



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

**ΕΚΤΕΤΑΜΕΝΟ ΦΙΛΤΡΟ ΚΑΛΜΑΝ ΓΙΑ ΤΗΝ ΒΡΑΧΥΠΡΟΘΕΣΜΗ
ΠΡΟΒΛΕΨΗ ΤΟΥ COVID-19**

του

ΑΘΑΝΑΣΙΟΥ ΤΖΟΒΑΡΑ

Διπλωματική Εργασία

Υπεβλήθη για την εκπλήρωση μέρους των απαιτήσεων για
την απόκτηση του Διπλώματος Μηχανολόγου Μηχανικού

Βόλος, 2022



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

**ΕΚΤΕΤΑΜΕΝΟ ΦΙΛΤΡΟ ΚΑΛΜΑΝ ΓΙΑ ΤΗΝ ΒΡΑΧΥΠΡΟΘΕΣΜΗ
ΠΡΟΒΛΕΨΗ ΤΟΥ COVID-19**

του

ΑΘΑΝΑΣΙΟΥ ΤΖΟΒΑΡΑ

Διπλωματική Εργασία

Υπεβλήθη για την εκπλήρωση μέρους των απαιτήσεων για
την απόκτηση του Διπλώματος Μηχανολόγου Μηχανικού

Βόλος, 2022

(Υπογραφή)

.....

Αθανάσιος Τζοβάρας
Υποψήφιος Μηχανολόγος Μηχανικός

© 2022 -- All Rights reserved

Copyright © 2022 --All rights reserved Αθανάσιος Τζοβάρας .

Με επιφύλαξη παντός δικαιώματος.

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. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανολόγων Μηχανικών της Πολυτεχνικής Σχολής του Πανεπιστημίου Θεσσαλίας δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα (Ν. 5343/32 αρ. 202 παρ. 2).

Εγκρίθηκε από τα Μέλη της Τριμελούς Εξεταστικής Επιτροπής:

Πρώτος Εξεταστής: Αμπουντώλας Κωνσταντίνος (Επιβλέπων)

Αναπληρωτής Καθηγητής, Τμήμα Μηχανολόγων Μηχανικών, Πανεπιστήμιο
Θεσσαλίας

Δεύτερος Εξεταστής: Παπαδημητρίου Κωνσταντίνος

Καθηγητής, Τμήμα Μηχανολόγων Μηχανικών, Πανεπιστήμιο Θεσσαλίας

Τρίτος Εξεταστής: Παντελής Δημήτριος

Καθηγητής, Τμήμα Μηχανολόγων Μηχανικών, Πανεπιστήμιο Θεσσαλίας

ΕΚΤΕΤΑΜΕΝΟ ΦΙΛΤΡΟ ΚΑΛΜΑΝ ΓΙΑ ΤΗΝ ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ ΤΟΥ COVID-19 ΤΖΟΒΑΡΑΣ ΑΘΑΝΑΣΙΟΣ

Τμήμα Μηχανολόγων Μηχανικών, Πανεπιστήμιο Θεσσαλίας, 2022

Επιβλέπων Καθηγητής: Δρ. Αμπουντώλας Κωνσταντίνος
Αναπληρωτής Καθηγητής Συστημάτων Αυτομάτου Ελέγχου στη Μηχανολογία

Περίληψη

Στα πλαίσια της διπλωματικής εργασίας, θα ασχοληθούμε με προβλέψεις και την εξέλιξη του νέου κορονοϊού (COVID-19). Για το σκοπό αυτό, θα χρησιμοποιήσουμε το μαθηματικό-επιδημιολογικό μοντέλο SIR (Susceptible-Infectious-Recovered) και διάφορες επεκτάσεις του. Συγκεκριμένα, στην έρευνα αυτή, θα χρησιμοποιηθεί μια επέκταση του απλού μοντέλου, το μοντέλο SIRD για διακριτό χρόνο (Susceptible-Infectious-Recovered-Died), το οποίο περιγράφει ικανοποιητικά μια επιδημία. Για την εκτίμηση/πρόβλεψη των χρονικά μεταβαλλόμενων παραμέτρων του μοντέλου SIRD, θα χρησιμοποιηθεί το Εκτεταμένο Φίλτρο Kalman (EKF) με συνθετικά δεδομένα για τον COVID-19 ενώ στη συνέχεια θα εφαρμοστεί το μοντέλο με ρεαλιστικά δεδομένα την επαρχία της Κίνας Χουμπέι. Στη συνέχεια θα γίνει σύγκριση με αντίστοιχη έρευνα που έχει γίνει στο παρελθόν με διαφορετική μέθοδο εκτιμήσεων, καθώς και με τα πραγματικά δεδομένα που παρατηρήθηκαν στο τέλος της περιόδου που θα μελετηθεί ενώ στο τέλος θα προταθούν άλλες μέθοδοι και ζητήματα που θα μπορούσαν να ερευνηθούν στο μέλλον.

Λέξεις-κλειδιά: πανδημία, COVID-19, Φίλτρο Κάλμαν, Εκτεταμένο Φίλτρο Κάλμαν, εκτιμήσεις, μακροπρόθεσμες προβλέψεις, επιδημιολογικά μοντέλα, SIR, SIRD, ρυθμός αναπαραγωγής.

EXTENDED KALMAN FILTER FOR THE SHORT-TERM PREDICTION OF COVID-19

TZOVARAS ATHANASIOS

Department of Mechanical Engineering, University of Thessaly, 2022

Supervisor: Dr. Ampountolas Konstantinos

Associate Professor of Automatic Control System in Engineering

Abstract

This project concerns long-term predictions of the coronavirus disease (COVID-19). For this purpose, the mathematical epidemiological model SIR (Susceptible-Infectious-Recovered) and its various extensions are used. Specifically, in this research, an extension of the simple model will be used, the SIRD model for discrete time (Susceptible-Infectious-Recovered-Died), which satisfactorily describes an epidemic. To estimate the time-varying parameters of the SIRD model, the Extended Kalman Filter (EKF) will be used with synthetic data for COVID-19, then the same model will be applied with realistic data in the Chinese province of Hubei. Then a comparison will be made with corresponding research that has been done in the past with a different estimation method, as well as with the actual data observed at the end of the period to be studied. Finally other methods and issues that could be researched in the future will be proposed.

Keywords: epidemic, COVID-19, Kalman Filter, Extended Kalman Filter, estimation, long-term prediction, epidemiological models, SIR, SIRD, reproduction number.

Πίνακας περιεχομένων

| | |
|--|----|
| ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ | 9 |
| 1.1 Βιβλιογραφική ανασκόπηση | 9 |
| 1.3 Βασικές έννοιες | 14 |
| ΚΕΦΑΛΑΙΟ 2: Επιδημιολογικά μοντέλα | 18 |
| 2.1 Τα επιδημιολογικά μοντέλα | 18 |
| 2.1.1 Συνεχές μοντέλο SIR | 20 |
| 2.1.2 Διακριτό Μοντέλο SIR | 22 |
| 2.2 Παραλλαγές του Επιδημιολογικού Μοντέλου SIR | 23 |
| ΚΕΦΑΛΑΙΟ 3: Φίλτρα Kalman | 26 |
| 3.1 Φίλτρο Kalman Διακριτού Χρόνου για γραμμικά συστήματα | 29 |
| 3.2 Το εκτεταμένο Φίλτρο Kalman για μη γραμμικά συστήματα | 31 |
| ΚΕΦΑΛΑΙΟ 4: Εφαρμογή του Εκτεταμένου φίλτρου Kalman στο επιδημιολογικό μοντέλο SIRD για τον COVID19 | 33 |
| 4.1 Μαθηματική Μοντελοποίηση | 33 |
| 4.2 Το Εκτεταμένο Φίλτρο Kalman για εκτίμηση παραμέτρων | 36 |
| 4.3 Εκτιμήσεις παραμέτρων με συνθετικά δεδομένα | 39 |
| 4.3.1 Αποτελέσματα | 40 |
| 4.4 Εκτιμήσεις παραμέτρων με πραγματικά δεδομένα | 47 |
| 4.4.1 Αποτελέσματα | 47 |
| 4.5 Συμπεράσματα-Σύγκριση Αποτελεσμάτων | 54 |
| ΚΕΦΑΛΑΙΟ 5: Σύνοψη-Συζήτηση | 56 |
| Βιβλιογραφία | 57 |
| Παράρτημα | 59 |

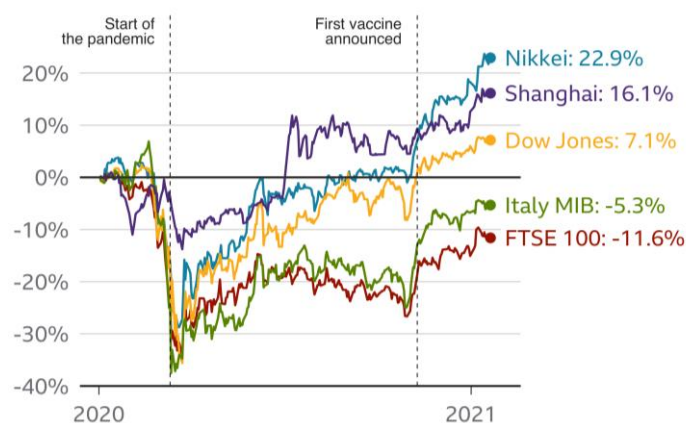
ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ

1.1 Βιβλιογραφική ανασκόπηση

Ο νεοεμφανιζόμενος ιός κορονοϊός (SARS-CoV-2), γνωστός και ως COVID-19, αναγνωρίστηκε ως πανδημία από τον Παγκόσμιο Οργανισμό Υγείας (WHO) στις 11 Μαρτίου 2020. Υπάρχουν πολλά αναπάντητα ερωτήματα σχετικά με αυτόν τον ιό μέχρι σήμερα και καμία οριστική θεραπεία λόγω της έλλειψης κατανόησης της παθογένεσης και του τρόπου πολλαπλασιασμού του, που είναι άγνωστες και δεν μπορούν να εντοπιστούν. Ο πρωταρχικός στόχος είναι να σταματήσει η εξάπλωσή του παγκοσμίως. Σε αυτή την εργασία στοχεύουμε να παρέχουμε προβλέψεις για τη διάδοσή του χρησιμοποιώντας ένα στοχαστικό μοντέλο Lotka–Volterra σε συνδυασμό με έναν εκτεταμένο αλγόριθμο φίλτρου Kalman (EKF) για τη μοντελοποίηση της δυναμικής του COVID-19. Τα αποτελέσματά μας δείχνουν τη σκοπιμότητα της χρήσης αυτού του μοντέλου για την πρόβλεψη της εξάπλωσης του ιού και την ικανότητα διαφορετικών μέτρων ελέγχου (π.χ. κοινωνική απόσταση, περιορισμός μετακινήσεων) για τη μείωση του αριθμού των προσβεβλημένων ατόμων.

Ο COVID-19 είναι ένας νεοεμφανιζόμενος ανθρώπινος μολυσματικός ιός που προέρχεται από τη Γουχάν της Κίνας και εξαπλώνεται ταχύτατα στην Κίνα και σε όλο τον κόσμο από τον Δεκέμβριο του 2019. Κηρύχθηκε παγκόσμια έκτακτη ανάγκη στις 31 Ιανουαρίου 2020 από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ) και στις 11 Μαρτίου η ασθένεια αναγνωρίστηκε ως παγκόσμια πανδημία. Στις 21 Ιανουαρίου 2020, ο ΠΟΥ πρότεινε ότι υπήρχε πιθανή παρατεταμένη μετάδοση από άνθρωπο σε άνθρωπο.

Αυτή η επιδημία έχει συγκλονίσει τον κόσμο και έχει δημιουργήσει μια σημαντική ποσότητα αβεβαιότητας στην καθημερινή ζωή. Με αυτό το ξέσπασμα να συνεχίζεται, έχει τρομερές συνέπειες στην παγκόσμια δημόσια υγεία και οικονομία όπως φαίνεται ενδεικτικά στην Εικόνα 1.1 και τα Σχήματα 1.1-1.4.

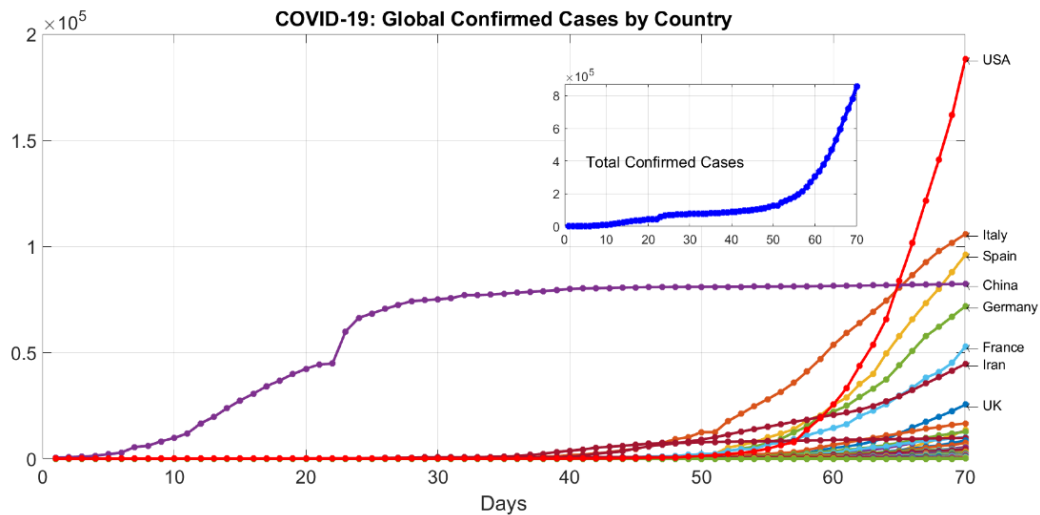


Source: Bloomberg, 24 January 2021, 00:01 GMT

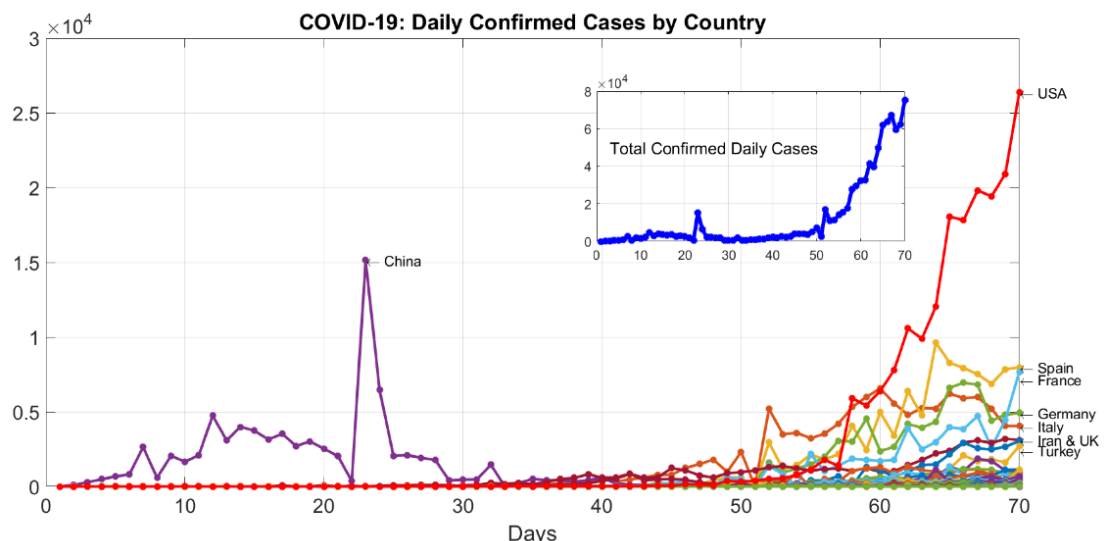
BBC

Εικόνα 1.1. Οι συνέπειες μετά από το ξέσπασμα του COVID-19 στην παγκόσμια οικονομία.

Ο αριθμός των επιβεβαιωμένων κρουσμάτων COVID-19 παγκοσμίως έως τις 31 Μαρτίου 2020 φαίνεται στο Σχήμα 1.1 που συλλέγεται από τον ιστότοπο του ΠΟΥ. Είναι σαφές ότι τα κρούσματα προήλθαν από την Κίνα, όπου έφθασαν στο αποκορύφωμά τους στα τέλη Φεβρουαρίου 2020 και σημείωσαν μια σταθερότητα στη συνέχεια, υποδηλώνοντας έναν πιθανό έλεγχο της εξάπλωσης. Ενώ οι ΗΠΑ έδειξαν ξεκάθαρα μια δραματική αύξηση στον αριθμό των κρουσμάτων φτάνοντας σε διάστημα 70ημερών περίπου τις 200.000 περιπτώσεις (κρούσματα). Εάν παρακολουθήσουμε τις περιπτώσεις μέρα με τη μέρα, όπως φαίνεται στο Σχήμα 1.2, δεν μπορούμε να βγάλουμε κανένα συμπέρασμα σχετικά με μια τάση, είναι όμως σαφές ότι υπάρχει μια εκθετική αύξηση συνολικά παγκοσμίως.

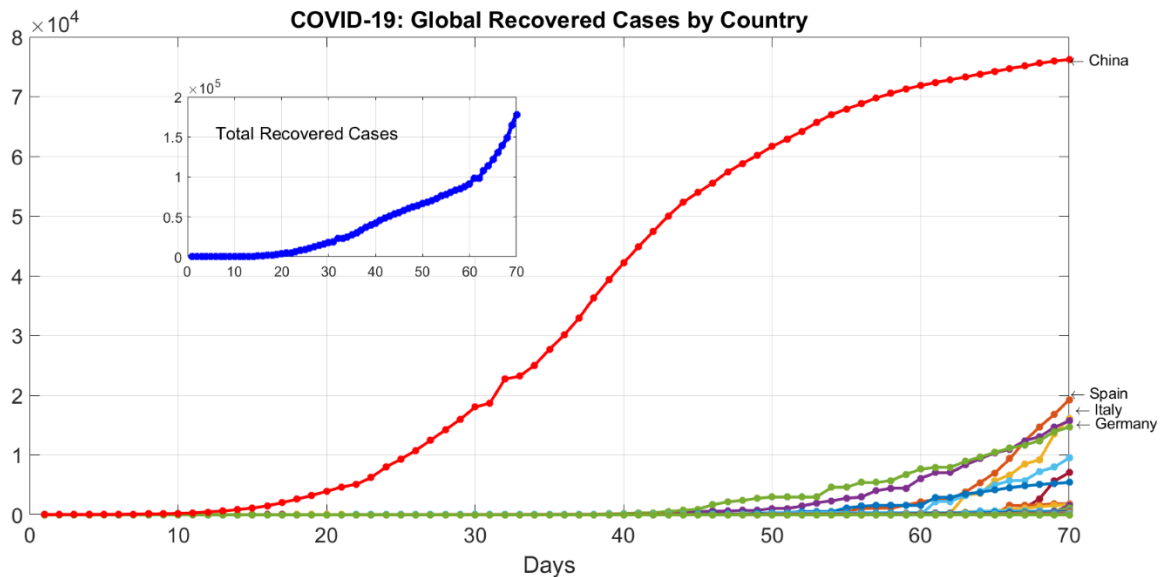


Σχήμα 1.1 Αθροιστική χρονοσειρά επιβεβαιωμένων κρουσμάτων παγκοσμίως για τον COVID-19 από 22 Ιανουαρίου έως 31 Μαρτίου.

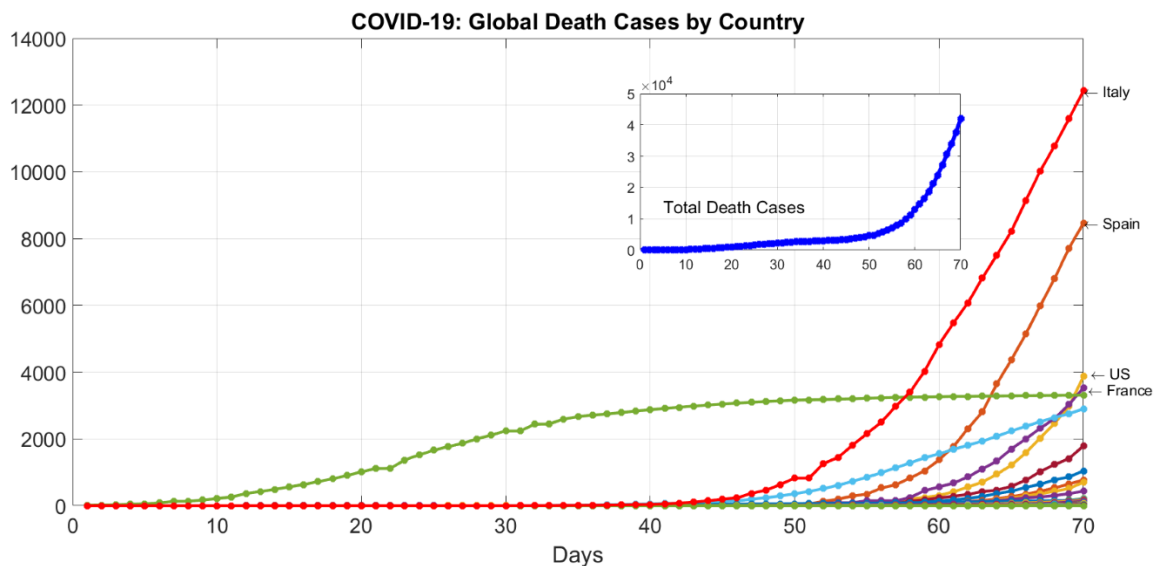


Σχήμα 1.2. Αριθμός ημερήσιων επιβεβαιωμένων κρουσμάτων COVID-19 ανά χώρα από 22 Ιανουαρίου έως 31 Μαρτίου.

Ωστόσο, τι συμβαίνει με εκείνους τους ανθρώπους που έχουν μολυνθεί και αναρρώσει; Τα δεδομένα αυτά συνοψίζονται στο Σχήμα 1.3 και δείχνουν ότι οι άνθρωποι αναρρώνουν από την ασθένειά τους και επιβιώνουν, ενώ άλλοι πεθαίνουν όπως φαίνεται στο Σχήμα 4. Οι τάσεις δείχνουν ότι ο αριθμός των ανθρώπων που επιβιώνουν ξεπερνά τον αριθμό εκείνων που τελικά πεθαίνουν, αλλά δεν υπάρχουν ακόμη αρκετά έγκυρα στοιχεία για τελικά συμπεράσματα ως προς τα δημογραφικά στοιχεία εκείνων που επιβιώνουν και πεθαίνουν.



Σχήμα 1.3. Αριθμός ημερήσιων επιβεβαιωμένων ανθρώπων που ανάρρωσαν από τον COVID-19 ανά χώρα από 22 Ιανουαρίου έως 31 Μαρτίου.



Σχήμα 1.4. Αριθμός ημερήσιων επιβεβαιωμένων θανάτων λόγω COVID-19 ανά χώρα από 22 Ιανουαρίου έως 31 Μαρτίου.

Δεδομένου ότι το ξέσπασμα είναι σχετικά πρόσφατο, δεν έχουν δημοσιευτεί πολλά άρθρα σχετικά με την πρόβλεψη του αντίκτυπου των μεταβλητών που μπορούν να μειώσουν τον συνολικό αριθμό των ατόμων που μολύνθηκαν χρησιμοποιώντας μαθηματικά μοντέλα. Οι Lin et al. υιοθέτησε ένα απλοποιημένο μοντέλο και εξέτασε διάφορες μεταβλητές, συμπεριλαμβανομένων κυβερνητικών ενεργειών, βασικών στοιχείων και άλλων, και εκτίμησε τις παραμέτρους χρησιμοποιώντας ένα μηχανιστικό μοντέλο που χρησιμοποιεί δεδομένα από την πανδημία γρίπης του 1918 στην Αγγλία. Δεν είναι σαφές ότι οι εξεταζόμενες μεταβλητές έχουν άμεση σχέση με το ξέσπασμα ούτε εάν οι τάσεις ισχύουν. Προτείνουμε τη χρήση ενός δυναμικού μοντέλου σε συνδυασμό με ένα φίλτρο Kalman για την εκτέλεση εκτίμησης των μεταβλητών που θεωρήσαμε ότι έχουν δείξει πρόσφατα ότι ελαχιστοποιούν την εξάπλωση του ιού.

Υπάρχει πολύ σημαντική πρακτική σημασία της επιδημιολογικής μοντελοποίησης μέσω μεθόδων προσομοίωσης σε υπολογιστή, η οποία περιλαμβάνει την κατανόηση της νόσου και της δυναμικής του μεταδοτικού ιού, π.χ. πρόβλεψη του αριθμού των νέων κρουσμάτων. Επίσης, το αποτέλεσμα της μοντελοποίησης και των προσομοιώσεων μπορεί να παρέχει ζωτικής σημασίας πληροφορίες για τις κυβερνήσεις και στους υπεύθυνους λήψης αποφάσεων. Ωστόσο, η απόδοση και η αξιοπιστία ενός μαθηματικού μοντέλου εξαρτώνται από τις αβεβαιότητες στο μοντέλο και τις παραμέτρους του μοντέλου. Ενώ ο ακριβής προσδιορισμός των παραμέτρων των μοντέλων, π.χ. το ποσοστό μόλυνσης και τα ποσοστά θνησιμότητας, είναι μια βασική άσκηση στη μαθηματική μοντελοποίηση, η ακριβής πρόβλεψη των παραμέτρων αυτών είναι μια δύσκολη πρόκληση.

Η πρόβλεψη της εξάπλωσης του COVID-19 παίζει σημαντικό ρόλο στην επιδημιολογική μελέτη και τις κυβερνητικές μάχες κατά της επιδημίας. Ωστόσο, οι υπάρχουσες μελέτες για την πρόβλεψη του COVID-19 κυριαρχούνται από σταθερές παραμέτρους του μοντέλου, που δεν μπορούν να αντικατοπτρίσουν την πραγματική κατάσταση της εξάπλωσης του COVID-19. Σε αυτή την εργασία παρουσιάζουμε μια μέθοδο για τη δυναμική πρόβλεψη της εξάπλωσης του COVID-19 λαμβάνοντας υπόψη τις παραμέτρους του μοντέλου που εξαρτώνται από το χρόνο. Αυτή η μέθοδος διακρίνει το επιδημιολογικό μοντέλο(SIRD) στον τομέα του χρόνου για να κατασκευάσει τη μη γραμμική εξίσωση κατάστασης-χώρου για δυναμική εκτίμηση της εξάπλωσης του COVID-19. Στη συνέχεια, αναπτύσσουμε ένα εκτεταμένο φίλτρο Kalman (ExKF) για την εκτίμηση της δυναμικής εξάπλωσης του COVID-19 με βάση τις εκτιμώμενες παραμέτρους του μοντέλου που χρησιμοποιήθηκαν σε αντίστοιχη έρευνα με συνθετικά δεδομένα.

Λόγω της εύκολης πρόσβασης στα δεδομένα σχετικά με τον COVID-19, αρκετοί ερευνητές και μοντελιστές επανεξέτασαν τα κλασικά επιδημιολογικά μοντέλα για να αξιολογήσουν την πρακτική εφαρμογή τους. Τα μαθηματικά μοντέλα, όπως το SIR (Susceptible-Infectious-Recovered), στην επιδημιολογία, εκφράζονται γενικά από ένα σύστημα συνήθων διαφορικών εξισώσεων (ODE). Πρόσφατες μελέτες για τη μοντελοποίηση του COVID-19 περιλαμβάνουν τη χρήση του βασικού μοντέλου SIR ή των επεκτάσεών του (τροποποιημένες) εκδόσεις όπως το SEIR (Susceptible-Exposed-Infectious-Recovered) , SIRD (Susceptible-Infectious-Recovered-Died) και SEIRD (Susceptible-Exposed-Infectious-Recovered-Died).

Καθώς τα διάφορα μαθηματικά μοντέλα μπορούν να προβλέψουν τη δυναμική διαφορών μεταδοτικών ιών, η αποτελεσματικότητα των μοντέλων αυτών εξαρτάται από τις παραμέτρους του μοντέλου. Πρόσφατα, έχουν προταθεί διάφορες μέθοδοι εκτίμησης παραμέτρων για διαφορετικά μοντέλα. Σε αυτή την έρευνα, θα χρησιμοποιήσουμε το Εκτεταμένο Φίλτρο

Kalman (ΕκΚΦ) στην εκτίμηση των χρονικά μεταβαλλόμενων παραμέτρων του μοντέλου με συνθετικά δεδομένα για τον COVID-19.

Επίσης θα συμπεριλάβουμε τον αναδρομικό αλγόριθμο ΕκΚΦ και θα παρουσιάσουμε τον τρόπο με τον οποίο αποδίδει το φίλτρο στην εκτίμηση των παραμέτρων του μοντέλου χρησιμοποιώντας συνθετικά δεδομένα. Ενώ στη συνέχεια θα γίνει και εφαρμογή της μεθόδου και με πραγματικά δεδομένα.

Οι παράμετροι του μαθηματικού μοντέλου μπορούν να εκτιμηθούν επιλύοντας αντίστροφα προβλήματα χρησιμοποιώντας δεδομένα παρατήρησης. Ορισμένες πρόσφατες έρευνες σχετικά με την αντίστροφη μοντελοποίηση COVID-19 με διαφορετικά μοντέλα περιλαμβάνουν έργα των Wan et al., Sun et al., Libotte et al. , Lobato et al. , Li et al. , Αναστασοπούλου κ.ά., Loli Piccolomini and Zama, Korolev , Ndaïrou et al., Ifguis et al., Engbert et al., Yang et al., Arroyo-Marioli et al. και Ghostine et al. Οι πρόσφατες μέθοδοι αντίστροφης μοντελοποίησης (inverse modelling) για την εκτίμηση παραμέτρων περιλαμβάνουν τις τεχνικές ελαχίστων τετραγώνων και διάφορους αλγόριθμους βελτιστοποίησης , τη μέθοδο διαφορικής εξέλιξης, τον αλγόριθμο στοχαστικής και πολλαπλών αντικειμενικών κλασμάτων αναζήτησης και μεθόδους αφομοίωσης δεδομένων.

Στην συγκεκριμένη έρευνα, για την περιγραφή της επιδημίας, θα χρησιμοποιήσουμε το μοντέλο SIRD (Susceptible-Infectious-Recovered-Died) , το οποίο αναπαριστά ικανοποιητικά την εξέλιξη μιας επιδημίας/πανδημίας. Όσον αφορά το μοντέλο SIRD, δεν λαμβάνει υπ' όψη διάφορους παράγοντες που επηρεάζουν την εξέλιξη μιας επιδημίας, όπως για παράδειγμα περιορισμό κυκλοφορίας, χρήση αντισηπτικών, ασυμπτωματικά κρούσματα κ.ά.

Κατά τη διάρκεια της πανδημίας COVID-19, υπήρχε μεγαλύτερος έλεγχος των μεμονωμένων μετακινήσεων λόγω των περιορισμών που επιβλήθηκαν από διάφορες κυβερνήσεις. Οι περιορισμοί περιλάμβαναν lockdown των πόλεων, κοινωνική απόσταση και μέτρα καραντίνας. Άλλα προληπτικά μέτρα ήταν η υγιεινή των χεριών και η χρήση μάσκας προσώπου. Λαμβάνοντας υπόψη την εφαρμογή περιορισμών και προληπτικών μέτρων, σε αυτή τη μελέτη, υποθέτουμε ότι το ποσοστό μόλυνσης, το ποσοστό ανάρρωσης και τα ποσοστά θνησιμότητας των περιπτώσεων COVID-19 ήταν όλα εξαρτώμενα από το χρόνο. Η εκτίμηση των χρονικά μεταβαλλόμενων παραμέτρων ενός μοντέλου μπορεί να είναι δύσκολη και το αντίστροφο πρόβλημα μπορεί να απαιτεί πιο σύνθετα μοντέλα.

Σκοπός αυτής της έρευνας είναι η αξιολόγηση της ικανότητας του Εκτεταμένου Φίλτρου Kalman να εκτιμά τις χρονικά μεταβαλλόμενες παραμέτρους του μοντέλου SIRD και να συγκρίνουμε τις με εκτιμήσεις με τις αντίστοιχες εκτιμήσεις που προέκυψαν από αντίστοιχη έρευνα που πραγματοποιήθηκε στο παρελθόν με την χρήση του Συνολικού Φίλτρου Kalman.



Εικόνα 1.2. Rudolf Emil Kalman (1930-2016).

1.2 Δομή διπλωματικής εργασίας.

Στο Κεφάλαιο 2 θα παρουσιάσουμε αρχικά το γενικευμένο μοντέλο SIR με τις διάφορες παραλλαγές του, ενώ στη συνέχεια θα αναλύσουμε και θα διατυπώσουμε το μαθηματικό μοντέλο SIRD.

Στο Κεφάλαιο 3 θα παρουσιάσουμε το Φίλτρο Kalman και το Εκτεταμένο Φίλτρο Kalman, ενώ στην συνέχεια θα μοντελοποιήσουμε το Εκτεταμένο Φίλτρο για τις ανάγκες της έρευνας.

Στο Κεφάλαιο 4 διατυπώνουμε το μαθηματικό μοντέλο με τις διάφορες παραδοχές που θα διαμορφώσουν την τελική μορφή του προβλήματος. Στην συνέχεια εφαρμόζουμε το Εκτεταμένο Φίλτρο Kalman για την εκτίμηση των χρόνο-μεταβαλλόμενων παραμέτρων του μοντέλου SIRD για διακριτό χρόνο χρησιμοποιώντας συνθετικά δεδομένα και στην συνέχεια πραγματικά δεδομένα για την επαρχία της Κίνας, Χουμπέι. Τέλος θα αναλύσουμε και σχολιάσουμε τα αποτελέσματα που θα προκύψουν καθώς και θα τα συγκρίνουμε με την αντίστοιχη έρευνα όπου εφαρμόστηκε το Συνολικό Φίλτρο Kalman.

Τα τελικά συμπεράσματα της διπλωματικής εργασίας και κατευθύνσεις για περαιτέρω έρευνα παρουσιάζονται στο 5

Το Παράρτημα περιλαμβάνει τον κώδικα σε γλώσσα προγραμματισμού Python που χρησιμοποιήσαμε για τους υπολογισμούς και τα αποτελέσματα που αναφέρονται στην έρευνα.

1.3 Βασικές έννοιες

Στοιχεία από τη θεωρία πιθανοτήτων

Παρακάτω θα δούμε κάποια από τα εργαλεία της θεωρίας των πιθανοτήτων που θα μας βοηθήσουν να κατανοήσουμε την έννοια του λευκού θορύβου (white noise), αλλά και τις ποσότητες που θα πρέπει να 'βελτιστοποιήσουμε' ώστε να πάρουμε την ζητούμενη εκτίμηση.

Πείραμα τύχης: Ένα πείραμα, δηλαδή ένα γεγονός ή μια διαδικασία της οποίας το αποτέλεσμα δεν μπορεί να προβλεφθεί ή να υπολογιστεί. Το σύνολο όλων των δυνατών αποτελεσμάτων ενός τέτοιου πειράματος ονομάζεται δειγματοχώρος Ω .

Κλασική Πιθανότητα: Ως πιθανότητα ενός γεγονότος ορίζουμε το πηλίκο του πλήθους των ευνοϊκών για την πραγματοποίηση του γεγονότος αποτελεσμάτων προς το σύνολο όλων των δυνατών αποτελεσμάτων.

Τυχαία Μεταβλητή: Μια συνάρτηση η οποία παίρνει τυχαίες τιμές. Για παράδειγμα, το αποτέλεσμα X της ρίψης ενός κέρματος είναι τυχαία μεταβλητή. Σε κάθε τιμή της τυχαίας μεταβλητής αντιστοιχεί μια πιθανότητα, π.χ. στο πείραμα της ρίψης ενός κέρματος έχουμε πιθανότητα $P(X=\text{κορόνα})=1/2$.

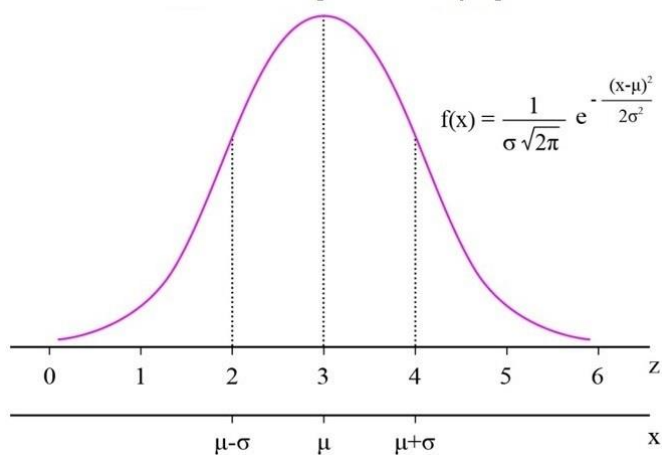
Στην περίπτωση που μιλάμε για συνεχείς τυχαίες μεταβλητές (με τις οποίες θα ασχοληθούμε στην έρευνα), η πιθανότητα η τυχαία μεταβλητή να πάρει μια συγκεκριμένη τιμή δίνεται από την $f(x)$, η οποία ονομάζεται συνάρτηση πυκνότητας πιθανότητας (probability density function-pdf). Η πιθανότητα η τυχαία μεταβλητή να πάρει μια τιμή μέσα σε ένα διάστημα $[a,b]$ δίνεται από το ολοκλήρωμα της συνάρτησης πυκνότητας πιθανότητας, δηλαδή $P(a < X < b) = \int_a^b f(x) dx$

Υπάρχουν πολλές χαρακτηριστικές συναρτήσεις κατανομής, που χρησιμοποιούνται ευρέως σε εφαρμογές και εκφράζουν διάφορα φυσικά φαινόμενα. Αυτή που θα μας απασχολήσει είναι η Κανονική (Normal) ή Gaussian κατανομή. Η συνάρτηση πυκνότητας πιθανότητάς έχει τη μορφή:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

όπου το ελληνικό γράμμα «μ» συμβολίζεται η μέση τιμή ενώ με το ελληνικό γράμμα «σ» συμβολίζεται η τυπική απόκλιση της κατανομής. Η μορφή της κατανομής, όπως θα δούμε και παρακάτω, θυμίζει καμπάνα, γι' αυτό ονομάζεται καμπανοειδής κατανομή, που σημαίνει ότι η πιθανότητα των γεγονότων γύρω από τη μέση τιμή είναι πολύ υψηλή σε σχέση με την πιθανότητα των ακραίων τιμών.

Κανονική κατανομή



Εικόνα 1.3. Διάγραμμα κανονικής κατανομής.

Παρατήρηση: Στην Κανονική Κατανομή τα μόνα μεγέθη που έχουν σημασία για τη συνάρτηση πυκνότητας πιθανότητας είναι η μέση τιμή και η τυπική απόκλιση. Καταλαβαίνουμε λοιπόν, πως αρκεί να προσδιορίσουμε τη μέση τιμή και η τυπική απόκλιση για να περιγράψουμε πλήρως μια τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή.

Μέση τιμή της τυχαίας μεταβλητής (Mean value) (μ): Η μέση τιμή είναι το πιο βασικό στατιστικό μέτρο και αποτελεί γενίκευση του μέσου όρου μιας ακολουθίας τιμών. Η ερμηνεία της είναι ότι μας δίνει ένα κέντρο βάρους γύρω από το οποίο παίρνει τιμές η τυχαία μεταβλητή.

Διασπορά ή Διακύμανση (variance) και τυπική απόκλιση (Standard Deviation Std): Η διασπορά μας δίνει πληροφορία για το πόσο απομακρυσμένες γύρω από τη μέση τιμή βρίσκονται οι τιμές της τυχαίας μεταβλητής και συμβολίζεται:

$$\text{Var}[X]=E[(X-\mu)^2]$$

Η τετραγωνική ρίζα της διασποράς ονομάζεται τυπική απόκλιση σ (standard deviation), δηλαδή $\sigma^2=\text{Var}[X]$.

Συνδιακύμανση ή Συνδιασπορά (Covariance): Η συνδιακύμανση μεταξύ δύο τυχαίων μεταβλητών X και Y μας δείχνει πως και εάν σχετίζονται οι 2 μεταβλητές μεταξύ τους, δηλαδή κατά πόσο έχουν την ίδια συμπεριφορά. Δίνεται από τον τύπο:

$$\text{Cov}(X,Y)=E[(X-\mu_x)(Y-\mu_y)]$$

Αν Για $X=Y$ έχουμε τον τύπο της διακύμανσης.

Αν $\text{Cov}(X,Y)=0$ οι δύο μεταβλητές ονομάζονται ανεξάρτητες.

Θεωρία στοχαστικών προτύπων

Στοχαστική διαδικασία : ονομάζεται ένα σύνολο τυχαίων μεταβλητών ορισμένων σε ένα χώρο πιθανοτήτων. Εάν είναι αριθμήσιμο το πλήθος τους τότε η διαδικασία συμβολίζεται με $X_t : X_1, X_2, X_3, \dots$. Εάν το πλήθος δεν είναι αριθμήσιμο, τότε η διαδικασία συμβολίζεται με $X(t): t \geq 0$. Στην πρώτη περίπτωση αναφερόμαστε σε διαδικασία σε διακριτό χρόνο, ενώ στη δεύτερη σε στοχαστική διαδικασία σε συνεχή χρόνο.

Στις περισσότερες περιπτώσεις η μεταβλητή t εκφράζει χρόνο και η στοχαστική διαδικασία συμβολίζει την εξέλιξη ενός γεγονότος με την πάροδο του, π.χ. ο αριθμός πωλήσεων ενός προϊόντος τους μήνες ενός χρόνου ή η στάθμη των βροχοπτώσεων σε μια περιοχή κάθε μήνα για έναν χρόνο.

Επειδή όπως είπαμε, οι στοχαστικές διαδικασίες εκφράζουν την εξέλιξη ενός φαινομένου με την πάροδο του χρόνου, τα X έχουν την ίδια συνάρτηση πιθανότητας (ή πυκνότητας πιθανότητας) της οποίας όμως τα στατιστικά χαρακτηριστικά μπορεί να διαφοροποιούνται από χρόνο σε χρόνο.

Λευκός Θόρυβος (White noise): Ως λευκό θόρυβο, ορίζουμε μια στοχαστική διαδικασία η οποία είναι Gaussian με μέση τιμή $\mu=0$ και ανεξάρτητη από στιγμή σε στιγμή. Ας αναλύσουμε κάθε μια ιδιότητα ξεχωριστά, ώστε να κατανοήσουμε πλήρως τα χαρακτηριστικά του θορύβου. Είναι Gaussian, δηλαδή κάθε τυχαία μεταβλητή σε κάθε χρονική στιγμή ακολουθεί την

κανονική συνάρτηση κατανομής με μέση τιμή ίση με 0. Επιπλέον, είναι ανεξάρτητη μεταξύ χρονικών στιγμών δηλαδή $Cov(X_{t_1}, X_{t_2}) = E[X_{t_1}, X_{t_2}] = 0$ για κάθε $t_1 \neq t_2$.

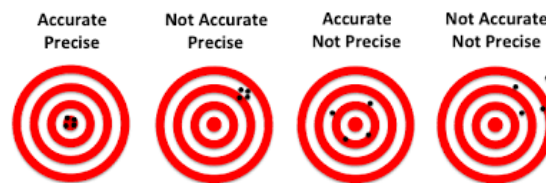
Στα μοντέλα που χρησιμοποιούμε, θεωρούμε πως οι διαταραχές εκφράζονται ως λευκοί θόρυβοι. Μελέτες τέτοιων διαταραχών σε συστήματα έχουν δείξει ότι όντως οι περισσότεροι θόρυβοι μετρήσεων ακολουθούν Gaussian κατανομή με μέση τιμή ίση με μηδέν. Μάλιστα έχει παρατηρηθεί πως το άθροισμα πολλών διαταραχών με διαφορετικές κατανομές τείνει να έχει Gaussian κατανομή. Επομένως ο λευκός θόρυβος αποτελεί μια πολύ καλή μοντελοποίηση των αληθινών διαταραχών.

Εκτίμηση (Estimate): Από μαθηματική σκοπιά, με τον όρο εκτίμηση εννοούμε ότι έχουμε έναν χονδρικό υπολογισμό της αξίας, του αριθμού, της ποσότητας ή της έκτασης κάποιας ποσότητας. Χρησιμοποιούμε εκτιμήσεις αριθμών για να κάνουμε ευκολότερα και γρηγορότερα υπολογισμούς όταν μας καλύπτει μια απάντηση που είναι πολύ κοντά στην πραγματική ποσότητα.

Ακρίβεια (Accuracy): Η ακρίβεια δείχνει πόσο κοντά είναι μια μέτρηση στην πραγματική τιμή.

Ευστοχία (Precision): Η ευστοχία περιγράφει πόση μεταβλητότητα υπάρχει σε έναν αριθμό μετρήσεων της ίδιας παραμέτρου. Η ακρίβεια και η ευστοχία αποτελούν τη βάση της εκτίμησης.

Αναδρομικός αλγόριθμος: Ο όρος αναδρομικός αλγόριθμος ή φίλτρο αναφέρεται στην ιδιότητα του να χρησιμοποιεί μέρος προηγούμενης πληροφορίας χωρίς να χρειάζεται όλη η προσφερόμενη πληροφορία να αποθηκευτεί για να επεξεργαστεί ξανά κάθε φορά που μια καινούργια μέτρηση παρουσιάζεται.



Εικόνα 1.4. Διαφορά ευστοχίας και ακρίβειας.

ΚΕΦΑΛΑΙΟ 2: Επιδημιολογικά μοντέλα

2.1 Τα επιδημιολογικά μοντέλα

Τα διαμερισματικά μοντέλα είναι μια πολύ γενική τεχνική μοντελοποίησης. Συχνά εφαρμόζονται στη μαθηματική μοντελοποίηση μολυσματικών ασθενειών. Ο πληθυσμός εκχωρείται σε διαμερίσματα με ετικέτες – για παράδειγμα, S, I ή R, (Ευπαθείς, Μολυνσμένοι, ή Αναρρώσαντες). Οι άνθρωποι μπορεί να προχωρούν μεταξύ των διαμερισμάτων. Η σειρά των ετικετών δείχνει συνήθως τα μοτίβα ροής μεταξύ των διαμερισμάτων.

Η προέλευση τέτοιων μοντέλων έγινε στις αρχές του 20ου αιώνα, με σημαντικά έργα του Ross το 1916, των Ross και Handson το 1917, Kermack και McKendrick το 1927 και Kental το 1956.

Τα μοντέλα συνήθως εκτελούνται με συνηθισμένες διαφορικές εξισώσεις (οι οποίες είναι ντετερμινιστικές), αλλά μπορούν επίσης να χρησιμοποιηθούν με ένα στοχαστικό (τυχαίο) πλαίσιο, το οποίο είναι πιο ρεαλιστικό αλλά πολύ πιο περίπλοκο στην ανάλυση.

Τα μοντέλα προσπαθούν να προβλέψουν ποσότητες όπως για παράδειγμα το πώς εξαπλώνεται μια ασθένεια ή ο συνολικός αριθμός μολύνσεων ή τη διάρκεια μιας επιδημίας και να εκτιμήσουν διάφορες επιδημιολογικές παραμέτρους. Τέτοια μοντέλα μπορούν να δείξουν πώς οι διαφορετικές παρεμβάσεις στη δημόσια υγεία μπορεί να επηρεάσουν την έκβαση της επιδημίας, π.χ. ποια είναι η πιο αποτελεσματική τεχνική για την έκδοση περιορισμένου αριθμού εμβολίων σε έναν δεδομένο πληθυσμό ή άλλα μέτρα και περιορισμούς που μπορεί να εφαρμόσει ένα κράτος.

Αρχικά θα παρουσιάσουμε το ευρέως διαδεδομένο συνεχές επιδημιολογικό μοντέλο SIR. Έστω, με βάση και αυτά που αναφέραμε παραπάνω, μια συγκεκριμένη ασθένεια, η οποία οδηγεί είτε σε ανοσία, μετά τη θεραπεία, είτε σε θάνατο.

Σε αυτή την περίπτωση ο πληθυσμός του είδους μας μπορεί να χωριστεί στις παρακάτω τρεις κατηγορίες:

1. Ευπαθείς (Susceptibles, S) : Άτομα του πληθυσμού, τα οποία είναι υγιείς αλλά μπορούν να μολυνθούν.
2. Μολυσμένοι (Infected, I) : Άτομα του πληθυσμού, τα οποία νοσούν (ασθενείς).
3. Διαγραμμένοι (Removed, R) : Άτομα του πληθυσμού τα οποία είτε έχουν θεραπευτεί και έχουν αποκτήσει ανοσία, είτε έχουν πεθάνει.

Σε μορφή διαγράμματος (θεωρούμε ότι τα βέλη δείχνουν την χρονική εξέλιξη) η χρονική μετάβαση για ένα άτομο του πληθυσμού είναι της μορφής:

$$S \rightarrow I \rightarrow R$$

Παρατηρούμε ότι η τελευταία κατηγορία (R) περιλαμβάνει τόσο όσους έχουν θεραπευτεί, αλλά και όσους έχουν πεθάνει. Εδώ κάνουμε την παραδοχή ότι και οι δύο ομάδες δεν επηρεάζονται πλέον με την ασθένεια, εφόσον οι πρώτοι έχουν αποκτήσει ανοσία, οπότε μπορούν να περιγραφούν με μια κοινή μεταβλητή. Αυτό μας επιτρέπει να δούμε το σύνολο του πληθυσμού, N, ως μια σταθερά, ανεξάρτητη του χρόνου, δηλαδή:

$$S(t) + I(t) + R(t) = N = N(t).$$

Μία παραδοχή που κάνουμε εδώ και θα χρησιμοποιήσουμε και στη συνέχεια είναι ότι το σύνολο του πληθυσμού N θεωρείται σταθερό, δηλαδή ανεξάρτητο του χρόνου αφού το διάστημα των μετρήσεων είναι πολύ μικρό (μία ημέρα) και οι μεταβολές της κατηγορίας δR είναι αμελητέες σε σύγκριση με το N.

Οι παραπάνω κατηγορίες, εύκολα περιγράφονται με τις αντίστοιχες μεταβλητές S(t), I(t), R(t) οι οποίες δηλώνουν τον αριθμό των ατόμων σε κάθε κατηγορία σε χρόνο dt. Στόχος μας, όπως σε κάθε μαθηματικό μοντέλο είναι η εύρεση των σχέσεων (εδώ συνήθων διαφορικών εξισώσεων), οι οποίες περιγράφουν τις παραπάνω πληθυσμιακές ομάδες.

2.1.1 Συνεχές μοντέλο SIR

Θεωρούμε σημαντικό να εξηγήσουμε αρχικά το γενικό μοντέλο SIR για ευκολότερη κατανόηση και στη συνέχεια να παρουσιάσουμε το μοντέλο SIRD που θα χρησιμοποιήσουμε στην συγκεκριμένη έρευνα.

Το πρώτο μοντέλο SIR έχει προταθεί από τους Kermack – McKendrick το 1927 και βασίζεται στις παρακάτω παραδοχές:

- Ο αριθμός των ασθενών/μολυσμένων (I) αυξάνει ανάλογα με τον αριθμό των επαφών μεταξύ ασθενών και ευπαθών με ένα γραμμικό όρο ως προς το γινόμενο της μορφής βSI .
- Ο αριθμός των ευπαθών (S) αντίστοιχα μειώνεται ανάλογα με τον αριθμό των επαφών μεταξύ ασθενών και ευπαθών με τον όρο βSI .
- Ο αριθμός των διαγραφμμένων (R) αυξάνει ανάλογα με τον αριθμό των ασθενών, κατά ένα ποσοστό γI .
- Ενώ αντίστοιχα με το ίδιο ποσοστό γI μειώνεται ο αριθμός των ασθενών.

Με βάση τα παραπάνω, οι Συνήθεις Διαφορικές Εξισώσεις του μοντέλου SIR είναι:

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (2.1)$$

$$\frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I \quad (2.2)$$

$$\frac{dR}{dt} = \gamma I \quad (2.3)$$

όπου β είναι ο ρυθμός μετάδοσης (transmission rate) [1/time], γ είναι ο ρυθμός διαγραφής (ανοσίας και θανάτου) [1/time] οι οποίες είναι θετικές σταθερές. Θεωρούμε επίσης ότι οι αρχικές συνθήκες του μοντέλου είναι:

$$S(0)=S_0, \quad I(0)=I_0 \quad \text{και} \quad R(0)=R_0, \quad S_0+I_0+R_0=N.$$

όπου S_0, I_0, R_0 είναι θετικές σταθερές.

Παρατηρούμε ότι εάν $I_0=0$ ή $S_0=0$ δεν υπάρχει εξέλιξη της επιδημίας/πανδημίας. Επίσης συνήθως θεωρούμε ότι $R_0=0$.

Βασικό ερώτημα σε κάθε επιδημιολογική μελέτη είναι η εύρεση δεδομένων για τον προσδιορισμό των παραμέτρων του προβλήματος, που αφορούν την ασθένεια και τον πληθυσμό ($\beta, \gamma, S_0, I_0, R_0$) εάν η ασθένεια θα εξαπλωθεί, θα γίνει δηλαδή επιδημία, ή όχι. Η εξέλιξη μιας ασθένειας εξαρτάται από τον λόγο $\rho = \gamma/\beta$ καθώς και από το ρυθμό αναπαραγωγής του φαινομένου (reproduction rate) που συμβολίζεται ως \mathcal{R} . Για να εξετάσουμε το συγκεκριμένο ερώτημα ως υπολογίσουμε την μεταβολή του πληθυσμού των ασθενών σε χρόνο $t=0$ δηλαδή στον χρόνο όπου παρατηρήθηκε για πρώτη φορά η ασθένεια, η οποία είναι:

$$\frac{dI}{dt} = I_0(\beta S_0 - \gamma) \quad (2.4)$$

Διακρίνουμε δύο περιπτώσεις για την παραπάνω παράγωγο:

- Παρατηρείτε ότι η παραπάνω παράγωγος θα είναι αρνητική εάν: $S_0 < \rho$,

τότε $dI/dt < 0$ για αρχικό χρόνο $t=0$ αλλά και για κάθε χρόνο $t > 0$. Το τελευταίο ισχύει καθώς: $dS/dt \leq 0$, δηλαδή $S \leq S_0$. Σε αυτήν λοιπόν την περίπτωση ο αριθμός των ασθενών μειώνεται συνεχώς και η επιδημία δεν εξελίσσεται.

- Αν αντίθετα : $S_0 > \rho$ η παράγωγος είναι θετική, δηλαδή σε αυτή την περίπτωση ο αριθμός των ασθενών, $I(t)$, αυξάνει με το χρόνο. Σε αυτή την περίπτωση έχουμε εξάπλωση της ασθένειας, δηλαδή η ασθένεια γίνεται επιδημία.

Βλέπουμε λοιπόν ότι ο λόγος $\rho = \gamma/\beta$, είναι η κρίσιμη παράμετρος για την εξάπλωση, ή όχι, της επιδημίας.

Επιπλέον μια άλλη σημαντική ποσότητα με την οποία περιγράφουμε μια επιδημία είναι ο ρυθμός αναπαραγωγής οποίος ορίζεται ως : $\mathcal{R} = \beta S_0/\gamma$

Ο ρυθμός αναπαραγωγής της ασθένειας, βιολογικά εκφράζει τον αριθμό δευτερογενών μολύνσεων από ένα μολυσμένο μέλος σε ολόκληρο τον πληθυσμό.

Αν $\mathcal{R} > 1$ υπάρχουν περισσότερες από μια μόλυνση και δημιουργείται η επιδημία. Σε αυτή την περίπτωση η αποτροπή της επιδημίας μπορεί να επιτευχθεί μέσω μείωσης του ρυθμού αναπαραγωγής \mathcal{R} . Αυτό γίνεται εφικτό, για παράδειγμα με τον εμβολιασμό, ο οποίος μειώνει τον αρχικό αριθμό του ευπαθούς πληθυσμού S_0 .

2.1.2 Διακριτό Μοντέλο SIR

Στη συνέχεια θα παρουσιάσουμε το διακριτό επιδημιολογικό μοντέλο SIR. Για τη μελέτη αυτού του μοντέλου εργαζόμαστε όπως και στα διακριτά μοντέλα που είδαμε παραπάνω.

Πιο συγκεκριμένα αν θεωρήσουμε και πάλι ως S_t , I_t και R_t τους πληθυσμούς σε χρονική στιγμή t , τότε κάνοντας τις ανάλογες παραδοχές με αυτές του συνεχούς μοντέλου SIR που είδαμε παραπάνω, το διακριτό SIR μοντέλο, δηλαδή οι εξισώσεις διαφορών που περιγράφουν τους πληθυσμούς την επόμενη χρονική στιγμή, έχουν τη μορφή:

$$S_{t+1} = S_t - \beta S_t I_t / N \quad (2.5)$$

$$I_{t+1} = I_t + \beta S_t I_t / N - \gamma I_t \quad (2.6)$$

$$R_{t+1} = R_t + \gamma I_t \quad (2.7)$$

Παρατήρηση: Σε κάθε διακριτό μοντέλο, ιδιαίτερη προσοχή απαιτείται στη σωστή επιλογή της χρονικής διάρκειας μεταξύ των δύο επαναλήψεων, Δt . Το βήμα αυτό μπορεί για παράδειγμα να είναι μια ώρα, μια μέρα ή μια εβδομάδα ανάλογα με το είδος της ασθένειας που μελετάμε (Στην συγκεκριμένη έρευνα το $\Delta t = 1$ ημέρα).

Το παραπάνω μοντέλο, μπορεί επίσης να γραφεί στη μορφή :

$$\Delta S = -\beta S_t I_t / N \quad (2.8)$$

$$\Delta I = \beta S_t I_t / N - \gamma I_t = I (\beta S - \gamma) \quad (2.9)$$

$$\Delta R = \gamma I_t \quad (2.10)$$

Και πάλι, μας ενδιαφέρει κυρίως η χρονική εξέλιξη του πληθυσμού των ασθενών ΔI . Εύκολα διαπιστώνουμε ότι η κρίσιμη παράμετρος είναι ο λόγος γ/β όπως αναφέραμε παραπάνω.

Παρόμοια λοιπόν με το συνεχές μοντέλο ορίζουμε τον συντελεστή αναπαραγωγής της ασθένειας ως: $\mathcal{R} = \beta \cdot S_0 / \gamma$

Αν $\mathcal{R} > 1$, τότε και το $\Delta I > 0$, επομένως θα έχουμε εξέλιξη της ασθένειας σε επιδημία.

2.2 Παραλλαγές του Επιδημιολογικού Μοντέλου SIR

Όπως αναφέρθηκε και παραπάνω το μοντέλο SIR είναι καλό για την περίπτωση ασθένειας στην οποία οι ασθενείς που θεραπεύονται αποκτούν ανοσία. Σε πολλές περιπτώσεις όμως αυτό δεν είναι καλή υπόθεση και χρειάζονται διαφορετικά μοντέλα. Παρακάτω θα δούμε, πολύ συνοπτικά, κάποιες παραλλαγές του μοντέλου SIR.

Το μοντέλο SI

Αν θέλουμε να μελετήσουμε κάποια ασθένεια, στην οποία δεν υπάρχει ίαση τότε η χρονική εξέλιξη, μεταξύ των διαφορετικών ομάδων, για ένα άτομο του πληθυσμού είναι της μορφής:

$$S \rightarrow I$$

Σε αυτή την περίπτωση οι εξισώσεις του μοντέλου θα έχουν τη μορφή:

$$S_{t+1} = S_t - \beta S_t I_t / N \quad (2.11)$$

$$I_{t+1} = I_t + \beta S_t I_t / N \quad (2.12)$$

Το μοντέλο SIS

Στη περίπτωση ασθένειας για την οποία δεν υπάρχει ανοσία,). Η πληθυσμιακή μετακίνηση μεταξύ των ομάδων είναι “διπλής διαδρομής”, με ροή από τους S στους I αλλά και από τους S στους I. Το μοντέλο μπορεί να περιγραφεί ως:

$$S \leftrightarrow I$$

Σε αυτή την περίπτωση οι εξισώσεις του μοντέλου θα έχουν τη μορφή:

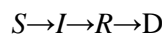
$$S_{t+1} = S_t - \beta S_t I_t / N + \gamma I_t \quad (2.13)$$

$$I_{t+1} = I_t + \beta S_t I_t / N - \gamma I_t \quad (2.14)$$

Το παραπάνω μοντέλο είναι αρκετά συνηθισμένο καθώς περιγράφει περιπτώσεις ασθενειών στις οποίες οι ασθενείς που θεραπεύονται δεν αποκτούν ανοσία, όπως συμβαίνει για παράδειγμα στην κοινή γρίπη.

Το μοντέλο SIRD

Το μοντέλο SIRD με το οποίο θα ασχοληθούμε αποτελεί παραλλαγή του γενικού μοντέλου SIR, προσθέτοντας μια επιπλέον πληθυσμιακή ομάδα για να περιγράψει τους νεκρούς ξεχωριστά από τους αναρρώσαντες. Οι ομάδες που διακρίνονται βάσει του μοντέλου αυτού είναι οι ευάλωτοι στην ασθένεια S, οι μολυσμένοι I που έχουν διαγνωσθεί, οι αναρρώσαντες R (Recovered) και οι νεκροί D (Deaths). Η πληθυσμιακή μετακίνηση μεταξύ των ομάδων είναι μονόδρομη, με ροή από τους S στους I και μετά είτε στους R είτε στους D. Το μοντέλο μπορεί να περιγράφει ως:



όπου:

Το μοντέλο SIRD περιγράφεται με το παρακάτω σύστημα συνήθων διαφορικών εξισώσεων (Ordinary Differential Equations):

$$\frac{dS}{dt} = -\beta IS/N \quad (2.15)$$

$$\frac{dI}{dt} = \beta IS/N - \gamma I - \delta I \quad (2.16)$$

$$\frac{dR}{dt} = \gamma I \quad (2.17)$$

$$\frac{dD}{dt} = \delta I \quad (2.18)$$

όπου:

N : Συνολικός πληθυσμός, $N = S + I + R + D = \text{σταθερό}$.

t : Χρόνος που έχει περάσει από την έναρξη της πανδημίας.

β : ο ρυθμός μετάδοσης (transmission rate) [1/time].

γ : ο ρυθμός ανάρρωσης (recovery rate) [1/time].

δ : ο ρυθμός θνησιμότητας (mortality rate) [1/time].

Ο ρυθμός αναπαραγωγής \mathcal{R} για τον μοντέλο SIRD υπολογίζεται ως εξής:

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \delta I = 0 \quad (2.19)$$

$$\Rightarrow \left(\frac{\beta S}{N} - \gamma - \delta\right)I = 0 \quad (2.20)$$

Ο μολυσμένος πληθυσμός δεν μπορεί να είναι $I=0$ επομένως:

$$\Rightarrow \left(\frac{\beta S}{N} - \gamma - \delta\right) = 0 \quad \Rightarrow \frac{\beta S}{N(\gamma + \delta)} = 1 \quad (2.21)$$

Επιπλέον στο ξεκίνημα της έξαρσης το σύνολο του πληθυσμού θεωρείται ευάλωτο στον ιό επομένως $S \approx N$ άρα τελικά, ο αρχικός ρυθμός αναπαραγωγής είναι:

$$\mathcal{R}_0 = \frac{\beta}{(\gamma + \delta)} \quad (2.22)$$

Όσον αφορά την συγκεκριμένη έρευνα θα χρησιμοποιήσουμε το διακριτό επιδημιολογικό μοντέλο SIRD το οποίο θα δούμε εκτενέστερα σε επόμενο κεφάλαιο και περιγράφεται από τις παρακάτω εξισώσεις:

$$S_{k+1} = S_k - \beta_k S_k I_k / N \quad (2.23)$$

$$I_{k+1} = I_k + \beta_k S_k I_k / N - \gamma_k I_k - \delta_k I_k \quad (2.24)$$

$$R_{k+1} = R_k + \gamma_k I_k \quad (2.25)$$

$$D_{k+1} = D_k + \delta_k I_k \quad (2.26)$$

ΚΕΦΑΛΑΙΟ 3: Φίλτρα Kalman

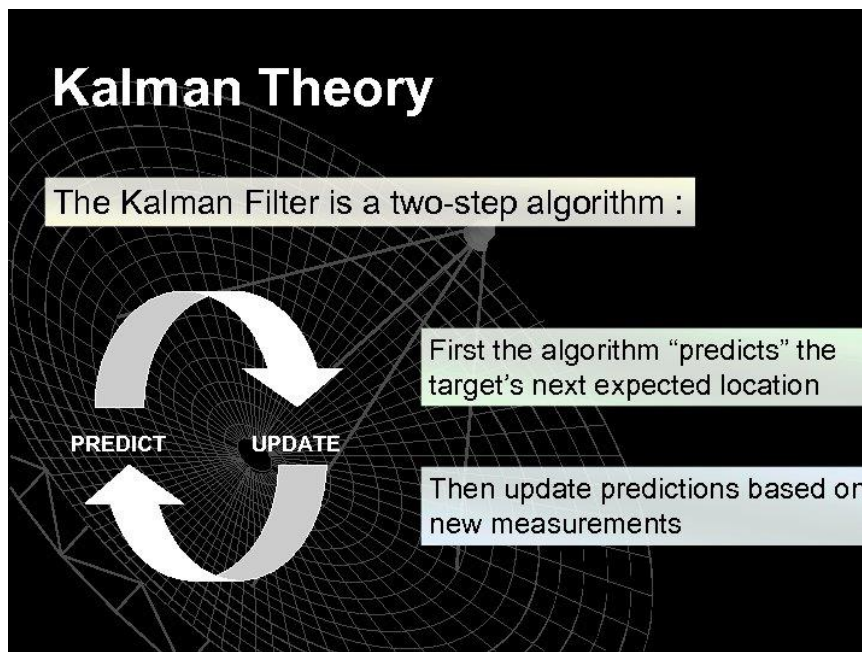
Γενικά για το Φίλτρο Kalman

Το Φίλτρο Kalman είναι ένας εκτιμητής του λεγόμενου γραμμικού προβλήματος ελαχίστων τετραγώνων όπου καλείται κανείς να υπολογίσει τη στιγμιαία “κατάσταση” (state) ενός γραμμικού δυναμικού συστήματος που διαταράσσεται από “λευκό θόρυβο” (white noise). Δηλαδή από μετρήσεις που σχετίζονται γραμμικά με τη δυναμική κατάσταση του συστήματος αλλά είναι αλλοιωμένες από “λευκό θόρυβο”. Ο εκτιμητής είναι στατιστικά βέλτιστος ως προς οποιαδήποτε συνάρτηση του τετραγώνου του σφάλματος εκτίμησης.

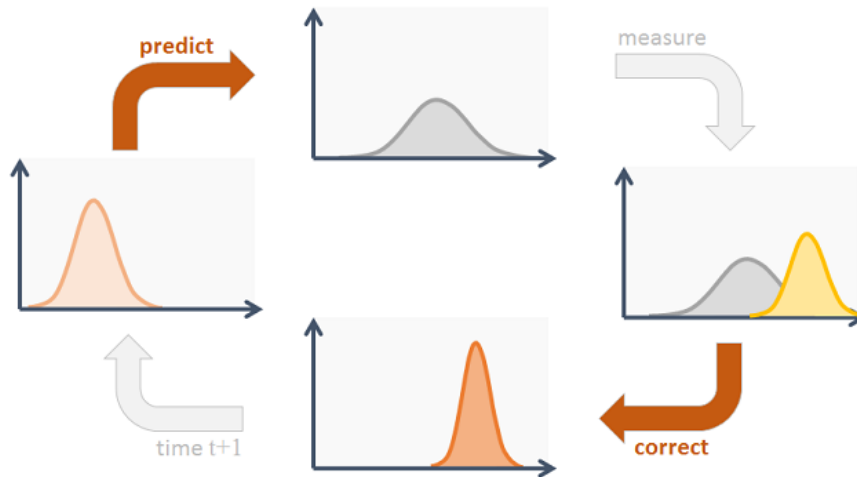
Τα περισσότερα σήματα αισθητήρων επηρεάζονται από κάποιο θόρυβο που μπορεί να ασκήσει αρνητική επίδραση στην απόδοση του αισθητήρα ή στον έλεγχο των μετρήσεων.

Έτσι λοιπόν, είναι απαραίτητο να υπάρχει κάποιο φίλτρο, ικανό να εξαλείψει κάθε παράγοντα παραμόρφωσης της εκτίμησης της κατάστασης του συστήματος. Το φίλτρο Kalman καλύπτει την παραπάνω αναγκαιότητα. Έτσι παρέχει βέλτιστη εκτίμηση των μεταβλητών που ενδιαφέρουν το σύστημα, κατά τέτοιο τρόπο ώστε να ελαχιστοποιείται το σφάλμα.

Το Φίλτρο Kalman αποτελεί μια από τις σημαντικότερες τεχνολογικές εφευρέσεις του περασμένου αιώνα. Είναι ένας αλγόριθμος που εφαρμόζεται σε συστήματα που δέχονται εξωτερικές φυσικές διαταραχές (θορύβους), με σκοπό την απαλοιφή των διαταραχών των μετρήσεων που γίνονται καθώς και τη δημιουργία μιας νέας εκτίμησης της κατάστασης του συστήματος.



Εικόνα 3.1. Η θεωρία του φίλτρου Kalman (α).



Εικόνα 3.2. Η θεωρία του φίλτρου Kalman (β).

Η σημασία του Φίλτρου Kalman

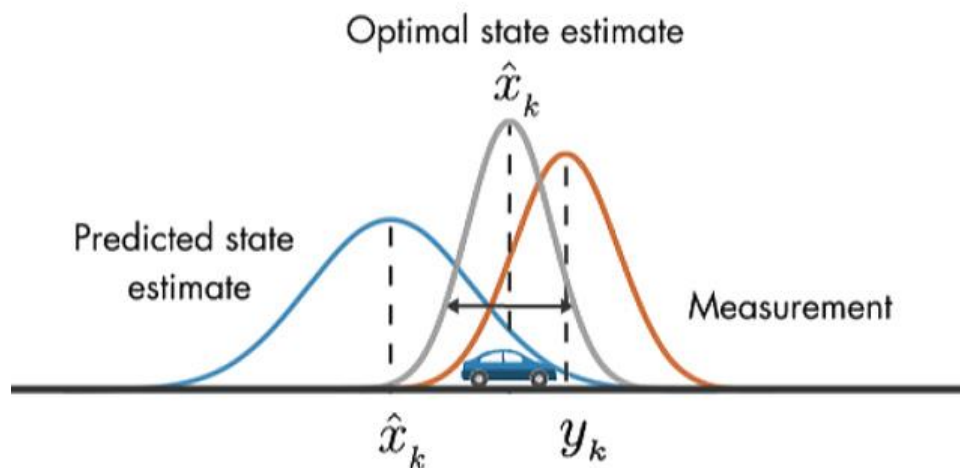
Η ανάπτυξη του φίλτρου, αποτελεί ένα από τα σημαντικότερα κατορθώματα του περασμένου αιώνα, τουλάχιστον από επιστημονική άποψη. Μια τέτοια δήλωση δεν είναι καθόλου υπερβολική, αφού το φίλτρο Kalman αποτέλεσε τη γέφυρα ανάμεσα στη θεωρητική μελέτη και στην πρακτική εφαρμογή. Δεδομένου πως κάθε διεργασία ή μέτρηση επηρεάζεται από θόρυβο, δε θα είχαμε τη δυνατότητα να αναπτύξουμε πολλές εφαρμογές αν δεν βρίσκαμε τρόπο να προβούμε σε αποφάσεις, αποβάλλοντας την αβεβαιότητα που δίνει ο θόρυβος. Το Φίλτρο Kalman εφαρμόζεται στη ναυπηγική, την εκτίμηση θέσης ενός αυτοκινήτου (GPS), την εξομάλυνση μηνυμάτων (ήχου είτε εικόνας) που έχουν φθαρεί από θόρυβο, σε Radar, γενικά σε οποιαδήποτε τεχνολογική διαδικασία απαιτεί υψηλή ακρίβεια δεδομένων. Επίσης μπορεί να χρησιμοποιηθεί σε οικονομικά μοντέλα, παραδείγματος χάρη τα Autoregressive-moving-average-models για τη μελέτη και την πρόβλεψη της συμπεριφοράς χρονοσειρών καθώς και στη μελέτη βιολογικών δομών, όπως μοντέλα πληθυσμών.

Ο Geir Evenson ήταν εκείνος που εισήγαγε το Συνολικό Φίλτρο Kalman (EnKF) το οποίο είναι μια τεχνική αφομοίωσης δεδομένων που μπορεί να χρησιμοποιηθεί για την ενημέρωση και διόρθωση τόσο των παραμέτρων του μοντέλου όσο και των μεταβλητών καταστάσεων με τις σχετικές αβεβαιότητες. Ο Yang et al. χρησιμοποίησε το EnKF για κοινή εκτίμηση παραμέτρων κατάστασης, όπου εκτιμήθηκε μία από τις παραμέτρους του μοντέλου SEIR, ο χρονικά μεταβαλλόμενος ρυθμός μόλυνσης. Ομοίως, οι Li et al. χρησιμοποίησε το Ensemble Adjustment Kalman Filter (EAKF) στον χώρο επαυξημένης παραμέτρου κατάστασης για να εκτιμήσει τις παραμέτρους ενός τροποποιημένου ντετερμινιστικού μοντέλου SEIR. Η

επαυξημένη παράμετρος κατάστασης μπορεί να προκαλέσει την αποτυχία του EnKF λόγω μιας ισχυρής μη γραμμικής σχέσης μεταξύ των παραμέτρων του μοντέλου και της κατάστασής του. Για να ξεπεραστεί αυτό, οι Engbert et al. υιοθέτησε μια προσέγγιση δύο σταδίων όπου το EnKF χρησιμοποιήθηκε αρχικά για την εκτίμηση των καταστάσεων, ακολουθούμενο από το συμπέρασμα με βάση την πιθανότητα μιας από τις παραμέτρους του μοντέλου SEIR.

Μια άλλη παραδοχή είναι πως ολόκληρος ο πληθυσμός είναι ευάλωτος στη νόσο στο πρώτο επίπεδο της επιδημίας. Ωστόσο, λιγότερα άτομα του συνόλου S (Susceptibles) είναι ευαίσθητα μέσω μέτρων ελέγχου, όπως ο περιορισμός των κινήσεων, η αυτοαπομόνωση και η κοινωνική απόσταση. Καθώς ο αρχικός αριθμός των ευαίσθητων ατόμων, $S(0)$, απαιτείται για την αντίστροφη μοντελοποίηση, η εκτίμηση του πραγματικού μεγέθους πληθυσμού υπό μελέτη μπορεί να είναι μια δύσκολη πρόκληση. Στις πρόσφατες μελέτες που εξετάζουν την αντίστροφη μοντελοποίηση του COVID-19, χρησιμοποιήθηκαν διάφορες υποθέσεις ή μέθοδοι για την επιλογή του συνολικού μεγέθους πληθυσμού, N . Αυτές περιλαμβάνουν την υπόθεση μιας σταθερής τιμής για το N , π.χ. πληθυσμό μιας πόλης ή μιας χώρας ή χρησιμοποιώντας μια κανονικοποιημένη έκδοση ενός διαμερισματικού μοντέλου.

Τα Φίλτρα Kalman στα οποία θα αναφερθούμε στην συγκεκριμένη έρευνα, αναφέρονται σε διακριτό χρόνο και θα τα χωρίσουμε σε εκείνα που χρησιμοποιούνται για γραμμικά συστήματα και σε μη-γραμμικά συστήματα.



Εικόνα 3.3. Η χρήση του Φίλτρου Kalman στα συστήματα πλοήγησης GPS για την εκτίμηση της ακριβούς τοποθεσίας ενός οχήματος.

3.1 Φίλτρο Kalman Διακριτού Χρόνου για γραμμικά συστήματα

Έστω ότι έχουμε την εξίσωση διαφορών $\mathbf{x}_k = \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{w}_{k-1}$ που περιγράφει το σύστημα διακριτού χρόνου, όπου με τον όρο \mathbf{x}_k συμβολίζουμε το $[13 \times 1]$ διάνυσμα κατάστασης, $\mathbf{A}_k = \mathbf{A}(t_{k+1}, t_k)$ είναι ο πίνακας μετάβασης $[13 \times 13]$ από τη χρονική στιγμή t_k στην t_{k+1} , ο οποίος προκύπτει από τη λύση της ομογενούς διαφορικής εξίσωσης και το \mathbf{w}_k είναι ένα διάνυσμα $n \times 1$, δηλαδή μια πολυδιάστατη στοχαστική διαδικασία με χαρακτηριστικά λευκού θορύβου, δηλαδή $E[\mathbf{w}] = 0$. Επίσης θεωρούμε $E[\mathbf{w}_k, \mathbf{w}_j^T] = \mathbf{Q} \delta(k-j)$ με τον πίνακα διακύμανσης \mathbf{Q} (process noise matrix) $[13 \times 13]$.

Το δ ονομάζεται δέλτα Kronecker για το οποίο ισχύει:

$$\delta_{ij} \equiv \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j. \end{cases}$$

Για το ίδιο σύστημα έχουμε ένα $[3 \times 1]$ διάνυσμα μετρήσεων y_k το οποίο περιγράφεται από την εξίσωση $y_k = \mathbf{C}_k \mathbf{x}_k + v_k$. Το y_k δηλαδή είναι ένας γραμμικός συνδυασμός των γνωστών καταστάσεων x_k όπου προστίθεται και ένα σφάλμα μετρήσεων v_k που επίσης είναι μια 3×1 πολυδιάστατη στατική στοχαστική διαδικασία με τα χαρακτηριστικά λευκού θορύβου $E[v] = 0$. Ο πίνακας \mathbf{C}_k ονομάζεται πίνακας μετρήσεων, αφού δίνει τις μετρήσεις των y_k από τα x_k , χωρίς το θόρυβο. Εξαρτάται από το k , γιατί στην γενική περίπτωση τον θεωρούμε χρονικά μεταβαλλόμενο (πράγμα που συμβαίνει στα πιο ανεπτυγμένα συστήματα).

Έστω ότι για το σύστημα μας έχουμε μια αρχική, εκ των προτέρων (a priori) εκτίμηση της κατάστασης στην χρονική στιγμή \mathbf{x}_k την οποία συμβολίζουμε ως $\hat{\mathbf{x}}_k^-$. Αναζητούμε μια βελτιωμένη εκτίμηση $\hat{\mathbf{x}}_k^+$, κάνοντας χρήση των μετρήσεων y_k . Επειδή στην πράξη μιλάμε για φίλτρα πεπερασμένης μνήμης, η εκτίμηση αναζητείται με τον αναδρομικό τύπο $\hat{\mathbf{x}}_k^+ = f(\hat{\mathbf{x}}_k^-, y_k)$ ως γραμμικός συνδυασμός της a priori εκτίμησης $\hat{\mathbf{x}}_k^-$ και των μετρήσεων y_k . Οι πίνακες θα προσδιοριστούν με κριτήριο την ελαχιστοποίηση του σφάλματος μεταξύ της εκτίμησης μας και της πραγματικής τιμής του διανύσματος x .

Ορίζουμε τον πίνακα συνδιασποράς (covariance matrix) ως:

$$\mathbf{P}_0 = E[\hat{\mathbf{x}}_k^+, \hat{\mathbf{x}}_k^{+T}] \quad [13 \times 13] \quad (3.1)$$

και τον πίνακα κέρδους (Kalman Gain Matrix):

$$\mathbf{K}_k = \mathbf{A}_k \mathbf{P}_k \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k \mathbf{C}_k^T + \mathbf{R}_k)^{-1} \quad [13 \times 3] \quad (3.2)$$

με τον $\mathbf{R}_k = E[v_k, v_k^T]$ (measurement noise matrix)

ενώ στη συνέχεια ο πίνακας συνδιασποράς ξανά υπολογίζεται ως:

$$\mathbf{P}_{k+1}^+ = [\mathbf{A}_k - \mathbf{K}_k \mathbf{C}_k] \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q}_k \quad [13 \times 13] \quad (3.3)$$

Παρατηρήσεις:

Ο βέλτιστος πίνακας κέρδους είναι αυτός που ελαχιστοποιεί το σφάλμα μεταξύ της εκτίμησης και της πραγματικής τιμής του διανύσματος x , δηλαδή την ποσότητα $(y_k - \hat{x}_k)$. Ο πίνακας \mathbf{P} πρέπει να είναι συμμετρικός και θετικά ορισμένος. Σε περίπτωση που δεν πληρούνται αυτές οι προϋποθέσεις πιθανόν να υπάρχει υπολογιστικό σφάλμα.

Γνωρίζοντας λοιπόν τους πίνακες μετάβασης \mathbf{A}_k [13x13], πίνακα μετρήσεων \mathbf{C}_k [3x13], τον αρχικό πίνακα συνδιασποράς \mathbf{P}_0 [13x13] και τους πίνακες θορύβων \mathbf{Q} [13x13], \mathbf{R} [3x3] είμαστε σε θέση να συνοψίσουμε το φίλτρο Kalman στα εξής βήματα (αλγόριθμο) :

1. Υπολογίζουμε τα \mathbf{P}_k^- και \hat{x}_k χρησιμοποιώντας τους $\mathbf{P}_{k-1}^+, \mathbf{A}_{k-1}, \mathbf{Q}_{k-1}$,
2. Υπολογίζουμε τον \mathbf{K}_k με χρήση του \mathbf{P}_k^- .
3. Υπολογίζουμε εκ νέου τον \mathbf{P}_k^+ με χρήση του \mathbf{K}_k που υπολογίσαμε στο προηγούμενο βήμα.
4. Υπολογίζουμε τη νέα εκτίμηση \hat{x}_k^+ με χρήση των \mathbf{K}_k , \hat{x}_k^- και των μετρήσεων y_k .

Η παραπάνω διαδικασία σε μορφή εξισώσεων εκφράζεται ως εξής:

- Έχουμε το σύστημα κατάστασης: $\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{w}_k$ με $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q})$ (3.4)

- και τις μετρήσεις: $\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k$ με $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R})$ (3.5)

όπου τα w, v είναι λευκού θορύβου Gaussian ακολουθίες με τα \mathbf{Q}, \mathbf{R} γνωστά.

Έστω η αρχική εκτίμηση $\hat{x}(0) = E[x_{k0}]$ τότε $\mathbf{P}(0) = E[\hat{x}(0), \mathbf{x}(0)]$.

- Υπολογίζουμε τον πίνακα κέρδους από τη σχέση (3.2):
 $\mathbf{K}_k = \mathbf{A}_k \mathbf{P}_k^- \mathbf{C}_k^T [\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}]^{-1}$ [13x3]
- Υπολογίζουμε εκ νέου τον πίνακα συνδιασποράς από τη σχέση (3.3):
 $\mathbf{P}_k^+ = [\mathbf{A}_k - \mathbf{K}_k \mathbf{C}_k] \mathbf{P}_k^- \mathbf{A}_k^T + \mathbf{Q}$
- Παίρνοντας τη μέτρηση y_k και γνωρίζοντας το x_k για τη δεδομένη χρονική στιγμή k , παίρνουμε την εκτιμώμενη τιμή \hat{x}_{k+1} από την εξίσωση :

$$\hat{x}_{k+1} = \hat{x}_k + \mathbf{K}_k [y_k - \mathbf{C}_k \hat{x}_k] \quad [13 \times 1] \quad (3.6)$$

Το πρώτο μέρος της παραπάνω σχέσης $\mathbf{A}_k \hat{x}_k$ ονομάζεται μοντέλο, ενώ το δεύτερο μέρος $\mathbf{K}_k [y_k - \hat{x}_k]$ ονομάζεται διορθωτικός παράγοντας. Η εκτιμώμενη τιμή που υπολογίζεται με την παραπάνω διαδικασία είναι η ζητούμενη ποσότητα που μας ενδιαφέρει.

Παρατήρηση:

Όπως αναφέραμε και νωρίτερα, σκοπός του πίνακα κέρδους είναι να ελαχιστοποιήσει την ποσότητα $y_k - \hat{x}_k$. Η διαφορά $y_k - \hat{x}_k$ είναι στην ουσία ο θόρυβος του συστήματος που θέλουμε να απαλείψουμε.

3.2 Το εκτεταμένο Φίλτρο Kalman για μη γραμμικά συστήματα

Στη θεωρία εκτίμησης, το εκτεταμένο φίλτρο Kalman (EKF) είναι η μη γραμμική (non-linear) έκδοση του φίλτρου Kalman. Ο αλγόριθμος που χρησιμοποιείται για το εκτεταμένο φίλτρο Kalman και η γενική τακτική του, είναι να διαμορφώνει όσο το δυνατόν πιο γραμμικά (linearize) την τελευταία τιμή του μέσου (mean) και της συνδιασποράς (covariance).

Στην περίπτωση καλώς ορισμένων μοντέλων μετάβασης, το EKF έχει θεωρηθεί το βασικότερο πρότυπο στη θεωρία της μη γραμμικής εκτίμησης κατάστασης.

Βασικό μειονέκτημα όμως του εκτεταμένου φίλτρου Kalman σε αντίθεση με το διακριτό φίλτρο, είναι πως οι διάφορες μεταβλητές που περιγράφουν το σύστημα παύουν να ακολουθούν κανονική κατανομή τη στιγμή που διαφοροποιούνται γραμμικά.

Για το λόγο αυτό το φίλτρο εκτιμά την κατάσταση του συστήματος προσεγγίζοντας την γραμμικά ευνοϊκότερη δυνατή συνθήκη, κατά Bayes. Τέλος όπως και στο γραμμικό Kalman φίλτρο, κάθε επανάληψη έχει δύο (2) κύκλους, μία πρόβλεψης και μία διόρθωσης. Η διαφορά είναι ότι κατά τον κύκλο πρόβλεψης η εκτίμηση του διανύσματος της κατάστασης γίνεται με το μη γραμμικό μοντέλο ενώ για την εκτίμηση του πίνακα σφάλματος συνδιασποράς γίνεται πρώτα γραμμικοποίηση.

Οι εξισώσεις οι οποίες περιγράφουν τον αλγόριθμο του εκτεταμένου Φίλτρου Kalman είναι παρόμοιες με εκείνες του Φίλτρου Kalman για γραμμικά συστήματα, με κάποιες διαφορές.

- Το μη γραμμικό σύστημα εξισώσεων κατάστασης και μετρήσεων έχει πλέον τη μορφή: $\mathbf{x}_{k+1} = \mathbf{f}[\mathbf{x}_k, \mathbf{w}_k]$ και $\mathbf{y}_k = \mathbf{g}[\mathbf{x}_k] + \mathbf{v}_k$ αντίστοιχα. Οι συναρτήσεις \mathbf{f} , \mathbf{g} εκφράζουν το μη γραμμικό σύστημα που εξετάζουμε. Στην δικιά μας περίπτωση η συνάρτηση \mathbf{f} είναι το σύστημα διαφορικών εξισώσεων που περιγράφουν το επιδημιολογικό μοντέλο SIRD, ενώ η συνάρτηση \mathbf{g} είναι ο τελεστής παρατήρησης που συνδέει τις παρατηρούμενες τιμές με τις τιμές κατάστασης του μοντέλου όπως θα δούμε αναλυτικότερα και στο επόμενο κεφάλαιο.

- Ο πίνακας κέρδους έχει την μορφή που ήδη γνωρίζουμε από την (3.2):

$$\mathbf{K}_k = \mathbf{A}_k \mathbf{P}_k \mathbf{C}_k^T [\mathbf{C}_k \mathbf{P}_k \mathbf{C}_k^T + \mathbf{R}]^{-1}$$

- Ο πίνακας συνδιασποράς έχει επίσης την μορφή που ήδη γνωρίζουμε από την (3.3): $\mathbf{P}_k^+ = [\mathbf{A}_k - \mathbf{K}_k \mathbf{C}_k] \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q}$, (υπενθυμίζοντας $\mathbf{P}_k \geq \mathbf{0}$ και συμμετρικός)

- Τέλος, για την νέα εκτίμηση έχουμε:

$$\hat{\mathbf{x}}_{k+1} = \mathbf{f}[\hat{\mathbf{x}}_k, 0] + \mathbf{K}_k [\mathbf{y}_k - \mathbf{g}[\hat{\mathbf{x}}_k]] \quad (3.7)$$

Στο Εκτεταμένο Φίλτρο Kalman ο πίνακας μετάβασης \mathbf{A}_k υπολογίζεται από την Jacobian της f ως προς το διάνυσμα \mathbf{x}_k , και θέτοντας ως $\mathbf{x}_k = \hat{\mathbf{x}}_k$ και $\mathbf{w}_k = 0$. Είναι δηλαδή:

$$\mathbf{A} = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{x}}_k} \quad [13 \times 13] \quad (3.8)$$

Όσον αφορά τον πίνακα μετρήσεων, υπολογίζεται από την Jacobian της $g[k]$ ως προς το διάνυσμα \mathbf{x}_k θέτοντας $\mathbf{x} = \hat{\mathbf{x}}_k$. Είναι δηλαδή:

$$\mathbf{C} = \left. \frac{\partial g}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{x}}_k} \quad [3 \times 13] \quad (3.9)$$

Μια σημαντική παρατήρηση για το Εκτεταμένο Φίλτρο Kalman είναι ότι οι τιμές του πίνακα $\hat{\mathbf{x}}_k$ δεν συγκλίνουν πάντα.

ΚΕΦΑΛΑΙΟ 4: Εφαρμογή του Εκτεταμένου φίλτρου Kalman στο επιδημιολογικό μοντέλο SIRD για τον COVID19

4.1 Μαθηματική Μοντελοποίηση

Σε αυτό το σημείο θεωρούμε σημαντικό να υπενθυμίσουμε το επιδημιολογικό μοντέλο SIRD, το οποίο θα χρησιμοποιηθεί στη συνέχεια.

Το SIRD (Susceptible, Infectious, Recovered, and Dead) είναι ένα μοντέλο τεσσάρων καταστάσεων που έχει χρησιμοποιηθεί ευρέως ως μέθοδος πρόβλεψης μολυσματικών ασθενειών. Στο μοντέλο SIRD, ο αριθμός των ευπαθών ατόμων (S), των μολυσμένων ατόμων (I), των αναρρωμένων ατόμων (R) και των νεκρών ατόμων (D) ποικίλλει με το χρόνο k ως εξής:

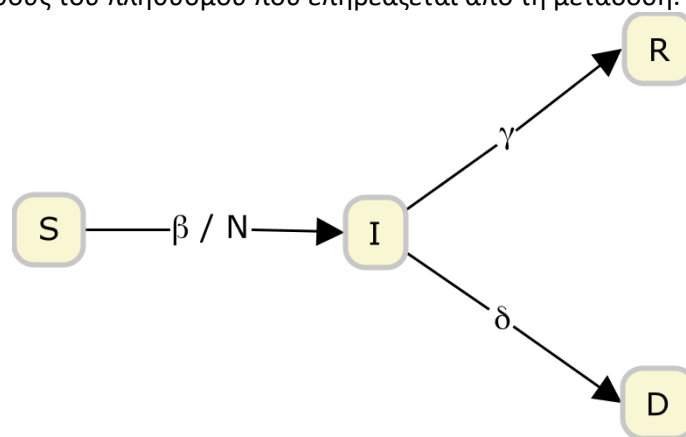
$$\frac{dS}{dk} = -\beta \frac{IS}{N} \quad (4.1)$$

$$\frac{dI}{dk} = \beta \frac{IS}{N} - \gamma I - \delta I \quad (4.2)$$

$$\frac{dR}{dk} = \gamma I \quad (4.3)$$

$$\frac{dD}{dk} = \delta I \quad (4.4)$$

όπου β είναι ο ρυθμός μετάδοσης (μόλυνση), γ είναι ο ρυθμός ανάρρωσης και δ είναι ο ρυθμός θανάτων. Η εκτίμηση του μοντέλου που χρησιμοποιεί δεδομένα COVID-19 χρησιμοποιεί δεδομένα για περιπτώσεις που έχουν διαγνωσθεί. Ως εκ τούτου, σε αυτό το εμπειρικό πλαίσιο, το δ μπορεί να θεωρηθεί καλύτερα ως το ποσοστό θνησιμότητας. Ακολουθώντας τους Calafiore et al. και Ianni και Rossi, το N ορίζεται ως το κλάσμα του συνολικού μεγέθους του πληθυσμού που επηρεάζεται από τη μετάδοση.



Σχήμα 4.1. Διάγραμμα ροής μοντέλου SIRD.

Το μοντέλο υποθέτει ότι κάθε άτομο που έχει ήδη μολυνθεί μπορεί να μεταδώσει τον ιό σε άτομα που είναι ευαίσθητα. Επιπλέον, η χρονική διάρκεια θεωρείται μικρή, ώστε να παραμελούνται οι γεννήσεις και οι θάνατοι που δεν σχετίζονται με τον ιό. Το μοντέλο SIRD

δεν λαμβάνει υπόψη τις επιπτώσεις της έκθεσης, της καραντίνας, του εγκλεισμού ή ενός ασυμπτωματικού πληθυσμού. Αυτό το μοντέλο είναι κατάλληλο για την περίπτωση χωρίς μέτρα προστασίας και περιορισμούς στις δραστηριότητες, π.χ. φορώντας μάσκες και μέτρα lockdown. Η ικανότητα του απλού μοντέλου SIRD να αποτυπώνει τη δυναμική του COVID-19 έχει αποδειχθεί από τους Fanelli και Piazza και Anastassopoulou et al.

Στο μοντέλο, $N_k = S_k + I_k + R_k + D_k$ που θεωρείται σταθερό. Η λύση του συστήματος ODE εξαρτάται από τις αρχικές συνθήκες $S_0 = S(0)$, $I_0 = I(0)$, $R_0 = R(0)$ και $D_0 = D(0)$ για τους ευαίσθητους, μολυσμένους, αναρρωμένους και νεκρούς πληθυσμούς, αντίστοιχα. Ωστόσο, ο αρχικός αριθμός του ευαίσθητου πληθυσμού, $S(0) = N(0) - I(0) - R(0) - D(0)$, είναι συνήθως άγνωστος αφού το $N_0 = N(0)$ είναι τυπικά άγνωστο. Αυτή η μελέτη υποθέτει ότι ολόκληρος ο πληθυσμός, είναι ευάλωτος στη νόσο στο πρώτο επίπεδο της επιδημίας και αφήνουμε $N = N_k = N(0)$.

Παράγοντες όπως οι περιορισμένες μετακινήσεις ατόμων, το lockdown των πόλεων, η κοινωνική απόσταση, τα μέτρα καραντίνας και τα προληπτικά μέτρα όπως η υγιεινή των χεριών και η χρήση μάσκας προσώπου επιτρέπουν στον ρυθμό μετάδοσης να ποικίλλει με την πάροδο του χρόνου.



Εικόνα 4.1. Ατομικά μέτρα προστασίας κατά του κορονοϊού.

Ως εκ τούτου, τα ποσοστά μετάδοσης, ανάρρωσης και θνησιμότητας επιτρεπόταν όλα να εξαρτώνται από το χρόνο σε αυτή τη μελέτη. Παρόμοια με τους Avila-Ponce de León et al. και Gupta et al., οι τρεις χρονικά μεταβαλλόμενες παράμετροι ορίστηκαν ως εξής:

Ο ρυθμός μόλυνσης: ο χρονικά μεταβαλλόμενος ρυθμός μόλυνσης πριν και μετά το lockdown περιγράφεται ως εξής:

$$\beta_k = \begin{cases} \beta_0, & k < t_{lockdown} \\ \beta_0 \exp\left(-\frac{k-t_{lockdown}}{\tau_\beta}\right) + \beta_1, & k \geq t_{lockdown} \end{cases} \quad (4.5)$$

Η $\beta(k)$ είναι συνάρτηση τριών χαρακτηριστικών σταθερών β_0 , β_1 και τ_β . Πριν από το lockdown, το $\beta(k) = \beta_0$ είναι μια σταθερά. Όταν το lockdown επιβάλλεται τη στιγμή $k = t_{lockdown}$, το $\beta(k)$ μειώνεται εκθετικά από $\beta_0 + \beta_1$ στην τελική τιμή β_1 με χαρακτηριστικό χρόνο μείωσης τ_β .

Ο ρυθμός ανάρρωσης: Με μια νέα ασθένεια όπως το COVID-19, το σύστημα υγειονομικής περίθαλψης και το ιατρικό προσωπικό πρέπει να μάθουν και να υιοθετήσουν νέες θεραπευτικές διαδικασίες, συμπεριλαμβανομένης της θεραπείας ασθενών με νέα συμπτώματα. Ως εκ τούτου, ο χρόνος ανάρρωσης για τους ασθενείς μπορεί να αλλάξει με την πάροδο του χρόνου. Σε αυτή τη μελέτη, η $\gamma(t)$ περιγράφεται από τη συνάρτηση :

$$\gamma_k = \gamma_0 + \frac{\gamma_1}{1 + \exp(-k + \tau_\gamma)} \quad (4.7)$$

όπου γ_0 είναι ο αρχικός ρυθμός ανάρρωσης και μετά από $k = \tau_\gamma$ ο τελικός ρυθμός ανάρρωσης γίνεται $\gamma_0 + \gamma_1$.

Το ποσοστό θνησιμότητας: το ποσοστό θνησιμότητας μπορεί επίσης να μειωθεί με την πάροδο του χρόνου λόγω παραγόντων όπως η προσαρμογή του παθογόνου και η ανάπτυξη προηγμένων θεραπειών και εμβολιασμών, συμπεριλαμβανομένων μη φαρμακευτικών παρεμβάσεων όπως η κοινωνική απόσταση, ο αποκλεισμός των πόλεων και η αύξηση της ευαισθητοποίησης του κοινού για την ασθένεια. Ο ρυθμός θανάτων $\delta(t)$ περιγράφεται χρησιμοποιώντας τη συνάρτηση:

$$\delta_k = \begin{cases} \delta_0, & k < t_{lockdown} \\ \delta_0 \exp\left(-\frac{k-t_{lockdown}}{\tau_\delta}\right) + \delta_1, & k \geq t_{lockdown} \end{cases} \quad (4.8)$$

όπου τη στιγμή $k = t_{lockdown}$, $\delta_0 + \delta_1$ είναι ο αρχικός ρυθμός θανάτων που μειώνεται εκθετικά στην τελική τιμή δ_1 με χαρακτηριστικό χρόνο μείωσης τ_δ .

Για να προσομοιώσει το μοντέλο SIRD μια συγκεκριμένη επιδημία με τις τρεις χρονικά μεταβαλλόμενες παραμέτρους, οι εννέα χαρακτηριστικές σταθερές (β_0 , β_1 , τ_β , γ_0 , γ_1 , τ_γ , δ_0 , δ_1 , τ_δ), πρέπει να εκτιμηθούν μέσω αντίστροφης μοντελοποίησης.

4.2 Το Εκτεταμένο Φίλτρο Kalman για εκτίμηση παραμέτρων

Εκτός από ένα επιδημιολογικό μοντέλο, η δυναμική μοντελοποίηση της πανδημίας COVID-19 απαιτεί επίσης έναν αναδρομικό αλγόριθμο σε πραγματικό χρόνο για την εκτίμηση της κατάστασης μετάδοσης. Η αναδρομική σχέση ελαχίστων τετραγώνων (LSM), όπως έχει αναφερθεί είναι μια παραδοσιακή μέθοδος για εκτίμηση παραμέτρων στην επιδημιολογική μοντελοποίηση. Μπορεί να δημιουργήσει βέλτιστη εκτίμηση κατάστασης μέσω της ελαχιστοποίησης της συνάρτησης κόστους γραμμικών ελαχίστων τετραγώνων που σχετίζεται με τις παρατηρήσεις του συστήματος. Ως βελτίωση της μεθόδου αυτής, το φίλτρο Kalman (KF) εισάγει την εξίσωση κατάστασης συστήματος στο LSM για να υπολογίσει τη διάδοση της κατάστασης του συστήματος μέσω μιας διαδικασίας πρόβλεψης. Μπορεί να επιτύχει βέλτιστη εκτίμηση κατάστασης στην ακρίβεια του ελάχιστου μέσου τετραγωνικού σφάλματος, ακόμη και αν δεν υπάρχουν παρατηρήσεις.

Οι Argyo-Marioli et al. χρησιμοποίησαν το Φίλτρο Kalman για να υπολογίσουν τον βασικό αριθμό αναπαραγωγής του COVID-19 με βάση μια γραμμική μορφή του μοντέλου SIR. Ωστόσο, τα LSM και KF μπορούν να εφαρμοστούν μόνο σε γραμμικά συστήματα, ενώ τα υπάρχοντα επιδημιολογικά μοντέλα για την πρόβλεψη του COVID-19 είναι μη γραμμικά. Τα περιορισμένα ελάχιστα τετράγωνα (CLS) και η αλυσίδα Markov Monte Carlo (MCMC) είναι η συνήθως χρησιμοποιούμενη μέθοδος εκτίμησης για μη γραμμικά επιδημιολογικά μοντέλα. Ωστόσο, δεδομένου ότι το CLS και το MCMC βασίζονται σε μέγιστες εκ των υστέρων εκτιμήσεις της συνάρτησης πυκνότητας πιθανότητας, η ακρίβεια και των δύο μεθόδων εξαρτάται σε μεγάλο βαθμό από το μέγεθος του δείγματος. Επιπλέον, και οι δύο μέθοδοι περιλαμβάνουν επίσης δαπανηρούς υπολογισμούς. Συνεπώς, είναι ακατάλληλα για τον χαρακτηρισμό τυχαίων αβεβαιοτήτων που εμπλέκονται στις παραμέτρους του επιδημιολογικού μοντέλου για την πρόβλεψη του COVID-19.

Μία βελτίωση του Φίλτρου Kalman για μη γραμμικά συστήματα, το Εκτεταμένο Φίλτρο Kalman (ExKF) γραμμικοποιεί δυναμικά το μη γραμμικό μοντέλο συστήματος για να χρησιμοποιήσει το παραδοσιακό KF για την εκτίμηση της κατάστασης. Σε σύγκριση με το CLS και το MCMC, το ExKF είναι ένας απλός επαναληπτικός αλγόριθμος με σημαντική υπολογιστική αποτελεσματικότητα για την μη γραμμική επιδημιολογική μοντελοποίηση. Μέχρι στιγμής, έχει υπάρξει πολύ περιορισμένη έρευνα σχετικά με τη χρήση του ExKF για επιδημιολογική μοντελοποίηση, ειδικά στον COVID-19. Ο Younes και ο Hassan ανέπτυξαν ένα ExKF με βάση το μοντέλο Lotka-Volterra για την εκτίμηση της εξάπλωσης του COVID-19. Ωστόσο, ο μηχανισμός αλληλεπίδρασης θηρευτή-θηράματος που περιγράφεται από το μοντέλο Lotka-Volterra έχει πολύ περιορισμένη ικανότητα να μοντελοποιεί τα πολύπλοκα χαρακτηριστικά της φυσικής διαδικασίας μετάδοσης του COVID-19. Ο Song et al. μελέτησε ένα νέο ExKF με βάση τη μέγιστη πιθανότητα για την εκτίμηση της εξάπλωσης του COVID-19. Ωστόσο, δεδομένου ότι αυτή η μέθοδος βασίζεται σε ένα ντετερμινιστικό επιδημιολογικό μοντέλο, δεν είναι σε θέση να χαρακτηρίσει τα стоχαστικά χαρακτηριστικά της εξάπλωσης του COVID-19. Επιπλέον, δεν λαμβάνει υπόψη ούτε τις επιπτώσεις της επαναμόλυνσης.

Παρακάτω θα παρουσιάσουμε μια στοχαστική μέθοδο για την εκτίμηση και την πρόβλεψη της πορείας της εξάπλωσης του COVID-19.

Θεωρούμε ένα διακριτό μη γραμμικό μοντέλο:

$$\mathbf{s}_{k+1} = \mathcal{S}(\mathbf{s}_k, \boldsymbol{\theta}_k) + \mathbf{w}_{k+1} \quad (4.10)$$

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \mathbf{v}_{k+1} \quad (4.11)$$

$$\mathbf{y}_{k+1} = \mathcal{M}(\mathbf{s}_{k+1}, \boldsymbol{\theta}_{k+1}) + \mathbf{e}_{k+1} \quad (4.12)$$

όπου $\mathbf{s}_k = [S_k \ I_k \ R_k \ D_k]$ είναι το διάνυσμα των μεταβλητών κατάστασης τη στιγμή $t = k$, το \mathcal{S} είναι ο μη γραμμικός τελεστής (επιδημιολογικό μοντέλο SIRD), $\boldsymbol{\theta}_k = [\beta_{0k}, \beta_{1k}, \tau_{\beta k}, \gamma_{0k}, \gamma_{1k}, \tau_{\gamma k}, \delta_{0k}, \delta_{1k}, \tau_{\delta k}]$ είναι το διάνυσμα των παραμέτρων που υποτίθεται ότι παραμένουν σταθερές στο χρόνο $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$ προσθέτοντας σε αυτές έναν «μικρό θόρυβο», το \mathbf{w}_{k+1} είναι ο θόρυβος του μοντέλου που θεωρούμε ότι ακολουθεί Gaussian ακολουθία με μηδενική μέση τιμή με πίνακα συνδιακύμανσης \mathbf{Q}_{k+1} , $[13 \times 13]$ το \mathbf{y}_{k+1} είναι το διάνυσμα της παρατήρησης (ενεργός αριθμός μολυσμένων περιπτώσεων, αθροιστικός αριθμός ανακτημένων περιπτώσεων και αθροιστικός αριθμός περιπτώσεων θανάτου), \mathcal{M} είναι ο τελεστής παρατήρησης που συνδέει τις παρατηρούμενες τιμές με τις τιμές κατάστασης του μοντέλου και το \mathbf{e}_{k+1} είναι ο θόρυβος παρατήρησης που υποτίθεται ότι ακολουθεί τον μηδενικό μέσο Gaussian θόρυβο με τον πίνακα συνδιακύμανσης \mathbf{R}_{k+1} $[3 \times 3]$.

Στο βήμα εκτίμησης (πρόβλεψης), οι μεταβλητές κατάστασης, \mathbf{s}_{k+1} και οι παράμετροι $\boldsymbol{\theta}_{k+1}$, αναπτύσσονται για να σχηματίσουν ένα διάνυσμα της μορφής:

$$\mathbf{x}_{k+1} = [\mathbf{s}_{k+1} \ \boldsymbol{\theta}_{k+1}]^T \quad (4.13)$$

Παρακάτω θα δούμε τον μη γραμμικό τελεστή \mathcal{S} (διακριτό κατά Euler μοντέλο SIRD) σύμφωνα με τον οποίο μεταβάλλεται το διάνυσμα μεταβλητών κατάστασης $\mathbf{s}_k = [S_k \ I_k \ R_k \ D_k]$:

$$S_{k+1} = S_k - \beta_k S_k I_k / N \quad (4.14)$$

$$I_{k+1} = I_k + \beta_k S_k I_k / N - \gamma_k I_k - \delta_k I_k \quad (4.15)$$

$$R_{k+1} = R_k + \gamma_k I_k \quad (4.16)$$

$$D_{k+1} = D_k + \delta_k I_k \quad (4.17)$$

Σημαντική υπενθύμιση: ο ρυθμός μεταβολής μετάδοσης, ανάρρωσης και θανάτων εξαρτάται από το t_{lockdown} καθώς οι εξισώσεις που περιγράφουν τις παραπάνω μεταβλητές διαφέρουν για το χρονικό διάστημα πριν και μετά από τον περιορισμό της κυκλοφορίας (καραντίνα).

Ο $i+1$ όρος πρόβλεψης του \mathbf{x}_{k+1} στη συνέχεια ενημερώνεται χρησιμοποιώντας τη διαφορά μεταξύ των “διαταραγμένων” από τον θόρυβο παρατηρήσεων και μετρήσεων σύμφωνα με την παρακάτω σχέση:

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{K}_{k+1} [\mathbf{y}_{k+1} - \mathbf{C}_k \hat{\mathbf{x}}_k] \quad (4.18)$$

Το Κέρδος Kalman υπολογίζεται από την σχέση:

$$\mathbf{K}_k = \mathbf{A}_k \mathbf{P}_k^- \mathbf{C}_k^T [\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}]^{-1} \quad [13 \times 3] \quad (4.19)$$

Σε κάθε επανάληψη στο Εκτεταμένο φίλτρο Kalman , κάθε μέλος του \mathbf{x}_{k+1} , δηλαδή για $i = 1, \dots, n$, ενημερώνεται αφομοιώνοντας διαταραγμένες παρατηρήσεις χρησιμοποιώντας την εξίσωση $\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{K}_{k+1} [\mathbf{y}_{k+1} - \hat{\mathbf{x}}_k]$. Στο εξής, μία επανάληψη ΕκΚΦ αντιστοιχεί σε έναν κύκλο αφομοίωσης. Η διαδικασία επαναλαμβάνεται με το ενημερωμένο σύνολο μέχρι να ικανοποιηθεί ένα κριτήριο διακοπής που ορίζεται από τον χρήστη, π.χ. διακοπή κριτηρίων που βασίζονται στον μέγιστο αριθμό επαναλήψεων ή ορίζοντας ένα όριο αλλαγής των τιμών των παραμέτρων μεταξύ δύο διαδοχικών επαναλήψεων ΕκΚΦ.

4.3 Εκτιμήσεις παραμέτρων με συνθετικά δεδομένα

Στην πρώτη περίπτωση δοκιμής, η απόδοση του Εκτεταμένου Φίλτρου Kalman αξιολογήθηκε χρησιμοποιώντας προσομοιωμένα δεδομένα με συνθετικές παρατηρήσεις. Ο παρακάτω πίνακας μας δείχνει τις παραμέτρους του μοντέλου, που χρησιμοποιούνται για τη δημιουργία των συνθετικών δεδομένων (παρατηρήσεις).

Οι τιμές των παραμέτρων του μοντέλου στον πίνακα αναφέρονται ως τιμές “στόχοι”. Το σύστημα διαφορικών εξισώσεων λύθηκε αριθμητικά για, $0 \leq k \leq 100$, με αρχικές τιμές $I_0 = 350$, $R_0 = 1$, $D_0 = 7$ και $S_0 = N - 350 - 7 - 1$, χρησιμοποιώντας τον λύτη ode45 του MATLAB (έκδοση R2016a). Το μέγεθος του πληθυσμού N και η ημέρα που ορίστηκε ο περιορισμός (καραντίνα) λήφθηκαν ως $N=60M$ και $t_{lockdown}=17$, αντίστοιχα. Οι “συνθετικές” παρατηρήσεις στη συνέχεια, καταγράφηκαν με εξαγωγή των τιμών κατάστασης I_k , R_k και D_k που αντιπροσωπεύουν τον αριθμό των μολυσμένων περιπτώσεων (Infected), τον αθροιστικό αριθμό των ανακτημένων περιπτώσεων (Recovered) και τον αθροιστικό αριθμό περιπτώσεων θανάτου (Died).

Για τις αρχικές τιμές με $k=0$ στην εκτίμηση των παραμέτρων ακολουθήσαμε το τρόπο με τον οποίο αντλήθηκαν οι τιμές στην προηγούμενη έρευνα ώστε να υπάρχει αντιστοιχία. Οι αρχικές εκτιμήσεις δημιουργήθηκαν τυχαία από ομοιόμορφη κατανομή (Uniform Distribution):

$$\begin{array}{lll} \beta_0 \sim \mathcal{U}(0.2, 0.6) & \beta_1 \sim \mathcal{U}(0.05, 0.15) & \tau_\beta \sim \mathcal{U}(10, 30) \\ \gamma_0 \sim \mathcal{U}(0.015, 0.045) & \gamma_1 \sim \mathcal{U}(0.02, 0.06) & \tau_\gamma \sim \mathcal{U}(12.5, 37.5) \\ \delta_0 \sim \mathcal{U}(0.005, 0.015) & \delta_1 \sim \mathcal{U}(0.01, 0.03) & \tau_\delta \sim \mathcal{U}(11, 33) \end{array}$$

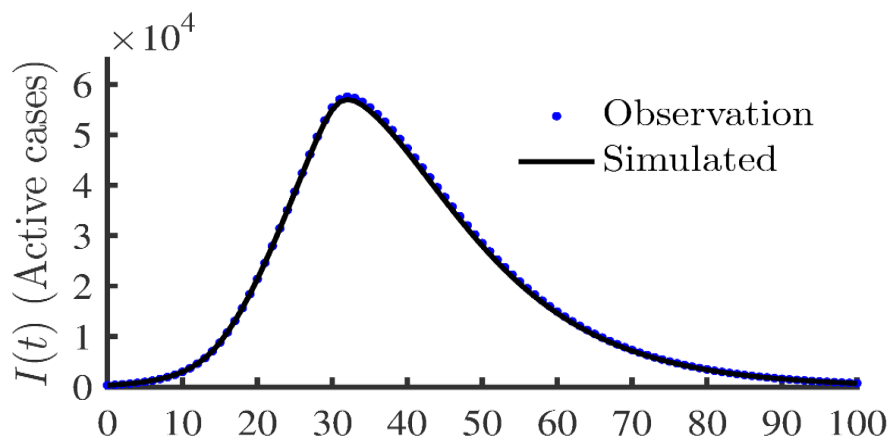
Το χρονικό διάστημα μεταξύ δύο βημάτων αφομοίωσης ExKF λήφθηκε ως $dt = dk = 1$ ημέρα. Ως εκ τούτου, οι παρατηρήσεις $y_{k,t}$, δηλαδή οι τιμές των I_k , R_k και D_k , θεωρήθηκαν γνωστές στο $k = 0, 1, \dots, 100$.

Στο βήμα αφομοίωσης (ανανέωσης), δημιουργούνται διαταραγμένες παρατηρήσεις χρησιμοποιώντας το διάνυσμα \hat{y}_k . Επομένως, οι διαταραγμένες παρατηρήσεις είναι $\hat{y}_k = \hat{y}_k + e_k$. Ο πρόσθετος θόρυβος $e_k \sim \mathcal{N}(0, \mathbf{R}_k)$ με $\mathbf{R}_k = \text{diag}[\sigma^2_I, \sigma^2_R, \sigma^2_D]$ [3x3], όπου τα σ_I , σ_R και σ_D είναι τα σφάλματα παρατήρησης που λαμβάνονται ως 10% των παρατηρούμενων τιμών I, R, D στην επανάληψη $t=k$.

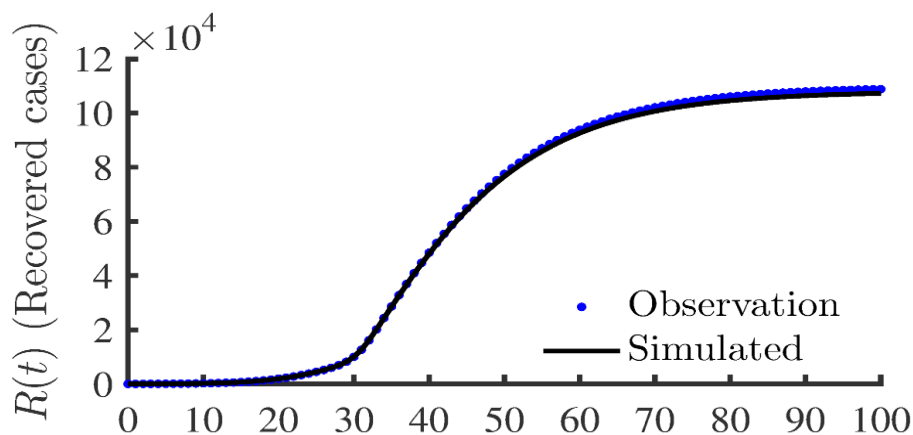
4.3.1 Αποτελέσματα

Θεωρούμε ως τιμές αναφοράς (ή τιμές “στόχους”), τα αποτελέσματα που πήραμε από την επίλυση του προβλήματος χρησιμοποιώντας το MATLAB. Στη συνέχεια θα κάνουμε σύγκριση των αποτελεσμάτων που πήραμε χρησιμοποιώντας το Εκτεταμένο Φίλτρο Kalman και της προηγούμενης έρευνας που έγινε για την εκτίμηση των μεταβλητών χρησιμοποιώντας το Συνολικό Φίλτρο Kalman.

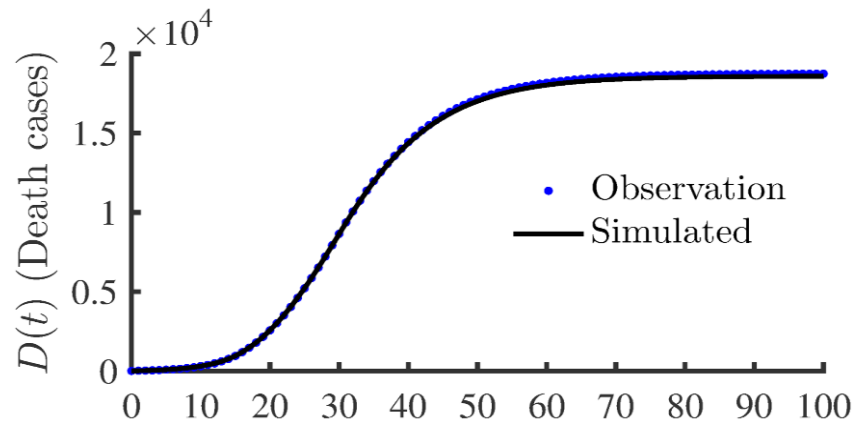
Αρχικά θα δούμε τα αποτελέσματα που προέκυψαν από το Συνολικό Φίλτρο Kalman (EnKF) στη προηγούμενη έρευνα τις βασικές ποσότητες που μας ενδιαφέρουν, δηλαδή τους Infected-Recovered-Died και στη συνέχεια τις εκτιμήσεις των παραμέτρων που προέκυψαν.



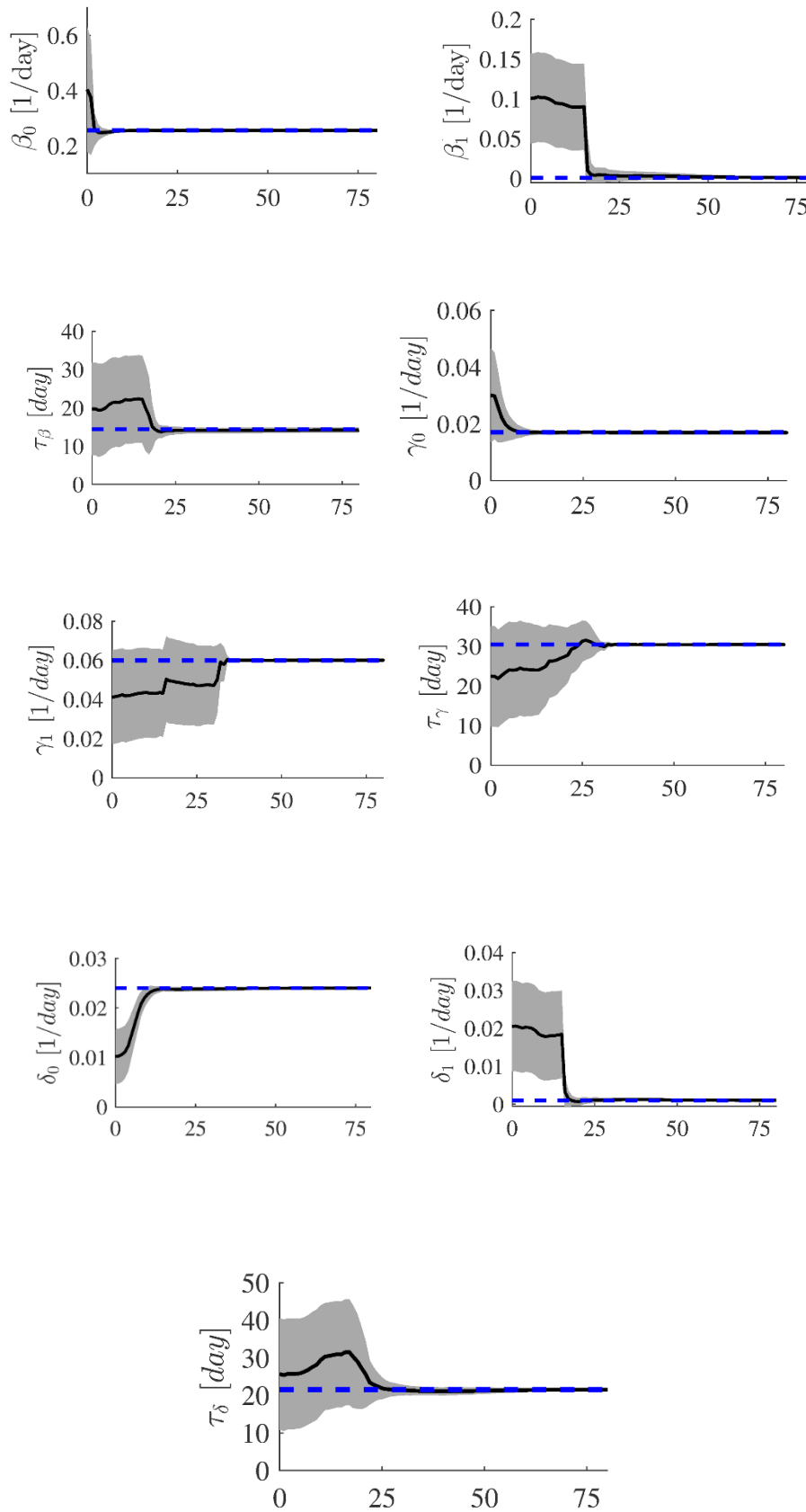
Σχήμα 4.1 Καμπύλη μολυσμένων.



Σχήμα 4.2 Καμπύλη αναρροσάντων.



Σχήμα 4.2 Καμπύλη νεκρών.



Σχήμα 4.3 Εκτιμήσεις παραμέτρων με EnKF.

| Παράμετρος | Τιμή | Μονάδα μέτρησης |
|---------------|-------|-----------------|
| β_0 | 0,256 | 1/ημέρα |
| β_1 | 0,001 | 1/ημέρα |
| τ_β | 14,39 | ημέρες |
| γ_0 | 0,017 | 1/ημέρα |
| γ_1 | 0,06 | 1/ημέρα |
| τ_γ | 30,5 | ημέρες |
| δ_0 | 0,024 | 1/ημέρα |
| δ_1 | 0,001 | 1/ημέρα |
| τ_δ | 21,6 | ημέρες |

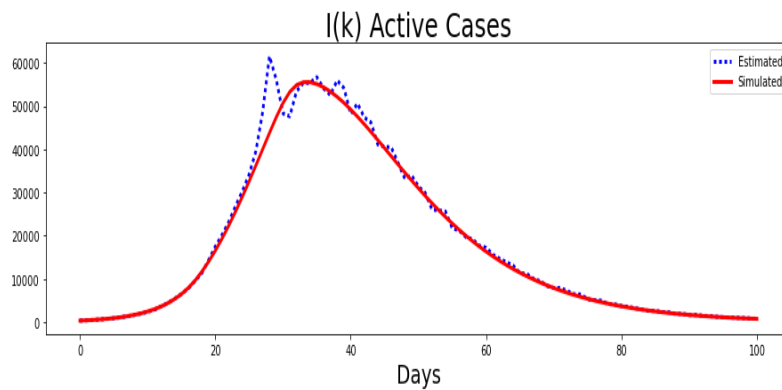
Πίνακας 4.1: Τιμές στις οποίες συγκλίνουν οι παράμετροι εφαρμόζοντας ΕπΚΦ με τη χρήση του Matlab (ode45).

Παρακάτω βλέπουμε το μέσο σχετικό απόλυτο σφάλμα (RMAE) των εκτιμήσεων που προέκυψαν από την έρευνα με την χρήση του Συνολικού Φίλτρου Kalman για την καθεμιά εκ των τριών πληθυσμιακών ομάδων που παρατηρούμε.

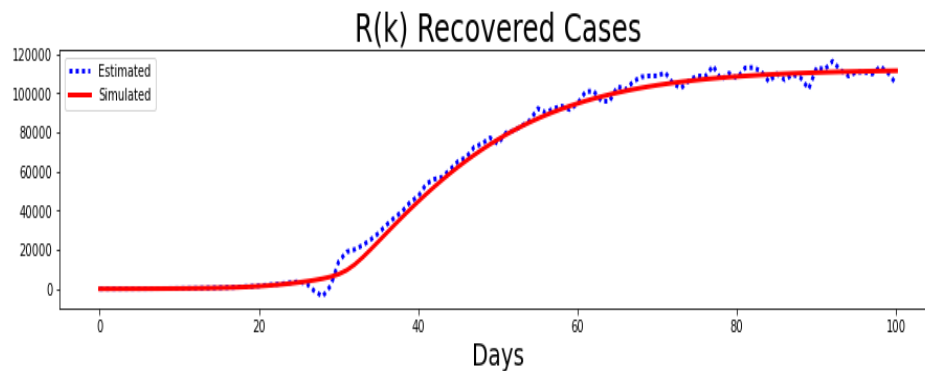
| | Active Cases (I) | Recovered Cases (R) | Death Cases (D) |
|----------------|------------------|---------------------|-----------------|
| RMAE(%) | 1.44 | 0.96 | 0.54 |

Πίνακας 4.2: Σχετικό Μέσο Απόλυτο Σφάλμα για τις πληθυσμιακές ομάδες I,R,D με ΕπΚΦ.

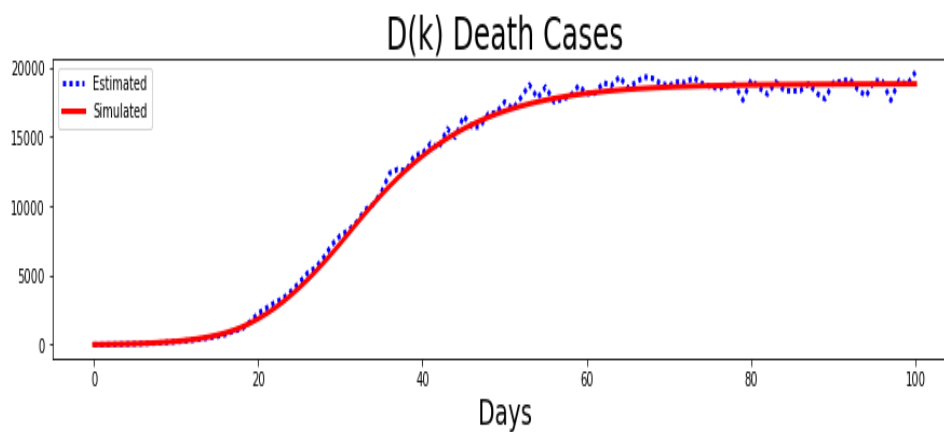
Στη συνέχεια θα δούμε την προσομοίωση που πραγματοποιήθηκε στα πλαίσια της παρούσας έρευνας εφαρμόζοντας το Εκτεταμένο Φίλτρο Kalman στο Διακριτοποιημένο κατά Euler επιδημιολογικό μοντέλο SIRD με την χρήση κώδικα σε γλώσσα προγραμματισμού Python.



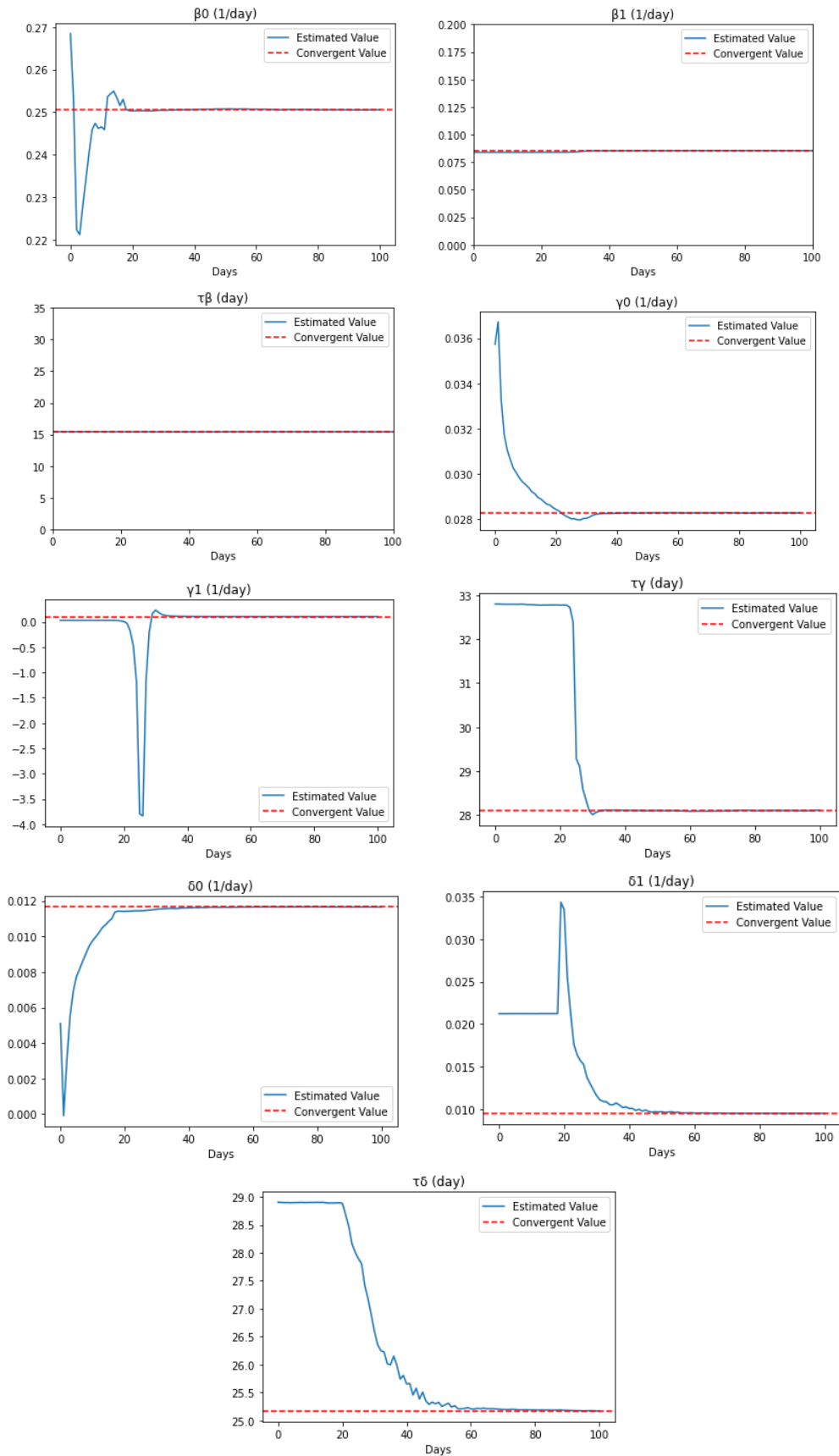
Σχήμα 4.4 Καμπύλη μολυσμένων.



Σχήμα 4.5 Καμπύλη αναρρωσάντων.



Σχήμα 4.6 Καμπύλη νεκρών.



Σχήμα 4.7 Εκτιμήσεις παραμέτρων με ExKF.

| Παράμετρος | Τιμή | Μονάδα μέτρησης |
|---------------|---------|-----------------|
| β_0 | 0.2505 | 1/ημέρα |
| β_1 | 0.0854 | 1/ημέρα |
| τ_β | 15.40 | ημέρες |
| γ_0 | 0.0283 | 1/ημέρα |
| γ_1 | 0.1033 | 1/ημέρα |
| τ_γ | 28.11 | ημέρες |
| δ_0 | 0.0117 | 1/ημέρα |
| δ_1 | 0.0096 | 1/ημέρα |
| τ_δ | 25.1683 | ημέρες |

Πίνακας 4.3: Τιμές στις οποίες συγκλίνουν οι παράμετροι εφαρμόζοντας ΕκΚΦ με τη χρήση της Python.

Τέλος βλέπουμε το μέσο σχετικό απόλυτο σφάλμα (RMAE) των εκτιμήσεων που προέκυψαν από την έρευνα με την χρήση του Εκτεταμένου Φίλτρου Kalman για την καθημέρα εκ των τριών πληθυσμιακών ομάδων που παρατηρούμε.

| | Active Cases (I) | Recovered Cases (R) | Death Cases (D) |
|----------------|------------------|---------------------|-----------------|
| RMAE(%) | 5.97 | 1.59 | 1.68 |

Πίνακας 4.4: Σχετικό Μέσο Απόλυτο Σφάλμα για τις πληθυσμιακές ομάδες I,R,D με ΕκΚΦ.

4.4 Εκτιμήσεις παραμέτρων με πραγματικά δεδομένα

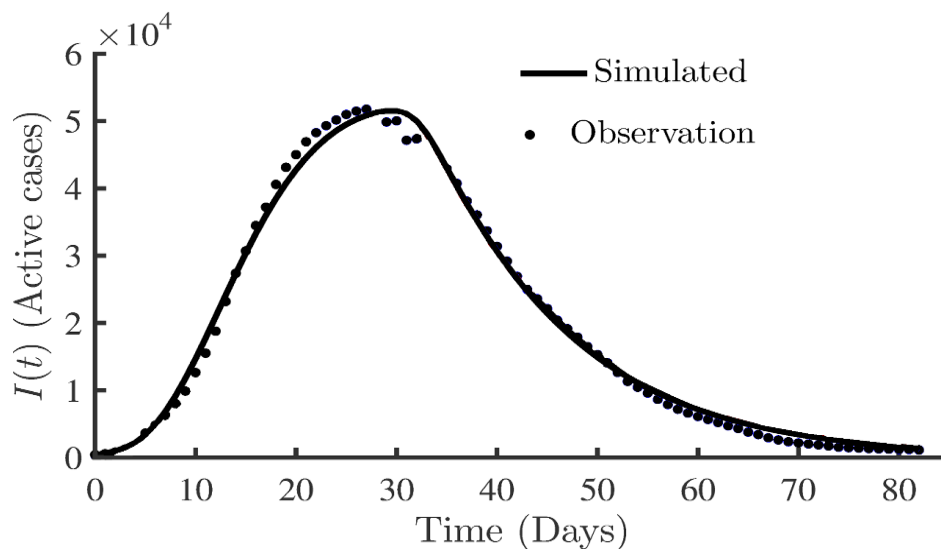
Στη δεύτερη περίπτωση δοκιμής, οι παράμετροι του μοντέλου με τις σχετικές αβεβαιότητες εκτιμήθηκαν χρησιμοποιώντας τα αναφερόμενα δεδομένα COVID-19 της επαρχίας Χουμπέι της Κίνας. Όπως και στη συνθετική περίπτωση, εννέα παράμετροι που περιγράφουν τις τρεις χρονικά μεταβαλλόμενες παραμέτρους του μοντέλου, $(\beta(k), \gamma(k)$ και $\delta(k))$, εκτιμήθηκαν χρησιμοποιώντας τα παρατηρούμενα δεδομένα $I(k), R(k)$ και $D(k)$.

Το μέγεθος του πληθυσμού της επαρχίας Χουμπέι λήφθηκε ως $N = 59$ εκατομμύρια (<https://data.stats.gov.cn/english/easyquery.htm?cn=E0103>). Και το σύνολο της αρχικής κατάστασης δημιουργείται όπως στη συνθετική περίπτωση με $s_0 = [S_0, I_0, R_0, D_0]$, όπου $S_0 = N - I_0 - R_0 - D_0$ με $I_0=431, R_0 = 28$ και $D_0 = 17$ από τα αναφερόμενα δεδομένα. Το t_{lockdown} έχει οριστεί σε 6. Παρόμοια με τη συνθετική περίπτωση, το αρχικό σύνολο για τις τιμές των παραμέτρων σχεδιάστηκε τυχαία από μια ομοιόμορφη κατανομή (Uniform Distribution) με το αρχικό εύρος τιμών να παρουσιάζεται παρακάτω.

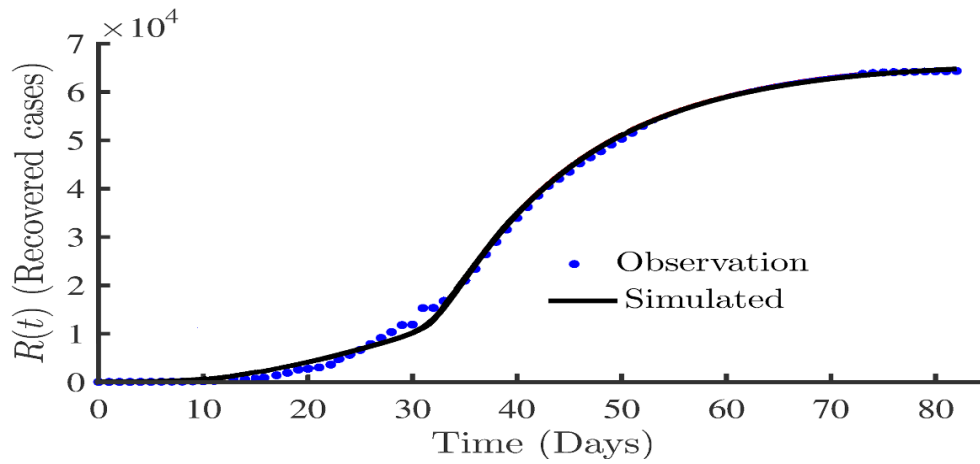
| | | |
|---|---|---------------------------------------|
| $\beta_0 \sim \mathcal{U}(0.1, 0.9)$ | $\beta_1 \sim \mathcal{U}(0.001, 0.002)$ | $\tau_\beta \sim \mathcal{U}(5, 20)$ |
| $\gamma_0 \sim \mathcal{U}(0.001, 0.002)$ | $\gamma_1 \sim \mathcal{U}(0, 0.001)$ | $\tau_\gamma \sim \mathcal{U}(7, 40)$ |
| $\delta_0 \sim \mathcal{U}(0, 0.001)$ | $\delta_1 \sim \mathcal{U}(0.001, 0.002)$ | $\tau_\delta \sim \mathcal{U}(7, 20)$ |

4.4.1 Αποτελέσματα

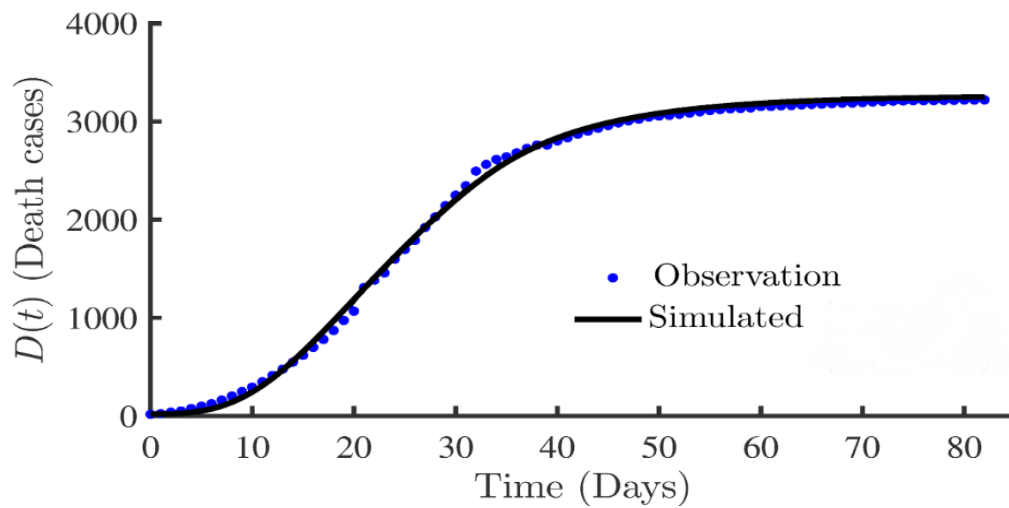
Όπως και στην προηγούμενη ενότητα αρχικά θα δούμε τα αποτελέσματα που προέκυψαν από το Συνολικό Φίλτρο Kalman (EnKF) με τις βασικές ποσότητες που μας ενδιαφέρουν, δηλαδή τους Infected-Recovered-Died και στη συνέχεια τις εκτιμήσεις των παραμέτρων που προέκυψαν.



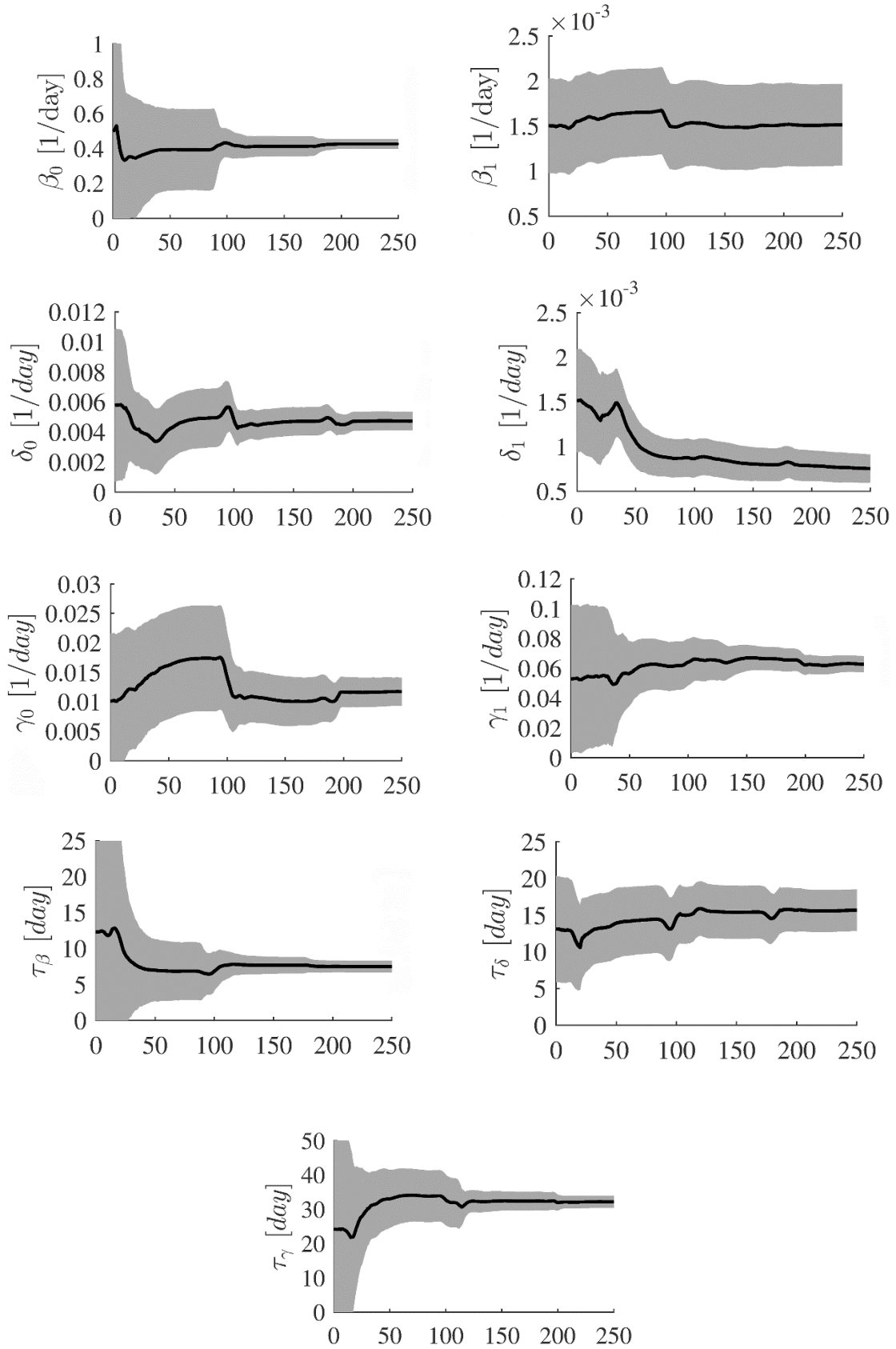
Σχήμα 4.8 Καμπύλη μολυσμένων.



Σχήμα 4.9 Καμπύλη αναρρωσάντων.



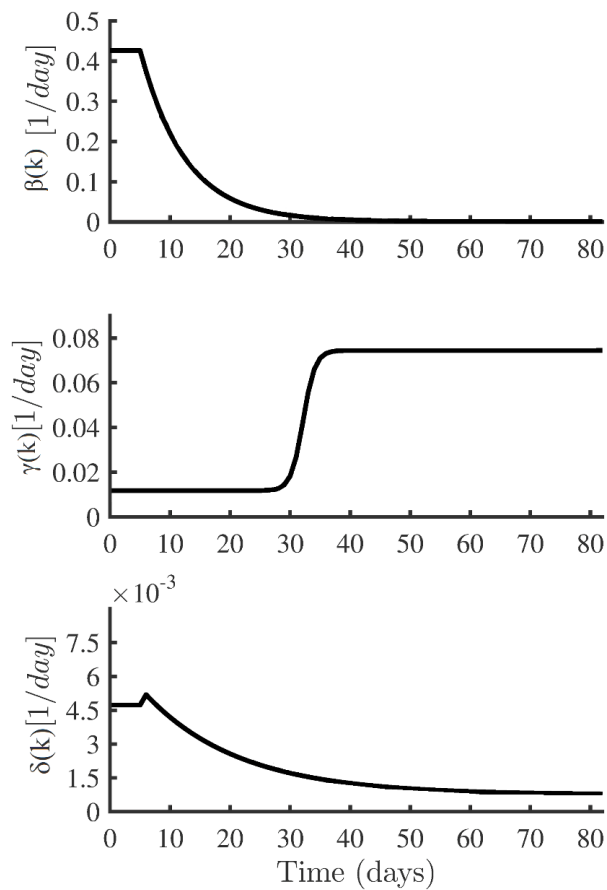
Σχήμα 4.10 Καμπύλη νεκρών.



Σχήμα 4.11 Εκτιμήσεις παραμέτρων (EnKF).

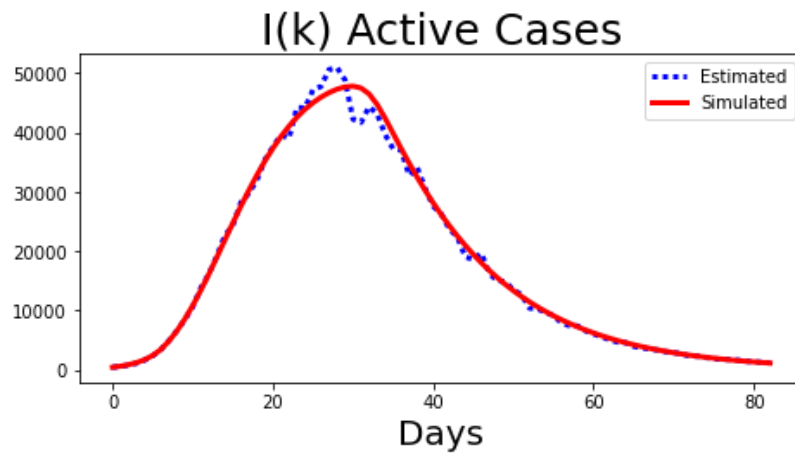
| Παράμετρος | Τιμή | Μονάδα μέτρησης |
|---------------|----------------|-----------------|
| β_0 | 0.4264±0.0133 | 1/ημέρα |
| β_1 | 0.0015±0.0002 | 1/ημέρα |
| τ_β | 7.4757±0.3995 | ημέρες |
| γ_0 | 0.0117±0.0012 | 1/ημέρα |
| γ_1 | 0.0628±0.0023 | 1/ημέρα |
| τ_γ | 31.1524±0.7933 | ημέρες |
| δ_0 | 0.0047±0.0003 | 1/ημέρα |
| δ_1 | 0.0008±0.0001 | 1/ημέρα |
| τ_δ | 16.3735±1.4076 | ημέρες |

Πίνακας 4.5: Τιμές στις οποίες συγκλίνουν οι παράμετροι με το αντίστοιχο σφάλμα τους εφαρμόζοντας το EnKF με τη χρήση του Matlab (ode45).

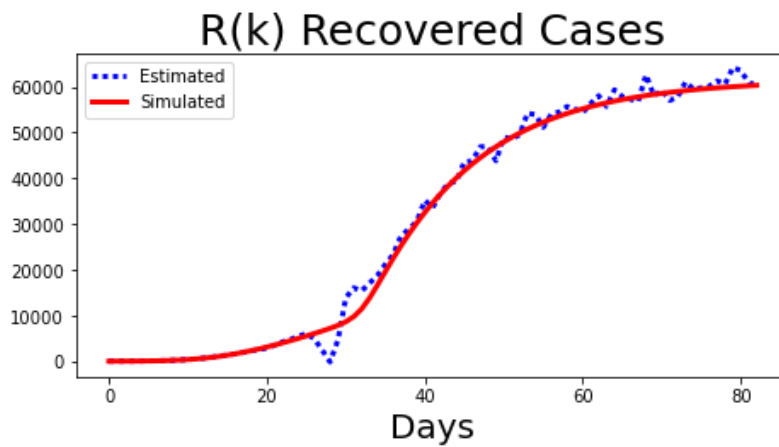


Σχήμα 4.12 Διαμόρφωση ρυθμού μεταβολής μετάδοσης, ανάρρωσης και θνησιμότητας (EnKF).

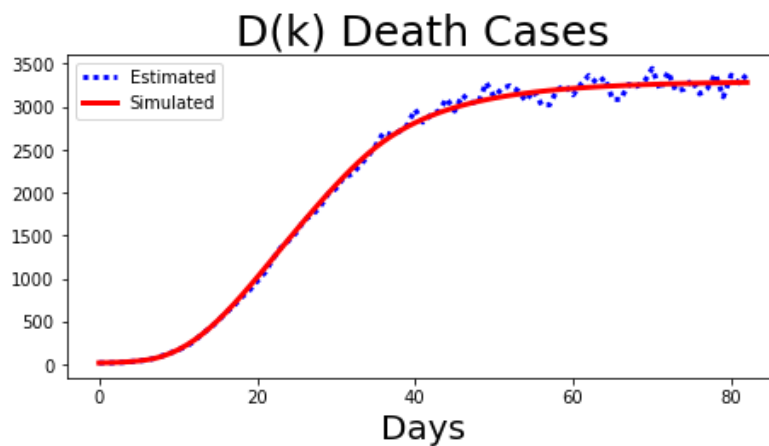
Παρακάτω θα παρουσιάσουμε τα αποτελέσματα που προέκυψαν εφαρμόζοντας το Εκτεταμένο Φίλτρο Kalman με τα πραγματικά δεδομένα για την επαρχία Χουμπί.



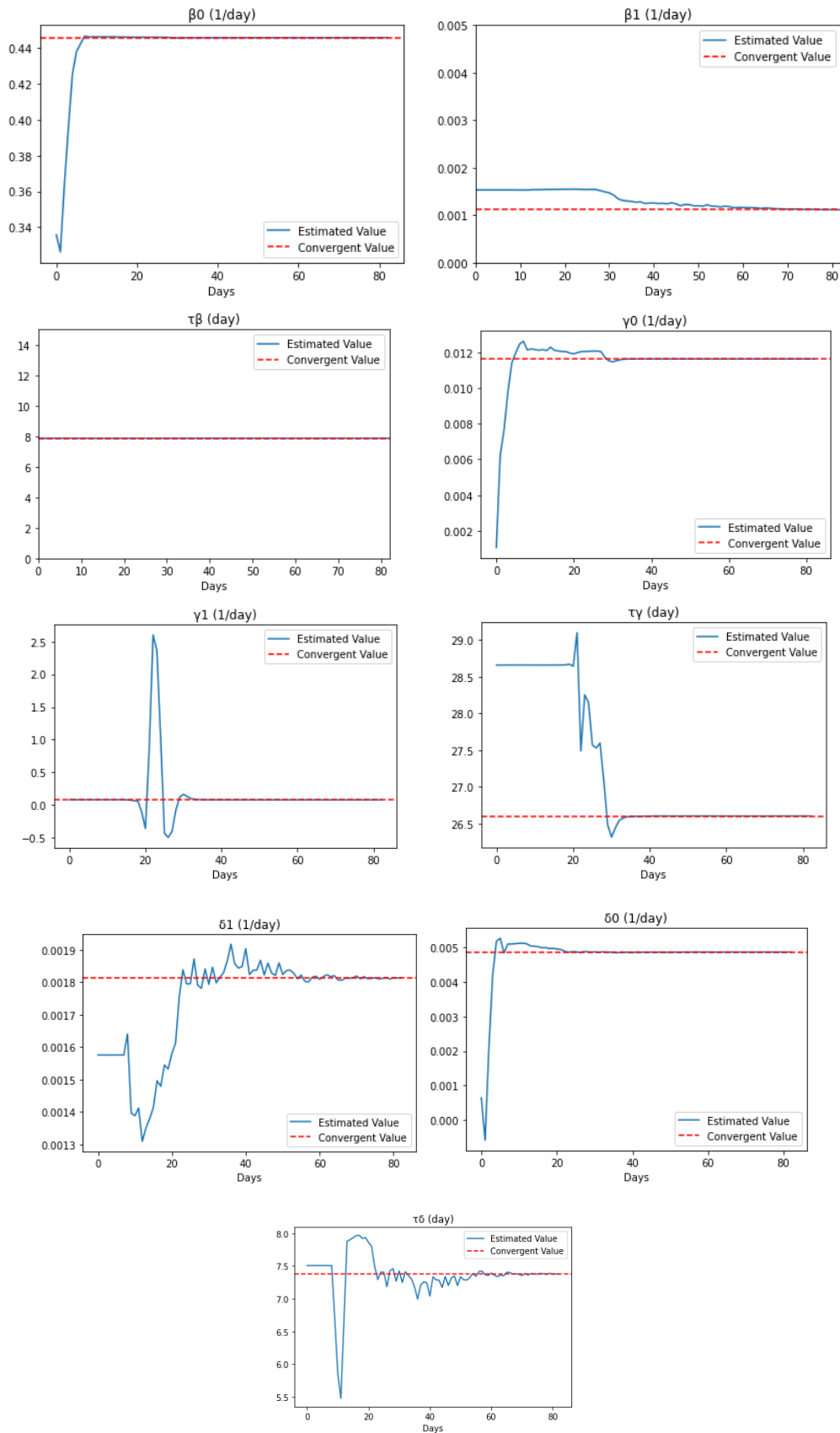
Σχήμα 4.13 Καμπύλη μολυσμένων.



Σχήμα 4.14 Καμπύλη αναρρωσάντων.



Σχήμα 4.15 Καμπύλη νεκρών.



Σχήμα 4.16 Εκτιμήσεις παραμέτρων (ExKF).

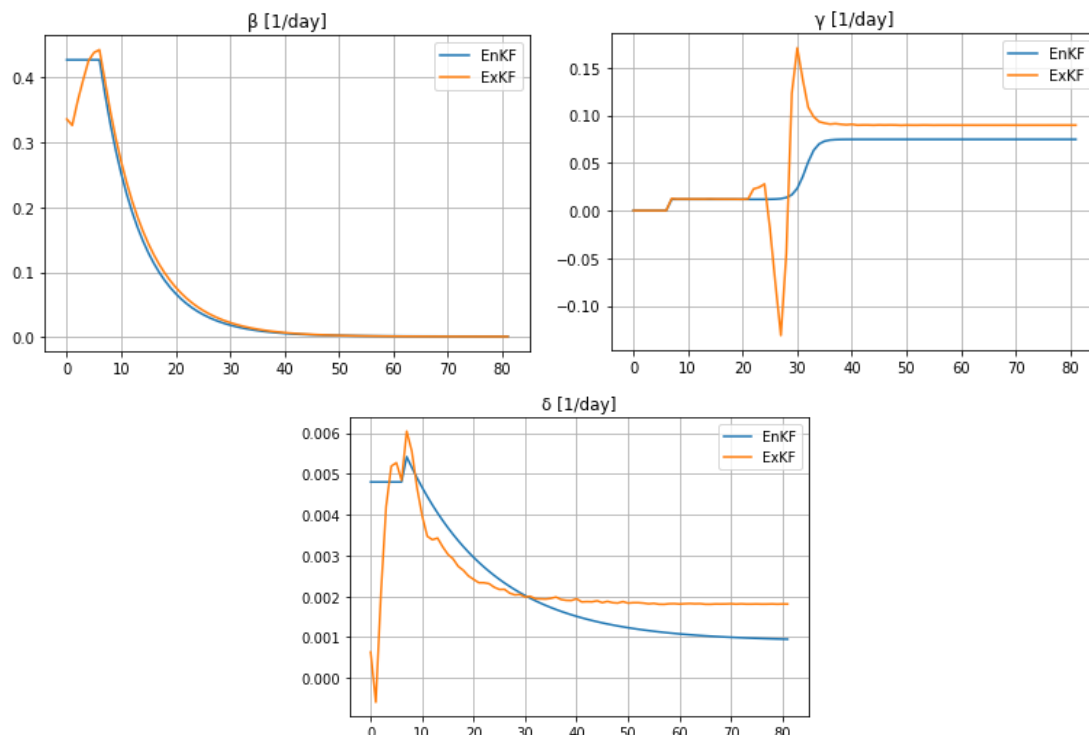
| Παράμετρος | Τιμή | Μονάδα μέτρησης |
|---------------|----------|-----------------|
| β_0 | 0.4467 | 1/ημέρα |
| β_1 | 0.0011 | 1/ημέρα |
| τ_β | 7.8714 | ημέρες |
| γ_0 | 0.0116 | 1/ημέρα |
| γ_1 | 0.0780 | 1/ημέρα |
| τ_γ | 26.6283 | ημέρες |
| δ_0 | 0.004873 | 1/ημέρα |
| δ_1 | 0.001815 | 1/ημέρα |
| τ_δ | 7.3712 | ημέρες |

Πίνακας 4.6: Τιμές στις οποίες συγκλίνουν οι παράμετροι εφαρμόζοντας ExKF με τη χρήση της Python.

Στη συνέχεια βλέπουμε το μέσο σχετικό απόλυτο σφάλμα (RMAE) των εκτιμήσεων που προέκυψαν από την έρευνα με την χρήση του Εκτεταμένου Φίλτρου Kalman για την καθεμιά εκ των τριών πληθυσμιακών ομάδων που παρατηρούμε. Τέλος θα δούμε πως διαμορφώνονται οι ρυθμοί μεταβολής μετάδοσης, ανάρρωσης και θνησιμότητας σε σχέση με το Συνολικό Φίλτρο.

| | Active Cases (I) | Recovered Cases (R) | Death Cases (D) |
|----------------|------------------|---------------------|-----------------|
| RMAE(%) | 3.508 | 8.043 | 2.660 |

Πίνακας 4.7: Σχετικό Μέσο Απόλυτο Σφάλμα για τις πληθυσμιακές ομάδες I,R,D με ExKF.



Σχήμα 4.17 Διαμόρφωση ρυθμού μεταβολής μετάδοσης, ανάρρωσης και θνησιμότητας (ExKF).

4.5 Συμπεράσματα-Σύγκριση Αποτελεσμάτων

Στην ενότητα αυτή θα μιλήσουμε γενικά για τα αποτελέσματα που προέκυψαν από τις δύο εφαρμογές που έγιναν χρησιμοποιώντας τα συνθετικά δεδομένα και τον τρόπο με τον οποίο προέκυψαν τα αποτελέσματα. Στη συνέχεια θα σχολιάσουμε εκτενέστερα τα αποτελέσματα που προέκυψαν από την δεύτερη δοκιμή στην οποία χρησιμοποιήθηκαν τα πραγματικά δεδομένα.

Με μια πρώτη ματιά παρατηρούμε και στις δύο μεθόδους τις καμπύλες προσομοίωσης (Simulated) να έχουν την ίδια μορφή παρουσιάζοντας όμοιες μέγιστες τιμές στις αντίστοιχες ημέρες. Η διαφορά στις δύο έρευνες είναι ο τρόπος με τον οποίο προέκυψαν τα διαγράμματα. Στην πρώτη περίπτωση για την επίλυση του μη γραμμικού συστήματος διαφορικών εξισώσεων (οι εξισώσεις του μοντέλου S-I-R-D) χρησιμοποιήθηκε το λογισμικό Matlab με την ενσωματωμένη συνάρτηση ode45, ενώ στην παρούσα έρευνα, για την επίλυση του μη γραμμικού μοντέλου κάναμε διακριτοποίηση κατά Euler. Η μέθοδος Euler είναι μια μέθοδος πρώτης τάξης, που σημαίνει ότι το τοπικό σφάλμα (σφάλμα ανά βήμα) είναι ανάλογο με το τετράγωνο του μεγέθους του βήματος και το συνολικό σφάλμα (σφάλμα σε μια δεδομένη στιγμή) είναι ανάλογο με το μέγεθος του βήματος.

Στα Σχήματα 4.4-4.7 όπου παρουσιάζονται τα διαγράμματα με την εφαρμογή του Εκτεταμένου Φίλτρου παρατηρούμε ότι υπάρχουν κάποιες “διαταραχές” (σφάλμα) στις εκτιμήσεις. Αυτό το σφάλμα οφείλεται τόσο στην μέθοδο με την οποία έγινε η διακριτοποίηση όπως αναφέραμε παραπάνω, όσο και στον τρόπο με τον οποίο λειτουργεί το Εκτεταμένο Φίλτρο Kalman. Υπενθυμίζουμε ότι στο Εκτεταμένο Φίλτρο Kalman ο πίνακας μετάβασης A_k (3.8) υπολογίζεται από την Jacobian της $f[k]$ και αντίστοιχα για τον πίνακα μετρήσεων C_k (3.9), υπολογίζεται από την Jacobian της $g[k]$. Οι γραμμικοποιήσεις που γίνονται αποφέρουν ένα επιπλέον σφάλμα το οποίο είναι αναμενόμενο.

Όσον αφορά τις εκτιμήσεις των χρονομεταβαλλόμενων παραμέτρων παρατηρούμε πως και στις δύο μεθόδους η σύγκλιση γίνεται σχετικά γρήγορα στις τελικές τιμές τους. Όσον αφορά το Σχήμα 4.3 του Συνολικού Φίλτρου, οι γκριζες σκιές που φαίνονται είναι το σύνολο των εκτιμήσεων που γίνονται ανά επανάληψη (ημέρα), ενώ η μαύρη γραμμή είναι η μέση τιμή των εκτιμήσεων αυτών.

Τέλος στους Πίνακες 4.2 και 4.4 φαίνονται τα Σχετικά Μέσα Απόλυτα Σφάλματα (RMAE) των τριών πληθυσμιακών ομάδων που παρατηρούμε. Όσον αφορά το Εκτεταμένο Φίλτρο παρατηρούμε ότι τα σφάλματα είναι μεγαλύτερα από τα αντίστοιχα του Συνολικού Φίλτρου. Παραπάνω αναφέραμε ότι στην έρευνά μας αναμέναμε να υπάρχει ένα σφάλμα ωστόσο το θεωρούμε αποδεκτό μιας και αυτό δεν είναι υψηλό. Έτσι θεωρούμε έγκυρα τα αποτελέσματα και το Εκτεταμένο Φίλτρο Kalman κατάλληλο για να προχωρήσουμε και στην έρευνα με τα πραγματικά δεδομένα για την επαρχία της Χουμπέι.

Στην δεύτερη δοκιμή που έγινε χρησιμοποιήθηκαν τα πραγματικά δεδομένα για την επαρχία της Κίνας, Χουμπέι, όπου οι αρχικές συνθήκες του προβλήματος ήταν γνωστές. Σε αυτή την περίπτωση η προσομοίωση έγινε για 82 ημέρες, από τις 22 Ιανουαρίου 2020 ($k=0$) έως 13 Απριλίου 2020 ($k=82$), όπως φαίνεται και στα διαγράμματα.

Τις σημαντικότερες πληροφορίες που πήραμε από αυτή την έρευνα τις βλέπει κανείς στο Σχήμα 4.17. Βλέποντας στα σχήματα το διάγραμμα ρυθμού μετάδοσης, ο ρυθμός ξεκινά από μια αρχική τιμή $\beta=\beta_0\approx 0.4478$. Εφαρμόζοντας τον περιορισμό κυκλοφορίας και των περιοριστικών μέτρων ο ρυθμός μετάδοσης μειώνεται σε $\beta=\beta_1\approx 0.0011$. Αυτή η μείωση είναι πολύ σημαντική, αν σκεφτεί κανείς ότι αρχικά ένας στους δύο μολυσμένους θα μετέδιδε τον ιό σε κάποιο υγιή άτομο. Ενώ μετά τους περιορισμούς μόνο ένας στους χίλιους μολυσμένους θα μεταδώσει τον ιό. Τέλος, το τ_β αντιπροσωπεύει το χρόνο καθίζησης και έχει υποτεθεί ότι σε $k = 3\tau_\beta$, ο ρυθμός μετάδοσης μειώνεται στο 90% της αρχικής του τιμής. Ο χαρακτηριστικός χρόνος καθίζησης στην έρευνα μας εκτιμήθηκε $\tau_\beta\approx 7.87$ που σημαίνει ότι σε 25 ημέρες περίπου από την εφαρμογή των περιορισμών αναμένουμε να δούμε τη σημαντική μείωση στον ρυθμό μετάδοσης. Πράγματι, από το Σχήμα 4.8 στην καμπύλη παρατήρησης (Observed) φαίνεται η μέγιστη τιμή να σημειώνεται περίπου την 30^η ημέρα ενώ στην συνέχεια ο αριθμός των μολυσμένων να μειώνεται καθημερινά. Αντίστοιχη συμπεριφορά στην καμπύλη των μολυσμένων παρατηρούμε και στο Σχήμα 4.13 της δικής μας έρευνας.

Όσον αφορά τον ρυθμό ανάρρωσης γ , αυτός ξεκινάει με $\gamma=\gamma_0\approx 0.0116$ και αυξάνεται έως $\gamma=\gamma_0+\gamma_1\approx 0.0896$ το οποίο αντιστοιχεί σε περίπου 12 ημέρες ($1/\gamma$) για τη θεραπεία ενός ασθενή. Ο χρόνος για την θεραπεία ενός ασθενή στην πραγματικότητα εξαρτάται από τα νέα φάρμακα τα οποία κατασκευάζονται και τις νέες μεθόδους θεραπείας που εφαρμόζονται. Αξίζει επίσης να υπενθυμίσουμε ότι ο ρυθμός ανάρρωσης (Εξ. 4.7) δεν εξαρτάται από την ημερομηνία εφαρμογής της καραντίνας. Από τον χαρακτηριστικό χρόνο αύξησης τ_γ συμπεραίνουμε ότι οι νέες μέθοδοι θεραπείας ή η παρασκευή νέων φαρμάκων ξεκίνησαν να εφαρμόζονται την 27^η ημέρα της πανδημίας με τα οφέλη αυτών να εμφανίζονται εντός πέντε ημερών. Με τον τρόπο αυτό ο ρυθμός ανάρρωσης αυξήθηκε, δηλαδή οι μέρες που χρειάζεται ένας ασθενής για να θεραπευτεί μειώθηκαν. Αντίστοιχα από την έρευνα όπου χρησιμοποιήθηκε το Συνολικό Φίλτρο προέκυψε ότι χρειάζονται περίπου 13 ημέρες ανάρρωσης, ενώ από τις επίσημες πληροφορίες του Παγκόσμιου Οργανισμού Υγείας (WHO) ο εκτιμώμενος χρόνος ανάρρωσης από τον COVID-19 είναι περίπου 2 εβδομάδες.

Τέλος, ο ρυθμός θνησιμότητας είναι αρχικά υψηλός $\delta=\delta_0\approx 0.004873$, λόγω έλλειψης ενημέρωσης για τη νόσο και ως εκ τούτου, ανιχνεύονται μόνο οι σοβαρές περιπτώσεις. Ωστόσο το ποσοστό θνησιμότητας μειώνεται σταδιακά και τελικά καταλήγει σε ένα μακροπρόθεσμο ποσοστό θνησιμότητας. Επιπλέον, με την εισαγωγή διαφόρων μη φαρμακευτικών παρεμβάσεων (όπως η κοινωνική απόσταση και το lockdown), ο αριθμός των νέων μολυσμένων μειώνεται δραστικά, γεγονός που με τη σειρά του μειώνει το ποσοστό θνησιμότητας σε $\delta=\delta_1\approx 0.001815$. Αντίστοιχα με τον χαρακτηριστικό χρόνο καθίζησης στον ρυθμό μετάδοσης, έτσι και εδώ, μετά από $k=3\tau_\delta$ ο ρυθμός θνησιμότητας θα έχει μειωθεί κατά 90% από την αρχική τιμή του.

ΚΕΦΑΛΑΙΟ 5: Σύνοψη-Συζήτηση

Σε αυτήν την διπλωματική εργασία μελετήθηκε η αξία των μακροπρόθεσμων προβλέψεων για την έξαρση της νόσου COVID-19. Συγκεκριμένα για την άντληση πληροφοριών, χρησιμοποιήθηκε το επιδημιολογικό μοντέλο SIRD που περιγράφει ικανοποιητικά μια πανδημία όπως αυτή που εξετάζουμε. Το μοντέλο αυτό αποτελείται από τις τέσσερις βασικές πληθυσμιακές ομάδες, καθώς και από τις χρονικά μεταβαλλόμενες παραμέτρους που συνδέουν τις ομάδες αυτές. Τόσο για την εκτίμηση των παραμέτρων, όσο και για την μακροπρόθεσμη εκτίμηση των μετρήσεων των πληθυσμιακών ομάδων χρησιμοποιήθηκε το Εκτεταμένο Φίλτρο Κάλμαν. Το Εκτεταμένο Φίλτρο Κάλμαν είναι ένας εκτιμητής που χρησιμοποιείται για μη γραμμικά προβλήματα, όπου καλείται κανείς να υπολογίσει τη στιγμιαία “κατάσταση” ενός δυναμικού συστήματος που διαταράσσεται από “λευκό θόρυβο”. Δηλαδή από μετρήσεις που σχετίζονται μη γραμμικά με τη δυναμική κατάσταση του συστήματος αλλά είναι αλλοιωμένες από “λευκό θόρυβο”. Η συγκεκριμένη εργασία βασίστηκε στη δομή και τα δεδομένα προηγούμενης έρευνας που δημοσιεύτηκε στο ηλεκτρονικό περιοδικό PLOS ONE όπου εκεί χρησιμοποιήθηκαν διαφορετικές μέθοδοι επίλυσης του προβλήματος. Τα αποτελέσματα τέτοιου είδους ερευνών βρίσκουν χρησιμότητα σε κρατικούς μηχανισμούς οι οποίοι καλούνται να πάρουν αποφάσεις ζωτικής σημασίας για την αποκλιμάκωση φαινομένων όπως αυτό της νόσου COVID-19.

Για παράδειγμα ο ρυθμός αναπαραγωγής, \mathcal{R}_0 είναι μια πολύ κρίσιμη παράμετρος σε κάθε επιδημία. Τα περισσότερα από τα στατιστικά μοντέλα του COVID-19 βασίζονται στο \mathcal{R}_0 . Βασικά ο ρυθμός αυτός εξηγεί τον αριθμό των κρουσμάτων που μπορεί να δημιουργήσει άμεσα κάθε μολυσμένο κρούσμα, λαμβάνοντας υπόψη όλα τα άτομα που είναι ευάλωτα στη μόλυνση. Είναι απλώς ένα όριο που χρησιμοποιείται για να προσδιορίσει τη φύση οποιασδήποτε ασθένειας εάν αυτή είναι επιδημική ($\mathcal{R}_0 > 1$) ή όχι ($\mathcal{R}_0 < 1$). Γενικά, όσο υψηλότερη είναι η τιμή του \mathcal{R}_0 , τόσο πιο δύσκολος είναι ο έλεγχος της επιδημίας. Από την Εξίσωση 2.22 που δίνει τον ρυθμό αναπαραγωγής για το μοντέλο SIRD ο αρχικός ρυθμός αναπαραγωγής για την επαρχία της Χουμπέι είναι $\mathcal{R}_0 \approx 27$ μια τιμή που προτρέπει τους ειδικούς σε μια δύσκολως ελεγχόμενη πανδημία που χρήζει άμεσα έλεγχο και μέτρα αποκλιμάκωσης. Έτσι, απλά θέτοντας τον περιορισμό κυκλοφορίας την 6^η μέρα της έξαρσης η πανδημία περιορίστηκε σημαντικά με τα αποτελέσματα της να είναι εμφανή σε διάστημα 3 μηνών, με τον ρυθμό αναπαραγωγής να μειώνεται σε $\mathcal{R} \approx 0.272$.

Μερικά από τα ερωτήματα που έχουν μείνει αναπάντητα και θα μπορούσαν να αποτελέσουν αντικείμενο περαιτέρω εργασίας είναι σχετικά με τον έλεγχο μιας πανδημίας. Στο Κεφάλαιο 4 αναφέραμε κάποια από τα μέτρα που έλαβαν οι κυβερνήσεις για την μείωση της έξαρσης της νόσου όπως η κοινωνική απόσταση και τα προληπτικά μέτρα όπως η υγιεινή των χεριών και η χρήση μάσκας προσώπου καθώς και η μείωση του αρχικά ευάλωτου πληθυσμού με την χορήγηση εμβολίων. Εφόσον ο έλεγχος του φαινομένου με τους τρόπους που αναφέραμε δεν συμπεριλήφθηκε στην παρούσα έρευνα, προτείνουμε την επέκταση αυτής για μελλοντική εργασία, λαμβάνοντας υπόψιν τον παράγοντα του ελέγχου εφαρμόζοντας το “Άοσμο” Φίλτρο Κάλμαν (Unscented Kalman Filter).

Βιβλιογραφία

1. Το άρθρο, στο οποίο βασίστηκε η παρούσα εργασία για την συλλογή των δεδομένων, πληροφοριών και σύγκριση των αποτελεσμάτων: PLOS ONE An application of the ensemble Kalman filter in epidemiological modelling. Rajnesh Lal, Weidong Huang, Zhenquan Li.
2. Lobato FS, Libotte GB, Platt GM. Identification of an Epidemiological Model to Simulate the COVID-19 Epidemic Using Robust Multiobjective Optimization and Stochastic Fractal Search. *Computational and Mathematical Methods in Medicine*. 2020;2020.
3. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*. 2020;15(3).
4. Wan K, Chen J, Lu C, Dong L, Wu Z, Zhang L. When will the battle against novel coronavirus end in Wuhan: A SEIR modeling analysis. *Journal of Global Health*. 2020;10(1). pmid:32257174
5. Ndaïrou F, Area I, Nieto JJ, Torres DF. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*. 2020;135:109846.
6. Sun D, Duan L, Xiong J, Wang D. Modeling and forecasting the spread tendency of the COVID-19 in China. *Advances in Difference Equations*. 2020;2020(1):1–16.
7. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020;368(6490):489–493. pmid:32179701
8. Ifguis O, El Ghoulani M, Ammou F, Moutcine A, Abdellah Z. Simulation of the Final Size of the Evolution Curve of Coronavirus Epidemic in Morocco using the SIR Model. *Journal of Environmental and Public Health*. 2020;2020.
9. Loli Piccolomini E, Zama F. Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS One*. 2020;15(8):e0237417.
10. Korolev I. Identification and Estimation of the SEIRD Epidemic Model for COVID-19. *Journal of Econometrics*. 2021;220(1):63–85.
11. Libotte GB, Lobato FS, Platt GM, Neto AJS. Determination of an optimal control strategy for vaccine administration in COVID-19 pandemic treatment. *Computer Methods and Programs in Biomedicine*. 2020;196:105664.
12. Yang Q, Yi C, Vajdi A, Cohnstaedt LW, Wu H, Guo X, et al. Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei province, China. *Infectious Disease Modelling*. 2020;5:563–574. pmid:32835146
13. Engbert R, Rabe MM, Kliegl R, Reich S. Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics. *Bulletin of Mathematical Biology*. 2021;83(1):1–16.
14. Arroyo-Marioli Francisco, Bullano Francisco, Kucinskis Simas, and Rondón-Moreno Carlos. Tracking r of COVID-19: A new real-time estimation using the Kalman filter. *PLoS One*, 16(1):e0244474, 2021. pmid:33439880
15. Ghostine Rabih, Gharamti Mohamad, Hassrouny Sally, and Hoteit Ibrahim. An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble Kalman filter. *Mathematics*, 9(6):636, 2021
16. Extended Kalman filter From Wikipedia, the free encyclopedia

17. ΦΙΛΤΡΑ KALMAN KAI RANDOM WALK SYSTEM από masciences.blogspot.com
18. Compartmental models in epidemiology From Wikipedia, the free encyclopedia
19. Infectious Disease Modelling: Beyond the Basic SIR Model Published in *Towards Data Science*
20. Numerical examples of Kalman Filter by www.KalmanFilter.NET
21. Ross R (1 February 1916). "An application of the theory of probabilities to the study of a priori pathometry.—Part I". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 92 (638): 204–230
22. Ross R, Hudson H (3 May 1917). "An application of the theory of probabilities to the study of a priori pathometry.—Part II". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 93 (650): 212–225.
23. Ross R, Hudson H (1917). "An application of the theory of probabilities to the study of a priori pathometry.—Part III". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 89 (621): 225–240
24. Kermack WO, McKendrick AG (1927). "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*.

Παράρτημα

Παρακάτω φαίνονται τμήματα του κώδικα που γράφηκαν στα πλαίσια της εργασίας για το Εκτεταμένο Φίλτρο Kalman σε γλώσσα προγραμματισμού Python στο προγραμματιστικό περιβάλλον Visual Studio Code.

```

y=np.zeros((3,1)) #the desired values for observation are (I,R,D)
y[0][0]=I[0]
y[1][0]=R[0]
y[2][0]=D[0]

xhat=np.zeros(13) #Matrix of estimated values
X=np.zeros(13)
P=np.eye(13) #Covariance matrix.Since our initial state vector is a guess, we will set a very high estimate uncertainty.
P=P*100 #The high estimate uncertainty results in a high Kalman Gain, giving a high weight to the measurement.
K=np.zeros((13,3)) #Kalman Gain matrix

C=np.zeros((3,13)) #Output matrix
C[0][1]=1
C[1][2]=1
C[2][3]=1
CT=C.transpose()
Q=np.zeros((13,13))
measurement_noise=np.zeros((3,))
process_noise=np.zeros((13,))

xhat=[s[0],i[0],r[0],d[0],b0[0],b1[0],tb[0],g0[0],g1[0],tg[0],d0[0],d1[0],td[0]] #matrix of estimated values
X=[S[0],I[0],R[0],D[0],B0[0],B1[0],TB[0],G0[0],G1[0],TG[0],D0[0],D1[0],TD[0]] #matrix of simulated values

xhat=np.array(xhat)
y=np.array(y)
X=np.array(X)
A=np.array((13,13))
II=np.eye(13) #Identity matrix (13*13)
tn=np.zeros(13)

t1=6 #time of lockdown

```

Solve the SIRD problem

```

def SIRD(z,N,t):
    if t <= t1:
        dSdt=-(z[4]*z[1]*z[0])/N
        dIdt=(z[4]*z[1]*z[0])/N -(z[7]+(z[8]/(1+np.exp(z[9]-t))))*z[1]-z[10]*z[1]
        dRdt=(z[7]+(z[8]/(1+np.exp(z[9]-t))))*z[1]
        dDdt=z[10]*z[1]
    else:
        dSdt=-(z[4]*np.exp((t1-t)/z[6])+z[5])*z[1]*z[0]/N
        dIdt=(z[4]*np.exp((t1-t)/z[6])+z[5])*z[1]*z[0]/N-((z[7]+(z[8]/(1+np.exp(z[9]-t))))*z[1])-(z[1]*(z[10]*np.exp((t1-t)/z[12])+z[11]))
        dRdt=(z[7]+(z[8]/(1+np.exp(z[9]-t))))*z[1]
        dDdt=z[1]*(z[10]*np.exp((t1-t)/z[12])+z[11])

    j=[dSdt,dIdt,dRdt,dDdt,0,0,0,0,0,0,0,0,0]
    return j

```

Noise of theta matrix

```

def theta_noise (r):
    for j in range (4,13):
        norm=np.random.normal(0)
        tn[j]=0.0001*r[j]*norm

    p=[0,0,0,0,tn[4],tn[5],tn[6],tn[7],tn[8],tn[9],tn[10],tn[11],tn[12]]
    return p

```

Main Program

```

for t in range(82):
    Q[0][0],Q[0][1],Q[0][2],Q[0][3]=0.0001*S[t]*S[t],0,0,0
    Q[1][0],Q[1][1],Q[1][2],Q[1][3]=0,0.001*I[t]*I[t],0,0
    Q[2][0],Q[2][1],Q[2][2],Q[2][3]=0,0,0.01*R[t]*R[t],0
    Q[3][0],Q[3][1],Q[3][2],Q[3][3]=0,0,0,0.01*D[t]*D[t]

    process_noise[0]=np.random.normal(0,np.sqrt(Q[0][0])) #noise of S
    process_noise[1]=np.random.normal(0,np.sqrt(Q[1][1])) #noise of I
    process_noise[2]=np.random.normal(0,np.sqrt(Q[2][2])) #noise of R
    process_noise[3]=np.random.normal(0,np.sqrt(Q[3][3])) #noise of D

    RR=np.diag((0.001*I[t]*I[t],0.001*R[t]*R[t],0.001*D[t]*D[t]))
    measurement_noise[0]=np.random.normal(0,np.sqrt(RR[0][0])) #noise of I
    measurement_noise[1]=np.random.normal(0,np.sqrt(RR[1][1])) #noise of R
    measurement_noise[2]=np.random.normal(0,np.sqrt(RR[2][2])) #noise of D

    dif=SIRD(X,t) #here we call the SIRD function to solve the SIRD model for simulated state
    X=X+dif
    y=C*(X-dif)+measurement_noise

    S[t+1]=N-X[1]-X[2]-X[3]
    I[t+1],R[t+1],D[t+1],B0[t+1],B1[t+1],TB[t+1]=X[1],X[2],X[3],X[4],X[5],X[6] #Update the simulated values matrix
    G0[t+1],G1[t+1],TG[t+1],D0[t+1],D1[t+1],TD[t+1]=X[7],X[8],X[9],X[10],X[11],X[12]

```

```

AT=A.transpose()
P=A@P@AT+Q #this is the prior estimate uncertainty (covariance) matrix of the current state (predicted at the previous state)
K=(P@CT)@np.linalg.inv(C@P@CT+RR) #Kalman Gain
TT=I-K@C
TT=TT.transpose()
KT=K.transpose()
P=(I-K@C)@P@TT+K@RR@KT #this is the estimate uncertainty (covariance) matrix of the current state

diff=SIRD(xhat,n,t)
t_d=theta_noise(xhat)

diff=np.array(diff)
t_d=np.array(t_d)

xhat=(xhat+diff+t_d)+K@(y-C*xhat) #estimate the new values

s[t+1]=n-xhat[1]-xhat[2]-xhat[3] #update all of the matrices (estimated)
i[t+1],r[t+1],d[t+1],b0[t+1],b1[t+1],tb[t+1]=xhat[1],xhat[2],xhat[3],xhat[4],xhat[5],xhat[6]
g0[t+1],g1[t+1],tg[t+1],d0[t+1],d1[t+1],td[t+1]=xhat[7],xhat[8],xhat[9],xhat[10],xhat[11],xhat[12]

```