



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Αναδρομική Εκτίμηση Παραμέτρων σε Γραμμικά και
Μη Γραμμικά Μοντέλα**
**Εφαρμογές στην πρόβλεψη ανάπτυξης καρκινικών όγκων και
την παρακολούθηση της θαλάσσιας κίνησης.**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΣΠΥΡΙΔΩΝΟΣ Β. ΠΑΤΜΑΝΙΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικού Υπολογιστών Α.Π.Θ.

ΕΠΙΒΛΕΠΩΝ:

Γ. Π. ΠΑΠΑΒΑΣΙΛΟΠΟΥΛΟΣ

Καθηγητής Ε.Μ.Π.

ΑΘΗΝΑ, Μάρτιος 2019



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Αναδρομική Εκτίμηση Παραμέτρων σε Γραμμικά και
Μη Γραμμικά Μοντέλα**
**Εφαρμογές στην πρόβλεψη ανάπτυξης καρκινικών όγκων και
την παρακολούθηση της θαλάσσιας κίνησης.**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΣΠΥΡΙΔΩΝΟΣ Β. ΠΑΤΜΑΝΙΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικού Υπολογιστών Α.Π.Θ.

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ: 1. Γ. Π. ΠΑΠΑΒΑΣΙΛΟΠΟΥΛΟΣ
2. Π. ΜΑΡΑΓΚΟΣ
3. Α. Χ. ΧΑΡΑΛΑΜΠΙΔΗΣ

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή στις 14/03/2019

..... 1. Γ. Π. ΠΑΠΑΒΑΣΙΛΟΠΟΥΛΟΣ Καθ. Ε.Μ.Π. (Επιβλέπων) 2. Π. ΜΑΡΑΓΚΟΣ Καθ. Ε.Μ.Π. 3. Α. Χ. ΧΑΡΑΛΑΜΠΙΔΗΣ Αναπ. Καθ. CentraleSupélec
---	--	--

..... 4. Δ. ΚΟΥΤΣΟΥΡΗΣ Καθ. Ε.Μ.Π. 5. Γ. ΜΑΤΣΟΠΟΥΛΟΣ Αναπ. Καθ. Ε.Μ.Π.
--	---

..... 6. Κ. ΝΙΚΗΤΑ Καθ. Ε.Μ.Π. 7. Α. Γ. ΣΤΑΦΥΛΟΠΑΤΗΣ Καθ. Ε.Μ.Π.
--------------------------------------	---

ΑΘΗΝΑ, Μάρτιος 2019

Στους γονείς μου Βασίλη και Οντέτα,
στον θείο μου Γιάνη,
και στον αδερφό μου Ηλία!

Copyright ©2019 Σπυρίδων Β. Πατμανίδης

"Με την επιφύλαξη κάθε νόμιμου δικαιώματος. All rights reserved"

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτηματικά που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

«Η έγκριση της παρούσης Διδακτορικής Διατριβής από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέως» (Ν. 5343/1932, άρθρο 202, παρ. 2)

Πρόλογος

Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (ΣΗΜΜΥ) του Εθνικού Μετσόβιου Πολυτεχνείου (Ε.Μ.Π.) στο πλαίσιο των δραστηριοτήτων του Εργαστηρίου Συστημάτων Αυτόματου Ελέγχου (Σ.Α.Ε.) υπό την επίβλεψη του καθηγητή Γεώργιου Π. Παπαβασιλόπουλου. Η διατριβή ξεκίνησε τον Ιανουάριο του 2014 και ολοκληρώθηκε τον Ιανουάριο του 2019.

Αντικείμενο της διατριβής είναι η αναδρομική εκτίμηση των αγνώστων παραμέτρων σε γραμμικά και μη γραμμικά μοντέλα μέσα από τις διαθέσιμες παρατηρήσεις του συστήματος. Στα πλαίσια της διατριβής αυτής μελετήθηκε η εκτίμηση παραμέτρων σε δυο μοντέλα, ένα μη γραμμικό κι ένα γραμμικό, από τα οποία το πρώτο χρησιμοποιήθηκε για την μελέτη της ανάπτυξης καρκινικών όγκων και το δεύτερο για την μελέτη της τροχιάς εμπορικών πλοίων. Επιπλέον, χρησιμοποιήθηκαν τεχνικές μηχανικής μάθησης για την ανάπτυξη δύο εφαρμογών λογισμικού, μια για την εκτίμηση των αγνώστων παραμέτρων του μοντέλου που περιγράφει την ανάπτυξη των καρκινικών όγκων και την πρόβλεψη της μελλοντικής εξέλιξης των όγκων, και μια για την παρακολούθηση της θαλάσσιας κίνησης, τον εντοπισμό μη κανονικών συμπεριφορών και την πρόβλεψη της μελλοντικής εμπορικών πλοίων.

Αξίζει να σημειωθεί ότι η εφαρμογή λογισμικού για την παρακολούθηση της θαλάσσιας κίνησης αναπτύχθηκε σε συνεργασία με την εταιρία ISI Hellas που ασχολείται με το σχεδιασμό, την ανάπτυξη, την ολοκλήρωση, την εγκατάσταση, τον έλεγχο και τη μετέπειτα υποστήριξη του λογισμικού και του εξοπλισμού για προϊόντα όπως Real-time Tactical Command, Control, Communication and Intelligence Systems, Standard Tactical Data Links, Tactical Data Link Planning Tools, Custom Data Links and Interoperability Solutions, Embedded Tactical Systems Trainers, Tactical Systems Simulators, Management Information Systems for Defense Applications, Military Message Preparation and Handling Systems, και άλλα. Τα αποτελέσματα της συγκεκριμένης μελέτης δημοσιεύθηκαν και παρουσιάστηκαν στο συνέδριο MED '16 που πραγματοποιήθηκε τον Ιούνιο του 2016 στην Αθήνα. Όσον αφορά την μελέτη της ανάπτυξης καρκινικών όγκων, τα πρώτα αποτελέσματα που προέκυψαν από τη συγκεκριμένη έρευνα δημοσιεύθηκαν και παρουσιάστηκαν στο συνέδριο IFAC 2017 που πραγματοποιήθηκε τον Ιούλιο του 2017 στην Τουλούζη,

ενώ στη συνέχεια ακολούθησε μια δεύτερη δημοσίευση στο διεθνές επιστημονικό περιοδικό *Computer Methods and Programs in Biomedicine*, Elsevier, τον Ιούλιο του 2018. Οι δύο τελευταίες δημοσιεύσεις προκάλεσαν το ενδιαφέρον και τα θετικά σχόλια αρκετών ερευνητών της διεθνούς κοινότητας που ασχολούνται με το συγκεκριμένο αντικείμενο.

Κατά τη διάρκεια της διδακτορικής μου έρευνας είχα την ευκαιρία να συνεργαστώ με αξιόλογους επιστήμονες και επαγγελματίες, να ανταλλάξω απόψεις, να αποκτήσω νέες γνώσεις και ανεκτίμητες εμπειρίες, γεγονός που συνέβαλε καθοριστικά στην προσωπική μου εξέλιξη και στη διαμόρφωση του χαρακτήρα μου. Μέσω της συμμετοχής μου σε διάφορες δραστηριότητες εντός και εκτός του εργαστηρίου, με οδηγό την περιέργεια και με εφόδιο τις γνώσεις που αποκόμισα από τις σπουδές μου, έμαθα στην πράξη πως σκέφτεται και δρα ένας μηχανικός, τόσο σε ερευνητικό-ακαδημαϊκό όσο και σε επαγγελματικό επίπεδο. Μαθήτευσα πλάι σε εξαιρετους επιστήμονες και ανθρώπους παίρνοντας εμπειρίες ανεκτίμητης αξίας. Στο σημείο λοιπόν αυτό, θα ήθελα να ευχαριστήσω όλους εκείνους που για χρόνια τώρα στάθηκαν δίπλα μου και με βοήθησαν να ολοκληρώσω αυτό το κεφάλαιο της ζωής μου.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Γεώργιο Παπαβασιλόπουλο, για τη συνεχή στήριξη και βοήθεια που μου παρείχε καθ' όλη τη διάρκεια της διδακτορικής μου διατριβής. Οι πολύτιμες συμβουλές του, σε συνδιασμό με τα κίνητρα και την ελευθερία επιλογής, δημιούργησαν για εμένα ένα ιδανικό περιβάλλον για έρευνα, για το οποίο του είμαι ειλικρινά ευγνώμων. Θα ήθελα επίσης να ευχαριστήσω τον κ. Αλέξανδρο Χ. Χαραλαμπίδη, αναπληρωτή καθηγητή στη CentraleSupélec, ο οποίος όχι μόνο συντέλεσε καθοριστικά στο να ξεκινήσω το διδακτορικό μου στο Εργαστήριο Σ.Α.Ε. της ΣΗΜΜΥ του Ε.Μ.Π., αλλά συνεπίβλεψε μαζί με τον κ. Παπαβασιλόπουλο τη διδακτορική μου διατριβή, προσφέροντας μου πολύτιμες συμβουλές και καθοδήγηση. Ένα ιδιαίτερο ευχαριστώ οφείλω επίσης στον Γιάννη Κορδόνη, τόσο για τη σημαντική του συμβολή στο ερευνητικό κομμάτι της διδακτορικής μου διατριβής, όσο και γιατί υπήρξε ένας πολύ καλός φίλος, πάντα πρόθυμος να βοηθήσει. Τέλος, θα ήθελα να ευχαριστήσω όλα τα υπόλοιπα μέλη του Εργαστηρίου Σ.Α.Ε., την Έλενα Σαρρή, τον Νίκο Χρυσαιθόπουλο, τον Παναγιώτη Κοντογιώργο και τον Αθανάσιο Λαγό, για την συμβολή τους στην καθημερινότητα μου σαν υποψήφιος διδάκτορας.

Θα ήθελα επιπλέον να ευχαριστήσω τον κ. Γεώργιο Δ. Μήτση, αναπληρωτή καθηγητή στο Πανεπιστήμιο McGill του Μόντρεαλ και τον κ. Roberto Chignola, καθηγητή στο Πανεπιστήμιο της Βερόνα, τόσο για τα πειραματικά δεδομένα καρκινικών όγκων που μου παρείχαν, όσο και για τις πολύτιμες συμβουλές τους που με βοήθησαν να αποκτήσω βασικές γνώσεις σχετικά με τη φυσιολογία και τη μελέτη του καρκίνου. Επιπλέον, θα ήθελα να ευχαριστήσω ιδιαίτερα την κοινότητα ανοιχτού λογισμικού για την ανιδιοτελή προσφο-

ρά τους σε εργαλεία και δεδομένα. Τέλος, ευχαριστώ τον κ. Πέτρο Μαραγκό, καθηγητή στο Ε.Μ.Π., για τη συμμετοχή του στην τριμελή συμβουλευτική επιτροπή.

Ιδιαίτερες ευχαριστίες θα ήθελα να εκφράσω προς τον Ειδικό Λογαριασμό Έρευνας και Κουδουλίων του Ιδρύματος για την υποτροφία που μου παρείχε και που αποτέλεσε σχεδόν αποκλειστική πηγή χρηματοδότησης για την εκπόνηση της διατριβής μου, καθώς επίσης και προς την Φοιτητική Λέσχη για την χρηματοδότηση της συμμετοχής μου σε δύο επιστημονικά συνέδρια.

Ευχαριστώ πολύ όλους μου τους φίλους που μου συμπαραστάθηκαν όλα αυτά τα χρόνια και με βοήθησαν σε μικρά και μεγάλα προβλήματα που κλήθηκα να αντιμετωπίσω στην καθημερινότητα μου. Ιδιαίτερα ευχαριστώ τον φίλο μου Κώστα Κωισταντόπουλο που υπήρξε πάντοτε παρών όταν τον χρειάστηκα. Τέλος, ευχαριστώ τους γονείς μου Βασίλη και Οντέτα, τον θείο μου Γιάννη, τον αδερφό μου Ηλία και τη σύντροφο μου Σοφιάνα για την συνεχή τους στήριξη σε οτιδήποτε και να χρειάστηκα καθ' όλη τη διάρκεια των σπουδών μου.

ΣΠΥΡΙΔΩΝ Β. ΠΑΤΜΑΝΙΔΗΣ

Μάρτιος 2019

Περίληψη

Αντικείμενο της παρούσας διατριβής είναι η εφαρμογή τεχνικών αναδρομικής εκτίμησης παραμέτρων σε γραμμικά και μη γραμμικά μοντέλα μέσω της αξιοποίησης της πληροφορίας που προσφέρουν οι διαθέσιμες παρατηρήσεις του συστήματος. Στην διατριβή αυτή μελετήθηκαν δύο εφαρμογές. Η πρώτη αφορά τη μοντελοποίηση της ανάπτυξης καρκινικών όγκων και τη δυνατότητα πρόβλεψης της μελλοντικής τους εξέλιξης με βάση τις διαθέσιμες μετρήσεις του όγκου τους. Η δεύτερη αφορά την παρακολούθηση της θαλάσσιας κίνησης, την ανίχνευση μη κανονικών συμπεριφορών στην κίνηση των πλοίων και τη δυνατότητα πρόβλεψης της μελλοντικής τους πορείας.

Όσον αφορά τη μοντελοποίηση της ανάπτυξης καρκινικών όγκων, παρουσιάζεται μια μεθοδολογία για την εκτίμηση των αγνώστων παραμέτρων του μη γραμμικού μοντέλου Gompertz, το οποίο είναι πολύ δημοφιλές στη βιβλιογραφία για την μακροσκοπική περιγραφή της ανάπτυξης καρκινικών όγκων. Το μοντέλο αυτό χαρακτηρίζεται από δυο παραμέτρους που αντιπροσωπεύουν τον ρυθμό αύξησης και την φέρουσα χωρητικότητα του καρκινικού όγκου. Σύμφωνα με στοιχεία που προκύπτουν από διάφορες μελέτες είναι πολύ πιθανό αυτές οι δύο παράμετροι να είναι εξατομικευμένες για κάθε ασθενή. Παρόλα αυτά, στη βιβλιογραφία, λόγω της δυσκολίας για ταυτόχρονη εκτίμηση και των δυο παραμέτρων, συνηθίζεται η φέρουσα χωρητικότητα να θεωρείται γνωστή και σταθερή ανάλογα με το είδος του οργανισμού (π.χ. άνθρωπος, ποντίκι, κλπ.) στο οποίο συνιαντάται ο καρκίνος. Η μεθοδολογία που παρουσιάζεται στην παρούσα διατριβή προτείνει την περιγραφή της ανάπτυξης του καρκινικού όγκου μέσω ενός στοχαστικού μοντέλου που βασίζεται στην εξίσωση Gompertz, τη θεώρηση τόσο του ρυθμού αύξησης όσο και της φέρουσας χωρητικότητας σαν μεταβλητές παραμέτρους και την εκτίμηση τους με τη χρήση τεχνικών που βασίζονται την αρχή της Μέγιστης Πιθανοφάνειας. Τα αποτελέσματα των προσομοιώσεων δείχνουν ότι οι τεχνικές που εφαρμόστηκαν είναι σε θέση να εκτιμήσουν με μεγάλη ακρίβεια τις άγνωστες παραμέτρους στα συνθετικά δεδομένα και να δώσουν στην πλειοψηφία των περιπτώσεων καλύτερες εκτιμήσεις της μελλοντικής εξέλιξης των όγκων σε σχέση με την συνήθη μεθοδολογία που προτείνεται στη βιβλιογραφία όσον αφορά τα πειραματικά δεδομένα.

Για την παρακολούθηση της θαλάσσιας κίνησης και την ανίχνευση των μη κανονικών συμπεριφορών στην κίνηση των πλοίων, αρχικά ορίστηκε ένα πλήθος από συμπεριφορές οι οποίες θεωρήθηκαν μη κανονικές. Στη συνέχεια αναπτύχθηκε μια εφαρμογή λογισμικού, η οποία χρησιμοποιεί δεδομένα από το Σύστημα Αυτόματης Αναγνώρισης με το οποίο είναι εξοπλισμένα τα πλοία και ανιχνεύει τις μη κανονικές συμπεριφορές. Επιπλέον, έγινε χρήση Γραμμικών Αυτοπαλίνδρομων μοντέλων για την εκτίμηση της μελλοντικής τροχιάς των πλοίων με βάση την παρελθοντική τους πορεία και τις γεωγραφικές θέσεις στις οποίες βρέθηκαν. Οι προβλέψεις αυτές χρησιμοποιήθηκαν επίσης για την ανίχνευση μη κανονικών συμπεριφορών από το λογισμικό που αναπτύχθηκε. Το λογισμικό αυτό εφαρμόστηκε επιτυχώς στην πράξη, σε δεδομένα που συλλέχθηκαν από πλοία στον Σαρωνικό Κόλπο.

Λέξεις Κλειδιά: Αναδρομική Εκτίμηση Παραμέτρων, Αναδρομική Εκτίμηση Καταστάσεων, Γραμμικά και Μη Γραμμικά Συστήματα, Μέγιστη Πιθανοφάνεια, Ελάχιστα Τετράγωνα, Φίλτρο Κάλμαν, Φίλτρα Σωματιδίων, Βιολογικά Συστήματα, Καρκίνος, Παρακολούθηση Θαλάσσιας Κίνησης.

Abstract

In this thesis, a variety of recursive estimation techniques for parameter and state estimation on linear and non-linear models were applied. These techniques make use of the information contained in the measurements in order to estimate the unknown parameters. Two separate applications were studied. The first one deals with tumor growth modeling and prediction based on the tumor's volume measurements, while the second one deals with maritime surveillance, vessel monitoring and abnormal behavior detection.

As regards tumor growth modeling, a methodology for the estimation of the unknown parameters of the non-linear Gompertz growth model is presented. This model is commonly used in the literature to describe the growth of macroscopic tumor measurements. According to the Gompertz model, the growth of a solid tumor is characterized by two main features, the growth rate and the carrying capacity. In the literature, when fitting experimental tumor data to the Gompertz equation, the carrying capacity of a tumor is considered as a species-specific parameter, and thus has a constant value. However, the theoretical assumption that any given tumor will approach the same asymptotic value has not been proven experimentally. On the other hand, the heterogeneity of tumor growth has been recognized to have clinical implications in the planning of effective treatment schedules. The methodology presented in this thesis considers both the growth rate and the carrying capacity to be subject-specific parameters and uses estimation techniques based on the principle of Maximum Likelihood in order to estimate their values. It has been observed that this approach can provide more accurate estimates of the unknown parameters for the synthetic data and more accurate predictions of the future growth for the experimental data.

On the topic of maritime surveillance, this thesis focuses on estimating the route of a vessel and detecting abnormal behaviors based on AIS data. A software application was created in order to detect these behaviors and notify the user whether or not further investigation is required. Linear Filtering using AR models is proposed in order to create trajectory forecasts. Furthermore, various alert criteria are set expressing different abnormalities. AIS data collected from Saronikos Bay were used and the results showed that the program could successfully monitor the bay in real time and report abnormalities in vessel behavior.

Keywords: Recursive Parameter Estimation, Recursive State Estimation, Linear and Non-Linear Systems, Maximum Likelihood, Least Squares, Kalman Filter, Particle Filters, Biological Systems, Cancer, Maritime Surveillance.

Περιεχόμενα

Πρόλογος	vii
Περίληψη	xi
Abstract	xiii
Κατάλογος Σχημάτων	xix
Κατάλογος Πινάκων	xxi
Πίνακας Συμβόλων, Ακρονομίων	xxiii
I Ανάπτυξη Καρκινικών Όγκων	1
1 Εισαγωγή	3
1.1 Μη Γραμμικά Στοχαστικά Συστήματα	3
1.2 Περιγραφή του προβλήματος	5
1.3 Καρκίνος	7
1.3.1 Μοντελοποίηση Ανάπτυξης Καρκινικών Όγκων	7
1.3.2 Εφαρμογές	9
2 Εκτίμηση Κατάστασης και Αγνώστων Παραμέτρων	11
2.1 Αναδρομική Μπεϋζιανή Εκτίμηση Κατάστασης	11
2.1.1 Φίλτρο Κάλμαν	13
2.1.2 Επεκτεταμένο Φίλτρο Κάλμαν	15
2.1.3 Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας	15
2.2 Αναδρομική Εκτίμηση Αγνώστων Παραμέτρων	17
2.2.1 Εκτιμητής Μέγιστης Πιθανοφάνειας	17
2.2.2 Εκτιμητής Μέγιστης εκ των Υστέρων Πιθανότητας	18

2.2.3	Μη-Γραμμικά Ελάχιστα Τετράγωνα	19
2.2.4	Αλγόριθμος Μεγιστοποίησης Προσδοκίας	19
3	Υλοποίηση Τεχνικών	21
3.1	Μη-Γραμμικά Ελάχιστα Τετράγωνα	21
3.2	Μέγιστη Πιθανοφάνεια στο απλοποιημένο μοντέλο	22
3.3	Μέγιστη Πιθανοφάνεια με αριθμητική ολοκλήρωση	23
3.4	Εκτιμητής Μέγιστης εκ των Υστέρων Πιθανότητας	26
3.5	Συνδυασμός Επεκτεταμένου Φίλτρου Κάλμαν και Μέγιστης Πιθανοφάνειας	27
3.6	Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας	29
3.7	Παράλληλο Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας	29
4	Προσομοιώσεις και Αποτελέσματα	33
4.1	Προσομοιώσεις με συνθετικά δεδομένα	33
4.1.1	Μετρικές	34
4.1.2	Περιπτώσεις	35
4.1.3	Υλοποίηση	36
4.1.4	Αποτελέσματα	37
4.2	Προσομοιώσεις με πειραματικά δεδομένα	50
4.2.1	Μετρικές	50
4.2.2	Υλοποίηση	51
4.2.3	Αποτελέσματα από ανάλυση Πολυκυτταρικών Σφαιροειδών Όγκων	52
4.2.4	Αποτελέσματα από Δερματικούς Όγκους	66
5	Ανάλυση Αποτελεσμάτων	75
5.1	Συνθετικά δεδομένα	76
5.2	Πειραματικά δεδομένα	78
5.2.1	Πολυκυτταρικά Σφαιροειδή Όγκων	79
5.2.2	Δερματικοί Όγκοι	81
6	Συμπεράσματα	83
II	Θαλάσσια Παρακολούθηση	87
7	Εισαγωγή	89
7.1	Παρακολούθηση Θαλάσσιας Κίνησης	89

Περιεχόμενα	xvii
7.2 Ανασκόπηση Βιβλιογραφίας	89
7.3 Περιγραφή του προβλήματος	90
8 Μεθοδολογία	93
8.1 Εκτίμηση Πορείας Πλοίου	93
8.2 Κριτήρια Συναγερμών	95
9 Εφαρμογή	99
10 Προσομοιώσεις, Αποτελέσματα και Συζήτηση	101
11 Συμπεράσματα	105
Βιβλιογραφία	107
Βιογραφικό Σημείωμα	115

Κατάλογος Σχημάτων

4.1	ΜΑΠΣ και τ.α. των παραμέτρων θ_1 και θ_2 για τον ΕΜΠ με αριθμητική ολοκλήρωση.	39
4.1	ΜΑΠΣ και τ.α. των παραμέτρων θ_1 και θ_2 για τον ΕΜΠ με αριθμητική ολοκλήρωση.	40
4.2	ΜΑΠΣ για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.	41
4.2	ΜΑΠΣ για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.	42
4.3	Τυπική απόκλιση για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.	43
4.3	Τυπική απόκλιση για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.	44
4.4	Καμπύλες ανάπτυξης για τυχαία επιλεγμένα αντικείμενα ελέγχου.	45
4.5	Μέση τιμή Καινοικοποιημένου Μέσου Τετραγωνικού Σφάλματος.	46
4.6	Εκτίμηση κατάστασεων συστήματος με το ΦΔΕΣ.	47
4.6	Εκτίμηση κατάστασεων συστήματος με το ΦΔΕΣ.	48
4.7	Χρόνος Εκτέλεσης για το ΦΔΕΣ.	49
4.8	Χρόνος εκτέλεσης ανά αριθμό μετρήσεων για τον ΕΜΠ με αριθμητική ολοκλήρωση.	52
4.9	Εξέλιξη ανάπτυξης αντιπροσωπευτικών Πολυκυτταρικών Σφαιροειδών Ογκων για κάθε σύνολο δεδομένων.	58
4.9	Εξέλιξη ανάπτυξης αντιπροσωπευτικών Πολυκυτταρικών Σφαιροειδών Ογκων για κάθε σύνολο δεδομένων.	59
4.10	Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή $ΜΓΕΤ_{ΜΕΤ}$	60
4.10	Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή $ΜΓΕΤ_{ΜΕΤ}$	61
4.11	Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή $ΜΓΕΤ_{ΜΤΜ}$	62
4.11	Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή $ΜΓΕΤ_{ΜΤΜ}$	63

4.12 Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ _{2p}	64
4.12 Μέση τιμή ΡΜΤΑ των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ _{2p}	65
4.13 Σύγκριση της ΡΜΤΑ των προβλέψεων μεταξύ του εκτιμητή ΜΓΕΤ και του ΕΜΠ.	68
4.14 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικιού CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ.	69
4.14 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικιού CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ.	70
4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικιού CM37.	71
4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικιού CM37.	72
4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικιού CM37.	73
4.16 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικιού CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ και του ΕΜΥΠ.	74
8.1 Κριτήρια Συναγερμών.	96
10.1 Πραγματική, επαναδειγματοληπτημένη και εκτιμώμενη πορεία για το πλοίο με MMSI 248551000.	103
10.2 Πραγματική, επαναδειγματοληπτημένη και εκτιμώμενη πορεία για το πλοίο με MMSI 514441000.	104

Κατάλογος Πινάκων

4.1	Ελάχιστη και μέγιστη τιμή για κάθε παράμετρο	34
4.2	Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PM- TA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του EMΠ έχει μικρότερη τιμή στο σύνολο δεδομένων MCF7.	55
4.3	Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PM- TA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του EMΠ έχει μικρότερη τιμή στο σύνολο δεδομένων U118.	56
4.4	Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PM- TA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του EMΠ έχει μικρότερη τιμή στο σύνολο δεδομένων 9L.	57
4.5	Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PM- TA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του EMΠ έχει μικρότερη τιμή στο σύνολο δεδομένων από δερματικούς όγ- κους ποντικών.	66

Πίνακας Συμβόλων, Ακροθυμίων

Ακρωνύμια / Συντομογραφίες

AR	AutoRegressive
ARMA	AutoRegressive Moving Average
ARMAX	AutoRegressive Moving Average with eXogenous terms
ARX	AutoRegressive with eXogenous terms
MMSI	Maritime Mobile Service Identity
ΑΜΕ	Αναδρομική Μπεϋζιανή Εκτίμηση
ΑΠΣ	Απόλυτο Ποσοστιαίο Σφάλμα
ΔΔΣ	Διαδοχική Δειγματοληψία Σπουδαιότητας
ΔΕΣ	Διαδοχική Επαναδειγματοληψία Σπουδαιότητας
ΕΜΠ	Εκτιμητής Μέγιστης Πιθανοφάνειας
ΕΜΥΠ	Εκτιμητής Μέγιστης εκ των Υστέρων Πιθανότητας
ΕΦΚ	Επεκτεταμένο Φίλτρο Κάλμαν
ΚΜΤΣ	Καινοικοποιημένο Μέσο Τετραγωνικό Σφάλμα
ΜΑΠΣ	Μέσο Απόλυτο Ποσοστιαίο Σφάλμα
ΜΓΕΤ	Μη-Γραμμικά Ελάχιστα Τετράγωνα
ΜΕΤ	Μέγιστη Εφικτή Τιμή
ΜΤΜ	Μέση Τιμή Μέσων

PMTA	Ρίζα Μέσης Τετραγωνικής Απόκλισης
ΣΑΑ	Σύστημα Αυτόματης Αναγνώρισης
ΦΔΕΣ	Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας
ΦΚ	Φίλτρο Κάλμαν
ΦΣ	Φίλτρα Σωματιδίων
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας
τ.α.	τυπική απόκλιση
τ.μ.	τυχαία μεταβλητή

Μέρος Ι

Ανάπτυξη Καρκινικών Όγκων

Κεφάλαιο 1

Εισαγωγή

1.1 Μη Γραμμικά Στοχαστικά Συστήματα

Τα μη γραμμικά και τα στοχαστικά συστήματα αποτελούν χωρίς αμφιβολία δύο από τις σημαντικότερες κατηγορίες συστημάτων με πολλές εφαρμογές σε διάφορα πεδία έρευνας του Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών. Τέτοια συστήματα για παράδειγμα χρησιμοποιούνται ευρέως σε εφαρμογές που σχετίζονται με τη Ρομποτική, τη Βιοϊατρική Μηχανική, τις Τηλεπικοινωνίες κλπ. [1].

Τα προβλήματα που έχουν σχέση το φιλτράρισμα μη γραμμικών στοχαστικών συστημάτων, αποτελούσαν πάντοτε μια περιοχή στην οποία δίνεται ιδιαίτερη βαρύτητα και προσελκύει το ενδιαφέρον από διάφορα πεδία έρευνας που σχετίζονται με την επιστήμη του μηχανικού. Όσον αφορά την επίλυση προβλημάτων μη γραμμικού φιλτραρίσματος, έχει αναπτυχθεί ένας μεγάλος αριθμός από υποβέλτιστες προσεγγίσεις, όπως για παράδειγμα το Επεκτεταμένο Φίλτρο Κάλμαν (Extended Kalman Filter), η γενικευμένη προσέγγιση ελαχίστων τετραγώνων (generalized Least Squares approach), το σταθμισμένο άθροισμα από Γκαουσιανές πυκνότητες (weighted sum of Gaussian densities) κλπ. Μεταξύ άλλων, μερικές πιο πρόσφατες προσεγγίσεις περιλαμβάνουν εκθετικά φραγμένα φίλτρα (exponentially bounded filters), προσεγγίσεις από Μαρκοβιανές αλυσίδες (approximations by Markov chains), φίλτρα ελάχιστης διακύμανσης (minimum variance filters), μετασχηματισμούς κυματομορφής (wavelet transforms), Φίλτρα Σωματιδίων (Particle Filters) κλπ. [1].

Στις μέρες μας, τα πρακτικά συστήματα που συναντώνται στην επιστήμη του μηχανικού παρουσιάζουν μεγάλο βαθμό πολυπλοκότητας, πράγμα που δημιουργεί σημαντική πρόκληση όσον αφορά την ανάλυση και τη σύνθεση τέτοιων συστημάτων. Μεταξύ άλλων, η μη γραμμικότητα και η στοχαστικότητα εξυπηρετούν σαν δύο βασικές πηγές ρεαλιστικότητας που οδηγούν σε μη αμελητέα πολυπλοκότητα και έχουν αποτελέσει σημαντικό

πεδίο έρευνας. Επιπλέον, λόγω των αναπόφευκτων σφαλμάτων μοντελοποίησης και της εξαρτημένης δυναμικής, μερικά ενδιαφέροντα φαινόμενα, όπως για παράδειγμα οι παράμετροι αβεβαιότητας, πρέπει να λαμβάνονται υπόψη κατά τη μελέτη των διαφόρων συστημάτων για να επιτευχθεί η αναμενόμενη απόδοση. Οι παραπάνω πηγές πολυπλοκότητας καθιστούν επιτακτική την ανάγκη για την ανάπτυξη τεχνικών φιλτραρίσματος και εκτίμησης παραμέτρων για διάφορα είδη περίπλοκων συστημάτων που να ανταποκρίνονται στις απαιτήσεις της σύγχρονης επιστήμης του μηχανικού. Συνεπώς, δεν προκαλεί καμία εντύπωση το γεγονός, ότι τα τελευταία χρόνια, το πρόβλημα του φιλτραρίσματος και της εκτίμησης παραμέτρων σε μη γραμμικά συστήματα έχει ερευνηθεί σε μεγάλο βάθος από πλήθος ερευνητών [1].

Όσον αφορά τη Βιοϊατρική Μηχανική, η χρήση μη γραμμικών συστημάτων γίνεται ολοένα και ευρύτερη. Αν και πολλά σημαντικά επιτεύγματα στην ιατρική οφείλονται στη χρήση γραμμικών συστημάτων, τα συστήματα αυτά αδυνατούν να περιγράψουν τη περίπλοκη δυναμική που παρουσιάζουν οι λειτουργίες του ανθρώπινου σώματος [2]. Οι μη γραμμικότητες είναι πανταχού παρούσες στην φυσιολογία και συχνά είναι απαραίτητες για να προάγουν σημαντικές όψεις των φυσιολογικών λειτουργιών [3]. Ακόμα και στην περίπτωση που είναι δυνατή η γραμμικοποίηση του συστήματος, αυτή οδηγεί σε απόρριψη σημαντικών πληροφοριών που περιέχονται στις μη γραμμικότητες [4]. Βέβαια, αν και η χρήση μη γραμμικών μοντέλων δεν είναι ακόμα σε θέση να εξηγήσει όλες τις πολύπλοκες λειτουργίες που συμβαίνουν στο ανθρώπινο σώμα, και η περαιτέρω βελτίωση και ανάπτυξη μη γραμμικών μοντέλων είναι απαραίτητες, εν τούτοις μας έχει βοηθήσει να κατανοήσουμε συμπεριφορές που τα γραμμικά συστήματα δεν μπορούν και να εισχύσουμε την κατανόηση μας για την περίπλοκη δυναμική των συστημάτων μέσα στο ανθρώπινο σώμα [2]. Ενδεικτικά αναφέρονται μερικά παραδείγματα εφαρμογής μη γραμμικών συστημάτων σε βιοϊατρικά συστήματα:

- **Αισθητηριακά Συστήματα:** Στην κατηγορία αυτή ανήκουν διάφοροι αισθητήρες, όπως το ακουστικό σύστημα [5, 6] και η ρετίνα [7, 8].
- **Βρόχοι Αντανακλαστικών:** Μη γραμμικά συστήματα έχουν χρησιμοποιηθεί για να μελετηθούν βρόχοι αντανακλαστικών για τον έλεγχο των άκρων [9, 10] και τη θέση του ματιού [11, 12].
- **Οργανικά Συστήματα:** Μη γραμμικοί βρόχοι ανάδρασης έχουν εφαρμοστεί σε μοντέλα μεταβλητότητας καρδιακού ρυθμού [13] και νεφρικής αυτορρύθμισης [14--16].
- **Μηχανισμοί Ιστών:** Βιολογικοί ιστοί, όπως οι πνευμονικοί ιστοί [17, 18] και οι σκελετικοί μύες [19--21], παρουσιάζουν μη γραμμική συμπεριφορά.

- Μοντέλα Δυναμικής Ασθενειών: Μη γραμμικά μοντέλα χρησιμοποιούνται ευρέως για τη μοντελοποίηση ασθενειών, όπως για παράδειγμα του ιού HIV/AIDS [22--24] και του καρκίνου [25--27].
- Μη γραμμικές διεργασίες: Επίπεδα σακχάρου και παραγωγή ινσουλίνης [28].

1.2 Περιγραφή του προβλήματος

Η γενικότερη περιγραφή, σε μορφή εξισώσεων κατάστασης-χώρου, ενός πεπερασμένων διαστάσεων συστήματος είναι:

$$x_{k+1} = f(x_k, u_k, w_k, \theta), \quad (1.1)$$

$$y_k = h(x_k, u_k, v_k, \theta), \quad (1.2)$$

όπου x_k είναι η κατάσταση του συστήματος, u_k είναι η κατάσταση της εισόδου ελέγχου του συστήματος και y_k η κατάσταση της εξόδου του συστήματος κατά τη χρονική στιγμή k . Με w_k συμβολίζεται ο θόρυβος της διεργασίας (ή αλλιώς διαταραχή), ενώ με v_k συμβολίζεται ο θόρυβος της μέτρησης. Τέλος, η μεταβλητή θ συμβολίζει ένα διάλυσμα αγνώστων παραμέτρων. Στην παρούσα εργασία, το σύστημα που μελετάται είναι μια ειδική μορφή του παραπάνω συστήματος, όπου η έξοδος y_k έχει γραμμική συσχέτιση με την κατάσταση x_k , δεν υπάρχει κάποια είσοδος ελέγχου στο σύστημα μας ($u_k = 0$), ενώ τόσο ο θόρυβος της διεργασίας όσο και ο θόρυβος της μέτρησης είναι προσθετικοί. Συνεπώς, το σύστημα που μελετάται έχει τη μορφή:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + w_k, \quad (1.3)$$

$$y_k = x_k + v_k. \quad (1.4)$$

Στα πλαίσια της παρούσας εργασίας, εκτός από τα σημεία που αναφέρεται κάτι διαφορετικό, οι τυχαίες μεταβλητές w_k και v_k , όπου $k \in \mathbb{N}$, θεωρούνται αμοιβαία ανεξάρτητες και καινοικά κατανομημένες ($w_k \sim \mathcal{N}(0, g_d(x_k, \theta_3, \theta_4))$ και $v_k \sim \mathcal{N}(0, g_m(x_k, \theta_5, \theta_6))$) με μηδενικό μέσο και άγνωστες παραμέτρους όσον αφορά τη διακύμανση. Οι μετρήσεις y_k είναι διαθέσιμες από τη στιγμή $k = 1$ και έπειτα, ενώ η αρχική κατάσταση x_0 θεωρείται γνωστή. Λόγω των τυχαίων στοιχείων w_k και v_k , στο παραπάνω σύστημα αντιστοιχεί και η στοχαστική περιγραφή:

$$x_{k+1} \sim p_{\theta_a}(x_{k+1}|x_k), \quad (1.5)$$

$$y_k \sim p_{\theta_b}(y_k|x_k), \quad (1.6)$$

όπου $p_\theta(x_{k+1}|x_k)$ είναι η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) που περιγράφει τη δυναμική για δοσμένες τιμές της x_k και $p_\theta(y_k|x_k)$ είναι η σ.π.π. που περιγράφει τις μετρήσεις. Σημειώνεται ότι στην εξίσωση (1.5) το διάνυσμα θ_a είναι $\theta_a = [\theta_1 \theta_2 \theta_3 \theta_4]^T$ και στην (1.6) το διάνυσμα θ_b είναι $\theta_b = [\theta_5 \theta_6]^T$. Σημειώνεται επίσης ότι $\theta_a \in \Theta_a$, όπου $\Theta_a \subseteq \mathbb{R}_{>0}^4$ και $\theta_b \in \Theta_b$, όπου $\Theta_b \subseteq \mathbb{R}_{>0}^2$.

Η παρούσα εργασία πραγματεύεται την εκτίμηση των παραμέτρων ενός μη γραμμικού συστήματος, που εκφράζει την εξέλιξη του μεγέθους στερεών καρκινικών όγκων και περιγράφεται από τις εξισώσεις (1.3)--(1.4). Κατά τα αρχικά στάδια, το μέγεθος των καρκινικών όγκων αυξάνεται εκθετικά. Αργότερα, καθώς το μέγεθος του όγκου μεγαλώνει, ο ρυθμός αύξησης μειώνεται και ο όγκος φτάνει στο οριακό του μέγεθος. Η συμπεριφορά αυτή μπορεί να προβλεφθεί με επιτυχία από τη συνάρτηση Gompertz [29]. Το μοντέλο Gompertz έχει χρησιμοποιηθεί ευρέως λόγω της απλότητας του και της ικανότητας του να περιγράφει επιτυχώς πειραματικά δεδομένα. Θεωρούμε ότι το διάνυσμα θ είναι:

$$\theta = \left[\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \right]^T, \quad (1.7)$$

όπου θ_1 (ρυθμός αύξησης) είναι μια σταθερά που σχετίζεται με την ικανότητα πολλαπλασιασμού των κυττάρων, θ_2 (φέρουσα χωρητικότητα) είναι η σταθερά που περιγράφει το μέγιστο μέγεθος του καρκινικού όγκου ($\lim_{k \rightarrow \infty} x_k = \theta_2$), θ_3 και θ_4 είναι δυο σταθερές που σχετίζονται με την τυπική απόκλιση του θορύβου της διεργασίας και θ_5 και θ_6 είναι δυο σταθερές που σχετίζονται με την τυπική απόκλιση του θορύβου της μέτρησης. Αν θεωρήσουμε επίσης ότι η συνάρτηση f είναι η συνάρτηση Gompertz [30], τότε η εξίσωση (1.3) μπορεί να γραφεί ως:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + w_k = \theta_2 \exp \left(\log \left(\frac{x_k}{\theta_2} \right) \exp \left(- \frac{1}{\theta_1} t_{k+1} \right) \right) + w_k, \quad (1.8)$$

όπου x_k (mm^3) είναι το μέγεθος του καρκινικού όγκου, t_{k+1} (μέρες) είναι το χρονικό διάστημα μεταξύ k και $k+1$, και $k \in \mathbb{N}$. Επίπλεον, όσον αφορά τις τυπικές αποκλίσεις των τ.μ. w_k και v_k , θεωρούμε ότι:

$$g_d(x_k, \theta_3, \theta_4) = \left(\theta_3 x_k^{\theta_4} \right)^2 \quad (1.9)$$

και

$$g_m(x_k, \theta_5, \theta_6) = \left(\theta_5 x_k^{\theta_6} \right)^2. \quad (1.10)$$

Το πρόβλημα που τίθεται προς επίλυση είναι εκτίμηση ενός $\hat{\theta}$ του διανύσματος παραμέτρων θ με βάση ένα σύνολο από N διαθέσιμες μετρήσεις $Y_N = [y_1, \dots, y_n]$ [31, 32].

1.3 Καρκίνος

Ο καρκίνος είναι μια ασθένεια που μπορεί να επηρεάσει τους περισσότερους ιστούς στο ανθρώπινο σώμα. Αναφέρεται στον αφύσικο και ανεξέλεγκτο πολλαπλασιασμό κυττάρων, με πιθανότητα εισβολής και διάδοσης στους περιβάλλοντες ιστούς. Σαν επιπλέον χαρακτηριστικό, σε πολλούς καρκίνους, αλλά όχι σε όλους, μπορεί να συμβεί μεταστάση καρκινικών κυττάρων από το αρχικό μέρος σε ένα μακρινό όπυ και εγκαθίστανται, ένα φαινόμενο το οποίο ονομάζεται μετάσταση. Ο καρκίνος περιλαμβάνει επίσης και καλοήθεις όγκους, που παρουσιάζουν ανεξέλεγκτο πολλαπλασιασμό κυττάρων, αλλά όχι εισβολή σε άλλους ιστούς. Οι καλοήθεις όγκοι στις περισσότερες των περιπτώσεων δεν απειλούν τη ζωή των ασθενών [33].

Ο καρκίνος είναι βασική αιτία νοσηρότητας και θνησιμότητας, επηρεάζοντας πληθυσμούς σε όλες τις χώρες και περιοχές. Ο αριθμός των ετήσιων νέων περιπτώσεων αναμένεται να ξεπεράσει τα 20 εκατομμύρια μέχρι το 2025, αριθμός πολύ μεγάλος σε σχέση με την εκτίμηση για το 2012 που κυμαινόταν στα 14.1 εκατομμύρια παγκοσμίως [34]. Η θεραπευτική αγωγή εξαρτάται από τον τύπο του καρκίνου και συνήθως αποτελείται από μια ή συνδυασμό περισσότερων θεραπειών, όπως εγχείρηση, χημειοθεραπεία, ραδιοθεραπεία, ανοσοθεραπεία και ορμονοθεραπεία.

1.3.1 Μοντελοποίηση Ανάπτυξης Καρκινικών Όγκων

Η μοντελοποίηση της ανάπτυξης των καρκινικών όγκων είναι ένα πεδίο στο οποίο έχει δοθεί ιδιαίτερη βαρύτητα και το οποίο έχει αποτελέσει αντικείμενο συστηματικής έρευνας. Στην εργασία των Araujo και McElwain [35] περιγράφεται εν συντομία η ιστορία της μελέτης της ανάπτυξης στερεών όγκων με την επιθεώρηση μερικών από των πιο σημαντικών μαθηματικών συνεισφορών από τις αρχές του εικοστού αιώνα.

Ειδικότερα, όταν τα δεδομένα περιγράφουν στερεά δομή όγκων, το πιο συνηθισμένο μοντέλο που χρησιμοποιείται για την μοντελοποίηση είναι το μοντέλο Gompertz. Λόγω του ότι τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα εργασία ανήκουν στην προηγούμενη κατηγορία, επιλέχθηκε να γίνει χρή του συγκεκριμένου μοντέλου. Η Γκομπερτζιανή εξίσωση έχει τις ρίζες στο μοντέλο θνησιμότητας που ανέπτυξε ο Benjamin Gompertz το 1825 [29], και έχει επίσης εφαρμοστεί στη μελέτη της ανάπτυξης σε τομείς της βιολογίας και των οικονομικών [36]. Έχει παρατηρηθεί ότι το μοντέλο Gompertz μπορεί να περιγράψει την ανάπτυξη διαφόρων ειδών όγκων [37]. Κάτι τέτοιο φαίνεται επίσης και σε διάφορες πρόσφατες έρευνες στις οποίες το μοντέλο Gompertz συγκρίνεται με άλλα κλασσικά μαθηματικά μοντέλα που χρησιμοποιούνται για την μοντελοποίηση της εξέλιξης καρκινικών όγκων, όπως για παράδειγμα το εκθετικό μοντέλο (exponential), το

μοντέλο Von Bertalanffy, ο νόμος δύναμης (power law) κλπ. [25--27]. Για το λόγο αυτό, το μοντέλο Gompertz αποτελεί μια συχνή επιλογή για τη μοντελοποίηση της ανάπτυξης των καρκινικών όγκων σε πολλές πρόσφατες πειραματικές μελέτες [38--44].

Σύμφωνα με το μοντέλο Gompertz, η ανάπτυξη ενός στερεού καρκινικού όγκου χαρακτηρίζεται από δύο βασικά γνωρίσματα, τον ρυθμό αύξησης και την φέρουσα χωρητικότητα. Ο ρυθμός αύξησης αναφέρεται στον χρόνο που χρειάζονται τα καρκινικά κύτταρα για να παράξουν νέα κύτταρα, ενώ η φέρουσα χωρητικότητα στο γεγονός ότι ο όγκος δεν μπορεί να μεγαλώνει απεριόριστα. Κατά τη διαδικασία ταιριάσματος πειραματικών δεδομένων στην εξίσωση Gompertz, αρκετές μελέτες [42--45] κάνουν χρήση των ευρημάτων των Brunton και Wheldon [46--48], σύμφωνα με τα οποία η φέρουσα χωρητικότητα ενός όγκου είναι μια παράμετρος που καθορίζεται από το είδος του οργανισμού. Παρόλα αυτά, ο θεωρητικός αυτός ισχυρισμός, ότι ο οποιοσδήποτε όγκος σε ένα συγκεκριμένο είδος θα πλησιάσει την ίδια ασυμπτωτική τιμή, δεν έχει αποδειχτεί πειραματικά. Αν και υπάρχουν σημαντικές ενδείξεις ότι η ανάπτυξη ενός καρκινικού όγκου είναι μια εξατομικευμένη διαδικασία του οργανισμού (και επομένως η φέρουσα χωρητικότητα δεν είναι ίδια για όλους τους οργανισμούς ενός είδους), η προσπάθεια να εκτιμηθεί η φέρουσα χωρητικότητα οδηγεί συνήθως σε μη ρεαλιστικά αποτελέσματα και συνεπώς προτιμάται η χρήση της σταθερής τιμής για τη φέρουσα χωρητικότητα [42].

Είναι γεγονός ότι δεν υπάρχει ένα καθολικό μοντέλο που να μπορεί να μοντελοποιήσει όλα τα είδη των καρκίνων. Μιας και κάθε μοντέλο δεσμεύεται από συγκεκριμένους περιορισμούς, η επιλογή του κατάλληλου μοντέλου σε κάθε περίπτωση γίνεται με βάση τον τύπο των δεδομένων που μελετώνται. Όσον αφορά το μοντέλο Gompertz, η κύρια κριτική που δέχεται οφείλεται στο ότι ο σχετικός ρυθμός αύξησης γίνεται αυθέραιτα μεγάλος για μικρούς καρκινικούς όγκους [25]. Επιπλέον, το συγκεκριμένο μοντέλο δεν μοντελοποιεί κάποιον βιολογικό μηχανισμό που αποτελεί θεμέλιο του καρκίνου, πράγμα που σημαίνει ότι δεν παρέχει κάποια γνώση όσον αφορά τις βιολογικές διαδικασίες. Έχει δειχθεί ότι είναι το καταλληλότερο για καρκίνους του μαστού και του πνεύμονα [25] και είναι επίσης ιδιαίτερα δημοφιλές για την περιγραφή καρκίνων σε μεγάλο μέγεθος ζώα, χωρίς όμως να είναι να ικανό να περιγράψει με ακρίβεια όλους τους τύπους καρκίνου. Είναι όμως ένα απλό μοντέλο, κατάλληλο για την μακροσκοπική μοντελοποίηση της ανάπτυξης καρκινικών όγκων. Το γεγονός ότι μπορεί να χρησιμοποιηθεί σε συνδιασμό με αλγόριθμους βέλτιστου ελέγχου κατά την βελτιστοποίηση της θεραπείας, έχει κάνει το μοντέλο Gompertz ιδιαίτερα δημοφιλές. Συν τοις άλλοις, έχει δειχτεί ότι είναι κατάλληλο για την πρόβλεψη της μέσης συμπεριφοράς της ανάπτυξης ενός όγκου [25, 37, 49, 50]. Περισσότερες πληροφορίες όσον αφορά τα πλεονεκτήματα και τους περιορισμούς μερικών από τα πιο δημοφιλή μοντέλα για την περιγραφή της ανάπτυξης καρκινικών όγκων

και πιθανούς λόγους για τους οποίους το μοντέλο Gompertz κατέλληξε να είναι το πιο ευρέως εφαρμόσιμο μπορούν να βρεθούν στην εργασία του Gerlee [51].

1.3.2 Εφαρμογές

Η χρήση της μαθηματικής μοντελοποίησης έχει προταθεί για τη βελτίωση του αποτελέσματος της χημειοθεραπείας και της ραδιοθεραπείας σε συνδυασμό με τεχνικές βέλτιστου ελέγχου [38--40, 52], για την πρόβλεψη της επιτυχίας ή αποτυχίας της ανοσοθεραπείας [53], και για να αποτυπώσει την δυνατότητα συνέργειας μεταξύ διαφορετικών θεραπευτικών μεθόδων [54]. Πρόσφατα ευρήματα δείχνουν ότι η χρήση μαθηματικής μοντελοποίησης μπορεί να προσφέρει χρήσιμες ενδείξεις σχετικά με τις επιπτώσεις της εγχείρησης όσον αφορά τη μετάσταση και να βοηθήσει στην πρόβλεψη του ρίσκου διάδοσης του καρκίνου [55]. Αναλυτική βιβλιογραφία, όσον αφορά το θέμα του πως η μαθηματική μοντελοποίηση μπορεί να εφαρμοστεί για να επιτευχθούν καλύτερα προγράμματα παροχής φαρμάκων, παρουσιάζεται στην εργασία των Michor και Beal [56].

Η μοντελοποίηση της ανάπτυξης καρκινικών όγκων είναι εγγενώς δύσκολη, μιας και ο καρκίνος είναι ένα περίπλοκο βιολογικό φαινόμενο. Το γεγονός ότι η μορφή, η ανάπτυξη και το αποτέλεσμα ποικίλει από ασθενή σε ασθενή καθιστά επιτακτική την ανάγκη για μοντελοποίηση της ανάπτυξης του καρκινικού όγκου σε ατομική βάση, έτσι ώστε να γίνει καλύτερα κατανοητή η βιολογία που βρίσκεται από πίσω και να σχεδιαστούν πιο αποτελεσματικές θεραπείες. Προς αυτή την κατεύθυνση, ο σχεδιασμός εξειδικευμένων τεχνικών για την επίλυση του προβλήματος της μοντελοποίησης της ανάπτυξης καρκινικών όγκων μπορεί να παρέχει πιο ακριβή μοντέλα και να βελτιώσει τη διαδικασία μοντελοποίησης.

Κεφάλαιο 2

Εκτίμηση Κατάστασης και Αγνώστων Παραμέτρων

Στο κεφάλαιο αυτό περιγράφεται η μεθοδολογία στην οποία βασίζονται οι υλοποιήσεις των τεχνικών που χρησιμοποιήθηκαν στην εργασία αυτή για την εκτίμηση της κατάστασης του συστήματος και των αγνώστων παραμέτρων του μοντέλου Gompertz.

2.1 Αναδρομική Μπεϋζιανή Εκτίμηση Κατάστασης

Εκτίμηση κατάστασης είναι η διαδικασία με την οποία αποδίδονται τιμές στις άγνωστες καταστάσεις του συστήματος με βάση μετρήσεις από το ίδιο το σύστημα και συγκεκριμένα κριτήρια. Η διαδικασία αυτή περιλαμβάνει μη τέλειες μετρήσεις, ενώ μπορούν να χρησιμοποιηθούν και τυχόν γνώσεις που αφορούν την ακρίβεια των συσκευών μέτρησης. Επιπλέον, στη διαδικασία αυτή χρησιμοποιούνται γνώσεις που αφορούν την τοπολογία του συστήματος.

Η Αναδρομική Μπεϋζιανή Εκτίμηση (Recursive Bayesian Estimation) είναι μια γενικευμένη πιθανοτική προσέγγιση για την δημιουργία της εκ των υστέρων (posterior) σ.π.π. των καταστάσεων με βάση όλη τη διαθέσιμη πληροφορία, συμπεριλαμβανομένου και ενός συνόλου ληφθέντων μετρήσεων. Από την στιγμή που η εκ των υστέρων σ.π.π. εμπεριέχει όλη τη διαθέσιμη στατιστική πληροφορία, μπορούμε να πούμε ότι αποτελεί την πλήρη λύση στο πρόβλημα εκτίμησης. Κατ' αρχήν, μια βέλτιστη (όσον αφορά κάποιο οποιοδήποτε κριτήριο) εκτίμηση μπορεί να ληφθεί από την εκ των υστέρων σ.π.π. Σε πολλά προβλήματα, μια εκτίμηση της κατάστασης είναι αναγκαία κάθε φορά που μια νέα μέτρηση γίνεται διαθέσιμη. Σε αυτές τις περιπτώσεις τα αναδρομικά φίλτρα είναι μια βολική λύση, μιας και τα δεδομένα που λαμβάνονται μπορούν να επεξεργαστούν διαδοχικά, χωρίς να είναι

απαραίτητες η αποθήκευση των δεδομένων στο σύνολο τους και η επανεπεξεργασία της υπάρχουσας πληροφορίας τη στιγμή που η νέα μέτρηση γίνεται διαθέσιμη. Στο στάδιο της πρόβλεψης χρησιμοποιείται το μοντέλο του συστήματος για να προβλεφθεί η σ.π.π. της κατάστασης από τη χρονική τιμή μιας μέτρησης στην επόμενη. Από τη στιγμή που η κατάσταση συνήθως υπόκειται σε άγνωστες διαταραχές (που μοντελοποιούνται σαν τυχαίος θόρυβος), η πρόβλεψη συνήθως παραμορφώνει και απλώνει τη σ.π.π. Ακολουθεί το στάδιο της ενημέρωσης (ή διόρθωσης) στο οποίο η σ.π.π. τροποποιείται με τη χρήση της τελευταίας μέτρησης χρησιμοποιώντας το Θεώρημα του Bayes [57].

Στην παρούσα εργασία ενδιαφερόμαστε για το πρόβλημα της Μη Γραμμικής Μπεϋζιανής Παρακολούθησης (Non-Linear Bayesian Tracking) σε συστήματα διακριτού χρόνου. Στη γενική μορφή του προβλήματος αυτού, θεωρούμε ότι το διάνυσμα κατάστασης $x_k \in \mathbb{R}^n$ εξελίσσεται με βάση το μοντέλο του παρακάτω συστήματος:

$$x_{k+1} = f_k(x_k, w_k), \quad (2.1)$$

όπου $f_k : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ είναι η συνάρτηση μετάβασης του συστήματος και $w_k \in \mathbb{R}^m$ ακολουθία λευκού θορύβου μηδενικού μέσου ανεξάρτητη από τις παρελθοντικές και τρέχουσες καταστάσεις, της οποίας όμως η σ.π.π. θεωρείται γνωστή. Οι μετρήσεις $y_k \in \mathbb{R}^p$ γίνονται γνωστές σε διακριτά χρονικά διαστήματα και σχετίζονται με το διάνυσμα κατάστασης μέσω της εξίσωσης:

$$y_k = h_k(x_k, v_k), \quad (2.2)$$

όπου $h_k : \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}^p$ είναι η συνάρτηση των μετρήσεων και $v_k \in \mathbb{R}^r$ ακολουθία λευκού θορύβου μηδενικού μέσου ανεξάρτητη από τις παρελθοντικές και τρέχουσες καταστάσεις, της οποίας η σ.π.π. θεωρείται επίσης γνωστή. Θεωρούμε επίσης ότι η αρχική σ.π.π. $p(x_1|Y_0) \equiv p(x_1)$ του διανύσματος κατάστασης είναι γνωστή μαζί με τις συναρτησιακές μορφές f_i και h_i για $i = 1, \dots, k$. Η διαθέσιμη πληροφορία τη χρονική στιγμή k είναι το σύνολο των μετρήσεων $Y_N = \{y_i : i = 1, \dots, k\}$ [58].

Η απαίτηση που προκύπτει είναι να κατασκευαστεί η σ.π.π. για την κατάσταση x_k με βάση τη διαθέσιμη πληροφορία που υπάρχει, δηλαδή η σ.π.π. $p(x_k|Y_k)$. Η συγκεκριμένη σ.π.π. μπορεί να προκύψει αναδρομικά σε δύο στάδια, το στάδιο της πρόβλεψης και το στάδιο της ενημέρωσης. Υποθέτουμε ότι η απαιτούμενη σ.π.π. $p(x_{k-1}|Y_{k-1})$ στο χρονικό σημείο $k - 1$ είναι γνωστή. Τότε, στο βήμα της πρόβλεψης, χρησιμοποιώντας το μοντέλο του συστήματος μπορεί να προκύψει η εκ των προτέρων σ.π.π. της κατάστασης στο

χρονικό σημείο k από την εξίσωση:

$$p(x_k|Y_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1})dx_{k-1}, \quad (2.3)$$

Σημειώνεται εδώ ότι στην εξίσωση (2.3), γίνεται χρήση του γεγονότος ότι $p(x_k|x_{k-1}) = p(x_k|x_{k-1}, Y_{k-1})$, μιας και η εξίσωση (2.1) περιγράφει μια Μαρκοβιανή διαδικασία πρώτης τάξης. Το πιθανοτικό μοντέλο της εξέλιξης της κατάστασης, $p(x_k|x_{k-1})$, ορίζεται από την εξίσωση του συστήματος και τα γνωστά στατιστικά του w_{k-1} .

Τη χρονική στιγμή k , μια νέα μέτρηση y_k γίνεται διαθέσιμη, και μπορεί να χρησιμοποιηθεί για να ενημερωθεί (στάδιο ενημέρωσης) η εκ των προτέρων σ.π.π. της κατάστασης μέσω του κανόνα του Bayes:

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(Y_k|Y_{k-1})}, \quad (2.4)$$

όπου η σταθερά κανονικοποίησης

$$p(y_k|Y_{k-1}) = \int p(y_k|x_k)p(x_k|Y_{k-1})dx_k, \quad (2.5)$$

εξαρτάται από τη συνάρτηση πιθανοφάνειας $p(y_k|x_k)$, που ορίζεται από την εξίσωση (2.2). Επομένως, από το στάδιο ενημέρωσης προκύπτει μέσω της εκ των προτέρων σ.π.π. (2.3) η εκ των υστέρων σ.π.π. (2.4).

Οι επαναληπτικές εξισώσεις (2.3) και (2.4) αποτελούν την τυπική λύση του προβλήματος της ΑΜΕ. Αναλυτική λύση σε αυτό το πρόβλημα είναι δυνατή μόνο για ένα σχετικά περιορισμένο αριθμό περιπτώσεων στις οποίες ισχύουν συγκεκριμένες προϋποθέσεις. Στην παρούσα εργασία, οι καταστάσεις, οι μετρήσεις και οι θόρυβοι είναι βαθμωτά μεγέθη. Εύκολα όμως μπορεί κανείς να χρησιμοποιήσει τις μεθόδους που περιγράφονται στη συνέχεια και με διανύσματα.

Για την εκτίμηση των καταστάσεων του συστήματος στην παρούσα εργασία έγινε χρήση του ΕΦΚ και του Φίλτρου Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας (Sequential Importance Resampling Filter). Επειδή, το ΕΦΚ βασίζεται στο Φίλτρο Κάλμαν (Kalman Filter), ακολουθεί μία σύντομη αναφορά σε αυτό.

2.1.1 Φίλτρο Κάλμαν

Το ΦΚ είναι ένα αποτελεσματικό αναδρομικό φίλτρο που εκτιμά την κατάσταση γραμμικών δυναμικών συστημάτων με θορυβώδεις μετρήσεις. Έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας και έχει χρησιμοποιηθεί ευρέως σε διάφορους τομείς. Πρό-

κειται για μια από τις περιορισμένες περιπτώσεις, όπου μπορεί να προκύψει αναλυτική επίλυση του προβλήματος της ΑΜΕ. Το ΦΚ προϋποθέτει οι f_k και h_k να είναι γραμμικές και w_k και v_k να είναι προσθετικοί Γκαουσσισιανοί θόρυβοι με γνωστή διακύμανση.

Ας υποθέσουμε ότι έχουμε ένα δυναμικό σύστημα που περιγράφεται από τις εξισώσεις:

$$x_{k+1} = A_k x_k + B_k + w_k, \quad (2.6)$$

$$y_{k+1} = C_k x_k + D_k + v_k, \quad (2.7)$$

όπου οι τ.μ. x_0 , w_k και v_k είναι αμοιβαία ανεξάρτητες και με γνωστές παραμέτρους. Επιπλέον, οι τ.μ. w_k και v_k έχουν μηδενικό μέσο και οι μήτρες συνδιακύμανσης τους συμβολίζονται με Q και R αντίστοιχα. Τέλος, οι μετρήσεις είναι διαθέσιμες από τη στιγμή $k = 1$ και έπειτα.

Υποθέτουμε επίσης ότι η τ.μ. x_k ακολουθεί την κανονική κατανομή με μέσο \hat{x}_k και μήτρα συνδιακύμανσης P_{x_k} . Από την (2.6) η οποία είναι γραμμική και γνωρίζοντας ότι οι γραμμικοί μετασχηματισμοί Γκαουσσισιανών τ.μ. είναι επίσης Γκαουσσισιανοί, έχουμε ότι η τ.μ. x_{k+1} ακολουθεί την κανονική κατανομή με μέσο \hat{x}_{k+1}^- και μήτρα συνδιακύμανσης \hat{P}_{k+1}^- που δίνονται από τις εξισώσεις:

$$\hat{x}_{k+1}^- = A_k \hat{x}_k + B_k, \quad (2.8)$$

$$P_{x_{k+1}}^- = A_k P_{x_k} A_k^T + Q. \quad (2.9)$$

Από την (2.6), επίσης σαν συνέπεια της γραμμικότητας, έχουμε ότι η εκτίμηση της y_{k+1} είναι:

$$\hat{y}_{k+1}^- = C_{k+1} \hat{x}_{k+1}^- + D_k, \quad (2.10)$$

ενώ η μήτρα συνδιακύμανσης της είναι:

$$P_{y_{k+1}}^- = C_{k+1} P_{x_{k+1}}^- C_{k+1}^T + R, \quad (2.11)$$

και η μήτρα ετεροσυνδιακύμανσης της x_{k+1} με την y_{k+1} είναι:

$$P_{x_{k+1}y_{k+1}}^- = P_{x_{k+1}}^- C_{k+1}^T. \quad (2.12)$$

Όταν γίνει λήψη της τιμής της μέτρησης y_{k+1} , αυτή μπορεί να χρησιμοποιηθεί για να γίνει διόρθωση της κατανομής. Το κέρδος Κάλμαν (Kalman Gain) δίνεται από την εξίσωση:

$$K_{k+1} = P_{x_{k+1}y_{k+1}}^- \left(P_{y_{k+1}}^- \right)^{-1}, \quad (2.13)$$

και χρησιμοποιείται για τον υπολογισμό των εκ των υστέρων τιμών του μέσου και της μήτρας συνδιακύμανσης της x_{k+1} , σύμφωνα με τις σχέσεις:

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - \hat{y}_{k+1}^-), \quad (2.14)$$

$$P_{x_{k+1}} = P_{x_{k+1}}^- - K_{k+1} P_{y_{k+1}}^- K_{k+1}^T. \quad (2.15)$$

Επομένως, το βήμα πρόβλεψης πραγματοποιείται βάσει των (2.8) και (2.9), ενώ το βήμα διόρθωσης βάσει των (2.14) και (2.15). Αφού επομένως είναι γνωστό ότι η x_0 είναι κανονικά κατανοημένη με μέσο \hat{x}_0 και μήτρα συνδιακύμανσης P_{x_0} , οι (2.9) ως (2.14) μπορούν να εφαρμοστούν αναδρομικά για να δώσουν τα στατιστικά χαρακτηριστικά της κατάστασης όταν είναι γνωστές οι μετρήσεις Y_N . Για περισσότερες πληροφορίες για το ΦΚ βλ. [59--63].

2.1.2 Επεκτεταμένο Φίλτρο Κάλμιν

Στο ΕΚΦ [64--66], οι εξισώσεις (2.8) και (2.10) γραμμικοποιούνται με τη χρήση του αναπτύγματος Taylor πρώτης τάξης. Πιο συγκεκριμένα, για το βήμα πρόβλεψης, αν \hat{x}_k είναι ο εκτιμώμενος μέσος της x_k ύστερα από το βήμα διόρθωσης την χρονική στιγμή k , τότε η (2.8) γραμμικοποιείται γύρω από την \hat{x}_k . Για το βήμα διόρθωσης, η (2.10) γραμμικοποιείται γύρω από την \hat{x}_{k+1}^- , όπου \hat{x}_{k+1}^- είναι ο εκτιμώμενος μέσος της x_{k+1} ύστερα από το βήμα πρόβλεψης την χρονικής στιγμής $k + 1$. Για να μπορεί να πραγματοποιηθεί κάτι τέτοιο, γίνεται η υπόθεση ότι οι Ιακωβιανές μήτρες των f_k και h_k είναι διαθέσιμες.

Το ΕΚΦ έχει χρησιμοποιηθεί ευρέως σε διάφορες εφαρμογές. Η απόδοση του είναι ικανοποιητική όταν οι μη γραμμικότητες είναι ήπιες. Διαφορετικά μπορεί να παρατηρηθούν ακόμα και φαινόμενα αστάθειας.

2.1.3 Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας

Το ΦΔΕΣ προτάθηκε από τους Gordon κ.α. [58] και είναι μια μέθοδος Monte Carlo που μπορεί να εφαρμοστεί στο πρόβλημα της AME. Οι μέθοδοι Monte Carlo είναι μια κατηγορία υπολογιστικών αλγορίθμων που χρησιμοποιούνται ευρέως σε προβλήματα, τα οποία είναι δύσκολο να επιλυθούν είτε αναλυτικά, είτε με τη χρήση κλασικών αριθμητικών μεθόδων. Βασίζονται στη χρήση τυχαίων αριθμών και στατιστικής, και χρησιμοποιούνται ιδιαίτερα στην επίλυση προβλημάτων φιλτραρίσματος που εμφανίζονται στην επεξεργασία σήματος και την Μπεϋζιανή στατιστική συναγωγή.

Το μεγάλο πλεονέκτημα της τεχνικής αυτής είναι ότι δεν υπόκειται σε περιορισμούς που αφορούν τις συναρτήσεις f_k και h_k (π.χ. γραμμικότητα) ή τις κατανομές των θο-

Algorithm 1 Διαδοχική Δειγματοληψίας Σπουδαιότητας (ΔΔΣ)

For times $k = 0, 1, 2, \dots$

- For $i = 1, \dots, N$, draw $x_k^{(i)} \sim \pi(x_k | x_{0:k-1}^{(i)}, y_{0:k})$ and set $x_{0:k}^{(i)} \triangleq (x_{0:k-1}^{(i)}, x_k^{(i)})$
- For $i = 1, \dots, N$, evaluate the importance weights up to a normalizing constant:

$$r_k^{*(i)} = r_{k-1}^{*(i)} \frac{p(y_k | x_k^{(i)}) p(x_k^{(i)} | x_{k-1}^{(i)})}{\pi(x_k | x_{0:k-1}^{(i)}, y_{0:k})} \quad (2.16)$$

- For $i = 1, \dots, N$, normalize the importance weights:

$$\tilde{r}_k^{(i)} = \frac{r_k^{*(i)}}{\sum_{j=1}^N r_k^{*(j)}} \quad (2.17)$$

ρύβων της διαδικασίας ή της μέτρησης. Οι μόνες προϋποθέσεις είναι να είναι γνωστές οι συναρτήσεις f και h , οι σ.π.π. $p(x_1)$ και $p(w_k)$ να είναι διαθέσιμες για δειγματοληψία και να είναι γνωστή η συνάρτησιακή μορφή της πιθανοφάνειας $p(y_k | x_k)$. Ο αλγόριθμος ΔΕΣ προκύπτει εύκολα από τον αλγόριθμο Διαδοχικής Δειγματοληψίας Σπουδαιότητας (Sequential Importance Sampling). Για τον λόγο αυτό θα περιγραφεί πρώτα ο αλγόριθμος ΔΔΣ και έπειτα οι τροποποιήσεις που χρειάζονται ώστε να προκύψει ο αλγόριθμος ΔΕΣ.

Από τη στιγμή που είναι αδύνατον να δειγματοληψήσουμε απευθείας από την εκ των υστέρων σ.π.π. της κατάστασης $p(x_{0:k} | y_{0:k})$, υιοθετούμε μια προσέγγιση Δειγματοληψίας Σπουδαιότητας (Importance Sampling) [67]. Θεωρούμε ότι τα δείγματα $\{x_{0:k}^{(i)}, i = 0, \dots, N\}$ επιλέγονται ανεξάρτητα από μια κανονικοποιημένη συνάρτηση σπουδαιότητας $\pi(x_{0:k} | y_{0:k})$. Μια εκτίμηση $\widehat{I}_N(f_k)$ της εκ των υστέρων πρόβλεψης $I_N(f_k)$ μπορεί να αποκτηθεί με τη χρήση Μπεϋζιανής Δειγματοληψίας Σπουδαιότητας [68]:

$$\widehat{I}_N(f_k) = \sum_i^N f_k(x_{0:k}^{(i)}) \tilde{r}_k^{(i)}, \quad (2.18)$$

$$\tilde{r}_k^{(i)} = \frac{r_k^{*(i)}}{\sum_{j=1}^N r_k^{*(j)}}, \quad (2.19)$$

όπου $r_k^{*(i)} = p(y_{0:k} | x_{0:k}) p(x_{0:k}) / \pi(x_{0:k} | y_{0:k})$ είναι το μη κανονικοποιημένο βάρος σπουδαιότητας. Υπό ασθενείς προϋποθέσεις η $\widehat{I}_N(f_k)$ συγκλίνει στην $I_N(f_k)$ [68]. Όμως, αυτή η μέθοδος δεν είναι αναδρομική. Ας υποθέσουμε όμως μια συνάρτηση σπουδαιότητας

της μορφής:

$$\pi(x_{0:k}|y_{0:k}) = \pi(x_0|y_0) \prod_{i=1}^k \pi(x_k|x_{0:k-1}, y_{0:k}). \quad (2.20)$$

Μια συνάρτηση σπουδαιότητας αυτής της μορφής επιτρέπει αναδρομικό υπολογισμό στο χρόνο των βαρών σπουδαιότητας καθώς οι διαδοχικές παρατηρήσεις y_k γίνονται διαθέσιμες. Προκύπτει έτσι ο αλγόριθμος 1 για τη ΔΔΣ [67].

Ο αλγόριθμος ΔΕΣ προκύπτει εύκολα από τον αλγόριθμο ΔΔΣ αν επιλεγεί ως πυκνότητα σπουδαιότητας η εκ των προτέρων σ.π.π. $p(x_k|x_{k-1}^{(i)})$ και πραγματοποιηθεί επαναδειγματοληψία του συνόλου των δειγμάτων $\{x_k^{(i)}\}$ σε κάθε χρονική στιγμή. Για περισσότερες πληροφορίες όσον αφορά τους αλγορίθμους ΔΔΣ και ΔΕΣ, τη Δειγματοληψία Σπουδαιότητας, τα μειονεκτήματα και τα πλεονεκτήματα των ΦΣ που αναφέρθηκαν και τη σημασία της επαναδειγματοληψίας βλ. [57, 58, 67]. Η υλοποίηση του αλγορίθμου ΔΕΣ που χρησιμοποιήθηκε στην παρούσα εργασία θα παρουσιαστεί αναλυτικά στο επόμενο κεφάλαιο.

2.2 Αναδρομική Εκτίμηση Αγνώστων Παραμέτρων

Η εκτίμηση παραμέτρων είναι ένα αντικείμενο το οποίο αναφέρεται σε πάρα πολλά βιβλία, με εύρος από πολύ θεωρητικές παρουσιάσεις στατιστικολόγων έως πιο πρακτικές αντιμετωπίσεις από χρήστες εφαρμοσμένης στατιστικής. Η θεωρία εκτίμησης συναντάται σε πολλά συστήματα επεξεργασίας σημάτων για εξαγωγή πληροφοριών. Αυτά τα συστήματα περιλαμβάνουν ραντάρ, ομιλία, επεξεργασία εικόνας, βιοϊατρική, τηλεπικοινωνίες, έλεγχο κλπ. Το κοινό σε όλα αυτά τα συστήματα είναι η ανάγκη για την εκτίμηση των τιμών κάποιας ομάδας παραμέτρων [61].

Για την εκτίμηση των αγνώστων παραμέτρων του συστήματος στην παρούσα εργασία έγινε χρήση του Εκτιμητή Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimator) [69, 70], του Εκτιμητή Μέγιστης εκ των Υστέρων Πιθανότητας (Maximum A Posteriori Probability Estimator) [61] και της μεθόδου των Μη-Γραμμικών Ελαχίστων Τετραγώνων (Non-Linear Least Squares).

2.2.1 Εκτιμητής Μέγιστης Πιθανοφάνειας

Ο ΕΜΠ βασίζεται στην αρχή της μέγιστης πιθανοφάνειας, η οποία είναι η πιο δημοφιλής προσέγγιση για την απόκτηση πρακτικών εκτιμητών [61]. Ορίζεται ως η τιμή $\hat{\theta}$ που

μεγιστοποιεί την συνάρτηση πιθανοφάνειας $p_\theta(Y_N)$ των μετρήσεων:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_\theta(Y_N). \quad (2.21)$$

Το διάνυσμα Y_N των μετρήσεων είναι φιξαρισμένο και η μεγιστοποίηση πραγματοποιείται στο χώρο επιτρεπτών τιμών του θ . Με βάση τον Κανόνα του Bayes προκύπτει ότι:

$$p_\theta(Y_N) = p_\theta(y_1) \prod_{k=2}^N p_\theta(y_k | Y_{k-1}), \quad (2.22)$$

όπου

$$p_\theta(y_k | Y_{k-1}) = \int p_\theta(y_k | x_k) p_\theta(x_k | Y_{k-1}) dx_k, \quad (2.23)$$

$$p_\theta(x_k | Y_{k-1}) = \int p_\theta(x_k | x_{k-1}) p_\theta(x_{k-1} | Y_{k-1}) dx_{k-1}, \quad (2.24)$$

και

$$p_\theta(x_k | Y_k) = \frac{p_\theta(y_k | x_k) p_\theta(x_k | Y_{k-1})}{p_\theta(y_k | Y_{k-1})}. \quad (2.25)$$

Γεικότερα, ο αναλυτικός υπολογισμός των παραπάνω ολοκληρωμάτων δεν είναι δυνατός. Στην περίπτωση όμως που εξετάζεται στην παρούσα εργασία, λόγω της μικρής διάστασης του x_k (είναι μονοδιάστατο), είναι δυνατό να χρησιμοποιηθεί αριθμητική ολοκλήρωση για τον υπολογισμό των ολοκληρωμάτων. Επιπλέον, μια ακόμα προσέγγιση για την επίλυση του προβλήματος είναι να εκτιμηθεί η κατάσταση x_k και να χρησιμοποιηθούν οι εκτιμήσεις \hat{x}_k για τον υπολογισμό του διανύσματος $\hat{\theta}$ που μεγιστοποιεί την πιθανοφάνεια $p_\theta(Y_N)$ των μετρήσεων [31, 32].

2.2.2 Εκτιμητής Μέγιστης εκ των Υστέρων Πιθανότητας

Ένα μειονέκτημα του ΕΜΠ είναι ότι αγνοεί οποιαδήποτε εμπειρική γνώση σε σχέση με τις άγνωστες παραμέτρους μπορεί να είναι διαθέσιμη. Η γνώση αυτή είναι δυνατόν να εκφραστεί από μια κατανομή η οποία ονομάζεται εκ των προτέρων (prior) και μπορεί να επηρεάζει τις εκτιμήσεις των αγνώστων παραμέτρων. Ένας αποτελεσματικός τρόπος για να χρησιμοποιηθεί τέτοιου είδους γνώση είναι η χρήση του ΕΜΥΠ. Ο εκτιμητής αυτός επιλέγει το διάνυσμα των αγνώστων παραμέτρων $\hat{\theta}$, έτσι ώστε να μεγιστοποιείται η εκ των υστέρων σ.π.π., η οποία είναι το γινόμενο της πιθανοφάνειας $p_\theta(Y_N)$ και της εκ των προτέρω σ.π.π. $p(\theta)$ [71]:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_\theta(Y_N) p(\theta). \quad (2.26)$$

2.2.3 Μη-Γραμμικά Ελάχιστα Τετράγωνα

Μια από τις γενικότερες μεθόδους που μπορεί να χρησιμοποιηθεί για την εκτίμηση αγνώστων παραμέτρων στην περίπτωση που τα δεδομένα που έχουν παρατηρηθεί μοντελοποιούνται από μια μη γραμμική συνάρτηση είναι η μέθοδος των Μη-Γραμμικών Ελαχίστων Τετραγώνων. Ένα πρόβλημα ΜΓΕΤ αναφέρεται συχνά στη βιβλιογραφία και σαν μη γραμμική παλινδρομηση [61]. Η μέθοδος αυτή είναι μια από τις πιο συνηθισμένες που προτείνεται στη βιβλιογραφία για την εκτίμηση των αγνώστων παραμέτρων του μοντέλου Gompertz, όταν αυτό χρησιμοποιείται για την μοντελοποίηση καρκινικών όγκων [42, 43].

Αν θεωρήσουμε λοιπόν ότι η εκτιμήτρια συνάρτηση των καταστάσεων του συστήματος είναι η:

$$\hat{x}_k = f(\hat{x}_{k-1}, \theta), \quad (2.27)$$

όπου \hat{x}_k είναι η εκτίμηση της κατάστασης του συστήματος με την τιμή της \hat{x}_0 να είναι γνωστή και θ το διάνυσμα των αγνώστων παραμέτρων, τότε η εκτίμηση $\hat{\theta}$ για το διάνυσμα των αγνώστων παραμέτρων προκύπτει από την ελαχιστοποίηση του αθροίσματος των τετραγώνων της διαφοράς μεταξύ των μετρήσεων y_k και των εκτιμήσεων \hat{x}_k :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{k=1}^N (y_k - \hat{x}_k)^2. \quad (2.28)$$

2.2.4 Αλγόριθμος Μεγιστοποίησης Προσδοκίας

Μια από τις πιο γνωστές μεθόδους για τον υπολογισμό των αγνώστων παραμέτρων που μεγιστοποιούν την πιθανοφάνεια σε πιθανοτικά μοντέλα με κρυμμένες μεταβλητές είναι ο αλγόριθμος Μεγιστοποίησης Προσδοκίας (Expectation Maximization) [72]. Ο αλγόριθμος της Μεγιστοποίησης Προσδοκίας δεν υλοποιήθηκε στην παρούσα εργασία, θα μπορούσε όμως να αποτελέσει επέκταση σε μελλοντική συνέχειά της. Για μια αναλυτικότερη περιγραφή σχετικά με τον τρόπο που μπορεί να εφαρμοστεί ο αλγόριθμος Μεγιστοποίησης Προσδοκίας για την εκτίμηση αγνώστων παραμέτρων σε μη γραμμικά μοντέλα κατάστασης χώρου, ο αναγνώστης παραπέμπεται στην εργασία των Schön κ.α. [70].

Κεφάλαιο 3

Υλοποίηση Τεχνικών

Σε αυτό το κεφάλαιο θα παρουσιαστούν οι μέθοδοι που υλοποιήθηκαν για την επίλυση του προβλήματος που συζητήθηκε στο υποκεφάλαιο 1.2. Αυτές οι μέθοδοι βασίζονται στη θεωρία που αναπτύχθηκε στο Κεφάλαιο 2 και σε συνδυασμό, είτε με κάποιες παραδοχές που γίνονται, είτε με παρατηρήσεις που αφορούν τη δομή του προβλήματος, καταφέρνουν να επιλύσουν το πρόβλημα που καλούνται να αντιμετωπίσουν.

Αρχικά παρουσιάζεται η υλοποίηση για τη μέθοδο των ΜΓΕΤ (υποκεφάλαιο 3.1). Σημειώνεται ότι η μέθοδος των ΜΓΕΤ εφαρμόστηκε στα πειραματικά δεδομένα για να γίνει σύγκριση με τις μεθόδους που προτείνονται στην εργασία αυτή. Στη συνέχεια παρουσιάζονται μια υλοποίηση του ΕΜΠ σε ένα απλοποιημένο μοντέλο (υποκεφάλαιο 3.2), μια υλοποίηση του ΕΜΠ στην οποία γίνεται χρήση αριθμητικής ολοκλήρωσης (υποκεφάλαιο 3.3) για τον υπολογισμό των ολοκληρωμάτων των εξισώσεων (2.23) και (2.24), μια υλοποίηση του ΕΜΥΠ (υποκεφάλαιο 3.4) και μια υλοποίηση στην οποία γίνεται συνδιασμός του ΕΜΠ και του ΕΦΚ (υποκεφάλαιο 3.5). Τέλος, παρουσιάζονται μια ακολουθιακή και μια παράλληλη υλοποίηση του ΦΔΕΣ (υποκεφάλαια 3.6, 3.7), οι οποίες χρησιμοποιήθηκαν για την εκτίμηση των καταστάσεων του συστήματος στις προσομοιώσεις με συνθετικά δεδομένα.

3.1 Μη-Γραμμικά Ελάχιστα Τετράγωνα

Κατά τη χρήση της μεθόδου των ΜΓΕΤ για την μοντελοποίηση καρκινικών όγκων με το μοντέλο Gompertz, η παράμετρος που περιγράφει την φέρουσα χωρητικότητα του όγκου συνηθίζεται να θεωρείται γνωστή και να φιξάρεται σε μια συγκεκριμένη τιμή που εξαρτάται από το είδος του οργανισμού που μελετάται. Έτσι, απομένει να υπολογιστεί μόνο η παράμετρος που περιγράφει το ρυθμό αύξησης. Αν λοιπόν φιξαριστεί η φέρουσα χωρητικότητα (σταθερά K) και με θ_1 συμβολίσουμε το ρυθμό αύξησης του μοντέλου

Gompertz, προκύπτει η εξίσωση:

$$\hat{x}_k = K \exp \left(\log \left(\frac{\hat{x}_{k-1}}{K} \right) \exp \left(- \frac{1}{\theta_1} t_k \right) \right). \quad (3.1)$$

Αν με $\hat{\theta}_1$ συμβολίσουμε την εκτίμηση για την άγνωστη παράμετρο θ_1 , όπου $\theta_1 \in \Theta_1$ με $\Theta_1 \subseteq \mathbb{R}_{>0}$, τότε η εκτίμηση $\hat{\theta}_1$ προκύπτει από την ελαχιστοποίηση του αθροίσματος των τετραγώνων της διαφοράς μεταξύ των μετρήσεων y_k και των εκτιμήσεων \hat{x}_k :

$$\hat{\theta}_1 = \arg \min_{\theta_1 \in \Theta_1} \sum_{k=1}^N (y_k - \hat{x}_k)^2. \quad (3.2)$$

3.2 Μέγιστη Πιθανοφάνεια στο απλοποιημένο μοντέλο

Στη συγκεκριμένη υλοποίηση γίνεται μια παραδοχή που απλοποιεί σημαντικά το πρόβλημα. Υποθέτουμε ότι ο θόρυβος της μέτρησης είναι κομμάτι του θορύβου της διαδικασίας και επομένως μπορούμε να θεωρήσουμε ότι οι μετρήσεις y_k που λαμβάνονται είναι οι καταστάσεις x_k του συστήματος, γεγονός που απλοποιεί τη λύση του προβλήματος, μιας και οι εξισώσεις (2.23) ως (2.25) δε χρειάζεται να υπολογιστούν. Επομένως, ο ΕΜΠ μπορεί να χρησιμοποιηθεί απευθείας, όπως στην περίπτωση που οι μετρήσεις προέρχονται από ένα σύστημα της μορφής:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + d_k, \quad (3.3)$$

$$y_k = x_k, \quad (3.4)$$

όπου $d_k \sim \mathcal{N}(0, \sigma_k^2)$ και $\sigma_k = \theta_3 x_k^{\theta_4} + \theta_5 x_k^{\theta_6}$. Το διάνυσμα $\hat{\theta}$ είναι η εκτίμηση του διανύσματος θ που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_\theta(Y_N)$ των μετρήσεων:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_\theta(Y_N). \quad (3.5)$$

Κάνοντας χρήση του Κανόνα του Bayes, η πιθανοφάνεια $p_\theta(Y_N)$ μπορεί να γραφτεί ως

$$p_\theta(Y_N) = \prod_{k=1}^N p_\theta(y_k | Y_{k-1}) = \prod_{k=1}^N p_\theta(x_k | x_{k-1}), \quad (3.6)$$

όπου η αρχική κατάσταση x_0 θεωρείται γνωστή και η συνάρτηση πιθανοφάνειας $p_\theta(x_k | x_{k-1})$ προέρχεται από την εξίσωση (3.3).

3.3 Μέγιστη Πιθανοφάνεια με αριθμητική ολοκλήρωση

Στη δεύτερη υλοποίηση του ΕΜΠ γίνεται χρήση αριθμητικής ολοκλήρωσης για τον υπολογισμό των ολοκληρωμάτων των εξισώσεων (2.23) και (2.24). Η μέθοδος αυτή εφαρμόζεται στο μοντέλο που περιγράφεται από τις εξισώσεις (1.3) και (1.4), και χρησιμοποιείται για να βρεθεί η εκτίμηση $\hat{\theta}$ για το διάνυσμα θ , η οποία μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_{\theta}(Y_N)$ των μετρήσεων:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_{\theta}(Y_N), \quad (3.7)$$

όπου

$$p_{\theta}(Y_N) = p_{\theta}(y_1) \prod_{k=2}^N p_{\theta}(y_k | Y_{k-1}), \quad (3.8)$$

Για τον υπολογισμό της συνάρτησης πιθανοφάνειας $p_{\theta}(Y_N)$ γίνεται χρήση των εξισώσεων (2.23) ως (2.25). Οι κατανομές των εξισώσεων (2.23) ως (2.25) μπορούν να υπολογιστούν επαναληπτικά. Για να υπολογιστούν τα ολοκληρώματα των εξισώσεων (2.23) και (2.24) χρησιμοποιείται αριθμητική ολοκλήρωση. Το διάστημα της ολοκλήρωσης είναι $[0, X_{max}]$, όπου X_{max} είναι η μέγιστη τιμή που μπορεί να φτάσει ένας καρκινικός όγκος. Για τον υπολογισμό των ορισμένων ολοκληρωμάτων που προκύπτουν γίνεται χρήση του κανόνα του τραπεζίου:

$$\int_a^b f(x) dx \approx (b - a) \left[\frac{f(a) + f(b)}{2} \right]. \quad (3.9)$$

Στην εξίσωση (3.9), το διάστημα ολοκλήρωσης $[a, b]$ επηρεάζει σημαντικά το σφάλμα προσέγγισης, λόγω της μη γραμμικότητας της συνάρτησης $f(x)$ στην οποία γίνεται η ολοκλήρωση. Για να μειωθεί το σφάλμα προσέγγισης, το ολοκλήρωμα που ορίζεται στο διάστημα $[0, X_{max}]$ μπορεί να χωριστεί σε επιμέρους ολοκληρώματα, έτσι ώστε το αρχικό ολοκλήρωμα να προκύπτει από το άθροισμα των επιμέρους ολοκληρωμάτων. Καθώς μεγαλώνει ο αριθμός των επιμέρους ολοκληρωμάτων (δηλ. όσο μικραίνει το διάστημα ολοκλήρωσης του κάθε ολοκληρώματος), αυξάνεται η ακρίβεια της προσέγγισης. Λόγω όμως του μεγέθους του διαστήματος $[0, X_{max}]$, ο αριθμός των επιμέρους ολοκληρωμάτων που χρειάζεται για να γίνει μια ακριβής προσέγγιση του αρχικού ολοκληρώματος είναι πολύ μεγάλος και καθιστά την προσέγγιση αυτή αναποτελεσματική, αφού ο αριθμός των πράξεων που απαιτείται σε κάθε επανάληψη είναι τεράστιος. Συνεπώς, η ανάγκη για εξεύρεση ενός τρόπου ώστε να περιοριστεί το διάστημα ολοκλήρωσης είναι επιτακτική, επειδή μπορεί να μειώσει τον αριθμό των επιμέρους ολοκληρωμάτων που χρειάζονται για να γίνει με ακρίβεια η προσέγγιση του αρχικού ολοκληρώματος και επομένως να μειωθεί ο χρόνος εκτέλεσης.

Χρησιμοποιώντας την Γκαουσιανή ιδιότητα του θορύβου, το διάστημα ολοκλήρωσης μπορεί να μειωθεί, έτσι ώστε να καταστεί εφαρμόσιμη η αριθμητική ολοκλήρωση με τη χρήση του κανόνα του τραπεζίου. Σε κάθε επαναληπτικό βήμα, αν η κατάσταση x_k θεωρηθεί τυχαία μεταβλητή με μέσο $\mu'_k = y_k$ και τυπική απόκλιση $\sigma'_k = \sigma_0 y_k^{e_0}$ ($x_k \sim \mathcal{N}(\mu'_k, (\sigma'_k)^2)$), τότε το διάστημα $[\mu'_k - 5\sigma'_k, \mu'_k + 5\sigma'_k] \cap [0, X_{max}]$, που είναι πολύ μικρότερο από το $[0, X_{max}]$, μπορεί να χρησιμοποιηθεί σαν διάστημα ολοκλήρωσης αντί του $[0, X_{max}]$. Το ολοκλήρωμα που ορίζεται στο νέο διάστημα μπορεί με τη σειρά του να χωριστεί σε επιμέρους ολοκληρώματα, που μπορούν να προσεγγιστούν κάνοντας χρήση του κανόνα του τραπεζίου. Σαν αποτέλεσμα, οι κατανομές των εξισώσεων (2.22) έως (2.25) μπορούν να προσεγγιστούν με πολύ μεγάλη ακρίβεια.

Ο Αλγόριθμος 2 δείχνει την αλγοριθμική αναπαράσταση του ΕΜΠ με τη χρήση αριθμητικής ολοκλήρωσης. Αρχικά, υπολογίζεται το διάστημα της ολοκλήρωσης και χωρίζεται σε επιμέρους διαστήματα, έτσι ώστε να εφαρμοστεί ο κανόνας του τραπεζίου σε καθένα από αυτά τα διαστήματα. Τα διανύσματα X_k περιέχουν τα όρια των διαστημάτων που χρησιμοποιούνται για να υπολογιστούν τα επιμέρους ολοκληρώματα για κάθε κατάσταση. Το μέγεθος του X_k ορίζεται από τον αριθμό των επιμέρους ολοκληρωμάτων. Συνεπώς, αν το διάστημα της ολοκλήρωσης χωριστεί σε M υποδιάστηματα, τότε το μέγεθος του X_k θα είναι $M + 1$ και

$$X_k[i] = \mu'_k - 5\sigma'_k + \frac{10\sigma'_k \cdot (i - 1)}{M}, \quad (3.10)$$

όπου $k = 1, 2, \dots, N$ και $i = 1, 2, \dots, M + 1$. Η κατανομή πιθανότητας του $p_\theta(x_1|Y_0)$ δίνεται από την εξίσωση

$$p_\theta(x_1|Y_0)[i] = \frac{1}{\sqrt{2\pi s_0^2}} \exp\left(-\frac{(X_1[i] - f(x_0, \theta_1, \theta_2))^2}{2s_0^2}\right), \quad (3.11)$$

όπου $s_0 = \theta_3 y_0^{\theta_4}$ και $i = 1, 2, \dots, M + 1$. Μετά από αυτή την αρχικοποίηση, ξεκινάει ένας επαναληπτικός βρόχος στον οποίο αρχικά υπολογίζεται η κατανομή πιθανότητας $p_\theta(y_k|x_k)$ που δίνεται από την εξίσωση

$$p_\theta(y_k|x_k)[i] = \frac{1}{\sqrt{2\pi\sigma_{k,i}^2}} \exp\left(-\frac{(y_k - X_k[i])^2}{2\sigma_{k,i}^2}\right), \quad (3.12)$$

όπου $\sigma_{k,i} = \theta_5 X_k[i]^{\theta_6}$, $k = 1, 2, \dots, N$ και $i = 1, 2, \dots, M + 1$. Στη συνέχεια υπολογίζεται το γινόμενο των $p_\theta(y_k|x_k)$ και $p_\theta(x_k|Y_{k-1})$ και αποθηκεύεται στο διάνυσμα p . Η πιθανοφάνεια

Algorithm 2 Εκτιμητής Μέγιστης Πιθανοφάνειας

```

1: procedure maximumLikelihood( $Y_N$ )
2:
3:    $\lambda = 1$ 
4:   Compute vector  $X_1$ :  $x_1 \sim \mathcal{N}(\mu'_1, (\sigma'_1)^2)$ 
5:   Compute vector  $p_\theta(x_1|Y_0)$ 
6:
7:   for  $k = 1; k \leq N; k++$  do
8:
9:     Compute vector  $p_\theta(y_k|x_k)$ 
10:    Compute  $p = p_\theta(y_k|x_k) p_\theta(x_k|Y_{k-1})$ 
11:    Compute  $p_\theta(y_k|Y_{k-1}) = \int_{x_k[1]}^{x_k[N+1]} p dX_k$ 
12:     $\lambda = \lambda \cdot p_\theta(y_k|Y_{k-1})$ 
13:    Compute  $p_\theta(x_k|Y_k) = \frac{p}{p_\theta(y_k|Y_{k-1})}$ 
14:    Compute vector  $X_{k+1}$ :  $x_{k+1} \sim \mathcal{N}(\mu'_{k+1}, (\sigma'_{k+1})^2)$ 
15:
16:    for  $i = 1; i \leq \text{length}(X_{k+1}) + 1; i++$  do
17:
18:      for  $j = 1; j \leq \text{length}(X_{k+1}) + 1; j++$  do
19:        Compute  $p_\theta(x_{k+1}|x_k)[i, j]$ 
20:      end for
21:
22:      Compute  $p_\theta(x_{k+1}|Y_k)[i] = \int_{x_k[1]}^{x_k[N+1]} p_\theta(x_{k+1}|x_k)[i, :] p_\theta(x_k|Y_k) dX_k$ 
23:
24:    end for
25:
26:  end for
27:
28:  Compute vector  $p_\theta(y_k|x_k)$ 
29:  Compute  $p = p_\theta(y_k|x_k) p_\theta(x_k|Y_{k-1})$ 
30:  Compute  $p_\theta(y_k|Y_{k-1}) = \int_{x_k[1]}^{x_k[N+1]} p dX_k$ 
31:   $\lambda = \lambda \cdot p_\theta(y_k|Y_{k-1})$ 
32:
33: end procedure

```

$p_\theta(y_k|Y_{k-1})$, η οποία είναι το ολοκλήρωμα του p στο διάστημα που ορίζει το X_k , υπολογίζεται με τη χρήση του κανόνα του τραπεζιού. Έπειτα, αφού υπολογιστεί η κατανομή πιθανότητας $p_\theta(x_k|Y_k)$, ξεκινάει ένας εσωτερικός βρόχος επανάληψης για να υπολογιστεί

η κατανομή πιθανότητας $p_\theta(x_{k+1}|x_k)$ σύμφωνα με την εξίσωση

$$p_\theta(x_{k+1}|x_k)[i, j] = \frac{1}{\sqrt{2\pi s_{k,j}^2}} \exp\left(-\frac{(X_k[i] - \hat{X}_{k+1}[j])^2}{2s_{k,j}^2}\right), \quad (3.13)$$

όπου $s_{k,j} = \theta_3 X_k[j]^{\theta_4}$, $\hat{X}_{k+1}[j] = f(X_k[j], \theta_1, \theta_2)$, $i = 1, 2, \dots, M+1$ και $j = 1, 2, \dots, M+1$. Τέλος, η κατανομή πιθανότητας $p_\theta(x_{k+1}|Y_k)$ υπολογίζεται ως το ολοκλήρωμα του γινομένου των $p_\theta(x_{k+1}|x_k)[i, 1 : M+1]$ και $p_\theta(x_k|Y_k)$ στο διάστημα που ορίζει το διάνυσμα X_k .

3.4 Εκτιμητής Μέγιστης εκ των Υστέρων Πιθανότητας

Σύμφωνα με το υποκεφάλαιο 2.2.2, ο ΕΜΥΠ υπολογίζει τις παραμέτρους που μεγιστοποιούν το γινόμενο της εκ των προτέρων σ.π.π. $p(\theta)$ με την πιθανοφάνεια $p_\theta(Y_N)$:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_\theta(Y_N)p(\theta). \quad (3.14)$$

Στη ανάλυση που έγινε χρησιμοποιήθηκε μια πολυμεταβλητή Γκαουσιανή κατανομή σαν εκ των προτέρων κατανομή $p(\theta)$, η οποία υπολογίστηκε με την συνάρτηση `fitgmdist` στο πρόγραμμα Matlab. Για να προκύψει η εκ των προτέρων κατανομή πιθανότητας, χρησιμοποιήθηκαν οι εκτιμήσεις $\hat{\theta}_1$ and $\hat{\theta}_2$ που προέκυψαν από τον ΕΜΠ για κάθε όγκο με όλες τις μετρήσεις διαθέσιμες. Η πιθανοφάνεια υπολογίστηκε χρησιμοποιώντας τον Αλγόριθμο 2. Για την υλοποίηση του ΕΜΥΠ, δημιουργήθηκε ένα δισδιάστατο πλέγμα μεγέθους 41×41 για κάθε μια κατανομή (εκ των προτέρων και πιθανοφάνεια) για ένα σύνολο παραμέτρων θ_1 και θ_2 . Οι τιμές των παραμέτρων που αντιστοιχούν στα χαρακτηριστικά του θορύβου (θ_3 έως θ_6) ορίστηκαν στις τιμές που εκτιμήθηκαν με τον ΕΜΠ. Η εκτίμηση $\hat{\theta}$ της εξίσωσης (3.14) μπορεί να υπολογιστεί από το πλέγμα της εκ των υστέρων κατανομής (γινόμενο εκ των προτέρων και πιθανοφάνειας). Ο βασικός λόγος που επιλέχθηκε να χρησιμοποιηθεί το πλέγμα για κάθε κατανομή ήταν η δυνατότητα οπτικοποίησης των κατανομών, η οποία μπορεί να οδηγήσει σε καλύτερη κατανόηση του προβλήματος. Από τη στιγμή που οι τιμές των κατανομών για κάθε στοιχείο του πλέγματος μπορούν να υπολογιστούν ανεξάρτητα, ενδείκνυται η χρήση παράλληλου προγραμματισμού για τον υπολογισμό τους, έτσι ώστε να μειωθεί ο χρόνος υπολογισμού. Για το λόγο αυτό χρησιμοποιήθηκε παράλληλος κώδικας σε γλώσσα CUDA [73]. Η παραλληλοποίηση του κώδικα είναι άμεση μιας και πρόκειται για όμοιους και ανεξάρτητους υπολογισμούς. Έτσι, το κάθε νήμα (thread) αναλαμβάνει να υπολογίσει την τιμή της κατανομής σε ένα σημείο του πλέγματος. Οι

υπολογισμοί για το δισδιάστατο πλέγμα απαιτούν λίγα μόνο δευτερόλεπτα όταν εκτελούνται παράλληλα.

3.5 Συνδυασμός Επεκτεταμένου Φίλτρου Κάλμαν και Μέγιστης Πιθανοφάνειας

Η συγκεκριμένη υλοποίηση είναι ένας συνδυασμός του ΕΦΚ και του ΕΜΠ, όπου το ΕΦΚ χρησιμοποιείται για των υπολογισμό των καταστάσεων x_k του συστήματος και ο ΕΜΠ για τον υπολογισμό των αγνώστων παραμέτρων. Για να μπορέσει το φίλτρο να προσεγγίσει τις μη γραμμικότητες της δυναμικής του συστήματος, δημιουργείται μια γραμμικοποίηση του μη γραμμικού μοντέλου κοντά στην τελευταία εκτίμηση. Όταν όλες οι εκτιμήσεις $\hat{X}_N = [\hat{x}_1, \dots, \hat{x}_n]$ της κατάστασης x_k γίνουν διαθέσιμες μπορεί να χρησιμοποιηθεί ο ΕΜΠ για να βρεθεί η εκτίμηση $\hat{\theta}_a$ του διανύσματος θ_a που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_{\theta_a}(\hat{X}_N)$ των εκτιμήσεων κατάστασης:

$$\hat{\theta}_a = \arg \max_{\theta_a \in \Theta_a} p_{\theta_a}(\hat{X}_N), \quad (3.15)$$

και μια εκτίμηση $\hat{\theta}_b$ του διανύσματος θ_b που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας των $p_{\theta_b}(Y_N)$ μετρήσεων:

$$\hat{\theta}_b = \arg \max_{\theta_b \in \Theta_b} p_{\theta_b}(Y_N). \quad (3.16)$$

Αν εφαρμοστεί ο κανόνας του Bayes, οι συναρτήσεις πιθανοφάνεια $p_{\theta_a}(\hat{X}_N)$ και $p_{\theta_b}(Y_N)$ μπορούν να γραφούν:

$$p_{\theta_a}(\hat{X}_N) = \prod_{k=1}^N p_{\theta_a}(\hat{x}_k | \hat{x}_{k-1}) \quad (3.17)$$

και

$$p_{\theta_b}(Y_N) = \prod_{k=2}^N p_{\theta_b}(y_k | \hat{x}_k), \quad (3.18)$$

όπου η $p_{\theta_a}(\hat{x}_k | \hat{x}_{k-1})$ προέρχεται από την (1.3), η $p_{\theta_b}(y_k | \hat{x}_k)$ από (1.4) και η αρχική κατάσταση \hat{x}_0 θεωρείται γνωστή.

Η υλοποίηση που χρησιμοποιήθηκε για το ΕΦΚ παρουσιάζεται στη συνέχεια. Για το σύστημα που περιγράφεται από τις εξισώσεις:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + w_k, \quad (3.19)$$

$$y_k = h(x_k) + v_k, \quad (3.20)$$

με

$$f(x_k, \theta_1, \theta_2) = \theta_2 \exp \left(\log \left(\frac{x_k}{\theta_2} \right) \exp \left(-\frac{1}{\theta_1} t_{k+1} \right) \right) \quad (3.21)$$

και

$$h(x_k) = x_k, \quad (3.22)$$

όπου $w_k \sim \mathcal{N}(0, q_k^2)$, $q_k = \theta_3 x_k^{\theta_4}$ και $v_k \sim \mathcal{N}(0, r_k^2)$, $r_k = \theta_5 x_k^{\theta_6}$, έχουμε:

$$x_{k+1} = f(\hat{x}_k) + \frac{\partial f}{\partial x_k} \Big|_{\hat{x}_k} (x_k - \hat{x}_k) + w_k \quad (3.23)$$

και

$$y_{k+1} = h(\hat{x}_k) + \frac{\partial h}{\partial x_k} \Big|_{\hat{x}_k} (x_k - \hat{x}_k) + v_k. \quad (3.24)$$

Συνεπώς, το βήμα πρόβλεψης είναι:

$$\hat{x}_{k+1}^- = f(\hat{x}_k), \quad (3.25)$$

$$P_{x_{k+1}}^- = \frac{\partial f}{\partial x_k} \Big|_{\hat{x}_k} P_{x_k} \frac{\partial f}{\partial x_k}^T \Big|_{\hat{x}_k} + q_k, \quad (3.26)$$

ενώ το βήμα διόρθωσης είναι:

$$\hat{y}_{k+1}^- = h(\hat{x}_{k+1}^-), \quad (3.27)$$

$$P_{y_{k+1}}^- = \frac{\partial h}{\partial x_k} \Big|_{\hat{x}_{k+1}^-} P_{x_{k+1}}^- \frac{\partial h}{\partial x_k}^T \Big|_{\hat{x}_{k+1}^-} + r_k, \quad (3.28)$$

$$P_{x_{k+1}y_{k+1}}^- = P_{x_{k+1}}^- \frac{\partial h}{\partial x_k}^T \Big|_{\hat{x}_{k+1}^-}, \quad (3.29)$$

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1} (y_{k+1} - \hat{y}_{k+1}^-), \quad (3.30)$$

$$P_{x_{k+1}} = P_{x_{k+1}}^- - K_{k+1} P_{y_{k+1}}^- K_{k+1}^T, \quad (3.31)$$

$$K_{k+1} = P_{x_{k+1}y_{k+1}}^- \left(P_{y_{k+1}}^- \right)^{-1}. \quad (3.32)$$

Επιπλέον, παρατηρήθηκε ότι όταν η διαδικασία που περιγράφηκε παραπάνω επαναλαμβάνεται, η εκτίμηση $\hat{\theta}$ συγκλίνει σε κάποια τιμή η οποία είναι καλύτερη εκτίμηση του διανύσματος θ σε σχέση με την πρώτη εκτίμηση. Η συμπεριφορά αυτή παρατηρείται διότι αρχικά το διάνυσμα $\hat{\theta}$ θεωρείται γνωστό για το ΕΦΚ, με αρχική τιμή την εκτίμηση που προκύπτει από την εφαρμογή της μεθόδου της Μέγιστης Πιθανοφάνειας στο απλοποιημένο μοντέλο. Επομένως, η εκτίμηση της κατάστασης γίνεται με βάση αυτό το διάνυσμα. Όμως, όταν οι εκτιμήσεις κατάστασης που προκύπτουν χρησιμοποιηθούν για

να μεγιστοποιήσουν τις συναρτήσεις πιθανοφάνειας των εξισώσεων (3.17) και (3.18) το νέο διάνυσμα $\hat{\theta}$ είναι καλύτερη εκτίμηση του διανύσματος θ από τις αρχικές συνθήκες που χρησιμοποιήθηκαν και επομένως όταν αυτό το διάνυσμα χρησιμοποιηθεί σαν αρχική συνθήκη στο ΕΦΚ, η εκτίμηση κατάστασης \hat{x}_k βελτιώνεται κ.ο.κ. Η διαδικασία αυτή επαναλαμβάνεται έως ότου να παρατηρηθεί ότι δύο διαδοχικές εκτιμήσεις δεν παρουσιάζουν σημαντική απόκλιση ή αφού συμπληρωθεί κάποιος προκαθορισμένος αριθμός επαναλήψεων.

3.6 Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας

Στην εργασία αυτή υλοποιήθηκε ένα ΦΔΕΣ το οποίο χρησιμοποιήθηκε για την εκτίμηση των καταστάσεων x_k του συστήματος. Ο αλγόριθμος 3 παρουσιάζει αναλυτικά την υλοποίηση που χρησιμοποιήθηκε. Το διάνυσμα θ με τις άγνωστες παραμέτρους υπολογίστηκε με τον ΕΜΠ που περιγράφεται στο υποκεφάλαιο 3.3. Επιπλέον σημειώνεται ότι η συνάρτηση `find(rand <= cumsum(P_r))` στην γραμμή 20 του αλγόριθμου 3 επιστρέφει τον πρώτο δείκτη m του διανύσματος P_r , έτσι ώστε να ισχύει ότι:

$$\text{rand} \leq \sum_{i=1}^m P_r[i], \quad (3.33)$$

όπου `rand` ένας ομοιόμορφα κατανεμημένος τυχαίος αριθμός που ανήκει στο διάστημα $(0, 1)$.

3.7 Παράλληλο Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας

Η υλοποίηση του παράλληλου ΦΔΕΣ έγινε στη γλώσσα προγραμματισμού CUDA. Η παραλληλοποίηση των βημάτων ενημέρωσης της κατάστασης και της μέτρησης, καθώς επίσης και του βήματος υπολογισμού των βαρών είναι άμεση, διότι οι διεργασίες που πρέπει να γίνουν σε κάθε δείγμα είναι ανεξάρτητες. Αντίθετα, η παραλληλοποίηση του βήματος επαναδειγματοληψίας για το ΦΔΕΣ είναι ελάχιστα πιο σύνθετη, μιας και απαιτείται μια συγκεντρωτική διεργασία, και πιο συγκεκριμένα ο υπολογισμός του αθροίσματος του διανύσματος των βαρών P_r . Για να γίνει αυτή η διεργασία χρησιμοποιήθηκε ο αλγόριθμος παράλληλης μείωσης διαδοχικής διευθυνσιοδότησης (Parallel Reduction: Sequential

Algorithm 3 Φίλτρο Διαδοχικής Επαναδειγματοληψίας Σπουδαιότητας - ΦΔΕΣ

```

1: procedure SIR( $Y_N, N_s, \theta$ ) ▷  $N_s$ : Number of samples
2:
3:   Initialize:  $\hat{x}_0 = x_0$ 
4:
5:   for  $k = 1; k \leq N; k++$  do
6:
7:     for  $j = 1; j \leq N_s; j++$  do
8:        $x_P[j] = f(\hat{x}_{k-1}, \theta_1, \theta_2) + w_j$ 
9:        $y_P[j] = x_P[j] + v_j$ 
10:       $\sigma = \theta_5(x_P[j])^{\theta_6}$ 
11:       $P_r[j] = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(-\frac{(y_k - y_P[j])^2}{2\sigma^2}\right)$ 
12:    end for
13:
14:    Normalize to form a probability distribution:
15:     $P_r = \frac{P_r}{\text{sum}(P_r)}$ 
16:
17:    Resample:
18:
19:    for  $j = 1; j \leq N_s; j++$  do
20:       $x_P[j] = x_P[\text{find}(\text{rand} \leq \text{cumsum}(P_r))]$ 
21:    end for
22:
23:    The final estimate the mean value of the estimates  $\hat{x}_k = \text{mean}(x_P)$ 
24:
25:  end for
26:
27: end procedure

```

Addressing) [74], με αποτέλεσμα να μειωθεί η πολυπλοκότητα υπολογισμού του αθροίσματος από $O(N)$ σε $O(\log(N))$. Τέλος, η παραλληλοποίηση του τελευταίου βήματος της διαδικασίας της επαναδειγματοληψίας είναι επίσης άμεση, μιας και αποτελείται από ανεξάρτητες διεργασίες.

Για την παραγωγή των καινικά κατανεμημένων τυχαίων αριθμών που χρειάστηκαν στο βήμα ενημέρωσης και των ομοίμορφα κατανεμημένων τυχαίων αριθμών που χρειάστηκαν στο βήμα επαναδειγματοληψίας, χρησιμοποιήθηκε η βιβλιοθήκη `curand_kernel` της CUDA [75], και πιο συγκεκριμένα έγινε χρήση των συναρτήσεων `curand_normal` και `curand_uniform`.

`!TEX root = ../thesis.tex`

Κεφάλαιο 4

Προσομοιώσεις και Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά οι προσομοιώσεις που έγιναν στο πρόγραμμα Matlab με τη χρήση συνθετικών αλλά και πειραματικών δεδομένων. Τα συνθετικά δεδομένα συνέβαλαν στο να ελεγχθεί η ικανότητα των μεθόδων να εκτιμούν τις άγνωστες παραμέτρους του μοντέλου που προτείνεται για τη μοντελοποίηση της ανάπτυξης των καρκινικών όγκων. Αφού επαληθεύτηκε ότι οι τεχνικές που υλοποιήθηκαν μπορούν να εκτιμήσουν τις άγνωστες παραμέτρους του μοντέλου, έγινε εφαρμογή σε πειραματικά δεδομένα με σκοπό την πρόβλεψη της μελλοντικής πορείας της εξέλιξης των όγκων σε σύγκριση με τις προβλέψεις που προέκυψαν από την υλοποίηση της μεθοδολογίας που προτείνεται συχνά στη βιβλιογραφία. Στο πρώτο μέρος αυτού του κεφαλαίου (υποκεφάλαιο 4.1), περιγράφεται ο τρόπος με τον οποίο δημιουργήθηκαν τα συνθετικά δεδομένα και οι περιπτώσεις που εξετάστηκαν, οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση των μεθόδων, καθώς επίσης και τα αποτελέσματα που προέκυψαν. Στο δεύτερο μέρος (υποκεφάλαιο 4.2), περιγράφονται τα πειραματικά δεδομένα και οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση των μεθόδων, καθώς επίσης και τα αποτελέσματα της ανάλυσης των δεδομένων.

4.1 Προσομοιώσεις με συνθετικά δεδομένα

Για να δημιουργηθεί το σύνολο των συνθετικών δεδομένων, προσομοιώθηκε η ανάπτυξη 100 καρκινικών όγκων με τη χρήση των εξισώσεων (1.3) και (1.4). Ο ρυθμός αύξησης (θ_1^j) και η φέρουσα χωρητικότητα (θ_2^j) για τον κάθε όγκο, καθώς επίσης και τα χαρακτηριστικά του θορύβου της διαδικασίας (θ_3^j, θ_4^j) και της μέτρησης (θ_5^j, θ_6^j), όπου $j = 1, \dots, 100$ είναι ο αύξων αριθμός του υποκειμένου ελέγχου, επιλέχθηκαν τυχαία από ομοιόμορφες κατανομές. Η ελάχιστη και η μέγιστη τιμή για την κάθε παράμετρο παρουσιάζονται στον Πίνακα 4.1 και επιλέχθηκαν ώστε τα δεδομένα που προκύπτουν να είναι

Παράμετροι	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
Ελάχιστο	6	320	1	0.3	2	0.2
Μέγιστο	16	800	4	0.5	6	0.5

Πίνακας 4.1 Ελάχιστη και μέγιστη τιμή για κάθε παράμετρο

παρόμοια με αυτά από πειραματικές μετρήσεις όγκων. Ο χρόνος δειγματοληψίας μεταξύ δύο διαδοχικών μετρήσεων είναι $T = 2$ (μέρες) και ο αριθμός των διαθέσιμων μετρήσεων είναι $N = 30$.

4.1.1 Μετρικές

Για να αξιολογηθεί η αποτελεσματικότητα των μεθόδων που περιγράφηκαν στο Κεφάλαιο 3 να εκτιμούν τις παραμέτρους του μοντέλου Gompertz και των χαρακτηριστικών του θορύβου στα συνθετικά δεδομένα, χρησιμοποιήθηκαν το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error) και η τυπική απόκλιση του Απόλυτου Ποσοστιαίου Σφάλματος (Absolute Percentage Error). Το ΜΑΠΣ δίνεται από την εξίσωση:

$$\text{ΜΑΠΣ} = \frac{1}{j} \sum_{i=1}^j \text{ΑΠΣ}_i(\%), \quad (4.1)$$

και η τ.α. του ΑΠΣ την εξίσωση:

$$\text{τ.α.} = \left(\frac{1}{j-1} \sum_{i=1}^j |\text{ΑΠΣ}_i - \text{ΜΑΠΣ}|^2 \right)^{1/2}. \quad (4.2)$$

Στις εξισώσεις (4.1) και (4.2)

$$\text{ΑΠΣ}_i = \left| \frac{A_i - E_i}{A_i} \right|, \quad (4.3)$$

όπου A_i είναι η πραγματική τιμή της παραμέτρου, E_i είναι η τιμή που δίνει ο εκτιμητής και $j (= 100)$ είναι ο συνολικός αριθμός όγκων που ελέγχονται. Για την αξιολόγηση της ικανότητας του ΦΔΕΣ να εκτιμά τις άγνωστες καταστάσεις του συστήματος γίνεται χρήση του Κανονικοποιημένου Μέσου Τετραγωνικού Σφάλματος (Normalized Mean Squared Error). Το ΚΜΤΣ δίνεται από την εξίσωση:

$$\text{ΚΜΤΣ} = \frac{1}{k} \sum_{i=1}^k \frac{(x_i - \hat{x}_i)^2}{\bar{x} \bar{x}_{est}}, \quad (4.4)$$

όπου k είναι ο αριθμός των διαθέσιμων μετρήσεων, x_i είναι η πραγματική τιμή της κατάστασης του συστήματος (η οποία είναι γνωστή για τα συνθετικά δεδομένα) και \hat{x}_i είναι η εκτίμηση για την κατάσταση x_i που δίνει το ΦΔΕΣ. Οι τιμές \bar{x} και \bar{x}_{est} δίνονται από τις εξισώσεις:

$$\bar{x}_i = \frac{1}{k} \sum_{i=1}^k x_i, \quad (4.5)$$

και

$$\bar{x}_{est} = \frac{1}{k} \sum_{i=1}^k \hat{x}_i. \quad (4.6)$$

4.1.2 Περιπτώσεις

Στην παρούσα εργασία ελέγχονται πέντε περιπτώσεις. Στις τέσσερις από αυτές σκοπός είναι να διαπιστωθεί αν οι μέθοδοι που υλοποιήθηκαν μπορούν να εκτιμήσουν τις άγνωστες παραμέτρους του μοντέλου στα συνθετικά δεδομένα, ενώ στην πέμπτη γίνεται προσπάθεια να εκτιμηθούν οι καταστάσεις του συστήματος. Σε όλες τις περιπτώσεις που εξετάζονται, ο ρυθμός αύξησης και η φέρουσα χωρητικότητα θεωρούνται άγνωστες. Στις τέσσερις πρώτες περιπτώσεις, η διαφοροποίηση γίνεται με βάση την διαθέσιμη πληροφορία όσον αφορά τον θόρυβο της διαδικασίας και τον θόρυβο της μέτρησης. Οι υλοποιήσεις που εφαρμόστηκαν στα συνθετικά δεδομένα είναι ο ΕΜΠ στο απλοποιημένο μοντέλο, ο ΕΜΠ με αριθμητική ολοκλήρωση, και ο συνδιαμός του ΕΦΚ με την αρχή της Μέγιστης Πιθανοφάνειας. Στην πέμπτη περίπτωση, όπου γίνεται προσπάθεια για εκτίμηση της κατάστασης του συστήματος, χρησιμοποιείται η εκτίμηση για τις άγνωστες παραμέτρους που προκύπτει από την περίπτωση 4 και με βάση αυτή γίνεται εκτίμηση της κατάστασης του συστήματος.

Περίπτωση 1 - Γνωστά χαρακτηριστικά θορύβου

Στην περίπτωση αυτή, οι άγνωστες παράμετροι είναι ο ρυθμός αύξησης (θ_1) και η φέρουσα χωρητικότητα (θ_2), ενώ τα χαρακτηριστικά του θορύβου διαδικασίας (θ_3, θ_4) και του θορύβου μέτρησης (θ_5, θ_6) θεωρούνται γνωστά. Οι παράμετροι που σχετίζονται με τον θόρυβο φιξάρονται στην πραγματική τους τιμή και το όρισμα που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_\theta(Y_n)$ είναι το $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2]^T$.

Περίπτωση 2 - Γνωστά χαρακτηριστικά θορύβου διαδικασίας

Στην περίπτωση αυτή, οι άγνωστες παράμετροι είναι οι $\theta_1, \theta_2, \theta_5$ και θ_6 , ενώ οι παράμετροι θ_3 και θ_4 θεωρούνται γνωστές. Οι παράμετροι θ_3, θ_4 φιξάρονται στην πραγματική

τους τιμή και το όρισμα που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_{\theta}(\mathbf{Y}_n)$ είναι το $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2; \hat{\theta}_3; \hat{\theta}_6]^T$.

Περίπτωση 3 - Γνωστά χαρακτηριστικά θορύβου μέτρησης

Στην περίπτωση αυτή, οι άγνωστες παράμετροι είναι οι $\theta_1, \theta_2, \theta_3$ και θ_4 , ενώ οι παράμετροι θ_5 και θ_6 θεωρούνται γνωστές. Οι παράμετροι θ_5, θ_6 φιξάρονται στην πραγματική τους τιμή και το όρισμα που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $p_{\theta}(\mathbf{Y}_n)$ είναι το $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2; \hat{\theta}_3; \hat{\theta}_4]^T$.

Περίπτωση 4 - Άγνωστα χαρακτηριστικά θορύβου

Στην περίπτωση αυτή, όλες οι παράμετροι ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ και θ_6) είναι άγνωστες και οι τεχνικές χρησιμοποιούνται όπως περιγράφηκαν στο Κεφάλαιο 3.

Περίπτωση 5 - Εκτίμηση Κατάστασης Συστήματος

Αρχικά, στην περίπτωση αυτή, οι παράμετροι ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ και θ_6) είναι άγνωστες και εκτιμώνται από τον ΕΜΠ με αριθμητική ολοκλήρωση (υποκεφάλαιο 3.3). Στη συνέχεια, οι εκτιμήσεις αυτές χρησιμοποιούνται από το ΦΔΕΣ (στην ακολουθιακή και στην παράλληλη υλοποίηση του), για να γίνει εκτίμηση των καταστάσεων του συστήματος.

4.1.3 Υλοποίηση

Οι προσομοιώσεις που αναφέρθηκαν στο κεφάλαιο αυτό και οι μέθοδοι που περιγράφηκαν στο προηγούμενο κεφάλαιο υλοποιήθηκαν στο πρόγραμμα Matlab (ver. R2015a, The MathWorks Inc., Natick, IL) με τη δημιουργία κατάλληλων συναρτήσεων. Οι προσομοιώσεις έγιναν με τη χρήση ενός επεξεργαστή Intel Core i7-6700K @ 4.00GHz και μνήμης 16GB DDR4 @ 3200 MHz. Για να μεγιστοποιηθούν οι συναρτήσεις πιθανοφάνειας (3.5), (3.7), (3.15) και (3.16) χρησιμοποιήθηκε η συνάρτηση `fmincon` [76]. Πιο συγκεκριμένα, η συνάρτηση `fmincon` χρησιμοποιήθηκε για να βρεθεί το όρισμα $\hat{\theta}$ που ελαχιστοποιεί τον αρνητικό λογάριθμο της συνάρτησης πιθανοφάνειας $p_{\theta}(\mathbf{D})$:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left(-\log (p_{\theta}(\mathbf{D})) \right), \quad (4.7)$$

όπου με \mathbf{D} συμβολίζεται το διάνυσμα δεδομένων \mathbf{Y}_N ή $\hat{\mathbf{X}}_N$ ανάλογα με τη μέθοδο που χρησιμοποιείται και τις παραμέτρους που εκτιμώνται. Η μεγιστοποίηση του λογαρίθμου της συνάρτησης πιθανοφάνειας είναι ισοδύναμη με τη μεγιστοποίηση της πιθανοφάνειας,

μιας και ο λογάριθμος είναι μια μονότονη συνάρτηση. Ο παράλληλος κώδικας σε CUDA υλοποιήθηκε επίσης στο πρόγραμμα Matlab [73].

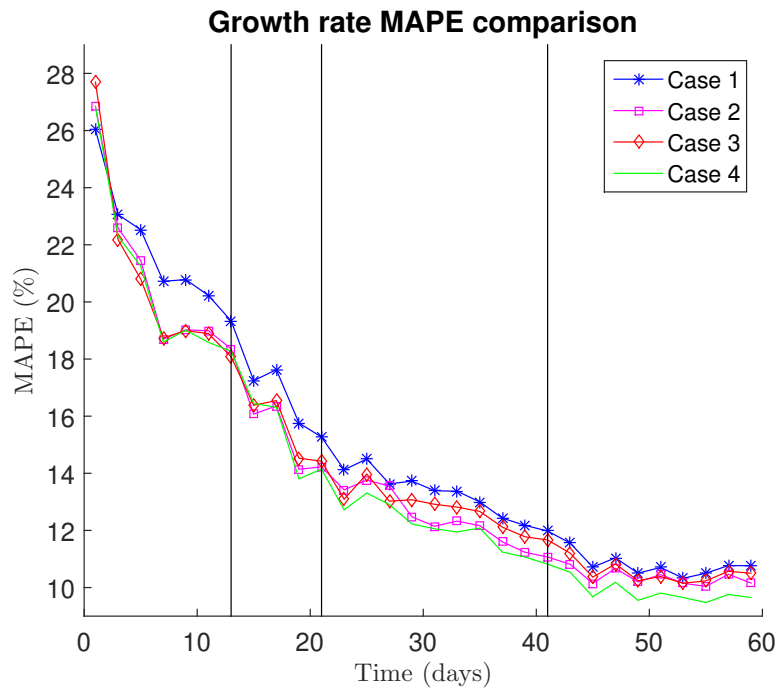
4.1.4 Αποτελέσματα

Στο σχήμα 4.1 παρουσιάζονται το ΜΑΠΣ και η τ.α. του ΑΠΣ για τον ρυθμό αύξησης και τη φέρουσα χωρητικότητα σε κάθε βήμα επανάληψης (δηλ. κάθε φορά που μια καινούργια μέτρηση του όγκου γίνεται διαθέσιμη) για κάθε μια από τις εξεταζόμενες περιπτώσεις. Η σύγκριση μεταξύ των διάφορων υλοποιήσεων που περιγράφηκαν στο Κεφάλαιο 3 γίνεται μόνο για την τέταρτη περίπτωση, όμως όμοια αποτελέσματα και συμπεράσματα προκύπτουν και κατά τη σύγκριση των μεθόδων όταν εφαρμοστούν στις άλλες περιπτώσεις. Στα σχήματα 4.2α' και 4.2β' παρουσιάζεται το ΜΑΠΣ για τον ρυθμό αύξησης και την φέρουσα χωρητικότητα αντίστοιχα, ενώ στο σχήμα 4.2γ' παρουσιάζονται τα ΜΑΠΣ για τις παραμέτρους του θορύβου. Στο σχήμα 4.3 παρουσιάζεται η τυπική απόκλιση του ΑΠΣ για τις παραπάνω παραμέτρους για κάθε μέθοδο που χρησιμοποιήθηκε. Οι τρεις κάθετες γραμμές δείχνουν το μέσο χρονικό διάστημα που χρειάζονται οι καρκινικοί όγκοι για να φτάσουν το 50% (περίπου 7 μετρήσεις -- 14 μέρες), το 75% (περίπου 11 μετρήσεις -- 22 μέρες) και το 100% (περίπου 21 μετρήσεις -- 42 μέρες) της φέρουσας χωρητικότητας τους.

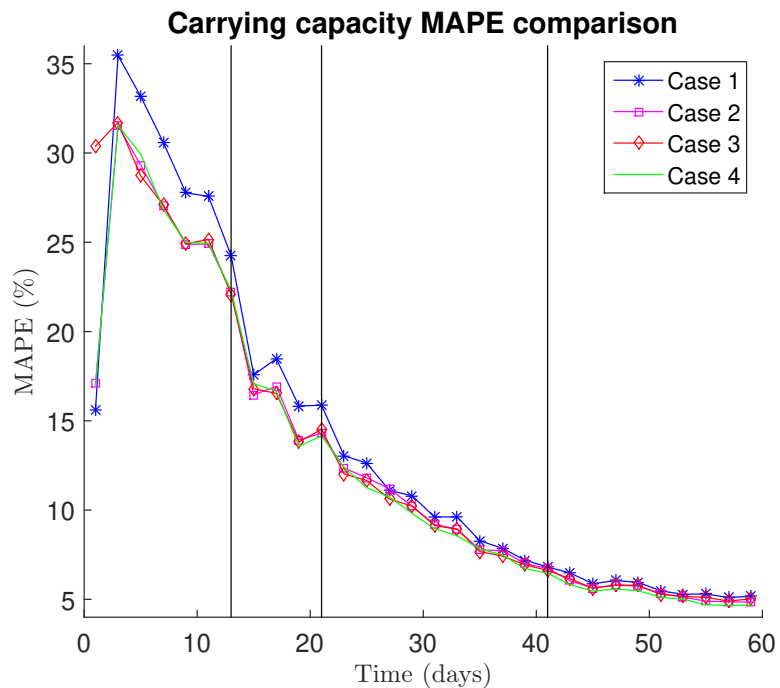
Στο σχήμα 4.4 φαίνεται η πρόοδος της ανάπτυξης για δύο τυχαία επιλεγμένα υποκείμενα ελέγχου, και η ικανότητα του μοντέλου Gompertz να δημιουργεί καμπύλες που να ταιριάζουν στα δεδομένα με βάση τις τιμές που έχουν εκτιμηθεί για τις άγνωστες παραμέτρους θ_1 και θ_2 . Οι καμπύλες αυτές δημιουργήθηκαν από εκτιμήσεις των αγνώστων παραμέτρων που προέκυψαν σε διαφορετικά χρονικά σημεία, δηλαδή υπολογίστηκαν για διαφορετικό αριθμό διαθέσιμων μετρήσεων. Οι τιμές των παραμέτρων που χρησιμοποιούνται για τις συγκεκριμένες απεικονίσεις είναι αυτές που υπολογίστηκαν με τη χρήση του ΕΜΠ με αριθμητική ολοκλήρωση, από την οποία προέκυψαν τα καλύτερα αποτελέσματα εκτίμησης σε σχέση με τις άλλες δύο μεθόδους. Στα σχήματα 4.4α' και 4.4β' η κόκκινη καμπύλη είναι αυτή που δημιουργήθηκε με τις παραμέτρους που υπολογίστηκαν από το πλήρες σετ μετρήσεων. Με πράσινο χρώμα απεικονίζεται η καμπύλη στην οποία οι παράμετροι επανυπολογίζονται κάθε φορά που μια νέα μέτρηση γίνεται διαθέσιμη, δηλαδή πρόκειται ουσιαστικά για πρόβλεψη ενός βήματος μπροστά. Για την μπλε καμπύλη, για την εκτίμηση λαμβάνονται υπόψη μόνο οι τρεις αρχικές μετρήσεις. Παρόμοια, για την μωβ, τη γαλάζια και την μαύρη καμπύλη, οι μετρήσεις που είναι διαθέσιμες έως ότου ο όγκος φτάσει στο 50%, 75% και 100% της φέρουσας χωρητικότητας αντίστοιχα, χρησιμοποιούνται για να παραχθεί η εκτίμηση για τις άγνωστες παραμέτρους και να προβλεφθεί η μελλοντική εξέλιξη της ανάπτυξης του όγκου.

Στο σχήμα 4.5 παρουσιάζεται η μέση τιμή του ΚΜΤΣ που προέκυψε από την εξίσωση (4.4) για τις εκτιμήσεις κατάστασης από ΦΔΕΣ, ενώ στο σχήμα 4.6 φαίνονται οι εκτιμήσεις των καταστάσεων του συστήματος που έδωσε το ΦΔΕΣ για ένα τυχαία επιλεγμένο υποκείμενο μελέτης σε τέσσερις διαφορετικές χρονικές στιγμές. Τέλος, στο σχήμα 4.7 παρουσιάζονται ο μέσος χρόνος εκτέλεσης του ΦΔΕΣ ανάλογα με τον αριθμό των διαθέσιμων μετρήσεων και ο χρόνος εκτέλεσης της ακολουθιακής και της παράλληλης υλοποίησης του ΦΔΕΣ ανάλογα με τον αριθμό των δειγμάτων (σωματιδίων) που χρησιμοποιήθηκαν.

Σχήμα 4.1 ΜΑΠΣ και τ.α. των παραμέτρων θ_1 και θ_2 για τον ΕΜΠ με αριθμητική ολοκλήρωση.

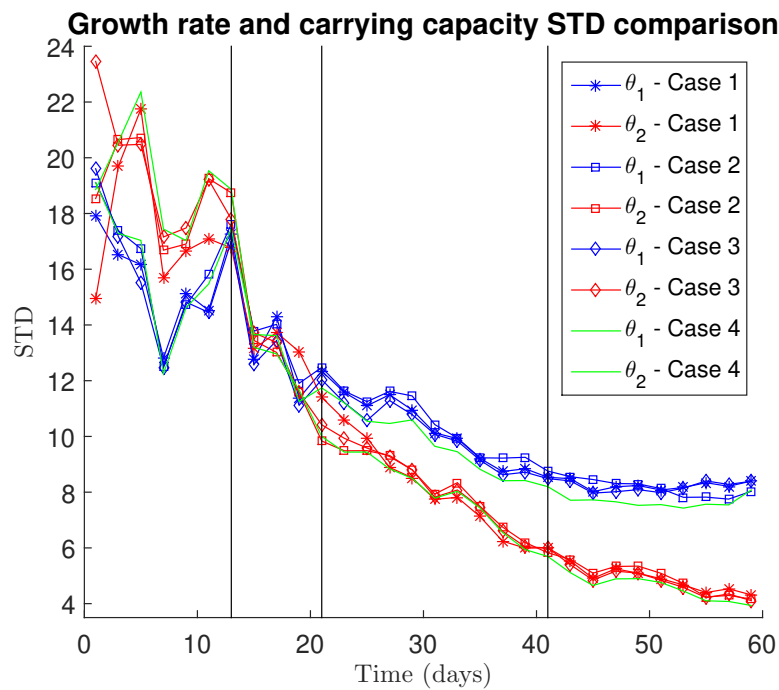


(α') ΜΑΠΣ του ρυθμού αύξησης (παράμετρος θ_1).

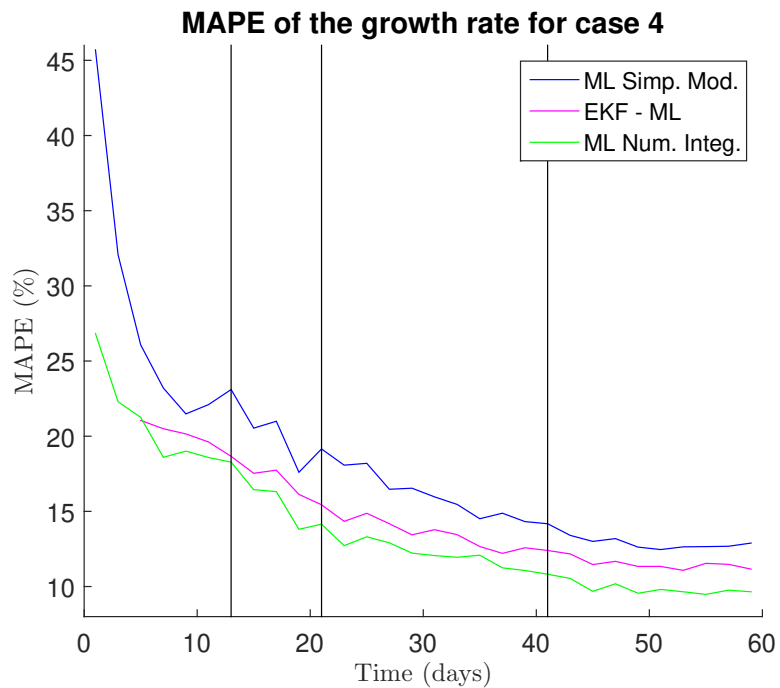


(β') ΜΑΠΣ της φέρουσας χωρητικότητας (παράμετρος θ_2).

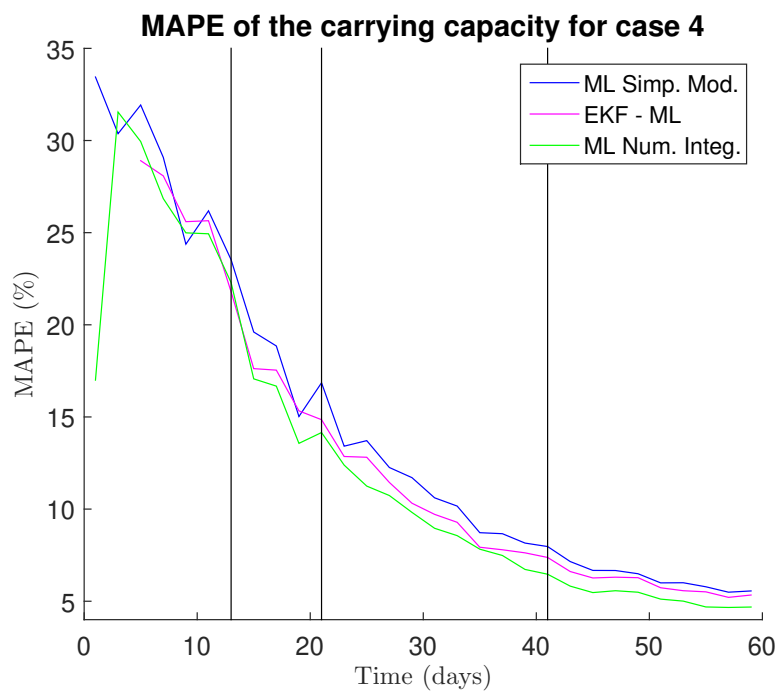
Σχήμα 4.1 ΜΑΠΣ και τ.α. των παραμέτρων θ_1 και θ_2 για τον ΕΜΠ με αριθμητική ολοκλήρωση.



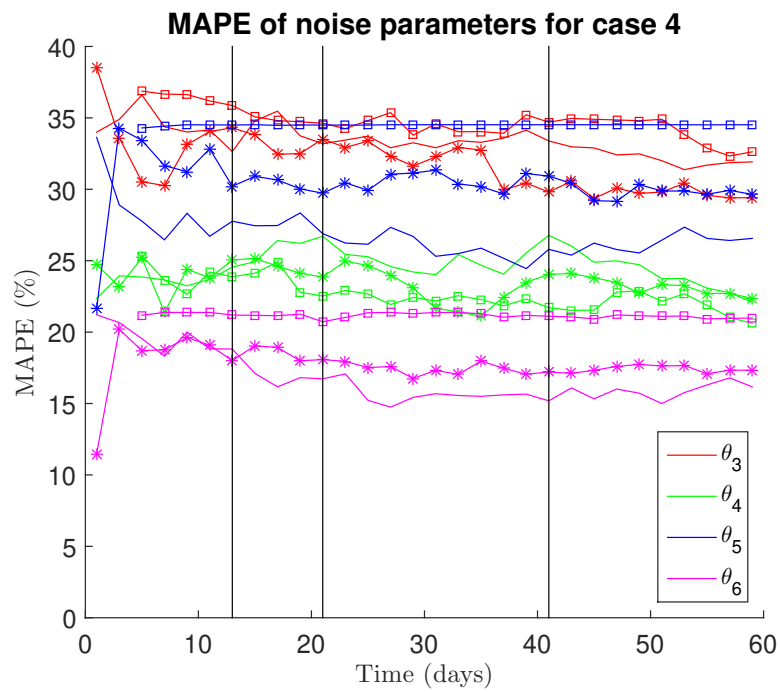
(γ) Τυπική απόκλιση του ΑΠΣ για τον ρυθμό αύξησης και την φέρουσα χωρητικότητα.

Σχήμα 4.2 ΜΑΠΣ για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.

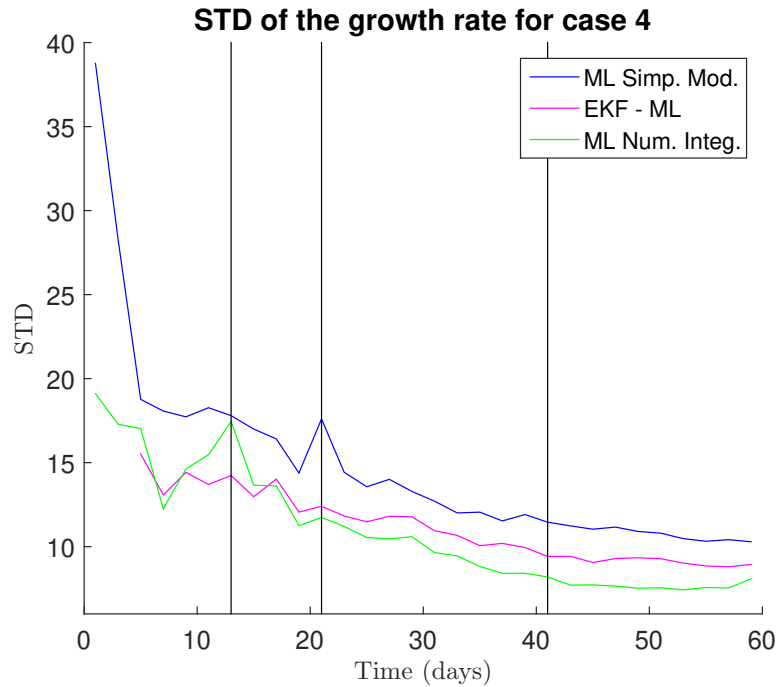
(α') ΜΑΠΣ του ρυθμού αύξησης (παράμετρος θ_1).



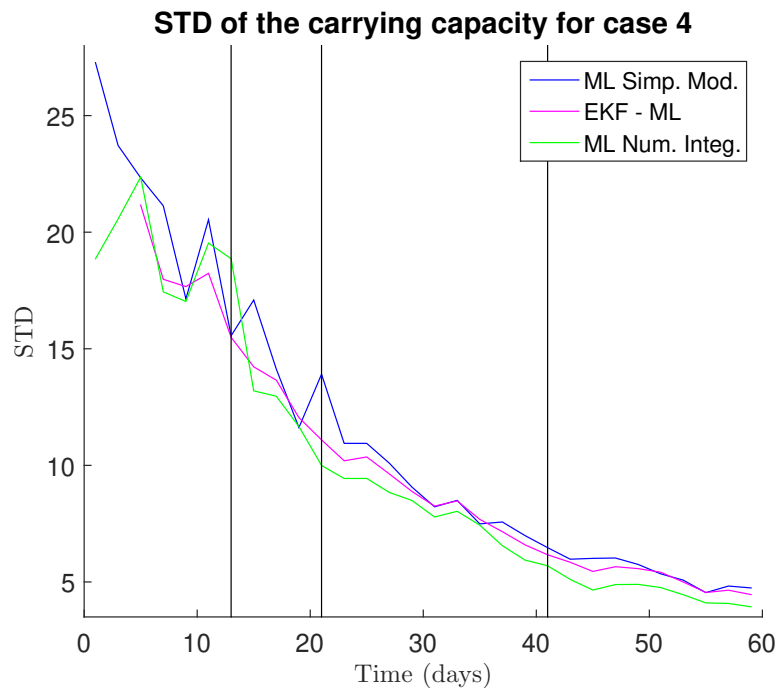
(β') ΜΑΠΣ της φέρουσας χωρητικότητας (παράμετρος θ_2).

Σχήμα 4.2 ΜΑΠΣ για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.

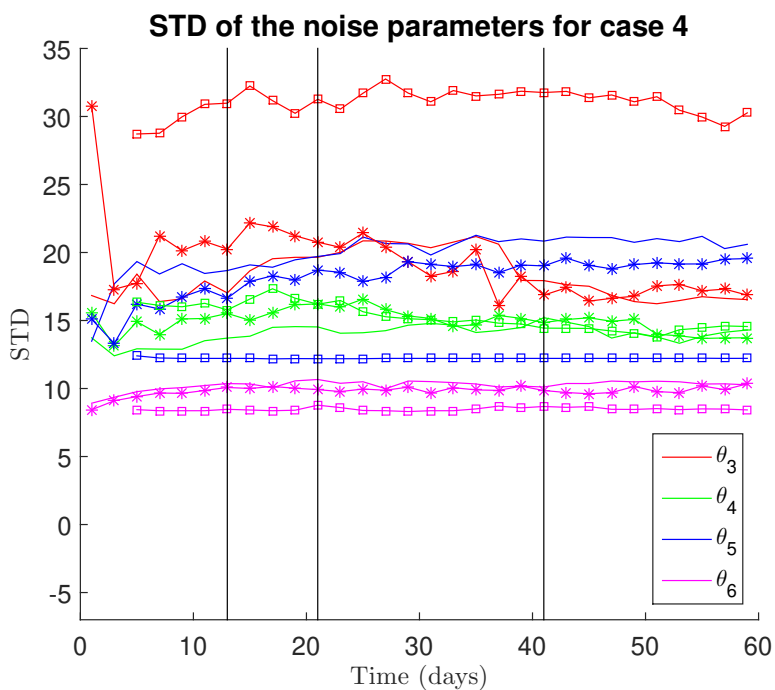
(γ) ΜΑΠΣ των χαρακτηριστικών του θορύβου (παραμέτρους θ_3 -- θ_6).

Σχήμα 4.3 Τυπική απόκλιση για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.

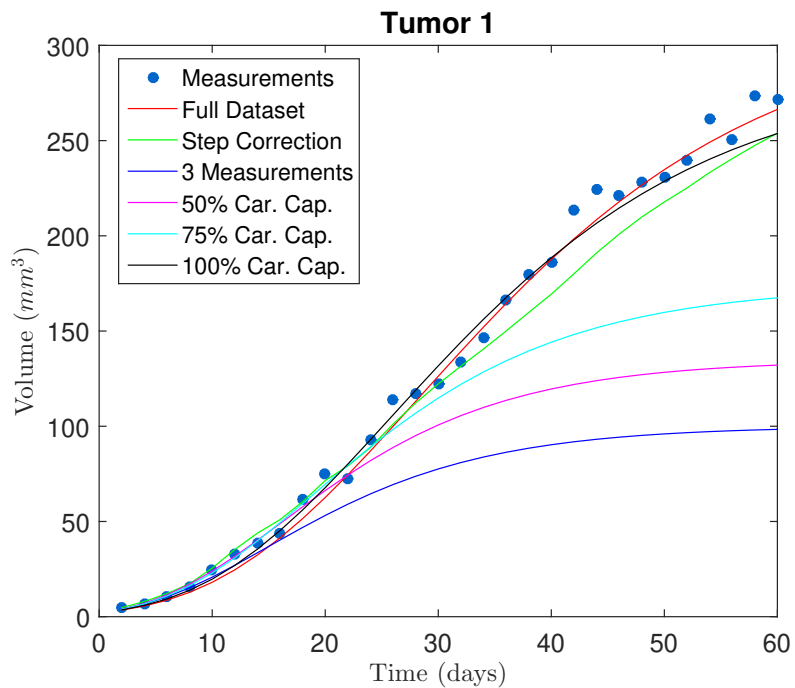
(α') Τυπική απόκλιση του ρυθμού αύξησης (παράμετρος θ_1).



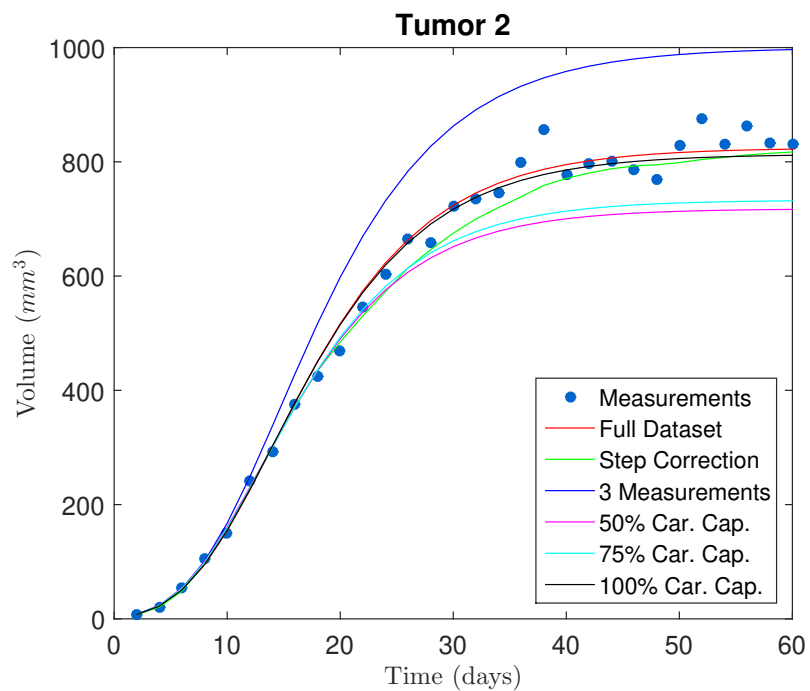
(β') Τυπική απόκλιση της φέρουσας χωρητικότητας (παράμετρος θ_2).

Σχήμα 4.3 Τυπική απόκλιση για όλες τις άγνωστες παραμέτρους στην περίπτωση 4.

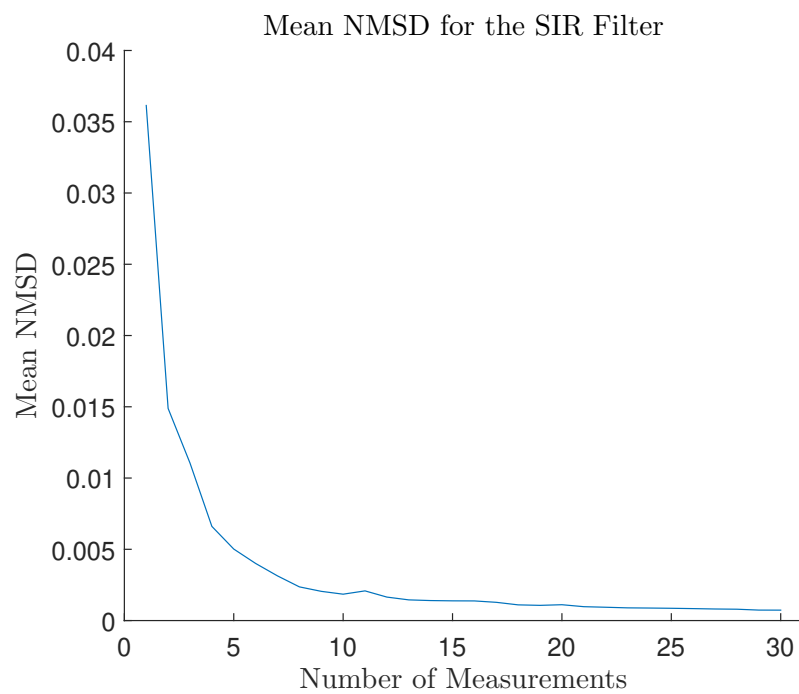
(γ) Τυπική απόκλιση των χαρακτηριστικών του θορύβου (παραμέτροι θ_3 -- θ_6).

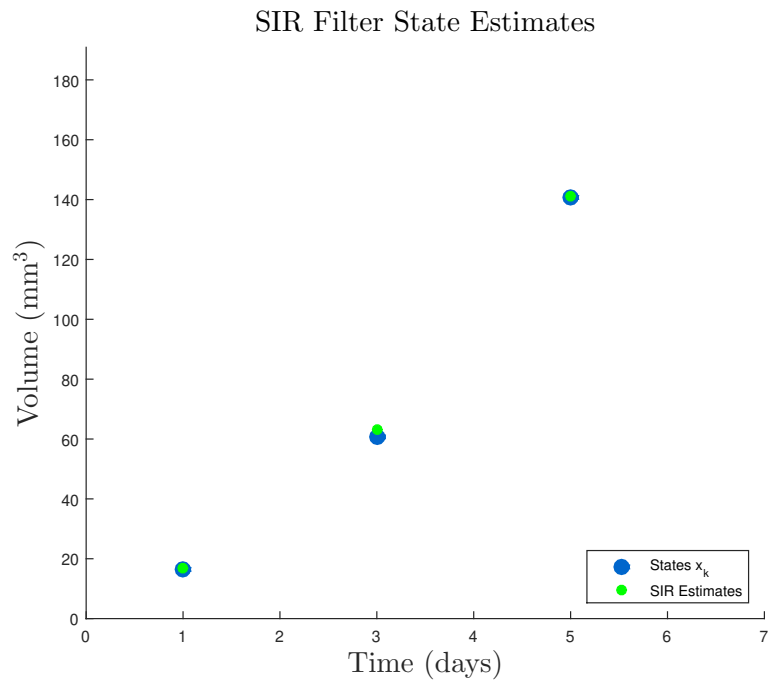
Σχήμα 4.4 Καμπύλες ανάπτυξης για τυχαία επιλεγμένα αντικείμενα ελέγχου.

(α') Υποκείμενο ελέγχου 1.

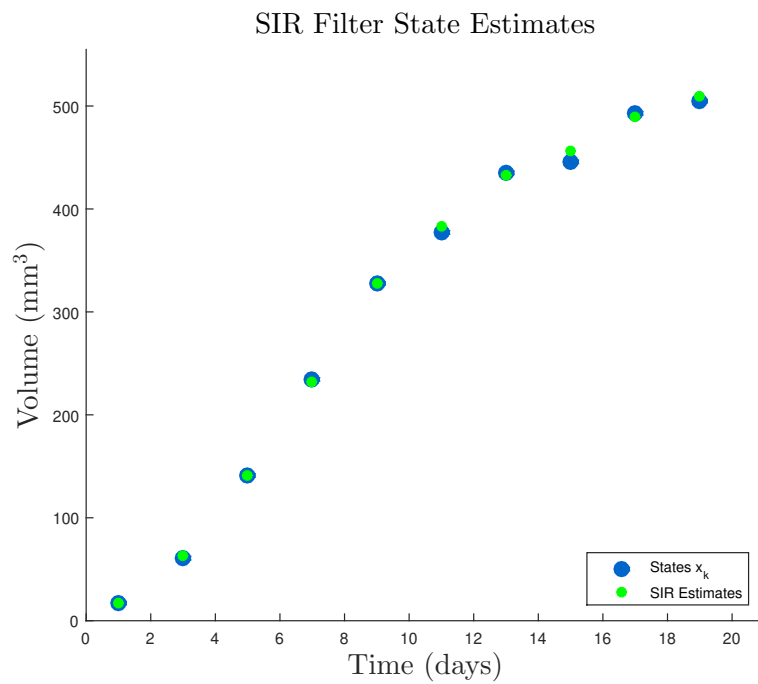


(β') Υποκείμενο ελέγχου 2.

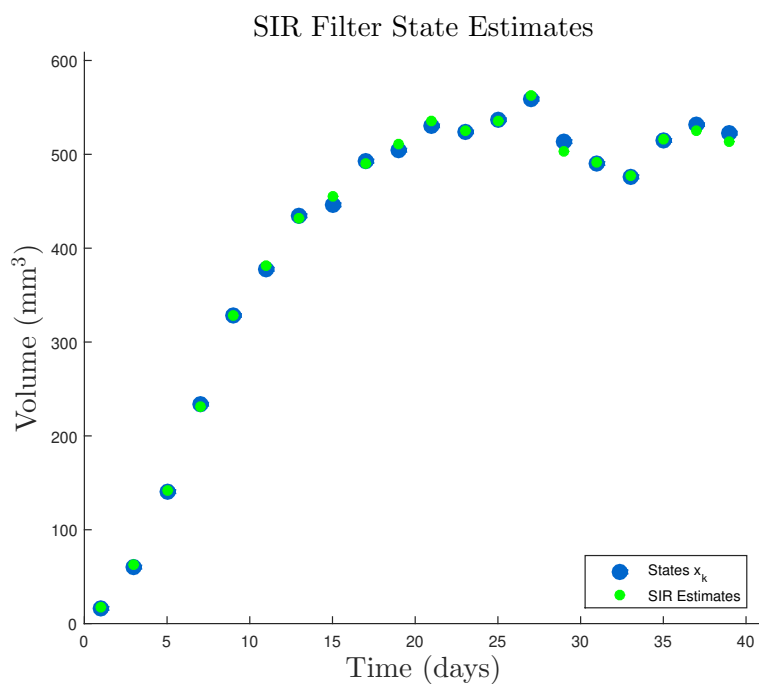
Σχήμα 4.5 Μέση τιμή Κανονικοποιημένου Μέσου Τετραγωνικού Σφάλματος.

Σχήμα 4.6 Εκτίμηση καταστάσεων συστήματος με το ΦΔΕΣ.

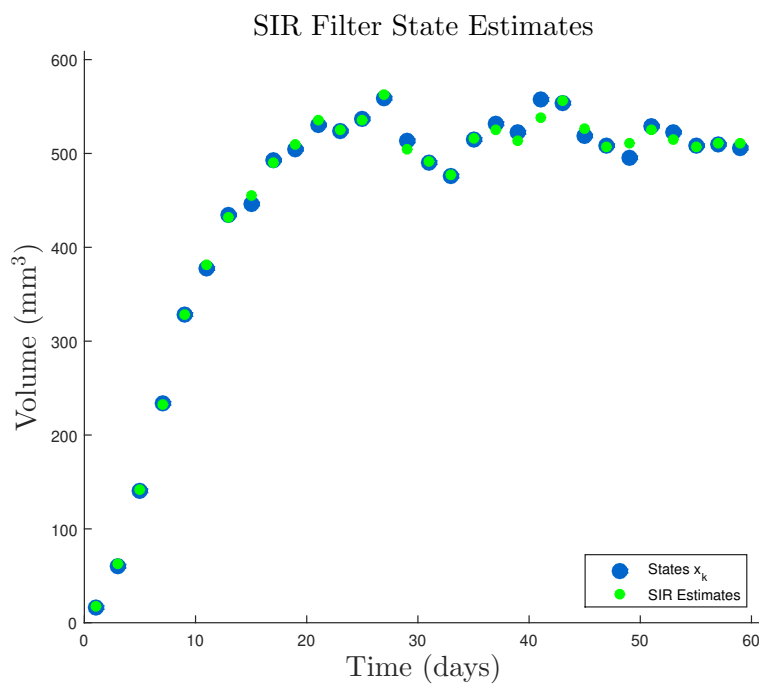
(α') Εκτίμηση καταστάσεων με τις 3 πρώτες μετρήσεις.



(β') Εκτίμηση καταστάσεων με τις 10 πρώτες μετρήσεις.

Σχήμα 4.6 Εκτίμηση καταστάσεων συστήματος με το ΦΔΕΣ.

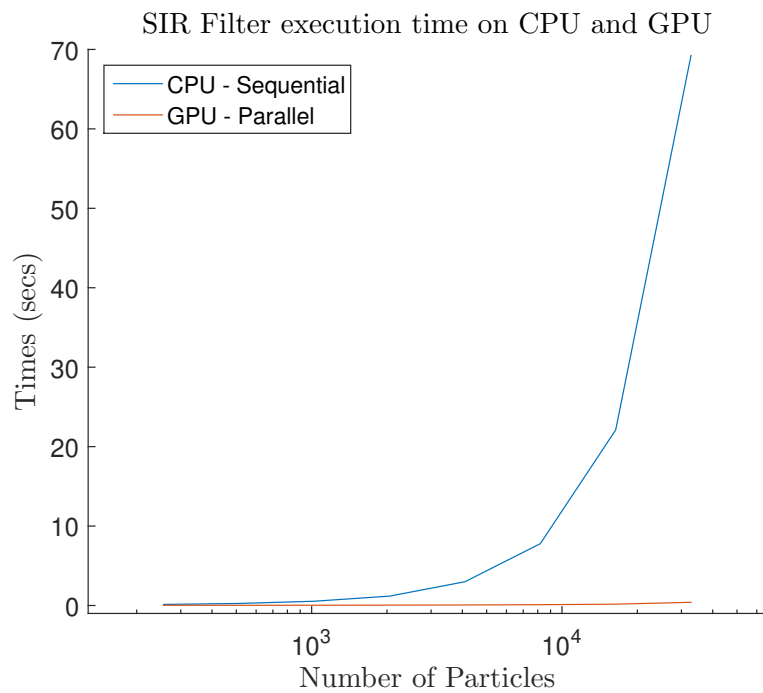
(γ') Εκτίμηση καταστάσεων με τις 20 πρώτες μετρήσεις.



(δ') Εκτίμηση καταστάσεων με τις 30 πρώτες μετρήσεις.

Σχήμα 4.7 Χρόνος Εκτέλεσης για το ΦΔΕΣ.

(α') Χρόνος εκτέλεσης ανά αριθμό διαθέσιμων μετρήσεων.



(β') Χρόνος εκτέλεσης στις CPU και GPU ανά αριθμό σωματιδίων για μια μέτρηση.

4.2 Προσομοιώσεις με πειραματικά δεδομένα

Όσον αφορά τα πειραματικά δεδομένα, χρησιμοποιήθηκαν δύο κατηγορίες δεδομένων. Στην πρώτη κατηγορία ανήκουν πειραματικά δεδομένα τα οποία αποκτήθηκαν σε δοκιμαστικό σωλήνα από πολυκυτταρικά σφαιροειδή όγκων (Multicell tumor spheroids). Πιο συγκεκριμένα, χρησιμοποιήθηκαν τρία σύνολα δεδομένων από αυτή την κατηγορία, δύο διαφορετικές κυτταρικές γραμμές γλιοβλαστώματος, η μια ανθρώπινης προέλευσης (U118 dataset) και η άλλη από ποντίκια (9L dataset) [44, 77, 78], καθώς και μια κυτταρική γραμμή καρκίνου του μαστού ανθρώπινης προέλευσης (MCF7 dataset)[79]. Το σύνολο δεδομένων U118 αποτελείται από 8 υποκείμενα ελέγχου, το 9L από 32, και το MCF7 από 5. Στη δεύτερη κατηγορία ανήκουν δεδομένα από δερματικούς όγκους τα οποία προήλθαν από εξ αρχής καρκινογένεση σε 8 εν ζώη διπλά διαγονιδιακά (double transgenic) ποντίκια (de novo carcinogenesis in vivo) [42]. Αναλυτικές πληροφορίες για τον τρόπο καλλιέργειας των πολυκυτταρικών σφαιροειδών μπορούν να βρεθούν στα [44, 79], ενώ για την καρκινογένεση στα διαγονιδιακά ποντίκια στο [42]. Σημειώνεται εδώ ότι η ακτίνα των σφαιροειδών μετρήθηκε από ψηφιακούς φωτομικρογράφους με τη χρήση του λογισμικού ανάλυσης ImageJ64-1.45i (<http://imagej.nih.gov>), υπολογίζοντας της περιοχή προβολής των σφαιροειδών και θεωρώντας σφαιρικό σχήμα [79], ενώ για τους δερματικούς όγκους των διαγονιδιακών ποντικίων μετρήθηκαν με τη χρήση παχύμετρου το μήκος και το πλάτος, και ο όγκος υπολογίστηκε θεωρώντας ελλειψοειδές σχήμα με τη χρήση της εξίσωσης:

$$V = \frac{\pi}{6}(xy)^{\frac{3}{2}} \quad (4.8)$$

4.2.1 Μετρικές

Όσον αφορά τα πραγματικά δεδομένα, είναι προφανές ότι δεν μπορεί να θεωρηθεί πραγματική τιμή για τις άγνωστες παραμέτρους, αφενός διότι το μοντέλο που χρησιμοποιείται αποτελεί προσέγγιση της ανάπτυξης των όγκων, και αφετέρου διότι ακόμα και να ήταν διαθέσιμο το μοντέλο που περιγράφει την πραγματική δυναμική της ανάπτυξης, οι πραγματικές τιμές των παραμέτρων δε θα μπορούσαν να είναι γνωστές. Για τον λόγο αυτό, η αξιολόγηση επρέπε να γίνει με διαφορετικό τρόπο απ' ότι στα συνθετικά δεδομένα. Έτσι, επιλέχθηκε να αξιολογηθεί η ικανότητα των μεθόδων να εκτιμούν παραμέτρους με βάση τις οποίες να γίνεται η πρόβλεψη της μελλοντικής ανάπτυξης των όγκων σε σχέση με τη μέθοδο των ΜΓΕΤ που χρησιμοποιείται στη βιβλιογραφία. Η μετρική που χρησιμοποιήθηκε για την αξιολόγηση της ικανότητας πρόβλεψης της μελλοντικής εξέλιξης ενός όγκου είναι η Ρίζα της Μέσης Τετραγωνικής Απόκλισης (Root-Mean-Square Deviation). Για δοσμένο σύνολο $Y_N = [y_1, \dots, y_N]$, αν οι k πρώτες μετρήσεις ($Y_{1:k} = [y_1, \dots, y_k]$),

$k < N$) χρησιμοποιηθούν για την εκτίμηση των παραμέτρων της δυναμικής της ανάπτυξης του όγκου $\hat{\theta}_1$ και $\hat{\theta}_2$, και με βάση αυτές προβλεφθεί η μελλοντική του εξέλιξη ($\hat{X}_{k+1:N} = [\hat{x}_{k+1}, \dots, \hat{x}_N]$) με τη χρήση της εξίσωσης:

$$\hat{x}_i = \hat{\theta}_2 \exp \left(\log \left(\frac{\hat{x}_{i-1}}{\hat{\theta}_2} \right) \exp \left(- \frac{1}{\hat{\theta}_1} t_i \right) \right), \quad (4.9)$$

όπου $k = 1, \dots, N - 1$ και $\hat{x}_1 = y_1$, τότε η PMTA δίνεται από την εξίσωση:

$$\text{PMTA} = \sqrt{\frac{\sum_{i=k+1}^N (y_i - \hat{x}_i)^2}{N - (k + 1)}}. \quad (4.10)$$

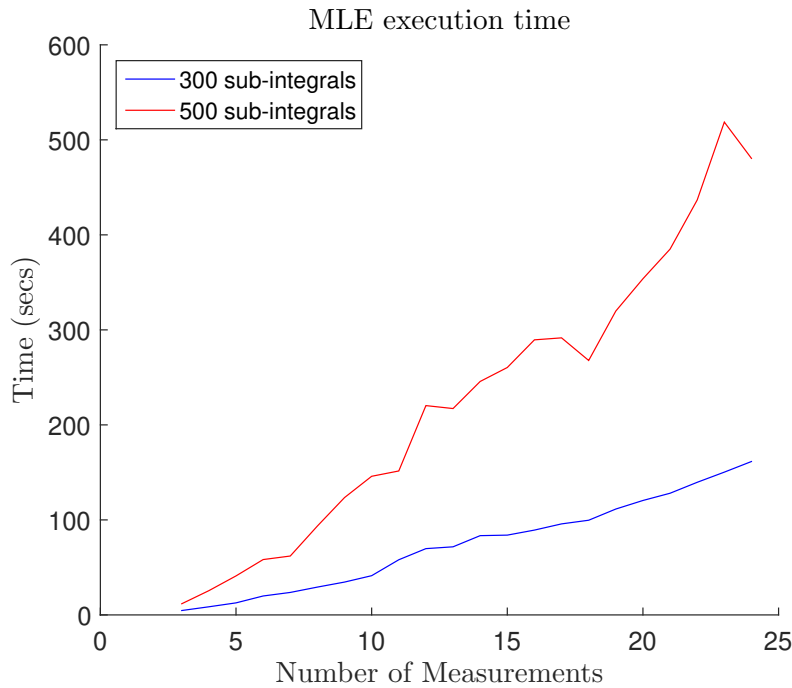
Σημειώνεται ότι η PMTA που υπολογίστηκε με τις εκτιμήσεις των ΜΓΕΤ θα αναφέρεται ως PMTA_{ET} (ή RMDS_{LS}) και η PMTA που υπολογίστηκε με τις εκτιμήσεις του ΕΜΠ θα αναφέρεται ως $\text{PMTA}_{\text{ΜΠ}}$ (ή RMDS_{ML}).

4.2.2 Υλοποίηση

Ομοίως με τα συνθετικά δεδομένα, οι προσομοιώσεις με τα πειραματικά δεδομένα υλοποιήθηκαν στο πρόγραμμα Matlab με τη δημιουργία κατάλληλων συναρτήσεων, και για να μεγιστοποιηθούν οι συναρτήσεις (3.2) και (3.7) χρησιμοποιήθηκε η συνάρτηση `fmincon` [76]. Οι προσομοιώσεις έγιναν με τη χρήση ενός επεξεργαστή Intel Core i7-6700K @ 4.00GHz και μνήμης 16GB DDR4 @ 3200 MHz. Όσον αφορά τον παράλληλο κώδικα σε γλώσσα CUDA, όπου χρησιμοποιήθηκαν παράλληλες συναρτήσεις εκτελέστηκαν σε μια NVIDIA TITAN Xp GPU.

Όσον αφορά την υλοποίηση του ΕΜΠ, αξίζει να σημειωθεί ότι το πλήθος των επιμέρους ολοκληρωμάτων που απαιτείται για να γίνει με ακρίβεια ο υπολογισμός των κατανομών (2.23) και (2.24), διαφέρει για τα σύνολα των συνθετικών και των πειραματικών δεδομένων. Για τα πειραματικά δεδομένα, ο αριθμός των υποδιαστημάτων που απαιτείται για τον υπολογισμό των εξισώσεων (3.10) έως (3.13) είναι σημαντικά υψηλότερος σε σχέση με τον αριθμό που απαιτείται για τα συνθετικά δεδομένα. Ενώ στα συνθετικά δεδομένα, 50 με 70 υποδιαστήματα είναι αρκετά, στις περισσότερες περιπτώσεις με πειραματικά δεδομένα, χρειάστηκαν 300 υποδιαστήματα για να υπολογιστούν με ακρίβεια οι εξισώσεις (3.10) έως (3.13), ενώ υπήρξαν και περιπτώσεις στις οποίες απαιτήθηκαν 500 υποδιαστήματα. Το σχήμα 4.8 δείχνει το χρόνο που απαιτείται για την εκτίμηση των άγνωστων παραμέτρων για 300 και 500 υποδιαστήματα.

Σχήμα 4.8 Χρόνος εκτέλεσης ανά αριθμό μετρήσεων για τον ΕΜΠ με αριθμητική ολοκλήρωση.



4.2.3 Αποτελέσματα από ανάλυση Πολυκυτταρικών Σφαιροειδών Όγκων

Στο υποκεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα που προέκυψαν από την ανάλυση πολυκυτταρικών σφαιροειδών. Εξετάστηκαν τρία διαφορετικά σύνολα δεδομένων, τα δύο ανθρώπινης προέλευσης και το τρίτο από ποντίκια. Από τα δεδομένα ανθρώπινης προέλευσης το ένα προέρχεται από κυτταρική γραμμή καρκίνου του μαστού (MCF7), ενώ το δεύτερο από κυτταρική γραμμή γλιωβλαστώματος (U118). Τα δεδομένα από ποντίκια προέρχονται επίσης από κυτταρική γραμμή γλιωβλαστώματος (9L). Τα διαγράμματα του Σχήματος 4.9 δείχνουν την ανάπτυξη δύο πολυκυτταρικών σφαιροειδών όγκων από κάθε σύνολο δεδομένων, έτσι ώστε να πληροφορήσουν τον αναγνώστη για το είδος των πειραματικών δεδομένων που χρησιμοποιήθηκαν στις προσομοιώσεις.

Στα προαναφερθέντα σύνολα δεδομένων εφαρμόστηκε ο ΕΜΠ με αριθμητική ολοκλήρωση και ο εκτιμητής ΜΓΕΤ, και έγινε σύγκριση της απόδοσης τους. Κατά την εφαρμογή του ΕΜΠ, τόσο ο ρυθμός αύξησης όσο και η φέρουσα χωρητικότητα θεωρούνται άγνωστες παράμετροι και εκτιμώνται σε κάθε χρονική στιγμή που μια νέα μέτρηση γίνεται διαθέσιμη. Όσον αφορά τον εκτιμητή ΜΓΕΤ, μαζί με την προτεινόμενη στη βιβλιογραφία μέθοδο όπου η φέρουσα χωρητικότητα θεωρείται φιξαρισμένη σε μια γνωστή τιμή,

δοκιμάστηκε επιπλέον και η περίπτωση που η φέρουσα χωρητικότητα θεωρείται άγνωστη παράμετρος. Η μέθοδος όπου η φέρουσα χωρητικότητα θεωρείται άγνωστη παράμετρος μπορεί να αναφέρεται στη συνέχεια και σαν ΜΓΕΤ_{2p} (ή LS_{2p}). Στην περίπτωση της φιξάρισμεινης φέρουσας χωρητικότητας, η τιμή της υπολογίστηκε με δύο τρόπους. Ο πρώτος προέρχεται από την βιβλιογραφία και είναι η Μέγιστη Εφικτή Τιμή. Σύμφωνα λοιπόν με την βιβλιογραφία, όπως αναφέρθηκε και στο Κεφάλαιο 1.3.1, κατά τη διαδικασία ταιριάσματος πειραματικών δεδομένων στην εξίσωση Gompertz, η φέρουσα χωρητικότητα ενός όγκου μπορεί να θεωρηθεί ως μια παράμετρος που ασυμπτωτικά έχει την ίδια τιμή για όλους τους οργανισμούς ενός συγκεκριμένου είδους. Επομένως, σαν ΜΕΤ καθορίστηκε η τιμή της μεγαλύτερης μέτρησης που παρατηρήθηκε σε κάθε σύνολο δεδομένων. Η μέθοδος αυτή μπορεί να αναφέρεται στη συνέχεια και σαν ΜΓΕΤ_{MET} (ή LS_{MAV}). Ο δεύτερος τρόπος για τον υπολογισμό της φέρουσας χωρητικότητας είναι η Μέση Τιμή των Μέσων. Αφού υπολογιστεί ο μέσος όρων των τελευταίων n μετρήσεων για κάθε σφαιροειδές, στη συνέχεια υπολογίζεται η μέση τιμή των μέσων όρων, και η φέρουσα χωρητικότητα φιξάρεται σε αυτή την τιμή. Η μέθοδος αυτή μπορεί να αναφέρεται στη συνέχεια και σαν ΜΓΕΤ_{MTM} (ή LS_{MM}). Η ΜΤΜ δίνεται από την εξίσωση:

$$MM = \frac{1}{nM} \sum_{i=1}^M \sum_{j=N-n}^N y_{i,j}, \quad (4.11)$$

όπου n είναι ο αριθμός των τελευταίων μετρήσεων που χρησιμοποιήθηκαν από κάθε σφαιροειδές και M είναι ο συνολικός αριθμός των σφαιροειδών του συνόλου δεδομένων. Στις προσομοιώσεις που έγιναν η τιμή για την παράμετρο n ήταν 10 μετρήσεις για τα σύνολα U118 και 9L, και 5 μετρήσεις για το σύνολο MCF7 λόγω του μικρού αριθμού μετρήσεων σε αυτό το σύνολο.

Οι εκτιμήσεις που προέκυψαν από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ χρησιμοποιήθηκαν στη συνέχεια για να γίνει πρόβλεψη της μελλοντικής πορείας του όγκου. Για να μετρηθεί η ακρίβεια των προβλέψεων χρησιμοποιήθηκε ΡΜΤΑ που δίνεται από την εξίσωση (4.10). Δεδομένου ότι $Y_N = [y_1, \dots, y_N]$ είναι το σύνολο μετρήσεων για ένα σφαιροειδές, οι k πρώτες μετρήσεις ($Y_{1:k} = [y_1, \dots, y_k], k < N$) χρησιμοποιούνται για να γίνει εκτίμηση των αγνώστων παραμέτρων της δυναμικής του όγκου, και στη συνέχεια οι εκτιμήσεις αυτές χρησιμοποιούνται για να προβλεφθεί η μελλονική ανάπτυξη του όγκου ($\hat{X}_{k+1:N} = [\hat{x}_{k+1}, \dots, \hat{x}_N]$). Στη συνέχεια υπολογίζεται η ΡΜΤΑ από την εξίσωση (4.10).

Στους πίνακες 4.2 - 4.4 φαίνεται το πλήθος των περιπτώσεων όπου η ΡΜΤΑ_{ΜΠ} ήταν μικρότερη από την ΡΜΤΑ_{ΕΤ}, δηλαδή οι περιπτώσεις στις οποίες ο ΕΜΠ έχε καλύτερη απόδοση από τον εκτιμητή ΜΓΕΤ. Στα σχήματα 4.10--4.12 παρουσιάζεται η μέση τιμή της ΡΜΤΑ_{ΜΠ} και της ΡΜΤΑ_{ΕΤ} στα χρονικά σημεία που μελετήθηκαν. Σε κάθε σχήμα

παρουσιάζεται η σύγκριση της μέσης τιμής της PMTA που προκύπτει από τη χρήση των εκτιμήσεων του ΕΜΠ με κάθεμια από τις υλοποιήσεις τους ΜΓΕΤ ξεχωριστά.

Πίνακας 4.2 Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PMTA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του ΕΜΠ έχει μικρότερη τιμή στο σύνολο δεδομένων MCF7.

MCF7												
Αριθμός Διαθέσιμων Μετρήσεων												
	2	3	4	5	6	7	8	9	10	11	12	13
	ML vs LS_{MAV}											
Πλήθος Όγκων	1	4	2	1	1	1	0	0	0	1	0	0
	ML vs LS_{MM}											
Πλήθος Όγκων	2	3	4	3	3	2	2	2	2	2	1	1
	ML vs LS_{2p}											
Πλήθος Όγκων	2	3	3	2	2	2	1	0	1	1	0	0

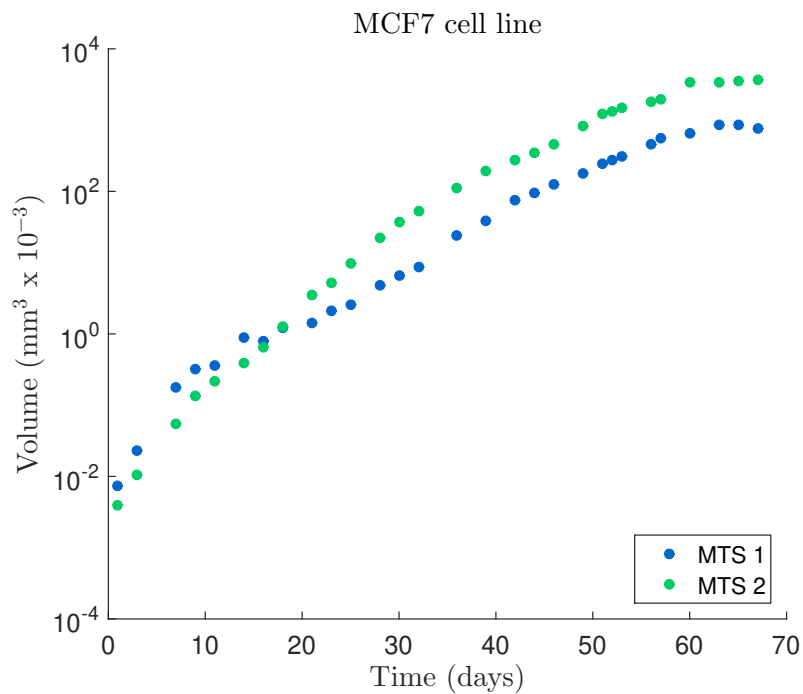
Πίνακας 4.3 Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PMTA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του ΕΜΠ έχει μικρότερη τιμή στο σύνολο δεδομένων U118.

U118										
Αριθμός Διαθέσιμων Μετρήσεων										
	2	3	4	5	6	7	8	9	10	11
	ML vs LS _{MAV}									
Πλήθος Όγκων	5	2	3	3	3	2	2	4	4	5
	ML vs LS _{MM}									
Πλήθος Όγκων	4	4	4	3	4	3	3	2	7	6
	ML vs LS _{2p}									
Πλήθος Όγκων	5	7	6	6	3	3	3	5	5	6
Αριθμός Διαθέσιμων Μετρήσεων										
	12	13	14	15	16	17	18	19	20	
	ML vs LS _{MAV}									
Πλήθος Όγκων	5	5	5	5	3	5	4	5	3	
	ML vs LS _{MM}									
Πλήθος Όγκων	4	4	5	5	4	4	4	5	4	
	ML vs LS _{2p}									
Πλήθος Όγκων	4	4	5	5	3	5	3	3	1	

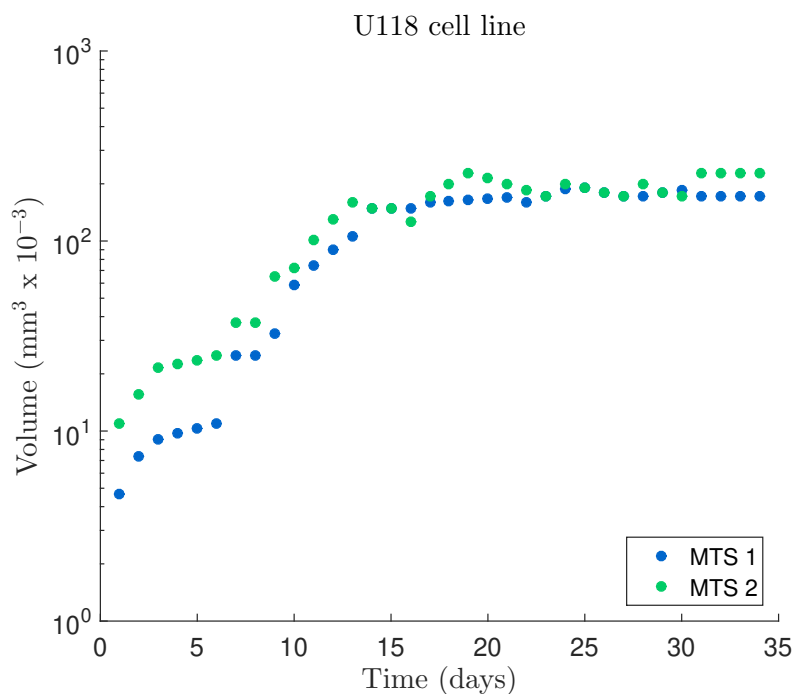
Πίνακας 4.4 Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η PMTA των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του ΕΜΠ έχει μικρότερη τιμή στο σύνολο δεδομένων 9L.

9L										
Αριθμός Διαθέσιμων Μετρήσεων										
	2	3	4	5	6	7	8	9	10	11
	ML vs LS _{MAV}									
Πλήθος Όγκων	31	29	29	32	32	31	30	30	31	30
	ML vs LS _{MM}									
Πλήθος Όγκων	4	9	7	7	5	5	5	6	9	8
	ML vs LS _{2p}									
Πλήθος Όγκων	13	16	18	18	16	22	17	17	16	16
	Αριθμός Διαθέσιμων Μετρήσεων									
	12	13	14	15	16	17	18	19	20	21
	ML vs LS _{MAV}									
Πλήθος Όγκων	27	27	28	26	27	27	27	25	27	29
	ML vs LS _{MM}									
Πλήθος Όγκων	10	14	16	16	16	15	16	17	22	23
	ML vs LS _{2p}									
Πλήθος Όγκων	16	16	12	13	12	12	8	13	18	19
	Αριθμός Διαθέσιμων Μετρήσεων									
	22	23	24	25	26	27	28	29	30	
	ML vs LS _{MAV}									
Πλήθος Όγκων	30	29	29	29	30	31	31	30	32	
	ML vs LS _{MM}									
Πλήθος Όγκων	25	23	25	21	24	23	24	23	24	
	ML vs LS _{2p}									
Πλήθος Όγκων	20	14	16	18	23	23	22	19	23	

Σχήμα 4.9 Εξέλιξη ανάπτυξης αντιπροσωπευτικών Πολυκυτταρικών Σφαιροειδών Όγκων για κάθε σύνολο δεδομένων.

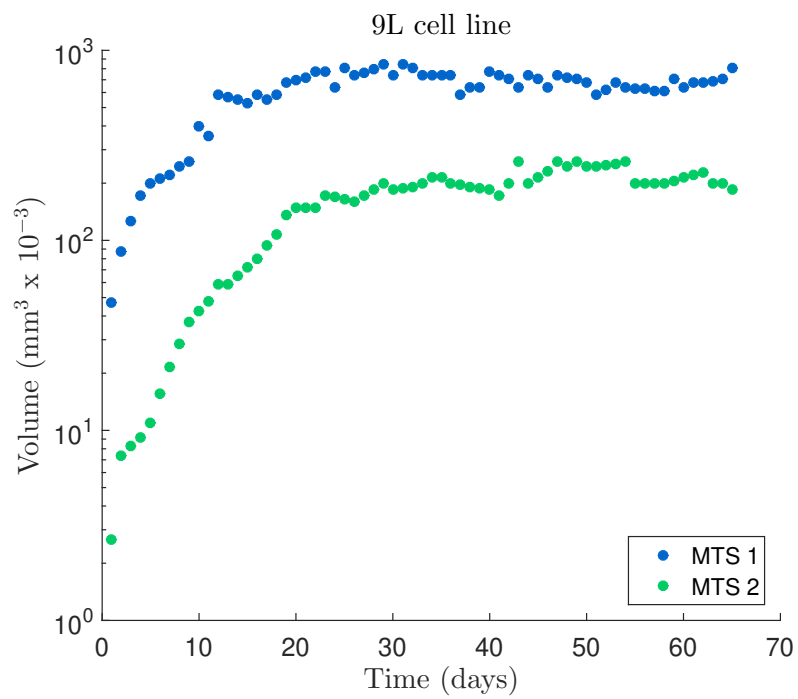


(α') Κυτταρική γραμμή MCF7.



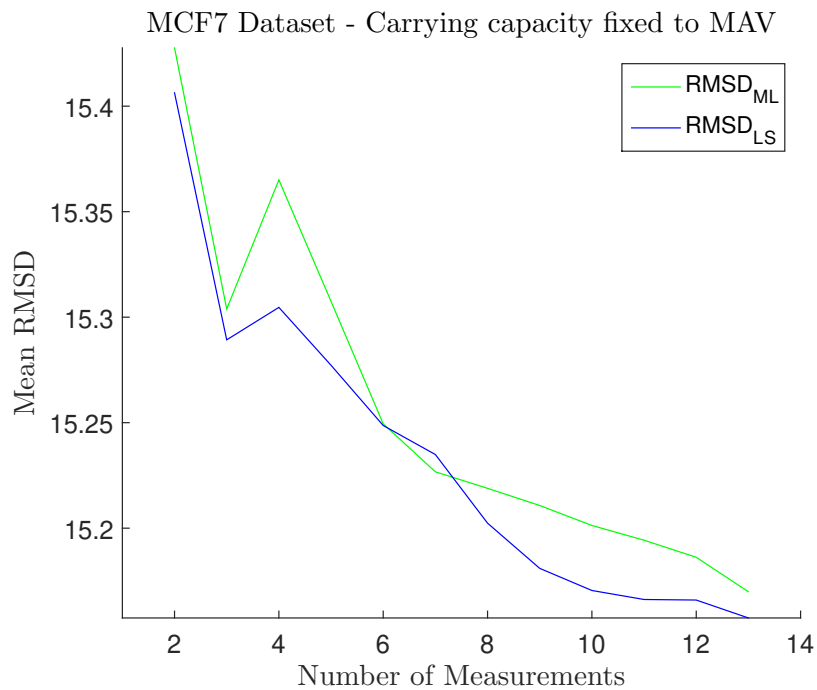
(β') Κυτταρική γραμμή U118.

Σχήμα 4.9 Εξέλιξη ανάπτυξης αντιπροσωπευτικών Πολυκυτταρικών Σφαιροειδών Όγκων για κάθε σύνολο δεδομένων.

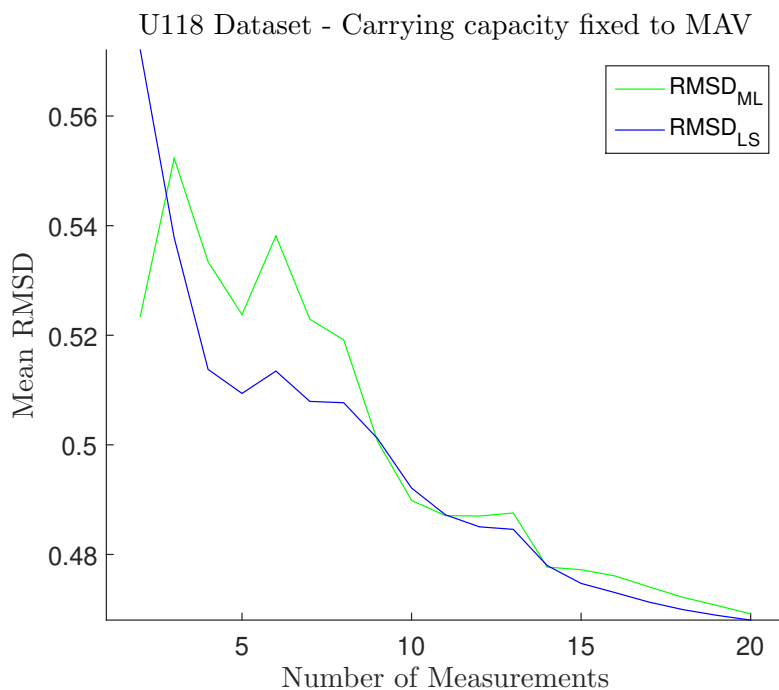


(γ) Κυτταρική γραμμή 9L.

Σχήμα 4.10 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{MET}.

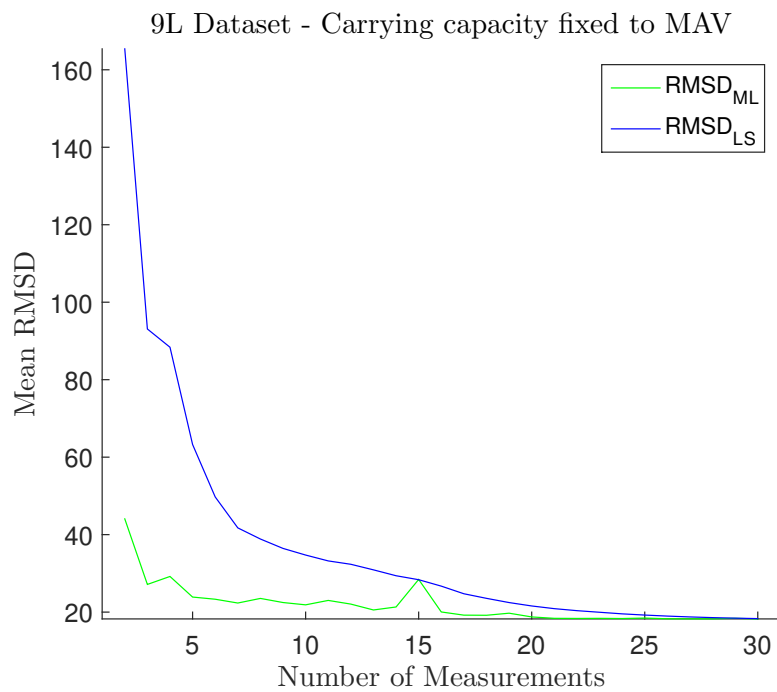


(α') Κυτταρική γραμμή MCF7.



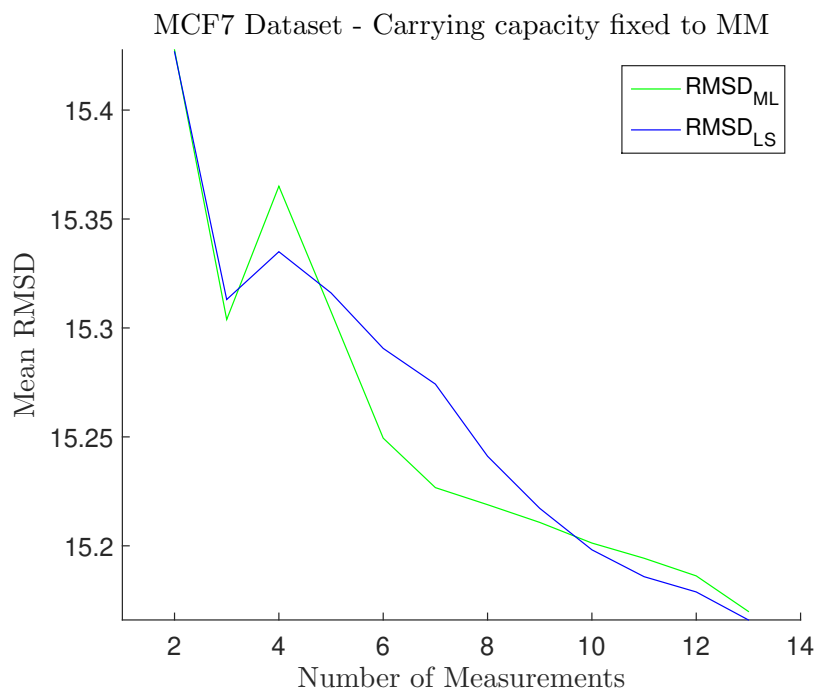
(β') Κυτταρική γραμμή U118.

Σχήμα 4.10 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{MET}.

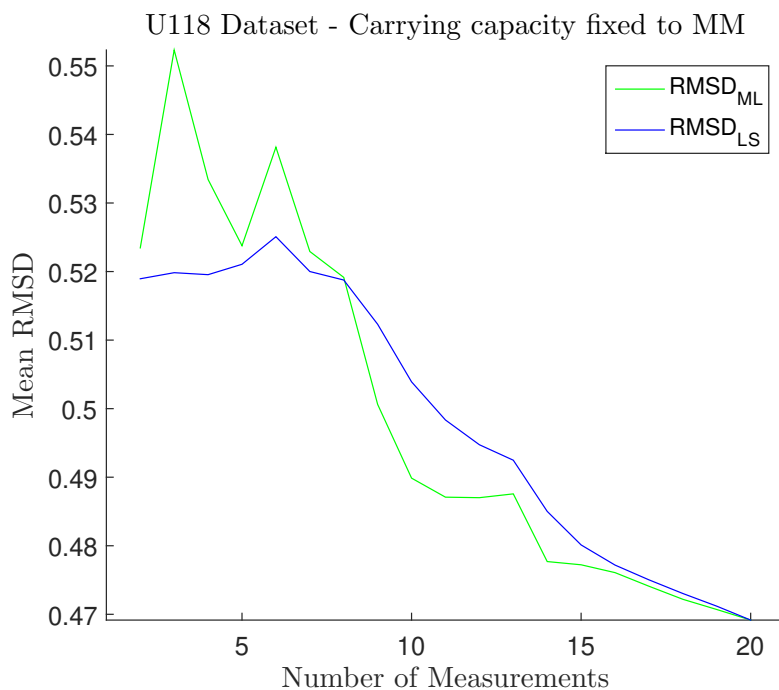


(γ') Κυτταρική γραμμή 9L.

Σχήμα 4.11 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{MTM}.

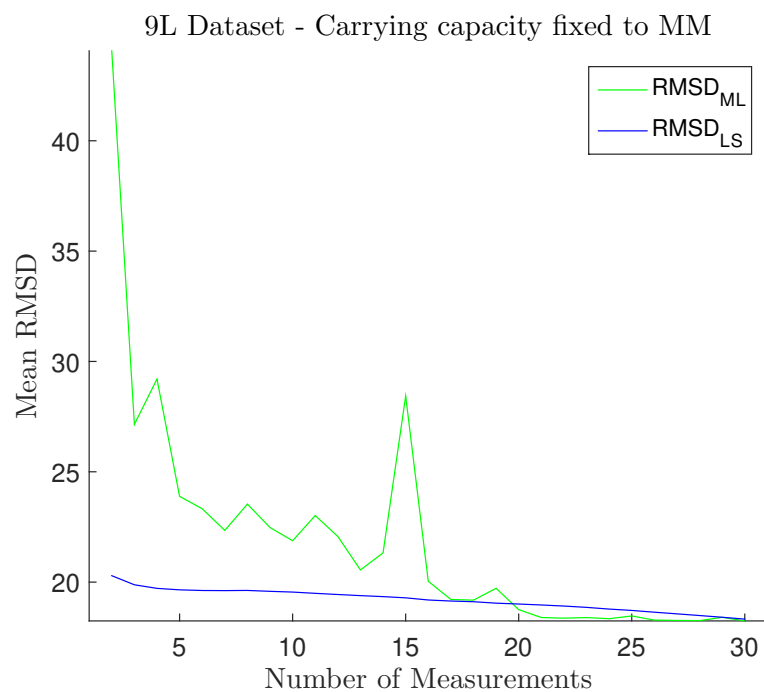


(α') Κυτταρική γραμμή MCF7.



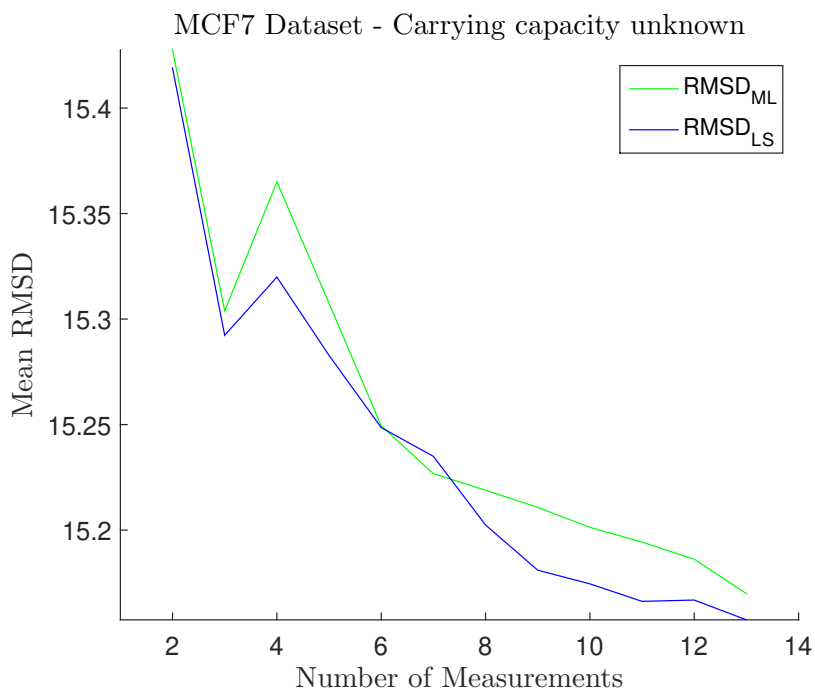
(β') Κυτταρική γραμμή U118.

Σχήμα 4.11 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{MTM}.

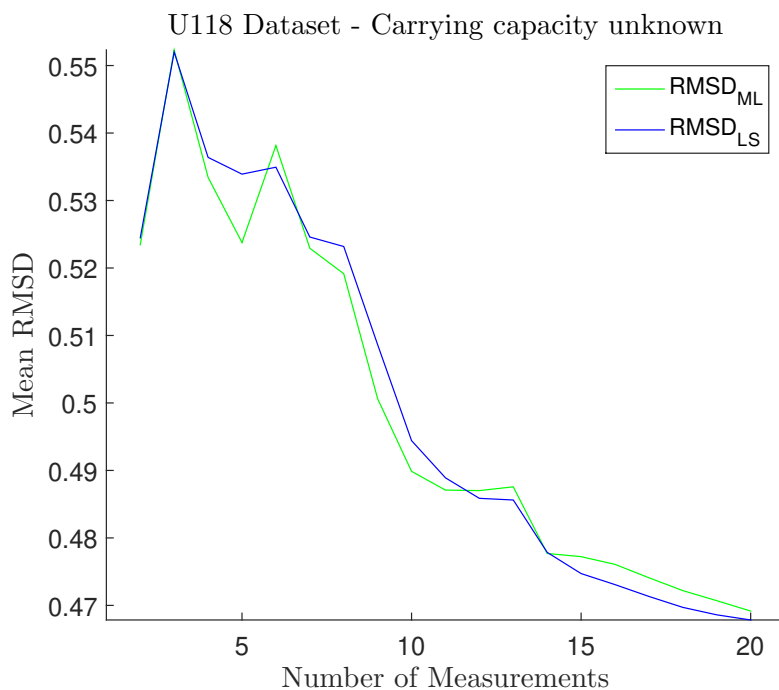


(γ') Κυτταρική γραμμή 9L.

Σχήμα 4.12 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{2p}.

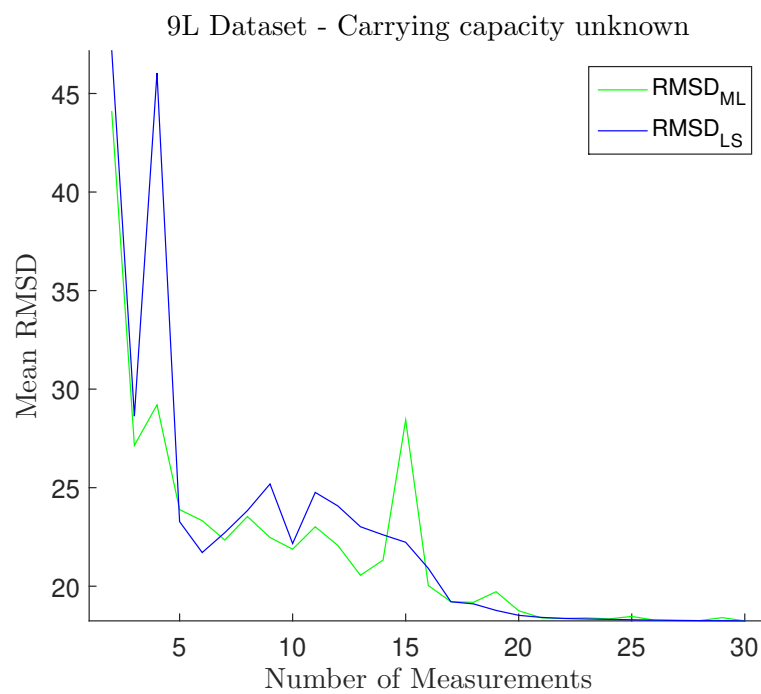


(α') Κυτταρική γραμμή MCF7.



(β') Κυτταρική γραμμή U118.

Σχήμα 4.12 Μέση τιμή PMTA των προβλέψεων με τη χρήση των εκτιμήσεων από τον ΕΜΠ και τον εκτιμητή ΜΓΕΤ_{2p}.



(γ) Κυτταρική γραμμή 9L.

4.2.4 Αποτελέσματα από Δερματικούς Όγκους

Σε αυτό το υποκεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν από την ανάλυση 24 όγκων που δεν έλαβαν θεραπεία σε 8 διπλά διαγονιδιακά ποντίκια (3 όγκους σε κάθε ποντίκι). Κάθε ποντίκι έχει μια ονομασία που ξεκινάει από τα γράμματα CM και ακολουθείται από έναν αριθμό (π.χ. CM37, CM38, κλπ). Οι επιτρεπτές τιμές για τις άγνωστες παραμέτρους θ_1 και θ_2 επιλέχθηκαν με βάση τη βιβλιογραφία [42], ενώ για τις παραμέτρους $\theta_3 - \theta_6$ επιλέχθηκαν με βάση τη διαδικασία της δοκιμής και του λάθους.

Στα παραπάνω δεδομένα εφαρμόστηκε ο ΕΜΠ με αριθμητική ολοκλήρωση, ο ΕΜΥΠ και ο εκτιμητής ΜΓΕΤ, και έγινε σύγκριση της απόδοσης τους. Κατά την εφαρμογή του ΕΜΠ και του ΕΜΥΠ, τόσο ο ρυθμός αύξησης όσο και η φέρουσα χωρητικότητα θεωρούνται άγνωστες παράμετροι και εκτιμώνται σε κάθε χρονική στιγμή που μια νέα μέτρηση γίνεται διαθέσιμη. Όσον αφορά τον εκτιμητή ΜΓΕΤ η φέρουσα χωρητικότητα φιξάρεται στην ΜΕΤ. Η μετρική που χρησιμοποιήθηκε για να μετρηθεί η ακρίβεια των προβλέψεων είναι η ΡΜΤΑ που δίνεται από την εξίσωση (4.10). Σε κάθε χρονική στιγμή, οι εκτιμήσεις από κάθε μέθοδο, χρησιμοποιήθηκαν για να εκτιμηθεί η μελλοντική εξέλιξη του όγκου.

Από τους 24 όγκους που μελετήθηκαν, παρατηρήθηκε ότι σε δύο περιπτώσεις (Όγκος 1 για το ποντίκι CM60 και Όγκος 2 για το ποντίκι CM79) οι εκτιμητές που χρησιμοποιήθηκαν παρείχαν πλήρως αναξιόπιστες προβλέψεις για την εξέλιξη των όγκων κατά τα πρώτα στάδια της παρατήρησης. Η συμπεριφορά αυτή οφείλεται πιθανότατα στην ασυνήθιστα αργή ανάπτυξη που παρουσίασαν οι δύο αυτοί όγκοι. Λόγω της μεγάλης απόκλισης των εκτιμήσεων από τις παρατηρούμενες τιμές, αποφασίστηκε να μην συμπεριληφθούν στη σύγκριση των μεθόδων, διότι η χρήση τους θα επηρέαζε σημαντικά την μέση τιμή της ΡΜΤΑ. Αξίζει όμως να σημειωθεί ότι και στις δύο περιπτώσεις, ο ΕΜΠ παρείχε πιο ακριβείς εκτιμήσεις σε σχέση με τον εκτιμητή ΜΓΕΤ. Για τους υπόλοιπους 22 όγκους, γίνεται στη συνέχεια σύγκριση μεταξύ του ΕΜΠ και του εκτιμητή ΜΓΕΤ. Ο πίνακας 4.5 δείχνει το πλήθος των όγκων για τους οποίους ο ΕΜΠ παρείχε πιο ακριβείς προβλέψεις όσον αφορά τη ΡΜΤΑ. Πιο συγκεκριμένα, με 3 διαθέσιμες μετρήσεις για την εκτίμηση των αγνώστων παραμέτρων ο ΕΜΠ παρείχε πιο ακριβείς προβλέψεις για 12 από τους 22 όγκους, για 4 διαθέσιμες μετρήσεις για 12 από τους 22 όγκους, κ.ο.κ.

Πίνακας 4.5 Αριθμός όγκων σε διαφορετικές χρονικές στιγμές για τους οποίους η ΡΜΤΑ των προβλέψεων που προέκυψαν από τη χρήση των εκτιμήσεων του ΕΜΠ έχει μικρότερη τιμή στο σύνολο δεδομένων από δερματικούς όγκους ποντικίων.

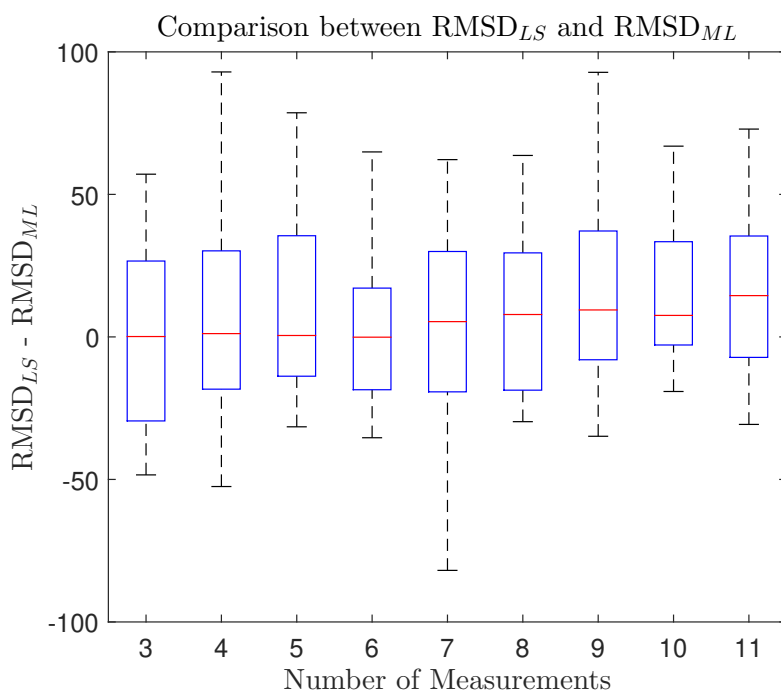
	Αριθμός διαθέσιμων μετρήσεων									
	3	4	5	6	7	8	9	10	11	
Πλήθος Όγκων	12	12	12	11	12	14	13	15	15	

Στο σχήμα 4.13 παρουσιάζεται το θηκόγραμμα (γραφικός τρόπος παρουσίασης πέντε περιληπτικών μέτρων μιας κατανομής ομαδοποιημένων δεδομένων) της διαφοράς μεταξύ της $PMTA_{ET}$ και $PMTA_{MPI}$ για κάθε όγκο στις διαφορετικές χρονικές στιγμές που μελετήθηκαν. Μέσα σε κάθε ορθογώνιο πλαίσιο, τοποθετείται με μια κόκκινη γραμμή η διάμεσος, ενώ οι επάνω και κάτω ακμές δείχνουν το 25^ο και 75^ο εκατοστημόριο αντίστοιχα. Από τα μέσα των άνω και κάτω βάσεων των ορθογωνίων πλαισίων αναπτύσσονται γραμμές που εκτείνονται μέχρι τα πιο ακραία σημεία που θεωρούνται οριακές τιμές. Πιθανές τιμές που βρίσκονται εκτός των οριακών τιμών, δηλώνονται επιπλέον, συνήθως με αστερίσκους ή κουκκίδες. Στο σχήμα 4.13 οι τιμές που βρίσκονται εκτός των οριακών τιμών δεν συμπεριλήφθηκαν λόγω του πολύ μικρού αριθμού τους. Στο σχήμα 4.13 παρατηρείται ότι η διαφορά μεταξύ της $PMTA_{ET}$ και $PMTA_{MPI}$ είναι μετατοπισμένη προς τον θετικό άνω ημιχώρο. Προκύπτει επομένως το συμπέρασμα ότι ο ΕΜΠ παρέχει γενικότερα πιο ακριβείς προβλέψεις όσον αφορά τη $PMTA$.

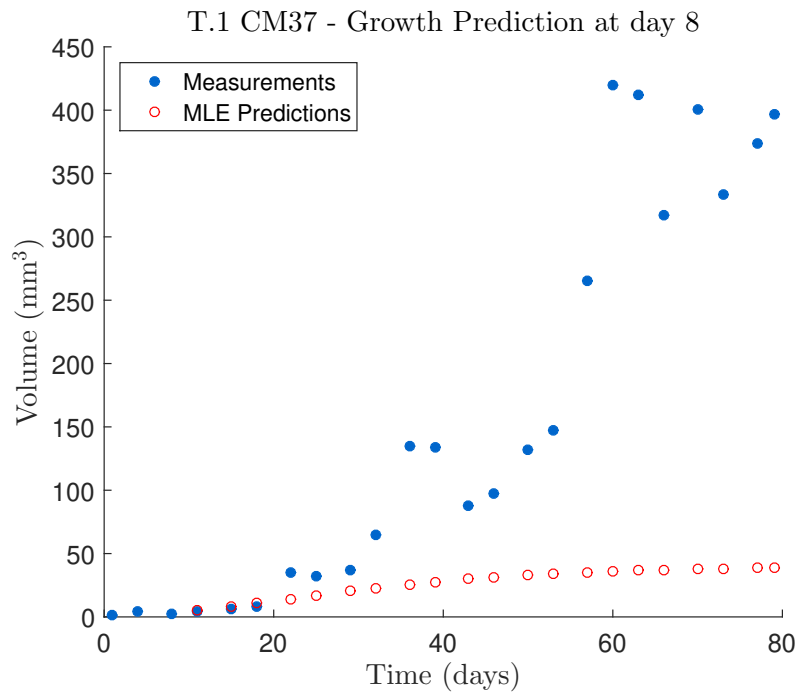
Στο σχήμα 4.14 παρουσιάζεται ενδεικτικά η εξέλιξη για έναν από τους όγκους του ποντικίου CM37, καθώς και οι προβλέψεις της εξέλιξης του όγκου με βάση τις εκτιμήσεις του ΕΜΠ σε διάφορες χρονικές στιγμές. Στο σχήμα 4.14α' φαίνονται οι προβλέψεις στο τέλος της πρώτης εβδομάδας, στο σχήμα 4.14β' οι προβλέψεις στο τέλος της δεύτερης εβδομάδας, στο σχήμα 4.14γ' οι προβλέψεις στο τέλος της τέταρτης εβδομάδας και στο σχήμα 4.14δ' οι προβλέψεις στο τέλος της έβδομης εβδομάδας. Όπως είναι αναμενόμενο, ο ΕΜΠ δεν είναι σε θέση να εκτιμήσει την εξέλιξη του όγκου στα αρχικά στάδια, μιας και η πληροφορία που υπάρχει στα δεδομένα σχετικά με το σημείο κορεσμού του όγκου, δηλαδή τη φέρουσα χωρητικότητα είναι σχεδόν αναύπαρκτη.

Στα σχήματα 4.15 και 4.16 φαίνονται οι κατανομές του ΕΜΥΠ και οι προβλέψεις που γίνονται με τη χρήση των παραμέτρων που εκτιμώνται από τον ΕΜΥΠ για έναν όγκο του ποντικίου CM37. Στο σχήμα 4.15α' φαίνεται η εκ των προτέρων κατανομή πιθανότητας για τις άγνωστες παραμέτρους που χρησιμοποιείται από τον ΕΜΥΠ, στα σχήματα 4.15β' και 4.15δ' οι κατανομές της πιθανοφάνειας και στα σχήματα 4.15γ' και 4.15ε' οι εκ των υστέρων κατανομές πιθανότητας. Στα σχήματα 4.15β' και 4.15δ' φαίνεται αυτό που αναφέρθηκε ωρίτερα σχετικά με την πληροφορία που παρέχουν τα δεδομένα σχετικά με τη φέρουσα χωρητικότητα. Σε κάποιες περιπτώσεις, όπως για παράδειγμα σε αυτές που φαίνονται στα σχήματα 4.15γ' και 4.15ε', ο ΕΜΥΠ καταφέρνει να ενσωματώσει την πληροφορία που εμπεριέχεται στην εκ των προτέρων κατανομή και να δώσει μια βελτιωμένη εκτίμηση (βλ. σχήμα 4.16) της μελλοντικής εξέλιξης του καρκινικού όγκου στο τέλος της πρώτης και της δεύτερης εβδομάδας.

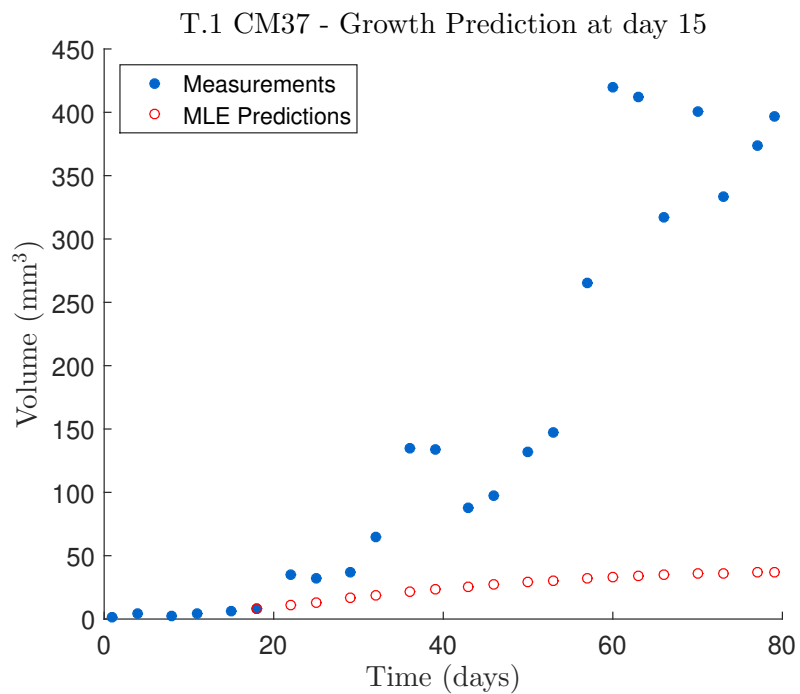
Σχήμα 4.13 Σύγκριση της PMTA των προβλέψεων μεταξύ του εκτιμητή ΜΓΕΤ και του ΕΜΠ.



Το θηκόγραμμα απεικονίζει την κατανομή της διαφοράς μεταξύ των $PMTA_{ET}$ και $PMTA_{MP}$ για κάθε όγκο στις διάφορες χρονικές στιγμές. Απο τη στιγμή που η διαφορά αυτή είναι μετατοπισμένη προς τον θετικό άνω ημιχώρο, συμπεραίνεται ότι γενικότερα ο ΕΜΠ παρέχει πιο ακριβείς προβλέψεις.

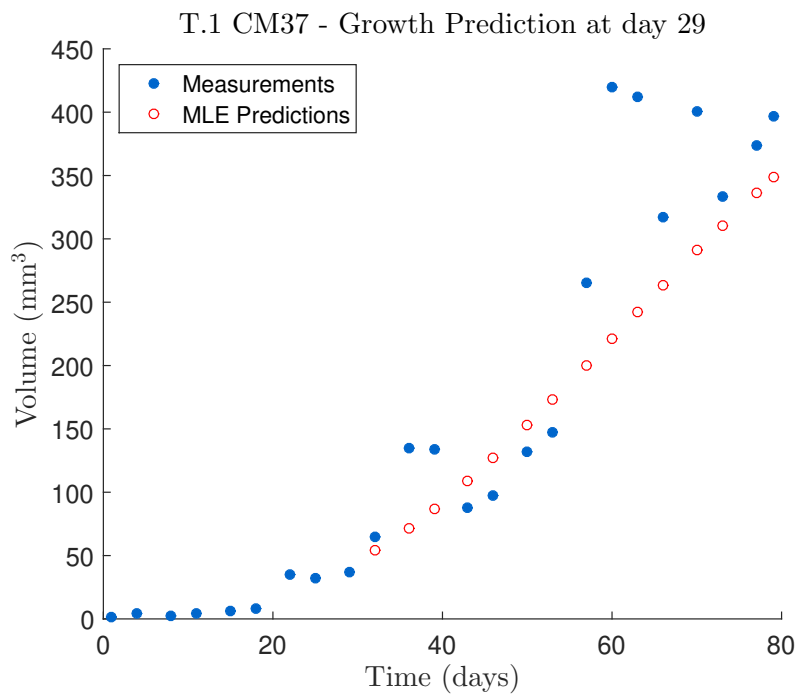
Σχήμα 4.14 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικιού CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ.

(α') 3 μετρήσεις.

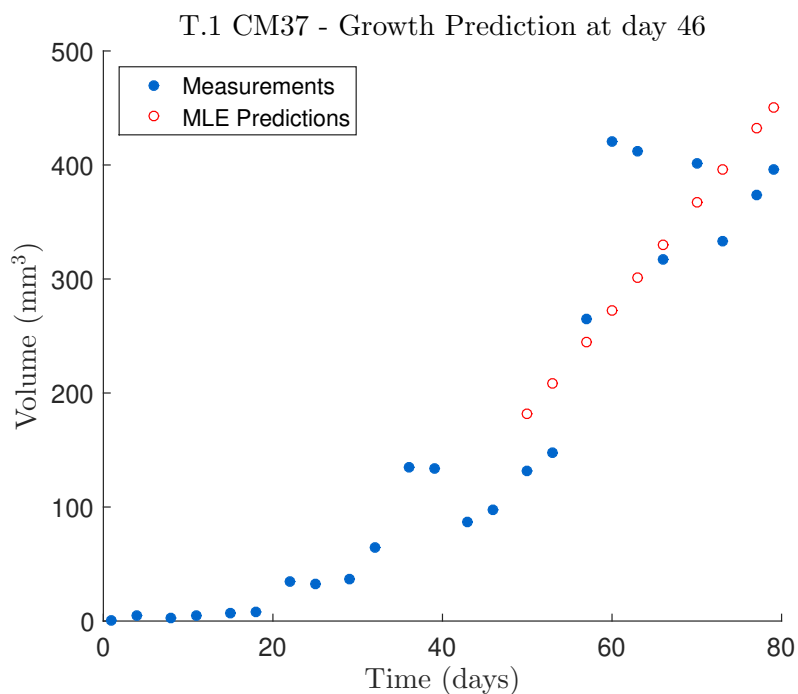


(β') 5 μετρήσεις.

Σχήμα 4.14 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικίου CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ.



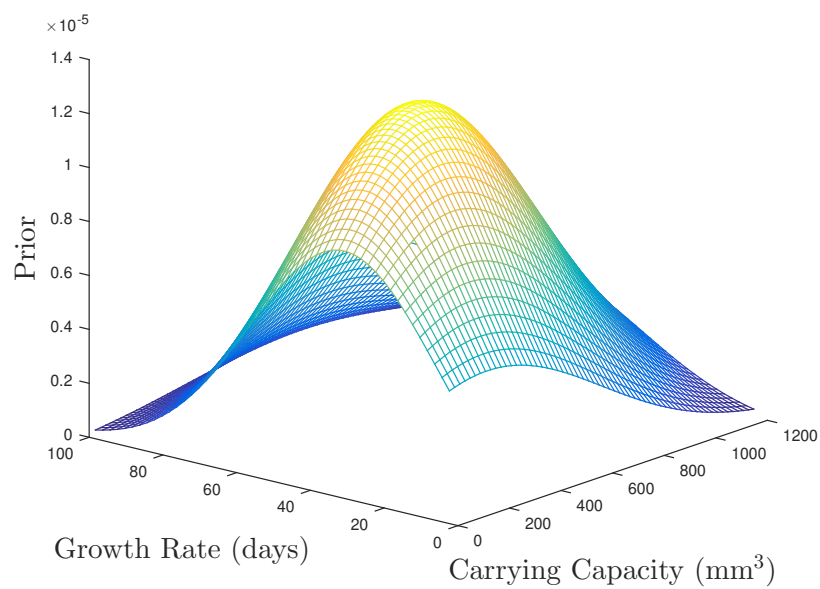
(γ) 9 μετρήσεις.



(δ) 14 μετρήσεις.

Σχήμα 4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικού CM37.

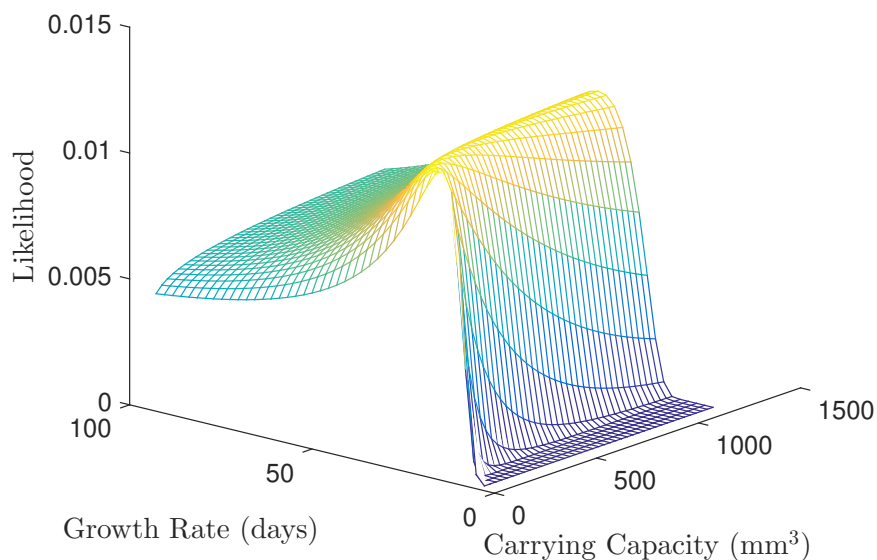
Prior distribution for the growth parameters



(α') Κατανομή για την εκ των προτέρων σ.π.π.

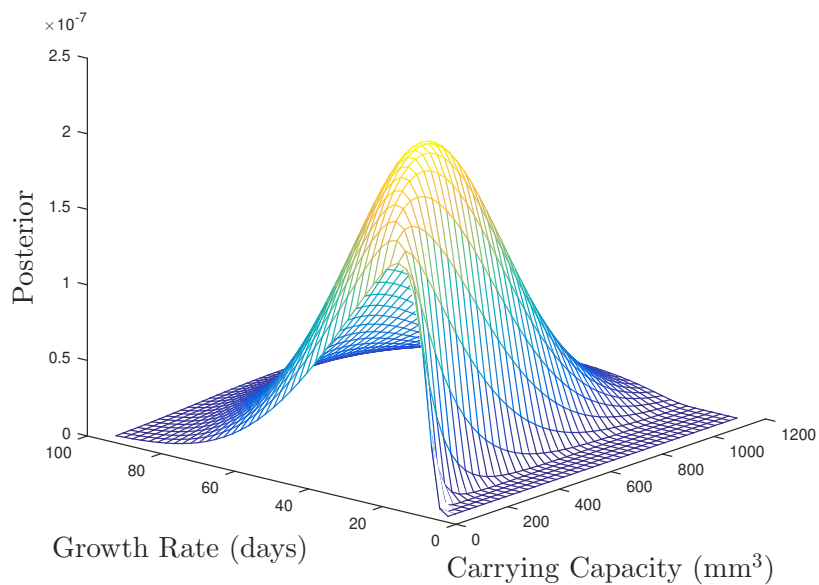
Σχήμα 4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικιού CM37.

T.1 CM37 - Likelihood Distribution at day 8



(β') Κατανομή για την συνάρτηση πιθανοφάνειας -- 3 Μετρήσεις

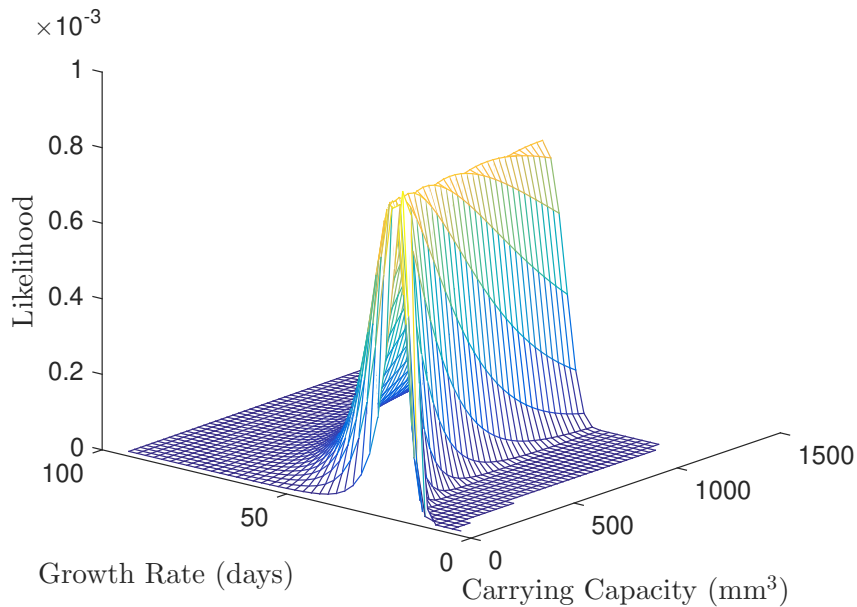
T.1 CM37 - Posterior Distribution at day 8



(γ') Κατανομή για την εκ των υστέρων σ.π.π. -- 5 Μετρήσεις

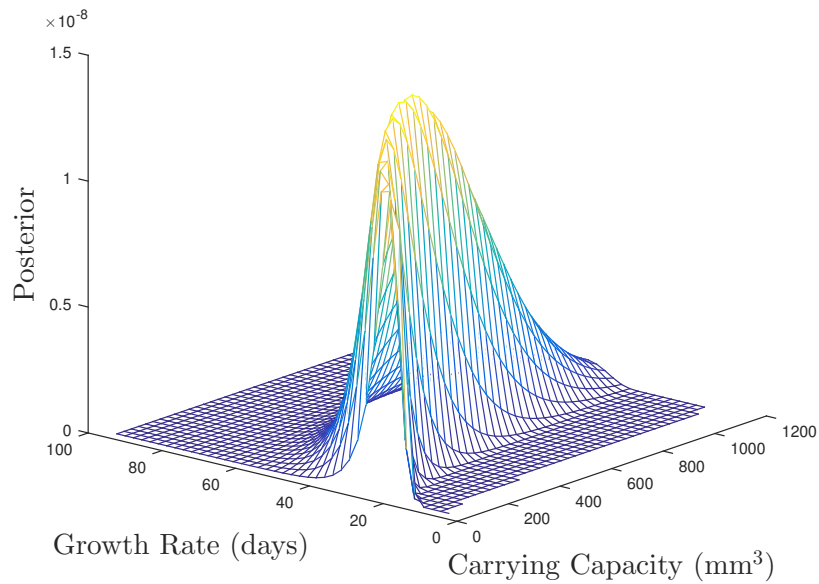
Σχήμα 4.15 Πλέγματα για τις εκ των προτέρων σ.π.π., συνάρτηση πιθανοφάνειας και εκ των υστέρων σ.π.π. για τον όγκο 1 του ποντικιού CM37.

T.1 CM37 - Likelihood Distribution at day 15



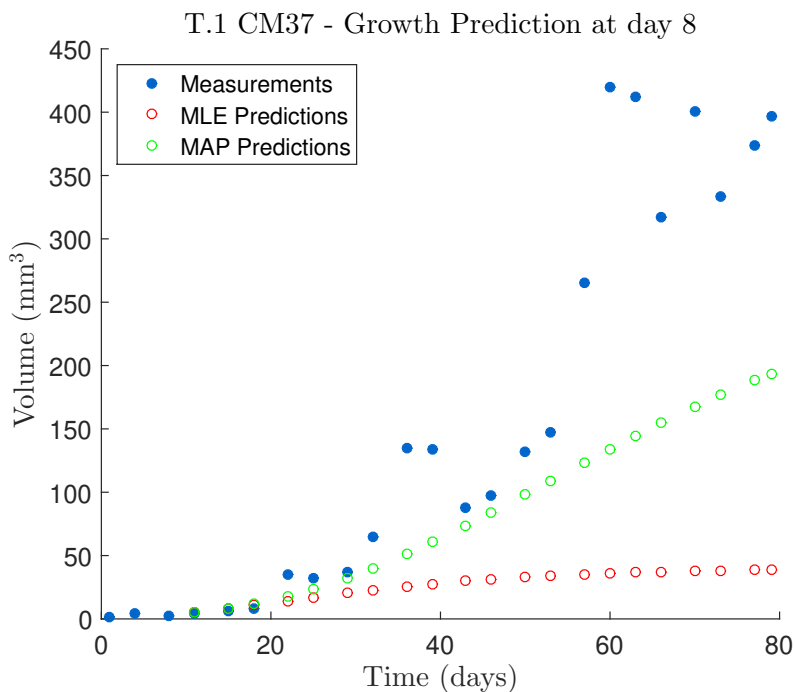
(δ) Κατανομή για την συνάρτηση πιθανοφάνειας -- 5 Μετρήσεις

T.1 CM37 - Posterior Distribution at day 15

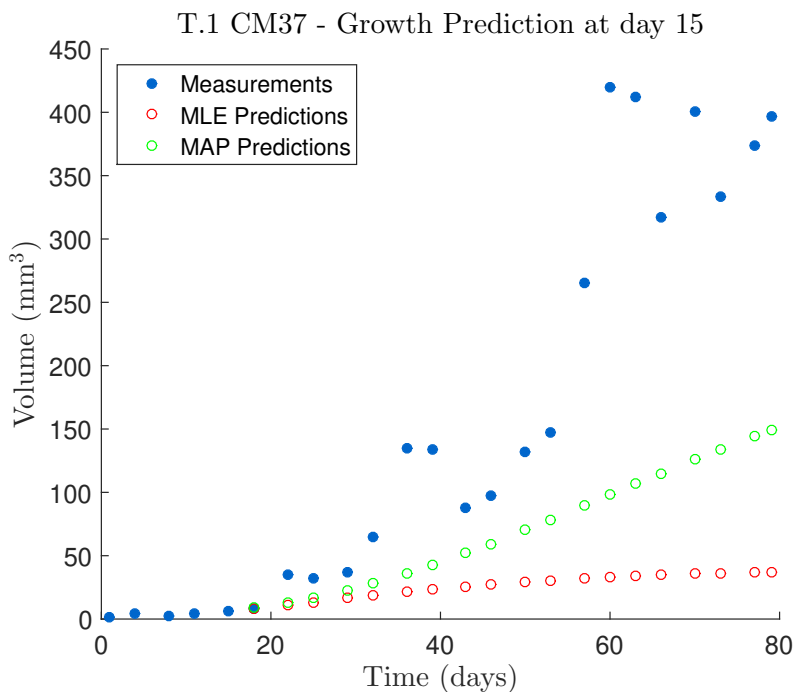


(ε) Κατανομή για την εκ των υστέρων σ.π.π. -- 5 Μετρήσεις

Σχήμα 4.16 Προβλέψεις εξέλιξης για τον όγκο 1 του ποντικίου CM37 με τη χρήση των εκτιμήσεων του ΕΜΠ και του ΕΜΥΠ.



(α') 3 μετρήσεις.



(β') 5 μετρήσεις.

Κεφάλαιο 5

Ανάλυση Αποτελεσμάτων

Στην εργασία αυτή μελετήθηκε η δυνατότητα εκτίμησης των παραμέτρων του μοντέλου Gompertz για την μοντελοποίηση και την πρόβλεψη της μελλοντικής εξέλιξης καρκινικών όγκων, με τη χρήση μεθόδων που βασίζονται στην αρχή της μέγιστης πιθανοφάνειας. Οι μέθοδοι αυτοί εφαρμόστηκαν τόσο σε συνθετικά, όσο και σε πειραματικά δεδομένα. Στο κεφάλαιο αυτό περιγράφονται τα συμπεράσματα που προέκυψαν από την ανάλυση που έγινε. Πριν την ανάλυση των αποτελεσμάτων, θα αναφερθούν κάποια γενικότερα θέματα που αφορούν τις μεθόδους που χρησιμοποιήθηκαν, τη μοντελοποίηση της ανάπτυξης των καρκινικών όγκων και την επιλογή του μαθηματικού μοντέλου.

Η επιλογή του μοντέλου Gompertz δεν αποκλείει τη χρήση κάποιου άλλου μοντέλου σε συνδιασμό με τις μεθόδους που περιγράφηκαν στο Κεφάλαιο 3, μιας και οι μέθοδοι δεν περιορίζονται από το μοντέλο που έχει επιλεγεί. Κάθε μέθοδος, με ελάχιστες τροποποιήσεις μπορεί να εφαρμοστεί σε συνδιασμό με κάποιο άλλο μοντέλο που πιθανόν να μπορεί να περιγράψει καλύτερα κάποιο συγκεκριμένο σύνολο δεδομένων από καρκινικούς όγκους. Επιπλέον, είναι προφανές ότι οι μέθοδοι που υλοποιήθηκαν μπορούν να χρησιμοποιηθούν σε οποιαδήποτε εφαρμογή έχει παρόμοια δομή με το πρόβλημα που περιγράφεται στην παρούσα εργασία.

Τέλος, πριν παρουσιαστούν τα ευρήματά, είναι σημαντικό να αναφερθεί η σημασία της έρευνας τόσο με συνθετικά, όσο και με πειραματικά δεδομένα. Το κύριο πλεονέκτημα των συνθετικών δεδομένων είναι ότι παρέχουν έναν εύκολο τρόπο για την επικύρωση των μεθόδων εκτίμησης. Σε ένα συνθετικό σύνολο δεδομένων, οι πραγματικές τιμές των παραμέτρων που πρέπει να εκτιμηθούν είναι γνωστές και ως εκ τούτου η ικανότητα των μεθόδων να εκτιμούν τις άγνωστες παραμέτρους μπορεί να εκτιμηθεί αντικειμενικά. Ένα άλλο πλεονέκτημα είναι ότι είναι εύκολο να δημιουργηθεί ένα σύνολο δεδομένων οποιουδήποτε μεγέθους το οποίο θα χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων. Ωστόσο, υπάρχουν και σημαντικά μειονεκτήματα. Ένα βασικό μειονέκτημα είναι ότι στην περι-

πτωση που μελετάται το μοντέλο που χρησιμοποιείται για τη δημιουργία συνθετικών δεδομένων είναι μια προσέγγιση της μέσης ανάπτυξης του όγκου και δεν αντιστοιχεί στη διαδικασία της πραγματικής ανάπτυξης. Ένα άλλο μπορεί να βρεθεί στο στοχαστικό μοντέλο και πιο συγκεκριμένα στον θόρυβο της διαδικασίας και της μέτρησης. Παρόλο που η Γκαουσιανή κατανομή είναι μια συνηθισμένη επιλογή για την μοντελοποίηση του θορύβου, δεν υπάρχει καμία εγγύηση ότι μπορεί περιγράψει την αβεβαιότητα στην περίπτωση που μελετάται. Τα πλεονεκτήματα και τα μειονεκτήματα αντιστρέφονται όταν πρέπει να αντιμετωπίσουμε τα πειραματικά δεδομένα. Σε αυτήν την περίπτωση, δεν μπορεί να αμφισβητηθεί η αξιοπιστία των δεδομένων (εκτός φυσικά από την περίπτωση που μετρώνται με λανθασμένο τρόπο ή γίνεται λάθος στις μετρήσεις). Ωστόσο, ακόμη και αν δεχθούμε ότι το στοχαστικό μοντέλο μπορεί να περιγράψει τη δυναμική του συστήματος, δεν μπορούμε ποτέ να είμαστε σίγουροι για τις πραγματικές παραμέτρους του μοντέλου.

5.1 Συνθετικά δεδομένα

Οι προσομοιώσεις που έγιναν με τη χρήση συνθετικών δεδομένων έδειξαν ότι οι τεχνικές που προτείνονται στο κεφάλαιο 3 μπορούν να εκτιμήσουν με ακρίβεια και τις δύο παραμέτρους που περιγράφουν τη δυναμική της εξέλιξης του όγκου. Επιπλέον, οι εκτιμήσεις της φέρουσας χωρητικότητας είναι ρεαλιστικές, ακόμα στην περίπτωση που ο αριθμός των διαθέσιμων μετρήσεων είναι σχετικά μικρός.

Στην εργασία αυτή μελετήθηκαν τέσσερις διαφορετικές περιπτώσεις οι οποίες διακρίνονται ανάλογα με τη διαθέσιμη πληροφορία που υπάρχει όσον αφορά τον θόρυβο. Οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση των αποτελεσμάτων είναι το ΜΑΠΣ και η τ.α. του ΑΠΣ. Τα αποτελέσματα που προέκυψαν από τις προσομοιώσεις για τις τέσσερις πρώτες περιπτώσεις που περιγράφηκαν στο Κεφάλαιο 4.1.2 παρουσιάζονται στο σχήμα 4.1. Η πιο σημαντική περίπτωση που μελετήθηκε είναι η τέταρτη, όπου τόσο ο θόρυβος της διαδικασίας όσο και ο θόρυβος της μέτρησης είναι άγνωστοι. Αυτή η περίπτωση αντιστοιχεί στο πρόβλημα που χρήζει προς αντιμετώπιση κατά την μελέτη πειραματικών δεδομένων. Οι υπόλοιπες τρεις περιπτώσεις μελετήθηκαν κυρίως για να ελεγχθεί αν η γνώση των χαρακτηριστικών του θορύβου μπορεί να οδηγήσει στην πιο ακριβή εκτίμηση των παραμέτρων της δυναμικής της ανάπτυξης του καρκινικού όγκου. Τα αποτελέσματα όμως των προσομοιώσεων δείχνουν ότι η γνώση της πραγματικής τιμής των παραμέτρων που αναφέρονται στον θόρυβο δεν βελτιώνει την εκτίμηση για τον ρυθμό αύξησης και την φέρουσα χωρητικότητα. Σε αντίθεση με αυτό που αναμένεται διαισθητικά, η θεώρηση των χαρακτηριστικών του θορύβου ως αγνώστων παραμέτρων και η προσπάθεια για εκτίμηση τους, οδηγεί σε εκτιμήσεις για αυτά που διαφέρουν από τις πραγματικές τιμές,

με ταυτόχρονη όμως ακριβέστερη εκτίμηση του ρυθμού αύξησης και της φέρουσας χωρητικότητας. Συνεπώς, αν και οι μέθοδοι δεν καταφέρνουν να εκτιμήσουν με ακρίβεια τα χαρακτηριστικά του θορύβου, η αδυναμία αυτή δεν επηρεάζει αρνητικά την ικανότητα τους για εκτίμηση της δυναμικής της ανάπτυξης του καρκινικού όγκου, που είναι και ο βασικός στόχος.

Ένα σημαντικό επίσης εύρημα είναι ότι ενώ οι εκτιμήσεις που αφορούν τον ρυθμό αύξησης και τη φέρουσα χωρητικότητα βελτιώνονται καθώς ο αριθμός των διαθέσιμων μετρήσεων αυξάνεται έως ότου προσεγγίσουν ασυμπτωτικά κάποια βέλτιστη τιμή εκτίμησης, δεν συμβαίνει το ίδιο και για τις εκτιμήσεις των χαρακτηριστικών του θορύβου. Παρατηρείται επίσης, όπως είναι αναμενόμενο, ότι αφού ο όγκος φτάσει την φέρουσα χωρητικότητα, οι νέες μετρήσεις που θα προκύψουν δεν προσφέρουν ιδιαίτερη πληροφορία όσον αφορά τον ρυθμό αύξησης, και συνεπώς από αυτό το σημείο και μετά η βελτίωση της εκτίμησης του ρυθμού αύξησης είναι μικρή. Παρομοίως, οι αρχικές μετρήσεις κατά τη διάρκεια της αύξησης του όγκου δεν μπορούν να χρησιμοποιηθούν για να εκτιμηθεί με μεγάλη ακρίβεια η φέρουσα χωρητικότητα του όγκου. Παρόλα αυτά, αν και οι αρχικές εκτιμήσεις της φέρουσας χωρητικότητας δεν είναι ακριβείς, οι εκτιμήσεις που προκύπτουν είναι ρεαλιστικές. Γενικότερα, η εκτίμηση της φέρουσας χωρητικότητας παρουσιάζει σημαντικότερη βελτίωση αφού ο όγκος έχει φτάσει περίπου το 75% του μέγιστου μεγέθους του. Όταν ο όγκος έχει φτάσει την τιμή της φέρουσας χωρητικότητας, τότε αυτή μπορεί να εκτιμηθεί με μεγάλη ακρίβεια.

Από τις μεθόδους που εφαρμόστηκαν, ο ΕΜΠ στο απλοποιημένο μοντέλο είναι λιγότερο ακριβής σε σχέση με τον ΕΜΠ με αριθμητική ολοκλήρωση και το συνδυασμό του ΕΦΚ και της Μέγιστης Πιθανοφάνειας κατά την εκτίμηση του ρυθμού αύξησης και της φέρουσας χωρητικότητας. Όμως, παρά το γεγονός ότι η μέθοδος αυτή υστερεί σε ακρίβεια, εξακολουθεί να είναι πολύ χρήσιμη, για την εκτίμηση των παραμέτρων στα συνθετικά δεδομένα, αφού οι εκτιμήσεις που παρέχει είναι αρκετά καλές και ο χρόνος εκτέλεσης της είναι εξαιρετικά μικρός. Οι εκτιμήσεις αυτές μπορούν να χρησιμοποιηθούν σαν αρχικές συνθήκες στις υπόλοιπες μεθόδους, διότι παρατηρήθηκε ότι κατ' αυτόν τον τρόπο μπορεί να μειωθεί ο χρόνος εκτέλεσης των άλλων δύο μεθόδων, διότι το να ξεκινήσει η διαδικασία βελτιστοποίησης από ένα σχετικά καλό σημείο εκτίμησης είναι καλύτερο από το να ξεκινήσει από κάποιο τυχαίο. Ειδικά, στον ΕΜΠ με αριθμητική ολοκλήρωση παρατηρήθηκε ότι ο χρόνος εκτέλεσης μπορεί να μειωθεί ακόμα και στο μισό, αν χρησιμοποιηθούν σαν αρχικές συνθήκες οι εκτιμήσεις του ΕΜΠ στο απλοποιημένο μοντέλο. Ο συνδυασμός του ΕΦΚ και της Μέγιστης Πιθανοφάνειας παρείχε καλύτερες εκτιμήσεις από τον ΕΜΠ στο απλοποιημένο μοντέλο, τόσο για τον ρυθμό αύξησης το όγκου, όσο και για την φέρουσα χωρητικότητα. Ο χρόνος εκτέλεσης της μεθόδου είναι μεγαλύτερος σε

σχέση με αυτόν του ΕΜΠ στο απλοποιημένο μοντέλο, αλλά παραμένει επίσης σε πολύ χαμηλά επίπεδα. Τις ακριβέστερες εκτιμήσεις για τις δύο προαναφερθείσες παραμέτρους, στην πλειοψηφία των περιπτώσεων, τις παρείχε ο ΕΜΠ με αριθμητική ολοκλήρωση. Βέβαια, αν και ο χρόνος εκτέλεσης της συγκεκριμένης μεθόδου αυξάνεται σημαντικά καθώς αυξάνεται το πλήθος το διαθέσιμων μετρήσεων, δεν φτάνει σε απαγορευτικά επίπεδα.

Όσον αφορά τα χαρακτηριστικά του θορύβου, καμία μέθοδος δεν φάνηκε να υπερτερεί με σταθερότητα έναντι των άλλων μεθόδων, αλλά σε γενικότερες γραμμές, ο ΕΜΠ με αριθμητική ολοκλήρωση απέδιδε καλύτερα στην εκτίμηση των παραμέτρων θ_3 και θ_6 , ενώ ο συνδυασμός του ΕΦΚ και της Μέγιστης Πιθανοφάνειας απέδιδε καλύτερα στην εκτίμηση των παραμέτρων θ_4 και θ_5 .

Τέλος, το ΦΔΕΣ ήταν σε θέση να παρέχει πολύ ακριβείς εκτιμήσεις για την κατάσταση του συστήματος. Ακόμα και στα αρχικά στάδια ανάπτυξης του όγκου που αριθμός των μετρήσεων είναι μικρός και η εκτίμηση των αγνώστων παραμέτρων από τον ΕΜΠ δεν είναι τόσο ακριβής, η εκτίμηση των καταστάσεων του συστήματος είναι σε γενικές γραμμές αρκετά ακριβής. Η μέση τιμή του ΚΜΤΣ ακόμα και με μικρό αριθμό μετρήσεων είναι κοντά στο μηδέν και συνεχίζει να μειώνεται καθώς αυξάνεται ο αριθμός των διαθέσιμων μετρήσεων και η εκτίμηση των αγνώστων παραμέτρων βελτιώνεται. Για την υλοποίηση του ΦΔΕΣ χρησιμοποιήθηκαν 4096 δείγματα (σωματίδια). Ο χρόνος εκτέλεσης βρίσκεται σε λογικά πλαίσια ακόμα και για την ακολουθιακή υλοποίηση. Βέβαια, αν και ήδη πριν τα 1000 δείγματα η παράλληλη υλοποίηση έχει αρχίσει να υπερτερεί όσον αφορά το χρόνο εκτέλεσης, η διαφορά στο χρόνο εκτέλεσης αρχίζει να γίνεται ιδιαίτερα αισθητή από τα 4096 δείγματα και έπειτα, όπου η εκθετική αύξηση του χρόνου εκτέλεσης της ακολουθιακής υλοποίησης γίνεται ιδιαίτερα αισθητή.

5.2 Πειραματικά δεδομένα

Η μοντελοποίηση ανάπτυξης όγκων είναι ένα εγγενώς δύσκολο πρόβλημα το οποίο πρέπει να βασίζεται σε ελλιπή πληροφόρηση λόγω έλλειψης κατάλληλων πειραματικών δεδομένων (π.χ. θορυβώδεις μετρήσεις, περιορισμένα σε αριθμό σύνολα δεδομένων, περιορισμοί στην διαδικασία κλπ.). Παρόλα αυτά, δεν πρέπει να υποτιμάται σε καμία περίπτωση η πληροφορία που μπορούν να δώσουν έστω και τα ελλιπή αυτά σύνολα δεδομένων.

Όπως αναφέρθηκε και στο Κεφάλαιο 4.2.1, ακόμα και να μπορούσαμε να δεχτούμε ότι το μοντέλο που χρησιμοποιείται περιγράφει επακριβώς τη δυναμική της εξέλιξης των όγκων, κατά την μελέτη πειραματικών δεδομένων η γνώση των πραγματικών τιμών των αγνώστων παραμέτρων θα ήταν αδύνατη. Συνεπώς, το ΜΑΠΣ, το οποίο χρησιμοποιήθη-

κε ως μετρική αξιολόγησης στις προσομοιώσεις με τα συνθετικά δεδομένα, δεν μπορεί να χρησιμοποιηθεί για να διαπιστωθεί η ικανότητα των μεθόδων να εκτιμούν τις άγνωστες παραμέτρους στα πειραματικά δεδομένα. Επειδή στην περίπτωση αυτή η μοναδική διαθέσιμη πληροφορία είναι οι μετρήσεις της διάστασης των όγκων και με δεδομένο ότι κύριος στόχος είναι η πρόβλεψη της μελλοντικής τους ανάπτυξης, επιλέχθηκε να χρησιμοποιηθεί σαν μετρική στα πειραματικά δεδομένα η PMTA των εκτιμήσεων από τις μετρήσεις.

Στα πειραματικά δεδομένα εξετάστηκε η απόδοση όλων των υλοποιήσεων που βασίζονται στη αρχή της Μέγιστης Πιθανοφάνειας (βλ. Κεφάλαιο 3). Στην μεγαλύτερη πλειοψηφία των περιπτώσεων που εξετάστηκαν, ο ΕΜΠ με αριθμητική ολοκλήρωση είναι σε θέση να δώσει εκτιμήσεις για τις άγνωστες παραμέτρους που οδηγούν σε ακριβέστερες μελλοντικές προβλέψεις. Αν και σε ορισμένες περιπτώσεις οι άλλες υλοποιήσεις (πέραν του ΕΜΠ με αριθμητική ολοκλήρωση) αποδίδουν καλύτερα, η απόδοση τους δεν είναι αξιόπιστη, επειδή εξαρτάται σε μεγάλο βαθμό από τις αρχικές συνθήκες που χρησιμοποιούνται για την εκτίμηση. Για τον λόγο αυτό, στο μέρος της ανάλυσης των πειραματικών δεδομένων, παρουσιάζεται μόνο η σύγκριση του ΕΜΠ με αριθμητική ολοκλήρωση σε σχέση με την μεθοδολογία που προτείνεται στη βιβλιογραφία.

Στη συνέχεια αυτού του κεφαλαίου θα περιγραφούν τα συμπεράσματα που προέκυψαν από την ανάλυση των διαθέσιμων πειραματικών δεδομένων.

5.2.1 Πολυκυτταρικά Σφαιροειδή Όγκων

Στο υποκεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα της σύγκρισης μεταξύ των επιδόσεων του ΕΜΠ με αριθμητική ολοκλήρωση που περιγράφεται στο Κεφάλαιο 3.3 με τον εκτιμητή ΜΓΕΤ. Όσον αφορά τον εκτιμητή ΜΓΕΤ, εφαρμόστηκαν οι τρεις διαφορετικές προσεγγίσεις που περιγράφηκαν στο Κεφάλαιο 4.2.3. Στην πρώτη και τη δεύτερη, η φέρουσα χωρητικότητα ήταν φιξαρισμένη στην ΜΕΤ και την ΜΤΜ αντιστοίχα, ενώ για την τρίτη θεωρήθηκε άγνωστη και έπρεπε να εκτιμηθεί.

Όσον αφορά το σύνολο δεδομένων MCF7, όλες οι προσεγγίσεις του εκτιμητή ΜΓΕΤ είχαν καλύτερη απόδοση από τον ΕΜΠ. Αυτό θα μπορούσε να αποτελεί μια ένδειξη ότι το στοχαστικό μοντέλο Gompertz που χρησιμοποιήθηκε δεν είναι κατάλληλο για την περιγραφή τέτοιου είδους όγκων. Ωστόσο, μια πιο προσεκτική ματιά στο σχήμα 4.8 δείχνει ότι η ανάπτυξη του όγκου μέχρι τη 13^η μέτρηση (που είναι το χρονικό διάστημα στο οποίο προσπαθούμε να υπολογίσουμε τον ρυθμό αύξησης και τη φέρουσα χωρητικότητα για να προβλέψουμε τη μελλοντική ανάπτυξη) απέχει πολύ από τη φάση κορεσμού, πράγμα που σημαίνει ότι δεν υπάρχει αρκετή πληροφορία σχετικά με τη φέρουσα χωρητικότητα. Αυ-

τό είναι ένα κρίσιμο σημείο, δεδομένου ότι ο ΕΜΠ υπολογίζει τη φέρουσα χωρητικότητα απευθείας από τα δεδομένα.

Στο σύνολο δεδομένων U118 μπορούμε να δούμε μια διαφορετική συμπεριφορά. Η προσέγγιση του εκτιμητή ΜΓΕΤ με τη φέρουσα χωρητικότητα να έχει φιξαριστεί στη ΜΕΤ ήταν καλύτερη από τον ΕΜΠ στα πρώτα στάδια ανάπτυξης του όγκου (όταν υπήρχαν λίγες διαθέσιμες μετρήσεις). Καθώς η ανάπτυξη του όγκου προχωρούσε, ο ΕΜΠ έδωσε ακριβέστερες προβλέψεις στην πλειοψηφία των περιπτώσεων, παρόλο που η μέση τιμή της $PMTA_{ET}$ είχε ελαφρώς χαμηλότερη τιμή από αυτήν της $PMTA_{MP}$. Η συμπεριφορά αυτή είναι αναμενόμενη από τη στιγμή που η φάση κορεσμού για αυτό το σύνολο δεδομένων έχει επιτευχθεί κατά το χρονικό διάστημα που εξετάζεται. Στην περίπτωση που η φέρουσα χωρητικότητα θεωρήθηκε άγνωστη για τον εκτιμητή ΜΓΕΤ, οι επιδόσεις του, ειδικότερα στα αρχικά στάδια, χειροτέρευσαν και ο ΕΜΠ παρουσίασε σε γενικότερες γραμμές καλύτερες επιδόσεις. Τέλος, η χρήση της ΜΤΜ ως φέρουσας χωρητικότητας είχε θετική επίδραση στα αρχικά στάδια ανάπτυξης και αρνητική στα μεταγενέστερα.

Σε αντίθεση με τα δύο προηγούμενα σύνολα δεδομένων, το σύνολο 9L ήταν μεγαλύτερο τόσο ως προς το πλήθος των σφαιροειδών, όσο και ως προς το διάστημα παρακολούθησης. Επιπλέον, σε αυτό το σύνολο, τα σφαιροειδή φτάνουν νωρίς στη φάση κορεσμού. Ως αποτέλεσμα, στο συγκεκριμένο σύνολο δεδομένων, ο ΕΜΠ δίνει καλύτερα αποτελέσματα σε σχέση με τον εκτιμητή ΜΓΕΤ. Θεωρώντας την φέρουσα χωρητικότητα ως άγνωστη παράμετρο, η απόδοση του εκτιμητή ΜΓΕΤ βελτιώθηκε, αλλά ο ΕΜΠ εξακολούθησε να παρέχει ελαφρώς καλύτερες προβλέψεις. Τέλος, η χρησιμοποίηση της ΜΤΜ ως φέρουσας χωρητικότητας βελτίωσε επιπλέον σε σημαντικό βαθμό την ικανότητα πρόβλεψης του εκτιμητή ΜΓΕΤ στα αρχικά στάδια της ανάπτυξης του όγκου, αλλά στα μεταγενέστερα στάδια ανάπτυξης, ο ΕΜΠ εξακολουθούσε να παρέχει πιο ακριβείς προβλέψεις.

Συνοψίζοντας, η κλασική προσέγγιση απέδωσε καλύτερες προβλέψεις για τον καρκίνο του μαστού και το γλοιοβλάστωμα ανθρώπινης προέλευσης, ενώ ο ΕΜΠ και η προσέγγιση του εκτιμητή ΜΓΕΤ με τη χρήση της ΜΤΜ παρουσίασαν καλύτερη απόδοση στην πρόβλεψη της ανάπτυξης για το γλοιοβλάστωμα των ποντικών. Διαπιστώθηκε ότι ο τύπος του όγκου και ο χρόνος παρακολούθησης είναι παράμετροι οι οποίες παίζουν σημαντικότατο ρόλο στην απόδοση του ΕΜΠ, καθώς ορίζουν την ανάπτυξη του όγκου και συνεπώς το αν τα δεδομένα περιέχουν ή όχι αρκετή πληροφορία για τη φάση κορεσμού. Βέβαια, λόγω του μικρού αριθμού σφαιροειδών σε κάθε σύνολο δεδομένων, περαιτέρω έρευνα είναι αναγκαία για να θεωρηθούν τα παραπάνω συμπεράσματα αξιόπιστα. Επιπλέον, είναι σημαντικό να εξεταστεί εάν άλλα μοντέλα (πέραν του μοντέλου Gompertz) μπορούν να παρέχουν ακριβέστερη μοντελοποίηση της ανάπτυξης και καλύτερες προ-

βλέψεις. Με άλλα λόγια, αντί του στοχαστικού μοντέλου των εξισώσεων (1.3)--(1.4), ο ΕΜΠ μπορεί να συνδυαστεί με μοντέλα που μπορούν να περιγράψουν συγκεκριμένους τύπους καρκινικών όγκων καλύτερα από το μοντέλο Gompertz. Αυτό θα μπορούσε να οδηγήσει σε μια πιο ακριβή πρόβλεψη ανάπτυξης για αυτό το είδος όγκων. Επιπλέον, δεδομένου ότι υπάρχουν ενδείξεις ότι η εκ των προτέρων πληροφορία μπορεί να βελτιώσει τις προβλέψεις ανάπτυξης, χρησιμοποιώντας εκτιμητές που λαμβάνουν υπόψη τυχόν προηγούμενες γνώσεις σχετικά με τις άγνωστες παραμέτρους έχουν τη δυνατότητα να παρέχουν ακριβέστερα αποτελέσματα.

5.2.2 Δερματικοί Όγκοι

Κατά τη μελέτη των δερματικών όγκων στα διπλά διαγωνιδιακά, το πιο σημαντικό εύρημα που προέκυψε είναι ότι στην πλειονότητα των περιπτώσεων ο ΕΜΠ με αριθμητική ολοκλήρωση και με άγνωστη την τιμή της φέρουσας χωρητικότητας, έδωσε καλύτερες προβλέψεις σε σύγκριση με τον εκτιμητή ΜΓΕΤ, όπου η φέρουσα χωρητικότητα είχε σταθερή τιμή. Όπως μπορούμε να δούμε στον πίνακα 4.5, στις πρώτες τέσσερις εβδομάδες (3 έως 9 μετρήσεις), η $PMTA_{MΠ}$ είχε μικρότερη τιμή από την $PMTA_{ΕΤ}$ στο μεγαλύτερο πλήθος περιπτώσεων που εξετάστηκαν. Επιπλέον, στο θηκόγραμμα του σχήματος 4.13 παρατηρείται ότι η διαφορά μεταξύ της $PMTA_{ΕΤ}$ και $PMTA_{MΠ}$ είναι μετατοπισμένη προς τον θετικό άνω ημιχώρο, γεγονός που σημαίνει ότι γενικότερα ο ΕΜΠ παρείχε πιο ακριβείς προβλέψεις όσον αφορά τη $PMTA$. Μετά την τέταρτη εβδομάδα, μπορούμε να διαπιστώσουμε ότι το πλήθος των περιπτώσεων όπου η $PMTA_{MΠ}$ έχει χαμηλότερη τιμή από την $PMTA_{ΕΤ}$ αυξάνεται ακόμα περισσότερο και ότι η διαφορά μεταξύ της $PMTA_{ΕΤ}$ και $PMTA_{MΠ}$ μετατοπίζεται ακόμα περισσότερο προς τον θετικό άνω ημιχώρο.

Η κύρια αδυναμία του ΕΜΠ είναι ότι η εκτίμηση της φέρουσας χωρητικότητας μπορεί να δώσει πολύ μικρές τιμές οι οποίες είναι μη ρεαλιστικές, ειδικά στα αρχικά στάδια της ανάπτυξης όπου δεν έχει ξεκινήσει η ταχεία ανάπτυξη του όγκου και τα δεδομένα δεν περιέχουν αρκετή πληροφορία για τη φάση κορεσμού. Το πρόβλημα των μη ρεαλιστικά μικρών εκτιμήσεων για την φέρουσα χωρητικότητα έχει επίσης παρατηρηθεί από τους Loizides κ.α. [42], όταν προσπάθησαν να θεωρήσουν αυτήν την παράμετρο άγνωστη. Τα Σχήματα 4.14α' και 4.14β' δείχνουν ένα παράδειγμα αυτού του προβλήματος. Μπορούμε να δούμε ότι το μέγεθος του όγκου δεν αυξάνεται σημαντικά κατά τις πρώτες δύομισή εβδομάδες (6 μετρήσεις), και όπως αναμένεται σε τέτοιες περιπτώσεις, ο εκτιμητής δεν θα είναι σε θέση να εκτιμήσει την φέρουσα χωρητικότητα του όγκου, αφού οι πληροφορίες που περιέχεται στις μετρήσεις είναι ανεπαρκής. Αργότερα, όταν ο όγκος αρχίζει να αναπτύσσεται γρήγορα, η φέρουσα χωρητικότητα μπορεί να εκτιμηθεί καλύτερα και οι προβλέψεις για τη μελλοντική ανάπτυξη του όγκου βελτιώνονται. Υπήρχαν επίσης πο-

λύ λίγες περιπτώσεις, όπου η φέρουσα χωρητικότητα έφθανε σε μη ρεαλιστικά υψηλές τιμές. Παρόλο που οι υψηλές εκτιμήσεις της φέρουσας χωρητικότητας δεν έχουν τόσο αρνητικές επιπτώσεις στις προβλέψεις όσο οι χαμηλές, δεδομένου ότι οι εκτιμήσεις του ρυθμού αύξησης μπορούν να εξισορροπήσουν το αποτέλεσμα, εξακολουθούν να παρέχουν μη βέλτιστες προβλέψεις.

Προκειμένου να βελτιωθούν οι εκτιμήσεις σε τέτοιες περιπτώσεις και να προκύψουν καλύτερες προβλέψεις, εφαρμόστηκε ο ΕΜΥΠ. Ο ΕΜΥΠ χρησιμοποιεί τις διαθέσιμες εμπειρικές γνώσεις σχετικά με τις άγνωστες παραμέτρους που περιέχονται στην εκ των προτέρων σ.π.π. (π.χ. Σχήμα 4.15α'), έτσι ώστε να βελτιώσει τις εκτιμήσεις που παρέχει. Χρησιμοποιώντας τις παραμέτρους που μεγιστοποιούν την εκ των υστέρων σ.π.π. των σχημάτων 4.15γ', 4.15ε' για την πρόβλεψη της εξέλιξης του όγκου, η ικανότητα πρόβλεψης σε μερικές περιπτώσεις βελτιώθηκε αισθητά στα αρχικά στάδια της μελέτης του όγκου. Απαραίτητη προϋπόθεση για να μπορέσει να έχει θετικά αποτελέσματα η εφαρμογή του ΕΜΥΠ είναι να περιέχεται στα δεδομένα κάποιος βαθμός πληροφορίας για τη φάση κορεσμού. Σε διαφορετική περίπτωση, η κατανομή που προκύπτει για την πιθανοφάνεια, μπορεί ακόμα και να ακυρώσει πρακτικά την επιροή της εκ των προτέρων σ.π.π. των αγνώστων παραμέτρων. Καθώς το μέγεθος του όγκου μεγάλωνει, οι προβλέψεις του ΕΜΥΠ στην πλειοψηφία των περιπτώσεων υπολείπονται, τόσο σε σχέση με αυτές του ΕΜΠ, όσο και με αυτές του εκτιμητή ΜΓΕΤ. Η συμπεριφορά αυτή οφείλεται πιθανότατα στο ότι λόγω της έλλειψης μεγαλύτερου συνόλου δεδομένων, πρέπει να βασιστούμε σε ένα μικρό αριθμό δεδομένων για να δημιουργήσουμε την εκ των προτέρων κατανομή για τις παραμέτρους ανάπτυξης, με αποτέλεσμα αυτή να είναι πιθανότατα μη αξιόπιστη. Ένα μεγαλύτερο σύνολο δεδομένων είναι απαραίτητο για να δημιουργηθεί μια πιο αξιόπιστη εκ των προτέρων κατανομή πιθανότητας για τις άγνωστες παραμέτρους, η οποία θα μπορούσε πιθανότατα να παρέχει καλύτερες εκτιμήσεις και προβλέψεις.

Κεφάλαιο 6

Συμπεράσματα

Στην παρούσα εργασία έγινε χρήση του μοντέλου Gompertz με σκοπό να μοντελοποιηθεί η διαδικασία της ανάπτυξης καρκινικών όγκων και να προβλεφθεί η μελλοντική πορεία της εξέλιξης τους. Υλοποιήθηκαν διάφορες αναδρομικές τεχνικές εκτίμησης παραμέτρων και καταστάσεων, με σκοπό την εκτίμηση των αγνώστων παραμέτρων της συνάρτησης Gompertz, των χαρακτηριστικών του θορύβου της διαδικασίας και της μέτρησης, και των καταστάσεων του συστήματος. Οι τεχνικές αυτές, αρχικά αξιολογήθηκαν με βάση την αποτελεσματικότητα και την ακρίβεια τους σε συνθετικά δεδομένα. Στη συνέχεια, εφαρμόστηκαν σε πειραματικά σύνολα δεδομένων και αξιολογήθηκαν με βάση την ικανότητα τους στην πρόβλεψη της μελλοντικής πορείας της εξέλιξης των όγκων σε σχέση με την συνηθισμένη μεθοδολογία που προτείνεται στη βιβλιογραφία.

Από τις προσομοιώσεις στα συνθετικά δεδομένα διαπιστώθηκε ότι οι άγνωστες παράμετροι που περιγράφουν τη δυναμική της ανάπτυξης των συνθετικών καρκινικών όγκων μπορούν να εκτιμηθούν με ακρίβεια, ακόμα και όταν η εκτίμηση για τις παραμέτρους του θορύβου δεν είναι ακριβής. Επίσης, παρατηρήθηκε ότι οι εκτιμήσεις της δυναμικής του όγκου μπορούν να είναι ρεαλιστικές, ακόμα και για μικρό αριθμό μετρήσεων. Όσον αφορά τις μεθόδους που υλοποιήθηκαν, ο ΕΜΠ με αριθμητική ολοκλήρωση παρείχε καλύτερες εκτιμήσεις για τις άγνωστες παραμέτρους σε σχέση με τις άλλες μεθόδους. Τέλος, το ΦΔΕΣ ήταν σε θέση να παρέχει αρκετά ακριβείς εκτιμήσεις για την κατάσταση του συστήματος σε εύλογο χρονικό διάστημα.

Κατά την εργασία με τα συνθετικά δεδομένα παρατηρήθηκε ότι η εφαρμογή εξειδικευμένων τεχνικών για την εκτίμηση των αγνώστων παραμέτρων παρείχε ακριβέστερα αποτελέσματα και καμπύλες που ταίριαζαν καλύτερα στα δεδομένα, σε σχέση με τη συνηθισμένη μέθοδο εκτίμησης που βασίζεται στον εκτιμητή των ΜΓΕΤ. Κρίθηκε συνεπώς επιτακτική η ανάγκη να ερευνηθεί εάν τα ίδια συμπεράσματα μπορούν να προκύψουν κατά την εφαρμογή των τεχνικών αυτών σε πειραματικά δεδομένα.

Όσον αφορά τα πειραματικά δεδομένα χρησιμοποιήθηκαν δύο τύποι δεδομένων, πολυκυτταρικά σφαιροειδή όγκων που αποκτήθηκαν σε δοκιμαστικούς σωλήνες και μετρήσεις από δερματικούς όγκους που προκλήθηκαν σε εν ζώη ποντίκια από μια ανοσοκατασταλτική και καρκινογενετική ουσία. Στα δεδομένα αυτά εφαρμόστηκε ο ΕΜΠ με αριθμητική ολοκλήρωση και συγκρίθηκε με την μέθοδο των ΜΓΕΤ, ως προς την ικανότητα πρόβλεψης της μελλοντικής εξέλιξης καρκινικών όγκων με βάση την εκτίμηση που έγινε για τις άγνωστες παραμέτρους του μοντέλου Gompertz. Επιπλέον, στα δεδομένα από τους δερματικούς όγκους εφαρμόστηκε ο ΕΜΥΠ για να ερευνηθεί εάν η διαθέσιμη εμπειρική πληροφορία που αφορά τις άγνωστες παραμέτρους μπορεί να βελτιώσει την πρόβλεψη της εξέλιξης του όγκου.

Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι ο ΕΜΠ με αριθμητική ολοκλήρωση που προτείνεται σε αυτή την εργασία και θεωρεί άγνωστες παραμέτρους τόσο τον ρυθμό αύξησης, όσο και την φέρουσα χωρητικότητα του μοντέλου Gompertz, ήταν σε θέση παρέχει καλύτερες προβλέψεις για την εξέλιξη των όγκων στα σφαιροειδή που προέρχονται από κυτταρική γραμμή ποντικών, αλλά υστερούσε σε απόδοση στις προβλέψεις της εξέλιξης των σφαιροειδών ανθρώπινης προέλευσης σε σχέση με την κλασική μεθοδολογία που βασίζεται στα ΜΓΕΤ και προτείνεται στη βιβλιογραφία. Για τους δερματικούς όγκους σε εν ζώη ποντίκια, ο ΕΜΠ με αριθμητική ολοκλήρωση κατάφερε να δώσει τις πιο ακριβείς προβλέψεις για την μελλοντική εξέλιξη των όγκων. Οι αξιολογήσεις των προβλέψεων στα πειραματικά δεδομένα έγινε με βάση την μετρική της ΡΜΤΑ. Τέλος, ο ΕΜΥΠ κατάφερε να βελτιώσει σε ορισμένες περιπτώσεις την πρόβλεψη της μελλοντικής εξέλιξης στα αρχικά στάδια μελέτης του όγκου, σημείο στο οποίο υστερούσε σημαντικά ο ΕΜΠ.

Αν και μεθολογία που αναπτύχθηκε δεν μπορεί να αποτελέσει μια γενική προσέγγιση που μπορεί να προβλέψει τον τρόπο με τον οποίο θα εξελιχθεί κάθε είδους όγκος κυρίως λόγω των περιορισμών που συνεπάγεται η χρησιμοποίηση του μοντέλου Gompertz, η ανάλυση των δεδομένων αυτής της εργασίας έδειξε πως η προσέγγιση που προτείνεται θα μπορούσε να χρησιμοποιηθεί επιτυχώς για συγκεκριμένους τύπους καρκινικών όγκων. Βέβαια, μιας και το πλήθος των πειραματικών δεδομένων που χρησιμοποιήθηκαν ήταν σχετικά μικρό, η περαιτέρω έρευνα σε μεγαλύτερα σύνολα δεδομένων για να επιβεβαιωθούν τα αποτελέσματα που προέκυψαν από την παρούσα εργασία είναι απαραίτητη. Σε κάθε περίπτωση όμως, τα αποτελέσματα είναι ιδιαίτερα ενθαρρυντικά όσον αφορά την ικανότητα της εξατομικευμένης πρόβλεψης να παρέχει ακριβείς εκτιμήσεις. Επιπλέον, οι αδυναμίες μοντελοποίησης που παρουσιάστηκαν κατά την ανάλυση των πειραματικών δεδομένων είναι δυνατόν να ξεπεραστούν με τη χρήση κάποιου άλλου μοντέλου που να περιγράφει καλύτερα τους συγκεκριμένους τύπους καρκινικών όγκων. Για το λόγο αυτό, θα είχε ιδιαίτερο ενδιαφέρον περαιτέρω έρευνα με τη χρήση διαφορετικών μοντέλων.

Οι τεχνικές που υλοποιήθηκαν δεν περιορίζονται από το μοντέλο Gompertz και μπορούν με μικρές τροποποιήσεις να χρησιμοποιηθούν με κάποιο άλλο μοντέλο, που πιθανώς να μπορεί να περιγράψει καλύτερα συγκεκριμένα είδη όγκων.

Είναι δεδομένο ότι η παραμικρή βελτίωση στην πρόβλεψη και περιγραφή της ανάπτυξης καρκινικών όγκων μπορεί να είναι πολύ σημαντική. Αν και το μεγάλο μειονέκτημα των εξατομικευμένων μεθόδων είναι η υπολογιστική τους πολυπλοκότητα, η σημερινή τεχνολογία μπορεί να παρακάμψει αυτό το εμπόδιο και να δώσει αποτελέσματα σε εύλογα χρονικά διαστήματα. Η εξατομικευμένη μοντελοποίηση της ανάπτυξης των όγκων, στην οποία λαμβάνεται υπόψη η ετερογένεια που μπορεί να παρουσιάσει ο κάθε όγκος, σε συνδιασμό με την εφαρμογή εξειδικευμένων τεχνικών για την εκτίμηση των αγνώστων παραμέτρων των μοντέλων, φαίνεται πως μπορεί να οδηγήσει σε πιο ακριβή πρόβλεψη της πορείας ενός όγκου και η χρήση της αρχής της Μέγιστης Πιθανοφάνειας είναι μια γενικότερη προσέγγιση που μπορεί να βοηθήσει προς αυτό τον σκοπό.

Μέρος ΙΙ

Θαλάσσια Παρακολούθηση

Κεφάλαιο 7

Εισαγωγή

7.1 Παρακολούθηση Θαλάσσιας Κίνησης

Η ραγδαία τεχνολογική εξέλιξη που παρατηρείται στις μέρες μας, έχει δώσει τη δυνατότητα στα πλοία (είτε πρόκειται για μικρά ταχύπλοα, είτε για μεγάλα πλοία) να εξοπλίζονται με ηλεκτρονικές συσκευές, οι οποίες σε πραγματικό χρόνο αποστέλουν δεδομένα που περιέχουν πληροφορίες σχετικά με την τοποθεσία και την κίνηση τους. Τα δεδομένα αυτά, αφού επεξεργαστούν με κατάλληλο τρόπο, μπορούν να χρησιμοποιηθούν για την παρακολούθηση της θαλάσσιας κίνησης και της αλληλεπίδρασης μεταξύ των πλοίων. Η παρακολούθηση της θαλάσσιας κίνησης είναι πολύ σημαντική, τόσο για την ασφάλεια των σκαφών και των επιβατών, όσο και για την ανίχνευση παράνομης δραστηριότητας (π.χ. λαθρεμπόριο, παράνομη μετανάστευση κ.ά.).

Το Σύστημα Αυτόματης Αναγνώρισης (Automatic Identification System) [80] είναι ένα ηλεκτρονικό σύστημα που μπορεί να ενσωματωθεί σε ένα πλοίο και να παρέχει δεδομένα σε πραγματικό χρόνο σχετικά με τη γεωγραφική θέση, την πορεία και την ταχύτητα του. Τα δεδομένα από το σύστημα αυτό μπορούν να χρησιμοποιηθούν για την παρακολούθηση της θαλάσσιας κίνησης. Όμως, ο όγκος των δεδομένων που παράγονται από το ΣΑΑ είναι τεράστιος, με αποτέλεσμα η χρήση αυτοματοποιημένων διαδικασιών για την επεξεργασία τους να είναι απαραίτητη.

7.2 Ανασκόπηση Βιβλιογραφίας

Στη βιβλιογραφία υπάρχει πληθώρα τεχνικών που προτείνονται για τη λύση του προβλήματος της παρακολούθησης της μη κανονικής συμπεριφοράς πλοίων. Οι Mascaro κ.ά. [81] προτείνουν τη χρήση Μπεϋζιανών Δικτύων, τα οποία εκπαιδεύονται με βάση πραγ-

ματικά δεδομένα που έχουν συλλεγεί από το ΣΑΑ και την παραγωγή μοντέλων σε δύο διαφορετικές χρονικές κλίμακες, τόσο για την παρακολούθηση σαν σύνολο, όσο και μεμονωμένα, με σκοπό την εκμάθηση της ανίχνευσης μη κανονικών συμπεριφορών. Οι Johanson και Falkman [82] προτείνουν επίσης τη χρήση Μπεϋζιανών Δικτύων για την ανίχνευση μη κανονικής συμπεριφοράς των πλοίων λόγω της ικανότητας που έχουν αυτά τα δίκτυα να συμπεριλάβουν γνώση από ειδικούς στα μοντέλα και της απλότητας στην κατανόηση και αντίληψη, και εφαρμόζουν την προσέγγιση αυτή σε συνθετικά δεδομένα. Η χρήση όμως των Μπεϋζιανών Δικτύων έχει το μειονέκτημα ότι τα δεδομένα του ΣΑΑ χρειάζονται προεπεξεργασία πριν από τη χρησιμοποίησή τους. Μια άλλη τεχνική, η οποία χρησιμοποιείται ευρέως, είναι η δημιουργία μοντέλων τροχιάς με τη χρήση Γενετικού Προγραμματισμού, η οποία όμως έχει το μειονέκτημα της αυξημένης πολυπλοκότητας. Οι Kowalska και Peel [83] προτείνουν τη χρήση ενός Μπεϋζιανού μοντέλου σε συνδυασμό με μια τεχνική Μηχανικής Μάθησης, η οποία ονομάζεται Ενεργής Μάθηση, με σκοπό τον υπολογισμό του μοντέλου. Οι Will κ.ά. [84] παρουσιάζουν ένα υπερσύγχρονο μη παραμετρικό μοντέλο παλινδρόμησης που βασίζεται σε Γκαουσιανές Διαδικασίες και κατασκευάζεται από το ΣΑΑ. Στην εργασία αυτή χρησιμοποιούνται επίσης K-D δέντρα με σκοπό τη μείωση της πολυπλοκότητας. Τέλος, μια ακόμα δημοφιλής λύση είναι ο συνδυασμός του ΣΑΑ με άλλα συστήματα παρακολούθησης, όπως για παράδειγμα ειδικά ραντάρ [85, 86].

7.3 Περιγραφή του προβλήματος

Είναι ευρέως γνωστό ότι στον Ευκλείδειο χώρο η πιο σύντομη απόσταση μεταξύ δύο σημείων είναι η ευθεία γραμμή. Αν η κίνηση των πλοίων στη θάλασσα προβληθεί σε ένα δισδιάστατο χώρο, τότε η κίνηση τους παρουσιάζει μεγάλα ευθύγραμμα τμήματα. Για το λόγο αυτό είναι δυνατόν να χρησιμοποιηθούν Γραμμικοί Εκτιμητές για να προβλεφθεί η πορεία τους. Επιπλέον, επειδή το ΣΑΑ είναι ένα σύστημα του οποίου η πληροφορία είναι δυνατόν να πλαστογραφηθεί, μπορεί να γίνει έλεγχος των δεδομένων που αποστέλλονται από τα πλοία, με σκοπό να ελεγχθεί η εγκυρότητα τους. Στην παρούσα εργασία επικεντρωνόμαστε στην εκτίμηση της μελλοντικής πορείας ενός πλοίου με βάση τις παρελθοντικές πληροφορίες σχετικά με την κίνηση και την τοποθεσία του και στον ορισμό περιπτώσεων συναγερμού με σκοπό την αναγνώριση μη κανονικής συμπεριφοράς από τα πλοία ή πλαστογραφία στα δεδομένων του ΣΑΑ.

Στο δεύτερο μέρος της εργασίας παρουσιάζονται ο αλγόριθμος που αναπτύχθηκε για την εκτίμηση της μελλοντικής πορείας των πλοίων, οι περιπτώσεις που ορίστηκαν να αποτελούν μη κανονικές συμπεριφορές και η εφαρμογή λογισμικού που αναπτύχθηκε για

να ελέγξει την αποτελεσματικότητα του προτεινόμενου αλγόριθμου. Οι προσομοιώσεις που πραγματοποιήθηκαν έγιναν με πραγματικά δεδομένα που συλλέχθηκαν από τον Σαρωνικό Κόλπο και δείχνουν ότι ο προτεινόμενος αλγόριθμος μπορεί να εφαρμοστεί στην πράξη και να δώσει εύρωστα αποτελέσματα. Επιπλέον, η μέθοδος που προτείνεται έχει το πλεονέκτημα ότι μπορεί να εφαρμοστεί για την παρακολούθηση της θαλάσσιας κίνησης σε πραγματικό χρόνο. Το εργαλείο που αναπτύχθηκε μπορεί να ενσωματωθεί εύκολα σε οποιοδήποτε σύστημα και να βοηθήσει τις αρχές στην ανίχνευση μη κανονικών συμπεριφορών στη θάλασσα. Επίσης, δείχνει ότι οι Γραμμικοί Εκτιμητές μπορούν να χρησιμοποιηθούν αποτελεσματικά για την πρόγνωση της πορείας των πλοίων. Ο Γραμμικός Εκτιμητής που χρησιμοποιήθηκε είναι το AR (AutoRegressive -- Αυτοπαλίνδρομο) μοντέλο, που είναι ειδική περίπτωση και βασικό συστατικό του ARMA (AutoRegressive Moving Average -- Αυτοπαλίνδρομο Κινητού Μέσου) μοντέλου.

Στο κεφάλαιο 8 παρουσιάζεται η μέθοδος για την πρόβλεψη της μελλοντικής πορείας του πλοίου, καθώς επίσης και τα διάφορα κριτήρια συναγερμού που ορίστηκαν για την παρακολούθηση της θαλάσσιας κίνησης και την ακεραιότητα των δεδομένων. Στο κεφάλαιο 9 περιγράφεται η εφαρμογή λογισμικού που δημιουργήθηκε με βάση την προτεινόμενη μεθοδολογία, στο κεφάλαιο 10 παρουσιάζονται και συζητούνται τα αποτελέσματα των προσομοιώσεων, και τέλος στο κεφάλαιο 11 παρουσιάζονται τα συμπεράσματα.

Κεφάλαιο 8

Μεθοδολογία

Το πρώτο μέρος αυτού του κεφαλαίου παρουσιάζει τη μέθοδο που χρησιμοποιήθηκε για να προβλέψει την μελλοντική πορεία ενός πλοίου με βάση τις προηγούμενες τοποθεσίες στις οποίες βρέθηκε. Στο δεύτερο μέρος παρουσιάζονται οι ορισμοί των ασυνεπειών που θεωρούνται περιπτώσεις συναγερμού.

8.1 Εκτίμηση Πορείας Πλοίου

Το ΣΑΑ παρέχει δεδομένα με τα οποία μπορεί να γίνει ταυτοποίηση ενός πλοίου, καθώς επίσης και εύρεση της τοποθεσίας, της ταχύτητας και της τροχιάς του. Για τον προσδιορισμό της τοποθεσίας του πλοίου χρησιμοποιούνται οι γεωγραφικές συντεταγμένες στις οποίες βρίσκεται, δηλαδή το γεωγραφικό μήκος και πλάτος. Εφαρμόζοντας Ανάλυση Χρονοσειρών σε αυτά τα δεδομένα είναι δυνατόν να δημιουργηθεί ένα μοντέλο και με βάση αυτό να γίνει πρόβλεψη των μελλοντικών τοποθεσιών στις οποίες αναμένεται να βρεθεί το πλοίο. Γενικότερα, η ραγδαία ανάπτυξη της υπολογιστικής ισχύος που έχει επιτευχθεί τις τελευταίες δεκαετίες έχει συμβάλει σημαντικά στη διάδοση της χρήσης τεχνικών επεξεργασίας ψηφιακών σημάτων διακριτού χρόνου σε μεγάλο πλήθος εφαρμογών, με αξιοσημείωτο παράδειγμα τον τομέα των τηλεπικοινωνιών [87, 88]. Στην παρούσα εργασία για την πρόβλεψη της μελλοντικής πορείας των πλοίων χρησιμοποιήθηκε το AR μοντέλο, ένα μοντέλο που χρησιμοποιείται ευρέως στην Ανάλυση Χρονοσειρών και είναι μια ειδική περίπτωση του ARMA μοντέλου [87, 89, 90]. Το μοντέλο ARMA είναι ένα θεμελιώδες μοντέλο στον τομέα της πρόγνωσης [91, 92], με εφαρμογή στις οικονομικές επιστήμες [93], την πρόγνωση του καιρού [94, 95], την πρόβλεψη ηλεκτρικού φορτίου [96], την επεξεργασία ψηφιακών σημάτων φωνής [97] κλπ.

Στο AR μοντέλο η έξοδος y εκφράζεται σαν ένας γραμμικός συνδυασμός των παρελθοντικών εξόδων y και μιας διαταραχής λευκού θορύβου $w(t)$:

$$y(t) = - \sum_{j=1}^{n_1} a_j y(t-j) + w(t). \quad (8.1)$$

Χρησιμοποιώντας τον συμβολισμό q σαν τελεστή ολίσθησης προς τα πίσω, το μοντέλο (8.1) μπορεί να εκφραστεί ως:

$$A(q)y(t) = w(t), \quad (8.2)$$

όπου

$$A(q) = 1 + a_1 q^{-1} + \dots + a_{n_1} q^{-n_1}, \quad (8.3)$$

Στην περίπτωση που στο μοντέλο περιέχονται οι εξωγενείς είσοδοι $u(t)$ το μοντέλο ονομάζεται αυτοπαλίνδρομο μοντέλο με εξωγενείς όρους (AutoRegressive with eXogenous terms) και η δομή του είναι:

$$A(q)y(t) = B(q)u(t) + w(t). \quad (8.4)$$

όπου

$$B(q) = (b_1 + b_2 q^{-1} + \dots + b_{n_2} q^{-n_2+1}) q^{-d} \quad (8.5)$$

και η μεταβλητή d αντιπροσωπεί τη χρονική καθυστέρηση.

Αν στο γραμμικό σύστημα που περιγράφεται από το ARX μοντέλο προστεθεί μια ανεξάρτητη ακολουθία θορύβου $\{w(t)\}$, τότε προκύπτει το αυτοπαλίνδρομο μοντέλο κινητού μέσου με εξωγενείς όρους (AutoRegressive Moving Average with eXogenous terms):

$$A(q)y(t) = B(q)u(t) + C(q)w(t) \quad (8.6)$$

όπου $C(q)$ είναι ένα φίλτρο της μορφής:

$$C(q) = I_0 + c_1 q^{-1} + \dots + c_{n_3} q^{-n_3} \quad (8.7)$$

Η ονομασία ARMA εισάγεται για το μοντέλο (8.6) όταν $u(t) \equiv 0$ και συνεπώς:

$$A(q)y(t) = C(q)w(t). \quad (8.8)$$

Τα δεδομένα που χρησιμοποιήθηκαν για τη δημιουργία του AR μοντέλου προέρχονται από δεδομένα που συλλέχθηκαν στον Σαρωνικό Κόλπο μέσω του ΣΑΑ. Πιο συγκεκριμένα, τα δεδομένα που χρειάζονται είναι η γεωγραφική θέση (μήκος και πλάτος), καθώς

επίσης και ο χρονικός ορίζοντας στον οποίο έγιναν οι μετρήσεις. Το σύστημα στέλνει δεδομένα ανά τακτά χρονικά διαστήματα, η συχνότητα αποστολής των οποίων εξαρτάται από την ταχύτητα με την οποία κινείται το πλοίο. Όμως, η ανάλυση χρονοσειρών που γίνεται, απαιτεί το χρονικό διάστημα μεταξύ των μετρήσεων να είναι σταθερό. Επομένως, για να μπορεί να γίνει ανάλυση χρονοσειρών στα δεδομένα του ΣΑΑ απαιτείται προεπεξεργασία των δεδομένων. Αφού γίνει επαναδειγματοληψία με τη χρήση γραμμικής παρεμβολής, οι νέες τιμές προέρχονται ανά τακτά χρονικά διαστήματα T . Μετά την προεπεξεργασία των δεδομένων το πρόβλημα προς επίλυση περιγράφεται ως εξής: Γνωρίζοντας την θέση ενός πλοίου τις χρονικές στιγμές $T, 2T, \dots, kT$, όπου $k \in \mathbb{N}$, να εκτιμηθεί η θέση του πλοίου τις χρονικές στιγμές $(k+1)T, \dots, (k+n)T$, όπου $n \in \mathbb{N}$.

Το παραπάνω πρόβλημα μπορεί να επιλυθεί με τη χρήση του AR μοντέλου. Η μαθηματική περιγραφή του μοντέλου είναι η εξής:

$$\hat{Y}_{k+1} = \sum_{i=m}^k A_i Y_i \quad (8.9)$$

όπου Y_k είναι το διάνυσμα θέσης του πλοίου τη χρονική στιγμή kT (2×1 διάνυσμα, όπου $y_{1,1}$ είναι το γεωγραφικό μήκος και $y_{2,1}$ είναι το γεωγραφικό μήκος), $m < k$ όπου $m \in \mathbb{N}$, και i είναι ο δείκτης των παρελθοντικών χρονικών στιγμών. Η μήτρα A_i έχει διάσταση 2×2 και οι τιμές της είναι πραγματικοί αριθμοί. Για τον υπολογισμό των στοιχείων $A_{i,(j,k)}$ των μητρών A_i χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων. Στη συνέχεια, οι θέσεις τις χρονικές στιγμές $(k+n)T$ μπορούν να εκτιμηθούν από την εξίσωση:

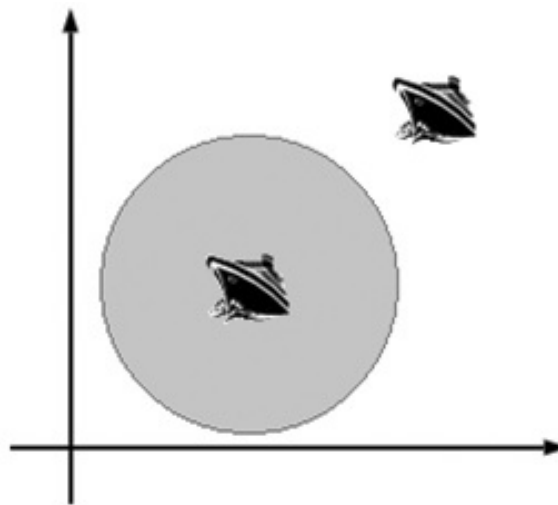
$$\hat{Y}_{k+n} = \sum_{i=0}^{n-2} A_{k-i} \hat{Y}_{k-i+n-1} + \sum_{i=0}^{k-m-n+1} A_{k-i-n+1} Y_{k-i} \quad (8.10)$$

8.2 Κριτήρια Συναγερμών

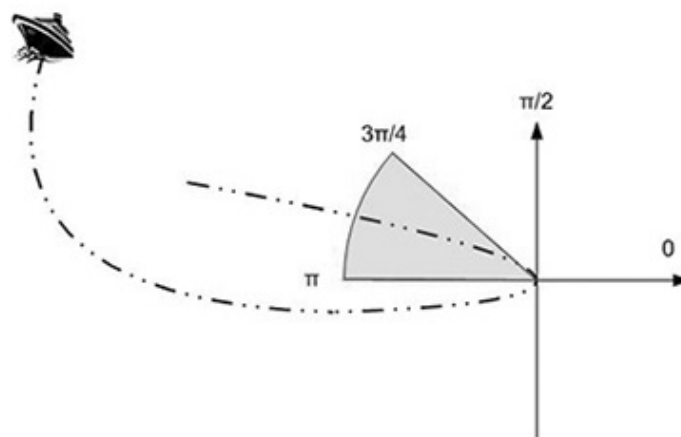
Στη συνέχεια θα οριστούν κάποια κριτήρια, η παραβίαση των οποίων συνιστά μη κανονική συμπεριφορά και παράγει συναγερμό. Αυτοί οι συναγερμοί μπορούν ελεγχθούν περαιτέρω από τις αρχές που παρακολουθούν την κίνηση στη θάλασσα και να οδηγήσουν σε πιο αποτελεσματική εποπτεία. Τα κριτήρια συναγερμών είναι τα ακόλουθα:

- Απόκλιση της εκτιμώμενης τοποθεσίας του πλοίου από την πραγματική τοποθεσία (Σχήμα 8.1α'). Αν η θέση που εκτιμήθηκε βρίσκεται εκτός του κύκλου που έχει κέντρο την πραγματική θέση του πλοίου και ακτίνα η οποία εξαρτάται από τον τύπο του πλοίου, τότε η ταχύτητα του πλοίου έχει αλλάξει. Αυτή η αλλαγή είναι πιθανόν να σημαίνει κάποιου είδους ανωμαλία η οποία να χρήζει περαιτέρω διερεύνησης.

Σχήμα 8.1 Κριτήρια Συναγερμών.



(α') Κριτήριο απόκλισης τοποθεσίας.



(β') Κριτήριο απόκλισης πορείας.

- Αλλαγή τροχιάς κίνησης σε ένα συγκεκριμένο εύρος γωνίας (π.χ. $3\pi/4, \pi$) η οποία έχει συντελεστεί σε ένα δεδομένο χρονικό διάστημα (π.χ. 5 -- 10 λεπτά) (Σχήμα 8.1β'). Η μεγάλη αλλαγή στην πορεία ενός πλοίου είναι πιθανόν να συνιστά μη κανονική συμπεριφορά.
- Η μέγιστη επιτρεπόμενη ταχύτητα για ένα πλοίο εξαρτάται από τον τύπο του. Αν για κάποιο λόγο η ταχύτητα ενός πλοίου ξεπεράσει αυτή τη μέγιστη επιτρεπόμενη τιμή, τότε σηματοδοτείται ένας συναγερμός.

Το ΣΑΑ παρέχει δεδομένα τόσο για την τοποθεσία (γεωγραφικό μήκος και πλάτος), όσο και για την κίνηση ενός πλοίου (ταχύτητα, πορεία). Η δυνατότητα αναγνώρισης της εγκυρότητας των δεδομένων αυτών είναι πολύ σημαντική. Χρησιμοποιώντας ξεχωριστά

τα δεδομένα τοποθεσίας και κίνησης, είναι δυνατόν να αναγνωρισθεί αν έχει γίνει κάποια πλαστογράφηση των δεδομένων, στις περιπτώσεις που ένα μόνο από τα δύο είδη δεδομένων έχει επεξεργαστεί, ή και τα δύο είδη έχουν υποστεί επεξεργασία αλλά τα αποτελέσματα τους δεν ταιριάζουν.

- Με βάση την ταχύτητα και την τοποθεσία, αν $v \cdot \Delta t \gg s = \|y_k - y_{k-1}\|$, τότε ενεργοποιείται ένας συναγερμός. y_k και y_{k-1} είναι οι τοποθεσίες του πλοίου τις χρονικές στιγμές k και $k-1$, v είναι η ταχύτητα του πλοίου και Δt είναι το χρονικό διάστημα μεταξύ k και $k-1$. Θεωρώντας την ταχύτητα σταθερή καθώς το πλοίο κινείται, η παραβίαση της παραπάνω συνθήκης σηματοδοτεί τον συναγερμό.

Για να υπολογιστεί η απόσταση μεταξύ δύο διαδοχικών τοποθεσιών χρησιμοποιείται ο αλγόριθμος Vincenty [98], που αναπτύχθηκε το 1975 από τον Thaddeus Vincenty. Πρόκειται για δύο συσχετιζόμενες επαναληπτικές μεθόδους που χρησιμοποιούνται στη γεωδαισία για να υπολογιστεί η απόσταση μεταξύ δύο σημείων στην επιφάνεια ενός σφαιροειδούς. Βασίζονται στην υπόθεση ότι η γη έχει το σχήμα ενός πεπλατυσμένου σφαιροειδούς, και για το λόγο αυτό καταφέρνουν να υπολογίσουν καλύτερα την απόσταση μεταξύ δύο σημείων που βρίσκονται στην επιφάνεια σε σχέση με τις μεθόδους που θεωρούν τη γη σαν σφαίρα.

- Η μέγιστη δυνατή ταχύτητα ενός πλοίου εξαρτάται από τον τύπο του. Αν ένα πλοίο κινείται με ταχύτητα μεγαλύτερη από τη μέγιστη εφικτή, τότε σηματοδοτείται ένας συναγερμός.
- Η έλλειψη δεδομένων από το ΣΑΑ για σχετικά μεγάλα χρονικά διαστήματα σηματοδοτεί έναν συναγερμό. Το ΣΑΑ είναι σχεδιασμένο να στέλνει τακτικά δεδομένα και επομένως η έλλειψη δεδομένων για ένα μεγάλο χρονικό διάστημα πιθανόν σημαίνει κάποια μη κανονικότητα.

Κεφάλαιο 9

Εφαρμογή

Στο κεφάλαιο αυτό περιγράφεται η εφαρμογή λογισμικού που δημιουργήθηκε για την παρακολούθηση της πορείας των πλοίων και την αναγνώριση των μη κανονικών συμπεριφορών με βάση στοιχεία που συλλέχθηκαν από το ΣΑΑ. Το περιβάλλον στο οποίο αναπτύχθηκε η εφαρμογή είναι το πρόγραμμα Matlab (ver. R2015a, The MathWorks Inc., Natick, IL) και τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από πραγματικές εκπομπές του ΣΑΑ από πλοία που έπλεαν στο Σαρωνικό Κόλπο. Για να ελεγχθεί η αξιοπιστία και η ακεραιότητα των αποτελεσμάτων έγιναν πολυάριθμα πειράματα με πλήθος από διαφορετικές βάσεις δεδομένων.

Τα δεδομένα εισόδου εισάγονται στο πρόγραμμα με τη μορφή μιας SQLite βάσης δεδομένων. Στη συνέχεια, αφού γίνει προεπεξεργασία της βάση δεδομένων τα δεδομένα για το κάθε πλοίο εισάγονται σε μια δομή του Matlab. Η προεπεξεργασία που αναφέρθηκε περιλαμβάνει το βήμα επαναδειγματοληψίας (Βήμα 2, Αλγόριθμος 4) που απαιτείται έτσι ώστε να μπορεί να εφαρμοστεί η ανάλυση χρονοσειρών στα δεδομένα του ΣΑΑ και την απόρριψη πληροφοριών που στέλνει το ΣΑΑ, αλλά δεν είναι χρήσιμες για την συγκεκριμένη εφαρμογή.

Αν το πλοίο που ελέγχεται δεν είναι αγκυροβολημένο, το πρόγραμμα συγκρίνει την πορεία του πλοίου που αναφέρεται από το ΣΑΑ με την πορεία που έχει εκτιμηθεί ωρίτερα και ελέγχει αν συμβαίνει κάποια από τις μη κανονικότητες που αναφέρθηκαν στο Κεφάλαιο 8.2. Στη συνέχεια χρησιμοποιεί τα δεδομένα τοποθεσίας των προηγούμενων λεπτών για να παράξει ένα νέο AR μοντέλο και με βάση αυτό εκτιμά την μελλοντική πορεία του πλοίου.

Πιο συγκεκριμένα για τη δημιουργία του AR μοντέλου απαιτείται συγκεκριμένος ελάχιστος αριθμός μετρήσεων. Το πρόγραμμα είναι ρυθμισμένο ώστε να περιμένει μέχρι να συμπληρωθεί το απαραίτητο πλήθος δεδομένων. Στην προκειμένη περίπτωση απαιτείται να έχουν περάσει δέκα λεπτά από την πρώτη μέτρηση του ΣΑΑ, ώστε μετά την

Algorithm 4 Ψευδοκώδικας Παρακολούθηση Θαλάσσιας Κίνησης

Step 1: Load Data**Step 2:** Resample Data**Step 3:** Check Alert Criteria**Step 4:** Report Alerts**Step 5:** Create AR Model and Forecast Future Positions**Step 6:** Return to **Step 1**

επαναδειγματοληψία να είναι διαθέσιμες 11 μετρήσεις (μια ανά κάθε λεπτό). Αν κατά τη διάρκεια των δέκα λεπτών από την πρώτη μέτρηση δεν υπάρξουν αρκετά δεδομένα, τότε σηματοδοτείται ένας συναγερμός και τα βήματα 2 έως 5 του αλγορίθμου 4 δεν εκτελούνται για το συγκεκριμένο πλοίο μέχρι να συλλεγεί επαρκές πλήθος δεδομένων.

Στο βήμα 3 του αλγορίθμου 4 ελέγχονται οι συνθήκες που έχουν οριστεί να σηματοδοτούν τους συναγερμούς. Αν κάποιο από τα κριτήρια πληρείται, τότε η ασυνέπεια αυτή καταγράφεται στο βήμα 4 του αλγορίθμου. Κάθε είδος συναγερμού έχει τον δικό του κωδικό. Κάθε συναγερμός αποτυπώνεται με ξεχωριστό αριθμό κλειδί μαζί με τον αριθμό ταυτότητας του πλοίου (Maritime Mobile Service Identity [80]), που επίσης είναι μοναδικός και με τις πληροφορίες για την τοποθεσία και τον χρόνο που συνέβη η ανωμαλία.

Τέλος, στο βήμα 5 του αλγορίθμου 4 δημιουργείται το AR μοντέλο για κάθε ένα πλοίο και εκτιμάται η μελλοντική πορεία του πλοίου, έτσι ώστε να μπορεί να γίνει σύγκριση της εκτίμησης με τα δεδομένα που θα προκύψουν από το ΣΑΑ κατά τα επόμενα δύο λεπτά.

Οι χρονικές παράμετροι και ο αριθμός των παρελθοντικών μετρήσεων επιλέχθηκαν ύστερα από πολυάριθμες δοκιμές και προσομοιώσεις, έτσι ώστε να εξυπηρετούν τις απαιτήσεις της συγκεκριμένης εφαρμογής παρακολούθησης. Η χρησιμοποίηση του προγράμματος σε διαφορετική εφαρμογή θα απαιτούσε διαφορετική παραμετροποίηση.

Κεφάλαιο 10

Προσομοιώσεις, Αποτελέσματα και Συζήτηση

Στο κεφάλαιο αυτό παρουσιάζονται τα πειραματικά αποτελέσματα από την λειτουργία της εφαρμογής που περιγράφηκε στο Κεφάλαιο 9. Το πρόγραμμα φορτώνει τα διαθέσιμα δεδομένα, ελέγχει τις συνθήκες που ορίστηκαν για την αναγνώριση των μη κανονικών συμπεριφορών, καταγράφει τις ανωμαλίες και εκτιμά την μελλοντική πορεία των πλοίων για τα επόμενα δύο λεπτά. Μετά από δυο λεπτά, όταν τα δεδομένα για το χρονικό διάστημα που έχει γίνει η εκτίμηση της τροχιάς γίνουν διαθέσιμα από το ΣΑΑ, το πρόγραμμα φορτώνει τα νέα δεδομένα και επαναλαμβάνει την προηγούμενη διαδικασία.

Τα σχήματα 10.1 και 10.2 παρουσιάζουν τα αποτελέσματα των προσομοιώσεων από δύο πλοία που επιλέχθηκαν τυχαία. Στα σχήματα 10.1α' και 10.2α' σημειώνεται με κύκλο η τροχιά των πλοίων, όπως αυτή προκύπτει από τα δεδομένα του ΣΑΑ και με σταυρό η τροχιά των πλοίων που προκύπτει μετά την επαναδειγματοληψία των δεδομένων. Εύκολα διαπιστώνεται ότι η πορεία που προκύπτει από την επαναδειγματοληψία ταιριάζει απόλυτα στην πραγματική πορεία των πλοίων και δεν υπάρχει κάποια απόκλιση. Το ίδιο συμβαίνει και για όλα τα υπόλοιπα πλοία που ελέγχθηκαν κατά τα πειράματα που πραγματοποιήθηκαν. Επομένως, συμπεραίνεται ότι η επαναδειγματοληψία δεν επηρεάζει την ακεραιότητα των δεδομένων.

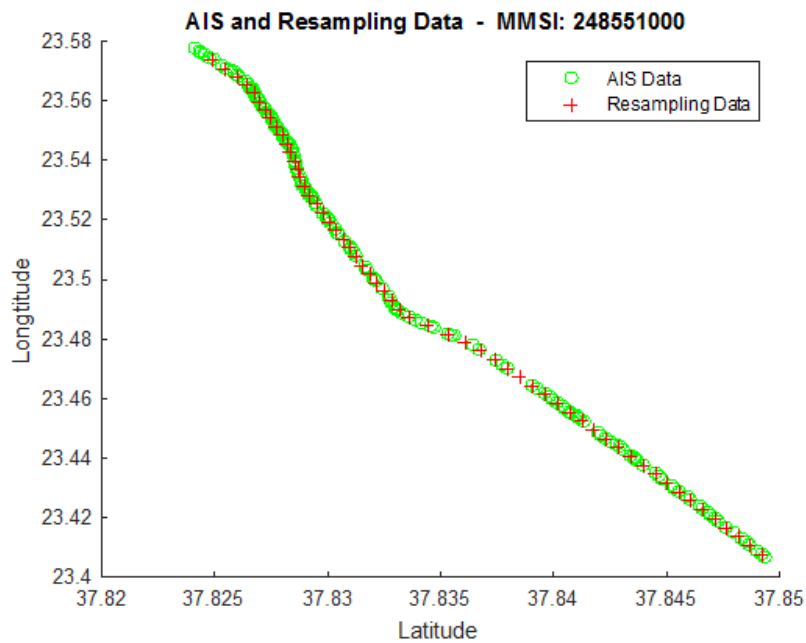
Στα σχήματα 10.1β' και 10.2β' φαίνεται η πορεία των πλοίων που προκύπτει από την επανδειγματοληψία και οι τοποθεσίες όπως αυτές εκτιμώνται από το AR μοντέλο. Η απουσία σταυρών (οι οποίοι δείχνουν την τοποθεσία του πλοίου) μπορεί να οδηγήσει σε αδυναμία δημιουργίας αντιπροσωπευτικού μοντέλου λόγω της έλλειψης δεδομένων για την αρχικοποίηση του AR μοντέλου, γεγονός που σηματοδοτεί έναν συναγερμό προς τον χρήστη. Οι προσομοιώσεις που έγιναν έδειξαν ότι τα μοντέλα που δημιουργούνται επιτυχάνουν να προβλέψουν την μελλοντική τοποθεσία των πλοίων στην πλειοψηφία των

περιπτώσεων. Εξαιρέσεις παρουσιάστηκαν στις περιπτώσεις που υπήρχε έλλειψη δεδομένων από το ΣΑΑ για παρατεταμένα χρονικά διαστήματα, η ταχύτητα των πλοίων παρουσίαζε διακυμάνσεις (μεγάλες τιμές επιτάχυνσης και επιβράδυνσης) και η πορεία του πλοίου έχει μεγάλες αποκλίσεις από τη γραμμική κατεύθυνση. Βέβαια, αυτές οι συμπεριφορές δεν αποτελούν συνηθισμένη συμπεριφορά για ένα πλοίο, και επομένως είναι επιθυμητό να μην μπορεί το μοντέλο που χρησιμοποιείται να τις προβλέψει, έτσι ώστε ο χρήστης να ειδοποιείται για την ύπαρξη πιθανής μη κανονικότητας.

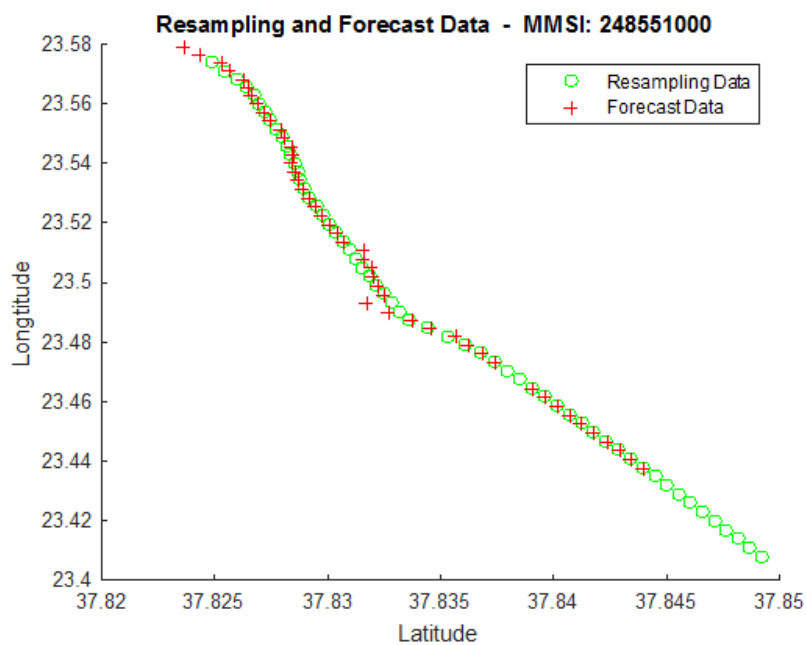
Στο σημείο αυτό είναι σημαντικό να τονιστεί η συνέπεια του προγράμματος, όσον αφορά την αναγνώριση των μη κανονικών συμπεριφορών. Το πρόγραμμα ελέγχθηκε για εκατοντάδες πλοία και υπήρχε σχεδόν απόλυτη επιτυχία στην αναγνώριση των ανωμαλιών. Ήταν επίσης ικανό να προβλέψει την μελλοντική τοποθεσία των πλοίων και να ανιχνεύσει την πλαστογραφία σε δεδομένα. Οι εκτιμώμενες τοποθεσίες παρουσίαζαν πολύ μικρή απόκλιση από την πραγματική θέση των πλοίων (περίπου 50-80 μέτρα), απόσταση μέσα σε επιτρεπτά όρια αν αναλογιστεί κανείς το μέγεθος των πλοίων. Λαιθασμένη πρόβλεψη θέσης ανιχνεύθηκε μόνο σε περιπτώσεις που υπήρχε απώλεια δεδομένων ή όταν τα πλοία εκτελούσαν μεγάλες στροφές καθώς κινούνταν. Στις περιπτώσεις αυτές βέβαια, όπως προαναφέρθηκε, είναι επιθυμητό το μοντέλο να μην μπορεί να προβλέψει αυτή τη συμπεριφορά.

Συνοψίζοντας, η συνεισφορά της παρούσας εργασίας είναι ότι δημιουργήθηκε ένα εργαλείο για την παρακολούθηση της συμπεριφοράς των πλοίων, τόσο για την παροχή προβλέψεων σχετικά με την τοποθεσία και την κίνηση του πλοίου, όσο και για τον ορισμό κάποιων βασικών μη κανονικών συμπεριφορών. Το εργαλείο αυτό πέρασε από πολυάριθμους ελέγχους και αποδείχτηκε ότι μπορεί να χρησιμοποιηθεί επιτυχώς για την παρακολούθηση της θαλάσσιας κίνησης και την ανίχνευση μη κανονικών συμπεριφορών.

Σχήμα 10.1 Πραγματική, επαναδειγματοληπτημένη και εκτιμώμενη πορεία για το πλοίο με MMSI 248551000.

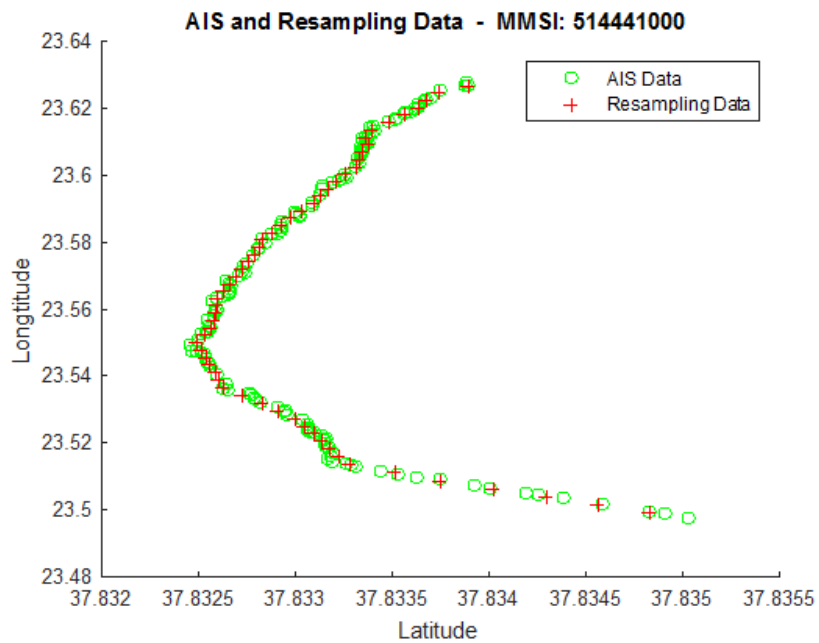


(α') Πραγματική και επαναδειγματοληπτημένη πορεία πλοίου.

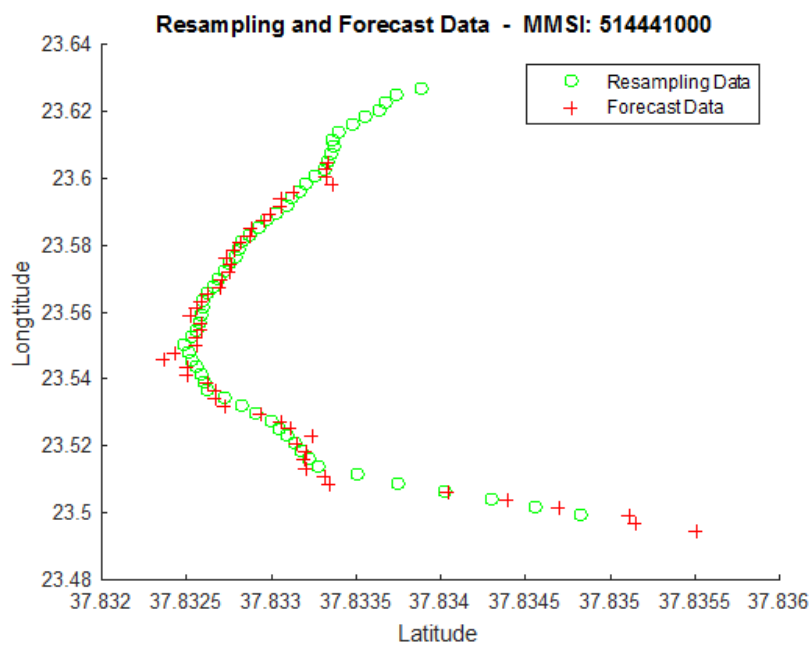


(β') Επαναδειγματοληπτημένη και εκτιμώμενη πορεία πλοίου.

Σχήμα 10.2 Πραγματική, επαναδειγματοληπτημένη και εκτιμώμενη πορεία για το πλοίο με MMSI 514441000.



(α') Πραγματική και επαναδειγματοληπτημένη πορεία πλοίου.



(β') Επαναδειγματοληπτημένη και εκτιμώμενη πορεία πλοίου.

Κεφάλαιο 11

Συμπεράσματα

Οι προσομοιώσεις που έγιναν σε πραγματικά δεδομένα έδειξαν ότι το εργαλείο που δημιουργήθηκε μπορεί να παρακολουθήσει επιτυχώς την θαλάσσια κίνηση, να προβλέψει την μελλοντική θέση των πλοίων και να ανιχνεύσει τις μη κανονικές συμπεριφορές που ορίστηκαν. Το γραμμικό AR μοντέλο αποδείχτηκε ότι μπορεί να χρησιμοποιηθεί για την επίλυση του προβλήματος της πρόβλεψης της τοποθεσίας του πλοίου και της εκτίμησης της τροχιάς του. Επιπλέον, ήταν δυνατή η αναγνώριση πλαστογραφίας στα δεδομένα του ΣΑΑ χωρίς την χρήση επιπρόσθετων συστημάτων ελέγχου (π.χ. ραντάρ), εκτός από τις περιπτώσεις που η πλαστογραφία γινόταν από άτομα που γνώριζαν επαρκώς πως να πλαστογραφήσουν τα δεδομένα. Βέβαια, για να εξακριβωθεί με απόλυτη σιγουριά αν τα δεδομένα έχουν υποστεί επεξεργασία είναι αναγκαία η χρήση επιπρόσθετων μέσων παρακολούθησης πέραν του ΣΑΑ. Συνοψίζοντας, μπορούμε να συμπεράνουμε με ασφάλεια ότι οι μέθοδοι και το εργαλείο που αναπτύχθηκαν έχουν την ικανότητα να συμβάλουν σημαντικά στην παρακολούθηση της θαλάσσιας κίνησης και στην πρόβλεψη της κίνησης των πλοίων, δίνοντας αξιόπιστα αποτελέσματα σε πραγματικό χρόνο.

Βιβλιογραφία

- [1] B. Shen, Z. Wang, and H. Shu. *Nonlinear Stochastic Systems with Incomplete Information*. Springer, London, UK, 1st edition, 2013.
- [2] J. P. Higgins. Nonlinear systems in medicine. *Yale Journal of Biology and Medicine*, 75(5--6):247--260, Sep.--Dec 2002.
- [3] V. Z. Marmarelis. *Nonlinear Dynamic Modeling of Physiological Systems*. Wiley-IEEE Press, Hoboken, New Jersey, 1st edition, 1994.
- [4] D. T. Westwick and R. E. Kearney. *Identification of Nonlinear Systems*. Wiley-IEEE Press, Piscataway, New Jersey, 1st edition, 2003.
- [5] L. H. Carney and M. Friedman. Nonlinear feedback models for the tuning of auditory nerve fibers. *Annals of Biomedical Engineering*, 24(3):440---450, May--Jun. 1996.
- [6] J. J. Eggermont. Wiener and volterra analyses applied to the auditory system. *Hearing Research*, 66(2):177---201, Apr. 1993.
- [7] K. I. Naka, H. M. Sakai, and I. Naohiro. Generation and transformation of second-order nonlinearity in catfish retina. *Annals of Biomedical Engineering*, 16(1):53--64, Jan. 1998.
- [8] M. C. Citron, R. C. Emerson, and W. R. Levick. Nonlinear measurement and classification of receptive fields in cat retinal ganglion cells. *Annals of Biomedical Engineering*, 16(1):65--77, Jan. 1998.
- [9] R. E. Kearney and I. W. Hunter. Nonlinear identification of stretch reflex dynamics. *Annals of Biomedical Engineering*, 16(1):79--94, Jan. 1998.
- [10] R. E. Kearney, R. B. Stein, and L. Parameswaran. Identification of intrinsic and reflex contributions to human ankle stiffness dynamics. *IEEE Transactions on Biomedical Engineering*, 44(6):493--504, Jun. 1997.
- [11] H. L. Galiana, H. L. Smith, and A. Katsarkas. Comparison of linear vs. nonlinear methods for analyzing the vestibulo-ocular reflex (vor). *Acta Oto-Laryngologica*, 115(5):585--596, Feb. 1995.
- [12] H. L. Galiana, H. L. Smith, and A. Katsarkas. Modelling non-linearities in the vestibulo-ocular reflex (vor) after unilateral or bilateral loss of peripheral vestibular function. *Experimental Brain Research*, 137(3--4):369--386, Apr. 2001.

- [13] K. H. Chon, T. J. Mullen, and R. J. Cohen. A dual-input nonlinear system analysis of autonomic modulation of heart rate. *IEEE Transactions on Biomedical Engineering*, 43(5):530--544, May 1996.
- [14] K. H. Chon, Y. M. Chen, N. H. Holstein-Rathlou, and V. Z. Marmarelis. Nonlinear system analysis of renal autoregulation in normotensive and hypertensive rats. *IEEE Transactions on Biomedical Engineering*, 45(3):342--353, Mar. 1998.
- [15] V. Z. Marmarelis, K. H. Chon, Y. M. Chen, D. J. Marsh, and N. H. Holstein-Rathlou. Nonlinear analysis of renal autoregulation under broadband forcing conditions. *Annals of Biomedical Engineering*, 21(6):591--603, Apr. 1993.
- [16] V. Z. Marmarelis, K. H. Chon, N. H. Holstein-Rathlou, and D. J. Marsh. Nonlinear analysis of renal autoregulation in rats using principal dynamic modes. *Annals of Biomedical Engineering*, 27(1):23--31, Jan. 1999.
- [17] G. N. Maksym, R. E. Kearney, and J. H. T. Bates. Nonparametric block-structured modeling of lung tissue strip mechanics. *Annals of Biomedical Engineering*, 26(2):242--252, Mar. 1998.
- [18] H. Yuan, D. T. Westwick, E. P. Ingenito, K. R. Lutchen, and B. Suki. Parametric and nonparametric nonlinear system identification of lung tissue strip mechanics. *Annals of Biomedical Engineering*, 27(4):548--562, Jul. 1999.
- [19] P. E. Crago. Muscle input-output model: The static dependence of force on length, recruitment, and firing period. *IEEE Transactions of Biomedical Engineering*, 39(8):871--874, Aug. 1992.
- [20] K. J. Hunt, M. Munih, N. N. Donaldson, and F. M. D. Barr. Investigation of the hammerstein hypothesis in the modeling of electrically stimulated muscle. *IEEE Transactions on Biomedical Engineering*, 45(8):998--1009, Aug. 1998.
- [21] M. Munih, K. Hunt, and N. Donaldson. Variation of recruitment nonlinearity and dynamic response of ankle plantarflexors. *Medical Engineering and Physics*, 22(2):97--107, Mar. 2000.
- [22] M. E. Brandt and G. Chen. Feedback control of a biodynamical model of hiv-1. *IEEE Transactions on Biomedical Engineering*, 48(7):754--759, Jul. 2001.
- [23] I. P. Kordonis, A. C. Charalampidis, and G. P. Papavassilopoulos. A nonlinear control approach for an hiv-1 model. *Proceedings of the 4th Panhellenic Conference of Electrical and Computer Engineering Students*, 25(3):353--361, Nov. 2010.
- [24] S. Gee, Z. Tian, and T. Lee. Nonlinear control of a dynamic model of hiv-1. *IEEE Transactions on Biomedical Engineering*, 25(3):353--361, Mar. 2005.
- [25] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. L. Ebos, L. Hlatky, and P. Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput. Biol.*, 10(8), Aug. 2014.

- [26] N. Hartung, S. Mollard, D. Barbolosi, A. Benabdallah, G. Chapuisat, G. Henry, S. Giacometti, A. Iliadis, J. Ciccolini, C. Faivre, and F. Hubert. Mathematical modeling of tumor growth and metastatic spreading: Validation in tumor-bearing mice. *Cancer Research*, 74(22):6397--6407, Nov. 2014.
- [27] E. A. Sarapata and L. G. de Pillis. A comparison and catalog of intrinsic tumor growth models. *Bulletin of Mathematical Biology*, 76(8):2010--2024, Aug. 2014.
- [28] A. C. Charalampidis, G. Pontikis, G. D. Mitsis, G. Dimitriadis, V. Lampadiari, V. Z. Marmarelis, A. Armaganidis, and G. P. Papavassilopoulos. Calibration of a microdialysis sensor and recursive glucose level estimation in icu patients using kalman and particle filtering. *Biomedical Signal Processing and Control*, 27:155--163, May 2016.
- [29] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513--583, 1825.
- [30] B. Dennis and J. M. Ponciano. Density-dependent state-space model for population-abundance data with unequal time intervals. *Ecology*, 95(8):2069--2076, Aug. 2014.
- [31] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, and G. P. Papavassilopoulos. Comparing methods for parameter estimation of the gompertz tumor growth model. *IFAC-PapersOnLine*, 50(1):12203--12209, Jul. 2017.
- [32] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, and G. P. Papavassilopoulos. Tumor growth modeling: Parameter estimation with maximum likelihood methods. *Computer Methods and Programs in Biomedicine*, 160:1--10, Jul. 2018.
- [33] E. Barillot, L. Calzone, P. Hupe, J. P. Vert, and A. Zinovyev. *Computational Systems Biology of Cancer*. CRC Press, Boca Raton, Florida, 2nd edition, 2012.
- [34] B. W. Stewart and C. P. Wild. *World Cancer Report 2014*. International Agency for Research on Cancer, Lyon, France, 2014.
- [35] R. P. Araujo and D. L. S. McElwain. A history of the study of solid tumour growth: The contribution of mathematical modelling. *Bulletin of Mathematical Biology*, 66(5):1039--1091, Sep. 2004.
- [36] C. P. Winsor. The gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences of the United States of America*, 18(1):1--8, Jan. 1932.
- [37] A. K. Laird. Dynamics of tumour growth. *British Journal of Cancer*, 18(3):490--502, Sep. 1965.
- [38] M. M. Hadjiandreou and G. D. Mitsis. Mathematical modeling of tumor growth, drug-resistance, toxicity, and optimal therapy design. *IEEE Transactions on Biomedical Engineering*, 61(2):415--425, Feb 2014.
- [39] P. Dua, V. Dua, and E. N. Pistikopoulos. Optimal delivery of chemotherapeutic agents in cancer. *Computers & Chemical Engineering*, 32(1):99--107, Jan. 2008.

- [40] B. Dominique and A. Iliadis. Optimizing drug regimens in cancer chemotherapy: a simulation study using a pk–pd model. *Computers in Biology and Medicine*, 31(3):157–172, May 2001.
- [41] A. Achilleos, C. Loizides, M. Hadjiandreou, T. Stylianopoulos, and G. D. Mitsis. Multiprocess dynamic modeling of tumor evolution with bayesian tumor-specific predictions. *Annals of Biomedical Engineering*, 42(5):1095–1111, May 2014.
- [42] C. Loizides, D. Iacovides, M. M. Hadjiandreou, G. Rizki, A. Achilleos, K. Strati, and G. D. Mitsis. Model-based tumor growth dynamics and therapy response in a mouse model of de novo carcinogenesis. *PLOS ONE*, 10(12):1–18, Dec. 2015.
- [43] R. Chignola, A. Schenetti, E. Chiesa, R. Foroni, S. Sartpris, A. Brendolan, G. Tridente, G. Andrighetto, and D. Liberati. Oscillating growth patterns of multicellular tumour spheroids. *Cell Proliferation*, 32(1):39–48, 1999.
- [44] R. Chignola and R. I. Foroni. Estimating the growth kinetics of experimental tumors from as few as two determinations of tumor size: implications for clinical oncology. *IEEE Transactions on Biomedical Engineering*, 52(5):808–815, May 2005.
- [45] A. Talkington and R. Durrett. Estimating tumor growth rates in vivo. *Bulletin of Mathematical Biology*, 77(10):1934–1954, Oct. 2015.
- [46] G. F. Brunton and T. E. Wheldon. Prediction of the complete growth pattern of human multiple myeloma from restricted initial measurements. *Cell Proliferation*, 10(6):591–594, Nov. 1977.
- [47] G. F. Brunton and T. E. Wheldon. Characteristic species dependent growth patterns of mammalian neoplasms. *Cell Proliferation*, 11(2):161–175, Jul. 1980.
- [48] G. F. Brunton and T. E. Wheldon. The gompertz equation and the construction of tumour growth curves. *Cell Proliferation*, 13(4):455–460, Jul. 1980.
- [49] T.E. Wheldon. *Mathematical models in cancer research*. Medical science series. A. Hilger, 1988.
- [50] L. Norton, R. Simon, H. D. Brereton, and A. E. Bogden. Predicting the course of gompertzian growth. *Nature*, 264:542–545, 1976.
- [51] P. Gerlee. The model muddle: In search of tumor growth laws. *Cancer Research*, 73(8):2407–2411, 2013.
- [52] R. C. Rockne and P. Frankel. *Advances in Radiation Oncology*. Springer, Cham, Switzerland, 2017.
- [53] C.F. Babbs. Predicting success or failure of immunotherapy for cancer: insights from a clinically applicable mathematical model. *American Journal of Cancer Research*, 2(2):204–213, Feb. 2012.
- [54] R. Serre, S. Benzekry, L. Padovani, C. Meille, N. André, J. Ciccolini, F. Barlesi, X. Muracciole, and D. Barbolosi. Mathematical modeling of cancer immunotherapy and its synergy with radiotherapy. *Cancer Research*, 76(17):4931–4940, Sep. 2016.

- [55] S. Benzekry, A. Tracz, M. Mastri, R. Corbelli, S. Barbolosi, and J. M. L. Ebos. Modeling spontaneous metastasis following surgery: An in vivo-in silico approach. *Cancer Research*, 76(3):535--547, Feb. 2016.
- [56] F. Michor and K. Beal. Improving cancer treatment via mathematical modeling: Surmounting the challenges is worth the effort. *Cell*, 163(5):1059--1063, Nov. 2015.
- [57] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174--188, Feb. 2002.
- [58] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2):107--113, Apr. 1993.
- [59] R. E. Kalman. A new approach to linear filtering and prediction problem. *ASME Journal of Basic Engineering*, 82(1):35--45, Mar. 1960.
- [60] P. S Maybeck. *Compensation of linear model inadequacies, in Stochastic Models, Estimation, and Control Volume 1*. Academic Press, London, UK, 1st edition, 1982.
- [61] S. M Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall PTR, Upper Saddle River, New Jersey, 1st edition, 1993.
- [62] A. C. Charalampidis and G. P. Papavassilopoulos. Computationally efficient kalman filtering for a class of nonlinear systems. *IEEE Transaction on Automatic Control*, 53(3):483--491, Sep. 2010.
- [63] A. C. Charalampidis. Introduction to kalman filtering. *Notes*, 2015.
- [64] G. L. Smith, S. F. Schmidt, and L. A. McGee. Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle. *Technical Report TR R-135, NASA*, 1962.
- [65] P. S Maybeck. *Compensation of linear model inadequacies, in Stochastic Models, Estimation, and Control Volume 2*. Academic Press, London, UK, 1st edition, 1982.
- [66] A. C. Charalampidis and G. P. Papavassilopoulos. Development and numerical investigation of new non-linear kalman filter variants. *IET Control Theory & Application*, 5(10):1155--1166, Jul. 2011.
- [67] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197--208, Jul. 2000.
- [68] J. Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317--1339, 1989.
- [69] L. Ljung. *Parameter Estimation Methods: System Identification Theory for the User*. Prentice Hall PTR, Upper Saddle River, New Jersey, 2nd edition, 1999.
- [70] T. B. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39--49, Jan. 2011.

- [71] H. W. Sorenson. *Parameter Estimation: Principles and problems*. Marcel Dekker, New York, 1980.
- [72] A. Dempster, N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1--38, Jan. 1977.
- [73] *Matlab Parallel Computing Toolbox User's Guide*. The MathWorks Inc., Natick, MA, 2017.
- [74] M. Harris. Optimizing parallel reduction in cuda. *NVIDIA Corporation*, 2007.
- [75] *CURAND LIBRARY Programming Guide*. NVIDIA Corporation, 2017.
- [76] *Matlab Optimization Toolbox User's Guide*. The MathWorks Inc., Natick, MA, 2017.
- [77] R. Chignola, A. Schenetti, E. Chiesa, R. Feroni, S. Sartpris, A. Brendolan, G. Tridente, G. Andrighetto, and D. Liberati. Oscillating growth patterns of multicellular tumour spheroids. *Cell Proliferation*, 32(1):39--48, 1999.
- [78] R. Chignola, A. Schenetti, G. Andrighetto, E. Chiesa, R. Feroni, S. Sartoris, G. Tridente, and D. Liberati. Forecasting the growth of multicell tumour spheroids: implications for the dynamic growth of solid tumours. *Cell Proliferation*, 33(4):219--229, 2000.
- [79] E. Milotti, V. Vyshemirsky, M. Sega, and R. Chignola. Interplay between distribution of live cells and growth dynamics of solid tumours. *Scientific Reports*, 2(990), 2012.
- [80] B. J. Tetreault. Use of the automatic identification system (ais) for maritime domain awareness (mda). In *Proceedings of OCEANS 2005 MTS/IEEE*, volume 2, pages 1590--1594, Sept 2005.
- [81] S. Mascaro, A. E. Nicholso, and K. B. Korb. Anomaly detection in vessel tracks using bayesian networks. *International Journal of Approximate Reasoning*, 55(1):84--98, 2014. Applications of Bayesian Networks.
- [82] F. Johansson and G. Falkman. Detection of vessel anomalies - a bayesian network approach. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 395--400, Dec. 2007.
- [83] K. Kowalska and L. Peel. Maritime anomaly detection using gaussian process active learning. In *2012 15th International Conference on Information Fusion*, pages 1164--1171, July 2012.
- [84] J. Will, L. Peel, and C. CLaxton. Fast maritime anomaly detection using kd-tree gaussian processes. In *Proceedings IMA Maths in Defence*, pages 1164--1171, 2011.
- [85] M. P. J. Butler. Project polar epsilon: joint space-based wide area surveillance and support capability. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, volume 2, pages 1194--1197, Jul. 2005.

- [86] F. Katsilieris, P. Braca, and S. Coraluppi. Detection of malicious ais position spoofing by exploiting radar information. In *Proceedings of the 16th International Conference on Information Fusion*, pages 1196--1203, Jul. 2013.
- [87] A. V Oppenheim and R. W. Schafer. *Discrete-time Signal Processing*. Prentice Hall, 3rd edition, 2009.
- [88] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.
- [89] L. Ljung. *Matlab & Simulink System Identification Toolbox User's Guide*. The MathWorks Inc., Natick, MA, 2014.
- [90] G. C Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Dover Publications, New York, 2009.
- [91] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis, Forecasting and Control*. Prentice Hall, Upper Saddle River, New Jersey, 3rd edition, 1994.
- [92] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159--175, 2003.
- [93] M. Rout, B. Majhi, R. Majhi, and G. Panda. Forecasting of currency exchange rates using an adaptive arma model with differential evolution based training. *Journal of King Saud University - Computer and Information Sciences*, 26(1):7--18, Jan 2014.
- [94] J. L. Torres, A. García., M. De Blas, and A. De Francisco. Forecast of hourly average wind speed with arma models in navarre (spain). *Solar Energy*, 79(1):65--77, Jul 2015.
- [95] P. Burlando, R. Rosso, L. G. Cadavid, and J. D. Salas. Forecasting of short-term rainfall using arma models. *Journal of Hydrology*, 144(1):193--211, Apr 1993.
- [96] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE Transactions on Power Systems*, 20(2):1035--1042, May 2005.
- [97] L. R Rabiner and R W. Schafer. *Theory and Applications of Digital Speech Processing*. Pearson, 1st edition, 2010.
- [98] T. Vincenty. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, 23(176):88--93, May 2005.

Βιογραφικό Σημείωμα

Προσωπικά Στοιχεία

Όνομα: **Σπυρίδων**
Επίθετο: **Πατμανίδης**
Πατρώνυμο: **Βασίλειος**
Τόπος | Ημ. γέννησης: **Βέροια | 23 Ιανουαρίου 1989**
Πληροφορίες επικοινωνίας
Διεύθυνση: **Προύσης 8, Βέροια, 59132**
Τηλέφωνο: **6972818983**
email: **spatmanid@gmail.com**

Εκπαίδευση

- 2014-2018 **Υποψήφιος Διδάκτορας στη Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών ΕΜΠ**
Πεδίο: **Αναδρομική Εκτίμηση Παραμέτρων σε Γραμμικά και Μη Γραμμικά Μοντέλα"**
Επιβλέπων: **Καθ. Γεώργιος Π. Παπαβασιλόπουλος**
- 2007-2013 **Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών ΑΠΘ**
Διπλωματική εργασία: **"Ανάπτυξη τεχνικών συνεργασίας για εξερεύνηση αγνώστων περιοχών σε Παιχνίδια Στρατηγικής Πραγματικού Χρόνου με χρήση Νευροεξέλιξης"**
- 2006 **Απόφοιτος 5ου Ενιαίου Λυκείου Σερρών**

Κατάλογος Δημοσιεύσεων

Άρθρα σε Επιστημονικά Περιοδικά

1. Patmanidis S., Charalampidis A. C., Kordonis I., Mitsis G. D. and Papavassilopoulos G. P. (2018). **Tumor Growth Modeling: Parameter estimation with Maximum Likelihood methods.** *Computer Methods and Programs in Biomedicine, Volume 160, Pages 1–10, July, DOI:10.1016/j.cmpb.2018.03.014*
2. Patmanidis S., Charalampidis A. C., Kordonis I., Mitsis G. D. and Papavassilopoulos G. P. (2018). **Individualized Tumor Growth Modeling and Prediction on Mice Skin Tumors with Maximum Likelihood Estimators.** *Under Preparation/Review.*
3. Patmanidis S., Chignola R., Charalampidis A. C. and Papavassilopoulos G. P. (2018). **A Comparison between Linear Regression and Maximum Likelihood Estimation for the Prediction of Tumor Growth on Experimental Data of Human and Rat Origin.** *Under Preparation/Review.*

Άρθρα σε Συνέδρια

1. Patmanidis S., Voulgaris I., Sarri E., Papavasileiou G., Papavassilopoulos G. P. (2016). **Maritime surveillance, vessel route estimation and alerts using AIS data.** *24th Mediterranean Conference on Control and Automation, IEEE, Athens, Greece, June.*
2. Patmanidis S., Charalampidis A. C., Kordonis I., Mitsis G. D. and Papavassilopoulos G. P. (2017). **Comparing Methods for Parameter Estimation of the Gompertz Tumor Growth Model.** *IFAC-PapersOnLine, Volume 50, Issue 1, Pages 12203--12209, Toulouse, France, July, DOI:10.1016/j.ifacol.2017.08.2289.*
3. Patmanidis S., Charalampidis A. C., Kordonis I. and Papavassilopoulos G. P. (2018). **Tumor Growth Modeling: State Estimation with Maximum Likelihood and Particle Filter Implementations.** *Under Preparation/Review.*