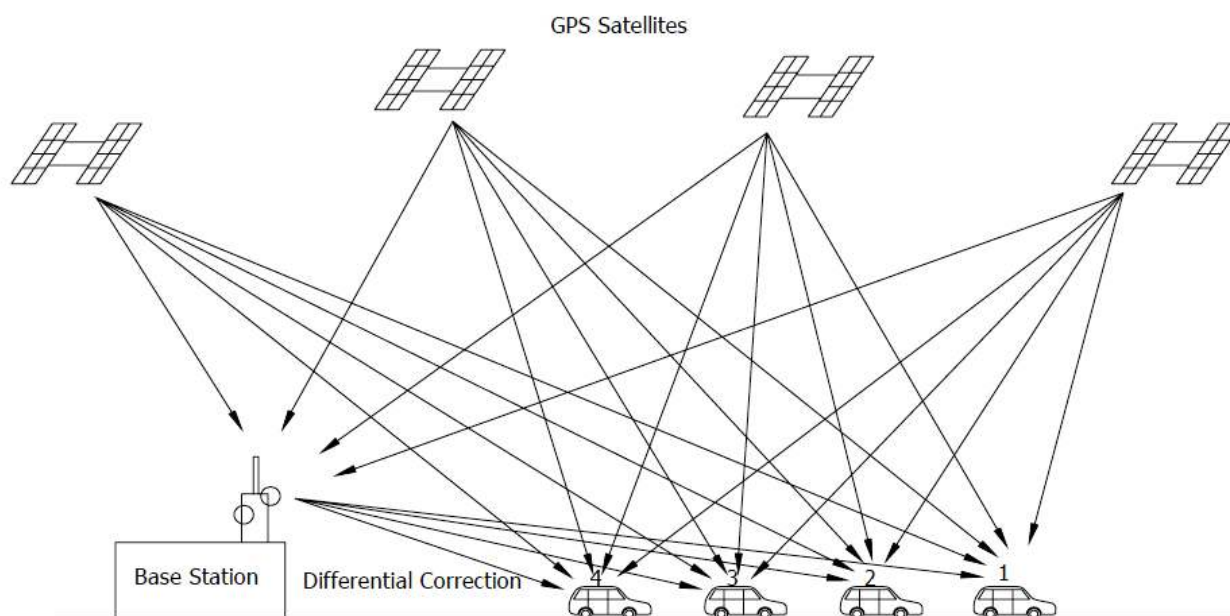




ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΚΤΙΜΗΣΗ ΠΡΟΤΥΠΩΝ ΑΚΟΛΟΥΘΙΑΣ ΟΧΗΜΑΤΩΝ ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



Εκπόνηση: Παπαθανασοπούλου Βασιλεία

Επιβλέπων: Κ. Αντωνίου, Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, 2012

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Αντωνίου Κωνσταντίνο, για την καθοδήγηση και βοήθειά του σε κάθε φάση της μεταπτυχιακής εργασίας μου, καθώς και όλους όσους συνέβαλαν στην ολοκλήρωση αυτής. Τέλος, θα ήθελα να ευχαριστήσω τον κύριο Vincenzo Punzo, επίκουρο καθηγητή του πανεπιστημίου της Νάπολης “Federico II”, για τη διάθεση δεδομένων ενός πειράματος προς ανάλυση σε αυτήν την εργασία.

Αθήνα, Οκτώβριος 2012

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	8
ABSTRACT	9
ΠΡΟΛΟΓΟΣ.....	10
1. ΕΙΣΑΓΩΓΗ	13
2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	16
2.1. Βασικοί ορισμοί.....	16
2.2. Κατηγορίες μοντέλων προσομοίωσης.....	17
2.3. Πρότυπα ακολουθίας οχημάτων (<i>Car-following models</i>).....	19
2.3.1. Γενική περιγραφή προτύπων ακολουθίας οχημάτων	19
2.3.2. Κατηγορίες προτύπων ακολουθίας οχημάτων	20
2.3.3. Το μοντέλο του Gipps.....	22
2.3.4. Εύρη τιμών για τις παραμέτρους του μοντέλου του Gipps	25
2.4. Τεχνολογίες συλλογής δεδομένων.....	30
2.5. Μέθοδοι διόρθωσης δεδομένων	33
3. ΜΕΘΟΔΟΛΟΓΙΑ.....	37
3.1. Γενικό μεθοδολογικό πλαίσιο	37
3.2. Μηχανική μάθηση	39
3.2.1. Εισαγωγή στην ιδέα της μηχανικής μάθησης	39
3.2.2. Τοπικά σταθμισμένη παλινδρόμηση (<i>loess</i>)	40
3.2.3. Ομαδοποίηση (<i>clustering</i>).....	42
3.2.4. Ταξινόμηση (<i>classification</i>)	44
3.2.5. Οφέλη και προκλήσεις.....	45
3.3. Αξιολόγηση μοντέλων.....	46
3.4. Πειραματική διάταξη.....	47
4. ΕΦΑΡΜΟΓΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ.....	48
4.1. Περιγραφή και ακρίβεια δεδομένων.....	49
4.2. Υπολογισμός παραγόμενων στοιχείων.....	50
4.3. Επιλογή δεδομένων για βαθμονόμηση.....	54
4.4. Εφαρμογή του μοντέλου <i>Gipps</i>	56
4.4.1. Ανάλυση ευαισθησίας παραμέτρων	56
4.4.2. Βαθμονόμηση μοντέλου	64
4.4.3. Έλεγχος εγκυρότητας μοντέλου.....	69
4.5. Εφαρμογή τοπικά σταθμισμένης παλινδρόμησης (<i>loess</i>).....	70
4.5.1. Ανάλυση ευαισθησίας τοπικά σταθμισμένης παλινδρόμησης (<i>loess</i>).....	70
4.5.3. Εφαρμογή τοπικά σταθμισμένης παλινδρόμησης (<i>loess</i>) και άμεση σύγκριση με το μοντέλο <i>Gipps</i>	75
4.5.4. Επιλογή διάφορων παραγόντων παλινδρόμησης	83
4.6. Σύγκριση μεθόδων και συμπεράσματα.....	88
5. ΕΦΑΡΜΟΓΗ ΣΕ ΔΕΔΟΜΕΝΑ NGSIM	89
5.1. Περιγραφή και ακρίβεια δεδομένων.....	89

5.2. Εύρη ταχυτήτων και δυνατότητα εκτίμησης	93
5.3. Υπολογισμός παραγόμενων στοιχείων	94
5.4. Εφαρμογή μοντέλου Gijrps και της προτεινόμενης μεθόδου	98
5.5. Αιτιολογία μειωμένης απόδοσης της μεθόδου loess για κάποιες σειρές δεδομένων	104
5.6. Συμπεράσματα	106
6. ΕΦΑΡΜΟΓΗ ΟΜΑΔΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ	108
6.1. Εφαρμογή	108
6.2. Αποτελέσματα και αιτιολόγηση	111
6.3. Συμπεράσματα	117
7. ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	118
7.1. Γενικά Συμπεράσματα	118
7.2. Προτάσεις για μελλοντική έρευνα	119
ΑΝΑΦΟΡΕΣ	122

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1.: Μοντελοποίηση της κυκλοφορίας	17
Σχήμα 2.2.: Ένα ψυχοσωματικό μοντέλο ακολουθίας οχημάτων	22
Σχήμα 2.3.: Οι θέσεις των οχημάτων στο μοντέλο του Gijrps	25
Σχήμα 2.4.: Μέθοδος διαφορικού GPS	32
Σχήμα 3.1.: Μεθοδολογία εκτίμησης προτύπων ακολουθίας οχημάτων με τη χρήση μεθόδων μηχανικής μάθησης.....	39
Σχήμα 3.1: Σκαρίφημα σχετικών αποστάσεων πειράματος	53
Σχήμα 5.1.: Αεροφωτογραφία με τη σχετική θέση της περιοχής μελέτης I-80 και του κτιρίου που ήταν τοποθετημένες οι κάμερες καταγραφής (αριστερό μέρος), σχηματική απεικόνιση των λωρίδων κυκλοφορίας στην περιοχή μελέτης (δεξιό μέρος).....	90
Σχήμα 5.2.: Ψηφιακή βιντεοκάμερα που καταγράφει τις τροχιές των οχημάτων στον αυτοκινητόδρομο I-80 του Σαν Φρανσίσκο από την κορυφή ενός κτιρίου	91
Σχήμα 5.3.: Σκαρίφημα σχετικών αποστάσεων μεταξύ των οχημάτων για τα δεδομένα NGSIM	97

ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ

Πίνακας 2.1.: Εύρη παραμέτρων του μοντέλου του Gijrps	26
Πίνακας 2.2.: Προτεινόμενα εύρη παραμέτρων από τους Ranjitkar et al. (2005).....	26
Πίνακας 2.3.: Επιλεγόμενα εύρη παραμέτρων	30
Πίνακας 4.1.: Ονομασία και χρονική διάρκεια σειρών δεδομένων.....	54

Πίνακας 4.2.: Επιλεγόμενες τιμές παραμέτρων για την εφαρμογή του μοντέλου του Gırps στη σειρά δεδομένων B1695	66
Πίνακας 4.3.: Βαθμονομημένες τιμές για τις παραμέτρους του Gırps σύμφωνα με τους Ciuffo et al., 2012.....	67
Πίνακας 4.4.: Πιθανοί συνδυασμοί παραμέτρων για την εφαρμογή του μοντέλου του Gırps και το αντίστοιχο σφάλμα RMSN.....	68
Πίνακας 4.5.: Έλεγχος εγκυρότητας του μοντέλου του Gırps για τους προτεινόμενους συνδυασμούς παραμέτρων για όλες τις σειρές δεδομένων	69
Πίνακας 4.6.: Ποσοστά απόκλισης σφάλματος μεταξύ της βέλτιστης και της επιλεγόμενης τιμής span για κάθε σειρά δεδομένων	73
Πίνακας 4.7.: Σύγκριση μοντέλου Gırps (g1) και μεθόδου loess (l1f) για χρόνο αντίδρασης $\tau=0.4$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN	75
Πίνακας 4.8.: Σύγκριση μοντέλου Gırps (G3) και μεθόδου loess (l2f) για χρόνο αντίδρασης $\tau=1.0$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN	76
Πίνακας 4.9.: Σύγκριση μεθόδου loess (l1s) με μοντέλο Gırps (g1) χρησιμοποιώντας λιγότερους παράγοντες για χρόνο αντίδρασης $\tau=0.4$ s	81
Πίνακας 4.10.: Σύγκριση μεθόδου loess (l2s) με μοντέλο Gırps (G3) χρησιμοποιώντας λιγότερους παράγοντες για χρόνο αντίδρασης $\tau=1.0$ s	82
Πίνακας 4.11.: Συνδυασμοί διάφορων παραγόντων ανά δυάδες για τη μέθοδο loess και το αντίστοιχο RMSN (%) για τις σειρές δεδομένων B1695, C621.....	84
Πίνακας 4.12.: RMSN (%) για διάφορους συνδυασμούς τριών παραγόντων στην εφαρμογή της μεθόδου loess	87
Πίνακας 5.1.: Δεδομένα NGSIM που θα χρησιμοποιηθούν για τον έλεγχο εγκυρότητας των μοντέλων	93
Πίνακας 5.2.: Σύγκριση RMSN (%) του μοντέλου Gırps και της μεθόδου loess για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=0.4$ s	99
Πίνακας 5.3.: Σύγκριση RMSN (%) του μοντέλου Gırps και της μεθόδου loess για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=1.0$ s	100
Πίνακας 6.1.: Αποτελέσματα εκτίμησης της ταχύτητας v_3 για όλες τις σειρές δεδομένων χρησιμοποιώντας ένα συνδυασμό μεθόδων μηχανικής μάθησης.....	112

ΠΙΝΑΚΑΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 2.1.: Εύρη παραμέτρων του μοντέλου του Gırps σύμφωνα με βιβλιογραφία	27
Διάγραμμα 2.2.: Εύρη τιμών για τον χρόνο αντίδρασης των οδηγών σύμφωνα με τη βιβλιογραφία..	29
Διάγραμμα 4.1.: Απεικόνιση της διανυόμενης απόστασης των τεσσάρων οχημάτων σε σχέση με το χρόνο	51
Διάγραμμα 4.2.: Εύρη τιμών της ταχύτητας v_3 για κάθε σειρά δεδομένων	54

Διάγραμμα 4.3.: Μη δυνατότητα εκτίμησης ταχυτήτων πέρα από αυτές που περιλαμβάνονται στο μοντέλο βαθμονόμησης.....	55
Διάγραμμα 4.4.: Ανάλυση ευαισθησίας του χρόνου αντίδρασης και του σφάλματος RMSN για τη σειρά δεδομένων B1695	57
Διάγραμμα 4.5.: Ανάλυση ευαισθησίας της επιθυμητής ταχύτητας και του σφάλματος RMSN για τη σειρά δεδομένων B1695	58
Διάγραμμα 4.6.: Ανάλυση ευαισθησίας της απόστασης και του σφάλματος RMSN για τη σειρά δεδομένων B1695	59
Διάγραμμα 4.7.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695.....	60
Διάγραμμα 4.8.: Ανάλυση ευαισθησίας της μέγιστης επιβράδυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695	60
Διάγραμμα 4.9.: Ανάλυση ευαισθησίας της εκτιμώμενης μέγιστης επιβράδυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695.....	61
Διάγραμμα 4.10.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695(επανεξέταση)	62
Διάγραμμα 4.11.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695(επανεξέταση)	62
Διάγραμμα 4.12: Επιλογή τιμής για την μέγιστη επιθυμητή επιτάχυνση	63
Διάγραμμα 4.13.: Επιλογή τιμής της παραμέτρου V_n (α' εκδοχή)	64
Διάγραμμα 4.14.: Επιλογή τιμής της παραμέτρου V (β' εκδοχή)	65
Διάγραμμα 4.15.: Επιρροή του RMSN για διάφορες τιμές span και degree, κατά την εφαρμογή της μεθόδου loess για χρόνο αντίδρασης $\tau=0.4$ s	71
Διάγραμμα 4.16.: Επιρροή του RMSN για διάφορες τιμές span και degree, κατά την εφαρμογή της μεθόδου loess για χρόνο αντίδρασης $\tau=1.0$ s	72
Διάγραμμα 4.17: Εύρη τιμών span για τις οποίες παρατηρείται το μικρότερο σφάλμα RMSN για κάθε σειρά δεδομένων	73
Διάγραμμα 4.18.: Μεταβολή RMSN ανάλογα με το βέλτιστο. το χειρότερο και το γενικά επιλεγόμενο span για κάθε σειρά δεδομένων	74
Διάγραμμα 4.19.: Σύγκριση RMSN του μοντέλου G1pps (g1) και της μεθόδου loess (l1f) για χρόνο αντίδρασης $\tau=0.4$ s	76
Διάγραμμα 4.20.: Σύγκριση μοντέλου G3pps (G3) και μεθόδου loess (l2f) για χρόνο αντίδρασης $\tau=1.0$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN	77
Διάγραμμα 4.21.: Σύγκριση της μεθόδου loess (l1f) και του μοντέλου G1pps (g1) για όλες τις σειρές δεδομένων με χρήση διάφορων μέτρων αξιολόγησης για χρόνο αντίδρασης $\tau=0.4$ s	79
Διάγραμμα 4.22.: Σύγκριση της μεθόδου loess (l2f) και του μοντέλου G3pps (G3) για όλες τις σειρές δεδομένων με χρήση διάφορων μέτρων αξιολόγησης για χρόνο αντίδρασης $\tau=1.0$	80

Διάγραμμα 4.23.: Σύγκριση RMSN του μοντέλου Gipps, της μεθόδου loess (v_3+v_2+D23)- I1f και της μεθόδου loess (v_3+v_2)- I1s για χρόνο αντίδρασης $\tau=0.4$ s.....	82
Διάγραμμα 4.24.: Σύγκριση RMSN του μοντέλου Gipps (G3), της μεθόδου loess (v_3+v_2+D23)- I2f και της μεθόδου loess (v_3+v_2)- I2s για χρόνο αντίδρασης $\tau=1.0$ s	83
Διάγραμμα 4.16.:RMSN (%) για διάφορους διπλούς συνδυασμούς παραγόντων στην εφαρμογή της μεθόδου loess	85
Διάγραμμα 4.27.: RMSN (%) για διάφορους διπλούς συνδυασμούς παραγόντων στην εφαρμογή της μεθόδου loess για κάθε σειρά δεδομένων	86
Διάγραμμα 4.28.: Απεικόνιση του σφάλματος RMSN (%) για διάφορους συνδυασμούς τριών παραγόντων στην εφαρμογή της μεθόδου loess.....	87
Διάγραμμα 5.1.: Εύρη ταχυτήτων για τα δεδομένα NGSIM	94
Διάγραμμα 5.2.: Σύγκριση RMSN (%) για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=0.4$ s.....	98
Διάγραμμα 5.3.: Σύγκριση RMSN (%) για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=1.0$ s.....	99
Διάγραμμα 5.4.: Σύγκριση του μοντέλου Gipps και της μεθόδου loess για τα NGSIM δεδομένα με τη χρήση των μέτρων αξιολόγησης RMSN, RMSPE, MPE,U	102
Διάγραμμα 5.5.: Σύγκριση του μοντέλου Gipps και της μεθόδου loess για τα NGSIM δεδομένα με τη χρήση των μέτρων αξιολόγησης U_M, U_S, U_C	103
Διάγραμμα 5.6.: Προφίλ ταχυτήτων για τη σειρά δεδομένων B1695 από την οποία προκύπτει το πρότυπο ακολουθίας οχημάτων με τη μέθοδο loess	104
Διάγραμμα 5.7.: Προφίλ ταχυτήτων για σειρές δεδομένων από τα NGSIM δεδομένα για τα οποία παρατηρήθηκε καλή εκτίμηση με τη μέθοδο loess	105
Διάγραμμα 5.8.: Προφίλ ταχυτήτων για σειρές δεδομένων από τα NGSIM δεδομένα για τα οποία δεν παρατηρήθηκε καλή εκτίμηση με τη μέθοδο loess	105
Διάγραμμα 5.9.: Συχνότητα εμφάνισης τιμών της προς εκτίμηση ταχύτητας v_3 για κάθε σειρά δεδομένων	106
Διάγραμμα 6.1.: Επιλογή βέλτιστου αριθμού κλάσεων	109
Διάγραμμα 6.2.: Απεικόνιση των αποτελεσμάτων της ομαδοποίησης για διάφορους αριθμούς κλάσεων	110
Διάγραμμα 6.3.: RMSN (%) ανάλογα με τον αριθμό κλάσεων για τη σειρά δεδομένων B1695	113
Διάγραμμα 6.4.: Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695(αριστερό μέρος), C621(δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 4.....	114
Διάγραμμα 6.5. : Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695(αριστερό μέρος), C621(δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 4.....	115
Διάγραμμα 6.6. : Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695(αριστερό μέρος), C621(δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 9.....	115

ΠΕΡΙΛΗΨΗ

Στα μέσα της δεκαετίας του '50 άρχισαν οι πρώτες έρευνες που αφορούσαν τα πρότυπα ακολουθίας οχημάτων και από τότε έχουν αναπτυχθεί πολυάριθμα μοντέλα. Οι τεχνολογικές εξελίξεις έχουν συμβάλει σημαντικά στη συλλογή υψηλής ακρίβειας δεδομένων για τροχιές οχημάτων, καθιστώντας έτσι εφικτή την πρόσβαση σε αυτού του είδους τα δεδομένα με σκοπό την ανάλυση τους για τη μελέτη της οδηγικής συμπεριφοράς και της καλύτερης δυνατής προσομοίωσής της. Στο πλαίσιο αυτής της εργασίας για την εκτίμηση προτύπων ακολουθίας οχημάτων προτείνεται μια εναλλακτική μεθοδολογία βασισμένη σε τεχνικές μηχανικής μάθησης όπως η τοπικά σταθμισμένη παλινδρόμηση, η ομαδοποίηση και η ταξινόμηση. Η προτεινόμενη μεθοδολογία εφαρμόζεται σε δεδομένα από ένα πείραμα που διεξήχθη στη Νάπολη της Ιταλίας, καθώς και στα NGSIM δεδομένα. Στα ίδια δεδομένα, επίσης, εφαρμόζεται το μοντέλο του Gipps, ένα από τα ευρέως χρησιμοποιούμενα μοντέλα ακολουθίας οχημάτων, αφού όμως προηγηθεί μια ανάλυση ευαισθησίας των παραμέτρων του και βαθμονόμηση του. Στη συνέχεια ακολουθεί σύγκριση μεταξύ της προτεινόμενης μεθοδολογίας και του μοντέλου του Gipps. Τα μοντέλα ακολουθίας οχημάτων που προκύπτουν από μαθηματικούς τύπους εξηγούν σαφώς τις σχέσεις που λαμβάνονται υπόψη, αλλά είναι περισσότερο αυστηρά στην εφαρμογή τους και επιβάλλουν περισσότερους περιορισμούς. Από την άλλη πλευρά, οι μέθοδοι μηχανικής μάθησης παρά το γεγονός ότι αδυνατούν να εξηγήσουν ποιοτικά τις συσχετίσεις που μοντελοποιούν, παρουσιάζουν μεγαλύτερη ευελιξία και μπορούν να παρέχουν καλύτερα αποτελέσματα στην εκτίμηση των προτύπων ακολουθίας οχημάτων.

ABSTRACT

Since the 1950' s car following models are well researched with varying approaches. Nowadays, technological advances have significantly contributed in data collection with high accuracy. Therefore, the availability of traffic data has improved and researchers could focus on their analysis in order to achieve a better simulation of drivers' behavior. In this study, an alternative methodology framework based on computational approaches is proposed for data driven estimation of car following models. Machine learning techniques such as locally weighted regression, classification and clustering are included. The proposed methodology is illustrated using both data from an experiment implemented in Naples, in Italy, and NGSIM data. The Gipps' model, one of the most extensively used car following models, is also applied to the same data, after a sensitivity analysis of its parameters and the calibration of the model are preceded. Then, a comparison between the proposed framework and Gipps' model is performed. Typical car following models are relied on a mathematical formula and are theoretically justified, though they are more restrictive. On the other hand, machine learning approaches may not provide as much insight into traffic flow theory as the traditional models do, though they are more flexible and allow the incorporation of additional information to the process of speed estimation. Machine learning methods could ensure reliability and improvement in data driven estimation of car following models.

ΠΡΟΛΟΓΟΣ

Στα μέσα της δεκαετίας του '50 άρχισαν οι πρώτες έρευνες που αφορούσαν τα πρότυπα ακολουθίας οχημάτων και από τότε έχουν αναπτυχθεί πολυάριθμα μοντέλα (Brackstone et al., 1999). Η ζήτηση των μοντέλων αυτών είναι αυξημένη καθώς αποτελούν το βασικό συνθετικό προγραμμάτων μικροσκοπικής προσομοίωσης που χρησιμοποιούνται ευρέως στον τομέα των μεταφορών για τη διαχείριση, έλεγχο και ασφάλεια της κυκλοφορίας (αναλύσεις χωρητικότητας, ανάλυση ατυχημάτων, επιδράσεις κυκλοφοριακών μελετών, σχεδιασμό διασταυρώσεων, κ.ά.). Τα μοντέλα ακολουθίας οχημάτων μοντελοποιούν τη μικροσκοπική συμπεριφορά των οδηγών και τις αλληλεπιδράσεις μεταξύ των οχημάτων και με αυτόν τον τρόπο συμβάλλουν στο σχεδιασμό συσκευών που βρίσκονται επί των οχημάτων και βοηθούν τους οδηγούς να απολαμβάνουν ασφαλείς και άνετες συνθήκες κυκλοφορίας. Επιπλέον, έχουν τη δυνατότητα να εξετάσουν τις αλληλεπιδράσεις των οδηγών με τις οδικές υποδομές ή ενδεχόμενα πολιτικά και τεχνολογικά μέτρα και να εκτιμηθεί η επιρροή τους. Γενικά η ανάπτυξη των κυκλοφοριακών μοντέλων συμβάλλει σημαντικά στην αξιολόγηση της λειτουργίας των κυκλοφοριακών συστημάτων είτε πρόκειται για τον εντοπισμό ελλείψεων είτε για προβλέψεις που αφορούν νέα δίκτυα μεταφορών. Πρόσφατα η συμβολή των προτύπων ακολουθίας οχημάτων είναι επίσης αξιόλογη στην ανάπτυξη ευφυών συστημάτων μεταφοράς.

Για τη μελέτη μοντέλων ακολουθίας οχημάτων απαραίτητη είναι η συλλογή και χρήση δεδομένων μεγάλης ακρίβειας για τροχιές οχημάτων που κινούνται σε σειρά. Η ραγδαία ανάπτυξη της τεχνολογίας έχει συμβάλει στη διαθεσιμότητα τέτοιων δεδομένων υψηλής ποιότητας. Πολλές έρευνες με σκοπό τη μελέτη μοντέλων ακολουθίας οχημάτων έχουν χρησιμοποιήσει δεδομένα που συλλέχθηκαν με το διαφορικό σύστημα εντοπισμού θέσης - διαφορικό GPS (DGPS - Differential GPS) ή την τεχνική κινηματικού προσδιορισμού σε πραγματικό χρόνο (RTK – Real Time Kinematic), μέθοδοι γνωστές για την υψηλή ακρίβεια που παρέχουν. Επίσης, δεδομένα με τροχιές οχημάτων διατίθενται από το πρόγραμμα NGSIM, καθιστώντας έτσι εφικτή την πρόσβαση σε αυτού του είδους τα δεδομένα με σκοπό την ανάλυση τους για τη μελέτη της οδηγικής συμπεριφοράς και της καλύτερης δυνατής προσομοίωσής της.

Η εξέλιξη της τεχνολογίας και των υπολογιστών έχει συμβάλλει επίσης στην ανάπτυξη των μεθόδων μηχανικής μάθησης, η χρήση των οποίων διευρύνεται συνεχώς σε πολλά επιστημονικά πεδία. Με τις μεθόδους μηχανικής μάθησης το σύστημα ανακαλύπτει συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων δημιουργώντας πρότυπα, χωρίς όμως να είναι σαφείς οι συσχετίσεις που προκύπτουν. Στον τομέα των μεταφορών οι μέθοδοι μηχανικής μάθησης έχουν χρησιμοποιηθεί σε αρκετές εφαρμογές όπως την εκτίμηση ή πρόβλεψη ταχυτήτων και την ταξινόμηση κυκλοφοριακών δεδομένων. Η χρήση τους όμως στον τομέα αυτόν παραμένει περιορισμένη, γεγονός το οποίο πιθανόν να οφείλεται στο ότι οι μέθοδοι αυτοί αποτελούν «μαύρο κουτί» για την επιστημονική κοινότητα. Ενώ οι μέθοδοι μηχανικής μάθησης παρέχουν καλά αποτελέσματα, είναι δυσνόητες ως προς τον τρόπο που τα εξάγουν και είναι αδύνατο να επικυρωθούν από εμπειρογνώμονες που βασίζονται στην εμπειρία και στις θεωρητικές γνώσεις τους.

Στο πλαίσιο αυτής της εργασίας προτείνεται μια εναλλακτική μεθοδολογία βασισμένη σε τεχνικές μηχανικής μάθησης με σκοπό την ανάπτυξη αξιόπιστων προτύπων ακολουθίας οχημάτων. Αρχικά παρουσιάζεται μια βιβλιογραφική ανασκόπηση για τα πρότυπα ακολουθίας οχημάτων που έχουν αναπτυχθεί κατά καιρούς, δίνοντας έμφαση στο μοντέλο του Gipps και τα εύρη των παραμέτρων του, καθώς χρησιμοποιείται σαν μοντέλο αναφοράς για σύγκριση με την προτεινόμενη μεθοδολογία που ακολουθεί στη συνέχεια της εργασίας. Επίσης, αναφέρονται οι τεχνολογίες που εφαρμόζονται για τη συλλογή δεδομένων, καθώς και οι μέθοδοι επεξεργασίας τους για τον περιορισμό των σφαλμάτων. Στη συνέχεια αναλύεται η προτεινόμενη μεθοδολογία που περιλαμβάνει ένα συνδυασμό μεθόδων μηχανικής μάθησης, όπως η τοπικά σταθμισμένη παλινδρόμηση, η ομαδοποίηση και η ταξινόμηση. Έπειτα, το μοντέλο του Gipps και η προτεινόμενη μεθοδολογία εφαρμόζεται σε μια σειρά δεδομένων από ένα πείραμα που διεξήχθη στην Ιταλία και εξετάζεται η εγκυρότητα τους στις υπόλοιπες σειρές δεδομένων του ίδιου πειράματος. Κατόπιν, η εγκυρότητα των δύο προσεγγίσεων ελέγχεται επίσης σε δεδομένα από το πρόγραμμα NGSIM και τέλος προκύπτουν τα γενικά συμπεράσματα από την παραπάνω ανάλυση.

Τα μοντέλα ακολουθίας οχημάτων που προκύπτουν από μαθηματικούς τύπους εξηγούν σαφώς τις σχέσεις που λαμβάνονται υπόψη, αλλά είναι περισσότερο αυστηρά στην εφαρμογή τους και επιβάλλουν περισσότερους περιορισμούς. Οι μέθοδοι

μηχανικής μάθησης παρά το γεγονός ότι αδυνατούν να εξηγήσουν ποιοτικά τις συσχετίσεις που μοντελοποιούν, παρέχουν καλύτερα αποτελέσματα από τα παραδοσιακά μοντέλα ακολουθίας οχημάτων και παρουσιάζουν μεγαλύτερη ευελιξία στη διαχείριση των δεδομένων. Επομένως, πρότυπα ακολουθίας οχημάτων που προκύπτουν με τη χρήση μεθόδων μηχανικής μάθησης θα μπορούσαν να χρησιμοποιηθούν για την ανάπτυξη μοντέλων προσομοίωσης που θα αναπαριστούν καλύτερα τη συμπεριφορά των οδηγών και γενικότερα την κίνηση των οχημάτων.

1. ΕΙΣΑΓΩΓΗ

Οι ολοένα αυξανόμενες απαιτήσεις για μεγαλύτερη ακρίβεια και ευελιξία στη προσομοίωση της οδηγικής συμπεριφοράς έχουν οδηγήσει στην ανάπτυξη πολλών προτύπων ακολουθίας οχημάτων κατά τις τελευταίες δεκαετίες. Αρκετά από αυτά αδυνατούν να παρέχουν μια αξιόπιστη εκτίμηση χωρίς να βασιστούν στην υπερπροσαρμογή των παραμέτρων τους (Ciuffo et al., 2012). Επιπλέον, η εκτίμηση της συμπεριφοράς των οδηγών σε απρόοπτες κυκλοφοριακές συνθήκες είναι ακόμα δυσκολότερη. Οι έρευνες συνεχίζονται έτσι ώστε τα μοντέλα προσομοίωσης με τη χρήση πρότυπων ακολουθίας οχημάτων να μπορούν να αναπαραστήσουν όσο το δυνατόν καλύτερα τη συμπεριφορά των οδηγών στο πραγματικό περιβάλλον.

Παράλληλα, η ανάπτυξη των μεθόδων μηχανικής μάθησης και η χρήση τους στον τομέα των μεταφορών για ταξινόμηση δεδομένων και για εκτίμηση ταχυτήτων προσφέρουν μεγαλύτερη ευελιξία, δίνοντας τη δυνατότητα εξέτασης περισσότερων παραμέτρων από αυτών που συμμετέχουν στους μαθηματικούς τύπους των διάφορων μοντέλων ή ακόμα διαφορετικών παραμέτρων ανάλογα με τα δεδομένα που διατίθενται. Ωστόσο, οι μέθοδοι αυτές στερούνται θεωρητικής τεκμηρίωσης και επεξήγησης. Δηλαδή οι μέθοδοι μηχανικής μάθησης είναι πιθανό να παρέχουν αξιόπιστα αποτελέσματα, χωρίς όμως να εξηγούν με ποιον ακριβώς τρόπο αυτά προκύπτουν. Το ερώτημα, όμως, που τίθεται είναι αν αυτό που έχει σημασία στη δημιουργία ενός μικρο-προσομοιωτή είναι η αναπαράσταση της συμπεριφοράς των οδηγών με μεγαλύτερη ακρίβεια ή η θεωρητική ανάλυση των προτύπων που χρησιμοποιούνται.

Στο πλαίσιο αυτό, η παρούσα διπλωματική εργασία έχει σαν στόχο την εκτίμηση προτύπων ακολουθίας οχημάτων με τη χρήση μεθόδων μηχανικής μάθησης και τη σύγκριση αυτών με τα παραδοσιακά πρότυπα ακολουθίας οχημάτων που προκύπτουν από μαθηματικούς τύπους. Οι μέθοδοι μηχανικής μάθησης που χρησιμοποιούνται είναι η τοπικά σταθμισμένη παλινδρόμηση, η ομαδοποίηση και η ταξινόμηση ενώ ως μοντέλο αναφοράς για τη δεύτερη κατηγορία προτύπων ακολουθίας οχημάτων χρησιμοποιείται το μοντέλο του Gipps λόγω της ευρείας χρήσης του. Η προτεινόμενη μεθοδολογία αποσκοπεί στην καλύτερη δυνατή αναπαράσταση της συμπεριφοράς των οδηγών.

Η δομή της παρούσας μεταπτυχιακής εργασίας είναι η παρακάτω:

Το **πρώτο κεφάλαιο** περιλαμβάνει τη βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε για τα πρότυπα ακολουθίας οχημάτων, δίνοντας έμφαση στο μοντέλο του Girps και στα εύρη των παραμέτρων του. Επίσης περιγράφονται οι τεχνολογίες που χρησιμοποιούνται για τη συλλογή αξιόπιστων δεδομένων με τροχίες οχημάτων, καθώς και οι μέθοδοι που εφαρμόζονται με σκοπό την εξάλειψη σφαλμάτων από αυτά.

Στο **δεύτερο κεφάλαιο** παρουσιάζεται η προτεινόμενη μεθοδολογία που βασίζεται σε μεθόδους μηχανικής μάθησης, όπως η τοπικά σταθμισμένη παλινδρόμηση (loess), η ομαδοποίηση (clustering) και η ταξινόμηση (classification). Οι τεχνικές αυτές και οι πιθανοί τρόποι εφαρμογής τους περιγράφονται αναλυτικά. Επιπλέον, επισημαίνονται τα μέτρα αποδοτικότητας που θα χρησιμοποιηθούν για την αξιολόγηση κάθε προσέγγισης στη συνέχεια της εργασίας.

Στο **τρίτο κεφάλαιο** επιχειρείται η σύγκριση του μοντέλου του Girps και της προτεινόμενης μεθοδολογίας χρησιμοποιώντας κάποια διαθέσιμα δεδομένα από ένα πείραμα που διεξήχθη στην πόλη Νάπολη της Ιταλίας (Punzo et al., 2005). Με τις δυο προσεγγίσεις γίνεται προσπάθεια εκτίμησης της ταχύτητας ενός οχήματος σε μια μελλοντική χρονική στιγμή έχοντας ως δεδομένα την ταχύτητα του ίδιου οχήματος και του μπροστινού του, καθώς και της μεταξύ τους απόστασης, σε μια προγενέστερη χρονική στιγμή. Αφού προηγηθεί μια ανάλυση ευαισθησίας για τις παραμέτρους του μοντέλου του Girps, γίνεται βαθμονόμηση του μοντέλου για μια αντιπροσωπευτική σειρά δεδομένων έτσι ώστε οι επιλεγόμενες τιμές να ελαχιστοποιούν το σφάλμα της εκτίμησης. Έπειτα, εξετάζεται η εγκυρότητα του βαθμονομημένου μοντέλου στις υπόλοιπες διαθέσιμες σειρές δεδομένων. Στη συνέχεια, εφαρμόζεται η μέθοδος loess εισάγοντας τα ίδια δεδομένα που εισήχθησαν στο μοντέλο του Girps και εξετάζεται η απόδοση της μεθόδου για τις διάφορες τιμές των παραμέτρων της προτεινόμενης μεθόδου (degree, span) και διαφορετικούς συνδυασμούς παραγόντων για όλες τις σειρές δεδομένων. Τέλος, συγκρίνονται τα αποτελέσματα που προέκυψαν από κάθε μέθοδο ξεχωριστά και διαπιστώνεται ποια μέθοδος παρέχει ακριβέστερη εκτίμηση.

Στο **τέταρτο κεφάλαιο** εξετάζεται η εγκυρότητα του μοντέλου του Girps και της προτεινόμενης μεθόδου στα διαθέσιμα δεδομένα NGSIM, θέτοντας ως τιμές παραμέτρων αυτές που προέκυψαν από το προηγούμενο κεφάλαιο. Γίνεται ξανά σύγκριση των δύο μεθόδων και προκύπτουν τα αντίστοιχα συμπεράσματα.

Στο **πέμπτο κεφάλαιο** εφαρμόζεται ολοκληρωμένη η προτεινόμενη μεθοδολογία, χρησιμοποιώντας ένα συνδυασμό της προτεινόμενης μεθόδου και των τεχνικών ομαδοποίησης και ταξινόμησης. Τέλος, συγκρίνεται η εφαρμογή της ολοκληρωμένης μεθοδολογίας με τη εφαρμογή μόνο της μεθόδου loess και την εφαρμογή του μοντέλου του Gijss.

Στο **έκτο κεφάλαιο** προκύπτουν γενικά συμπεράσματα από τη σύγκριση των προαναφερόμενων προσεγγίσεων και προτείνεται ποια από αυτές είναι ικανή να παρέχει την πιο αξιόπιστη εκτίμηση. Τέλος, τίγονται θέματα για περαιτέρω ανάλυση και γίνονται προτάσεις για μελλοντική έρευνα.

2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

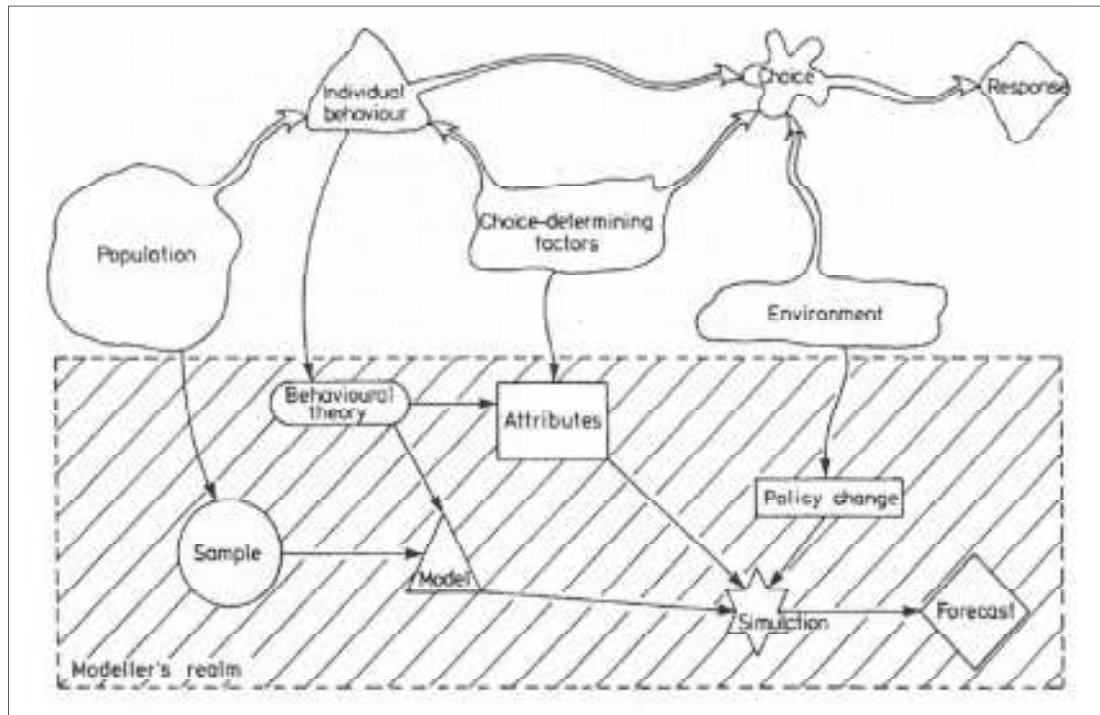
Στο παρόν κεφάλαιο παρουσιάζεται η βιβλιογραφική ανασκόπηση σχετικά με τα πρότυπα ακολουθίας οχημάτων. Αρχικά δίνονται βασικοί ορισμοί για το σύστημα, το μοντέλο και τη προσομοίωση και έπειτα αναφέρονται οι κατηγορίες μοντέλων προσομοίωσης που υπάρχουν. Στη συνέχεια παρουσιάζονται τα πιο γνωστά μοντέλα ακολουθίας οχημάτων και αναλύεται το μοντέλο του Gipps που θα χρησιμοποιηθεί σαν μοντέλο αναφοράς στη συνέχεια της εργασίας. Τέλος, περιγράφονται οι μέθοδοι που χρησιμοποιούνται συνήθως για συλλογή δεδομένων με τροχιές οχημάτων, καθώς και οι μέθοδοι που χρησιμοποιούνται για τη διόρθωσή τους με σκοπό τον περιορισμό των σφαλμάτων.

2.1. Βασικοί ορισμοί

Ένα *σύστημα* υπάρχει, λειτουργεί και εξελίσσεται στο χρόνο και στο χώρο, π.χ. μια κυκλοφοριακή κατάσταση (Sargent, 1999).

Ένα *μοντέλο* είναι η απλοποιημένη αναπαράσταση του συστήματος σε κάποιο συγκεκριμένο σημείο του χώρου ή του χρόνου με σκοπό την καλύτερη κατανόηση του πραγματικού συστήματος. Έτσι, υπάρχουν μοντέλα που αναπαριστούν π.χ. την κίνηση των οχημάτων (Sargent, 1999).

Ένα *μοντέλο προσομοίωσης της κυκλοφορίας* ορίζεται ως ένα πακέτο ανάλυσης στον υπολογιστή που αναπαριστά τις κινήσεις οχημάτων σε ένα μοντελοποιημένο οδικό δίκτυο και ποσοτικοποιεί την απόδοση του δικτύου ως αποτέλεσμα των κινήσεων των συνολικών οχημάτων που συμμετέχουν (Institute of Transportation Engineers California, 2004).



Σχήμα 2.1.: Μοντελοποίηση της κυκλοφορίας
 Πηγή: Ortuzar and Willumsen, 1995

2.2. Κατηγορίες μοντέλων προσομοίωσης

Τα μοντέλα προσομοίωσης μπορούν να κατηγοριοποιηθούν ανάλογα με το επίπεδο λεπτομέρειας που η κυκλοφοριακή ροή μοντελοποιείται, σε μικροσκοπικά, μεσοσκοπικά και μακροσκοπικά (Olstam and Tapani, 2004).

Στα *μικροσκοπικά* μοντέλα όλα τα οχήματα ή τα μέρη ενός συστήματος περιγράφονται μεμονωμένα. Τα μοντέλα αυτά μπορούν να συμπεριλαμβάνουν τις αλληλεπιδράσεις μεταξύ των οχημάτων ή μεταξύ των οχημάτων και του οδικού δικτύου. Κάθε όχημα στο οδικό δίκτυο περιγράφεται από μερικές παραμέτρους όπως την προέλευση του, τον προορισμό του, την επιθυμητή ταχύτητα, την επιτάχυνση και την επιβράδυνση που μπορεί να επιφέρει, τον τύπο του οχήματος, την υπομονή του οδηγού κ.ά. (Bellemans et al., 2002). Λόγω της μεγάλης λεπτομέρειας που προσφέρουν, θεωρούνται κατάλληλα για την αξιολόγηση «ευφών συστημάτων αναφοράς» και την ανάλυση μικρών οδικών δικτύων. Ωστόσο, παρουσιάζουν μεγαλύτερη δυσκολία στην εφαρμογή τους και απαιτούν περισσότερο χρόνο και κόστος (Burghout et al., 2005). Τα μικροσκοπικά μοντέλα ανάλογα με τη συμπεριφορά των οδηγών που εξετάζουν χωρίζονται σε μοντέλα αποδοχής κενού,

μοντέλα προσαρμογής ταχύτητας, μοντέλα αλλαγής λωρίδας κυκλοφορίας, μοντέλα προσπέρασης και μοντέλα ακολουθίας οχημάτων.

Τα μοντέλα αποδοχής κενού καθορίζουν τις ελάχιστες αποδεκτές αποστάσεις με τα γειτονικά οχήματα σε οδικές διασταυρώσεις ή σημεία συγχώνευσης λωρίδων κυκλοφορίας. Τα μοντέλα προσαρμογής ταχύτητας αναφέρονται στην υιοθέτηση της ταχύτητας σχεδιασμού της οδού από το όχημα σε κάθε θέση του στο οδικό δίκτυο. Τα μοντέλα αλλαγής λωρίδας κυκλοφορίας περιγράφουν τη συμπεριφορά των οδηγών όταν αποφασίζουν να αλλάξουν λωρίδα κυκλοφορίας σε έναν οδικό σύνδεσμο με πολλές λωρίδες κυκλοφορίας. Αντίστοιχα, τα μοντέλα προσπέρασης ελέγχουν ην οδική συμπεριφορά κατά την προσπέραση σε αγροτικές οδούς δύο λωρίδων. Τέλος, υπάρχουν τα μοντέλα ακολουθίας οχημάτων τα οποία εξετάζουν τις αλληλεπιδράσεις των οχημάτων με τα προπορευόμενα οχήματα που βρίσκονται στην ίδια λωρίδα κυκλοφορίας (Olstam and Tapani, 2004). Τα μοντέλα αυτά έχουν αποτελέσει θέμα μελέτης από πολλούς ερευνητές (Bellemans et al., 2002).

Τα *μακροσκοπικά* κυκλοφοριακά μοντέλα χρησιμοποιούν συγκεντρωτικές μεταβλητές για να περιγράψουν μια κυκλοφοριακή κατάσταση. Τέτοιου είδους μοντέλα προσομοιώνουν την κίνηση σαν μια συνεχόμενη ροή και χρησιμοποιούν θεωρίες που εφαρμόζονται για την κίνηση ενός υγρού (Burghout et al., 2005). Μακροσκοπικές μετρήσεις αποτελούν η ταχύτητα, η κυκλοφοριακή ροή και η πυκνότητα της κυκλοφορίας (Boxill and Yu, 2000). Ένα μακροσκοπικό μοντέλο ορίζει τη σχέση μεταξύ της κυκλοφοριακής πυκνότητας, της μέσης ταχύτητας και της ροής. Ένα τέτοιο μοντέλο είναι εύκολο να οριστεί και να εφαρμοστεί εξαιτίας των λίγων παραμέτρων που πρέπει να οριστούν (Bellemans et al., 2002). Τα μοντέλα αυτά θεωρούνται κατάλληλα για την ανάλυση μεγάλων οδικών δικτύων (Olstam and Tapani, 2004). Ωστόσο, αν και βελτιώνεται η υπολογιστική απόδοσή του μοντέλου, μειώνεται η λεπτομέρεια ανάλυσης (Burghout et al., 2005).

Τα *μεσοσκοπικά* μοντέλα έχουν κοινά στοιχεία τόσο με τα μακροσκοπικά όσο και τα μικροσκοπικά μοντέλα (Boxill and Yu, 2000). Το επίπεδο λεπτομέρειάς τους βρίσκεται ανάμεσα στα επίπεδα των μακροσκοπικών και των μικροσκοπικών μοντέλων. Τα μεσοσκοπικά μοντέλα προσομοίωσης περιγράφουν μεμονωμένα οχήματα αλλά σε συγκεντρωτικό επίπεδο, συνήθως με σχέσεις ταχύτητας-πυκνότητας και θεωρητικές προσεγγίσεις για την αναμονή σε ουρά (Burghout et al., 2005). Οι οδικοί σύνδεσμοι αποτελούνται από ένα κινούμενο μέρος και ένα μέρος που αναμένει

στην ουρά (Ben-Akiva et al., 2002, Ben-Akiva et al., 2010). Τα οχήματα κινούνται στο κινούμενο μέρος με μια ταχύτητα η οποία καθορίζεται από τις επικρατούσες συνθήκες, συνήθως την κυκλοφοριακή πυκνότητα. Από την άλλη πλευρά τα οχήματα που αναμένουν στην ουρά καθορίζονται από την χωρητικότητα εξόδου του συγκεκριμένου οδικού τμήματος (Antonίου and Koutsopoulos, 2007).

2.3. Πρότυπα ακολουθίας οχημάτων (Car-following models)

2.3.1. Γενική περιγραφή προτύπων ακολουθίας οχημάτων

Ένα πρότυπο ακολουθίας οχημάτων ελέγχει τη συμπεριφορά των οδηγών σε σχέση με το προπορευόμενο όχημα στην ίδια λωρίδα κυκλοφορίας. Ένα όχημα περιορίζεται από την κίνηση του μπροστινού οχήματος, διότι η οδήγηση του με την επιθυμητή ταχύτητα μπορεί να οδηγήσει σε σύγκρουση. Όταν ένα όχημα δεν περιορίζεται από κάποιο προπορευόμενο όχημα τότε θεωρείται ότι κινείται ελεύθερα με την επιθυμητή ταχύτητα του οδηγού. Οι ενέργειες του οχήματος που ακολουθεί ένα άλλο καθορίζονται από την επιτάχυνση του οχήματος, αν και σε μερικά μοντέλα όπως αυτό του Gipps (1981), οι ενέργειες του οχήματος που ακολουθεί καθορίζονται από την ταχύτητα του. Μερικά πρότυπα ακολουθίας οχημάτων περιγράφουν τη συμπεριφορά των οδηγών μόνο στην περίπτωση που ακολουθούν κάποιο άλλο όχημα, ενώ συμπεριλαμβάνουν όλες τις καταστάσεις. Κάθε πρότυπο ακολουθίας οχημάτων πρέπει να ορίζει τόσο την κατάσταση του οχήματος όσο και τις ενέργειες που γίνονται σε κάθε κατάσταση. Τα περισσότερα μοντέλα ακολουθίας οχημάτων περιγράφουν τη συμπεριφορά των οδηγών σε διάφορες καταστάσεις. Ένας συνήθης διαχωρισμός που γίνεται είναι ο εξής. Μία κατάσταση αντιστοιχεί σε συνθήκες ελεύθερης οδήγησης, μια άλλη στην περίπτωση που ένα όχημα ακολουθεί ένα άλλο και μια ακόμη για επείγουσα επιβράδυνση. Τα οχήματα σε ελεύθερη κατάσταση δεν έχουν περιορισμούς και προσπαθούν να επιτύχουν την επιθυμητή ταχύτητα, ενώ τα οχήματα που ακολουθούν άλλο όχημα περιορίζονται από την ταχύτητα του. Τα οχήματα σε κατάσταση επείγουσας επιβράδυνσης, υποβάλλονται σε απότομη πέδηση προκειμένου να αποφευχθεί μια σύγκρουση (Olstam and Tapani, 2004).

2.3.2. Κατηγορίες προτύπων ακολουθίας οχημάτων

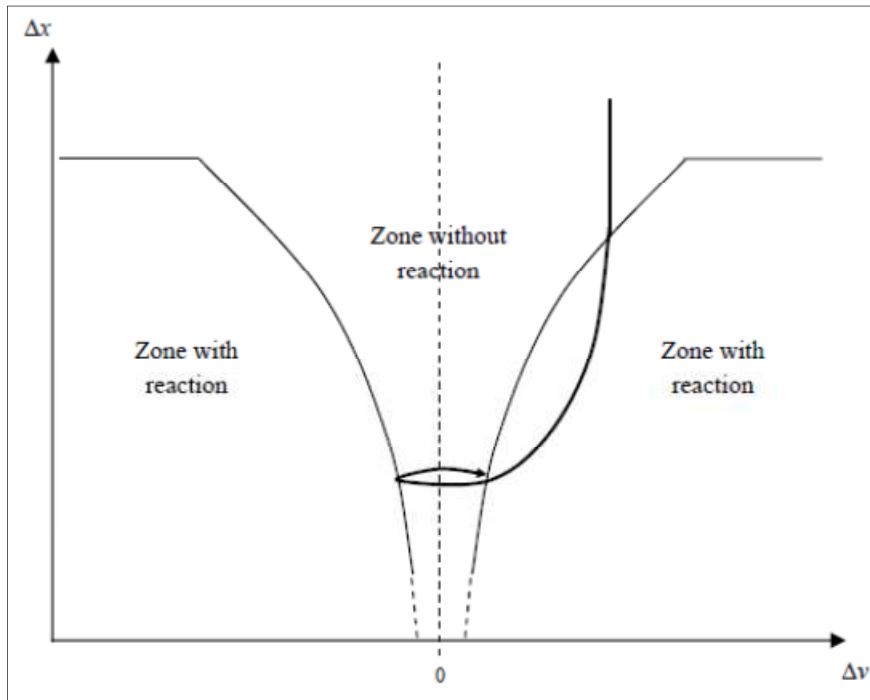
Σύμφωνα με τους Olstam and Tapani (2004) τα πρότυπα ακολουθίας οχημάτων χωρίζονται σε κατηγορίες ανάλογα με τη λογική που χρησιμοποιούν. Οι κατηγορίες αυτές είναι τα μοντέλα Gazis-Herman-Rothery, τα μοντέλα απόστασης ασφαλείας, ψυχοσωματικά μοντέλα και τα μοντέλα ασαφούς λογικής. Στη συνέχεια ακολουθεί η περιγραφή τους σύμφωνα με τους Olstam and Tapani (2004).

Τα μοντέλα Gazis-Herman-Rothery (GHR) ή γενικά πρότυπα ακολουθίας οχημάτων έκαναν την εμφάνιση τους το 1958 και από τότε έχουν υποστεί αρκετές βελτιώσεις. Τα μοντέλα αυτά ελέγχουν μόνο την πραγματική συμπεριφορά των οδηγών που ακολουθούν κάποιο άλλο όχημα. Η βασική σχέση μεταξύ του προπορευόμενου οχήματος και του οχήματος που ακολουθεί είναι μια συνάρτηση τύπου δράσης-αντίδρασης. Η επιτάχυνση του οχήματος που ακολουθεί είναι ανάλογη με την ταχύτητα του προπορευόμενου οχήματος, τη διαφορά της ταχύτητας των δύο οχημάτων και την απόσταση ανάμεσα στα δύο οχήματα (Brackstone and McDonald, 1999). Το μοντέλο επίσης περιλαμβάνει τρεις παραμέτρους που ελέγχουν τις αναλογίες. Το μοντέλο μπορεί αν είναι συμμετρικό στην περίπτωση που οι τιμές των παραμέτρων που χρησιμοποιούνται είναι ίδιες για καταστάσεις επιτάχυνσης και επιβράδυνσης ή μη συμμετρικό στην περίπτωση που χρησιμοποιούνται διαφορετικές τιμές.

Στα μοντέλα απόστασης ασφαλείας, ο οδηγός του οχήματος που ακολουθεί ένα άλλο όχημα οφείλει να τηρεί πάντα μια απόσταση ασφαλείας από το μπροστινό όχημα. Σύμφωνα με τους Hoogendoorn and Bovy (2001) ο κανόνας «Ένας κανόνας για την ακολουθία ενός οχήματος σε απόσταση ασφαλείας είναι να υπάρχει ανάμεσα στο όχημα που ακολουθεί και το προπορευόν όχημα μια απόσταση ίση με το μήκος ενός οχήματος για κάθε δέκα μίλια την ώρα ταχύτητα που ταξιδεύει» αποτελεί ένα απλό παράδειγμα ενός μοντέλου απόστασης ασφαλείας. Ωστόσο, η απόσταση ασφαλείας συνήθως καθορίζεται από τις εξισώσεις κίνησης του Νεύτωνα. Σε μερικά μοντέλα, η απόσταση αυτή υπολογίζεται ως η απαραίτητη απόσταση να αποφευχθεί μια σύγκρουση αν το προπορευόν όχημα επιβραδύνει απότομα. Το πρώτο μοντέλο αυτής της κατηγορίας παρουσιάστηκε από τους Kometsani and Sasaki το 1959 (Brackstone and McDonald, 1998). Το 1981 ο Gipps υπέβαλε μια βελτιωμένη έκδοση του πρωτότυπου μοντέλου. Σε αυτή τη βελτιωμένη έκδοση του μοντέλου το όχημα που ακολουθεί δεν πρόκειται να συγκρουστεί με το προπορευόμενό του εάν η χρονική

απόσταση από αυτό είναι μεγαλύτερη ή ίση με $3T/2$, όπου T είναι ο χρόνος αντίδρασης, και η εκτίμηση του οδηγού από το πίσω όχημα για την επιβράδυνση του μπροστινού οχήματος είναι μεγαλύτερη ή ίση από την πραγματική επιβράδυνση του μπροστινού οχήματος.

Το 1963 ο Michaels παρουσίασε μια νέα προσέγγιση των μοντέλων ακολουθίας οχημάτων (Brackstone and McDonald, 1999). Τα μοντέλα που χρησιμοποιούν αυτήν την προσέγγιση χαρακτηρίζονται ψυχοσωματικά. Τα GHR μοντέλα που περιγράφηκαν παραπάνω θεωρούν ότι ο οδηγός του πίσω οχήματος αντιδρά σε μικρές αλλαγές της σχετικής ταχύτητας και σε οποιαδήποτε ενέργεια του προπορευόμενου οχήματος ακόμα και αν η απόσταση μεταξύ των δυο οχημάτων είναι πολύ μεγάλη ή η αντίδραση του οχήματος που ακολουθεί δεν υφίσταται μόλις η σχετική ταχύτητα μηδενιστεί. Αυτό μπορεί να διορθωθεί είτε επεκτείνοντας τα GHR μοντέλα προσθέτοντας περισσότερες καταστάσεις- συνθήκες οδήγησης είτε χρησιμοποιώντας ένα ψυχοσωματικό μοντέλο. Τα ψυχοσωματικά μοντέλα χρησιμοποιούν τα κατώτερα όρια ή σημεία που ο οδηγός αλλάζει τη συμπεριφορά του. Οι οδηγοί μπορούν να αντιδρούν σε αλλαγές της απόστασης ή της σχετικής ταχύτητας μόνο όταν επιτυγχάνονται αυτά τα κατώτερα όρια (Leutzbach, 1988). Τα κατώτερα όρια και οι καταστάσεις που ορίζουν, παρουσιάζονται σε ένα σχετικό διάγραμμα διαστήματος-ταχύτητας για το ζεύγος των δυο οχημάτων. Ένα τέτοιο παράδειγμα φαίνεται στο σχήμα 2.2. Χαρακτηριστικά παραδείγματα ψυχοσωματικών προτύπων οχημάτων είναι αυτά που αναπτύχθηκαν από τους Wiedemann and Reiter (1992) and Fritzsche (1994).



Σχήμα 2.2.: Ένα ψυχοσωματικό μοντέλο ακολουθίας οχημάτων
 Πηγή: Olstam and Tarani, 2004

Τα μοντέλα ασαφούς λογικής (fuzzy logic) χρησιμοποιούν ασαφή σύνολα για να ποσοτικοποιήσουν, για παράδειγμα τι σημαίνει «πολύ κοντά». Τα σύνολα αυτά εισάγονται σε λογικούς κανόνες, π.χ. αν «πολύ κοντά», τότε να γίνει χρήση της επείγουσας επιβράδυνσης. Στα προηγούμενα μοντέλα που περιγράφηκαν θεωρείται γνωστή η ακριβής ταχύτητα, η απόσταση από τα άλλα οχήματα και ό,τι άλλο απαιτείται. Αντιθέτως, στα μοντέλα ασαφούς λογικής οι οδηγοί θεωρούνται ικανοί να κρίνουν αν η ταχύτητα του προπορευόμενου οχήματος είναι χαμηλή, μέτρια, υψηλή ή πολύ υψηλή. Τα ασαφή σύνολα είναι πιθανό να υπερκαλύπτονται και σε αυτές τις περιπτώσεις χρησιμοποιείται μια συνάρτηση πιθανότητας προκειμένου να αποφασιστεί π.χ. αν ο οδηγός θεωρεί την ταχύτητα του μπροστινού οχήματος χαμηλή ή μέτρια. Έχουν γίνει αρκετές προσπάθειες στη δημιουργία και χρήση τέτοιου είδους μοντέλων, όπως για παράδειγμα οι Al-Shihabi and Mourant (2003).

2.3.3. Το μοντέλο του Gipps

Το μοντέλο του Gipps (1981) αποτελεί ένα από τα πιο διαδεδομένα μοντέλα που βασίζονται στην απόσταση ασφαλείας. Χρησιμοποιείται καθημερινά καθώς αποτελεί τη βάση για διάφορα λογισμικά μικροσκοπικής προσομοίωσης, όπως AIMSUN

(Barceló J., 2002), DRACULA (Liu R., 2010), SISTM 1 (Liu R., 2010), SIGSIM (Silcock J.P., 1993), SITRAS (Hidas P., 1998). Για τον λόγο αυτό κρίνεται σκόπιμη η ανάλυσή του. Σύμφωνα με αυτό το μοντέλο, η ταχύτητα του οχήματος (n) που ακολουθεί ένα προπορευόν όχημα ($n-1$) υπόκειται σε τρεις περιορισμούς. Πρώτον, η ταχύτητα του οχήματος n δεν υπερβαίνει την ταχύτητα με την οποία ο οδηγός του οχήματος επιθυμεί να κινείται (V_n). Δεύτερον, το όχημα επιταχύνει αρχικά γρήγορα ώσπου να πλησιάσει την επιθυμητή ταχύτητα και έπειτα όταν την πλησιάσει η επιτάχυνση μειώνεται στο μηδέν. Στην περίπτωση που τα δύο οχήματα βρίσκονται μακριά μεταξύ τους, δηλαδή συμπεριφέρονται όπως σε κατάσταση ελεύθερης ροής, οι δύο παραπάνω συνθήκες συνοψίζονται στην εξίσωση:

$$u_n(t+\tau) \leq u_n(t) + 2.5 \cdot a_n \cdot \tau \cdot \left(1 - \frac{u_n(t)}{V_n}\right) \cdot \sqrt{\left(0.025 + \frac{u_n(t)}{V_n}\right)} \quad (2.1)$$

Η τρίτη συνθήκη ισχύει σε κυκλοφοριακές συνθήκες (π.χ. κυκλοφοριακή συμφόρηση) που τα οχήματα κινούνται κοντά μεταξύ τους και καθορίζει τη συμπεριφορά του πίσω οχήματος ενώ επιβραδύνει. Θεωρείται δεδομένο ότι ο οδηγός του οχήματος που ακολουθεί ένα άλλο θα προσαρμόσει την ταχύτητα του, έτσι ώστε να μπορεί να τηρήσει μια απόσταση ασφαλείας, ακόμα και αν το προπορευόμενο όχημα σταματήσει ξαφνικά. Επομένως, σύμφωνα με τη συνθήκη αυτή στην περίπτωση που η κίνηση του πίσω οχήματος περιορίζεται από την κίνηση του προπορευόμενου οχήματος ισχύει:

$$u_n(t+\tau) \leq b_n \cdot \tau + \sqrt{\left(b_n^2 \cdot \tau^2 - b_n \left[2 \cdot [x_{n-1}(t) - s_{n-1} - x_n(t)] - u_n(t) \cdot \tau - \frac{u_{n-1}(t)^2}{\hat{b}}\right]\right)} \quad (2.2)$$

Όπως φαίνεται από την εξίσωση (2.2) η ταχύτητα του οχήματος που ακολουθεί επηρεάζεται από το χρόνο αντίδρασης του οδηγού, την απόσταση του οχήματος που ακολουθεί από το προπορευόν όχημα, τις ταχύτητες των δυο οχημάτων και τον ρυθμό επιβράδυνσης που οι δυο οδηγοί επιθυμούν να έχουν. Επίσης, ο Gipps (1981) επισήμανε ότι στον χρόνο αντίδρασης του οδηγού θα πρέπει να προστεθεί ένα επιπλέον περιθώριο ασφαλείας θ το οποίο θα εξασφαλίζει το έγκαιρο σταμάτημα του οχήματος σε περίπτωση που η αντίδραση του οδηγού καθυστερήσει για κάποιο λόγο. Το περιθώριο ασφαλείας θεωρείται σταθερό και ίσο με το μισό του χρόνου αντίδρασης ($\tau/2$) και έχει συμπεριληφθεί στην εξίσωση (2.2) (Rakha and Gao, 2010). Επίσης, ο οδηγός του οχήματος n προσαρμόζει την ταχύτητα του σύμφωνα με την μέγιστη εκτιμώμενη πέδηση που μπορεί να επιφέρει ο οδηγός του οχήματος $n-1$ αλλά

μπορεί να ασκήσει μεγαλύτερη επιβράδυνση αν αυτό κριθεί αναγκαίο. Επομένως ακόμα και αν η θέληση επιβράδυνσης του μπροστινού οδηγού υποτιμηθεί από τον οδηγό του οχήματος που ακολουθεί, το μοντέλο δεν καταρρίπτεται.

Συνολικά, σύμφωνα με τους παραπάνω περιορισμούς η ταχύτητα του οχήματος n σε χρόνο $(t+\tau)$ μπορεί να υπολογιστεί από την εξίσωση (2.3).

$$u_n(t+\tau) = \min \left\{ \begin{array}{l} u_n(t) + 2.5 \cdot a_n \cdot \tau \cdot \left(1 - \frac{u_n(t)}{V_n}\right) \cdot \sqrt{\left(0.025 + \frac{u_n(t)}{V_n}\right)} \\ b_n \cdot \tau + \sqrt{\left(b_n^2 \cdot \tau^2 - b_n \left[2 \cdot [x_{n-1}(t) - s_{n-1} - x_n(t)] - u_n(t) \cdot \tau - \frac{u_{n-1}(t)^2}{\hat{b}}\right]\right)} \end{array} \right\} \quad (2.3)$$

Όπου:

a_n : η μέγιστη επιτάχυνση την οποία ο οδηγός του οχήματος n επιθυμεί να αποκτήσει (m/s^2)

b_n : η μέγιστη πέδηση την οποία ο οδηγός του οχήματος n επιθυμεί να επιφέρει ($b < 0$) προκειμένου να αποφύγει μια σύγκρουση, $b_n < 0$ (m/s^2)

\hat{b} : η εκτιμώμενη μέγιστη πέδηση που ο οδηγός του προπορευόμενου οχήματος ($n-1$) επιθυμεί να επιφέρει. Πρόκειται για μια εκτίμηση καθώς είναι αδύνατο ο οδηγός του δεύτερου οχήματος να γνωρίζει ακριβώς πόσο θα φρενάρει ο οδηγός του μπροστινού οχήματος (m/s^2).

$s_{n-1} = L_{n-1} + \text{Safety}$, δηλαδή το μέγεθος του προπορευόμενου οχήματος ($n-1$) συμπεριλαμβανομένου του μήκους του και της απόστασης ασφαλείας την οποία το όχημα n δεν επιθυμεί να παραβιάσει ακόμη και όταν τα δύο οχήματα είναι σταματημένα (m)

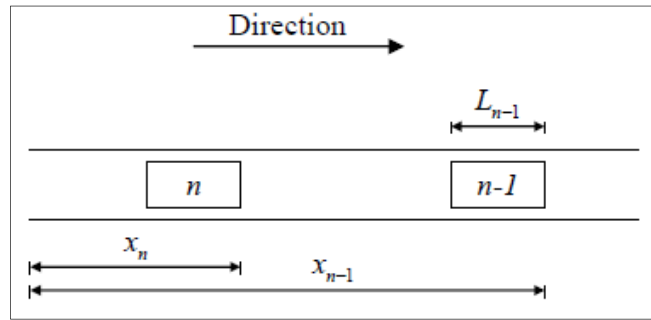
V_n : η ταχύτητα με την οποία ο οδηγός του οχήματος n επιθυμεί να κινείται (m/s)

$x_n(t)$, $x_{n-1}(t)$: η θέση του μπροστινού μέρους του αντίστοιχου οχήματος (n ή $n-1$) την χρονική στιγμή t (m), (Σχήμα 2.3.)

$u_{n-1}(t)$: η ταχύτητα του προπορευόμενου οχήματος ($n-1$) την χρονική στιγμή t (m/s)

$u_n(t)$: η ταχύτητα του πίσω οχήματος (n) την χρονική στιγμή t (m/s)

τ : ο χρόνος αντίδρασης (μια σταθερά για όλα τα οχήματα) (s)



Σχήμα 2.3.: Οι θέσεις των οχημάτων στο μοντέλο του Gipps
Πηγή: Olstam and Tarani, 2004

Παρατηρώντας καλύτερα την εξίσωση 2.3. και το σχήμα 2.3. γίνεται εύκολα αντιληπτό ότι τα απαραίτητα δεδομένα για την εφαρμογή του μοντέλου είναι οι ταχύτητες των οχημάτων n , $n-1$ καθώς και η μεταξύ τους απόσταση, όπως αποδεικνύεται από τη σχέση (2.4):

$$x_{n-1} - x_n - s_{n-1} = x_{n-1} - x_n - (L_{n-1} + \text{Safety}) = (x_{n-1} - x_n - L_{n-1}) - \text{Safety} = D_{23} - \text{Safety}$$

$$x_{n-1} - x_n - s_{n-1} = D_{23} - \text{Safety} \quad (2.4)$$

Πολλοί ερευνητές έχουν προσπαθήσει να τροποποιήσουν το παραπάνω μοντέλο. Αξιοσημείωτες είναι οι αναλύσεις των Wilson (2000) και Rakha et al. (2009). Ωστόσο, το μοντέλο που περιγράφεται από την εξίσωση (2.3) παραμένει το πιο ευρέως χρησιμοποιημένο. Η διαδεδομένη χρήση του αποδίδεται στο σαφές περιεχόμενο και τις ξεκάθαρες υποθέσεις του (Ciuffo et al., 2012). Λεπτομερής ανάλυση του μοντέλου και της εξέλιξης του μέχρι σήμερα έγινε από τους Ciuffo et al. (2012), οι οποίοι κρίνουν αναγκαία τη βαθμονόμηση των παραμέτρων του μοντέλου.

2.3.4. Εύρη τιμών για τις παραμέτρους του μοντέλου του Gipps

Για την εφαρμογή του μοντέλου του Gipps απαιτείται η βαθμονόμηση των παραμέτρων του αναφέρθηκαν στην προηγούμενη ενότητα. Το μοντέλο έχει βαθμονομηθεί με τις ακόλουθες τιμές (Gipps, 1981):

- Το a_n παίρνει τιμές από μια κανονική κατανομή $N(1.7, 0.3^2)$
- Η μέγιστη επιθυμητή επιβράδυνση b_n ισούται με $-2.0a_n$, προς αποφυγήν τρακαρίσματος όμως είναι μεγαλύτερη.
- Η εκτιμώμενη μέγιστη επιβράδυνση \hat{b} ισούται με $\min(-3.0, (b_n - 3.0)/2)$

- Η απόσταση s_{n-1} παίρνει τιμές από μια κανονική κατανομή $N(6.5, 0.3^2)$
- Η επιθυμητή ταχύτητα V_n παίρνει τιμές από την κανονική κατανομή $N(20.0, 3.2^2)$
- Ο χρόνος αντίδρασης τ ισούται με $2/3$ s

Επειδή, όμως, το 95% των τιμών μιας κανονικής τυχαίας μεταβλητής βρίσκεται στο διάστημα $(\mu-3\sigma, \mu+3\sigma)$, μπορεί να θεωρηθεί ότι τα προτεινόμενα εύρη από τον Girps (1981) για τις παραμέτρους είναι:

Παράμετροι	Κατώτερη τιμή	Ανώτερη τιμή	Μέση τιμή
$\alpha_n(\text{m/s}^2)$	0.8	2.6	1.7
$b_n(\text{m/s}^2)$	-1.6	-5.2	-3.4
$\hat{b}(\text{m/s}^2)$	-3.0	-4.1	-3.2
$s_{n-1}(\text{m})$	5.6	7.4	6.5
$V_n(\text{m/s})$	10.4	29.6	20
$\tau(\text{s})$	$2/3$	-	$2/3$

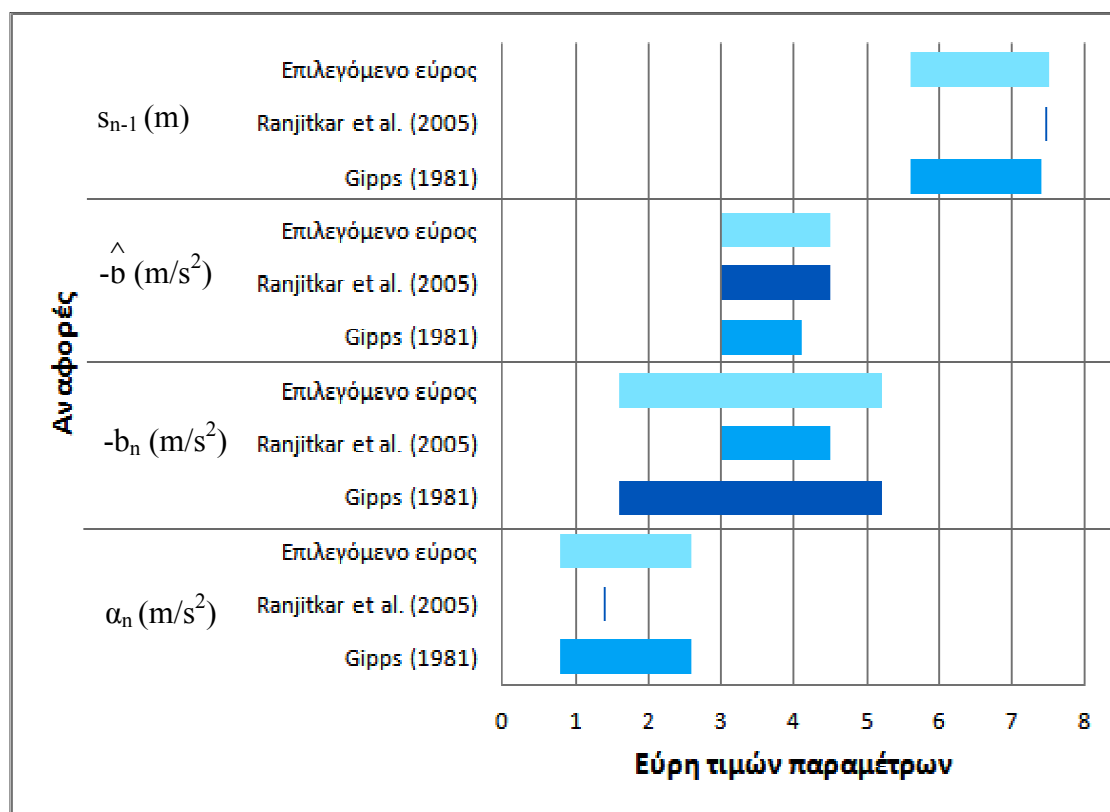
*Πίνακας 2.1.: Εύρη παραμέτρων του μοντέλου του Girps
Πηγή: Girps, 1981*

Επειδή, όμως, η παραπάνω βαθμονόμηση έγινε το 1981, πρέπει να ελεγχθεί αν οι τιμές αυτές είναι πλέον κατάλληλες. Σύμφωνα με τους Ranjitkar et al. (2005b), κάποια γενικά όρια ή τιμές για τις παραμέτρους είναι τα ακόλουθα:

Παράμετροι	Κατώτερη τιμή	Ανώτερη τιμή	Μέση τιμή
$\alpha_n(\text{m/s}^2)$	1.5	-	1.5
$b_n(\text{m/s}^2)$	-3.0	-4.5	-3.75
$\hat{b}(\text{m/s}^2)$	-3.0	-4.5	-3.75
$s_{n-1}(\text{m})$	7.5	-	7.5
$V_n(\text{m/s})$	20	25	22.5
$\tau(\text{s})$	0.5	3.0	1.75

*Πίνακας 2.2.: Προτεινόμενα εύρη παραμέτρων από τους Ranjitkar et al. (2005)
Πηγή: Ranjitkar et al., 2005(b)*

Τα εύρη που αναφέρονται στους πίνακες 2.2.,2.3. καθώς και το επιλεγόμενο εύρος που τα υπερκαλύπτει απεικονίζονται στο παρακάτω διάγραμμα. Για την ταχύτητα V_n επιλέγεται το εύρος που αναφέρεται στον πίνακα 2.1 ως πιο εκτενές και δεν απεικονίζεται στο διάγραμμα 2.1, καθώς παίρνει μεγαλύτερες τιμές και δεν είναι δυνατή η απεικόνιση του στο ίδιο διάγραμμα μαζί με τις υπόλοιπες παραμέτρους.



Διάγραμμα 2.1.: Εύρη παραμέτρων του μοντέλου του Gipps σύμφωνα με βιβλιογραφία

Σύμφωνα με τον Gipps (1981) η παράμετρος s_n δεν επηρεάζει σημαντικά την ταχύτητα ή την κυκλοφοριακή ροή καθώς η απόσταση αυτή είναι πολύ μικρή σε σχέση με τις αποστάσεις που μεσολαβούν ανάμεσα στα οχήματα όταν κινούνται, με εξαίρεση τις καταστάσεις έντονης κυκλοφοριακής συμφόρησης που τα οχήματα πλησιάζουν αρκετά. Η εφαρμογή του μοντέλου επηρεάζεται, όμως, από τις τιμές της ταχύτητας V_n , του χρόνου αντίδρασης τ , και της αναλογίας των επιβραδύνσεων b_n / \hat{b} .

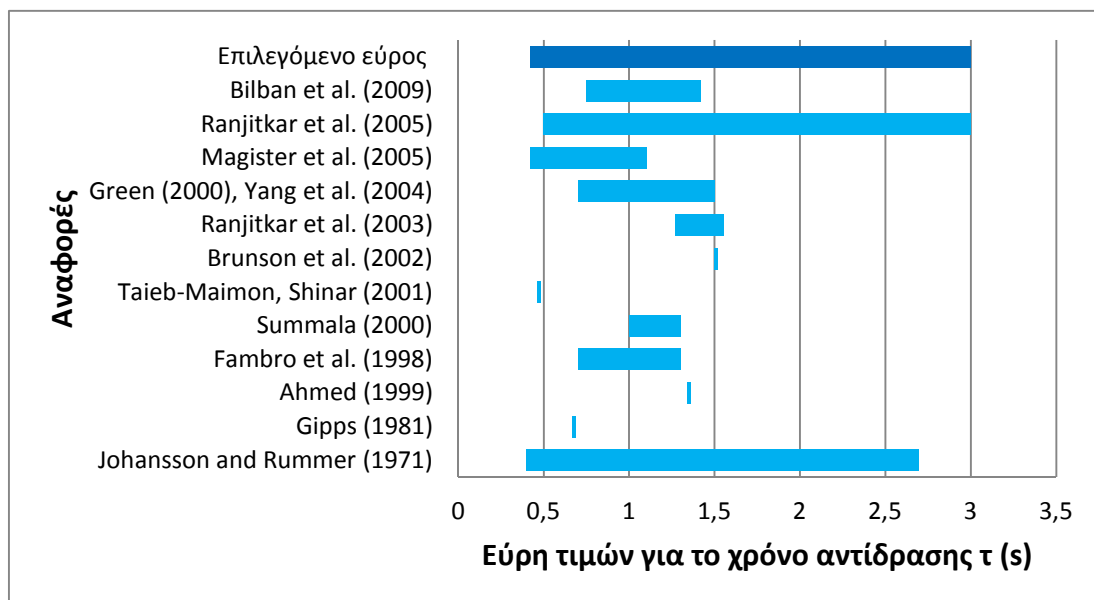
Όσον αφορά το χρόνο αντίδρασης, πολλές μελέτες έχουν διεξαχθεί με σκοπό τον υπολογισμό του είτε με πειράματα στο εργαστήριο είτε με προσομοιωτές οδήγησης. Ως χρόνος αντίδρασης ενός οδηγού αναφέρεται η διάρκεια από τη στιγμή που ο

οδηγός παρατηρεί ένα συμβάν μέχρι τη στιγμή που θα πατήσει το πεντάλ του φρένου (Yang et al., 2004). Οι Johansson and Rummel (1971) βάζοντας 300 άτομα να πατήσουν ένα πετάλι κατόπιν ενός θορύβου εκτίμησαν ότι ο χρόνος αντίδρασης κυμαίνεται από 0.4 έως 2.7 s, με έναν μέσο όρο 1 s. Επιπλέον, η μέση τιμή των χρόνων αντίδρασης όπως αυτή εκτιμήθηκε από τον Ahmed (1999) είναι 1.34 s. Από την άλλη πλευρά, οι Taieb-Maimon and Shinar (2001) επισημαίνουν ότι ο μέσος χρόνος αντίδρασης μετρήθηκε 0.47 s υπό εργαστηριακές συνθήκες, ενώ σύμφωνα με τους Brunson S. J. et al. (2002) θεωρείται να είναι 1.5 s. Οι Fambro et al. (1998) υπολόγισαν το μέσο χρόνο αντίδρασης για μη αναμενόμενα και αναμενόμενα συμβάντα αντίστοιχα 1.3 και 0.7 s. Ο Ozaki (1993) ανέπτυξε μία γραφική μέθοδο για τον υπολογισμό του χρόνου αντίδρασης σε σχέση με την ταχύτητα και την επιτάχυνση και υπέθεσε ότι ο χρόνος αντίδρασης μπορεί να σχετίζεται με την απόσταση, την ταχύτητα και την επιτάχυνση του προπορευόμενου οχήματος. Οι Ranjitkar et al. (2003) εφαρμόζοντας τη μέθοδο αυτή στην ανάλυση συμπεριφορών ακολουθίας οχημάτων και συλλέγοντας κάποια πειραματικά δεδομένα, εκτίμησαν ο μέσος χρόνος αντίδρασης των οδηγών κυμαίνεται από 1.27 έως 1.55 s.

Μια άλλη μελέτη που χρησιμοποιεί έναν προσομοιωτή αλλά και πραγματικές συνθήκες οδήγησης υπέδειξε ότι ο χρόνος αντίδρασης των οδηγών σε έναν αναμενόμενο κίνδυνο σε πραγματικές συνθήκες οδήγησης είναι κατά μέσο όρο 0.42 s, ενώ σε έναν μη αναμενόμενο κίνδυνο είναι 1.1 s (Magister et al., 2005). Ο Green (2000) και οι Yang et al (2004) επισημαίνουν ότι ο χρόνος αντίδρασης μπορεί να κυμαίνεται από 0.70- 1.5 s. Επίσης, ο Green (2000) αναφέρει ότι η ανταπόκριση σε μη αναμενόμενα σήματα αλλά λογικά να συμβούν όπως τα φώτα των φρένων ενός προπορευόμενου οχήματος, είναι περίπου 1.25 s, ενώ ο χρόνος αντίδρασης σε αναπάντεχα συμβάντα υπολογίζεται περίπου σε 1.5 s. Επίσης, ο Summala (2000) αναφέρει ότι οι οδηγοί μπορούν να αντιδράσουν σε ένα εμπόδιο που θα συναντήσουν σε περίπου 1.0-1.3 s, ενώ οι οι Ranjitkar et al. (2005) ορίζουν ένα γενικό εύρος 0.5-3 s για το χρόνο αντίδρασης. Σε μια λίγο πιο πρόσφατη μελέτη οι Bilban et al (2009) μέτρησαν το μέσο χρόνο αντίδρασης 0.965 s και αναφέρουν ότι ο ελάχιστος χρόνος αντίδρασης που επετεύχθη στο πλαίσιο της έρευνας τους ήταν 0.75 s, ενώ ο υψηλότερος 1.42 s.

Τα διαφορετικά αποτελέσματα οφείλονται στον τρόπο μέτρησης του χρόνου αντίδρασης στην κάθε έρευνα. Άλλωστε, υπάρχουν πολλοί παράγοντες που

επηρεάζουν το χρόνο αντίδρασης ενός οδηγού όπως η ηλικία, οι κυκλοφοριακές συνθήκες, η ορατότητα, η συνομιλία με συνεπιβάτες ή η χρήση κινητού τηλεφώνου. Τα διάφορα προτεινόμενα εύρη τιμών που προαναφέρθηκαν περιγράφονται στο διάγραμμα 2.2. Σύμφωνα με την παραπάνω βιβλιογραφία και θέλοντας να οριστεί ένα μεγάλο και λογικό εύρος τιμών για το χρόνο αντίδρασης στο πλαίσιο της εργασίας αυτής, θα μπορούσε να θεωρηθεί ότι ο χρόνος αντίδρασης κυμαίνεται από 0.4 έως 3.0 s. Ωστόσο, αξιοσημείωτο είναι ότι το κατώτερο όριο που λαμβάνεται είναι σύμφωνα με το διάγραμμα 2.2 ένα εύρος 0.5-1.5 s συμπεριλαμβάνει την πλειοψηφία των ερευνών που αναφέρθηκαν. Επίσης, το κατώτερο όριο μπορεί να χαρακτηριστεί αρκετά χαμηλό, αλλά σύμφωνα με τη βιβλιογραφία αποδεκτό.



Διάγραμμα 2.2.: Εύρη τιμών για τον χρόνο αντίδρασης των οδηγών σύμφωνα με τη βιβλιογραφία

Συνολικά, συνδυάζοντας τα εύρη τιμών από τους Gipps (1981) και Ranjitkar et al. (2005) καθώς και τη σχετική βιβλιογραφία για το χρόνο αντίδρασης, όπως αυτά προέκυψαν από την παραπάνω ανάλυση, τα επιλεγόμενα εύρη που θα εξεταστούν στη συνέχεια της εργασίας αναφέρονται στον πίνακα 2.3. Ωστόσο, κάποια κατώτατα όρια όπως αυτό του χρόνου αντίδρασης μπορούν να χαρακτηριστούν ως ακραία.

Παράμετροι	Κατώτερη τιμή	Ανώτερη τιμή	Μέση τιμή
τ (s)	0.4	3	1.7
V_n (m/s)	10.4	29.6	20
s_{n-1} (m)	5.6	7.5	6.6
α_n (m/s ²)	0.8	2.6	1.7
b_n (m/s ²)	-5.2	-1.6	-3.4
\hat{b} (m/s ²)	-4.5	-3.0	-3.8

Πίνακας 2.3.: Επιλεγόμενα εύρη παραμέτρων

2.4. Τεχνολογίες συλλογής δεδομένων

Για την ανάλυση προτύπων ακολουθίας οχημάτων απαιτείται η συλλογή δεδομένων για τροχιές οχημάτων από το πεδίο. Για τη συλλογή δεδομένων υπάρχουν δύο προσεγγίσεις. Η πρώτη περιλαμβάνει την παρακολούθηση των οχημάτων με τη χρήση καμερών, ενώ η δεύτερη τη χρήση οχημάτων εξοπλισμένων με κατάλληλα όργανα. Κάθε μέθοδος από αυτές που παρουσιάζονται παρακάτω έχει τα πλεονεκτήματα και τα μειονεκτήματα της. Ωστόσο, αναλόγως την εφαρμογή, την επιθυμητή ακρίβεια και το σκοπό του πειράματος μπορεί να επιλεγθεί η κατάλληλη μέθοδος (Punzo et al., 2005).

➤ Παρακολούθηση οχημάτων με κάμερες

Όσον αφορά στην πρώτη προσέγγιση οι κάμερες μπορεί να είναι τοποθετημένες σε σταθερά και υπερυψωμένα σημεία ή σε εναέριες πλατφόρμες (Hoogendoorn et al., 2005; Treiterer, 1975). Η προσέγγιση αυτή επιτρέπει τη συλλογή δεδομένων για τροχιές πολλών οχημάτων αλλά σε ένα περιορισμένο οδικό τμήμα (συνήθως μήκους έως 500 m). Με τη μέθοδο αυτή μπορούν να παρατηρηθούν συμπεριφορές οδηγών σχετικά με την ακολουθία οχημάτων, την αλλαγή λωρίδας κυκλοφορίας και την αποδοχή κενού. Επίσης, η χρήση αυτής της προσέγγισης δεν επηρεάζει καθόλου την ροή της κυκλοφορίας. Ωστόσο, η καταγραφή και η ανάλυση της οδηγικής συμπεριφοράς υπόκεινται σε χωρικούς περιορισμούς. Ένα ακόμα μειονέκτημα της μεθόδου είναι ότι η πειραματική διάταξη και η επεξεργασία των δεδομένων είναι

χρονοβόρες. Η ακρίβεια της μεθόδου αυτής κυμαίνεται από 10 έως 100 cm, ενώ το ποσοστό ανίχνευσης από 80% έως 98% (NGSIM, 2004).

➤ *Παρακολούθηση οχημάτων με ειδικούς αισθητήρες*

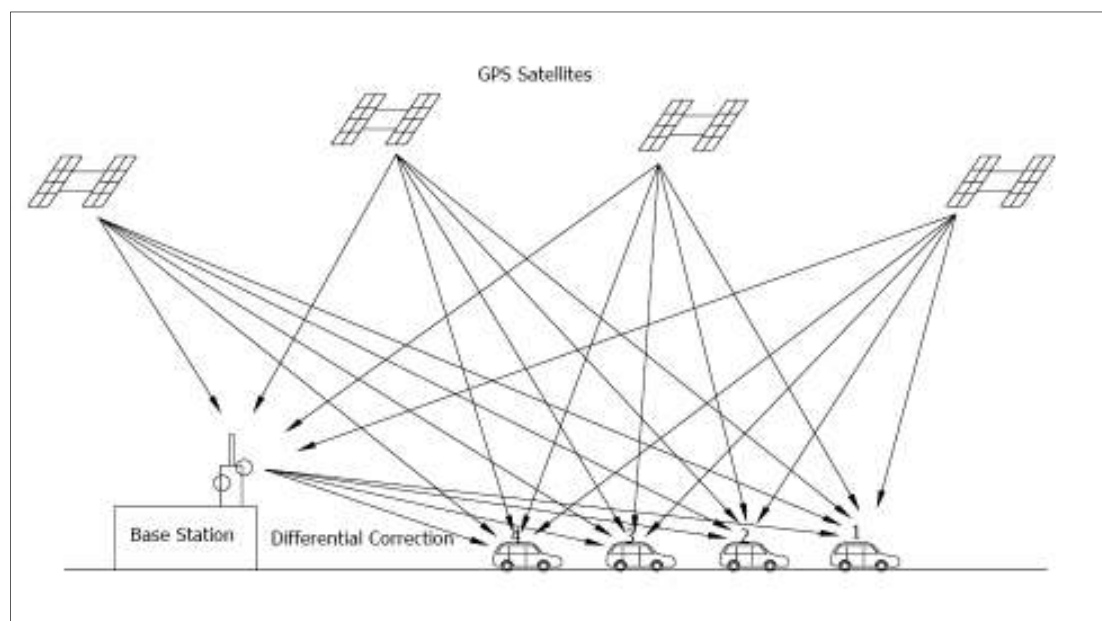
Η δεύτερη προσέγγιση συλλογής δεδομένων περιλαμβάνει τη χρήση πιο σύγχρονης τεχνολογίας. Μία εκδοχή αυτής της προσέγγισης είναι η χρήση οχημάτων εξοπλισμένων με τουλάχιστον δύο όργανα και συγκεκριμένα έναν αισθητήρα ταχύτητας (π.χ. ένα λέιζερ ταχύμετρο) και ένα αισθητήρα σχετικής απόστασης και ταχύτητας (π.χ. ένα ραντάρ). Αυτή η τεχνική επιτρέπει την καταγραφή δεδομένων από την τροχιά ενός οχήματος που ακολουθεί ένα άλλο για μεγάλη διαδρομή, επιτρέποντας παράλληλα την άγνοια του οδηγού για τη συμμετοχή του στο πείραμα. Η μέθοδος αυτή παρέχει τη δυνατότητα ανάλυσης της συμπεριφοράς των οδηγών σε ποικίλες κυκλοφοριακές συνθήκες και σε διαφορετικά οδικά δίκτυα. Ωστόσο, περιορίζει την ανάλυση της οδηγικής συμπεριφοράς μόνο ως προς τη λογική της ακολουθίας ενός ζεύγους οχημάτων. Η ακρίβεια των ραντάρ κυμαίνεται περίπου από ± 0.2 m για τις αποστάσεις και ± 0.4 m/s για τη σχετική ταχύτητα (Brackstone et al., 2002).

➤ *Παρακολούθηση οχημάτων με τεχνολογία GPS*

Μια εναλλακτική μέθοδος που συγκαταλέγεται στην δεύτερη προσέγγιση συλλογής δεδομένων βασίζεται στην ανάπτυξη του Παγκόσμιου Συστήματος Εντοπισμού Θέσης (Global Positioning System). Το παγκόσμιο σύστημα εντοπισμού θέσης (GPS- Global Positioning System) αποτελεί ένα δορυφορικό σύστημα εντοπισμού θέσης με ακρίβεια που κυμαίνεται από μερικά χιλιοστά έως μερικά μέτρα ανάλογα με τον εξοπλισμό και τις τεχνικές που χρησιμοποιούνται. Ένας δέκτης GPS τοποθετημένος στο έδαφος υπολογίζει τις παγκόσμιες συντεταγμένες σε γεωγραφικό μήκος, γεωγραφικό πλάτος και υψόμετρο, λαμβάνοντας υπόψη τις αποστάσεις από τουλάχιστον τέσσερις δορυφόρους με τριγωνισμό (Ranjitkar et al., 2005a).

Η ακρίβεια του GPS μπορεί να βελτιωθεί σημαντικά με τη διόρθωση των σφαλμάτων τα οποία είναι ίδια για όλους τους δέκτες στην περιοχή του πειράματος. Τα σφάλματα αυτά δεν μπορούν να εντοπιστούν με τη χρήση ενός απλού δέκτη GPS, εκτός και να η θέση του είναι ήδη γνωστή. Στην τεχνική του *διαφορικού GPS (Differential GPS-DGPS)*, ένας δέκτης τοποθετείται σε μια γνωστή θέση και ονομάζεται ως σταθμός βάσης με σκοπό τη μέτρηση των σφαλμάτων αυτών και την αναμετάδοση των

διορθωμένων σημάτων στους υπόλοιπους φορητούς δέκτες που βρίσκονται σε κάποια απόσταση, ο αριθμός των οποίων δεν είναι περιορισμένος (Ranjitkar et al., 2005a). Η μέθοδος αυτή περιγράφεται επίσης στο σχήμα 2.3.



Σχήμα 2.4.: Μέθοδος διαφορικού GPS
Πηγή: Ranjitkar P., Suzuki, H. and Nakatsuji, T. , 2005 (α)

Η τεχνική του *κινηματικού εντοπισμού σε πραγματικό χρόνο (Real Time Kinematic-RTK)* αποτελεί ένα είδος διαφορικού GPS με ακρίβεια σε επίπεδο χιλιοστών. Η χρήση της μεθόδου έχει επιβεβαιωθεί έναντι των μετρήσεων που γίνονται με χρήση μηχανικών εξοπλισμών. Για την εφαρμογή της μεθόδου απαιτείται, εκτός από τους κινητούς δέκτες GPS και το σταθμό βάσης, μία ασύρματη ζεύξη μεταξύ τους (radio link). Η ζεύξη πραγματοποιείται χρησιμοποιώντας δύο συσκευές – μία με το δέκτη – οι οποίες αποτελούνται από ένα μόντεμ και ένα πομποδέκτη. Η διαφορά με το διαφορικό GPS έγκειται στο ότι ο σταθμός αναφοράς συνδέεται με τον ένα πομποδέκτη και εκπέμπει διορθώσεις, τις οποίες λαμβάνει ο δεύτερος πομποδέκτης που είναι συνδεδεμένος με τον κινητό δέκτη και τις μεταφέρει σε αυτόν. Ο κινητός δέκτης με τη σειρά του, χρησιμοποιεί τις διορθώσεις αυτές για να επιλύσει επιτόπου τις ασάφειες φάσης και να δώσει τα αποτελέσματα την ίδια στιγμή (Παραδείσης, 2000). Η τεχνική αυτή απαιτεί την εγκατάσταση ενός σταθμού βάσης και η

αποτελεσματικότητα της μεθόδου περιορίζεται από την απόσταση ανάμεσα στο σταθμό βάσης και τους κινητούς δέκτες (περίπου 10 km).

Η αξιοπιστία της μεθόδου μπορεί να βελτιωθεί ακόμα περισσότερο με τη χρήση μόνιμων σταθμών αναφοράς, όπως με την τεχνική εικονικών σταθμών αναφοράς. Η τεχνική αυτή βασίζεται στην ύπαρξη δικτύου μόνιμων δεκτών αναφοράς τοποθετημένους σε απόσταση μέχρι 80 km μακριά. Αυτοί τροφοδοτούν έναν κεντρικό υπολογιστή μέσω διαδικτυακής επικοινωνίας. Ο υπολογιστής επεξεργάζεται τα δεδομένα αυτά και πραγματοποιεί τις κατάλληλες διορθώσεις. Ο κινητός δέκτης στέλνει τη θέση του μέσω κινητού τηλεφώνου στον κεντρικό υπολογιστή και λαμβάνει τις διορθώσεις για τη θέση του. Η βασική αρχή είναι η παρεμβολή των δεδομένων ενός πλήθους σταθμών αναφοράς για να προκύψουν τα δεδομένα διορθώσεων για τους κινητούς δέκτες, γεγονός που μειώνει σημαντικά τα συστηματικά σφάλματα των RTK μετρήσεων. Αυτή η τεχνική παρέχει καλύτερη απόδοση και δεν απαιτεί την εγκατάσταση ενός σταθμού αναφοράς όπως η μέθοδος του κινηματικού εντοπισμού σε πραγματικό χρόνο. Ωστόσο, η αμφίδρομη επικοινωνία που απαιτείται αυξάνει το κόστος εφαρμογής και περιορίζει την ισχύ της μεθόδου (Ranjitkar et al., 2005a).

Όποια μέθοδος και αν εφαρμοστεί, τα οχήματα μιας φάλαγγας εξοπλίζονται με δέκτες GPS που καταγράφουν τις τροχιές των οχημάτων. Με αυτόν τρόπο επιτρέπεται η ανάλυση της συμπεριφοράς των οδηγών για πολλά οχήματα που κινούνται στη σειρά ταυτόχρονα, δεδομένου ότι καταγράφεται ο χρόνος της κάθε μέτρησης. Επιπλέον, οι καταγραμμένες θέσεις των οχημάτων χαρακτηρίζονται από μεγάλη ακρίβεια, στην περίπτωση του διαφορικού GPS η αναμενόμενη ακρίβεια είναι 10mm. Η συχνότητα των μετρήσεων (έως 20 Hz) είναι επίσης υψηλή σε σχέση με τις υπόλοιπες μεθόδους. Το βασικό μειονέκτημα της μεθόδου είναι ότι οι χρήστες γνωρίζουν για τη συμμετοχή τους στο πείραμα. Επιπλέον, αν το πείραμα διεξάγεται σε πραγματικό περιβάλλον είναι πιθανή η παρεμβολή άλλων οχημάτων στη φάλαγγα (Punzo et al., 2005).

2.5. Μέθοδοι διόρθωσης δεδομένων

Οι τεχνικές που χρησιμοποιούνται για τη διόρθωση των δεδομένων από τα σφάλματα των μετρήσεων είναι οι ακόλουθες (Punzo et al., 2005).

➤ *Τοπική παλινδρόμηση (local regression)*

Η μέθοδος της τοπικής παλινδρόμησης, ενισχυμένη με κάποια προκαταρκτικά όρια αποκοπής για τα σφάλματα μετρήσεων, χρησιμοποιείται αντί για την απλή παλινδρόμηση, όταν τα δεδομένα (π.χ. δεδομένα τροχιών) απαιτούν ευέλικτες λειτουργικές μορφές. Οι συντεταγμένες από κάθε δέκτη GPS επεξεργάζονται και το αποτέλεσμα που προκύπτει είναι η αλληλουχία στιγμιαίων ταχυτήτων ανά χρονική στιγμή μέτρησης. Συγκεκριμένα, η διαδικασία που ακολουθείται απαρτίζεται από τα εξής βήματα:

- Εξάλειψη των δεδομένων με σφάλμα μεγαλύτερο από ένα επιτρεπτό κατώτερο όριο
- Παρεμβολή των δεδομένων που λείπουν
- Υπολογισμός των καμπυλόγραμμων τετμημένων από τις θέσεις τους
- Αναζήτηση του καλύτερου πολυωνύμου παρεμβολής με τη μέθοδο των ελαχίστων τετραγώνων για σημεία που βρίσκονται μέσα σε πλαίσιο ενός καθορισμένου πλάτους (local regression)
- Υπολογισμός της πρώτης και δεύτερης παραγώγου για το κεντρικό σημείο του πλαισίου (στιγμιαία ταχύτητα και επιτάχυνση)
- Επανάληψη των δυο προηγούμενων βημάτων ώσπου η ταχύτητα και η επιτάχυνση να είναι εντός των περιορισμών
- Επανάληψη των ίδιων βημάτων για όλα τα σημεία του διανύσματος των καμπυλόγραμμων τετμημένων

Η μέθοδος που περιγράφηκε μπορεί να εξομαλύνει τα προφίλ των ταχυτήτων όταν τα σήματα δεν είναι πολύ διαταραγμένα. Ωστόσο, υπάρχει αβεβαιότητα ως προς την επιλογή των παραμέτρων της τοπικής παλινδρόμησης και του κατώτερου ορίου που τίθεται. Επίσης, δεν παρέχει μεγάλη συνέπεια στην εκτίμηση ταχυτήτων των οχημάτων.

➤ *Επιλεκτικά φίλτρα*

Η δεύτερη μέθοδος περιλαμβάνει την εφαρμογή βαθυπερατών επιλεκτικών φίλτρων (π.χ. χρήση του φίλτρου Savitzky- Golay για την εκτίμηση επιταχύνσεων) και κατά κύριο λόγο την εφαρμογή ενός τέταρτης τάξης φίλτρου Butterworth. Η ανάλυση φάσματος δεν επιτρέπει την αναγνώριση της συχνότητας αποκοπής, η οποία

προκύπτει από την προσομοίωση (ίση με 0.5 Hz). Τα αποτελέσματα που εξάγει η μέθοδος συνοψίζονται ποιοτικά σε δυο βασικά σημεία. Για ταχύτητες που παρατηρήθηκαν και είναι χαμηλότερες από 0.3 m/s (δηλαδή όταν το όχημα είναι στάσιμο), ο θόρυβος δεν εξαλείφεται με την εφαρμογή του φίλτρου. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί θέτοντας αυτές τις ταχύτητες ως μηδενικές πριν την εφαρμογή του φίλτρου. Όσον αφορά σε ταχύτητες μεγαλύτερες από 0.3 m/s, ο θόρυβος συνήθως έχει υψηλότερες συχνότητες από το σήμα του GPS και κατά συνέπεια ο θόρυβος περιορίζεται από το φίλτρο. Παρόλα αυτά μερικές φορές παρατηρείται ένας πολωμένος θόρυβος του σήματος των μετρήσεων. Σε αυτήν την περίπτωση, αν και το φίλτρο περιορίζεται στην υψηλή συχνότητα του σήματος, εξάγει μια αναξιόπιστη εκτίμηση.

➤ *Φίλτρο Kalman*

Το φίλτρο Kalman αποτελεί μία μέθοδο (με εφαρμογή διαδοχικών συνορθώσεων) που επιτρέπει πρόβλεψη, φιλτράρισμα σε πραγματικό χρόνο και εξομάλυνση παραμέτρων που μεταβάλλονται με το χρόνο (Αγάτζα-Μπαλοδήμου, 2008). Η ανάλυση που ακολουθεί βασίζεται στους Anderson and Moore (1979) και στους Punzo et al. (2005). Για την εφαρμογή της μεθόδου απαραίτητη είναι η ανάπτυξη των εξισώσεων κατάστασης του συστήματος καθώς και των εξισώσεων μέτρησης.

$$x(k+1) = A(k) \cdot x(k) + B(k) \cdot u(k) + D(k) \cdot \gamma(k) \quad (2.5)$$

$$y(k) = C(k) \cdot x(k) + \zeta(k) \quad (2.6)$$

Όπου:

$x(k)$, $u(k)$ και $y(k)$: οι μεταβλητές της κατάστασης, των εισαγόμενων και εξαγόμενων δεδομένων αντίστοιχα

$\gamma(k)$ και $\zeta(k)$: αναπαριστούν το θόρυβο του συστήματος και των μετρήσεων

A, B, C, D: οι πίνακες των συντελεστών των μεταβλητών του συστήματος, των δεδομένων που εισάγονται, των μετρήσεων και των σφαλμάτων του μοντέλου, για τη χρονική στιγμή k.

Τα σφάλματα του συστήματος και των μετρήσεων είναι στοχαστικές διαδικασίες με μέσο όρο μηδέν και γνωστό πίνακα συμμεταβλητοτήτων.

$$E[\gamma(k)] = 0 \quad (2.7)$$

$$E[\zeta(k)] = 0 \quad (2.8)$$

$$E[\gamma(k_i)\gamma(k_j)^T] = Q(k_i)\delta_{ij} \quad (2.9)$$

$$E[\zeta(k_i)\zeta(k_j)^T] = R(k_i)\delta_{ij} \quad (2.10)$$

Όπου $Q(k_i) \geq 0$ και $R(k_i) \geq 0$ είναι γνωστοί και $\delta_{ij}=0$ αν $i \neq j$ ή $\delta_{ij}=1$ αν $i=j$.

Οι εκτιμήσεις του φίλτρου προκύπτουν από την επίλυση της ακόλουθης εξίσωσης:

$$\hat{x}_F(k) = \hat{x}_P(k) + K(k) \cdot [y(k) - C(k) \cdot \hat{x}_P(k)] \quad (2.11)$$

Όπου \hat{x}_P είναι η πρόβλεψη κατά ένα βήμα του k και η οποία έχει προκύψει από την εξίσωση:

$$\hat{x}_P(k) = A(k-1) \cdot \hat{x}_F(k-1) + B(k-1) \cdot u(k-1) + D(k-1) \cdot \gamma(k-1) \quad (2.12)$$

Ο πίνακας $K(k)$ ονομάζεται πίνακας κέρδους (Kalman gain) και είναι:

$$K(k) = P_p(k) \cdot C(k)^T + [C(k) \cdot P_p(k) \cdot C(k)^T + R(k)]^{-1} \quad (2.13)$$

Ο πίνακας αυτός ικανοποιεί δύο απαιτήσεις, τη χρήση διαθέσιμων μετρήσεων που θα προσαρμόσουν την εκτίμηση του μοντέλου σε μια μελλοντική κατάσταση και δεν θα υποβαθμίσουν την εκτίμηση αυτή λόγω σφαλμάτων στις μετρήσεις. Ο πίνακας κέρδους είναι ανάλογος με τον πίνακα συμμεταβλητοτήτων του υπολογισμένου σφάλματος $P_p(k)$, το οποίο πρέπει να υπολογίζεται εκ νέου σε κάθε βήμα από την ακόλουθη εξίσωση:

$$P_p(k+1) = A(k) \cdot P_F(k) \cdot A(k)^T + D(k) \cdot Q(k) \cdot D(k)^T \quad (2.14)$$

όπου

$$P_F(k) = [I - K(k) \cdot C(k)] \cdot P_p(k) \quad (2.15)$$

3. ΜΕΘΟΔΟΛΟΓΙΑ

Στη κεφάλαιο αυτό περιγράφονται οι μέθοδοι μηχανικής μάθησης ως εναλλακτικές μέθοδοι για την ανάπτυξη πρότυπων ακολουθίας οχημάτων. Αρχικά περιγράφεται το γενικό μεθοδολογικό πλαίσιο που θα ακολουθηθεί στη συνέχεια της εργασίας. Επίσης, αναλύονται οι μέθοδοι της τοπικά σταθμισμένης παλινδρόμησης, της ομαδοποίησης και της ταξινόμησης. Τέλος, προκύπτουν κάποια αρχικά συμπεράσματα από τη σύγκριση των μεθόδων μηχανικής μάθησης και των μοντέλων προσομοίωσης που βασίζονται σε μαθηματικούς τύπους.

3.1. Γενικό μεθοδολογικό πλαίσιο

Η μεθοδολογία απαρτίζεται από δυο προσεγγίσεις και κάθε προσέγγιση από δύο μέρη. Στην πρώτη προσέγγιση χρησιμοποιείται η τοπικά σταθμισμένη παλινδρόμηση, ενώ στην δεύτερη ένας συνδυασμός μεθόδων μηχανικής μάθησης. Στο πρώτο μέρος κάθε προσέγγισης διαμορφώνεται ένα πρότυπο μοντέλο, ενώ στο δεύτερο περιγράφεται η εφαρμογή του.

Αρχικά λαμβάνονται και ελέγχονται τα διαθέσιμα δεδομένα από το πεδίο. Στη συγκεκριμένη εργασία τα διαθέσιμα δεδομένα δεν υπόκεινται σε διορθώσεις, όπως περιγράφηκε στην ενότητα 2.2 αλλά χρησιμοποιήθηκαν οι μετρήσεις απευθείας από το πεδίο. Δεν κρίθηκε αναγκαίο, δεδομένου ότι για τη σύγκριση δυο προσεγγίσεων τα σφάλματα δεν επηρεάζουν το αποτέλεσμα της σύγκρισης αφού και στις δυο προσεγγίσεις υπεισέρχονται τα ίδια ακριβώς σφάλματα. Ωστόσο, τα δεδομένα ελέγχονται, έτσι ώστε να μην υπάρχουν χονδροειδή λάθη. Αυτό επιτυγχάνεται ελέγχοντας τις τροχιές των οχημάτων που κινούνται σε σειρά. Αναμφισβήτητα οι τροχιές των οχημάτων δεν πρέπει να τέμνονται σε καμία περίπτωση, καθώς αυτό θα σήμαινε ότι τα οχήματα συγκρούονται. Αφού, ελεγχθούν τα δεδομένα ή διορθωθούν από τα σφάλματα των μετρήσεων (όπως γίνεται σε άλλες μελέτες), ακολουθείται το επόμενο βήμα της μεθοδολογίας.

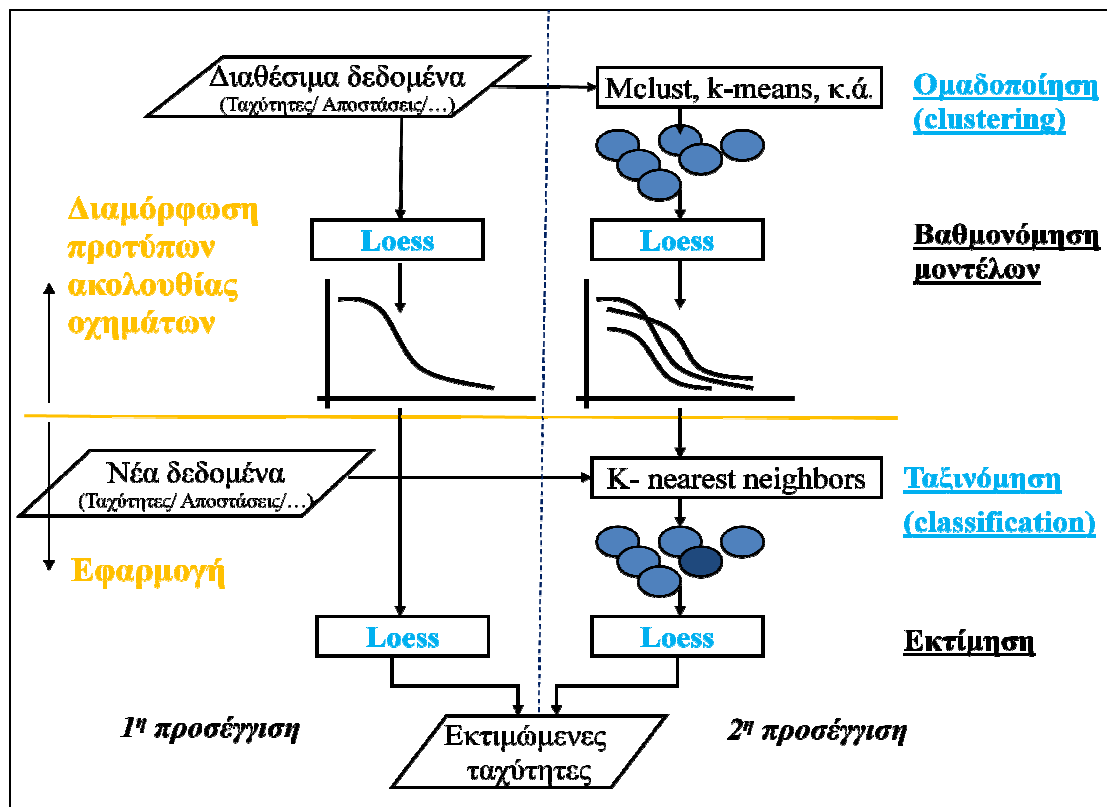
Για τα διαθέσιμα δεδομένα εκτιμάται ποια θα είναι η συμπεριφορά του οχήματος (πιο συγκεκριμένα η ταχύτητα του) την επόμενη χρονική στιγμή, επιχειρείται δηλαδή η εκτίμηση προτύπων ακολουθίας οχημάτων με τη χρήση μεθόδων μηχανικής μάθησης. Πρόκειται για μη παραμετρικές προσεγγίσεις, αφού δεν περιλαμβάνουν μια συγκεκριμένη σχέση. Κατά συνέπεια μπορούν να θεωρηθούν ως γενικεύσεις

προσεγγίσεων με πολλές καταστάσεις κυκλοφορίας (Antoniou and Koutsopoulos, 2006).

Η *πρώτη προσέγγιση* περιλαμβάνει την εφαρμογή της μεθόδου loess σε μια σειρά δεδομένων. Η πληροφορία που παράγεται σχετικά με το πώς επηρεάζεται η ταχύτητα ενός οχήματος από το προπορευόμενο όχημα ή ακόμα και το όχημα που το ακολουθεί αποθηκεύεται στο σύστημα (βαθμονόμηση μοντέλου - calibration). Στην ουσία προκύπτει ένα πρότυπο ακολουθίας οχημάτων που βασίζεται στη σειρά δεδομένων που εισήχθη. Στη συνέχεια, θεωρείται μια νέα σειρά δεδομένων, στην οποία θα εφαρμοστεί το πρότυπο που δημιουργήθηκε με σκοπό να γίνει μια εκτίμηση των ταχυτήτων που αφορούν τα νέα δεδομένα (estimation). Αναμφισβήτητα όσο περισσότερες καταστάσεις περιλαμβάνει η αρχική σειρά δεδομένων με την οποία δημιουργήθηκε το πρότυπο, τόσο πιο αντιπροσωπευτικό θα είναι το πρότυπο που δημιουργήθηκε και κατά συνέπεια θα εκτιμά καλύτερα τις ταχύτητες στις νέες σειρές μετρήσεων.

Η *δεύτερη προσέγγιση* περιλαμβάνει την εφαρμογή ενός συνδυασμού των μεθόδων μηχανικής μάθησης και προτείνεται γιατί τα διαθέσιμα δεδομένα προφανώς θα συμπεριλαμβάνουν διάφορες κυκλοφοριακές καταστάσεις. Επομένως πριν εφαρμοστεί η μέθοδος της τοπικά σταθμισμένης παλινδρόμησης, θα ήταν φρόνιμο να χρησιμοποιηθεί η διαδικασία της ομαδοποίησης των αρχικών δεδομένων. Συγκεκριμένα, με την ομαδοποίηση αναγνωρίζονται οι διαφορετικές καταστάσεις και τα δεδομένα χωρίζονται σε ομάδες ανάλογα με τις ομοιότητες ή διαφορές που έχουν (clustering). Στη συνέχεια εφαρμόζεται η μέθοδος της τοπικά σταθμισμένης παλινδρόμησης σε κάθε κλάση που δημιουργήθηκε ξεχωριστά, έτσι ώστε να προκύψουν πιο συγκεκριμένα πρότυπα για κάθε κατάσταση που θα προσαρμόζονται καλύτερα στα εκάστοτε δεδομένα (fitting). Μέχρι εδώ αναλύθηκε η φάση της βαθμονόμησης των μοντέλων (calibration). Στη συνέχεια, όταν προκύψουν νέες μετρήσεις, αυτές ταξινομούνται στις αντίστοιχες κλάσεις ανάλογα με ποια κλάση έχουν τα περισσότερα κοινά χαρακτηριστικά (classification). Έπειτα, το κατάλληλο ευέλικτο μοντέλο επιλέγεται και ανακτάται για κάθε κλάση ξεχωριστά. Τέλος, με την εφαρμογή των κατάλληλων μοντέλων προκύπτει η εκτίμηση των ταχυτήτων που αφορούν τα νέα δεδομένα (estimation).

Η συνολική μεθοδολογία που περιγράφηκε παραπάνω παρουσιάζεται στο σχήμα 3.1.



Σχήμα 3.1.: Μεθοδολογία εκτίμησης προτύπων ακολουθίας οχημάτων με τη χρήση μεθόδων μηχανικής μάθησης

Πηγή: Antoniou and Koutsopoulos, 2006, *Ίδια επεξεργασία*

3.2. Μηχανική μάθηση

3.2.1. Εισαγωγή στην ιδέα της μηχανικής μάθησης

Τα μοντέλα προσομοίωσης που αναφέρθηκαν στις προηγούμενες ενότητες χρησιμοποιούν μια συγκεκριμένη παραμετρική σχέση της ταχύτητας ενός οχήματος και της ταχύτητας του προπορευόμενου οχήματος ή της απόστασής του από αυτό ή κάποιου άλλου μεγέθους. Ωστόσο, η χρήση μιας συγκεκριμένης σχέσης σε όλες τις περιπτώσεις μπορεί να μην είναι κατάλληλη. Από την άλλη πλευρά, σχέσεις που προσαρμόζονται στα δεδομένα, ανάλογα με τις επικρατούσες συνθήκες μπορεί να είναι περισσότερο αποδοτικές για τη δυναμική προσομοίωση της κυκλοφορίας. Τέτοιες εναλλακτικές προσεγγίσεις μπορούν να προκύψουν με τη βοήθεια της μηχανικής μάθησης (Antoniou and Koutsopoulos, 2006).

Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση (machine learning), (Βλαχάβας κ.ά., 2005).

Με τις μεθόδους μηχανικής μάθησης το σύστημα ανακαλύπτει συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων δημιουργώντας πρότυπα, χωρίς όμως να είναι σαφείς οι συσχετίσεις που προκύπτουν. Παρακάτω, περιγράφονται τρεις εναλλακτικές προσεγγίσεις μηχανικής μάθησης και συγκεκριμένα η γεωγραφικά σταθμισμένη παλινδρόμηση, η ομαδοποίηση και η ταξινόμηση. Οι προσεγγίσεις αυτές χαρακτηρίζονται μη παραμετρικές και δεν χρησιμοποιούν μια μοναδική σχέση όπως τα μοντέλα που περιγράφηκαν σε παραπάνω ενότητες.

3.2.2. Τοπικά σταθμισμένη παλινδρόμηση (loess)

Η ανάλυση και η βιβλιογραφική ανασκόπηση που ακολουθεί για την τοπικά σταθμισμένη παλινδρόμηση βασίζεται στους Antoniou και Koutsopoulos (2006).

Η τοπικά σταθμισμένη παλινδρόμηση (locally weighted regression-loess) προτάθηκε για πρώτη φορά από τον Cleveland (1978) και αργότερα από τους Cleveland et al. (1988). Οι Cleveland και Devlin (1988) εντόπισαν τα πεδία εφαρμογής της συγκεκριμένης μεθόδου και συγκεκριμένα πρότειναν τη μέθοδο αυτή για την ανάλυση γραφικών δεδομένων, για τον έλεγχο παραμετρικών μοντέλων που εφαρμόζονται σε δεδομένα καθώς και την άμεση χρήση της τοπικής παλινδρόμησης αντί για τη χρήση παραμετρικών εξισώσεων.

Στον τομέα των μεταφορών, οι Sun et al. (2003, 2004) εφάρμοσαν την τοπική παλινδρόμηση για βραχυπρόθεσμη πρόβλεψη της κυκλοφορίας και επισήμαναν ότι ήταν καλύτερη μέθοδος από την μέθοδο του εγγύτερο γείτονα και τη μέθοδο kernel. Έπειτα, οι Toledo et al. (2006) χρησιμοποίησαν τη μέθοδο αυτή για να σχηματίσουν συνεχείς τροχιές οχημάτων έχοντας στη διάθεση τους δεδομένα για τη θέση αυτών και κατά συνέπεια μπόρεσαν να διαμορφώσουν το προφίλ ταχύτητας και επιτάχυνσης για τα οχήματα αυτά. Στη συνέχεια εφάρμοσαν τη μεθοδολογία αυτή σε μια σειρά δεδομένων και τα αποτελέσματα χρησιμοποιήθηκαν για τη βαθμονόμηση μοντέλων ακολουθίας οχημάτων.

Η τοπικά σταθμισμένη παλινδρόμηση αποτελεί μια γενίκευση της μεθόδου του κ-εγγύτερου γείτονα. Αν και η τελευταία δίνει μια εκτίμηση $g(x)$ σε ένα συγκεκριμένο σημείο, με την τοπικά σταθμισμένη παλινδρόμηση προκύπτει μια εκτίμηση $g(x)$ που αναφέρεται στην περιοχή που περιβάλλει το σημείο αυτό. Για την εκτίμηση αυτή μπορούν να χρησιμοποιηθούν διάφορα είδη λειτουργικών μορφών, όπως γραμμικές ή

τετραγωνικές συναρτήσεις ή πολλαπλά νευρωνικά δίκτυα. Παρακάτω, αναλύεται η τοπικά σταθμισμένη παλινδρόμηση σύμφωνα με τους Cleveland και Devlin (1988).

Η τοπικά σταθμισμένη παλινδρόμηση $y_i = \hat{g}(x_i) + \varepsilon_i$, όπου ε_i σφάλματα που υπολείπονται, παρέχει μια εκτίμηση $g(x)$ της επιφάνειας παλινδρόμησης για οποιαδήποτε τιμή x στον p - διαστάσεων χώρο των ανεξάρτητων μεταβλητών. Έστω q ένας ακέραιος $1 \leq q \leq n$, τότε για την εκτίμηση της $g()$ στο x χρησιμοποιούνται οι q παρατηρήσεις για τις οποίες οι τιμές x_i βρίσκονται κοντά στο x . Σε καθένα από αυτά τα σημεία δίνεται ένα βάρος σύμφωνα με την απόσταση του από το x . Τα σημεία που βρίσκονται κοντά στο x έχουν μεγάλο βάρος ενώ στα πιο απομακρυσμένα σημεία δίνεται μικρότερο βάρος. Μια συνάρτηση των ανεξάρτητων μεταβλητών προσαρμόζεται στην εξαρτημένη μεταβλητή με τη βοήθεια των ελαχίστων τετραγώνων με αυτά τα βάρη. Τότε $y = g(x)$ είναι η τιμή της χρησιμοποιούμενης συνάρτησης στο x .

Μια συνάρτηση απόστασης p στο χώρο των ανεξάρτητων μεταβλητών απαιτείται για την τοπικά σταθμισμένη παλινδρόμηση προκειμένου να καθοριστούν τα κατάλληλα βάρη. Σε περιπτώσεις μιας ανεξάρτητης μεταβλητής, η συνάρτηση p μπορεί να οριστεί ως ο τύπος της Ευκλείδειας απόστασης. Για περιπτώσεις πολλαπλής παλινδρόμησης είναι σημαντική η αξιολόγηση κάθε μεταβλητής με μια κλίμακα πριν την εφαρμογή μιας συνάρτησης απόστασης.

Η τοπικά σταθμισμένη παλινδρόμηση απαιτεί επίσης μια συνάρτηση βάρους και ο προσδιορισμός του μεγέθους της περιοχής που θα λαμβάνεται υπόψη. Μια συνάρτηση βάρους που χρησιμοποιείται συνήθως είναι η εξής:

$$W(u) = \min \left\{ \begin{array}{l} (1 - u^3)^3, 0 \leq u \leq 1 \\ 0 \end{array} \right\} \quad (3.1)$$

Τότε, το βάρος της παρατήρησης (y_i, x_i) είναι:

$$w_i(x) = W[p(x, x_i) / d(x)] \quad (3.2)$$

Όπου $d(x)$ είναι η απόσταση του q -εγγύτερου x_i στο x .

Η αντικειμενική συνάρτηση που πρέπει να ελαχιστοποιηθεί είναι το άθροισμα των τετραγώνων των σταθμισμένων υπολοίπων.

$$\sum_{i=1}^n w_i \cdot \varepsilon^2 \quad (3.3)$$

3.2.3. Ομαδοποίηση (clustering)

Ο σκοπός της ομαδοποίησης είναι η κατάταξη δεδομένων άγνωστης δομής σε ομάδες. Υπάρχουν αρκετές προσεγγίσεις ομαδοποίησης αλλά παρακάτω αναλύονται δύο από αυτές, η μέθοδος k-means και η μέθοδος που βασίζεται στο συνδυασμό διαφόρων μοντέλων.

Μέθοδος k-means

Ο αλγόριθμος k-means (MacQueen, 1967, Hartigan, 1975, Hartigan, 1979) επιλέγει έναν προκαθορισμένο αριθμό κέντρων ομάδων (cluster), τα οποία ελαχιστοποιούν τις αποστάσεις μεταξύ των ομάδων από τα κέντρα αυτά. Έστω ότι $h = \langle \mu_1, \dots, \mu_k \rangle$ περιγράφει τα κέντρα των k ομάδων. Ο υπολογισμός της υπόθεσης μέγιστης πιθανότητας για τη μέση τιμή μιας κανονικής κατανομής, δεδομένου ότι οι παρατηρήσεις είναι x_1, x_2, \dots, x_m , είναι απλός. Ωστόσο, ο αλγόριθμος k-means περιλαμβάνει συνδυασμό k διαφορετικών κανονικών κατανομών, χωρίς να υποδεικνύει από ποιες παρατηρήσεις προήλθε η κάθε κατανομή και κατά συνέπεια εμπεριέχει κρυμμένες μεταβλητές. Κάθε παρατήρηση μπορεί να γραφτεί ως μια πλειάδα $\langle x_i, z_{i1}, z_{i2}, \dots, z_{ik} \rangle$ όπου z_{ij} ισούται με 1 αν η x_i δημιουργήθηκε από την j κανονική κατανομή, διαφορετικά ισούται με 0. Σε αυτήν την περίπτωση το x_i αποτελεί την μεταβλητή που παρατηρήθηκε (ή το διάνυσμα των μεταβλητών) και οι z_{ij} αποτελούν τις κρυμμένες μεταβλητές. Η υπόθεση μέγιστης πιθανότητας εντοπίζεται κατόπιν επαναλαμβανόμενες επανεκτιμήσεις των αναμενόμενων τιμών των z_{ij} δεδομένης της τρέχουσας υπόθεσης h . Τέλος, μια νέα υπόθεση μέγιστης πιθανότητας υπολογίζεται ξανά με τις αναμενόμενες τιμές των κρυμμένων μεταβλητών. Για παράδειγμα, σε περίπτωση που ο προκαθορισμένος αριθμός των ομάδων (cluster) είναι 2, τότε $h = \langle \mu_1, \mu_2 \rangle$ όπου μ_1 και μ_2 είναι κάποιες αυθαίρετες αρχικές τιμές. Για την ολοκλήρωση της διαδικασίας ακολουθούν τα εξής βήματα ώσπου ο αλγόριθμος να καταλήξει σε μια σταθερή τιμή για το h :

- Υπολογίζονται οι αναμενόμενες τιμές $E[z_{ij}]$ για κάθε κρυμμένη μεταβλητή z_{ij} δεδομένου η τρέχουσα υπόθεση h ισχύει.

- Υπολογίζεται μια νέα υπόθεση μέγιστης πιθανότητας $h' = \langle \mu_1', \mu_2' \rangle$, θεωρώντας ότι η τιμή κάθε κρυμμένης μεταβλητής είναι η αναμενόμενη τιμή της $E[z_{ij}]$ που υπολογίστηκε στο προηγούμενο βήμα. Τέλος, η νέα υπόθεση αντικαθιστά την αρχική.

Μέθοδος βασισμένη στο συνδυασμό μοντέλων

Σύμφωνα με τους Fraley and Raftery (2002) η μέθοδος ομαδοποίησης που προτείνουν βασίζεται στο συνδυασμό μοντέλων. Η μέθοδος περιλαμβάνει τρία βασικά στοιχεία, αρχικοποίηση σύμφωνα με τα μοντέλα κατά συγκεντρωτική ιεραρχική ομαδοποίηση, εκτίμηση της μέγιστης πιθανότητας μέσω ενός αλγορίθμου μεγιστοποίησης και επιλογή του μοντέλου και του αριθμού των κλάσεων χρησιμοποιώντας τον συντελεστή Bayes και το κριτήριο BIC (Bayesian Information Criterion), Swartz (1978). Η συγκεντρωτική ιεραρχική ομαδοποίηση που βασίζεται σε μοντέλα αποτελεί μια προσέγγιση για τον υπολογισμό της μέγιστης πιθανότητας για ομαδοποίηση. Έχοντας ως δεδομένα $y = (y_1, \dots, y_n)$ τότε η πιθανότητα για ένα συνδυασμό μοντέλων με G επιμέρους μοντέλα είναι:

$$L_{CL}(q_1, \dots, q_G; l_1, \dots, l_n / y) = \prod_{i=1}^n f_{l_i}(y_i / q_{l_i}) \quad (3.4)$$

Όπου l_i είναι ετικέτες που υποδεικνύουν μια μοναδική ομάδα που θα ανήκει η κάθε παρατήρηση, παίρνοντας την τιμή k αν η παρατήρηση y_i ανήκει στην ομάδα k .

Στη συγκεντρωτική ιεραρχική ομαδοποίηση κάθε στάδιο συγχώνευσης αντιστοιχεί σε ένα μοναδικό αριθμό κλάσεων και έναν μοναδικό διαχωρισμό των στοιχείων. Ένας συγκεκριμένος διαχωρισμός που δίνεται μπορεί να μετασχηματιστεί σε μεταβλητές, που μπορούν να χρησιμοποιηθούν ως όροι πιθανοτήτων σε ένα βήμα M του αλγορίθμου για την εκτίμηση των παραμέτρων, κατά την αρχικοποίηση του αλγορίθμου. Σε συνδυασμό με τους συντελεστές Bayes και το κριτήριο BIC για την επιλογή του μοντέλου προκύπτει μια στρατηγική ομαδοποίησης με τα εξής βήματα:

- Καθορίζεται ένας μέγιστος αριθμός ομάδων και ένας συνδυασμός μοντέλων
- Παρουσιάζεται μια ιεραρχική συγκέντρωση που μεγιστοποιεί την πιθανότητα ομαδοποίησης για κάθε μοντέλο και γίνονται οι αντίστοιχες ομαδοποιήσεις έως και M ομάδες.

- Σε κάθε μοντέλο εφαρμόζεται ο αλγόριθμος μεγιστοποίησης και κάθε αριθμό ομάδων $2, \dots, M$ ξεκινάει η ομαδοποίηση από την ιεραρχική συγκέντρωση
- Υπολογίζεται ο συντελεστής BIC για την περίπτωση της μιας κλάσης και για το συνδυασμό των μοντέλων με τις βέλτιστες παραμέτρους του αλγορίθμου για $2, \dots, M$ κλάσεις. Παρόλο που καθορίζεται ο βέλτιστος αριθμός των κλάσεων, είναι πιθανό να επιλεγεί μικρότερος αριθμός, αποσκοπώντας στην ανάπτυξη πιο απλών μοντέλων.

3.2.4. Ταξινόμηση (classification)

Οι μέθοδοι ομαδοποίησης συνήθως συνοδεύονται από μια μέθοδο ταξινόμησης, έτσι ώστε να μπορούν να εφαρμοστούν. Οι μέθοδοι της ομαδοποίησης και της ταξινόμησης έχουν χρησιμοποιηθεί από πολλούς ειδικούς. Πιο συγκεκριμένα, ο El Faouzi (2004) με τη χρήση των μεθόδων αυτών αναζητάει το καλύτερο μοντέλο εκτίμησης. Έπειτα, οι El Faouzi and Lefevre (2006) επιχειρούν μια εκτίμηση του χρόνου μετακίνησης και οι Wang et al. (2005) κατηγοριοποιούν τη συμπεριφορά ακολουθίας οχημάτων σε πολλές καταστάσεις. Επίσης, οι Antoniou and Koutsopoulos (2006) εφαρμόζουν τη μεθοδολογία που αναλύθηκε στην ενότητα 3.1. Τέλος, οι Azimi and Zhang (2010) ομαδοποιούν τις συνθήκες κυκλοφοριακής ροής σε αυτοκινητόδρομο βασιζόμενοι στα χαρακτηριστικά της κυκλοφορίας.

Οι μέθοδοι ταξινόμησης χρησιμοποιούνται προκειμένου νέες μετρήσεις να ταξινομηθούν στην πιο κατάλληλη ομάδα από αυτές που προέκυψαν με την ταξινόμηση που προηγήθηκε. Το γενικό ζήτημα της ομαδοποίησης είναι ότι δίνονται K τάξεις, M μεγέθη και N ομαδοποιημένα σημεία δεδομένων και ζητείται να καθοριστεί ένα σύνολο κανόνων με σκοπό την ταξινόμηση μελλοντικών δεδομένων ενός μεγέθους. Κάθε σημείο δεδομένων αντιστοιχεί σε ένα ζεύγος $x_j \in R_M$ και $G_j \in \{1, \dots, K\}$, όπου x_j είναι το διανυσματικό μέγεθος και G_j είναι η τάξη του αντίστοιχου σημείου δεδομένων και απαιτείται ένας κανόνας που θα καθορίζει $G:R_M \rightarrow \{1, \dots, K\}$ (Roughan et al., 2004).

Υπάρχουν αρκετές μέθοδοι ομαδοποίησης, όπως η μέθοδος του εγγύτερου γείτονα, η ομαδοποίηση με τη χρήση νευρωνικών δικτύων, η μέθοδος της γραμμικής διαχωριστικής ανάλυσης. Παρακάτω αναλύεται η μέθοδος του εγγύτερου γείτονα ως απλή και πολύ διαδεδομένη.

Σύμφωνα με τη μέθοδο του εγγύτερου γείτονα όλες οι παρατηρήσεις αντιστοιχούν σε σημεία του n -διάστατου χώρου (Mitchell, 1997). Με αυτή τη μέθοδο κάθε νέο σημείο δεδομένων εισάγεται στην τάξη των εγγύτερων γειτόνων του από τα ήδη ομαδοποιημένα δεδομένα. Οι εγγύτεροι γείτονες ορίζονται σύμφωνα με την Ευκλείδεια απόσταση. Επομένως η τάξη της νέας παρατήρησης $G(x)$ είναι ίδια με την τάξη G_i των σημείων x_i (από τα ομαδοποιημένα δεδομένα) τα οποία ελαχιστοποιούν την απόσταση $\|x_i - x\|$. Πιο αναλυτικά, έστω x μια παρατήρηση που περιγράφεται από την πλειάδα $\langle a_1(x), a_2(x), \dots, a_n(x) \rangle$ όπου το $a_r(x)$ υποδηλώνει την τιμή της r ιδιότητας του x . Ιδιότητες του x μπορούν να είναι η πυκνότητα, η ροή της κίνησης, ο χρόνος και άλλα μεγέθη. Η απόσταση ανάμεσα σε δύο παρατηρήσεις ορίζεται ως εξής:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n [a_r(x_i) - a_r(x_j)]^2} \quad (3.5)$$

Στην μέθοδο του εγγύτερου γείτονα μπορεί να χρησιμοποιηθεί μια συνάρτηση διακριτών ή πραγματικών τιμών. Στην περίπτωση των διακριτών τιμών, όπως όταν μια παρατήρηση x_q πρέπει να ταξινομηθεί σε μια ομάδα, ο αλγόριθμος που χρησιμοποιείται συλλέγει τις k παρατηρήσεις από το σετ δεδομένων που είναι κοντά στην παρατήρηση αυτή και επιστρέφει:

$$f(x_q) \leftarrow \arg \max_{u \in V} \sum_{i=1}^k \delta[u, f(x_i)] \quad (3.6)$$

Όπου:

$$\delta(a, b) = \begin{cases} 1, & a = b \\ 0 & \end{cases} \quad (3.7)$$

3.2.5. Οφέλη και προκλήσεις

Από την ανάλυση που προηγήθηκε φαίνεται ότι τόσο οι μέθοδοι που βασίζονται σε μαθηματικά μοντέλα όσο και οι μέθοδοι μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για τη μελέτη και εκτίμηση της κίνησης οχημάτων. Ωστόσο, η κάθε εναλλακτική μέθοδος έχει αντίστοιχα πλεονεκτήματα και μειονεκτήματα. Από τη μια πλευρά τα μαθηματικά μοντέλα επεξηγούν ακριβώς με ποιον τρόπο γίνεται η εκτίμηση αλλά ενδεχομένως να απαιτούν περισσότερο χρόνο. Από την άλλη πλευρά

οι μέθοδοι μηχανικής μάθησης είναι πιο ευέλικτες, πιο γρήγορες αλλά δεν εξηγούν με ποιον τρόπο γίνεται η εκτίμηση καθώς γίνεται μηχανικά με τον υπολογιστή. Το ζητούμενο όμως είναι ποιες μέθοδοι είναι πιο ακριβείς και παρέχουν καλύτερα και πιο αξιόπιστα αποτελέσματα. Σύμφωνα με τους Antoniou και Koutsopoulos (2006), οι μέθοδοι μηχανικής μάθησης μπορούν να παρέχουν βελτιωμένη ακρίβεια στον υπολογισμό ταχυτήτων σε σχέση με τα μοντέλα που βασίζονται σε μαθηματικούς τύπους (συγκεκριμένα σχέση ταχύτητας-πυκνότητας) και ίσως αποδειχτούν να είναι πιο κατάλληλες για την εφαρμογή τους σε μοντέλα προσομοίωσης. Στη συνέχεια της εργασίας εξετάζεται αν για τα διαθέσιμα δεδομένα η εκτίμηση της ταχύτητας είναι πιο ακριβής με το μοντέλο του Girps που βασίζεται σε έναν μαθηματικό τύπο ή με την εφαρμογή μηχανικών μεθόδων.

3.3. Αξιολόγηση μοντέλων

Στη συνέχεια της εργασίας, η προτεινόμενη μεθοδολογία θα εφαρμοστεί στα διαθέσιμα δεδομένα. Ωστόσο, για να μπορεί να αξιολογηθεί ως μέθοδος, το μοντέλο του Girps εφαρμόζεται επίσης στα ίδια δεδομένα και επιχειρείται σύγκριση των δύο προσεγγίσεων.

Η αξιολόγηση κάθε προσέγγισης, τόσο του μοντέλου Girps όσο και της προτεινόμενης μεθοδολογίας που αναλύθηκε παραπάνω στην ενότητα 3.1., γίνεται με τα προτεινόμενα μέτρα αποδοτικότητας που βρέθηκαν στη βιβλιογραφία και παρουσιάζονται παρακάτω. Οι τύποι υπολογισμού τους είναι οι εξής:

$$RMSN = \frac{\sqrt{N \cdot \sum_{n=1}^N (Y_n^{obs} - Y_n^{sim})^2}}{\sum_{n=1}^N Y_n^{obs}} \quad (3.8)$$

$$RSMPE = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^N \left(\frac{Y_n^{sim} - Y_n^{obs}}{Y_n^{obs}} \right)^2} \quad (3.9)$$

$$MPE = \frac{1}{N} \cdot \sum_{n=1}^N \left(\frac{Y_n^{sim} - Y_n^{obs}}{Y_n^{obs}} \right) \quad (3.10)$$

$$U = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^{sim} - Y_n^{obs})^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n^{sim})^2 + \frac{1}{N} \sum_{n=1}^N (Y_n^{obs})^2}} \quad (3.11)$$

$$U^M = \frac{(\bar{Y}_n^{sim} - \bar{Y}_n^{obs})}{\frac{1}{N} \sum_{n=1}^N (Y_n^{sim} - Y_n^{obs})^2} \quad (3.12)$$

$$U^S = \frac{(\sigma^{sim} - \sigma^{obs})^2}{\frac{1}{N} \sum_{n=1}^N (Y_n^{sim} - Y_n^{obs})^2} \quad (3.13)$$

$$U^C = \frac{2(1-p)\sigma_{sim}\sigma_{obs}}{\frac{1}{N}\sum_{n=1}^N(Y_n^{sim} - Y_n^{obs})^2} \quad (3.14)$$

Όπου Y^{obs} είναι οι πραγματικές παρατηρήσεις και Y^{sim} είναι οι εκτιμήσεις του μοντέλου και N ο αριθμός των συνολικών παρατηρήσεων κατά την περίοδο αναφοράς.

Χρησιμοποιούνται διάφορα μέτρα αξιολόγησης, καθώς αξιολογούν από διαφορετική άποψη τα αποτελέσματα. Το κανονικοποιημένο μέσο τετραγωνικό σφάλμα (normalized root mean square error- RMSN) υπολογίζει τη διαφορά μεταξύ των πραγματικών παρατηρήσεων και των εκτιμήσεων της κάθε μεθόδου και με αυτόν τον τρόπο αξιολογεί το συνολικό λάθος, τις επιδόσεις της κάθε μεθόδου (Pindyck and Rubinfeld, 1997), (Huang et al., 2009), (Barcelo, 2010). Το μέσο τετραγωνικό σφάλμα εκτίμησης (root mean square percentage error- RMSPE) δίνει μεγαλύτερες τιμές για μεγάλες αποκλίσεις, ενώ μικρότερες τιμές για μικρές αποκλίσεις. Το μέσο σφάλμα εκτίμησης (mean percent error- MPE) υποδεικνύει την ύπαρξη συστηματικής υπό/υπέρ-τίμησης της εκτίμησης (Toledo, 2003). Τα % σφάλματα προτιμώνται καθώς δίνουν η αξιολόγησή τους αναφέρεται σε σχέση με τη μέση μέτρηση (Vikrant et al., 2009)

Επίσης, το μέτρο του συντελεστή ανισότητας Theil (U) έχει εφαρμοστεί για την επιβεβαίωση εγκυρότητας κυκλοφοριακών μοντέλων, συνδυάζοντας τρεις αναλογίες του μέτρου του Theil, πόλωσης (U_M), μεταβλητότητας (U_S) και συμμεταβλητότητας (U_C). Εξ ορισμού οι τρεις αυτές αναλογίες έχουν άθροισμα ίσο με 1. Το U_M αντιπροσωπεύει το συστηματικό σφάλμα, το U_S υποδεικνύει την ικανότητα προσομοίωσης στην επανάληψη της μεταβλητότητας των παρατηρήσεων, το U_C μετράει το σφάλμα που παραμένει. Τα τρία αυτά μέτρα παίρνουν τιμές ανάμεσα σε 0 και 1, όπου το 0 αντιστοιχεί στην ιδανική προσαρμογή, ενώ το 1 στην χειρότερη για τα μέτρα U_S , U_M (Theil, 1978).

3.4. Πειραματική διάταξη

Ακολουθεί συνοπτική περιγραφή και κωδικοποίηση της πειραματικής διάταξης:

- Σειρές δεδομένων για βαθμονόμηση: B1695
- Σειρές δεδομένων για έλεγχο εγκυρότητας: C621, A358, A172, C168, C171

- Τύποι μορφών μοντέλων: Μοντέλο του Gijbbs, προτεινόμενο μοντέλο με πολλές παραμέτρους, προτεινόμενο μοντέλο με λίγες παραμέτρους, προτεινόμενο μοντέλο με εφαρμογή ομαδοποίησης δεδομένων
- Μέτρα αξιολόγησης: RMSN, RMSPE, MPE, U, U_S, U_M, U_C
- Παράμετροι loess: span, degree
- Παράμετροι Gijbbs: τ , V_n , α_n , s_{n-1} , b_n και \hat{b}
- Κωδικοποίηση μοντέλων:
 - Μοντέλο Gijbbs g1*- Βαθμονομημένο μοντέλο Gijbbs για $\tau=0.4$ s που επιλέγεται
 - Μοντέλο Gijbbs g2*- Βαθμονομημένο μοντέλο Gijbbs για $\tau=0.4$ s που δεν επιλέγεται
 - Μοντέλο Gijbbs G3*- Βαθμονομημένο μοντέλο Gijbbs για $\tau=1.0$ s που επιλέγεται
 - Μοντέλο Gijbbs G4*- Βαθμονομημένο μοντέλο Gijbbs για $\tau=1.0$ s που δεν επιλέγεται
 - Μοντέλο loess full*- Προτεινόμενο μοντέλο loess με πολλές παραμέτρους
 - Μοντέλο loess 1 full (l1s)*- Προτεινόμενο μοντέλο loess με πολλές παραμέτρους για $\tau=0.4$ s
 - Μοντέλο loess 2 full (l2s)*- Προτεινόμενο μοντέλο loess με πολλές παραμέτρους για $\tau=1.0$ s
 - Μοντέλο loess simple*- Απλουστευμένο προτεινόμενο μοντέλο με λίγες παραμέτρους
 - Μοντέλο loess 1 simple (l1s)*- Απλουστευμένο προτεινόμενο μοντέλο με λίγες παραμέτρους για $\tau=0.4$ s
 - Μοντέλο loess 2 simple (l2s)*- Απλουστευμένο προτεινόμενο μοντέλο με λίγες παραμέτρους για $\tau=1.0$ s

4. ΕΦΑΡΜΟΓΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο παρόν κεφάλαιο αντικείμενο μελέτης αποτελεί η εκτίμηση της ταχύτητας ενός οχήματος τόσο με τη χρήση μαθηματικών μοντέλων, συγκεκριμένα του μοντέλου

Girps, όσο και με τη χρήση μεθόδων μηχανική μάθησης, όπως η τοπικά σταθμισμένη παλινδρόμηση. Για τον σκοπό αυτό χρησιμοποιούνται κάποια διαθέσιμα δεδομένα που προέκυψαν από ένα πείραμα στην πόλη Naples της Ιταλίας. Ειδικότερα, επιχειρείται μια αξιολόγηση και σύγκριση των δύο παραπάνω προτεινόμενων μεθόδων ως προς την αποτελεσματικότητά τους αναφορικά με τα διαθέσιμα δεδομένα.

4.1. Περιγραφή και ακρίβεια δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν προέκυψαν από μια σειρά πειραμάτων που διεξήχθη σε οδούς που περιβάλλουν την πόλη Νάπολη της Ιταλίας σε πραγματικές κυκλοφοριακές συνθήκες τον Οκτώβριο του 2002. Τα δεδομένα συλλέχτηκαν από 4 οχήματα που κινούνται σε σειρά υπό διάφορες κυκλοφοριακές συνθήκες. Αξιοσημείωτο είναι ότι τα οχήματα κινούνταν σε οδούς μίας λωρίδας ανά κατεύθυνση και κατά συνέπεια η συμπεριφορά των οδηγών δεν επηρεάζεται από αλλαγές λωρίδας. Όλα τα οχήματα ήταν εξοπλισμένα με δέκτες GPS που κατέγραφαν τη θέση του κάθε οχήματος ανά 0.1s. Πρόκειται για συσκευές διπλής συχνότητας GPS + GLONASS με οριζόντια ακρίβεια 10 mm+1.0 ppm και υψομετρική ακρίβεια 15 mm+1.0 ppm. Εξαιτίας περιβαλλοντικών συνθηκών στα παραπάνω δεδομένα υπήρχαν κενά, δηλαδή για κάποια χρονικά διαστήματα του πειράματος δεν καταγράφηκε μέτρηση. Ωστόσο, για τον σκοπό της εργασίας, προτιμήθηκε να αποκοπούν όσο το δυνατόν μεγαλύτερα κομμάτια με συνεχείς πραγματικές μετρήσεις παρά να χρησιμοποιηθεί κάποια μέθοδος γραμμικής ή πολυωνυμικής παρεμβολής για την εκτίμηση των μετρήσεων που έλειπαν. Κατά συνέπεια τα δεδομένα που χρησιμοποιούνται στην εργασία είναι δεδομένα που προέκυψαν από το πεδίο χωρίς διορθώσεις. Αναλυτική περιγραφή των δεδομένων διατίθεται από τους Punzo et al (2005).

Συνήθως υπάρχουν τρεις καταστάσεις οδήγησης, η πρώτη αναφέρεται στην ελεύθερη οδήγηση, η δεύτερη στην κατάσταση που ένα όχημα περιορίζεται από το μπροστινό του όχημα και η τρίτη για την περίπτωση επείγουσας επιβράδυνσης με σκοπό την αποφυγή ατυχήματος (Olstam and Tapani, 2004), (Koutsopoulos and Farah, 2012). Αντικείμενο της εργασίας αυτής αποτελεί η δεύτερη κατάσταση, καθώς τα δεδομένα αναφέρονται σε τέσσερα οχήματα που κινούνται σε σειρά. Πιο συγκεκριμένα η ταχύτητα του τρίτου οχήματος σε σειρά είναι η ζητούμενη ταχύτητα προς εκτίμηση.

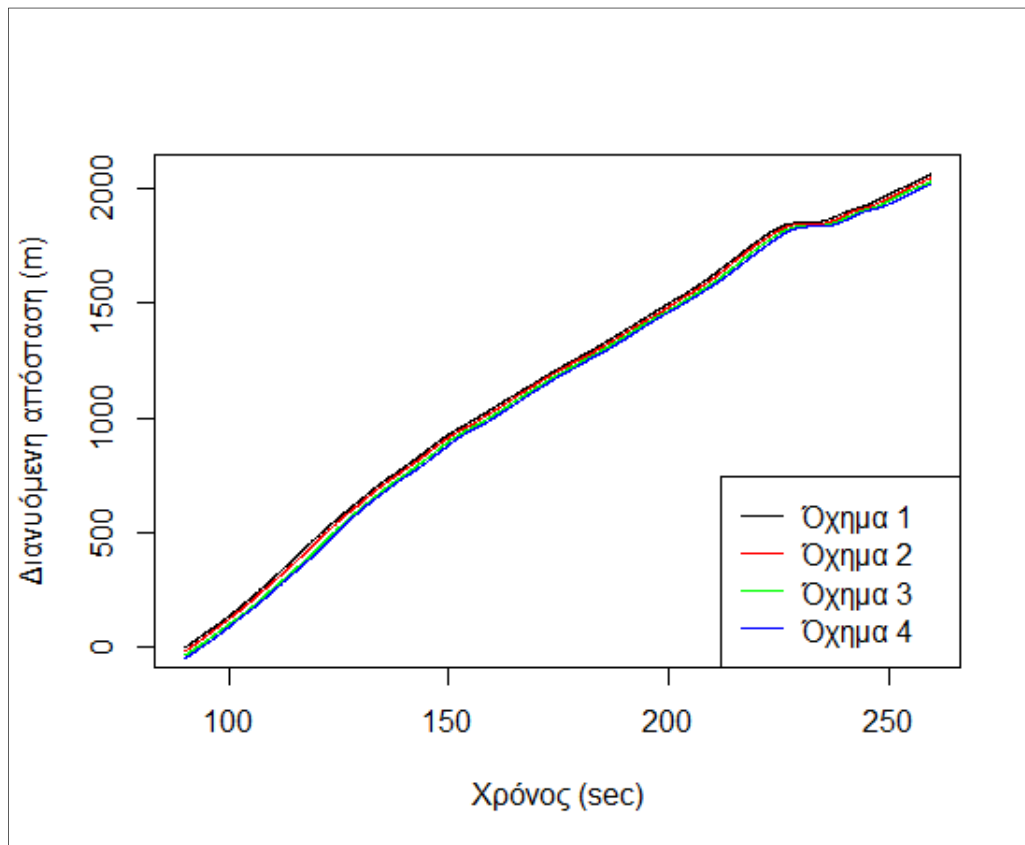
4.2. Υπολογισμός παραγόμενων στοιχείων

Από τα δεδομένα που διατίθενται χρησιμοποιήθηκαν οι μετρήσεις που αφορούν τη θέση των οχημάτων κάθε χρονική στιγμή (X, Y, Z). Σύμφωνα με τις συντεταγμένες αυτές υπολογίστηκαν οι αποστάσεις μεταξύ των οχημάτων, η διανύμενη απόσταση ανά 0.1s για το κάθε όχημα, η ταχύτητα τους και η σχετική τους θέση στη διαδρομή που διανύουν.

Επειδή οι συντεταγμένες που μετρήθηκαν αναφέρονται στην κεραία του κάθε οχήματος, αρχικά υπολογίζεται η απόσταση μεταξύ των κεραιών δύο οχημάτων i, j από τον τύπο:

$$d_{ij}(t) = \sqrt{(X_j(t) - X_i(t))^2 + (Y_j(t) - Y_i(t))^2 + (Z_j(t) - Z_i(t))^2} \quad (4.1)$$

Για την εύρεση της *πραγματικής απόστασης* (D_{ij}) μεταξύ των οχημάτων από την απόσταση που υπολογίστηκε από τη σχέση (4.1) πρέπει να αφαιρεθεί ακόμα η απόσταση του πίσω μέρους του προπορευόμενου οχήματος από την κεραία του και η απόσταση του μπροστινού μέρους του πίσω οχήματος από την κεραία του, όπως φαίνεται στο σχήμα 4.1. που ακολουθεί παρακάτω. Ελέγχθηκε η απόσταση μεταξύ των οχημάτων να είναι πάντα θετική και οι τροχιές τους να μην τέμνονται (Διάγραμμα 4.1.). Σε διαφορετική περίπτωση θα σήμαινε ότι υπάρχει σύγκρουση των οχημάτων και προφανώς θα υπήρχε κάποιο χονδροειδές σφάλμα στα δεδομένα.



Διάγραμμα 4.1.: Απεικόνιση της διανυόμενης απόστασης των τεσσάρων οχημάτων σε σχέση με το χρόνο

Η απόσταση που διανύθηκε ανάμεσα σε δυο διαδοχικές χρονικές στιγμές ($t_2-t_1=0.1s$) για το κάθε όχημα υπολογίζεται από τη σχέση 4.2.

$$s_i(t_2) = \sqrt{(X_i(t_2) - X_i(t_1))^2 + (Y_i(t_2) - Y_i(t_1))^2 + (Z_i(t_2) - Z_i(t_1))^2} \quad (4.2)$$

Η ταχύτητα του κάθε οχήματος για κάθε χρονική στιγμή υπολογίζεται από τη σχέση (4.3).

$$v_i(t_2) = \frac{s_i(t_2)}{\Delta t}, \text{ όπου } \Delta t=0.1s \quad (4.3)$$

Επίσης, υπολογίζεται η χιλιομετρική θέση (ss) του κάθε οχήματος για τη διαδρομή που διανύουν, δηλαδή πρόκειται για το άθροισμα των αποστάσεων που διανύουν ανάμεσα σε διαδοχικές χρονικές στιγμές. Ως αρχή έναρξης της χιλιομετρικής μέτρησης ορίζεται το σημείο έναρξης της διαδρομής για το πρώτο όχημα, στο σημείο

$ss = 0$, όπως φαίνεται στο σχήμα 4.1. Επομένως, για το όχημα 1, η χιλιομετρική θέση του υπολογίζεται από τη σχέση 4.4.

$$ss_1(t_i) = \sum_0^{t_i} s_1(t_i) \quad (4.4)$$

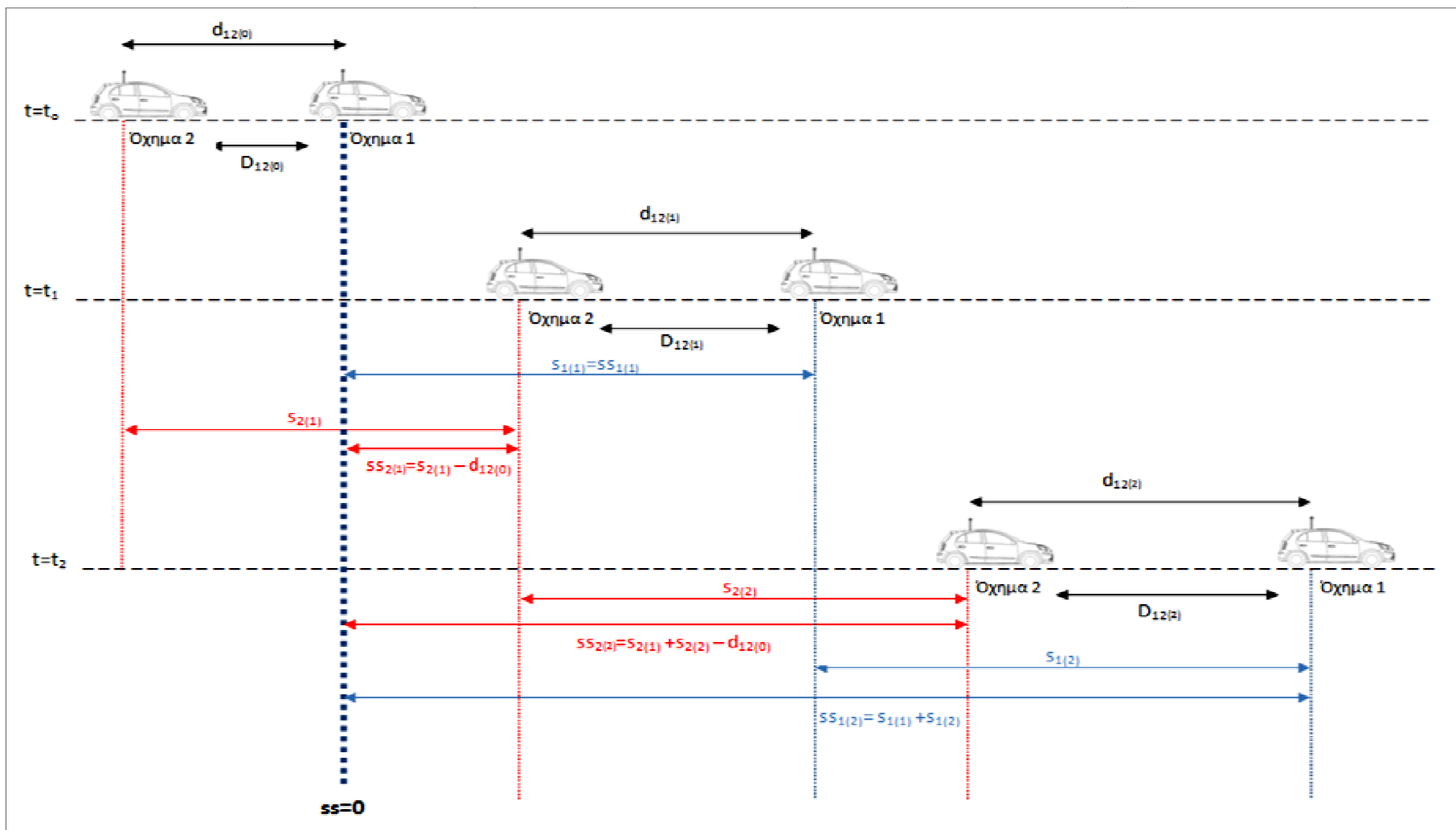
Επειδή, όμως, τα υπόλοιπα οχήματα ξεκινούν πριν από το σημείο έναρξης της χιλιομετρικής μέτρησης, για τον ορισμό της χιλιομετρικής θέσης τους θα αθροιστούν οι διανυόμενες αποστάσεις των επιμέρους χρονικών διαστημάτων όπως έγινε για το όχημα 1, αλλά θα αφαιρεθεί η απόσταση που διένυσαν μέχρι να φτάσουν στο σημείο έναρξης, δηλαδή την αρχική απόσταση που είχαν από το πρώτο όχημα τη χρονική στιγμή t_0 , σύμφωνα με το σχήμα 4.1. Επομένως, προκύπτουν οι σχέσεις:

$$ss_2(t_i) = \sum_0^{t_i} s_2(t_i) - d_{12}(0) \quad (4.5)$$

$$ss_3(t_i) = \sum_0^{t_i} s_3(t_i) - d_{13}(0) \quad (4.6)$$

$$ss_4(t_i) = \sum_0^{t_i} s_4(t_i) - d_{14}(0) \quad (4.7)$$

Σημειώνεται ότι οι παραπάνω υπολογισμοί έγιναν με τη βοήθεια του λογισμικού R project for statistical computing, ενώ χρησιμοποιήθηκε η διεπαφή Rstudio. Πλέον έχει δημιουργηθεί ένα αρχείο στο οποίο έχουν καταγραφεί και υπολογιστεί οι συντεταγμένες των οχημάτων, οι χιλιομετρικές τους θέσεις, οι ταχύτητες τους καθώς και οι μεταξύ τους αποστάσεις για κάθε χρονική στιγμή.



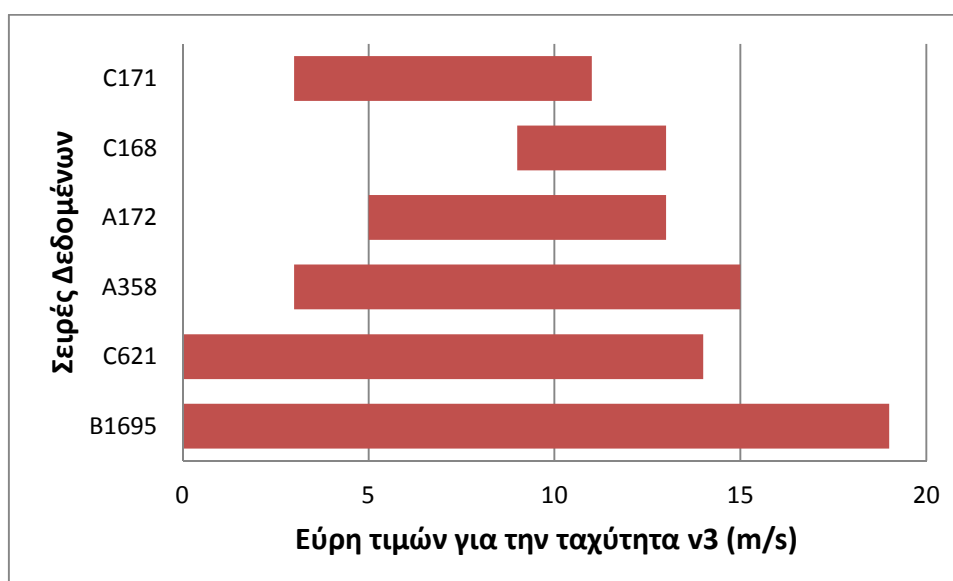
Σχήμα 3.1: Σκαρίφημα σχετικών αποστάσεων πειράματος

4.3. Επιλογή δεδομένων για βαθμονόμηση

Ως σειρά δεδομένων για βαθμονόμηση στις δύο προτεινόμενες μεθοδολογίες επιλέγεται το B1695 επειδή αποτελεί το εκτενέστερο σε μέγεθος αρχείο δεδομένων που διατίθεται καθώς και περιλαμβάνει το μεγαλύτερο εύρος ταχυτήτων. Η επιλογή αυτή γίνεται έτσι ώστε το μοντέλο που θα προκύψει να είναι περισσότερο αντιπροσωπευτικό και να μπορεί να προβλέψει περισσότερες καταστάσεις. Συγκεκριμένα, το μέγεθος και τα εύρη ταχυτήτων για κάθε σειρά δεδομένων φαίνονται στον πίνακα 4.1. και στο διάγραμμα 4.2.:

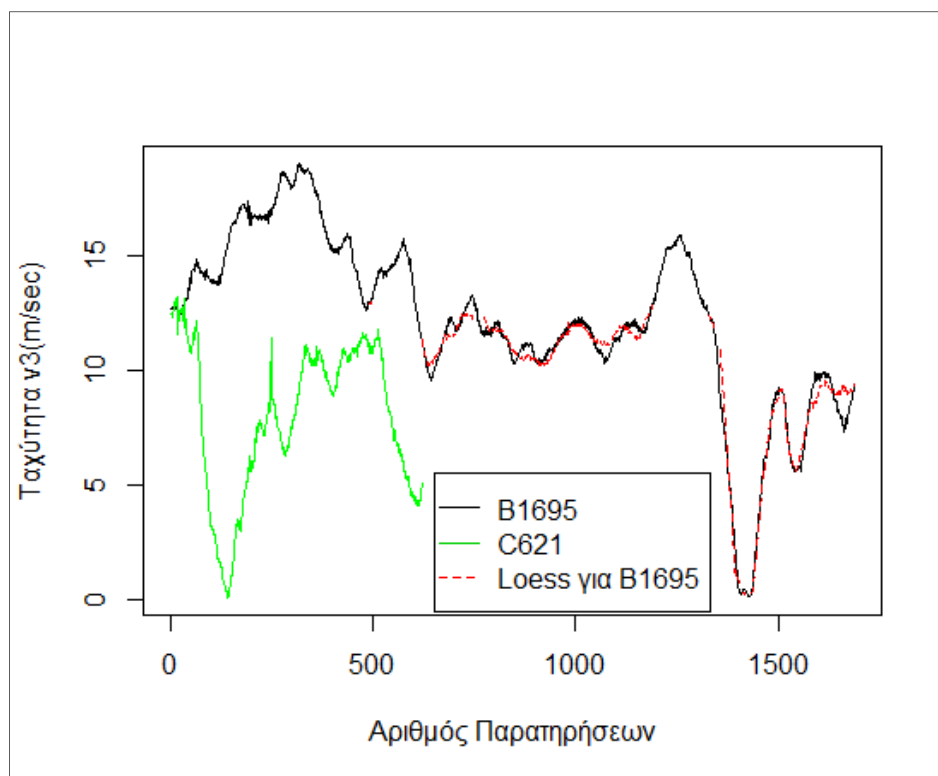
Σειρές δεδομένων	Πλήθος παρατηρήσεων	Χρονική διάρκεια δεδομένων (s)
B1695	1695	169.4
C621	621	62.0
A358	358	35.7
A172	172	17.1
C168	168	16.7
C171	171	17.0

Πίνακας 4.1.: Ονομασία και χρονική διάρκεια σειρών δεδομένων



Διάγραμμα 4.2.: Εύρη τιμών της ταχύτητας v3 για κάθε σειρά δεδομένων

Όσον αφορά στην εφαρμογή της προτεινόμενης μεθοδολογίας με την τοπικά σταθμισμένη παλινδρόμηση (loess), δημιουργούνται κάποια πρότυπα και σύμφωνα με αυτά γίνεται η εκτίμηση στα νέα δεδομένα που εισάγονται. Κατά συνέπεια αν επιλεγόταν ως μοντέλο βαθμονόμησης κάποιο από τα υπόλοιπα αρχαία δεδομένων και γινόταν προσπάθεια επιβεβαίωσης της εγκυρότητας του μοντέλου αυτού στο B1695, τότε δεν θα ήταν δυνατή η εκτίμηση των ταχυτήτων που δεν συναντώνται στο μοντέλο βαθμονόμησης, δηλαδή οι ταχύτητες πάνω από 15 m/s δεν θα μπορούσαν να προβλεφθούν. Για παράδειγμα, στο παρακάτω σχήμα γίνεται εκτίμηση της ταχύτητας ενός οχήματος από τη σειρά δεδομένων B1695 χρησιμοποιώντας ως πρότυπο μοντέλο αυτό που προέκυψε σύμφωνα με τη σειρά δεδομένων C621. Όπως φαίνεται στο σχήμα, οι ταχύτητες που δεν υπάρχουν στο C621 δεν μπορούν να προβλεφθούν με το loess για το B1695.



Διάγραμμα 4.3.: Μη δυνατότητα εκτίμησης ταχυτήτων πέρα από αυτές που περιλαμβάνονται στο μοντέλο βαθμονόμησης

4.4. Εφαρμογή του μοντέλου Gipps

4.4.1. Ανάλυση ευαισθησίας παραμέτρων

Στην ενότητα αυτή εξετάζονται όλοι οι πιθανοί συνδυασμοί παραμέτρων με κατώτερα και ανώτερα όρια αυτά, τα οποία επιλέχθηκαν σύμφωνα με τη βιβλιογραφία στην ενότητα 2.3.4 και παρουσιάστηκαν στον πίνακα 2.3., και αναζητείται ο βέλτιστος συνδυασμός τους για τη σειρά δεδομένων B1695, έτσι ώστε να επιτευχθεί η καλύτερη δυνατή εκτίμηση της ταχύτητας v_3 του τρίτου οχήματος. Δηλαδή αναζητείται ο συνδυασμός των παραμέτρων που δίνει το μικρότερο σφάλμα, έτσι ώστε το μοντέλο αυτό να χρησιμοποιηθεί ως μοντέλο αναφοράς για τη σύγκριση με την προτεινόμενη μεθοδολογία. Στο σημείο αυτό χρησιμοποιείται ως μέτρο αξιολόγησης το μέσο κανονικοποιημένο τετραγωνικό σφάλμα, καθώς αξιολογεί τη συνολική διαδικασία της προσομοίωσης.

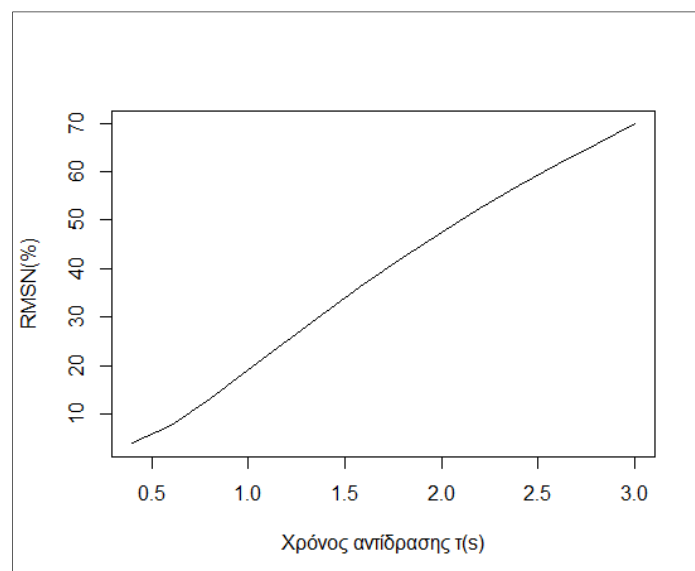
Ξεκινώντας από τις μέσες τιμές των παραμέτρων για το επιλεγόμενο εύρος ($\tau=1.7$ s, $V_n=20$ m/s, $\alpha_n=1.7$ m/s², $s_{n-1}=6.6$ m, $b_n=-3.4$ m/s² και $\hat{b}=-3.8$ m/s²) εξετάζεται πώς η κάθε παράμετρος επιδρά στην εφαρμογή του μοντέλου για τα δεδομένα αυτά. Για τις μέσες τιμές το σφάλμα RMSN που προκύπτει είναι 39.5%. Αναμφισβήτητα το σφάλμα είναι πολύ υψηλό και δεν έχει κάποια φυσική σημασία αλλά θα χρησιμοποιηθεί ως σημείο αναφοράς για το κατά πόσο αυτό μπορεί να βελτιωθεί με την τροποποίηση της κάθε παραμέτρου ξεχωριστά.

Στην ανάλυση ευαισθησίας που ακολουθεί ακολουθούν δύο εκδοχές. Στην πρώτη οι τιμές των επιβραδύνσεων και της επιτάχυνσης δεν συνδέονται με κάποια σχέση αλλά εξετάζονται η καθεμιά ξεχωριστά. Στη δεύτερη εκδοχή, οι επιβραδύνσεις δεν παίρνουν ανεξάρτητες τιμές αλλά υπολογίζονται από τους τύπους $b_n=-2\alpha_n$ και $\hat{b}=\min(-3.0, (b_n-3.0)/2)$, όπως είχε προτείνει ο Gipps (1981). Επίσης, οι Spyropoulou and Karlaftis (2008) στην εφαρμογή τους πρότειναν τον υπολογισμό των b , \hat{b} από αυτές τις δύο εξισώσεις που όρισε ο Gipps (1981), ενώ θεωρούν ότι σε περιπτώσεις με διαφορετικά όρια ταχύτητας θα ήταν καλύτερη η λήψη κάποιων σταθερών τιμών.

➤ **Εκδοχή α'**

Χρόνος αντίδρασης τ

Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και αλλάζοντας μόνο το χρόνο αντίδρασης, ο οποίος παίρνει τιμές 0.4-3 s προκύπτει το εξής διάγραμμα 4.4. Όπως είναι φανερό, ο χρόνος αντίδρασης διαδραματίζει καθοριστικό ρόλο, καθώς αλλάζοντας μόνο αυτήν την παράμετρο το σφάλμα μειώνεται δραματικά. Αυτό ήταν αναμενόμενο καθώς η παράμετρος αυτή συμμετέχει και στα δύο σκέλη της εξίσωσης του Girps. Επίσης, παρατηρείται ότι όσο μεγαλύτερος είναι ο χρόνος αντίδρασης, τόσο μικρότερο είναι το σφάλμα για την εκτίμηση της ταχύτητας v_3 . Για την τιμή 0.4 του χρόνου αντίδρασης, το σφάλμα γίνεται 10 φορές μικρότερο, 3.9%. Επομένως, το μοντέλο έχει την καλύτερη απόδοση για την τιμή 0.4.

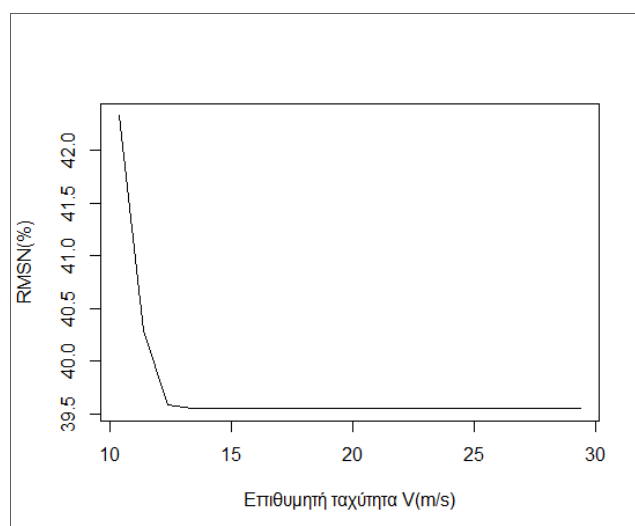


Διάγραμμα 4.4.: Ανάλυση ευαισθησίας του χρόνου αντίδρασης και του σφάλματος RMSN για τη σειρά δεδομένων B1695

Επειδή η τιμή 0.4 είναι ακραία, είναι αμφίβολο αν πρέπει να επιλεγεί η τιμή αυτή. Στο σημείο αυτό, όμως, αξίζει να αναφέρουμε ότι αναζητείται ο συνδυασμός των παραμέτρων που δίνει το μικρότερο σφάλμα, έτσι ώστε να βρεθεί το ελάχιστο σφάλμα που μπορεί να προκύψει από την εφαρμογή του μοντέλου του Girps. Ωστόσο, στη συνέχεια της εργασίας πέρα από την τιμή 0.4, θα εξεταστεί επίσης το μοντέλο αν ο χρόνος αντίδρασης πάρει μια πιο λογική τιμή, όπως 1 s.

Επιθυμητή Ταχύτητα V

Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και αλλάζοντας μόνο την επιθυμητή ταχύτητα V , η οποία παίρνει τιμές μέσα στο εύρος 10.4-29.6 m/s προκύπτει το διάγραμμα 4.5. Όπως παρατηρείται από το παρακάτω διάγραμμα η ταχύτητα δεν έχει τόσο μεγάλη επίδραση στο σφάλμα RMSN όπως ο χρόνος. Επίσης, αξιοσημείωτο είναι ότι το σφάλμα μένει ανεπηρέαστο για τις τιμές της ταχύτητας από 13 έως 29.6 m/s, ενώ για μικρότερες τιμές το σφάλμα αυξάνεται. Κατά συνέπεια, δεν εφικτή η επιλογή της τιμής της που δίνει το ελάχιστο σφάλμα, εάν δεν καθοριστούν πρώτα οι υπόλοιπες παράμετροι.

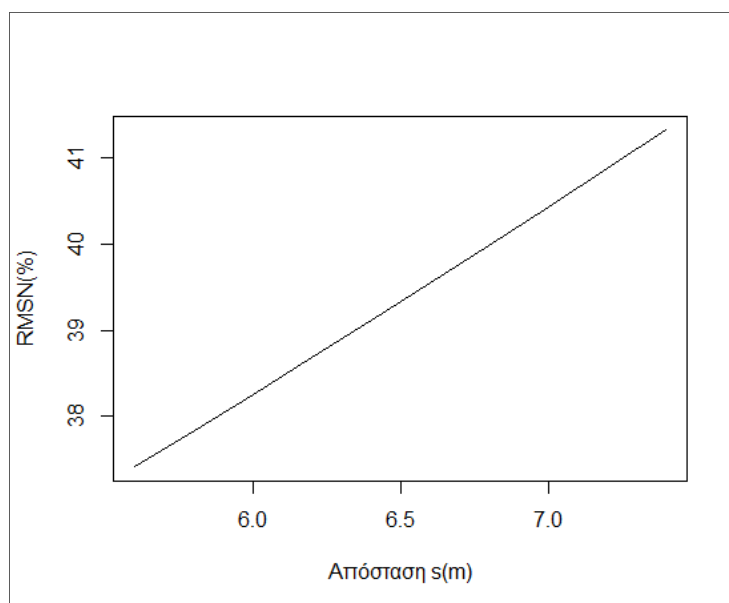


Διάγραμμα 4.5.: Ανάλυση ευαισθησίας της επιθυμητής ταχύτητας και του σφάλματος RMSN για τη σειρά δεδομένων B1695

Απόσταση s_{n-1}

Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και πειραματιζόμενοι μόνο με την απόσταση s_{n-1} , η οποία παίρνει τιμές μέσα στο εύρος 5.6-7.5m προκύπτει το διάγραμμα 4.6. Όπως παρατηρείται η παράμετρος αυτή έχει μικρή σχετικά επίδραση στο σφάλμα RMSN. Σύμφωνα με το διάγραμμα αυτό όσο μικρότερη είναι η τιμή της απόστασης, τόσο μειώνεται το σφάλμα. Κατά συνέπεια, για την καλύτερη απόδοση του μοντέλου σε αυτά τα δεδομένα, η τιμή της απόστασης

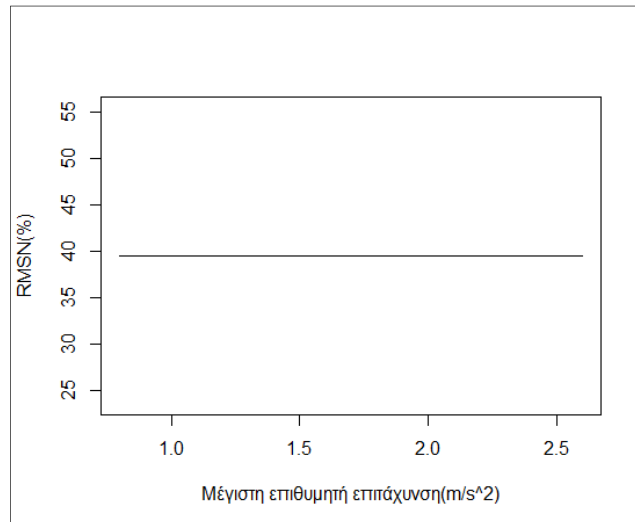
θα είναι 5.6 m. Για τη βαθμονόμηση του μοντέλου σίγουρα δεν θα επιλεγόταν μια ακραία τιμή. Ωστόσο, για τον σκοπό της εργασίας μπορεί να επιλεγθεί.



Διάγραμμα 4.6.: Ανάλυση ευαισθησίας της απόστασης και του σφάλματος RMSN για τη σειρά δεδομένων B1695

Μέγιστη επιθυμητή επιτάχυνση

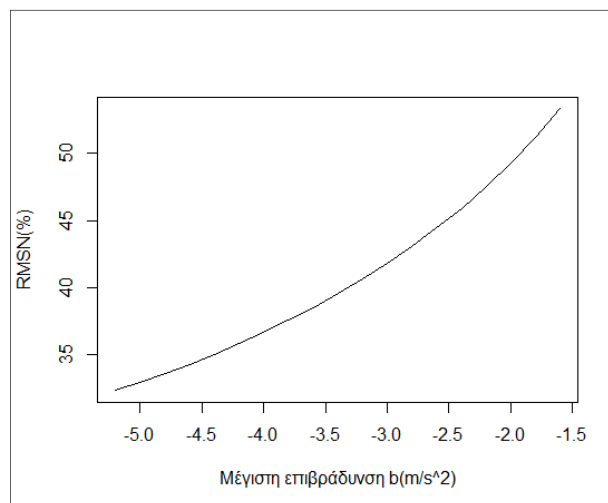
Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και αλλάζοντας μόνο τη μέγιστη επιθυμητή επιτάχυνση a , η οποία παίρνει τιμές μέσα στο εύρος 0.8-2.6 προκύπτει το εξής διάγραμμα 4.7. Όπως παρατηρείται η παράμετρος δεν έχει καμία επίδραση στο σφάλμα RMSN. Σύμφωνα με το διάγραμμα αυτό για όλες τις τιμές της επιτάχυνσης, το σφάλμα παραμένει σταθερό. Το γεγονός αυτό οφείλεται στο ότι η παράμετρος αυτή συναντάται μόνο στο πρώτο σκέλος της εξίσωσης και είναι πιθανό για αυτές τις τιμές των παραμέτρων η τιμή της ταχύτητας να προβλέπεται κατά κύριο λόγο από το δεύτερο σκέλος της εξίσωσης. Κατά συνέπεια θα πρέπει να καθοριστούν πρώτα οι υπόλοιπες παράμετροι και έπειτα να εξεταστεί πότε η επιτάχυνση ασκεί επίδραση στο σφάλμα RMSN του μοντέλου.



Διάγραμμα 4.7.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695

Μέγιστη επιθυμητή επιβράδυνση

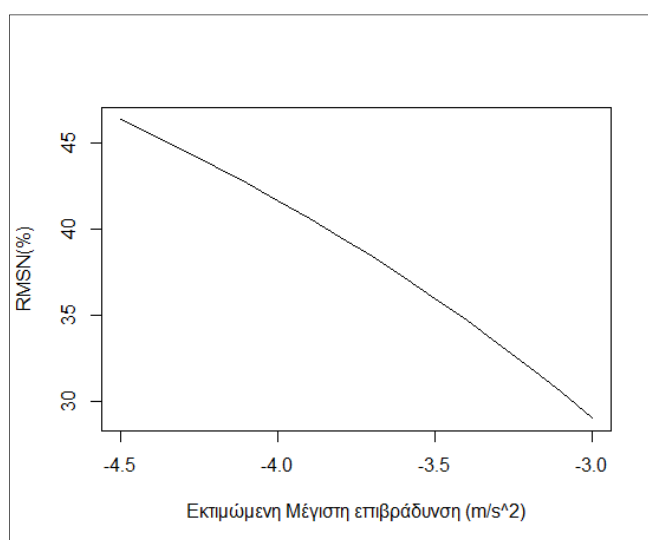
Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και αλλάζοντας μόνο τη μέγιστη πέδηση b_n , η οποία παίρνει τιμές μέσα στο εύρος -1.6 έως -5.2 προκύπτει το διάγραμμα 4.8. Όπως παρατηρείται από το παρακάτω διάγραμμα η πέδηση b_n έχει σχετικά μικρή επίδραση στο σφάλμα RMSN. Όπως φαίνεται όσο μεγαλύτερη είναι η τιμή της επιβράδυνσης, τόσο μειώνεται το σφάλμα. Κατά συνέπεια, για την καλύτερη απόδοση του μοντέλου σε αυτά τα δεδομένα, η τιμή της επιβράδυνσης b_n θα είναι 5.2 m/s^2 .



Διάγραμμα 4.8.: Ανάλυση ευαισθησίας της μέγιστης επιβράδυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695

Εκτιμώμενη μέγιστη επιβράδυνση

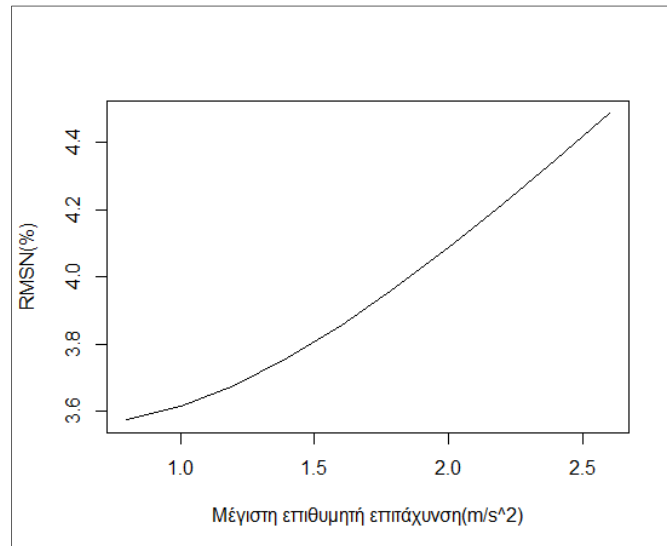
Αφήνοντας τις άλλες παραμέτρους σταθερές στις μέσες τιμές τους και αλλάζοντας μόνο την εκτιμώμενη μέγιστη πέδηση \hat{b} , η οποία παίρνει τιμές μέσα στο εύρος -3 έως -4.5 προκύπτει το διάγραμμα 4.9. Όπως παρατηρείται από το παρακάτω διάγραμμα όσο μειώνεται η εκτιμώμενη επιβράδυνση \hat{b} , τόσο μειώνεται το σφάλμα RMSN. Κατά συνέπεια, για την καλύτερη απόδοση του μοντέλου σε αυτά τα δεδομένα, η τιμή της επιβράδυνσης \hat{b} θα είναι 3 m/s^2 .



Διάγραμμα 4.9.: Ανάλυση ευαισθησίας της εκτιμώμενης μέγιστης επιβράδυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695

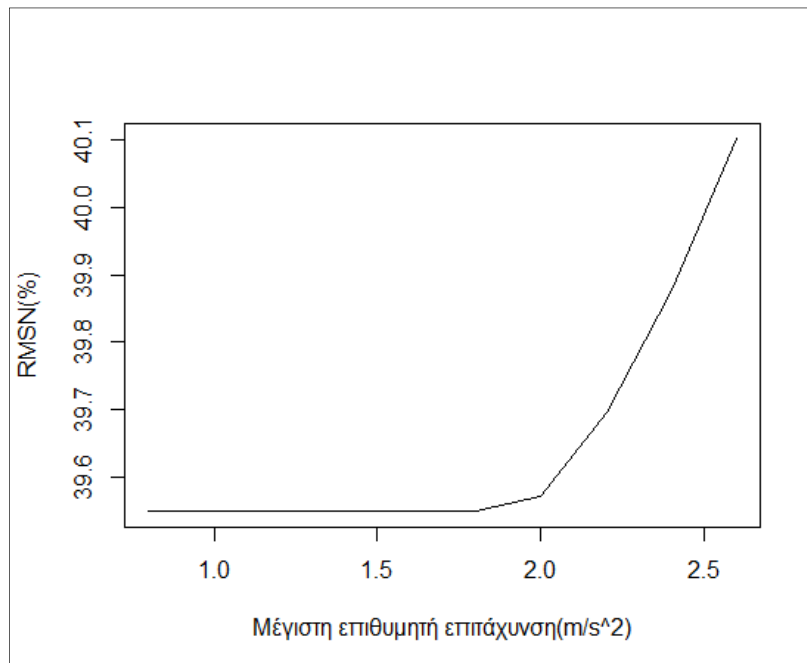
Τώρα πρέπει να εξεταστούν ξανά οι παράμετροι V_n , a_n για τις οποίες δεν υπάρχει ξεκάθαρη εικόνα. Αυτές οι δύο παράμετροι συμμετέχουν στη μία από τις δύο εξισώσεις του Gipps και επειδή φαίνεται ότι δεν επηρεάζουν είναι πιθανό η συμπεριφορά τους να εξαρτάται από τις τιμές των υπόλοιπων παραμέτρων.

Διατηρώντας τις μέσες τιμές των παραμέτρων αλλά θέτοντας τη βέλτιστη τιμή για κάθε παράμετρο από τις υπόλοιπες ξεχωριστά, παρατηρείται η συμπεριφορά της επιτάχυνσης. Συγκεκριμένα, αφήνοντας τις μέσες τιμές των παραμέτρων και θέτοντας τη βέλτιστη τιμή για το χρόνο αντίδρασης $\tau=0.4 \text{ s}$, η συμπεριφορά της επιτάχυνσης παρουσιάζεται στο διάγραμμα 4.10.



Διάγραμμα 4.10.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695(επανεξέταση)

Διατηρώντας τις μέσες τιμές των άλλων παραμέτρων αλλά θέτοντας $V_n=13\text{m/s}$, παρατηρείται ξανά η συμπεριφορά της επιτάχυνσης. Σύμφωνα με το διάγραμμα 4.11, για τιμές της επιτάχυνσης από 0.8 έως 1.8 m/s^2 το σφάλμα παραμένει σταθερό, ενώ για μεγαλύτερες τιμές αυξάνεται.

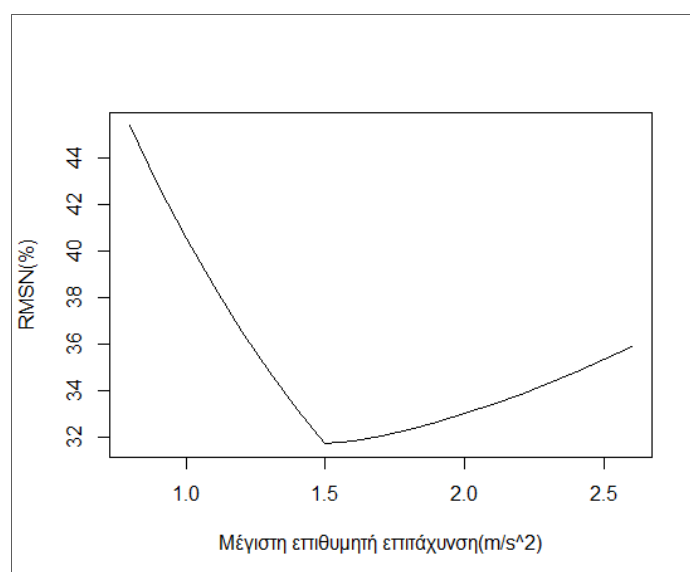


Διάγραμμα 4.11.: Ανάλυση ευαισθησίας της μέγιστης επιθυμητής επιτάχυνσης και του σφάλματος RMSN για τη σειρά δεδομένων B1695(επανεξέταση)

Αντιθέτως, επαναλαμβάνοντας την ίδια διαδικασία για τις υπόλοιπες παραμέτρους, η επιτάχυνση φαίνεται να μην ασκεί καμία επίδραση στο RMSN. Επομένως, η τιμή του χρόνου αντίδρασης και της ταχύτητας είναι αυτές που επηρεάζουν την επιλογή της τιμής της επιτάχυνσης. Αυτό ήταν αναμενόμενο αφού στο πρώτο σκέλος της εξίσωσης συμμετέχουν αυτά τα τρία μεγέθη και είναι λογικό να αλληλεπιδρούν. Έτσι, από τα δύο παραπάνω διαγράμματα επιλέγεται ως τιμή της επιτάχυνσης αυτή που δίνει το μικρότερο σφάλμα RMSN, δηλαδή $a_n=0.8 \text{ m/s}^2$.

➤ Εκδοχή β'

Τώρα λαμβάνοντας υπόψη ότι οι επιβραδύνσεις υπολογίζονται από τους τύπους $b_n = -2a_n$ και $\hat{b} = \min(-3.0, (b_n-3.0)/2)$, επαναλαμβάνεται η ίδια διαδικασία, δηλαδή κρατώντας σταθερά τις μέσες τιμές και εναλλάσσοντας κάθε παράμετρο ξεχωριστά, εξετάζεται πότε ελαχιστοποιείται το σφάλμα RMSN. Οι παράμετροι τ , V_n , s_{n-1} έχουν ακριβώς την ίδια συμπεριφορά που περιγράφηκε παραπάνω στην εκδοχή α' εκτός από την μέγιστη επιθυμητή επιτάχυνση που η συμπεριφορά της περιγράφεται στο διάγραμμα 4.13. Οι δύο επιβραδύνσεις υπολογίζονται αυτόματα συναρτήσει της επιτάχυνσης. Επομένως από το διάγραμμα είναι φανερό ότι η τιμή της επιτάχυνσης που ελαχιστοποιεί το σφάλμα RMSN είναι $a_n=1.5 \text{ m/s}^2$ και κατά συνέπεια οι τιμές για τις επιβραδύνσεις είναι $b_n=-3.0 \text{ m/s}^2$ και $\hat{b}=-3.0 \text{ m/s}^2$.

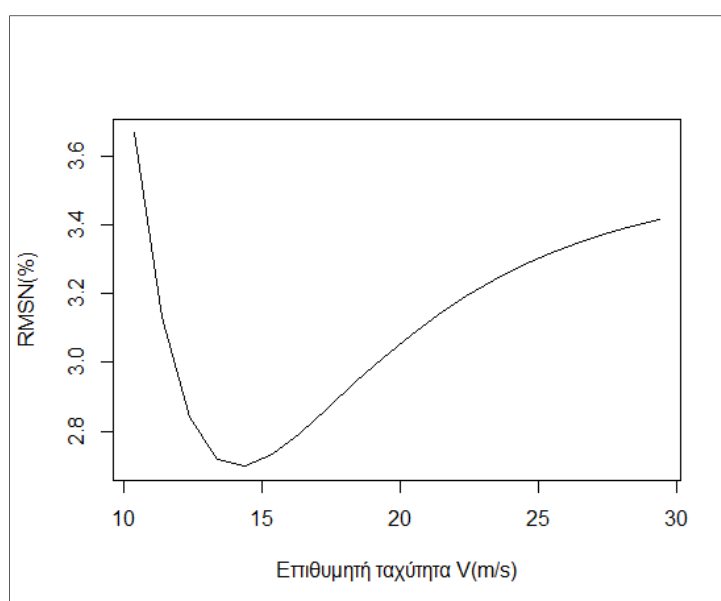


Διάγραμμα 4.12: Επιλογή τιμής για την μέγιστη επιθυμητή επιτάχυνση

4.4.2. Βαθμονόμηση μοντέλου

➤ Εκδοχή α'

Έχοντας καταλήξει για την εκδοχή α' για τις τιμές όλων των υπόλοιπων παραμέτρων και πιο συγκεκριμένα θέτοντας $\tau=0.4$ s, $a_n=0.8$ m/s², $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s² και $\hat{b}=-3.0$ m/s², η συμπεριφορά της ταχύτητας V παρουσιάζεται στο διάγραμμα 4.12. Σύμφωνα με το διάγραμμα αυτό η τιμή της ταχύτητας $V_n=14$ m/s ελαχιστοποιεί το σφάλμα για αυτά τα δεδομένα και αυτόν το συνδυασμό παραμέτρων.

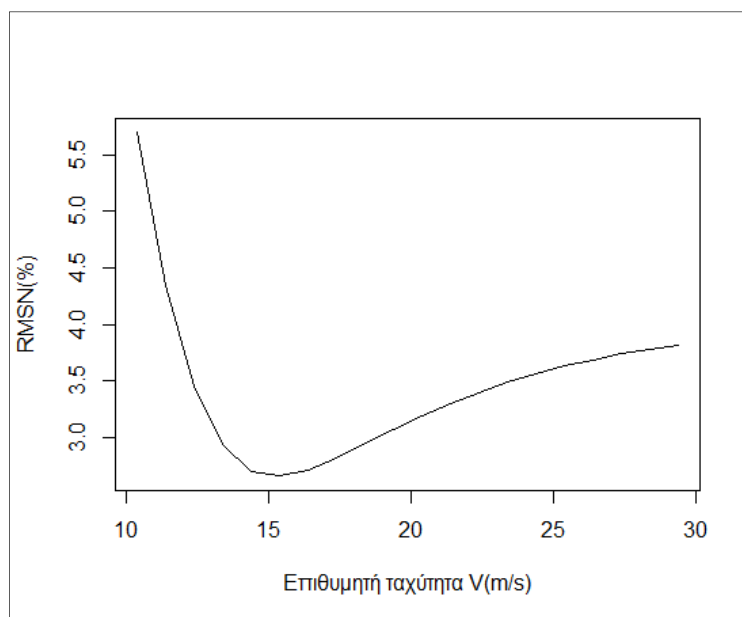


Διάγραμμα 4.13.: Επιλογή τιμής της παραμέτρου V_n (α' εκδοχή)

Η ίδια ακριβώς διαδικασία ακολουθήθηκε θεωρώντας τον χρόνο αντίδρασης $\tau=1$ s στοχεύοντας σε μια πιο ρεαλιστική επιλογή μοντέλου. Οι τιμές που προέκυψαν σε αυτήν την περίπτωση είναι $\tau=1.0$ s, $a_n=1.6$ m/s², $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s², $\hat{b}=-3.0$ m/s² και $V_n=16$ m/s.

➤ *Εκδοχή β'*

Για την εκδοχή β' οι τιμές των παραμέτρων που ελαχιστοποιούν το σφάλμα RMSN είναι $\tau=0.4\text{s}$, $s_{n-1}=5.6\text{m}$, $b_n=-3.0\text{m/s}^2$ και $\hat{b}=-3.0\text{m/s}^2$ και $\alpha_n=1.5\text{m/s}^2$. Για αυτές τις τιμές βρίσκεται η βέλτιστη τιμή της ταχύτητας V_n (Διάγραμμα 4.14). Οπότε σε αυτή την περίπτωση οι τιμές που δίνουν το μικρότερο σφάλμα για τη σειρά δεδομένων B1695 είναι $\tau=0.4\text{ s}$, $\alpha_n=1.5\text{ m/s}^2$, $s_{n-1}=5.6\text{ m}$, $b_n=-3.0\text{ m/s}^2$, $\hat{b}=-3.0\text{ m/s}^2$ και $V_n=15\text{ m/s}$.



Διάγραμμα 4.14.: Επιλογή τιμής της παραμέτρου V (β' εκδοχή)

Αντίστοιχα, για χρόνο αντίδρασης $\tau=1.0\text{ s}$, επαναλαμβάνοντας την ίδια διαδικασία, προκύπτουν οι τιμές $\tau=1.0\text{ s}$, $\alpha_n=2.3\text{ m/s}^2$, $s_{n-1}=5.6\text{m}$, $b_n=-4.6\text{m/s}^2$, $\hat{b}=-3.8\text{m/s}^2$ και $V_n=18\text{m/s}$.

Συνολικά από την παραπάνω ανάλυση οι τέσσερις συνδυασμοί που προέκυψαν παρουσιάζονται στον πίνακα 4.2. Οι επιλογές 1,2 αναφέρονται στις βέλτιστες και πιθανόν ακραίες τιμές των παραμέτρων, ενώ οι επιλογές 3,4 θεωρούνται πιο ρεαλιστικές, δεδομένου ότι για το χρόνο αντίδρασης τέθηκε μια πιο ρεαλιστική τιμή.

Παράμετροι	$\tau=0.4$ s		$\tau=1.0$ s	
	Επιλογή 1	Επιλογή 2	Επιλογή 3	Επιλογή 4
a_n (m/s ²)	0.8	1.5	1.6	2.3
b_n (m/s ²)	-5.2	-3	-5.2	-4.6
\hat{b} (m/s ²)	-3	-3	-3.0	-3.8
s_{n-1} (m)	5.6	5.6	5.6	5.6
V_n (m/s)	14	15	16	18
τ (s)	0.4	0.4	1.0	1.0

Πίνακας 4.2.: Επιλεγόμενες τιμές παραμέτρων για την εφαρμογή του μοντέλου του Girps στη σειρά δεδομένων B1695

Ενδεικτικά το μοντέλο του Girps μπορεί να εφαρμοστεί με τις μέσες τιμές των παραμέτρων όπως πρότεινε ο Girps (1981) ή όπως πρότειναν οι Ranjitkar et al. (2005), οι οποίες αναφέρονται στους πίνακες 2.1, 2.2. Επίσης, στο σημείο αυτό αξίζει να αναφερθεί ότι οι Ciuffo et al. (2012) αναλύοντας τα ίδια δεδομένα με αυτά που χρησιμοποιούνται σε αυτήν την εργασία και θέλοντας να βαθμονομήσουν το μοντέλο του Girps προσπάθησαν να βρουν τις βέλτιστες τιμές των παραμέτρων για τις σειρές δεδομένων «B» και «C». Αν και πρόκειται για τα ίδια δεδομένα οι Ciuffo et al. (2012) δεν είχαν αποκόψει τμήματα των δεδομένων με ελλιπείς μετρήσεις όπως έχει γίνει σε αυτήν την εργασία αλλά είχαν χρησιμοποιήσει όλα τα δεδομένα αφού πρώτα όμως είχαν εφαρμόσει ειδικά φίλτρα για την κάλυψη των μετρήσεων που έλειπαν και τη διόρθωση τους από τα σφάλματα. Χρησιμοποιώντας ένα πολύ εκτεταμένο εύρος τιμών για τις παραμέτρους κατέληξαν ότι για το κλασικό μοντέλο του Girps και τις σειρές δεδομένων «B» και «C» οι βέλτιστες τιμές που προσδιόρισαν φαίνονται στον πίνακα 4.3. Θα ήταν σκόπιμο να δοκιμαστούν οι τιμές αυτές και τελικά μετά από σύγκριση να επιλεγούν οι καλύτερες δυνατές.

Σειρά δεδομένων	Παράμετροι μοντέλου Gipps					
	a_n (m/s ²)	b_n (m/s ²)	\hat{b} (m/s ²)	s_{n-1} (m)	V_n (m/s)	τ (s)
B	4.73	-4.93	-8.00	Μήκος οχήματος+2.0	25	0.20
C	7.02	-5.62	-8.00	Μήκος οχήματος+1.0	19.25	0.60
Μέση τιμή	5.88	-5.28	-8.00	Μήκος οχήματος+1.5	22.13	0.40

Πίνακας 4.3.: Βαθμονομημένες τιμές για τις παραμέτρους του Gipps σύμφωνα με τους Ciuffo et al., 2012

Σημείωση: Ως μήκος οχήματος μπορεί να θεωρηθεί το μέσο μήκος των 4 οχημάτων που συμμετείχαν στο πείραμα, το οποίο ήταν 3.72 m.

Συνοψίζοντας, από την παραπάνω ανάλυση προέκυψαν οι πιθανοί συνδυασμοί παραμέτρων που φαίνονται στον πίνακα 4.4., καθώς και το RMSN για τη σειρά B1695 ανά περίπτωση. Στον παρακάτω πίνακα δε συμπεριλαμβάνονται οι μέσες τιμές από τους Ranjitkar et al. (2005), καθώς δίνουν πολύ μεγάλο σφάλμα για τη σειρά δεδομένων B1695. Μελετώντας τον παρακάτω πίνακα το μοντέλο του Gipps παρέχει καλύτερα αποτελέσματα όταν οι παράμετροι παίρνουν τις τιμές που προσδιορίστηκαν από την ανάλυση ευαισθησίας (Επιλογή 1, 2).

Παράμετροι	Επιλογή 1 (g1)	Επιλογή 2 (g2)	Επιλογή 3 (G3)	Επιλογή 4 (G4)	Gipps (1981)- Μέσες τιμές	Ciuffo et al. (2012) προτεινόμενες τιμές για τη σειρά δεδομένων B	Ciuffo et al. (2012) προτεινόμενες τιμές για τη σειρά δεδομένων C	Ciuffo et al. (2012)- Μέσες τιμές για τις σειρές δεδομένων B,C
α_n (m/s ²)	0.8	1.5	1.6	2.3	1.7	4.73	7.02	5.88
b_n (m/s ²)	-5.2	-3	-5.2	-4.6	-3.4	-4.93	-5.62	-5.28
\hat{b} (m/s ²)	-3	-3	-3.0	-3.8	-3.2	-8.00	-8.00	-8.00
s_{n-1} (m)	5.6	5.6	5.6	5.6	6.5	5.72	4.72	5.22
V_n (m/s)	14	15	16	18	20	25	19.25	22.13
τ (s)	0.4	0.4	1.0	1.0	2/3	0.2	0.6	0.4
RMSN (%)	2.7	2.7	4.9	9.6	5.7	6.0	10.1	7.4

Πίνακας 4.4.: Πιθανοί συνδυασμοί παραμέτρων για την εφαρμογή του μοντέλου του Gipps και το αντίστοιχο σφάλμα RMSN

Ωστόσο, ο βέλτιστος συνδυασμός για μια σειρά δεδομένων δεν αποτελεί απαραίτητα το βέλτιστο συνδυασμό για μια άλλη σειρά δεδομένων, λόγω της υπερπροσαρμογής (over-fitting). Κατά συνέπεια στην επόμενη ενότητα θα δοκιμαστούν όλοι συνδυασμοί του παραπάνω πίνακα και θα επιλέγει εκείνος που δίνει το ελάχιστο δυνατό σφάλμα για την πλειοψηφία των δεδομένων. Άλλωστε ένας λόγος για τον μεγάλο αριθμό των προτεινόμενων κατά καιρούς “car-following” μοντέλων είναι η αδυναμία τους να προβλέψουν τη ροή της κυκλοφορίας και τις αλληλεπιδράσεις των οδηγών χωρίς να βασιστούν στην “υπερπροσαρμογή (over-fitting)” των παραμέτρων στα εκάστοτε δεδομένα, η οποία είναι περιττή και χωρίς φυσική σημασία (Ciuffo et al., 2012).

4.4.3. Έλεγχος εγκυρότητας μοντέλου

Στην προηγούμενη ενότητα για τη σειρά δεδομένων B1695 προέκυψαν οι βέλτιστες τιμές των παραμέτρων, δηλαδή αυτές που δίνουν το μικρότερο σφάλμα συγκρίνοντας τις πραγματικές μετρήσεις με τις τιμές της ταχύτητας που προκύπτουν από το μοντέλο Gipps. Στη συνέχεια θα εξεταστεί η εγκυρότητα του κάθε μοντέλου στις υπόλοιπες σειρές δεδομένων και θα επιλεγεί αυτό με τις καλύτερες τιμές των παραμέτρων, δηλαδή αυτές που δίνουν το μικρότερο σφάλμα για την πλειοψηφία των δεδομένων που διατίθενται. Για τους συνδυασμούς του πίνακα τώρα εξετάζεται η εγκυρότητα του κάθε μοντέλου.

	RMSN(%)								
	Σειρές Δεδομένων	Επιλογή 1 (g1)	Επιλογή 2 (g2)	Επιλογή 3 (G3)	Επιλογή 4 (G4)	Gipps (1981)- Μέσες τιμές	Ciuffo et al. (2012) προτεινόμενες τιμές για τη σειρά δεδομένων B	Ciuffo et al. (2012) προτεινόμενες τιμές για τη σειρά δεδομένων C	Ciuffo et al. (2012)- Μέσες τιμές για τις σειρές δεδομένων B,C
<i>Calibration</i>	B1695	2.7	2.7	4.9	9.6	5.7	6.0	10.1	7.4
<i>Validation</i>	C621	6.6	5.8	14.4	21.9	15.5	9.2	15.6	12.5
	A358	2.7	3.5	12.7	16.6	11.3	9.2	8.7	6.9
	A172	4.6	6.1	16.0	23.2	15.3	11.7	15.1	13.9
	C168	2.3	2.7	4.9	18.8	10.4	5.9	20.1	12.4
	C171	7.2	10.2	31.6	44.0	32.3	12.0	29.6	19.7

Πίνακας 4.5.: Έλεγχος εγκυρότητας του μοντέλου του Gipps για τους προτεινόμενους συνδυασμούς παραμέτρων για όλες τις σειρές δεδομένων

Παρατηρώντας τα αποτελέσματα του πίνακα 4.5, γίνεται αντιληπτό ότι ο χρόνος αντίδρασης διαδραματίζει καθοριστικό ρόλο στην εφαρμογή του μοντέλου και στην εκτίμηση Επομένως, όπως φαίνεται από τον παραπάνω πίνακα η επιλογή 1 (μοντέλο g1) παρέχει την καλύτερη εκτίμηση με το μικρότερο σφάλμα για την πλειοψηφία των δεδομένων και θα χρησιμοποιηθεί σαν μοντέλο αναφοράς για τη σύγκριση με τη

μέθοδο που θα εφαρμοστεί στη συνέχεια. Επίσης, κρίνεται σκόπιμο να γίνει σύγκριση και με την Επιλογή 3 (μοντέλο G3), για χρόνο αντίδρασης $\tau=1.0$ s, για μια πιο ρεαλιστική προσέγγιση.

4.5. Εφαρμογή τοπικά σταθμισμένης παλινδρόμησης (loess)

Με τη βοήθεια της τοπικά σταθμισμένης παλινδρόμησης (locally weighted regression) επιχειρείται η εκτίμηση της ταχύτητας του τρίτου σε σειρά οχήματος. Για να επιτευχθεί αυτό πρέπει αρχικά να αποφασιστεί σύμφωνα με ποιους παράγοντες θα γίνει η εκτίμηση, δηλαδή ποιοι παράγοντες θα χρησιμοποιηθούν στην παλινδρόμηση, π.χ. η ταχύτητα του μπροστινού ή του πίσω οχήματος, η απόσταση του τρίτου οχήματος από το μπροστινό ή το πίσω όχημα, κ.α. Στην πραγματικότητα αναζητούνται εκείνα τα μεγέθη που επηρεάζουν περισσότερο την ταχύτητα του τρίτου οχήματος, έτσι ώστε όταν αυτά χρησιμοποιηθούν στην παλινδρόμηση να μπορεί να επιτευχθεί αξιόπιστη εκτίμηση της ζητούμενης ταχύτητας. Για τον σκοπό αυτό θα δοκιμαστούν διάφοροι συνδυασμοί και θα επιλεγεί ο καλύτερος δυνατός.

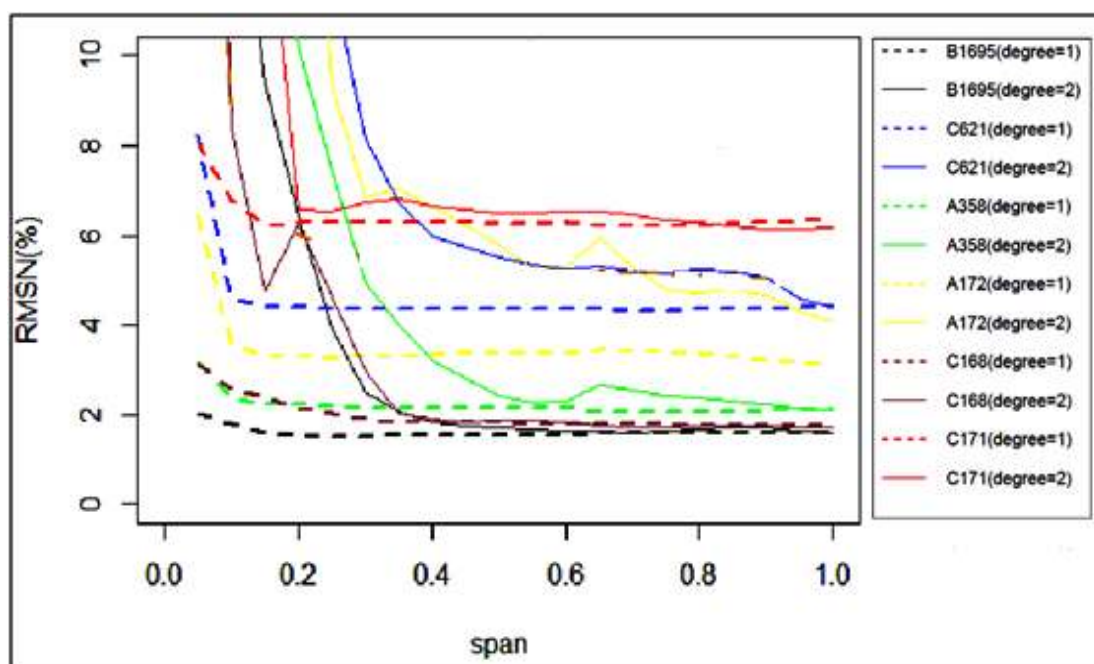
Επιπλέον, πρέπει να καθοριστούν οι τιμές για τις παραμέτρους *span* (ή α) και *degree* της παλινδρόμησης στο R. Οι τιμές αυτές επιλέγονται έτσι ώστε η καμπύλη που προκύπτει από την παλινδρόμηση να ταιριάζει καλύτερα στα δεδομένα. Η παράμετρος *span* καθορίζει πόσο ομαλή θα είναι η καμπύλη και η τιμή της κυμαίνεται από 0 (κυματοειδής καμπύλη) έως 1 (ομαλή καμπύλη), (Wickham H., 2009), (SAS Institute Inc.,1999). Συγκεκριμένα, ένα *span* α σημαίνει ότι σε κάθε τοπική προσαρμογή της καμπύλης λαμβάνονται υπόψη $\alpha \times 100\%$ των δεδομένων (Irizarry and Bravo, 2010), δηλαδή καθορίζει το μέγεθος της «περιοχής επιρροής». Η παράμετρος *degree* παίρνει τιμές 1 ή 2 και καθορίζει το βαθμό των τοπικών πολυωνύμων που θα χρησιμοποιηθεί σε κάθε τοπική παλινδρόμηση (SAS Institute Inc.,1999). Η τιμή 1 αναφέρεται σε γραμμική συνάρτηση, ενώ η τιμή 2 σε τετραγωνική συνάρτηση. Η τιμή 0 είναι επίσης επιτρεπτή αλλά δεν συνιστάται λόγω των περιορισμένων δοκιμών της.

4.5.1. Ανάλυση ευαισθησίας τοπικά σταθμισμένης παλινδρόμησης (loess)

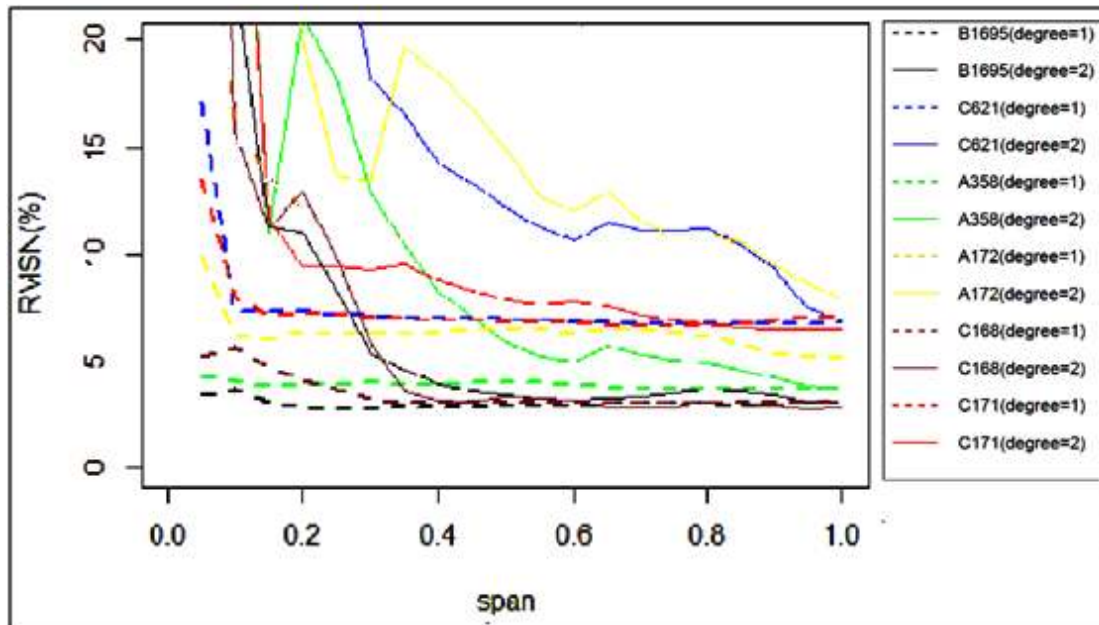
Για την επιλογή κατάλληλων τιμών *span* και *degree* είναι χρήσιμο να εξεταστεί η απόδοση του μοντέλου για τις διάφορες τιμές τους και να γίνει η σχετική σύγκριση. Αυτό επιχειρείται για όλες τις σειρές δεδομένων λαμβάνοντας υπόψη για την

παλινδρόμηση τους παράγοντες που χρησιμοποιεί το μοντέλο του Gijbbs, προκειμένου να ακολουθήσει η σύγκριση των δυο μεθόδων.

Στα διάγραμμα 4.15., 4.16 απεικονίζεται το μέσο κανονικοποιημένο τετραγωνικό σφάλμα που προκύπτει από την τοπικά σταθμισμένη παλινδρόμηση, εφαρμόζοντας την για διάφορες τιμές span και degree. Τα δεδομένα που εισάγονται στην παλινδρόμηση είναι οι ταχύτητες των οχημάτων 2, 3, καθώς και η μεταξύ τους απόσταση σε χρόνο t_i , τα οποία απαιτούνταν επίσης για την εφαρμογή του μοντέλου του Gijbbs. Το επιθυμητό εξαγόμενο αποτέλεσμα είναι η εκτίμηση της ταχύτητας του τρίτου οχήματος σε χρόνο $t_i + \tau$, όπου τ ο χρόνος αντίδρασης. Συγκεκριμένα στο διάγραμμα 4.15. ο χρόνος αντίδρασης θεωρήθηκε 0.4 s, ενώ στο διάγραμμα 4.16. ο χρόνος αντίδρασης έλαβε την τιμή 1.0 s. Οι τιμές αυτές είναι όμοιες με αυτές που ερευνήθηκε το μοντέλο του Gijbbs, έτσι ώστε να είναι δυνατή η σύγκριση των δυο προσεγγίσεων στη συνέχεια. Η εκτίμηση γίνεται χρησιμοποιώντας το πρότυπο που προκύπτει από τη σειρά δεδομένων B1695.



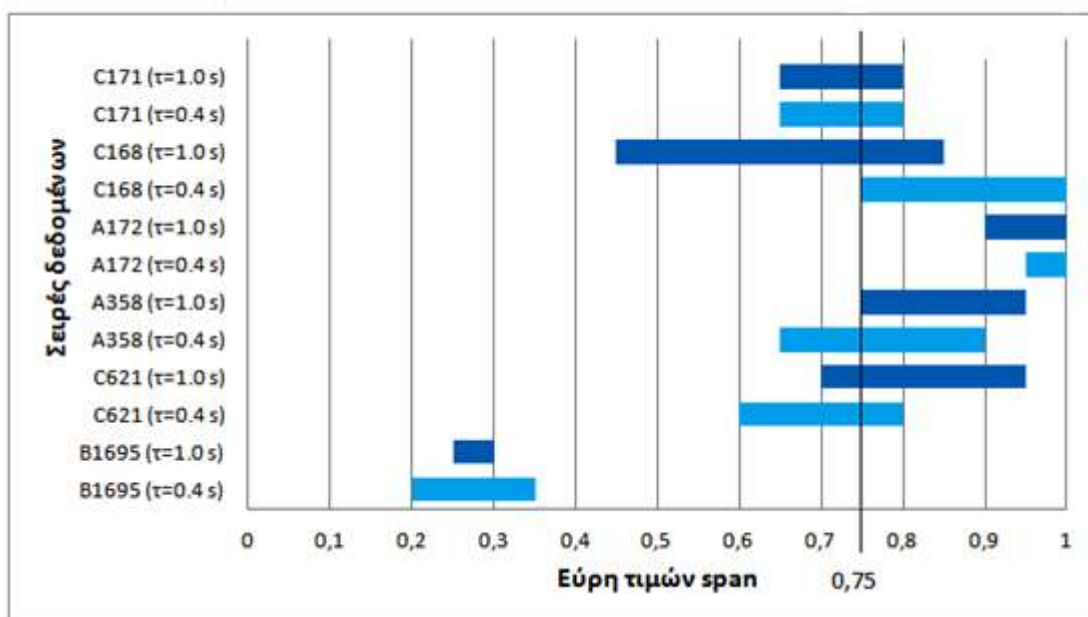
Διάγραμμα 4.15.: Επιρροή του RMSN για διάφορες τιμές span και degree, κατά την εφαρμογή της μεθόδου loess για χρόνο αντίδρασης $\tau=0.4$ s



Διάγραμμα 4.16.: Επιρροή του RMSN για διάφορες τιμές span και degree, κατά την εφαρμογή της μεθόδου loess για χρόνο αντίδρασης $\tau=1.0$ s

Στα παραπάνω διαγράμματα οι διακεκομμένες γραμμές απεικονίζουν το RMSN εκτίμησης της v_3 με τη μέθοδο loess για κάθε σειρά δεδομένων με degree=1 και κάθε τιμή του span, ενώ οι συνεχείς γραμμές απεικονίζουν ακριβώς το ίδιο αλλά για degree=2. Είναι φανερό ότι οι διακεκομμένες γραμμές (degree=1) είναι πιο ομαλές και αντιστοιχούν σε μικρότερο RMSN σε σχέση με τις συνεχείς γραμμές (degree=2). Συνεπώς, για την παράμετρο degree του loess προτιμάται η τιμή 1.

Όσον αφορά στην παράμετρο span, παρατηρώντας τις διακεκομμένες γραμμές διακρίνεται εύκολα ότι για όλες τις σειρές δεδομένων που εξετάζονται το σφάλμα είναι σχεδόν σταθερό για τις τιμές 0.4 έως 1.0, τόσο όταν ο χρόνος αντίδρασης ληφθεί 0.4 s όσο και όταν ληφθεί 1.0 s. Επομένως, εξαιρώντας τις πολύ μικρές τιμές, το span δεν φαίνεται να επηρεάζει σημαντικά τα αποτελέσματα. Ωστόσο, στο διάγραμμα 4.17 απεικονίζονται τα εύρη τιμών για την παράμετρο span που παρατηρήθηκε το μικρότερο σφάλμα για κάθε σειρά δεδομένων. Για τη σειρά B1695 η οποία χρησιμοποιείται ως πρότυπο για την εκτίμηση ήταν αναμενόμενο ένα μικρό span να επιφέρει μικρότερο σφάλμα RMSN. Για τις υπόλοιπες σειρές δεδομένων το «ιδανικό span» εντοπίζεται σε μεγαλύτερες τιμές κυρίως ανάμεσα στο εύρος 0.6-0.9. Επομένως, μπορεί να επιλέγει η τιμή 0.75 ως πιο κατάλληλη για τα δεδομένα. Επίσης, η τιμή 0.75 είναι η προκαθορισμένη τιμή από το πρόγραμμα.



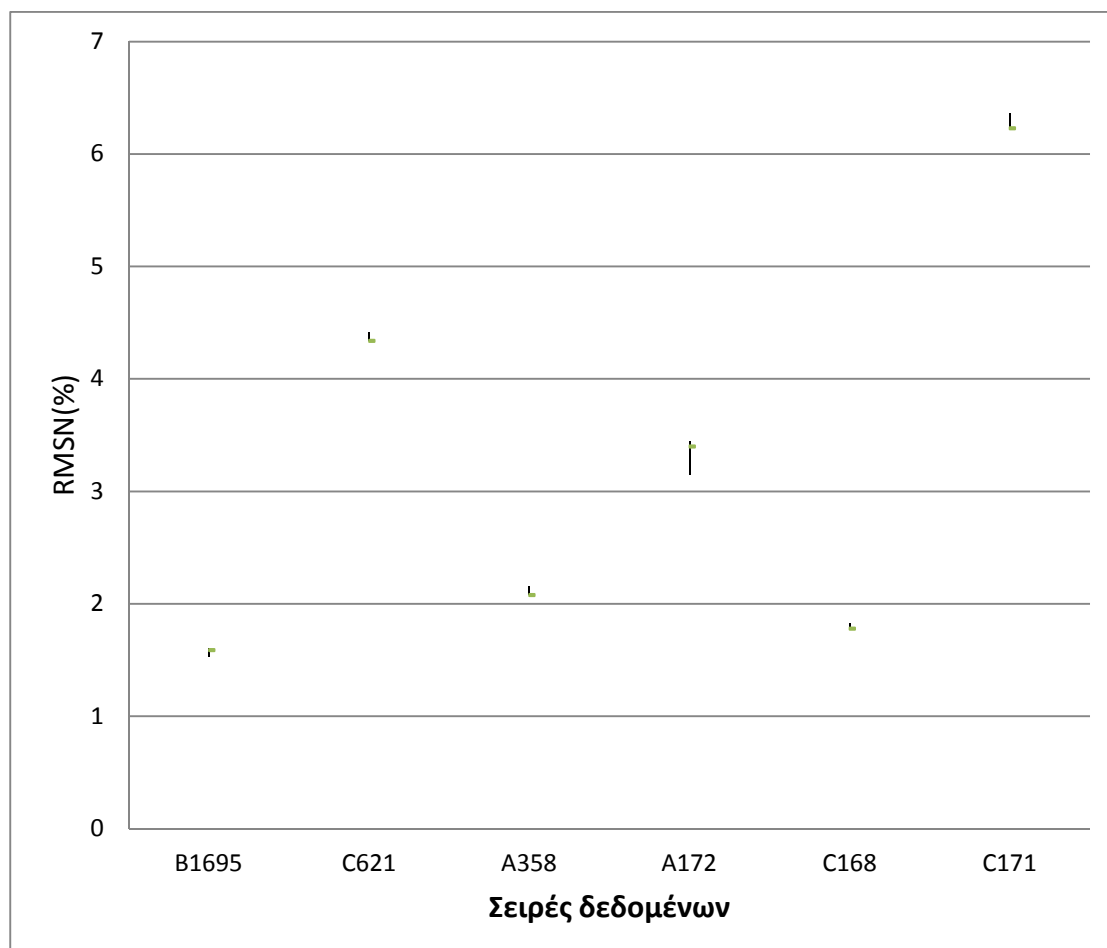
Διάγραμμα 4.17: Εύρη τιμών span για τις οποίες παρατηρείται το μικρότερο σφάλμα RMSN για κάθε σειρά δεδομένων

Στη συνέχεια, παρουσιάζεται το πόσο επιδρούν οι διαφορετικές τιμές του span στο αποτέλεσμα. Αναλυτικότερα στον πίνακα 4.6 υπολογίστηκε το σφάλμα RMSN για τη μέθοδο loess, θεωρώντας το χρόνο αντίδρασης 0.4 s και θέτοντας αρχικά τη βέλτιστη τιμή span για κάθε σειρά δεδομένων και έπειτα την επιλεγόμενη τιμή span =0.75. Τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα και γίνεται σύγκριση, έτσι ώστε να γίνει σαφές ότι το span επηρεάζει καθόλου ή ελάχιστα την αποτελεσματικότητα της μεθόδου.

Σειρές δεδομένων	Βέλτιστο span	RMSN(%) για το βέλτιστο span	RMSN(%) για span=0.75	Ποσοστό(%) απόκλισης σφάλματος
B1695	0.25	1.53	1.59	3.92
C621	0.75	4.34	4.34	0
A358	0.75	2.08	2.08	0
A172	1.00	3.15	3.40	7.94
C168	0.85	1.77	1.78	0.56
C171	0.65	6.23	6.23	0

Πίνακας 4.6: Ποσοστά απόκλισης σφάλματος μεταξύ της βέλτιστης και της επιλεγόμενης τιμής span για κάθε σειρά δεδομένων

Στο διάγραμμα 4.18. απεικονίζεται πόσο μεγαλύτερο ή μικρότερο σφάλμα προκύπτει για κάθε σειρά δεδομένων με span 0.75 σε σχέση με το «ιδανικό span» ή το «χειρότερο span» της κάθε σειράς δεδομένων. Το «ιδανικό span» ή το «χειρότερο span» αναφέρονται στο span (ανάμεσα στο εύρος 0.4-1.0) με το οποίο προκύπτει το μικρότερο ή το μεγαλύτερο σφάλμα RMSN αντίστοιχα. Ο δείκτης που φαίνεται στο διάγραμμα αντιστοιχεί στην τιμή του RMSN που προκύπτει για span=0.75. Όπως είναι φανερό πρόκειται για πολύ μικρές διαφοροποιήσεις που δεν επηρεάζουν την εκτίμηση. Επομένως, οι προκαθορισμένες τιμές span=0.75 και degree=1 είναι κατάλληλες για τα διαθέσιμα δεδομένα (SAS Institute Inc., 2009). Επίσης, οι διαφορές του RMSN ανάλογα με το span είναι του ίδιου μεγέθους όταν ο χρόνος αντίδρασης θεωρηθεί 1.0 s.



Διάγραμμα 4.18.: Μεταβολή RMSN ανάλογα με το βέλτιστο, το χειρότερο και το γενικά επιλεγόμενο span για κάθε σειρά δεδομένων

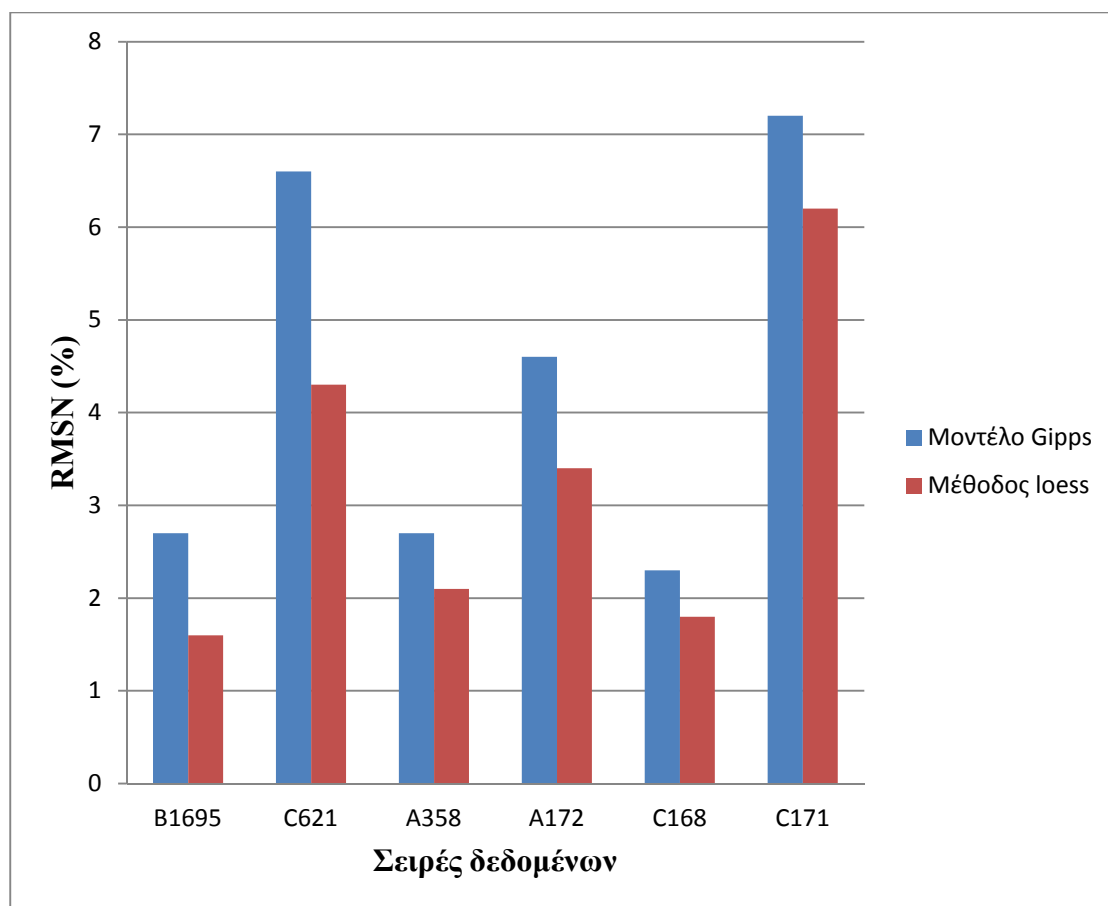
4.5.3. Εφαρμογή τοπικά σταθμισμένης παλινδρόμησης (loess) και άμεση σύγκριση με το μοντέλο Gírps

Για άμεση σύγκριση με το μοντέλο Gírps θα πρέπει για την εφαρμογή του loess να χρησιμοποιηθούν οι ίδιοι παράγοντες που χρησιμοποιήθηκαν στην εφαρμογή του μοντέλου Gírps. Στο μοντέλο του Gírps σύμφωνα με την εξίσωση 2.3 πέρα από τις παραμέτρους του μοντέλου χρησιμοποιήθηκαν η ταχύτητα του προπορευόμενου οχήματος (όχημα 2), η ταχύτητα του οχήματος που ακολουθεί (όχημα 3) και η μεταξύ τους απόσταση σε χρόνο t_i , προκειμένου να εκτιμηθεί η ταχύτητα του οχήματος 3 σε χρόνο $t_i+\tau$, όπου τ ο χρόνος αντίδρασης των οδηγών.

Κατά συνέπεια, για την εφαρμογή της μεθόδου loess θα πρέπει να ληφθούν υπόψη, επίσης, οι ίδιοι παράγοντες, δηλαδή οι ταχύτητες των δύο οχημάτων (v_2, v_3) και η μεταξύ τους απόσταση (D_{23}) σε χρόνο t_i για την εκτίμηση της ταχύτητας του τρίτου οχήματος σε χρόνο $t_i+\tau$ και συγκεκριμένα $t_i+0.4$ ή $t_i+1.0$, καθώς οι τιμές αυτές επιλέχθηκαν για το χρόνο αντίδρασης κατά την εφαρμογή του μοντέλου του Gírps στις ενότητες 4.4.1 και 4.4.2. Για την εφαρμογή της μεθόδου loess τίθενται οι τιμές των παραμέτρων της $span=0.75$ και $degree=1$ όπως επιλέχθηκαν στην προηγούμενη ενότητα. Για το μοντέλο του Gírps οι τιμές των παραμέτρων είναι αυτές που ελαχιστοποιούν το σφάλμα του: $\tau=0.4$ s, $\alpha_n=0.8$ m/s², $V_n=14$ m/s, $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s² και $\hat{b}=-3.0$ m/s² (μοντέλο g1) ή για μια πιο ρεαλιστική προσέγγιση $\tau=1.0$ s, $\alpha_n=1.6$ m/s², $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s², $\hat{b}=-3.0$ m/s² και $V_n=16$ m/s (μοντέλο G3), όπως προέκυψε από την ανάλυση ευαισθησίας. Τα αποτελέσματα παρουσιάζονται στους πίνακες 4.7., 4.8 και αντίστοιχα στα διαγράμματα 4.19., 4.20.

Σειρές δεδομένων	RMSN (%) για επιλεγμένο μοντέλο Gírps (μοντέλο g1- επιλογή 1 από πίνακα 4.5.)	RMSN(%) για μοντέλο loess (μοντέλο llf)	Βελτίωση εκτίμησης (%)
B1695	2.7	1.6	40.7
C621	6.6	4.3	34.8
A358	2.7	2.1	22.2
A172	4.6	3.4	26.1
C168	2.3	1.8	21.7
C171	7.2	6.2	13.9

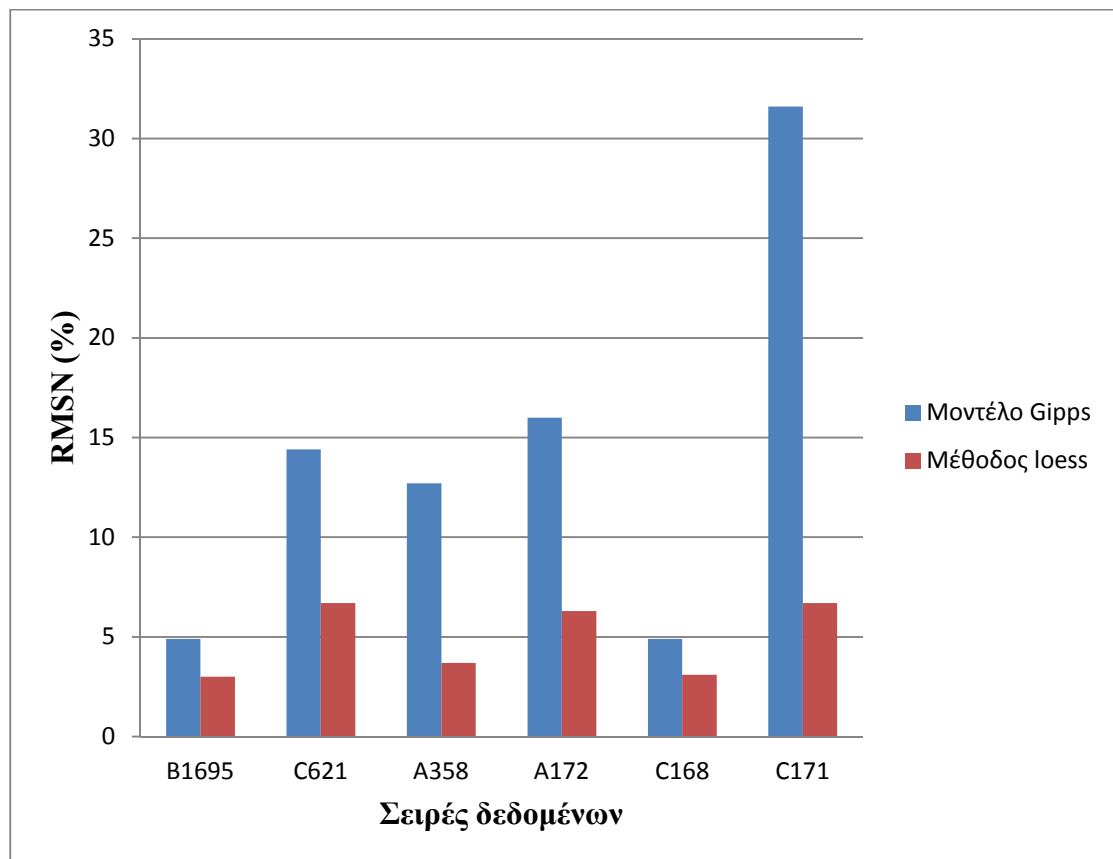
Πίνακας 4.7.: Σύγκριση μοντέλου Gírps (g1) και μεθόδου loess (llf) για χρόνο αντίδρασης $\tau=0.4$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN



Διάγραμμα 4.19.: Σύγκριση RMSN του μοντέλου Gipps (g1) και της μεθόδου loess (l1f) για χρόνο αντίδρασης $\tau=0.4$ s

Σειρές δεδομένων	RMSN (%) για επιλεγμένο μοντέλο Gipps (μοντέλο G3- επιλογή 3 από πίνακα 4.5.)	RMSN(%) για μοντέλο loess (loess l2f)	Βελτίωση εκτίμησης (%)
B1695	4.9	3.0	38.8
C621	14.4	6.7	53.5
A358	12.7	3.7	70.9
A172	16.0	6.3	60.6
C168	4.9	3.1	36.7
C171	31.6	6.7	78.8

Πίνακας 4.8.: Σύγκριση μοντέλου Gipps (G3) και μεθόδου loess (l2f) για χρόνο αντίδρασης $\tau=1.0$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN

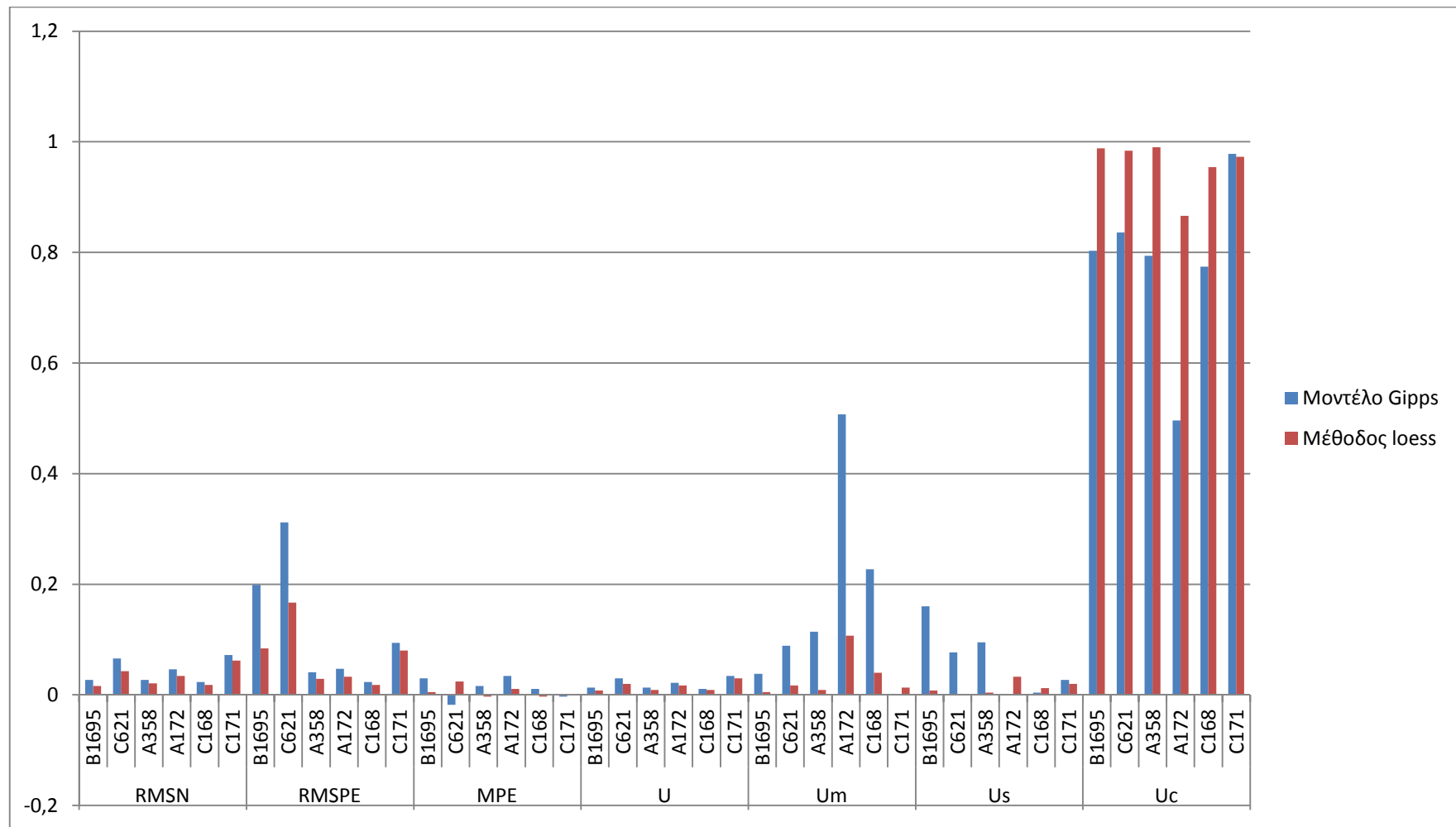


Διάγραμμα 4.20.: Σύγκριση μοντέλου Gijps ($G3$) και μεθόδου loess ($l2f$) για χρόνο αντίδρασης $\tau=1.0$ s χρησιμοποιώντας ως μέτρο αξιολόγησης το RMSN

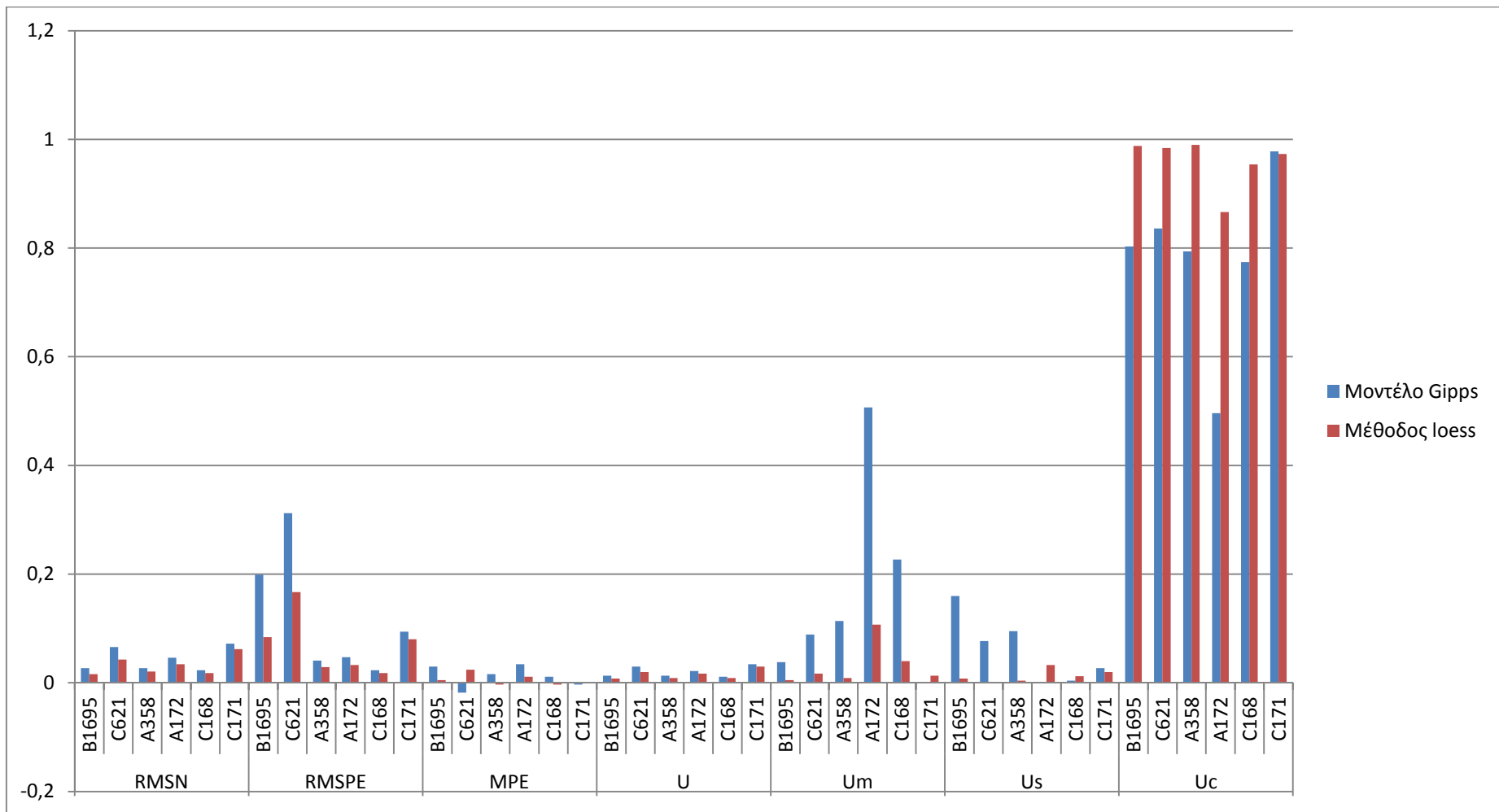
Σύμφωνα με τα παραπάνω αποτελέσματα η μέθοδος loess παρέχει πιο αξιόπιστη εκτίμηση για όλες τις σειρές δεδομένων των πινάκων 4.7, 4.8 σε σχέση με το μοντέλο του Gijps, με εξαίρεση μία για την οποία το σφάλμα RMSN για την εκτίμηση της $v3$ είναι ακριβώς το ίδιο και με τις δυο μεθόδους (Πίνακας 4.7). Επίσης, αξίζει να αναφερθεί ότι η μέθοδος loess υπερτερεί του μοντέλου του Gijps στην ταχύτητα υπολογισμού.

Στο διάγραμμα 4.21., από τη σύγκριση των τιμών του RMSN συμπεραίνουμε ότι η μέθοδος loess προσεγγίζει καλύτερα τις πραγματικές παρατηρήσεις και έχει μικρότερο συνολικό σφάλμα από το μοντέλο του Gijps. Το RMSPE είναι μικρότερο για τη μέθοδο loess για όλες τις σειρές δεδομένων. Το μέτρο MPE δεν δείχνει να υπάρχει μεγάλη συστηματική υποτίμηση ή υπερτίμηση της εκτίμησης για κάποια από τις δύο μεθόδους. Το μέτρο αξιολόγησης U επιβεβαιώνει την εγκυρότητα της μεθόδου loess, καθώς για την πλειοψηφία των δεδομένων έχει μικρότερη τιμή για τη μέθοδο loess σε σχέση με το μοντέλο του Gijps. Τέλος, εξετάστηκαν οι επιμέρους αναλογίες του

συντελεστή Theil που όσο οι τιμές τους πλησιάζουν στο 1, τόσο καλύτερη προσαρμογή υπάρχει στα δεδομένα. Σύμφωνα με τις τιμές του U_M , η μέθοδος loess παρέχει αποτελέσματα με μικρότερο συστηματικό σφάλμα από το μοντέλο Gijbbs για την πλειοψηφία των σειρών δεδομένων. Το μέτρο U_M υποδηλώνει καλύτερα αποτελέσματα για τη μέθοδο loess χωρίς όμως να δίνει ξεκάθαρη εικόνα και από το μέτρο U_c η μέθοδος loess παρουσιάζει καλύτερη συμπεριφορά από το μοντέλο του Gijbbs. Στο διάγραμμα 4.22. η συμπεριφορά του μοντέλου Gijbbs εμφανίζεται χειρότερη από τη μέθοδο loess σύμφωνα με όλα τα μέτρα αξιολόγησης.



Διάγραμμα 4.21.: Σύγκριση της μεθόδου loess (lf) και του μοντέλου GIPPS (g1) για όλες τις σειρές δεδομένων με χρήση διάφορων μέτρων αξιολόγησης για χρόνο αντίδρασης $\tau=0.4$ s

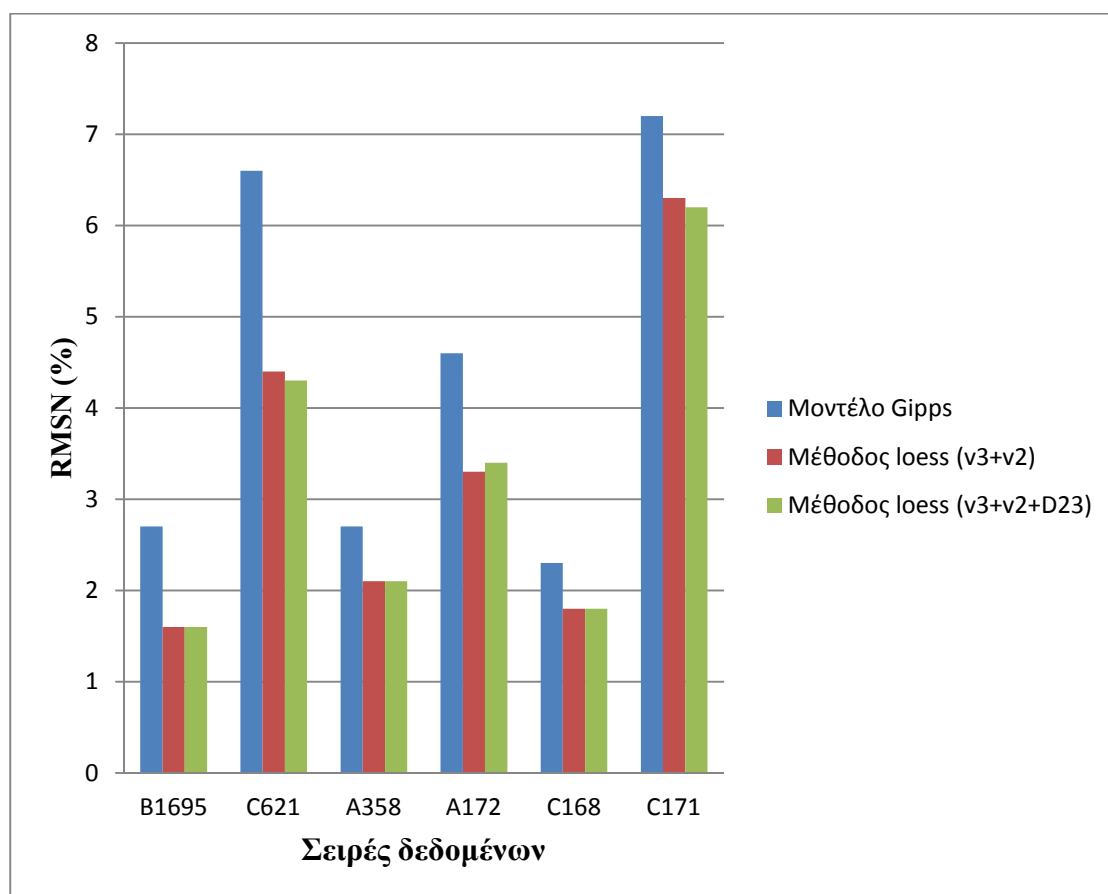


Διάγραμμα 4.22.: Σύγκριση της μεθόδου loess (l2f) και του μοντέλου G3 για όλες τις σειρές δεδομένων με χρήση διάφορων μέτρων αξιολόγησης για χρόνο αντίδρασης $\tau=1.0$

Τέλος για μια καλύτερη σύγκριση των δύο μεθόδων δοκιμάζεται η μέθοδος loess ακόμη μια φορά αλλά χρησιμοποιώντας λιγότερους από τους παράγοντες που χρησιμοποιεί το μοντέλο του Gipps. Οι παράγοντες που χρησιμοποιούνται σε αυτήν την παλινδρόμηση είναι μόνο οι ταχύτητες των δύο οχημάτων σε χρόνο t_i για την εκτίμηση της ταχύτητας του τρίτου οχήματος σε χρόνο $t_i+0.4s$ και $t_i+1.0s$ (μοντέλο 11s, μοντέλο 12s). Οι παράμετροι του loess ορίζονται $span=0.75$ και $degree=1$ όπως παραπάνω. Τα αποτελέσματα παρουσιάζονται στους πίνακες 4.9., 4.10. Παρά το γεγονός ότι στην παλινδρόμηση χρησιμοποιήθηκαν λιγότεροι παράγοντες, η μέθοδος loess παρέχει καλύτερη εκτίμηση από το μοντέλο Gipps. Τέλος συγκρίνοντας την μέθοδο loess όπως εφαρμόστηκε προηγουμένως με 3 παράγοντες (loess-full) και όπως εφαρμόστηκε με 2 παράγοντες (loess-simple) γίνεται αντιληπτό ότι ο τρίτος παράγοντας (απόσταση D23) δεν προσφέρει σημαντική βελτίωση στην εκτίμηση (Διαγράμματα 4.23., 4.24.).

Σειρές δεδομένων	RMSN (%) για μοντέλο Gipps (μοντέλο g1-Επιλογή 1 από πίνακα 4.5.)	RMSN (%) για απλουστευμένο μοντέλο loess (μοντέλο 11s)	Βελτίωση εκτίμησης(%)
B1695	2.7	1.6	40.7
C621	6.6	4.4	33.3
A358	2.7	2.1	22.2
A172	4.6	3.3	28.3
C168	2.3	1.8	21.7
C171	7.2	6.3	12.5

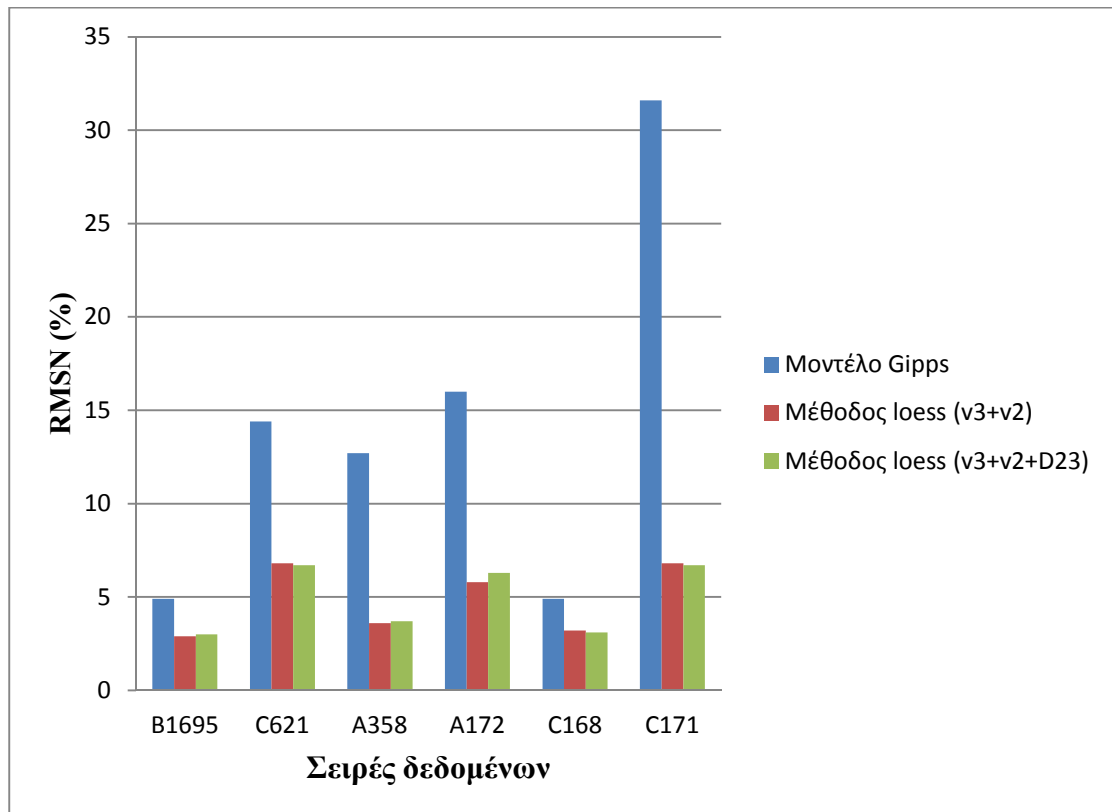
Πίνακας 4.9.: Σύγκριση μεθόδου loess (11s) με μοντέλο Gipps (g1) χρησιμοποιώντας λιγότερους παράγοντες για χρόνο αντίδρασης $\tau=0.4 s$



Διάγραμμα 4.23.: Σύγκριση RMSN του μοντέλου Gipps, της μεθόδου loess (v3+v2+D23)- 11f και της μεθόδου loess (v3+v2)- 11s για χρόνο αντίδρασης $\tau=0.4$ s

Σειρές δεδομένων	RMSN (%) για μοντέλο Gipps (μοντέλο G3-Επιλογή 3 από πίνακα 4.5.)	RMSN (%) για απλουστευμένο μοντέλο loess (μοντέλο l2s)	Βελτίωση εκτίμησης (%)
B1695	4.9	2.9	40.8
C621	14.4	6.8	52.8
A358	12.7	3.6	71.7
A172	16.0	5.8	63.8
C168	4.9	3.2	34.7
C171	31.6	6.8	78.5

Πίνακας 4.10.: Σύγκριση μεθόδου loess (l2s) με μοντέλο Gipps (G3) χρησιμοποιώντας λιγότερους παράγοντες για χρόνο αντίδρασης $\tau=1.0$ s



Διάγραμμα 4.24.: Σύγκριση RMSN του μοντέλου G3, της μεθόδου loess (v3+v2+D23)- 12f και της μεθόδου loess (v3+v2)- 12s για χρόνο αντίδρασης $\tau=1.0$ s

4.5.4. Επιλογή διάφορων παραγόντων παλινδρόμησης

Σε αντίθεση με το μοντέλο του G3 που πρέπει αναγκαστικά να εφαρμοστεί ένας τύπος και να έχουν συλλεχθούν τα απαραίτητα δεδομένα, με τη μέθοδο loess μπορούν να γίνουν διαφορετικοί συνδυασμοί των παραγόντων και να δώσουν το επιθυμητό αποτέλεσμα είτε με την ίδια ακρίβεια ή μικρότερη αναλόγως τα δεδομένα που διατίθενται. Για το λόγο αυτό στη συνέχεια της εργασίας κρίνεται σκόπιμο να εξεταστούν διάφοροι συνδυασμοί των παραγόντων και επιχειρείται η αναζήτηση του βέλτιστου συνδυασμού τους. Με σκοπό την καλύτερη δυνατή εκτίμηση της ταχύτητας του τρίτου οχήματος (v_3) σε χρόνο $t_i + \tau$ (όπου τ ο χρόνος αντίδρασης του οδηγού) παρακάτω αναζητείται ο βέλτιστος συνδυασμός των εξής παραμέτρων σε χρόνο t_i :

- η ταχύτητα του πρώτου οχήματος που κινείται σε σειρά (v_1).
- η ταχύτητα του δεύτερου οχήματος (v_2).
- η ταχύτητα του τρίτου οχήματος (v_3)
- η ταχύτητα του τέταρτου οχήματος (v_4)
- η απόσταση του τρίτου οχήματος από το μπροστινό του (D23)

- η απόσταση του τρίτου οχήματος από το πίσω όχημα (D34)

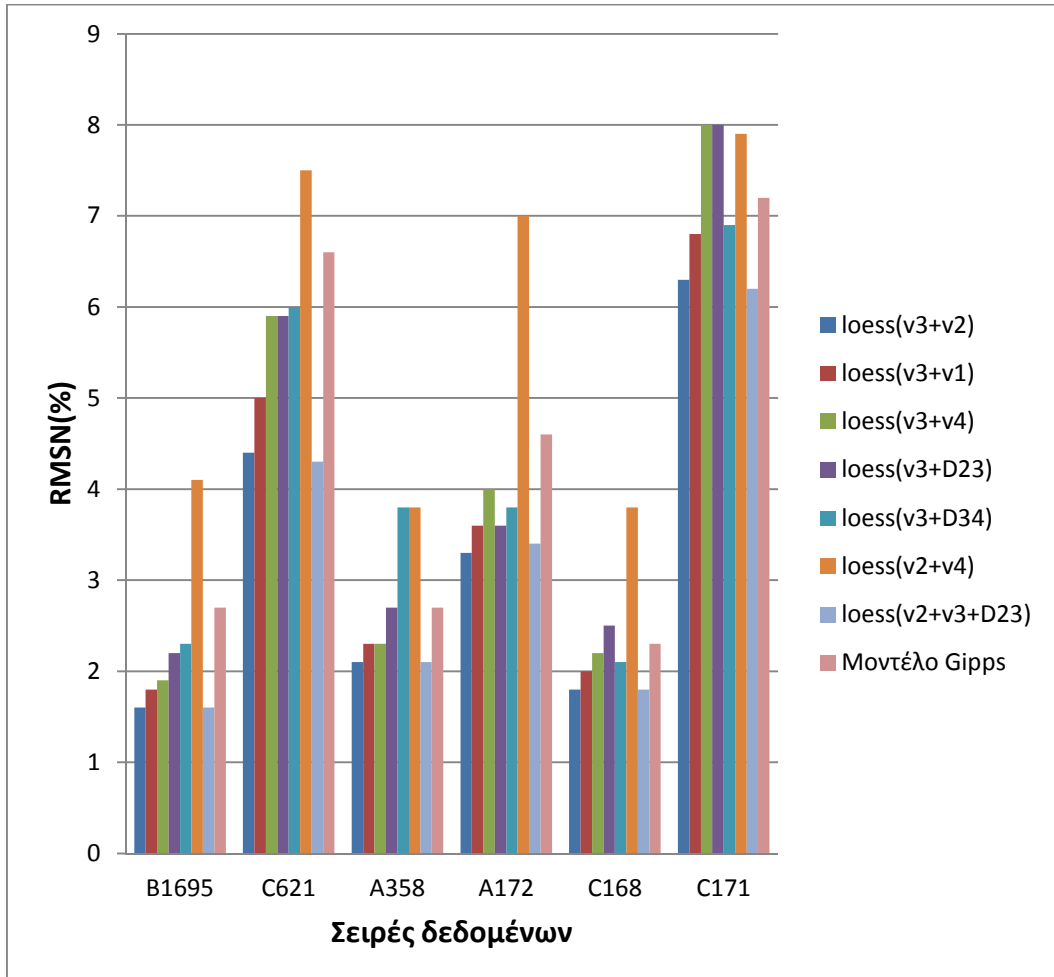
Πρώτα δοκιμάζονται όλοι οι πιθανοί συνδυασμοί των παραπάνω παραμέτρων ανά δυάδες στη σειρά δεδομένων B1695 για χρόνο αντίδρασης $\tau=0.4s$, degree 1 και span 0.75 όπως επιλέχθηκαν στις προηγούμενες ενότητες. Στον πίνακα 4.11. παρουσιάζονται τα αποτελέσματα. Όπως είναι φανερό οι συνδυασμοί των παραγόντων επιδρούν σχεδόν με τον ίδιο τρόπο στην αποτελεσματικότητα της μεθόδου και για τις δύο σειρές δεδομένων. Οι πέντε πρώτοι συνδυασμοί όπου συμμετέχει η ταχύτητα του τρίτου οχήματος υπερέχουν αρκετά σε σύγκριση με τους υπόλοιπους. Επομένως, για την εκτίμηση της ταχύτητας του τρίτου οχήματος σε χρόνο $t_i + \tau$ αρκεί να είναι γνωστή η ταχύτητα του ίδιου οχήματος σε χρόνο t_i και ένας παράγοντας ακόμα. Σε περίπτωση που δεν είναι γνωστή ο επόμενος καλύτερος συνδυασμός είναι η ταχύτητες του μπροστινού και του πίσω οχήματος.

Συνδυασμοί παραμέτρων για το loess	RMSN (%) για B1695	RMSN (%) για C621
v3+v2	1.6	4.4
v3+v1	1.8	5.0
v3+v4	1.9	5.9
v3+D23	2.2	5.9
v3+D34	2.3	6.0
v2+v4	4.1	7.5
v2+D34	4.8	9.0
v2+v1	5.3	9.9
v2+D23	5.3	10.0
v1+v4	5.3	11.6
D23+v4	6.1	12.5
v1+D23	6.7	14.8
v1+D34	6.7	15.1
D23+D34	8.0	22.2
D34+v4	8.2	16.5

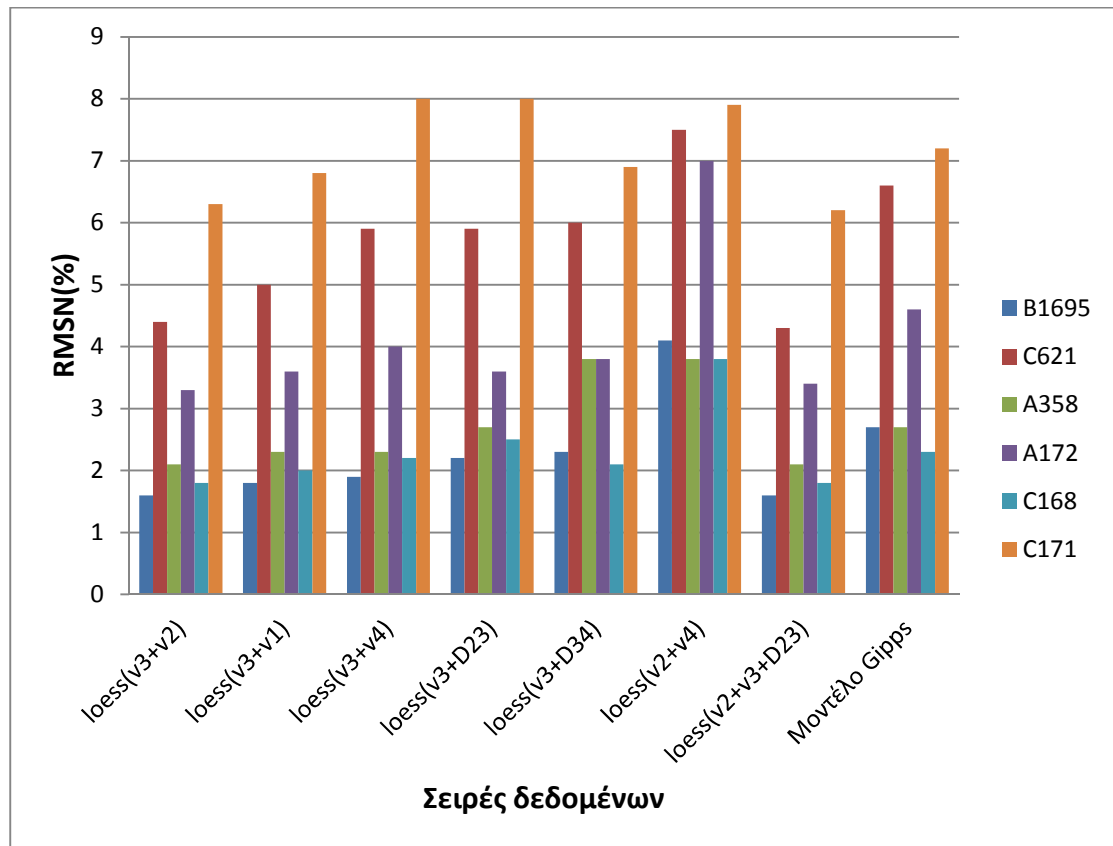
Πίνακας 4.11.: Συνδυασμοί διάφορων παραγόντων ανά δυάδες για τη μέθοδο loess και το αντίστοιχο RMSN (%) για τις σειρές δεδομένων B1695, C621

Στα διαγράμματα 4.26., 4.27. απεικονίζονται για όλες τις σειρές δεδομένων το σφάλμα RMSN για το μοντέλο του Gipps, για το loess με τις παραμέτρους του Gipps

και για τα 6 πρώτα loess του πίνακα 4.11. Σε όλες τις περιπτώσεις το loess με τις παραμέτρους v_2+v_4 είναι το χειρότερο, ενώ στην πλειοψηφία των περιπτώσεων το καλύτερο είναι το loess που λαμβάνει υπόψη τις παραμέτρους του Gipps.



Διάγραμμα 4.16.: RMSN (%) για διάφορους διπλούς συνδυασμούς παραγόντων στην εφαρμογή της μεθόδου loess

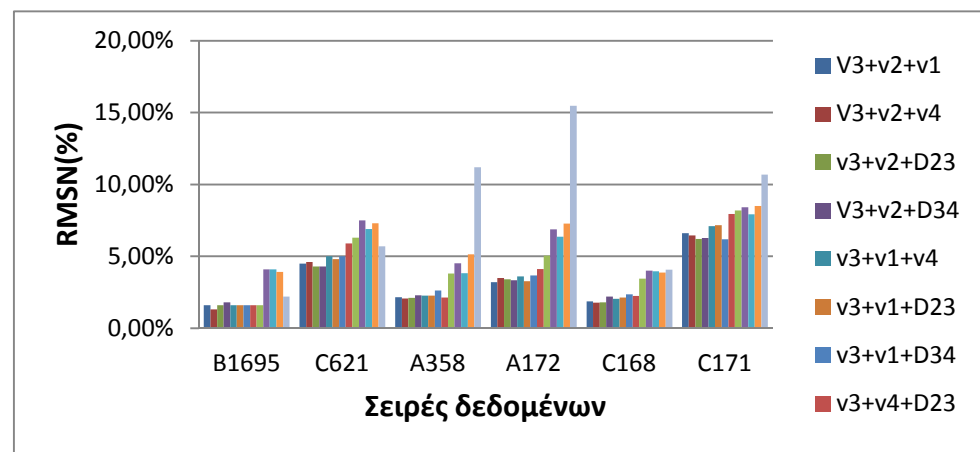


Διάγραμμα 4.27.: RMSN (%) για διάφορους διπλούς συνδυασμούς παραγόντων στην εφαρμογή της μεθόδου loess για κάθε σειρά δεδομένων

Στη συνέχεια για τους έξι πρώτους καλύτερους συνδυασμούς του πίνακα 4.11. επιχειρήθηκε να προστεθεί ένας τρίτος παράγοντας για να διαπιστωθεί αν βελτιώνεται η απόδοση της μεθόδου. Τα αποτελέσματα φαίνονται στον πίνακα 4.12 και στα διαγράμματα 4.28., 4.29. Ωστόσο, με την προσθήκη ενός τρίτου παράγοντα παρατηρείται ότι δεν υπήρξαν μεγάλες διαφοροποιήσεις στο σφάλμα RMSN και ο καλύτερος συνδυασμός τριών παραγόντων για την πλειοψηφία των σειρών δεδομένων είναι ο συνδυασμός που προκύπτει από την εξίσωση του Gijpss (v3, v2, D23). Παρόλα αυτά οι υπόλοιποι συνδυασμοί (όπως v3+v2+v1, v3+v2+v4, v3+v2+D34, κ.ά.) που παρουσιάζονται δίνουν επίσης καλά αποτελέσματα χωρίς μεγάλες αποκλίσεις. Στο σημείο αυτό φαίνεται η ευελιξία της μεθόδου, καθώς δεν απαιτεί συγκεκριμένα δεδομένα αλλά μπορεί να χρησιμοποιήσει όποια δεδομένα είναι διαθέσιμα.

	V3+v2+v1	V3+v2+v4	v3+v2+D23	V3+v2+D34	v3+v1+v4	v3+v1+D23	v3+v1+D34	v3+v4+D23	v3+v4+D34	v2+v4+v1	v2+v4+D23	v2+v4+D34	v3+D23+D34
B1695	1.60	1.30	1.6	1.80	1.60	1.60	1.60	1.60	1.60	4.10	4.10	3.90	2.20
C621	4.50	4.60	4.3	4.30	5.00	4.80	5.00	5.90	6.30	7.50	6.90	7.30	5.70
A358	2.2	2.1	2.1	2.3	2.3	2.3	2.6	2.1	3.8	4.5	3.8	5.1	11.2
A172	3.2	3.5	3.4	3.3	3.6	3.3	3.7	4.1	5.0	6.9	6.4	7.3	15.5
C168	1.9	1.8	1.8	2.2	2.0	2.1	2.3	2.2	3.4	4.0	3.9	3.9	4.1
C171	6.6	6.4	6.2	6.3	7.1	7.2	6.2	7.9	8.2	8.4	7.9	8.5	10.7

Πίνακας 4.12.: RMSN (%) για διάφορους συνδυασμούς τριών παραγόντων στην εφαρμογή της μεθόδου loess



Διάγραμμα 4.28.: Απεικόνιση του σφάλματος RMSN (%) για διάφορους συνδυασμούς τριών παραγόντων στην εφαρμογή της μεθόδου loess

Στη συνέχεια έγινε προσπάθεια να προστεθεί μια τέταρτη παράμετρος αλλά το σφάλμα σχεδόν σε όλες τις περιπτώσεις έμεινε σταθερό ή αυξήθηκε και κατά συνέπεια προτιμώνται τα πιο απλά μοντέλα δεδομένου ότι απαιτούν λιγότερα δεδομένα. Επομένως, ο συνδυασμός δύο ή τριών παραμέτρων αρκεί για να εφαρμοστεί η μέθοδος loess και παρέχει αξιόπιστη εκτίμηση στα διαθέσιμα δεδομένα. Ωστόσο, θα μπορούσε να εξεταστεί η αλληλεπίδραση με άλλους παράγοντες πέρα από αυτούς που παρουσιάζονται στο πλαίσιο αυτής της εργασίας, όπως παραμέτρους του οδικού δικτύου.

4.6. Σύγκριση μεθόδων και συμπεράσματα

Από την ανάλυση που προηγήθηκε προέκυψε ότι η μέθοδος loess παρέχει καλύτερα αποτελέσματα από το μοντέλο του Gijbbs, ακόμα και για λιγότερα δεδομένα από αυτά που απαιτεί ο μαθηματικός τύπος του Gijbbs. Επίσης, από την εφαρμογή της μεθόδου loess διαπιστώνεται ότι πρόκειται για μια μέθοδο που παρέχει ευελιξία ακόμα και στα δεδομένα που απαιτούνται. Όποια δεδομένα και αν εισάγονται κατά την εφαρμογή της μεθόδου loess, δημιουργούνται ανάλογα πρότυπα που μπορούν να εφαρμοστούν για την εκτίμηση ταχυτήτων σε νέα δεδομένα. Αν και με τη μέθοδο loess δεν γίνεται πλήρως κατανοητό πως από τα δεδομένα που εισάγονται προκύπτει η επιθυμητή εκτίμηση, το γεγονός ότι η εκτίμηση που παράγεται είναι πιο ακριβής την καθιστά προτιμότερη από το μοντέλο του Gijbbs ενισχύοντας έτσι το συμπέρασμα ότι οι μέθοδοι μηχανικής μάθησης μπορούν να προσφέρουν πιο αξιόπιστα αποτελέσματα σε σχέση με τα παραδοσιακά μαθηματικά μοντέλα. Όμως, το γεγονός ότι το μοντέλο του Gijbbs τεκμηριώνει τα αποτελέσματα που εξάγει βασίζόμενο στη λογική θεωρία της ακολουθίας οχημάτων, το καθιστά πιο εύκολα εμπιστεύσιμο από την επιστημονική κοινότητα.

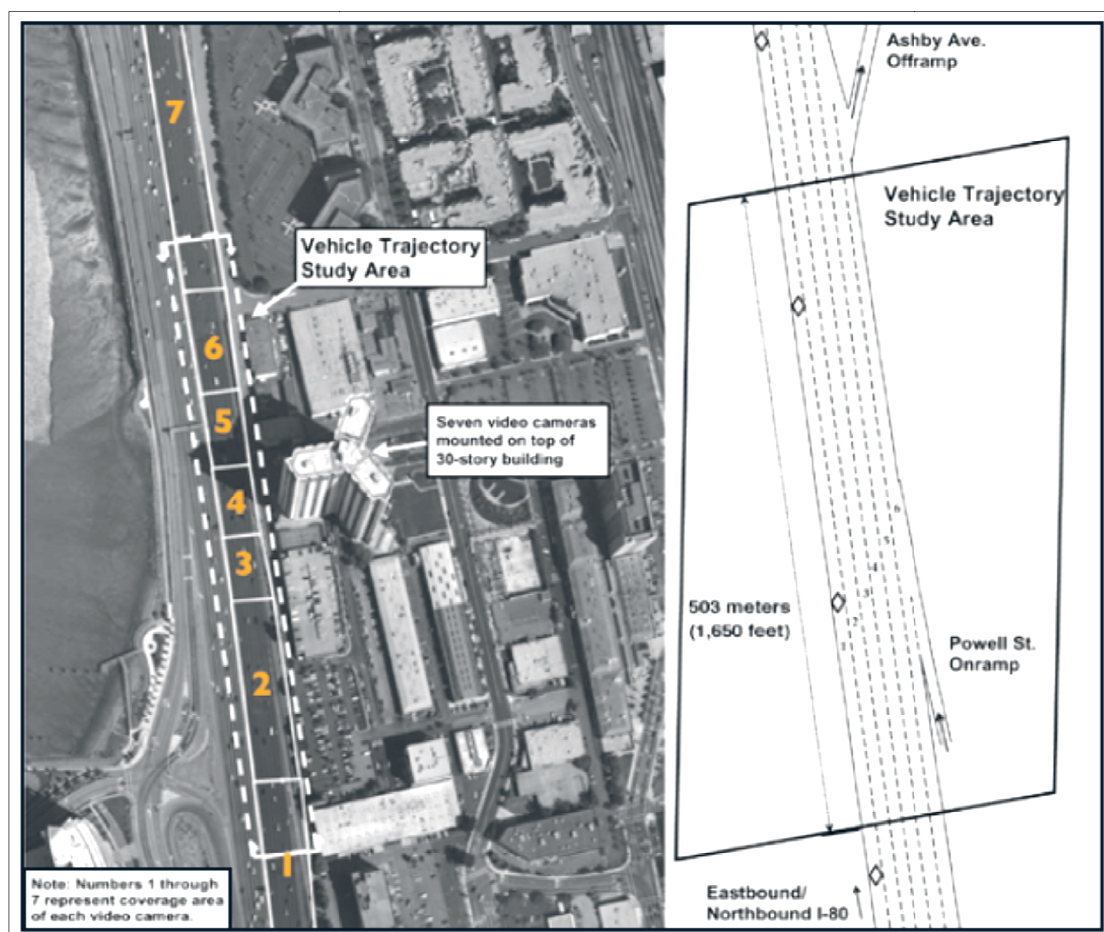
5. ΕΦΑΡΜΟΓΗ ΣΕ ΔΕΔΟΜΕΝΑ NGSIM

Στο κεφάλαιο αυτό επιχειρείται η επιβεβαίωση της προτεινόμενης μεθόδου χρησιμοποιώντας άλλα διαθέσιμα δεδομένα. Στο προηγούμενο κεφάλαιο διαπιστώθηκε ότι η μέθοδος loess παρέχει καλύτερη εκτίμηση σε σχέση με το μοντέλο Gipps για τα δεδομένα που χρησιμοποιήθηκαν. Στη συνέχεια ελέγχεται αν στα νέα δεδομένα που χρησιμοποιούνται η μέθοδος loess είναι επίσης προτιμότερη.

5.1. Περιγραφή και ακρίβεια δεδομένων

Τα δεδομένα που χρησιμοποιούνται σε αυτό το κεφάλαιο προέρχονται από το πρόγραμμα «Next Generation SIMulation (NGSIM) program» (US Department of Transportation, 2012). Στο πλαίσιο αυτού του προγράμματος συλλέγονται δεδομένα με τροχιές οχημάτων σε πραγματικό περιβάλλον με σκοπό την κατανόηση και έρευνα της συμπεριφοράς των οδηγών σε μικροσκοπικό επίπεδο. Συγκεκριμένα κάποια λεπτομερή δεδομένα από τροχιές οχημάτων συλλέχτηκαν στον ανατολικό σύνδεσμο Interstate 80 (I-80) στο Σαν Φρανσίσκο στην περιοχή Emeryville στις 13 Απριλίου 2005. Η περιοχή μελέτης είχε μήκος περίπου 500 m και αποτελούταν από 6 λωρίδες κυκλοφορίας (Σχήμα 5.1.). Επτά σύγχρονες ψηφιακές κάμερες τοποθετήθηκαν στην κορυφή ενός κτιρίου με 30 ορόφους δίπλα στον αυτοκινητόδρομο και κατέγραφαν τα διερχόμενα από την περιοχή μελέτης οχήματα (Σχήμα 5.2.). Με τη χρήση ενός ειδικού λογισμικού (NG-VIDEO) από το βίντεο προέκυπταν τα δεδομένα για τις τροχιές των οχημάτων. Στα δεδομένα αυτά περιγραφόταν η ακριβής θέση κάθε οχήματος στην περιοχή μελέτης ανά 1/10 s καθώς και η σχετική θέση ως προς τα υπόλοιπα οχήματα. Δεδομένα συνολικής διάρκειας 45 λεπτών είναι διαθέσιμα σε ένα σετ δεδομένων χωρισμένο σε τρεις περιόδους των 15 λεπτών και συγκεκριμένα σύμφωνα με την ώρα καταγραφής, 4:00 μ.μ. έως 4:15 μ.μ., 5:00 μ.μ. έως 5:15 μ.μ., 5:15 μ.μ. έως 5:30 μ.μ. Τα δεδομένα αυτά αναφέρονται κατά κύριο λόγο σε καταστάσεις κυκλοφοριακής συμφόρησης. Το πλήρες σετ των δεδομένων I-80 που περιγράφεται παραπάνω είναι ελεύθερα διαθέσιμο στην ιστοσελίδα του προγράμματος NGSIM (<http://ngsim.fhwa.dot.gov>). Τα δεδομένα αυτά έχουν χρησιμοποιηθεί σε πολλές μελέτες για τη βαθμονόμηση και επιβεβαίωση εγκυρότητας (calibration/validation) ήδη υπάρχοντων μοντέλων και αλγορίθμων (Beyrani and Chung, 2011). Για τον σκοπό αυτό χρησιμοποιούνται στη συνέχεια του κεφαλαίου αυτού. Αξιοσημείωτο είναι επίσης ότι μόνο τα έτη 2007-2008 περισσότερες από 30 μελέτες χρησιμοποίησαν

τα δεδομένα NGSIM (Punzo, 2011). Ωστόσο, λίγες μελέτες αναφέρθηκαν στην ακρίβεια των NGSIM δεδομένων (Hamdar and Mahmassani, 2008, Thiemann et al., 2008, Punzo et al., 2011). Αν και ο τρόπος που η Cambridge Systematics Inc στο NGSIM υπολόγισε τις ταχύτητες και επιταχύνσεις από τις τροχιές των οχημάτων και μείωσε τα σφάλματα τους δεν είναι γνωστός, οι παραπάνω μελέτες υποδεικνύουν την ύπαρξη θορύβου και σφαλμάτων στα δεδομένα αυτά (Bevrani and Chung, 2011). Στο πλαίσιο της εργασίας αυτής δεν εξετάζεται η ύπαρξη θορύβου στα δεδομένα, καθότι αν υπάρχουν σφάλματα θα συμπεριληφθούν και στις δύο μεθόδους (μοντέλο GIPRS, προτεινόμενη μέθοδος) και κατά συνέπεια δεν θα επηρεαστεί η σύγκρισή τους παρά μόνο ίσως το αποτέλεσμα της κάθε μεθόδου ξεχωριστά.



Σχήμα 5.1.: Αεροφωτογραφία με τη σχετική θέση της περιοχής μελέτης I-80 και του κτιρίου που ήταν τοποθετημένες οι κάμερες καταγραφής (αριστερό μέρος), σχηματική απεικόνιση των λωρίδων κυκλοφορίας στην περιοχή μελέτης (δεξιό μέρος)

Πηγή: NGSIM, 2006



*Σχήμα 5.2.: Ψηφιακή βιντεοκάμερα που καταγράφει τις τροχιές των οχημάτων στον αυτοκινητόδρομο I-80 του Σαν Φρανσίσκο από την κορυφή ενός κτιρίου
Πηγή: NGSIM, 2006*

Αναλυτικά τα δεδομένα διατίθενται σε 18 στήλες που περιλαμβάνουν τον αριθμό ταυτοποίησης του οχήματος, τον αύξοντα αριθμό για κάθε δέκατο του δευτερολέπτου (κατά αύξουσα σειρά από τη στιγμή έναρξης), το συνολικό αριθμό των δεκάτων του δευτερολέπτου που το όχημα εμφανίζεται σε κάθε σειρά δεδομένων, το χρόνο (milliseconds), τις συντεταγμένες X, Y του μπροστινού μέρους κάθε οχήματος σε τοπικό σύστημα αναφοράς (πόδια), τις συντεταγμένες X, Y του μπροστινού μέρους κάθε οχήματος σε παγκόσμιο σύστημα αναφοράς (πόδια), το μήκος του οχήματος (πόδια), το πλάτος του οχήματος (πόδια), τον τύπο του οχήματος (αυτοκίνητο, φορτηγό ή μοτοσυκλέτα), την ταχύτητα του οχήματος (πόδια/s), την επιτάχυνση του οχήματος (πόδια/s²), τον αριθμό λωρίδας κυκλοφορίας, τον αριθμό του προπορευόμενου οχήματος, τον αριθμό οχήματος που ακολουθεί, την απόσταση μεταξύ του μπροστινού μέρους ενός οχήματος και του μπροστινού μέρους του προπορευόμενου οχήματος, τον απαιτούμενο χρόνο ώστε το μπροστινό μέρος του οχήματος να βρεθεί στη θέση του μπροστινού μέρους του προπορευόμενου οχήματος (s).

Από τα διαθέσιμα δεδομένα επιλέγονται τυχαία 17 τετράδες οχημάτων που κινούνται σε σειρά, το οποίο γίνεται εύκολα αντιληπτό ελέγχοντας τον αριθμό του μπροστινού και πίσω οχήματος. Τα οχήματα αυτά εξετάζονται μόνο για τη χρονική διάρκεια που κινούνται στην ίδια λωρίδα και βρίσκονται σε ακολουθία το ένα πίσω από το άλλο. Με αυτόν τον τρόπο προκύπτουν τετράδες οχημάτων που κινούνται σε σειρά αντίστοιχα με αυτές που εξετάστηκαν στο προηγούμενο κεφάλαιο. Για κάθε όχημα χρησιμοποιούνται τα εξής στοιχεία:

- ο αριθμός ταυτοποίησης του οχήματος
- ο τύπος του οχήματος (επιλέγονται μόνο αυτοκίνητα)
- ο χρόνος (ms)
- συντεταγμένη X σε παγκόσμιο σύστημα αναφοράς (πόδια)
- συντεταγμένη Y σε παγκόσμιο σύστημα αναφοράς (πόδια)
- μήκος του οχήματος
- η ταχύτητα του οχήματος (πόδια/s)
- η απόσταση μεταξύ του μπροστινού μέρους ενός οχήματος και του μπροστινού μέρους του προπορευόμενου οχήματος
- ο αριθμός ταυτοποίησης του μπροστινού οχήματος
- ο αριθμός ταυτοποίησης του οχήματος που ακολουθεί
- ο αριθμός της λωρίδας κυκλοφορίας

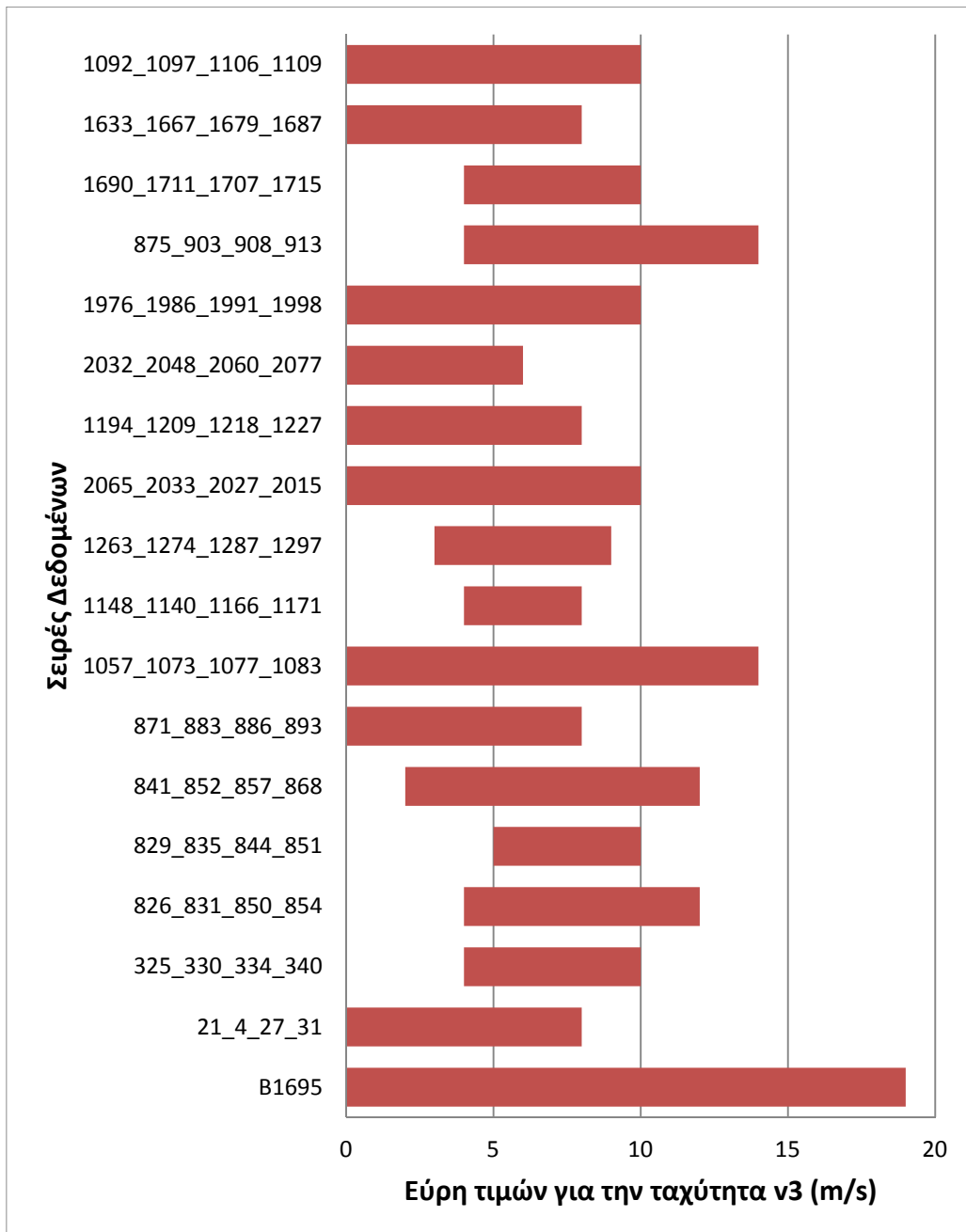
Στον παρακάτω πίνακα παρουσιάζονται οι τετράδες οχημάτων που επιλέχθηκαν τυχαία από το διαθέσιμο σετ δεδομένων.

Τετράδες οχημάτων που κινούνται σε σειρά	Διάρκεια (s)	Ωρα καταγραφής
21_4_27_31	27.7	4:00 μ.μ. έως 4:15 μ.μ.
826_831_850_854	41.1	
829_835_844_851	10.2	
841_852_857_868	47.6	
875_903_908_913	51.8	
1057_1073_1077_1083	30.3	
1690_1711_1707_1715	48.6	
871_883_886_893	50.2	5:00 μ.μ. έως 5:15 μ.μ.
1194_1209_1218_1227	86.3	
1976_1986_1991_1998	102.7	
2032_2048_2060_2077	24.4	
1092_1097_1106_1109	90.3	
325_330_334_340	37.3	5:15 μ.μ. έως 5:30 μ.μ.
1148_1140_1166_1171	20.2	
1263_1274_1287_1297	13	
2065_2033_2027_2015	81.9	
1633_1667_1679_1687	125.3	

Πίνακας 5.1.: Δεδομένα NGSIM που θα χρησιμοποιηθούν για τον έλεγχο εγκυρότητας των μοντέλων

5.2. Εύρη ταχυτήτων και δυνατότητα εκτίμησης

Με τα νέα δεδομένα θα ελεγχθεί ξανά η εγκυρότητα της μεθόδου loess και του μοντέλου GIPRS. Πρέπει, όμως, να προηγηθεί ένας έλεγχος για τη δυνατότητα εκτίμησης των ταχυτήτων στα νέα δεδομένα σύμφωνα με το πρότυπο που προέκυψε από τη σειρά B1695 των προηγούμενων δεδομένων. Για να είναι εφικτό αυτό θα πρέπει το εύρος ταχυτήτων που περιλαμβάνει η σειρά δεδομένων B1695 να συμπεριλαμβάνει τις ταχύτητες από τις νέες σειρές δεδομένων. Για τον σκοπό αυτό δημιουργήθηκε το διάγραμμα 5.1. Από το διάγραμμα αυτό είναι φανερό ότι η σειρά B1695 περιλαμβάνει το πιο εκτεταμένο εύρος και επομένως η εκτίμηση των ταχυτήτων στα νέα δεδομένα θα είναι εφικτή.



Διάγραμμα 5.1.: Εύρη ταχυτήτων για τα δεδομένα NGSIM

5.3. Υπολογισμός παραγόμενων στοιχείων

Πριν από κάθε υπολογισμό τα δεδομένα που αναφέρονται σε μήκος μετατρέπονται από πόδια σε m (1 πόδι = 0.3048m) και ο χρόνος από millisecond σε second (1 ms = 10^{-3} s). Τα δεδομένα που απαιτούνται για την εφαρμογή του μοντέλου Gipps και της μεθόδου Ioess έχουν δοθεί ήδη υπολογισμένα (οι ταχύτητες των οχημάτων και οι μεταξύ τους αποστάσεις). Για την εφαρμογή όμως του μοντέλου Gipps υπολογίστηκε ακόμα η σχετική θέση των οχημάτων ορίζοντας ως αρχή έναρξης της μέτρησης τη

θέση που βρίσκεται το μπροστινό μέρος του πρώτου οχήματος την πρώτη χρονική στιγμή (σχήμα 5.1). Οι ταχύτητες των οχημάτων έχουν δοθεί υπολογισμένες. Για τον υπολογισμό της πραγματικής απόστασης αφαιρείται το μήκος του οχήματος από την απόσταση μεταξύ των μπροστινών μέρων των οχημάτων που δίνεται. Ελέγχθηκε η πραγματική απόσταση μεταξύ των οχημάτων να είναι πάντα θετική. Σε διαφορετική περίπτωση θα σήμαινε ότι υπάρχει σύγκρουση των οχημάτων και προφανώς θα υπήρχε κάποιο χονδροειδές σφάλμα στα δεδομένα

Για τον υπολογισμό της σχετικής θέσης ακολουθείται η ίδια διαδικασία με το κεφάλαιο 3.2. Η απόσταση που διανύθηκε ανάμεσα σε δυο διαδοχικές χρονικές στιγμές ($t_2-t_1=0.1s$) για το κάθε όχημα υπολογίζεται από τη σχέση 5.1.

$$s_i(t_2) = \sqrt{(X_i(t_2) - X_i(t_1))^2 + (Y_i(t_2) - Y_i(t_1))^2} \quad (5.1)$$

Στη συνέχεια υπολογίζεται η σχετική θέση (ss) του κάθε οχήματος για τη διαδρομή που διανύουν, δηλαδή πρόκειται για το άθροισμα των αποστάσεων που διανύουν ανάμεσα σε διαδοχικές χρονικές στιγμές. Ως αρχή έναρξης της μέτρησης ορίζεται το σημείο έναρξης της διαδρομής για το πρώτο όχημα, στο σημείο $ss = 0$, όπως φαίνεται στο σχήμα 5.2. Επομένως, για το όχημα 1 η σχετική θέση του υπολογίζεται από τη σχέση 5.2.

$$ss_1(t_i) = \sum_0^{t_i} s_1(t_i) \quad (5.2)$$

Επειδή, όμως, τα υπόλοιπα οχήματα ξεκινούν πριν από το σημείο έναρξης της χιλιομετρικής μέτρησης, για τον ορισμό της χιλιομετρικής θέσης τους θα αθροιστούν οι διανυόμενες αποστάσεις των επιμέρους χρονικών διαστημάτων όπως έγινε για το όχημα 1, αλλά θα αφαιρεθεί η απόσταση που διένυσαν μέχρι να φτάσουν στο σημείο έναρξης, δηλαδή την αρχική απόσταση που είχαν από το πρώτο όχημα τη χρονική στιγμή t_0 , σύμφωνα με το σχήμα 5.3. Επομένως, προκύπτουν οι σχέσεις:

$$ss_2(t_i) = \sum_0^{t_i} s_2(t_i) - d_{12}(0) \quad (5.3)$$

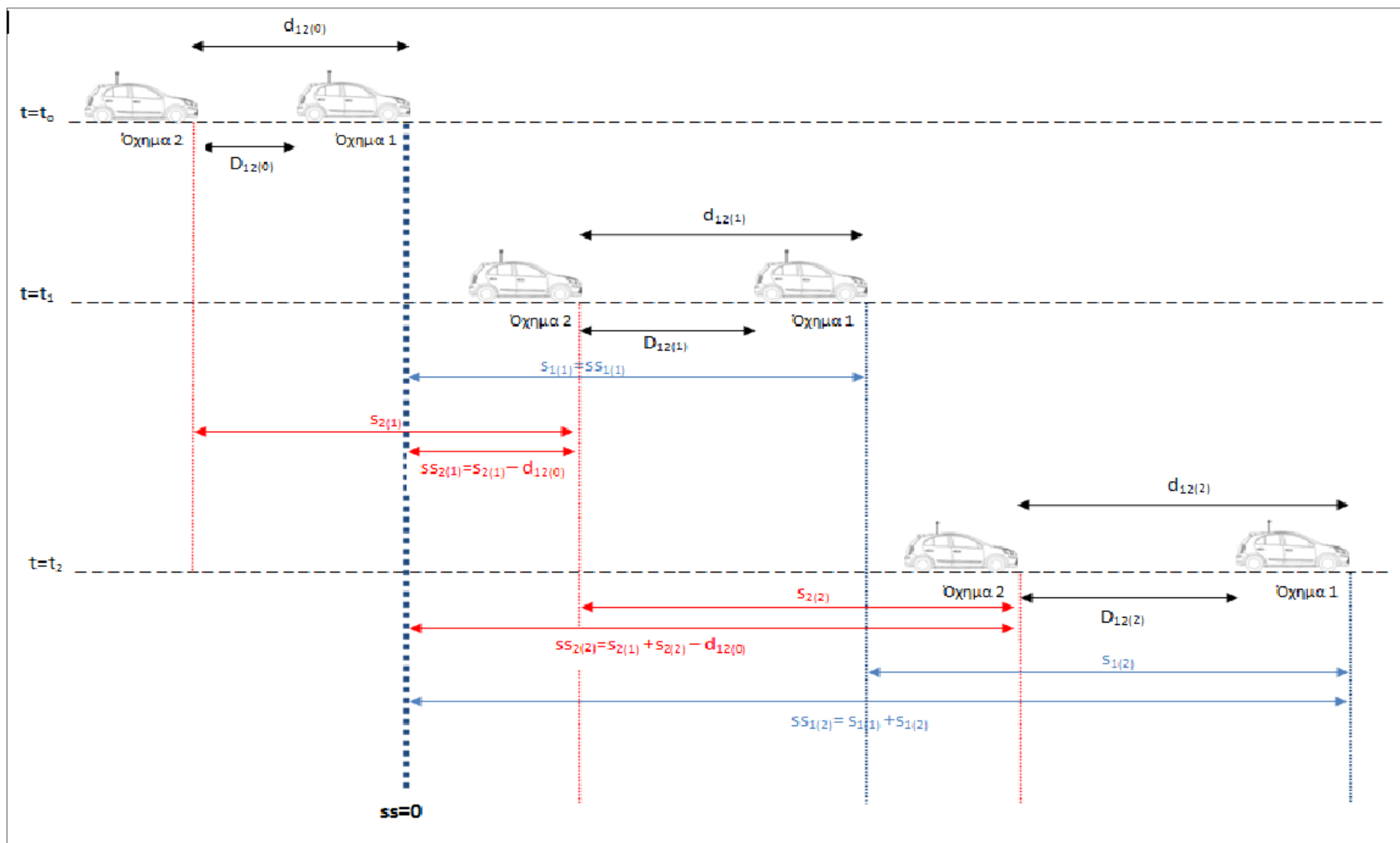
$$ss_3(t_i) = \sum_0^{t_i} s_3(t_i) - d_{13}(0) \quad (5.4)$$

$$ss_4(t_i) = \sum_0^{t_i} s_4(t_i) - d_{14}(0) \quad (5.5)$$

Όπου οι αποστάσεις d_{1j} υπολογίζονται από τον τύπο:

$$d_{1j}(t) = \sqrt{(X_j(t) - X_1(t))^2 + (Y_j(t) - Y_1(t))^2} \quad (5.6)$$

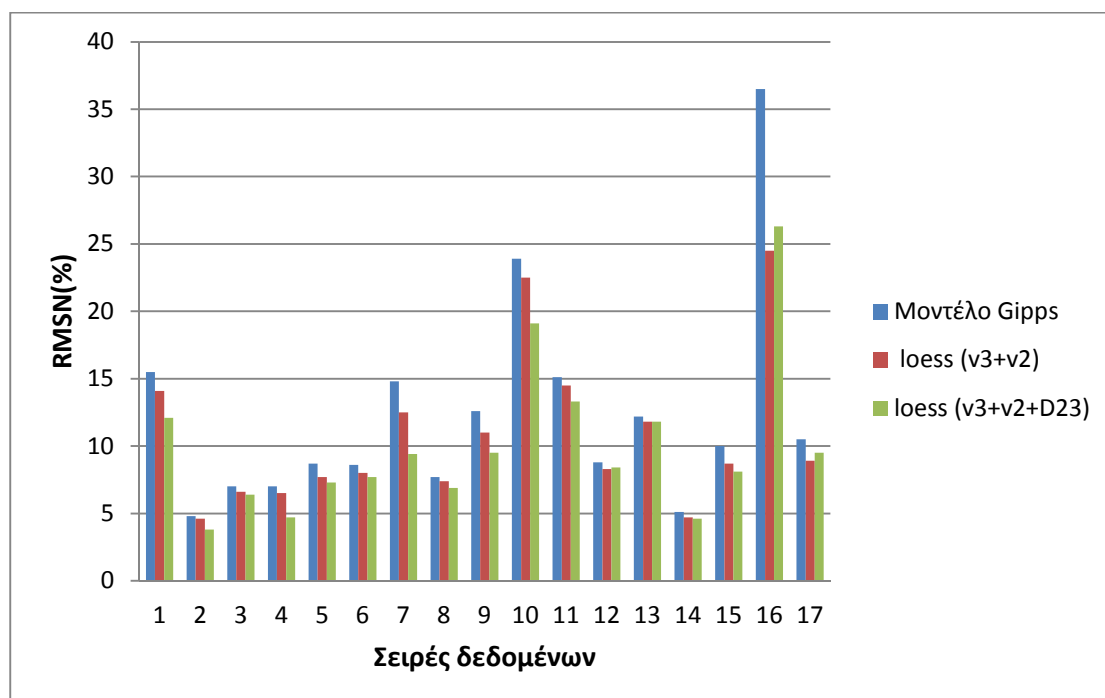
Σημειώνεται ότι οι παραπάνω υπολογισμοί έγιναν με τη βοήθεια του λογισμικού R project for statistical computing, ενώ χρησιμοποιήθηκε η διεπαφή Rstudio.



Σχήμα 5.3.: Σκαρίφημα σχετικών αποστάσεων μεταξύ των οχημάτων για τα δεδομένα NGSIM

5.4. Εφαρμογή μοντέλου Gipps και της προτεινόμενης μεθόδου

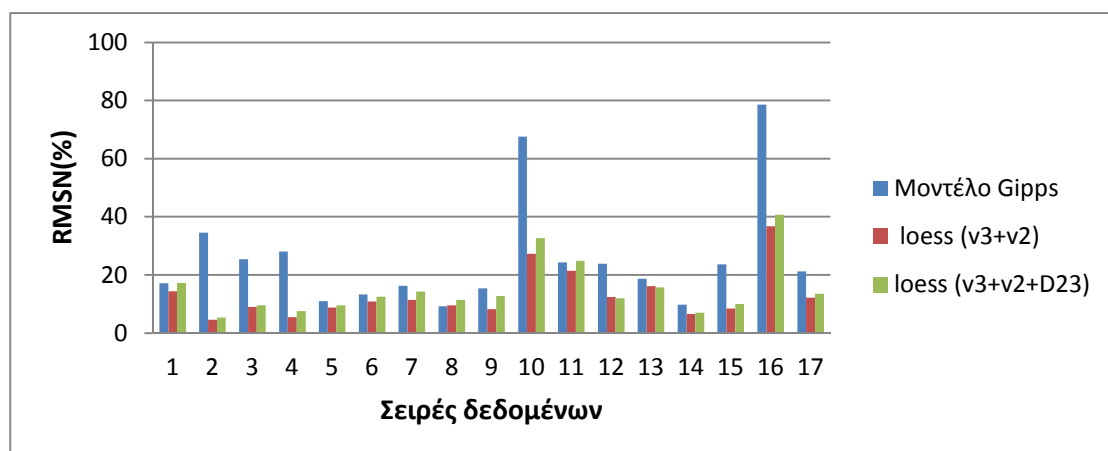
Για τα δεδομένα του πίνακα 5.1. επιχειρείται να εφαρμοστεί το μοντέλο του Gipps για τις ίδιες παραμέτρους που εφαρμόστηκε στα δεδομένα του προηγούμενου κεφαλαίου και πιο συγκεκριμένα για τις τιμές που ελαχιστοποιούσαν το σφάλμα του: $\tau=0.4$ s, $\alpha_n=0.8$ m/s², $V_n=14$ m/s, $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s² και $\hat{b}=-3.0$ m/s² ή για μια πιο ρεαλιστική προσέγγιση $\tau=1.0$ s, $\alpha_n=1.6$ m/s², $s_{n-1}=5.6$ m, $b_n=-5.2$ m/s², $\hat{b}=-3.0$ m/s² και $V_n=16$ m/s. Η μέθοδος loess εφαρμόζεται, επίσης, για $\tau=0.4$ s ή $\tau=1.0$ s, span=0.75 και degree=1, όπως επιλέχτηκε για τα δεδομένα του προηγούμενου κεφαλαίου. Οι παράγοντες που θα ληφθούν υπόψη για την παλινδρόμηση είναι οι ίδιοι που χρησιμοποιεί το μοντέλο του Gipps (v3, v2, D23) ώστε να είναι άμεση η σύγκριση τους ως προς τα δεδομένα που εισάγονται. Επίσης, έγινε μια δοκιμή με λιγότερους παράγοντες από αυτούς που απαιτεί το μοντέλο του Gipps (συγκεκριμένα τους v3,v2). Τα αποτελέσματα παρουσιάζονται στο διάγραμμα 5.2 και πιο αναλυτικά στον πίνακα 5.2 για $\tau=0.4$ s, ενώ για χρόνο αντίδρασης $\tau=1$ s τα αποτελέσματα φαίνονται στο διάγραμμα 5.3. και πίνακα 5.3 αντίστοιχα.



Διάγραμμα 5.2.: Σύγκριση RMSN (%) για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=0.4$ s

A/A	Σειρές δεδομένων	Εφαρμογή του μοντέλου Gipp's (g1) RMSN(%)	Εφαρμογή loess (v3, v2, D23)- (μοντέλο If) RMSN(%)	Βελτίωση (%) σε σχέση με το μοντέλο Gipp's	Εφαρμογή loess (v3, v2) - (μοντέλο 11s) RMSN(%)	Βελτίωση (%) σε σχέση με το μοντέλο Gipp's
1	21_4_27_31	15.5	12.1	21.9	14.1	9.0
2	325_330_334_340	4.8	3.8	20.8	4.6	4.2
3	826_831_850_854	7	6.4	8.6	6.6	5.7
4	829_835_844_851	7	4.7	32.9	6.5	7.1
5	841_852_857_868	8.7	7.3	16.1	7.7	11.5
6	871_883_886_893	8.6	7.7	10.5	8.0	7.0
7	1057_1073_1077_1083	14.8	9.4	36.5	12.5	15.5
8	1148_1140_1166_1171	7.7	6.9	10.4	7.4	3.9
9	1263_1274_1287_1297	12.6	9.5	24.6	11.0	12.7
10	2065_2033_2027_2015	23.9	19.1	20.1	22.5	5.9
11	1194_1209_1218_1227	15.1	13.3	11.9	14.5	4.0
12	2032_2048_2060_2077	8.8	8.4	4.5	8.3	5.7
13	1976_1986_1991_1998	12.2	11.8	3.3	11.8	3.3
14	875_903_908_913	5.1	4.6	9.8	4.7	7.8
15	1690_1711_1707_1715	10	8.1	19.0	8.7	13.0
16	1633_1667_1679_1687	36.5	24.5	32.9	26.3	27.9
17	1092_1097_1106_1109	10.5	8.9	15.2	9.5	9.5

Πίνακας 5.2.: Σύγκριση RMSN (%) του μοντέλου Gipp's και της μεθόδου loess για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=0.4$ s



Διάγραμμα 5.3.: Σύγκριση RMSN (%) για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=1.0$ s

A/A	Σειρές δεδομένων	Εφαρμογή του μοντέλου Gipp's (G3) RMSN(%)	Εφαρμογή loess (v3, v2, D23)- (μοντέλο l2f) RMSN(%)	Βελτίωση (%) σε σχέση με το μοντέλο Gipp's	Εφαρμογή loess (v3, v2)- (μοντέλο l2s) RMSN(%)	Βελτίωση (%) σε σχέση με το μοντέλο Gipp's
1	21_4_27_31	17.2	14.4	16.3	17.3	-0.6
2	325_330_334_340	34.5	4.6	86.7	5.4	84.3
3	826_831_850_854	25.4	9	64.6	9.6	62.2
4	829_835_844_851	28	5.5	80.4	7.6	72.9
5	841_852_857_868	11	8.8	20.0	9.6	12.7
6	871_883_886_893	13.3	10.9	18.0	12.5	6.0
7	1057_1073_1077_1083	16.3	11.4	30.1	14.3	12.3
8	1148_1140_1166_1171	9.2	9.5	-3.3	11.4	-23.9
9	1263_1274_1287_1297	15.4	8.2	46.8	12.8	16.9
10	2065_2033_2027_2015	67.6	27.3	59.6	32.7	51.6
11	1194_1209_1218_1227	24.3	21.4	11.9	24.9	-2.5
12	2032_2048_2060_2077	23.9	12.4	48.1	12	49.8
13	1976_1986_1991_1998	18.7	16.2	13.4	15.7	16.0
14	875_903_908_913	9.8	6.6	32.7	7	28.6
15	1690_1711_1707_1715	23.7	8.5	64.1	10	57.8
16	1633_1667_1679_1687	78.6	36.7	53.3	40.7	48.2
17	1092_1097_1106_1109	21.2	12.2	42.5	13.5	36.3

Πίνακας 5.3.: Σύγκριση RMSN (%) του μοντέλου Gipp's και της μεθόδου loess για τα δεδομένα NGSIM για χρόνο αντίδρασης $\tau=1.0$ s

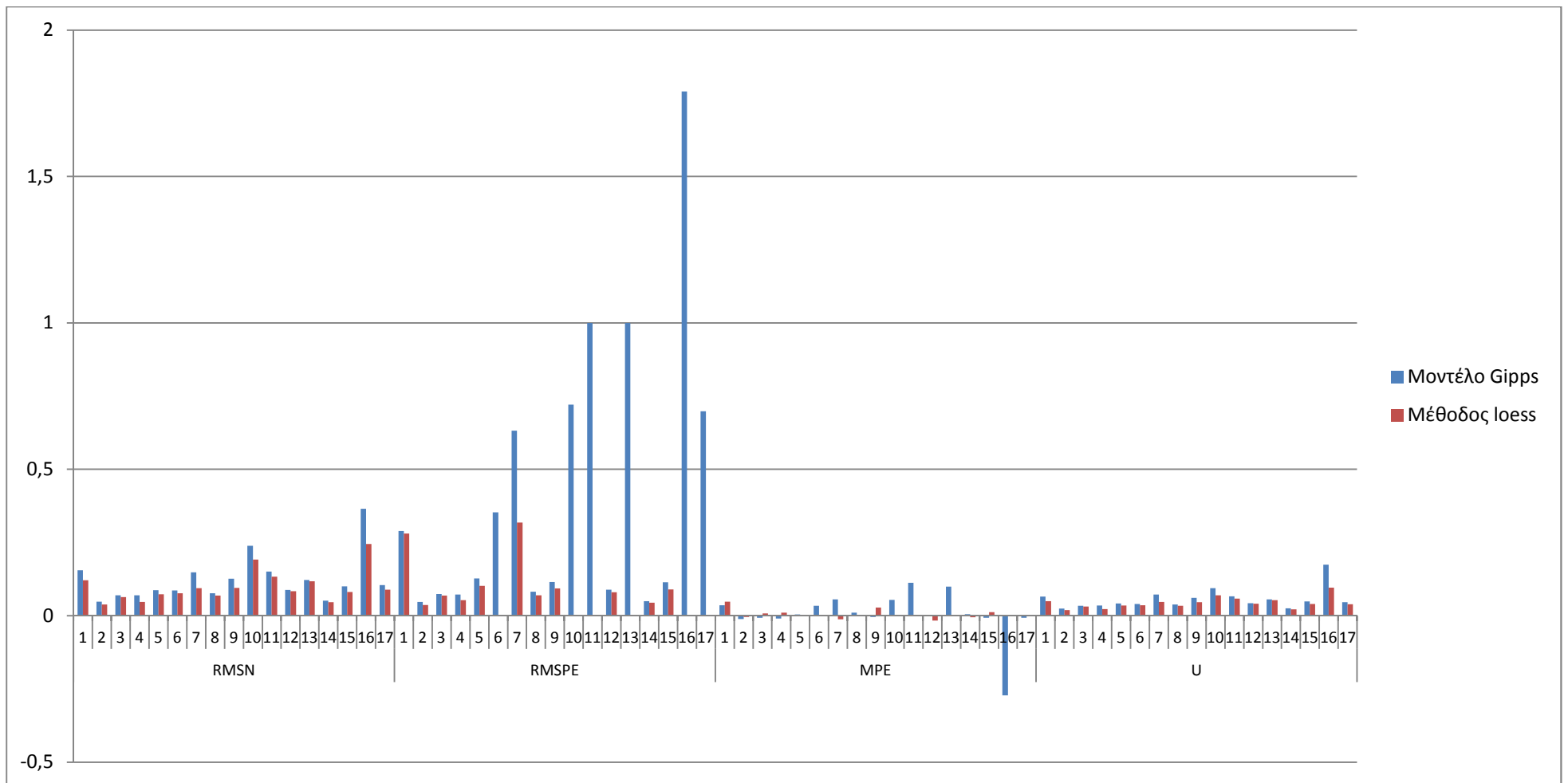
Σύμφωνα με τους πίνακες 5.2., 5.3 και τα διαγράμματα 5.2., 5.3. η μέθοδος loess παρείχε καλύτερα αποτελέσματα από το μοντέλο Gipp's για όλες τις σειρές δεδομένων που εξετάστηκαν. Αξιοσημείωτο είναι ότι το ίδιο επετεύχθη ακόμα και όταν χρησιμοποιήθηκαν μόνο οι ταχύτητες των δυο οχημάτων. Ωστόσο, για όλες τις σειρές δεδομένων η μέθοδος loess έδωσε καλύτερα αποτελέσματα όταν λήφθηκαν υπόψη και οι τρεις παράγοντες. Όταν ο χρόνος θεωρήθηκε 1 s τα σφάλματα αυξήθηκαν σε όλες τις περιπτώσεις τόσο για μοντέλο Gipp's όσο και για τη μέθοδο loess. Παρόλα αυτά στο μοντέλο Gipp's αυξήθηκαν δραματικά.

Επίσης, είναι πιθανό το μοντέλο του Girps να έδινε καλύτερα αποτελέσματα αν άλλαζαν οι τιμές των παραμέτρων. Ωστόσο, κάτι τέτοιο είναι χρονοβόρο και δεν απαιτήθηκε από τη μέθοδο loess. Επομένως, οι τιμές των παραμέτρων φαίνεται ότι επηρεάζουν την απόδοση του μοντέλου και η βέλτιστη απόδοση του πιθανόν να βασίζεται στην υπερπροσαρμογή τους στα εκάστοτε δεδομένα. Το γεγονός αυτό καθιστά το μοντέλο δυσλειτουργικό αν για κάθε νέα σειρά δεδομένων που προκύπτει απαιτείται εκ νέου βαθμονόμηση των παραμέτρων του. Αντιθέτως στην εφαρμογή της μεθόδου loess οι παράμετροι span και degree δεν διαδραματίζουν τόσο καθοριστικό ρόλο στην απόδοση της.

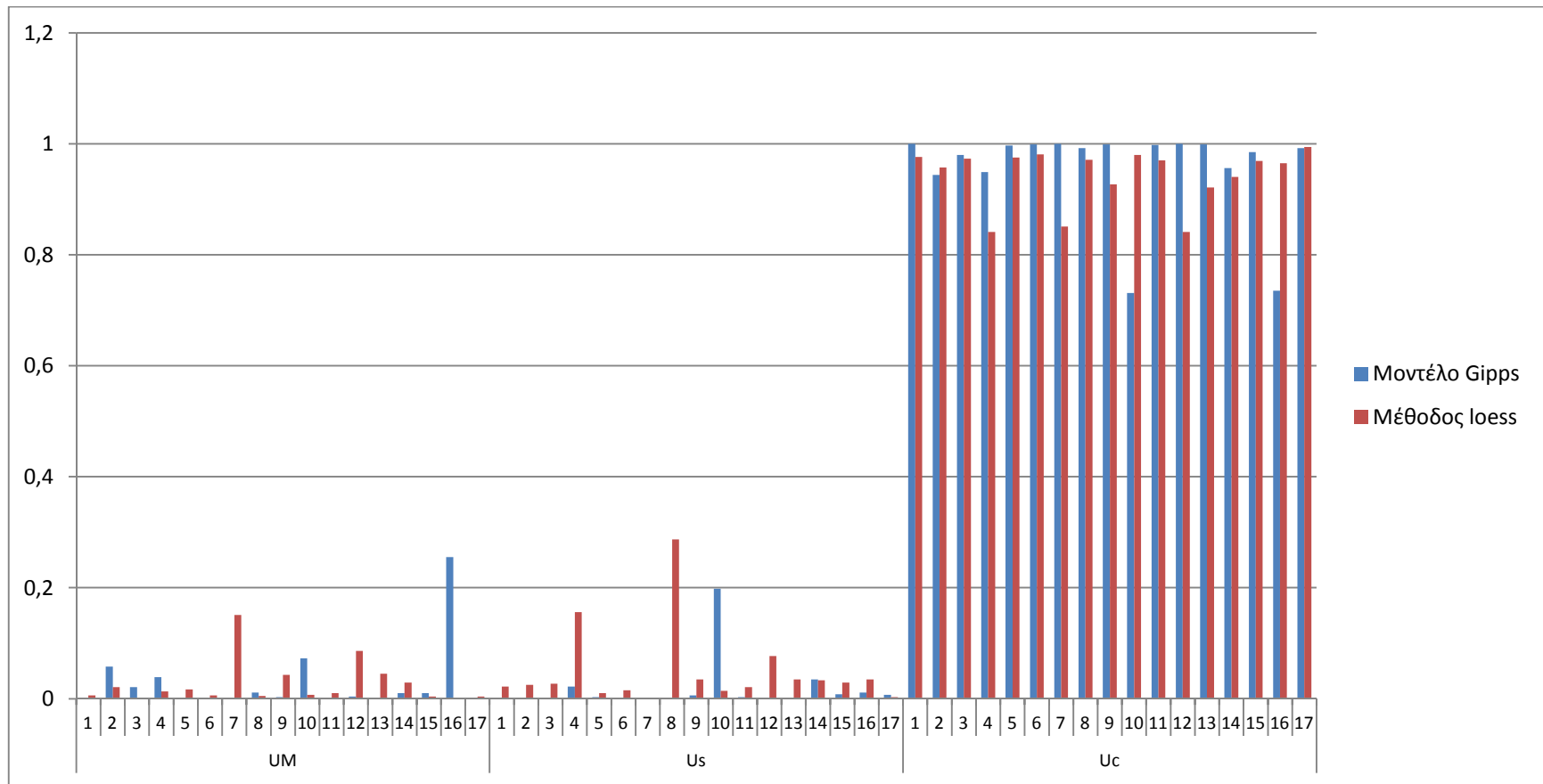
Τέλος, για μια καλύτερη σύγκριση των δυο μεθόδων υπολογίζονται και άλλα μέτρα αξιολόγησης και παρουσιάζονται στα διάγραμμα 5.4., 5.5. Προτιμάται η αξιολόγηση της μεθόδου για την ακραία τιμή του χρόνου αντίδρασης που το μοντέλο του Girps παρουσιάζει καλύτερα αποτελέσματα. Έτσι αν για τις ακραίες τιμές που ελαχιστοποιούν το σφάλμα του μοντέλου του Girps, η μέθοδος loess παρέχει συγκριτικά καλύτερη εκτίμηση, τότε για πιο λογικές τιμές του χρόνου αντίδρασης η μέθοδος loess παρουσιάζει ακόμα μεγαλύτερη βελτίωση σε σχέση με τα αποτελέσματα του μοντέλου του Girps. Έτσι, αν η μέθοδος loess παρουσιάζεται καλύτερη για χρόνο αντίδρασης $\tau=0.4$ s, τότε το ίδιο θα ισχύει για χρόνο αντίδρασης $\tau=1.0$ s που το μοντέλο Girps εμφανίζει χειρότερη συμπεριφορά.

Στο διάγραμμα 5.4., από τη σύγκριση των τιμών του RMSN συμπεραίνουμε ότι η μέθοδος loess προσεγγίζει καλύτερα τις πραγματικές παρατηρήσεις και έχει μικρότερο συνολικό σφάλμα από το μοντέλο του Girps. Το RMSPE υποδεικνύει μεγάλες αποκλίσεις για το μοντέλο του Girps για αρκετές σειρές δεδομένων, ενώ δεν παρατηρούνται τόσο μεγάλες αποκλίσεις με τη μέθοδο loess. Το μέτρο MPE δεν δείχνει να υπάρχει μεγάλη συστηματική υποτίμηση ή υπερτίμηση της εκτίμησης για κάποια από τις δύο μεθόδους, με εξαίρεση λίγες σειρές δεδομένων για το μοντέλο του Girps. Το μέτρο αξιολόγησης U επιβεβαιώνει την εγκυρότητα της μεθόδου loess, καθώς για την πλειοψηφία των δεδομένων έχει μικρότερη τιμή για τη μέθοδο loess σε σχέση με το μοντέλο του Girps.

Τέλος, στο διάγραμμα 5.5. εξετάστηκαν οι επιμέρους αναλογίες του συντελεστή Theil που όσο οι τιμές των U_M , U_S πλησιάζουν στο 1, τόσο καλύτερη προσαρμογή υπάρχει στα δεδομένα. Από τα μέτρα αυτά δεν υπάρχει σαφής εικόνα για το ποια μέθοδος είναι καλύτερη.



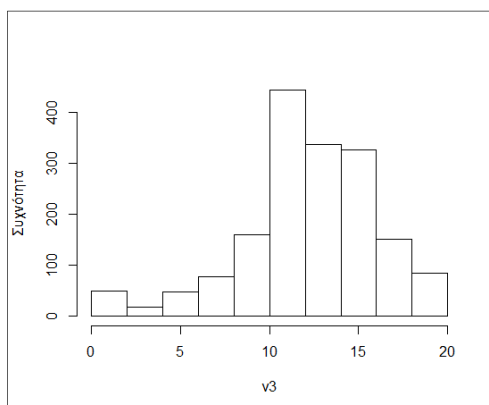
Διάγραμμα 5.4.: Σύγκριση του μοντέλου Gyrps και της μεθόδου loess για τα NGSIM δεδομένα με τη χρήση των μέτρων αξιολόγησης RMSN, RMSPE, MPE, U



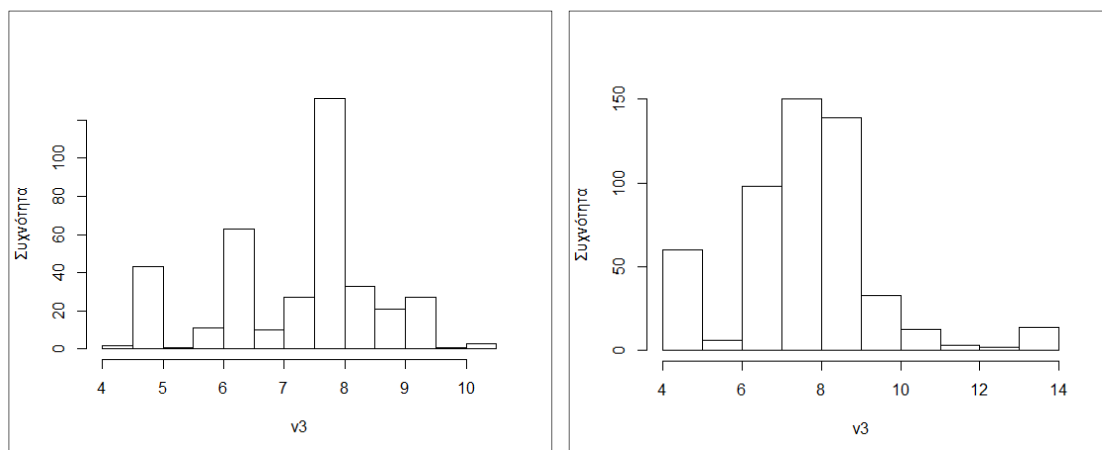
Διάγραμμα 5.5.: Σύγκριση του μοντέλου Gipps και της μεθόδου loess για τα NGSIM δεδομένα με τη χρήση των μέτρων αξιολόγησης U_M , U_S , U_C

5.5. Αιτιολογία μειωμένης απόδοσης της μεθόδου loess για κάποιες σειρές δεδομένων

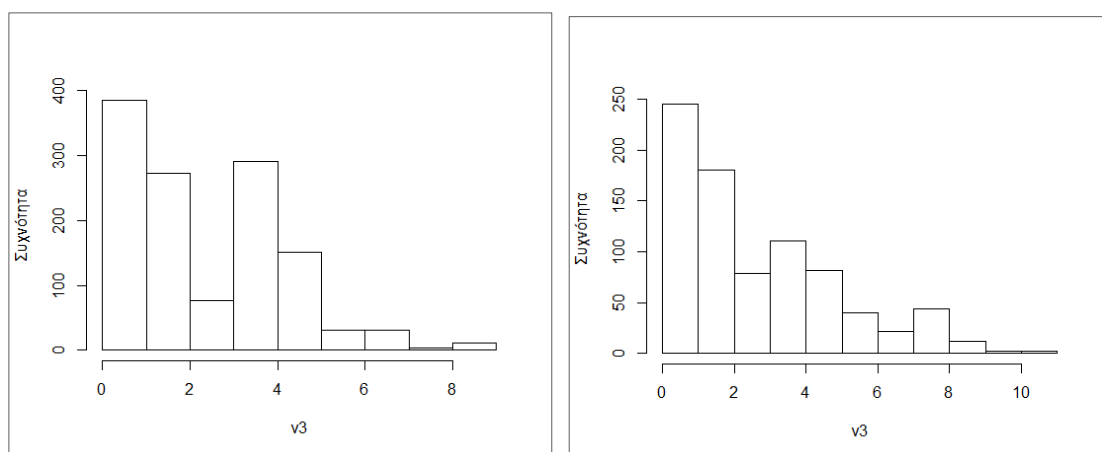
Όπως παρατηρείται σε κάποιες σειρές δεδομένων το σφάλμα της εκτίμησης είναι πολύ μικρό, ενώ σε άλλες μεγαλύτερο. Συγκεκριμένα σε 5 από τις 17 σειρές δεδομένων παρατηρήθηκαν σφάλματα εκτίμησης RMSN μεγαλύτερα από 10%. Τα μεγαλύτερα σφάλματα εξηγούνται παρατηρώντας τα προφίλ των ταχυτήτων. Στη σειρά δεδομένων 1633_1667_1679_1687 παρατηρήθηκε το μεγαλύτερο σφάλμα εκτίμησης, καθώς υπάρχουν πολλές μηδενικές ταχύτητες λόγω τις κυκλοφοριακής συμφόρησης, κάτι που δεν παρατηρήθηκε στο B1695 σύμφωνα με το οποίο γίνεται η εκτίμηση. Παρατηρώντας τα διαγράμματα 5.6, 5.7, 5.8 είναι φανερό ότι στις σειρές δεδομένων που παρατηρήθηκε μεγάλο σφάλμα υπήρχε μεγάλη συχνότητα από πολύ μικρές ταχύτητες 0-2 m/s, δεδομένου ότι τα στοιχεία αυτά λήφθηκαν σε συνθήκες κυκλοφοριακής συμφόρησης. Αντιθέτως, στη σειρά B1695 στην οποία βασίζεται το πρότυπο εκτίμησης, υπήρχε μεγάλη συχνότητα εμφάνισης των ταχυτήτων 6-10 m/s και δεν υπήρχαν τόσο μικρές ταχύτητες σε μεγάλη συχνότητα. Επομένως η χρήση μιας σειράς δεδομένων με παρόμοια χαρακτηριστικά θα μπορούσε να επιφέρει καλύτερα αποτελέσματα.



Διάγραμμα 5.6.: Προφίλ ταχυτήτων για τη σειρά δεδομένων B1695 από την οποία προκύπτει το πρότυπο ακολουθίας οχημάτων με τη μέθοδο loess

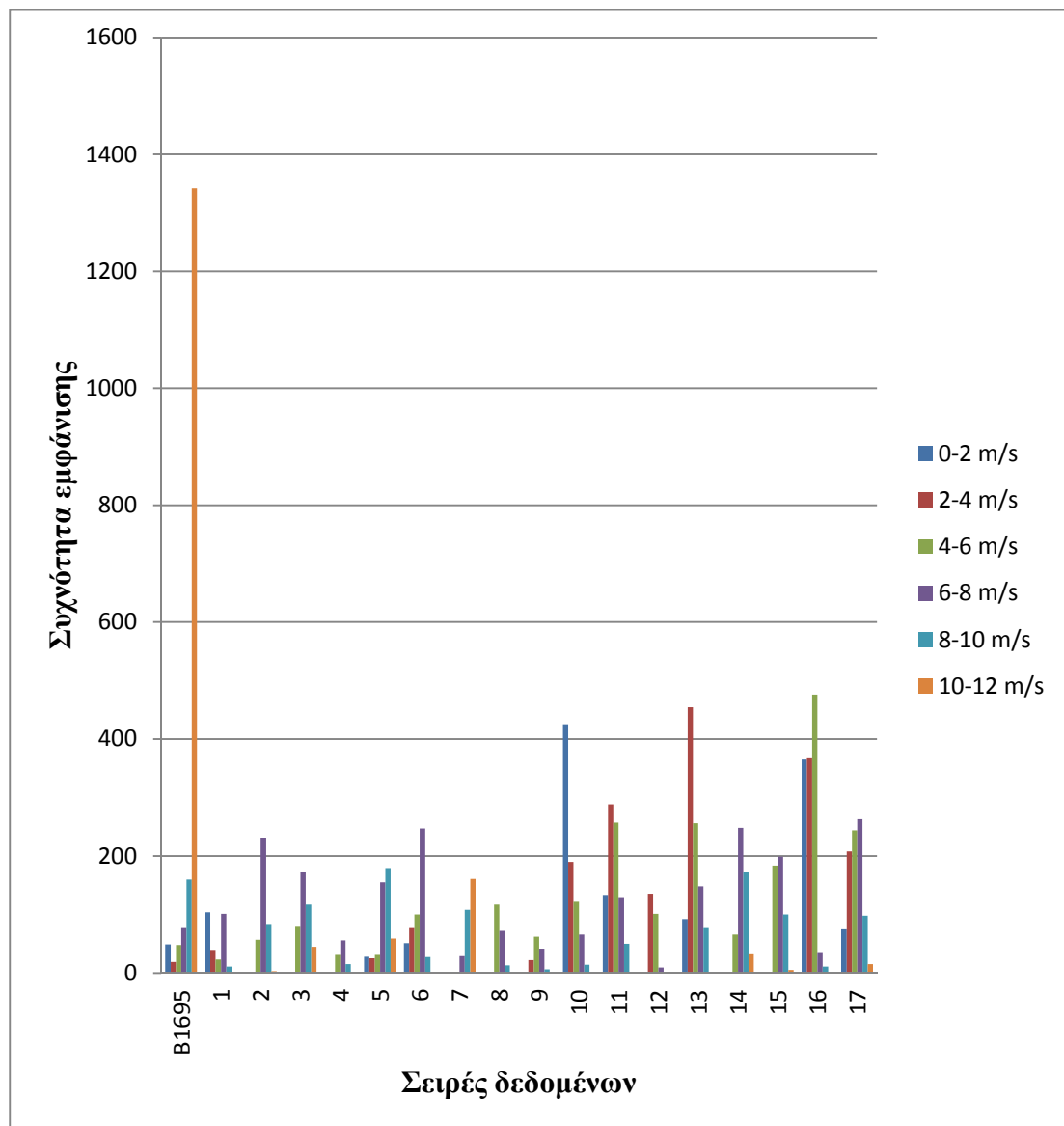


Διάγραμμα 5.7.: Προφίλ ταχυτήτων για σειρές δεδομένων από τα NGSIM δεδομένα για τα οποία παρατηρήθηκε καλή εκτίμηση με τη μέθοδο loess



Διάγραμμα 5.8.: Προφίλ ταχυτήτων για σειρές δεδομένων από τα NGSIM δεδομένα για τα οποία δεν παρατηρήθηκε καλή εκτίμηση με τη μέθοδο loess

Πιο αναλυτικά στο διάγραμμα 5.9. παρουσιάζεται η συχνότητα εμφάνισης κάθε τιμής της προς εκτίμηση ταχύτητας v_3 για κάθε σειρά δεδομένων. Όπως παρατηρείται για τις σειρές δεδομένων 1,10,11,13,16 για τις οποίες υπάρχουν αρκετές ταχύτητες μεγέθους 0-2 ή 2-4 m/s, η εκτίμηση με τη μέθοδο loess δεν ήταν πολύ καλή. Αυτό ήταν αναμενόμενο, εφόσον το πρότυπο εκτίμησης δημιουργήθηκε σύμφωνα με τη σειρά δεδομένων B1695, στην οποία δεν υπάρχουν αρκετές ταχύτητες τέτοιου μεγέθους.



Διάγραμμα 5.9.: Συχνότητα εμφάνισης τιμών της προς εκτίμηση ταχύτητας v_3 για κάθε σειρά δεδομένων

5.6. Συμπεράσματα

Με τα NGSIM δεδομένα επιβεβαιώνεται η καλύτερη αξιοπιστία της μεθόδου loess έναντι του μοντέλου Gyrps. Το μοντέλο Gyrps είναι πιθανό να μπορούσε να εκτιμήσει τις ταχύτητες των νέων δεδομένων με αλλαγή στις τιμές των παραμέτρων του. Το γεγονός αυτό καθιστά το μοντέλο του Gyrps χρονοβόρο και δίνει πλεονέκτημα στη μέθοδο loess, η οποία κατά την εφαρμογή της περιλαμβάνει μόνο δύο παραμέτρους (span, degree) που δεν επηρεάζουν αρκετά το αποτέλεσμα. Για όλες τις σειρές των δεδομένων NGSIM η μέθοδος loess είχε καλύτερη απόδοση από το μοντέλο του Gyrps. Έτσι, αποδεικνύεται η ευελιξία της μεθόδου. Ωστόσο, σε κάποιες σειρές

δεδομένων από αυτές που χρησιμοποιήθηκαν παρατηρήθηκαν σχετικά μεγάλα σφάλματα, γεγονός το οποίο οφείλεται στη διαφορετική φύση των δεδομένων. π.χ. δεδομένα σε συνθήκες κυκλοφοριακής συμφόρησης. Όταν οι κυκλοφοριακές συνθήκες ήταν παρόμοιες με αυτές της σειράς δεδομένων σύμφωνα με την οποία δημιουργήθηκε το πρότυπο μοντέλο, τότε η εκτίμηση των ταχυτήτων ήταν καλύτερη. Το ζήτημα αυτό μπορεί να αντιμετωπιστεί χρησιμοποιώντας ως πρότυπη σειρά δεδομένων κάποια που να περιλαμβάνει παρόμοιες συνθήκες.

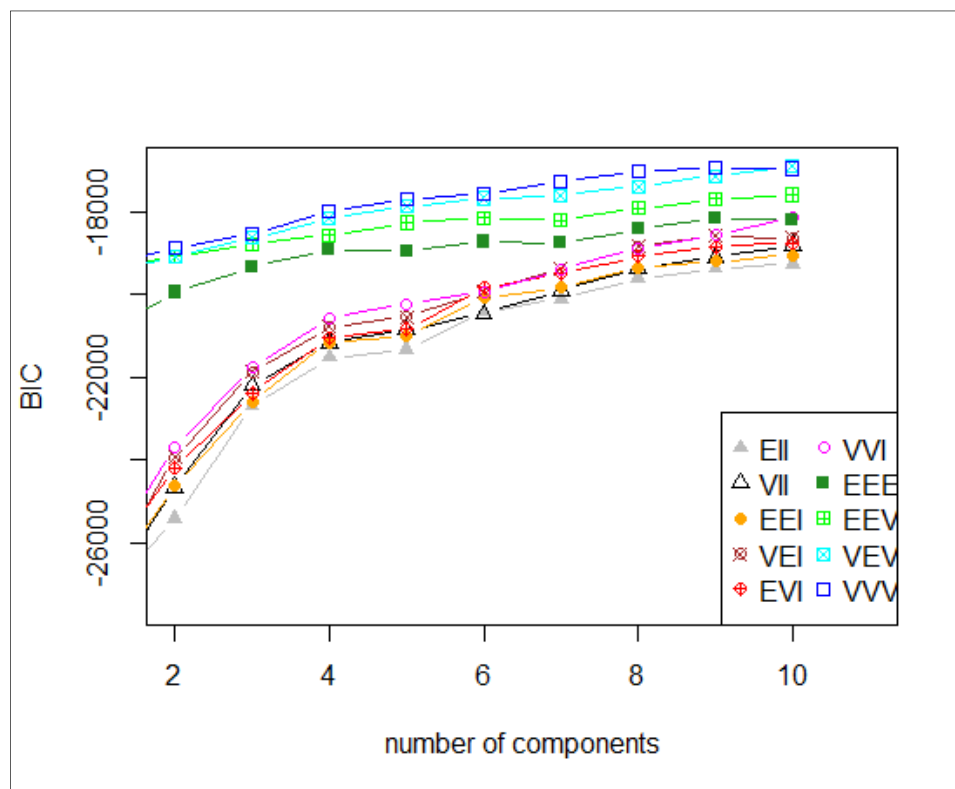
6. ΕΦΑΡΜΟΓΗ ΟΜΑΔΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ

Στο κεφάλαιο αυτό επιχειρείται μια δεύτερη προσέγγιση εκτίμησης προτύπων ακολουθίας οχημάτων χρησιμοποιώντας ένα συνδυασμό μεθόδων μηχανικής μάθησης και συγκεκριμένα της μεθόδου loess και των τεχνικών ομαδοποίησης και ταξινόμησης. Ελέγχεται αν η προσέγγιση αυτή μπορεί να παρέχει καλύτερα αποτελέσματα από την πρώτη προσέγγιση που περιλαμβάνει μόνο την εφαρμογή της μεθόδου loess και προκύπτουν τα σχετικά συμπεράσματα.

6.1. Εφαρμογή

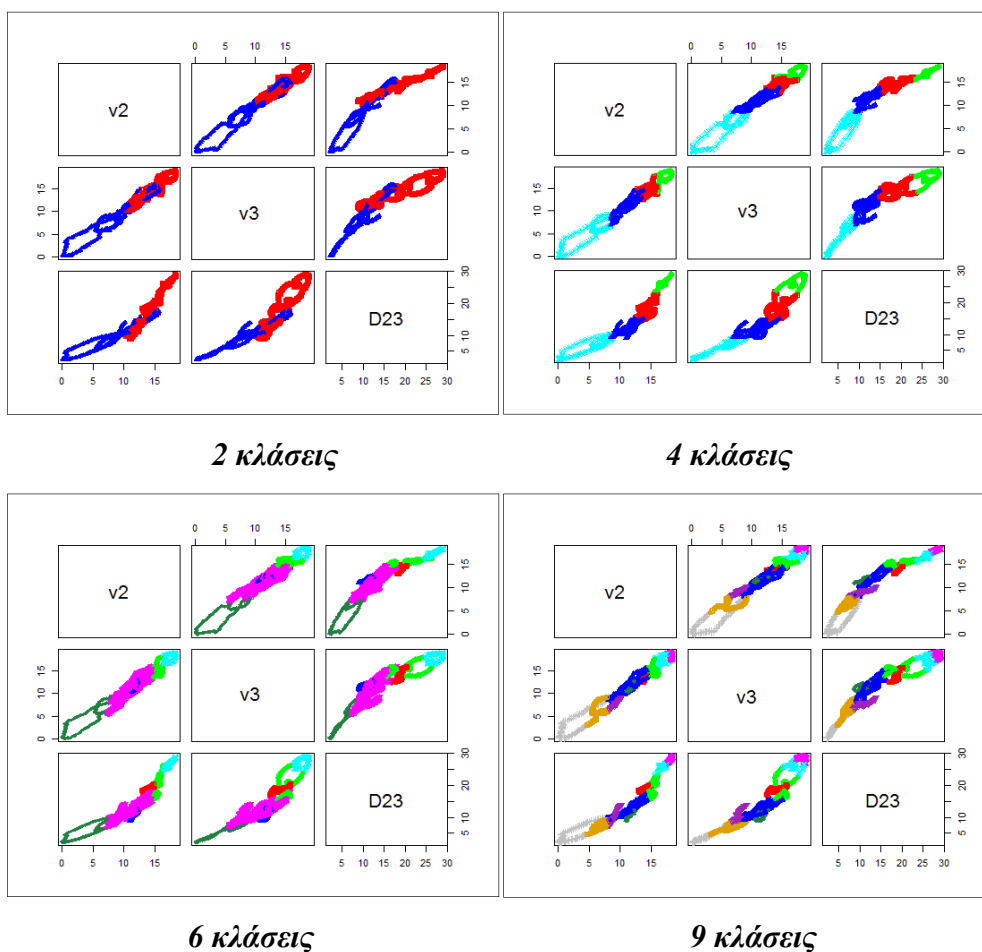
Αρχικά, για τη σειρά δεδομένων B1695 σύμφωνα με την οποία θα διαμορφωθεί ένα πρότυπο ακολουθίας οχημάτων, εφαρμόζεται μια ομαδοποίηση των δεδομένων της. Η διαδικασία της ομαδοποίησης βασίζεται στο συνδυασμό διάφορων μοντέλων, όπως περιγράφηκε στην ενότητα 3.2.3. Για την εφαρμογή της χρησιμοποιήθηκε το πακέτο Mclust (Fraley and Raftery, 2002) σε υπολογιστικό περιβάλλον Rstudio. Συγκεκριμένα, με την ομαδοποίηση αναγνωρίζονται οι διαφορετικές καταστάσεις (π.χ. κατάσταση επιτάχυνσης, κατάσταση επιβράδυνσης, κλπ.) και τα δεδομένα χωρίζονται σε ομάδες ανάλογα με τις ομοιότητες ή διαφορές που έχουν (clustering). Για την ομαδοποίηση οι παράγοντες που λήφθηκαν υπόψη είναι οι ταχύτητες των δύο οχημάτων v_2, v_3 και η μεταξύ τους απόσταση D23, καθώς στο προηγούμενο κεφάλαιο κρίθηκαν ως οι πιο καθοριστικοί για τη συμπεριφορά των οδηγών για τα δεδομένα που εξετάστηκαν. Με τον αλγόριθμο της ομαδοποίησης εξετάστηκαν διάφοροι συνδυασμοί μοντέλων, έτσι ώστε να επιλεχθεί ο βέλτιστος, καθώς και αναζητήθηκε ο βέλτιστος αριθμός κλάσεων. Με τη χρήση του προαναφερόμενου πακέτου υπολογίστηκε ο δείκτης BIC σύμφωνα με τον οποίο προτείνεται να επιλεχθεί ο αριθμός κλάσεων που ελαχιστοποιεί την τιμή του δείκτη αυτού. Στο διάγραμμα 6.1. παρουσιάζεται μια απεικόνιση των αποτελεσμάτων της ταξινόμησης για τους διάφορους συνδυασμούς μοντέλων και για διαφορετικό αριθμό κλάσεων (components), όσον αφορά στη σειρά δεδομένων B1695. Η μικρότερη τιμή του δείκτη BIC εμφανίζεται για τις 9 κλάσεις, όπως φαίνεται στο διάγραμμα 6.1. Επίσης, παρατηρείται ότι μεταξύ 7 και 9 κλάσεων οι διαφοροποιήσεις του δείκτη BIC είναι πολύ μικρές και κατά συνέπεια η προσαρμογή στα δεδομένα είναι σχεδόν παρόμοια για αυτές τις περιπτώσεις. Όταν ο αριθμός των κλάσεων μειωθεί ακόμα περισσότερο, για παράδειγμα η κατανομή των δεδομένων μέχρι και σε 4 κλάσεις, δεν προκαλεί

μεγάλη απώλεια σε σχέση με το βέλτιστο αριθμό κλάσεων. Επομένως, αξίζει να δοκιμαστούν και μικρότεροι αριθμοί κλάσεων αποσκοπώντας στην ανάπτυξη πιο απλών μοντέλων.



Διάγραμμα 6.1.: Επιλογή βέλτιστου αριθμού κλάσεων

Στο διάγραμμα 6.2. παρουσιάζονται τα αποτελέσματα της ομαδοποίησης για διάφορους αριθμούς κλάσεων. Όπως παρατηρείται από το διάγραμμα, όσο μικρότερος είναι ο αριθμός των κλάσεων τόσο πιο απλή είναι η ομαδοποίηση που γίνεται. Για μικρούς αριθμούς κλάσεων (π.χ. 2 ή 4) τα χαρακτηριστικά της κάθε κλάσης είναι περισσότερο διακριτά και εύκολα αναγνωρίσιμα. Αντιθέτως, για μεγαλύτερο αριθμό κλάσεων (π.χ. 6 ή 9) οι κυκλοφοριακές καταστάσεις ανά κλάση εμφανίζονται κάπως συγκεχυμένες.



Διάγραμμα 6.2.: Απεικόνιση των αποτελεσμάτων της ομαδοποίησης για διάφορους αριθμούς κλάσεων

Έπειτα, εφαρμόζεται η μέθοδος της τοπικά σταθμισμένης παλινδρόμησης σε κάθε κλάση που δημιουργήθηκε ξεχωριστά, έτσι ώστε να προκύψουν πιο συγκεκριμένα πρότυπα για κάθε κατάσταση που θα προσαρμόζονται καλύτερα στα εκάστοτε δεδομένα (fitting). Για την εφαρμογή της μεθόδου loess χρησιμοποιήθηκαν οι παράμετροι $\text{span}=0.75$, $\text{degree}=1$ και ως παράγοντες λήφθηκαν υπόψη οι ταχύτητες των δύο οχημάτων v_2, v_3 και η μεταξύ τους απόσταση D23, όπως είχαν επιλεγεί στο προηγούμενο κεφάλαιο. Με τον παραπάνω τρόπο επιτυγχάνεται η βαθμονόμηση του μοντέλου (calibration) σύμφωνα με τη σειρά δεδομένων B1695.

Στη συνέχεια οι υπόλοιπες διαθέσιμες σειρές δεδομένων ταξινομούνται στις αντίστοιχες κλάσεις ανάλογα με ποια κλάση έχουν τα περισσότερα κοινά χαρακτηριστικά (classification). Συγκεκριμένα, τα δεδομένα από τις υπόλοιπες σειρές δεδομένων ταξινομούνται στις ήδη υφιστάμενες κλάσεις που δημιουργήθηκαν από τη σειρά δεδομένων B1695 και δεν έγινε εκ νέου ομαδοποίηση (clustering) σε αυτές. Για

την ταξινόμηση χρησιμοποιήθηκε η μέθοδος του εγγύτερου γείτονα η οποία περιγράφηκε στην ενότητα 3.2.4. και υλοποιείται με τη χρήση του αλγορίθμου *k*-nearest neighbors. Κατά την εφαρμογή του αλγορίθμου αυτού επιλέγεται μια παράμετρος *k* η οποία ορίζει την περιοχή μελέτης σε κάθε επανάληψη που γίνεται, δηλαδή σε κάθε επανάληψη θα ληφθούν υπόψη οι *k* εγγύτεροι γείτονες. Ως τιμή της παραμέτρου αυτής ορίστηκε η τιμή 5, δεδομένου ότι δοκιμάστηκαν διάφορες τιμές αλλά δεν έδειξαν να επηρεάζουν το αποτέλεσμα της ταξινόμησης σε σημαντικό βαθμό. Έπειτα, το κατάλληλο ευέλικτο μοντέλο επιλέγεται και ανακτάται για κάθε κλάση ξεχωριστά και προκύπτει η εκτίμηση των ταχυτήτων που αφορούν τα νέα δεδομένα (estimation). Τα αποτελέσματα παρουσιάζονται στην επόμενη ενότητα του κεφαλαίου.

6.2. Αποτελέσματα και αιτιολόγηση

Τα αποτελέσματα που προέκυψαν παρουσιάζονται στον πίνακα 6.1. Το αναμενόμενο αποτέλεσμα θα ήταν να μειωθεί το σφάλμα εκτίμησης με τη δεύτερη προσέγγιση, δηλαδή αναμενόταν όσο θα αυξάνεται ο αριθμός των κλάσεων να προκύπτει πιο ακριβής εκτίμηση, καθώς τα μοντέλα εφαρμόζονται σε πιο ομοιογενή δεδομένα. Από την άλλη πλευρά, όμως, σε αυτήν την περίπτωση χρησιμοποιούνται λιγότερα δεδομένα (ανά κατηγορία). Το αναμενόμενο αποτέλεσμα δεν επετεύχθη για όλες τις σειρές δεδομένων. Συγκεκριμένα, σε τρεις σειρές δεδομένων υπήρξε μείωση του σφάλματος για την περίπτωση των 4 κλάσεων, ενώ σε άλλες τρεις σειρές δεδομένων δεν υπήρξε καθόλου βελτίωση με τη δεύτερη προσέγγιση. Οι σειρές δεδομένων που εξετάζονται σε αυτήν την εργασία αποτελούνται από λίγες παρατηρήσεις και κατά συνέπεια οι τυχόν περιορισμοί που σχετίζονται με το πλήθος των δεδομένων μπορεί να μην επηρέαζαν το αποτέλεσμα αν υπήρχε πλήθος δεδομένων. Το θέμα αυτό διερευνάται στη συνέχεια εξετάζοντας κάποιες σειρές δεδομένων ξεχωριστά.

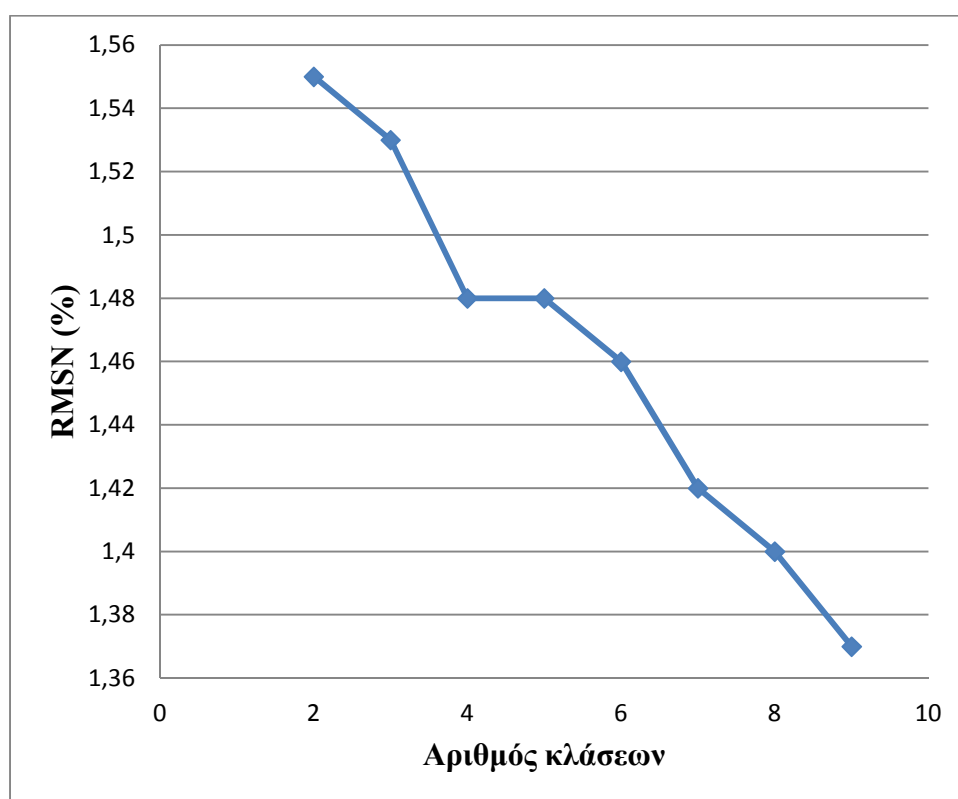
Σειρές δεδομένων	RMSN (%)									
	Μοντέλο του Gijpss	Μέθοδος loess	Προσέγγιση με ομαδοποίηση- ταξινόμηση (Αριθμός κλάσεων)							
			2	3	4	5	6	7	8	9
B1695	2.70	1.59	1.55	1.53	1.48	1.48	1.46	1.42	1.40	1.37
C621	6.60	4.34	4.37	4.41	3.99	4.59	4.54	4.10	3.99	5.36
A358	2.70	2.08	<u>2.10</u>	2.31	3.10	2.34	2.25	2.33	2.33	3.61
A172	4.60	3.40	3.48	3.06	2.44	3.14	2.85	3.30	3.13	9.39
C168	2.30	1.78	<u>1.87</u>	2.04	1.95	2.02	2.05	2.09	2.08	2.08
C171	7.20	6.23	<u>6.31</u>	6.60	7.35	6.50	6.47	8.22	8.19	8.73

Πίνακας 6.1.: Αποτελέσματα εκτίμησης της ταχύτητας v_3 για όλες τις σειρές δεδομένων χρησιμοποιώντας ένα συνδυασμό μεθόδων μηχανικής μάθησης

Όσον αφορά στη σειρά δεδομένων B1695, αποτελεί τη μοναδική σειρά δεδομένων που επιτυγχάνεται το επιθυμητό αποτέλεσμα. Όπως παρατηρείται στον παραπάνω πίνακα και στο διάγραμμα 6.3., όσο αυξάνεται ο αριθμός των κλάσεων τόσο το σφάλμα RMSN (%) μειώνεται, δηλαδή τόσο η εκτίμηση πλησιάζει την τιμή της πραγματικής ταχύτητας. Η σειρά B1695 περιλαμβάνει 1695 παρατηρήσεις και αποτελεί την σειρά με τα περισσότερα δεδομένα. Επίσης, η σειρά δεδομένων B1695 είναι αυτή που χρησιμοποιήθηκε για να δημιουργηθεί το πρότυπο μοντέλο και επομένως στο πρότυπο μοντέλο υπάρχουν τόσες κυκλοφοριακές καταστάσεις όσες στη σειρά B1695, καθώς επίσης υπάρχουν επαρκή δεδομένα ανά κλάση για να παραχθεί η κατάλληλη εκτίμηση της ταχύτητας. Επομένως, όλες οι καταστάσεις μιας σειράς δεδομένων αντιστοιχίζονται με κάποιες καταστάσεις του πρότυπου μοντέλου και υπάρχει μεγάλη ομοιότητα μεταξύ αυτών, καθώς και επάρκεια δεδομένων ανά κλάση.

Αυτό, όμως, δε συμβαίνει με τις υπόλοιπες σειρές δεδομένων, οι οποίες συμπεριλαμβάνουν λιγότερα δεδομένα και πιθανόν να περιλαμβάνουν καταστάσεις που δεν εμπεριέχονται στο πρότυπο μοντέλο, με αποτέλεσμα να μη γίνεται σωστή εκτίμηση. Συγκεκριμένα, από τον πίνακα 6.1. φαίνεται ότι το πλήθος των παρατηρήσεων που περιλαμβάνει η κάθε σειρά δεδομένων επηρεάζει την εφαρμογή της προτεινόμενης μεθοδολογίας. Για τη σειρά δεδομένων C621 που περιλαμβάνει 621 παρατηρήσεις, παρατηρείται ότι για την περίπτωση των 4, των 7 και των 9 κλάσεων επιτυγχάνεται μείωση του σφάλματος RMSN (%) σε σχέση με την εφαρμογή της πρώτης προσέγγισης, δηλαδή την εφαρμογή μόνο της μεθόδου loess. Επομένως η εφαρμογή της μεθοδολογίας στη σειρά αυτή φαίνεται να πλησιάζει το επιθυμητό αποτέλεσμα. Για τη σειρά A358 που περιλαμβάνει λιγότερες παρατηρήσεις, η μέθοδος

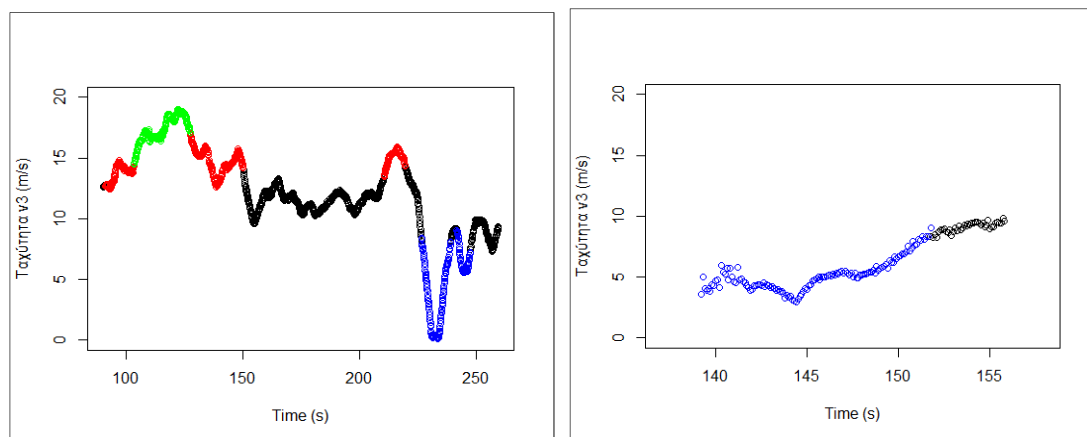
δεν αποδίδει τόσο καλά. Για την σειρά δεδομένων A358 δεν επιτυγχάνεται μείωση του σφάλματος αλλά το σφάλμα διατηρείται κυρίως στα ίδια επίπεδα με αυτό που προέκυψε από τη μέθοδο loess. Τέλος, όταν το πλήθος των παρατηρήσεων μειώνεται ακόμα περισσότερο, όπως συμβαίνει για τις υπόλοιπες σειρές δεδομένων, η απόδοση της μεθόδου συνήθως δεν είναι καλή. Τέλος, αξίζει να σημειωθεί ότι σε αρκετές περιπτώσεις, παρά τα ελάχιστα δεδομένα, η μεθοδολογία παρέχει εκτίμηση των ταχυτήτων περίπου ίδια με αυτή που προκύπτει από τη μέθοδο loess και καλύτερη από αυτή του μοντέλου του Gipps. Για παράδειγμα αυτό παρατηρείται για τη σειρά δεδομένων C168 μεταξύ 2 και 6 κλάσεων, καθώς και για τις σειρές C171 και A358 για την ομαδοποίηση με 2 κλάσεις.



Διάγραμμα 6.3.: RMSN (%) ανάλογα με τον αριθμό κλάσεων για τη σειρά δεδομένων B1695

Στη συνέχεια εξετάζεται πιο αναλυτικά μια σειρά δεδομένων για την οποία δεν υπήρξε καθόλου βελτίωση με τη δεύτερη προσέγγιση για κανέναν από τους εξεταζόμενους αριθμούς κλάσεων, όπως η C171. Στο διάγραμμα 6.4. απεικονίζονται οι ταχύτητες v_3 που εκτιμώνται για τη κάθε σειρά δεδομένων ανάλογα με την κλάση

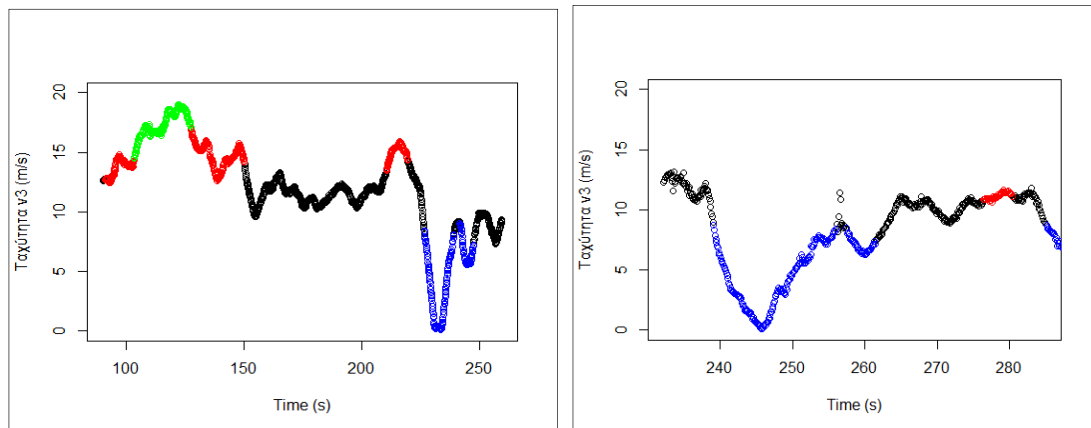
στην οποία ανήκουν. Παρατηρώντας την εξέλιξη της ταχύτητας και στις δυο περιπτώσεις είναι φανερό ότι δεν υπάρχει μεγάλη ομοιότητα. Στη σειρά δεδομένων B1695 υφίστανται 4 κλάσεις δεδομένων, ενώ στη σειρά C171 υπάρχουν μόνο δύο. Αυτές οι δύο κλάσεις δεν παρουσιάζουν μεγάλη ομοιότητα όσον αφορά τον τρόπο που εξελίσσεται η ταχύτητα με τις αντίστοιχες κλάσεις της σειράς B1695. Ωστόσο, τα δεδομένα καταχωρούνται στην κλάση με την οποία έχουν τις περισσότερες ομοιότητες, ακόμη και αν αυτό σημαίνει ότι οι ομοιότητες δεν είναι πολλές. Επομένως, στην περίπτωση αυτή περιορίζονται από την κλάση τους και δεν βελτιώνεται η εκτίμηση τους. Στο σημείο αυτό γίνεται κατανοητό ότι για τη σωστή απόδοση της δεύτερης προσέγγισης απαιτείται να υπάρχει αντιστοιχία των καταστάσεων ή συνθηκών των νέων δεδομένων με αυτών που δημιουργήθηκε το αρχικό πρότυπο ακολουθίας οχημάτων. Για το σκοπό αυτό θα έπρεπε να έχει χρησιμοποιηθεί μια εκτεταμένη σειρά δεδομένων που θα καλύπτει όσο το δυνατόν περισσότερες περιπτώσεις της κυκλοφορίας.



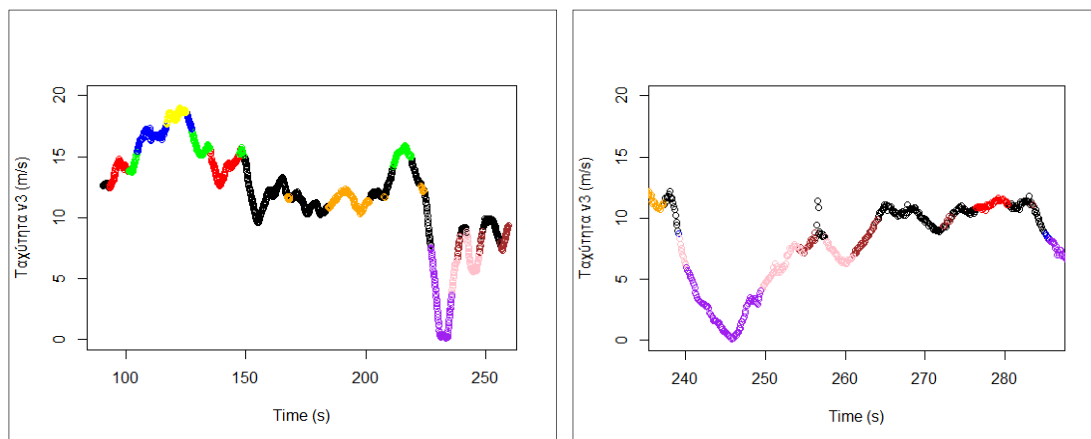
Διάγραμμα 6.4.: Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695 (αριστερό μέρος), C621 (δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 4

Στη συνέχεια εξετάζεται η σειρά C621 για την περίπτωση με τις 4 κλάσεις που δίνει καλύτερο αποτέλεσμα από τη μέθοδο loess και την περίπτωση με τις 9 κλάσεις που δίνει το χειρότερο αποτέλεσμα. Με αυτόν τον τρόπο, θα γίνει κατανοητό πότε η προσέγγιση με το συνδυασμό της μεθόδου loess και των τεχνικών της ομαδοποίησης και της ταξινόμησης παρέχουν καλύτερο αποτέλεσμα. Από τα διαγράμματα 6.5. και 6.6. μπορεί να εξηγηθεί γιατί συμβαίνει αυτό. Στο διάγραμμα 6.5. που αναφέρεται στις 4 κλάσεις, τα δεδομένα φαίνονται καλύτερα κατανοημένα, ενώ στην περίπτωση των

9 κλάσεων στο διάγραμμα 6.6. τα δεδομένα χωρίζονται σε μικρότερες κλάσεις με πιο ιδιαίτερα χαρακτηριστικά. Έτσι, καλύτερη εκτίμηση θα υπάρξει μόνο όταν στα νέα δεδομένα συναντώνται επίσης παρατηρήσεις με αυτά τα χαρακτηριστικά. Σε διαφορετική περίπτωση τα νέα δεδομένα καταχωρούνται αναγκαστικά σε κάποια κλάση που δεν βοηθά την εκτίμηση της επιθυμητής ταχύτητας.



Διάγραμμα 6.5. : Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695 (αριστερό μέρος), C621 (δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 4



Διάγραμμα 6.6. : Οι προς εκτίμηση ταχύτητες v_3 για τις σειρές δεδομένων B1695 (αριστερό μέρος), C621 (δεξιό μέρος) ανάλογα με την κλάση που ανήκουν 1 έως 9

Επιπλέον, στο διάγραμμα 6.5. συναντάται ένα ακόμα πρόβλημα. Για την κλάση που απεικονίζεται με κόκκινο χρώμα, στην περίπτωση της σειράς δεδομένων B1695 οι ταχύτητες κυμαίνονται περίπου από 12 έως 16 m/s, ενώ στη σειρά C621 για την ίδια κλάση οι ταχύτητες που πρέπει να εκτιμηθούν είναι μικρότερες από 12 m/s. Για το

λόγο αυτό όταν εφαρμόστηκε η μέθοδος loess δεν μπόρεσε να εκτιμήσει τις ταχύτητες αυτές, καθώς δεν υπήρχαν τέτοιες τιμές στο πρότυπο μοντέλο.

Το πρόβλημα αυτό εντοπίστηκε σε μεγάλο βαθμό στα διαθέσιμα δεδομένα NGSIM. Για παράδειγμα, στη σειρά B1695 δεν υπάρχουν πολλές χαμηλές τιμές ταχύτητας. Στα NGSIM δεδομένα καταστάσεις συνεχούς εναλλαγής επιτάχυνσης και επιβράδυνσης συναντώνται σε χαμηλές ταχύτητες. Στην περίπτωση αυτή τέτοιες καταστάσεις είναι πιθανό να ενταχθούν σε κάποια κλάση που συμβαίνει κάτι παρόμοιο για τη σειρά B1695 αλλά σε υψηλότερες ταχύτητες. Με αποτέλεσμα όταν εφαρμόζεται η μέθοδος loess, δεν μπορεί να παράγει εκτίμηση για τιμές της ταχύτητας που δεν υπάρχουν στο πρότυπο μοντέλο. Ως εκ τούτου δεν μπορούσαν να προκύψουν αξιόλογα αποτελέσματα από την εφαρμογή της δεύτερης προσέγγισης στα NGSIM δεδομένα λόγω διαφορετικής φύσης των δεδομένων και κατά συνέπεια δεν κρίθηκε σημαντικό να παρατεθούν στο σημείο αυτό. Πιο συγκεκριμένα, τα δεδομένα NGSIM καταγράφηκαν σε ώρες αιχμής υπό συνθήκες κυκλοφοριακής συμφόρησης. Τα πρότυπα ακολουθίας οχημάτων που προκύπτουν για κάθε κλάση των δεδομένων της σειράς B1695 είναι αρκετά πιο εξειδικευμένα από αυτό που προκύπτει μόνο από την εφαρμογή της μεθόδου loess σε όλα τα δεδομένα μαζί. Έτσι, ένα εξειδικευμένο πρότυπο αντιστοιχεί σε εξειδικευμένες καταστάσεις κυκλοφορίας. Σε καταστάσεις κυκλοφοριακής συμφόρησης τα πρότυπα ακολουθίας οχημάτων αναμένονται να είναι πολύ διαφορετικά. Για παράδειγμα, οι αποστάσεις μεταξύ των οχημάτων αναμένονται να είναι πολύ μικρότερες σε σχέση με αυτές που υπάρχουν όταν τα οχήματα κινούνται σε κανονική ροή. Επιπλέον, οι ταχύτητες που παρατηρούνται σε συνθήκες κυκλοφοριακής συμφόρησης είναι πολύ χαμηλότερες. Στη σειρά δεδομένων B1695 δεν συναντώνται συνθήκες κυκλοφοριακής συμφόρησης και κατά συνέπεια δεν προκύπτει από την ομαδοποίηση των δεδομένων της κάποιο αντίστοιχο εξειδικευμένο πρότυπο. Επομένως, αρκετά από τα NGSIM δεδομένα καταχωρούνται αναγκαστικά σε κάποια κλάση που προέκυψε από τη σειρά δεδομένων B1695 χωρίς όμως να παρουσιάζουν σημαντικές ομοιότητες με αυτήν. Έτσι το πρότυπο ακολουθίας οχημάτων που επιλέγεται να εκτιμήσει τις ταχύτητες δεν θεωρείται κατάλληλο.

6.3. Συμπεράσματα

Η προτεινόμενη μεθοδολογία που συνιστάται από ένα συνδυασμό μεθόδων μηχανικής μάθησης και συγκεκριμένα της μεθόδου loess και των τεχνικών ομαδοποίησης και ταξινόμησης (δεύτερη προσέγγιση) δεν απορρίπτεται αλλά ούτε επιβεβαιώνεται στο πλαίσιο της εργασίας αυτής. Το γεγονός ότι σε κάποιες σειρές δεδομένων επιφέρει καλύτερο αποτέλεσμα από την απλή μέθοδο loess (πρώτη προσέγγιση) ενισχύει την πιθανότητα βελτίωσης της εκτίμησης προτύπων ακολουθίας οχημάτων και επιβάλλει την περαιτέρω έρευνα για την προσέγγιση αυτή σε δεδομένα μεγαλύτερης έκτασης. Επιπλέον, το επιθυμητό αποτέλεσμα, δηλαδή ότι με την αύξηση του αριθμού των κλάσεων επιτυγχάνεται καλύτερη εκτίμηση επιβεβαιώνεται από τη σειρά δεδομένων B1695, με την οποία διαμορφώθηκε το πρότυπο. Επίσης, το πλήθος των παρατηρήσεων φαίνεται να επηρεάζει την εφαρμογή της μεθοδολογίας, καθώς όσο αυτό αυξάνεται η απόδοση της μεθόδου βελτιώνεται.

Αν θεωρητικά υπάρχει μια μεγάλη βάση ομαδοποιημένων δεδομένων που να περιλαμβάνει διάφορες κυκλοφοριακές καταστάσεις για διάφορες τιμές ταχύτητας είναι πιθανό κάθε νέα σειρά μετρήσεων που θα ταξινομείται σε αυτή τη βάση δεδομένων να βρίσκεται σε σχεδόν πλήρη αντιστοιχία με τις υπάρχουσες κλάσεις και επομένως να προκύπτουν εκτιμήσεις μεγαλύτερης ακρίβειας από αυτές που παρέχονται από τα παραδοσιακά μοντέλα που βασίζονται σε μαθηματικούς τύπους. Αξιοσημείωτο είναι ότι η μέθοδοι μηχανικής μάθησης παρουσιάζουν μεγάλη ευελιξία και μπορούν να προσαρμοστούν και να βελτιωθούν στην εκτίμηση προτύπων ακολουθίας οχημάτων, το οποίο όμως απαιτεί να ερευνηθεί περισσότερο προτού οι μέθοδοι αυτές ενταχθούν στην εκτίμηση και κατά προέκταση στην πρόβλεψη προτύπων για την προσομοίωση κυκλοφοριακών συστημάτων.

7. ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Στο παρόν τελευταίο κεφάλαιο της μελέτης παρουσιάζονται τα γενικά συμπεράσματα από την ανάλυση που προηγήθηκε όσον αφορά στη σύγκριση του μοντέλου του Girps και των μεθόδων μηχανικής μάθησης ως προς την εκτίμηση προτύπων ακολουθίας οχημάτων. Επίσης προτείνονται θέματα για μελλοντική έρευνα και ανάλυση.

7.1. Γενικά Συμπεράσματα

Στο πλαίσιο της εργασίας αυτής εξετάστηκαν οι μέθοδοι μηχανικής μάθησης ως εναλλακτικές μέθοδοι για την εκτίμηση προτύπων ακολουθίας οχημάτων. Η μέθοδος της τοπικά σταθμισμένης παλινδρόμησης (loess) εφαρμόστηκε στα διαθέσιμα δεδομένα και παρείχε για την πλειοψηφία των δεδομένων καλύτερα αποτελέσματα σε σχέση με το μοντέλο του Girps. Οι ταχύτητες που εκτιμήθηκαν με τη μέθοδο loess πλησίαζαν περισσότερο τις πραγματικές ταχύτητες που παρατηρήθηκαν. Έπειτα, η μέθοδος της σταθμισμένης παλινδρόμησης δοκιμάστηκε να συνδυαστεί με τις τεχνικές της ομαδοποίησης και της ταξινόμησης για την απόδοση καλύτερης εκτίμησης. Η μεθοδολογία αυτή φαίνεται να επιτυγχάνει μόνο όταν υπάρχει επάρκεια δεδομένων και όταν υπάρχουν αρκετές ομοιότητες μεταξύ των αντίστοιχων κλάσεων του πρότυπου μοντέλου και των νέων δεδομένων.

Οι μέθοδοι μηχανικής μάθησης παρουσιάζουν μεγάλη ευελιξία και ταχύτητα ως προς τη διαχείριση των δεδομένων, καθώς δεν απαιτείται ειδική βαθμονόμηση αρκετών παραμέτρων όπως συνήθως χρειάζεται στα παραδοσιακά μοντέλα. Επίσης, παρέχουν τη δυνατότητα εξέτασης νέων παραμέτρων που επιδρούν στη συμπεριφορά των οδηγών και δεν περιορίζονται σε αυτές που ορίζουν οι μαθηματικοί τύποι των παραδοσιακών μοντέλων. Το γεγονός ότι η εκτίμηση που παράγεται με τη μέθοδο loess είναι πιο ακριβής την καθιστά προτιμότερη από το μοντέλο του Girps ενισχύοντας έτσι το συμπέρασμα ότι οι μέθοδοι μηχανικής μάθησης μπορούν να προσφέρουν αξιόπιστα αποτελέσματα. Όμως, το μοντέλο του Girps έχει το προτέρημα ότι τεκμηριώνει τα αποτελέσματα που εξάγει βασιζόμενο στη θεωρία της ακολουθίας οχημάτων. Αντιθέτως, η μέθοδοι μηχανικής μάθησης, παρά την ευελιξία τους, στερούνται θεωρητικής τεκμηρίωσης και ο τρόπος με τον οποίο εκτιμώνται τα πρότυπα ακολουθίας οχημάτων δεν γίνεται κατανοητός από τον ανθρώπινο νου. Η

διαδικασία εκτίμησης των προτύπων αποτελεί ένα «μαύρο κουτί» και είναι φυσικό επακόλουθο οι επιστήμονες να φοβούνται να εμπιστευτούν μεθόδους που δεν μπορούν να εξηγήσουν. Αυτό γίνεται ακόμα πιο δύσκολο όταν τίγονται θέματα ασφάλειας. Η ένταξη των μεθόδων μηχανικής μάθησης στη διαδικασία της προσομοίωσης απαιτεί αρκετές μελέτες ακόμα που θα επιβεβαιώνουν την εγκυρότητα τους. Η ανάγκη για όσο το δυνατόν μεγαλύτερη ακρίβεια και αξιοπιστία, όμως, κρίνει απαραίτητη τη μελέτη αυτή.

Θεωρητικά αν υπάρχει μια βάση δεδομένων με στοιχεία που καλύπτουν διάφορες κυκλοφοριακές καταστάσεις τότε θα μπορούν να εκτιμηθούν αξιόπιστα πρότυπα με τις μεθόδους μηχανικής μάθησης. Στη συλλογή δεδομένων θα συμβάλει σημαντικά η εξέλιξη της τεχνολογίας. Ωστόσο, η ύπαρξη μιας βάσης δεδομένων που θα περιλαμβάνει όλα τα πιθανά σενάρια θεωρείται ουτοπική δεδομένου ότι οι συμπεριφορές των οδηγών εξαρτώνται από πολλούς παράγοντες και οι κυκλοφοριακές συνθήκες μπορούν να είναι πολλές και διαφορετικές. Κατά συνέπεια, το βασικό ερώτημα που προκύπτει από αυτή την εργασία είναι πόσα δεδομένα πρέπει τελικά να συλλεχθούν για τη μελέτη της συμπεριφοράς των οδηγών σε ποικίλες κυκλοφοριακές συνθήκες, έτσι ώστε να θεωρούνται αντιπροσωπευτικά για τον πραγματικό κόσμο. Ο πραγματικός κόσμος εξελίσσεται, οι συμπεριφορές των ανθρώπων αλλάζουν, νέες βελτιωμένες οδικές υποδομές διατίθενται και νέα οχήματα με καλύτερες δυνατότητες κυκλοφορούν συνεχώς. Ως εκ τούτου, κρίνεται αναγκαία η συνεχής συλλογή και μελέτη δεδομένων με σκοπό την εξέλιξη της μοντελοποίησης οχημάτων και την καλύτερη δυνατή εκτίμηση προτύπων της συμπεριφοράς των οδηγών σε ποικίλες κυκλοφοριακές καταστάσεις.

7.2. Προτάσεις για μελλοντική έρευνα

Στην παρούσα εργασία μελετήθηκαν οι αλληλεπιδράσεις μεταξύ οχημάτων που βρίσκονται σε ακολουθία. Ωστόσο, υπάρχουν ακόμα πολλές άλλες προοπτικές του θέματος που δεν έχουν εξεταστεί. Η εξέταση των αλληλεπιδράσεων μεταξύ των οχημάτων και των οδικών υποδομών αποτελεί ένα θέμα που θα είχε ιδιαίτερο ενδιαφέρον. Οι μέθοδοι μηχανικής μάθησης δεν ορίζουν ποιες θα πρέπει να είναι οι εξεταζόμενες παράμετροι, όπως συμβαίνει με τα παραδοσιακά μοντέλα τα οποία ακολουθούν συγκεκριμένο μαθηματικό τύπο και περιορίζουν τις εξεταζόμενες παραμέτρους. Με αυτόν τον τρόπο οι μέθοδοι μηχανικής μάθησης δίνουν τη

δυνατότητα να εξεταστούν νέες παράμετροι που είναι πιθανό να επηρεάζουν τη συμπεριφορά των οδηγών. Επομένως, ως παράμετροι μπορούν να ληφθούν, πέρα από τις ταχύτητες των οχημάτων και τις μεταξύ τους αποστάσεις, κάποια χαρακτηριστικά της οδού, όπως η καμπυλότητα ή η ύπαρξη διασταυρώσεων με στοπ ή φωτεινό σηματοδότη. Επίσης, άλλοι παράμετροι που είναι πιθανό να επηρεάζουν τη συμπεριφορά των οδηγών είναι αναμφισβήτητα τα χαρακτηριστικά των ίδιων των οδηγών, όπως η ηλικία, ο χρόνος αντίδρασης, η εμπειρία, κλπ. Θα μπορούσαν για παράδειγμα να διαμορφωθούν διαφορετικά πρότυπα ακολουθίας οχημάτων ανάλογα με την ηλικία του οδηγού.

Η εκτίμηση του χρόνου αντίδρασης των οδηγών αποτελεί ένα θέμα που έχει εξεταστεί αρκετά. Ωστόσο, δεν υπάρχουν ακόμα κάποια σαφή λογικά όρια από τη βιβλιογραφία και απαιτείται επιπλέον έρευνα δεδομένου ότι ο χρόνος αντίδρασης διαδραματίζει καθοριστικό ρόλο στη διαμόρφωση προτύπων ακολουθίας οχημάτων και γενικότερα στην προσομοίωση των κυκλοφοριακών συστημάτων. Οι ακραίες μικρές τιμές προτείνεται να αποφεύγονται, έτσι ώστε να αφήνεται στον οδηγό ένα μικρό περιθώριο καθυστέρησης ή λάθους.

Όσον αφορά στη μεθοδολογία που εμπεριέχει ένα συνδυασμό μεθόδων μηχανικής μάθησης και εξετάστηκε σύντομα στο έκτο κεφάλαιο της παρούσας εργασίας, κρίνεται σκόπιμη η διεξοδική ανάλυση του. Για τον σκοπό αυτό προτείνεται η χρήση πιο εκτεταμένων σειρών δεδομένων ειδικά για τη διαμόρφωση ενός πρότυπου ακολουθίας οχημάτων. Επιπλέον, κρίνεται σκόπιμο να εξεταστούν οι παράμετροι που επηρεάζουν περισσότερο τη συμπεριφορά των οδηγών στην κάθε κλάση διαφορετικά. Για παράδειγμα αν τα δεδομένα είναι χωρισμένα σε δυο κλάσεις, μια που αναφέρεται κυρίως σε συνθήκες επιτάχυνσης και μία άλλη που αναφέρεται σε συνθήκες επιβράδυνσης. Είναι πιθανό, σε κάθε κατάσταση (κάθε κλάση) η συμπεριφορά του οδηγού να επηρεάζεται περισσότερο από διαφορετικές παραμέτρους. Αν αυτό ισχύει τότε σε κάθε κλάση η μέθοδος loess θα έπρεπε να εφαρμοστεί λαμβάνοντας υπόψη τους αντίστοιχους παράγοντες. Οι διαφορετικές κυκλοφοριακές καταστάσεις, επίσης, μπορούν να γίνουν πολύ συγκεκριμένες, όπως για παράδειγμα να εξεταστεί τι συμβαίνει σε καταστάσεις επείγουσας επιβράδυνσης προς αποφυγήν ατυχήματος.

Η εκτίμηση των προτύπων ακολουθίας οχημάτων αποτελεί ένα πολύ σημαντικό κομμάτι στη διαδικασία πρόβλεψης ταχυτήτων και κατά συνέπεια στην προσομοίωση ενός κυκλοφοριακού συστήματος. Επομένως, απαιτείται μεγάλη ακρίβεια και

αξιοπιστία στην εκτίμηση των προτύπων ακολουθίας οχημάτων και ιδιαίτερα όταν τίγονται θέματα ασφάλειας. Στην εργασία αυτή παρουσιάστηκε σαν εναλλακτική εκδοχή η χρήση των μεθόδων μηχανικής μάθησης. Αναμφισβήτητα, τα συμπεράσματα που προέκυψαν πρέπει να ελεγχθούν και να εφαρμοστούν σε πολλά δεδομένα ακόμη, καθώς και να ακολουθήσουν άλλες σχετικές μελέτες, προτού οι μέθοδοι μηχανικής μάθησης ενταχθούν επίσημα στον τομέα της προσομοίωσης κυκλοφοριακών συστημάτων.

Στην ελληνική γλώσσα

Βλαχάβας Ι., Κεφαλάς Π., Βασιλειάδης Ν., Κόκορας Φ., Σακελλαρίου (2005). Η “Τεχνητή Νοημοσύνη”, Β’ Έκδοση, Γαρταγάνη, Θεσσαλονίκη, 2005.

Παραδείσης Δ. (2000). Σημειώσεις Δορυφορικής Γεωδαισίας, Κέντρο Δορυφόρων Διονύσου, Τομέας Τοπογραφίας, Ε.Μ.Π., Αθήνα.

Αγάτζα- Μπαλοδήμου Α.Μ. (2008). Σημειώσεις «Φίλτρο Kalman», ΔΠΜΣ Γεωπληροφορική

Ξενόγλωσσες

Ahmed K. I. (1999). Modeling Drivers’ Acceleration and Lane Changing Behavior, PhD thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1999.

Al-Shihabi T. and Mourant R.R. (2003). Toward More Realistic Driving Behavior Models for Autonomous Vehicles in Driving Simulators, 82nd Annual Meeting of the Transportation Research Board, Washington. D. C.

Anderson B.D.O., Moore J.B. (1979). Optimal Filtering, Prentice-Hall, Inc.

Antoniou C. and H. N. Koutsopoulos (2006). Estimation of Traffic Dynamics Models with Machine Learning Methods. Transportation Research Record: Journal of the Transportation Research Board, No. 1965, pp. 103-111, Washington D.C., 2006.

Azimi M. and Y. Zhang (2010). Categorizing Freeway Flow Conditions Using Clustering Methods, Proceedings of the 89th Annual Meeting of the Transportation Research Board, Washington, D.C.

Barcelo J. (2002). Dynamic Network Simulation with AIMSUN. Proc., International Symposium on Transport Simulation, Yokohama, Aug. Kluwer Academic Publishers.

Barcelo J. (2010). Fundamentals of Traffic Simulation, International Series in Operations Research & Management Science, Vol. 145 1st Edition, 2010.

Bellemans T. De Schutter B. and De Moor B. (2002). "Models for traffic control", Journal A., Vol 43, no 3-4, pp. 13-22, 2002.

Ben-Akiva M., Bierlaire M., Koutsopoulos H. N. and Mishalani R. (2002). Real-time simulation of traffic demand-supply interactions within DynaMIT, in M. Gendreau and P. Marcotte (eds). Transportation and network analysis: current trends. Miscellanea in honor of Michael Florian, pp. 19-36, Kluwer Academic Publishers, Boston/Dordrecht/London.

Ben-Akiva M., H. N. Koutsopoulos, C. Antoniou and R. Balakrishna (2010). "Traffic Simulation with DynaMIT". In J. Barcelo (ed.) "Fundamentals of traffic simulation". (ISBN: 978-1-4419-6141-9), Springer.

Bevrani K., Chung E. (2011). Car following model improvement for traffic safety metrics reproduction. In Proceedings of the Australasian Transport Research Forum 2011, PATREC, Adelaide Hilton Hotel, Adelaide, SA, pp. 1-14.

Bilban M., Alenka Vojvoda and Janez Jerman (2009). Age Affects Drivers' Response Times, Coll. Antropol. 33 (2009) 2: 467–471.

Boxill S.A. and L. Yu (2000). An Evaluation of Traffic Simulation Models for Supporting ITS Development. Center for Transportation Training and Research Texas Southern University.

Brackstone M. and McDonald M. (1999). "Car-following: a historical review." Transportation Research F 2: 181–196.

Brackstone M., Sultan B., McDonald M. (2002). Motorway driver behaviour: studies on car-following. Transportation Research Part F 5. 329-344.

Brunson S. J., E. M. Kyle, N.C. Phamdo and G. R. Preziotti (2002). Alert algorithm development program: NHTSA rear-end collision alert algorithm. DOT HS 809-526, Final Report, The John Hopkins University in cooperation with the NHTSA and General Motors Corporation.

Burghout W., H. N. Koutsopoulos, I. Andréasson (2005). Hybrid Mesoscopic-Microscopic Traffic Simulation, *Transportation Research Record*, Vol. 1934, No. -1. (2005), pp. 218-255.

Cleveland W. S. (1978). Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*, Vol. 74, 1978, pp. 829–836.

Cleveland W. S., and S. J. Devlin (1988a). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association*, Vol. 83, 1988, pp. 596–610.

Cleveland W. S., S. J. Devlin, and E. Grosse (1988b). Regression by Local Fitting: Methods. Properties and Computational Algorithms. *Journal of Econometrics*, Vol. 37, 1988, pp. 87–114.

Ciuffo B., Punzo V., Montanino M. (2012). 30 years of the Gipps' car-following model: applications, developments and new features. *TRB 2012 Annual Meeting*, Paper number: 12-3350.

El Faouzi N.-E. (2004). Data-driven aggregative schemes for multisource estimation Fusion: a road travel time application. In *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2004*. (Belur V. Dasarathy. ed.), *Proceedings of SPIE Vol. 5434*, pp. 351-359 (SPIE. Bellingham. WA. 2004).

El Faouzi N.-E. and E. Lefevre (2006). Classifiers and Distance-Based Evidential Fusion for Road Travel Time Estimation. In *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006*. (Belur V. Dasarathy. ed.), *Proceedings of SPIE Vol. 6242*.

Fambro D. B., R. J. Koppa, D. L. Picha, and K. Fitzpatrick (1988). Driver Perception-Brake Response in Stopping Sight Distance Situations. In *Transportation Research Record 1628*, TRB, National Research Council, Washington. D.C., 1998, pp. 1–7.

Fraley C. and A. E. Raftery (2002). Model-Based Clustering. Discriminant Analysis. and Density Estimation. *Journal of the American Statistical Association*, Vol. 97, No. 458, pp. 611-631.

Fritzsche H-T. (1994). "A model for traffic simulation." *Transportation Engineering Contribution 5*: 317–321.

Gipps P.G. (1981). *A Behavioral Car Following Model for Computer Simulation*. *Transportation Research B* 15, pp 105-111.

Green M. (2000). "How Long Does It Take to Stop?" *Methodological Analysis of Driver Perception-Brake Times.* *Transportation Human Factors*, vol. 2, no. 3, pp. 195–216, 2000.

Hamdar S. H. & Mahmassani H. S. (2008). *Driver car-following behavior: From discrete event process to continuous set of episodes*. *Transportation Research Board Annual Meeting*.

Hartigan J. A. (1975). *Clustering Algorithms*, John Wiley and Sons, New York, 1975.

Hartigan J. A., and M. A. Wong (1979). *A K-Means Clustering Algorithm*. *Applied Statistics*, Vol. 28, 1979, pp. 100–108.

Hidas P. (1998). *A car-following model for urban traffic simulation*. *Traffic Engineering and Control*, 39, pp. 300-305.

Hoogendoorn S., Bovy P.(2001). *State-of-the-art of vehicular traffic flow modelling*. *Proceedings of the Institution of Mechanical Engineer, Part I: Journal of Systems and Control Engineering* 215(4), 283–303.

Hoogendoorn S. P., Van Zuylen H. J., Schreuder M., Gorte B., and Vosselman G. (2003). *Microscopic Traffic Data Collection by Remote Sensing*. *Proceedings of the 82nd annual meeting TRB*, Washington, D.C.

Huang E., Antoniou C., Wen Y., Ben-Akiva M., Lopes J., Bento J. (2009). "Real-time multi-sensor multi-source network data fusion using dynamic traffic assignment models." *Intelligent Transportation Systems, 2009, ITSC '09, 12th International IEEE Conference on*, 2009, 1-6, © 2009 IEEE.

Institute of Transportation Engineers California Border Section Highway Capacity Task Force (2004). A report on the use of traffic simulation models in the san diego region.

- Irizarry A.R. and Hector Corrada Bravo (2010). Lecture 7: Smoothing March, 2010.
- Johansson G., and K. Rummel (1971). Drivers' Brake Reaction Time. *Human Factors*, Vol. 13, No. 1, 1971, p. 23.
- Koutsopoulos H.N., Haneen F. (2012). Latent class model for car following behavior. *Transportation Research Part B* 46 (2012) 563–578.
- Leutzbach W. (1988). *Introduction to the theory of traffic flow*. Berlin, Springer Verlag.
- Liu R. (2010). *Traffic simulation with DRACULA*. *Fundamentals of Traffic Simulation* (Ed. J Barcelo), Springer.
- Liu R. (2010). *A Motorway Simulation Model*. Leaflet, pp. LF2061, SISTM. Transport Research Laboratory.
- MacQueen J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proc., 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (L. M. Le Cam and J. Neuman, eds.), University of California Press. Berkeley, 1967, pp. 281–297.
- Magister T., R. Krulec, M. Batista, and L. Bogdanoviæ (2005). The Driver Reaction Time Measurement Experiences. *Proceedings of Innovative Automotive Technology (IAT'05) Conference*, Bled, Slovenia, 2005.
- Mitchell T. (1997). *Machine Learning*, McGraw Hill.
- NGSIM (2004). *Task E.3: High-Level Data Plan. Final Report*, prepared for Federal Highway Administration, by Cambridge Symatics, Inc., Massachusetts Institute of Technology.
- NGSIM (2006). Halkias J., Colyar J., *Interstate 80 Freeway Dataset. Fact Sheet*. U.S. Department of Transportation, Federal Highway Administration, December, 2006.
- Ozaki H. (1993). Reaction and Anticipation in the Car-Following Behavior *Transportation and Traffic Theory. Proc., 12th International Symposium on the*

Theory Traffic Flow and Transportation (C. F. Daganzo. ed.), Berkeley, Calif., 1993, pp. 349–362.

Olstam J., J. Tapani A. (2004). Comparison of Car-following models. Swedish National Road and Transport Research Institute, VTI meddelande 960A · 2004.

Ortuzar J., Willumsen L.G. (1995). Modelling Transport, Second Edition, 1995.

Pindyck R.S., and D.L. Rubinfeld (1997). Econometric Models and Economic Forecasts, 4th ed. Irwin McGraw-Hill. Boston, Mass., 1977.

Punzo V., Formisano D.J., Torrieri V. (2005). A non-stationary Kalman Filter for the estimation of accurate multiple Car-Following Data, proceedings of the 84th annual meeting TRB, Washington, D. C.

Punzo V., Simonelli (2005). Analysis and Comparison of Car-following Models Using Real Traffic Microscopic Data, proceedings of the 84th annual meeting TRB, Washington, D. C.

Punzo V., Borzacchiello M. T. & Ciuffo B. (2011). On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. Transportation Research Part C: Emerging Technologies. In Press, Corrected Proof.

Punzo V., Ciuffo B., and Montanino. M (2012). May we trust results of car following models calibration based on trajectory data? In Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, D.C.

Rakha H., Wang W. (2009). Procedure for calibrating Gipps car-following model. Transportation Research Record, 2124, pp. 113-124.

Rakha H., Gao Y.(2010). Calibration of steady-state car-following models using macroscopic loop detector data. Final report VT-2008-01, prepared for Virginia Transportation Research Council, Virginia Tech Transportation Institute, May, 2010.

Ranjitkar P., T. Nakatsuji, Y. Azuta, and G. S. Gurusinghe (2003). Stability Analysis Based on Instantaneous Driving Behavior Using Car-Following Data. In Transportation Research Record: Journal of the Transportation Research Board, No.

1852, Transportation Research Board of the National Academies, Washington. D.C., 2003, pp. 140–151.

Ranjitkar P., Suzuki H. and Nakatsuji T. (2005a). Microscopic traffic data with real-time kinematic global positioning system. Proceedings of Annual Meeting of Infrastructure Planning and Management, Japan Society of Civil Engineer, Miyazaki, Preprint CD, December, 2005.

Ranjitkar P., Kawamura A., H. and Nakatsuji T. (2005b). Car-following models: An experiment based benchmarking. Journal of the Eastern Asia Society for Transportation Studies, Vol.6, pp.1582-1596, 2005.

Roughan M., Sen S., Spatscheck O., Duffield N.(2004). Class-of-service mapping for QoS: a statistical signature-based approach to IP traffic classification. Proceeding IMC '04 Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. Pages 135 – 148.

Sargent G.R. (1999). Validation and verification of simulation models. Proceedings of the 1999 Winter Simulation Conference P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

SAS Institute Inc. (1999). User's Guide. Version 8, Cary. NC: SAS Institute Inc.

SAS Institute Inc. (2009). SAS/STAT. 9.2. User's Guide, Second Edition, Cary. NC: SAS Institute Inc.

Silcock. J.P. (1993). SIGSIM version 1.0 users guide. Working Paper, Centre for Transport Studies, University of London.

Spyropoulou I., Karlaftis M. (2008). Parameters related to modeling intelligent speed adaptation systems with the employment of a microscopic traffic model. Proceedings of European Conference on Human centred design for intelligent transport systems. Lyon, France, pp. 163-171.

Summala H. (2000). Brake reaction times and driver behavior analysis. Transportation Human Factors, 2(3), pp. 217-226.

Sun H., H. X. Liu, H. Xiao, R. R. He, and B. Ran (2003). Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1836, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 143–150.

Sun H., C. Zhang, B. Ran. and K. Choi (2004). Prediction Intervals for Traffic Time Series. Presented at 83rd Annual Meeting of the Transportation Research Board. Washington, D.C., 2004.

Schwarz G., 1978. "Estimating the dimension of a model". *Annals of Statistics* 6(2): 461-464.

Taieb-Maimon M. and D. Shinar (2001). Minimum and comfortable driving headways: reality versus perception. *Human Factors*, Vol. 43, No. 1, pp. 159-172.

Theil H. (1978). *Introduction to Econometrics*, Prentice Hall, New Jersey.

Thiemann C., Treiber M. & Kesting A. (2008). Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data. *Transportation Research Record*, 90-101.

Toledo T. (2003). *Integrated Driving Behaviour Modelling*. PhD thesis, Massachusetts Institute of Technology.

Treiterer J. (1975). *Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques*. Transportation Research Center, Department of Civil Engineering, Ohio State University, Final Report EES 278.

Vikrant V., Antoniou C., Wen Y., Ben-Akiva M. (2009). Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance, *Transportation Research Board*, 2009, Vol. 2090, pp.1-9.

Wang. L., J. Rong and X. Liu (2005). The Classification of Car-Following Behavior in Urban Expressway Based on Fuzzy Clustering Analysis. *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Wickham H. (2009). *Ggplot2: Elegant Graphics for Data Analysis (Use R!)*

Wiedemann R. and Reiter U. (1992). Microscopic traffic simulation: the simulation system MISSION, background and actual state. Project ICARUS (V1052) Final Report, Brussels, CEC. 2: Appendix A.

Wilson R.E. (2000). An analysis of Gipps' car-following model of highway traffic. IMA Journal of Applied Mathematics, 66, pp. 509-537.

Yang X. et al. (2004). "A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning." Proc. 1st Annual Int'l. Conf. Mobile and Ubiquitous Syst: Networking and Services.

Χρήσιμες διευθύνσεις στο διαδίκτυο

US Department of Transportation (2012). "NGSIM – Next Generation Simulation."
<http://www.ngsim.fhwa.dot.gov>

