σφάλματος των συνοπτικών μέτρων σφάλματος για την μέθοδο GRU. ....	- 175 -
<b>Πίνακας 50</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο GRU. ....	- 175 -
<b>Πίνακας 51</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο GRU ως προς την μέθοδο αναφοράς. ....	- 176 -

<b>Πίνακας 52</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο GRU ως προς την μέθοδο αναφοράς. ....	- 176 -
<b>Πίνακας 53</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο CNN. ....	- 181 -
<b>Πίνακας 54</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο CNN. ....	- 182 -
<b>Πίνακας 55</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο CNN. ....	- 182 -
<b>Πίνακας 56</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο CNN. ....	- 182 -
<b>Πίνακας 57</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο CNN ως προς την μέθοδο αναφοράς. ....	- 183 -
<b>Πίνακας 58</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο CNN ως προς την μέθοδο αναφοράς. ....	- 183 -
<b>Πίνακας 59</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Dilated CNN. ....	- 188 -
<b>Πίνακας 60</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Dilated CNN. ....	- 189 -
<b>Πίνακας 61</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Dilated CNN. ....	- 189 -
<b>Πίνακας 62</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Dilated CNN. ....	- 189 -
<b>Πίνακας 63</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Dilated CNN ως προς την μέθοδο αναφοράς. ....	- 190 -
<b>Πίνακας 64</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Dilated CNN ως προς την μέθοδο αναφοράς. ....	- 190 -
<b>Πίνακας 65</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο CNN-LSTM. ....	- 195 -
<b>Πίνακας 66</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο CNN-LSTM. ....	- 196 -
<b>Πίνακας 67</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο CNN-LSTM. ....	- 196 -
<b>Πίνακας 68</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο CNN-LSTM. ....	- 196 -
<b>Πίνακας 69</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο CNN-LSTM ως προς την μέθοδο αναφοράς. ....	- 197 -
<b>Πίνακας 70</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο CNN-LSTM ως προς την μέθοδο αναφοράς. ....	- 197 -
<b>Πίνακας 71</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Conv-LSTM. ....	- 202 -
<b>Πίνακας 72</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Conv-LSTM. ....	- 203 -
<b>Πίνακας 73</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Conv-LSTM. ....	- 203 -
<b>Πίνακας 74</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Conv-LSTM. ....	- 203 -
<b>Πίνακας 75</b> Ποσοστιαία μείωση (ακρίβεια 1ου δεκαδικού) για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Conv-LSTM ως προς την μέθοδο αναφοράς. ....	- 204 -
<b>Πίνακας 76</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Conv-LSTM ως προς την μέθοδο αναφοράς. ....	- 204 -
<b>Πίνακας 77</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder-Decoder LSTM. ....	- 209 -
<b>Πίνακας 78</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder LSTM. ....	- 210 -
<b>Πίνακας 79</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder LSTM. ....	- 210 -

<b>Πίνακας 80</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder LSTM. ....	- 210 -
<b>Πίνακας 81</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder-Decoder LSTM ως προς την μέθοδο αναφοράς.....	- 211 -
<b>Πίνακας 82</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder-Decoder LSTM ως προς την μέθοδο αναφοράς. ....	- 211 -
<b>Πίνακας 83</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder-Decoder CNN-LSTM. ....	- 216 -
<b>Πίνακας 84</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder CNN-LSTM. -	217 -
<b>Πίνακας 85</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder CNN-LSTM. -	217 -
<b>Πίνακας 86</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder CNN-LSTM.....	- 217 -
<b>Πίνακας 87</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder-Decoder CNN-LSTM ως προς την μέθοδο αναφοράς. ....	- 218 -
<b>Πίνακας 88</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder-Decoder CNN-LSTM ως προς την μέθοδο αναφοράς.....	- 218 -
<b>Πίνακας 89</b> Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder Decoder Conv-LSTM. ....	- 223 -
<b>Πίνακας 90</b> Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder Conv-LSTM. -	224 -
<b>Πίνακας 91</b> Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder Conv-LSTM. -	224 -
<b>Πίνακας 92</b> Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder Conv-LSTM. ....	- 224 -
<b>Πίνακας 93</b> Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder Decoder Conv-LSTM ως προς την μέθοδο αναφοράς. ....	- 225 -
<b>Πίνακας 94</b> Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder Decoder Conv-LSTM ως προς την μέθοδο βάσης. ....	- 225 -



# 1 ΕΙΣΑΓΩΓΗ

---

## 1.1 ΓΕΝΙΚΑ

Ο όρος «έξυπνη πόλη» (*smart city*) χρησιμοποιείται πλέον κατά κόρον και γίνεται όλο και πιο διαδεδομένος, καθώς ο αριθμός των πόλεων και ο πληθυσμός σε αυτές αυξάνεται ραγδαία λόγω της έντονης αστικοποίησης η οποία παρατηρείται ήδη από τα μέσα του περασμένου αιώνα και της φαινόμενης αύξησης του παγκόσμιου πληθυσμού [1]. Ο όρος «έξυπνη πόλη» αναφέρεται συνήθως στην χρήση πληροφοριακών και συστημάτων επικοινωνίας για την κατανόηση, ανάλυση και συλλογή σημαντικών δεδομένων με σκοπό τον σχεδιασμό αποτελεσματικών ευφυών συστημάτων που σκοπό έχουν την βέλτιστη οργάνωση μίας πόλης και σε επίπεδο μεταφορών [2]. Με την ανάπτυξη τεχνολογιών άμεσα συσχετισμένων με τις «έξυπνες πόλεις» (π.χ. η εξάπλωση των IoT συσκευών), όλο και περισσότερα «φαινόμενα» γίνονται μετρήσιμα, μέσω των δεδομένων που γίνονται διαθέσιμα μέσω αυτών των τεχνολογιών. Αποτέλεσμα των παραπάνω, είναι η ολοένα αυξανόμενη διαθεσιμότητα και ποιότητα των δεδομένων που περιγράφουν ικανοποιητικά πλέον το πολύπλοκο αυτό οικοσύστημα που ονομάζουμε «πόλη».

Ανάμεσα στους αξιοσημείωτους στόχους των έξυπνων πόλεων, είναι η κατασκευή έξυπνων μεταφορικών συστημάτων και έξυπνων συστημάτων αστικής διαχείρισης, τα οποία μπορούν να επηρεάσουν σημαντικά την ζωή των κατοίκων στις πόλεις μελλοντικά. Τα Προχωρημένα Συστήματα Διαχείρισης Κυκλοφορίας (Advanced Traffic Management Systems – ATMSs), τα προχωρημένα συστήματα πληροφόρησης ταξιδιωτών (Advanced Traveller Information Systems – ATISs) και τα Ευφυή Συστήματα Μεταφορών (Intelligent Transportation Systems – ITSs) ενσωματώνουν συστήματα πληροφοριών, επικοινωνίας και πολλά άλλα συστήματα, τα οποία χρησιμοποιούνται στον κλάδο των μεταφορών με σκοπό την δημιουργία ενός ενοποιημένου συστήματος το οποίο αποτελείται από ανθρώπους, δρόμους (και γενικά υποδομές) και οχήματα. Τα συστήματα αυτά απαρτίζονται από δαιδαλώδες, πλήρως λειτουργικό, πραγματικού χρόνου, ακριβές, και αποτελεσματικό πλαίσιο διαχείρισης μεταφορών [3].

Για τα ATMSs, ATISs και ITSs είναι ύψιστης σημασίας η πρόβλεψη των επόμενων πιθανών καταστάσεων της κυκλοφορίας με μεγάλη ακρίβεια, καθώς τέτοιου είδους πληροφορία μπορεί να συμβάλλει ουσιαστικά στην αποτροπή μη αποδεκτών καταστάσεων όπως κυκλοφοριακή

συμφόρηση αλλά και πολλών άλλων μη επιθυμητών φαινομένων στο κυκλοφοριακό δίκτυο μιας πόλης. Ο στόχος της πρόβλεψης κυκλοφοριακών καταστάσεων είναι η πρόβλεψη μελλοντικών κυκλοφοριακών μεταβλητών όπως φόρτος, πυκνότητα και ταχύτητα του δικτύου, η οποία θα βασίζεται σε ιστορικά ή παρατηρούμενα κυκλοφοριακά δεδομένα. Λόγω της αύξησης του παγκόσμιου πληθυσμού και της έντονης αστικοποίησης, όπως προαναφέρθηκε, που παρατηρείται ανά τον κόσμο και συνεπώς της χρήσης όλων των μέσων μεταφοράς και κυρίως του αυτοκινήτου, δημιουργούνται θέματα που σχετίζονται με τις κυκλοφοριακές συνθήκες στα μεγάλα αστικά κέντρα και αφορούν πλέον όλες τις μεγάλες πόλεις. Η κυκλοφοριακή συμφόρηση που δημιουργείται, προκαλεί προβλήματα σε κοινωνικό, οικονομικό και περιβαλλοντικό επίπεδο και οι αρμόδιοι οργανισμοί προσπαθούν για δεκαετίες να μετριάσουν το πρόβλημα της κυκλοφοριακής συμφόρησης. Οι προσπάθειες αυτές κινούνται σε 3 άξονες:

1. Βελτίωση των ήδη υπάρχουσών υποδομών και περαιτέρω επέκταση αυτών.
2. Μετακίνηση των χρηστών του δικτύου από το αυτοκίνητο σε Μέσα Μαζικής Μεταφοράς.
3. Μείωση της κυκλοφοριακής συμφόρησης μέσω διαχείρισης της κυκλοφορίας.

Η πρώτη κατεύθυνση περιορίζεται σημαντικά από οικονομικούς και τοπολογικούς παράγοντες καθώς η κατασκευή και τροποποίηση του κυκλοφοριακού δικτύου θα έχει υψηλά κόστη ενώ υπάρχουν και φυσικοί περιορισμοί όπως έλλειψη χώρου. Επιπροσθέτως, αλλαγές στην χωρητικότητα του δικτύου μπορούν να επιφέρουν κάποιες μη αναμενόμενες αλλαγές οδηγώντας σε μη επιθυμητά αποτελέσματα διαφορετικά από εκείνα που είχαν προβλεφθεί [4]. Η δεύτερη είναι απλά θέμα κοινωνικής πολιτικής, ενώ η τελευταία βελτιώνεται συνεχώς ανά τα χρόνια κυρίως λόγω της ανάπτυξης νέων μαθηματικών μοντέλων, υψηλών επιδόσεων υπολογιστικές μονάδες αλλά και της ολοένα αυξανόμενης διαθεσιμότητας κυκλοφοριακών (και όχι μόνο) δεδομένων από συσκευές που είναι ενσωματωμένες είτε στις υποδομές μίας πόλης είτε στα οχήματα, αλλά και της συνεχούς βελτίωσης τεχνολογιών που σχετίζονται με την συλλογή αυτών των δεδομένων. Έτσι, δεδομένα από αισθητήρες στους δρόμους είναι πλέον διαθέσιμα με εξαιρετικά λεπτομερή ανάλυση, τα οποία δημοσιεύονται σε τακτικά ενημερωμένες δημόσιες βιβλιοθήκες (ανοικτά δεδομένα) και υπάρχει ακόμα και πρόσβαση σε δεδομένα με τη μορφή δεδομένων πλωτών αυτοκινήτων (floating car data), κινητικότητας πεζών, κυκλοφοριακών μετρήσεων για ποδήλατα κ.α. Γίνεται επομένως κατανοητό ότι η ικανότητα ακριβούς πρόβλεψης αλλαγών σε κυκλοφοριακά χαρακτηριστικά είναι μείζονος σημασίας στην αποτελεσματική διαχείριση κυκλοφορίας όσον αφορά στην αντιμετώπιση της κυκλοφοριακής συμφόρησης.

## 1.2 ΠΡΟΒΛΕΨΗ ΚΥΚΛΟΦΟΡΙΑΣ

Υπάρχουν δύο κύριες κατευθύνσεις που υιοθετούνται για την πρόβλεψη κυκλοφορίας στους δρόμους:

1. Η καθοδηγούμενη από μοντέλα προσέγγιση (*model – driven approach*).
2. Η καθοδηγούμενη από δεδομένα προσέγγιση (*data – driven approach*).

Η καθοδηγούμενη από μοντέλα προσέγγιση αφορά στην προσπάθεια μοντελοποίησης υπολογιστικά του κυκλοφοριακού δικτύου, συνήθως μέσω προγραμμάτων προσομοίωσης, και ανάλυσης της απόδοσης και της συμπεριφοράς των οδηγών σε μικροσκοπικό συνήθως επίπεδο [5]. Για την καλύτερη δυνατή πρόβλεψη, η τοπολογία του δικτύου αναπαρίσταται λεπτομερώς συμπεριλαμβάνοντας κόμβους, αριθμό λωρίδων, όρια ταχυτήτων, φωτεινή σηματοδότηση κα. Όλη αυτή η πληροφορία αναπαρίσταται υπό μορφή γράφου με τους κόμβους να αναπαριστούν τους κόμβους (κυκλικούς ή μη) και τις διασταυρώσεις, ενώ οι ακμές τους δρόμους. Κατά την διάρκεια της προσομοίωσης, το σύστημα ενεργοποιεί όλες τις οντότητες σύμφωνα με την δοσμένη τοπολογία του δικτύου και τις αλληλεπιδράσεις μεταξύ εικονικών υποδομών και οδηγών, αλλά ακόμα και αλληλεπιδράσεις οδηγών με οδηγών και πεζούς σύμφωνα με εξελιγμένα μοντέλα θεωρία κυκλοφορίας. Έπειτα μελετώντας την κίνηση ενός οχήματος και αναλύοντας τις κυκλοφοριακές συνθήκες σε μία συγκεκριμένη περιοχή, μπορούμε να κάνουμε πρόβλεψη για την μελλοντική κατάσταση του δικτύου. Η καθοδηγούμενη από μοντέλα προσέγγιση συνήθως χρησιμοποιείται για μακροπρόθεσμη πρόβλεψη κυκλοφορίας δηλαδή για αλλαγές στην κατάσταση του δικτύου που έχουν ορίζοντα πρόβλεψης μήνες ή ακόμα και χρόνια χωρίς να λαμβάνεται στο ενδιάμεσο διάστημα οποιασδήποτε μορφής πραγματικού χρόνου δεδομένα [6].

Η δεύτερη προσέγγιση ωστόσο είναι εκείνη που ανθίζει ερευνητικά αντλώντας την αίγλη της από την επανάσταση της πληροφορίας που παρατηρείται τις τελευταίες δεκαετίες. Αυτή η στροφή περνώντας από τις αναλυτικές μεθόδους σε εκείνες που είναι οδηγούμενες από τα δεδομένα οφείλεται και στην άνθιση σύγχρονων μεθόδων ανάλυσης δεδομένων και άντλησης πληροφορίας δανεισμένες από την περιοχή της Υπολογιστικής Νοημοσύνης (*Computational Intelligence – CI*) οι οποίες θα αναλυθούν πιο αναλυτικά παρακάτω στο κείμενο. Επανερχόμενοι στις σύγχρονες τάσεις της διαχείρισης κυκλοφορίας, η δημιουργία αποδοτικών ITS συστημάτων βασίζεται στην συνεχή ροή πληροφορόρησης για τις κυκλοφοριακές συνθήκες στο δίκτυο οι οποίες αλλάζουν συνεχώς στο χρόνο και είναι προβλέψιμες στην φύση

τους [7]. Άρα υπάρχει η εγγενής ανάγκη αλλά και η δυνατότητα (μέσω των δεδομένων που είναι διαθέσιμα) για βραχυπρόθεσμη πρόβλεψη (*short – term prediction*), για την οποία καταλληλότερα είναι τα μοντέλα που είναι καθοδηγούμενα από τα δεδομένα (2<sup>η</sup> προσέγγιση). Σχεδόν πέντε δεκαετίες (από τις αρχές τις δεκαετίας του 80') τώρα, η επιστημονική κοινότητα ασχολείται με το πρόβλημα της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας, όπου με τον όρο βραχυπρόθεσμη εννοούμε ορίζοντα έως 2 ώρες περίπου, και είναι κυρίως επικεντρωμένη στην ανάπτυξη μεθοδολογιών που μπορούν να χρησιμοποιηθούν για την μοντελοποίηση διαφόρων χαρακτηριστικών κυκλοφορίας και να εκτιμήσουν τις αναμενόμενες κυκλοφοριακές συνθήκες του δικτύου όπως αναφέρθηκε σε μικρό ορίζοντα. Δεδομένης της πολυπλοκότητας του προβλήματος της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας και των πολυάριθμων ερευνών που έχουν δημοσιευθεί πάνω στο αντικείμενο, γίνεται εύκολα αντιληπτό ότι η επιλογή του καταλληλότερου μοντέλου για μία συγκεκριμένη εφαρμογή δεν είναι καθόλου τετριμμένη.

### **1.3 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΜΟΝΤΕΛΩΝ ΒΡΑΧΥΠΡΟΘΕΣΜΗΣ ΠΡΟΒΛΕΨΗΣ**

#### **ΚΥΚΛΟΦΟΡΙΑΣ**

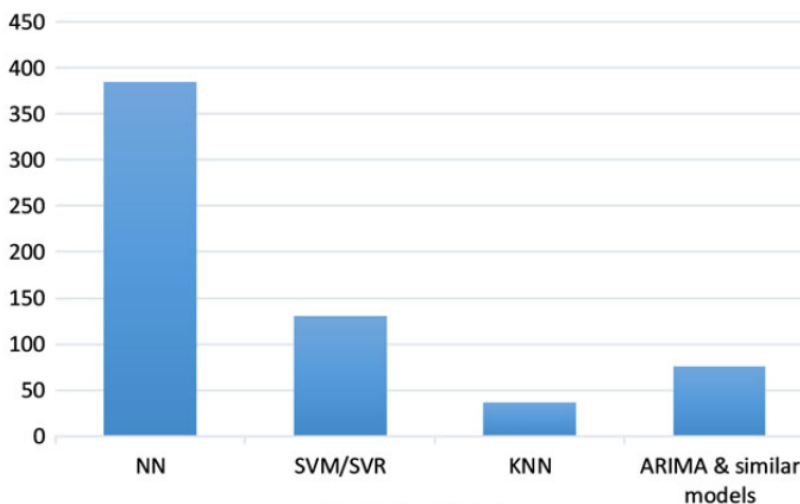
Η βραχυπρόθεσμη πρόβλεψη κυκλοφορίας πραγματοποιείται χρησιμοποιώντας δύο ομάδες μοντέλων:

1. Παραμετρικά μοντέλα
2. Μη παραμετρικά μοντέλα

Ο όρος «παραμετρικά» υποδηλώνει ότι μόνο οι παράμετροι του μοντέλου χρειάζεται να εκτιμηθούν μέσω των διαθέσιμων δεδομένων, ενώ η δομή του μοντέλου είναι προκαθορισμένη. Γνωστό πλεονέκτημα αυτών των μεθόδων είναι ότι για την «εκμάθηση» των παραμέτρων τους απαιτούνται συνήθως λιγότερα δεδομένα σε σχέση με τα μη παραμετρικά μοντέλα [8]. Ορισμένα παραμετρικά μοντέλα έχουν δείξει καλά αποτελέσματα όσον αφορά στην ακρίβεια της προβλέψης και την υπολογιστική πολυπλοκότητα. Η διασημότερη οικογένεια παραμετρικών μοντέλων είναι τα Αυτοπαλίνδρομα Κινητού Μέσου Ολοκληρωμένα μοντέλα (Autoregressive Integrated Moving Average – ARIMA) και πολλές παραλλαγές αυτών.

Ο όρος «μη παραμετρικά» δεν χρησιμοποιείται για να υποδηλώσει ότι σε αυτή την κατηγορία μοντέλων δεν απαιτείται η εκμάθηση παραμέτρων αλλά ότι ο αριθμός και η φύση των παραμέτρων

αυτών είναι ευέλικτα και όχι προκαθορισμένα. Η δομή του μοντέλου και των παραμέτρων του μοντέλου καθορίζονται από τα δεδομένα. Αυτός είναι και ο λόγος που συνήθως χρειάζονται περισσότερα δεδομένα για την «εκμάθηση» των μη παραμετρικών μοντέλων. Ένα πλεονέκτημα τους είναι ότι μπορούν να μοντελοποιήσουν τις δύσκολες, δυναμικές και μη γραμμικές διαδικασίες που χαρακτηρίζουν το φαινόμενο της κυκλοφορίας σε δρόμους. Επιπροσθέτως, καμία εκ των προτέρων γνώση για αυτές τις διαδικασίες δεν χρειάζεται καθώς όπως προαναφέρθηκε μαθαίνονται από τα δεδομένα. Ένα αρνητικό είναι ότι αφού η εκμάθηση βασίζεται εξ' ολοκλήρου στα δεδομένα, περιπτώσεις που δεν συναντώνται στα δεδομένα εκπαίδευσης, όπως κάποια κυκλοφοριακά και μη συμβάντα, είναι δύσκολο να μοντελοποιηθούν [8]. Από τα πιο γνωστά μη παραμετρικά μοντέλα είναι τα μοντέλα Νευρωνικών Δικτύων με πληθώρα παραλλαγών (Neural Networks, NN), η παλινδρόμηση διανυσμάτων υποστήριξης (Support Vector Regression, SVR), η παλινδρόμηση των k κοντινότερων γειτόνων (K-Nearest Neighbors Regression – KNN). Όπως αναφέρεται και στην εργασία των *Do et al.* [9] τα μοντέλα Νευρωνικών Δικτύων είναι τα δημοφιλέστερα καθώς υπάρχουν στην βιβλιογραφία εκατοντάδες επιστημονικές εργασίες όπως απεικονίζεται και στην Εικόνα 1.



**Εικόνα 1** Αριθμός εργασιών ανά κατηγορία μοντέλων (μέχρι το 2018) [9]

Τα τελευταία χρόνια η επιστημονική κοινότητα έχει κάνει στροφή και χρησιμοποιεί προηγμένα μοντέλα Νευρωνικών Δικτύων και βαθιές αρχιτεκτονικές τα οποία αποτελούν μία καινούρια υποκατηγορία των μη παραμετρικών μοντέλων και επέκταση των μοντέλων Νευρωνικών Δικτύων τα οποία ονομάζονται μοντέλα Βαθιάς Μηχανικής Μάθησης (Deep Learning Models). Σε αυτό το

σημείο να αναφερθεί ότι από εδώ και πέρα κάνουμε διάκριση στο κείμενο μεταξύ των μοντέλων βαθιάς μηχανικής μάθησης και των κλασικών μοντέλων μηχανικής μάθησης καθώς η φιλοσοφία τους κάπως διαφοροποιείται.

#### **1.4 ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ ΚΥΚΛΟΦΟΡΙΑΣ**

Είναι σημαντικό σε αυτό το σημείο να γίνει κατανοητό ποια είναι τα ζητήματα και οι πτυχές του προβλήματος της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας που απασχολούν τους ερευνητές τα τελευταία χρόνια. Η υπό-ενότητα αυτή θα βασιστεί κυρίως στα σχόλια και συμπεράσματα της εργασίας των *Lana et al.* [11] η οποία αποτελεί μία λεπτομερή ανασκόπηση των ανασκοπήσεων του προβλήματος της πρόβλεψης κυκλοφορίας μέχρι τα μέσα του 2018. Τα τελευταία χρόνια, μία συλλογή από έρευνες έχει ταξινομήσει και ομαδοποιήσει την υπάρχουσα βιβλιογραφία υπό διαφορετικές οπτικές γωνίες. Από τις πρώτες βιβλιογραφικές ανασκοπήσεις είναι εκείνη των *Van Arem et al.* [12] στην οποία οι συγγραφείς μελέτησαν εφαρμογές της πρόβλεψης κυκλοφορίας σε δυναμική διαχείριση κυκλοφορίας αντιμετωπίζοντας την ως μία οικονομική διαδικασία προσφοράς και ζήτησης. Η εν λόγω βιβλιογραφική ανασκόπηση επικεντρώνεται σε:

1. Ορίζοντα πρόβλεψης
2. Μεθοδολογία πρόβλεψης
3. Τεχνικές αξιολόγησης επίδοσης του μοντέλου
4. Παραδείγματα πραγματικών εφαρμογών

Στην εργασία των *Vlahogianni et al.* [13] τονίζονται τρεις βασικές πτυχές της πρόβλεψης κυκλοφορίας:

1. Ο σκοπός της εφαρμογής, δηλαδή το είδος του δρόμου στον οποίο θα γίνει η πρόβλεψη και ο τύπος της εφαρμογής στον οποίο θα χρησιμοποιηθεί η πρόβλεψη.
2. Ο προσδιορισμός της εξόδου, δηλαδή η επιλογή του ορίζοντα πρόβλεψης και το βήμα της πρόβλεψης.
3. Οι μεταβλητές που θα χρησιμοποιηθούν στην μοντελοποίηση της πρόβλεψης όπως π.χ. κυκλοφοριακός φόρτος, πυκνότητα, μέση ταχύτητα, χρόνος διαδρομής.

Διαφορετική ταξινόμηση έχει γίνει από την εργασία των *Van Hinsbergen et al.* [8] με βάση τις μεθόδους πρόβλεψης που είχαν χρησιμοποιηθεί σε προηγούμενες ερευνητικές εργασίες τα οποία είναι τα:

1. Απλοϊκά μοντέλα (μοντέλα που βασίζονται μόνο σε ιστορικές τιμές), κατηγορία που προσέθεσαν οι συγγραφείς στις επόμενες 2 που είναι οι επικρατέστερες στην βιβλιογραφία.
2. Παραμετρικά μοντέλα
3. Μη Παραμετρικά μοντέλα

Επιπροσθέτως, πήραν υπόψιν τον ορίζοντα πρόβλεψης, τον σκοπό της εφαρμογής, την υπολογιστική ταχύτητα και την ακρίβεια, τα οποία έρχονται σε συμφωνία με πολλές άλλες θεωρήσεις σε άλλες ερευνητικές εργασίες. Οι *Bolshinsky et al.* [14] εστιάζουν στις μεθόδους πρόβλεψης και εκτός από τον ορίζοντα πρόβλεψης υπογραμμίζουν την συσχέτιση με την προέλευση των δεδομένων, και αναφέρονται στη σημασία μη κυκλοφοριακών μεταβλητών (εξωγενών μεταβλητών) οι οποίες μπορούν να επηρεάσουν τις κυκλοφοριακές συνθήκες όπως οι καιρικές συνθήκες, οι περίοδοι διακοπών, διάφορες διοργανώσεις, συμβάντα και εργασίες εντός του κυκλοφοριακού δικτύου καθώς και εποχικούς παράγοντες. Εισάγοντας αυτή την επιπλέον πληροφορία στα μοντέλα πρόβλεψης μπορεί να οδηγήσει σε βελτιωμένες προβλέψεις. Από τις πληρέστερες βιβλιογραφικές ανασκοπήσεις είναι εκείνη των *Vlahogianni et al.* [15] όπου οι συγγραφείς συγκρίναν τον σκοπό της εφαρμογής, τον ορίζοντα πρόβλεψης, την πηγή εισόδου και την μεθοδολογική προσέγγιση για βιβλιογραφικές εργασίες πάνω στο θέμα από το 2004 έως το 2014.

Παραπάνω αναφέραμε κάποια από τα πιο σημαντικά ζητήματα που αφορούν το πρόβλημα της πρόβλεψης κυκλοφορίας. Γενικά, οι βιβλιογραφικές ανασκοπήσεις που έχουν διεξαχθεί πάνω στο θέμα έχουν μέχρι στιγμής αναλύσει το πρόβλημα κάτω από διαφορετικά κριτήρια. Ο Πίνακας 1 συνοψίζει τα πιο σημαντικά κριτήρια με βάση αυτές τις βιβλιογραφικές ανασκοπήσεις. Όπως φαίνεται τα πιο συχνά κριτήρια που χρησιμοποιούνται είναι:

- Μέθοδος πρόβλεψης
- Κλίμακα εφαρμογής
- Μεταβλητές εξόδου (ορίζοντας πρόβλεψης, βήμα)

**Πίνακας 1** Βασικότερα κριτήρια για το πρόβλημα πρόβλεψης κυκλοφορίας σύμφωνα με το [11].

<b>Κριτήριο</b>	<b>Τύποι</b>
Μέθοδος Πρόβλεψης	Απλοϊκά Παραμετρικά Μη παραμετρικά
Ορίζοντας Πρόβλεψης	Ορίζοντας Πρόβλεψης Βήμα Πρόβλεψης
Κλίμακα Πρόβλεψης	Μοναδικό Σημείο Τμήμα Οδού Τμήμα ή ολόκληρο το δίκτυο
Πλαίσιο Πρόβλεψης	Αστικό Υπαίθριο Αυτοκινητόδρομος
Πηγή Δεδομένων	Κινητό Τηλέφωνο GPS-FCD Κάμερα Δημόσιες Συγκοινωνίες Πληθοποριστικά Δεδομένα Κοινωνικά Δίκτυα Αισθητήρες
Εξωγενείς Παράγοντες	Διακοπές Καιρός Ημερολογιακοί Παράγοντες Ημερήσιοι Περίοδοι Εποχικότητες Ειδικά Συμβάντα Περιοδικά Συμβάντα Εργασίες στους Δρόμους Κυκλοφοριακά Συμβάντα Ιδιοσυγκρασία Περιοχών
Μεταβλητή Πρόβλεψης	Κυκλοφοριακός φόρτος Κυκλοφοριακή Πυκνότητα Μέση Ταχύτητα Χρόνος Διαδρομής
Μονο-Πολυμεταβλητή	-
Επίδοση Πρόβλεψης	-
Είδος Βελτιστοποίησης	-
Υπολογιστική Προσπάθεια	-
Ικανότητα Γενίκευσης	-
Σκοπός Εφαρμογής	ATIS, AMTS, Logistic
Ανάλυση Δεδομένων	-
Εξόρυξη από Ροή Δεδομένων	Online, Offline



Από τα παραπάνω γίνεται κατανοητό ότι η πρόβλεψη κυκλοφορίας αποτελεί ένα πολυπαραγοντικό πρόβλημα με πολλές παραμέτρους να εμπλέκονται στην βελτιστοποίηση της διαδικασίας της πρόβλεψης. Εκτός από τα προαναφερθέντα ζητήματα, οι *Van Lint et al.* [16] καταλήγουν ότι δεν υπάρχει μία καθολική μέθοδος για κάθε δεδομένη κατάσταση που να είναι καλύτερη από όλες τις άλλες, αποτέλεσμα που έρχεται σε συμφωνία θεώρημα No Free Lunch [17], αναφέροντας πως είναι πρόκληση η δημιουργία μίας μεθόδου που να επιλέγει το «σωστό» μοντέλο βασισμένη στα χαρακτηριστικά της πρόβλεψης. Οι *Catillo et al.* [18] διατυπώνουν το πρόβλημα ως πρόβλημα μαθηματικής βελτιστοποίησης. Στην εργασία τους οι *Bolshinsky et al.* [14] τονίζουν την δυσκολία σύγκρισης μεθόδων, δυσκολία που ενισχύεται από το γεγονός ότι δεν υπάρχουν σύνολα αναφοράς κυκλοφοριακών δεδομένων (benchmark datasets) ([11]) ώστε να μπορούν να δημιουργηθούν μεταδεδομένα από τις ήδη υπάρχουσες ερευνητικές εργασίες. Τέλος, από όσο γνωρίζουμε σε καμία από τις ανασκοπήσεις μέχρι και το 2019 δεν συγκρίνονται σύγχρονες μέθοδοι από τον κλάδο της βαθιάς μηχανικής μάθησης.

## **1.5 Η ΕΠΕΛΑΣΗ ΤΗΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Αλλάζοντας κλίμα θα αναφερθούμε συντόμως στην επέλαση της βαθιάς μηχανικής μάθησης στον κόσμο της πληροφορικής και της ανάλυσης δεδομένων. Το γεγονός αυτό είναι εμφανές και στον ευρύτερο κλάδο της ανάλυσης κυκλοφοριακών δεδομένων, αφού η επιστημονική κοινότητα έχει κάνει στροφή όσον αφορά την πρόβλεψη κυκλοφορίας σε μεθόδους βαθιάς μηχανικής μάθησης. Ομολογουμένως, στον τομέα της επιστήμης υπολογιστών ο κλάδος της υπολογιστικής νοημοσύνης και πιο συγκεκριμένα ο υποκλάδος της (Βαθιά Μηχανική Μάθηση) είναι από τους πιο γοργά αναπτυσσόμενους επιστημονικούς κλάδους τα τελευταία χρόνια, με χιλιάδες επιστημονικές εργασίες ανά έτος, λόγω των εξαιρετικών αποτελεσμάτων που έχει να επιδείξει. Άλλος ένα λόγος που συντέλεσε σε αυτήν την ανάπτυξη είναι η εξέλιξη των υπολογιστών σε επίπεδο υλικού με την χρήση των Μονάδων Επεξεργασίας Γραφικών (Graphics processing unit – GPU) ως επεξεργαστικές μονάδες για την εκπαίδευση των μοντέλων βαθιάς μηχανικής μάθησης γεγονός που την καθιστά πλέον πραγματικά εφικτή.

Μία από τις βασικές δυσκολίες σε πολλές εφαρμογές τεχνητής νοημοσύνης είναι ότι πολύ παράγοντες μεταβλητότητας επηρεάζουν κάθε ένα μέρος από τα δεδομένα που μπορούμε να παρατηρήσουμε. Στις περισσότερες εφαρμογές απαιτείται να «διαχωρίσουμε» τους παράγοντες

εκείνους της μεταβλητότητας που μας ενδιαφέρουν και να απορρίψουμε εκείνους που δεν βοηθούν στην επίλυση του προβλήματος. Βεβαίως, είναι πολύ δύσκολο να εξαχθούν τέτοιου υψηλού επιπέδου, αφηρημένα χαρακτηριστικά από ανεπεξέργαστα δεδομένα. Όταν η πολυπλοκότητα της εξαγωγής τέτοιου είδους αναπαραστάσεων των δεδομένων για την επίλυση του αρχικού προβλήματος αυξάνεται, φαίνεται η εκμάθηση αναπαράστασης εκ πρώτης όψεως να μην βοηθάει [10].

Η Βαθιά Μηχανική Μάθηση λύνει αυτό το κεντρικό πρόβλημα στην μάθηση αναπαράστασης με το να εισάγει αναπαραστάσεις οι οποίες εκφράζονται μέσω άλλων απλούστερων αναπαραστάσεων. Ένα απλό παράδειγμα μεθόδου βαθιάς μηχανικής μάθησης είναι το βαθύ δίκτυο εμπρόσθιας μετάδοσης ή κυρίως γνωστό και ως πολυεπίπεδο περσέπτρον (multilayer perceptron – MLP). Ένα MLP είναι απλώς μία μαθηματική συνάρτηση που απεικονίζει κάποιο σύνολο δεδομένων εισόδου σε ένα σύνολο δεδομένων εξόδου. Η συνάρτηση αυτή σχηματίζεται συνθέτοντας άλλες απλούστερες συναρτήσεις. Μπορούμε να σκεφτούμε ότι κάθε εφαρμογή μίας διαφορετικής μαθηματικής συνάρτησης επιστρέφει μία νέα αναπαράσταση της εισόδου [10].

Η ιδέα μάθησης της σωστής αναπαράστασης των δεδομένων είναι μόνο η μία πλευρά των μεθόδων βαθιάς μηχανικής μάθησης. Η άλλη οπτική είναι ό,τι το «βάθος» επιτρέπει στον υπολογιστή να μάθει ένα υπολογιστικό πρόγραμμα πολλών βημάτων. Κάθε επίπεδο της αναπαράστασης μπορεί να το σκεφτεί κανείς ως την κατάσταση της μνήμης του υπολογιστή αφού εκτελεστεί άλλο ένα σύνολο από οδηγίες παράλληλα. Δίκτυα με μεγαλύτερο βάθος μπορούν να εκτελέσουν περισσότερες οδηγίες στην σειρά. Οι ακολουθιακές οδηγίες προσφέρουν μεγάλη ισχύ, διότι αργότερα, οι οδηγίες μπορούν να «ψάξουν» πίσω στα αποτελέσματα προηγούμενων χρονικά οδηγιών [10].

Εν κατακλείδι, η βαθιά μηχανική μάθηση αποτελεί ένα συγκεκριμένο είδος μηχανικής μάθησης που επιτρέπει σε ένα υπολογιστικό σύστημα να βελτιώνεται με την εμπειρία και τα δεδομένα. Πρόκειται για εξαιρετικά ισχυρές και ευέλικτες μεθόδους, οι οποίες μπορούν να μαθαίνουν να αναπαριστούν τον κόσμο σαν μία εμφωλευμένη ιεραρχία από πλαίσια, με κάθε πλαίσιο να καθορίζεται σε σχέση με απλούστερα πλαίσια, και αφηρημένες αναπαραστάσεις που υπολογίζονται σε σχέση με λιγότερο αφηρημένες αναπαραστάσεις. Λεπτομέρειες για την αρχιτεκτονική τους, την φιλοσοφία τους και τον τρόπο που λειτουργούν θα δοθούν σε επόμενο κεφάλαιο.

## 1.6 ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Στην παρούσα διπλωματική εργασία επιδιώκεται μία συγκριτική ανάλυση 15 διαφορετικών μεθόδων πρόβλεψης από τις οποίες 10 αφορούν σε μεθόδους βαθιάς μηχανικής μάθησης ενώ οι υπόλοιπες 5 αφορούν σε γνωστές στατιστικές μεθόδους χρονοσειρών (κάνοντας χρήση αυτοματοποιημένων διαδικασιών) για βραχυπρόθεσμη πρόβλεψη μονομεταβλητών κυκλοφοριακών χρονοσειρών (μέση ταχύτητα διαδρομής) και για διάφορες τιμές του ορίζοντα πρόβλεψης. Απώτερος σκοπός είναι η ανάδειξη δυσκολιών που σχετίζονται με την εκπαίδευση τέτοιων μεθόδων (Βαθιάς Μηχανικής Μάθησης). Εξετάστηκαν επίσης, δύο μέθοδοι προεπεξεργασίας δεδομένων, δανεισμένες από την κλασσική ανάλυση χρονοσειρών, και πως αυτές επηρεάζουν την επίδοση των μεθόδων βαθιάς μηχανικής μάθησης. Ως μέθοδος αξιολόγησης της προβλεπτικής ικανότητας των μοντέλων χρησιμοποιήθηκε η μέθοδος κυλιόμενης αρχής πρόβλεψης χωρίς επανεκπαίδευση, για 5 διαφορετικές μετρικές σφάλματος. Το πρόβλημα της πρόβλεψης διατυπώνεται ως πρόβλημα αυτοπαλινδρόμησης ( $AR(p)$ ) χωρίς εξωγενείς μεταβλητές. Τέλος, υπολογίζονται μέσω μίας μετρικής οι υπολογιστικές απαιτήσεις για κάθε μοντέλο, εξετάστηκε η σχέση της προσαρμογή του μοντέλου με την προβλεπτική του ικανότητα, και η ευαισθησία των μη παραμετρικών μοντέλων στον χώρο των παραμέτρων που εξετάστηκαν.

### 1.6.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Εφαρμόζονται 10 διαφορετικά μοντέλα πρόβλεψης βαθιάς μηχανικής μάθησης και 5 διαφορετικά μοντέλα στατιστικών μεθόδων καθώς και 1 μέθοδος αναφοράς ανάλυσης κυκλοφοριακών χρονοσειρών μέσης ταχύτητας διαδρομής από ανεπεξεργαστα GPS δεδομένα.
2. Αξιολογούνται οι 15 μέθοδοι χρησιμοποιώντας την μέθοδο αξιολόγησης κυλιόμενης αρχής πρόβλεψης χωρίς επανεκπαίδευση για 5 διαφορές μετρικές σφάλματος για τις οποίες βρέθηκαν τα εξής:
  - Όλες οι εξεταζόμενες μέθοδοι έχουν καλύτερη προβλεπτική ικανότητα από την μέθοδο αναφοράς.
  - Φαίνεται ότι οι μέθοδοι βαθιάς μηχανικής μάθησης υπερτερούν των στατιστικών μεθόδων, χωρίς η επιλογή της μετρικής σφάλματος να επηρεάζει την επιλογή της καλύτερης μεθόδου.

3. Εξετάζονται δύο μέθοδοι προεπεξεργασίας δεδομένων, δανεισμένες από την κλασσική ανάλυση χρονοσειρών, την αποεποχικοποίηση των δεδομένων και την αφαίρεση τάσης, και πως αυτές επηρεάζουν την επίδοση μοντέλων βαθιάς μηχανικής μάθησης. Για τον υπολογισμό της 1<sup>ης</sup> έγινε χρήση της μεθόδου “LOESS” ενώ για τον υπολογισμό της 2<sup>ης</sup> χρησιμοποιήθηκε η μέθοδος των κυβικών σπλινών. Τα αποτελέσματα καταδεικνύουν ότι και οι 2 μέθοδοι προεπεξεργασίας είναι στατιστικά σημαντικές για την βελτίωση της απόδοσης των μεθόδων βαθιάς μηχανικής μάθησης που εξετάστηκαν. Μάλιστα, η αφαίρεση της εποχικής τάσης βρέθηκε να έχει σημαντική μειωτική επιρροή στο σφάλμα για όλες τις μεθόδους.
4. Συγκρίνεται η υπολογιστική πολυπλοκότητα των μεθόδων βαθιάς μηχανικής μάθησης με την αντίστοιχη των στατιστικών μεθόδων. Από την σύγκριση προέκυψε ότι η υπολογιστική πολυπλοκότητα των πρώτων είναι τρεις τάξεις μεγέθους μεγαλύτερη.

### **1.6.2 Οργάνωση κειμένου**

Εργασίες σχετικές με το αντικείμενο της διπλωματικής παρουσιάζονται στο Κεφάλαιο 2. Το Κεφάλαιο 3 περιλαμβάνει την συλλογή δεδομένων, την δημιουργία του συνόλου δεδομένων και τον καθαρισμό των δεδομένων. Στο Κεφάλαιο 4 αναπτύσσεται το θεωρητικό υπόβαθρο για όλες τις μεθόδους που υλοποιήθηκαν. Στο Κεφάλαιο 5 περιγράφεται η μεθοδολογική προσέγγιση που ακολουθήθηκε. Στο Κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα. Και τέλος, στο Κεφάλαιο 7 δίνονται τα συμπεράσματα καθώς και προτάσεις για περαιτέρω έρευνα.

## 2 ΣΥΝΤΟΜΗ ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

---

Οι ραγδαίες τεχνολογικές εξελίξεις στις αρχές του 21<sup>ου</sup> αιώνα έχουν δώσει πνοή και φαίνεται να δημιουργούν νέους ορίζοντες στον κλάδο της πρόβλεψης κυκλοφορίας και να προσφέρονται πλέον εναλλακτικές μέθοδοι έναντι των κλασικών γεγονόσ που τονίζεται έντονα στην βιβλιογραφία ([20], [22], [23], [25]-[38]). Οι μέθοδοι που χρησιμοποιούνται κατά κόρον είναι τα ανατροφοδοτούμενα νευρωνικά δίκτυα, τα συνελκτικά νευρωνικά δίκτυα και πληθώρα παραλλαγών τους. Ωστόσο, οι περισσότερες πρότερες ερευνητικές εργασίες είναι κυρίως στατιστικής φύσεως και στηρίζονται στις κλασσικές στατιστικές μεθόδους πρόβλεψης χρονοσειρών.

Στην βιομηχανία έχει δοθεί ιδιαίτερη έμφαση στις μεθόδους βαθιάς μηχανικής μάθησης, με πολλές εφαρμογές και στον τομέα των μεταφορών ([19], [20], [21], [22], [23]). Πράγματι, τα νευρωνικά δίκτυα βάθους και οι τελευταίες παραλλαγές τους αντιμετωπίζονται ως τα καλύτερα και πιο υποσχόμενα μοντέλα σε επίπεδο επιδόσεων. Παρόλα αυτά εντοπίζονται και πρόσφατες ερευνητικές εργασίες [24] στις οποίες χρησιμοποιούνται κλασσικές στατιστικές μέθοδοι πρόβλεψης χρονοσειρών και έχουν πολύ καλά αποτελέσματα.

Από την υποενότητα 1.5 έγινε ήδη εμφανές ότι στην βιβλιογραφία παρατηρείται πληθώρα ερευνητικών ανασκοπήσεων για το πρόβλημα της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας όπου σχολιάζονται μέθοδοι που έχουν χρησιμοποιηθεί γενικά και τονίζονται τα θετικά τους και τα αρνητικά τους. Παρόλα αυτά ενώ πολλές ερευνητικές εργασίες αναφέρουν την αποτελεσματικότητα των διαφόρων μεθόδων που βασίζονται στα τεχνητά νευρωνικά δίκτυα, λίγη έρευνα έχει γίνει όσον αφορά στην λεπτομερή σύγκριση των πλέον σύγχρονων μεθόδων βαθιάς μηχανικής μάθησης με τις κλασσικές στατιστικές μεθόδους χρονοσειρών για μονομεταβλητές χρονοσειρές με σκοπό την βαθύτερη κατανόηση τους και το πως «συμπεριφέρονται» και αλληλοεπιδρούν (σε επίπεδο βελτιστοποίησης των υπερπαραμέτρων) με τις ιδιοσυγκρασίες που διέπουν τις κυκλοφοριακές χρονοσειρές. Η διαδικασία σύγκρισης είναι απαιτητική και δύσκολη καθώς δεν υπάρχει κάποια καθολική μέθοδος και δύσκολα μπορούν να εξαχθούν μεταδεδομένα από την υπάρχουσα βιβλιογραφία αλλά και συμπεράσματα από τις βιβλιογραφικές ανασκοπήσεις λόγω της έλλειψης συνόλων αναφοράς. Στην βιβλιογραφία επομένως δεν εμφανίζονται συχνά εκτενείς ποσοτικές συγκριτικές αναλύσεις και κυρίως των πιο πρόσφατων μεθόδων βαθιάς μηχανικής μάθησης. Μία από τις πρώτες συγκριτικές αναλύσεις είναι εκείνη των *Clark et al.* [39]

όπου οι συγγραφείς συγκρίνουν απλά νευρωνικά δίκτυα με διάφορες στατιστικές μεθόδους πρόβλεψης χρονοσειρών για ορίζοντα μίας ώρας. Γενικά αναφέρουν ότι οι στατιστικές μέθοδοι έχουν καλύτερα αποτελέσματα με μικρή όμως διαφορά από τα νευρωνικά δίκτυα και ότι άλλοι παράγοντες είναι καθοριστικοί για την επιλογή του μοντέλου και όχι τα μοντέλα τα ίδια. Αργότερα, οι Kirby *et al.* [40] συγκρίνουν Αυτοπαλίνδρομα Κινητού Μέσου Ολοκληρωμένα μοντέλα (Autoregressive Integrated Moving Average – ARIMA) με απλά νευρωνικά δίκτυα και ένα τρίτο μοντέλο το ATHENA [41] με κυκλοφοριακά δεδομένα που αφορούν σε αυτοκινητόδρομο. Οι συγγραφείς καταλήγουν ότι τα δύο πρώτα μοντέλα έχουν την ίδια ακρίβεια, με το τρίτο να είναι το καλύτερο. Τέλος, σχολιάζουν ότι η ακρίβεια του μοντέλου να μην είναι σημαντικός παράγοντας όσον αφορά την επιλογή του μοντέλου αλλά δεν πρέπει να στηρίζεται κανείς αποκλειστικά μόνο σε αυτό.

Οι Smith *et al.* [42] διεξήγαγαν σύγκριση μεταξύ τεσσάρων διαφορετικών μοντέλων, το μοντέλο ιστορικών μέσων (Historical Averages), το μοντέλο ARIMA, το μοντέλο νευρωνικών δικτύων και τέλος εκείνο της μη παραμετρικής παλινδρόμησης. Τα αποτελέσματα έδειξαν σημαντική υπεροχή της τελευταίας μεθόδου, η οποία όπως χαρακτηριστικά αναφέρουν είναι όχι μόνο η εύκολη στην υλοποίηση αλλά μπορεί να μεταφερθεί εύκολα σε δεδομένα άλλων περιοχών. Οι Smith *et al.* [43] σε μετέπειτα εργασία τους επανεξετάζουν την περίπτωση της μη παραμετρικής παλινδρόμησης χρησιμοποιώντας διαφορετικές μεθόδους βελτιστοποίησης και την συγκρίνουν με το εποχικό μοντέλο ARIMA σε μονομεταβλητές κυκλοφοριακές χρονοσειρές. Καταλήγουν ότι το εποχικό μοντέλο ARIMA είναι ανώτερο της μη παραμετρικής παλινδρόμησης, γεγονός που καταδεικνύει την στοχαστική συμπεριφορά της κυκλοφορίας. Στην εργασία τους οι Tang *et al.* [44] εξετάζουν τα ίδια μοντέλα με το [42] και προσθέτουν το γκαουσιανό μοντέλο μέγιστης πιθανοφάνειας το οποίο καταλήγουν ότι είναι και το καλύτερο σε ακρίβεια σε σχέση με τα υπόλοιπα. Στην εργασία των Vlahogianni *et al.* [45] γίνεται σύγκριση μεταξύ διαφόρων παραλλαγών μοντέλων ARIMA και νευρωνικών δικτύων για μονομεταβλητές και πολυμεταβλητές χρονοσειρές. Τα νευρωνικά δίκτυα φαίνεται να υπερτερούν έναντι των κλασικών στατιστικών μεθόδων ιδιαίτερα για υψηλότερης ανάλυσης δεδομένα (μεγαλύτερη συχνότητα). Τέλος, η εργασία των Karlaftis *et al.* [46] αποτελεί μία εξαιρετική ποιοτική σύγκριση μεταξύ στατιστικών μεθόδων πρόβλεψης κυκλοφοριακών χρονοσειρών και νευρωνικών δικτύων. Γενικά, η πλειοψηφία των παραπάνω άρθρων χρησιμοποιεί ως μετρικές αξιολόγησης της προβλεπτικής ικανότητας των υπό εξέταση μεθόδων το Μέσο Απόλυτο Σφάλμα (Mean Absolute

Error – MAE) και το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error – MAPE). Παρόλα αυτά τονίζεται ότι και άλλες μετρικές αξιολόγησης πρέπει να λαμβάνονται υπόψη όπως ο χρόνος και η προσπάθεια για την ανάπτυξη της μεθόδου, η μεταφερσιμότητα των αποτελεσμάτων, οι ικανότητες και η εξειδίκευση που απαιτείται και πολλά άλλα.

Από όσο γνωρίζουμε μέχρι και σήμερα η πιο εκτενής ποσοτικά συγκριτική ανάλυση είναι η εργασία του *Howell* [47]. Ο συγγραφέας συγκρίνει 21 διαφορετικά μοντέλα για πρόβλεψη κυκλοφορίας για πολυμεταβλητή βραχυπρόθεσμη πρόβλεψη κυκλοφορίας με δεδομένα από 411 αισθητήρες σε αστικό κυκλοφοριακό δίκτυο σε πόλη της Μεγάλης Βρετανίας. Στα μοντέλα που εξετάστηκαν περιλαμβάνονται 2 μοντέλα βαθιάς μηχανικής μάθησης και πιο συγκεκριμένα 2 παραλλαγές ανατροφοδοτούμενων νευρωνικών δικτύων (Long Short Term Memory – LSTM και Gated Recurrent Units – GRU) αλλά και αρκετές μέθοδοι μηχανικής μάθησης καθώς και 1 πολυμεταβλητό στατιστικό μοντέλο πρόβλεψης χρονοσειρών (VAR). Τα μοντέλα συγκρίνονται ως προς την προβλεπτική τους ικανότητα αλλά λαμβάνονται υπόψη και πιο ρεαλιστικοί παράγοντες όπως ο χρόνος εκπαίδευσης και πρόβλεψης, η ευαισθησία των υπερπαραμέτρων και η πολυπλοκότητα της εφαρμογής κάθε μοντέλου.

Η τελευταία συγκριτική ανάλυση που παρατηρείται στην βιβλιογραφία είναι η εργασία των *Bratsas et al.* [48]. Οι συγγραφείς συγκρίνουν 3 μεθόδους μηχανικής μάθησης και πιο συγκεκριμένα την μέθοδο Random Forests, Support Vector Regression και Multilayer Perceptron σε probe δεδομένα από την πόλη της Θεσσαλονίκης για 3 διαφορετικά σενάρια. Τα αποτελέσματα δείχνουν ότι στις σταθερές συνθήκες με μικρή μεταβλητότητα το Support Vector Regression μοντέλο έχει την καλύτερη ακρίβεια ενώ στο σενάριο όπου οι συνθήκες είναι πιο ασταθείς και τα δεδομένα παρουσιάζουν μεγαλύτερη μεταβλητότητα το Multilayer Perceptron ήταν το καλύτερο όσον αφορά την ακρίβεια αλλά και τον αριθμό των σχεδόν μηδενικών σφαλμάτων.

Αξίζει να αναφερθούν σε αυτό το σημείο σε αυτό το σημείο και κάποιες συγκριτικές αναλύσεις που έχουν βρεθεί στην βιβλιογραφία γενικά για πρόβλεψη χρονοσειρών με σκοπό τον εντοπισμό των προβλημάτων που εμφανίζονται γενικά στον επιστημονικό κλάδο της πρόβλεψης χρονοσειρών μιας και η βραχυπρόθεσμη πρόβλεψη κυκλοφορίας αποτελεί υποκλάδο της. Οι *Ahmed et al.* [49] χρησιμοποιούν δεδομένα από τον M3 διαγωνισμό και συγκεκριμένα 1000 μονομεταβλητές χρονοσειρές με συχνότητα ενός μήνα. Εξετάστηκαν 8 διαφορετικά μέθοδοι μηχανικής μάθησης από τις οποίες οι 2 καλύτερες, ήταν η μέθοδος πολυεπίπεδου περσέπτρον και

η παλινδρόμηση γκαουσιανής διαδικασίας. Επιπροσθέτως, εξετάστηκαν και διάφοροι μέθοδοι προεπεξεργασίας δεδομένων που συναντάμε κυρίως στις κλασσικές στατιστικές μεθόδους πρόβλεψης χρονοσειρών. Στην εργασία τους οι *Makridakis et al.* [50] χρησιμοποιούνται και πάλι 1045 χρονοσειρές με μηνιαία συχνότητα μέτρησης από τον διαγωνισμό M3. Οι συγγραφείς συγκρίνουν 8 κλασσικές στατιστικές μεθόδους πρόβλεψης χρονοσειρών με τις πιο διάσημες μεθόδους μηχανικής μάθησης να περιλαμβάνονται στην ανάλυση. Ουσιαστικά, η εργασία αυτή αποτελεί μία επέκταση και επικαιροποίηση της προηγούμενης εργασίας ([49]) η οποία διεξάχθηκε 8 χρόνια νωρίτερα. Εξετάστηκαν διάφορες μέθοδοι πρόβλεψης πολλών βημάτων για τις μεθόδους μηχανικής μάθησης καθώς και αρκετές μέθοδοι προεπεξεργασίας των δεδομένων. Τέλος, η πρόβλεψη διεξάχθηκε για διάφορες τιμές του ορίζοντα πρόβλεψης. Τα αποτελέσματα έδειξαν ότι όλες οι στατιστικές μέθοδοι ήταν ανώτερες των μεθόδων μηχανικής μάθησης και για όλες τις τιμές του ορίζοντα πρόβλεψης. Στην εργασία τους οι *Cerqueira et al.* [51] ορμώμενοι από την προηγούμενη συγκριτική ανάλυση([50]) θεωρούν ότι έχει εισαχθεί μεροληψία από τους συγγραφείς υπέρ των στατιστικών μεθόδων με το να χρησιμοποιούν μικρά σύνολα δεδομένων. Προς υποστήριξη αυτής της πεποίθησης, συγκρίνουν σχεδόν τις ίδιες μεθόδους αλλά χρησιμοποιώντας ένα σύνολο δεδομένων 100 χρονοσειρών με πάνω από 1000 παρατηρήσεις η κάθε μία και καταλήγουν ότι για ένα αριθμό παρατηρήσεων και πάνω οι μέθοδοι μηχανικής μάθησης υπερτερούν των κλασσικών στατιστικών μεθόδων.

Ανακαλώντας την υποενότητα 1.5 και την σύντομη βιβλιογραφική ανασκόπηση που έγινε, παρατηρείται ότι σε καμία συγκριτική ανάλυση δεν περιλαμβάνονται οι πλέον σύγχρονες μέθοδοι βαθιάς μηχανικής μάθησης για πρόβλεψη χρονοσειρών. Στην γενικότερη βιβλιογραφία πάνω στο αντικείμενο παρατηρείται η τάση προς αύξηση της πολυπλοκότητας των μοντέλων χρησιμοποιώντας όλο και πιο καινούριες μεθόδους λαμβάνοντας υπόψη σαν μετρική αξιολόγησης του μοντέλου μόνο την ακρίβεια. Παρότι είναι σημαντική η χρήση νέων μεθόδων, θεωρείται ακόμα πιο σημαντική η δημιουργία ουσιαστικών συστάσεων για την αποτελεσματική επιλογή μεθόδων ανάλογα με το πρόβλημα που καλείται να επιλύσει ο ερευνητής. Η διπλωματική αποτελεί μία αφετηρία η οποία έχει ως απώτερο σκοπό την δημιουργία μίας μεγάλης βάσης μεταδεδομένων από την εφαρμογή πληθώρας μεθόδων που σε συνδυασμό με τεχνικές μεταμάθησης και αυτοματοποιημένων μεθόδων εκπαίδευσης να οδηγήσει σε μία πλήρως αυτοματοποιημένη διαδικασία για βραχυπρόθεσμη πρόβλεψη κυκλοφορίας.



### **3 ΠΡΟΕΤΟΙΜΑΣΙΑ ΚΑΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ**

---

Η επιλογή του συνόλου των δεδομένων είναι γενικά σημαντική παράμετρος στο πρόβλημα της πρόβλεψης κυκλοφορίας (data source criterion [11]) καθώς ανάλογα με την πηγή των δεδομένων αναμένεται να είναι και διαφορετική η φύση των δεδομένων (μεταβλητότητα, συχνότητα μέτρησης, κυκλοφοριακές μεταβλητές που μπορούν να εξαχθούν κ.α.) και άρα έχει επίπτωση στην επιλογή της μεθόδου. Για τις ανάγκες της διπλωματικής θα χρησιμοποιηθούν μόνο ενδογενείς κυκλοφοριακές μεταβλητές χωρίς καμία εξωγενή μεταβλητή.

#### **3.1 ΜΕΘΟΔΟΙ ΣΥΛΛΟΓΗΣ ΚΥΚΛΟΦΟΡΙΑΚΩΝ ΔΕΔΟΜΕΝΩΝ**

Στης 1<sup>ης</sup> γενιάς συστήματα ATMS και ITS, η κύρια πηγή δεδομένων προερχόταν από αισθητήρες που βρίσκονταν σε σταθερή θέση, και είχαν την δυνατότητα να ανιχνεύουν την παρουσία των κοντινών αυτοκινήτων. Αρχικά, η πιο δημοφιλής πηγή ήταν εκείνη των φωρατών, πλέον όμως υπάρχουν πολύ διαφορετικοί τύποι αισθητήρων που είναι διαθέσιμοι σε ένα κυκλοφοριακό δίκτυο [52]. Παρακάτω θα αναφέρουμε κάποιους από τους πιο γνωστούς.

Οι επαγωγικοί βρόγχοι είναι ενσωματωμένοι στους δρόμους και έχουν την δυνατότητα να εντοπίζουν την παρουσία αγωγίμου μετάλλου μετρώντας την διαφορά του μαγνητικού πεδίου. Οι μαγνητικοί αισθητήρες εντοπίζουν την παρουσία σιδηρούχων μετάλλων μετρώντας την αλλαγή που προκαλούν στο μαγνητικό πεδίο της γης. Οι αισθητήρες εικόνας-βίντεο αναλύουν την εικόνα από κάμερες παρακολούθησης της κυκλοφορίας εντός του κυκλοφοριακού δικτύου και εξάγουν δεδομένα κυκλοφοριακής ροής σε επίπεδο λωρίδας. Οι αισθητήρες ραντάρ μικροκυμάτων εκπέμπουν ηλεκτρομαγνητικά κύματα και προσλαμβάνουν πίσω υπερήχους από τα αντικείμενα ενδιαφέροντος. Οι αισθητήρες ραντάρ λέιζερ είναι τύπος ενεργητικών αισθητήρων οι οποίοι μεταδίδουν υπέρυθρες ακτίνες σάρωσης σε κοντινό φάσμα υπερήχων σε μία ή περισσότερες λωρίδες. Τέλος, οι αισθητήρες ήχου είναι παθητικού τύπου αισθητήρες και χρησιμοποιούν διαφορετικές τεχνικές επεξεργασίας ηχητικού σήματος για να υπολογίσουν την κυκλοφοριακή πυκνότητα ή τον κυκλοφοριακό φόρτο [53].

Η έλευση των εφοδιασμένων με GPS συσκευές κινητών τηλεφώνων και οχημάτων έχει δώσει νέα πνοή στην συλλογή κυκλοφοριακών δεδομένων. Πλέον, είναι δυνατή η συλλογή πιο λεπτομερών δεδομένων και με μεγαλύτερη ανάλυση σε σχέση με τις προαναφερθείσες μεθόδους σταθερού

σημείου. Οι πραγματικού χρόνου και οι ιστορικές κυκλοφοριακές τροχιές που εξάγονται από τις εφοδιασμένες συσκευές GPS έχουν δώσει την δυνατότητα για καλύτερες προβλέψεις όσον αφορά την κυκλοφοριακή ροή οι οποίες μπορούν να χρησιμοποιηθούν από τα συστήματα ATMS και ITS [53].

### 3.2 Η ΠΗΓΗ ΤΟΥ ΑΝΕΠΕΞΕΡΓΑΣΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

Το σύνολο των δεδομένων που χρησιμοποιήθηκε για τις ανάγκες της διπλωματικής εργασίας είναι ένα ανοικτό σύνολο δεδομένων (open dataset) το οποίο προσφέρεται από την πλατφόρμα της εταιρείας DiDi Express, από οδηγούς ταξί που είναι συμβεβλημένοι με την εταιρία (DiDi Premier drivers), (Πηγή: <https://gaia.didichuxing.com>) και αφορά δεδομένα τροχιών εντός του δεύτερου δακτυλίου της πόλης Xi'An (Κίνα). Η DiDi είναι εταιρία που εδράζεται στην Αυστραλία και δραστηριοποιείται στο τομέα παροχής υπηρεσιών ridesharing. Το διάστημα μετρήσεων του στίγματος από gps συσκευές εντός των οχημάτων είναι 2 – 4 δευτερόλεπτα. Τα στίγματα gps έχουν οριοθετηθεί εντός των φυσικών δρόμων κατά τρόπο ώστε τα δεδομένα τροχιών και οι πραγματικοί δρόμοι να μπορούν να αντιστοιχηθούν. Να σημειωθεί ότι για λόγους προστασίας δεδομένων οι πληροφορίες που αφορούν τους οδηγούς και τα ταξίδια έχουν κρυπτογραφηθεί και ανωνυμοποιηθεί. Το χωρικό εύρος των μετρήσεων δίνεται από τις συντεταγμένες του ορθογωνίου που ακολουθούν: [108.92309,34.279936], [109.008833,34.278608], [109.009348,34.207309], [108.921859,34.204946] και φαίνεται στην Εικόνα 2 και 3.



**Εικόνα 2** Ο 2<sup>ος</sup> δακτύλιος της πόλης Xi'an (Κίνα) και το χωρικό εύρος των μετρήσεων (Πηγή: <https://gaia.didichuxing.com>)



**Εικόνα 3** Ο 2<sup>ος</sup> δακτύλιος της πόλης Xi'an (Κίνα) (Πηγή: <https://www.openstreetmap.org/>)

Το χρονικό εύρος είναι οι μήνες Οκτώβριος (1-31) έως και Νοέμβριο (1-30) του έτους 2016 (συνολικά 61 ημέρες). Τα δεδομένα δίνονται σε δύο αρχεία, ένα για κάθε μήνα, και κάθε ένα από αυτά έχει ομαδοποιηθεί σε αρχεία που αφορούν μετρήσεις μίας ημέρας. Τα αρχεία που αφορούν σε κάθε μήνα συνολικά δίνονται σε μορφή tar.gz και κάθε αρχείο έχει όγκο σε συμπιεσμένη μορφή  $\approx 5.4\text{gb}$  (συνολικά  $\approx 11\text{gb}$ ). Οι συντεταγμένες των σημείων δίνονται σε GCJ-02 σύστημα συντεταγμένων (κινέζικο σύστημα). Στην Εικόνα 4 δίνεται ένα δείγμα από το σύνολο των ακατέργαστων δεδομένων.

Field	Type	Sample	Comment
Driver ID	String	glox.jrrlltBMvCh8nxqktdr2dt opmlH	Anonymized
Order ID	String	jkkt8kxniovlFuns9qrrlvst@iq nkwz	Anonymized
Time Stamp	String	1501584540	Unix timestamp, in seconds
Longitude	String	104.04392	GCJ-02 Coordinate System
Latitude	String	104.04392	GCJ-02 Coordinate System

**Εικόνα 4** Δείγμα από το σύνολο των ακατέργαστων δεδομένων (Πηγή: <https://gaia.didichuxing.com>)

### 3.3 ΔΗΜΙΟΥΡΓΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

Για τις ανάγκες της διπλωματικής θα χρησιμοποιηθούν δεδομένα από GPS συσκευές εντός οχημάτων δηλαδή δεδομένα από κινούμενους αισθητήρες. Σε αντίθεση με του αισθητήρες σταθερής θέσης, οι εν λόγω αισθητήρες βρίσκονται συνέχεια σε κίνηση αν ο ιδιοκτήτης κινεί το όχημα. Χρησιμοποιώντας φορμαλισμό, τα δεδομένα από αισθητήρες GPS μπορούν να αναπαρασταθούν ως μία διατεταγμένη χρονικά ακολουθία μετρήσεων όπου κάθε δείγμα συντεταγμένης GPS συσχετίζεται με ένα χρονικό δείκτη:

$$\overline{\mathbf{pos}} = \{pos_1 \rightarrow pos_2 \rightarrow \dots \rightarrow pos_t \rightarrow \dots\} \quad (3.1)$$

και κάθε  $pos_t$  περιέχει το γεωγραφικό πλάτος και μήκος καθώς και την χρονοσήμανση  $t$ . Οπότε το σύνολο δεδομένων είναι ένα σύνολο από ακολουθίες γεωγραφικού πλάτους και μήκους με χρονοσήμανση  $t = t_{start}^{id}, \dots, t_{end}^{id}$ , με  $t_{start}^{id}$  να είναι ο χρόνος εκκίνησης ενός ταξιδιού με  $id$  το  $id$  του οδηγού και  $t_{end}^{id}$  ο χρόνος που τελειώνει το ταξίδι του, ο ίδιος οδηγός. Από την περιγραφή που έχουμε για το σύνολο δεδομένων ισχύει ότι  $2 \text{ sec} \leq t_{i+1}^{id} - t_i^{id} \leq 4 \text{ sec}$ .

Η περιοχή του 2<sup>ου</sup> δακτυλίου της πόλης Χί'αν περιλαμβάνει χιλιάδες δρόμους και όπως αναφέραμε το σύνολο δεδομένων είναι 11gb οπότε θα περιοριστούμε σε 1 μόνο δρόμο. Στην θεωρία κυκλοφορίας, με τον όρο χρονική άθροιση (temporal aggregation) συνήθως εννοούμε τον μέσο όρο μία κυκλοφοριακής μεταβλητής σε μεγάλα σταθερού μήκους χρονικά διαστήματα, π.χ. 5λεπτα, 15λεπτα, ημέρες, μήνες κλπ. Όπως αναφέρεται και στην εργασία των *Vlahogianni et al.* [54] η χρονική άθροιση αποτελεί σημαντική παράμετρο για την πρόβλεψη κυκλοφορίας και μπορεί να οδηγήσει στην εξάλειψη των χαρακτηριστικών μακροπρόθεσμης μνήμης και της ετερογένειας της διασποράς. Για τον δρόμο επιλογής θα υπολογιστεί η μέση ταχύτητα διαδρομής ανά 5λεπτο, δηλαδή τα μοντέλα που θα αναπτυχθούν είναι scalar – based (μονομεταβλητή κυκλοφοριακή χρονοσειρά). Η επιλογή στηρίχθηκε στην βιβλιογραφία καθώς πολλά πρόσφατα άρθρα επιλέγουν αυτή την ανάλυση στα δεδομένα τους. Παρόλα αυτά η κατάλληλη επιλογή χρονικής άθροισης χρήζει βελτιστοποίησης και αφήνεται για περαιτέρω έρευνα. Παρακάτω ακολουθεί η διαδικασία που ακολουθήθηκε για την δημιουργία του συνόλου των δεδομένων. Για την υλοποίηση της διαδικασίας αυτής αναπτύχθηκε κώδικας σε προγραμματιστική γλώσσα Python.

Τα βήματα που ακολουθήσαμε είναι:

1. Για κάθε ένα από τα 61 αρχεία έγινε μετατροπή των συντεταγμένων από σύστημα αναφοράς GCJ-02 σε σύστημα αναφοράς WGS-84.
2. Οι χρονοσημάνσεις μετατράπηκαν σε τοπική ώρα Κίνας και για τα 61 αρχεία.
3. Ορίσαμε το ορθογώνιο εντός του κυκλοφοριακού δικτύου της πόλης Χί'Αν που περιλαμβάνει την υπό εξέταση αστική αρτηρία με συντεταγμένες:
  - Σημείο 1: 34.248, 108.9375
  - Σημείο 2: 34.248, 108.93645
  - Σημείο 3: 34.241, 108.93645
  - Σημείο 4: 34.241, 108.9375

Να σημειωθεί ότι το ορθογώνιο που σχηματίζεται περιλαμβάνει και τις 2 κατευθύνσεις. Για τις ανάγκες της διπλωματικής θα γίνει χρήση της μίας κατεύθυνσης και συγκεκριμένα μόνο την λωρίδα που η κατεύθυνσή της είναι βόρεια.

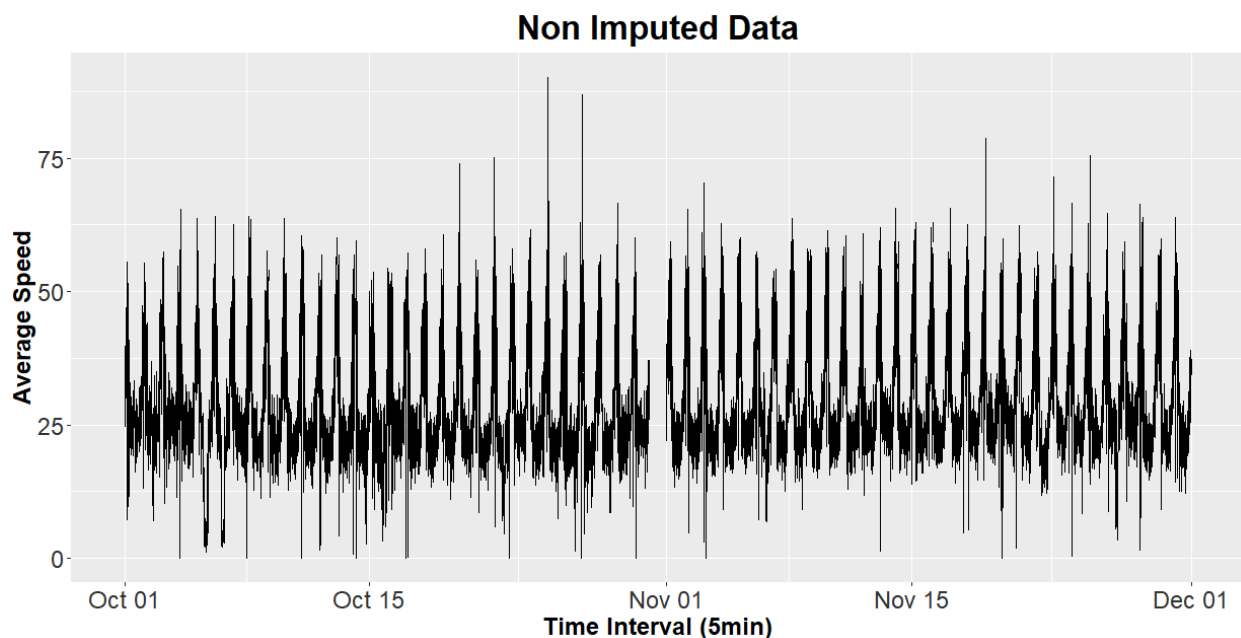
4. Διατηρήθηκαν μόνο τα σημεία που βρίσκονται εντός της περιοχής ενδιαφέροντος (εντός του ορθογωνίου που ορίζεται από τα 4 σημεία του (3)). Συνολικά προέκυψαν περίπου 13,5 εκατομμύρια σημεία για τους μήνες Οκτώβριο και Νοέμβριο (61 μέρες αρχεία).
5. Ομαδοποιούνται τα σημεία ενδιαφέροντος ανά driver id, order id και timestamp. Έπειτα ομαδοποιήθηκαν σε 17568 5λεπτα που αντιστοιχεί στην διαμέριση των 61 ημερών σε συνεχόμενα 5λεπτα.
6. Για κάθε ένα 5λεπτο υπολογίζουμε ποια σημεία ενδιαφέροντος βρίσκεται η χρονοσήμανση τους εντός του χρονικού του εύρους του 5λέπτου.
7. Για κάθε 5λεπτο και για κάθε σημείο εντός του 5λεπτου και ανά οδηγό και ταξίδι υπολογίζουμε για τα συνεχόμενα σημεία την ταχύτητα του οδηγού από τον τύπο  $u = \frac{\delta x}{\delta t}$ , όπου  $\delta x$  είναι η γεωδαισική απόσταση των 2 συνεχόμενων σημείων και  $\delta t$  η διαφορά των χρονοσημάνσεων των σημείων.
8. Υπολογίζουμε σε κάθε διάστημα 5 λεπτών την μέση ταχύτητα του οδηγού και έπειτα την μέση ταχύτητα όλων των οδηγών που είχαν κομμάτι ταξιδιού εντός του κάθε διαστήματος.

Εκτελώντας την παραπάνω διαδικασία προκύπτει το σύνολο δεδομένων που έχει μορφή διανύσματος, μήκους 17568 που ισούται με τον αριθμό των 5λέπτων σε διάστημα 61 ημερών,

όπου κάθε «κελί» έχει ως τιμή την μέση ταχύτητα διαδρομής των οδηγών του αντίστοιχου 5λέπτου. Το σύνολο δεδομένων που δημιουργήθηκε περιέχει ελλείπουσες τιμές (missing data) καθώς και ακραίες τιμές με τα οποία θα ασχοληθούμε στην επόμενη υποενότητα.

### 3.4 ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Για την οπτικοποίηση των δεδομένων χρησιμοποιήθηκε η γλώσσα προγραμματισμού R και το εξαιρετικό πακέτο οπτικοποίησης ggplot2 [55]. Στο Διάγραμμα 1 φαίνεται το σύνολο δεδομένων μετά την διαδικασία που περιγράφηκε στην υποενότητα 3.3



**Διάγραμμα 1** Το σύνολο των πραγματικών δεδομένων (ανά 5 λεπτά).

#### 3.4.1 Οι Συνιστώσες μίας Χρονοσειράς – Ανάλυση Αυτοσυσχέτισης και Μερικής Αυτοσυσχέτισης

Γενικά μία χρονοσειρά αποτελείται από τέσσερις κύριες συνιστώσες οι οποίες μπορούν να διαχωριστούν από τα παρατηρούμενα δεδομένα. Αυτές οι συνιστώσες είναι: η συνιστώσα τάσης (trend), η κυκλική (cycle), η εποχική (seasonality) και η τυχαία συνιστώσα (irregular) [56].

Η γενική τάση μίας χρονοσειράς να αυξάνεται, να μειώνεται ή να μένει σταθερή για ένα μεγάλο χρονικό διάστημα ονομάζεται τάση (trend). Η συνιστώσα τάσης μπορεί να αντιμετωπισθεί ως η μακροπρόθεσμη κίνηση μίας χρονοσειράς. Για παράδειγμα, μία χρονοσειρά που σχετίζεται με την αύξηση του πληθυσμού, των αριθμό των σπιτιών σε μία πόλη κλπ. θα έχει ανοδική τάση, ενώ καθοδική τάση μπορεί να παρατηρηθεί σε χρονοσειρές που σχετίζονται με θνησιμότητα, επιδημίες κλπ. [56].

Οι εποχικές μεταβολές σε μία χρονοσειρά είναι οι διακυμάνσεις που παρουσιάζονται εντός ενός χρόνου κατά την διάρκεια μίας εποχής. Αυτή η γενική περιγραφή της εποχικότητας, έχει διαφορετική ερμηνεία στο πρόβλημα της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας όσον αφορά στον χρόνο και στην εποχή. Κάποιοι σημαντικοί παράγοντες που προκαλούν αυτές τις εποχικές διακυμάνσεις είναι: το κλίμα, οι καιρικές συνθήκες, συνήθειες κλπ. Για παράδειγμα, οι πωλήσεις παγωτού αυξάνονται το καλοκαίρι, ενώ οι πωλήσεις των μάλλινων ρούχων αυξάνονται τον χειμώνα [56].

Η κυκλική συνιστώσα σε μία χρονοσειρά περιγράφει τις μεσοπρόθεσμες αλλαγές τις χρονοσειράς οι οποίες προκαλούνται από κάποιες περιστάσεις, οι οποίες επαναλαμβάνονται σε κύκλους. Η διάρκεια ενός κύκλου εκτείνεται για μεγαλύτερη χρονική διάρκεια, συνήθως 2 ή περισσότερα χρόνια [56].

Η τυχαία συνιστώσα σε μία χρονοσειρά προκαλείται από μη προβλέψιμες επιρροές, οι οποίες δεν είναι κανονικές και επίσης δε επαναλαμβάνονται με κάποιο συγκεκριμένο μοτίβο. Αυτές οι διακυμάνσεις μπορούν να προκληθούν από διάφορα συμβάντα όπως πόλεμος, σεισμός, πλημμύρες κα. Δεν έχει βρεθεί κάποια στατιστική μέθοδος που να μπορεί να «μετρήσει» αυτές τις τυχαίες διακυμάνσεις σε μία χρονοσειρά [56].

Παρατηρώντας το Διάγραμμα 1 μπορούμε να διακρίνουμε τα εξής. Αρχικά, φαίνεται να υπάρχει κάποια επαναληψιμότητα στα δεδομένα δηλαδή παρατηρούμε ότι με κάποια συχνότητα την οποία θα ανακαλύψουμε αργότερα περνάμε συνεχώς από υψηλές τιμές της μέσης ταχύτητας σε χαμηλές τιμές και ακολούθως πάλι σε υψηλές κοκ. Άρα τα δεδομένα μας έχουν εποχική συνιστώσα. Επιπροσθέτως, δεν φαίνεται εκ πρώτης όψεως να υπάρχει κάποια τάση στα δεδομένα (trend), φαίνεται να έχουν σχετικά σταθερή διασπορά. Τέλος, έντονη είναι και η παρουσία των ελλειπουσών τιμών (missing data).

Η στατιστική συσχέτιση συνοψίζει το πόσο ισχυρή είναι η σχέση μεταξύ 2 μεταβλητών. Μπορούμε να υποθέσουμε ότι η κατανομή κάθε μεταβλητής είναι κανονική. Εάν ισχύει αυτό, μπορεί να γίνει χρήση του συντελεστή συσχέτισης του Pearson ο οποίος υποδεικνύει αν δύο μεταβλητές είναι συσχετισμένες και πόσο. Οι τιμές που μπορεί να πάρει ο συντελεστής συσχέτισης του Pearson είναι μεταξύ -1 και 1. Η τιμή 0 υποδηλώνει ότι δεν υπάρχει συσχέτιση μεταξύ των 2 μεταβλητών. Όσον αφορά δεδομένα χρονοσειρών μπορεί να υπολογιστεί η συσχέτιση παρατηρήσεων με παρατηρήσεις από προηγούμενες χρονικές στιγμές, τα οποία ονομάζονται υστερήσεις. Επειδή η συσχέτιση των παρατηρήσεων μίας χρονοσειράς υπολογίζεται με παρατηρήσεις της ίδιας χρονοσειράς σε προηγούμενες χρονικές στιγμές, ονομάζεται αυτοσυσχέτιση. Έτσι, όπως ακριβώς η συσχέτιση μετρά το κατά πόσο υφίσταται γραμμική σχέση μεταξύ δύο μεταβλητών, η αυτοσυσχέτιση μετρά την γραμμική σχέση μεταξύ τιμών υστέρησης μίας χρονοσειράς η οποία συμβολίζεται με  $r_k$ , και γράφεται ως

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (3.2)$$

όπου T είναι το μήκος της χρονοσειράς (αριθμός παρατηρήσεων).

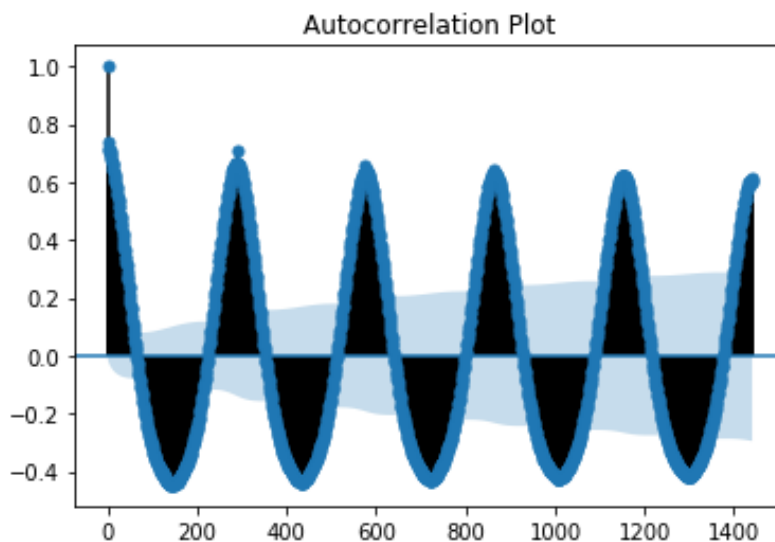
Όταν τα δεδομένα έχουν τάση, οι αυτοσυσχετίσεις για μικρές υστερήσεις τείνουν να είναι μεγάλες και θετικές διότι οι κοντινές παρατηρήσεις χρονικά είναι κοντινές και σε «απόσταση». Έτσι η συνάρτηση αυτοσυσχέτισης σε χρονοσειρές που έχουν τάση τείνει να έχει θετικές τιμές οι οποίες φθίνουν αργά όσο η υστέρηση αυξάνεται [57].

Όταν τα δεδομένα έχουν εποχική συνιστώσα, οι αυτοσυσχετίσεις θα είναι μεγαλύτερες για τις υστερήσεις εποχικότητας (δηλαδή τα πολλαπλάσια της εποχικής συχνότητας) σε σχέση με τις άλλες υστερήσεις. Εάν τα δεδομένα έχουν και συνιστώσα τάσης και συνιστώσα εποχικότητας τότε μπορούμε να δούμε τον συνδυασμό των επιδράσεων [57].

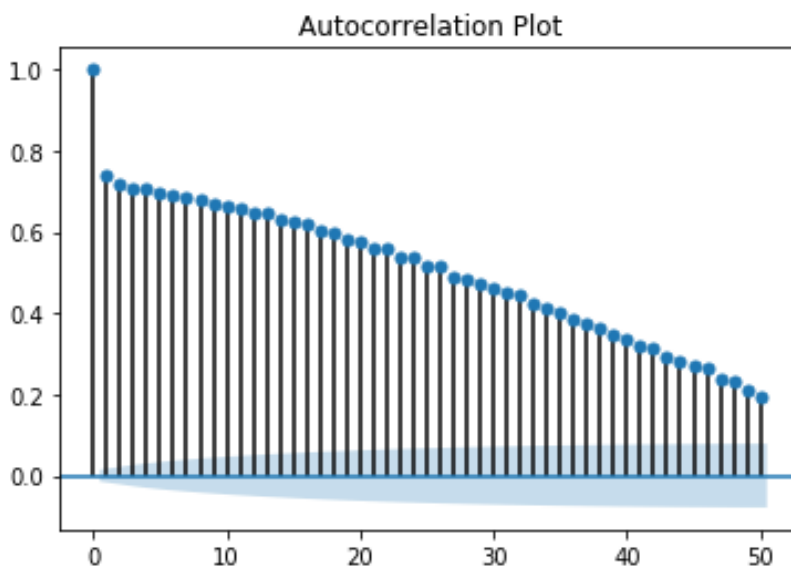
Στο Διάγραμμα 2 φαίνεται η συνάρτηση αυτοσυσχέτισης του συνόλου δεδομένων για υστερήσεις έως 1440. Όπως μπορούμε να παρατηρήσουμε για πολλαπλάσια του 288 εμφανίζονται επαναλαμβανόμενα οι μέγιστες τιμές της συνάρτησης αυτοσυσχέτισης άρα η εποχική συχνότητα των δεδομένων είναι η υστέρηση με τιμή 288. Αυτό σημαίνει ότι υπάρχει επαναλαμβανόμενο μοτίβο κάθε 288 5λεπτα, δηλαδή κάθε μία μέρα. Στο Διάγραμμα 3 φαίνεται πάλι η συνάρτηση αυτοσυσχέτισης των δεδομένων αλλά για υστέρηση ίση με 50 όπου φαίνονται καλύτερα οι



αυτοσυσχετίσεις για συνεχόμενες υστερήσεις. Τα διαστήματα εμπιστοσύνης που έχουν οριστεί στο 95% σχηματίζονται σαν ένας κώνος. Ουσιαστικά, οι αυτοσυσχετίσεις που «πέφτουν» έξω από τον κώνο θεωρούνται στατιστικά σημαντικές και δεν οφείλονται στην τυχαιότητα.



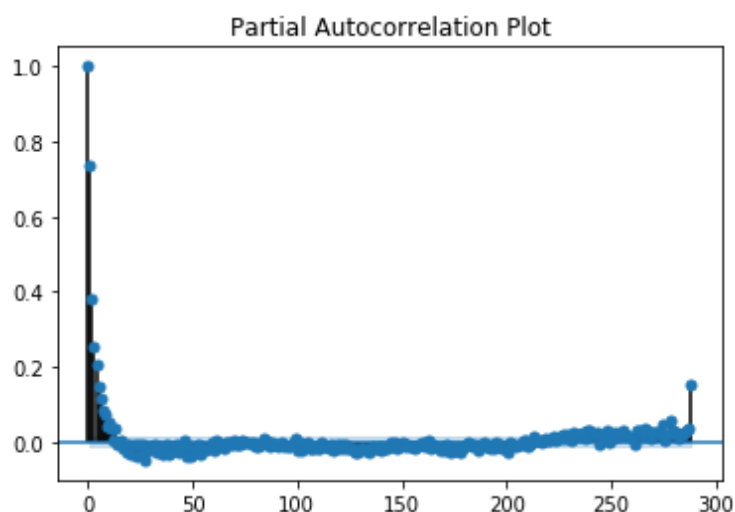
**Διάγραμμα 2** Το Διάγραμμα αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 1440.



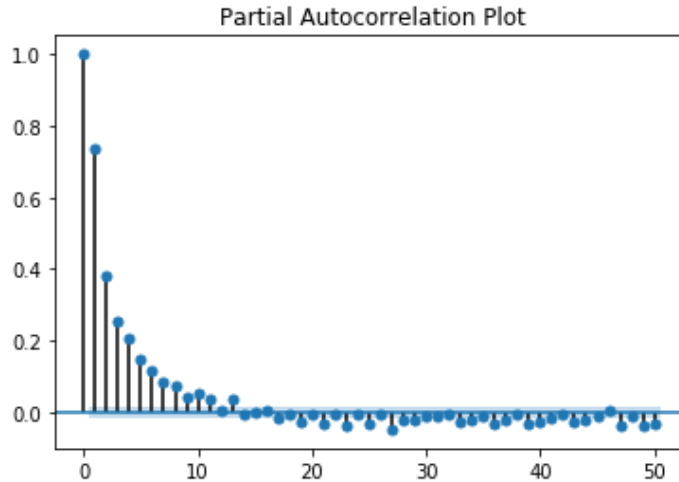
**Διάγραμμα 3** Το Διάγραμμα αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 50.

Η μερική αυτοσυσχέτιση είναι μία περιγραφή της σχέσης μεταξύ μίας παρατήρησης μίας χρονοσειράς με παρατηρήσεις σε προηγούμενες χρονικές στιγμές με τις συσχετίσεις των ενδιάμεσων παρατηρήσεων να έχουν αφαιρεθεί. Ουσιαστικά, η μερική αυτοσυσχέτιση για υστέρηση  $k$  είναι η αυτοσυσχέτιση που προκύπτει από την αφαίρεση των αυτοσυσχετίσεων λόγω των ενδιάμεσων όρων των μικρότερων υστερήσεων [58]. Η αυτοσυσχέτιση μεταξύ μίας παρατήρησης και της παρατήρησης σε προηγούμενη χρονική στιγμή αποτελείται από την άμεση συσχέτιση των 2 παρατηρήσεων καθώς και τις έμμεσες συσχετίσεις των ενδιάμεσων υστερήσεων. Αυτές οι έμμεσες συσχετίσεις είναι ένας γραμμικός συνδυασμός της συσχέτισης της παρατήρησης με όλες τις ενδιάμεσες παρατηρήσεις χρονικά. Αυτές οι έμμεσες συσχετίσεις είναι που αφαιρούνται στην μερική αυτοσυσχέτιση.

Στο Διάγραμμα 4 φαίνεται το Διάγραμμα μερικής αυτοσυσχέτισης του συνόλου δεδομένων για υστέρηση έως 288. Παρατηρείται ότι σταδιακά οι αυτοσυσχετίσεις μειώνονται ενώ για υστέρηση ίση με 288 παρατηρείται μεγάλη τιμή της μερικής αυτοσυσχέτισης γεγονός που ενισχύει την ύπαρξη της εποχικότητας με συχνότητα ίση με 288. Στο Διάγραμμα 5 φαίνεται το Διάγραμμα μερικής αυτοσυσχέτισης του πραγματικού συνόλου δεδομένων για τιμή υστέρησης έως 50 όπου φαίνεται πιο καθαρά το μοτίβο της μερικής αυτοσυσχέτισης.



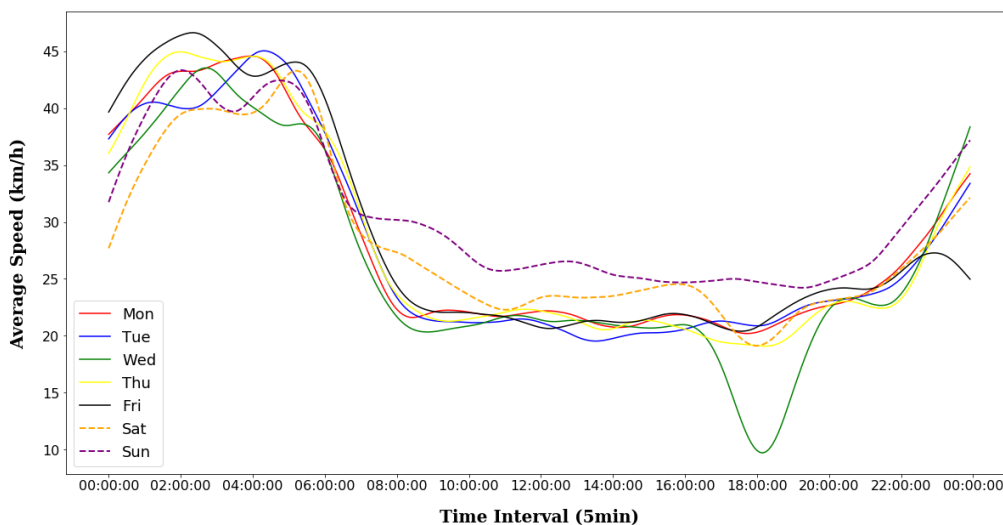
**Διάγραμμα 4** Το Διάγραμμα μερικής αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 288.



**Διάγραμμα 5** Το Διάγραμμα μερικής αυτοσυσχέτισης του συνόλου δεδομένων για τιμή υστέρησης έως 50.

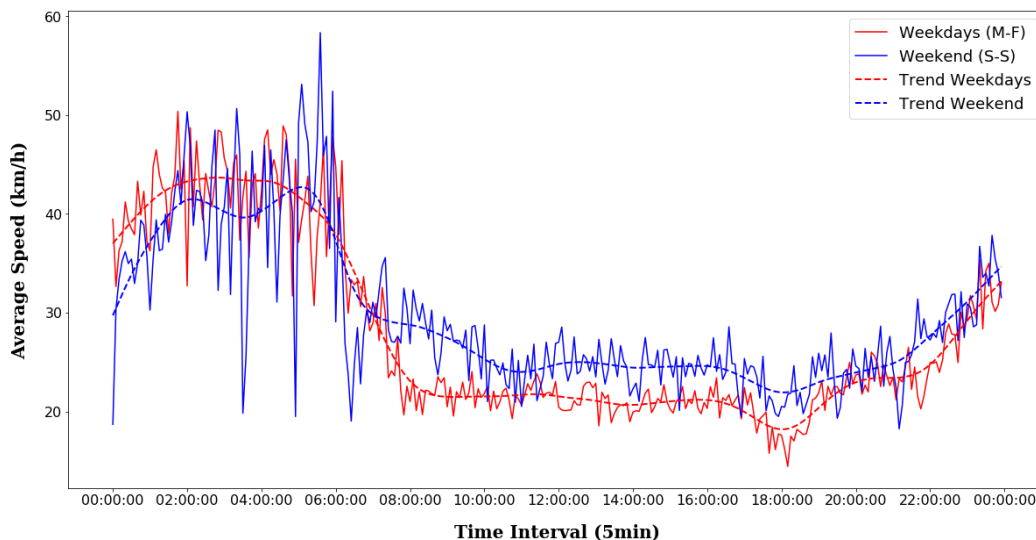
### 3.4.2 Αναλύοντας σε βάθος το σύνολο δεδομένων

Από τα παραπάνω προκύπτει ότι εμφανίζονται επαναλαμβανόμενα μοτίβα ανά ημέρα όποτε θα γίνει απεικόνιση της μέσης ταχύτητα ανά ημέρα (μ.ο. ημέρας) αφού εφαρμοστεί ομαλοποίηση κινητού μέσου (moving average smoothing) ανά ημέρα στο σύνολο δεδομένων. Όπως φαίνεται στο Διάγραμμα 6 επιβεβαιώνεται η ύπαρξη των επαναλαμβανόμενων μοτίβων δηλαδή παρατηρείται ότι για κάθε μέρα τις βραδινές ώρες έχουμε υψηλότερες ταχύτητες οι οποίες μειώνονται σταδιακά και παίρνουν την χαμηλότερη τιμή τους κατά τις ώρες αιχμής. Επιπροσθέτως, τις ημέρες του σαββατοκύριακου στην διάρκεια των ωρών αιχμής παρατηρούνται υψηλότερες ταχύτητες.



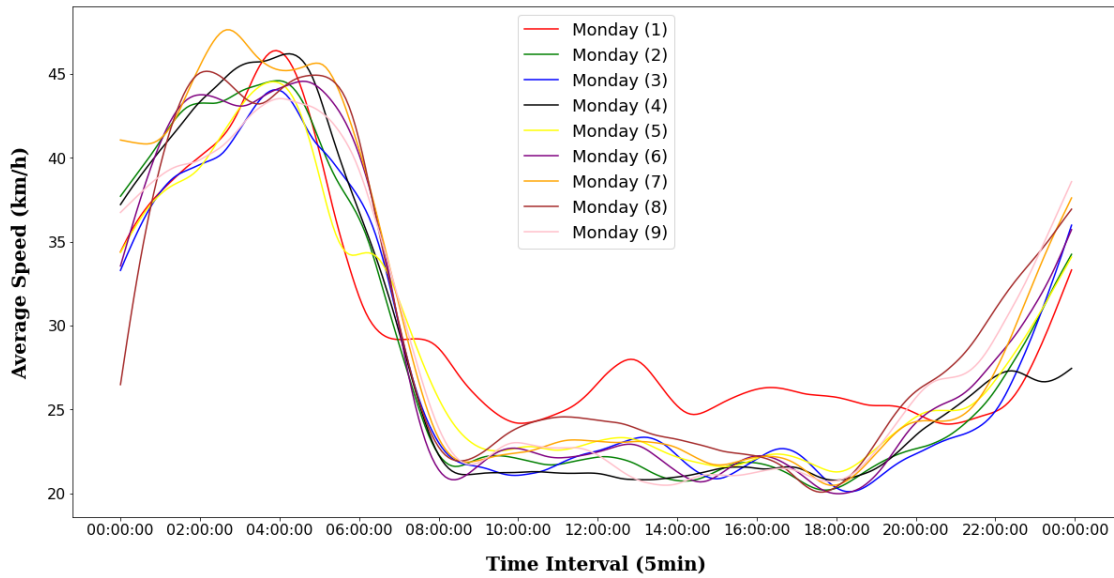
**Διάγραμμα 6** Μέση ταχύτητα ανά ημέρα με ομαλοποίηση κινητού μέσου.

Στο Διάγραμμα 7 έχει γίνει ομαδοποίηση τις ημέρες της εβδομάδας σε καθημερινές (Δευτέρα έως Παρασκευή) και σε ημέρες σαββατοκύριακου (Σάββατο και Κυριακή). Για τις δύο ομάδες έχει υπολογιστεί η μέση τιμή της μέσης ταχύτητας ανά 5λεπτο και επιπλέον έχει εφαρμοσθεί και πάλι ομαλοποίηση κινητού μέσου στις τιμές αυτές. Για τις ημέρες εντός του σαββατοκύριακου παρατηρείται λίγο υψηλότερη μέση ταχύτητα μετά τις 7:00 πμ.

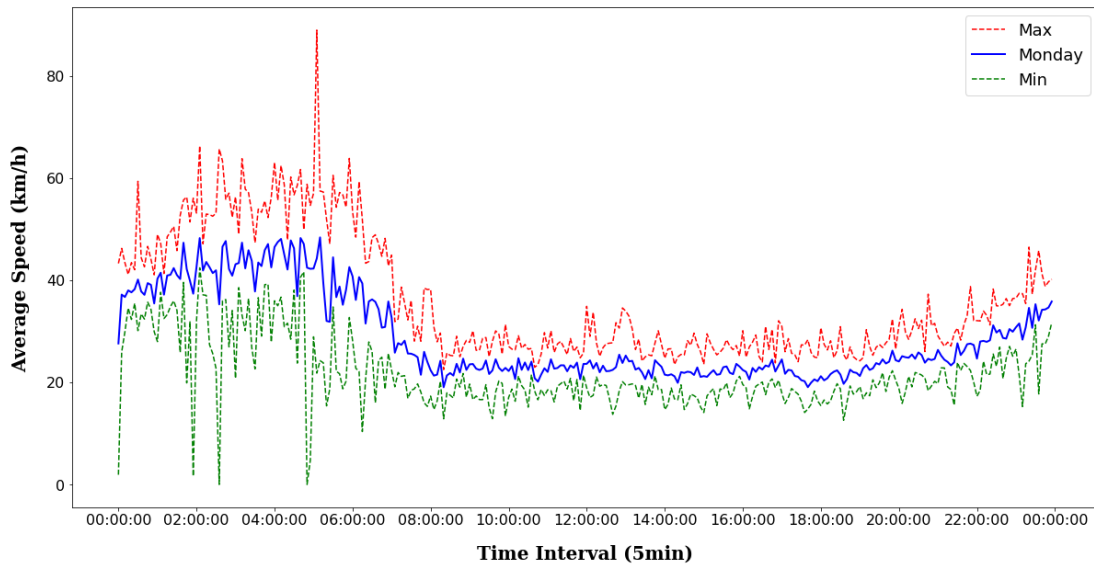


**Διάγραμμα 7** Μέση ταχύτητα για καθημερινές και σαββατοκύριακο (μέσοι όροι και ομαλοποιημένες τιμές κινητού μέσου).

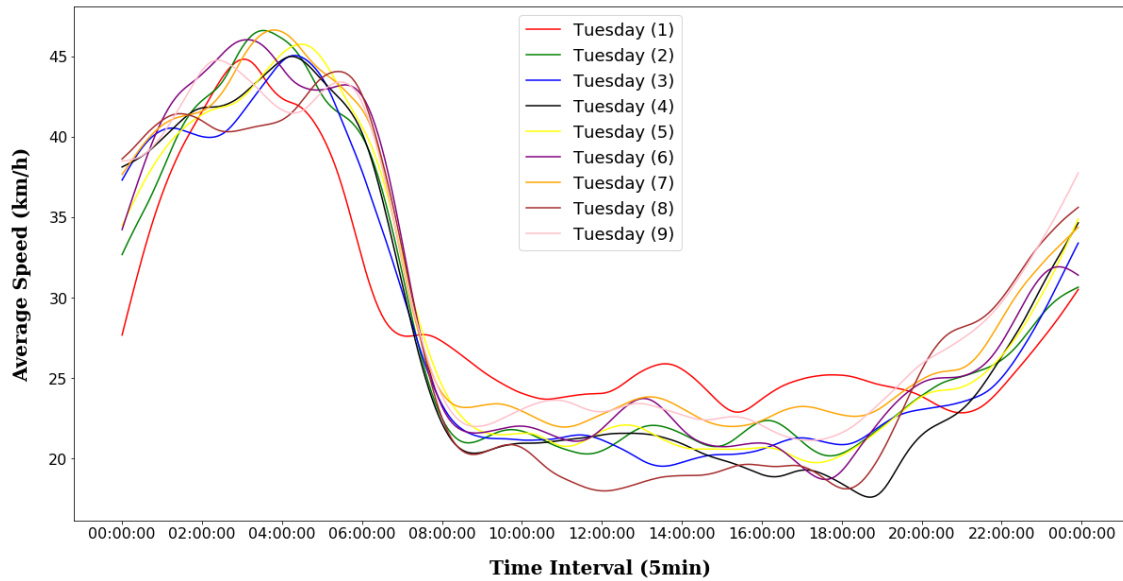
Στα επόμενα Διαγράμματα (8-21) φαίνεται η ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο ανά ημέρα (Δευτέρα – Κυριακή) για όλες τις μέρες του συνόλου δεδομένων και η μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο ανά ημέρα (Δευτέρα – Κυριακή) για όλες τις μέρες του συνόλου δεδομένων. Γενικά, παρατηρείται ότι την 1<sup>η</sup> Τετάρτη και Πέμπτη στο σύνολο δεδομένων τις ώρες 11:00 – 20:00 είναι πολύ χαμηλές, χαμηλότερες σε σχέση με τις υπόλοιπες «αντίστοιχες» μέρες (υπόλοιπες Τετάρτες και Πέμπτες). Αναζητώντας στο google για τις συγκεκριμένες ημερομηνίες, βρέθηκε ότι υπήρχαν εορτασμοί στην πόλη Xi'an και για τις 2 ημέρες. Τέλος, για κάποιες μέρες παρατηρείται πτώση της μέσης ταχύτητας τις απογευματινές ώρες αιχμής (17:00 – 20:00) και ότι υπάρχουν κάποιες μηδενικές ταχύτητες όπως και κάποιες πολύ υψηλές ταχύτητες για αστική αρτηρία. Με την αντιμετώπιση αυτών των τιμών θα ασχοληθούμε στην επόμενη υποενότητα.



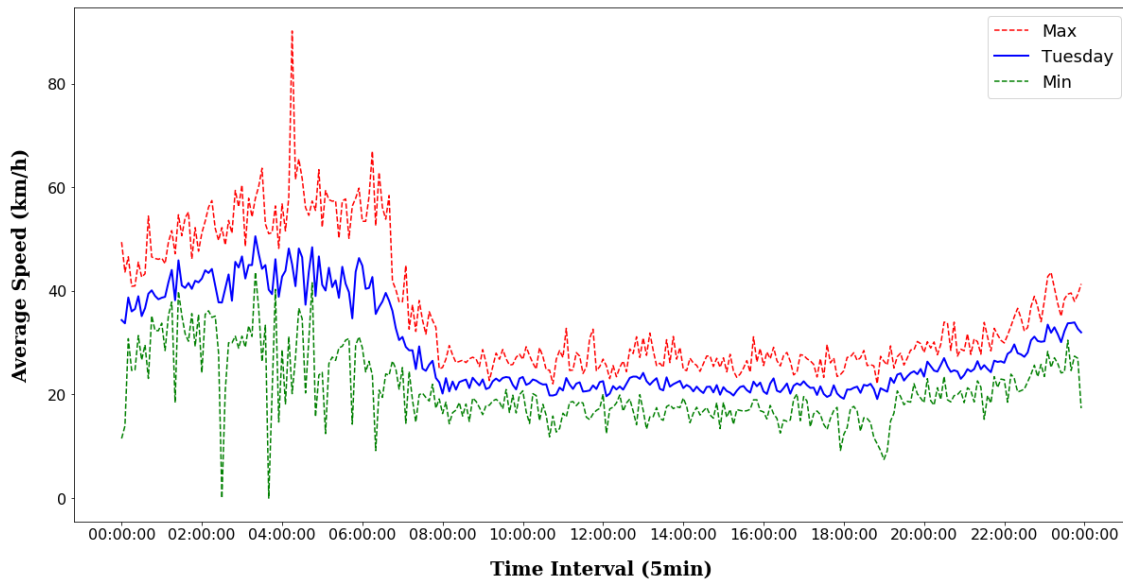
**Διάγραμμα 8** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Δευτέρες του συνόλου δεδομένων.



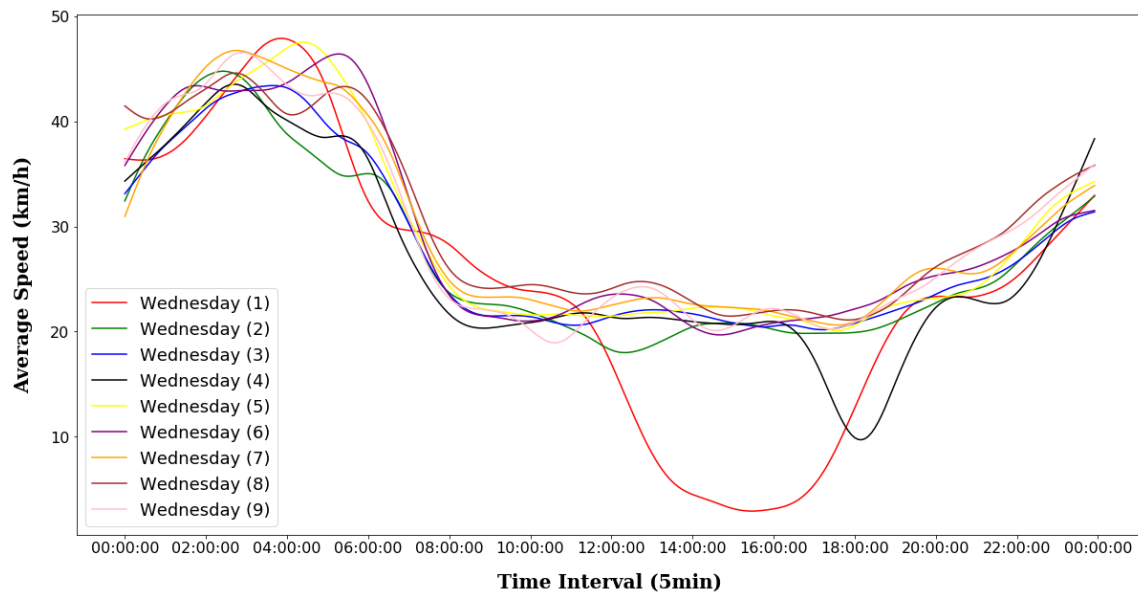
**Διάγραμμα 9** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Δευτέρες του συνόλου δεδομένων.



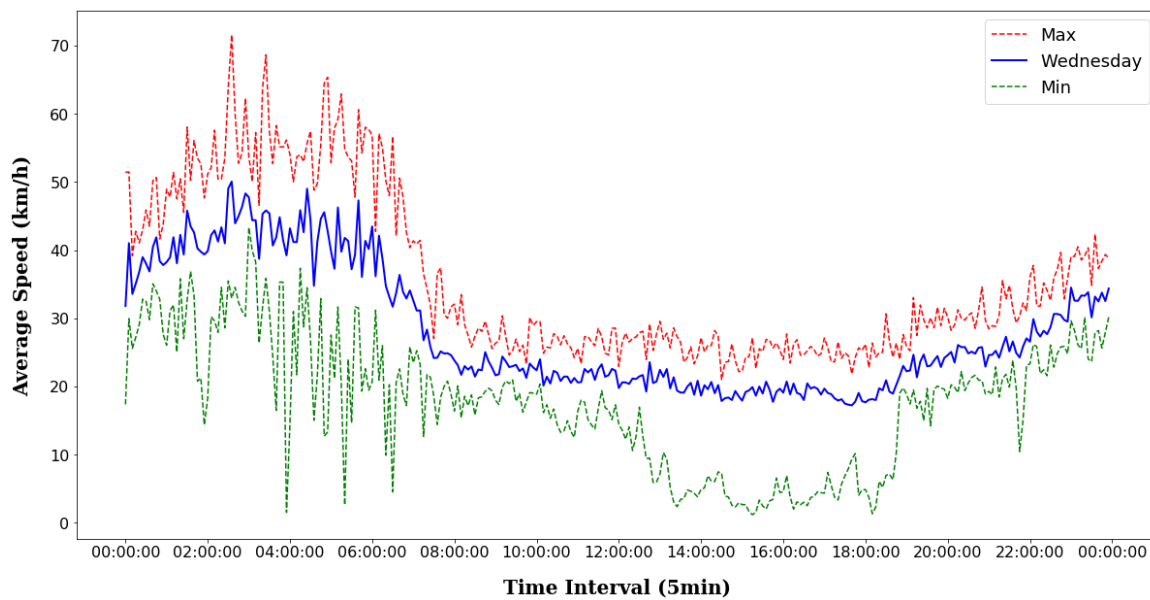
**Διάγραμμα 10** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Τρίτες του συνόλου δεδομένων.



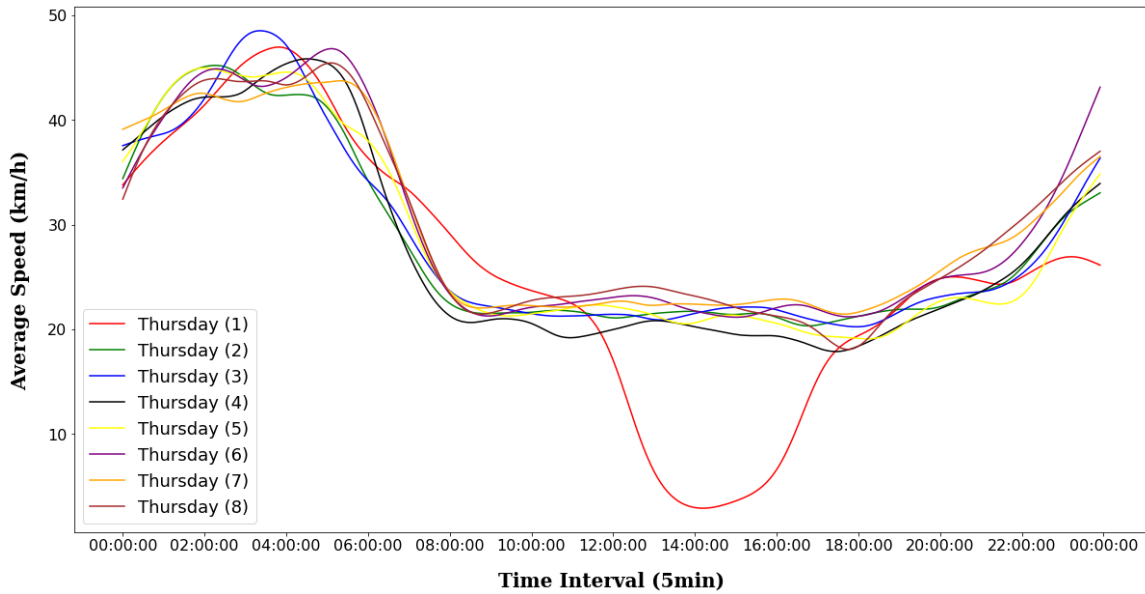
**Διάγραμμα 11** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Τρίτες του συνόλου δεδομένων.



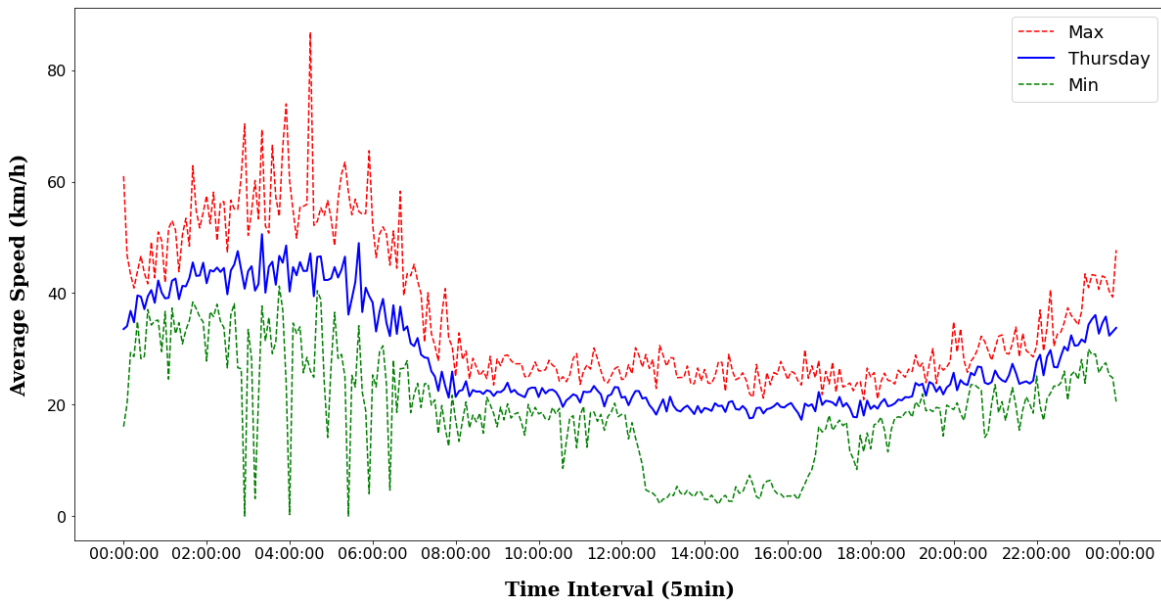
**Διάγραμμα 12** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Τετάρτες του συνόλου δεδομένων.



**Διάγραμμα 13** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Τετάρτες του συνόλου δεδομένων.

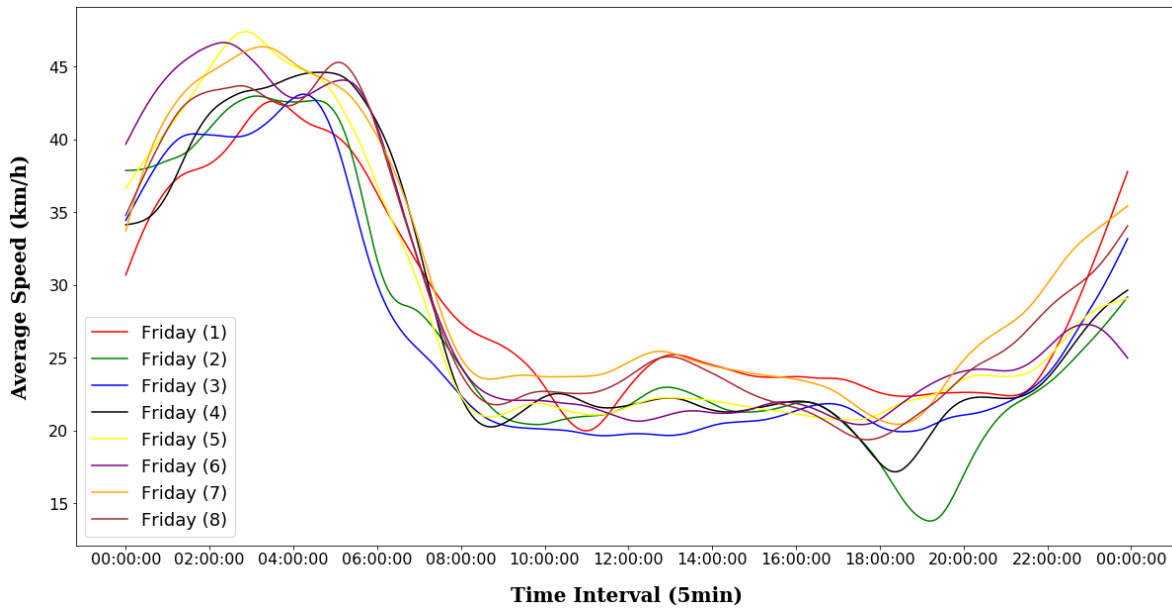


**Διάγραμμα 14** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Πέμπτες του συνόλου δεδομένων.

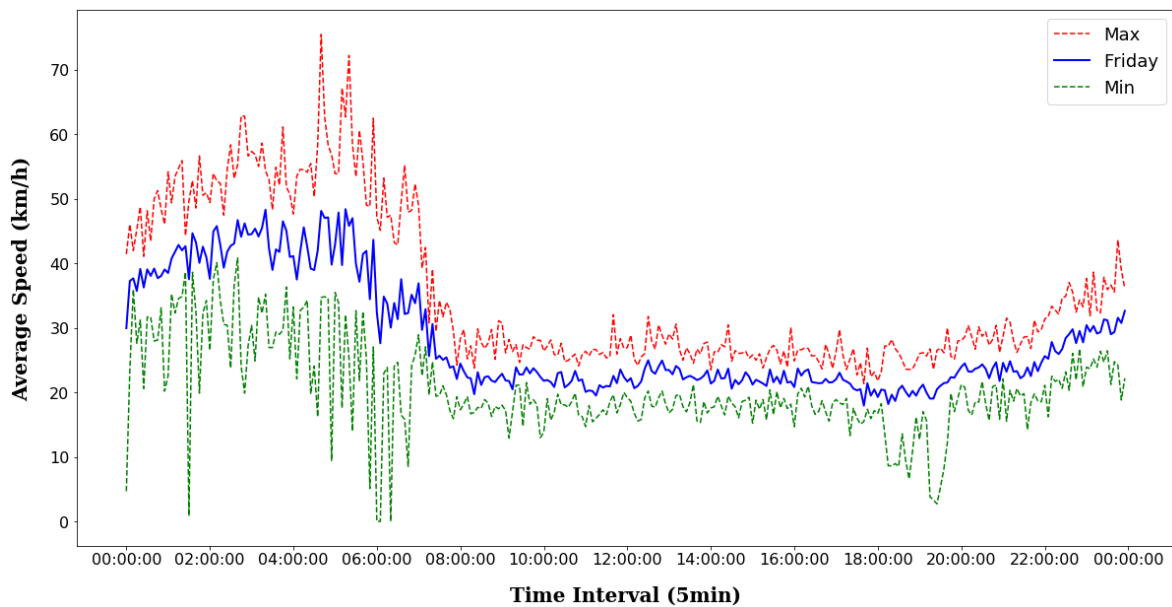


**Διάγραμμα 15** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Πέμπτες του συνόλου δεδομένων.

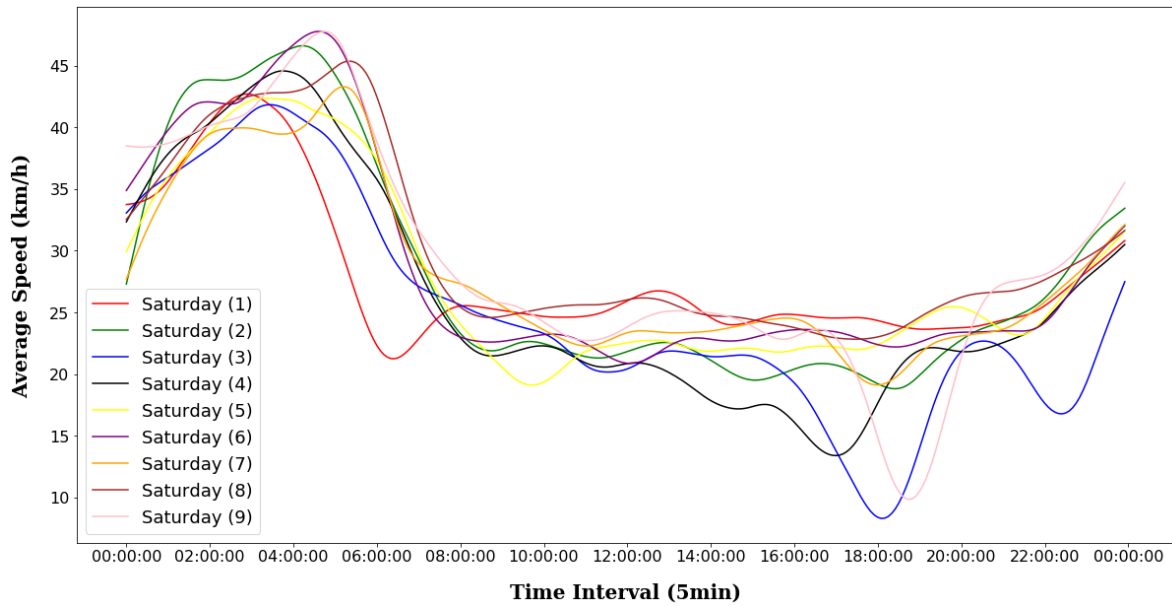




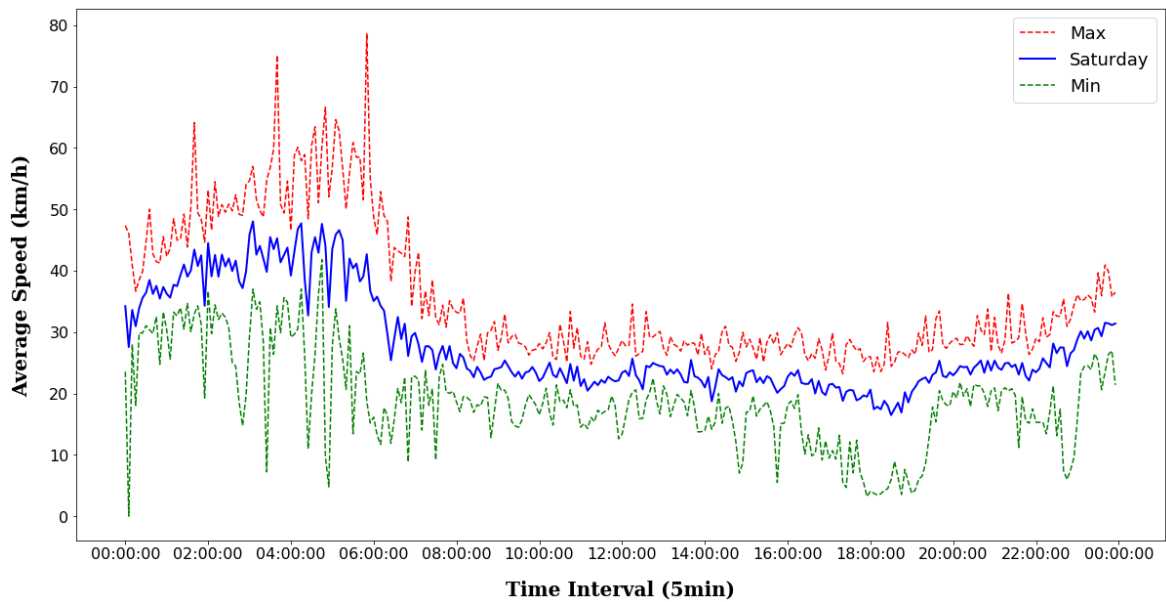
**Διάγραμμα 16** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλες τις Παρασκευές του συνόλου δεδομένων.



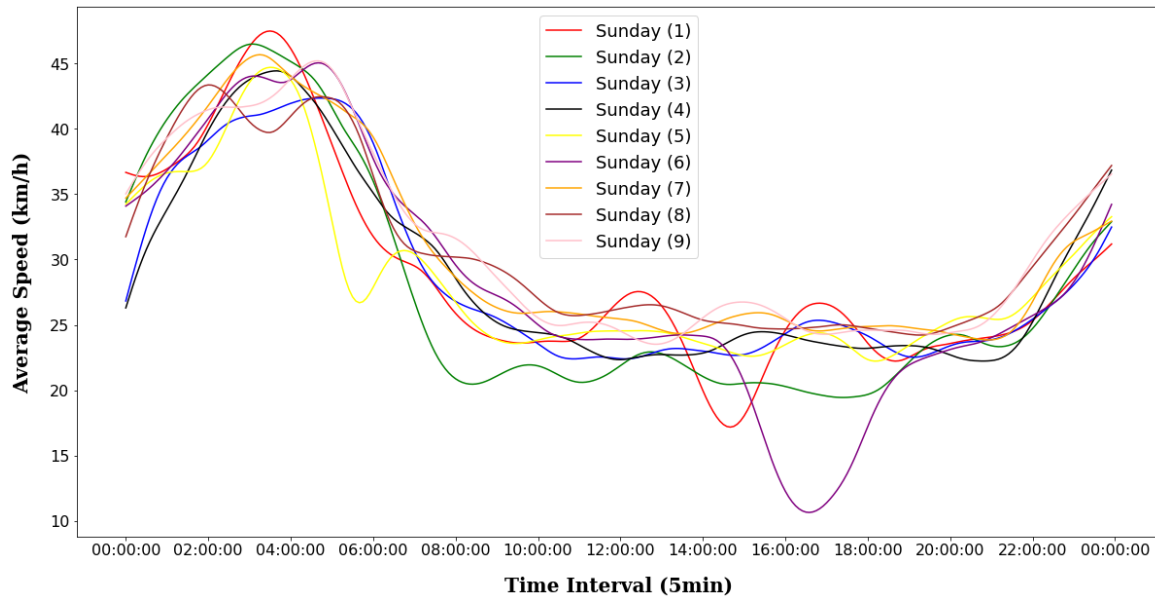
**Διάγραμμα 17** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Παρασκευές του συνόλου δεδομένων.



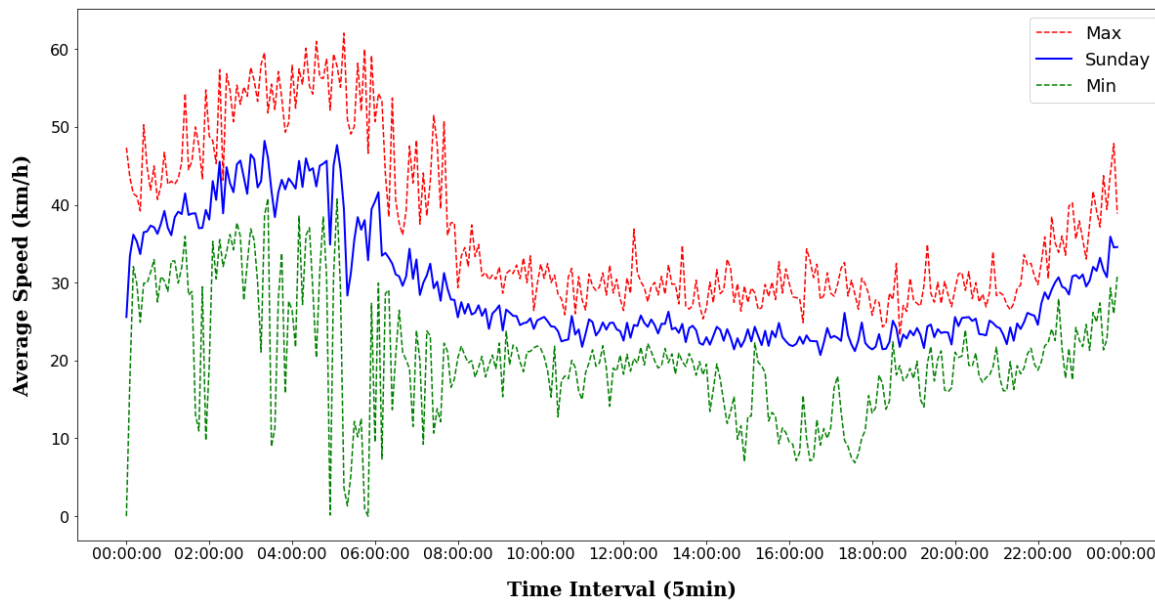
**Διάγραμμα 18** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλα τα Σάββατα του συνόλου δεδομένων.



**Διάγραμμα 19** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τα Σάββατα του συνόλου δεδομένων.



**Διάγραμμα 20** Ομαλοποιημένη (κινητού μέσου) μέση ταχύτητα ανά 5λεπτο για όλα τις Κυριακές του συνόλου δεδομένων.



**Διάγραμμα 21** Μέση, μέγιστη και ελάχιστη μέση ταχύτητα ανά 5λεπτο για όλες τις Κυριακές του συνόλου δεδομένων.

### 3.5 ΣΤΑΣΙΜΟΤΗΤΑ

Μία διαδικασία  $\{x(t), t = 0, 1, 2, \dots\}$  ισχυρά στάσιμη (strongly stationary) ή αυστηρά στάσιμη (strictly stationary) εάν η από κοινού συνάρτηση κατανομής πιθανότητας των  $\{x_{t-s}, x_{t-s+1}, \dots, x_t, \dots, x_{t+s-1}, x_{t+s}\}$  είναι ανεξάρτητη των  $t$  και  $s$ . Έτσι, σε μία ισχυρά στάσιμη χρονοσειρά η από κοινού κατανομή οποιουδήποτε υποσυνόλου τυχαίων μεταβλητών της διαδικασίας είναι ανεξάρτητες του χρόνου.

Σε πολλές πρακτικές εφαρμογές, η υπόθεση της ισχυρής στασιμότητας δεν είναι πάντα απαραίτητη και έτσι απαιτείται μία πιο ασθενή μορφή της. Μία στοχαστική διαδικασία θα λέμε ότι είναι ασθενώς στάσιμη τάξεως  $k$  εάν οι στατιστικές ροπές της μέχρι και την τάξη  $k$  εξαρτώνται μόνο από την διαφορά του χρόνου μεταξύ 2 συμβάντων και όχι από την χρονική στιγμή των συμβάντων των δεδομένων που χρησιμοποιούνται για την εκτίμηση τους. Για παράδειγμα, μία στοχαστική διαδικασία  $\{x(t), t = 0, 1, 2, \dots\}$  είναι 2<sup>ης</sup> τάξης στάσιμη εάν η μέση τιμή και η διασπορά της είναι ανεξάρτητες του χρόνου και η συνδιασπορά  $Cov(x_t, x_{t-s})$  εξαρτάται μόνο από το  $s$ . Είναι σημαντικό να τονιστεί σε αυτό το σημείο ότι ούτε η ισχυρή ούτε η ασθενής στασιμότητα υποδηλώνει η μία την άλλη. Παρόλα αυτά, μία ασθενώς στάσιμη διαδικασία η οποία ακολουθεί την κανονική κατανομή είναι επίσης ισχυρά στάσιμη.

Μία χρονοσειρά η οποία έχει συνιστώσα τάσης (δηλαδή χρονικά εξαρτώμενο μέσο) και εποχική συνιστώσα δεν είναι στάσιμη. Έχει ήδη σχολιαστεί ότι η χρονοσειρά του πραγματικού συνόλου δεδομένων έχει εποχική συνιστώσα. Δύο πολύ γνωστοί έλεγχοι οι οποίοι εξετάζουν αν η χρονοσειρά είναι στάσιμη ως προς την τάση δηλαδή ως προς τον μέσο είναι του Dickey και Fuller [59] και των Kwiatkowski–Phillips–Schmidt–Shin (KPSS) [60].

- Επαυξημένος έλεγχος Dickey-Fuller:

Εξετάζει την μηδενική υπόθεση ύπαρξης μοναδιαίας ρίζας στο δείγμα μίας χρονοσειράς. Η εναλλακτική υπόθεση είναι συνήθως στασιμότητα ή στασιμότητα λόγω τάσης. Ουσιαστικά, πρόκειται για μία επαυξημένη εκδοχή του κλασσικού ελέγχου Dickey-Fuller για μεγαλύτερου μεγέθους και πιο πολύπλοκες χρονοσειρές. Το στατιστικό του επαυξημένου ελέγχου Dickey-Fuller που χρησιμοποιείται για τον έλεγχο της μοναδιαίας ρίζας είναι αρνητικός αριθμός. Όσο μικρότερη τιμή («πιο» αρνητική) παίρνει τόσο ισχυρότερη είναι η απόρριψη της υπόθεσης ότι υπάρχει μοναδιαία ρίζα (μηδενικός έλεγχος) για κάποιο δοσμένο επίπεδο εμπιστοσύνης.

- Έλεγχος KPSS:

Χρησιμοποιείται για τον έλεγχο της μηδενικής υπόθεσης ότι μία χρονοσειρά είναι στάσιμη γύρω από μία αιτιοκρατική τάση (π.χ. στάσιμη όσον αφορά της τάση). Σε αντίθεση με τους περισσότερους ελέγχους μοναδιαίας ρίζας, η παρουσία της μοναδιαίας ρίζας δεν είναι η μηδενική υπόθεση αλλά η εναλλακτική. Επιπροσθέτως, στον έλεγχο KPSS, η απουσία μοναδιαίας ρίζας δεν είναι απόδειξη ότι η σειρά είναι στάσιμη αλλά από κατασκευής του ελέγχου ότι είναι στάσιμη ως προς τη τάση. Αυτή είναι σημαντική διάκριση καθώς μία χρονοσειρά μπορεί να μην είναι στάσιμη, αλλά να μην έχει μοναδιαία ρίζα δηλαδή να είναι στάσιμη ως προς την τάση.

Παρακάτω ακολουθούν τα αποτελέσματα (της κονσόλας, R) και για τους 2 ελέγχους:

- Επαυξημένος έλεγχος Dickey-Fuller  $\rightarrow H_0$ : ύπαρξη μοναδιαίας ρίζας,  $H_1$ : στασιμότητα

```
ADF Statistic: -15.204514150772724
p-value: 5.76770573554268e-28
Critical Values:
  1%, -3.430723238497956
Critical Values:
  5%, -2.861704957013964
Critical Values:
 10%, -2.5668578023342805
```

Παρατηρούμε ότι το  $p - value$  είναι πολύ μικρότερο του 0.05 άρα μπορούμε να απορρίψουμε την μηδενική υπόθεση έναντι της εναλλακτικής δηλαδή η χρονοσειρά είναι στάσιμη ως προς της τάση.

- Έλεγχος KPSS  $\rightarrow H_0$ : στασιμότητα,  $H_1$ : ύπαρξη μοναδιαίας ρίζας

```
KPSS Statistic: 0.276664
p-value: 0.100000
Critical Values:
 10%, 0.347
Critical Values:
  5%, 0.463
Critical Values:
 2.5%, 0.574
Critical Values:
  1%, 0.739
```

Παρατηρούμε ότι το  $p - value$  είναι μεγαλύτερο του 0.05 άρα δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση έναντι της εναλλακτικής δηλαδή η χρονοσειρά είναι στάσιμη ως προς την

(αιτιοκρατική) τάση. Οι 2 έλεγχοι συμφωνούν ότι δεν υπάρχει τάση στα δεδομένα. Βέβαια η χρονοσειρά δεν είναι στάσιμη καθώς έχουμε δείξει ότι έχει εποχική συνιστώσα.

### **3.6 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ**

#### **3.6.1 Τι είναι η προεπεξεργασία δεδομένων;**

Η προεπεξεργασία των δεδομένων είναι ένα από τα πιο σημαντικά βήματα της διαδικασίας εξόρυξης γνώσης από δεδομένα ο οποίος είναι ο ευρύτερος κλάδος στον οποίο περιλαμβάνεται και το πρόβλημα της πρόβλεψης κυκλοφορίας. Η φράση «*garbage in, garbage out*» είναι πολύ γνωστή στον κλάδο της εξόρυξης δεδομένων και της μηχανικής μάθησης. Πολλές φορές κατά την συλλογή των δεδομένων μπορεί να καταλήξουμε με τιμές που να βρίσκονται εκτός αποδεκτών ορίων, ελλείπουσες τιμές κα. Μία διαδικασία ανάλυση δεδομένων όπου δεν έχουν διορθωθεί τα προαναφερθέντα προβλήματα μπορεί να οδηγήσει σε παραπλανητικά αποτελέσματα. Γι' αυτό το λόγο η σωστή αναπαράσταση και η ποιότητα των δεδομένων είναι πρώτιστης και ύψιστης σημασίας πριν συνεχιστεί η ανάλυση τους. Πολύ συχνά η διαδικασία προεπεξεργασίας των δεδομένων θεωρείται από τις πιο σημαντικές φάσεις ενός πρότζεκτ μηχανικής μάθησης.

Εάν υπάρχει μη σχετική και πλεοναστική πληροφορία καθώς και μη αξιόπιστα δεδομένα με θόρυβο τότε η εξόρυξη γνώσης κατά την διάρκεια της φάσης εκπαίδευσης (ή εκτίμησης των παραμέτρων) γίνεται πιο δύσκολη. Η προετοιμασία των δεδομένων και τα διάφορα βήματα φιλτραρίσματος μπορούν να πάρουν σημαντικό χρόνο. Η προεπεξεργασία των δεδομένων περιλαμβάνει:

- Καθαρισμό των δεδομένων
- Κανονικοποίηση των δεδομένων
- Μετασχηματισμός των δεδομένων
- Εξαγωγή και επιλογή χαρακτηριστικών

Το αποτέλεσμα της προεπεξεργασίας δεδομένων είναι το τελικό σύνολο δεδομένων.

#### **3.6.2 Ελλείπουσες τιμές και τρόποι αντιμετώπισης**

Όπως φάνηκε καθαρά στο Διάγραμμα 1 είναι έντονη η παρουσία των ελλειπουσών τιμών (*missing data*), συνήθως συμβολίζονται με NA (*non available*), και αποτελούν τις τιμές εκείνες οι οποίες

δεν είναι διαθέσιμες για κάποιες μεταβλητές. Το φαινόμενο των ελλειπουσών τιμών είναι συχνό και μπορεί να έχει σημαντικές επιπτώσεις όσον αφορά τα συμπεράσματα που εξάγονται από τα δεδομένα. Υπάρχουν διάφοροι τύποι ελλειπουσών τιμών με διαφορετικές επιπτώσεις στην ισχύ των εξαγόμενων συμπερασμάτων. Αυτοί είναι:

- Ελλείπουσες τιμές τελείως τυχαία (missing completely at random).

Οι τιμές σε ένα σύνολο δεδομένων είναι ελλείπουσες τελείως τυχαία αν τα συμβάντα που οδήγησαν στη συγκεκριμένη τιμή να λείπει είναι ανεξάρτητη και από τις παρατηρήσιμες μεταβλητές και από τις μη παρατηρήσιμες παραμέτρους ενδιαφέροντος και συμβαίνουν τελείως στην τύχη [61]. Όταν οι ελλείπουσες τιμές είναι τελείως τυχαίες τότε η ανάλυση στα δεδομένα είναι αμερόληπτη. Η περίπτωση αυτή συνήθως είναι σπάνια.

- Ελλείπουσες τιμές τυχαία (missing at random).

Οι τιμές σε ένα σύνολο δεδομένων είναι ελλείπουσες τυχαία όταν η έλλειψη δεν είναι τυχαία, αλλά μπορεί να αντικατασταθεί από μεταβλητές για τις οποίες υπάρχει πλήρης πληροφόρηση. Η περίπτωση αυτή είναι μία υπόθεση που δεν μπορεί να επιβεβαιωθεί στατιστικά, αλλά θα πρέπει να έχει την βάση της στην ισχύ της λογικής αυτής της υπόθεσης. Ανάλογα με την μέθοδο ανάλυσης που θα επιλεγεί, με αυτά τα δεδομένα μπορεί να εισαχθεί μεροληψία στις παραμέτρους της ανάλυσης λόγω του ενδεχομενικού κενού των κελιών (θέσεις όπου λείπουν τα δεδομένα) [62]. Παρόλα αυτά, αν οι παράμετροι εκτιμηθούν με μέγιστη πιθανοφάνεια πλήρους πληροφόρησης (Full Information Maximum Likelihood) τότε οι εκτιμήτριες των παραμέτρων είναι συνεπής.

- Ελλείπουσες τιμές όχι τυχαία (missing not at random).

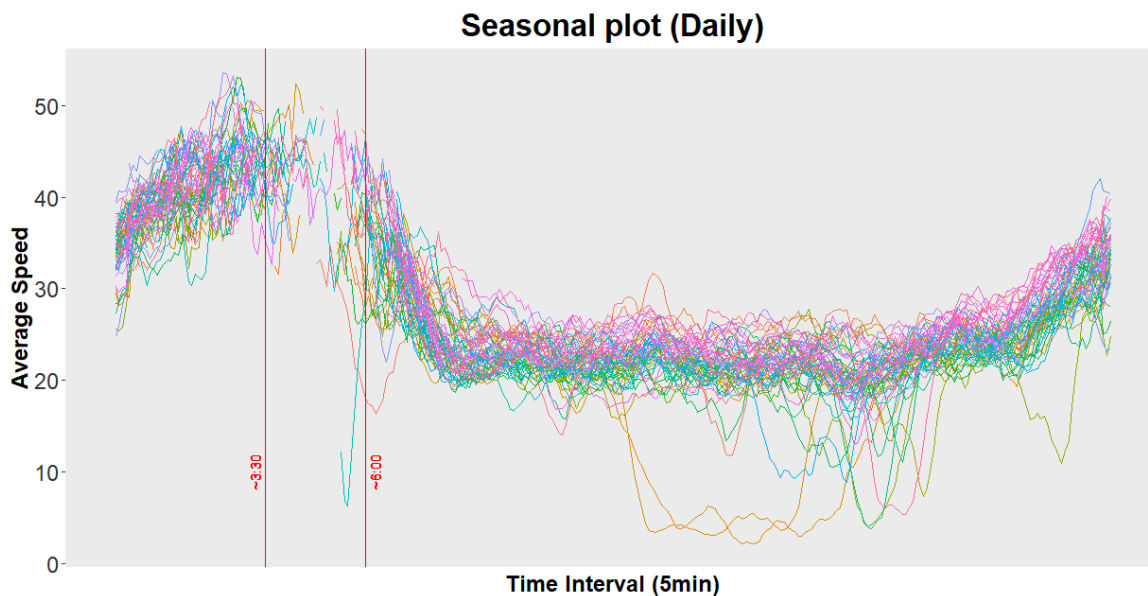
Οι τιμές σε ένα σύνολο δεδομένων είναι ελλείπουσες όχι τυχαία όταν δεν ανήκουν σε καμία από τις 2 άλλες κατηγορίες, δηλαδή η τιμή της μεταβλητής που λείπει είναι συσχετισμένη με την αιτία για την οποία λείπει.

Οι ελλείπουσες τιμές μειώνουν την αντιπροσωπευτικότητα του δείγματος και μπορούν να αλλοιώσουν την συμπερασματολογία όσον αφορά τον πληθυσμό. Γενικά υπάρχουν τρεις τεχνικές αντιμετώπισης των ελλειπουσών τιμών:

1. Συμπλήρωση δεδομένων (data imputation), όπου οι τιμές συμπληρώνονται με διάφορες τεχνικές ανάλογα το πρόβλημα.

2. Παράλειψη δεδομένων (data omission), όπου τα δείγματα που περιλαμβάνουν ελλείπουσες τιμές αφαιρούνται από την ανάλυση.
3. Επιλογή μεθόδων ανάλυσης των δεδομένων που δεν επηρεάζονται από τις ελλείπουσες τιμές.

Επανερχόμενοι στο σύνολο δεδομένων της διπλωματικής εργασίας και στις μεθόδους που χρησιμοποιήθηκαν για την βραχυπρόθεσμη πρόβλεψη κυκλοφορίας οι μέθοδοι της βαθιάς μηχανικής μάθησης είναι ρωμαλέες στο φαινόμενο των ελλειπουσών τιμών σε αντίθεση με τις κλασσικές στατιστικές μεθόδους πρόβλεψης χρονοσειρών. Στο Διάγραμμα 22 φαίνεται ότι οι ελλείπουσες τιμές εμφανίζονται κυρίως τις βραδινές ώρες (3:30 – 6:00) το οποίο λογικά οφείλεται στην μειωμένη κυκλοφορία των ταξί εκείνες τις ώρες.

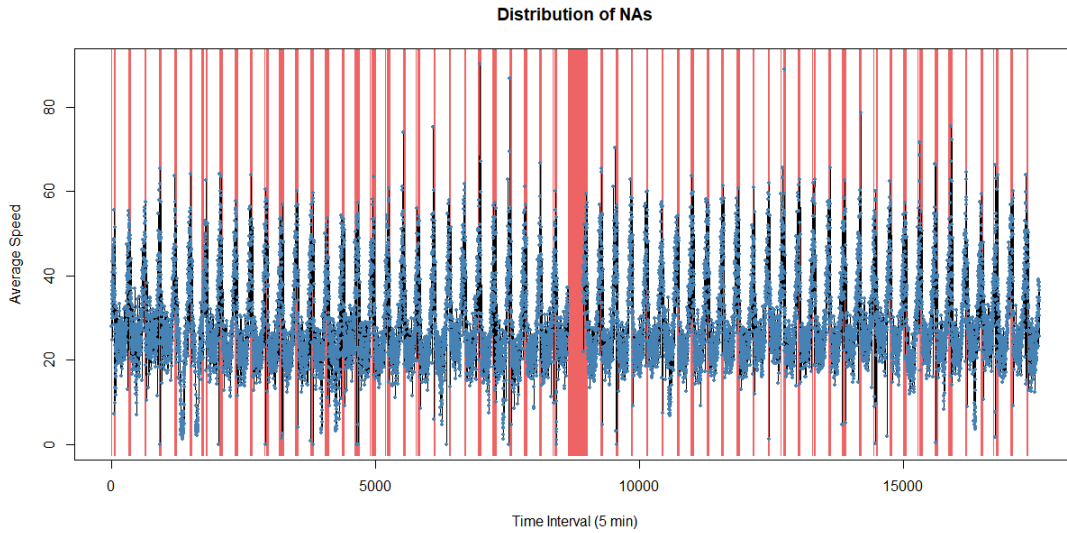


**Διάγραμμα 22** Το σύνολο δεδομένων ανά ημέρα (ομαλοποιημένες τιμές κινητού μέσου)

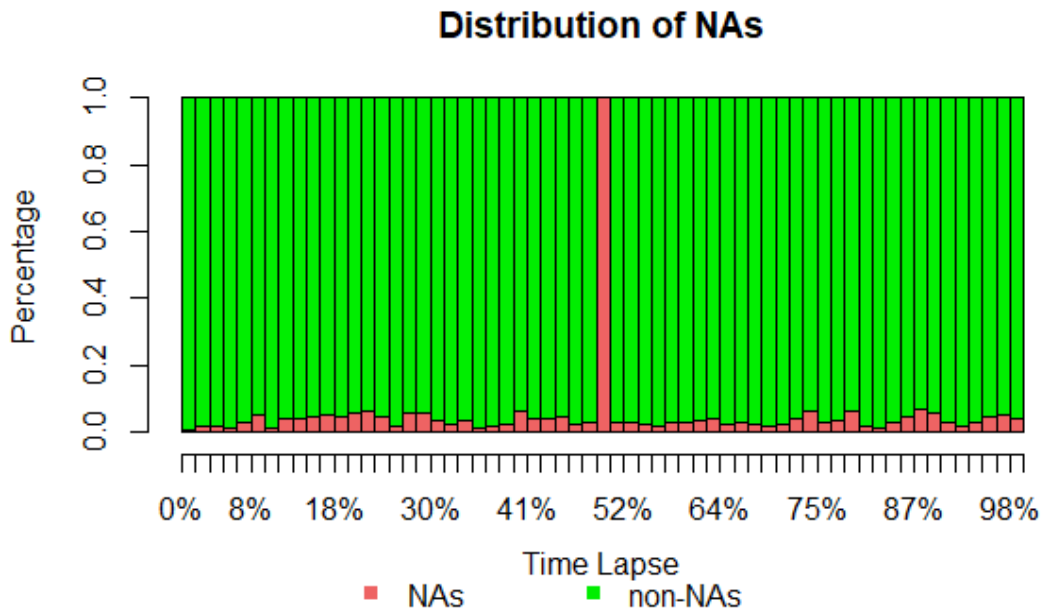
Από την ανάλυση προέκυψε ότι συνολικά στο σύνολο δεδομένων υπάρχουν 897 5λεπτα στα οποία δεν υπάρχουν διαθέσιμες τιμές από τα 17568 συνολικά 5λεπτα. Στο Διάγραμμα 23 φαίνεται η κατανομή των ελλειπουσών τιμών στο σύνολο δεδομένων όπου παρατηρείται ότι τις βραδινές ώρες εμφανίζονται συστηματικά τιμές ελλείπουσες τιμές καθώς και για μία ολόκληρη μέρα από τις 31-10-2016 00:00:00 έως τη 01-11-2016 00:00:00, πιθανόν από κάποιο τεχνικό πρόβλημα της εταιρείας, όπως φαίνεται φαίνεται στο Διάγραμμα 24. Είναι γνωστή λοιπόν η αιτία των τιμών που επαναλαμβανόμενα λείπουν από το σύνολο δεδομένων άρα οι ελλείπουσες τιμές εμφανίζονται όχι τυχαία (κατηγορία 3). Παρόλα αυτά θα χρησιμοποιηθεί η τεχνική της συμπλήρωσης δεδομένων



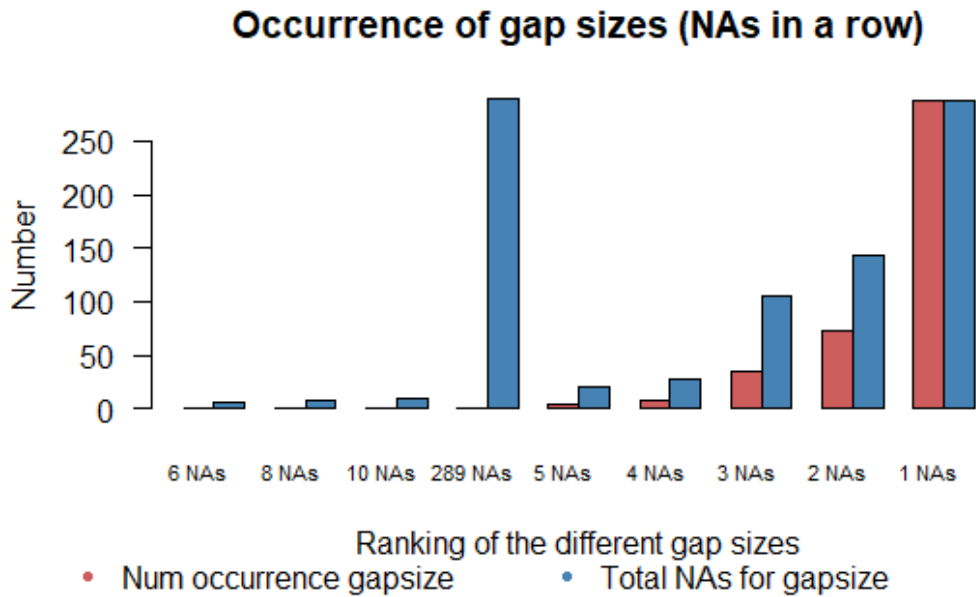
για την αντιμετώπιση τους. Για την συμπλήρωση των δεδομένων χρησιμοποιήθηκε η βιβλιοθήκη *imputeTS* σε προγραμματιστικό περιβάλλον *R* η οποία «εξειδικεύεται» αποκλειστικά στην συμπλήρωση δεδομένων για δεδομένα μονομεταβλητών χρονοσειρών και θεωρείται από τις καλύτερες διαθέσιμες βιβλιοθήκες γι' αυτό τον σκοπό.



**Διάγραμμα 23** Κατανομή ελλειπουσών τιμών στο σύνολο δεδομένων.



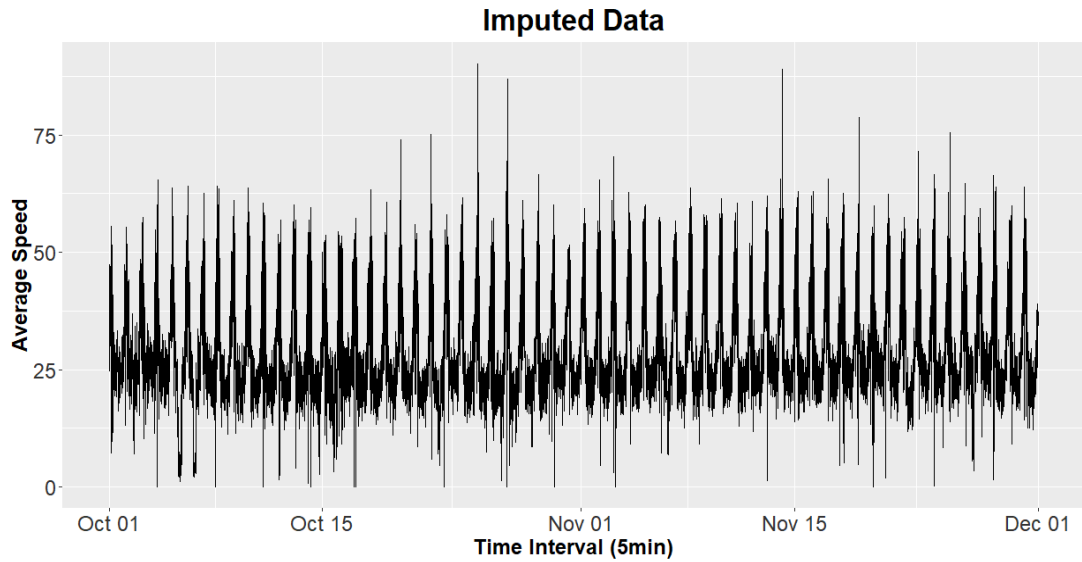
**Διάγραμμα 24** Ποσοστό ελλειπουσών τιμών ανά ημέρα.



**Διάγραμμα 25** Οι συχνότερες εμφανίσεις ακολουθιών ελλειπουσών τιμών στο σύνολο δεδομένων.

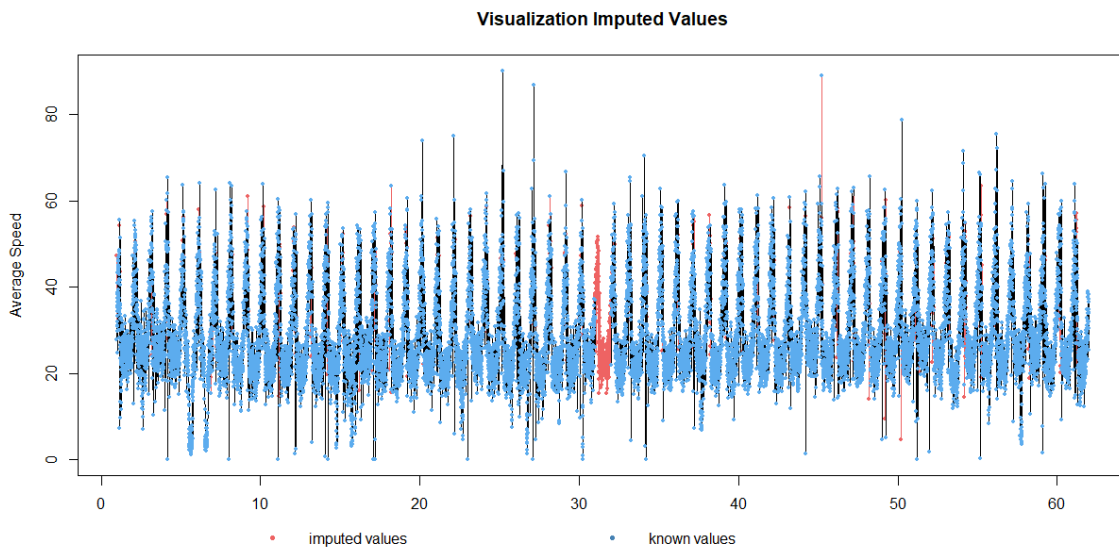
Στο Διάγραμμα 25 φαίνονται οι συχνότερες ακολουθίες ελλειπουσών τιμών στο σύνολο δεδομένων. Η συντριπτική πλειοψηφία είναι μία τιμή να λείπει ακολουθώντας σχεδόν φθίνουσα πορεία εμφανίσεων όσο αυξάνεται το μέγεθος της ακολουθίας. Και πάλι γίνεται εμφανές ότι από τα δεδομένα λείπει μία ολόκληρη μέρα δηλαδή μία ακολουθία 289 (288 + 1) 5λέπτων όπου δεν υπάρχουν διαθέσιμα δεδομένα για την μέση ταχύτητα.

Το δεδομένα έχουν εποχική συνιστώσα και μέσω του Διαγράμματος αυτοσυσχέτισης έγινε εμφανές πως η εποχική συχνότητα είναι ίση με 288. Επιλέχθηκε λοιπόν η μέθοδος *na\_seasplit* η οποία χωρίζει το σύνολο δεδομένων σε εποχές και έπειτα κάνει συμπλήρωση για κάθε ένα από τα υποσύνολα που προκύπτουν ξεχωριστά. Σαν μέθοδος για την συμπλήρωση εντός των υποσυνόλων ορίστηκε η μέθοδος της παρεμβολής. Να σημειωθεί ότι δοκιμάστηκε πληθώρα άλλων μεθόδων, με την μέθοδο που εν τέλει παρουσιάζεται εδώ να δίνει τα πιο ρεαλιστικά αποτελέσματα. Δεν θα γίνει περαιτέρω σχολιασμός για τις υπόλοιπες μεθόδους συμπλήρωσης καθώς είναι εκτός του σκοπού της παρούσας διπλωματικής εργασίας,

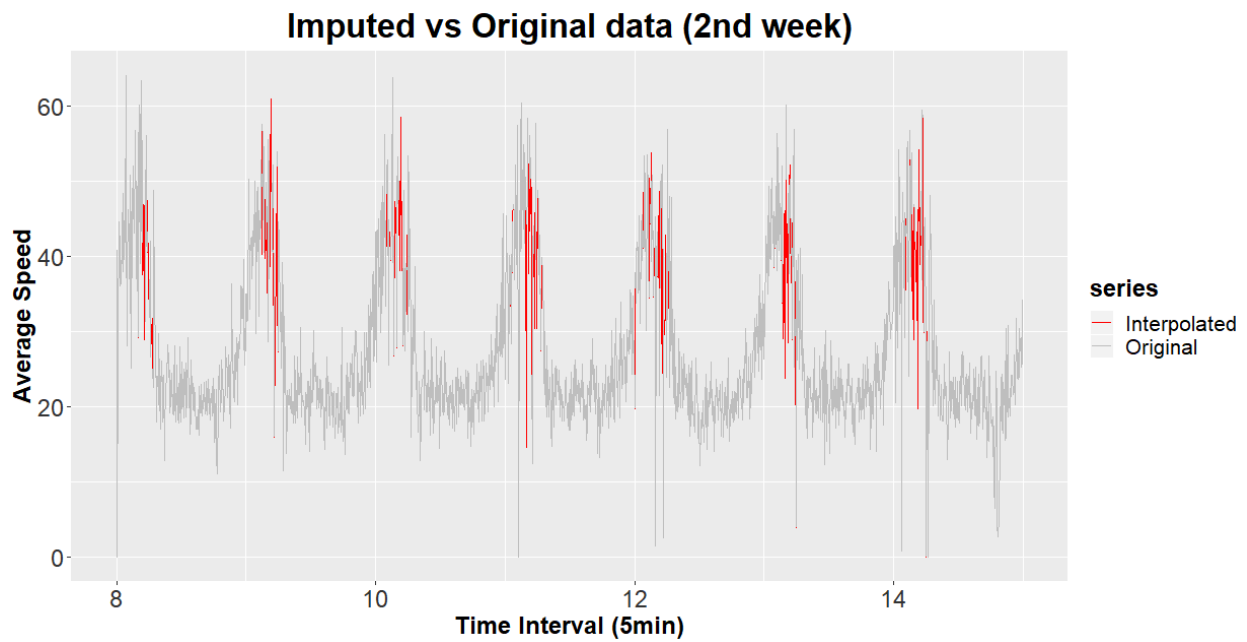


**Διάγραμμα 26** Το συμπληρωμένο σύνολο δεδομένων.

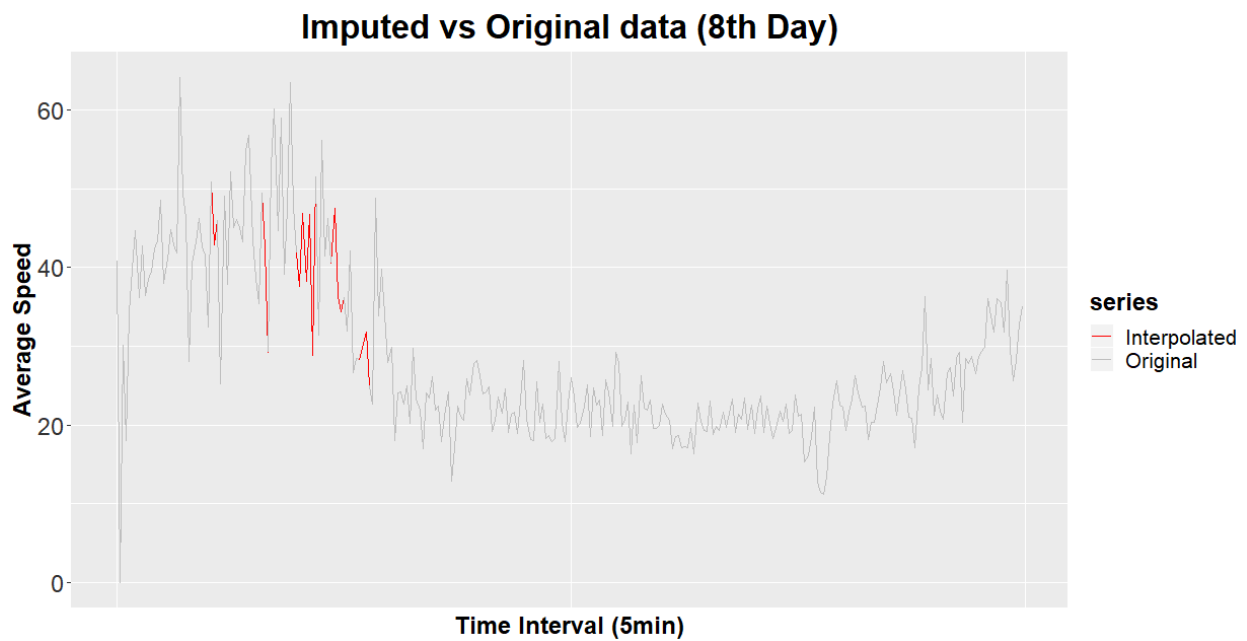
Στο Διάγραμμα 26 απεικονίζεται το σύνολο δεδομένων μετά την συμπλήρωση των ελλειπουσών τιμών, ενώ στο Διάγραμμα 27 φαίνεται το συμπληρωμένο σύνολο δεδομένων με το αρχικό σύνολο δεδομένων.



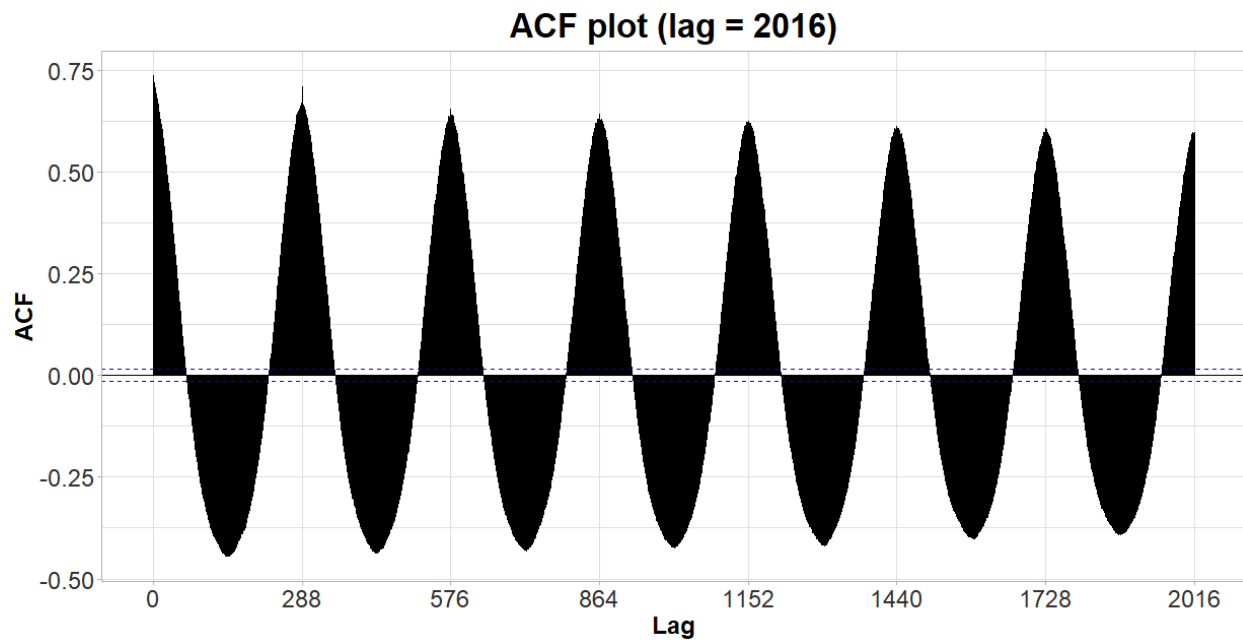
**Διάγραμμα 27** Το συμπληρωμένο και το αρχικό σύνολο δεδομένων.



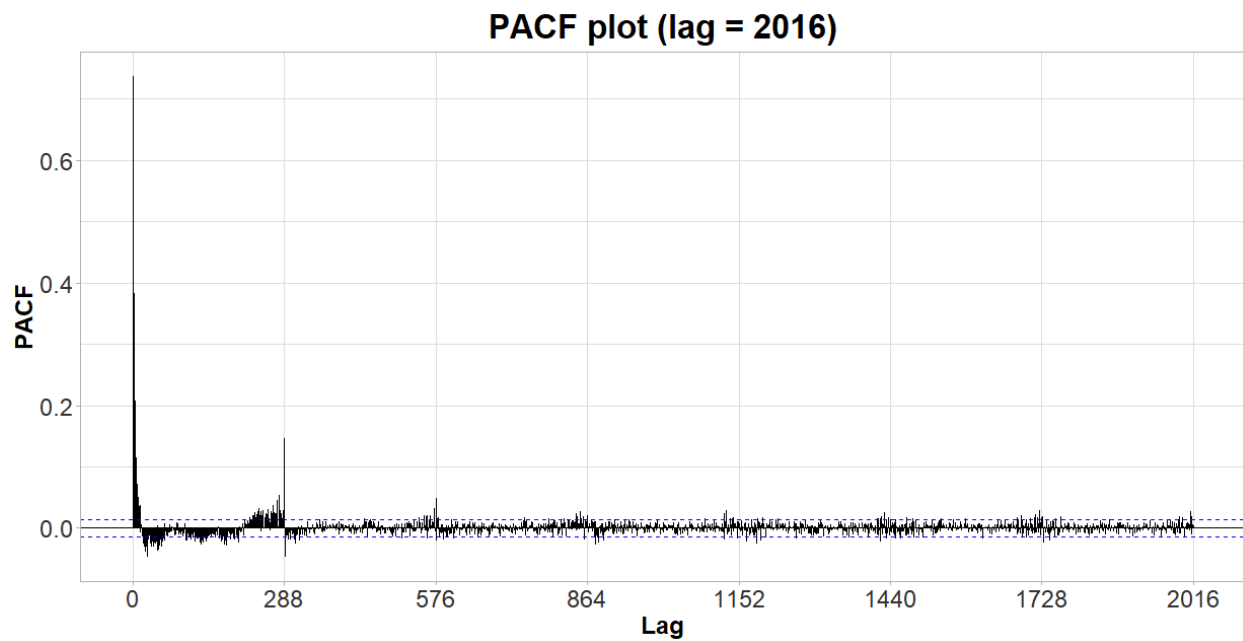
**Διάγραμμα 28** Το συμπληρωμένο σύνολο δεδομένων για την 2<sup>η</sup> εβδομάδα.



**Διάγραμμα 29** Το συμπληρωμένο σύνολο δεδομένων για την 8<sup>η</sup> ημέρα.



**Διάγραμμα 30** Το Διάγραμμα αυτοσυσχέτισης για το συμπληρωμένο σύνολο δεδομένων.



**Διάγραμμα 31** Το Διάγραμμα μερικής αυτοσυσχέτισης για το συμπληρωμένο σύνολο δεδομένων.

Τόσο για το Διάγραμμα αυτοσυσχέτισης (Διάγραμμα 30) όσο και για το Διάγραμμα μερικής αυτοσυσχέτισης (Διάγραμμα 31) η εικόνα δεν έχει αλλάξει όπως και ήταν αναμενόμενο άλλωστε. Έχοντας ολοκληρώσει την συμπλήρωση του συνόλου δεδομένων μέσω της συμπλήρωσης των ελλειπουσών τιμών συνεχίζουμε την επεξεργασία των δεδομένων με τον εντοπισμό των ακραίων τιμών.

### **3.6.3 Εντοπισμός ακραίων τιμών και τρόποι αντιμετώπισης**

Εντοπισμός των ακραίων τιμών είναι η αναγνώριση των παρατηρήσεων οι οποίες δημιουργούν υποψίες με το να διαφέρουν σημαντικά και σε υπερθετικό βαθμό από την πλειοψηφία των δεδομένων. Οπότε ακραία τιμή είναι ένα σημείο του συνόλου δεδομένων το οποίο διαφέρει σημαντικά από τις υπόλοιπες. Μία ακραία τιμή μπορεί να οφείλεται είτε στην μεταβλητότητα των μετρήσεων ή να οφείλεται σε σφάλμα μέτρησης όπου συνήθως τα τελευταία διαγράφονται από το σύνολο των δεδομένων. Μία ακραία τιμή μπορεί να προκαλέσει σοβαρά προβλήματα σε μία στατιστική ανάλυση [63], [64].

Οι ακραίες τιμές μπορούν να εμφανιστούν στην τύχη σε οποιαδήποτε κατανομή, αλλά συνήθως υποδεικνύουν είτε σφάλμα μέτρησης είτε ότι ο πληθυσμός έχει κατανομή με παχιές ουρές. Στην 1<sup>η</sup> περίπτωση είτε αφαιρούνται από το σύνολο δεδομένων είτε γίνεται χρήση μεθόδων που είναι ρωμαλέες σε ακραίες τιμές, ενώ στην 2<sup>η</sup> περίπτωση υποδεικνύουν ότι η κατανομή των δεδομένων είναι λοξή και πρέπει να δοθεί προσοχή σε μεθόδους που υποθέτουν κανονική κατανομή.

Οι ακραίες τιμές όντας οι πιο ακραίες παρατηρήσεις ενός δείγματος μπορεί να περιλαμβάνουν τον δειγματικό μέγιστο ή το δειγματικό ελάχιστο ή και τα 2. Παρόλα αυτά το δειγματικό μέγιστο ή ελάχιστο δεν σημαίνει ότι είναι πάντα ακραία τιμή διότι μπορεί να μην είναι ασυνήθιστα μακριά από τις υπόλοιπες παρατηρήσεις. Υπάρχουν δύο τύποι ακραίων τιμών: μονομεταβλητές και πολυμεταβλητές. Οι μονομεταβλητές ακραίες τιμές εμφανίζονται όταν τα δεδομένα μας έχουν κατανομή σε μονοδιάστατο χώρο καταστάσεων. Οι πολυμεταβλητές ακραίες τιμές «ζουν» σε n-διάστατο χώρο καταστάσεων και είναι δύσκολο κοιτώντας κατανομές σε n διαστάσεις για τον άνθρωπο να καταλάβει ποιες είναι ακραίες τιμές γι' αυτό και είναι απαραίτητη η ανάπτυξη μοντέλων που να αναλαμβάνουν τέτοιες διαδικασίες [65].

Οι ακραίες τιμές μπορούν να διαφέρουν στην φύση τους ανάλογα με το περιβάλλον στο οποίο παρουσιάζονται: σημειακές ακραίες τιμές, ακραίες τιμές συνάφειας ή συγκεντρωτικές ακραίες

τιμές. Οι σημειακές ακραίες τιμές είναι μεμονωμένα σημειακά δεδομένα που βρίσκονται μακριά από την υπόλοιπη κατανομή. Οι ακραίες τιμές συνάφειας μπορεί να είναι θόρυβος στα δεδομένα, όπως για παράδειγμα σημεία στίξης σε μία ανάλυση κειμένου ή σήμα θορύβου σε αναγνώριση φυσικής γλώσσας. Οι συγκεντρωτικές ακραίες τιμές μπορεί να είναι υποσύνολα «καινοτομιών» στα δεδομένα όπως π.χ. ένα σήμα το οποίο να υποδεικνύει την ανακάλυψη ενός νέου φαινομένου [65].

Τα ακραία σημεία μπορούν να έχουν πολλές αιτίες ανωμαλίας. Παροδική δυσλειτουργία μίας συσκευής μέτρησης, λάθος στην μετάδοση μίας μέτρησης ή ενός σήματος είναι κάποιες συνήθεις αιτίες. Ακραίες τιμές μπορούν να προκύψουν λόγω αλλαγών στην δυναμική ενός συστήματος, σε κακόβουλη συμπεριφορά, ανθρώπινο λάθος, σφάλμα μηχανήματος ή λόγω φυσικών αποκλίσεων του πληθυσμού.

Δεν υπάρχει κάποιος αυστηρός μαθηματικός ορισμός για το τι είναι ακραία παρατήρηση. Το αν μία παρατήρηση είναι ακραία είναι κριτήριο υποκειμενικό. Υπάρχει πληθώρα μεθόδων στην βιβλιογραφία για εντοπισμό ακραίων παρατηρήσεων [66]:

- Βασισμένες στην κατανομή (distribution based), όπου οι ακραίες τιμές είναι παρατηρήσεις οι οποίες αποκλίνουν από μία δοσμένη κατανομή (κανονική, πουασσόν κλπ.).
- Βασισμένες στο βάθος, οι οποίες βασίζονται ουσιαστικά στον υπολογισμό διαφορετικού επιπέδου  $k-d$  κυρτών θηκών, όπου οι ακραίες παρατηρήσεις είναι αντικείμενα που βρίσκονται στο εξωτερικό αυτών των θηκών.
- Βασισμένες στην απόσταση, όπου υπολογίζεται κάποια κρίσιμη απόσταση. Όταν μία παρατήρηση βρίσκεται πάνω από αυτή την κρίσιμη απόσταση, τότε θεωρείται ότι η παρατήρηση ανήκει στο σύνολο των ακραίων παρατηρήσεων.
- Βασισμένες στην ομαδοποίηση, όπου γίνεται χρήση αλγόριθμων ομαδοποίησης και οι ακραίες τιμές προκύπτουν ως υποπροϊόν της διαδικασίας ομαδοποίησης.
- Βασισμένες στην κατανομή πυκνότητας πιθανότητας, όπου γίνεται εντοπισμός τοπικά ακραίων παρατηρήσεων «κοιτώντας» τοπικά γύρω από ένα αντικείμενο την πυκνότητα.

Η επιλογή της μεθόδου για την αντιμετώπιση των ακραίων τιμών πρέπει να εξαρτάται από την αιτία που προκάλεσε την εμφάνισή της. Κάποιοι εκτιμητές είναι εξαιρετικά «ευαίσθητοι» σε ακραίες παρατηρήσεις. Θα αναφερθούν οι τρεις πιο γνωστές προσεγγίσεις:

### 1. Διατήρηση των ακραίων παρατηρήσεων (retention)

Ακόμα και ένα μοντέλο που χρησιμοποιεί την υπόθεση της κανονικής κατανομής είναι κατάλληλο για τα δεδομένα προς ανάλυση, για μεγάλου μεγέθους δείγματα και όσον αφορά στην παρουσία ακραίων παρατηρήσεων δεν θα πρέπει να γίνεται η διαγραφή τους αυτόματα. Η οποιαδήποτε εφαρμογή θα μπορούσε να κάνει χρήση μεθόδων συσταδοποίησης οι οποίες να είναι ρωμαλέες σε ακραίες παρατηρήσεις έτσι ώστε να μοντελοποιούνται δεδομένα στα οποία είναι φυσιολογική η εμφάνιση ακραίων παρατηρήσεων.

### 2. Εξαίρεση των ακραίων παρατηρήσεων (exclusion)

Η διαγραφή των ακραίων παρατηρήσεων είναι αμφιλεγόμενη διαδικασία και αμφισβητείται από πολλούς επιστήμονες. Ενώ υπάρχουν μαθηματικά κριτήρια τα οποία προσφέρουν μία αντικειμενική και ποσοτική μέθοδο για διαγραφή δεδομένων, δεν δημιουργούν παρόλα αυτά μία διαδικασία η οποία να είναι επιστημονικής και αυστηρά μεθοδολογικής φύσεως, ειδικά για μικρού μεγέθους σύνολα δεδομένων ή σε περιπτώσεις όπου δεν μπορεί να γίνει η υπόθεση της κανονικής κατανομής. Η μέθοδος διαγραφής των ακραίων παρατηρήσεων υιοθετείται κυρίως σε επιστημονικές περιοχές όπου το υποβόσκων μοντέλο της διαδικασίας που παρατηρείται μέσω μετρήσεων και η κατανομή του σφάλματος αυτών των μετρήσεων είναι γνωστά.

Οι δύο κύριες προσεγγίσεις που ακολουθούνται στην εξαίρεση των ακραίων παρατηρήσεων είναι η περικοπή (truncation) και η αποδοκιμασία (winsorizing). Στην πρώτη περίπτωση έχουμε πλήρη διαγραφή των δεδομένων (ακραίων παρατηρήσεων) ενώ στην δεύτερη προχωράμε σε αντικατάσταση των δεδομένων συνήθως με τους κοντινότερους «φυσιολογικούς» γείτονες της εκάστοτε ακραίας παρατήρησης.

### 3. Μη κανονικές κατανομές (non normal distributions)

Σε αυτήν την περίπτωση θα πρέπει να μελετηθεί η περίπτωση όπου η υποβόσκουσα κατανομή των δεδομένων δεν είναι ούτε προσεγγιστικά κανονική, και να εξεταστεί η περίπτωση κατανομών με παχιές ουρές. Για παράδειγμα, όταν γίνεται δειγματοληψία από μία Cauchy κατανομή, η δειγματική διασπορά αυξάνεται όσο αυξάνεται το μέγεθος του δείγματος, και ο δειγματικός μέσος αποτυγχάνει να συγκλίνει καθώς αυξάνονται οι παρατηρήσεις και έτσι οι ακραίες παρατηρήσεις αναμένονται σε πολύ μεγαλύτερο βαθμό σε σχέση με την κανονική κατανομή. Ακόμα και για πολύ



μικρές διαφορές στο πάχος των ουρών μίας κατανομής μπορεί να δημιουργήσει τεράστιες διαφορές στον αναμενόμενο αριθμό ακραίων τιμών στο δείγμα.

Στο σύνολο δεδομένων εκ πρώτης όψεως παρατηρούνται κάποιες τιμές οι οποίες μπορούν με βάση την λογική να χαρακτηριστούν ως ακραίες τιμές όπως ταχύτητες οι οποίες είναι μεγαλύτερες των 60 km/h (69 παρατηρήσεις) όπως και οι μηδενικές ταχύτητες (16 παρατηρήσεις) οι οποίες είναι πιθανό να οφείλονται στην φύση της συλλογής δεδομένων. Δεν πρέπει να ξεχνάμε ότι η συλλογή δεδομένων έγινε από *gps* συσκευές από οδηγούς ταξί. Άρα οι μηδενικές ταχύτητες είναι πιθανό να υποδεικνύουν ότι σε ένα 5λεπτο είχαμε κάλυψη από ένα οδηγό ο οποίος να είναι σταματημένος στον δρόμο ενδιαφέροντος. Το ίδιο και για τις μεγάλες ταχύτητες είναι πιθανή και πάλι η κάλυψη από ένα οδηγό ο οποίος προσπέρασε τον δρόμο ενδιαφέροντος με μεγάλη ταχύτητα. Πρέπει να τονιστεί ότι δεν μας ενδιαφέρει ένα μοντέλο να «μάθει» τέτοιες συμπεριφορές και να προβλέπει τέτοιες τιμές οι οποίες δεν αποτελούν γενίκευση της συμπεριφοράς των οδηγών ως προς την ταχύτητα αλλά αφορούν μόνο μεμονωμένα περιστατικά και οφείλονται στην περιορισμένη κάλυψη του δικτύου (κυρίως τις βραδινές ώρες).

Πέραν όμως από αυτές τις προφανείς μη κανονικές τιμές, λόγω της ιδιομορφίας της συλλογής των δεδομένων μπορεί να υπάρχει και κάποια ανεξήγητη μεταβλητότητα στα δεδομένα η οποία να χρειάζεται κάποιου είδους «ομαλοποίηση». Για τις ανάγκες της διπλωματικής έγινε χρήση της βιβλιοθήκης *anomalize* [67] σε προγραμματιστικό περιβάλλον *R*. Η εν λόγω βιβλιοθήκη «εξειδικεύεται» στον εντοπισμό και τον καθαρισμό ακραίων τιμών για δεδομένα χρονοσειρών. Θεωρήθηκε ότι η παρουσία των ακραίων παρατηρήσεων οφείλεται σε φυσιολογικές αποκλίσεις του πληθυσμού (υψηλές ταχύτητες, σταματημένα οχήματα) οι οποίες εκδηλώνονται στο σύνολο δεδομένων λόγω του τρόπου συλλογής των δεδομένων.

Για την επιλογή της μεθόδου και πάλι διαδραμάτισε σημαντικό ρόλο η παρουσία της έντονης εποχικής συνιστώσας του συνόλου δεδομένων και στην αποτελεσματικότητά της μεθόδου με βάση τις αναφορές στην επιστημονική κοινότητα. Η διαδικασία που ακολουθείται είναι η εξής:

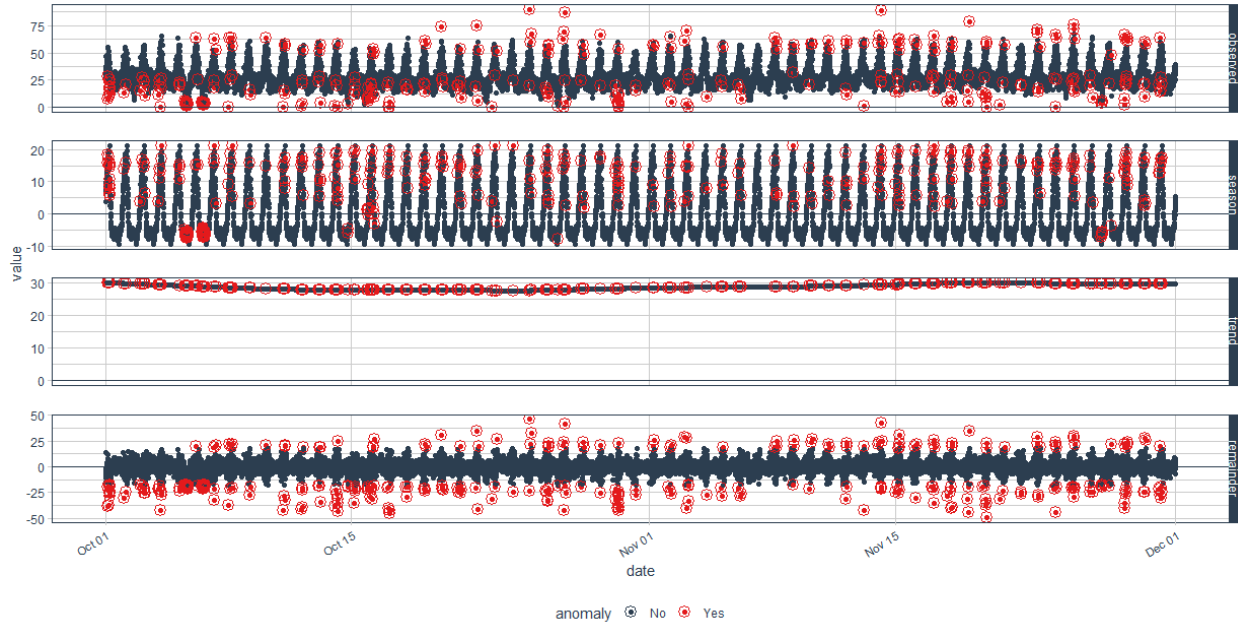
- Αρχικά, γίνεται ανάλυση σε συνιστώσες της χρονοσειράς με την οποία αφαιρείται από τα δεδομένα η συνιστώσα τάσης και η εποχική συνιστώσα. Αυτή η διαδικασία πραγματοποιείται με την μέθοδο “LOESS” που αποτελεί ένα είδος τοπικής παλινδρόμησης μέσω της συνάρτησης *time\_decompose()*.

- Έπειτα, αφού αφαιρεθούν οι δύο συνιστώσες οι ακραίες παρατηρήσεις εντοπίζονται μέσω των υπολοίπων με χρήση της συνάρτησης *anomalize()* η οποία κάνει χρήση της μεθόδου “gesd” γενικευμένων ακραίων studentized αποκλίσεων (generalized extreme studentized deviation) [68] και είναι η ίδια μέθοδος με τη πολύ γνωστή βιβλιοθήκη του twitter “AnomalyDetection”. Σε σχέση με την άλλη ευρέως γνωστή μέθοδο του εσωτερικού εύρους τεταρτημορίου (inner quantile range method, “iqr”) που χρησιμοποιείται και από το πακέτο *forecast*, η “gesd” να μεν θεωρείται πιο αργή αλλά πολύ μεγαλύτερης ακρίβειας.
- Αφού εντοπιστούν οι ακραίες παρατηρήσεις χρησιμοποιείται η εντολή *clean\_anomalies()* και αντικαθίστανται οι ακραίες παρατηρήσεις κάνοντας χρήση της συνιστώσας τάσης και της εποχικής συνιστώσας.

Όπως γίνεται αντιληπτό η μέθοδος που χρησιμοποιήθηκε για την αντιμετώπιση των ακραίων τιμών είναι εξαίρεση (exclusion) με αποδοκιμασία (winsorizing). Στο Διάγραμμα 32 φαίνονται με κόκκινο κύκλο τα σημεία του συνόλου δεδομένων που αναγνωρίστηκαν ως ακραίες παρατηρήσεις. Είναι εμφανές ότι τα σημεία που αναγνωρίστηκαν ως ακραίες παρατηρήσεις είναι κατά κύριο λόγο σημεία τα οποία παίρνουν ασυνήθιστα μεγάλες ή μικρές τιμές. Συνολικά βρέθηκαν 306 ακραίες παρατηρήσεις ενώ ο μέγιστος αριθμός που δόθηκε στον αλγόριθμο για την εύρεση ακραίων παρατηρήσεων ήταν 3.513 ( $\text{max\_anomns} = 0.2$ ). Στην παράμετρο *alpha* του αλγορίθμου που ελέγχει το πλάτος του «κανονικού» εύρους δόθηκε η τιμή 0.05 η οποία θεωρείται συντηρητική προτείνεται όμως γενικά σε αρκετά forum ανάλυσης δεδομένων από διακεκριμένους καθηγητές. Το 70% των ακραίων παρατηρήσεων είναι τις βραδινές ώρες μεταξύ 00:00:00 και 06:00:00 γεγονός που ενισχύει τον ισχυρισμό μας για την συμπεριφορά των οδηγών ταξί και για την κάλυψη του κυκλοφοριακού δικτύου τις νυκτερινές ώρες.

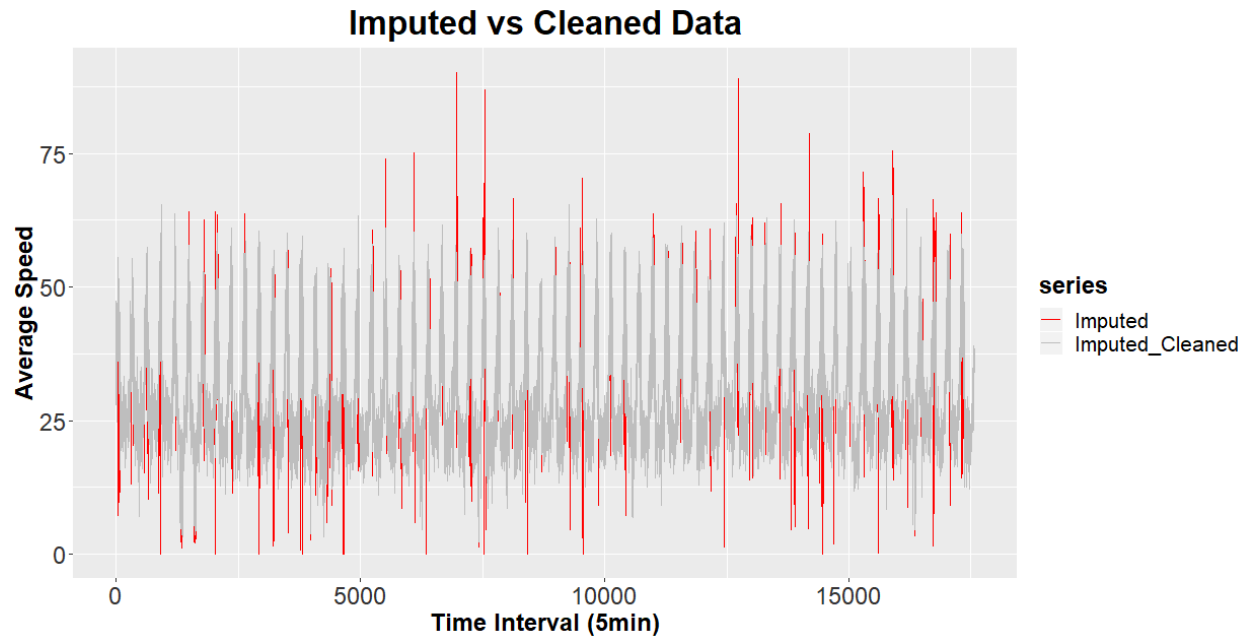
Στο Διάγραμμα 33 απεικονίζεται το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με τις ακραίες παρατηρήσεις να έχουν αντικατασταθεί. Να σημειωθεί ότι εκτός από τον αποκλεισμό πιθανόν σημείων μόχλευσης εμφανίζεται και σταθεροποίηση της διασποράς των δεδομένων το οποίο διαδραματίζει σημαντικό ρόλο στις κλασσικές στατιστικές μεθόδους λόγω της απαίτησης της ασθενούς στασιμότητας της χρονοσειράς.

Στα Διαγράμματα 34 (2<sup>η</sup> εβδομάδα) και 35 (8<sup>η</sup> ημέρα) έχοντας «μεγεθυνθεί» όλο και περισσότερο φαίνεται και γραφικά ότι οι ακραίες παρατηρήσεις που εντόπισε ο αλγόριθμος είναι κυρίως τις νυκτερινές ώρες όπου εντοπίζεται και μεγαλύτερη μεταβλητότητα των δεδομένων.

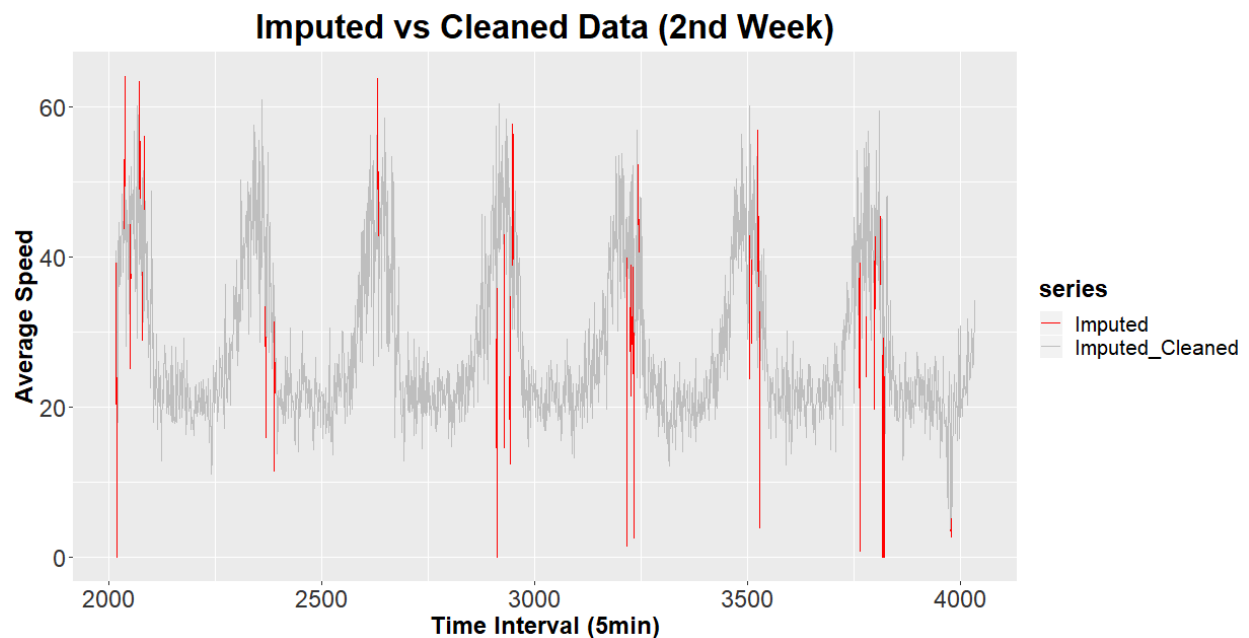


**Διάγραμμα 32** Εντοπισμός των ακραίων παρατηρήσεων του συνόλου δεδομένων με την μέθοδο «gesd».

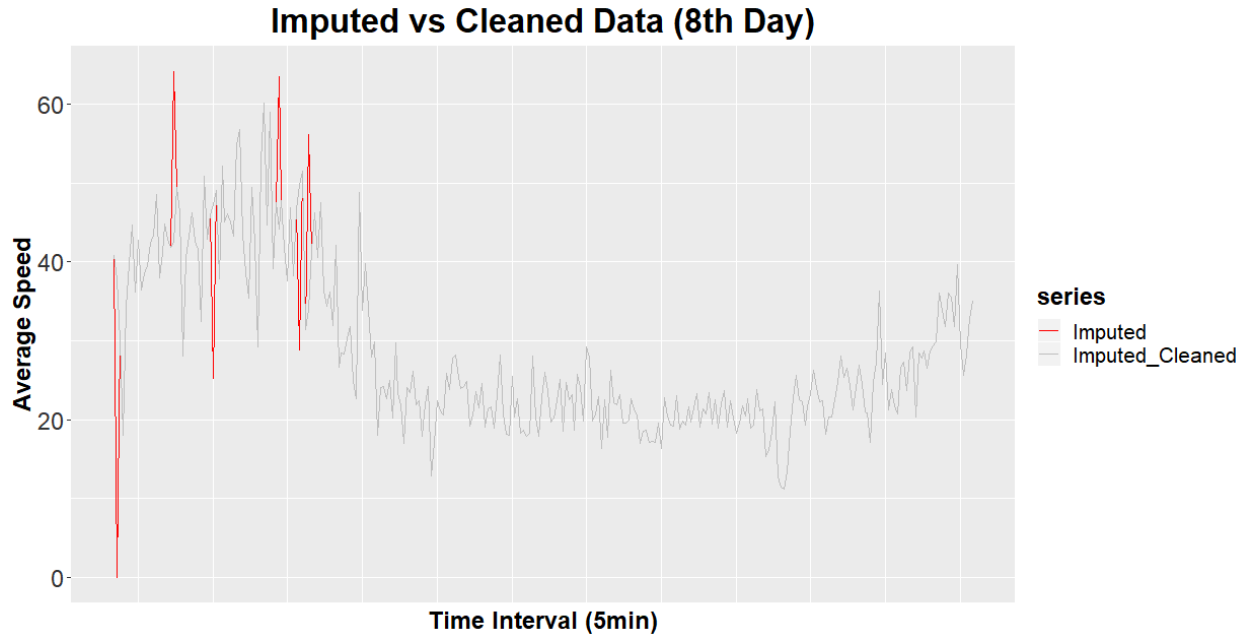
Να σημειωθεί ότι στην παράμετρο του αλγορίθμου *frequency* δόθηκε η επιλογή «auto» και όπως ήταν αναμενόμενο ως συχνότητα εποχικότητας επιστράφηκε η τιμή 288, όπως εξάλλου είχε παρατηρηθεί και μέσω του διαγράμματος αυτοσυσχέτισης. Τα δεδομένα που επιστρέφονται μετά και απ' αυτό το βήμα είναι και που θα χρησιμοποιηθούν για την εκπαίδευση των μοντέλων. Είναι σημαντικό σε αυτό το σημείο να σχολιαστεί ότι η συμπλήρωση δεδομένων και η αντικατάσταση των ακραίων παρατηρήσεων έγινε καθολικά στο σύνολο δεδομένων. Αυτό σημαίνει ότι έχει εισαχθεί στο μοντέλο αυτό που αποκαλείται στην στατιστική χρονοσειρών μεροληψία εμπρόσθιου κοιτάγματος (look-ahead bias). Πιο συγκεκριμένα, όταν αργότερα χωριστούν τα δεδομένα σε σύνολο εκπαίδευσης και σύνολο ελέγχου, καθώς η προεπεξεργασία των δεδομένων έγινε καθολικά, σημαίνει οτι έχουμε «κοιτάζει» το σύνολο ελέγχου και άρα έχει εισαχθεί μελλοντική πληροφορία στο μοντέλο. Παρόλα αυτά τα 2 αυτά βήματα προεπεξεργασίας είναι σημαντικά για την σωστή υλοποίηση των μεθόδων πρόβλεψης που επιλέχθηκαν και κυρίως των κλασικών μεθόδων στατιστικής. Αφού έχουμε καταλήξει με το τελικό σύνολο δεδομένων θα προχωρήσουμε στη λεπτομερή ανάλυση του με σκοπό την αναγνώριση των τελικών χαρακτηριστικών και των ιδιοσυγκρασιών που το διέπουν (στασιμότητα κλπ.).



**Διάγραμμα 33** Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με τις ακραίες παρατηρήσεις να έχουν αντικατασταθεί.



**Διάγραμμα 34** Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με αντικατεστημένες τις ακραίες παρατηρήσεις για την 2<sup>η</sup> εβδομάδα.



**Διάγραμμα 35** Το συμπληρωμένο σύνολο δεδομένων και το συμπληρωμένο σύνολο δεδομένων με αντικατεστημένες τις ακραίες παρατηρήσεις για την 8<sup>η</sup> ημέρα.

### 3.7 ΑΝΑΛΥΣΗ ΣΕ ΣΥΝΙΣΤΩΣΕΣ – ΕΠΟΧΙΚΗ ΑΠΟΣΥΝΘΕΣΗ

Σε προηγούμενη ενότητα μιλήσαμε για τις συνιστώσες από τις οποίες μπορεί να αποτελείται μία χρονοσειρά. Εδώ θα δούμε τις μεθόδους με τις οποίες γίνεται η ανάλυση συνιστωσών και πως μπορούν να χρησιμοποιηθούν στην ανάπτυξη μοντέλων πρόβλεψης. Η προσθετική ανάλυση μπορεί να γραφεί ως,

$$y_t = S_t + T_t + R_t \quad (3.3)$$

όπου  $y_t$  είναι τα δεδομένα,  $S_t$  είναι εποχική συνιστώσα,  $T_t$  είναι η συνιστώσα τάσης-κύκλου, και  $R_t$  είναι η υπολείπουσα συνιστώσα, όλες την περίοδο  $t$ . Εναλλακτικά, μπορεί να υποτεθεί πολλαπλασιαστική ανάλυση η οποία γράφεται ως,

$$y_t = S_t \times T_t \times R_t \quad (3.4)$$

Η προσθετική ανάλυση είναι καταλληλότερη εάν το μέγεθος των εποχικών διακυμάνσεων, ή της διακύμανσης γύρω από την τάση-κύκλο δεν ποικίλει ανάλογα με το επίπεδο της χρονοσειράς. Όταν η διασπορά στο εποχικό μοτίβο, ή γύρω από την τάση-κύκλο εμφανίζεται να είναι ανάλογη του επιπέδου της χρονοσειράς τότε η πολλαπλασιαστική ανάλυση είναι πιο κατάλληλη [57].

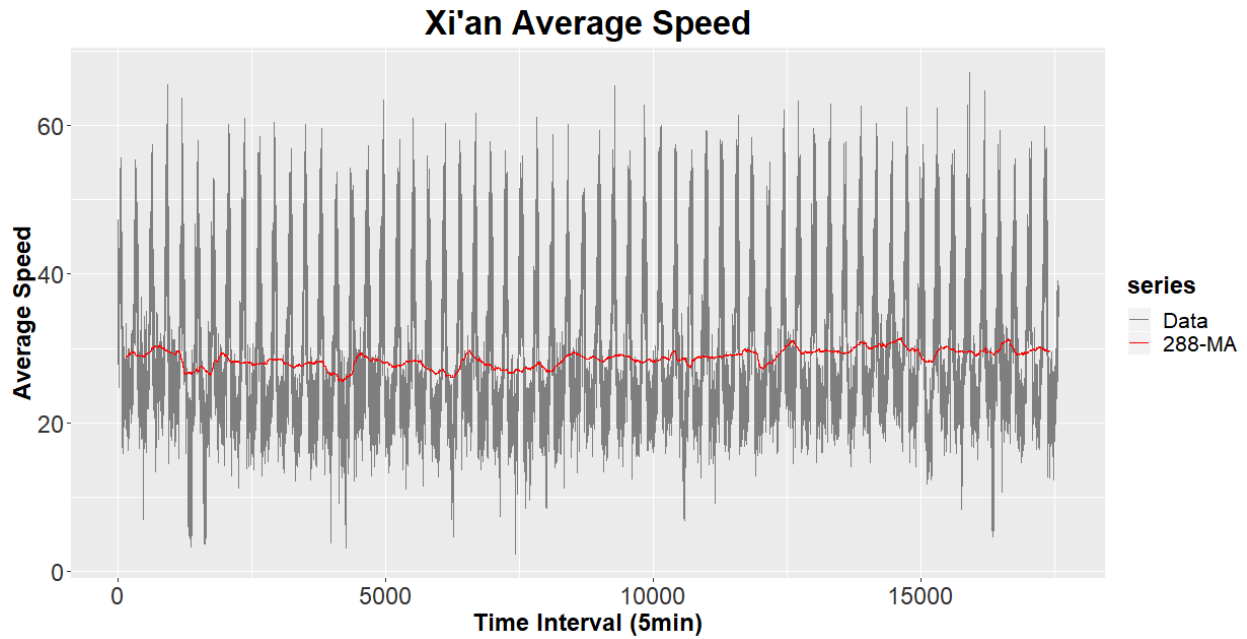
Αν από τα αρχικά δεδομένα αφαιρεθεί η εποχική συνιστώσα, τότε οι τιμές που προκύπτουν από αυτήν την αφαίρεση ονομάζονται τα εποχιακά προσαρμοσμένα δεδομένα. Στην προσθετική ανάλυση συνιστωσών, τα εποχιακά προσαρμοσμένα δεδομένα δίνονται από τον τύπο  $y_t - S_t$ , ενώ για την πολλαπλασιαστική είναι  $y_t/S_t$ . Εάν η διακύμανση λόγω της εποχικότητας δεν είναι πρωτεύουσας σημασίας, τότε τα εποχιακά προσαρμοσμένα δεδομένα μπορούν να φανούν πολύ χρήσιμα, δηλαδή με την αφαίρεση της διακύμανσης λόγω εποχικότητας επιτρέπεται στο μοντέλο να επικεντρωθεί στην μη-εποχική διακύμανση.

Η μέθοδος που χρησιμοποιήθηκε για την ανάλυση συνιστωσών είναι η μέθοδος STL (Seasonal and Trend decomposition using "LOESS" - STL decomposition) [69], η οποία είναι μέθοδος υπολογισμού της εποχικής και της συνιστώσας τάσης με χρήση της μεθόδου LOESS, η οποία θεωρείται κατάλληλη για την εκτίμηση μη γραμμικών σχέσεων. Πρόκειται για μία πολύπλευρη και ρωμαλέα μέθοδο ανάλυσης συνιστωσών μίας χρονοσειράς που σε σχέση με τις υπόλοιπες μεθόδους που χρησιμοποιούνται (κλασσική MA, SEATS, X11) θεωρείται καλύτερη και πιο ευέλικτη [57] αφού:

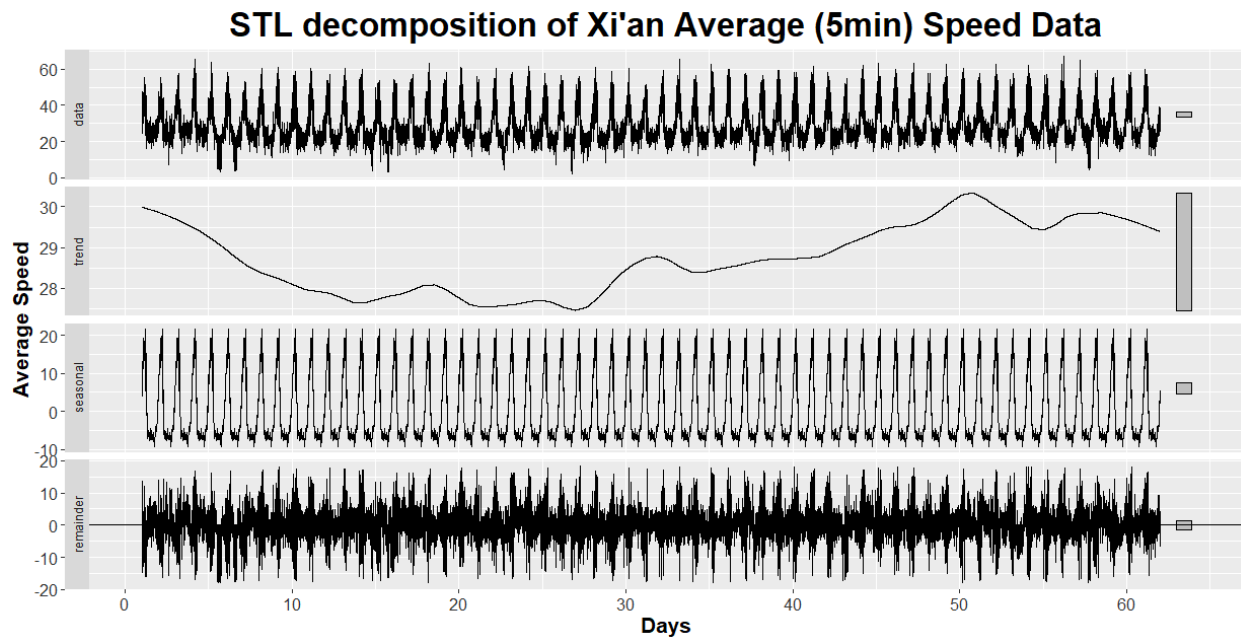
- Μπορεί να χρησιμοποιηθεί για υψηλότερης συχνότητας εποχικότητα και όχι μόνο για μηνιαία και 4μηνιαία εποχικότητα.
- Η εποχική συνιστώσα επιτρέπεται να αλλάζει με το χρόνο και ο ρυθμός αλλαγής μπορεί να καθοριστεί και να ελεγχθεί.
- Η ομαλότητα της συνιστώσας τάσης-κύκλου μπορεί επίσης να ελεγχθεί.
- Είναι ρωμαλέα στην εμφάνιση ακραίων παρατηρήσεων έτσι ώστε να μην επηρεάζεται η εκτίμηση της εποχικής και της συνιστώσας τάσης-κύκλου από την περιστασιακή εμφάνιση ασυνήθιστων παρατηρήσεων. Παρόλα αυτά θα επηρεαστεί η υπολείπουσα συνιστώσα.

Μία απλή γραφική μέθοδος για να εντοπιστεί αν υπάρχει τάση στα δεδομένα είναι η εφαρμογή ομαλοποίησης κινητού μέσου στα δεδομένα χρησιμοποιώντας ως παράμετρο υστέρησης την εποχική συχνότητα δηλαδή εδώ ως τιμή υστέρησης επιλέχθηκε το 288. Στο Διάγραμμα 36 φαίνεται το τελικό σύνολο δεδομένων και οι ομαλοποιημένες τιμές (κόκκινο). Δεν φαίνεται εκ πρώτης όψεως τουλάχιστον να υπάρχει σημαντική τάση στα δεδομένα. Για την μέθοδο αποσύνθεσης STL χρησιμοποιήθηκε η συνάρτηση `stl` της βιβλιοθήκης *forecast* σε προγραμματιστικό περιβάλλον R. Οι δύο κύριες παράμετροι της συνάρτησης είναι το παράθυρο

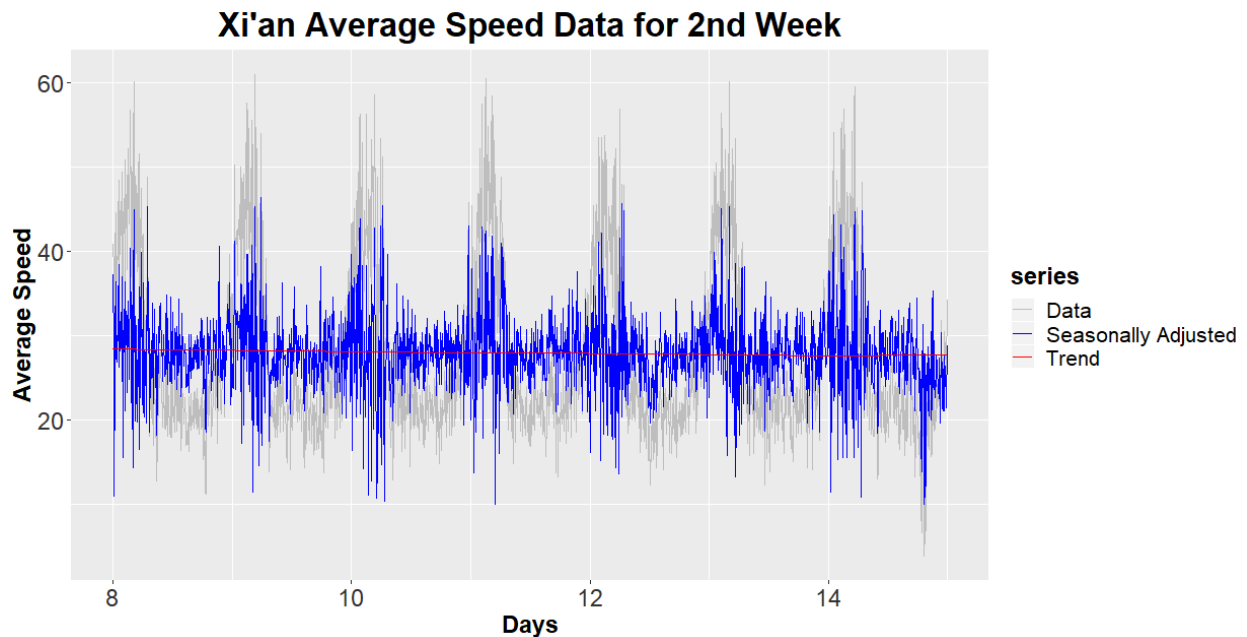
τάσης-κύκλου και το εποχικό παράθυρο. Αυτές οι δύο παράμετροι ελέγχουν πόσο ραγδαία η εποχική και η συνιστώσα τάσης-κύκλου μπορούν να αλλάζουν.



**Διάγραμμα 36** Τελικό σύνολο δεδομένων και ομαλοποιημένα δεδομένα με κινητό μέσο (τιμή υστέρησης = 288).



**Διάγραμμα 37** Ανάλυση συνιστωσών (τάσης, εποχική, υπόλοιπο) του τελικού συνόλου δεδομένων με χρήση της μεθόδου "LOESS".



**Διάγραμμα 38** Το τελικό σύνολο δεδομένων, τα εποχιακά προσαρμοσμένα δεδομένα και η συνιστώσα τάσης για την 2<sup>η</sup> εβδομάδα.

Το εποχικό παράθυρο ορίστηκε να είναι περιοδικό και υπολογίστηκε από την συνάρτηση 289 (πρέπει να είναι μονός αριθμός) ενώ το παράθυρο τάσης-κύκλου τέθηκε 2017 ( $7 * 288 + 1$  - μονός αριθμός υποχρεωτικά) διότι ζητείται η συνιστώσα τάσης- κύκλου να είναι πιο ομαλή. Στο Διάγραμμα 37 φαίνονται οι συνιστώσες που επιστρέφονται από την συνάρτηση. Στο Διάγραμμα 38 φαίνονται μαζί το τελικό σύνολο δεδομένων, τα εποχιακά προσαρμοσμένα δεδομένα και η συνιστώσα τάσης για την 2<sup>η</sup> εβδομάδα για καλύτερη εποπτεία. Δημιουργούνται υποψίες ότι ίσως έχει μείνει κάποιο ποσοστό εποχικότητας στα εποχιακά προσαρμοσμένα δεδομένα.

### 3.8 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΤΕΛΙΚΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΩΝ ΕΠΟΧΙΑΚΑ ΠΡΟΣΑΡΜΟΣΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Στην συνέχεια θα διεξαχθεί μία σειρά από ελέγχους στο τελικό σύνολο δεδομένων και στα εποχιακά προσαρμοσμένα δεδομένα για να εντοπιστεί αν υπάρχει τάση στα δεδομένα και εποχικότητα. Η συνάρτηση *ndiffs()* της βιβλιοθήκης *forecast* υπολογίζει των αριθμών των διαφορών που απαιτούνται για να γίνει μία χρονοσειρά στάσιμη (ως προς την τάση), κάνοντας



επαναλαμβανόμενους στατιστικούς ελέγχους για την ύπαρξη μοναδιαίας ρίζας. Οι στατιστικοί έλεγχοι οι οποίοι είναι διαθέσιμοι είναι ο έλεγχος KPSS, ο έλεγχος ADF και ο έλεγχος Phillips-Perron [70]. Εάν η συνάρτηση επιστρέψει 0 για δοσμένο έλεγχο σημαίνει ότι η χρονοσειρά είναι στάσιμη με βάση τον αντίστοιχο έλεγχο. Η συνάρτηση *nsdiffs()* υπολογίζει τον αριθμό των εποχικών διαφορών ώστε να γίνει μία χρονοσειρά στάσιμη (ως προς την εποχικότητα) κάνοντας επαναλαμβανόμενους στατιστικούς ελέγχους για την ύπαρξη εποχικής μοναδιαίας ρίζας. Οι στατιστικοί έλεγχοι οι οποίοι είναι διαθέσιμοι είναι ο έλεγχος SEAS [71] στον οποίο μετρείται η «ένταση» της εποχικής συνιστώσας, ο έλεγχος των Canova-Hansen [72], ο έλεγχος Hylleberg, Engle, Granger & Yoo [73] και τέλος ο έλεγχος Osborn-Chui-Smith-Birchenhall [74]. Εάν η συνάρτηση επιστρέψει 0 για δοσμένο έλεγχο σημαίνει ότι η χρονοσειρά είναι στάσιμη (δεν έχει εποχική συνιστώσα) με βάση τον αντίστοιχο έλεγχο.

Για το τελικό σύνολο δεδομένων:

Όσον αφορά τον έλεγχο στασιμότητας ως προς την συνιστώσα τάσης – έλεγχος μοναδιαίας ρίζας και για τους 3 διαφορετικούς ελέγχους που αναφέρθηκαν πιο πάνω η συνάρτηση *ndiffs()* επέστρεψε την τιμή 0. Όσον αφορά τον έλεγχο στασιμότητας ως προς την εποχική συνιστώσα – έλεγχος εποχικής μοναδιαίας ρίζας και για τους 4 διαφορετικούς ελέγχους που αναφέρθηκαν πιο πάνω η συνάρτηση *nsdiffs()* επέστρεψε την τιμή 1.

Για το εποχιακά προσαρμοσμένο σύνολο δεδομένων:

Όσον αφορά τον έλεγχο στασιμότητας ως προς την συνιστώσα τάσης – έλεγχος μοναδιαίας ρίζας για τον έλεγχο kpss η συνάρτηση *ndiffs()* επέστρεψε την τιμή 1 ενώ για τους υπόλοιπους ελέγχους επέστρεψε την τιμή 0. Όσον αφορά τον έλεγχο στασιμότητας ως προς την εποχική συνιστώσα – έλεγχος εποχικής μοναδιαίας ρίζας και για τους 4 διαφορετικούς ελέγχους που αναφέρθηκαν πιο πάνω η συνάρτηση *nsdiffs()* επέστρεψε την τιμή 0.

Άρα χρησιμοποιώντας την ανάλυση συνιστωσών και αφαιρώντας την εποχική συνιστώσα, τα εποχιακά προσαρμοσμένα δεδομένα δεν έχουν πλέον εποχικότητα και υπάρχουν ενδείξεις ότι και η συνιστώσα τάσης είναι στατιστικά μη σημαντική.

Η ανάλυση σε συνιστώσες μίας χρονοσειράς μπορεί να χρησιμοποιηθεί και για την μέτρηση της «ισχύς» της τάσης και της εποχικότητας σε μία χρονοσειρά [75]. Όσον αφορά στην προσθετική παραδοχή των συνιστωσών, για να θεωρηθεί η συνιστώσα τάσης ισχυρή, τα εποχιακά

προσαρμοσμένα δεδομένα θα πρέπει να έχουν πολύ μεγαλύτερη διακύμανση από την υπολείπουσα συνιστώσα. Γι' αυτόν ακριβώς τον λόγο ο λόγος  $Var(R_t)/Var(T_t + R_t)$  πρέπει να είναι σχετικά μικρός. Για δεδομένα με μικρή ή χωρίς τάση, οι δύο διακυμάνσεις πρέπει να είναι σχεδόν οι ίδιες [57]. Έτσι ορίζεται η ισχύς της συνιστώσας τάσης ως:

$$F_t = \max\left(0, 1 - \frac{Var(R_t)}{Var(T_t + R_t)}\right) \quad (3.5)$$

Η παραπάνω ποσότητα θα δώσει μία μετρική της ισχύς της συνιστώσας τάσης μεταξύ του 0 και 1. Επειδή η διασπορά της υπολείπουσας συνιστώσας μπορεί περιστασιακά να είναι μεγαλύτερη από την διασπορά των εποχιακά προσαρμοσμένων δεδομένων, τίθεται η ελάχιστη δυνατή τιμή του  $F_t$  να είναι ίση με μηδέν. Η ισχύς της εποχικής συνιστώσας ορίζεται όμοια, αλλά συναρτηθεί των δεδομένων στα οποία έχει αφαιρεθεί η συνιστώσα τάσης και όχι στα εποχιακά προσαρμοσμένα δεδομένα:

$$F_S = \max\left(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)}\right) \quad (3.6)$$

Μία χρονοσειρά με εποχική ισχύ  $F_S$  κοντά στο 0 δεν έχει σχεδόν καθόλου εποχικότητα, ενώ μία χρονοσειρά με ισχυρή εποχικότητα θα έχει τιμή  $F_S$  κοντά στο 1 διότι η ποσότητα  $Var(R_t)$  θα είναι πολύ μικρότερη από την ποσότητα  $Var(S_t + R_t)$  [57].

Οι παραπάνω δοσμένες σχέσεις υπολογίστηκαν στην R για το τελικό σύνολο δεδομένων και τα αποτελέσματα είναι τα ακόλουθα:

$$F_t = 0.029 \text{ (3ο δεκαδικό)}$$

$$F_S = 0.752 \text{ (3ο δεκαδικό)}$$

Επιβεβαιώνεται και με αυτόν τον έλεγχο ότι τα δεδομένα έχουν ισχυρή εποχική συνιστώσα και σχεδόν καθόλου συνιστώσα τάσης.

Στην συνέχεια θα γίνει μία συνοπτική συζήτηση για την χρησιμότητα των μαθηματικών μετασχηματισμών που χρησιμοποιούνται στην στατιστική. Πολλές φορές μετασχηματίζοντας τα ιστορικά δεδομένα μπορεί να οδηγήσει σε ένα απλούστερο πρόβλημα πρόβλεψης. Σκοπός αυτών των μετασχηματισμών είναι να διορθώνουν τα μοτίβα των δεδομένων και να τα κάνουν πιο

συναφή σε όλο το μήκος της χρονοσειράς οδηγώντας πολλές φορές σε πιο ακριβείς προβλέψεις [57].

Αν τα δεδομένα έχουν διακύμανση η οποία αυξάνεται ή μειώνεται κατά μήκος της χρονοσειράς, τότε κάποιος μετασχηματισμός μπορεί να φανεί χρήσιμος. Για παράδειγμα, ο λογαριθμικός μετασχηματισμός είναι συχνά χρήσιμος. Αν συμβολίσουμε τις αρχικές παρατηρήσεις με  $y_1, \dots, y_T$  και τις μετασχηματισμένες με  $w_1, \dots, w_T$  τότε ισχύει ότι  $w_t = \log(y_t)$ . Οι λογάριθμοι είναι χρήσιμοι πολλές φορές και λόγω της ευκολίας στην ερμηνεία τους, δηλαδή αλλαγές στην λογαριθμοποιημένη τιμή είναι σχετικές (ποσοστά) αλλαγές στην αρχική κλίμακα. Άλλο ένα θετικό χαρακτηριστικό των λογαριθμικών μετασχηματισμών είναι ότι περιορίζουν τις προβλέψεις να παραμένουν θετικές στην αρχική κλίμακα [57].

Πολλές φορές και άλλοι τύποι μετασχηματισμών είναι χρήσιμοι, παρότι όχι τόσο ερμηνεύσιμοι. Για παράδειγμα, τετραγωνικές και κυβικές ρίζες μπορεί να αποτελούν χρήσιμους μετασχηματισμούς, οι οποίοι αποκαλούνται μετασχηματισμοί εις την δύναμη (power transformations) διότι μπορούν να γραφούν ως  $w_t = y_t^p$ . Μία χρήσιμη οικογένεια μετασχηματισμών είναι οι Box-Cox μετασχηματισμοί οι οποίοι περιλαμβάνουν και μετασχηματισμούς εις την δύναμη και λογαριθμικούς μετασχηματισμούς και εξαρτώνται από τη παράμετρο  $\lambda$ . Οι εν λόγω μετασχηματισμοί ορίζονται ως:

$$w_t = \begin{cases} \log(y_t) & \text{αν } \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda} & \text{διαφορετικά.} \end{cases} \quad (3.7)$$

Ο λογάριθμος στον μετασχηματισμό Box-Cox είναι πάντα ο φυσικός λογάριθμος (βάση η σταθερά του *Euler*). Αν το  $\lambda = 0$ , τότε χρησιμοποιείται ο λογαριθμικός μετασχηματισμός με φυσικό λογάριθμο, και αν  $\lambda \neq 0$  τότε χρησιμοποιείται μετασχηματισμός εις την δύναμη και ακολουθείται συνήθως υπό κάποια απλή κλίμακα. Αν το  $\lambda = 1$ , τότε  $w_t = y_t - 1$ , και τα μετασχηματισμένα δεδομένα απλά μετατοπίζονται κατά μία μονάδα προς τα κάτω χωρίς να υπάρχει καμία αλλαγή στο σχήμα της χρονοσειράς. Για όλες τις υπόλοιπες τιμές της παραμέτρου  $\lambda$ , η χρονοσειρά αλλάζει σχήμα [57].

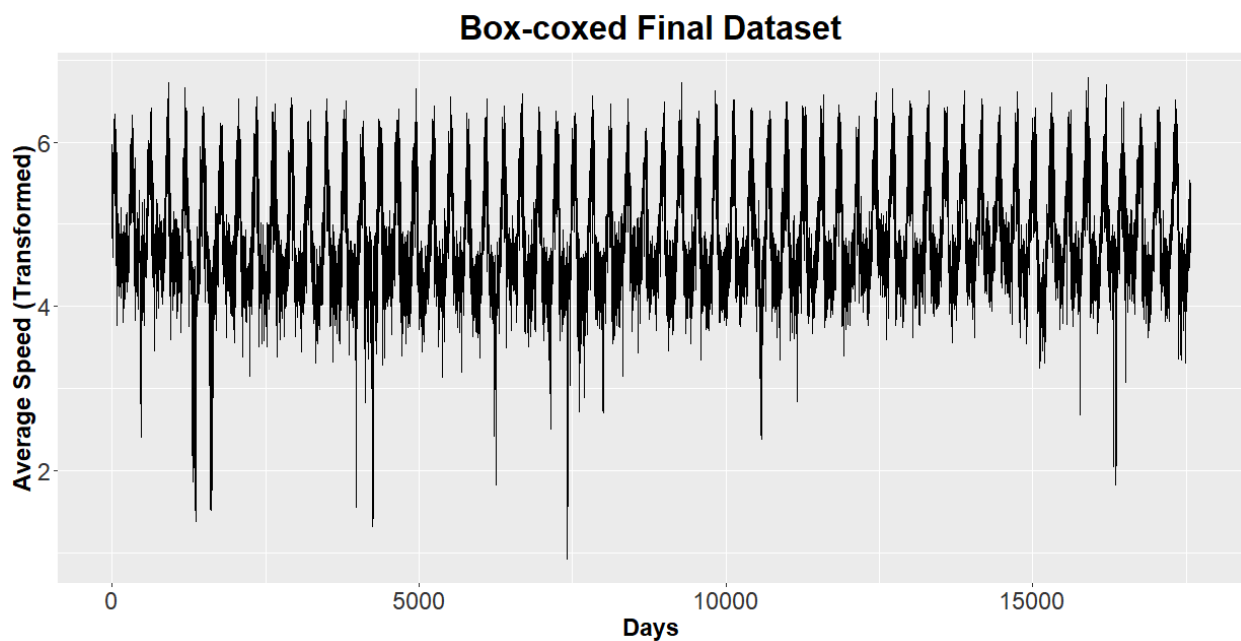
Στην βιβλιοθήκη *forecast* υπάρχει η εντολή *BoxCox.lambda()* που υπολογίζει αυτόματα την τιμή της παραμέτρου  $\lambda$  έτσι ώστε να ελαχιστοποιείται ο συντελεστής διακύμανσης

χρησιμοποιώντας την μέθοδο Guerrero [76]. Εάν για το  $\lambda$  επιστραφεί η τιμή 1 τότε δεν χρειάζεται να γίνει κάποιος μετασχηματισμός στα δεδομένα. Εφαρμόζοντας στο τελικό σύνολο δεδομένων και στα εποχιακά προσαρμοσμένα δεδομένα την παραπάνω συνάρτηση, επιστρέφονται οι εξής τιμές για το  $\lambda$  σε κάθε μία περίπτωση (με ακρίβεια 3<sup>ου</sup> δεκαδικού):

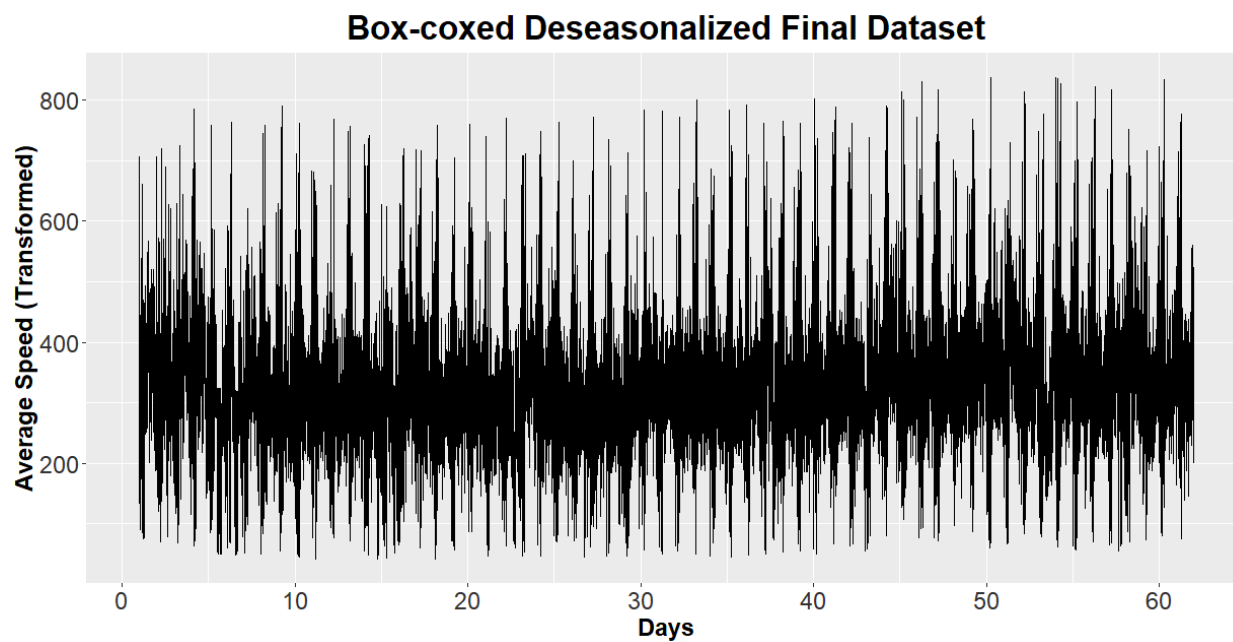
Τελικό σύνολο δεδομένων:  $\lambda = 0.212$

Εποχιακά προσαρμοσμένα δεδομένα:  $\lambda = 1.908$

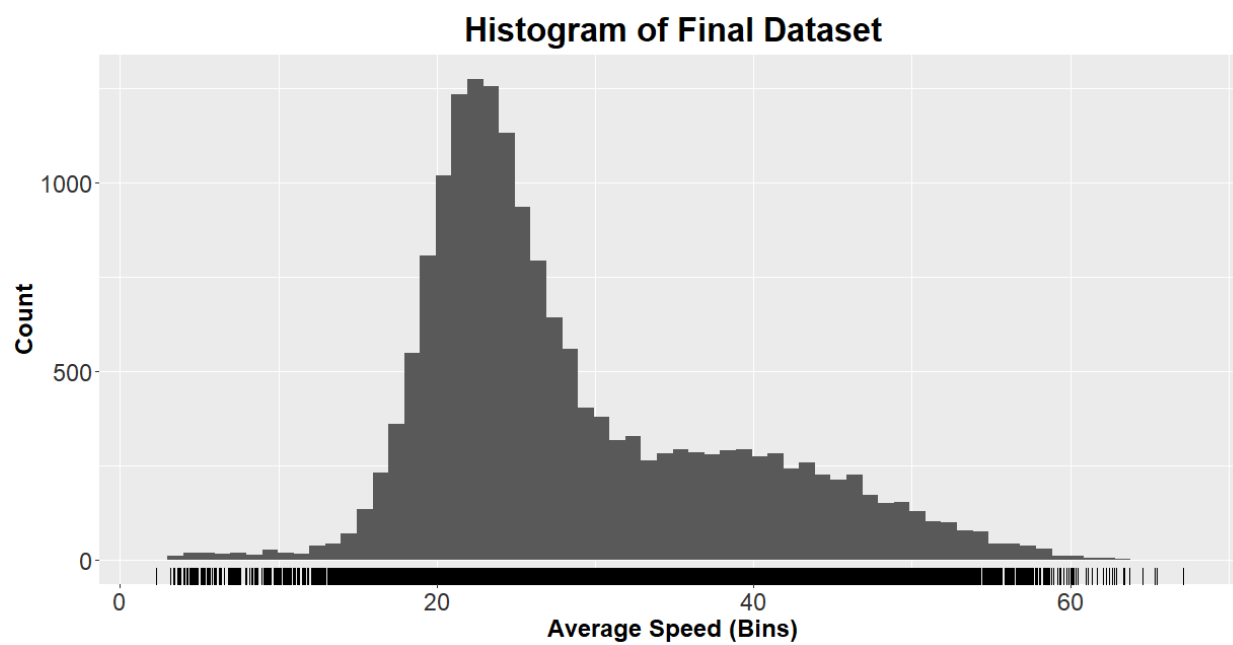
Στο Διάγραμμα 39 και 40 φαίνονται το μετασχηματισμένο με την μέθοδο Box-Cox τελικό σύνολο δεδομένων και τα μετασχηματισμένα με την μέθοδο Box-Cox εποχιακά προσαρμοσμένα δεδομένα. Στο Διάγραμμα 40 παρατηρούμε ότι το εύρος των δεδομένων κατά μήκος της χρονοσειράς είναι σταθερό σε σχέση με τα δεδομένα του Διαγράμματος 39. Η αφαίρεση της εποχικής συνιστώσας από τα δεδομένα, αρχικά, και μετασχηματίζοντας τα δεδομένα κατά Box-Cox, έπειτα, παρατηρείται ότι σταθεροποιούνται κάπως οι ιδιότητες της χρονοσειράς.



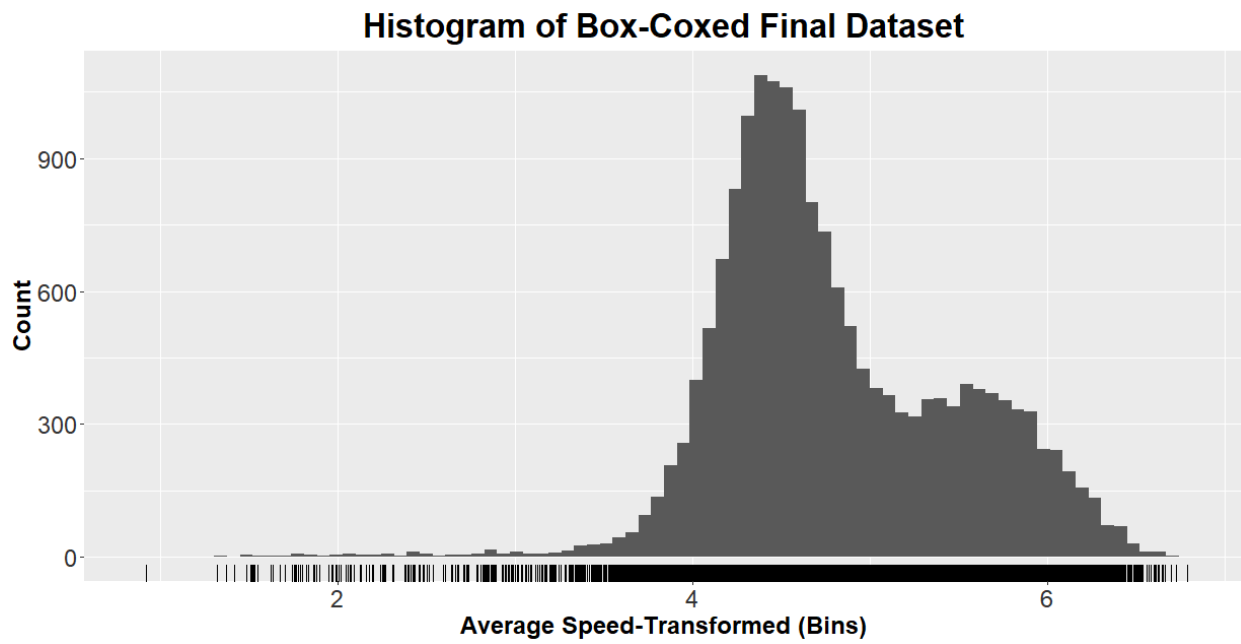
**Διάγραμμα 39** Το τελικό σύνολο δεδομένων μετασχηματισμένο κατά *Box-Cox*.



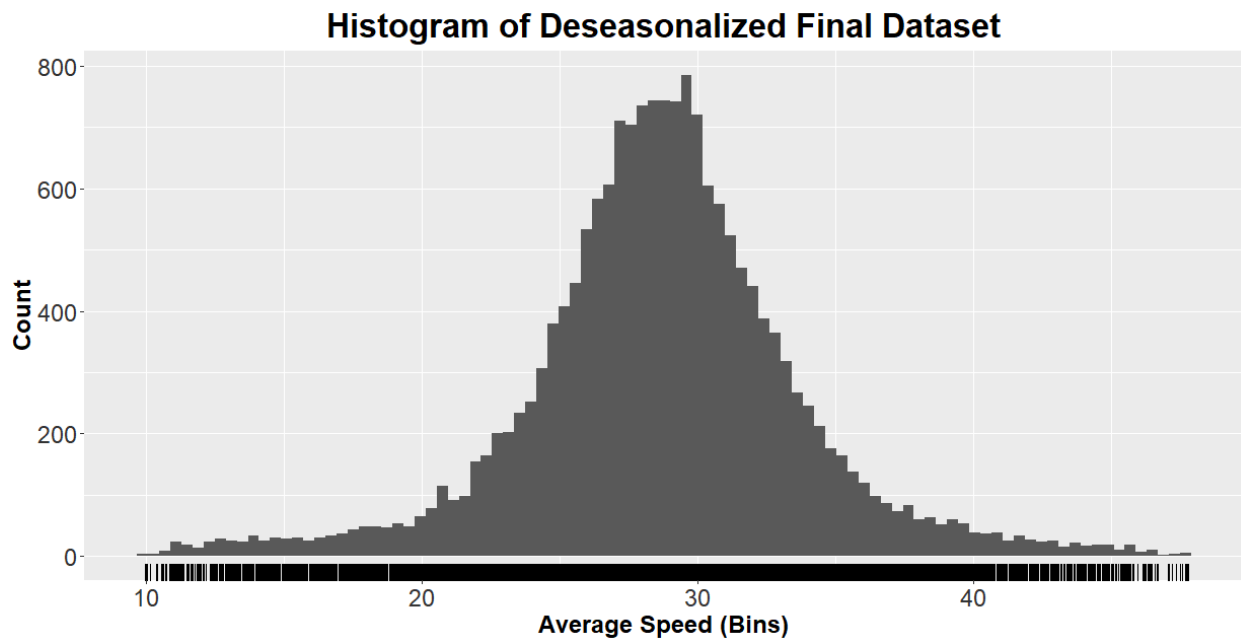
**Διάγραμμα 40** Τα εποχικά προσαρμοσμένα δεδομένα μετασχηματισμένα κατά *Box-Cox*.



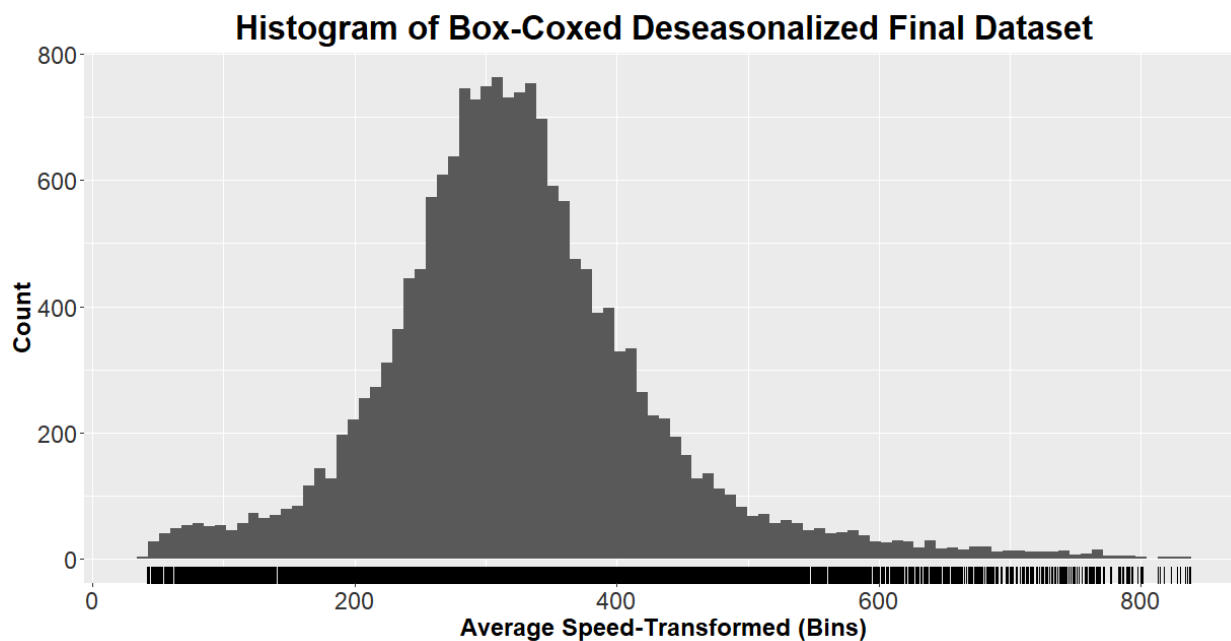
**Διάγραμμα 41** Το ιστόγραμμα του τελικού συνόλου δεδομένων.



**Διάγραμμα 42** Το ιστόγραμμα του τελικού συνόλου δεδομένων μετασχηματισμένο κατά *Box-Cox*.



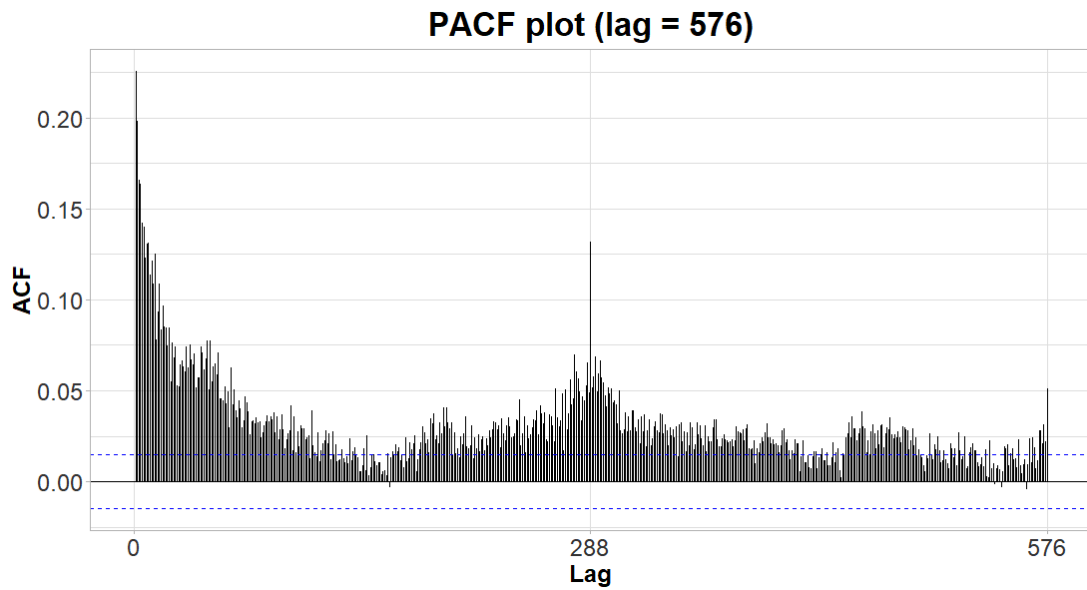
**Διάγραμμα 43** Το ιστόγραμμα των εποχιακά προσαρμοσμένων δεδομένων.



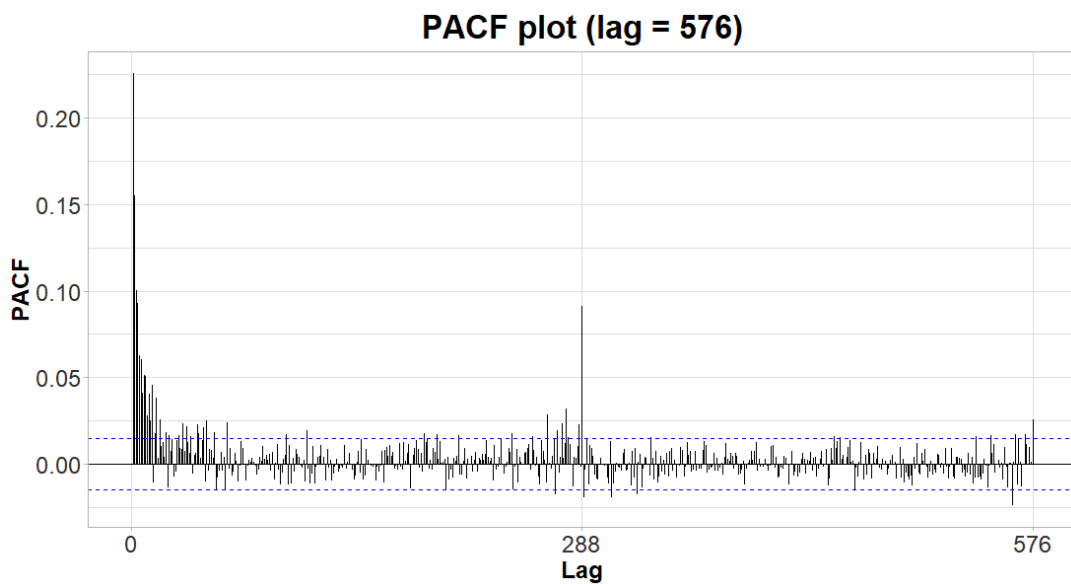
**Διάγραμμα 44** Το ιστόγραμμα των εποχιακά προσαρμοσμένων δεδομένων μετασχηματισμένο κατά *Box-Cox*.

Στα Διαγράμματα 41 και 42 παρατηρούμε το ιστόγραμμα του τελικού συνόλου δεδομένων και το μετασχηματισμένο τελικό σύνολο δεδομένων. Στο 1<sup>ο</sup> από τα δύο διαγράμματα φαίνεται να έχουμε κάπως λοξή κατανομή ενώ μετά τον μετασχηματισμό τείνει προς δικόρυφη κατανομή (ίσως μίξη κανονικών κατανομών). Αυτό συμβαίνει διότι ουσιαστικά υπάρχουν 2 ομάδες ταχυτήτων, τις υψηλές ταχύτητες τις βραδινές ώρες και τις χαμηλές ταχύτητες τις ώρες αιχμής. Στα Διαγράμματα 43 και 44 φαίνεται το ιστόγραμμα των εποχιακά προσαρμοσμένων δεδομένων. Παρατηρούμε ότι αφαιρώντας την εποχική συνιστώσα τα δεδομένα πλέον ακολουθούν κανονική κατανομή ιδιότητα που είναι επιθυμητή.

Στην συνέχεια επαναλήφθηκε το διάγραμμα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για τα εποχιακά προσαρμοσμένα δεδομένα και τα αντίστοιχα μετασχηματισμένα για τον έλεγχο συσχετίσεων που πιθανώς να έχουν απομείνει μεταξύ των υστερήσεων. Όπως φαίνεται στα παρακάτω 4 Διαγράμματα (45-48) έχουν μείνει συσχετίσεις τις οποίες θα πρέπει να εξηγήσουν οι μέθοδοι πρόβλεψης.

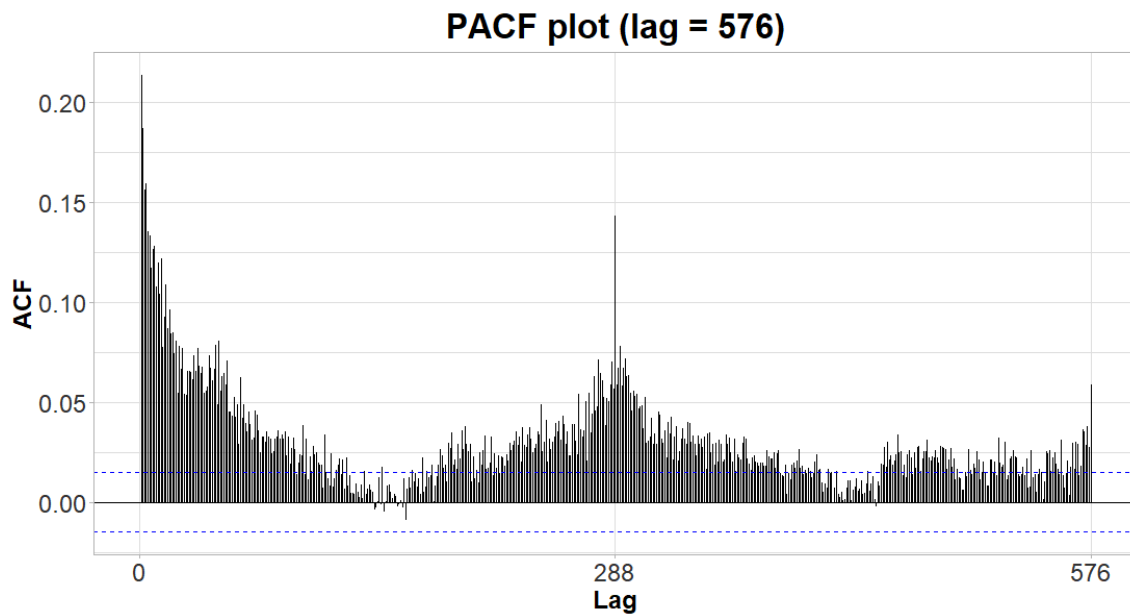


**Διάγραμμα 45** Το Διάγραμμα αυτοσυσχέτισης για τα εποχιακά προσαρμοσμένα δεδομένα.

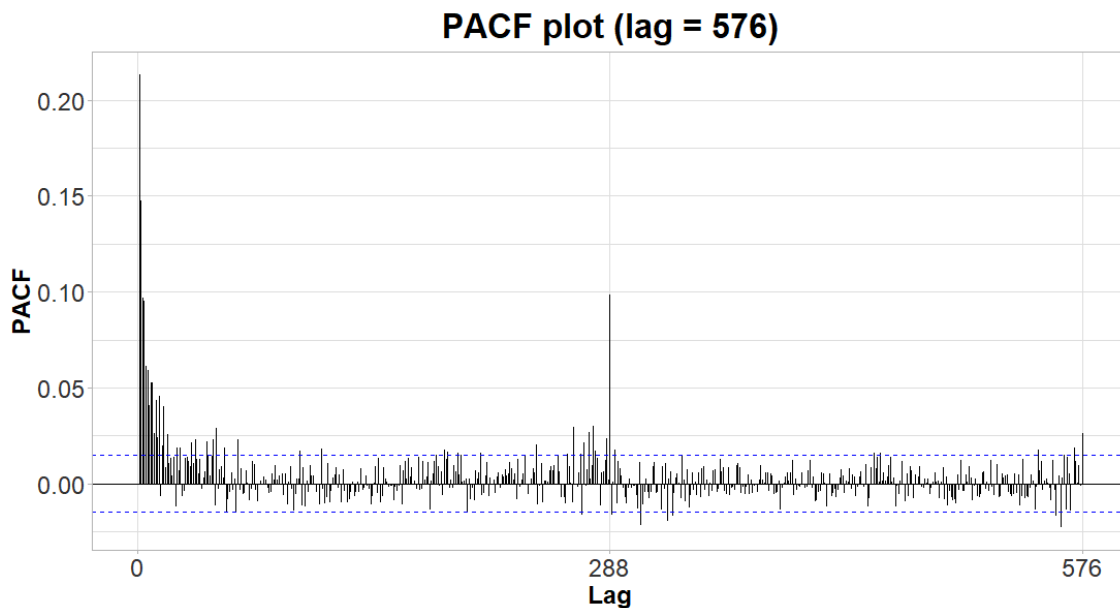


**Διάγραμμα 46** Το Διάγραμμα μερικής αυτοσυσχέτισης για τα εποχιακά προσαρμοσμένα δεδομένα.





**Διάγραμμα 47** Το Διάγραμμα αυτοσυσχέτισης για τα μετασχηματισμένα εποχιακά προσαρμοσμένα δεδομένα.

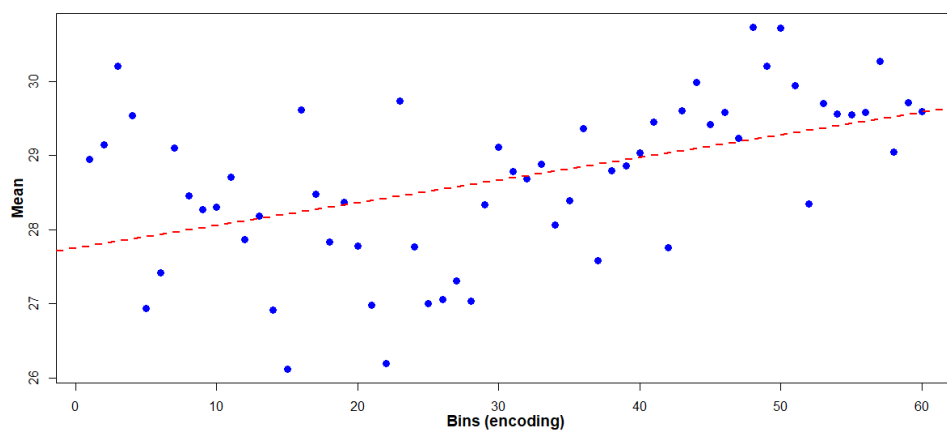


**Διάγραμμα 48** Το Διάγραμμα μερικής αυτοσυσχέτισης για τα μετασχηματισμένα εποχιακά προσαρμοσμένα δεδομένα.

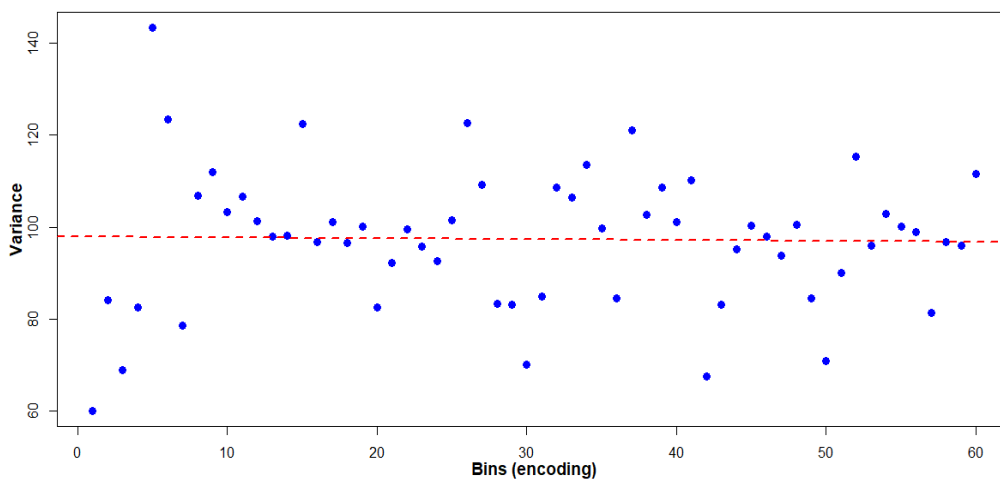
Θα γίνει εφαρμογή μίας σχετικά πρόχειρης μεθόδου για την εξέταση της ύπαρξης σταθερής διασποράς στα δεδομένα κοιτώντας κυρίως αν υπάρχει κάποια τάση αύξησης ή μείωσης της διασποράς κατά μήκος της χρονοσειράς, καθώς είναι κάπως θολές οι συστάσεις που δίνονται από την επιστημονική κοινότητα για αυτόν τον έλεγχο. Επί τη ευκαιρία θα επαναληφθεί η ίδια διαδικασία και για τον μέσο. Η διαδικασία αυτή πραγματοποιήθηκε για το τελικό σύνολο δεδομένων, τα εποχιακά προσαρμοσμένα δεδομένα, μετασχηματισμένων και μη, και για τα δύο σύνολα. Τα βήματα που ακολουθήθηκαν είναι τα εξής:

- Αρχικά τα δεδομένα χωρίστηκαν σε 60 υποσύνολα από το αρχικό σύνολο, σεβόμενοι την χρονική σειρά των δεδομένων.
- Για κάθε ένα από τα υποσύνολα υπολογίζεται η μέση τιμή και η διασπορά.
- Προσαρμόζεται μία ευθεία για τον εντοπισμό ενδεχόμενης τάσης.

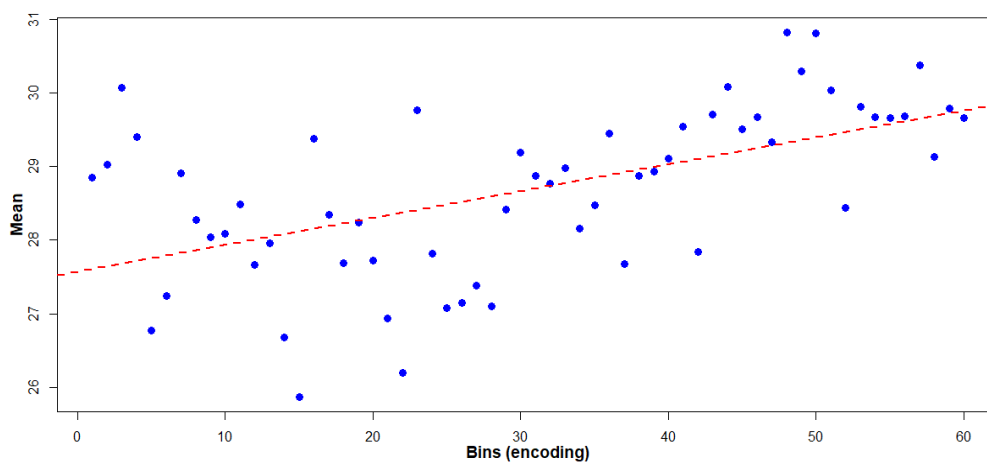
Και για τα 3 διαφορετικά σύνολα όσον αφορά τον μέσο φαίνεται να υπάρχει κάποια αυξητική τάση η οποία βέβαια στους στατιστικούς ελέγχους μοναδιαίας ρίζας βρέθηκε να είναι στατιστικά μη σημαντική. Όσον αφορά την διασπορά το τελικό σύνολο δεδομένων και το μετασχηματισμένο εποχιακά προσαρμοσμένο σύνολο δεδομένων φαίνεται να έχει σχετικά σταθερή διασπορά ενώ των εποχιακά προσαρμοσμένων δεδομένων φαίνεται να έχει μία μικρή μειωτική τάση. Ωστόσο αυτός ο γραφικός έλεγχος είναι ενδεικτικός και πολλές φορές όχι τόσο αξιόπιστος. Βέβαια η διαδικασία ελέγχου σταθερής διασποράς πέρα από τις προφανείς περιπτώσεις είναι αρκετά περίπλοκη και ασαφής και ξεφεύγει εκτός των πλαισίων της διπλωματικής εργασίας.



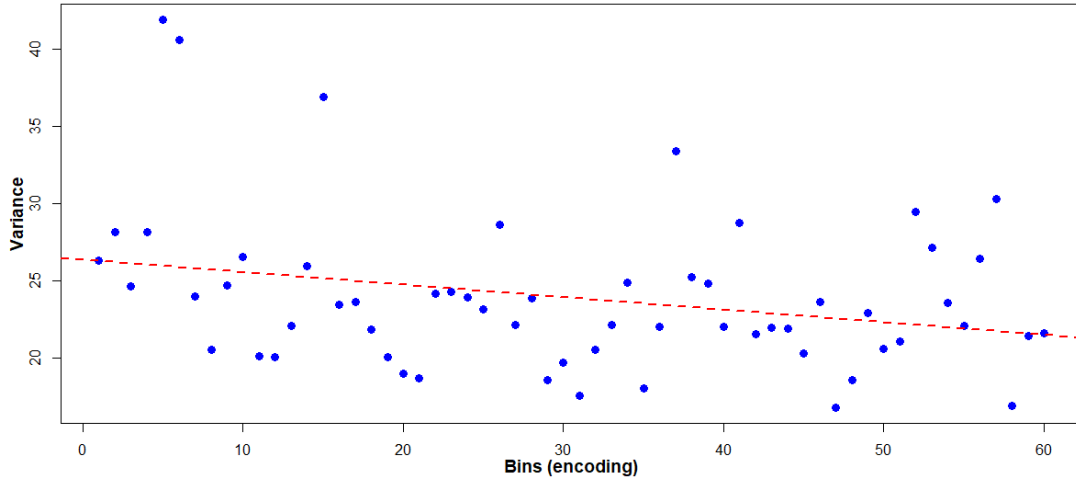
**Διάγραμμα 49** Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το τελικό σύνολο δεδομένων.



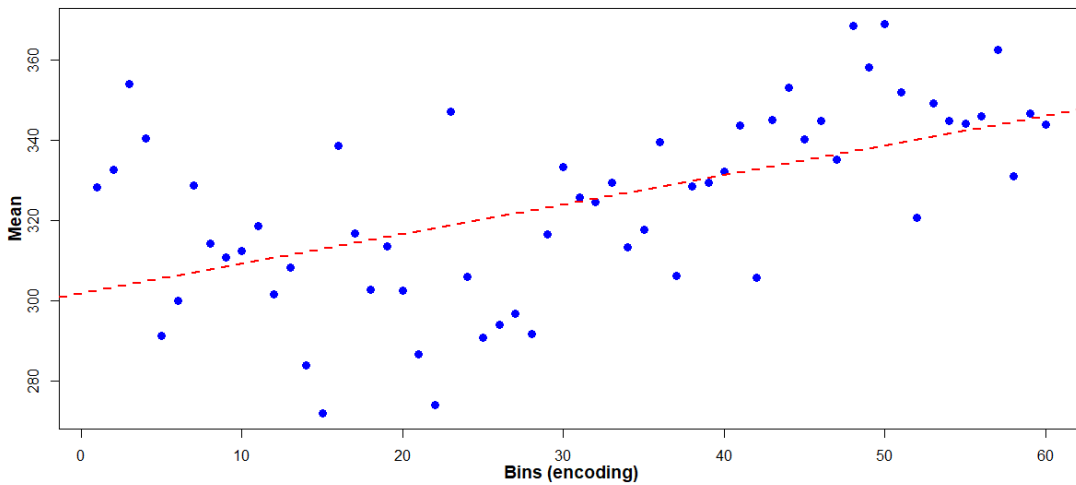
**Διάγραμμα 50** Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το τελικό σύνολο δεδομένων.



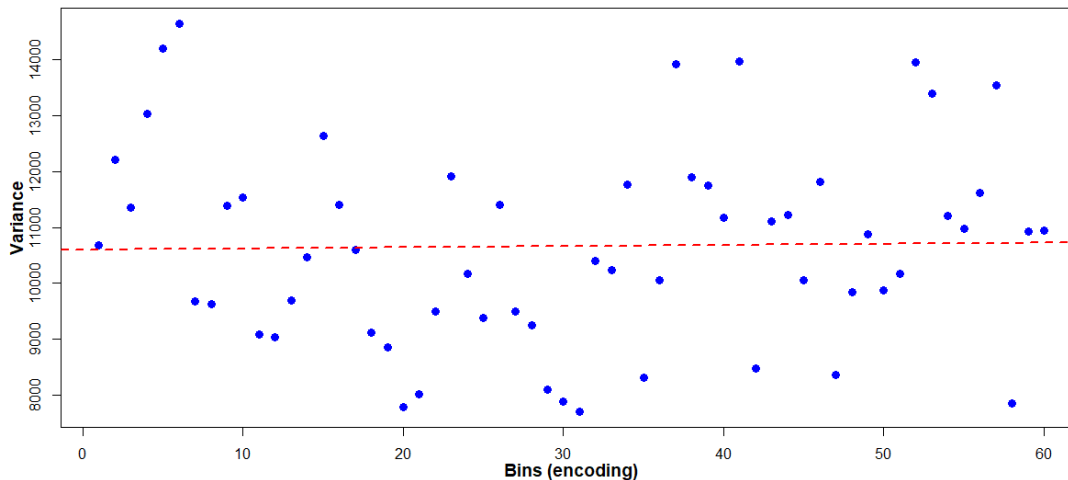
**Διάγραμμα 51** Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχικά προσαρμοσμένων δεδομένων.



**Διάγραμμα 52** Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων δεδομένων.



**Διάγραμμα 53** Ο μέσος για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων και μετασχηματισμένων δεδομένων.



**Διάγραμμα 54** Η διασπορά για κάθε ένα από τα υποσύνολα στα οποία χωρίστηκε το σύνολο των εποχιακά προσαρμοσμένων και μετασχηματισμένων δεδομένων.

### 3.9 ΠΡΟΒΛΕΨΙΜΟΤΗΤΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

Πολλές φορές είναι χρήσιμο να υπάρχει κάποια ένδειξη για το πόσο δύσκολο ή εύκολο είναι να γίνει πρόβλεψη σε μία χρονοσειρά. Η ικανότητα να μπορεί κάποιος με ακρίβεια εκ των προτέρων να προσδιορίσει την προβλεψιμότητα μίας χρονοσειράς είναι σημαντικό πλεονέκτημα καθώς μπορεί να δώσει καθοδήγηση για την πιθανή ακρίβεια της πρόβλεψης καθώς και για τον βαθμό δυσκολίας και προσπάθειας που απαιτείται για την επίλυση του προβλήματος [77]. Στην εργασία τους οι *Yentes et al.* [77] έδειξαν ότι χαμηλές τιμές της εντροπίας στο σύνολο εκπαίδευσης συσχετίζονται με χαμηλές τιμές του σφάλματος στο σύνολο ελέγχου όσον αφορά τουλάχιστον την μετρική σφάλματος MASE. Χρησιμοποιώντας την βιβλιοθήκη `graspa` και την εντολή `approx_entropy()`, που βασίζεται στο [77], υπολογίστηκε η εντροπία των δεδομένων σε 4 διαφορετικά σύνολα:

- Μία συνημιτονοειδή καμπύλη με μικρό ποσοστό τυχαίου θορύβου (αναμένεται μικρή εντροπία).
- Το τελικό σύνολο δεδομένων.
- Τα εποχιακά προσαρμοσμένα δεδομένα.
- Μία ακολουθία με γκαουσιανό θόρυβο (αναμένεται μεγάλη εντροπία).

Τα αποτελέσματα για κάθε ένα από τα παραπάνω σύνολα είναι:

- Εντροπία συνημιτονοειδούς καμπύλης: 0.327
- Εντροπία τελικού συνόλου δεδομένων: 1.494
- Εντροπία εποχιακά προσαρμοσμένων δεδομένων: 2.045
- Εντροπία γκαουσιανού θορύβου: 2.754

Παρατηρούμε ότι αφού αφαιρεθεί από τα δεδομένα η εποχική συνιστώσα, τα εποχιακά προσαρμοσμένα δεδομένα έχουν μεγάλη εντροπία δηλαδή η πρόβλεψη εκπαιδύοντας ένα μοντέλο με αυτά τα δεδομένα δεν θα είναι εύκολο πρόβλημα καθώς βρίσκεται πιο κοντά στην τιμή της εντροπίας του γκαουσιανού θορύβου. Όπως ήταν λογικό και οι 2 τιμές της εντροπίας για το τελικό σύνολο δεδομένων βρίσκονται ανάμεσα σε μία σχεδόν ντετερμινιστική συνάρτηση (συνημιτονοειδή καμπύλη με τυχαίο θόρυβο) και στο διάνυσμα που περιέχει τιμές γκαουσιανού θορύβου.

## 4 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΜΕΘΟΔΩΝ ΠΡΟΒΛΕΨΗΣ

---

Στο κεφάλαιο 4 θα γίνει συζήτηση για την προτυποποίηση του προβλήματος, για τα μοντέλα που χρησιμοποιήθηκαν για το πρόβλημα της βραχυπρόθεσμης πρόβλεψης κυκλοφορίας αναφέροντας και κάποια θεωρητικά στοιχεία εν συντομία, για την μέθοδο και τις μετρικές αξιολόγησης της προβλεπτικής ικανότητας των μοντέλων, της προσαρμογής στα δεδομένα και της υπολογιστικής και χρονικής πολυπλοκότητας κάθε μοντέλου. Επιπροσθέτως, θα αναφέρουμε τις μεθόδους προεπεξεργασίας των δεδομένων (βλέπε και κεφάλαιο 3) που χρησιμοποιήθηκαν για κάθε μέθοδο. Όσον αφορά τις μεθόδους πρόβλεψης θα χωριστούν σε τρεις ομάδες. Στην πρώτη ομάδα είναι μία μόνο μέθοδος, η μέθοδος αναφοράς. Ένα μοντέλο θα γίνεται αποδεκτό αν έχει καλύτερη προβλεπτική ικανότητα από την μέθοδο αναφοράς. Στην 2<sup>η</sup> ομάδα ανήκουν οι κλασσικές στατιστικές μέθοδοι πρόβλεψης χρονοσειρών για την υλοποίηση των οποίων θα ακολουθηθεί πλήρως αυτοματοποιημένη διαδικασία. Στην ομάδα 2 αναπτύχθηκαν 5 διαφορετικά μοντέλα. Τέλος, στην 3<sup>η</sup> ομάδα ανήκουν οι μέθοδοι βαθιάς μηχανικής μάθησης. Στην ομάδα 3 αναπτύχθηκαν 10 διαφορετικές μέθοδοι. Να σημειωθεί για λόγους επιστημονικής συνέπειας τα ονόματα των μεθόδων θα αναφέρονται στην αγγλική γλώσσα.

### 4.1 ΜΕΘΟΔΟΙ ΒΡΑΧΥΠΡΟΘΕΣΜΗΣ ΠΡΟΒΛΕΨΗΣ ΚΥΚΛΟΦΟΡΙΑΣ

Αρχικά θα γίνει αναφορά ονομαστικά στις μεθόδους κάθε ομάδας και στην συνέχεια θα ακολουθήσει η ανάλυση ανά ομάδα κάθε μίας μεθόδου.

#### **Μέθοδος Αναφοράς (ομάδα 1)**

- Μέθοδος Seasonal Naive

#### **Στατιστικές Μέθοδοι (ομάδα 2)**

- Μέθοδος Arima + Εποχική αποσύνθεση (STL+Arima)
- Δυναμική Αρμονική Παλινδρόμηση με σφάλματα ARIMA.
- Μέθοδος ETS + Εποχική αποσύνθεση (STL+ETS)
- Μέθοδος TBATS (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components)
- Μέθοδος Theta

## Μέθοδοι Βαθιάς Μηχανικής Μάθησης (ομάδα 3)

- Multi-layer Perceptron (MLP)
- Long-Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Convolutional Neural Networks (CNN)
- Dilated Convolutional Neural Networks (Dil-CNN)
- CNN-LSTM
- Convolutional LSTM (Conv-LSTM)
- Encoder-Decoder LSTM
- Encoder-Decoder CNN-LSTM
- Encoder-Decoder Conv-LSTM

Τέλος, θα αναφερόμαστε σε κάθε ομάδα είτε ως ομάδα 1,2,3 είτε με την κλάση μεθόδων που την αποτελούν.

## 4.2 ΜΕΘΟΔΟΣ ΑΝΑΦΟΡΑΣ

### 4.2.1 Μέθοδος Seasonal Naive

Η μέθοδος αυτή ανήκει στην ομάδα των μεθόδων αναφοράς και εφαρμόζεται συνήθως πρώτη κατά την προσέγγιση ενός προβλήματος πρόβλεψης. Κάθε άλλη μέθοδος πρέπει να έχει καλύτερη προβλεπτική ικανότητα από την εν λόγω μέθοδο για να γίνει αποδεκτή. Η εποχική μέθοδος Naive προτιμάται έναντι της «απλής» μεθόδου Naive όταν τα δεδομένα έχουν έντονη εποχική συνιστώσα. Στην περίπτωση αυτής της μεθόδου μία πρόβλεψη είναι ίση με την τιμή της τελευταίας παρατήρησης από την ίδια εποχή της προηγούμενης περιόδου, δηλαδή για εποχική συχνότητα ίση με 288 η πρόβλεψη της μέσης ταχύτητας για το 5λεπτο 13:05 της ημέρας Πέμπτης, με διαθέσιμα τα δεδομένα μέχρι και τη 13:00, ισούται με την τιμή του αντίστοιχου 5λέπτου για 288 5λεπτα πριν, δηλαδή 13:05 της ημέρας Τετάρτης. Μαθηματικά η πρόβλεψη για τον χρόνο  $T + h$  μπορεί να γραφεί ως:

$$\hat{Y}_{T+h|T} = Y_{T+h-m(k+1)} \quad (4.1)$$



όπου  $m$  είναι η εποχική συχνότητα και  $k$  είναι το ακέραιο μέρος της διαίρεσης  $\frac{h-1}{m}$  (δηλαδή ο αριθμός των πλήρων ημερών στην περίοδο πρόβλεψης πριν την χρονική στιγμή  $T + h$  [57]). Για την εφαρμογή αυτής της μεθόδου έγινε χρήση της βιβλιοθήκης *forecast* στην R και της εντολής *snaiive()*.

### 4.3 ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ

Αρχικά θα γίνει η θεωρητική περιγραφή των διαδικασιών *Arima* και στην συνέχεια θα εξειδικεύσουμε στις μεθόδους που χρησιμοποιήθηκαν οι οποίες βασίζονται σε εκείνη καθώς και στις υπόλοιπες μεθόδους της ομάδας 2.

#### 4.3.1 Αυτοπαλινδρομες διαδικασίες

Σε ένα μοντέλο πολυμεταβλητής παλινδρόμησης, γίνεται η πρόβλεψη της εξαρτημένης μεταβλητής χρησιμοποιώντας έναν γραμμικό συνδυασμό από άλλες μεταβλητές [57]. Σε ένα μοντέλο αυτοπαλινδρόμησης, κάνουμε πρόβλεψη για την μεταβλητή ενδιαφέροντος χρησιμοποιώντας παρελθοντικές τιμές της ίδιας μεταβλητής, εξ' ου και ο όρος αυτοπαλινδρόμηση. Ένα μοντέλο αυτοπαλινδρόμησης τάξεως  $p$  μπορεί να γραφεί ως,

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (4.2)$$

όπου  $\varepsilon_t$  είναι λευκός θόρυβος. Η παραπάνω σχέση μοιάζει με πολλαπλή παλινδρόμηση αλλά ως μη εξαρτημένες μεταβλητές χρησιμοποιούνται οι τιμές υστέρησης της  $y_t$ . Το παραπάνω μοντέλο θα αναφέρεται ως *AR(p)* μοντέλο. Συνήθως τα αυτοπαλινδρομα μοντέλα περιορίζονται σε στάσιμα δεδομένα, όπου για να επιτευχθεί αυτό εισάγονται κάποιοι περιορισμοί στις τιμές των παραμέτρων που απαιτούνται. Για τάξη  $p \geq 3$  οι περιορισμοί γίνονται πολύ πιο περίπλοκοι, παρόλα αυτά σε προγραμματιστικό περιβάλλον της R, η βιβλιοθήκη *forecast* κατά τον υπολογισμό του μοντέλου κάνει αυτόματα τους ελέγχους των περιορισμών όσο σύνθετο και αν είναι το μοντέλο.

#### 4.3.2 Διαδικασίες Κινητού Μέσου

Αντί της χρήσης παρελθοντικών τιμών της μεταβλητής πρόβλεψης στην μέθοδο της αυτοπαλινδρόμησης μπορούν να χρησιμοποιηθούν σαν ανεξάρτητες μεταβλητές το σφάλμα πρόβλεψης προηγούμενων χρονικών στιγμών σε μία παλινδρόμηση [57]. Δηλαδή,

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4.3)$$

όπου ο όρος  $\varepsilon_t$  είναι λευκός θόρυβος. Το μοντέλο αυτό ονομάζεται  $MA(q)$ , δηλαδή είναι η μέθοδος κινητού μέσου τάξεως  $q$ . Να σημειωθεί ότι η τιμή του  $\varepsilon_t$  δεν είναι παρατηρήσιμη οπότε δεν είναι η συνηθισμένη μέθοδος παλινδρόμησης. Κάθε τιμή της εξαρτημένης μεταβλητής  $y_t$  είναι ένας σταθμισμένος κινητός μέσος των παρελθοντικών (κάποιων) σφαλμάτων πρόβλεψης. Ας δούμε τώρα πως συσχετίζονται οι αυτοπαλίνδρομες διαδικασίες και η μέθοδος του κινητού μέσου. Είναι δυνατό να γραφεί μία στάσιμη  $AR(p)$  διαδικασία σαν μία  $MA(\infty)$  διαδικασία. Χρησιμοποιώντας επανειλημμένα αντικατάσταση εντός της αναδρομικής εξίσωσης για ένα  $AR(1)$  μοντέλο προκύπτει,

$$\begin{aligned} y_t &= \varphi_1 y_{t-1} + \varepsilon_t = \varphi_1 (\varphi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varphi_1^2 y_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t = \\ &= \varphi_1^3 y_{t-3} + \varphi_1^2 \varepsilon_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t \text{ κλπ.} \end{aligned}$$

Δοσμένου ότι  $-1 < \varphi_1 < 1$ , η τιμή του  $\varphi_1^k$  θα γίνεται όλο και μικρότερη όσο το  $k$  γίνεται μεγαλύτερο. Άρα τελικά προκύπτει,

$$y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \varphi_1^3 \varepsilon_{t-3} + \dots \quad (4.4)$$

το οποίο είναι το  $MA(\infty)$  μοντέλο.

Το αντίστροφο αποτέλεσμα ισχύει αν εισαχθούν κάποιοι περιορισμούς στις  $MA$  παραμέτρους έτσι ώστε το  $MA$  μοντέλο να είναι αντιστρέψιμο (invertible). Έχοντας εξασφαλίσει τα πιο πάνω ισχύει ότι κάθε αντιστρέψιμο  $MA(q)$  μοντέλο μπορεί να γραφεί ως ένα  $AR(\infty)$  μοντέλο. Η ιδιότητα της αντιστρεψιμότητας δεν είναι χρήσιμη μόνο για να μπορεί ένα  $MA(q)$  μοντέλο να γραφεί ως ένα  $AR(\infty)$  μοντέλο αλλά δίνει στο μοντέλο και κάποιες επιθυμητές μαθηματικές ιδιότητες. Μια  $MA$  διαδικασία είναι αντιστρέψιμη εάν  $|\theta| < 1$ , το οποίο οδηγεί στο ότι οι πιο πρόσφατες παρατηρήσεις έχουν μεγαλύτερο βάρος και άρα μεγαλύτερη επιρροή σε σχέση με πιο παρελθοντικές παρατηρήσεις.

Οι περιορισμοί που εισάγονται στις παραμέτρους μίας διαδικασίας  $MA$  έτσι ώστε να είναι αντιστρέψιμη είναι παρόμοιοι με τους αντίστοιχους περιορισμούς ώστε μία  $AR$  διαδικασία να είναι στάσιμη. Για τάξη  $q \geq 3$  οι περιορισμοί γίνονται πολύ περίπλοκοι, παρόλα αυτά σε προγραμματιστικό περιβάλλον της  $R$ , η βιβλιοθήκη `forecast` κατά τον υπολογισμό του μοντέλου κάνει αυτόματα τους ελέγχους των περιορισμών όσο σύνθετο και αν είναι το μοντέλο.

### 4.3.3 Διαφόριση

Μία χρονοσειρά θεωρείται στάσιμη όταν οι ιδιότητες τις δεν εξαρτώνται από τον χρόνο στον οποίο γίνεται η παρατήρηση. Έτσι, χρονοσειρές οι οποίες εμφανίζουν τάση ή και ακόμα και εποχικότητα δεν είναι στάσιμες, διότι αυτά τα δύο προφανώς επηρεάζουν την τιμή της χρονοσειράς σε διαφορετικούς χρόνους.

Μία μέθοδος μετατροπής μίας χρονοσειράς από μη στάσιμη σε στάσιμη είναι η μέθοδος διαφόρισης. Η διαφόριση μπορεί να βοηθήσει στη σταθεροποίηση του μέσου μίας χρονοσειράς με την απομάκρυνση των αλλαγών στο επίπεδο της χρονοσειράς πράγμα που οδηγεί στην απομάκρυνση (ή σημαντική μείωση) της τάσης και της εποχικότητας. Η διαφορίσιμη χρονοσειρά αναπαριστά τις αλλαγές μεταξύ συνεχόμενων παρατηρήσεων της αρχικής χρονοσειράς και μπορεί να γραφεί ως (διαφόριση 1<sup>ης</sup> τάξεως),

$$y'_t = y_t - y_{t-1} \quad (4.5)$$

Η ολοκληρωμένη χρονοσειρά θα έχει μόνο  $T - 1$  τιμές, καθώς δεν είναι δυνατό να υπολογιστεί η διαφορά  $y'_1$  για την 1<sup>η</sup> παρατήρηση. Πολλές φορές μπορεί τα διαφορίσιμα δεδομένα να μην είναι στάσιμα και να χρειάζεται να διαφορίσουμε τα δεδομένα και 2<sup>η</sup> φορά (διαφόριση 2<sup>ης</sup> τάξης) δηλαδή

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad (4.6)$$

Η χρονοσειρά  $y''_t$  θα έχει  $T - 2$  τιμές. Συνήθως, στην πράξη απαιτείται μέχρι διαφόριση 2<sup>ης</sup> τάξης.

Ένας άλλος τύπος διαφόρισης είναι η εποχική διαφόριση, η οποία είναι η διαφορά μεταξύ μίας παρατήρησης και της προηγούμενης παρατήρησης από την αμέσως προηγούμενη εποχή. Ο σκοπός της είναι να μετατρέψει μία χρονοσειρά από μη στάσιμη σε στάσιμη λόγω της εποχικότητας. Γράφουμε,

$$y'_t = y_t - y_{t-m} \quad (4.7)$$

όπου  $m$  είναι η συχνότητα της εποχικότητας. Πολλές φορές είναι απαραίτητο να εφαρμοστεί στα δεδομένα και κανονική και εποχική διαφόριση. Προτείνεται σε χρονοσειρές με έντονη εποχικότητα να γίνεται πρώτα η εποχική διαφόριση διότι πολλές φορές το αποτέλεσμα είναι μία στάσιμη χρονοσειρά και ως προς τη τάση γεγονός που οδηγεί στην μη χρήση της κανονικής διαφόρισης ενώ το ανάποδο δεν ισχύει.

#### 4.3.4 Μη-εποχικά μοντέλα ARIMA

Αν συνδυαστούν οι διαδικασίες αυτοπαλινδρόμησης, κινητού μέσου και η ιδέα της διαφορίσης τότε προκύπτει το μη εποχικό ολοκληρωμένο αυτοπαλίνδρομο κινητού μέσου μοντέλο (AutoRegressive Integrated Moving Average –ARIMA). Η ολοκλήρωση αφορά την αντίστροφη διαδικασία της διαφορίσης. Το μοντέλο ARIMA γράφεται ως,

$$y'_t = c + \varphi_1 y'_{t-1} + \dots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4.8)$$

όπου  $y'_t$  είναι η χρονοσειρά των διαφορών (1<sup>ης</sup> τάξης ή 2<sup>ης</sup> τάξης). Το παραπάνω μοντέλο συμβολίζεται ως με  $ARIMA(p, d, q)$  όπου,

1.  $p$  είναι η τάξη του αυτοπαλίνδρομου μέρους.
2.  $d$  είναι η τάξη της διαφορίσης.
3.  $q$  είναι η τάξη του μέρους του κινητού μέσου.

Οι ίδιες συνθήκες στασιμότητας και αντιστρεψιμότητας που περιγράφηκαν για τις αυτοπαλίνδρομες διαδικασίες και τις διαδικασίες κινητού μέσου διέπουν και τις διαδικασίες ARIMA. Η μέθοδος ARIMA δέχεται μη στάσιμα δεδομένα όσον αφορά την τάση. Δεν μπορεί να χειριστεί δεδομένα με μη σταθερή διασπορά και δεδομένα που έχουν εποχικότητα. Βέβαια στην περίπτωση που τα δεδομένα είναι στάσιμα ως προς την τάση τότε ακολούθως η μέθοδος ARIMA μειώνεται στην μέθοδο ARMA, η οποία είναι ισοδύναμη με την 1<sup>η</sup> παραλείποντας όμως την διαδικασία της διαφορίσης και της ολοκλήρωσης. Ένα μοντέλο ARMA μπορεί να γραφεί ως,

$$y_t = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4.9)$$

όπου  $y_t$  είναι η αρχική χρονοσειρά.

Μία επέκταση των μοντέλων ARIMA τα οποία είναι ικανά για την αντιμετώπιση επιπλέον της εποχικότητας είναι το εποχικό ARIMA μοντέλο (Seasonal ARIMA) το οποίο σε συντομογραφία γράφεται SARIMA. Ένα μοντέλο SARIMA σχηματίζεται εάν προσθέσουμε επιπλέον εποχικούς όρους στην εξίσωση που περιγράφει το μοντέλο ARIMA. Δηλαδή συμβολίζεται ως  $ARIMA(p, d, q)(P, D, Q)_m$ . Η πρώτη παρένθεση αναπαριστά το μη εποχικό μέρος του μοντέλου ενώ η δεύτερη το εποχικό μέρος και  $m$  είναι ο αριθμός των παρατηρήσεων που περιλαμβάνονται σε μία εποχή.

#### 4.3.5 Η μέθοδος Arima με Εποχική αποσύνθεση (STL+Arima)

Η εποχική εκδοχή της μεθόδου ARIMA (SARIMA) είναι σχεδιασμένη να λειτουργεί καλά για μικρότερες περιόδους. Αυτό συμβαίνει διότι υπάρχουν  $m-1$  παράμετροι που πρέπει να εκτιμηθούν για τις αρχικές εποχικές καταστάσεις όπου  $m$  είναι η εποχική περίοδος. Έτσι για μεγάλες τιμές του  $m$  (εδώ 288) η εκτίμηση γίνεται σχεδόν αδύνατη. Οπότε ακολουθήθηκε η εξής διαδικασία:

- 1) Αρχικά γίνεται ανάλυση στις βασικές συνιστώσες με την μέθοδο STL.
- 2) Αφαιρείται η εποχική συνιστώσα από τα δεδομένα.
- 3) Εφαρμόζεται η μέθοδος ARIMA στα εποχιακά προσαρμοσμένα δεδομένα χρησιμοποιώντας αυτοματοποιημένη διαδικασία μέσω της συνάρτησης *auto.arima()* και έπειτα γίνεται η πρόβλεψη.
- 4) Η πρόβλεψη της εποχικής συνιστώσας γίνεται μέσω της μεθόδου *seasonal naive*.
- 5) Οι τελικές προβλέψεις είναι το αποτέλεσμα του αθροίσματος των προβλέψεων για τα εποχιακά προσαρμοσμένα δεδομένα και των προβλέψεων για την εποχική συνιστώσα.

#### 4.3.6 Δυναμική Αρμονική Παλινδρόμηση με σφάλματα ARIMA

Άλλη μία μέθοδος που προτείνεται σε περιπτώσεις όπου η εποχική περίοδος είναι μεγάλη (μεγάλος αριθμός παρατηρήσεων ανά περίοδο) είναι η δυναμική αρμονική παλινδρόμηση με σφάλματα ARIMA. Έτσι, το εποχικό μοτίβο προτυποποιείται μέσω όρων Fourier, ενώ η βραχυπρόθεσμη δυναμική της χρονοσειράς προτυποποιείται μέσω των σφαλμάτων της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA. Η εν λόγω μέθοδος μπορεί να γραφεί ως,

$$y_t = a + \sum_{k=1}^K \left[ a_k \sin\left(\frac{2\pi kt}{m}\right) + \beta_k \cos\left(\frac{2\pi kt}{m}\right) \right] + N_t \quad (4.10)$$

όπου  $N_t \sim \text{ARIMA}(p, d, q)$ .

Τα θετικά αυτής της μεθόδου είναι [78]:

- Μπορεί να περιγράψει οποιοδήποτε μήκος εποχικότητας.
- Μπορεί να περιγράψει περισσότερες από μία διαφορετικές εποχικότητες, καθώς μπορεί να γίνει χρήση όρων Fourier με διαφορετικές συχνότητες.

- Η ομαλότητα του εποχικού μοτίβου μπορεί να ελεγχθεί μέσω της παραμέτρου  $K$ , η οποία είναι ο αριθμός Fourier ζευγαριών ημιτόνου και συνημιτόνου. Για μικρότερη τιμή του  $K$  έχουμε ομαλότερο εποχικό μοτίβο. Επιπροσθέτως, η τιμή της παραμέτρου  $K$  μπορεί να επιλεγεί με αναζήτηση πλέγματος χρησιμοποιώντας το κριτήριο AIC.
- Η βραχυπρόθεσμη δυναμική των δεδομένων μπορεί εύκολα να περιγραφεί από ένα απλό μοντέλο ARIMA.

Το μόνο πραγματικό μειονέκτημα σε σχέση πάντα με την εποχική μέθοδο ARIMA είναι ότι η εποχικότητα θεωρείται από το μοντέλο ότι είναι σταθερή, δηλαδή το εποχικό μοτίβο δεν επιτρέπεται να αλλάζει με τον χρόνο. Στην πράξη όμως και για μικρού μήκους χρονοσειρές (σε όρους χρονικού εύρους) η εποχικότητα είναι εντυπωσιακά σταθερή συνήθως και γι' αυτό το λόγο δεν αποτελεί σημαντικό πρόβλημα στις εφαρμογές.

Τα βήματα που ακολούθηθηκαν για την δυναμική αρμονική παλινδρόμηση με σφάλματα ARIMA είναι τα εξής:

1. Γίνεται χρήση της συνάρτησης *auto.arima()* όπου στο όρισμα *xreg* για της εξωγενείς μεταβλητές ορίζεται η σειρά Fourier μέσω συνάρτησης που έχει ως όρισμα  $k$  των αριθμό των σειρών που θα χρησιμοποιηθούν.
2. Η υπερπαραμέτρος  $k$  ορίζεται για εύρος τιμών από 1 έως 10 και η συνάρτηση *auto.arima()* τρέχει εμφωλευμένα για τις 10 διαφορετικές τιμές του  $k$ .
3. Επιλέγεται η παραμετροποίηση με το μικρότερο AIC.

Να σημειωθεί σε αυτό το σημείο ότι το μοντέλο δεν είναι ARIMAX δηλαδή δεν είναι ARIMA με εξωγενής μεταβλητές (covariates). Παρόλα αυτά η μέθοδος ARIMAX με εξωγενή μεταβλητή σειρές Fourier και η δυναμική παλινδρόμηση με σφάλματα που ακολουθούν διαδικασία ARIMA είναι εφάμιλλες όσον αφορά την προβλεπτική τους ικανότητα με την μέθοδο δυναμικής παλινδρόμησης να προκρίνεται σε επίπεδο ερμηνείας των παραμέτρων του μοντέλου ως προς την εξωγενή μεταβλητή. Άτυπα βέβαια θα αναφερόμαστε στην μέθοδο ως ARIMAX.

Πως λειτουργεί η συνάρτηση *auto.arima()* της βιβλιοθήκης *forecast* [57]:

Η συνάρτηση *auto.arima()* χρησιμοποιεί μια παραλλαγή του αλγορίθμου Hyndman-Khandakar [79] ο οποίος συνδυάζει ελέγχους μοναδιαίας ρίζας, ελαχιστοποίηση του κριτηρίου AIC και του MLE (Maximum Likelihood Estimation) ώστε να εκτιμήσει τις βέλτιστες

παραμέτρους της μεθόδου ARIMA. Τα βήματα που ακολουθεί ο αλγόριθμος (προκαθορισμένη διαδικασία) είναι τα εξής:

1. Ελέγχεται αν χρειάζεται μετασχηματισμός Box-Cox των δεδομένων με αυτόματο υπολογισμό της παραμέτρου  $\lambda$  του μετασχηματισμού, με σκοπό την σταθεροποίηση της διασποράς των δεδομένων.
2. Υπολογίζεται ο αριθμός των διαφορίσεων που απαιτούνται ( $0 \leq d \leq 2$ ) χρησιμοποιώντας επαναλαμβανόμενους ελέγχους KPSS, με σκοπό την σταθεροποίηση του μέσου.
3. Έπειτα επιλέγονται οι παράμετροι  $p$  και  $q$  μέσω της ελαχιστοποίησης του κριτηρίου AICc [106] και αφού έχουν εφαρμοστεί οι απαιτούμενες διαφορίσεις στα δεδομένα. Αντί να γίνει έλεγχος πλέγματος η οποία είναι ιδιαίτερα χρονοβόρα διαδικασία, ο αλγόριθμος χρησιμοποιεί διαδικασία ελέγχου κατά βήματα (stepwise procedure) για την «εξερεύνηση» του χώρου των παραμέτρων του μοντέλου. Αυτό γίνεται ως εξής:
  - α) Εκτιμώνται αρχικά 4 μοντέλα:
    - $ARIMA(0, d, 0)$
    - $ARIMA(2, d, 2)$
    - $ARIMA(1, d, 0)$
    - $ARIMA(0, d, 1)$
  - β) Το καλύτερο μοντέλο από τα 4 που εκτιμήθηκαν στο βήμα 3α) (με βάση το κριτήριο AICc) επιλέγεται ως το τρέχων μοντέλο της διαδικασίας κατά βήματα.
  - γ) Ελέγχονται διάφορες παραλλαγές του τρέχοντος μοντέλου, δηλαδή:
    - Ψάξε για της παραμέτρους  $p$  και  $q$  σε κάθε βήμα (+1/-1).
    - Συμπεριέλαβε ή όχι την σταθερά  $c$ .
    - Το καλύτερο μοντέλο που εντοπίζεται σε κάθε βήμα γίνεται το τρέχων για το επόμενο βήμα της διαδικασίας.
  - δ) Επανάλαβε το βήμα 3γ) έως ότου να μην μπορεί να βρεθεί παραμετροποίηση του μοντέλου με χαμηλότερο AICc.

#### 4.3.7 Εισαγωγή στις μεθόδους Exponential Smoothing

Οι μέθοδοι αυτές εμπίπτουν στην ευρύτερη κατηγορία των *Exponential Smoothing* μεθόδων χρησιμοποιώντας *state – space* προσέγγιση και λογική [81].

Η απλούστερη από τις μεθόδους exponential smoothing είναι η απλή exponential smoothing μέθοδος (Simple Exponential Smoothing – SES), η οποία θεωρείται κατάλληλη για δεδομένα που δεν έχουν ξεκάθαρη τάση ή εποχικότητα. Η μέθοδος αυτή σε μορφή συνιστωσών (component form) γράφεται ως:

$$\text{Εξίσωση πρόβλεψης (Forecast equation): } \hat{y}_{t+h|t} = l_t \quad (4.11)$$

$$\text{Εξίσωση Ομαλοποίησης (Smoothing equation): } l_t = \alpha y_t + (1 - \alpha)l_{t-1} \quad (4.12)$$

όπου  $l_t$  είναι το επίπεδο (ή διαφορετικά η ομαλοποιημένη τιμή) της χρονοσειράς την χρονική στιγμή  $t$ . Θέτοντας  $h = 1$  παίρνουμε τις προσαρμοσμένες τιμές, ενώ θέτοντας  $t = T$  παίρνουμε τις πραγματικές προβλέψεις πέρα των δεδομένων εκπαίδευσης.

Η μόνη συνιστώσα που περιλαμβάνεται στην μέθοδο είναι το επίπεδο (level),  $l_t$ . Η αναπαράσταση σε μορφή συνιστωσών για τις μεθόδους Exponential Smoothing περιλαμβάνουν πάντα την εξίσωση πρόβλεψης (forecast equation) και μία εξίσωση ομαλοποίησης (Smoothing equation) για κάθε μία από τις συνιστώσες τις οποίες περιλαμβάνει το εκάστοτε μοντέλο.

Ο Holt [82] επέκτεινε την απλή exponential smoothing μέθοδο επιτρέποντας την πρόβλεψη σε δεδομένα με τάση. Αυτή η μέθοδος περιλαμβάνει την εξίσωση πρόβλεψης και άλλες δύο συναρτήσεις ομαλοποίησης, μία που αφορά στο επίπεδο της χρονοσειράς και μία που αφορά στην τάση της χρονοσειράς οι οποίες γράφονται ως,

$$\text{Εξίσωση πρόβλεψης (Forecast equation): } \hat{y}_{t+h|t} = l_t + hb_t \quad (4.13)$$

$$\text{Εξίσωση Επιπέδου (Level equation): } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4.14)$$

$$\text{Εξίσωση Τάσης (Trend equation): } b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (4.15)$$

όπου  $l_t$  είναι η εκτίμηση του επιπέδου της χρονοσειράς την χρονική στιγμή  $t$ ,  $b_t$  της τάσης της χρονοσειράς την χρονική στιγμή  $t$ ,  $\alpha$  είναι η παράμετρος ομαλοποίησης του επιπέδου ( $0 \leq \alpha \leq 1$ ) και  $\beta^*$  είναι η παράμετρος ομαλοποίησης της τάσης ( $0 \leq \beta^* \leq 1$ ). Σε αντιπαράθεση με την απλή μέθοδο exponential smoothing, η εξίσωση επιπέδου είναι ο σταθμισμένος μέσος της παρατήρησης  $y_t$  και της ενός βήματος μπροστά πρόβλεψης την χρονική στιγμή  $t$  η οποία δίνεται από την πρόσθεση των όρων  $l_{t-1}$  και  $b_{t-1}$ . Η εξίσωση τάσης την χρονική στιγμή  $t$  είναι ξανά ο σταθμισμένος όρος της εκτιμημένης τάσης την χρονική στιγμή  $t$  το οποίο δίνεται από το  $l_t - l_{t-1}$



και την εκτίμηση της τάσης την χρονική στιγμή  $t-1$ . Η εξίσωση πρόβλεψης δεν είναι πλέον επίπεδη αλλά έχει τάση. Η  $h$  βήματα μπροστά πρόβλεψη ισούται με το τελευταίο εκτιμημένο επίπεδο συν  $h$  φορές την τελευταία εκτιμημένη τιμή της τάσης. Οι προβλέψεις λοιπόν είναι γραμμικές συναρτήσεις του ορίζοντα  $h$ .

Το πρόβλημα με την παραπάνω μέθοδο είναι ότι οι προβλέψεις που κάνει έχουν σταθερή τάση (αύξουσα ή φθίνουσα) αόριστα στο «μέλλον». Είναι σπάνιες οι περιπτώσεις στις οποίες η τάση σε μία χρονοσειρά θα είναι σταθερά αύξουσα ή φθίνουσα και η εμπειρία έχει δείξει ότι τέτοιες μέθοδοι τείνουν να κάνουν υπερεκτίμηση ή υποεκτίμηση των προβλέψεων ειδικά για μεγάλους ορίζοντες. Γι' αυτό το λόγο οι Gardner & McKenzie [83] εισήγαγαν στο μοντέλο μία παράμετρο η οποία αποσβένει την τάση η οποία μετά από κάποιο χρόνο στο «μέλλον» καταλήγει να είναι ευθεία γραμμή. Άρα σε σχέση με την μέθοδο χωρίς απόσβεση αυτή την φορά εισάγεται επιπλέον στις εξισώσεις η παράμετρος απόσβεσης με  $\varphi: 0 < \varphi < 1$ ,

$$\text{Εξίσωση πρόβλεψης (Forecast equation): } \hat{y}_{t+h|t} = l_t + (\varphi + \varphi^2 + \dots + \varphi^h)b_t \quad (4.16)$$

$$\text{Εξίσωση Επιπέδου (Level equation): } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \varphi b_{t-1}) \quad (4.17)$$

$$\text{Εξίσωση Τάσης (Trend equation): } b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\varphi b_{t-1} \quad (4.18)$$

Αν το  $\varphi = 1$ , τότε η απόσβεση δεν έχει επίδραση και η μέθοδος είναι ισοδύναμη με την γραμμική μέθοδο του Holt. Για τιμές μεταξύ του 0 και 1, η  $\varphi$  αποσβένει την τάση έτσι ώστε να προσεγγίζει κάποια σταθερά σε κάποια χρονική στιγμή στο μέλλον.

Περαιτέρω επέκταση για να μπορούν να έχουν οι χρονοσειρές εποχική συνιστώσα έγινε από τους Holt [84] και Winters [85]. Η εποχική μέθοδος Holt-Winters περιλαμβάνει τη εξίσωση και 3 εξισώσεις ομαλοποίησης, μία για το επίπεδο  $l_t$  της χρονοσειράς, μία για την εποχική συνιστώσα  $s_t$ , και τέλος μία για την τάση  $b_t$  με αντίστοιχες παραμέτρους  $\alpha, \beta^*$  και  $\gamma$ . Με  $m$  συμβολίζουμε την συχνότητα της εποχικότητας.

Υπάρχουν 2 παραλλαγές αυτής τη μεθόδου οι οποίες διαφέρουν στον τρόπο προσέγγισης της εποχικής συνιστώσας. Η προσθετική μέθοδος προτιμάται όταν η εποχικές διακυμάνσεις είναι περίπου σταθερές κατά μήκος της χρονοσειράς, ενώ η πολλαπλασιαστική μέθοδος προτιμάται όταν οι εποχικές διακυμάνσεις μεταβάλλονται ανάλογα με το επίπεδο της χρονοσειράς. Θα παρουσιαστεί μόνο η προσθετική μέθοδος εδώ.

Το Holt-Winters προσθετικό μοντέλο σε μορφή συνιστωσών γράφεται ως,

$$\text{Εξίσωση πρόβλεψης (Forecast equation): } \hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (4.19)$$

$$\text{Εξίσωση Επιπέδου (Level equation): } l_t = a(y_t - s_{t-m}) + (1 - a)(l_{t-1} + b_{t-1}) \quad (4.20)$$

$$\text{Εξίσωση Τάσης (Trend equation): } b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (4.21)$$

$$\text{Εξίσωση Εποχικότητας (Seasonal equation): } s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (4.22)$$

όπου  $k$  είναι το ακέραιο μέρος της διαίρεσης  $\frac{h-1}{m}$ , το οποίο εξασφαλίζει ότι η εκτίμηση της εποχικότητας που θα χρησιμοποιηθεί για πρόβλεψη προέρχεται από την τελευταία διαθέσιμη εποχή του δείγματος. Η εξίσωση επιπέδου είναι ο σταθμισμένος μέσος μεταξύ της εποχικά προσαρμοσμένης παρατήρησης ( $y_t - s_{t-m}$ ) και την μη εποχικής πρόβλεψης ( $l_{t-1} + b_{t-1}$ ) την χρονική στιγμή  $t$ . Η εξίσωση τάσης είναι πανομοιότυπη με την αντίστοιχη της γραμμικής μεθόδου του Holt. Η εποχική εξίσωση είναι ο σταθμισμένος μέσος μεταξύ του τρέχοντος εποχικού δείκτη ( $y_t - l_{t-1} - b_{t-1}$ ) και του εποχικού δείκτη της αμέσως προηγούμενης εποχής. Αντίστοιχα, μπορεί στις παραπάνω μεθόδους να προστεθεί απόσβεση.

Κάθε μοντέλο αποτελείται από μία εξίσωση μέτρησης (measurement equation) η οποία περιγράφει τα παρατηρούμενα δεδομένα και από κάποιες εξισώσεις καταστάσεις (state equations) οι οποίες περιγράφουν πως οι μη παρατηρούμενες συνιστώσες ή καταστάσεις (επίπεδο, τάση, εποχικότητα) μεταβάλλονται στον χρόνο. Γι' αυτό και εμπίπτουν στην ευρύτερη κατηγορία των state-space μοντέλων.

Για κάθε ένα από τα 9 μοντέλα που προκύπτουν από τους διάφορους δυνατούς συνδυασμούς είναι δυνατό να χρησιμοποιηθούν προσθετικά σφάλματα είτε πολλαπλασιαστικά σφάλματα. Για την διάκριση των διαφορετικών μοντέλο γίνεται χρήση μίας αναπαράσταση με 3 γράμματα. Έτσι, κάθε μοντέλο θα συμβολίζεται με ETS(.,.,.) για (Σφάλμα, Τάση, Εποχικότητα). Οι δυνατότητες για κάθε συνιστώσα είναι:

1. Σφάλμα = {A, M} όπου,
  - A = προσθετικά σφάλματα (additive errors)
  - M = πολλαπλασιαστικά σφάλματα (multiplicative errors)
2. Τάση = {N, A, A<sub>d</sub>} όπου,
  - N = χωρίς τάση (None)

- $A$  = προσθετική τάση (additive trend)
- $A_d$  = προσθετική τάση με απόσβεση (additive damped trend)

3. Εποχικότητα =  $\{N, A, M\}$  όπου,

- $N$  = χωρίς εποχικότητα (None)
- $A$  = προσθετική εποχικότητα (additive seasonality)
- $M$  = πολλαπλασιαστική εποχικότητα (multiplicative seasonality)

Σε αυτό το σημείο αξίζει να σχολιαστεί ότι για τις μεθόδους exponential smoothing με εποχική συνιστώσα ισχύει το ίδιο πρόβλημα που αναφέρθηκε πιο πάνω για την μέθοδο SARIMA, δηλαδή δεν λειτουργούν καλά σε επίπεδο υπολογισιμότητας για μεγάλες εποχικές συχνότητες.

#### 4.3.8 Η μέθοδος ETS με Εποχική αποσύνθεση (STL+ETS)

Η διαδικασία που ακολουθήθηκε είναι η εξής:

- 1) Αρχικά γίνεται ανάλυση στις βασικές συνιστώσες με την μέθοδο STL.
- 2) Αφαιρείται η εποχική συνιστώσα από τα δεδομένα.
- 3) Εφαρμόζεται η μέθοδος ETS στα εποχιακά προσαρμοσμένα δεδομένα χρησιμοποιώντας αυτοματοποιημένη διαδικασία μέσω της συνάρτησης *ets()* και έπειτα γίνεται η πρόβλεψη.
- 4) Η πρόβλεψη της εποχικής συνιστώσας γίνεται μέσω της μεθόδου *seasonal naive*.
- 5) Οι τελικές προβλέψεις είναι το αποτέλεσμα του αθροίσματος των προβλέψεων για τα εποχιακά προσαρμοσμένα δεδομένα και των προβλέψεων για την εποχική συνιστώσα.

Πως λειτουργεί η συνάρτηση *ets()* της βιβλιοθήκης *forecast* [57]:

Ένα θετικό των μεθόδων ETS είναι ότι για την επιλογή του καλύτερου μοντέλου είναι δυνατό να χρησιμοποιηθούν κριτήρια πληροφορίας. Προκαθορισμένα επιλέγει το καλύτερο μοντέλο χρησιμοποιώντας το κριτήριο AICc. Εάν δοθεί ως όρισμα μόνο το σύνολο εκπαίδευσης τότε επιλέγει εντελώς αυτοματοποιημένα το καταλληλότερο μοντέλο.

#### 4.3.9 Μέθοδος TBATS

Η μέθοδος TBATS [86] αποτελεί μία state space μέθοδο η οποία μπορεί να μοντελοποιήσει εποχικότητες υψηλών συχνοτήτων και αντιμετωπίζει πολλά άλλα προβλήματα που παρουσιάστηκαν στις προαναφερθείσες μεθόδους. Η μέθοδος TBATS ουσιαστικά αποτελεί γενίκευση του μοντέλου Holt-Winters. Η μέθοδος περιορίζεται σε γραμμικά ομοσκεδαστικά μοντέλα επιτρέποντας κάποιες μορφές μη γραμμικότητας μέσω του μετασχηματισμού Box-Cox.

Αυτό περιορίζει την προσέγγιση σε χρονοσειρές που παίρνουν μόνο θετικές τιμές όμως οι κυκλοφοριακές σειρές εν γένει περιορίζονται σε θετικές τιμές. Ο συμβολισμός  $y_t^{(\omega)}$  χρησιμοποιείται για την αναπαράσταση των μετασχηματισμένων κατά Box-Cox δεδομένων με παράμετρο  $\omega$ , όπου  $y_t$  είναι η παρατήρηση την χρονική στιγμή  $t$ . Άρα επεκτείνεται η μέθοδος Holt-Winters συμπεριλαμβάνοντας τον μετασχηματισμό Box-Cox, και σφάλματα ARMA το οποίο γράφεται ως (με απλή εποχικότητα),

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^\omega - 1}{\omega}, & \omega \neq 0 \\ \log y_t, & \omega = 0 \end{cases} \quad (4.23)$$

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + s_{t-1} + d_t \quad (4.24)$$

$$l_t = l_{t-1} + \phi b_{t-1} + a d_t \quad (4.25)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t \quad (4.26)$$

$$s_t = \sum_{j=1}^k s_{j,t} \quad (4.27)$$

$$s_{j,t} = s_{j,t-1} \cos \lambda_j + s_{j,t-1}^* \sin \lambda_j + \gamma_1 d_t \quad (4.28)$$

$$s_{j,t}^* = -s_{j,t-1} \sin \lambda_j + s_{j,t-1}^* \cos \lambda_j + \gamma_2 d_t \quad (4.29)$$

$$d_t = \sum_{i=1}^p \phi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (4.30)$$

όπου  $m$  είναι η εποχική περίοδος,  $l_t$  είναι το επίπεδο τοπικά την χρονική στιγμή  $t$ ,  $b$  είναι η μακροπρόθεσμη τάση,  $b_t$  είναι η βραχυπρόθεσμη τάση την χρονική στιγμή  $t$ , το  $s_t$  αναπαριστά την εποχική συνιστώσα την χρονική στιγμή  $t$ , το  $d_t$  συμβολίζει μία διαδικασία  $ARMA(p, q)$  και το  $\varepsilon_t$  είναι μία διαδικασία γκαουσιανού λευκού θορύβου με μέση τιμή 0 και σταθερή διασπορά  $\sigma^2$ . Στην αναζήτηση μίας ευέλικτης και φειδωλής προσέγγισης, στην μέθοδο TBATS εισάγεται τριγωνομετρική αναπαράσταση της εποχικής συνιστώσας η οποία βασίζεται σε σειρές Fourier. Τα  $\alpha$ ,  $\beta$  είναι παράμετροι ομαλοποίησης της μεθόδου καθώς και  $\gamma_1$ ,  $\gamma_2$  είναι επίσης παράμετροι ομαλοποίησης και  $\lambda_j = \frac{2\pi j}{m}$ . Το στοχαστικό επίπεδο της εποχικής συνιστώσας συμβολίζεται με

$s_{j,t}$ , και η στοχαστική αύξηση του επιπέδου της εποχικής συνιστώσας η οποία είναι απαραίτητη για να περιγράψει την αλλαγή της εποχικής συνιστώσας με τον χρόνο συμβολίζεται με  $s_{j,t}^*$ . Ο αριθμός των αρμονικών όρων που απαιτούνται για την εποχική συνιστώσα συμβολίζεται με  $k$ . Όπως είναι εμφανές έχουμε αποσβένουσα τάση με παράμετρο απόσβεσης  $\phi$ , με πρόσθετη μία μακροπρόθεσμη τάση  $b$ . Αυτή η πρόσθεση οδηγεί τις προβλέψεις μελλοντικών τιμών της βραχυπρόθεσμης τάσης  $b_t$  να συγκλίνει στην μακροπρόθεσμη τάση και όχι στο μηδέν. Παρόλο που ο παράγοντας απόσβεσης συμπεριλαμβάνεται και στην εξίσωση μέτρησης και στην εξίσωση τάσης και επιπέδου, εάν παραληφθούν οι ίδιες ακριβώς προβλέψεις επιστρέφονται.

Το διάνυσμα  $TBATS(\omega, \phi, p, q, \{m, k\})$  καθορίζει το εκτιμημένο μοντέλο. Κάποια από τα πλεονεκτήματα της μεθόδου TBATS είναι:

- 4 Δέχεται μεγαλύτερο χώρο παραμέτρων αυξάνοντας την πιθανότητα για καλύτερες προβλέψεις.
- 5 Επιτρέπει την τον συνδυασμό εμφωλευμένων και μη εμφωλευμένων πολλαπλών εποχικών συνιστωσών.
- 6 Μπορεί να διαχειριστεί κάποια βασικά μη γραμμικά χαρακτηριστικά που εμφανίζονται συχνά σε πραγματικές χρονοσειρές.
- 7 Επιτρέπει την διαχείριση οποιουδήποτε είδους αυτοσυσχέτισης στα κατάλοιπα.
- 8 Περιλαμβάνει μία πολύ απλούστερη, αλλά αποτελεσματική εκτίμηση των παραμέτρων της διαδικασίας.

Όσον αφορά την επιλογή του καλύτερου μοντέλου μεταξύ των μοντέλων που κάθε φορά εκτιμώνται χρησιμοποιείται το κριτήριο πληροφορίας AIC (ή AICc). Η επιλογή της υπερπαραμέτρου  $k$  που είναι ο αριθμός των αρμονικών της εποχικής συνιστώσας ακολουθείται η εξής διαδικασία:

- Χρησιμοποιώντας κατάλληλη μέθοδο γίνεται αφαίρεση της τάσης των πρώτων εποχών των μετασχηματισμένων δεδομένων.
- Έπειτα προσεγγίζονται τα δεδομένα που προκύπτουν από το προηγούμενο βήμα χρησιμοποιώντας την γραμμική παλινδρόμηση  $\sum_{j=1}^k \alpha_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)$ . Ξεκίνα με ένα αρμονικό όρο και βαθμιαία αύξανε κατά ένα τους αρμονικούς όρους, ελέγχοντας κάθε φορά την στατιστική σημαντικότητα του νέου όρου με F έλεγχο.

- Για όλες τις τιμές του  $k$  που «πέρασαν» τον  $F$  έλεγχο, προσαρμόζεται το τρέχων μοντέλο στα δεδομένα για όλες τις τιμές του  $k$  και υπολογίζεται η τιμή του AIC του κάθε μοντέλου.
- Το βέλτιστο  $k$  είναι εκείνο που χρησιμοποιήθηκε στο μοντέλο με το μικρότερο AIC.

Τα σφάλματα μοντελοποιούνται με διαδικασία ARMA οπότε πρέπει να υπολογιστούν οι παράμετροι  $p$  και  $q$  της διαδικασίας. Η διαδικασία είναι η εξής:

- Αρχικά, προσαρμόζεται ένα μοντέλο χωρίς ARMA όρο.
- Έπειτα, χρησιμοποιείται ο αυτοματοποιημένος ARIMA αλγόριθμος [79] στα κατάλοιπα του μοντέλου του προηγούμενου βήματος έτσι ώστε να προσδιοριστούν οι καλύτερες τιμές για τις παραμέτρους  $p$  και  $q$  (υπόθεση: τα κατάλοιπα είναι στάσιμα).
- Το επιλεγμένο μοντέλο προσαρμόζεται ξανά με σφάλματα ARMA( $p,q$ ), όπου οι παράμετροι της μεθόδου ARMA εκτιμώνται από κοινού με τις υπόλοιπες παραμέτρους.
- Ο όρος ARMA διατηρείται μόνο εάν το AIC του μοντέλου είναι μικρότερο από το αντίστοιχο χωρίς ARMA όρο.

Η κύρια διαφορά της μεθόδου TBATS με την δυναμική παλινδρόμηση, πέρα από την καλύτερη αναζήτηση στον χώρο, είναι ότι επιτρέπει στην εποχική συνιστώσα μεγαλύτερη ευελιξία δηλαδή επιτρέπει στην εποχικότητα να μεταβάλετε με αργούς ρυθμούς στον χρόνο, ενώ στην δυναμική παλινδρόμηση οι όροι Fourier «εξαναγκάζουν» το εποχικό μοτίβο να επαναλαμβάνεται περιοδικά χωρίς να μεταβάλετε. Ένα από τα μειονεκτήματα της μεθόδου είναι ότι ο υπολογισμός των παραμέτρων του καλύτερου μοντέλου μπορεί να είναι αργός, ειδικά για χρονοσειρές μεγάλου μήκους.

Για την υλοποίηση της μεθόδου έγινε χρήση της βιβλιοθήκης *forecast*. Η συνάρτηση που υλοποιεί την μέθοδο TBATS είναι η *tbats()*, η οποία υλοποιεί μία πλήρως αυτοματοποιημένη διαδικασία. Ως όρισμα στην συνάρτηση δόθηκε το τελικό σύνολο δεδομένων.

#### 4.3.10 Μέθοδος Theta

Η μέθοδος Theta [87] βασίζεται στην ιδέα της τροποποίησης των τοπικών καμπυλοτήτων της χρονοσειράς. Αυτή η αλλαγή λαμβάνεται από έναν συντελεστή, ο οποίος ονομάζεται Theta-συντελεστής, και εφαρμόζεται απευθείας στην διαφώριση 2<sup>ης</sup> τάξης της χρονοσειράς,

$$X''_{new}(\theta) = \theta X''_{data} \quad (4.31)$$

όπου  $X''_{data} = X_t - 2X_{t-1} + X_{t-2}$  την χρονική στιγμή  $t$ .

Αν οι τοπικές καμπυλώσεις μειώνονται σταδιακά τότε η χρονοσειρά αποπληθωρίζεται. Όσο μικρότερη είναι η τιμή του συντελεστή Theta, τόσο μεγαλύτερος είναι ο ρυθμός του αποπληθωρισμού. Στην ακραία περίπτωση που η τιμή του συντελεστή Theta είναι 0, η χρονοσειρά μετασχηματίζεται σε μία ευθεία γραμμικής παλινδρόμησης. Η προοδευτική μείωση των διακυμάνσεων εξαλείφει της απόλυτες διαφορές μεταξύ των όρων της εξαγόμενης χρονοσειράς και συσχετίζεται, σε ποσοτικούς όρους, με την εμφάνιση μακροπρόθεσμων τάσεων στα δεδομένα. Αντίθετα, αν η τοπική καμπύλωση είναι αυξημένη ( $\theta > 1$ ), τότε η χρονοσειρά είναι διεσταλμένη. Όσο μεγαλύτερος είναι ο βαθμός της διαστολής, τόσο μεγαλύτερη είναι αύξηση της βαχυπρόθεσμης συμπεριφοράς.

Ακολουθώντας αυτή την διαδικασία, ένα σύνολο από νέες χρονοσειρές δημιουργείται, οι οποίες ονομάζονται Theta γραμμές. Η τοποθέτηση αυτών των γραμμών σε σχέση με τα αρχικά δεδομένα μπορεί να γίνει με αρκετούς διαφορετικούς τρόπους. Εάν η προσαρμογή πραγματοποιείται με OLS διαδικασία εκτίμησης τότε ο μέσος και η κλίση των Theta γραμμών παραμένουν ίδια συγκρινόμενα πάντα με τα αντίστοιχα των αρχικών δεδομένων.

Η γενική εφαρμογή της μεθόδου γίνεται ως εξής. Η αρχική χρονοσειρά αποσυντίθεται σε δύο ή περισσότερες Theta γραμμές. Κάθε Theta γραμμή προεκβάλεται ξεχωριστά και οι προβλέψεις απλά συνδυάζονται. Οποιαδήποτε μέθοδος πρόβλεψης μπορεί να χρησιμοποιηθεί για την προεκβολή της Theta γραμμής. Επιπροσθέτως μπορεί διαφορετικός συνδυασμός των Theta γραμμών να χρησιμοποιηθεί για κάθε ορίζοντα πρόβλεψης. Η απλούστερη περίπτωση είναι εκείνη στην οποία η αρχική χρονοσειρά αποσυντίθεται σε 2 Theta γραμμές, για παράδειγμα,  $\theta=0$  και  $\theta=2$ .

$$\text{Δεδομένα} = \frac{1}{2}(L(\theta = 0) + L(\theta = 2)) \quad (4.32)$$

όπου  $L(\theta = 0)$  και  $L(\theta = 2)$  είναι οι Theta γραμμές για τιμές τις παραμέτρου  $\theta=0$  και  $\theta=2$ , αντίστοιχα.

Η 1<sup>η</sup> γραμμή Theta ( $\theta=0$ ) είναι η ευθεία της γραμμικής παλινδρόμησης των δεδομένων και η 2<sup>η</sup> περιέχει τις διαφορές ακριβώς 2 φορές της αρχικής χρονοσειράς. Αυτή είναι η περίπτωση όπου δύο, ακραίες και συμμετρικές στο 1, Theta γραμμές συντίθενται. Η 1<sup>η</sup> συνιστώσα  $L(\theta = 0)$  περιγράφει την χρονοσειρά μέσω μίας γραμμικής τάσης. Η δεύτερη γραμμή  $L(\theta = 2)$ , έχει

διπλασιάζει τις τοπικές καμπυλώσεις τονίζοντας έτσι την βραχυπρόθεσμη συμπεριφορά της χρονοσειράς. Η 1<sup>η</sup> Theta γραμμή προεκβάλεται με τον συνηθισμένο τρόπο που χρησιμοποιείται για γραμμική τάση, ενώ η 2<sup>η</sup> με την απλή exponential smoothing μέθοδο. Ένας απλός συνδυασμός των 2 προβλέψεων δίνει την τελική πρόβλεψη για την Theta μέθοδο για την χρονοσειρά.

Εν κατακλείδι περιγράφουμε τα βήματα του αλγορίθμου για εποχικές χρονοσειρές:

1. Η χρονοσειρά αποσυντίθεται σε 2 Theta γραμμές, την ευθεία της γραμμικής παλινδρόμησης ( $\theta=0$ ) και η Theta γραμμή για  $\theta=2$ .
2. Η 1<sup>η</sup> Theta γραμμή προεκβάλεται με τον συνηθισμένο τρόπο που χρησιμοποιείται για γραμμική τάση, ενώ η 2<sup>η</sup> με την απλή exponential smoothing μέθοδο.
3. Οι προβλέψεις που παράγονται από το 3<sup>ο</sup> βήμα για τις 2 γραμμές συνδυάζονται με ίσα βάρη.

Το δυνατό σημείο της μεθόδου είναι η αποσύνθεση (ανάλυση σε Theta γραμμές) των αρχικών δεδομένων. Οι δύο συνιστώσες περιλαμβάνουν πληροφορίες, οι οποίες είναι χρήσιμες για την διαδικασία της πρόβλεψης αλλά χάνονται ή δεν μπορούν να τις εκμεταλλευτούν εξ' ολοκλήρου οι υπάρχουσες μέθοδοι όταν εφαρμόζονται απευθείας στα αρχικά δεδομένα.

Για την υλοποίηση της μεθόδου έγινε χρήση της βιβλιοθήκης *forecast*. Η συνάρτηση που υλοποιεί την μέθοδο Theta είναι η *thetaf()*. Ως όρισμα στην συνάρτηση δόθηκε το τελικό σύνολο δεδομένων.

## 4.4 ΜΕΘΟΔΟΙ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Στην ομάδα 3 ανήκουν οι μέθοδοι βαθιάς μηχανικής μάθησης. Η 1<sup>η</sup> μέθοδος θα παρουσιαστεί με μεγαλύτερη λεπτομέρεια καθώς αποτελεί την βάση όλων των υπολοίπων και επιπροσθέτως κάποιες από τις αρχές που την διέπουν, διέπουν και τις υπόλοιπες μεθόδους.

### 4.4.1 Η Μέθοδος Multi-layer Perceptron (MLP)

Τα πολύ-επίπεδα περσέπτρον (multi-layer perceptrons) ή διαφορετικά τα νευρωνικά δίκτυα εμπρόσθιας μετάδοσης είναι η πεμπουσία των μεθόδων βαθιάς μηχανικής μάθησης. Ο στόχος ενός εμπρόσθιου δικτύου είναι να προσεγγίσει μία συνάρτηση  $f^*$ . Ορίζει, δηλαδή, μία απεικόνιση  $f(x; w)$  και μαθαίνει τις τιμές των παραμέτρων  $w$  οι οποίες οδηγούν στην καλύτερη προσέγγιση της συνάρτησης. Ονομάζονται δίκτυα διότι τυπικά αναπαρίστανται από την σύνθεση πολλών διαφορετικών συναρτήσεων. Το μοντέλο αποτελεί ουσιαστικά ένα κατευθυνόμενο ακυκλικό

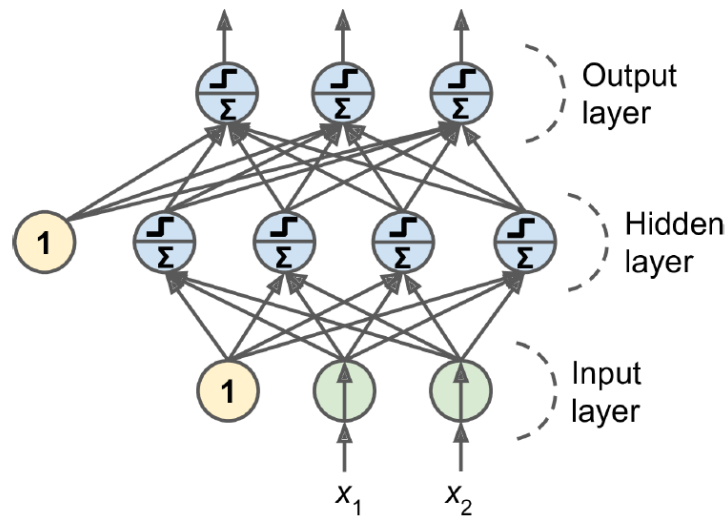


γράφημα το οποίο περιγράφει πως αυτές οι συναρτήσεις συνδέονται μεταξύ τους [10]. Ένα μοντέλο MLP αποτελείται από ένα επίπεδο εισόδου, ένα ή περισσότερα επίπεδα που αποτελούνται από λογικές μονάδες κατωφλίου (Threshold Logic Units – TLUs), τα οποία ονομάζονται κρυφά επίπεδα (hidden layers), και ένα τελευταίο επίπεδο που αποτελείται από λογικές μονάδες κατωφλίου και ονομάζεται επίπεδο εξόδου. Ο αριθμός των κρυφών επιπέδων ορίζει το βάθος του νευρωνικού δικτύου, ενώ ο αριθμός των νευρών – μονάδων το πλάτος του. Τα επίπεδα που βρίσκονται κοντά στο επίπεδο εισόδου συνήθως τα λέμε χαμηλότερα επίπεδα ενώ εκείνα που βρίσκονται κοντά στο επίπεδο εξόδου τα ονομάζουμε ανώτερα επίπεδα.

Όλα τα επίπεδα εκτός του επιπέδου εξόδου περιλαμβάνουν ένα νευρώνα κατωφλίου και είναι πλήρως συνδεδεμένα με τα επόμενα επίπεδα. Κάθε μία σύνδεση μεταξύ 2 επιπέδων «συνοδεύεται» από ένα βάρος,  $w_{ij}$ , όπου  $i$  είναι ο αριθμός του νευρώνα από όπου ξεκινάει η σύνδεση και  $j$  ο αριθμός του νευρώνα που καταλήγει η σύνδεση. Τα βάρη αυτά αποτελούν και τις παραμέτρους του μοντέλου. Η είσοδος σε κάθε νευρώνα είναι ο γραμμικός συνδυασμός των βαρών των συνδέσεων που καταλήγουν στον νευρώνα και των εξόδων του νευρώνα του προηγούμενου επιπέδου. Εκτός από τα επίπεδα εισόδου και εξόδου, η έξοδος από κάθε νευρώνα κάθε κρυφού επιπέδου είναι η είσοδος του επόμενου επιπέδου μετασχηματισμένη από μία μη γραμμική συνάρτηση, Στην Εικόνα 5 φαίνεται η αρχιτεκτονική ενός MLP με δύο εισόδους, ένα κρυφό επίπεδο, και τρεις νευρώνες εξόδου.

Επειδή το σήμα «ρέει» μόνο προς μία κατεύθυνση, δηλαδή από το επίπεδο εισόδου προς το επίπεδο εξόδου, αυτή η αρχιτεκτονική είναι ένα παράδειγμα νευρωνικού δικτύου εμπρόσθιας τροφοδότησης. Επίσης, όταν ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network) περιλαμβάνει αρκετά κρυφά επίπεδα, ονομάζεται βαθύ νευρωνικό δίκτυο (Deep Neural Network).

Κατά την εκπαίδευση του νευρωνικού δικτύου, οδηγούμε την  $f(x)$  προς την  $f^*(x)$ . Τα δεδομένα εκπαίδευσης μας εφοδιάζουν με θορυβώδη, προσεγγιστικά παραδείγματα της  $f^*(x)$ , δηλαδή εκτιμήσεις σε διάφορα σημεία του συνόλου εκπαίδευσης. Κάθε παράδειγμα  $x$  συνοδεύεται από μία ετικέτα (label)  $y \approx f^*(x)$ . Τα παραδείγματα εκπαίδευσης καθορίζουν επακριβώς τι πρέπει το επίπεδο εξόδου να κάνει σε κάθε σημείο  $x$ . Είναι προφανές ότι η έξοδος πρέπει να είναι ένα σημείο  $y$  που να είναι «κοντά» στο  $x$ , ως προς μία συνάρτηση κόστους στην έξοδο του νευρωνικού. Η συμπεριφορά των ενδιάμεσων επιπέδων δεν καθορίζεται άμεσα από τα δεδομένα εκπαίδευσης, και αυτός είναι και ο λόγος που ονομάζονται κρυφά.



**Εικόνα 5** Η αρχιτεκτονική του MLP με δύο εισόδους, ένα κρυφό επίπεδο, και τρεις νευρώνες εξόδου [88].

Ο αλγόριθμος που χρησιμοποιείται για την εκπαίδευση του δικτύου, δηλαδή την εκμάθηση των βαρών, είναι ο αλγόριθμος οπίσθιας μετάδοσης γνωστός στην βιβλιογραφία ως αλγόριθμος backpropagation. Βασίζεται στην γνωστή μέθοδο κατάβασης κλίσης και χρησιμοποιεί ένα έξυπνο και αποδοτικό τρόπο υπολογισμού των κλίσεων (παραγώγων) αυτοματοποιημένα (automatic differentiation). Χρειάζεται μόνο δύο περάσματα στο δίκτυο (ένα εμπρόσθιο, ένα οπίσθιο) για να υπολογίσει την παράγωγο του σφάλματος του δικτύου αναφορικά με κάθε μία παράμετρο του δικτύου. Ουσιαστικά, μαθαίνει πως πρέπει κάθε βάρος κάθε σύνδεσης και κάθε όρος κατωφλίου να μεταβληθεί ώστε να μειωθεί το σφάλμα. Αφού έχουν υπολογιστεί οι παράγωγοι του σφάλματος εφαρμόζει ένα κλασσικό βήμα καθοδικής κλίσης, και η διαδικασία επαναλαμβάνεται μέχρι ο αλγόριθμος να συγκλίνει σε κάποια λύση. Η διαδικασία που ακολουθεί, αναλυτικά, είναι η εξής [88]:

- Κάθε φορά διαχειρίζεται μία μικρή παρτίδα από το σύνολο εκπαίδευσης το οποίο ονομάζεται mini-batch (π.χ. 32 παρατηρήσεις). Το μέγεθος αυτής της παρτίδας μάλιστα αποτελεί σημαντική υπερπαραμέτρο της μεθόδου. Ο αλγόριθμος «βλέπει» όλο το σύνολο εκπαίδευσης πολλές φορές μέσω των mini-batch. Κάθε ένα πέρασμα από όλα τα δεδομένα ονομάζεται εποχή (epoch).
- Κάθε mini-batch περνάει μέσω του επιπέδου εισόδου, στο πρώτο κρυφό επίπεδο. Ο αλγόριθμος υπολογίζει την έξοδο του κάθε νευρώνα σε αυτό το επίπεδο (για κάθε ένα

στιγμιότυπο του mini-batch). Το αποτέλεσμα εισάγονται στο επόμενο επίπεδο κοκ. μέχρι το τελευταίο επίπεδο.

- Ακολουθώντας, ο αλγόριθμος μετράει το σφάλμα εξόδου του δικτύου, χρησιμοποιώντας μία συνάρτηση κόστους συγκρίνοντας την πραγματική με την παραγόμενη έξοδο.
- Υπολογίζει αναλυτικά πόσο κάθε σύνδεση εξόδου συνέβαλε στο σφάλμα μέσω του κανόνα της αλυσίδας.
- Συνεχίζει κάνοντας το ίδιο με κάθε σύνδεση από το προηγούμενο επίπεδο δηλαδή μετράει πόσο συμβάλλει κάθε νευρώνας κάθε επιπέδου στο σφάλμα πάλι χρησιμοποιώντας τον κανόνα της αλυσίδας μέχρι να φτάσει στο επίπεδο εισόδου.
- Τέλος, ο αλγόριθμος πραγματοποιεί ένα καθοδικό βήμα κλίσης ώστε να αλλάξει την τιμή των βαρών κατά τρόπο ώστε να μειώσει το σφάλμα του δικτύου.

Ένας τρόπος να κατανοήσει κανείς τα νευρωνικά δίκτυα εμπρόσθιας μετάδοσης είναι να σκεφτεί πως μπορούν να ξεπεραστούν οι περιορισμοί των γραμμικών μοντέλων. Η επέκταση των τελευταίων ώστε να μπορούν να αναπαραστήσουν μη γραμμικές συναρτήσεις του  $x$ , μπορεί να επιτευχθεί εφαρμόζοντας το γραμμικό μοντέλο όχι απευθείας στο  $x$  αλλά σε μία μετασχηματισμένη είσοδο  $\varphi(x)$ , όπου  $\varphi$  είναι ένας μη γραμμικός μετασχηματισμός.

Ανακεφαλαιώνοντας τα σημαντικότερα στοιχεία που απαρτίζουν ένα νευρωνικό δίκτυο, τα οποία είναι:

- Συνάρτηση κόστους στο επίπεδο εξόδου.
- Συνάρτηση ενεργοποίησης στους νευρώνες του επιπέδου εξόδου.
- Συνάρτηση ενεργοποίησης των νευρώνων των κρυφών επιπέδων.
- Αρχιτεκτονική του δικτύου (βάθος, πλάτος).

Το πρόβλημα της πρόβλεψης χρονοσειρών χρησιμοποιώντας βαθιά μηχανική μάθηση στην απλή μορφή του μοντελοποιείται ως πρόβλημα παλινδρόμησης και πιο συγκεκριμένα ως πρόβλημα αυτοπαλινδρόμησης δηλαδή ως μία διαδικασία  $AR(p)$ . Αυτό σημαίνει ότι η έξοδος του επιπέδου εξόδου είναι συνεχής μεταβλητή και επηρεάζει την επιλογή των παραπάνω αναφερθέντων σημείων.

Η συνάρτηση κόστους [10] είναι από τα σημαντικά συστατικά στοιχεία ενός νευρωνικού δικτύου. Ευτυχώς, οι συναρτήσεις κόστους είναι περίπου ίδιες με εκείνες που χρησιμοποιούνται στα

γραμμικά μοντέλα. Χωρίς να προχωρήσουμε σε περισσότερη λεπτομέρεια, έγινε χρήση του μέσου τετραγωνικού σφάλματος (Mean Squared Error – MSE), η οποία αποτελεί και κλασική επιλογή για προβλήματα παλινδρόμησης. Λύνοντας το παρακάτω πρόβλημα βελτιστοποίησης,

$$f^* = \operatorname{argmin}_f \mathbb{E}_{x,y \sim p_{data}} \|y - f(x)\|^2 \quad (4.33)$$

παίρνουμε

$$f^*(x) = \mathbb{E}_{y \sim p_{data}(y|x)}[y] \quad (4.34)$$

εφόσον αυτή η συνάρτηση ανήκει στην κλάση των συναρτήσεων πάνω στην οποία βελτιστοποιείται το πρόβλημα. Δηλαδή, αν το μοντέλο εκπαιδευόταν σε άπειρα δείγματα από την πραγματική κατανομή των δεδομένων, ελαχιστοποιώντας το μέσο τετραγωνικό σφάλμα εκτιμώνται οι προβλέψεις της μέσης τιμής του  $y$  για κάθε  $x$ .

Η επιλογή της συνάρτησης κόστους συνδέεται άμεσα με την επιλογή της συνάρτησης ενεργοποίησης των μονάδων εξόδου [10]. Θα υποθέσουμε ότι το νευρωνικό δίκτυο εμπρόσθιας μετάδοσης μας τροφοδοτεί με κρυφά χαρακτηριστικά τα οποία ορίζονται ως  $h = f(x; \theta)$ . Ο ρόλος του επιπέδου εξόδου είναι να εφαρμόσει έναν επιπλέον μετασχηματισμό από τα χαρακτηριστικά που παίρνει σαν είσοδο με στόχο την ολοκλήρωση της διαδικασίας που το νευρωνικό δίκτυο καλείται να υλοποιήσει. Στην διπλωματική έγινε χρήση γραμμικών μονάδων εξόδου (Linear Units – LU) οι οποίες βασίζονται σε έναν αφινικό μετασχηματισμό χωρίς να εισάγεται καμία μη γραμμικότητα.

Συνήθως, αυτού του τύπου οι μονάδες χρησιμοποιούνται για να παράγουν τον μέσο μίας δεσμευμένης κανονικής κατανομής:

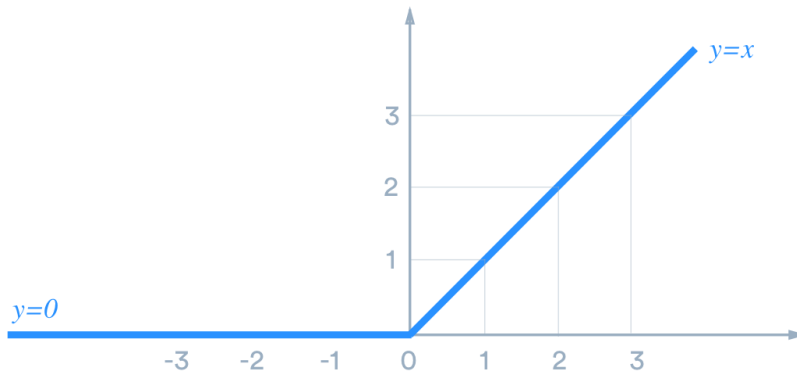
$$p(y|x) = \mathcal{N}(y; \hat{y}, I) \quad (4.35)$$

Σε αυτή την περίπτωση η μεγιστοποίηση του λογαρίθμου της πιθανοφάνειας είναι ισοδύναμη με την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος.

Εξίσου σημαντική είναι και η επιλογή της συνάρτησης ενεργοποίησης των κρυφών επιπέδων του δικτύου. Η πιο διαδεδομένη και αποδεκτή ως προκαθορισμένη συνάρτηση στις περισσότερες βιβλιοθήκες βαθιάς μηχανικής μάθησης είναι η συνάρτηση διορθωμένης γραμμικής μονάδας (Rectified Linear Unit – ReLU) γνωστή και ως ReLU η οποία ορίζεται από την σχέση,

$$\varphi(z) = \begin{cases} 0, & z \leq 0 \\ z, & z > 0 \end{cases} = \max\{0, z\} \quad (4.36)$$

και η μορφή της φαίνεται στην Εικόνα 6.



**Εικόνα 6** Διορθωμένη γραμμική μονάδα (*ReLU*)

Αυτού του τύπου οι συναρτήσεις είναι εύκολο να βελτιστοποιηθούν λόγω της ομοιότητάς τους με τις γραμμικές μονάδες. Η μόνη διαφορά μεταξύ των γραμμικών μονάδων και των διορθωμένων γραμμικών μονάδων είναι ότι η τελευταία είναι 0 στο μισό πεδίου ορισμού της. Αυτό κάνει τις παραγώγους να είναι μεγάλες όταν η μονάδα είναι ενεργή. Η 2<sup>η</sup> παράγωγος είναι σχεδόν παντού μηδέν, ενώ όπου είναι ενεργός ο νευρώνας είναι παντού 1. Αυτό σημαίνει ότι η κατευθυνόμενη παράγωγος είναι πολύ πιο χρήσιμη για μάθηση σε σύγκριση με συναρτήσεις ενεργοποίησης που εμφανίζουν επιρροές 2<sup>ης</sup> τάξης.

Άλλο μία σημαντική υπερπαραμέτρος είναι ο καθορισμός της αρχιτεκτονικής του δικτύου. Ο όρος αρχιτεκτονική αναφέρεται στην συνολική κατασκευή του δικτύου: δηλαδή πόσες μονάδες πρέπει να έχει το δίκτυο και πώς πρέπει να συνδέονται μεταξύ τους. Τα περισσότερα νευρωνικά δίκτυα οργανώνονται σε ομάδες από μονάδες αυτό που ονομάσαμε επίπεδο. Οι περισσότερες αρχιτεκτονικές νευρωνικών δικτύων οργανώνουν αυτά τα επίπεδα σε αλυσιδωτή δομή, με κάθε επίπεδο να είναι συνάρτηση του επιπέδου που προηγείται. Σε αυτή την δομή το 1<sup>ο</sup> επίπεδο δίνεται ως,

$$h^{(1)} = \varphi^{(1)}(W^{(1)}x + b^{(1)}) \quad (4.37)$$

ενώ το 2<sup>ο</sup> επίπεδο δίνεται από την σχέση

$$h^{(2)} = \varphi^{(2)}(W^{(2)}h^{(1)} + b^{(2)}) \quad (4.38)$$

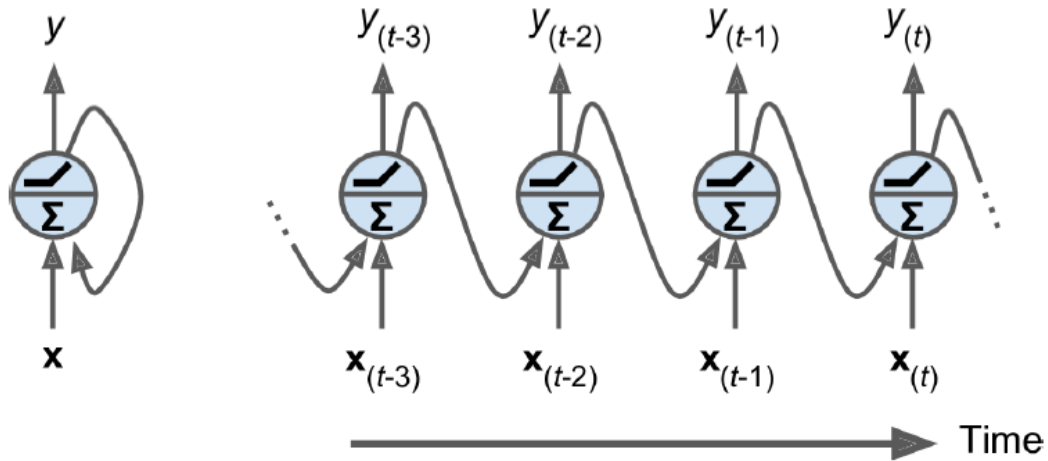
κοκ.

Για τις βασισμένες σε αλυσιδωτή δομή αρχιτεκτονικές, οι κύριες παράμετροι προς επιλογή είναι το βάθος και το πλάτος του δικτύου. Σε πολλές περιπτώσεις ακόμα και ένα κρυφό επίπεδο μπορεί να είναι αρκετό για προσαρμογή στο σύνολο εκπαίδευσης. Είναι γενικά αποδεκτό ότι η καλύτερη αρχιτεκτονική για ένα πρόβλημα πρέπει να βρεθεί μέσω πειραμάτων καθοδηγούμενων από το σφάλμα του συνόλου αξιολόγησης.

Ανακεφαλαιώνοντας, τα νευρωνικά δίκτυα εμπρόσθιας μετάδοσης είναι αρχιτεκτονικές πλήρως συνδεδεμένες τις οποίες μπορούμε να βλέπουμε σαν αποδοτικές μη γραμμικές συναρτήσεις που χρησιμοποιούνται για προσέγγιση συνάρτησης, ελαχιστοποιώντας το σφάλμα προσέγγισης μέσω της μεθόδου κατάβασης κλίσης.

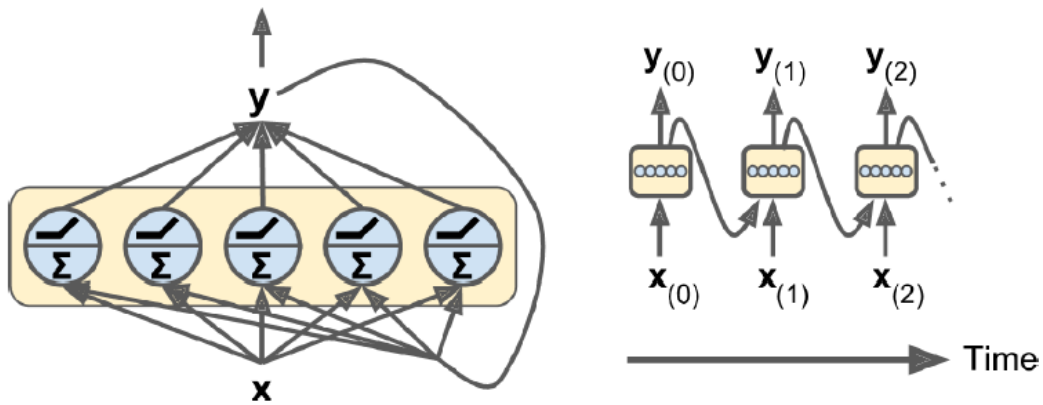
#### 4.4.2 Τα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα

Τα ανατροφοδοτούμενα νευρωνικά δίκτυα (recurrent neural networks-RNNs) [10], [88] είναι μία οικογένεια νευρωνικών δικτύων που «εξειδικεύονται» στην επεξεργασία ακολουθιακών δεδομένων, όπως π.χ. χρονοσειρές. Μοιάζουν με τα νευρωνικά δίκτυα εμπρόσθιας μετάδοσης, με την διαφορά ότι έχουν συνδέσεις που πηγαίνουν προς τα πίσω. Η βασική ιδέα είναι η διαμοίραση των ίδιων παραμέτρων (sharing parameters) σε διαφορετικά μέρη του μοντέλου. Η ιδέα αυτή επιτρέπει την επέκταση και την εφαρμογή του μοντέλου σε παραδείγματα διαφορετικών μορφών (π.χ. διαφορετικού μήκους δεδομένα στην είσοδο) και την δυνατότητα της γενίκευσης. Στην Εικόνα 7 φαίνεται η απλούστερη μορφή ενός RNN, το οποίο αποτελείται από ένα νευρώνα, παράγει μία έξοδο και επιστρέφει την έξοδο πάλι στον εαυτό του. Σε κάθε χρονική στιγμή  $t$ , αυτός ο ανατροφοδοτούμενος νευρώνας έχει ως είσοδο το  $x_{(t)}$  καθώς και την δική του έξοδο την προηγούμενη χρονική στιγμή,  $y_{(t-1)}$ . Στο 1<sup>ο</sup> βήμα (χρονική στιγμή 1) δεν υπάρχει έξοδος από κάποια προηγούμενη χρονική στιγμή άρα παίρνει την τιμή 0. Το δίκτυο μπορεί να αναπαρασταθεί με βάση τον χρονικό άξονα κάνοντας «ξεδίπλωμα» του δικτύου στον χρόνο. Πρόκειται για τον ίδιο νευρώνα ο οποίος απεικονίζεται μία φορά για κάθε χρονική στιγμή.



**Εικόνα 7** Ένας ανατροφοδοτούμενος νευρώνας (αριστερά) "ξεδιπλωμένος" στο χρόνο (δεξιά).

Μπορεί εύκολα να δημιουργηθεί ένα επίπεδο από ανατροφοδοτούμενους νευρώνες. Σε κάθε χρονική στιγμή  $t$ , κάθε νευρώνας δέχεται το διάνυσμα εισόδου  $\mathbf{x}(t)$  και το διάνυσμα εξόδου από την προηγούμενη χρονική στιγμή,  $\mathbf{y}(t-1)$ , όπως φαίνεται στην Εικόνα 8.



**Εικόνα 8** Ένα επίπεδο ανατροφοδοτούμενων νευρώνων (αριστερά) ξεδιπλωμένο στον χρόνο (δεξιά) [88].

Κάθε ανατροφοδοτούμενος νευρώνας έχει δύο σύνολα παραμέτρων: ένα για τις εισόδους  $\mathbf{x}(t)$  και ένα άλλο για τις εξόδους του προηγούμενου βήματος που δέχεται ως είσοδο,  $\mathbf{y}(t-1)$ . Τα διανύσματα βαρών θα συμβολίζονται με  $w_x$  και  $w_y$ , αντίστοιχα. Εάν θεωρηθεί όλο το επίπεδο ως ένα σύνολο, μπορούν να τοποθετηθούν όλα τα διανύσματα βαρών σε δύο πίνακες βαρών,  $W_x$  και  $W_y$ . Το διάνυσμα εξόδου ολόκληρου του ανατροφοδοτούμενου επιπέδου δίνεται από την σχέση,

$$y_{(t)} = \varphi(W_x^T x_{(t)} + W_y^T y_{(t-1)} + b) \quad (4.39)$$

όπου  $b$  είναι το διάνυσμα κατωφλίου (step) και  $\varphi(\cdot)$  είναι η συνάρτηση ενεργοποίησης (π.χ. ReLU).

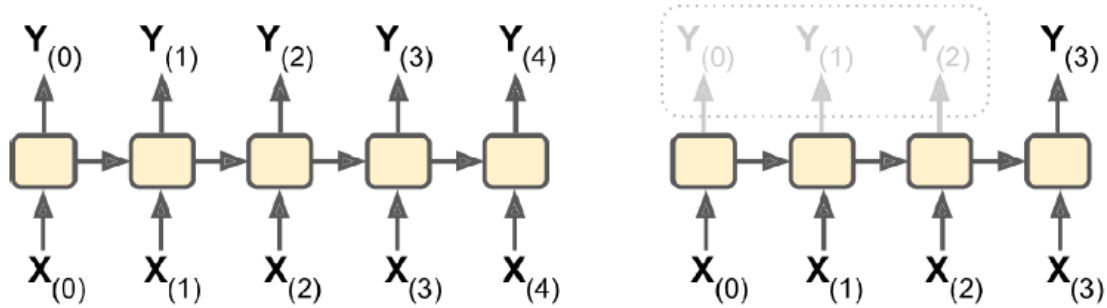
Εφόσον η έξοδος ενός ανατροφοδοτούμενου νευρώνα την χρονική στιγμή  $t$  είναι συνάρτηση όλων των εισόδων από τις προηγούμενες χρονικές στιγμές, το δίκτυο έχει μίας μορφής μνήμη. Ένα τμήμα του νευρωνικού δικτύου που διατηρεί μία κατάσταση διά μήκους κάποιων χρονικών στιγμών ονομάζεται κελί μνήμης ή απλά κελί. Ένας απλός ανατροφοδοτούμενος νευρώνας, ή ακόμα και ένα επίπεδο από ανατροφοδοτούμενους νευρώνες, αποτελεί ένα πολύ βασικό κελί, το οποίο είναι ικανό να μάθει μοτίβα περιορισμένου μήκους.

Γενικά, η κατάσταση ενός κελιού την χρονική στιγμή  $t$ , συμβολίζεται με  $h_{(t)}$ , και είναι μία συνάρτηση κάποιων εισόδων εκείνης της χρονικής στιγμής και της κατάστασης της προηγούμενης χρονικής στιγμής:  $h_{(t)} = f(h_{(t-1)}, x_{(t)})$ . Η έξοδος του κελιού την χρονική στιγμή  $t$ , συμβολίζεται με  $y_{(t)}$ , και είναι επίσης συνάρτηση της εισόδου την χρονική στιγμή  $t$  και των προηγούμενων καταστάσεων. Στην περίπτωση των βασικών κελιών η κατάσταση και η έξοδος συμπίπτουν αλλά για πολύπλοκες αρχιτεκτονικές κελιών αυτό δεν ισχύει. Ένα δίκτυο RNN μπορεί να πάρει ως είσοδο μία ακολουθία και να παράγει μία ακολουθία στην έξοδο. Αυτός ο τύπος δικτύου ονομάζεται ακολουθία-σε-ακολουθία (sequence-to-sequence) και είναι πολύ χρήσιμος για την πρόβλεψη χρονοσειρών. Εναλλακτικά, μπορεί να δοθεί στο δίκτυο μία ακολουθία από εισόδους αγνοώντας όλες τις εξόδους εκτός από εκείνη που αντιστοιχεί στην τελευταία έξοδο. Αυτός ο τύπος ονομάζεται ακολουθία-σε-διάνυσμα (sequence-to-vector).

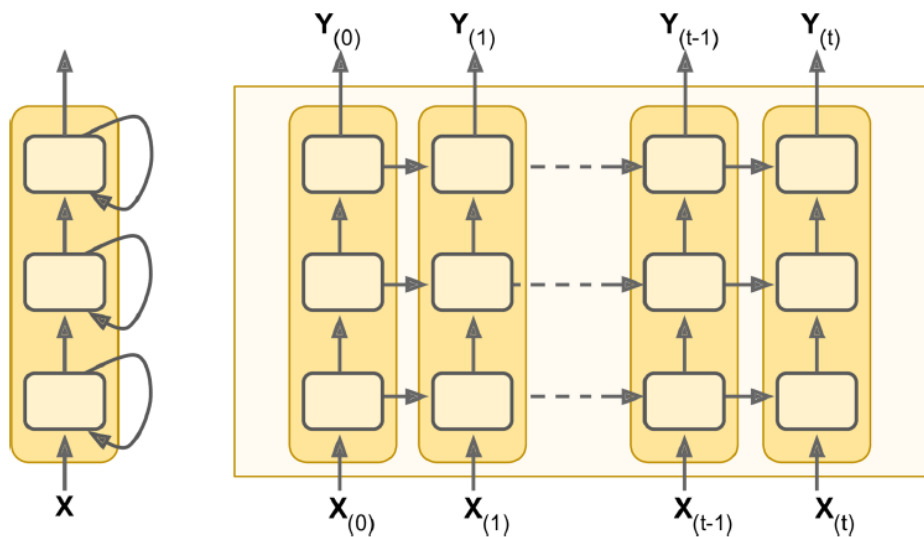
Για την εκπαίδευση ενός δικτύου RNN, ξεδιπλώνεται το δίκτυο στον χρόνο και χρησιμοποιείται ο αλγόριθμος backpropagation που αναφέραμε για νευρωνικά δίκτυα εμπρόσθιας μετάδοσης. Αυτή η στρατηγική ονομάζεται backpropagation στον χρόνο (backpropagation through time-BPTT). Λειτουργώντας σε πλαίσιο διαμοιρασμένων παραμέτρων τα  $W$  και  $b$  είναι ίδια για όλες τις χρονικές στιγμές και έτσι ο αλγόριθμος κάνει την διόρθωση στις παραμέτρους αθροίζοντας πάνω σε όλες τις χρονικές στιγμές.

Συνήθης πρακτική είναι να «κολλάμε» πολλά επίπεδα κελιών έτσι ώστε να έχουμε μία βαθιά αρχιτεκτονική, την οποία ονομάζουμε δίκτυο RNN με βάθος. Στην Εικόνα 10 φαίνεται ένα βαθύ δίκτυο RNN.





Εικόνα 9 Ακολουθία-σε-ακολουθία (αριστερά) και ακολουθία-σε-διάνυσμα (δεξιά) δίκτυο [88].



Εικόνα 10 Δίκτυο RNN με βάθος (αριστερά) ξεδιπλωμένο στον χρόνο (δεξιά) [88].

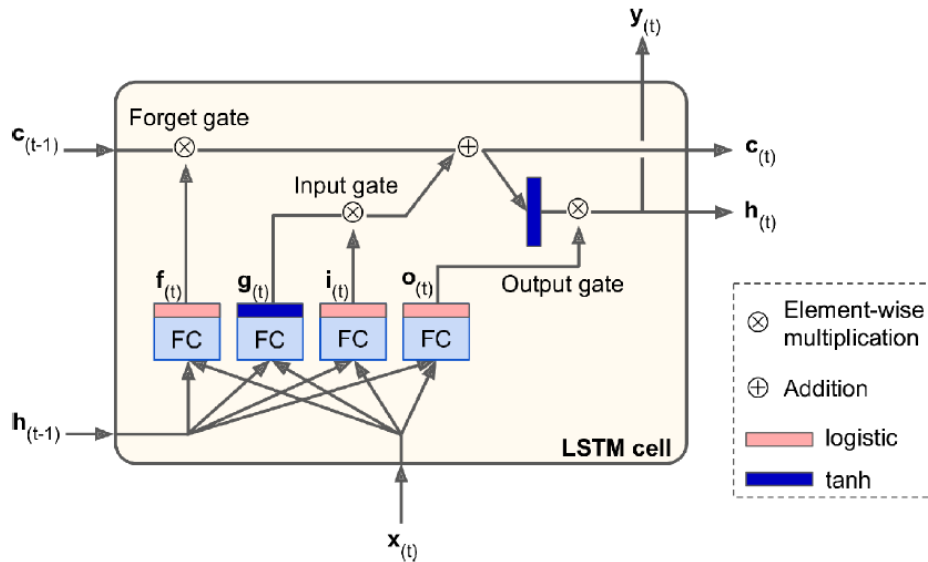
Ένα σημαντικό πρόβλημα που συναντάται στην εκπαίδευση τέτοιων δικτύων είναι οι εξαφανιζόμενες και εκρηγνύμενες κλίσεις. Πολλές φορές οι κλίσεις γίνονται όλο και πιο μικρές καθώς ο αλγόριθμος προχωράει προς τα χαμηλότερα επίπεδα. Αυτό έχει ως αποτέλεσμα η ανανέωση μέσω της κατάβασης κλίσης να μην ανανεώνει τα βάρη στις συνδέσεις των χαμηλότερων επιπέδων, και η εκπαίδευση του δικτύου να μην συγκλίνει ποτέ σε μία καλή λύση. Το πρόβλημα αυτό ονομάζεται εξαφανιζόμενες κλίσεις. Από την άλλη υπάρχουν περιπτώσεις όπου συμβαίνει ακριβώς το αντίθετο, δηλαδή οι κλίσεις να γίνονται όλο και πιο μεγάλες, με αποτέλεσμα τα βάρη των επιπέδων να υπόκεινται σε εξαιρετικά μεγάλες ανανεώσεις βαρών και ο αλγόριθμος τελικά να αποκλίνει. Αυτό το φαινόμενο ονομάζεται εκρηγνύμενες κλίσεις. Γενικά,

μπορεί να σχολιαστεί ότι οι βαθιές αρχιτεκτονικές πάσχουν από τέτοια φαινόμενα ασταθών κλίσεων λόγω του βάθους που έχουν. Διαφορετικά επίπεδα μπορούν να εκπαιδεύονται με τρομακτικά διαφορετικές ταχύτητες.

#### 4.4.3 Το κελί LSTM

Για την αντιμετώπιση των ασταθών κλίσεων προτάθηκε ένα νέο είδος κελιού το Long Short-Term Memory [88], [89] το οποίο έχει δεχθεί αρκετές βελτιώσεις τα τελευταία χρόνια. Το κελί LSTM μπορεί να χρησιμοποιηθεί σαν ένα απλό κελί στην θέση του συμβατικού κελιού, με την διαφορά ότι θα έχει πολύ καλύτερες επιδόσεις γενικά και η εκπαίδευση θα συγκλίνει πολύ πιο γρήγορα και θα μπορεί να «μάθει» της μακροπρόθεσμες εξαρτήσεις πολύ πιο γρήγορα. Η αρχιτεκτονική του κελιού φαίνεται στην Εικόνα 11.

Παρατηρώντας λίγο πιο μακροσκοπικά το κελί LSTM, μοιάζει με ένα απλό κελί με την διαφορά ότι η «κατάσταση» του χωρίζεται σε 2 διανύσματα  $h_{(t)}$  και  $c_{(t)}$ , όπου, το διάνυσμα  $h_{(t)}$  αναπαριστά τη βραχυπρόθεσμη κατάσταση του κελιού την χρονική στιγμή  $t$ , ενώ το διάνυσμα  $c_{(t)}$  τη μακροπρόθεσμη κατάσταση του κελιού την χρονική στιγμή  $t$ .



Εικόνα 11 Το κελί LSTM [88].

Η κεντρική ιδέα είναι ότι το δίκτυο μπορεί να μάθει τι να αποθηκεύσει στο διάνυσμα μακροπρόθεσμης κατάστασης, τι να διαγράψει και τι να διαβάσει από αυτό. Καθώς το διάνυσμα μακροπρόθεσμης κατάστασης  $c_{(t-1)}$  διαπερνά το κελί από τα αριστερά προς τα δεξιά, περνάει

πρώτα από μία πύλη απώλειας μνήμης (forget gate), «πετάει» κάποια μνήμη, μετά προσθέτει μνήμη μέσω ενός τελεστή άθροισης από την πύλη εισόδου (input gate). Το αποτέλεσμα είναι το διάνυσμα μακροπρόθεσμης μνήμης  $c_{(t)}$  το οποίο «στέλνεται» απευθείας έξω από το κελί, χωρίς περαιτέρω μετασχηματισμούς. Έτσι σε κάθε χρονική στιγμή κάποιες μνήμες αφαιρούνται και κάποιες άλλες προστίθενται. Επιπροσθέτως, μετά την εφαρμογή του τελεστή άθροισης, η μακροπρόθεσμη κατάσταση αντιγράφεται και μετασχηματίζεται μέσω της συνάρτησης της υπερβολικής εφαπτομένης ( $\tanh$ ) με το αποτέλεσμα να φιλτράρεται μέσω της πύλης εξόδου (output gate). Το τελευταίο παράγει την βραχυπρόθεσμη κατάσταση  $h_{(t)}$  η οποία είναι και η έξοδος του κελιού,  $y_{(t)}$ , την χρονική στιγμή  $t$ .

Το ερώτημα που φυσικά γεννάται από τα παραπάνω είναι πως δημιουργούνται οι νέες μνήμες και πως η κάθε πύλη λειτουργεί. Η είσοδοι που χρησιμοποιούνται εντός του κελιού LSTM γι' αυτές τις διαδικασίες είναι το διάνυσμα εισόδου  $x_{(t)}$  και το διάνυσμα της βραχυπρόθεσμης κατάσταση της προηγούμενης χρονικής στιγμής  $h_{(t-1)}$ . Αυτά τα 2 διανύσματα δίνονται ως είσοδοι σε 4 πλήρως συνδεδεμένα επίπεδα. Όλα εξυπηρετούν ένα διαφορετικό σκοπό:

- Το κυρίως επίπεδο είναι εκείνο που έχει ως έξοδο το διάνυσμα  $g_{(t)}$ . Ο ρόλος του είναι να αναλύσει την τρέχουσα είσοδο  $x_{(t)}$  και την προηγούμενη βραχυπρόθεσμη κατάσταση  $h_{(t-1)}$ . Η διαδικασία έχει ως εξόδους το διάνυσμα  $y_{(t)}$  και το  $h_{(t)}$ . Στο κελί LSTM η έξοδος αυτού του βασικού επιπέδου χρησιμοποιείται για την 1<sup>η</sup> ανανέωση της μακροπρόθεσμης κατάστασης όπου τα πιο σημαντικά κομμάτια της εξόδου αποθηκεύονται στην μακροπρόθεσμη κατάσταση ενώ τα υπόλοιπα παραλείπονται.
- Τα υπόλοιπα 3 επίπεδα είναι ελεγκτές των πυλών (gate controllers). Χρησιμοποιούν ως συνάρτηση ενεργοποίησης την λογιστική άρα οι έξοδοι τους έχουν εύρος από 0 έως 1. Όλες οι έξοδοι δίνονται ως είσοδοι σε τελεστές πολλαπλασιασμού ανά στοιχείο, έτσι εάν στην έξοδο παράγονται μηδενικά τότε η πύλη είναι κλειστή ενώ στην περίπτωση των άσσων, η πύλη είναι τελείως ανοικτή. Πιο συγκεκριμένα:
  - Η πύλη απώλειας μνήμης (ελέγχεται από την  $f_{(t)}$ ) ελέγχει ποια κομμάτια της μακροπρόθεσμης κατάστασης πρέπει να σβηστούν.
  - Η πύλη εισόδου (ελέγχεται από την  $i_{(t)}$ ) και ελέγχει ποια κομμάτια της εξόδου  $g_{(t)}$  πρέπει να προστεθούν στην μακροπρόθεσμη κατάσταση.

- Η πύλη εξόδου (ελέγχεται από την  $o_{(t)}$ ) ελέγχει ποια κομμάτια της βραχυπρόθεσμης κατάστασης πρέπει να διαβαστούν και να δοθούν ως έξοδοι την τρέχουσα χρονική στιγμή, για το  $h_{(t)}$  και το  $y_{(t)}$ .

Εν συντομία, ένα κελί LSTM μπορεί να αναγνωρίζει μία σημαντική είσοδο (αυτός είναι ο ρόλος της πύλης εισόδου), να το αποθηκεύει στο διάνυσμα μακροπρόθεσμης κατάστασης, να το διατηρεί για όσο χρειάζεται (αυτός είναι ο ρόλος της πύλης απώλειας μνήμης), και τέλος να το εξάγει όποτε χρειάζεται. Αυτό εξηγεί γιατί αυτές οι αρχιτεκτονικές έχουν τεράστια επιτυχία στο να μαθαίνουν μοτίβα σε δεδομένα χρονοσειρών. Τέλος, οι εξισώσεις που χρησιμοποιούνται για τον υπολογισμό των παραπάνω δίνονται από τις σχέσεις (4.40) – (4.45):

$$i_{(t)} = \sigma(W_{xi}^T x_{(t)} + W_{hi}^T h_{(t-1)} + b_i) \quad (4.40)$$

$$f_{(t)} = \sigma(W_{xf}^T x_{(t)} + W_{hf}^T h_{(t-1)} + b_f) \quad (4.41)$$

$$o_{(t)} = \sigma(W_{xo}^T x_{(t)} + W_{ho}^T h_{(t-1)} + b_o) \quad (4.42)$$

$$g_{(t)} = \sigma(W_{xg}^T x_{(t)} + W_{hg}^T h_{(t-1)} + b_g) \quad (4.43)$$

$$c_{(t)} = f_{(t)} \otimes c_{(t-1)} + i_{(t)} \otimes g_{(t)} \quad (4.44)$$

$$y_{(t)} = h_{(t)} = o_{(t)} \otimes \tanh(c_{(t-1)}) + i_{(t)} \otimes g_{(t)} \quad (4.45)$$

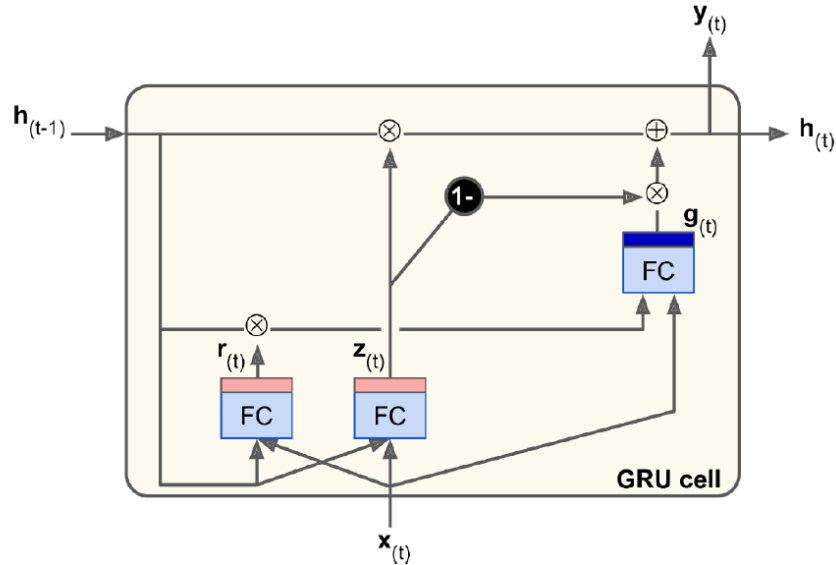
Οι παράμετροι των παραπάνω εξισώσεων είναι:

- $W_{xi}, W_{xf}, W_{xo}, W_{xg}$  είναι οι πίνακες των βαρών για κάθε ένα από τα 4 επίπεδα όσον αφορά τις συνδέσεις τους με το διάνυσμα εισόδου  $x_{(t)}$ .
- $W_{hi}, W_{hf}, W_{ho}, W_{hg}$  είναι οι πίνακες των βαρών για κάθε ένα από τα 4 επίπεδα όσον αφορά τις συνδέσεις τους με το διάνυσμα εισόδου  $h_{(t-1)}$ .
- $b_i, b_f, b_o, b_g$  είναι οι όροι κατωφλίου για κάθε ένα από τα 4 επίπεδα.

#### 4.4.4 Το κελί GRU

Το κελί GRU [88] (Gated Recurrent Unit) προτάθηκε 1<sup>η</sup> φορά από τους *Cho et al.* [90]. Η αρχιτεκτονική του κελιού φαίνεται στην Εικόνα 12. Πρόκειται για μία απλοποιημένη έκδοση του

κελιού LSTM και φαίνεται να έχει τις ίδιες επιδόσεις με το τελευταίο. Λόγω της περίπλοκης λειτουργίας των κελιών LSTM εισάγονται πολλές καινούριες παράμετροι τις οποίες πρέπει να μάθει ο αλγόριθμος, το οποίο κάνει την διαδικασία μάθησης να είναι αρκετά χρονοβόρα.



**Εικόνα 12** Το κελί GRU [88].

Οι κύριες απλοποιήσεις συνοψίζονται στα παρακάτω σημεία:

- Τα 2 διανύσματα κατάστασης ενώνονται σε ένα διάνυσμα  $h(t)$ .
- Υπάρχει μόνο ένας ελεγκτής πύλης  $z(t)$  ο οποίος ελέγχει την πύλη απώλειας μνήμης και την πύλη εισόδου. Εάν ο ελεγκτής πύλης εξάγει την τιμή 1, η πύλη απώλειας μνήμης είναι ανοικτή και η πύλη εισόδου είναι κλειστή, ενώ αν η τιμή που εξάγεται είναι το 0 συμβαίνει ακριβώς το αντίθετο.
- Δεν υπάρχει πύλη εξόδου. Το ενοποιημένο διάνυσμα κατάστασης είναι η έξοδος σε κάθε χρονική στιγμή. Παρόλα αυτά υπάρχει ένας νέος ελεγκτής πύλης  $r(t)$  ο οποίος ελέγχει ποιο μέρος της προηγούμενης κατάστασης θα φανεί στο κυρίως επίπεδο ( $g(t)$ ).

Οι εξισώσεις που περιγράφουν τις λειτουργίες του κελιού είναι:

$$z(t) = \sigma(W_{xz}^T x(t) + W_{hz}^T h(t-1) + b_z) \quad (4.46)$$

$$r_{(t)} = \sigma(W_{xr}^T x_{(t)} + W_{hr}^T h_{(t-1)} + b_r) \quad (4.47)$$

$$g_{(t)} = \tanh(W_{xg}^T x_{(t)} + W_{hg}^T (r_{(t)} \otimes h_{(t-1)})) + b_g \quad (4.48)$$

$$h_{(t)} = z_{(t)} \otimes h_{(t-1)} + (1 - z_{(t)}) \otimes g_{(t)} \quad (4.49)$$

#### 4.4.5 Τα Συνελικτικά Νευρωνικά Δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα [10], [88], [91] αποτελούν ένα εξειδικευμένο είδος νευρωνικών δικτύων για την ανάλυση δεδομένων τα οποία έχουν τοπολογία τύπου πλέγματος. Οι χρονοσειρές είναι ένα τέτοιο παράδειγμα δεδομένων, οι οποίες προτυποποιούνται ως πλέγμα μίας διάστασης με δειγματοληψία σε συγκεκριμένα χρονικά διαστήματα. Γίνεται χρήση μίας μαθηματικής πράξης που ονομάζεται συνέλιξη, η οποία είναι μία γραμμική πράξη, εξ'ού και η προέλευση του ονόματός τους. Τα συνελικτικά νευρωνικά δίκτυα είναι απλά νευρωνικά δίκτυα τα οποία χρησιμοποιούν συνέλιξη στην θέση του συμβατικού πολλαπλασιασμού πινάκων σε τουλάχιστον ένα επίπεδο. Η πράξη της συνέλιξης ουσιαστικά εκμεταλλεύεται 3 σημαντικές ιδέες οι οποίες μπορούν να βοηθήσουν ένα σύστημα μηχανικής μάθησης. Αυτές είναι:

- Αραιές αναπαραστάσεις.
- Διαμοιρασμός παραμέτρων.
- Εξισωτικές αναπαραστάσεις.

Η πράξη της συνέλιξης συνήθως συμβολίζεται με έναν αστερίσκο (\*):

$$s(t) = (x * w)(t) \quad (4.50)$$

Στην ορολογία των συνελικτικών νευρωνικών δικτύων, το 1<sup>ο</sup> όρισμα (x) της συνέλιξης αναφέρεται ως είσοδος ενώ το 2<sup>ο</sup> όρισμα (w) ως μάσκα (kernel), και τέλος η έξοδος ως χάρτης χαρακτηριστικού (feature map). Στις εφαρμογές μηχανικής μάθησης, η είσοδος είναι συνήθως ένας πολυδιάστατος πίνακας από δεδομένα και ο πυρήνας είναι επίσης ένας πολυδιάστατος πίνακας από παραμέτρους τις οποίες πρέπει να μάθει ο αλγόριθμος μάθησης.

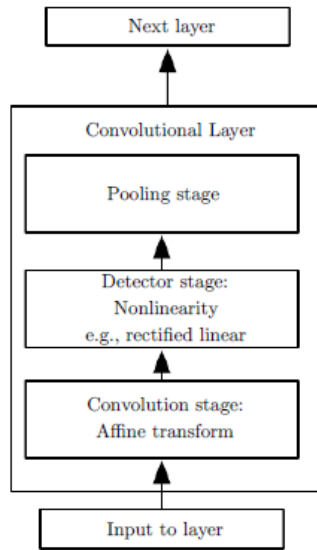
Η διακριτή συνέλιξη μπορεί να αντιμετωπιστεί σαν πολλαπλασιασμός με ένα πίνακα. Παρόλα, αυτά πολλά κελιά του πίνακα είναι περιορισμένα στο να είναι ίδια με άλλα κελιά. Για παράδειγμα, στην μονομεταβλητή διακριτή συνέλιξη, κάθε γραμμή του πίνακα είναι περιορισμένη να είναι ίση με την από πάνω γραμμή μετατοπισμένη κατά ένα κελί. Αυτός ο πίνακας είναι γνωστός ως πίνακας

Toerplitz. Επιπροσθέτως ο πίνακας της συνέλιξης είναι συνήθως αραιός, δηλαδή η πλειοψηφία των κελιών του είναι 0. Αυτό συμβαίνει διότι η μάσκα είναι ένας πίνακας πολύ μικρότερος από τον πίνακα εισόδου.

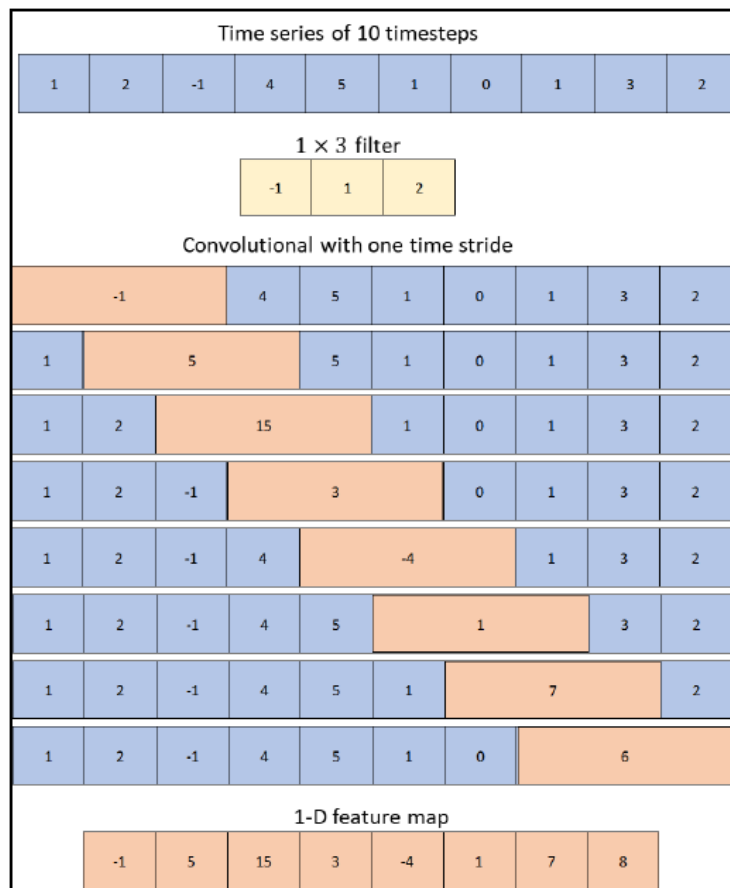
Ένα τυπικό επίπεδο ενός συνελκτικού νευρωνικού δικτύου αποτελείται από 3 στάδια. Στο 1<sup>ο</sup> στάδιο, το επίπεδο εφαρμόζει πολλές συνέλιξεις παράλληλα για να παράγει ένα σύνολο από γραμμικές ενεργοποιήσεις. Στο 2<sup>ο</sup> στάδιο, κάθε γραμμική ενεργοποίηση «περνάει» μέσω μίας μη γραμμικής συνάρτησης, όπως π.χ. η συνάρτηση ReLU. Και τέλος, στο 3<sup>ο</sup> στάδιο χρησιμοποιούμε μία συνάρτηση συγκέντρωσης (pooling function) για την περαιτέρω τροποποίηση της εξόδου του συνελκτικού επιπέδου. Στην Εικόνα 13 φαίνονται σχηματικά τα 3 στάδια-συστατικά ενός τυπικού συνελκτικού επιπέδου.

Μία συνάρτηση συγκέντρωσης αντικαθιστά την έξοδο του δικτύου σε μία συγκεκριμένη τοποθεσία με ένα στατιστικό σύνοψης (summary statistic) των κοντινών εξόδων. Για παράδειγμα, η συνάρτηση max pooling επιστρέφει την μεγαλύτερη από τις εξόδους εντός μίας τετραγωνικής γειτονιάς. Η συγκέντρωση (pooling) βοηθάει στην μετατροπή μίας αναπαράστασης να είναι προσεγγιστικά αμετάβλητη σε μικρές αλλαγές στην είσοδο. Η χρήση του pooling σε χωρικές περιοχές δημιουργεί αναπαραστάσεις που είναι αμετάβλητες όσον αφορά στη μεταφορά (translation). Αν όμως εφαρμοστεί σε εξόδους ξεχωριστά παραμετροποιημένων συνέλιξεων, τα χαρακτηριστικά μπορούν να μάθουν σε ποιους μετασχηματισμούς η αναπαράσταση να γίνει αμετάβλητη.

Για μονομεταβλητές χρονοσειρές χρησιμοποιούνται μονοδιάστατα συνελκτικά επίπεδα με πολλούς μονοδιάστατους πυρήνες να ολισθαίνουν κατά μήκος της ακολουθίας, με κάποιο δοσμένο βήμα, παράγοντας έτσι ένα μονοδιάστατο χάρτη χαρακτηριστικού ανά μάσκα. Κάθε μάσκα μαθαίνει να εντοπίζει ένα μοναδικό μικρού μήκους ακολουθιακό μοτίβο, όχι μεγαλύτερο από το μέγεθος του πυρήνα. Στην Εικόνα 14 φαίνεται ένα παράδειγμα μονοδιάστατης συνέλιξης για μονομεταβλητές χρονοσειρές. Ένα πρόβλημα με την κλασσική συνέλιξη και τα δεδομένα χρονοσειρών είναι ότι εάν χρησιμοποιηθούν πάνω από ένα συνελκτικά επίπεδα τότε εισάγεται στο μοντέλο look-ahead μεροληψία και αυτό γιατί στην κλασσική συνέλιξη δεν λαμβάνεται υπόψη η κατεύθυνση της συνέλιξης. Αυτό γίνεται εμφανές στην Εικόνα 15.



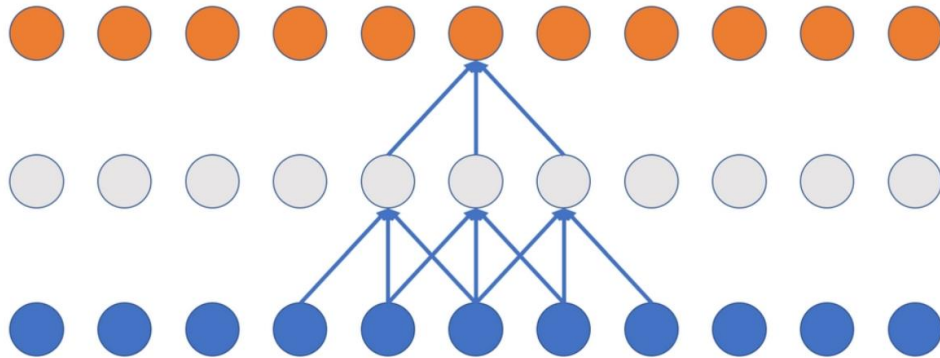
**Εικόνα 13** Τα 3 στάδια ενός τυπικού συνελκτικού επιπέδου.



**Εικόνα 14** Παράδειγμα μονοδιάστατης συνελίξης για μονομεταβλητές χρονοσειρές [92].



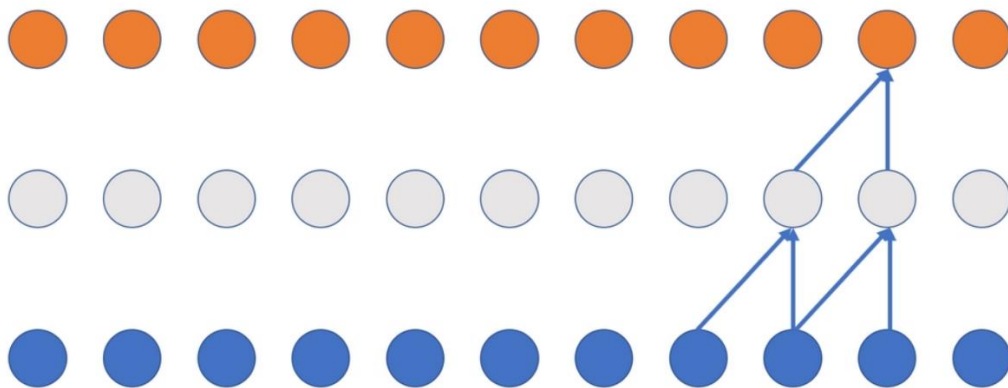
### Standard Convolution



**Εικόνα 15** Look-ahead μεροληψία στην κλασσική συνέλιξη [94].

Όπως το συνελκτικό φίλτρο ολισθαίνει κοιτάει τόσο στο μέλλον όσο και στο παρελθόν όπως φαίνεται στην Εικόνα 15. Γι' αυτό τον λόγο έχει εισαχθεί η αιτιώδης συνέλιξη (causal convolution) η οποία λαμβάνει υπόψη την κατεύθυνση με την οποία κινείται το φίλτρο, δηλαδή κινεί το φίλτρο προς τα δεξιά κατά τρόπο ώστε η έξοδος την χρονική στιγμή  $t$  να αντλεί πληροφορία μόνο από την χρονική στιγμή  $t$  και πριν. Στην Εικόνα 16 απεικονίζεται η αρχιτεκτονική της αιτιώδους συνέλιξης.

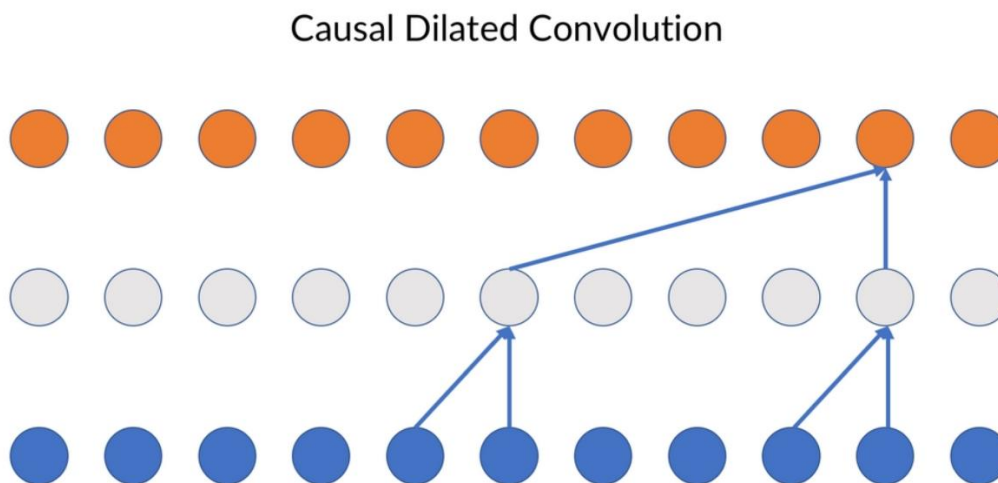
### Causal Convolution



**Εικόνα 16** Αιτιώδης συνέλιξη [94].

#### 4.4.6 Τα Διεσταλμένα Αιτιώδη Συνελκτικά Νευρωνικά Δίκτυα

Σε μία επιστημονική εργασία του 2016 [93] εισάχθηκε μία νέα αρχιτεκτονική με το όνομα WaveNet. Οι συγγραφείς «κόλλησαν» μονοδιάστατα αιτιώδη συνελκτικά επίπεδα, διπλασιάζοντας σε κάθε επίπεδο τον ρυθμό της διαστολής (dilation). Με τον όρο διαστολή εννοούμε πόσο διεσπαρμένες είναι οι εισοδοί του κάθε νευρώνα δηλαδή η μάσκα έχει πρόσβαση κάθε  $n$  στοιχεία, όπως φαίνεται στην Εικόνα 17. Το 1<sup>ο</sup> συνελκτικό επίπεδο έχει ως εισόδους μία χρονική στιγμή πίσω, το 2<sup>ο</sup> συνελκτικό επίπεδο 2 χρονικές στιγμές πίσω, το 3<sup>ο</sup> συνελκτικό επίπεδο 4 χρονικές στιγμές πίσω κ.ο.κ. Με αυτόν τον τρόπο τα χαμηλότερα επίπεδα μαθαίνουν τα βραχυπρόθεσμα μοτίβα ενώ τα πιο ψηλά επίπεδα μαθαίνουν τα μακροπρόθεσμα μοτίβα, με κάθε επίπεδο να αφοσιώνεται σε διαφορετική χρονική ανάλυση. Αποδεικνύεται ότι η εκθετική αύξηση της διαστολής είναι η πιο αποδοτική από την σκοπιά της βραχυπρόθεσμης και μακροπρόθεσμης μνήμης.

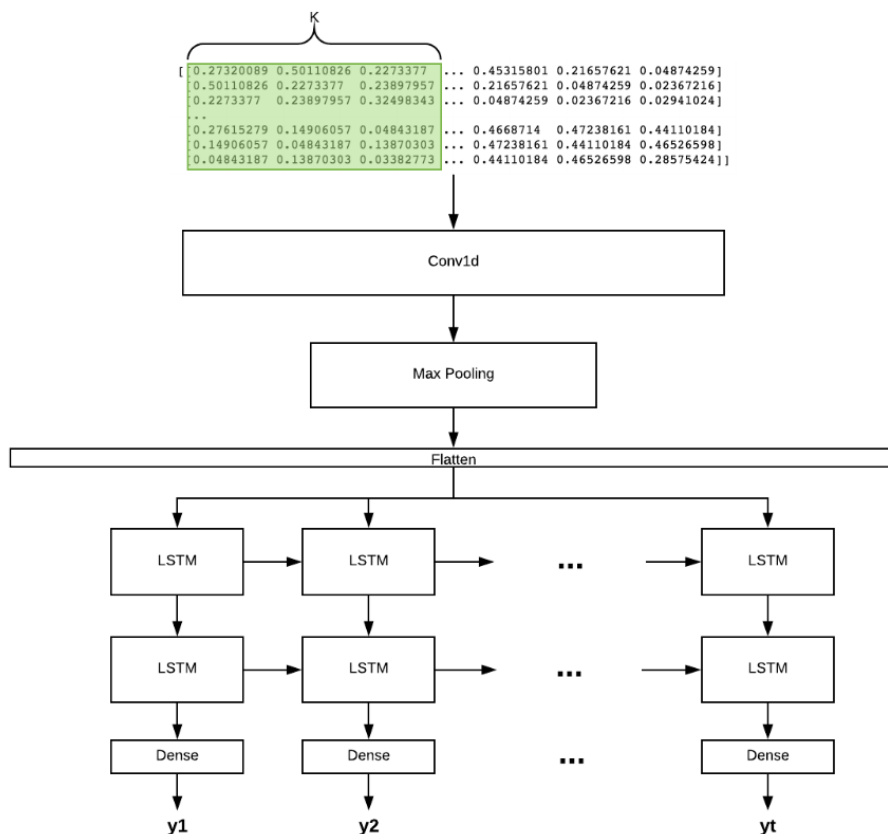


**Εικόνα 17** Αιτιώδης διεσταλμένη συνέλιξη [94].

#### 4.4.7 Η Μέθοδος CNN-LSTM

Με τον όρο υβριδικές μέθοδοι εννοούμε εκείνες τις μεθόδους που συνδυάζουν άλλες μεθόδους. Σε αυτές συγκαταλέγεται και η μέθοδος CNN-LSTM. Η μέθοδος αυτή αποτελείται από ένα ένα ή περισσότερα αιτιώδη συνελκτικά επίπεδα και η έξοδος καταλήγει σε ένα ανατροφοδοτούμενο νευρωνικό δίκτυο με κελιά LSTM. Ουσιαστικά, χρησιμοποιούνται αρχικά τα συνελκτικά μπλοκ για την εξαγωγή των σημαντικότερων χαρακτηριστικών από την ακολουθία των δεδομένων σεβόμενοι την χρονική σειρά των δεδομένων, τα οποία έπειτα εισάγονται στο δίκτυο LSTM προς

την περαιτέρω εκμάθηση των μακροπρόθεσμων και βραχυπρόθεσμων μοτίβων σε αυτά. Στην Εικόνα 18 φαίνεται ένα παράδειγμα της αρχιτεκτονικής ενός τέτοιου μοντέλου.



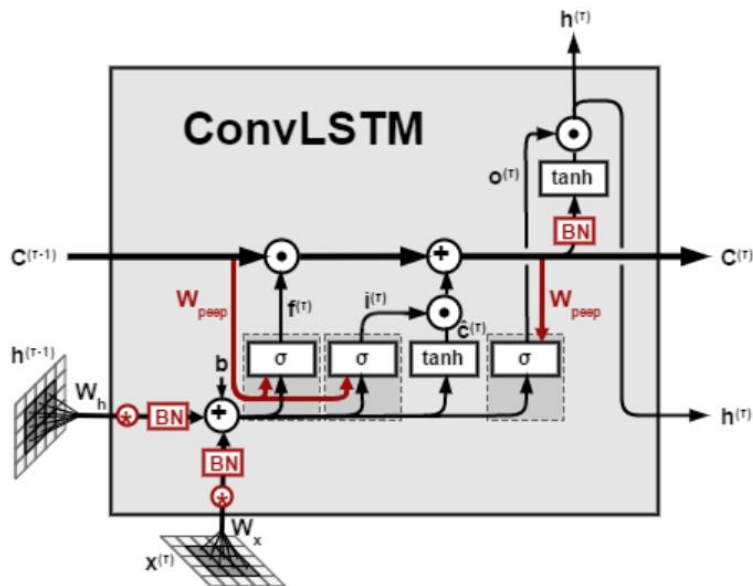
Εικόνα 18 Δίκτυο CNN-LSTM.

#### 4.4.8 Η Μέθοδος των Συνελκτικών LSTM δικτύων

Τα Conv-LSTM είναι επέκταση των κλασικών LSTM που προτάθηκε από τους *Shi et al.* [96] έτσι ώστε το κελί LSTM να έχει ενσωματωμένες συνελκτικές δομές στην μετάβαση εισόδου-σε-κατάσταση και κατάστασης-σε-κατάσταση. Ουσιαστικά, αντικαθίσταται η πράξη του πολλαπλασιασμού πινάκων με την πράξη της συνέλιξης σε κάθε πύλη του κελιού LSTM. Στην Εικόνα 19 φαίνεται η δομή ενός κελιού Conv-LSTM.

Σε ένα τυπικό κελί LSTM, οι ελεγκτές πυλών μπορούν να «κοιτάξουν» μόνο στην είσοδο  $x_{(t)}$  και στην προηγούμενη βραχυπρόθεσμη κατάσταση  $h_{(t-1)}$ . Πολλές φορές θεωρείται καλή ιδέα να δοθεί περισσότερη ευελιξία στο μοντέλο με το να επιτραπεί στο κελί να κοιτάξει στο διάνυσμα μακροπρόθεσμης κατάστασης. Αυτή η ιδέα προτάθηκε από τους *Gers et al.* [97], οι οποίοι

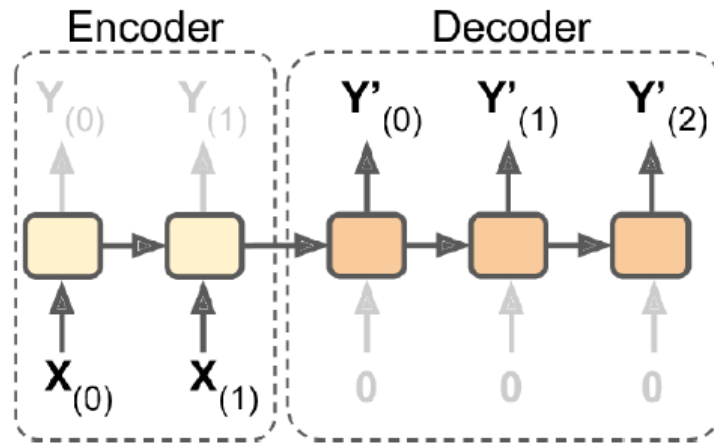
πρότειναν την παραλλαγή του κελιού LSTM με τις επιπλέον συνδέσεις, οι οποίες ονομάζονται peephole συνδέσεις. Η προηγούμενη μακροπρόθεσμη κατάσταση  $c_{(t-1)}$  προστίθεται σαν είσοδος στην πύλη απώλειας μνήμης και εισόδου, και το τρέχον διάνυσμα μακροπρόθεσμης κατάστασης  $c_{(t)}$  προστίθεται σαν είσοδος στον ελεγκτή της πύλης εξόδου.



Εικόνα 19 Το κελί Conv-LSTM.

#### 4.4.9 Μέθοδοι Κωδικοποιητή-Αποκωδικοποιητή

Πολλές φορές στην πρόβλεψη χρονοσειρών απαιτείται η πρόβλεψη πολλών βημάτων (multistep prediction). Κατάλληλη για αυτή την προσέγγιση θεωρείται η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder). Όπως είχαμε περιγράψει προηγούμενως υπάρχουν αρχιτεκτονικές όπως ακολουθία-σε-διάνυσμα και ακολουθία-σε-ακολουθία. Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή αποτελείται από ένα δίκτυο, τον κωδικοποιητή (encoder), που αντιστοιχεί μία ακολουθία (δεδομένα εισόδου) σε διάνυσμα, ακολουθούμενο από ένα δεύτερο δίκτυο, τον αποκωδικοποιητή (decoder), που αντιστοιχεί το διάνυσμα εξόδου του κωδικοποιητή σε μία ακολουθία μήκους  $n$  (πρόβλεψη  $n$  βημάτων). Συνολικά επομένως, γίνεται μία αντιστοίχιση από ακολουθία σε ακολουθία. Η γενική ιδέα είναι ότι πρώτα κοιτάει το μοντέλο όλη την ακολουθία και μετά παράγει την πρόβλεψη. Η αρχιτεκτονική του κωδικοποιητή-αποκωδικοποιητή φαίνεται στην Εικόνα 20.



Εικόνα 20 Αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή [88].

Για τις ανάγκες της διπλωματικής αναπτύχθηκαν 3 αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή συνδυάζοντας διαφορετικές αρχιτεκτονικές εκτός του κωδικοποιητή και του αποκωδικοποιητή. Αυτές είναι:

- **Encoder-Decoder LSTM**

Σε αυτή τη μέθοδο ο κωδικοποιητής και ο αποκωδικοποιητής αποτελούνται κάθε ένας από ένα δίκτυο LSTM.

- **Encoder-Decoder CNN-LSTM**

Σε αυτή τη μέθοδο ο κωδικοποιητής αποτελείται από ένα δίκτυο CNN, ενώ ο αποκωδικοποιητής από ένα δίκτυο LSTM.

- **Encoder-Decoder Conv-LSTM**

Σε αυτή τη μέθοδο ο κωδικοποιητής αποτελείται από ένα δίκτυο Conv-LSTM, ενώ ο αποκωδικοποιητής από ένα δίκτυο LSTM.

Γενικά, υπάρχουν 3 μέθοδοι για την επίτευξη προβλέψεων πολλών βημάτων μπροστά για τις μεθόδους μηχανικής μάθησης [105], [107]. Αυτές είναι συνοπτικά:

- **Επαναληπτική πρόβλεψη (Iterative forecasting)**

Στην επαναληπτική πρόβλεψη το μοντέλο εκπαιδεύεται στο να κάνει προβλέψεις ενός βήματος μπροστά μόνο. Κάνοντας αυτό είναι δυνατό να γίνουν περαιτέρω προβλέψεις χρησιμοποιώντας την πρόβλεψη από το 1<sup>ο</sup> βήμα αντί της πραγματικής τιμής στην αντίστοιχη χρονική στιγμή, παίρνοντας την πρόβλεψη του 2<sup>ου</sup> βήματος κοκ. Όπως γίνεται κατανοητό όσο αυξάνεται ο ορίζοντας πρόβλεψης, οι νέες προβλέψεις εξαρτώνται από την ακρίβεια των προηγούμενων, που σημαίνει ότι η ακρίβεια πιο μακροπρόθεσμων προβλέψεων θα μειώνεται σταδιακά. Πλεονέκτημα της μεθόδου είναι η απλότητα της και η ευκολία στον να υπολογιστεί.

- **Απευθείας πρόβλεψη (Direct forecasting)**

Στην απευθείας πρόβλεψη το μοντέλο παράγει προβλέψεις N βημάτων μπροστά και όχι ενός βήματος όπως στην προηγούμενη περίπτωση. Αυτό μπορεί να επιτευχθεί για όλες τις μεθόδους προστίθοντας ένα πυκνό δίκτυο 2 επιπέδων συνήθως όπου το τελευταίο επίπεδο στην έξοδο του έχει N νευρώνες, ένα για κάθε ορίζοντα. Έτσι, ο 1<sup>ος</sup> νευρώνας εξόδου είναι υπεύθυνος για τον ορίζοντα πρόβλεψης  $h=1$ , ο 2<sup>ος</sup> νευρώνας είναι υπεύθυνος για τον ορίζοντα πρόβλεψης  $h=2$ , κοκ.

- **Πρόβλεψη πολλαπλών μοντέλων (multi-neural network forecasting)**

Σε αυτήν την μέθοδο πολλαπλά μοντέλα εκπαιδεύονται ανά ορίζοντα πρόβλεψης. Για την πρόβλεψη N βημάτων μπροστά, εκπαιδεύονται N διαφορετικά μοντέλα που κάθε ένα «εξειδικεύεται» σε έναν ορίζοντα πρόβλεψης.

Για τις ανάγκες της διπλωματικής χρησιμοποιήθηκε για όλες τις μεθόδους βαθιάς μηχανικής μάθησης η μέθοδος της απευθείας πρόβλεψης με το προτελευταίο επίπεδο να έχει 100 νευρώνες και το τελευταίο να έχει 12, αριθμός ίσος με τα βήματα της πρόβλεψης. Τέλος, για την υλοποίηση των μεθόδων της ομάδας 3 έγινε χρήση της γλώσσας Python, και πιο συγκεκριμένα του API Keras με Backend το Tensorflow.

## 5 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

---

Στο κεφάλαιο 5 περιγράφεται η μεθοδολογική προσέγγιση που ακολουθήθηκε για την εφαρμογή των μεθόδων που περιγράψαμε στο προηγούμενο κεφάλαιο η οποία περιλαμβάνει τα εξής:

- Περιγραφή του προβλήματος.
- Δημιουργία συνόλου εκπαίδευσης, αξιολόγησης και ελέγχου.
- Μέθοδος αξιολόγησης της προβλεπτικής ικανότητας.
- Μετρικές αξιολόγησης.
- Μέθοδοι προεπεξεργασίας για τις μεθόδους βαθιάς μηχανικής μάθησης.
- Βελτιστοποίηση υπερπαραμέτρων και κανονικοποίηση.

### 5.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Επιλύεται ένα πρόβλημα πρόβλεψης για μονομεταβλητές κυκλοφοριακές χρονοσειρές. Υπάρχει λοιπόν ένα χαρακτηριστικό (μία τιμή) ανά χρονική στιγμή, η μέση ταχύτητα διαδρομής της οδού ανά 5 λεπτά. Τα δεδομένα αφορούν 2 μήνες (Οκτώβριος, Νοέμβριος 2016), συνολικά 61 μέρες, δειγματοληψίας ανά 5 λεπτο. Να σημειωθεί ότι η δειγματοληψία είναι ομοιόμορφη στον χρόνο, το οποίο μεταφράζεται σε 17568 διαθέσιμα 5λεπτα. Όποτε το τελικό σύνολο δεδομένων είναι ένα διάνυσμα με 17568 τιμές όπου κάθε τιμή περιγράφει την μέση ταχύτητα της οδού στο αντίστοιχο 5λεπτο, αποθηκευμένες με χρονική σειρά ( $1^{\text{η}}$  θέση =  $1^{\circ}$  5λεπτο, ...,  $n^{\text{η}}$  θέση =  $n^{\circ}$  5λεπτο κοκ).

Για την προτυποποίηση του προβλήματος θα χρησιμοποιηθούν μόνο ενδογενείς μεταβλητές, δηλαδή δεν θα χρησιμοποιηθούν εξωγενείς μεταβλητές όπως καιρός, αργίες κλπ. Τα δεδομένα φαίνεται να παρουσιάζουν χρονοεξαρτώμενα μοτίβα, όπως τάση και εποχικότητα, και άρα έχουν δομή (structured data). Σκοπός είναι η πρόβλεψη της μέσης ταχύτητας για πάνω από ένα χρονικό βήμα (5λεπτο) άρα απαιτείται πρόβλεψη πολλών βημάτων (multi-step prediction). Αυτό σημαίνει ότι η έξοδος των μεθόδων ανά παρατήρηση θα είναι ένα μοναδιάστατο διάνυσμα με θέσεις όσα είναι τα 5λεπτα που προβλέπουμε μπροστά στον χρόνο. Η βραχυπρόθεσμη πρόβλεψη κυκλοφορίας συνήθως αφορά πρόβλεψη μέχρι 2 ώρες μπροστά. Εδώ επιλέγεται πρόβλεψη για 12 5λεπτα μπροστά στον χρόνο, δηλαδή ο ορίζοντας πρόβλεψης εκτείνεται χρονικά από 5 λεπτά έως 60 λεπτά μπροστά. Να σημειωθεί ότι από εδώ και πέρα στο κείμενο θα αναφερόμαστε στον ορίζοντα πρόβλεψης σε όρους βημάτων (1-12).

## 5.2 ΔΗΜΙΟΥΡΓΙΑ ΣΥΝΟΛΟΥ ΕΚΠΑΙΔΕΥΣΗΣ, ΑΞΙΟΛΟΓΗΣΗΣ ΚΑΙ ΕΛΕΓΧΟΥ

Για την αξιολόγηση της προβλεπτικής ικανότητας ενός μοντέλου πρέπει να ελέγξουμε τις επιδόσεις του σε δεδομένα τα οποία δεν έχει «δει», δηλαδή σε δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την φάση της εκπαίδευσης. Στην γλώσσα της μηχανικής μάθησης ονομάζεται γενίκευση. Στους περισσότερους αλγόριθμους μάθησης γίνεται μία τυχαία μετάθεση δεδομένα πριν χωριστούν σε δύο ξένα μεταξύ τους υποσύνολα. Το ένα ονομάζεται σύνολο εκπαίδευσης (training set) και χρησιμοποιείται για την επιλογή του καταλληλότερου μοντέλου, την εκμάθηση των παραμέτρων, την επιλογή των μεταβλητών κα. ενώ το άλλο ονομάζεται σύνολο ελέγχου (test set) το οποίο δεν συμμετεχει στην διαδικασία της εκπαίδευσης και χρησιμοποιείται για την τελική αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου. Στο πρόβλημα της πρόβλεψης χρονοσειρών βασικό ζητούμενο είναι η εκμάθηση του παρελθόντος με σκοπό την πρόβλεψη του μέλλοντος. Γίνεται λοιπόν εύκολα αντιληπτό ότι για δεδομένα χρονοσειρών εκ φύσεως δεν επιτρέπεται η τυχαία μετάθεση που αναφέρθηκε προηγουμένως καθώς αυτά συνοδεύονται από μία χρονική δομή η οποία είναι καίριας σημασίας να διατηρηθεί.

Παραλείποντας λοιπόν το τυχαίο ανακάτεμα, γίνεται ο διαχωρισμός των δεδομένων σε ένα σύνολο εκπαίδευσης το οποίο περιλαμβάνει τις 57 πρώτες ημέρες (16416 5λεπτα) και ένα σύνολο ελέγχου που περιλαμβάνει τις τελευταίες 4 μέρες (Σάββατο, Κυριακή, Δευτέρα, Τρίτη-1152 5λεπτα). Σε αυτό το σημείο υπάρχουν κάποιες διαφορές μεταξύ των μεθόδων στατιστικής και υπολογιστικής νοημοσύνης. Η επιλογή του καταλληλότερου μοντέλου για την ίδια οικογένεια μεθόδων γίνεται με χρήση του κριτηρίου AIC. Αυτό επιτρέπει απευθείας την χρήση όλου του συνόλου εκπαίδευσης. Έχει αποδειχθεί εξάλλου εμπειρικά ότι για αυτοματοποιημένες διαδικασίες επιλογής μεθόδων το κριτήριο πληροφορίας AIC είναι καταλληλότερο σε σχέση με άλλες μεθόδους όπως πρόβλεψη εκτός δείγματος (out-of-sample).

Για τις μεθόδους βαθιάς μηχανικής μάθησης η διαδικασία αυτή αλλάζει. Συνήθως το σύνολο εκπαίδευσης διασπάται περαιτέρω σε ένα νέο σύνολο εκπαίδευσης και ένα σύνολο αξιολόγησης. Κατά την εκπαίδευση των μεθόδων βαθιάς μηχανικής μάθησης χρησιμοποιηθεί μία μέθοδος που ονομάζεται πρόωρο σταμάτημα (early stopping) και γι' αυτό θα εργαστούμε λίγο διαφορετικά. Θα γίνει ο διαχωρισμός του συνόλου εκπαίδευσης σε νέα 3 σύνολα: το «προσωρινό» σύνολο εκπαίδευσης, το σύνολο αξιολόγησης και το «προσωρινό» σύνολο ελέγχου της προβλεπτικής ικανότητας του μοντέλου.



Κατά την διαδικασία της επιλογής των καλύτερων παραμέτρων και υπερπαραμέτρων κάθε μεθόδου, τα μοντέλα θα εκπαιδεύονται πάνω στο «προσωρινό» σύνολο εκπαίδευσης. Το σύνολο αξιολόγησης θα χρησιμοποιείται για την εφαρμογή του πρόωρου σταματήματος για την αποφυγή της υπερεκπαίδευσης των μοντέλων, ενώ το «προσωρινό» σύνολο ελέγχου προβλεπτικής ικανότητας θα χρησιμοποιείται για την αξιολόγηση κάθε διαφορετικά παραμετροποιημένου μοντέλου ως προς την προβλεπτική του ικανότητα σε δεδομένα που δεν έχει δει. Αφού για κάθε ομάδα μοντέλων επιλεγεί η καλύτερη παραμετροποίηση, κάθε τελικό μοντέλο θα εκπαιδευτεί στο αρχικό σύνολο εκπαίδευσης και έπειτα θα ελέγχεται στο αρχικό σύνολο ελέγχου ως προς την προβλεπτική του ικανότητα.

Άλλη μία σημαντική διαφορά είναι ο τρόπος με τον οποίο τα δεδομένα εισάγονται στο μοντέλο για κάθε μία ομάδα. Στην ομάδα 1 και ομάδα 2 που αποτελούν κλασσικές μεθόδους χρονοσειρών τα δεδομένα εισάγονται «ολόκληρα», ως ένα διάνυσμα. Στον κλάδο της μηχανικής μάθησης η πρόβλεψη χρονοσειράς αντιμετωπίζεται ως πρόβλημα επιβλεπόμενης μάθησης. Αυτό επιτυγχάνεται με την χρήση «παρελθοντικών» τιμών ως μεταβλητές εισόδου και τις επόμενες ως προς αυτές («μελλοντικές») τιμές ως μεταβλητές εξόδου. Έστω μία χρονοσειρά με 1000 παρατηρήσεις  $X = \{t_1, \dots, t_{1000}\}$ . Θα δημιουργηθεί ένα σύνολο από παρατηρήσεις  $(\vec{x}_i, \vec{y}_i)$ . Έστω τώρα ότι επιζητείται η προτυποποίηση μίας AR(8) διαδικασίας και πρόβλεψη 12 βημάτων μπροστά. Αυτό σημαίνει ότι δοσμένων 8 παρατηρήσεων σε χρονική σειρά, θα γίνεται πρόβλεψη για τις επόμενες 12 παρατηρήσεις (χρονικά). Όποτε γράφουμε,

$$1^{\text{η}} \text{ παρατήρηση: } x_1 = (t_1, \dots, t_8), \quad y_1 = (t_9, \dots, t_{20})$$

$$2^{\text{η}} \text{ παρατήρηση: } x_2 = (t_2, \dots, t_9), \quad y_2 = (t_{10}, \dots, t_{21})$$

⋮

$$981^{\text{η}} \text{ παρατήρηση: } x_{981} = (t_{981}, \dots, t_{988}), \quad y_{981} = (t_{989}, \dots, t_{1000})$$

Το νέο σύνολο δεδομένων θα αποτελείται από 980 παρατηρήσεις που κάθε μία δημιουργείται προχωρώντας κάθε φορά 1 βήμα σε κάθε επόμενη παρατήρηση. Χρησιμοποιείται δηλαδή η μέθοδος του κυλιόμενου παραθύρου με κύλιση ενός βήματος (ενός 5λέπτου). Αν με  $p$  συμβολίσουμε την τάξη της AR διαδικασίας δηλαδή πόσο πίσω κοιτάει το μοντέλο στον χρόνο (παράθυρο) και με  $h$  των αριθμών των βημάτων που ζητούνται για πρόβλεψη, τότε για  $n$  παρατηρήσεις στο διάνυσμα τις χρονοσειράς προκύπτουν τα ζευγάρια,

$$(x_i, y_i), \quad i = 1, \dots, (n - p - h + 1)$$

όπου

$$x_i = \{t_i, t_{i+1}, \dots, t_{i+p-1}\} \in \mathbb{X} \subset \mathbb{R}^p$$

$$y_i = \{t_{i+p}, t_{i+p+1}, \dots, t_{i+p+h-1}\} \in \mathbb{Y} \subset \mathbb{R}^h$$

Ο σκοπός είναι η κατασκευή ενός μοντέλου  $f: \mathbb{X} \rightarrow \mathbb{Y}$ , όπου με  $f$  συμβολίζεται η συνάρτηση παλινδρόμησης. Με άλλα λόγια, η αρχή πίσω από αυτήν την προσέγγιση είναι ο υπολογισμός της δεσμευμένης κατανομής της  $i$ -οστής παρατήρησης δοσμένων των  $p$  προηγούμενων τιμών:  $f(y_i | x_i)$ . Με Η χρονική συσχέτιση μοντελοποιείται χρησιμοποιώντας τις παρελθοντικές τιμές της χρονοσειράς ως επεξηγηματικές μεταβλητές, και ουσιαστικά, θα μπορούσε να χρησιμοποιηθεί οποιαδήποτε μέθοδο μηχανικής μάθησης που λύνει ένα πρόβλημα γραμμικής παλινδρόμησης.

### 5.3 ΜΕΘΟΔΟΣ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΠΡΟΒΛΕΠΤΙΚΗΣ ΙΚΑΝΟΤΗΤΑΣ

Για την αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου χρησιμοποιήθηκε μία παραλλαγή της μεθόδου walk-forward αξιολόγησης, στην γλώσσα της στατιστικής και των χρονοσειρών ονομάζεται μέθοδος κυλιόμενης αρχής (rolling origin). Γενικά, οι ερευνητές που ασχολούνται με μεθόδους πρόβλεψης συμφωνούν ότι η ακρίβεια του μοντέλου πρέπει να κρίνεται με βάση δεδομένα ελέγχου εκτός δείγματος (out-of-sample) και όχι με μεθόδους καλής προσαρμογής στα ιστορικά δεδομένα (in-sample δεδομένα). Κάποιοι από τους λόγους είναι οι εξής [101]:

- Το σφάλμα στο σύνολο εκπαίδευσης τείνει να υποεκτιμά το σφάλμα της πρόβλεψης (out-of-sample). Η επιλογή μεθόδου και η εκτίμηση των παραμέτρων είναι σχεδιασμένες να εκπαιδεύουν μία διαδικασία πρόβλεψης στα ιστορικά δεδομένα. Όμως δεν είναι φυσιολογικό να περιμένει κανείς ότι οι ιδιοσυγκρασίες του παρελθόντος θα επιμείνουν στο μέλλον, και ότι οι ιδιοσυγκρασίες του μέλλοντος έχουν αποκαλυφθεί ήδη στο παρελθόν.
- Οι περισσότερες στατιστικές μέθοδοι βασίζονται σε διαδικασίες ανανέωσης, όπου η πρόβλεψη γίνεται σαν να βρισκόμαστε πάντα ένα χρονικό βήμα πίσω. Σε αυτές τις περιπτώσεις τα μέτρα καλής προσαρμογής βασίζονται στο σφάλμα πρόβλεψης ενός βήματος μπροστά (one-step-ahead forecast error). Αυτό δημιουργεί πρόβλημα σε προβλήματα πρόβλεψης όπου μας

ενδιαφέρει να γίνει πρόβλεψη πολλών βημάτων, και αυτό διότι η έρευνα έχει δείξει ότι τα σφάλματα για το μακρινό μέλλον είναι μεγαλύτερα από εκείνα του ενός βήματος μπροστά.

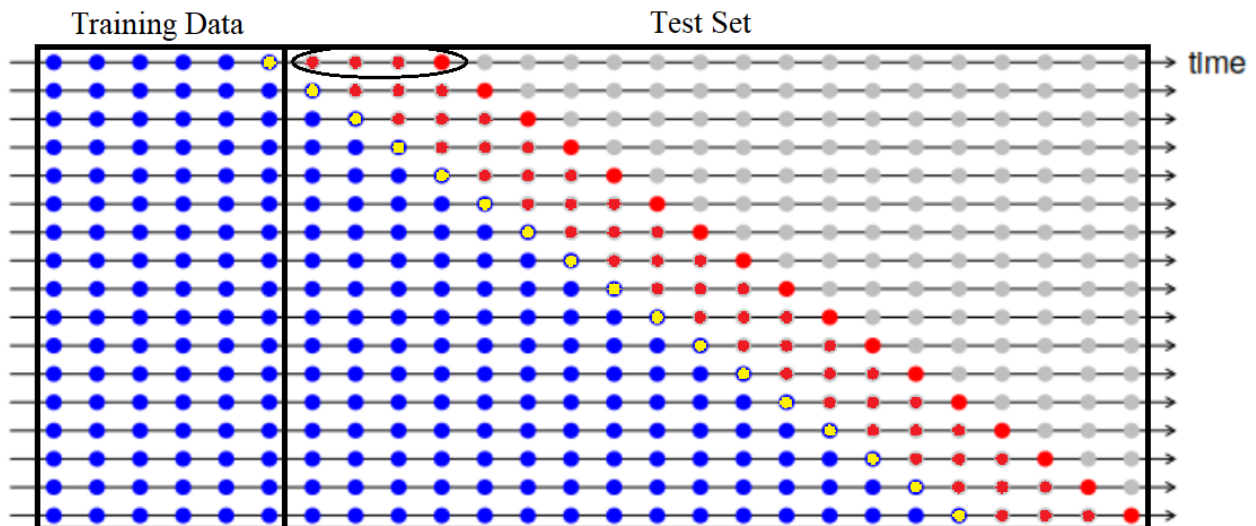
Για τους παραπάνω λόγους ο διαχωρισμός των δεδομένων σε σύνολο εκπαίδευσης και σύνολο ελέγχου συναντάται στις περισσότερες εφαρμογές πρόβλεψης δεδομένων. Η τελευταία χρονική στιγμή στο σύνολο εκπαίδευσης, δηλαδή το σημείο από όπου ξεκινάει η πρόβλεψη, ονομάζεται αρχή πρόβλεψης (forecasting origin). Ο αριθμός των χρονικών περιόδων μεταξύ της αρχής πρόβλεψης και του χρονικού σημείου μέχρις ότου γίνεται η πρόβλεψη ονομάζεται ορίζοντας πρόβλεψης (forecasting horizon). Ο μεγαλύτερος (σε χρονική απόσταση) ορίζοντας πρόβλεψης ονομάζεται πρόβλεψη N βημάτων μπροστά (N-step-ahead forecast).

Στην αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου χρησιμοποιώντας το σύνολο ελέγχου μπορεί να χρησιμοποιηθεί είτε μία αρχή πρόβλεψης είτε πολλές. Στην πρώτη περίπτωση, η οποία ονομάζεται αξιολόγηση σταθερής αρχής (fixed origin), κινούμαστε ως εξής: βρισκόμαστε στην χρονική στιγμή T, κάνουμε προβλέψεις για της χρονικές περιόδους T+1, T+2,...,T+N. Από αυτές τις τιμές εξάγεται οποιαδήποτε πληροφορία για το σφάλμα πρόβλεψης του μοντέλου. Αυτή η μέθοδος έχει αρκετά αρνητικά. Αρχικά εξάγεται μία πρόβλεψη για κάθε ορίζοντα πρόβλεψης άρα και ένα σφάλμα πρόβλεψης ανά ορίζοντα μόνο. Δεύτερον, οι προβλέψεις που γίνονται από ένα σταθερό σημείο είναι ευαίσθητες σε ιδιосуγκρασίες που σχετίζονται με την συγκεκριμένη αρχή πρόβλεψης. Τέλος, τα στατιστικά σύνολοι είναι μείγμα μακροπρόθεσμων και βραχυπρόθεσμων σφαλμάτων πρόβλεψης.

Όλα αυτά μπορούν να αποφευχθούν διαλέγοντας να ανανεώνουμε την αρχή πρόβλεψης κάθε φορά που γίνεται διαθέσιμη μία μέτρηση. Η μέθοδος αυτή ονομάζεται αξιολόγηση με κυλιόμενη αρχή. Αυτή είναι και η μέθοδος που επιλέχθηκε για την αξιολόγηση της προβλεπτικής ικανότητας των μοντέλων που αναπτύχθηκαν στην διπλωματική εργασία. Θα χρησιμοποιηθεί πιο συγκεκριμένα μια μικρή παραλλαγή αυτής, δανειζόμενοι κάποια στοιχεία από μία μέθοδο που χρησιμοποιείται στην μηχανική μάθηση και ονομάζεται cross validation για χρονοσειρές.

Η μέθοδος θα ονομαστεί αξιολόγηση με κυλιόμενη αρχή σταθερού ορίζοντα χωρίς επανεκπαίδευση. Χωρίς επανεκπαίδευση εννοούμε ότι ενώ μεταφέρεται κάθε φορά η αρχή πρόβλεψης κατά ένα χρονικό βήμα μπροστά, δεν εκπαιδεύεται ξανά το μοντέλο στα καινούρια διαθέσιμα δεδομένα, δηλαδή στο παλιό σύνολο δεδομένων επανυψημένο κατά μία ακόμα παρατήρηση. Παρόλα αυτά θα ανανεώνονται κάποιες πληροφορίες όπως εποχικότητα και τάση.

Ας υποθεθεί ότι μία δοσμένη χρονοσειρά έχει χωριστεί σε ένα σύνολο εκπαίδευσης  $TR = \{1, \dots, T\}$  και ένα σύνολο ελέγχου  $Te = \{T + 1, \dots, T + N\}$ . Έστω τώρα ότι έχει επιλεγεί να γίνει πρόβλεψη  $h$  βημάτων μπροστά, κοιτάζοντας  $h$  βήματα πίσω (AR( $p$ )). Η πρώτη out-of sample πρόβλεψη που θα γίνει είναι εκείνη με αρχή πρόβλεψης την χρονική στιγμή  $T$  και τα  $h$  βήματά της θα αφορούν στις χρονικές περιόδους  $\{T + 1, \dots, T + h\}$ . Ακολούθως η 2<sup>η</sup> πρόβλεψη έχει αρχή πρόβλεψης την χρονική στιγμή  $T+1$  προβλέποντας τις χρονικές περιόδους  $\{T + 2, \dots, T + 1 + h\}$  κοκ. μέχρι ο ορίζοντας  $h$  βημάτων μπροστά της τελευταίας πρόβλεψης να ισούται με  $T + N$ . Με αυτήν την μέθοδο επιτυγχάνεται ίδιος αριθμός προβλέψεων για κάθε ορίζοντα, ο οποίος ισούται με  $N - h + 1$ . Εφόσον πλέον έχουν εξασφαλιστεί πολλαπλές προβλέψεις για κάθε ορίζοντα άρα και πολλαπλά σφάλματα πρόβλεψης, τα αποτελέσματα για την ακρίβεια της μεθόδου ανά ορίζοντα είναι πιο ρωμαλέα. Στην Εικόνα 21 φαίνεται σχηματικά η μέθοδος αξιολόγησης με κυλιόμενη αρχή σταθερού ορίζοντα χωρίς επανεκπαίδευση.



**Εικόνα 21** Αξιολόγηση με κυλιόμενη αρχή σταθερού ορίζοντα χωρίς επανεκπαίδευση.

Ψάχνοντας στην βιβλιογραφία όσον αφορά τις προγραμματιστικές γλώσσες Python και R μόνο η βιβλιοθήκη `forecast` προσφέρει την δυνατότητα της ανανέωσης της αρχής πρόβλεψης χωρίς επανεκπαίδευση. Είναι σημαντικό να τονιστεί ότι πρόκειται για σημαντική διαδικασία διότι επιτρέπει την ουσιαστική σύγκριση μεταξύ μεθόδων της στατιστικής σχολής και της μηχανικής μάθησης. Στην βιβλιογραφία συχνά προτείνεται η μέθοδος `cross-validation` για χρονοσειρές κατά

την οποία ορίζεται ένα ελάχιστο μήκος του συνόλου εκπαίδευσης από τα δεδομένα στο οποίο κάθε φορά προστίθεται μία παρατήρηση. Μετά από κάθε τέτοια πρόσθεση επανεκπαιδεύεται το μοντέλο και γίνεται πρόβλεψη N βημάτων μπροστά. Αυτή η μέθοδος για μεθόδους βαθιάς μηχανικής μάθησης είναι δυσβάστακτη όσον αφορά τον υπολογιστικό φόρτο και γι' αυτό το λόγο αποφεύχθηκε.

## 5.4 ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ

### 5.4.1 Μετρικές Σφάλματος

Το σφάλμα της πρόβλεψης ορίζεται ως η διαφορά μεταξύ μίας παρατηρούμενης τιμής και της πρόβλεψης της. Με τον όρο σφάλμα δεν εννοείται κάποιου είδους λάθος αλλά το μέρος της παρατηρούμενης τιμής που δεν μπορεί να προβλεφθεί [57], και γράφεται ως,

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T} \quad (5.1)$$

όπου το σύνολο εκπαίδευσης δίνεται από  $\{y_1, \dots, y_T\}$  και το σύνολο ελέγχου από  $\{y_{T+1}, y_{T+2}, \dots\}$ . Να τονιστεί σε αυτό το σημείο ότι τα σφάλματα πρόβλεψης διαφέρουν από τα κατάλοιπα. Πρώτον, τα κατάλοιπα αφορούν το σύνολο εκπαίδευσης ενώ τα σφάλματα πρόβλεψης το σύνολο ελέγχου. Και δεύτερον, τα κατάλοιπα βασίζονται στις προβλέψεις ενός βήματος μπροστά (στα στατιστικά μοντέλα τουλάχιστον), ενώ τα σφάλματα πρόβλεψης μπορεί να αφορούν προβλέψεις πολλών βημάτων μπροστά.

Μπορούμε να μετρηθεί η ακρίβεια πρόβλεψης ενός μοντέλου συνοψίζοντας τα σφάλματα με διάφορους τρόπους χρησιμοποιώντας διάφορες μετρικές αξιολόγησης. Χρησιμοποιήθηκαν 5 διαφορετικές μετρικές αξιολόγησης οι οποίες είναι [57], [102]:

- Το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error-MAE) που ορίζεται ως:

$$MAE = \frac{\sum_{t=1}^N e_t}{N} \quad (5.2)$$

Είναι στην ίδια κλίμακα με τα δεδομένα και είναι διάσημη μετρική κυρίως επειδή υπολογίζεται εύκολα και είναι ερμηνεύσιμη. Μία μέθοδος πρόβλεψης που ελαχιστοποιεί το MAE οδηγεί σε προβλέψεις της διάμεσου.

- Το Ριζικό Μέσο Τετραγωνικό Σφάλμα (Root Mean Squared Error-RMSE) που ορίζεται ως:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N e_t^2}{N}} \quad (5.3)$$

Και αυτή η μετρική είναι στην ίδια κλίμακα με τα δεδομένα. Είναι δυσκολότερη στον υπολογισμό της και στην ερμηνεία της. Μία μέθοδος πρόβλεψης που ελαχιστοποιεί το RMSE οδηγεί σε προβλέψεις του μέσου.

- Το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error-MAPE) το οποίο ορίζεται ως:

$$MAPE = \frac{\sum_{t=1}^N |p_t|}{N} \quad (5.4)$$

όπου

$$p_t = \frac{100e_t}{y_t} \quad (5.5)$$

είναι το ποσοστιαίο σφάλμα. Τα ποσοστιαία σφάλματα έχουν το πλεονέκτημα ότι δεν έχουν μονάδες και εκφράζουν το σφάλμα υπό μορφή ποσοστού. Οι μετρικές που βασίζονται στα ποσοστιαία σφάλματα έχουν το μειονέκτημα ότι δεν ορίζονται για  $y_t = 0$ , και παίρνουν εξαιρετικά μεγάλες τιμές όταν το  $y_t$  είναι κοντά στο 0. Επίσης, δίνουν μεγαλύτερο βάρος στα αρνητικά σφάλματα από ότι στα θετικά σφάλματα.

- Το Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (symmetric Mean Absolute Percentage Error-sMAPE) το οποίο ορίζεται ως:

$$sMAPE = \frac{\sum_{t=1}^N \left( \frac{200|y_t - \hat{y}_t|}{y_t + \hat{y}_t} \right)}{N} \quad (5.6)$$

Αυτή η μορφή διορθώνει το πρόβλημα με τα αρνητικά σφάλμα όμως παραμένουν τα προβλήματα με την τιμή του  $y_t$ .

- Το Μέσο Απόλυτο Σφάλμα υπό Κλίμακα (Mean Absolute Scaled Error-MASE) το οποίο ορίζεται ως:

$$MASE = \text{mean}(|q_j|) \quad (5.7)$$

όπου για χρονοσειρές με εποχικότητα,

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{m+1}^T |y_t - y_{t-m}|} \quad (5.8)$$

Επειδή ο αριθμητής και ο παρονομαστής περιέχουν τιμές στην κλίμακα των αρχικών δεδομένων το  $q_j$  είναι ανεξάρτητο της κλίμακας των δεδομένων. Ένα σφάλμα υπό κλίμακα είναι μικρότερο του 1 αν προέρχεται από μία πρόβλεψη που είναι καλύτερη από την μέση seasonal naïve πρόβλεψη υπολογισμένη στα δεδομένα εκπαίδευσης. Και μεγαλύτερη από το 1 αν ισχύει το αντίστροφο.

#### 5.4.2 Υπολογιστική Πολυπλοκότητα

Η υπολογιστική πολυπλοκότητα (Computational Complexity-CC) χρησιμοποιείται για να καθοριστεί ο χρόνος που χρειάζεται για την εκπαίδευση ενός μοντέλου και για να γίνουν οι προβλέψεις που ζητούνται. Άρα το CC μπορεί να οριστεί ως ο μέσος απαιτούμενος υπολογιστικός χρόνος που χρειάζεται ένα μοντέλο για να προβλέψει μία χρονοσειρά, διαιρεμένος με τον αντίστοιχο χρόνο που χρειάζεται η seasonal naïve μέθοδος για να εκτελέσει την ίδια εργασία. Έτσι καταλήγουμε σε μία σχετική μετρική, η οποία υποδεικνύει τον επιπρόσθετο, αναλογικά χρόνο που απαιτείται για να γίνουν προβλέψεις χρησιμοποιώντας μία πιο πολύπλοκη μέθοδο. Το CC ορίζεται ως:

$$(CC) = \frac{\text{Υπολογιστικός Χρόνος Μοντέλου}}{\text{Υπολογιστικός Χρόνος Seasonal Naive μεθόδου}} \quad (5.9)$$

#### 5.4.3 Προσαρμογή Μοντέλου

Για να καθοριστεί το πόσο καλά το μοντέλο έχει προσαρμοστεί στα δεδομένα του συνόλου εκπαίδευσης ορίζεται το μέγεθος προσαρμογή μοντέλου (Model Fitting-MF) για προβλέψεις εντός του συνόλου εκπαίδευσης και ενός βήματος μπροστά, και δίνεται από την παρακάτω σχέση,

$$MF = \frac{n \sum_{t=1}^n (y_t - \hat{y}_t)^2}{(\sum_{t=1}^n y_t)^2} * 100\% \quad (5.10)$$

το οποίο είναι το μέσο τετραγωνικό σφάλμα κανονικοποιημένο από το τετράγωνο της μέσης τιμής της χρονοσειράς που εξετάζεται.

## 5.5 ΜΕΘΟΔΟΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑΣ ΓΙΑ ΤΙΣ ΜΕΘΟΔΟΥΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ

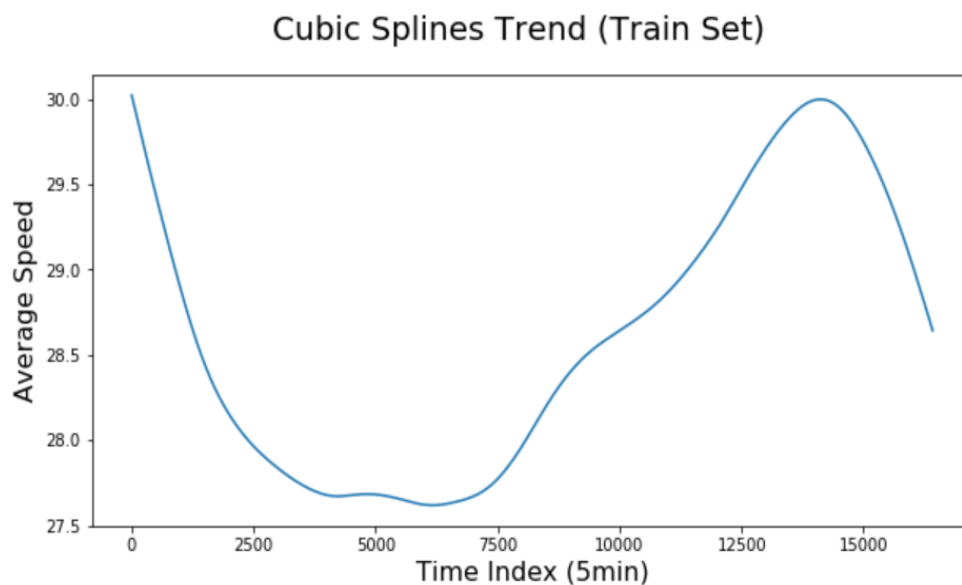
### ΜΑΘΗΣΗΣ

Στην βιβλιογραφία της μηχανικής μάθησης υπάρχουν αντικρουόμενες απόψεις με κάποιες μελέτες να υποστηρίζουν ότι οι μέθοδοι μηχανικής μάθησης μπορούν να εφαρμοστούν στα αρχικά δεδομένα μαθαίνοντας όποιο μοτίβο περιέχεται σε αυτά χωρίς κάποια επεξεργασία για την τάση ή την εποχικότητα [103], ενώ άλλες μελέτες έχουν καταλήξει στο αντίθετο αποτέλεσμα, όπου ισχυρίζονται ότι χωρίς κατάλληλη προεπεξεργασία ως προς τα προαναφερόμενα, οι μέθοδοι μηχανικής μάθησης μπορούν να γίνουν ασταθείς και να μην επιστρέψουν τα καλύτερα δυνατά αποτελέσματα [104]. Φαίνεται λογικό ότι αφαιρώντας την εποχικότητα ή την τάση, το μοντέλο μπορεί να επικεντρωθεί στο στοχαστικό κομμάτι των δεδομένων, και υπάρχει πιθανότητα να δίνει καλύτερα αποτελέσματα. Προς διερεύνηση αυτού για τις μεθόδους βαθιάς μηχανικής μάθησης για βραχυπρόθεσμη πρόβλεψη μονομεταβλητών κυκλοφοριακών χρονοσειρών, εξετάζονται οι παρακάτω περιπτώσεις υπό μορφή υπερπαραμέτρων:

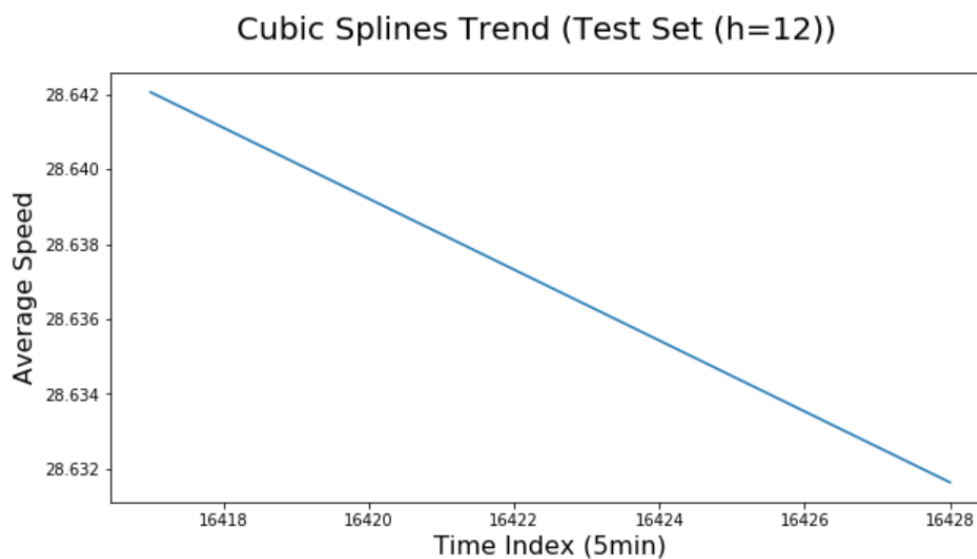
- **Αρχικά δεδομένα:** Καμία προεπεξεργασία στα δεδομένα.
- **Αποεποχικοποίηση των δεδομένων:** Υπολογίζεται η εποχική συνιστώσα των δεδομένων αφαιρείται από τα δεδομένα και η εκπαίδευση των μοντέλων γίνεται πάνω στα εποχιακά προσαρμοσμένα δεδομένα. Τέλος, για την πρόβλεψη εποχικής συνιστώσας χρησιμοποιείται η *seasonal naïve* μέθοδος, οι προβλέψεις της οποίας προστίθενται στις προβλέψεις του μοντέλου για τα εποχιακά προσαρμένα δεδομένα.
- **Αφαίρεση της τάσης από τα δεδομένα:** Υπολογίζεται η συνιστώσα της τάσης χρησιμοποιώντας την μέθοδο των κυβικών σπλινών, η οποία ακολούθως αφαιρείται από τα δεδομένα. Η εκπαίδευση του μοντέλου πραγματοποιείται στα δεδομένα χωρίς τάση. Οι τελικές προβλέψεις προκύπτουν προσθέτοντας την προβλεπόμενη τάση στις προβλέψεις που επιστρέφει το μοντέλο. Στην Εικόνα 22 φαίνεται ο υπολογισμός της συνιστώσας της τάσης με χρήση κυβικών *splines* για το σύνολο δεδομένων, ενώ στην Εικόνα 23 φαίνεται η πρόβλεψη για ορίζοντα πρόβλεψης 12 βημάτων στο σύνολο ελέγχου.
- **Μετασχηματισμός δεδομένων:** Μετατροπή δεδομένων σε κλίμακα  $[0,1]$  (*scaling*) με χρήση *min – max* μεθόδου μετατροπής κλίμακας.
- **Συνδυασμός των 3 πιο πάνω εναλλακτικών.**



Σημαντικό είναι να αναφερθεί ότι για τον υπολογισμό των παραπάνω ακολουθήθηκε κυλιόμενη λογική. Δηλαδή ακολουθώντας την λογική του 5.3 κάθε φορά που κινείται η αρχή (origin) της πρόβλεψης, υπολογίζεται ξανά η εποχικότητα και η τάση, καθώς πρόκειται για εξαιρετικά «φθηνούς» υπολογισμούς.



**Εικόνα 22** Υπολογισμός τάσης των δεδομένων εκπαίδευσης με χρήση κυβικών σπλινών.



**Εικόνα 23** Πρόβλεψη τάσης για ορίζοντα 12 βημάτων στο σύνολο ελέγχου.

## 5.6 ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΤΩΝ ΥΠΕΡΠΑΡΑΜΕΤΡΩΝ ΚΑΙ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Στο κεφάλαιο 4 τονίστηκε ότι όσον αφορά τις στατιστικές μεθόδους πρόβλεψης η διαδικασία που ακολουθήθηκε ήταν πλήρως αυτοματοποιημένη μέσω της βιβλιοθήκης `forecast`. Δεν συμβαίνει το ίδιο με τις μεθόδους βαθιάς μηχανικής μάθησης (ομάδα 3) όπου θα πρέπει να γίνει επιλογή των υπερπαραμέτρων προς αναζήτηση του καταλληλότερου μοντέλου. Ακολουθήθηκε εξαντλητική αναζήτηση τύπου πλέγματος (`grid search`) στον χώρο των υπερπαραμέτρων που ορίστηκε για κάθε μέθοδο. Αρχικά, έχουν ήδη εισαχθεί κάποιες υπερπαραμέτροι οι οποίες είναι κοινές για όλες τις μεθόδους και αφορούν στην τάξη  $p$  της AR διαδικασίας, δηλαδή πόσο πίσω στο χρόνο θα επιτραπεί στο μοντέλο να «κοιτάει», σε εκείνες που αναφέρονται στην παράγραφο 5.5, τέλος στην εισαγωγή μεθόδων κανονικοποίησης στο μοντέλο προς αποφυγή της υπερεκπαίδευσης. Αν και οι μέθοδοι κανονικοποίησης μπορεί να είναι διαφορετικές μεταξύ μεθόδων, αναφέρεται στις κοινές καθώς για όλες τις μεθόδους τέθηκε να είναι `true` ή `false`, εφαρμόζοντας καθολικά κάποιους συνδυασμούς τεχνικών. Επίσης για όλες τις μεθόδους ο αριθμός των εποχών ορίστηκε να είναι 100, δηλαδή ο αλγόριθμος βλέπει το πολύ 100 φορές τα δεδομένα πριν τερματιστεί η εκπαίδευσή του. Ως αλγόριθμος βελτιστοποίησης ορίστηκε να είναι ο `adam` [108], ο οποίος γενικά θεωρείται η καλύτερη προκαθορισμένη επιλογή και πρόκειται για αυτοματοποιημένη μέθοδο χωρίς να χρειάζεται ΡΥΘΜΙΣΗ κάποια από τις παραμέτρους της μεθόδου. Οι τιμές για αυτές τις υπερπαραμέτρους ορίστηκαν ως εξής:

- Παράθυρο: [8, 10, 12]
- Αφαίρεση εποχικότητας: [True, False]
- Αφαίρεση τάσης: [True, False]
- Μετατροπή υπό κλίμακα: [True, False]
- Κανονικοποίηση: [True, False]
- Εποχές: [100]
- Αλγόριθμος βελτιστοποίησης: ["adam"]

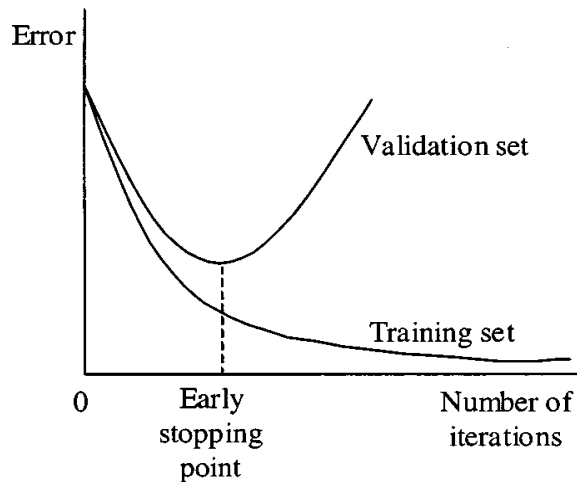
Ήδη από τις παραπάνω τιμές που ορίστηκαν για τις κοινές υπερπαραμέτρους έχουν δημιουργηθεί 48 διαφορετικοί συνδυασμοί υπερπαραμέτρων.

### 5.6.1 Κανονικοποίηση

Πρόκειται για μία τεχνική που έχει ως στόχο την βελτίωση της ικανότητας γενίκευσης ενός μοντέλου αποφεύγοντας την υπερεκπαίδευση του. Η υπερεκπαίδευση συμβαίνει όταν το αποτέλεσμα της εκπαίδευσης αντιστοιχεί πολύ κοντά ή ακριβώς στο σύνολο στο οποίο εκπαιδεύτηκε ένα μοντέλο, έχοντας «μάθει» τις ιδιοσυγκρασίες του συνόλου εκπαίδευσης. Αυτό οδηγεί ένα μοντέλο στο να μην μπορεί να γενικεύσει, π.χ. για πρόβλημα πρόβλεψης να δίνει κακές προβλέψεις σε δεδομένα που δεν έχει «δει» κατά την εκπαίδευση (σύνολο ελέγχου). Αυτό το φαινόμενο είναι μη επιθυμητό καθώς μπορεί να οδηγήσει σε πολύ μικρές τιμές ακρίβειας. Έτσι ενώ η προσαρμογή του μοντέλου μπορεί να είναι πολύ καλή, το εμπειρικό σφάλμα είναι μεγάλο. Αν οι μετρήσεις έχουν θόρυβο, τότε το φαινόμενο της υπερεκπαίδευσης είναι ακόμα πιο έντονο. Η κανονικοποίηση ουσιαστικά εισάγει μία ποινή, υπό διάφορες μορφές, έτσι ώστε να διερευνώνται συγκεκριμένες περιοχές του χώρου της συνάρτησης που χρησιμοποιήθηκε για την κατασκευή του μοντέλου, πράγμα που μπορεί να οδηγήσει σε βελτίωση της ικανότητας γενίκευσης του.

### 5.6.2 Πρόωρο Σταμάτημα

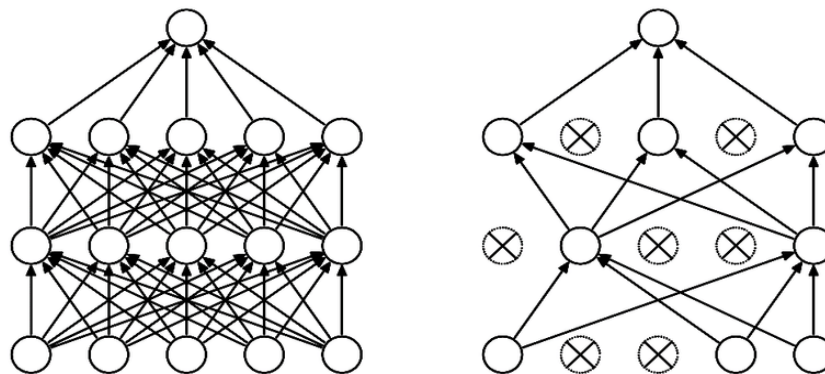
Το πρόωρο σταμάτημα (early stopping) [109] είναι μία τεχνική κανονικοποίησης στον χρόνο. Διασθητικά, μία διαδικασία εκπαίδευσης όπως η κατάβαση κλίσης θα τείνει να μαθαίνει όλο και περισσότερο πολύπλοκες συναρτήσεις καθώς ο αριθμός των επαναλήψεων αυξάνεται. Με το να επιβάλλεται κανονικοποίηση στον χρόνο, η πολυπλοκότητα του μοντέλου μπορεί να ελεγχθεί, βελτιώνοντας έτσι την ικανότητα γενίκευσης. Στην πράξη, η μέθοδος του πρόωρου σταματήματος εφαρμόζεται κατά την διάρκεια της εκπαίδευσης πάνω σε ένα σύνολο εκπαίδευσης και μετρώντας της ακρίβεια σε ένα στατιστικά ανεξάρτητο σετ αξιολόγησης. Το εκπαιδευμένο μοντέλο έπειτα εξετάζεται στο σύνολο ελέγχου. Στην Εικόνα 24 φαίνεται σχηματικά πως περίπου λειτουργεί η μέθοδος. Στον άξονα των  $x$  απεικονίζεται ο αριθμός των εποχών ενώ στον άξονα τον  $y$  το σφάλμα. Καθώς προχωράει η εκπαίδευση του μοντέλου σε κάθε εποχή, εκτός από το σφάλμα στο σύνολο εκπαίδευσης υπολογίζεται το σφάλμα στο σύνολο αξιολόγησης. Στο σημείο εκείνο όπου συνεχίζει η πτώση του σφάλματος εκπαίδευσης αλλά αρχίζει η αύξηση του σφάλματος αξιολόγησης, σταματάει η εκπαίδευση του μοντέλου. Ουσιαστικά εντοπίζουμε μέσω αυτής της διαδικασίας το σημείο στο οποίο ξεκινάει η υπερεκπαίδευση του μοντέλου. Ένα αρνητικό αυτής της μεθόδου, είναι ότι μειώνονται τα δεδομένα του αρχικού συνόλου εκπαίδευσης, καθώς διασπάται περαιτέρω σε σύνολο εκπαίδευσης και αξιολόγησης, για να μπορεί να υλοποιηθεί η μέθοδος.



**Εικόνα 24** Η μέθοδος του πρόωρου σταματήματος.

### 5.6.3 Dropout

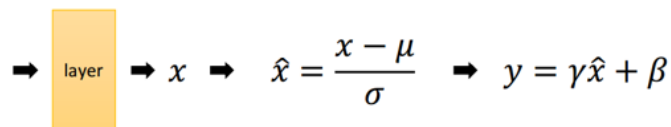
Η μέθοδος dropout [110] είναι μία μέθοδος κανονικοποίησης που αποσκοπεί στην μείωση του φαινομένου της υπερεκπαίδευσης με το να αποτρέπει πολύπλοκες προσαρμογές στα δεδομένα. Οι μέθοδοι ensemble για νευρωνικά δίκτυα με διαφορετικές παραμετροποιήσεις ανά μοντέλο είναι γνωστό ότι μειώνουν την υπερεκπαίδευση, με αντάλλαγμα όμως την αυξημένη υπολογιστική προσπάθεια και εκπαίδευση πολλών μοντέλων. Παρόλα αυτά ένα μόνο μοντέλο μπορεί να προσομοιωθεί σαν να αποτελείται από ένα μεγάλο αριθμό διαφορετικών αρχιτεκτονικών δικτύων «πετώντας» εκτός εκπαίδευσης τυχαία ένα ποσοστό κόμβων κατά την διάρκεια της εκπαίδευσης του μοντέλου. Αυτή η διαδικασία είναι υπολογιστικά «φθηνή» και εξαιρετικά αποτελεσματική σαν μέθοδος κανονικοποίησης για μείωση της υπερεκπαίδευσης και βελτιώνει το σφάλμα γενίκευσης σε δίκτυα βαθιάς αρχιτεκτονικής. Στην Εικόνα 25 φαίνεται πως λειτουργεί η μέθοδος dropout.



**Εικόνα 25** Νευρωνικό δίκτυο πριν (αριστερά) και μετά την εφαρμογή του dropout (δεξιά).

#### 5.6.4 Κανονικοποίηση Παρτίδων

Η κανονικοποίηση των παρτίδων (Batch normalization) [111] είναι μία τεχνική για την βελτίωση της ταχύτητας, της επίδοσης και της ευαισθησίας των τεχνητών νευρωνικών δικτύων. Χρησιμοποιείται για την κανονικοποίηση ενός επιπέδου με την προσαρμογή και τον μετασχηματισμό υπό κλίμακα των ενεργοποιήσεων. Η μέθοδος αρχικά προτάθηκε για την επίλυση του προβλήματος της αλλαγής της εσωτερικής μεταβλητότητας. Κατά την διάρκεια της φάσης εκπαίδευσης ενός δικτύου, καθώς οι παράμετροι των προηγούμενων επιπέδων αλλάζουν, η κατανομή των εισόδων στα τρέχοντα επίπεδα αλλάζει αντίστοιχα, με τέτοιο τρόπο ώστε το τρέχον επίπεδο χρειάζεται στιγμιαία να αναπροσαρμοστεί σε καινούριες κατανομές. Αυτό το πρόβλημα είναι αρκετά έντονο στις βαθιές αρχιτεκτονικές καθώς μικρές αλλαγές στα πιο ρηχά κρυφά επίπεδα θα αυξάνουν καθώς προχωράνε εντός του δικτύου, με αποτέλεσμα να καταλήγουν σε μεγάλη αλλαγή της κατανομής στα βαθύτερα δίκτυα. Γι' αυτό το λόγο, η μέθοδος αυτή προτείνεται για να μειωθούν πιθανών μη ζητούμενες αλλαγές και έχει ως αποτέλεσμα την επιτάχυνση της διαδικασίας εκπαίδευσης και την παραγωγή πιο αξιόπιστων μοντέλων. Επιπροσθέτως, κάποια από τα θετικά που επιφέρει η μέθοδος είναι ότι μπορούν να χρησιμοποιηθούν μεγαλύτερες τιμές για το ρυθμό μάθησης χωρίς να δημιουργούνται προβλήματα εξαφανιζόμενων ή εκρηγνυόμενων κλίσεων, κανονικοποιεί το δίκτυο κατά τρόπο που αυξάνει την ικανότητα γενίκευσης, δεν χρειάζεται η χρήση του dropout για την μείωση της υπερεκπαίδευσης, και τέλος το δίκτυο επίσης γίνεται πιο ρωμαλέο στις διαφορετικές αρχικοποιήσεις των βαρών και στις διαφορετικές τιμές του ρυθμού εκμάθησης. Στην Εικόνα 26 φαίνεται πως λειτουργεί η μέθοδος της κανονικοποίησης των παρτίδων σχηματικά.



- $\mu$ : mean of  $x$  in mini-batch
- $\sigma$ : std of  $x$  in mini-batch
- $\gamma$ : scale
- $\beta$ : shift
- $\mu, \sigma$ : functions of  $x$ , analogous to responses
- $\gamma, \beta$ : parameters to be learned, analogous to weights

**Εικόνα 26** Το επίπεδο batch normalization σχηματικά.

### 5.6.5 L2 ποινικοποίηση

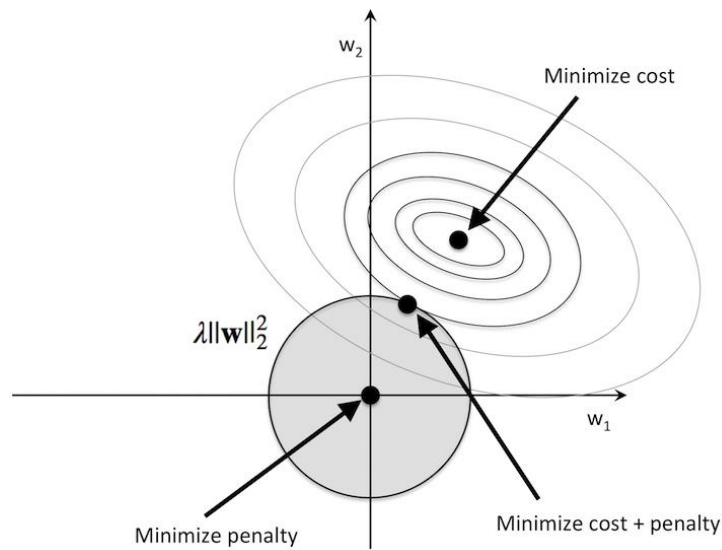
Κατά την εκμάθηση μίας γραμμικής συνάρτησης  $f$ , η οποία παραμετροποιείται από ένα άγνωστο διάνυσμα  $w$  τέτοιο ώστε  $f(x) = w \cdot x$ , μπορεί να προστεθεί η νόρμα L2 του διανύσματος  $w$  στην εξίσωση κόστους επιβάλλοντας μικρότερες νόρμες. Αυτό ονομάζεται L2 ποινικοποίηση [112] και πρόκειται από τις πιο γνωστές μεθόδους κανονικοποίησης, η οποία γράφεται ως,

$$\min_w \sum_{i=1}^n V(\hat{x}_i \cdot w, \hat{y}_i) + \lambda \|w\|_2^2 \quad (5.11)$$

Στην περίπτωση μίας γενικής συνάρτησης, αντικαθιστούμε με την νόρμα της συνάρτησης στον αναπαραγόμενο πυρήνα του Hilbert χώρου,

$$\min_f \sum_{i=1}^n V(f(\hat{x}_i), \hat{y}_i) + \lambda \|f\|_{\mathcal{H}}^2 \quad (5.12)$$

Αφού η L2 νόρμα είναι παντού διαφορίσιμη, προβλήματα μάθησης που χρησιμοποιούν τέτοιου τύπου κανονικοποίηση μπορούν να λυθούν με μεθόδους κατάβασης κλίσης. Στην Εικόνα 27 φαίνεται η υπό συζήτηση μέθοδος.



**Εικόνα 27** Η μέθοδος της L2 ποινικοποίησης.

Στο Πίνακα 2 φαίνονται οι παραμετροποιήσεις που επιλέχθηκαν για τις μεθόδους βαθιά μηχανικής μάθησης καθώς και την δομή κανονικοποίησης που επιβλήθηκε στην δομή του κάθε μοντέλου.

Πίνακας 2 Παραμετροποιήσεις και δομή κανονικοποίησης για τις μεθόδους βαθιάς μηχανικής μάθησης.

Μέθοδος	Υπερπαράμετροι										Δομή Κανονικοποίησης				Αριθμός Παραμετροποιήσεων	
	Κρυφά Επίπεδα	Αριθμός Συνελκτικών ή Επίπεδων	Αριθμός Νευρώνων /επίπεδο	Τύπος Κελιού LSTM /επίπεδο	GRU	Αριθμός Φίλτρων	Μέγεθος Μίσκας	Τύπος Padding	Ρυθμός Διαστολής	Μέγεθος Batch	Πρόωρο Σταμάτημα	Επίπεδο Dropout	Ποσοστό Dropout	L2 ποινή		Batch Normalization
MLP	[1,2,3]	-	[100,200]	-	-	-	-	-	-	[32,64]	Ναι	1 μετά από κάθε κρυφό επίπεδο	0.1	1ο κρυφό επίπεδο	-	576
LSTM	[1,2]	-	-	[100,200]	-	-	-	-	-	[64,128]	Ναι	1 μετά από κάθε κρυφό επίπεδο	0.1	προτελευταίο πλήρως συνδεδεμένο επίπεδο	-	384
GRU	[1,2]	-	-	-	[100,200]	-	-	-	-	[64,128]	Ναι	1 μετά από κάθε κρυφό επίπεδο	0.1	προτελευταίο πλήρως συνδεδεμένο επίπεδο	-	384
CNN	-	[2] σε 1 μπλοκ	-	-	-	[16,32,64]	[2,3]	"causal"	-	[64,128]	Ναι	προτελευταίο πλήρως συνδεδεμένο επίπεδο	0.1	-	Μετά το συνελκτικό μπλοκ	576
Dilated CNN	-	[3] σε 2 μπλοκ	-	-	-	[32,64]	[2,3]	"causal"	1,2,4	[64,128]	Ναι	προτελευταίο πλήρως συνδεδεμένο επίπεδο	0.1	-	Μετά από κάθε συνελκτικό μπλοκ	384
CNN-LSTM	[1]	[2] σε 1 μπλοκ	-	[100]	-	[32,64]	[2,3]	"causal"	-	[64,128]	Ναι	τελευταίο κρυφό + προτελευταίο πλήρως συνδεδεμένο επίπεδο	0.1	-	Μετά από κάθε συνελκτικό μπλοκ	384
Conv-LSTM	[1]	-	-	[100]	-	[32,64]	[2,3]	-	-	[64,128]	Ναι	τελευταίο κρυφό + προτελευταίο πλήρως συνδεδεμένο επίπεδο	0.1	-	-	384
Encoder-Decoder LSTM	[1] Encoder [1] Decoder	-	-	[100,200]	-	-	-	-	-	[64,128]	Ναι	Encoder + Decoder	0.1	προτελευταίο πλήρως συνδεδεμένο επίπεδο	-	192
Encoder-Decoder CNN-LSTM	[1] Decoder	[2] σε 1 μπλοκ στον Encoder	-	[100]	-	[32,64]	[2,3]	"causal"	-	[64,128]	Ναι	Decoder	0.1	προτελευταίο πλήρως συνδεδεμένο επίπεδο	Encoder	384
Encoder-Decoder Conv-LSTM	[1] Encoder [1] Decoder	-	-	[100]	-	[32,64]	[2,3]	-	-	[64,128]	Ναι	Encoder + Decoder	0.1	προτελευταίο πλήρως συνδεδεμένο επίπεδο	-	384

Οι παραπάνω τιμές των παραμέτρων ορίζουν για κάθε μοντέλο τον χώρο αναζήτησης για την επιλογή του καλύτερου μοντέλου. Το τελικό σύνολο δεδομένων που προέκυψε από το κεφάλαιο 3 έχει χωριστοί σε ένα σύνολο εκπαίδευσης και σε ένα σύνολο ελέγχου. Για την εκπαίδευση των μεθόδων βαθιάς μηχανικής μάθησης θα χρειαστεί περαιτέρω διαχωρισμός του συνόλου εκπαίδευσης, διατηρώντας και πάλι την χρονική σειρά των δεδομένων. Το σύνολο εκπαίδευσης θα χωριστεί περαιτέρω σε ένα προσωρινό σύνολο εκπαίδευσης, σε ένα σύνολο αξιολόγησης και σε ένα προσωρινό σύνολο ελέγχου. Αυτό γίνεται διότι το σύνολο αξιολόγησης θα χρησιμοποιηθεί στην διαδικασία του πρόωρου σταματήματος, και για να μην εισαχθεί περαιτέρω look-ahead μεροληψία στο μοντέλο, η προβλεπτική ικανότητα κάθε παραμετροποίησης θα αξιολογηθεί στο προσωρινό σύνολο ελέγχου που θα ακολουθεί χρονικά του συνόλου αξιολόγησης. Το προσωρινό σύνολο εκπαίδευσης θα αποτελείται από τις 14112 πρώτες παρατηρήσεις χρονικά, οι επόμενες 1152 παρατηρήσεις ανήκουν στο σύνολο αξιολόγησης, και τέλος οι τελευταίες 1152 παρατηρήσεις στο «προσωρινό» σύνολο ελέγχου. Η μέθοδος αξιολόγησης των μοντέλων (ανά παραμετροποίηση) είναι η ίδια που περιγράφηκε στην παράγραφο 5.3, αλλά η επιλογή της καλύτερης παραμετροποίησης ανά μέθοδο έγινε με βάση το μέσο ριζικό του μέσου τετραγωνικού σφάλματος (Average RMSE) για όλους του ορίζοντες πρόβλεψης (1-12 βήματα), δηλαδή

$$\text{Av. RMSE} = \frac{\sum_{h=1}^{12} \text{RMSE}_h}{12} \quad (5.13)$$

Αφού επιλεγεί η καλύτερη παραμετροποίηση ανά μέθοδο εκπαιδεύεται πλέον το μοντέλο σε όλο το σύνολο εκπαίδευσης. Εδώ υπάρχει ο κίνδυνος για υπερεκπαίδευση οπότε για κάθε παραμετροποίηση διατηρείται η εποχή στην οποία σταμάτησε η εκπαίδευση. Αν για όλες τις μεθόδους τέθηκε ο μέγιστος αριθμός εποχών να είναι 100 μπορεί σε κάποια παραμετροποίηση η εκπαίδευση να σταματάει νωρίτερα λόγω του πρόωρου σταματήματος. Αυτό είναι και το καλύτερο που μπορεί να επιτευχθεί καθώς είναι σημαντικό το τελικό μοντέλο να εκπαιδευτεί μέχρι και την τελευταία χρονική στιγμή (πρώτη αρχή πρόβλεψης), πριν την έναρξη του συνόλου ελέγχου.

Μετά την διαδικασία της βελτιστοποίησης των υπερπαραμέτρων ανά μέθοδο και της επιλογής του καλύτερου μοντέλου ανά μέθοδο, η εκπαίδευση στα δεδομένα εκπαίδευσης επαναλήφθηκε 20 φορές λόγω της εξάρτησης των μεθόδων από την τυχαία αρχικοποίηση των βαρών του δικτύου.



## 6 ΑΠΟΤΕΛΕΣΜΑΤΑ

---

Στο κεφάλαιο 6 θα σχολιαστούν τα αποτελέσματα ανά ομάδα και ανά μέθοδο και έπειτα θα παρουσιαστούν τα αποτελέσματα της συγκριτικής ανάλυσης μεταξύ όλων των μεθόδων.

### 6.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΕΘΟΔΟΥ ΑΝΑΦΟΡΑΣ

Στην ομάδα 1 ανήκει η μέθοδος Seasonal Naive η οποία αποτελεί και την μέθοδο αναφοράς όπου κάθε μέθοδος πρέπει να έχει καλύτερα αποτελέσματα από εκείνη για να γίνεται αποδεκτή για περαιτέρω σύγκριση. Στον Πίνακα 3 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 4 δίνονται κάποια συνοπτικά αποτελέσματα τα οποία αφορούν το μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), το μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), το μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και την μέσο σφάλμα (ανά μετρική). Κάθε μία από αυτές ορίζεται ως εξής:

- Μέσο Βραχυπρόθεσμο Σφάλμα:

$$\frac{\sum_{h=1}^4(\text{error})_h}{4} \quad (6.1)$$

- Μέσο Μεσοπρόθεσμο Σφάλμα:

$$\frac{\sum_{h=5}^8(\text{error})_h}{4} \quad (6.2)$$

- Μέσο Μακροπρόθεσμο Σφάλμα:

$$\frac{\sum_{h=9}^{12}(\text{error})_h}{4} \quad (6.3)$$

- Μέσο σφάλμα (όλοι οι ορίζοντες):

$$\frac{\sum_{h=1}^{12}(\text{error})_h}{12} \quad (6.4)$$

Οι παραπάνω υπολογισμοί γίνονται για όλες τις μεθόδους (ομάδα 1, 2 και 3) και θα αναφέρονται ως συνοπτικά μέτρα σφάλματος.

**Πίνακας 3** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Seasonal Naive.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	4.792	4.776	4.771	4.781	4.785	4.785	4.787	4.788	4.787	4.778	4.774	4.780
<b>RMSE</b>	6.571	6.537	6.535	6.544	6.546	6.546	6.546	6.547	6.547	6.539	6.537	6.540
<b>MAPE</b>	18.941	18.908	18.895	18.921	18.937	18.938	18.941	18.947	18.944	18.917	18.909	18.921
<b>sMAPE</b>	16.908	16.859	16.847	16.878	16.892	16.893	16.897	16.903	16.900	16.877	16.867	16.882
<b>MASE</b>	0.987	0.931	0.927	0.923	0.921	0.913	0.908	0.899	0.888	0.874	0.862	0.856

**Πίνακας 4** Συνοπτικά μέτρα σφάλματος για την μέθοδο Seasonal Naive.

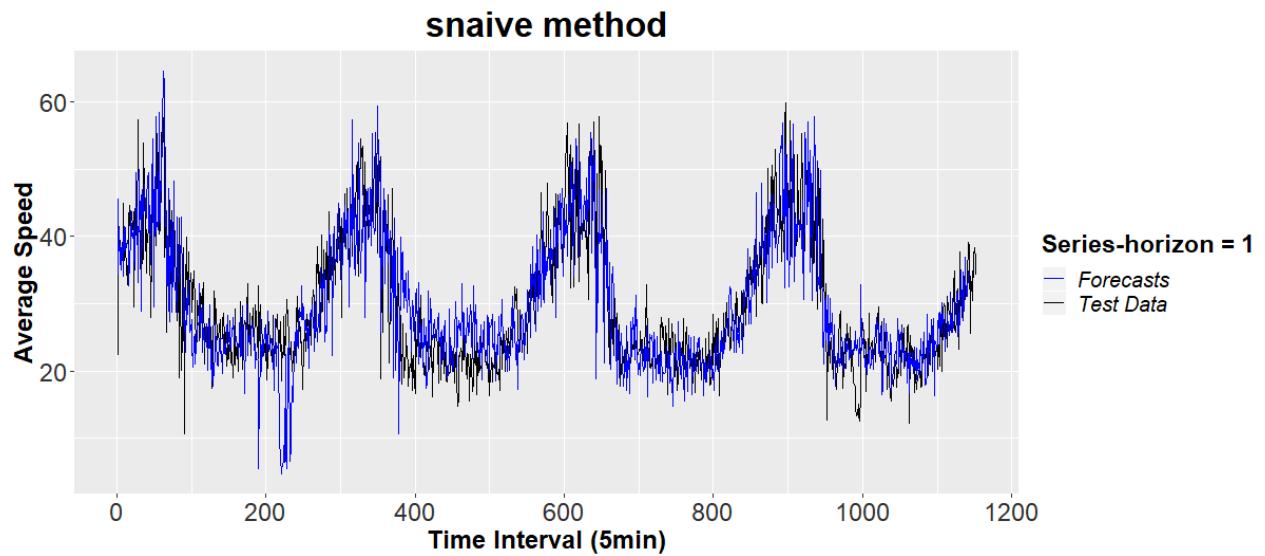
	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	4.780	4.786	4.780	4.782
<b>RMSE</b>	6.547	6.546	6.541	6.545
<b>MAPE</b>	18.916	18.941	18.923	18.927
<b>sMAPE</b>	16.873	16.896	16.881	16.884
<b>MASE</b>	0.942	0.910	0.870	0.907

**Πίνακας 5** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Seasonal Naive.

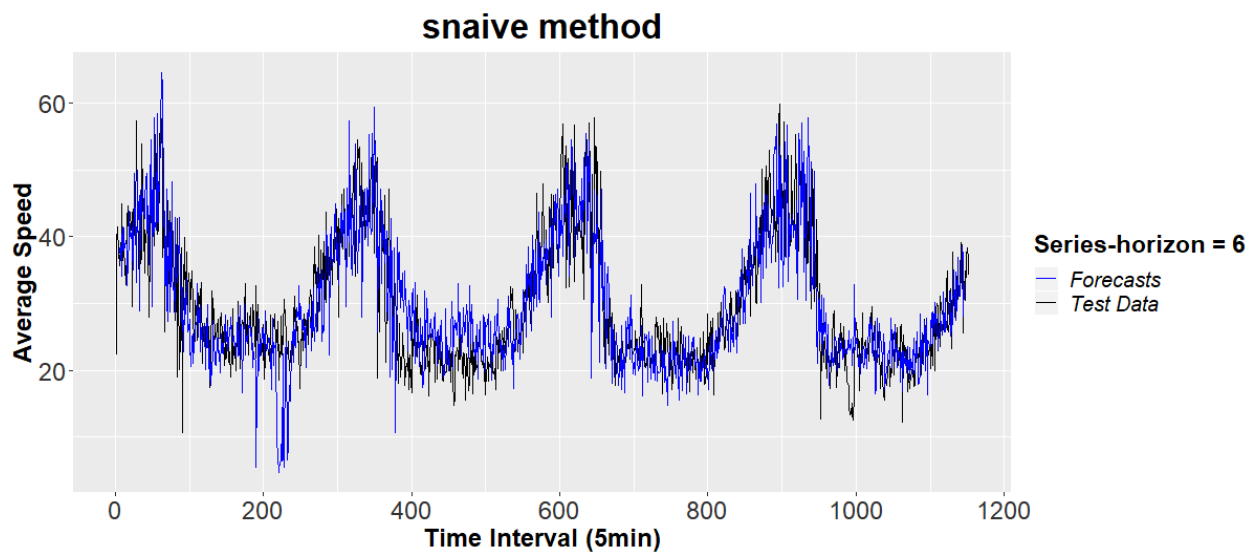
<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
0.007 sec	1.216 sec	1.223 sec	1	5.323

Στον Πίνακα 5 φαίνονται οι μετρικές χρόνου οι οποίες για τις μεθόδους της ομάδας 1 και 2 είναι ο χρόνος εκπαίδευσης (Train Time), ο χρόνος πρόβλεψης (Prediction Time) και ο συνολικός χρόνος (Total Time). Επιπροσθέτως, ο Πίνακας 5 περιλαμβάνει ακόμα 2 μετρικές οι οποίες ορίστηκαν στην παράγραφο 5.4, την υπολογιστική πολυπλοκότητα (CC) και την προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης.

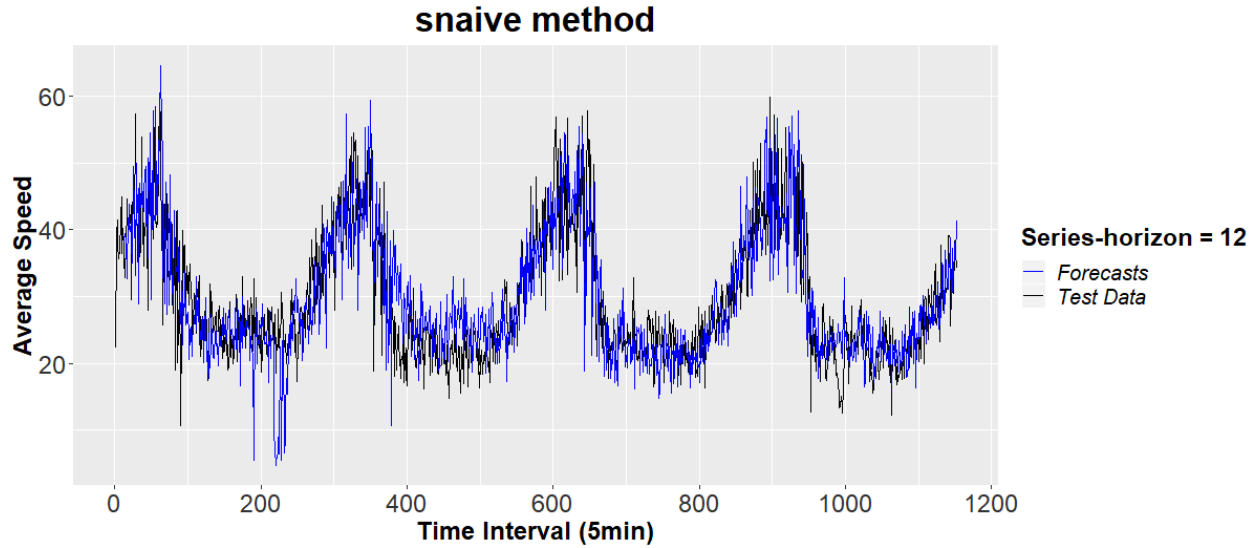
Στα Διαγράμματα 55, 56 και 57 φαίνονται ενδεικτικά τις προβλέψεις της μεθόδου Seasonal Naive για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει «πιάσει» την εποχικότητα των δεδομένων, επαναλαμβάνοντας όμως κάποιες ιδιοσυγκρασίες του παρελθόντος οι οποίες εμφανώς δεν επαναλαμβάνονται επακριβώς στο μέλλον, όπως κάποιες απότομες πτώσεις της ταχύτητας την 1<sup>η</sup> ημέρα της πρόβλεψης.



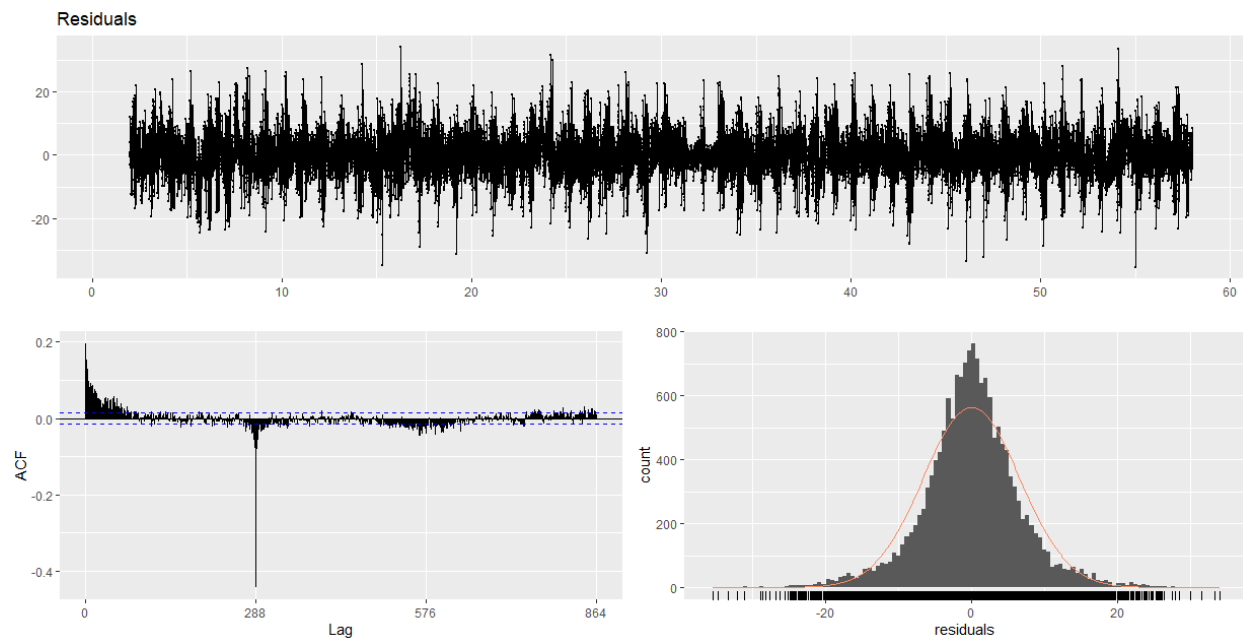
**Διάγραμμα 55** Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 56** Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 57** Προβλέψεις της μεθόδου Seasonal Naive για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



**Διάγραμμα 58** Γραφικός έλεγχος των καταλοίπων της μεθόδου Seasonal Naive.

Παρατηρείται στο Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 58) των καταλοίπων ότι έχει μείνει σημαντική πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος Seasonal Naive.

## 6.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ

Υπενθυμίζουμε ότι στις στατιστικές μεθόδους (ομάδα 2) ανήκουν οι ακόλουθες 5 μέθοδοι:

- Μέθοδος Arima + Εποχική αποσύνθεση (STL+ARIMA)
- Δυναμική Αρμονική Παλινδρόμηση με σφάλματα ARIMA.
- Μέθοδος ETS + Εποχική αποσύνθεση (STL+ETS)
- Μέθοδος TBATS (Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components)
- Μέθοδος Theta

### 6.2.1 Μέθοδος ARIMA με εποχική αποσύνθεση (STL+ARIMA)

Το μοντέλο που επιστράφηκε για τα εποχικά προσαρμοσμένα δεδομένα είναι ένα ARIMA(4,1,1). Αυτό μεταφράζεται σε ένα ARIMA μοντέλο το οποίο έχει 4 AR στοιχεία, 1 MA στοιχείο. Τα δεδομένα έχουν τάση σύμφωνα με τον KPSS έλεγχο οπότε το μοντέλο προσαρμόστηκε στα δεδομένα στα οποία έχει εφαρμοσθεί διαφορίση 1<sup>ης</sup> τάξης. Προφανώς δεν υπάρχουν εποχικοί όροι, καθώς έχει αφαιρεθεί η εποχικότητα από τα δεδομένα. Επίσης, επιλέχθηκε από το μοντέλο μετασχηματισμός Box-Cox με  $\lambda = 1.16833$ . Οι συντελεστές που υπολογίστηκαν και επιστράφηκαν από την κονσόλα της R για το μοντέλο ARIMA(4,1,1) είναι:

```

Coefficients:
      ar1      ar2      ar3      ar4      ma1
  0.1159  0.0800  0.0377  0.0477 -0.9694
s.e.  0.0086  0.0085  0.0085  0.0085  0.0036

```

**Πίνακας 6** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο STL+ARIMA.

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
<b>MAE</b>	3.386	3.421	3.451	3.485	3.515	3.531	3.545	3.563	3.573	3.566	3.574	3.582
<b>RMSE</b>	4.568	4.604	4.630	4.670	4.696	4.716	4.734	4.748	4.765	4.753	4.761	4.773
<b>MAPE</b>	11.258	11.419	11.563	11.690	11.811	11.876	11.924	11.986	12.028	12.019	12.052	12.084
<b>sMAPE</b>	11.371	11.520	11.662	11.796	11.919	11.975	12.022	12.082	12.120	12.108	12.142	12.172
<b>MASE</b>	0.697	0.667	0.671	0.673	0.676	0.674	0.673	0.669	0.663	0.652	0.646	0.641

Στον Πίνακα 6 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 7 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 7** Συνοπτικά μέτρα σφάλματος για την μέθοδο STL+ARIMA.

	Short	Medium	Long	Average
<b>MAE</b>	3.436	3.538	3.574	3.516
<b>RMSE</b>	4.618	4.724	4.763	4.702
<b>MAPE</b>	11.483	11.899	12.046	11.809
<b>sMAPE</b>	11.587	12.000	12.136	11.908
<b>MASE</b>	0.677	0.673	0.650	0.667

Η μέθοδος ARIMA με εποχική αποσύνθεση είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 8 και 9, και ενδεικτικά να αναφέρουμε ότι έχει επιτευχθεί μείωση στο μέσο RMSE κατά 28.2%.

**Πίνακας 8** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο STL+ARIMA ως προς την μέθοδο αναφοράς.

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
<b>MAE</b>	29.3	28.4	27.7	27.1	26.5	26.2	25.9	25.6	25.4	25.4	25.1	25.1
<b>RMSE</b>	30.5	29.6	29.2	28.6	28.3	28.0	27.7	27.5	27.2	27.3	27.2	27.0
<b>MAPE</b>	40.6	39.6	38.8	38.2	37.6	37.3	37.0	36.7	36.5	36.5	36.3	36.1
<b>sMAPE</b>	32.7	31.7	30.8	30.1	29.4	29.1	28.9	28.5	28.3	28.3	28.0	27.9
<b>MASE</b>	29.4	28.4	27.6	27.1	26.6	26.2	25.9	25.6	25.3	25.4	25.1	25.1

Στον Πίνακα 9 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η βελτίωση που σημειώθηκε δεν ήρθε χωρίς «κόστος» καθώς η εν λόγω μέθοδος χρειάστηκε 66 φορές παραπάνω υπολογιστικό

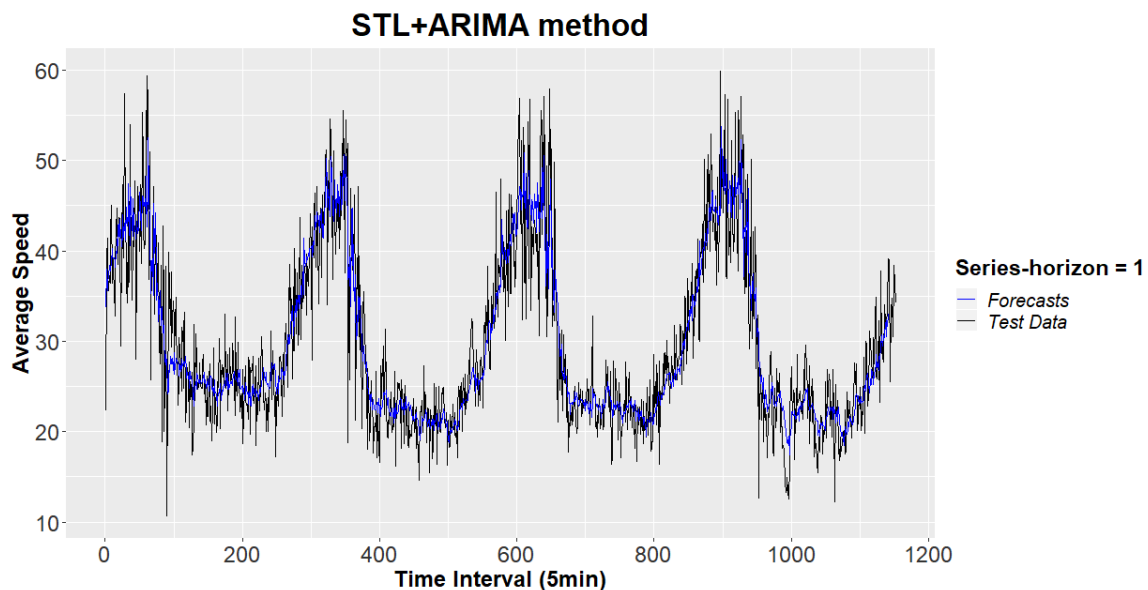
χρόνο σε σχέση με την μέθοδο αναφοράς. Τέλος, σημειώνεται βελτίωση (μείωση σχεδόν 50%) και στην προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης.

**Πίνακας 9** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο STL+ARIMA ως προς την μέθοδο αναφοράς.

	Short	Medium	Long	Average
<b>MAE</b>	28.1	26.1	25.2	26.5
<b>RMSE</b>	29.5	27.8	27.2	28.2
<b>MAPE</b>	39.3	37.2	36.3	37.6
<b>sMAPE</b>	31.3	29.0	28.1	29.5
<b>MASE</b>	28.1	26.0	25.3	26.5

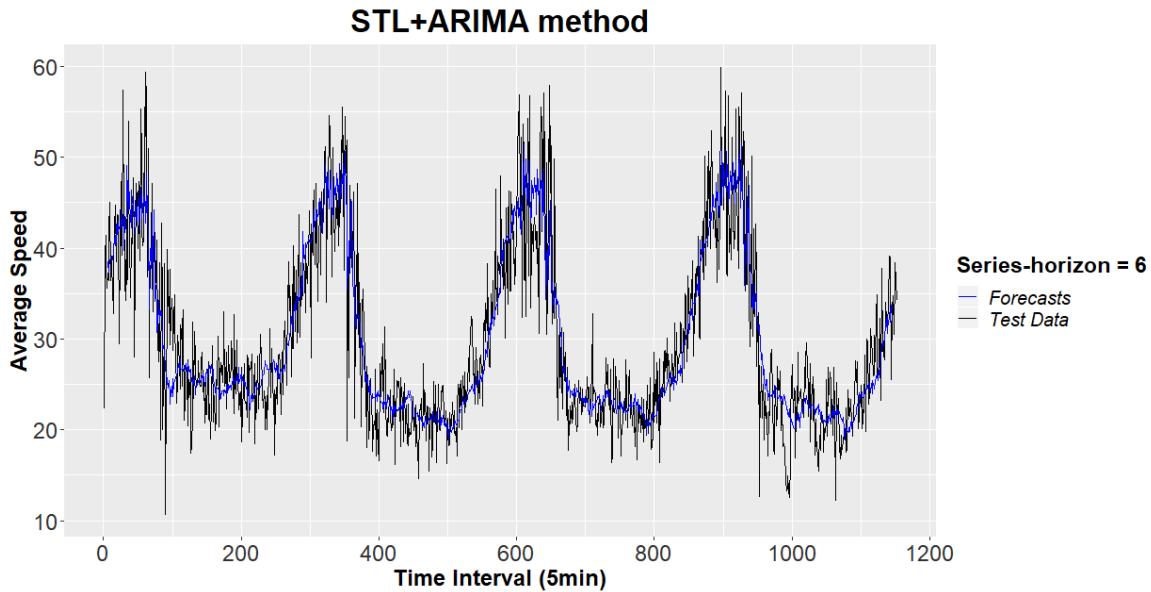
**Πίνακας 10** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο STL+ARIMA.

Train Time	Prediction Time	Total Time	CC	Model Fitting
8.6853 sec	145.5395 sec	154.2249 sec	66	2.766

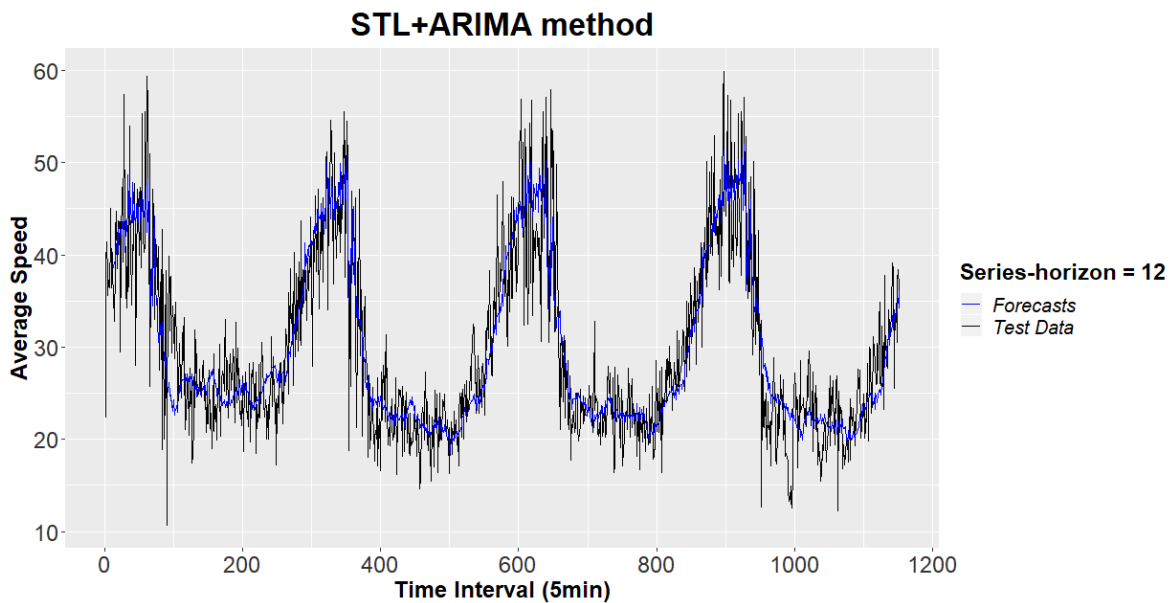


**Διάγραμμα 59** Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.

Στα Διαγράμματα 59, 60 και 61 δίνονται ενδεικτικά οι προβλέψεις της μεθόδου STL+ARIMA για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος φαίνεται να δίνει καλύτερες προβλέψεις σε σχέση με την μέθοδο βάσης, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις.

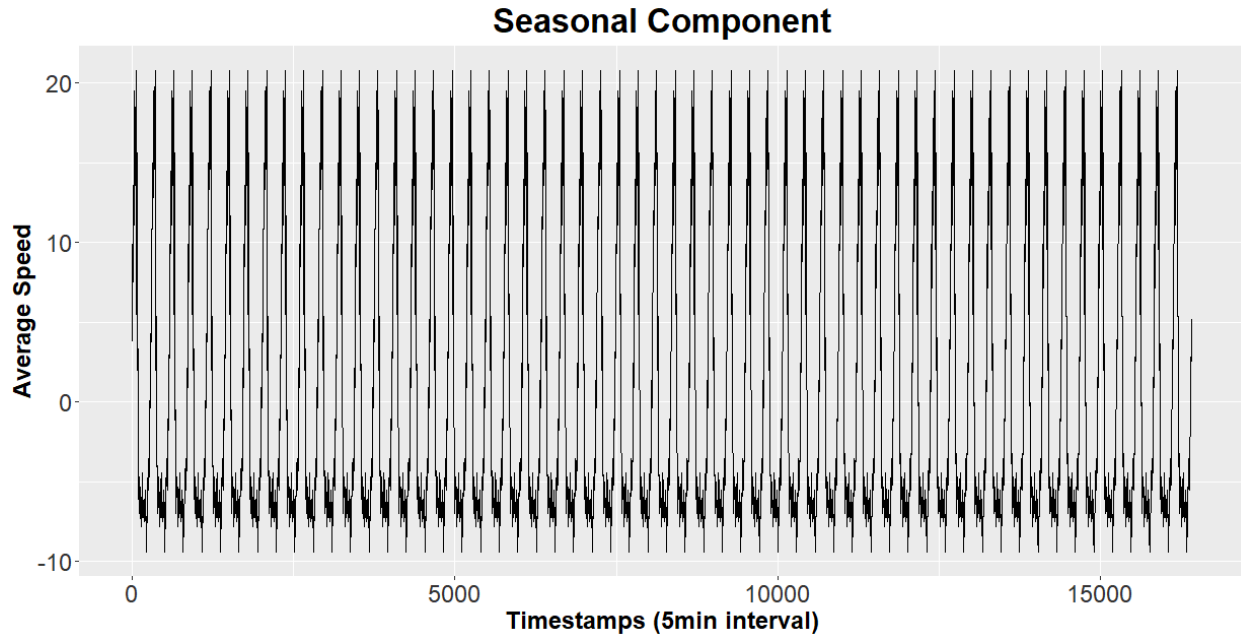


**Διάγραμμα 60** Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.

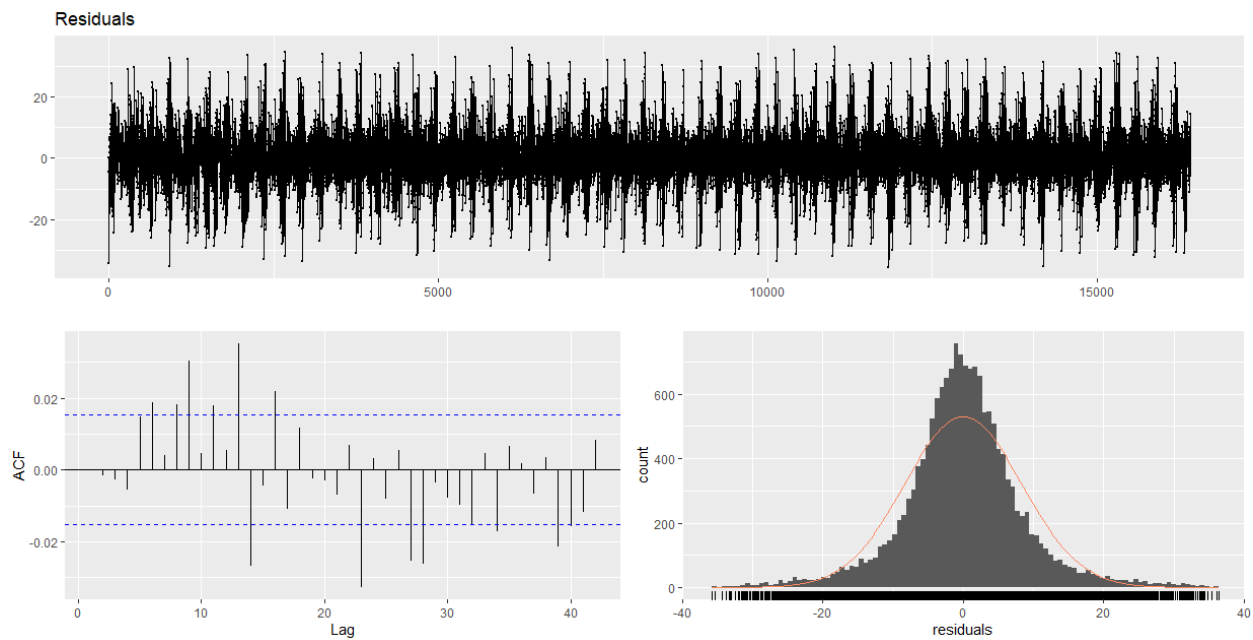


**Διάγραμμα 61** Προβλέψεις της μεθόδου STL+ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.





Διάγραμμα 62 Η εποχική συνιστώσα που υπολογίστηκε για την μέθοδο STL+ARIMA με μέθοδο LOESS.



Διάγραμμα 63 Γραφικός έλεγχος των καταλοίπων της μεθόδου STL+ARIMA.

Στο Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 63) των καταλοίπων φαίνεται να έχει μείνει κάποια πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος STL+ARIMA.

Βέβαια δεδομένου ότι το διάστημα εμπιστοσύνης έχει οριστεί να είναι 95%, το 5% των αυτοσυσχετίσεων που εμφανίζονται ως σημαντικές μπορεί να οφείλεται στην τυχαιότητα.

### 6.2.2 Δυναμική Αρμονική Παλινδρόμηση με σφάλματα ARIMA

Το μοντέλο που επιστράφηκε από την συνάρτηση *auto.arima()* με εξωγενείς *fourier* όρους είναι το ακόλουθο:

$$y_t = 6.02 \times 10^{-6} + \sum_{k=1}^9 \left[ a_k \sin\left(\frac{2\pi kt}{288}\right) + \beta_k \cos\left(\frac{2\pi kt}{288}\right) \right] + N_t \quad (6.5)$$

όπου  $N_t \sim \text{ARIMA}(5,1,1)$ .

Οι συντελεστές των όρων *fourier* της δυναμικής παλινδρόμησης φαίνονται στον Πίνακα 11, ενώ οι όροι της διαδικασίας ARIMA που ακολουθούν τα σφάλματα της δυναμικής παλινδρόμησης φαίνονται ακολούθως στον Πίνακα 12.

**Πίνακας 11** Συντελεστές των όρων *fourier* της δυναμικής παλινδρόμησης.

	$\alpha$	$\beta$
<b>1</b>	0.5593007	0.5096047
<b>2</b>	0.3037954	-0.00813119
<b>3</b>	-0.04093231	-0.09818516
<b>4</b>	-0.02112128	0.02417593
<b>5</b>	0.06036352	0.02498863
<b>6</b>	0.02871521	0.00605661
<b>7</b>	-0.05072084	0.01372564
<b>8</b>	-0.00170829	-0.02036829
<b>9</b>	-0.00666127	-0.01539865

**Πίνακας 12** Συντελεστές της διαδικασίας ARIMA που ακολουθούν τα σφάλματα της δυναμικής παλινδρόμησης.

<b>ar1</b>	<b>ar2</b>	<b>ar3</b>	<b>ar4</b>	<b>ar5</b>	<b>m1</b>
0.2044605	0.1259952	0.05332578	0.06353472	0.03293424	-0.9759184

Αυτό μεταφράζεται σε ένα ARIMA μοντέλο το οποίο έχει 5 AR στοιχεία, 1 MA στοιχείο. Έχει βρεθεί τάση στα δεδομένα σύμφωνα με τον KPSS έλεγχο, οπότε το μοντέλο προσαρμόστηκε στα δεδομένα στα οποία έχει εφαρμοσθεί διαφορίση 1<sup>ης</sup> τάξης. Και πάλι δεν υπάρχουν εποχικοί όροι στο μοντέλο ARIMA καθώς η εποχικότητα έχει μοντελοποιηθεί μέσω της δυναμικής παλινδρόμησης. Επίσης, επιλέχθηκε από το μοντέλο μετασχηματισμός Box-Cox με  $\lambda = 0.2200494$ .

Στον Πίνακα 13 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 14 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 13** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.432	3.492	3.521	3.549	3.578	3.591	3.607	3.629	3.646	3.637	3.646	3.654
<b>RMSE</b>	4.688	4.721	4.765	4.789	4.801	4.806	4.822	4.830	4.854	4.835	4.845	4.861
<b>MAPE</b>	11.427	11.684	11.827	11.939	12.077	12.140	12.199	12.276	12.334	12.317	12.356	12.387
<b>sMAPE</b>	11.432	11.664	11.803	11.923	12.060	12.112	12.169	12.243	12.298	12.280	12.319	12.347
<b>MASE</b>	0.707	0.681	0.684	0.685	0.688	0.685	0.684	0.681	0.676	0.665	0.659	0.654

**Πίνακας 14** Συνοπτικά μέτρα σφάλματος για την μέθοδο της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.499	3.601	3.646	3.582
<b>RMSE</b>	4.741	4.815	4.848	4.801
<b>MAPE</b>	11.719	12.173	12.349	12.080
<b>sMAPE</b>	11.705	12.146	12.311	12.054
<b>MASE</b>	0.689	0.685	0.664	0.679

**Πίνακας 15** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA.

<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
877.797 sec	228.708 sec	1106.506 sec	471	2.877

Η μέθοδος της δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 16 και 17. Στον Πίνακα 15 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η βελτίωση αυτή λοιπόν δεν ήρθε χωρίς «κόστος» καθώς η εν λόγω μέθοδος χρειάστηκε 471 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Ενδεικτικά να αναφερθεί ότι έχουμε επιτευχθεί μείωση στο μέσο RMSE κατά 26.6%.

**Πίνακας 16** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA ως προς την μέθοδο αναφοράς.

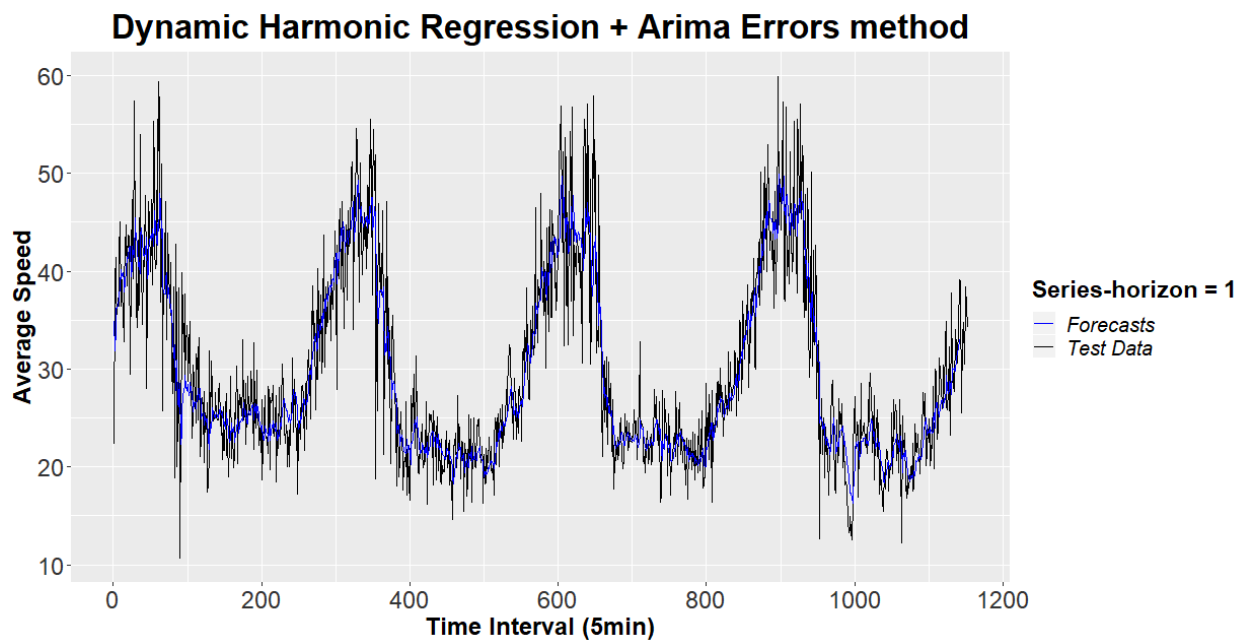
	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	28.4	26.9	26.2	25.8	25.2	25.0	24.7	24.2	23.8	23.9	23.6	23.6
<b>RMSE</b>	28.7	27.8	27.1	26.8	26.7	26.6	26.3	26.2	25.9	26.1	25.9	25.7
<b>MAPE</b>	39.7	38.2	37.4	36.9	36.2	35.9	35.6	35.2	34.9	34.9	34.7	34.5
<b>sMAPE</b>	32.4	30.8	29.9	29.4	28.6	28.3	28.0	27.6	27.2	27.2	27.0	26.9
<b>MASE</b>	28.4	26.8	26.2	25.8	25.3	25.0	24.6	24.2	23.8	23.9	23.6	23.6

**Πίνακας 17** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος μέθοδο δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA ως προς την μέθοδο αναφοράς.

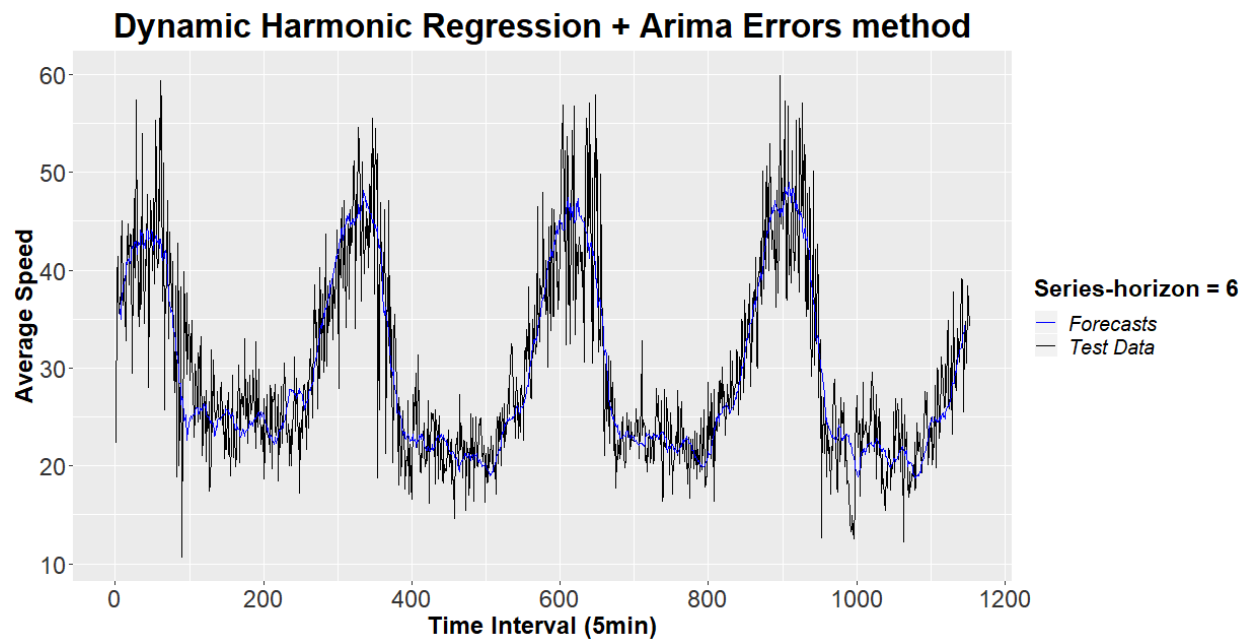
	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	26.8	24.8	23.7	25.1
<b>RMSE</b>	27.6	26.4	25.9	26.6
<b>MAPE</b>	38.0	35.7	34.7	36.2
<b>sMAPE</b>	30.6	28.1	27.1	28.6
<b>MASE</b>	26.8	24.8	23.7	25.1

Από τους πίνακες ποσοστιαίων μειώσεων του σφάλματος παρατηρείται ότι είναι μικρότερα τα ποσοστά για την εν λόγω μέθοδο σε σχέση με τα αντίστοιχα της μεθόδου STL+ARIMA. Δηλαδή η μέθοδος STL+ARIMA δίνει καλύτερες προβλέψεις έχοντας χαμηλότερη μέση τιμή για όλες τις μετρικές σφάλματος. Επιπροσθέτως, υπάρχει βελτίωση και στην προσαρμογή του μοντέλου για την μέθοδο STL+ARIMA και παρότι την καλύτερη ακρίβεια της είναι και πιο «φθηνή» όσον αφορά το υπολογιστικό κόστος σε σχέση με την μέθοδο της δυναμικής παλινδρόμησης με σφάλματα Arima, και πιο συγκεκριμένα 7 φορές περίπου πιο «γρήγορη».

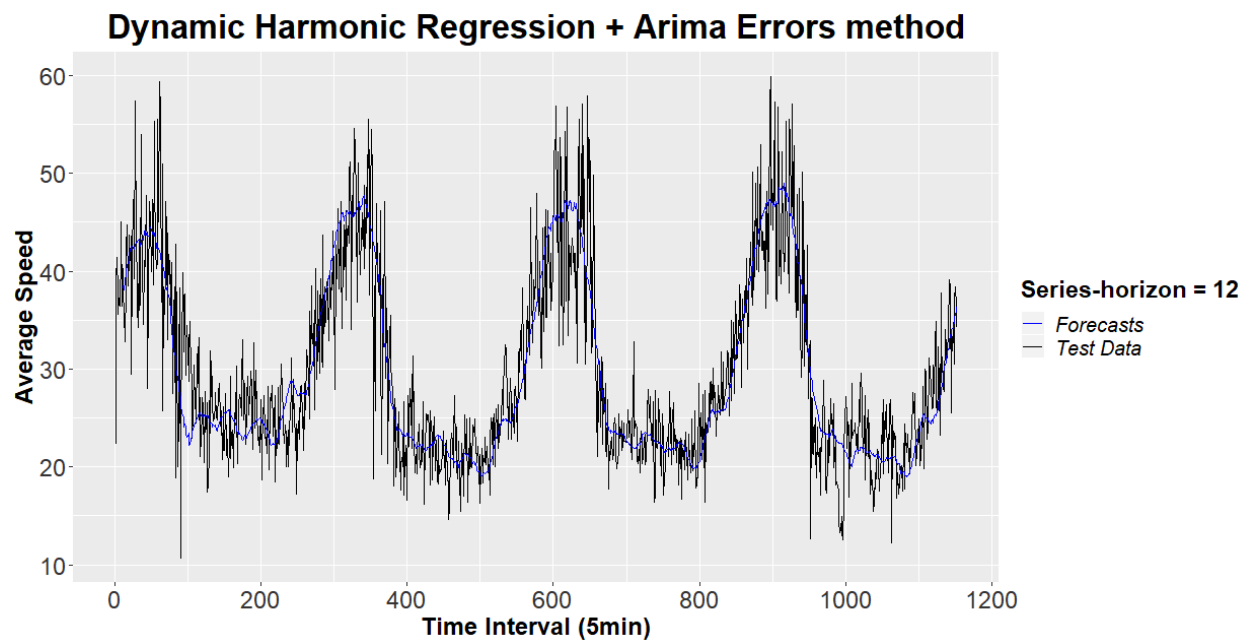
Στα Διαγράμματα 64, 65 και 66 δίνονται γραφικά οι προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει καλύτερες προβλέψεις συγκριτικά με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Σε σύγκριση με την μέθοδο STL+ARIMA δεν μπορούν να σχολιαστούν οπτικά τουλάχιστον σημαντικές διαφορές.



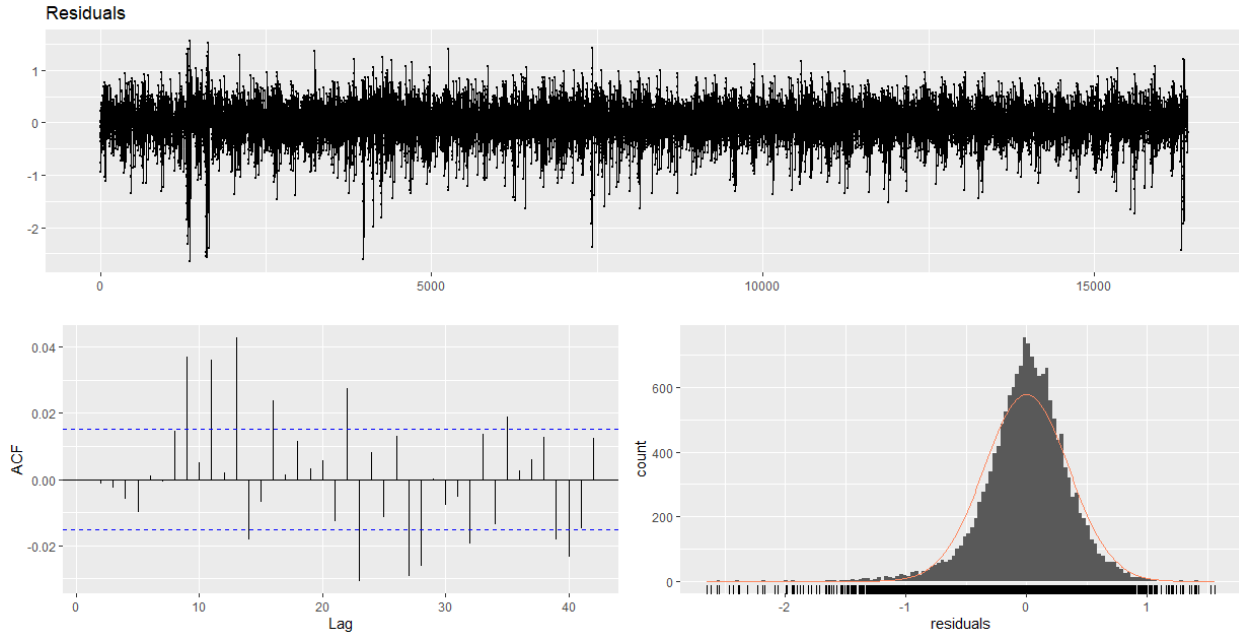
**Διάγραμμα 64** Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 65** Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 66** Προβλέψεις της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



**Διάγραμμα 67** Γραφικός έλεγχος των καταλοίπων της μεθόδου δυναμικής αρμονικής παλινδρόμησης με σφάλματα ARIMA.

Από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 67) των καταλοίπων δημιουργείται η υποψία ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος. Βέβαια δεδομένου ότι το διάστημα εμπιστοσύνης έχει οριστεί στο 95%, το 5% των αυτοσυσχετίσεων που εμφανίζονται σημαντικές ενδεχομένως να οφείλεται στην τυχαιότητα.

### 6.2.3 Μέθοδος ETS με εποχική αποσύνθεση (STL+ETS)

Το μοντέλο που εκτιμήθηκε από την συνάρτηση `ets()` είναι το  $ETS(A, N, N)$  δηλαδή το μοντέλο simple exponential smoothing με προσθετικά σφάλματα. Από την κωδικοποίηση που παρατέθηκε παρατηρούμε ότι δεν βρέθηκε τάση και εποχικότητα στα δεδομένα εκπαίδευσης. Επίσης έγινε χρήση του μετασχηματισμού Box-Cox με  $\lambda = 1.16833$ . Το μοντέλο μπορεί να γραφεί ως:

$$\text{Εξίσωση πρόβλεψης (Forecast equation): } \hat{y}_{t+h|t} = l_t$$

$$\text{Εξίσωση Ομαλοποίησης (Smoothing equation): } l_t = 0.084y_t + (1 - 0.084)l_{t-1}$$

Η διορθωμένη μορφή ως προς το σφάλμα της εξίσωσης ομαλοποίησης δίνεται από την σχέση:

$$l_t = l_{t-1} + 0.084(y_t - l_{t-1}) = l_{t-1} + 0.084e_t$$

όπου  $e_t = y_t - l_{t-1} = y_t - \hat{y}_{t|t-1}$ .

Μπορεί επίσης να ξαναγραφεί υπό τη μορφή  $y_t = l_{t-1} + e_t$ , έτσι ώστε κάθε παρατήρηση να μπορεί να αναπαρασταθεί ως άθροισμα του προηγούμενου επιπέδου και του σφάλματος. Επιπλέον, υποθέτοντας ότι για τα σφάλματα ισχύει ότι  $e_t = \varepsilon_t \sim NID(0, \sigma^2)$ , καταλήγουμε στις τελικές εξισώσεις που περιγράφουν το μοντέλο με προσθετικά σφάλματα οι οποίες είναι,

$$y_t = l_{t-1} + \varepsilon_t \quad (\text{εξίσωση μέτρησης}) \quad (6.6)$$

$$l_t = l_{t-1} + 0.084\varepsilon_t \quad (\text{εξίσωση κατάστασης}) \quad (6.7)$$

Η εποχική συνιστώσα για αυτή την μέθοδο είναι ίδια με την αντίστοιχη της STL+ARIMA μεθόδου (Διάγραμμα 62). Στον Πίνακα 18 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήσαμε ανά ορίζοντα πρόβλεψης. Στον Πίνακα 19 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 18** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο STL+ETS.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.404	3.437	3.471	3.519	3.560	3.588	3.609	3.634	3.655	3.636	3.647	3.662
<b>RMSE</b>	4.598	4.631	4.664	4.710	4.743	4.772	4.800	4.822	4.849	4.828	4.841	4.865
<b>MAPE</b>	11.365	11.507	11.652	11.830	11.988	12.094	12.174	12.263	12.344	12.310	12.359	12.418
<b>sMAPE</b>	11.473	11.601	11.741	11.924	12.083	12.179	12.257	12.345	12.421	12.382	12.435	12.488
<b>MASE</b>	0.701	0.670	0.674	0.679	0.685	0.685	0.685	0.682	0.678	0.665	0.659	0.656

**Πίνακας 19** Συνοπτικά μέτρα σφάλματος για την μέθοδο STL+ETS.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.458	3.598	3.650	3.569
<b>RMSE</b>	4.651	4.784	4.846	4.760
<b>MAPE</b>	11.589	12.130	12.358	12.025
<b>sMAPE</b>	11.685	12.216	12.432	12.111
<b>MASE</b>	0.681	0.684	0.664	0.677



**Πίνακας 20** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο STL+ETS.

<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
2.697 sec	125.139 sec	127.837 sec	54	2.802

Η μέθοδος ETS με εποχική αποσύνθεση είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 20 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η βελτίωση που σημειώθηκε δεν ήρθε χωρίς «κόστος» καθώς η εν λόγω μέθοδος χρειάστηκε 54 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 21 και 22. Επιτεύχθηκε μείωση στο μέσο RMSE κατά 27.3%.

**Πίνακας 21** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο STL+ETS ως προς την μέθοδο αναφοράς.

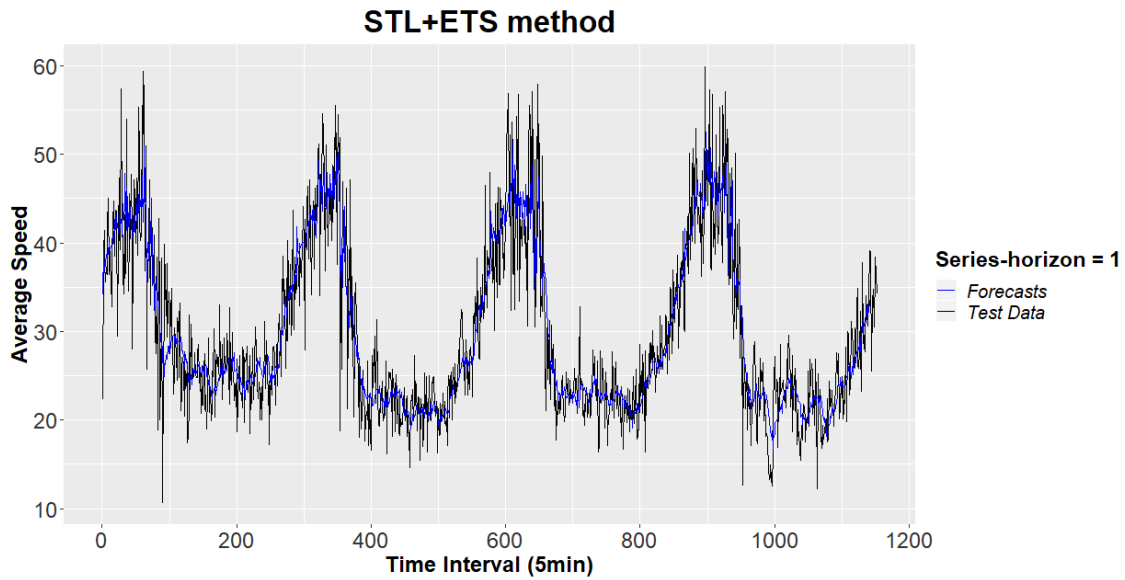
	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.0	28.0	27.3	26.4	25.6	25.0	24.6	24.1	23.6	23.9	23.6	23.4
<b>RMSE</b>	30.0	29.2	28.6	28.0	27.5	27.1	26.7	26.3	25.9	26.2	25.9	25.6
<b>MAPE</b>	40.0	39.1	38.3	37.5	36.7	36.1	35.7	35.3	34.8	34.9	34.6	34.4
<b>sMAPE</b>	32.1	31.2	30.3	29.4	28.5	27.9	27.5	27.0	26.5	26.6	26.3	26.0
<b>MASE</b>	29.0	28.0	27.3	26.4	25.6	25.0	24.6	24.1	23.6	23.9	23.6	23.4

**Πίνακας 22** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο STL+ETS ως προς την μέθοδο αναφοράς.

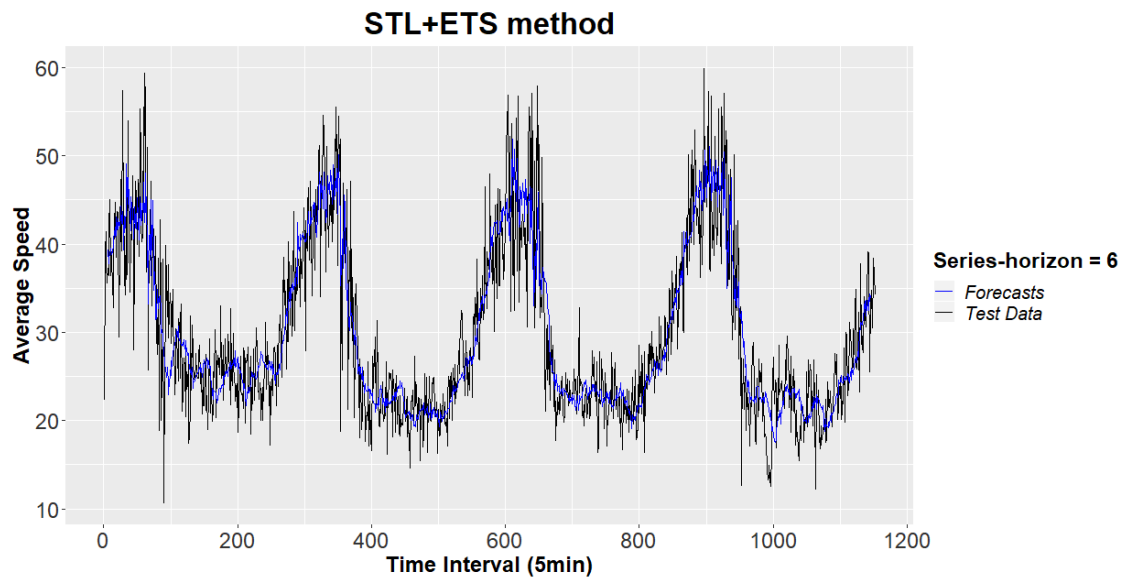
	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	27.7	24.8	23.6	25.4
<b>RMSE</b>	29.0	26.9	25.9	27.3
<b>MAPE</b>	38.7	36.0	34.7	36.5
<b>sMAPE</b>	30.7	27.7	26.4	28.3
<b>MASE</b>	27.7	24.8	23.6	25.4

Από τους Πίνακες ποσοστιαίων μειώσεων του σφάλματος παρατηρούμε ότι η μέθοδος STL+ETS είναι καλύτερη από την μέθοδο της δυναμικής αρμονικής παλινδρόμησης και λίγο χειρότερη από

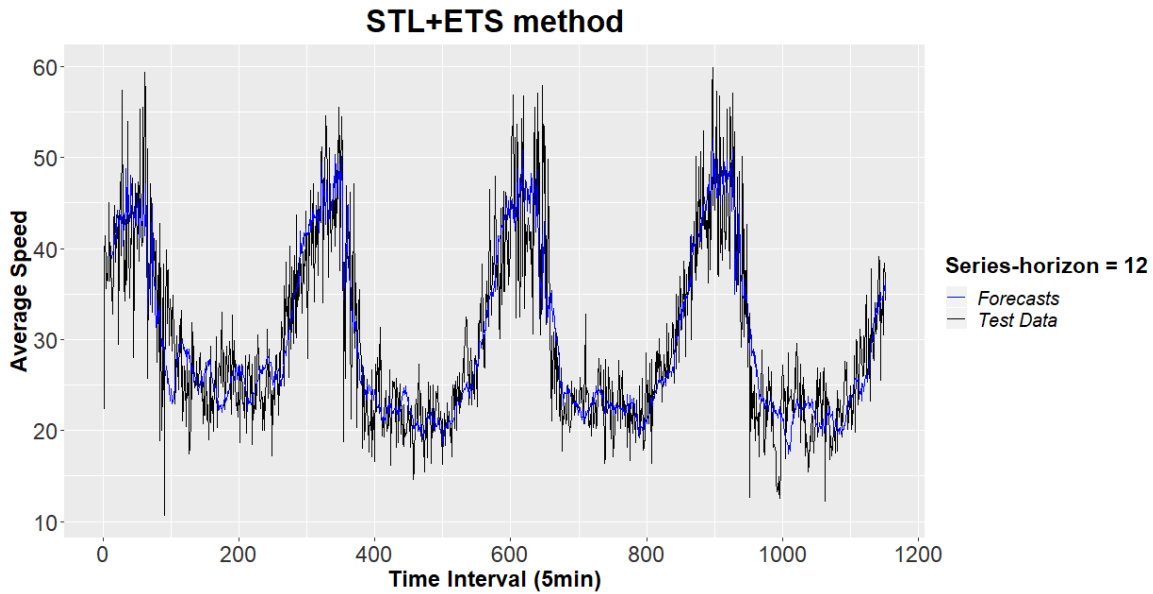
την μέθοδο STL+ARIMA. Το ίδιο ισχύει και με την προσαρμογή του μοντέλου. Όσον αφορά το υπολογιστικό κόστος είναι πιο «γρήγορη» από την μέθοδο STL+ARIMA, οριακά.



**Διάγραμμα 68** Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.

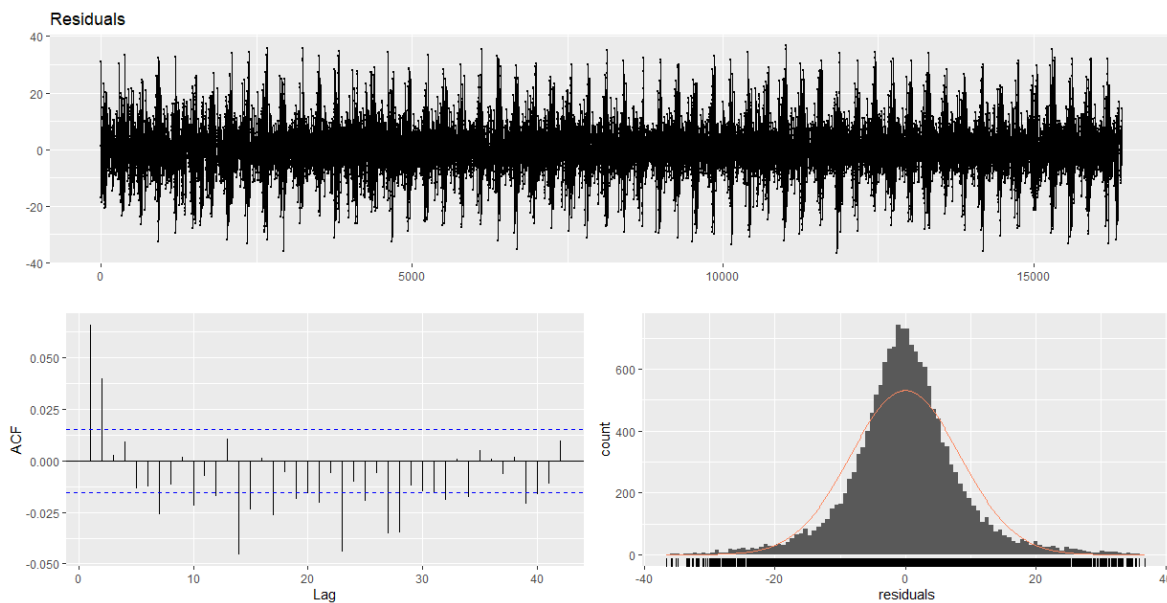


**Διάγραμμα 69** Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 70** Προβλέψεις της μεθόδου STL+ETS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.

Στα Διαγράμματα 68 , 69 και 70 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις.



**Διάγραμμα 71** Γραφικός έλεγχος των καταλοίπων της μεθόδου STL+ETS.

Στο Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 71) των καταλοίπων παρατηρείται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος. Βέβαια δεδομένου ότι το διάστημα εμπιστοσύνης είναι 95% αυτό επιτρέπει να σχολιαστεί ότι το 5% των αυτοσυσχετίσεων που εμφανίζονται σημαντικές μπορεί να οφείλεται στην τυχαιότητα.

#### 6.2.4 Μέθοδος TBATS

Η εντολή `tbats()` επιστρέφει το ακόλουθο μοντέλο:

$$y_t^{(\omega)} = \frac{y_t^{0.613} - 1}{0.613} \quad (6.8)$$

$$y_t^{(\omega)} = l_{t-1} + s_{t-1} + d_t \quad (6.9)$$

$$l_t = l_{t-1} + 0.0862d_t \quad \mu\epsilon \quad l_0 = 10.509 \quad (6.10)$$

$$s_t = \sum_{j=1}^7 s_{j,t} \quad (6.11)$$

$$s_{j,t} = s_{j,t-1} \cos \frac{2\pi j}{288} + s_{j,t-1}^* \sin \frac{2\pi j}{288} + 0.0105d_t \quad (6.12)$$

$$s_{j,t}^* = -s_{j,t-1} \sin \frac{2\pi j}{288} + s_{j,t-1}^* \cos \frac{2\pi j}{288} + 0.006d_t \quad (6.13)$$

$$d_t = \sum_{i=1}^2 \varphi_i d_{t-i} + \sum_{i=1}^3 \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad \mu\epsilon \quad \varepsilon_t \sim NID(0, \sigma^2) \quad (6.14)$$

όπου  $d_t \sim ARMA(2,3)$  με τους συντελεστές των όρων της που δίνονται στον Πίνακα 23.

**Πίνακας 23** Οι συνιστώσες τον ARMA σφαλμάτων της TBATS μεθόδου.

$\varphi_1$	$\varphi_2$	$\theta_1$	$\theta_2$	$\theta_3$
0.275	0.656	-0.062	-0.538	-0.063

με αρχικές τιμές για τις παραμέτρους  $s_{j,0}$  και  $s_{j,0}^*$ ,  $j = 1, \dots, 7$  το διάνυσμα (εναλλάξ) που δίνεται στον Πίνακα 24.

**Πίνακας 24** Αρχικές τιμές των παραμέτρων της εποχικότητας για την μέθοδο TBATS.

4.3910	-0.2994	-2.9927	0.2793	0.4998	0.0368	-0.0609	4.5626	3.2349	-0.1249	-0.5277	0.5945	0.4224	-0.6464
--------	---------	---------	--------	--------	--------	---------	--------	--------	---------	---------	--------	--------	---------

Στον Πίνακα 25 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 26 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

Η μέθοδος Tbats είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 27 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Αυτή η βελτίωση δεν ήρθε χωρίς «κόστος» καθώς η εν λόγω μέθοδος χρειάστηκε 67 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο αναφοράς. Οι σχετικές βελτιώσεις του σφάλματος δίνονται στους Πίνακες 28 και 29. Ενδεικτικά να αναφέρουμε ότι η μείωση στο μέσο RMSE είναι 23.3%.

**Πίνακας 25** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Tbats.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.482	3.546	3.598	3.677	3.723	3.775	3.837	3.871	3.930	3.918	3.956	3.978
<b>RMSE</b>	4.748	4.816	4.895	4.948	4.995	5.023	5.070	5.098	5.160	5.137	5.171	5.216
<b>MAPE</b>	11.607	11.883	12.117	12.437	12.654	12.867	13.087	13.221	13.421	13.415	13.563	13.660
<b>sMAPE</b>	11.648	11.895	12.117	12.434	12.641	12.835	13.046	13.176	13.372	13.358	13.513	13.597
<b>MASE</b>	0.717	0.692	0.699	0.710	0.716	0.720	0.728	0.727	0.729	0.717	0.715	0.712

**Πίνακας 26** Συνοπτικά μέτρα σφάλματος για την μέθοδο Tbats.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.576	3.802	3.945	3.774
<b>RMSE</b>	4.852	5.047	5.171	5.023
<b>MAPE</b>	12.011	12.957	13.515	12.828
<b>sMAPE</b>	12.024	12.924	13.460	12.803
<b>MASE</b>	0.704	0.723	0.718	0.715

**Πίνακας 27** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Tbats.

<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
140.887 sec	15.993 sec	156.880 sec	67	2.97

Από τους Πίνακες ποσοστιαίων μειώσεων του σφάλματος συνάγεται ότι η μέθοδος Tbats επιστρέφει τα χειρότερα αποτελέσματα από όλες τις μεθόδους της ομάδας 2 που έχουν εξεταστεί έως τώρα, για όλους τους ορίζοντες προβλέψεις και ακολούθως για όλα τα συνοπτικά μέτρα. Έχει επίσης την χειρότερη προσαρμογή στα δεδομένα εκπαίδευσης. Όσον αφορά το υπολογιστικό κόστος είναι ίδιο σχεδόν με το αντίστοιχο της μεθόδου STL+ARIMA.

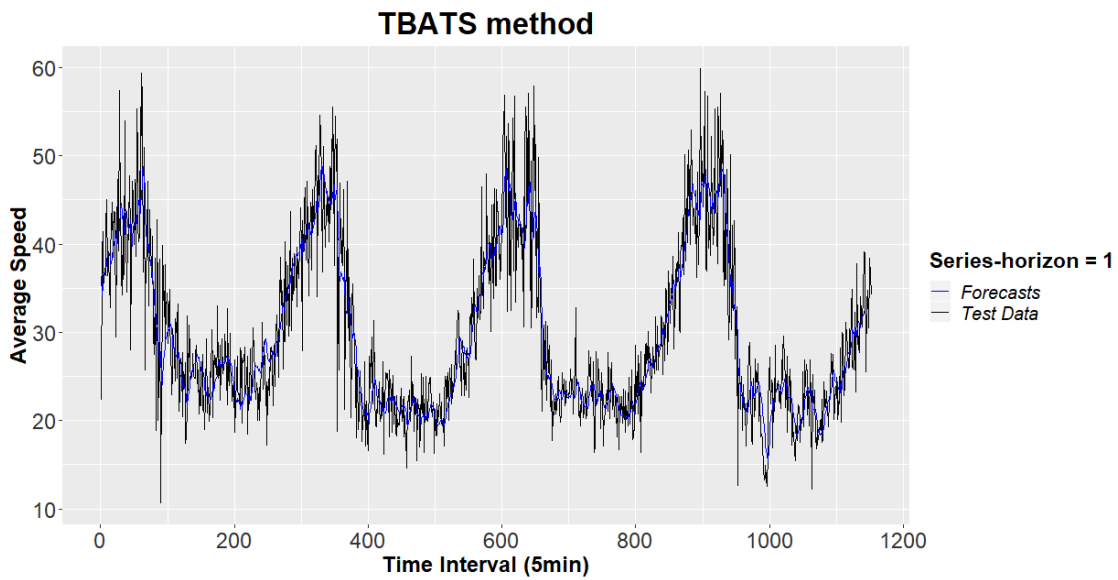
**Πίνακας 28** Ποσοστιαία μείωση για όλες τις μετρικές ράσφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Tbats ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	27.3	25.8	24.6	23.1	22.2	21.1	19.8	19.1	17.9	18.0	17.1	16.8
<b>RMSE</b>	27.7	26.3	25.1	24.4	23.7	23.3	22.5	22.1	21.2	21.4	20.9	20.2
<b>MAPE</b>	38.7	37.2	35.9	34.3	33.2	32.1	30.9	30.2	29.2	29.1	28.3	27.8
<b>sMAPE</b>	31.1	29.4	28.1	26.3	25.2	24.0	22.8	22.1	20.9	20.8	19.9	19.5
<b>MASE</b>	27.3	25.7	24.6	23.1	22.2	21.1	19.8	19.2	17.9	18.0	17.1	16.8

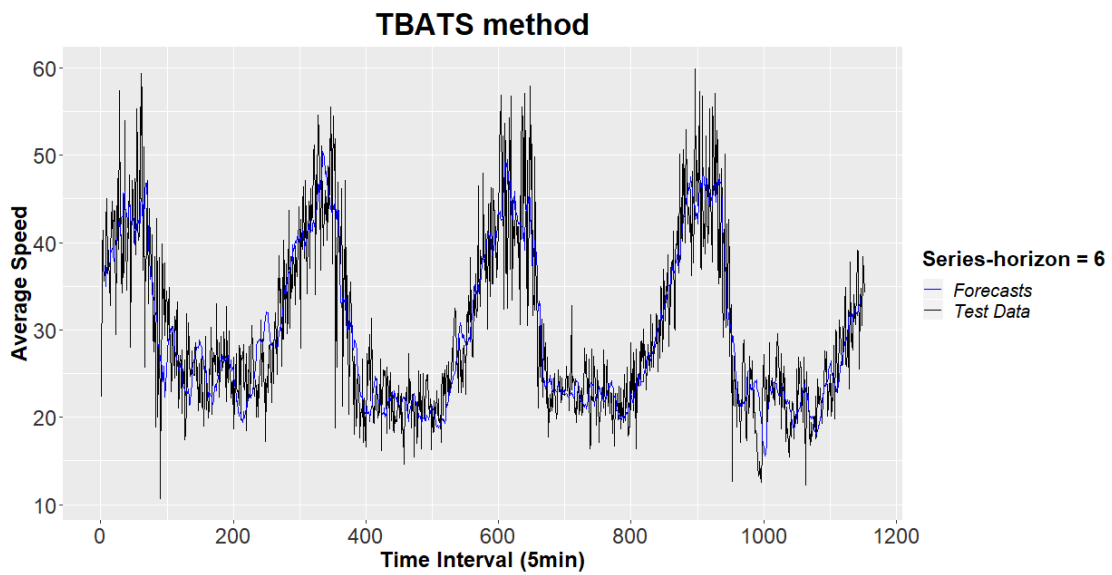
**Πίνακας 29** Συνοπτικά μέτρα σφάλματος για όλους τους ορίζοντες για την μέθοδο Tbats.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	25.2	20.6	17.5	21.1
<b>RMSE</b>	25.9	22.9	20.9	23.3
<b>MAPE</b>	36.5	31.6	28.6	32.2
<b>sMAPE</b>	28.7	23.5	20.3	24.2
<b>MASE</b>	25.2	20.6	17.5	21.2

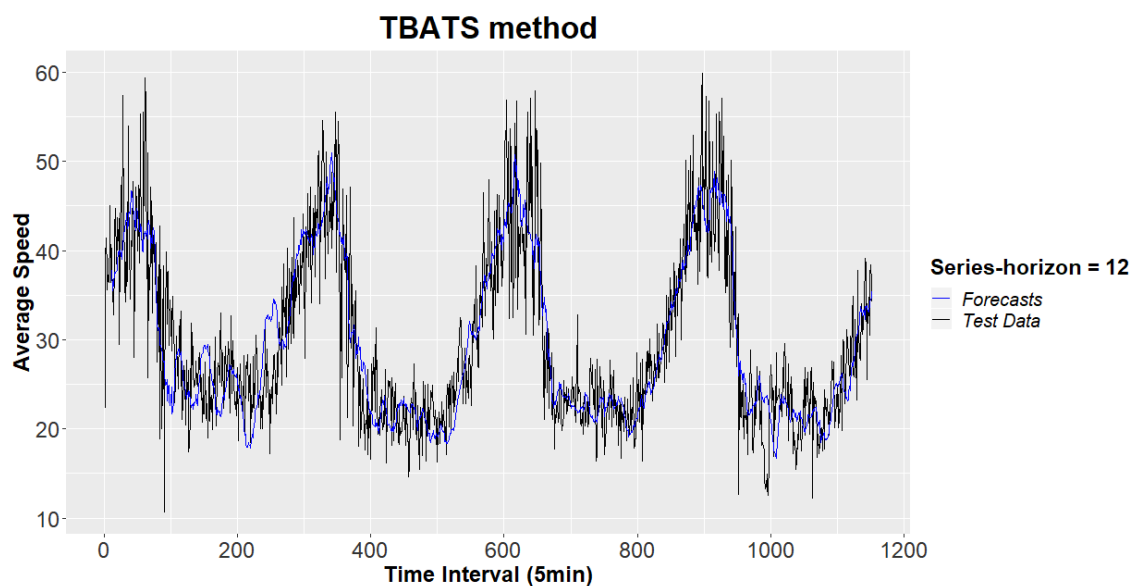
Στα Διαγράμματα 72, 73 και 74 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει καλύτερες προβλέψεις σε σχέση με την μέθοδο βάσης, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις.



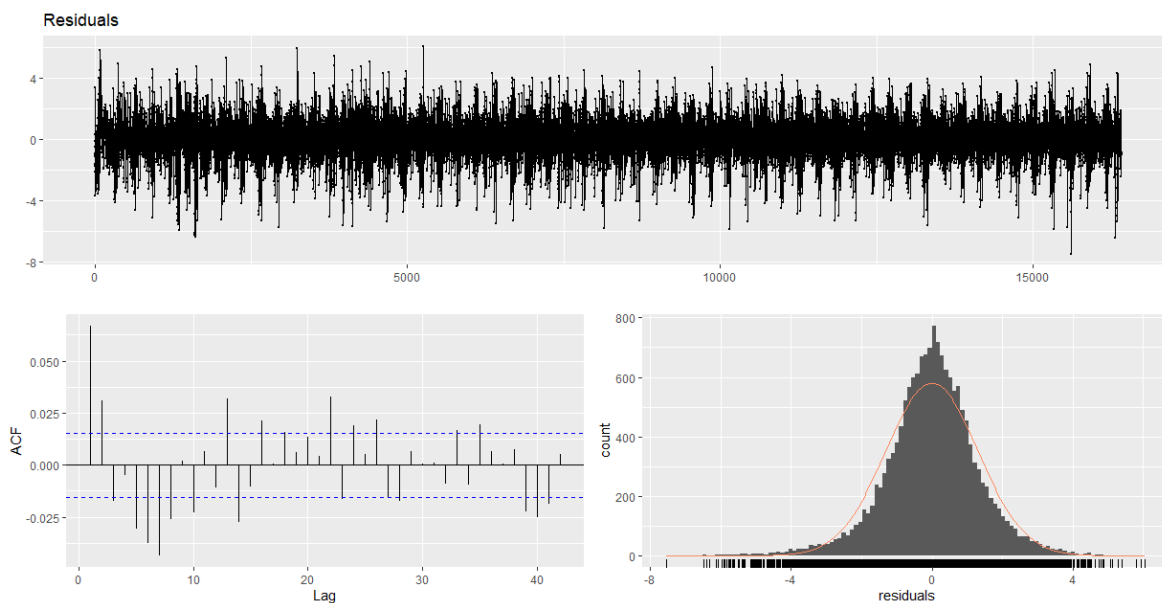
Διάγραμμα 72 Προβλέψεις της μεθόδου Tbats για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



Διάγραμμα 73 Προβλέψεις της μεθόδου Tbats για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 74** Προβλέψεις της μεθόδου TbatS για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



**Διάγραμμα 75** Γραφικός έλεγχος των καταλοίπων της μεθόδου TbatS.

Από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 75) των καταλοίπων φαίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος. Βέβαια



δεδομένου ότι το διάστημα εμπιστοσύνης είναι 95%, το 5% των αυτοσυσχετίσεων που εμφανίζονται σημαντικές μπορεί να οφείλεται στην τυχαιότητα.

### 6.2.5 Μέθοδος Theta

Η εντολή `thetaf()` δεν επιστρέφει κάποια πληροφορία όσον αφορά την δομή του μοντέλου πληροφορίες. Στον Πίνακα 30 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήσαμε ανά ορίζοντα πρόβλεψης. Στον Πίνακα 31 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 30** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Theta.

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
<b>MAE</b>	3.421	3.481	3.515	3.581	3.629	3.655	3.691	3.721	3.729	3.706	3.727	3.745
<b>RMSE</b>	4.637	4.689	4.728	4.787	4.824	4.857	4.899	4.927	4.958	4.920	4.939	4.969
<b>MAPE</b>	11.402	11.650	11.814	12.066	12.250	12.359	12.488	12.586	12.626	12.579	12.660	12.748
<b>sMAPE</b>	11.486	11.715	11.868	12.125	12.310	12.404	12.530	12.628	12.664	12.611	12.698	12.771
<b>MASE</b>	0.705	0.679	0.683	0.691	0.698	0.697	0.700	0.698	0.692	0.678	0.673	0.671

**Πίνακας 31** Συνοπτικά μέτρα σφάλματος για την μέθοδο Theta.

	Short	Medium	Long	Average
<b>MAE</b>	3.500	3.674	3.727	3.633
<b>RMSE</b>	4.710	4.877	4.947	4.844
<b>MAPE</b>	11.733	12.421	12.653	12.269
<b>sMAPE</b>	11.798	12.468	12.686	12.317
<b>MASE</b>	0.689	0.699	0.678	0.689

**Πίνακας 32** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Theta.

Train Time	Prediction Time	Total Time	CC	Model Fitting
0.130 sec	131.204 sec	131.334 sec	56	2.776

Η μέθοδος Theta είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 32 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η βελτίωση που αναφέρθηκε πιο πάνω δεν ήρθε χωρίς «κόστος» καθώς η εν λόγω μέθοδος χρειάστηκε 56 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 33 και 34. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 26%.

**Πίνακας 33** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Theta ως προς την μέθοδο αναφοράς.

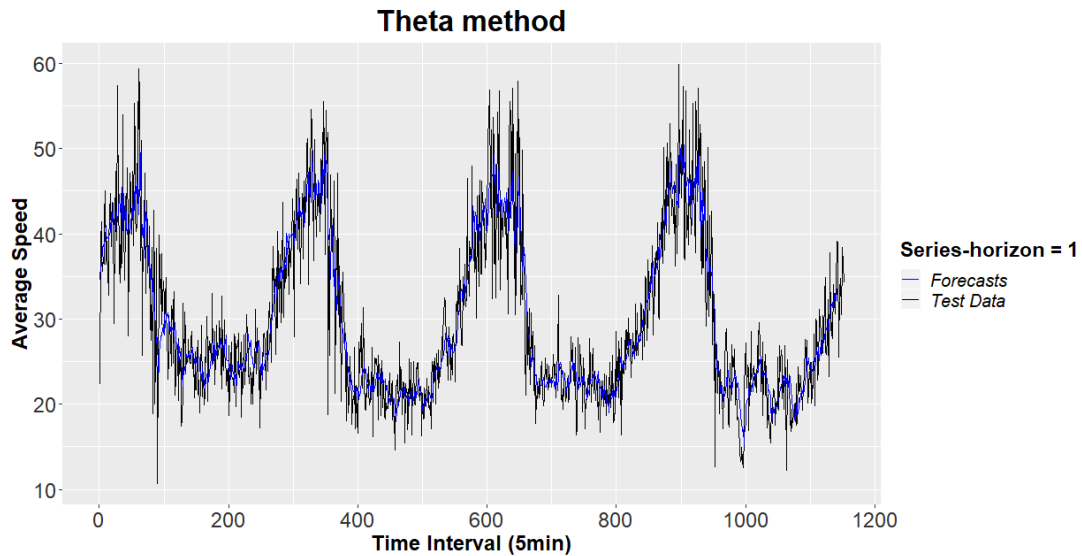
	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	28.6	27.1	26.3	25.1	24.2	23.6	22.9	22.3	22.1	22.4	21.9	21.7
<b>RMSE</b>	29.4	28.3	27.7	26.9	26.3	25.8	25.2	24.7	24.3	24.8	24.4	24.0
<b>MAPE</b>	39.8	38.4	37.5	36.2	35.3	34.7	34.1	33.6	33.4	33.5	33.0	32.6
<b>sMAPE</b>	32.1	30.5	29.6	28.2	27.1	26.6	25.8	25.3	25.1	25.3	24.7	24.4
<b>MASE</b>	28.6	27.1	26.3	25.1	24.2	23.6	22.9	22.3	22.1	22.4	21.9	21.7

**Πίνακας 34** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Theta ως προς την μέθοδο αναφοράς.

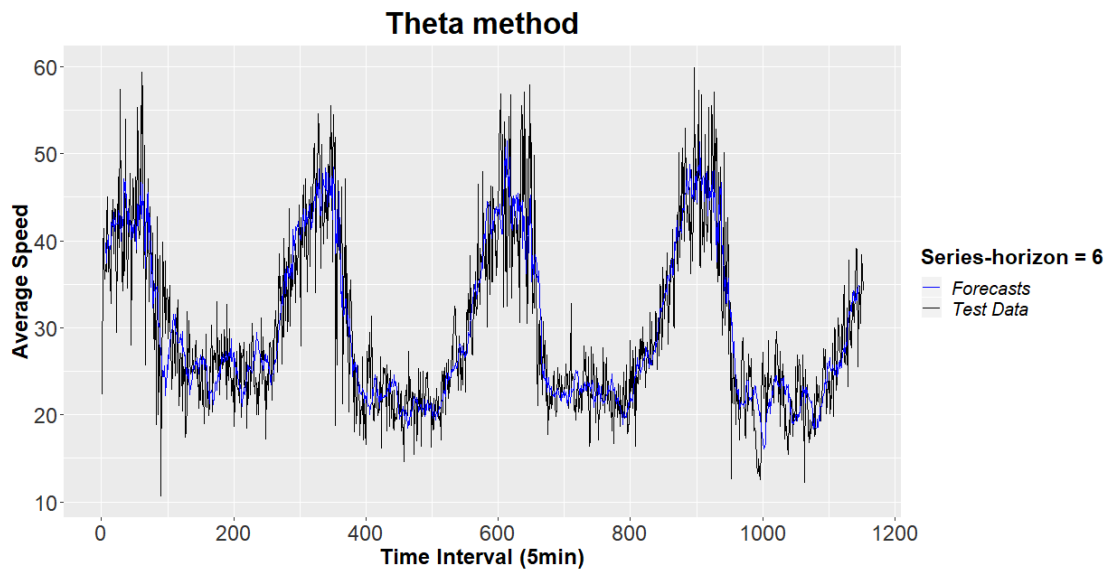
	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	26.8	23.2	22.0	24.0
<b>RMSE</b>	28.1	25.5	24.4	26.0
<b>MAPE</b>	38.0	34.4	33.1	35.2
<b>sMAPE</b>	30.1	26.2	24.9	27.0
<b>MASE</b>	26.8	23.2	22.0	24.1

Από τους Πίνακες ποσοστιαίων μειώσεων του σφάλματος παρατηρείται ότι η μέθοδος Theta είναι η 2<sup>η</sup> χειρότερη μέθοδος. Ενώ έχει καλύτερο βραχυπρόθεσμο μέσο σφάλμα από την μέθοδο της δυναμικής αρμονικής παλινδρόμησης στο μεσοπρόθεσμο και μακροπρόθεσμο μέσο σφάλμα είναι χειρότερη. Έχει την 2<sup>η</sup> καλύτερη προσαρμογή στα δεδομένα εκπαίδευσης μετά την μέθοδο STL+ARIMA. Όσον αφορά το υπολογιστικό κόστος είναι η πιο «γρήγορη» μέθοδος (λίγο καλύτερη από την μέθοδο STL+ETS). Στα Διαγράμματα 76, 77 και 78 φαίνονται ενδεικτικά τις

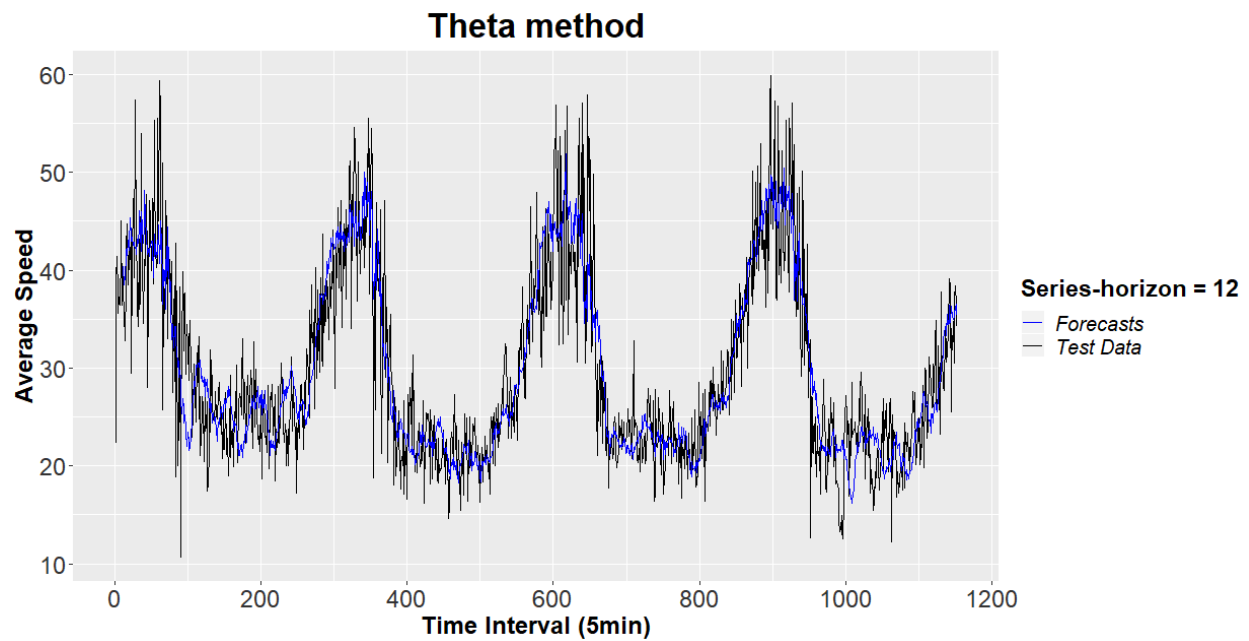
προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται δίνει καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις.



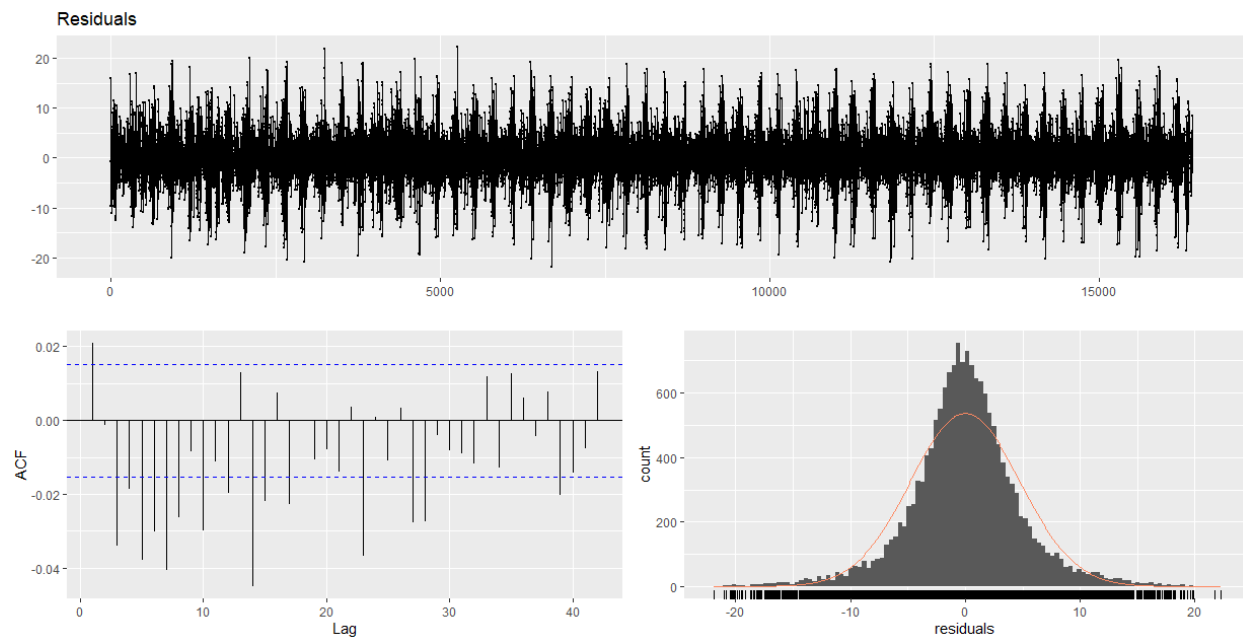
**Διάγραμμα 76** Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 77** Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 78** Προβλέψεις της μεθόδου Theta για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



**Διάγραμμα 79** Γραφικός έλεγχος των καταλοίπων της μεθόδου Theta.

Από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 79) των καταλοίπων μπορεί να εξαχθεί ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος. Το διάστημα εμπιστοσύνης είναι 95% οπότε το 5% των αυτοσυσχετίσεων που εμφανίζονται σημαντικές μπορεί να οφείλεται στην τυχαιότητα.

Εν κατακλείδει, όσον αφορά τις στατιστικές μεθόδους (ομάδα 2) με εξαίρεση ίσως την μέθοδο TbatS δεν σημειώνονται μεγάλες αποκλίσεις στην προβλεπτική ικανότητα των μοντέλων. Το ίδιο ισχύει και για την προσαρμοστικότητα στα δεδομένα εκπαίδευσης. Η καλύτερη από τις μεθόδους της ομάδας 2 είναι η STL+ARIMA μέθοδος, ενώ η χειρότερη είναι η μέθοδος TBATS.

### 6.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Υπενθυμίζουμε ότι στην ομάδα 3 ανήκουν οι ακόλουθες 10 μέθοδοι:

- Multi-layer Perceptron (MLP)
- Long-Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Convolutional Neural Networks (CNN)
- Temporal Convolutional Neural Networks (TCNN)
- CNN-LSTM
- Convolutional LSTM (Conv-LSTM)
- Encoder-Decoder LSTM
- Encoder-Decoder CNN-LSTM
- Encoder-Decoder Conv-LSTM
- 

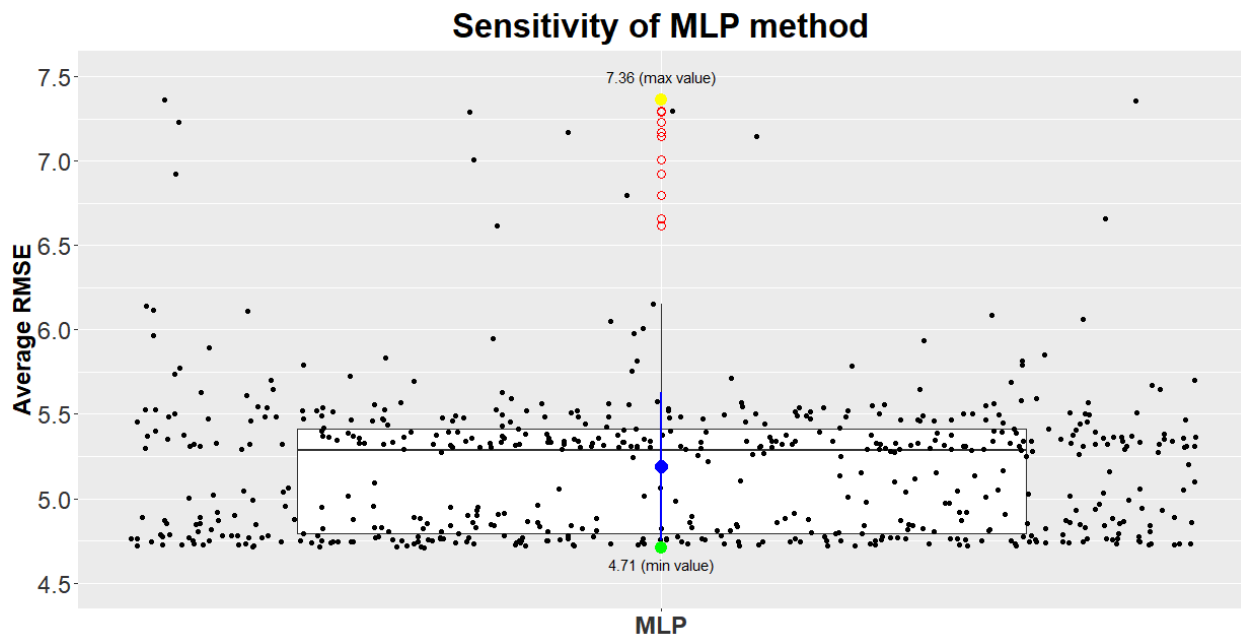
#### 6.3.1 Μέθοδος Multi-layer Perceptron (MLP)

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προέκυψε είναι το εξής (ως προς τις υπερπαραμέτρους):

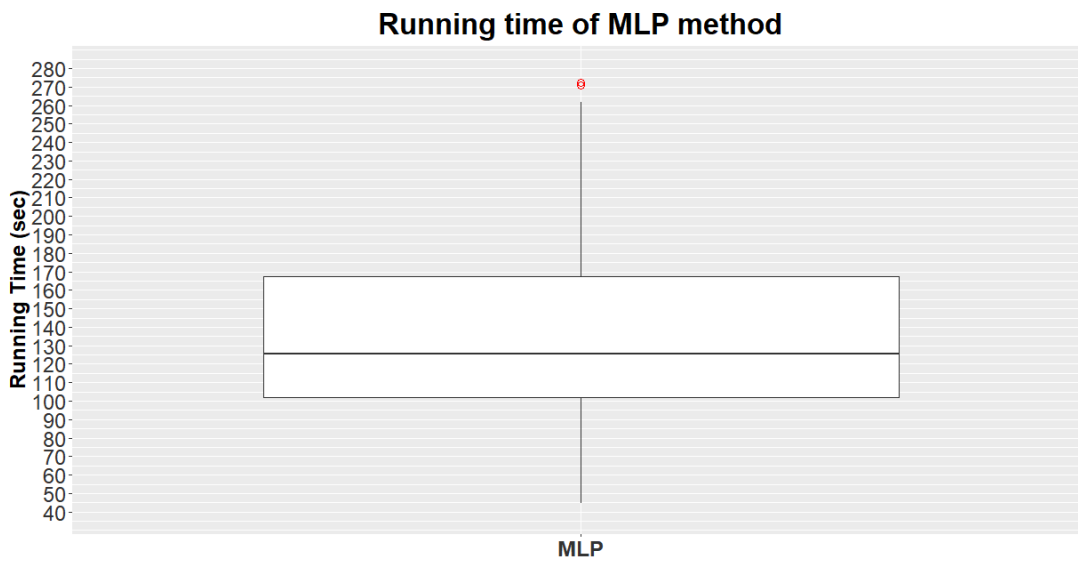
- **Order p:** 8
- **Deseasonalization:** True

- **Detrending:** False
- **Scaling:** True
- **Regularization:** True
- **Layers:** 1
- **Nodes:** 100
- **Batch Size:** 64

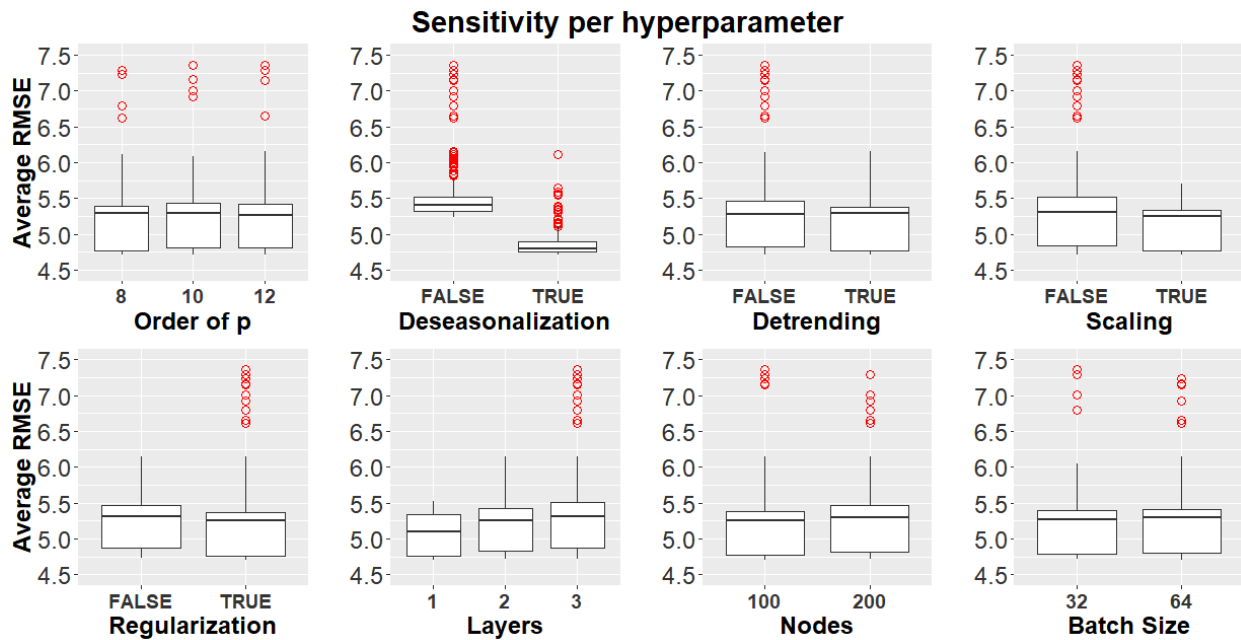
με μέσο RMSE 4.71, αριθμό υπερπαραμέτρων 2112 και χρόνο εκπαίδευσης 90.36 δευτερόλεπτα. Στο Διάγραμμα 80 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.71 – 7.36). Στο Διάγραμμα 81 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο Multi-layer Perceptron (MLP) ο οποίος εκτείνεται από 44.63 δευτερόλεπτα έως 272.42 δευτερόλεπτα και μέση τιμή περί τα 136.05 δευτερόλεπτα.



**Διάγραμμα 80** Ευαισθησία της μεθόδου Multi-layer Perceptron (MLP).



**Διάγραμμα 81** Χρόνος ανά παραμετροποίηση για την μέθοδο Multi-layer Perceptron (MLP).



**Διάγραμμα 82** Ευαισθησία της μεθόδου Multi-layer Perceptron (MLP) ανά υπερπαραμέτρο.

Στο Διάγραμμα 82 φαίνεται η ευαισθησία της μεθόδου Multi-layer Perceptron (MLP) ανά υπερπαραμέτρο υπό μορφή boxplot. Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
window         1   0.07    0.07    1.164  0.2812
Layers         1   6.33    6.33  102.188 < 2e-16 ***
Nodes         1   0.27    0.27    4.302  0.0385 *
Batch         1   0.00    0.00    0.037  0.8467
Deseasonalization 1  62.29   62.29 1005.206 < 2e-16 ***
Detrending     1   0.96    0.96   15.433 9.6e-05 ***
Scaling        1   5.65    5.65   91.138 < 2e-16 ***
Regularization 1   0.07    0.07    1.083  0.2985
Residuals     567  35.14    0.06
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Layers, Deseasonalization, Detrending και Scaling. Οι συντελεστές από την ανάλυση φαίνονται στον Πίνακα 35.

**Πίνακας 35** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Multi-layer Perceptron (MLP).

<b>(Intercept)</b>	5.2721098005
<b>Window</b>	0.0068512280
<b>Layers</b>	0.1284145308
<b>Nodes</b>	0.0004302464
<b>Batch</b>	0.0001254166
<b>DeseasonalizationTRUE</b>	-0.6576976165
<b>DetrendingTRUE</b>	-0.0814928632
<b>ScalingTRUE</b>	-0.1980382826
<b>RegularizationTRUE</b>	-0.0215880023

Στον Πίνακα 36 από τον συντελεστή για κάθε υπερπαραμέτρο εξάγονται τα εξής συμπεράσματα. Αρχικά, όσο αυξάνεται ο αριθμός των επιπέδων αυξάνεται το μέσο RMSE. Ωστόσο, η πιο καίρια υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την



μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο διάγραμμα 82. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.52, ενώ με εποχική αποσύνθεση είναι 4.86 (βλέπε και διάγραμμα 81). Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο.

**Πίνακας 36** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Multi-layer Perceptron (MLP).

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.414	3.460	3.487	3.501	3.516	3.537	3.555	3.570	3.567	3.555	3.575	3.572
<b>RMSE</b>	4.636	4.681	4.701	4.702	4.726	4.761	4.770	4.789	4.784	4.762	4.783	4.775
<b>MAPE</b>	11.526	11.725	11.888	11.954	12.042	12.126	12.187	12.232	12.222	12.206	12.289	12.271
<b>sMAPE</b>	11.474	11.654	11.790	11.861	11.941	12.021	12.073	12.124	12.127	12.103	12.179	12.172
<b>MASE</b>	0.703	0.675	0.677	0.676	0.676	0.675	0.674	0.670	0.662	0.650	0.646	0.640

**Πίνακας 37** Συνοπτικά μέτρα σφάλματος για την μέθοδο Multi-layer Perceptron (MLP).

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.466	3.544	3.567	3.526
<b>RMSE</b>	4.680	4.761	4.776	4.739
<b>MAPE</b>	11.773	12.147	12.247	12.056
<b>sMAPE</b>	11.695	12.040	12.145	11.960
<b>MASE</b>	0.683	0.674	0.649	0.669

**Πίνακας 38** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Multi-layer Perceptron (MLP).

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
78364.018 sec	57.294 sec	9.396 sec	78430.708 sec	33233	2.688

Στον Πίνακα 36 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 37 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

Η μέθοδος Multi-layer Perceptron (MLP) είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 38 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 78430 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο αναφοράς. Οι σχετικές βελτιώσεις του σφάλματος δίνονται στους Πίνακες 39 και 40. Ενδεικτικά να αναφέρουμε ότι έχει επιτευχθεί μείωση στο μέσο RMSE κατά 27.6%.

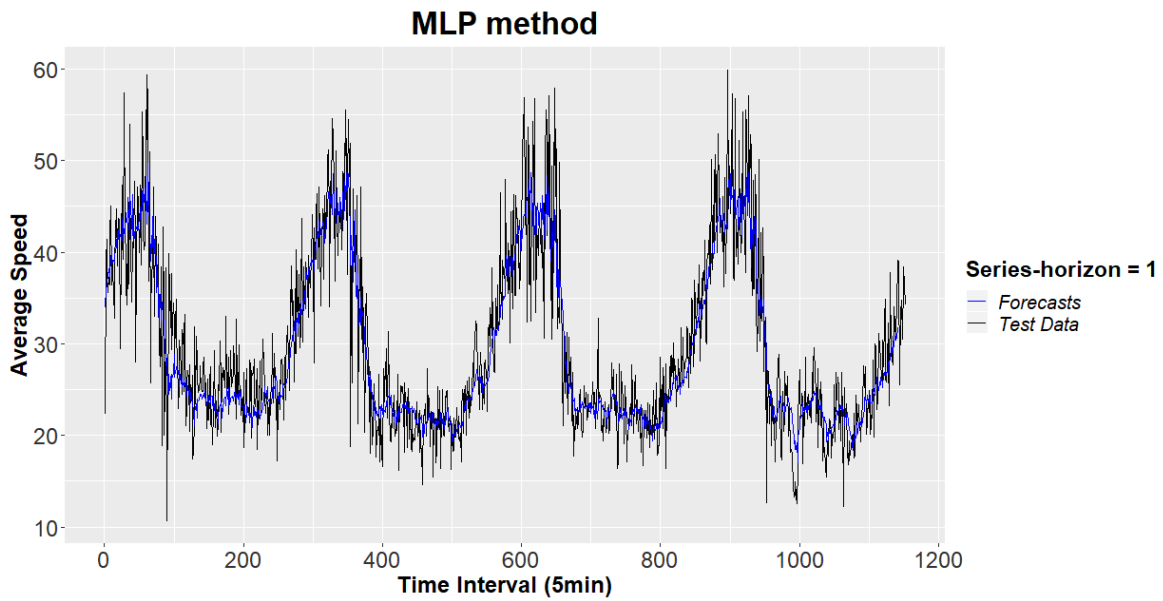
**Πίνακας 39** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Multi-layer Perceptron (MLP) ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	28.8	27.5	26.9	26.8	26.5	26.1	25.7	25.4	25.5	25.6	25.1	25.3
<b>RMSE</b>	29.5	28.4	28.1	28.1	27.8	27.3	27.1	26.8	26.9	27.2	26.8	27.0
<b>MAPE</b>	39.1	38.0	37.1	36.8	36.4	36.0	35.7	35.4	35.5	35.5	35.0	35.1
<b>sMAPE</b>	32.1	30.9	30.0	29.7	29.3	28.8	28.5	28.3	28.2	28.3	27.8	27.9
<b>MASE</b>	28.8	27.5	26.9	26.8	26.6	26.1	25.7	25.5	25.5	25.6	25.1	25.3

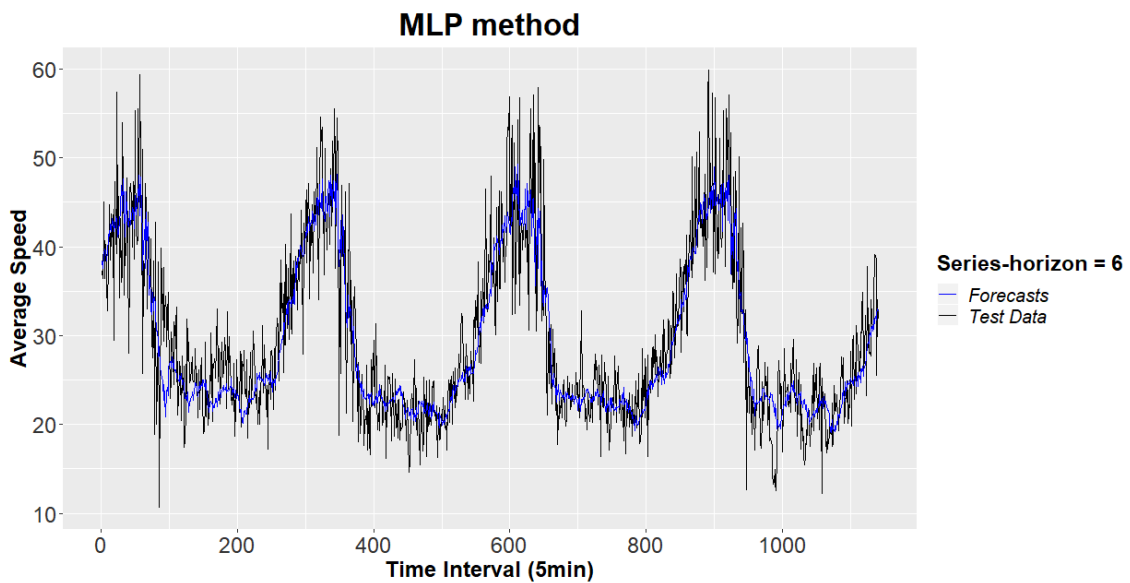
**Πίνακας 40** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Multi-layer Perceptron (MLP) ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	27.5	25.9	25.4	26.3
<b>RMSE</b>	28.5	27.3	27.0	27.6
<b>MAPE</b>	37.8	35.9	35.3	36.3
<b>sMAPE</b>	30.7	28.7	28.1	29.2
<b>MASE</b>	27.5	25.9	25.4	26.3

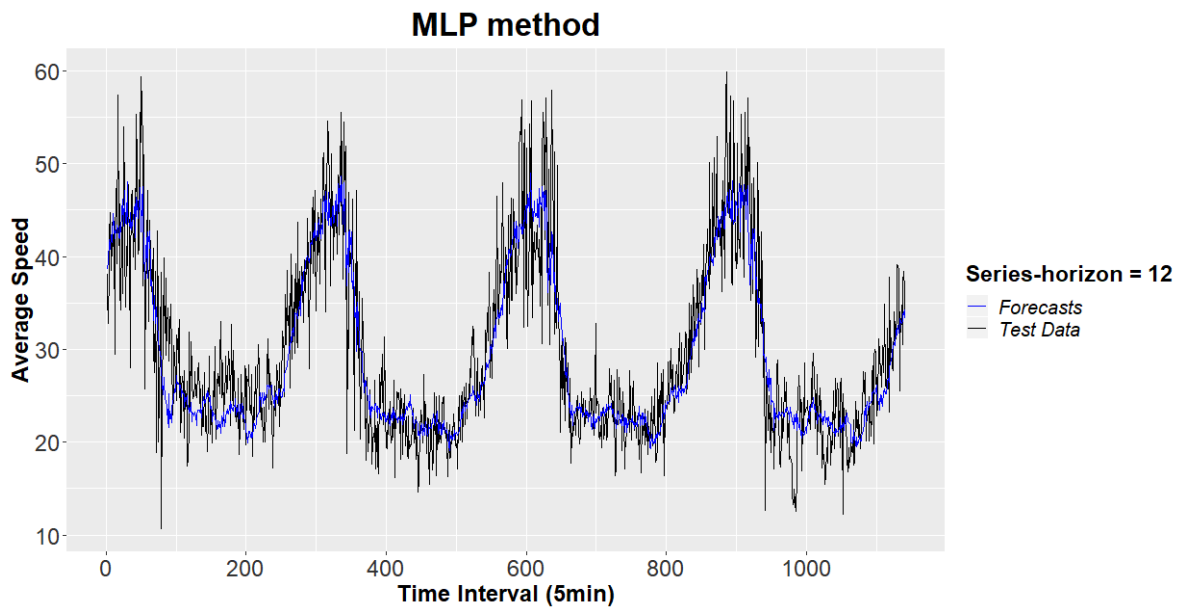
Στα Διαγράμματα 83, 84 και 85 δίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Παρατηρείται η υποτίμηση της πρόβλεψης του επιπέδου της χρονοσειράς για την 1η μέρα του συνόλου ελέγχου, γεγονός που φαίνεται να εντείνεται καθώς αυξάνεται η τιμή του ορίζοντα πρόβλεψης. Τέλος, από το διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 86) των καταλοίπων συνάγεται ότι έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



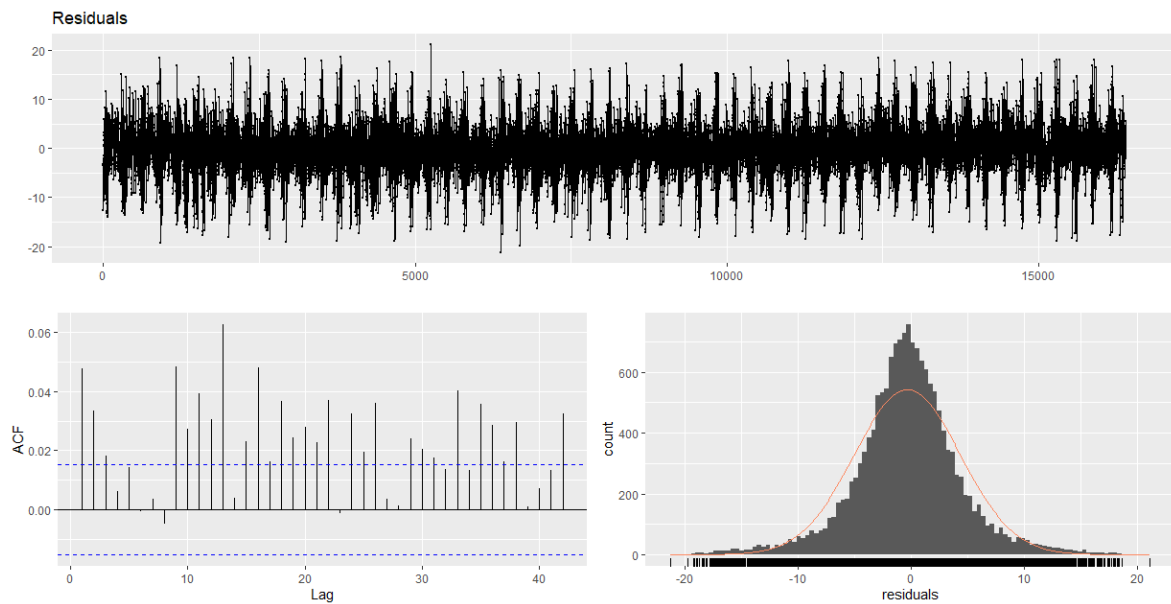
**Διάγραμμα 83** Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 84** Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 85** Προβλέψεις της μεθόδου Multi-layer Perceptron (MLP) για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



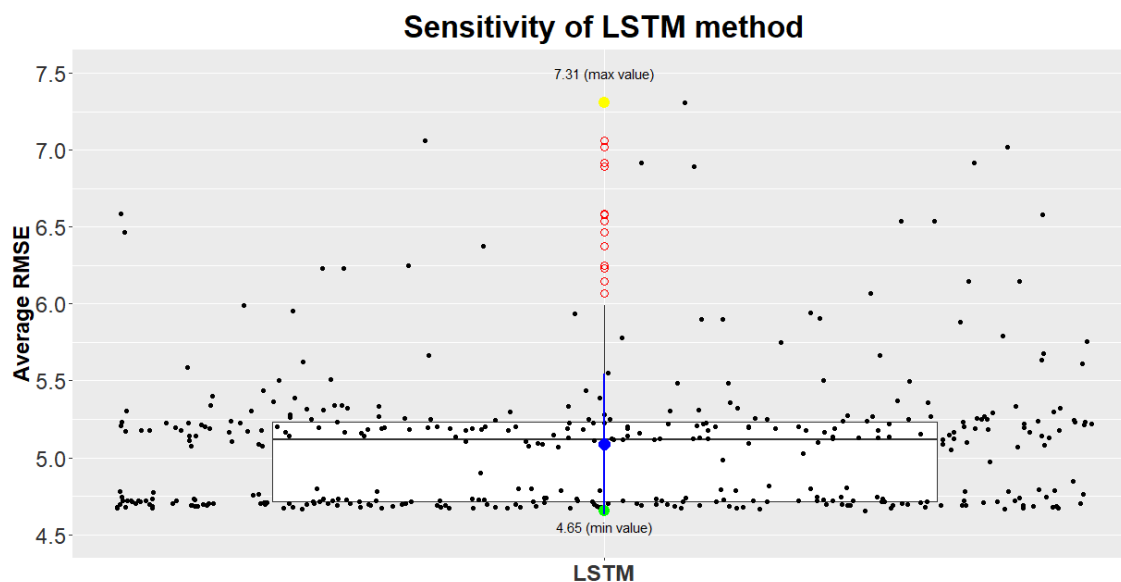
**Διάγραμμα 86** Γραφικός έλεγχος των καταλοίπων της μεθόδου Multi-layer Perceptron (MLP).

### 6.3.2 Μέθοδος Long-Short Term Memory (LSTM)

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** False
- **Layers:** 1
- **Nodes:** 200
- **Batch Size:** 64

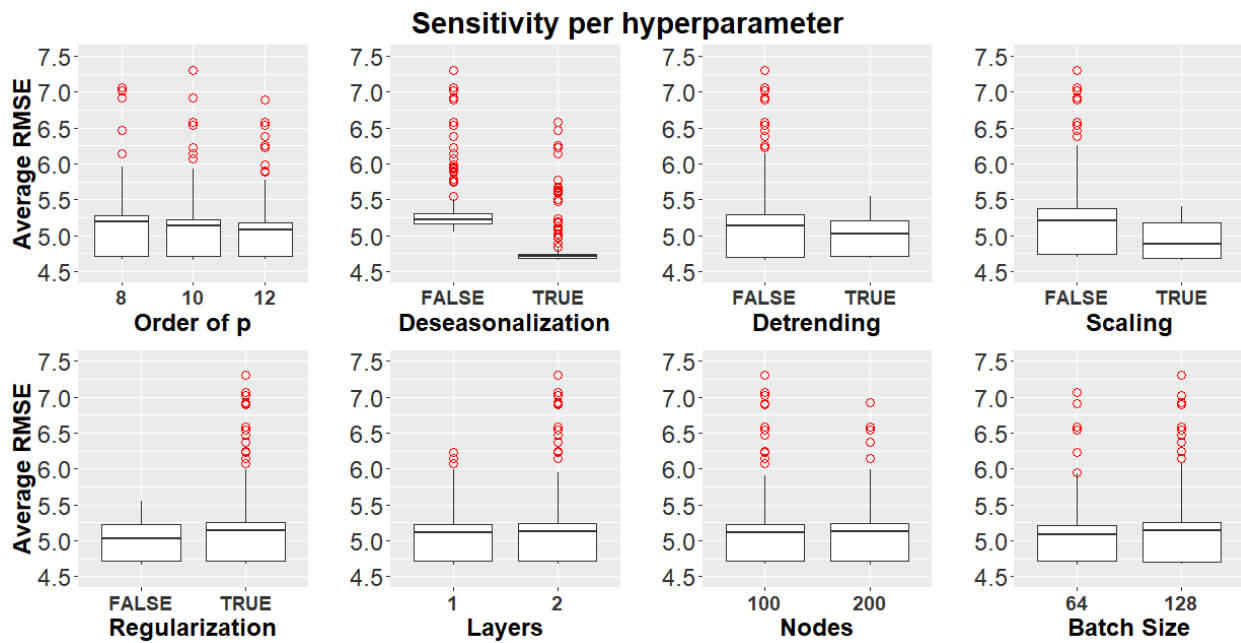
με μέσο RMSE 4.65, αριθμό υπερπαραμέτρων 182912 και χρόνο εκπαίδευσης 405.35 δευτερόλεπτα. Στο Διάγραμμα 87 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.65 – 7.30). Στο Διάγραμμα 88 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο LSTM ο οποίος εκτείνεται από 64.35 δευτερόλεπτα έως 2048.16 δευτερόλεπτα και μέση τιμή περί τα 369.02 δευτερόλεπτα.



Διάγραμμα 87 Ευαισθησία της μεθόδου LSTM.



Διάγραμμα 88 Χρόνος ανά παραμετροποίηση για την μέθοδο LSTM.



Διάγραμμα 89 Ευαισθησία της μεθόδου LSTM ανά υπερπαραμέτρο.

Στο Διάγραμμα 89 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαράμετρο υπό μορφή boxplot. Διεξάχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq F value    Pr(>F)
window         1   0.10   0.102    1.165    0.2811
Layers         1   0.55   0.547    6.228    0.0130 *
Nodes          1   0.11   0.106    1.209    0.2723
Batch          1   0.29   0.291    3.320    0.0692 .
Deseasonalization 1 27.05 27.053 308.233 < 2e-16 ***
Detrending     1   4.69   4.694   53.481 1.59e-12 ***
Scaling        1   9.34   9.336  106.368 < 2e-16 ***
Regularization 1   4.73   4.725   53.837 1.36e-12 ***
Residuals     375 32.91   0.088
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαράμετροι είναι οι Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 41.

**Πίνακας 41** Συντελεστές της ανάλυσης ANOVA για την μέθοδο LSTM.

<b>(Intercept)</b>	5.4626590777
<b>Window</b>	-0.0099928525
<b>Layers</b>	0.0754552295
<b>Nodes</b>	-0.0003324205
<b>Batch</b>	0.0008608077
<b>DeseasonalizationTRUE</b>	-0.5308479177
<b>DetrendingTRUE</b>	-0.2211212392
<b>ScalingTRUE</b>	-0.3118427472
<b>RegularizationTRUE</b>	0.2218567075

Παρατηρούμε στον Πίνακα 41 ότι η πιο σημαντική υπερπαράμετρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 89. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.35, ενώ με εποχική αποσύνθεση είναι 4.82. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο.

Στον Πίνακα 42 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 43 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 42** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο LSTM.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.372	3.405	3.431	3.474	3.489	3.497	3.516	3.520	3.524	3.510	3.520	3.521
<b>RMSE</b>	4.572	4.593	4.630	4.668	4.680	4.682	4.700	4.700	4.699	4.675	4.682	4.687
<b>MAPE</b>	11.200	11.354	11.517	11.699	11.788	11.853	11.926	11.950	11.968	11.951	11.989	12.015
<b>sMAPE</b>	11.321	11.466	11.596	11.778	11.851	11.895	11.963	11.978	11.995	11.969	12.009	12.021
<b>MASE</b>	0.695	0.664	0.667	0.670	0.671	0.667	0.667	0.661	0.654	0.642	0.636	0.630

**Πίνακας 43** Συνοπτικά μέτρα σφάλματος για την μέθοδο LSTM.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.421	3.506	3.519	3.482
<b>RMSE</b>	4.616	4.690	4.686	4.664
<b>MAPE</b>	11.442	11.879	11.981	11.767
<b>sMAPE</b>	11.540	11.922	11.999	11.820
<b>MASE</b>	0.674	0.667	0.641	0.660

**Πίνακας 44** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
141704.094 sec	467.239 sec	16.84 sec	142188.175 sec	60249	2.706

Η μέθοδος LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 44 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 60249 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του



σφάλματος δίνονται στους Πίνακες 45 και 46. Η μείωση που επιτεύχθηκε για το μέσο RMSE είναι της τάξεως του 28.7%.

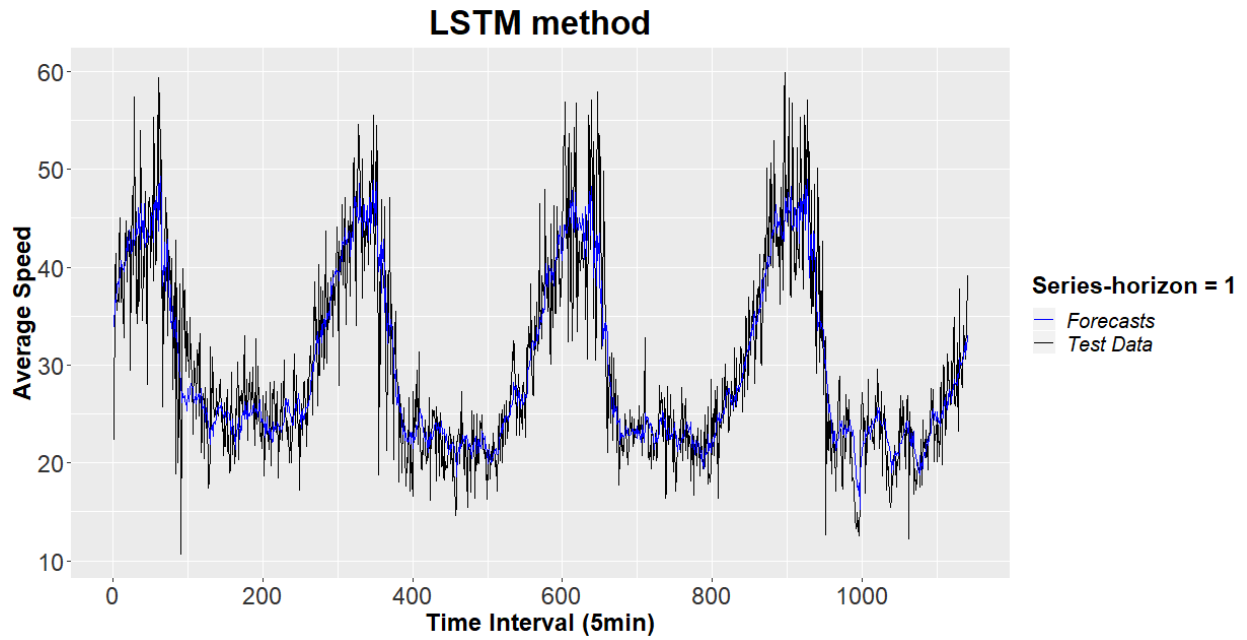
**Πίνακας 45** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.6	28.7	28.1	27.3	27.1	26.9	26.5	26.5	26.4	26.5	26.3	26.3
<b>RMSE</b>	30.4	29.7	29.2	28.7	28.5	28.5	28.2	28.2	28.2	28.5	28.4	28.3
<b>MAPE</b>	40.9	40.0	39.0	38.2	37.8	37.4	37.0	36.9	36.8	36.8	36.6	36.5
<b>sMAPE</b>	33.0	32.0	31.2	30.2	29.8	29.6	29.2	29.1	29.0	29.1	28.8	28.8
<b>MASE</b>	29.6	28.7	28.1	27.4	27.1	26.9	26.5	26.5	26.4	26.5	26.2	26.4

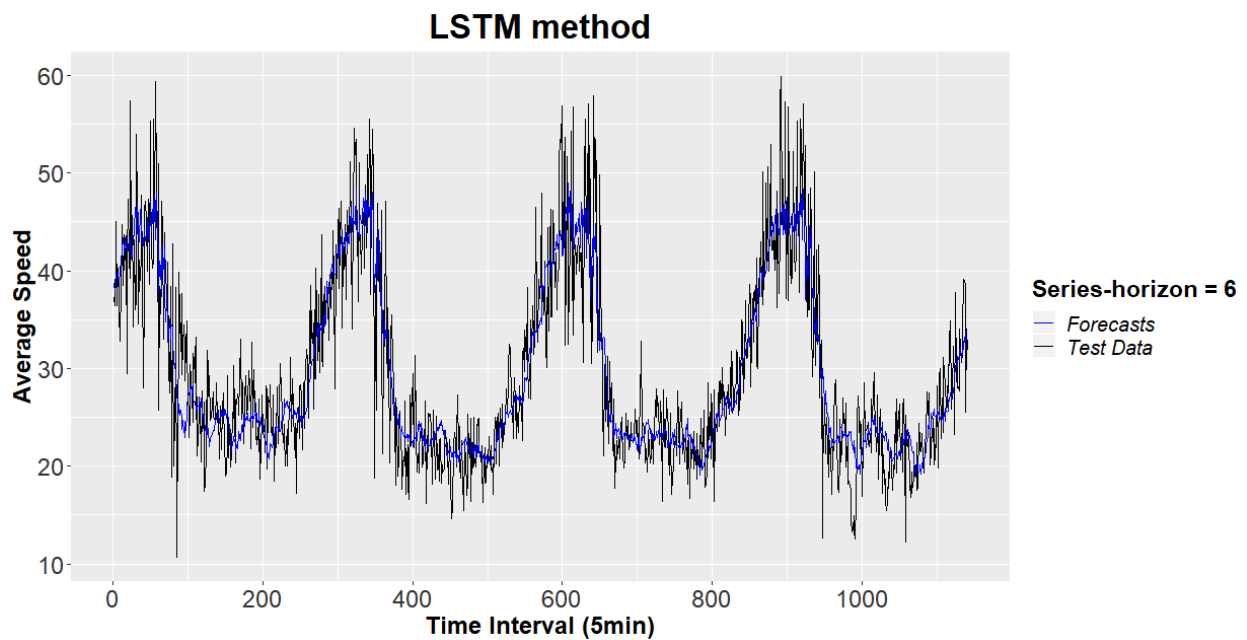
**Πίνακας 46** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο LSTM ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.4	26.8	26.4	27.2
<b>RMSE</b>	29.5	28.3	28.4	28.7
<b>MAPE</b>	39.5	37.3	36.7	37.8
<b>sMAPE</b>	31.6	29.4	28.9	30.0
<b>MASE</b>	28.5	26.7	26.4	27.2

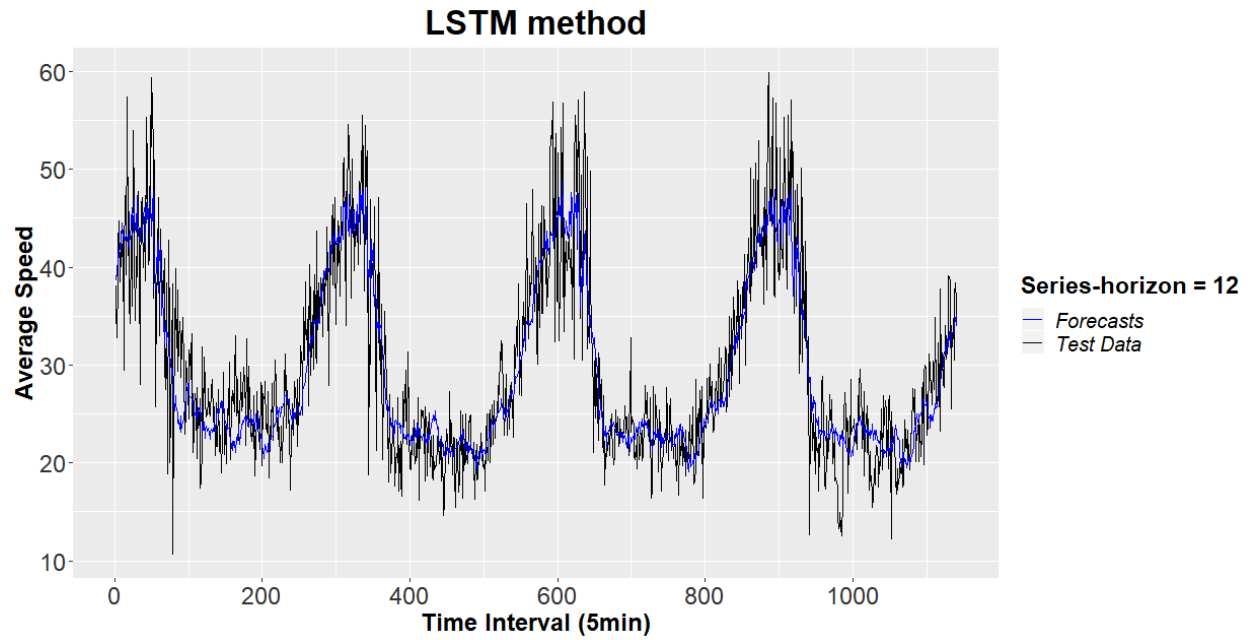
Στα Διαγράμματα 90, 91 και 92 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Παρατηρείται και εδώ η υποτίμηση της πρόβλεψης του επιπέδου της χρονοσειράς για την 1<sup>η</sup> μέρα του συνόλου ελέγχου, γεγονός που φαίνεται να εντείνεται καθώς αυξάνεται η τιμή του ορίζοντα πρόβλεψης. Τέλος, στο Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 93) των καταλοίπων φαίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



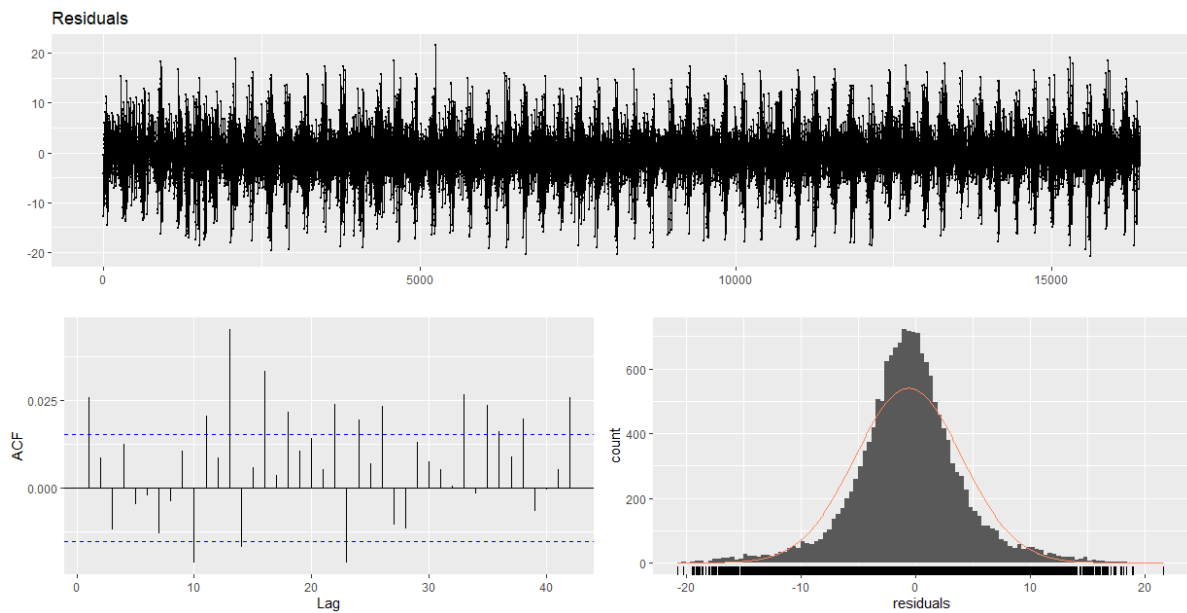
**Διάγραμμα 90** Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 91** Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 92** Προβλέψεις της μεθόδου LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



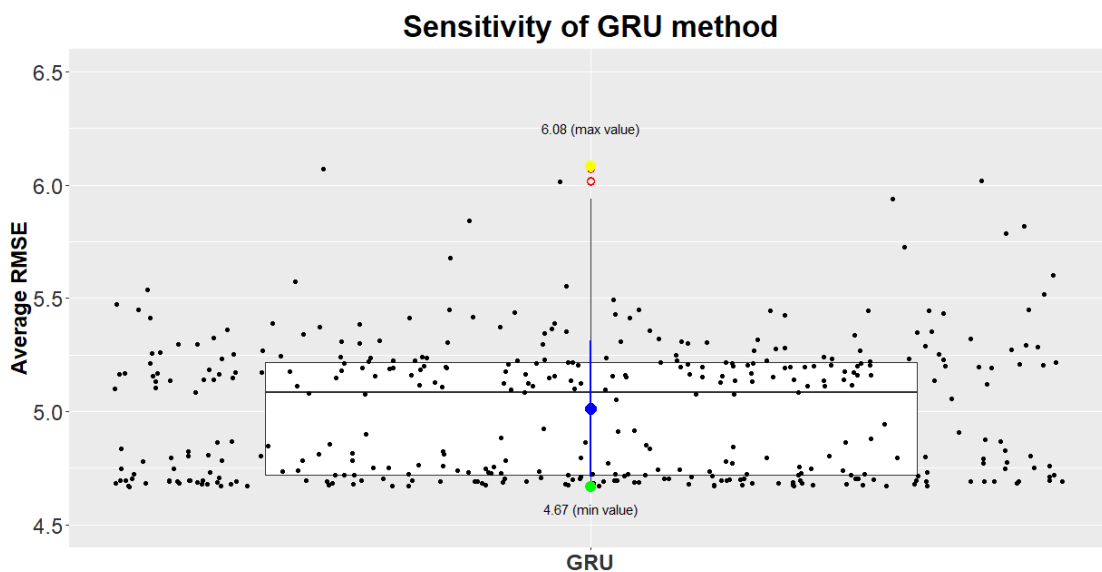
**Διάγραμμα 93** Γραφικός έλεγχος των καταλοίπων της μεθόδου LSTM.

### 6.3.3 Μέθοδος Gated Recurrent Unit (GRU)

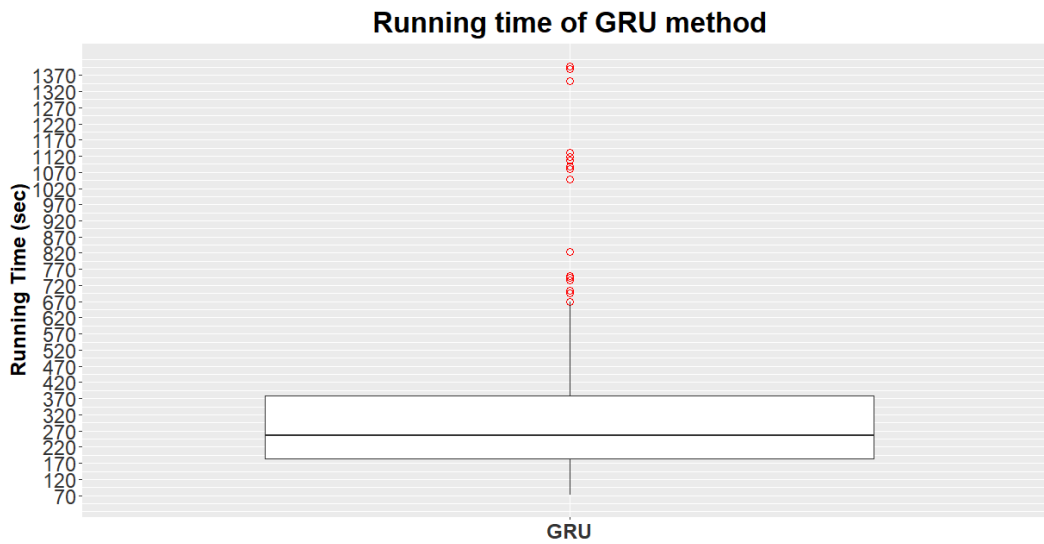
Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** True
- **Layers:** 1
- **Nodes:** 100
- **Batch Size:** 64

με μέσο RMSE 4.67, αριθμό παραμέτρων 41912 και χρόνο εκπαίδευσης 320.64 δευτερόλεπτα. Στο Διάγραμμα 94 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.67 – 6.08). Φαίνεται καθαρά ότι η μέθοδος GRU έχει σημαντικά μικρότερη ευαισθησία από τις 2 προηγούμενες μεθόδους που έχουν εξεταστεί μέχρι στιγμής.

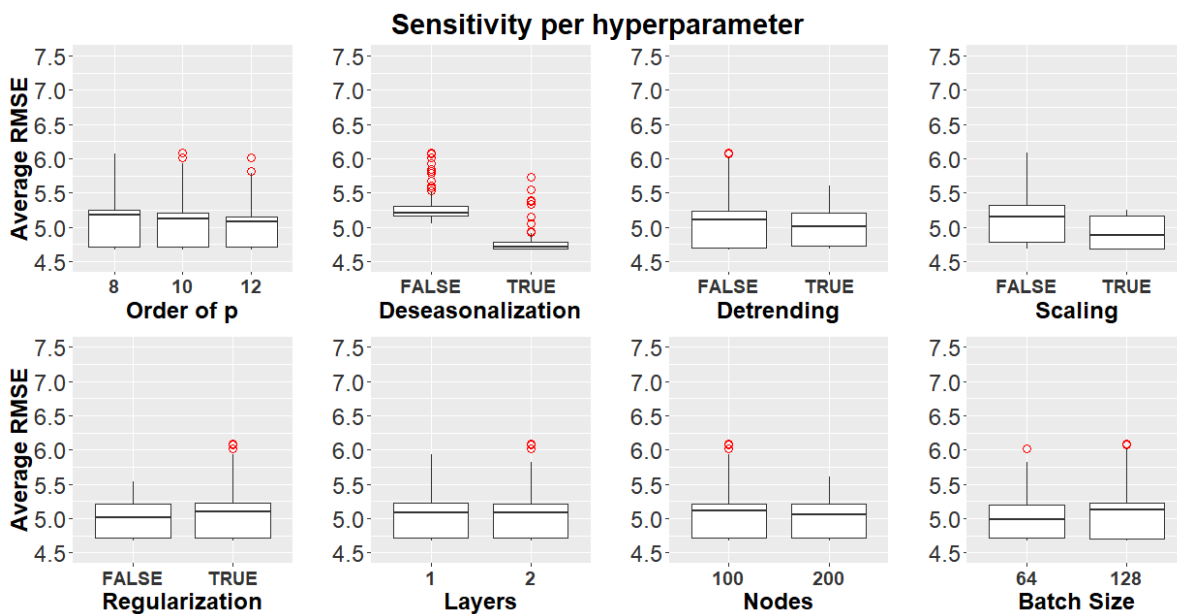


Διάγραμμα 94 Ευαισθησία της μεθόδου GRU.



**Διάγραμμα 95** Χρόνος ανά παραμετροποίηση για την μέθοδο GRU.

Στο Διάγραμμα 95 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο GRU ο οποίος εκτείνεται από 72.91 δευτερόλεπτα έως 1398.48 δευτερόλεπτα και μέση τιμή περί τα 314.78 δευτερόλεπτα.



**Διάγραμμα 96** Ευαισθησία της μεθόδου GRU ανά υπερπαραμέτρο.

Στο Διάγραμμα 96 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαράμετρο υπό μορφή boxplot. Διεξήχθη επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
Window         1  0.080   0.080    4.459  0.0354 *
Layers         1  0.018   0.018    1.015  0.3144
Nodes         1  0.084   0.084    4.672  0.0313 *
Batch         1  0.041   0.041    2.276  0.1323
Deseasonalization 1 24.725 24.725 1380.523 < 2e-16 ***
Detrending     1  0.174   0.174    9.689  0.0020 **
Scaling        1  2.675   2.675  149.345 < 2e-16 ***
Regularization 1  0.323   0.323   18.053 2.71e-05 ***
Residuals     375  6.716   0.018
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαράμετροι είναι οι Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 47.

**Πίνακας 47** Συντελεστές της ανάλυσης ANOVA για την μέθοδο GRU.

<b>(Intercept)</b>	5.4232801215
<b>Window</b>	-0.0088310015
<b>Layers</b>	0.0137586394
<b>Nodes</b>	-0.0002952430
<b>Batch</b>	0.0003219524
<b>DeseasonalizationTRUE</b>	-0.5074987409
<b>DetrendingTRUE</b>	-0.0425168180
<b>ScalingTRUE</b>	-0.1669199395
<b>RegularizationTRUE</b>	0.0580352183

Από τον Πίνακα 47 συμπεραίνεται ότι η πιο σημαντική υπερπαράμετρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό

φαίνεται και σχηματικά στο Διάγραμμα 96. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.27, ενώ με εποχική αποσύνθεση είναι 4.76. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο.

Στον Πίνακα 48 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκε ανά ορίζοντα πρόβλεψης. Στον Πίνακα 49 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 48** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο GRU.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.364	3.403	3.427	3.458	3.484	3.504	3.521	3.525	3.537	3.524	3.530	3.540
<b>RMSE</b>	4.568	4.595	4.619	4.646	4.672	4.688	4.702	4.709	4.715	4.699	4.707	4.721
<b>MAPE</b>	11.202	11.371	11.513	11.646	11.770	11.872	11.942	11.961	11.992	11.987	12.012	12.058
<b>sMAPE</b>	11.309	11.463	11.583	11.722	11.832	11.913	11.973	11.992	12.029	12.011	12.035	12.067
<b>MASE</b>	0.693	0.664	0.666	0.667	0.670	0.669	0.668	0.662	0.656	0.645	0.638	0.634

**Πίνακας 49** Συνοπτικά μέτρα σφάλματος των συνοπτικών μέτρων σφάλματος για την μέθοδο GRU.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.413	3.509	3.533	3.485
<b>RMSE</b>	4.607	4.693	4.711	4.670
<b>MAPE</b>	11.433	11.887	12.012	11.777
<b>sMAPE</b>	11.519	11.928	12.036	11.828
<b>MASE</b>	0.672	0.667	0.643	0.661

**Πίνακας 50** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο GRU.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
120875.028 sec	330.769 sec	17.225 sec	121223.022 sec	51366	2.721

Από τα αποτελέσματα συμπεραίνεται ότι η μέθοδος GRU είναι καλύτερη από την μέθοδο Seasonal Naïve για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 50 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 51366 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 51 και 52. Ενδεικτικά, η μείωση στο μέσο RMSE είναι της τάξεως του 28.6%.

**Πίνακας 51** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο GRU ως προς την μέθοδο αναφοράς.

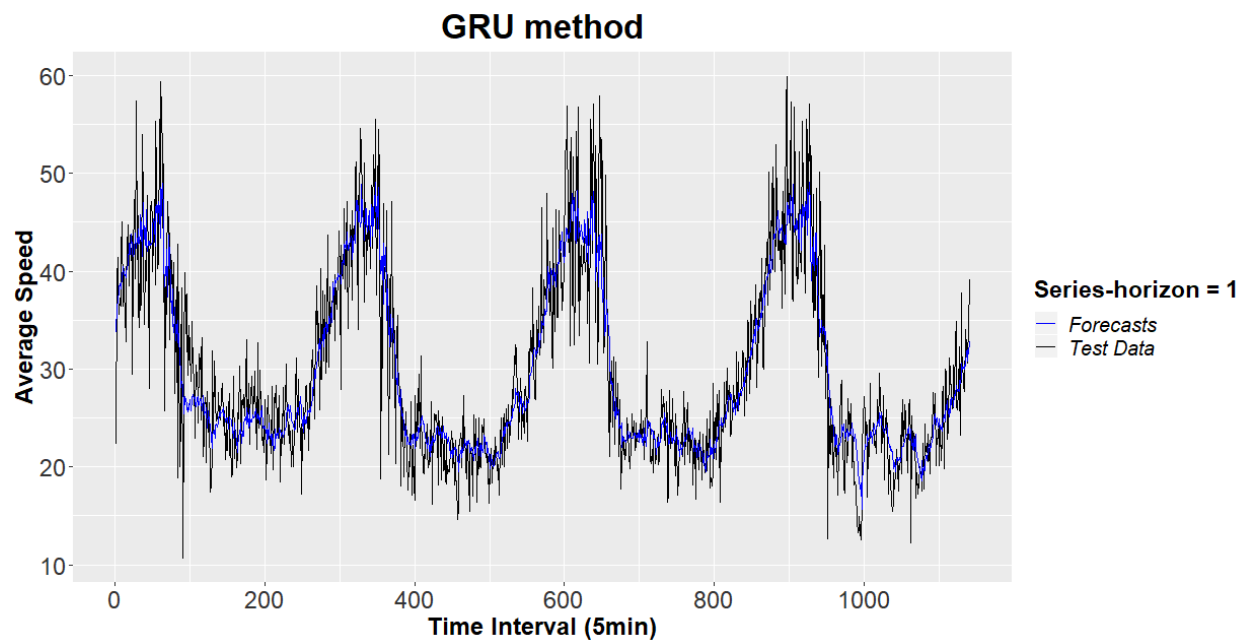
	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.8	28.8	28.2	27.7	27.2	26.8	26.4	26.4	26.1	26.2	26.1	25.9
<b>RMSE</b>	30.5	29.7	29.3	29.0	28.6	28.4	28.2	28.1	28.0	28.1	28.0	27.8
<b>MAPE</b>	40.9	39.9	39.1	38.5	37.8	37.3	37.0	36.9	36.7	36.6	36.5	36.3
<b>sMAPE</b>	33.1	32.0	31.2	30.5	30.0	29.5	29.1	29.1	28.8	28.8	28.6	28.5
<b>MASE</b>	29.8	28.7	28.2	27.7	27.2	26.8	26.4	26.4	26.1	26.2	26.0	26.0

**Πίνακας 52** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο GRU ως προς την μέθοδο αναφοράς.

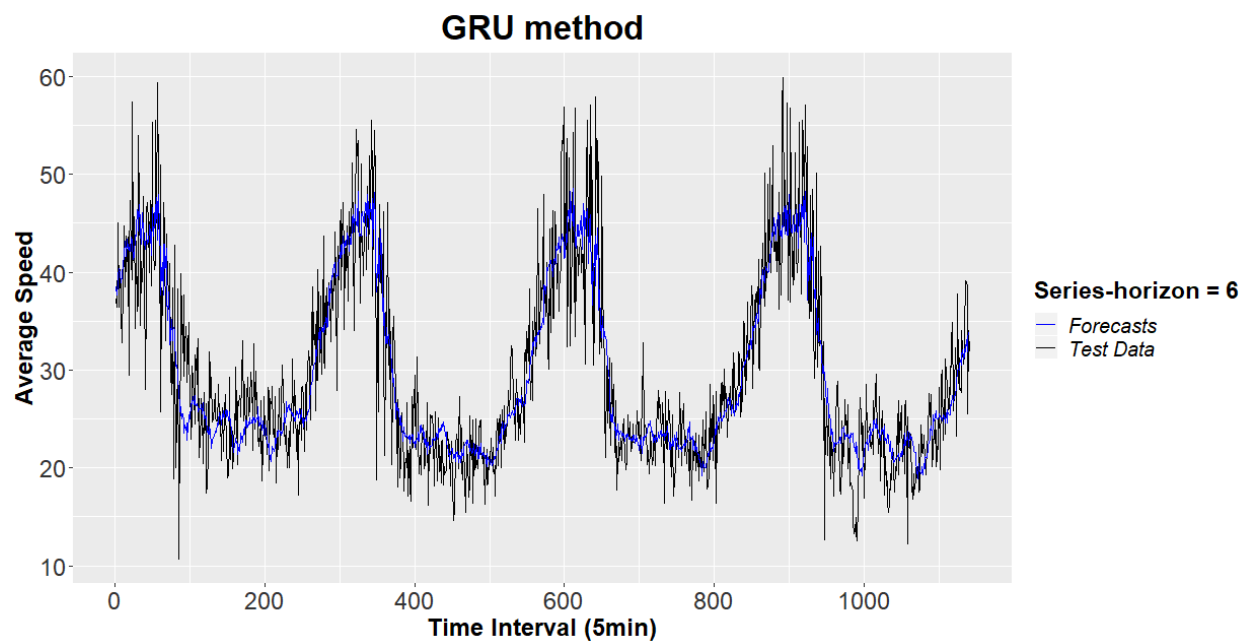
	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.6	26.7	26.1	27.1
<b>RMSE</b>	29.6	28.3	28.0	28.6
<b>MAPE</b>	39.6	37.2	36.5	37.8
<b>sMAPE</b>	31.7	29.4	28.7	29.9
<b>MASE</b>	28.6	26.7	26.1	27.1

Στα Διαγράμματα 97, 98 και 99 βλέπουμε ενδεικτικά τις προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Τέλος, από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 100) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.

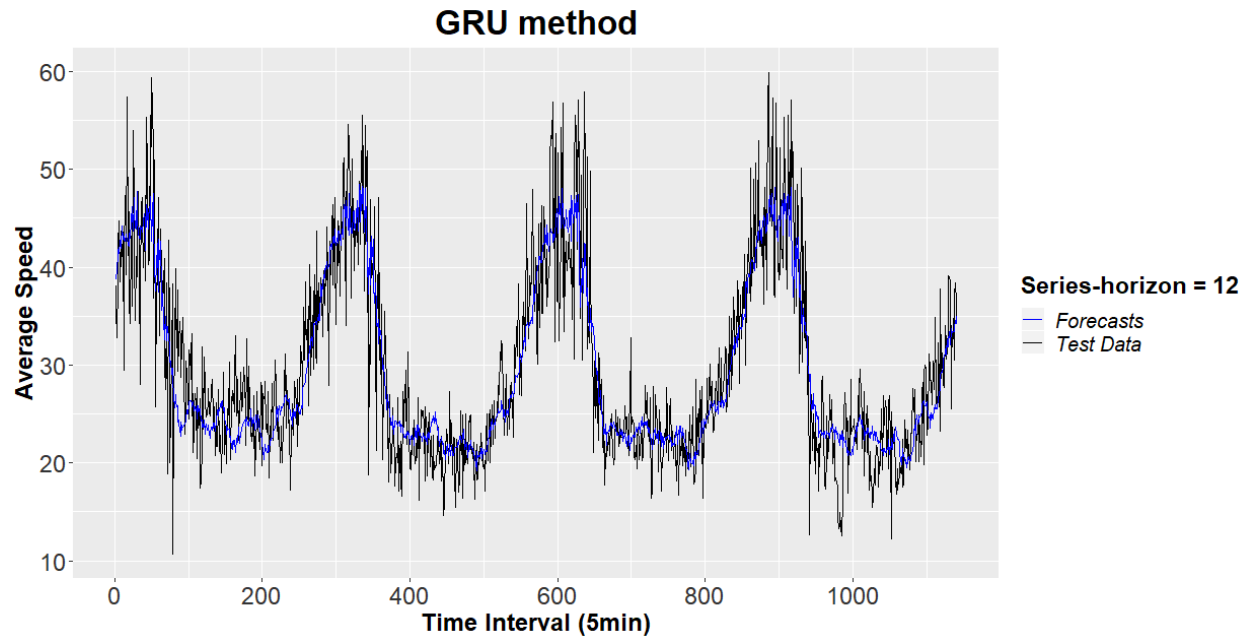




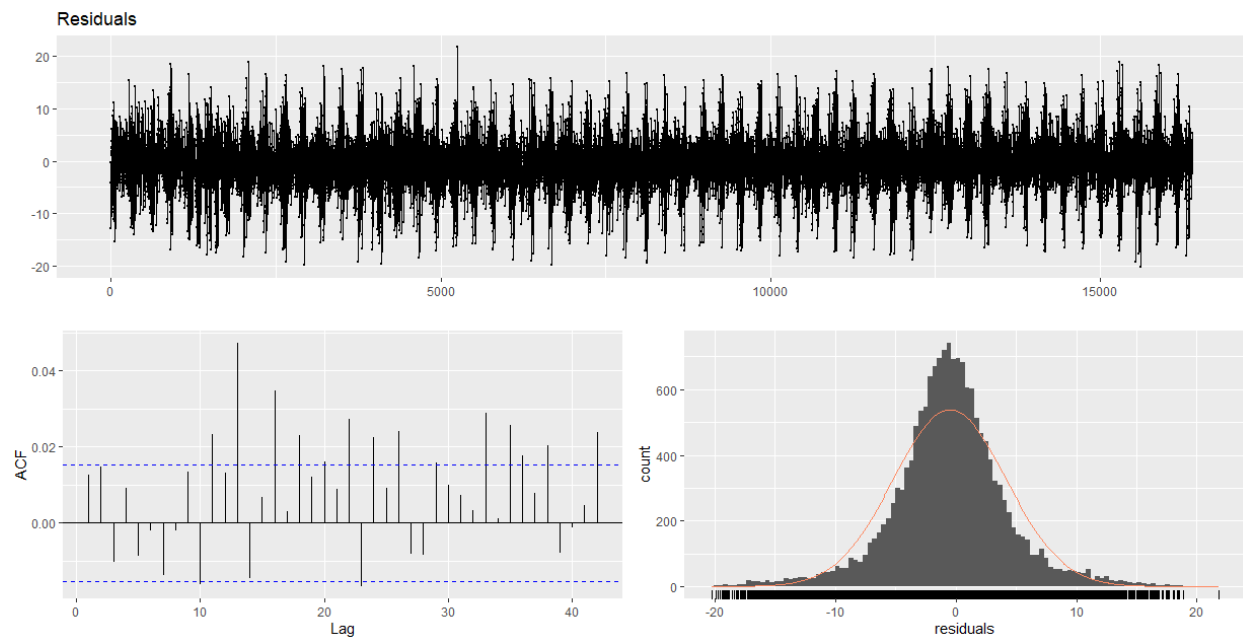
**Διάγραμμα 97** Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 98** Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 99** Προβλέψεις της μεθόδου GRU για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



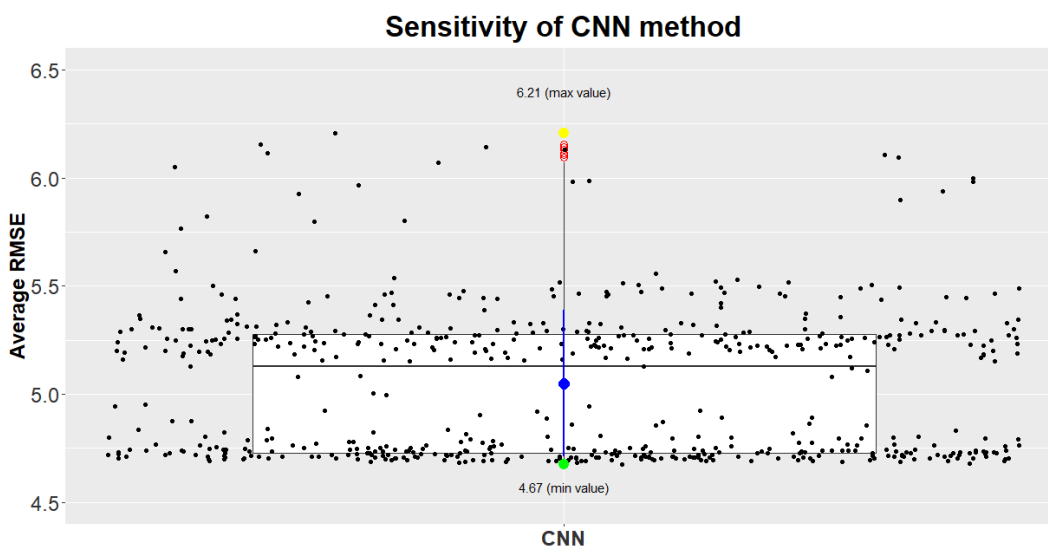
**Διάγραμμα 100** Γραφικός έλεγχος των καταλοίπων της μεθόδου GRU.

### 6.3.4 Μέθοδος Convolutional Neural Networks (CNN)

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

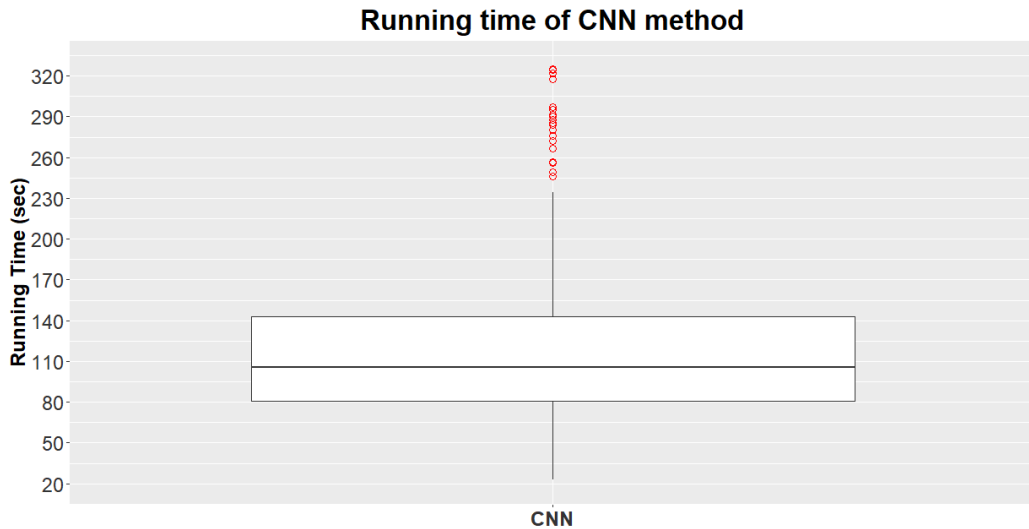
- **Order p:** 12
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** True
- **Block:** 1
- **Filters:** 32
- **Kernel:** 2
- **Batch Size:** 128

με μέσο RMSE 4.67, αριθμό παραμέτρων 16288 και χρόνο εκπαίδευσης 83.48 δευτερόλεπτα. Στο Διάγραμμα 101 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.67 – 6.21). Φαίνεται καθαρά ότι η μέθοδος CNN έχει σημαντικά μικρότερη ευαισθησία, όπως η μέθοδος GRU, σε σχέση με τις 2 πρώτες μεθόδους που εξετάστηκαν.

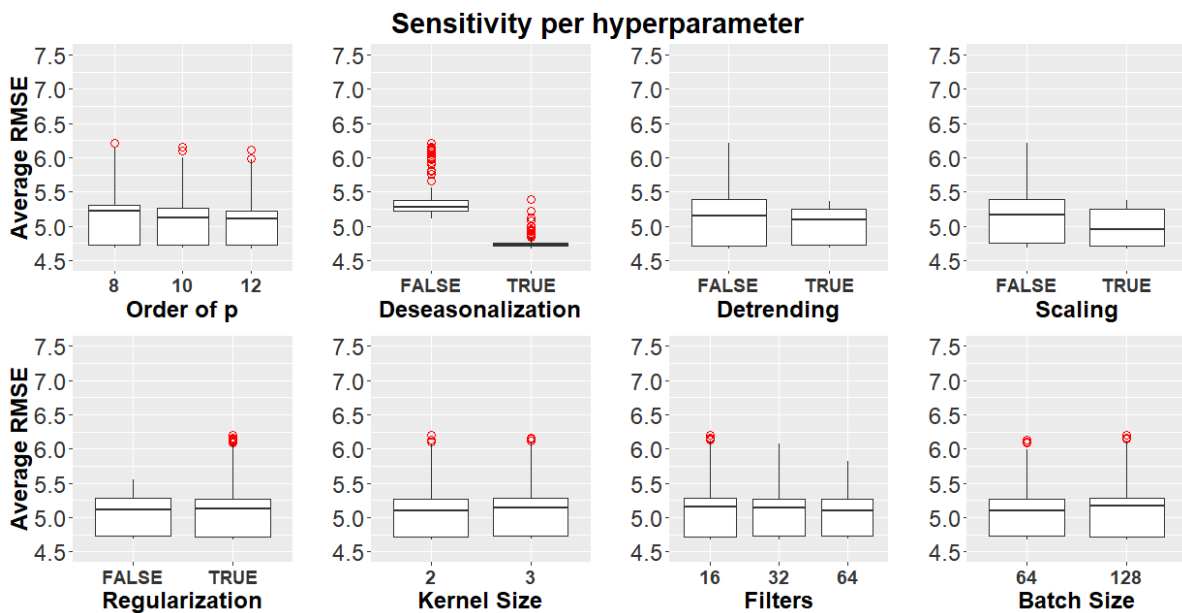


Διάγραμμα 101 Ευαισθησία της μεθόδου CNN.

Στο διάγραμμα 102 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο GRU ο οποίος εκτείνεται από 22.68 δευτερόλεπτα έως 324.96 δευτερόλεπτα και μέση τιμή περί τα 177.73 δευτερόλεπτα. Στο διάγραμμα 103 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαραμέτρο υπό μορφή boxplot. Διεξήχθηκε επίσης ανάλυση απονα για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων.



**Διάγραμμα 102** Χρόνος ανά παραμετροποίηση για την μέθοδο CNN.



**Διάγραμμα 103** Ευαισθησία της μεθόδου CNN ανά υπερπαραμέτρο.

Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
Window         1   0.22    0.22    12.266 0.000498 ***
Filters         1   0.16    0.16     8.785 0.003165 **
Kernels         1   0.02    0.02     1.258 0.262482
Batch           1   0.02    0.02     0.929 0.335661
Deseasonalization 1  51.04   51.04 2884.416 < 2e-16 ***
Detrending      1   1.26    1.26   71.461 2.4e-16 ***
Scaling         1   2.89    2.89  163.302 < 2e-16 ***
Regularization  1   0.19    0.19   10.882 0.001032 **
Residuals      567  10.03    0.02
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Window, Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 53.

**Πίνακας 53** Συντελεστές της ανάλυσης ANOVA για την μέθοδο CNN.

<b>(Intercept)</b>	5.5487123982
<b>Window</b>	-0.0118872642
<b>Filters</b>	-0.0008232290
<b>Kernels</b>	0.0124340098
<b>Batch</b>	0.0001669027
<b>DeseasonalizationTrue</b>	-0.5953591896
<b>DetrendingTrue</b>	-0.0937095524
<b>ScalingTrue</b>	-0.1416593266
<b>RegularizationTrue</b>	0.0365677313

Από τον Πίνακα 53 συμπεραίνεται ότι η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο διάγραμμα 103. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.35, ενώ με εποχική αποσύνθεση είναι 4.75. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο.

Στον Πίνακα 54 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 56 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 54** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο CNN.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.371	3.402	3.452	3.481	3.499	3.515	3.527	3.536	3.542	3.530	3.540	3.539
<b>RMSE</b>	4.576	4.598	4.655	4.679	4.695	4.709	4.716	4.724	4.729	4.716	4.720	4.721
<b>MAPE</b>	11.254	11.409	11.655	11.786	11.895	11.968	12.027	12.054	12.056	12.032	12.076	12.084
<b>sMAPE</b>	11.315	11.462	11.669	11.804	11.888	11.957	12.004	12.034	12.053	12.025	12.067	12.066
<b>MASE</b>	0.694	0.664	0.671	0.672	0.673	0.671	0.669	0.664	0.657	0.646	0.639	0.634

**Πίνακας 55** Συνοπτικά μέτρα σφάλματος για την μέθοδο CNN.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.426	3.519	3.537	3.494
<b>RMSE</b>	4.627	4.711	4.722	4.687
<b>MAPE</b>	11.526	11.986	12.062	11.858
<b>sMAPE</b>	11.563	11.971	12.053	11.862
<b>MASE</b>	0.675	0.669	0.644	0.663

**Πίνακας 56** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο CNN.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
67815.348 sec	79.205 sec	12.002 sec	67906.555 sec	28774	2.669

Από τα αποτελέσματα συμπεραίνεται ότι η μέθοδος CNN είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 56 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 28774 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την

μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους πίνακες 57 και 58. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.4%.

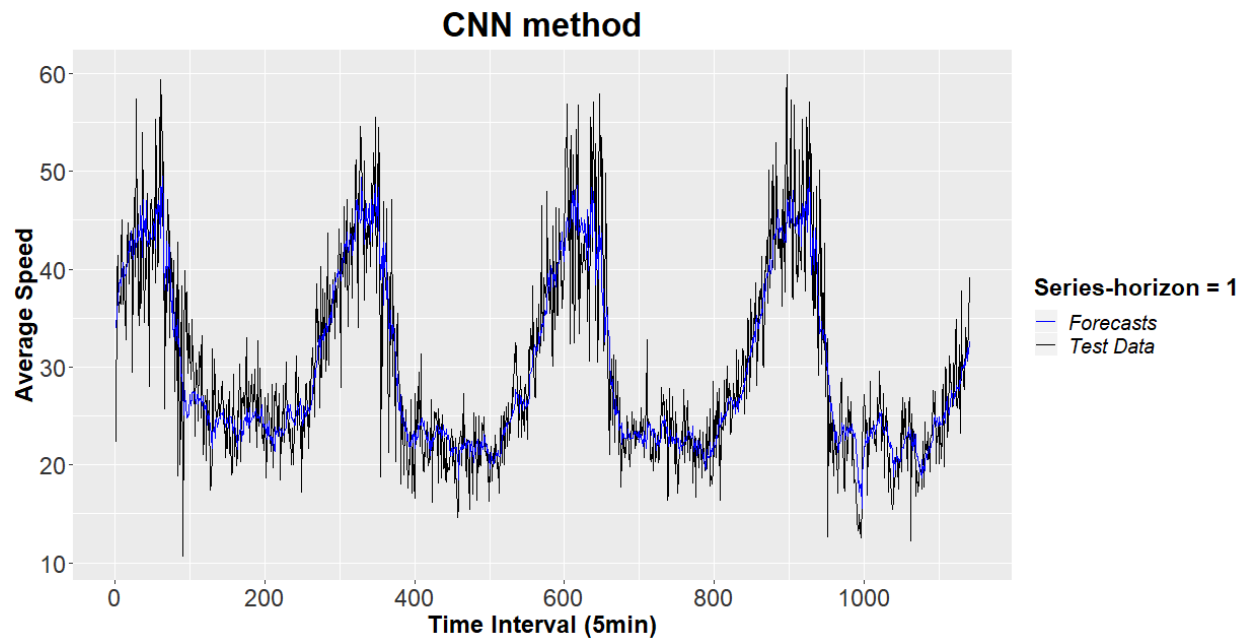
**Πίνακας 57** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο CNN ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.7	28.8	27.7	27.2	26.9	26.5	26.3	26.1	26.0	26.1	25.8	26.0
<b>RMSE</b>	30.4	29.7	28.8	28.5	28.3	28.1	28.0	27.8	27.8	27.9	27.8	27.8
<b>MAPE</b>	40.6	39.7	38.3	37.7	37.2	36.8	36.5	36.4	36.4	36.4	36.1	36.1
<b>sMAPE</b>	33.1	32.0	30.7	30.1	29.6	29.2	29.0	28.8	28.7	28.8	28.5	28.5
<b>MASE</b>	29.6	28.7	27.7	27.2	26.9	26.5	26.3	26.2	26.0	26.1	25.8	26.0

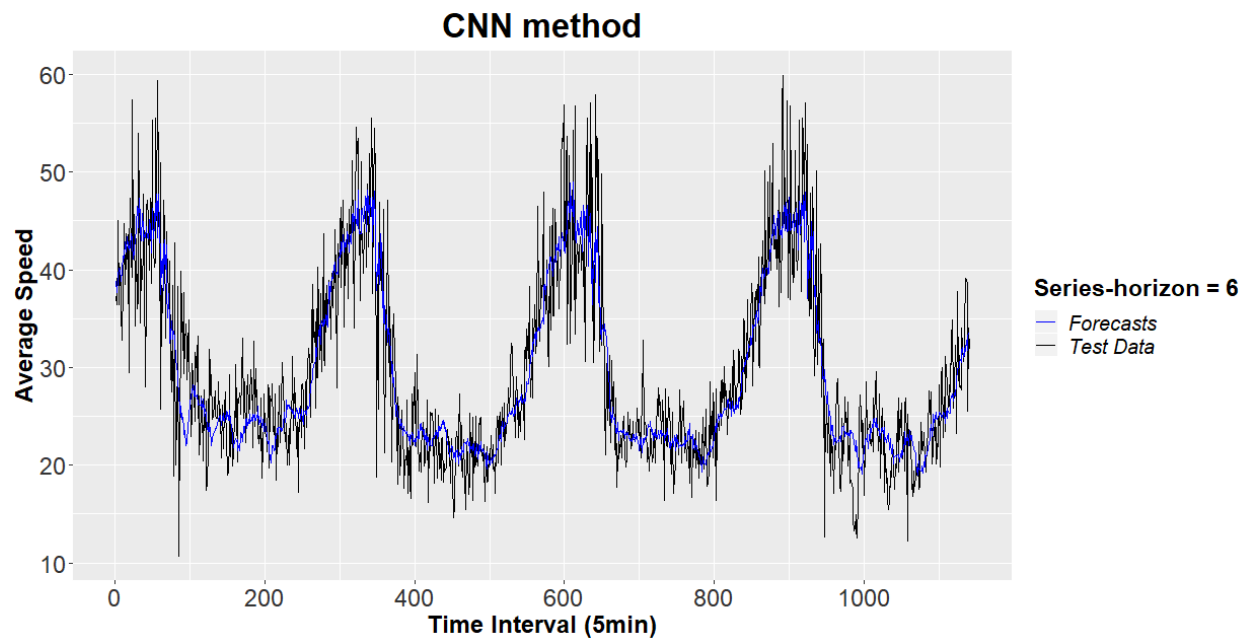
**Πίνακας 58** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο CNN ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.3	26.5	26.0	26.9
<b>RMSE</b>	29.3	28.0	27.8	28.4
<b>MAPE</b>	39.1	36.7	36.3	37.3
<b>sMAPE</b>	31.5	29.2	28.6	29.7
<b>MASE</b>	28.3	26.5	26.0	26.9

Στα Διαγράμματα 104, 105 και 106 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει κάνει πολύ καλύτερες προβλέψεις συγκριτικά με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Όπως παρατηρείται για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1η ημέρα πρόβλεψης. Τέλος, παρατηρούμε στο Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 107) των καταλοίπων ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.

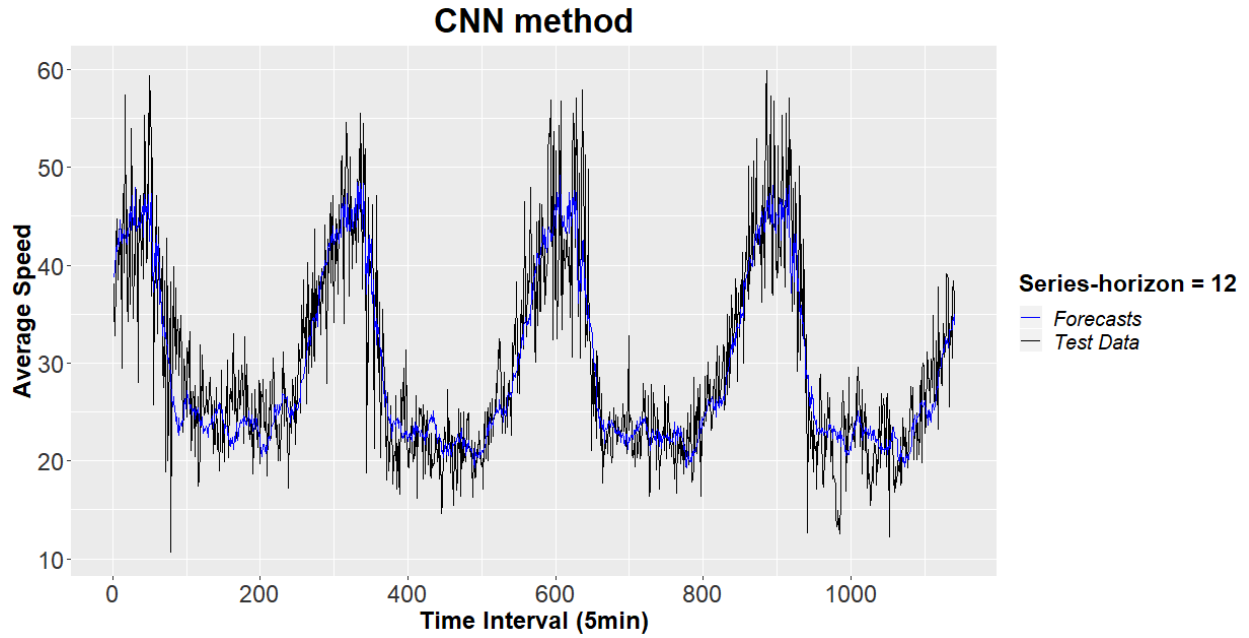


**Διάγραμμα 104** Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.

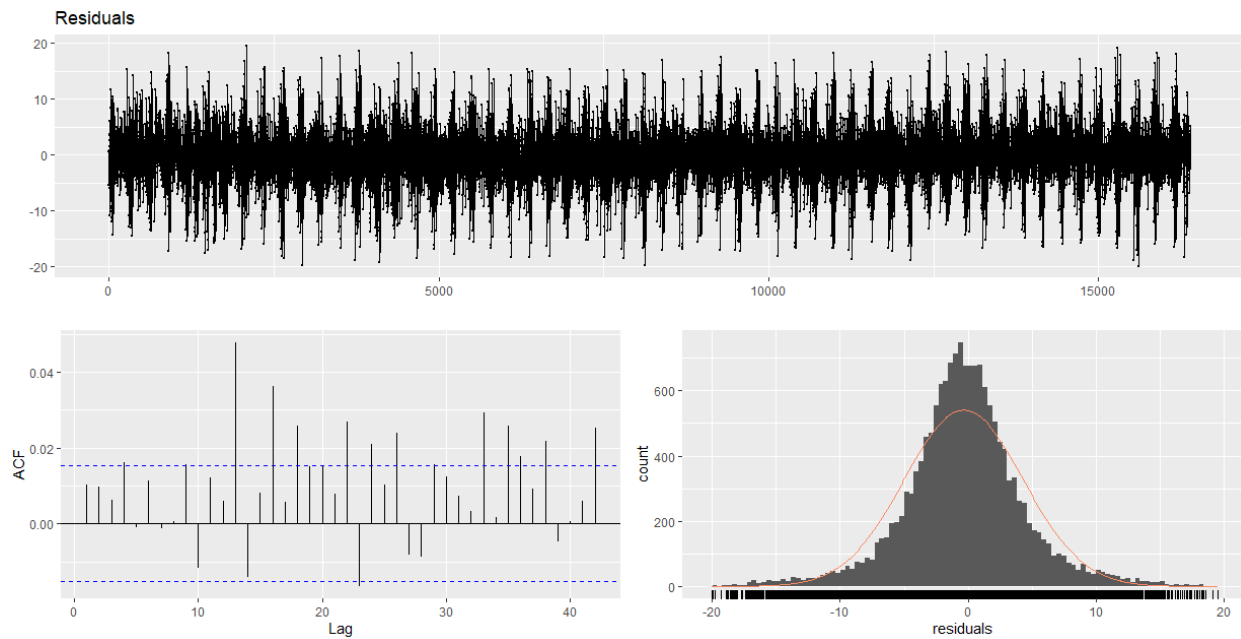


**Διάγραμμα 105** Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.





**Διάγραμμα 106** Προβλέψεις της μεθόδου CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



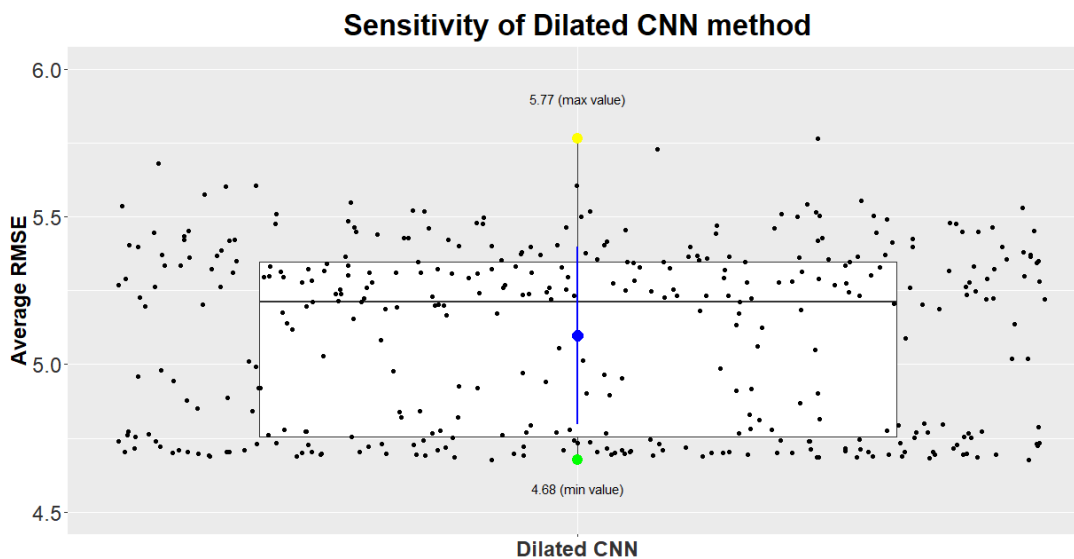
**Διάγραμμα 107** Γραφικός έλεγχος των καταλοίπων της μεθόδου CNN.

### 6.3.5 Μέθοδος Dilated Convolutional Neural Networks (Dilated CNN)

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

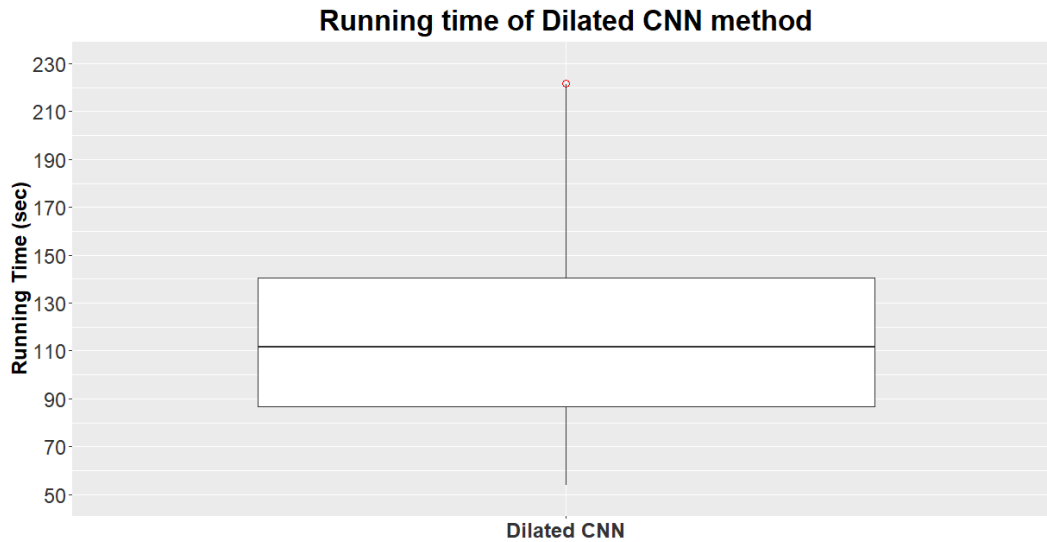
- **Order p:** 12
- **Deseasonalization:** True
- **Detrending:** True
- **Scaling:** True
- **Regularization:** True
- **Block:** 2
- **Filters:** 32
- **Kernel:** 2
- **Batch Size:** 64

με μέσο RMSE 4.67, αριθμό παραμέτρων 50208 και χρόνο εκπαίδευσης 168.53 δευτερόλεπτα. Στο Διάγραμμα 108 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.67 – 5.77). Φαίνεται καθαρά ότι η μέθοδος Dilated CNN έχει την μικρότερη ευαισθησία από τα μοντέλα που έχουν εξεταστεί έως τώρα.

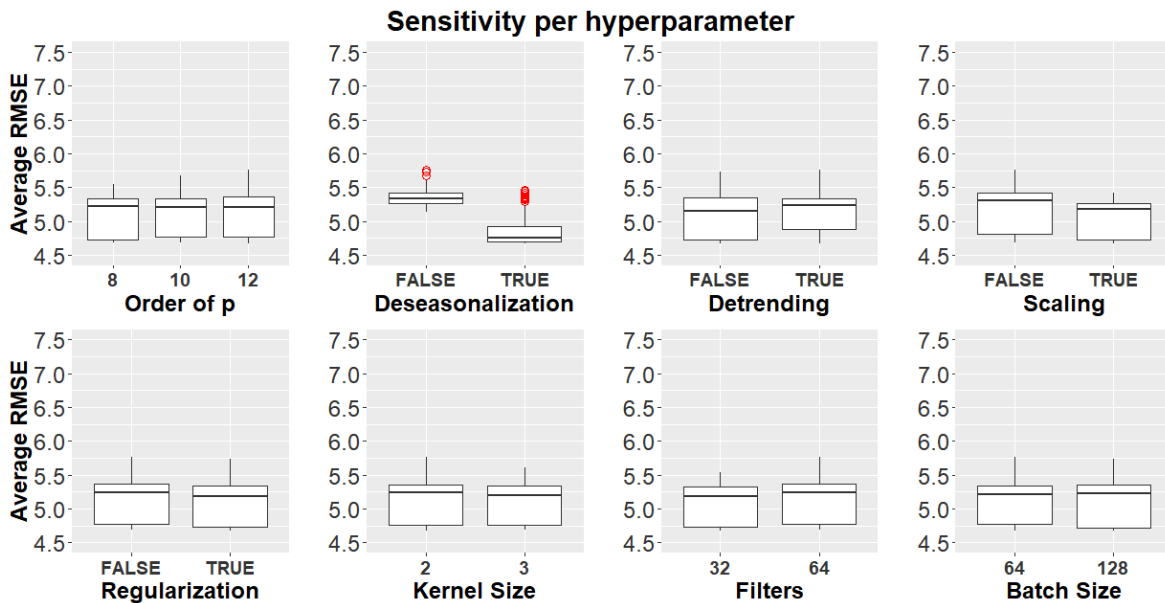


Διάγραμμα 108 Ευαισθησία της μεθόδου Dilated CNN.

Στο Διάγραμμα 109 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 53.9 δευτερόλεπτα έως 221.81 δευτερόλεπτα και μέση τιμή περί τα 114.8 δευτερόλεπτα. Στο Διάγραμμα 110 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαραμέτρο υπό μορφή boxplot.



**Διάγραμμα 109** Χρόνος ανά παραμετροποίηση για την μέθοδο Dilated CNN.



**Διάγραμμα 110** Ευαισθησία της μεθόδου Dilated CNN ανά υπερπαραμέτρο.

Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
window          1  0.189    0.189   11.299 0.000855 ***
Filters          1  0.714    0.714   42.690 2.09e-10 ***
Kernels          1  0.055    0.055    3.285 0.070719 .
Batch            1  0.078    0.078    4.678 0.031179 *
Deseasonalization 1 24.345   24.345 1455.849 < 2e-16 ***
Detrending       1  0.666    0.666   39.805 7.92e-10 ***
Scaling          1  2.153    2.153  128.742 < 2e-16 ***
Regularization   1  0.239    0.239   14.320 0.000179 ***
Residuals       375  6.271    0.017
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Window, Filters, Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 59.

**Πίνακας 59** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Dilated CNN.

<b>(Intercept)</b>	5.2458118646
<b>Window</b>	0.0135833794
<b>Filters</b>	0.0026947879
<b>Kernels</b>	-0.0239205443
<b>Batch</b>	-0.0004460363
<b>DeseasonalizationTRUE</b>	-0.5035787996
<b>DetrendingTRUE</b>	0.0832677815
<b>ScalingTRUE</b>	-0.1497510511
<b>RegularizationTRUE</b>	-0.0499433152

Απο τον Πίνακα 59 συμπεραίνεται ότι και πάλι η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο διάγραμμα 110. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.35,

ενώ με εποχική αποσύνθεση είναι 4.85. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 60 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 61 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 60** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Dilated CNN.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.382	3.417	3.451	3.493	3.509	3.529	3.536	3.551	3.552	3.543	3.542	3.548
<b>RMSE</b>	4.595	4.620	4.637	4.679	4.695	4.715	4.723	4.737	4.738	4.724	4.726	4.733
<b>MAPE</b>	11.279	11.462	11.643	11.815	11.899	11.990	12.014	12.077	12.075	12.054	12.064	12.089
<b>sMAPE</b>	11.361	11.526	11.690	11.861	11.940	12.019	12.039	12.098	12.101	12.077	12.082	12.106
<b>MASE</b>	0.697	0.666	0.670	0.674	0.675	0.673	0.671	0.667	0.659	0.648	0.640	0.635

**Πίνακας 61** Συνοπτικά μέτρα σφάλματος για την μέθοδο Dilated CNN.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.436	3.531	3.546	3.504
<b>RMSE</b>	4.633	4.717	4.730	4.693
<b>MAPE</b>	11.550	11.995	12.070	11.872
<b>sMAPE</b>	11.609	12.024	12.091	11.908
<b>MASE</b>	0.677	0.671	0.646	0.665

**Πίνακας 62** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Dilated CNN.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
44082.016 sec	61.769 sec	37.719 sec	44181.504 sec	18721	2.669

Παρατηρούμε ότι η μέθοδος Dilated CNN είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 62 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν

λόγω μέθοδος χρειάστηκε περίπου 18721 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 63 και 64. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.3%.

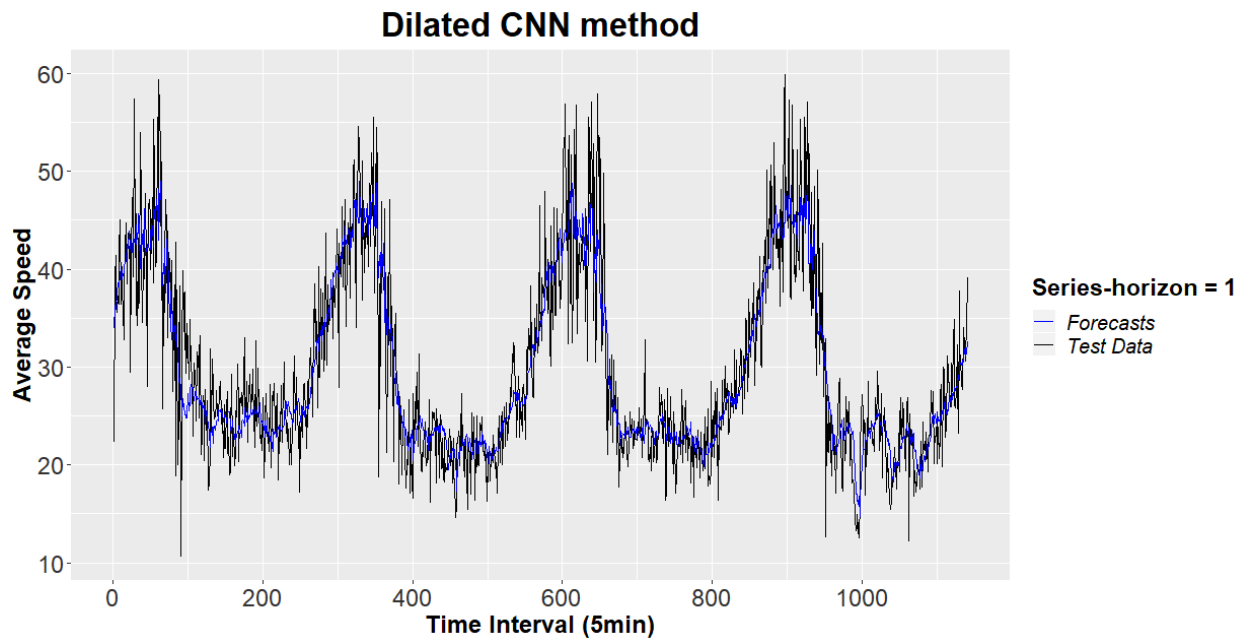
**Πίνακας 63** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Dilated CNN ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.4	28.5	27.7	26.9	26.7	26.3	26.1	25.8	25.8	25.8	25.8	25.8
<b>RMSE</b>	30.1	29.3	29.0	28.5	28.3	28.0	27.9	27.7	27.6	27.8	27.7	27.6
<b>MAPE</b>	40.5	39.4	38.4	37.6	37.2	36.7	36.6	36.3	36.3	36.3	36.2	36.1
<b>sMAPE</b>	32.8	31.6	30.6	29.7	29.3	28.8	28.8	28.4	28.4	28.4	28.4	28.3
<b>MASE</b>	29.4	28.4	27.7	27.0	26.7	26.3	26.1	25.8	25.8	25.8	25.8	25.8

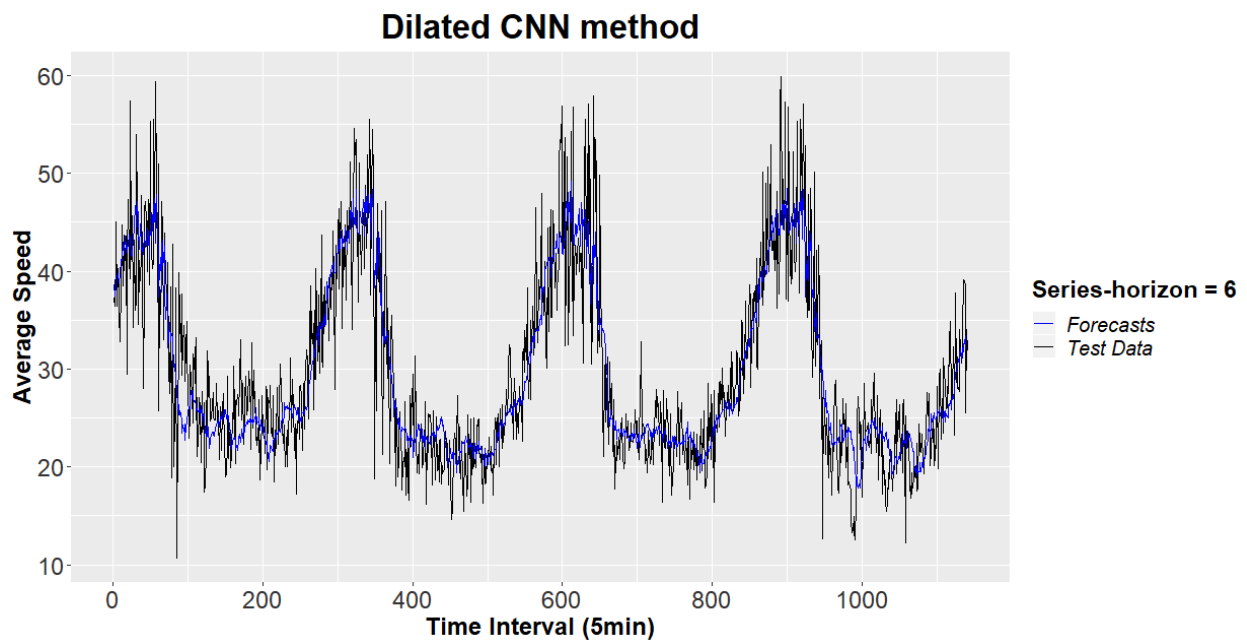
**Πίνακας 64** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Dilated CNN ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.1	26.2	25.8	26.7
<b>RMSE</b>	29.2	27.9	27.7	28.3
<b>MAPE</b>	38.9	36.7	36.2	37.3
<b>sMAPE</b>	31.2	28.8	28.4	29.5
<b>MASE</b>	28.1	26.2	25.8	26.7

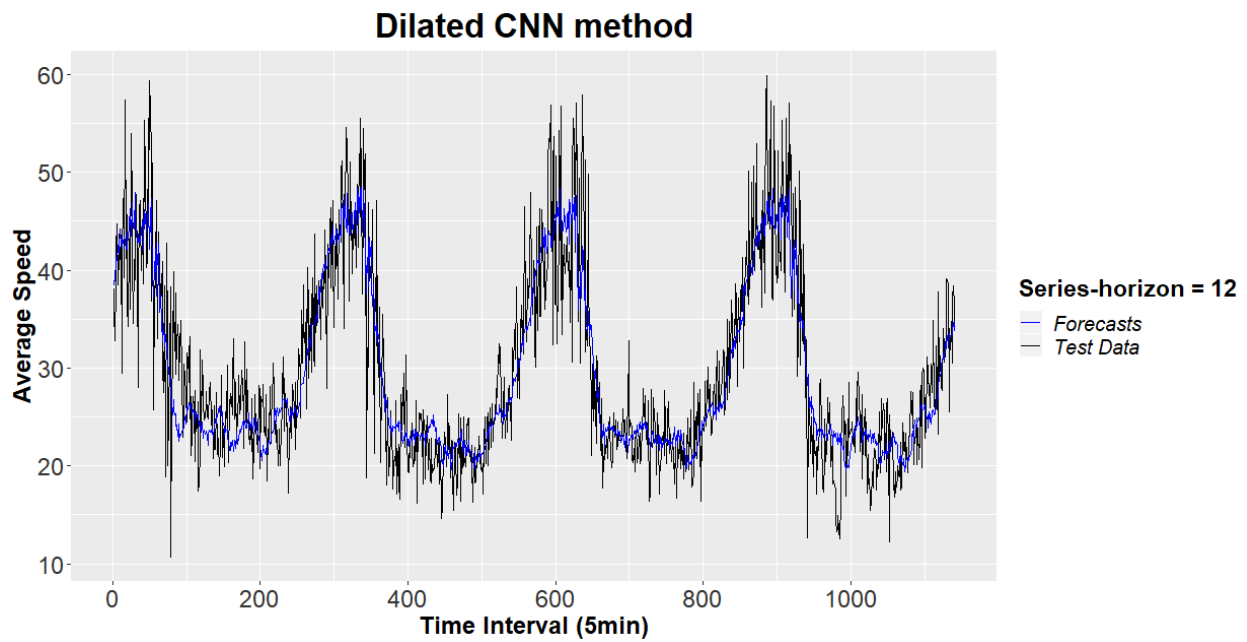
Στα Διαγράμματα 111, 112 και 113 φαίνονται ενδεικτικά τις προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει πολύ καλύτερες προβλέψεις συγκριτικά με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Όπως παρατηρούμε για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1η ημέρα πρόβλεψης. Τέλος, από το διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 114) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



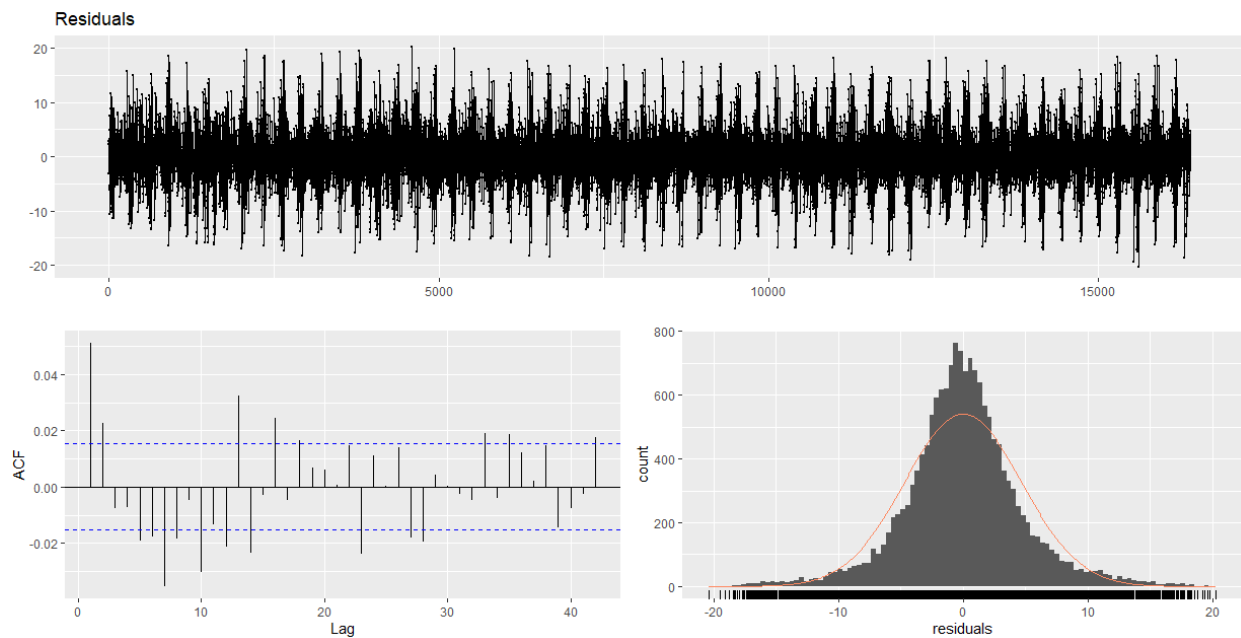
**Διάγραμμα 111** Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 112** Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 113** Προβλέψεις της μεθόδου Dilated CNN για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



**Διάγραμμα 114** Γραφικός έλεγχος των καταλοίπων της μεθόδου Dilated CNN.

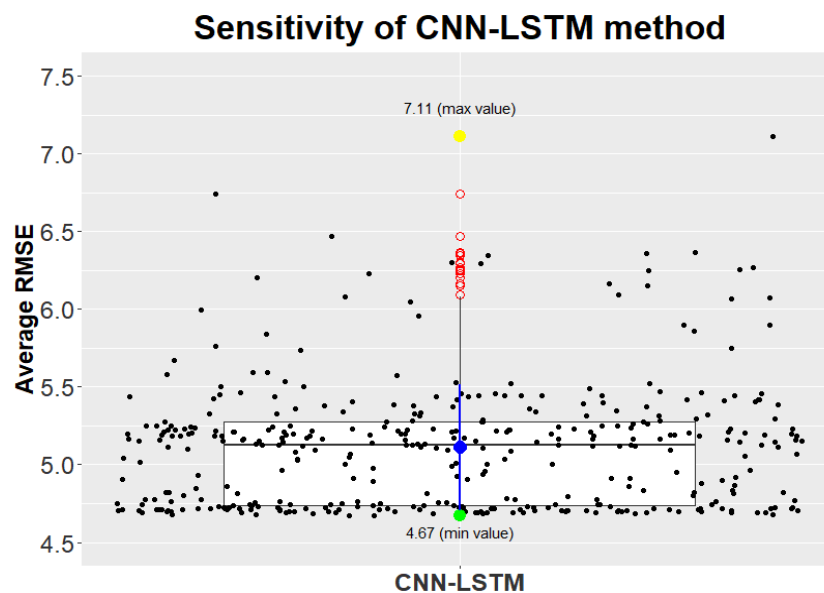


### 6.3.6 Μέθοδος CNN-LSTM

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

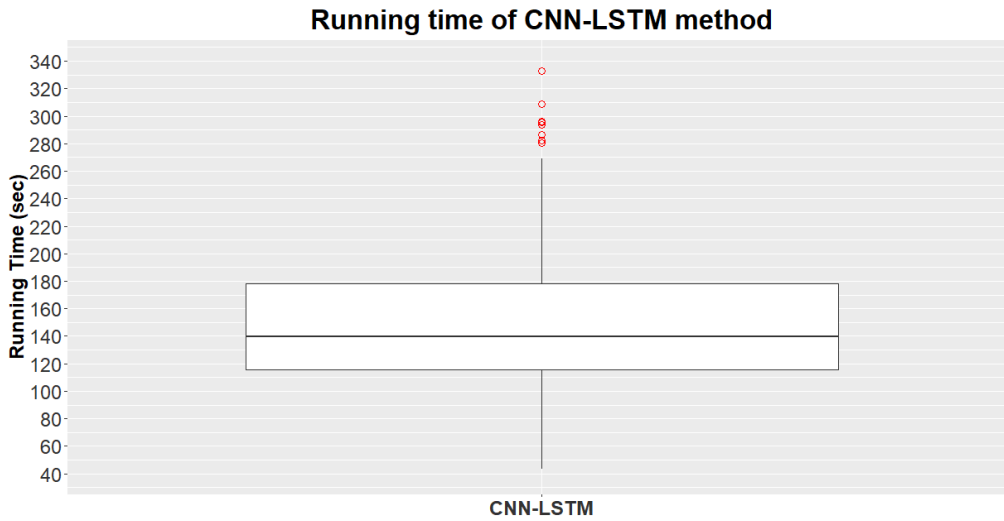
- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** True
- **Block:** 1
- **Filters:** 64
- **Kernel:** 3
- **Layers:** 1
- **Nodes:** 100
- **Batch Size:** 128

με μέσο RMSE 4.67, αριθμό παραμέτρων 192320 και χρόνο εκπαίδευσης 82.77 δευτερόλεπτα. Στο Διάγραμμα 115 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.67 – 7.11).

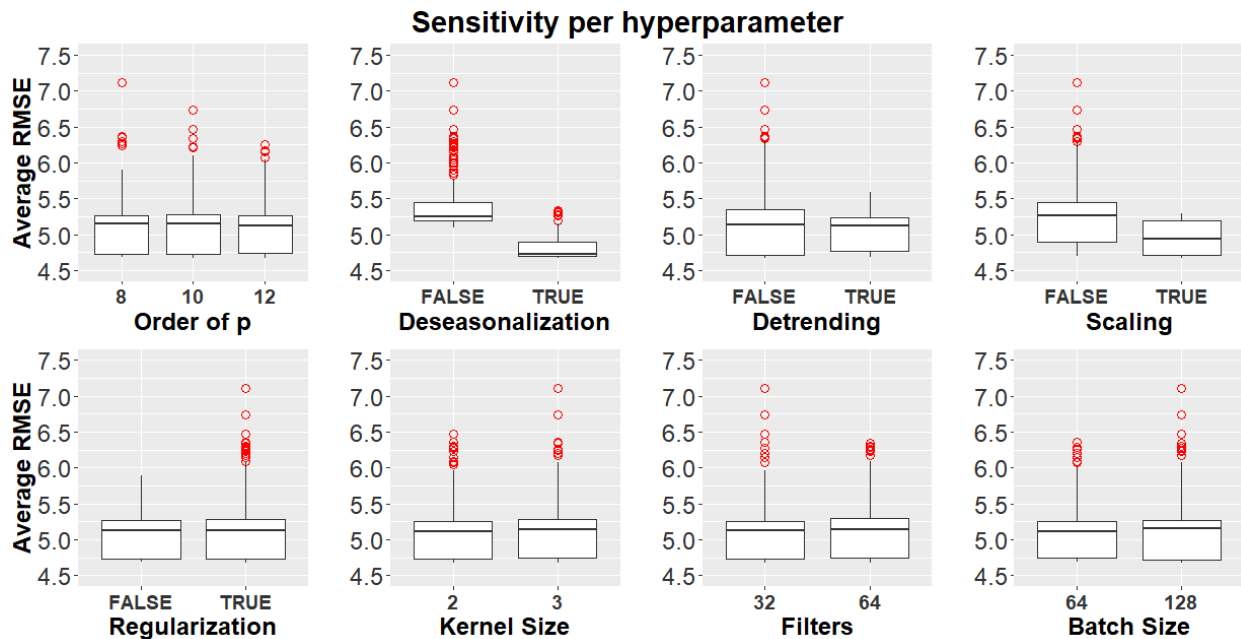


Διάγραμμα 115 Ευαισθησία της μεθόδου CNN-LSTM.

Στο Διάγραμμα 116 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 43.62 δευτερόλεπτα έως 332.79 δευτερόλεπτα και μέση τιμή περί τα 147.98 δευτερόλεπτα. Στο Διάγραμμα 117 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαραμέτρο υπό μορφή boxplot.



Διάγραμμα 116 Χρόνος ανά παραμετροποίηση για την μέθοδο CNN-LSTM.



Διάγραμμα 117 Ευαισθησία της μεθόδου CNN-LSTM ανά υπερπαραμέτρο.

Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq F value    Pr(>F)
window          1   0.03    0.03    0.567    0.452
Filters          1   0.02    0.02    0.505    0.478
Kernels          1   0.12    0.12    2.716    0.100
Batch            1   0.01    0.01    0.234    0.628
Deseasonalization 1 34.19   34.19 750.967 < 2e-16 ***
Detrending       1   1.01    1.01  22.218 3.44e-06 ***
Scaling          1  10.01   10.01 219.922 < 2e-16 ***
Regularization   1   0.71    0.71  15.654 9.09e-05 ***
Residuals       375  17.08    0.05
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 65.

**Πίνακας 65** Συντελεστές της ανάλυσης ANOVA για την μέθοδο CNN-LSTM.

<b>(Intercept)</b>	5.5007893886
<b>Window</b>	-0.0050233235
<b>Filters</b>	0.0004837442
<b>Kernels</b>	0.0358917775
<b>Batch</b>	0.0001647850
<b>DeseasonalizationTRUE</b>	-0.5968198925
<b>DetrendingTRUE</b>	-0.1026559206
<b>ScalingTRUE</b>	-0.3229735341
<b>RegularizationTRUE</b>	0.0861690623

Παρατηρούμε στον Πίνακα 65 ότι για ακόμα μία φορά η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 117. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι

5.40, ενώ με εποχική αποσύνθεση είναι 4.80. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 66 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον πίνακα 67 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 66** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο CNN-LSTM.

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
<b>MAE</b>	3.384	3.415	3.440	3.473	3.502	3.510	3.524	3.535	3.538	3.533	3.530	3.536
<b>RMSE</b>	4.599	4.623	4.647	4.675	4.698	4.709	4.722	4.730	4.727	4.716	4.713	4.716
<b>MAPE</b>	11.368	11.530	11.670	11.826	11.951	12.003	12.061	12.105	12.107	12.113	12.103	12.121
<b>sMAPE</b>	11.361	11.504	11.631	11.785	11.907	11.949	11.998	12.043	12.051	12.047	12.046	12.066
<b>MASE</b>	0.697	0.666	0.668	0.670	0.674	0.670	0.669	0.664	0.656	0.646	0.638	0.633

**Πίνακας 67** Συνοπτικά μέτρα σφάλματος για την μέθοδο CNN-LSTM.

	Short	Medium	Long	Average
<b>MAE</b>	3.428	3.518	3.534	3.493
<b>RMSE</b>	4.636	4.715	4.718	4.690
<b>MAPE</b>	11.599	12.030	12.111	11.913
<b>sMAPE</b>	11.570	11.974	12.052	11.866
<b>MASE</b>	0.675	0.669	0.643	0.663

**Πίνακας 68** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο CNN-LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
56824.856 sec	92.785 sec	14.733 sec	56932.374 sec	24124	2.659

Η μέθοδος CNN-LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον πίνακα 68 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η

προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 24124 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο αναφοράς. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 69 και 70. Η ποσοστιαία μείωση που επιτεύχθηκε στο μέσο RMSE της τάξεως του 28.3%.

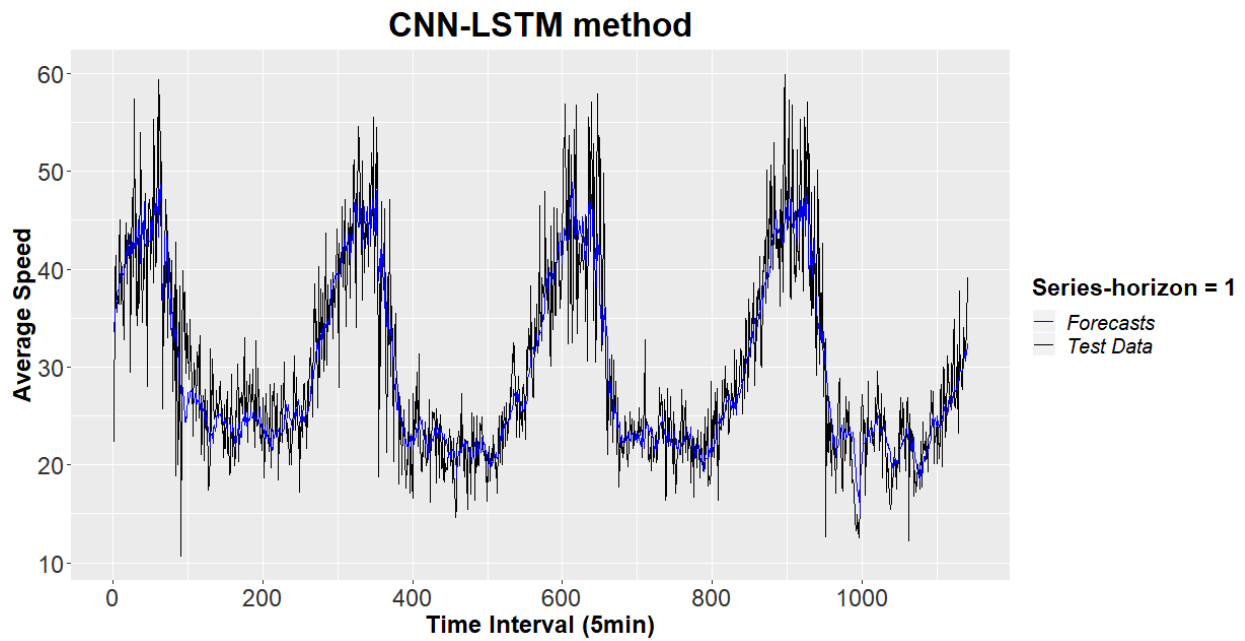
**Πίνακας 69** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο CNN-LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.4	28.5	27.9	27.4	26.8	26.7	26.4	26.2	26.1	26.1	26.1	26.0
<b>RMSE</b>	30.0	29.3	28.9	28.6	28.2	28.1	27.9	27.7	27.8	27.9	27.9	27.9
<b>MAPE</b>	40.0	39.0	38.2	37.5	36.9	36.6	36.3	36.1	36.1	36.0	36.0	35.9
<b>sMAPE</b>	32.8	31.8	31.0	30.2	29.5	29.3	29.0	28.8	28.7	28.6	28.6	28.5
<b>MASE</b>	29.4	28.5	27.9	27.4	26.9	26.7	26.4	26.2	26.1	26.0	26.0	26.1

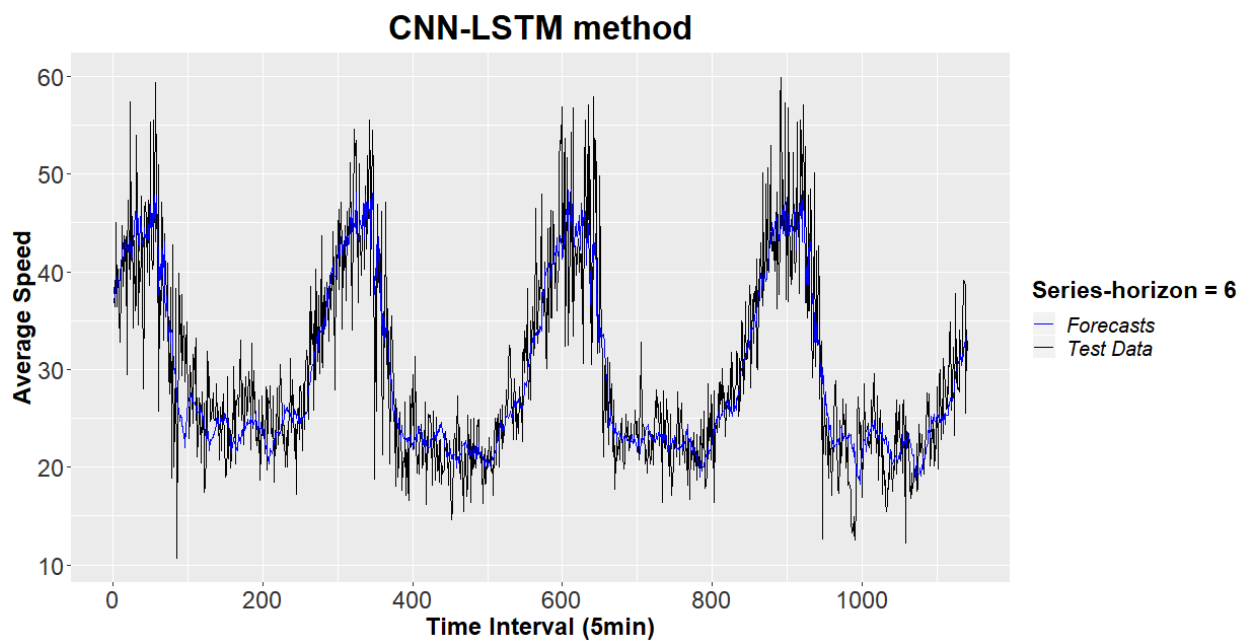
**Πίνακας 70** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο CNN-LSTM ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.3	26.5	26.1	26.9
<b>RMSE</b>	29.2	28.0	27.9	28.3
<b>MAPE</b>	38.7	36.5	36.0	37.1
<b>sMAPE</b>	31.4	29.1	28.6	29.7
<b>MASE</b>	28.3	26.5	26.1	27.0

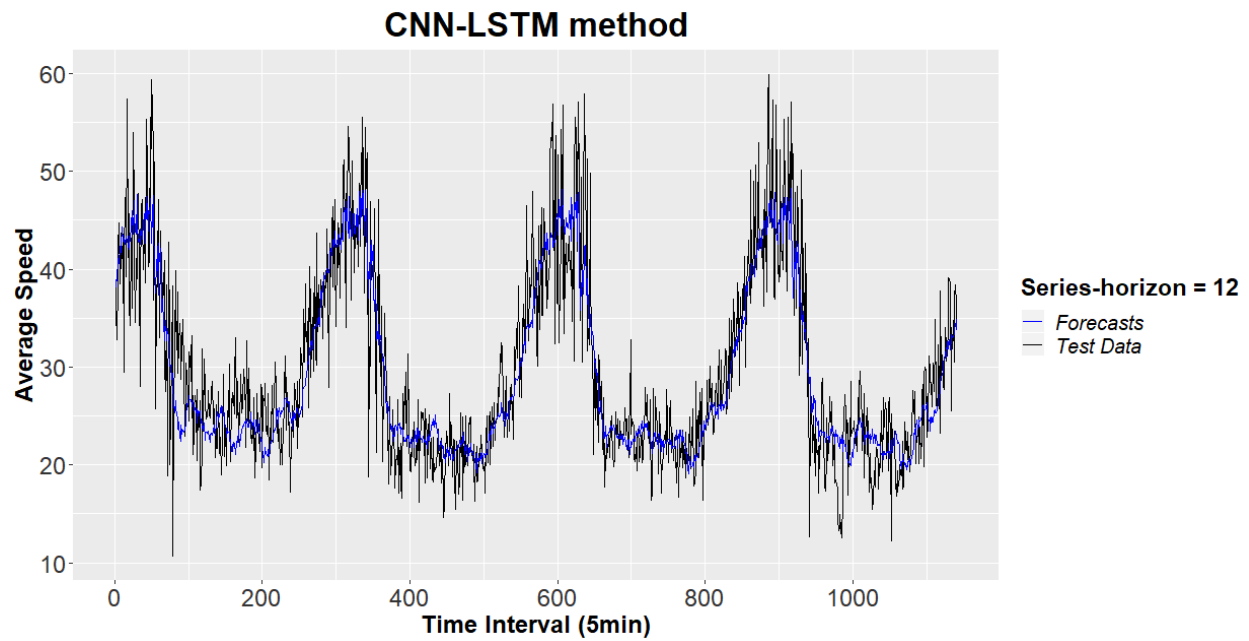
Στα Διαγράμματα 118 119 και 120 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει κάνει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο βάσης, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1η ημέρα πρόβλεψης. Τέλος, από το διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 121) των καταλοίπων ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



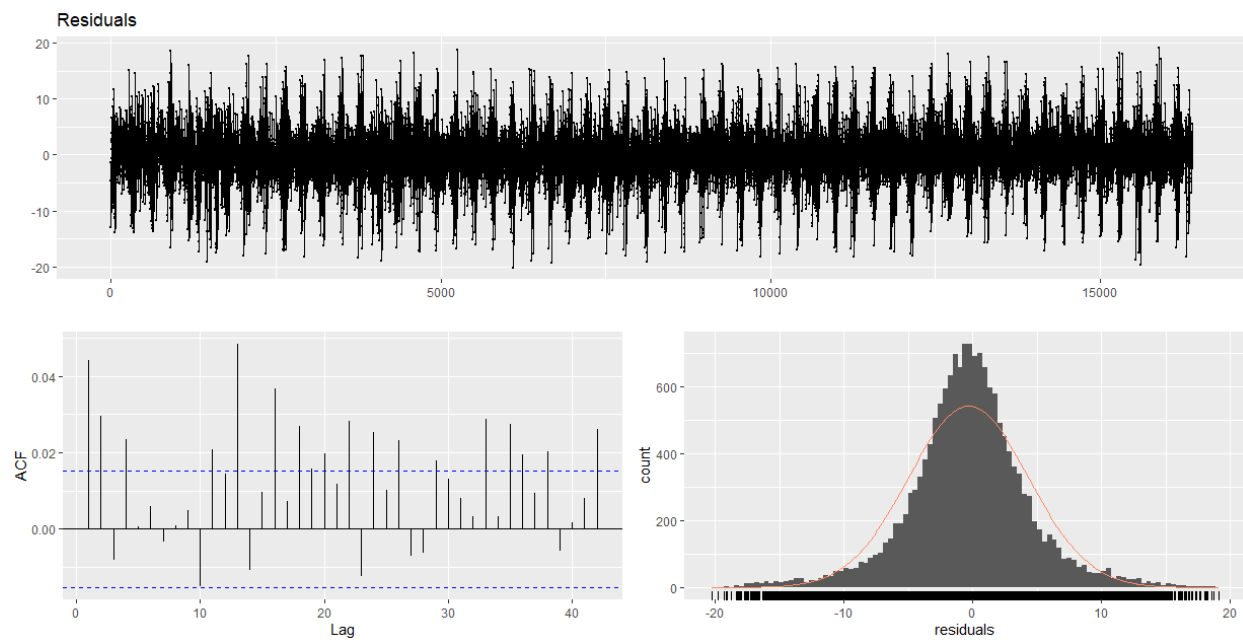
**Διάγραμμα 118** Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 119** Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 120** Προβλέψεις της μεθόδου CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



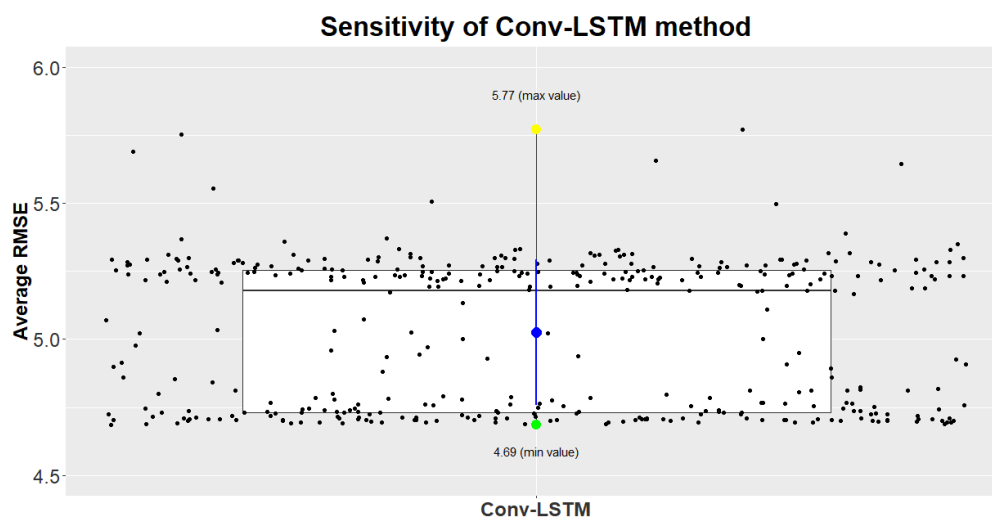
**Διάγραμμα 121** Γραφικός έλεγχος των καταλοίπων της μεθόδου CNN-LSTM.

### 6.3.7 Μέθοδος Convolutional LSTM (Conv-LSTM)

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** True
- **Layers:** 1
- **Filters:** 64
- **Kernel:** 4
- **Nodes:** 100
- **Batch Size:** 128

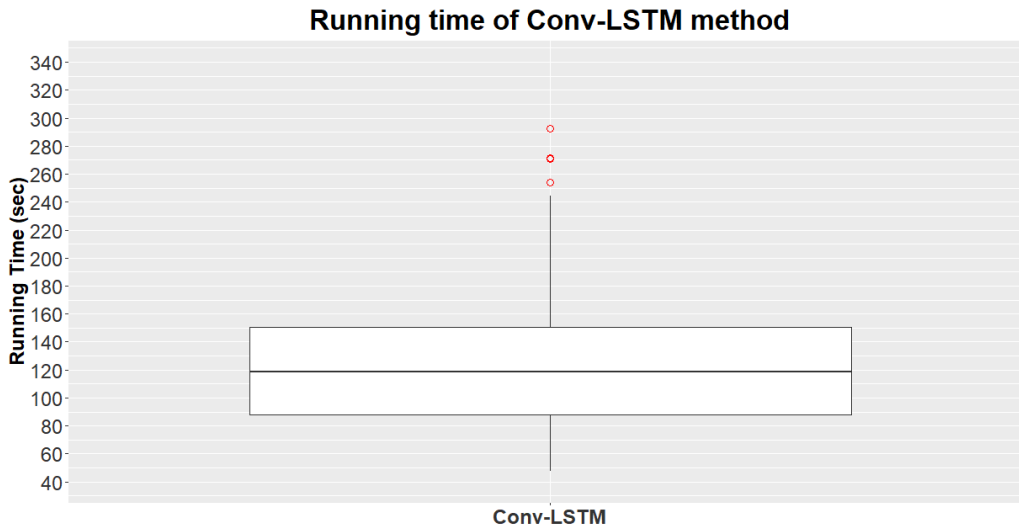
με μέσο RMSE 4.68, αριθμό παραμέτρων 100128 και χρόνο εκπαίδευσης 91.92 δευτερόλεπτα. Στο Διάγραμμα 122 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.68 – 5.77). Η μέθοδος Conv-LSTM μαζί με την μέθοδο Dilated CNN έχουν την μικρότερη ευαισθησία από τα μοντέλα που έχουν εξεταστεί μέχρι τώρα.



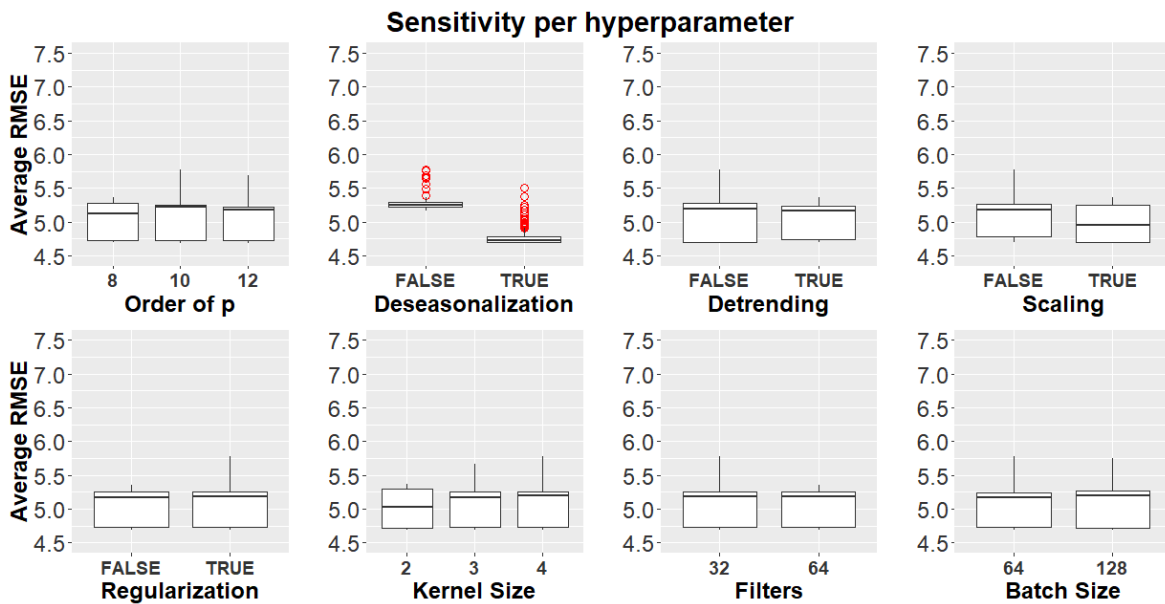
Διάγραμμα 122 Ευαισθησία της μεθόδου Conv-LSTM.



Στο Διάγραμμα 123 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 47.37 δευτερόλεπτα έως 292.68 δευτερόλεπτα και μέση τιμή περί τα 124.04 δευτερόλεπτα. Στο Διάγραμμα 124 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαράμετρο υπό μορφή boxplot.



**Διάγραμμα 123** Χρόνος ανά παραμετροποίηση για την μέθοδο Conv-LSTM.



**Διάγραμμα 124** Ευαισθησία της μεθόδου Conv-LSTM ανά υπερπαράμετρο.

Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
window         1  0.026   0.026    2.691 0.10174
Filters         1  0.033   0.033    3.377 0.06689 .
kernels         1  0.054   0.054    5.526 0.01926 *
Batch           1  0.001   0.001    0.090 0.76411
Deseasonalization 1 23.005  23.005 2350.867 < 2e-16 ***
Detrending      1  0.034   0.034    3.520 0.06143 .
Scaling         1  0.780   0.780   79.754 < 2e-16 ***
Regularization  1  0.073   0.073    7.453 0.00663 **
Residuals      375  3.670   0.010
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Deseasonalization, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 71.

**Πίνακας 71** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Conv-LSTM.

<b>(Intercept)</b>	5.269417e+00
<b>Window</b>	-2.953823e-04
<b>Filters</b>	-5.798251e-04
<b>Kernels</b>	2.146714e-02
<b>Batch</b>	4.737421e-05
<b>DeseasonalizationTRUE</b>	-4.895245e-01
<b>DetrendingTRUE</b>	-1.894098e-02
<b>ScalingTRUE</b>	-9.016451e-02
<b>RegularizationTRUE</b>	2.756220e-02

Παρατηρούμε στον Πίνακα 71 ότι και πάλι η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 124. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.27,

ενώ με εποχική αποσύνθεση είναι 4.78. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 72 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 73 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 72** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Conv-LSTM.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.383	3.419	3.447	3.489	3.508	3.514	3.532	3.534	3.544	3.546	3.548	3.551
<b>RMSE</b>	4.600	4.635	4.660	4.706	4.726	4.723	4.730	4.735	4.736	4.739	4.739	4.738
<b>MAPE</b>	11.256	11.439	11.604	11.796	11.916	11.934	12.000	12.010	12.049	12.075	12.085	12.102
<b>sMAPE</b>	11.352	11.510	11.652	11.820	11.915	11.945	12.014	12.020	12.054	12.075	12.088	12.094
<b>MASE</b>	0.697	0.667	0.670	0.673	0.675	0.670	0.670	0.663	0.657	0.649	0.641	0.636

**Πίνακας 73** Συνοπτικά μέτρα σφάλματος για την μέθοδο Conv-LSTM.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.434	3.522	3.547	3.501
<b>RMSE</b>	4.650	4.729	4.738	4.706
<b>MAPE</b>	11.524	11.965	12.078	11.855
<b>sMAPE</b>	11.583	11.974	12.078	11.878
<b>MASE</b>	0.677	0.670	0.646	0.664

**Πίνακας 74** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Conv-LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
47631.245 sec	211.489 sec	17.175 sec	47859.910 sec	20280	2.670

Παρατηρούμε ότι η μέθοδος Conv-LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 74 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική

πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 20280 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο αναφοράς. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 75 και 76. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.1%.

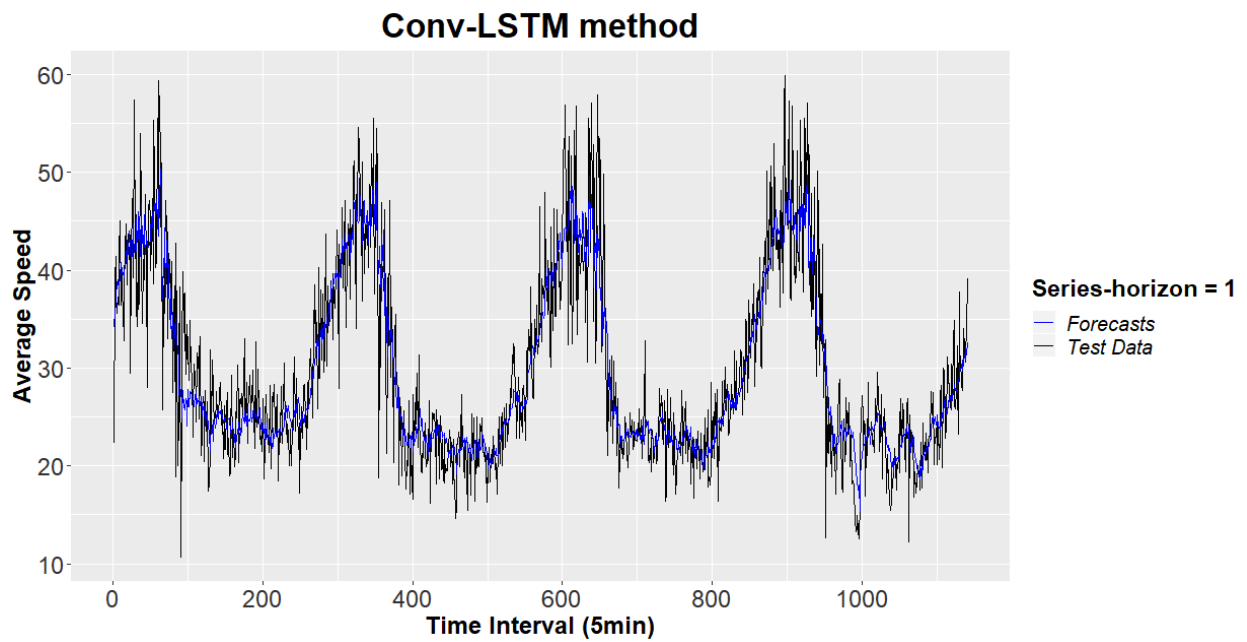
**Πίνακας 75** Ποσοστιαία μείωση (ακρίβεια 1ου δεκαδικού) για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Conv-LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.4	28.4	27.8	27.0	26.7	26.6	26.2	26.2	26.0	25.8	25.7	25.7
<b>RMSE</b>	30.0	29.1	28.7	28.1	27.8	27.8	27.7	27.7	27.7	27.5	27.5	27.6
<b>MAPE</b>	40.6	39.5	38.6	37.7	37.1	37.0	36.6	36.6	36.4	36.2	36.1	36.0
<b>sMAPE</b>	32.9	31.7	30.8	30.0	29.5	29.3	28.9	28.9	28.7	28.5	28.3	28.4
<b>MASE</b>	29.4	28.4	27.8	27.0	26.7	26.6	26.2	26.2	26.0	25.8	25.7	25.7

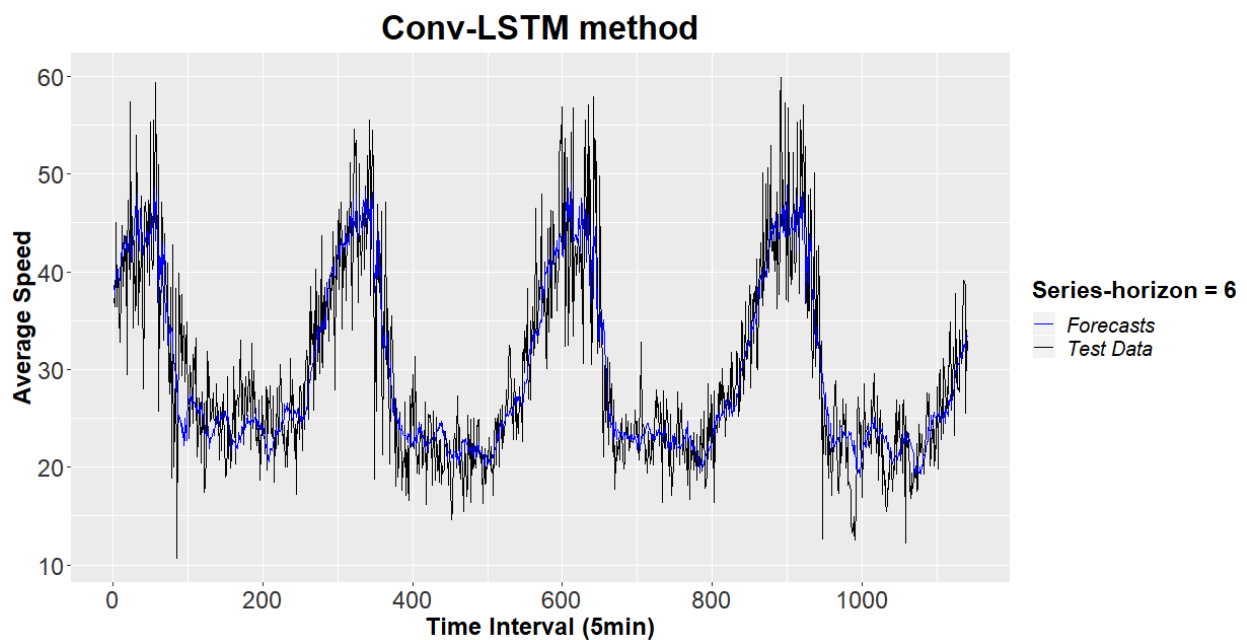
**Πίνακας 76** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Conv-LSTM ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.2	26.4	25.8	26.8
<b>RMSE</b>	29.0	27.8	27.6	28.1
<b>MAPE</b>	39.1	36.8	36.2	37.4
<b>sMAPE</b>	31.3	29.1	28.5	29.6
<b>MASE</b>	28.2	26.4	25.8	26.8

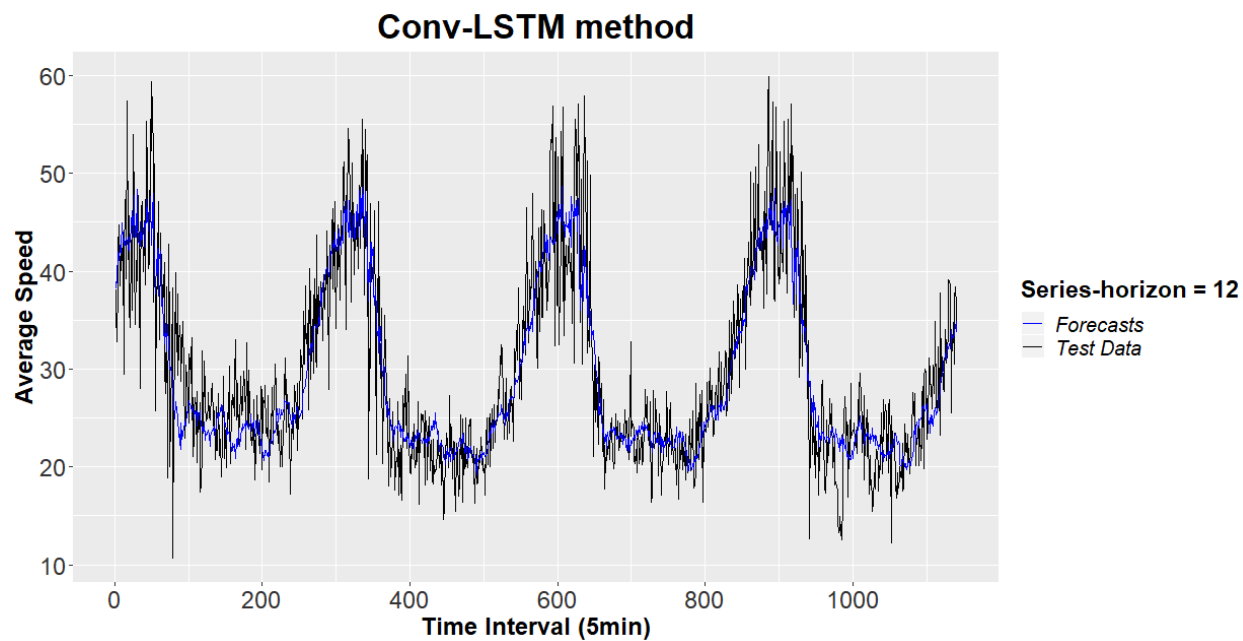
Στα διαγράμματα 125, 126 και 127 φαίνονται ενδεικτικά τις προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει κάνει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1η ημέρα πρόβλεψης. Τέλος, από το διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 128) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



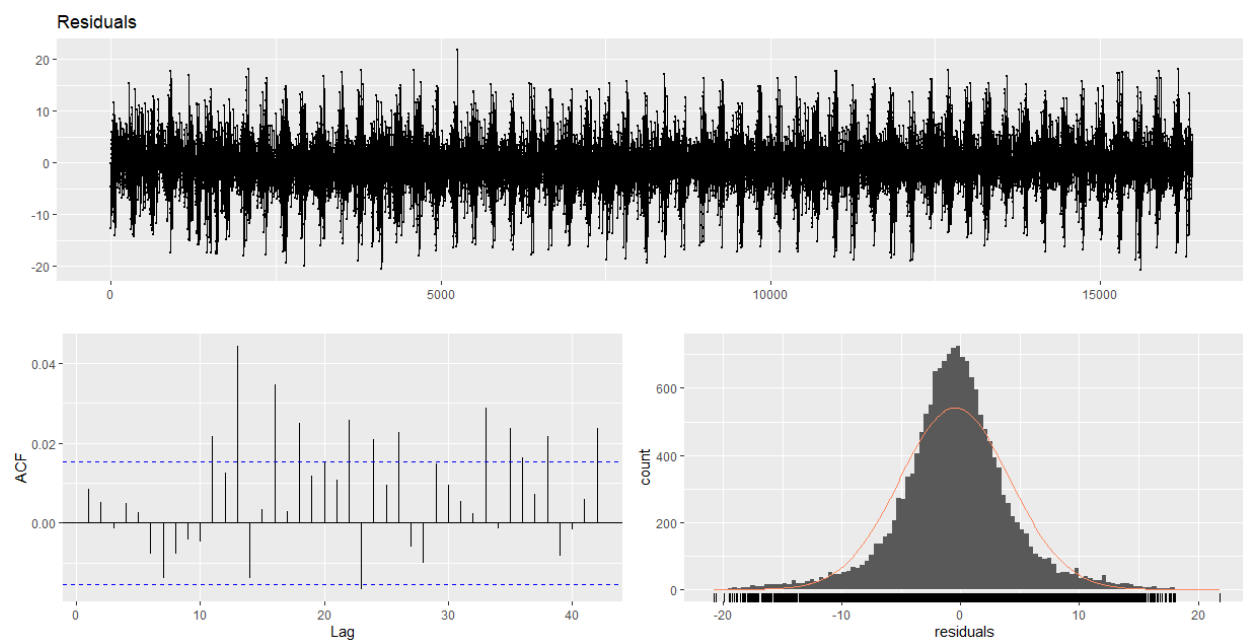
**Διάγραμμα 125** Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 126** Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 127** Προβλέψεις της μεθόδου Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



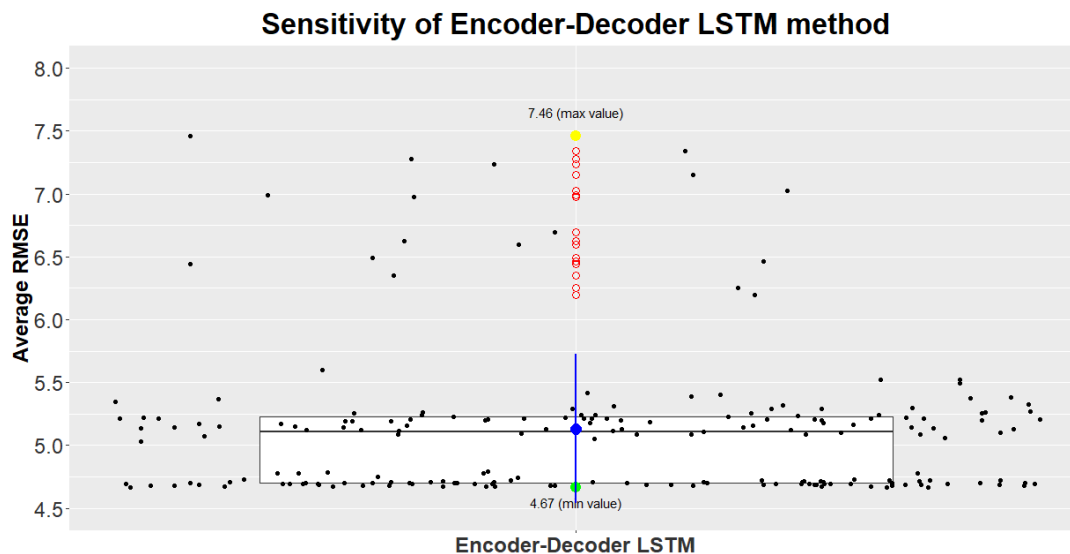
**Διάγραμμα 128** Γραφικός έλεγχος των καταλοίπων της μεθόδου Conv-LSTM.

### 6.3.8 Μέθοδος Encoder-Decoder LSTM

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

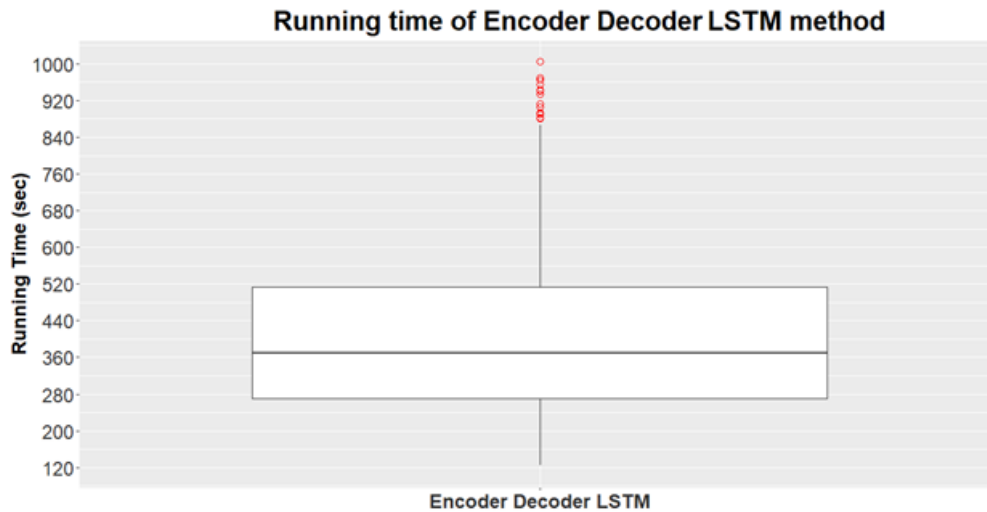
- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** False
- **Regularization:** True
- **Layers:** 1
- **Nodes:** 200
- **Batch Size:** 64

με μέσο RMSE 4.66, αριθμό παραμέτρων 502601 και χρόνο εκπαίδευσης 941.088 δευτερόλεπτα. Στο Διάγραμμα 129 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.66 – 7.46). Η μέθοδος Encoder Decoder LSTM έχει την μεγαλύτερη ευαισθησία συγκριτικά με τα μοντέλα που έχουν εξεταστεί έως τώρα.

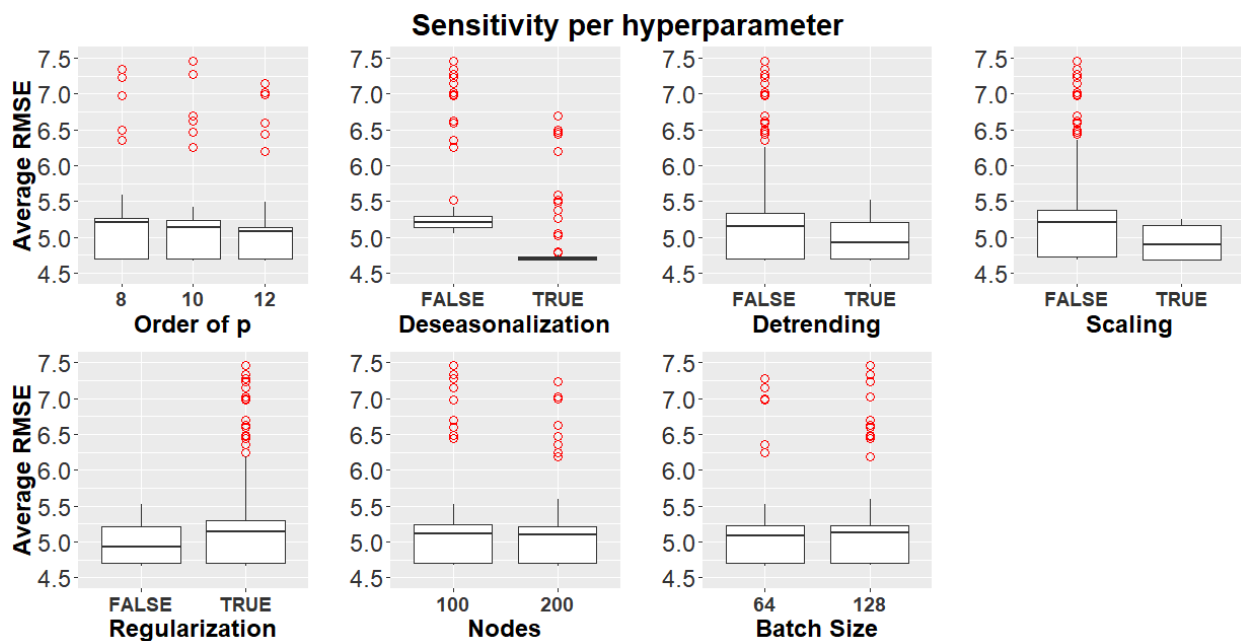


Διάγραμμα 129 Ευαισθησία της μεθόδου Encoder Decoder LSTM.

Στο Διάγραμμα 130 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 124.33 δευτερόλεπτα έως 1004.23 δευτερόλεπτα και μέση τιμή περί τα 431.76 δευτερόλεπτα. Στο Διάγραμμα 131 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαράμετρο υπό μορφή boxplot.



**Διάγραμμα 130** Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder LSTM.



**Διάγραμμα 131** Ευαισθησία της μεθόδου Encoder-Decoder LSTM ανά υπερπαράμετρο.



Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq F value    Pr(>F)
window         1  0.072   0.072    0.430    0.513
Nodes          1  0.078   0.078    0.462    0.497
Batch          1  0.222   0.222    1.315    0.253
Deseasonalization  1 16.222  16.222  96.284 < 2e-16 ***
Detrending     1  5.887   5.887  34.941 1.62e-08 ***
Scaling        1  8.029   8.029  47.657 7.97e-11 ***
Regularization  1  5.534   5.534  32.849 4.01e-08 ***
Residuals     184 31.000   0.168
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 77.

**Πίνακας 77** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder-Decoder LSTM.

<b>(Intercept)</b>	5.7091035970
<b>Window</b>	-0.0118914179
<b>Nodes</b>	-0.0004028275
<b>Batch</b>	0.0010616599
<b>DeseasonalizationTRUE</b>	-0.5813349994
<b>DetrendingTRUE</b>	-0.3502031689
<b>ScalingTRUE</b>	-0.4089894274
<b>RegularizationTRUE</b>	0.3395565294

Από τα αποτελέσματα του Πίνακα 77 συμπεραίνεται ότι για ακόμη μία φορά η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 131. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι 5.42, ενώ με εποχική αποσύνθεση είναι 4.84. Ακολουθούν τα

αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 78 φαίνονται οι 5 μετρικές σφάλματος ανά ορίζοντα πρόβλεψης. Στον Πίνακα 79 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 78** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder LSTM.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.361	3.400	3.425	3.457	3.476	3.494	3.512	3.520	3.530	3.518	3.520	3.535
<b>RMSE</b>	4.566	4.593	4.617	4.644	4.658	4.672	4.687	4.696	4.705	4.685	4.688	4.710
<b>MAPE</b>	11.211	11.372	11.489	11.623	11.713	11.798	11.872	11.909	11.949	11.931	11.947	12.004
<b>sMAPE</b>	11.300	11.458	11.578	11.714	11.802	11.873	11.939	11.971	12.008	11.986	12.007	12.059
<b>MASE</b>	0.692	0.663	0.665	0.667	0.669	0.667	0.666	0.661	0.655	0.644	0.636	0.633

**Πίνακας 79** Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder LSTM.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.411	3.501	3.526	3.479
<b>RMSE</b>	4.605	4.678	4.697	4.660
<b>MAPE</b>	11.424	11.823	11.958	11.735
<b>sMAPE</b>	11.513	11.896	12.015	11.808
<b>MASE</b>	0.672	0.666	0.642	0.660

**Πίνακας 80** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
82897.182 sec	893.064 sec	27.985 sec	83818.231 sec	35516	2.728

Τα αποτελέσματα δείχνουν ότι και η μέθοδος Encoder Decoder LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 8 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 35516 φορές παραπάνω υπολογιστικό χρόνο

σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνεται στους Πίνακες 81 και 82. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.8%.

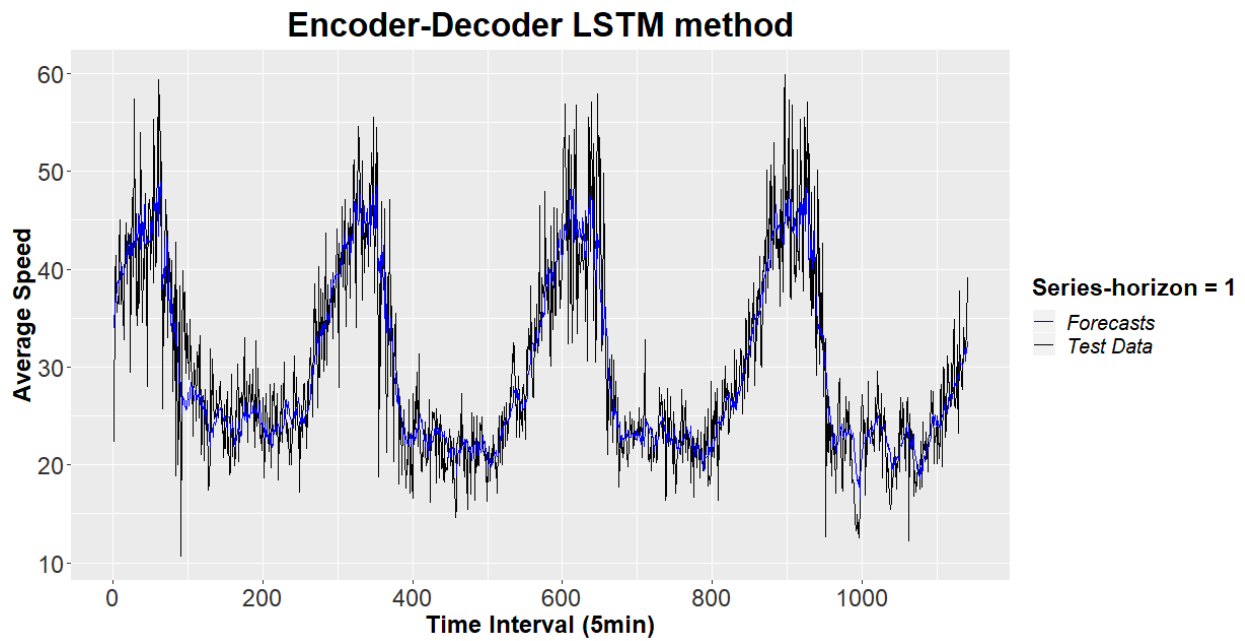
**Πίνακας 81** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder-Decoder LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.9	28.8	28.2	27.7	27.4	27.0	26.6	26.5	26.3	26.4	26.3	26.0
<b>RMSE</b>	30.5	29.7	29.4	29.0	28.8	28.6	28.4	28.3	28.1	28.4	28.3	28.0
<b>MAPE</b>	40.8	39.9	39.2	38.6	38.1	37.7	37.3	37.1	36.9	36.9	36.8	36.6
<b>sMAPE</b>	33.2	32.0	31.3	30.6	30.1	29.7	29.3	29.2	28.9	29.0	28.8	28.6
<b>MASE</b>	29.9	28.8	28.2	27.7	27.4	27.0	26.6	26.5	26.3	26.4	26.2	26.1

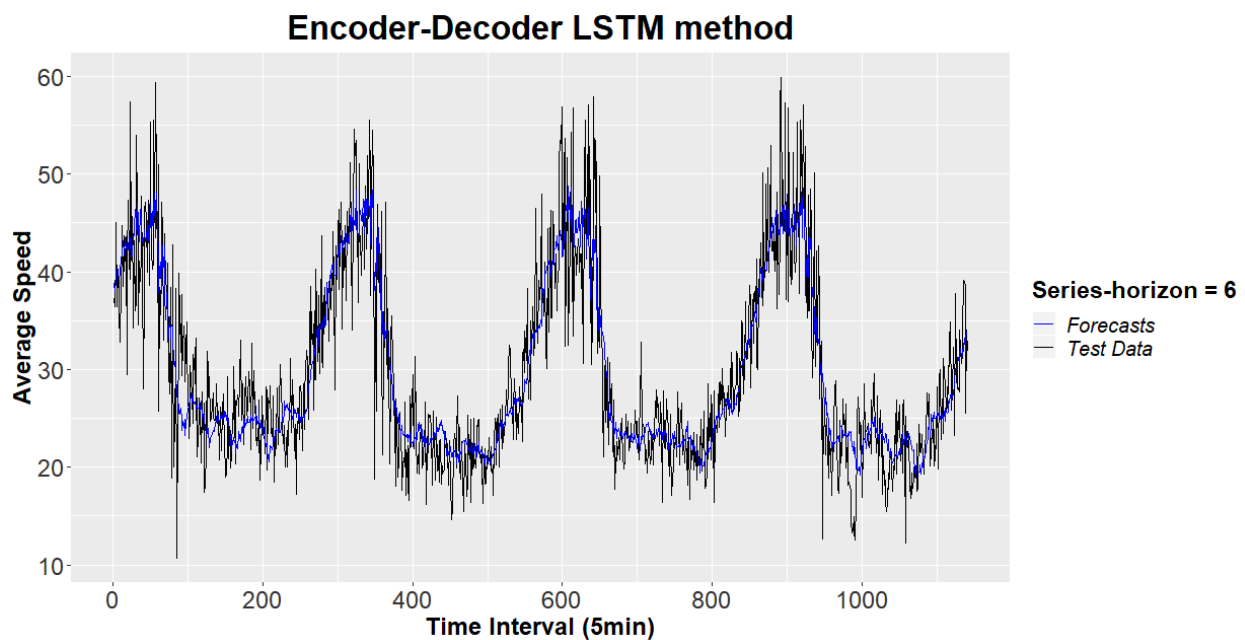
**Πίνακας 82** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder-Decoder LSTM ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.6	26.9	26.2	27.2
<b>RMSE</b>	29.7	28.5	28.2	28.8
<b>MAPE</b>	39.6	37.6	36.8	38.0
<b>sMAPE</b>	31.8	29.6	28.8	30.1
<b>MASE</b>	28.7	26.9	26.2	27.3

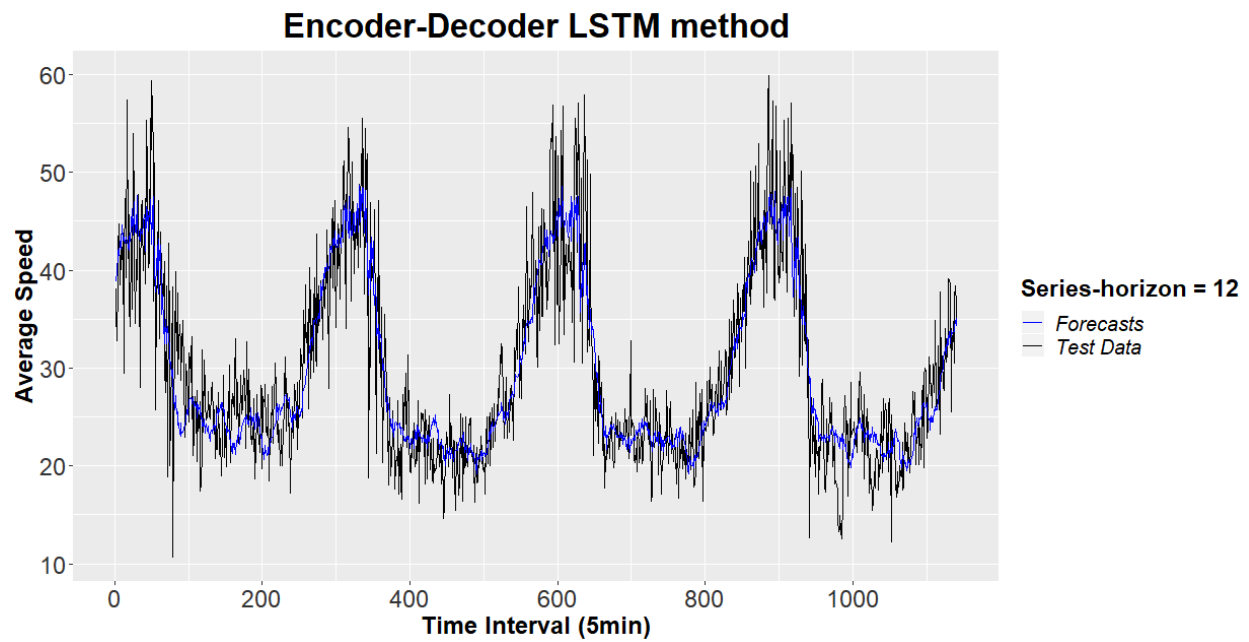
Στα Διαγράμματα 132, 133 και 134 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει δώσει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο αναφοράς, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1η ημέρα πρόβλεψης. Τέλος, από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 135) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



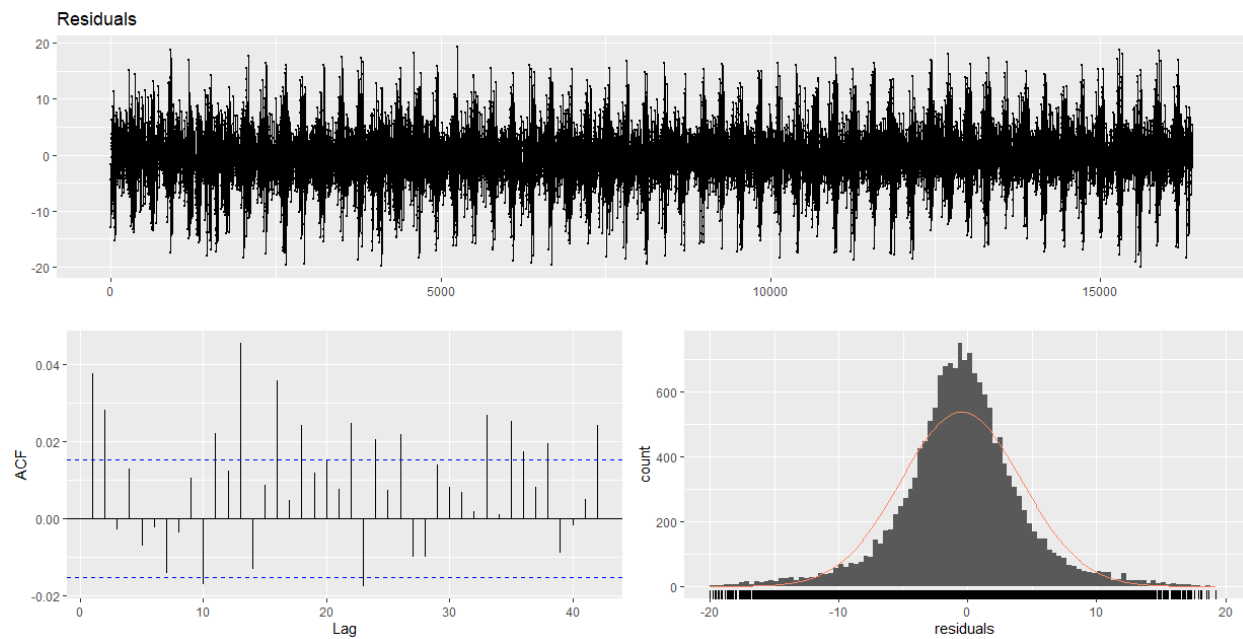
**Διάγραμμα 132** Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 133** Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 134** Προβλέψεις της μεθόδου Encoder-Decoder LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



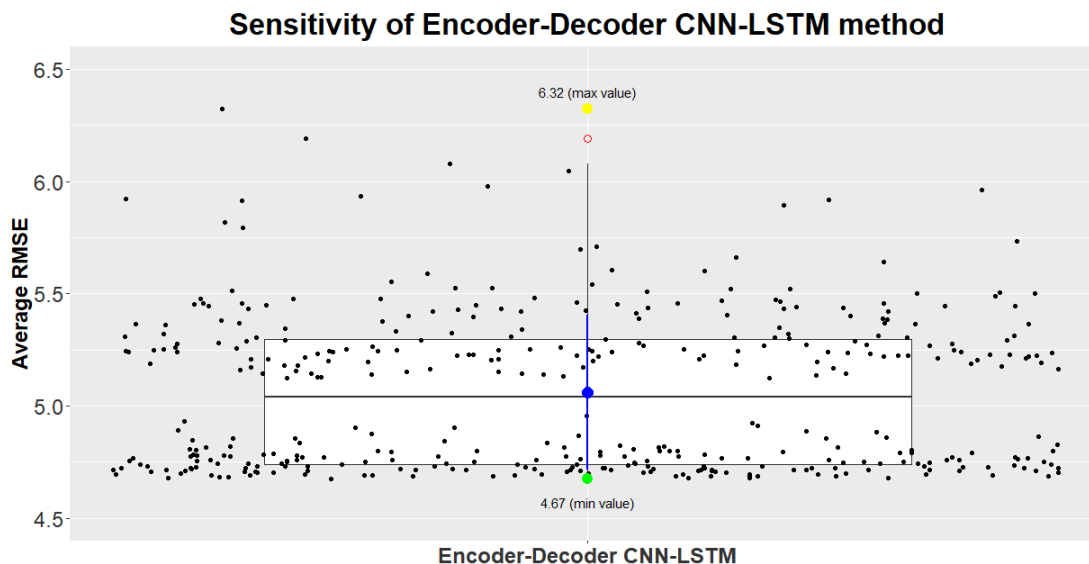
**Διάγραμμα 135** Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder LSTM.

### 6.3.9 Μέθοδος Encoder-Decoder CNN-LSTM

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

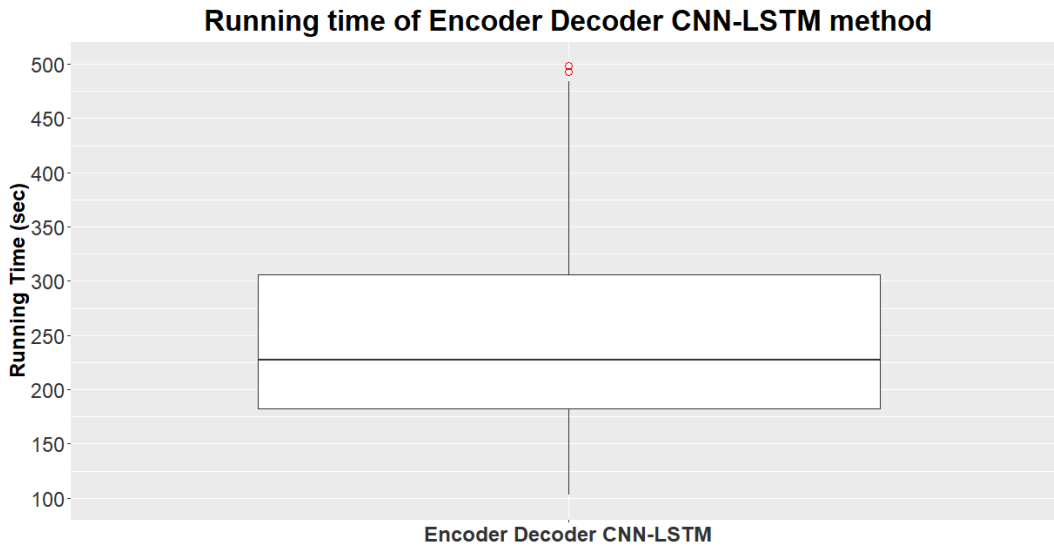
- **Order p:** 12
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** False
- **Block:** 1
- **Filters:** 32
- **Kernel:** 2
- **Layers:** 1
- **Nodes:** 100
- **Batch Size:** 128

με μέσο RMSE 4.67, αριθμό παραμέτρων 116777 και χρόνο εκπαίδευσης 182.16 δευτερόλεπτα. Στο Διάγραμμα 136 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.67 – 6.32). Παρατηρείται και πάλι η συγκέντρωση των παρατηρήσεων σε 2 περιοχές.

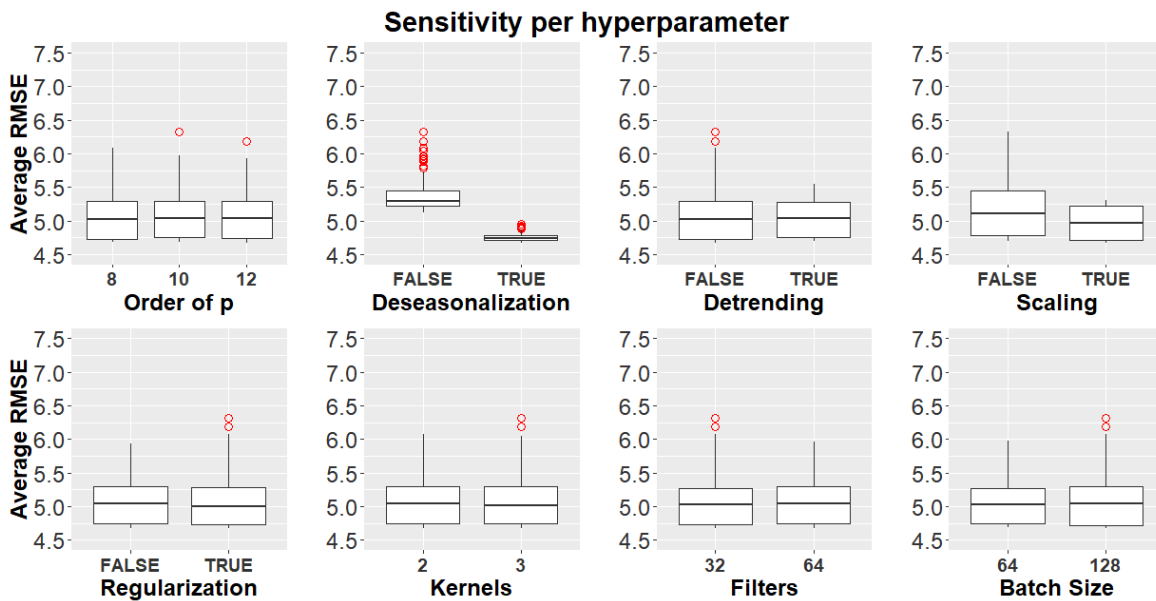


Διάγραμμα 136 Ευαισθησία της μεθόδου Encoder-Decoder CNN-LSTM.

Στο Διάγραμμα 137 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 102.73 δευτερόλεπτα έως 498.69 δευτερόλεπτα και μέση τιμή περί τα 248.9 δευτερόλεπτα. Στο Διάγραμμα 138 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαραμέτρο υπό μορφή boxplot.



**Διάγραμμα 137** Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder CNN-LSTM.



**Διάγραμμα 138** Ευαισθησία της μεθόδου Encoder-Decoder CNN-LSTM ανά υπερπαραμέτρο.

Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
window         1   0.01    0.01    0.666 0.41491
Filters         1   0.00    0.00    0.000 0.99610
Kernels         1   0.02    0.02    1.279 0.25890
Batch           1   0.01    0.01    0.418 0.51858
Deseasonalization 1 35.87   35.87 2236.689 < 2e-16 ***
Detrending      1   0.18    0.18   11.245 0.00088 ***
Scaling         1   3.12    3.12  194.830 < 2e-16 ***
Regularization  1   0.05    0.05    3.263 0.07165 .
Residuals     375   6.01    0.02
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Deseasonalization, Detrending και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 83.

**Πίνακας 83** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder-Decoder CNN-LSTM.

<b>(Intercept)</b>	5.448391e+00
<b>Window</b>	-3.230223e-03
<b>Filters</b>	1.977674e-06
<b>Kernels</b>	1.461513e-02
<b>Batch</b>	1.304989e-04
<b>DeseasonalizationTRUE</b>	-6.112999e-01
<b>DetrendingTRUE</b>	-4.334380e-02
<b>ScalingTRUE</b>	-1.804179e-01
<b>RegularizationTRUE</b>	2.334935e-02

Παρατηρούμε στον Πίνακα 84 ότι για ακόμη μία φορά η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 138. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι



5.36, ενώ με εποχική αποσύνθεση είναι 4.75. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 84 φαίνονται οι 5 μετρικές σφάλματος που χρησιμοποιήθηκαν ανά ορίζοντα πρόβλεψης. Στον Πίνακα 85 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 84** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder CNN-LSTM.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.403	3.417	3.446	3.481	3.511	3.528	3.532	3.543	3.550	3.544	3.558	3.556
<b>RMSE</b>	4.617	4.625	4.651	4.683	4.707	4.726	4.735	4.742	4.746	4.739	4.744	4.750
<b>MAPE</b>	11.574	11.630	11.756	11.916	12.051	12.134	12.158	12.210	12.233	12.230	12.283	12.290
<b>sMAPE</b>	11.452	11.533	11.664	11.821	11.951	12.021	12.041	12.085	12.109	12.100	12.160	12.160
<b>MASE</b>	0.701	0.666	0.669	0.672	0.676	0.673	0.670	0.665	0.658	0.648	0.643	0.637

**Πίνακας 85** Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder CNN-LSTM.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.437	3.529	3.552	3.506
<b>RMSE</b>	4.644	4.728	4.744	4.705
<b>MAPE</b>	11.719	12.138	12.259	12.039
<b>sMAPE</b>	11.617	12.025	12.132	11.925
<b>MASE</b>	0.677	0.671	0.647	0.665

**Πίνακας 86** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder CNN-LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
95576.467 sec	171.297 sec	21.715 sec	95769.479 sec	40580	2.684

Από τα αποτελέσματα συμπεραίνεται ότι η μέθοδος Encoder Decoder CNN-LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 86 φαίνονται οι μετρικές χρόνου καθώς και η υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα

εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 40580 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο αναφοράς. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 87 και 88. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.1%.

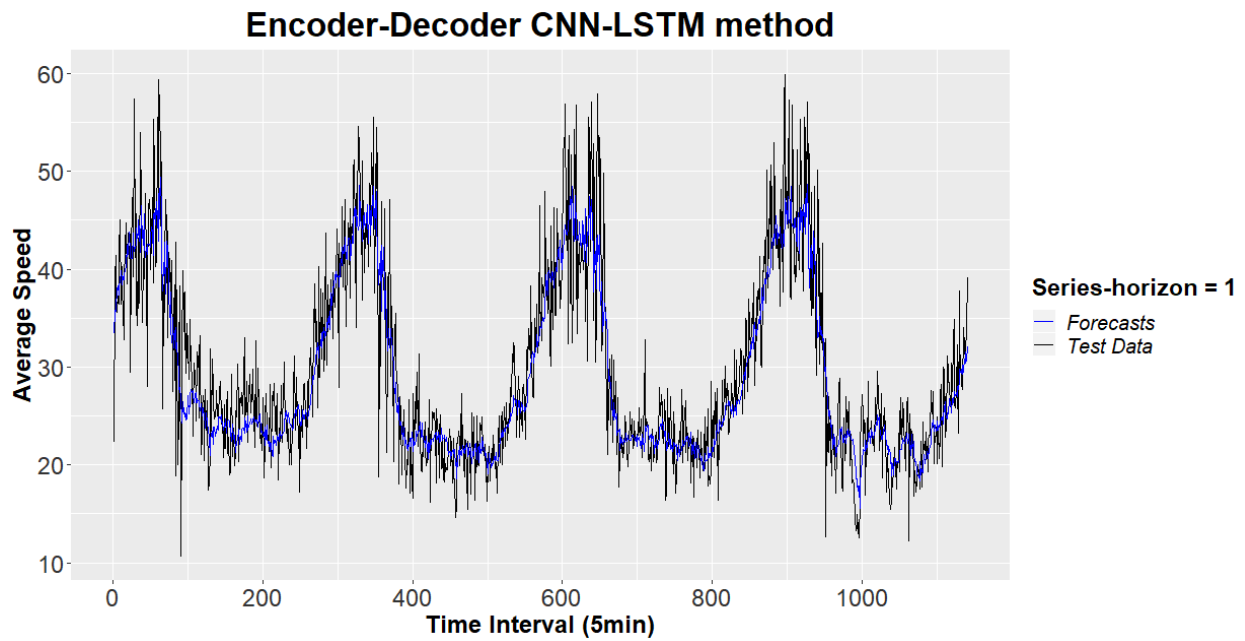
**Πίνακας 87** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder-Decoder CNN-LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	29.0	28.5	27.8	27.2	26.6	26.3	26.2	26.0	25.8	25.8	25.5	25.6
<b>RMSE</b>	29.7	29.3	28.8	28.4	28.1	27.8	27.7	27.6	27.5	27.5	27.4	27.4
<b>MAPE</b>	38.9	38.5	37.8	37.0	36.4	35.9	35.8	35.6	35.4	35.4	35.0	35.0
<b>sMAPE</b>	32.3	31.6	30.8	30.0	29.2	28.8	28.7	28.5	28.3	28.3	27.9	28.0
<b>MASE</b>	29.0	28.4	27.8	27.2	26.7	26.3	26.2	26.0	25.9	25.8	25.4	25.6

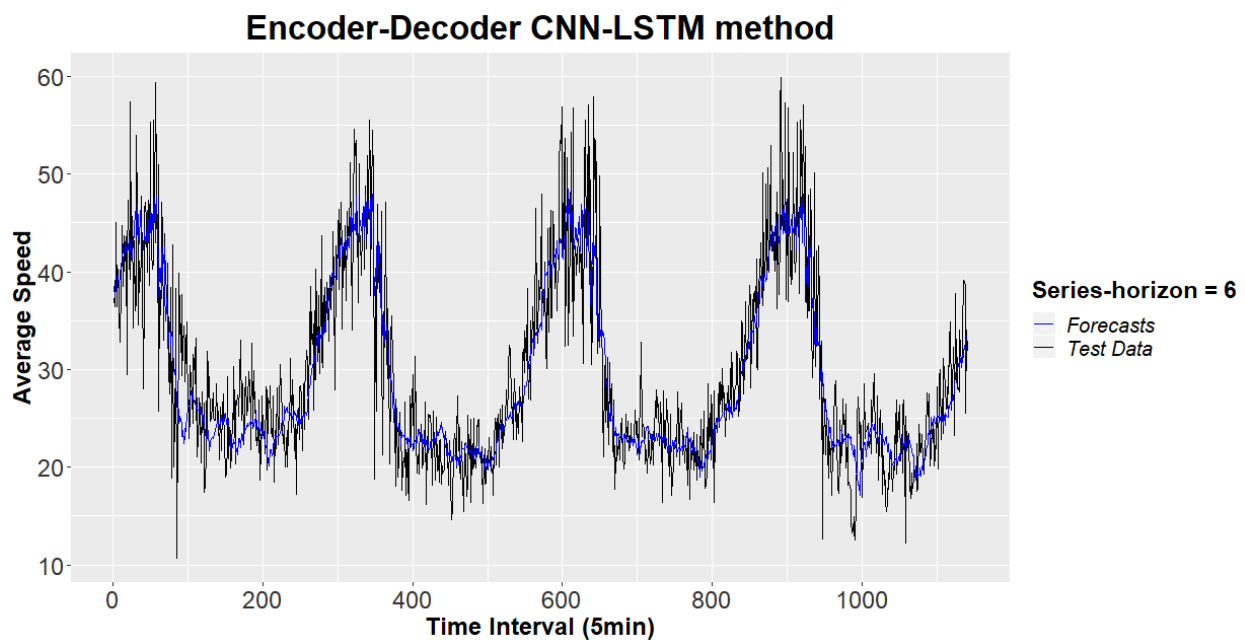
**Πίνακας 88** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder-Decoder CNN-LSTM ως προς την μέθοδο αναφοράς.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	28.1	26.3	25.7	26.7
<b>RMSE</b>	29.1	27.8	27.5	28.1
<b>MAPE</b>	38.0	35.9	35.2	36.4
<b>sMAPE</b>	31.1	28.8	28.1	29.4
<b>MASE</b>	28.1	26.3	25.7	26.7

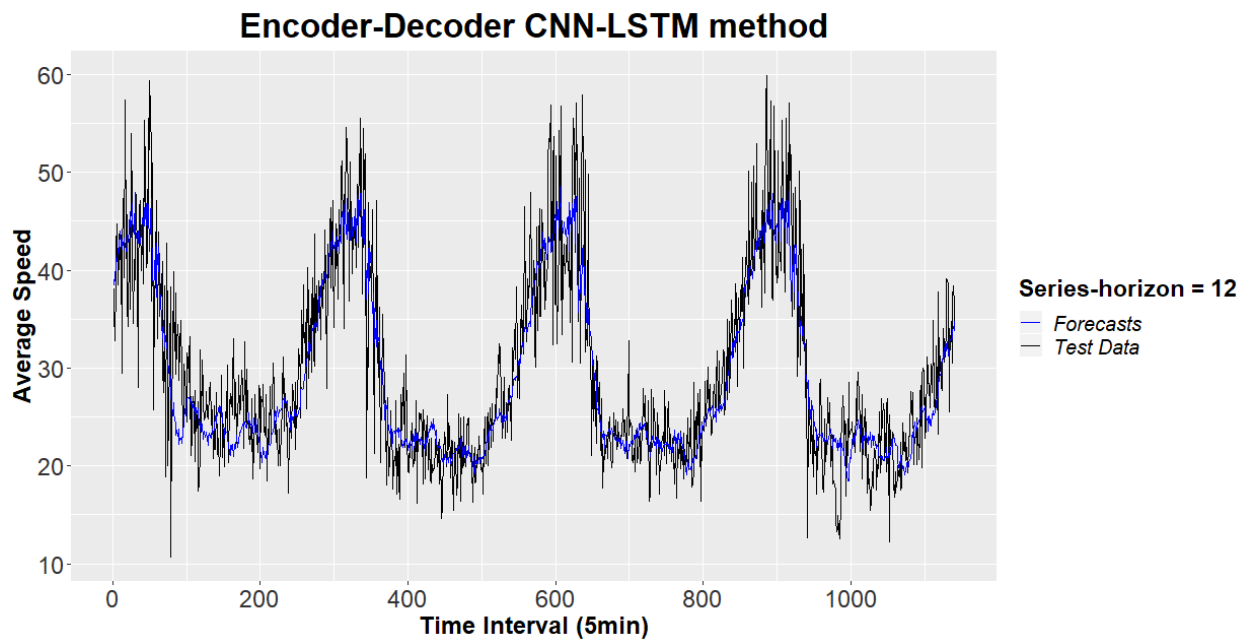
Στα Διαγράμματα 139, 140 και 141 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει κάνει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο βάσης, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1<sup>η</sup> ημέρα πρόβλεψης. Τέλος, από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 142) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



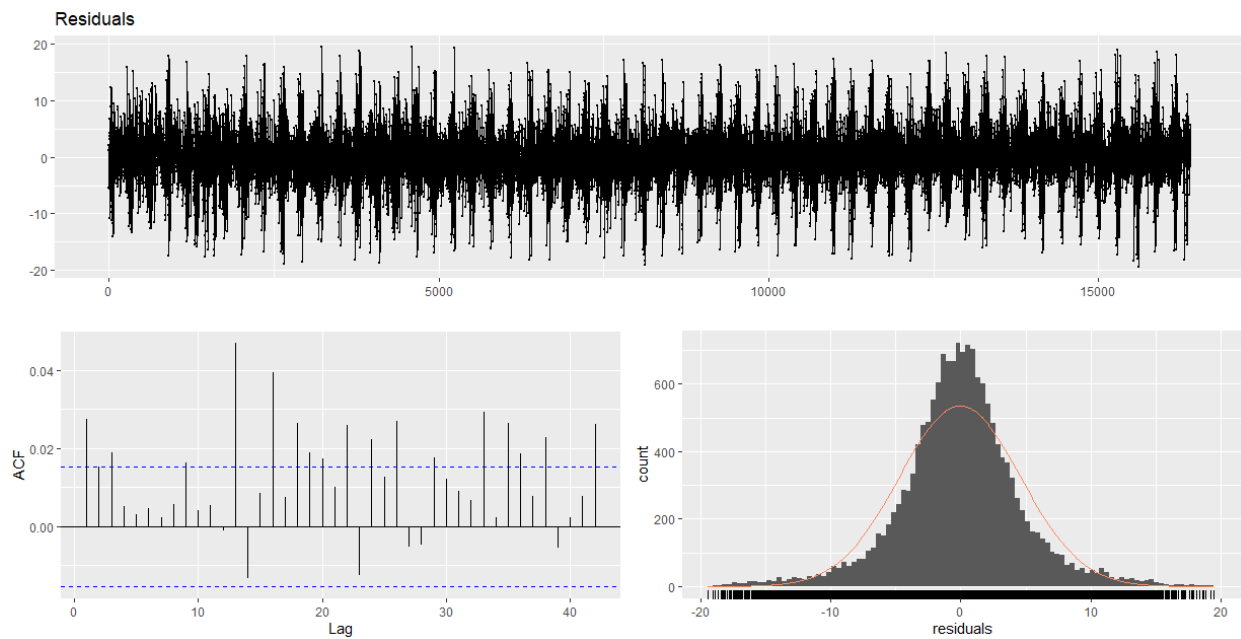
**Διάγραμμα 139** Προβλέψεις της μεθόδου Encoder-Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 140** Προβλέψεις της μεθόδου Encoder Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



**Διάγραμμα 141** Προβλέψεις της μεθόδου Encoder Decoder CNN-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



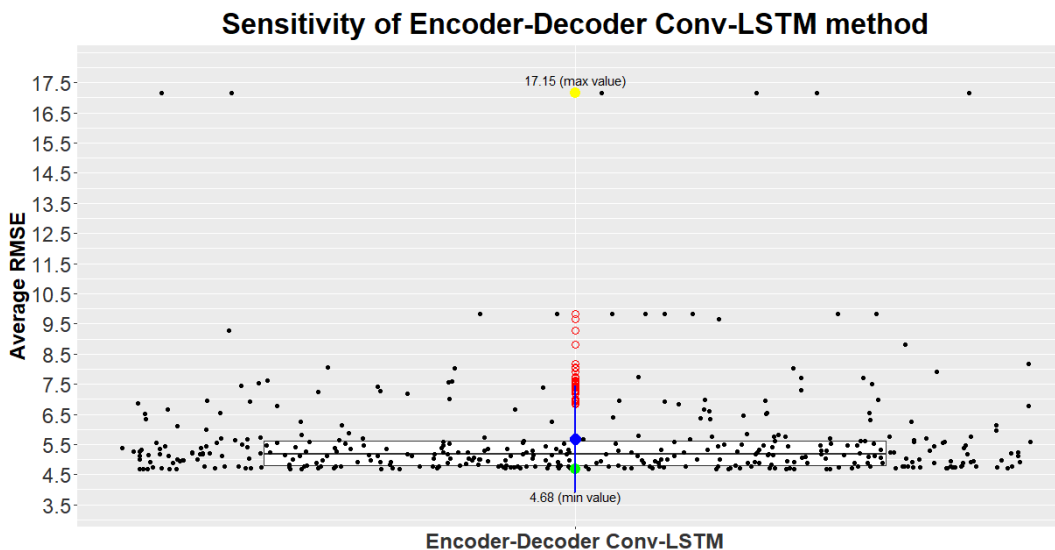
**Διάγραμμα 142** Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder CNN-LSTM.

### 6.3.10 Μέθοδος Encoder-Decoder Conv-LSTM

Από την βελτιστοποίηση των παραμέτρων το βέλτιστο μοντέλο που προκύπτει είναι το εξής (ως προς τις υπερπαραμέτρους):

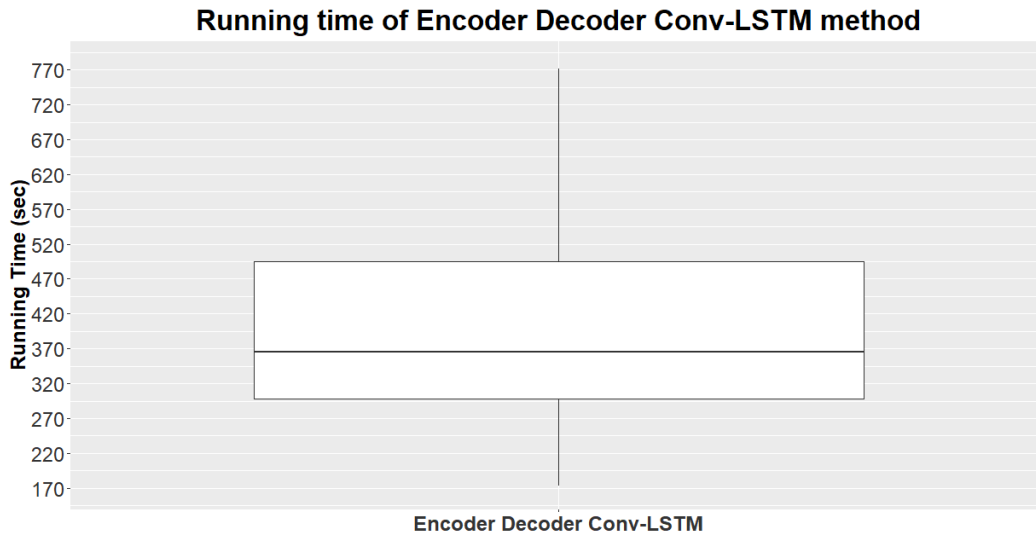
- **Order p:** 10
- **Deseasonalization:** True
- **Detrending:** False
- **Scaling:** True
- **Regularization:** False
- **Encoder Layers:** 1
- **Filters:** 32
- **Kernel:** 2
- **Decoder Layers:** 1
- **Nodes:** 100
- **Batch Size:** 128

με μέσο RMSE 4.68, αριθμό παραμέτρων 187177 και χρόνο εκπαίδευσης 305.93 δευτερόλεπτα. Στο Διάγραμμα 143 φαίνεται η ευαισθησία της μεθόδου στις υπερπαραμέτρους που ελέγχθηκαν. Το μέσο RMSE ανήκει στο διάστημα (4.68 – 17.15). Η εν λόγω μέθοδος έχει την μεγαλύτερη ευαισθησία συγκριτικά με τις μεθόδους της ομάδας 3.

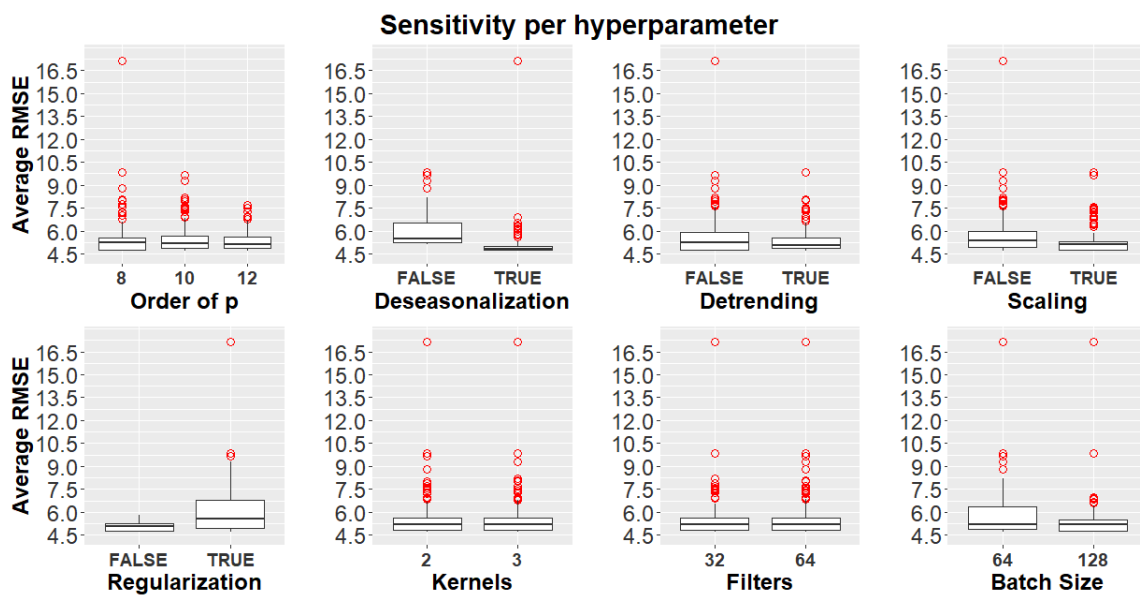


**Διάγραμμα 143** Ευαισθησία της μεθόδου Encoder Decoder Conv-LSTM.

Στο Διάγραμμα 144 φαίνεται ο χρόνος ανά παραμετροποίηση για την μέθοδο ο οποίος εκτείνεται από 172.92 δευτερόλεπτα έως 771.58 δευτερόλεπτα και μέση τιμή περί τα 398.42 δευτερόλεπτα. Στο Διάγραμμα 145 φαίνεται η ευαισθησία της μεθόδου ανά υπερπαραμέτρο υπό μορφή boxplot.



**Διάγραμμα 144** Χρόνος ανά παραμετροποίηση για την μέθοδο Encoder-Decoder Conv-LSTM.



**Διάγραμμα 145** Ευαισθησία της μεθόδου Encoder Decoder Conv-LSTM ανά υπερπαραμέτρο.

Διεξήχθηκε επίσης ανάλυση ANOVA για να διαπιστωθεί η στατιστική σημαντικότητα των υπερπαραμέτρων. Τα αποτελέσματα που επιστράφηκαν από την κονσόλα της R για την ανάλυση ANOVA για τις υπερπαραμέτρους που εξετάστηκαν είναι τα ακόλουθα:

```

              Df Sum Sq Mean Sq F value    Pr(>F)
Window         1   33.7    33.65   13.874 0.000225 ***
Filters         1    0.0     0.03    0.013 0.908048
Kernels         1    0.0     0.00    0.001 0.974238
Batch           1   16.0    15.99    6.594 0.010619 *
Deseasonalization 1   48.3    48.33   19.926 1.07e-05 ***
Detrending      1   17.1    17.07    7.038 0.008319 **
Scaling         1   43.4    43.39   17.890 2.95e-05 ***
Regularization  1  148.1   148.10  61.058 5.64e-14 ***
Residuals     375  909.6     2.43
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Για επίπεδο σημαντικότητας 0.01 στατιστικά σημαντικές υπερπαραμέτροι είναι οι Window, Deseasonalization, Detrending, Scaling και Regularization. Οι συντελεστές από την ανάλυση δίνονται στον Πίνακα 89.

**Πίνακας 89** Συντελεστές της ανάλυσης ANOVA για την μέθοδο Encoder Decoder Conv-LSTM.

<b>(Intercept)</b>	8.4178828141
<b>Window</b>	-0.1812876552
<b>Filters</b>	-0.0005741224
<b>Kernels</b>	-0.0051366056
<b>Batch</b>	-0.0063777831
<b>DeseasonalizationTRUE</b>	-0.7095500316
<b>DetrendingTRUE</b>	-0.4216959046
<b>ScalingTRUE</b>	-0.6723208528
<b>RegularizationTRUE</b>	1.2420745848

Παρατηρούμε στον Πίνακα 89 ότι για ακόμη μία φορά η πιο σημαντική υπερπαραμέτρος είναι η εποχική αποσύνθεση η οποία έχει το μεγαλύτερο συντελεστή άρα και την μεγαλύτερη επιρροή. Αυτό φαίνεται και σχηματικά στο Διάγραμμα 145. Η μέση τιμή χωρίς εποχική αποσύνθεση είναι

6.03, ενώ με εποχική αποσύνθεση είναι 5.31. Ακολουθούν τα αποτελέσματα για το βέλτιστο μοντέλο. Στον Πίνακα 90 φαίνονται οι 5 μετρικές σφάλματος ανά ορίζοντα πρόβλεψης. Στον πίνακα 91 δίνονται τα συνοπτικά αποτελέσματα τα οποία αφορούν στο μέσο βραχυπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μεσοπρόθεσμο σφάλμα (ανά μετρική), στο μέσο μακροπρόθεσμο σφάλμα (ανά μετρική) και στο μέσο σφάλμα (ανά μετρική).

**Πίνακας 90** Μετρικές σφάλματος ανά ορίζοντα για την μέθοδο Encoder-Decoder Conv-LSTM.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	3.415	3.434	3.451	3.480	3.514	3.523	3.537	3.546	3.540	3.539	3.546	3.555
<b>RMSE</b>	4.634	4.648	4.669	4.691	4.721	4.734	4.749	4.754	4.745	4.737	4.743	4.749
<b>MAPE</b>	11.638	11.701	11.797	11.938	12.073	12.126	12.188	12.228	12.219	12.226	12.267	12.301
<b>sMAPE</b>	11.492	11.578	11.675	11.815	11.948	11.992	12.046	12.082	12.068	12.073	12.117	12.148
<b>MASE</b>	0.703	0.670	0.670	0.672	0.676	0.672	0.671	0.666	0.657	0.647	0.640	0.636

**Πίνακας 91** Συνοπτικά μέτρα σφάλματος για την μέθοδο Encoder-Decoder Conv-LSTM.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	3.445	3.530	3.545	3.507
<b>RMSE</b>	4.661	4.740	4.743	4.715
<b>MAPE</b>	11.768	12.154	12.253	12.058
<b>sMAPE</b>	11.640	12.017	12.101	11.919
<b>MASE</b>	0.679	0.671	0.645	0.665

**Πίνακας 92** Μετρικές χρόνου, υπολογιστικής πολυπλοκότητας και προσαρμογής του μοντέλου για την μέθοδο Encoder-Decoder Conv-LSTM.

<b>Tuning Time</b>	<b>Train Time</b>	<b>Prediction Time</b>	<b>Total Time</b>	<b>CC</b>	<b>Model Fitting</b>
152991.459 sec	355.595 sec	25.195 sec	153372.249 sec	64988	2.694

Παρατηρούμε ότι η μέθοδος Encoder Decoder Conv-LSTM είναι καλύτερη από την μέθοδο Seasonal Naive για όλες τις μετρικές και για όλους τους ορίζοντες πρόβλεψης και ακολούθως για όλες τις μετρικές σύνοψης. Στον Πίνακα 92 φαίνονται οι μετρικές χρόνου καθώς και η



υπολογιστική πολυπλοκότητα (CC) και η προσαρμογή του μοντέλου (MF) στα δεδομένα εκπαίδευσης. Η εν λόγω μέθοδος χρειάστηκε περίπου 64988 φορές παραπάνω υπολογιστικό χρόνο σε σχέση με την μέθοδο βάσης. Οι σχετικές βελτιώσεις του σφάλματος φαίνονται στους Πίνακες 93 και 94. Η μείωση που επιτεύχθηκε στο μέσο RMSE είναι της τάξεως του 28.0%.

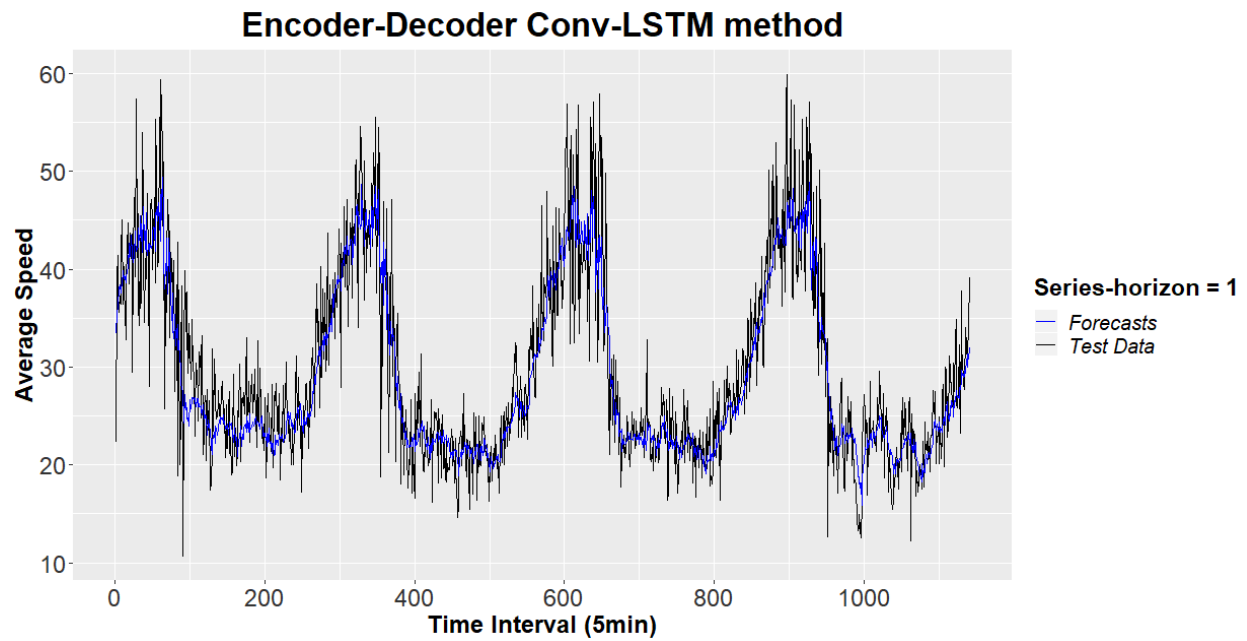
**Πίνακας 93** Ποσοστιαία μείωση για όλες τις μετρικές σφάλματος ανά ορίζοντα πρόβλεψης για την μέθοδο Encoder Decoder Conv-LSTM ως προς την μέθοδο αναφοράς.

	<b>h=1</b>	<b>h=2</b>	<b>h=3</b>	<b>h=4</b>	<b>h=5</b>	<b>h=6</b>	<b>h=7</b>	<b>h=8</b>	<b>h=9</b>	<b>h=10</b>	<b>h=11</b>	<b>h=12</b>
<b>MAE</b>	28.7	28.1	27.7	27.2	26.6	26.4	26.1	25.9	26.1	25.9	25.7	25.6
<b>RMSE</b>	29.5	28.9	28.5	28.3	27.9	27.7	27.4	27.4	27.5	27.6	27.4	27.4
<b>MAPE</b>	38.6	38.1	37.6	36.9	36.2	36.0	35.7	35.5	35.5	35.4	35.1	35.0
<b>sMAPE</b>	32.0	31.3	30.7	30.0	29.3	29.0	28.7	28.5	28.6	28.5	28.2	28.0
<b>MASE</b>	28.7	28.1	27.7	27.2	26.6	26.4	26.1	26.0	26.1	25.9	25.7	25.6

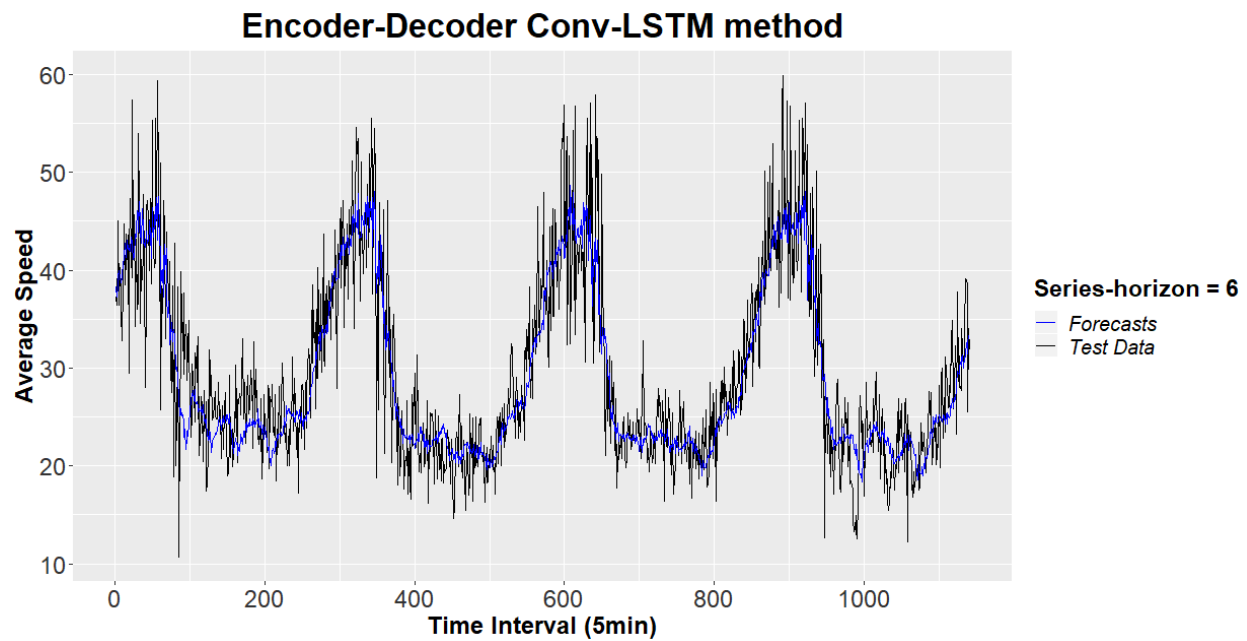
**Πίνακας 94** Ποσοστιαία μείωση των συνοπτικών μέτρων σφάλματος για την μέθοδο Encoder Decoder Conv-LSTM ως προς την μέθοδο βάσης.

	<b>Short</b>	<b>Medium</b>	<b>Long</b>	<b>Average</b>
<b>MAE</b>	27.9	26.2	25.8	26.7
<b>RMSE</b>	28.8	27.6	27.5	28.0
<b>MAPE</b>	37.8	35.8	35.2	36.3
<b>sMAPE</b>	31.0	28.9	28.3	29.4
<b>MASE</b>	27.9	26.2	25.8	26.7

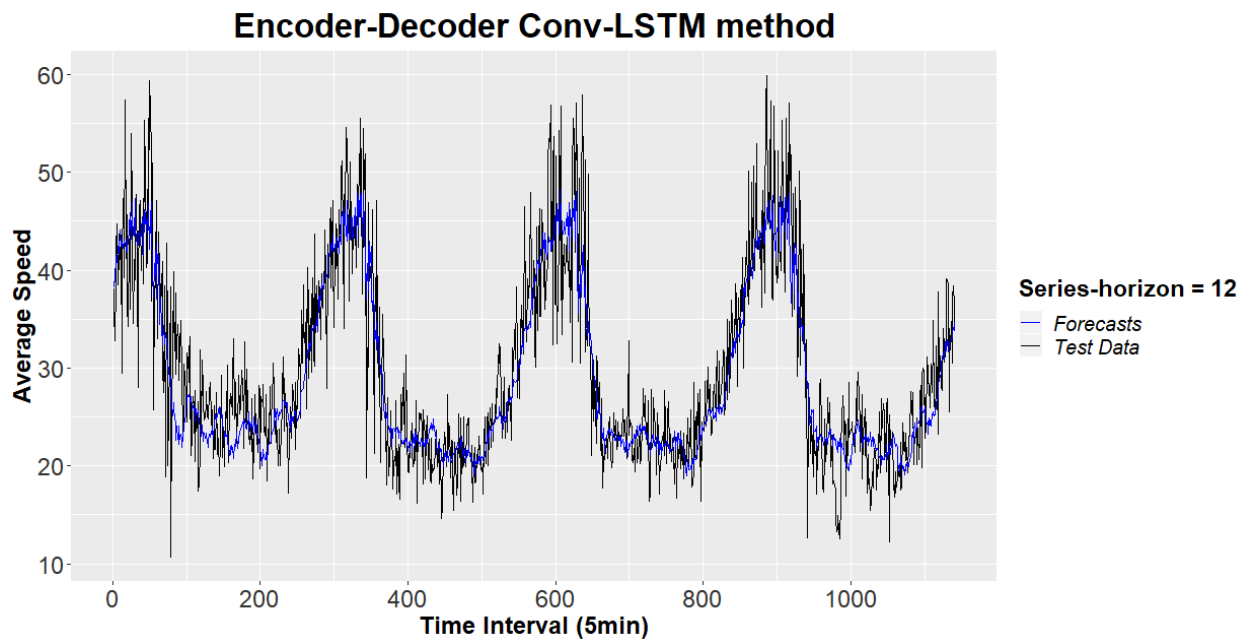
Στα Διαγράμματα 146, 147 και 148 φαίνονται ενδεικτικά οι προβλέψεις της μεθόδου για ορίζοντες πρόβλεψης 1, 6 και 12. Η μέθοδος όπως φαίνεται έχει κάνει πολύ καλύτερες προβλέψεις σε σχέση με την μέθοδο βάσης, και οι προβλέψεις έχουν πολύ μικρότερες διακυμάνσεις. Για ορίζοντα πρόβλεψης ίσο με 12 η πρόβλεψη υποεκτιμά το επίπεδο της χρονοσειράς για την 1<sup>η</sup> ημέρα πρόβλεψης. Τέλος, από το Διάγραμμα αυτοσυσχέτισης (κάτω αριστερά-Διάγραμμα 149) των καταλοίπων συμπεραίνεται ότι ίσως έχει μείνει πληροφορία στα κατάλοιπα που δεν μπορεί να εξηγήσει η μέθοδος.



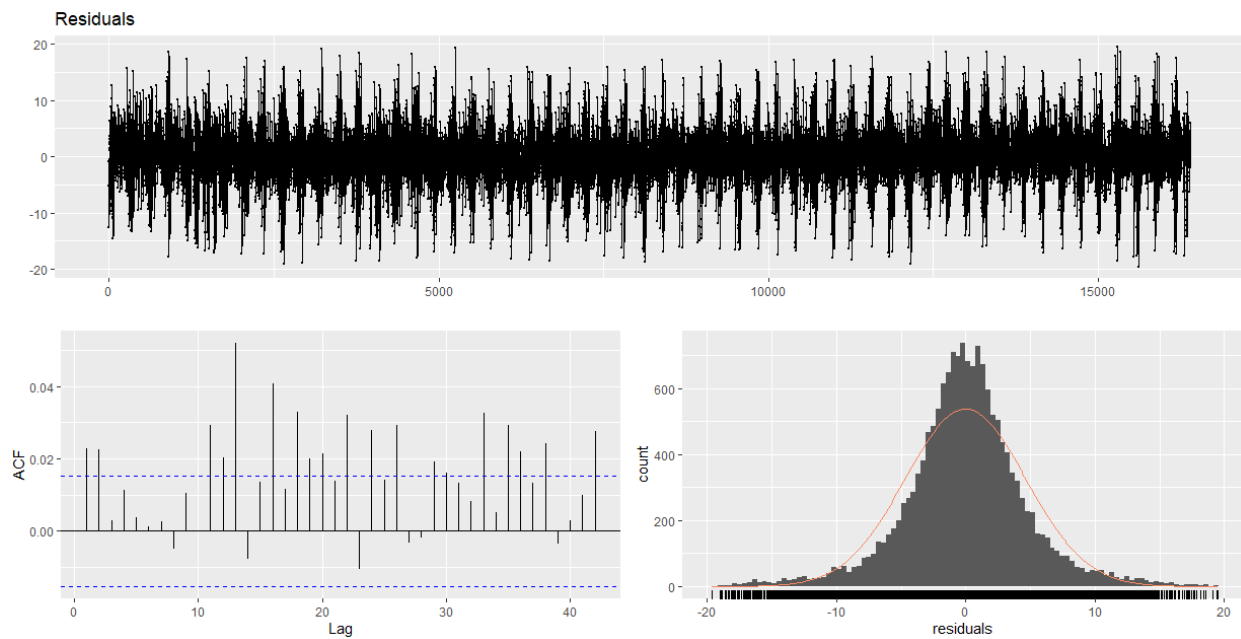
**Διάγραμμα 146** Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 147** Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 6.



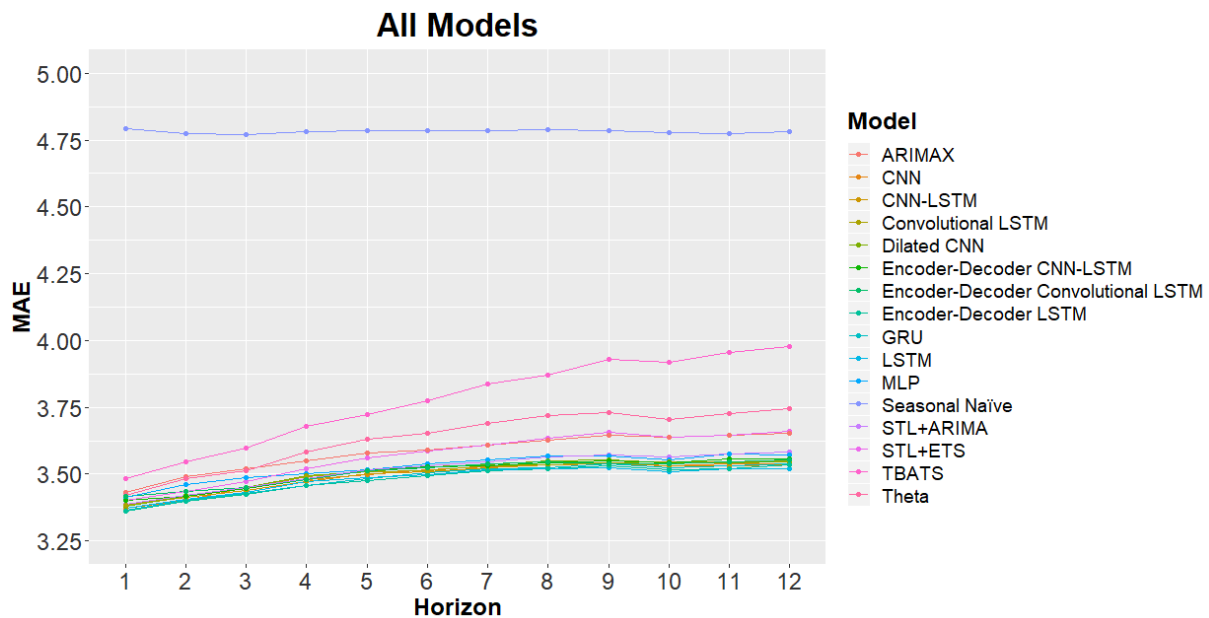
**Διάγραμμα 148** Προβλέψεις της μεθόδου Encoder Decoder Conv-LSTM για το σύνολο ελέγχου για ορίζοντα πρόβλεψης ίσο με 12.



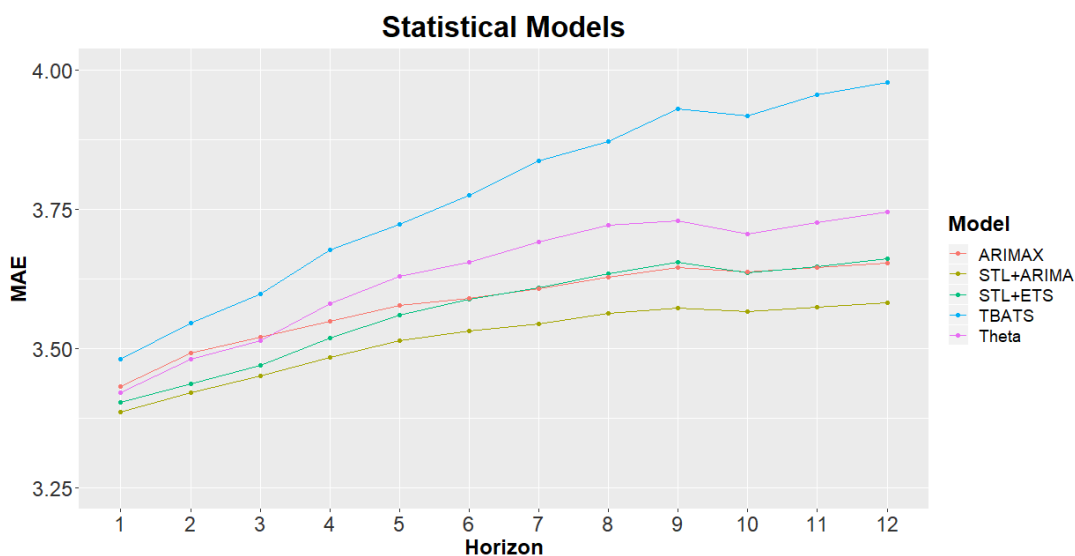
**Διάγραμμα 149** Γραφικός έλεγχος των καταλοίπων της μεθόδου Encoder-Decoder Conv-LSTM.

## 6.4 ΣΥΓΚΡΙΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

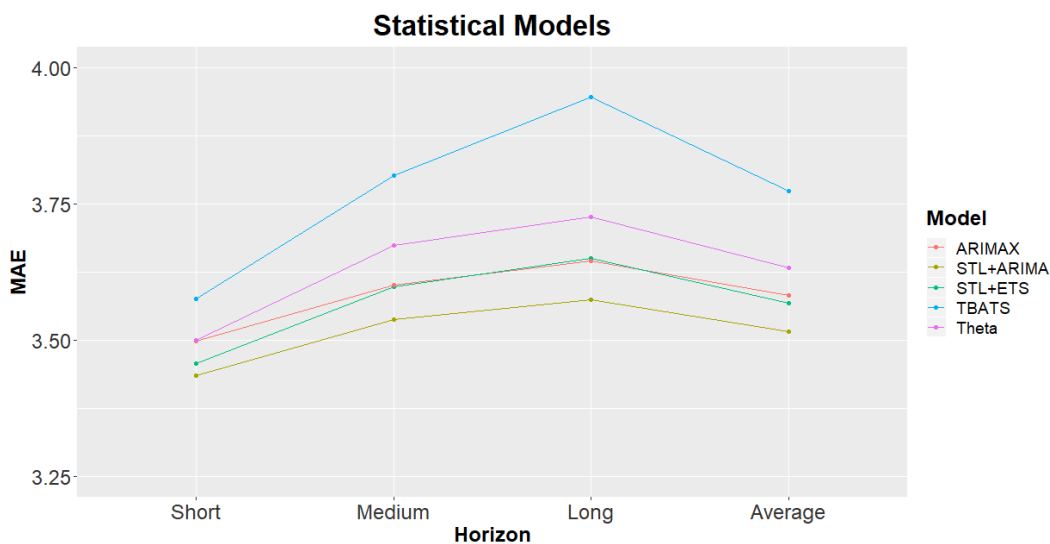
Στο Διάγραμμα 150 που δίνεται για την μετρική MAE φαίνεται ότι όλες οι μέθοδοι έχουν μικρότερο σφάλμα από την μέθοδο αναφοράς. Όποτε όλα τα μοντέλα γίνονται αποδεκτά. Από το Διάγραμμα 151 συμπεραίνεται ότι το μοντέλο STL+ARIMA είναι το καλύτερο μοντέλο με βάση την μετρική MAE για όλους τους ορίζοντες πρόβλεψης από όλα τα στατιστικά μοντέλα της ομάδας 2. Από τα στατιστικά μοντέλα το χειρότερο είναι το TBATS για όλους τους ορίζοντες πρόβλεψης. Παρατηρούμε ότι εκτός του μοντέλου TBATS όλα τα υπόλοιπα φαίνεται να έχουν την ίδια κλίση όσον αφορά την χειροτέρευση του MAE καθώς αυξάνεται ο ορίζοντας πρόβλεψης. Όπως είναι φυσιολογικό η ίδια εικόνα διατηρείται και για τα συνοπτικά μέτρα για την μετρική MAE, όπως φαίνεται και στο Διάγραμμα 152. Επιπροσθέτως, όσο αυξάνεται ο ορίζοντας πρόβλεψης το MAE αυξάνεται, καθώς αυξάνεται η αβεβαιότητα μας για το μέλλον, για όλα τα μοντέλα. Τέλος, αυτή η μονοτονία ως προς την αύξηση του MAE καθώς αυξάνει ο ορίζοντας πρόβλεψης, «σπάει» για ορίζοντα πρόβλεψης ίσο με 10.



Διάγραμμα 150 Μετρική MAE ανά ορίζοντα για όλες τις μεθόδους.

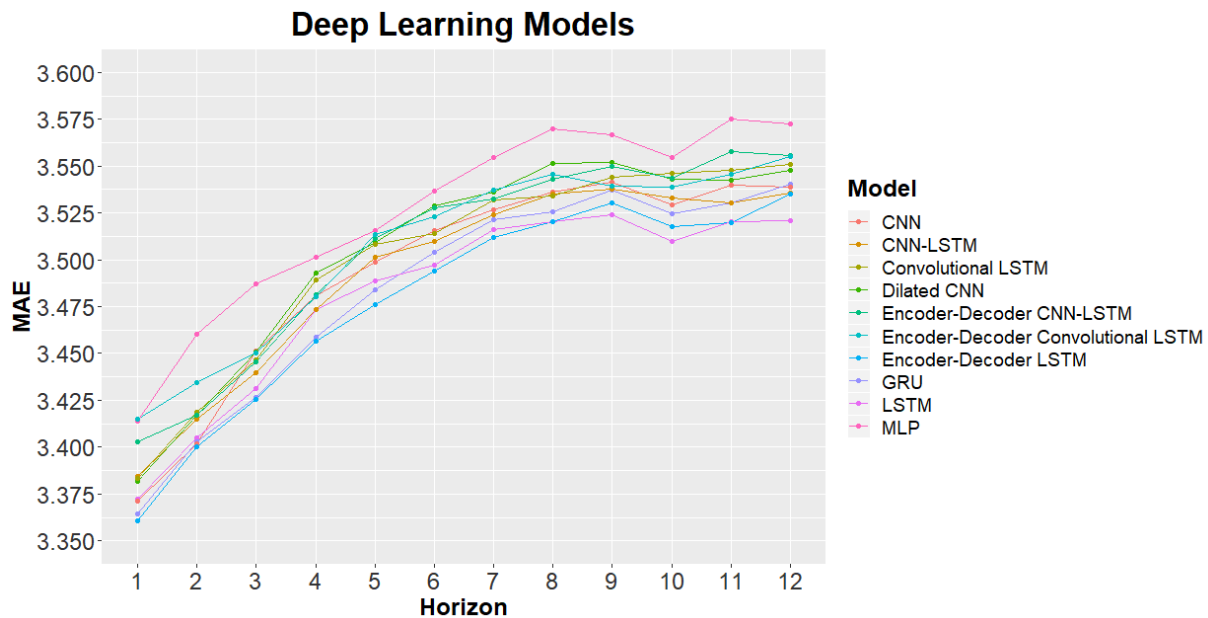


**Διάγραμμα 151** Μετρική MAE ανά ορίζοντα για τις στατιστικές μεθόδους.



**Διάγραμμα 152** Συνοπτικά μέτρα για την μετρική σφάλματος MAE για τις στατιστικές μεθόδους.

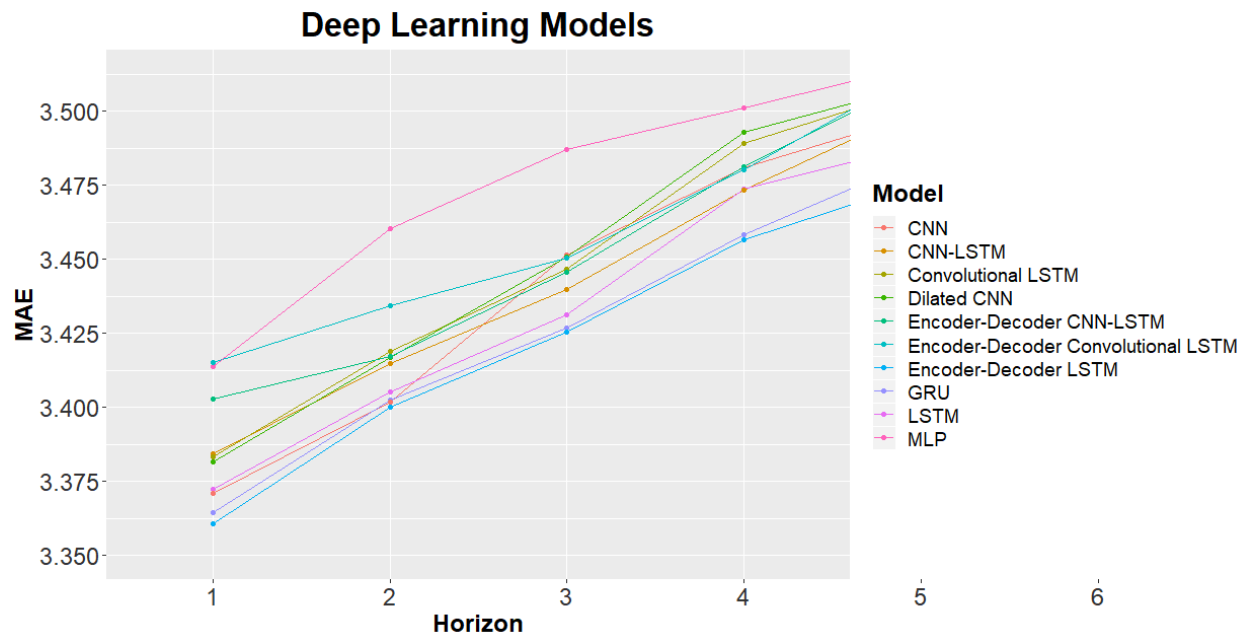
Στο Διάγραμμα 153 φαίνεται η μετρική MAE ανά ορίζοντα πρόβλεψης για τις μεθόδους βαθιάς μηχανικής μάθησης (ομάδα 3). Το χειρότερο μοντέλο είναι το MLP για όλους τους ορίζοντες πρόβλεψης εκτός του ορίζοντα πρόβλεψης βήματος 1 για τον οποίο το μεγαλύτερο MAE είναι οριακά αυτό που αντιστοιχεί στην μέθοδο Encoder-Decoder Convolutional LSTM.



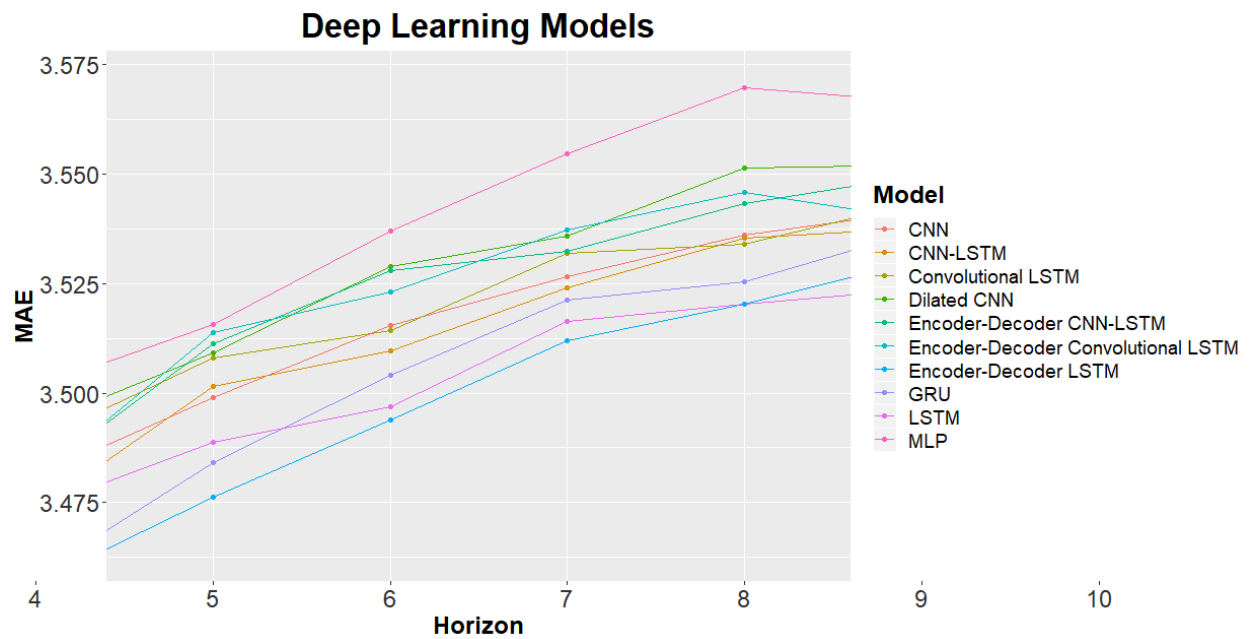
**Διάγραμμα 153** Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.

Πολλές φορές το μοντέλο MLP θεωρείται ως η μέθοδος αναφοράς των μεθόδων βαθιάς μηχανικής μάθησης και παρατηρούμε ότι με βάση αυτό το κριτήριο όλες οι εκλεπτυσμένες μέθοδοι που χρησιμοποιήθηκαν είναι καλύτερες σε επιδόσεις όσον αφορά στην πρόβλεψη, και άρα είναι αποδεκτές.

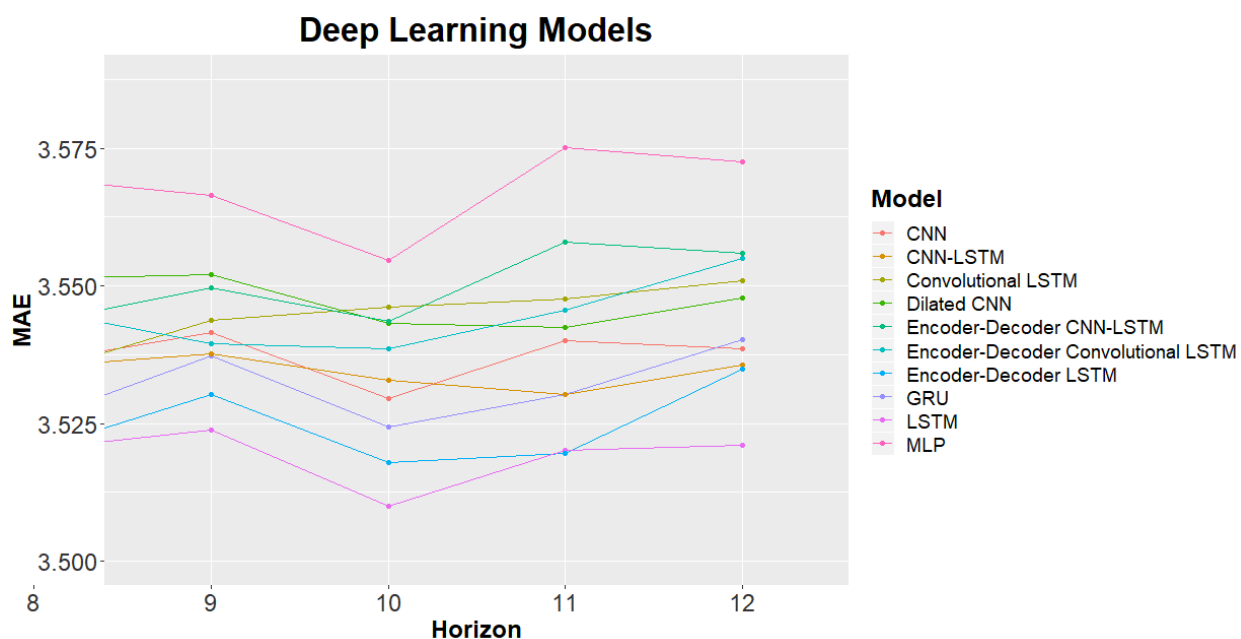
Τα Διαγράμματα 154, 155, 156 που ακολουθούν αποτελούν μεγέθυνση του Διαγράμματος 153 για τις βραχυπρόθεσμες (1, 2, 3, 4 βήματα), τις μεσοπρόθεσμες (5, 6, 7, 8 βήματα) και τις μακροπρόθεσμες τιμές (9, 10, 11, 12 βήματα) του ορίζοντα πρόβλεψης αντίστοιχα για καλύτερη εποπτεία των αποτελεσμάτων της ομάδας 3. Πιο ξεκάθαρη είναι η εικόνα στο Διάγραμμα 157, όπου φαίνονται τα συνοπτικά μέτρα για την μετρική σφάλματος MAE για τις μεθόδους βαθιάς μηχανικής μάθησης. Εκεί φαίνεται ξεκάθαρα ότι το χειρότερο μοντέλο είναι το MLP. Ενώ το καλύτερο μοντέλο για βραχυπρόθεσμη και μεσοπρόθεσμη πρόβλεψη είναι το Encoder-Decoder LSTM, ενώ για μακροπρόθεσμη πρόβλεψη είναι το LSTM. Όσον αφορά όμως το μέσο όρο του MAE για όλους τους ορίζοντες πρόβλεψης το καλύτερο μοντέλο είναι το Encoder-Decoder LSTM.



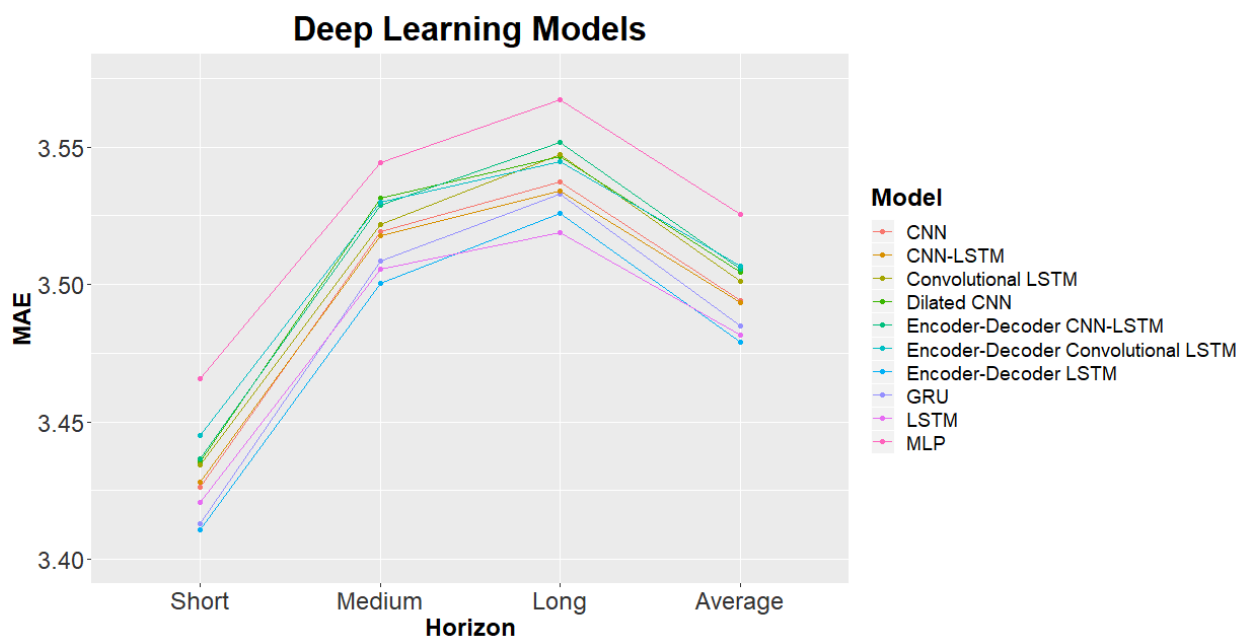
**Διάγραμμα 154** Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 πρώτες τιμές του ορίζοντα πρόβλεψης.



**Διάγραμμα 155** Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 μεσαίες τιμές του ορίζοντα πρόβλεψης.

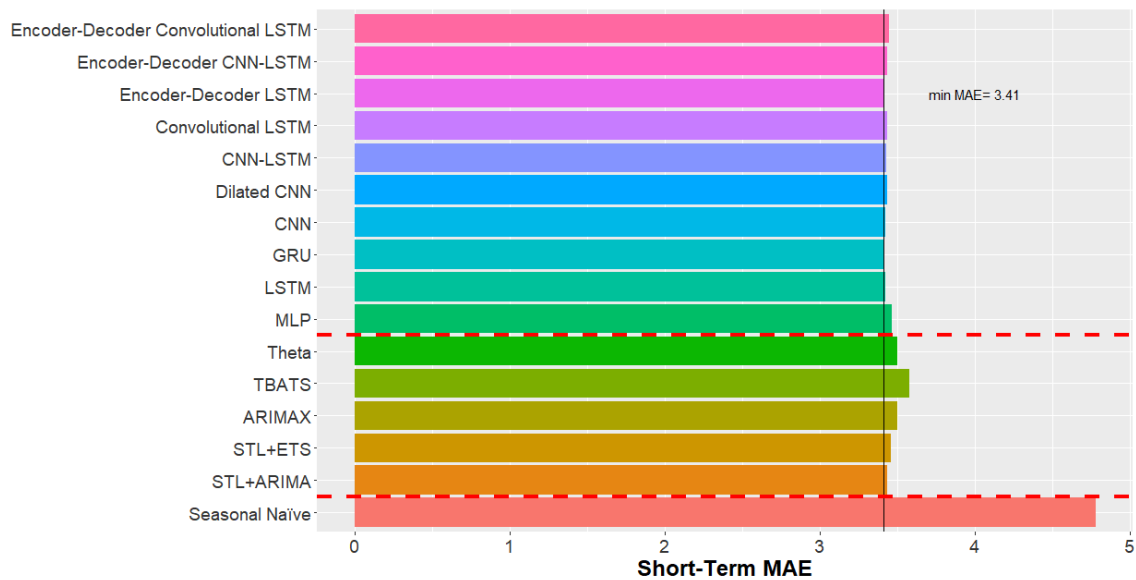


**Διάγραμμα 156** Μετρική MAE ανά ορίζοντα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης, μεγέθυνση στις 4 τελευταίες τιμές του ορίζοντα πρόβλεψης.



**Διάγραμμα 157** Συνοπτικά μέτρα για την μετρική σφάλματος MAE για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.

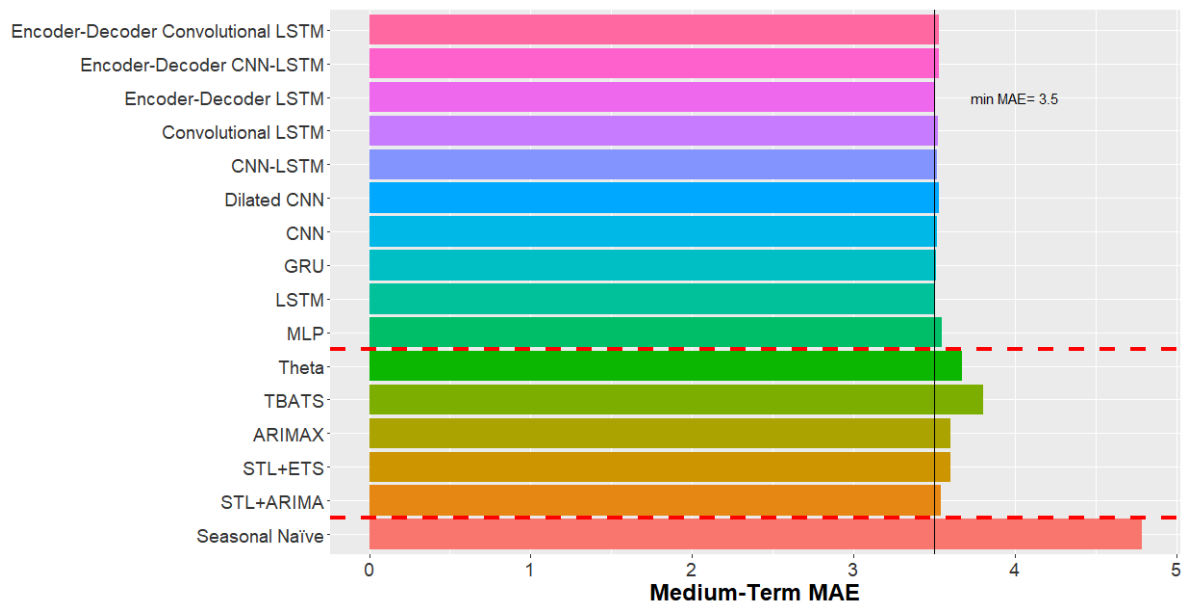




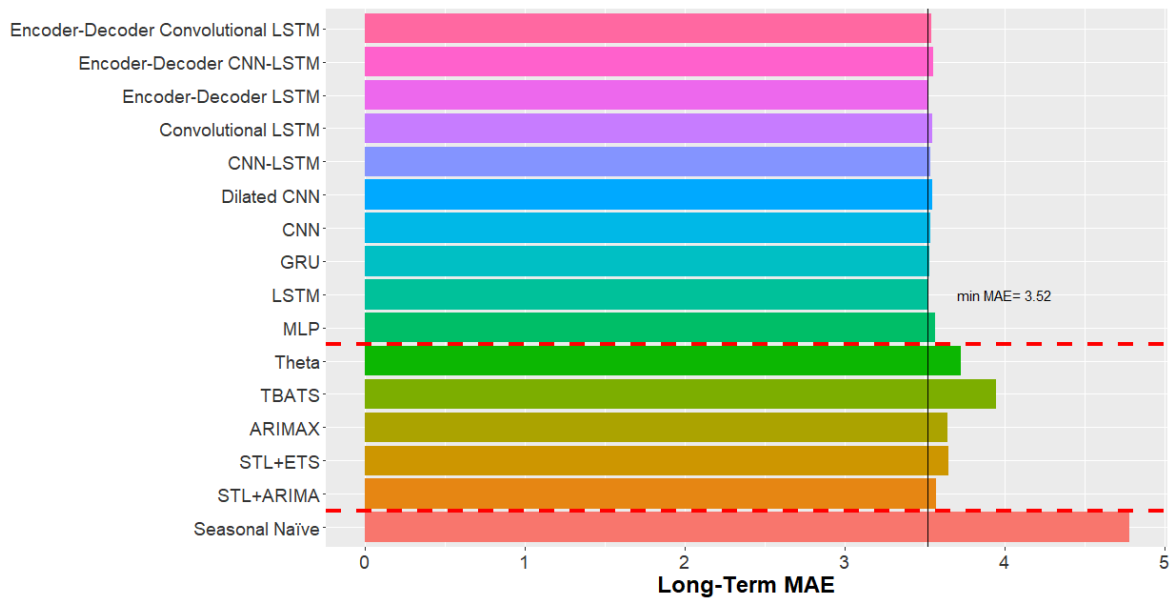
**Διάγραμμα 158** Μετρική MAE για βραχυπρόθεσμο μέσο ορίζοντα ανά μέθοδο.

Στο Διάγραμμα 158 φαίνεται η μετρική MAE για τον βραχυπρόθεσμο μέσο ορίζοντα ανά μέθοδο. Παρατηρούμε ότι η μέθοδος Encoder-Decoder LSTM έχει το μικρότερο βραχυπρόθεσμο MAE με τιμή 3.41 από όλες τις μεθόδους, με τις μεθόδους της ομάδας 3 να είναι κοντά εκτός την μέθοδο MLP, και την μέθοδο STL+ARIMA από τις μεθόδους της ομάδας 2.

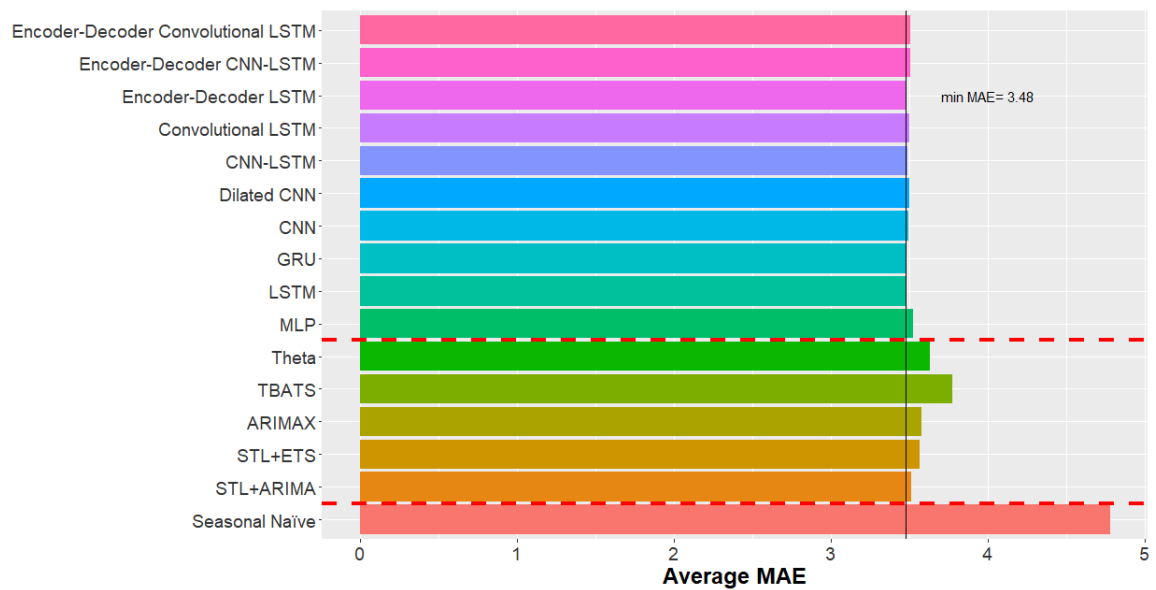
Στο Διάγραμμα 159 φαίνεται η μετρική MAE για τον μεσοπρόθεσμο μέσο ορίζοντα ανά μέθοδο. Παρατηρούμε ότι και πάλι η μέθοδος Encoder-Decoder LSTM έχει το μικρότερο μεσοπρόθεσμο MAE από όλες τις μεθόδους με τιμή 3.5 (έχουμε εύλογα αύξηση του σφάλματος καθώς αυξάνεται ο ορίζοντας) και η διαφορά με την καλύτερη μέθοδο από την ομάδα 2 να αυξάνεται. Τέλος, στο Διάγραμμα 160 φαίνεται η μετρική MAE για τον μακροπρόθεσμο μέσο ορίζοντα ανά μέθοδο. Σε αυτήν την περίπτωση η μέθοδος LSTM έχει το μικρότερο μακροπρόθεσμο MAE από όλες τις μεθόδους με τιμή 3.52 (έχουμε εύλογα αύξηση του σφάλματος καθώς αυξάνεται ο ορίζοντας) και η διαφορά με την καλύτερη μέθοδο από την ομάδα 2 (STL+ARIMA) αυξάνεται ακόμα περισσότερο. Να σχολιαστεί σε αυτό το σημείο ότι για τα αναλυτικά διαγράμματα έγινε χρήση της μετρικής σφάλματος MAE καθώς αποτελεί ίσως την πιο φυσικά ερμηνεύσιμη μετρική.



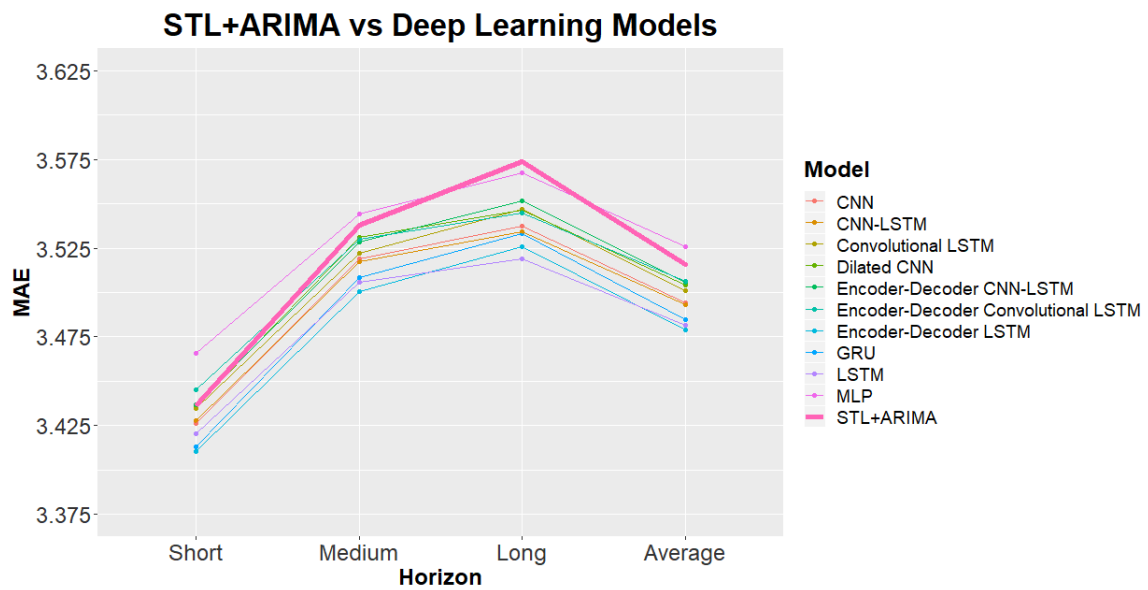
**Διάγραμμα 159** Μετρική MAE για μεσοπρόθεσμο μέσο ορίζοντα ανά μέθοδο.



**Διάγραμμα 160** Μετρική MAE για μακροπρόθεσμο μέσο ορίζοντα ανά μέθοδο.



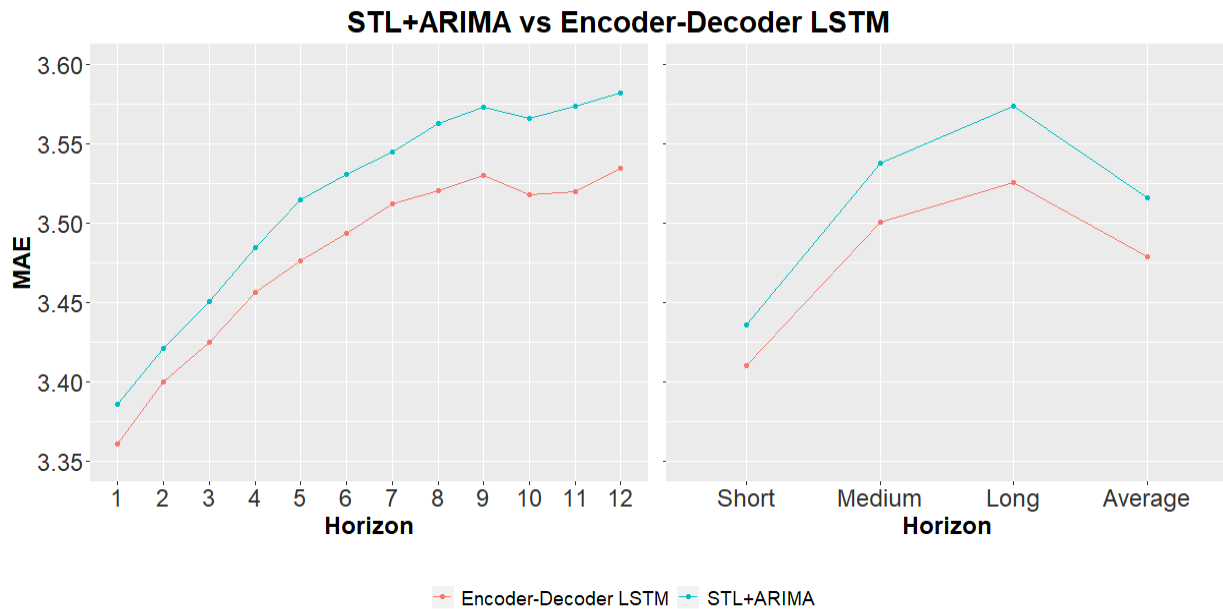
Διάγραμμα 161 Μέσο MAE ανά μέθοδο.



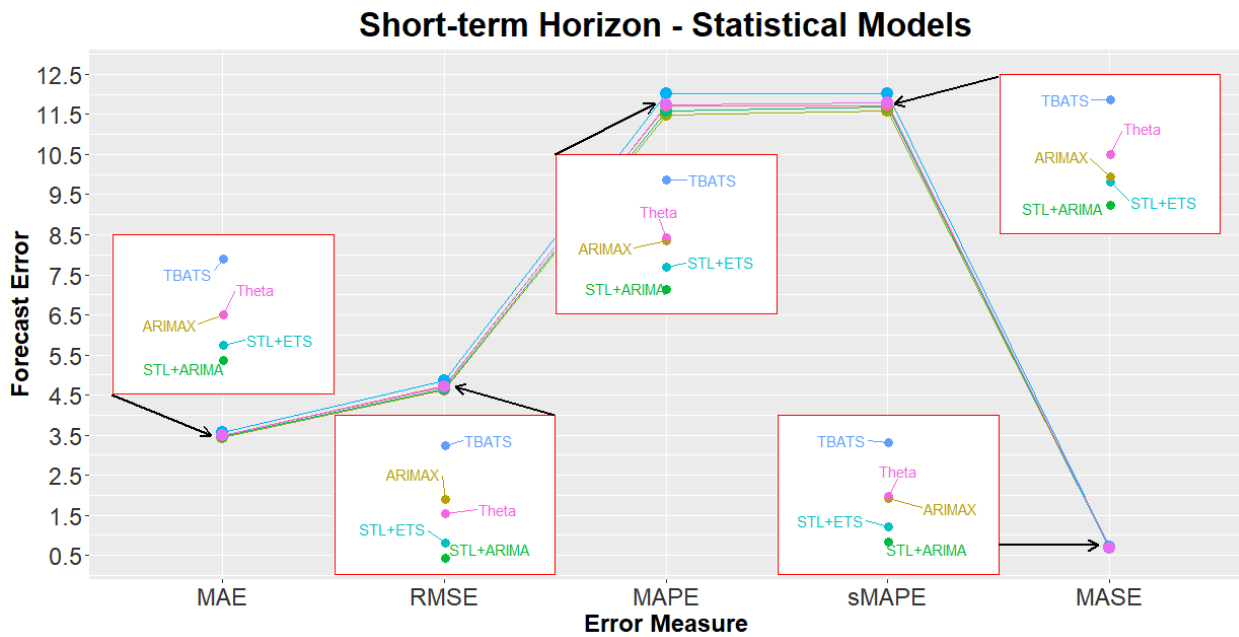
Διάγραμμα 162 Σύγκριση της μεθόδου STL+ARIMA με τις μεθόδους Βαθιάς Μηχανικής Μάθησης για τα συνοπτικά μέτρα της μετρικής σφάλματος MAE.

Στο Διάγραμμα 161 φαίνεται το μέσο MAE ανά μέθοδο. Η μέθοδος Encoder-Decoder LSTM έχει το μικρότερο μέσο MAE από όλες τις μεθόδους με τιμή 3.48. Πιο καθαρή είναι η εικόνα στα Διαγράμματα 162 και 163. Στο 1<sup>ο</sup> παρατηρούμε ότι η καλύτερη από τις στατιστικές μεθόδους (ομάδα 2), η STL+ARIMA, για βραχυπρόθεσμο μέσο ορίζοντα είναι καλύτερη από την μέθοδο MLP και την μέθοδο Encoder-Decoder Conv-LSTM αλλά όχι από τις υπόλοιπες μεθόδους της ομάδας 3. Για μεσοπρόθεσμο μέσο ορίζοντα είναι καλύτερη μόνο από την μέθοδο MLP, γεγονός που αλλάζει για τον μακροπρόθεσμο ορίζοντα όπου είναι χειρότερη από όλες τις μεθόδους της ομάδας 3. Τέλος, για το μέσο MAE η μέθοδος STL+ARIMA είναι καλύτερη μόνο από την μέθοδο MLP. Το συμπέρασμα είναι ότι οι μέθοδοι Βαθιάς Μηχανικής Μάθησης φαίνεται να είναι ανώτερες έστω και με μικρή διαφορά. Σημαντική παρατήρηση είναι ότι όσο αυξάνεται ο ορίζοντας πρόβλεψης αυξάνεται και η κλίση του διαγράμματος για την μέθοδο STL+ARIMA σε σχέση με τις υπόλοιπες μεθόδους εξαιρουμένης της μεθόδου MLP. Αυτό σημαίνει ότι η μέθοδος STL+ARIMA χειροτερεύει ως προς το MAE για μεγαλύτερους ορίζοντες πρόβλεψης με μεγαλύτερο ρυθμό, το οποίο όμως χρήζει περαιτέρω έρευνας. Στο διάγραμμα 163 φαίνεται η αύξηση της «απόστασης» της μεθόδου STL+ARIMA από την μέθοδο Encoder-Decoder LSTM όσο αυξάνεται η τιμή του ορίζοντα πρόβλεψης. Για του λόγου το αληθές σημειώνεται μείωση του βραχυπρόθεσμου, μεσοπρόθεσμου και μακροπρόθεσμου MAE κατά 0.8%, 1.1% και 1.7%, αντίστοιχα, σε σχέση με την μέθοδο STL+ARIMA. Για το μέσο MAE υπάρχει βελτίωση κατά 1.1%. Μπορεί να σχολιαστεί σε αυτό το σημείο ότι οι μέθοδοι βαθιάς μηχανικής μάθησης είναι πολλά υποσχόμενες όσον αφορά σε προβλέψεις για μεγαλύτερες τιμές του ορίζοντα πρόβλεψης.

Σκοπός των επόμενων Διαγραμμάτων (164, 165, 166, 167 και 168, 169, 170, 171) είναι η διερεύνηση της επιρροής που έχει η επιλογή της μετρικής σφάλματος στην επιλογή της καλύτερης μεθόδου όσον αφορά την προβλεπτική της ικανότητα. Η διαδικασία χωρίστηκε για λόγους ευκρίνειας ανά ομάδα μεθόδων (ομάδα 2 και 3). Στο Διάγραμμα 164 φαίνεται το σφάλμα για τις στατιστικές μεθόδους και για τις 5 διαφορετικές μετρικές σφάλματος που χρησιμοποιήθηκαν για βραχυπρόθεσμο ορίζοντα πρόβλεψης. Παρατηρούμε ότι για όλες τις μετρικές σφάλματος, καλύτερη είναι η μέθοδος STL+ARIMA. Οι μόνες εναλλαγές που υπάρχουν είναι για τις μεθόδους ARIMAX και Theta με την 1<sup>η</sup> να είναι καλύτερη για όλες τις μετρικές εκτός την μετρική RMSE.

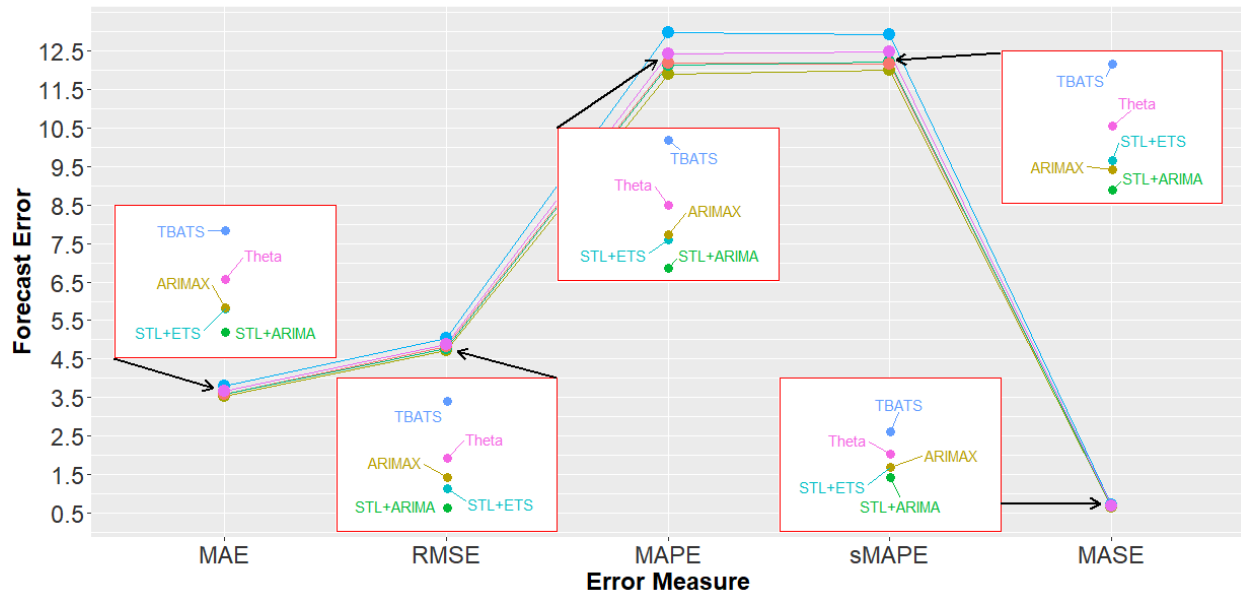


**Διάγραμμα 163** Σύγκριση της μεθόδου STL+ARIMA με την μέθοδο Encoder-Decoder LSTM ανά ορίζοντα και για τα συνοπτικά μέτρα της μετρικής σφάλματος MAE.



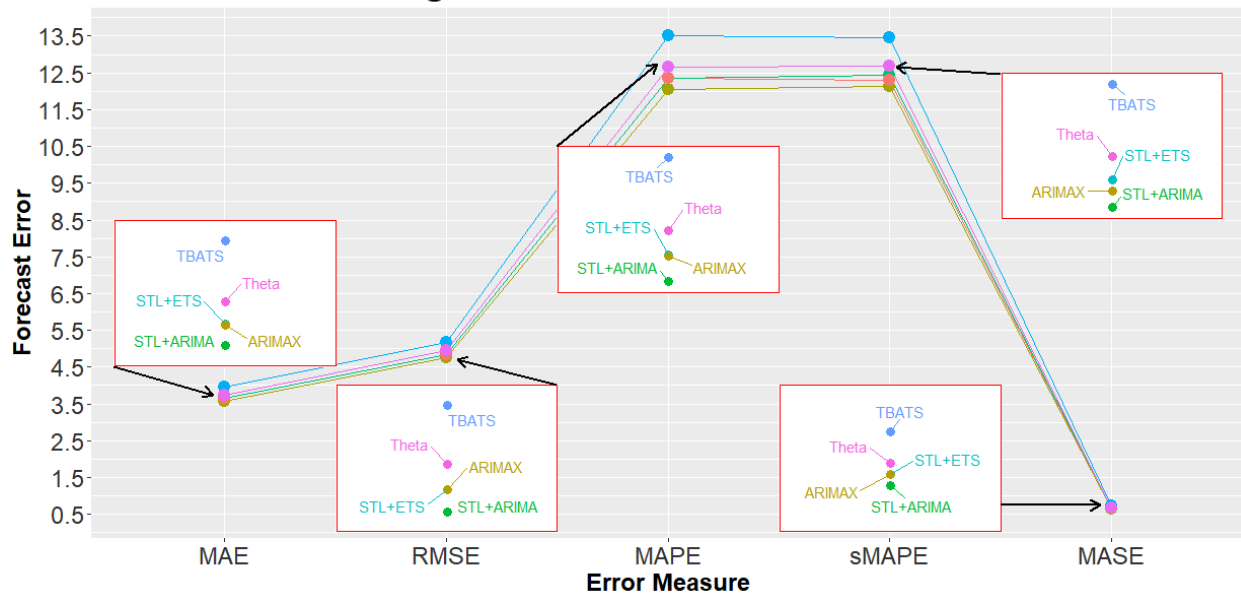
**Διάγραμμα 164** Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για βραχυπρόθεσμο ορίζοντα πρόβλεψης.

### Medium-term Horizon - Statistical Models



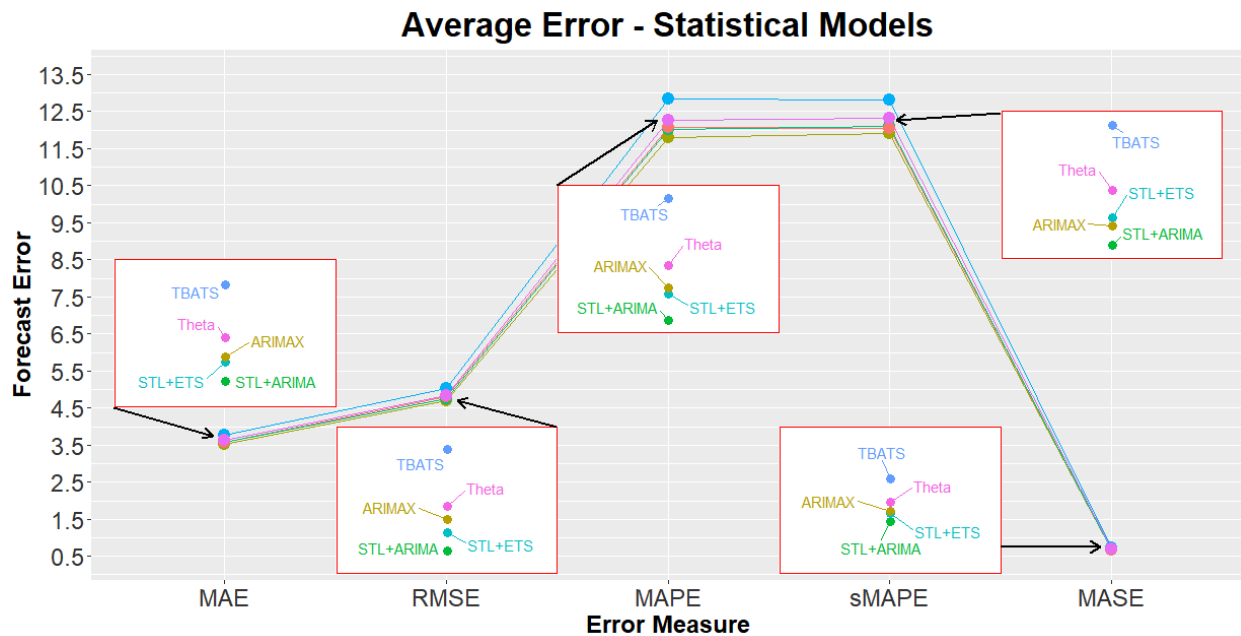
Διάγραμμα 165 Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης.

### Long-term Horizon - Statistical Models



Διάγραμμα 166 Το σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης.

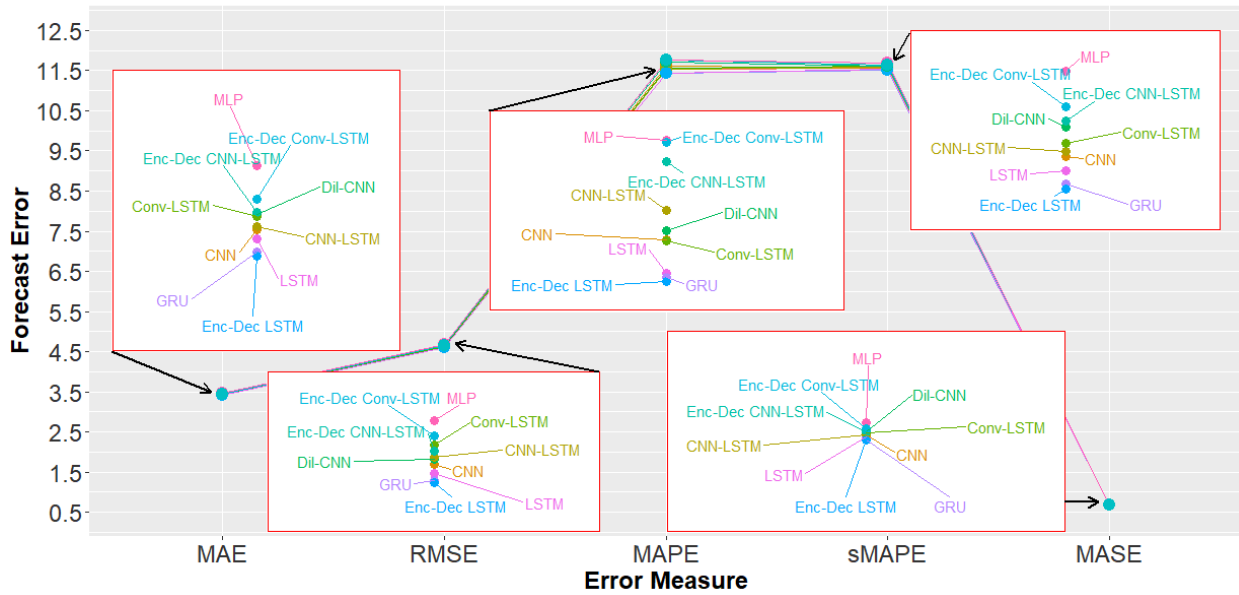
Παρόμοια ισχύουν, βλέποντας το Διάγραμμα 165 όπου απεικονίζεται το σφάλμα για τις μεθόδους της ομάδας 2 ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης. Η μονη εναλλαγή είναι για τις μεθόδους ARIMAX και STL+ETS, όπου η 1<sup>η</sup> είναι ανώτερη της 2<sup>ης</sup> μόνο για την μετρική sMAPE. Στο διάγραμμα 166 φαίνεται το σφάλμα για τις μεθόδους της ομάδας 2 ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης. Και πάλι η μόνη εναλλαγή που παρατηρείται είναι για τις μεθόδους ARIMAX και STL+ETS, με την 2<sup>η</sup> να είναι καλύτερη μόνο για την μετρική RMSE. Τέλος, στο Διάγραμμα 167 για το μέσο σφάλμα για όλους τους ορίζοντες, ισχύουν τα αντίστοιχα που σχολιάστηκαν για το Διάγραμμα 166.



**Διάγραμμα 167** Το μέσο σφάλμα για τις στατιστικές μεθόδους ανά μετρική σφάλματος.

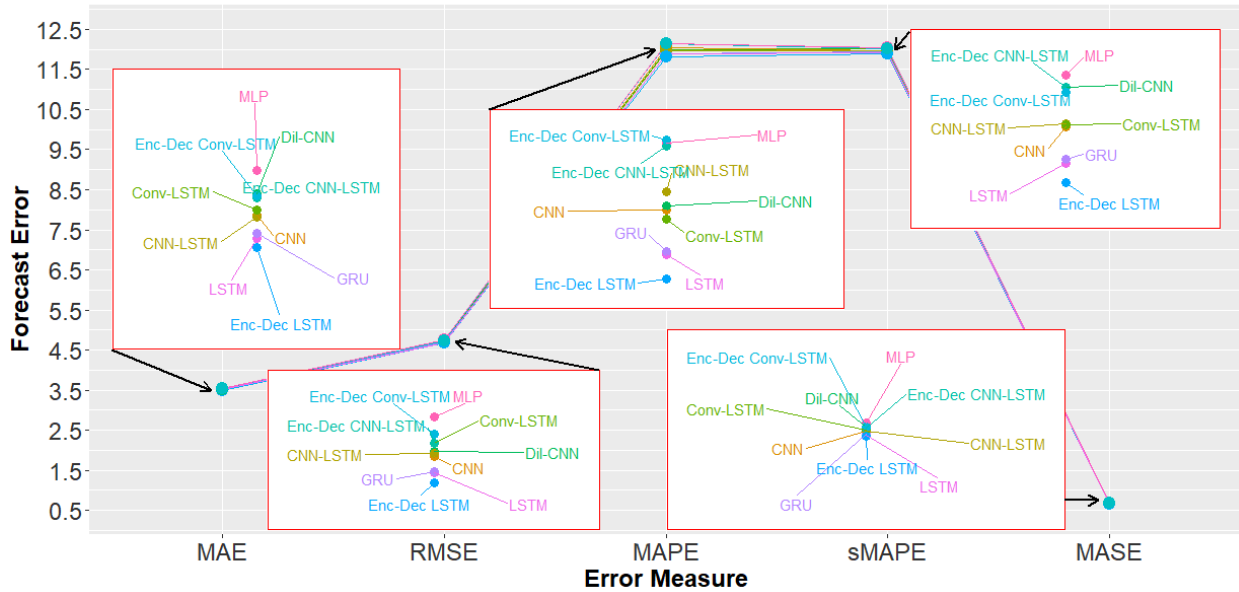
Στο Διάγραμμα 168 φαίνεται το σφάλμα για τις μεθόδους της ομάδας 3, και για τις 5 διαφορετικές μετρικές σφάλματος, για βραχυπρόθεσμο ορίζοντα πρόβλεψης. Για όλες τις μετρικές καλύτερη από τις μεθόδους της ομάδας 3 είναι η μέθοδος Encoder-Decoder LSTM, ενώ χειρότερη η μέθοδος MLP. Τα αποτελέσματα των μετρικών MAE, MASE, sMAPE σε όλες τις μεθόδους συμφωνούν, ενώ για τις μετρικές MAPE και RMSE εμφανίζουν κάποιες έντονες ενάλλαγες χωρίς να υπάρχει βέβαια αλλαγή για τα 3 καλύτερα μοντέλα και για το 4<sup>ο</sup> καλύτερο να αλλάζει θέση μόνο για την μετρική MAPE (η οποία γενικά δεν προτείνεται πλέον στην βιβλιογραφία).

### Short-term Horizon - Deep Learning Models



Διάγραμμα 168 Το σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος για βραχυπρόθεσμο ορίζοντα πρόβλεψης.

### Medium-term Horizon - Deep Learning Models

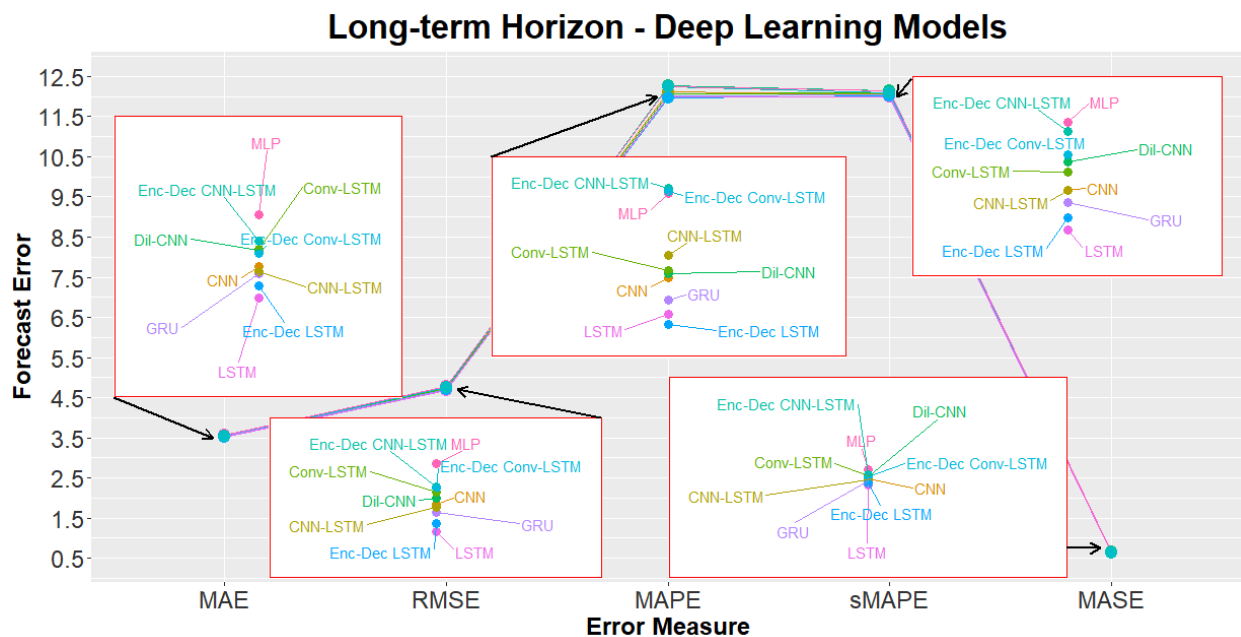


Διάγραμμα 169 Το σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης.

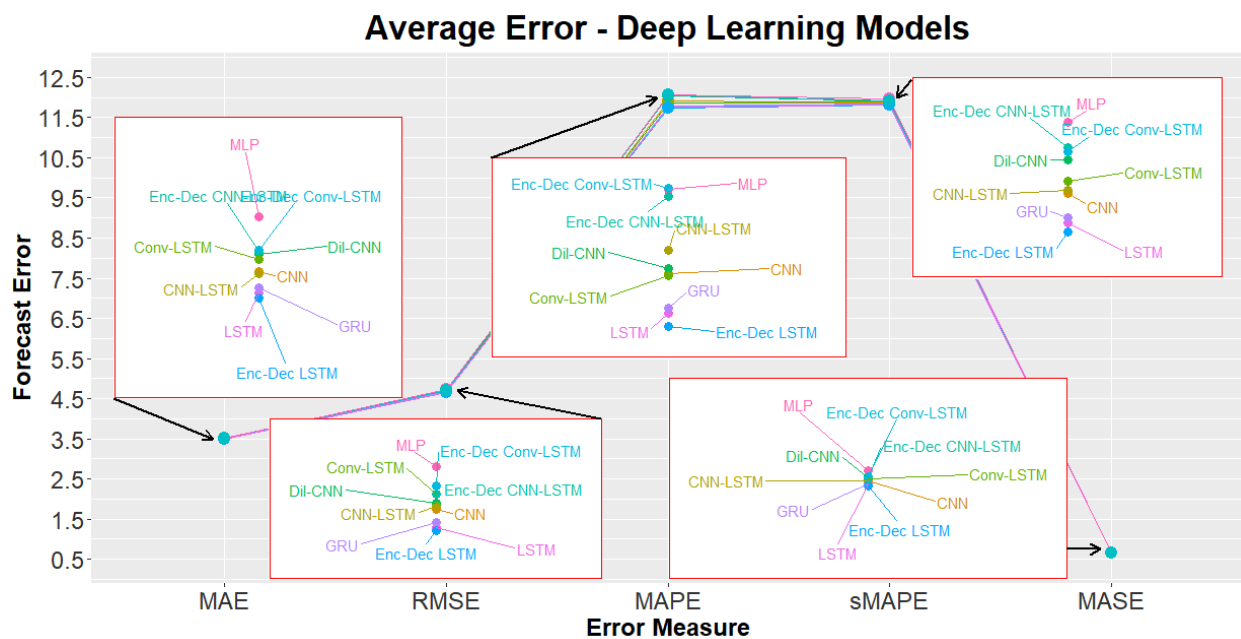


Στο Διάγραμμα 169 φαίνεται το σφάλμα για τις μεθόδους της ομάδας 3 ανά μετρική σφάλματος για μεσοπρόθεσμο ορίζοντα πρόβλεψης, όπου και σε αυτήν την περίπτωση καλύτερη μέθοδος είναι το Encoder-Decoder LSTM. Χειρότερη είναι μέθοδος MLP για όλες τις μετρικές εκτός της μετρικής MAPE όπου το μεγαλύτερο σφάλμα έχει η μέθοδος Encoder-Decoder Conv-LSTM. Και πάλι οι τρεις πρώτες μέθοδοι είναι οι ίδιες για όλες τις μετρικές. Ωστόσο, παρατηρούνται κάποιες περιορισμένες εναλλαγές κυρίως για την μέθοδο Dilated CNN και Conv-LSTM, και για την μετρική MAPE.

Στο Διάγραμμα 170 απεικονίζεται το σφάλμα για τις μεθόδους της ομάδας 3 ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης. Εδώ σημειώνεται αλλαγή στην θέση του καλύτερου μοντέλου (όπως σχολιάστηκε προηγουμένως και για το μέσο MAE), δηλαδή, για όλες τις μετρικές εκτός του MAPE καλύτερη μέθοδος είναι η LSTM. Και πάλι όμως οι 3 καλύτερες μέθοδοι παραμένουν οι ίδιες για όλες τις μετρικές. Σημειώνονται βέβαια κάποιες εναλλαγές κυρίως για τις ίδιες μεθόδους παρόμοιες με εκείνες που σχολιάστηκαν για το Διάγραμμα 169.



**Διάγραμμα 170** Το σφάλμα για τις Βαθιάς Μηχανικής Μάθησης μεθόδους ανά μετρική σφάλματος για μακροπρόθεσμο ορίζοντα πρόβλεψης.

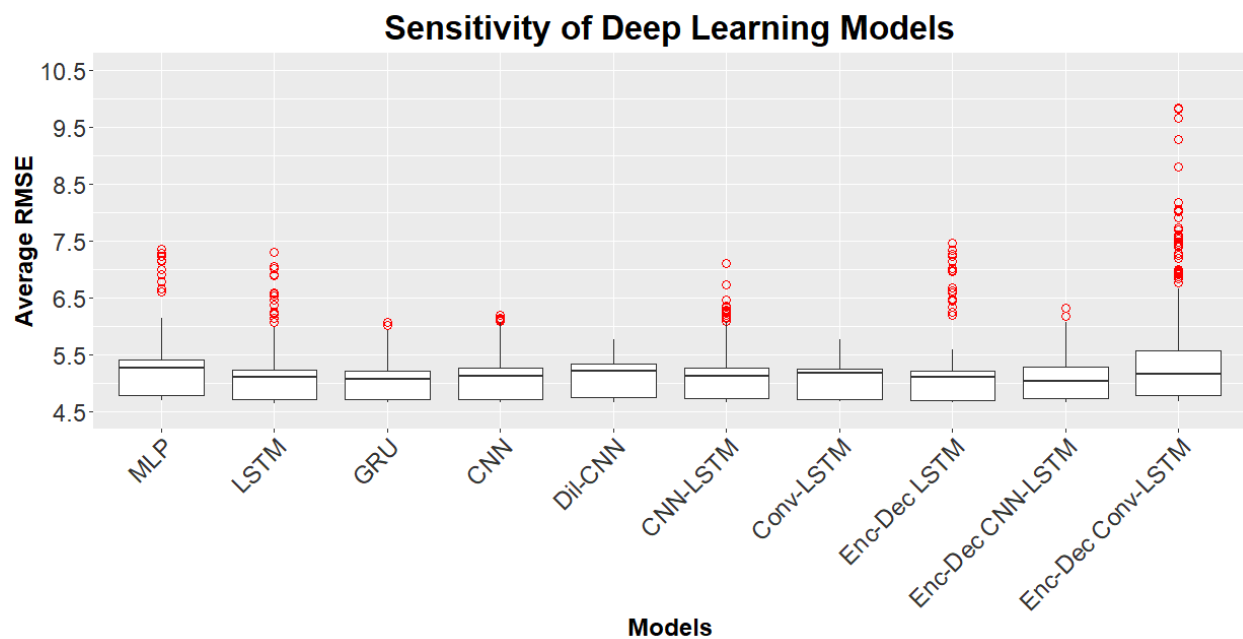


**Διάγραμμα 171** Το μέσο σφάλμα για τις μεθόδους Βαθιάς Μηχανικής Μάθησης ανά μετρική σφάλματος.

Στο Διάγραμμα 171 φαίνεται το μέσο σφάλμα για τις μεθόδους της ομάδας 3 ανά μετρική σφάλματος. Τα ίδια ισχύουν και για την περίπτωση του μέσου σφάλματος με καλύτερη την μέθοδο Encoder-Decoder LSTM και να ακολουθούν οι μέθοδοι LSTM και GRU.

Το τελικό συμπέρασμα είναι ότι δεν αλλάζει πολύ το τοπίο ως προς την αξιολόγηση της προβλεπτικής ικανότητας των μεθόδων χρησιμοποιώντας διαφορετικές μετρικές σφάλματος. Αυτό βέβαια ισχύει δεδομένου του προβλήματος της διπλωματικής. Για περισσότερες χρονοσειρές με διαφορετικά δεδομένα (πχ. κυκλοφοριακός φόρτος) μπορεί να αλλάξει το παραπάνω συμπέρασμα, και αφήνεται για περαιτέρω έρευνα.

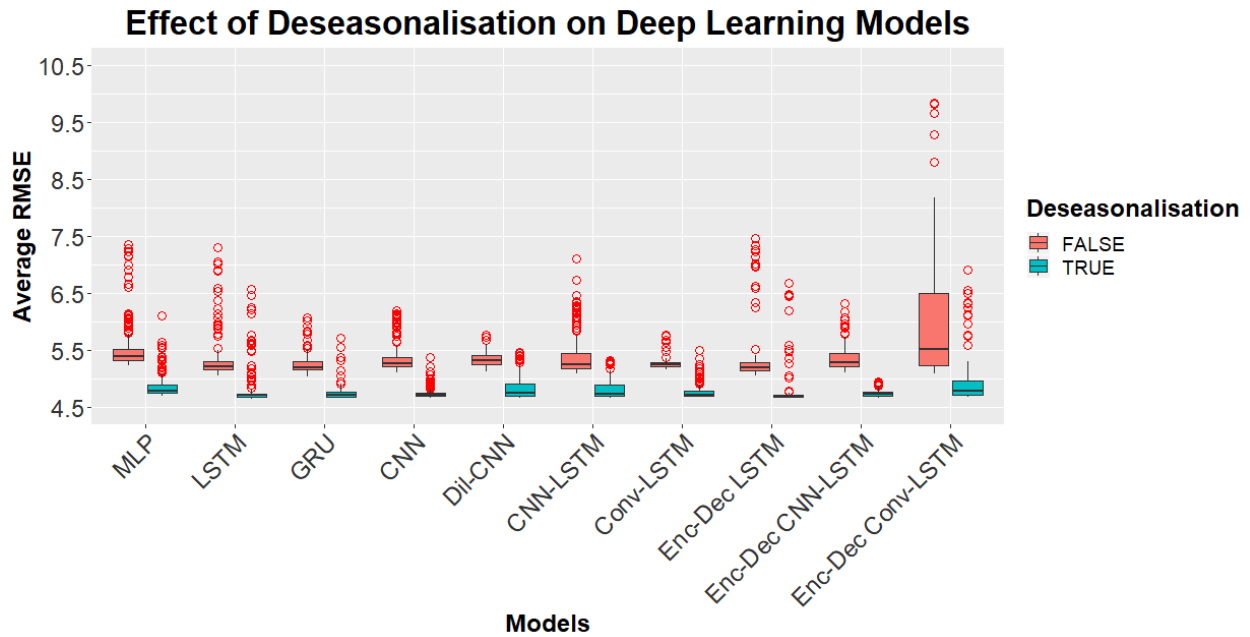
Στο Διάγραμμα 172 φαίνεται η ευαισθησία των μεθόδων της ομάδας 3. Οι μέθοδοι με την μικρότερη ευαισθησία είναι οι Conv-LSTM, Dilated CNN, GRU και CNN. Οι μέθοδοι με την μεγαλύτερη ευαισθησία είναι οι Encoder-Decoder Conv-LSTM, Encoder-Decoder LSTM και LSTM. Φαίνεται λοιπόν ότι οι δύο καλύτερες μέθοδοι όσον αφορά την προβλεπτική ικανότητα κατέχουν από τις μεγαλύτερες ευαισθησίες.



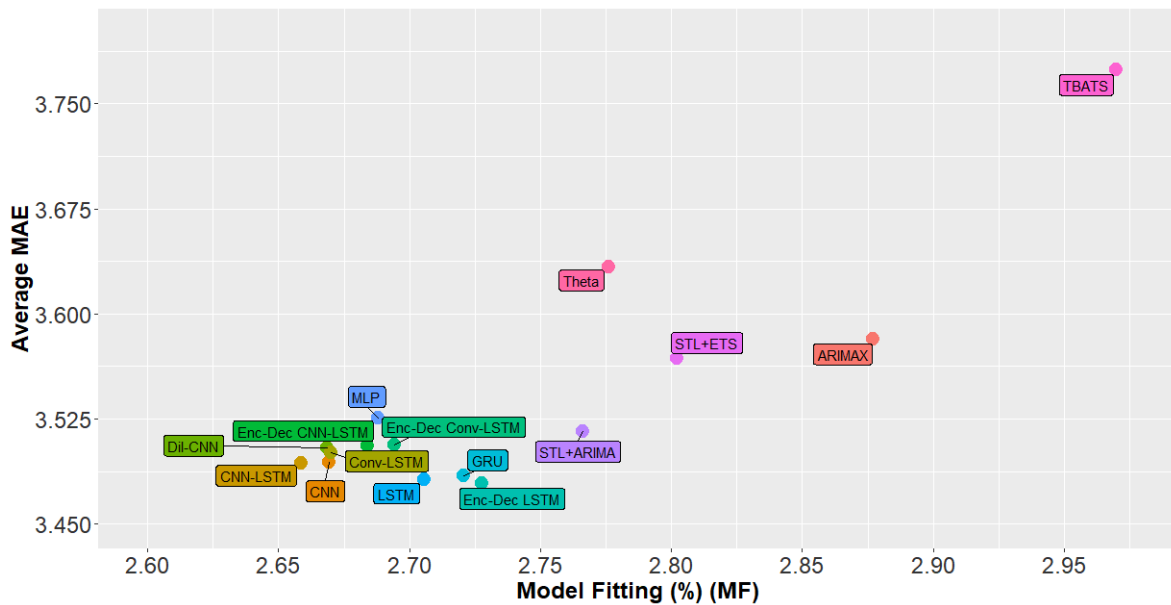
**Διάγραμμα 172** Ευαισθησία των μεθόδων Βαθιάς Μηχανικής Μάθησης στο χώρο παραμέτρων που εξετάστηκε κάθε μία μέθοδος.

Από τα σημαντικότερα αποτελέσματα της διπλωματικής φαίνεται στο Διάγραμμα 173. Είναι ξεκάθαρο ότι η αποεποχικοποίηση των δεδομένων έχει σημαντική μειωτική επιρροή στο σφάλμα πρόβλεψης για όλες τις μεθόδους βαθιάς μηχανικής μάθησης. Παρότι το θεώρημα καθολικής προσέγγισης δηλώνει ότι ένα νευρωνικό δίκτυο μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση για αυθαίρετο αριθμό νευρώνων, προκύπτει από τα αποτελέσματα της διπλωματικής ότι δεν ισχύει τουλάχιστον για μονομεταβλητές κυκλοφοριακές χρονοσειρές. Είναι ένα αποτέλεσμα που τα τελευταία δύο χρόνια έχει αρχίσει να αντιλαμβάνεται η επιστημονική κοινότητα [113]. Η εποχικότητα για κυκλοφοριακές χρονοσειρές είναι μία συνιστώσα η οποία είναι σταθερή στον χρόνο συνήθως και ίσως αποτελεί την πιο αιτιοκρατική συνιστώσα της χρονοσειράς. Η αφαίρεση της τάσης προέκυψε στατιστικά σημαντική με μειωτική επίδραση στην πλειοψηφία των μεθόδων (εκτός των μεθόδων Dilated CNN και Conv-LSTM). Παρόλα αυτά, στα περισσότερα βέλτιστα μοντέλα για κάθε μέθοδο δεν έχει επιλεγεί (FALSE). Αυτό βέβαια μπορεί να οφείλεται στο γεγονός ότι η παρουσία της τάσης δεν ήταν τόσο έντονη στο σύνολο δεδομένων. Μαθαίνοντας ξεχωριστά την εποχικότητα με την μέθοδο LOESS και αφήνοντας το νευρωνικό δίκτυο να επικεντρωθεί μόνο στην βραχυπρόθεσμη συσχέτιση των δεδομένων και στην ενδεχόμενη τάση η οποία δεν ήταν καθόλου έντονη, τα αποτελέσματα βελτιώθηκαν εμφανώς για όλες τις μεθόδους

και μάλιστα εάν δεν είχε χρησιμοποιηθεί ο πιο πάνω συνδυασμός μεθόδων προεπεξεργασίας και μηχανικής μάθησης, όλα τα αποτελέσματα για τις μεθόδους βαθιάς μηχανικής μάθησης θα ήταν χειρότερα με εκείνα των στατιστικών μεθόδων.

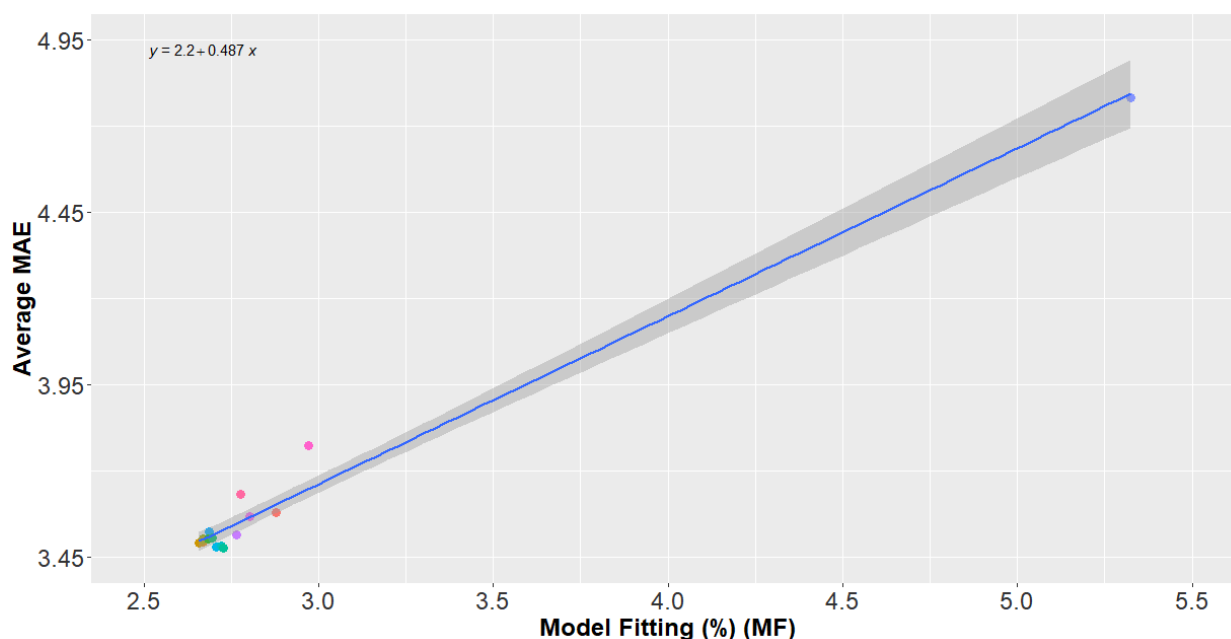


**Διάγραμμα 173** Επιρροή της αποεποχικοποίησης των δεδομένων στο μέσο RMSE.

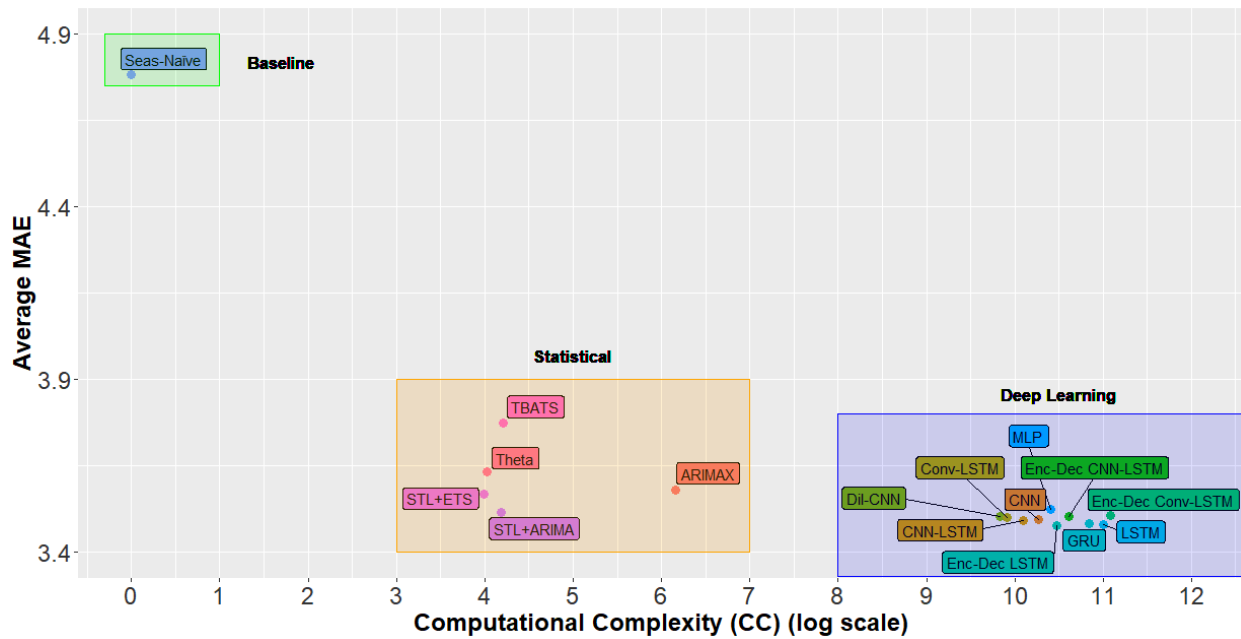


**Διάγραμμα 174** Προσαρμογή κάθε μεθόδου σε σχέση με το μέσο MAE.

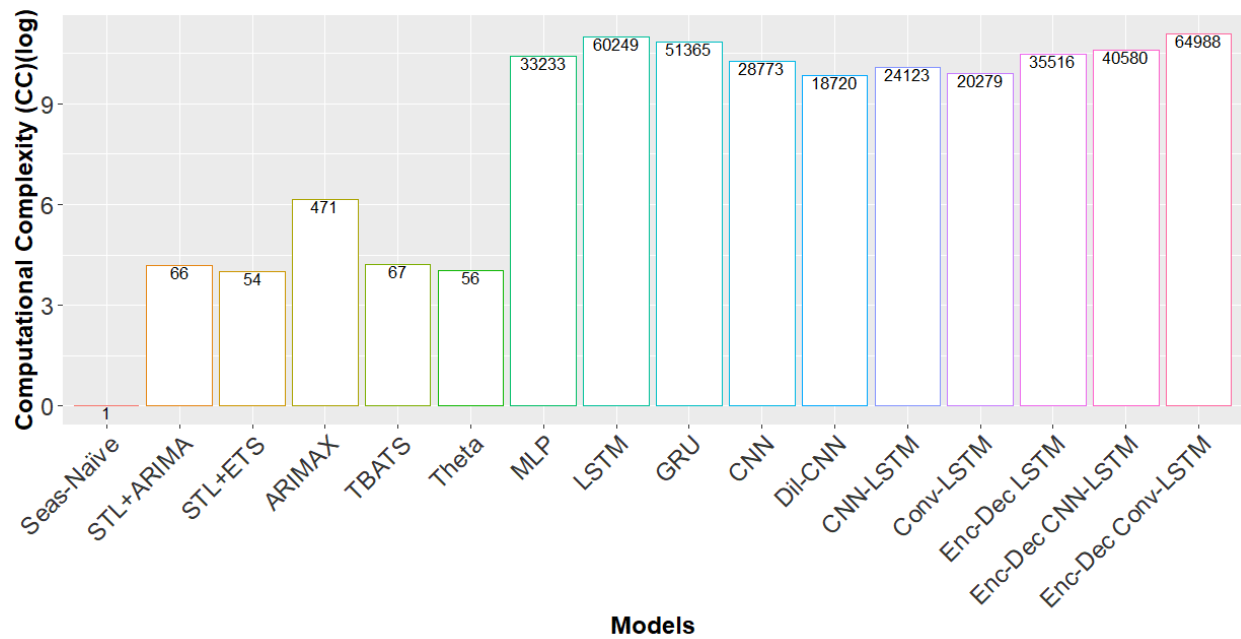
Στο Διάγραμμα 174 φαίνεται η προσαρμογή του μοντέλου σε σχέση με το μέσο MAE. Παρατηρούμε ότι όλες οι μέθοδοι της ομάδας 2 έχουν χειρότερη προσαρμογή σε σχέση με τις μεθόδους της ομάδας 3. Οι 3 καλύτερες μέθοδοι έχουν ποσοστιαία προσαρμογή μεταξύ 2.7 και 2.73, είναι δηλαδή πολύ κοντά. Στο Διάγραμμα 175 φαίνεται η ευθεία της γραμμικής παλινδρόμησης της προσαρμογής του μοντέλου ως προς το μέσο MAE. Όσο αυξάνεται το ποσοστό της προσαρμογής (η προσαρμογή εκφράζεται με όρους σφάλματος), δηλαδή μειώνεται η προσαρμογή, τόσο αυξάνεται το MAE. Μπορεί να ειπωθεί άτυπα ότι για όλες τις μεθόδους βαθιάς μηχανικής μάθησης δεν έγινε υπερεκπαίδευση, το οποίο βέβαια επιβεβαιώνεται και κατά την διαδικασία της εκπαίδευσης μέσω του πρόωρου σταματήματος.



**Διάγραμμα 175** Γραμμική παλινδρόμηση για την προσαρμογή μοντέλου ως προς το μέσο MAE.

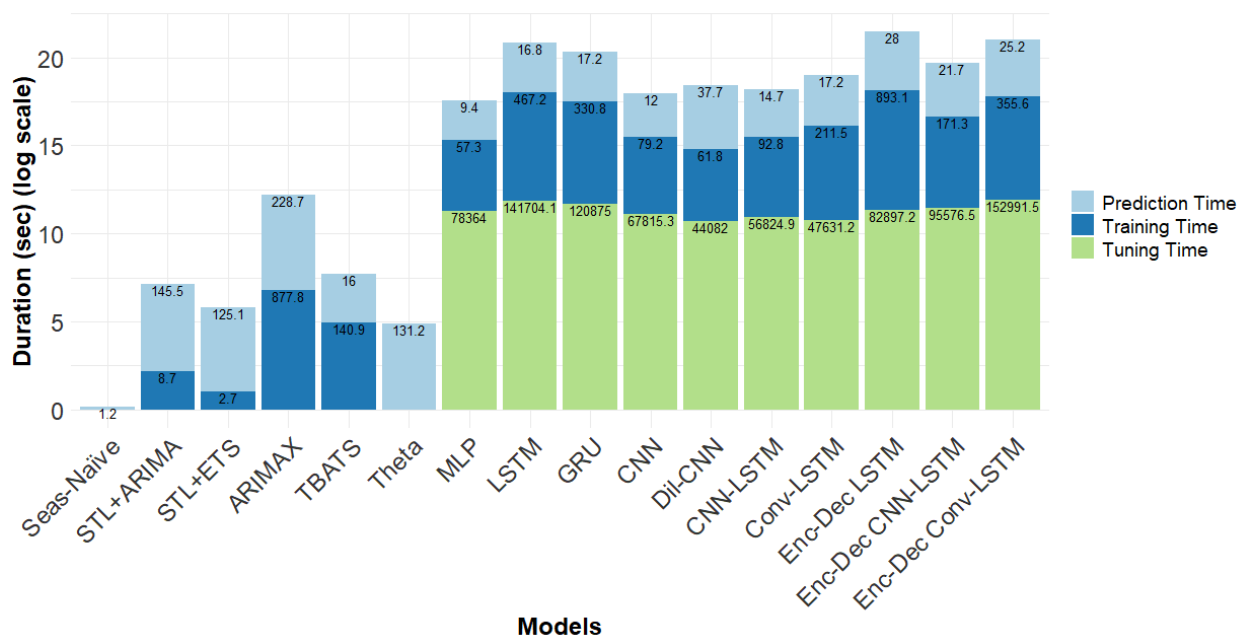


Διάγραμμα 176 Υπολογιστική πολυπλοκότητα όλων των μεθόδων σε συνάρτηση με το μέσο MAE.

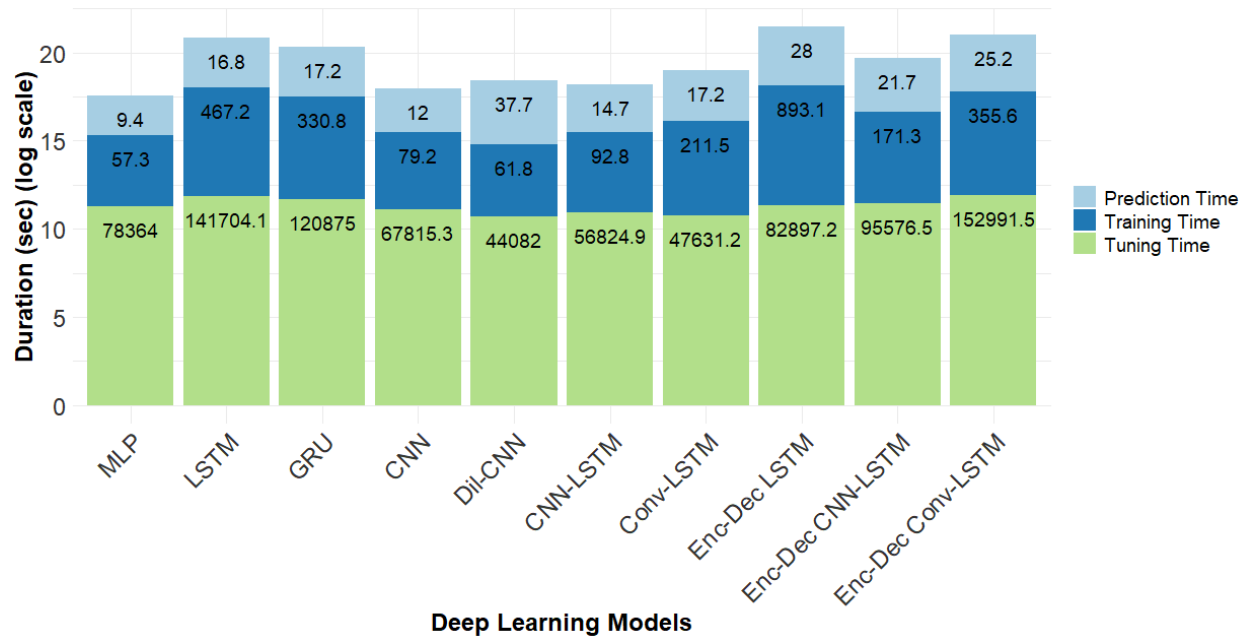


Διάγραμμα 177 Υπολογιστική πολυπλοκότητα ανά μοντέλο σε μορφή barplot.

Στο Διάγραμμα 176 φαίνεται η υπολογιστική πολυπλοκότητα ως προς το μέσο MAE (ο άξονας x είναι σε λογαριθμική κλίμακα). Παρατηρούμε ότι όλες οι μέθοδοι βαθιάς μηχανικής μάθησης έχουν μεγαλύτερη υπολογιστική πολυπλοκότητα, και πιο συγκεκριμένα, 3 τάξεις μεγέθους μεγαλύτερη σε σύγκριση με τις στατιστικές μεθόδους. Άρα το κέρδος στο μέσο MAE ήρθε με μεγάλο υπολογιστικό κόστος. Στο Διάγραμμα 177 φαίνεται πιο καθαρά η υπολογιστική πολυπλοκότητα ανά μέθοδο (ο άξονας y είναι σε λογαριθμική κλίμακα). Από τις στατιστικές μεθόδους η πιο «φθηνή» είναι η μέθοδος STL+ETS και ακολουθούν η μέθοδος Theta, STL+ARIMA, TBATS και ARIMAX. Από τις μεθόδους βαθιάς μηχανικής μάθησης η πιο «φθηνή» μέθοδος είναι η Dilated CNN και ακολουθεί η μέθοδος Conv-LSTM, CNN-LSTM, CNN, MLP, Encoder-Decoder LSTM, Encoder-Decoder CNN-LSTM, GRU, LSTM και τέλος η μέθοδος Encoder-Decoder Conv-LSTM (με αύξουσα σειρά υπολογιστικής πολυπλοκότητας). Στο Διαγράμματα 178 και 179 φαίνονται οι χρόνοι βελτιστοποίησης υπερπαραμέτρων, εκπαίδευσης και πρόβλεψης για όλες τις μεθόδους και για τις μεθόδους βαθιάς μηχανικής μάθησης, αντίστοιχα, για λόγους καλύτερης εποπτείας. Η συντριπτική πλειοψηφία του χρόνου των μεθόδων της ομάδας 3 αφιερώνεται στη βελτιστοποίηση των υπερπαραμέτρων.



**Διάγραμμα 178** Χρόνοι βελτιστοποίησης υπερπαραμέτρων, εκπαίδευσης και πρόβλεψης για όλες τις μεθόδους.

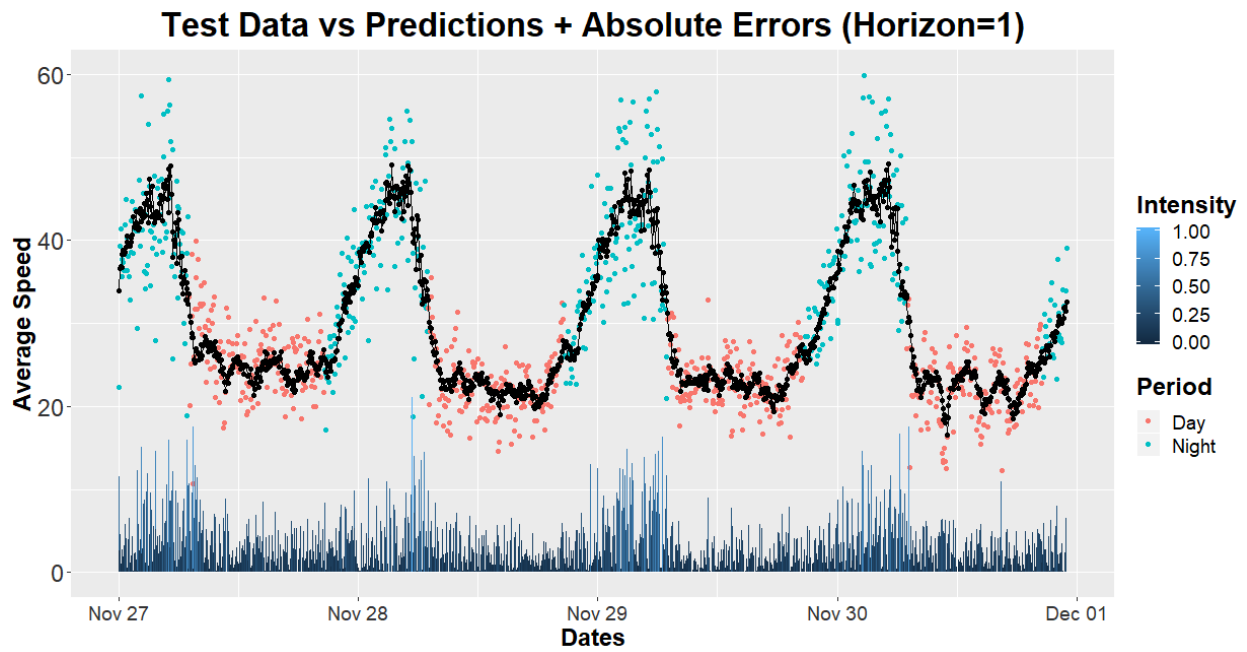


**Διάγραμμα 179** Χρόνοι βελτιστοποίησης υπερπαραμέτρων, εκπαίδευσης και πρόβλεψης για τις μεθόδους Βαθιάς Μηχανικής Μάθησης.

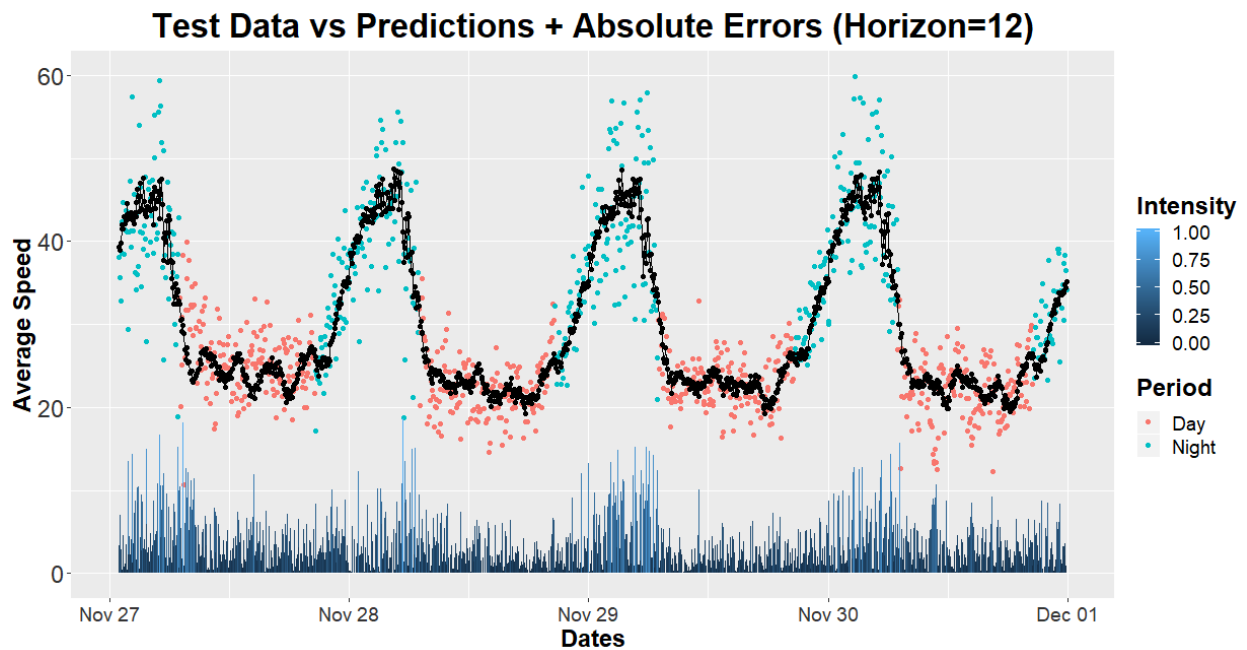
Οι πιο «ακριβές» μέθοδοι όσον αφορά τον χρόνο βελτιστοποίησης των υπερπαραμέτρων, με φθίνουσα σειρά χρόνου, είναι οι Encoder-Decoder Conv-LSTM, LSTM, GRU, Encoder-Decoder CNN-LSTM και Encoder-Decoder LSTM. Ενώ για τον χρόνο εκπαίδευσης η πιο «ακριβή» μέθοδος είναι η Encoder-Decoder LSTM. Όσον αφορά τον χρόνο πρόβλεψης οι μέθοδοι Βαθιάς Μηχανικής Μάθησης απαιτούν λιγότερο χρόνο για την μέθοδο αξιολόγησης με κυλιόμενη αρχή. Να σχολιαστεί σε αυτό το σημείο ότι για τις στατιστικές μεθόδους δεν διατίθεται χρόνος βελτιστοποίησης υπερπαραμέτρων καθώς λόγω της αυτοματοποιημένης διαδικασίας που χρησιμοποιήθηκε, δεν ήταν δυνατή η διάκριση με τον χρόνο εκπαίδευσης. Ακόμα και αθροιστικά όμως οι χρόνοι παραμένουν πολύ μικρότεροι σε τάξη μεγέθους σε σύγκριση με τους αντίστοιχους των μεθόδων βαθιάς μηχανικής μάθησης.

Είναι έκδηλο ότι παρόλο που όλες οι μέθοδοι της ομάδας 3 (εκτός της MLP) είναι καλύτερες σε προβλεπτική ικανότητα (μέσο MAE) από τις στατιστικές μεθόδους (ομάδα 2) το «κόστος» είναι πολύ μεγαλύτερο (κατά 3 τάξεις μεγέθους) που μας κάνει να είμαστε σκεπτικοί για τις μεθόδους βαθιάς μηχανικής μάθησης. Στα Διαγράμματα 180 και 181 δίνονται οι προβλέψεις του καλύτερου μοντέλου όσον αφορά την μέση προβλεπτική ικανότητα.





**Διάγραμμα 180** Προβλέψεις της μεθόδου Encoder-Decoder LSTM για ορίζοντα πρόβλεψης ίσο με 1.



**Διάγραμμα 181** Προβλέψεις της μεθόδου Encoder-Decoder LSTM για ορίζοντα πρόβλεψης ίσο με 12.

Στα Διαγράμματα 180 και 181 απεικονίζονται οι προβλέψεις της μεθόδου Encoder-Decoder LSTM για ορίζοντα πρόβλεψης (σε βήματα) ίσο με 1 και 12, αντίστοιχα, μαζί με τα πραγματικά δεδομένα. Στον άξονα x φαίνονται υπό μορφή barplot τα απόλυτα σφάλματα ανά πρόβλεψη. Παρατηρούμε ότι τα μεγαλύτερα σφάλματα εμφανίζονται και για τις 4 μέρες πρόβλεψης τις νυκτερινές ώρες. Αυτό συμβαίνει διότι φαίνεται να αλλάζει η κατανομή των δεδομένων τις βραδινές ώρες έχοντας μεγαλύτερη μεταβλητότητα.

## 7 ΣΥΜΠΕΡΑΣΜΑΤΑ

---

### 7.1 ΓΕΝΙΚΑ

Σκοπός της παρούσας διπλωματικής ήταν η διερεύνηση και σύγκριση διαφόρων μεθόδων βραχυπρόθεσμης πρόβλεψης μονομεταβλητών κυκλοφοριακών χρονοσειρών, μεθόδων δανεισμένων από τους κλάδους της στατιστικής και της βαθιάς μηχανικής μάθησης. Η πολυπλοκότητα των μεθόδων και η εξείκωση που απαιτείται κάνει πολύ απαιτητική την διαδικασία της πρόβλεψης και γι' αυτό η επιστημονική κοινότητα έχει στραφεί σε αυτοματοποιημένες διαδικασίες, γεγονός που είναι πιο έντονο για τις στατιστικές μεθόδους χρονοσειρών. Για το λόγο αυτό, έγινε χρήση αυτοματοποιημένων στατιστικών μεθόδων πρόβλεψης και εκλεπτυσμένων μεθόδων βαθιάς μηχανικής μάθησης που δεν είχαν βρεθεί σε άλλες συγκριτικές μελέτες σε κυκλοφοριακή πρόβλεψη, με στόχο την διερεύνηση της προβλεπτικής τους ικανότητας, των δυσκολιών που μπορούν να προκύψουν κατά την εκπαίδευση αυτών των μεθόδων και πως θα μπορούσαν πιθανώς να συνδυαστούν με κάποιες ιδέες δανεισμένες από τον κλάδο των στατιστικών χρονοσειρών. Πρόκειται για μία ιδέα η οποία έχει γίνει δημοφιλής τα τελευταία 2 χρόνια στο ευρύτερο κλάδο της πρόβλεψης χρονοσειρών και δεν έχει σχολιαστεί επαρκώς από τους επιστήμονες που ασχολούνται με την πρόβλεψη κυκλοφοριακών χρονοσειρών.

Παρόλο που η επιστημονική κοινότητα έχει στραφεί σε κυκλοφοριακές προβλέψεις σε επίπεδο δικτύου η εμπειρία έχει δείξει ότι λόγω της αυξημένης πολυπλοκότητας πολλές φορές τα αποτελέσματα σε τέτοιου τύπου προβλέψεων μπορεί να είναι χειρότερα σε σχέση με την μεμονωμένη ανά σύνδεσμο (οδικό τμήμα) πρόβλεψη (μονομεταβλητές κυκλοφοριακές χρονοσειρές). Τα πρώτα μπορούν να χρησιμοποιηθούν συνδυαστικά με τα δεύτερα μέσω μεθόδων ensemble, δίνοντας πολύ καλύτερα αποτελέσματα.

### 7.2 ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΒΑΣΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

Επιλέχθηκαν 16 διαφορετικές μέθοδοι, 1 μέθοδος αναφοράς (Seasonal Naive), 5 στατιστικές μέθοδοι και 10 μέθοδοι βαθιάς μηχανικής μάθησης. Η επιλογή των μεθόδων καλύπτει πλήρως τις ευρέως χρησιμοποιούμενες μεθόδους στη βιβλιογραφία, συνεπώς η ανάλυση και συγκριτική αξιολόγηση μπορεί να θεωρηθεί πλήρης σε σχέση με τις απαιτήσεις της βιβλιογραφίας. Για την εκπαίδευση των στατιστικών μεθόδων χρησιμοποιήθηκαν αυτοματοποιημένες μέθοδοι μέσω του

πακέτου forecast της R, ενώ για την εκπαίδευση των μεθόδων βαθιάς μηχανικής μάθησης έγινε χρήση του API Keras (backend: Tensorflow) σε γλώσσα προγραμματισμού Python. Για την βελτιστοποίηση των υπερπαραμέτρων των μεθόδων βαθιάς μηχανικής μάθησης έγινε εξαντλητική αναζήτηση πλέγματος. Να σημειωθεί ότι η εκπαίδευση των στατιστικών μεθόδων διεξάχθηκε σε συμβατικό υπολογιστή ενώ η αντίστοιχη των μεθόδων βαθιάς μηχανικής μάθησης στο Google Cloud (GPU, 11gb).

Για την αξιολόγηση των μεθόδων έγινε χρήση της μεθόδου κυλιόμενης αρχής χωρίς επανεκπαίδευση με σταθερό ορίζοντα πρόβλεψης, ενώ ταυτόχρονα χρησιμοποιήθηκαν 5 διαφορετικές μετρικές σφάλματος πρόβλεψης (MAE, RMSE, MAPE, sMAPE και MASE) για να εξεταστεί το κατά πόσο η επιλογή αυτή επηρεάζει την επιλογή της καλύτερης μεθόδου. Στην αξιολόγηση της κάθε μεθόδου προστέθηκαν ακόμα δύο μετρικές, η υπολογιστική πολυπλοκότητα, που «περιγράφει» το επιπλέον υπολογιστικό κόστος (πολλαπλασιαστικά) της μεθόδου ως προς την μέθοδο αναφοράς (Seasonal Naive), καθώς και η προσαρμογή της μεθόδου στα δεδομένα εκπαίδευσης.

Εξετάστηκε, επιπροσθέτως, η επιρροή κάποιων πλαισίων δανεισμένων από την βιβλιογραφία των στατιστικών μεθόδων, όπως αφαίρεση της τάσης και της εποχικότητας από τα δεδομένα που θα μπορούσαν ενδεχομένως να έχουν στις μεθόδους βαθιάς μηχανικής μάθησης, οι οποίες τέθηκαν ως υπερπαραμέτροι κατά την διαδικασία εύρεσης του βέλτιστου μοντέλου ανά μέθοδο (ομάδα 3 μόνο).

Η ανάλυση έδειξε ότι γενικά οι μέθοδοι βαθιάς μηχανικής μάθησης έχουν καλύτερη προβλεπτική ικανότητα σε σχέση με τις στατιστικές μεθόδους (εξαιρουμένης της μεθόδου MLP). Η μέγιστη μείωση που επιτεύχθηκε ήταν για μακροπρόθεσμο ορίζοντα πρόβλεψης για την μετρική MAE κατά 1.7%. Το αν και κατά πόσο είναι σημαντική αυτή η μείωση εξαρτάται από το πρόβλημα και το τι ορίζει ο ερευνητής κάθε φορά. Επίσης φαίνεται, οι μέθοδοι βαθιάς μηχανικής μάθησης να έχουν μικρότερο ρυθμό αύξησης του σφάλματος καθώς αυξάνεται η τιμή του ορίζοντα πρόβλεψης συγκριτικά με τις στατιστικές μεθόδους, το οποίο χρήζει περαιτέρω έρευνας. Η ταξινόμηση των μεθόδων με βάση την προβλεπτική τους ικανότητα δεν αλλάζει δραματικά για καμία από τις μετρικές σφάλματος καθώς όλες συμφωνούν για τα 3 ακόμα και 4 καλύτερα μοντέλα για όλους τους ορίζοντες πρόβλεψης και για τα συνοπτικά μέτρα σφάλματος. Άρα εδώ συμπεραίνεται ότι η

επιλογή της μετρικής σφάλματος δεν επηρεάζει την επιλογή της καλύτερης μεθόδου, εξαιρουμένης ίσως της μετρικής σφάλματος MAPE.

Τα αρνητικά σημεία των μεθόδων βαθιάς μηχανικής μάθησης φαίνεται να είναι τα εξής:

- Πολύ μεγάλο υπολογιστικό κόστος σε σχέση με τις στατιστικές μεθόδους ( $\sim 3^{ηs}$  τάξης μεγέθους μεγαλύτερο).
- Μεγαλύτερη εξειδίκευση δεδομένης της πλήρως αυτοματοποιημένης διαδικασίας που υπάρχουν διαθέσιμες για τις στατιστικές μεθόδους.
- Μεγαλύτερες ανάγκες σε υπολογιστικό υλικό.

Για την υλοποίηση των στατιστικών μεθόδων χρειάστηκαν μόλις 500 γραμμές κώδικα στην R διαδικαστικού χαρακτήρα, σε αντίθεση με τις μεθόδους βαθιάς μηχανικής μάθησης που χρειάστηκαν υπερπενταπλάσιο αριθμό γραμμών και αρκετά πιο απαιτητικό κώδικα. Επιπροσθέτως, αυτό το μεγάλο επιπλέον υπολογιστικό κόστος προήλθε χρησιμοποιώντας cloud πλατφόρμα της Google διότι τα μοντέλα βαθιάς μηχανικής μάθησης δεν μπορούν να εκπαιδευτούν σε συμβατικό υπολογιστή, σε αντίθεση με το χαμηλό υπολογιστικό κόστος των μεθόδων στατιστικής το οποίο ήρθε σε συνδυασμό με την χρήση ενός συμβατικού laptop (CPU: i5, RAM: 8gb).

Σημαντικό εύρημα της διπλωματικής που συμπληρώνει περαιτέρω την βιβλιογραφία είναι ότι οι μέθοδοι βαθιάς μηχανικής μάθησης δεν μπορούν να μάθουν ικανοποιητικά όλες τις συνιστώσες από τις οποίες αποτελείται μία χρονοσειρά, όπως πχ. η εποχικότητα. Από τα αποτελέσματα προκύπτει ότι η προβλεπτική ικανότητα των μεθόδων βαθιάς μηχανικής μάθησης βελτιώνεται εάν εφαρμοστούν απευθείας στα εποχιακά προσαρμοσμένα δεδομένα. Υποτίθεται ότι το ίδιο θα συνέβαινε αν τα δεδομένα είχαν έντονη συνιστώσα τάσης, το οποίο προτείνεται για περαιτέρω έρευνα. Ακόμα και με την πεποίθηση ότι ένα πιο πολύπλοκο μοντέλο βαθιάς μηχανικής μάθησης θα μπορούσε να μάθει την χρονοσειρά χωρίς καμία «υποβοήθηση», αυτό θα οδηγούσε πολύ πιθανά σε ακόμα μεγαλύτερο υπολογιστικό κόστος, καθώς η εκμάθηση των συνιστωσών (εποχικότητα και τάση) μίας χρονοσειράς είναι πολύ εύκολες διαδικασίες χρησιμοποιώντας στατιστικές μεθόδους.

### 7.3 ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΕΡΕΥΝΑ

Η διπλωματική εργασία ήταν μόνο η αρχή για την ανάδειξη κάποιων βασικών ζητημάτων που αφορούν την βραχυπρόθεσμη πρόβλεψη κυκλοφοριακών μονομεταβλητών χρονοσειρών. Αν και η επιστημονική κοινότητα έχει κάνει στροφή προς μεθόδους πρόβλεψης σε επίπεδο δικτύου χωρίς να έχει επιλύσει βασικά ζητήματα προβλέψεων σε επίπεδο αρτηρίας, γεγονός που θα βοηθούσε και στο 1<sup>ο</sup> πρόβλημα. Για παράδειγμα, αν ισχύει ότι η αποεποχικοποίηση των δεδομένων βοηθάει τις μεθόδους βαθιάς μηχανικής μάθησης για μονομεταβλητές χρονοσειρές, για ποιο λόγο να μην βοηθάει και σε επίπεδο δικτύου. Εξηγώντας το παραπάνω, προτείνεται να γίνει αποεποχικοποίηση σε επίπεδο δικτύου προς διερεύνηση του παραπάνω και να γίνει η πρόβλεψη στο εποχιακά προσαρμοσμένο δίκτυο. Το ίδιο θα μπορούσε να γίνει και για την συνιστώσα τάσης η οποία όμως δεν εμφανίζεται τόσο έντονη στις κυκλοφοριακές χρονοσειρές. Επιπροσθέτως, μπορεί να διερευνηθεί ποια είναι η καλύτερη μέθοδος αποεποχικοποίησης και αφαίρεσης τάσης για τις μεθόδους βαθιάς μηχανικής μάθησης.

Σημαντικό είναι επίσης όλη αυτή η προσπάθεια να γίνει σε σύνολα δεδομένων αναφοράς ώστε να μπορούν να προκύψουν μεταδεδομένα από διάφορες συγκριτικές μελέτες τα οποία να είναι αξιόπιστα. Είναι γνωστό ότι στην πρόβλεψη χρονοσειρών πολλές εμπειρικές ερευνητικές εργασίες άλλαξαν τον κλάδο καθώς πολλές φορές τα αποτελέσματα τους ήταν αντίθετα από εκείνο που υποδείκνυε έως τότε η θεωρία. Άρα είναι σημαντικό να γίνει το ίδιο και για τις μεθόδους βαθιάς μηχανικής μάθησης.

Λόγω της αυξημένης εξειδίκευσης και υπολογιστικής πολυπλοκότητας που απαιτείται για την βελτιστοποίηση των μεθόδων βαθιάς μηχανικής, κάνοντας χρήση εξαντλητικών μεθόδων, είναι ύψιστης σημασίας η δημιουργία αυτοματοποιημένων διαδικασιών οι οποίες θα βασίζονται σε τεχνικές μεταμάθησης και εκλεπτυσμένων μεθόδων αναζήτησης υπερπαραμέτρων όπως η μπεϋζιανή βελτιστοποίηση που αποκτάει ολοένα και περισσότερο προσοχή στον κλάδο της βαθιάς μηχανικής μάθησης. Αυτή η διαδικασία γίνεται με την εξαγωγή σημαντικών χαρακτηριστικών από κυκλοφοριακές χρονοσειρές, οι οποίες χρησιμοποιούνται για την ομαδοποίηση τους σε κοινές ομάδες ανάλογα με τον βαθμό ομοιότητας τους και κάνοντας μία μετανάλυση στα αποτελέσματα που έχει κάθε μέθοδος για την συσχέτιση του σφάλματος και των υπερπαραμέτρων με τα προαναφερθέντα χαρακτηριστικά.

Δεν είναι επίσης καθόλου προφανές πως επηρεάζουν τα παρακάτω μεγέθη τις μεθόδους βαθιάς μηχανικής μάθησης:

- Το μέγεθος του παραθύρου στο οποίο επιτρέπεται σε ένα μοντέλο να «κοιτάξει» πίσω σε μία κυκλοφοριακή χρονοσειρά.
- Η χρονική ανάλυση των δεδομένων, δηλαδή δεδομένα ανά 1λεπτο, 5λεπτο, 10λεπτο κλπ., καθώς με τις μεθόδους βαθιάς μηχανικής μάθησης ανοίγονται νέες δυνατότητες για υψηλές αναλύσεις στα δεδομένα.
- Πόσα δεδομένα πίσω στο παρελθόν χρειαζόμαστε για την εκπαίδευση ενός μοντέλου βαθιάς μηχανικής μάθησης.

Άλλο ένα πρόβλημα προκύπτει από το γεγονός ότι οι κυκλοφοριακές χρονοσειρές δεν αποτελούνται από μία κατανομή, αντιθέτως αναλόγως την κατάσταση του κυκλοφοριακού δικτύου η κατανομή αυτή αλλάζει. Θα ήταν λοιπόν η ανάπτυξη μεθόδων (τύπου concept drift) οι οποίες είναι ικανές να εντοπίζουν τις αλλαγές στην κατανομή της κυκλοφορίας (βράδυ ή πρωί, συμβάντα στο δίκτυο κα.) και με βάση αυτό να γίνεται αλλαγή στο μοντέλο που χρησιμοποιείται για πρόβλεψη. Για παράδειγμα, σε καταστάσεις συμφόρησης θα μπορούσε να χρειάζεται η μείωση του μήκους του παραθύρου και η αύξηση της ανάλυσης των δεδομένων. Όλα αυτά τα προβλήματα παραμένουν ανοιχτά, και θεωρούμε ότι είναι σημαντικό να διερευνηθούν και μάλιστα σε πλαίσιο πραγματικού χρόνου χρησιμοποιώντας ροές δεδομένων, και όχι απαραίτητα «στατικά» δεδομένα, έτσι ώστε να υπάρχει περαιτέρω αξία και σε επίπεδο διαχείρισης κυκλοφορίας πραγματικού χρόνου.





- [1] Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075), 462-465. doi: 10.1038/nature04292
- [2] Qin, H., Li, H. & Zhao, X. (2010). “Development status of domestic and foreign smart city”, *Glob. Presence* 9 50–52.
- [3] An, S., Lee, B., & Shin, D. (2011). A Survey of Intelligent Transportation Systems. 2011 Third International Conference On Computational Intelligence, Communication Systems And Networks.
- [4] Qi, Y. & Ishak, S. (2014). A Hidden Markov Model for short term prediction of traffic conditions on freeways. *Transportation Research Part C: Emerging Technologies*, 43, pp.95-111.
- [5] Blatnig, S. (2008). “Microscopic Traffic Simulation with Intelligent Agents”, Ph.D. dissertation, Alpen-Adria-Universitat Klagenfurt at Fakultat fur Technische Wissenschaften.
- [6] Barros, J., Araujo, M. & Rossetti, R. (2015). Short-term real-time traffic prediction methods: A survey. 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS).
- [7] Lieu, H. C. (2000). Traffic estimation and prediction system. *TR News*, 208, pp. 3–6.
- [8] van Hinsbergen, C., van Lint, J., & Sanders, F. (2007). “Short term traffic prediction models”, in *Proc. 14th World Congress Intelligent Transportation Systems*, Beijing.
- [9] Do, L., Taherifar, N., & Vu, H. (2018). Survey of neural network - based models for short - term traffic state prediction. *Wires Data Mining And Knowledge Discovery*, 9(1).
- [10] Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning*. Massachusetts: MIT Press.
- [11] Lana, I., Del Ser, J., Velez, M., & Vlahogianni, E. (2018). Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intelligent Transportation Systems Magazine*, 10(2), 93-109.

- [12] Van Arem, B., Kirby, H., Van Der Vlist, M., & Whittaker, J. (1997). Recent advances and applications in the field of short-term traffic forecasting. *International Journal Of Forecasting*, 13(1), 1-12.
- [13] Vlahogianni, E., Golias, J., & Karlaftis, M. (2004). Short - term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, 24(5), 533-557.
- [14] Bolshinsky E., & Freidman R. (2012). Traffic flow forecast survey. Technion—Israel Institute of Technology, Tech. Rep.
- [15] Vlahogianni, E., Karlaftis, M., & Golias, J. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.
- [16] van Lint, H. J., & van Hinsbergen, C. P. (2012). Short-term traffic and travel time prediction models. *Artif. Intell. Applicat. Crit. Transport. Issues*, vol. 22, pp. 22–41.
- [17] Wolpert, D., & Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions On Evolutionary Computation*, 1(1), 67-82.
- [18] Castillo, E., Grande, Z., Calviño, A., Szeto, W., & Lo, H. (2015). A State-of-the-Art Review of the Sensor Location, Flow Observability, Estimation, and Prediction Problems in Traffic Networks. *Journal of Sensors*, 2015, 1-26.
- [19] Asencio-Cortés, G., Florido, E., Troncoso, A., & Martínez-Álvarez, F. (2016). A novel methodology to predict urban traffic congestion with ensemble learning. *Soft Computing*, 20(11), 4205-4216.
- [20] Yi, H., Jung, H., & Bae, S. (2017). Deep neural networks for traffic flow prediction. Paper presented at the 2017 I.E. International Conference on Big Data and Smart Computing (BigComp), Jeju, South Korea (pp. 328–331). IEEE.
- [21] Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors*, 17(4), 818.

- [22] Fouladgar, M., Parchami, M., Elmasri, R., & Ghaderi, A. (2017). Scalable deep traffic flow neural networks for urban traffic congestion prediction. 2017 International Joint Conference on Neural Networks (IJCNN), 2251-2258.
- [23] Polson, N., & Sokolov, V. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1-17.
- [24] Gao Y., Zheng L., Zhao D., Wu Y., Wang G. (2018) Short Term Prediction Models of Mobile Network Traffic Based on Time Series Analysis. In: Gu X., Liu G., Li B. (eds) *Machine Learning and Intelligent Communications. MLICom 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 226. Springer, Cham
- [25] Pieper, F. K. A. H. S., & Sauer, P. (2017). Traffic prediction using a deep learning paradigm. Paper presented at the EDBT/ICDT 2017 Joint Conference, Venice, Italy.
- [26] Epelbaum, T., Gamboa, F., Loubes, J. M., & Martin, J. (2017). Deep learning applied to road traffic speed forecasting. arXiv preprint arXiv:1710.08266.
- [27] Kanestrøm, P. (2017). Traffic flow forecasting with deep learning (master's thesis). NTNU.
- [28] Tian, Y., & Pan, L. (2015). Predicting short-term traffic flow by long short-term memory recurrent neural network. Paper presented at the 2015 I.E. International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, China (pp. 153–158). IEEE.
- [29] Epelbaum, T., Gamboa, F., Loubes, J. M., & Martin, J. (2017). Deep learning applied to road traffic speed forecasting. arXiv preprint arXiv:1710.08266.
- [30] Song, C., Lee, H., Kang, C., Lee, W., Kim, Y. B., & Cha, S. W. (2017). Traffic speed prediction under weekday using convolutional neural networks concepts. Paper presented at the 2017 I.E. Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA (pp. 1293–1298). IEEE.
- [31] Tu, W., Xiao, F., Fu, L., & Pan, G. (2017). A deep learning model for traffic flow state classification based on smart phone sensor data. arXiv preprint arXiv:1709.08802.
- [32] Dai, X., Fu, R., Lin, Y., Li, L., & Wang, F. Y. (2017). DeepTrend: A deep hierarchical neural network for traffic flow prediction. arXiv preprint arXiv:1707.03213.

- [33] Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. Paper presented at the Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China (pp. 324–328). IEEE.
- [34] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122.
- [35] Jia, Y., Wu, J., & Du, Y. (2016). Traffic speed prediction using deep learning method. Paper presented at the 2016 I.E. 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil (pp. 1217–1222).
- [36] Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865–873.
- [37] Zang, D., Wang, D., Cheng, J., Tang, K., & Li, X. (2017). Traffic parameters prediction using a Three-Channel convolutional neural network. In Z. Shi, B. Goertzel, & J. Feng (Eds.), *International conference on intelligence science* (pp. 363–371). Cham, Switzerland: Springer.
- [38] Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68–75.
- [39] Clark, S.D., Dougherty, M.S., & Kirby, H.R. (1993). THE USE OF NEURAL NETWORKS AND TIME SERIES MODELS FOR SHORT TERM TRAFFIC FORECASTING: A COMPARATIVE STUDY.
- [40] Kirby, H. R., Watson, S. M., & Dougherty, M. S. (1997). Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International Journal of Forecasting*, 13(1), 43–50.
- [41] Danech-Pajouh, M., & Aron, M. (1991). TRAFFIC MANAGEMENT. RTS ENGLISH ISSUE NUMBER 6. ATHENA: A METHOD FOR SHORT-TERM INTER-URBAN MOTORWAY TRAFFIC FORECASTING.
- [42] Smith, B. L., & Demetsky, M. J. (1997). Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, 123(4), 261–266.

- [43] Smith, B., Williams, B., & Keith Oswald, R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10(4), 303-321.
- [44] Tang, Y., Lam, W., & Ng, P. (2003). Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong. *Journal Of Transportation Engineering*, 129(3), 271-277.
- [45] Vlahogianni, E., & Karlaftis, M. (2013). Testing and Comparing Neural Network and Statistical Approaches for Predicting Transportation Time Series. *Transportation Research Record: Journal Of The Transportation Research Board*, 2399(1), 9-22.
- [46] Karlaftis, M., & Vlahogianni, E. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387-399.
- [47] Howell, Shaun. (2018). Meta-analysis of machine learning approaches to short-term urban traffic prediction.
- [48] Bratsas, C., Koupidis, K., Salanova, J., Giannakopoulos, K., Kaloudis, A., & Aifadopoulou, G. (2019). A Comparison of Machine Learning Methods for the Prediction of Traffic Speed in Urban Places. *Sustainability*, 12(1), 142.
- [49] Ahmed, N., Atiya, A., Gayar, N., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, 29(5-6), 594-621.
- [50] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.
- [51] Cerqueira, Vitor & Torgo, Luís & Soares, Carlos. (2019). Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters.
- [52] Klein, L., Mills, M., & Gibson, D. (2006). *Traffic detector handbook*. McLean, VA: U.S. Dept. of Transportation, Federal Highway Administration, Research, Development, and Technology, Turner-Fairbank Highway Research Center.

- [53] Nagy, A., & Simon, V. (2018). Survey on traffic prediction in smart cities. *Pervasive And Mobile Computing*, 50, 148-163.
- [54] Vlahogianni, E., & Karlaftis, M. (2011). Temporal aggregation in traffic data: implications for statistical characteristics and model choice. *Transportation Letters*, 3(1), 37-49.
- [55] Wickham, H., & Sievert, C. *ggplot2*.
- [56] Adhikari, R., & Agrawal, R. (2013). *An Introductory Study on Time Series Modeling and Forecasting*. Saarbrücken: LAP LAMBERT Academic Publishing.
- [57] Hyndman, R., & Athanasopoulos, G. (2015). *Forecasting principles and practice*. USA: O texts.
- [58] Cowpertwait, P., & Metcalfe, A. (2009). *Introductory time series with R*. New York, N.Y.: Springer.
- [59] Cochrane, H., J. (1997). *Time Series for Macroeconomics and Finance*, Graduate School of Business, University of Chicago.
- [60] Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root". *Journal of Econometrics*. 54 (1–3): 159–178.
- [61] Polit DF Beck CT (2012). *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 9th ed. Philadelphia, USA: Wolters Klower Health, Lippincott Williams & Wilkins.
- [62] Little, Roderick J. A.; Rubin, Donald B. (2002), *Statistical Analysis with Missing Data* (2nd ed.), Wiley.
- [63] Grubbs, F. E. (1969). "Procedures for detecting outlying observations in samples". *Technometrics*. 11 (1): 1–21.
- [64] Maddala, G. S. (1992). "Outliers". *Introduction to Econometrics* (2nd ed.). New York: MacMillan. pp. 89. ISBN 978-0-02-374545-4.
- [65] <https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques>

- [66] Jin, W., Tung, A., & Han, J. (2001). Mining top-n local outliers in large databases. Proceedings Of The Seventh ACM SIGKDD International Conference On Knowledge Discovery And Data Mining - KDD '01.
- [67] <https://cran.r-project.org/web/packages/anomalize/anomalize.pdf>
- [68] Rosner, B., (1983), "Percentage Points for a Generalized ESD Many-Outlier Procedure", *Technometrics*, 25(2), pp. 165-172.
- [69] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–33.
- [70] Phillips, P.C.B. and Perron, P. (1988) "Testing for a unit root in time series regression", *Biometrika*, 72(2), 335-346
- [71] Wang, X, Smith, KA, Hyndman, RJ (2006) "Characteristic-based clustering for time series data", *Data Mining and Knowledge Discovery*, 13(3), 335-364.
- [72] Canova F and Hansen BE (1995) "Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability", *Journal of Business and Economic Statistics* 13(3):237-252.
- [73] Hylleberg S, Engle R, Granger C and Yoo B (1990) "Seasonal integration and cointegration.", *Journal of Econometrics* 44(1), pp. 215-238.
- [74] Osborn DR, Chui APL, Smith J, and Birchenhall CR (1988) "Seasonality and the order of integration for consumption", *Oxford Bulletin of Economics and Statistics* 50(4):361-377.
- [75] Wang, X., Smith, K. A., & Hyndman, R. J. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, 13(3), 335–364.
- [76] Guerrero, V.M. (1993) Time-series analysis supported by power transformations. *Journal of Forecasting*, 12, 37–48.
- [77] Yentes, J.M., N. Hunt, K.K. Schmid, J.P. Kaipust, D. McGrath, N. Stergiou (2012). The Appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.*

- [78] <https://robjhyndman.com/hyndsight/longseasonality/>
- [79] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(1), 1–22.
- [80] <https://robjhyndman.com/hyndsight/arimax/>
- [81] Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer-Verlag.
- [82] Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.
- [83] Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31(10), 1237–1246.
- [84] Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.
- [85] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.
- [86] De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *J American Statistical Association*, 106(496), 1513–1527.
- [87] Assimakopoulos, V. and Nikolopoulos, K. (2000) The theta model: a decomposition approach to forecasting. *International Journal of Forecasting* 16, 521–530.
- [88] Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media, Incorporated.
- [89] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.



- [90] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Proceedings of The 2014 Conference On Empirical Methods in Natural Language Processing (EMNLP).
- [91] LeCun, Y. (1989). Generalization and network design strategies. Technical Report CRG-TR-89-4, University of Toronto.
- [92] Pal, A., & Prakash, P. (2017). Practical time-series analysis.
- [93] Oord, Aaron van den; Dieleman, Sander; Zen, Heiga; Simonyan, Karen; Vinyals, Oriol; Graves, Alex; Kalchbrenner, Nal; Senior, Andrew; Kavukcuoglu, Koray (2016). WaveNet: A Generative Model for Raw Audio. 1609. arXiv:1609.03499
- [94] Klaas, J. (2019). Machine Learning for Finance. Birmingham: Packt Publishing, Limited.
- [95] Noshaba Cheema, Somayeh Hosseini, Janis Sprenger, Erik Herrmann, Han Du, Klaus Fischer, and Philipp Slusallek. (2018). Dilated temporal fully-convolutional network for semantic segmentation of motion capture data. In Proceedings of the Conference on Computer Animation (SCA Posters '18). Eurographics Association, Goslar, DEU, 5–6.
- [96] Shi, Xingjian & Chen, Zhourong & Wang, Hao & Yeung, Dit-Yan & Wong, Wai Kin & WOO, Wang-chun. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.
- [97] Gers, F., & Schmidhuber, J. (2000). Recurrent nets that time and count. Proceedings Of The IEEE-INNS-ENNS International Joint Conference On Neural Networks. IJCNN 2000. Neural Computing: New Challenges And Perspectives For The New Millennium.
- [98] Billah, B., Hyndman, R. J. & Koehler, A. B. (2005), 'Empirical information criteria for time series forecasting model selection', Journal of Statistical Computation and Simulation 522 75, 831–840.
- [99] Quinn, J., McEachen, J., Fullan, M., Gardner, M., & Drummy, M. Dive into deep learning.

- [100] Floris Takens. Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80, chapter Detecting strange attractors in turbulence, pages 366–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 1981.
- [101] Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437-450.
- [102] Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
- [103] Gorr WL. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*. 10(1):1±4.
- [104] Zhang GP, & Qi M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 501-514.
- [105] HamzacËebi C., Akay D., & Kutay F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications*, 36(2, Part 2), 3839-3844.
- [106] Sakamoto Y., Ishiguro M., & Kitagawa G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.
- [107] Bontempi, G., Ben Taieb, S., & Le Borgne, Y. (2013). Machine Learning Strategies for Time Series Forecasting. *Business Intelligence*, 62-77.
- [108] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
- [109] Yuan, Y., Rosasco, L., & Caponnetto, A. (2007). "On Early Stopping in Gradient Descent Learning". *Constructive Approximation*. 26 (2): 289–315.
- [110] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580.

[111] Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15). JMLR.org, 448–456.

[112] Bishop, Christopher M. (2007). Pattern recognition and machine learning, New York: Springer.

[113] <https://eng.uber.com/m4-forecasting-competition/>