

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΡΟΒΛΕΨΗ ΘΕΣΗΣ
ΣΕ
ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ**

Αριστείδης Θ. Μπούρδας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Ιούλιος 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΡΟΒΛΕΨΗ ΘΕΣΗΣ
ΣΕ
ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ**

Αριστείδης Θ. Μπούρδας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Ιούλιος 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Ιωάννης Θεοδωρίδης, Καθηγητής Τμήματος Πληροφορικής (Επιβλέπων)
- Πελέκης Νίκος, Λέκτορας Τμήματος Στατ. & Ασφαλ. Επιστήμης
- Κοφίδης Ελευθέριος, Επίκουρος Καθηγητής Τμήματος Στατ. & Ασφαλ. Επιστήμης

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**PREDICTION OF FUTURE
LOCATION IN MOBILITY DATA**

Aristeides T. Mpourdas

M.Sc. Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfillment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece
July 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Θέλω να αφιερώσω την παρούσα διπλωματική εργασία στους γονείς, τον αδερφό μου και την κοπελάρα μου, για την υλική και ηθική στήριξη, τη συμπαράσταση και τη συνεχή τους ενθάρρυνση.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Ευχαριστίες

Θέλω να ευχαριστήσω όλους όσους με την κατανόηση και την υπομονή τους με βοήθησαν στην εκπόνηση αυτής της διπλωματικής εργασίας.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Περίληψη

Σκοπός της διπλωματικής εργασίας μας ήταν να προβλέψουμε τη θέση σε κινούμενα δεδομένα. Συγκεκριμένα, να προβλέψουμε τη θέση ενός κινούμενου αντικειμένου με τη βοήθεια των προηγούμενων χωρικών και χρονικών καταγραφών του.

Στη συνέχεια, έπρεπε να επιλέξουμε τη μέθοδο με την οποία θα κάναμε την πρόβλεψή μας, με γνώμονα ότι θα έπρεπε να βασιζόταν σε μια στατιστική μέθοδο. Η μέθοδος που επιλέξαμε ήταν οι χρονοσειρές. Θεωρήσαμε το γεωγραφικό πλάτος και μήκος σαν δύο χρονοσειρές, οπότε χρησιμοποιώντας υποδείγματα χρονοσειρών κάναμε την πρόβλεψή μας για μερικές περιόδους. Τέλος, χρησιμοποιώντας ένα μέτρο απόδοσης που δημιουργήσαμε αξιολογήσαμε την απόδοση της πρόβλεψής μας. Τα αποτελέσματά μας ήταν ικανοποιητικά κυρίως για βραχυπρόθεσμες προβλέψεις.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Abstract

The aim of this master thesis was to predict the future position of moving data. Specifically, was to predict the future position of a moving object based on its previous spatiotemporal data.

Then we had to choose the method by which we were going to make our prediction, based on the fact that this method must be a statistical method. The method that we have chosen is Time Series Analysis. Supposing that the longitude and the latitude was two time series, we have used a time series model in order to make a prediction for some periods. Furthermore, we have used a performance measure that we have created for testing the accuracy of our prediction. Our results was satisfactory mainly for short-term forecasts.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΑΙΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Περιεχόμενα

Κατάλογος Πινάκων.....	xv
Κατάλογος Εικόνων	xvii
Κατάλογος Συντομογραφιών	xix
ΕΙΣΑΓΩΓΗ	1
ΚΕΦΑΛΑΙΟ 1.....	3
1.1 «Πρόβλεψη θέσης σε δεδομένα κίνησης» και σχετική βιβλιογραφία	3
1.2 Συναφείς εργασίες.....	3
ΚΕΦΑΛΑΙΟ 2.....	11
2.1 Γραφική απεικόνιση κεντρικής ιδέας	11
2.2 Αναλυτική περιγραφή	12
2.3 Βασικές έννοιες χωροχρονικών δεδομένων.....	12
2.4 Χρονοσειρές.....	13
ΚΕΦΑΛΑΙΟ 3.....	25
3.1 Γενικές πληροφορίες των δεδομένων	25
3.2 Παρουσίαση των δεδομένων μας.....	26
3.3 Εύρεση κατάλληλων μοντέλων χρονοσειρών	27
3.3.1 Υπόδειγμα για την χρονοσειρά Longitude.....	30
3.3.2 Υπόδειγμα για την χρονοσειρά Latitude.....	33
3.4 Πρόβλεψη τιμών ανά χρονοσειρά.....	37
3.4.1 Πρόβλεψη για την χρονοσειρά Longitude.....	37
3.4.2 Πρόβλεψη για την χρονοσειρά Latitude	40
3.4.3 Πρώτα συμπεράσματα	43
3.5 Πρόβλεψη μελλοντικής θέσης.....	44
3.6 Αξιολόγηση της πρόβλεψης.....	46
3.7 Δευτερεύουσα πρακτική εφαρμογή	48
ΚΕΦΑΛΑΙΟ 4.....	53
4.1 Συμπεράσματα	53
ΠΑΡΑΡΤΗΜΑ 1	55
ΠΑΡΑΡΤΗΜΑ 2	59
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	65

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Κατάλογος Πινάκων

Πίνακας 3.3-1. Ο πίνακας των αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος	27
Πίνακας 3.3-2. Τύπος του υποδείγματος που εκτιμήθηκε ARIMA(2,2,0) για την χρονοσειρά Longitude	30
Πίνακας 3.3-3. Αποτελέσματα ελέγχων για ARIMA(2,2,0) για την χρονοσειρά Longitude	30
Πίνακας 3.3-4. Οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος για την χρονοσειρά Longitude	30
Πίνακας 3.3-5. Ο πίνακας των αυτοσυσχετίσεων της χρονοσειράς των σφαλμάτων..	32
Πίνακας 3.3-6. Οι τιμές των παραμέτρων του ARIMA(3,2,0) υποδείγματος για την χρονοσειρά Longitude	33
Πίνακας 3.3-7. Οι τιμές των παραμέτρων του ARIMA(2,2,1) υποδείγματος για την χρονοσειρά Longitude	33
Πίνακας 3.3-8. Τύπος του υποδείγματος που εκτιμήθηκε ARIMA(1,1,0) για την χρονοσειρά Latitude	34
Πίνακας 3.3-9. Αποτελέσματα ελέγχων για ARIMA(1,1,0) για την χρονοσειρά Latitude.....	34
Πίνακας 3.3-10. Οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος για την χρονοσειρά Latitude	34
Πίνακας 3.3-11. Ο πίνακας των αυτοσυσχετίσεων της χρονοσειράς των σφαλμάτων	36
Πίνακας 3.3-12. Οι τιμές των παραμέτρων του ARIMA(2,1,0) υποδείγματος για την χρονοσειρά Latitude	37
Πίνακας 3.3-13. Οι τιμές των παραμέτρων του ARIMA(1,1,1) υποδείγματος για την χρονοσειρά Latitude	37
Πίνακας 3.4-1. Οι προβλεπόμενες τιμές της χρονοσειράς Longitude μαζί με τα πάνω και κάτω όρια	38
Πίνακας 3.4-2. Οι προβλεπόμενες και πραγματικές τιμές της χρονοσειράς Longitude	39
Πίνακας 3.4-3. Οι προβλεπόμενες τιμές της χρονοσειράς Latitude μαζί με τα πάνω και κάτω όρια	41
Πίνακας 3.4-4. Οι προβλεπόμενες και πραγματικές τιμές της χρονοσειράς Latitude .	42
Πίνακας 3.7-1. Πίνακας των 50 επιλεγμένων ταξί.....	49
Πίνακας 3.7-2. Τα δημοφιλέστερα υποδείγματα τροχιών.....	49
Πίνακας 3.7-3. Αποτελέσματα καλής εφαρμογής των υποδειγμάτων χρονοσειρών ...	50
Πίνακας 3.7-4 Αποτελέσματα καλής εφαρμογής των υποδειγμάτων χρονοσειρών	51
Πίνακας 0-1. Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(0,1,1)	59
Πίνακας 0-2: Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(1,1,0)	59
Πίνακας 0-3. Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(0,1,2)	60
Πίνακας 0-4. Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(2,1,0)	60

Πίνακας 0-5. Έλεγχος των όρων μέσω t-test ARIMA(0,1,1)	61
Πίνακας 0-6. Έλεγχος των όρων μέσω t-test ARIMA(1,1,0)	62
Πίνακας 0-7. Έλεγχος των όρων μέσω t-test ARIMA(0,1,2)	63
Πίνακας 0-8. Έλεγχος των όρων μέσω t-test ARIMA(2,1,0)	64

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Κατάλογος Εικόνων

Εικόνα 1.2-1. Τροχιά αντικειμένου αναπαριστάμενη από MBB (Sigal Elnekave et al. (2007)).....	3
Εικόνα 1.2-2. Ανακριβής πρόβλεψη	5
Εικόνα 1.2-3. Παράδειγμα ενός T-pattern (Anna Monreale et al. (2009))	6
Εικόνα 1.2-4. Network constrained trajectories (Yan, Z. (2010))	8
Εικόνα 2.1-1. Σχεδιάγραμμα διπλωματικής.....	11
Εικόνα 3.1-1. Ιστογράμματα της χρονικής και γεωγραφικής απόστασης μεταξύ δύο σημείων (Yuan, J et al. (2011)).....	25
Εικόνα 3.2-1. Ιστογράμματα της χρονικής και γεωγραφικής απόστασης μεταξύ δύο σημείων για το ταξί με id 366	26
Εικόνα 3.3-1. Διαγράμματα αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος.....	28
Εικόνα 3.3-2. Διαγράμματα αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος σε πρώτες διαφορές.....	28
Εικόνα 3.3-3. Διάγραμμα αυτοσυσχετίσεων για το γεωγραφικό πλάτος σε δεύτερες διαφορές	29
Εικόνα 3.3-4. Διαγράμματα μερικών αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος με δεύτερες διαφορές και δεξιά για το γεωγραφικό μήκος σε πρώτες διαφορές	29
Εικόνα 3.3-5. Γραφική απεικόνιση του υποδείγματος.....	31
Εικόνα 3.3-6. Διαγράμματα αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων	31
Εικόνα 3.3-7. Γραφική απεικόνιση του υποδείγματος.....	35
Εικόνα 3.3-8. Διαγράμματα αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων	35
Εικόνα 3.4-1. Γραφική απεικόνιση της πρόβλεψης με το υπόδειγμα APIMA(2,2,0)	38
Εικόνα 3.4-2. Γραφική σύγκριση προβλεπόμενων με πραγματικών τιμών της χρονοσειράς Longitude	39
Εικόνα 3.4-3. Η γραφική απεικόνιση της διαφοράς των προβλεπόμενων με των πραγματικών τιμών της χρονοσειράς Longitude	40
Εικόνα 3.4-4. Γραφική απεικόνιση της πρόβλεψης με το υπόδειγμα APIMA(1,1,0)	40
Εικόνα 3.4-5. Γραφική σύγκριση προβλεπόμενων με πραγματικών τιμών της χρονοσειράς Latitude	41
Εικόνα 3.4-6. Η γραφική απεικόνιση της διαφοράς των προβλεπόμενων με των πραγματικών τιμών της χρονοσειράς Latitude.....	42
Εικόνα 3.4-7. Χάρτης που απεικονίζει τις δύο λανθασμένες καταγραφές	43
Εικόνα 3.5-1. Γραφική απεικόνιση προβλεπόμενης πορείας.....	44
Εικόνα 3.5-2. Γραφική απεικόνιση προβλεπόμενης πορείας με τα πάνω και κάτω όρια της πρόβλεψης.....	45

Εικόνα 3.5-3. Απεικόνιση σε πραγματικό χάρτη των πραγματικών αλλά και των προβλεπόμενων θέσεων του αντικειμένου.....	46
Εικόνα 3.6-1. Γραφική απεικόνιση του μέτρου	46
Εικόνα 3.6-2. Απόσταση της προβλεπόμενης από την πραγματική θέση, καθ' όλη την διάρκεια της τροχιάς του ταξί 366	47
Εικόνα 3.6-3. Απόδοση της μεθόδου που χρησιμοποιήσαμε.....	48

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Κατάλογος Συντομογραφιών

MBB	<i>Minimal Bounding Box</i>
GPS	<i>Global Position System</i>
V-ARIMA	<i>Vector ARIMA</i>
ST-ARIMA	<i>Spatiotemporal ARIMA</i>
Traj-ARIMA	<i>Trajectory ARIMA</i>
ε.σ.	<i>Επίπεδο σημαντικότητας</i>

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΕΙΣΑΓΩΓΗ

Η ραγδαία τεχνολογική ανάπτυξη των τελευταίων δεκαετιών έχει επιφέρει πολλές αλλαγές στην καθημερινότητα μας, ως προς την βελτίωση της. Μια τέτοια αλλαγή είναι η όλο και μεγαλύτερη χρήση φορητών συσκευών με ενσωματωμένο δέκτη GPS (*Παγκόσμιο σύστημα εντοπισμού θέσης*), όπως έξυπνα κινητά τηλέφωνα και φορητοί υπολογιστές. Η χρήση τέτοιων συσκευών, όπως είναι επακόλουθο δημιουργούν ίχνη με την τρέχουσα τοποθεσία των χρηστών των συσκευών αυτών, συγκεκριμένα με το γεωγραφικό πλάτος και μήκος, αλλά και τον χρόνο. Τέτοιου τύπου δεδομένα, καλούνται χωροχρονικά (*spatio-temporal*).

Η ευρεία ύπαρξη πλέον χωροχρονικών δεδομένων δημιούργησε την ανάγκη ανάλυσή τους για την εξαγωγή γνώσης από αυτά. Οι πιο διαδεδομένες εφαρμογές, χρήσης τέτοιων δεδομένων, είναι για την εξαγωγή συχνών τροχιών κίνησης και για την πρόβλεψη της μελλοντικής θέσης. Συγκεκριμένα με την εξαγωγή συχνών τροχιών εννοούμε, για παράδειγμα, την εύρεση συχνών σημείων ενδιαφέροντος σε μια πόλη, όπως επίσης και την εύρεση συχνής χρήσης ενός δρόμου. Μερικοί τομείς στους οποίους απευθύνεται η παραπάνω μέθοδος, είναι ο τομέας του μάρκετινγκ για την βελτιστοποίηση και την καλύτερη στόχευση στην προώθηση προϊόντων. Ένας άλλος τομέας είναι και ο τομέας των υπηρεσιών πλοήγησης, ο οποίος χρησιμοποιεί την συγκεκριμένη μέθοδο για την διευκόλυνση των οδηγών, παρουσιάζοντάς τους την κίνηση των δρόμων, ώστε να αποφύγουν διαδρομές με μεγάλη κίνηση, για την ευκολότερη και γρηγορότερη άφιξη στον προορισμό τους. Πέρα λοιπόν από την εύρεση συχνών τροχιών, η άλλη εφαρμογή των χωροχρονικών δεδομένων, όπως ήδη έχουμε αναφέρει είναι και για την πρόβλεψη της μελλοντικής θέσης. Με τον όρο πρόβλεψη μελλοντικής θέσης εννοούμε, με την χρήση της πληροφορίας που ήδη έχουμε από ένα κινούμενο αντικείμενο να προβλέψουμε ποια θα είναι η θέση του μετά από κάποιο συγκεκριμένο χρονικό διάστημα. Ένας τομέας που απαντά η παραπάνω μέθοδος είναι ο τομέας των διασωστικών φορέων. Για παράδειγμα έστω ένα αεροπλάνο, πλοίο ή οποιοδήποτε άλλο κινούμενο αντικείμενο σταματήσει να εκπέμπει σήμα με τις συντεταγμένες του, οι φορείς που είναι υπεύθυνοί για την εύρεσή τους, με την χρήση της προηγούμενης πληροφορίας μπορούν να προβλέψουν την μελλοντική θέση τους σε κάποιο συγκεκριμένο χρονικό διάστημα μετά την τελευταία θέση τους.

Στην συγκεκριμένη διπλωματική εργασία εμείς θα εστιάσουμε στην εφαρμογή μιας στατιστικής μεθόδου με σκοπό την πρόβλεψη της μελλοντικής θέσης ενός αντικειμένου.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

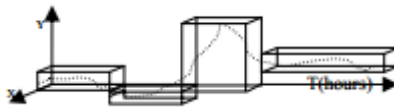
ΚΕΦΑΛΑΙΟ 1

1.1 «Πρόβλεψη θέσης σε δεδομένα κίνησης» και σχετική βιβλιογραφία

Υπάρχουν πάρα πολλές μελέτες, οι οποίες έχουν ασχοληθεί με το συγκεκριμένο πρόβλημα, δηλαδή την πρόβλεψη θέσης σε δεδομένα κίνησης. Οι περισσότερες από αυτές χρησιμοποιούν μοντέλα, που βασίζονται στα συχνά πρότυπα και στους κανόνες συσχέτισης. Συγκεκριμένα μερικές μελέτες προβλέπουν την μελλοντική θέση χρησιμοποιώντας όλα τα δεδομένα των κινούμενων αντικειμένων μιας βάσης δεδομένων, ενώ κάποιες άλλες μελέτες βασίζονται μόνο στα δεδομένα του αντικειμένου που θέλουν για να προβλέψουν την μελλοντική του θέση.

1.2 Συναφείς εργασίες

Οι Sigal Elnekave et al. (2007) περιγράφουν μια μέθοδο με την οποία χρησιμοποιώντας το κέντρο της συστάδας (*cluster's centroid*), συσταδοποιημένων χωροχρονικών δεδομένων, προβλέπουν την μελλοντική θέση κινούμενων αντικειμένων. Πιο αναλυτικά, αναπαριστούν μια τροχιά (*trajectory*) ως σύνολο πλαισίων, όπου κάθε πλαίσιο (MBB^1) αντιπροσωπεύει ένα διάστημα οριοθετημένο από το χρόνο και την τοποθεσία (εικόνα 1.2-1). Το διάστημα οριοθέτησης του πλαισίου (MBB) είναι ορισμένο από την αρχή με την ιδιότητα, ότι όταν το διάστημα αυτό είναι μεγάλο να κερδίζει σε ταχύτητα επεξεργασίας, αλλά να χάνει σε ακρίβεια, ενώ αντίθετα όταν είναι μικρό να κερδίζει σε ακρίβεια, αλλά να χάνει σε ταχύτητα επεξεργασίας.



Εικόνα 1.2-1. Τροχιά αντικειμένου αναπαριστάμενη από MBB (Sigal Elnekave et al. (2007))

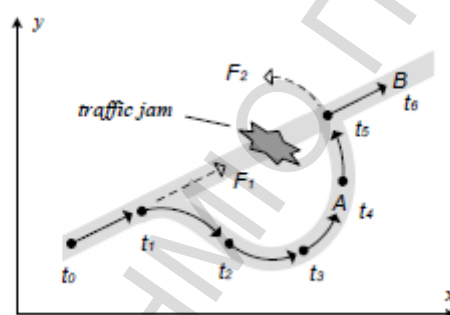
Από εκεί και πέρα οι συγγραφείς, την υπόλοιπη διαδικασία αλλά και την τελική πρόβλεψη την στηρίζουν σε αυτά τα πλαίσια (MBB). Συνεχίζοντας λοιπόν τη διαδικασία τους, ορίζουν ως μέτρο ομοιότητας μεταξύ δύο τροχιών, το άθροισμα των ομοιοτήτων μεταξύ δύο επικαλυπτόμενων πλαισίων, διαιρούμενο από το σύνολο των πλαισίων των κάθε συγκρινόμενων τροχιών.

¹ Minimal Bounding Box

Για να εντοπίσουν τις ομοιότητες μεταξύ των πλαισίων χρησιμοποιούν δύο εμπειρικά μέτρα. Το πρώτο μέτρο ονομάζεται «ελάχιστες αποστάσεις» (“*minimal distances*”), όπου ορίζει την απόσταση μεταξύ δύο πλαισίων μιας τροχιάς, ως το κατώτατο όριο της απόστασης μεταξύ των αρχικών δεδομένων. Το δεύτερο μέτρο ονομάζεται «δεδομένα βασισμένα στην ποσότητα» (“*data-amount-based*”), το οποίο πολλαπλασιάζει το μέτρο «ελάχιστες αποστάσεις» με τη διαφορά των δύο συγκρινόμενων πλαισίων, ώστε να διασφαλίσουν την ορθότητα των αποτελεσμάτων. Στην συνέχεια, για να συσταδοποιήσουν τις τροχιές, χρησιμοποιούν μια παραλλαγή του αλγορίθμου K-Means που να μπορεί να χειρίζεται οριοθετημένα διαστήματα και όχι διανύσματα. Κάθε κέντρο μιας συσταδοποιημένης τροχιάς αναπαριστά όλες τις τροχιές που ανήκουν σε αυτή την συστάδα, στην συγκεκριμένη περίπτωση τα πλαίσια. Επιπλέον για να βελτιστοποιήσουν την διαδικασία πρόβλεψης, αφαίρεσαν πλαίσια από τις συστάδες που δεν ήταν συχνά, μέσω ενός αλγορίθμου που είχε ως σκοπό την εύρεση του βέλτιστου αριθμού των μη συχνών τροχιών. Φτάνοντας στην διαδικασία πρόβλεψης, αυτό που κάνανε για να προβλέψουν τη μελλοντική θέση ενός κινούμενου αντικειμένου ήταν να ψάχνουν σε κάθε κέντρο μιας συστάδας, για ένα πλαίσιο που να περιέχει την μελλοντική χρονική στιγμή, μεταξύ του χρονικού ορίου ενός πλαισίου, επιστρέφοντας σαν x και y το μέγιστο και το ελάχιστο όριο αντίστοιχα. Εάν δεν υπάρχει ένα τέτοιο πλαίσιο, τότε θα δώσει σαν αποτέλεσμα το κοντινότερο πλαίσιο στη χρονική στιγμή που ψάχνουν. Επίσης στην περίπτωση που βρεθούν παραπάνω από ένα πλαίσια, πρώτον ο αλγόριθμος θα πρέπει να επιστρέψει όλα τα πλαίσια με τις συντεταγμένες τους ή δεύτερον να επιστρέψει το πλαίσιο με τα περισσότερα δεδομένα που περιέχονται σε αυτό. Να αναφέρουμε ότι οι συγγραφείς χρησιμοποίησαν την πρώτη μέθοδο γιατί διασφαλίζει την μη εξαίρεση πιθανών πλαισίων. Τέλος για την αξιολόγηση της μεθόδου τους χρησιμοποίησαν ένα σετ δεδομένων σαν δεδομένα εκπαίδευσης του αλγορίθμου τους και άλλο σετ σαν δεδομένα δοκιμής. Στην συνέχεια όρισαν σαν *recall* το ποσοστό μιας επιτυχημένης πρόβλεψης, όπου μια μελλοντική τοποθεσία βρίσκονταν μέσα στο προβλεπόμενο πλαίσιο και σαν *precision* 1 (πιθανότητα 100%) εάν η μελλοντική τοποθεσία βρεθεί μεταξύ των προβλεπόμενων πλαισίων, ενώ σε κάθε άλλη περίπτωση υπολογίζεται ως τον αριθμό 1 διαιρούμενο με την μικρότερη απόσταση μεταξύ της μελλοντικής και της προβλεπόμενης θέσης.

Επίσης οι Hoyoung Jeung et al. (2008) παρουσιάζουν μια υβριδική μέθοδο πρόβλεψης μελλοντικής θέσης τόσο για βραχυπρόθεσμα όσο και για μακροπρόθεσμα. Η συγκεκριμένη μέθοδος βασίζεται στις συχνές τροχιές (*trajectory patterns*), κάτι που έχει σαν αποτέλεσμα να έχει πιο ακριβείς προβλέψεις για μελλοντικές μακρινές στιγμές, σε σχέση με τα μοντέλα που κάνουν

την πρόβλεψη με την βοήθεια μαθηματικών συναρτήσεων. Για να καταλάβουμε καλύτερα τα εμπόδια που συναντάνε οι άλλοι αλγόριθμοι θα παραθέσουμε δύο παραδείγματα. Έστω ότι ακόμα και να ξέρουμε ότι η Jane ήταν στο σπίτι στις 9:00μμ και περνάει από ένα εμπορικό κέντρο τώρα (9:05μμ), είναι ανούσιο να προβλέψουμε την θέση της τις απογευματινές ώρες βασιζόμενοι στις παραπάνω κινήσεις της. Δηλαδή ότι οι πρόσφατες κινήσεις ενός αντικειμένου είναι χρήσιμες για την πρόβλεψη της μελλοντικής θέσης του μόνο στο κοντινό μέλλον. Επίσης ακόμα και αν μέσω μαθηματικών μεθόδων προβλέψουμε την μελλοντική θέση ενός αντικειμένου, όπως κάνουν οι υπόλοιποι αλγόριθμοι, η πραγματική κίνηση του αντικειμένου είναι εντελώς διαφορετική στην πραγματικότητα. Δηλαδή έστω ότι ένα αυτοκίνητο θέλει να πάει από το ένα σημείο της πόλης στο άλλο, αλλά στην μέση της διαδρομής βρίσκει κίνηση, οπότε είναι αναγκασμένο να πάρει άλλο δρόμο (εικόνα 1.2-2). Αυτό έχει σαν αποτέλεσμα ο αλγόριθμος να αποτύχει στην πρόβλεψή του.

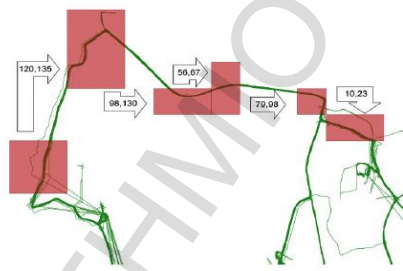


Εικόνα 1.2-2. Ανακριβής πρόβλεψη

Σε αυτό το σημείο έρχεται ο *Hybrid* αλγόριθμος και λειτουργεί ως εξής. Στην αρχή χωρίζει όλες τις τροχιές σε τοποθεσίες και αντίστοιχες χρονικές στιγμές και δημιουργεί ομάδες G_t που αντιστοιχούν σε όλες τις τοποθεσίες που βρέθηκε το κινούμενο αντικείμενο σε κάθε χρονική στιγμή t . Μέσω των ομάδων G_t και εφαρμόζοντας κάποια μέθοδο συσταδοποίησης προκύπτουν πυκνές συστάδες R_t που αντιστοιχούν στις τοποθεσίες που το κινούμενο αντικείμενο μπορεί να εμφανιστεί την χρονική στιγμή t . Γενικά ονομάζουμε τις G_t συστάδες ως συχνές περιοχές στην t στιγμή (*frequent region at t*). Επειδή μπορεί να υπάρχει πάνω από μια τοποθεσίες πιθανές σε κάθε χρονική στιγμή, για να τις ξεχωρίζουν χρησιμοποιείται ένας εκθέτης j που υποδηλώνει τον αριθμό της τοποθεσίας. Συγκεκριμένα στον *Hybrid* αλγόριθμο για να ανακαλύψουν τις συχνές περιοχές χρησιμοποιείται ο *DBSCAN* για μέθοδο συσταδοποίησης (*clustering*). Στην συνέχεια για να ανακαλυφθούν οι συχνές τροχιές από τις συχνές περιοχές που δημιουργήθηκαν χρησιμοποιείται μια τροποποίηση του αλγορίθμου *apriori* που κλαδεύει όλες τις τροχιές που

στο δεύτερο μέρος του κανόνα συσχέτισης δεν έχουν κάποια περιοχή που αντιστοιχεί σε μελλοντική χρονική στιγμή και επίσης κλαδεύει τους κανόνες συσχέτισης που περιέχουν πάνω από ένα στοιχείο στο δεύτερο μέλος του κανόνα εφόσον θέλουμε την πρόβλεψη για μια χρονική στιγμή. Έτσι λοιπόν μειώνεται ο όγκος των εξαγόμενων συχνών τροχιών και ταυτόχρονα και ο χρόνος που χρειάζεται για να εξαχθεί το αποτέλεσμα. Συνοπτικά ο *Hybrid* αλγόριθμος για την πρόβλεψη της μελλοντικής θέσης, αυτό που κάνει, είναι να αλλάζει συμπεριφορά ανάλογα με το αν θέλουμε μακροπρόθεσμη ή βραχυπρόθεσμη πρόβλεψη. Δηλαδή στην μακροπρόθεσμη χρησιμοποιούνται οι εξαγόμενες τροχιές, ενώ για τις βραχυπρόθεσμες προβλέψεις χρησιμοποιείται η συνάρτηση της κίνησης (*motion function*) του κινούμενου σώματος.

Οι Anna Monreale et al. (2009) περιγράφουν μια μέθοδο πρόβλεψης μελλοντικής θέσης κινούμενων αντικειμένων σε ορισμένο επίπεδο ακρίβειας χρησιμοποιώντας δέντρο απόφασης. Σε αντίθεση με τους Sigal Elnekave et al. (2007) που περιγράψαμε την μέθοδο τους πιο πάνω, εδώ κάθε τροχιά ορίζεται σαν ένα σύνολο μερικών προτύπων (*patterns*) (εικόνα 1.2-3).



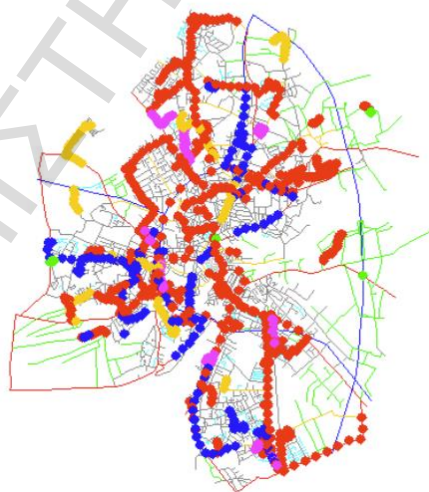
Εικόνα 1.2-3. Παράδειγμα ενός T-pattern (Anna Monreale et al. (2009))

Συγκεκριμένα δημιουργούν τα πρότυπα τροχιάς (*Trajectory Pattern*) μέσω ενός αλγορίθμου, ο οποίος εξάγει ένα σύνολο συχνών τροχιών από πυκνές χωρικές περιοχές οι οποίες έχουν εξαχθεί από τροχιές βάση δύο παραμέτρων, τη σ και τη τ . Η παράμετρος σ είναι το κατώτατο στήριγμα (*support*), που στην περίπτωση των προτύπων τροχιών ορίζεται και ως παράμετρος χωρικής πυκνότητας και η παράμετρος τ είναι ένα διάστημα χρόνου. Στην πράξη ο αλγόριθμος προτύπων τροχιών διαχωρίζει τον χώρο που δουλεύουν, μέσω ενός πλέγματος με κελιά μεγέθους ορισμένα από το χρήστη. Στην συνέχεια η πυκνότητα κάθε κελιού υπολογίζεται βάση κάθε τροχιάς ξεχωριστά. Τέλος ένα σύνολο από συχνές περιοχές εξάγεται μέσω ενός απλού ιεραρχικού αλγορίθμου λαμβάνοντας υπόψη μόνο τα κελιά με πυκνότητα μεγαλύτερη από σ . Με άλλα λόγια κάθε εξαγόμενο πρότυπο τροχιών είναι μια συνοπτική περιγραφή συχνών συ-

μπεριφορών, βάση του χώρου και του χρόνου. Στην συνέχεια χρησιμοποίησαν τα πρότυπα τροχιών σαν κανόνες συσχέτισης για να κατασκευάσουν ένα δέντρο πρότυπων τροχιών (*T-pattern Tree*). Όπου κάθε κόμβος του δέντρου αποτελείται από μια μεταβλητή *id*, η οποία είναι ένα μοναδικό διαχωριστικό από τους άλλους κόμβους, μια μεταβλητή *Region* που ονοματίζει την περιοχή του πρότυπου τροχιάς καθώς και μια μεταβλητή *support* η οποία αποτελεί το στήριγμα του πρότυπου τροχιάς. Επίσης για κάθε ακμή αντιστοιχεί ο χρόνος μεταβολής μεταξύ τροχιών προτύπων. Η πρόβλεψη τους βασίζεται σε τρία βασικά βήματα, πρώτον χρησιμοποιώντας το δέντρο πρότυπων τροχιών ψάχνουν τις καλύτερες αντιστοιχίες, στην εξαγωγή κάποιων πιθανών τοποθεσιών και τέλος στην πρόβλεψη. Συγκεκριμένα στον εντοπισμό των αντιστοιχιών χρησιμοποιούν δύο μέτρα. Το πρώτο μέτρο είναι το σκορ ακρίβειας (*punctual score*), το οποίο υπολογίζει την χωροχρονική απόσταση μεταξύ του τμήματος της τροχιάς (που οριοθετείται στο χρόνο χρησιμοποιώντας τον προηγούμενο χρόνο μετάβασης) και του κόμβου μιας διαδρομής. Το δεύτερο μέτρο είναι το σκορ διαδρομής (*path score*) και είναι η συγκέντρωση όλων των σκορ ακρίβειας κατά μήκος μιας διαδρομής. Αξίζει να αναφέρουμε ότι χρησιμοποίησαν τρεις διαφορετικούς τρόπους υπολογισμού του σκορ διαδρομής. Το *avgScore* είναι το φυσικό μέτρο, το οποίο προσπαθεί την γενίκευση της έννοιας της ομοιότητας ως η μέση απόσταση μεταξύ τροχιάς και κόμβου. Το *sumScore* το οποίο βασίζεται στην έννοια του βάθους, δηλαδή δίνει προτεραιότητα στην μεγαλύτερη διαδρομή που ταιριάζει με την τροχιά. Και τέλος το *maxScore*, το οποίο είναι και το πιο αισιόδοξο, εάν μια τροχιά έχει καλή αντιστοιχία με ένα κόμβο, τότε αυτό έχει προτεραιότητα σε σχέση με τις άλλες. Έτσι την καλύτερη αντιστοιχία την έχουμε με το μεγαλύτερο σκορ διαδρομής με τουλάχιστον μια αποδεκτή πρόβλεψη. Στην συνέχεια ορίζουν μια συνάρτηση αξιολόγησης για την εκτίμηση της προβλεπτικής ικανότητας της συλλογής πρότυπων τροχιών πριν από την δημιουργία του δέντρου πρότυπων τροχιών και την ποιότητα της απόδοσης της πρόβλεψης. Για το σκοπό αυτό όρισαν τρία μέτρα, την ακρίβεια, το μέσο σφάλμα και τέλος το ποσοστό πρόβλεψης. Το μέτρο της ακρίβειας ορίζεται ως το ποσοστό των σωστών προβλεπόμενων περιοχών διαιρούμενο από τον συνολικό αριθμό των τροχιών. Το μέσο σφάλμα το οποίο είναι η μέση χωροχρονική απόσταση μεταξύ των πραγματικών τροχιών, στο ίδιο διάστημα, και της περιοχής που προβλέφθηκε. Και τέλος το ποσοστό πρόβλεψης το οποίο ορίζεται ως ο αριθμός των τροχιών που έχουν προβλέψει διαιρούμενο από το συνολικό αριθμό των τροχιών.

Τέλος μία ακόμα συναφή εργασία, η οποία είναι πιο κοντά στην δικιά μας αντιμετώπιση του προβλήματος της πρόβλεψης θέσης σε δεδομένα κίνησης, δηλαδή στην πρόβλεψη θέσης ενός

κινούμενου αντικειμένου με την βοήθεια ενός στατιστικού μοντέλου είναι η εργασία του Yan, Z. (2010). Συγκεκριμένα, χρησιμοποιεί μια παραλλαγή του υποδείγματος *ARIMA*, των χρονοσειρών, για την πρόβλεψη θέσης ενός κινούμενου αντικειμένου σε ένα δίκτυο δρόμων (*Network-Constrained Trajectory*). Ξεκινάει, λοιπόν περιγράφοντας τις ιδιότητες και βασικές παραμέτρους ενός υποδείγματος *ARIMA*. Στην συνέχεια, αναφέρει για μια ειδική έκδοση του υποδείγματος *ARIMA*, το *Vector ARIMA (V-ARIMA)*, το οποίο είναι μια παραλλαγή του *ARIMA* με κύρια διαφορά ότι το *ARIMA* είναι για μια μονοδιάστατη χρονοσειρά, ενώ αντίθετα το *V-ARIMA* χρησιμοποιείται πολυδιάστατες χρονοσειρές (*vector*). Τέλος, πριν εισάγει για πρώτη φορά την έννοια του *traj-ARIMA (trajectory ARIMA)*, εισάγει μια ακόμη παραλλαγή του *ARIMA* το *Space-time ARIMA*, το οποίο υπόδειγμα βασίζεται στο *V-ARIMA*. Συγκεκριμένα, είναι ένα υπόδειγμα, το οποίο δίνει έμφαση σε χωρικές διαστάσεις με την έννοια χωρικών συσχετίσεων (*spatial correlations*) και όχι μόνο σε χρονικές συσχετίσεις (*temporal correlations*). Με απλά λόγια το υπόδειγμα *ST-ARIMA* εκφράζει κάθε παρατήρηση συναρτήσει του χρόνου και της θέσης, σαν γραμμικό συνδυασμό των προηγούμενων παρατηρήσεων, με υστερήσεις (*lags*) για χρόνο και για χώρο. Στην συνέχεια, περιγράφει τι είναι τροχιά (*trajectory*) καθώς και τι είναι δίκτυο που χρησιμοποιείται σαν περιοριστικό μέσο για τις τροχιές (*Network constrain for trajectories*)(εικόνα 1.2-4).



Εικόνα 1.2-4. Network constrained trajectories (Yan, Z. (2010))

Επίσης ένα δίκτυο δρόμων κυκλοφορίας, μοντελοποιείται σαν να αποτελείται από πολλά κομμάτια δρόμων, όπου κάθε ένωση μεταξύ των δρόμων θεωρούνται κορυφές. Πλέον, με βάση το δίκτυο της κυκλοφορίας, ορίζει ένα μέτρο, το *trajectory flow*, με το οποίο μετράει την κίνηση

και την μέση ταχύτητα των κινητών για κάθε κομμάτι του δικτύου (διαφορετικούς δρόμους) και τροποποιεί το υπόδειγμα *V-ARIMA* λαμβάνοντάς τα υπόψη για την πρόβλεψη της επόμενης θέσης του κινητού.

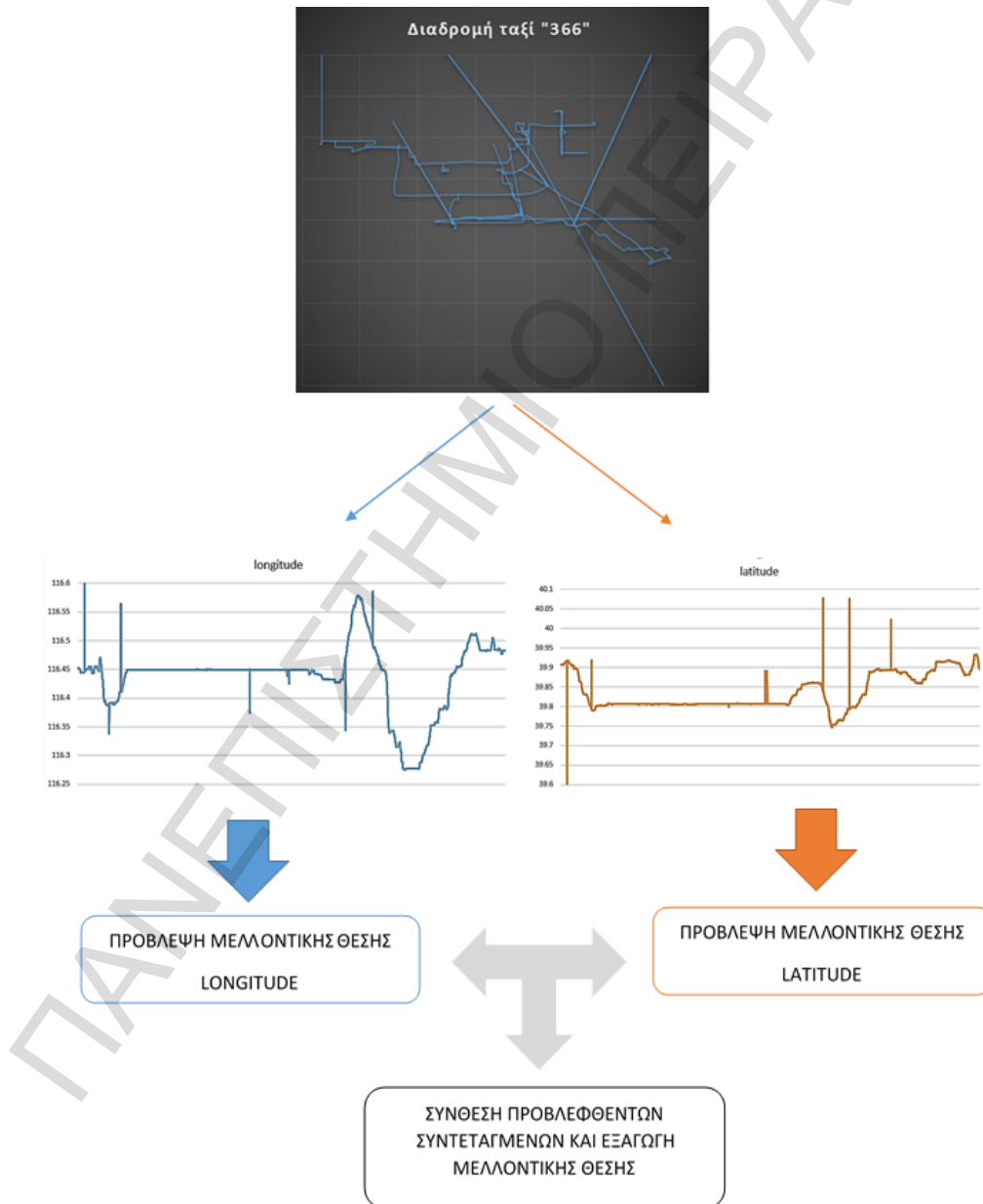
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΚΕΦΑΛΑΙΟ 2

2.1 Γραφική απεικόνιση κεντρικής ιδέας

Σκοπός της συγκεκριμένης διπλωματικής εργασίας, όπως ήδη έχουμε αναφέρει, είναι να προβλέψουμε την μελλοντική θέση ενός κινούμενου αντικειμένου, χρησιμοποιώντας σαν δεδομένα το γεωγραφικό πλάτος και μήκος του αντικειμένου αυτού που καταγράφονται σε σταθερό χρονικό διάστημα. Η πρόβλεψη θα γίνει με τη χρήση μοντέλου χρονοσειρών.



Εικόνα 2.1-1. Σχεδιάγραμμα διπλωματικής

Αναλυτικά χρησιμοποιώντας τα δεδομένα T-Drive trajectory, τα οποία περιέχουν τροχιές ταξί, θα εφαρμόσουμε ξεχωριστά για το γεωγραφικό μήκος και πλάτος μια μέθοδο χρονοσειρών. Συγκεκριμένα θα πάρουμε την τροχιά ενός ταξί όπως φαίνεται και στην (Εικόνα 2.1-1), θα την διασπάσουμε σε γεωγραφικό μήκος και πλάτος και στην συνέχεια θα προσπαθήσουμε να βρούμε το καλύτερο μοντέλο χρονοσειρών για κάθε μια από τις παραπάνω χρονοσειρές ξεχωριστά. Αφού το βρούμε και κάνουμε την πρόβλεψη για το γεωγραφικό μήκος και πλάτος, θα τα συνδυάσουμε ξανά και θα προκύψει η πρόβλεψη για τη μελλοντική θέση. Τέλος θα συγκρίνουμε την προβλεπόμενη με την πραγματική θέση, για να δούμε πόσο ικανοποιητική ήταν η πρόβλεψή μας.

2.2 Αναλυτική περιγραφή

Πριν ξεκινήσουμε λοιπόν την αναλυτική περιγραφή, αλλά και την πρακτική εφαρμογή της μεθόδου που μόλις περιγράψαμε από πάνω, για την καλύτερη κατανόησή της θα παρουσιάσουμε τις βασικές έννοιες αλλά και τους τύπους των μεθόδων που θα χρησιμοποιήσουμε για την πρόβλεψη της μελλοντικής θέσης ενός κινούμενου αντικειμένου.

2.3 Βασικές έννοιες χωροχρονικών δεδομένων

Οι έννοιες οι οποίες είναι απαραίτητες για την ομαλότερη εισαγωγή του αναγνώστη στην συγκεκριμένη διπλωματική εργασία, θα παρουσιαστούν παρακάτω.

Τα δεδομένα λοιπόν που θα χρησιμοποιήσουμε, είναι χωροχρονικά (*spatio-temporal*). Χωροχρονικά δεδομένα είναι αυτά, τα οποία μας δίνουν πληροφορία για τον χώρο και το χρόνο. Εμείς θα ασχοληθούμε με δεδομένα, που προέρχονται από δέκτες GPS, δηλαδή η πληροφορία μας είναι το γεωγραφικό πλάτος και μήκος, δηλαδή οι καρτεσιανές συντεταγμένες, καταγεγραμμένα σε διαδοχικές χρονικές στιγμές.

Τροχιά (*trajectory*) ενός κινούμενου αντικειμένου είναι, η αλληλουχία τοποθεσιών σε καταγεγραμμένες χρονικές στιγμές.

$$T = \langle x_0, y_0, t_0 \rangle, \dots, \langle x_n, y_n, t_n \rangle \quad (2.1)$$

Όπου t_i ($i = 0, \dots, n$) αναπαριστά την χρονική στιγμή που έγινε η καταγραφή και $\forall_{0 < i < n}$, $t_i < t_{i+1}$ και (x_i, y_i) ανήκουν R^2 . Γενικά κάθε σειτ δεδομένων $\langle x_i, y_i, t_i \rangle$ αναπαριστά την τοποθεσία του αντικειμένου (x_i, y_i) την χρονική στιγμή t_i .

2.4 Χρονοσειρές

Μια στοχαστική διαδικασία είναι ένα σύνολο από τυχαίες μεταβλητές $\{Y_t\}$, όπου t παίρνει τιμές από ένα συγκεκριμένο σύνολο C . Στην περίπτωση μας, των χρονοσειρών (*time series*), το σύνολο C αναφέρεται σε συγκεκριμένες χρονικές στιγμές. Για κάθε τιμή t του συνόλου C , ορίζεται μια τυχαία μεταβλητή Y_t και οι τιμές που παίρνουν οι τυχαίες μεταβλητές αυτές, για τις διαφορετικές τιμές του t , δημιουργούν την χρονοσειρά. Δηλαδή μια χρονοσειρά T με τιμές (y_1, \dots, y_T) , είναι ένα δείγμα του διανύσματος των τυχαίων μεταβλητών (Y_1, \dots, Y_T) , που αντιστοιχούν στις χρονικές στιγμές $t = 1, \dots, T$ και η παρατηρηθείσα χρονοσειρά είναι αποτέλεσμα μιας στοχαστικής διαδικασίας.

Οι χρονοσειρές λοιπόν μπορούν να χωριστούν σε στάσιμες και μη στάσιμες. Οι στάσιμες χρονοσειρές είναι αυτές οι οποίες τα χαρακτηριστικά τους δεν αλλάζουν σε συνάρτηση με το χρόνο. Αντίθετα εάν τα χαρακτηριστικά μιας χρονοσειράς αλλάξουν με το χρόνο, τότε έχουμε μη στάσιμη. Συγκεκριμένα, έχουμε αναφέρει ότι κάθε στοχαστική χρονοσειρά y_1, \dots, y_T μπορεί να θεωρηθεί σαν να έχει δημιουργηθεί από ένα σύνολο τυχαίων μεταβλητών που προέρχονται από την ίδια κατανομή. Ομοίως μια μελλοντική παρατήρηση y_{T+1} μπορεί να θεωρηθεί ότι προέρχεται από μια δεσμευμένη συνάρτηση κατανομής $p(y_{T+1} | y_1, \dots, y_T)$, δηλαδή η πιθανότητα της κατανομής για την y_{T+1} δοσμένων των παλιών παρατάσεων y_1, \dots, y_T . Ορίζουμε λοιπόν μια στάσιμη χρονοσειρά, εκείνη στην οποία

$$p(y_t, \dots, y_{t+k}) = p(y_{t+m}, \dots, y_{t+k+m}) \quad (2.2)$$

$$p(y_t) = p(y_{t+m}) \quad (2.3)$$

για κάθε t, m .

Επίσης για κάθε χρονοσειρά που είναι στάσιμη, ο μέσος της ορίζεται ως

$$\mu_y = E(y_t) \quad (2.4)$$

και ισχύει $E(y_t) = E(y_{t+m})$ για κάθε t, m . Ακόμα η διακύμανση της χρονοσειράς είναι

$$\sigma_y^2 = E[(y_t - \mu_y)^2] \quad (2.5)$$

, αλλά επειδή είναι στάσιμη ισχύει ότι $E[(y_t - \mu_y)^2] = E[(y_{t+m} - \mu_y)^2]$ για κάθε t, m και τέλος για κάθε υστέρηση (lag) k η συνδιακύμανση της χρονοσειράς είναι,

$$\gamma_k = Cov(y_t, y_{t+k}) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)] \quad (2.6)$$

και επειδή όπως έχουμε πει είναι στάσιμη, ισχύει ότι $Cov(y_t, y_{t+k}) = Cov(y_{t+m}, y_{t+m+k})^3$.

Εάν μια χρονοσειρά είναι στάσιμη, η συνάρτηση πιθανότητας $p(y_t)$ είναι ίδια για κάθε χρονική στιγμή t και η μορφή της (τις περισσότερες φορές) μπορεί να αναγνωρισθεί με ένα απλό ιστόγραμμα των παρατηρήσεων y_1, \dots, y_T . Τέλος η εκτίμηση του μ_y μπορεί να γίνει όπως γίνεται με τον απλό μέσο, δηλαδή

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t \quad (2.7)$$

, αλλά και η διακύμανση υπολογίζεται όπως την απλή διακύμανση, δηλαδή ως

$$\hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2 \quad (2.8)$$

Καθώς συνήθως είναι αδύνατον να εκτιμήσουμε την κατανομή μιας χρονοσειράς, η συνάρτηση αυτοσυσχέτισης (*autocorrelation function*) θα φανεί πολύ χρήσιμη γιατί δίνει μια μερική περιγραφή της χρονοσειράς για σκοπούς μοντελοποίησης. Η συνάρτηση αυτοσυσχέτισης μας λέει πόσο συσχέτιση υπάρχει ανάμεσα στα γειτονικά δεδομένα μιας χρονοσειράς y_t . Ορίζουμε σαν αυτοσυσχέτιση με υστέρηση k ως

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sqrt{E[(y_t - \mu_y)^2]E[(y_{t+k} - \mu_y)^2]}} = \frac{Cov(y_t, y_{t+k})}{\sigma_{y_t} \sigma_{y_{t+k}}} \quad (2.9)$$

Για μια στάσιμη χρονοσειρά η διακύμανση στον χρόνο t είναι ίδια με στον χρόνο $t + k$, οπότε η σχέση (2.9) γίνεται

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sigma_y^2} \quad (2.10)$$

Παρατηρούμε ότι ο αριθμητής της σχέσης (2.10) είναι η συνδιακύμανση μεταξύ y_t και y_{t+k} και είναι γ_k οπότε

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.11)$$

και συνεπώς

$$\rho_0 = 1 \quad (2.12)$$

Ας υποθέσουμε τώρα μια απλή χρονοσειρά

$$y_t = \varepsilon_t \quad (2.13)$$

όπου ε_t είναι ανεξάρτητες κατανομημένες τυχαίες μεταβλητές με μέσο μηδέν. Τότε εύκολα από την σχέση (2.10) μπορούμε να δούμε ότι η συνάρτηση αυτοσυσχέτισης για αυτή την διαδικασία δίνεται από $\rho_0 = 1$ και $\rho_k = 0$ για κάθε $k > 0$. Η διαδικασία της σχέσης (2.13) ονομάζεται λευκός θόρυβος (*white noise*) και δεν υπάρχει καλύτερο μοντέλο που μπορεί να κάνει πρόβλεψη από το $\hat{y}_{T+l} = 0$ για κάθε l . Συνεπώς εάν η συνάρτηση της αυτοσυσχέτισης είναι μηδέν ή κοντά στο μηδέν, δεν υπάρχει τιμή ή είναι πολύ μικρή που να περιγράφει την πρόβλεψη της χρονοσειράς αυτής (Pindyck, R. S. et al., 1998).

Στην πραγματικότητα είναι πού λίγες οι φορές που έχουμε μια στάσιμη χρονοσειρά. Ευτυχώς όμως, αν και τις περισσότερες φορές δεν είναι στάσιμες, κάνοντας τις πρώτες διαφορές ή περισσότερες, οδηγούμαστε σε μια στάσιμη χρονοσειρά. Τέτοιες μη στάσιμες χρονοσειρές ονομάζονται ομοιογενείς. Ο αριθμός που μια χρονοσειρά πρέπει να αφαιρεθεί πριν γίνει στάσιμη ονομάζεται τάξη ομοιογένειας (*order of homogeneity*). Συνεπώς αν y_t είναι πρώτης τάξης ομοιογενείς, η χρονοσειρά

$$w_t = y_t - y_{t-1} = \Delta y_t \quad (2.14)$$

είναι στάσιμη. Εάν y_t τυχαίνει να είναι δεύτερης τάξης ομοιογενείς μη στάσιμης διαδικασίας, η χρονοσειρά

$$w_t = \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \quad (2.15)$$

είναι στάσιμη.

Στην συνέχεια θα αναφέρουμε κάποιους βασικούς κανόνες της συνάρτησης αυτοσυσχέτισης για μια στάσιμη στοχαστική διαδικασία. Έστω y_t μια στάσιμη διαδικασία και L_t μια οποιαδήποτε γραμμική συνάρτηση της y_t και για κάθε υστέρηση (*lag*) στην y_t , για παράδειγμα,

$$L_t = a_1 y_t + a_2 y_{t-1} + \dots + a_k y_{t-k+1} \quad (2.16)$$

Τώρα, εφόσον η y_t είναι στάσιμη, οι συνδιακυμάνσεις (*covariances*) της y_t είναι στάσιμες, και

$$\text{Cov}(y_{t+i}, y_{t+j}) = \gamma_{|i-j|} \quad (2.17)$$

ανεξαρτήτως του t . Τότε, τετραγωνίζοντας και τα δύο μέλη της εξίσωσης (2.16), βλέπουμε ότι η διακύμανση της L_t δίνεται από τον τύπο

$$\text{Var}(L_t) = \sum_{i=1}^{\kappa+1} \sum_{j=1}^{\kappa+1} a_i a_j \gamma_{|i-j|} \quad (2.18)$$

Εάν τα a δεν είναι μηδέν, η διακύμανση του L_t πρέπει να είναι μεγαλύτερη του μηδέν και συνεπώς πρέπει να έχουμε, για κάθε i και j

$$\gamma_{|i-j|} > 0, \text{ για κάθε } i = j \quad (2.19)$$

Τώρα, για n παρατηρήσεις, γράφουμε τις συνδιακυμάνσεις του y_t σαν πίνακα:

$$\Gamma_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \dots & \gamma_0 \end{bmatrix} \quad (2.20)$$

Αυτός ο πίνακας πρέπει να είναι θετικά ορισμένος γιατί η διακύμανση του L_t είναι πάντα μεγαλύτερη από το μηδέν. Παρατηρούμε ότι

$$\Gamma_n = \sigma_y^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{bmatrix} = \sigma_y^2 P_n \quad (2.21)$$

όπου P_n είναι ο πίνακας των αυτοσυσχετίσεων και είναι θετικά ορισμένος. Συνεπώς, ο προσδιοριστικός P_n και τα κύρια παράγωγα πρέπει να είναι μεγαλύτερα από το μηδέν.

Αν για παράδειγμα, υπολογίσουμε την περίπτωση για $n=2$. Η συνθήκη για τον προσδιοριστικό πίνακα P_n γίνεται

$$\det \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} > 0 \quad (2.22)$$

Συνεπάγεται

$$1 - \rho_1^2 > 0 \quad (2.23)$$

Ομοίως, για $n=3$, μπορεί εύκολα να δούμε ότι οι τρεις συνθήκες πρέπει να είναι

$$-1 < \rho_1 < 1 \quad (2.24)$$

$$-1 < \rho_2 < 1 \quad (2.25)$$

ούτω καθεξής. Ονομάζουμε τις σειρές $\alpha_1, \alpha_2, \alpha_3, \dots$ μερική αυτοσυνδιακύμανση και μπορούμε να βασιστούμε για την εύρεση της τάξης της αυτοπαλίνδρομης διαδικασίας. Συγκεκριμένα αν η πραγματική τάξη της διαδικασίας είναι p , πρέπει να έχουμε $\alpha_j \approx 0$ για $j > 0$.

Για να εξετάσουμε πότε συγκεκριμένα το α_j είναι μηδέν, μπορούμε να χρησιμοποιήσουμε το γεγονός ότι ακολουθεί περίπου κανονική κατανομή με μέσο μηδέν και διακύμανση $1/T$. Συνεπώς μπορούμε να ελέγξουμε πότε είναι στατιστικά σημαντική, για παραδείγματος χάρη σε επίπεδο σημαντικότητάς 5%, κοιτώντας πότε ξεπερνάει $2/\sqrt{T}$ σε έκταση.

Στην πραγματικότητα τις περισσότερες φορές δουλεύουμε με μη στάσιμες χρονοσειρές, δηλαδή ότι τα χαρακτηριστικά τους αλλάζουν με την πάροδο του χρόνου. Γι' αυτό κατασκευάζουμε μοντέλα για μη στάσιμες χρονοσειρές οι οποίες μπορούν να μετατρέπον σε στάσιμες παίρνοντας τις πρώτες ή περισσότερες διαφορές. Μπορούμε να πούμε ότι y_t είναι μια ομογενής μη στάσιμη τάξης d εάν

$$w_t = \Delta^d y_t \quad (2.32)$$

είναι μια στάσιμη χρονοσειρά. Εδώ Δ ορίζεται οι διαφορές, παραδείγματος χάρη,

$$\Delta y_t = y_t - y_{t-1} \text{ και } \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \quad (2.33)$$

και ούτω καθεξής.

Παρατηρούμε ότι εάν έχουμε μια σειρά w_t , μπορούμε να πάμε πίσω στην y_t αθροίζοντας την w_t d φορές. Οπότε γράφεται ως

$$y_t = \sum_d w_t \quad (2.34)$$

Όπου Σ είναι το σύμβολο του αθροίσματος:

$$\sum w_t = \sum_{i=-\infty}^t w_i \quad (2.35)$$

$$\sum_2 w_t = \sum_{j=-\infty}^t \sum_{i=-\infty}^j w_i \quad (2.36)$$

και ούτω καθεξής.

Σε αυτό το σημείο να σημειώσουμε ότι το άθροισμα Σ είναι η αντίστροφη διαδικασία των διαφορών Δ . Οπότε $\Delta y_t = y_t - y_{t-1}$, μπορεί να γραφεί σαν $\Delta = 1 - B$, και $\Sigma = \Delta^{-1} = (1 - B)^{-1}$.

Στον υπολογισμό του παραπάνω αθροίσματος μιας πραγματικής χρονοσειράς, ξεκινάμε με την πρώτη παρατήρηση της κανονικής σειράς (*original undifferenced*) (y_0) και μετά προσθέτουμε τις διαδοχικές τιμές της αφαιρεμένης χρονοσειράς (*differenced*). Συνεπώς, εάν $w_t = \Delta y_t$, μπορούμε να υπολογίσουμε την y_t από τον παρακάτω τύπο,

$$y_t = \sum w_t = \sum_{i=-\infty}^t w_i = \sum_{i=-\infty}^0 w_i + \sum_{i=0}^t w_i = y_0 + w_1 + w_2 + \dots + w_t \quad (2.37)$$

Εάν η y_t είχε αφαιρεθεί δύο φορές, θα είχαμε $w_t = \Delta^2 y_t$, μπορούμε να υπολογίσουμε την y_t από την w_t , προσθέτοντας την w_t δύο φορές.

Μετά, αφού πλέον έχουμε την αφαιρεμένη χρονοσειρά της y_t ώστε να παράξουμε την στάσιμη χρονοσειρά w_t , μπορούμε να μοντελοποιήσουμε την w_t σαν *ARMA* διαδικασία. Εάν $y_t = \Delta^d y_t$ και w_t είναι *ARMA*(p, q) διαδικασία, τότε λέμε ότι η y_t είναι μια ενσωματωμένη αυτοπαλίνδρομη –κινητού μέσου διαδικασία τάξης (p, d, q), ή απλά *ARIMA*(p, d, q). Μπορούμε να γράψουμε την εξίσωση για την διαδικασία *ARIMA*(p, d, q),

$$\varphi(B)\Delta^d y_t = \delta + \theta(B)\varepsilon_t \quad (2.38)$$

με

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \quad (2.39)$$

και

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p \quad (2.40)$$

Ονομάζουμε το $\varphi(B)$ τον αυτοπαλίνδρομο παράγοντα και $\theta(B)$ τον παράγοντα κινητού μέσου.

Παρατηρούμε ότι ο μέσος της $w_t = \Delta^d y_t$ δίνεται από τον τύπο

$$\mu_w = \frac{\delta}{1 - \varphi_1 - \varphi_2 - \dots - \varphi_p} \quad (2.41)$$

Συνεπώς, εάν το δ δεν ισούται με μηδέν, τότε η ενσωματωμένη (*integrated*) χρονοσειρά y_t θα έχει μια μη τυχαία (*deterministic*) τάση.

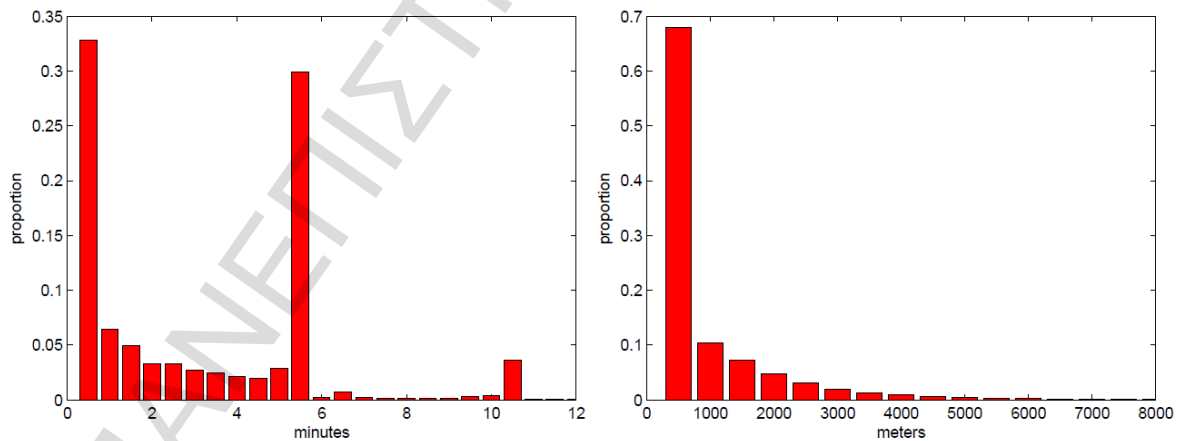
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΚΕΦΑΛΑΙΟ 3

Σε αυτό το κεφάλαιο θα παρουσιάσουμε τα δεδομένα τα οποία θα χρησιμοποιήσουμε για το πρακτικό μέρος της διπλωματικής αυτής, αλλά και την αναλυτική καθαυτού πρακτική εφαρμογή της μεθόδου που επιλέξαμε, δηλαδή ένα μοντέλο χρονοσειρών, για την πρόβλεψη της μελλοντικής θέσης ενός κινούμενου αντικειμένου.

3.1 Γενικές πληροφορίες των δεδομένων

Το σετ δεδομένων που θα χρησιμοποιήσουμε περιέχουν *GPS* τροχιές από 10.357 ταξί της πόλης του Πεκίνου κατά την χρονική διάρκεια 2 Φεβρουαρίου με 8 Φεβρουαρίου 2008. Ο συνολικός αριθμός των καταγραφών είναι περίπου 15 εκατομμύρια και η συνολική απόσταση των τροχιών φτάνει τα 9 εκατομμύρια χιλιόμετρα. Στην (εικόνα 3.1-1) βλέπουμε την κατανομή του χρονικού αλλά και γεωγραφικού διαστήματος μεταξύ δύο σημείων. Η μέση χρονική διαφορά είναι 177 δευτερόλεπτα και η μέση απόσταση είναι 623 μέτρα. Κάθε αρχείο από το σετ δεδομένων, που έχει την ονομασία του από το *ID* του κάθε ταξί, περιέχει μόνο την τροχιά ενός ταξί.



Εικόνα 3.1-1. Ιστογράμματα της χρονικής και γεωγραφικής απόστασης μεταξύ δύο σημείων (Yuan, J et al. (2011))

Τέλος παρουσιάζουμε ένα δείγμα των αρχείων

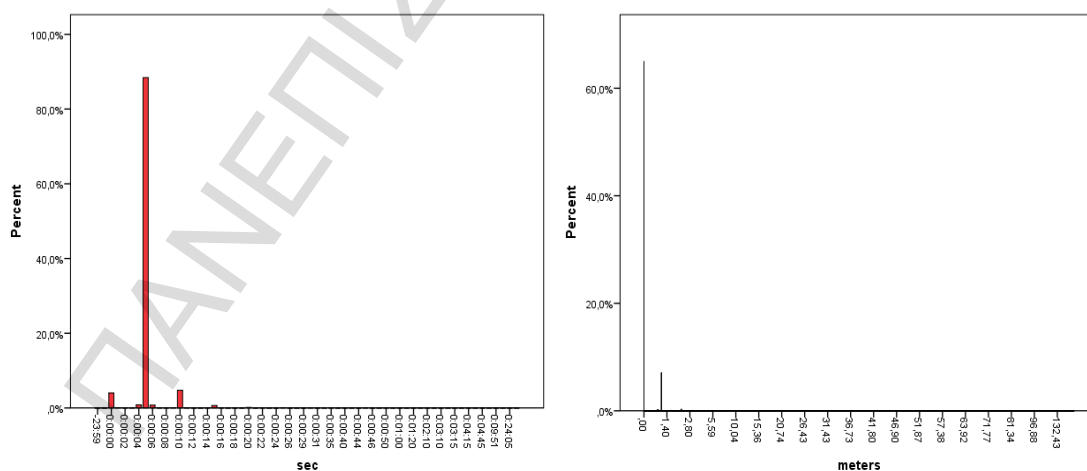
366,	2008-02-02	13:30:48,	116.45353,	39.90732
366,	2008-02-02	13:30:53,	116.45348,	39.90729
366,	2008-02-02	13:30:58,	116.45334,	39.90725
366,	2008-02-02	13:31:03,	116.4533,	39.90722
366,	2008-02-02	13:31:08,	116.45327,	39.90722
366,	2008-02-02	13:31:13,	116.4532,	39.90725
366,	2008-02-02	13:31:18,	116.45309,	39.9076
366,	2008-02-02	13:31:23,	116.453,	39.9077
366,	2008-02-02	13:31:28,	116.45281,	39.9076
366,	2008-02-02	13:31:33,	116.45271,	39.90767

Όπου κάθε γραμμή έχει τα ακόλουθα στοιχεία, τα οποία χωρίζονται με κόμμα:

taxi id, date time, longitude, latitude

3.2 Παρουσίαση των δεδομένων μας

Τα δεδομένα που επιλέξαμε για να εφαρμόσουμε την μέθοδό μας για την πρόβλεψη της μελλοντικής θέσης, προέρχονται από το παραπάνω σετ δεδομένων που αναλύσαμε και συγκεκριμένα από το σετ δεδομένων² του ταξί με id 366. Στην (εικόνα 3.2-1) βλέπουμε την κατανομή του χρονικού αλλά του γεωγραφικού διαστήματος μεταξύ δύο συνεχόμενων σημείων. Η μέση



Εικόνα 3.2-1. Ιστογράμματα της χρονικής και γεωγραφικής απόστασης μεταξύ δύο σημείων για το ταξί με id 366

² Το σετ δεδομένων μας βρίσκεται στο Παράρτημα 1

χρονική διαφορά είναι 5 δευτερόλεπτα και η μέση απόσταση είναι 38,95 μέτρα, ανάμεσα σε δύο σημεία.

Σε αυτό το σημείο είναι καλό να αναφέρουμε ότι ο τύπος³ που χρησιμοποιήσαμε για να βρούμε την απόσταση μεταξύ δύο σημείων είναι ο παρακάτω:

$$\Delta\hat{\sigma} = 2\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_s\cos\varphi_f\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \quad (3.1)$$

3.3 Εύρεση κατάλληλων μοντέλων χρονοσειρών

Σε αυτή την ενότητα πλέον, θα επιλέξουμε και θα εφαρμόσουμε το κατάλληλο μοντέλο των χρονοσειρών. Συγκεκριμένα, με την βοήθεια του στατιστικού προγράμματος SPSS, πρώτα θα ελέγξουμε και θα υπολογίσουμε, ξεχωριστά την χρονοσειρά του γεωγραφικού πλάτους και του γεωγραφικού μήκους των πρώτων 100 εγγραφών.

Θα ξεκινήσουμε λοιπόν ελέγχοντας αν οι χρονοσειρές μας δεν είναι λευκός θόρυβος και αν είναι στάσιμες.

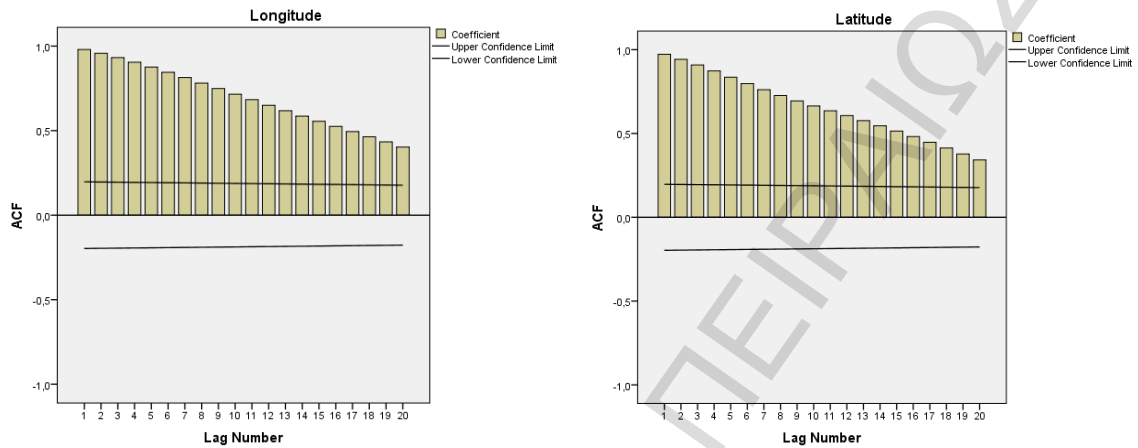
Autocorrelations						Autocorrelations					
Series: Longitude						Series: Latitude					
Lag	Autocorrelation	Std. Error ^a	Box-Ljung Statistic			Lag	Autocorrelation	Std. Error ^a	Box-Ljung Statistic		
			Value	df	Sig. ^b				Value	df	Sig. ^b
1	,980	,099	98,962	1	,000	1	,973	,099	97,468	1	,000
2	,957	,098	194,298	2	,000	2	,942	,098	189,859	2	,000
3	,932	,098	285,666	3	,000	3	,909	,098	276,682	3	,000
4	,905	,097	372,663	4	,000	4	,873	,097	357,656	4	,000
5	,876	,097	454,995	5	,000	5	,835	,097	432,582	5	,000
6	,845	,096	532,505	6	,000	6	,797	,096	501,525	6	,000
7	,814	,095	605,121	7	,000	7	,761	,095	565,027	7	,000
8	,781	,095	672,801	8	,000	8	,727	,095	623,549	8	,000
9	,749	,094	735,662	9	,000	9	,694	,094	677,545	9	,000
10	,716	,094	793,774	10	,000	10	,664	,094	727,508	10	,000
11	,683	,093	847,227	11	,000	11	,635	,093	773,783	11	,000
12	,650	,093	896,220	12	,000	12	,606	,093	816,414	12	,000
13	,618	,092	940,928	13	,000	13	,577	,092	855,387	13	,000
14	,586	,092	981,654	14	,000	14	,546	,092	890,713	14	,000
15	,555	,091	1018,664	15	,000	15	,514	,091	922,410	15	,000
16	,525	,091	1052,135	16	,000	16	,481	,091	950,518	16	,000
17	,494	,090	1082,174	17	,000	17	,447	,090	975,112	17	,000
18	,464	,090	1108,939	18	,000	18	,413	,090	996,313	18	,000
19	,434	,089	1132,609	19	,000	19	,377	,089	1014,239	19	,000
20	,403	,089	1153,327	20	,000	20	,342	,089	1029,135	20	,000

a. The underlying process assumed is independence (white noise).
b. Based on the asymptotic chi-square approximation.

Πίνακας 3.3-1. Ο πίνακας των αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος

³Latitude, Longitude, And Great Circles: <http://www.cpearson.com/excel/LatLong.aspx>

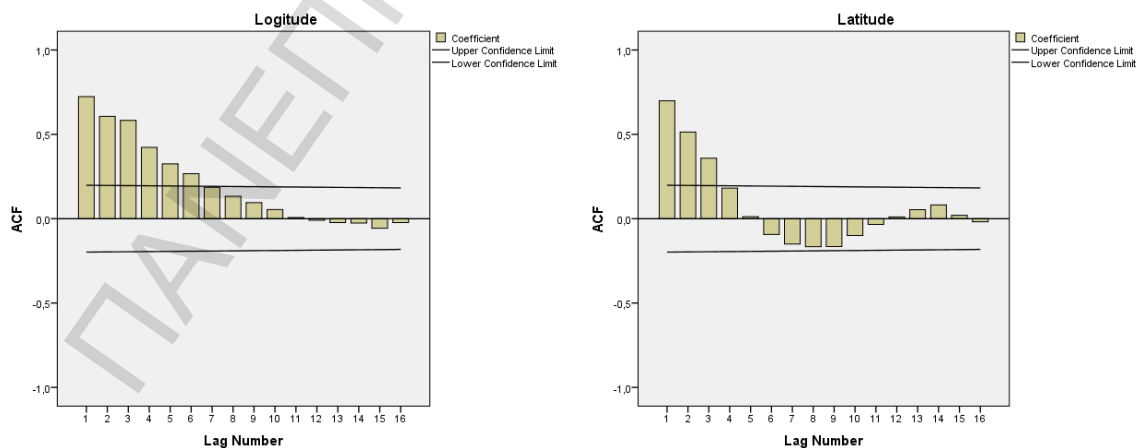
Στον πίνακα των αυτοσυσχετίσεων (εικόνα 3.3-1) το SPSS δίνει τις τιμές της στατιστικής Q των Box και Ljung με τα αντίστοιχα P-Values για τον έλεγχο της μηδενικής υπόθεσης ότι η χρονοσειρά είναι λευκός θόρυβος, όπου βλέπουμε ότι απορρίπτουμε την αρχική μας υπόθεση και ότι άρα τα δεδομένα μας δεν είναι λευκός θόρυβος, δηλαδή συσχετίζονται μεταξύ τους.



Εικόνα 3.3-1. Διαγράμματα αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος

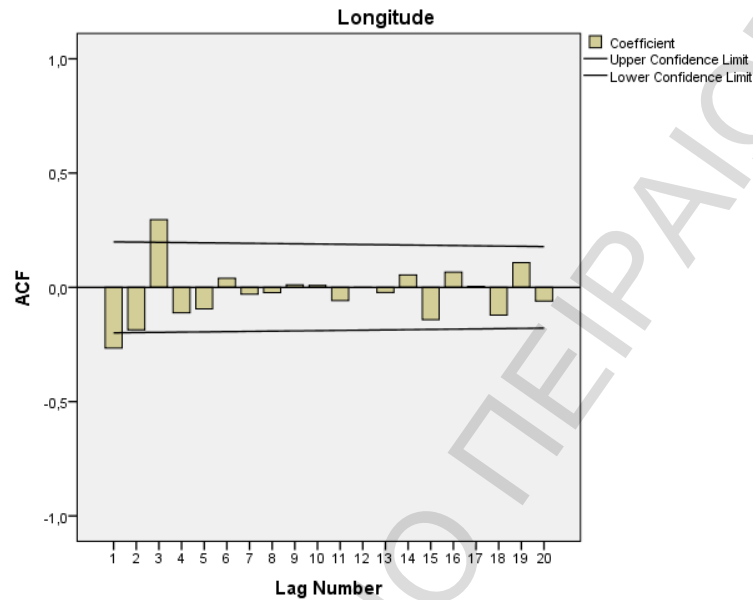
Παρατηρούμε ότι οι αυτοσυσχετίσεις της χρονοσειράς φθίνουν με αργό ρυθμό και δεν συγκλίνουν στο μηδέν οπότε μπορεί να θεωρηθεί μη στάσιμη.

Στην συνέχεια θα υπολογίσουμε τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις των χρονοσειρών μας για τις πρώτες διαφορές.



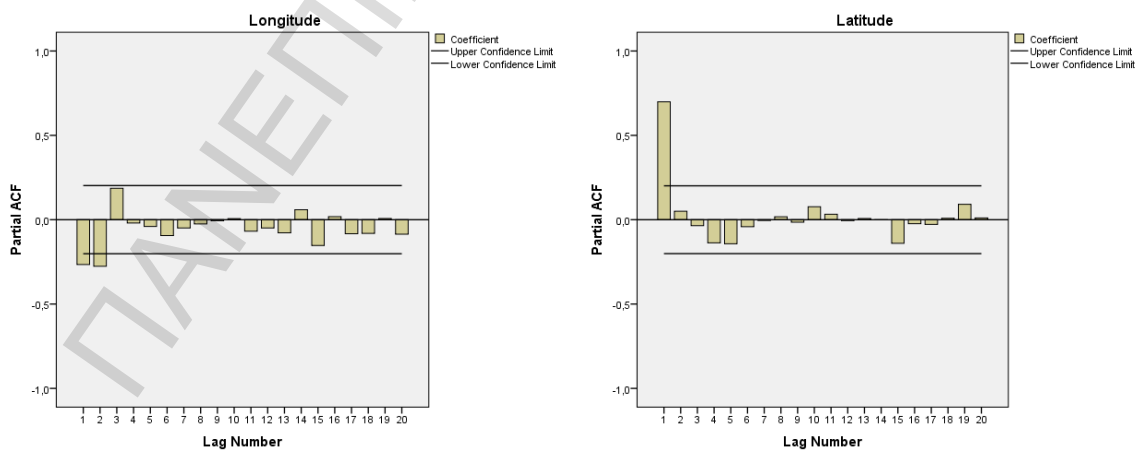
Εικόνα 3.3-2. Διαγράμματα αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος και δεξιά για το γεωγραφικό μήκος σε πρώτες διαφορές

Παρατηρούμε στην (εικόνα 3.3-2) ότι οι αυτοσυσχετίσεις της χρονοσειράς *Longitude* σε πρώτες διαφορές δεν φθίνουν γρήγορα οπότε δεν είναι στάσιμη και χρειαζόμαστε να την ελέγξουμε σε δεύτερες διαφορές. Αντίθετα, οι αυτοσυσχετίσεις της χρονοσειράς *Latitude* φθίνουν γρήγορα και μηδενίζονται, οπότε μπορούμε να θεωρήσουμε ότι η χρονοσειρά είναι στάσιμη.



Εικόνα 3.3-3. Διάγραμμα αυτοσυσχετίσεων για το γεωγραφικό πλάτος σε δεύτερες διαφορές

Πλέον μπορούμε να θεωρήσουμε, βλέπε (εικόνα 3.3-3), ότι και η χρονοσειρά *Longitude* είναι στάσιμη γιατί οι αυτοσυσχετίσεις φθίνουν γρήγορα και μηδενίζονται.



Εικόνα 3.3-4. Διαγράμματα μερικών αυτοσυσχετίσεων αριστερά για το γεωγραφικό πλάτος με δεύτερες διαφορές και δεξιά για το γεωγραφικό μήκος σε πρώτες διαφορές

Τέλος για την χρονοσειρά *Longitude* παρατηρούμε δύο μη μηδενικές μερικές αυτοσυσχετίσεις (εικόνα 3.3-4) και άρα το μοντέλο βάση για την περαιτέρω εξέταση θα είναι το *ARIMA(2,2,0)*. Στην περίπτωση της χρονοσειράς *Latitude* παρατηρούμε μια μη μηδενική μερική αυτοσυσχέτιση και άρα το μοντέλο *ARIMA(1,1,0)* θα χρησιμοποιηθεί σαν μοντέλο βάσης για την συνέχεια.

3.3.1 Υπόδειγμα για την χρονοσειρά *Longitude*

Στην συνέχεια βλέπουμε την εκτίμηση του υποδείγματος που βρήκαμε πιο πάνω, πιο συγκεκριμένα:

Model Description		
Model ID	Longitude	Model Type
Model_1		ARIMA(2,2,0)

Πίνακας 3.3-2. Τύπος του υποδείγματος που εκτιμήθηκε *ARIMA(2,2,0)* για την χρονοσειρά *Longitude*

Model Statistics						
Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Longitude-Model_1	0	-17,988	10,481	16	,840	0

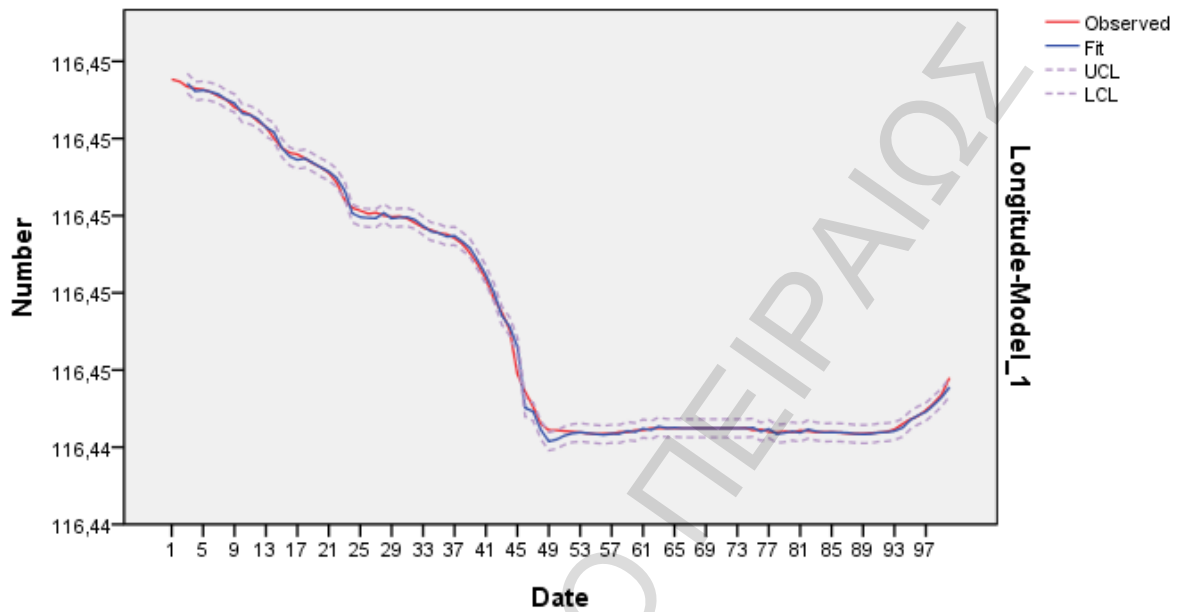
Πίνακας 3.3-3. Αποτελέσματα ελέγχων για *ARIMA(2,2,0)* για την χρονοσειρά *Longitude*

Στον (πίνακα 3.3-3) βλέπουμε την τιμή του κριτηρίου *Normalized BIC* και την τιμή της στατιστικής *Q* των *Box* και *Ljung* για 17 χρονικές υστερήσεις με το *P-Value* για τον έλεγχο της μηδενικής υπόθεσης ότι η χρονοσειρά των σφαλμάτων είναι λευκός θόρυβος.

ARIMA Model Parameters					Estimate	SE	t	Sig.
Longitude-Model_1	Longitude	No Transformation	AR	Lag 1	-,350	,100	-3,506	,001
				Lag 2	-,285	,100	-2,853	,005
				Difference		2		

Πίνακας 3.3-4. Οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος για την χρονοσειρά *Longitude*

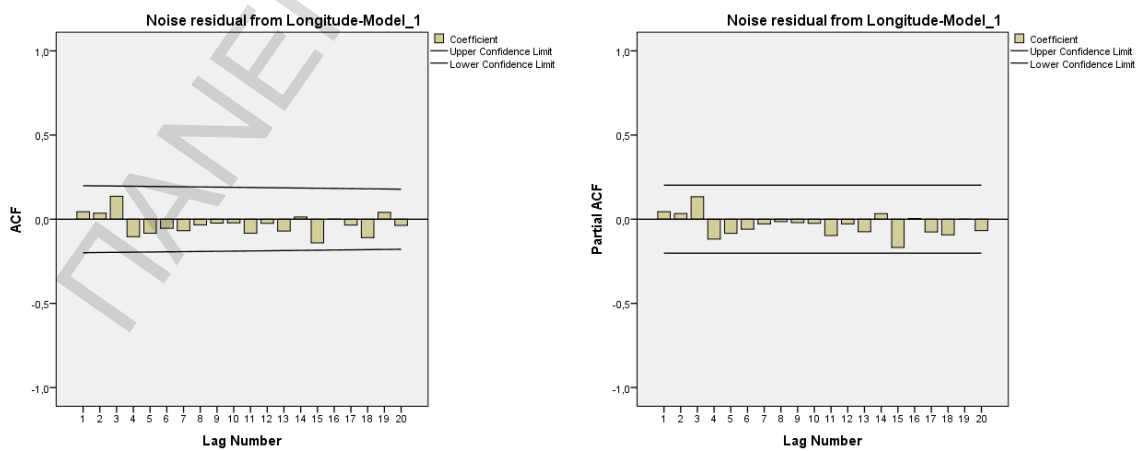
Αντίστοιχα στον (πίνακα 3.3-4) βλέπουμε για τους δύο όρους του αυτοπαλίνδρομου μοντέλου, τις τιμές της t στατιστικής για τον έλεγχο της στατιστικής σημαντικότητας και το P -Value του ελέγχου. Στην περίπτωση μας είναι στατιστικά σημαντικοί και οι δύο όροι.



Εικόνα 3.3-5. Γραφική απεικόνιση του υποδείγματος

Στην (εικόνα 3.3-5) απεικονίζονται οι τιμές της χρονοσειράς του *Longitude* μαζί με τις εκτιμήσεις τους και τα κάτω όρια του διαστήματος εμπιστοσύνης των εκτιμήσεων.

Συνεχίζοντας με τον έλεγχο του υποδείγματος θα ελέγξουμε τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις της χρονοσειράς των σφαλμάτων.



Εικόνα 3.3-6. Διαγράμματα αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων

Από τα διαγράμματα των αυτοσυσχετίσεων και των μερικών αυτοσυσχετίσεων των σφαλμάτων (εικόνα 3.3-6) διαπιστώνουμε ότι δεν υπάρχουν συσχετίσεις που να υπερβαίνουν τα όρια των διαστημάτων εμπιστοσύνης.

Autocorrelations

Series: Noise residual from Longitude-Model_1

Lag	Autocorrelation	Std. Error ^a	Box-Ljung Statistic		
			Value	df	Sig. ^b
1	,045	,099	,202	1	,653
2	,036	,099	,332	2	,847
3	,136	,098	2,239	3	,524
4	-,104	,098	3,357	4	,500
5	-,083	,097	4,082	5	,538
6	-,054	,097	4,392	6	,624
7	-,068	,096	4,889	7	,673
8	-,033	,096	5,010	8	,757
9	-,023	,095	5,068	9	,828
10	-,022	,095	5,120	10	,883
11	-,083	,094	5,896	11	,880
12	-,025	,094	5,964	12	,918
13	-,070	,093	6,531	13	,925
14	,013	,093	6,550	14	,951
15	-,141	,092	8,882	15	,884
16	,000	,091	8,882	16	,918
17	-,035	,091	9,027	17	,939
18	-,109	,090	10,481	18	,915
19	,041	,090	10,686	19	,934
20	-,037	,089	10,856	20	,950

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Πίνακας 3.3-5. Ο πίνακας των αυτοσυσχετίσεων της χρονοσειράς των σφαλμάτων

Επίσης ελέγχουμε τα σφάλματα να είναι λευκός θόρυβος κάτι που γίνεται με την στατιστική Q των *Box* και *Ljung*. Παρατηρούμε από τον (πίνακα 3.3-5) ότι πράγματι τα σφάλματα είναι λευκός θόρυβος.

Επιπρόσθετα για να είμαστε ακόμα πιο πολύ σίγουροι για την ορθότητα του υποδείγματος που επιλέξαμε θα υπολογίσουμε μερικά διευρυμένα υποδείγματα με σκοπό να βρούμε ότι οι συντελεστές τους δεν είναι στατιστικά σημαντικοί. Γι' αυτό τον λόγο τα διευρυμένα μοντέλα που θα εξετάσουμε είναι τα $ARIMA(3,2,0)$ και $ARIMA(2,2,1)$.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Longitude-Model_1	Longitude	No Transformation	AR	Lag 1	-,295	,103	-2,854	,005
				Lag 2	-,218	,105	-2,068	,041
				Lag 3	,184	,103	1,780	,078
				Difference	2			

Πίνακας 3.3-6. Οι τιμές των παραμέτρων του $ARIMA(3,2,0)$ υποδείγματος για την χρονοσειρά *Longitude*

Παρατηρούμε από τον (πίνακα 3.3-6) ότι ο επιπρόσθετος όρος που βάλαμε δεν είναι στατιστικά σημαντικός κάτι που αποδεικνύει την ορθότητα της επιλογής του μοντέλου μας.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Longitude-Model_1	Longitude	No Transformation	AR	Lag 1	-,754	,239	-3,153	,002
				Lag 2	-,405	,098	-4,111	,000
				Difference	2			
				MA	Lag 1	-,452	,258	-1,751

Πίνακας 3.3-7. Οι τιμές των παραμέτρων του $ARIMA(2,2,1)$ υποδείγματος για την χρονοσειρά *Longitude*

Ομοίως και στο διευρυμένο μοντέλο $ARIMA(2,2,1)$ βλέπουμε ότι ο επιπλέον όρος που προσθέσαμε δεν είναι στατιστικά σημαντικός.

Άρα το υπόδειγμα που θα χρησιμοποιήσουμε για την εκτίμηση της μελλοντικής τιμής για την χρονοσειρά *Longitude* είναι το $ARIMA(2,2,0)$.

3.3.2 Υπόδειγμα για την χρονοσειρά *Latitude*

Αφού βρήκαμε το υπόδειγμα για την χρονοσειρά *Longitude*, θα συνεχίσουμε για την εύρεση του υποδείγματος για την χρονοσειρά *Latitude* με ακριβώς τον ίδιο τρόπο.

Model Description

			Model Type
Model ID	Latitude	Model_1	ARIMA(1,1,0)

Πίνακας 3.3-8. Τύπος του υποδείγματος που εκτιμήθηκε ARIMA(1,1,0) για την χρονοσειρά Latitude

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Latitude-Model_1	0	-17,281	8,970	17	,941	0

Πίνακας 3.3-9. Αποτελέσματα ελέγχων για ARIMA(1,1,0) για την χρονοσειρά Latitude

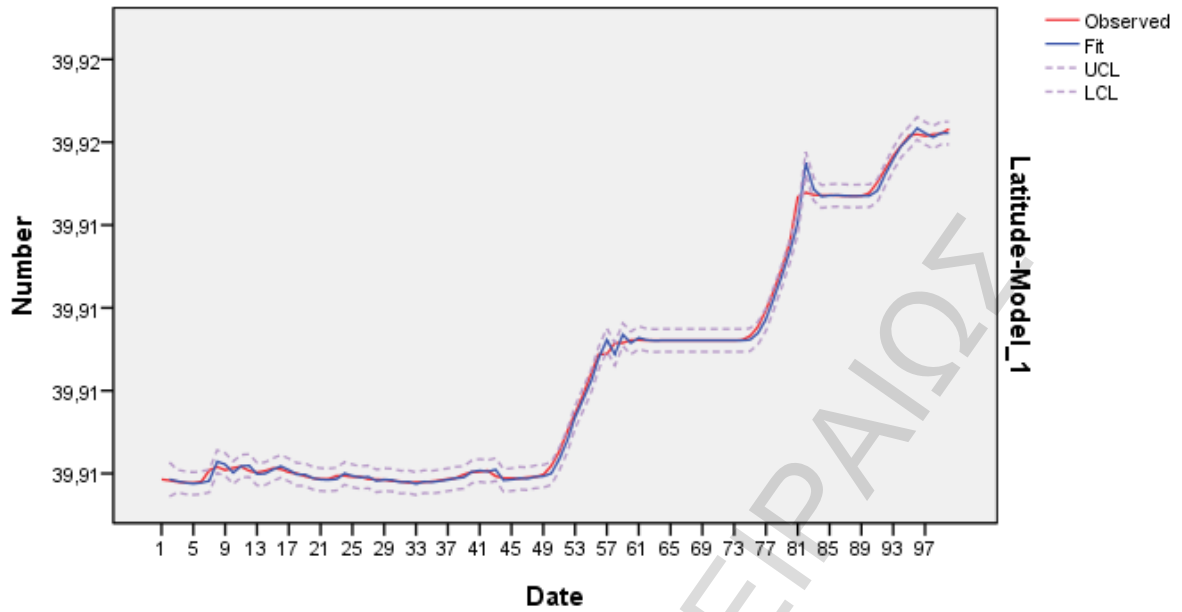
Στον (πίνακα 3.3-9) βλέπουμε την τιμή του κριτηρίου *Normalized BIC* και την τιμή της στατιστικής *Q* των *Box* και *Ljung* για 17 χρονικές υστερήσεις με το *P-Value* για τον έλεγχο της μηδενικής υπόθεσης ότι η χρονοσειρά των σφαλμάτων είναι λευκός θόρυβος.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Latitude-Model_1	Latitude	No Transformation	AR Lag 1	,745	,067	11,140	,000
			Difference	1			

Πίνακας 3.3-10. Οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος για την χρονοσειρά Latitude

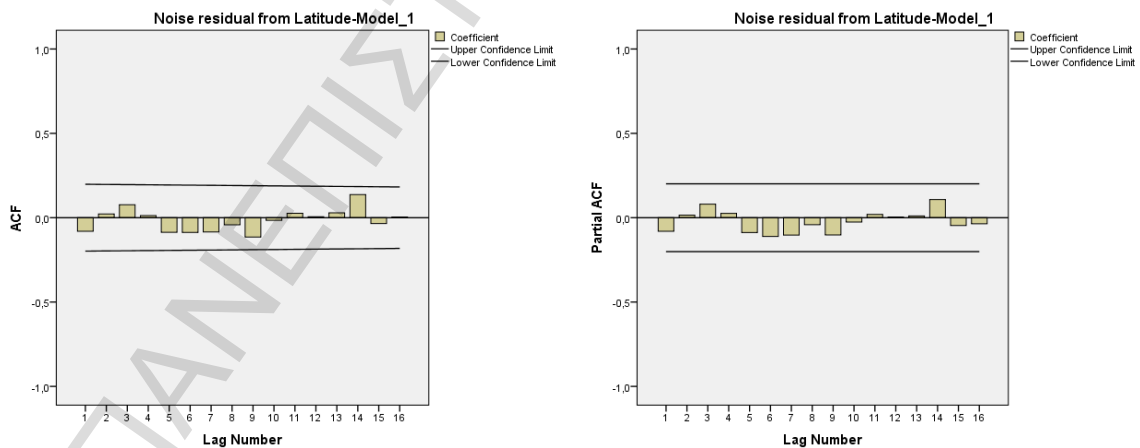
Στο ARIMA Model Parameters περιλαμβάνονται οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος, δηλαδή στην περίπτωση που εξετάζεται η τιμή του όρου του αυτοπαλίνδρομου πρώτης τάξης μαζί με το τυπικό σφάλμα της εκτίμησης, την τιμή της *t* στατιστικής για τον έλεγχο της στατιστικής του σημαντικότητας και το *P-Value* του ελέγχου.



Εικόνα 3.3-7. Γραφική απεικόνιση του υποδείγματος

Στην (εικόνα 3.3-7) απεικονίζονται οι τιμές της χρονοσειράς του *Latitude* μαζί με τις εκτιμήσεις τους και τα κάτω όρια του διαστήματος εμπιστοσύνης των εκτιμήσεων.

Συνεχίζοντας με τον έλεγχο του υποδείγματος θα ελέγξουμε τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις της χρονοσειράς των σφαλμάτων.



Εικόνα 3.3-8. Διαγράμματα αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων

Από τα διαγράμματα των αυτοσυσχετίσεων και των μερικών αυτοσυσχετίσεων των σφαλμάτων (εικόνα 3.3-8) διαπιστώνουμε ότι δεν υπάρχουν συσχετίσεις που να υπερβαίνουν τα όρια των διαστημάτων εμπιστοσύνης.

Autocorrelations

Series: Noise residual from Latitude-Model_1

Lag	Autocorrelation	Std. Error ^a	Box-Ljung Statistic		
			Value	df	Sig. ^b
1	-,081	,099	,664	1	,415
2	,022	,098	,713	2	,700
3	,076	,098	1,323	3	,724
4	,013	,097	1,340	4	,855
5	-,087	,097	2,139	5	,830
6	-,087	,096	2,960	6	,814
7	-,085	,096	3,741	7	,809
8	-,042	,095	3,936	8	,863
9	-,115	,095	5,403	9	,798
10	-,015	,094	5,430	10	,861
11	,026	,094	5,507	11	,904
12	,006	,093	5,511	12	,939
13	,029	,093	5,607	13	,959
14	,137	,092	7,821	14	,898
15	-,035	,092	7,963	15	,925
16	,003	,091	7,964	16	,950

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Πίνακας 3.3-11. Ο πίνακας των αυτοσυσχετίσεων της χρονοσειράς των σφαλμάτων

Επίσης ελέγχουμε τα σφάλματα να είναι λευκός θόρυβος κάτι που γίνεται με την στατιστική Q των *Box* και *Ljung*. Παρατηρούμε από τον (πίνακα 3.3-11) ότι πράγματι τα σφάλματα είναι λευκός θόρυβος.

Επιπρόσθετα για να είμαστε ακόμα πιο πολύ σίγουροι για την ορθότητα του υποδείγματος που επιλέξαμε θα υπολογίσουμε μερικά διευρυμένα υποδείγματα με σκοπό να βρούμε ότι οι συντελεστές τους δεν είναι στατιστικά σημαντικοί. Γι' αυτό τον λόγο τα διευρυμένα μοντέλα που θα εξετάσουμε είναι τα ARIMA(2,1,0) και ARIMA(1,1,1).

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Latitude-Model_1	Latitude	No Transformation	AR Lag 1	,685	,101	6,748	,000
			Lag 2	,079	,101	,776	,440
			Difference	1			

Πίνακας 3.3-12. Οι τιμές των παραμέτρων του ARIMA(2,1,0) υποδείγματος για την χρονοσειρά Latitude

Παρατηρούμε από τον (πίνακα 3.3-12) ότι ο επιπρόσθετος όρος που βάλαμε δεν είναι στατιστικά σημαντικός κάτι που αποδεικνύει την ορθότητα της επιλογής του μοντέλου μας.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Latitude-Model_1	Latitude	No Transformation	AR Lag 1	,787	,083	9,504	,000
			Difference	1			
			MA Lag 1	,099	,136	,727	,469

Πίνακας 3.3-13. Οι τιμές των παραμέτρων του ARIMA(1,1,1) υποδείγματος για την χρονοσειρά Latitude

Ομοίως και στο διευρυμένο μοντέλο $ARIMA(1,1,1)$ βλέπουμε ότι ο επιπλέον όρος που προσθέσαμε δεν είναι στατιστικά σημαντικός.

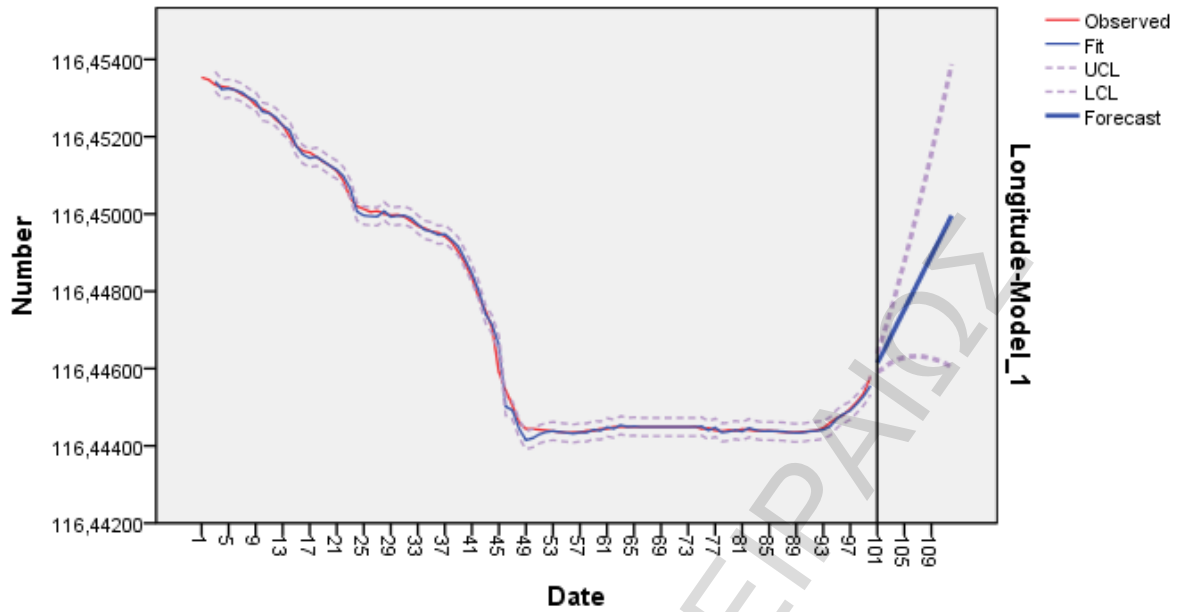
Άρα το υπόδειγμα που θα χρησιμοποιήσουμε για την εκτίμηση της μελλοντικής τιμής για την χρονοσειρά *Latitude* είναι το $ARIMA(1,1,0)$.

3.4 Πρόβλεψη τιμών ανά χρονοσειρά

Πλέον μπορούμε να κάνουμε την πρόβλεψη μελλοντικής θέσης, με τα υποδείγματα χρονοσειρών που βρήκαμε στην προηγούμενη υποενότητα, για τα δεδομένα μας που χρησιμοποιούμε. Δηλαδή για την χρονοσειρά *Longitude* θα χρησιμοποιήσουμε το υπόδειγμα $ARIMA(2,2,0)$ και για την χρονοσειρά *Latitude* το υπόδειγμα $ARIMA(1,1,0)$.

3.4.1 Πρόβλεψη για την χρονοσειρά Longitude

Συγκεκριμένα για το υπόδειγμα της χρονοσειράς *Longitude* θα προβλέψουμε μέχρι ένα λεπτό μετά την τελευταία καταγραφή που είναι στις 13:39:03 ή διαφορετικά θα προβλέψουμε μέχρι την στιγμή 112.



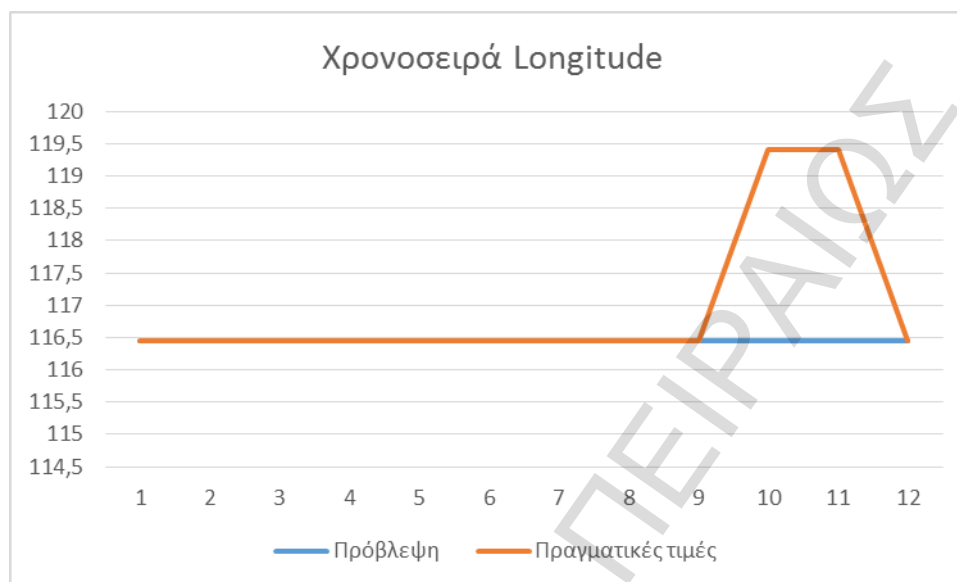
Εικόνα 3.4-1. Γραφική απεικόνιση της πρόβλεψης με το υπόδειγμα APIMA(2,2,0)

Από την παραπάνω (εικόνα 3.4-1) μπορούμε να δούμε την γραφική απεικόνιση των μελλοντικών τιμών της χρονοσειράς *Longitude*, μαζί με τα διαστήματα εμπιστοσύνης των τιμών αυτών. Για να δούμε όμως σε σύγκριση με τα κανονικά δεδομένα πόσο κοντά στις πραγματικές τιμές βρέθηκαν οι προβλεπόμενες τιμές.

Forecast			
	Model		
	Longitude-Model_1		
	Forecast	UCL	LCL
101	116,44615	116,44639	116,44592
102	116,44647	116,44693	116,44602
103	116,44683	116,44750	116,44615
104	116,44718	116,44812	116,44624
105	116,44752	116,44876	116,44629
106	116,44787	116,44943	116,44632
107	116,44822	116,45012	116,44632
108	116,44857	116,45083	116,44631
109	116,44892	116,45156	116,44627
110	116,44926	116,45232	116,44621
111	116,44961	116,45309	116,44614
112	116,44996	116,45388	116,44604

Πίνακας 3.4-1. Οι προβλεπόμενες τιμές της χρονοσειράς *Longitude* μαζί με τα πάνω και κάτω όρια

Παραπάνω βλέπουμε τις τιμές της χρονοσειράς *Longitude* που προβλέψαμε με το υπόδειγμα *ARIMA(2,2,0)*.

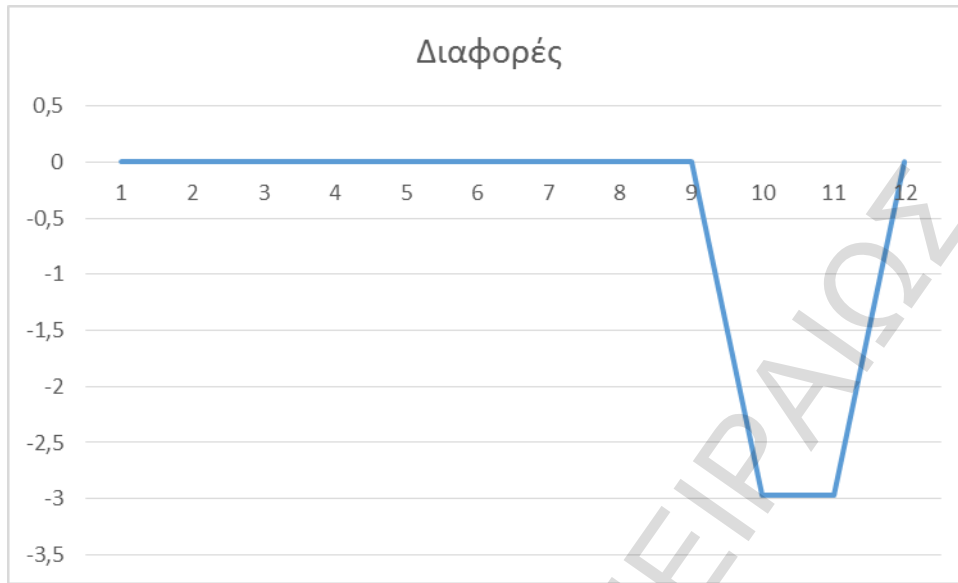


Εικόνα 3.4-2. Γραφική σύγκριση προβλεπόμενων με πραγματικών τιμών της χρονοσειράς *Longitude*

Επειδή από το παραπάνω διάγραμμα (εικόνα 3.4-2) δεν βλέπουμε ξεκάθαρα τις διαφορές μεταξύ της προβλεπόμενης και της πραγματικής τιμής θα παρουσιάσω παρακάτω τον συγκριτικό πίνακα με τις τιμές μαζί με ένα σχεδιάγραμμα με τις διαφορές (πίνακας 3.4-2) και (εικόνα 3.4-3).

Πρόβλεψη	Πραγματικές τιμές	Διαφορές
116,44615	116,44603	0,00012
116,44647	116,44598	0,00049
116,44683	116,44625	0,00058
116,44718	116,44686	0,00032
116,44752	116,44686	0,00066
116,44787	116,44686	0,00101
116,44822	116,44686	0,00136
116,44857	116,44686	0,00171
116,44892	116,44686	0,00206
116,44926	119,41647	-2,96721
116,44961	119,41647	-2,96686
116,44996	116,44686	0,0031

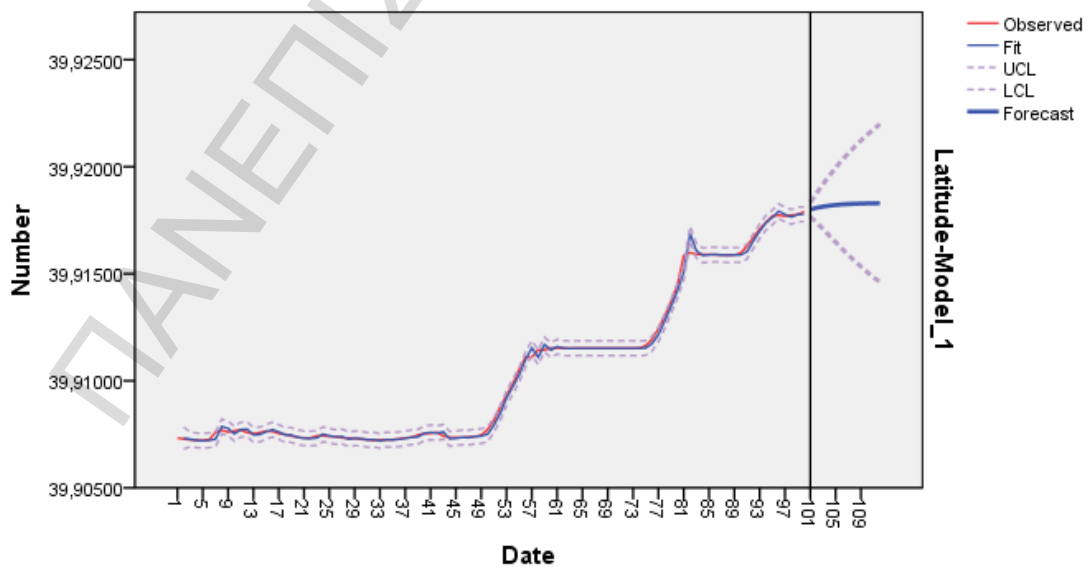
Πίνακας 3.4-2. Οι προβλεπόμενες και πραγματικές τιμές της χρονοσειράς *Longitude*



Εικόνα 3.4-3. Η γραφική απεικόνιση της διαφοράς των προβλεπόμενων με των πραγματικών τιμών της χρονοσειράς Longitude

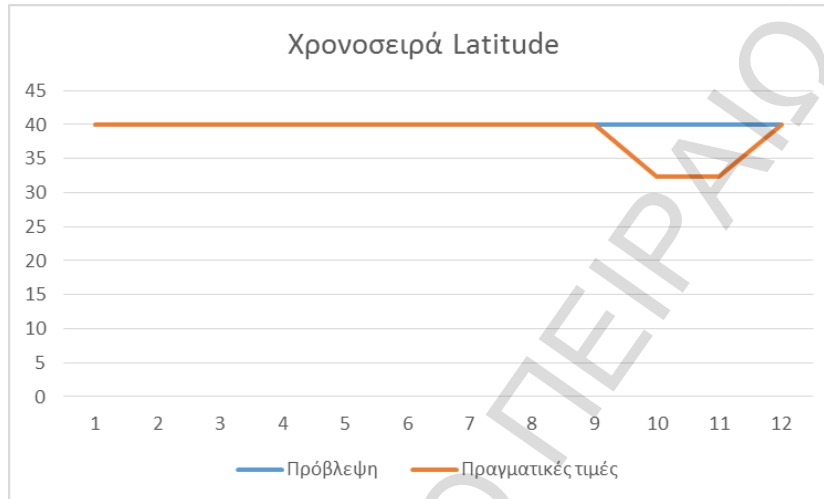
3.4.2 Πρόβλεψη για την χρονοσειρά Latitude

Στην συνέχεια θα παρουσιάσουμε ομοίως και την πρόβλεψη της χρονοσειράς *Latitude*, συγκεκριμένα παρακάτω (εικόνα 3.4-4) βλέπουμε την γραφική απεικόνιση των μελλοντικών τιμών της χρονοσειράς *Latitude*, μαζί με τα διαστήματα εμπιστοσύνης των τιμών αυτών.



Εικόνα 3.4-4. Γραφική απεικόνιση της πρόβλεψης με το υπόδειγμα APIMA(1,1,0)

Στον (πίνακα 3.4-3) βλέπουμε και τις τιμές της πρόβλεψης με την χρήση του υποδείγματος $ARIMA(1,1,0)$ μαζί με τα διαστήματα εμπιστοσύνης.



Εικόνα 3.4-5. Γραφική σύγκριση προβλεπόμενων με πραγματικών τιμών της χρονοσειράς Latitude

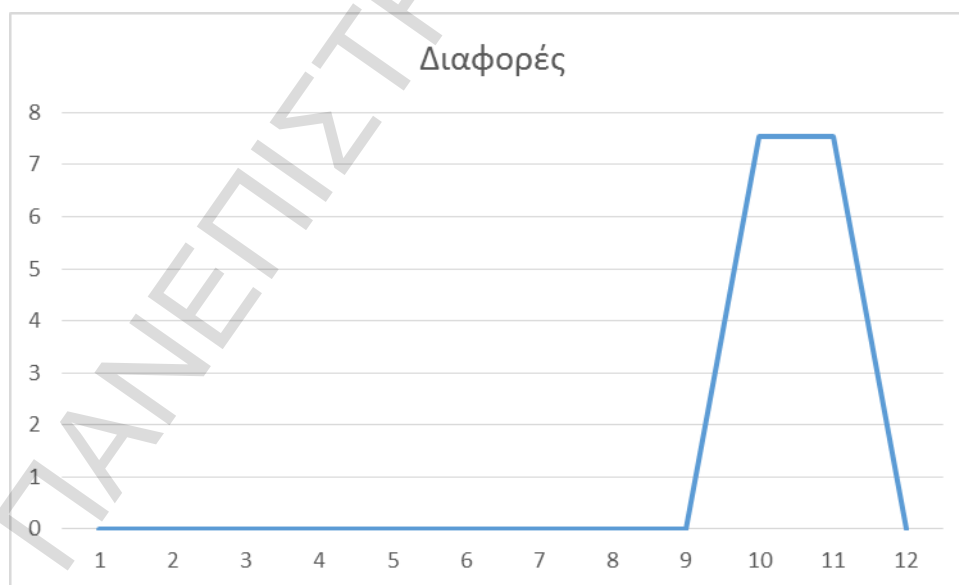
Forecast			
	Model		
	Latitude-Model_1		
	Forecast	UCL	LCL
101	39,91800	39,91835	39,91766
102	39,91808	39,91877	39,91739
103	39,91814	39,91919	39,91709
104	39,91818	39,91958	39,91678
105	39,91821	39,91996	39,91647
106	39,91824	39,92030	39,91617
107	39,91826	39,92063	39,91588
108	39,91827	39,92094	39,91560
109	39,91828	39,92122	39,91533
110	39,91829	39,92149	39,91508
111	39,91829	39,92175	39,91483
112	39,91830	39,92199	39,91460

Πίνακας 3.4-3. Οι προβλεπόμενες τιμές της χρονοσειράς Latitude μαζί με τα πάνω και κάτω όρια

Επειδή από το παραπάνω διάγραμμα (εικόνα 3.4-5) δεν βλέπουμε ξεκάθαρα τις διαφορές μεταξύ της προβλεπόμενης και της πραγματικής τιμής θα παρουσιάσω παρακάτω τον συγκριτικό πίνακα με τις τιμές μαζί με ένα σχεδιάγραμμα με τις διαφορές (πίνακας 3.4-4) και (εικόνα 3.4-6).

Πρόβλεψη	Πραγματικές τιμές	Διαφορές
39,918	39,91775	0,00025
39,91808	39,91778	0,0003
39,91814	39,91752	0,00062
39,91818	39,91706	0,00112
39,91821	39,91706	0,00115
39,91824	39,91706	0,00118
39,91826	39,91706	0,0012
39,91827	39,91706	0,00121
39,91828	39,91706	0,00122
39,91829	32,36623	7,55206
39,91829	32,36623	7,55206
39,9183	39,91706	0,00124

Πίνακας 3.4-4. Οι προβλεπόμενες και πραγματικές τιμές της χρονοσειράς Latitude

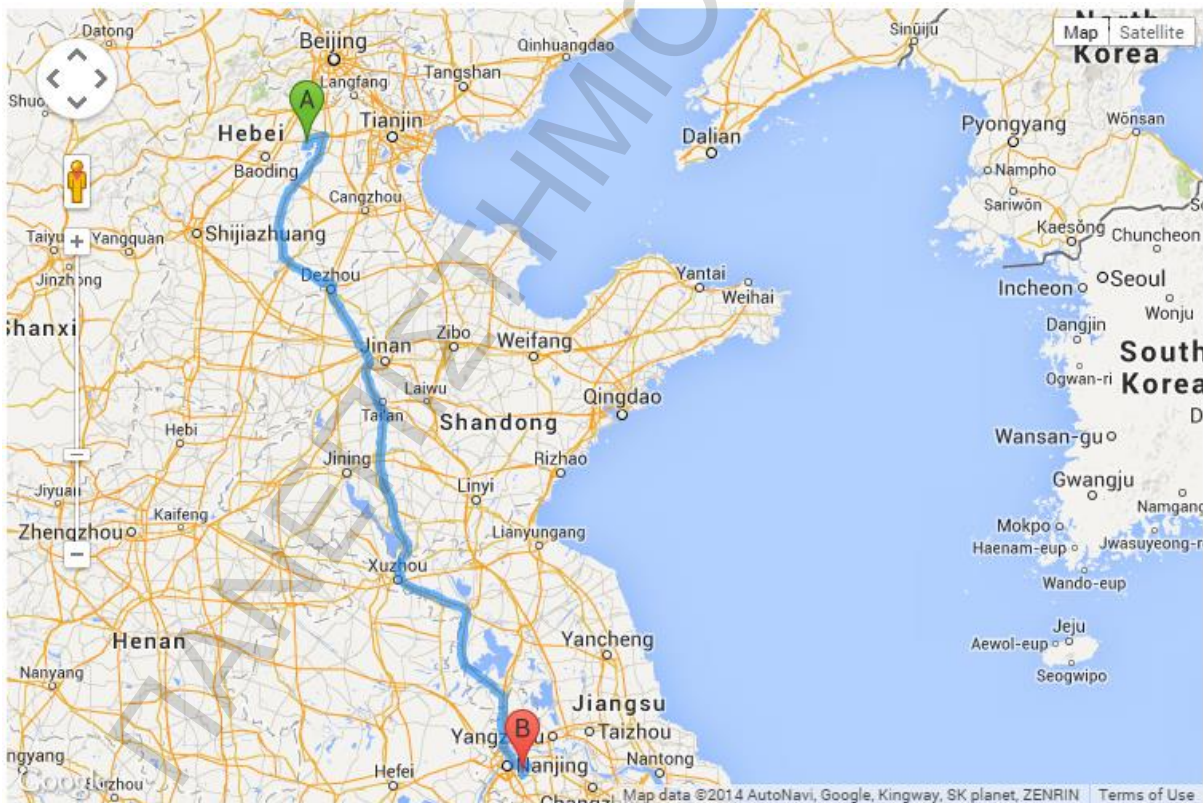


Εικόνα 3.4-6. Η γραφική απεικόνιση της διαφοράς των προβλεπόμενων με των πραγματικών τιμών της χρονοσειράς Latitude

3.4.3 Πρώτα συμπεράσματα

Σε αυτό το σημείο θα ήθελα να αναφέρω ότι, παρατηρώντας την (εικόνα 3.4-2) και τον (πίνακα 3.4-2), για την χρονοσειρά *Longitude*, βλέπουμε ότι από την 104 μέχρι την 109, αλλά και την 112 οι πραγματικές τιμές είναι ίδιες, δηλαδή 116,44686. Για τις καταγραφές 110 και 111 παρατηρούμε ένα τεράστιο άλμα στις τιμές, 119,41647. Το ίδιο ακριβώς πρόβλημα το παρατηρούμε και στην χρονοσειρά *Latitude*. Συγκεκριμένα, έχουμε την ίδια τιμή 39,91706 για τις καταγραφές 104-109 και 112, ενώ για τις καταγραφές 110 και 111 έχουμε την τιμή 32,36623.

Άρα παρατηρούμε την μετακίνηση του αντικειμένου από το σημείο A του χάρτη⁴ που αντιστοιχεί στις συντεταγμένες (116,44686-39,91706) στο σημείο B με συντεταγμένες (119,41647-32,36623) μόλις σε 5 δευτερόλεπτα. Η πραγματική απόσταση των δύο σημείων αντιστοιχεί σε 1204 χιλιόμετρα, απόσταση που δεν μπορεί να καλυφθεί σε 5 δευτερόλεπτα. Άρα θεωρούμε ότι οι καταγραφές 110 και 111 που παρουσιάζουν το αντικείμενο στην θέση B είναι λάθος του *GPS*, άρα τις αφαιρούμε από την πρόβλεψή μας.



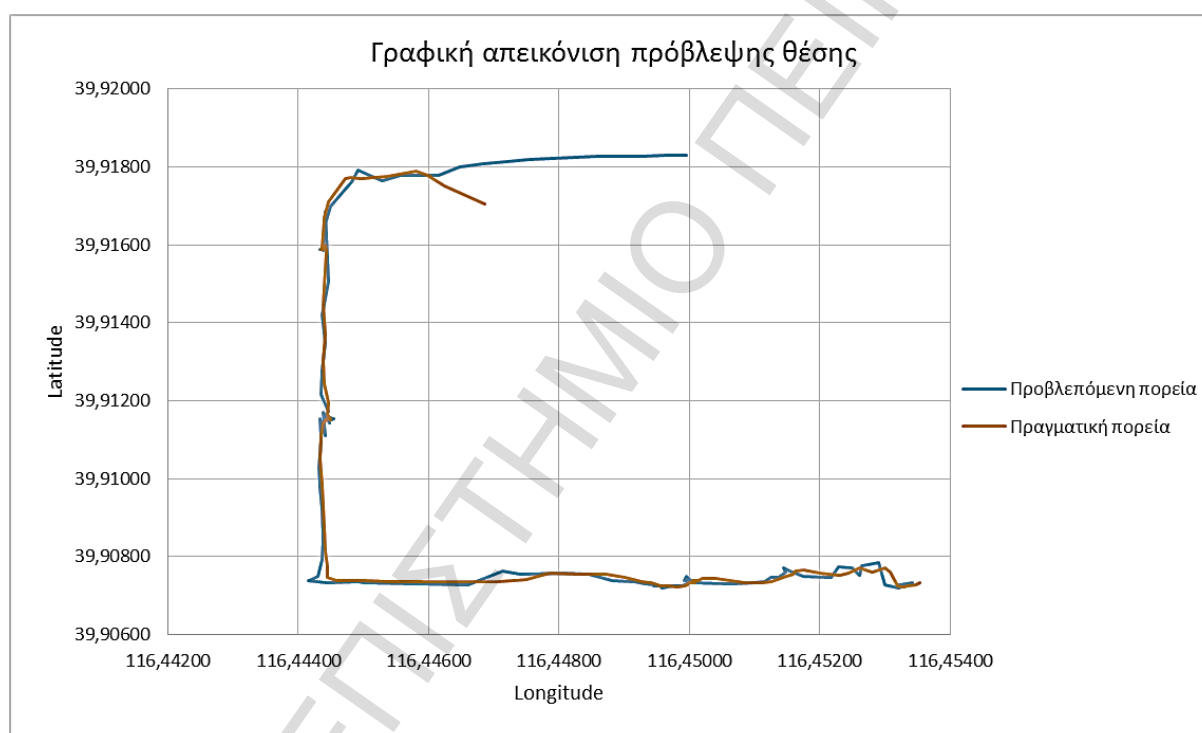
Εικόνα 3.4-7. Χάρτης που απεικονίζει τις δύο λανθασμένες καταγραφές

⁴Google Driving Directions: <http://www.gps-coordinates.net/driving-directions>

Οπότε θα θεωρήσουμε ότι το αντικείμενο από την καταγραφή 104 έως 112 ότι είναι στάσιμό στο σημείο A (116,44686-39,91706).

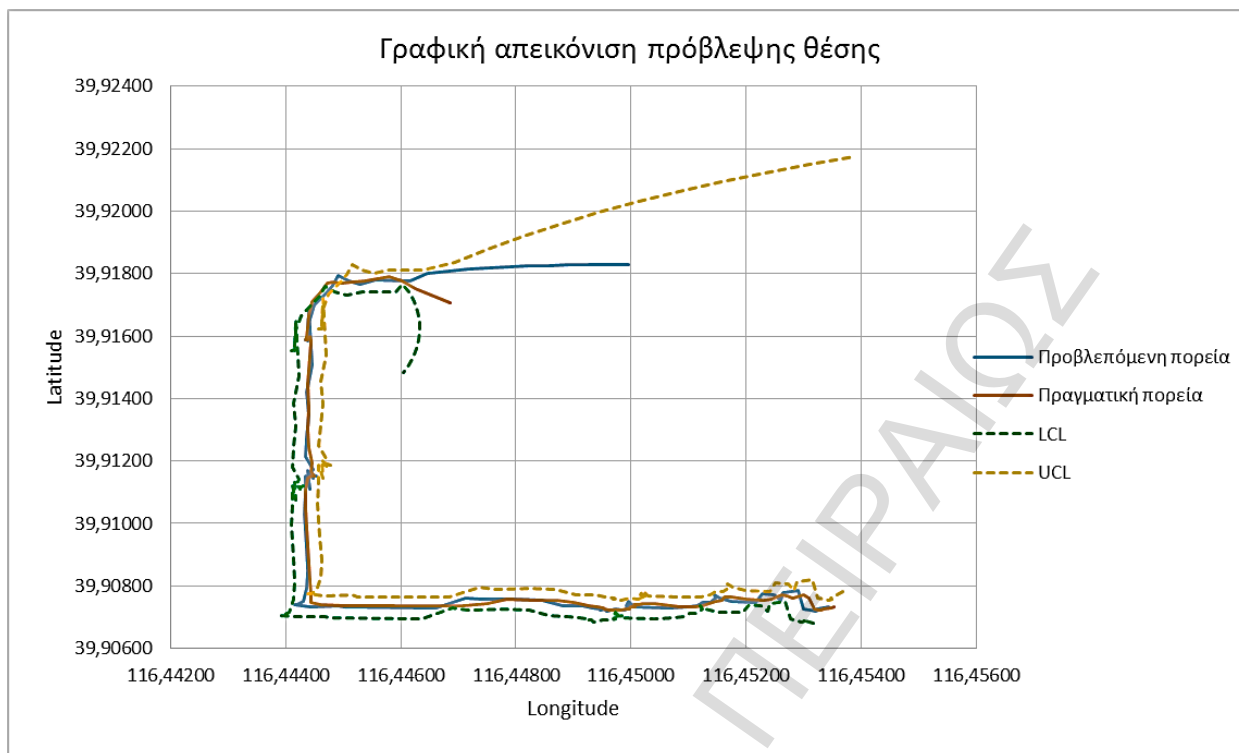
3.5 Πρόβλεψη μελλοντικής θέσης

Αφού προβλέψαμε τις τιμές ανά χρονοσειρά, δηλαδή για την *Longitude* και *Latitude* χρονοσειρά, πλέον θα ενώσουμε ξανά τις τιμές αυτές, για να συνθέσουμε την προβλεπόμενη θέση, αλλά και τη τροχιά του αντικειμένου μας και θα την συγκρίνουμε με την πραγματική θέση και την πραγματική τροχιά.



Εικόνα 3.5-1. Γραφική απεικόνιση προβλεπόμενης πορείας

Πράγματι από την παραπάνω γραφική απεικόνιση (εικόνα 3.5-1) βλέπουμε την τροχιά του αντικειμένου που ακολουθεί στην πραγματικότητα, αλλά και την προβλεπόμενη τροχιά, που βρέθηκε με την χρήση των χρονοσειρών. Αυτό που παρατηρούμε είναι, ότι ενώ στην αρχή είμαστε κοντά στην πραγματική θέση, όσο μεγαλώνει ο χρόνος πρόβλεψης η απόκλιση της πραγματικής θέσης του αντικειμένου σε σύγκριση με την προβλεπόμενη μεγαλώνει.



Εικόνα 3.5-2. Γραφική απεικόνιση προβλεπόμενης πορείας με τα πάνω και κάτω όρια της πρόβλεψης

Πλέον παρατηρώντας τα πάνω και κάτω όρια της πρόβλεψης (εικόνα 3.5-2) ενισχύουμε τα προηγούμενα συμπεράσματα. Δηλαδή, κοιτώντας τα πάνω και κάτω όρια βλέπουμε ότι όσο μεγαλώνει η περίοδος πρόβλεψης τόσο μεγαλώνουν και αυτά κάτι που οδηγεί στο συμπέρασμα ότι η πρόβλεψη γίνεται πιο αναξιόπιστη. Αντίθετα όταν είναι να προβλέψουμε πολύ μικρό χρονικό διάστημα η πρόβλεψή μας είναι πιο εύστοχη. Σε αυτό το σημείο να αναφέρω ότι παρά την αστοχία πρόβλεψης για μεγάλα χρονικά διαστήματα, η πραγματική θέση βρίσκεται πάντα μέσα στα όρια εμπιστοσύνης μας. Αυτό είναι λογικό να συμβαίνει γιατί η πρόβλεψή μας έγινε με ένα μαθηματικό μοντέλο και συγκεκριμένα με την βοήθεια των χρονοσειρών. Τα υποδείγματα των χρονοσειρών είναι αρκετά καλά στην πρόβλεψη μιας χρονικής περιόδου μόνο, κάτι που φαίνεται και στην πρόβλεψή μας.

Τέλος για να δώσουμε στον αναγνώστη της παρούσης διπλωματικής εργασίας, μια καλύτερη οπτική παρουσίαση των αποτελεσμάτων της μεθόδου που χρησιμοποιήσαμε για την πρόβλεψη της μελλοντικής θέσης, παραθέτουμε την παρακάτω εικόνα (εικόνα 3.5-3). Αυτό που εύκολα μπορούμε να διακρίνουμε είναι ότι για τις πρώτες δύο περιόδους η πρόβλεψή μας τα πήγε πολύ καλά, όπως είναι λογικό, επειδή χρησιμοποιούμε μαθηματικό μοντέλο. Αντίθετα στην συνέχεια, άρχισε να χάνει ακρίβεια, πόσο μάλλον που στην περίπτωσή μας το ταξί σταμάτησε να

κινείται για τις υπόλοιπες 10 περιόδους, κάτι που φαίνεται και στην πρόβλεψή μας που συνεχίζει να προβλέπει, ενώ στην πραγματικότητα έχει σταματήσει.



Εικόνα 3.5-3. Απεικόνιση σε πραγματικό χάρτη των πραγματικών αλλά και των προβλεπόμενων θέσεων του αντικείμενου

3.6 Αξιολόγηση της πρόβλεψης

Τα υποδείγματα χρονοσειρών που χρησιμοποιήσαμε για την πρόβλεψη της τροχιάς του ταξί με κωδικό 366, δηλαδή το $ARIMA(2,2,0)$ για το γεωγραφικό μήκος και το $ARIMA(1,1,0)$ για το γεωγραφικό πλάτος, λόγω της φύσης τους αποδείχθηκαν ικανά για βραχυπρόθεσμη πρόβλεψη. Παρατηρήθηκε ότι η πρόβλεψη είναι πιο ακριβείς για 1 έως 2 περιόδους πρόβλεψης, ενώ όσο κάνουμε πρόβλεψη για μακρινότερες χρονικά περιόδους η ακρίβεια της πρόβλεψης φθίνει.

Για τον λόγο αυτό δημιουργήσαμε ένα μέτρο, μαθηματικής φύσης, για να αξιολογήσουμε την πραγματική απόδοση της πρόβλεψης της μελλοντικής θέσης με την χρήση των υποδειγμάτων χρονοσειρών.

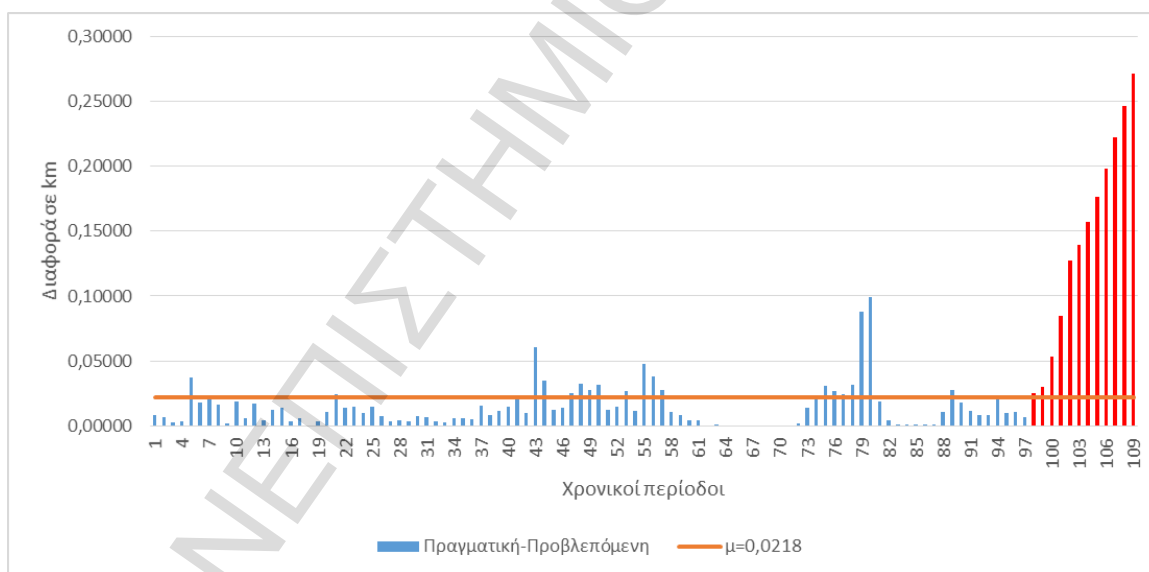


Εικόνα 3.6-1. Γραφική απεικόνιση του μέτρου

Συγκεκριμένα, θεωρούμε το κέντρο του κύκλου την πραγματική θέση του αντικειμένου, σύμφωνα με τις συντεταμένες του, και η ακτίνα του κύκλου, που την συμβολίσαμε με μ , είναι η μέση απόσταση ανά δύο διαδοχικών με βάση των 100 σημείων που χρησιμοποιήσαμε για την κατασκευή των υποδειγμάτων του ταξί με κωδικό 366 (εικόνα 3.6-1). Όταν η προβλεπόμενη θέση είναι εντός της περιφέρειας του κυκλικού δίσκου τότε θεωρούμε ότι έχουμε επιτυχία μεγαλύτερη ή ίση του 100%, με ίση πάνω στα όρια του κύκλου. Αντίθετα, για πρόβλεψη που απέχει μεγαλύτερη απόσταση από μ , που στην περίπτωση μας την έχουμε ονομάσει μ' , από την πραγματική τιμή, η ακρίβεια μειώνεται αναλογικά με την απόσταση μ' .

$$\text{Μέτρο} = \frac{\mu}{\mu'} * 100 \quad (3.2)$$

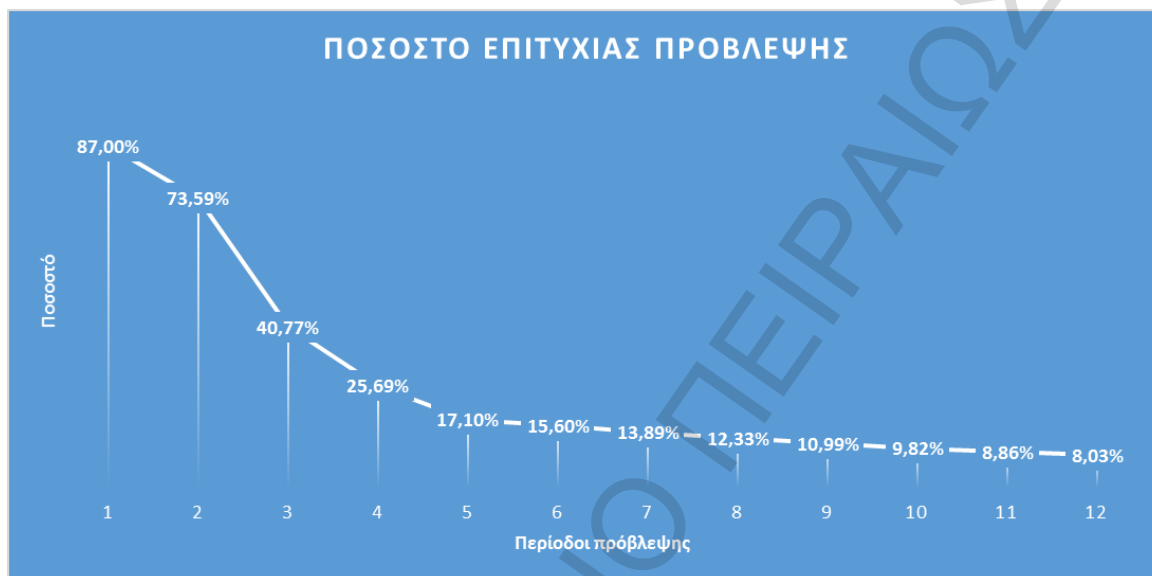
Οπότε, βάση του πάνω μέτρου (τύπος 3.2) και με μέση απόσταση $\mu=0,0218$ χιλιόμετρα, μπορούμε να υπολογίσουμε την ακριβή απόδοση της προβλεπτικής ικανότητας των υποδειγμάτων χρονοσειρών που χρησιμοποιήσαμε.



Εικόνα 3.6-2. Απόσταση της προβλεπόμενης από την πραγματική θέση, καθ'όλη την διάρκεια της τροχιάς του ταξί 366

Από το παραπάνω γράφημα (γράφημα 3.6-2) βλέπουμε την απόσταση της πραγματικής τιμής μεταξύ της προβλεπόμενης τιμής σε σχέση με την μέση απόσταση μ . Όλες οι γραμμές αναπαριστούν την διαφορά αυτή, με τις μπλε γραμμές μας να είναι στις χρονικές περιόδους

στις οποίες βασιστήκαμε για να κατασκευάσουμε το μοντέλο, ενώ οι κόκκινες γραμμές αναπαριστούν την πρόβλεψη για 12 περιόδους. Βάση λοιπόν του παραπάνω γραφήματος, βλέπουμε ότι η μέθοδός μας είναι αρκετά ακριβής για τις δύο πρώτες προβλέψεις (δύο πρώτες κόκκινες γραμμές) ενώ στην συνέχεια χάνει σε απόδοση.



Εικόνα 3.6-3. Απόδοση της μεθόδου που χρησιμοποιήσαμε

Πράγματι κοιτώντας το παραπάνω γράφημα (εικόνα 3.6-3) καταλήγουμε και μαθηματικά στο ίδιο συμπέρασμα, που μας προϊδέαζαν οι γραφικές απεικονίσεις των προβλέψεων, δηλαδή ότι η μέθοδός μας τα πάει περίφημα για δύο περιόδους πρόβλεψης με ποσοστό απόδοσης 87% και 73,59% αντίστοιχα, ενώ από την τρίτη περίοδο και μετά βλέπουμε μια πτωτική απόδοση κάτι που είναι λογικό, από την στιγμή που χρησιμοποιούμε ένα καθαρά μαθηματικό μοντέλο για τις προβλέψεις μας.

3.7 Δευτερεύουσα πρακτική εφαρμογή

Σε αυτή την ενότητα θα προσπαθήσουμε να γενικεύσουμε ένα υπόδειγμα χρονοσειρών για πολλές χρονοσειρές, που στην περίπτωσή μας είναι τροχιές ταξί, για την πρόβλεψη μελλοντική θέσης κινούμενων αντικειμένων.

Θα ξεκινήσουμε λοιπόν επαναπροσδιορίζοντας την πρακτική εφαρμογή που κάναμε για μια τροχιά ενός ταξί. Συγκεκριμένα, αυτό που θέλαμε να κάνουμε είναι να βρούμε 4 δημοφιλή υποδείγματα χρονοσειρών από 50 τροχιές ταξί και στην συνέχεια να εφαρμόσουμε αυτά τα

υποδείγματα και για τις 50 τροχιές ταξί. Αυτή ήταν η αρχική μας ιδέα, αλλά για να δούμε πως ανταποκρίνεται στην πραγματικότητα, τα 4 δημοφιλέστερα υποδείγματα χρονοσειρών των 50 τροχιών ταξί, τα εφαρμόσαμε σε 10 τυχαία επιλεγμένες τροχιές ταξί από τις 50.

ID									
366	2126	4396	6166	7376	8526	9946	1356	1776	2866
1336	2146	5316	6656	7396	8696	16	1386	1946	2876
1466	2486	5856	6876	7416	8736	136	1526	2226	2926
1566	3866	5866	7056	7806	9206	146	1746	2566	3096
1796	4346	5896	7146	8126	9556	996	1756	2596	9336

Πίνακας 3.7-1. Πίνακας των 50 επιλεγμένων ταξί

Για να επιλέξουμε αυτά τα 50 ταξί, για να εφαρμόσουμε την ιδέα μας, επιλέχθηκαν τυχαία 50 από αυτά που ο ID αριθμός τους τελείωνε σε 6. Έτσι προέκυψαν τα παραπάνω 50 ταξί (πίνακας 3.6-1).

Τα δημοφιλέστερα υποδείγματα των 50 τροχιών, τα οποία τα βρήκα με την βοήθεια του στατιστικού προγράμματος *SPSS*, ήταν τα εξής,

ARIMA(0,1,1)	ARIMA(1,1,0)	ARIMA(0,1,0)	ARIMA(0,1,2)	ARIMA(2,1,0)
9	25	32	9	5

Πίνακας 3.7-2. Τα δημοφιλέστερα υποδείγματα τροχιών

Από τον παραπάνω πίνακα βλέπουμε ότι τα δημοφιλέστερα υποδείγματα είναι τα $ARIMA(0,1,1)$, $ARIMA(1,1,0)$, $ARIMA(0,1,0)$, $ARIMA(0,1,2)$ και $ARIMA(2,1,0)$. Σε αυτό το σημείο να αναφέρω ότι το υπόδειγμα $ARIMA(0,1,0)$ δεν έχει καμία προβλεπτική ικανότητα, οπότε και αφαιρέθηκε για την περαιτέρω μελέτη.

Πλέον έχοντας τα υποδείγματα χρονοσειρών, αυτό που κάναμε ήταν να τα εφαρμόσουμε και για τις 10 πρώτες τροχιές ταξί ξεχωριστά.

	ARIMA(1,1,0)			ARIMA(0,1,1)		
	Ljung-Box Q(18)	t-test	Residuals	Ljung-Box Q(18)	t-test	Residuals
Longitude_366			x	x		
Latitude_366				x		
Longitude_1336				x		
Latitude_1336						
Longitude_1466			x			x
Latitude_1466					x	
Longitude_1566		x			x	
Latitude_1566		x			x	
Longitude_1796		x			x	
Latitude_1796						
Longitude_2126		x				x
Latitude_2126			x		x	
Longitude_2146			x			x
Latitude_2146			x	x		
Longitude_2486		x			x	
Latitude_2486		x			x	
Longitude_3866		x			x	
Latitude_3866		x			x	
Longitude_4346						
Latitude_4346						x
Ποσοστό αποδοχής των υποδειγμάτων		35%			15%	

Πίνακας 3.7-3. Αποτελέσματα καλής εφαρμογής των υποδειγμάτων χρονοσειρών

Τις τεχνικές και τις μεθόδους που χρησιμοποιήσαμε για να βρούμε το κατάλληλο υπόδειγμα για την τροχιά του ταξί 366 (βλέπε υποενότητα 3.3), τις χρησιμοποιήσαμε και εδώ για κάθε ένα από τα δημοφιλέστερα υποδείγματα που εφαρμόστηκαν στις 10 πρώτες τροχιές των ταξί.

	ARIMA(0,1,2)			ARIMA(2,1,0)		
	Ljung-Box Q(18)	t-test	Residuals	Ljung-Box Q(18)	t-test	Residuals
Longitude_366	x				x	
Latitude_366			x		x	
Longitude_1336	x					
Latitude_1336		x			x	
Longitude_1466		x			x	
Latitude_1466		x			x	
Longitude_1566		x			x	
Latitude_1566		x			x	
Longitude_1796		x			x	
Latitude_1796		x			x	
Longitude_2126		x			x	
Latitude_2126		x			x	
Longitude_2146						x
Latitude_2146						x
Longitude_2486		x			x	
Latitude_2486		x			x	
Longitude_3866		x			x	
Latitude_3866		x			x	
Longitude_4346		x			x	
Latitude_4346					x	
Ποσοστό αποδοχής των υποδειγμάτων		15%			5%	

Πίνακας 3.7-4 Αποτελέσματα καλής εφαρμογής των υποδειγμάτων χρονοσειρών

Συγκεκριμένα, με την τιμή της στατιστικής Q των Box και Ljung, είδαμε ότι για το υπόδειγμα $ARIMA(1,1,0)$ τα σφάλματα των χρονοσειρών είναι λευκός θόρυβος σε ε.σ. 5% ⁵, οπότε μπορούμε να συνεχίσουμε στον επόμενο έλεγχο. Στον έλεγχο $t-test$, κοιτάμε αν είναι στατιστικά σημαντικά οι παράμετροι των υποδειγμάτων. Δηλαδή στην περίπτωση μας, στο υπόδειγμα χρονοσειρών $ARIMA(1,1,0)$, βλέπουμε ότι για τις χρονοσειρές που στον πίνακα έχει σημειωθεί με x ένας ή περισσότεροι όροι του υποδείγματος δεν είναι στατιστικά σημαντικοί (σε ε.σ. 5% ⁶). Στο επόμενο στάδιο ελέγξαμε ξανά για το αν τα σφάλματα των επιλεγθέντων χρονοσειρών είναι λευκός θόρυβος. Έτσι με τον autocorrelation πίνακα είδαμε και εδώ ότι δεν πέρασαν τον έλεγχο όλες οι χρονοσειρές μας (επίσης αυτές που δεν πέρασαν τον έλεγχο σε ε.σ. 5% σημειώνονται με x στους παραπάνω πίνακες). Σε αυτό το σημείο να αναφέρω ότι αν μια χρονοσειρά δεν πέρασε ένα έλεγχο, με την σειρά που τους αναφέραμε, δεν συνεχιζόταν ο

⁵ Βλέπε Παράρτημα 2

⁶ Βλέπε Παράρτημα 2

έλεγχός της στους επόμενους ελέγχους, για τον λόγο ότι απορριπτόταν σε προηγούμενο στάδιο το υπόδειγμα.

Οπότε με την μεθοδολογία που περιγράψαμε, μέσω των δύο παραπάνω πινάκων (3.6-2 και 3.6-3 πίνακας) βλέπουμε ότι για το υπόδειγμα χρονοσειρών $ARIMA(1,1,0)$, πέρασαν όλους τους ελέγχους 7 στις 20 χρονοσειρές δηλαδή το 35%, αλλά στην πραγματικότητα μόνο δύο τροχιές (γεωγραφικό πλάτος και μήκος ενός ταξί). Αντίστοιχα και για το υπόδειγμα $ARIMA(0,1,1)$, μόνο τρεις χρονοσειρές πέρασαν όλους τους ελέγχους, 15%, και καμία τροχιά. Επίσης για το υπόδειγμα $ARIMA(0,1,2)$, τρεις χρονοσειρές πέρασαν όλους τους ελέγχους, 15%, και μία τροχιά και τέλος για το $ARIMA(2,1,0)$ πέρασε μόνο μία τροχιά όλους τους ελέγχους, 5%, και όπως είναι φυσικό καμία τροχιά.

Άρα καταλήξαμε στο συμπέρασμα ότι δεν μπορούμε να χρησιμοποιήσουμε ένα υπόδειγμα χρονοσειρών για όλες τις χρονοσειρές που επιλέξαμε, γιατί όπως είδαμε από τα παραπάνω κάθε υπόδειγμα μετά βίας εφαρμοζόταν σε μία τροχιά και έτσι δεν μπορούσαμε να συγκρίνουμε τα αποτελέσματα και την απόδοση του υποδείγματος.

Από την φύση του πάντως, το υποδείγματος ARIMA είναι κατάλληλο να εφαρμόζεται σε κάθε μια χρονοσειρά ξεχωριστά, αν θέλαμε να εφαρμόσουμε ένα υπόδειγμα χρονοσειρών για πολλές χρονοσειρές μαζί ίσως θα ήταν καλύτερο το Vector ARIMA, αλλά σε αυτή την περίπτωση θα έπρεπε κάθε όχημα να ξεκινούσε από το ίδιο σημείο εκκίνησης, την ίδια χρονική στιγμή κάτι που δεν το γνωρίζουμε εμείς για τα δικά μας δεδομένα.

ΚΕΦΑΛΑΙΟ 4

Πλέον φτάσαμε στο τέλος της διπλωματικής εργασίας και σε αυτό το σημείο καλό θα ήταν να αναφέρουμε τις απόψεις και τα συμπεράσματά μας.

4.1 Συμπεράσματα

Σκοπός της διπλωματικής εργασίας μας ήταν να προβλέψουμε την θέση σε κινούμενα δεδομένα. Συγκεκριμένα να προβλέψουμε την θέση ενός κινούμενου αντικειμένου με την βοήθεια των προηγούμενων χωρικών και χρονικών καταγραφών του. Στην περίπτωση μας χρησιμοποιήσαμε δεδομένα *GPS*, δηλαδή καταγραφές γεωγραφικού πλάτους και μήκους ανά τακτές και συνεχόμενες χρονικές στιγμές, ταξί της πόλης του Πεκίνου.

Στην συνέχεια έπρεπε να επιλέξουμε την μέθοδο με την οποία θα κάναμε την πρόβλεψή μας, με γνώμονα ότι θα έπρεπε να βασιζόταν σε μια στατιστική μέθοδο. Η μέθοδος που επιλέξαμε ήταν οι χρονοσειρές. Επιλέξαμε τις χρονοσειρές για το λόγο ότι οραματιστήκαμε τα δεδομένα μας σαν το *DNA*, δηλαδή όπως αποτελείτε το *DNA* από δύο επιμήκων αλύσεων, έτσι και η κίνηση του αντικειμένου μας αποτελείται από δύο χρονοσειρές, του γεωγραφικού μήκους και πλάτους. Έτσι “διασπάσαμε” την κίνηση του αντικειμένου σε αυτές τις δύο χρονοσειρές και με τα δεδομένα που είχαμε προσπαθήσαμε να βρούμε το καλύτερο μοντέλο για καθεμία χρονοσειρά ξεχωριστά. Αφού βρήκαμε τα υποδείγματα των χρονοσειρών συνεχίσαμε στην πρόβλεψη των μελλοντικών περιόδων. Στην συνέχεια ξανά συνθέσαμε τις χρονοσειρές που βρήκαμε και δημιουργήσαμε έτσι ξανά την προβλεπόμενη θέση του αντικειμένου μας.

Τα συμπεράσματα που βγάλαμε από την μέθοδο που εφαρμόσαμε, είναι ότι με την συγκεκριμένη στατιστική μέθοδο, δηλαδή των χρονοσειρών, η πρόβλεψή μας τα πήγε πολύ καλά στην εκτίμηση της μελλοντικής θέσης του αντικειμένου μας, ιδιαίτερα σε λίγες περιόδους. Όπως κάθε μαθηματικό μοντέλο, έτσι και τα υποδείγματα των χρονοσειρών, που χρησιμοποιήσαμε στην παρούσα διπλωματική εργασία, είναι αρκετά καλά για την πρόβλεψη βραχυπρόθεσμων μελλοντικών θέσεων του αντικειμένου. Στην περίπτωση μας είδαμε ότι μέχρι δύο περιόδους η πρόβλεψή μας ήταν αρκετά κοντά στην πραγματική θέση του αντικειμένου. Αντίθετα μετά από κάποιες περιόδους η απόκλιση μεταξύ της προβλεπόμενης και της πραγματικής θέσης άρχισε να μεγαλώνει, και από ότι είδαμε στην αναπαράσταση του χάρτη (εικόνα 3.5-3) δεν “βγήκαμε” πολύ έξω από την πραγματική πορεία.

Επίσης προσπαθήσαμε να γενικεύσουμε την μέθοδό μας, δηλαδή να εφαρμόσουμε μόνο ένα υπόδειγμα χρονοσειρών για κάθε χρονοσειρά που είχαμε και να προβλέψουμε την μελλοντική θέση του αντικειμένου μας. Δυστυχώς η γενίκευση αυτή δεν έφερε ικανοποιητικά αποτελέσματα, με αποτέλεσμα να μην συνεχίσουμε την περαιτέρω εφαρμογή της.

Τέλος μετά από τα συγκεκριμένα αποτελέσματα που βγάλαμε από την χρήση υποδειγμάτων χρονοσειρών και παράλληλα διαπιστώνοντας την κύρια αδυναμία τους, δηλαδή την αδυναμία στις μακροπρόθεσμες προβλέψεις, θα προτείναμε στους μελλοντικούς ερευνητές να μελετήσουν την εφαρμογή υποδειγμάτων χρονοσειρών σε συσταδοποιημένα δεδομένα βάση σημείων ενδιαφέροντος, ώστε να βελτιώσουν την απόδοση της πρόβλεψης και σε μακροπρόθεσμες προβλέψεις.

ΠΑΡΑΡΤΗΜΑ 1

366	2/2/2008	13:30:48	116,45353	39,90732
366	2/2/2008	13:30:53	116,45348	39,90729
366	2/2/2008	13:30:58	116,45334	39,90725
366	2/2/2008	13:31:03	116,4533	39,90722
366	2/2/2008	13:31:08	116,45327	39,90722
366	2/2/2008	13:31:13	116,4532	39,90725
366	2/2/2008	13:31:18	116,45309	39,9076
366	2/2/2008	13:31:23	116,453	39,9077
366	2/2/2008	13:31:28	116,45281	39,9076
366	2/2/2008	13:31:33	116,45271	39,90767
366	2/2/2008	13:31:38	116,45262	39,90771
366	2/2/2008	13:31:48	116,45244	39,90758
366	2/2/2008	13:31:53	116,4523	39,90753
366	2/2/2008	13:31:58	116,45201	39,90758
366	2/2/2008	13:32:03	116,45176	39,90766
366	2/2/2008	13:32:08	116,45163	39,90763
366	2/2/2008	13:32:13	116,45159	39,90754
366	2/2/2008	13:32:18	116,45148	39,9075
366	2/2/2008	13:32:23	116,45135	39,90742
366	2/2/2008	13:32:28	116,45125	39,90736
366	2/2/2008	13:32:33	116,45111	39,90733
366	2/2/2008	13:32:38	116,45086	39,90733
366	2/2/2008	13:32:43	116,45041	39,90743
366	2/2/2008	13:32:48	116,4502	39,90743
366	2/2/2008	13:32:53	116,45013	39,9074
366	2/2/2008	13:32:58	116,45005	39,9074
366	2/2/2008	13:33:03	116,45008	39,90733
366	2/2/2008	13:33:08	116,45	39,90732
366	2/2/2008	13:33:13	116,44997	39,90731
366	2/2/2008	13:33:18	116,44998	39,90727
366	2/2/2008	13:33:23	116,44993	39,90726
366	2/2/2008	13:33:28	116,44981	39,90722
366	2/2/2008	13:33:33	116,44969	39,90724
366	2/2/2008	13:33:38	116,44963	39,90724
366	2/2/2008	13:33:43	116,44955	39,90726
366	2/2/2008	13:33:48	116,44952	39,90729
366	2/2/2008	13:33:53	116,44942	39,90733
366	2/2/2008	13:33:58	116,44927	39,90736
366	2/2/2008	13:34:03	116,44902	39,90747
366	2/2/2008	13:34:08	116,44872	39,90754
366	2/2/2008	13:34:13	116,44835	39,90755
366	2/2/2008	13:34:18	116,44785	39,90759
366	2/2/2008	13:34:23	116,4475	39,90742

366	2/2/2008	13:34:28	116,44706	39,90736
366	2/2/2008	13:34:38	116,4459	39,90736
366	2/2/2008	13:34:43	116,44543	39,90735
366	2/2/2008	13:34:48	116,44505	39,90738
366	2/2/2008	13:34:53	116,4446	39,9074
366	2/2/2008	13:34:58	116,44444	39,90746
366	2/2/2008	13:35:03	116,44444	39,90773
366	2/2/2008	13:35:08	116,44442	39,90816
366	2/2/2008	13:35:13	116,4444	39,90876
366	2/2/2008	13:35:18	116,44438	39,90932
366	2/2/2008	13:35:23	116,44436	39,90987
366	2/2/2008	13:35:28	116,44434	39,91052
366	2/2/2008	13:35:33	116,44435	39,9111
366	2/2/2008	13:35:33	116,44435	39,9111
366	2/2/2008	13:35:38	116,44439	39,91144
366	2/2/2008	13:35:38	116,44439	39,91144
366	2/2/2008	13:35:43	116,44444	39,91153
366	2/2/2008	13:35:43	116,44444	39,91153
366	2/2/2008	13:35:48	116,44449	39,91152
366	2/2/2008	13:35:53	116,44449	39,91152
366	2/2/2008	13:35:58	116,44449	39,91152
366	2/2/2008	13:36:03	116,44449	39,91152
366	2/2/2008	13:36:08	116,44449	39,91152
366	2/2/2008	13:36:13	116,44449	39,91152
366	2/2/2008	13:36:18	116,44449	39,91152
366	2/2/2008	13:36:23	116,44449	39,91152
366	2/2/2008	13:36:28	116,44449	39,91152
366	2/2/2008	13:36:33	116,44449	39,91152
366	2/2/2008	13:36:38	116,44449	39,91152
366	2/2/2008	13:36:43	116,44449	39,91152
366	2/2/2008	13:36:48	116,4445	39,91153
366	2/2/2008	13:36:53	116,44444	39,91165
366	2/2/2008	13:36:58	116,44446	39,91194
366	2/2/2008	13:37:03	116,4444	39,91243
366	2/2/2008	13:37:08	116,44438	39,91303
366	2/2/2008	13:37:13	116,44441	39,9137
366	2/2/2008	13:37:18	116,44439	39,91448
366	2/2/2008	13:37:28	116,44443	39,91585
366	2/2/2008	13:37:33	116,44442	39,91598
366	2/2/2008	13:37:38	116,44439	39,91591
366	2/2/2008	13:37:43	116,44439	39,9159
366	2/2/2008	13:37:48	116,44439	39,9159
366	2/2/2008	13:37:53	116,44438	39,91589
366	2/2/2008	13:37:58	116,44436	39,91588

366	2/2/2008	13:38:03	116,44435	39,91588
366	2/2/2008	13:38:08	116,44435	39,91588
366	2/2/2008	13:38:13	116,44437	39,91597
366	2/2/2008	13:38:18	116,44438	39,91629
366	2/2/2008	13:38:23	116,44444	39,91669
366	2/2/2008	13:38:28	116,44446	39,91709
366	2/2/2008	13:38:33	116,44446	39,9174
366	2/2/2008	13:38:38	116,44472	39,9177
366	2/2/2008	13:38:43	116,44481	39,91774
366	2/2/2008	13:38:48	116,44496	39,91769
366	2/2/2008	13:38:53	116,44514	39,91774
366	2/2/2008	13:38:58	116,44536	39,91776
366	2/2/2008	13:39:03	116,4458	39,9179
366	2/2/2008	13:39:08	116,44603	39,91775
366	2/2/2008	13:39:13	116,44598	39,91778
366	2/2/2008	13:39:18	116,44625	39,91752
366	2/2/2008	13:39:23	116,44686	39,91706
366	2/2/2008	13:39:28	116,44686	39,91706
366	2/2/2008	13:39:33	116,44686	39,91706
366	2/2/2008	13:39:38	116,44686	39,91706
366	2/2/2008	13:39:43	116,44686	39,91706
366	2/2/2008	13:39:48	116,44686	39,91706
366	2/2/2008	13:39:53	119,41647	32,36623
366	2/2/2008	13:39:53	119,41647	32,36623
366	2/2/2008	13:39:58	116,44686	39,91706

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΡΑΡΤΗΜΑ 2

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Longitude_366-Model_1	0	-17,596	61,119	17	,000	0
Latitude_366-Model_2	0	-16,950	36,627	17	,004	0
Longitude_1466-Model_3	0	-11,143	16,529	17	,487	0
Latitude_1466-Model_4	0	-14,027	13,763	17	,684	0
Longitude_1566-Model_5	0	-12,143	6,465	17	,989	0
Latitude_1566-Model_6	0	-11,081	14,169	17	,655	0
Longitude_1796-Model_7	0	-13,707	24,054	17	,118	0
Latitude_1796-Model_8	0	-14,273	7,763	17	,971	0
Longitude_2126-Model_9	0	-11,835	23,568	17	,132	0
Latitude_2126-Model_10	0	-12,814	10,462	17	,883	0
Longitude_2146-Model_11	0	-10,762	25,235	17	,090	0
Latitude_2146-Model_12	0	-9,973	31,801	17	,016	0
Longitude_2486-Model_13	0	-12,347	11,822	17	,811	0
Latitude_2486-Model_14	0	-12,147	21,149	17	,220	0
Longitude_3866-Model_15	0	-13,501	17,414	17	,427	0
Latitude_3866-Model_16	0	-13,065	11,822	17	,811	0
Longitude_4346-Model_17	0	-12,572	6,454	17	,990	0
Latitude_4346-Model_18	0	-11,655	18,009	17	,388	0
Longitude_1336-Model_19	0	-10,054	48,265	17	,000	0
Latitude_1336-Model_20	0	-11,361	7,753	17	,972	0

Πίνακας 0-1: Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(0,1,1)

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Longitude_366-Model_1	0	-17,981	19,779	17	,286	0
Latitude_366-Model_2	0	-17,281	8,970	17	,941	0
Longitude_1466-Model_3	0	-11,174	17,607	17	,414	0
Latitude_1466-Model_4	0	-14,039	13,112	17	,729	0
Longitude_1566-Model_5	0	-12,144	6,603	17	,988	0
Latitude_1566-Model_6	0	-11,083	14,405	17	,638	0
Longitude_1796-Model_7	0	-13,719	23,781	17	,126	0
Latitude_1796-Model_8	0	-14,280	7,093	17	,982	0
Longitude_2126-Model_9	0	-11,858	23,694	17	,128	0
Latitude_2126-Model_10	0	-12,816	10,047	17	,902	0
Longitude_2146-Model_11	0	-10,829	25,229	17	,090	0
Latitude_2146-Model_12	0	-10,151	25,483	17	,084	0
Longitude_2486-Model_13	0	-12,353	11,710	17	,817	0
Latitude_2486-Model_14	0	-12,148	21,580	17	,201	0
Longitude_3866-Model_15	0	-13,509	16,933	17	,459	0
Latitude_3866-Model_16	0	-13,068	10,573	17	,878	0
Longitude_4346-Model_17	0	-12,583	5,431	17	,996	0
Latitude_4346-Model_18	0	-11,827	5,835	17	,994	0
Longitude_1336-Model_19	0	-10,138	24,533	17	,106	0
Latitude_1336-Model_20	0	-11,358	8,383	17	,958	0

Πίνακας 0-2: Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(1,1,0)

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Longitude_366-Model_1	0	-17,629	38,528	16	,001	0
Latitude_366-Model_2	0	-17,058	17,498	16	,354	0
Longitude_1466-Model_3	0	-11,220	9,334	16	,899	0
Latitude_1466-Model_4	0	-13,989	12,643	16	,699	0
Longitude_1566-Model_5	0	-12,090	6,430	16	,983	0
Latitude_1566-Model_6	0	-11,035	11,949	16	,747	0
Longitude_1796-Model_7	0	-13,807	7,152	16	,970	0
Latitude_1796-Model_8	0	-14,219	7,438	16	,964	0
Longitude_2126-Model_9	0	-11,790	23,531	16	,100	0
Latitude_2126-Model_10	0	-12,763	9,233	16	,903	0
Longitude_2146-Model_11	0	-11,116	3,892	16	,999	0
Latitude_2146-Model_12	0	-10,223	10,870	16	,817	0
Longitude_2486-Model_13	0	-12,321	10,775	16	,823	0
Latitude_2486-Model_14	0	-12,103	20,923	16	,181	0
Longitude_3866-Model_15	0	-13,472	15,372	16	,498	0
Latitude_3866-Model_16	0	-13,014	9,906	16	,872	0
Longitude_4346-Model_17	0	-12,527	4,960	16	,996	0
Latitude_4346-Model_18	0	-11,777	4,146	16	,999	0
Longitude_1336-Model_19	0	-10,084	31,673	16	,011	0
Latitude_1336-Model_20	0	-11,305	7,771	16	,955	0

Πίνακας 0-3. Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(0,1,2)**Model Statistics**

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
Longitude_366-Model_1	0	-17,962	15,166	16	,513	0
Latitude_366-Model_2	0	-17,231	9,757	16	,879	0
Longitude_1466-Model_3	0	-11,252	8,931	16	,916	0
Latitude_1466-Model_4	0	-13,992	13,519	16	,635	0
Longitude_1566-Model_5	0	-12,093	6,648	16	,979	0
Latitude_1566-Model_6	0	-11,045	11,991	16	,745	0
Longitude_1796-Model_7	0	-13,814	7,703	16	,957	0
Latitude_1796-Model_8	0	-14,228	7,306	16	,967	0
Longitude_2126-Model_9	0	-11,803	23,902	16	,092	0
Latitude_2126-Model_10	0	-12,761	9,632	16	,885	0
Longitude_2146-Model_11	0	-10,929	19,554	16	,241	0
Latitude_2146-Model_12	0	-10,142	22,841	16	,118	0
Longitude_2486-Model_13	0	-12,321	11,567	16	,773	0
Latitude_2486-Model_14	0	-12,115	22,446	16	,129	0
Longitude_3866-Model_15	0	-13,474	15,635	16	,479	0
Latitude_3866-Model_16	0	-13,018	8,895	16	,918	0
Longitude_4346-Model_17	0	-12,527	5,273	16	,994	0
Latitude_4346-Model_18	0	-11,778	4,419	16	,998	0
Longitude_1336-Model_19	0	-10,167	13,615	16	,627	0
Latitude_1336-Model_20	0	-11,307	7,687	16	,958	0

Πίνακας 0-4. Έλεγχος για το αν είναι λευκός θόρυβος τα σφάλματα ARIMA(2,1,0)

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Longitude_366-Model_1	Longitude_366	No Transformation	Difference	1			
			MA Lag 1	-,700	,074	-9,449	,000
Latitude_366-Model_2	Latitude_366	No Transformation	Difference	1			
			MA Lag 1	-,580	,083	-7,003	,000
Longitude_1466-Model_3	Longitude_1466	No Transformation	Difference	1			
			MA Lag 1	-,157	,100	-1,577	,118
Latitude_1466-Model_4	Latitude_1466	No Transformation	Difference	1			
			MA Lag 1	-,211	,100	-2,116	,037
Longitude_1566-Model_5	Longitude_1566	No Transformation	Difference	1			
			MA Lag 1	-,086	,101	-,851	,397
Latitude_1566-Model_6	Latitude_1566	No Transformation	Difference	1			
			MA Lag 1	-,080	,101	-,792	,430
Longitude_1796-Model_7	Longitude_1796	No Transformation	Difference	1			
			MA Lag 1	-,089	,101	-,886	,378
Latitude_1796-Model_8	Latitude_1796	No Transformation	Difference	1			
			MA Lag 1	-,234	,098	-2,377	,019
Longitude_2126-Model_9	Longitude_2126	No Transformation	Difference	1			
			MA Lag 1	-,366	,094	-3,878	,000
Latitude_2126-Model_10	Latitude_2126	No Transformation	Difference	1			
			MA Lag 1	-,123	,101	-1,220	,226
Longitude_2146-Model_11	Longitude_2146	No Transformation	Difference	1			
			MA Lag 1	-,598	,082	-7,328	,000
Latitude_2146-Model_12	Latitude_2146	No Transformation	Difference	1			
			MA Lag 1	-,345	,095	-3,621	,000
Longitude_2486-Model_13	Longitude_2486	No Transformation	Difference	1			
			MA Lag 1	-,112	,100	-1,111	,269
Latitude_2486-Model_14	Latitude_2486	No Transformation	Difference	1			
			MA Lag 1	-,048	,101	-,477	,634
Longitude_3866-Model_15	Longitude_3866	No Transformation	Difference	1			
			MA Lag 1	-,134	,100	-1,343	,182
Latitude_3866-Model_16	Latitude_3866	No Transformation	Difference	1			
			MA Lag 1	-,142	,100	-1,419	,159
Longitude_4346-Model_17	Longitude_4346	No Transformation	Difference	1			
			MA Lag 1	-,247	,098	-2,521	,013
Latitude_4346-Model_18	Latitude_4346	No Transformation	Difference	1			
			MA Lag 1	-,541	,089	-6,100	,000
Longitude_1336-Model_19	Longitude_1336	No Transformation	Difference	1			
			MA Lag 1	-,290	,097	-2,994	,003
Latitude_1336-Model_20	Latitude_1336	No Transformation	Difference	1			
			MA Lag 1	-,245	,109	-2,261	,026

Πίνακας 0-5. Έλεγχος των όρων μέσω t-test ARIMA(0,1,1)

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Longitude_366-Model_1	Longitude_366	No Transformation	AR Lag 1	,807	,063	12,842	,000
			Difference	1			
Latitude_366-Model_2	Latitude_366	No Transformation	AR Lag 1	,745	,067	11,140	,000
			Difference	1			
Longitude_1466-Model_3	Longitude_1466	No Transformation	AR Lag 1	,262	,097	2,690	,008
			Difference	1			
Latitude_1466-Model_4	Latitude_1466	No Transformation	AR Lag 1	,252	,098	2,582	,011
			Difference	1			
Longitude_1566-Model_5	Longitude_1566	No Transformation	AR Lag 1	,096	,101	,955	,342
			Difference	1			
Latitude_1566-Model_6	Latitude_1566	No Transformation	AR Lag 1	,099	,101	,986	,327
			Difference	1			
Longitude_1796-Model_7	Longitude_1796	No Transformation	AR Lag 1	,160	,100	1,606	,111
			Difference	1			
Latitude_1796-Model_8	Latitude_1796	No Transformation	AR Lag 1	,253	,098	2,591	,011
			Difference	1			
Longitude_2126-Model_9	Longitude_2126	No Transformation	AR Lag 1	,404	,093	4,354	,000
			Difference	1			
Latitude_2126-Model_10	Latitude_2126	No Transformation	AR Lag 1	,140	,101	1,390	,168
			Difference	1			
Longitude_2146-Model_11	Longitude_2146	No Transformation	AR Lag 1	,665	,075	8,887	,000
			Difference	1			
Latitude_2146-Model_12	Latitude_2146	No Transformation	AR Lag 1	,566	,083	6,820	,000
			Difference	1			
Longitude_2486-Model_13	Longitude_2486	No Transformation	AR Lag 1	,146	,100	1,462	,147
			Difference	1			
Latitude_2486-Model_14	Latitude_2486	No Transformation	AR Lag 1	,063	,101	,623	,534
			Difference	1			
Longitude_3866-Model_15	Longitude_3866	No Transformation	AR Lag 1	,175	,099	1,762	,081
			Difference	1			
Latitude_3866-Model_16	Latitude_3866	No Transformation	AR Lag 1	,163	,100	1,634	,105
			Difference	1			
Longitude_4346-Model_17	Longitude_4346	No Transformation	AR Lag 1	,280	,097	2,893	,005
			Difference	1			
Latitude_4346-Model_18	Latitude_4346	No Transformation	AR Lag 1	,662	,075	8,804	,000
			Difference	1			
Longitude_1336-Model_19	Longitude_1336	No Transformation	AR Lag 1	,435	,091	4,796	,000
			Difference	1			
Latitude_1336-Model_20	Latitude_1336	No Transformation	AR Lag 1	,233	,110	2,129	,036
			Difference	1			

Πίνακας 0-6. Έλεγχος των όρων μέσω t-test ARIMA(1,1,0)

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Longitude_366-Model_1	Longitude_366	No Transformation	Difference	1			
			MA Lag 1	-.806	,101	-7,980	,000
			Lag 2	-.238	,101	-2,352	,021
Latitude_366-Model_2	Latitude_366	No Transformation	Difference	1			
			MA Lag 1	-.675	,096	-7,052	,000
			Lag 2	-.348	,096	-3,625	,000
Longitude_1466-Model_3	Longitude_1466	No Transformation	Difference	1			
			MA Lag 1	-.082	,094	-.863	,390
			Lag 2	-.373	,095	-3,947	,000
Latitude_1466-Model_4	Latitude_1466	No Transformation	Difference	1			
			MA Lag 1	-.241	,101	-2,379	,019
			Lag 2	-.146	,107	-1,363	,176
Longitude_1566-Model_5	Longitude_1566	No Transformation	Difference	1			
			MA Lag 1	-.074	,101	-.729	,468
			Lag 2	-.062	,101	-.607	,545
Latitude_1566-Model_6	Latitude_1566	No Transformation	Difference	1			
			MA Lag 1	-.063	,101	-.621	,536
			Lag 2	-.090	,101	-.887	,377
Longitude_1796-Model_7	Longitude_1796	No Transformation	Difference	1			
			MA Lag 1	-.105	,094	-1,109	,270
			Lag 2	-.376	,095	-3,974	,000
Latitude_1796-Model_8	Latitude_1796	No Transformation	Difference	1			
			MA Lag 1	-.237	,101	-2,338	,021
			Lag 2	-.045	,101	-.448	,655
Longitude_2126-Model_9	Longitude_2126	No Transformation	Difference	1			
			MA Lag 1	-.399	,101	-3,929	,000
			Lag 2	-.083	,102	-.811	,419
Latitude_2126-Model_10	Latitude_2126	No Transformation	Difference	1			
			MA Lag 1	-.145	,102	-1,421	,158
			Lag 2	-.078	,102	-.760	,449
Longitude_2146-Model_11	Longitude_2146	No Transformation	Difference	1			
			MA Lag 1	-1,081	,076	-14,201	,000
			Lag 2	-.681	,076	-8,914	,000
Latitude_2146-Model_12	Latitude_2146	No Transformation	Difference	1			
			MA Lag 1	-.511	,089	-5,765	,000
			Lag 2	-.504	,089	-5,666	,000
Longitude_2486-Model_13	Longitude_2486	No Transformation	Difference	1			
			MA Lag 1	-.121	,100	-1,207	,231
			Lag 2	-.176	,100	-1,754	,083
Latitude_2486-Model_14	Latitude_2486	No Transformation	Difference	1			
			MA Lag 1	-.046	,101	-.459	,647
			Lag 2	-.087	,101	-.860	,392
Longitude_3866-Model_15	Longitude_3866	No Transformation	Difference	1			
			MA Lag 1	-.145	,100	-1,444	,152
			Lag 2	-.163	,100	-1,630	,106
Latitude_3866-Model_16	Latitude_3866	No Transformation	Difference	1			
			MA Lag 1	-.144	,101	-1,418	,159
			Lag 2	-.078	,101	-.771	,443
Longitude_4346-Model_17	Longitude_4346	No Transformation	Difference	1			
			MA Lag 1	-.269	,101	-2,659	,009
			Lag 2	-.106	,101	-1,050	,297
Latitude_4346-Model_18	Latitude_4346	No Transformation	Difference	1			
			MA Lag 1	-.668	,098	-6,844	,000
			Lag 2	-.437	,099	-4,402	,000
Longitude_1336-Model_19	Longitude_1336	No Transformation	Difference	1			
			MA Lag 1	-.282	,099	-2,862	,005
			Lag 2	-.249	,099	-2,521	,013
Latitude_1336-Model_20	Latitude_1336	No Transformation	Difference	1			
			MA Lag 1	-.250	,113	-2,218	,029
			Lag 2	-.013	,113	-.116	,908

Πίνακας 0-7. Έλεγχος των όρων μέσω t-test ARIMA(0,1,2)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Longitude_366-Model_1	Longitude_366	No Transformation	AR	Lag 1	,651	,102	6,376	,000
				Lag 2	,195	,103	1,897	,061
				Difference	1			
Latitude_366-Model_2	Latitude_366	No Transformation	AR	Lag 1	,685	,101	6,748	,000
				Lag 2	,079	,101	,776	,440
				Difference	1			
Longitude_1466-Model_3	Longitude_1466	No Transformation	AR	Lag 1	,169	,095	1,780	,078
				Lag 2	,349	,095	3,678	,000
				Difference	1			
Latitude_1466-Model_4	Latitude_1466	No Transformation	AR	Lag 1	,226	,101	2,232	,028
				Lag 2	,103	,108	,949	,345
				Difference	1			
Longitude_1566-Model_5	Longitude_1566	No Transformation	AR	Lag 1	,089	,101	,877	,383
				Lag 2	,073	,101	,725	,470
				Difference	1			
Latitude_1566-Model_6	Latitude_1566	No Transformation	AR	Lag 1	,085	,100	,852	,396
				Lag 2	,135	,101	1,338	,184
				Difference	1			
Longitude_1796-Model_7	Longitude_1796	No Transformation	AR	Lag 1	,101	,094	1,067	,289
				Lag 2	,368	,094	3,905	,000
				Difference	1			
Latitude_1796-Model_8	Latitude_1796	No Transformation	AR	Lag 1	,237	,101	2,336	,022
				Lag 2	,063	,101	,622	,535
				Difference	1			
Longitude_2126-Model_9	Longitude_2126	No Transformation	AR	Lag 1	,390	,102	3,832	,000
				Lag 2	,035	,102	,339	,736
				Difference	1			
Latitude_2126-Model_10	Latitude_2126	No Transformation	AR	Lag 1	,135	,102	1,317	,191
				Lag 2	,044	,103	,427	,670
				Difference	1			
Longitude_2146-Model_11	Longitude_2146	No Transformation	AR	Lag 1	,918	,094	9,775	,000
				Lag 2	-,373	,094	-3,971	,000
				Difference	1			
Latitude_2146-Model_12	Latitude_2146	No Transformation	AR	Lag 1	,443	,099	4,455	,000
				Lag 2	,212	,100	2,132	,036
				Difference	1			
Longitude_2486-Model_13	Longitude_2486	No Transformation	AR	Lag 1	,123	,100	1,229	,222
				Lag 2	,153	,100	1,528	,130
				Difference	1			
Latitude_2486-Model_14	Latitude_2486	No Transformation	AR	Lag 1	,053	,100	,531	,597
				Lag 2	,150	,101	1,490	,140
				Difference	1			
Longitude_3866-Model_15	Longitude_3866	No Transformation	AR	Lag 1	,150	,100	1,489	,140
				Lag 2	,143	,100	1,428	,156
				Difference	1			
Latitude_3866-Model_16	Latitude_3866	No Transformation	AR	Lag 1	,150	,101	1,483	,141
				Lag 2	,076	,101	,754	,452
				Difference	1			
Longitude_4346-Model_17	Longitude_4346	No Transformation	AR	Lag 1	,271	,101	2,668	,009
				Lag 2	,033	,102	,330	,742
				Difference	1			
Latitude_4346-Model_18	Latitude_4346	No Transformation	AR	Lag 1	,718	,101	7,101	,000
				Lag 2	-,084	,101	-,832	,407
				Difference	1			
Longitude_1336-Model_19	Longitude_1336	No Transformation	AR	Lag 1	,310	,097	3,188	,002
				Lag 2	,282	,097	2,895	,005
				Difference	1			
Latitude_1336-Model_20	Latitude_1336	No Transformation	AR	Lag 1	,251	,113	2,233	,028
				Lag 2	-,081	,113	-,720	,473
				Difference	1			

Πίνακας 0-8. Έλεγχος των όρων μέσω t-test ARIMA(2,1,0)

BIBΛΙΟΓΡΑΦΙΑ

- [1]. Brockwell, P. J., & Davis, R. A. (Eds.). (2002). *Introduction to time series and forecasting (Vol. 1)*. Taylor & Francis.
- [2]. Elnekave, S., Last, M., & Maimon, O. (2007). Predicting future locations using clusters' centroids. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* (p. 55). ACM.
- [3]. Google Driving Directions: <http://www.gps-coordinates.net/driving-directions>
- [4]. Jeung, H., Liu, Q., Shen, H. T., & Zhou, X. (2008). A hybrid prediction model for moving objects. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on* (pp. 70-79). IEEE.
- [5]. Latitude, Longitude, And Great Circles: <http://www.cpearson.com/excel/LatLong.aspx>
- [6]. Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 637-646). ACM.
- [7]. Peña, D. (2001). Prediction and model selection. *A Course in Time Series Analysis*, 111-135.
- [8]. Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts (Vol. 4)*. New York: McGraw-Hill.
- [9]. T-Drive trajectory data sample: <http://research.microsoft.com/apps/pubs/?id=152883>
- [10]. Yan, Z. (2010). Traj-ARIMA: a spatial-time series model for network-constrained trajectory. In *Proceedings of the Second International Workshop on Computational Transportation Science* (pp. 11-16). ACM.
- [11]. Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2011). Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 316-324). ACM.
- [12]. Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G., & Huang, Y. (2010). T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 99-108). ACM.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ